



**UNIVERSIDAD
TORCUATO DI TELLA**

Master in Management + Analytics

**Predicción de arribos y programación de turnos
en una estación de servicio**

Por

Octavio Eneas Biasoli

Tutor: **Javier Marengo, PhD**

Diciembre, 2020

RESUMEN

La programación de turnos en estaciones de servicio consiste en la asignación de horarios de trabajo rotativos al personal de manera tal que pueda atenderse al público de la mejor manera posible. Ésta es una tarea que influye directamente en los costos y en la calidad de atención de las estaciones de servicio, ya que determina el tamaño de la nómina de empleados de la empresa y la cantidad de trabajadores que van a estar disponibles para atender al público a cada hora. Sin embargo, no se le suele dar la importancia que debería tener, ya que su realización suele ser asignada a un encargado de personal que los lleva a cabo de manera manual y sin ningún tipo de análisis.

Esta situación genera que haya empresas con una cantidad de empleados excesiva e innecesaria, los cuales, a la vez, suelen estar mal asignados, debido a que sobran en algunos días y después faltan en otros con mayor demanda. Por consiguiente, muchos empresarios se encuentran pagando en este momento sueldos de trabajadores que no necesitan, perdiendo clientes por brindarles una mala atención y/o padeciendo las consecuencias de un mal clima laboral dado el malestar que genera en los empleados la imprevisibilidad de las correcciones constantes de horarios.

La pregunta principal de este trabajo consiste en cómo predecir la demanda y optimizar los turnos de trabajo en una estación de servicio, haciendo especial foco en las implicancias que esto tiene sobre el tiempo de espera del cliente, la utilización del sistema de atención al público y cantidad de personal necesario para atender correctamente.

Para ello, se ha tomado como objeto de estudio una estación de servicio ubicada en San Justo, provincia de Buenos Aires, para comparar su situación actual, realizando horarios de manera manual, con la programación de horarios que se propone en este trabajo. Para llevarla a cabo, se ha realizado un análisis de los datos históricos de la empresa basado en la teoría de colas, se han utilizado dos de los algoritmos de *machine learning* más potentes de la actualidad, y se ha empleado un *solver* de gran capacidad que permitió armar los horarios de manera óptima, los cuales se encuentran sujetos a determinadas restricciones legales y de la industria.

Palabras Clave: teoría de colas – predicción de arribos – *machine learning* – programación de horarios – optimización – estación de servicio – tiempo de espera – utilización del sistema – cantidad de personal.

ABSTRACT

Shift scheduling at gas stations consists of assigning rotating work hours to staff in such a way that the customers are always properly served and handled. This is a task that directly impacts on the cost and quality of the service provided by these businesses since it defines the quantity of employees on payroll and the amount of employees that are available at any given moment. Nonetheless, it is not handled with the importance that it should, because its crafting is usually given to a manager who makes them manually and with no analysis whatsoever.

Therefore many businesses are unnecessarily overstaffed, and sometimes with uneven shifts, too crowded or understaffed. As a result, many entrepreneurs find themselves paying employees they do not need, losing customers due to poor service and/or dealing with the consequences of a bad work environment on account of the discomfort that this shifts mismanagement generates to the staff.

The main question of this dissertation is how to predict demand and optimize work shifts in a gas station, with special focus on the implications that this has on customer waiting time, the use of the customer service system and number of staff needed to serve properly.

In order to answer this question, a gas station located in San Justo, Buenos Aires, was taken as the object of study to compare its current situation, in which the shifts were made manually, with the schedule proposed in this dissertation. To carry it out, an analysis of the historical data of the company based on queuing theory has been carried out, two of the currently most powerful machine learning algorithms have been used, and a high-performance solver has been used that allowed to set the schedules in an optimal way, which are subject to certain legal and industry restrictions.

Keywords: queueing theory – arrival prediction – machine learning – shifts scheduling – optimization – gas station – waiting time – system use – number of staff.

INDICE

RESUMEN	ii
ABSTRACT	iii
INTRODUCCION.....	6
LITERATURA RELACIONADA	8
ANALYTICS DESCRIPTIVO.....	13
CAPITULO 1: PREPARACION Y ANALISIS DE DATOS HISTORICOS	13
1.1 Análisis exploratorio.....	13
1.2 Análisis teoría de colas	15
1.3 Datos de operaciones (línea base)	19
ANALYTICS PREDICTIVO	26
CAPITULO 2: PREDICCION DE ARRIBOS.....	26
2.1 Estructura de los datos.....	26
2.2 Análisis exploratorio.....	27
2.3 Conjunto de entrenamiento, validación y testeo.....	33
2.4 Modelos	34
2.5 Métrica de evaluación.....	34
2.6 Ingeniería de atributos	35
2.7 Optimización de hiperparámetros.....	36
2.8 Resultados.....	44
ANALYTICS PRESCRIPTIVO.....	49
CAPITULO 3: PROGRAMACION DE HORARIOS	49
3.1 Modelado y técnica de resolución	49
3.2 Modelo original	52
3.3 Modelo ajustado	58
3.4 Modelo relajado.....	64

3.5 Propuesta de solución	68
3.6 Análisis de resultados	69
CONCLUSIONES.....	74
BIBLIOGRAFIA	76
ANEXO I: CODIGO ZIMPL (MODELO ORIGINAL)	78
ANEXO II: CODIGO EN R (<i>GRADIENT BOOSTING MACHINE</i>).....	80

INTRODUCCION

La programación de horarios (*scheduling* o *timetabling* en inglés) consiste en el proceso de construir un calendario de trabajo para el personal de una organización, de manera tal que la misma pueda satisfacer la demanda para sus bienes o servicios. Dada su complejidad, la programación de horarios es objeto de estudio de la investigación operativa, más específicamente de la optimización combinatoria, ya que el armado de éste implica tomar un conjunto de decisiones que se impactan entre sí para optimizar un objetivo determinado.

La gestión de horarios rotativos en estaciones de servicio suele ser un tema al que no se le da la importancia que debería tener, a pesar de que es una tarea que influye directamente en los costos y calidad de atención de la empresa. En general, la misma es delegada a un encargado de personal que resuelve la problemática de manera manual y sin ningún tipo de análisis. Los horarios generalmente son asignados por intuición, y usos y costumbres del rubro, sin tener en cuenta la real afluencia de clientes ni la cantidad de personal necesario para atender dicha demanda. De esta manera, suelen ocurrir días en donde hay más empleados trabajando que los necesarios, que luego faltan en días con mayor movimiento. O incluso, suceden días en donde, por algún error humano, falta personal para cubrir los puestos mínimos de trabajo, los cuales rápidamente hay que completar haciendo una llamada de último momento.

Estos problemas de gestión generan costos adicionales que no suelen ser identificados por los empresarios, haciendo que erróneamente releguen el tema en cuestión. Por ello, muchos de ellos se encuentran en este momento pagando sueldos de empleados que no necesitan, perdiendo clientes por brindarles una mala atención y/o padeciendo las consecuencias de un mal clima laboral debido al malestar que la imprevisibilidad genera en los mismos empleados.

Por consiguiente, la pregunta principal de este trabajo consiste en cómo predecir la demanda y optimizar los turnos de trabajo en una estación de servicio, haciendo especial foco en las implicancias que esto tiene sobre el tiempo de espera del cliente, la utilización del sistema de atención al público y cantidad de personal necesario para atender correctamente. El propósito de este trabajo es desarrollar una herramienta de planificación de horarios de personal para una estación de servicio ubicada en San Justo, provincia de

Buenos Aires, utilizando el *solver*¹ SCIP² para ello, de manera tal que pueda asegurarse una óptima calidad de atención al público con la menor cantidad de personal posible. Se busca que el modelo planteado respete una serie de restricciones *hard*, en función de lo establecido por la Ley de Contrato de Trabajo y el Convenio Colectivo de Empleados de Estaciones de Servicio vigente, a la vez que se le sugiere que satisfaga una serie de restricciones *soft*, ligadas a las buenas prácticas empresariales.

Dicha herramienta será un prototipo, y servirá como prueba de concepto para cuantificar la mejora en el tiempo de espera en cola, la mejor utilización del sistema evitando saturaciones, y el ahorro en sueldos de personal, si existiera.

El presente documento se estructura en tres capítulos. En el primer capítulo se realiza un análisis de los datos históricos de la empresa, el cual consta de dos partes. Por un lado, se lleva a cabo un estudio basado en la teoría de colas, para determinar el modelo aplicable a la estación de servicio y utilizarlo para calcular el tiempo de espera de los clientes, a la vez que se determina la tasa de servicio³ de la empresa. Por el otro, se utiliza esta información para realizar un análisis de los datos de operaciones de la empresa, el cual será utilizado como línea base para comparar con los modelos de optimización realizados en este trabajo. En el segundo capítulo se desarrolla la predicción de arribos, justamente para poder definir la tasa de arribos⁴ a cada hora, e incorporarla a los modelos de optimización. Para ello, se utilizan dos de los algoritmos más poderosos de la actualidad, *Random Forest* y *Gradient Boosting Machine*, seleccionando a aquel que mejor rendimiento obtuvo para tomar sus datos. Por último, en el tercer capítulo, se efectúa la programación de diversos horarios, tomando al mejor de ellos como propuesta final, y se lo compara con el modelo manual realizado por la estación de servicio.

¹ Un *solver* es una herramienta que permite resolver un problema matemático, existiendo diversas alternativas en el mercado, tanto comerciales como no comerciales.

² SCIP es uno de los *solver* no comerciales más rápidos de la actualidad. Para más información puede consultarse <https://www.scipopt.org/>.

³ La tasa de servicio es la cantidad de autos por hora que puede atender un servidor.

⁴ La tasa de arribos es la cantidad de clientes por hora que llegan al establecimiento.

LITERATURA RELACIONADA

La planificación de horarios de personal es un tema que ha sido muy estudiado durante las últimas décadas por los académicos especializados en el área de operaciones. En este apartado, se traen a colación una serie de trabajos realizados por estudiantes y profesionales de todo el mundo, que contribuyen desde diferentes ángulos al desarrollo de esta temática.

En el año 1974 se publicó *Scheduling Full-Time and Part-Time Staff to Meet Cyclic Requirements* escrito por Kenneth, que fue uno de los artículos más antiguos que se encontró al respecto. El objeto de estudio en este *paper* es una empresa que trabaja siete días a la semana, planteándose como objetivo determinar la cantidad de personal necesario en cada día de la semana para satisfacer la demanda con la menor cantidad de personal posible.

Para ello, se plantean las restricciones del caso, siendo una de las más relevantes el hecho de exigirle al modelo que fije dos días de descanso consecutivos para los empleados, y se lleva a cabo la resolución, la cual es manual. Justamente, una de las características más distintivas de este documento es que presenta una serie de pasos bajo los cuales cualquiera podría llevar a cabo la planificación de su personal, aunque quedando sujeto al modelado planteado por el autor, el cual es muy específico al caso planteado.

En el año 1997, Beaumont publica *Scheduling staff using mixed integer programming*. Un trabajo ya de mayor complejidad que el anterior, con el objetivo de determinar la cantidad de personal necesario en cada turno de cada día de una empresa de servicios australiana que trabaja las 24 horas del día los 365 días del año.

Para ello, se recolectaron datos de las llamadas telefónicas recibidas desde 01/10/1989 hasta 30/09/1990 inclusive, para estimar la demanda promedio en cada uno de los períodos en los cuales se iba a llevar a cabo la planificación del personal. Además, se determinó arbitrariamente la tasa de servicio de los empleados, la cual se estableció en tres llamadas por hora, y se calculó tanto el costo de cada empleado como el costo de hacer esperar a los clientes.

Junto con una serie de restricciones, siendo la más curiosa para comentar la exigencia al modelo que no cree más de cinco horarios rotativos distintos para administrar, se planteó la función objetivo que busca minimizar la sumatoria entre el costo de personal, el costo de transporte de éste y el costo de hacer esperar a los clientes. De esta manera, se

consiguió una reducción de los costos totales en un 17% con relación a la situación anterior que se encontraba la empresa.

En el mismo año, Thornton y Sattar publican *Nurse Rostering and Integer Programming Revisited*, trabajo cuyo objeto de estudio es un hospital en Costa de Oro, Australia. La motivación de este artículo se derivó de la importancia que tienen las enfermeras para un hospital, ya que no solo son el costo más importante dentro de la estructura de costos de la empresa, sino que también influyen directamente sobre el cuidado y la atención de los pacientes. Mantener a las enfermeras satisfechas implica tener una menor rotación de personal calificado, el cual es escaso en este rubro, y una mejor calidad de atención al público.

Con el objetivo de maximizar la satisfacción general del personal, se llevó a cabo una asignación de turnos para una empresa que trabaja las 24 horas los 365 días del año, pero teniendo en cuenta la particularidad de que se le pidió a las enfermeras que establezcan un orden de preferencia sobre las distintas combinaciones de horarios posibles.

Asimismo, se crearon dos tipos de restricciones. Por un lado, restricciones *soft*, las cuales pueden ser violadas por el modelo, pero sufriendo una penalidad. Algunas de ellas son: cantidad máxima y mínima de turnos sin días de descanso, preferencia por turno mañana, tarde o noche y cantidad máxima y mínima de turnos por semana. Por otro lado, se delimitaron restricciones *hard*, las cuales no pueden ser violadas por el modelo. La más relevante de ellas es la cantidad máxima y mínima de personal que debe haber para cada turno.

En función de estos datos, se formuló una función objetivo que cumpla con las restricciones establecidas, pero buscando minimizar la insatisfacción general o, dicho de otra manera, maximizando la satisfacción de las enfermeras y la calidad de atención al público.

En el año 2012, Mudra publica *Staff Scheduling in Health Care Systems*, que es un texto con lineamientos similares al anterior pero que toma como objeto de estudio el sistema de salud en general. “Mientras que en otras áreas podemos aceptar horarios de menor calidad, la demanda sobre la calidad de los horarios del personal médico es inflexible” (Mudra, 2012).

En su trabajo, Mudra planifica un horario de cuatro semanas para un hospital que trabaja 24 horas los 365 días del año. Las restricciones *hard* más relevantes que utiliza para ello

son: cantidad fija determinada de empleados para cada día, cantidad mínima de días de trabajo por empleado por semana, cantidad máxima de días de trabajo corridos por empleado y tiempo mínimo entre turnos. Por otro lado, aplica una serie de restricciones *soft* dadas por el personal, como por ejemplo agrupar los días de trabajo y balancear la cantidad de descansos en días de semana y fines de semana.

En función de esto se dispone una función objetivo que minimice la penalidad total, al igual que en el trabajo de Thornton y Sattar. La intención es maximizar los requerimientos del personal médico para evitar la rotación de éste, y optimizar la calidad de atención a los pacientes.

En el año 2020 Aydas O., Ross, Scanlon y Aydas B. publicaron *New results on integrated nurse staffing and scheduling: The medium-term context for intensive care units*, cuyo objeto de estudio son las unidades de terapia intensiva en un hospital en Estados Unidos, las cuales trabajan 24 horas los 365 días del año. El objetivo de este *paper* es reducir los costos de personal a la vez que se balancean los riesgos de asignar empleados de más o de menos a cada turno.

Para ello, este artículo presenta la novedad de realizar un pronóstico de demanda entrenando distintos modelos de regresión a partir de datos históricos del hospital, cuyos resultados posteriormente se utilizan como *input* en el modelo de optimización. Asimismo, se fijaron una serie de restricciones, siendo las más notables el tiempo de descanso mínimo entre asignaciones para una misma persona, la cantidad máxima de horas semanales de trabajo, asignación rotativa en fines de semana y horarios fijos para algunos empleados. En función de estos datos, la función objetivo de este trabajo busca minimizar la cantidad de personal necesario en cada turno.

Cada uno de los artículos mencionados anteriormente presentan similitudes y diferencias con el presente trabajo, las cuales se detallan a continuación.

A pesar de realizar un postulado mucho más sencillo desde el punto de vista computacional, el documento publicado por Kenneth coincide en el fraccionamiento semanal de siete días bajo el cual se planifican los horarios y en el hecho de buscar satisfacer la demanda con la menor cantidad de personal posible.

Por su parte, Beaumont incorpora un elemento clave para este trabajo que es el análisis histórico de la demanda. Asimismo, considera más de un turno diario para realizar la asignación de personal, la cual la realiza para los 365 días del año, y posteriormente brinda

la herramienta creada a la empresa para que pueda seguir utilizándola sin ayuda de terceros. Todos elementos en común con este trabajo.

Sin embargo, la función objetivo que presenta en su artículo busca minimizar la sumatoria entre el costo del personal, el costo de transporte de éste y el costo de hacer esperar a los clientes, lo cual es una evidente diferencia. Lo que se busca en el caso de la estación de servicio es minimizar el costo del personal, pero sujeto a un umbral de atención considerado como óptimo, el cual implica tener permanentemente el sistema por debajo de una utilización del 80%. Determinar el costo de hacer esperar a los clientes no solo es realmente muy complejo, sino que atenta contra la propia empresa al tomar bajo consideración la posibilidad de atender peor dado que es más barato que tomar más personal.

Thornton y Sattar van más allá a la hora de modelar del problema e incluyen restricciones *soft*, mediante las cuales le sugieren al modelo que cumpla con determinados requisitos, padeciendo una penalidad en caso de no hacerlo. En su artículo, comentan que crean horarios con preferencia turno mañana, turno tarde y turno noche, en lugar de horarios que roten continuamente, para poder así satisfacer las necesidades del personal médico. Además, le piden al modelo que tenga en cuenta la cantidad mínima y máxima de turnos sin descanso, para evitar que se junten muchos días de trabajo sin un solo día libre. Lógicamente, estas consideraciones también son importantes para la planificación de horarios de la estación de servicio.

Sin embargo, con estos autores existe una diferencia importante en cuanto a la función objetivo. El presente trabajo busca minimizar la cantidad de personal a la vez que se brinda una atención de calidad y rápida, estando el foco puesto en la satisfacción del cliente, y no en la satisfacción del personal, tal como plantean Thornton y Sattar.

Mismos comentarios se tienen con relación al documento de Mudra.

Por último, el artículo de Aydas O., Ross, Scanlon y Aydas B. incorpora cuestiones que también están presentes en el modelo planteado para la estación de servicio. La realización de un pronóstico de la demanda mediante un modelo de regresión que tome como *input* datos históricos de la empresa, es justamente uno de los pilares de este trabajo. Asimismo, la función objetivo persigue la minimización de la cantidad de personal necesario para satisfacer las restricciones impuestas.

No obstante, la diferencia que se presenta radica en cuanto al objeto de estudio de cada uno de los trabajos. Mientras que estos autores se enfocaron en las unidades de terapia intensiva de un hospital en Estados Unidos, en este trabajo el objeto de estudio es una estación de servicio Shell en Argentina. Se trata de dos industrias con algunos aspectos en común, dado que ambas prestan servicios, pero en el detalle presentan características muy distintas que deben ser tenidas en cuenta en cada uno de los modelos de optimización planteados.

ANALYTICS DESCRIPTIVO

CAPITULO 1: PREPARACION Y ANALISIS DE DATOS HISTORICOS

1.1 Análisis exploratorio

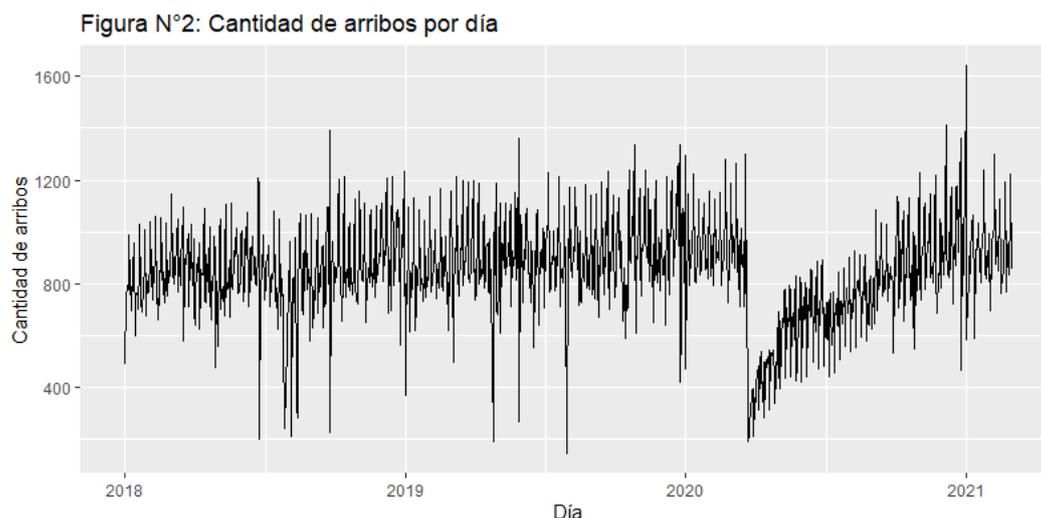
Se cuenta con el detalle de los despachos realizados por el establecimiento desde el 01/01/2018 hasta el 28/02/2021, los cuales representan los arribos de los clientes a cada hora. La Figura N°1 da una pequeña muestra de la información que el sistema de la estación de servicio arroja.

Figura N°1: Información almacenada por el sistema de la estación de servicio

FECHA	HORA	MANGUERA	PRODUCTO	LITROS	PRECIO	TOTAL
1/1/2018	06:47:55	10	VP DIESEL	3.96	25.28	100.08
1/1/2018	06:48:18	8	SUPER	50.76	23.65	1,200.40
1/1/2018	06:50:17	2	VPOWER	11.03	27.21	299.99
1/1/2018	06:56:34	14	VPOWER	18.38	27.21	499.98
1/1/2018	07:03:48	8	SUPER	8.46	23.65	199.98
1/1/2018	07:04:59	2	VPOWER	23.89	27.21	649.97
1/1/2018	07:20:35	16	SUPER	12.68	23.65	299.98
1/1/2018	07:27:11	9	F. DIESEL	14.15	21.30	301.29
1/1/2018	07:42:44	16	SUPER	26.64	23.65	630.11
1/1/2018	07:52:58	7	VP DIESEL	11.87	25.28	300.00

Utilizando RStudio, y haciendo especial foco en datos como la fecha, la hora, la manguera utilizada y la cantidad de litros despachados, se llevó a cabo el análisis exploratorio y de teoría de colas que se presenta a continuación.

Para tener un primer contacto con la cantidad arribos diarios de los clientes, y si los mismos tienen un comportamiento que, a primera vista, pueda ser útil para predecir los arribos de los próximos meses, se confeccionó la Figura N°2. En ella se puede apreciar una cantidad de arribos ligeramente creciente desde el 2018 hasta principios del 2020 cuando, por efecto de la cuarentena implementada por Argentina, se produce un desplome abismal en los despachos realizados. Con el correr de los meses se contempla un crecimiento acelerado de los mismos hasta alcanzar valores normales a partir de septiembre de ese año.



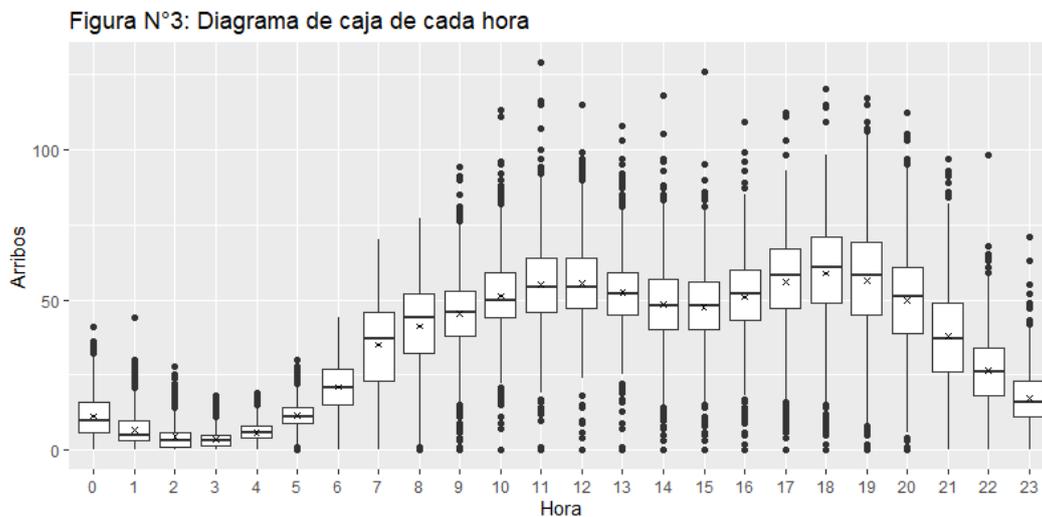
En este sentido, para comenzar a tomar conocimiento sobre la distribución de los arribos a cada hora, se realizó el diagrama de caja que se puede ver en la Figura N°3, en donde rápidamente se puede sacar una serie de conclusiones sobre los datos.

Desde las 00:00 horas hasta las 05:00 horas inclusive, los arribos promedio no superan los 20 autos por hora, los cuales toman una carrera alcista a partir de las 06:00 horas, donde la demanda crece rápidamente hasta alcanzar un pico a las 12:00 horas, con una cantidad de 56 autos por hora en promedio. Luego, la misma desciende levemente hasta las 15:00 horas a un promedio de 48 autos por hora, para volver a tomar una tendencia alcista que toca su segundo pico del día a las 18:00 horas, con una demanda promedio de 59 autos.

Además, se contempla que la cantidad de arribos que pueden suceder a cada hora se encuentra entre rangos muy amplios, lo que dificulta la tarea de planificación de horarios del personal. Por ejemplo, la diferencia entre el primer cuartil y el tercer cuartil de los arribos a las 19:00 horas es de 24 autos, lo que haría imposible atender al público correctamente en caso de errar en la planificación. Y eso por no mencionar la diferencia entre el valor mínimo y máximo de este horario, el cual es de 117 autos.

Por ello, es fundamental tener en cuenta que el mes que se esté evaluando, el momento del mes en que uno se encuentre, el día de la semana que se trate, el número de semana del año, la existencia de feriados, la cuarentena, la temperatura, las precipitaciones, entre otras causales, hacen que la cantidad de arribos a cada hora tome valores tan distintos como se ve en el gráfico a continuación.

Esto acompaña la idea que se propone en este trabajo, que básicamente consiste en profesionalizar la planificación de horarios en este tipo de negocios. Es imperiosa la necesidad de contar con una herramienta que pueda contemplar todas estas variables para evitar errores humanos que puedan perjudicar a la empresa, tanto por contratar personal en exceso para satisfacer la demanda, como por asignarlo incorrectamente a los turnos de trabajo.



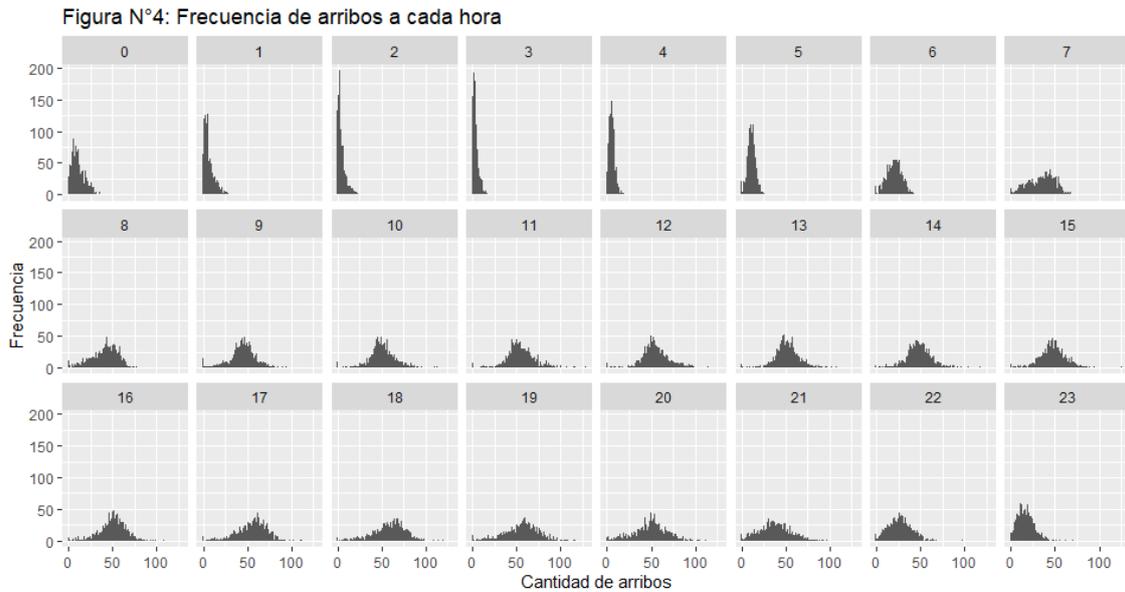
1.2 Análisis teoría de colas

La teoría de colas consiste en el estudio matemático de las filas de espera dentro de un sistema. Se entiende como fila de espera a una hilera formada por uno o varios clientes que esperan recibir un servicio, la cual se genera debido a un desequilibrio temporal entre la demanda de un servicio y la capacidad del sistema para suministrarlo. Existen diversos modelos de filas de espera para analizar las operaciones de una empresa, siendo uno de los más comunes el modelo M/M/S. Para poder ser utilizado, dicho modelo requiere de un proceso de arribos *Poisson*, tiempos de servicio exponenciales y disciplina PEPS (Primero Entrar, Primero Salir), que consiste en que el primer cliente que llega al establecimiento sea el primero en ser atendido.

Este último requisito es sencillo de validar, ya que los clientes se van ordenando en filas en función de su orden de llegada, y es respetando ese orden la manera en que se los atiende en la estación de servicio. No se aplica ningún método de selección aleatoria de los clientes, sino que el primero que llega es el primero que se retira.

Ahora, para determinar si los restantes datos con los que se cuenta aplican a un modelo M/M/S dentro de la teoría de colas, se calculó la frecuencia de arribos a cada hora para

concluir gráficamente si los mismos siguen una distribución *Poisson*. En la Figura N°4 se puede observar la frecuencia de arribos a cada hora, y rápidamente se puede dilucidar que, al menos en la mayor parte del día, no se sigue una distribución *Poisson*.



Para validar esto numéricamente, se computó el tiempo entre arribos de los datos de llegada, y en función de cada hora se cuantificó el promedio y el desvío estándar, pudiendo determinar de esta manera el coeficiente de variación de cada hora. Si el coeficiente de variación de alguna de ellas fuera igual a uno, se podría concluir que en ese momento del día los arribos siguen una distribución *Poisson*.

A partir de la Figura N°5 se puede concluir que en ningún momento del día los arribos siguen una distribución *Poisson*, ya que los tiempos entre arribos no siguen una distribución exponencial. Esto queda confirmado al ver que ninguno de los coeficientes de variación calculados es igual a uno, aunque sí es cierto que, tal como se apreciaba gráficamente, a las 11 horas y a las 12 horas el proceso de arribos es muy similar a esta distribución buscada. A pesar de ello, no alcanza para validar el modelo desde el punto de vista de los arribos.

Figura N°5: Tiempo entre arribos (en minutos)

HORA	PROMEDIO	DESV. EST.	COEF. VAR.
0	4.79	6.01	1.25
1	7.67	10.10	1.32
2	11.16	14.09	1.26
3	14.51	16.66	1.15

4	10.77	12.75	1.18
5	5.45	6.53	1.20
6	2.91	3.58	1.23
7	1.71	2.46	1.44
8	1.45	2.30	1.59
9	1.32	1.89	1.43
10	1.17	1.46	1.25
11	1.08	1.18	1.09
12	1.07	1.12	1.05
13	1.14	1.25	1.10
14	1.23	1.38	1.12
15	1.25	1.48	1.18
16	1.18	1.50	1.27
17	1.07	1.49	1.39
18	1.01	1.44	1.43
19	1.05	1.71	1.63
20	1.19	1.97	1.66
21	1.55	2.70	1.74
22	2.24	3.53	1.57
23	3.33	4.23	1.27

Por otro lado, dado que en la estación de servicio no se poseen datos sobre el tiempo de servicio, se infirió el mismo a partir de los datos de los arribos de la siguiente manera.

En primer lugar, a partir del dato de la manguera utilizada para realizar cada despacho, se ordenaron los mismos en función de la cara de cada surtidor en que fueron realizados. En la Figura N°6 se puede observar el conjunto de mangueras que compone cada cara de cada uno de los surtidores de la empresa.

Figura N°6: *Layout* de la playa de combustibles líquidos



Luego, una vez realizada esta separación, se procedió a calcular el tiempo entre despachos, pero solamente se tuvo en cuenta aquellos que se encontraban dentro de un

tiempo de servicio considerado como normal, y se descartó el resto. Se estimó como tiempo de servicio normal a todo aquel que fuera menor o igual a tres segundos por litro despachado más cinco minutos en concepto de tiempo de cobro. Cualquier tiempo superior a este valor fue descartado por presuntamente incorporar demoras no relacionadas al tiempo de servicio. De esta manera, se arribó al tiempo de servicio por cara.

Ahora, teniendo en cuenta que cada playero puede atender hasta dos caras de un surtidor sin alterar su capacidad de atención, debido a que el surtidor puede despachar combustible sin necesidad del playero, se consideró que el tiempo de servicio es de la mitad del valor obtenido. Esta presunción de que un playero puede atender hasta dos caras de un surtidor, comúnmente realizada en el rubro por usos y costumbres, fue igualmente validada a través de las cámaras de seguridad del establecimiento, en donde se pudo observar que el dato obtenido era un fiel reflejo de la realidad.

A pesar de poder haber realizado dicho análisis a través de una medición de campo, se prefirió llevarlo a cabo a través de la totalidad de los datos históricos de la empresa por considerarse un resultado más robusto, ya que iba a contemplar un promedio “perfecto” de todas las transacciones realizadas por ésta. Esto se debe a que, a la hora de atender a un cliente, el mismo puede encontrarse en una moto, en un auto, en una camioneta, o en un camión, puede elegir pagar con efectivo, con tarjeta de crédito o débito, o con Mercado Pago, puede querer que le revisen los fluidos del vehículo o no, y puede elegir o no utilizar su tarjeta de puntos. Todas estas situaciones generan tiempos de servicio distintos, las cuales deberían ser contempladas en su proporción correcta para que la medición de campo arroje un promedio que represente correctamente la realidad. En este caso, se consideró que los datos históricos ya contaban con ello, por lo que fueron los elegidos para realizar el cálculo.

No obstante, se insiste en que los resultados obtenidos a través de los datos históricos fueron corroborados a través de observaciones realizadas en el establecimiento para determinar su veracidad. Por lo que, a pesar de no haber utilizado la observación de campo para obtener el dato del tiempo de servicio, sí fue utilizada para validar lo calculado.

La Figura N°7 detalla los valores obtenidos para una cara del surtidor y los calculados para el playero, que a la hora de aplicar el modelo de fila de espera va a ser considerado como el servidor del sistema.

Figura N°7: Tiempo de servicio (minutos)

UNIDAD	PROMEDIO	DESV. EST.	COEF. VAR.
Una cara	3.24	1.50	0.46
Dos caras (playero)	1.62	1.06	0.65

Al igual que lo sucedido con el tiempo entre arribos, se rechaza la posibilidad de que el tiempo de servicio siga una distribución exponencial, dado que el coeficiente de variación correspondiente al playero dista de ser 1. De esta manera, queda invalidada la posibilidad de aplicar un modelo M/M/S dentro de la teoría de colas, quedando sujeto a aplicar el modelo aproximado propuesto por Hirotaka Sakasegawa (Sakasegawa, 1977).

Asimismo, se utilizan los datos de la figura anterior para determinar la tasa de servicio de cada servidor, la cual va a ser un *input* clave a la hora de realizar la programación de turnos. La Figura N°8 aclara que por cada servidor que se incorpore al sistema, es decir, por cada playero adicional que se incorpore a la atención al público, se van a poder atender hasta 37.02 autos adicionales.

Figura N°8: Tasa de servicio

UNIDAD	MU (Cl/Min)	MU (Cl/Hs)
Una cara	0.31	18.51
Dos caras (playero)	0.62	37.02

1.3 Datos de operaciones (línea base)

La Figura N°9 detalla la cantidad de personal que la empresa asignó a cada uno de los turnos del día durante el mes de febrero del año 2021.

Figura N°9: Planilla de horarios de la estación de servicio

	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
NOCHE	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
MAÑANA	3	2	2	2	3	3	3	2	2	2	2	2	3	2	2	2	2	2	3	3	2	2	2	2	2	2	3	2	
TARDE	2	2	3	4	2	3	2	2	3	2	3	3	3	3	4	3	4	3	4	3	3	4	3	4	4	4	4	3	3

Por usos y costumbres, las estaciones de servicio suelen utilizar tres horarios distintos. Un turno mañana que se lleva a cabo entre las 06:00 horas y las 14:00 horas, un turno tarde que se efectúa desde las 14:00 horas hasta las 22:00 horas y, por último, un turno noche se lleva a cabo desde las 22:00 horas hasta las 06:00 horas del día siguiente.

En función de los datos relevados en la Figura N°1, Figura N°5, Figura N°7, Figura N°8 y Figura N°9, se creó la tabla que se muestra en la Figura N°10, la cual resume cómo le ha ido a la empresa operando de manera manual en el mes de febrero del año 2021.

En las primeras columnas de la tabla se detalla información con respecto a la fecha y la hora a la que corresponden los arribos que posteriormente se presentan, a la vez que se encuentra el coeficiente de variación del tiempo entre arribos de cada una de las horas del día. Asimismo, la séptima columna indica la cantidad de servidores (playeros) que son utilizados a cada hora, información que fue obtenida de la planificación manual de la empresa. A partir de la cantidad de servidores a cada hora, se procede a calcular la tasa de servicio por minuto y por hora, que será de utilidad para calcular el tiempo de espera en minutos⁵ y la utilización del sistema⁶, respectivamente.

Figura N°10: Análisis planificación de horarios manual

AÑO	MES	DIA	HORA	ARRIBOS		CANTIDAD SERVIDORES	SERVICIO		UTILIZACION	TIEMPO DE ESPERA (MIN)	
				TASA	COEF. VAR.		TASA	POR MINUTO			POR HORA
2021	FEBRERO	1	0	9	1.25	2	1.23	74.04	12%	0.04	
2021	FEBRERO	1	1	10	1.32	2	1.23	74.04	14%	0.06	
2021	FEBRERO	1	2	0	1.26	2	1.23	74.04	0%	0.00	
2021	FEBRERO	1	3	1	1.15	2	1.23	74.04	1%	0.00	
2021	FEBRERO	1	4	4	1.18	2	1.23	74.04	5%	0.01	
2021	FEBRERO	1	5	9	1.20	2	1.23	74.04	12%	0.04	
2021	FEBRERO	1	6	30	1.23	3	1.85	111.06	27%	0.07	
2021	FEBRERO	1	7	28	1.44	3	1.85	111.06	25%	0.07	
2021	FEBRERO	1	8	32	1.59	3	1.85	111.06	29%	0.11	
2021	FEBRERO	1	9	32	1.43	3	1.85	111.06	29%	0.10	
2021	FEBRERO	1	10	48	1.25	3	1.85	111.06	43%	0.20	
2021	FEBRERO	1	11	58	1.09	3	1.85	111.06	52%	0.28	
2021	FEBRERO	1	12	49	1.05	3	1.85	111.06	44%	0.16	
2021	FEBRERO	1	13	60	1.10	3	1.85	111.06	54%	0.31	
2021	FEBRERO	1	14	55	1.12	2	1.23	74.04	74%	1.72	
2021	FEBRERO	1	15	37	1.18	2	1.23	74.04	50%	0.54	
2021	FEBRERO	1	16	53	1.27	2	1.23	74.04	72%	1.80	
2021	FEBRERO	1	17	65	1.39	2	1.23	74.04	88%	6.46	
2021	FEBRERO	1	18	78	1.43	2	1.23	74.04	105%	100.00	
2021	FEBRERO	1	19	54	1.63	2	1.23	74.04	73%	2.91	
2021	FEBRERO	1	20	56	1.66	2	1.23	74.04	76%	3.53	
2021	FEBRERO	1	21	37	1.74	2	1.23	74.04	50%	1.03	
2021	FEBRERO	1	22	20	1.57	2	1.23	74.04	27%	0.24	
2021	FEBRERO	1	23	10	1.27	2	1.23	74.04	14%	0.05	

⁵ Se calcula a partir de la fórmula que plantea Hirotaka Sakasegawa en su modelo aproximado.

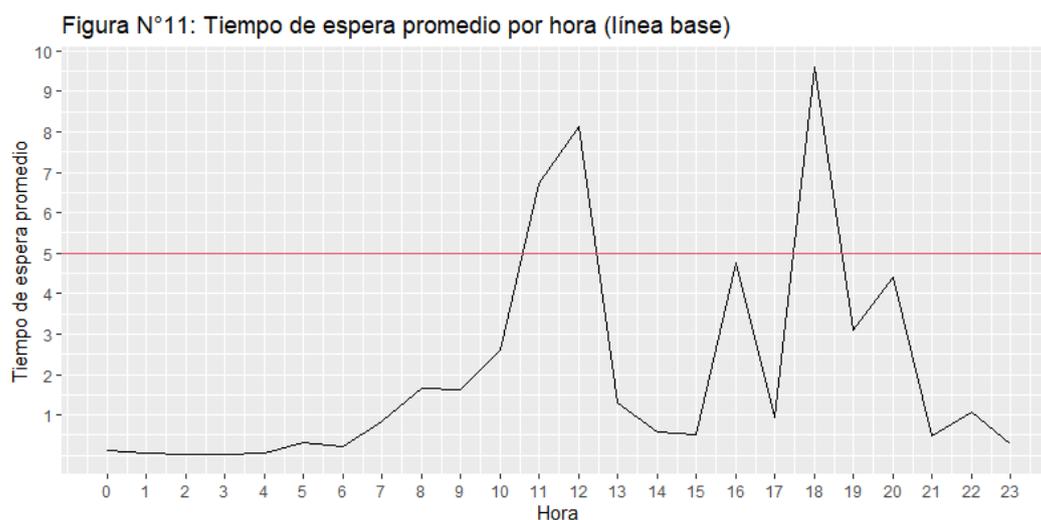
⁶ Consiste en una división entre la tasa de arribos por hora y la tasa de servicio por hora.

El contraste entre la planilla de horarios realizada por la estación de servicio para el mes de febrero del año 2021 y los arribos sucedidos en dicho mes será utilizado para determinar si la herramienta de planificación de horarios propuesta en este trabajo implica una mejora para el establecimiento. Para arribar a una conclusión al respecto, serán analizadas las siguientes métricas:

- Tiempo de espera en cola
- Utilización del sistema
- Cantidad de personal

El tiempo de espera en cola es una métrica de calidad clave para las empresas de servicios, y su correcta gestión se convierte en un activo de gran valor para estas compañías. En el caso de las estaciones de servicio es fundamental prestarle atención, ya que la excesiva demora en el tiempo de atención puede significar la pérdida de un cliente. Todo estacionero⁷ debe tener en cuenta que nadie carga combustible por gusto, sino que más bien por una necesidad, por lo que cuanto menos tarde un cliente en irse del establecimiento más valorará la atención recibida.

En la Figura N°11 se detalla el tiempo de espera en cola promedio en cada uno de los momentos del día a lo largo del mes de febrero del año 2021.



Este gráfico deja en evidencia dos grandes falencias en cuanto a la planificación manual de los horarios del personal.

⁷ En la jerga, suele llamársele estacionero a la persona que es dueña y, generalmente, administradora general de una o más estaciones de servicio.

En primer lugar, se nota que la rigidez de turnos que se utilizan (06:00 – 14:00 horas, 14:00 – 22:00 horas y 22:00 – 06:00 horas) carece de la flexibilidad adecuada para captar los distintos picos de demanda a lo largo del día. Específicamente, las 11:00 horas, 12:00 horas y 18:00 horas son momentos del día que necesitarían especial atención para evitar que el tiempo de espera aumente tanto.

En segundo lugar, más allá de la rigidez de los turnos utilizados, sale a la luz que hay un problema de asignación de personal, ya que en los horarios mencionados anteriormente debería haber un playero adicional para evitar que los tiempos de espera en cola se disparen. Habiendo realizado un análisis de la capacidad de atención de la empresa, puede afirmarse que el cuello de botella es totalmente atribuible a este problema de planificación de horarios más que a una falta de la capacidad instalada del establecimiento. Esto se debe a que la capacidad máxima de la empresa, que consiste en tener trabajando a tres playeros (servidores) de manera simultánea, permite atender el 99.95% de las demandas históricas, o el 99.32% de las demandas históricas si se establece como tope una utilización del 80% de la capacidad máxima.

En el 4.46% de las 672 horas que componen al mes de febrero del año 2021 el tiempo de espera fue superior a cinco minutos, lo que significa que en una hora cada 22 horas la estación de servicio brindó una calidad de atención al público que se encuentra por debajo de los estándares de la industria. Por otro lado, el tiempo de espera promedio a lo largo del mes fue de 1.91 minutos.

Teniendo en cuenta estos indicadores, posiblemente haya más de un cliente pensando en cambiar de estación de servicio la próxima vez que necesite cargar combustible, por lo que es imperioso profesionalizar esta manera de planificar los horarios. Para llevar a cabo esta profesionalización, es necesario conocer cómo se genera el tiempo de espera en cola, y uno de los factores que más influye es el nivel de utilización del sistema.

El nivel de utilización del sistema es una métrica muy importante dentro de la teoría de colas, ya que rápidamente permite dilucidar el nivel de saturación de éste. Es fundamental mantenerla siempre en valores menores al 80%, ya que una vez superado este umbral el impacto sobre el tiempo de espera es muy significativo.

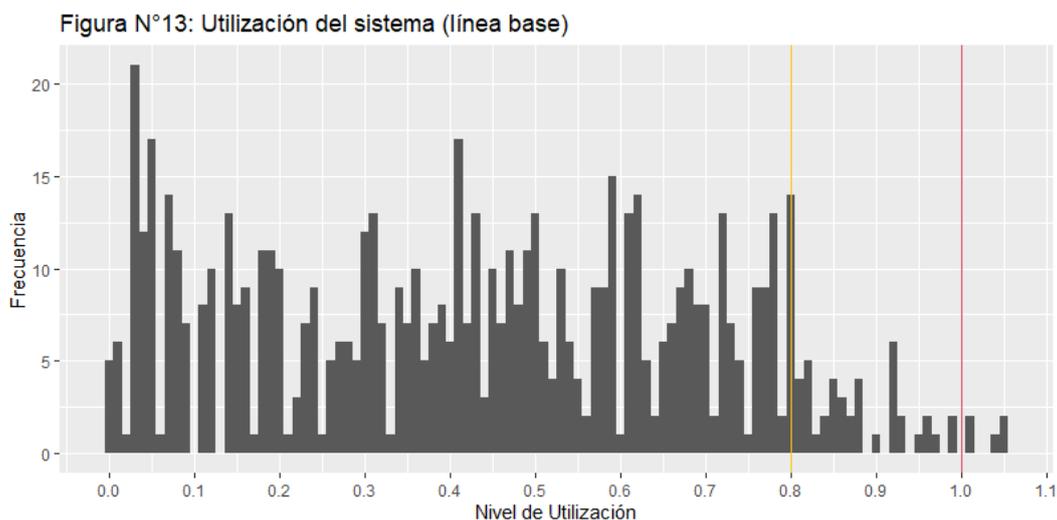
En la Figura N°12 se puede observar el impacto que tiene el aumento del nivel de utilización del sistema sobre el tiempo de espera para ser atendido en la estación de servicio durante el mes de febrero del año 2021. En ella, solo se expone el tiempo de

espera de aquellas utilizaciones inferiores a 100%, dado que cuando este valor se supera, en realidad se trata de un sistema inestable que tiende a valores de tiempo de espera infinitos. Por lo que, en esos casos, se asigna un tiempo de espera de 100 minutos, en concepto de penalidad por haberse saturado la capacidad máxima de atención del sistema.



Este gráfico ha sido realizado teniendo en cuenta los datos del mes completo, que incluye distintas combinaciones de coeficientes de variación de tiempo entre arribos y cantidad de servidores, generando que en algunos momentos el tiempo de espera disminuya levemente a pesar de estar aumentando la utilización del sistema. Sin perjuicio de ello, queda en evidencia que el tiempo de espera para ser atendido aumenta exponencialmente en la medida que el nivel de utilización del sistema supera el 80%.

Uno de los errores más graves que pueden cometerse a la hora de gestionar una fila de espera es permitir que el nivel de utilización del sistema supere el 100%, ya que no solo los tiempos de espera se vuelven infinitos, sino que éste se vuelve inestable e imposible de gestionar. Esto es justamente lo que se identificó en la estación de servicio en más de una oportunidad, lo cual se resume en la Figura N°13, que permite identificar la frecuencia de utilizaciones a lo largo del período bajo análisis. A partir de una delgada línea roja se precisan cuántas veces la estación de servicio ha visto su sistema desbordado por haber superado la demanda su capacidad máxima de atención al público. Asimismo, con una línea naranja se distinguen las utilizaciones superiores al 80%, las cuales también representan una alarma para la empresa.



El nivel de utilización promedio del sistema es de 43.24%. Sin embargo, el 6.85% de los horarios del mes se encuentran en un nivel de utilización superior al 80%, de los cuales el 10.87% se encuentra por encima del 100%. Evidentemente, es una métrica que debe ser mejorada para poder optimizar la calidad de atención del público, y eso es lo que se buscará con la herramienta de planificación de horarios que se propone en este trabajo.

Es muy importante tener en consideración la cantidad de empleados (o servidores) utilizados en un modelo de fila de espera porque existe un *trade off* constante entre la calidad de servicio que se presta y la eficiencia de los recursos que se tiene. Naturalmente, brindar un servicio de alta calidad, en donde se atienda a los clientes de manera casi inmediata, implica contratar mayor personal para tener un sistema holgado que difícilmente supere una utilización superior al 80%, lo que involucra un costo superior. En contraposición, se podría argumentar que se prefiere utilizar al máximo los recursos disponibles resignando calidad de atención.

Para atender al público durante el mes de febrero del año 2021, la empresa ha contado con 204 turnos (ver Figura N°9), lo que es equivalente a 51 turnos por semana. A modo informativo, se aclara que un turno es un período de tiempo de ocho horas corridas bajo las cuales un empleado presta sus servicios a la empresa de manera ininterrumpida. Por semana, un playero trabaja seis turnos y tiene un día libre. De esta manera, puede calcularse que la empresa ha requerido de 8.5 empleados por semana (= 51 turnos por semana / 6 turnos por empleado) para desempeñarse de la manera que lo hizo. El motivo por el cual el número obtenido no es entero se debe a que uno de los empleados tuvo catorce días de vacaciones, justo la mitad de los días del mes bajo análisis.

En conclusión, se deriva de la planificación manual de horarios de personal que, a pesar de que en líneas generales la misma pareciera ser correcta, a la hora de entrar en detalles existen varias falencias de esta metodología.

Por un lado, si se analizara únicamente el nivel promedio de utilización del sistema a lo largo del mes, se podría concluir que el mismo se encuentra en niveles óptimos, ya que es del 43.24%. Ahora, habiendo tomado conocimiento sobre las implicancias de niveles de utilización superiores al 80%, se puede notar que en el 6.85% de los casos el sistema de atención al público se encuentra exigido más allá de lo recomendado. Y ni hablar de los momentos en los que el sistema se encuentra saturado, que representan el 10.87% de los casos superiores al 80%, o el 0.74% del total. Evidentemente hay un punto de mejora en esta métrica.

Por otro lado, el tiempo de espera para ser atendido también se encuentra en un valor promedio más que adecuado, siendo éste de 1.91 minutos por cliente. Ahora, en el 4.46% de las horas del mes la realidad es que el mismo es superior a cinco minutos, alcanzando valores exponenciales tal como se veían en la Figura N°12. Si una estación de servicio quiere cuidar a sus clientes, debe trabajar firmemente en la cantidad de minutos que tarda en atenderlos, y eso es lo que se busca en este trabajo al plantear una programación de horarios a partir de un modelo de optimización.

Por último, es crucial tener en cuenta que la performance de la estación de servicio fue conseguida a través de 8.5 empleados por semana, valor que será juzgado más adelante cuando el modelo de optimización determine cuánto personal necesita para cumplir con la demanda. Como se mencionó anteriormente, tener en cuenta la cantidad de personal que la empresa dispuso para atender a la demanda es clave para entender la inversión que la misma realizó para conseguir estos resultados.

La Figura N°14 trae a colación un resumen de los resultados obtenidos por la estación de servicio durante el mes de febrero en el año 2021, en cuanto a nivel de utilización del sistema, tiempo de espera en ser atendido y cantidad de personal utilizado para ello.

Figura N°14: Resultados línea base

UTILIZACION			TIEMPO DE ESPERA		CANTIDAD DE PERSONAL
PROMEDIO	> 80%	> 100%	PROMEDIO	> 5 MIN	
43.24%	6.85%	0.74%	1.91 min	4.46%	8.50 empleados

ANALYTICS PREDICTIVO

CAPITULO 2: PREDICCION DE ARRIBOS

2.1 Estructura de los datos

La predicción de arribos se llevó a cabo fundamentalmente con los datos brindados por la estación de servicio, detallados en la Figura N°1, que permitieron conocer la cantidad de arribos que sucedían a cada hora desde 01/01/2018 hasta 28/02/2021. Además, se obtuvo información de la empresa en cuanto a la hora exacta a la que se recibían los camiones de combustible y los días anteriores a un aumento de precio. Asimismo, se incorporó información del Ministerio del Interior de la República Argentina, del Convenio Colectivo de Trabajo 317-99 de Estaciones de Servicio y del Servicio Meteorológico Nacional en cuanto a feriados, ceses de actividades obligatorios y temperatura y precipitaciones, respectivamente.

Con toda esta información se creó un *dataset* con 27.624 observaciones, 16 variables predictoras y una variable de respuesta. La Figura N°15 aclara brevemente de qué se trata cada una de las variables predictoras, siendo la variable de respuesta la cantidad de clientes que ingresan a la estación de servicio a cada hora.

Figura N°15: Descripción de las variables del *dataset*

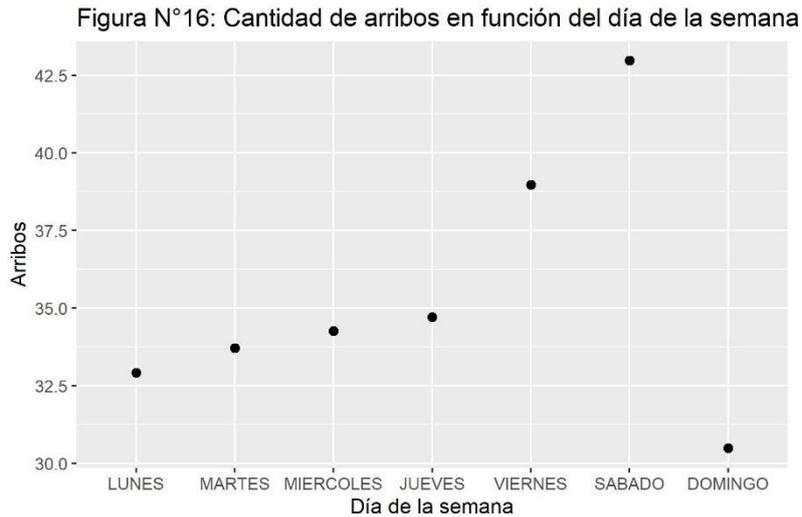
VARIABLE	DESCRIPCION	TIPO
Año	Año al que corresponden los arribos.	Factor
Mes	Mes al que corresponden los arribos.	Factor
Día	Día al que corresponden los arribos.	Factor
Hora	Hora a la que corresponden los arribos.	Factor
Dia.Sem	Día de la semana al que corresponden los arribos.	Factor
Num.Sem	Número de semana del año a la que corresponden los arribos.	Factor
Camión	Toma valor "SI" si a esa hora se descargó el camión de combustible. "NO" en caso contrario.	Factor
Finde	Toma valor "SI" si es fin de semana. "NO" en caso contrario.	Factor
Feriado	Toma valor "SI" si es feriado. "NO" en caso contrario.	Factor

Estación	Estación climática del año a la que corresponden los arribos.	Factor
C.Precio	Toma valor "SI" si es el día anterior a un aumento de precios. "NO" en caso contrario.	Factor
Cerrado	Toma valor "SI" si el establecimiento se encuentra cerrado por imposición del convenio colectivo de trabajo. "NO" en caso contrario.	Factor
Pandemia	Toma valores del 1 al 5 en función del grado de acatamiento a las restricciones de movilidad dadas por la cuarentena implementada por el Covid-19. Toma valor 0 si no hay pandemia.	Factor
Temperatura	Temperatura a cada hora.	Numérica
Precipitaciones	Precipitaciones a cada hora.	Numérica
Tendencia	Días transcurridos desde 01/01/2018	Numérica

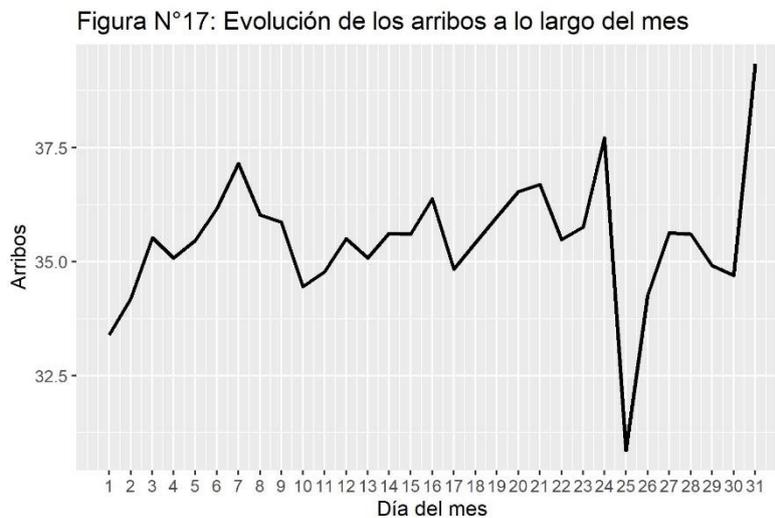
2.2 Análisis exploratorio

De manera complementaria a lo realizado con los primeros datos históricos recabados de la compañía, se procede a realizar un análisis exploratorio del *dataset* completo que, como se comentó, incluye datos procesados de diversas fuentes de la empresa, del Ministerio del Interior de la República Argentina, del Convenio Colectivo de Trabajo 317-99 de Estaciones de Servicio y del Servicio Meteorológico Nacional.

Cuando se empieza a prestar atención a las variables que pueden llegar a ser significativas para predecir la cantidad de arribos por hora en una estación de servicio, rápidamente se puede dilucidar que los días de la semana influyen, y mucho. Siguiendo la Figura N°16, es muy curioso observar que, en promedio, los arribos por hora de los lunes, martes, miércoles y jueves son muy similares, pero con una pequeña tendencia alcista del primero al último. Ahora, ya en vísperas del fin de semana, los viernes se produce un salto importante en el movimiento del establecimiento, el cual llega a su punto máximo los sábados. Por su parte, los domingos representan el día de la semana con menor tráfico, siendo el mismo apenas superior a 30 autos por hora.



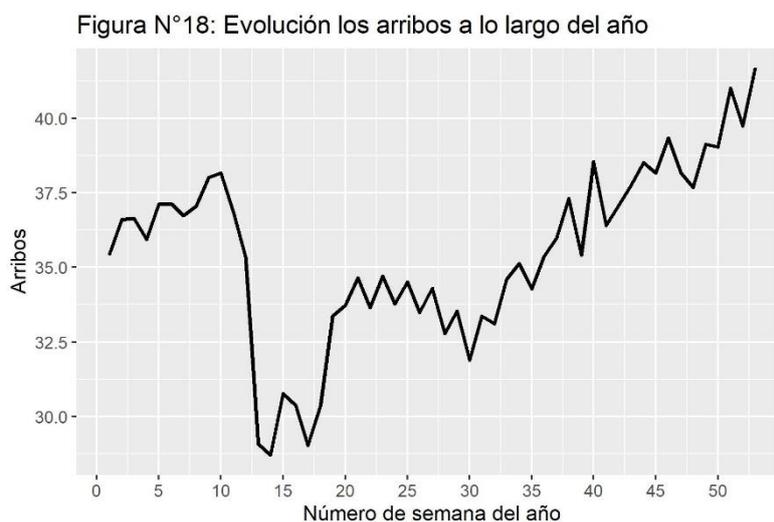
Así como los diferentes días de la semana tienen un efecto sobre los hábitos de consumo de los clientes de la estación de servicio, se buscó determinar si cada uno de los días que componen al mes también lo tiene. En general, los dueños de estaciones de servicio suelen considerar que la primera quincena del mes es mejor, en cuanto a cantidad de movimiento de la empresa, que la segunda quincena del mes. Sin embargo, en la Figura N°17 puede observarse que la cantidad de arribos promedio por hora a lo largo de cada día del mes es de lo más variado, y no presenta un patrón tan marcado. Si se analiza el gráfico por quincenas, se puede advertir que la primera quincena del mes suele tener una cantidad de arribos más estable, comenzando con su menor valor promedio el día 1 y alcanzando su pico el día 7. Por el otro lado, la segunda quincena se contempla de mayor varianza, teniendo su menor valor promedio el día 25 y su pico el día 31, valores que, a la vez, son el piso y techo del mes.



Siguiendo la línea de la evolución de los arribos promedio por hora, se identificó que, a lo largo de cada semana del año, los mismos también presentaban variaciones. Sorprendentemente, según se puede ver en la Figura N°18, la diferencia entre la semana en la que más autos por hora van a la estación de servicio, y la que menos, es superior a 10 autos por hora.

Posiblemente, la abrupta caída que se presenta entre las semanas 10 y 20 sea, en parte, por la etapa más estricta de la cuarentena implementada en el año 2020 pero, sin perjuicio de ello, las variaciones que se presentan son notables. Es importante destacar que el *dataset* incluye datos desde 01/01/2018 hasta 28/02/2021 inclusive, por lo que los efectos más duros de cuarentena son, en gran parte, balanceados por los períodos sin ella.

Asimismo, es llamativo contemplar el repunte de arribos promedio por hora que presenta la empresa a partir de la semana 30, alcanzando su valor promedio más alto en la última semana del año. El mayor movimiento que generan las fiestas de fin de año evidentemente tiene un impacto en el consumo de combustible y, por ende, en la necesidad de acudir a estaciones de servicio para poder seguir utilizando los vehículos.



Retomando nuevamente el tema de la cuarentena implementada por la República Argentina en el año 2020, se investigó específicamente el impacto de ésta sobre la estación de servicio bajo análisis. Como es de público conocimiento, la cuarentena redujo fuertemente la movilidad en el país y, consecuentemente, el consumo de combustible.

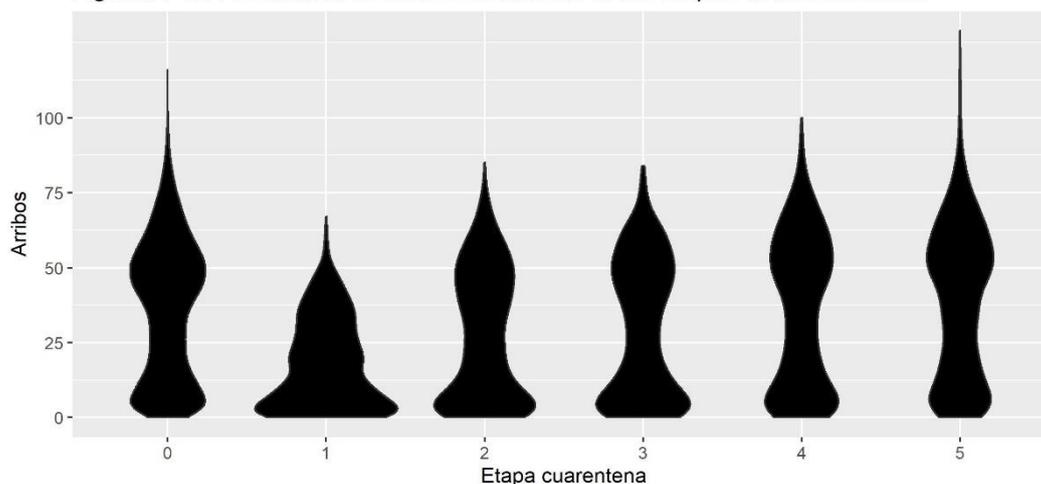
Para ello se llevó a cabo el fraccionamiento del *dataset* en seis etapas, las cuales fueron definidas de manera arbitraria junto con el estacionero, en función de la apreciación del cumplimiento de esta en la calle. La etapa 0 corresponde a aquel período de tiempo que

no fue parte de la cuarentena, mientras que la etapa 1 comienza con la vigencia del Decreto de Necesidad y Urgencia firmado por el Presidente de la Nación el día 20/03/2020 y finaliza el 31/05/2020. Las siguientes etapas corresponden a fraccionamientos de aproximadamente dos meses, siendo la etapa 2 desde 01/06/2020 hasta 31/07/2020, la etapa 3 desde 01/08/2020 hasta 30/09/2020, la etapa 4 desde 01/10/2020 hasta 30/11/2020 y la etapa 5 desde 01/12/2020 hasta 28/02/2021.

De esta manera se arriba a la Figura N°19, en la que puede contemplarse el enorme impacto de la etapa 1 en la distribución de los arribos. Con una base mucho más ancha que la etapa 0, se interpreta una evidente reducción de la cantidad de arribos por hora, encontrándose la mayor cantidad de los arribos por debajo de los 25 autos por hora. Asimismo, en esta primera etapa la cantidad máxima de arribos por hora supera ligeramente los 60 autos, a diferencia de la etapa anterior que se encuentra alrededor de los 115 autos. Sin lugar a duda, fue una etapa durísima para este tipo de negocios, dada la gran estructura que deben soportar.

Con respecto a las etapas siguientes, se ve un gran cambio en la distribución de los datos de la etapa 1 a la etapa 2, que con el correr de los meses fue adoptando una forma más parecida a la original, antes de la cuarentena. Sin embargo, puede contemplarse que la etapa 5 posee una distribución mucho menos concentrada que la etapa 0, alcanzando valores incluso superiores a ésta. Nuevamente, esta mayor dispersión de los datos reivindica la necesidad de planificar los horarios del personal de la manera más profesional posible, para así poder predecir con la mayor certeza posible estos valores tan dispares.

Figura N°19: Frecuencia de arribos en función de las etapas de la cuarentena

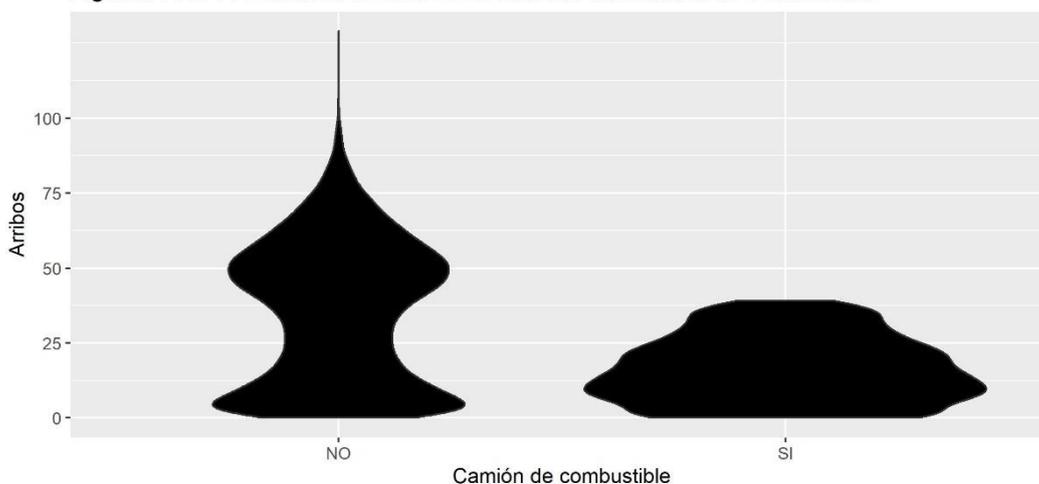


Otra de las variables que demostró tener un impacto importante en la predicción de arribos, es la presencia o no del camión cisterna, que es aquel que se presenta en la estación de servicio para realizar la descarga de combustible y llenar los tanques de almacenamiento de la empresa. Esta variable pudo ser incorporada gracias a que la estación de servicio cuenta con un registro de la hora de llegada y de la hora de salida del camión cisterna.

Dado que se trata de una operación riesgosa, la petrolera Shell les exige a sus estaciones de servicio que dejen de atender al público durante la descarga del camión cisterna o, al menos, que trabajen con una atención al público mínima. En ese momento, el principal foco de la empresa es la seguridad por sobre todas las cosas.

Es así como el impacto de la Figura N°20 termina siendo notable en cuanto a la distribución de los arribos en los momentos que hay, o en los que no hay, un camión cisterna descargando combustible. Al tener que cerrar la estación de servicio o, al menos, atender al público en niveles mínimos, se puede ver como la distribución de los arribos en ese momento pasa a tener un valor máximo de 33 autos por hora, encontrándose su valor promedio en 16 autos por hora. En este caso, la base de la distribución se vuelve realmente muy ancha, acumulando muchos valores pequeños, a diferencia de la distribución cuando no hay un camión cisterna descargando combustible, la cual es más delgada y alcanza valores máximos de hasta 130 autos por hora aproximadamente.

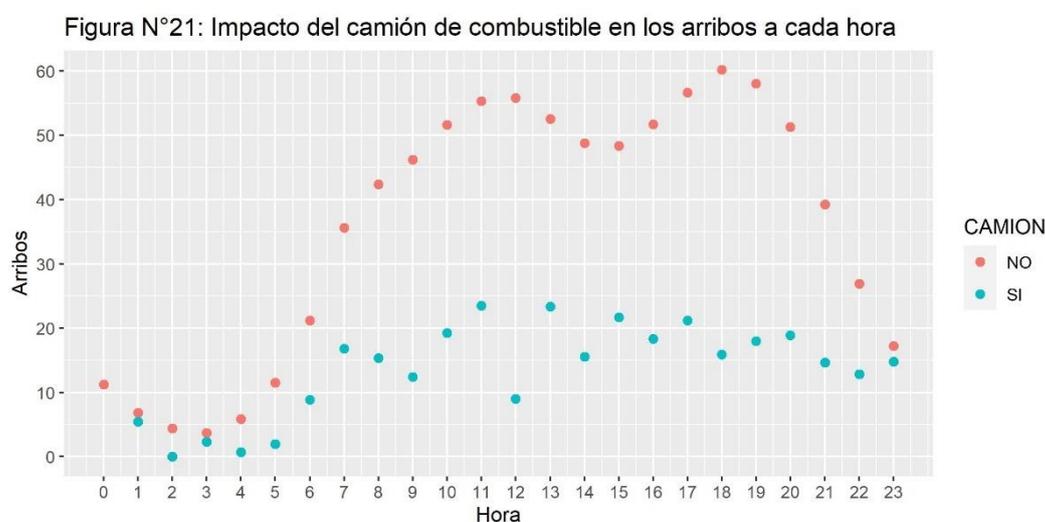
Figura N°20: Frecuencia de arribos en función del camión de combustible



Dado el gran impacto que tiene esta variable a nivel general, se buscó conocer su efecto específico a cada hora, confeccionando de esta manera la Figura N°21.

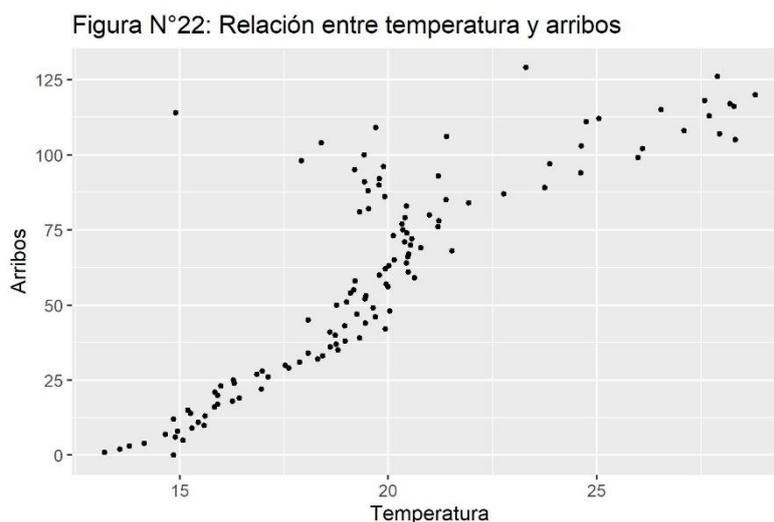
En rojo pueden verse los arribos promedio a cada hora de manera similar a lo presentado en la Figura N°3, en donde a partir de las 06:00 se visualiza un aumento significativo en la cantidad de arribos, que toca el primer pico del día a las 12:00 horas y el segundo a las 18:00 horas, siendo este último el más significativo del día. Luego, se registra un descenso constante de la cantidad de arribos por hora, hasta que vuelve a iniciar el ciclo al día siguiente.

En azul figuran los arribos promedio a cada hora en función de la hora en la que estuvo el camión cisterna en el establecimiento. A pesar de que a las 00:00 horas no hay registro de éste en todo el período analizado, puede percibirse que el efecto que tiene descargar el camión por la noche, desde las 23:00 horas hasta las 05:00 horas, es mucho menor que hacerlo durante el día. En los horarios más concurridos del día, a las 12:00 horas y a las 18:00 horas, se presenta una disminución en la cantidad de arribos de aproximadamente 45 autos. Un impacto realmente importante.



Por último, a partir de datos obtenidos del Servicio Meteorológico Nacional, en cuanto a temperatura y precipitaciones de cada hora del período bajo análisis, se pudo identificar la relativa importancia que tiene la temperatura para predecir los arribos por hora.

La Figura N°22 demuestra una correlación positiva entre la temperatura a cada hora y la cantidad de arribos que suceden en la misma, la cual es de 0.3169. Esto implica que la relación entre estas variables es baja, pero existente al fin, tomándose en consideración esta nueva variable para predecir los arribos de los clientes a cada hora. Sin embargo, es importante aclarar que la correlación no implica causalidad, por lo que no puede decirse que el aumento de la cantidad de arribos se deba a una mayor temperatura, ya que una correlación estadística no alcanza para inferir una relación causal.



2.3 Conjunto de entrenamiento, validación y testeo

El *dataset* se dividió en tres partes.

Por un lado, se creó el conjunto de testeo, compuesto por la totalidad de los datos desde 01/02/2021 hasta 28/02/2021 inclusive, que fue separado y utilizado únicamente para verificar la performance final del modelo cuando se consideró que el mismo ya se había entrenado lo suficiente. Estos datos no fueron utilizados en ningún momento para validar las decisiones durante la etapa de entrenamiento para evitar realizar *overfitting* sobre los mismos. Asimismo, se asumen desconocidos a los fines de este trabajo, ya que la predicción de los arribos durante este período de tiempo es la que se utilizará como *input* para el modelo de optimización planteado.

Por otro lado, se creó un conjunto de validación, compuesto por el 25% de los datos entre 01/02/2020 y 31/01/2021 inclusive, seleccionados de manera aleatoria. De esta manera, se quiere tener en cuenta la temporalidad de los datos, buscando que el modelo sea preciso a la hora de predecir aquellos más recientes en el tiempo, sobre todo, teniendo en cuenta la pandemia que azotó el mundo. Al cambiar de alguna manera los hábitos de consumo de los clientes como, por ejemplo, la cantidad de ellos que utilizan sus vehículos y que posteriormente necesitan cargar combustible, se consideró conveniente armar el conjunto de validación con estas características. Este conjunto de validación simuló ser un grupo de datos desconocidos durante la etapa de entrenamiento, que sirvió para ir confirmando cada una de las decisiones que se tomaron en cuanto a ingeniería de atributos y selección de modelos.

Por último, se creó un conjunto de entrenamiento, compuesto por los datos desde 01/01/2018 hasta 31/01/2021 inclusive que no fueron seleccionados al azar para formar parte del conjunto de validación, los cuales fueron utilizados para que los modelos empleados desarrollen su poder predictivo.

Una vez concluida la etapa de entrenamiento, el conjunto de validación fue incluido dentro del conjunto de entrenamiento para así preparar a los modelos con la totalidad de los datos desde 01/01/2018 hasta 31/01/2021 inclusive.

2.4 Modelos

El *input* más importante del modelo de optimización consta de la cantidad de autos que van a acercarse a la estación de servicio a cada hora del día, para así poder determinar la cantidad de playeros que van a ser necesarios a cada momento para atenderlos. Para realizar esta estimación de la manera más precisa posible, serán utilizadas diversas técnicas de *machine learning*. En concreto, se utilizarán dos de los algoritmos de aprendizaje supervisado más potentes de la actualidad: *Random Forest (RF)* y *Gradient Boosting Machine (GBM)*.

Random Forest es una combinación de árboles de decisión creados a partir de un mismo *dataset* pero con un muestreo doble en simultáneo: de variables predictoras y de observaciones. Esto genera una doble ganancia, ya que al tiempo que se generan mejoras en la capacidad predictiva al “decorrelar” los árboles se gana eficiencia computacional (Breiman, 2001). Para generar la variable de respuesta que será utilizada en el modelo de optimización, esta combinación de árboles de decisión creados lo que hace es promediar la opinión de cada uno de ellos en una única respuesta.

Gradient Boosting Machine, en lugar de ensamblar un conjunto de modelos, lo que hace es construir árboles de decisión secuenciales simples, los cuales van perfeccionando poco a poco la predicción realizada (Friedman, 1999). A diferencia del algoritmo anterior, la variable de respuesta se va generando de manera secuencial a partir del perfeccionamiento que se logra con el aporte de cada uno de los árboles de decisión que se van creando.

2.5 Métrica de evaluación

Para determinar el modelo más conveniente, la configuración de hiperparámetros a utilizar, la selección de variables y todo tipo de decisiones en cuanto a ingeniería de

atributos durante la etapa de entrenamiento, se utilizó como métrica de evaluación el error cuadrático medio.

El error cuadrático medio es una medida del error que existe entre un conjunto de datos predicho y un conjunto de datos observado. El mismo calcula el valor medio de la diferencia al cuadrado entre ambos conjuntos. De esta manera, se caracteriza por no tomar valores negativos y por destacar aquellos errores de mayor magnitud, dado que los eleva al cuadrado.

2.6 Ingeniería de atributos

La ingeniería de atributos es el acto de extraer rasgos de los datos brutos y transformarlos en formatos apropiados para el modelo de *machine learning* que se vaya a utilizar (Zheng y Casari, 2018). De hecho, la performance de los modelos de *machine learning* es muy dependiente de la manera en que se representen los datos o cómo estos sean utilizados (Bengio, 2013).

En primer lugar, lo que se realizó en este sentido fue un filtrado de los horarios que no tenían arribos, entre las 07:00 horas y las 21:00 horas, ambas inclusive, sin existencia de camión de combustible descargando, sin ser día feriado, sin estar cerrada la estación de servicio y sin encontrarse en pandemia. Se encontraron 82 observaciones que se debían a horas en los que el sistema había fallado desde 01/01/2018 hasta 28/02/2021 y no había registrado las visitas de los clientes, dado que los surtidores habían tenido que ser pasados a modo manual cada vez que esto sucedía.

Por otro lado, a partir de la fecha de cada día se crearon nuevas variables, tales como el día de la semana al que corresponde la observación, si la misma es en fin de semana o no, la estación del año a la que pertenece y el número de semana del año. Además, el día fue dividido en seis momentos distintos de igual longitud (mañana, mediodía, tarde, noche, medianoche y trasnoche) para intentar captar patrones adicionales.

Sin embargo, con el correr de los distintos modelos y el análisis de la mejora o no de su poder predictivo, algunas de las variables creadas en un primer momento fueron quitadas. Aquella variable que dividía al día en seis partes, por ejemplo, no supo agregar valor a los modelos y fue descartada.

Se probaron diversas combinaciones de variables de esta manera, y el conjunto que mejor poder predictivo demostró fue el que se presentó en la Figura N°15.

Para poder conectar temporalmente los datos, se incorporó la variable “tendencia”, la cual cuenta la cantidad de días transcurridos desde el primer día del *dataset* hasta el último, justamente para identificar aumentos o disminuciones de arribos a lo largo del tiempo.

Finalmente, se tipificó a cada variable de la manera correspondiente, a la vez que se decidió no escalar las variables numéricas dado que no aportaban valor a los modelos utilizados.

2.7 Optimización de hiperparámetros

La optimización de hiperparámetros consiste en encontrar la mejor configuración de parámetros posible para el modelo que se busca emplear, y es una instancia clave a la hora de entrenar un modelo de *machine learning* porque tiene un efecto directo sobre la performance de éste.

A pesar de ser la búsqueda manual y *grid search* dos de las estrategias más comunes para optimizar hiperparámetros, en este trabajo se llevó a cabo la estrategia de *random search*, ya que es más eficiente y hasta es capaz de encontrar mejores configuraciones que los métodos anteriores (Bergstra, 2012). A diferencia de *grid search*, *random search* analiza un espectro más amplio de valores de cada hiperparámetro, a la vez que ahorra costo computacional al no entrenar cada una de las combinaciones posibles entre éstos. Dada su robustez ante hiperparámetros irrelevantes, son aquellos más significativos para la predicción que se busca realizar los que terminan definiendo la configuración más adecuada.

El proceso de optimización de hiperparámetros se llevó a cabo en dos etapas.

En primer lugar, se determinó un amplio rango de posibles valores para cada uno de los hiperparámetros de cada modelo. En función de ello, se configuraron al azar 40 modelos de *Random Forest* y otros 40 modelos de *Gradient Boosting Machine*, para así obtener una primera aproximación de las configuraciones más adecuadas para cada algoritmo. La Figura N°23 y la Figura N°24 detallan los rangos utilizados para *Random Forest* y *Gradient Boosting Machine*, respectivamente.

Figura N°23: Primer rango de hiperparámetros para *Random Forest*

HIPERPARAMETRO	RANGO
Mtry	[1:15]
Ntree	[1000:15000]

Sample	[0.01:0.99]
Maxnodes	[100:5000]
Nodesize	[1:500]

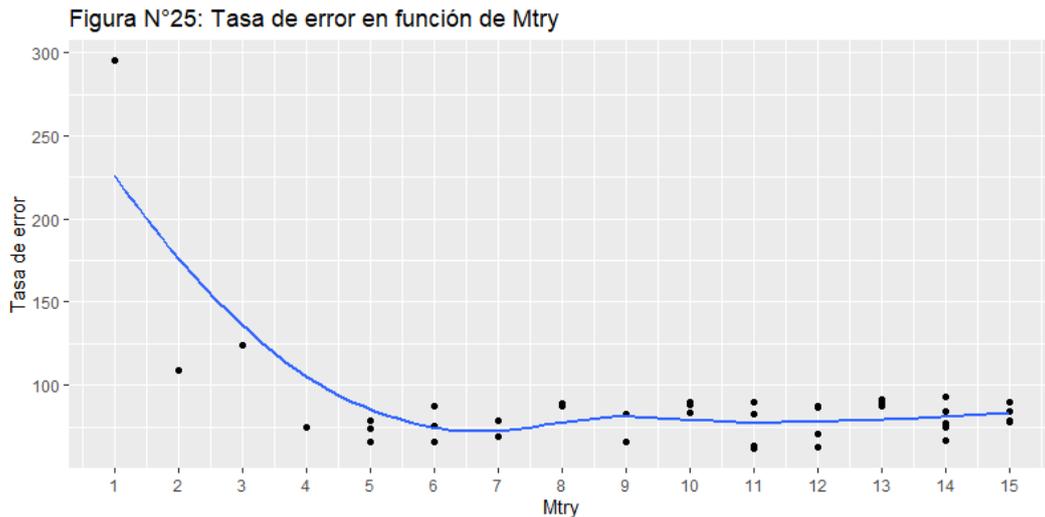
Figura N°24: Primer rango de hiperparámetros para GBM

HIPERPARAMETRO	RANGO
N.trees	[1000:15000]
Shrinkage	[0.001:0.1]
Interaction.depth	[1:49]
N.minobsinnode	[1:500]
Bag.fraction	[0.01:0.99]

En virtud de la cantidad de modelos entrenados para cada algoritmo, se realizó un análisis gráfico de sensibilidad de cada hiperparámetro para tomar conocimiento del impacto de cada uno de ellos sobre la tasa de error en el conjunto de validación. A pesar de que en este análisis no se respeta la regla *ceteris paribus*, que implica modificar una variable dejando constantes las demás, puede igualmente identificarse el efecto de cada hiperparámetro dado que son aquellos más robustos los que terminan haciendo una diferencia en la performance del modelo.

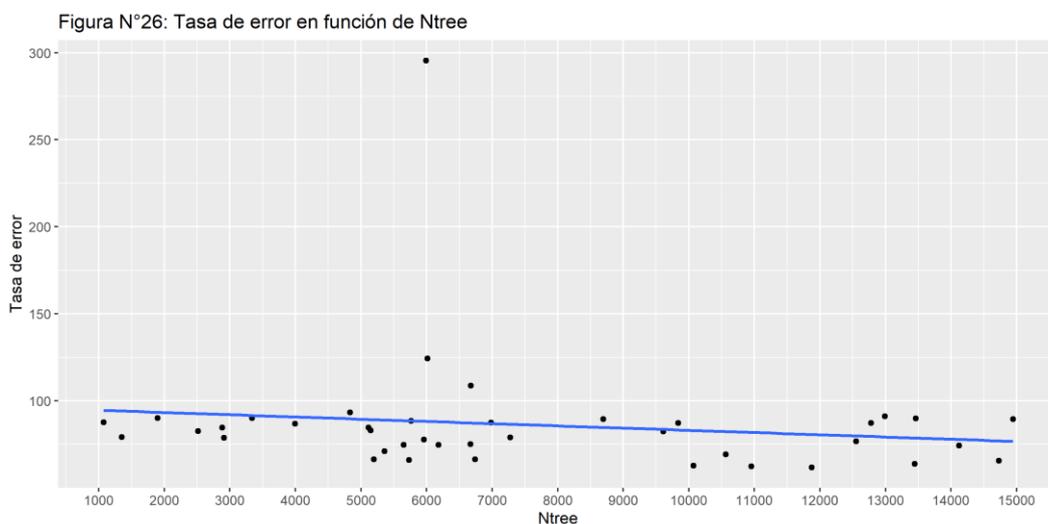
El primer hiperparámetro de *Random Forest* a configurar es *Mtry*, el cual determina la cantidad de variables, elegidas de manera aleatoria, que cada árbol va a utilizar a la hora de ser entrenado. Como se puede ver en la Figura N°25, en la medida que más *features* se seleccionan de manera aleatoria mejor es la capacidad predictiva del modelo. Sin embargo, esto sucede hasta un determinado punto, ya que luego la tasa de error se mantiene en valores constantes. En este sentido, *Mtry* demuestra ser uno de los hiperparámetros más influyentes, debido a la gran disminución marginal de la tasa de error que genera en un primer momento.

Específicamente en el entrenamiento realizado, los mejores resultados se obtuvieron al elegir 11 variables predictoras de manera aleatoria.



Ntree fija la cantidad de árboles que van a ser entrenados para realizar una predicción, los cuales posteriormente van a ser promediados para decidir la cantidad de autos que van a acercarse al establecimiento a cada hora. En el entrenamiento realizado se utilizó un amplio rango de árboles, entre 1000 y 15000, pero no tuvo un efecto significativo en la mejora de la tasa de error en el conjunto de validación. De hecho, como se puede contemplar en la Figura N°26, la línea de tendencia trazada en color azul se encuentra casi constante. A pesar de que los datos presentan una tendencia negativa, que implica una mejora de la performance del modelo en función de la incorporación de más árboles al mismo, la misma mucho menos empinada en relación con la anterior.

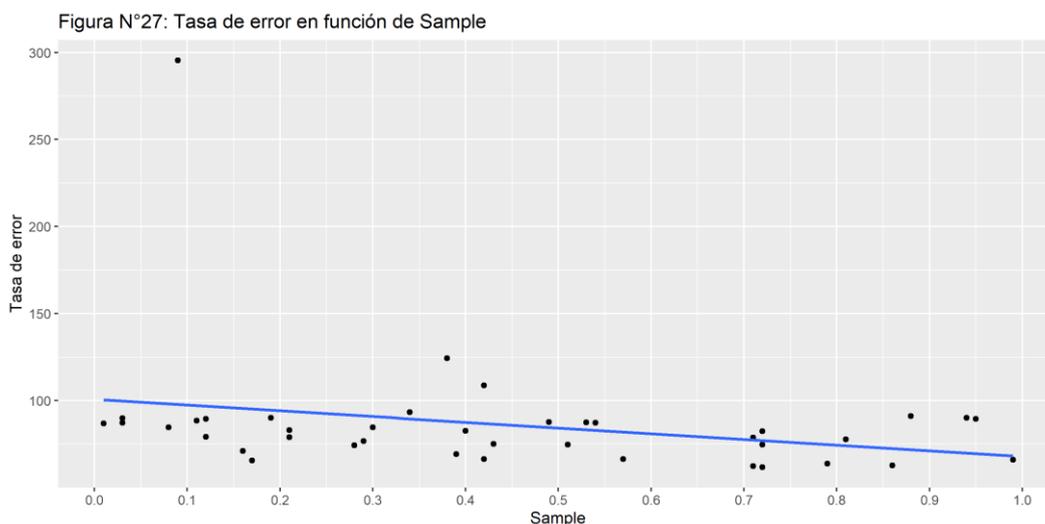
Sin perjuicio de ello, la cantidad de árboles que obtuvo el mejor rendimiento en el entrenamiento realizado fue de 11876.



La Figura N°27 pertenece a los resultados de Sample, que corresponde al hiperparámetro que determina el porcentaje de las observaciones que van a ser elegidas de manera

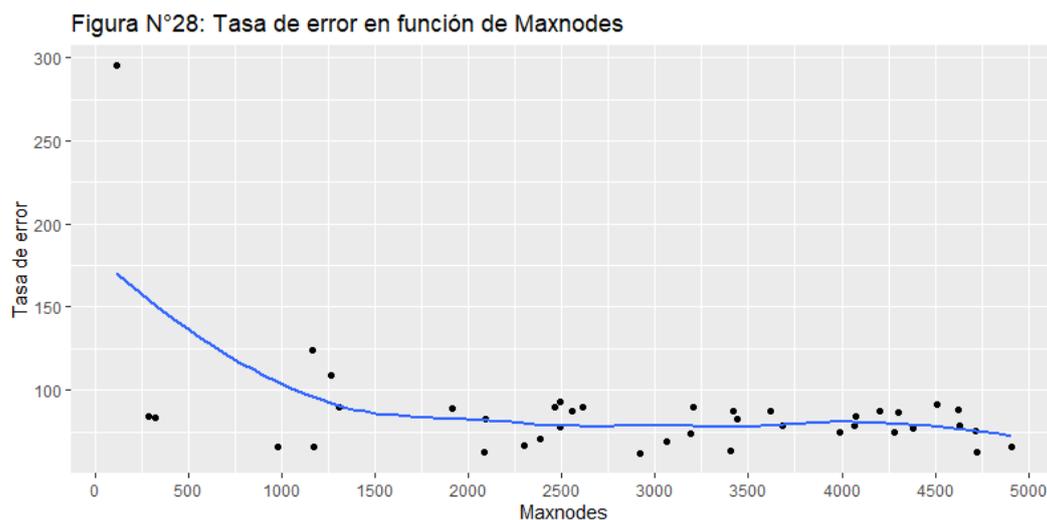
aleatoria para entrenar cada uno de los árboles. De esta manera, al igual que con Mtry, se consigue “decorrelar” a los árboles a la vez que se gana eficiencia computacional. En el gráfico, se aprecia una reducción de la tasa de error en el conjunto de entrenamiento en la medida que aumenta el valor de Sample, teniendo esta configuración un impacto más significativo que Ntree a lo largo del rango establecido.

El modelo de *Random Forest* que mejor resultado obtuvo se llevó a cabo con un Sample 0.72.



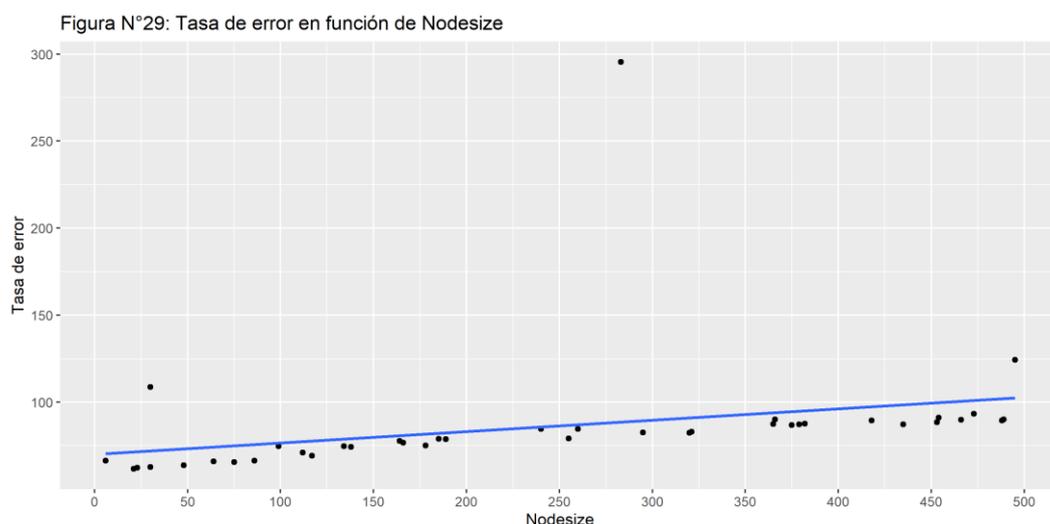
Maxnodes es el otro hiperparámetro que demostró tener mucha importancia a la hora de optimizar un modelo de *Random Forest*, el cual determina la cantidad máxima de nodos que puede tener cada uno de los árboles que va a ser entrenado. En la Figura N°28 se puede observar una línea de tendencia muy similar a que presentada con Mtry. Evidentemente, a mayor cantidad de nodos máximos que se le permite tener a un árbol, mayor es la capacidad de predicción de éste y, por ende, menor es su tasa de error en el conjunto de validación. No obstante, al igual que lo visto con Mtry, esta disminución de la tasa de error se genera solo en un primer momento, pasando a línea de tendencia a tomar una pendiente prácticamente constante para los valores posteriores.

En el entrenamiento efectuado, la cantidad de nodos que obtuvo la menor tasa de error fue de 2918.



El último hiperparámetro configurado en *Random Forest* fue *Nodesize*, que especifica la cantidad mínima de observaciones que pueden tener los nodos terminales de cada uno de los árboles entrenados. A diferencia de lo visto hasta el momento, este hiperparámetro presenta una línea de tendencia ascendente, lo que significa que, a mayor cantidad mínima de observaciones en cada nodo terminal, mayor es la tasa de error del modelo. Esto se detalla en la Figura N°29.

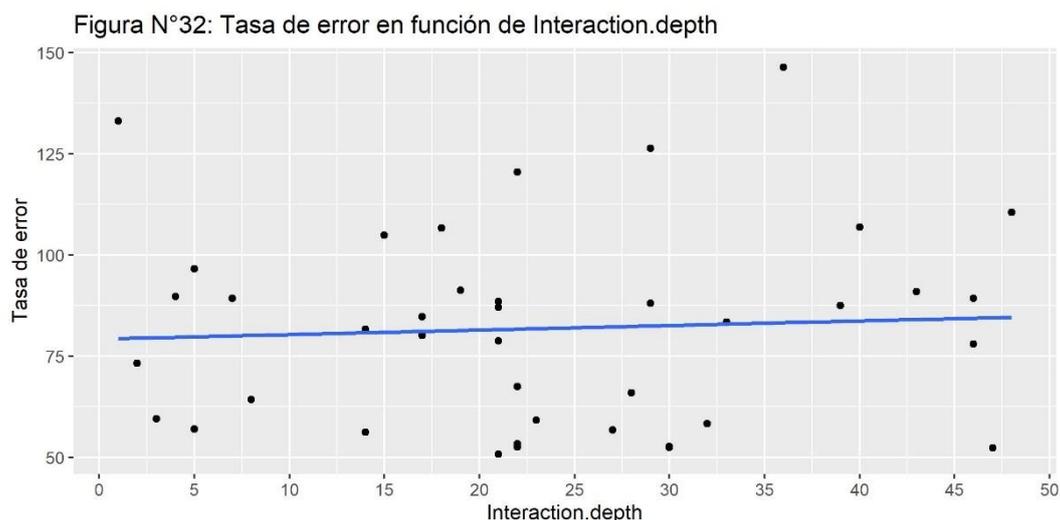
En esta primera etapa de optimización de hiperparámetros, el valor óptimo de *Nodesize* fue 21.



Por otro lado, el primer hiperparámetro a optimizar de *Gradient Boosting Machine* corresponde a *N.trees*, el cual especifica la cantidad de árboles que van a ser entrenados de manera secuencial para realizar la predicción. En función del entrenamiento llevado a cabo, la Figura N°30 muestra una línea de tendencia con pendiente positiva, indicando que, a mayor cantidad de árboles mayor es la tasa de error del modelo. Sin embargo, dada

De manera similar a lo ocurrido con `N.trees`, el hiperparámetro `Interaction.depth` presenta datos muy dispersos y una línea de tendencia prácticamente constante, demostrando no tener mayor injerencia sobre la performance del modelo, lo cual se aprecia en la Figura N°32. Por lo tanto, para el caso bajo análisis, la cantidad de cortes de cada árbol parece no tener un impacto específico sobre el resultado del modelo, ya que mucho va a depender de cómo estén configurados los demás parámetros también.

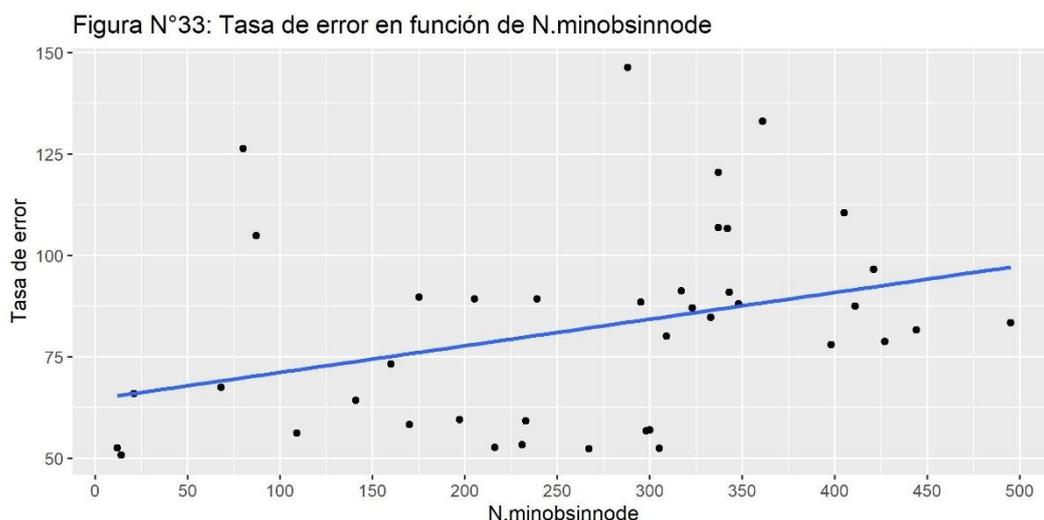
La mejor configuración de `Interaction.depth` en la primera etapa de entrenamiento fue de 21.



Al igual que con *Random Forest*, el hiperparámetro que determina la cantidad mínima de observaciones que puede tener cada nodo terminal presenta una marcada tendencia positiva, lo que significa que, a mayor cantidad mínima de observaciones en cada nodo terminal, mayor es la tasa de error sobre el conjunto de validación. La Figura N°33 detalla esta situación y, se puede ver en ella el importante salto que se produce en la tasa de error a partir de las 325 observaciones mínimas en cada nodo terminal.

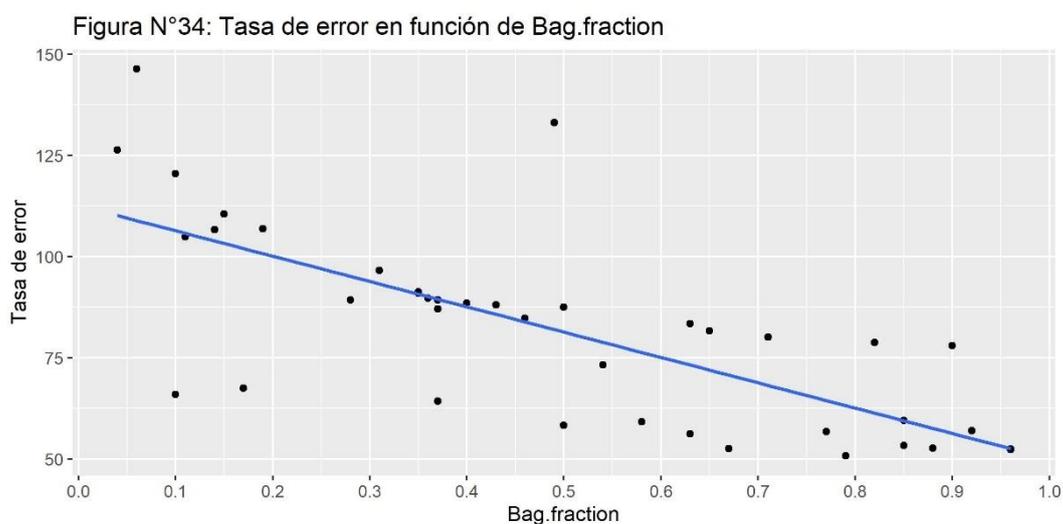
Esta pérdida en la capacidad predictiva que se produjo en ambos modelos al incrementar el tamaño de los nodos terminales se debe a que éstos comienzan a incurrir en *underfitting*, dado que pierden la capacidad de identificar los distintos patrones en los datos. Al tener tantos datos juntos en un mismo nodo terminal, ambos modelos dejaron de poder identificar ciertas particularidades que éstos le presentaban, dado que ahora se encuentran todos juntos como si fueran lo mismo.

En la primera etapa de optimización de hiperparámetros de *Gradient Boosting Machine*, la cantidad mínima de observaciones en cada nodo terminal óptima fue de 14.



El último hiperparámetro que se optimizó de *Gradient Boosting Machine* fue *Bag.fraction*, que corresponde a la proporción de observaciones que van a ser utilizadas para entrenar cada modelo. Salvando las diferencias entre algoritmos, se observa una evidente mejora en la capacidad predictiva del modelo en la medida en que más datos se le facilitan al mismo, al igual que con *Random Forest*. La línea de tendencia azul de la Figura N°34 sentencia que, sin dudas, este hiperparámetro es uno de los más influyentes a la hora de predecir la cantidad de arribos a cada hora en la estación de servicio.

La proporción óptima de observaciones encontrada para entrenar el modelo en esta primera etapa fue de 0.79.



En segundo lugar, teniendo en cuenta los hiperparámetros del modelo con mejor performance de cada algoritmo, se volvieron a crear nuevos rangos de hiperparámetros, pero esta vez más acotados, con límites superiores e inferiores al 20% del mismo. De esta manera, se corrieron ocho modelos al azar de *Random Forest* y otros ocho modelos de

Gradient Boosting Machine, para así encontrar la configuración de hiperparámetros más adecuada para cada algoritmo.

La Figura N°35 y Figura N°36 detallan los nuevos rangos utilizados para determinar finalmente la configuración óptima de cada algoritmo.

Figura N°35: Segundo rango de hiperparámetros para *Random Forest*

HIPERPARAMETRO	RANGO
Mtry	[9:13]
Ntree	[9501:14251]
Sample	[0.58:0.86]
Maxnodes	[2334:3502]
Nodesize	[17:25]

Figura N°36: Segundo rango de hiperparámetros para GBM

HIPERPARAMETRO	RANGO
N.trees	[4591:6887]
Shrinkage	[0.0296:0.0444]
Interaction.depth	[17:25]
N.minobsinnode	[11:17]
Bag.fraction	[0.63:0.95]

Es así como se dispuso que, para la situación bajo análisis, la configuración óptima para *Random Forest* es de 9, 10434, 0.73, 2862 y 22, valores correspondientes a los hiperparámetros Mtry, Ntree, Sample, Maxnodes y Nodesize, respectivamente.

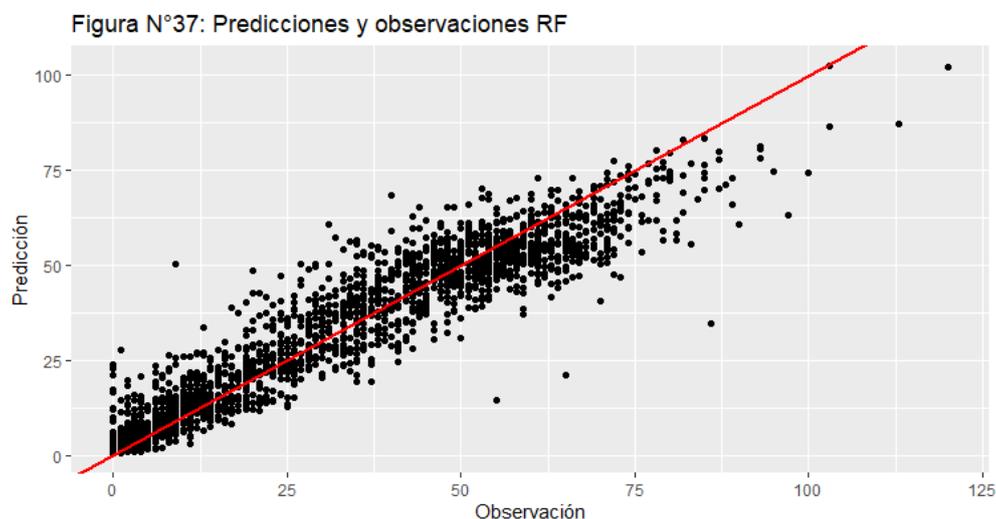
Por otro lado, la configuración óptima para *Gradient Boosting Machine* se estableció en 6575, 0.039, 17, 16, 0.92, valores correspondientes a los hiperparámetros N.trees, Shrinkage, Interaction.depth, N.minobsinnode y Bag.fraction, respectivamente.

2.8 Resultados

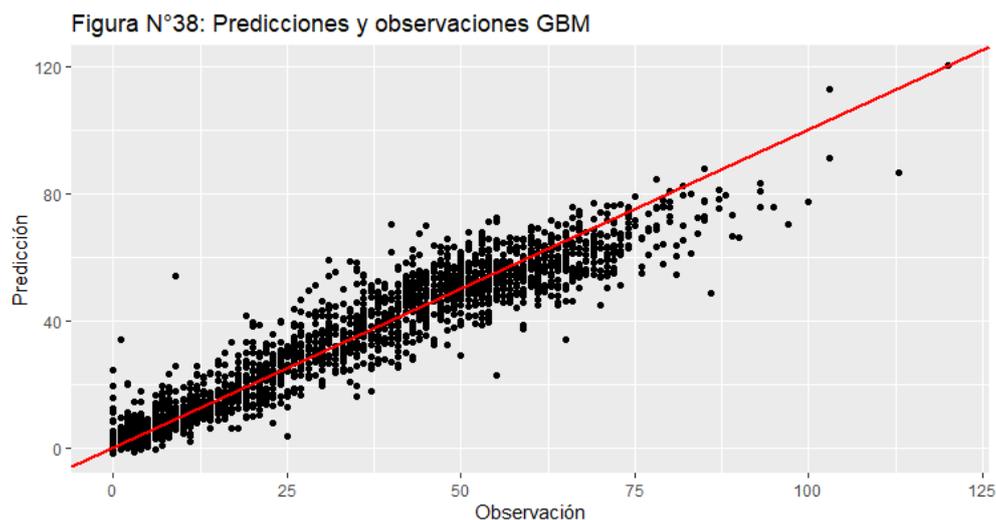
En esta sección se expondrán los resultados obtenidos con la muestra de entrenamiento tanto de *Random Forest* como de *Gradient Boosting Machine*, seleccionando de esta manera el mejor modelo de ambos. Este modelo será utilizado para llevar a cabo la predicción sobre los datos desde 01/02/2021 hasta 28/02/2021, que corresponde al

período temporal bajo análisis en este trabajo. Asimismo, la predicción será llevada a cabo utilizando los datos de entrenamiento y de validación en conjunto, para aprovechar de esta manera la totalidad de la información con la que se dispone. Posteriormente, se utilizarán estos datos como *input* para llevar a cabo la optimización de los horarios del personal de la estación de servicio.

Con la muestra de entrenamiento, el algoritmo de *Random Forest* obtuvo un error cuadrático medio de 60.42 sobre los datos de validación o, lo que es equivalente, un error promedio de 7.77 autos en su predicción correspondiente a cada hora. En la Figura N°37 puede verse un gráfico de dispersión que compara lo predicho con la realidad en cada uno de los datos de validación. Asimismo, se incorpora una línea roja que orienta en la precisión de cada una de las predicciones, ya que cuanto más cercana se encuentre la misma sobre la línea, más precisa es.



Por otro lado, con la muestra de entrenamiento, *Gradient Boosting Machine* demostró tener un error cuadrático medio de 50.04 sobre los datos de validación, o de 7.03 autos por hora. En la Figura N°38 se encuentra el gráfico de dispersión de este algoritmo, que compara las predicciones realizadas con las observaciones de los datos de validación. Además, como se mencionó anteriormente, la línea roja ayuda a identificar la tendencia que deben acompañar los datos para que sean considerados como predicciones correctas.

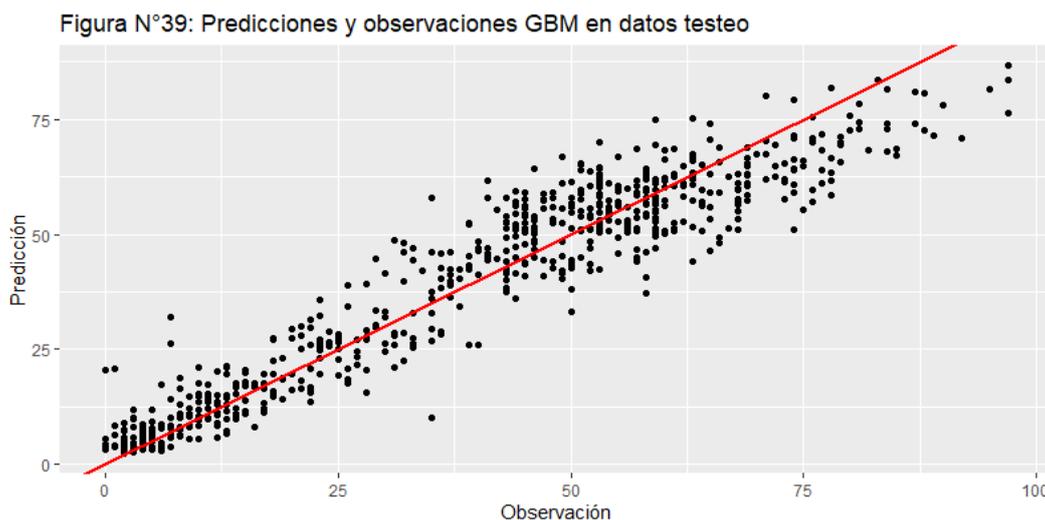


Dado que *Gradient Boosting Machine* demostró tener un desempeño un 17% superior al que obtuvo *Random Forest*, es el algoritmo que fue empleado para realizar la predicción del mes de febrero del año 2021. Posteriormente, estos datos serán utilizados como *input* en la herramienta de planificación de horarios que se busca desarrollar para la estación de servicio.

El error cuadrático medio de *Gradient Boosting Machine* sobre los datos de testeo, que corresponden al mes de febrero de 2021, fue de 57.46, o de 7.58 autos por hora. Una performance incluso superior a la que tuvo *Random Forest* sobre los datos de validación, que son aquellos que fueron utilizados para ir optimizando los modelos.

En la Figura N°39 se encuentra nuevamente un gráfico de dispersión entre las predicciones de *Gradient Boosting Machine* y las observaciones, pero esta vez con los datos del conjunto de testeo. De esta manera, se advierte nuevamente un correcto ajuste de las predicciones del algoritmo con relación a los datos reales, el cual es ligeramente peor al obtenido con los datos de validación, pero igualmente útil para el modelo de optimización de horarios.

Es importante destacar que, como se mencionó en la sección 2.3 (entrenamiento, validación y testeo), los datos de testeo no fueron utilizados en ningún momento de la etapa de entrenamiento. Es recién en esta instancia que fueron utilizados para realizar las predicciones correspondientes, y evaluar el error cuadrático medio sobre las mismas. Por ello, es que el modelo arroja una tasa de error ligeramente superior a la que resulta de los datos de validación.



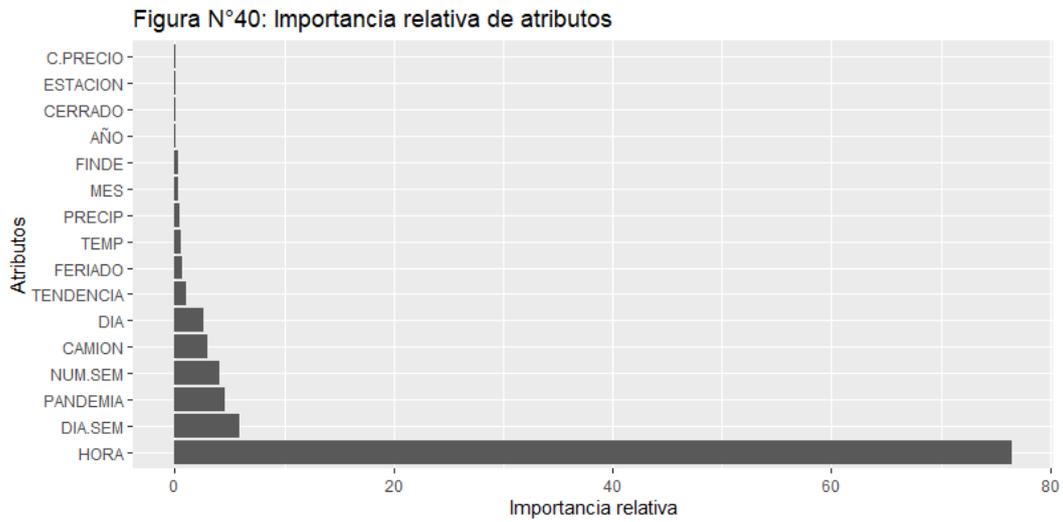
Al momento de predecir la cantidad de autos que van a ingresar a la estación de servicio a cada hora, no todas las variables tienen la misma entidad. Algunas variables predictoras son más relevantes a la hora de disminuir la tasa de error lo máximo posible, y esto es algo que *Gradient Boosting Machine* permite conocer.

En la Figura N°40 se encuentran los 16 *features* que fueron presentados en la Figura N°15, ordenados de menor a mayor según su relevancia, en función de lo determinado por *Gradient Boosting Machine*. Tal como era de esperar, la hora es la variable más significativa para saber cuántos autos se acercarán a la empresa a cada hora, pero, sin embargo, también se encontraron otras características que influyen en la demanda y vale la pena destacar.

Siguiendo el gráfico en orden de relevancia, se encuentra un segundo conjunto de variables que orientan en cuanto a la cantidad de autos que habrá que atender a cada hora: el día de la semana, la etapa de la cuarentena, el número de semana del año, la presencia o no del camión de combustible y el número de día de mes. En la práctica, esto es algo que intuitivamente los operadores de estaciones de servicio conocen, pero el hecho de que el algoritmo pueda cuantificar su importancia y darle una jerarquía, resulta sumamente útil. De hecho, es a partir de esta capacidad que tienen los algoritmos, de explicar y entender los hechos de una manera más precisa de lo que podría hacer un propio administrador de estaciones de servicio, que se busca llevar la planificación de horarios del personal a otro nivel en esta industria.

VARIABLES como la tendencia presente en los datos, la presencia de un día feriado, la temperatura y las precipitaciones se hallan en un conjunto ya intermedio, donde ayudan al modelo en su predicción, pero en menor grado que las variables mencionadas

anteriormente. Por último, el mes del año, el hecho de que sea fin de semana o no, los días que el establecimiento debe permanecer cerrado por ley, la estación del año y los días de cambio de precio, a pesar de que igualmente ayudan al modelo, son parte de las variables menos importantes para predecir la cantidad de arribos a cada hora en la estación de servicio.



ANALYTICS PRESCRIPTIVO

CAPITULO 3: PROGRAMACION DE HORARIOS

3.1 Modelado y técnica de resolución

El término “modelo” se usa generalmente para una estructura que se ha construido con el propósito de exhibir atributos y características de otros objetos (o situaciones). Generalmente, solo algunos de esos atributos y características son retenidos en el modelo, dependiendo de su finalidad, dado que éste solo es una representación simplificada de la realidad, aunque lo suficientemente precisa como para que las conclusiones (soluciones) obtenidas sean también válidas en la realidad. En el área de la investigación operativa, que es el ámbito de aplicación de este trabajo, estos modelos suelen ser matemáticos en el simbolismo algebraico que se utilizará para reflejar las relaciones internas en el objeto que se modela (Williams, 2013).

La investigación operativa consiste en el estudio y resolución computacional de modelos matemáticos para la toma de decisiones que afecten positivamente a los objetivos de una organización. En otras palabras, es simplemente un enfoque científico para la toma de decisiones, que busca diseñar y operar mejor un sistema, a menudo bajo condiciones que requieren la asignación de recursos escasos (Winston, 2004). Dentro de ella, existe la rama de optimización combinatoria, que se aboca al estudio y resolución computacional de modelos matemáticos en los que se optimiza una función sobre un conjunto definido en términos combinatorios, que es justamente lo que se busca hacer con la estación de servicio. Para ello, en este trabajo se plantea un modelo de programación lineal, el cual resuelve de manera determinística un problema de naturaleza estocástica.

La programación lineal es el campo de la programación matemática dedicado a maximizar o minimizar (optimizar) una función lineal, denominada función objetivo, de forma tal que las variables de dicha función estén sujetas a una serie de restricciones expresadas mediante un sistema de ecuaciones o inecuaciones. Dado que es habitual que algunos parámetros de los problemas de optimización combinatoria estén sujetos a incerteza, tal como los arribos de los clientes en el caso de la estación de servicio, en muchas ocasiones se los puede modelar por medio de variables aleatorias. Esto es a lo que se conoce como optimización estocástica. Sin embargo, debido a que la resolución computacional de este tipo de planteos es más difícil que para aquellos determinísticos, cuyos coeficientes están determinados de antemano, en muchas oportunidades se encara

su resolución por medio de modelos determinísticos, que contienen realizaciones concretas de las variables aleatorias del modelo estocástico. En esta oportunidad, los coeficientes fueron determinados en el segundo capítulo de este documento.

Para plantear un modelo que represente de manera simplificada, aunque precisa, la realidad en la que se desenvuelve la planificación de horarios en una estación de servicio, se han utilizado tres conjuntos de datos, un parámetro, tres variables, una función objetivo y siete restricciones. En el Anexo I se detalla el código utilizado que se describe a continuación, el cual permite alcanzar soluciones que son absolutamente factibles en la práctica.

El primer conjunto utilizado corresponde a la cantidad de empleados que se le disponen al *solver* para asignar a cada horario, la cual es de ocho. Para ayudar al modelo a arribar a una solución óptima con el menor *gap* posible, el conjunto correspondiente a la cantidad de empleados óptimos es dada como *input*, la cual es obtenida del mismo modelo al realizar pruebas de factibilidad. No se asigna un conjunto de empleados superior al óptimo para ayudar al *solver* a obtener una solución, y que esta sea la mejor posible.

Además, se empleó un conjunto respectivo a las 672 horas que componen al mes de febrero del año 2021, en lugar de uno de días y otro de las horas del día, para simplificar la escritura del modelo en cuanto la cantidad de conjuntos a tratar, y el cumplimiento de las doce horas obligatorias de descanso entre turnos. El día y la hora a la que debe ingresar el empleado al establecimiento es fácilmente interpretable al exportar el resultado obtenido a un Excel, mediante la aplicación de una fórmula determinada.

Un último conjunto, de horas de ingreso prohibidas, es facilitado al modelo para evitar que los empleados ingresen o se retiren del establecimiento entre las 01:00 horas y las 05:00 horas inclusive. Algo que va a permitir que los horarios sean altamente aceptados, tanto por el personal como por los empresarios.

El parámetro utilizado toma como *input* aquello realizado en la etapa de predicción de arribos, justamente para determinar la cantidad de autos que van a ingresar al establecimiento a lo largo de las 672 horas que componen al mes de febrero. Esta información es clave ya que, junto con la capacidad de atención de cada servidor determinada en el análisis de teoría de colas, posibilita que se cree una restricción que no permita que la utilización del sistema supere el 80%. Supuesto básico sobre el que se basa este trabajo.

Las tres variables empleadas son binarias. La primera de ellas se encarga de asignar al empleado a la tarea, la segunda fija la hora de ingreso de cada empleado, el cual posteriormente trabajará por ocho horas corridas, y la tercera premia al modelo por fijar a un empleado el mismo horario de ingreso que el día anterior.

A partir de ello, se creó una función objetivo que minimiza la cantidad de empleados necesarios a la vez que le asigna al personal el horario más estable posible. Todo esto sujeto al cumplimiento de una serie de restricciones que obedecen a la Ley de Contrato de Trabajo, al Convenio Colectivo de Empleados de Estaciones de Servicio y a las buenas prácticas empresariales, que hacen que este planteo sea posible en la realidad. Las mismas serán explicadas en detalle al momento de la presentación matemática del modelo.

Durante el proceso de modelado, se identificó la posibilidad de plantear un modelo más exigente aun en cuanto a la utilización máxima del sistema. Por lo que, además de haber un modelo que plantee un horario de personal en donde la utilización del sistema no supere el 80%, se planteará un segundo modelo un 5% mejor, en donde la utilización nunca supere el 76%. Esto es posible dado a las características particulares de demanda del mes bajo análisis, que permiten exigir al modelo un poco más allá de lo que la teoría ya considera como adecuado.

A su vez, tal como se reconoció esta alternativa superadora, también se tomó conocimiento de la posibilidad de plantear un modelo más relajado en cuanto a la máxima utilización del sistema, pero que permita ahorros de personal más significativos en detrimento del tiempo de espera del cliente. Este tercer modelo, en donde la utilización máxima del sistema es de 90%, también será planteado para analizar sus efectos.

Para resolver el modelo de programación lineal entera determinístico, el mismo es escrito a través del lenguaje de modelado ZIMPL⁸. Por medio de su interprete, es ZIMPL quien compone el problema en formato de programación lineal, el cual posteriormente es resuelto por el *solver* SCIP. Esto implica un proceso iterativo entre el lenguaje de modelado y el *solver*, en donde se comienza con un primer planteo sencillo, y luego se va complejizando hasta arribar al resultado esperado.

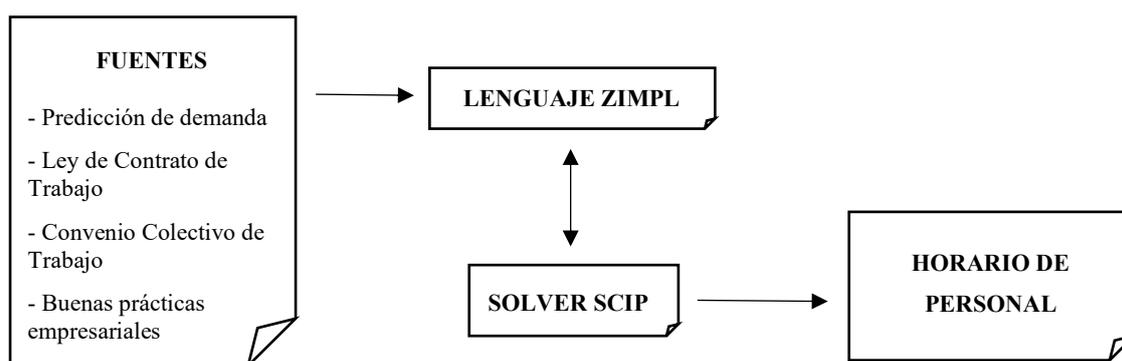
Asimismo, para obtener el resultado final, se tuvo que emplear una heurística muy sencilla, dado que el *solver* no finalizaba por sus propios medios, que consistió en

⁸ ZIMPL es un lenguaje de modelado que se caracteriza por ser flexible para realizar cambios y adaptarse fácilmente a la problemática que se busca resolver. Para más información puede consultarse <https://zimpl.zib.de/>.

establecer un tiempo máximo de resolución de cinco minutos y quedarse con la mejor solución hasta ese momento. Como se verá más adelante, las brechas obtenidas entre el resultado óptimo y el presentado son realmente insignificantes, siendo la solución obtenida prácticamente óptima y, por ende, considerada como el horario del personal para el mes de febrero del año 2021.

La Figura N°41 presenta un diagrama conceptual sobre cómo se constituye la herramienta de planificación de horarios de personal para el mes de febrero del año 2021 en cada uno de los modelos realizados.

Figura N°41: Diagrama conceptual



3.2 Modelo original

El modelo original es un modelo que atiende a una de las premisas más importantes dentro de la teoría de colas, que consiste en nunca superar el 80% de utilización de la capacidad de atención de la empresa, justamente para evitar tiempos de espera exponenciales y asegurar una excelente calidad de atención. Enfocado en minimizar la cantidad de personal necesario para atender la demanda predicha por el algoritmo *Gradient Boosting Machine*, este modelo contempla las restricciones planteadas por la Ley de Contrato de Trabajo vigente en la República Argentina, el Convenio Colectivo de Empleados de Estaciones de Servicio y ciertas consideraciones que hacen a las buenas prácticas empresariales.

A continuación, se especifica su modelo matemático.

Conjuntos:

- (1) $E = \{1, \dots, 8\}$
- (2) $H = \{0, \dots, 671\}$
- (3) $H_0 = \{1, 2, 3, 4, 5, 17, 18, 19, 20, 21, \dots, 669\}$

(1) Empleados disponibles para atender al público.

(2) Horas en el mes de febrero del año 2021.

(3) Horas del mes en las que el ingreso al establecimiento está prohibido, ya sea porque el ingreso es entre las 01:00 horas y las 05:00 horas, ambas inclusive, o porque el egreso es en dichas horas.

Parámetro:

- A_h : Cantidad de arribos a la hora h .

Variables:

- X_e : Asignación del empleado e a la tarea.
- Y_{eh} : Asignación del empleado e a la hora de ingreso h .
- Z_{eh} : Asignación de premio al empleado e en la hora de ingreso h . El premio consiste en tener la misma hora de ingreso que el día anterior.

Función Objetivo:

- (1) $\text{Min } \sum_{e \in E} X_e - \sum_{e \in E} \sum_{h \in H} 0.001 * Z_{eh}$

(1) Minimizar la cantidad de empleados necesarios para atender al público a la vez que se asignan la mayor cantidad de premios.

Sujeto a:

- (1) $A_h \leq 0.80 * \sum_{e \in E} \sum_{h-7}^h Y_{eh} * 37.02 \quad \forall h \in H$
- (2) $\sum_{h \in H} Y_{eh} = 24 * X_e \quad \forall e \in E$
- (3) $\sum_{h-19}^h Y_{eh} \leq 1 \quad \forall e \in E y \forall h \in H$
- (4) $\sum_{h-167}^h Y_{eh} \leq 6 \quad \forall e \in E y \forall h \in H$
- (5) $Y_{eh_0} = 0 \quad \forall e \in E y \forall h_0 \in H_0$
- (6) $Y_{eh-24} \geq Z_{eh} \quad \forall e \in E y \forall h \in H$
- (7) $Z_{eh} \leq Y_{eh} \quad \forall e \in E y \forall h \in H$
- (8) $X_e = \{0,1\} \quad \forall e \in E$
- (9) $Y_{eh} = \{0,1\} \quad \forall e \in E y \forall h \in H$
- (10) $Z_{eh} = \{0,1\} \quad \forall e \in E y \forall h \in H$

(1) La capacidad de atención al público de la empresa, la cual es de 37.02 autos por empleado por hora, no puede superar el 80% de utilización.

(2) Los empleados que sean asignados deben trabajar exactamente 24 días en el mes.

(3) Los empleados deben tener al menos 12 horas de descanso entre el final de una jornada y el inicio de otra.

(4) Los empleados pueden tener un máximo de 6 días de trabajo sin descanso.

(5) Los empleados no pueden ingresar a trabajar en los horarios prohibidos, ya sea porque el inicio o el final de la jornada laboral es en la madrugada.

(6) (7) Se ligan las variables para premiar al modelo cuando la hora ingreso de un empleado de un determinado día es igual a la hora de ingreso del día anterior.

(8) La variable X debe ser binaria.

(9) La variable Y debe ser binaria.

(10) La variable Z debe ser binaria.

Luego de plantear el modelo matemático presentado anteriormente, el *solver* SCIP arroja la solución que se representa en la Figura N°42. La misma se consigue al aplicar una heurística consistente en establecer un tiempo de resolución máximo de cinco minutos, y presenta un *gap* de 0.65% con respecto a la solución óptima.

El formato elegido para presentar los datos consiste en una tabla que podría ser perfectamente utilizada en las operaciones diarias de una estación de servicio, dada su sencillez y fácil interpretabilidad. En ella, todo el personal puede tomar conocimiento rápidamente de qué debe hacer cada día del mes, siendo los números la hora de ingreso de cada día y la “F” su día de franco.

Asimismo, otra gran ventaja que se puede visualizar es lo estable que son los horarios del personal, lo cual se consiguió gracias a las restricciones *soft* incorporadas al modelo. Si hay algo que uno se puede quedar tranquilo con este horario es que, dada la demanda predicha por el algoritmo, no hay una combinación de turnos que tenga un horario tan estable para cada empleado como el que se muestra a continuación. Lógicamente, este atributo es muy valorado por el personal que trabaja bajo jornadas rotativas, lo que permite que se cree un mejor clima laboral en la empresa.

Otra particularidad que se encuentra presente en el horario, y que contribuye al buen clima laboral, es que ningún empleado ingresa al establecimiento o egresa del mismo entre las 01:00 horas y las 05:00 horas, dando la posibilidad a los playeros de conseguir con mayor facilidad transporte público, en caso de necesitarlo, y mejores horas de sueño.

Por último, se podrá contemplar que el modelo atiende perfectamente todas las restricciones *hard* planteadas, lo que lo hace absolutamente válido y aplicable en la realidad por cumplir con todos los requisitos legales.

Figura N°42: Output modelo original

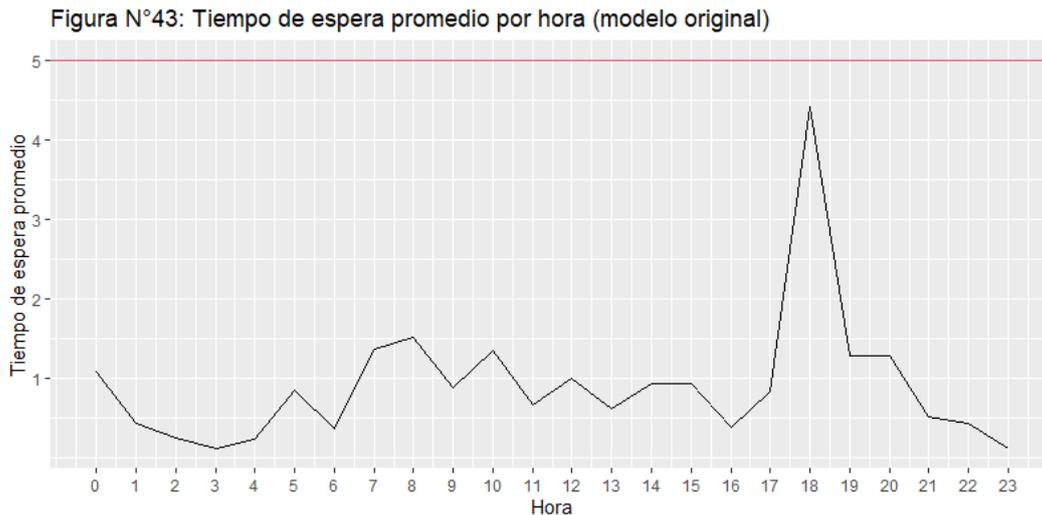
	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28		
EMPLEADO 1	16	16	16	16	22	22	F	16	16	16	16	22	22	F	16	16	16	16	22	22	F	16	16	16	F	22	22	22	
EMPLEADO 2	8	8	F	8	8	8	8	8	8	F	8	8	8	8	8	8	F	8	8	8	8	8	8	8	F	9	9	9	9
EMPLEADO 3	F	16	16	16	16	16	16	F	16	16	16	16	16	16	F	16	16	16	16	16	16	F	16	16	16	16	16	16	16
EMPLEADO 4	0	0	0	0	0	0	F	0	0	0	0	0	0	F	0	0	0	0	0	0	F	0	0	0	0	0	0	0	F
EMPLEADO 5	0	F	9	9	16	16	16	16	F	9	9	16	16	16	16	F	9	9	16	16	16	16	F	9	16	16	16	16	
EMPLEADO 6	14	14	14	14	14	14	F	14	14	14	14	14	14	F	14	14	14	14	14	14	F	14	14	14	14	14	14	14	F
EMPLEADO 7	6	6	6	F	8	8	8	6	6	6	F	8	8	8	6	6	6	F	8	8	8	6	F	8	8	8	8	8	
EMPLEADO 8	6	F	6	6	6	6	6	6	F	6	6	6	6	6	F	6	6	6	6	6	6	F	6	6	6	6	6	6	6

De la misma manera que lo realizado para los datos de operaciones tomados como línea base, se procede a evaluar la performance del modelo original en función de tres métricas: tiempo de espera en cola, utilización del sistema y cantidad de personal.

El tiempo de espera en cola a partir de la asignación de personal que plantea el modelo original otorga resultados asombrosos a lo largo del día, oscilando en valores alrededor de un solo minuto de espera, o menos, en promedio a cada hora. El único momento del día en que el tiempo de espera en cola es significativamente superior al resto de los horarios es a las 18:00 horas, siendo el mismo de aproximadamente 4.50 minutos.

El tiempo de espera promedio para ser atendido se ha fijado en 0.92 minutos por cliente, lo que a priori es un gran valor. Sin embargo, es importante tener en cuenta que dichos valores son promedios, por lo que, de todas formas, en algunos momentos del mes el tiempo de espera es superior a los cinco minutos, el cual es un valor de referencia que en caso de ser superado resulta nocivo para la capacidad de retención de clientes del negocio. De las 672 horas que componen al mes de febrero del año 2021, en solo el 1.93% de ellas el tiempo de espera fue superior a cinco minutos.

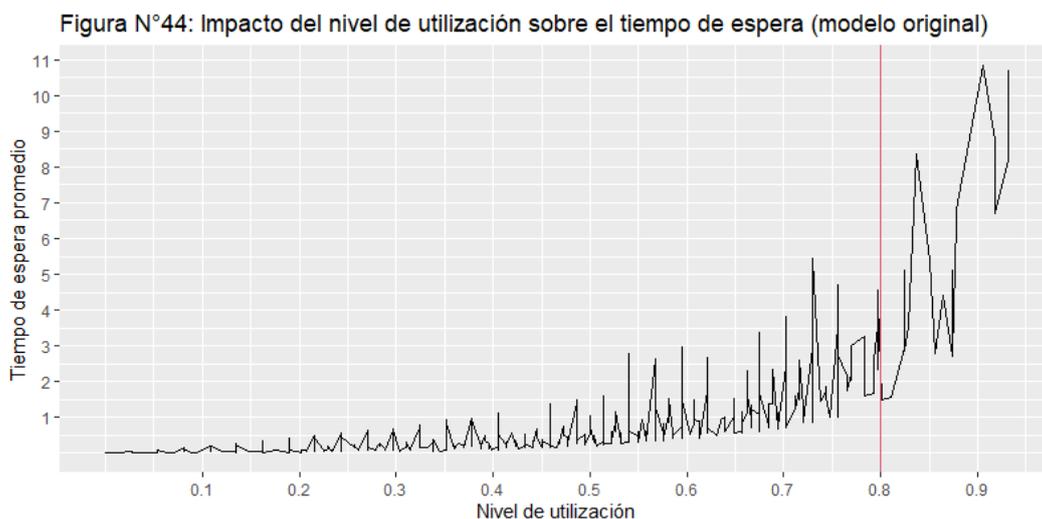
La Figura N°43 permite observar el tiempo de espera promedio a cada hora, y se puede ver que el modelo realizado mantiene por debajo de este umbral tan importante de cinco minutos de espera a todos los horarios del día.



De manera complementaria, puede verse en la Figura N°44 cómo va oscilando el tiempo de espera promedio en la medida en que aumenta la utilización del sistema en las 672 horas que componen al mes de febrero del año 2021.

Dado que la predicción realizada por *Gradient Boosting Machine* no es exacta, hay unas pocas horas en el mes en donde la utilización supera el 80%, generando que los tiempos de espera promedio alcancen casi los 11 minutos. Sin embargo, tal como se concluyó anteriormente, esto no afecta significativamente la performance global del modelo original a lo largo del mes bajo análisis, ya que todos los tiempos de espera promedio a cada hora se encuentran por debajo de los cinco minutos.

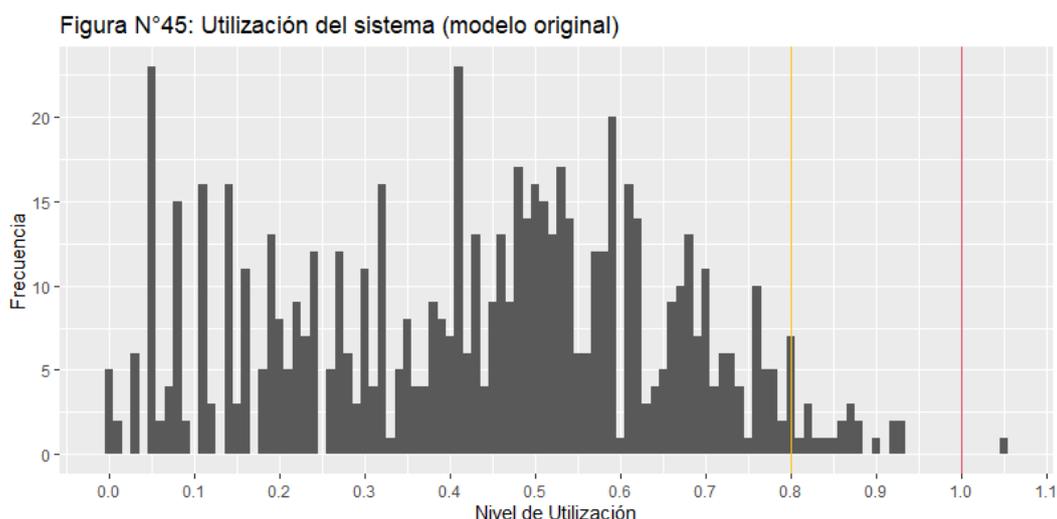
Es valioso dejar en claro que este gráfico solo contiene el tiempo de espera de aquellas utilidades inferiores a 100%, dado que cuando este valor se iguala o se supera, se trata de un sistema inestable con tiempos de espera que tienden a infinito. En esos casos, el tiempo de espera asignado es de 100 minutos en concepto de penalidad por ello.



Tal como se ha mencionado anteriormente, es importante recordar que la utilización del sistema es una métrica que permite conocer cuán exigida se encuentra la estructura organizacional destinada a atender al público, y que su correcta gestión es clave para conseguir que el tiempo de espera sea el menor posible. Para ello, es fundamental que la utilización del sistema se mantenga siempre por debajo del 80%, debido a que cuando comienza a superarse este valor el tiempo de espera se dispara rápidamente. Ni hablar de cuando la misma supera el 100%, ya que resulta en un sistema inestable e imposible de gestionar, con valores infinitos como tiempos de espera.

La Figura N°45 detalla la frecuencia de los distintos niveles de utilización a lo largo de todo el mes bajo análisis, a lo que se le adiciona una delgada línea naranja para visualizar fácilmente las utilizations superiores al 80% y otra roja para las utilizations superiores al 100%.

El nivel de utilización promedio del sistema es de 43.41%. Ahora, en el gráfico se puede observar que en algunos pocos momentos del mes el nivel de utilización es superior al 80%, fracción que representa el 3.13% del total, de los cuales el 4.80% se encuentra por encima del 100%. En otras palabras, solo en una oportunidad la capacidad de atención de la empresa fue superada por la demanda del momento, generando ese único dato que se encuentra a la derecha de la línea roja.



El personal es el concepto de mayor peso dentro de la estructura de costos de una estación de servicio, por lo que su correcta gestión es indispensable para asegurar la rentabilidad del negocio. Asimismo, es el personal también quien, a través de su atención de calidad, conquista y retiene a los clientes de la empresa, que son quienes generan los ingresos del establecimiento y, por ende, su rentabilidad. De esta manera, surgen dos objetivos

contrapuestos que ponen a prueba a todo estacionero, interpe­lándolo en un dilema que no es nada fácil de resolver: cómo maximizar la calidad de atención a la vez que se minimizan los costos de personal.

El modelo planteado atiende estos dos objetivos de manera simultánea, justamente al buscar minimizar la cantidad de empleados a la vez que se mantiene el nivel de utilización del sistema en valores no superiores al 80%. Es así como el *solver* arriba a la conclusión que la demanda del mes de febrero del año 2021 puede atenderse perfectamente con ocho empleados. Este dato es fundamental para todo empresario, ya que lo orienta en cuanto a la inversión necesaria para atender correctamente a sus clientes y le garantiza que la misma ha sido calculada de manera profesional.

En función de las métricas evaluadas anteriormente, se concluye que el modelo original presenta resultados muy alentadores, que motivan a la utilización de herramientas más sofisticadas para la realización de los horarios del personal dentro de las estaciones de servicio. De esta manera, consigue ser un modelo que combina perfectamente la teoría y la práctica, consiguiendo tiempos de espera mínimos al respetar los supuestos previstos en la teoría de colas, a la vez que obedece todo requisito legal que hace a la operatoria en una estación de servicio. Representa un evidente avance para la industria.

La Figura N°46 resume los resultados obtenidos por el modelo original en el mes bajo análisis en cuanto a la utilización del sistema, el tiempo de espera para ser atendido y la cantidad de personal necesario para ello.

Figura N°46: Resultados modelo original

UTILIZACION			TIEMPO DE ESPERA		CANTIDAD DE PERSONAL
PROMEDIO	> 80%	> 100%	PROMEDIO	> 5 MIN	
43.41%	3.13%	0.15%	0.92 min	1.93%	8 empleados

3.3 Modelo ajustado

Dadas las características de demanda del mes de febrero, se identificó durante el proceso de modelado que el modelo original puede ser exigido un 5% más, surgiendo de esta manera un modelo ajustado que consigue mejores resultados sin invertir recursos adicionales. Ahora, en lugar de establecer el nivel máximo de utilización del sistema en 80%, el mismo es fijado en 76%.

El modelo ajustado es un modelo que, al igual que el original, busca minimizar la cantidad de personal necesario para responder a la demanda predicha por el algoritmo *Gradient*

Boosting Machine, respetando en todo momento las restricciones planteadas por la Ley de Contrato de Trabajo vigente en la República Argentina, el Convenio Colectivo de Empleados de Estaciones de Servicio y ciertas consideraciones que hacen a las buenas prácticas empresariales.

El modelo matemático del modelo ajustado es idéntico al del modelo original, excepto por una restricción. A continuación, se detalla la única diferencia que presenta.

Sujeto a:

$$\bullet \quad (1) A_h \leq 0.76 * \sum_{e \in E} \sum_{h-7}^h Y_{eh} * 37.02 \quad \forall h \in H$$

(1) La capacidad de atención al público de la empresa, la cual es de 37.02 autos por empleado por hora, no puede superar el 76% de utilización.

La planilla de horarios del modelo ajustado presenta las mismas características de formato y modelado que la del modelo original, con la única diferencia de tolerar una utilización máxima del sistema de hasta 76% inclusive. Esto genera que los turnos de los empleados deban ser reorganizados de una manera más eficiente que la original, buscando que la demanda predicha por el algoritmo *Gradient Boosting Machine* nunca exija al sistema más allá del umbral mencionado anteriormente.

Al igual que en el modelo original, para arribar al resultado fue necesario aplicar una heurística de tiempo máximo de resolución, obteniendo así un *gap* de 0.47% con respecto al resultado óptimo.

En la Figura N°47 puede verse en detalle el horario de ingreso de cada uno de los empleados que el modelo ajustado determina que son necesarios para satisfacer la demanda estimada.

Figura N°47: Output modelo ajustado

	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
EMPLEADO 1	0	0	6	14	14	14	F	0	0	6	14	14	14	F	0	6	14	14	14	14	F	0	6	14	14	14	14	F
EMPLEADO 2	8	16	16	16	16	F	8	8	16	16	16	16	F	8	8	16	16	16	F	8	8	8	16	16	16	F	8	8
EMPLEADO 3	14	F	0	0	0	0	0	14	F	0	0	0	0	0	F	0	0	0	0	0	0	F	0	0	0	0	0	0
EMPLEADO 4	8	8	8	8	8	8	F	8	8	8	8	8	8	F	8	8	8	8	8	8	F	12	8	8	8	8	8	F
EMPLEADO 5	16	16	16	16	F	16	16	16	16	16	16	F	16	16	16	16	16	F	16	16	16	16	16	16	F	16	16	16
EMPLEADO 6	F	8	8	8	8	8	8	F	8	8	8	8	8	8	F	8	8	8	8	8	8	F	8	8	8	8	8	8
EMPLEADO 7	6	6	F	6	6	6	6	6	6	F	6	6	6	6	6	F	6	6	6	6	6	6	F	6	6	6	6	6
EMPLEADO 8	16	16	16	F	16	16	16	16	16	16	F	16	16	16	16	16	F	16	16	16	16	16	16	F	16	16	16	16

Para identificar el mejor modelo aplicable para el mes de febrero del año 2021, se procede a evaluar la performance del modelo ajustado en función de estas tres métricas que se vienen contemplando a lo largo de todo el trabajo: tiempo de espera en cola, utilización del sistema y cantidad de personal. Estos indicadores serán contrastados con los del modelo original para considerar más claramente los pros y contras de implementar el modelo ajustado en lugar del modelo original.

En la Figura N°48 se hace una comparación entre la performance del modelo original y del modelo ajustado, detallando el tiempo de espera promedio a cada hora en azul y en rojo, respectivamente.

Desde la 00:00 horas hasta las 05:00 horas inclusive, ambos modelos demuestran resultados prácticamente idénticos, siendo ligeramente superiores los del modelo original. Sin embargo, a partir de las 06:00 horas hasta las 15:00 horas inclusive, el modelo ajustado hace valer la reorganización de los horarios realizada y presenta tiempos de espera promedio mucho más competitivos que los del modelo original, alcanzando diferencias de hasta más de un minuto entre los valores de un modelo y otro.

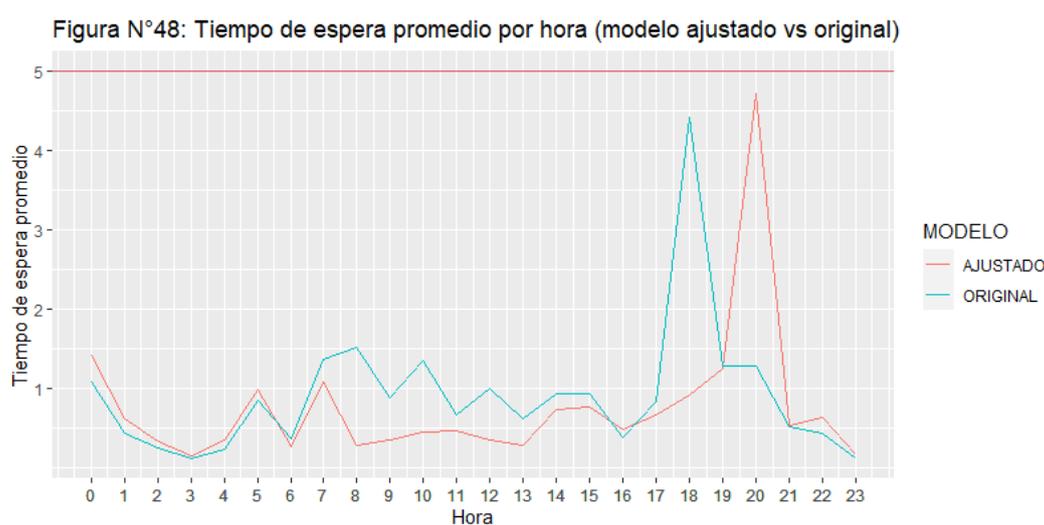
En el resto de los horarios del día las diferencias entre modelos son insignificantes, salvo a las 18:00 horas y a las 20:00 horas, que es cuando los modelos presentan sus mayores falencias. A las 18:00 horas el modelo original exhibe un pico de casi 4.50 minutos, que es el horario en el que el sistema se satura por única vez en el mes, no siendo así el caso del modelo ajustado, que presenta un tiempo de espera promedio de aproximadamente 1 minuto en ese mismo horario. No obstante, a las 20:00 horas sucede lo contrario, mientras que el modelo original presenta un valor ligeramente superior a 1 minuto, el tiempo de espera promedio del modelo ajustado supera por poco los 4.50 minutos.

Esto se debe a que en la reorganización de horarios el modelo ajustado ha sabido sortear la dificultad de las 18:00 horas, pero ha acabado viéndose saturado una única vez en el mes a las 20:00 horas, siendo penalizado por ello por un valor muy elevado que hace que el tiempo de espera promedio resulte en un número tan alto. En síntesis, le ocurrió lo mismo que al modelo original, pero en un horario distinto.

El tiempo de espera promedio del modelo ajustado se ha situado en 0.76 minutos a lo largo del mes bajo análisis, obteniendo una ventaja considerable con respecto al modelo original que consiguió atender a los clientes en 0.92 minutos en promedio. Asimismo, el modelo ajustado presenta tiempos de espera superiores a cinco minutos en solo el 0.74%

de los casos, a diferencia del modelo original que llega a 1.93%. Esta diferencia de rendimiento se origina por la holgura adicional en la capacidad que el modelo ajustado tiene por sobre el modelo original, dado que en su planteo tolera capacidades máximas de hasta un 76%, permitiendo que el tiempo de espera no se dispare tan rápidamente en casos donde la predicción del algoritmo *Gradient Boosting Machine* haya subestimado la demanda.

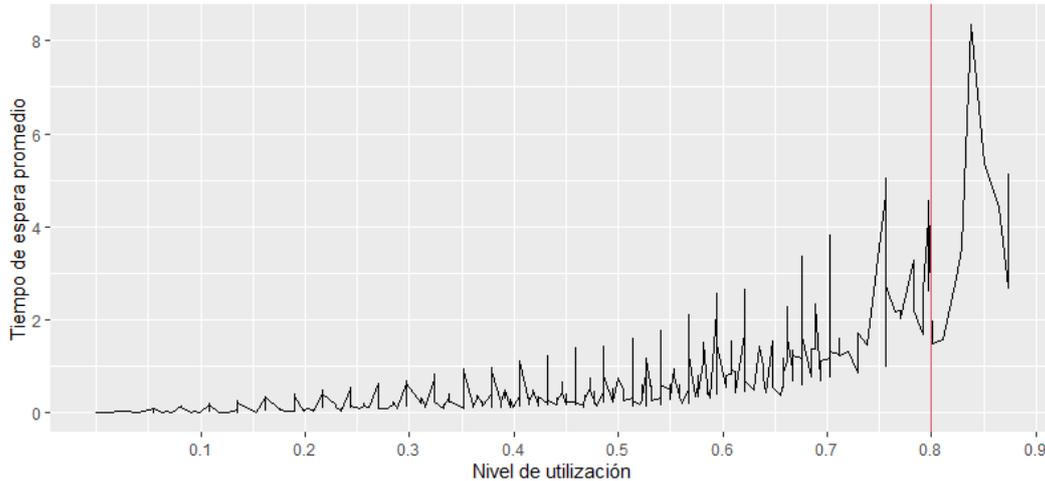
Sin perjuicio de lo mencionado anteriormente, es notable destacar que ninguno de los modelos presenta un tiempo de espera promedio superior a cinco minutos en cada una de las horas que componen al día.



La Figura N°49 muestra el rendimiento del modelo ajustado a lo largo de todo el mes de febrero del año 2021 ordenado por los distintos niveles de utilización que tuvo el sistema, excepto aquellos mayores a 100%, y el tiempo de espera promedio que se originó como consecuencia de ello. El tiempo de espera de las utilidades superiores a 100% no puede ser expuesto en este gráfico porque el mismo tiende a infinito, dado que se trata de un sistema inestable que, en realidad, no puede ser gestionado. En esos casos se asigna como tiempo de espera una penalidad de 100 minutos.

A partir del gráfico se puede visualizar que el modelo ajustado presenta una muy buena gestión de los tiempos de espera en cola a lo largo de todo el mes, superando ligeramente en una única oportunidad los ocho minutos como tiempo de espera promedio. De esta manera se puede comprender mejor cómo el modelo arriba a tiempos de espera promedio en la Figura N°48 tan bajos.

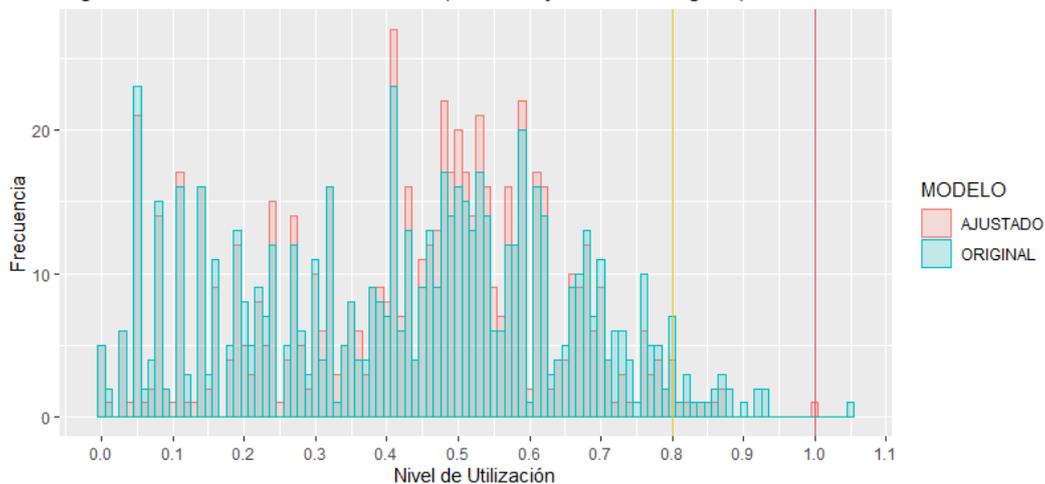
Figura N°49: Impacto del nivel de utilización sobre el tiempo de espera (modelo ajustado)



La Figura N°50 detalla la frecuencia de los distintos niveles utilización del modelo original y del modelo ajustado, los cuales se exhiben en azul y rojo, respectivamente. El modelo ajustado demuestra realizar una asignación más eficiente de los recursos que el modelo original, concentrando los niveles de utilización especialmente entre el 40% y el 60%. Esto permite liberar al sistema de niveles de utilización superiores que concluyen inevitablemente en tiempos de espera elevados, empeorando la calidad de atención del establecimiento.

El nivel promedio de utilización del sistema ha sido de 42.52% para el modelo ajustado, superando al modelo original que se ubicaba en 43.41%. Asimismo, el modelo ajustado superó el 80% de utilización en solo el 1.49% de las horas del mes, a diferencia del modelo original que lo hizo en el 3.13% de éstas. Asimismo, ambos modelos superaron el 100% de utilización en una oportunidad, lo que representa el 0.15% del total.

Figura N°50: Utilización del sistema (modelo ajustado vs original)



Como ya se ha mencionado anteriormente, el personal en una estación de servicio es clave a la hora de brindar una atención de calidad al cliente, pero a la vez, es la variable que más influye dentro de la estructura de costos de la empresa. Por ello, se debe ser muy cuidadoso al momento de determinar el equilibrio justo entre calidad de atención a brindar y cantidad de personal a contratar, priorizándose en este caso la calidad de atención que se brinda al público, aunque buscando minimizar la cantidad de personal necesaria para ello.

El modelo ajustado, al igual que el modelo original, requiere de ocho empleados para obtener la performance descrita a lo largo de este capítulo, siendo la única diferencia entre éstos la asignación de horarios que se ha realizado del personal. Es esa única, pero gran diferencia, lo que ha ocasionado las significativas diferencias de rendimiento entre cada propuesta de solución.

A merced de las métricas analizadas, el modelo ajustado ha evidenciado resultados muy superiores a los del modelo original, los cuales ya eran bastante alentadores. Exigir al modelo de optimización al máximo de sus posibilidades en función de la demanda predicha para el mes de febrero, ha permitido que, sin realizar una inversión adicional de personal, se realice una asignación mucho más eficiente que la anterior. De esta manera, se consigue que el establecimiento obtenga las mejores métricas de performance posibles para el mes bajo análisis.

Este modelo ha demostrado adaptarse a medida a las necesidades de la estación de servicio, yendo incluso más allá de lo que la teoría de colas considera como adecuado, a la vez que atiende a todo requisito legal que hace a la operatoria en una estación de servicio. Definitivamente, es el planteo más adecuado en caso de que el objetivo del empresario sea minimizar el tiempo de espera en cola de los clientes del negocio, ya que sin necesidad de invertir recursos adicionales puede organizar al personal de una manera más eficiente que el modelo original.

Asimismo, es notable destacar que no es posible obtener una mejor relación entre calidad de atención público e inversión necesaria en personal que la determinada por el modelo ajustado. Para poder alcanzar mejores métricas de rendimiento, necesariamente se debería contratar personal adicional, incrementando los costos de la empresa más allá de los actuales sin justificación, dado que la performance conseguida es muy competitiva.

La Figura N°51 resume los resultados mencionados de las métricas analizadas para ambos modelos: utilización del sistema de atención al público, tiempo de espera en cola para ser atendido y cantidad de personal necesario.

Figura N°51: Comparación resultados modelo original y modelo ajustado

	UTILIZACION			TIEMPO DE ESPERA		CANTIDAD DE PERSONAL
	PROMEDIO	> 80%	> 100%	PROMEDIO	> 5 MIN	
Original	43.41%	3.13%	0.15%	0.92 min	1.93%	8 empleados
Ajustado	42.52%	1.49%	0.15%	0.76 min	0.74%	8 empleados

3.4 Modelo relajado

El modelo relajado surge como una alternativa más económica para las estaciones de servicio, la cual consigue ahorros en gastos de personal en detrimento de la calidad de atención al público. Para ello, este modelo admite una máxima utilización del sistema más relajada que los modelos anteriores, la cual se sitúa en 90%.

El objetivo de este modelo consiste en minimizar la cantidad de personal necesario para satisfacer la demanda predicha por el algoritmo *Gradient Boosting Machine*. Asimismo, contempla las restricciones planteadas por la Ley de Contrato de Trabajo vigente en la República Argentina, el Convenio Colectivo de Empleados de Estaciones de Servicio y una serie de cuestiones que hacen a las buenas prácticas empresariales.

El modelo matemático del modelo relajado es idéntico al del modelo original, excepto por un conjunto y una restricción. A continuación, se detallan únicamente las diferencias que presenta.

Conjunto:

- (1) $E = \{1, \dots, 7\}$

(1) Empleados disponibles para atender al público.

Sujeto a:

- (1) $A_h \leq 0.90 * \sum_{e \in E} \sum_{h-7}^h Y_{eh} * 37.02 \quad \forall h \in H$

(1) La capacidad de atención al público de la empresa, la cual es de 37.02 autos por empleado por hora, no puede superar el 90% de utilización.

La planilla de horarios del modelo relajado ha sido confeccionada siguiendo las mismas características de formato y modelado que los modelos anteriores, con la diferencia que admite una utilización máxima del sistema de 90%. Consecuentemente, esta mayor

flexibilidad posibilita un planteo con una menor cantidad de empleados que los modelos anteriores, priorizando el ahorro en costos de personal antes que la calidad de atención al público.

Para arribar al resultado se ha apelado a una heurística consistente en establecer un tiempo máximo de resolución de cinco minutos, consiguiendo un *gap* de 0.57% con respecto al resultado óptimo.

La Figura N°52 especifica el horario de ingreso de cada uno de los empleados que deberían contratarse para atender correctamente al público respetando las restricciones establecidas.

Figura N°52: Output modelo relajado

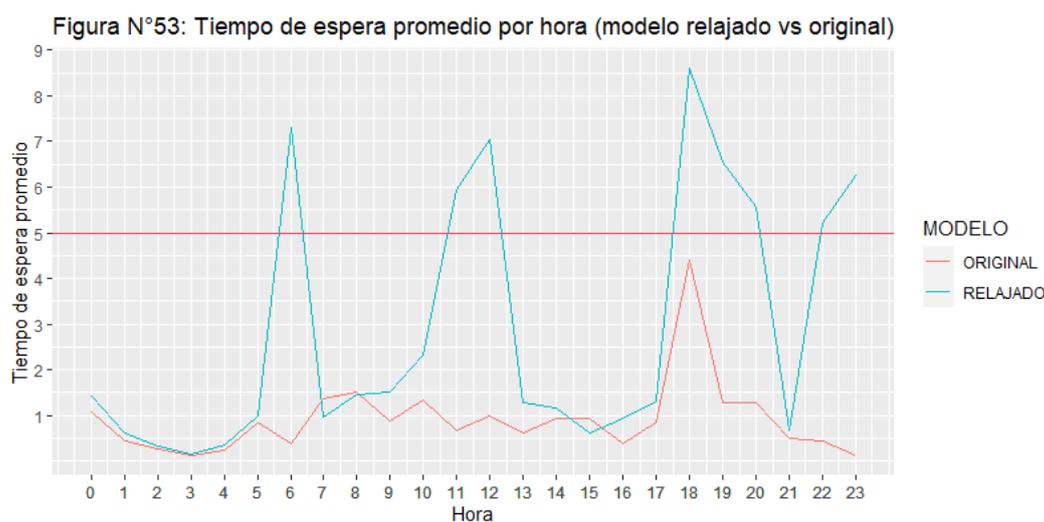
	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
EMPLEADO 1	F	7	7	7	7	7	7	F	7	7	7	7	7	7	F	7	7	7	7	7	7	F	7	7	7	7	8	8
EMPLEADO 2	0-22	F	14	14	14	14	F	0	0	15	15	15	15	F	6	15	15	15	15	15	F	6	16	16	16	F	6	6
EMPLEADO 3	6	6	6	22	22	22	F	6	6	6	23	23	23	F	6	6	6	23	23	23	F	6	6	6	23	23	23	F
EMPLEADO 4	14	14	F	14	14	14	16	16	F	14	14	14	14	22	22	F	14	14	14	14	22	22	F	14	14	15	14	22
EMPLEADO 5	F	15	15	15	15	15	15	F	15	15	15	15	15	15	F	15	15	15	15	15	15	F	15	15	15	15	15	15
EMPLEADO 6	8	22	22	F	8	8	8	8	22	22	F	8	8	14	14	22	22	F	8	8	14	14	22	F	0	14	14	14
EMPLEADO 7	14	14	F	6	6	6	6	14	14	F	6	6	6	6	14	14	F	6	6	6	6	14	14	F	6	6	6	6

Consistentemente a lo realizado con cada modelo, se procede a analizar la performance del modelo relajado en función del tiempo de espera en cola, utilización del sistema y cantidad de personal que requiere para ello. Asimismo, se compararán estos resultados con los del modelo original para poder contemplar más claramente los beneficios y perjuicios de aplicar este horario en lugar del primero planteado.

La Figura N°53 especifica el tiempo de espera promedio por hora del modelo relajado y del modelo original, los cuales se encuentran en azul y rojo, respectivamente. Desde las 00:00 horas hasta las 05:00 horas los modelos no presentan diferencias significativas, pero resulta evidente que el planteo realizado por el modelo relajado afecta de manera muy negativa la calidad de atención al público a lo largo del día, ya que a partir de las 06:00 horas los tiempos de espera comienzan a dispararse. Las principales falencias del modelo relajado se aprecian a las 06:00 horas, 11:00 horas, 12:00 horas, 18:00 horas, 19:00 horas, 20:00 horas, 22:00 horas y 23:00 horas, arrojando tiempos de espera promedio entre cinco y nueve minutos. Estos indicadores resultan bastante impactantes si uno considera que el modelo original no supera los cinco minutos de espera promedio

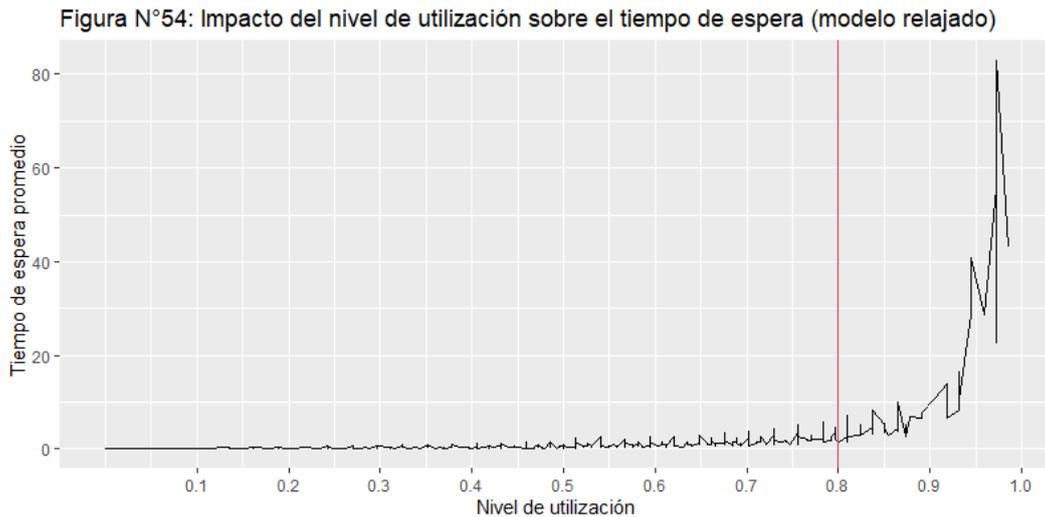
en ningún momento del día, siendo bastante superior al modelo relajado desde el punto de vista de la calidad de atención al público.

El tiempo de espera promedio para ser atendido del modelo relajado asciende a 2.86 minutos, un valor que triplica los 0.92 minutos que consigue el modelo original. Asimismo, a lo largo del mes, el modelo relajado supera la barrera de los cinco minutos de espera en el 6.25% de los casos, ocasionando como consecuencia que en un tercio del día sus tiempos de espera promedio se encuentren en valores tan elevados. Por otro lado, el modelo original lo hace en solo el 1.93% de las horas del mes, y en promedio nunca termina superando el umbral en cuestión.



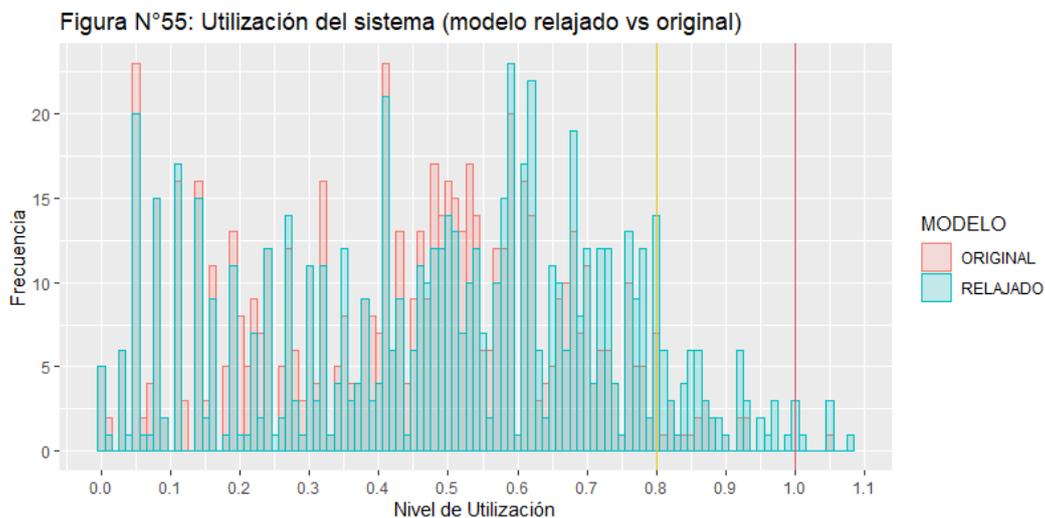
Para entender mejor por qué los tiempos de espera promedio a cada hora terminan siendo tan altos en el modelo relajado, se presenta la Figura N°54 que muestra el impacto que tiene en el tiempo de espera el hecho de tener niveles de utilización muy elevados. Nuevamente, se recuerda que los datos presentados corresponden a niveles de utilización inferiores a 100%, ya que cuando se supera este valor el sistema se vuelve inestable e imposible de gestionar, con tiempos de espera que tienden a infinito. En esos casos, se le asigna al modelo una penalidad de 100 minutos como tiempo de espera.

Al permitir niveles de utilización del sistema más relajados, y exigir tanto la capacidad de la empresa de atender al público, este modelo recorre completamente los niveles de utilización superiores a 80%, obteniendo tiempos de espera que llegan a superar los 80 minutos. Sin dudas, todas esas horas del mes que tienen tiempos de espera tan elevados hacen que los valores promedio que se mostraron en la Figura N°53 destaquen de una manera tan negativa, y sitúen al modelo relajado en un escalón inferior que el modelo original en cuanto a la calidad de atención al público.



La Figura N°55 precisa la frecuencia de los distintos niveles de utilización que el modelo relajado desarrolla a lo largo del mes, las cuales son comparadas con las del modelo original, y se muestran en azul y rojo, respectivamente. Tal como era de esperar, el modelo relajado se destaca por sobre el modelo original en los niveles de utilización más elevados, superándolo en prácticamente todos los niveles desde el 60% en adelante. Esto es porque al priorizar contratar menos personal y hacer un uso más eficiente de sus recursos, el sistema se ve saturado más rápidamente, derivando también en tiempos de espera superiores como se vio anteriormente.

La utilización promedio del modelo relajado ha sido superior a la del modelo original, la cual se ubica en 49.82% y 43.41%, respectivamente. A su vez, el modelo relajado supera el 80% de utilización en el 8.78% de las 672 horas del mes, y el 100% en el 0.74%, mientras que el modelo original lo hace en el 3.13% y 0.15%, respectivamente.



Con respecto a la cantidad de personal, es en esta métrica en la que el modelo relajado se diferencia positivamente del modelo original y del modelo ajustado, ya que justamente en el planteo se resigna calidad de atención al público para poder disminuir el costo que implica contratar tanto personal. Es una solución que prioriza la eficiencia en el uso de los recursos por sobre todas las cosas. A diferencia del modelo original y del modelo ajustado, el modelo relajado requiere solamente de siete empleados para atender al público con las implicancias que se han comentado, lo que significa un ahorro en personal de 12.50%.

En función de lo analizado, el modelo relajado ha dejado en evidencia el enorme impacto que tiene la decisión de la cantidad de personal a ser contratado en la calidad de atención al público, ya que solo por contratar una persona menos que los modelos anteriores, los indicadores de calidad de atención han sido comparativamente mucho peores. Evidentemente, la disyuntiva entre calidad de atención y costo de personal a la que se ve sometida todo estacionero es una problemática muy sensible a la disposición de las variables.

A pesar de que el modelo relajado no respeta de manera estricta lo que plantea la teoría de colas en cuanto a la máxima utilización a la que debe ser sometido un sistema para disminuir los tiempos de espera, es un planteo totalmente válido en la práctica que debe ser considerado. Especialmente, dada la reducción de personal que consigue, el cual es el gasto más importante dentro de la estructura de costos de una estación de servicio.

La Figura N°56 contiene un resumen comparativo entre el modelo relajado y el modelo original con respecto a la utilización del sistema, el tiempo de espera para ser atendido y la cantidad de personal necesario para ello.

Figura N°56: Comparación resultados modelo original y modelo relajado

	UTILIZACION			TIEMPO DE ESPERA		CANTIDAD DE PERSONAL
	PROMEDIO	> 80%	> 100%	PROMEDIO	> 5 MIN	
Original	43.41%	3.13%	0.15%	0.92 min	1.93%	8 empleados
Relajado	49.82%	8.78%	0.74%	2.86 min	6.25%	7 empleados

3.5 Propuesta de solución

A lo largo de este capítulo se han probado distintos modelos para determinar la cantidad de personal adecuado para el mes bajo análisis y su óptima asignación de turnos, para poder así atender la demanda de la estación de servicio de la mejor manera posible. Cada uno de ellos ha presentado distintas características en cuanto a la calidad de atención

brindada y su costo relacionado, pero siempre respetando todos los requisitos legales necesarios para poder ser aplicados en la realidad.

Para disponer cuál es el mejor modelo de todos es fundamental tener en cuenta que la carga de combustibles no es algo que el cliente haga por placer, sino que es algo que éste debe hacer para poder continuar con su día, y quiere hacerlo lo más rápido posible. En función de ello, es de vital importancia considerar el tiempo que le va a demorar este servicio desde que llega al establecimiento hasta que se retira del mismo.

En este trabajo, el tiempo de servicio no es una variable que se encuentre bajo análisis en vísperas de ser optimizada, pero sí el tiempo de espera para ser atendido, que es el que se consideró a cada momento. Por lo que, para hacer de la experiencia de los clientes la mejor posible, el modelo que debe considerarse es aquel que reduzca su tiempo de espera lo máximo posible en cada momento del día y, por ende, el tiempo que el consumidor va a permanecer en el establecimiento.

Dicho todo esto, se considera que el mejor modelo a aplicar en la estación de servicio es el modelo ajustado, ya que justamente es el que mejor calidad de atención brinda entre todos los modelos.

3.6 Análisis de resultados

En esta sección se procede a efectuar una comparación de rendimiento entre el mejor modelo que se pudo confeccionar, mediante la programación de horarios realizada en función de las predicciones llevadas a cabo por el algoritmo *Gradient Boosting Machine*, y la planificación manual realizada por la empresa. El objetivo es determinar si una programación de horarios realizada de manera profesional puede contribuir positivamente a la organización de la empresa y, eventualmente, reemplazar los procedimientos de trabajo actuales con ánimos de mejorar la calidad de atención y reducir los costos.

Antes de comenzar con el análisis, se recuerda que tanto el modelo manual (línea base), como el modelo profesional (modelo ajustado), han sido confeccionados en igualdad de condiciones. Es decir, ambos fueron preparados con información desde enero del año 2018 hasta enero del año 2021 inclusive, buscando con ella realizar el mejor planteo posible para el mes de febrero del mismo año.

Se utilizarán las mismas métricas que han sido empleadas a lo largo de este trabajo, y que se caracterizan por poner foco en los aspectos más relevantes a tener en cuenta a la hora de armar una estructura de atención al público:

- Tiempo de espera en cola
- Utilización del sistema
- Cantidad de personal

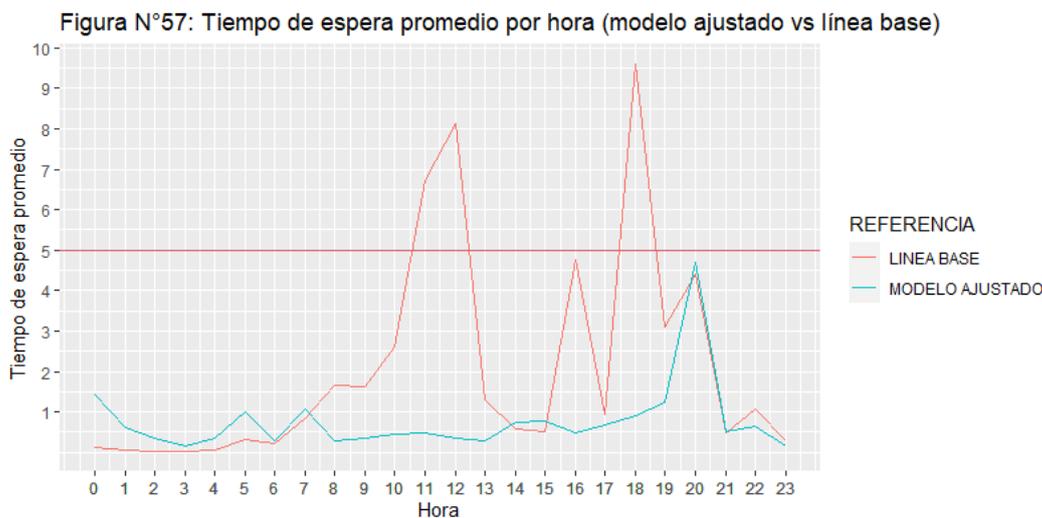
La Figura N°57 detalla el tiempo de espera promedio para ser atendido a cada hora del día para el modelo ajustado y para la línea base, los cuales se encuentran en azul y rojo, respectivamente. Ya a primera vista se evidencian diferencias muy significativas entre modelos, mostrando resultados sorprendentes del modelo profesional con relación al modelo manual. La capacidad del modelo ajustado de mantener el tiempo de espera promedio de los clientes por debajo del minuto y medio a lo largo de prácticamente todo el día, genera un impacto visual muy fuerte, que se agiganta al contrastarlo con los resultados conseguidos por la línea base.

Específicamente, en modelo ajustado consigue diferencias significativas con respecto a la línea base en la franja horaria de 08:00 horas a 13:00 horas, sorteando exitosamente el pico de las 11:00 horas y 12:00 horas que padece el modelo manual. De igual manera, a las 16:00 horas, 18:00 horas y 19:00 horas, el modelo profesional asigna al personal perfectamente para poder evitar sin dificultades los cuellos de botella que se generan con la planificación manual en estos horarios. El resto de los horarios del día se desarrollan sin diferencias elocuentes, dado que, en general, no representan momentos de dificultad para la empresa. El único horario en el que la línea base presenta ciertas dificultades y el modelo propuesto no pudo superarlo, sino igualarlo, es a las 20:00 horas.

De más está decir que, a diferencia de la línea base, el modelo propuesto no supera en ningún momento del día los cinco minutos de espera promedio, los cuales, en caso de superarse, significan una amenaza para la atención al público de calidad. El momento del día de mayor espera del modelo propuesto es a las 20:00 horas con poco más de 4.50 minutos en promedio, siendo el segundo mayor tiempo a las 00:00 horas con tan solo 1.50 minutos en promedio.

El tiempo de espera promedio a lo largo del mes comparativo entre modelos resulta notable, consiguiendo el modelo profesional mejoras superiores al 60% con respecto al modelo manual, siendo el tiempo alcanzado por cada uno de ellos de 0.76 minutos y 1.91 minutos, respectivamente. Consecuentemente, al manejar tiempos de espera promedio inferiores, el modelo ajustado solo hace esperar a los clientes más de cinco minutos en el

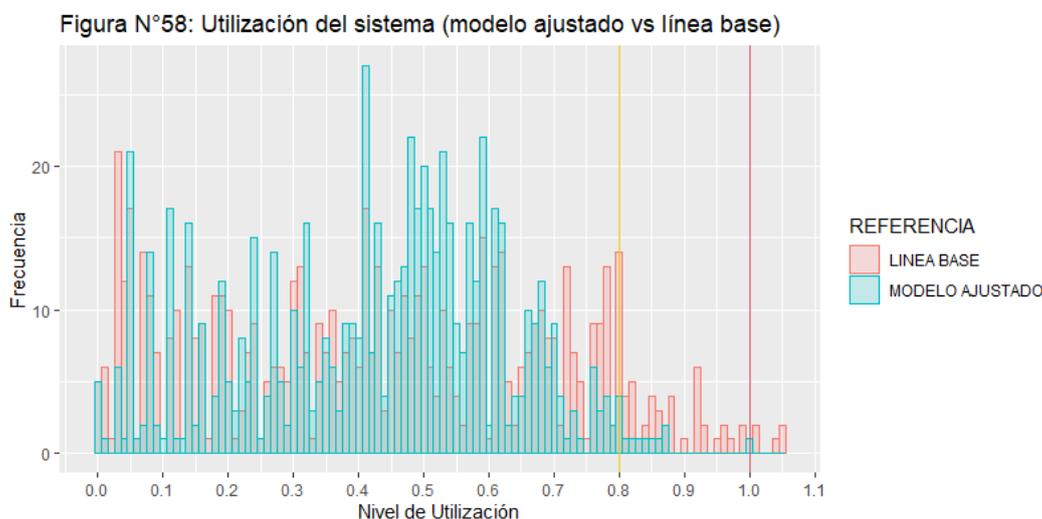
0.74% de casos, a diferencia del horario manual que lo hace en el 4.46%. Esto implica una mejora de casi el 85% en este sentido.



La Figura N°58 compara la utilización del sistema del modelo ajustado y de la línea base, detallándose cada uno en azul y rojo, respectivamente. Asimismo, se destaca mediante una línea vertical naranja y otra roja la utilización igual a 80% y 100%, para poder identificar más rápidamente las zonas de saturación del sistema. Rápidamente logra identificarse un contraste claro entre modelos, situándose el modelo propuesto prácticamente en su totalidad por debajo de la línea naranja, mientras que el modelo manual se destaca justamente por exigir más al sistema, llevándolo a valores que rondan el 80% de utilización, o más, en reiteradas oportunidades. La versatilidad del modelo ajustado lo lleva a reconocer los momentos de mayor demanda, destinando mayor cantidad de recursos en ese horario, permitiendo así que la estación de servicio disponga de la capacidad justa para el momento justo, concentrando de esta manera los niveles de utilización entre 40% y 60%. En cambio, la rigidez del modelo manual hace que, si se mira el gráfico en detalle, la estación de servicio tenga una frecuencia de cada nivel de utilización bastante parejo a lo largo de toda la escala hasta el 80%, disminuyendo la misma, pero volviendo a ser ciertamente uniforme de 80% a 100%.

Precisamente, la utilización promedio del modelo profesional fue casi idéntica a la de la línea base, ubicándose en 42.52% contra 43.24%, respectivamente. Sin embargo, es en la cantidad de veces que cada sistema se ve sobre exigido por la demanda en donde surge la gran diferencia entre ambos, lo cual se debe a la flexibilidad y capacidad de anticipación de la demanda que posee el modelo propuesto. La capacidad de atención al público del modelo ajustado supera el 80% solamente en el 1.49% del mes, y el 100% en el 0.15%,

los cuales son los mejores valores que se han conseguido en este trabajo, mientras que la línea base lo hace en el 6.85% y 0.74%, respectivamente. En síntesis, el modelo ajustado mejora en casi un 80% la cantidad de veces que el sistema supera el 80% de utilización, y prácticamente mitiga los colapsos de éste.



Con respecto a la cantidad de empleados, resulta sorprendente que todas las mejoras que se han visto en los dos gráficos anteriores se hayan conseguido con menos personal del que ha utilizado la empresa para confeccionar lo que se tomó como línea base, pero afortunadamente es cierto. La capacidad del modelo propuesto de adelantarse a la demanda futura, y asignar específicamente al personal en función de ella, permite ser mucho más eficiente en su utilización.

Asimismo, esto alivia enormemente la presión de los estacioneros de definir la cantidad de personal necesario para atender a los clientes, ya que, ya sea por exceso o por escasez de éstos, siempre se termina perjudicando a algún agente de interés de la empresa. Cuando la empresa contrata empleados de más, son los socios los que se ven perjudicados al ver una menor rentabilidad de la compañía sin justificación alguna, mientras que cuando se disponen menos empleados de los necesarios, son los clientes los primeros que notan un detrimento en la calidad de atención. Encontrar el punto justo es algo realmente difícil de hacer mediante intuición, por no decir imposible, por lo que esta herramienta puede ayudar enormemente en la tarea.

El modelo realizado mediante técnicas profesionales requiere únicamente de ocho empleados para atender al público, a diferencia del modelo manual que se utiliza actualmente en la estación de servicio, que requiere de 8.5 empleados. Es importante aclarar que la empresa, en realidad, se encuentra abonando nueve sueldos, pero dado que

un empleado estuvo justo la mitad del mes de vacaciones, es que se computa medio empleado en lugar de uno entero.

En síntesis, queda en evidencia que la rigidez y la incorrecta asignación de personal fueron las falencias más graves que se le han notado al modelo manual cuando se lo analizó en detalle al comienzo de este trabajo. Inevitablemente, estas características derivan en todo lo que un administrador de una estación de servicio debería evitar, que es ineficiencia en la utilización de recursos y mala calidad de atención al público. Tal vez, esto haya sido una tarea casi imposible hace décadas atrás, pero hoy en día, los avances del conocimiento permiten combinar distintas herramientas para obtener los excelentes resultados que se han visto. La predicción de la demanda a través de la utilización de algoritmos y la asignación de personal mediante la utilización de un *solver* debería comenzar a convertirse en una herramienta competitiva en el mercado, para volverse con los años un supuesto básico de la industria.

La Figura N°59 confronta los resultados entre el modelo ajustado y la línea base desde la perspectiva del nivel de utilización del sistema, el tiempo de espera de los clientes y la cantidad de personal requerido por cada modelo. Sintetizando, en definitiva, la facultad del modelo profesional de atender mejor a los clientes con una menor cantidad de recursos.

Figura N°59: Comparación resultados modelo ajustado y línea base

	UTILIZACION			TIEMPO DE ESPERA		CANTIDAD DE PERSONAL
	PROMEDIO	> 80%	> 100%	PROMEDIO	> 5 MIN	
Modelo Ajustado	42.52%	1.49%	0.15%	0.76 min	0.74%	8 empleados
Línea Base	43.24%	6.85%	0.74%	1.91 min	4.46%	8.50 empleados

CONCLUSIONES

La pregunta principal de este trabajo se basó en cómo hacer para predecir la demanda y optimizar los turnos de trabajo en una estación de servicio, teniendo en cuenta las implicancias de esto sobre el tiempo de espera del cliente, la utilización del sistema de atención al público y la cantidad de personal necesario para ello. Para responder esta pregunta, se trabajó con una estación de servicio ubicada en San Justo, provincia de Buenos Aires, con el objetivo de desarrollar una herramienta de planificación de horarios de personal, de manera tal que pueda asegurarse una óptima calidad de atención al público con la menor cantidad de personal posible.

Para ello, fue necesario llevar a cabo un análisis previo de los datos históricos de la empresa. Basado en la teoría de colas, se analizó cuál era el modelo aplicable a las características de la estación de servicio que permita calcular el tiempo de espera para ser atendido a cada hora, y se determinó la tasa de servicio de cada playero. Asimismo, se emplearon los datos históricos para determinar la performance de la compañía realizando los horarios de manera manual en el mes de febrero del año 2021, situación que fue considerada como línea base para las comparaciones futuras.

Luego, se efectuó la predicción de arribos mediante dos de los algoritmos más potentes de la actualidad, *Random Forest* y *Gradient Boosting Machine*, seleccionando el mejor de ellos para realizar la predicción utilizada en la programación de turnos. De esta manera, se tomó conocimiento del último dato necesario para planificar el horario de personal de manera profesional, la tasa de arribos, la cual se estima con un error de solo 7.58 autos por hora.

A partir de estos datos y otros obtenidos de fuentes externas, tales como la Ley de Contrato de Trabajo, el Convenio Colectivo de Empleados de Estaciones de Servicio y ciertas prácticas habituales en el rubro, se programaron modelos de horarios con diversas características, los cuales fueron comparados con la línea base. El mejor de ellos, el modelo ajustado, redujo en un 60% el tiempo de espera promedio, en un 85% la cantidad de veces que el tiempo de espera es superior a los cinco minutos y en un 80% los niveles de utilización superiores a 80%. Todo ello, reduciendo también la cantidad de personal necesario para atender al público en un 6%.

Por consiguiente, queda demostrado el enorme impacto de migrar de una planificación de horarios manual a una profesionalizada, habiendo conseguido mejoras significativas en

todas las métricas consideradas en este trabajo. En la actualidad, seguramente haya más de un estacionero pagando por sueldos de empleados que no necesita, perdiendo clientes por brindarles una mala atención y/o padeciendo las consecuencias de un mal clima laboral debido al malestar que la imprevisibilidad de las prácticas actuales genera en los mismos empleados. Por ello, acompañar el cambio, adaptando a la empresa a las novedades tecnológicas que van sucediendo con el correr de los años, es necesario tanto para optimizar el rendimiento de la inversión como para garantizar la supervivencia de la empresa en el mercado.

La programación de horarios realizada en este trabajo puede ser llevada a cabo por cualquier empresa que así lo desee, ya que el código de ZIMPL facilitado en el Anexo I puede ser utilizado por cualquier estación de servicio que se encuentre sujeta a las mismas normas legales que la analizada. Sin embargo, es importante tener en cuenta que la predicción de arribos sí es específica a las características de cada empresa, por lo que cada compañía debería realizar la propia. Para facilitar la tarea, se puede encontrar en el Anexo II el código en R utilizado en este trabajo para realizar la predicción del mes de febrero del año 2021. Tener en cuenta que éste será útil luego de haber realizado la ingeniería de atributos y la optimización de hiperparámetros correspondiente.

Se invita a aquel que lo desee a seguir explorando en esta área, utilizando nuevos algoritmos para predecir los arribos de clientes y mejorar las métricas de error conseguidas en este trabajo, o incluso a plantear modelos de optimización distintos al considerado. Podría analizarse la posibilidad de minimizar el tiempo de espera global, en lugar de la cantidad de empleados, o asignarles un valor monetario a todas las variables bajo análisis y armar una función objetivo que lo minimice. Solo por dar algunos ejemplos.

Más allá de las variantes y mejoras que se le puedan hacer a este trabajo, lo que sí queda claro es que la planificación manual de horarios es una práctica que debe ser reemplazada de manera urgente, siendo la planificación profesional de turnos la única manera de conseguir una mejora significativa en la calidad de atención al público a la vez que se reducen los costos de la compañía. El momento de cambiar es ahora. ¡Adelante!

BIBLIOGRAFIA

Aydas, O.; Ross, A.; Scanlon, M.; Aydas, B. (2020). New results on integrated nurse staffing and scheduling: The medium-term context for intensive care units. *Journal of the Operational Research Society* (en prensa). Recuperado de: <https://orsociety.tandfonline.com/doi/full/10.1080/01605682.2020.1806742#.YGo9qOhKg2w>.

Beaumont, N. (1997). Sheduling staff using mixed integer progamming. *European Journal of Operational Research*. Vol N°98. Pp 473-484. Recuperado de: <https://www.sciencedirect.com/science/article/abs/pii/S0377221797000556>.

Bengio, Y.; Courville, A.; Vincent. P. (2013). Representation Learning: A Review and New Perspectives. *IEEE Transactions on Software Engineering*. Vol N°35. Pp 1798-1828. Recuperado de: https://www.researchgate.net/publication/240308775_Representation_Learning_A_Review_and_New_Perspectives.

Bergstra, J.; Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*. Vol N°13. Pp 281-305. Recuperado de: <https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf>.

Breiman, L. (2001). Random Forests. *Kluwer Academic Publishers*. Vol N°45. Pp 5-32. Recuperado de: <https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf>.

Friedman, J (1999). Stochastic Gradient Boosting. *Computational Statistics & Data Analysis*. Vol N°38. Pp 367-378. Recuperado de: <https://www.sciencedirect.com/science/article/abs/pii/S0167947301000652>.

Kenneth, R. (1974). Scheduling Full-Time and Part-Time Staff to Meet Cyclic Requirements. *Journal of the Operational Research Society*. Vol N°25. Pp 65-76. Recuperado de: <https://orsociety.tandfonline.com/doi/abs/10.1057/jors.1974.7#.YHGLIehKg2w>.

Mudra, S. (2012). Staff Scheduling in Health Care Systems. *IOSR Journal of Mechanical and Civil Engineering (IOSRJMCE)*. Vol N°1. Pp 28-40. Recuperado de: <https://docplayer.net/11329988-Staff-scheduling-in-health-care-systems.html>.

Sakasegawa, H. (1977). An Aproximation Formula. *Ann. Inst. Statist. Math*. Vol N°29. Pp 67-75. Recuperado de:

<https://docs.google.com/viewer?a=v&pid=forums&srcid=MDg5NjE4Njk5NjQ0MTI0NzAwNzUBMDAzNDMzNzA3MTkwNzM4NjE0NTIBQ2lwQ05STXRDZ0FKATAuMQEBdjI>.

Thornton, J.; Sattar, A. (1997). Nurse Rostering and Integer Programming Revisited. *International Conference on Computational Intelligence and Multimedia Applications*. Griffith University Gold Coast, Australia, 49-58. Recuperado de: <https://www.semanticscholar.org/paper/Nurse-Rostering-and-Integer-Programming-Revisited-Thornton/9d0b5eb5749a6699a4c43bf8084d02113454b40b#citing-papers>.

Williams, P. H. (2013). *Model Building in Mathematical Programming*. (5th.ed.). United Kingdom: Wiley.

Winston, W. L. (2004). *Operations Research: Applications and algorithms*. (4th.ed.). United States: Thomson Learning.

Zheng, A.; Casari, A. (2018). *Feature Engineering for Machine Learning*. (1st.ed.). United States of America: O'Reilly.

ANEXO I: CODIGO ZIMPL (MODELO ORIGINAL)

```

##### CONJUNTOS #####

# Conjunto de empleados

set E := {1 .. 8};

# Conjunto de horas en el mes

set H := {0 .. 671};

# Conjunto de horas prohibidas

set H0 := {read "Noingreso.txt" as "<1n>" skip 1};

##### PARAMETROS #####

# Arribos

param a[H] := read "Arribos.csv" as "<3n> 4n" skip 1;

##### VARIABLES #####

# Asignación del empleado E a la tarea

var x[E] binary;

# Asignación del empleado E a la hora H

var y[E*H] binary;

# Asignación de premios

var z[E*H] binary;

##### FUNCION OBJETIVO #####

# Minimizar la cantidad de empleados necesarios

minimize fobj: sum <e> in E: x[e] - sum <e,h> in E*H: 0.001 * z[e,h];

##### RESTRICCIONES #####

# Capacidad al 80% o menos

subto cap: forall <h> in H:

    a[h] <= 0.80 * (sum <e,hp> in E*H with h-7 <= hp and hp <= h: y[e,hp] * 37.02);

# Turnos mensuales por empleado

```

subto turnos: forall $\langle e \rangle$ in E:

$$\text{sum } \langle h \rangle \text{ in H: } y[e,h] == 24 * x[e];$$

12 horas entre turnos

subto descansodiarario: forall $\langle e,h \rangle$ in E*H:

$$\text{sum } \langle hp \rangle \text{ in H with } h-(12+8-1) \leq hp \text{ and } hp \leq h: y[e, hp] \leq 1;$$

Cantidad máxima de días sin descanso (6 días)

subto descansosemanal: forall $\langle e,h \rangle$ in E*H:

$$\text{sum } \langle hp \rangle \text{ in H with } h-(168-1) \leq hp \text{ and } hp \leq h: y[e, hp] \leq 6;$$

Ingreso prohibido

subto noingreso: forall $\langle e, h0 \rangle$ in E*H0:

$$y[e,h0] == 0;$$

Horario estable (1/2)

subto mismoingreso1: forall $\langle e,h \rangle$ in E*H:

$$\text{sum } \langle hp \rangle \text{ in H with } h-24 \leq hp \text{ and } hp \leq h-24: y[e, hp] \geq z[e,h];$$

Horario estable (2/2)

subto mismoingreso2: forall $\langle e,h \rangle$ in E*H:

$$z[e,h] \leq y[e,h];$$

ANEXO II: CODIGO EN R (*GRADIENT BOOSTING MACHINE*)

```
library(gbm)
library(ggplot2)

# ENTRENAMIENTO CON MEJORES HIPERPARAMETROS EN TRAIN + VAL #
gbm.reg.train = gbm(ARRIBOS~.,
  data = train,
  distribution = "gaussian",
  n.trees = best.M.trees_2,      # M (overfitting)
  shrinkage = best.Shrinkage_2,  # lambda (regularización)
  interaction.depth = best.Interaction.depth_2,  # Complejidad de cada árbol.
  n.minobsinnode = best.N.minobsinnode_2,  # Otro parámetro de complejidad.
  train.fraction = 1,    # delta
  bag.fraction = best.Bag.fraction_2,  # eta (Bagging sobre la fracción de train)
  cv.folds = 5,
  verbose = T)

best.iter = gbm.perf(gbm.reg.train,method="cv")

# ERROR CUADRATICO MEDIO SOBRE DATOS DE TESTEO #
gbm.pred.test = predict(gbm.reg.train, n.trees = best.iter, newdata = Test)
gbm.ECM.test = sum((gbm.pred.test - Test$ARRIBOS)^2) / nrow(Test)
gbm.ECM.test # (out of sample)

# GRAFICO CON GGLOT2 #
ggplot(Test, aes(x = ARIBOS, y = gbm.pred.test)) + geom_point() +
  geom_abline(intercept = 0, slope = 1, color = "red", size = 1) + labs(title = "Figura N°39:
  Predicciones y observaciones GBM en datos testeo", x = "Observación", y =
  "Predicción")
```