

Informe Final: Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit.



Alumno: Emilio Rodrigo Nordenflycht Lübbert

Universidad: Universidad Adolfo Ibáñez

Carrera: Ingeniería Civil Industrial

Empresa y área: Falabella Flexit

Profesor: Fernando Vásquez Acuña

Resumen ejecutivo

This report details the development of a predictive model for job applications within Falabella stores using the Flexit application. The methodology involves data cleaning, model implementation, and evaluation, with a focus on iterative refinement over a three-month period.

The effectiveness of ARIMA and SARIMAX models has been demonstrated, particularly when incorporating exogenous variables like hours allocated to stores. Results reveal varying sensitivities of stores to the availability of hiring hours, suggesting that certain stores may have their labor needs met more efficiently through the application. The potential for stores to satisfy the full labor demand through Flexit rises as a possibility especially in the periods of beginning and end of year where the demographic of the app has more free time to work on part-time jobs.

The chosen model manages to return reliable results for most of the stores of Falabella, helping the team to determine the budget for hours for each store and the capacity to know if the app can satisfy the needs of the store in terms of personnel.

Looking ahead, continuous model refinement is recommended, involving the inclusion of more historical data and expansion to additional Falabella stores. Further exploration of new exogenous variables and assessing the model's adaptability to significant changes in the work environment are suggested for future iterations.

In summary, the project has successfully established a robust framework for predicting job applications, with results indicating the varying responsiveness of stores to offered hours. Future steps should prioritize ongoing model improvement, expansion to new stores, and enhanced adaptability to changing labor conditions.

Contenido

Resumen ejecutivo.....	2
Introducción.....	4
Contexto.....	4
Descripción del problema.....	6
Objetivos.....	7
Objetivo General.....	7
Objetivos específicos.....	7
Estado del arte.....	8
Solución escogida.....	12
Metodología.....	15
Metodología por objetivo.....	15
Metodología general.....	18
Riesgos y mitigaciones.....	18
Métricas de desempeño.....	20
Desarrollo.....	23
Trabajo preliminar.....	23
ARIMA.....	24
Ingresar data previa.....	27
SARIMAX.....	28
Modelo múltiples tiendas.....	29
Resultados.....	31
Evaluación económica.....	32
Conclusiones.....	33
Anexo.....	35
Modelo 1) (2,0,2).....	35
Modelo 2) (7,0,2).....	35
Modelo 3) (2,0,1).....	36

Introducción

Contexto

Flexit es una aplicación creada en el año 2018 dentro de Falabella Retail para solucionar varios dolores que la empresa tenía presente en sus tiendas Falabella al igual que el resto de las tiendas dentro de “Grupo Falabella”, este dolor venía enfocado a sus trabajadores en tienda y más específico a la flexibilidad de estos. Dentro del mundo de retail y en especial en las tiendas del tamaño de Falabella el costo de personal es uno necesario, para recibir y atender clientes al igual que mantener el funcionamiento de las tiendas, y costoso por lo que no tener trabajadores en tienda para el funcionamiento de esta al igual que tener muchos en horas de baja actividad en tienda son dos problemas siempre presentes dentro de Falabella. Debido a estos dolores nace la idea de Flexit una aplicación de contratación que da la habilidad a gerentes de recursos humanos de tiendas asociadas a poder hacer postulaciones de trabajo de un día a otro para hacer distintas labores dentro de tienda, especificando horarios necesarios y generando contratos vigentes por jornadas laborales.

Actualmente Flexit se encuentra siendo usado en alrededor de 150 tiendas a lo largo del país donde se encuentran todas las tiendas Falabella de Chile, con una mayor concentración en la región Metropolitana, tiene una base cerca de 160 mil usuarios donde se han trabajado más de 1.3 Millones de horas contando desde 2021 donde cerca de 970 mil han sido usadas dentro de este año 2023 por lo que se podría considerar que aumentó considerablemente el uso de Flexit durante este año. Generando unas ganancias a Flexit cerca de 195 mil UF desde su implementación en las tiendas Falabella y otras tiendas del grupo Falabella a lo largo del país.

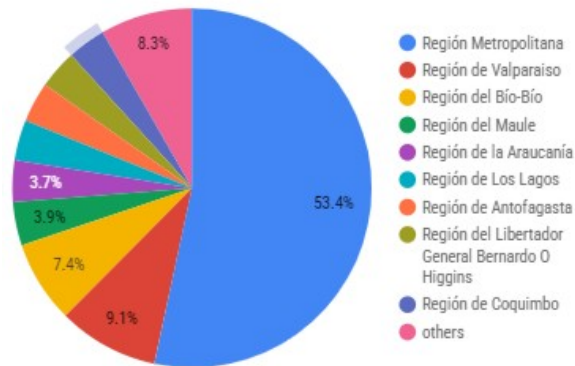


Ilustración 1: Distribución usuarios¹

Con una base de usuarios con edad promedio 23 años Flexit es una aplicación con público joven y activo que se encuentra como una de las maneras más efectivas para encontrar funcionarios externos para un número creciente de tiendas que buscan sumarse a la aplicación, con intereses futuros en expandirse a países fuera de Chile como Colombia o Perú, Flexit se ha vuelto una aplicación de suma importancia en Falabella.

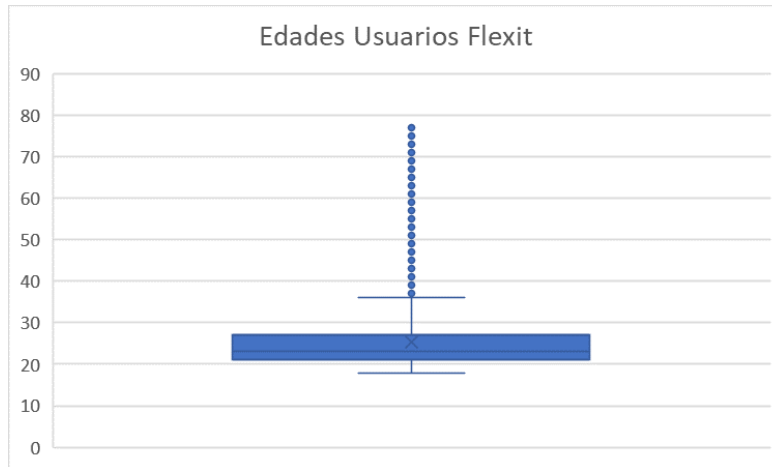


Ilustración 2: Distribución de edades usuarios Flexit²

¹ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

² Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

Descripción del problema

Debido al crecimiento considerable de la aplicación durante el año 2023, más horas son otorgadas a las tiendas para la contratación de usuarios Flexit. Las horas otorgadas a Flexit vienen de la cantidad de horas otorgadas a las tiendas para repartir entre sus operarios contratados, pero debido a la flexibilidad en horas y ahorro en costo por hora (el costo de contratación de usuarios Flexit es menor que el de operarios contratados), las tiendas prefieren el uso de usuarios a diferencia de usar personal ya contratado en tienda. La aplicación a experimentado un crecimiento exponencial, pasando de tener cerca de cinco mil usuarios entre los años 2021 y 2022 a tener casi veinticinco mil usuarios desde el comienzo de 2023 a finales de noviembre del mismo año.

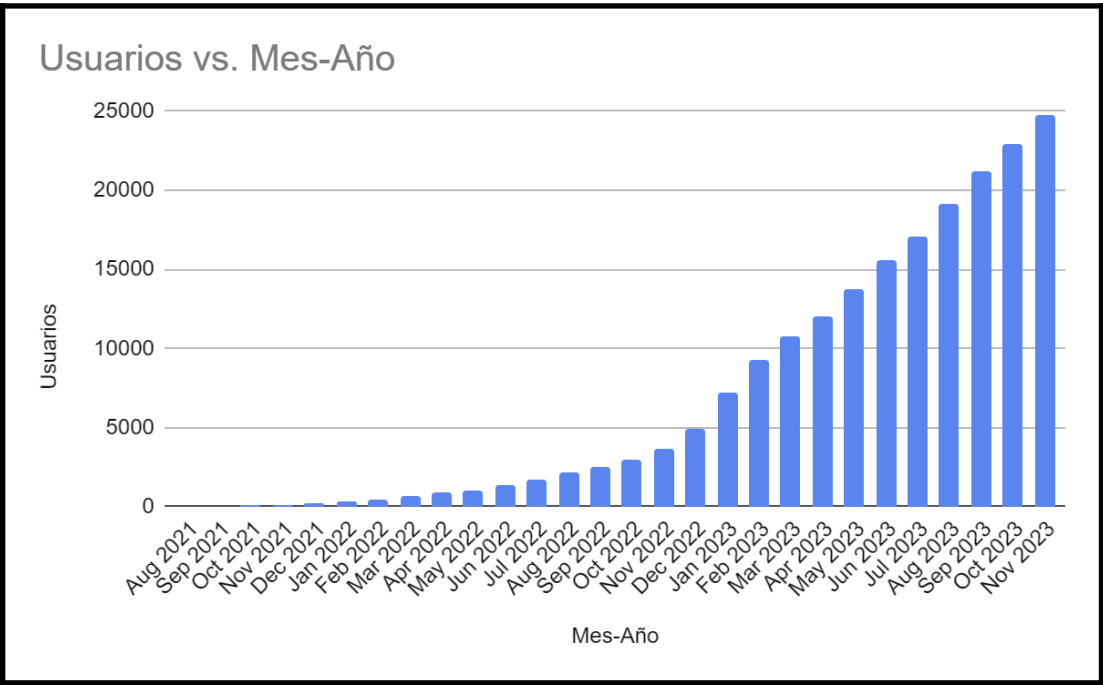


Tabla 1: Cantidad de usuarios Flexit³

Esto no significa que todas las tiendas prefieren depender de Flexit para ocupar todas las vacantes que tienen por día, esto se debe a varias razones, pero las dos más importantes detectadas al hablar gerentes de recursos humanos de las tiendas Falabella son la falta de experiencia de algunos usuarios Flexit en comparación a operarios que llevan trabajando en tienda durante

³ Emilio Nordenflycht, 2023, Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit

periodos de tiempo más extenso o la incertidumbre sobre la tasa de llenado de los usuarios a ofertas postuladas a través de la aplicación. Esta última razón nace de falta de conocimiento sobre cuál es la máxima cantidad de usuarios que llegarán a llenar las ofertas en caso de dar más vacantes a usuarios Flexit, este problema no solo los tiene las tiendas, el equipo Flexit es incapaz de determinar la cantidad de usuarios que postulan a las ofertas del mes que viene al momento de asignar horas a las tiendas a comienzos de cada mes. Este dolor de no saber la cantidad de horas óptimas afecta las tasas de llenado de las tiendas de Flexit, ya que se otorgan horas a las tiendas para uso de Flexit sin saber si las ofertas realizadas por las tiendas podrán llenarse. Por otro lado, otorgar menos horas puede implicar una demanda insatisfecha al tener usuarios dispuestos a trabajar que postulan a ofertas de trabajo, pero no quedan por falta de vacantes pedidas por parte de la tienda.

Este problema presenta un mayor dolor en tiendas Falabella fuera de la región Metropolitana, al tener menos conocimiento de cómo se comportan las bases de usuarios en regiones al igual que tener tasas de llenado peores a las de tiendas de Santiago.

Lo que se busca encontrar la cantidad de horas a destinar a las tiendas pertenecientes que trabajan con Flexit, si estas deben ser mayor a las que se están destinando actualmente debido a que la aplicación puede satisfacer la demanda laboral de las tiendas o deben ser menor, debido a que Flexit obtiene una ganancia de 0,16 UF por hora trabajada a través de la aplicación el foco actual del equipo es en maximizar las ganancias al otorgar horas sin sobreestimar la capacidad de la aplicación para satisfacer las cantidades de usuarios que piden las tiendas.

Objetivos

Objetivo General

Predecir las cantidades de postulaciones a ofertas para periodos futuros, durante un periodo de 3 meses, utilizando data histórica de Flexit para cada tienda de Falabella fuera de la región metropolitana.

Objetivos específicos

1. Mezclar y limpiar data sets.

- Estudio y limpieza de distintos datasets de Flexit para trabajar en el proyecto.
- 2. Sacar métricas para ingresar al modelo.
 - Métricas como valores input para el modelo al igual que ajustes de auto regresión, integración, promedio móvil, temporalidad, tendencia y ruido.
- 3. Poner a prueba distintos modelos para evaluar la mejor alternativa.
 - Evaluar que tipo de modelo se adapta mejor a los datos históricos de la aplicación poniendo en prueba sus capacidades predictivas.
- 4. Aplicar modelo para todas las tiendas deseadas.
 - Se crea un modelo para tienda que busca predecir en el alcance del proyecto teniendo en cuenta lo aprendido en la etapa tres.
- 5. Trabajar con valores predichos para encontrar cantidades óptimas a entregar a tiendas.
 - Poner a prueba la calidad de la predicción para cada tienda para luego entregar esta información de eventos en el próximo mes al equipo Flexit.
- 6. Entrega y automatización del modelo.
 - Limpieza y automatización del código de modelo final a entregar al equipo asegurándose de su funcionamiento con poca mantención requerida.

Estado del arte

La implementación de machine learning para la creación de modelos predictivos no es algo nuevo y ha sido implementado a lo largo de más de 30 años en distintos rubros debido a su capacidad de adaptación a distintas necesidades en varias industrias, para el trabajo de investigación se buscó el uso de machine learning dentro de la industria del retail , una de las razones para esto es que durante la investigación se encontró que el problema presentando a pesar de ser bastante común, como cualquier otro problema de predicción, es bastante único debido a que se trata de ver

dentro de una aplicación de contratación donde solo se pudo encontrar similitud en Uber⁴ con su modelo híbrido ocupan NN o Neural Networks para trabajar el ruido de sus time series de manera independiente.

En la industria retail comparte varias similitudes con el problema que está experimentando Flexit al tener que trabajar con datos en series temporales para entidades que aumentan o disminuyen sus ventas a lo largo del año, dependiendo de factores de temporadas, externos, tendencia y variables aleatorias, por lo que ejemplos de modelos predictivos en base a análisis de series temporales al igual que modelos en base a regresiones pueden ser aplicados a la situación encontrada en Flexit a modo de predicción de cantidad de postulaciones.

Uno de los ejemplos más importantes encontrados en la literatura sobre este tema es el estudio de hecho en “University of Applied Sciences Upper Austria” en análisis predictivo para la demanda de productos en retail, el estudio realiza un análisis extensivo en distintos trabajos realizados a lo largo de 20 años remarcando las fortalezas y debilidades en sus aplicaciones de distintos tipos de modelos predictivos para sus series temporales, se hace explicaciones de los modelos más populares para realizar forecasting haciendo énfasis importante en sus capacidades para trabajar con ciertos tipos de data al igual que maneras en las que uno puede adaptar estos modelos para poder obtener resultados favorables.

Los modelos más remarcables y aplicables al modelo de Flexit vienen a ser los componentes de ARIMA, esto es debido a que ARIMA es aplicable a series temporales de una variable, como lo sería ventas de cierto producto y en este caso postulaciones a tienda, al igual que ser altamente adaptable a la tendencia, factores de temporada y variables exógenas al modelo.

ARIMA tiene 3 componentes:

⁴ Nikolay Laptev, Jason Yosinski, Li Erran Li, Slawek Smyl, 2017. *Time-series Extreme Event Forecasting with Neural Networks at Uber*.

- Autoregression (AR): Se refiere a un modelo que muestra una variable cambiante que retrocede por sus propios valores rezagados o anteriores.
- Integrated (I): Representa la diferenciación de observaciones sin procesar para permitir la serie temporal se vuelve estacionaria (es decir los valores de los datos se reemplazan por la diferencia entre los valores de los valores anteriores).
- Moving Average (MA): Incorpora la dependencia entre una observación y un error residual de un modelo de media móvil aplicado a observaciones rezagadas⁵.

Cada componente de ARIMA ocupa parámetros modificables para el tipo de modelo que se quiere crear y esto dependiendo del comportamiento de los datos en la serie temporal.

Luego de realizar un análisis extensivo en el tema el estudio empieza a explicar cómo fue evolucionando su modelo predictivo para la demanda de verduras entre enero 2017 a diciembre 2019 donde se va iterando con distintas técnicas como ARIMA, SARIMA, SARIMAX y métodos de “Neuronal Networks” para lograr obtener mejores resultados predictivos, para esto se ocupan evaluadores de modelos “mean absolute percentage error” (MAPE) y “root mean square deviation” (RMSE) donde un se logró un forecasting altamente preciso con la utilización de los modelos previamente planteados, teniendo que adaptarse a cada tipo de verdura por distintos factores exógenos que afectan la demanda de algunas de estas como días de promociones, disponibilidad y popularidad por temporada entre otros⁶.

Dentro de la literatura sobre temas de forecasting, se puede encontrar un estudio del 2010 sobre el uso un tipo de ARIMA llamado ARIMAX que logra incluir variables exógenas en el apartado “X” del modelo para poder determinar la venta de ropa para niños musulmanes en periodos de días festivos, siendo la variable exógena los días festivos dentro del calendario islámico sin la necesidad de implementar el modelo SARIMAX que incluye el componente temporal en series de datos de datos no estacionarios⁷. Siguiendo este ejemplo se puede confirmar que es factible aplicar

⁵ Adam Hayes, 2022. *Autoregressive Integrated Moving Average Prediction Model*.

<https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp>.

⁶ Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. (2022). Predictive analytics for demand forecasting—a comparison of SARIMA and LSTM in retail SCM. *Procedia Computer Science*, 200, 993-1003.

⁷ Lee, M. H., & Hamzah, N. (2010). Calendar variation model based on ARIMAX for forecasting sales data with Ramadhan effect. In *Proceedings of the Regional Conference on Statistical Sciences* (Vol. 10, No. 2010, pp. 349-361).

variables externas dentro de un modelo predictivo y que esto mejoraría los resultados al momento de predecir eventos futuros.

Más recientemente en un caso de investigación sobre la demanda de un hotel anónimo en Estados Unidos se utilizó la temperatura anual como variable exógena para poder realizar “predicciones condicionales” al tener información futura de una variable con correlación elevada a la que se busca predecir, en este caso la demanda por semana de registros al hotel, se puede entrenar el modelo junto a esta variable exógena y luego otorgar al modelo para mejorar resultados si la variable exógena, en este caso temperatura, es lo suficientemente significativa para explicar el comportamiento de la variable a predecir.

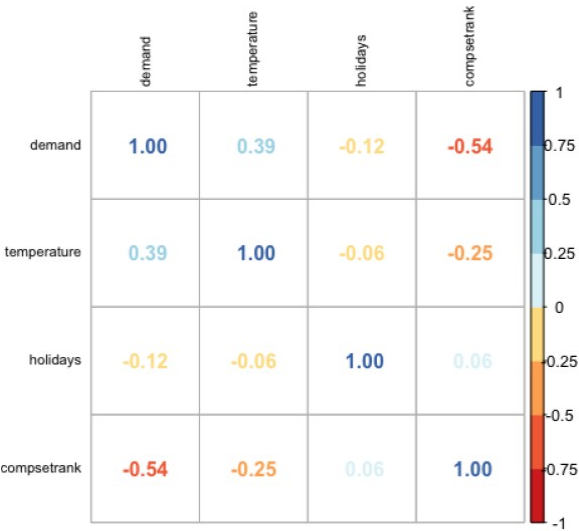


Tabla 2: Correlación variables Hotel⁸

En esta tabla de correlación del estudio del hotel, se compararon algunas variables, como la “temperature”, “holidays” y “compsetrank” a pesar de no tener correlaciones tan altas con la variable a predecir, aportan a la predicción a futuro debido a que son variables conocidas y pueden ayudar a realizar predicciones condicionales “Interestingly, the SARIMAX(1, 0, 1)(0, 1, 2, 7) model, whose AIC is 10,143.059, includes the exogenous variables as external regressors, namely temperature weekday, holidays, and competitive set ranking, it turns out that it is the best one. We employed our fitted SARIMAX model on the test set and obtained that the test MAPE is

⁸ Ampountolas, A. (2021). Modeling and forecasting daily hotel demand: A comparison based on sarimax, neural networks, and garch models. *Forecasting*, 3(3), 580-595.

7.333%, which is significantly more robust than any other proposed model.”⁹ Se puede apreciar que las variables otorgan mayor precisión al modelo al igual que permitir la capacidad del modelo de realizar predicciones condicionales con variables conocidas hacia el futuro, una herramienta de suma importancia para el modelo que se busca crear.

Solución escogida

Como fue visto previamente la solución al problema de incertidumbre en las postulaciones futuras vendría siendo algún tipo de modelo predictivo que logre estimar la cantidad de postulaciones que experimentará una tienda dependiendo del día y mes. Para esto no solo es necesario un modelo predictivo debido a que se necesita trabajar con pronóstico condicional. El pronóstico condicional es cuando se tiene el conocimiento de una variable que va a ocurrir en un futuro y se entrena el modelo con esta variable en ocurrencias del pasado, debido a esto se necesitará escoger un modelo que pueda ser no solo entrenado con variables exógenas, sino que se pueda interactuar para ingresar variables exógenas conocidas sobre un periodo que se quiera pronosticar y que otorgue valores distintos dependiendo del tipo de valor de las variables exógenas al igual que la data histórica.

Se tiene en evidencia el estudio del 2021, para referencias más recientes y similares al problema planteado debido a la similitud que presentan los tipos de datos y el foco del proyecto planteado en el trabajo “Modeling and Forecasting Daily Hotel Demand: A Comparison Based on SARIMAX, Neural Networks, and GARCH Models” que busca realizar pronósticos condicionales con la temperatura. En este trabajo se pone a prueba distintos modelos para poder pronosticar de manera condicional y probar su efectividad bajo métricas como MAPE (Mean absolute percentage error) para medir el error del modelo, menor el índice mejor el modelo.

⁹ Ampountolas, A. (2021). Modeling and forecasting daily hotel demand: A comparison based on sarimax, neural networks, and garch models. *Forecasting*, 3(3), 580-595.

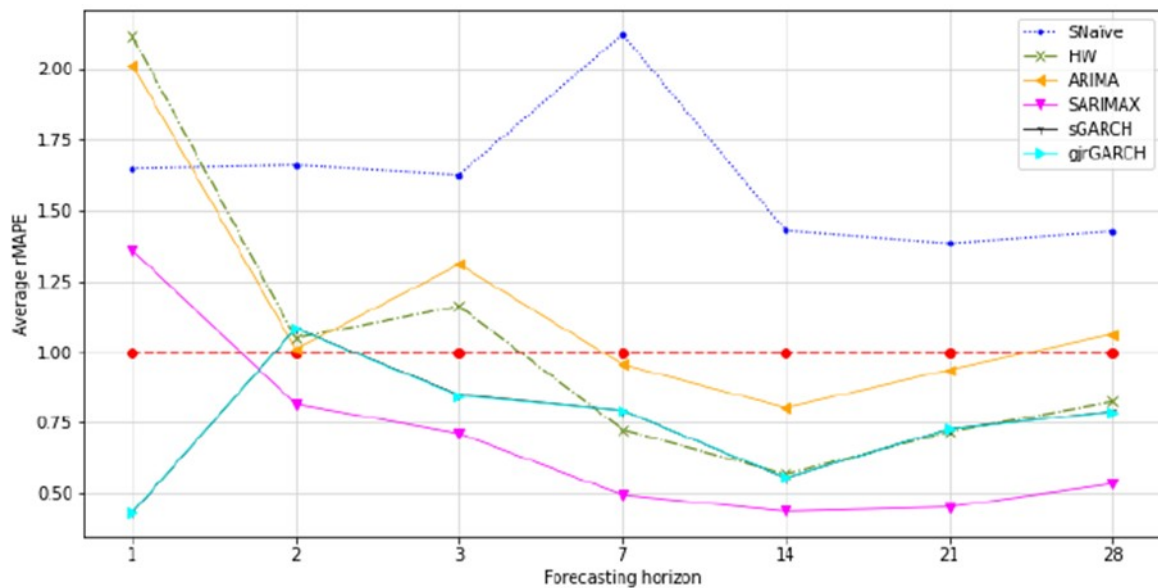


Tabla 3: MAPE promedio de distintos modelos¹⁰

Este gráfico muestra la evolución del error de los modelos a lo largo de la predicción, siendo días los horizontes partiendo desde 1 a 28 días, se utiliza MAPE escalado para mostrar el rendimiento del modelo a lo largo de la predicción, mientras menor sea el índice de MAPE menor es el error de la predicción en relación a los valores verdaderos en la prueba de “cross-validation”. Se puede observar que modelos como “Naive” no logran detectar el nivel de complejidad de la data trabajada y modelos como GARCH y ARIMA junto con sus distintas variables (SARIMAX, girGARCH) otorgan resultados bastante competitivos. Dentro de la selección de modelo se tiene en cuenta que actualmente estos modelos pueden servir para la solución del problema por lo que tienen que compararse para determinar el modelo que se adapte mejor a las necesidades de este proyecto.

Basado en investigación y puesta a prueba de los distintos modelos se llegó a una evaluación de parámetros importantes a considerar en un modelo predictivo para este proyecto, se usó una evaluación del 1 al 5 con escala de Likert representando la valoración de estos parámetros:

- Complejidad: Debido a la poca experiencia del equipo Flexit con modelos predictivos y para la longevidad del proyecto se tendrá en cuenta la complejidad de crear y mantención del modelo al pasar del tiempo siendo 1 muy complejo y 5 muy sencillo.

¹⁰ Ampountolas, A. (2021). Modeling and forecasting daily hotel demand: A comparison based on sarimax, neural networks, and garch models. *Forecasting*, 3(3), 580-595.

- **Flexibilidad:** A causa de la naturaleza de los datos de Flexit que no siguen un patrón marcado en torno a ciclos del año al igual que cada tienda tiene un comportamiento distinto se considera la flexibilidad de adaptación del modelo a data irregular como un factor a tomar en cuenta.
- **Automatización:** El modelo busca ser realizado para múltiples tiendas y que sea mantenido en el tiempo con poca experiencia, esto provoca que sus parámetros tendrán que ajustarse cada vez que el modelo se actualice por lo que la automatización se tiene en cuenta al momento de entregar el proyecto.
- **Efectividad:** Cerca de ser el factor más importante es la efectividad de este modelo, hay modelos que por su simplicidad o falta de complejidad no logran pronosticar data con más volatilidad o cambios de tendencia, por lo que es importante que el modelo sea robusto a cambios o outliers en la data.

	Complejidad	Flexibilidad	Automatización	Efectividad	Puntaje
ARIMA	2	3	4	3	12
SARIMAX	1	4	5	4	14
Naïve	5	2	5	2	14
Holt-Winters	2	3	4	4	13
GARCH	1	3	3	4	11

Tabla 4: Valoración de modelos¹¹

Como se puede observar en la tabla de valoración, los modelos con mayor puntaje en efectividad son SARIMAX, HW y GARCH donde claramente SARIMAX se presenta como una mejor opción al tener mayor flexibilidad por su sistema de parámetros que puede abarcar más tipos de datas al igual que su alto nivel de automatización por sus librerías como “auto. arima” que permite adaptar el modelo a distintos tipos de data. Todos estos beneficios no vienen sin un problema que sería el nivel de complejidad donde SARIMAX y Garch son considerados modelos complicados de trabajar por su complejidad y necesidad de mantención. A pesar de estos impedimentos Garch es un

¹¹ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

modelo efectivo para predicciones dentro del mundo financiero y SARIMAX para uso en escenarios reales donde variables exógenas afectan la variable estudiada.

En base a estas observaciones se toma la decisión de que SARIMAX se adapta más a las necesidades del proyecto por su capacidad, automatización y entrenamiento con variables exógenas.

Metodología

Metodología por objetivo

A continuación, una explicación de las distintas metodologías que se aplicarán a las distintas etapas del proyecto para realizar el modelo predictivo:

1. **Mezclar y limpiar data sets:**

Para este paso, y el resto del proyecto, se utilizará Python en Jupyter Notebook para trabajar con los data sets de la aplicación Flexit donde se tiene registro de las distintas interacciones de usuarios, sea individuos Flexits o tiendas. Se decidió trabajar en este lenguaje por su facilidad para trabajar con machine learning y que el equipo Flexit tiene experiencia con el lenguaje asegurando la perpetuidad del proyecto luego de su entrega. En el caso del equipo Flexit guarda la información en formato Excel y se divide en varios periodos los tipos de ofertas que hacen las tiendas junto con todas sus características, también se guardan todas las postulaciones de los usuarios Flexits sean aceptadas o rechazadas con todas sus características.

Los datasets al ser provenientes del mundo real pueden venir alterados, con variables faltantes o repetida por lo que su limpieza es prioritaria para empezar a trabajar los datos y tener una buena calidad de resultados. Para el trabajo previo se va a limpiar valores nulos, repetidos y variables que no sean necesarias para el modelo ya que los datasets son de gran tamaño por lo que reducir sus dimensiones para su trabajo servirá para la rapidez de la ejecución de códigos.

2. **Sacar métricas para ingresar al modelo.**

Cómo se trabajará con modelos predictivos se debe trabajar con una variable que se quiera predecir, para esto se busca la cantidad de aplicaciones realizadas a las ofertas por semana y por tienda, esto es posible por el código identificador de las ofertas donde sus primeros 3 dígitos indica la tienda y el resto 4 dígitos indicaría la oferta.

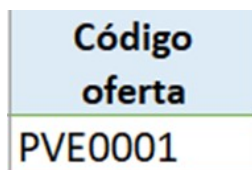


Ilustración 2 Ejemplo Código de Oferta

Luego de determinar la cantidad de aplicaciones se elimina las “Falsas aplicaciones”, esto se habló con el equipo Flexit que consistiría en postulaciones hechas por una persona durante un mismo día para jornadas de trabajo durante la misma jornada laboral en distintas tiendas, por lo que las postulaciones que caigan bajo esa condición no se contarán aquellas que sean “Rechazadas” en caso de tener otra postulación “Firmada” (que significa que fue aceptado en la postulación) para no inflar el número de aplicaciones y medir una cantidad real de aplicaciones que llenarían las postulaciones de las tiendas.

3. Poner a prueba distintos modelos para evaluar la mejor alternativa.

Para trabajar con modelos predictivos se debe comparar el rendimiento de distintos modelos con los índices RMSE y MAPE por lo que se debe evaluar modelos con distintos parámetros de entrenamiento, con posibles metodologías como “seasonal decompose” para quitar la temporalidad de la data de esa manera poder trabajar con data no estacionaria¹², agregando datasets fuera del año 2023 y con ciclo semanal o mensual. Para esto se utilizará una tienda de prueba donde se irá probando modelos ARIMA, SARIMA y SARIMAX para determinar cuál otorga los mejores resultados al momento de predecir al igual que sirva para las necesidades del equipo Flexit. La literatura sobre modelos predictivos deja bastante evidencia de que no existe una única solución para todo, en todos los proyectos llevados a cabo en esta área siempre se necesita una etapa de prueba para determinar el mejor modelo a ocupar para la data trabajada, por lo que es

¹²Ritu Santra,2023, *Stationarity in Time Series* <https://medium.com/@ritusantra/stationarity-in-time-series-887eb42f62a9>

de suma importancia evaluar todas las opciones con el fin de tener el menor error posible al momento de predecir.

Este proceso implica estudio de la data disponible sobre las postulaciones a ofertas de tiendas a lo largo de Chile e información sobre la construcción de modelos predictivos para poder generar un código robusto que pueda predecir de manera efectiva sin necesidad de mucho input por parte del equipo, de esta manera se disminuye el riesgo de falla humana.

4. Aplicar modelo para todas las tiendas deseadas.

Luego de encontrar la mejor manera de realizar el modelo se repetirá este proceso en distintas tiendas para crear un código final que contenga la información y parámetro de cada tienda que se busca predecir.

5. Trabajar con valores predichos para encontrar cantidades óptimas a entregar a tiendas.

Al momento de lograr predecir valores con el modelo, se aplica la tasa de aceptación por tienda que viene siendo la cantidad de postulaciones que se logran volver usuarios contratados, esta tasa es única por tienda ya que desde el momento de postular a firmar pueden ocurrir múltiples escenarios como el rechazo de oferta, cancelación de oferta, no queden por selección o que se llenen los cupos. Con la cantidad de potenciales usuarios contratados se estima por parte del equipo Flexit si aumentar o disminuir horas destinadas en base a la necesidad de personal de cada tienda.

6. Entrega y automatización del modelo entregable.

Luego de encontrar un modelo óptimo para la implementación en la tienda de prueba se buscará aplicar este modelo a 28 tiendas Falabella, se crearán dos códigos que recorran las bases de postulaciones para cada tienda y entrenen los modelos en base a los parámetros necesarios para cada tienda. El primer código será para la prueba de estos códigos donde se harán “cross-validation” para los distintos modelos y otorgando un “output” de índices RMSE y MAPE para que se puede ir revisando la calidad de los códigos al integrar data nueva cada mes, de esta manera se puede observar si la implementación de nueva data para el entrenamiento del modelo está

funcionando de manera correcta al igual que mostrar la posibilidad de error para cada modelo, información importante al momento de considerar que tan acertadas son las predicciones otorgadas en relación a un escenario real. El segundo código será el cual otorgue predicciones condicionales en base a la variable exógena otorgada, al momento de predecir con los modelos se deberá proporcionar las horas a otorgar a tienda esto será en formato mes o día.

Metodología general

El desarrollo del proyecto, por su carácter predictivo, tendrá la particularidad de que se tendrá que iterar a lo largo de 3 meses para perfeccionar la calidad del modelo por lo que se deberá volver a pasos previos en el trabajo para realizar cambios de métricas y editar tiempos en caso de que los valores otorgados no representan las aplicaciones reales. Se realizarán iteraciones para confirmar el funcionamiento de las predicciones con “cross-validation” para poder entregar el mejor modelo posible con la data entregada. Se probarán distintas técnicas y métodos para mejorar los resultados e ir comparando los índices de cada uno de estos modelos.

Riesgos y mitigaciones

Debido a la naturaleza de este proyecto los riesgos van orientados a la ejecución del código al igual que la mantención de este, a continuación, se encuentra la matriz de probabilidad impacto:

Leve	1 a 2
Moderado	3 a 4
Alto	6 a 9

Tabla 5: Indicador de riesgo¹³

¹³ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

Probabilidad de ocurrencia	Impacto		
	Leve	Moderado	Alto
Leve	1	2	3
Moderada	2	4	6
Alta	3	6	9

Tabla 6: Matrix probabilidad de ocurrencia¹⁴

Es una valoración del 1 al 9 indicando la relevancia del riesgo que podría presentar un suceso previsto hacia la integridad y funcionamiento correcto del proyecto a largo plazo.

Luego se tiene la matriz de riesgos donde se tienen en cuenta los riesgos más remarcables durante el desarrollo de este proyecto junto con mitigaciones para estos posibles escenarios:

Riesgo	Probabilidad	Impacto	Valorización	Clasificación de riesgo	Mitigaciones
Fallas en el modelo por input no estandarizado	Moderada	Leve	2	Bajo	Revisar datos de input para que estén acordes con las especificaciones del modelo.
Poco conocimiento de machine learning dentro del equipo para arreglar fallos inesperados en el modelo	Leve	Moderado	2	Bajo	Automatizar el proceso para tener de poco a nula interacción fuera de ingresar inputs. En caso de fallas más catastróficas contactar al equipo TI
Abandono del proyecto luego de su implementación	Leve	Alto	3	Moderada	Entablar la importancia y mejora que presentara el proyecto al pasar el tiempo con la retroalimentación del modelo con nueva información.
Falla de estimación de escenarios reales por presencia de outliers fuera de calendario y de historial.	Alta	Moderado	6	Alta	Experiencia del equipo Flexit sobre el flujo de postulaciones que ayude a entender sucesos que ocurran en escenarios no pronosticados.

Tabla 7: Matriz de riesgo¹⁵

En esta matriz se puede observar a simple vista problemas técnicos que son simples de solucionar como sería el trabajo de input y fallos inesperados generales, en ambos casos se requiere un poco de trabajo en el código donde el equipo Flexit es capaz de operar. En otro escenario esta uno más grave que es el abandono del proyecto por motivos inesperados, para esto la mitigación vendría siendo tratar de impulsar conversaciones con el equipo para transmitir la importancia del proyecto al igual como esta mejora en el tiempo. Esta mejora sería en una primera instancia por el simple funcionamiento de los modelos en series temporales, los cuales mejoran cuando se les agrega más información histórica al pasar del tiempo y otra mejora en el tiempo es sobre la necesidad de este

¹⁴ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

¹⁵ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

proyecto cuando Flexit se expanda a otras empresas será de alto valor tener información de cómo se comporta la base de usuarios cerca de las tiendas Falabella (ejemplo actualmente se está entablando conversaciones para entrar en Sodimac e información de la frecuencia de trabajo de la base de usuarios en tiendas cercanas a tiendas Sodimac podrá servir para pronosticar escenarios pilotos).

El escenario más grave está asociado a las limitaciones del mismo modelo, en caso de escenarios en el mundo real que afecten el funcionamiento normal del flujo de trabajo de las personas a nivel región o país podrán afectar de manera importante las predicciones del modelo y este cambio no se verá identificado hasta que se actualice el modelo al final del mes con la información de este escenario. En el caso de que ocurra un evento que afecte de esta manera el modelo se necesitará del equipo Flexit para que identifique el evento y actúe de manera acorde a la experiencia que han obtenido de eventos similares en el pasado para estimar que tanto afectara las predicciones planteadas del mes.

Métricas de desempeño

Como se había planteado en un momento, los modelos se evaluarán con RMSE y MAPE que son dos métricas de desempeño de modelos predictivos ocupadas en machine learning

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{y_t}$$

Ecuación 1: MAPE

- \hat{y}_t es el valor actual observado en el tiempo t.
- y_t es el valor predicho por el modelo en el tiempo t.
- n es el número de observaciones.

MAPE calcula el error porcentual promedio entre las observaciones reales y las predicciones. Se utiliza cuando se desea expresar el error en términos porcentuales, lo que facilita la interpretación de la precisión del modelo. La fórmula permite estimar la calidad del modelo, mientras menor sea la diferencia mejor es el modelo:

MAPE	Interpretation
<10	Highly accurate forecasting
10-20	Good forecasting
20-50	Reasonable forecasting
>50	Inaccurate forecasting

Source: Lewis (1982, p. 40)

Ilustración 3: Interpretación MAPE

También se tiene RMSE que sirve para ver la distancia promedio entre valores verdaderos y valores predichos, este indicador es útil para la comparación entre modelos por lo que servirá al momento de determinar qué modelo se acomoda más a la serie temporal, al igual que MAPE mientras menor sea este valor, mejor es el modelo en relación con otros modelos comparados.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2}$$

Ecuación 2: RMSE

- \hat{y}_t es el valor actual observado en el tiempo t .
- y_t es el valor predicho por el modelo en el tiempo t .
- n es el número de observaciones.

RMSE calcula la raíz cuadrada del error cuadrático promedio entre las observaciones reales y las predicciones. Es sensible a errores grandes, ya que penaliza más fuertemente las desviaciones grandes entre las observaciones reales y las predicciones.

Para esto la tienda Falabella Viña del Mar fue escogida como la tienda a probar el modelo inicial para luego aplicar al resto de las tiendas, esto viene a que la tienda de Viña contiene data histórica previa a 2023 que podría servir para mejorar la predicción del modelo, al igual que ser de las tiendas con menor rendimiento en Tasas de llenado (Ts) de ofertas Flexit (cerca de 75% durante

2023 hasta la fecha), pero con considerable uso de la app siendo la tienda numero 11 con más horas trabajadas a través de la aplicación en todo el año.

Otra métrica considerar para la evaluación de modelo es el AIC, o Criterio de Información de Akaike, es una medida utilizada para comparar modelos estadísticos. El AIC toma en cuenta la calidad del ajuste del modelo y penaliza por la complejidad del modelo. La idea es encontrar un equilibrio entre la precisión del modelo y su simplicidad. El AIC se calcula mediante la siguiente fórmula:

$$AIC = 2k - 2 \ln(L)$$

donde:

- k es el número de parámetros en el modelo.
- L es la función de verosimilitud del modelo.

El AIC sirve al momento de designar parámetros, a diferencia del MAPE y RMSE que son utilizados sobre las predicciones de modelos en etapas de “cross-validation”, de esta manera se pueden determinar los mejores valores para ingresar al modelo en la etapa de determinación de parámetros.

Para índices del modelo se ocuparán valores obtenidos del modelo base ARIMA para lograr establecer un “baseline” de modelo predictivo buscado a mejorar. Estos valores serian:

	Actuales	Esperados
MAPE	49	20
RMSE	63	30
AIC	1959	<1000

Ilustración 4: Valores actuales y esperados del modelo predictivo¹⁶

¹⁶ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

Por otro lado, se busca aumentar la cantidad de usuarios firmados que se viene refiriendo a los usuarios que son contratados por la aplicación, se demostrara la ganancia potencial de los usuarios firmados con el ingreso por hora trabajada (IH) que viene siendo la ganancia para Flexit por cada hora trabajada en tienda, representando 0,14UF por hora.

Desarrollo

Trabajo preliminar

Antes del trabajo en el modelo predictivo se debe hacer una limpieza de los múltiples Excel de Flexit, estos se separan en reportes trimestrales de ofertas al igual que reportes semestrales de postulaciones de usuarios, se vinculan estos Excel limpiándose de información no relevante como datos personales de usuarios, cargos de trabajo y otra información de empresa.

Se saca las horas de cada oferta descontando horas de colación cuando es necesario para sacar las horas reales pedidas por cada oferta, valor importante si se decide trabajar con horas de contratación como variable exógena a ingresar al modelo.

	offer_code	Fecha Inicio	Tienda_x	Postulaciones	Horas reales pedidas
0	VNA0941	2023-01-05	Falabella Viña del Mar	7.0	30.0
7	VNA0942	2023-01-05	Falabella Viña del Mar	10.0	25.0
17	VNA0943	2023-01-06	Falabella Viña del Mar	18.0	30.0
19	VNA0944	2023-01-06	Falabella Viña del Mar	11.0	25.0
46	VNA0971	2023-01-09	Falabella Viña del Mar	36.0	18.0

Tabla 8: Referencia de blend de datos¹⁷.

En base a la clave principal de oferta “offer_code” se puede separar crear un dataset con todas las ofertas por tienda para luego sacar de este las tiendas con las que se quiere trabajar, se reúne la cantidad de horas pedias y postulaciones a cada oferta por día, al igual que llenar días no registrados con valor “0” para poder representar el flujo real de la aplicación. Se trabajan con 28

¹⁷ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

tiendas distintas, pero para periodo de análisis de modelos ARIMA, se trabajará con la tienda Viña del Mar.

ARIMA

Como fue visto en el estado del arte ARIMA tiene tres componentes, “Autoregressive”, “Integrated” y “Moving average”, para estos componentes se deben escoger parámetros que indiquen con que cantidad de “lags” o días en este caso se deben considerar en el pasado, a primera revisión se ve la autocorrelación que mide la relación lineal entre los valores pasados y valores presentes al igual que ver la autocorrelación parcial. Con la autocorrelación se puede ver el parámetro “Autoregressive” y con la autocorrelación parcial se puede ver el parámetro “Moving Average”.

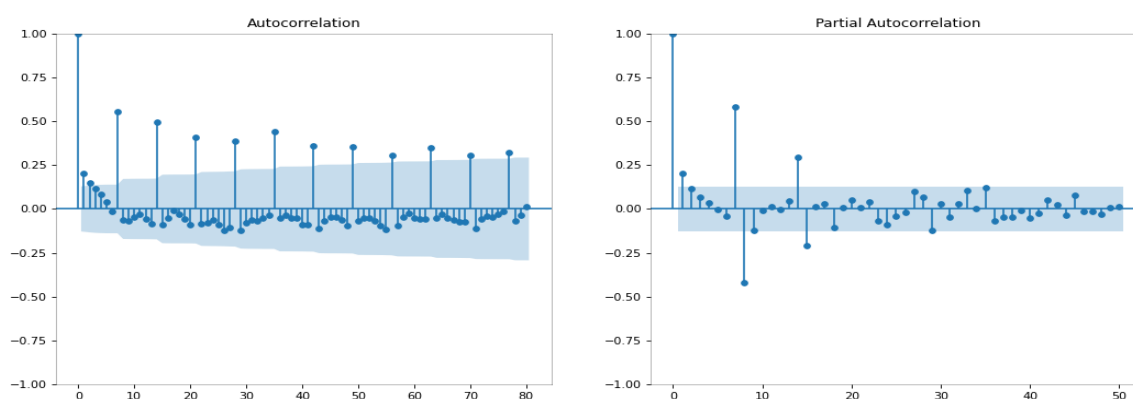


Tabla 9: AC y PAC¹⁸

Observando las graficas otorgadas por el código se ve claramente que hay una temporalidad semana al existir “lags” significativos cada 7 días en el grafico AC, se consideran significativos aquellos que salen del área azul. En el grafico PAC no se logra observar “lags” significativos a la larga, por lo que el grafico no nos otorga suficiente información para escoger el siguiente parámetro, a pesar de esto se logra identificar la existencia de un flujo semanal marcado.

Después de esto se puede aplicar una técnica va probando distintos parámetros para el modelo:

¹⁸ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*


```

# Búsqueda de parametros
for params in param_combinations:
    order = params[:3]

    try:
        model = sm.tsa.SARIMAX(df_train,
                                order=order,
                                seasonal_order=7)
        result = model.fit(dis=False)
        aic = result.aic

        if not math.isinf(result.zvalues.mean()):
            print(order, seasonal_order, aic)

            if aic < best_aic:
                best_aic = aic
                best_params = params

            else:
                print(order, seasonal_order, 'no converge')

    except:
        continue

```

Ilustración 5: Código de búsqueda de parámetros¹⁹

Con esto se escoge el modelo con el mejor índice AIC, luego se lograron extraer tres modelos ARMA:

- 1) (2,0,2)
- 2) (7,0,2)
- 3) (2,0,1)

Para probar la calidad de estos se puso a prueba con el método de “cross-validation” donde se trabajo con el 80% de los días como base de entrenamiento y el resto 20% como base de prueba. De los tres modelos entrenados solo el modelo 2) logro percibir los altos y bajos de la data, a pesar de esto su índice de error se mantiene elevado con un MAPE≈ 49 y RMSE≈ 63 debido a que no logra predecir el aumento considerable de postulaciones en el periodo de prueba.

Luego se prueba otro método de “auto.arima” que escoge el mejor modelo en base a el índice AIC al igual que permite agregar la temporalidad si existe en los datos, en este caso la temporalidad es 7 ósea semanal como se encontró previamente.

¹⁹ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

```

model = auto_arima(df_train,
                   seasonal=True,
                   m=7,
                   suppress_warnings=True)
model.fit(df_train)

model.summary()

```

SARIMAX Results						
Dep. Variable: y			No. Observations: 200			
Model:	SARIMAX(1, 0, 0)x(1, 0, [1], 7)			Log Likelihood	-974.968	
Date:	Mon, 25 Sep 2023			AIC	1959.936	
Time:	17:53:58			BIC	1976.427	
Sample:	01-09-2023 - 07-27-2023			HQIC	1966.610	
Covariance Type: opg						
	coef	std err	z	P> z	[0.025	0.975]
intercept	0.8386	0.722	1.162	0.245	-0.576	2.253
ar.L1	0.5249	0.070	7.458	0.000	0.387	0.663
ar.S.L7	0.9738	0.018	53.658	0.000	0.938	1.009
ma.S.L7	-0.7708	0.070	-10.960	0.000	-0.909	-0.633

Ilustración 6: Auto ARIMA en Viña del Mar²⁰

Se observa que el modelo pasa a ser (1,0,0) x (1,0,1,7) con parámetros temporales agregados que sirven a detectar ciclos dentro de la data, que en este caso serían semanales. Nuevamente se realiza una prueba otorgando índices MAPE≈ 46 y RMSE≈61, índices levemente mejores que la prueba anterior. Al momento de graficar ambas predicciones con las postulaciones actuales se observa que la diferencia no es mayor, a pesar de esto se aprende que el modelo producido de auto. arima es igual o mejor que aplicar parámetros realizando un estudio de la data.

²⁰ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

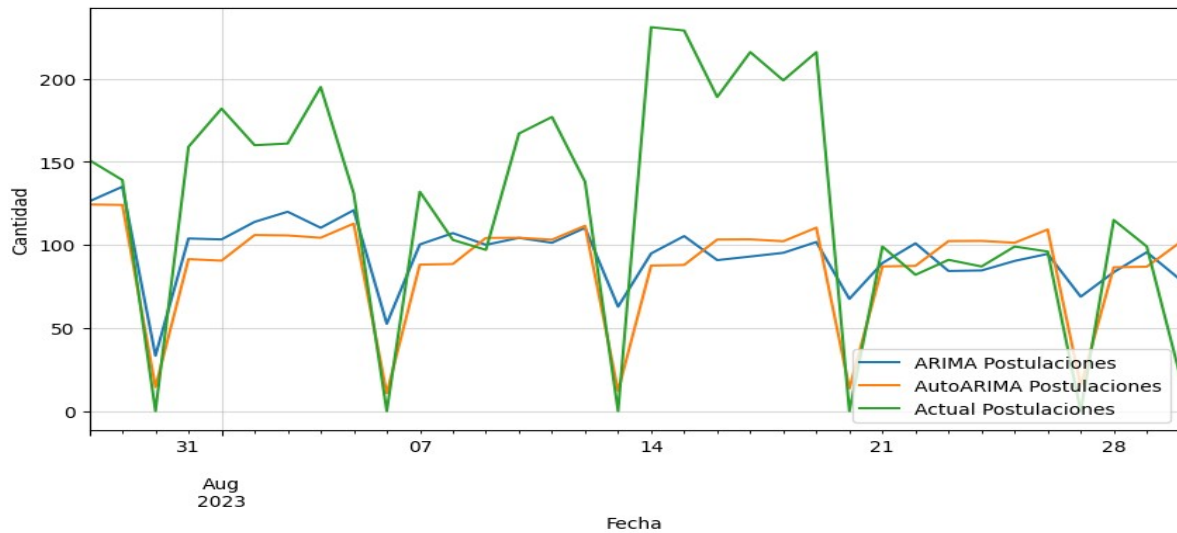


Ilustración 7: Grafica ARIMA/ Auto.ARIMA²¹

Ingresar data previa

Para tratar de buscar una manera de poder predecir periodos no conocidos en el modelo se trato de trabajar con data del año pasado entrenar el modelo con tendencia y temporalidad de meses que se buscan predecir. En este caso se limpiaron data sets con información del año pasado donde se puedo agregar postulaciones del segundo semestre de 2022 logrando tener una base de datos de 543 días, pero esto no logra mejorar los resultados de la predicción solo logra empeorarla con $MAPE \approx 68$ y $RMSE \approx 75$. Esto puede ser a base de que la data del año pasado no es representativa de lo que ocurre actualmente en la aplicación, un aumento de postulaciones debido al aumento de usuarios en la aplicación y un uso más ocasional de la aplicación por parte de las tiendas a diferencia de la actualidad donde se ocupa regularmente.

²¹ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

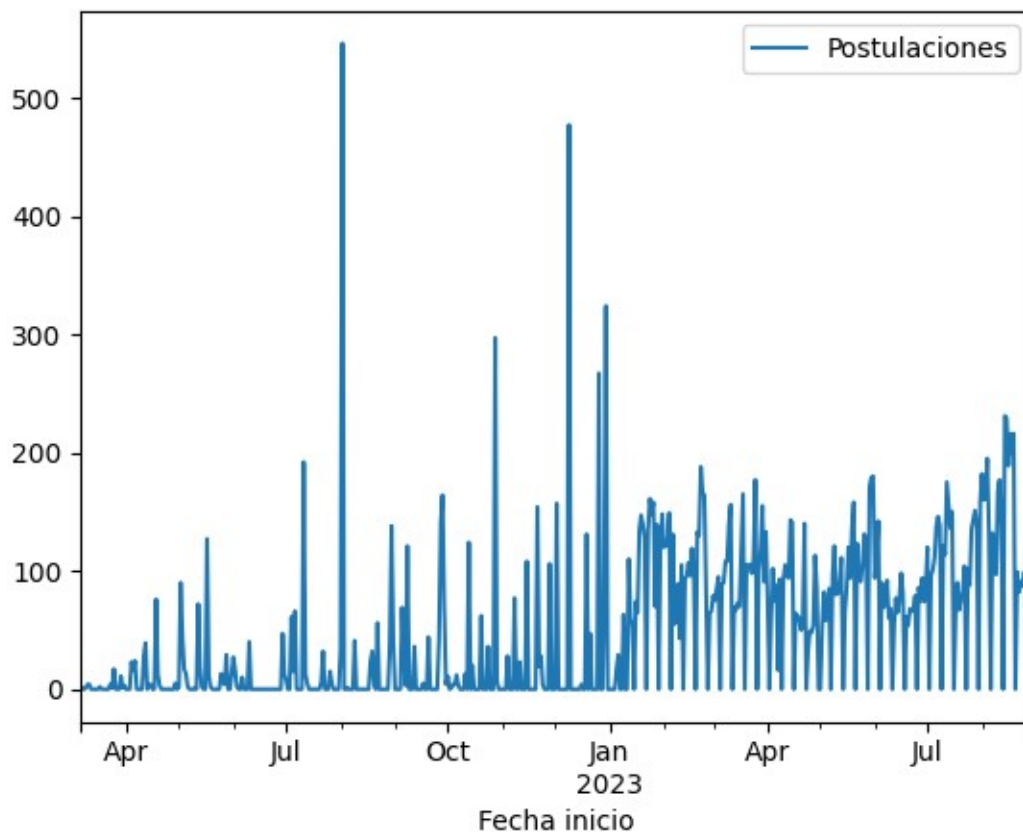


Ilustración 8: Modelo con data 2022²²

SARIMAX

Luego de probar técnicas de estimación de parámetros con la data de Viña del Mar Falabella se ve la posibilidad de agregar variables exógenas, o variables fuera de data que tienen a efecto en la variable a predecir. Para esto se ingresa la cantidad de horas destinadas a las tiendas, de esta manera realizar una predicción condicional de las postulaciones a futuro, debido a que la correlación de la cantidad de horas destinadas con la cantidad de postulaciones es alta en gran parte de las tiendas con un promedio de 0,65.

²² Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

```

Correlación entre 'Postulaciones' y 'Horas reales pedidas' en Falabella Puerto Montt: 0.5575313110483873
Correlación entre 'Postulaciones' y 'Horas reales pedidas' en Falabella Plaza Antofagasta: 0.6369273006194872
Correlación entre 'Postulaciones' y 'Horas reales pedidas' en Falabella Chillán: 0.7587199861949677
Correlación entre 'Postulaciones' y 'Horas reales pedidas' en Falabella Temuco: 0.5235080920770179
Correlación entre 'Postulaciones' y 'Horas reales pedidas' en Falabella Curicó: 0.5027004117758226
Correlación entre 'Postulaciones' y 'Horas reales pedidas' en Falabella Pucón: 0.7580809101588349
Correlación entre 'Postulaciones' y 'Horas reales pedidas' en Falabella Los Angeles: 0.7035947329374419
Correlación entre 'Postulaciones' y 'Horas reales pedidas' en Falabella Melipilla: 0.5227624341895772
Correlación entre 'Postulaciones' y 'Horas reales pedidas' en Falabella Punta Arenas: 0.35872860992499106
Correlación entre 'Postulaciones' y 'Horas reales pedidas' en Falabella Valparaíso: 0.7177060846746492
Correlación entre 'Postulaciones' y 'Horas reales pedidas' en Falabella Rancagua: 0.5757253626980126

```

Ilustración 9: Extracto de correlaciones tienda²³

Debido a esto las predicciones mejoran considerablemente con $MAPE \approx 42$ y $RMSE \approx 59$ al igual que permitir realizar predicciones condicionales fuera de la data.

Modelo múltiples tiendas

A base de la experimentación previa se puede crear un modelo para las 28 tiendas ocupando datasets respectivos para cada tienda junto con sus variables exógenas, usando auto. arima se logran seleccionar todos los parámetros óptimos para las tiendas. Obteniendo resultados de prueba favorables para las tiendas.

Tienda	RMSE	MAPE
Falabella Viña del Mar	59.5953	42.0034
Falabella El Trébol	49.8749	41.3372
Falabella Quilpué	29.1208	21.3461
Falabella Concepción	27.8412	24.4855
Falabella Valparaíso	27.6149	17.9044
Falabella Curicó	25.1675	20.6802
Falabella Puerto Montt	23.2355	20.3031
Falabella Talca	22.6969	16.6707
Falabella Valdivia	22.5636	18.2208
Falabella Chillán	18.0967	12.8943
Falabella Los Angeles	17.9375	14.2496
Falabella Plaza Antofagasta	17.6104	12.477
Falabella Rancagua	16.7162	13.1526
Falabella La Calera	16.5325	13.8991
Falabella Melipilla	14.6634	13.084
Falabella Copiapó	13.9422	10.3894
Falabella Osorno	13.0884	10.8807
Falabella La Serena	11.7813	8.2722
Falabella Temuco	11.1544	8.23153
Falabella Iquique	10.8944	8.30154
Falabella San Felipe	9.75469	7.65623
Falabella Ovalle	8.96908	7.80596
Falabella Punta Arenas	6.95008	5.80822

Ilustración 10: Extracto índices todas las tiendas²⁴

²³ Emilio Nordenflycht, 2023, Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit

²⁴ Emilio Nordenflycht, 2023, Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit

Luego de esto se crea el código dedicado al entrenamiento de la data completa al igual que a la creación de predicciones a postulaciones dentro del mes de noviembre.

```
# Diccionario para almacenar las predicciones condicionales para cada tienda y variación de horas
predicciones_condicionales_tiendas = {}

# Itera sobre cada entrada en el diccionario de modelos ajustados
for nombre_tienda, model_fit in modelos_arima.items():
    # Encuentra la última fecha en los datos de la tienda
    ultima_fecha_tienda = dataframes_tiendas[nombre_tienda].index.max()

    # Calcula el primer día del mes siguiente
    primer_dia_siguiente_mes = ultima_fecha_tienda + pd.DateOffset(days=1)
    fecha_inicio_siguiente_mes = pd.date_range(start=primer_dia_siguiente_mes, periods=30, freq='D') # Cambiar periodos en función del mes a predecir

    # Obtiene la cantidad de horas mensuales proporcionadas para la tienda actual
    horas_mensuales_proporcionadas = plan_horas_por_tienda[nombre_tienda]
    # Calcula las horas diarias proporcionadas
    horas_diarias_proporcionadas = (horas_mensuales_proporcionadas/30)

    # Crea un DataFrame para el siguiente mes con las horas diarias proporcionadas
    data_siguiente_mes = pd.DataFrame({'Fecha Inicio': fecha_inicio_siguiente_mes, 'Horas reales pedidas': horas_diarias_proporcionadas})

    # Realiza una única predicción condicional para todo el mes
    prediccion_condicional = model_fit.predict(n_periods=len(data_siguiente_mes), X=data_siguiente_mes["Horas reales pedidas"].values.reshape(-1, 1))

    # Almacena la predicción en el diccionario por nombre_tienda
    predicciones_condicionales_tiendas[nombre_tienda] = prediccion_condicional
```

Ilustración 11: Extracto de código predicciones

Lo que se obtendrá del resultado dependerá del input por parte del equipo Flexit, en este caso se ingresó data de horas a otorgar a tiendas en formato mensual, por lo que aplico un “Downsampling” una técnica usada en modelos predictivos cuando se entrena un modelo con variables exógenas en un formato (en este caso días) pero al momento de alimentar el modelo con la variable exógena del periodo a predecir esta se encuentra en otro formato (en este caso mes).

Debido a que varia se crearon dos tipos de código, uno que tome el valor de las horas a otorgar a tienda en formato día y otro mes, para esto se deberá escoger que tipo de información se esta ejecutando para que el modelo lo pueda interpretar de manera correcta. Para esto el equipo deberá cambiar a “True” el formato deseado:

```
[ ] mes=True
    dia= False
```

Ilustración 12: Formato horas²⁵

²⁵ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

Resultados

El modelo logra predecir valores con bajos índices de error para la mayoría de las tiendas, estos valores pueden ser presentados en formato diario o mensual, en el caso de predicción de noviembre se obtuvieron datos mensuales de postulaciones a futuro.

Tienda	Sumatoria de Predicciones
Falabella Puerto Montt	1634.62
Falabella Plaza Antofagasta	1591.76
Falabella Chillán	2164.38
Falabella Temuco	645.604
Falabella Curicó	1005.46
Falabella Pucón	114.386
Falabella Los Angeles	667.249
Falabella Melipilla	421.834
Falabella Punta Arenas	500.506
Falabella Valparaíso	1439.56
Falabella Rancagua	1426.79
Falabella Arica	224.658

Ilustración 13: Extracto de postulaciones mensuales²⁶

Al momento de obtener resultados predichos se logra sacar un KPI simple de Postulaciones dividido en horas otorgadas para poder percibir que tienda son mas sensibles a la disponibilidad de horas de contratación de la tienda.

²⁶ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

```

KPI para Falabella Puerto Montt: 1.17
KPI para Falabella Plaza Antofagasta: 0.53
KPI para Falabella Chillán: 0.94
KPI para Falabella Temuco: 0.81
KPI para Falabella Curicó: 1.12
KPI para Falabella Pucón: 0.29
KPI para Falabella Los Angeles: 0.74
KPI para Falabella Melipilla: 0.70
KPI para Falabella Punta Arenas: 0.42
KPI para Falabella Valparaíso: 1.44
KPI para Falabella Rancagua: 0.48
KPI para Falabella Arica: 0.37
KPI para Falabella Concepción: 0.51
KPI para Falabella La Serena: 0.27
KPI para Falabella Valdivia: 0.95
KPI para Falabella San Fernando: 0.39
KPI para Falabella Iquique: 0.48
KPI para Falabella Castro: 0.33
KPI para Falabella Copiapó: 0.37
KPI para Falabella Calama: 0.48
KPI para Falabella Quilpué: 1.16
KPI para Falabella Viña del Mar: 1.74
KPI para Falabella El Trébol: 1.34
KPI para Falabella San Felipe: 0.50
KPI para Falabella Talca: 0.50
KPI para Falabella Ovalle: 1.23
KPI para Falabella Osorno: 0.90
KPI para Falabella La Calera: 1.62

```

Ilustración 14: KPI tiendas²⁷

Se puede observar que hay tiendas cuyos usuarios son más sensibles al aumento de oferta laboral por parte de la tienda que otros, en estas tiendas es posible que las necesidades de oferta laboral puedan ser satisfechas completamente por la aplicación.

Evaluación económica

Debido a la naturaleza de este proyecto no existe un costo de implementación del proyecto debido al logro de automatizar el proceso no se necesita personal experimentado para la implementación y utilización de la herramienta predictiva.

El fin del proyecto buscaba determinar si Flexit puede satisfacer una mayor demanda de las tiendas, actualmente la cantidad de horas otorgadas a las tiendas viene siendo cerca del 20% del total de horas para la tienda. El incentivo está en tener más usuarios Flexit trabajando en tienda debido al ingreso de 0,16 UF por hora trabajada. Para caso de estudio se tomo en referencia el ingreso obtenido de ser aplicado el modelo predictivo en el mes de noviembre contando con dos

²⁷ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

escenarios, un escenario optimista donde todos los usuarios trabajan 9 horas y uno pesimista donde trabajan 5 horas.

Tienda	Postulaciones Totale	Gente contratada	Gente contratada Flex	Escenario optimista (UF)	Escenario pesimista (UF)
Falabella Puerto Montt	10332	255	51	294	163
Falabella Plaza Antofagasta	12105	405	81	467	259
Falabella Chillán	20078	3800	760	4378	2432
Falabella Temuco	9686	255	51	294	163
Falabella Curicó	11333	3800	760	4378	2432
Falabella Pucón	1301	255	51	294	163
Falabella Los Angeles	12255	3145	629	3623	2013
Falabella Melipilla	5722	405	81	467	259
Falabella Punta Arenas	2846	255	51	294	163
Falabella Valparaíso	21322	255	51	294	163
Falabella Rancagua	9494	255	51	294	163

Ilustración 15: Extracto ingresos Noviembre²⁸

Debido al periodo donde se realizó la prueba que es noviembre donde las postulaciones aumentan en comparación al resto del año Flexit podría satisfacer la necesidad de personal de 26 de las 28 tiendas, generando un ingreso de 39.648 UF en el escenario optimista y 21.926 UF en el pesimista.

La ganancia potencial varía de mes en mes por lo que es necesario aplicar el modelo al comienzo de cada mes para poder tener valores acertados.

Se hace nota de que entregar valores de horas a destinar a cada tienda diarios mejora la predicción del modelo debido a que al agrupar los datos mensuales no se tiene en cuenta si las postulaciones son para ofertas a comienzo o final del mes por lo que asumir que se llenarían todas es un supuesto aplicado en los cálculos.

Conclusiones

En este informe se ha detallado el desarrollo de un modelo predictivo para las postulaciones a ofertas laborales en tiendas Falabella mediante la aplicación Flexit. La metodología abarca desde la limpieza de datos hasta la implementación y evaluación de modelos, destacando la iteración continua a lo largo de tres meses.

²⁸ Emilio Nordenflycht, 2023, *Modelo predictivo para las postulaciones a ofertas de trabajo de la aplicación Flexit*

Se ha demostrado la eficacia de modelos ARIMA y SARIMAX, especialmente al incluir variables exógenas como las horas destinadas a las tiendas. Los resultados muestran la sensibilidad de las tiendas a la disponibilidad de horas de contratación.

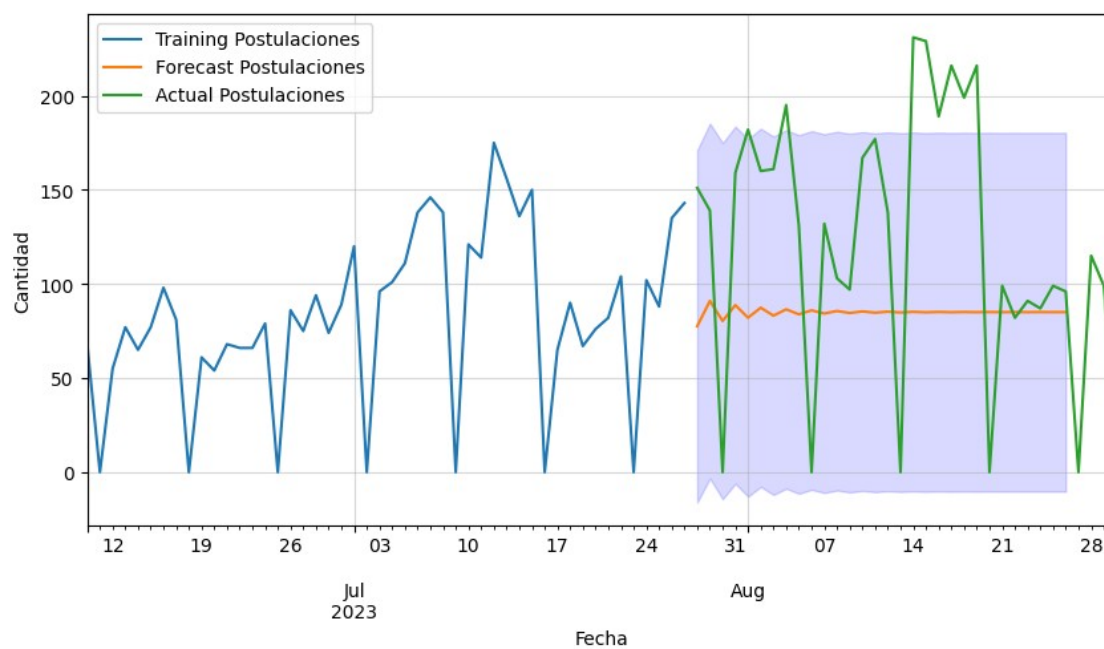
Durante el desarrollo del proyecto se buscó bajar el error de la tienda de prueba Falabella Viña del Mar, a pesar de agregar variables exógenas y adaptar el modelo lo mejor posible el error siguió siendo alto por lo que no termina siendo un modelo satisfactorio para la tienda, esto se explica por el periodo de prueba del “cross-validation” en el mes de agosto el modelo no logro identificar un pic en postulaciones demostrando las limitaciones del modelo cuando outliers provocados por eventos en el mundo real no son detectados.

Para recomendaciones en el corto plazo se sugiere empezar a otorgar mas horas a las tiendas para realizar pruebas en el mundo real sobre la capacidad predictiva del modelo. Al igual que en el mediano plazo se recomienda cambiar el formato del modelo a 12 periodos para poder predecir teniendo en cuenta el comportamiento de las postulaciones ocurridas en el mismo mes del año pasado, de esta manera mejorar la capacidad de detectar temporalidad del modelo.

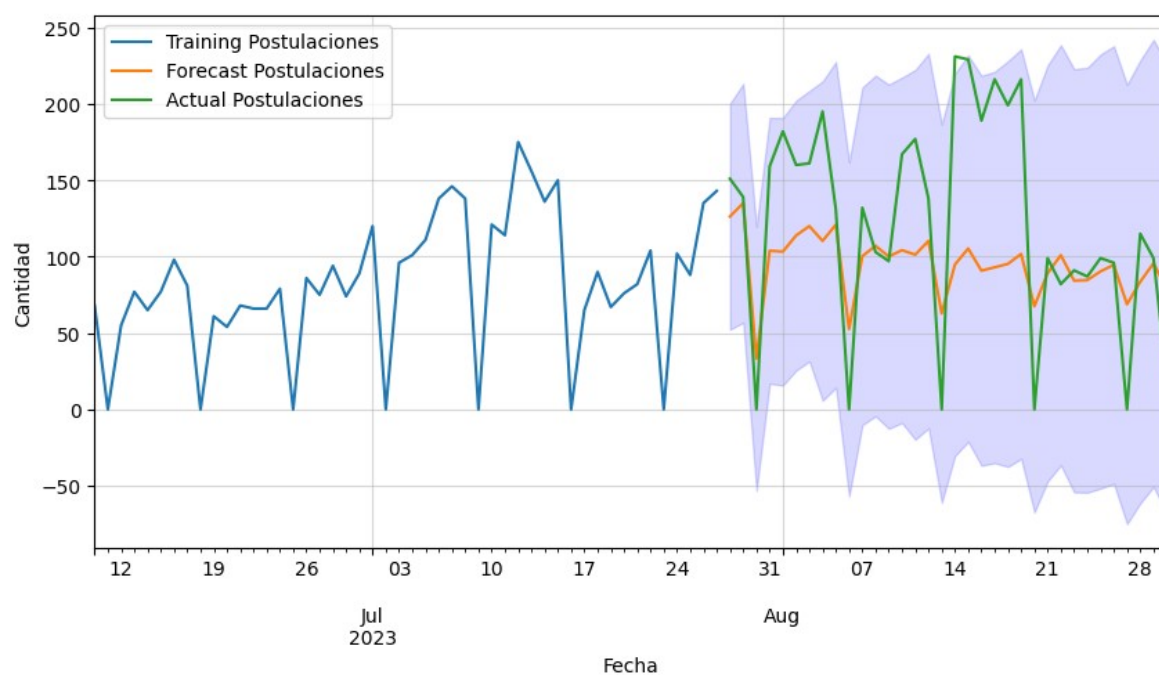
Para el futuro se sugiere considerar la incorporación de más datos históricos y la expansión a otras tiendas Falabella, de esta manera mejorando la capacidad predictiva del modelo. Además, se propone explorar nuevas variables exógenas y evaluar la capacidad del modelo para adaptarse a cambios significativos en el entorno laboral. Esta herramienta puede seguir otorgando valor al equipo Flexit si se decide mantener y utilizar el modelo para necesidades futuras.

Anexo

Modelo 1) (2,0,2)



Modelo 2) (7,0,2)



Modelo 3) (2,0,1)

