

Autor: Nicolás Castillo - Sebastián Arango - Camilo Delgado - Johan Rodríguez

PROYECTO FINAL ENTREGA 1 - CIENCIA DE DATOS APLICADA

1 DEFINICIÓN DE LA PROBLEMÁTICA Y ENTENDIMIENTO DEL NEGOCIO

Para el desarrollo del proyecto de ciencia de datos, hemos seleccionado a OPAIN S.A., empresa encargada de la concesión del aeropuerto El Dorado, con responsabilidades de administración, modernización y explotación comercial de este importante aeropuerto. En general, los aeropuertos cuentan con tres fuentes principales de ingresos: ingresos por vuelos comerciales, ingresos por vuelos de carga y las ventas de establecimientos dentro del aeropuerto.

Aunque históricamente las dos primeras fuentes han representado la mayoría de los ingresos, existe un creciente interés en potenciar el segmento comercial. Sobre todo, teniendo en cuenta que según un estudio de la consultora McKinsey que indica que para el año 2024 en la mayoría de las aerolíneas la deuda superará los ingresos lo que llevará a un incremento de los precios de tiquetes (Bower et al, 2022). Esto a su vez, representaría una disminución de los ingresos por parte de vuelos comerciales a nivel global. Actualmente, el aeropuerto El Dorado ha tenido limitadas estrategias para potenciar los ingresos provenientes de los más de 200 comercios situados en sus instalaciones. De ahí surge la necesidad de fusionar datos operativos con datos comerciales, para identificar oportunidades de mejora en esta área.

El propósito principal de este proyecto es descubrir y sugerir tácticas para aumentar las ventas de los establecimientos ubicados en el Aeropuerto Internacional El Dorado, basándonos en información básica de los vuelos. Utilizaremos técnicas de regresión para identificar patrones y comportamientos de los usuarios, lo que facilitará la toma de decisiones y la implementación de medidas más efectivas. Las métricas de negocio claves que servirán para evaluar nuestras propuestas serán:

- Ventas en tiendas cercanas a las puertas de embarque (1.5 horas antes del despegue)
- Número de transacciones en establecimientos cercanos a las puertas de embarque (1.5 horas antes del despegue).

2 IDEACIÓN

A continuación, se describe la propuesta de producto a desarrollar a partir del análisis de los procesos y problemáticas actuales y las necesidades y expectativas del cliente.

Usuarios potenciales del producto de datos

- **OPAIN S.A.:** como entidad administradora del Aeropuerto Internacional El Dorado, tiene interés en aumentar los ingresos comerciales y mejorar la experiencia de los pasajeros. Los resultados y recomendaciones generadas por el producto de datos serán fundamentales para ayudar en la toma de decisiones estratégicas y operativas.
- **Gerentes de tiendas y sus equipos de marketing:** tendrán acceso a la información de cómo la programación de vuelos impacta las ventas en sus tiendas.
- **Analistas de datos:** los profesionales encargados de analizar los resultados generados por el producto de datos serán parte importante del proceso. Sus observaciones y conclusiones contribuirán a la mejora continua de las estrategias operativas y comerciales.

Procesos actuales y problemáticas asociadas

En cuanto a procesos actualmente no se hace un uso de los datos operativos o comerciales para obtener un valor agregado. Como problemáticas se identificaron las siguientes:

- **Operaciones del aeropuerto:** asignación subóptima de vuelos, congestiones o subutilización de algunas áreas.
- **Operaciones de tiendas:** falta de correlación entre los horarios de mayor afluencia de pasajeros y las operaciones de las tiendas.

Propuesta de producto de datos

Se propone el desarrollo de un sistema que brinde al usuario información estimada del aumento de las ventas que pueden llegar a generar determinados vuelos de acuerdo con sus características tales como el destino, la hora de salida y la cantidad de pasajeros a bordo. Esto se traduce en una herramienta de ayuda para la toma de decisiones estratégicas para realizar la asignación de las puertas de abordaje. Este sistema se basará en un modelo de machine learning entrenado con datos históricos operacionales y comerciales.

Componentes tecnológicos:

- Interfaz de visualización
- Modelo de regresión

3 RESPONSIBLE

El presente proyecto plantea diversas implicaciones éticas, de privacidad, confidencialidad, transparencia y aspectos regulatorios que deben ser cuidadosamente considerados.

Es importante destacar que, en la fase inicial del proyecto, llevamos a cabo diversas sesiones de trabajo virtuales en las cuales definimos los aspectos fundamentales relacionados con la privacidad y la confidencialidad de la información. Estas sesiones representaron una etapa preliminar, donde se acotó la firma de acuerdos de confidencialidad, y posteriormente, *OPAIN S.A* se centró en la gestión de entrega de la data operativa y comercial.

Adicionando, se mencionan principales consideraciones éticas y legales que se deben tener en cuenta a lo largo del desarrollo del proyecto:

- **Transparencia en los Resultados de IA:** En caso de que se decida incorporar modelos de IA en el prototipo que formará parte de la entrega final, es esencial proporcionar una comprensión clara de cómo operan estos modelos y de qué manera impactan en el proceso de toma de decisiones.
- **Confidencialidad de los Datos Comerciales:** Los datos comerciales son valiosos y sensibles. La empresa debe garantizar que estos datos se encuentren protegidos contra el acceso no autorizado y su uso indebido. También deben establecerse políticas claras de acceso a estos datos dentro de la organización.
- **Privacidad de los Pasajeros y Empleados:** OPAIN S.A. debe asegurarse de que la recopilación y el uso de datos no violen la privacidad de los pasajeros y empleados del aeropuerto. Esto implica la necesidad de obtener el consentimiento adecuado cuando sea necesario y garantizar la anonimización de los datos personales.
- **Política de Privacidad, Aviso de Privacidad y Consentimiento Informado:**

OPAIN S.A. debe desarrollar una política de privacidad que explique en detalle cómo se recopilan, almacenan y utilizan los datos personales de los pasajeros y empleados del aeropuerto. Además, debe proporcionar un aviso de privacidad fácilmente accesible para todas las partes interesadas que describa claramente los derechos de las personas con respecto a sus datos personales. El consentimiento informado de las personas antes de la recopilación de datos es esencial, especialmente si se utilizan para análisis y toma de decisiones. Este consentimiento debe ser específico, voluntario y permitir a las personas revocarlo en cualquier momento, garantizando así que estén informadas y tengan control sobre el uso de sus datos personales.

4 ENFOQUE ANALÍTICO

Teniendo en cuenta que el producto a desarrollar tiene como objetivo brindar al usuario información estimada de las ventas que puede generar un vuelo para la toma de decisiones, se tiene la siguiente hipótesis a desarrollar durante el semestre:

“A partir de la información categórica y temporal de un vuelo es posible estimar el valor monetario de las ventas que generaran los usuarios que viajen en este”

Para poder comprobar tal hipótesis, se plantea un modelo de regresión, en el cual por medio de características estáticas y categóricas del vuelo como lo son su destino, muelle de salida y aerolínea, así como características temporales cualitativas como la hora de salida se pueda estimar el valor de ventas que generarían sus pasajeros en un momento determinado. Con tal fin se usa una estrategia compuesta de los siguientes pasos, los cuales suceden luego del proceso de recolección de datos:

1. **Análisis univariado:** Análisis individual de las variables o columnas seleccionadas, que se segmenta en los siguientes pasos:
 - 1.1. **Análisis de variables cualitativas:** Evaluación de la cardinalidad (número de categorías) y distribución de las categorías (análisis de Pareto y diagrama de barras para hallar valores más frecuentes).
 - 1.2. **Análisis de variables cuantitativas:** Generación de estadísticos de

tendencia central (media, mediana), medidas de dispersión (desviación estándar), evaluación de percentiles 5%, 25%, 50% (mediana) 75% y 95% para establecer la centralidad de la distribución, valor mínimo y máximo para establecer si hay datos atípicos, y gráfica de histograma para establecer el sesgo de la distribución y si existen múltiples modas.

2. **Análisis bivariado:** Usando como variable de comparación el valor de ventas estimado para el vuelo se hace un análisis de poder de predicción de cada una de las otras variables contra esta.

- 2.1. **Análisis bivariado de variables cualitativas:** Diagramas de cajas donde el eje X es la variable categórica y el eje Y es el valor de ventas estimado. Adicionalmente se aplica el test de Kruskal-Wallis para determinar si los grupos son diferentes respecto al valor de ventas.

- 2.2. **Análisis bivariado de variables cuantitativas:** Diagramas de dispersión X es la variable cuantitativa y el eje Y es el valor de ventas estimado. Si esta relación parece ser lineal se aplica la correlación de Pearson y en caso contrario la de Spearman.

- 2.3. **Análisis temporal:** Para unas variables categóricas se hace una gráfica de heatmap, donde el eje X es una variable de tiempo (hora o día del año) y el eje Y tiene los valores de la variable categórica para observar si para cada categoría el valor de ventas se acumula en algún periodo de tiempo.

Después de tener estos resultados se plantea como modelo preliminar un modelo de RandomForestRegressor, el cual consiste en un ensamble de árboles que segmentan la información numérica y categórica hasta llegar a un valor estimado de ventas.

5 RECOLECCIÓN DE DATOS

A medida que se avanzaba en las sesiones de trabajo con *OPAIN S.A.* durante las últimas semanas, se ha llevado a cabo un proceso de análisis que ha culminado en la identificación y consenso en torno a los datos de mayor relevancia. Dentro de este proceso, se ha determinado que los datos relacionados con los vuelos de salida son de particular interés y poseen un valor estratégico significativo. Esto se debe a que, durante el período previo a la salida de un vuelo, los pasajeros suelen pasar una considerable cantidad de tiempo en las áreas de espera del aeropuerto. Durante este intervalo, es más probable que los viajeros se involucren en actividades como compras, lo que incluye la adquisición de souvenirs, productos de moda, accesorios, y la experiencia gastronómica en los restaurantes y locales de comida disponibles en el aeropuerto.

Teniendo en cuenta lo anteriormente mencionado, se puede comentar que el enfoque en los vuelos de salida como punto focal de estudio se fundamenta en la idea de que, al comprender y adaptarse a las necesidades y preferencias de los pasajeros en este momento crítico, se puede mejorar la satisfacción del cliente, aumentar los ingresos generados en el aeropuerto y fomentar una experiencia de viaje más placentera en general. A través de la recopilación y análisis inteligente de estos datos, estaremos en posición de tomar decisiones informadas que

beneficiarán tanto a los pasajeros como a las partes interesadas involucradas en la operación del aeropuerto.

Descripción

El Dataset proporcionado de vuelos de salida comprende un período de tres años, desde 2020 hasta 2023, y consta de 344,146 filas y 31 columnas. Entre los atributos más destacados se encuentra información relevante como destino, tipo de vuelo, muelle, fecha, tipo de aerolínea y el número de pasajeros que salen. Estos datos son fundamentales para analizar patrones de vuelo, eficiencia operativa y preferencias de los pasajeros en el aeropuerto durante ese período. El desarrollo técnico de este punto se presenta en la sección **5. Recolección de datos** del notebook adjunto: **ENTREGA_UNO_PUNTOS_5_Y_6.ipynb**

6 ENTENDIMIENTO DE LOS DATOS

El reporte del análisis exploratorio de datos y calidad de los datos se presenta en la sección **6. Entendimiento de los datos** del notebook adjunto: **ENTREGA_UNO_PUNTOS_5_Y_6.ipynb**

7 CONCLUSIONES E INSIGHTS

- Este proyecto subraya el valor de los datos operativos y comerciales como activos estratégicos. Utilizar análisis de datos avanzados y técnicas de ciencia de datos puede proporcionar a OPAIN S.A. información vital para tomar decisiones más informadas y eficientes.
- Del análisis exploratorio inicial podemos observar que variables como la hora del día, la cantidad de pasajeros en el muelle y el tipo de aerolínea tienen un impacto significativo sobre las transacciones y las ventas. En los próximos avances esperaríamos evidenciar que estas variables tengan un impacto mayor en comparación al resto de los features observados. Adicional a esto, el año 2020 al ser impactado por la pandemia del Covid-19 tuvo unas ventas extremadamente irregulares. Por esto, lo mejor será no tomar datos de este año. Además, al enfocarse en los vuelos de salida, se busca mejorar la experiencia del pasajero. Durante este período previo a la salida, los pasajeros son más propensos a involucrarse en actividades como compras, lo que brinda oportunidades significativas para aumentar las ventas comerciales.
- **Optimización de Asignación de Puertas:** La asignación estratégica de puertas de embarque, considerando la coordinación con los horarios de mayor afluencia de pasajeros, puede prevenir congestiones y mejorar la experiencia del cliente. Esto probablemente aumentaría las ventas en las tiendas cercanas a estas puertas.
- **Sincronización de Horarios:** Sincronizar las operaciones de las tiendas con los horarios de mayor afluencia de pasajeros puede aumentar significativamente las oportunidades de venta y mejorar la eficiencia comercial. Esto podría lograrse mediante la programación de promociones y eventos especiales en momentos estratégicos.

8 MOCKUP

Mockup conceptual de la herramienta a desarrollar (*Es una representación conceptual y no refleja la interfaz final del producto*)

https://www.canva.com/design/DAFuoRRkrc4/XBKCnspP7GvuYoYQk9oa_w/view?utm_content=DAFuoRRkrc4&utm_campaign=designshare&utm_medium=link&utm_source=publishsharelink#1

9 BIBLIOGRAFÍA

- Bower, et al. (2022). Back to the future? Airline sector poised for change post-COVID-19. McKinsey & Company
- Krishnan, K. (2013). Data Warehousing in the Age of Big Data. Newnes.
- Pinto, J. K. (2015). Gerencia de proyectos (Tercera edición). Pearson.
- Nicholas, J. M. (2017). Project Management for Engineering, Business and Technology (5a Ed.). Partes 1, 2 y 3.
- Project Management Institute. (2013). A Guide to the Project Management Body of Knowledge (PMBOK® Guide) (5a Ed.). PMI.org. (6a ed. a partir de septiembre de 2017).