

## Prueba - Predicción del precio de vuelos

En esta prueba validaremos nuestros conocimientos del módulo, pondrás a prueba todo lo visto hasta el momento. Para lograrlo, necesitarás aplicar los conceptos vistos de calidad de datos, análisis exploratorio, preprocesamiento de datos y sobre modelamiento y evaluación. Pondrás a prueba tus habilidades como Data Scientist utilizando los archivos **business.csv** y **economy.csv**.

Lee todo el documento antes de comenzar el desarrollo **Individual**, para asegurarte de tener el máximo de puntaje y enfocar bien los esfuerzos.

### Descripción

En esta emocionante prueba final de Data Science, nos adentraremos en un desafiante problema del mundo real relacionado con la industria de la aviación. Nuestro objetivo es desarrollar un modelo predictivo capaz de estimar el precio de un boleto de avión basándonos en diversas características del vuelo. Este problema representa un escenario común en la industria de viajes, donde las aerolíneas y las agencias de viajes buscan ofrecer a los pasajeros tarifas competitivas y precisas.

Imagina que trabajas en una agencia de viajes en línea que desea ofrecer a los clientes una herramienta que les permita obtener estimaciones precisas del precio de los boletos de avión para su próximo viaje. Sin embargo, el precio de un boleto de avión puede variar significativamente según múltiples factores, como la aerolínea, el destino, la temporada, la clase de servicio y la anticipación de la reserva. Aquí es donde entra en juego el Data Science.

El desafío radica en desarrollar un modelo de Machine Learning que pueda predecir el precio de un boleto de avión en función de estas diversas variables. Esto implica limpiar, explorar y transformar los datos, seleccionar características relevantes, entrenar y afinar modelos predictivos y evaluar su rendimiento de manera precisa.

Esta prueba final de Data Science te desafiará a aplicar todos los conocimientos adquiridos durante el curso para abordar un problema del mundo real. Al final de este desafío, habrás desarrollado un modelo de Machine Learning capaz de estimar con precisión el precio de un boleto de avión, lo que puede tener un impacto significativo en la industria de viajes y mejorar la experiencia de los viajeros.

Prepárate para sumergirte en este emocionante proyecto y demostrar tus habilidades en Data Science para abordar un problema complejo y relevante. ¡Buena suerte!

Considerando los datos entregados en las bases de datos, realiza las siguientes actividades:

1. Describe con detenimiento el problema de negocio que se desea resolver, y cómo se va a hacer esto. Indica la metodología, tareas a realizar, variable objetivo a predecir, etc.
2. Carga los dataset entregados y genera un reporte de calidad de los mismos. Indica qué estrategias se van a utilizar para aquellos puntos encontrados (Indicar nulos, outliers, valores perdidos, que se hará con esto, etc.)
3. Genera un análisis exploratorio de los dataset entregados, un análisis univariado y bivariado. Prioriza los gráficos más importantes y entrega una conclusión a partir de estos.
4. Realiza un análisis de correlaciones entre las diferentes variables existentes, identificando cuáles son las variables más importantes para la predicción de la variable objetivo. Utiliza las técnicas vistas en el curso.
5. Realiza las transformaciones necesarias para realizar el modelamiento posterior y crea las variables que estimes convenientes con ayuda del análisis previo y la expertise del negocio.
6. Genera una función que encapsule el tratamiento de datos necesario, para entregar un dataset limpio y procesado a partir del dataset original.
7. Elige al menos 3 modelos candidatos para resolver el problema. A partir de esto, genera un conjunto de entrenamiento y prueba, para luego entrenar los diferentes modelos.
8. Elige una grilla de hiperparametros y luego optimízalos, buscando la mejor combinación para cada grilla. Guardar los modelos entrenados.
9. Define al menos 3 métricas para evaluar los modelos entrenados y genera gráficos de comparación. Elige un baseline para ver qué tan buena es tu opción respecto a ese baseline y concluye.
10. Genera una conclusión final respecto a qué tan útiles son los resultados encontrados para resolver el problema propuesto y define cuáles podrían ser los próximos pasos para el proyecto.

## Requerimientos

1. Comprende un problema de negocio y sus datos, a partir de un contexto dado y un análisis exploratorio de los datos involucrados considerando indicadores estadísticos, gráficos y análisis de correlaciones. A partir de esto, selecciona los datos más relevantes y las variables objetivo. **(3 puntos)**
2. Crea una función para limpiar el dataset, realizar las transformaciones necesarias para el modelamiento posterior y crea las variables necesarias para el análisis. **(1 punto)**
3. Elige y entrena modelos de resolución del problema, ajusta sus hiperparámetros y los guarda. **(3 puntos)**
4. Evalúa los modelos con al menos 3 métricas y utilizando gráficos de comparación. Concluye al respecto y define pasos futuros. **(3 puntos)**



¡Mucho éxito!

### Consideraciones y recomendaciones

¡Importante! Como vimos en todo el curso, este es un framework iterativo y a pesar de que el trabajo está planteado de una forma lineal, no tiene que hacerse necesariamente así ya que muchas veces hay que volver unos pasos atrás para poder proceder de manera correcta. Por ejemplo; ¿qué ocurre si tengo que transformar variables para ver las correlaciones o visualizaciones?