

## Predicción de la demanda futura de hotel.

### Resumen del Proyecto:

El objetivo de este proyecto es predecir la demanda futura de un hotel utilizando técnicas de aprendizaje automático. A lo largo del proyecto, se han realizado diversas tareas, incluyendo la importación de bibliotecas, limpieza de datos y la implementación de un modelo de regresión logística para predecir cancelaciones.

### Sección 1: Importación de Bibliotecas y Carga de Datos:

En esta sección, se importaron las bibliotecas necesarias, se configuraron las opciones de visualización y se cargaron los datos desde el archivo CSV. Esto estableció las bases para el proyecto.

### Sección 2: Comprobación de Valores Nulos en el Conjunto de Datos:

Primero realizamos un análisis del dataset y la estructura de sus columnas, para comprender el orden y la estructura de los datos. Se verificó la presencia de valores nulos en el conjunto de datos para identificar áreas que requieren imputación de datos faltantes.

### Sección 3: Limpieza de Datos:

Todas las columnas tienen los datos predefinidos excepto country, agent, company para estas columnas definimos una regla para asignar los valores faltantes. En esta sección, se realizaron operaciones para reemplazar valores vacíos y eliminar filas específicas del conjunto de datos. Esto mejoró la calidad del conjunto de datos al abordar valores nulos y filas irrelevantes.

### Sección 4: Separación de Datos por Tipo de Hotel y Eliminación de Reservas Canceladas:

Para entender la naturaleza de los datos y de las implicaciones que podría tener el análisis de datos en las operaciones del hotel, realizamos algunas segmentaciones en la data para obtener ciertas características. Se separaron los datos en dos conjuntos: "Resort Hotel" y "City Hotel". Las reservas canceladas se eliminaron para centrarse en las reservas confirmadas.

### Sección 5: Análisis de Huéspedes por Países y Visualización:

Se obtuvo información sobre la procedencia de los huéspedes y se creó un gráfico circular para visualizar la distribución de huéspedes por países. Claramente se puede observar en la figura 1 que la mayoría de huéspedes provienen de Portugal y países europeos, por

lo que sería una buena estrategia basada en data aumentar la infraestructura para esos países o buscar crecimiento en los países donde tienen menos presencia.

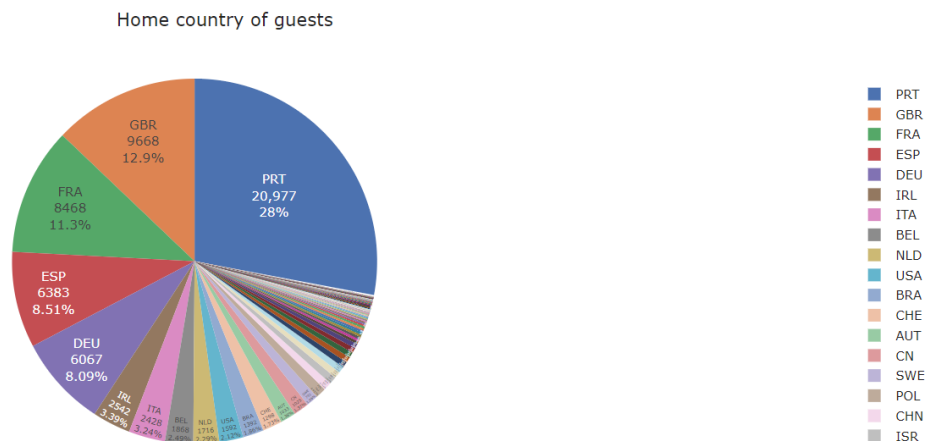


Figura 1. Países de origen de los huéspedes.

## Sección 6: Cálculo de la Tarifa Diaria por Persona:

Realizamos un análisis respecto al precio, para esto incluiremos solamente información de adultos y niños sin los bebés, también asumimos el euro como la moneda utilizada en el hotel. Se calculó la tarifa diaria por persona para mejorar la comprensión de la estructura de precios de los hoteles. La respuesta del código fue la siguiente:

De todas las reservas sin cancelar, el promedio de precios es:  
 Resort hotel: 47.49 € por noche y persona.  
 City hotel: 59.27 € por noche y persona.

Figura 2. Promedio de precios de las reservas que no fueron canceladas

Podemos observar que el precio de reserva es más elevado para el resort que para el hotel en la ciudad. Analizamos igualmente la variación de precios de las reservas a en todos los meses del año, con el fin de predecir futuras demandas en dichos meses. El resultado obtenido se mostrará en la siguiente gráfica, comparando el comportamiento de los precios en ambas sedes del hotel:

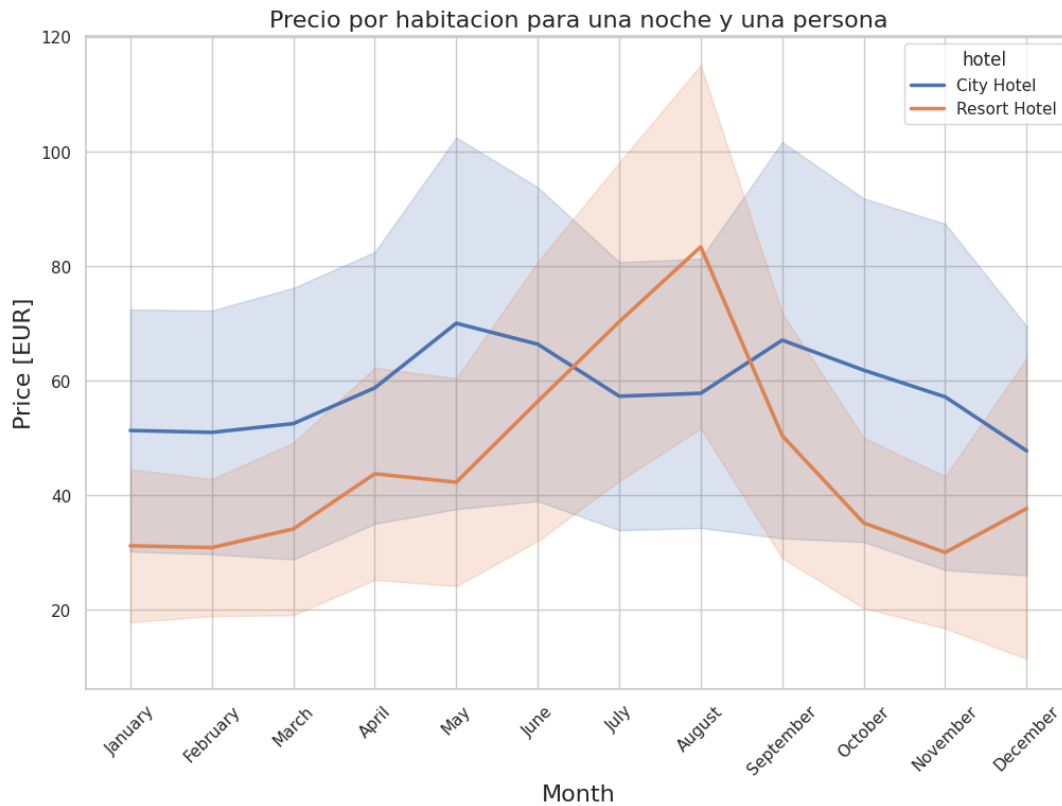


Figura 3. Precios de reservas por habitación para una noche por persona a lo largo del año.

Luego de ver la variación de precios en las dos sedes del hotel, procedimos a seguir depurando los datos que estaban vacíos o nulos. Además, eliminamos las columnas cuya cantidad de valores únicos era bastante alta, lo cual iba a incrementar el número de posibilidades, haciendo el código menos eficaz pues le aporta información que por el momento no necesitamos para nuestras predicciones. Posterior a esto dividimos aleatoriamente los datos correspondientes a entrenamiento del modelo y los que tomaremos para hacerle pruebas al mismo.

## Sección 7: Implementación de un Modelo de Regresión Logística:

Se implementó un modelo de regresión logística para predecir cancelaciones de reservas. El modelo se entrenó, evaluó y se proporcionaron métricas de rendimiento. Los resultados del modelo se muestran a continuación:

```

Classification Report is:
      precision    recall  f1-score   support

      0       0.77       0.88       0.82     15926
      1       0.66       0.46       0.54       7916

 accuracy      0.74     23842
  macro avg       0.71     23842
 weighted avg       0.73     23842

Confusion Matrix:
[[14064 1862]
 [ 4309 3607]]

Training Score:
74.02478922863975

Accuracy Score of Logistic Regression is: 0.7411710426977602

```

Figura 4. Métricas del modelo de regresión logística.

El informe de clasificación muestra que el modelo de regresión logística alcanza una precisión general del 74%. No obstante, se evidencia un desempeño desigual en la clasificación de las reservas de hotel, ya sea como canceladas o no canceladas. Para la clase de reservas no canceladas (0), el modelo logra una precisión del 77% y un recall del 88%, lo que indica una efectividad en la identificación de reservas que no se cancelarán. Por otro lado, en la clase de reservas canceladas (1), la precisión es del 66% y el recall del 46%, lo que denota una menor precisión en la predicción de cancelaciones. La puntuación F1 muestra un valor del 82% para las reservas no canceladas y del 54% para las reservas canceladas. Al observar la matriz de confusión, se revela que se cometieron 1,862 falsos positivos y 4,309 falsos negativos. La puntuación de entrenamiento del modelo se sitúa en el 74.02%. En resumen, estos resultados indican que el modelo es más preciso en la predicción de reservas que no se cancelarán, aunque su precisión disminuye en el caso de reservas canceladas.

#### Conclusiones:

Hasta este punto del proyecto, se ha realizado una sólida preparación de datos y se ha implementado un modelo de regresión logística para predecir cancelaciones. El modelo implementado tuvo una mayor precisión en uno de los aspectos (reservas no canceladas) a predecir, lo cual es un tema a mejorar en la siguiente entrega. Los próximos pasos podrían incluir la mejora del modelo, la exploración de otras técnicas de aprendizaje automático y un análisis más profundo de los factores que influyen en la demanda de hoteles con el fin de lograr unas mejores métricas y que nuestro modelo sea más seductor a nivel empresarial.