

AUTORA: ROCIO BUSTAMANTE



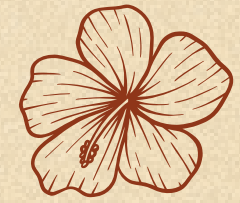
Inn

HOTEL BOUTIQUE

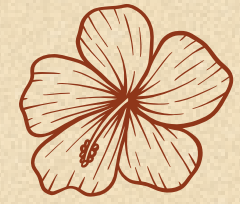
¿Cómo lograr que la mayoría
de las reservas se concreten?



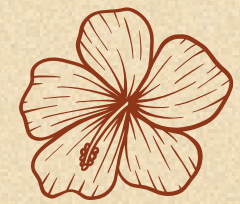
AGENDA



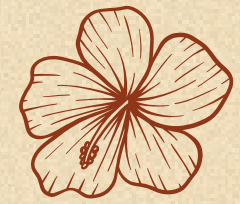
Contexto y audiencia.



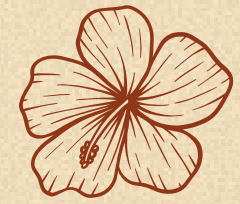
Objetivo.



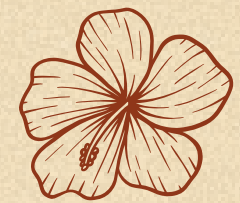
Hipótesis y preguntas de interés.



Análisis exploratorio.



Insights.



Conclusiones.



Contexto y audiencia.

El hotel enfrenta pérdidas directas de ingresos, costos adicionales de distribución y necesitan ajustar los precios para revender habitaciones vacías debido a las cancelaciones de reservas. Las tecnologías de reserva online han exacerbado este problema. Esto genera un impacto significativo en la rentabilidad y la operación del hotel.

Procederé a analizar los datos del hotel desde el 01/01/2017 al 31/12/2018.

La resolución de este problema puede ser de gran utilidad para cadenas de hotelería y turismo.



Objetivo

El creciente número de cancelaciones entre los años 2017 y 2018 exige una solución basada en un modelo de aprendizaje automático que pueda predecir si las futuras reservas tienen tendencia a ser canceladas.

Creando así políticas rentables para cancelaciones y reembolsos, y brindar una mejora en el servicio.



Hipótesis

Teniendo una primera vista de la cantidad de reservas en los años 2017 y 2018 en el hotel, se puede ver que aproximadamente un 30% de las reservas fueron canceladas. Esto puede ocurrir por distintos factores, los cuales se van a analizar para poder obtener respuestas a las siguientes preguntas.





Preguntas de interés



"¿Por qué vía de mercado las reservas son más probables a ser canceladas?"



"¿Qué meses tendrán más demanda de reservas?"




"Las reservas que son canceladas, ¿Son mayormente realizadas por clientes del hotel?"



"¿Cuáles son los factores que influyen en las cancelaciones?"



Metadata

- NÚMERO DE FILAS: 36275
 - NÚMERO DE COLUMNAS: 19
 - TIPOS DE VARIABLES: Flotantes (1), Enteros (13), Objetos (5)
 - VARIABLES DE INTERES: "booking_status" (reservas canceladas y no canceladas, VARIABLE TARGET), "rescan" (filtro de reservas canceladas), "huespedes" (cantidad total de huespedes), "arrival month" (mes en que se realiza la reserva), "market_segment_type" (segmentador de mercado).
 - Los datos fueron extraídos de Kaggle.
- 

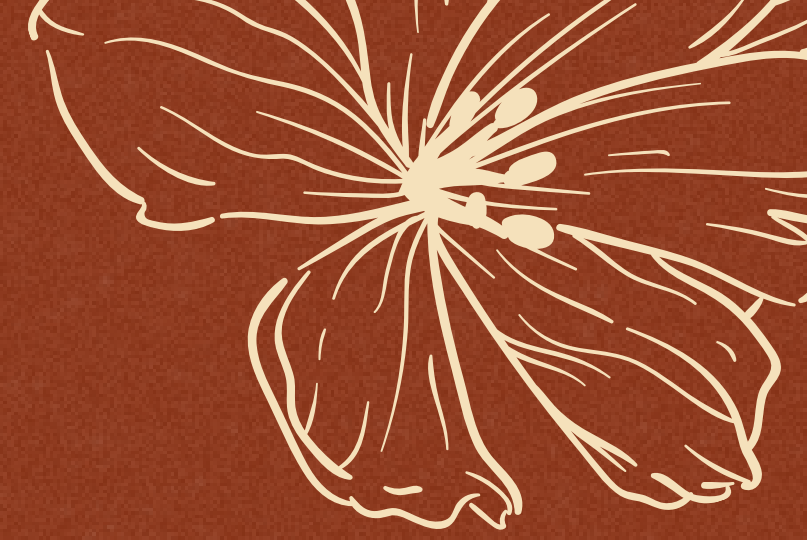
Análisis exploratorio

Estado de reserva por año

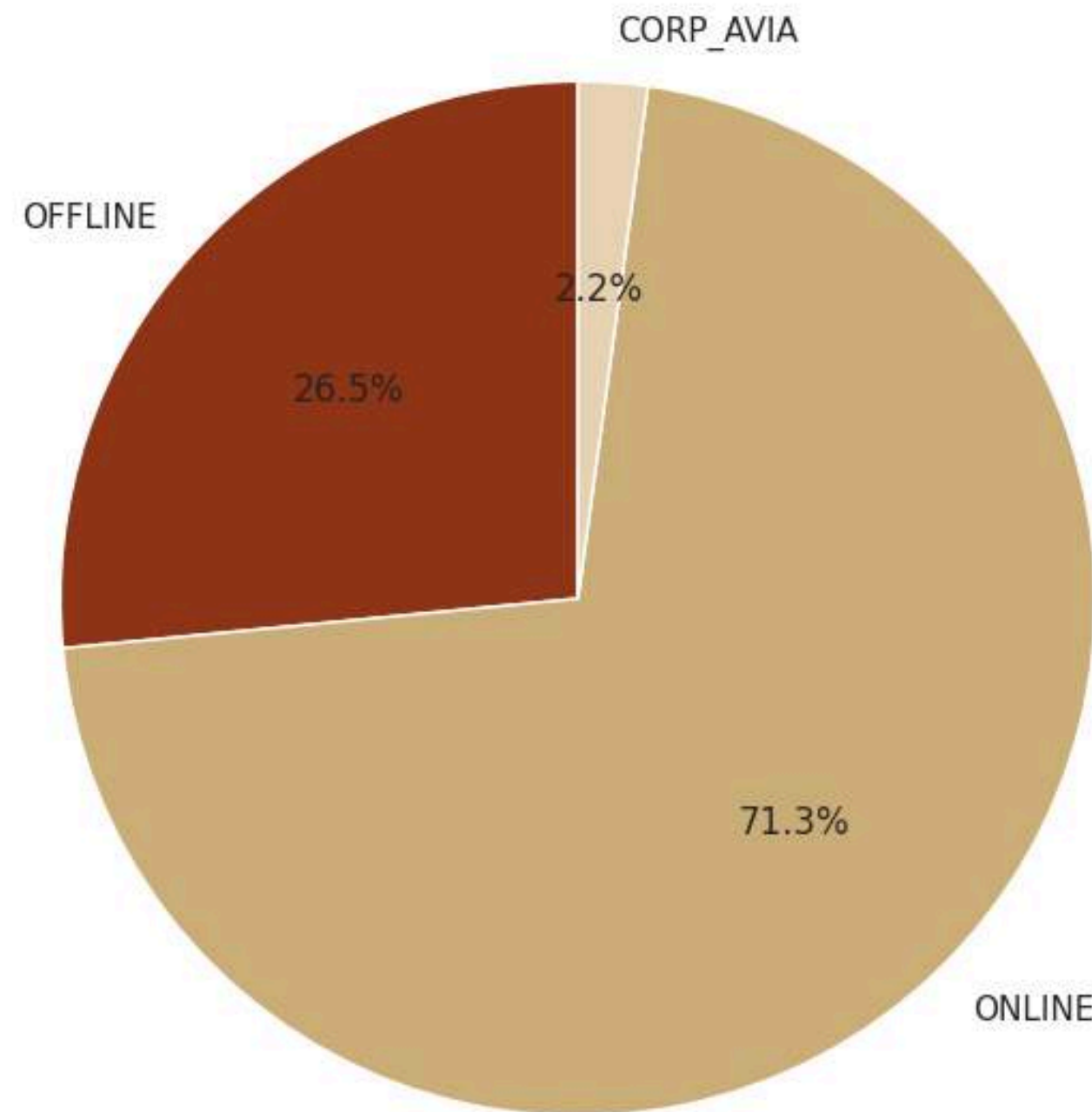
Si observamos el gráfico de la izquierda podemos darnos cuenta de que hubieron muchas más reservas no canceladas, que canceladas tanto en 2017, como en 2018. Y además podemos ver un incremento de reservas en el año 2018.



Análisis exploratorio



Porcentajes de reservas canceladas según el segmentador de mercado



¿Por qué vía de mercado las reservas son más probables a ser canceladas?

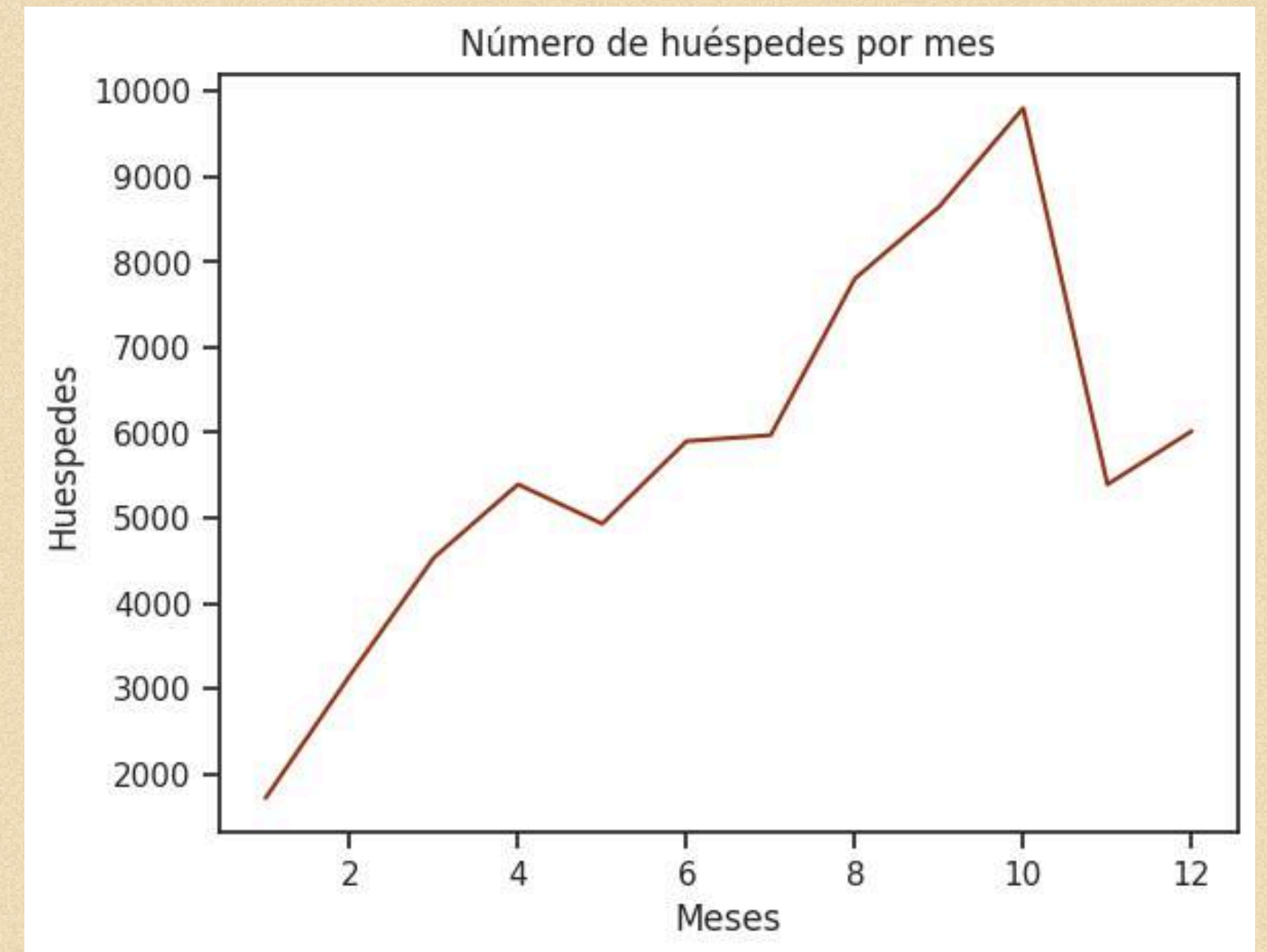
Con la ayuda del gráfico se puede ver que la mayor cantidad de reservas canceladas se realizan de manera online.



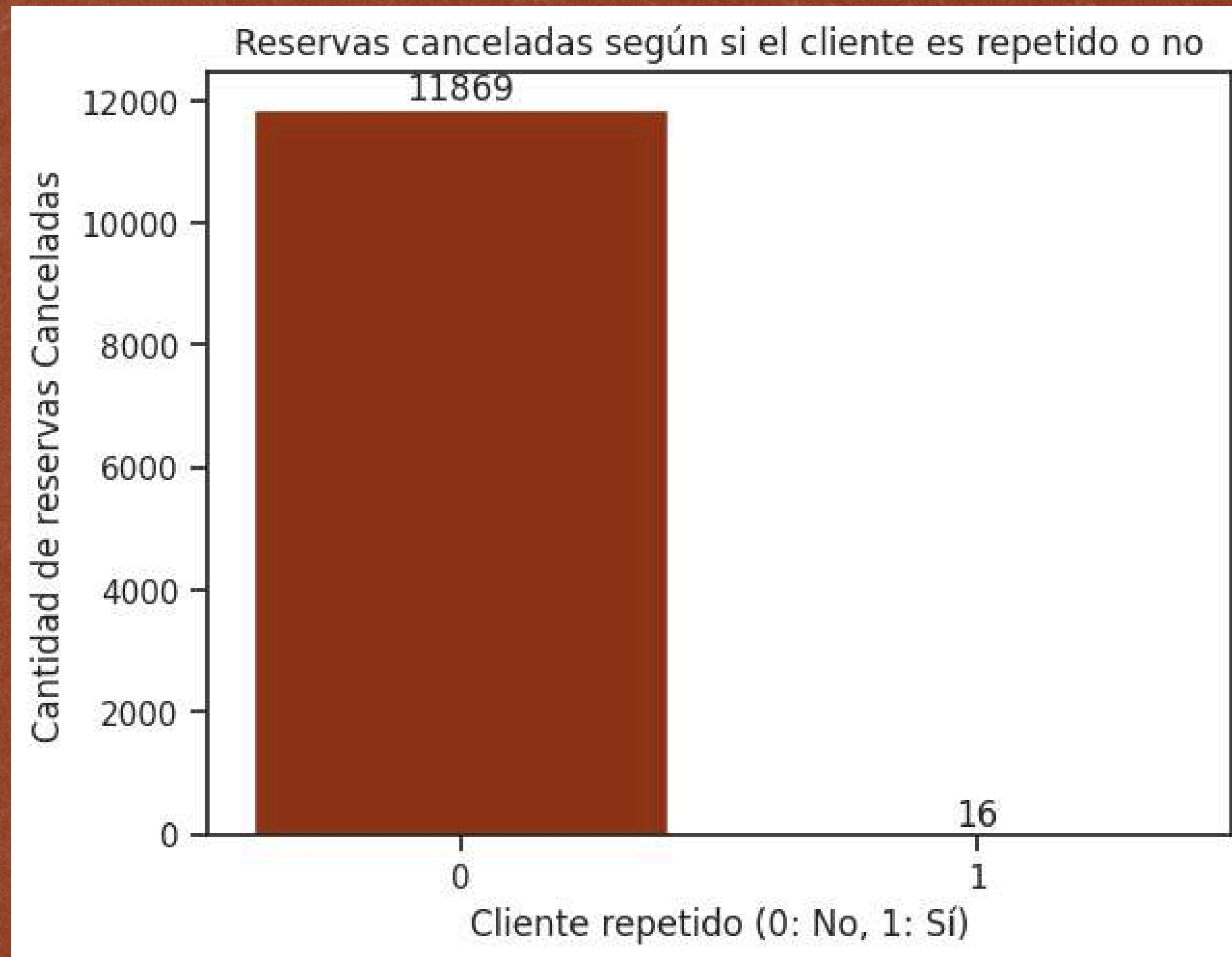
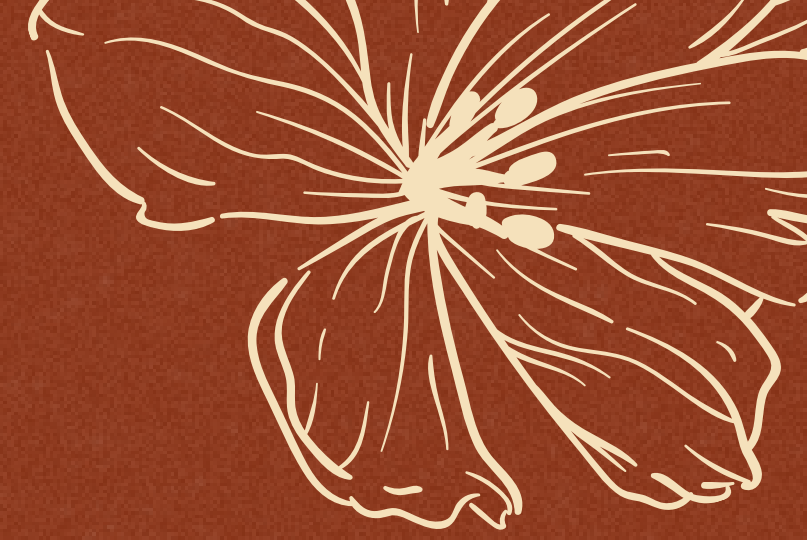
Análisis exploratorio

¿Qué meses tendrán más demanda de reservas?

En el gráfico de líneas podemos ver que en el mes de octubre hubieron más huéspedes en el hotel. Por lo que es un mes con alta demanda y los clientes son más fieles a las reservas.



Análisis exploratorio



Las reservas que son canceladas, ¿Son mayormente realizadas por clientes del hotel?

Podemos darnos cuenta de que por lo general las reservas que son canceladas es por personas que no son clientes del hotel.



Insights

- La cantidad de reservas canceladas son 11885, guardadas en la variable "rescan".
- La columna "arrival month" nos ayudará a poder filtrar el mes con más demanda de reservas.
- "Online" es la manera en que se realiza la mayor cantidad de reservas y cancelaciones.
- "Complementary", "Corporate" y "Aviation" son segmentos de mercado en los que menos se realizaron reservas.
- Las reservas son canceladas por personas que no son clientes fijos del Hotel.

Modelos elegidos para realizar pruebas:

- Regresión logística y Random Forest.

Resultado de métricas:

- Regresión logística:

Matriz de confusión: array ([[0, 2428] [0, 4827]])

Cálculo la tasa de verdaderos positivos:
100.0

Accuracy: 0.6653342522398346

- Random Forest:

Recall: 1.0

Accuracy: 1.0

Insights

- Modelo elegido: **Regresión Logística**

Esta vez se realizó con la variable target y con la variable "caract" (que se creó con las características de las reservas). Además se le realizó una optimización de hiper parámetros "GridSearchCV".

Resultado de las métricas:

Matriz de confusión: array([[1308, 1120], [472, 4355]])

Cálculo la tasa de verdaderos positivos: 92.42125984251969

Accuracy: 0.780565127498277

- El modelo final funciona mejor que los modelos anteriores. Se realizó un Cross Validation que tuvo un rendimiento del 78% en promedio en los datos de prueba, en diferentes divisiones de la misma, y las métricas dan mejores resultados. La curva de aprendizaje nos deja ver que el modelo acierta los datos de manera correcta y sin overfitearse.

Conclusiones

Luego de las visualizaciones que se realizaron con la base de datos, noté que la mayor cantidad de cancelaciones las realizan vía online usuarios que no son clientes fijos. Por lo que se puede utilizar un algoritmo de clasificación, que me permita diferenciar los usuarios de reservas online que tienden a cancelar, de los usuarios que no cancelan, según sus cualidades. Por ende el algoritmo al detectar esas cualidades podría ofrecer ofertas en las habitaciones, promociones, o mejoras en el servicio, y así no tener pérdidas.

