☐ Presentation of Hotelling booking trend and Cancellation

Name :Akash Singh
date:25/08/2024
M:8709628348

Objective:

- **Analysis hotel booking data to identify trends and patterns.**
- **Understand factors affecting cancellations and customer demographics.**
- **Provide actionable insights for pricing, marketing, and customer segmentation.**

Data Overview:

- **Dataset:** Hotel bookings (2014-2017)
- **Key Features:** Booking date, customer demographics, stay details, cancellation status, and more.

Data Cleaning:

Handled missing values and inconsistencies.

Removed outliers and corrected errors.

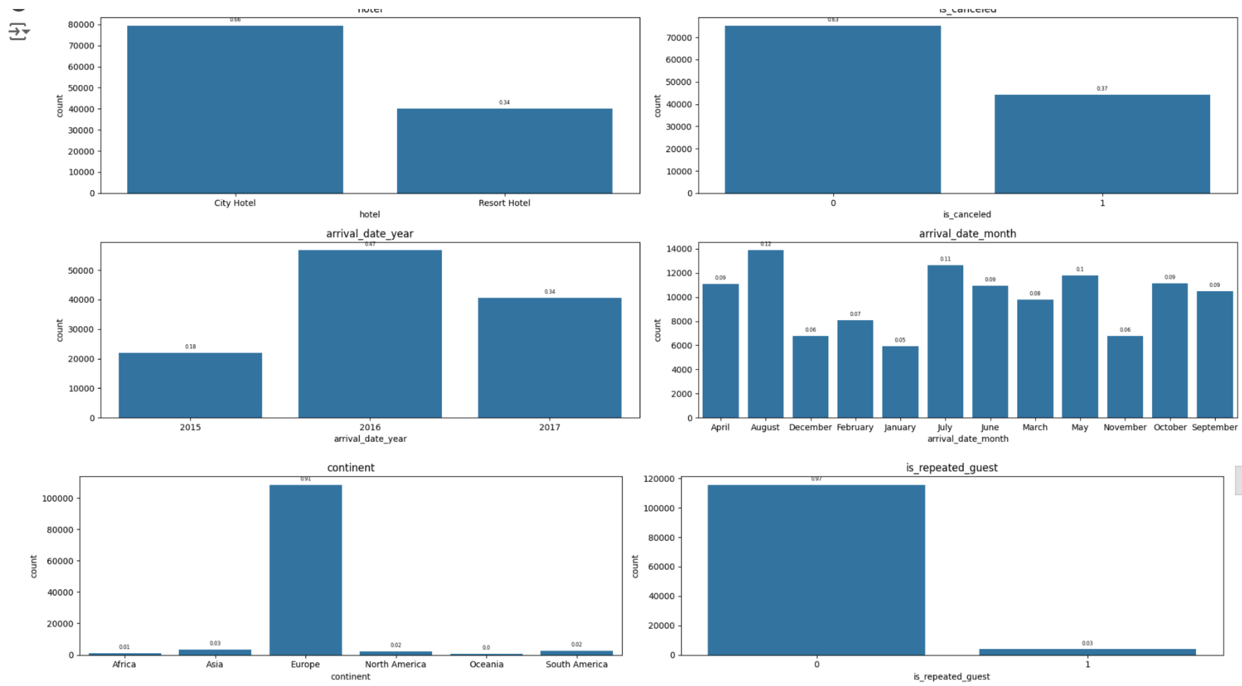
Engineered features for improved model performance.

meal	0
country	488
market_segment	0
distribution_channel	0
is_repeated_guest	0
previous_cancellations	0
previous_bookings_not_canceled	0
reserved_room_type	0
assigned_room_type	0
booking_changes	0
deposit_type	0
agent	16340
company	112593

```
#Filling the missing values
data['children'].fillna(data['children'].mode().values[0],inplace=True)
data['country'].fillna(data['country'].mode().values[0],inplace=True)
data['agent'].fillna(data['agent'].mode().values[0],inplace=True)
data['company'].fillna('Unknown',inplace=True)
```

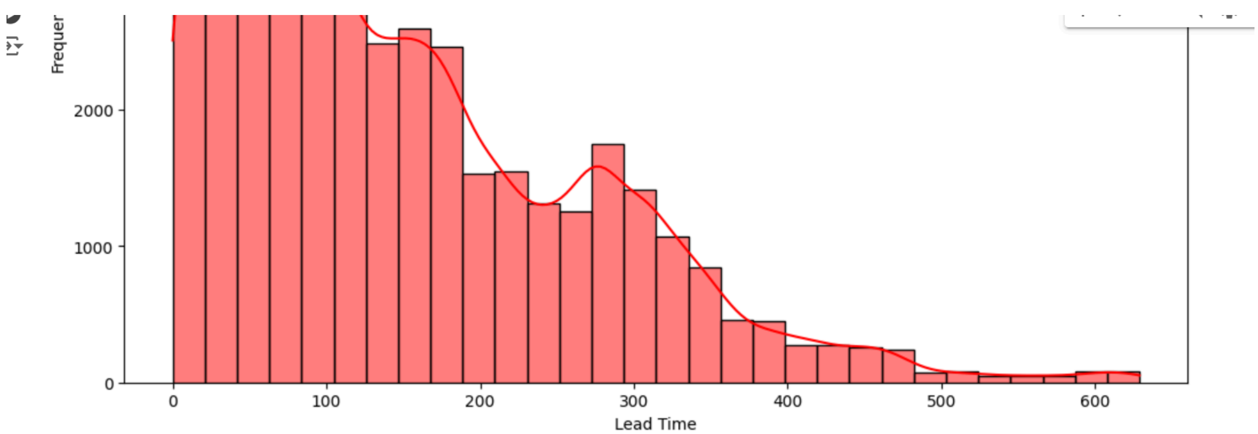
EDA Insights:

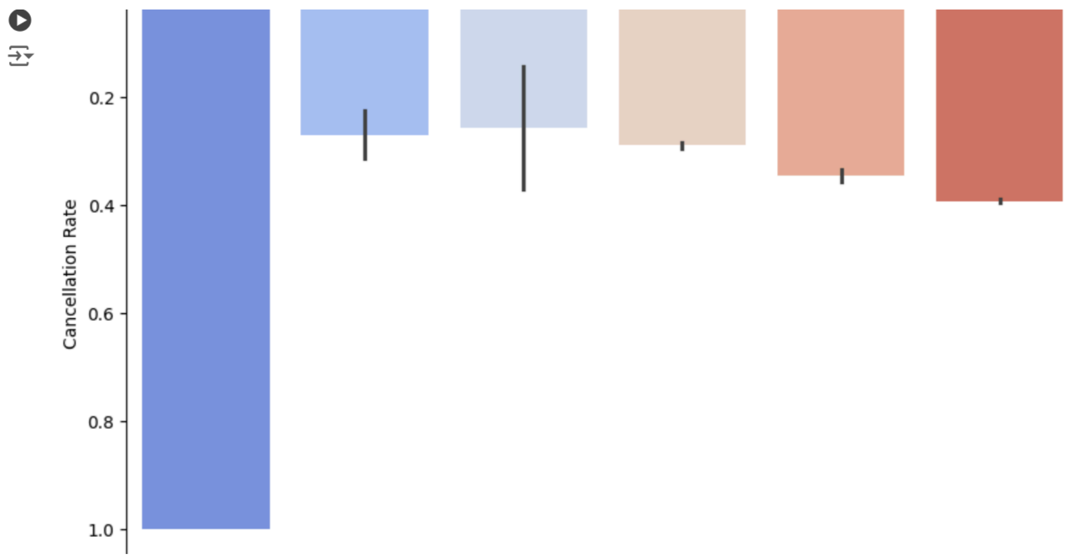
- **Booking Trends:** Seasonal patterns and peak periods.
- **Customer Demographics:** Age, family size, nationality.
- **Visualizations:** Trends in bookings and cancellations.



Cancellation Patterns:

- **Trends:** High cancellation rates in specific periods.
- **Factors:** Booking lead time, room type, and market segment influence cancellations.





Hypothesis Testing:

- **Hypotheses Tested:** Factors affecting cancellation rates.
- **Results:** Significant variables include booking lead time and room type.

```
#results
print(f"Chi2 Statistic: {chi2}")
print(f"P-value: {p}")
print(f"Degrees of Freedom: {dof}")
print(f"Expected Frequencies Table:\n{ex}")
```

```
Chi2 Statistic: 5321.73290110073
P-value: 0.0
Degrees of Freedom: 1
Expected Frequencies Table:
[[59620.31885418 35077.68114582]
 [15545.68114582  9146.31885418]]
```

```
[ ] #hence p value is less then 0.05
```

Hence the p-value is less than 0.05, we reject the null hypothesis and conclude that bookings made more than 6 months in advance have a higher cancellation rate.

```
data['day_of_week'] = data['reservation_status_date'].dt.day_
data['booking_day'] = data['day_of_week'].apply(lambda x: 'We
#average adr
adr_by_day = data.groupby('booking_day')['adr'].mean()
#ADR
weekday_adr = data[data['booking_day'] == 'Weekday']['adr']
weekend_adr = data[data['booking_day'] == 'Weekend']['adr']

# t-test
from scipy.stats import ttest_ind
t_stat, p_value = ttest_ind(weekday_adr, weekend_adr)
print(f'T-statistic: {t_stat}, P-value: {p_value}')
```

```
T-statistic: -15.206866062360545, P-value: 3.56448899560107e-
```

value of 3.56448899560107 × 1 0 − 52 3.56448899560107×10 −52

Neekday bookings have a higher ADR than weekend bookings.

Model Overview:

- **Models Used:** Logistic Regression, Random Forest.
- **Performance Metrics:** Accuracy, precision, recall, F1 score.
- **Key Findings:** Model performance and predictive power.

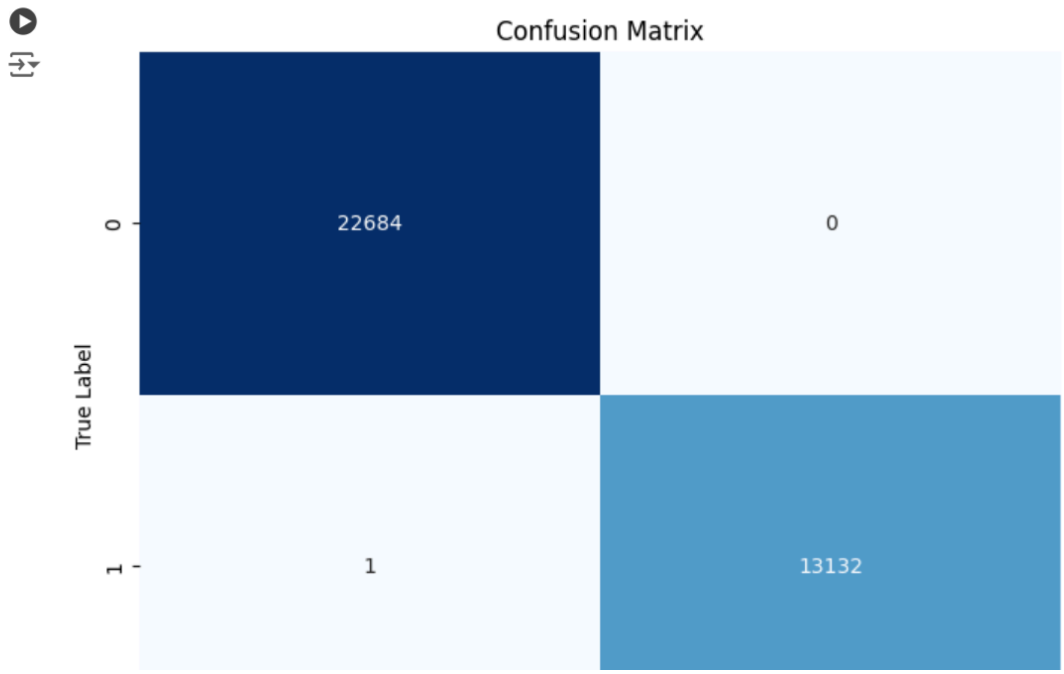
```
# initialize and train model
lr_model = LogisticRegression(max_iter=1000)
lr_model.fit(X_train_scaled, y_train)

# predictions
lr_predictions = lr_model.predict(X_test_scaled)

#report
print("Logistic Regression (Scaled):")
print("Accuracy:", accuracy_score(y_test, lr_predictions))
print("Precision:", precision_score(y_test, lr_predictions))
print("Recall:", recall_score(y_test, lr_predictions))
print("F1 Score:", f1_score(y_test, lr_predictions))
```

```
Logistic Regression (Scaled):
Accuracy: 0.989558031102549
Precision: 0.9998432970304787
Recall: 0.971674407979898
F1 Score: 0.9855576150756874
```

Logistic Regression (Scaled): Accuracy: 0.989558031102549 Precision: 0.9998432970304787 Recall: 0.971674407979898 F1 Score: 0.9855576150756874



Recommendations:

- **Pricing Strategies:** Implement dynamic pricing based on trends.
- **Customer Segmentation:** Target high-value and repeat guests.
- **Marketing Focus:** Optimize channels and geographic targeting.
- City hotel have to change the policy of pricing to decrease the cancellation.