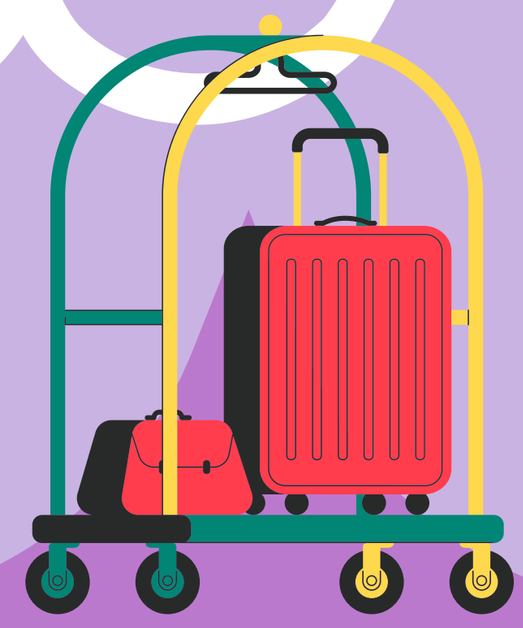
Hotel Reservation Cancelation Prediction



Group Members:
Sakshi Lade - 202201040218, Palak Yerawar - 202201040195, Surabhi Kharkate - 202201040215

PROBLEM STATEMENT

Develop a machine learning solution to predict hotel reservation cancellations using historical data and key factors, enabling proactive management and improving revenue and customer experience.

DATA COLLECTION & PREPROCESSING

Data Collection:
The dataset includes historical hotel reservation records with details like guest count, booking duration, room type, and booking status.

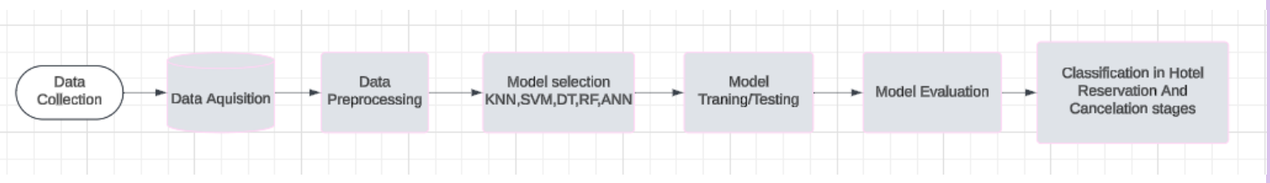
Cleaning:
The dataset was complete with no missing values.
- Irrelevant columns, such as `Booking_ID`, were removed.

Encoding:
Categorical features (`type_of_meal_plan`, `room_type_reserved`, `market_segment_type`) were transformed using one-hot encoding.
The target variable, `booking_status`, was encoded as:
- `Canceled` → 1
- `Not_Canceled` → 0

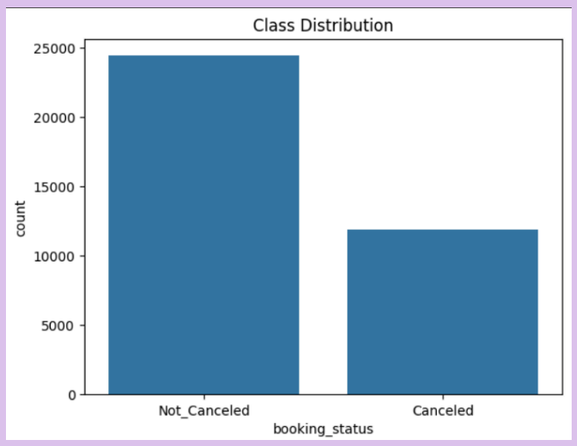
MODEL BUILDING

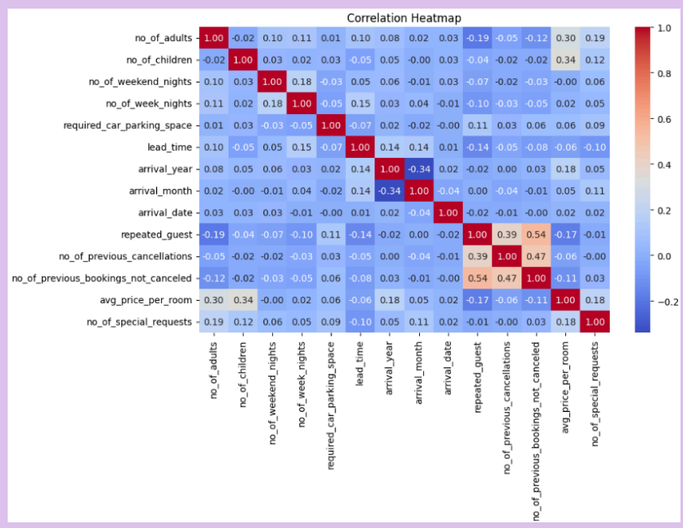
Data Collection
↓
Collect historical booking data from PMS and OTAs.
↓
Clean data, encode features, and scale numerical values.(Data preprocessing)
↓
Feature Engineering
Add features like lead time and seasonal trends.
↓
Train models (Logistic Regression, Random Forest, XGBoost).
↓
Assess performance using AUC-ROC and Precision/Recall.
↓
Deploy using Flask for real-time predictions.
↓
Monitoring
Regularly retrain with updated data.

FLOWCHART



GRAPH ANALYSIS





MLOPS WORKFLOW

- Data Collection: Automate data ingestion from PMS and OTA systems.
- Preprocessing: Clean, encode, and scale using pipelines.
- Model Training: Train and evaluate Logistic Regression, Random Forest, XGBoost.
- Deployment: Deploy the best model with Streamlit Model using Joblib
- Monitoring: Track accuracy and feature drift.
- Retraining: Automate retraining with fresh data.

TOOL FRAMEWORK

- Data Handling : Pandas, NumPy
- Data Visualization : Matplotlib, Seaborn
- Model Development : Scikit-learn, XGBoost, TensorFlow/Keras (optional for advanced modeling)
- Model Deployment : Streamlit,Joblib
- MLOps & Monitoring : MLflow, Kubernetes, Prometheus
- Version Control & CI/CD : Git, GitHub Actions

CONCLUSION

The project revealed higher cancellations for couples, weeknight bookings, longer lead times, and online reservations, with peaks in 2018's July and October. The Decision Tree Classifier, with 85% accuracy, offers an effective tool to reduce cancellations and improve hotel revenue and customer satisfaction.