AIRLINES CUSTOMER SATISFACTION

INTRODUCTION

The dataset provided by Invistico Airlines contains valuable information about their customers' experiences and satisfaction levels. With the aim of predicting future customer satisfaction and improving service quality, this dataset encompasses various customer attributes and feedback on different aspects of their flights. The primary objectives of this dataset are to predict customer satisfaction and identify areas for service improvement.

BUSINESS UNDERSTANDING

Airlines face intense competition, making customer satisfaction a critical factor for loyalty and revenue growth. The project aims to identify the key drivers of customer satisfaction, build a predictive model to classify customers as "Satisfied" or "Dissatisfied," and leverage the findings to guide business decisions. Key stakeholders include:

- **Airline Management:** To enhance service delivery and improve satisfaction scores.
- **Customer Service Team:** To address pain points like delays or poor service quality.
- **Marketing Team:** To develop targeted campaigns for retaining dissatisfied customers.
- **Operations Team:** To optimize flight schedules and resource allocation for better service.

DATA UNDERSTANDING

The dataset comprises 129,880 rows and 23 columns, with no duplicated records but 393 missing values in the 'Arrival Delay in Minutes' column. The data distribution is generally balanced across features. However, rating features revealed an issue with a disproportionately low number of '0' ratings, which could hinder analysis. To resolve this, '0' ratings were combined with '1' ratings, ensuring a consistent scale from 1 to 5 for meaningful analysis. This adjustment allows for a more robust interpretation of the dataset.


Correlation Heatmap
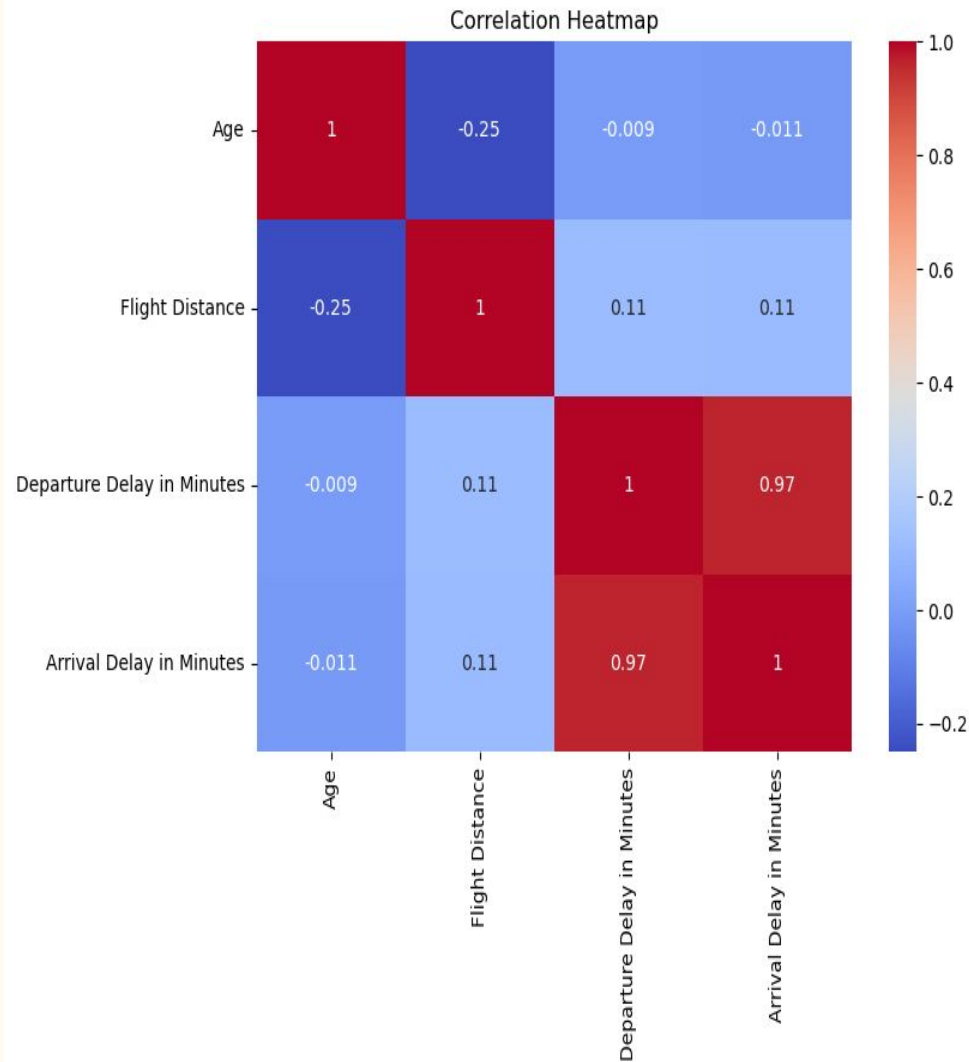
EXPLORATORY DATA ANALYSIS (EDA)

The dataset revealed no duplicate records and 393 missing values in the **'Arrival Delay in Minutes'** column. The target variable, **'Satisfaction'**, is moderately balanced between satisfied and dissatisfied customers. Numerical features like **'Age'** and **'Flight Distance'** displayed distinct distributions, while delays were highly correlated and significantly influenced dissatisfaction. Categorical features such as **'Gender'**, **'Customer Type'**, and **'Class'** showed meaningful trends, with returning customers and Economy Class being predominant.. Correlation analysis and visualizations highlighted key patterns, with service quality and delays being major drivers of satisfaction. These insights set the stage for effective predictive modeling. In the countplot we observed that "Inflight entertainment" plays a pivotal role in customer satisfaction.

Satisfaction vs inflight entertainment
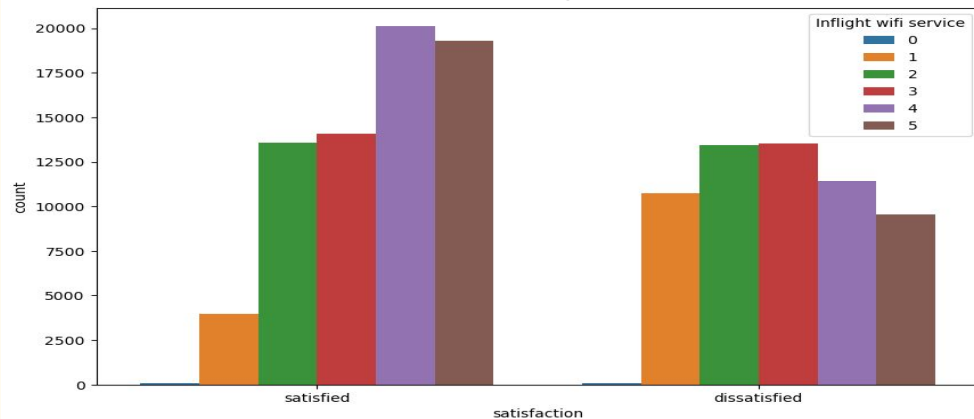
Distribution of Inflight entertainment Ratings
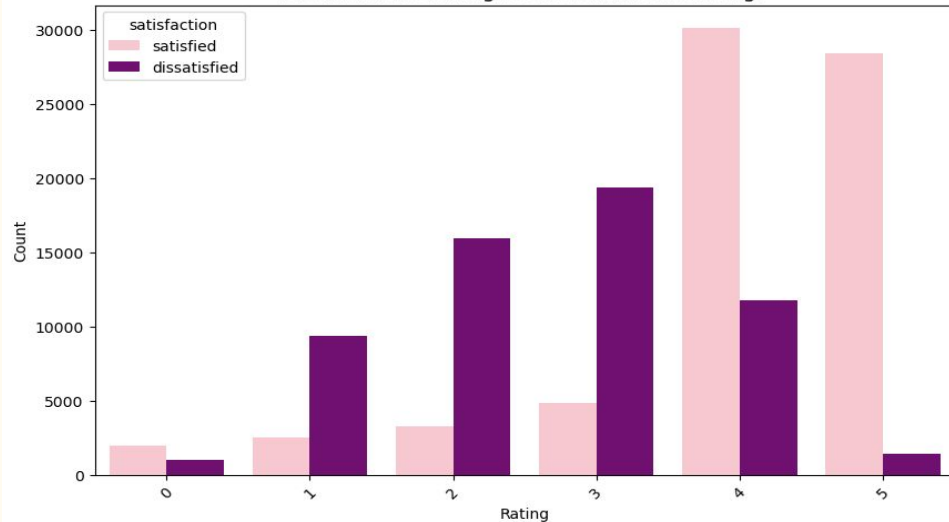
PREPROCESSING

These are essential steps taken to prepare raw data for analysis and modeling in our analysis they included;

Handled Missing Values:Imputed missing values in the **'Arrival Delay in Minutes'** column 393 entries.

Encoded Categorical Variables:

Label Encoding: Applied to **'Gender'**, **'Customer Type'**, and **'Type of Travel'** for binary categories.

One-Hot Encoding: Applied to **'Class'** to create new columns for each class Economy, Business, Economy Combine Low Ratings:Merged '0' and '1' ratings in certain features **'Seat Comfort'** to maintain a consistent 1-5 scale.

Scaled Numerical Features:Standardized **'Age'**, **'Flight Distance'**, and **'Departure Delay in Minutes'** using **StandardScaler** to normalize feature ranges.

Split Data:Divided the dataset into training 70% and testing 30% subsets for model development and evaluation.
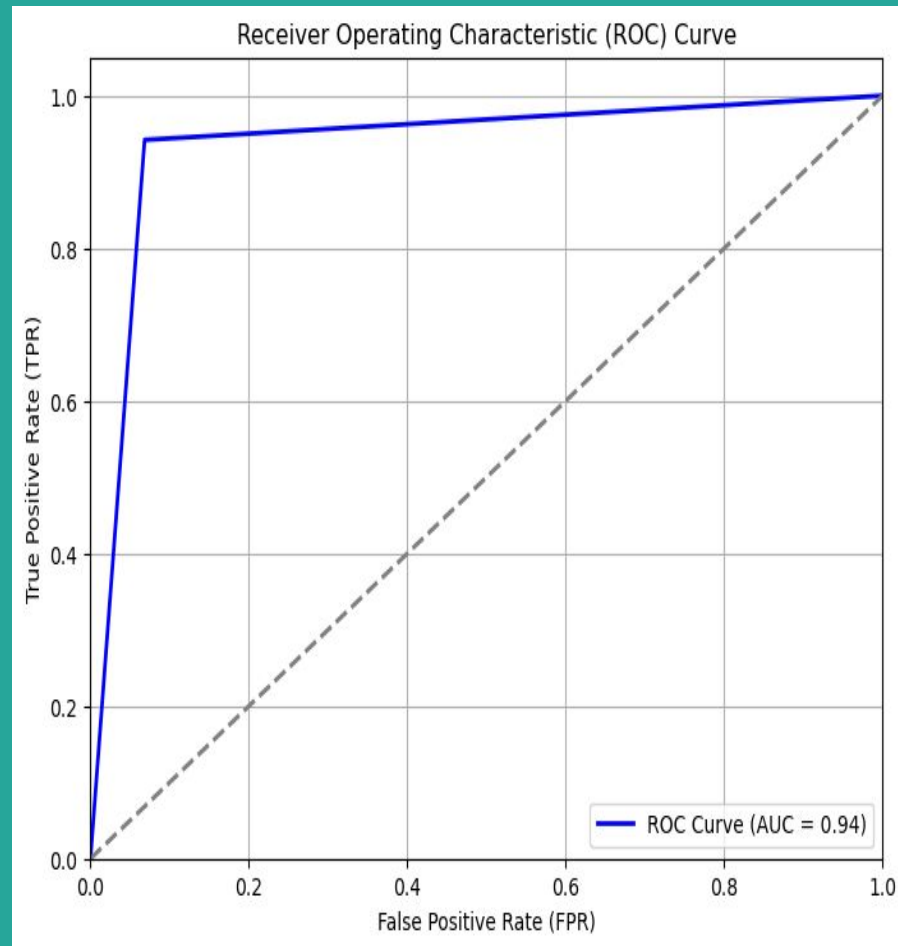
MODELING

In this analysis, three models were developed and evaluated: Logistic Regression, Decision Tree, and Random Forest. Each model was trained on a dataset aimed at predicting customer satisfaction and assessed using metrics like accuracy and ROC-AUC.

1. **Logistic Regression**:

Accuracy: 83.68%Logistic Regression provided a straightforward and interpretable model. However, its performance was limited due to its linear nature, making it less effective in capturing complex relationships in the data.

2. **Decision Tree**:

Accuracy: 93.84%;The Decision Tree outperformed Logistic Regression by handling non-linear data and identifying critical decision points. However, it showed a tendency to overfit, which may affect its generalization on unseen data.


Receiver Operating Characteristic (ROC) Curve

MODELING

3. **Random Forest**:

Accuracy: 95.92% Random Forest emerged as the best-performing model. By combining multiple decision trees, it reduced overfitting and improved accuracy and robustness. This ensemble approach excelled at capturing complex patterns in the dataset, making it the most reliable model for predicting customer satisfaction.

Overall, the **Random Forest model** is the most suitable for this task due to its superior accuracy and ability to generalize better than the other models. It provides actionable insights for stakeholders while ensuring reliable predictions.
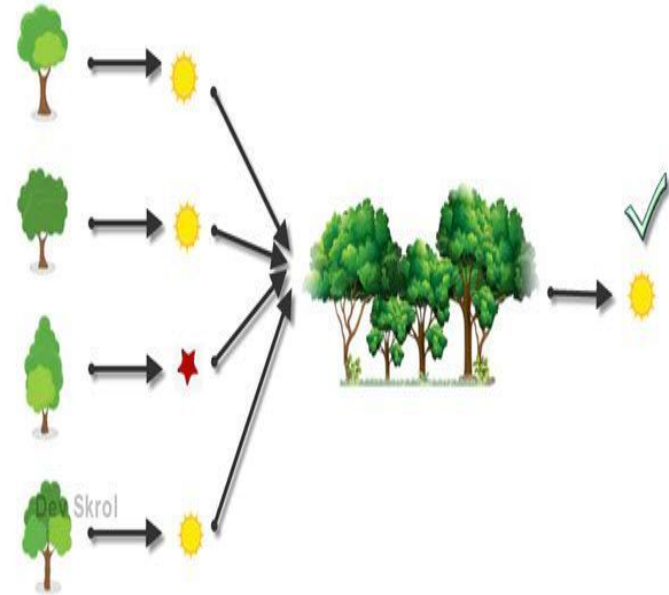
Recommendations

1. The most influential factors contributing to customer satisfaction are **Seat Comfort**, **Inflight Entertainment**, and **Food and Drink**. Prioritizing improvements in these areas can lead to a direct boost in customer satisfaction.
2. The **Operations Team** should focus on optimizing flight schedules to reduce delays and ensure smooth transitions from check-in to boarding.
3. The **Marketing Team** can use the insights to create campaigns targeting customers who are most likely to be dissatisfied, offering them
4. Regularly gather customer feedback through surveys, social media, and review platforms. Use this feedback to update models and continuously adapt service offerings.
5. Empowering customer service representatives to offer personalized assistance could increase satisfaction, particularly in situations involving complaints or long wait times.

Conclusion

In conclusion, this analysis equips Invistico Airlines with insights to prioritize areas for improvement, specifically enhancing Inflight Entertainment and seat comfort. By addressing these aspects, the airline can foster higher customer satisfaction, bolster customer loyalty, and elevate its overall service quality.