Airline Food Demand Prediction - Final Report

1. Problem Statement

Airlines face a critical challenge in minimizing food waste while maintaining passenger satisfaction. The goal of this project is to predict the total food demand for a flight based on characteristics such as flight duration, passenger count, class ratios, and flight type (International/Domestic). An accurate prediction model helps optimize catering logistics and reduce costs.

2. Dataset Description

A synthetic dataset of 5,500 flight records was generated to simulate real-world scenarios. Key features include:

- flight_duration (1-12 hours)

- passenger_count (50-300)

- business_class_ratio (0-1)

- is_international (Binary)

- total_food_demand (Target)

Constraints applied:

- International flights are >= 3 hours.

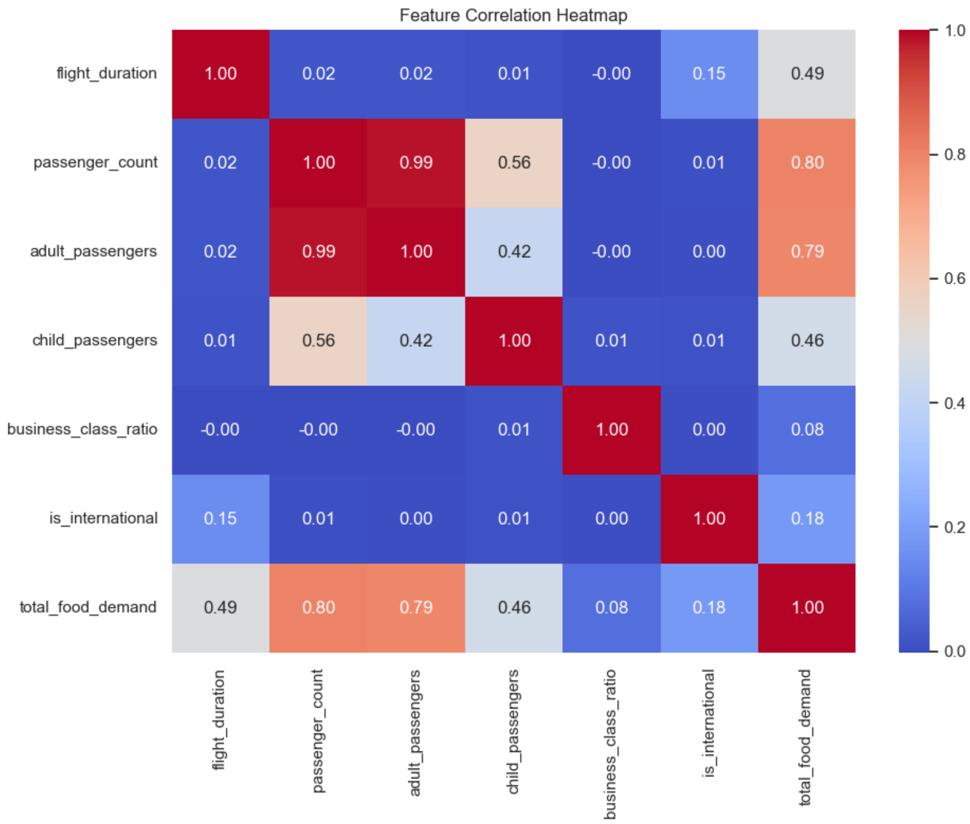
- Total food demand depends on passenger count, duration multiple, and class weights.

- Logical consistency (e.g., adults + children = total passengers).

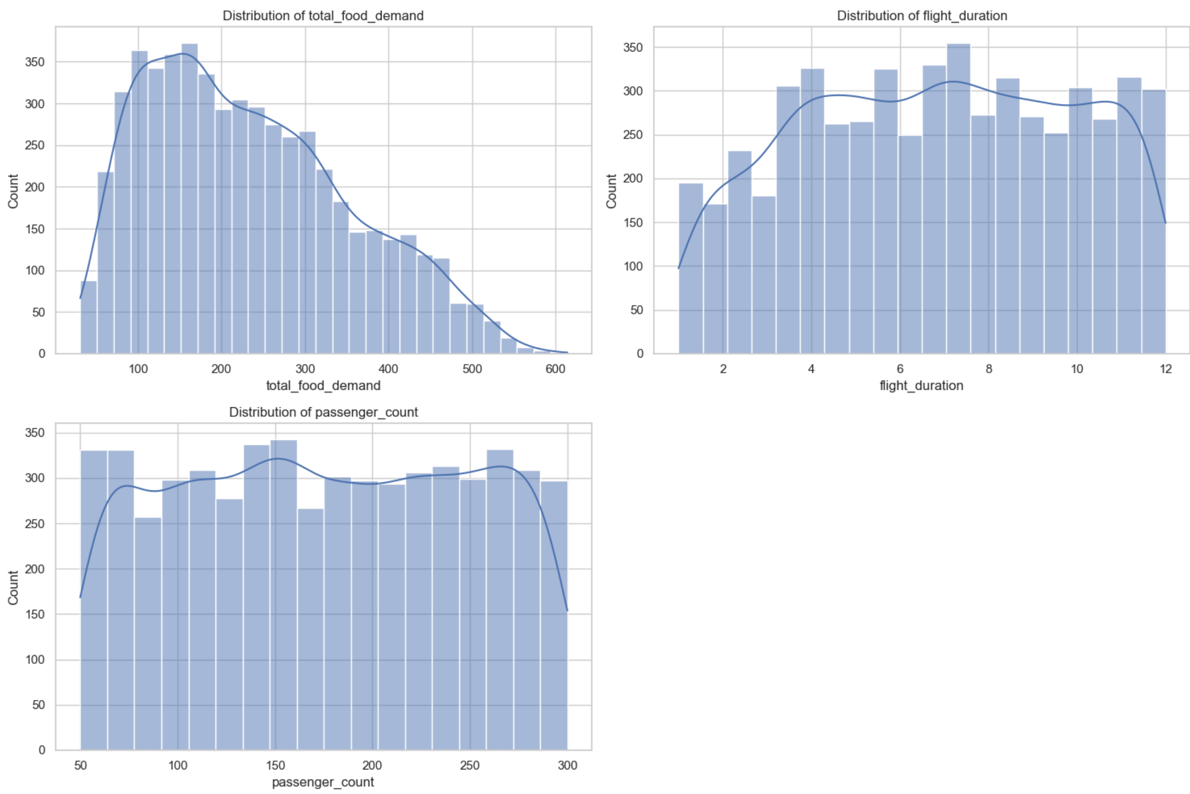
3. Exploratory Data Analysis (EDA)

We performed EDA to understand feature distributions and correlations. The heatmap below shows the correlation between features.

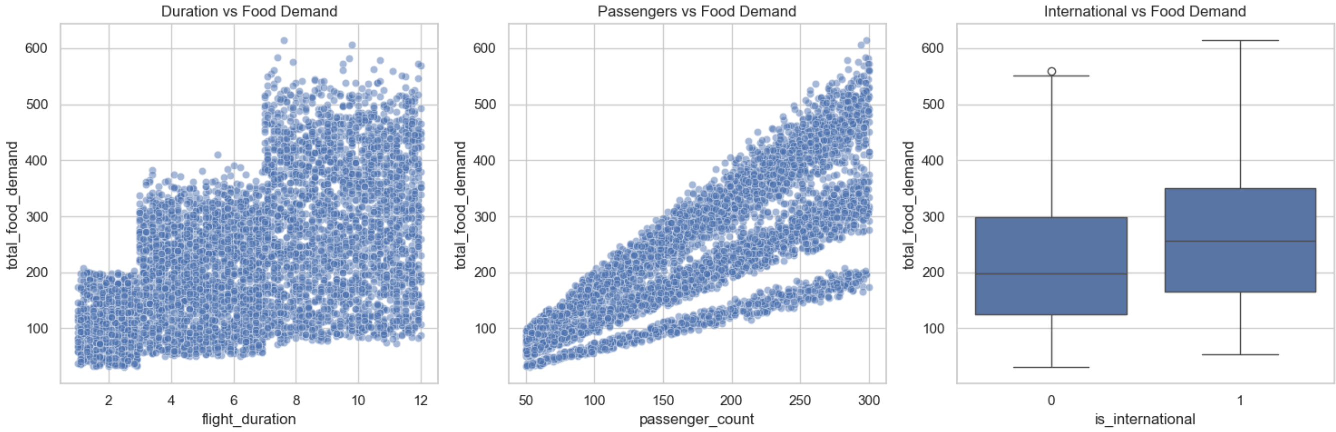
Airline Food Demand Prediction - Final Report



Distributions of key variables and their relationships with the target were also analyzed.



Airline Food Demand Prediction - Final Report



Airline Food Demand Prediction - Final Report

4. Methodology

We implemented three models to predict food demand:

1. Baseline (Mean Predictor): To establish a minimum performance benchmark.

2. Linear Regression: A simple interpretable model.

3. Random Forest Regressor: A powerful ensemble method capable of capturing non-linear relationships.

The data was split into 80% training and 20% testing sets.

5. Results

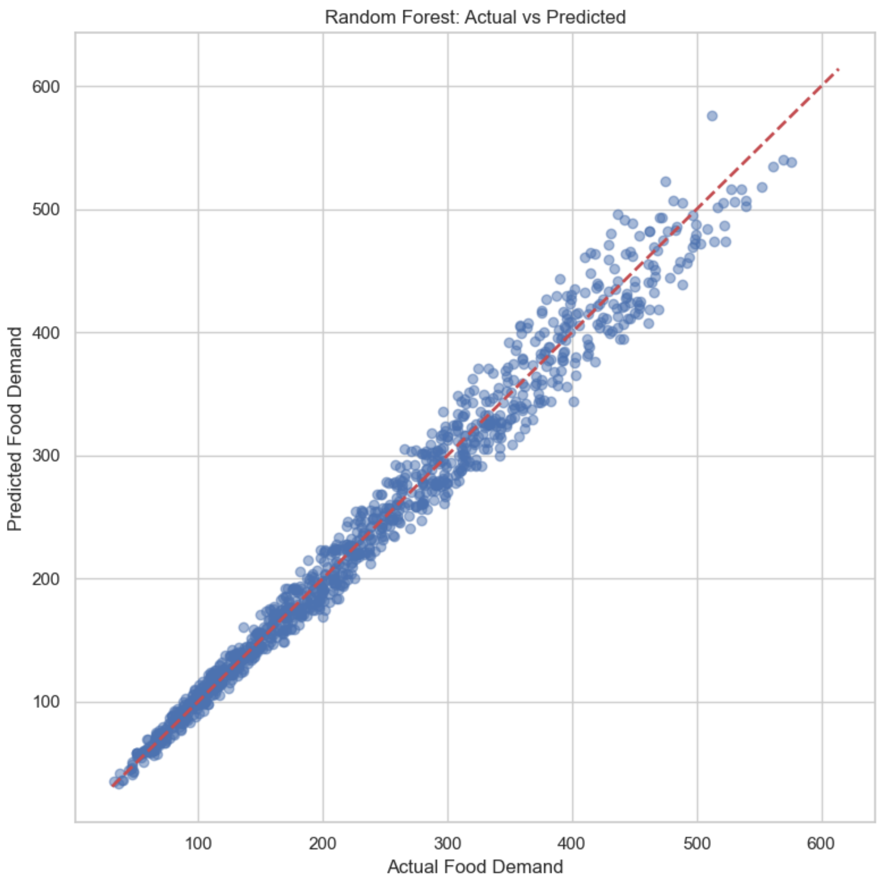
The models were evaluated using MAE, RMSE, and R-squared metrics.

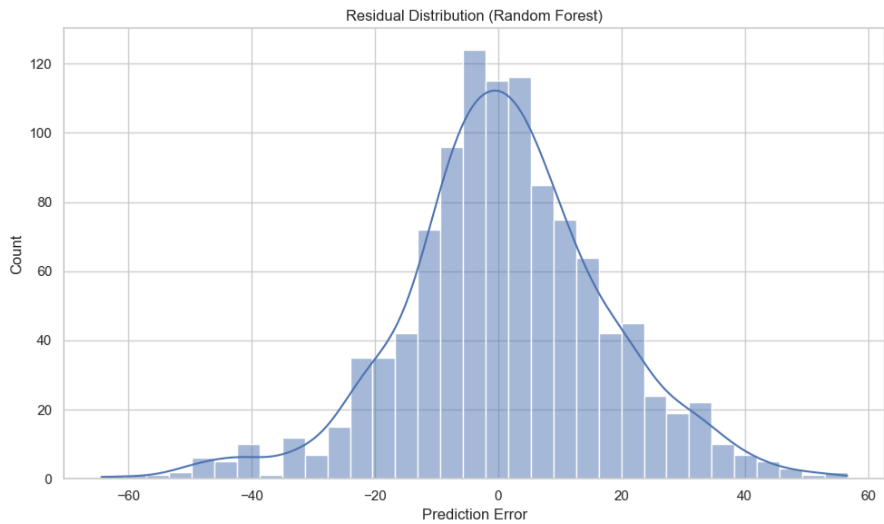
Model	R2	MAE	RMSE
Baseline	-0.0005	101.644	121.3061
Linear Regression	0.8949	31.1593	39.3218
Random Forest	0.981	12.6004	16.7216

The Random Forest model significantly outperformed the others, achieving the highest R2 score and lowest error metrics. This indicates strong non-linear patterns in the data that Linear Regression missed.

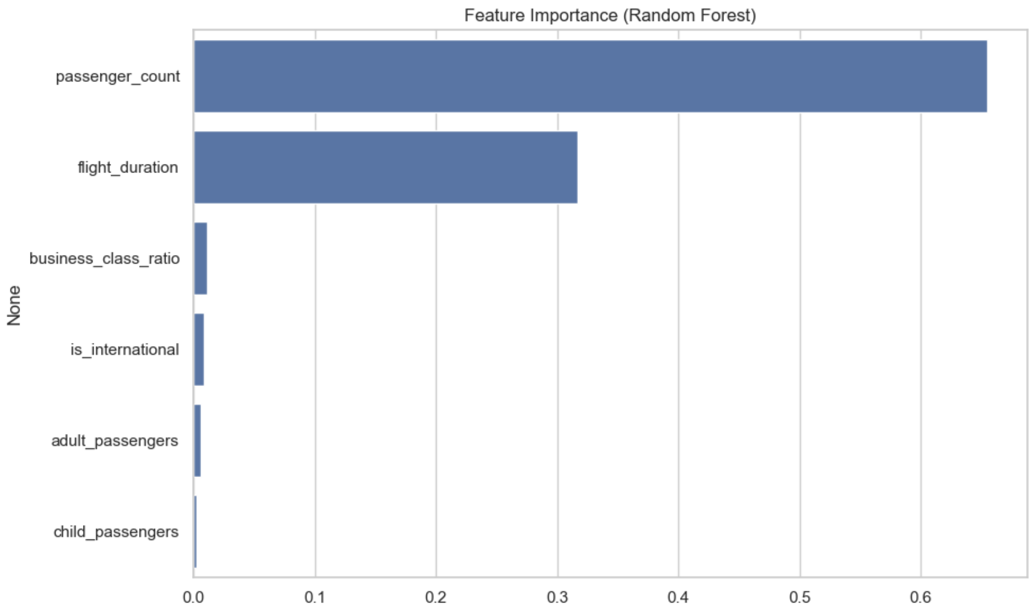
Model Analysis (Random Forest)

Airline Food Demand Prediction - Final Report





Airline Food Demand Prediction - Final Report



6. Conclusion

The project successfully developed a high-accuracy model for predicting airline food demand. The Random Forest model proved to be the best candidate. Feature importance analysis revealed that Passenger Count and Flight Duration are the most critical factors.

Implementing this model could save the airline significantly by reducing food wastage and avoiding under-catering penalties.