



# HOTEL RECOMMENDATION

Zhihan Wang, Junrui Ye



# PROJECT GOAL

- Analyze the impact of different factors in hotel booking
- Hotel Booking Prediction
- Gross Booking in USD (GBU) Prediction

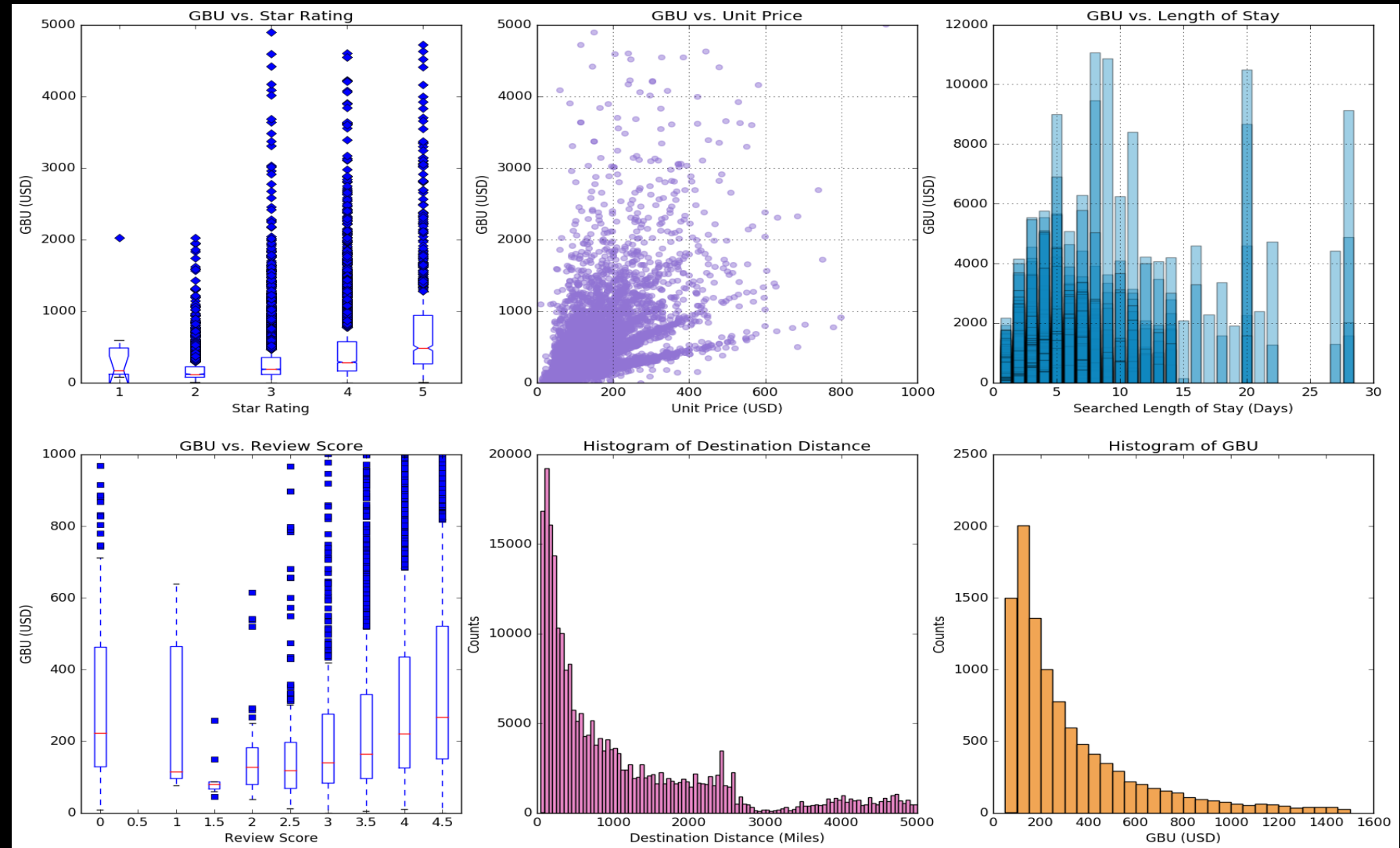


# DATASET

- Data Description
  - Hotel booking information from Expedia (Kaggle)
- Data Cleaning
  - Missing data
  - Corrupted data
- Key Features
  - Hotel features: Star rating, hotel unit price, review score
  - Customer features: Historical booking price of the customer, and search features like length of stay, number of rooms.

# DESCRIPTIVE STATISTICS

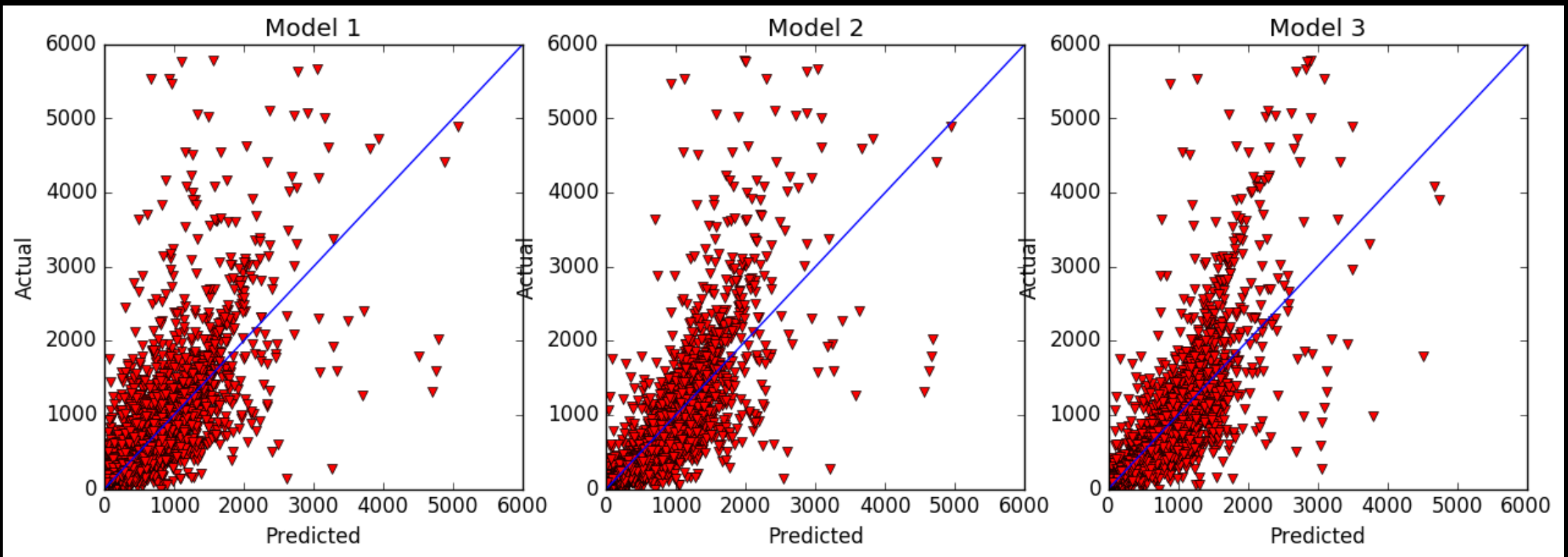
- Data Trend and Feature Correlation



# DATA ANALYSIS

- Feature Engineering:
  - Clustering: visitor star rating, visitor historical booking price, property star rating, property review score, property location score, price, booking window, position
- Model Picking
  - Logistic Regression
    - Create balanced input data set
    - Cross validation score: avg 0.58.

# DATA ANALYSIS: GBU PREDICTION








# NEXT STEP

- Increase prediction accuracy (for booking prediction and booking value prediction)
- Further Feature Selection/Engineering
  - Advanced Feature Analysis (further research)
  - Random Forest
  - Feature Normalization

The image features a solid black background. At the top, there is a decorative horizontal band with a wavy, fluid appearance. This band is composed of several overlapping layers of color, including bright yellow, orange, red, and a hint of cyan or light blue, creating a sense of motion and depth.

THANK YOU!