Airbnb — New User Bookings

DMML Project Proposal

Reza Almassi – Jad Mahmoud

August 23, 2025

Academic Year 2024-2025

Problem Description

Context:
Airbnb operates in 190+ countries and 34,000+ cities. Predicting where a new user will book first enables personalized content, reduces time to first booking, and improves demand forecasting.

Objective:
Predict the **top 5 travel destinations** (ranked by relevance) for each new user based on their profile and behavior.

Motivation:

- Improves user engagement and revenue.
- Realistic dataset with mixed features + session logs.
- Strong use case for full KDD pipeline.

Planned KDD Steps:

- EDA and cleaning (missing values, outliers)
- Feature engineering (session aggregates, time features)
- Models: Logistic Regression, LightGBM/XGBoost
- Validation: stratified split, NDCG@5

Dataset Description

Source: Kaggle — Airbnb New User Bookings
`https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings`

Main File: `train_users.csv` (≈200k rows)
Key Features (16):

- user_id, account creation date
- first active timestamp, date of first booking
- gender, age
- signup method & source
- language preference
- marketing attribution (channel, campaign)
- signup app, first device type, first browser

Target: `country_destination` (12 possible classes)

Evaluation & References

Evaluation Metric: NDCG@5

For each user, predict up to 5 countries (ranked).

Ground truth relevance = 1 for correct country, else 0.

$\text{DCG@5} = \sum_{i=1}^{5} \frac{2^{rel_i}-1}{\log_2(i+1)}$, IDCG=1 $\Rightarrow$ NDCG@5 $\in [0, 1]$.

Higher score if true destination appears earlier in top-5.

References

Wu, Y.; Zhou, Z. *New User Booking Prediction for Airbnb Historical Data*. UCSD CSE258 Project Report.
https://jmcauley.ucsd.edu/cse258/projects/fa15/045.pdf

Zhang, K. *The Prediction of Booking Destination On Airbnb Dataset*. UCSD CSE258 Project Report.
https://jmcauley.ucsd.edu/cse258/projects/fa15/038.pdf

Avireddy, S. *Predicting Airbnb user destination using user demographic and session data*. UCSD CSE258 Project Report.
https://jmcauley.ucsd.edu/cse258/projects/fa15/040.pdf

Ulfsson, H. *Predicting Airbnb user's desired travel destinations*. KTH Thesis.
https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1108334

Airbnb Tech. *Understanding User Booking Intent at Airbnb*.
https://airbnb.tech/wp-content/uploads/sites/19/2024/12/Understanding-User-Booking-Intent-at-Airbnb.pdf