Temporal Airbnb Seasonality and Modeling

Course: EAS 510 – Basics of AI
Assignment: Extra Credit – Temporal Airbnb Seasonality and Modeling

---

1. Introduction

This project analyzes temporal dynamics in Airbnb markets using InsideAirbnb calendar and listings data. The primary goals are to (1) construct a night-level panel dataset, (2) explore seasonal patterns in price and booking probability, and (3) evaluate predictive models using temporally valid train/validation/test splits. Two prediction targets are considered: nightly price (regression) and booking status (classification).

---

2. Datasets & Inputs

For each city and snapshot date, two files are used:

- **listings.csv** – static listing attributes (capacity, room type, review scores, etc.)
- **calendar.csv** – night-level availability and price information

The analysis notebook (`airbnb-temporal-modeling.ipynb`) loads each snapshot pair, verifies shapes and data types, and merges them into a unified panel dataset as required by the assignment rubric.

---

3. Part 1: Night-Level Panel Dataset Construction

Process

- All `listings.csv` and `calendar.csv` files are loaded successfully with evidence shown via shapes and sample rows.
- A **left join on `listing_id`** produces one row per listing per date.
- Key cleaning steps:

    - `price` converted from strings (e.g., "$120.00") to numeric values.
    - `is_booked` constructed from `available` (`f → 1`, `t → 0`).
    - `date` parsed into a proper datetime format.
- Time-based features created:
    - `month`, `day_of_week`, `week_of_year`, `day_of_year`
    - `is_weekend` indicator

This results in a reusable, reproducible panel dataset suitable for exploratory analysis and modeling.

---

4. Part 2: Seasonality Analysis

Exploratory analysis shows clear temporal patterns:

- **Prices** tend to increase during peak travel months and weekends.
- **Booking probability** exhibits milder seasonality, suggesting demand is influenced by factors beyond calendar effects alone.
- Weekend nights consistently show higher prices and higher booking rates compared to weekdays.

These findings motivate the inclusion of time-based features in predictive models.

---

5. Part 3: Temporal Train / Validation / Test Split

To avoid temporal leakage, a strictly time-ordered split is used:

- **Training:** January – September
- **Validation:** October – November
- **Test:** December – February

Feature matrices exclude identifiers and raw dates, using only listing attributes and engineered time features. Separate datasets are constructed for price regression and booking classification.

---

6. Part 4: Modeling and TensorBoard Analysis

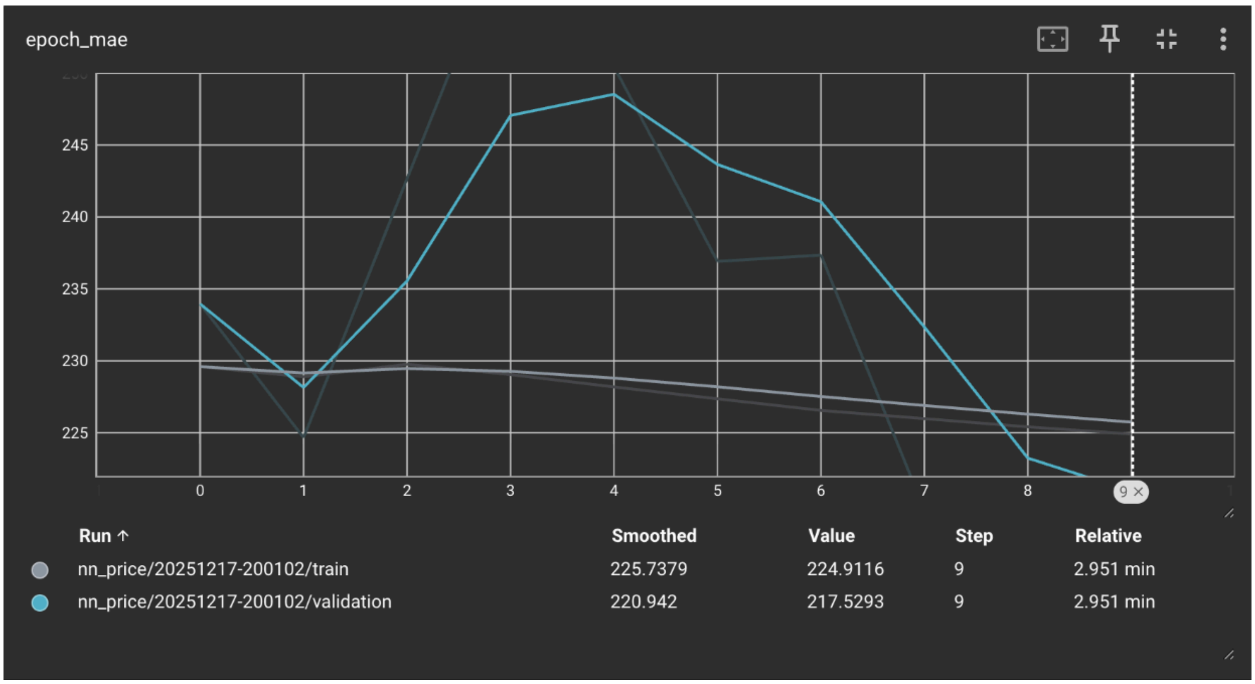
Two Neural Network models are trained:

- **Price Model:** Regression with MSE loss
- **Booking Model:** Classification with binary cross-entropy loss

TensorBoard is used to log training and validation metrics for both models, with separate log directories.

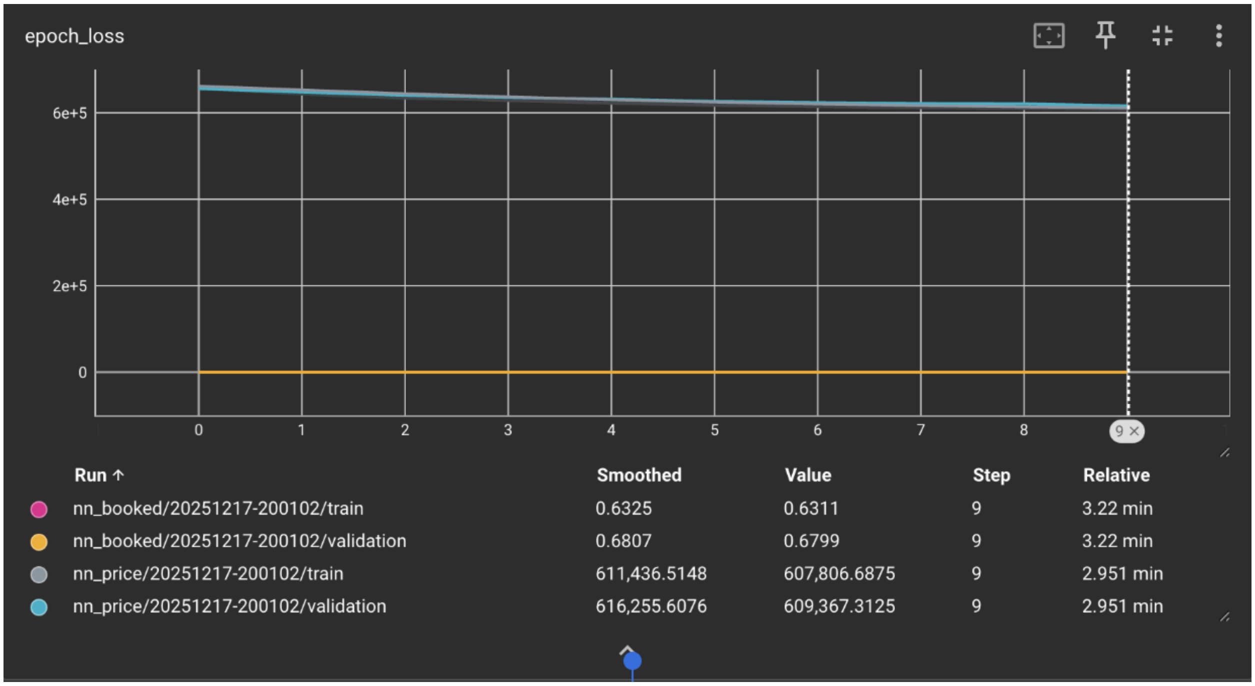
6.1 Price Model (Regression)

MAE over epochs:



The MAE curve shows an initial increase followed by a gradual decline, indicating the model is learning but with some instability early in training. Validation MAE ultimately decreases below training MAE, suggesting limited overfitting and reasonable generalization.

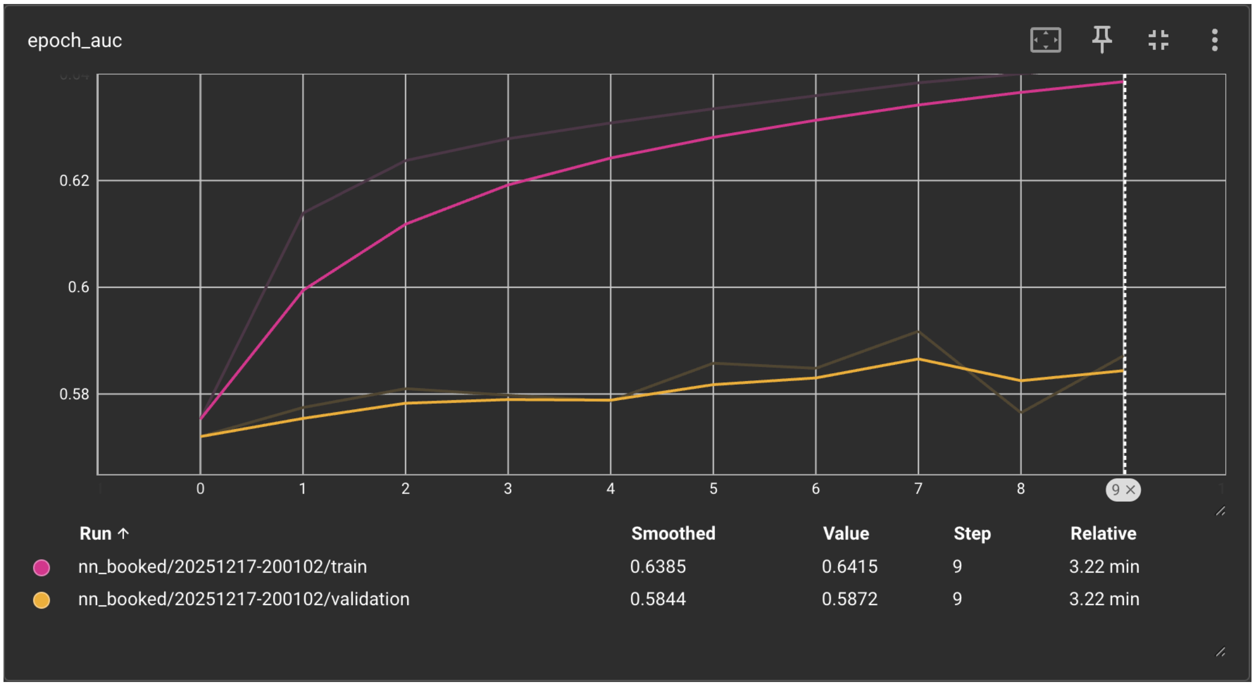
Loss over epochs:



Both training and validation loss steadily decrease and remain close, indicating stable optimization and no severe divergence between datasets.

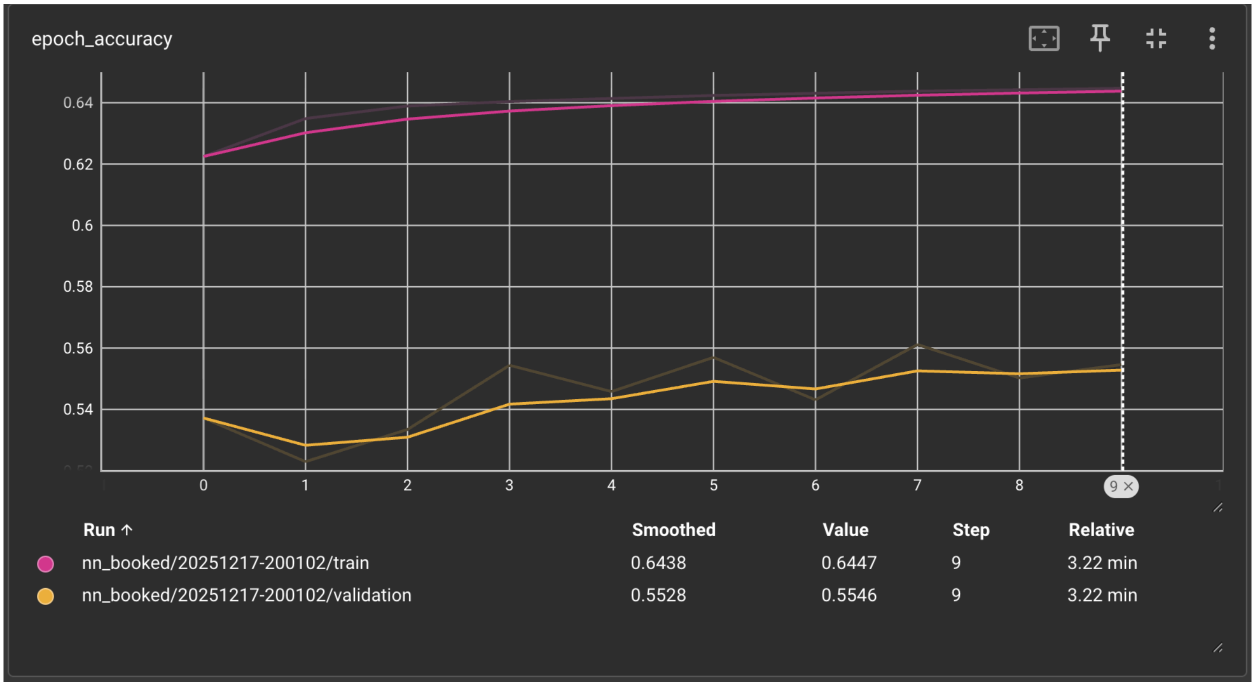
6.2 Booking Model (Classification)

AUC over epochs:



Training AUC improves steadily, while validation AUC increases more slowly and plateaus around ~0.58–0.59. This gap suggests mild overfitting and highlights that booking behavior is harder to predict than price.

Accuracy over epochs:



Training accuracy reaches approximately 64%, while validation accuracy stabilizes near 55%. This modest performance indicates limited signal in the available features for predicting bookings.

---

7. Interpretation and Comparison

Overall, **price prediction** shows stronger and more stable learning curves than **booking prediction**. Prices are more directly tied to seasonality and listing attributes, while bookings depend on additional unobserved factors (traveler intent, competition, events).

TensorBoard curves confirm:

- Stable training dynamics for the price model
- Mild overfitting and weaker generalization for the booking model

This aligns with final validation and test metrics and supports the conclusion that nightly price is the more predictable target.

---

8. Business Insights

From a business perspective:

- **Price forecasting** is more reliable and can directly support revenue management and dynamic pricing strategies.
- **Booking probability prediction**, while useful, likely requires richer behavioral or market-level features to achieve stronger performance.

---

9. Conclusion

This project demonstrates a complete temporal modeling workflow: dataset construction, seasonality analysis, temporally valid modeling, and TensorBoard-based interpretation. Results highlight the importance of respecting time structure in Airbnb data and show that prices exhibit clearer, more learnable temporal patterns than booking outcomes.
