STAT540 Final Project Instruction for Fall 2024

Important Due Dates:

Final Project Proposal Due Date: November 11, 2024 before class

Final Project Write-Up Due Date: December 11, 2024 before 5PM

Final project (30% of Grade; report 20%, presentation 10%): The format should be single-spaced, font size 12 and no more than 10 pages (figures, tables and references included). You must work on it **independently**. **Do not discuss** your analysis with your classmates. You can discuss with me if you need helps.

Please hand in a hard copy of your final project report (compiled pdf file from R markdown) before the due date and email your R code to Yen-Yi Ho (hoyen@stat.sc.edu).

Data Availability

Please find the city you picked https://people.stat.sc.edu/hoyen/STAT540/Exam/FinalProjectCity.pdf. The data can be download https://insideairbnb.com/get-the-data/. The data dictionary can be found https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHlNyGInUvHg2BoUGoNRIGa6Szc4/edit

Analysis of Airbnb Data

To get started, the websites listed in the Reference sections provide very helpful information. A search in Kaggle will also present many analyses using Airbnb data. The papers listed in (15, 16) in the reference section provides more ideas for airbnb data analysis. Pick some topics of analysis to form a meaningful conclusion about your city from your analysis.

- Explore: What is the distribution of prices across a city's neighborhoods? How does it change when you segment it further by room_type?
- Visualize: Create a map with a dot for each listing in a city and add a color scale based on price on the dots.
- Analyze: How do listings that require a minimum stay of a week or longer differ from those that don't?
- Identify factors that impact listing prices, such as property type, location, or amenities.
- Determine if there are significant differences in average ratings between different neighbourhoods or listing types.
- Analyse the impact of seasonal trends on occupancy rates and pricing.
- Anomaly Detection: Identify unusual patterns in data, such as fraudulent listings or reviews, to maintain platform integrity and trust.

Some potential ideas are listed below.

Some ideas for analysis

- Listing Characteristics. Room types by neighbor. What types of airbnb listing are popular in a neighborhood? Why? Are people in certain cities more inclined to share their room/house? Compare hotel room versus room in a home.

- Location Analysis. Where are popular locations for airbnb listings and why? What are the expensive airbnb neighborhoods? What are the most economical locations for travellers? Heatmaps: Create heatmaps to visualize the concentration of listings in different neighborhoods. Analyze whether there is a spatial correlation in listing prices, i.e., whether nearby listings tend to have similar prices. Insight: Identify hot spots for Airbnb rentals and areas where there is potential for growth. This can be useful for potential investors or for existing hosts looking to expand.

- Price Analysis. What are the factors have the most significant impact on airbnb listing price ? Average or median price by locations. What determine better than avarage price in a neigborhood? Number of reviews, Amenities, neighborhoods, local attractions, host characteristics, customer reviews, availability, occupancy, room types, listing characteristics, day of the week, holidays, ...etc. Do you observe price changes for a listing? Determine how sensitive bookings are to price changes.

- Avalability Analysis. Are there a seasonal trend of the number of listings? Plot the average price and availability of listings over time, highlighting any seasonal trends (e.g., higher prices during summer or holidays). Seasonal Decomposition: Use time series decomposition to break down the trend, seasonal, and residual components of your data. Insights: Determine the peak seasons for bookings and how prices adjust accordingly. This can help hosts optimize their pricing strategy throughout the year.

- Occupancy Analysis. Are there a trend (seasonal, location, price, ...) of occupancy? Average length of stay in your city. Visit frequency of the different locations. Booking by room types and property type.

- Amenity Analysis. What are some of the essential amenities in airbnb listing in your city? What are the amenities that could attract customer? For example, wifi, dedicated workspace, dryer/washer, kitchen, ...etc.

- Hosting Analysis. What are the average number of listings per host? Explore host characteristics in your city. Host with multiple listings in various neighborhoods? Trend in host earnings.

- Review Analysis. Trend in review. Does good review associated with higher listing price? What are the "words" that manners in booking? Examine the relationship between review scores and features like cleanliness, location, communication, and price. Text Analysis: Perform analysis on guest reviews to identify common themes in positive and negative reviews. Discover which aspects of the service (e.g., cleanliness or communication) are most strongly associated with higher ratings and customer satisfaction.

- Comparisons between various neighborhoods. Are there any discrepancies in listing characteristics, min/max stay, host communications, reviews,listing price, host characteristics, review, ...etc?

For example your analysis goal can be. . .

Finding best location to buy a new property for investment that host on Airbnb. Answering these questions for each borough

- Number of properties that host on Airbnb

- Average price per night

- Estimate Occupancy day per year

- Calculate annual income, annual expense and annual return, initial expense etc.

- Calculate return of investment rate and total profit

Reference

1. Analyzing Airbnb Data: project developed by Wondrous Raichu for INFO 2950: Introduction to Data Science at Cornell University https://pages.github.coecis.cornell.edu/info2950-s23/project-wondrous-raichu/eda.html

2. https://mohamedirfansh.github.io/Airbnb-Data-Science-Project/ This site has some really cool plots.

2. https://rpubs.com/shahworld/airbnb

3. https://rpubs.com/noearambula/finalproject

4. https://www.kaggle.com/code/djonafegnem/airbnb-data-analysis-in-r

5. https://www.kaggle.com/code/faressayah/analysis-of-airbnb-data-new-york-city

6. https://rpubs.com/siddharth2711/612101 (beautiful plots)

5. https://medium.com/airbnb-engineering/using-r-packages-and-education-to-scale-data-science-at-airbnb-906faa58e12d

6. https://pjournal.github.io/boun01g-r-sizlar/nyc_rsizlar.html

7. https://medium.com/@svinetas/exploratory-analysis-of-nyc-airbnb-data-via-r-244215bc90bc

8. https://www.linkedin.com/pulse/exploratory-data-analysis-airbnb-listings-using-r-murughan-mudaliyar

9. https://www.kaggle.com/code/jimmyoboni/analyzing-airbnb-pricing-in-new-york-city-r

10. https://www.kaggle.com/code/qusaybtoush1990/airbnb-analysis-dataset

11. https://www.kaggle.com/code/fehmifratpolat/istanbul-airbnb-data-analysis-and-visualization

12. https://www.datacamp.com/datalab/datasets/dataset-r-airbnb-listings

13. https://nycdatascience.com/blog/student-works/r-shiny/visualizing-nyc-airbnb-data/

14. https://github.com/XIAOL96/AirBnb-Data-Analysis-in-R/blob/master/text-mining.R (Word cloud example)

15. Airbnb research: an analysis in tourism and hospitality journals.

16. On the determinants of Airbnb location and its spatial distribution.