131 Project Memo

Cynthia Cao

2022-10-02

Final Project Memo

- The dataset is about a price prediction of Airbnb. Airbnb is an app used for booking houses, cabins, condos, or other unique stays for the trip. The dataset includes the prices and other house information of existed house owner.

- Link: https://www.kaggle.com/datasets/stevezhenghp/airbnb-price-prediction?resource=download

- It includes 29 predictors with 74112 observations.

- I will be working with both numeric and characteristic variables.

- No missing data.

- I will predicting the price of Airbnb per night. Question: How much will cost per night if I book on Airbnb.

- Response variable: Price/night (how much will cost of one night booking)

- I think it's better to use regression approach.

- Predictors: log_price, accommodates, bathrooms, host_response_rate, host_since, last_review, lattitude, longtitude, number_of_reviews, review_scores_rating, bedrooms, beds

- My goal will be combination of predictive. I want to predict the price with certain circumstances.

- Timeline: week 2-7: complete the first draft week 8: get advise from professor and TAs week 9-10: review and revise

Question: My dataset is based on both characteristic and numerical variables, if I choose to use the regression approach, can I also include the characteristic variables? or I can only include the variables with number?