MGS 616 – Spring 2024

Homework 1 – Due 2/18 @ 11:59pm

NOTE: The homework file should be submitted using UB Learns, with the assignment submission tool and make sure the file is named following the below protocol:

File naming protocol: MGS 616_HW1_GroupName

This is a **group assignment**. You should make sure that your name is on the homework document (or on the file name!). The submission should be in the form of a report that includes the analysis you create as well as any code or workflow you create. Don't forget to include any analysis considerations. DO NOT SUBMIT LINKS TO GOOGLE DOCS

Assignment:

The data set for this assignment is made available by Airbnb. It contains data about listings in the Boston, MA area. There are 3,583 listings in the data set. (The number of columns is reduced significantly from the original data set). Each row represents a single listing and contains information about the host of property and the property's characteristics. The goal is to visualize the input variables to get familiar with the data set. In a future assignment, you will be using this data set to predict the nightly price of a property.

Note:

- Description of the data set: The description of each variable is provided on the next page. For the purposes of this assignment, we will focus on the visualization of input variables.
- Please make sure the correct role and level of each variable is as shown on the next page.

Write-up:

To complete this assignment, draw upon what you have learned in the associated exercise and create a write-up. Your write-up should include the following:

1. Create a histogram for the PRICE variable:
   a) Provide the histogram of the variable.
   b) Discuss whether transformation is needed for the variable, and why.

2. Create a histogram for the BEDROOMS variable:
   a) Provide the histogram of the variable.
   b) Discuss whether transformation is needed for the variable, and why.

3. Create a boxplot for the BEDS variable:
   a) Provide the boxplot of the variable.
   b) Report the min whisker, max whisker, and mean of the variable.
   c) Discuss whether there are outliers and provide the value of the farthest outlier, if any

4. Create a boxplot for BEDS (as X) and PRICE (as Y):
   a) Provide the boxplot.
   b) Discuss two insights you gathered from the boxplot (your insights might concern a relationship, anomaly, skewness, outliers, etc.)
   (NOTE: How missing values are represented in the plot might affect how you interpret results.)

5. A scatterplot between PRICE (as Y) and REVIEW_ SCORES_RATING (as X):
   a) Provide the scatterplot
   b) Comment on the observed relationship (if there isn't a discernible relationship, explain why you can't identify a relationship from the chart.)

Description of the variables:

Variable:	Description:	Role	Level
id	ID of the Airbnb host	ID	Nominal
host_days	Number of days the host is a member (calculated until 1/1/2017)	Input	Interval
host_response_time	How quickly the host responds to inguiries. Categorical: a few days or more, within a day, within a few hours, within an hour, N/A	Input	Nominal
host_response_rate	Rate at which host responded to inquiries (percentage value)	Input	Interval
host_acceptance_rate	Rate at which host accepts reservation requests	Input	Interval
host_is_superhost	Is the host a superhost (1=Yes, 0=No)	Input	Binary
host_listings_count	The number of listings of the host	Input	Interval
host_identity_verified	Whether the host is verified or not (1=Yes, 0=No)	Input	Binary
neighbourhood_cleansed	Name of the neighborhood (25 categories)	Input	Nominal
city	Name of the city (34 categories including "missing")	Input	Nominal
property_type	Type of the property. 13 categories (Apartment, House, Boat, etc.)	Input	Nominal
room_type	Room type. 3 categories. Entire home/apt, Private room, Shared room	Input	Nominal
accommodates	Number of people that can be accommodated	Input	Interval
bathrooms	Number of bathrooms	Input	Interval
bedrooms	Number of bedrooms	Input	Interval
beds	Number of beds	Input	Interval
bed_type	Type of bed. 5 categories. Airbed, Couch, Futon, Pull-out Sofa, Real Bed	Input	Nominal
price	Price per night	Target*	Interval
security_deposit	Whether the host requires a security deposit (1=Yes, 0=No)	Input	Binary
cleaning_fee	Whether the host requires a cleaning fee (1=Yes, 0=No)	Input	Binary
guests_included	Number of guests included in price	Input	Interval
extra_people	Additional charge per person	Input	Interval
minimum_nights	Minimum number of nights for a reservation	Input	Interval
review_scores_rating	Overall rating of the property	Input	Interval
review_scores_accuracy	Rating for the accuracy of description	Input	Interval
review_scores_cleanliness	Rating for the cleanliness of the property	Input	Interval
review_scores_checkin	Rating for the check in experience	Input	Interval
review_scores_communication	Rating for the hosts communication with guests	Input	Interval
review_scores_location	Rating for the location of the property	Input	Interval
review_scores_value	Rating for the value of the property	Input	Interval
instant_bookable	Whether the property can be booked in an instance (1=Yes, 0=No)	Input	Binary
cancellation_policy	The cancellation policy of the host (flexible, moderate, strict, super strict_30)	Input	Nominal
reviews_per_month	Number of reviews per month for the property	Input	Interval