Task 6.1

Airbnb Amsterdam Data

- Source of Data: The data is available on the Kaggle Website and I have provided the link for the dataset https://www.kaggle.com/datasets/erikbruin/airbnb-amsterdam
- Data Collection: The files are downloaded from insideairbnb.com and give a snapshot of the Amsterdam situation on December 6th, 2018
- Data Content: The 'listings' file contains all the advertisements in Amsterdam on December 6th, 2018 (20k). The listings_details file contains additional variables. The calendar has 365 records for each listing. It specifies the whether the listing is available on a particular day (365 days ahead), and the price on that day. In addition, a reviews and reviews_detail file is available of people's reviews. And the neighborhoods list is given of the area.
- Data Profile: For the simplicity of the task I have merged both listings and listings_details to a dataframe listings 34 columns and 20030 rows

Index	Columns	Description	Time Variant/Invariant	Data Type
1	name	Name of customer	Invariant	object
2	host_id	ID of the host	Invariant	integer
3	host_name	Name of host	Invariant	object
4	neighbourhood	Neighborhood in Amsterdam	Invariant	object
5	latitude	Latitude	Invariant	float
6	longitude	Longitude	Invariant	float
7	Room_type	Type of rental	Invariant	object
8	price	Price of rental	Invariant	integer
9	minimum_nights	Minimum nights of stay	Time Variant	integer
10	number_of_reviews	How many customers review the property	Invariant	integer
11	last_review	Date of last review	Invariant	object
12	reviews_per_month	How many reviews per month	Invariant	float
13	calculated_host_listings_count	Number of hosts	Invariant	integer
14	Availability_365	If available all year long	Time Variant	integer
15	property_type	Type of property	Invariant	object

16	accommodates	If it accommodates long term	Invariant	integer
17	first_review	First review date	Invariant	object
18	review_scores_value	Score values of reviews	Invariant	float
19	review_scores_cleanliness	Cleanliness review scores	Invariant	float
20	review_scores_location	Location review scores	Invariant	float
21	review_scores_accuracy	Accuracy review scores	Invariant	float
22	review_scores_communication	Communication review scores	Invariant	float
23	review_scores_checkin	Check in review scores	Invariant	float
24	review_scores_rating	Rating review scores	Invariant	float
25	maximum_nights	Maximum nights of stay	Invariant	integer
26	listing_url	URL listing	Invariant	object
27	host_is_superhost	If host is a superhost	Invariant	object
28	host_about	The host bio	Time Variant	object
29	host_response_time	The response time of the host	Invariant	object
30	host_response_rate	The response rate of the host	Invariant	object
31	street	street of the property	Invariant	object
32	weekly_price	Weekly price of property	Invariant	object
33	monthly_price	Monthly price of the property	Invariant	object
34	market	On market or not	Invariant	object

- Limitations and ethics:
    1. Limitations: some of the data is still missing for further analysis, we can either impute or get rid of all together.
    2. Ethics: PLA security is requiring getting rid of the personal information, but up to the stakeholder to and the analysis conducted for the case study.
- Questions to explore:
    1. What are the average monthly prices for Amsterdam?

2. What are the most expensive neighborhoods in Amsterdam?
3. What are the average weekly prices for Amsterdam?
4. What is the average response rate of the hosts?