



Capstone Project

-Project Plan-

Mentor:

Maryam Najimigoshtasb

Member:

Pablo Martin Ruiz Diaz



Content

Dataset	2
Research Topic Title	2
Abstract	2
Data Processing Procedure	3
Analytical Objectives	3
Work Breakdown Structure	4
Gantt Graph	4



Dataset

Source: *Kaggle*

Title: *Inside AirBnB - USA*

Link: <https://www.kaggle.com/datasets/konradb/inside-airbnb-usa>

Research Topic Title

Analysis of the Airbnb datasets in the US to help the company to promote their listings and, for the customers, to find them more efficiently.

Abstract

I, as data analyst for Airbnb, am requested to give suggestions to the company about the market in the U.S. to make future marketing campaigns. On the other side, help customers to find the best place for their travels. To make this possible, the company shared diverse datasets for multiple popular cities, where each folder we may look at the following csv files:

- Calendar
- Listings
- Listings_detailed
- Neighbourhood
- Reviews
- Reviews_detailed

As a consequence of these, my plan is to focus on analyzing them in detail to better understand the U.S. market, such as word clouds across cities regarding customers' comments (for hosts) or description of the place (for customers), and show the fluctuation of incomes per city and get a forecast for a possible future.



To make all of these possible, I will create a database with all the raw data to figure them out. Then, I will create a database (OLAP) with all the data in simple tables with their corresponding relationships, and on them, applying data cleansing. Last but not least, create diverse visualizations using Tableau Desktop to get insights.

Data Processing Procedure

- MySQL, to create and store data.
- Python, to ETL processes between sources, data cleaning and machine learning (NLP, Regression techniques and so on).
- Tableau Desktop, for data visualization.

Analytical Objectives

- Word clouds of listings' names, descriptions and comments.
- Fluctuation of incomes across cities in 2023 and forecast for 2024.
- Categorization of room types from reviews by city and overall.
- Clustering by location versus prices to make future inversion for existing or new hosts.
- Tops cities by type of rooms, reviews, availability and price.
- Recommendation of type of room by location.
- Recommendation to hosts to raise the quality of their shelters (such as popular amenities).
- History of popularity of cities regarding number of comments.



Work Breakdown Structure

Week 1 - Create databases and ETL processes.

Week 2 and 3 - Data cleansing, descriptive and predictive analysis.

Week 4 and 5 - Data visualization and findings.

Week 6 - Final presentation.

Gantt Graph

