**1. Project Overview**

The project plan must include the following:

**· A Project Title.**

Analysing the Impact of Location, Amenities, and Host Behaviour on Airbnb Pricing and Occupancy Rates in London, England

**· A short summary of the project topic and background.**

The objective of this project is to study how location, amenities, and hosts affect Airbnb pricing and occupancy rate in London, England. Using data from Inside Airbnb, the study will examine the causes of variation in rental performance in one of the world’s most active short term rental markets.

**· A Research Question.**

How effectively can machine learning models recommend rental items to users based on their preferences and browsing history?

**· The Project Objectives.**

* To examine how different London neighborhoods affect Airbnb listing prices and occupancy levels.
* To determine which amenities correlate with higher pricing and increased occupancy rates.
* To investigate how host responsiveness and listing accuracy impact guest reviews and occupancy.

**· Reference List**

Dhillon, J., Eluri, N.P., Kaur, D., Chhipa, A., Gadupudi, A., Eravi, R.C. and Pirouz, M., 2021. Analysis of airbnb prices using machine learning techniques. In: *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*. [online] IEEE. pp.0297–0303. Available at: <<https://ieeexplore.ieee.org/abstract/document/9376144/>> [Accessed 10 February 2025].

Rezazadeh Kalehbasti, P., Nikolenko, L. and Rezaei, H., 2021. Airbnb Price Prediction Using Machine Learning and Sentiment Analysis. In: A. Holzinger, P. Kieseberg, A.M. Tjoa and E. Weippl, eds. *Machine Learning and Knowledge Extraction*, Lecture Notes in Computer Science. [online] Cham: Springer International Publishing. pp.173–184. <https://doi.org/10.1007/978-3-030-84060-0_11>.

Zhu, A., Li, R. and Xie, Z., 2020. Machine learning prediction of new york airbnb prices. In: *2020 Third International Conference on Artificial Intelligence for Industries (AI4I)*. [online] IEEE. pp.1–5. Available at: <<https://ieeexplore.ieee.org/abstract/document/9253078/>> [Accessed 10 February 2025].

**2. Project Plan: Task List and/or Project Timeline**

| **Task No.** | **Task Description** | **Start Date** | **End Date** | **Notes** |
| --- | --- | --- | --- | --- |
| 1 | Data Collection: Download and organize Airbnb data from Inside Airbnb. | Feb 10, 2025 | Feb 15, 2025 | Ensure data is current and relevant to the study period. |
| 2 | Data Cleaning: Process and clean the data for analysis. | Feb 16, 2025 | Feb 28, 2025 | Handle missing values, outliers, and ensure data consistency. |
| 3 | Literature Review: Research existing studies on factors influencing Airbnb metrics. | Feb 10, 2025 | Mar 8, 2025 | Conducted concurrently with data-related tasks. |
| 4 | Exploratory Data Analysis (EDA): Perform initial analysis to identify trends. | Mar 1, 2025 | Mar 12, 2025 | Use visualizations to understand data distributions and relationships. |
| 5 | Model Development: Develop statistical models to assess the impact of variables. | Mar 13, 2025 | Apr 10, 2025 | Focus on regression models to quantify effects. |
| 6 | Results Interpretation: Analyze model outputs to draw conclusions. | Apr 11, 2025 | Apr 20, 2025 | Relate findings to project objectives. |
| 7 | Report Writing: Compile findings into a comprehensive report. | Apr 21, 2025 | May 2, 2025 | Include methodology, results, discussions, and conclusions. |
| 8 | Final Review and Submission: Review the report and submit. | May 3, 2025 | May 5, 2025 | Ensure all components are complete and polished. |

**3. Data Management Plan**

**Overview of the Dataset**

Inside Airbnb is a platform that gathers publicly available information about Airbnb listings, thus forming the data set. Details like listing locations, prices, available amenities, information about the host and occupancy rates are included in the data. This information is gathered in order to produce insights into how Airbnb is affecting housing markets, as well as communities.

**Data Collection**

The source of data will be from Inside Airbnb’s website, where the very same data set for London is available directly on their platform.

**Metadata**

The dataset is typically provided in CSV format, with the following specifications:

* **File Format**: CSV
* **Number of Records**: Approximately 90,000 listings
* **Data Size**: Approximately 50 MB

**Document Control**

The project’s code and documentation will be managed in a GitHub repository. The weekly changes and progress will be committed into the repository.

**ReadMe File**

The ReadMe file will include:

* Project overview
* Instructions for reproducing analyses
* Dependencies and system requirements
* Contact information for further inquiries

**Security and Storage**

The data in the GitHub repository along with the corresponding code will be backed up by the secure cloud storage service, such as One Drive, every week. Sensitive data will have restricted access only to authorised personnel and will be shared via a secure channel for confidentiality purposes.

Ethical requirements: You must address each of the following issues and state how your specific dataset meets these requirements, give evidence when possible (e.g. screenshots or references):

1. Does the data come under GDPR requirements?

The information included in the dataset is publicly available and does not contain personal data covered by GDPR regulations.

2. Does the project conform to UH ethical policies?

The ethical guidelines of not harming individuals and communities are aligned with the project.

3. Do you have permission to use the data for your proposed research project?

The data is public, and research on it is permissible.

4. Are you assured that the data was collected ethical (i.e. by the original people who gathered/collected/ collated/made the data)?

Inside Airbnb aggregates data publicly available in such a way that the data collection is following ethical collection principles.