Project Plan

Analysing Airbnb Data in Sydney

Georgia Platt

Jackson Scown

Gia Huy Lieu

Table of Contents

[1.0 Introduction 3](#_Toc144471257)

[1.1 Background 3](#_Toc144471258)

[1.2 Scope 3](#_Toc144471259)

[1.3 Document contents 3](#_Toc144471260)

[2.0 Work Breakdown Structure 5](#_Toc144471261)

[3.0 Activity Definition & Estimation 6](#_Toc144471262)

[4.0 Gantt Chart 11](#_Toc144471263)

# Introduction

## Background

*The Sydney AirBnB Data Analysis project aims to analyse a detailed dataset containing information regarding AirBnB listings in Sydney Australia. The project involves various tasks, including retrieving and reporting listing information for accommodation in Manly between 2012 and 2023, creating a price distribution chart, conducting keyword-based searches, analysing customer comments related to cleanliness, and listing the suburbs in order of average price and again with average rating. The goal is to provide valuable insights into the AirBnB market in Sydney and enhance decision-making for hosts, guests, and stakeholders.*

## Scope

1. Data Collection and Preparation:

Obtain the Sydney AirBnB dataset, clean and preprocess the data for analysis. This involves handling missing values, data validation, and data formatting. Need to clean “name” column of dataset.

1. Data Retrieval and Reporting:

Develop a system to retrieve and display information about Manly listings between 2012 and 2015. This includes listing details such as property name, room type, price, number of reviews, etc.

1. Price Distribution Chart:

Generate a visual representation (chart) of the distribution of property prices between 2012 and 2023. This chart will help understand the range and spread of prices.

1. Keyword-based Search:

Implement a search functionality to retrieve records containing specific user-entered keywords (e.g., "pool," "pet"). This feature enables users to find listings that match their preferences.

1. Cleanliness Analysis:

Analyse customer reviews to identify factors related to cleanliness. Develop a method to extract relevant reviews and quantify the frequency of cleanliness-related keywords.

1. Time Series Analysis:

Perform an innovative analysis, which explores trends in booking patterns over time. Analyse with the assumption that the last review was completed the day the booking came to an end.

## Document contents

1. Introduction:

Problem Statement:

Describe the objective of the project, which is to analyse a comprehensive dataset of AirBnB listings in Sydney to provide valuable insights for hosts, guests, and stakeholders.

Project Scope:

Define the scope of the project, including tasks such as data retrieval for Manly listings, price distribution chart creation, keyword-based searches, cleanliness analysis, and time series analysis.

Document Overview:

Provide an outline of the sections covered in the document.

1. Work Breakdown Structure (WBS):

WBS Overview:

Present a hierarchical breakdown of the project tasks, starting from high-level components (e.g., Data Collection) and drilling down into sub-tasks (e.g., Data Cleaning, Data Preprocessing).

WBS Diagram:

Include a visual representation of the WBS with task relationships and dependencies.

1. Activity Definition and Estimation:

Task Definitions:

Detail the tasks identified in the WBS, including a brief description of each task's purpose.

Task Duration Estimation:

Provide estimated durations for each task based on team expertise, historical data, and potential complexities.

1. Gantt Chart:

Project Timeline:

Present the timeline of the project with tasks plotted over time.

Milestones:

Highlight key milestones, such as Data Collection Completion, Analysis Phases, and final reporting.

Dependencies:

Depict task dependencies to show the order in which tasks need to be executed.

# Work Breakdown Structure

# Updated Activity Definition & Estimation

|  |  |
| --- | --- |
| **1.0** | **INITIATION (11 Days)** |
| **1.1** | **Define Project Scope and Objectives (3 Days)** |
| Outline the scope, objective and goals of the project through identifying the boundaries and constraints of the whole project. | |
| **1.2** | **Assign roles to team members (2 Day)** |
| Establish the roles and responsibilities of each team member and establish forms of communication for the project. | |
| **1.3** | **Develop project management plan (2 Days)** |
| Created a detailed project plan that will outline the tasks and timelines of the whole project. Create a WBS for all phases of the project. | |
| **1.4** | **Risk management (Ongoing)** |
| Identify potential risks and uncertainties and develop strategies to mitigate the identified risks. Duration of this will continue throughout the whole project. | |
| **1.5** | **Outline User Requirements (3 Days)** |
| Identify the requirements of the user outlining a clear list of features that the whole project will need. | |
| **1.6** | **Outline Software Requirements (3 Days)** |
| Identify the requirements of the software outlining a clear list of features that the whole project will need. | |
| **2.0** | **DATA COLLECTION (7 Days)** |
| **2.1** | **Obtain Sydney Airbnb dataset (1 Days)** |
| Download the necessary datasets from Kaggle and organise the required data into git | |
|  |  |
| **2.2** | **Risk management plan (2 Day)** |
| Develop a plan to help identify, manage and mitigate potential risks. | |
| **2.3** | **Project cost estimate (1 Day)** |
| Estimate the overall estimate of the project cost factoring things like costs of required software and resource costs. | |
| **2.4** | **Project timeline (3 Days)** |
| Plan the project timeline with the key milestones and deadlines, created a roadmap through the use of a Gantt chart. | |
| **2.5** | **Clean and pre-process the data (3 Days)** |
| **2.5.1** | **Validate data (2 Days)** |
| Examine data and determine overall accuracy and consistency, identifying any anomalies or inconsistencies. | |
| **2.5.2** | **Handle missing values (1 Days)** |
| Create handling methods to deal with missing data points applying techniques to fill in missing values. | |
| **2.5.3** | **Format data for analysis (1 Days)** |
| Organise data into a format that will allow further analysis, reducing the amount of excel documents required further creating suitable format for modelling. | |
| **3.0** | **CREATION (30 Days)** |
| **3.1** | **Software Design flowchart/plan (3 Days)** |
| Create flowcharts and diagrams outlining the architecture of the software. Define the structure and its components. | |
| **3.2** | **Identify System components/functions (3 Days)** |
| List and define the individual components and functions of the system, identifying the key functionalities. | |
| **3.3** | **Suburb Specific Listing information (3 Days)** |
| **3.3.1** | **Design Search Functionality (2 Days)** |
| Design the logic for the search functionality an define the process for user interaction. | |
| **3.3.2** | **Implement/Test algorithm (2 Days)** |
| Implement the search algorithm based on the design testing that the algorithm is accurate and efficient. | |
| **3.4** | **Price Distribution task (3 Days)** |
| **3.4.1** | **Extract price data for chart (2 Days)** |
| Gather and organise price data from the excel data further preparing the data for chart creation. | |
| **3.4.2** | **Design and generate chart (2 Days)** |
| Create a chart that visually represent the price distribution formatting the chart and generating its content. | |
| **3.5** | **Keyword-based Search (8 Days)** |
| **3.5.1** | **Design search functionality (2 Days)** |
| Design the logic for the search functionality an define the process for user interaction. | |
| **3.5.2** | **Implement algorithm (2 Days)** |
| Develop and code the search algorithm based on a keyword-based search function implementing functional code that utilises the keyword search. | |
| **3.5.3** | **Test and validate algorithm (4 Days)** |
| Ensure the algorithm is completely working for all potential search functions | |
| **3.6** | **Cleanliness Analysis (8 Days)** |
| **3.6.1** | **Extract Customer Reviews Data (1 Days)** |
| Collect and organise customer review data and prepare the dataset | |
| **3.6.2** | **Identify cleanliness related words (2 Days)** |
| Identify key words related to cleanliness from the reviews analysing the text and extracting relevant words. | |
| **~~3.6.3~~** | **~~Quantify frequency of keywords (3 Days)~~** |
| ~~Calculate the frequency of the identified keywords and analyse the occurrence of these words in the reviews.~~ | |
| **3.7** | **Suburb Cost/Rating Analysis (2 Days)** |
| **3.7.1** | **Extract review price/rating (1 Days)** |
| Extract and collect review data related to prices and ratings identifying the relevant attributes for analysis. | |
| **3.7.2** | **Test and validate algorithm (2 Days)** |
| Develop and implement algorithms to process the price and rating data, validating the accuracy and performance | |
| **3.7.3** | **Display data visually (1 Days)** |
| Create visualisations of the data exploring relationships between price and rating for the data within different suburbs. | |
| **3.8** | **User interface design (15 Days)** |
| **3.8.1** | **Structural Design (8 Days)** |
| Plan and create the structural design of the interface defining layout and overall navigational flow. | |
| **3.8.2** | **Visual Design (8 Days)** |
| Design the visual elements of the interface ensure consistent colour schemes, typography and graphical elements ensuring a visually appealing and user-friendly experience. | |
| **4.0** | **VALIDATION (4 Days)** |
| **4.1** | **Analyse insights (2 Days)** |
| Analyse the insights and findings from the different phases identifying trends patterns and actionable insights. | |
| **4.2** | **Test full program functionality (3 Days)** |
| **4.2.1** | **Identify & address bugs and issues (2 Days)** |
| Test the entire program functionality to ensure all requirements are met. Identify and address any bugs, glitches or issues. | |
| **4.2.2** | **Create use cases for software (2 Days)** |
| Develop a detailed use cases and scenarios for software testing and ensure coverage of all varying user interactions. | |
| **5.0** | **CLOSURE (4 Days)** |
| **5.1** | **Review project against initial objectives (1 Days)** |
| Evaluate project against the initial goals and objectives verifying that everything created meets the initial requirements | |
| **5.2** | **Handover Final Software Version (1 Days)** |
| Prepare final version of the software with all relevant information and documents for the final handover | |
| **5.3** | **Software Testing Report (2 Days)** |
| Create a comprehensive report outlining the software testing phases, documenting all processes, result and remaining issues. | |

Changed the times to reflect how long it actually took when doing the project, whilst the order that the project slightly differed there were no additional steps nor were there any changes to the original design we had gone with. Didn’t quantify the Frequency of keywords in 3.6.3.

# Pre-Project Gantt Chart



# Post-Project Gantt Chart



When reviewing the initial plan to the final result many differences can be observed. Some of these differences include that getting the ui up and running took nowhere near as long as we initially thought, the issue was more finding the time to actually get in and create the interface and design the code. Furthermore, there were a number of processes that were not quite in the order that we ended up executing them in, for example the creation of the visual interface was done first and the code was then built to make that interface work, rather than building code and placing an interface on front of it. The order of creation for the 4 functionalities were also done in reverse with the price distribution, suburb cost/rating and cleanliness analysis all completed before the main page with the keyword-based search and filtering.