Software Design Document

Sydney Airbnb Open Data

Adithya Srinivas Bellamkonda (s5295811)

**Melvin Chavara (s5295081)**

Table of Contents

[1.0 System Vision 3](#_Toc46748622)

[1.1 Problem Background 3](#_Toc46748623)

[1.2 System Overview 3](#_Toc46748624)

[1.3 Potential Benefits 3](#_Toc46748625)

[2.0 Requirements 4](#_Toc46748626)

[2.1 User Requirements 4](#_Toc46748627)

[2.2 Software Requirements 4](#_Toc46748628)

[2.3 Use Cases 4](#_Toc46748629)

[3.0 System Components and Software Design 5](#_Toc46748630)

[3.1 System Components 5](#_Toc46748631)

[3.2 Software Design 5](#_Toc46748632)

[4.0 User Interface Design 6](#_Toc46748633)

# System Vision

## Problem Background

The Sydney Airbnb dataset is a sizable collection of homestays listing activity in Sydney, New South Wales, Australia, and it is a helpful resource for understanding the dynamic short-term rental market in this thriving city. This dataset, which was produced as a part of the InsideAirbnb initiative, is updated each month to make sure it remains a current and valuable source of information for researchers and data analysts. It offers an exhaustive overview of the city's Airbnb ecosystem and covers a wide range of data categories.

These categories include listings, which provide thorough descriptions of properties and average review scores; calendar data, which offers pricing and availability information on a daily basis; reviews, which provide insightful reviews and comments from visitors; summary information on listings; and neighbourhood data, which enables geospatial filtering. This dataset offers an exceptional opportunity to investigate issues related to the use of Airbnb, its effects on various Sydney neighbourhoods, property price trends, guest satisfaction, and numerous other aspects of short-term lodging in one of Australia's most populated cities.

## System Overview

Users looking to better understand Sydney, Australia's Airbnb market can use the software we are creating for this project as a powerful tool. The data extraction from the Sydney Airbnb dataset can be analysed and visualised more easily thanks to its simple graphical user interface. Our software's primary function is to give users the opportunity to investigate significant suburbs of the Airbnb market, with an emphasis on occupancy rates in various Sydney suburbs. Users can choose specific time periods and suburbs of interest using this core functionality, which offers insights into the usage trends of Airbnb listings over time. Users can maximize their travel experiences by booking accommodations when and where they make informed decisions based on a visual representation of occupancy rates.

The software is designed in a user-friendly, modular manner with separate components for each analytical tool. It also emphasizes interaction, allowing users to customize their queries and engage with the data. The architecture of the software enables simple expansion and adaptation to include additional analysis tools or accommodate future updates to the Airbnb dataset. With a focus on usability and functionality, our software aims to provide users with insightful information about the Sydney Airbnb market, assisting them in navigating the wide range of listings and neighbourhoods and helping them make decisions when booking their stays in this dynamic Australian city.

## Potential Benefits

The software we are creating has a great deal of potential to benefit many different stakeholders and users. First of all, it gives tourists and potential Airbnb guests vital information about the Sydney homestay industry, allowing them to worry less about their stay in Sydney. Users can book their stays more intelligently and affordably by visualising occupancy rates, pricing trends, and guest feedback, resulting in a more pleasurable and hassle-free journey. By enabling users to filter listings based on particular amenities or cleanliness standards and book accommodations in popular suburbs during prime times, this information can help users' overall satisfaction during their stay in Sydney. Additionally, property owners and Airbnb hosts can benefit greatly from using our software. With the aid of this tool, hosts can gain an easier understanding by comprehending market trends, perfecting pricing plans, and spotting potential improvement areas based on feedback from visitors. As a result, hosts may see an increase in bookings and income. Additionally, researchers and analysts can use our software to conduct in-depth analyses of the Sydney Airbnb market. By doing so, they can learn important details about the city's tourism trends, the effects of short-term rentals on specific neighbourhoods, and much more. Our software could ultimately help Sydney's Airbnb ecosystem become more knowledgeable, effective, and data-driven while also serving as a model for similar programs in other areas.

# Requirements

## User Requirements

Firstly, the software certainly will provide a useful tool for travellers and Airbnb guests to improve their stay in Sydney. These users can quickly enter their travel dates and choose a particular Sydney suburb to access vital data on occupancy rates. They can then decide when to make hotel reservations with confidence, ensuring a smooth and pleasurable stay in the city. Additionally, our software enables visitors to filter listings based on desired amenities, such as pools or pet friendliness, and access guest reviews and comments regarding cleanliness, further enhancing their decision-making process. Our software enables visitors to Sydney to make the most of their visit by providing visualisations and data-driven insights.

Secondly, our software must have a lot to offer researchers and analysts in the hospitality and travel sectors. Researchers can collect historical data on occupancy rates, pricing trends, and guest sentiments using its data extraction and analysis capabilities for market or academic research. Making data-driven decisions can be aided by the ability to visualise occupancy rate trends and spot patterns. To find specific data subsets for their research projects, researchers can conduct keyword-based searches. Furthering our understanding of the Sydney Airbnb market is our investigation of pricing distribution and its changes over time.

Above all, our software will be an essential tool for Sydney's landlords and Airbnb hosts to use to improve their listings. Hosts can improve their property management strategies by using data-driven decisions to examine occupancy rates and pricing trends in their respective neighbourhoods. They can review customer feedback on cleanliness and other factors to improve the overall guest experience and increase bookings. Additionally, our software provides insights into market demand and competition, enabling hosts to adjust their pricing strategies appropriately and maintain their competitiveness in Sydney's dynamic Airbnb market.

## Software Requirements

The program must meet a number of functional specifications in order to give users reliable tools for conducting in-depth analyses of data from the Sydney Airbnb dataset. These prerequisites are as follows:

**Rule 1. Information about Listings can be retrieved by User-Selected Time and Suburb.**

* Users will be able to choose a particular time frame and neighbourhood using the software.
* Once all listings that match the chosen criteria have been retrieved, detailed information about each listing will be reported.

**Rule 2. Create a Price Distribution Chart for the User-Selected Period.**

* Users will be able to choose a time period that interests them.
* The computer program will create a graph showing how the distribution of real estate prices changed over the selected time period.

**Rule 3. Keyword-Based Record Retrieval**

* The ability for users to enter a keyword (e. g. , "pool," or "pet") connected to on-site amenities.
* All records with the user-specified keyword in their descriptions will be retrieved by the software.

**Rule 4. Analyse the cleanliness feedback from customers.**

* The software will analyse customer reviews regarding elements related to cleanliness.
* It will recognize and tally remarks containing words or phrases linked to cleanliness (e. g. clean, orderly, and hygienic).
* These keywords' selection will be supported by evidence of their connection to cleanliness.

**Rule 5. Measure the occupancy rates in the suburbs.**

* The software will have the ability to calculate and examine occupancy rates for various Sydney suburbs.
* To gain insights into the usage trends of Airbnb listings over time, users can choose suburbs of interest.

## Use Cases & Use Case Diagrams

|  |  |  |
| --- | --- | --- |
| Use Case Number | Actor | Scenario |
| 1 | Social Media Traveller | 1. Traveller opens the analysis tool and selects the ‘Retrieve Listings Information’ feature. 2. They choose their desired travel time period and specify the suburb they would like to stay in. 3. The software retrieves and displays a list of available Airbnb listings in the selected suburb during the chosen dates. 4. The traveller reviews details such as; property, prices, amenities, and reviews to make an informed decision on the homestay. |
| 2 | Airbnb Host | 1. The host uses the “Keyword Based retrieval” feature. 2. They enter the keyword "smoke-alarm" to find listings that offer this amenity. 3. The software retrieves and presents a list of relevant listings. 4. The host reviews the listings, identifies competitors with similar amenities, and adjusts their pricing and listing descriptions accordingly to attract more guests. |
| 3 | Data Research Analyst | 1. The analyst opens the program and selects the ‘Price Distribution Chart’ feature. 2. They specify the time period they would like to generate the chart for. 3. The software generates and displays a visual chart that allows the analyst to see the property prices during the selected time period. 4. The analyst interprets the chart to help identify variations in pricing trends and will use the data to create an analysis report. |

A diagram of a system

Description automatically generated

Figure 2.3 - Use Case Diagram

# Software Design and System Components

## Software Design

A diagram of a system

Description automatically generated

## System Components

### Functions

|  |  |  |
| --- | --- | --- |
| Function Name | Type | Details |
| loadData | Function | The Sydney Airbnb dataset is loaded into the program by this function. It extracts information from the dataset for analysis using user-selected parameters, such as the time period and suburb, as input. |
| displayListings | Function | A display function that presents the information of all Airbnb listings within the specified suburb and time period. It formats and shows the data in a user-friendly manner, including property descriptions, prices, and amenities, to assist travellers in making informed decisions. |
| generatePriceDistrubutionPriceChart | Function | This function generates a chart to visualize the distribution of property prices during the user-selected period. It takes into account the data analysis requirements and creates a chart to illustrate pricing trends and variations for data research. |
| retrieveKeywordMatches | Function | This function is responsible for retrieving all records that contain a user-entered keyword, which could be related to property amenities (e.g., "pool," "smoke-alarm"). It searches through the dataset and presents the matching listings, enabling users to filter data based on specific criteria. |
| analyseCleanlinessComments | Function | The cleanliness function performs an analysis of customer comments related to cleanliness. It counts and presents the number of comments containing specific cleanliness-related keywords or phrases (e.g., "clean," "tidy," "hygiene"). The selection of these keywords is justified based on their relevance to cleanliness. |
| measureOccupancyRates | Function | This function calculates and analyses occupancy rates for different Sydney suburbs. Users can select suburbs of interest, and the function computes the occupancy rate based on booking and availability data. It provides insights into the utilisation patterns of Airbnb listings over time. |

### Data Structures / Data Sources

Data Structures:

* **AirbnbDataset**: Represents the Sydney Airbnb dataset and includes methods for data loading and retrieval.
* **Listing**: A class representing individual property listings, including attributes like description, price, amenities, and cleanliness comments.
* **DataVisualisation**: A class for generating charts and visualizations, such as price distribution charts.
* **KeywordSearch**: A class for handling keyword-based record retrieval from the dataset.
* **CleanlinessAnalysis**: A class for analysing comments related to cleanliness and counting relevant keywords.
* **OccupancyRateAnalyser**: A class responsible for measuring and analysing occupancy rates for different suburbs.

Algorithms**:**

**Price Distribution Analysis:** This algorithm computes the distribution of property prices within a specified time period using statistical methods like histograms or kernel density estimation.

**Keyword Matching**: The algorithm matches user-entered keywords with property listings' descriptions or attributes to identify relevant records.

**Cleanliness Comment Analysis**: It counts the occurrences of cleanliness-related keywords or phrases in guest comments, allowing for cleanliness sentiment analysis.

**Occupancy Rate Calculation:** This algorithm calculates occupancy rates by comparing booked nights to available nights, providing insights into the utilization patterns of Airbnb listings.

### Detailed Design

Algorithm for Price Distribution Chart Generation:

Input: Dataset: A list of property listings with associated prices. Time Period: The selected time period for analysis. Chart Type: The desired type of chart (e.g., histogram, box plot).

Output: A visual chart displaying the distribution of property prices.

# User Interface Design

Penpot, an open-source user interface designer renowned for its adaptability and teamwork features, served as our main tool during this design phase.

Our design process begins with a thorough analysis of the software's functional requirements to ensure that the interface perfectly matches the desired user interaction and functionality. To make the interface user-friendly and usable by a variety of users, including travellers, tourists, and Airbnb hosts, we place great emphasis on user-centric design principles.

The necessity of simple and clear navigation to various functions, the significance of visual clarity in data presentation (e. g. , charts, and graphs), and the importance of including interactive elements for user input, such as date selection and keyword entry.

The particular design elements in the following subsections, explaining how each element of the interface was created to effectively satisfy the specifications and user needs. Our goal is to develop a software interface that not only offers strong data analysis capabilities but also improves the user experience overall, enabling users to extract insightful data quickly and easily from the Sydney Airbnb dataset.

## Structural Design

For the structural design of our software interface, we have carefully considered the navigational and information structure to ensure an intuitive and user-friendly experience. The structure is designed to support efficient access to different functions and data analysis capabilities while maintaining logical groupings of information.

Navigation Structure:

1. **Main Menu:**
   * At the top of the interface, we have a main menu that serves as the primary navigation hub. It includes options such as "Retrieve Listings," "Generate Price Chart," "Retrieve Keyword Matches," "Analyse Cleanliness," and "Measure Occupancy Rates."
2. **Sub-Menus:**
   * Each main menu option leads to a sub-menu that allows users to specify parameters for their desired analysis. For example, selecting "Retrieve Listings" would lead to a sub-menu where users can input the time period and suburb of interest.
3. **Interactive Elements:**
   * We have incorporated interactive elements like date pickers, keyword input fields, and suburb selection dropdowns within the sub-menus to enhance user input and customization.
4. **Results Display:**
   * After users make selections and initiate the analysis, the results are displayed in a separate section below the sub-menus. Here, users can view listings, charts, keyword matches, cleanliness analysis results, or occupancy rate information, depending on their selected function.

Information Structure:

1. **Listings Information:**
   * The information related to Airbnb listings, including property descriptions, prices, and amenities, is presented in an easy to read and structured format. Users can scroll through the listings and click on individual listings for more details.
2. **Chart Visualization:**
   * Chart visualizations, such as price distribution charts, are displayed clearly and with accompanying labels and legends for easy interpretation.
3. **Keyword Matches:**
   * Keyword matches results are presented in table format, highlighting relevant counts.
4. **Cleanliness Analysis:**
   * Cleanliness analysis results are presented in table format, highlighting relevant comments and counts.
5. **Occupancy Rates:**
   * Occupancy rate data for different suburbs is organized in a table or chart format, making it easy for users to compare and analyse.

Justification of Design Choices:

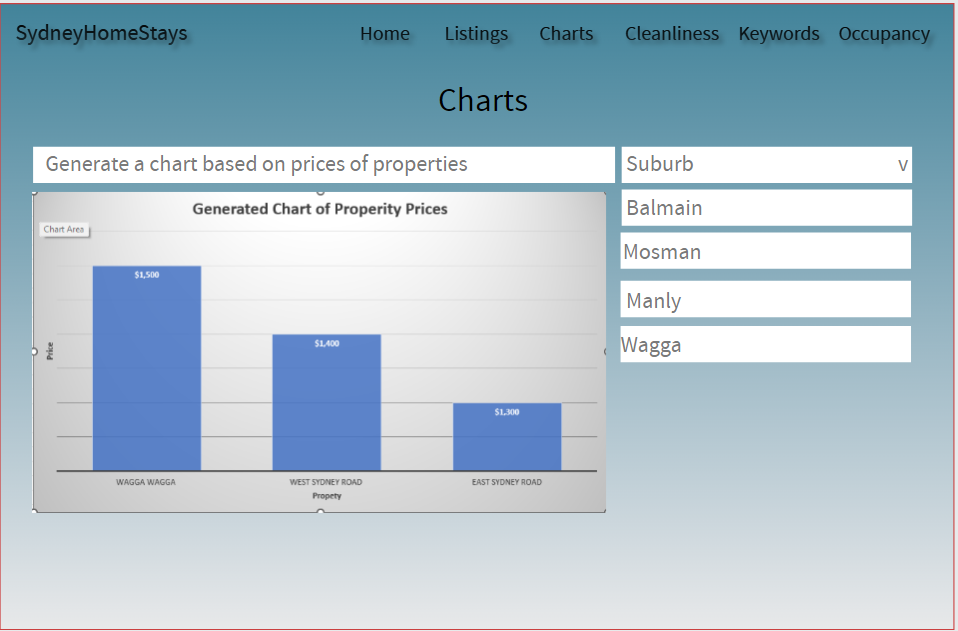
We deliberately chose this build design to achieve a smooth and user-friendly interface. By creating a clear separation between navigation and results, our design ensures that users can easily navigate the software without unnecessary confusion. The combination of submenus and interactive elements allows users to customize the analysis and enter specific query parameters, thus increasing the level of personalization. The logical organization of information in the interface simplifies the user's ability to quickly find the data they need without encountering complexity. In addition, the inclusion of clear labels and visual elements greatly improves data presentation, which ultimately improves the overall user experience and overall usability of the software.

## Visual Design

**Main Menu:**  
  
The main menu is supported by navigation elements that will allow the user to select whichever feature they need to use; they can choose between Home, Listings, Charts, Cleanliness, Keywords and Occupancy. The main menu will also have elements such as pre-generated charts and tables to help users get a better understanding of the website and its features.

**Listings Feature (Ignore Photo):**A screenshot of a home listing

Description automatically generated  
The listings feature will allow users to search for the listings and will generate things such as property name, address name, price etc. With this feature users can enter specific criteria that allows them to easier seek accommodation for the homestay and streamlines the process.

**Chart Feature:**

The chart feature will allow the users to generate a chart that based on the prices of properties that you can select via the search bar, you can also specify the target suburb.

**Keyword Search Feature:**

A screenshot of a keyword search

Description automatically generated

The keyword search feature will allow users to search for a property via keywords and be separated by a comma so they can search for multiple keywords if they wish. It will then generate a list based on the selected Airbnb keywords inputted by the user.

**Cleanliness Feature:  
A screenshot of a search engine

Description automatically generated**

The cleanliness feature will allow users to generate a report based on the selected property. The report will find key words in the comment of the data to list on it to help the user selected their accommodation better.

**Occupancy Rates Feature:**

A screenshot of a computer

Description automatically generated

The occupancy rate feature allows the user to generate a chart based on the occupancy rates of each suburb which allows them to see the popularity of the suburb. This could be used to either research the occupancy rates on trends and prices for analysts or to see if the prices in different suburbs affect the travellers accommodation.