|  |
| --- |
| System Airbnb Data System Development Executive Summary |
| Shahab Moheb (s5264189), Vaibhav Jha (s5295470), Jakob Karl (s5289718)2810ICT Software TechnologiesDate 8/10/2023 |

# Abstract

This report presents the evaluation of a project aimed at enhancing data analysis capabilities for the Sydney Airbnb dataset. The project involved the development of five primary analysis features: filtering by date, suburb, keyword, tracking cleanliness-related comments, and creating a property price distribution chart from three data files: calander\_dec18.csv, listings\_dec18.csv, and reviews\_dec18.csv. Additionally, an advanced filter feature was introduced to improve user-friendliness.

The suburb filter enables users to search for listings in a specific suburb, while the property price distribution chart visualizes listing prices over a selected date range. The 'Keyword' analysis allows users to find listing IDs containing specified keywords within defined date parameters. Customer comments related to cleanliness are tallied using a predetermined set of cleanliness-related words. Finally, the filter by date and advanced filter features enables users to refine search results based on specific date ranges, price limits, and review ratings. This report provides insights into the successful implementation of these features and their contributions to data analysis within the Sydney Airbnb dataset.

# Introduction

The purpose of this report is to evaluate the completed project. The project consisted of developing 5 analysis features to filter through three data files. The three data files were named calander\_dec18.csv, listings\_dec18.csv, and reviews\_dec18.csv. These files all related to the Sydney Airbnb dataset and provided a large data pool of listings and their costs, location, listing id, host names, room types, etc. The project was focused on developing 5 analysis features being filter by date, filter by suburb, filter by keyword, display the number of comments relating to cleanliness and develop a price distribution chart. However in the development process a sixth analysis feature was added to improve user friendliness which was called the advanced filter.

# **Analysis 1 Listing’s in Specified Suburb**

The search by suburb analysis feature of the project allowed for a user to search for a desired suburb. The system then filtered all results of the data that contained that suburbs name in the datasets suburb\_clensed column. This provided the user with all listing results that were in that specified suburb.

# **Analysis 2 Property Price Distribution Chart**

The property price distribution chart provides the user with a developed chart that has taken all the prices from the calander\_dec18.csv file of each listing and developed a graph that shows all the prices and how frequent a listing was listed at that price. The analysis feature also allows the user to select the time period between January 1st 2018 and December 31st 2019 and only get results between the two dates they choose to select.

# **Analysis 3 Retrieving Record of ‘Keyword’**

The retrieve record by ‘Keyword’ analysis feature provided users with the option to search for a ‘keyword’ which may have been a city, room type, allowed occupants, etc. The user is also able to input dates between January 1st, 2018, and December 31st ,2019, that they wish to search between. The analysis was then performed over all three .csv files and provided a result of all listing ID’s that had mention of the users specified ‘keyword’.

# **Analysis 4 Customer Comments Related to Cleanliness**

The customer comments related to cleanliness analysis function was quite simple as it was a tally of how many users had left reviews related to cleanliness. To determine cleanliness a stored array of words which the developers determined was appropriate. The chosen words were 'clean', 'tidy', 'spotless', 'dirt', 'dust', 'hygiene', 'sanitize', and 'neat'. These words were chosen as they all come to mind when someone thinks of a clean space. Once the user chooses to check how many comments relate to cleanliness the data is filtered and all record of each word is counted and added to the final tally which is then displayed to the user.

# **Analysis 5 Filtering by Date (newest-oldest)**

The filter by date analysis feature allows the user to search by date newest to oldest within a specified time between January 1st, 2018, and December 31st ,2019. The analysis is then done on the calander\_dec18.csv file and filters all listings that were posted on a date between the two specified time periods.

# **Analysis 6 Advanced Filter (newest-oldest)**

The advanced filter analysis function allows for the user to do a more precise search to filter the data. This analysis feature prompts the user to enter an amount in $ as a minimum and maximum for the search result to filter through as well as a review rating limit. The user can also choose to filter in ascending order and by price or review rating. Once the user has input all features the advanced analysis then filters the data from the listings\_dec18.csv and calander\_dec18.csv files and displays all results to the user.