

Review from Property Transaction in UK

Chow Chun Hin 230047216

Tang Wai Nok 230181860

Tsoi Ho Chun 230490841

Contents

[**1** **Introduction** 2](#_Toc183810608)

[**2** **Setting and Utilization of Github** 3](#_Toc183810609)

[**3** **Roles** 5](#_Toc183810610)

[**4** **Product Backlog items** 5](#_Toc183810611)

[**5** **Sprint Backlog items** 6](#_Toc183810612)

[**6** **Data Pre-processing** 7](#_Toc183810613)

[**7** **Answering the client's concerns** 11](#_Toc183810614)

[**7.1** **Generate (Last 20 rows) processed renting post records.** 11](#_Toc183810615)

[**7.2** **What kind of property is having the most number of the reception on average?** 11](#_Toc183810616)

[**7.3** **What is the contribution of house type in the record? What is the most common type of `property in the UK?** 12](#_Toc183810617)

[**7.4** **What is the value distribution of the number of bathroom between the flat house and terraced house?** 13](#_Toc183810618)

[**7.5** **What kind of property is contain the second most turnover?** 14](#_Toc183810619)

[**7.6** **Is there any relationship between the number of bedrooms, the number of bathrooms and the average price of a different property?** 15](#_Toc183810620)

[**7.7** **Analyze** 16](#_Toc183810621)

[**7.7.1** **Average Price vs. Average Number of Bedrooms** 16](#_Toc183810622)

[**7.7.2** **Average Price vs. Average Number of Bathrooms** 16](#_Toc183810623)

[**7.7.3** **Conclusion** 16](#_Toc183810624)

[**8** **Extra suggestion and Conclusion** 17](#_Toc183810625)

1. **Introduction**

**Background**

In recent years, immigration to the UK has become a growing concern. This has led to a noticeable rise in demand for rental properties, especially in the lettings market. To better understand the factors affecting rental transactions in the UK, a client asked us to conduct a feasibility study. This report aims to answer key questions about rental transactions by analyzing data from two days of rental property listings in the UK.

**Throughout this report we will address the following key questions:**

1. Generate (5 to 10 rows) processed renting post records.
2. What kind of property has the most number of bathrooms on average?
3. What is the contribution of house type in the record?
4. What is the proportion of room facilities on average in the detached house?
5. Which date has the more renting post?
6. Is there any relationship between the number of bedrooms and bathrooms?
7. **Setting and Utilization of Github**

Create a New Repository:

Click on the "+" icon in the top right corner and select "New repository."

Enter RepresentativeName\_4887Project as the repository name.

choose visibility (private).

Click "Create repository."

**Clone the Repository Locally**:

* Open terminal.
* Run the command:
  + git clone <https://github.com/cupkappu/RepresentativeName_4887Project.git>
  + cd RepresentativeName\_4887Project
  + git touch README.md
  + git add .
  + git commit -m "init"
  + git push origin main

After we initialize the repository, we structed the folders in the order below.

|  |
| --- |
| . ├── code # Codes of the project │ ├── analyze.sas # Code of SAS analyze │ ├── Data\_Pre-processing.rmd # SAS code explaination │ └── SAS\_code\_explan.md # R Code of Data preparation ├── dataset # Datasets(before-process and after-process dataset) │ ├── Housing2122.json # Before-process dataset │ └── Project\_Housing.csv # After-process dataset ├── outputs # The SAS analyze output │ ├── analyze-results.html # SAS html output │ └── images # The SAS Output Images │ ├── Q1.png │ ├── Q2.png │ ├── Q3.png │ ├── Q4.png │ ├── Q5.png │ └── Q6.png └── README.md |

The final Git Commit Log Tree is showed below.

|  |
| --- |
|  |

1. **Roles**

|  |  |  |
| --- | --- | --- |
| Name | Role | Responsibilities |
| Chow Chun Hin, Tsoi Ho Chun, Tang Wai Nok | Development | Responsible for coding and programming tasks related to the project. |
| Chow Chun Hin, Tsoi Ho Chun, Tang Wai Nok | Scrum Master | Supports the team in self-organization and cross-functional collaboration. |
| Chow Chun Hin, Tsoi Ho Chun, Tang Wai Nok | Product Owner | Ensuring requirements are understood and prioritized. |

1. **Product Backlog items**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sprint | Task | Category | Priority | Status |
| 1 | Setting Up GitHub Repository | Collaboration | Low | Completed |
| 2 | Data Pre-processing | Data Processing | High | Completed |
| 3 | Client Concerns Analysis Using SAS | Client Analysis | High | Completed |
| 4 | Report | Report Writing | Medium | Completed |

1. **Sprint Backlog items**

|  |
| --- |
|  |

1. **Data Pre-processing**
   1. **Convert the data source format (JSON) to the dataframe.**

We utilized the jsonlite package to import the raw JSON dataset into R and converted it into a dataframe for analysis.

|  |
| --- |
| A screenshot of a computer code  Description automatically generated |

* 1. **Tidy up the date format to become readable.**

We used the gsub function to replace "nd," "st," "th," and "rd" in the dates with an empty string (""). Afterward, we converted the date column in the dataset to the Date type and reformatted it to the dd/mm/yyyy format.

|  |
| --- |
| A computer code with white text  Description automatically generated |

* 1. **Fill all missing data with 0s except the sq. column.**

We used the is.na() function to check for missing values in the sq. column of the dataset and replaced them with zeros (0).

|  |
| --- |
| A computer screen shot of white text  Description automatically generated |

* 1. **Aggregate the number of corresponding facilities (Bed, Baths, Reception).**

We applied the same approach to three columns in total. Using the mapply function, we processed the pairs of columns (bed, beds), (reception, receptions), and (bath, baths). For each pair, we checked which column contained a value, and if neither column had a value, we assigned a 0.

|  |
| --- |
|  |

* 1. **Extract Property type from “types” column.**

**Data Cleaning Process**

We used the str\_extract function from the stringr package to extract the word between "bed" and "for sale" in phrases like *"4 bed flat for sale"*, which allowed us to capture the flattype (e.g., *flat*). Since some entries followed a different pattern, such as *"Studio for sale"*, we applied a second regular expression. We identified missing flattype values using is.na() and re-applied a regular expression to extract the word before "for sale" as the flattype.

|  |
| --- |
|  |

|  |
| --- |
|  |

After calculating the total (Q4), we removed redundant columns, including 'baths', 'beds', 'bed', 'receptions', 'reception', and 'bath', as they were no longer needed.

|  |
| --- |
|  |

**Data cleaning**

After extracting the flattype, we noticed some values had leading or trailing spaces. We used the str\_trim() function to remove these spaces.

|  |
| --- |
|  |

During plotting, we found inconsistencies in the flattype values due to uppercase and lowercase variations (e.g., *flat* vs. *Flat*). To resolve this, we used the tolower() function to convert all entries to lowercase, ensuring consistent categories.

|  |
| --- |
|  |

We identified rows with almost no meaningful data, where most columns were empty. Using the is.na() function, we checked if these rows were entirely blank and deleted them accordingly.

|  |
| --- |
|  |

* 1. **Output pre-processed data to CSV format called ”Project\_Housing.csv”**

We used the write.csv() function to save the cleaned dataset to the file path ../dataset/Project\_Housing.csv.

1. **Answering the client's concerns**
   1. **Generate (Last 20 rows) processed renting post records.**

This query retrieves the last 20 records from the ASM.Housing dataset by ordering the entries in descending order using a monotonic function, which gives a unique number to each row.

|  |
| --- |
|  |

* 1. **What kind of property is having the most number of the reception on average?**

This query calculates the average number of receptions for each property type (FlatType). It groups the data by property type and sorts the result to find which type has the highest average number of receptions. The outobs=1 limits the result to the top entry.

|  |
| --- |
|  |

The average number of receptions by property type for a "link-detached house" is 2.

* 1. **What is the contribution of house type in the record? What is the most common type of `property in the UK?**

This code creates a frequency table of property types, counting how many entries correspond to each type. The results are then visualized in a pie chart to clearly show the distribution of property types, helping identify the most common type.

|  |
| --- |
|  |

The most common property type is flat, with a frequency of 5,994 entries. The pie chart visually represents the distribution of property types, showing the contributions of each type.

* 1. **What is the value distribution of the number of bathroom between the flat house and terraced house?**

This query extracts the number of bathrooms for flats and terraced houses, organizing the data by property type. The box plot visualizes the distribution of bathroom counts, allowing comparison between the two property types.

|  |
| --- |
|  |

* Flats: Typically have a wider range of bathroom counts which is 0-4, with some outliers. Lower Median.
* Terraced Houses: Same as Flats, have a wider range of bathroom counts(0-4), with some outliers but higher median.
  1. **What kind of property is contain the second most turnover?**

This query counts the number of records for each property type to determine turnover. By ordering the results in descending order and limiting the output to two entries, it identifies the property type with the second most records.

|  |
| --- |
|  |

The terraced house ranks second, with 1,119 records.

* 1. **Is there any relationship between the number of bedrooms, the number of bathrooms and the average price of a different property?**

This set of queries calculates the average price, number of bedrooms, and number of bathrooms for each property type. The results are visualized using scatter plots to explore potential relationships between average price and the average number of bedrooms and bathrooms, aiding in understanding pricing trends based on property features.

|  |
| --- |
|  |

* 1. **Analyze**
     1. **Average Price vs. Average Number of Bedrooms**

The scatter plot suggests a positive correlation between the number of bedrooms and the average price. As the average number of bedrooms increases, the average price tends to rise as well. This indicates that properties with more bedrooms generally command higher prices.

* + 1. **Average Price vs. Average Number of Bathrooms**

Similarly, the second scatter plot shows a positive correlation between the number of bathrooms and the average price. Properties with more bathrooms also tend to have higher average prices, reinforcing the idea that additional amenities can significantly influence property values.

* + 1. **Conclusion**

Both plots indicate that there is a relationship between the number of bedrooms, the number of bathrooms, and the average price of properties. More bedrooms and bathrooms are associated with higher average prices, which aligns with common real estate trends.

1. **Extra suggestion and Conclusion**

Based on the analysis of the client's questions, several important insights and recommendations can be derived for both the marketing strategy and property offerings.

The "link-detached house" stands out with the highest average number of receptions, suggesting that this property type appeals strongly to families or buyers who prioritize social space. This feature can be highlighted in marketing materials, focusing on the spaciousness and versatility of the home. Additionally, the "flats" category, which appears to be the most common property type, indicates a competitive market. To distinguish properties in this segment, the emphasis should be placed on location, high-quality finishes, or desirable amenities, which can attract buyers looking for added value or luxury within a more affordable framework.

The "terraced houses" category reveals a higher median number of bathrooms, suggesting that this property type may appeal to buyers who value comfort and privacy. For these properties, marketing efforts should focus on the luxury and comfort aspects, particularly highlighting the larger bathrooms as a key selling point. In contrast, marketing for "link-detached houses" could emphasize family-friendly features such as spacious living areas, gardens, and communal spaces, further tapping into the appeal for those seeking a home suitable for family life.

When considering property offerings, it is clear that properties with more bedrooms and bathrooms generally command higher prices. This finding suggests that larger homes, such as "terraced houses," which feature more bathrooms, are particularly appealing to high-income buyers seeking luxury and more space. For budget-conscious customers, smaller properties with fewer bedrooms and bathrooms should be offered at a competitive price to meet the demand for more affordable housing options. By aligning property offerings with the preferences of different buyer segments—either focusing on larger, more luxurious homes for high-income buyers or emphasizing affordability for budget-conscious buyers—the marketing strategy can be optimized to cater to both ends of the market effectively.