*Please delete all the cursive text before submission. It is here just for your reference*.

*Further: data set – DS, research question – RQ*

*The mark (****x words****) after each subchapter states the word count limit. This indicates the expected amount of information which you can exceed by 10% without losing the mark.*

7COM1079-0901-2024 - Team Research and Development Project

Final report title: (*the topic of your research.)*

Group ID:

Dataset number: DS161

Prepared by: *[Name and ID of submitting student first],*

*[Name and ID of other group members]*

*Esika Arifin Rumky ID: 23027070*

*Jason Banks ID: 20050575*

***Please make sure*** *the document spelled correctly (including image labels, section headings, and table of contents). Please use correct punctuation.*

*Make sure your report is grammatically correct.*

University of Hertfordshire

Hatfield, 2024

Table of Contents *add page numbers here*

1. Introduction
   1. Problem statement and research motivation
   2. The data set
   3. Research question
   4. Null hypothesis and alternative hypothesis (~~H0/H1~~)

1. Background research
   1. Research papers (~~at least 3 relevant to your topic / DS~~)
   2. Why RQ is of interest (~~research gap and future directions according to the literature~~)

1. Visualisation
   1. Appropriate plot for the RQ *output of an R script (~~NOT a screenshot~~)*
   2. Additional information relating to understanding the data (optional)
   3. Useful information for the data understanding

1. Analysis
   1. Statistical test used to test the hypotheses and output
   2. The null hypothesis is rejected /not rejected based on the p-value

1. Evaluation – group’s experience at 7COM1079
   1. What went well
   2. Points for improvement
   3. Group’s time management
   4. Project’s overall judgement
   5. Comment on GitHub log output

1. Conclusions
   1. Results explained.
   2. Interpretation of the results
   3. Reasons and/or implications for future work, limitations of your stud

1. Reference list

Harvard (author, date) format.

1. Appendices
2. R code used for analysis and visualisation.
3. GitHub log output.

*The list below outlines the chapter/subchapter numbers, names, word count limits, and explanations of what to write in each section.*

**1. Introduction**

##### 1.1 Problem Statement and Research Motivation

Affordable housing is a growing concern in urban areas, particularly in London's borough, where median house prices have skyrocketed in recent years. This poses challenges for policy-makers who strive to balance urban growth with affordable living conditions and families who live in these areas or commute to them for work.

Understanding how factors such as population density, employment rates, and greenspace percentage influence housing costs is vital. Prior research, such as Smith et al. (2020), emphasizes the importance of examining these socioeconomic and environmental variables to gain deeper insights into housing market dynamics.

##### 1.2 The Dataset

The "London Borough Profiles 2016" dataset (Dataset ID: DS161) is interesting because it provides comprehensive insights into London's boroughs' socioeconomic and demographic characteristics. It allows for analysing how population density, employment rates, and greenspace influence key outcomes such as house prices, providing a deeper understanding of urban living conditions and economic trends. Our team chose this dataset to explore whether crowded neighbourhoods correlate with higher house prices.

##### 1.3 Research Question

**Is there a correlation between the house price and population density in London boroughs?** This question will be addressed by employing statistical analysis and visualization techniques using Rstudio to explore the relationships between the independent variables (population density per hectare, 2016) and the dependent variable (median house prices, 2014).

##### 1.4 Hypotheses

* **Null Hypothesis (H₀):** There is no correlation between the house price and population density in London boroughs.
* **Alternative Hypothesis (H₁):** There is a correlation between house price and population density in London boroughs.

To test these hypotheses Spearman’s Rho which is a non-parametric statistical method will be utilized due to the non-normal distribution of the data. This approach ensures robust results, enabling an accurate assessment of the relationships between population density and housing costs.

**2.Background research**

**2.1 Research Papers**

Research shows important insights about how population density and housing prices are connected, especially in London’s boroughs. Ahmed et al. (2021) looked at the differences in population density and housing price levels across London. They found that boroughs with higher housing prices tend to have more people living in each area. Their study used simple comparisons and supports the idea behind this research, which is also focused on comparing population density and housing prices.

Maxwell (2013) ~~in "Structural and Spatial Determinants of London House Prices~~" 🡨 [[get rid of this?]] studied how different factors, like population density, affect housing prices in London. This study gives useful details about how the housing market works in London and fits well with the current research. Maxwell’s work highlights how space and structure influence housing prices across different boroughs.

Johnston et al. (2016) examined how higher house prices and crowded areas have impacted different groups of people in London. The study, ~~"House Price Increases and Higher Density Housing Occupation,"🡨~~ [[get rid of this?]] set forth to investigate social differences, especially with respect to non-white households. This research helps demonstrate the impact of population density on housing prices and social changes, thus enhancing the relevance of this research.

Together, these studies provide a strong foundation for this work, addressing a crucial gap that allows understanding how population density and housing prices are different among London's boroughs. [[get rid of study names or add name for each reference in this section?]]

**2.2 ~~Why RQ is of interest~~** **Why this Research Question is of interest** [[maybe this should be title?]]

The research question holds importance because it identifies a specific gap in the understanding of the relation of population density with housing prices within London's boroughs. There have been earlier studies that tried to evaluate housing prices and density but mostly talked about broad trends and hardly even compared the differences between the particular boroughs. The present research takes a closer look at the variations in population density across high-and low-priced boroughs. This localized analysis would provide practical insights to urban planning and housing policies. The study fills the gap in actual area-specific research as it pertains to London. Such studies are crucial for informing the policymakers on how to better address challenges relating to affordability and population density.

### 

### 3. Visualization

#### 3.1 Appropriate Plot for the RQ Output of an R Script Appropriate [[reword this??]]

The histogram [Fig 3.1.1] [[“created in Rstudio” 🡨 add something like that to this paragraph]] shows the distribution of the median house prices within the different London boroughs during 2014. This distribution is selected to carry the frequencies of house prices into some ranges [[what does this mean?]], which helps in revealing the patterns in the housing market. The density curve overlay provides a good understanding along with this [[“along”, what does this mean? An understanding of what? Understanding of the distribution of house prices?]]. [[mention that it isn’t normal distribution, and is skewed to the left]]

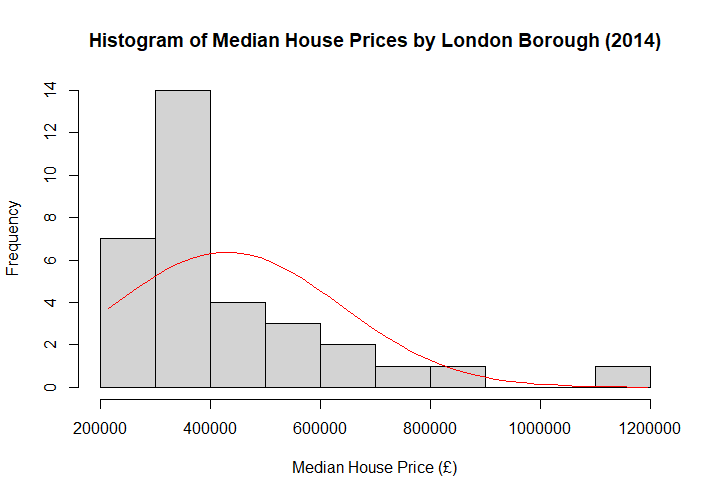


Fig: 3.1.1

The scatter plot [Fig 3.1.2] provides a sense of the association between population density (per-hectare) and median prices of houses in each borough (£). A fitted trend line indicates the weak positive correlation between the two. ~~This was a chosen plot that one wanted to test for probable associations and to present outliers by borough~~. [[ 🡨 this sentence doesn't make much sense]] [[this was the plot chosen as it best represented the correlation between the two variables along with any outliers.]]

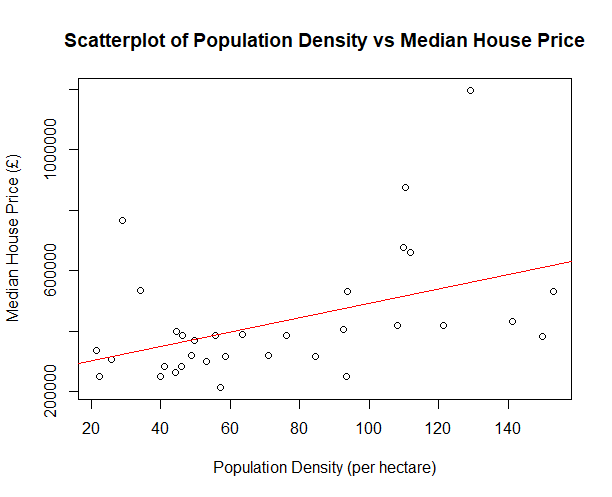


Fig: 3.1.2

#### 3.2 Additional Information Relating to Understanding the Data

The histogram (Figure 3.1.1) shows that most boroughs have lower house prices, with only a few boroughs in the higher price ranges, indicating a right-skewed distribution.

The scatterplot (Figure 3.1.2) indicates that the rising house price is generally associated with increased population density. Beyond this, the outliers may point to other influencing factors.

#### 3.3 Useful Information for Data Understanding

1. Figure 3.1.1 presents the histogram reflecting skewed house pricings populating the lowest price categories with a tail extending for high prices.
2. The scatterplot, demonstrated in Figure 3.1.2, however, showed a weak positive relation from many densely populated cities with high residential prices, but poorly demonstrated outliers dominating the complexity of the housing market itself.

**References**

Ahmed, et al. (2021). *The effects of population density on housing price tiers*. [Link to paper](https://scholars.csus.edu/esploro/outputs/graduate/The-effects-of-population-density-on/99257831021701671)

Maxwell, H. (2013). *Structural and Spatial Determinants of London House Prices*. [Link to paper](https://www.academia.edu/24939620/Structural_and_Spatial_Determinants_of_London_House_Prices?)

Johnston, R., ~~& Others.~~ [[should be “et al.” right?]] (2016). *House Price Increases and Higher Density Housing Occupation*. [Link to paper](https://research-information.bris.ac.uk/en/publications/house-price-increases-and-higher-density-housing-occupation-the-r?)

Why RQ is of interest (research gap and future directions according to the literature) **(100 word**s) [[why is this here?]]

1. Visualisation
   1. Appropriate plot for the RQ *output of an R script (NOT a screenshot)* (**50 words)**

* *Explain the choice of the plot.*
* *Anything on the plot from R is not counted towards word count limit*
* *Make sure that the plot is from output of an R script (NOT a screenshot)*
* *Make sure that the plot has a caption or title, X and Y-axis labels, legend if appropriate and units.*
* *Make sure the title or caption and axis labels are informative.*
  1. Additional information relating to understanding the data (optional) (**50 words)**
* *Per plot: explain the purpose and insights.*
  1. Useful information for the data understanding (**50 words)**
* *Summarise key observations from the plot.*

1. Analysis
   1. Statistical test used to test the hypotheses and output (**75 words)**

* *Explain the choice of the test.*
* *Make sure the test is appropriate for the RQ and data.*
  1. The null hypothesis is rejected /not rejected based on the p-value (**100 words)** *(interpret the results)*

1. Evaluation – group’s experience at 7COM1079
   1. What went well **(75 words)**
   2. Points for improvement **(75 words)**
   3. Group’s time management (**50 words)**
   4. Project’s overall judgement (**50 words)**
   5. Note any changes to group since submission of Assignment 1. Add new or amended GitHub Ids for new members **(75 words, write only if applies to your group arrangements)**
   6. Comment on the GitHub log output **(50 words)**

*Please comment on the GitHub log output, and refer to it as being placed into**Appendix B.*

*From your Git log, select the three most significant commits during this project and include the following for each:*

1. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*
2. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*
3. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*

1. Conclusions
   1. Results explained (**75 words)**
   2. Interpretation of the results (**75 words)**

* *Interpretation of what the results mean in terms of your RQ and the effect this may have on your population and the wider context of your topic.*
  1. Reasons and/or implications for future work, limitations of your study (**50 words)**

1. Reference list ***(not included in the work count)***

Harvard (author, date) format.

1. Appendices
2. R code used for analysis and visualisation ***(not included in the word count)***

Analysis.R code with the appropriate statistics to test the hypotheses.

* ***No word count****, but ensure the code is without redundant lines, well-commented and produces the correct output.*
* *Make sure it runs (look in Rscript.log for output from a statistical test)*
* *It should compute appropriate statistics to test the hypotheses*

1. GitHub log output.