**TEAM INFORMATION (Group 53)**

|  |  |
| --- | --- |
| **Team Members** | **Experience** |
| Singh, Ram C (rsingh451) | Currently a data analyst at a bank with prior experience as a process engineer in the semiconductor industry. Graduated with a materials engineering degree. |
| Gan, Hong Yee (hgan32) | Quality staff engineer with experience in the semiconductor industry. Graduated with a materials engineering degree. Completed deep learning defect image recognition proof of concept |
| Lim, Reginald Edbert K (rlim37) | A finance manager with experience in the private equity space. Graduated with an economics bachelor’s degree and a chartered accountant. |
| Rachman, Arif (arachman6) | A data analyst with experience in the biotech industry. Graduated with an IT engineering bachelor's degree and an MBA degree. |
| Tai, Jin Yao (jtai30) | Industrial Engineer in the semiconductor industry. Graduated with a chemical engineering bachelor's degree. |

**OBJECTIVE/PROBLEM (5 points)**

**Project Title: Changing Demographics and its Impact on Local Public Housing Resale Prices in Singapore**

**Background Information on chosen project topic:**

The Housing & Development Board (HDB) serves as Singapore's public housing authority and is entrusted with the task of developing housing estates in a manner that ensures affordability, quality, and a pleasant living environment. Currently, HDB flats houses more than 80% of Singapore’s resident population [1].

Prime Minister Lee acknowledges that there are supply-demand imbalances in HDB markets [2] which has resulted in higher resale prices for the past 3 years. The escalating housing prices have sparked concerns regarding housing affordability which is especially pressing for young couples who are eager to establish their families and desire to own their own homes. Unfortunately, the rapid increase in housing prices has outpaced salary growth, resulting in diminished affordability in public housing.

**Problem Statement (clear and concise statement explaining purpose of your analysis and investigation):**

Our project hopes to facilitate better policy planning and outcomes through the study of how the changing demographic structure in Singapore affects the prices of resale public housing flats.

**State your Primary Research Question (RQ):**

How do local population demographics affect the resale public housing prices?

**Add some possible Supporting Research Questions (2-4 RQs that support problem statement):**

1. Does the rate of increase in a particular age group result in a greater than proportionate rate of increase in public housing prices?
2. Does an ageing population result in overall higher public housing prices?

**Business Justification:** **(Why is this problem interesting to solve from a business viewpoint? Try to quantify the financial, marketing, or operational aspects and implications of this problem, as if you were running a company, non-profit organization, city or government that is encountering this problem.)**

One of the main objectives of the government is to ensure that public housing remains affordable for all, regardless of background, age, or educational qualifications.

One metric to measure house affordability is the Housing Price to Income Ratio (PIR) [3].

|  |  |
| --- | --- |
| Period | PIR |
| 2001 – 2020 | 4.1 |
| 2021 | 4.5 |
| 2022 Q1 | 4.8 |

The table shows a steep increase of PIR from 2020 to 2022 Q1. Generally, a PIR ranging from 4.1 to 5.0 indicates that housing is seriously unaffordable and going beyond this range signifies that housing is severely unaffordable [4].

A predictive model can be constructed to determine the demographic age group that influences housing prices most significantly, so that policymakers can implement policies tailored to the needs of this age group. Hence, this project aims to assist policy makers in devising more effective and efficient policies to tackle the rising public housing prices, focusing on the segments of the population that matter the most.

**DATASET/PLAN FOR DATA (4 points)**

**Dataset 1:** The primary dataset under investigation is the HDB resale data obtained from Data.gov.sg. Data.gov.sg serves as a government centralized platform hosting diverse government datasets across domains like the economy, education, environment, finance, health, infrastructure, society, technology, and transport of Singapore. Please find the snapshot of the dataset [5].

A picture containing text, screenshot, font, number

Description automatically generated

**Dataset 2:** Complementing the primary dataset is the consumer price index (CPI) dataset, which encompasses quarterly Singapore inflation rate from 1970 to 2022. This dataset is sourced from The World Bank, a globally recognized international organization [6].



**Dataset 3:** Another dataset to be utilized is the Population and Population Structure of Singapore, which includes information such as the Singapore Citizen population, Permanent Residence population, population growth, and population breakdown from 1950 to 2022 [7].

A picture containing text, font, number, line

Description automatically generated

**Key Variables: (which ones will be considered independent and dependent? Are you going to create new variables?** **What variables do you hypothesize beforehand to be most important?)**

Dependent: Adjusted Public housing Price (Accounted for Inflation)

Independent: Population growth in different age group (Numerical), Location (Categorical), Size of flat (Numerical), Age group- Size (Interaction), Age group-location (Interaction)

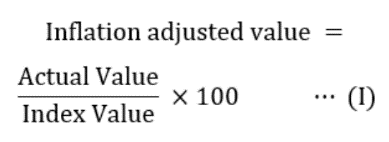
We plan to introduce interaction terms to assess the impact of various age groups on size and location of the flat and hypothesize that location and size of the flat will likely be the most important variables in determining the resale prices of the flat.

**APPROACH/METHODOLOGY (8 points)**

**Planned Approach (In paragraph(s), describe the approach you will take and what are the models you will try to use? Mention any data transformations that would need to happen. How do you plan to compare your models? How do you plan to train and optimize your model hyper-parameters?))**

The first step is to perform data cleaning which includes addressing missing data through techniques like last observation carried forward (LOCF), handling duplicate values and outliers’ treatment.

Next the datasets will be merged using a shared key; the year-month identifier. Considering the extensive timeframe covered by the dataset, it is crucial to adjust for inflation using the Consumer Price Index (CPI) to obtain the real prices. The formula for calculating the inflation-adjusted price is [8]:



Next, we will conduct feature engineering such as encoding categorical variables and creation of new interaction variables as well as feature selection using correlation matrix, VIF and PCA.

The dataset will then be divided into training, testing, and validation subsets using a split ratio of 70:15:15 and stratified across different time periods to increase the robustness of the model. Before modelling, the data is scaled to balance the impact of all predictors. The potential models to be utilized include multiple linear regression, polynomial regression, non-linear regression, regression trees and/or random forest.

After the models have been trained, we will conduct hyperparameter tuning using methods such as random search and GridSearch. The performance of the models is then evaluated using validation data using a variety of techniques such as adjusted R-squared, root mean squared error (RMSE), cross validation and information criteria such as Akaike Information Criteria (AIC) or Bayesian Information Criteria (BIC). Finally, we will estimate the quality of the chosen model using the test data.

**Anticipated Conclusions/Hypothesis (what results do you expect, how will your approach lead you to determining the final conclusion of your analysis) Note: At the end of the project, you do not have to be correct or have acceptable accuracy, the purpose is to walk us through an analysis that gives the reader insight into the conclusion regarding your objective/problem statement**

**Hypothesis:** Our hypothesis suggests that changes in age group population increases the demand in public housing. As a result, this increased demand relative to housing supply will drive up housing prices.

**What business decisions will be impacted by the results of your analysis? What could be some benefits?**

This project primarily benefits the government and policymakers as it enables them to utilize predictive analysis for formulating effective housing policies. By comprehending the relationship between demographics, inflation, and HDB housing prices, policymakers can implement appropriate measures to ensure housing affordability and maintain stability in the housing market. For instance, if the senior age group has a significant influence on housing prices and the aging population is increasing, policymakers may consider constructing more short-term lease retirement homes to cater to the needs of active, independent seniors.

Conversely, for investors, property developers, homebuyers, and sellers, this analysis provides valuable insights that facilitate the evaluation of long-term viability and profitability, risk management, and the adjustment of strategies to make well-informed decisions regarding property acquisitions and portfolio management. It also enhances public understanding of the resale market for HDB flats in Singapore.

**PROJECT TIMELINE/PLANNING (2 points)**

**Project Timeline/Mention key dates you hope to achieve certain milestones by:**

* 06/12/23 to 06/19/23 Prepare Proposal Report
* 06/18/23 to 06/25/23 Prepare Proposal Video
* 06/23/23 to 06/25/23 Data Cleaning and Feature Engineering
* 06/26/23 to 07/02/23 Exploratory Data Analysis
* 07/03/23 to 07/05/23 Prepare Progress Report
* 07/06/23 to 07/15/23 Modelling and Model Evaluation
* 07/15/23 to 07/19/23 Prepare Final Report
* 07/20/23 to 07/22/23 Prepare Video and Code Submission

**Appendix (any preliminary figures or charts that you would like to include):**

**References**

[1] “HDB | Public Housing – A Singapore Icon - Housing & Development Board,” [*https://www.hdb.gov.sg/about-us/our-role/public-housing-a-singapore-icon*](https://www.hdb.gov.sg/about-us/our-role/public-housing-a-singapore-icon) (accessed June 1, 2023).

[2] “Commentary: Supply-demand imbalances in HDB market will need time to clear,” [*https://www.channelnewsasia.com/commentary/hdb-public-housing-bto-affordability-accessibility-supply-demand-3259186*](https://www.channelnewsasia.com/commentary/hdb-public-housing-bto-affordability-accessibility-supply-demand-3259186)(accessed June 15, 2023)

[3] “The Unassuming Economist,” *The Unassuming Economist A Look at Housing Affordability in Asia Comments* [*https://unassumingeconomist.com/2019/12/a-look-at-housing-affordability-in-asia/*](https://unassumingeconomist.com/2019/12/a-look-at-housing-affordability-in-asia/) (accessed June 16, 2023)

[4] “Is Public Housing Still Affordable For The Average Singaporean Couple?” [*https://blog.seedly.sg/is-public-housing-still-affordable-for-the-average-singaporean-couple/*](https://blog.seedly.sg/is-public-housing-still-affordable-for-the-average-singaporean-couple/) (accessed June 17, 2023)

[5] “Resale Flat Prices” *Data.gov.sg.* [*https://data.gov.sg/dataset/resale-flat-prices*](https://data.gov.sg/dataset/resale-flat-prices) (accessed June 12, 2023)

[6] “A Global Database of Inflation,” *World Bank.* [*https://www.worldbank.org/en/research/brief/inflation-database*](https://www.worldbank.org/en/research/brief/inflation-database) (accessed June 12, 2023)

[7] “Population and Population Structure,” *Base.* [*https://www.singstat.gov.sg/find-data/search-by-theme/population/population-and-population-structure/latest-data*](https://www.singstat.gov.sg/find-data/search-by-theme/population/population-and-population-structure/latest-data) (accessed June 18, 2023)

[8] “How To Adjust For Inflation In Monetary Data Sets,” *Time Series Analysis, Regression, and Forecasting* [*https://timeseriesreasoning.com/contents/inflation-adjustment/#:~:text=As%20we%20have%20seen%2C%20you,multiplying%20the%20result%20by%20100.&text=This%20is%20an%20important%20formula*](https://timeseriesreasoning.com/contents/inflation-adjustment/#:~:text=As%20we%20have%20seen%2C%20you,multiplying%20the%20result%20by%20100.&text=This%20is%20an%20important%20formula) (accessed June 17, 2023)