**Report on HDB Flat Prices Analysis (1990-2021)**

**Author:**

**1. Mohamed Waleed - 41910017**

**2. Ahmed Nasser - 41920015**

**3. Omar Mahmoud Abdelkarim - 41910312**

**4. Omar Nasr Elden - 41910045**

**5. Ibrahim Emam Mahmoud - 41910270**

**6. Mostafa Mohamed Ibraheem – 41810276**

Introduction:

Provide a brief overview of the dataset and its components, including the source of the data and the purpose of the analysis.

Data Overview:

1 Data Files:

ALL Prices 1990-2021 Mar.csv: Contains over 800k rows of transactions for HDB resale flats.

Balas Table.csv: Ratios of leasehold land value to freehold land value for each year of remaining lease.

MAS Core Inflation.csv: Core CPI index values from January 1990 to February 2021.

complete.csv: Unique block addresses with latitude and longitude coordinates.

gni per capita.csv: Singapore's GNI per capita in nominal S$ for 1990-2020.

HDB machine learning.xlsx: Analysis of variables affecting resale prices using linear regression.

2 Data Preparation:

- Geocoding of addresses for accuracy.

- Handling missing or inconsistent data.

Data Observation:

- Identify and discuss any trends or patterns observed in the dataset.

Data Analysis:

1 Descriptive Statistics:

- Check for null values.

- Explore data types.

- Determine mean and maximum values.

2 Feature Importance:

- Identify unuseful features.

Data Preparation for Processing:

1 Encoding:

* Discuss encoding techniques applied to categorical data.

2 Scaling:

- Explain the difference between MaxAbsScaler and StandardScaler.

Model Building:

1 Simple Linear Model:

1 Improving Model:

- Discuss steps taken to improve the model.

- Introduce the concept of the pipeline.

2 The Effects of Degree:

- Explore the impact of polynomial degree in the model.

2 Ridge Model

3 Lasso Model

4 Elastic-Net Model

Model Comparison:

1 Metrics:

- Define and discuss metrics used for model evaluation.

2 Comparative Analysis:

- Compare the performance of different models.

Conclusion:

Summarize key findings, insights, and recommendations based on the analysis.