BRIEF REPORT

I. SUMMARY OF THE TOPIC

Based on data columns such as: AREA, INT_SQFT, BUILDTYPE, REG_FEE,... From there put into a regression model to predict the price of the house.

II. OVERVIEW OF DATASET

- The Chennai Housing Sales Price dataset is a dataset of real estate prices in Chennai, India. This dataset contains information about real estate selling prices and real estate attributes such as: Area, real estate type, real estate condition, public amenities and real estate selling price. (From 2004 to 2015).
- Dataset includes 7109 rows with 22 descriptive attributes.
- Link: https://www.kaggle.com/datasets/kunwarakash/chennai-housing-sales-price

III. DATA MINING PROCESS

- Data preprocessing: Statistics on attributes in the dataset, handling null data, converting data types,...
- Data visualization: Visualize the correlation matrix.
- Algorithms used: Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable.
- Evaluation: Function to calculate R Square and RMSE index.

IV. CONCLUSION

- State the advantages and limitations of the algorithm model building process.
- Propose future development directions.