# Projected **PRICE PREDICTION** in 5 years

## TEAM 5

**Urvashi Dube**
**Sudhish Subramaniam**

# Table of CONTENTS



# Business UNDERSTANDING

One of the finest ways to diversify net worth is to invest in land, even if it is undeveloped raw property.

It is one of the biggest steps to invest in a land. Taking this steps needs a lot of understanding and research.

In this project, the primary goal is to identify the relationship between the locality of the house, previous land value, previous improvement value, current improvement value and current land value to predict the future land value in 5 years of land in Vancouver City.

# Data UNDERSTANDING AND PREPARATION

•The dataset consists of 218563 rows (data points) and 29 columns (features). The data is owned by the City of Vancouver, BC Assessment, Finance Risk, and Supply Chain Management – Revenue Services.

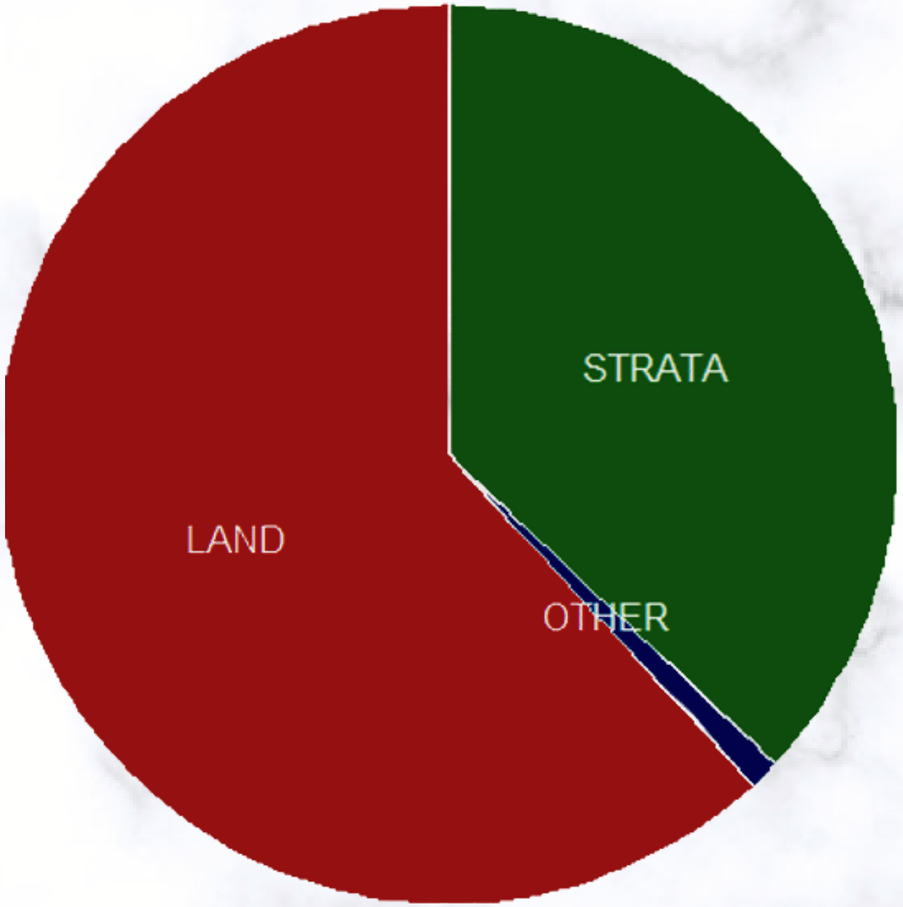
•The dataset is obtained from the open data portal of the City of Vancouver. The open data portal provides data to the public for data analysis.

•Considering the problem statement, the narrative lines are not required for processing, hence are dropped.

•To impute the null values in the dataset, all the null values in the numeric columns are replaced with the mean of the column and year columns are replaced with the mode(most frequent element) of the column.

# The ANALYSIS

# CURRENT LAND VALUE OF EACH LEGAL TYPE AND EACH ZONE

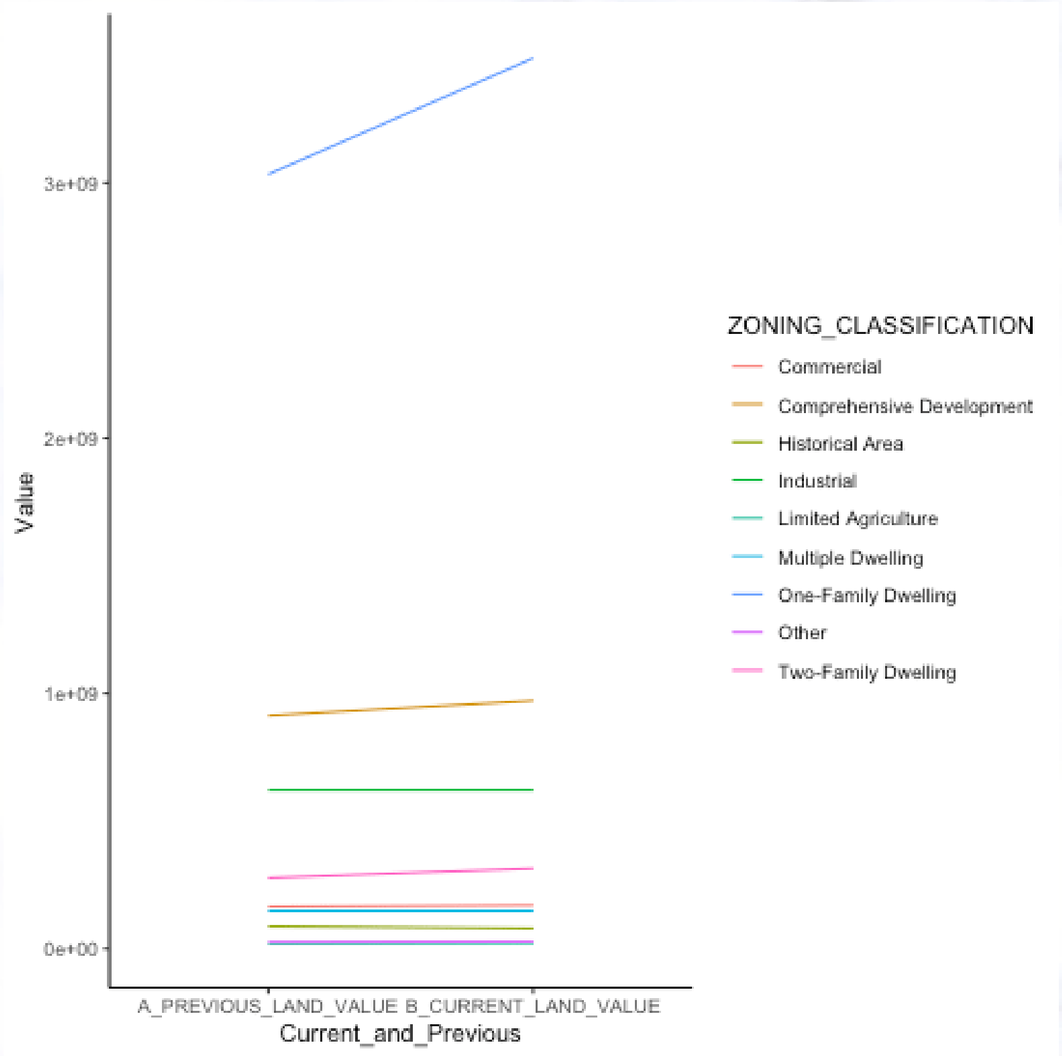




Land has the maximum current land value as compared to Strata and other types of properties

It is evident that one family dwelling is the most expensive property in Vancouver as compared to the others.

# ZONEWISE CURRENT LAND VALUE VS PREVIOUS LAND VALUE

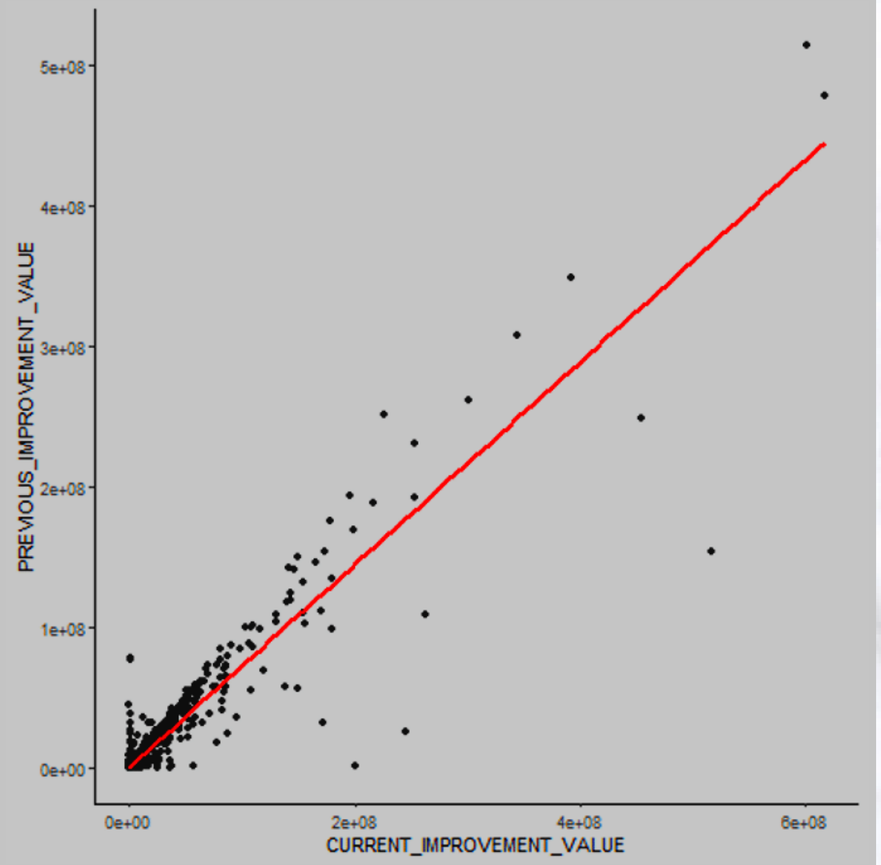



With the highest Land Value, multiple dwelling has shown an increase in the land value in the past years

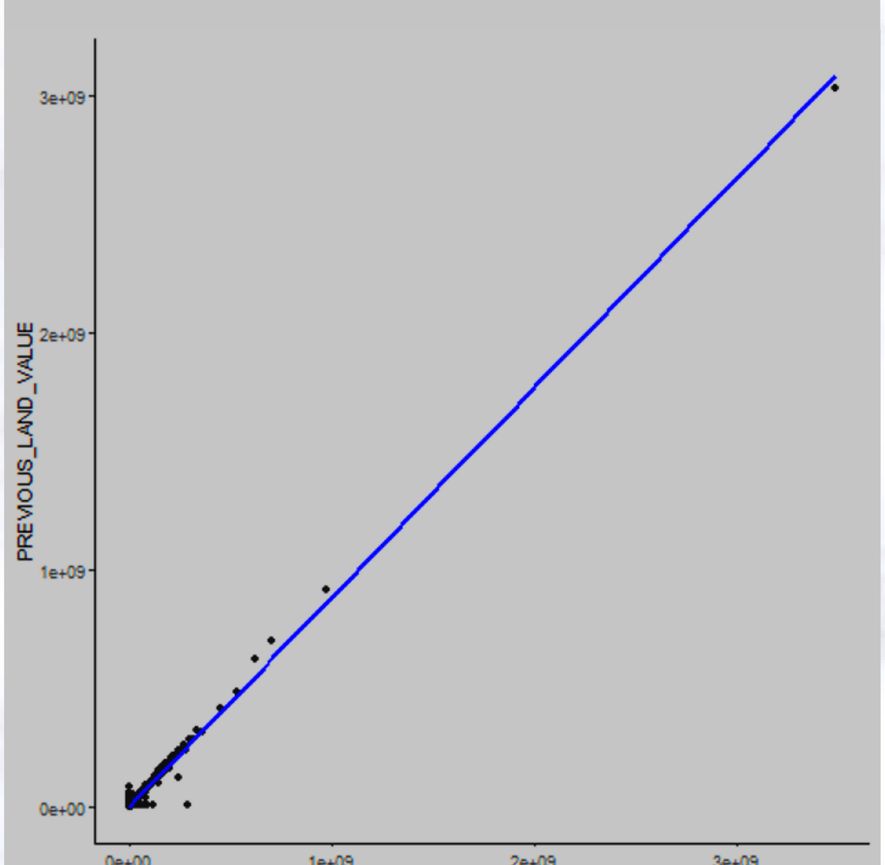
Comprehensive Developments and two family dwellings have also shown a significant increase in the land value

Apart from them, There is not much change in the land Values of other properties in the past years.

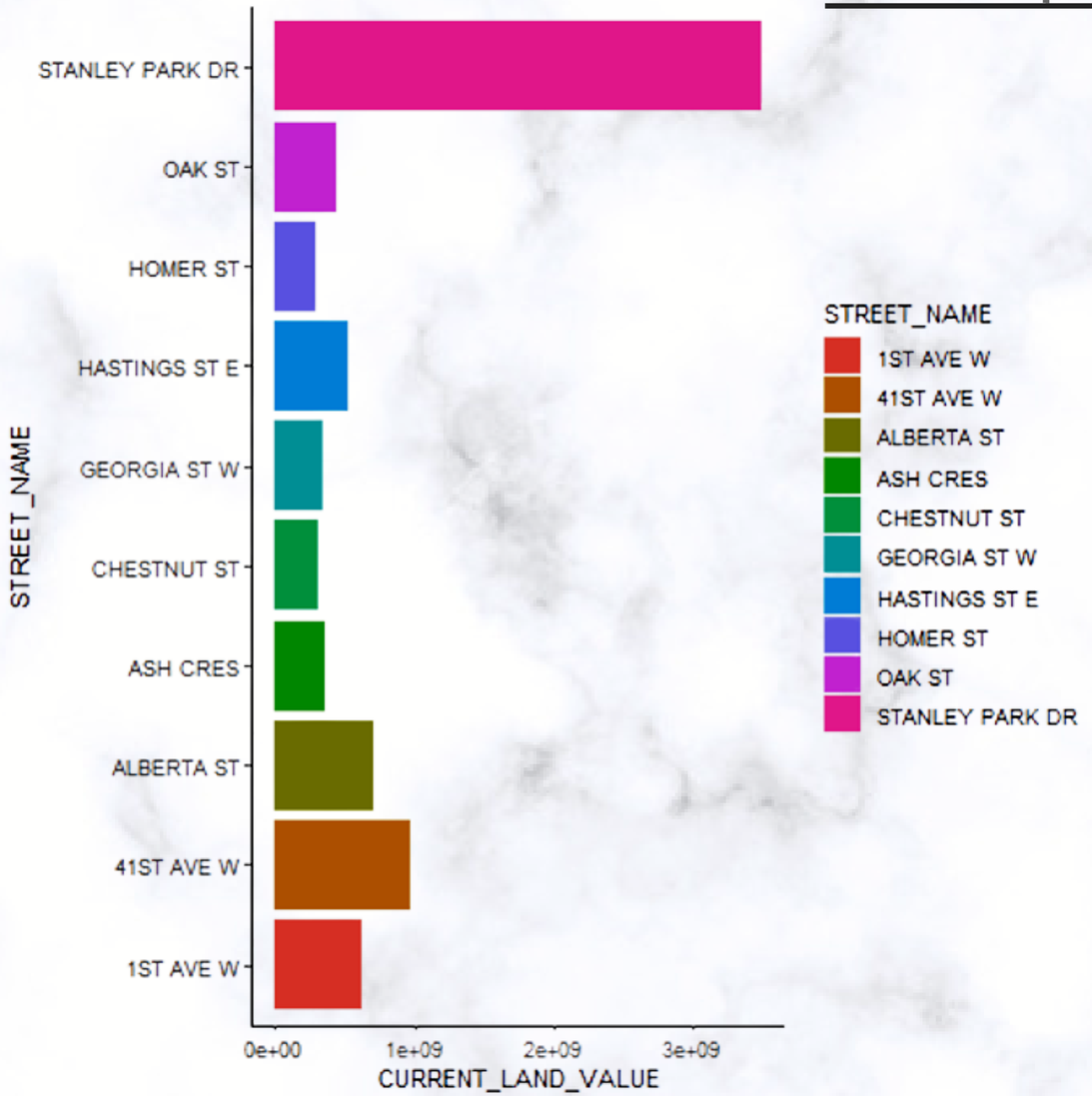
# CURRENT LAND VALUE VS PREVIOUS LAND VALUE OF EACH PROPERTY AND ZONE




It is evident that there is a linear relationship between current improvement value and previous improvement value as well as between current land value and the previous land value

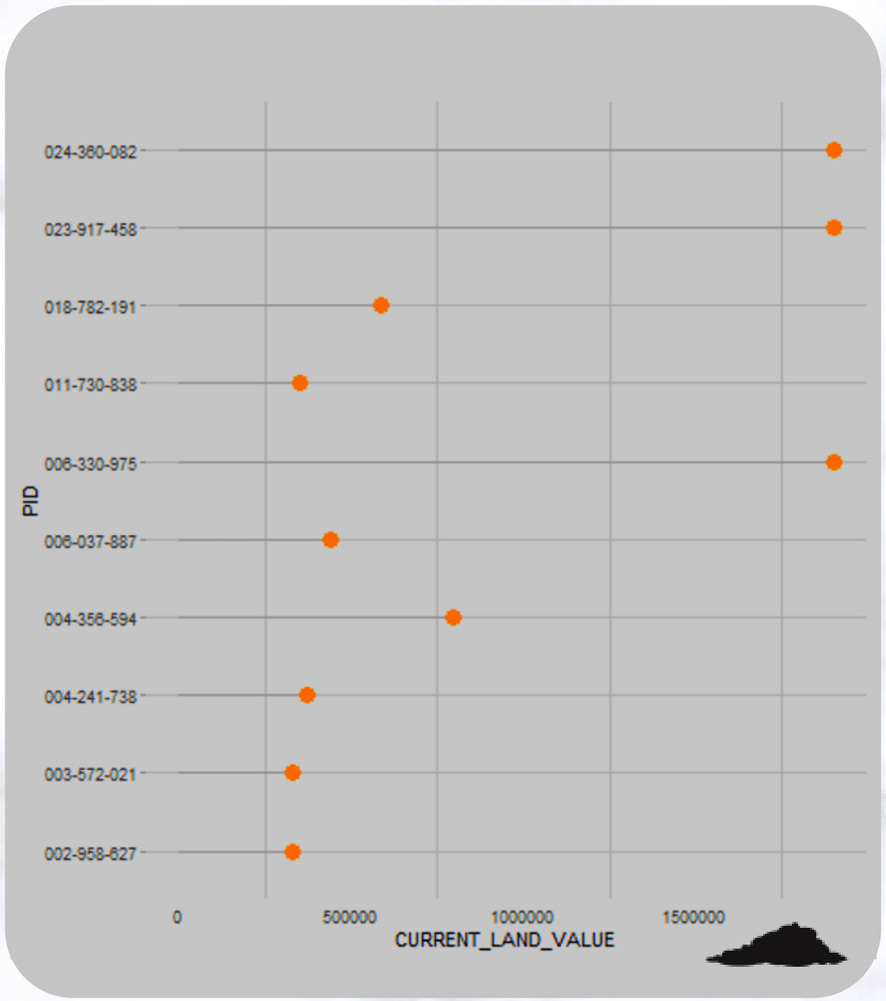


# TOP 10 EXPENSIVE STREETS OF VANCOUVER



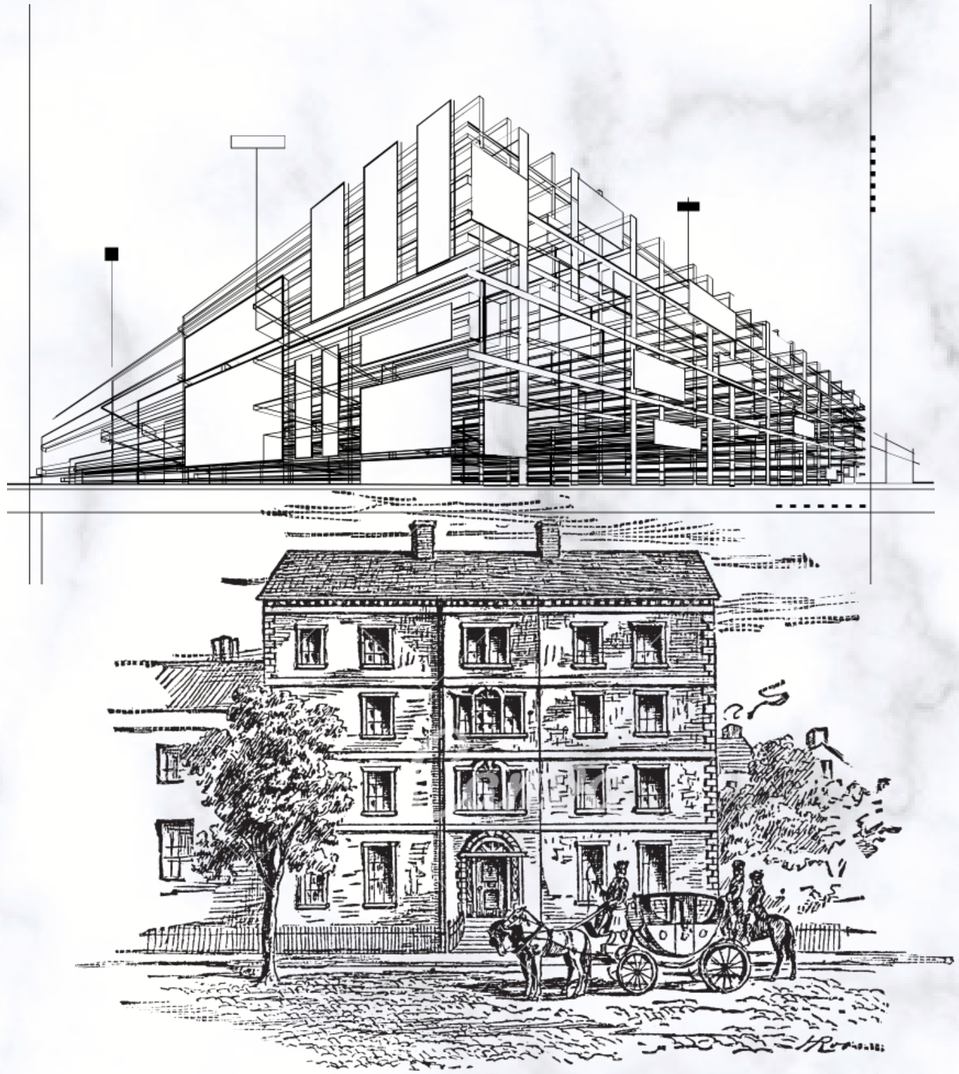

Properties in Stanley Park are most expensive And properties at chestnut Street are the least Expensive in Vancouver. 41st Avenue and Alberta Ave W also follows.

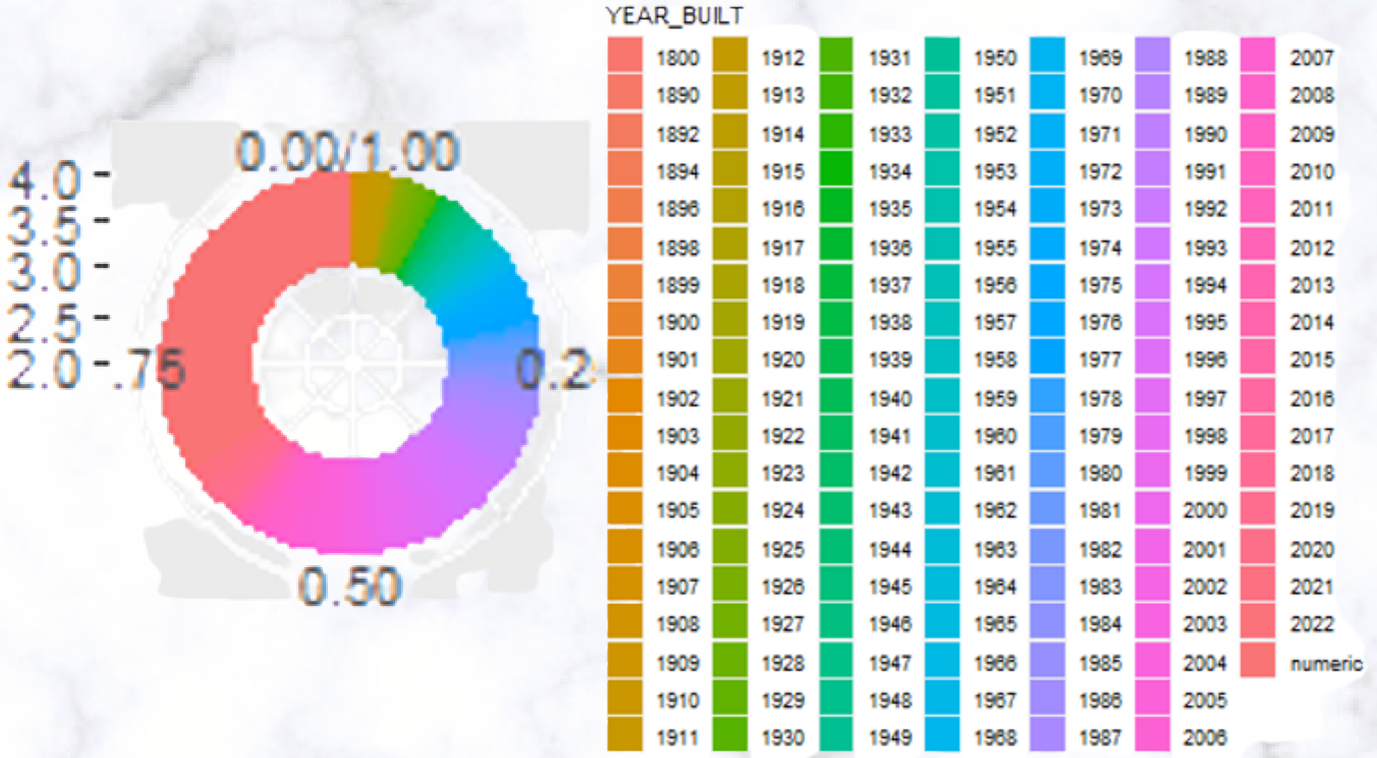
# TOP 10 MOST EXPENSIVE PROPERTIES OF VANCOUVER




Out of the top 10 most expensive properties of Vancouver, 024-360-082 property is the most expensive property and 003-572-021 is the lease expensive. These properties are in the prime location of the city with the highest current improvement value.

# RECENCY OF PROPERTIES OF VANCOUVER





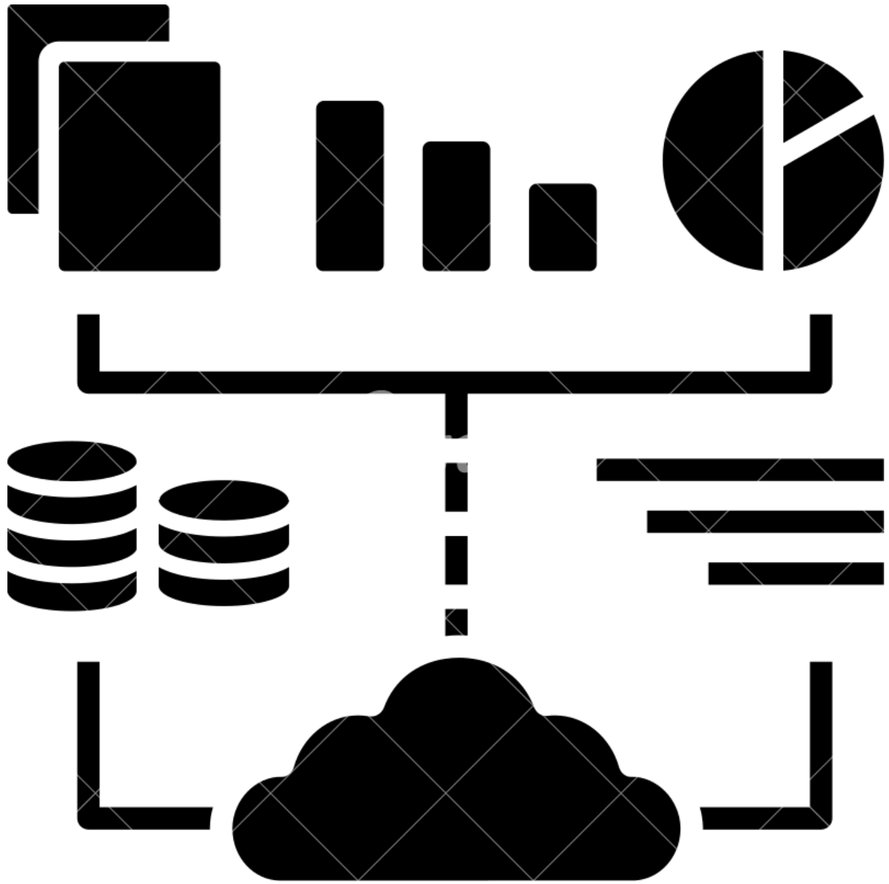

This graph clearly tells us more than 40 % of the properties in Vancouver are recently built(2007-2022).

# DATA MODELING

We have applied two multiple linear regression models to predict the future land value and future improvement value of a particular land.
We have used the Previous land value, previous improvement value, year built and Big improvement year to train and predict the current land value of the property.

We have used the PID, TAX_ASSESSMENT_YEAR, YEAR_BUILT, BIG_IMPROVEMENT_YEAR, PREVIOUS_IMPROVEMENT_VALUE, CURRENT_IMPROVEMENT_VALUE to predict the future improvement value.

On applying the other models on the train dataset, it is concluded that multiple linear regression gives the maximum accuracy.



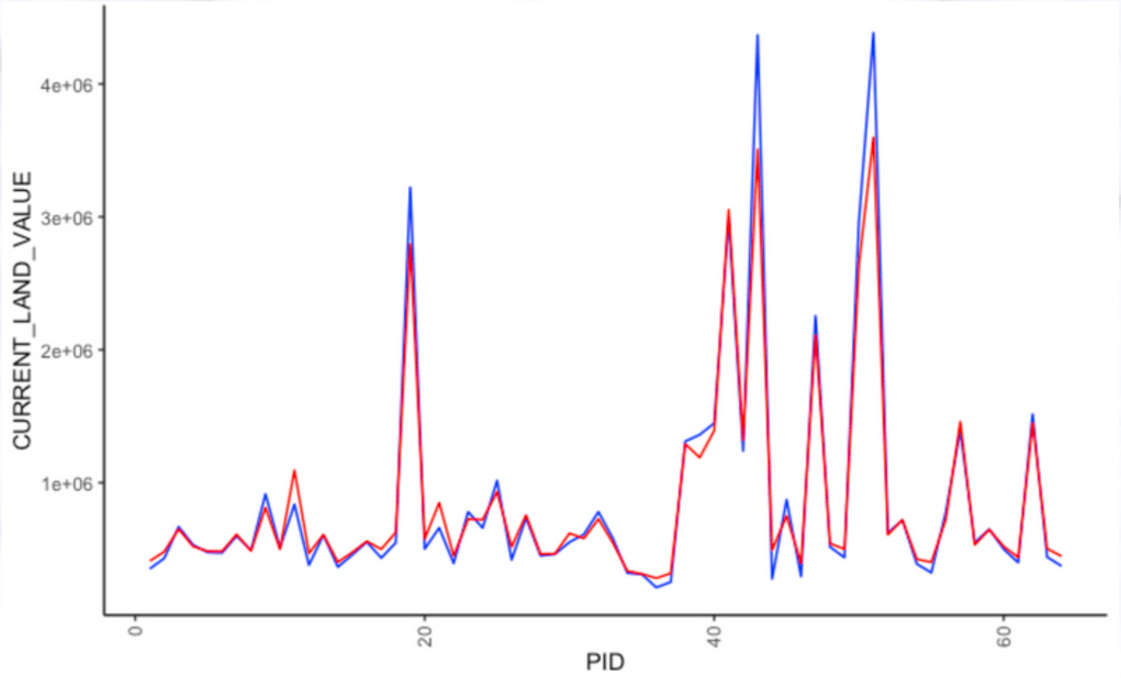
# FUTURE LAND VALUE AFTER 5 YEARS

The blue line indicates the current land values, and the red line shows the predicted land values.

The graph indicates that for the PID taken into consideration, for some properties, there will be a significant drop in the land value.

For some, there will be a slight increase, but for the majority ones, there will be no change in the value. This is because no new constructions or developments are predicted in those areas.

A significant drop in the land value, happens if there is an increase in the land value of another property in a street. The hike in the land value can be explained by the construction of new buildings, increased transportation routes and more schools, colleges, and supermarkets nearby



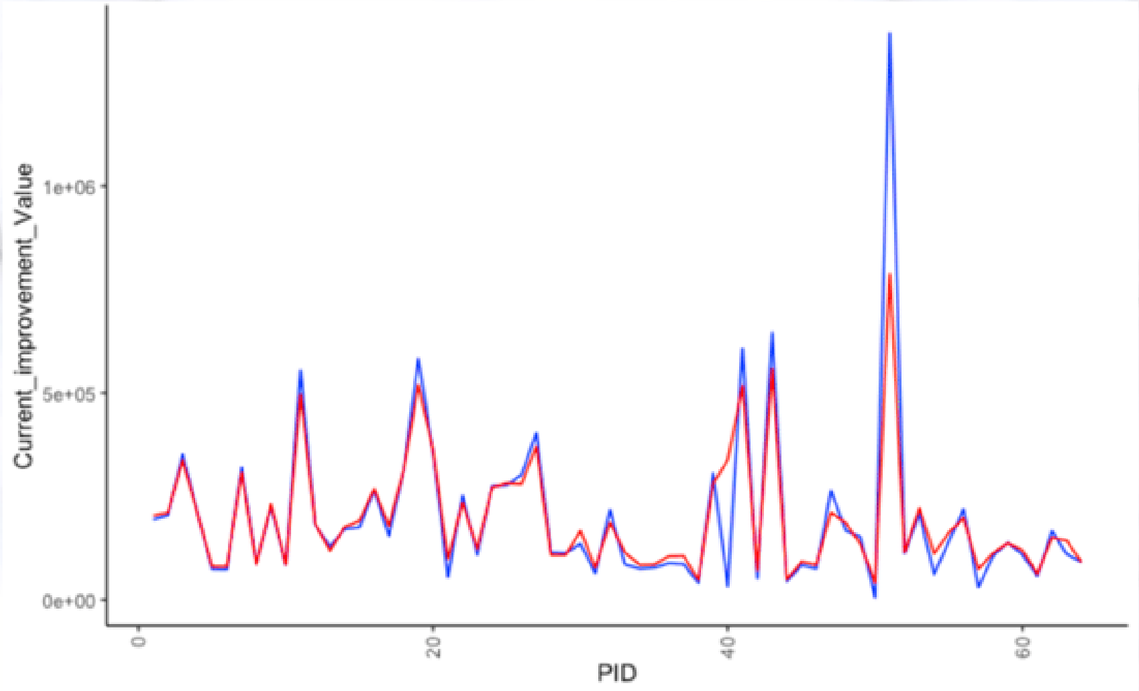
# FUTURE IMPROVEMENT VALUES

This graph shows the predicted and current improvement values.

The blue line shows the current improvement value, and the red line shows the predicted improvement value.

Considering the graph, there has been a significant decrease in the improvement value in the upcoming years.

The main reason behind this downfall can be the recession in countries, the downfall of the labour market, the lack of skilled labour, the unavailability of fruitful properties and the impact of the COVID-19 pandemic



# MODEL SUMMARY

```
Residuals:
    Min      1Q  Median      3Q     Max
-919773 -265442  -71719  166863 2309176

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                        2.675e+07  5.306e+06   5.041 4.55e-06 ***
df_for_pred$PID                    9.841e+03  3.629e+03   2.712  0.00871 **
df_for_pred$PREVIOUS_IMPROVEMENT_VALUE  4.291e+00  4.197e-01  10.224 9.17e-15 ***
df_for_pred$YEAR_BUILT            -1.361e+04  2.661e+03  -5.115 3.47e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 507700 on 60 degrees of freedom
Multiple R-squared:  0.6917,    Adjusted R-squared:  0.6762
F-statistic: 44.86 on 3 and 60 DF,  p-value: 2.458e-15
```

Summary of Linear Regression Model

```
R2(df_test_rf, df_rf_pred$prediction)
```

```
[1] 0.9833344
```

Summary of Random Forest Regression

Scores of :

- Multiple linear regression: **0.6917**
- Random Forest regression: **0.9833**

It is evident from both R2 scores of both models that Random Forest regression has better results as compared to the linear regression model.

# RECOMMENDATIONS

Sentiment Analysis could be applied on the narrative lines to get more insights and features.

Advanced Machine learning algorithms could be applied to get more deep insights about the problem.

More parameters could be considered for evaluation. Other datasets could be combined to get more insights on the dataset.

More data could be incorporated to increase the accuracy of the machine learning model

# FINAL CONCLUSION

The model has not only predicted the future land value but also the growth of the area in which the land is developed.

Builders will be helped by this, as with result of the model great insights of locality, tax rate, land quality and land value can be drawn. The model is also used to predict the next improvement value of the property.

With the dataset it is easier now to identify the most developing and promising property in Vancouver city.

# REFERENCES


•https://wowa.ca/vancouver-housing-market

•https://vancouver.citynews.ca/2022/08/12/bc-real-estate-price-decline-questions/

•https://opendata.vancouver.ca/explore/embed/dataset/property-tax-report/table/?refine.report_year=2022&dataChart=eyJxdWVyaWVzIjpbeyJjb25maWciOnsiZGF0YXNldCI6InByb3BlcnR5LXRheC1yZXBvcnQiLCJvcHRpb25zIjp7fX0sImNoYXJ0cyI6W3siYWxpZ25Nb250aCI6dHJ1ZSwidHlwZSI6ImNvbHVtbiIsImZ1bmMiOiJBVkciLCJ5QXhpcyI6ImN1cnJlbnRfbGFuZF92YWx1ZSIsInNjaWVudGlmaWNEaXNwbGF5Ijp0cnVlLCJjb2xvciI6IiMwMjc5QjEifV0sInhBeGlzIjoibGVnYWxfdHlwZSIsIm1heHBvaW50cyI6NTAsInNvcnQiOiIifV0sInRpbWVzY2FsZSI6IiIsImRpc3BsYXlMZWdlbmQiOnRydWUsImFsaWduTW9udGgiOnRydWV9
•https://opendata.vancouver.ca/explore/dataset/property-tax-report/information/?refine.report_year=2022&dataChart=eyJxdWVyaWVzIjpbeyJjb25maWciOnsiZGF0YXNldCI6InByb3BlcnR5LXRheC1yZXBvcnQiLCJvcHRpb25zIjp7fX0sImNoYXJ0cyI6W3siYWxpZ25Nb250aCI6dHJ1ZSwidHlwZSI6ImNvbHVtbiIsImZ1bmMiOiJBVkciLCJ5QXhpcyI6ImN1cnJlbnRfbGFuZF92YWx1ZSIsInNjaWVudGlmaWNEaXNwbGF5Ijp0cnVlLCJjb2xvciI6IiMwMjc5QjEifV0sInhBeGlzIjoibGVnYWxfdHlwZSIsIm1heHBvaW50cyI6NTAsInNvcnQiOiIifV0sInRpbWVzY2FsZSI6IiIsImRpc3BsYXlMZWdlbmQiOnRydWUsImFsaWduTW9udGgiOnRydWV9


# APPENDIX

| | |
|---|---|
| Linear Regression | Regression analysis model |
| Random Forest Regression | Regression analysis model |
| R2 Score | Performance metric for regression model |
| RMSE Score | Performance metric for regression model |
| Correlation | Evaluation of association between two or more variables |

# THANK YOU

Urvashi Dube
Sudhish Subramaniam