Team Members: Urvashi Dube, Sudhish Subramaniam

- **Business Problem:**

  Projected price prediction for the property in 5 years

- **Motivation**

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

  The benchmark price of homes in Metro Vancouver stands at $1,155,300, representing a 2.1% monthly decline but a 3.9% yearly increase in September 2022 [1]. A B.C. real estate expert is downplaying a possible decline in average home prices nationally by the end of next year [2]. The rise and fall of Vancouver houses are predicted in these articles, but the accurate predictions can be achieved only through machine learning and analysis of most recent data. The information will be useful to the public, who plan to make an investment or plan to buy house for living.

  The model will not only predict the future rate 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 (from the age of the building) and land value can be drawn. The model will also predict the next big improvement year.

- **Data**

  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. Each row in the dataset represents a land, the features (columns) of the dataset include: Property Identifier (PID), Legal_Type, Folio, Land_Coordinate, Zoning_District, Zoning_Classification, Lot, Plan, Block, District_Lot, From_Civic_Number, To_Civic_Number, Street_Name, Property_Postal_Code, Narrative_Legal_Line1, Narrative_Legal_Line2, Narrative_Legal_Line3, Narrative_Legal_Line4, Narrative_Legal_Line5, Current_Land_Value, Current_Improvement_Value, Tax_Assessment_Year, Previous_Land_Value, Previous_Improvement_Value, Year_Built, Big_Improvement_Year, Tax_Levy, Neighbourhood_Code, Report_Year

- **Source of Data**

  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. The link to the data and its info is mentioned in the references [3] and [4].

- **Who are the Stakeholders?**

  Builders: people who wish to develop building projects in Vancouver. It will help them to identify the most developing and profitable area.

  Public: People who wish to purchase land of their own in Vancouver either for their stay or as an investment.

Shop Owners, entrepreneurs and Companies: They can identify a suitable area of Vancouver for their business.

- **Analysis:**

  Exploratory Data Analysis:

  ➔ Scatter plot between PID, Zoning District and zoning classification to understand the groups in the dataset.
  ➔ Scatter plot between PID and street name to understand the groups in the dataset.
  ➔ Scatter plot between PID and Tax assessment year to understand the previous and new data of taxes.
  ➔ Scatter plot between PID and Year built to understand the aged and latest buildings.
  ➔ Box plot of a current land value by tax assessment year to analyse the highest and lowest values.

  Main Analysis:
  ➔ Polynomial Regression to learn the pattern between previous land value and current land value to predict the future land value (Value of the land in 5 years) using features PID, zone classification, Street name, previous land value, current land value, year built,
  ➔ Linear Regression to predict the future improvement value (after 5 years) using features PID, tax assessment year, current improvement value, previous improvement value, and year built.
  ➔ Linear Regression to predict the next big improvement year using the PID, big improvement year, tax assessment year, current improvement value, previous improvement value, and year built.
  ➔ Logistic Regression to predict if there will be an increase in the value of the land (Yes/ No). This will be achieved consider the following features: PID, tax assessment year, current improvement value, previous improvement value, and year built.

Multiple models with various variables will be performed, compared, and assessed for each of the aforementioned models using accuracy measures including RMSE (linear regression), precision, accuracy, and recall (logistic regression).

References:

[1] https://wowa.ca/vancouver-housing-market

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

[3] **Dataset**: https://opendata.vancouver.ca/explore/embed/dataset/property-tax-report/table/?refine.report_year=2022&dataChart=eyJxdWVyaWVzIjpbeyJjb25maWciOnsiZGF0YXNldCI6InByb3BlcnR5LXRheC1yZXBvcnQiLCJvcHRpb25zIjp7X0sImNoYXJ0cyI6W3siYWxpZ25Nb250aCI6dHJ1ZSwidHlwZSI6ImNvbHVtbiIsImZ1bmMiOiJBVkciLCJ5QXhpcyI6ImN1cnJlbnRfbGFuZF92YWx1ZSIsInNjaWVudGlmaWNEaXNwbGF5Ijp0cnVlLCJjb2xvciI6IiMwMjc5QjEifV0sInhBeGlzIjoibGVnYWxfdHlwZSIsIm1heHBvaW50cyI6NTAsInNvcnQiOiIifV0sInRpbWVzY2FsZSI6IiIsImRpc3BsYXlMZWdlbmQiOnRydWUsImFsaWduTW9udGgiOnRydWV9

[4] **Dataset Information:** https://opendata.vancouver.ca/explore/dataset/property-tax-report/information/?refine.report_year=2022&dataChart=eyJxdWVyaWVzIjpbeyJjb25maWciOnsiZGF0YXNldCI6InByb3BlcnR5LXRheC1yZXBvcnQiLCJvcHRpb25zIjp7X0sImNoYXJ0cyI6W3siYWxpZ25Nb250aCI6dHJ1ZSwidHlwZSI6ImNvbHVtbiIsImZ1bmMiOiJBVkciLCJ5QXhpcyI6ImN1cnJlbnRfbGFuZF92YWx1ZSIsInNjaWVudGlmaWNEaXNwbGF5Ijp0cnVlLCJjb2xvciI6IiMwMjc5QjEifV0sInhBeGlzIjoibGVnYWxfdHlwZSIsIm1heHBvaW50cyI6NTAsInNvcnQiOiIifV0sInRpbWVzY2FsZSI6IiIsImRpc3BsYXlMZWdlbmQiOnRydWUsImFsaWduTW9udGgiOnRydWV9