Big Data Analytics

Task 1

# Description

The housing dataset contains the prices and other attributes of almost

Your task is to perform an Ex- ploratory Data Analysis on the dataset.

# Dataset

The dataset contains following information attributes::

**Rooms:**Number of rooms **Price:**Price in Australian dollars **Date:**Date sold

**Type:**House type (h=house, u=unit/duplex, t=townhouse) **Distance:**Distance from Central Business District in KMs **Regionname:**General Region (West, North West, North, etc.) **Propertycount:**Number of properties that exist in the suburb **Bathroom:**Number of bathrooms

**Car:**Number of carspots **Landsize:**Land size in Metres **BuildingArea:**Building size in Metres **YearBuilt:**Year the house was built

# Task

Your task is to perform EDA and calculate the strength of relationships between the variables of the dataset. Consider below as a guideline:

1)Your task is to clean the dataset and prepare it for analysis by e.g. removing/replacing NAs, outliers, and incorrect values.

2) Begin your analysis with a summary of the variables (use basic statis- tical methods). Brie pie chart, bar chart, histogram, scatter plot. Each plot should display di variables. Each plot must have a title and meaningful labels.

3) Focus your analysis on the price variable:

1. Show the histogram of the price variable. Describe it brie In- clude summary statistics like mean, median, and variance.
2. Group houses by some price ranges (like low, medium, high, etc.) and summarise those groups separately.
3. Explore prices for di house types. You might want to use the boxplot.
4. How di attributes are correlated with the price? Which

variables are correlated the most with price?

4) List the frequencies of houses for various types. Create scatter plots and colour the house price by landsize and type.