

## PHASE-2 PROJECT.

### **Project Overview**

For this project, you will use multiple linear regression modeling to analyze house sales in a northwestern county.

It will work in providing stakeholders with information on how to increase their return on investment and the factors that contribute to getting high returns depending on the house features and the house ratings.

### **Introduction:**

The real estate market is a significant sector of the economy, with many homeowners buying and selling properties as investments. As such, it's essential for stakeholders in the real estate industry to maximize their returns on these investments. One way to achieve this is by providing advice to homeowners about how home renovations might increase the estimated value of their homes, and by what amount. This is where multiple linear regression modeling comes into play.

In this project, we will use multiple linear regression modeling to analyze house sales in a northwestern county. Our goal is to provide insights into the relationship between various home features and the sale price of the houses. By identifying the features that have the most significant impact on the sale price, we can help real estate stakeholders make informed decisions about home renovations and investments.

We will be using the King County House Sales dataset for our analysis, which contains information about various home features and the sale price of the houses. We will perform exploratory data analysis, handle missing data and outliers, and assess the assumptions of linear regression before interpreting the results. Our ultimate aim is to help stakeholders in the real estate industry make better-informed decisions that will enable them to fetch more returns on their investments.

**Problem Statement:**

The problem statement for this project is to build the best predictive model for house prices in King County, Washington, using various features. Our goal in this project is to analyze the relationship between various home features and the sale price of houses in a northwestern county. We have identified four potential independent variables that we believe may have a significant impact on the sale price of the houses: square footage of the house above ground level, grade, year built, and a number of bedrooms.

Our aim is to provide insights and advice to stakeholders in the real estate industry about how they can improve their returns on investments by focusing on the features that have the most significant impact on the sale price of the houses.

**Objectives.**

Objective 1: To determine the relationship between the square footage of the house above ground level and the sale price of the houses in a northwestern county using multiple linear regression modeling.

Objective 2: To examine the relationship between the overall grade of the house and the sale price of the houses in a northwestern county using multiple linear regression modeling.

Objective 3: To explore the relationship between the year built and the sale price of the houses in a northwestern county using multiple linear regression modeling.

Objective 4: To investigate the relationship between the number of bedrooms and the sale price of the houses in a northwestern county using multiple linear regression modeling.

## **Challenges.**

- The lack of affordable housing is the rapid growth of the county's population. As more people move into the area, the demand for housing increases, which can drive up prices and limit the availability of affordable option.
- The grading system can be skewed, certain properties are overvalued or undervalued, based on factors that are not related to the quality of the property.
- Scarcity of available units for residence within Kings County which has driven up the cost of house units between the various grades ie: average, good, excellent and luxurious.
- The disparities in the housing market play out on a subregional basis within King County Properties located in desirable areas, such as waterfront or downtown areas are limited and highly valued.
- Limited supply of land in the region, particularly in desirable areas close to job centers and transportation. The scarcity of land can limit the number of available units and drive up prices.