Data Science Project Proposal: Linear Regression on USA Housing Data

Project Title: **Predicting Housing Prices using Linear Regression**

Project Overview:
In this data science project, we will conduct a linear regression analysis on a dataset of USA housing prices. The goal is to build a predictive model that can estimate the housing prices based on various features such as location, square footage, number of bedrooms, and more. This analysis will provide valuable insights into the factors influencing housing prices and allow us to make predictions for future property values.

Objectives:

1. Collect and preprocess the USA housing dataset.
2. Perform exploratory data analysis to understand the distribution of variables and identify potential outliers.
3. Split the dataset into training and testing sets.
4. Build a linear regression model using relevant features to predict housing prices.
5. Train the model on the training set and evaluate its performance using appropriate metrics.
6. Validate the model's accuracy on the testing set.

Data Source: The dataset will be obtained from a reliable source such as a real estate agency, housing database, or government dataset. It will contain information about various housing features and their corresponding prices.

Timeline:

- Data Collection and Preprocessing: 1 week
- Exploratory Data Analysis: 1 week
- Model Building and Training: 2 weeks
- Model Evaluation and Fine-Tuning: 1 week
- Finalising Project Report: 1 week

Project Team:
The project will be led by a data scientist with expertise in machine learning and housing market analysis. The team may include data engineers for preprocessing and visualisation specialists for exploratory data analysis.

Conclusion:
By completing this project, we aim to create a valuable predictive model that can assist stakeholders in understanding housing price trends and making informed decisions related to real estate investments and property transactions.