The Proposal

A Problem

The housing market has a big problem due to a lot of fluctuating prices which are influenced by many different factors like location, square feet, the number of bedrooms in the house and number of bathrooms, as well as amenities. Being able to accurately predict house prices can be really challenging, this leads to a lot of difficulties for buyers, sellers, and real estate agents in making some informed decisions.

A Proposed Method of Solution Involving a Machine Learning Algorithm

I would like to propose a linear regression be employed in the design of a predictive model on house prices. It looks into and analyzes the past data to establish a relation between housing features and prices, which then provides correct price predictions based on input variables.

Potential Stakeholders

- Homebuyers: Anyone interested in knowing the fair home market prices.
- Real Estate Agents: Agents who can use the model to provide data-driven price recommendations to clients.
- Investors: People investing in real estate and are looking to determine the value of a potential investment property.
- Housing Market Analysts: Analysts who attempt to understand trends and predict what the pricing would be in the future

Potential Obstacles

- Data Availability: The task of acquiring a dataset that is comprehensive enough to capture a variety of features of houses and their sale prices.
- Data Quality: Ascertains that this dataset is clean, accurate, and representative of the market.
- Overfitting: It is the risk of fitting the model too closely to the training data, hence generalizing badly on new, unseen data.

What is Novel About Your Problem and Your Potential Approach

The novelty of this approach lies in the way that linear regression is not used just to predict house prices, but also to help us understand which factors are most important in determining those prices. We can improve this method by including interactions between features and seeing complex relationships, making it stronger than standard models.

The Plan

1. **Where do you plan to get your datasets?**
    - I will source datasets from websites such as Kaggle, the U.S. Census Bureau, and real estate websites.
    - **Storage Plan:** The collected data will be stored in a CSV file and saved in google drive.
    - **How do you plan to organize your data?** The dataset will be organized in tabular format, with columns for variables such as price, location, square footage, number of bedrooms, number of bathrooms, and amenities.
2. **How do you plan to analyze your data? What are you looking for?**
    - I will be utilizing exploratory data analysis to visualize different relationships between the housing features and the prices. I will also be looking for any correlations and patterns that would be able to inform the linear regression model.
3. **Based on your problem statement, what type of model do you plan to use?**
    - I will use a linear regression model.
4. **How will you determine if your model is accurate?**
    - I will evaluate model accuracy using metrics such as R-squared, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). I could also use cross validation to assess model performance on unseen data.
5. **Are there similar models/papers/code online?**
    - Yes there are similar models out there that have tried to address this problem as well.