PHYS243 Project Proposal

Group 4
James Siefert, Liam Hsieh, Panawannage Fernando, and Christopher Moussa

---

Dataset Link
Angi 1, **Angi 2**, **Angi 3**, **Angi 4**, **Angi 5**, **Angi 6**

Project Description & Background

This project will investigate the value increase associated with various home improvements in order to help homeowners maximize their return on investment (ROI). By analyzing a real estate dataset from Kaggle, we will identify how various *changeable* features of a house (square footage, number of bedrooms/bathrooms, number of stories, presence of a guest room, central air conditioning, and furnishing status) impact the house's price while also controlling for *non-changeable* features (such as proximity to a main road, presence of a basement, hot water heating, parking spaces, and preferred area status). These features are considered *non-changeable* due to practical constraints.

The results of this project will provide insights for homeowners about which improvements are likely to provide the best "bang for their buck". Also, the housing market plays a crucial role in today's national economy, representing one of the most significant sectors in terms of financial transactions and investment opportunities. The valuation and prediction of housing prices have garnered significant attention from researchers, policymakers, and real estate professionals, as it directly impacts numerous stakeholders, including homeowners, buyers, sellers, investors, and government institutions.

Proposed Project Implementation

1. **Pricing Model**: We will preprocess the Kaggle dataset by separating out the home features from the price. The dataset will then be split into training and testing sets, and we will train various machine learning models to predict house prices based on the features. The model will be validated using the testing set and we will utilize the model that gives us the best results.
2. **Home Improvement Cost Model**: Utilizing data from Angi's website, we'll create a cost matrix, including factors like conversion type, square footage, impacted home features, and high/low cost estimates. We'll employ machine learning to mold this matrix data into a predictive model, capable of accommodating various project parameters. Limited data and necessary assumptions will be explicitly acknowledged in our report.
3. **Monte Carlo Simulations**: To account for uncertainty in improvement costs, we'll employ Monte Carlo simulations. These simulations will generate a distribution of potential costs for each home improvement, offering a more comprehensive view of possible outcomes compared to single-point estimates.
4. **ROI Analysis**: Combining our pricing model and the results from the Monte Carlo simulations, we will develop a predictive model for ROI. This model will estimate the potential ROI for different improvements, given a house's non-changeable features and considering the uncertainty tied to the cost estimates.

5. **Tool Development**: Create an interactive tool that helps homeowners input their current house features and receive estimated ROIs for various potential improvements. The tool will present results in a user-friendly format, clearly explaining the underlying assumptions and calculations to help homeowners make informed decisions.

Deliverables

1. **Jupyter Notebook**: This will contain the Python code for the entire project, including home pricing model development, improvement cost analysis, and ROI analysis. It will include detailed comments explaining the purpose and function of each section of the code.
2. **Interactive ROI Tool**: A user-friendly tool with graphical user interface that will allow homeowners to input their current house features and see estimated ROIs (with uncertainty) for various potential improvements. The tool will have options that allow the user to customize their analysis.
3. **Final Project Report**: Each member will submit their own unique version of the final report which will be 4-10 pages in length, and include an introduction, problem statement, applicability of machine learning, methodology, results, and conclusions.