Project Abstract

Content-Based Recommender System for Real Estate

Team members: **Firdose Shaik** **Jenvith Manduva** **Sri Charan Namana** **Sri Charitha Narra**

Understanding the Dataset:
In the realm of real estate, the deluge of property listings can be overwhelming. Our project leverages a publicly available dataset comprising diverse property listings, including flats and houses. Each entry in the dataset meticulously details various attributes such as location, price, area size, number of bedrooms, amenities, and more. The dataset's richness allows our team to hypothesize about the underlying patterns, preferences, and also bedrock for our exploratory analysis that drive user choices in real estate. By examining these measurements, we aim to craft a personalized recommendation engine that aligns with potential buyers' aspirations, thereby making the property selection process less daunting and more focused.
Dataset link: https://www.kaggle.com/datasets/pradeepsapparapu/bengaluru-house-datacsv

We might be taking more data depending on the requirement for this project.
The dataset's comprehensiveness ensures the feasibility of our project. Our team, armed with keen analytical skills, intends to dissect this data to uncover correlations and trends that will inform our recommendation algorithms. This deep dive into the data will allow us to pose interesting hypotheses and explore them through both low-risk methods, ensuring steady progress, and high-risk, high-reward approaches that could potentially revolutionize how real estate recommendations are made.

Goals for Exploratory Data Analysis (EDA):
As the leader of a resource-conscious team, our EDA goals are **twofold**: to gain actionable insights from the data and to lay the groundwork for an effective recommender system. We will embark on a journey through the dataset, starting with univariate analysis to understand the distribution of individual variables, moving to bivariate analysis to explore the relationships between property prices and features like location and size, and concluding with multivariate analysis to decode the complex interplay of multiple variables.

Specifically, our EDA will aim to:

1. Identify the most influential property features that drive user interest.
2. Predict the market trends based on property attributes.
3. Understand user behavior patterns to tailor property recommendations.

To achieve these objectives, we will apply three key procedures from our class:
Correlation Analysis: To identify the relationships between different property features.
Cluster Analysis: To segment properties into distinct categories based on their features.
Principal Component Analysis (PCA): To reduce dimensionality and focus on the most relevant features.
We are cognizant that not all procedures will yield successful outcomes, and some questions may remain unanswered. However, our strategy includes a mix of methods to ensure that some will provide the insights we seek. Our plan is to document both our successes and the reasons behind the failures in the Milestone report, providing a clear understanding of our project's trajectory.

The project is meticulously planned to match the caliber of our team without overreaching our capabilities. By balancing ambition with practicality, we aim to demonstrate that our project is not only viable but positioned to make a significant impact on the real estate market through the power of data science.