

Project Proposal

Group2

Caroline Bordin, Erjola Lluka, Jihye Yoon, Sughra Shadab, Supreet Ahuja

Project overview

The project aims to develop machine learning models do

- **Price Prediction:** predict property prices based on features such as location, size, amenities, and market trends. (Supreet Ahuja, Erjola Lluka)
- **Recommendation Systems:** suggest properties to potential buyers or renters based on their preferences, search history. (Jihye Yoon)
- **Market Analysis:** identify trends, hotspots, and investment opportunities.(Caroline Bordin)
- **Visualization and data analysis on Tableau** (Sughra Shadab)

Based on key features such as location, number of bedrooms and bathrooms. Leveraging a dataset sourced from Kaggle, we intend to create robust predictive models that **can assist home buyers, sellers and real estate professionals in making informed decisions**

Project plan

Day1, Day 2 (Nov 13, 15)		Day 3(Nov 16)		Day 5 (Nov 22)		Final Day (Nov 22)	
Clean data	Supreet	Work on model	Everyone	Showing visualization to team	Everyone		
				Sending readme to Caroline	Everyone		
proposal	Jihye	Day 4, (Nov 20)		Final editing README	Caroline	Organizing Github repository	Jihye
Save data to sql	Jihye	Showing progress To a team	Jihye, Caroline, Supreet, Erjola	Ppt file	Jihye, Caroline, Supreet, Erjola	final discussion for presentation	Everyone
Select machine learning model	Everyone	Push a model to github	Jihye, Caroline, Supreet, Erjola	Make a script	Everyone	Rehearsal	Everyone
		Work on Tableau	Sughra	Work on Tableau	Sughra	presentation	Everyone

1. Data Collection and Understanding:

- Load and inspect the data from the given CSV file.(Supreet)
- Check for missing values or outliers and handle them.(Supreet)
- Understand the meaning of each column, such as address, price, longitude, latitude, number of rooms, and number of bathrooms.(Everyone)
- Save the data to SQL (Jihye)

2. Data Preprocessing:(Everyone)

- Decide how to use address information. For example, you may use only the city name or extract geographical features.
- Convert categorical data to numerical data
- Scale continuous data like price.

3. Data Visualization and Exploratory Data Analysis (EDA):(Everyone)

- Perform statistical and visual analysis of the data to identify patterns or interesting relationships.
- Examine the distribution of prices, correlations between features, etc.

4. Select Machine Learning Model:(Everyone)

- Depending on the nature of the problem, consider treating it as a regression problem since you want to predict property prices.
- Consider algorithms such as linear regression, decision trees, random forests, or gradient boosting.

5. Data Splitting:(Everyone)

- Split the entire dataset into training and testing sets to evaluate the model's generalization performance.

6. Model Training:(Everyone)

- Train the selected machine learning model using the training dataset.

7. Model Evaluation:(Everyone)

- Evaluate the model's performance using the test dataset. Metrics like Mean Squared Error (MSE) or R-squared can be used.

8. Model Tuning:(Everyone)

- Adjust hyperparameters of the model or try different models to improve performance.

9. Results Interpretation:(Everyone)

- Interpret the model's prediction results and understand the impact of each feature on property prices.
- Readme file.(Caroline)

10. Presentation preparation(Everyone)

- Decide how we are going to present the model and make a presentation file.
- Work on own script
- Rehearsal for a presentation

Technologies and libraries we plan to use

- Python Pandas
- Python Matplotlib
- HTML/CSS/
- SQL Database
- Tableau
- Scikit-learn