**Project:** Renta

**Introduction:**

Pakistan is a country with a population of over 220 million people, and about 25% of the population lives in rented housing. Finding suitable rental properties that meet the tenants' needs and are affordable for them is a challenging task, and it becomes even more challenging during these difficult times. This problem affects both tenants and property owners, as tenants struggle to find suitable rental properties, and property owners struggle to find suitable tenants for their properties.

This project aims to address this problem by developing a model that helps tenants find affordable rental places that meet their specific needs and preferences. Since the scope of this project is quite broad, we will also be continuing this as our FYP.

**Why Renta:**

The problem of finding suitable rental properties in Pakistan is essential to solve for several reasons:

1. **Access to housing is a fundamental human right:** Everyone should have access to safe and affordable housing. When people struggle to find suitable rental properties, it can lead to homelessness, poor living conditions, and inadequate housing.
2. **Economic stability:** The rental market plays a crucial role in the overall economic stability of the country. When tenants can't find suitable properties, it affects their ability to work and contribute to the economy. Similarly, when property owners struggle to find suitable tenants, it affects their income and ability to maintain their properties.
3. **Social impact:** The lack of suitable rental properties can have a significant social impact, especially on vulnerable populations such as low-income families, single-parent households, and elderly people. These populations may have difficulty finding suitable housing, which can lead to social exclusion and marginalization.

Overall, solving the problem of finding suitable rental properties is essential for the well-being of individuals, the economy, and society as a whole.

**Literature Review:**

Following are the papers that we studied that matches to our project.

1. **"Predicting Airbnb Listing Prices for New York City Using Machine Learning Techniques" by Oza and Srivastava:**

This paper uses multiple machine learning techniques, including linear regression and decision trees, to predict Airbnb listing prices in NYC. They use features such as the number of bedrooms, bathrooms, and amenities to predict the price. Their work is related to ours in terms of predicting Airbnb listing prices in NYC using machine learning techniques.

1. "**Airbnb Price Prediction using Machine Learning" by Liu and Hu:**

This paper also uses machine learning techniques to predict Airbnb listing prices. They use features such as the number of bedrooms, bathrooms, and reviews to predict the price. Their approach is similar to ours in terms of using features related to the property and user experience.

1. **"Predicting the Price of Airbnb Listings Using Machine Learning" by Jang and Lee:**

This paper uses a combination of machine learning techniques, including linear regression and gradient boosting, to predict Airbnb listing prices in NYC. They use features such as the location, property type, and amenities to predict the price. Their approach is similar to ours in terms of using a combination of machine learning techniques and features related to the property.

**Furthermore we also studied.**

**Predicting Price of Airbnb Listings in NYC:**

**The objective of the analysis is to:**

* estimate listing price based on provided information
* derive additional useful and interesting insights

**Part 1** deals with taking the existing dataset, performing data cleaning, feature engineering and running preliminary analysis. The product of this part is a data file for Machnie Learning analysis.

**Part 2 -** The Machine Learning part of the project applies machine learning algorithms to predict price of lisings based on various input variables.

**How our project differs:**

Our work differs from these related works in terms of the features used and the machine learning techniques applied. The location from where we collected data was completely different.

**Dataset:**

All the datasets we found online were either outdated or of different countries. Whereas to get the best results from our application we needed a latest dataset on rental properties and that too in Pakistan. So our only option was to collect data from Scratch

**Size of Dataset:** The dataset used in this project was collected from zameen.com and initially consisted of 2500 observations. However, after performing data cleaning using Python, BigQuery, and Excel formulas, the size of the dataset was reduced to 2000 observations.

Distribution of Classes: The data was distributed into several classes, including area, city, latitude, longitude, size of the area, number of baths, number of bedrooms, type, and rent. However, it is unclear if any class imbalances exist as no specific information regarding the distribution of classes was provided.

**Preprocessing:**

Data cleaning is a critical step in any data analysis project to ensure the accuracy and reliability of the results. For this project, several preprocessing techniques were used to clean the data, including:

1. **Removing duplicate values:** Duplicate values can cause bias in the data and lead to inaccurate results. Hence, all duplicate values were removed from the dataset.
2. **Handling missing values:** Missing values in the dataset can result in incorrect results or biased analysis. Therefore, missing values were either replaced or removed from the dataset based on the specific feature.
3. **Correcting data types:** The data types of some features were incorrect, leading to incorrect results. Hence, the data types were corrected to match the feature's nature.
4. **Removing outliers:** Outliers can significantly impact the analysis and lead to incorrect results. Hence, outliers were removed from the dataset to improve the accuracy of the results.

Overall, the preprocessing techniques used helped to improve the accuracy and reliability of the data, ensuring that the results of the analysis were valid.

**Baseline:**

Our baseline approach involves using location-based classifiers to identify rental properties that meet the specific location preferences and financial constraints of tenants. This approach will help identify properties that are within the preferred location and rent budget of the tenant. We will use location-based algorithms such as K-nearest neighbors (KNN) and collaborative filtering to implement this approach.

**Oracle:**

Since there cannot be a location that is 100% ideal for a person, we will be using humans as an oracle to determine the most suitable rental location based on majority votes. We will collect feedback from multiple human evaluators, who will assess the rental locations based on various factors such as proximity to public transportation, shopping centers, schools, and other amenities. The location that receives the most votes will be selected as the most suitable rental location.

Overall, our approach will help tenants find rental properties that meet their specific location preferences and financial constraints. The baseline approach will help narrow down the search to suitable rental properties, while the oracle approach will provide a human touch to ensure that the final selection is the most suitable rental location for the tenant.

**Main Approach:**

Our project has two main features:

1. **First,** we will suggest the user the most optimal rental locations based on the their preferred locations and price range. The preferred location can be anything from where the user want his/her rental location to be closet from. It can either be user’s office location, school, or most visited location etc
2. **Secondly,** we also offer a feature for people who are looking to rent out their property and want’s a rough estimate of the rental price of their property.

We took two main approaches for our project:

**First Approach:** Our first feature was that we had to suggest user the most optimum rental location that is closest to the user’s choice of locations and also is in range of the price suggested by the user.

For this we implemented **K-Nearest Neighbor** algorithm. We found the nearest neighbors i.e., the nearest locations by training a module on latitudes and longitudes of a location. As user can have more than one preferred location so we found out all the nearest neighbors from all preferred locations and then we take the intersection to find some common areas.

Since, we have to find the closest neighbors (locations) that are also in user’s range of price, we first filter out the areas whose prices were in range of user’s choice. Next we took the intersection of both areas (one produced by location, second produced by rent price) which gave us the most ideal location for user both in terms of vicinity and price.

**Second Approach:** To implement our second feature, we used the concepts of linear regression and random forest. Our dataset was in both numerical and categorical so we had to encode our features which were categorical like “**Areas, City”** to one-hot coding to ensure that the linear regression correctly works.

**Evaluation Metric:**

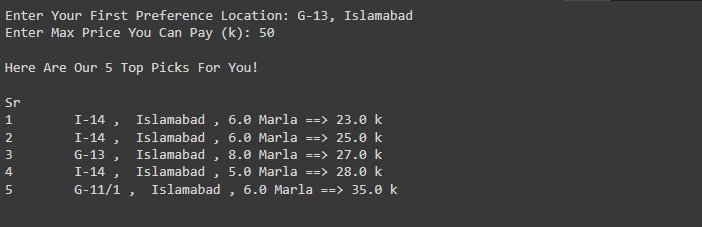
Since KNN was used for finding the nearest locations instead of classifying features into a single output point so we tried to implement our evaluation matric in the form of maps. From maps it was visually evident whether the rental location were nearest to the preferred locations or not which give us an idea how accurate our model was.

Whereas for linear regression our evaluation metrics was solely based on the accuracy score which can ensure us that our predicted location is accurate.

**Results & Analysis:**

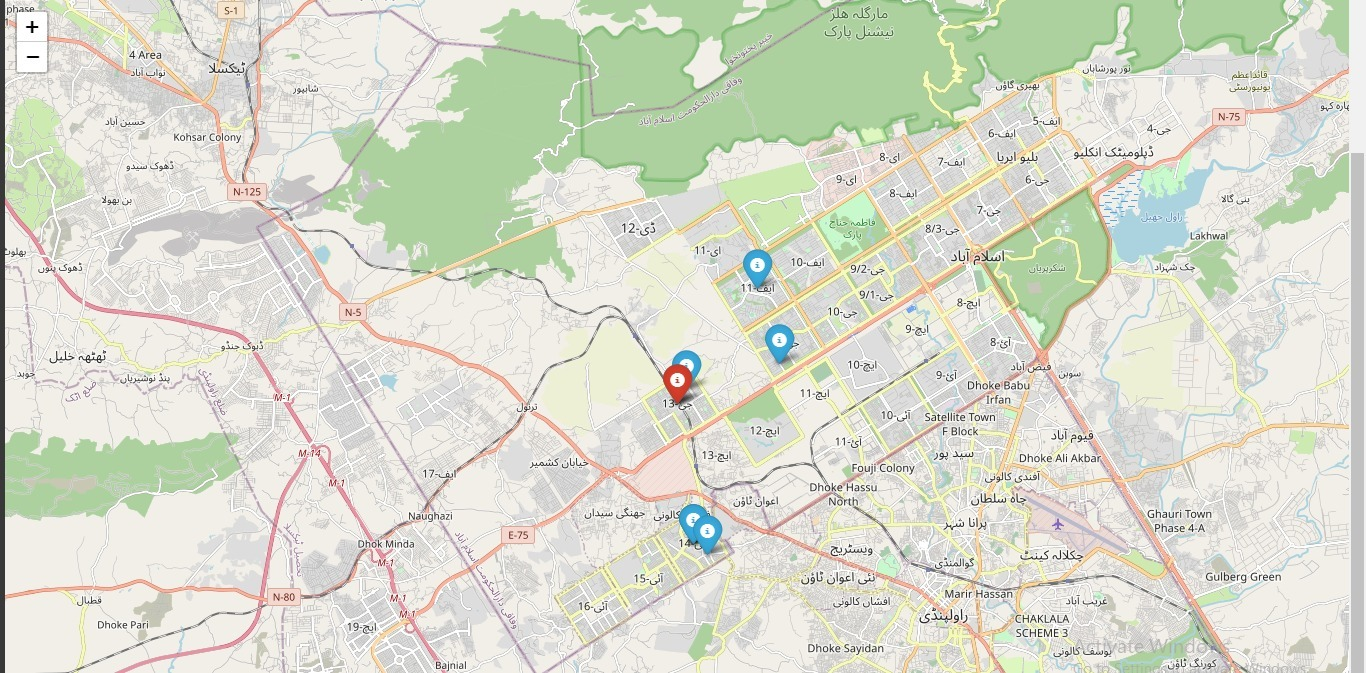
**Text based result for property prediction feature.**

Here user entered preferred location and price range and then via KNN we suggested him the 5 suitable rental locations based on his choice of location and price.



**Map based result for property prediction feature.**

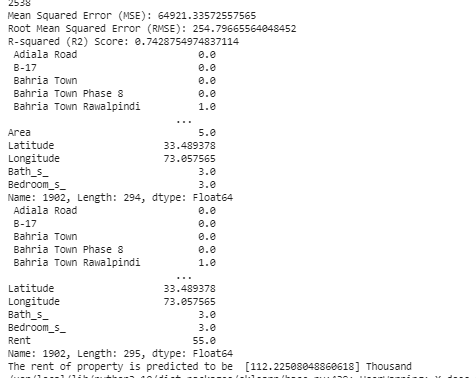
Here we visualize the coordinates of the 5 suitable rental locations and user preferred location on the map, whereas the blue icons shows suitable rental locations and red ones shows users choice of location.



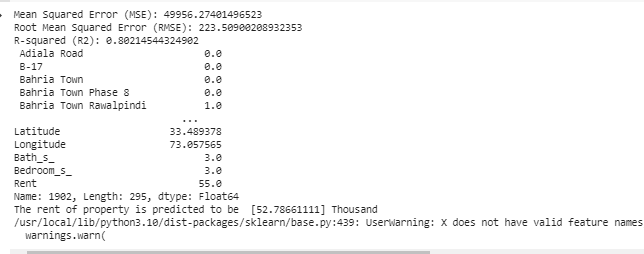
**Rental Property Prize Prediction:**

We have applied Bayesian Ridge ,Linear Regression and random forest to predict the rent of the property based on their feature .Random Forest gave us the best prediction .While training We have saved two to three rows so we can predict the rental prize on them after the model is trained.

**Bayesian Ridge:**

****

**Random Forest:**

****

**Linear Regression:**

**A screenshot of a computer

Description automatically generated with medium confidence**

**Prediction of rent On a data Row:**

**Bayesian Ridge:**

****

**Random Forest:**

****

The Random Forest network predict the rent, the rent should be 55 thousand but it is giving 52.7 thousand. That is a good estimator.

**Linear Regression:**

****

The Linear Regression predict the rent, the rent should be 110 thousand but it is giving 93 thousand.

**Future Work:**

Since we will be continuing this project as our FYP and we would be implementing it with many feature. Given below are ideas we are thinking to add.

1. **Incorporate more features:** Currently, the recommender only suggests location and price. we could consider incorporating additional features such as apartment size, number of bedrooms, and amenities (e.g., pool, gym, parking) to provide a more comprehensive recommendation to the user.
2. **Improve the accuracy of the model:** The recommender's accuracy can be improved by using more advanced algorithms such as deep learning or ensemble models. These models can learn complex patterns and relationships in the data and provide more accurate recommendations to the user.
3. **Incorporate user feedback:** The system should be designed to incorporate user feedback. Users should be able to rate and provide feedback on the recommended locations, which can be used to improve the recommendation algorithm in the future.
4. **Provide additional information:** The system can provide additional information about the recommended locations such as crime rate, school ratings, and local amenities. This can help users make an informed decision about the recommended location.

By incorporating these suggestions, the rental location recommender can provide a more comprehensive and personalized recommendation to the user, leading to increased user satisfaction and engagement.

**References:**

References to research paper that we studied:

1. Oza, H., & Srivastava, S. (2019). Predicting Airbnb Listing Prices for New York City Using Machine Learning Techniques. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 2686-2693). IEEE.
2. Liu, X., & Hu, X. (2018). Airbnb Price Prediction using Machine Learning. arXiv preprint arXiv:1812.04557.
3. Jang, Y., & Lee, J. H. (2018). Predicting the Price of Airbnb Listings Using Machine Learning. In 2018 12th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS) (pp. 500-504). IEEE.