**Project Progress Report I**

**Project Title: [Ontario rental market analysis]**

**Project Phase**: Exploratory Data Analysis (EDA), and Visualization

**Reporting Period:** [January 2024 to February, 2024]

**Project Overview:** we have large Ontario dataset which gives us very useful insights. In this project firstly we did EDA to clean dataset as we have missing values and outliers in this dataset, so we remove them. Also, we remove some columns which we didn’t need. Then, we visualize our dataset using different charts and graphs. Also, we did auto EDA to get meaningful insights from dataset.

**Accomplishments:**

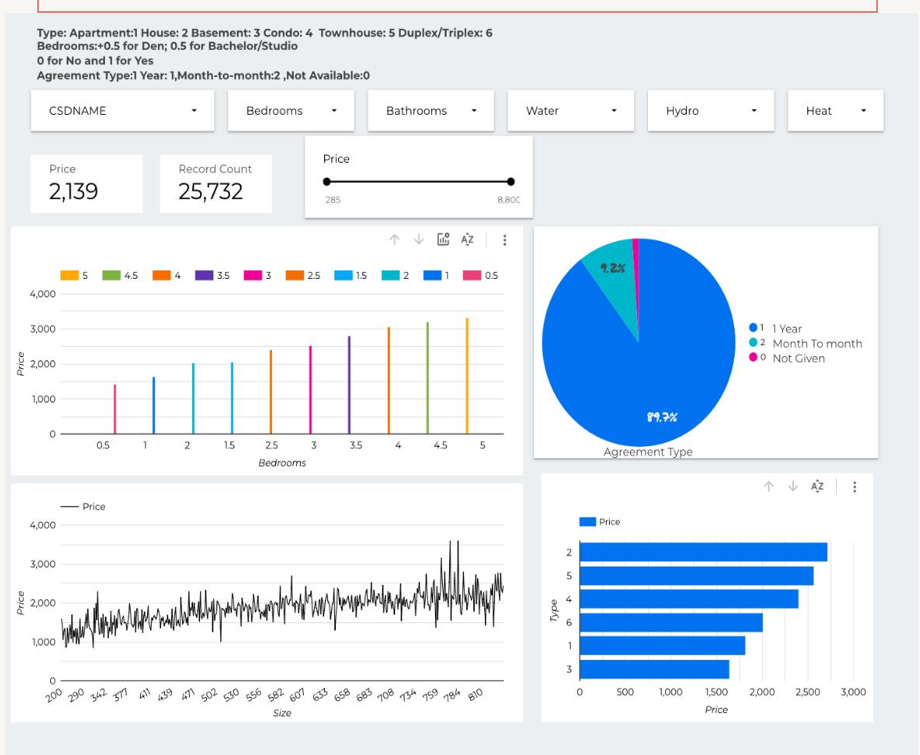
The collected Ontario rental dataset from Kijiji have around 25732 rows and 18 columns that gives detailed information about the properties. As, it has size, number of bedrooms and bathrooms, longitude, latitude, facilities like heat, electricity, and water, CSDUID, type of a property, agreement type, URL and many more. Also, we saw co relation between these columns which shows specific patterns to get insights.

**Exploratory Data Analysis (EDA) Highlights:**

To clean dataset we did EDA, firstly we remove columns like URL, Date posted from dataset as we don’t need them to visualize or get insights. Then, we change the datatype of each column to numeric or fix the categories for instance we have column bedrooms which is in categorical form, so we convert it into numeric. Also, we have missing values in this dataset, so we fill them with mean or mode as needed and we have outliers in our dataset which we remove especially in size column. As we took a close look of size column, we saw that we have size in negative form or we have very small which seems a mistake as it doesn’t relate with the type of property which considered as outlier, so we make changes in this column with our own generalization.

**Data Visualization:**

For visualization we use looker studio and make bar graphs, line graphs, pie charts and we use map to see correlation between columns and trends in our dataset. Also, we use a slider of price which we can use to adjust the price and see which type of property is more popular on particular location. Here we have one picture of our looker dashboard.



**Challenges Encountered:**

**Deal with Outliers:**

The main challenge we have to encountered was detecting outliers as its very difficult to observe data closely and detect if there is mistake or a rare value in column and then should know how to deal with it or we have to remove it.

A black background with white text

Description automatically generated

**Deal with missing values:**

Then we next problem in dataset was missing values as we have many values in this dataset which we to fill with logical values.

A screen shot of a computer code

Description automatically generated

**Deal with different dtypes**

Moreover, we have categorical features in our dataset which we have to convert to numeric to perform all tasks and to get insights.

A screenshot of a computer code

Description automatically generated

**Next Steps:**

**Data Refinement:**

Resolving Data discrepancies: To get rid of discrepancies, make sure that data types and formatting are consistent between columns.   
• To ensure proper analysis, the dataset was cleaned by addressing outliers, missing values, and data type transformations.   
Investigating Different Data Sources:   
• To improve the dataset and offer more thorough insights into Ontario rents, it was thought to integrate new data sources.   
• With a specific focus on gathering information about tiny villages in Ontario, the analysis's granularity and relevancy were improved.

**Advanced Analysis:**

After our data has been thoroughly cleansed, the following steps are involved: • Predicting rental costs in local areas around Ontario using a predictive model.   
• Applying cutting-edge analytical approaches to find hidden trends or patterns in the data and produce useful insights.   
• Setting priorities for research on the particular traits and dynamics of small-town rental markets in order to customize tactics.

**Stakeholder Engagement:**

Presented the results of the Exploratory Data Analysis (EDA) to stakeholders at a recent session.   
• Involved stakeholders, highlighting the emphasis on small community rental markets, to get their opinions and views on the analysis's findings.   
To promote focused and efficient decision-making, it was discussed how important it is to comprehend the unique demands and dynamics of tiny communities.

• To guarantee accuracy in analysis and modeling, further remove data discrepancies and clean the dataset.   
• To improve the depth of insights, investigate additional data sources, such as economic and demographic indices.

**Conclusion:**

We have been working on Kijiji's dataset of small community rental apartments in Ontario for the past month, both for exploratory data analysis (EDA) and visualization. After carefully gathering, analyzing, and exploring the data, we were able to provide insights into the types of properties, their amenities, and rental costs. We successfully shrunk the dataset for analysis in spite of challenges like managing outliers and missing values.

Going forward, we'll concentrate on improving the data and looking for other sources to support our conclusions. In addition to including stakeholders to make sure our study satisfies their objectives, we want to employ predictive modeling approaches to precisely forecast rental costs. By utilising modern analytics and stakeholder engagement, we want to enhance the small community rental market in Ontario.

**Attachments:**

[List any documents, charts, or presentations attached with the report]

**References**

Cite your sources in APA style.