Project : A Progress Report
Project Title: Small Community Housing Data

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

Reporting Period: [January/2024 to April/2024]

Project Overview:
Project is to extract insightful information from rental housing in Ontario province. To achieve the goal, we identified and observed trends, patterns, and correlations between attributes in the data that covers different property kinds, locations, and amenities to help make strategy.
To ensure consistency and integrity of the dataset, we clean and prepare our data for data mining. The process begins by applying feature engineering steps into the dataset. We identified missing values and outliers. next, categorical values are converted (eg. bedroom:2) While converting the values, we explored different encoding techniques such as one hot encoding, ordinary encoding.
Next, we start looking into the features of the dataset. We generate meaningful visualization using python and google's looker studio. Visualization helps us finding the patterns and correlation in the data.
We identified the attributes which are important in terms of renting a property. Also, we looked at different areas in Ontario based on its population. The objective is to see what is happening in the small communities in Ontario. We considered areas where population is less than 10,000 as a small community.

Accomplishments:

Explain:
The Ontario rental dataset from Kijiji is a comprehensive collection of information regarding rental properties across the region. It covers essential details such as the number of bedrooms, property types, size dimensions, rental prices, amenities, location specifics, and rental agreement terms. This dataset is a valuable resource for individuals and property managers who are seeking rental accommodations. Potential renters can utilize the dataset to assess available rental options based on their preferences and needs. Meanwhile, property managers can leverage the data to understand market trends, optimize pricing strategies, and tailor property offerings to meet tenant demands effectively. Comprehensive nature and diverse range of data points make the data valuable tool for navigating the dynamic rental market landscape of the region.

Exploratory Data Analysis (EDA) Highlights:
Explain how you:
In our analysis, we identified issues in the bedroom, bathroom, and size columns. Functions like pd.to_numeric () ensure uniformity of data types. Data inconsistencies, including missing values, are resolved using imputation techniques to maintain dataset integrity. Specifically, for the size column, we used string manipulation functions such as str. replace () to clean non-standardized formats and convert them into a numeric format for analysis. Our Exploratory Data Analysis (EDA) involved
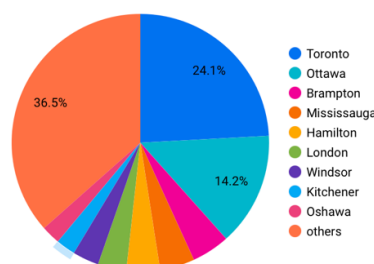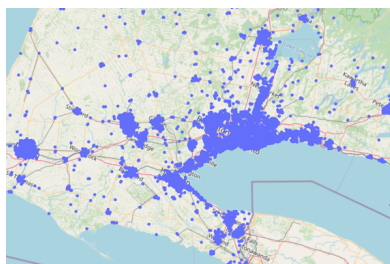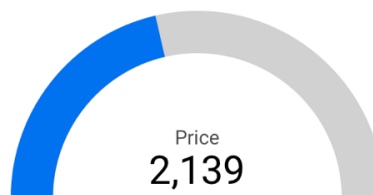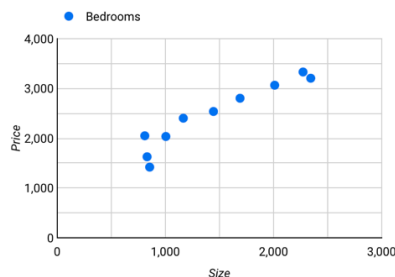
selecting small communities based on population specific criteria and comprehensive examinations of various attributes. Through visualization techniques, we identified key trends, outliers, and patterns within the data, providing insights into population demographics, housing preferences, and rental prices. Also, we explored count, frequency, and price distributions. This allows a deeper understanding of rental property and market dynamics in small communities in Ontario.

Data Visualization:

We used Google's Looker Studio as visualization tool to assist us with our visualizations because it allows us to play with the graph to a greater extent than we could with Code, and it's simple to use even for non-programmers. We generate interactive graphs and tables by utilizing Looker's wide range of visualization options. We ensured clarity and usability through iterative refinement based on feedback from stakeholders. Bar charts that display the total number of rentals in an area, pi-charts that indicate which properties are the largest and account out of 100% for the rentals in an area, geographical maps that display properties with specifics, and price gauges that display the average cost of a rental in any area. It varies depending on the area you choose to view, and a histogram that illustrates the relationship between size, bedrooms, and price is provided are some examples of the visualizations available.

Looker Dashboard - https://lookerstudio.google.com/reporting/67ad996b-7378-44e7-bd22-28b263068a47/page/O5lpD

Some Examples of Visualizations:

Challenges Encountered:

Data Quality:

```
lower_limit = 400
upper_limit = 9000
obj_df.loc[(obj_df['Size'] < lower_limit) | (obj_df['Size'] > upper_limit) | (obj_df['Size'] < 0), 'Size'] = np.nan

# Replace NaN values with the mean of the column
obj_df['Size'].fillna(obj_df['Size'].mean(),inplace=True)
```

There are some outliers in some columns we establish upper and lower limits to identify values outside the expected range. Any values exceeding these limits are considered outliers and replaced with NaN values. Subsequently, we compute the mean of the column and replace the NaN values with this mean value.

```
[ ]   str_num_conv = {"Bedrooms":{"Bedrooms: 2": 2,"Bedrooms: 1": 1,"Bedrooms: 3": 3,"Bedrooms: 4": 4 ,
                        "Bedrooms: 1 + Den": 1.5 ,"Bedrooms: Bachelor/Studio": 0.5 ,
                        "Bedrooms: 2 + Den": 2.5 ,"Bedrooms: 3 + Den": 3.5 ,"Bedrooms: 4 + Den": 4.5 ,"Bedrooms: 5+": 5},
      }
```

Next, we encounter an issue in the bedroom section, where the data contains a mix of numeric and string values. To address this issue, we utilize a dictionary to convert the bedroom section into a consistent numeric format.

Resource Constraints: We did not have any.

Data Visualizations:

We haven't had many issues, but the one we did have was that while all of the graphs are significant and related to one another, we are unable to visualize them all, so it was difficult to choose which to display.

Next Steps:

Data Refinement:

Address data inconsistencies and clean the dataset for accurate analysis.
Explore additional data sources to enhance the richness of collected information.

Advanced Analysis:

Carry out research and elaborate on how to:
Implement advanced analytical techniques such as predictive modeling to forecast price trends and any other classification you may think appropriate.
Investigate correlations between population and other parameters.

Stakeholder Engagement:

Give details on the last session with stakeholders to present EDA findings and gather their insights.

Provide suggestions from the stakeholder here.
Mention if more information and what info is required from clients.

Conclusion:
In conclusion, our project successfully analyzed the Ontario rental housing dataset from Kijiji, providing valuable insights for strategic decision-making. Through data cleaning and exploratory analysis, we identified trends, patterns, and correlations, particularly in small communities. Challenges such as outlier handling were addressed, and stakeholder engagement facilitated valuable feedback. Moving forward, refining the dataset and exploring advanced analytical techniques will enhance our insights, contributing to informed decision-making in the real estate sector.

Attachments:

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

References
pandas - Python Data Analysis Library. (n.d.). https://pandas.pydata.org/
Looker Studio Overview. (n.d.). https://lookerstudio.google.com/navigation/reporting