Name: Hassan Adil
Cohort: DA-07 Green

atom**camp**

EXECUTIVE REPORT

1. Purpose:

The primary purpose of this data analysis project is to conduct an exploratory data analysis (EDA) of real estate listings from Zameen.com in Pakistan. The core problem statement addressed is: "What drives property prices in Pakistan?" By thoroughly analyzing the provided dataset the project aims to identify key factors influencing property values, understand market trends and uncover relationships between various property attributes and their prices.

The overall objective is to provide actionable insights that can assist stakeholders, including real estate agents, buyers, sellers, and developers, in making informed decisions within the Pakistani real estate market.

2. Significance:
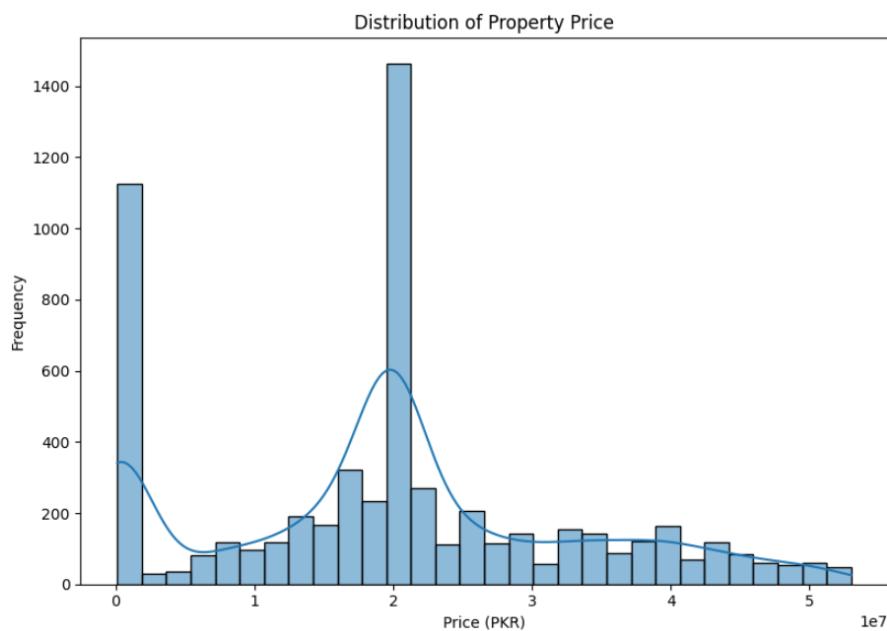
- Offer a clearer picture of fair market values, enabling more informed negotiations and investment decisions.
- Equip agents with insights into critical pricing factors helping them better advise clients and streamline their sales processes.
- Highlight lucrative areas and property types, assisting in strategic planning for new developments and optimizing investment portfolios.
- Introduce a standardized approach to understanding property valuation, potentially fostering greater trust and efficiency in the market.

3. Key Findings/ Insights:
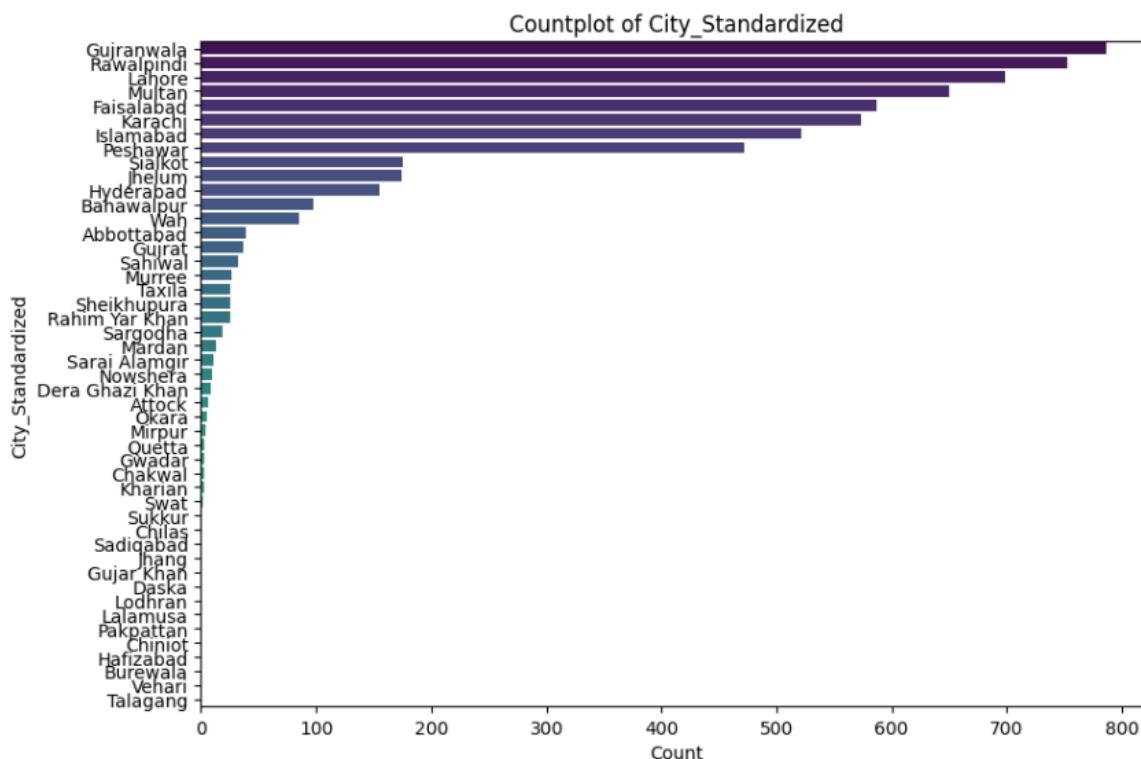
- The initial dataset presented significant challenges, including numerous missing values in critical columns 'Built in year', 'Kitchens', 'Description' inconsistent data types 'Price', 'Area', 'Bedrooms', 'Bathrooms' initially as objects, and varied categorical entries 'City', 'Type' requiring **extensive standardization** through fuzzy matching.
- The 'Price' column exhibited a strong **right-skewed distribution**, with a median price of 19.9 million PKR significantly lower than the mean, indicating the presence of high-value outliers and necessitating median imputation for missing values.
- Applying the IQR method for **outlier removal** on 'Price', 'Built in year', and 'Kitchens' resulted in a substantial reduction of approximately 6,955 rows from the dataset from 12,997 to 6,042 rows. This refined the dataset to better represent typical market properties, especially for price analysis.
- 'House' and 'Flat' emerged as the **most common property types** highlighting their prominence in the Pakistani real estate market.
- Lahore, Karachi, and Islamabad were identified as the cities with the highest number of property listings, underscoring their status as major real estate hubs.
- The creation of 'Age_of_Property' (derived from 'Built in year') and 'Price_per_SqFt' (calculated from 'Price' and 'Area') proved crucial for gaining deeper insights into property valuation relative to age and size.

- Property prices vary significantly across different cities. Major metropolitan areas like Lahore and Karachi generally exhibit **higher average prices** and a wider price range compared to smaller cities.
- 'Bedrooms' and 'Bathrooms' show only a **weak positive correlation** with 'Price'. This suggests that while these features contribute, they are not the primary drivers of price. Factors like location, property condition, and amenities likely play a more substantial role.
- There is an observable trend indicating that newer properties (lower 'Age_of_Property') tend to have **higher prices**. This highlights a market preference for modern construction and potentially better amenities or structural integrity.
- The calculated 'Price_per_SqFt' is a critical metric for comparing property values on a **standardized basis**. Its variability, even within similar property types or cities, indicates that specific micro-locations, finishes and unique features heavily influence perceived value.
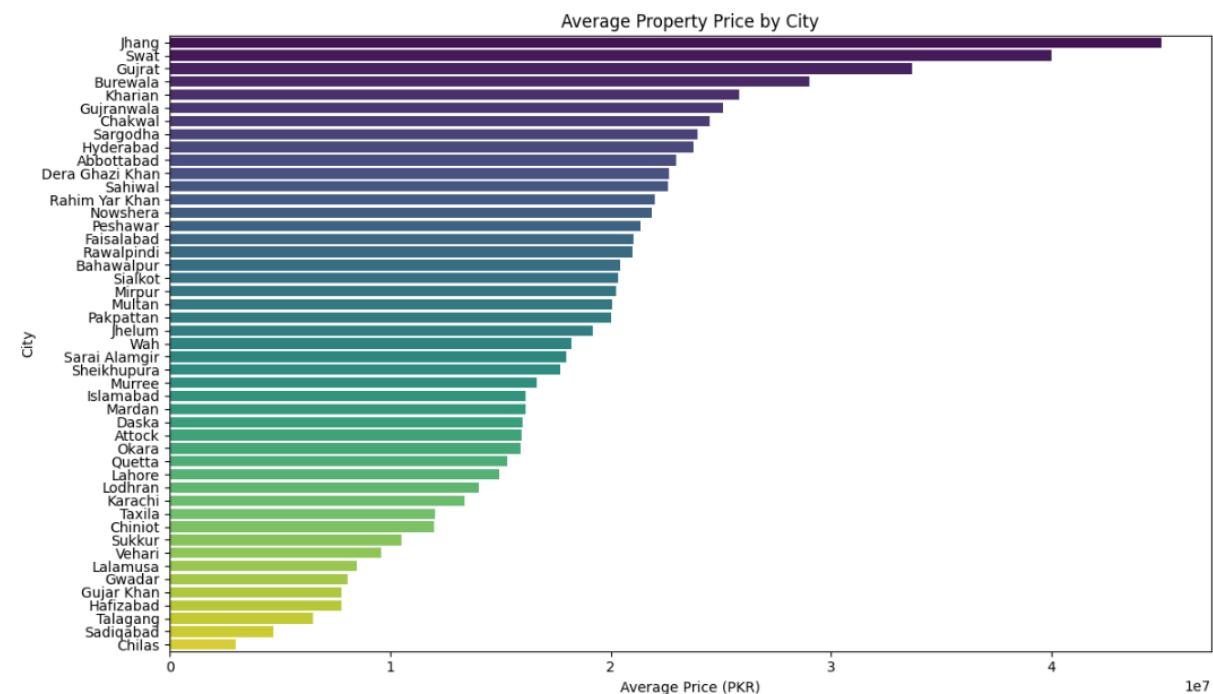
4. **Key Visuals:**

Insight: The 'tail' of the distribution extends significantly to the right, signifying the presence of a few properties with exceptionally high prices. These are likely luxury properties or those in prime locations.

Insight: Gujranwala is the city with the highest number of property listings which typically correspond to the major real estate hubs or areas with the most active market.

Insight: It clearly highlights which cities command the highest average prices (Jhang) and which are more affordable (Chilas).

5. **Conclusion/ Price Influencers:**

- Property prices are heavily skewed towards lower values with the median price being a more accurate representation than the mean.
- Prices vary significantly by city with major metropolitan areas like Lahore and Karachi showing higher average prices.
- 'House' and 'Flat' dominate the market with diverse pricing while 'Penthouses' and 'Farm Houses' command premium prices.
- 'Bedrooms' and 'Bathrooms' showed only a weak positive correlation with 'Price', suggesting other factors (location, condition, amenities) are more influential.
- Newer properties lower 'Age_of_Property' tend to have higher prices, indicating a market preference for modern construction.
- 'Price_per_SqFt' is a critical metric for standardized comparison highlighting the influence of micro-locations and unique features on perceived value.