atomcamp

Real Estate Market Insights: An Exploratory Analysis of Zameen.com Listings in Pakistan

Project Objective:

The objective of this project is to **analyze property listings on Zameen.com** and generate actionable insights to support real estate investment decisions. While **guiding questions and directions are provided**, it's **necessary to apply your own analytical skills** to explore the data, interpret trends and provide meaningful insights.

The project aims to:

- **Identify pricing trends** across different property types and locations.
- **Compare neighborhoods** based on property value, demand and amenities.
- **Assess listing quality** to determine reliability and completeness.
- **Provide actionable insights** that can help investors make informed decisions.

This project **clearly requires participants to think independently** use their analytical knowledge and **extract insights on provided guidance** demonstrating practical data interpretation and visualization skills.

Deliverables:

1. **A Colab/ Jupyter Notebook** with end-to-end analysis
2. **A GitHub repository** with code, README, and insights summary (this is optional)
3. **An executive report or PDF** summarizing key findings and visualizations

Rubric:

Code with comments: 50
Report: 50

atomcamp

Dataset Description:

From your file, we have the following features:

Dataset link

- Listing titles, location (city, area)
- Price
- Property type
- Area
- Number of beds/baths
- Date of posting
- Description text

Project Sections

1. Problem Statement

●Define the business question: What drives property prices in Pakistan?

2. Data Understanding & Preprocessing

- Use: pandas, numpy
- Tasks:

    View basic info

    Handle duplicates

    Clean messy columns (e.g., remove commas, PKR symbols)

Convert area units (marla/kanal/sqft → uniform sqft), if required

3. Missing Values Treatment

- Tasks:

  Identify missing values using .isnull().sum()

  Impute with mode/mean/forward-fill for categorical/numerical

  Justify treatment

4. Data Cleaning & Consistency

- Use: FuzzyWuzzy, custom functions
- Tasks:

  Standardize city names using Fuzzy Matching

  Detect and correct inconsistent property types

  Remove outliers using IQR/Z-score

5. Feature Engineering

- Create new features to drive more insightful analysis

6. Univariate & Bivariate Analysis

- Use: matplotlib, seaborn, visualization library
- Tasks:

  Plot distributions of prices, area

Correlation heatmap

Violin plots / box plots to compare prices by city, property type, bedrooms

7. Insights & Recommendations

- Use: domain logic + stats

9. Conclusion & Next Steps

- Summary of learnings
- Suggestions for stakeholders (investors)

GitHub Structure Suggestion:

```
zameen-eda-project/
├── data/
│   └── Zameen.com.xlsx
├── notebook/
│   └── zameen_eda_final.ipynb
├── outputs/
│   └── visuals/
├── report/
│   └── Zameen_EDA_Insights_Report.pdf
└── README.md
```