Take-Home Assignment: Data Understanding, EDA & FastAPI

Overview

As part of our interview process, we are sharing a short take-home assignment focused on data understanding, exploratory data analysis (EDA), and implementing a minimal FastAPI application.

Dataset

You will receive:

- Mumbai_Property_Data.xlsx

Objective

This assignment evaluates:

- Data understanding

- Insight generation via EDA

- A basic ML model exposed through FastAPI

Tasks

1. Data Understanding

Explain:

- What each column represents

- Numerical vs categorical features

- Target variable chosen and why

- Data quality issues or assumptions

2. Exploratory Data Analysis (EDA)

- Missing values, duplicates, outliers

- Univariate and bivariate analysis

- Insights on locality pricing, area vs price, suspicious values

3. Minimal ML Model

- Predict price or price_per_sqft

- Use one model

- Basic preprocessing

- Show one metric (R^2 or RMSE)

4. Minimal FastAPI

Endpoint:

POST /predict

Input:

```
{
  "locality": "Andheri West",
  "area_sqft": 900,
  "bedrooms": 2,
  "bathrooms": 2,
  "furnishing": "Semi-Furnished"
}
```

Output:

```
{
```

"predicted_price": <numeric value>

}

Submission Instructions

- Public GitHub repository
- Expected structure:

eda_and_model.ipynb

main.py

README.md

requirements.txt (optional)

README Expectations

- Dataset understanding
- EDA insights
- Target variable
- Model and metric
- Steps to run FastAPI

Time & Policy

- Suggested time: 4–5 hours
- Submission window: 48 hours
- Partial submissions allowed

What We Don’t Expect

- Perfect accuracy
- Advanced ML

- Over-engineered APIs

Best regards,

HR