"Analysis of Housing Prices Based on Neighbourhood Factors Using SQL"

This project analyzes how various neighborhood and environmental factors affect housing prices. Using SQL, we explore data on crime rates, pollution, distance to employment hubs, and education quality to understand their impact on median house prices.

The dataset contains **511 records** and **13 attributes**, including:

- Crime_Rate: Crime incidents per capita
- Industrial_Area_percentage: % of area used for industry
- air_pollution_index: Pollution level
- average_room_per_house: Average number of rooms per dwelling
- Student_Teacher_Ratio: Education quality indicator
- Median_House_Price_Lakhs: Target variable representing median housing prices

Tools Used

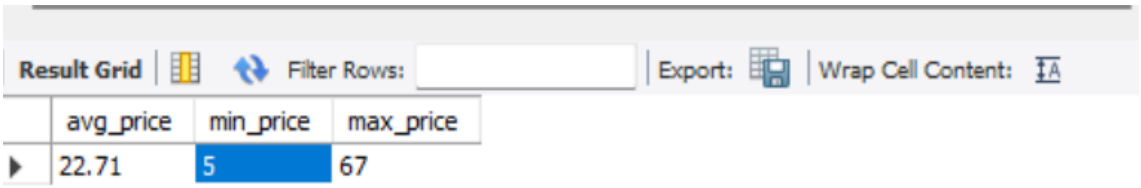
- Database: MySQL / PostgreSQL / SQLite (whichever you used)
- Language: SQL
- Environment: MySQL Workbench / VS Code / DBeaver, etc.
- File: data.csv (imported as table name bostonhousing)

Here are my SQL queries

- **Query 1: Average and Range of House Prices**

```
SELECT
  ROUND(AVG(Median_House_Price_Lakhs), 2) AS avg_price,
  ROUND(MIN(Median_House_Price_Lakhs), 2) AS min_price,
  ROUND(MAX(Median_House_Price_Lakhs), 2) AS max_price
FROM bostonhousing;
```

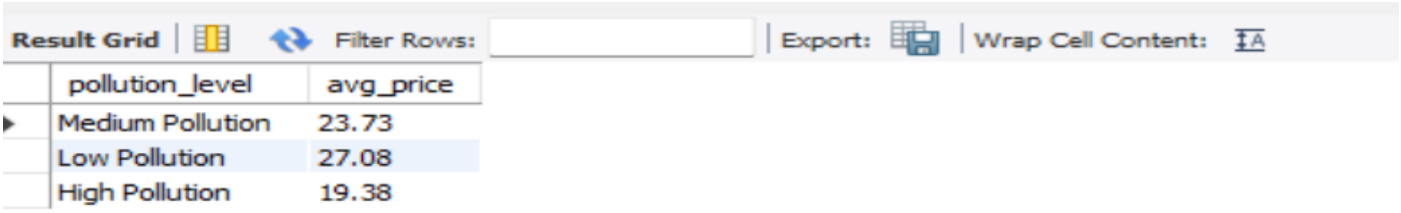
Output :



- Effect of Pollution on Housing Prices

```
SELECT
  CASE
    WHEN air_pollution_index < 0.45 THEN 'Low Pollution'
    WHEN air_pollution_index BETWEEN 0.45 AND 0.55 THEN 'Medium Pollution'
    ELSE 'High Pollution'
  END AS pollution_level,
  ROUND(AVG(Median_House_Price_Lakhs), 2) AS avg_price
FROM bostonhousing
GROUP BY pollution_level;
```

Output :



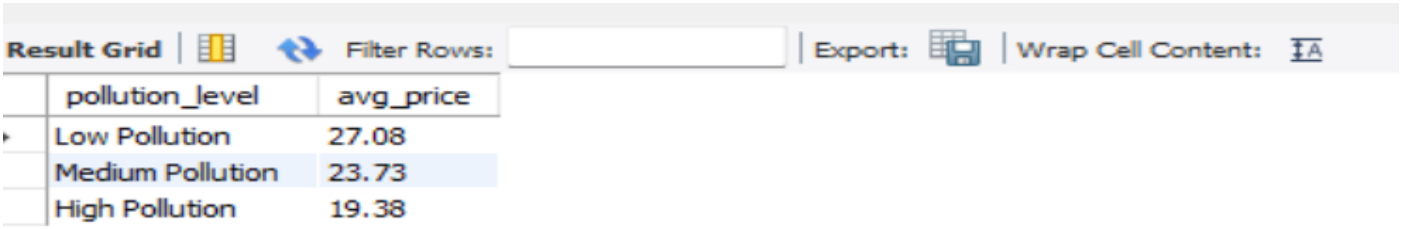
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

pollution_level	avg_price
Medium Pollution	23.73
Low Pollution	27.08
High Pollution	19.38

- **Effect of Air Pollution on Housing Prices**

```
SELECT
  CASE
    WHEN air_pollution_index < 0.45 THEN 'Low Pollution'
    WHEN air_pollution_index BETWEEN 0.45 AND 0.55 THEN 'Medium Pollution'
    ELSE 'High Pollution'
  END AS pollution_level,
  ROUND(AVG(Median_House_Price_Lakhs), 2) AS avg_price
FROM bostonhousing
GROUP BY pollution_level
ORDER BY avg_price DESC;
```

Output :



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

pollution_level	avg_price
Low Pollution	27.08
Medium Pollution	23.73
High Pollution	19.38

-- Multi-Factor Analysis (Crime + Pollution)

```
SELECT
  CASE
    WHEN Crime_Rate < 0.05 THEN 'Low Crime' ELSE 'High Crime' END AS crime_category,
  CASE
    WHEN air_pollution_index < 0.5 THEN 'Clean Air' ELSE 'Polluted' END AS pollution_category,
  ROUND(AVG(Median_House_Price_Lakhs), 2) AS avg_price
FROM bostonhousing
GROUP BY crime_category, pollution_category
ORDER BY avg_price DESC;
```

Output :

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

crime_category	pollution_category	avg_price
Low Crime	Clean Air	28.58
High Crime	Clean Air	25.73
High Crime	Polluted	20.36
Low Crime	Polluted	19.67

Key Insights / Findings

- Low crime rate areas have **higher average house prices**.
- Houses closer to employment hubs are **more expensive**.
- Better student-teacher ratios (education quality) **positively influence** housing prices.
- Air pollution and crime are **negatively correlated** with property values.

Conclusion

The SQL analysis reveals that housing prices are strongly affected by environmental and social factors. Safer, cleaner, and better-educated neighborhoods tend to have higher property values. SQL was effective for uncovering patterns and relationships in real estate data.

Chart 1: Shows how house price changes with crime rate.

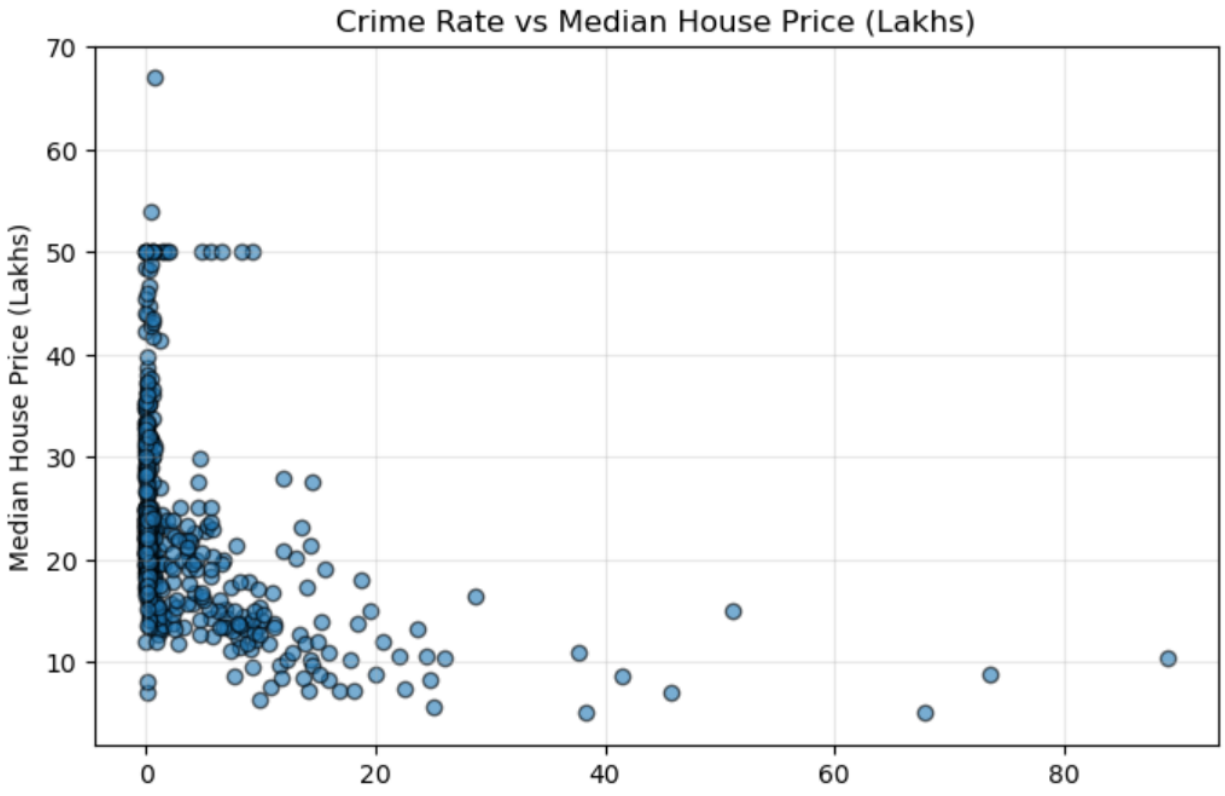


Chart 2: Compares average house prices across pollution levels

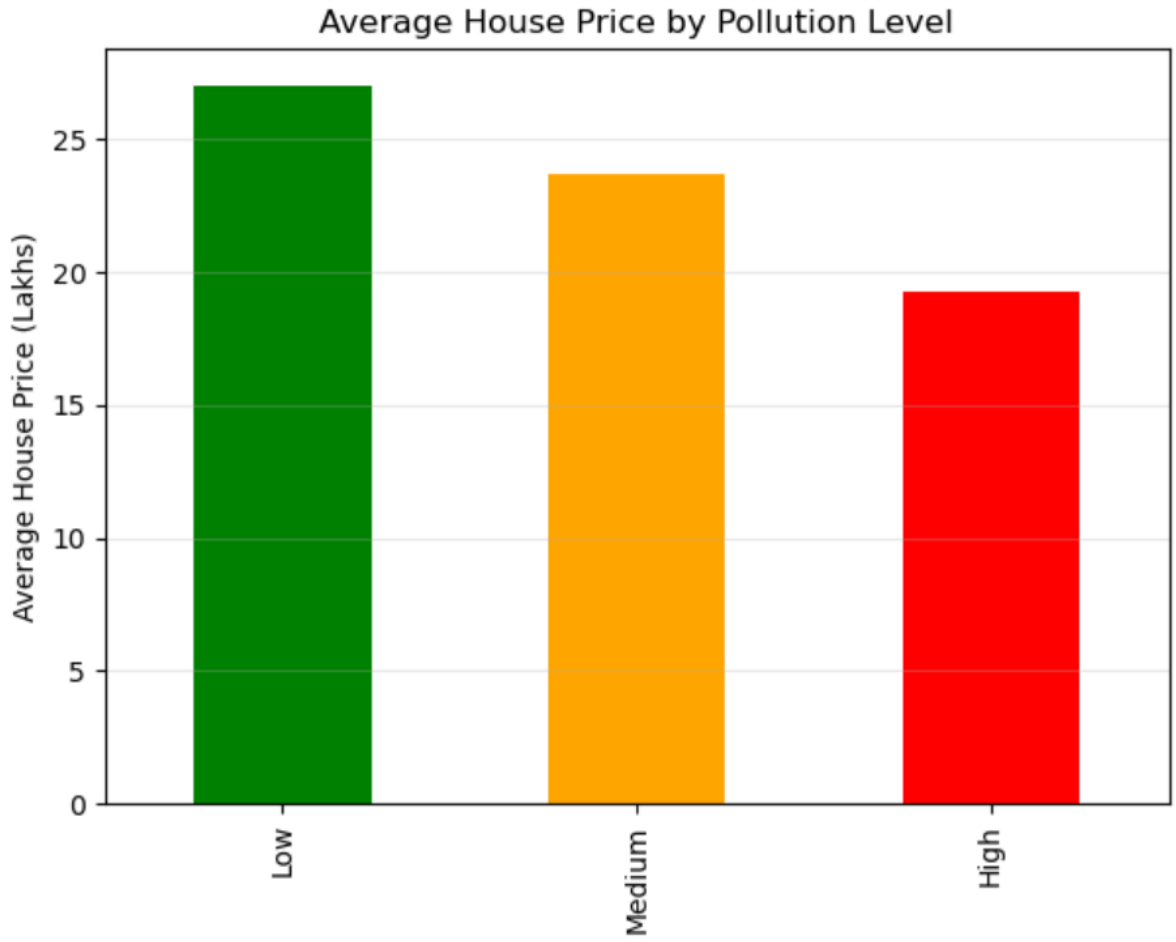


Chart 3: Displays the overall income group distribution in your dataset.