



Group 3 Project 3: Yau Shu Wong, Rock David Adeline, Fatema Girnary, and Lisa Shimano

# Perth Metro Area Real Estate Analysis



Extract, Transform, Load (ETL)



Impact of structural features on House Price



Parameters that impact House Price



HTML Visualisation

# Ethical Considerations



## Privacy and Data Protection

No personal or sensitive information of individuals (such as property owners or students) is included or can be inferred from the data.



## Bias and Fairness

The datasets are representative of the broader population and do not disproportionately focus on specific demographics or regions.



## Transparency and Accountability

The data is used only for the intended purpose of analysing factors influencing property purchase decisions.



## Informed Consent and Permissions

Consent was obtained for the specific purpose of the analysis - WA Police Force was contacted to seek permission to scrape their website for crime statistics.

Using Publicly available data



## Accuracy and Reliability

The results do not inadvertently disadvantage any particular group, such as low-income families or minority communities.





# E xtract

## Data Extraction:

- Sourced data from CSV & Excel files, and Web scraping
- Minimal review and cleansing - remove irrelevant data



# T ransform

## Data Transformation:

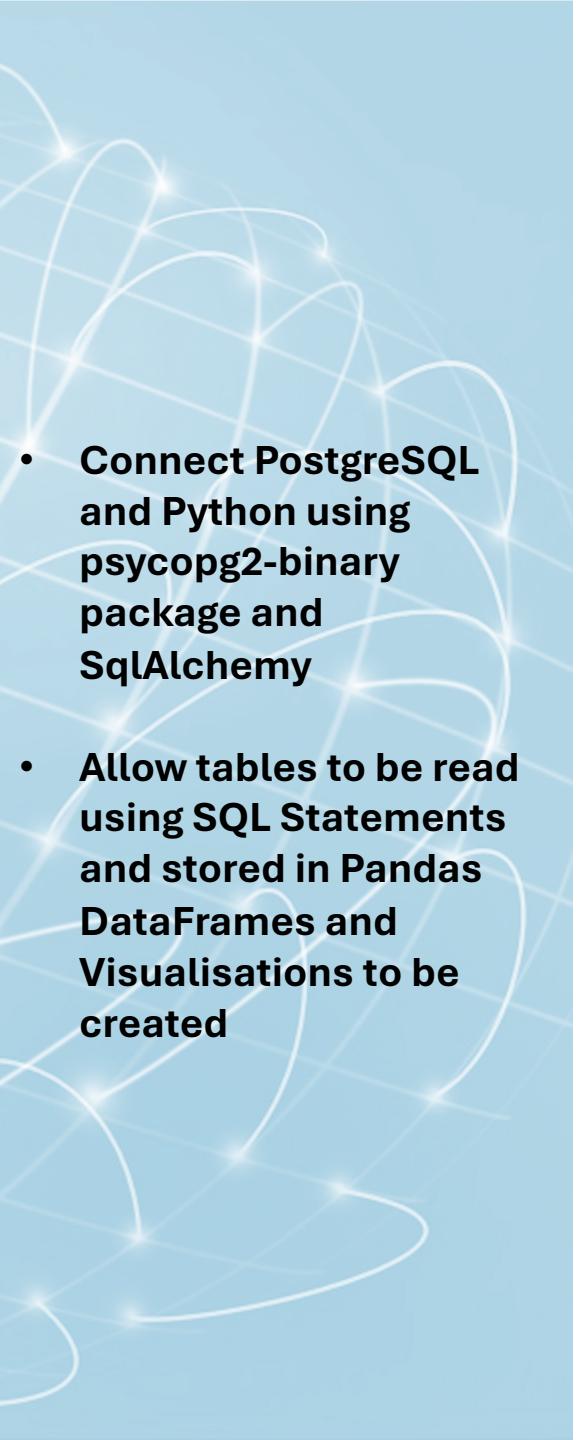
- Data cleansing - removing duplicates & correcting errors
- Data filtering - selecting relevant data and removing redundant data
- Data validation – check completeness and source missing data
- Transform to align on key columns – school names & suburb



# Analyzing Data L oad

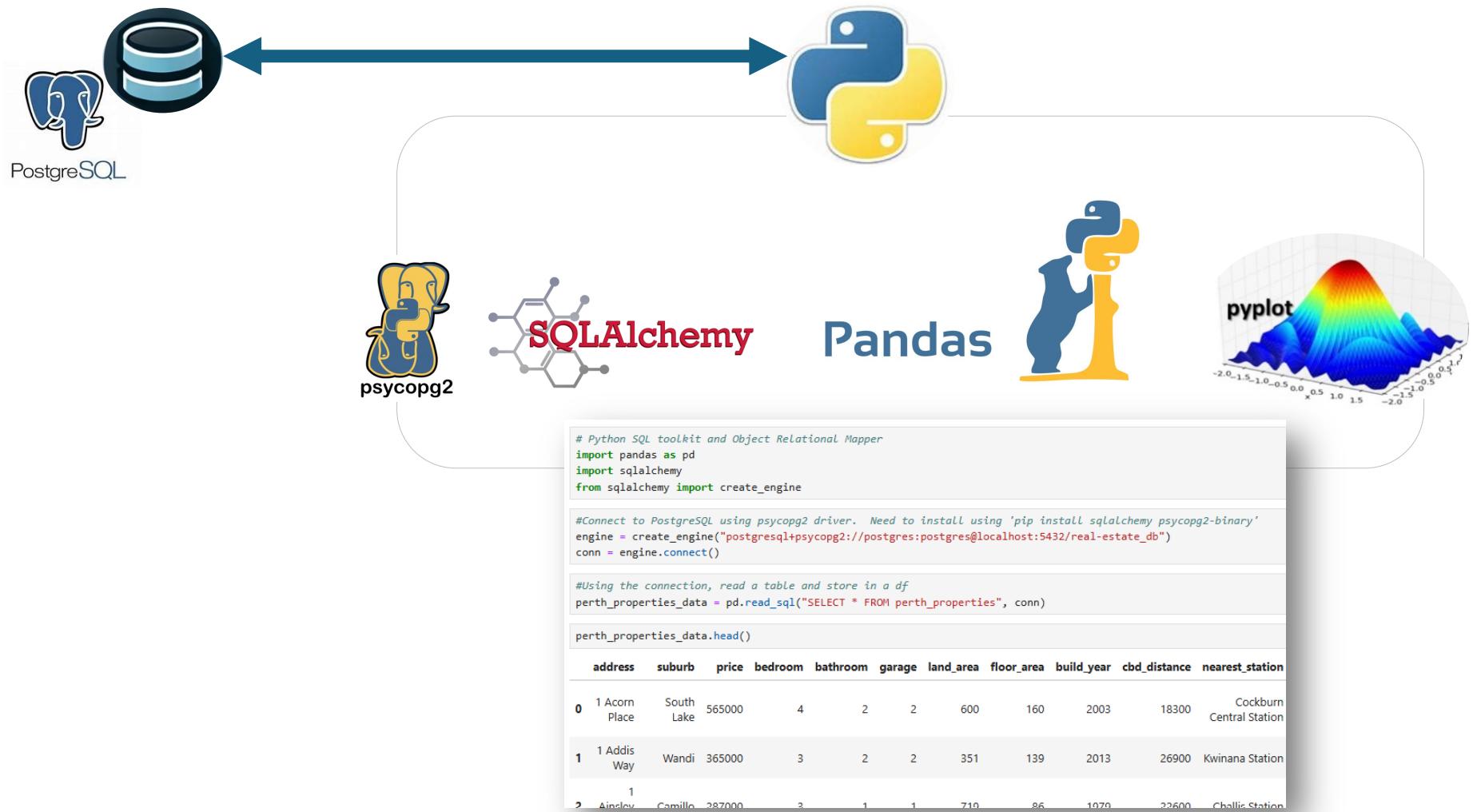
## Data Loading:

- PostgreSQL – create, update database, schema & tables
- PostgreSQL – insert Data via Import
- SQL scripts – SELECT & JOINS to get the correct set

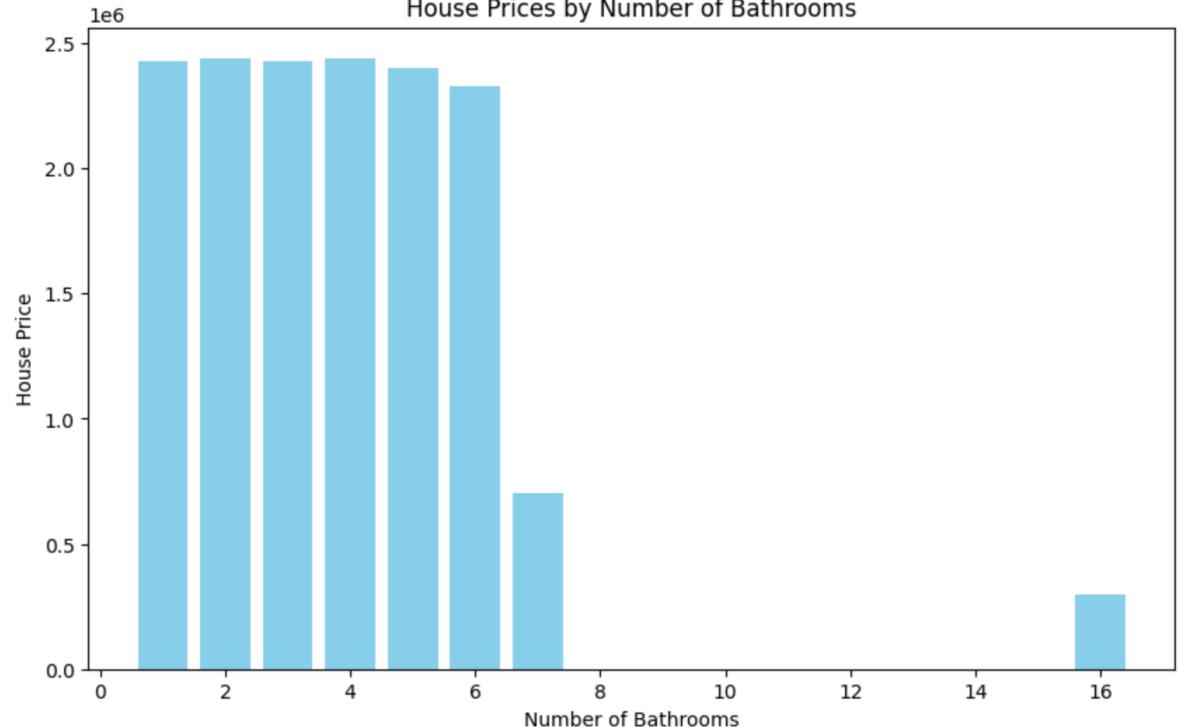


- Connect PostgreSQL and Python using **psycopg2-binary** package and **SqlAlchemy**
- Allow tables to be read using **SQL Statements** and stored in **Pandas** **DataFrames** and **Visualisations** to be created

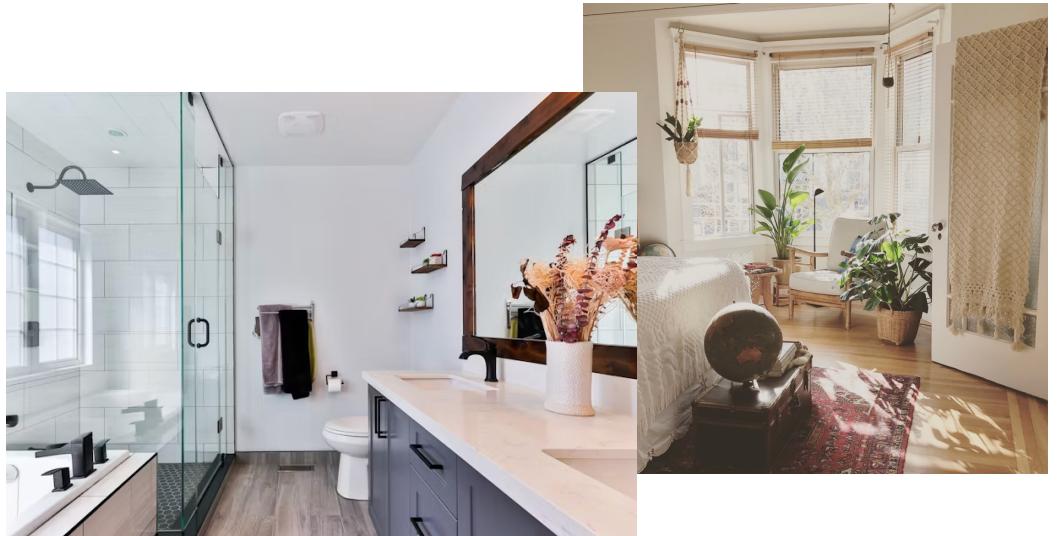
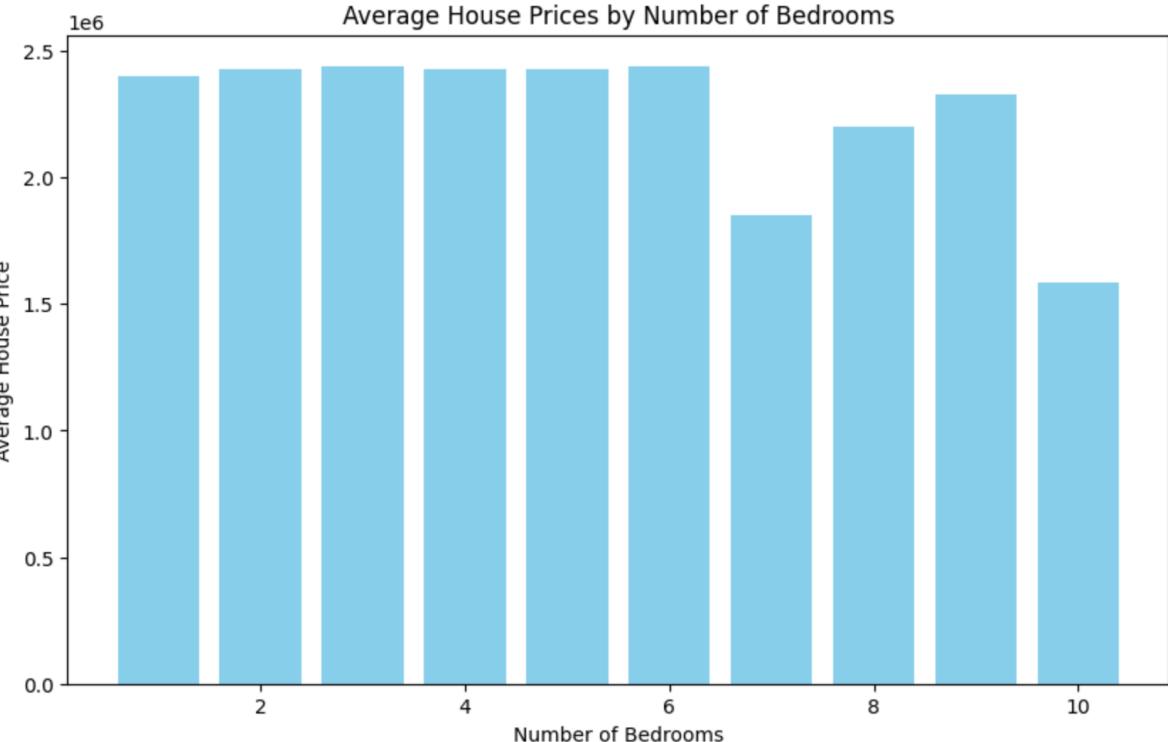
# Connecting PostgreSQL to Python



### House Prices by Number of Bathrooms



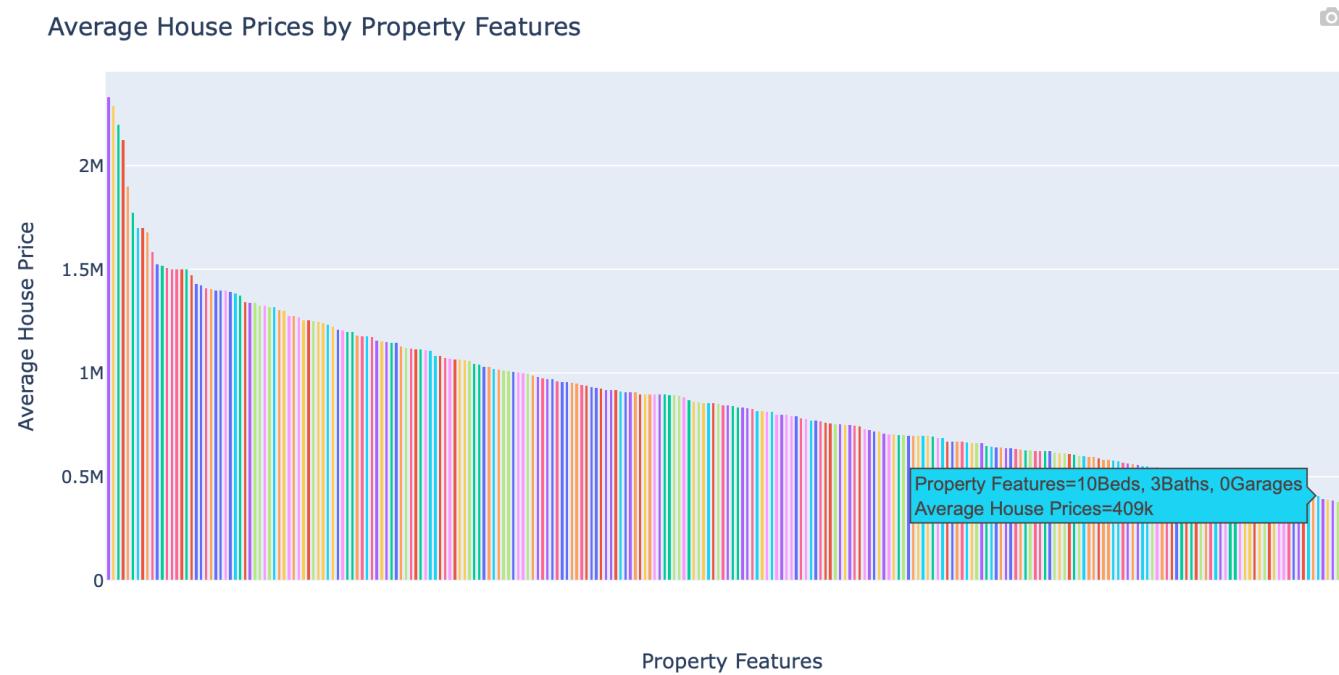
### Average House Prices by Number of Bedrooms



# Average House Prices by Property Features

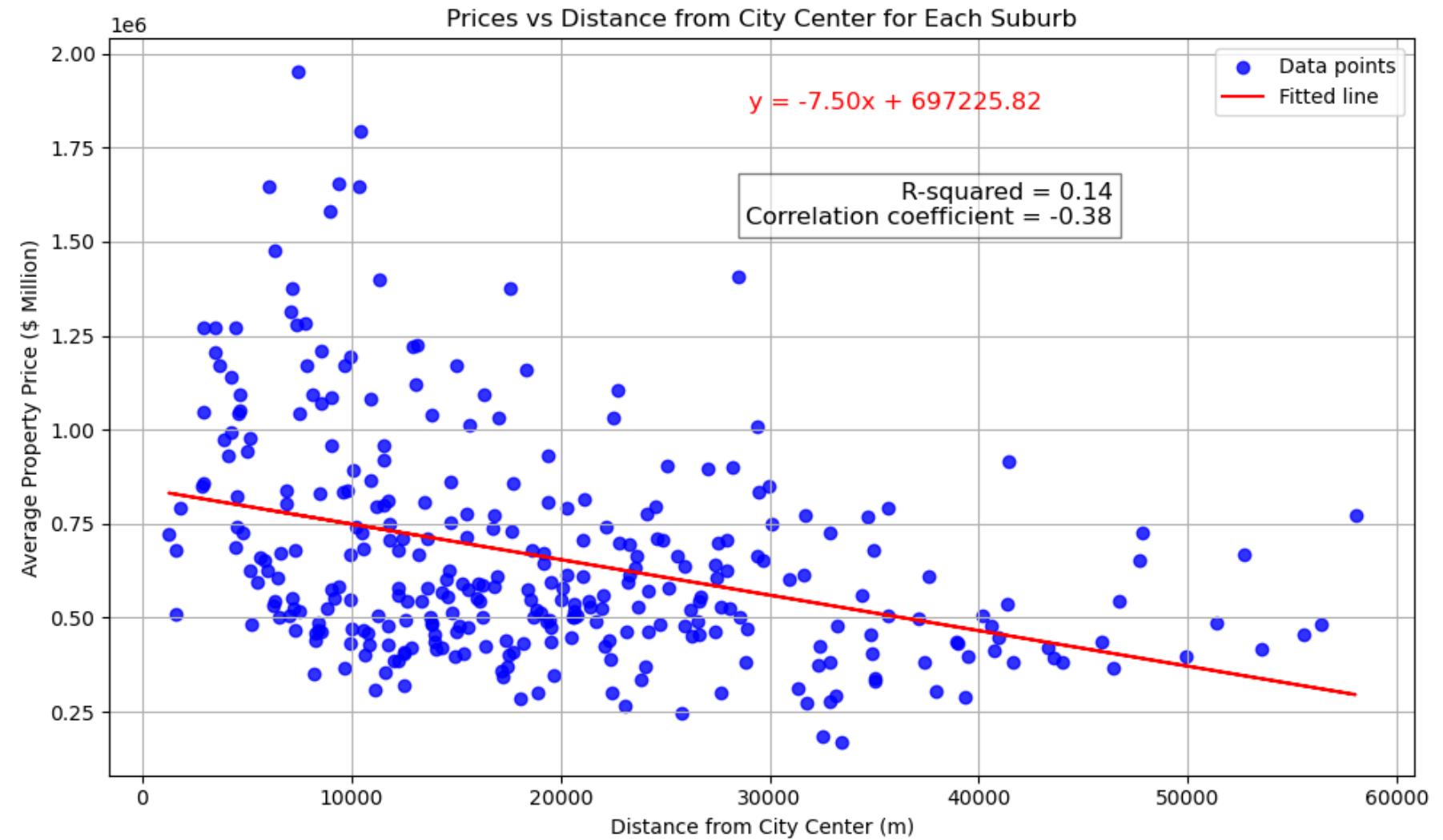


- Property Features
- 1Beds, 1Baths, 0Garages
  - 1Beds, 1Baths, 1Garage
  - 1Beds, 1Baths, 2Garages
  - 1Beds, 1Baths, 3Garages
  - 1Beds, 1Baths, 4Garages
  - 1Beds, 1Baths, 6Garages
  - 1Beds, 2Baths, 0Garages
  - 1Beds, 2Baths, 2Garages
  - 1Beds, 3Baths, 0Garages
  - 1Beds, 3Baths, 2Garages
  - 1Beds, 7Baths, 0Garages
  - 2Beds, 1Baths, 0Garages
  - 2Beds, 1Baths, 1Garage
  - 2Beds, 1Baths, 2Garages
  - 2Beds, 1Baths, 3Garages
  - 2Beds, 1Baths, 4Garages
  - 2Beds, 1Baths, 5Garages



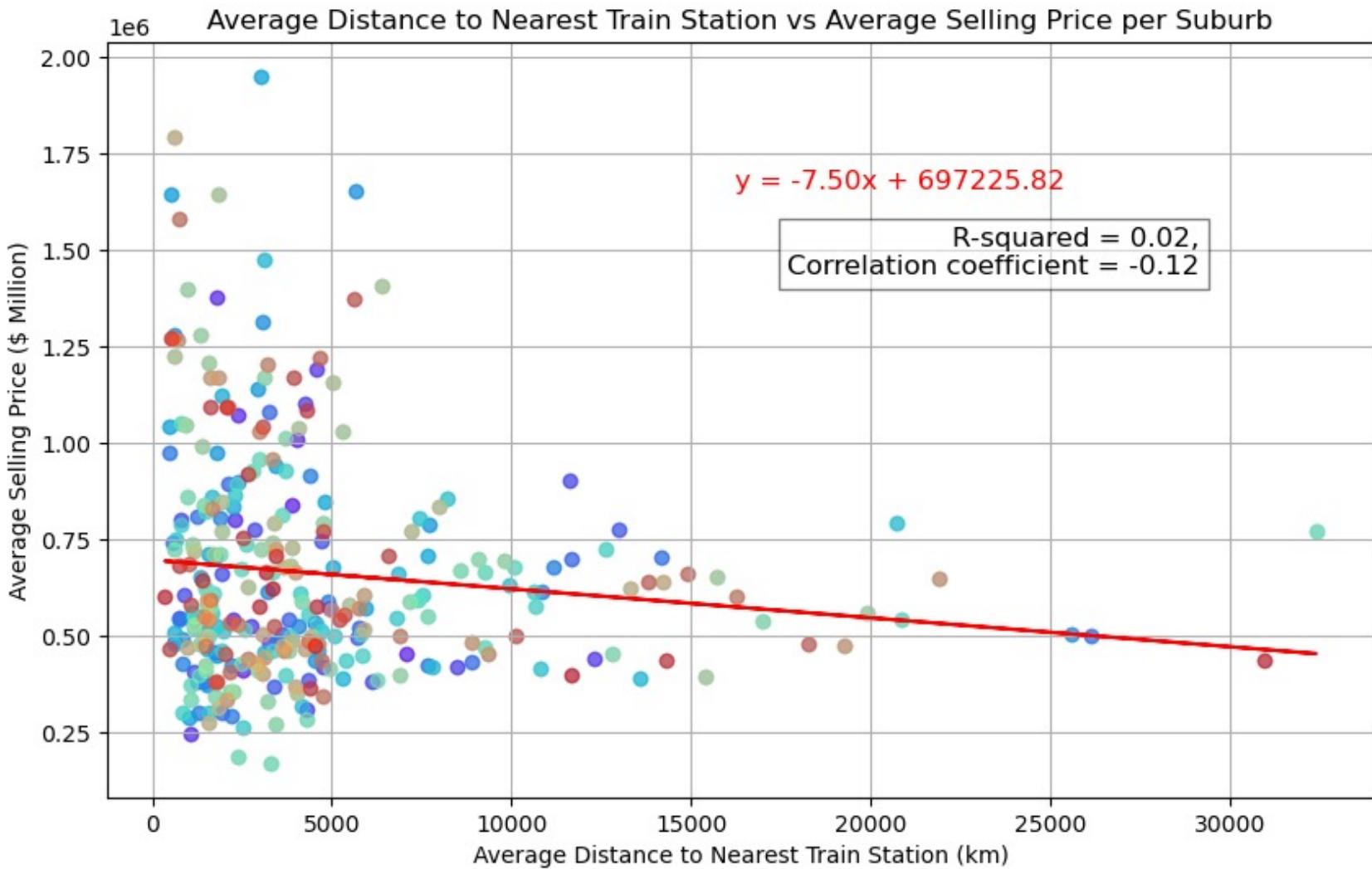
- Property Features
- 1Beds, 1Baths, 0Garages
  - 1Beds, 1Baths, 1Garage
  - 1Beds, 1Baths, 2Garages
  - 1Beds, 1Baths, 3Garages
  - 1Beds, 1Baths, 4Garages
  - 1Beds, 1Baths, 6Garages
  - 1Beds, 2Baths, 0Garages
  - 1Beds, 2Baths, 2Garages
  - 1Beds, 3Baths, 0Garages
  - 1Beds, 3Baths, 2Garages
  - 1Beds, 7Baths, 0Garages
  - 2Beds, 1Baths, 0Garages
  - 2Beds, 1Baths, 1Garage
  - 2Beds, 1Baths, 2Garages
  - 2Beds, 1Baths, 3Garages
  - 2Beds, 1Baths, 4Garages
  - 2Beds, 1Baths, 5Garages

# Price vs Distance to CBD

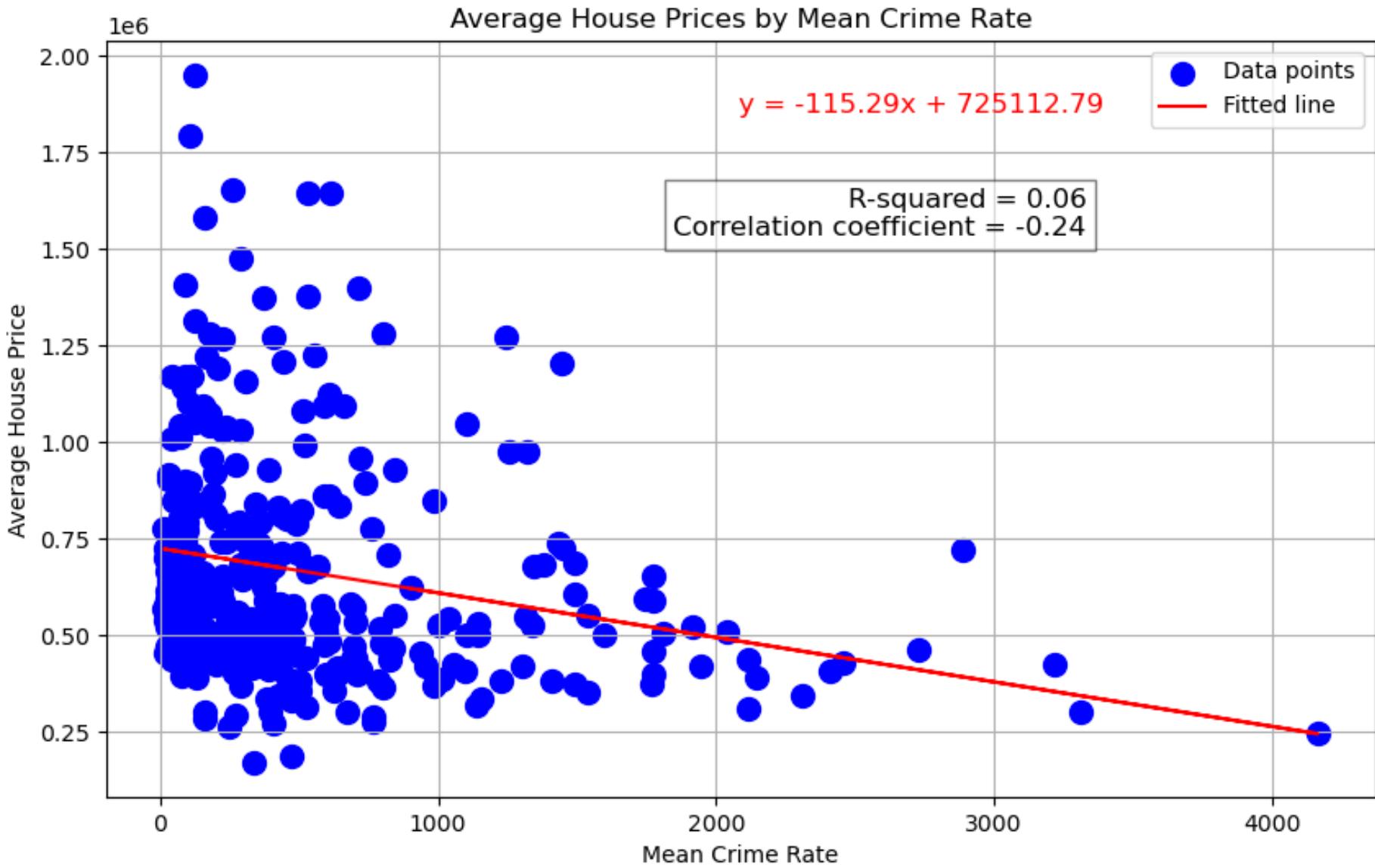




# Price vs Distance to Closest Train Station

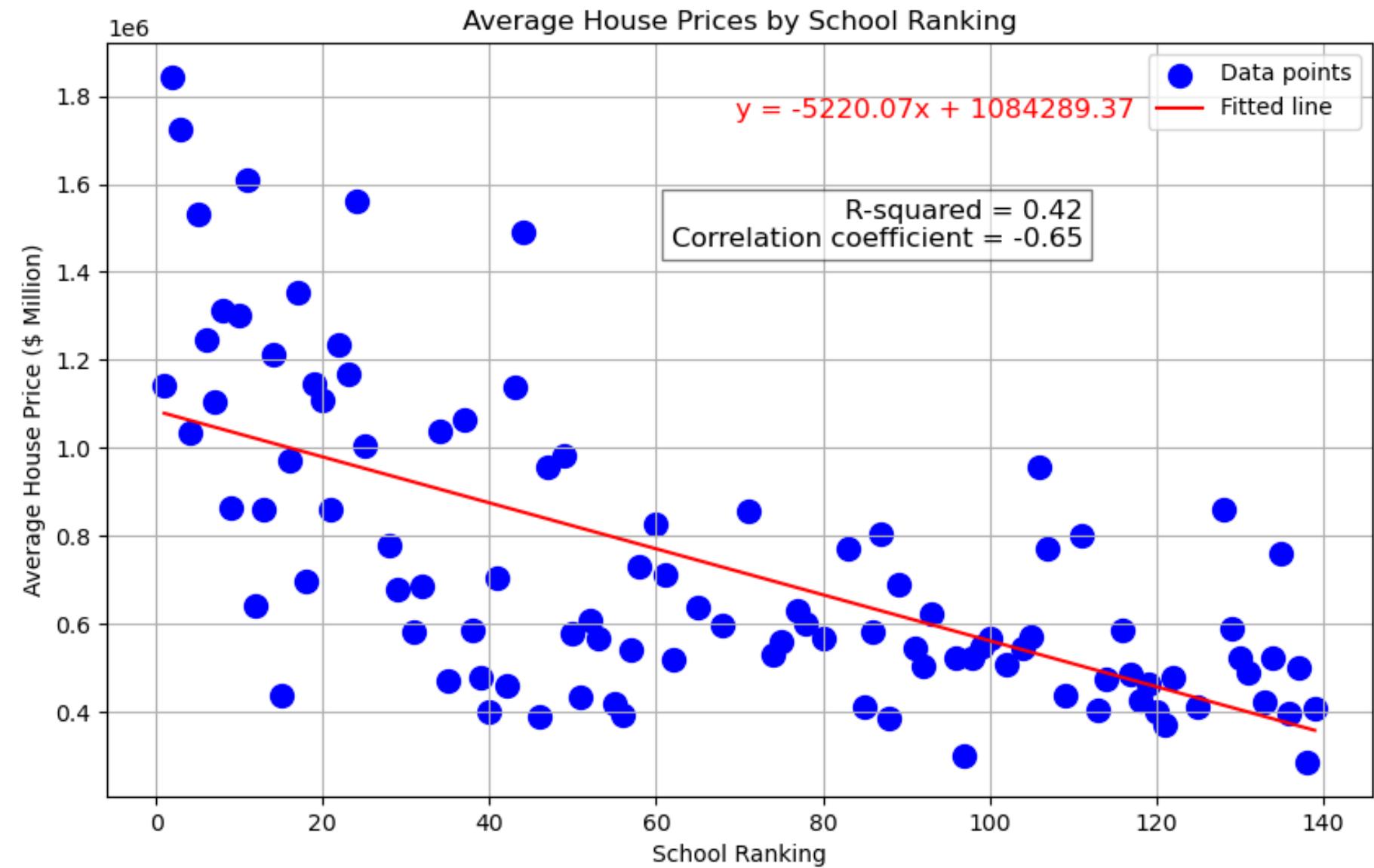


# Price vs Crime



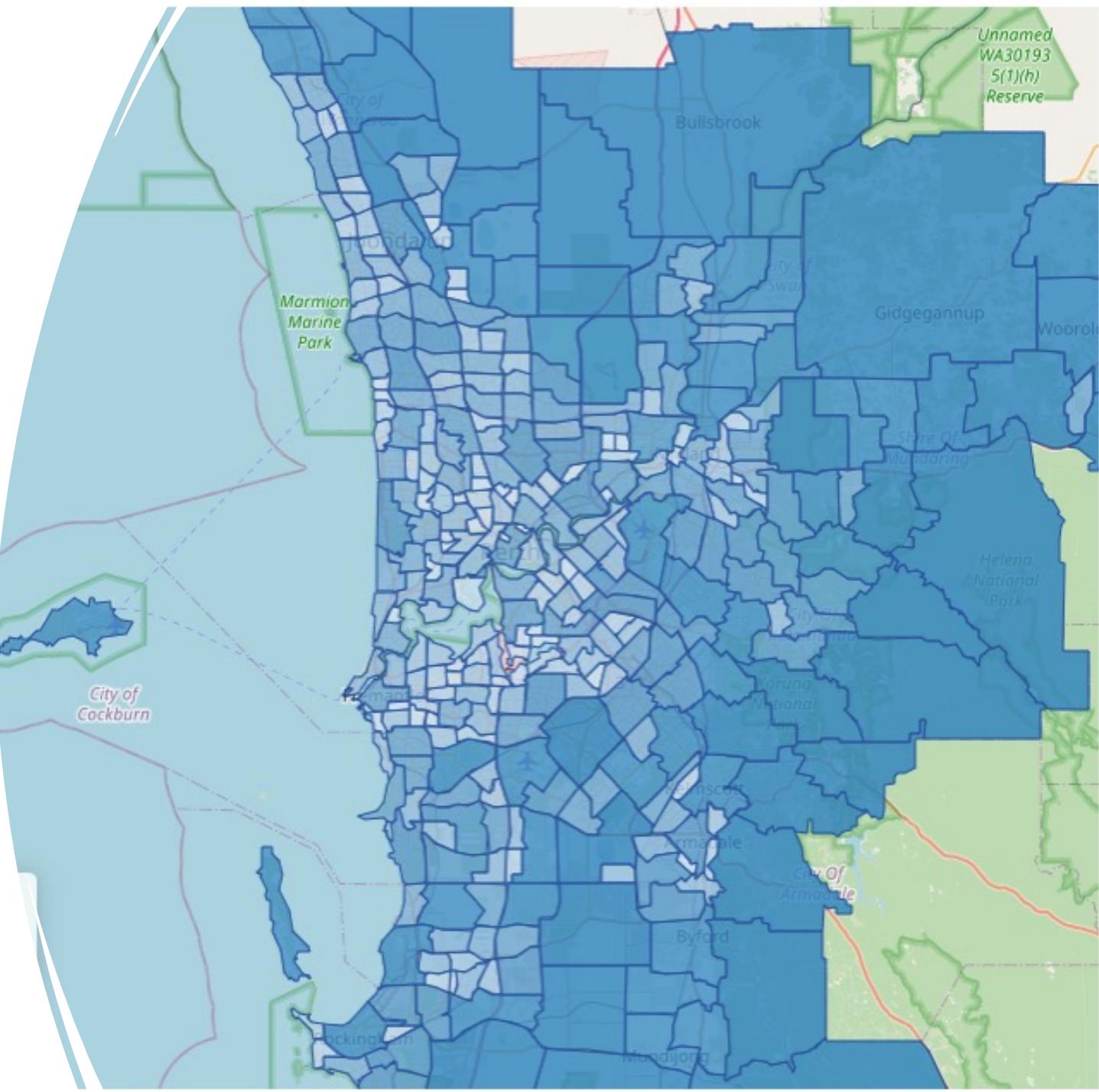


# Prices vs School Rank



# HTML Visualisation

---



# SUMMARY

## LIMITATIONS

- Data Quality and Availability
- Variable Selection
- External Factors
- Measurement Errors
- Causal Inference