Cherrybrook Real Estate Data Analysis Project

By: Nipunn Khurana

1. Project Overview

This project focuses on analysing the housing market in Cherrybrook, NSW, using real-world property data. The goal was to uncover insights into pricing trends, land size distribution, school catchment impacts, and sales agent performance. The analysis followed the data analytics pipeline: data extraction, cleaning, transformation, and visualization.

2. Data Extraction using Selenium

To begin with, I extracted raw property listing data from Domain.com.au using Selenium, a Python automation tool that simulates human interaction with web pages. I built a custom script that:

1. Navigated through multiple listing pages for Cherrybrook.
2. Captured relevant data fields such as address, price, sold date, link, property type, number of bedrooms, bathrooms, car spaces, and land area.
3. Stored the data in a structured format (CSV) for further analysis.

```python
import time

class DomainListings:
    def __init__(self, driver_path):
        self.driver_path = driver_path
        self.driver = None

    def start_browser(self):
        chrome_options = Options()
        chrome_options.add_argument("--start-maximized")

        # ✅ Block images to speed up page load
        prefs = {"profile.managed_default_content_settings.images": 2}
        chrome_options.add_experimental_option("prefs", prefs)

        service = Service(self.driver_path)
        self.driver = webdriver.Chrome(service=service, options=chrome_options)
        print("Browser started successfully (with images blocked).")

    def navigate_to_url(self, url):
        if self.driver is None:
            print("Error: Browser is not started. Call 'start_browser' first.")
            return
        self.driver.get(url)
        print(f"Navigated to {url}")

    def click_sold_button_and_apply_filter(self):
        try:
            wait = WebDriverWait(self.driver, 5)

            # Step 1: Click the 'Sold' button
            sold_button = wait.until(EC.element_to_be_clickable((By.CSS_SELECTOR, 'button[data-testid="sold-navigation"]')))
            sold_button.click()
            print("✅ Clicked the 'Sold' button.")

            # Step 2: Click the 'Filters' button
            filters_button = wait.until(EC.element_to_be_clickable((By.CSS_SELECTOR, 'button[data-testid="search-filters-button-desktop"]')))
            filters_button.click()
            print("✅ Clicked the 'Filters' button.")
```

This method enabled me to gather up-to-date, location-specific real estate information which would otherwise not be readily available in public datasets.

3. Data Cleaning and Manipulation in SQL

After collecting the raw dataset, I imported the CSV file into a SQL database to perform data cleaning and preprocessing. Key steps included:

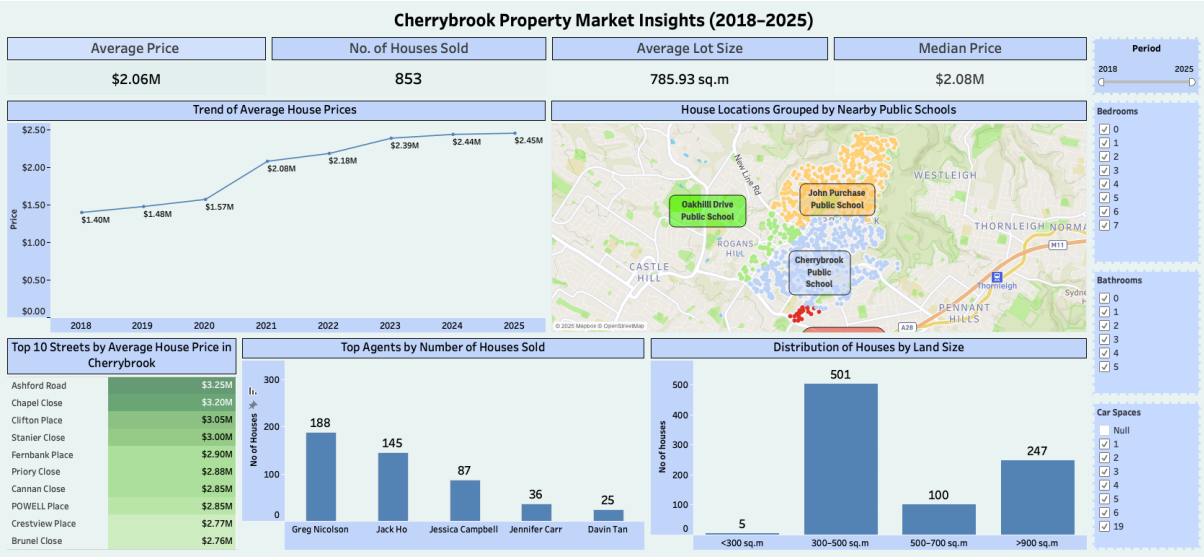
1. Parsing and standardizing address components.
2. Extracting numeric values from text fields such as price and area.
3. Handling missing values, such as estimating area where missing or filtering out incomplete records.
4. Formatting dates and calculating useful derived fields like price per square metre.
5. Using SQL for data wrangling ensured accuracy, efficiency, and reusability for multiple queries and filters.

```sql
-- Task 1.
ALTER TABLE cherrybrook
DROP COLUMN Page, Listing_Number

-- Task 2.

UPDATE cherrybrook
SET
  Car_Spaces = NULLIF(Car_Spaces, 'N/A'),
  Area = NULLIF(Area, 'N/A')
WHERE Car_Spaces = 'N/A' OR Area = 'N/A';

-- Car_Spaces has 3 N/A values, Area has 207 N/A values, Agent_Name has 131 missing values,
-- For Car_spaces, we will replace 'N/A' values with NULL
-- For Area, replace 'N/A' by NULL
-- For Agent_Name, dont replace 'N/A'

-- Task 3.

WITH CTE AS (
  SELECT *, ROW_NUMBER() OVER (PARTITION BY Link, Address, Price ORDER BY (SELECT NULL)) AS rn
  FROM cherrybrook
)
DELETE FROM CTE
WHERE rn > 1;

```

4. Visualization in Tableau

Once the dataset was cleaned and transformed, I imported it into Tableau Public to design an interactive dashboard. The dashboard was structured to answer key stakeholder questions and includes the following components:

1. Trend line of average property prices from 2018 to 2025
2. Geographical map of houses categorized by public school catchment
3. Bar charts showing agent-wise and area-wise property sales
4. Top 10 streets in Cherrybrook based on average price
5. KPIs like average price, median price, number of houses sold, and average land size



You can view the live dashboard here:

🔗 Tableau Public – [Cherrybrook Insights Dashboard]

5. Key Insights and Findings

Based on the dashboard visualizations, several important trends were identified:

- 🏠 The average house price in Cherrybrook increased significantly from $1.4M in 2018 to $2.45M in 2025, showing a clear upward trend in property value.
- 🗺️ School catchment areas had a noticeable effect on property distribution and clustering, especially around Cherrybrook Public, John Purchase, and Oakhill Drive Public School.
- 📊 Most houses were sold in the 300–500 sq.m range, followed by >900 sq.m, suggesting a strong demand for medium-sized lots.
- 👨‍💼 Greg Nicolson and Jack Ho were the leading agents in Cherrybrook by volume of houses sold.

📌 Top-priced streets include Ashford Road, Chapel Close, and Clifton Place, with average prices exceeding $3 million.

6. Conclusion

This end-to-end project demonstrates how real estate data can be scraped, structured, cleaned, and visualized to extract meaningful business insights. The combination of Selenium, SQL, and Tableau created a robust pipeline for analysis, and the findings can support homebuyers, real estate agents, or investors in making informed decisions.