General Sir John Kotelawala Defense University

Faculty of Management, Social Science and Humanities

Department of Languages

BSc in Applied Data Science Communication

Group 11

SQL For Data Science/MF1224

Year 1: Semester 2

Group assignment

30.11.2023

D/ADC/23/0019-B.K.M.M.Hansanee

D/ADC/23/0037- P.A.M.T Athukorala

D/ADC/23/0041-P.R.Withanage

D/ADC/23/0047-M.L.B.T.S Perera

Greater Manchester Property

Values

Dashboard Design

CONTENT

INTRODUCTION

The Greater Manchester Dataset is a comprehensive source of real estate information that captures the nuances of the real estate industry in this dynamic United Kingdom town. This dataset, which was built by carefully compiling data from many sources over a number of years, serves as the foundation for an in-depth examination of the Greater Manchester real estate market. The dataset, which includes information on property sales, pricing, geographic characteristics, and transaction volumes, is intended to provide academics, business professionals, and policymakers with a comprehensive viewpoint.

The dataset has been methodically arranged into tables, views, and stored procedures by utilizing T-SQL procedures. This has allowed for the efficient manipulation and retrieval of relevant information. Our goal as we analyze this dataset is to uncover important information on sales volume, price distribution, geographic patterns, and sales trends in order to provide a comprehensive picture of the Greater Manchester real estate market.

02.EXPLORATION OF DATA

2.1 Data set review

A data set review is a thorough examination of a set of data that looks at the relevance, quality, and any biases of the data. It involves evaluating the consistency, accuracy, and completeness of the data. Scholars and analysts carefully review the dataset's documentation to comprehend its source and constraints. This procedure guarantees the accuracy of data for a range of uses, including research, statistical analysis, and machine learning, enabling well-informed decision-making based on reliable and solid data.

1. Transaction unique identifier A reference number is generated automatically recording each published sale. The number is unique and will change each time a sale is recorded.

2. Price The sale price is stated on the transfer deed.

3. Date of Transfer Date when the sale was completed, as stated on the transfer deed.

Postcode This is the postcode used at the time of the original transaction. Note that postcodes can be reallocated, and these changes are not reflected in the Price Paid Dataset.

4. Property Type
   D = Detached, S = Semi-Detached, T = Terraced, F = Flats/Maisonettes, O = Other

5. Old/New Indicates the age of the property and applies to all price-paid transactions, residential and non-residential.
   Y = a newly built property, N = an established residential building

6. Duration
   F = Freehold, L= Leasehold etc.

7. PAON (Primary Addressable Object Name) Typically, the house number or name.

8. SAON (Secondary Addressable Object Name) Where a property has been divided into separate units (for example, flats), the PAON (above) will identify the building and a SAON will be specified that identifies the separate unit/flat.

9. Street

10. Locality

11. Town/City

12. District

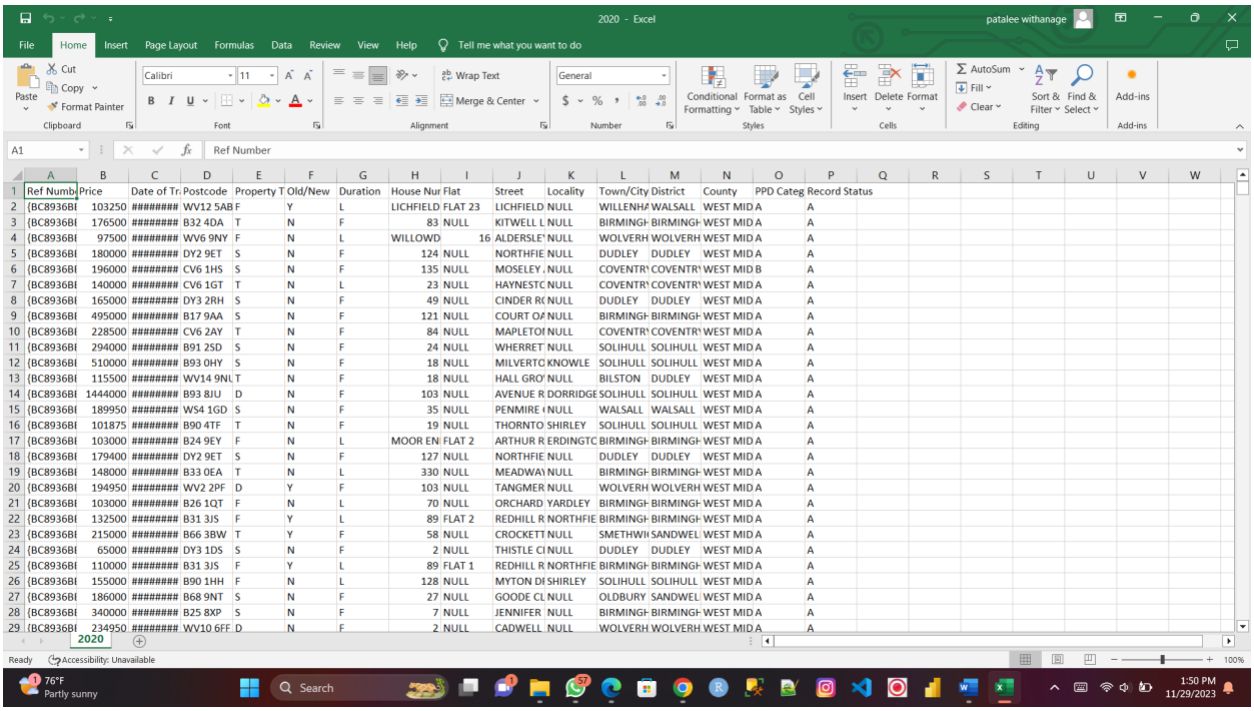
13. Country

14. PPD Category Type

    Indicates the type of Price Paid transaction.

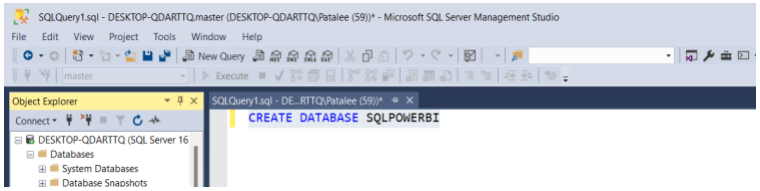
15. Record Status monthly file only Indicates additions, changes, and deletions to the records.
    A = Addition, C = Change, D = Delete

Dataset: https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads
Metadata: https://www.gov.uk/guidance/about-the-price-paid-data



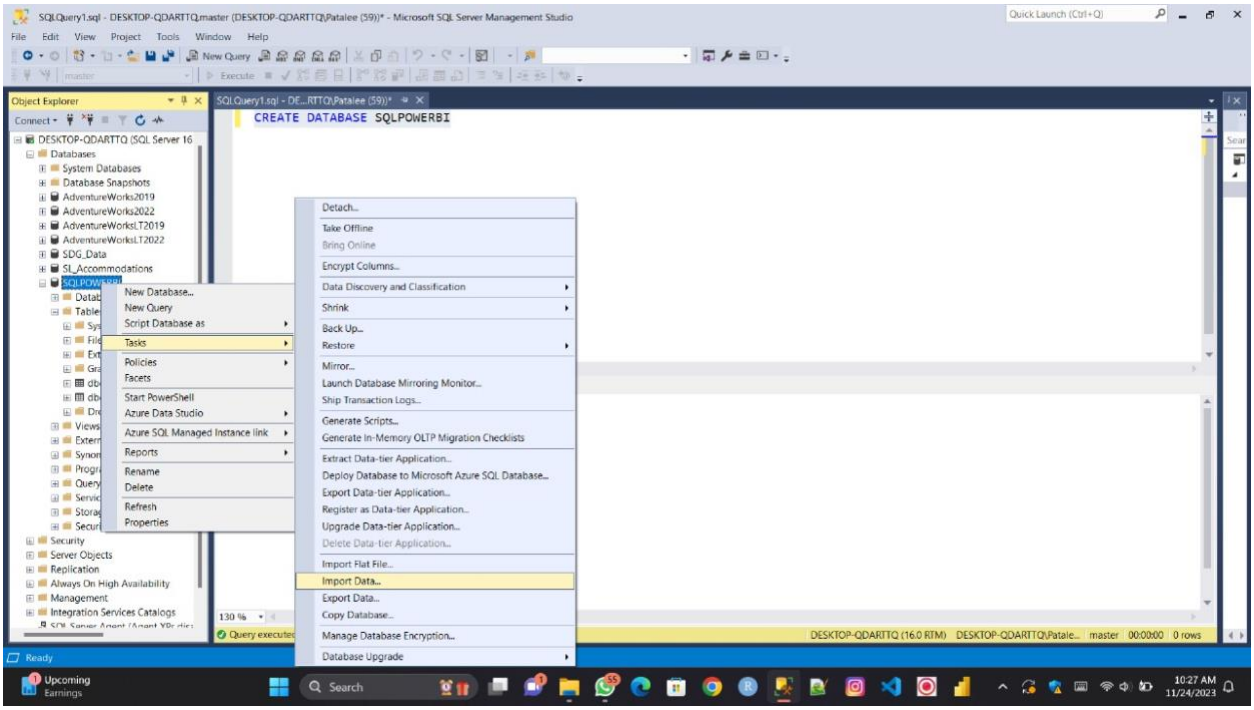
2.2 Importing Data from Excel to a Microsoft SQL Server Database

- ✓ Use the provided link to get CSV data files for the following years: 2019, 2020, 2021, and 2022. https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads
- ✓ View each of those four CSV data files independently.
- ✓ Open SSMS and connect to SQL server instance.
- ✓ Create a database named as 'SQLPOWERBI'.
- ✓ In the toolbar located and click the button labelled new query.



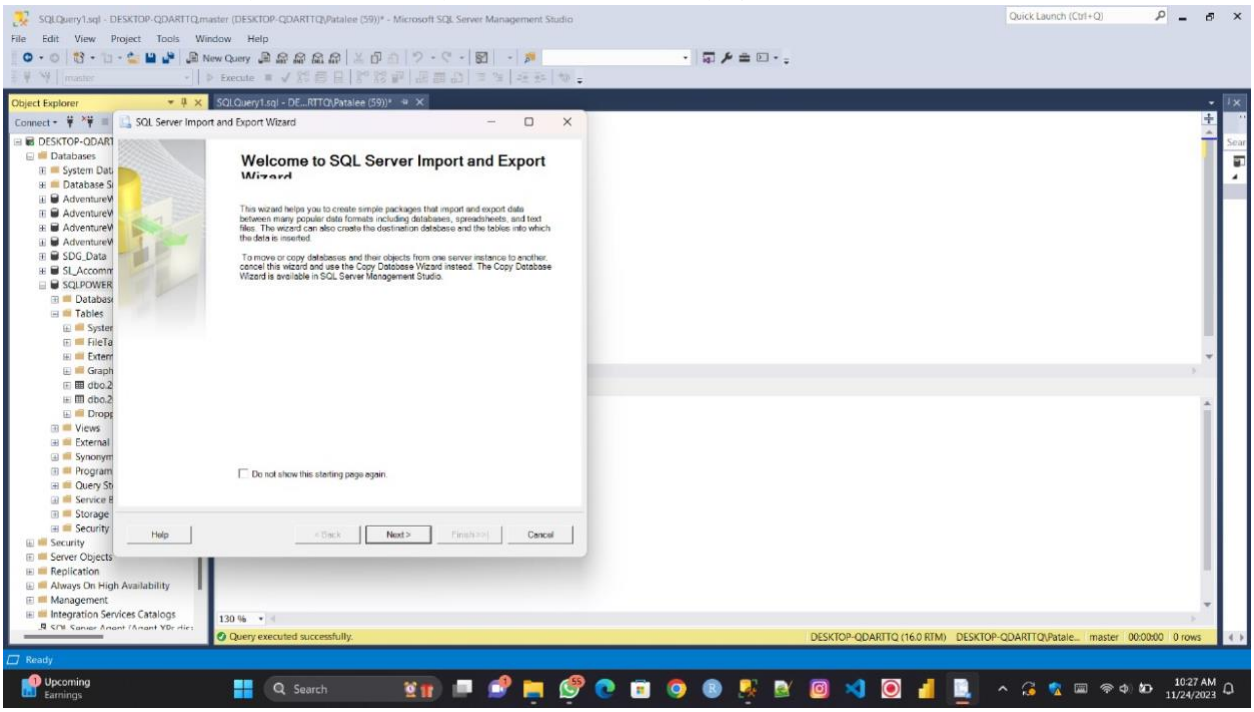
- ✓ Now enter the data following code into a query editor window and execute.

- ✓ Expand the object explorer server tree
- ✓ Expand the Database folder.
- ✓ Right click on ‘SQLPOWERBI’ Database.
- ✓ Select Tasks.
- ✓ Move to Import Data.

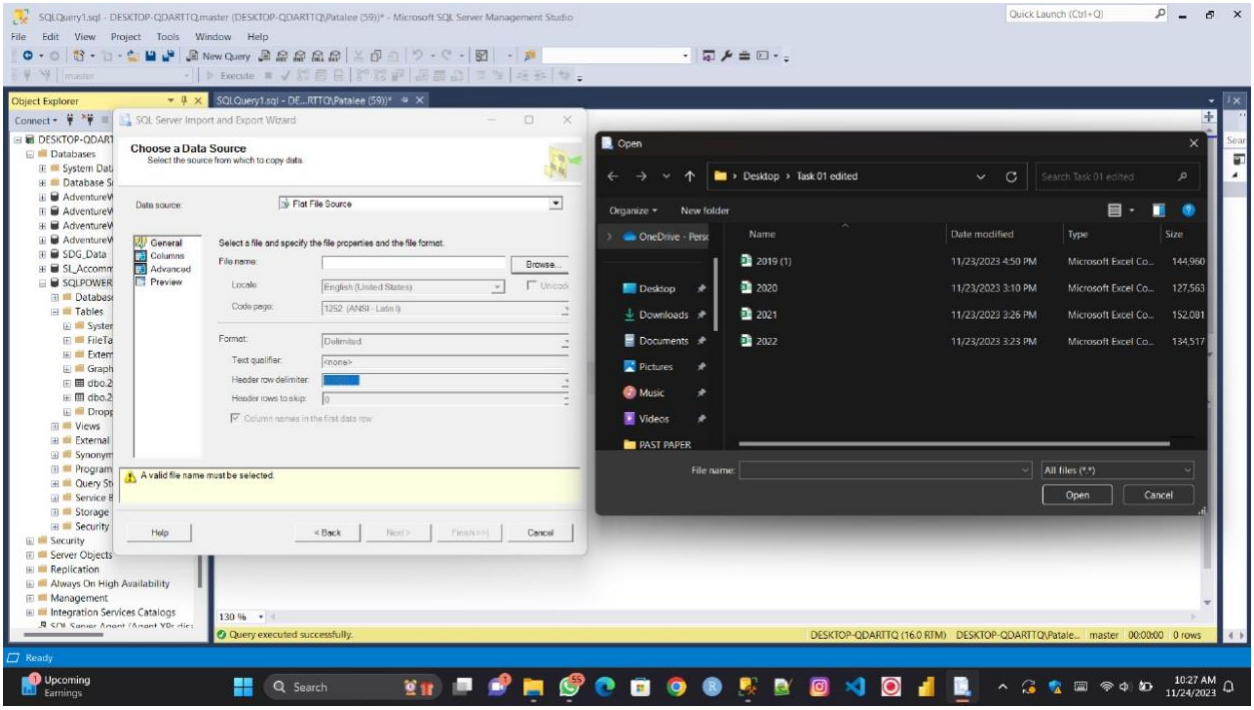


- ✓ Now click next on Import data and Export Wizard Welcome page.

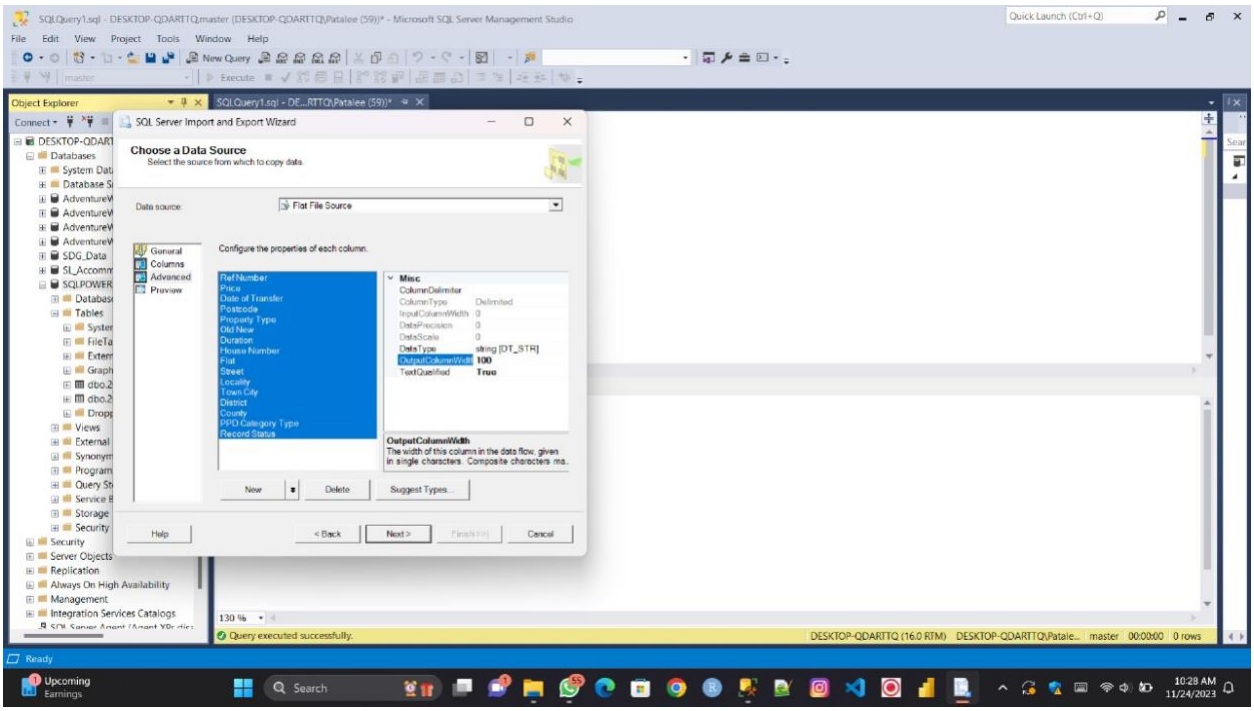
✓ Now click next on Import data and Export Wizard Welcome page



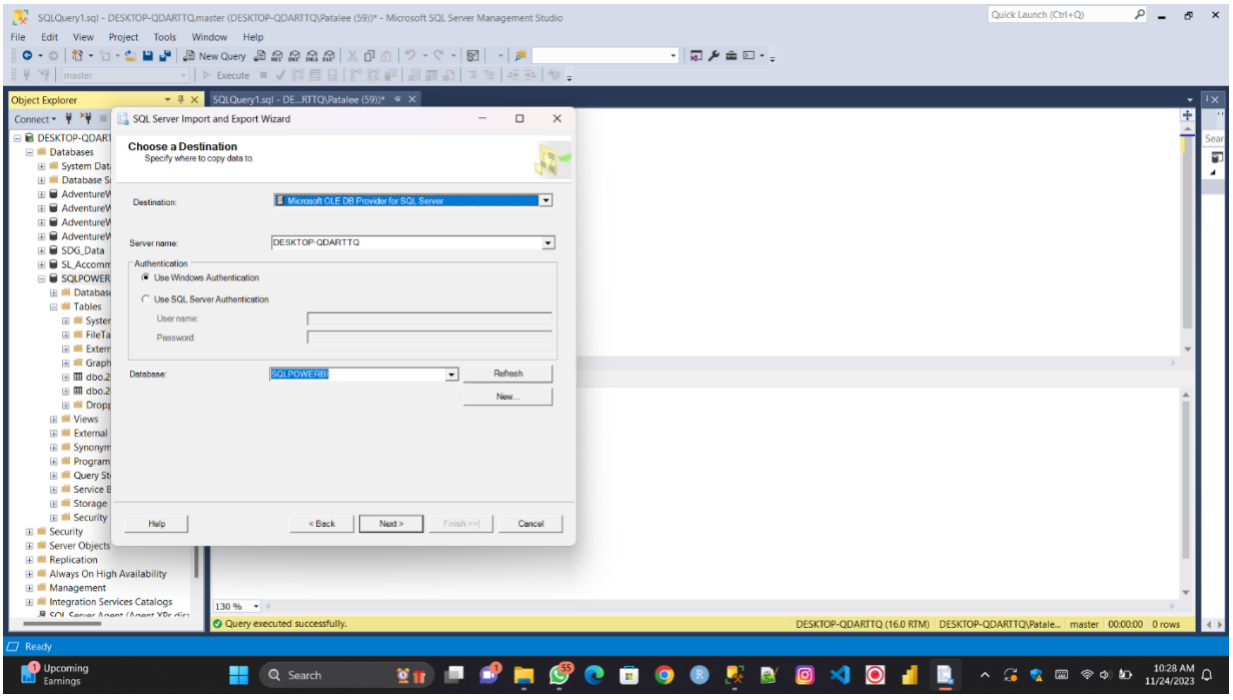
✓ You can select flat file source as the data source and enter or browse for the file.



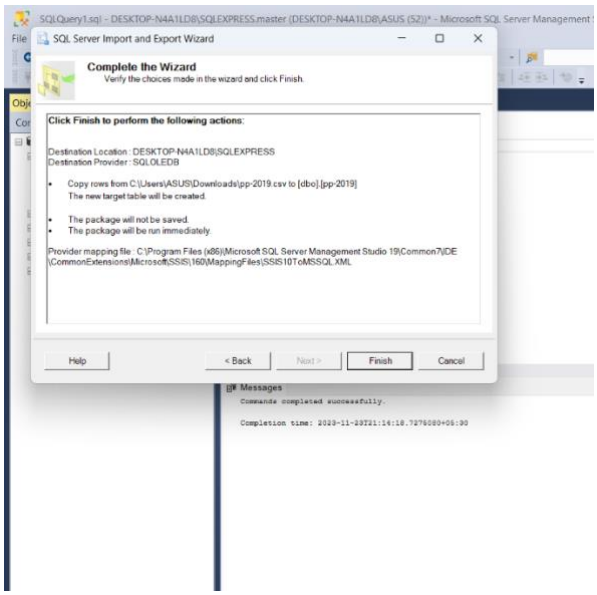
✓ Go to Advanced and change all the columns from 50 to 100



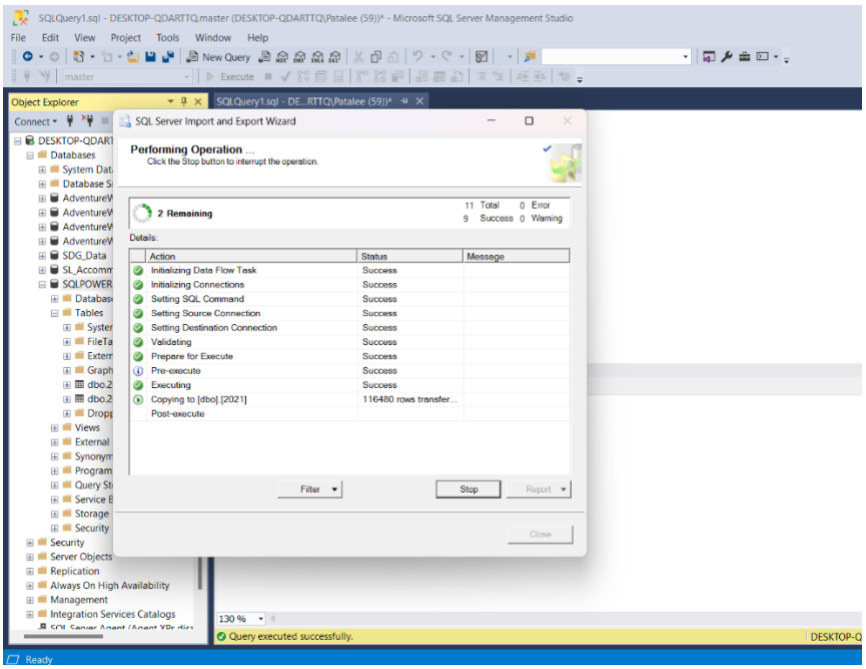
✓ Click Next to move forward and choose a destination.



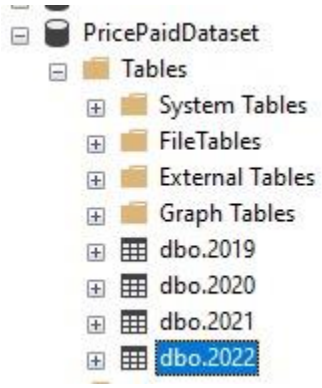
✓ Click Next and Accept the default.
✓ Click finish



✓ The execution dialog box appears, if all the data has loaded click close

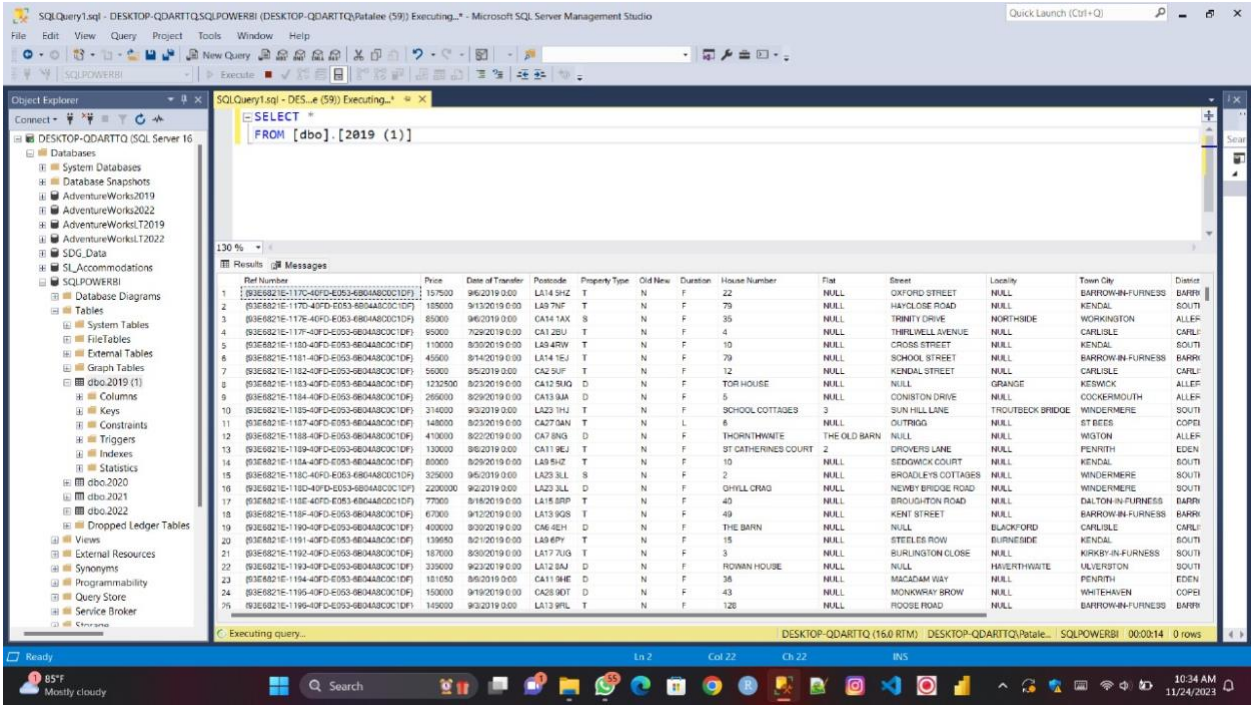


✓ Now you can see the new table in ‘SQLPOWERBI’ database.



✓ You must repeat the same procedure for three times to add the three CSV files

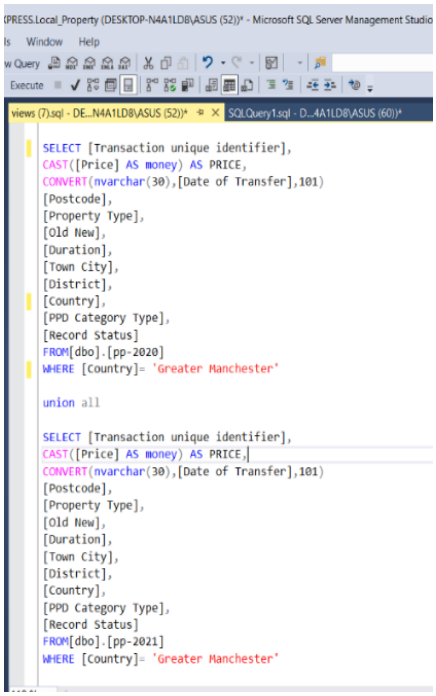
✓ In a new query, type and execute the following code to see the imported data and check the other two tables as follows



2.3Create view

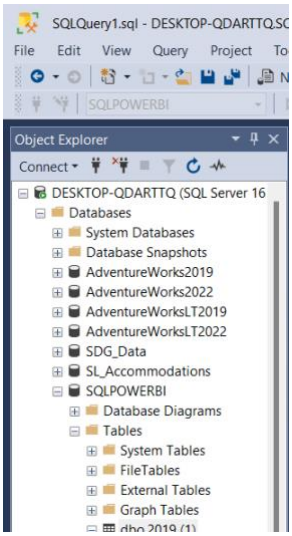
Using this technique, was able to eliminate a few columns and retrieve the inconsistent data under these stages.

- Under a new query, enter the following code to create a view by attaching all four table.



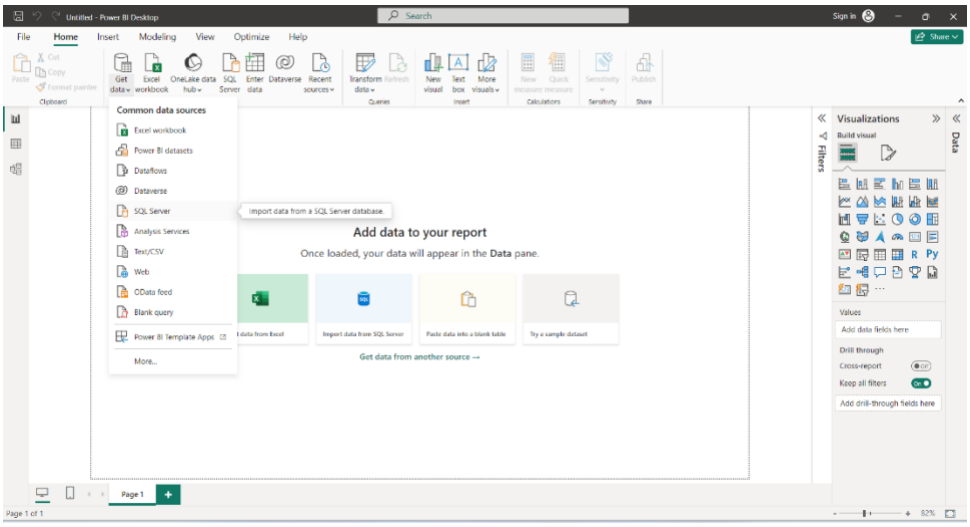
```
SELECT [Transaction unique identifier],
CAST([Price] AS money) AS PRICE,
CONVERT(nvarchar(30),[Date of Transfer],101)
[Postcode],
[Property Type],
[Old New],
[Duration],
[Town City],
[District],
[Country],
[PPD Category Type],
[Record Status]
FROM[dbo].[pp-2020]
WHERE [Country]= 'Greater Manchester'

union all

SELECT [Transaction unique identifier],
CAST([Price] AS money) AS PRICE,
CONVERT(nvarchar(30),[Date of Transfer],101)
[Postcode],
[Property Type],
[Old New],
[Duration],
[Town City],
[District],
[Country],
[PPD Category Type],
[Record Status]
FROM[dbo].[pp-2021]
WHERE [Country]= 'Greater Manchester'
```

- You can see the view under the views, in 'SQLPOWERBI' data bases



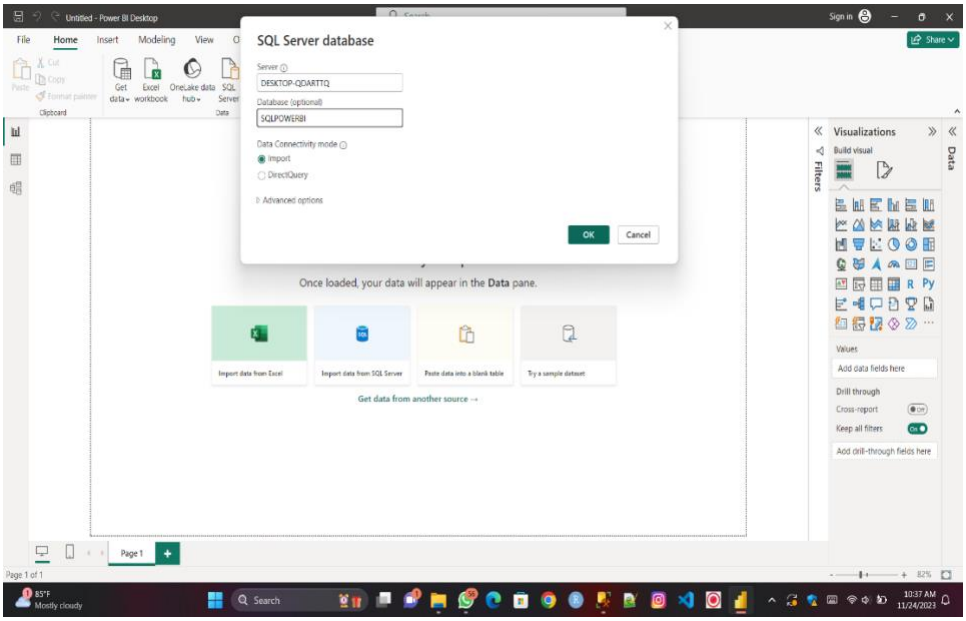
2.4Data from Microsoft SQL Server to Power BI

- Firstly, Open the Power BI in the desktop.
- In the Home tab, go to Get data and connect to data from multiple sources.

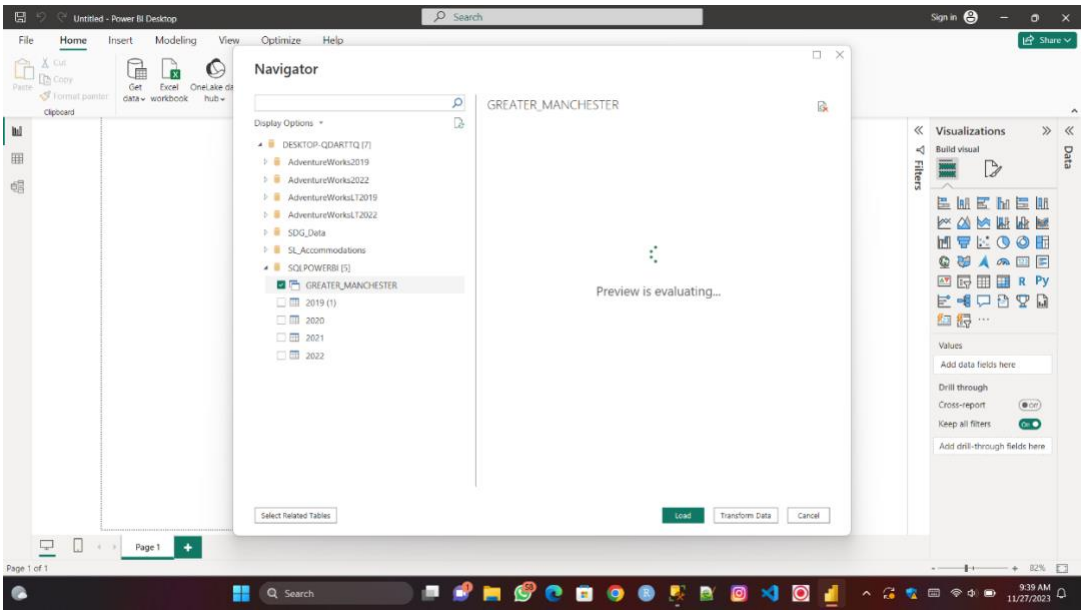


- Select the Microsoft SQL Server.

- Connect the MSS and add the 'SQLPOWERBI' Database.



- Select the View and upload it in the Power BI data tab.



03.Dashboard design and implementation

3.1 Dashboard purpose

The process of building a Power BI dashboard for the Greater Manchester Property Price Monitor entails creating graphics with the intention of offering useful information about the regional real estate market. Users of the dashboard, including researchers, policymakers, and industry stakeholders, should be able to better understand the Greater Manchester area's sales volume, pricing distribution, sales trends, and geographic patterns.

1. **Sales Trend Analysis:**
    - **Visual:** Line chart showing monthly property sales over time.
    - **Purpose:** Understand the overall trend in property sales to identify peak months, seasonal patterns, or market trends.
2. **Price Distribution:**
    - **Visual:** Histogram showing the distribution of property prices.
    - **Purpose:** Analyze the range and frequency of property prices to identify the most common price brackets and outliers.

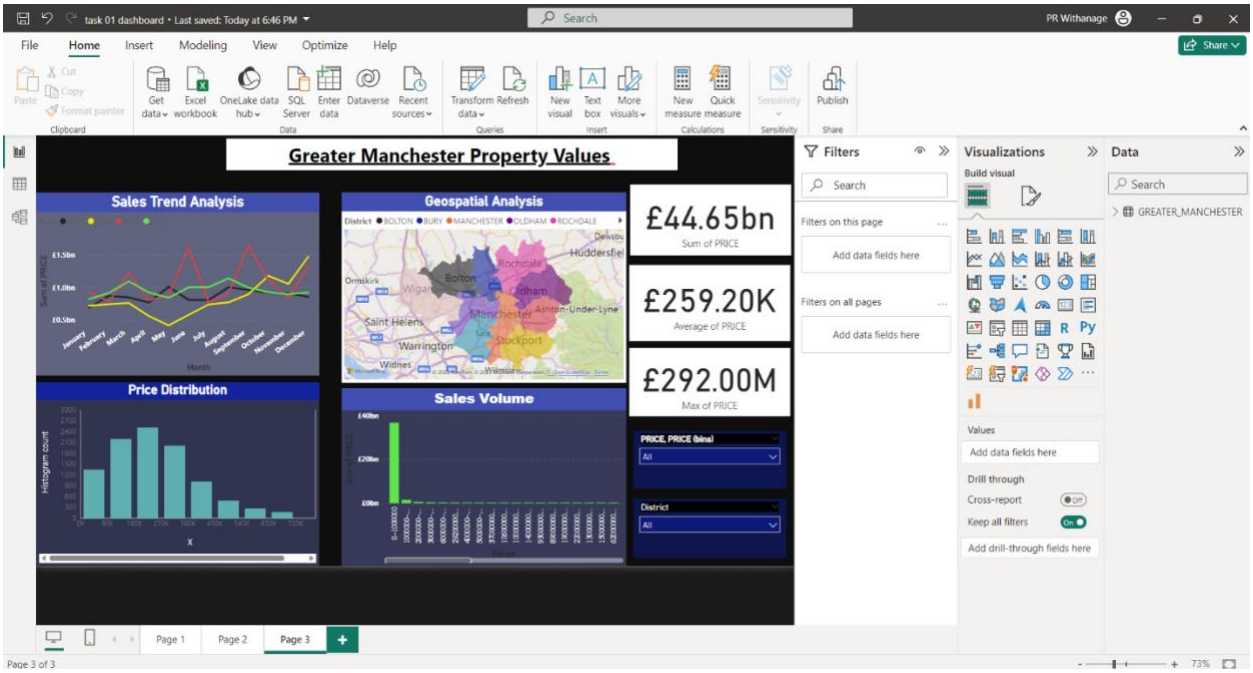
3. **Geospatial Analysis:**

    - **Visual:** Map visualization showing sales distribution across regions.
    - **Purpose:** Identify areas with higher or lower sales activity, helping users understand the geographical dynamics of the property market.

4. **Sales Volume:**

    - **Visual:** Column or bar chart representing the number of sales in different price brackets.
    - **Purpose:** Analyze the volume of sales within specific price ranges, providing insights into market segmentation and buyer preferences.

3.2 Dashboard draw up



Dashboard illustration

By analyzing a large amount of data, this study aims to shed light on the intricate dynamics of the real estate market. In order to provide insight into the intricacies of sales across many locations and time periods, its goal is to identify trends, patterns, and distribution in the property sales landscape over a substantial length of time.

This study employs Power BI to visually illustrate the intricacies of the real estate market. It displays geographic patterns, pricing distributions, sales trends, and price-sales correlations. Extremely helpful for stakeholders, policymakers, and investors to make educated decisions.

Investors can make well-informed judgments by analyzing sales trends, price variations, and regional discrepancies; policy officials can use this information to evaluate and control markets.

Methodology

The analysis can use a technique that includes several critical steps:

1. Data acquisition: Data collection from government databases and transfer to a local database using the Greater Manchester 2019, 2020, 2021, and 2022 rates paid datasets.

2. Data Cleaning: Systematic handling of collected data to ensure accuracy, uniformity and reliability. This step addresses the issue of missing values and removes duplicate entries.

3. Data visualization with Power BI: Using Power BI features to generate visual representations such as line graphs to depict sales trends over time, histograms to depict price distributions, maps to display geographic sales patterns, and bar/bar charts to compare sales volume. across different categories. Pricing in parentheses

The methodology involves data processing, visualization, analysis, interpretation and reporting with the ultimate goal of deriving significant insights from the real estate sales data set.

Analysis overview

• Analysis of sales trends

A line chart depicting monthly property sales displays recurring patterns, allowing for seasonal high points, low points and periods with potential for increased sales to be identified.

• Price allocation

A historical chart illustrating the distribution of property prices shows pricing patterns within the market. By analyzing this representation, it is possible to identify existing price ranges or any irregularities in the market.

• Spatial analysis

A map representation showing the spread of sales across different locations offers a geographic perspective of sales activity. A high or low concentration of sales in specific locations is evident, and it is possible to identify areas of significant market activity and regions that may require additional study or targeted tactics.

• Quantity of goods sold

A bar/bar chart depicting sales across several price ranges provides valuable insight into the correlation between price and sales volume. Examining this figure, it is clear that there are relationships between different price ranges and sales volume.

Visualizations offer multiple perspectives on real estate dynamics, serving as a valuable tool for investors and policymakers to improve understanding, identify impacts, and inform decision-making. They hinder more examinations and proposals.

Insights and Implication

1. Implications:

• Rational decision making:

- Investors: Analytical information on sales patterns, price dispersion and geographic disparities provides valuable guidance for making strategic real estate investments.

- Government officials responsible for creating and implementing policies: Understanding the fluctuations and forces that drive the market is critical to creating policies that promote long-term growth in the real estate industry.

• Market Analysis:

- Identifying opportunities: Sales trend research helps investors identify certain periods of potential growth that can result in higher returns on investment.

- Policy intervention: Geographical analysis supports policy makers in allocating resources to places with limited sales activity with the aim of promoting growth and progress.

2. Limitations:

• Data quality: -

- Incomplete data: The presence of inaccurate or missing data points can have a negative impact on the accuracy of trend analysis or price distribution.

- Limited data set: If the data set covers a limited time frame, it may fail to capture long-term patterns and developments.

• Difficulty understanding and explaining meaning: - Subjectivity: Interpretations of trends can be subjective, which can affect decision-making by influencing human perceptions.

- External factors: Factors not taken into account, such as fluctuations in the economy or changes in policies, can affect the results of the analysis. In the Power BI dashboard

Recommendations for real estate investors:

1. The act or process of forming mental images or representations. Description: • Strategic Action Cards: This section offers practical insight and advice for real estate investors in a concise card format.

• Cards are classified based on specific proposition types such as market timing, target investing or risk avoidance.

• Key Insights: Each card includes brief descriptions or concise bullet points that outline actionable recommendations as a result of the analysis.

2. A more nuanced perspective:

• The display is designed to be action-oriented, providing investors with clear and practical advice. These suggestions are categorized for easy reference.

• Precise Recommendations: Categorized cards enable investors to quickly find and focus on strategies relevant to their investment objectives.

• Easy-to-understand information: short explanations or bullet points ensure a quick understanding of the basic aspects of each proposed technology.

Recommendations for policy makers:

1. Description of visualization:

• Policy-oriented tiles: This section presents policy-oriented insights and recommendations in a tile-based format.

• Tile Categories: Each tile corresponds to a different policy domain, such as regional development, regulatory changes or market stabilization measures.

• Brief Description: Tiles consist of brief summaries of essential points, summarizing potential policy actions arising from the analysis.

The purpose of the visualizations in the Power BI dashboard is to simplify complex analysis into concise, actionable propositions for real estate investors and policy makers. This helps in making well-informed decisions and developing strategic plans in the real estate market.

Conclusion

This study details the use of T-SQL procedures to progressively integrate Greater Manchester property databases into a local database between 2019 and 2022. To provide efficient data administration and user-friendliness, this approach required the creation of tables, views, and stored procedures.

Using this large database, a comprehensive Energy BI dashboard was developed to satisfy user needs and accurately assess the local real estate market. For effective market analysis, a basic visualization dashboard should include the following:

A dynamic line chart visually depicts monthly property sales trends over time.

A clear and understandable representation of sales distribution across various areas can be obtained by using map visualization, whereas volume of goods sold is represented by a bar or bar chart that shows sales volumes in various price categories. This illustrates the relationship between price ranges and sales. Property price distribution primarily highlights price ranges and unusual patterns in the market.

Combining many sources of data makes for a thorough examination of the property market in Greater Manchester.

The Power BI dashboard enables users to access actionable information to make well-informed decisions in the real estate sector.