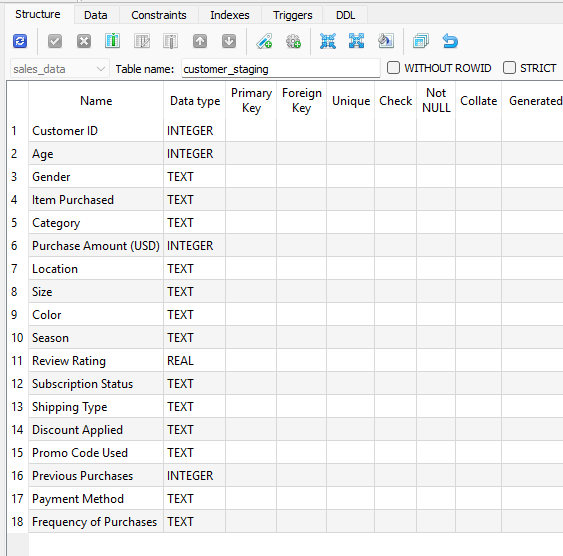
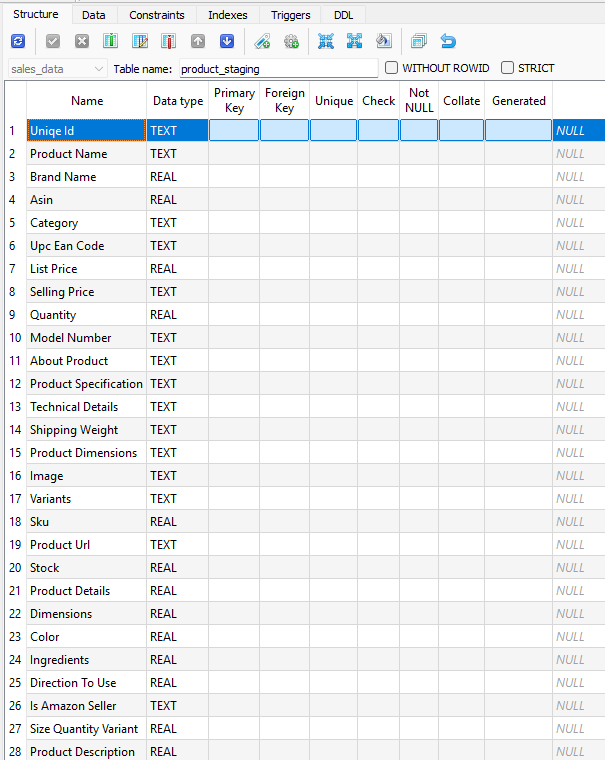
**Documentation**

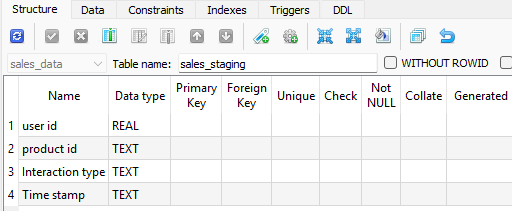
1. Background
   1. This project to create a datamart whose data comes from https://www.kaggle.com/datasets/datascientist97/e-commerece-sales-data-2024.
   2. ETL process using python script and sqlite as data storage for datamart.
   3. Automate jobs using task scheduler in windows system by running batch file that calls python script
   4. Datamart results in sqlite databases are used to build dashboards with Power BI
2. ETL Process
   1. Data collecting from E-commerce Sales Data 2024 dataset on Kaggle. consist of
      1. customer\_details.csv
      2. E-commerece sales data 2024.csv
      3. product\_details.csv
   2. Data extraction using python script. This stage is a process
      1. Checking data
      2. Cleaning data
      3. Remove duplicate
      4. Remove unused column
      5. Transform data
   3. Load data into sqlite databases from pandas dataframes, the steps are
      1. Create table for datamart purpose (staging,dimension and fact table).
      2. Insert into table staging for temporary data
      3. Load data from table staging to datamart (dimension and fact table).
3. Databases Structure and Datawarehouse Schema
   1. Staging Area
      1. Customer\_staging



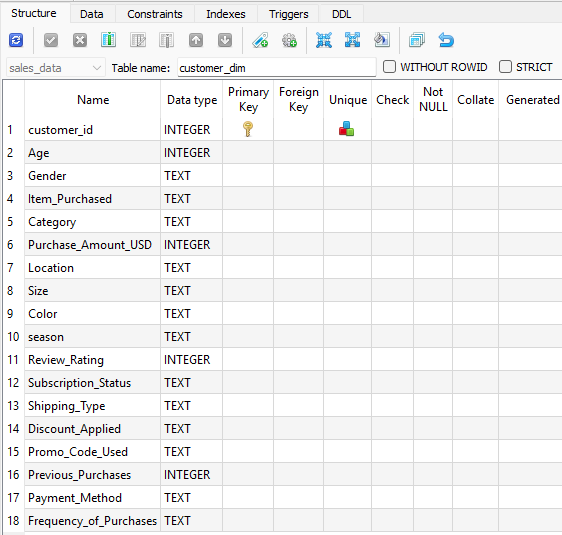
* + 1. Product Staging



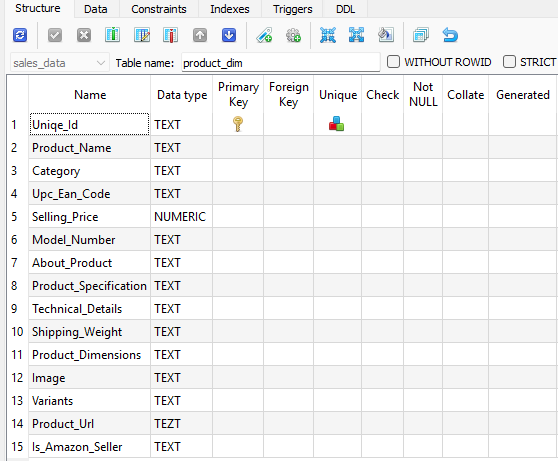
* + 1. Sales Staging



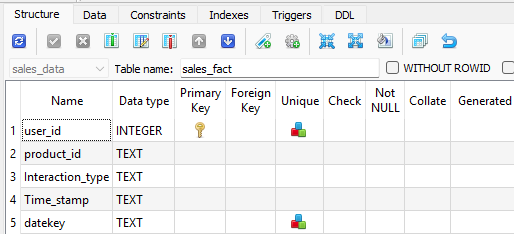
* 1. Fact Area
     1. Customer\_dim



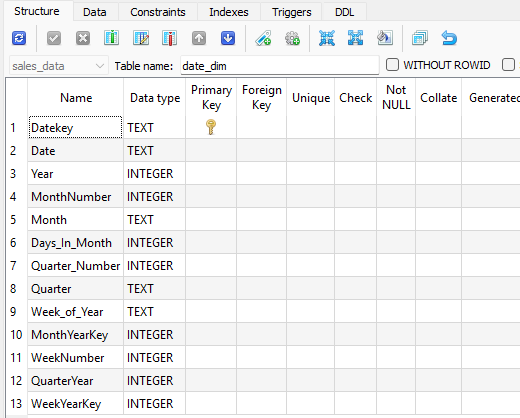
* + 1. Product\_dim



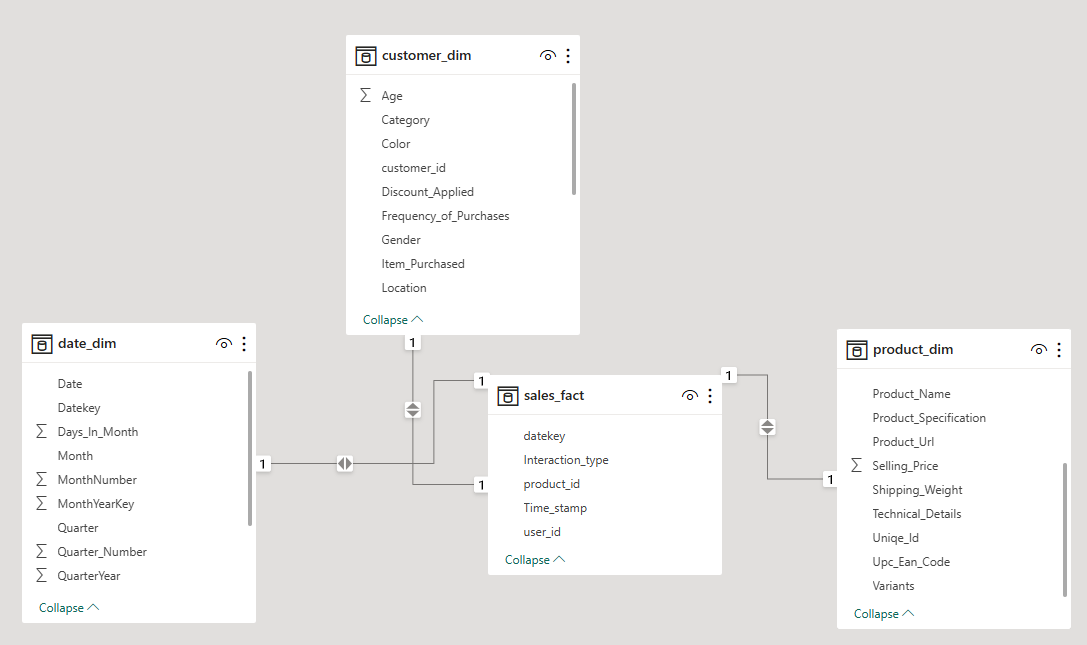
* + 1. Sales\_fact



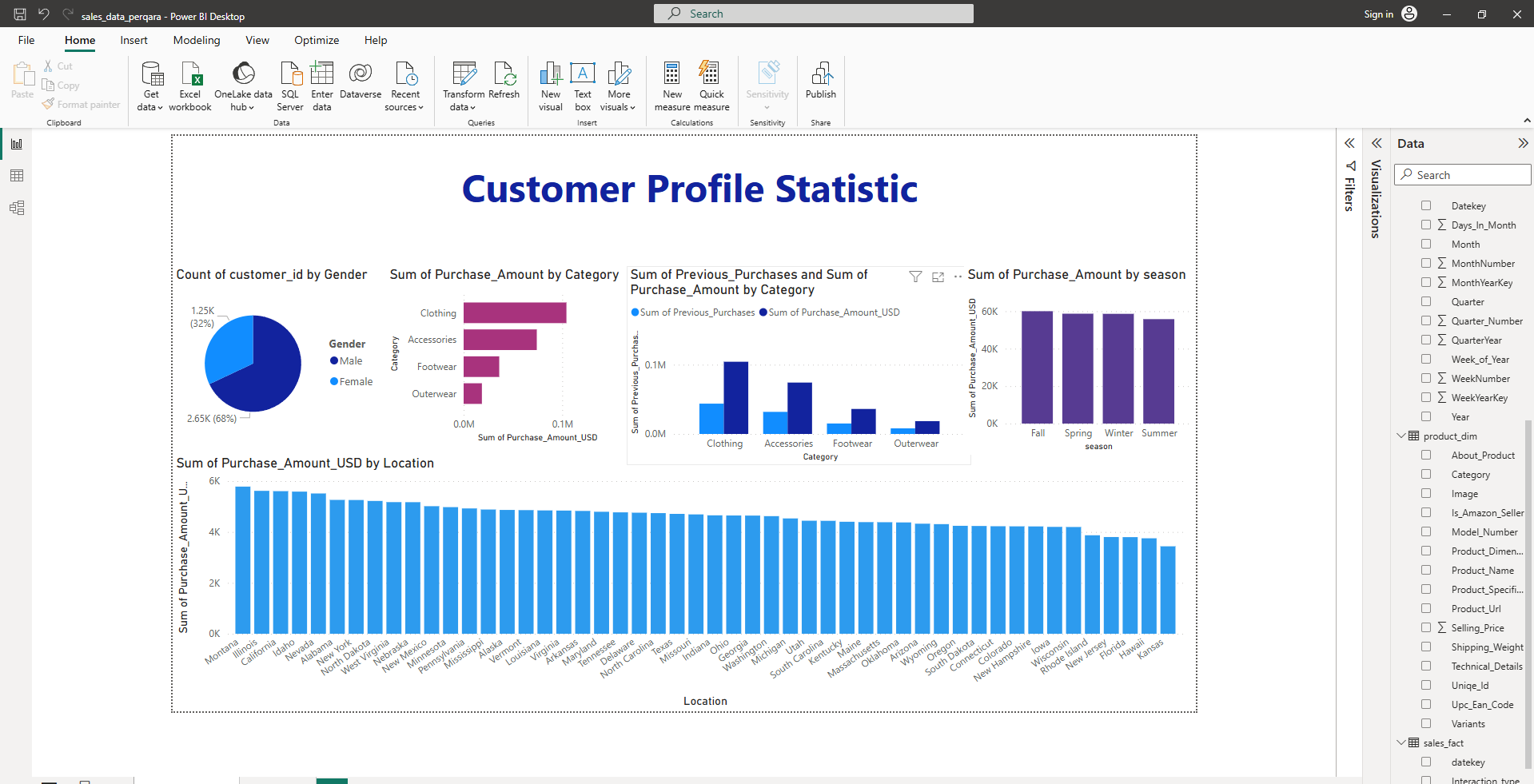
* + 1. Date\_dim



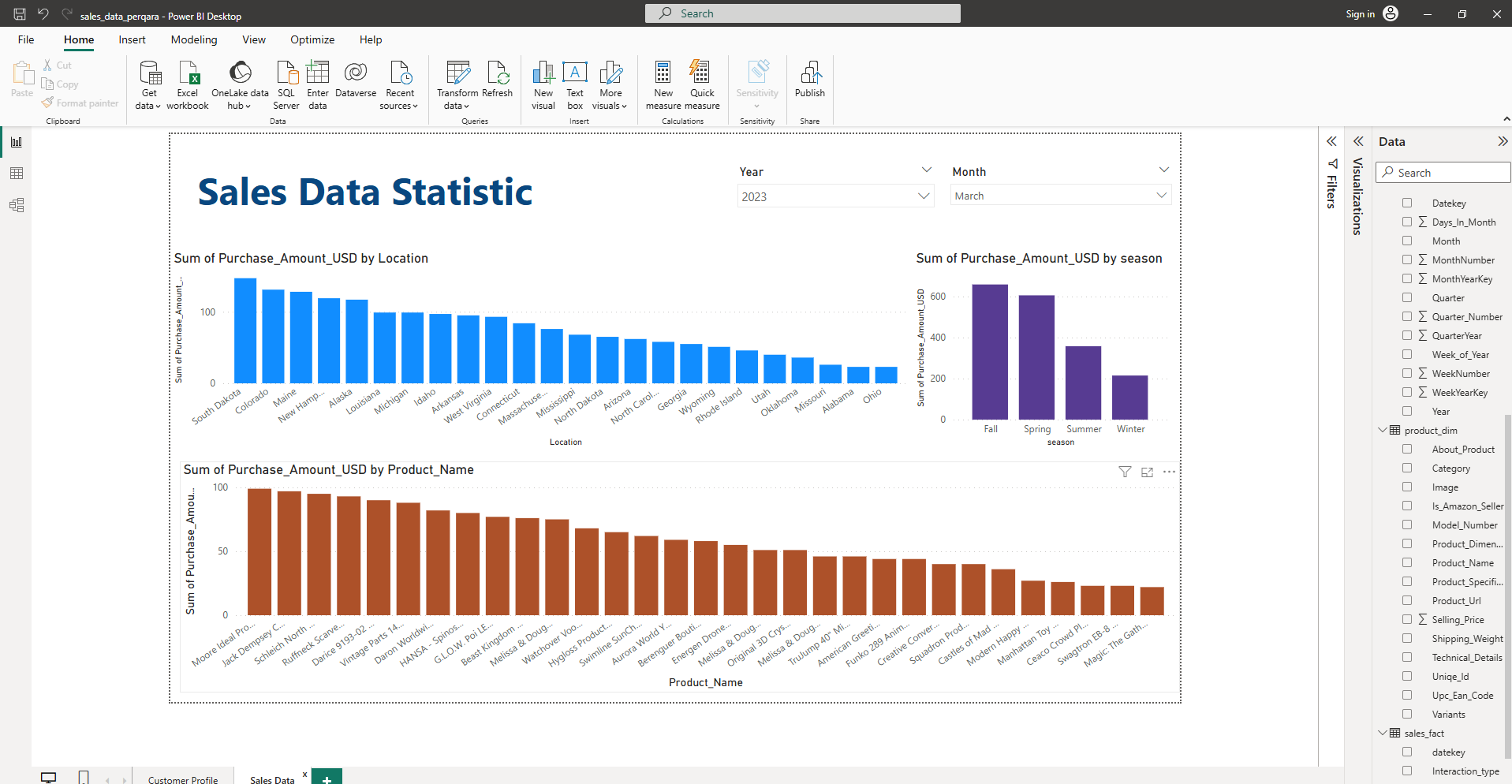
* 1. Datawarehouse using star schema design, this is model view capture from power BI



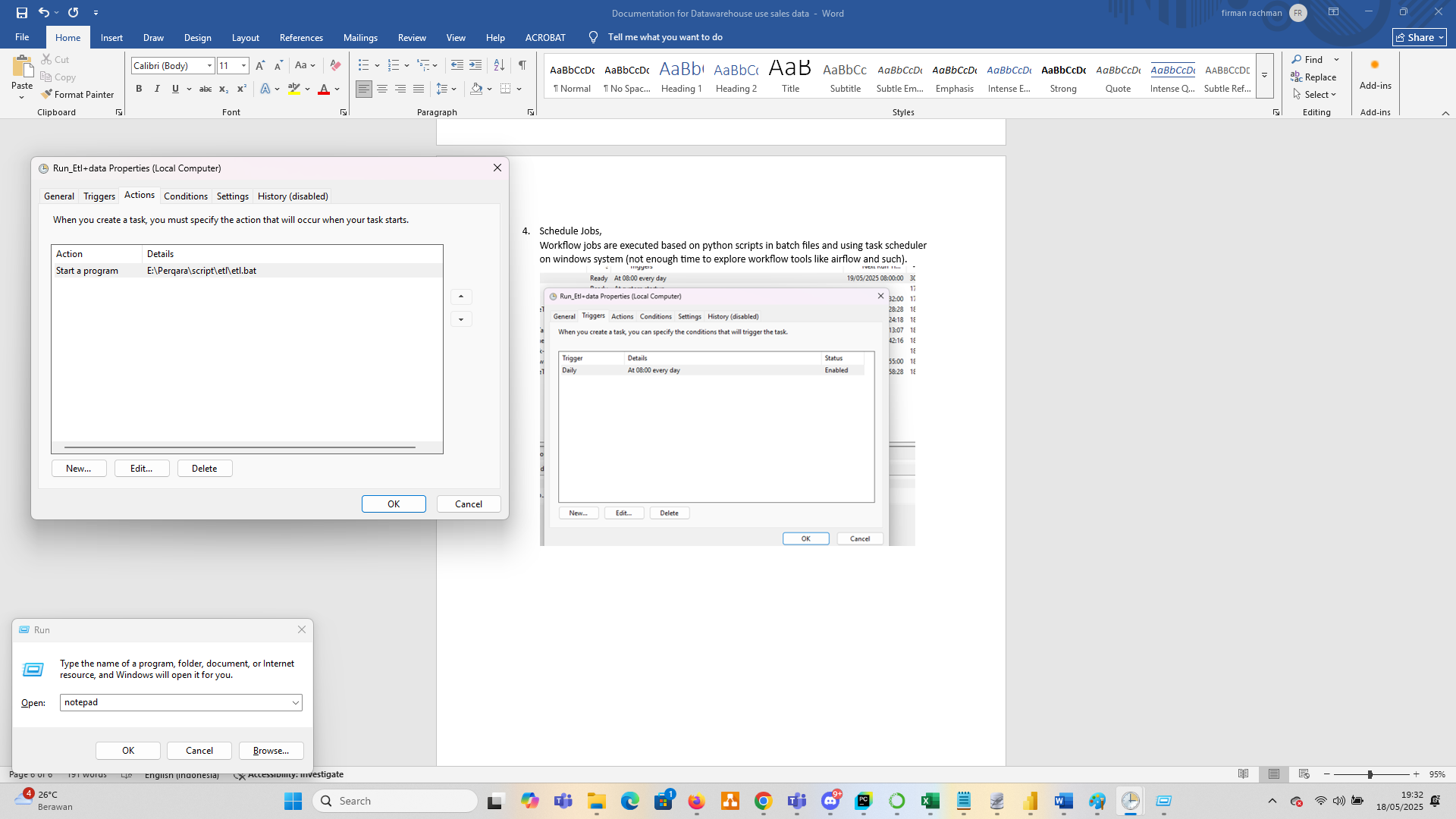
1. Report and Visualization Data, this is sample from Power BI Report Design
   1. Customer Profile

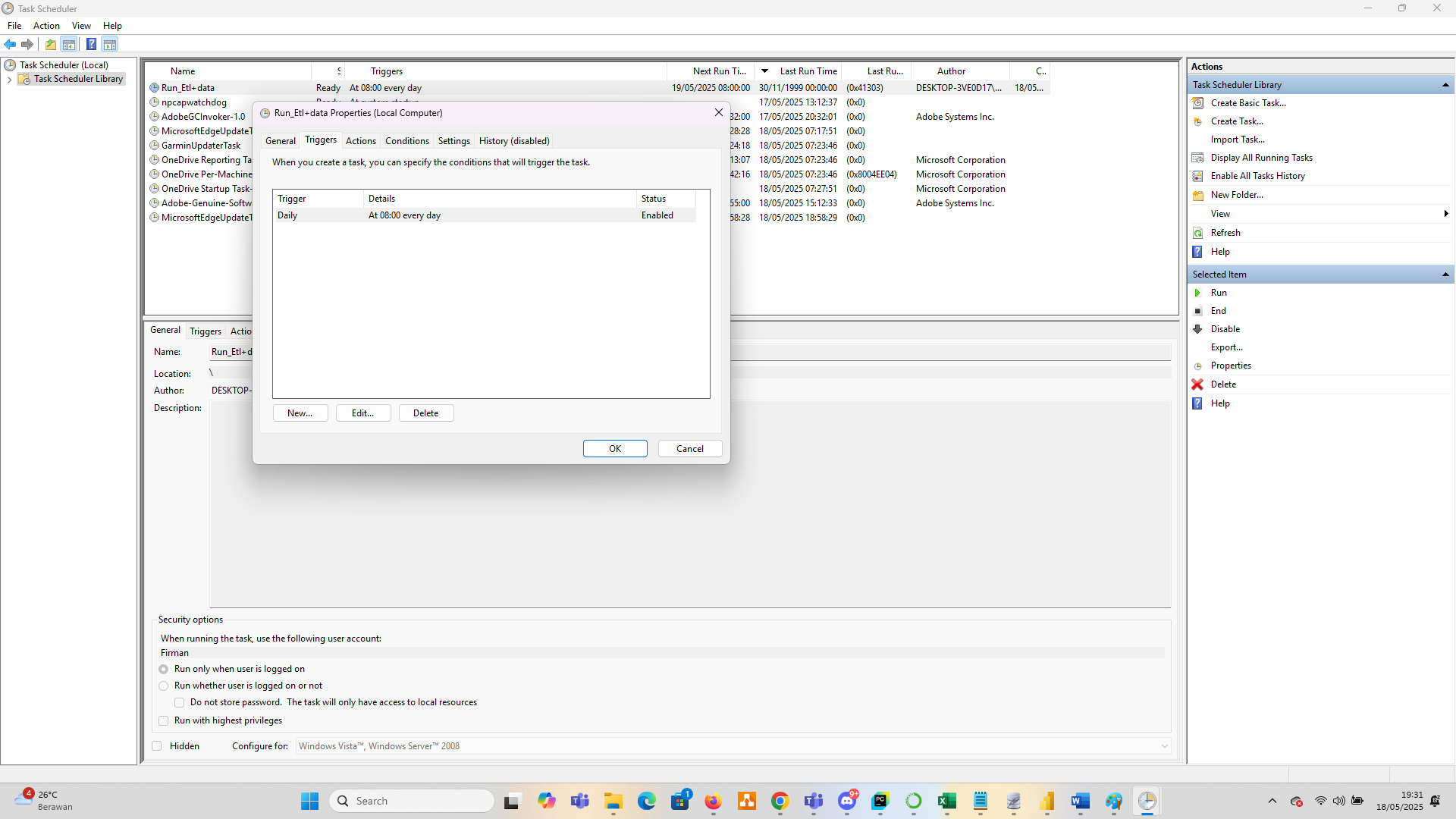


* 1. Interactive Sales Data Statistic Dashboard



1. Schedule Jobs,   
   Workflow jobs are executed based on python scripts in batch files and using task scheduler on windows system (not enough time to explore workflow tools such as airflow).





1. Attachment File, Check the supporting documents, including:
   1. data.zip (source data)
   2. exctract\_data.zip (python and batch file script)
   3. script\_databases\_sqlite.zip (sql file script)
   4. sales\_data\_perqara\_report.zip (Power BI report)
2. Installation instruction
   1. extract source data (data.zip)
   2. create sqlite database with the name sales\_data and then execute sql file to generate table
   3. extract python and batch file script and modified code for adjust data source location and sqlite databases
   4. modified connection in power BI report to access your datamart in sqlite databases
   5. Modified batch file to adjus script python location
   6. Testing batch file to run program
   7. Create task schedule in windows system to run program autamatically (call batch file).