

RETAIL INVENTORY MANAGEMENT AND FORECASTING

PROJECT





OVERVIEW

1

Our Team

2

About Project

3

Objectives

4

Project Steps

1

OUR TEAM



**MOHAMED
MAGDY**

Database



**ABDULLAH
HATEM**

ETL



**MAHMOUD
SAMIR**

ETL



**AHMED
YASSER**

ML on Azure



**MANSOUR
ELASHRY**

Dashboard

2

ABOUT PROJECT



This project focuses on delivering a comprehensive analysis of business data to improve decision-making through a combination of database management, data transformation, machine learning, and data visualization.

3

OBJECTIVES

01

Efficient data storage and retrieval through well-structured SQL tables.

02

Streamlined data processing with ETL workflows using SSIS.

03

Predictive analytics powered by machine learning models in Azure.

04

Clear, actionable insights provided through an interactive Power BI dashboard.



PROJECT STEPS

4

STEP 1

Creating Tables with SQL

SQL was used to create structured tables that store customer, order, and product data, ensuring easy data management and querying.

STEP 2

ETL using SSIS

Data from various sources was extracted, transformed, and loaded (ETL) into a centralized data warehouse using SQL Server Integration Services (SSIS), ensuring the data is clean and ready for analysis.

STEP 3

Machine Learning with Python in Azure

Python scripts were implemented in Azure to build machine learning models, enabling predictive analysis and deeper insights into business trends.

STEP 4

Power BI Dashboard

Finally, a Power BI dashboard was created to visualize key metrics and performance indicators, making data easily accessible and understandable for stakeholders.

03 STEP 1

Database Creation

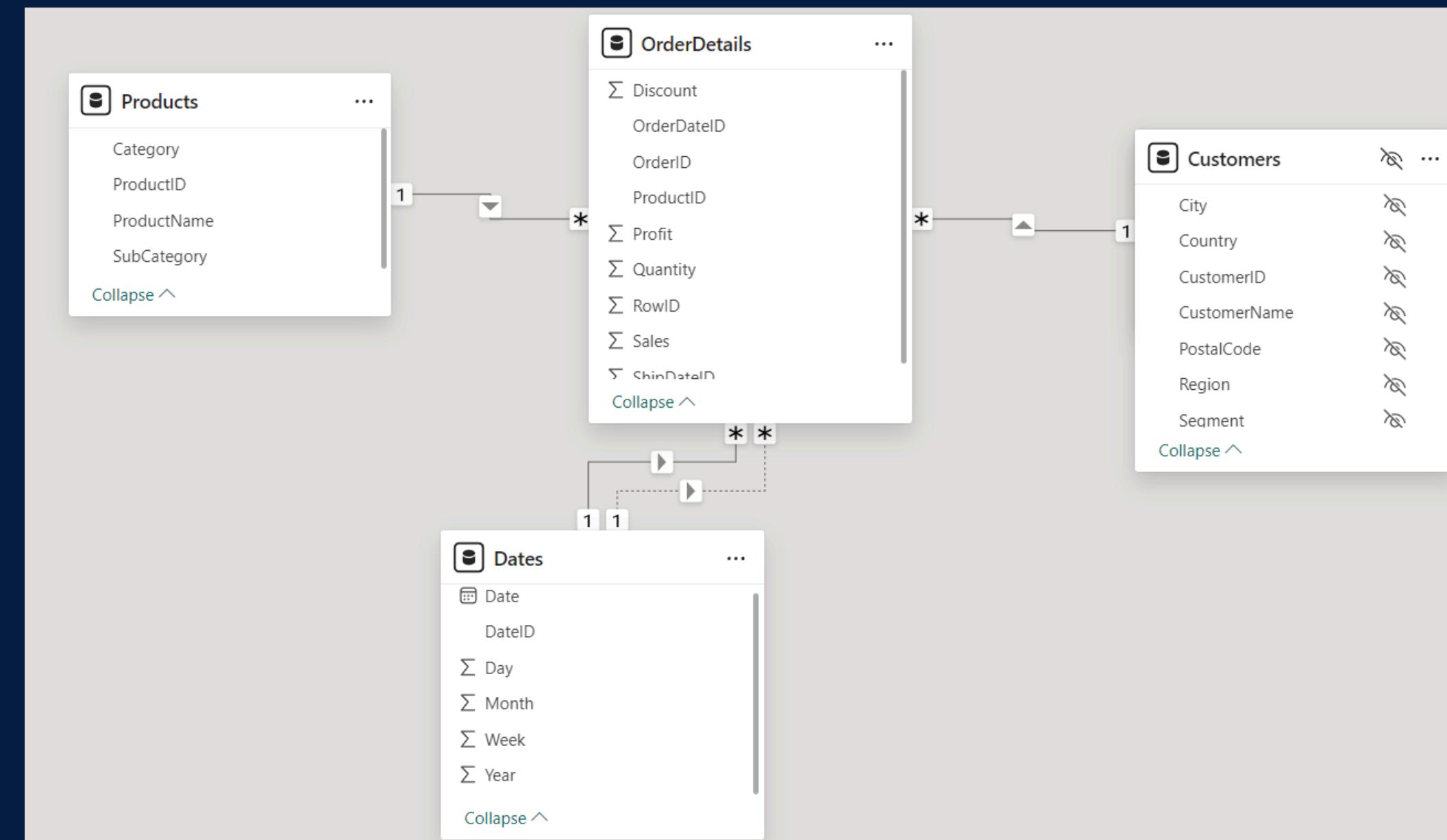
- 1) Imported the dataset as CSV file to SQL Server environment
- 2) Normalized the dataset into smaller tables based on relevant details
- 3) Used insert based on select statement to avoid the hassel of bulk insertion

```
CREATE TABLE Customers (
    CustomerID NVARCHAR(10) PRIMARY KEY,
    CustomerName NVARCHAR(255),
    Segment NVARCHAR(50),
    Country NVARCHAR(50),
    City NVARCHAR(100),
    Region NVARCHAR(50),
    PostalCode NVARCHAR(20)
);
```

```
FOREIGN KEY (ProductID) REFERENCES Product(ProductID),
);
INSERT INTO Customers (CustomerID, CustomerName, Segment, Country, City, Region, PostalCode)
SELECT
    Customer_ID,
    Customer_Name,
    Segment,
    Country,
    City,
    Region,
    ISNULL(Postal_Code, '00000') -- Ensure no NULLs in PostalCode
FROM superstore_cleaned;
```

03 STEP 1

Database Schema



03 STEP 1

Database Exploration

```
Exploration.sql -...(MAGDY\muham (52)) ✎ X
--What is the total sales amount for each product in the Pr
SELECT p.ProductName, SUM(od.Sales) AS TotalSales
FROM Products p
JOIN OrderDetails od ON p.ProductID = od.ProductID
GROUP BY p.ProductName;

--How many orders were placed by each customer?
SELECT c.CustomerName, COUNT(o.OrderID) AS NumberOfOrders
FROM Customers c
LEFT JOIN Orders o ON c.CustomerID = o.CustomerID
GROUP BY c.CustomerName;

--What are the top 5 products by total profit?
SELECT p.ProductName, SUM(od.Profit) AS TotalProfit
FROM Products p
JOIN OrderDetails od ON p.ProductID = od.ProductID
GROUP BY p.ProductName
ORDER BY TotalProfit DESC
OFFSET 0 ROWS FETCH NEXT 5 ROWS ONLY;
```

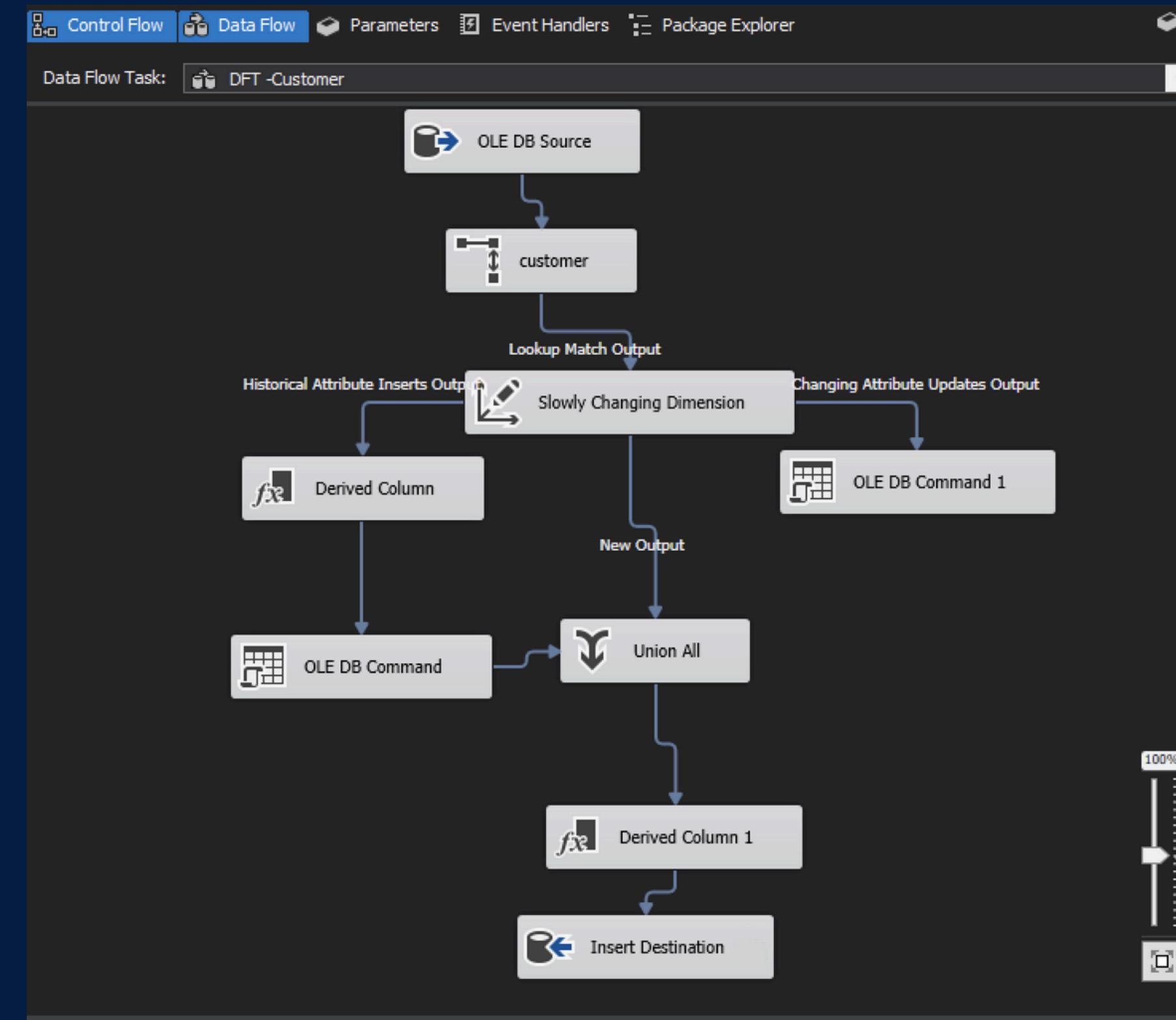
	ProductName	TotalProfit
1	Canon imageCLASS 2200 Advanced Copier	3919.99
2	Fellowes PB500 Electric Punch Plastic Comb Binding Mac...	2414.89
3	Ativa V4110MDD Micro-Cut Shredder	2400.97
4	GBC DocuBind TL300 Electric Binding System	2107.93
5	Canon imageCLASS MF7460 Monochrome Digital Laser ...	1995.99

Top 5 products by total profit

03 STEP 2

ETL

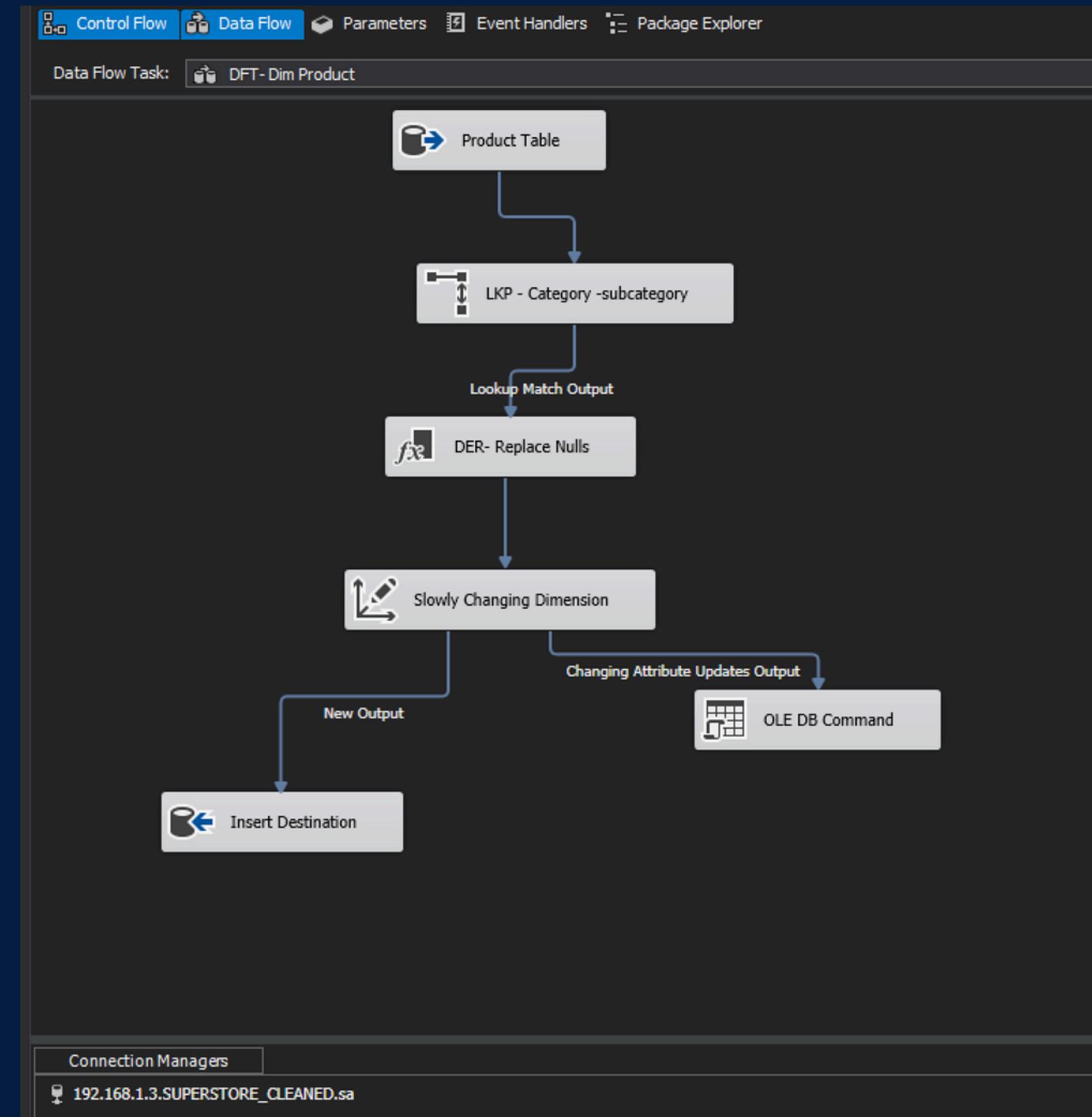
Dim customer package
Cleaning Data & Selecting
Data types



03 STEP 2

ETL

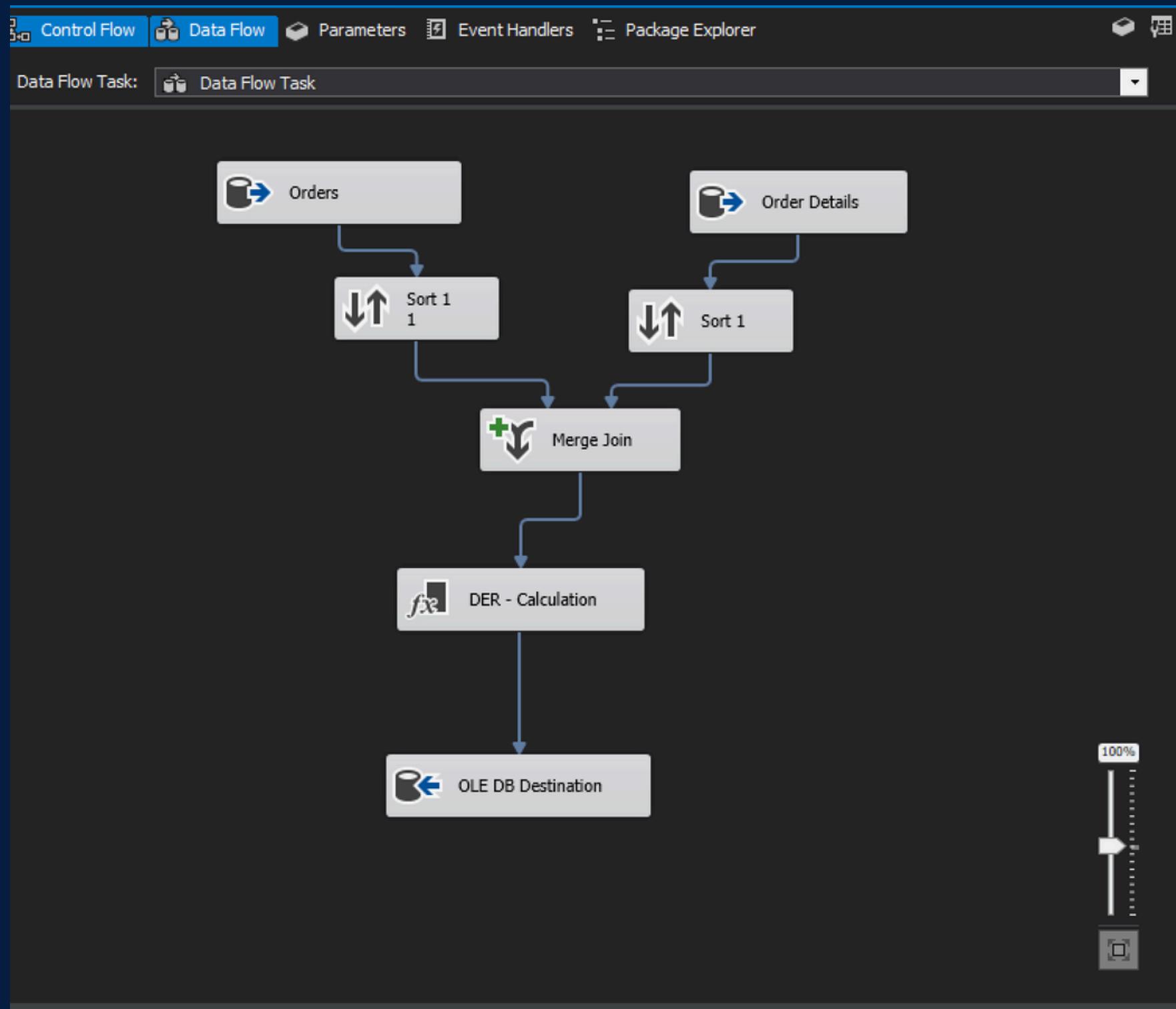
Dim product package
Cleaning Data & Selecting
Data types.



03 STEP 2

ETL

Fact sales package
Merge & Calculation
process



03 STEP 3

ML on Azure

- import library
- show dataset

Azure AI | Machine Learning Studio

Egypt University of Informatics > WC001 > Notebooks

MLmodel.ipynb

Compute: ai304121212048951 - Stopped

No kernel connected

Viewing Last saved 9 minutes ago

```
Requirement already satisfied: joblib>=1.2.0 in /anaconda/envs/azureml_py38/lib/python3.10/site-packages (from scikit-learn) (1.2.0)
Requirement already satisfied: six>=1.5 in /anaconda/envs/azureml_py38/lib/python3.10/site-packages (from python-dateutil>=2.7.3->pandas) (1.16.0)
```

1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import warnings
5 import scipy.stats as stats
6
7 warnings.filterwarnings('ignore')
8 warnings.filterwarnings("ignore", message="A NumPy version >=1.16.5 and <1.23.0 is required for this version of SciPy")
9
10 from sklearn.preprocessing import MinMaxScaler
11 from sklearn.model_selection import train_test_split
12 from sklearn.preprocessing import StandardScaler, OneHotEncoder
13 from sklearn.compose import ColumnTransformer
14 from sklearn.pipeline import Pipeline
15 from sklearn.linear_model import LinearRegression
16 from sklearn.tree import DecisionTreeRegressor
17 from sklearn.ensemble import RandomForestRegressor
18 from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
19 from sklearn.impute import SimpleImputer

Notebooks

Automated ML
Designer
Prompt flow

Assets

Data
Jobs
Components
Pipelines
Environments
Models
Endpoints

1
2 import mltable
3 from azure.ai.ml import MLClient
4 from azure.identity import DefaultAzureCredential
5
6 ml_client = MLClient.from_config(credential=DefaultAzureCredential())
7 data_asset = ml_client.data.get("super", version="1")
8
9 tbl = mltable.load(f'azurerm:/{{data_asset.id}}')
10
11 df = tbl.to_pandas_dataframe()
12 df

[23]

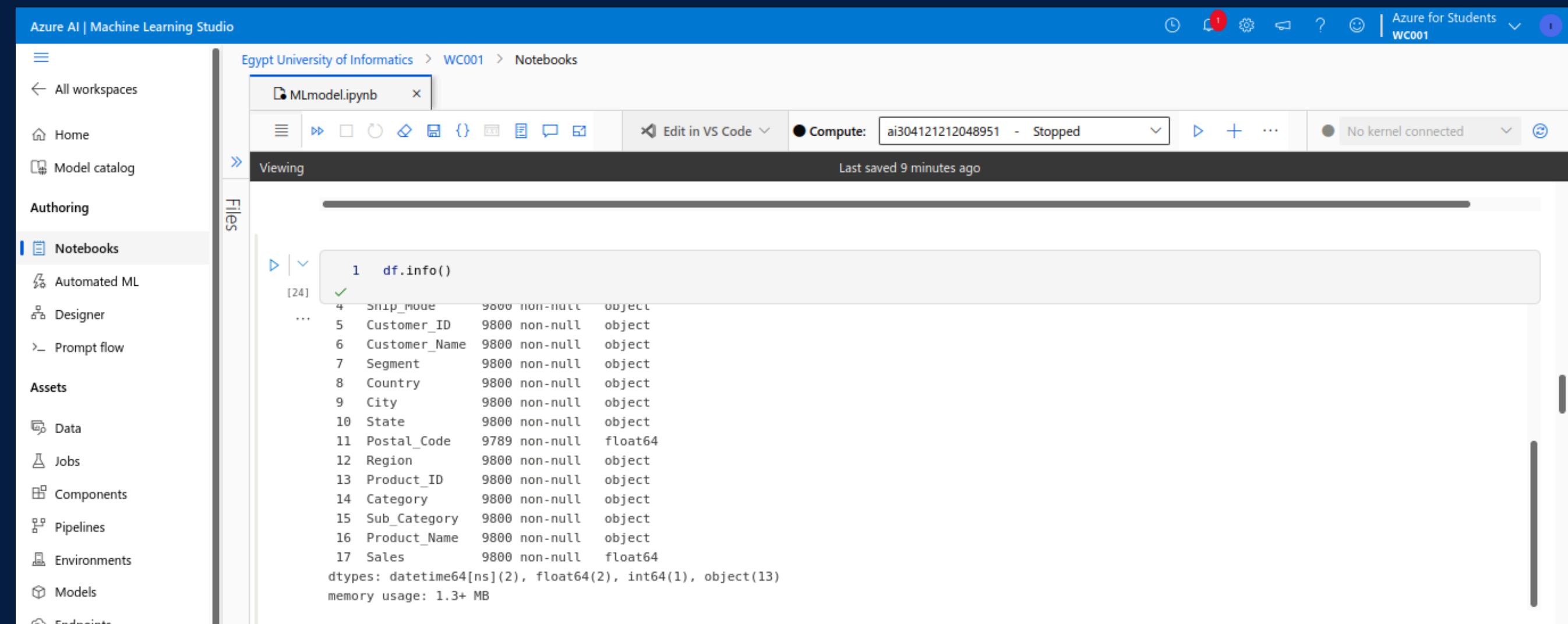
Found the config file in: /config.json

Row_ID	Order_ID	Order_Date	Ship_Date	Ship_Mode	Customer_ID	Customer_Name	Segment	Country	City	State	Postal_Code	Region	Product_ID	Category	Sub_Category	Product_Name	
0	1	CA-2017-152156	2017-11-08	2017-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase
1	2	CA-2017-152156	2017-11-08	2017-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs,...
2	3	CA-2017-138688	2017-06-12	2017-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036.0	West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Tumblers

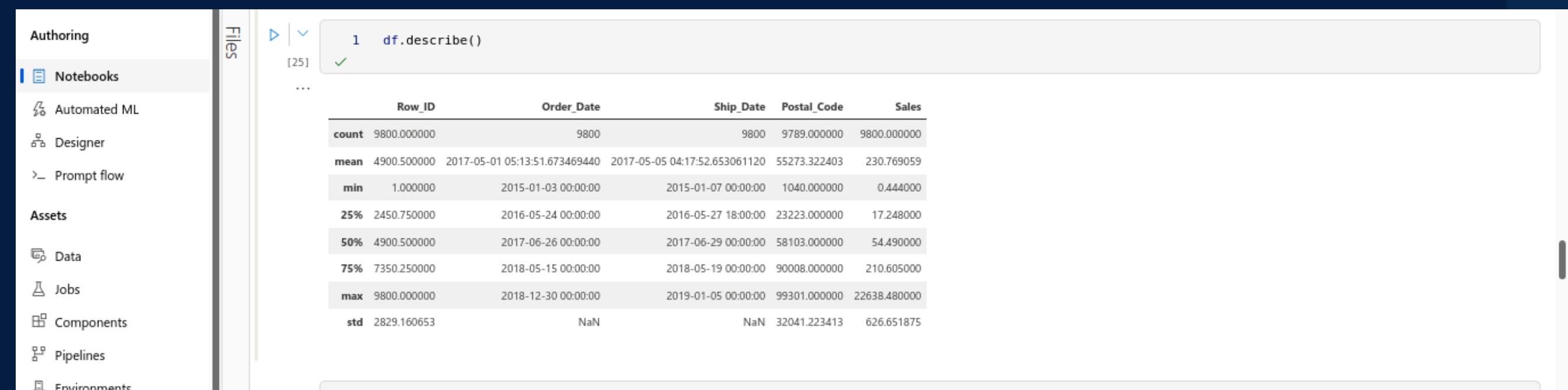
03 STEP 3

ML on Azure

- data info



```
df.info()
[24]
1   df.info()
...
4   Ship_Mode      9800 non-null  object
5   Customer_ID    9800 non-null  object
6   Customer_Name   9800 non-null  object
7   Segment        9800 non-null  object
8   Country         9800 non-null  object
9   City            9800 non-null  object
10  State           9800 non-null  object
11  Postal_Code    9789 non-null  float64
12  Region          9800 non-null  object
13  Product_ID     9800 non-null  object
14  Category        9800 non-null  object
15  Sub_Category   9800 non-null  object
16  Product_Name   9800 non-null  object
17  Sales           9800 non-null  float64
dtypes: datetime64[ns](2), float64(2), int64(1), object(13)
memory usage: 1.3+ MB
```



```
df.describe()
[25]
1   df.describe()
...
Row_ID          Order_Date       Ship_Date  Postal_Code      Sales
count  9800.000000             9800        9800  9789.000000  9800.000000
mean   4900.500000  2017-05-01 05:13:51.673469440  2017-05-05 04:17:52.653061120  55273.322403  230.769059
min    1.000000   2015-01-03 00:00:00  2015-01-07 00:00:00  1040.000000  0.444000
25%   2450.750000   2016-05-24 00:00:00  2016-05-27 18:00:00  23223.000000  17.248000
50%   4900.500000   2017-06-26 00:00:00  2017-06-29 00:00:00  58103.000000  54.490000
75%   7350.250000   2018-05-15 00:00:00  2018-05-19 00:00:00  90008.000000  210.605000
max   9800.000000   2018-12-30 00:00:00  2019-01-05 00:00:00  99301.000000  22638.480000
std   2829.160653                    NaN        NaN  32041.223413  626.651875
```

03 STEP 3

ML on Azure

- Data Preparation
- Train-Test Split
- Identifying Numeric and Categorical Features
- Preprocessing Pipelines
- Defining Models
- Training and Evaluation

The screenshot shows the Azure AI | Machine Learning Studio interface. On the left, there's a sidebar with sections like 'All workspaces', 'Home', 'Model catalog', 'Authoring' (which is expanded to show 'Notebooks', 'Automated ML', 'Designer', 'Prompt flow', 'Assets' (with 'Data', 'Jobs', 'Components', 'Pipelines', 'Environments', 'Models', 'Endpoints'), and 'Manage'), and 'Compute' (set to 'ai304121212048951 - Stopped'). The main area is a Jupyter Notebook titled 'MLmodel.ipynb'. The code in the notebook is as follows:

```
1
2
3 rmse_results = {}
4 mae_results = {}
5 r2score_results = {}
6
7 X = df.drop(columns=['Sales'])
8 y = df['Sales']
9
10 # Split the data into training and testing sets
11 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
12
13 # Preprocessing steps for numeric and categorical features
14 numeric_features = X_train.select_dtypes(include=[np.number]).columns.tolist()
15 categorical_features = X_train.select_dtypes(include=['object']).columns.tolist()
16
17 numeric_transformer = Pipeline(steps=[
18     ('imputer', SimpleImputer(strategy='mean')),
19     ('scaler', StandardScaler())
20 ])
21
22 categorical_transformer = Pipeline(steps=[
23     ('imputer', SimpleImputer(strategy='most_frequent')),
24     ('encoder', OneHotEncoder(handle_unknown='ignore'))
25 ])
26
27 preprocessor = ColumnTransformer(transformers=[
28     ('num', numeric_transformer, numeric_features),
29     ('cat', categorical_transformer, categorical_features)
30 ])
31
32
33 models = {
34     'Linear Regression': LinearRegression(),
35     'Decision Tree': DecisionTreeRegressor(random_state=42),
36     'Random Forest': RandomForestRegressor(random_state=42)
37 }
38
39 results = {}
40
41 for model_name, model in models.items():
42     # Create a pipeline
43     model_pipeline = Pipeline(steps=[
44         ('preprocessor', preprocessor),
45         ('model', model)
46     ])
47
48     # Fit the model
49     model_pipeline.fit(X_train, y_train)
50
51     # Make predictions on the test data
52     y_pred = model_pipeline.predict(X_test)
53
54     # Evaluate the model using MSE, R-squared, MAE
55     rmse = mean_squared_error(y_test, y_pred)
56     mae = mean_absolute_error(y_test, y_pred)
57     r2 = r2_score(y_test, y_pred)
58
59     # Store the results
60     rmse_results[model_name] = rmse
61     mae_results[model_name] = mae
62     r2score_results[model_name] = r2
63
64     print("Model : ", model_name)
65     print(f"RMSE : {rmse:.2f}")
66     print(f"MAE : {mae:.2f}")
67     print(f"r2 : {r2:.2f}")
68     print("-----")
```

This part of the screenshot continues the Jupyter Notebook 'MLmodel.ipynb'. The sidebar remains the same. The code continues from the previous snippet:

```
69
```

03 STEP 4

The DAX Measures

✓ Key Measures

✓ Matrics

Profit

Sales

✓ Previous Year

Profit PY

Sales PY

✓ VS Previous Year

VS PY - Profit

VS PY - Sales

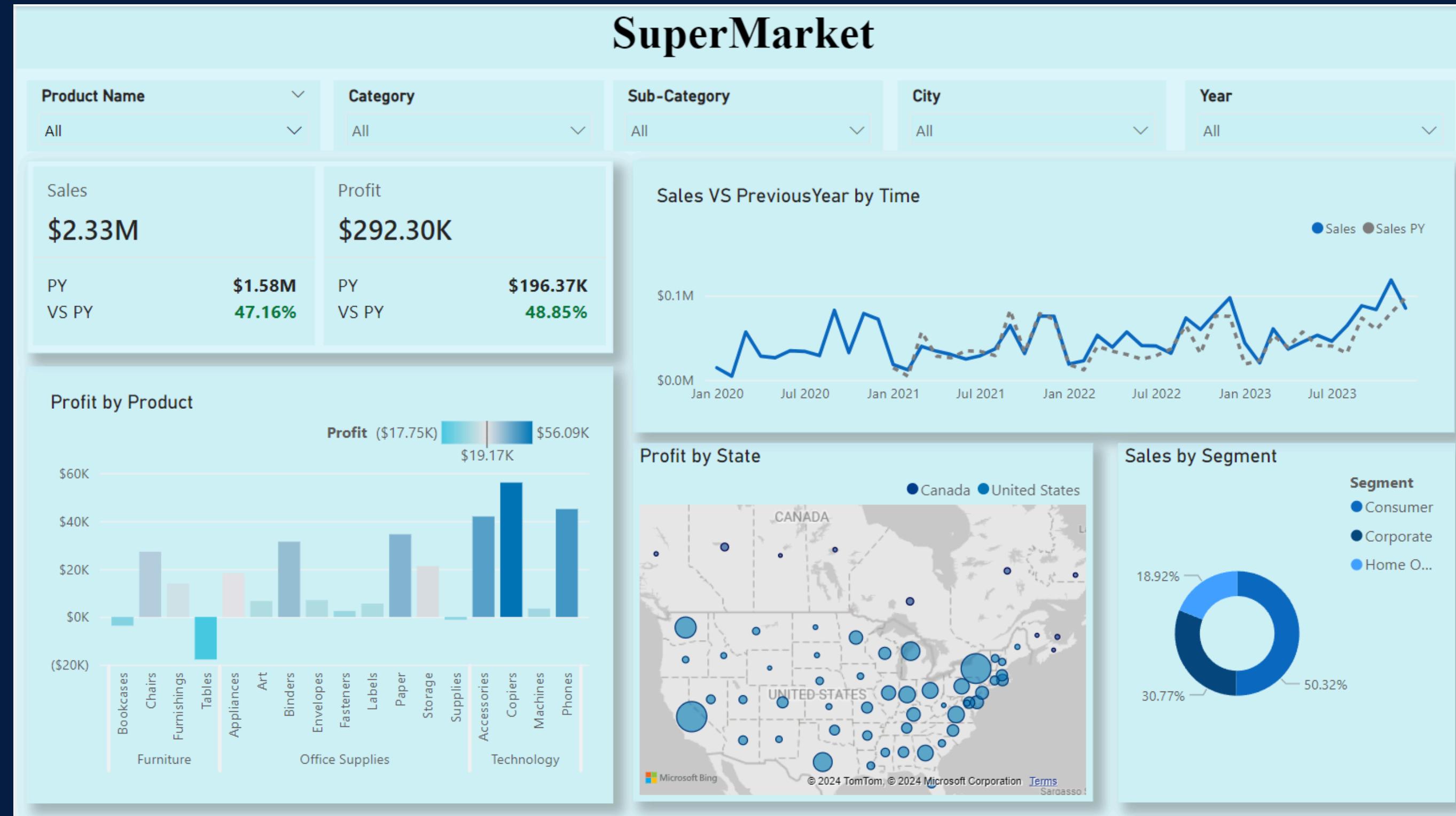
```
1 Sales = SUM(Orders[Sales])
```

```
1 Sales PY =
2 CALCULATE(
3     [Sales],
4     SAMEPERIODLASTYEAR('Date Table'[Date]))
```

```
1 VS PY - Sales =
2 DIVIDE( [Sales] - [Sales PY] , [Sales PY])
```

03 STEP 4

The Dashboard



03 STEP 4

The Key Points

Interactive Visualization

Allows for dynamic filtering by product, category, sub-category, city, and year.

Key Performance

Displays total sales, profit, and year-over-year comparisons.

Sales and Profit Trends

Graphical insights into sales over time, segmented by product and region.

Geographical Breakdown

Visual map of profit by state with a comparison between the U.S. and Canada.

Customer Segmentation

Insights into sales distribution across consumer, corporate, and home office segments.

THANK YOU

