SUPER-STORE DATA WAREHOUSE

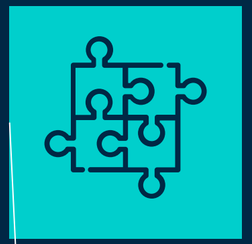
PRESENTED BY :

Ahmed Bensalah
Omar Bouattour

SUPERVIZED BY:

MRS Senda Bouaziz
MR Ali Salem

TABLE OF CONTENTS



01

PROBLEMS & SOLUTION

02

OUR PROCESS

03

ANALYSES

OUR STORE

The store represents a multi-category retail store that sells a variety of products, ranging from office supplies to furniture and technology.

UNDERSTANDING THE PROBLEM

In a retail environment, data often originates from multiple sources like sales transactions, customer databases, inventory systems, and shipping logs. These datasets may be stored in different formats .

Combining and analyzing this data directly can be challenging due to inconsistencies and fragmentation.

As the dataset grows , running complex queries to analyze sales trends, customer behavior, or inventory performance becomes slow and inefficient in traditional transactional databases.

WHAT IS THE SOLUTION ?

Data Warehouse

A data warehouse solves these challenges by providing a **centralized**, **scalable repository** that integrates data from multiple sources

WHY DATA WAREHOUSE ?

Time Savings

Cost Savings

Enhanced Data Quality

Performance Optimization

TECHNOLOGIES USED

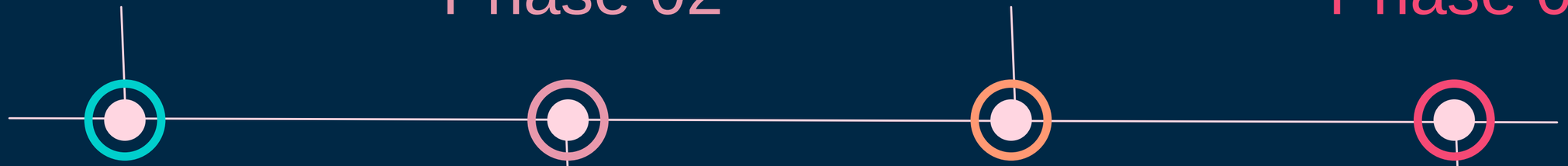
POSTGRE SQL

POWER BI

PENTAHO

PYTHON

OUR PROCESS



Phase 01

Design

Relational and dimensional models

Phase 02

ETL Pipelines

Create multiple ETL pipelines with PDI and Python

Phase 03

OLAP Cube

Create the OLAP cube using Pentaho Schema workbench

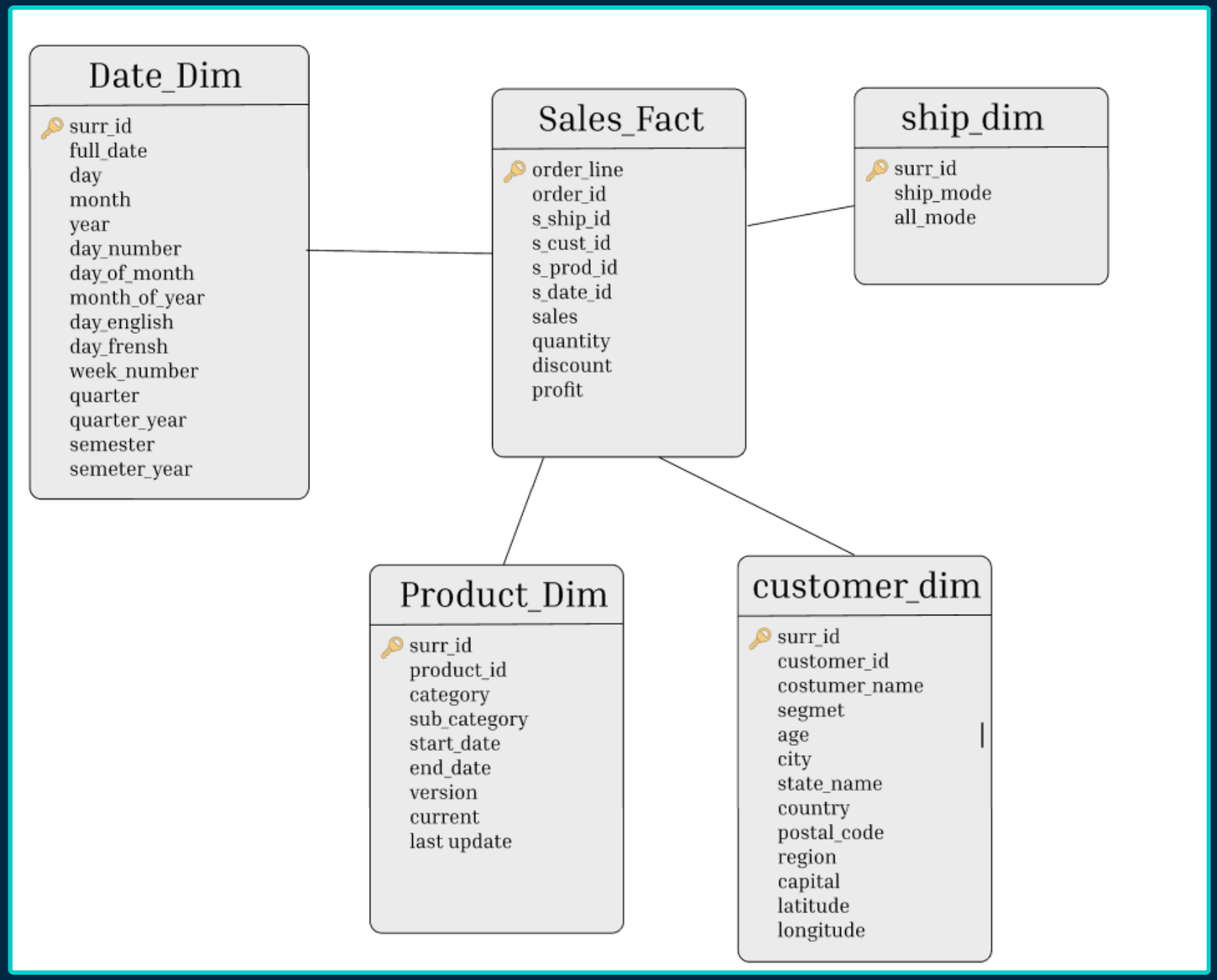
Phase 04

MDX queries

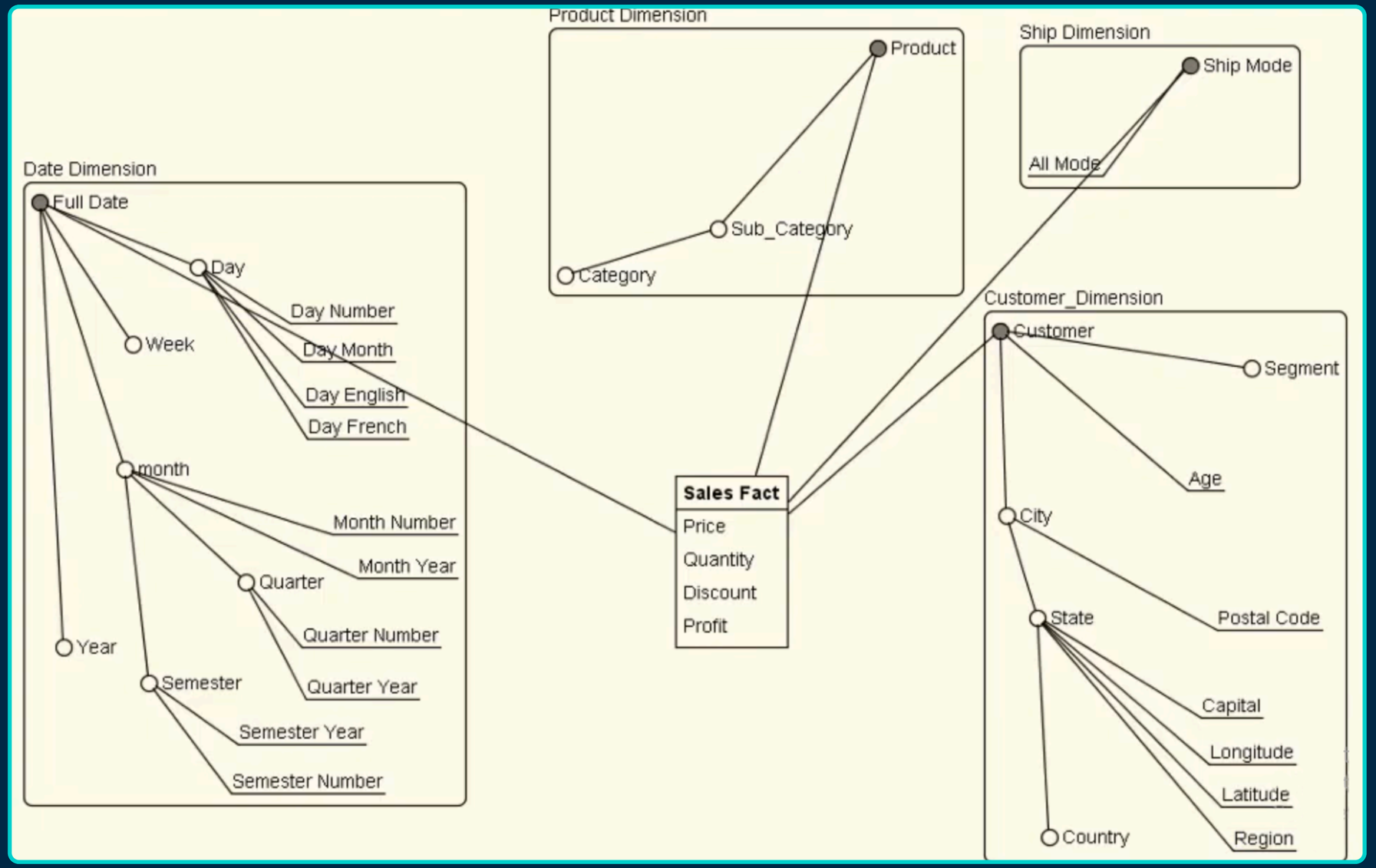
Prepare some MDX queries for the analysis

1) DESIGN

Relational schema Data Warehouse



Dimensional Model



2) ETL PIPELINES

WITH PENTAHO DATA INTEGRATION

LIVE DEMO

WITH PYTHON

https://github.com/OmarBouat/Dw_Sales_Pr_AhmedBS_OmarB/tree/main

3) OLAP CUBE

WITH PENTAHO SCHEMA WORKBENCH

LIVE DEMO

4) MDX QUERIES

WHAT IS MDX ?

+ MDX :(Multidimensional Expressions) is a query language used to interact with multidimensional data structures, typically in OLAP (Online Analytical Processing) .

Query 1 : Top 5 Customers by Sales

This query lists the top 5 customers by their total sales.

```
SELECT
    [Measures].[Sales] ON COLUMNS,
    TOPCOUNT([Customer].[Customer_name].Members, 5, [Measures].[Sales]) ON ROWS
FROM [Sales_psw]
```

```
Axis #0:
{}
Axis #1:
{[Measures].[Sales]}
Axis #2:
{[Customer].[Sean Miller]}
{[Customer].[Tamara Chand]}
{[Customer].[Raymond Buch]}
{[Customer].[Tom Ashbrook]}
{[Customer].[Adrian Barton]}
Row #0: 25 043,05
Row #1: 19 052,218
Row #2: 15 117,339
Row #3: 14 595,62
Row #4: 14 473,571
```

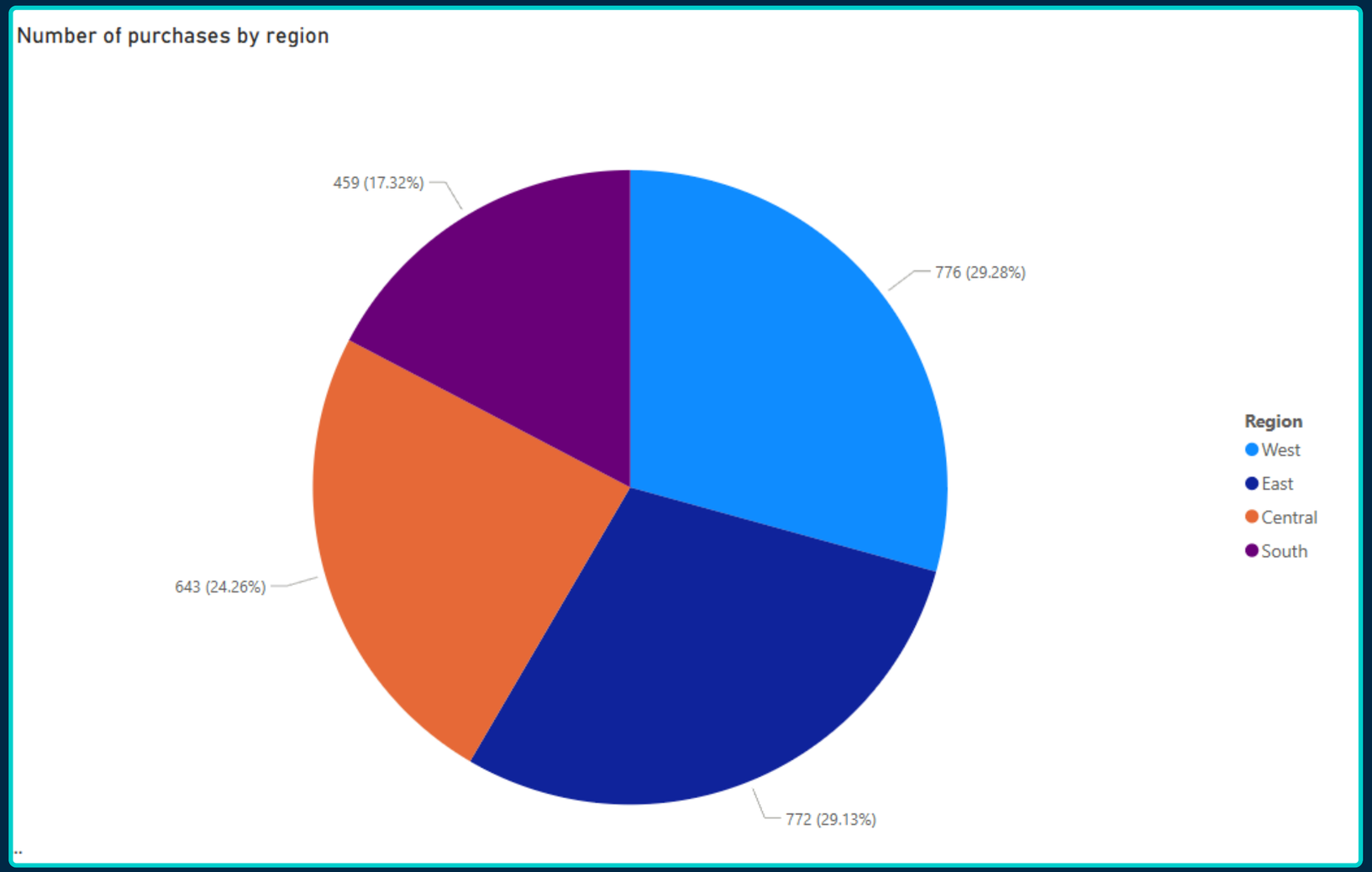
Query 2 : Profit by Shipping Mode

This query retrieves the profit for each shipping mode.

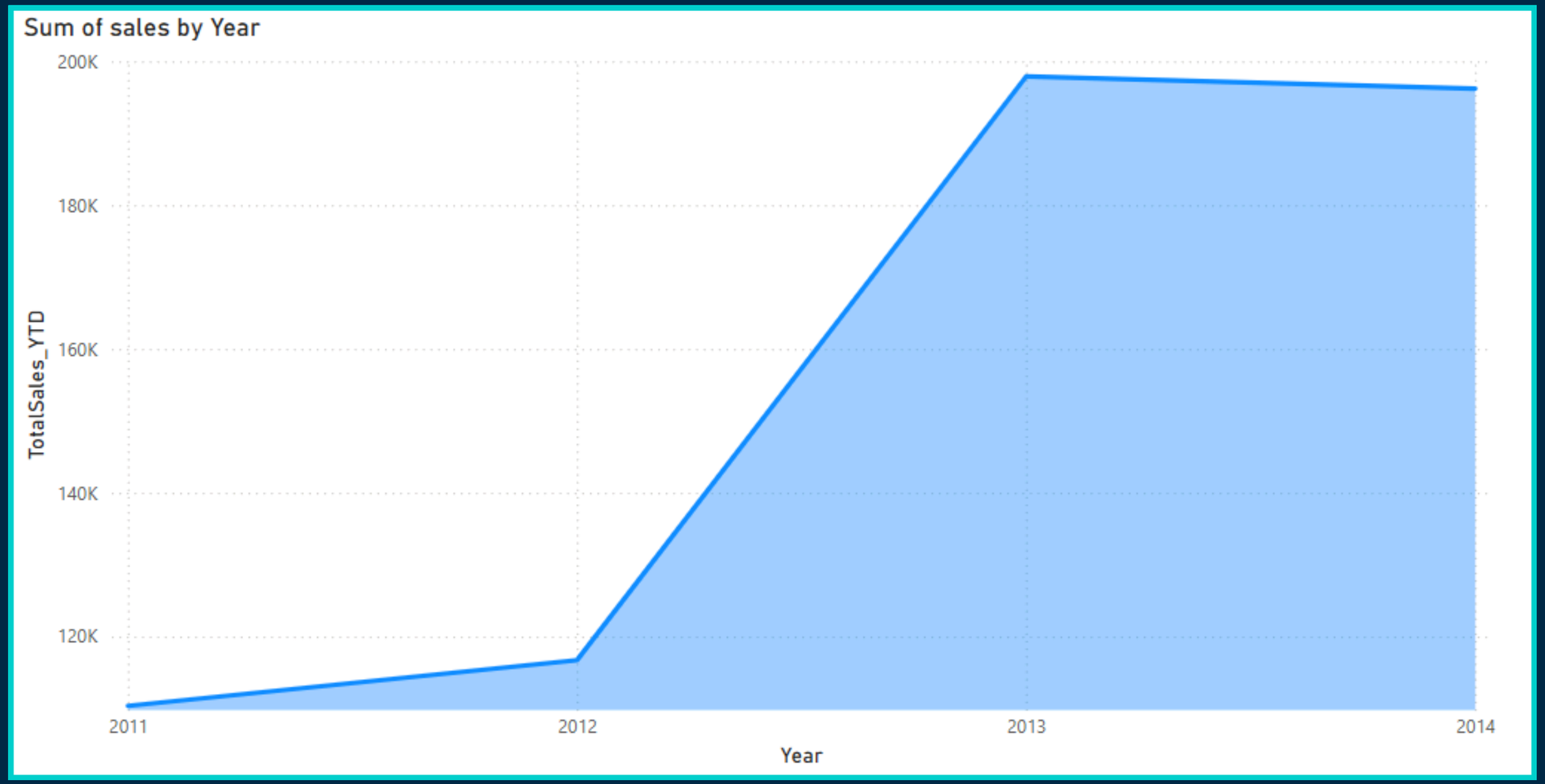
```
SELECT
    [Measures].[Profit] ON COLUMNS,
    [Ship].[Mode].Members ON ROWS
FROM [Sales_psw]
```

```
Axis #0:
{}
Axis #1:
{[Measures].[Profit]}
Axis #2:
{[Ship].[First Class]}
{[Ship].[Same Day]}
{[Ship].[Second Class]}
{[Ship].[Standard Class]}
Row #0: 48 760,836
Row #1: 15 749,247
Row #2: 57 178,268
Row #3: 164 474,964
```

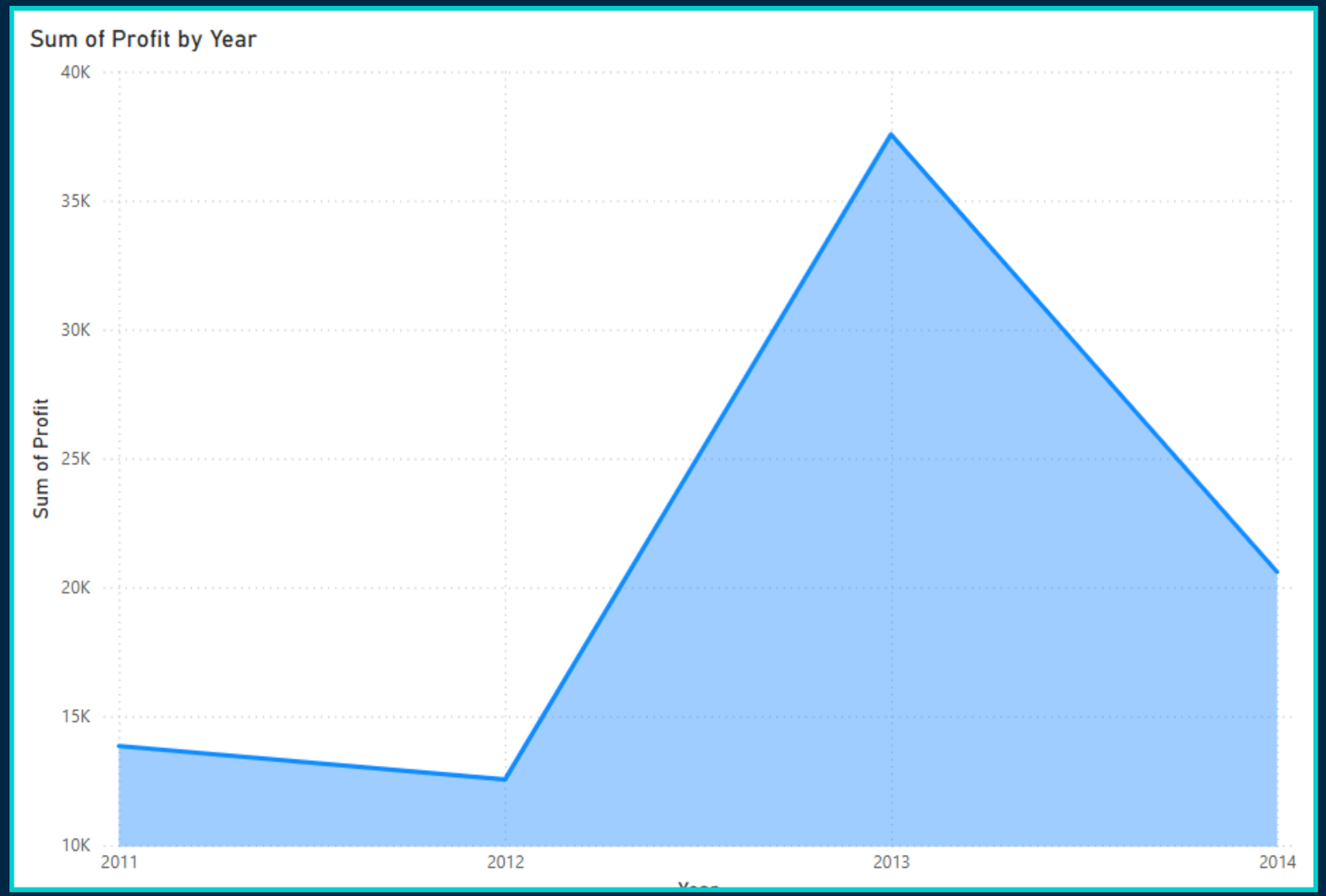
Analysis with Power BI



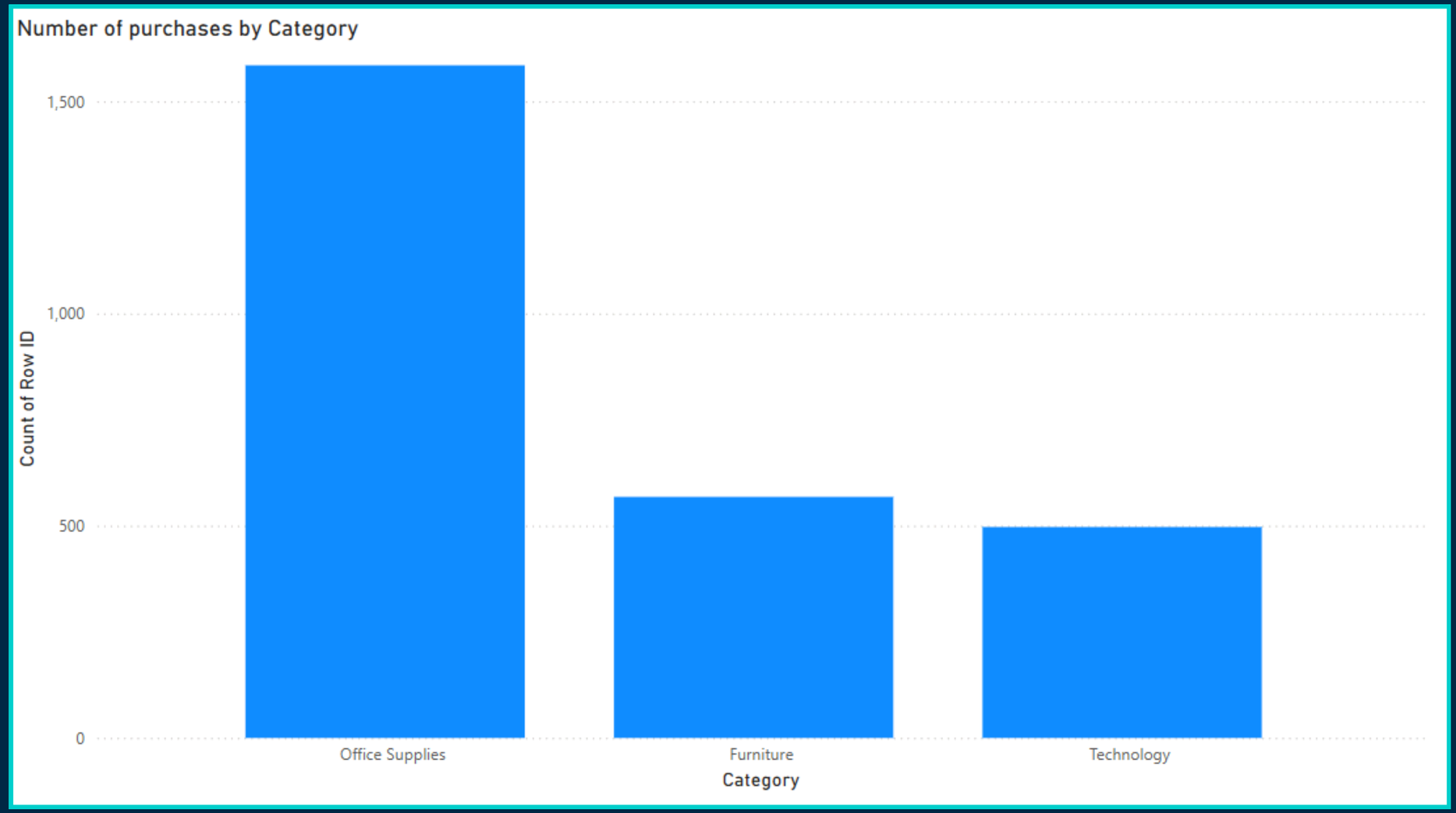
Analysis with Power BI



Analysis with Power BI



Analysis with Power BI



THANKS FOR YOUR ATTENTION