Comprehensive Sales Performance and Growth Analysis

Introduction

Background

Analyzing sales data is pivotal in today's competitive business landscape, offering insights into market trends, customer preferences, and overall business performance.

Objective

The main objective is to analyze year-over-year sales growth, aiming to uncover trends and patterns that can inform business strategies.

Data Overview

Data Sources

The project utilizes the "Adventure Works in Excel Tables" dataset from Kaggle, consisting of comprehensive sales-related data, converted and loaded into a MySQL database.

Data Description

The MySQL database contains several tables including Products, Employees, Regions, Targets, SalesPersonRegion, Resellers, and Sales. This data specifically pertains to the sales of bicycles and their components across various regions: Australia, Canada, Central, France, Germany, Northeast, Northwest, Southeast, Southwest, and the United Kingdom.

Methodology

Database Design

The Adventure Works database schema is designed to efficiently manage and relate sales data. The schema includes the following tables and key relationships:

Products Table: Stores product details.

Primary Key: ProductKey

Relationships: Connected to the Sales table.

Employees Table: Contains employee and salesperson information.

Primary Key: EmployeeKey

Relationships: Linked to the Sales, Targets, and SalesPersonRegion tables.

Regions Table: Details sales territories.

Primary Key: SalesTerritoryKey

Relationships: Linked to the SalesPersonRegion and Sales tables.

Sales Table: Records sales transactions.

Composite Key: Includes SalesOrderNumber, ProductKey, ResellerKey, EmployeeKey, SalesTerritoryKey

Relationships: Ties together Products, Employees, Regions, and Resellers.

Resellers Table: Information about resellers.

Primary Key: ResellerKey

Relationships: Connected to the Sales table.

Targets and SalesPersonRegion Tables: Define targets for employees and link employees to their sales regions.

This schema ensures data integrity and supports complex queries involving multiple joins and aggregations, critical for this project's analysis goals.

Data Preparation

Initial data preparation, involving normalization and cleaning, was conducted in a Jupyter Notebook on Kaggle before downloading the data for MySQL database insertion.

Query Development

Developed SQL queries with joins, window functions, and aggregations for data retrieval and analysis.

Analysis

Year-over-Year Growth Analysis

Involved calculating the year-over-year sales growth for products, regions, and salespersons. A significant increase in total sales was observed in 2018, followed by a notable decrease in 2020.

Key Findings

In 2020, the highest number of salespeople not meeting their sales targets was observed, with 24 individuals falling short of their goals.

The top 3 regions where salespeople most frequently failed to meet targets were Southwest (80% - 4 out of 5 salespeople), Northwest (80% - 4 out of 5 salespeople), and Canada (75% - 3 out of 4 salespeople).

These findings highlight specific areas and periods where sales strategies and targets may need to be reassessed.

Additional Analysis Components

Sales Audit Log, Sales Summary View, Stored Procedures and Triggers

These components were created for efficiently managing data integrity, automated logging of sales activities, and streamlining data access.

Other Key Findings

Additional insights included the identification of top-selling products, sales performance metrics of salespersons, and sales trends over the years.

Sales Summary View

Established a SalesSummaryView to streamline the access and analysis of aggregated sales data.

Stored Procedures and Triggers

Implemented stored procedures and triggers to manage data insertion and summary updates efficiently.

Conclusions

The project yielded valuable insights into sales trends and efficiency, informing strategic business decisions. Notable were the findings of the significant sales increase in 2018 and the decrease in 2020.

Reflections and Learning

The project enhanced expertise in SQL, database management, and data analysis. The ability to interpret the EXPLAIN statement improved, aiding query optimization.

Future Work

Suggested areas for future analysis include investigating the causes of sales fluctuations and incorporating predictive analytics for forecasting.

Appendix

A. Complete SQL Code for Database Setup and Data Loading

Here are the complete SQL statements used for setting up the database schema and loading the data:

Table Creation:

```
CREATE TABLE Products (...);

CREATE TABLE Employees (...);

CREATE TABLE Regions (...);

-- Additional CREATE TABLE statements for Targets, SalesPersonRegion, Resellers, Sales...
```

Data Loading from CSV Files

```
LOAD DATA INFILE '...products.csv' INTO TABLE Products ...;

LOAD DATA INFILE '...employees.csv' INTO TABLE Employees ...;

-- Additional data loading statements...
```

B. SQL Code for Query Optimization (Indexes)

```
ALTER TABLE Employees ADD INDEX idx_title (Title);

-- Additional ALTER TABLE statements for creating indexes on other tables.
```

C. SQL Code for Complex Queries

```
-- Example of a complex query: Top 3 selling products in each Region
```

```
SELECT main.Region, main.Product, main.TotalSales
FROM (
  SELECT R.Region, P.Product, SUM(S.Sales) AS TotalSales, ...
) AS main
WHERE main.SalesRank <= 3;
-- Additional SELECT statements for analysis, such as Year-over-Year growth.
```

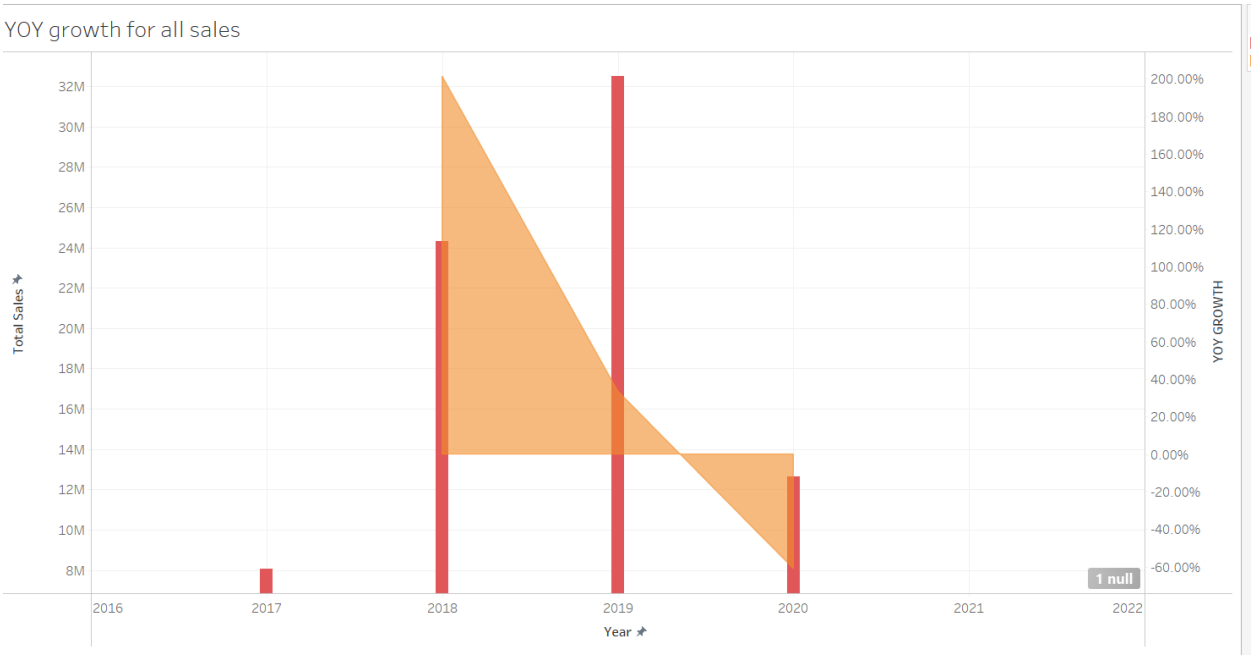
D. SQL Code for Stored Procedures and Triggers

```
CREATE PROCEDURE InsertSaleAndUpdateSummary(...);
CREATE TRIGGER LogUnrealisticSales ...;
-- Additional stored procedure and trigger definitions...
```

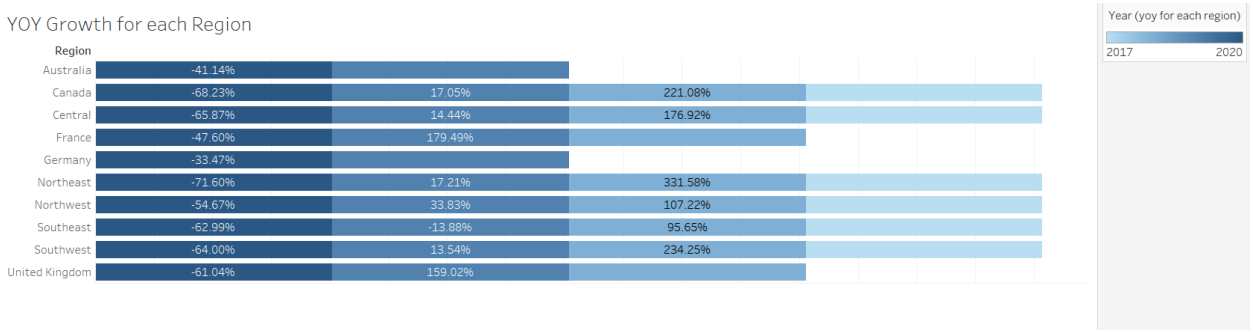
E. Visualization

Tableau visualizations that illustrate key findings from the data analysis.

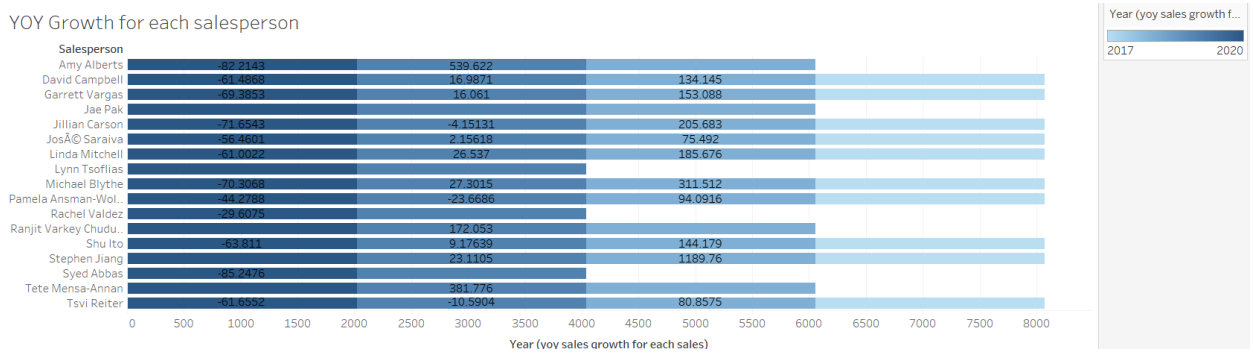
1st YOY growth for all sales:



2nd YOY Growth for each Region



3rd YOY Growth for each salesperson



Visualization in Tableau is available here:

https://public.tableau.com/app/profile/artur.melnyk/viz/YOYGrowthAnalysis/Dashboard1