**Project** **Report**

**On**

**Sales Performance Analysis using SQL and PowerBI**

Submitted

In Partial Fulfilments of

**Bachelor of Computer Applications (Artificial Intelligence & Machine Learning)**

1. **Submitted by:**

Name: Bhupender raghav

Roll No: 23/SCA/BCA (AI&ML)/010

**Under the Supervision of:**

Dr Priyanka Dadhich



**School of Computer Applications**

**Manav Rachna International Institute of Research and Studies**

**(DEEMED TO BE UNIVERSITY)**

Sector-43, Aravalli Hills

Faridabad – 121001

**Academic Session: 2023-2026**

**Declaration**

I do hereby declare that this project work entitled “Sales Performance Analysis using SQL and PowerBI” submitted by me for the partial fulfilment of the requirement for the award of **Bachelor of Computer Applications** is a record of my own work. The report embodies the finding based on my study and observation and has not been submitted earlier for the award of any degree or diploma to any Institute or University.

**SIGNATURE**

Name: Bhupender Raghav

Roll No: 23/SCA/BCA (AI/ML)/010

Date: 16/07/2025

**Certificate from the Guide**

This is to certify that the project report entitled **“Sales Performance Analysis using SQL and PowerBI**” submitted in partial fulfilment of the degree of **MASTER OF COMPUTER APPLICATIONS** to Manav Rachna International Institute of Research and Studies, Faridabad is carried out by **Mr. Bhupender Raghav** ,  
Roll No, 23/SCA/BCA(AI&ML)/010 under my guidance.

**Project Guide**

Mr. Sanskar Vats

**Head of Department**

Dr. Suhail Javed Qureshi

**ACKNOWLEDGEMENT**

I gratefully acknowledge for the assistance, cooperation, guidance and clarification provided by Mr. **Sanskar Vats** during the development of **Sales Performance Analysis using SQL and PowerBI.** My extreme gratitude to **Dr. Raj Kumar, Associate Professor & TPO** who guided us throughout the project. Without his willing disposition, spirit accommodation, frankness, timely clarification and above all faith in us, this project could not have been completed in due time. His readiness to discuss all important matters at work deserves special attention of.

I would like to extend my sincere gratitude to **Prof. (Dr.) Suhail Javed Quraishi – HOD, Prof. (Dr.) Rashmi Agrawal – Associate Dean and Prof. (Dr.) Brijesh Kumar – Dean** for their valuable teachings and advice. I want to thank all the department faculty members for their cooperation and support. I want to thank non-teaching staff of the department for their cooperation and support.

This opportunity is a big milestone in my career development. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement, to attain desired career objectives. I hope to continue cooperation with all of you in the future

|  |  |
| --- | --- |
| **Section** | **Title** |
| I. | **Cover Page** |
| II. | **System study** |
| III. | **Feasibility Study** |
| IV. | **Project Monitoring System (Gantt Chart)** |
| V. | **System Analysis** |
| VI. | **System Design** |
| VII. | **Input / Output Design** |
| VIII. | **System Testing** |
| IX. | **System Implementation** |
| X. | **Documentation** |
| XI. | **Scope of the Project** |
| XII. | **Bibliography** |

**1. Introduction**

Data analysis has become an indispensable tool for businesses aiming to understand their performance, identify trends, and make informed decisions. In the realm of sales, effective analysis can reveal customer behavior, product popularity, and areas for improvement, directly impacting revenue and strategic planning. This project focuses on performing a comprehensive sales performance analysis using a real-world dataset, leveraging the power of SQL for data cleaning and transformation, and PowerBI for interactive visualization.

**About the Project:** This project has been developed as a practical exercise in data analysis, demonstrating the end-to-end process from raw data to actionable insights. It simulates a business scenario where sales data needs to be processed, analyzed, and presented in an easily digestible format for stakeholders.

**Aim & Objectives:** The primary goal of this project is to analyze sales data to uncover key performance indicators, identify trends, and provide insights that can drive business decisions. The system is intended to:

* Extract and clean raw sales data from a relational database using SQL.
* Transform and prepare data for effective visualization and analysis.
* Develop an interactive sales dashboard in PowerBI to visualize key metrics.
* Identify top-performing products and customers.
* Analyze sales trends over time and against budget.
* Present insights through a user-friendly and interactive dashboard.

**Manpower:** This project was developed and implemented by a single student ([Your Name]) under the supervision of faculty members in the [Your Department Name] Department. All development, testing, and documentation tasks were completed independently by the student, simulating the role of a data analyst.

**2. System Study**

**Existing System along with Limitations:**

In many organizations, especially those without dedicated data analysis tools, sales reporting might rely on:

* **Manual Spreadsheet Analysis:** Data is extracted into Excel or similar spreadsheets and manually manipulated, filtered, and charted.
  + **Limitations:** Time-consuming, prone to human error, difficult to scale with large datasets, limited interactivity, challenging to update frequently, static reports.
* **Basic Database Queries:** Simple SQL queries might be run to pull raw data, but further analysis and visualization are often manual.
  + **Limitations:** Lack of visual insights, requires technical expertise to extract information, not easily shareable with non-technical stakeholders.
* **Ad-hoc Reporting:** Reports are generated only when specific questions arise, leading to reactive rather than proactive decision-making.
  + **Limitations:** Missed opportunities, delayed insights, inconsistent reporting standards.

**Proposed System along with Advantages:**

To overcome the limitations of traditional sales reporting, this project proposes a robust, automated (for data preparation), and interactive sales analysis solution. The system leverages SQL for efficient data cleaning and transformation, and PowerBI for dynamic and insightful visualizations.

**Key Features of the Proposed System:**

* **Automated Data Preparation (SQL):** SQL scripts are used to clean, transform, and join multiple data tables, ensuring data quality and readiness for analysis. This reduces manual effort and errors.
* **Interactive Dashboards (PowerBI):** PowerBI enables the creation of dynamic, multi-page dashboards that allow users to drill down into data, apply filters, and explore insights interactively.
* **Comprehensive Sales Metrics:** The system provides key sales performance indicators such as total sales, sales by product category, top customers, and budget vs. actuals.
* **Trend Analysis:** Visualizations clearly show sales trends over time, enabling identification of seasonal patterns or growth areas.
* **Geographic Insights:** Sales data can be visualized on maps to understand regional performance.
* **Shareable Reports:** Dashboards can be exported as static PDFs for easy sharing with stakeholders who may not have PowerBI access.

**Advantages Over the Existing System:**

|  |
| --- |
|  |

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Existing System** | **Proposed System** |
| 1. | Manual & error-prone | Automated data cleaning & transformation |
| 2. | Static reports | Interactive & dynamic dashboards |
| 3. | Limited insights | Comprehensive metrics & trend analysis |
| 4. | Time-consuming updates | Efficient data refresh |
| 5. | Poor scalability | Handles large datasets effectively |
| 6. | Difficult to share | Easily shareable interactive/static reports |

**3. Feasibility Study**

A feasibility study was conducted to evaluate the practicality of implementing the proposed Sales Analysis System. This assessment covered technical, behavioral, and economic aspects to ensure the system is achievable and beneficial.

**Technical Feasibility:**

The system is technically feasible as it utilizes widely adopted and robust technologies:

* **SQL (Microsoft SQL Server):** A powerful relational database management system for data manipulation.
* **PowerBI:** A leading business intelligence tool for data visualization and interactive dashboards.
* **CSV/Excel:** Standard data formats for input and output.

The project relies on proven tools with extensive documentation and community support, ensuring that technical challenges can be effectively addressed. The data is structured and suitable for relational database operations and visualization.

**Conclusion:** The system is technically sound and relies on mature, well-supported technologies.

**Behavioral Feasibility:**

Behavioral feasibility assesses the acceptance and usability of the system by its end-users (e.g., sales managers, business analysts).

* **User-Friendly Interface:** PowerBI dashboards are designed to be intuitive and interactive, requiring minimal training for users to navigate and extract insights.
* **Actionable Insights:** The visualizations are tailored to answer common business questions related to sales performance, making the system directly relevant and valuable to decision-makers.
* **Accessibility:** Reports can be shared in various formats (interactive PowerBI files, static PDFs), catering to different user needs and technical proficiencies.

**Conclusion:** The system is user-friendly and aligns well with the analytical needs of business stakeholders.

**Economic Feasibility:**

Economic feasibility analyzes the cost-effectiveness of the proposed system.

* **Open-Source/Freemium Tools:** SQL Server (Developer Edition/Express) and PowerBI Desktop are available for free, minimizing software licensing costs for development.
* **No Commercial Licenses:** The core tools used do not require expensive commercial licenses for personal or academic project use.
* **Resource Efficiency:** The analysis can be performed on standard computing hardware, requiring no specialized infrastructure.
* **High ROI (Potential):** For a business, implementing such a system can lead to improved sales strategies, better resource allocation, and increased revenue, far outweighing the initial development effort.

**Conclusion:** The system is highly economical and cost-efficient for both academic and potential enterprise usage.

**4. Project Monitoring System (Gantt Chart)**

The Gantt chart visually represents the project timeline, including key phases: Planning, Data Acquisition & Cleaning, Analysis & Modeling, Visualization, and Documentation. Each task is scheduled across a defined number of weeks to ensure timely completion and progress tracking.

**Gantt Chart Overview (Example for 6 Weeks):**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Activity** | **Week 1** | **Week 2** | **Week 3** | **Week 4** | **Week 5** | **Week 6** |
| Requirement Analysis | ✓ |  |  |  |  |  |
| Data Acquisition & Cleaning |  | ✓ | ✓ |  |  |  |
| Data Analysis & Modeling |  |  | ✓ | ✓ |  |  |
| Dashboard Visualization |  |  |  | ✓ | ✓ |  |
| Report Documentation |  |  |  |  | ✓ | ✓ |
| Final Review & Presentation |  |  |  |  |  | ✓ |

**5. System Analysis**

System analysis involves breaking down the overall project into structured components to understand its behavior, identify key requirements, and design how data flows through the system. This stage ensures that the final solution aligns with the project’s objectives and technical feasibility.

**a) Requirement Specification**

**Hardware Requirements:**

|  |  |
| --- | --- |
| **Component** | **Specification** |
| Processor | Intel i5 equivalent or higher |
| RAM | Minimum 8 GB (16 GB recommended for large datasets in PowerBI) |
| Storage | Minimum 256 GB SSD (for faster performance) |
| Network | Internet Connection (for downloading software and datasets) |

**Software Requirements:**

|  |  |
| --- | --- |
| **Software** | **Version / Type** |
| Operating System | Windows 10/11 (PowerBI Desktop is Windows-only) |
| Database Management System | Microsoft SQL Server (e.g., Developer Edition or Express) |
| Business Intelligence Tool | Microsoft PowerBI Desktop |
| Text Editor | Visual Studio Code or similar (for SQL scripts) |
| Spreadsheet Software | Microsoft Excel (for budget data) |

**Functional Requirements:**

* The system must extract sales data from the AdventureWorksDW2019 database.
* SQL queries must clean and prepare DimDate, DimCustomer, DimProduct, and FactInternetSales tables.
* The cleaned data must be exportable to CSV format.
* PowerBI must import the CSV files and the Excel budget file.
* PowerBI data model must correctly establish relationships between tables.
* The dashboard must display total sales, sales trends over time, and sales by product category.
* The dashboard must identify and display top 10 customers by sales.
* The dashboard must identify and display top 10 best-selling products.
* The dashboard must compare actual sales against budget.
* The dashboard must be interactive, allowing filtering by year, month, product category, and customer city.
* The dashboard should be exportable as a PDF document.

**Non-Functional Requirements:**

* **Accuracy:** All calculations and visualizations must be accurate and reflect the underlying data correctly.
* **Performance:** The PowerBI dashboard should load and respond quickly to user interactions.
* **Usability:** The dashboard must be intuitive and easy for non-technical users to understand and navigate.
* **Maintainability:** SQL scripts and PowerBI models should be well-organized and commented for future modifications.
* A diagram of a company

  AI-generated content may be incorrect.**b) System Flowchart**
* **Level 0 DFD – High-Level System Overview:**
* A diagram of sales analysis system

  AI-generated content may be incorrect.
* **Level 1 DFD – Data Processing Breakdown:**
* A diagram of a diagram

  AI-generated content may be incorrect.
* A diagram of a diagram

  AI-generated content may be incorrect.**Level 2 DFD – Detailed Data Cleaning & Transformation Logic:**

**d) Entity-Relationship Diagram (ERD) Overview**

While this project does not create a new database schema, it leverages the existing relational structure of the AdventureWorksDW2019 data warehouse. The key tables and their relationships, as utilized in the SQL queries for cleaning and subsequent PowerBI modeling, are conceptually as follows:

* **FactInternetSales**: The central fact table containing sales transaction details.
  + Attributes: ProductKey, OrderDateKey, CustomerKey, SalesAmount, SalesOrderNumber, etc.
* **DimDate**: Dimension table for date information.
  + Attributes: DateKey, FullDateAlternateKey (Date), EnglishDayNameOfWeek (Day), EnglishMonthName (Month), CalendarYear (Year), etc.
  + Relationship: FactInternetSales.OrderDateKey relates to DimDate.DateKey.
* **DimCustomer**: Dimension table for customer information.
  + Attributes: CustomerKey, FirstName, LastName, Gender, DateFirstPurchase, Customer City (from DimGeography).
  + Relationship: FactInternetSales.CustomerKey relates to DimCustomer.CustomerKey.
* **DimProduct**: Dimension table for product information.
  + Attributes: ProductKey, ProductName, Product Subcategory, Product Category, Product Color, Product Size, Product Status.
  + Relationship: FactInternetSales.ProductKey relates to DimProduct.ProductKey.
* **DimGeography**: Dimension table for geographic information (joined with DimCustomer).
  + Attributes: GeographyKey, City.
  + Relationship: DimCustomer.GeographyKey relates to DimGeography.GeographyKey.

The SalesBudget (from Excel) table is linked to DimDate in PowerBI based on the Date column for budget vs. actuals analysis.

**6. System Design**

System design defines the architecture, components, interfaces, and data flow of the system to meet the specified requirements. For this project, the system is designed with data integrity, clarity, and interactive visualization in mind, using SQL for data preparation and PowerBI for dashboard creation.

**a) File / Data Design**

The system processes data from a SQL Server database and an Excel file, producing cleaned CSV files as intermediaries, which are then consumed by PowerBI.

**Key Data Components & Flow:**

1. **Raw Data Source:**
   * AdventureWorksDW2019.bak: The primary dataset residing in Microsoft SQL Server.
   * sales\_budget\_data.xlsx: An Excel file containing monthly sales budget data.
2. **SQL Scripts (sales\_analysis\_queries.sql):**
   * Contains queries to select, clean, and transform relevant columns from DimDate, DimCustomer, DimProduct, and FactInternetSales.
   * Joins tables (e.g., DimCustomer with DimGeography, DimProduct with DimProductSubcategory and DimProductCategory).
   * Filters FactInternetSales and DimDate for the last two years of data.
   * Renames columns for clarity (e.g., FullDateAlternateKey to Date).
   * Handles NULL values (e.g., Product Status using ISNULL).
3. **Intermediate Data (CSV Exports):**
   * cleaned\_dim\_customer.csv
   * cleaned\_dim\_date.csv
   * cleaned\_dim\_product.csv
   * cleaned\_fact\_internet\_sales.csv
   * These CSVs serve as the cleaned and prepared input for PowerBI, ensuring consistency and ease of import.
4. **PowerBI Project File (sales\_dashboard\_project.pbix):**
   * Imports the four cleaned CSV files and the sales\_budget\_data.xlsx file.
   * Establishes relationships between these tables in the PowerBI data model.
   * Contains DAX (Data Analysis Expressions) formulas for creating calculated columns and measures (e.g., Total Sales, Sales vs. Budget).
   * Houses the interactive dashboard pages and their visualizations.

**b) Dashboard Metric Design (PowerBI)**

The PowerBI dashboard is designed to present key sales metrics and insights through various visualizations. The "metrics" here refer to the calculated measures and aggregated data displayed on the dashboard.

**Key Metrics and Visualizations:**

* **Total Sales:** A card visualization showing the overall sales amount.
* **Sales by Product Category:** A donut chart or bar chart showing the distribution of sales across major product categories (Bikes, Accessories, Clothing).
* **Sales by Product Subcategory:** A bar chart allowing drill-down to see sales within specific product subcategories.
* **Sales by Month (Budget vs. Actual):** A line chart comparing actual monthly sales against the set budget, highlighting performance over time.
* **Sales by Customer City:** A map visualization showing sales distribution geographically, with bubble sizes indicating sales volume.
* **Top 10 Customers by Sales:** A bar chart listing the top customers and their respective sales amounts.
* **Top 10 Best-Selling Products:** A bar chart displaying the products with the highest sales.

Each visualization is designed to be clear, concise, and interactive, allowing users to filter by various dimensions (e.g., Year, Month, Product Category, Customer City) to gain deeper insights.

**7. Input / Output Design**

This section covers how data is presented and interacted with within the PowerBI environment and as final reports.

**a) PowerBI Interface Design**

The "input forms" in this context refer to the PowerBI Desktop interface used for building the dashboard, and the "output forms" are the interactive dashboard pages themselves.

**PowerBI Desktop Interface (Design View):**

* **Fields Pane:** Displays all loaded tables and their columns, allowing drag-and-drop functionality to create visualizations.
* **Visualizations Pane:** Offers a wide range of chart types (bar, line, pie, map, cards, tables) and formatting options.
* **Filters Pane:** Enables setting up interactive filters for the entire report, specific pages, or individual visuals.
* **Data View:** Allows inspecting the raw and transformed data tables.
* **Model View:** Crucial for establishing and managing relationships between tables.

**b) Report & Dashboard Visualization**

The core output of this project is a multi-page interactive PowerBI dashboard, which can also be exported as a static PDF report.

**Interactive PowerBI Dashboard (3 Pages):**

1. **Page 1 - AdventureWorks Sales Overview:**
   * **Purpose:** Provides a high-level summary of sales performance.
   * **Visuals:**
     + Donut chart for "Sales by Product Category" showing the proportion of sales from Bikes, Accessories, and Clothing. (e.g., Bikes 93.7%, Accessories 4.3%, Clothing 2.1%).
     + Line chart for "Budget vs Sales by Month" displaying monthly sales against the budget for 2021 and 2022.
     + Map visualization for "Sales by Customer City" showing sales volume across different cities globally.
     + Bar charts for "Top 10 Customers By Sales" and "Top 10 Best Selling Products" providing quick insights into key performers.
   * A screenshot of a computer

     AI-generated content may be incorrect.**Key Insights:** Overall sales distribution, performance against budget, major sales regions, and top customers/products.
2. **Page 2 - Sales by Customer:**
   * **Purpose:** Focuses on customer-centric sales analysis.
   * **Visuals:**
     + Detailed bar chart of "Top 10 Customers By Sales" with exact sales figures.
     + Table showing monthly sales breakdown for each of the top customers.
     + Map visualization of "Sales by Customer City" to see customer concentration.
   * **Key Insights:** Who are the most valuable customers, their purchasing patterns over time, and their geographic distribution.

A screenshot of a computer

AI-generated content may be incorrect.

1. **Page 3 - Sales by Product:**
   * **Purpose:** Provides in-depth analysis of product performance.
   * **Visuals:**
     + Detailed bar chart of "Top 10 Best Selling Products" with exact sales figures.
     + Table showing monthly sales breakdown for each of the top products.
     + Filters for "Product Category," "Product Subcategory," and "Product Name" for granular analysis.
   * **Key Insights:** Which products are driving sales, their performance trends, and the ability to analyze specific product lines.

A screenshot of a computer

AI-generated content may be incorrect.**PDF Report (sales\_performance\_dashboard.pdf):** A static export of the PowerBI dashboard, providing a snapshot of the key visualizations and data for easy sharing and printing. This ensures that insights can be disseminated even to those without PowerBI Desktop.

**8. System Testing**

System testing is a critical phase that ensures the sales analysis system performs as expected under various conditions. For this project, testing was conducted at multiple levels: verifying SQL query outputs, validating data loading in PowerBI, and confirming dashboard accuracy and interactivity.

**a) Preparation of Test Data**

The primary test data used was the provided AdventureWorksDW2019 dataset, along with the sales\_budget\_data.xlsx file. This real-world-like data allowed for realistic testing scenarios.

* **SQL Data:** The .bak file was restored in SQL Server, providing access to the FactInternetSales, DimDate, DimCustomer, DimProduct, and DimGeography tables.
* **Budget Data:** The sales\_budget\_data.xlsx file provided the monthly budget figures.

**b) Testing with Live Data**

The SQL queries were executed against the live SQL Server database, and the resulting CSVs were loaded into PowerBI. The PowerBI dashboard was then tested for functionality and accuracy.

**Observations:**

* SQL queries successfully extracted and transformed data, producing the expected CSV files.
* PowerBI imported all CSVs and the Excel file without errors.
* Relationships between tables in PowerBI were correctly established and validated.
* All dashboard visualizations loaded correctly and displayed data as anticipated.
* Filters (Year, Month, Product Category, Customer City) functioned interactively, updating visuals in real-time.
* Calculated measures (e.g., Total Sales, Budget vs. Sales) showed accurate values.

**c) Test Cases with Results**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test Case ID** | **Description** | **Input** | **Expected Output** | **Actual Result** | **Status** |
| TC01 | SQL Query for DimDate filtering | sales\_analysis\_queries.sql | Dates from 2020 onwards | Success | ✅ |
| TC02 | SQL Query for DimCustomer join & gender map | sales\_analysis\_queries.sql | Customer details with city & gender (Male/Female) | Success | ✅ |
| TC03 | SQL Query for DimProduct status handling | sales\_analysis\_queries.sql | Product list with 'Current' or 'Outdated' status | Success | ✅ |
| TC04 | PowerBI Data Load & Relationships | All 4 CSVs + Excel budget | All tables loaded, relationships active | Success | ✅ |
| TC05 | Total Sales Calculation | Dashboard "Total Sales" card | Correct sum of SalesAmount | Success | ✅ |
| TC06 | Top 10 Customers Visualization | Filter by "2022" | Bar chart showing top 10 customers for 2022 | Success | ✅ |
| TC07 | Budget vs. Sales Chart | Line chart for "Budget and Monthly Sales" | Sales line tracking budget line, showing variances | Success | ✅ |
| TC08 | Interactive Filtering | Select "Bikes" in Product Category filter | All visuals update to show only Bike sales | Success | ✅ |

**Error Handling Tests (Conceptual):**

* **Missing Data Files:** PowerBI would display errors during data refresh.
* **Incorrect SQL Syntax:** SQL Server would return query execution errors.
* **Broken PowerBI Relationships:** Visuals would show blank or incorrect data.

**Conclusion:** System testing confirms that the Sales Analysis project is:

* Functionally reliable in data processing and visualization.
* Accurate in its calculations and insights.
* Interactive and user-friendly, meeting the project objectives.

**9. System Implementation**

System implementation involves deploying the developed system in a production-like environment. In this project, the sales analysis system was implemented by preparing the data using SQL and building interactive dashboards in PowerBI Desktop.

**a) System Requirements**

**Hardware Requirements:**

|  |  |
| --- | --- |
| **Component** | **Specification** |
| Processor | Intel Core i5 or equivalent |
| RAM | Minimum 8 GB (16 GB recommended) |
| Storage | 256 GB SSD (for performance) |
| Network | Internet access (for software downloads) |

**Software Requirements:**

|  |  |
| --- | --- |
| **Software** | **Version / Type** |
| Operating System | Windows 10/11 |
| Database Management System | Microsoft SQL Server |
| Business Intelligence Tool | Microsoft PowerBI Desktop |
| Text Editor | Visual Studio Code or similar |
| Spreadsheet Software | Microsoft Excel |

**b) Implementation Steps**

1. **SQL Server Setup & Data Restoration:**
   * Installed Microsoft SQL Server.
   * Restored the AdventureWorksDW2019.bak database backup file into SQL Server Management Studio (SSMS).
   * Executed the Update\_AdventureWorksDW\_Data.sql script (from TechTalkCorner) to update date columns in the dataset for consistency.
2. **SQL Query Development & Execution:**
   * Wrote and refined SQL queries in sales\_analysis\_queries.sql to clean and prepare the DimDate, DimCustomer, DimProduct, and FactInternetSales tables.
   * Executed these queries to generate the cleaned datasets.
3. **CSV Export:**
   * Exported the results of the cleaned SQL queries into four separate CSV files: cleaned\_dim\_customer.csv, cleaned\_dim\_date.csv, cleaned\_dim\_product.csv, and cleaned\_fact\_internet\_sales.csv.
4. **PowerBI Desktop Setup:**
   * Installed PowerBI Desktop.
5. **Data Import into PowerBI:**
   * Imported the four cleaned CSV files and the sales\_budget\_data.xlsx file into PowerBI Desktop.
6. **Data Modeling in PowerBI:**
   * Established and verified relationships between the imported tables (e.g., FactInternetSales linked to DimDate, DimCustomer, DimProduct).
7. **Data Transformation (Power Query):**
   * Performed additional transformations in Power Query (PowerBI's data transformation tool) such as renaming columns, changing data types, and handling any remaining data quality issues to ensure data was in a format suitable for visualization.
8. **Dashboard Creation:**
   * Designed and built the three interactive dashboard pages ("AdventureWorks Sales Overview," "Sales by Customer," and "Sales by Product") using various PowerBI visualizations (cards, bar charts, line charts, donut charts, maps).
   * Created necessary DAX measures for calculations like total sales, budget variance, etc.
9. **Report Export:**
   * Exported the completed PowerBI dashboard as a static PDF report (sales\_performance\_dashboard.pdf).
   * Saved the PowerBI project file (sales\_dashboard\_project.pbix).

**Final Outcome:** The implemented system provides a comprehensive and interactive solution for sales performance analysis, transforming raw database data into actionable business insights.

**10. Documentation**

Documentation plays a crucial role in ensuring the usability, maintainability, and scalability of a data analysis project. It provides clear guidance to users and developers regarding project setup, usage, and troubleshooting.

**a) User Documentation**

This part of the documentation is aimed at users or stakeholders who will consume the sales analysis reports and dashboards.

**Dashboard Usage Instructions:**

* **Accessing the Dashboard:** The interactive dashboard can be accessed by opening the sales\_dashboard\_project.pbix file with PowerBI Desktop. A static version is available in sales\_performance\_dashboard.pdf.
* **Navigating Pages:** Use the page tabs at the bottom of the PowerBI Desktop interface (or the PDF pages) to switch between "AdventureWorks Sales Overview," "Sales by Customer," and "Sales by Product."
* **Applying Filters:** Utilize the filter panes on the right side of the dashboard (in PowerBI Desktop) to filter data by "Year," "Month," "Product Category," "Customer City," etc. Click on visual elements (e.g., a bar in a chart) to cross-filter other visuals on the page.
* **Drill-Down:** Some visuals may support drill-down functionality (e.g., from Product Category to Product Subcategory). Look for the drill-down arrows on the top right of the visual.
* **Exporting Data:** In PowerBI Desktop, data from any visual can be exported to CSV by right-clicking the visual and selecting "Export data."

**b) Developer Documentation**

This includes internal logic and how to modify or extend the system.

**SQL Script Overview (sales\_analysis\_queries.sql):**

* **Purpose:** Cleans, transforms, and joins raw AdventureWorks data into a format suitable for PowerBI.
* **Sections:**
  + DimDate Cleaning: Selects relevant date attributes, renames FullDateAlternateKey to Date, and filters for CalendarYear >= 2020.
  + DimCustomer Cleaning: Joins DimCustomer with DimGeography to include Customer City, renames columns, and maps Gender codes to 'Male'/'Female'.
  + DimProduct Cleaning: Joins DimProduct with DimProductSubcategory and DimProductCategory, renames columns, and handles NULL values in Product Status by replacing them with 'Outdated'.
  + FactInternetSales Cleaning: Selects key sales attributes and filters data to include only the last two years based on OrderDateKey.
* **Dependencies:** Requires access to an AdventureWorksDW2019 database instance.

**PowerBI Model Overview (sales\_dashboard\_project.pbix):**

* **Data Sources:** Links to the four CSV files and the sales\_budget\_data.xlsx file. Ensure these files are in the expected directory or update their paths if moved.
* **Data Model:** Defines relationships between tables (e.g., one-to-many from dimension tables to fact table).
* **Measures & Calculated Columns (DAX):**
  + Total Sales = SUM(FactInternetSales[SalesAmount])
  + Sales vs Budget = [Total Sales] - SUM('Budget'[Budget])
  + ... (and other measures for specific KPIs)
* **Power Query Steps:** Review "Transform data" in PowerBI Desktop to understand the applied cleaning and transformation steps within Power Query.

**c) Troubleshooting Guidelines**

|  |  |  |
| --- | --- | --- |
| **Issue** | **Cause** | **Solution** |
| PowerBI visuals show blank data | Broken relationships or data loading errors | Check PowerBI "Model View" for relationships; refresh data sources. |
| Incorrect calculations in dashboard | Incorrect DAX formulas or data types | Review DAX measures; verify column data types in Power Query. |
| SQL query errors | Syntax errors or incorrect table/column names | Review sales\_analysis\_queries.sql for typos; verify database schema. |
| CSV files not updating | SQL export process failed | Manually run SQL queries and re-export CSVs. |
| Dashboard slow to load/interact | Large dataset or complex DAX/visuals | Optimize DAX formulas; consider data aggregation or filtering for performance. |

**Summary:** This documentation enables anyone with basic data analysis knowledge to:

* Understand the project's data flow and design.
* Interact with the PowerBI dashboard effectively.
* Modify or extend the SQL queries and PowerBI model.
* Troubleshoot common issues.

**11. Scope of the Project**

The Sales Performance Analysis project has been designed as a practical and insightful solution for understanding sales trends and key performance indicators. This project offers wide applicability across academic and business environments for foundational data analysis.

**Current Scope**

The current implementation provides the following capabilities:

* **Comprehensive Data Cleaning & Transformation:** Utilizes SQL to refine raw sales, customer, product, and date data into a clean, usable format.
* **Interactive Sales Dashboard:** Presents a multi-page PowerBI dashboard with key visualizations for sales overview, customer analysis, and product performance.
* **Key Performance Indicators (KPIs):** Displays total sales, sales by category/subcategory, top customers, top products, and budget vs. actual sales.
* **Trend and Geographic Analysis:** Visualizes sales trends over time and geographically using maps.
* **Shareable Reports:** Provides an exportable PDF version of the dashboard for easy distribution.

**Future Scope**

The system has strong potential for extension and integration into a more automated and advanced analytics pipeline. Possible enhancements include:

* **Automated Data Refresh:** Implement a Python script or other automation tools to automatically extract data from the SQL database, run cleaning scripts, and refresh the PowerBI dataset without manual intervention. This aligns with the "Future Project Aim" mentioned in your README.md.
* **Automated Report Distribution:** Develop a mechanism (e.g., Python script) to automatically export the PowerBI dashboard as a PDF and email it to concerned stakeholders on a scheduled basis.
* **Predictive Analytics:** Integrate machine learning models (e.g., using Python) to forecast future sales, identify potential churn customers, or recommend personalized products.
* **Advanced Customer Segmentation:** Beyond top customers, segment customers based on purchasing behavior (e.g., RFM analysis) for targeted marketing.
* **Integration with Other Data Sources:** Combine sales data with marketing, inventory, or website analytics data for a more holistic business view.
* **User Access & Security:** For enterprise deployment, implement robust user authentication and authorization for PowerBI reports.
* **Alerting System:** Set up alerts within PowerBI or external tools to notify stakeholders of significant deviations from sales targets or unusual trends.

**Conclusion:** The current system successfully fulfills key sales analysis needs, providing valuable insights through interactive visualizations. It is built with future extensibility in mind, with clear pathways to evolve into a fully automated and more sophisticated business intelligence solution.

**12. Bibliography**

The following references and resources were used during the research, development, and documentation phases of this project:

**Books and Research Papers:**

* [Add relevant books on SQL, Data Warehousing, Business Intelligence, or Data Analysis, e.g., "The Data Warehouse Toolkit" by Ralph Kimball, "SQL for Data Analysis" by Cathy O'Neil and Rachel Schutt]

**Websites and Online Resources:**

* Microsoft Learn - AdventureWorks Sample Databases: <https://learn.microsoft.com/en-us/sql/samples/adventureworks-install-configure?view=sql-server-ver15&tabs=ssms>
* Microsoft Power BI Documentation: <https://docs.microsoft.com/en-us/power-bi/>
* SQL Documentation (e.g., for specific functions used): <https://docs.microsoft.com/en-us/sql/sql-server/>
* TechTalkCorner SQL Script for AdventureWorks Data Update: <https://github.com/techtalkcorner/SampleDemoFiles/blob/master/Database/AdventureWorks/Update_AdventureWorksDW_Data.sql>

**Tools Used:**

* Microsoft SQL Server Management Studio (SSMS)
* Microsoft PowerBI Desktop
* Visual Studio Code (for SQL scripting)
* Microsoft Excel (for budget data)

This bibliography reflects a mix of technical documentation, online resources, and tools that contributed to the planning, design, and development of the Sales Performance Analysis system.