Supply chain data analysis

Presented by:

- **Mariam Yasser Nabih AST 44**
- **Zahra Mahmoud Ahmed AST 40**
- **Alaa Walid Hussein AST 43**
- **Aya Ahmed Mohamed AST 42**
- **Rawan Ayman Saed AST 41**

Project Documentation:

Table of Content:

1. Project Planning:

- **Project Proposal: overview:**

Effective supply chain management is essential for minimizing costs, improving delivery performance, and enhancing customer satisfaction. This project focuses on analyzing customer orders, product performance, and shipping data to identify key patterns that impact business efficiency.

Objectives:

Analyze customer purchasing behaviors and classify them based on the country. Identify and assess shipping delays to improve fulfillment efficiency. Evaluate product sales performance and profitability based on discounts and profit margins. Provide business recommendations for optimizing supply chain operations.

Scope:

This project involves:

Data Collection & Cleaning – Processing order, customer, product, and shipping data.

EDA– Identifying trends in sales, customers, and deliveries.

Customer Segmentation – Categorizing customers based on order frequency.

Shipping Analysis – Detecting and classifying shipment delays.

Profitability Insights – Assessing product margins and sales trends.

- **Project plan:**

Week 1: Build Data Model, Data Cleaning and Preprocessing

- Data Preprocessing: Build a data model and clean and preprocess the data.
- Tools: Python (pandas), power query
- deliverables: Cleaned dataset ready for analysis. Data preprocessing notebook. Week 2:

Analysis of Questions Phase

- Determine Data Analysis Questions
- Deliverables:

- Sales & Profitability Analysis:

1. What is the trend of total sales over the years? Does the declining order trend indicate seasonal variations, market saturation, or other factors?
2. Which product categories contribute the most to revenue? How can the business optimize inventory and pricing for underperforming categories?
3. Which cities generate the highest sales? How can these insights be used for targeted marketing or expansion?
4. How does revenue loss impact overall profitability? What percentage of losses come from fraud versus cancellations, and how can they be minimized?

Customer & Order Insights:

5. What is the distribution of customers across different regions? Does a specific region contribute more to cancellations or revenue loss?
6. How do different delivery statuses (advance, on-time, canceled) impact customer retention? Can we identify patterns among frequently canceled orders?
7. Are there specific customer segments that order more frequently or contribute significantly to revenue? How can this help with customer retention strategies?

Supply Chain & Shipping Performance: 8. Which

regions face the highest delivery delays?

9. Is there a correlation between order delays and revenue loss? Are high-profit regions also experiencing frequent shipping delays?
10. What actions can be taken to improve order fulfillment based on current trends in delays and cancellations?

Week 3: Forecasting Questions Phase

- Determine a set of forecasting questions and answer them using the trends
- Deliverables: Visualization using power bi, SQL query answering forecasting questions.

Week 4: Visualization Dashboard and Final Presentation

- Visualization dashboard.
- Tools: Power bi

2. Requirements gathering:

- According to this Data the need To identify inefficiencies, reduce order delays, and optimize logistics and inventory management.
- Sales & Marketing Teams – To analyze top-performing products, understand customer behavior, and target high-revenue cities for better marketing campaigns.
- Finance Team – To track revenue trends, assess profitability, and minimize revenue loss from fraud and cancellations.
- Customer Service Teams – To monitor delivery performance, reduce cancellation rates, and improve customer satisfaction.
- Executives & Decision-Makers – To get a high-level overview of sales, revenue, and operational performance for strategic planning and business growth.

3. System analysis and Design:

- Tables in excel creating DIM and Fact tables
- Database Design & Data Modeling: Logical & Physical Schema – Tables, attributes, keys, and normalization considerations.
- **Fact Table (Order Fact Table)**: Stores transactional data, including sales, order details, revenue, and profit.
- **Customer Dimension** – Contains customer details like ID, location, and demographics.
- **Product Dimension** – Stores product information such as category, name, and price.
- **Shipping Dimension** – Tracks shipment details, including delays, order status, and logistics.
- Data Modeling Approach
- **Star Schema**: The database follows a star schema structure, with the fact table at the center linked to dimension tables, enabling efficient querying and reporting.
- **Indexing & Optimization**: Indexing was implemented to improve query performance for real-time analysis.

4. Implementation:

1. Data Collection & Integration

Extracted raw data from multiple sources, including Order Fact Table, Customer DIM, Product DIM, and Shipping DIM. - Integrated the datasets to ensure consistency and proper relationships between fact and dimension tables.

2. Data Cleaning & Preprocessing

Handled missing values and inconsistencies using Python (Pandas, NumPy). - Removed duplicates and standardized formats for customer names, product categories, and order statuses. - Addressed data integrity issues in fields such as dates, order IDs, and revenue calculations.

3. Data Transformation & Modeling

Designed a star schema database structure for efficient querying and reporting. - Merged datasets to create a final cleaned dataset (main data after cleaning.xlsx) for analysis. – Applied business logic to classify order statuses, detect revenue loss due to fraud/cancellations, and track delivery delays.

4. Data Analysis & Insights Extraction

Conducted exploratory data analysis (EDA) to uncover trends in sales, revenue, and customer distribution. - Measured order trends, revenue loss, delivery performance, and regional sales patterns. - Identified key insights, such as top-performing product categories, high-revenue cities, and common fraud patterns.

5. Visualization & Dashboard Creation

Built an interactive Power BI dashboard to present findings in a structured format. Designed visual representations for sales trends, revenue breakdowns, delivery performance, and geographical distribution. enabled stakeholders to filter and drill down into specific insights for decision-making.

5. Testing and quality assurance:

Final Review & Optimization

- Validated results against expected business logic to ensure accuracy.
- Optimized queries and data models for better performance.
- Prepared the final documentation to summarize findings and recommendations.

6. Challenges and solutions:

Error Handling & Debugging

- Identified and fixed incorrect data mappings, calculation errors, and missing values.

- Addressed inconsistent date formats and incorrect regional categorization.
- Implemented data validation checks to prevent errors from propagating in future updates
- date format and country abbreviation.

Solutions:
Used tools like Power Query and Python to fix the date format. And using SQL for fixing the country abbreviation for retrieving data.

7. Final presentation and reports:

Final deliverables