WideWorldImporters

Data Analysis Project

Prepared by:

1. **AbdelRahman AbdelMoez Anwar**
2. **Fatma Ali Khaled**
3. **Youssef Mohamed Farag**
4. **Noha Soliman Mohamed**

Supervised by: Eng. Ahmed Alaa

Group: ALX2_DAT1_G1 Google Data Analyst Specialist

Tools Used: SQL Server, Power BI, Python, Excel

---

1 Project Planning & Management

Project Proposal

Our team will conduct a comprehensive analysis using the **WideWorldImporters** database. This project will focus on **Sales, HR, Supply Chain, and Marketing** using **SQL Server, Power BI, Python, and Excel** to generate valuable business insights.

Project Plan

📅 Timeline Breakdown (2 Months Plan)

Week	Tasks
Week 1	Database understanding, task assignments, planning.
Week 2	SQL queries for data extraction, initial cleaning.

Week	Tasks
Week 3	Exploratory Data Analysis (EDA) – Finding trends.
Week 4	Power BI dashboard creation.
Week 5	Refining insights, testing dashboard usability.
Week 6	Documentation drafting.
Week 7	Final reporting, business recommendations.
Week 8	Presentation preparation, submission.

Task Assignments & Roles

Team Member	Responsibility
AbdelRahman AbdelMoez Anwar	Data Visualization (Power BI) & Business Insights
Fatma Ali Khaled	Data Cleaning & Processing (Python, Excel)
Youssef Mohamed Farag	SQL Queries for Data Extraction
Noha Soliman Mohamed	Documentation & Presentation

Risk Assessment & Mitigation

- **Data Integrity Issues** → Use **Python (Pandas) & SQL** for cleaning.
- **Time Constraints** → Plan tasks efficiently to meet deadlines.
- **Visualization Complexity** → Ensure dashboards focus on **KPIs & trends** for clarity.

KPIs for Success

- **Sales:** Revenue growth, best-selling products, customer retention rate.
- **HR:** Employee turnover, performance analysis.
- **Supply Chain:** Inventory turnover, supplier efficiency.
- **Marketing:** Campaign effectiveness, customer acquisition cost.

② Literature Review

Feedback & Evaluation

- Lecturer's assessment of the project, identifying strengths and areas for improvement.

Suggested Improvements

- Enhancements in data analysis techniques, additional insights, and optimization.

Final Grading Criteria

- **Documentation:** Clear and structured reporting.
- **Implementation:** Effective use of SQL, Python, and Power BI.
- **Testing:** Data accuracy and visualization effectiveness.
- **Presentation:** Delivery of findings and storytelling clarity.

---

③ Requirements Gathering

Stakeholder Analysis

We focus on **sales managers, HR specialists, supply chain teams, and marketing executives**, ensuring that our analysis aligns with their needs.

User Stories & Use Cases

- *"As a Sales Manager, I want to see revenue trends by region to optimize product distribution."*
- *"As an HR Manager, I need to track employee turnover trends to reduce hiring costs."*

Functional & Non-Functional Requirements

- **Functional:**
    - SQL-based data extraction

- Power BI dashboarding
- Python-powered data cleaning
- **Non-Functional:**
  - Fast report generation
  - Secure data access

4️⃣ System Analysis & Design

Problem Statement & Objectives

We aim to answer critical business questions:

- **Sales:** What products generate the most revenue?
- **HR:** What factors contribute to high employee turnover?
- **Supply Chain:** Which suppliers are most reliable?
- **Marketing:** What campaigns drive the highest ROI?

Database Structure & Schema Overview

The **WideWorldImporters** database includes key schemas:

- **Sales:** Orders, customers, transactions.
- **Application:** System-wide settings and user information.
- **Purchasing:** Suppliers, purchase orders.
- **Warehouse:** Inventory, stock movements.

Data Flow & Process

1. Extract raw data using **SQL Server**.
2. Clean and preprocess data with **Python (Pandas)**.
3. Visualize insights using **Power BI** and **Excel Reports**.

Entity-Relationship Diagram (ERD)

- **Sales:** Customers ↔ Orders ↔ Products
- **HR:** Employees ↔ Departments ↔ Salaries

- **Supply Chain:** Suppliers ↔ Inventory ↔ Warehouse

5️⃣ Data Cleaning & Preprocessing

Handling Data Issues

- **Missing Data:** Use **Pandas & SQL** to fill gaps with averages or interpolations.
- **Duplicate Entries:** Remove using **SQL DISTINCT** and **Pandas drop_duplicates()**.
- **Data Formatting:** Standardize dates, currency conversions.

6️⃣ Data Analysis & Insights

Sales Analysis

- **Top 5 Best-Selling Products**
- **Sales Revenue Trends (Monthly, Yearly)**
- **Customer Retention Rate by Region**

HR Analysis

- **Employee Turnover Rate**
- **Department-wise Salary Distribution**
- **Performance vs. Retention Trends**

Supply Chain Analysis

- **Supplier Reliability Score**
- **Inventory Turnover by Product Category**
- **Delivery Time Analysis**

Marketing Analysis

- **Campaign Performance (CTR, ROI, Conversions)**
- **Customer Segmentation Based on Purchases**

7️⃣ Data Visualization & Reporting

- **Power BI Dashboards:** Interactive charts with filters (date, category, region).
- **Excel Reports:** Pivot tables for detailed analysis.

8️⃣ System Deployment & Integration

Technology Stack

- **SQL Server:** Backend database.
- **Python (Pandas, Matplotlib, Seaborn):** Data cleaning, statistical analysis.
- **Power BI:** Dashboarding.
- **Excel:** Pivot tables, automated reporting.

Deployment Plan

- Schedule **automated SQL queries** to refresh data.
- Publish **Power BI dashboards** for end-user access.

9️⃣ Additional Deliverables

- **Testing:** Query performance, dashboard usability checks.
- **Final Report:** Executive summary, key findings, recommendations.
- **Presentation:** Stakeholder-friendly storytelling with visual insights.