DATA WAREHOUSE

**A data-driven approach to empower e-commerce decision-making through a robust data warehouse and insightful visualizations.**

**Authors:**

Kainat Bibi [Student ID: B2231046]

Labiba Ilyas [Student ID: B2231050]

Minam Muzaffar [Student ID: B2231055]

**Supervisor:**

Ms. Ambreen

**Institute:**

Benazir Bhutto Shaheed University Lyari Department of

Computer Science and Information Technology

**Submission Date:**

15-November, 2024

**For E Commerce**

Abstract:

* Project goal: Transform e-commerce business through a data-driven approach.
* Focus areas:
  + Customer behavior
  + Product performance
  + Business trends
* Key components:
  + Centralized data warehouse developed in Snowflake.
  + Data integration from multiple sources using SQL and Python for ETL processes.
* Data handling:
  + Ensured data consistency and accuracy.
  + Developed a well-structured data model.
* Reporting tools:
  + Interactive dashboards created in Tableau.
* Insights generated:
  + Customer preferences
  + Sales patterns
  + Target segments
* Outcomes:
  + Enabled data-backed decision-making for stakeholders.
  + Optimized marketing strategies.
  + Enhanced customer experiences.
  + Contributed to business growth and success.

Table of Contents

1. Introduction ............................................... 3

1.1 Background ............................................. 3

2. System Analysis .................................... 3

2.1 Data Sources............... 3

2.2 Data Extraction.............................. 3

2.3 Data Cleaning and Preprocessing ............... 3

2.4 Data Transformation and Integration………3

3. Data Warehouse Design........................................ 5

3.1 Data Modeling……………5

3.2 Snowflake Schema ................................... 5

4. Implementation……… 5

4.1 Data Visualization and Analysis ................. 5

5. Conclusion ................................................ 5

5.1 Summary of Key Findings…….5 - 6

6. References ................................................ 7

7. ER Model ................................................... 7

Introduction

Background:

In today's competitive e-commerce landscape, data-driven decision-making has become essential for businesses to thrive. By leveraging the power of data analytics, organizations can gain valuable insights into customer behavior, product performance, and market trends.

To harness the full potential of data, we embarked on a project to establish a robust data warehouse and visualization platform for our e-commerce store. This initiative aimed to consolidate data from various sources, transform it into actionable insights, and empower decision-makers with data-driven intelligence.

By implementing a data-driven approach, we sought to improve operational efficiency, enhance customer experience, and drive overall business growth.

System Analysis

2. Data Acquisition and Preparation

2.1 Data Sources

The primary data source for this project is our e-commerce website's database. This database contains a wealth of information, including customer demographics, purchase history and product details

2.2 Data Extraction

SQL queries were utilized to extract relevant data from the e-commerce website's database. These queries were designed to efficiently retrieve specific information, such as customer profiles, order details, and product sales data.

2.3 Data Cleaning and Preprocessing

The extracted data underwent a rigorous cleaning and preprocessing process to ensure data quality and consistency. This involved:

Handling missing values, Outlier detection and removal, data normalization and validation

2.4 Data Transformation and Integration

The cleaned and preprocessed data was transformed into a suitable format for loading into the data warehouse. This involved:

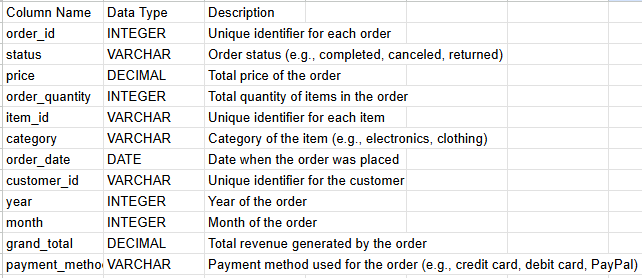
Data integration: Combining data from multiple sources into a unified dataset.

Data transformation: Reshaping and restructuring data to fit the data warehouse schema.

Data enrichment: Enhancing the data with additional information, such as product categories, customer segments, and time-based attributes.

System Design

3. Data Warehouse Design

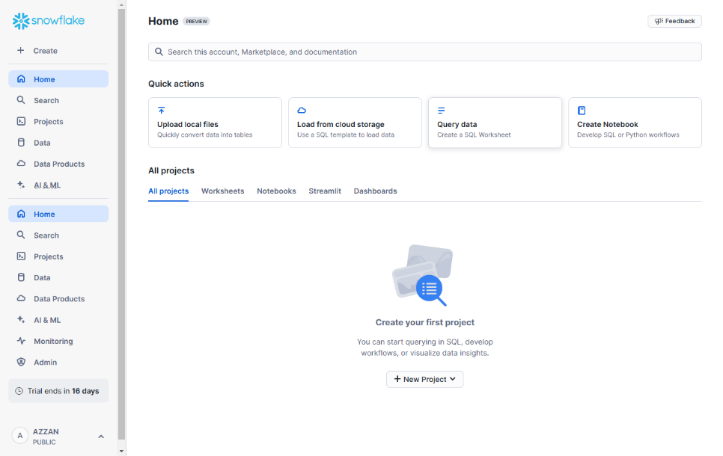


3.1 Data Modeling

A robust data model was designed to organize and structure the data effectively. This involved identifying the key entities and their relationships, such as customers, products, orders, and transactions. By carefully designing the data model, we ensured that data was stored in a consistent and efficient manner, facilitating efficient querying and analysis.

3.2 Snowflake Schema

The Snowflake schema was chosen as the data warehouse architecture due to its flexibility and scalability. This schema organizes data into a hierarchical structure, with normalized dimensions and fact tables. By adopting the Snowflake schema, we optimized data storage and query performance, enabling faster and more efficient data analysis.



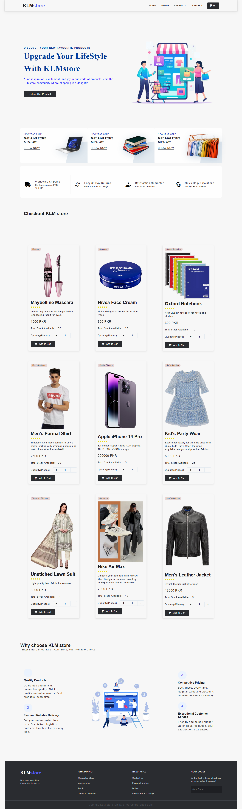
3.3 Data Loading and Refreshment

The extracted and transformed data was loaded into the Snowflake data warehouse using efficient loading techniques. Automated data loading and refreshment processes were established to ensure that the data warehouse remains up-to-date with the latest information from the e-commerce store. Regular data refreshes were scheduled to maintain data accuracy and consistency.

Implementation

4. Data Visualization and Analysis

4.1 To bring this data-driven initiative to life, a combination of technologies was employed. Snowflake was selected as the robust data warehouse platform to store and manage our e-commerce data. SQL and Python were instrumental in extracting, transforming, and loading (ETL) data from the website into the data warehouse. The e-commerce website itself was developed using HTML, CSS, and JavaScript to provide a seamless user experience.



5. Conclusion:

5.1 Summary of Key Findings:

By leveraging the power of data analytics and visualization, we have gained valuable insights into our e-commerce store's performance.

Key findings include:

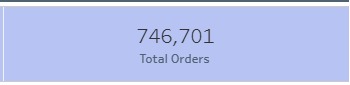
* A total of 114,219 customers have made purchases.



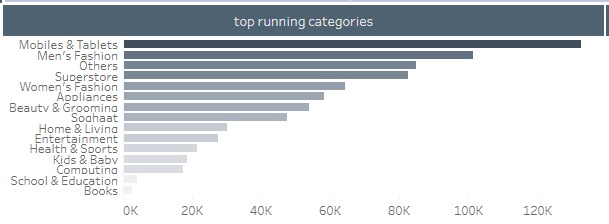
* The total sales generated over the period amount to 4,950,238,930 rupees.



* A total of 746,701 orders have been processed.



* The top-performing product categories are Mobile & Tablets, Men's Fashion, Women's Fashion, Appliances, and Beauty & Grooming.



* The Mobile & Tablets category has been the most successful, generating 2,440,790,600 rupees in sales and processing 132,695 orders over the three-year period.



* These insights will enable us to make informed decisions to optimize our business strategies and drive future growth.

6. REFERENCE:

* + Keggle
  + Chatgpt

7. ER MODEL

