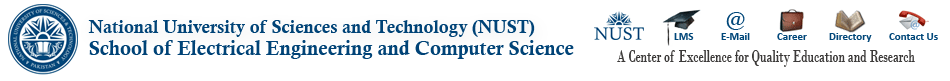
****

# High Impact Skills Development Program

# in Artificial Intelligence, Data Science, and Blockchain

***NAME: SHABBIR HUSSAIN***

***REG NO: SK-23-039***

***Module: Data Mining***

***SUBMITTED TO: Dr.Bilal Ali***

***SECTION: “A”***

***Email address : 0786shabbirhussain@gmail.com***

***github profile link : <https://github.com/101SkShabbirHussainAryanKhaN?tab=repositories>***

**Project Report: SQL Data Analysis for Online Retail**

**Introduction**

In this project, we conducted a series of SQL data analysis tasks to gain insights into an online retail dataset. The dataset contained information about sales transactions, including details about products, customers, orders, and more. The primary objectives of this project were as follows:

1. **Customer Segmentation by Purchase Frequency :**

We aimed to categorize customers into high, medium, and low-frequency segments based on their purchase history. This segmentation helps identify loyal customers and those who may need more attention.

**2. Average Order Value by Country :**

We calculated the average order value for each country to understand where the most valuable customers are located.

**3. Customer Churn Analysis :**

We identified customers who haven't made a purchase in the last 6 months to assess churn. This analysis helps in retaining customers by re-engaging with those who have become inactive.

**4. Product Affinity Analysis :**

We determined which products are often purchased together by calculating the correlation between product purchases. This can be used for cross-selling and product recommendations.

5. **Time-based Analysis** :

We explored trends in customer behavior over time, focusing on monthly sales patterns to understand seasonal variations.

***Methodology and Procedural Aspects***

**Customer Segmentation by Purchase Frequency**

We started by calculating the purchase frequency for each customer. The procedural aspects involved:

* Using the `COUNT` function to count the number of distinct invoices for each customer.
* Grouping the results by `CustomerID`.
* Defining segmentation criteria based on purchase frequency (high, medium, low).
* Using a `CASE` statement to assign segmentation labels.
* The results provided us with a clear segmentation of customers based on their purchase frequency, which can be used for targeted marketing efforts.

**Average Order Value by Country**

To calculate the average order value by country, we followed these steps:

* Aggregated order values by summing the product of `Quantity` and `UnitPrice` for each invoice.
* Grouped the results by `Country`.
* Calculated the average order value using the `AVG` function.
* Ordered the results to identify the countries with the highest average order values.
* This analysis helps in targeting marketing efforts and understanding where the most valuable customers are located.

**Customer Churn Analysis**

For the customer churn analysis, we performed the following steps:

* Calculated the most recent purchase date for each customer using the `MAX` function.
* Filtered customers whose last purchase date was more than 6 months ago.
* The resulting list provided us with customers who may be considered churned or inactive. Strategies can be devised to re-engage with these customers.

**Product Affinity Analysis**

To find products that are often purchased together, we executed a product affinity analysis:

* Created a Common Table Expression (CTE) to find pairs of products purchased together.
* Calculated the correlation coefficient by normalizing the count of purchases.
* The results showed which products have strong associations and can be bundled together or recommended to customers.

**Time-based Analysis**

To explore trends in customer behavior over time, we conducted monthly sales analysis:

* Used the `DATE\_FORMAT` function to extract the year and month from the `InvoiceDate`.
* Aggregated data by month.
* Calculated metrics such as order count, total quantity sold, and total revenue for each month.
* The analysis revealed monthly sales patterns, helping in understanding seasonal trends.

**Conclusion**

In this project, we successfully conducted a series of SQL data analysis tasks to gain valuable insights into customer behavior, purchase patterns, and product associations. These insights can inform marketing strategies, customer retention efforts, and product recommendations.

These SQL queries provided a robust foundation for analyzing and visualizing the data, helping the business make data-driven decisions to improve sales and customer satisfaction.

This report summarizes the project's objectives, methodology, and procedural aspects of the SQL queries performed. Screenshots of executed queries have not been incorporated, but the descriptions provided should give a clear understanding of the tasks completed in this analysis.