**SUMMARY OF SQL COHORT ANALYSIS PROJECT**

**Introduction**

In this project we explored an online retail dataset and created a Retention Cohort Analysis in SQL and Tableau. Cohorts are formed by grouping users or customers based on a common characteristic and tracking their behavior overtime to see patterns and trends.

This data set is gotten from the UCI Machine Learning Repository.

**Website**: [**https://archive.ics.uci.edu/dataset/352/online+retail**](https://archive.ics.uci.edu/dataset/352/online+retail).

**Key** **Columns**: Quantity, InvoiceDate, CustomerID,

The tool used for this project is the **Microsoft** **SQL Server.**

We worked on the following skills:

* Understanding Cohort and Cohort Analysis
* Handling missing values
* Month extracted from date
* Assigning Cohort to each transaction
* Assigning Cohort Index to each transaction
* Create Cohort Table for retention rate
* Visualize the Cohort Table in Tableau
* Interpret the Retention Rate

**Data Cleaning**

We started by cleaning our table step by step using CTEs by identifying the null values. After cleaning we got 397,884 out of 541,909 rows. We also checked for duplicates using one of the Windows Function called ROW\_NUMBER().

**Beginning of Cohort Analysis**

The data required for the Cohort Analysis are:

* Unique Identifier (CustomerID)
* Initial Start Date
* Revenue Data (any data linked to purchase eg. quantity, price, amount)

**Creation of Cohort Group Column**

Described as the last time a customer made a purchase. To get the data we used a Windows Function of MIN() and date format of DATEFROMPARTS(). This is how we got our Cohort Date.

**Creation of Cohort Index Column**

This is an integer representation of the number of months that has passed since the customer first purchase by subtracting the Year & Month of the main table and Year & Month of the Cohort Group. And with the use of Sub-Query, we were able to derive our Index with this formular **Cohort\_Index = Year\_Difference \* 12 + Month\_Difference + 1**. The 1 is the Cohort Index, this indicates that a customer purchased it’s second purchase in the same month after it’s first and if we changed the one to 2, it indicates that a customer purchased its second purchase in the following month.

The next thing I did was to save this data separately and create a Cohort Retention in Tableau. We can also create a Cohort Retention in SQL.

The following steps shows:

In the next steps we derived a DISTINCT in the **CustomerID**, **Cohort\_Date** and **Cohort\_Index** from the saved data and PIVOT the data to see our **Cohort Retention Table**. Lastly, we created a **Cohort Retention Rate** so we can see the percentage of customers that came back the following month.

**Insights**

* I noticed from our **Cohort Retention Rate** and **Cohort Retention Table** that in the first Cohort Period which was 2010/12/01, 50% of the customers came back to purchase our goods in the 12th Month which records the highest percentage of customers that came back since 2010/12/01.
* In a consistent trend in every penultimate month, I noticed that customers usually come back much more than the rest. This trend records approximately more than 75% across our Retention Table.

**Potential Impacts**

If this pattern is consistent, it means customers are disengaging for awhile but return just before they fully churn. This suggests that customers may be reacting to specific triggers, such as renewal discounts or promotional offers.

**Solution**

If retention efforts (e.g., discount, personalized outreach) are driving this behavior, the company might need to refine these strategies to retain customers earlier in the cycle instead of just in the penultimate month.