**SUBSCRIPTION COHORT ANALYSIS**

*SITUATION AND HINT*

**THE SITUATION**

You’ve just been hired as an analytics consultant for Mavenflix, a subscription-based streaming platform with customers all over the world.

**THE ASSIGNMENT**

To better understand customer retention and churn rates, you’ve been handed a dataset that contains subscription data from **Sept 2022 – Sept 2023**.

Your goal is to **transform the data into a monthly cohort analysis** that Mavenflix leadership can use to identify churn trends and patterns.

**THE OBJECTIVES**

Profile and QA (Qualitive Assurance) the raw data

Reshape and prepare the data to support the cohort analysis

Build a monthly cohort visualization to show patterns and trends

***Objective 1***

Profile & QA the data

Your first objective is to profile and QA the data by **removing unpaid customer** subscriptions and understanding duplicate customer\_id instances.

a) Profile and QA the data and ensure data types and column headers are all accurate

b) Remove all customer records where the subscription was not actually paid.

c) KNOWLEDGE CHECK: What does it mean that there are duplicate instances of the same customer identification number in the customer\_id column?

- Due to the customer's re-subscription, they requested cancellation. After that, they re-subscribed, resulting in the data containing this ID multiple times.

***Objective 2 (DONE)***

Prepare the data for analysis

Your second objective is to transform the data into a usable format by creating a record for each month of a customer's active subscription.

a) Transform the canceled\_date column and replace null values using a date literal (e.g. #date) based on the last day of September 2023 and confirm the data type

b) Add a new start of month column based on created\_date called "Created Date (SOM)"

c) Add a new column named "Month Span" to show the number of months between created\_date and canceled\_date (eg customer\_id 154536156 = 14)

d) Add a new column called "Month List", using ListNumbers, to expand "Month Span" into multiple records (eg if "Month Span" = 14, then 14 records will be added with the same customer\_id )

e) Expand the resulting list and update the data types as necessary

***Objective 3***

Build a cohort visualization

Your final objective is to create a **heatmap** to analyze monthly retention rates across the subscription periods in the dataset.

a) Create a matrix that shows "Created Date (SOM)" as rows and "Month List" as columns

b) Create a measure that shows the volume of customers retained by month, called [Customer Retention Volume], and add it to the values field

c) Create a measure that shows the percentage of customers retained by month called [Customer Retention %], and swap it out with [Customer Retention Volume]

d) Add background color conditional formatting and adjust the visual to improve readability

e) Based on your visual, what recommendations might you suggest to the leadership team?

***Final Step***

Final Project Question

In which month were the most new subscriptions created? (Enter the full month name (text only)) (JULY 2023)

**DEFINITIONS**

The cycle billing is 1 month.

If the created\_date = canceled\_date, it will be calculated as a full month (03/09/2023 to 04/09/2023 it is 1 month, 05/06/2023 to 05/09/2023 is 3 months)

If the day of created\_date >= the day of canceled\_date, the month span will be calculated as CREATED\_DATE – CANCELED\_DATE

If the day of created\_adate < the day of canceled\_date, the month span will be calculated as CREATED\_DATE – CANCELED\_DATE + 1

STEPS TO SOLVE THIS PROJECT

*TOOLS: SQL SERVER AND POWER BI*

* Set the date for type MM/DD/YYYY
* Remove all customers that unpaid for the subscription
* Replace all null values on canceled\_date column with the last day of September in 2023 (09/30/2023)

\*\*\* Processing duplicate customers

* **Step 1: Cleaning and Calculating**
* **Cleaning**

+/ This is the big problem: duplicate customers are the customers who re-subscribe for this service after they cancel their subscription.

+/ Classify these duplicate customers:

* ***Rule 1:*** If the customer has (cancellation date – created date) <= 1 month is the retention customer.
* ***Rule 2:*** If the customer has (cancellation date – created date) > 1 month is the win-back customer.
* ***Rule 3:*** Depending on the duration of time subscription:

\*\*\* *For example*, if the customer has a created date is on 03/09/2023 and the cancellation date is on 05/11/2023, then he/she re-subscribes this service on 06/29/2023. The question is “This customer will be classified for Retention Customer or Win-Back Customer?”. Let’s breakdown it.

++ From 03/09/2023 to 04/08/2023: 1 month

++ From 04/09/2023 to 05/08/2023: 1 month

**NOTE**: Because this customer cancels the subscription on 05/11/2023 so he/she will be charged for full the next month is May. If he/she cancels by or on 05/09/2023, he/she won’t be charged for the next month.

++ From 05/09/2023 to 06/09/2023: 1 month (because the customer won’t be charged on the cancellation date which is after the end date of circle billing one)

* This indicates that the customer used this service for 3 months (March, April and May). Although the number of months from 05/11/2023 to 06/29/2023 is 2 months, this is compatible with ***Rule 2*** but this customer will be classified as **Retention Customer**. Because he/she continues to use this in June after 3 used months March, April and May.
* **Calculating**

+/ Calculating Month Span

+/ Creating Month List

* **Step 2: Consolidating**
* If duplicate customers are classified as retention customers, they will be consolidated into one row
* If duplicate customers are classified as win-back customers, retain duplicate customer row numbers, and the second-time or third time they subscribe, this month they will be a new customer.

\*\*\* Processing one-time subscription customers: calculating Month Span and creating Month List.

* Last, compiling duplicate customers table and one-time subscription customers table table into Table (dataset). Visualizing this dataset with Power BI.