Subscription Churn
Team 5

Javeed Ahmed

Afreen Sakina Tariq

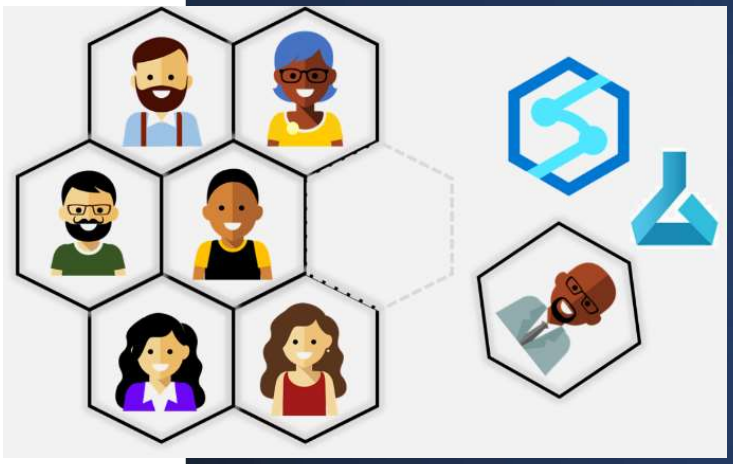
Hanna Uzair

Humza

Sep 1st, 2022

Team - Background

Sakina Igaadi - From Morocco, from Beni Mellal and she work as a teacher, and is interested in data science. She holds a license in human resource management.
Batool Abidi – Grasduated from undergrad and want to expand her skills.
Uzair Khalil – Comes formCivil Engineering background and is planning to enter data science field.
Afreen Nikhat - Computer science Engineer. Experienced in Web Designing and website Mgmt.
Hanna Khan – Worked as BI & Reporting Analyst, Finance & Corp. Mgmt.
Javeed Ahmed – Service Delivery Manager at an IT Software & Service Organization



Project - Background

Analyze the Churn Rate for a Subscription-based personal finance business and make projections about the future.

Customer Churn is important in any business. The ability to predict that a particular customer might churn could impact revenue, market share, and reputation

A minimum churn rate and higher retention rate bring stability to any business; therefore, we aim to help the business minimize the churn rate.

Predictable, recurring revenue

Higher customer lifetime value

Build strong customer relationship

Objectives

Calculate	Predict	Predict	Build
• Calculate average churn rate over the last 9 months for each price tier.	• Predict the number of currently active subscriptions that will still be active next month.	• Predict the number of currently active subscriptions that will still be active in 3 months.	• Build a separate model that predicts tenure based on price tier, source, and country. Provide actionable insights to the business.

5 Questions (3 Descriptive and 2 Predictive)

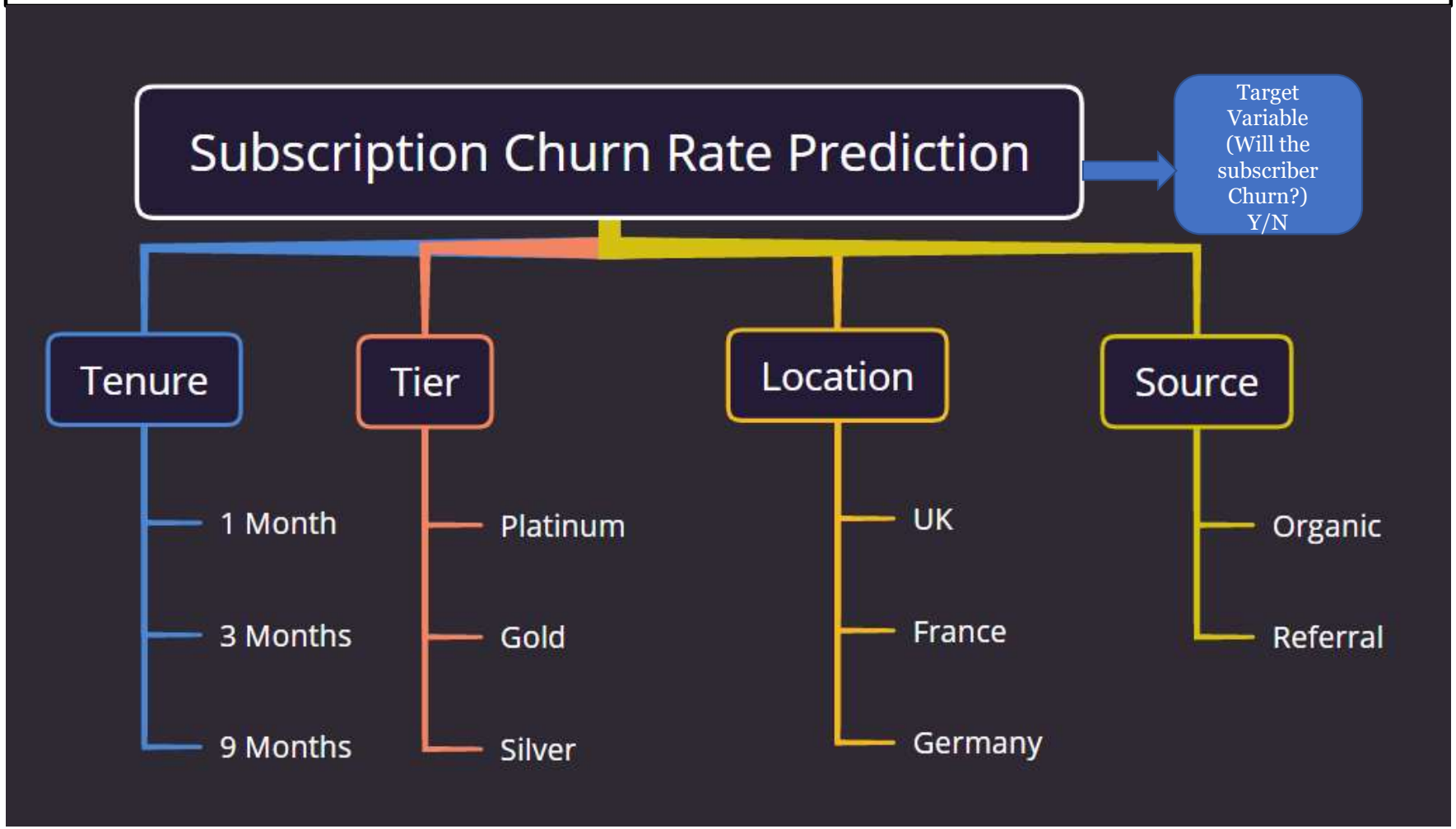
Descriptive

- Which tier had the most Churn in the past 9 months?
- Which Source contributed to the highest number of Subscriptions?
- Which country did most subscribers come from?

Predictive

- What is the likelihood that the number of currently active subscriptions will still be active next month?
- What is the likelihood that the subscriptions in the platinum price tier would have the maximum churn rate?

Analytical Base Table – Subscription Churn



As-is-state & To-be-state

As Is State - Currently, the business doesn't have a system in place to determine the subscription churn rate.

Loosing customer is a concern for businessas they do not want to lose the revenue

To Be State - Create a model to determine/predict the subscription churn rate for the business and avoid revenue loss

Communication Channel

- Whatsapp
- Slack
- Google Meet
- Google Drive

Subscription Churn – Presentation – 2

Team Members

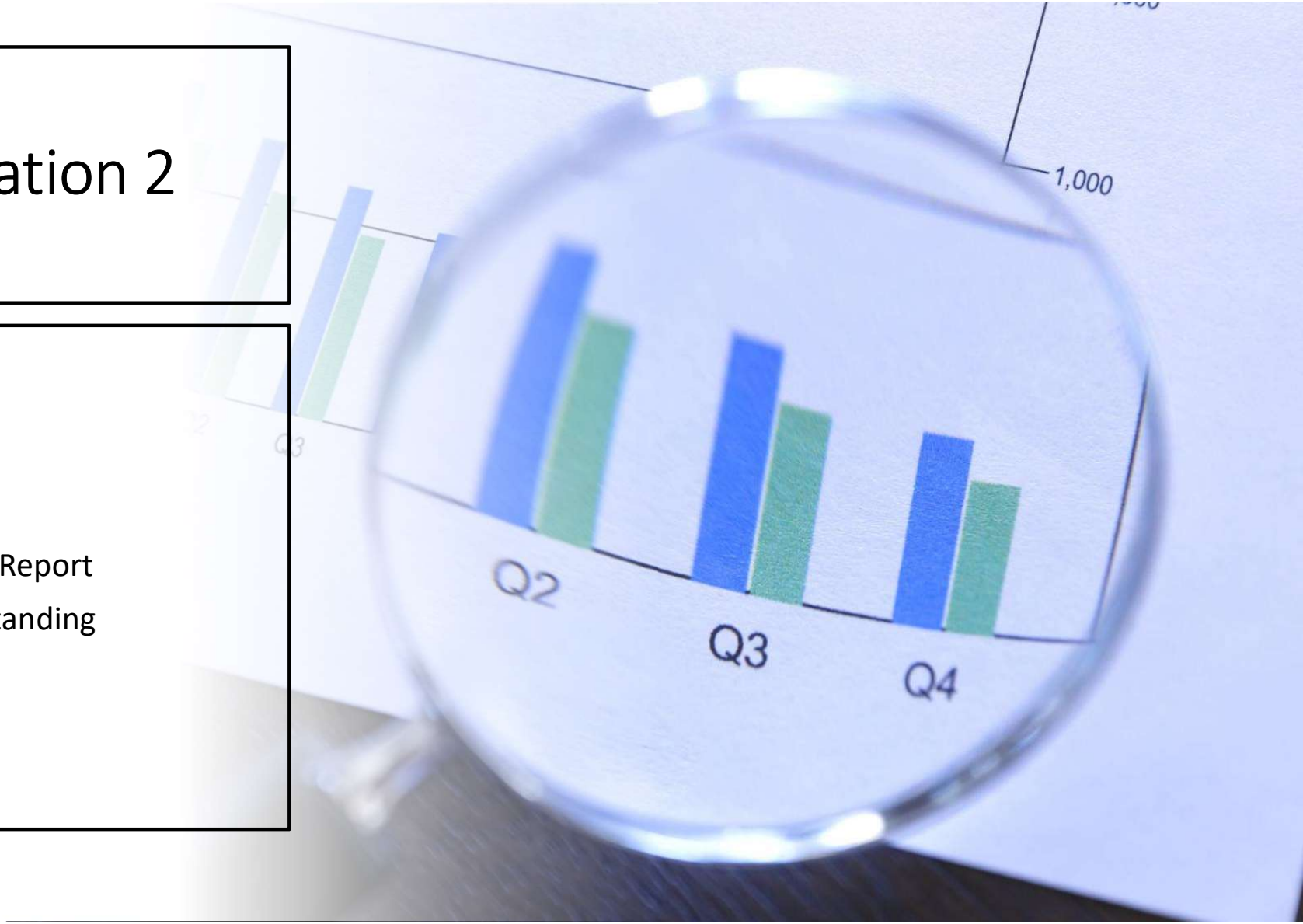
- **Javeed Ahmed**
- **Afreen**
- **Sakina**
- **Hanna**

Oct 6th, 2022

Presentation 2

Agenda

- EDA
- Data Viz
- Data Quality Report
- Data Understanding
- Appendix



Exploratory Data Analysis

Dataset Statistics	
Number of Variables	6
Number of Observations	99,944
Missing Cells	None
Duplicate Rows	None
Variables Types	
Categorical	3
Numeric	3

Data Dictionary

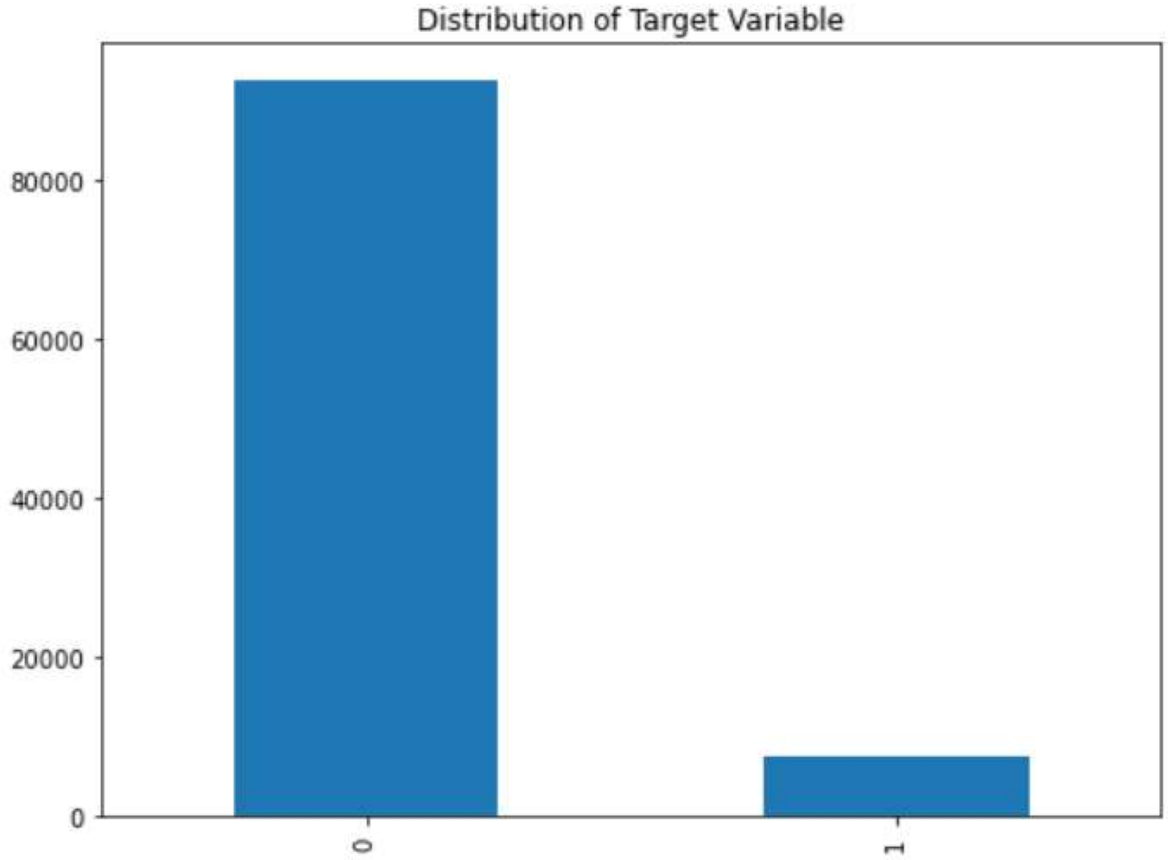
- **member_id** - Unique ID of the user.
- **tier** - Price tier (Silver, Gold, or Platinum).
- **country** - Member country.
- **source** - Original acquisition channel.
- **tenure** - Number of cycles billed. Min is 1. Max is 9.
- **active** - Is the subscription still active?

DATA QUALITY REPORT

Salient Features		Data Observations/Description	
Data Type	Categorial , Numeric,	Record Count	99,944
	Categorical		
3	Source (3) – Organic, Referral, Partnership Country (8) – US,UK, Germany, Netherlands, France, Spain, Italy, Canada Tier (3) – Gold, Silver, Platinum	Skewness	Active, Tenure- Positively skewed to the right.
3	Numeric Member ID, Tenure (9 Months) Active – Active or Not active	Outliers	Tenure, Active – Outliers to the right of the distribution
6	Total Features	Duplicates	None
TV	Active (Churn) – Binary 0/1	Missing Values	None
TV(D)	Churn Rate	Special Characters	None

Data Understanding

- Target Variable (TV) is Imbalanced
- Churn subscribers(0) = 92.5%
- Active (1) = 7.4%

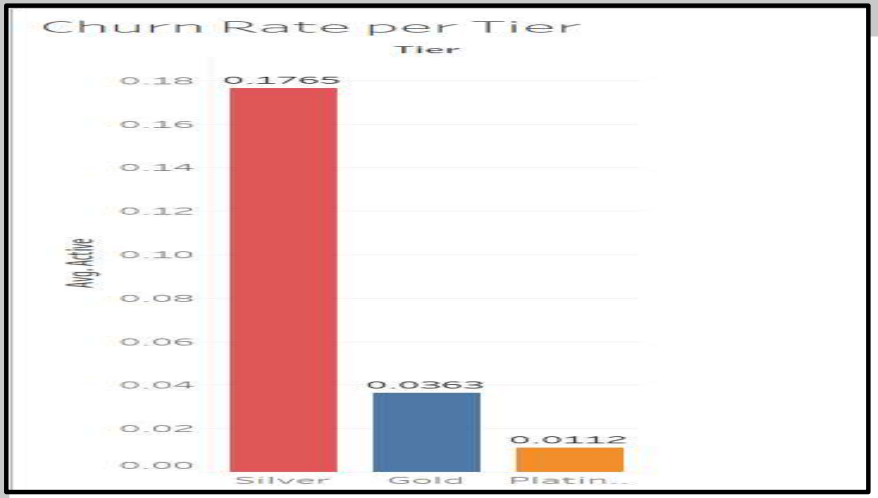


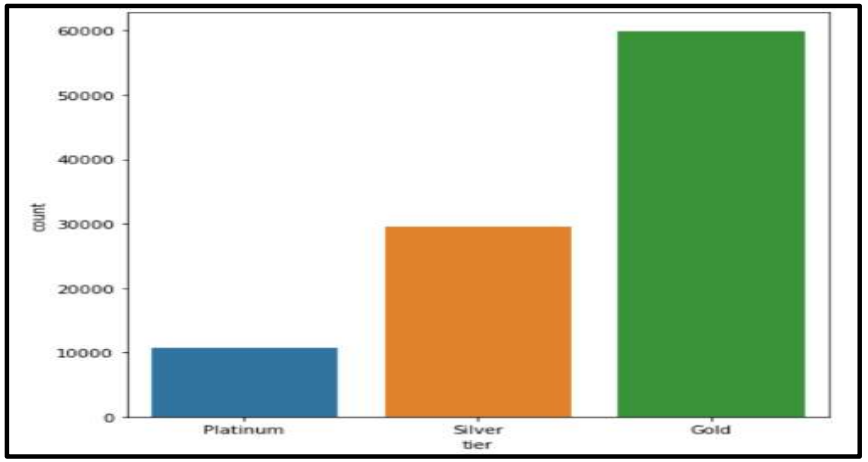
Data Understanding – Churn Rate per Tier

Which tier had the most Churn in the past 9 months?

- Silver tier had the highest churn rate of 17.65%
- Gold had the highest subscriber count.

Details	Gold	Platinum	Silver	Grand Total
Churn Count	57669	10516	24270	92455
Active Count	2171	120	5204	7495
Active Rate (%)	3.62	1.12	17.65	22.39
Grand Total	**59840**	**10636**	**29474**	**99950**

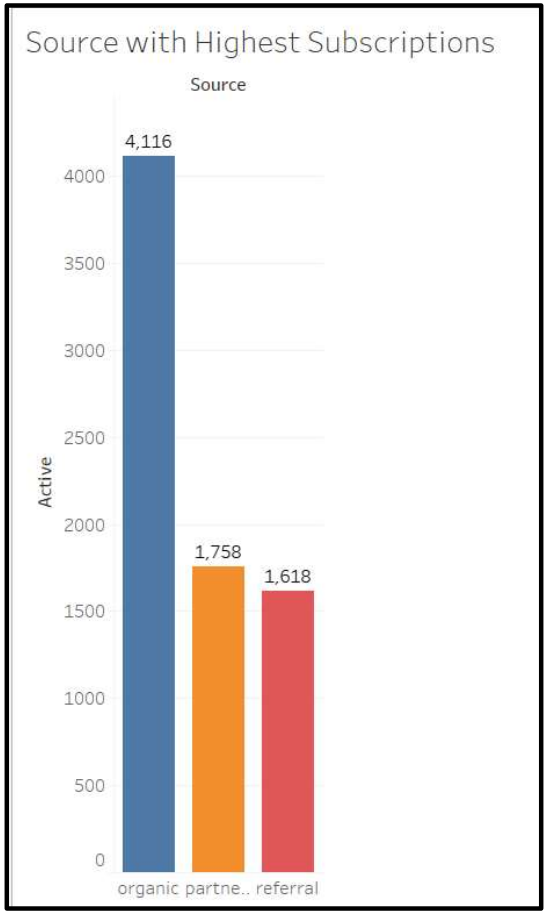




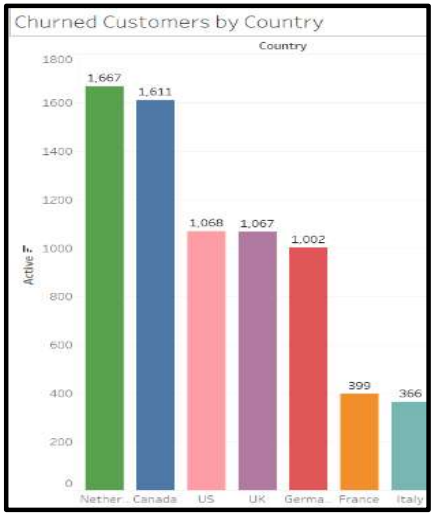
Questions from Presentation 1

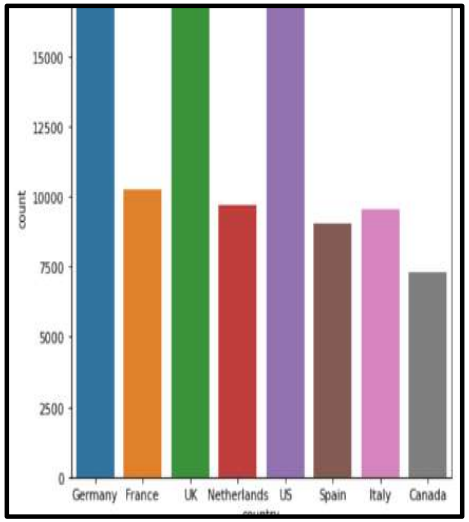
Which Source contributed to the highest number of Subscriptions?

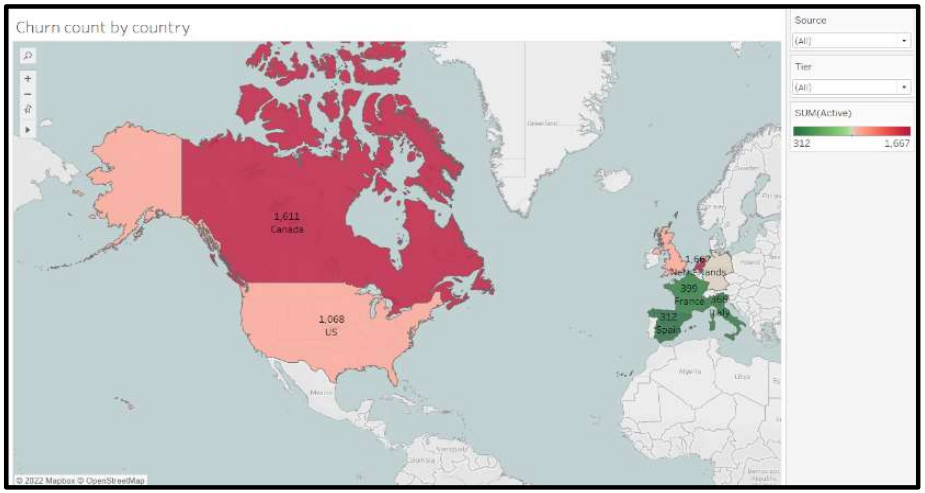
- Organic source had 82,873 Subscribers and 4116 Subscribers are active within the 9-month period.



Questions Answered - Data Visualization







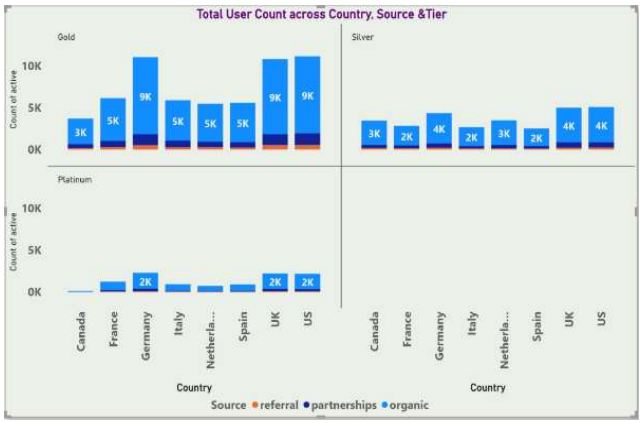
Which country did most subscribers come from?

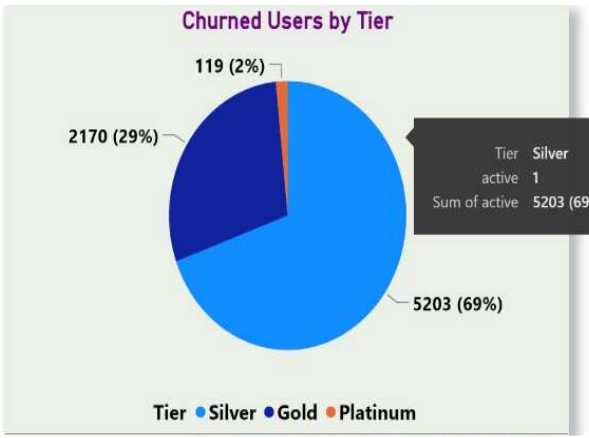
- Most Subscribers came from the US
- Netherlands had the highest churn 9694 subscribed and 1667 are active

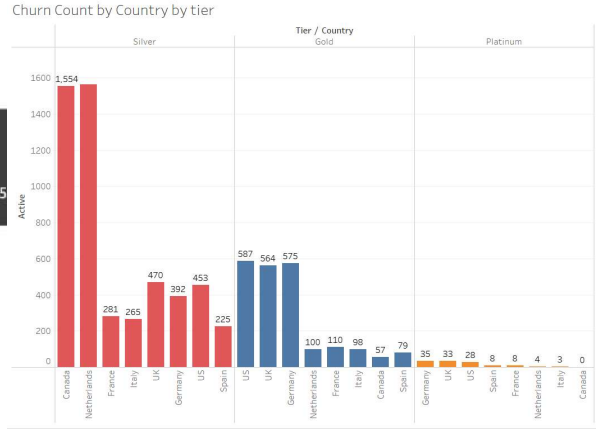
Which country had the most subscribers (count)?

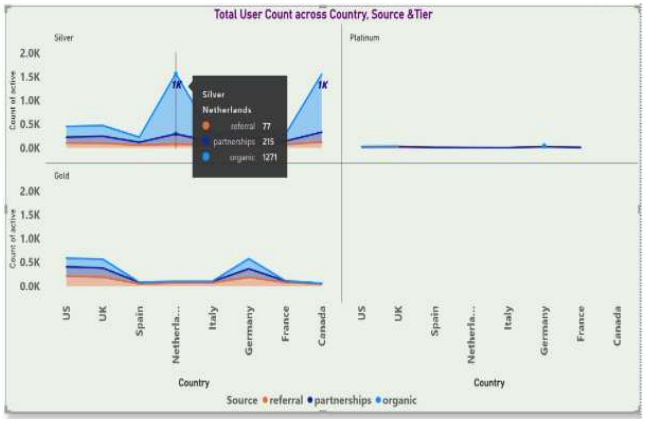
Countries	Total Count	Active Count
Germany	17689	1002
Netherlands	9693	1667
UK	18033	1067
US	18425	1068
Grand Total	**63840**	**4804**

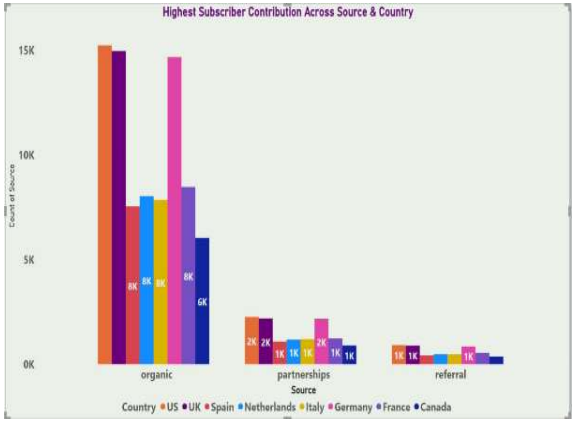
Appendix


Total User Count across Country, Source &Tier
Gold
Silver
Platinum
Count of active
Canada
France
Germany
Italy
Netherla...
Spain
UK
US
Country
Source referral partnerships organic


Churned Users by Tier
119 (2%)
2170 (29%)
5203 (69%)
Tier Silver
active 1
Sum of active 5203 (69.45
Tier Silver Gold Platinum

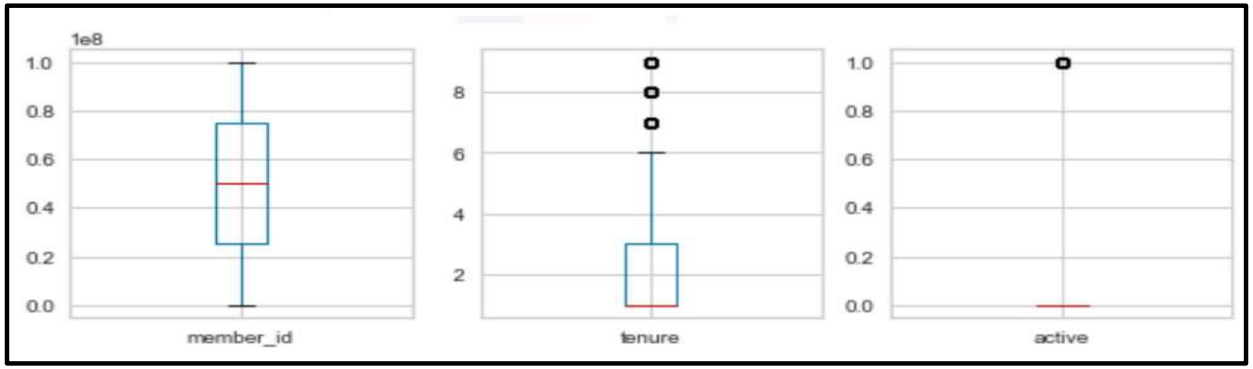

Churn Count by Country by tier
Tier / Country
Silver
Gold
Platinum
Active
1,554
281
265
470
392
453
225
587
564
575
100
110
98
57
79
35
33
28
8
8
4
3
0
Canada
Netherlands
France
Italy
UK
Germany
US
Spain


Total User Count across Country, Source &Tier
Silver
Platinum
Gold
Count of active
Silver
Netherlands
referral 77
partnerships 215
organic 1271
US
UK
Spain
Netherla...
Italy
Germany
France
Canada
Country
Source referral partnerships organic

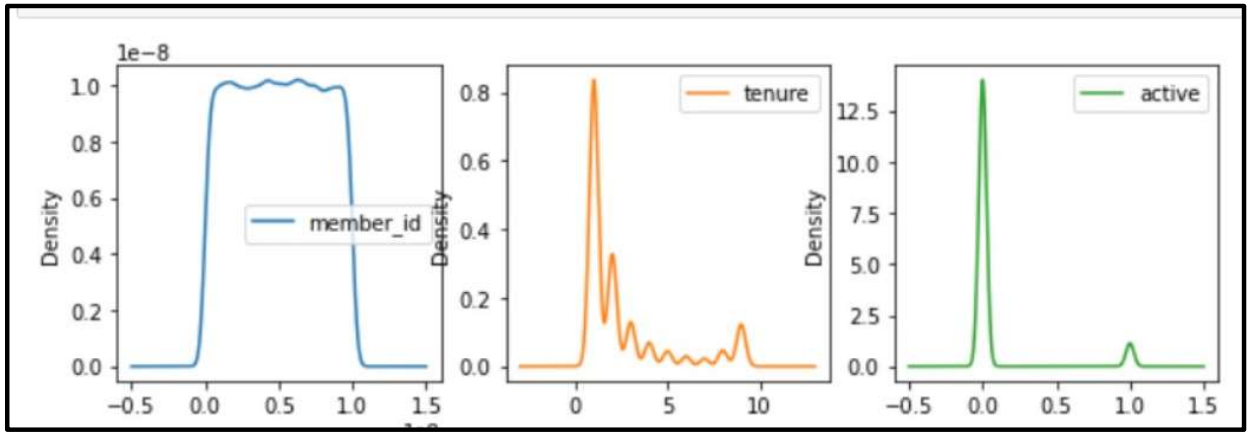

Highest Subscriber Contribution Across Source & Country
Count of Source
organic
partnerships
referral
Source
Country US UK Spain Netherlands Italy Germany France Canada

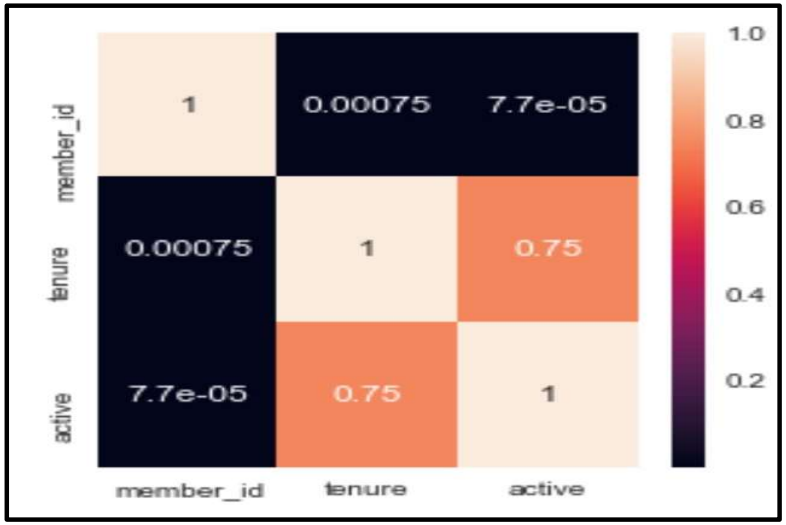
Appendix – Data Visualization

Outliers in Tenure depicts that most customers started churning in later months



Positively skewed to the right depicting outliers





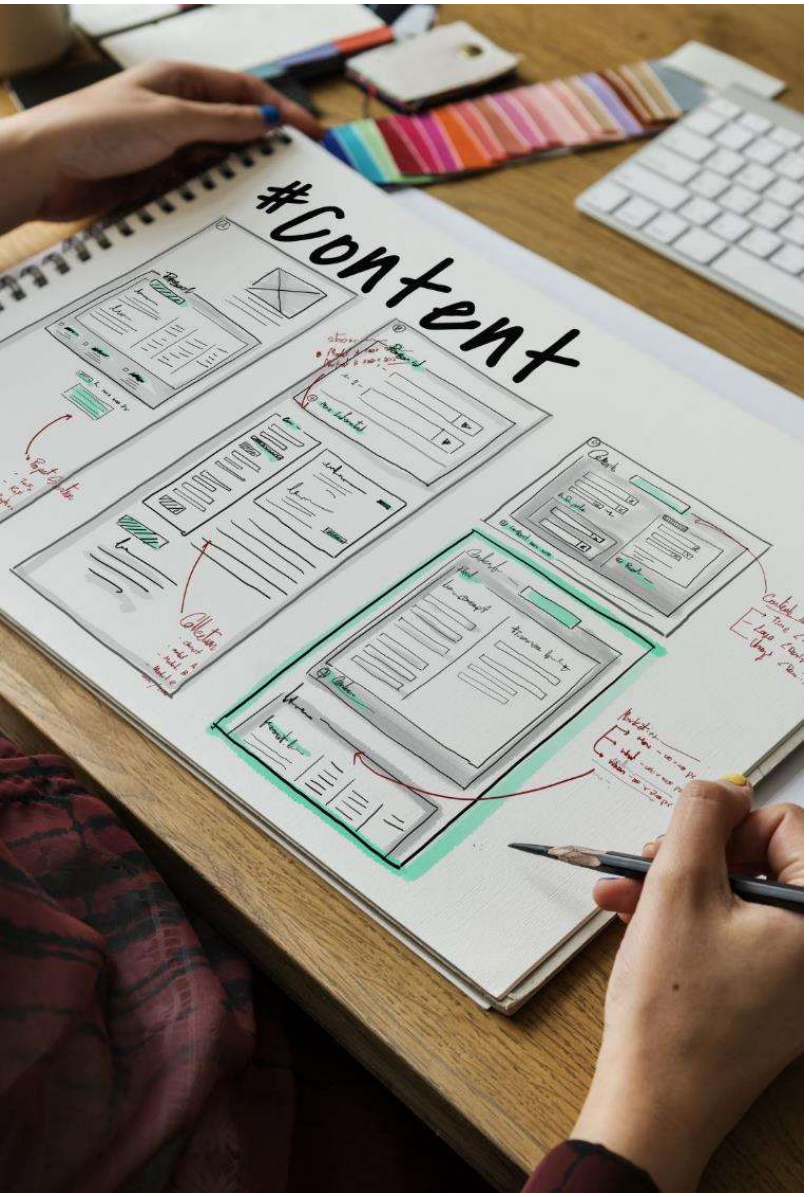
Tenure and Active features are highly correlated which depicts that most customers churned during the later months.

Subscription Churn – Presentation – 3

Team Members

- **Javeed Ahmed**
- **Afreen**
- **Hanna**

Nov 7th, 2022



Agenda

- Data Preparation
- Modelling Technique
- Data Ready for Modelling
- Appendix

Data Preparation

- Total Number of Features = 6
- Numeric Features = 3
- Categorical Features = 3

- Replace method (mapping) to transform the ordinal column (tier)
- Use one hot encoding (Dummies) to transform the nominal features.
- Dropped member ID

Data Preparation

Scale our Data except not required for Decision Tree
Use Precision and Recall as evaluation metric before balancing the data
Use SMOTE (Synthetic Minority Over sampling technique) to Balance out TV and check accuracy.
Use Binary Classification Algorithms like DT, NB, SVM, LR to predict our TV.

Subscription Churn – Presentation – 4

Team Members

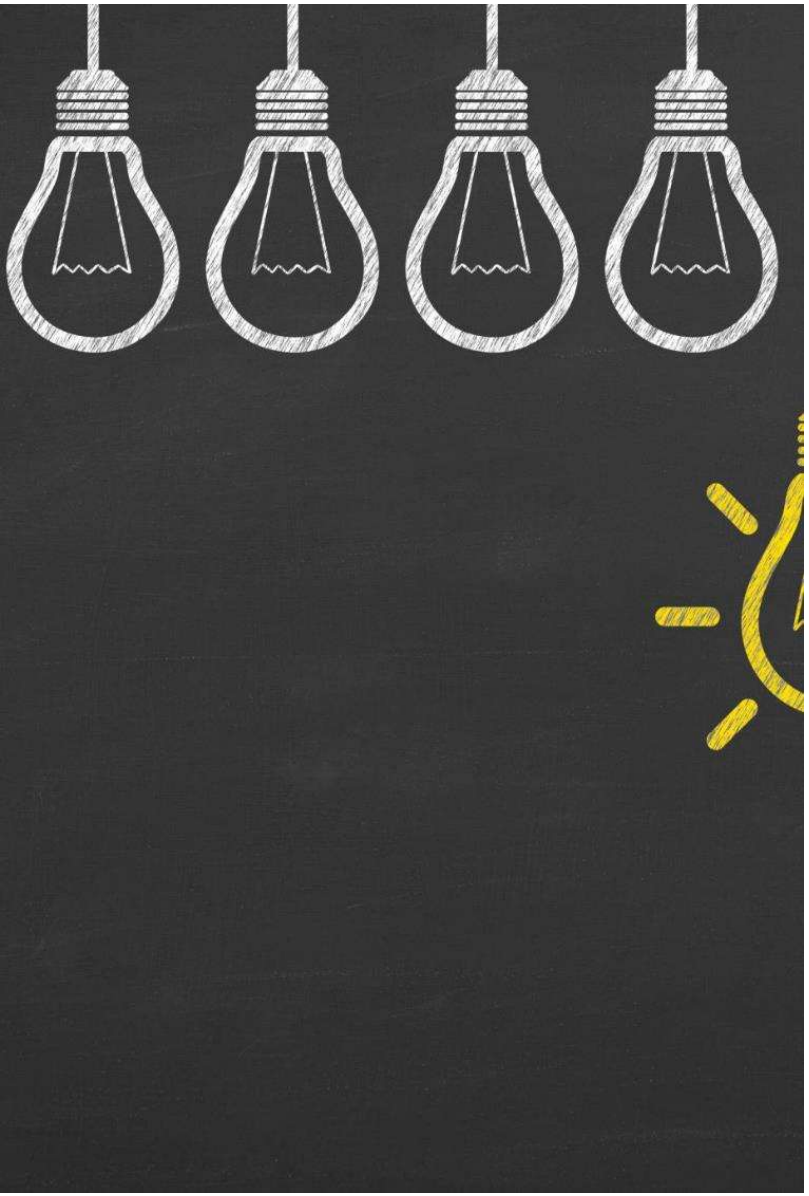
- **Javeed Ahmed**
- **Afreen**
- **Hanna**

Dec 9th, 2022

Subscription Churn
Team 5 – Presentation 4

Javeed Ahmed
Afreen
Hanna

Dec 9th, 2022



Agenda

- Data Modelling
- Model Comparison
- Model Implementation
- Appendix

Decision Tree

Used DT Classifier to train our Data without scaling our Balancing our Data

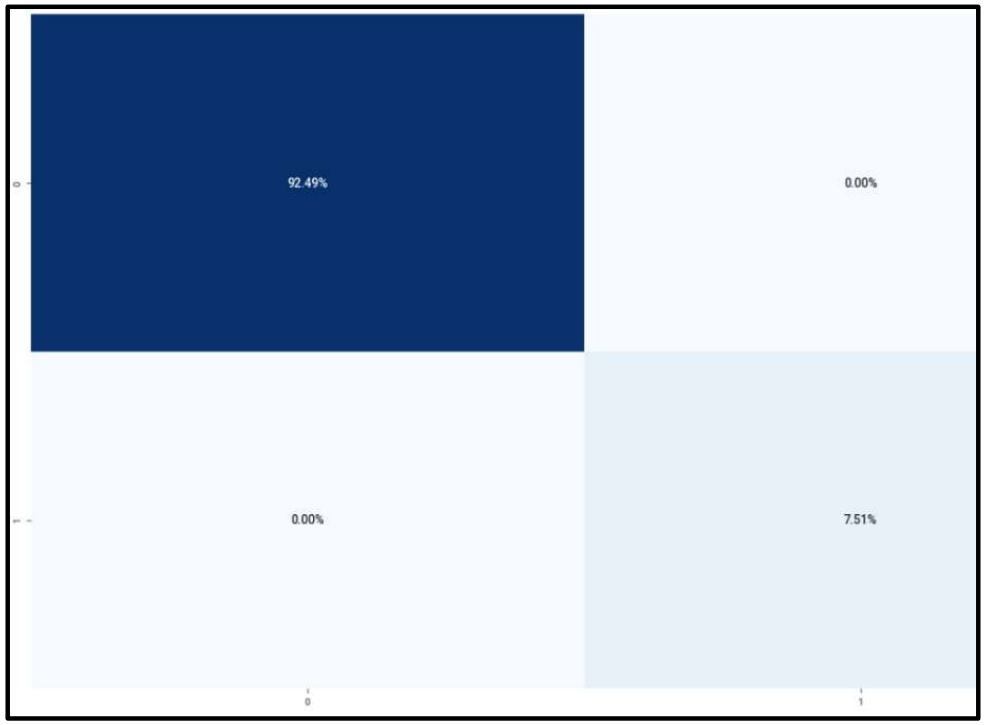
Achieved an accuracy of 100% as expected because data is imbalanced as it might have captured the same class only.

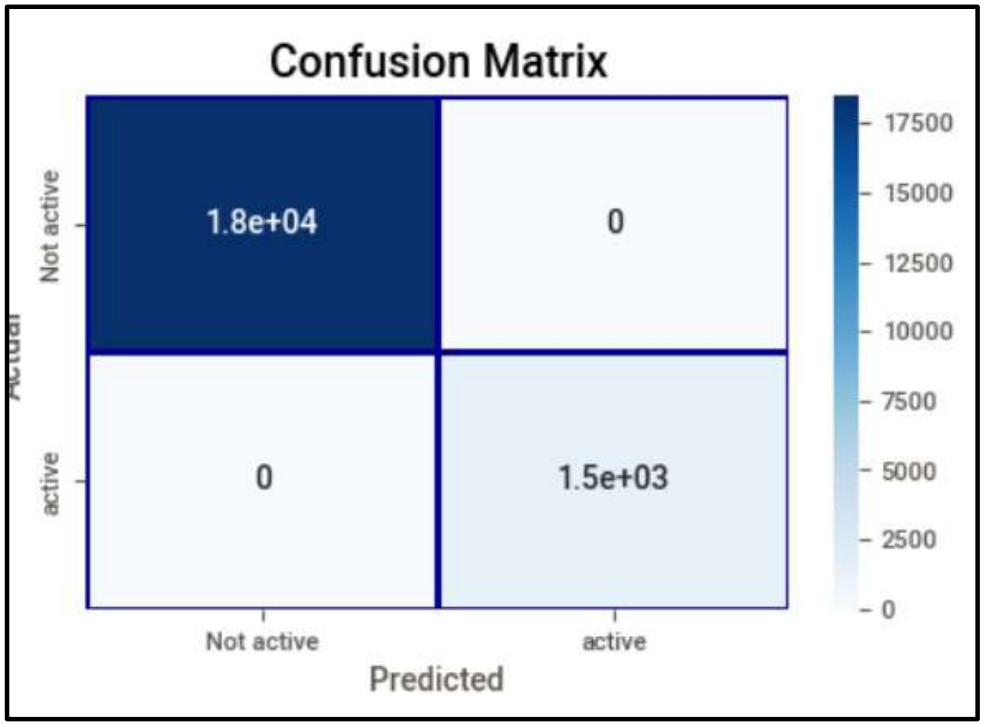
Precision and Recall to be determined, also 100%

Will check for accuracy once again after balancing and scaling the data

Naïve Bayes

Accuracy = 100%





Model Comparison

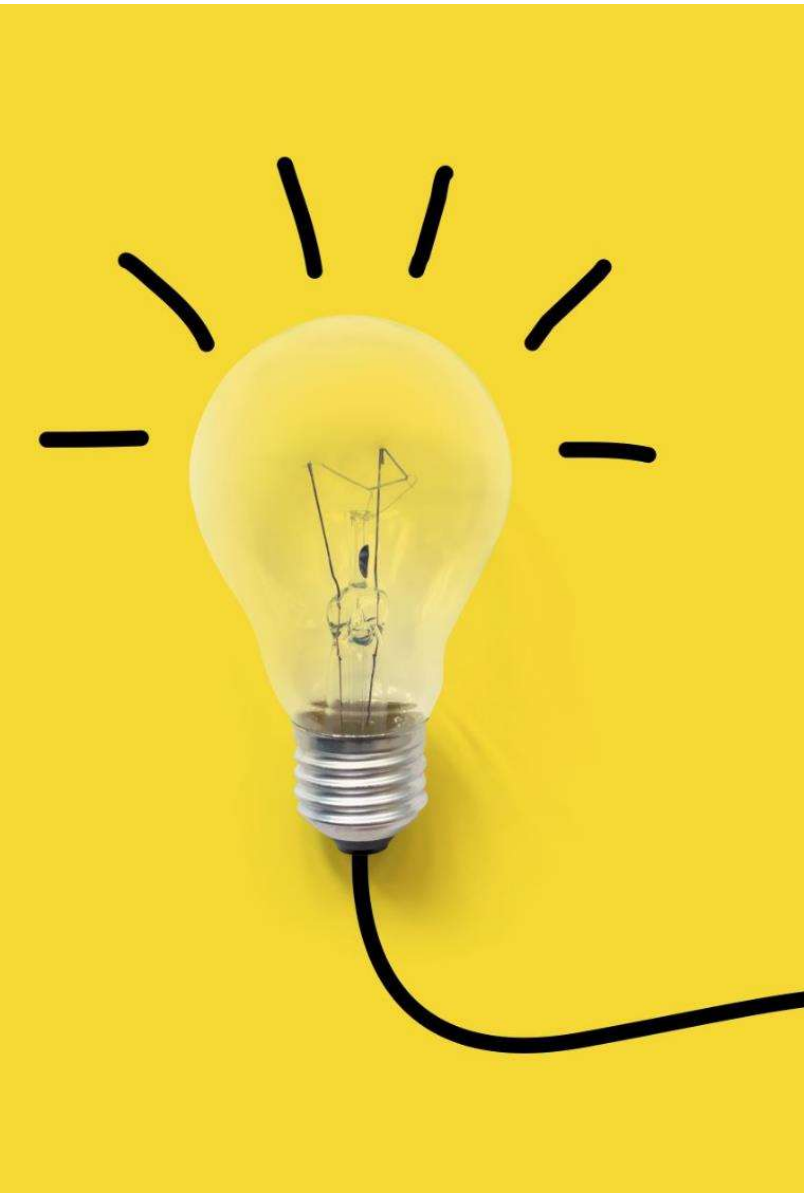
- Run another Classification Models and pick the best model
- Optimize the best model by using Hyper parameter tunning

Subscription Churn – Presentation – 5

Team Members

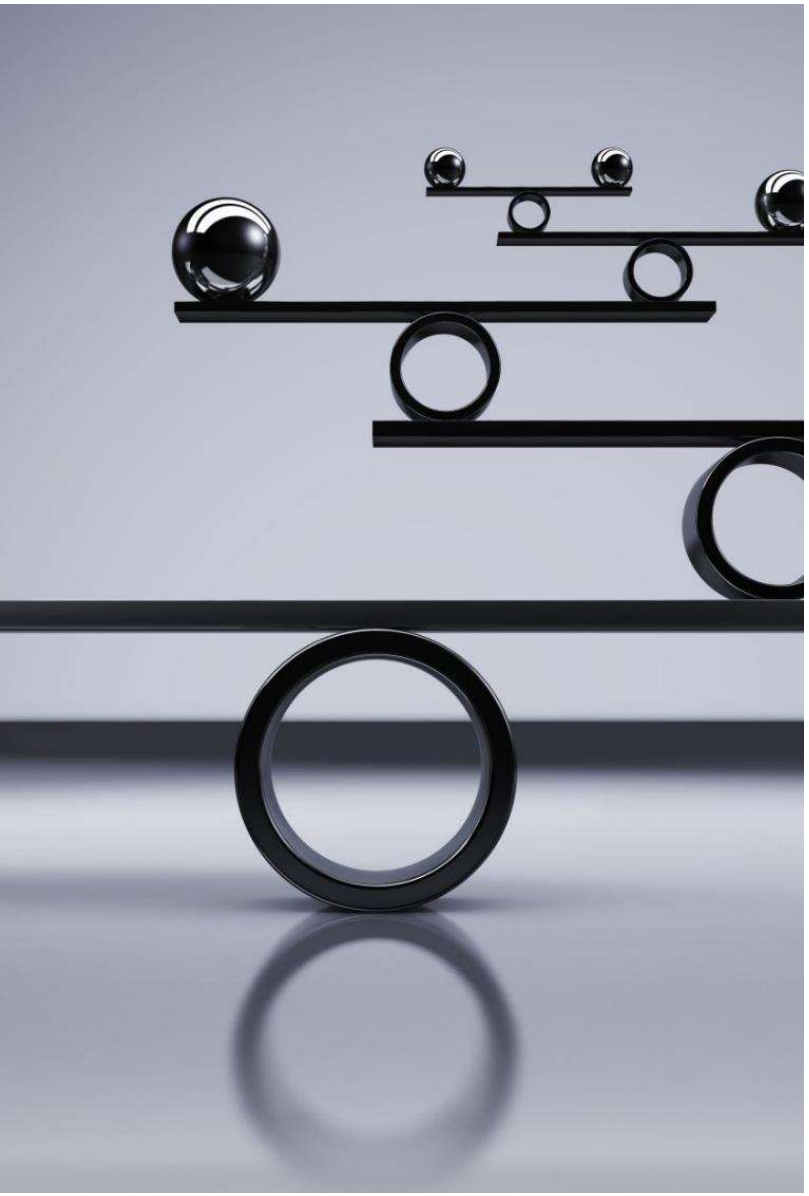
- **Javeed Ahmed**
- **Afreen**
- **Hanna**

Jan 6th, 2023



Agenda

- End to End Model Delivery
- Action Plan
- Client Recommendation
- Appendix

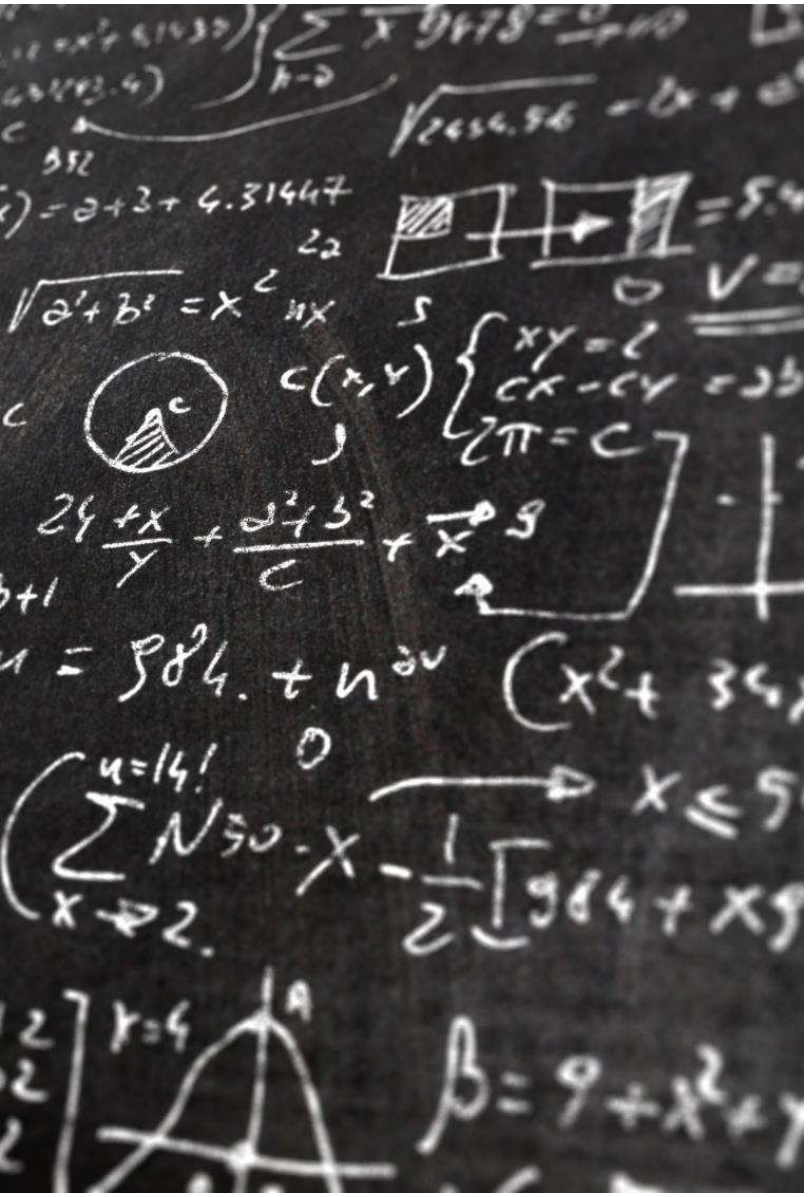


Model Delivery

- End to End Binary Classification problem where we predicted if a customer would churn in 3 months or 9 months. The model prediction was 100% because we had class imbalance, so we looked at Precision which is True Positive / True Positive + False positive got a value we got a value of 1.

- Only 1 Feature Tenure has the highest impact on our TV that is Churn.

```
****************** Accuracy Score Prediction ********************
*                                                                 *
*        Dummy Classifer - Baseline               : 92.51%        *
*        Decision Tree  Accuracy                  : 100.00%       *
*        Naive Bayes  Accuracy                    : 100.00%       *
*        Logistic Regression Accuracy             : 92.62%        *
*        Random Forest Accuracy                   : 100.00%       *
*        SVM Prediction Accuracy                  : 100.00%       *
*        Kernalized SVM Accuracy                  : 100.00%       *
*        KNeighbors Accuracy                      : 99.80%        *
*                                                                 *
*******************************************************************
```

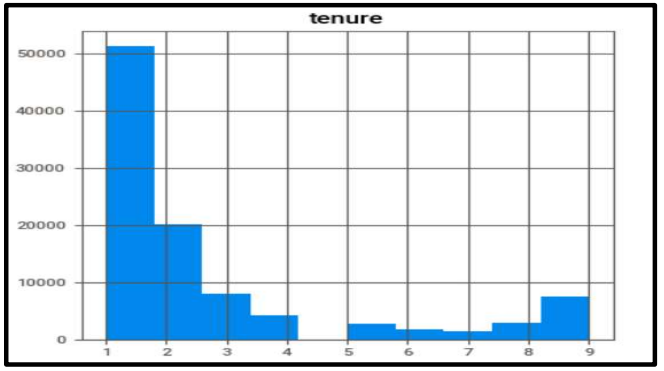
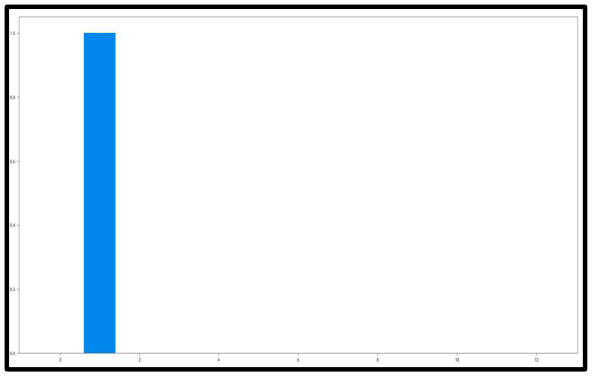
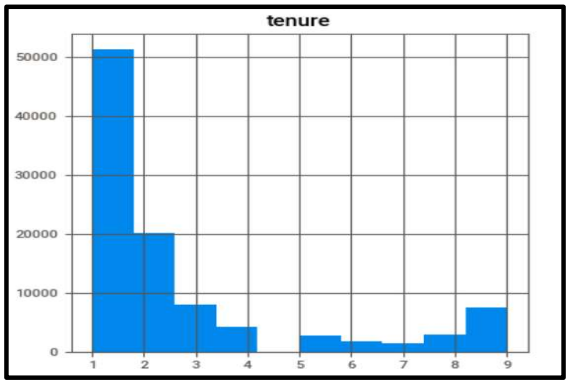


Action Plan

- In our use case in which we observe more data points of one class than of another, the accuracy is no longer a useful metric. Using a different sample size and balancing the data using SMOTE achieved 100% accuracy, meaning that 92% of customers will churn.
- Extract More Features or look for revised data
- Look into different features that is impacting churn rate.
- Churners are generally more recent subscribers, with a median subscription of approximately two months. This information can be used to pinpoint milestones during the customer subscription journey. For example, once a user reaches two months into their subscription, businesses should consider providing promotions, incentives, and features that will encourage them to stay.

Appendices

	Tier	Count	Churn Rate against Tier	Overall Churn Rate against Tier
Active	Gold	57668		
	Platinum	10515		
	Silver	24269		
	Total	92452		
Churned	Gold	2170	3.76%	2.2%
	Platinum	119	1.13%	0.1%
	Silver	5203	21.44%	5.2%
	Total	7492		
	Grand Total	99944		7.5%

Churn Count by Country by tier

	Tier		
Country	Silver	Gold	Platin..
Canada	1,554	57	0
France	281	110	8
Germany	392	575	35
Italy	265	98	3
Netherlands	1,563	100	4
Spain	225	79	8
UK	470	564	33
US	453	587	28