*Institute of Business Administration, Karachi*

**CS 416**

**Business Intelligence   
Final Exam  
Spring 2023**

**Date:** 21st May 2023  
**Duration:** 3 hours  
**Total Marks: 25%**

**Instructors:**  *Dr. Tariq Mahmood (Class # 70594)*  
*Ms. Abeera Tariq (Class # 70689)*

**REPORT TEMPLATE**

Chosen Dataset: **Predicting Credit Card Customer Segmentation**

Link: <https://www.kaggle.com/datasets/thedevastator/predicting-credit-card-customer-attrition-with-m?resource=download>

Colab Notebook Link: https://colab.research.google.com/drive/1NXDdxa\_knIRiZKZX9vDSz8peY\_aphBbd?usp=sharing

Team Members:

|  |  |  |
| --- | --- | --- |
| Sno. | Name | ERP |
| 1 | Muhammad Mujtaba Khan | 18624 |
| 2 | Sarmed Ahmed Usmani | 19673 |

**Section A: Problem Statement**

We have basically made our analysis on customer churn and segmentation over the variables:

* Education\_Level
* Gender
* Months\_on\_book
* Months\_Inactive\_12\_mon
* Avg\_Utilization\_Ratio
* Income\_Category
* Card\_Category
* Total\_revolving\_Balance

**Section B: Data Cleaning and EDA**

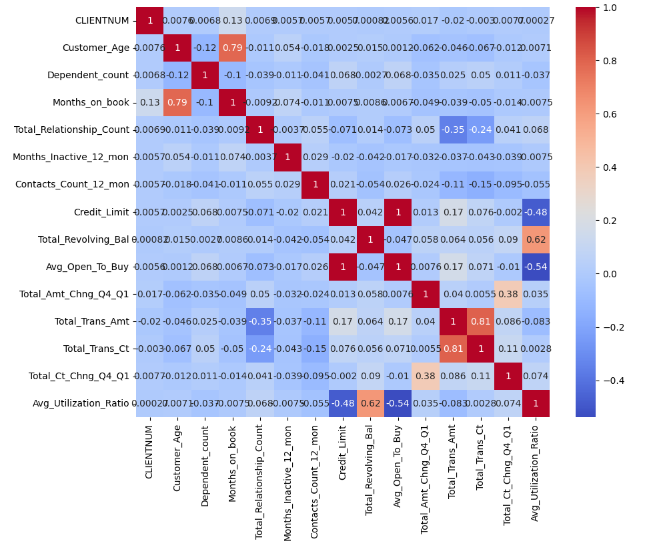
* In the Dataset, there are no missing/null value and also redundancy as there is unique column of “CLIENTNUM”.
* We have dropped these columns as they were irrelevant to our analysis:

1. Naive\_Bayes\_Classifier\_Attrition\_Flag\_Card\_Category\_Contacts\_Count\_12\_mon\_Dependent\_count\_Education\_Level\_Months\_Inactive\_12\_mon\_1
2. Naive\_Bayes\_Classifier\_Attrition\_Flag\_Card\_Category\_Contacts\_Count\_12\_mon\_Dependent\_count\_Education\_Level\_Months\_Inactive\_12\_mon\_2

* Numerical columns having outlier:

1. Months\_on\_Book
2. Credit\_Limit
3. Total\_Revolving\_Bal
4. Avg\_Open\_To\_Buy
5. Total\_Amt\_Chng\_Q4\_Q1
6. Total\_Trans\_Amt
7. Total\_Ct\_Chng\_Q4\_Q1
8. Avg\_Utilization\_Ratio

* Heatmap of the correlation matrix

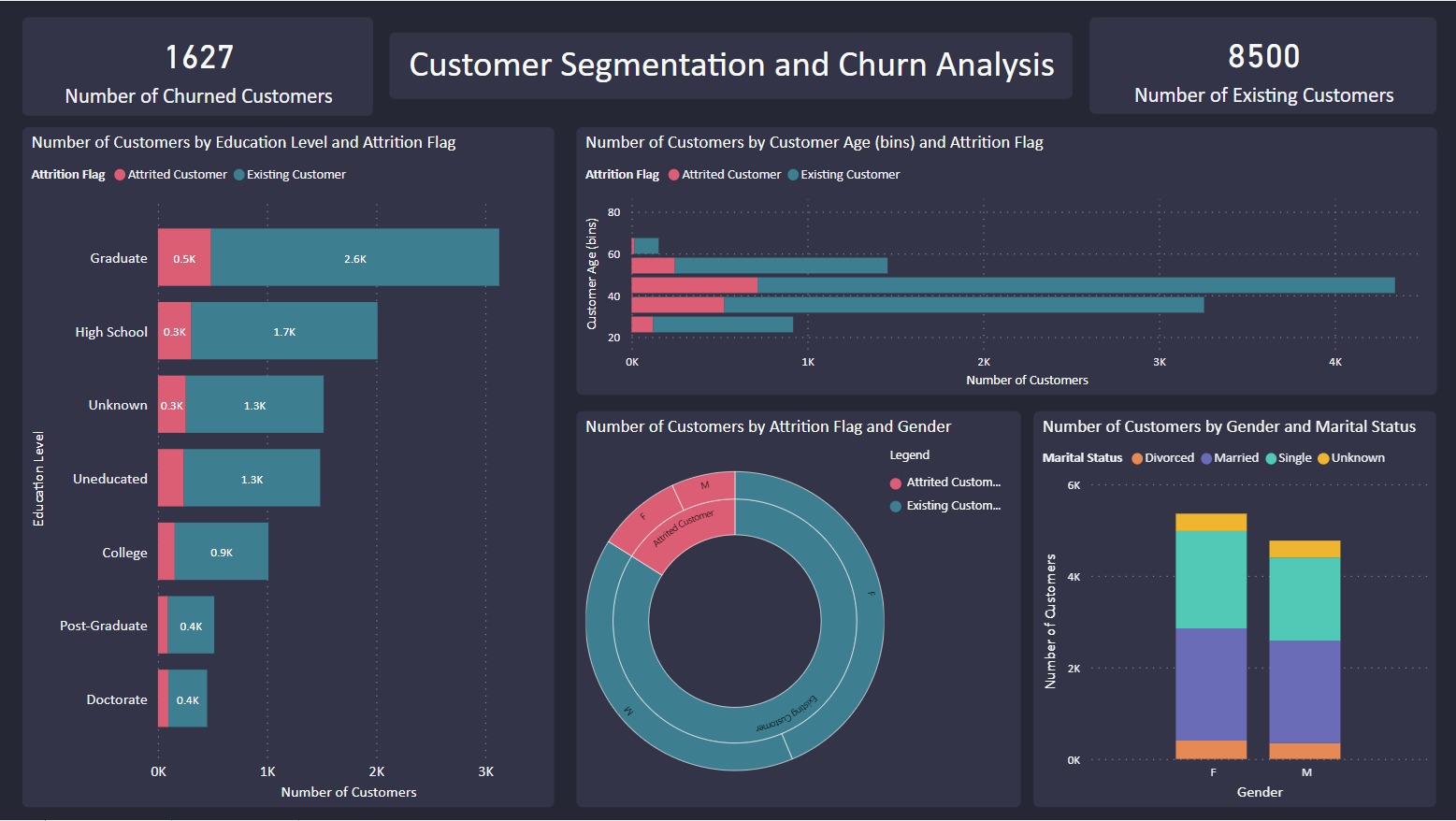


1. Avg\_utilization\_Ratio is negatively correlated with Credit\_Limit.
2. Avg\_utilization\_Ratio is negatively correlated with Avg\_Open\_To\_Buy.
3. Avg\_utilization\_Ratio is positively correlated with Total\_Revolving\_Bal.
4. Total\_Trans\_Amt is negatively correlated with Total\_Relationship\_Count.
5. Total\_Trans\_Ct is negatively correlated with Total\_Relationship\_Count.
6. Total\_Trans\_Ct is positively correlated with Total\_Trans\_Amt.
7. Avg\_utilization\_Ratio is positively correlated with Total\_Revolving\_Bal.
8. Customer\_Age is positively correlated with Months\_on\_book.

* We have used ANOVA and Chi-Square test on our columns for doing Bivariate analysis, the results are shown in the main function of Colab notebook file.

**Section C: Dashboards and Stories of Customer Profiles**

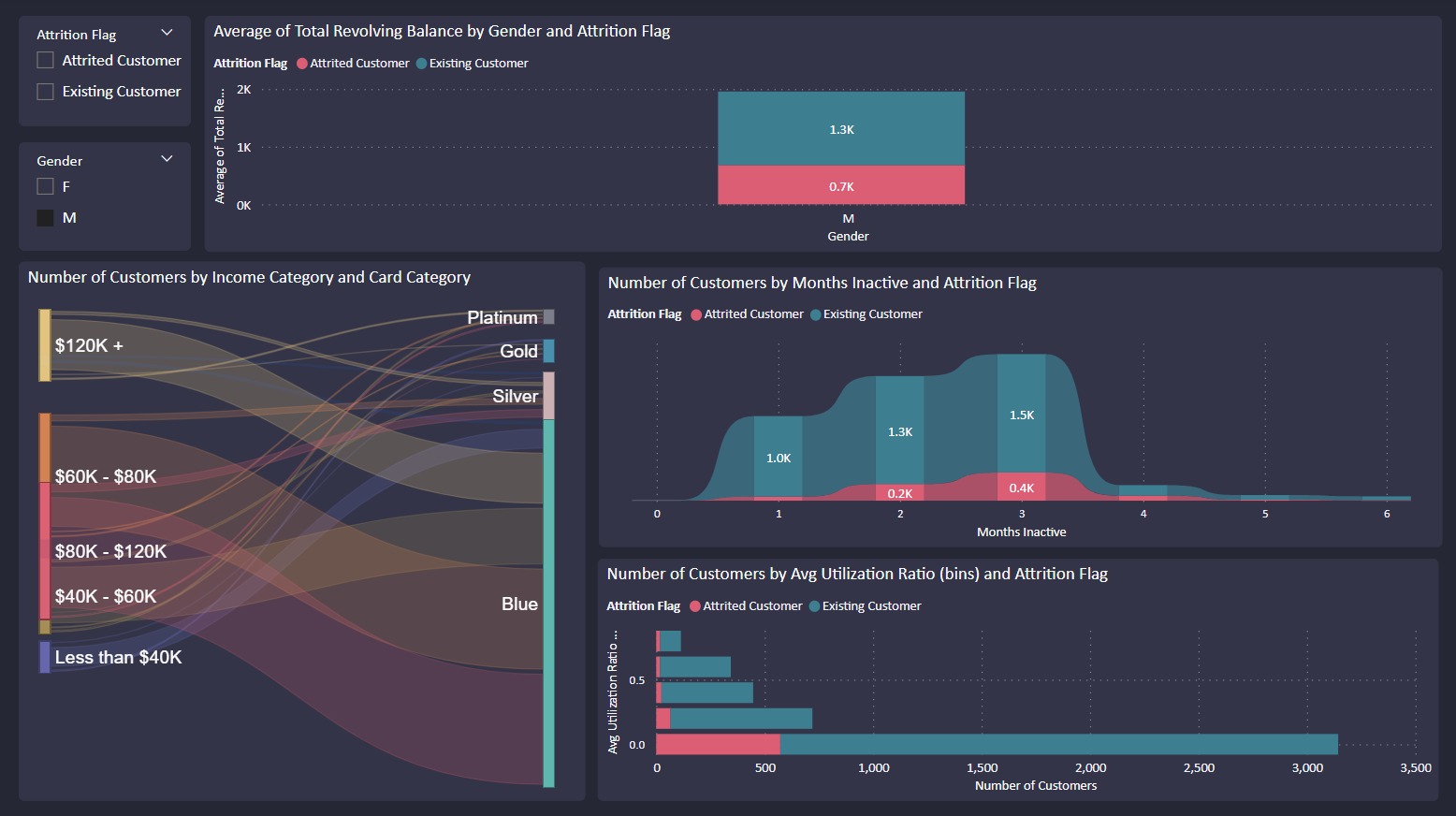
Dashboard 1:



Insights:

* 16% of our customer base has churned, indicating a significant churn rate.
* The majority of our customers have a Graduate education level. Additionally, the age bins between 35.40 and 44.80 have the highest number of customers.
* The attribution rate across all education level categories and customer age bins is proportional to the number of customers in each category. In other words, the number of churned customers is not biased towards any specific category.
* Customers who are either married or single tend to open the highest number of accounts compared to other marital status categories.

Dashboard 2:



Insights:

* The average total revolving balance for both genders is $1800.
* Existing customers have an average total revolving balance of $2500, which is significantly higher than the average balance of churned customers, which is approximately $1300.
* According to the Sankey chart, the highest number of blue card holders belongs to the income category of less than $40K, accounting for approximately 93% of the total blue card holders.
* When analyzing gender differences, we find that the income category of $80K - $120K is the highest for males, while for females, the highest category is less than $40K.
* The number of customers who have been inactive for 4, 5, or 6 months is relatively low compared to those who have been inactive for 1, 2, or 3 months. Among the inactive periods, the highest customer count is associated with an inactive period of '3' months.
* The bin range of the average utilization ratio from 0 to 0.2 has the highest number of customers, precisely 5334 customers. Furthermore, this bin range also exhibits the highest attribution rate at 22.3%.

**Section D: Final Set of Insights – What is the answer to the problem statement?**

* Approximately 16% of our customers have churned, indicating a significant churn rate.
* The majority of our customers have a Graduate education level, and the age bins between 35.40 and 44.80 have the highest number of customers.
* The attribution rate for churned customers is independent of any particular category, indicating that the number of attributions is not biased towards a specific category.
* Customers who are married or single tend to open the most accounts compared to other marital status categories.
* The average total revolving balance for both genders is $1800.
* Existing customers have a significantly higher average total revolving balance of around $2500 compared to churned customers, which is approximately $1300.
* In the Sankey chart, the income category of less than $40K has the highest number of blue card holders, accounting for about 93% of total blue card holders.
* For males, the income category of $80K - $120K is the highest, while for females, the category of less than $40K is the highest.
* The number of customers inactive for 4, 5, or 6 months is relatively low compared to those inactive for 1, 2, or 3 months. The highest customer count is associated with an inactive period of 3 months.
* The bin range of the average utilization ratio from 0 to 0.2 has the highest number of customers, specifically 5334 customers. Additionally, this bin range also exhibits the highest attribution rate at 22.3%.

**Section E: References**

If you used any links, make sure to enlist them in this section.

Colab Notebook Link: https://colab.research.google.com/drive/1NXDdxa\_knIRiZKZX9vDSz8peY\_aphBbd?usp=sharing