**SMARTBANK’s DATA ANALYSIS REPORT**

This report outlines the foundational steps taken toward building a predictive model for customer churn at Smart Bank. It includes data selection, exploratory data analysis (EDA), and data cleaning and preprocessing, ensuring the dataset is prepared for model development.

**RATIONALE**

To determine the root cause to why customers are leaving Smart Bank to other competitors, I used all data sets in the Customer Chern Data excel file, to determine t which groups where the highest and therefore the most likely to move to competitor companies. Using all the data sources was necessary to gain a better understanding on who was leaving and predict why they may have left, to quickly solve the problem reducing the chances of remaining customers from leaving. To predict this accurately All data sources need to be evaluated and processed for analysis. Using the 5 data sources I was able to create 9 data visualisation and analysis, creating 5 more data sets, which were processed and grouped based on what needed analysing.

**📊** Data Sets Created Include:

1. **TimeLine**
2. **Churn\_based\_on\_service\_usage**
3. **Churn\_based\_on\_transactions**
4. **Churn\_based\_Democraphics**
5. **Churn\_based\_customer\_service**

**Timeline**

This data set contains the customer’s ID the Last time they logged in and the Churn status. The data was used to create an interactive timeline on when customers left them and when they left the least in focusing on the year 2023 which can then be selected to only show specific month ranges, this data can then be used as a tracking tool, to see which changes Smart Bank implemented in those specific months, in which they can create a new plan to continue with the change or remove the change completely, additionally the time frame can be used to identify the changed competitors implemented in their system etc.

**Churn\_based\_on\_service\_usage**

This data set contains the customer’s ID, their Service Usage and lastly their churn Status. This data was used to determine which service that had the most and least churns, From the data set Mobile Apps had the most churn cases, which therefore translates to the need for updates and changes to the bank’s mobile app, as there could be something causing customers to leave. The service with the least churn cases was the website, so even though it may still need checking, it isn’t at the top of the bank’s concerns.

**Churn\_based\_on\_transactions**

This data set includes, the customer’s ID, the amount they spent on their product, the price range, the product purchased and lastly the church status. The data here was use to get two different insights, which I used to visualise two different groups, the church amounts based on the price range and the product, which were visualised differently using a bar line and a pie chart, and from the data, the price range that had the least amount of churn was 0-50, and the product that had the most was 150-200, which either means that the price range 150-200 have products that aren’t worth that price or the products at the range aren’t what customers at Smart Bank want.

**Churn\_based\_Democraphics**

This data set includes the customer’s ID, their age, their age group, their marital status, their income status and lastly their churn status. With the data gathered I was able to create 3 different insights, which were all visualised using bar charts, which were based on the customers age group, marital status and income status, and from the data it was determined that the age group that had the least amounts of churns were 15-20 and the age group that had the most was 40-45, the top three with the must churns were all 40+, which means that perhaps Smart Bank isn’t pleasing to the older generation, which could mean that it isn’t as easy to use as it should be, leading the older generation elsewhere.

The second insight was marital status, the highest churn rate was with married individuals, which means that the bank isn’t considering married couple, maybe the bank needs to introduce shared accounts or mortgage options the group with the least churn rates are single customers.

The third insight was Income status, and the income level that had the most amounts of churn levels were customers with a low-income level, therefor the bank needs to introduce new concepts perhaps interest on customers that save etc. and the customers with the least amounts of churn levels were medium earners.

**Churn\_based\_customer\_service**

This data set includes the customer’s ID, their interaction type, the resolution status and lastly their churn status. With this data I was able to create to insights using a bar chart and a pie chart. The first insight focuses on churn levels based on the interaction the customer had, and the interaction with the most churn cases was feedback, so customer service and help desk employees need to be retrained on how to help customers resolve certain issues, or if the problem is with their customer service skilled, they could either be retrained or new employees could be hired, however if the processed is given to a bot, the bot will need updating.

Surprisingly for the second insight the resolution status that had the most churn cases were the resolved ones, which either means that employees are stating that unsolved cases have been solved, or they weren’t solved in a way the customer wanted, and was still sent as solved, this is a serious issue, and if not resolved in due time some employees will need to be replaced as the results push the belief that employees aren’t following proper customer service protocols. So calls may need to be recorded for monitoring.

**VISUALISATION AND STATISTICAL SUMMARIES FROM THE EDA**

A screenshot of a computer

AI-generated content may be incorrect.

**TimeLine**

A screenshot of a graph

AI-generated content may be incorrect.

**Service Usage**

A screenshot of a spreadsheet

AI-generated content may be incorrect.

**Transactional Data**

A screenshot of a computer

AI-generated content may be incorrect.

**Demographic**

A screenshot of a spreadsheet

AI-generated content may be incorrect.

**Interaction**

**♻️ THE DATA CLEANING PROCESS STEPS TAKEN**

To process the data sets, I used the automatically collected data from the bank and created new data sets based on it manually, while still using automated techniques to reduce gathering and processing time and to make the data collected more accurate by minimising human error.

I put all the data sets in a table format in order to create pivot tables and easily visualise the data accurately and to create a consistent format for easier understanding and using.

After all the data was gathered and put in tables I determined the best visualisation format to use whether that be bar chart, line chart or pie chart, I made sure to choose the data visualisation format that would show the data collected more accurately.

Then I colour coded all the different data sets, so that they would be more easily identifiable, in the dashboard.

Lastly after double checking all the data created, making sure it’s all relevant and useful to determine why churn happens and making it easy to predict what could be done to resolve churn I then created a dashboard including all the visualised data.

**📊 THE CLEANED AND PROCESSED DATA SET READY FOR MODEL BUILDING**

A screenshot of a data analysis dashboard

AI-generated content may be incorrect.

**FINISHED DASHBOARD – Ready for machine learning models.**

**Prepared by:** Stephane U.G

**Date:** 10/07/2025