

## Task 6.1 - Sourcing Open Data

### Data Source - [Bank Customer Churn Dataset](#)

#### Summary

This banking dataset captures essential customer information, including credit scores, demographics (country and gender), age, tenure with the bank, account balance, number of products held, credit card status, active membership, estimated salary, and churn rate. It provides insights into customer behavior and financial interactions, offering a comprehensive view for analyzing factors influencing customer retention and financial health within a banking context.

#### Why I chose data of Bank.

I chose banking data for my project because it provides a rich and complex dataset to analyze. I am particularly interested in studying how social, environmental, and economic factors influence the quality of life at both global and national levels over time. Banking data offers numerous variables that allow me to explore these dynamics thoroughly and test various hypotheses related to economic development and social well-being.

#### Limitations and ethical considerations

The bank dataset comprises customer information such as credit scores, demographics (country and gender), age, tenure with the bank, account balance, number of products held, credit card status, active membership status, estimated salary, and churn rate. Descriptive statistical analysis can reveal key insights: summary statistics like mean, median, and standard deviation elucidate the distribution and central tendencies of numerical variables such as credit scores, age, and balances. Frequency counts uncover the prevalence of categorical variables like country and gender. However, limitations include challenges in accessing complete and accurate data due to potential sample biases or incomplete entries, impacting data reliability. Ethical considerations include safeguarding customer privacy and ensuring fair data usage to avoid biases and uphold transparency and accountability.

[Data Profile](#) for Bank customers

**Key questions?**

What is the distribution of customers across different countries in the dataset?

Is there any discernible pattern or relationship between credit scores and countries in the dataset?

What are the key summary statistics (mean, median, standard deviation, etc.) for credit scores in the dataset?

How does the churn rate differ between customers who hold credit cards and those who do not?

How does the distribution of account balances vary across different countries?