**COMP5310 – Project Stage 1**

Student Name – Veena Vemula

Student ID – 490484287

Uni key – vvem2507

**Problem –**

Customer churn also known as customer attrition occurs when customers or subscriber’s stop’s relationship with a company or service. Banks, insurance companies, streaming services companies and telcom service companies, often use customer churn analysis and customer churn rates as one of their key business metrics because the cost of retaining existing customers is far less than acquiring a new customer - earning business from new customers means working leads all the way through the sales funnel, utilizing your marketing and sales resources throughout the process.

Churn in banking can occur in any of the following but not limited to –

* Account Closure
* Deposit closure
* Credit card closure
* Fasttag for toll gates

Banks usually make a distinction between voluntary churn and involuntary churn. Voluntary churn occurs due to a decision by the customer to switch to another bank . Involuntary churn occurs due to circumstances such as a customer's relocation to a long-term care, death, or the relocation to a distant location. In most applications, involuntary reasons for churn are excluded from the analytical models. The focus is to concentrate on voluntary churn by analysing the customer bank relationship which bank controls like account Interactions, billing interactions credit cards/rewards, Loan Customers, and demographics.

Our Problem at hand is to identify customers that are most likely to continue using banking services and the ones that are highly unlikely.

**Data**

Taking the problem in hand, I have searched for several different churn data sets available on public domain’s based on criteria -amount of data ( 27000 rows with 31 different attributes), target variables , supporting features for analysis in proposed approach like customer account usage details, App users, Loan customers and Loan details and other relevance features keeping mind the age-old IT wisdom garbage in – garbage out. The data set has been selected from Kaggle located at URL <https://www.kaggle.com/santoshd3/bank-customers>.

**Data Cleaning** - After Importing the data into python, it has been observed that continous features and categorical features have null values in both recognisable and unrecognizable format by tool. These have been replaced by mean and mode. Features have been examined carefully for their contribution and removed any unnecessary as an example Zodiac sign.

A picture containing diagram

Description automatically generated