Bank churn prediction

Problem statement

Banks as financial institutions operate by contractual or subscription model when customers are paying for the services like deposit and credit management, brokerage and so on. When customers stop using bank services or cancel their regular payments they can be concerned as churn.

Churn is a measurement of the percentage of accounts that cancel or choose not to renew their subscriptions if we are considering a subscription or contractual model. Usually it is measured for a specific snapshot of time as month, quarter, year. Non contractual models can include other parameters taken into account for churn definition like - not using credit card for several days or weeks, not performing specific actions with product for required time. Such cases should be defined by company management, business department and data analysts working on such problems.

Each customer contributes to profit growth and losing customers can be a great problem for any business as the cost of gaining new customers can be 10x higher than measures required to retain existing ones. It can require additional budget spending on marketing, advertisement, sales activities etc. Another problem with losing customers is that long-term relationships with customers can generate more profit for the company than newly acquired ones.

The main benefit of churn rate management is that it can contribute to profit growth through reallocation resources for measures to keep existing customers and improve their satisfaction rate for provided services. So churn rate management is a part of the retention management program.

Dataset

Since customers are the most valuable assets of most banking institutions, it is advantageous for banks to know what leads a client towards the decision to leave the company. Provided dataset has 14 explanatory columns and 1 outcome for churn customers. Data is provided as some timeframe snapshot and has data for 10000 users. CVS file was downloaded from Kaggle.

Scope of solution space

I have to verify data parameters, check what information and what quality it provides. Then I have to perform explanatory data analysis altogether with statistical inference testing.
Once data is cleaned and ready for modeling I'll have to explore different classification models which can help with prediction whether customers stay with the bank or leave it.

Deliverables

A GitHub repo including:

- Datasets
- Jupyter Notebooks for each step of the analysis and modeling process
- A slide deck
- A project report