

NAME: Yogya Sharma (ys5250@nyu.edu)

*Henrietta, a Data Science Product Manager, has just joined TelCo, one of the largest telecommunication firms. Telco is having a major problem with churn in their wireless business. In the mid-Atlantic region, 20% of cell-phone customers leave when their contracts expire, and it is getting increasingly difficult to acquire new customers. They call her in to help understand the problem and devise a solution. Marketing has designed a special retention offer.*

***Specifically, your task is to help Henrietta devise a precise, step-by-step plan for how the analyst/tech team should use TelCo's vast data resource to decide which customers to target with the special retention offer prior to the expiration of their contracts. Be specific as to what data to use and how to use them, and specifically how the team should decide on the set of customers to target to best reduce churn for a particular incentive budget. Use your better judgment as to what data Telco would have.***

***What's the plan?***

## Objective:

To reduce churn by targeting customers who are most likely to leave when their contracts expire, with a retention plan.

## Approach towards the solution:

By implementing a machine learning workflow with steps from collecting data, analyzing data, to generating a predictive model and validating it by measuring business impact in real time.

Steps:

1. The team should collect all the relevant data.

Attributes to be focused on:

- Demographic data: age, location, yearly income, profession.
- Contract/Usage data: average call duration, average internet data usage, type of contract with the TelCo.

Age, location, income, and profession can give an insight into the affordability, and potential usage of the plan that a customer takes up. Demographics generally tend to show the buying/selling patterns of the users in the area which can help the team to reduce the statistical bias in the data like reducing class imbalance and addressing any potential data drift.

The contract details and usage patterns can be used to model any new plans and then target users with a retention policy, giving a better spread in the spectrum of affordability.

2. Now that we have a clean and balanced dataset, we need to analyze it. So, the team should start focusing on identifying any trends and patterns in the data which could give an insight into the reasons of why customers leave.

Two things that the team needs to focus on here are 'feature engineering' and 'feature importance'. This will be a highly important step as this will help the team to know exactly how much each variable is affecting the reason for the customers to stay or leave. And on basis of this and the other exploratory analysis, some new features can be engineered to benefit the quality of the data. This will reduce the redundancy in the analysis and help expedite the process.

3. After the data has been analyzed, the team needs to create a predictive model and train it on the data that has been generated after the previous step. It should be done by splitting the data into train, test, and validation sets. Validation data helps to increase the accuracy of the model as it is trained and before it is tested on the test data.

Different metrics such as accuracy, recall, F1 score etc. can be used to measure the stability of the model.

4. Now the prediction from this model can be used to find the exact group of people to be targeted with the retention campaign.
5. As the telecom company will have the data flowing in real time with current users and their retention, the best way to keep a check on the campaign will be to generate the predictions in real time as well. The best way to do it will be to deploy the generated trained model as an instance on the cloud and keep validating its accuracy as the campaign proceeds to keep track of the campaigns impact and checking if the model needs to be trained more to get better predictions.

## **Conclusion:**

These steps will help Henrietta and team to get insight and then create impact with the campaign with constant updates as the model is being trained and validated on real time data, thus constantly improving its accuracy, and keeping up with any new and emerging trends.