A paper with writing on it

Description automatically generated

**Project**: Perform an Exploratory Data Analysis on customer churn using a Telco dataset.

Thus far we have examined the following:

|  |  |
| --- | --- |
| **Data Set Description** | **Link** |
| Ecommerce Customer Churn Analysis and Prediction | <https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction/data> |
| Telco Customer churns | <https://www.kaggle.com/datasets/blastchar/telco-customer-churn> |
| Bank customer churn | <https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset/data> |
| Grid Dynamics | Repo (includes data): <https://github.com/griddynamics/rnd-gcp-starter-kits/tree/main/churn-prevention-vertex-ai>  Paper: <https://blog.griddynamics.com/churn-analytics-in-the-technology-and-telecom-industries-using-google-vertex-ai-a-reference-notebook/> |
|  |  |

After looking at the data sets for the four sets above, we are leaning to the last one, “Grid Dynamics” in the table above. It has the most amount of data columns to evaluate as well as also having more continuous data, which would make for better Exploratory Data Analysis.

**Project Team:** David Gerhart, Cody Kushing, Doug Francis

**Project “Catechism”**

1. What are we trying to do.
   1. Perform Exploratory Data analysis for a customer chur analysis.
   2. After Exploratory Data Analysis, if time permits, perform an analysis of churn to build a predictive model. This should enable us to probabilistically predict who is most likely to churn. In turn, this can be used to build a strategy of how to retain as many customers as possible.
2. How is it done today. What are the limits to the current practice.
   1. We speculate that for medium and large size businesses, they would already be doing ML techniques as well as Exploratory Data Analysis. We further speculate that companies are probably using other non-AI data mining techniques.
   2. Currently the dataset we are looking at does not have any mitigation data in it. That is, if they employed different retention programs, were they effective and what demographics influence the success of the programs.
3. What is new in our approach? Why do you think it will be successful?
   1. We hope to expand on our exploratory data analysis and use machine learning to find patterns in the data that we were not able to identify.
   2. We believe a combination of manual data analysis and Machine learning will reveal new patterns in the churn data. There are members of our team who have had a long-standing interest in customer churn and have created retention programs. We are very interested in the insights that AI will provide.
4. “Who cares”? If you are successful, what difference does it make.
   1. Any business that depends on recurring revenue from its customers’ needs to minimize churn. Insights gained through churn analysis and the success of various retention techniques may apply to multiple industries.
   2. Being able to probabilistically know which customers have the highest probability of churning would allow a company to better target their limited customer retention spending.
5. What are the risks?
   1. If the data collected does not have the key variables, then the analysis would not yield a good predictive model.
   2. Since no company will in truth be using our analysis, there is little risk.
6. How much will it cost?
   1. There are no direct project costs since this is a student exercise. The cost will just be in terms of time. No other expenses will be incurred.
7. How long will it take?
   1. The project must be completed by February 13th.
8. What are the mid-term and final exams to check for success?
   1. Initial data cleaning and exploration
   2. Creation of 6 or 8 exploratory graphs
   3. Drafting of conclusions and follow up analysis.
   4. Creation of

\*\* Ecommerce Customer Churn Analysis and Prediction

<https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction/data>

zzz

\*\* Telco Customer churns

<https://www.kaggle.com/datasets/blastchar/telco-customer-churn> <- the data on kaggle.

Someone's analysis they have completed.

<https://365datascience.com/tutorials/python-tutorials/how-to-build-a-customer-churn-prediction-model-in-python/#5>

Bank Customer Churn Dataset

<https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset/data>

Grid Dynamics:

Paper: <https://blog.griddynamics.com/churn-analytics-in-the-technology-and-telecom-industries-using-google-vertex-ai-a-reference-notebook/>

Repo: <https://github.com/griddynamics/rnd-gcp-starter-kits/tree/main/churn-prevention-vertex-ai>

Notebook: https://github.com/griddynamics/rnd-gcp-starter-kits/blob/main/churn-prevention-vertex-ai/notebooks/Churn\_Predictions\_for\_Telco\_with\_Vertex\_AI.ipynb