Assessing the Risk of Telco Customer Churn using Predictive Algorithms and Models

Background & Introduction

- Focusing on customer churn analysis in the pivotal Telco industry
  - Customers have high expectations for quality telcom services at all times

Problem: Telecommunications companies are losing their customers

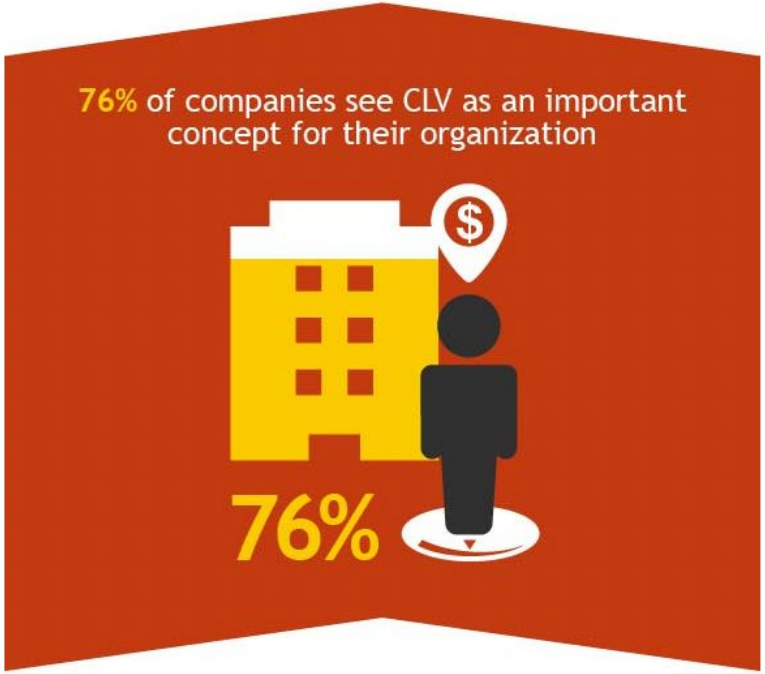
- Churn is often used as a company's key operational indicators as customer retention is crucial to companies
- A loyal customer is worth far more than a new one in the long run

Thesis: Our group aims to produce a customer churn model that can accurately predict when a customer will churn, reducing profit loss for the company

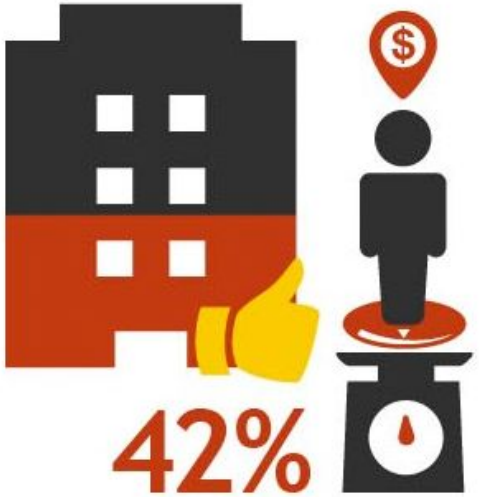
How to calculate Churn Rate?

- With more prediction APIs, more businesses are able to gain access to this modelling tool

What about customers?



Only 42% of companies are able to measure Customer Lifetime Value (CLV) accurately.



Motivation

erv… …er churn

ons… …vior and

wil… …client

g m… …ar

ervi…

Literature Review

A Customer Churn Prediction Model in Telecom Industry Using Boosting *SCI 2014《IEEE Transactions on Industrial Informatics》*

Key Learning points

- Boosting can be used to further increase the accuracy of a simple classification model

A Big Data Clustering Algorithm for Mitigating the Risk of Customer Churn
SCI 2016 IEEE Transactions on Industrial Informatics

Key Learning Points

- Classification can be used (predictive analysis) instead of clustering as we have a target column in our data

Evaluation of customer behavior with temporal centrality metrics for churn prediction of prepaid contracts
SCI 2020《EXPERT SYSTEMS WITH APPLICATIONS》

Key Learning points

- Social network and finance data are both extremely important
- Customer attributes are crucial in arriving at a more accurate analysis

Study on the Prediction of Imbalanced Bank Customer Churn Based on Generative Adversarial Network
SCI 2020《Journal of Physics Conference》

Key learning points

- *SMOTE* can be used for unbalanced data
- Combination of 2 other methods to arrive at the best solution

Choosing the right dataset

Initially, we picked Dataset 1 from Kaggle - **Telcom Customer Churn.**
Limited at 7043 records and imbalance at 26.6%

Dataset 2 WA_Fn-UseC_-Telco-Customer-Churn.csv From : IBM Business Analytics Community	Similar in number of records however, it has a higher degree of missing data as compared to dataset 1 at 0.35%.

Dataset 3 Telcom User Churn Dataset From : DataFountain Platform	Unreliable as it is provided by a personal user and has lower number of records.

usually confidential and **possible to use oversampling** .

Dataset: Telco Customer Churn

7043 rows (customers), 21 columns (features, customer's attributes)

Description	Column Name
Customers who left within the last month	Churn [Target]
Services that each customer has signed up for	Phone, Multiple Lines, Internet, Online Security, Online Backup, Device Protection, Tech Support and Streaming TV and Movies
Customer account information	How long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
Demographic info about customers	gender, age range, and if they have partners and dependents

https://www.kaggle.com/blastchar/telco-customer-churn/version/1

Literature Review Part 2

What has been done?	What are we going to do?
Using and comparing **1-3 models** for comparison with **machine learning algorithm**	Construct more models using a bigger variety of algorithms including deep learning methods and use voting classifier afterwards.
Strong focus on machine learning **prediction modelling**	Do more in-depth EDA and visualisation for each and every individual attribute against the target column - **Churn or Not Churn.**
Basic methods for dealing with the imbalanced data	Compare/use with the different oversampling methods available to improve the data set
Focus on good **prediction model and performance**	With both business indicator analysis and improvement of prediction modeling, we can have more in-depth suggestions for the ‘company's operation

Tools and Resources

EDA

- matplotlib
- seaborn
- plotly
  - offline
  - graph_objs
  - figure_factory

Model Evaluation

- sklearn.metrics

Data Preparation

- numpy
- pandas
- scikit-learn
  - preprocessing
  - decomposition
- imblearn

  SMOTE (SVC SMOTE, KMeans SMOTE)

Modeling Analysis

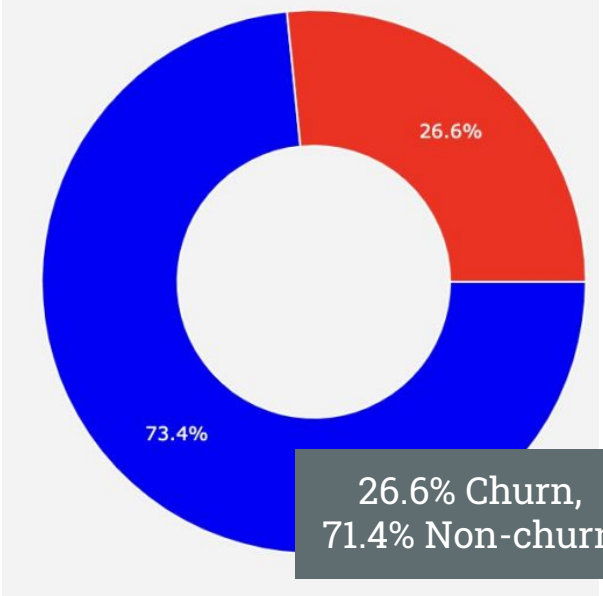
- sklearn.model_selection
- sklearn.neighbors
- sklearn.svm
- sklearn.ensemble

adaboost,gbdt,extratree...

- sklearn.tree
- xgboost
- lightbgm
- keras

Preliminary Results & Evaluation

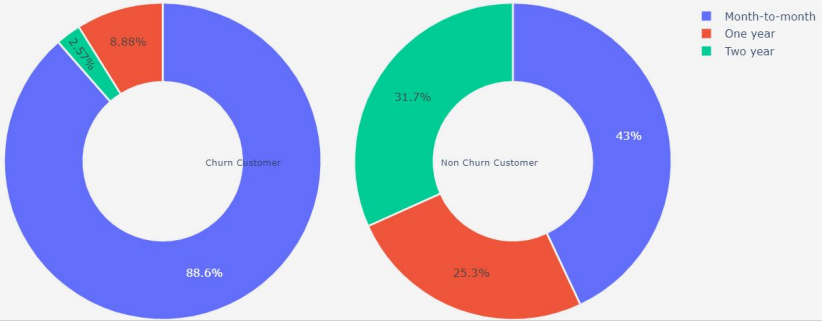
churn vs non churn customers



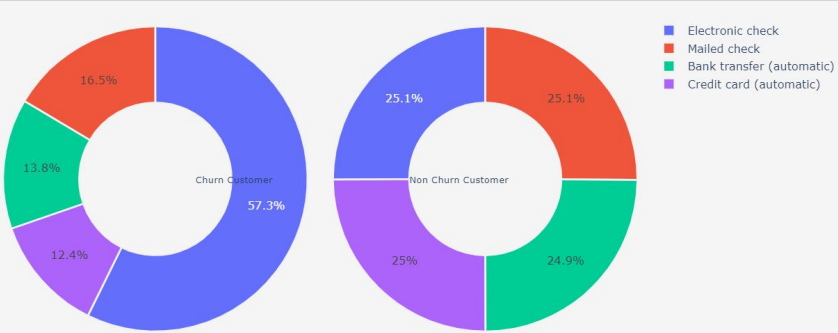
26.6% Churn,
71.4% Non-churn

Afterwards, we did EDA on the data by visualising the target class proportion of each individual attribute.

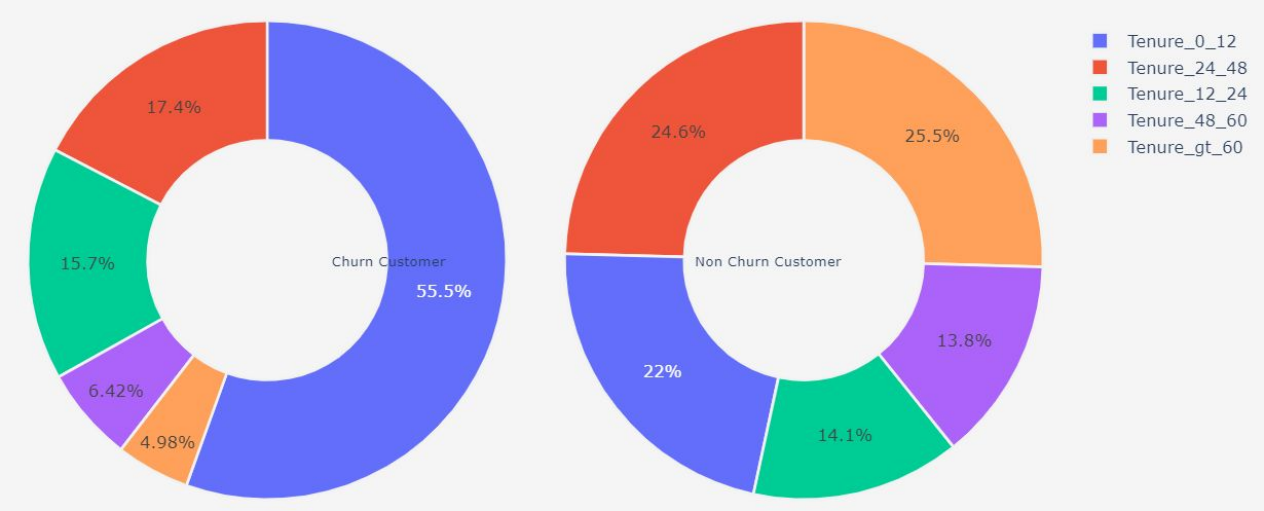
Divided using the type of the length of **Contract**

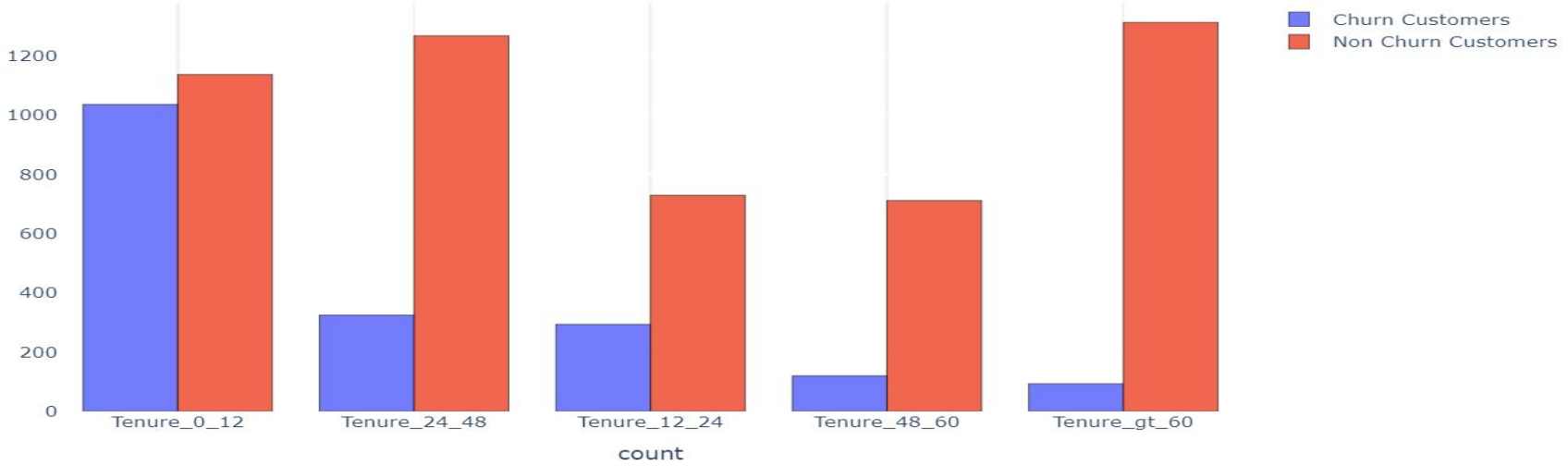


Divided using the type of **Payment Method**



Focusing on the the relationship between the **Tenure variable and the target class - Customer Churn**

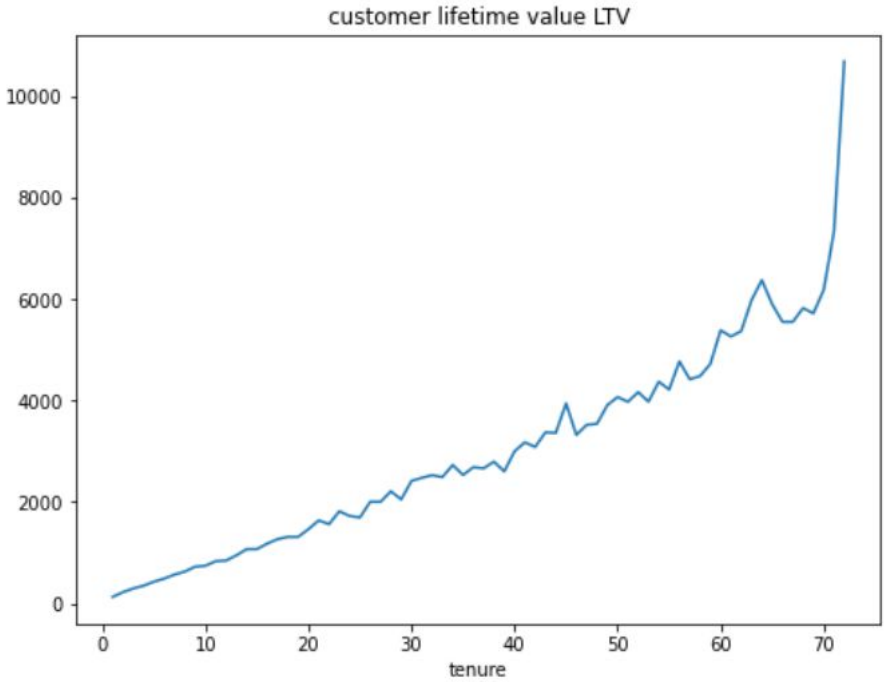




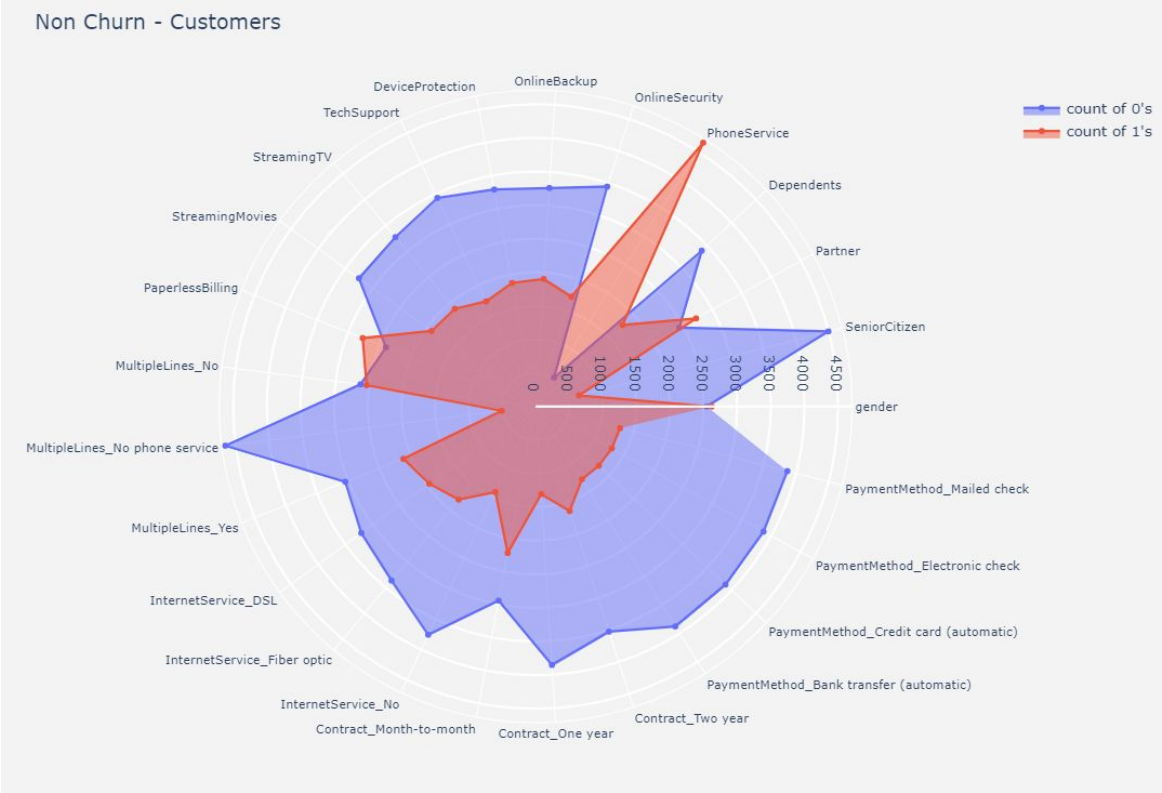
In the business perspective…

LTV =
Total historical payment (total charges) + Remaining charges

- The **longer** the customers stays with the Telco, the **higher** the customer lifetime value.
- Sharp peak at **72 months!**
- Therefore, it is very important to be able to retain customers longer.



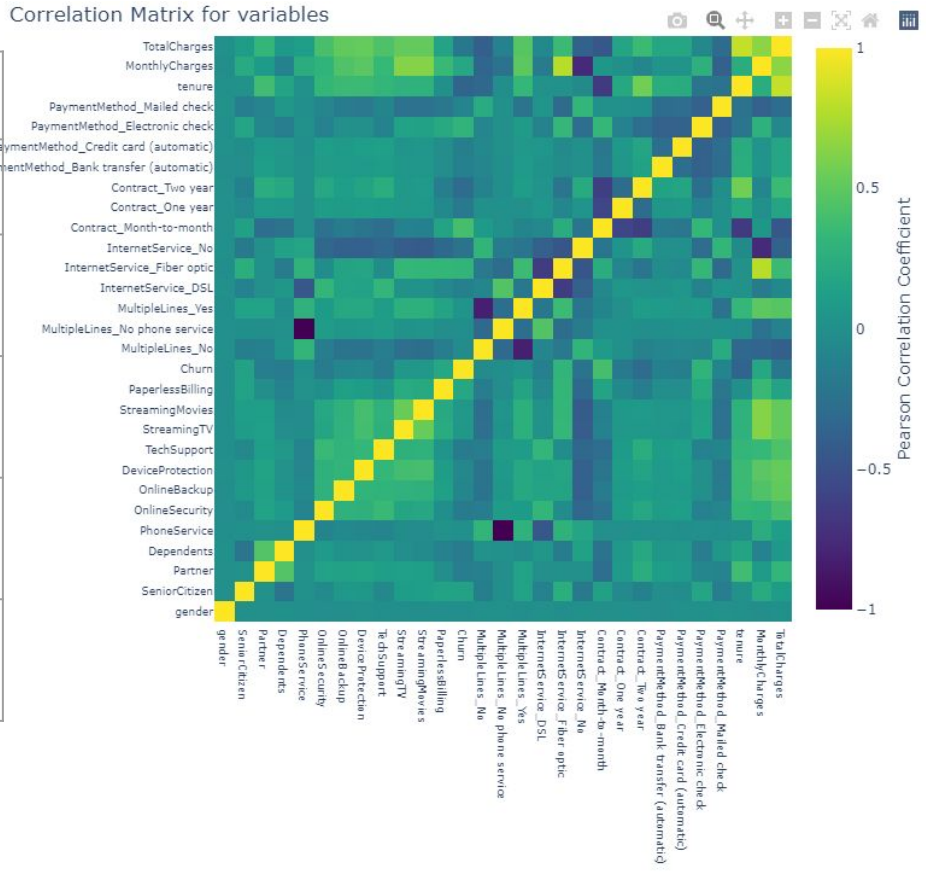
Constructing a Customer Persona



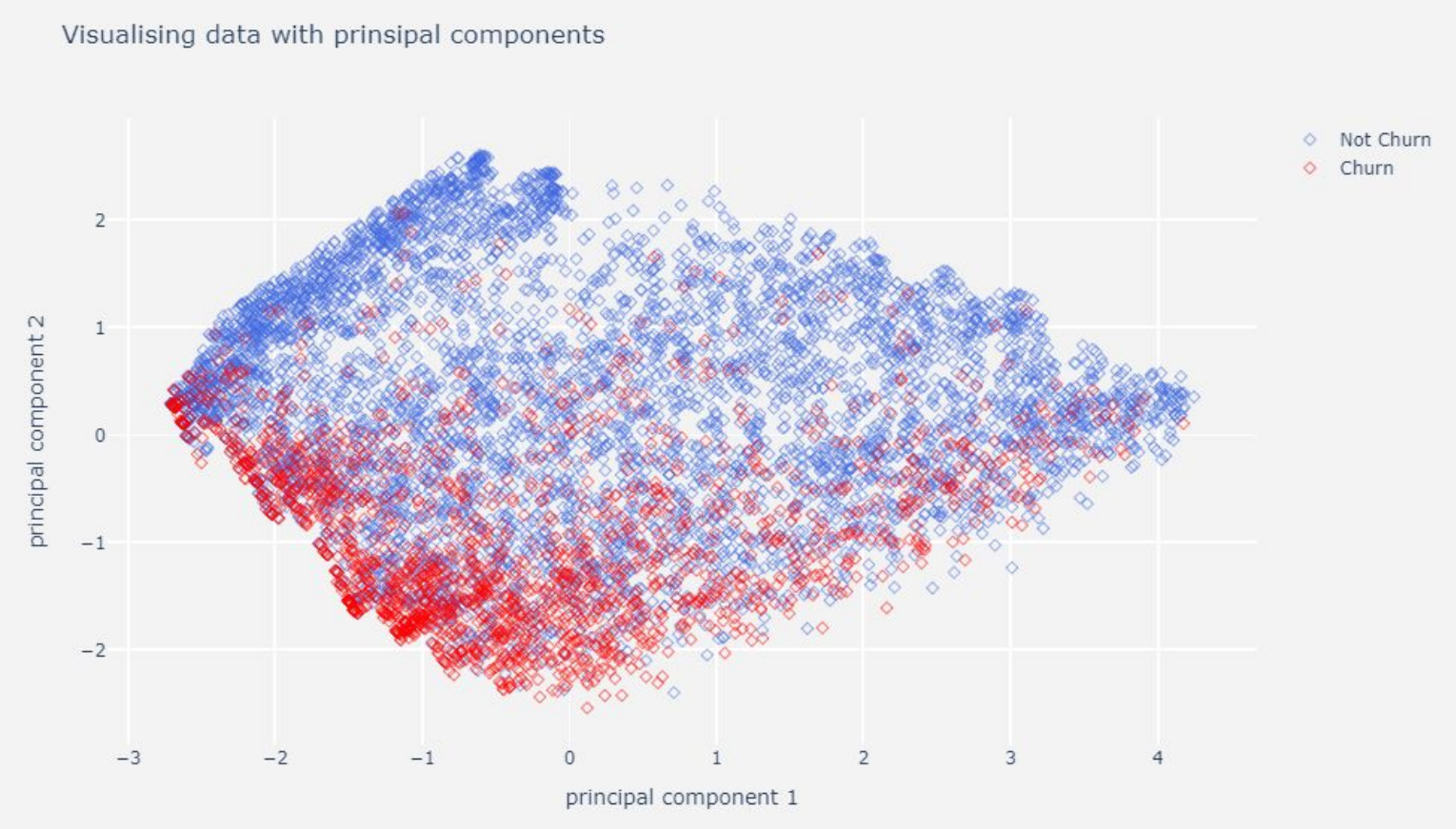
Correlation and PCA - feature engineering

Noteworthy observations:

Variable 1	Variable 2	Correlation
Totalcharges	Tenure	0.826
Internet Services	Tenure	0.78
Streaming Movies	Monthly Charges	0.62
Internet Services	Monthly Charges	-0.72
Monthly Contract	Tenure	-0.65



Correlation and PCA - feature engineering



Going Forward…

1) Dealing with Unbalanced Data

- SMOTE, SVC SMOTE, KNN SMOTE
  - Oversampling Techniques
  - Synthetic samples for minority class

2) Single Classification Model

- KNN
- Random Forest
- Deep Learning (ANN)
- XGBoost
- LightGBM

3) Comparing Models

- Recall Rate
- Precision Rate
- Accuracy Rate
- AUC Score
- Voting Classifier

Project Milestones and Tentative Schedule

Project Start Date	1/19/2022 (Wednesday)	Display Week	9
Project Lead	Zijiao		

TASK	LEAD	START	END	DAYS	% DONE
Final Proposal Presentation				-	
Method	Zijiao	Thu 3/24/22	Mon 4/04/22	12	0%
Results and Discussion	Glendys	Thu 3/24/22	Mon 4/04/22	12	0%
Conclusion and Future Work	Aiwei	Thu 3/24/22	Mon 4/04/22	12	0%
Reference	Jia Yi	Thu 3/24/22	Mon 4/04/22	12	0%
Beautify slides	Jolene	Mon 4/04/22	Tue 4/05/22	2	0%
Last check and clean up	Jolene	Tue 4/05/22	Wed 4/06/22	2	0%
Final Proposal Slides Submission (7.30am)	Jolene	Wed 4/06/22	Wed 4/06/22	1	0%
Final Report				-	
Method	Zijiao	Thu 3/17/22	Wed 4/06/22	21	0%
Results and Discussion	Glendys	Thu 3/17/22	Wed 4/06/22	21	0%
Conclusion and Future Work	Aiwei	Thu 3/17/22	Wed 4/06/22	21	0%
Reference	Jia Yi	Thu 3/17/22	Wed 4/06/22	21	0%
Last check, clean up and formatting	Jia Yi	Wed 4/06/22	Fri 4/08/22	3	0%
Final Report Submission (3.30pm)	Jolene	Fri 4/08/22	Fri 4/08/22	1	0%

References

Literature Review:

Lu, N., Lin, H., Lu, J., & Zhang, G. (2012). A customer churn prediction model in telecom industry using boosting. *IEEE Transactions on Industrial Informatics*, *10*(2), 1659–1665.

Bi, W., Cai, M., Liu, M., & Li, G. (2016). A big data clustering algorithm for mitigating the risk of customer churn. *IEEE Transactions on Industrial Informatics*, *12*(3), 1270–1281.

Li, B., & Xie, J. (2020, October). Study on the Prediction of Imbalanced Bank Customer Churn Based on Generative Adversarial Network. In *Journal of Physics: Conference Series* (Vol. 1624, No. 3, p. 032054). IOP Publishing.

Calzada-Infante, L., Óskarsdóttir, M., & Baesens, B. (2020). Evaluation of customer behavior with temporal centrality metrics for churn prediction of prepaid contracts. *Expert Systems with Applications*, *160*, 113553.

Others:

Campbell, P. (2021). *Customer Churn Models: Lowering CAC, Maximizing Retention.* ProfitWell. Retrieved January 22, 2022, from

https://www.profitwell.com/customer-churn/models

Dorard, L., & Patel, N. (2021). *How to Improve Your Subscription Based Business by Predicting Churn.* Neil Patel. Retrieved January 22, 2022, from

https://neilpatel.com/blog/improve-by-predicting-churn/

KDnuggets. (2017, March 1). *What is Customer Churn Modeling? Why is it valuable?* KDnuggets. Retrieved January 22, 2022, from

https://www.kdnuggets.com/2017/03/datascience-customer-churn-modeling.html

Clark, B. (2020, March 16). How to Improve Customer Retention in the Telecom Industry – Acquire. Acquire; Acquire.

https://acquire.io/blog/improve-customer-retention-in-telecom-industry/

Saleh, K., & Shukairy, A. (2021, November 11). *Customer Acquisition Vs.Retention Costs [Infographic].* Invesp. Retrieved February 22, 2022, from

THANK YOU!

Any questions?