Telecom Customer Churn Analysis & Prediction

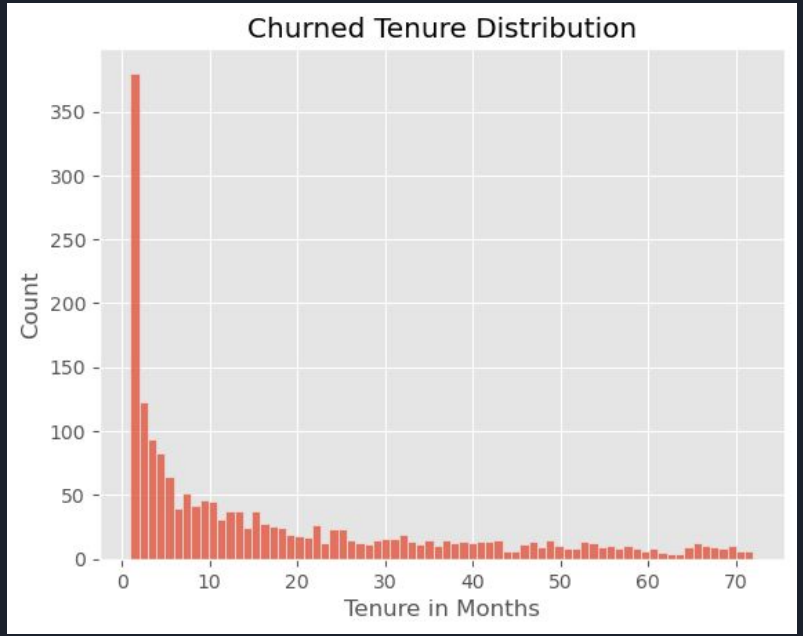
By Vince Coletti

The Problem

How do we reduce customer churn to increase revenue?

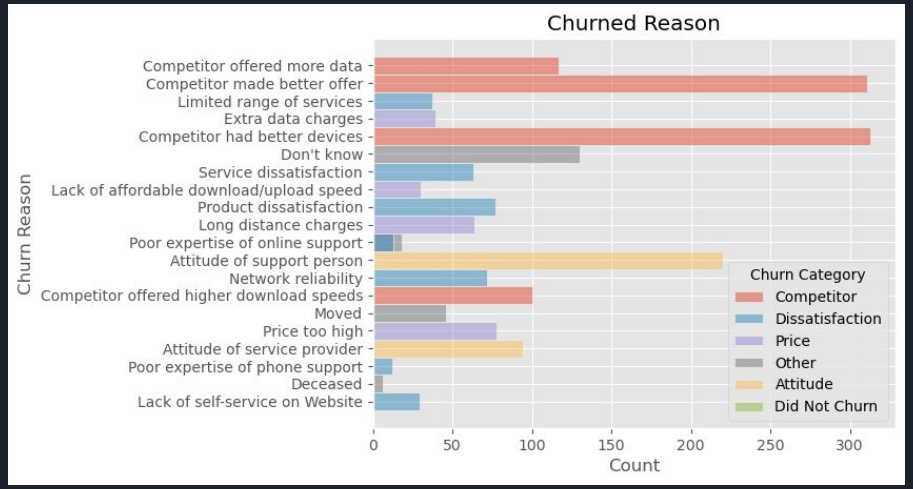
The Solutions

1) Reinforce customer's purchasing decision as soon as they sign up.



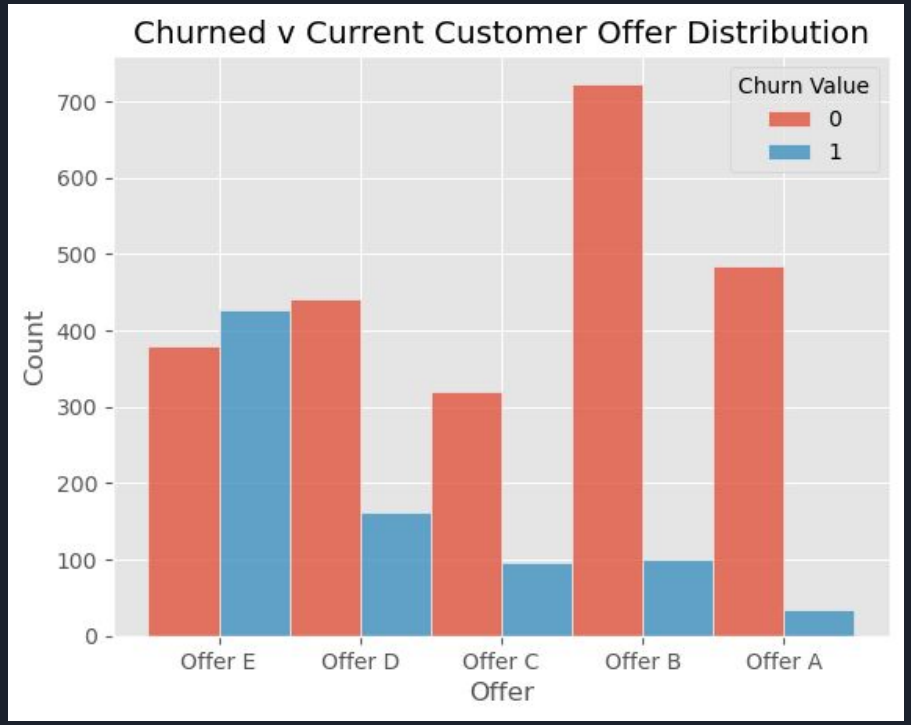
The Solutions

2) Improve some of our offerings, improve pricing, and improve customer support training.



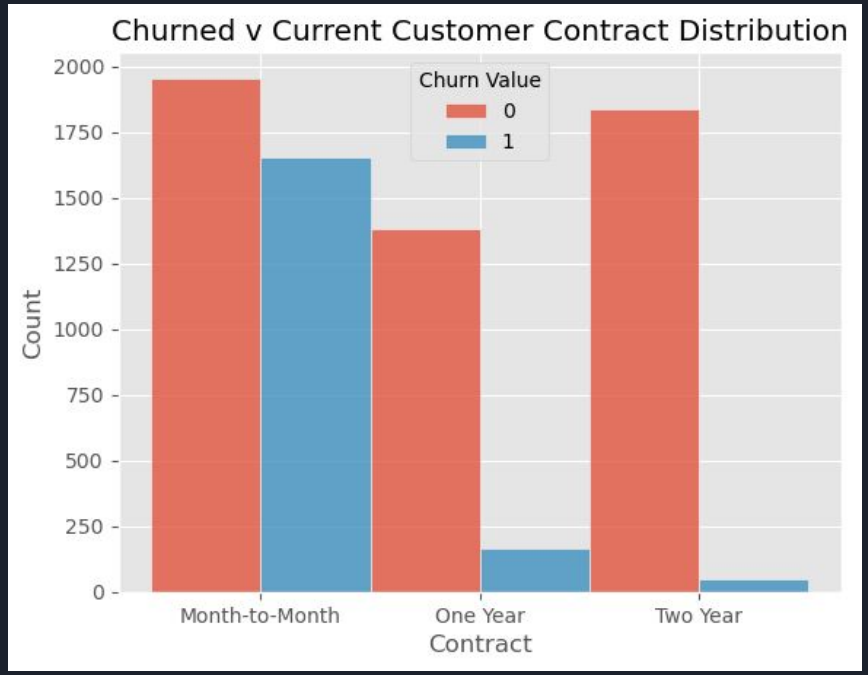
The Solutions

3) Phase out Offer E immediately.



The Solutions

4) Incentivize customers to sign up for 1-2 year contracts instead of month to month.



The Data

	Satisfaction Score	Churn Value	Churn Score	CLTV	Churn Category	Churn Reason	Employee Related Cancellation	Product/Service Related Cancellation	Competitor Cancellation	Price Cancellation	Referred a Friend	Number of Referrals	Tenure in Months
0	3	1	91	5433	Competitor	Competitor offered more data	0	0	1	0	No	0	1
1	3	1	69	5302	Competitor	Competitor made better offer	0	0	1	0	Yes	1	8
2	2	1	81	3179	Competitor	Competitor made better offer	0	0	1	0	No	0	18
3	2	1	88	5337	Dissatisfaction	Limited range of services	0	1	0	0	Yes	1	25
4	2	1	67	2793	Price	Extra data charges	0	0	0	1	Yes	1	37
5	1	1	95	4638	Competitor	Competitor had better devices	0	0	1	0	No	0	27
6	2	1	76	3964	Other	Don't know	0	0	0	0	Yes	1	1
7	1	1	91	5444	Dissatisfaction	Service dissatisfaction	1	1	0	0	Yes	6	58
8	2	1	91	5717	Dissatisfaction	Limited range of services	0	1	0	0	No	0	15
9	2	1	81	4419	Price	Lack of affordable download/upload speed	0	0	0	1	No	0	7

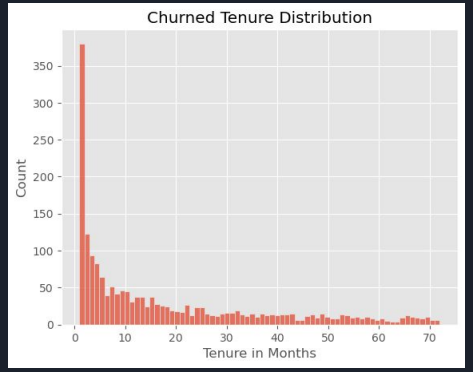
Data Source:
https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco-customer-churn-1113

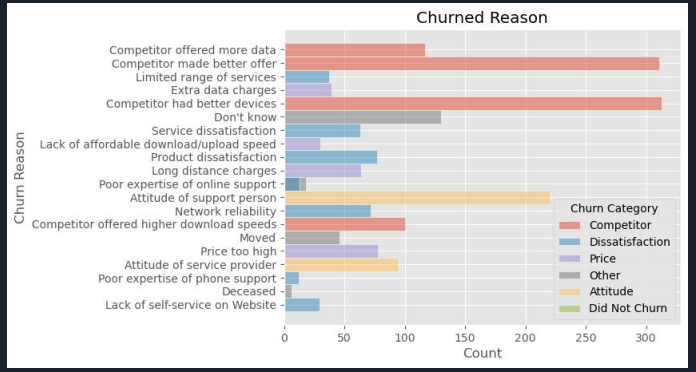
Data Cleaning

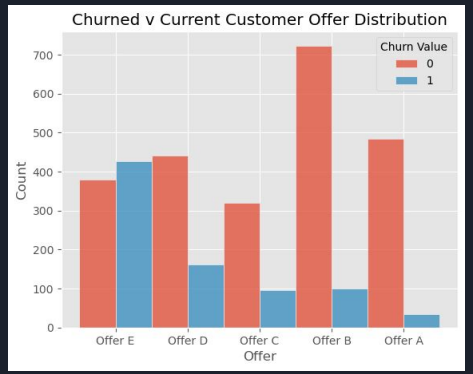
Original dataset came in 4 different Excel files, had 7,043 rows, and 62 columns. Most of the data was categorical, and there were missing values.

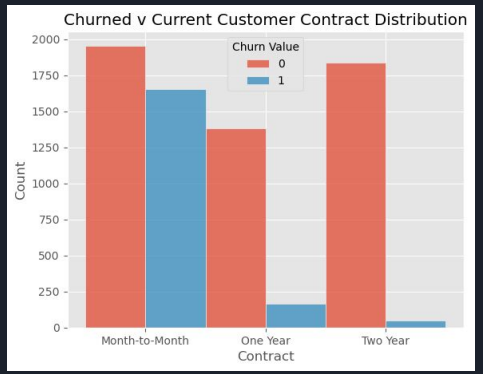
- Joined tables
- Missing values were handled appropriately
- Dropped columns that provided no further information.
- 'Zip Codes' values were converted to strings to ensure that any potential leading 0's were not dropped.
- Our target variable is naturally ‘Churn Value’ as that is what we are trying to predict.

Exploratory Data Analysis





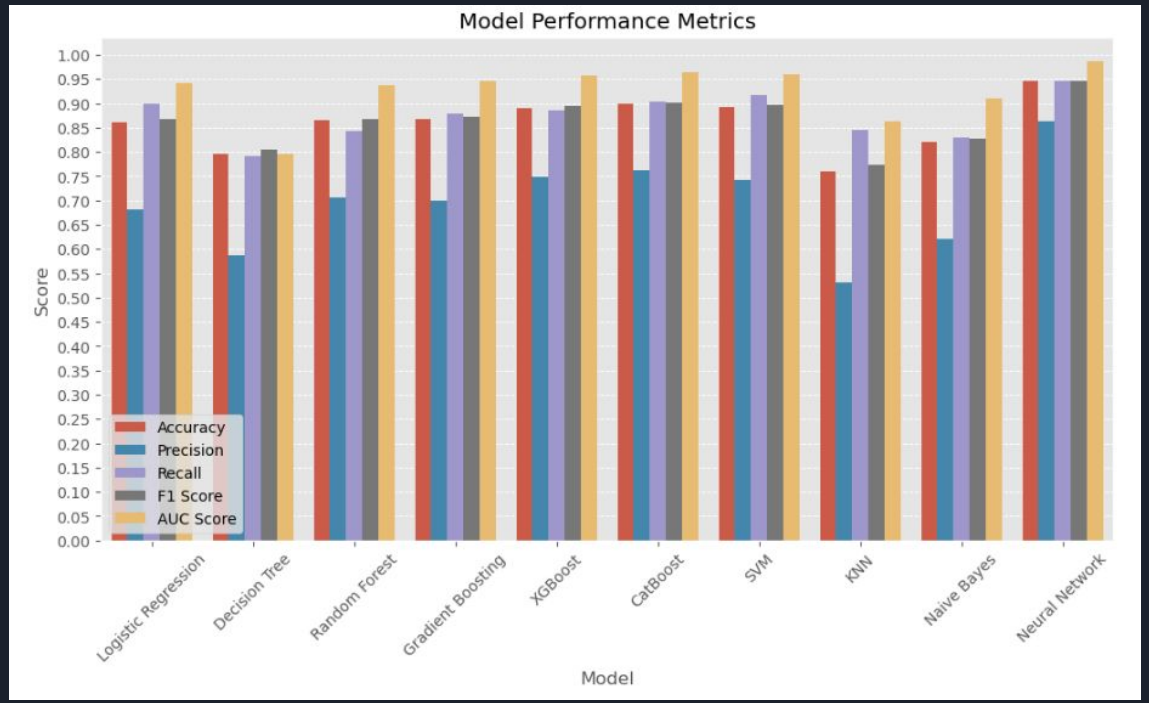




Data Preprocessing

- One hot encoding
- Drop churn reason & churn category features
- Split the data
- Scale the data
- Principal Component Analysis
- Undersampling

Initial Modeling



Model Hyperparameter Tuning

Logistic Regression

- Accuracy: 86%
- Recall: 90%

Best Cross-Validation Score: 0.8846

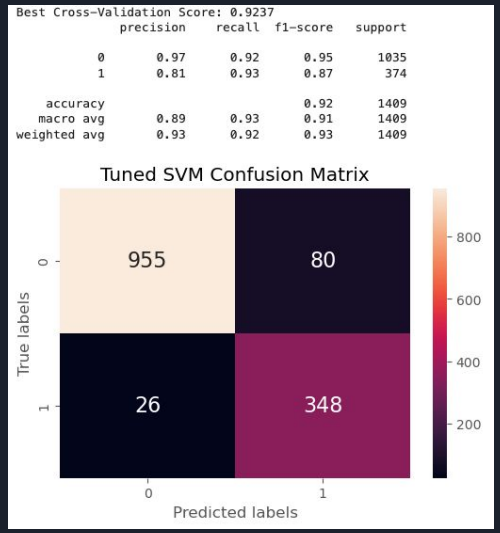
	precision	recall	f1-score	support
0	0.96	0.85	0.90	1035
1	0.68	0.90	0.78	374
accuracy			0.86	1409
macro avg	0.82	0.87	0.84	1409
weighted avg	0.89	0.86	0.87	1409

SVM

- Accuracy: 92%
- Recall: 93%

Best Cross-Validation Score: 0.9237

	precision	recall	f1-score	support
0	0.97	0.92	0.95	1035
1	0.81	0.93	0.87	374
accuracy			0.92	1409
macro avg	0.89	0.93	0.91	1409
weighted avg	0.93	0.92	0.93	1409

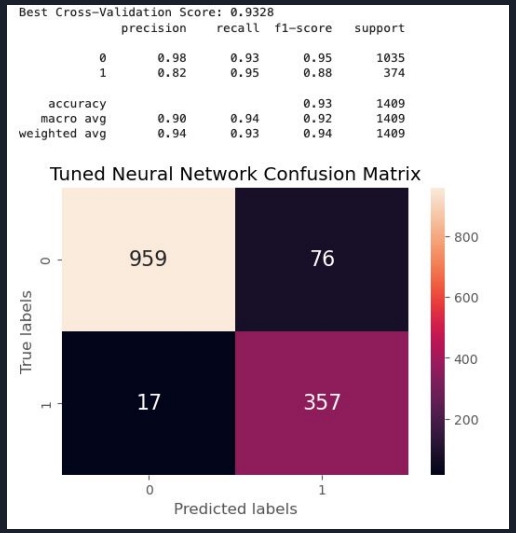


Neural Network

- Accuracy: 93%
- Recall: 95%

Best Cross-Validation Score: 0.9328

	precision	recall	f1-score	support
0	0.98	0.93	0.95	1035
1	0.82	0.95	0.88	374
accuracy			0.93	1409
macro avg	0.90	0.94	0.92	1409
weighted avg	0.94	0.93	0.94	1409

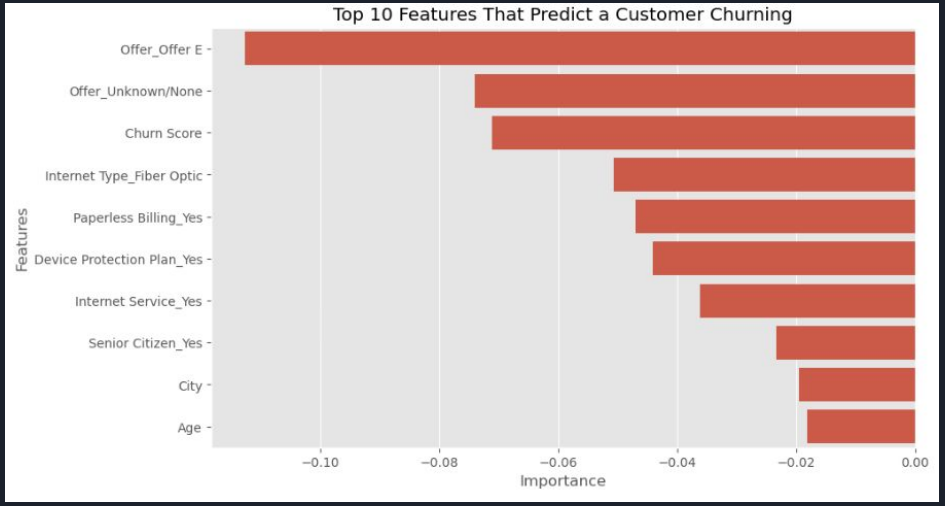


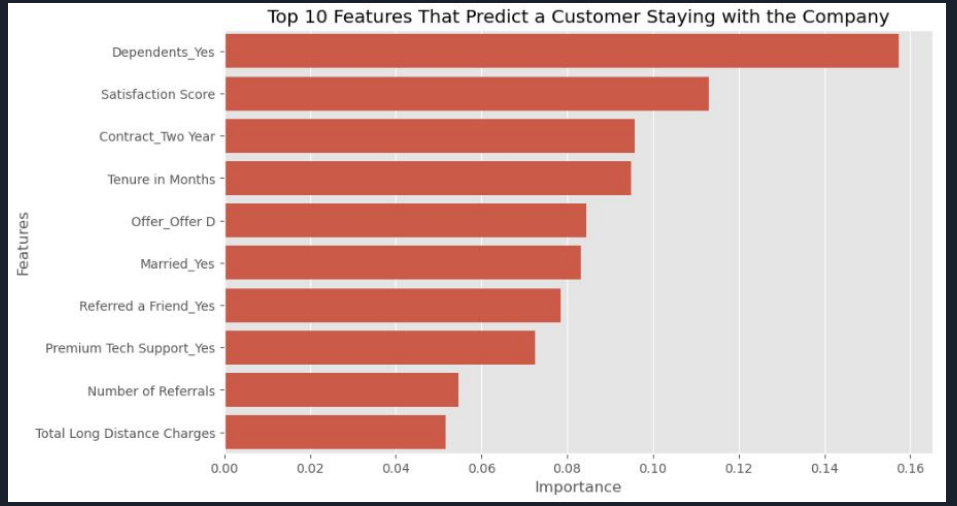
Neural Network Prediction Scoring

93% Accuracy

95% Recall

What makes a churned customer vs. a loyal one?





Future Research

- After making recommended changes, collect more data to confirm if churn decreases and how the models change.
- Collect more survey data from *loyal* customers to see why they say they remain with the company.
- For competitor related churn reasons, it would be useful to know what exactly the competitor offered that was better than our offering and/or which competitor devices the customer switched because of.

Questions?