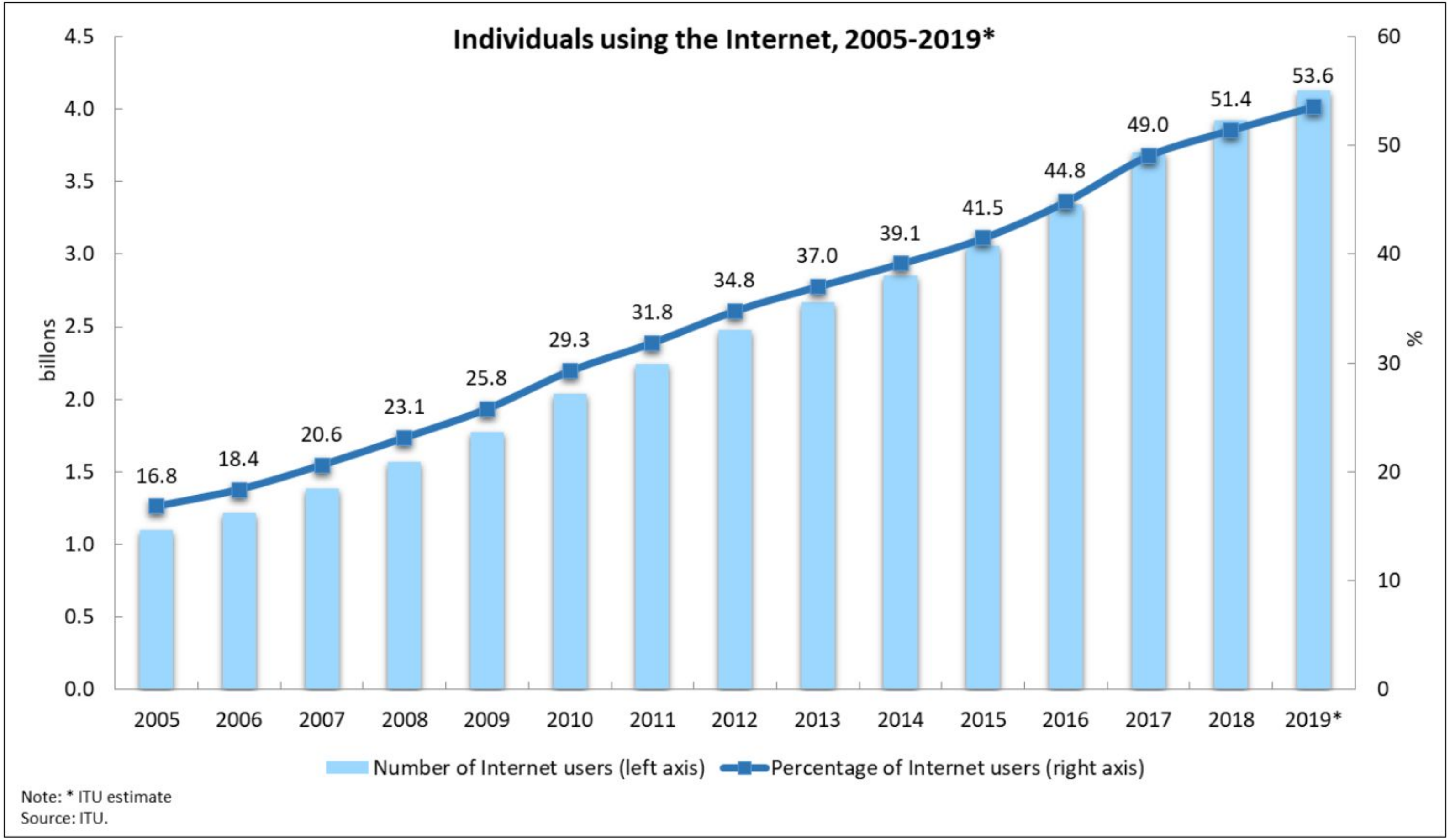
Predicting Telecom Churn Rates

Presentation Outline

- Explain research / importance of research
- Dataset explanation
- Data analysis
- Final thoughts

Technology is Always Improving


Individuals using the Internet, 2005-2019*
billons
%
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
60
50
40
30
20
10
0
16.8
18.4
20.6
23.1
25.8
29.3
31.8
34.8
37.0
39.1
41.5
44.8
49.0
51.4
53.6
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019*
Number of Internet users (left axis)
Percentage of Internet users (right axis)
Note: * ITU estimate
Source: ITU.

Research Information

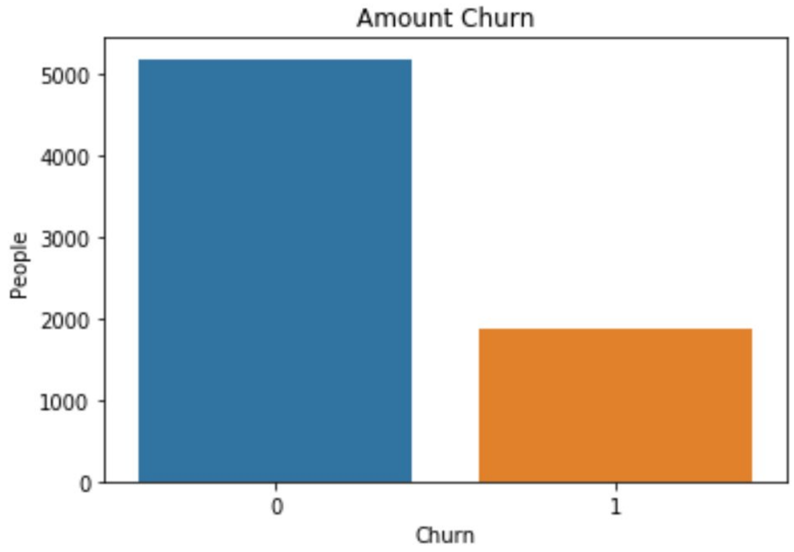
- Telecom Churn rates are described as the percentage of customers leaving a service provider over a period of time
- Telecommunication industry has many people relying on them for satisfactory services
- Aim to create a classification model that predicts if customers will churn or not

The Data

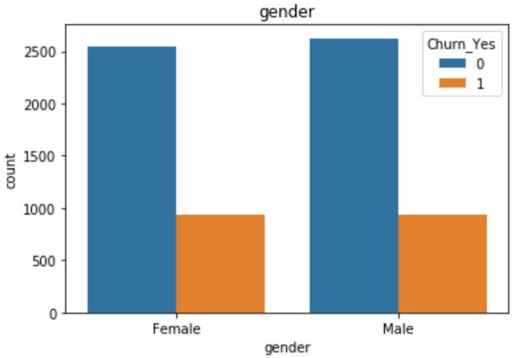
- The dataset was created by Zagarsuren Sukhbaatar and obtained from kaggle
- The information used in the dataset was obtained from IBM Watson Analytics

About the Data

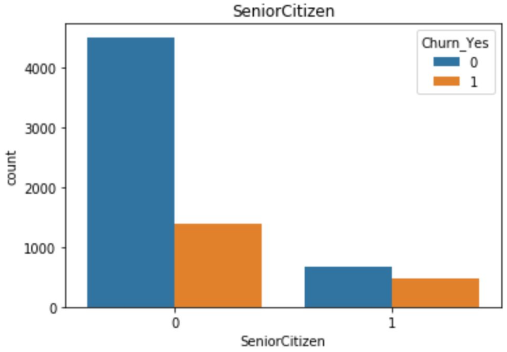
- Data for 7043 total customers
  - 5,174 did not churn
  - 1,869 did churn
- Dataset contained 21 features for each customer (including if customer churned)



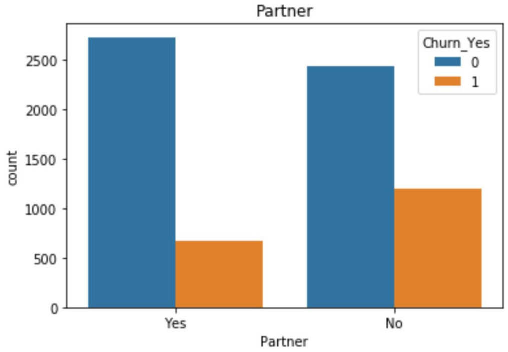
gender	Churn_Yes
Female	0.269209
Male	0.261603



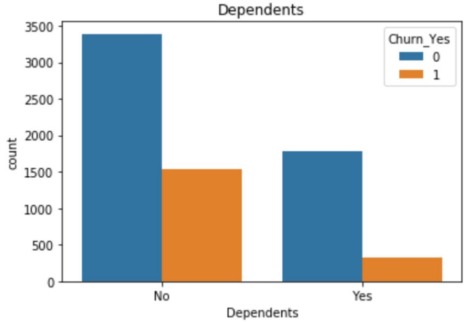
SeniorCitizen	Churn_Yes
0	0.236062
1	0.416813



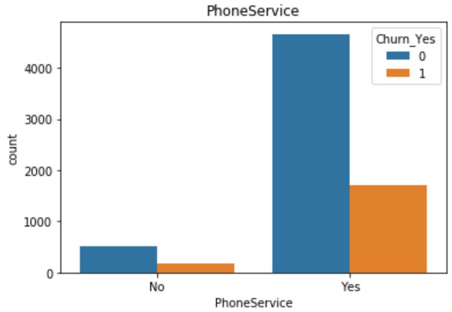
Partner	Churn_Yes
No	0.329580
Yes	0.196649



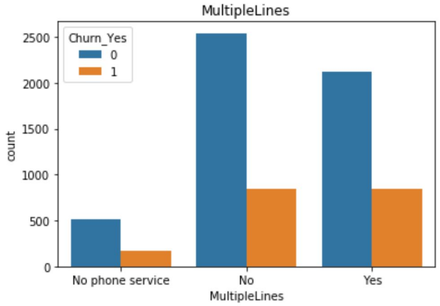
Dependents	Churn_Yes
No	0.312791
Yes	0.154502



PhoneService	Churn_Yes
No	0.249267
Yes	0.267096



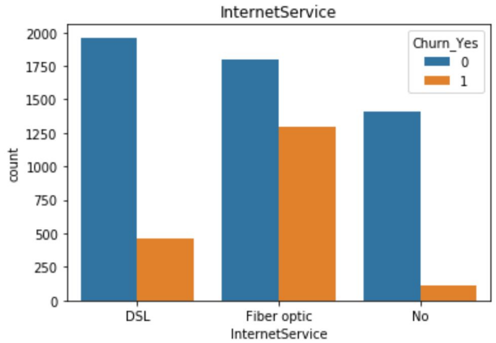
MultipleLines	Churn_Yes
No	0.250442
No phone service	0.249267
Yes	0.286099

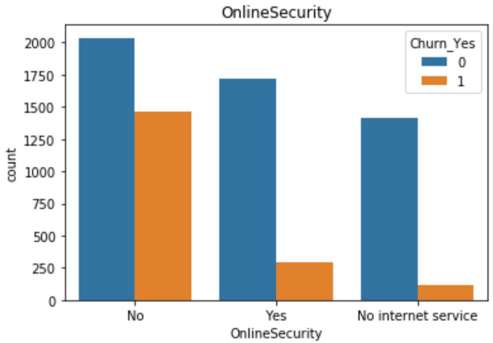


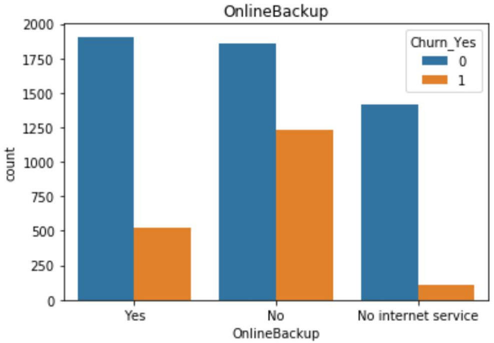
	Churn_Yes
InternetService	
DSL	0.189591
Fiber optic	0.418928
No	0.074050

	Churn_Yes
OnlineSecurity	
No	0.417667
No internet service	0.074050
Yes	0.146112

	Churn_Yes
OnlineBackup	
No	0.399288
No internet service	0.074050
Yes	0.215315



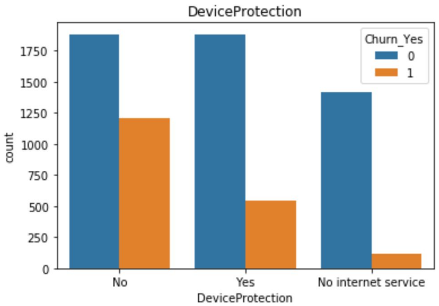


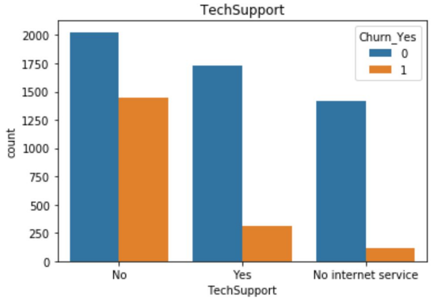


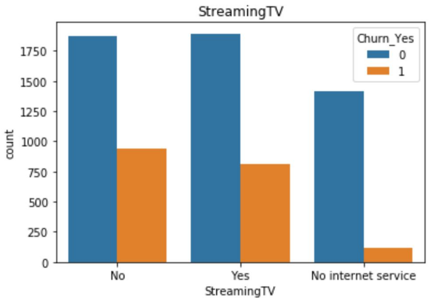
	Churn_Yes
DeviceProtection	
No	0.391276
No internet service	0.074050
Yes	0.225021

	Churn_Yes
TechSupport	
No	0.416355
No internet service	0.074050
Yes	0.151663

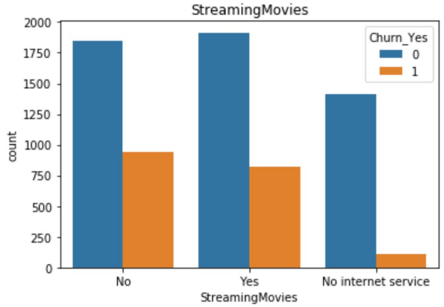
	Churn_Yes
StreamingTV	
No	0.335231
No internet service	0.074050
Yes	0.300702



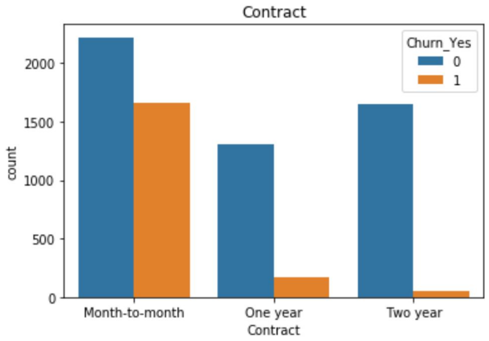




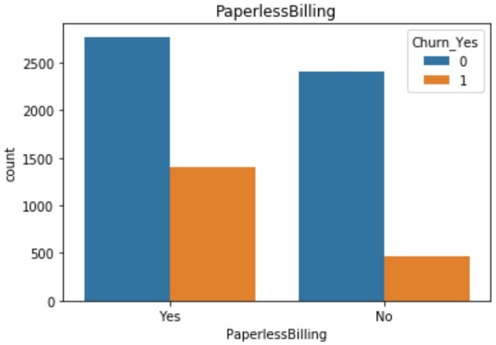
StreamingMovies	Churn_Yes
No	0.336804
No internet service	0.074050
Yes	0.299414



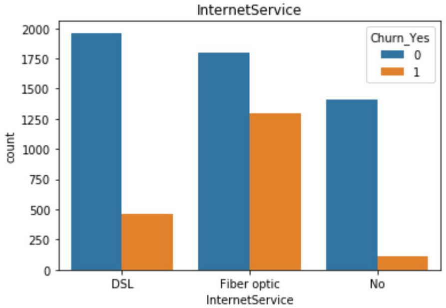
Contract	Churn_Yes
Month-to-month	0.427097
One year	0.112695
Two year	0.028319



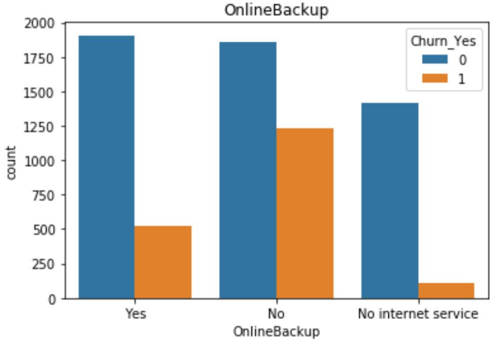
PaperlessBilling	Churn_Yes
No	0.163301
Yes	0.335651



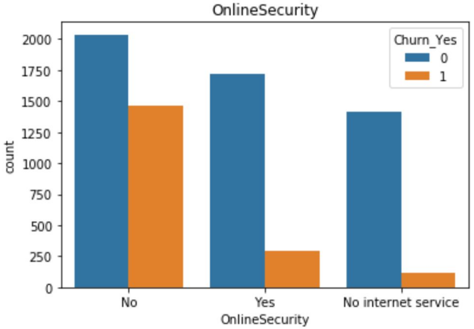
	Churn_Yes
InternetService	
DSL	0.189591
Fiber optic	0.418928
No	0.074050



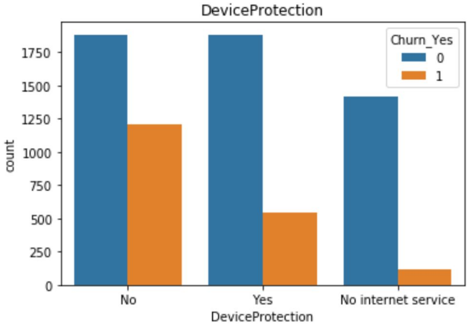
	Churn_Yes
OnlineBackup	
No	0.399288
No internet service	0.074050
Yes	0.215315



	Churn_Yes
OnlineSecurity	
No	0.417667
No internet service	0.074050
Yes	0.146112



	Churn_Yes
DeviceProtection	
No	0.391276
No internet service	0.074050
Yes	0.225021



KNN Model Comparison

K Nearest Neighbor Classifier

	Predicted No Churn	Predicted Churn
Actual No Churn	890	145
Actual Churn	169	205

	precision	recall	f1-score	support
0	0.84	0.86	0.85	1035
1	0.59	0.55	0.57	374
accuracy			0.78	1409
macro avg	0.71	0.70	0.71	1409
weighted avg	0.77	0.78	0.77	1409

K Nearest Neighbor Classifier with K = 10

	Predicted No Churn	Predicted Churn
Actual No Churn	921	114
Actual Churn	165	209

	precision	recall	f1-score	support
0	0.85	0.89	0.87	1035
1	0.65	0.56	0.60	374
accuracy			0.80	1409
macro avg	0.75	0.72	0.73	1409
weighted avg	0.79	0.80	0.80	1409

K Nearest Neighbor Classifier with PCA

	Predicted No Churn	Predicted Churn
Actual No Churn	925	110
Actual Churn	196	178

	precision	recall	f1-score	support
0	0.83	0.89	0.86	1035
1	0.62	0.48	0.54	374
accuracy			0.78	1409
macro avg	0.72	0.68	0.70	1409
weighted avg	0.77	0.78	0.77	1409

SVC Model Comparison

Support Vector Classifier

	Predicted No Churn	Predicted Churn
Actual No Churn	964	71
Actual Churn	188	186

	precision	recall	f1-score	support
0	0.84	0.93	0.88	1035
1	0.72	0.50	0.59	374
accuracy			0.82	1409
macro avg	0.78	0.71	0.74	1409
weighted avg	0.81	0.82	0.80	1409

Support Vector Classifier with K = 10

	Predicted No Churn	Predicted Churn
Actual No Churn	948	87
Actual Churn	187	187

	precision	recall	f1-score	support
0	0.84	0.92	0.87	1035
1	0.68	0.50	0.58	374
accuracy			0.81	1409
macro avg	0.76	0.71	0.73	1409
weighted avg	0.79	0.81	0.80	1409

Support Vector Classifier with PCA

	Predicted No Churn	Predicted Churn
Actual No Churn	912	123
Actual Churn	160	214

	precision	recall	f1-score	support
0	0.85	0.88	0.87	1035
1	0.64	0.57	0.60	374
accuracy			0.80	1409
macro avg	0.74	0.73	0.73	1409
weighted avg	0.79	0.80	0.80	1409

Random Forest Model Comparison

Random Forest Classifier

	Predicted No Churn	Predicted Churn
Actual No Churn	935	100
Actual Churn	183	191

	precision	recall	f1-score	support
0	0.84	0.90	0.87	1035
1	0.66	0.51	0.57	374
accuracy			0.80	1409
macro avg	0.75	0.71	0.72	1409
weighted avg	0.79	0.80	0.79	1409

Random Forest Classifier with K = 10

	Predicted No Churn	Predicted Churn
Actual No Churn	943	92
Actual Churn	180	194

	precision	recall	f1-score	support
0	0.84	0.91	0.87	1035
1	0.68	0.52	0.59	374
accuracy			0.81	1409
macro avg	0.76	0.71	0.73	1409
weighted avg	0.80	0.81	0.80	1409

Random Forest Classifier with PCA

	Predicted No Churn	Predicted Churn
Actual No Churn	948	87
Actual Churn	187	187

	precision	recall	f1-score	support
0	0.84	0.92	0.87	1035
1	0.68	0.50	0.58	374
accuracy			0.81	1409
macro avg	0.76	0.71	0.73	1409
weighted avg	0.79	0.81	0.80	1409

Gradient Boosted Model Comparison

Gradient Boosted Classifier

	Predicted No Churn	Predicted Churn
Actual No Churn	923	112
Actual Churn	174	200

	precision	recall	f1-score	support
0	0.84	0.89	0.87	1035
1	0.64	0.53	0.58	374
accuracy			0.80	1409
macro avg	0.74	0.71	0.72	1409
weighted avg	0.79	0.80	0.79	1409

Gradient Boosted Classifier with K = 10

	Predicted No Churn	Predicted Churn
Actual No Churn	924	111
Actual Churn	186	188

	precision	recall	f1-score	support
0	0.83	0.89	0.86	1035
1	0.63	0.50	0.56	374
accuracy			0.79	1409
macro avg	0.73	0.70	0.71	1409
weighted avg	0.78	0.79	0.78	1409

Gradient Boosted Classifier with PCA

	Predicted No Churn	Predicted Churn
Actual No Churn	917	118
Actual Churn	181	193

	precision	recall	f1-score	support
0	0.84	0.89	0.86	1035
1	0.62	0.52	0.56	374
accuracy			0.79	1409
macro avg	0.73	0.70	0.71	1409
weighted avg	0.78	0.79	0.78	1409

Concluding Thoughts

- After conducting SelectKBest and PCA, majority of the models precision and recall scores improved
    - Some features contained redundant information
    - PCA and SelectKBest did not result with excessive loss of information
- Each model had its own pros and cons, but specifically focusing on determining true positives it appears the K Nearest Neighbor with SelectKBest had the most optimal outcome
    - Had more correctly predicted churned customers
    - Also still successfully predicting customers who didn’t churn
- The Gradient Boosted Model performed better without PCA and SelectKBest
    - However, results were lackluster compared to other models
    - Execution time took longer for this model when compared to others

Concluding Thoughts

- Using the best model (KNN with SelectKBest)
  - Correctly predicted 209 out of 374 customers would churn
  - Correctly predicted 921 out of 1035 customers would not churn
- This model can be used to
  - Help telecommunications companies determine which customers will stay and which will leave their services
  - Determine what features play a bigger role in keeping customers satisfied


Thank You!