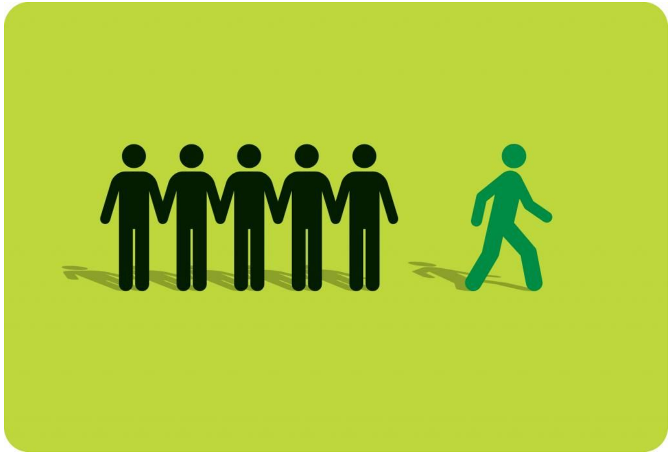
Predicting Customer Churn

And Identifying Retention Strategies to Reduce it

By: Jason Powell

Agenda

1. Problem
2. Data Source
3. Data Exploration
4. Feature Engineering & Selection
5. Modeling
6. Implementation
7. Conclusions

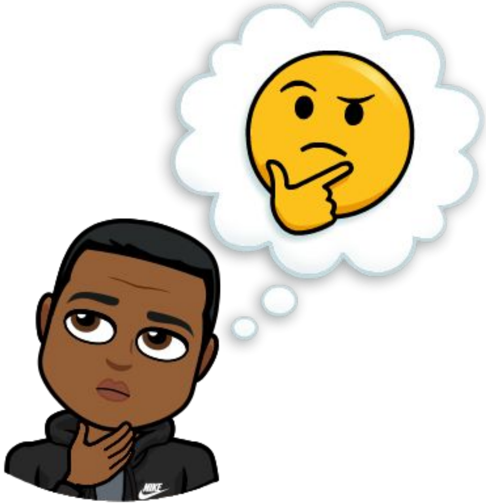


C
H
U
R
N

Image Source: Sujan Patel

Problem

1 Churn is a problem that afflicts every company. Building a model that accurately predicts churn will be an aid to any company's bottom line.



Data Source

Kaggle:

Telco Customer Churn

Data Exploration

Data Exploration

1

21 Features

7,043 Records

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...

Data Exploration

2

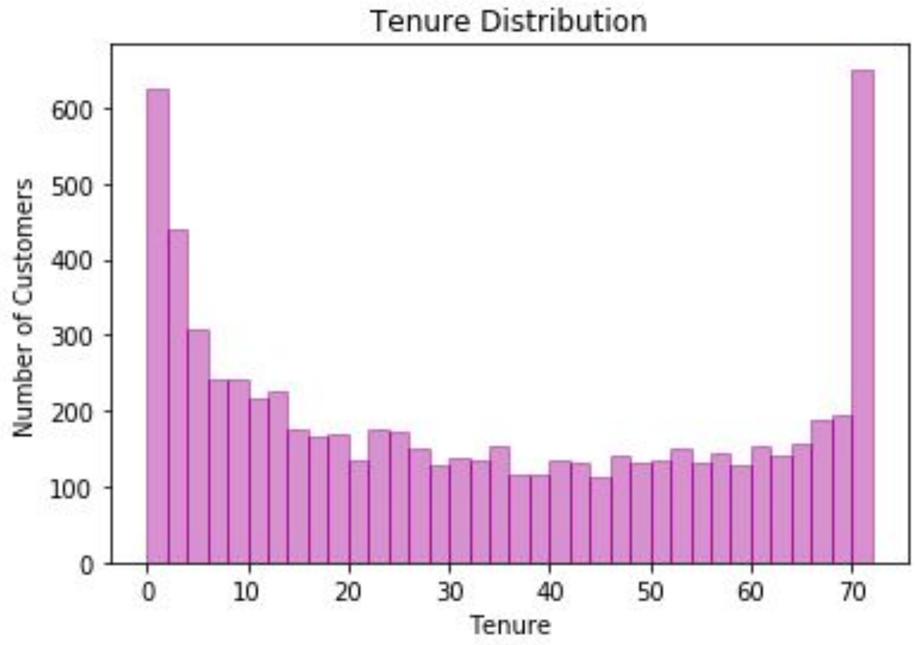
Data Types:

- ★ Nominal: 18
- ★ Ratio: 3

```
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
customerID          7043 non-null object
gender              7043 non-null int64
SeniorCitizen       7043 non-null int64
Partner             7043 non-null int64
Dependents          7043 non-null int64
tenure              7043 non-null int64
PhoneService        7043 non-null int64
MultipleLines       7043 non-null object
InternetService     7043 non-null object
OnlineSecurity      7043 non-null object
OnlineBackup        7043 non-null object
DeviceProtection    7043 non-null object
TechSupport         7043 non-null object
StreamingTV         7043 non-null object
StreamingMovies     7043 non-null object
Contract            7043 non-null object
PaperlessBilling    7043 non-null int64
PaymentMethod       7043 non-null object
MonthlyCharges      7043 non-null float64
TotalCharges        7043 non-null float64
Churn               7043 non-null int64
```

Data Exploration

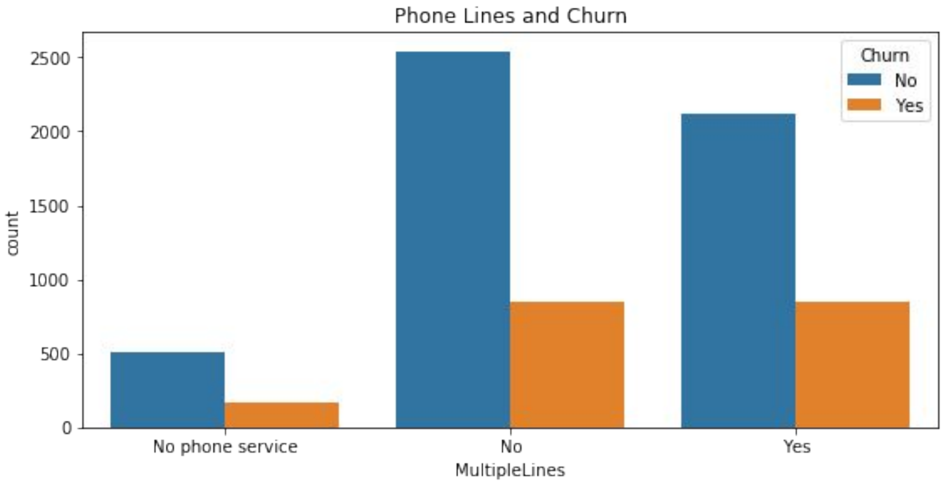
3 **Customer Tenure**

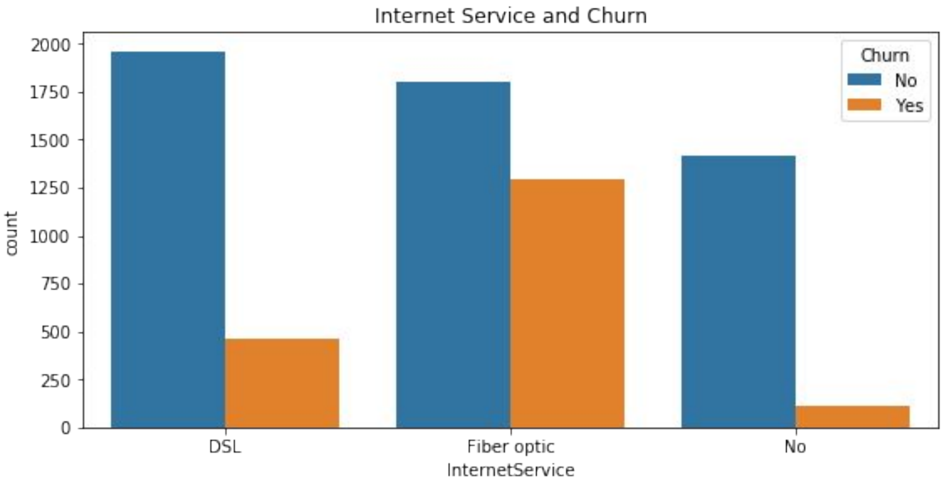



```
The average customer tenure is 32.37 months.
```

Data Exploration

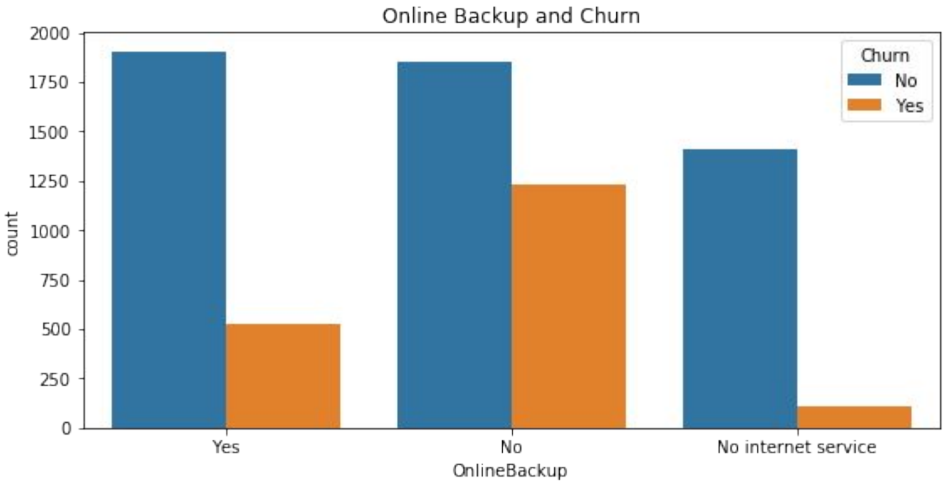
4 **Core Products and Churn**





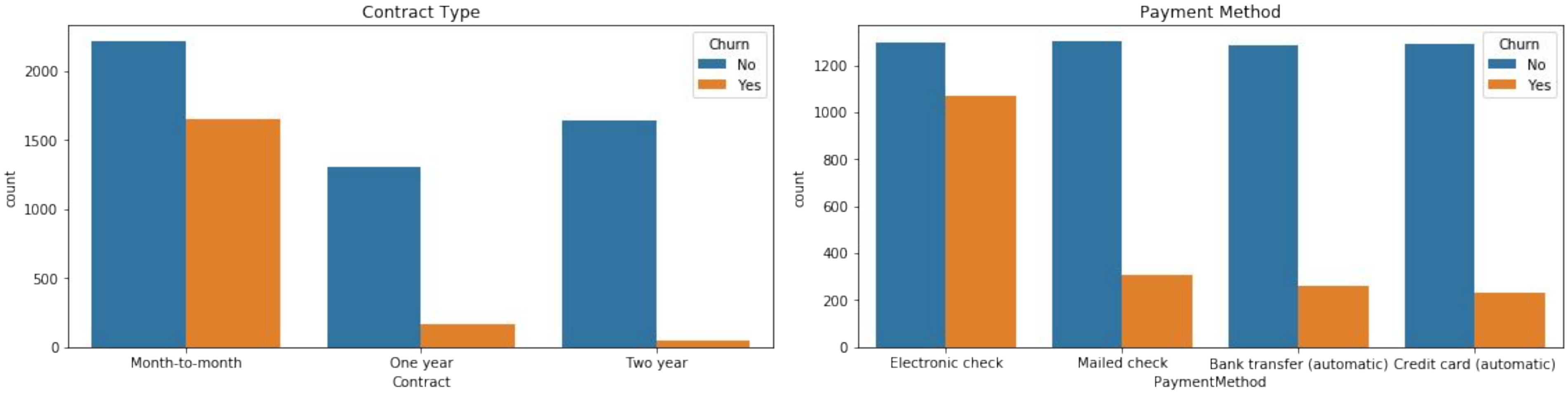
Data Exploration

5 **Non-Core Products and Churn**



Data Exploration

6 **Contract Type, Payment Methods and Churn**



Feature Engineering & Selection

Feature Engineering

1. Converting categorical binary features to numerical

	gender	Partner	Dependents	PaperlessBilling	PhoneService	Churn
0	Female	Yes	No	Yes	No	No
1	Male	No	No	No	Yes	No
2	Male	No	No	Yes	Yes	Yes
3	Male	No	No	No	No	No
4	Female	No	No	Yes	Yes	Yes

→

	gender	Partner	Dependents	PaperlessBilling	PhoneService	Churn
0	0	1	0	1	0	0
1	1	0	0	0	1	0
2	1	0	0	1	1	1
3	1	0	0	0	0	0
4	0	0	0	1	1	1

Feature Engineering

2 One-Hot Encoding

	MultipleLines	InternetService	Contract	PaymentMethod
0	No phone service	DSL	Month-to-month	Electronic check
1	No	DSL	One year	Mailed check
2	No	DSL	Month-to-month	Mailed check
3	No phone service	DSL	One year	Bank transfer (automatic)
4	No	Fiber optic	Month-to-month	Electronic check
5	Yes	Fiber optic	Month-to-month	Electronic check
6	Yes	Fiber optic	Month-to-month	Credit card (automatic)
7	No phone service	DSL	Month-to-month	Mailed check
8	Yes	Fiber optic	Month-to-month	Electronic check
9	No	DSL	One year	Bank transfer (automatic)

[
'MultipleLines_No',
'MultipleLines_No phone service',
'MultipleLines_Yes',
'InternetService_DSL',
'InternetService_Fiber optic',
'InternetService_No',
'Contract_Month-to-month',
'Contract_One year',
'Contract_Two year',
'PaymentMethod_Bank transfer (automatic)',
PaymentMethod_Credit card (automatic)',
'PaymentMethod_Electronic check',
'PaymentMethod_Mailed check'
]

Feature Engineering

3

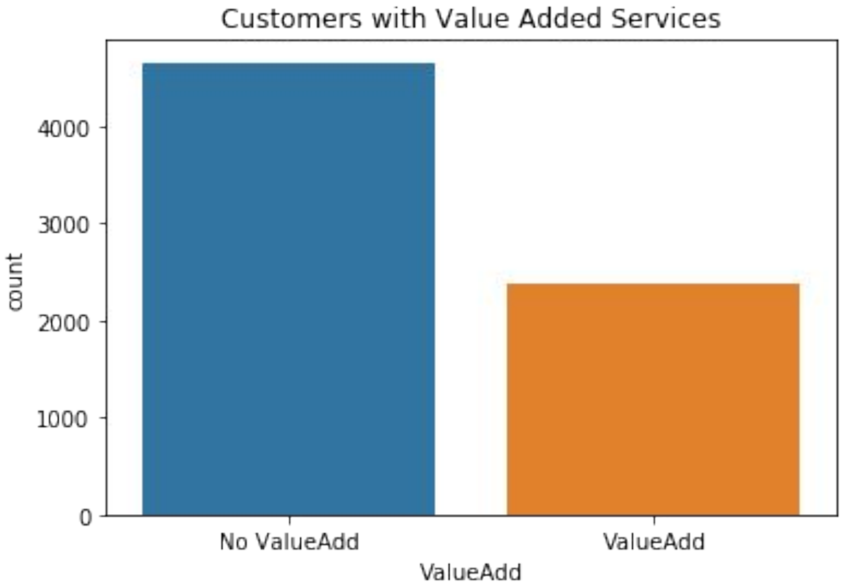
Create the ‘ValueAdd’ Feature

DeviceProtection_Yes	OnlineSecurity_Yes	OnlineBackup_Yes	StreamingTV_Yes	StreamingMovies_Yes	ValueAdd
1	1	0	1	1	1
1	0	1	1	1	1
0	1	0	0	0	0
0	0	0	0	0	0
1	1	0	1	1	1

Data Exploration

4

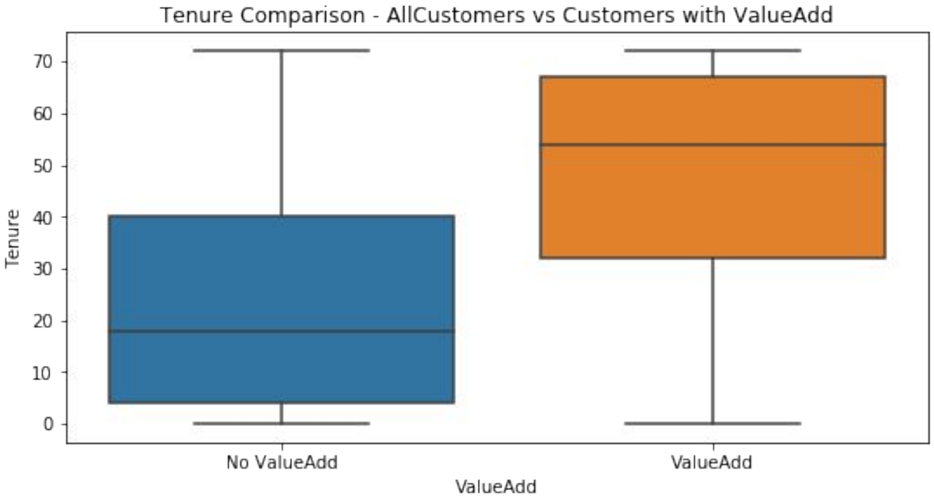
~ 33% of customers make use of value added services.

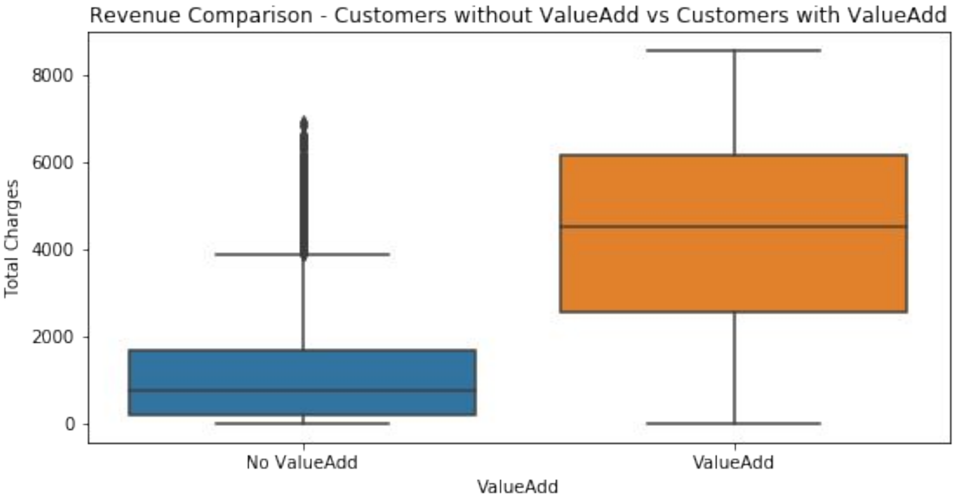


Feature Engineering

5

Impact of Value Added Services



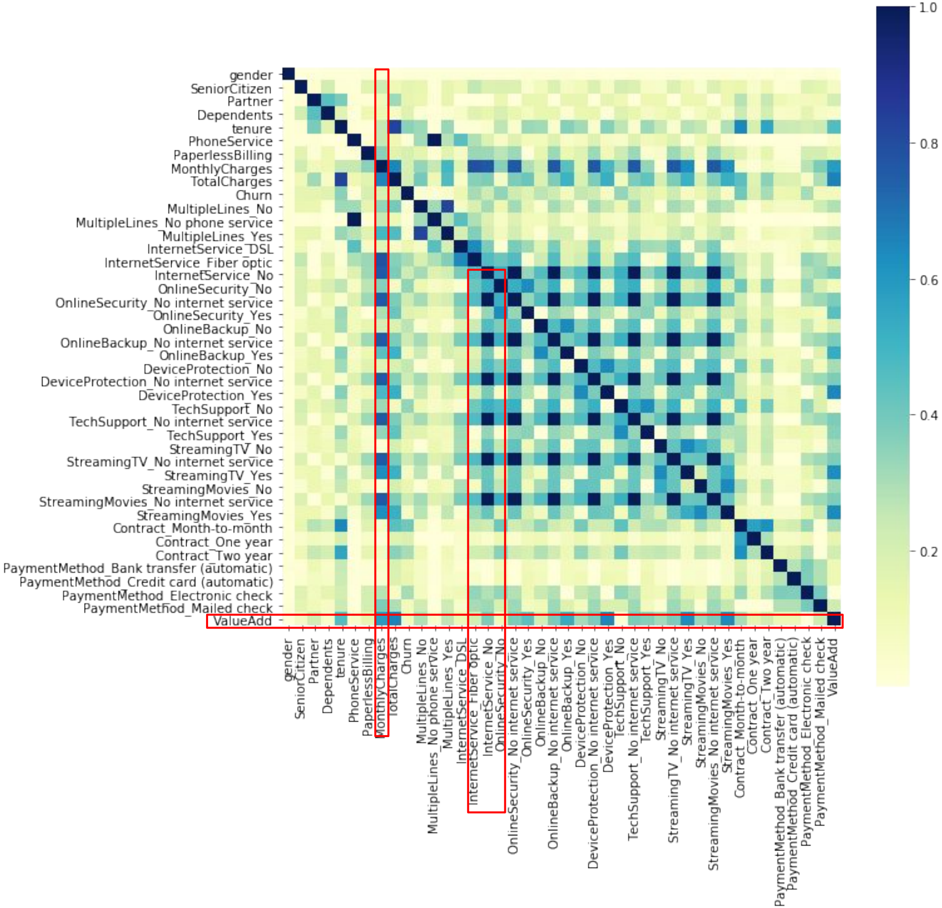


Feature Selection

1

Correlation Matrix

- ★ Drop all features that include “no internet”
- ★ Drop TotalCharges



Models

Models - Approach

1. Problem Type: Classification
2. Data Type: Mainly categorical.
3. Candidate Models: Logistic Regression, Random Forest, Bernoulli Naive Bayes, Neural Network (MLP)

Logistic Regression

Accuracy: 80.5
CV Accuracy: 80.3% (+/- 2.1%)

	0	1	All
Churn			
0	4647	527	5174
1	849	1020	1869
All	5496	1547	7043

- Sensitivity: 89.81%
- Specificity: 54.57%
- T1 Error Rate: 7.48%
- T2 Error Rate: 12.05%

Random Forest Classifier

Accuracy: 86.1 %
CV Accuracy: 80.1% (+/- 2.6%)

	0	1	All
Churn			
0	4833	341	5174
1	638	1231	1869
All	5471	1572	7043

- **Sensitivity: 93.41%**
- **Specificity: 65.86%**
- **T1 Error Rate: 4.84%**
- **T2 Error Rate: 9.05%**

Multilayered Perceptron Network (MLP)

Accuracy: 80.2%
CV Accuracy: 80.0% (+/- 2.2%)

	0	1	All
Churn			
0	4515	659	5174
1	733	1136	1869
All	5248	1795	7043

- Sensitivity: 87.26%
- Specificity: 60.78%
- T1 Error Rate: 9.35%
- T2 Error Rate: 10.41%

Model Selection

3 **Selection:** Random Forest

Models - Results

1. Churners: 1532 (22%)

2. Characteristics:
    - Pay month-to-month via electronic check
    - Have fiber optic internet
    - Do not use value added services
    - Do not have dependents

Implementation

Implementation

1. The model generates churn probabilities, not predictions.
2. The probabilities are inputs to an algorithm that aids decision making.

Conclusions

Conclusions

1. Companies should focus on how long customers stay, not whether or not they stay.
2. Value added services are sticky.
3. More data about the customers is needed.

Questions?

Thank you.

e: powell.jasonm@gmail.com