

# Cell Company's Churn Conundrum

Tim, Kyle, Calvin - 3/27/20

Background: <https://www.uswitch.com/mobiles/guides/history-of-mobile-phones/>

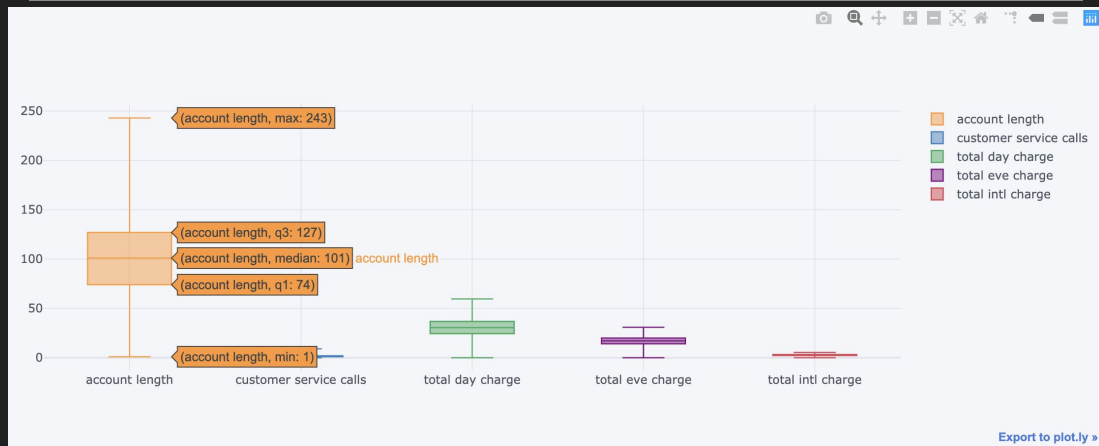
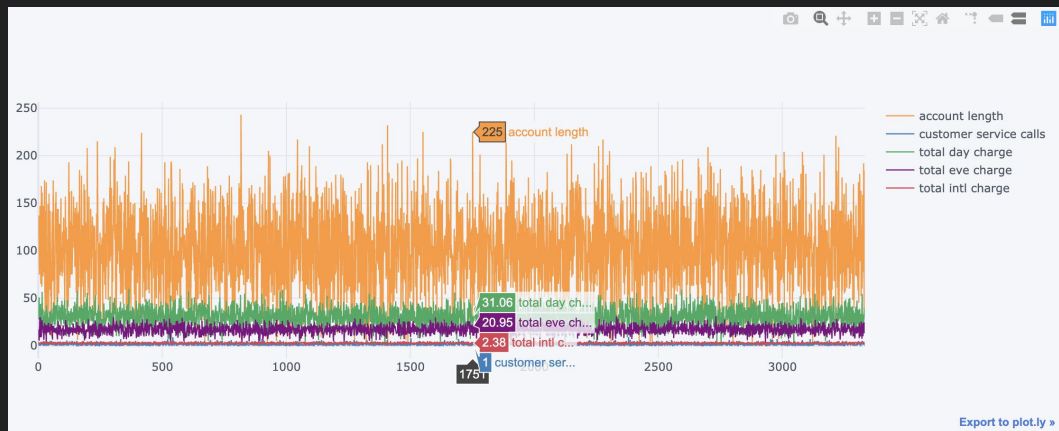
# Problem Domain

1. Customers leaving a service providing company is not good (for that company)
2. Some customers bring in a lot of business and when they leave it hurts the company
3. A company's best interest to figure out who is staying and who might leave
4. This is called customer churn and is generally measured as a percentage called churn rate
5. In this case for a cell phone service provider, it would be nice to determine this
6. A model to help find when a customer is getting to that point is best
7. Company can provide incentives to the leaving customers to retain their business

# Overview of Methodology

1. Load data and explore what columns are or aren't useful
2. First model: Dummy Classifier
3. Second model: Random Forest Classifier model
4. Third model: Gradient Boosting
5. Between the 3 models we tried, we also tried various settings for each

# Data Exploration



# Base Model - Dummy Classifier

1. Dummy Classifier model gave us a baseline to improve
2. Tried 3 models with no modified settings to compare
3. Right away had improvements vs Dummy model

```
=====
KNeighborsClassifier
***Results***
      precision    recall  f1-score   support

   False       0.89       0.99       0.94       709
    True       0.87       0.27       0.41       125

 accuracy          0.88
 macro avg         0.88       0.63       0.68       834
weighted avg         0.88       0.88       0.86       834

=====
RandomForestClassifier
***Results***
      precision    recall  f1-score   support

   False       0.91       0.99       0.95       709
    True       0.89       0.45       0.60       125

 accuracy          0.91
 macro avg         0.90       0.72       0.77       834
weighted avg         0.91       0.91       0.90       834

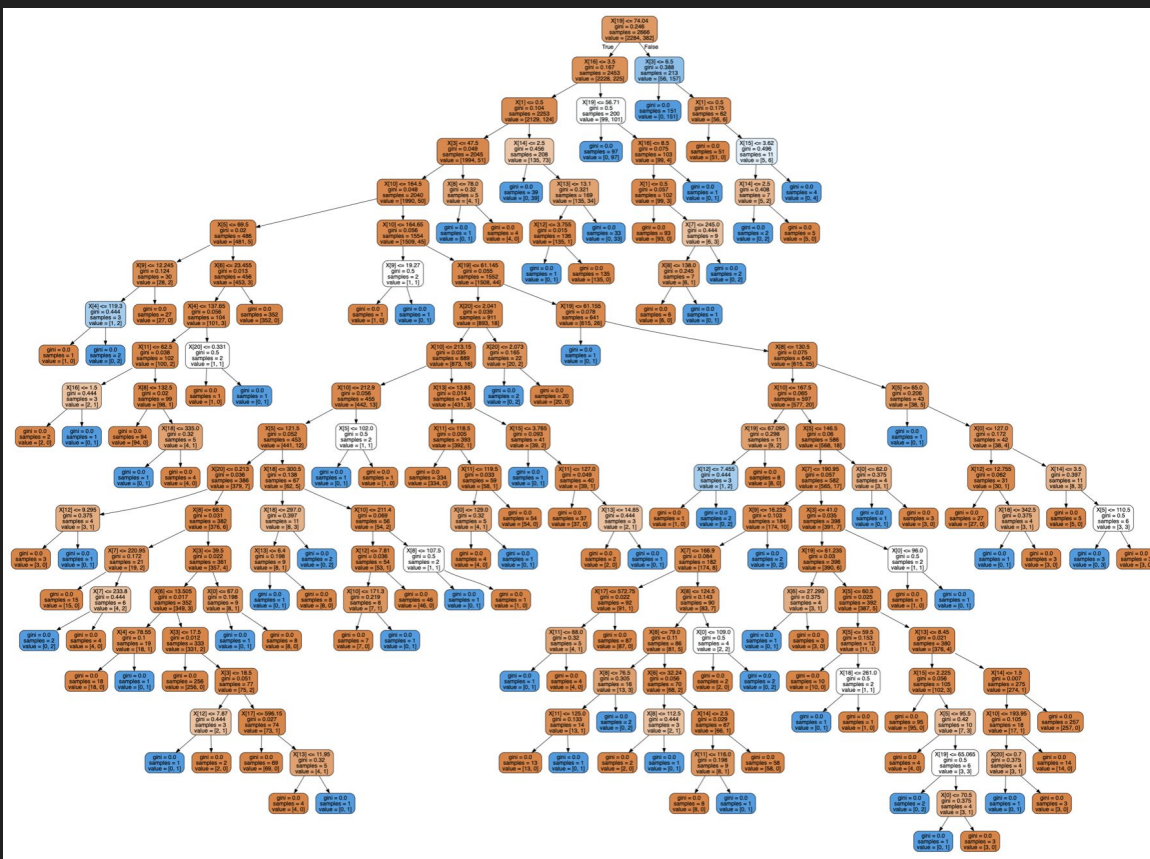
=====
GradientBoostingClassifier
***Results***
      precision    recall  f1-score   support

   False       0.92       0.99       0.95       709
    True       0.86       0.50       0.63       125

 accuracy          0.91
 macro avg         0.89       0.74       0.79       834
weighted avg         0.91       0.91       0.90       834

=====
```

# Second Model - Random Forest Classifier



# Final Model - Gradient Boosting

1. Predict if customers would cancel their accounts in the near future
2. Can't say with certainty which factor is most important
3. Can identify about 88% of customers who may leave
4. Pushing this further would have forced our model to be perfect
5. This causes issues when trying to generalize the model with new data

# Specific Insights

- Feature Importances
  - Total Cost 0.45
  - Customer service Calls 0.13
  - International Plan 0.11
- More Data
  - Duration of plans
  - Sentiment Analysis of CS calls





# Conclusion & Recommendations

- Successfully able to predict 88% of potential customer churn
- Found out what problems are most correlated with customers leaving
- Focus our efforts to increase customer retention
- Recommend collecting additional customer data for analysis:
  - Credit score
  - Regular payments
- 'total cost' being the highest coefficient in determining the churn
- Make sure the customers are on the correct plan for their usage