# Telecom Customer Churn Visualisation & Analysis

Kieran Baker

## Executive Summary

This is an analysis of the ‘Telecom Customer Churn Prediction’ dataset from Kaggle.com, which consists of various demographic and account-related attributes of roughly 7000 customers and former customers of an unnamed telecommunications company, along with their churn – whether or not the customer left the company in the last month.

The goal of the report was to visualise the data and provide analysis to determine which attributes contribute to churn and so should be the main focus of the company when deciding where to spend resources to retain customers.

The analysis identified four especially significant variables:

- Age – being a senior citizen makes a customer more likely to churn
- Having a partner makes a customer less likely to churn
- A higher monthly charge makes a customer more likely to churn
- Having multiple phone lines makes a customer more likely to churn

## Initial Data Analysis

### Summary

Each row of the data set represents a customer, and the columns contain information about:

- Demographics – gender, age (whether or not they are a senior citizen), if they have a partner, if they have dependents.
- Account – tenure (number of months they have been a customer), contract type, payment method, whether they have paperless billing, monthly charges and total charges
- Services – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Churn

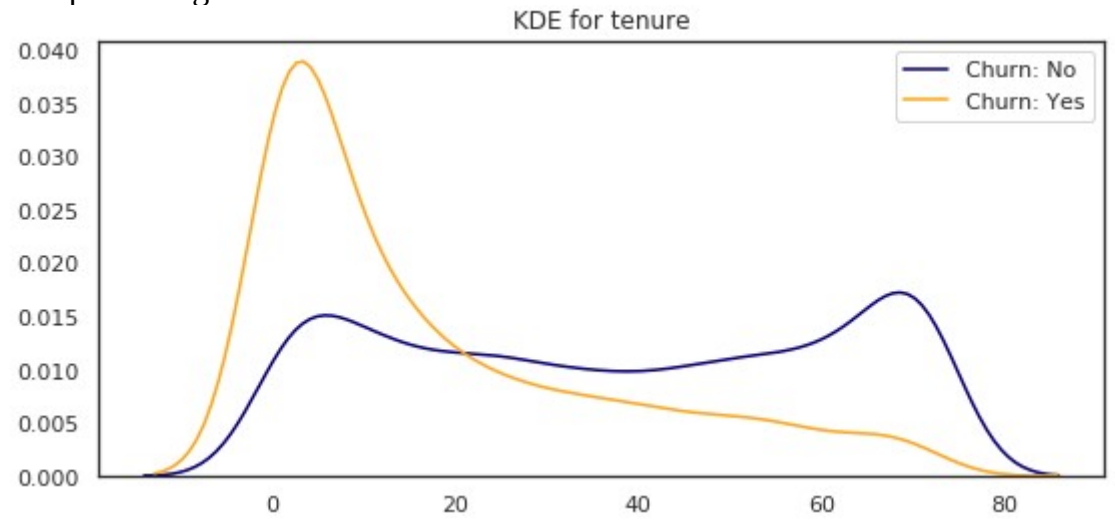
### Missing Data

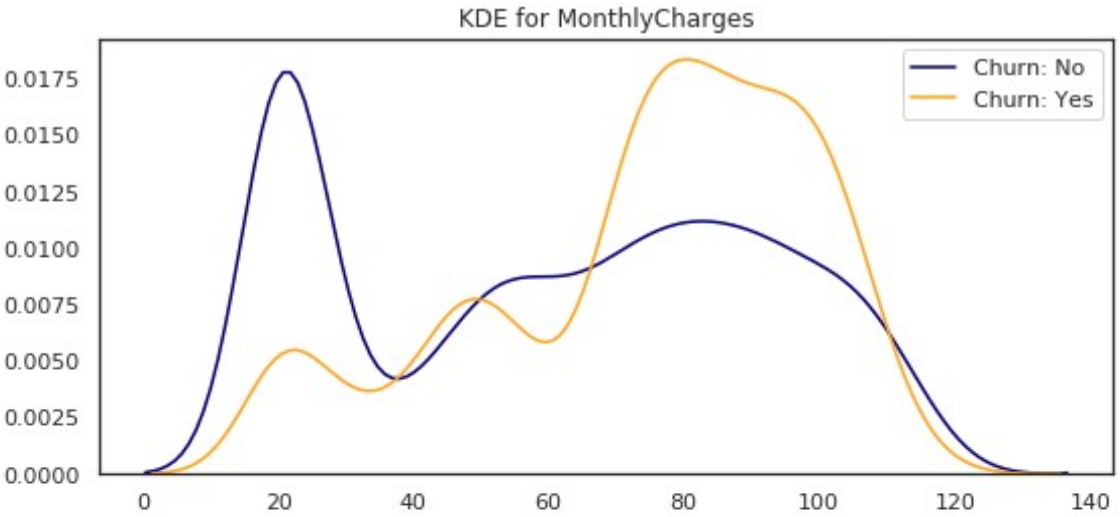
Since this is originally a sample data set from IBM there are really no missing values or much cleaning to be done at all. However the total charges column has 11 missing values, which presumably correspond to customers who joined the company in the last month and have not yet been billed. Therefore these values were imputed with 0.

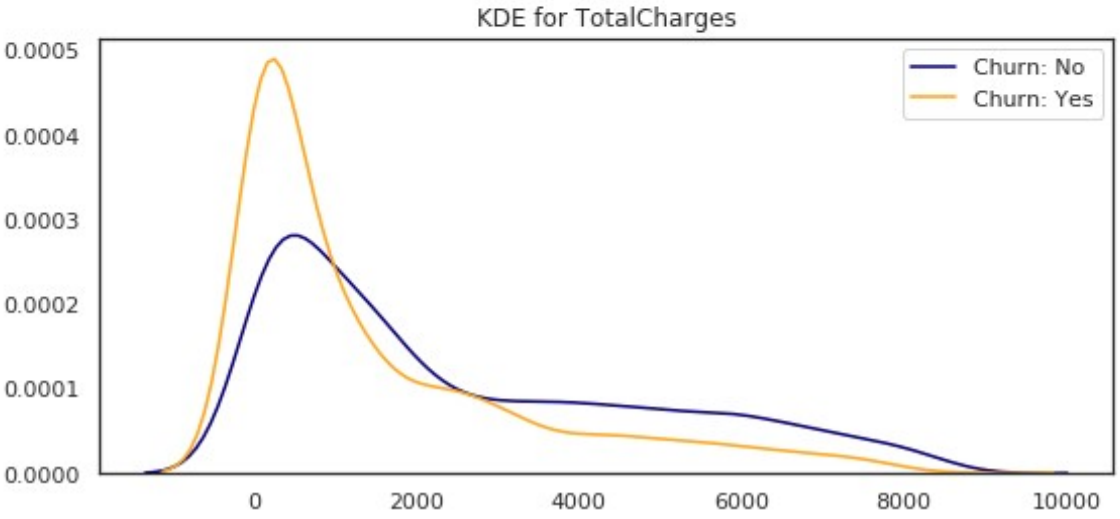
### Numerical Features

Only three columns are numerical – tenure, and the monthly and total charges. We can use kernel density estimation to estimate the probability density function for each for churn and non-churn customers to get an idea of how these features affect churn.

The graphs give three insights: recent customers (low tenure) are more likely to churn, customers with high monthly charges are more likely to churn, and total charges is not useful for predicting churn.



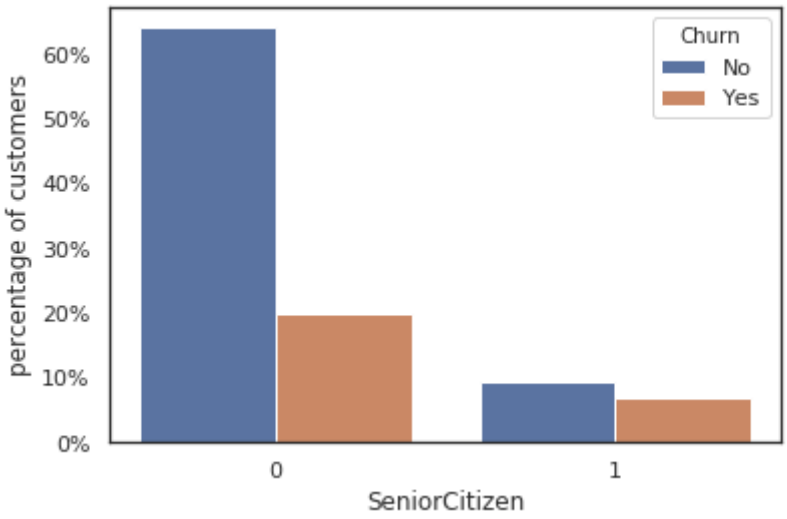



## Categorical Features

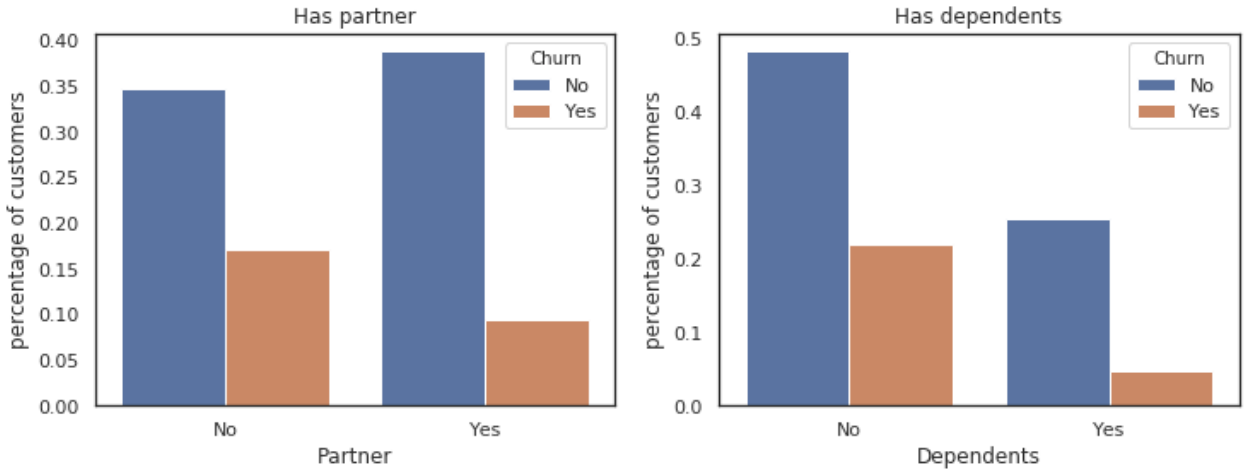
The remaining 16 features are categorical, with six being binary. We present charts for some of the more important features to how they affect churn.

### Senior Citizens



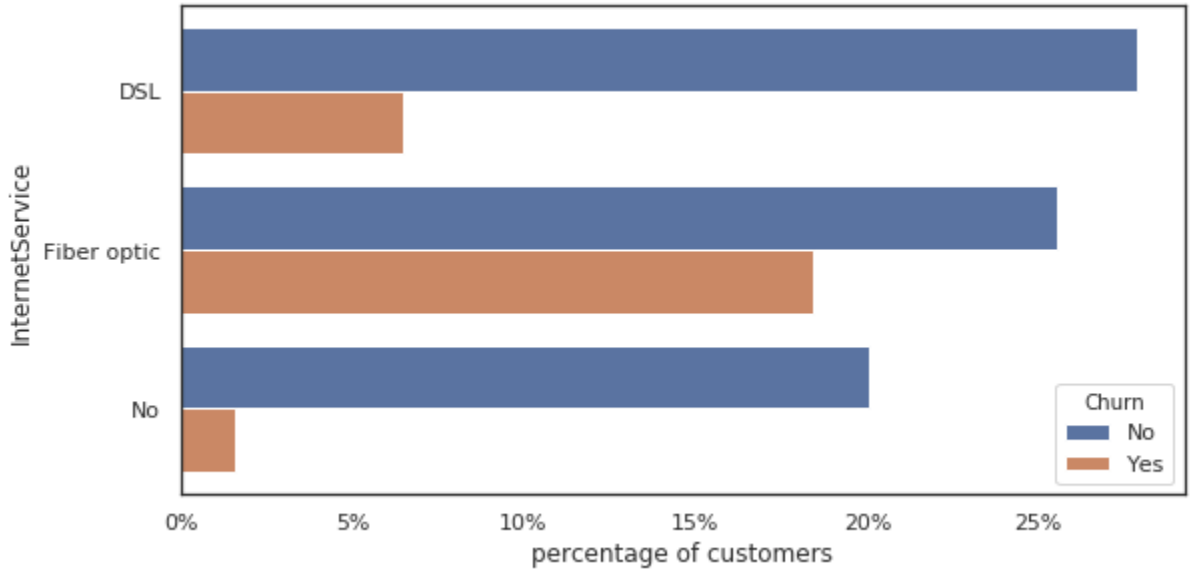

Senior citizens represent a small percentage of customers, but have roughly double the churn rate of non-senior citizens.

### Partner and Dependents



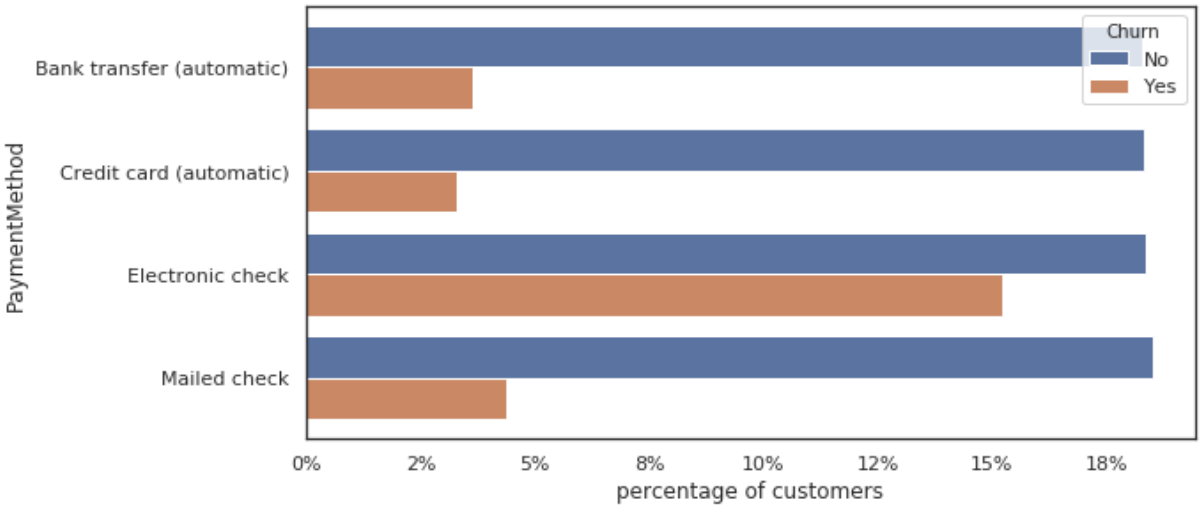

Customers with no partner and customers with no dependents are more likely to churn.

## Internet Services




Customers with fiber have an extremely high churn rate, while those with no internet have an extremely low churn rate. This latter fact is especially surprising considering one would probably expect no internet service to correlate at least somewhat positively with senior citizens, this would be interesting to look into in a more in depth analysis.

## Payment Method




Electronic checking is very highly correlated to churn. This is an important insight as this is the most common payment method with around 35% of customers.

## Correlation and Relationships

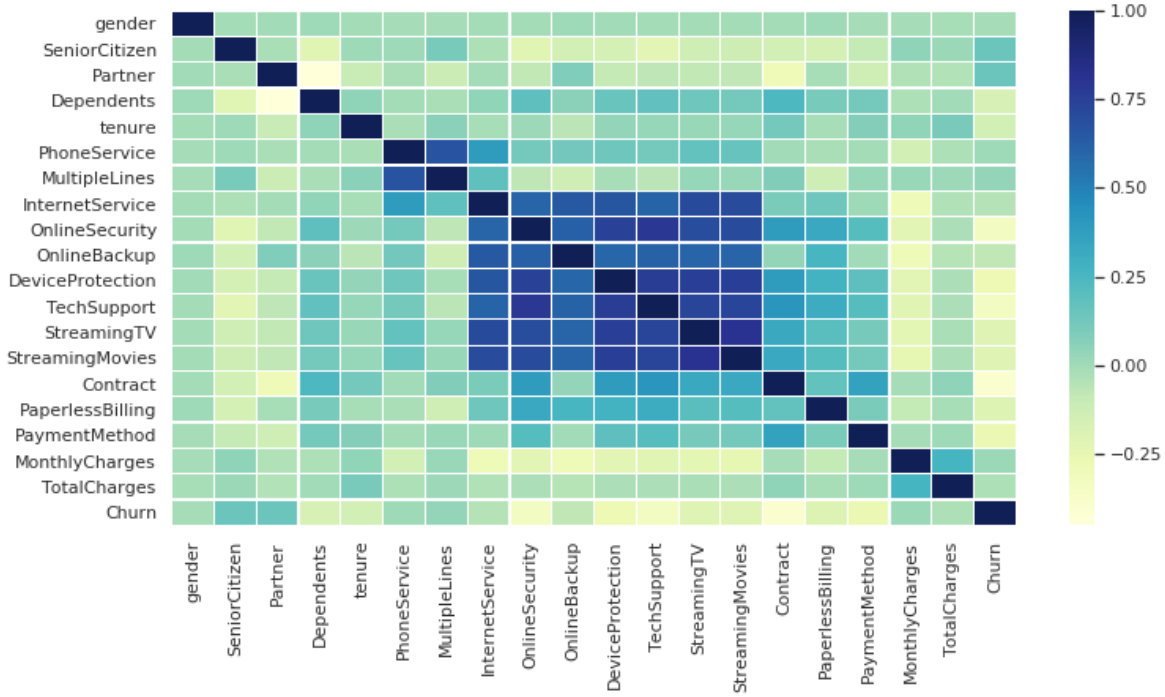



*Illustration 9: Correlation heatmap of variables*

It is perhaps unsurprising that the central service variables are so highly correlated (e.g. a customer who pays for tech support would be more likely to also pay for online back ups).

## Analysis, Reasoning and Conclusion

The features most closely correlated to churn are senior citizen, partner, monthly charges and multiple lines. Interestingly being a senior citizen is not particularly strongly correlated with having internet access.

Given the constraint on the length of this report I decided to only look into two-way relationships between the various features and customer churn, as this is the most important question for the telecom business.

Of the four variables identified, three would seem to have obvious explanations. Senior citizens are more likely to be retired and so have more time to shop around for different providers. Customers with a partner presumably would have to consult with that partner about changing providers and so those without one would be more likely to do so because they can make ‘spur of the moment’ decisions about it. Paying more each month provides an obvious incentive to switch to a different provider if they can pay less.

The most interesting feature identified is multiple lines – it might be thought that this causes customers to churn simply because they are paying more, but the heatmap shows monthly charges and multiple lines have almost zero correlation. This suggests there is something about the company’s service for multiple-line customers that a large number of them find unsatisfactory, and this should be the first thing the business looks into.