Churn Analysis of a Telecommunication Company

Springboard Data Science Capstone Project

Presented by: Cuili Huang

Agenda

01 Introduction

02 Data Acquisition and Cleaning

03 Data Exploration

Agenda

Introduction

Churn Rate

The percentage of service subscribers who discontinue their subscriptions within a given time period.

Introduction

Who Cares About Churn Rate?

Companies whose revenues are based on subscriptions (such as telecommunications, newspapers, SAAS (Software as a service), etc.

Introduction

The Use of Churn Analysis:

Customers Retention Strategy

Data Acquisition and Cleaning

Data Acquisition

- Dataset acquired from Kaggle
- Contains the demographic information of the customers, services subscribed, monthly payments, their tenures, etc.
- Contains 21 variables and 7043 observations

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
customerID          7043 non-null object
gender              7043 non-null object
SeniorCitizen       7043 non-null int64
Partner             7043 non-null object
Dependents          7043 non-null object
tenure              7043 non-null int64
PhoneService        7043 non-null object
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 object
PaymentMethod       7043 non-null object
MonthlyCharges      7043 non-null float64
TotalCharges        7043 non-null object
Churn               7043 non-null object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
None
```

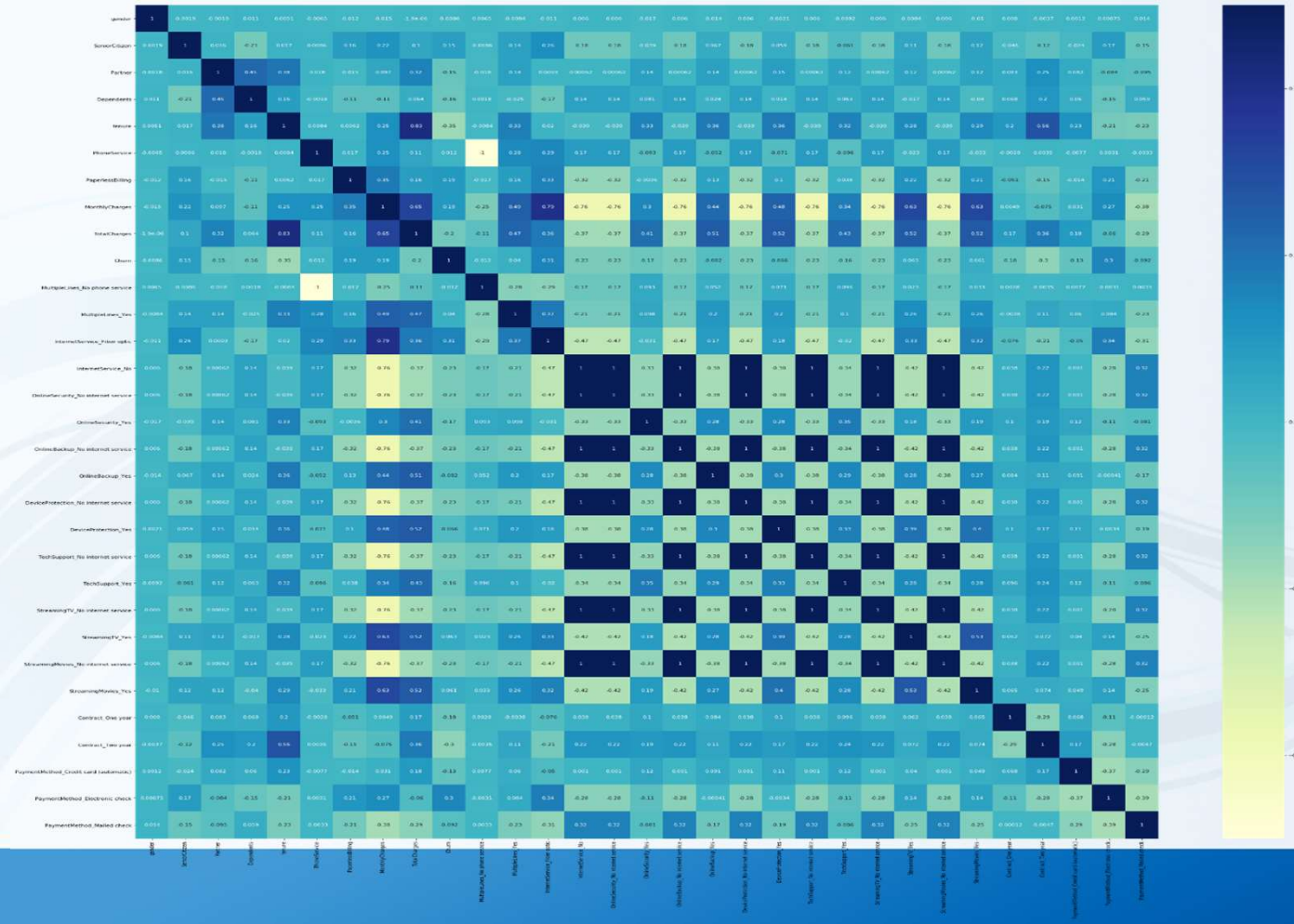
Data Acquisition and Cleaning

Data Cleaning

- Value count categorical variables
- Calculate statistics for numerical variables and try to identify outliers
- Handle missing values (fill up with median value)
- Convert categorical variables into dummy variables, correct data type (from “string” to “float”)

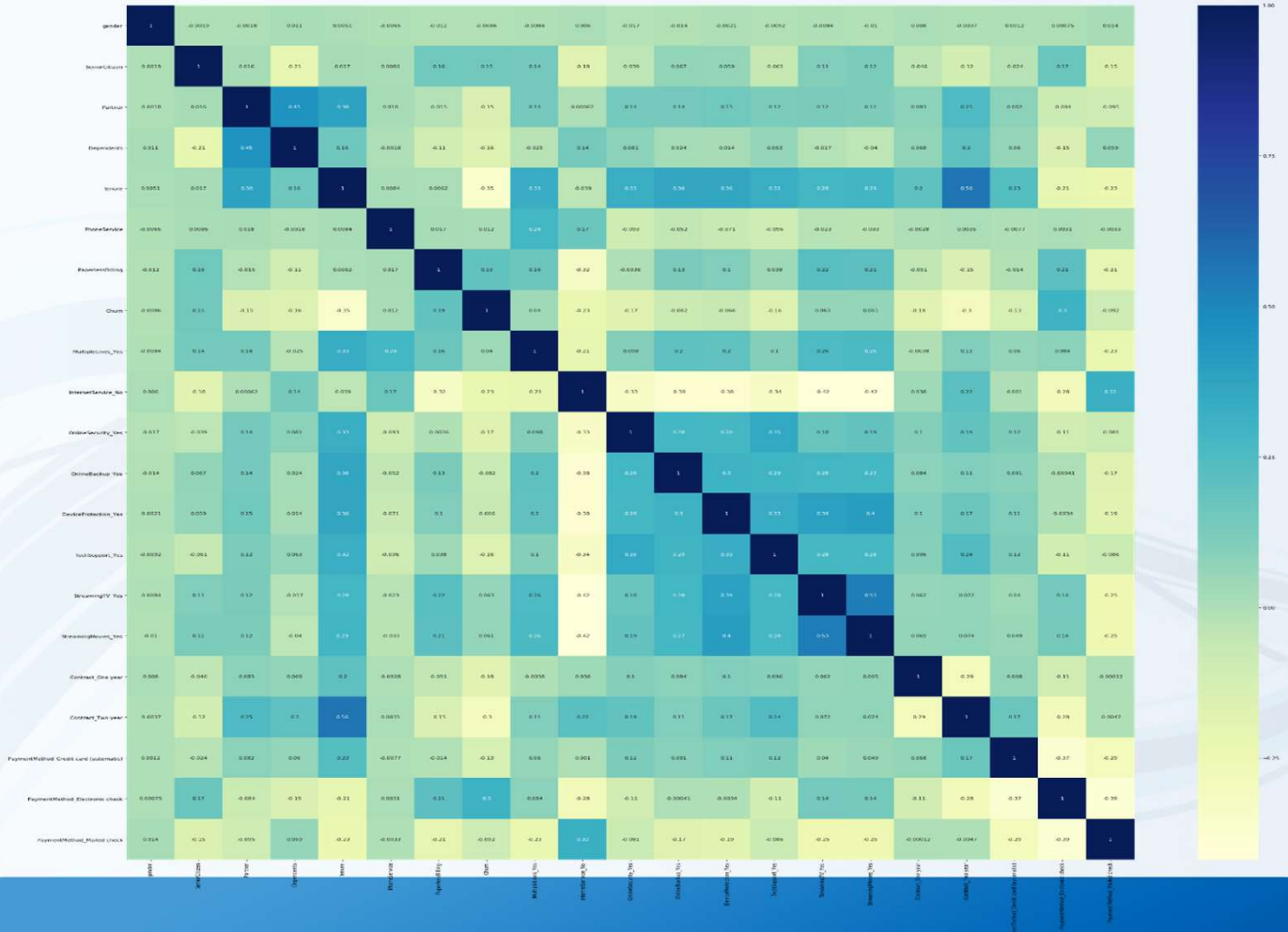
Data Exploration

Plot correlation heatmap to investigate the collinearity among features



Data Exploration

Delete some collinear features and re-plot the correlation heatmap



THANKS