

Customer Churn Prediction

Results from investigating last years customers

Andreas Ährlund-Richter Data Analyst

01-03-2024

(OBS case presentation, synthetic data:
<https://www.kaggle.com/datasets/yeanzc/telco-customer-churn-ibm-dataset>)

Summary

- Last year was a customer lifetime loss of about 4 million dollar, looking at mean retained customer lifetime spending
- A predictor with 50% accuracy and 80% recall (could detect 80% of churning) was made
- Half of Single Household Fiber Optic owners churn, almost 50% of all churners, campaigns aimed at these groups is a good idea

Outline

1. Customer profile and data
2. Predictor
3. Important factors for churn risk.

Customer Data

Data profile

- 7000 customers active last year
- 1869 churned
- Subscription date currently unavailable

Customer Data

Customer profile

- About 1/5 of the customers are seniors, about half are single households
- About half of customers pay by monthly subscription instead of yearly contracts
- Non-churned customer lifetime spending of 855 dollar

Customer Data

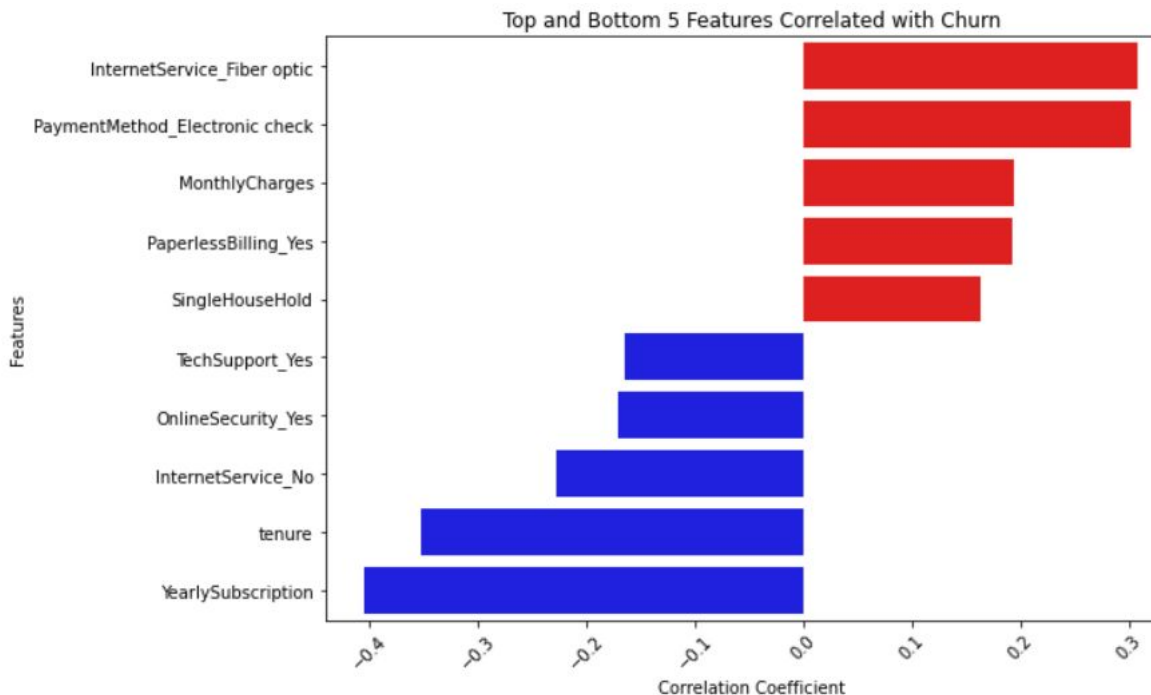
Churn profile

- About 1/5 of the customers are seniors
- About half of customers pay by monthly subscription instead of yearly contracts
- Non-churned customer lifetime spending of 855 dollar

Most important factors at a glance

Red is correlated with **churning**, blue with **staying subscribed**

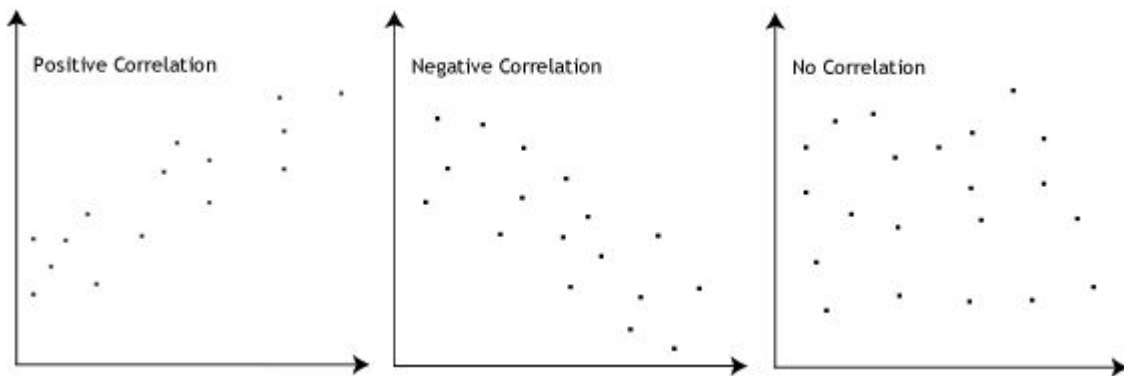
- Yearly subscriptions is a success
- InternetService_Fiber optic needs investigating
- Singles Churn more



Most important factors at a glance

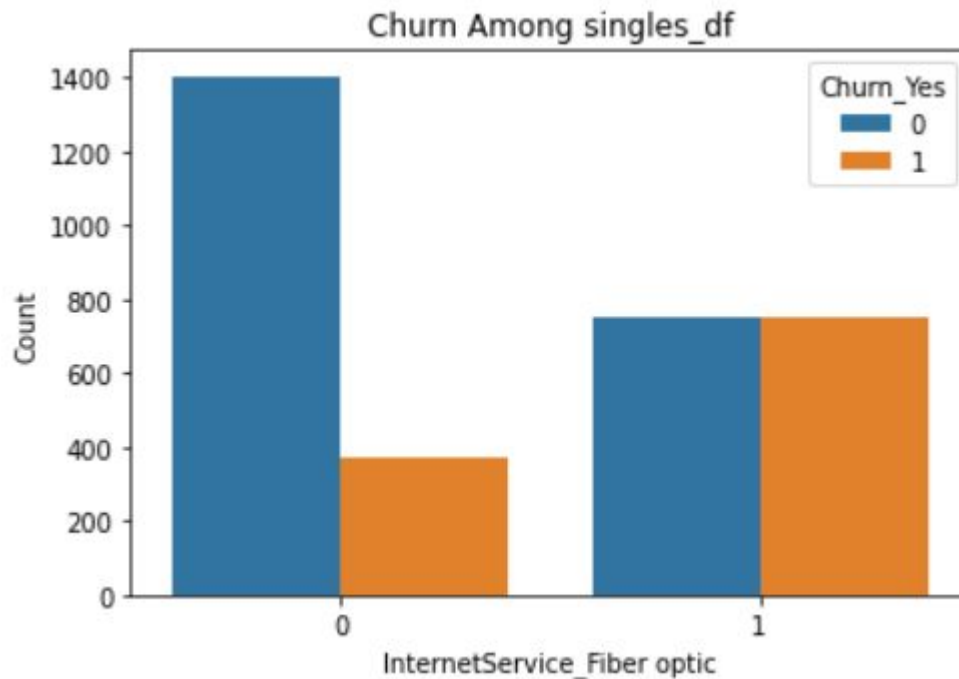
Pearson Correlation recap:

If variables of two types tend to be large or small at the same time, negatively or positively, you have correlation.



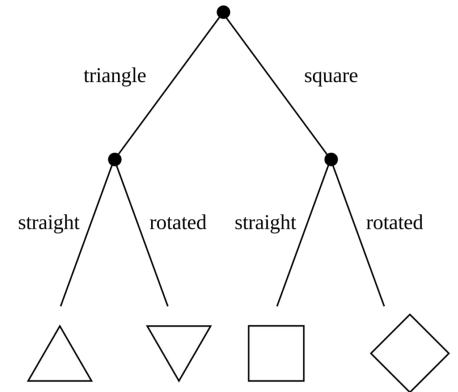
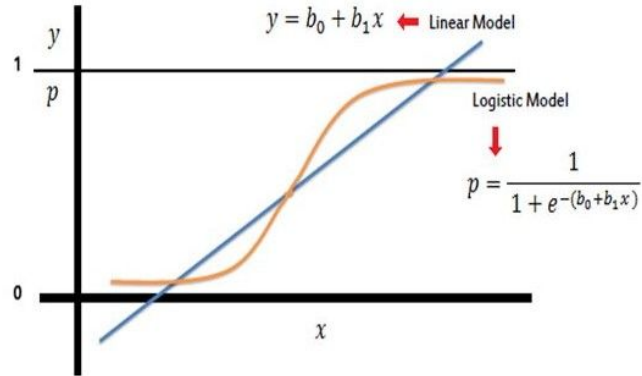
Almost half of single Fiber optic owners churn

Singlehouseholds with Fiber Optic internet services churn very frequently, and constitute almost half of all churn-events.



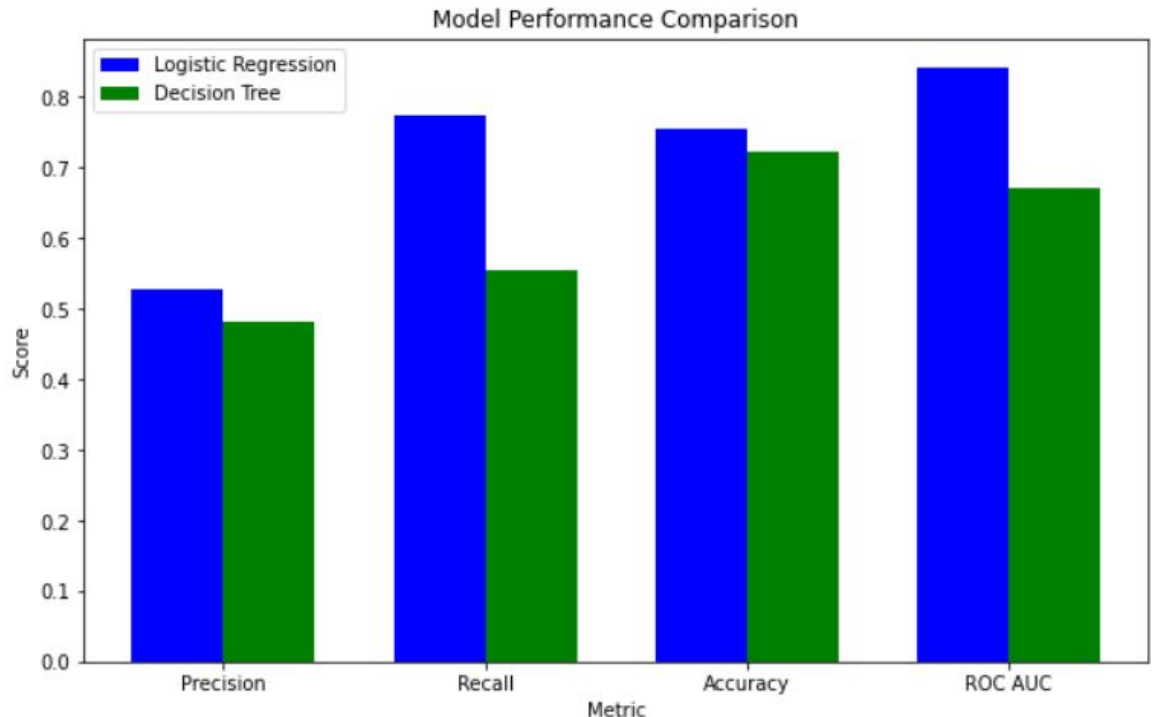
Predictor Analysis

- Two statistical models used for mixed type of data were compared, decision trees and logistic regression



Predictor Analysis

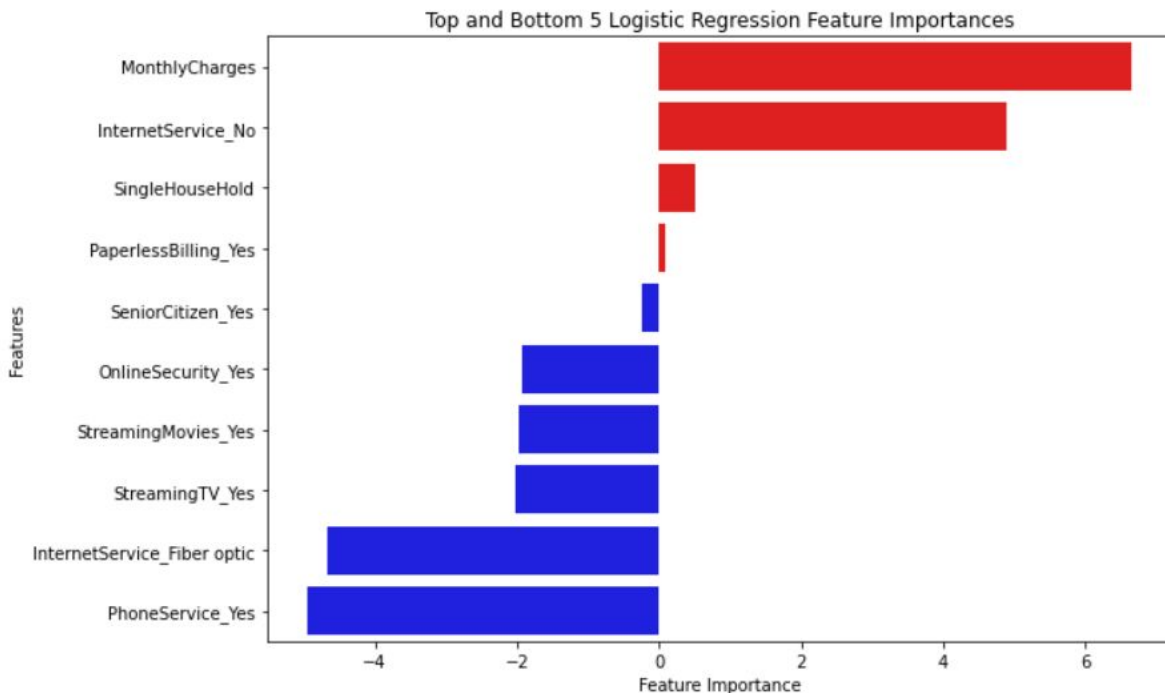
- Logistic Regression scores higher in all categories
- Recall, or churn cases detected is high at 80% of cases
- False positives is also high, around half of predictions false positives



Predictor Results

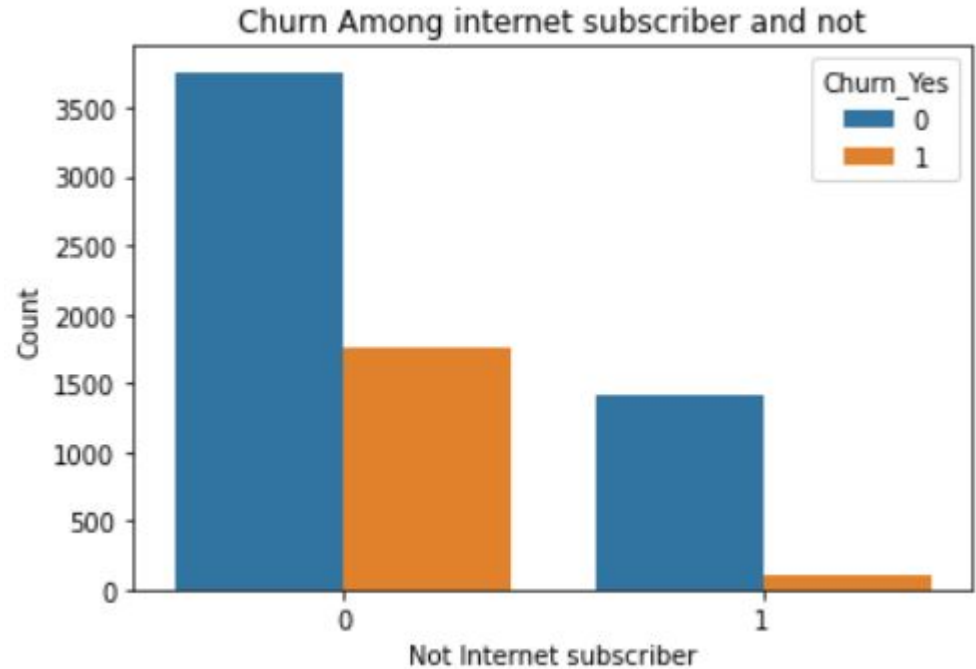
Red is correlated with **churning**, blue with **staying subscribed**

- Logistic regression identified no internet service as important factor not included in Pearson correlation



Predictor Results

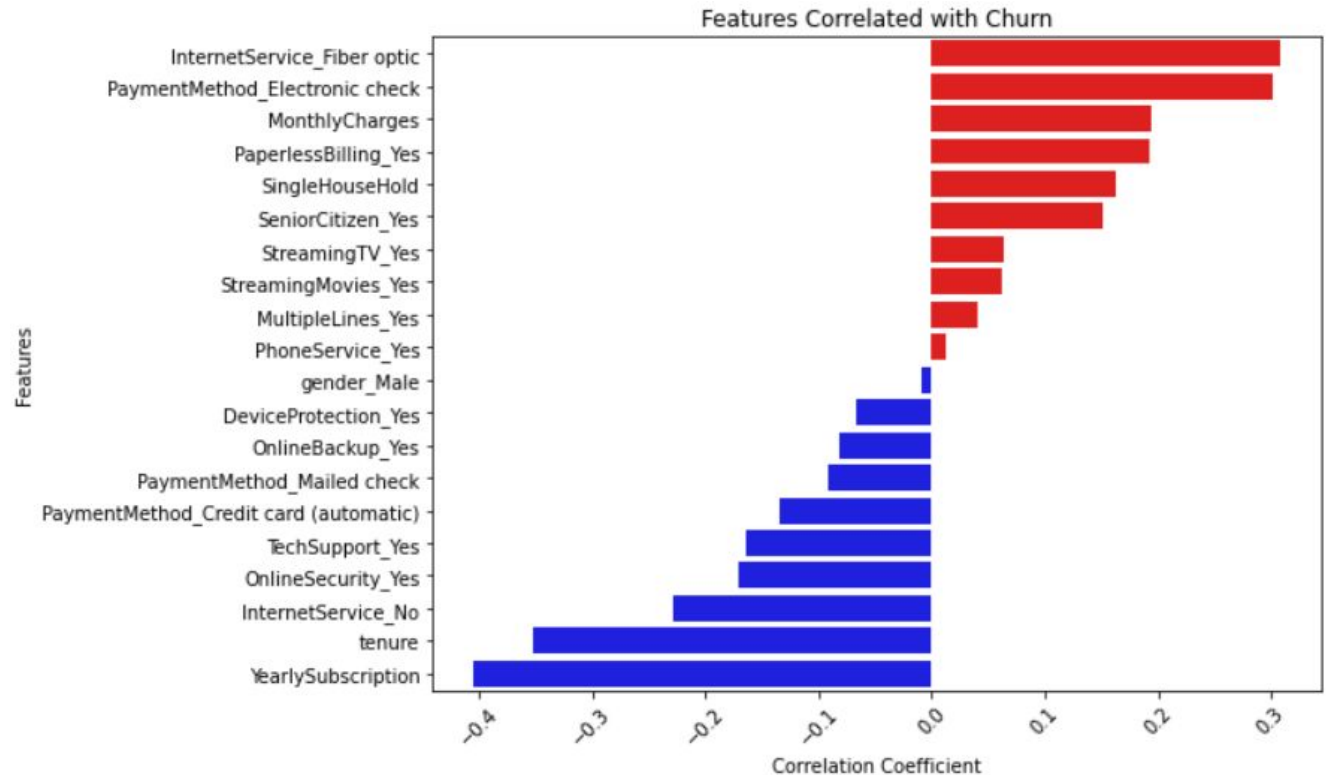
- Logistic regression identified phoneservice as important factor not included in pearson correlation, pattern not obvious



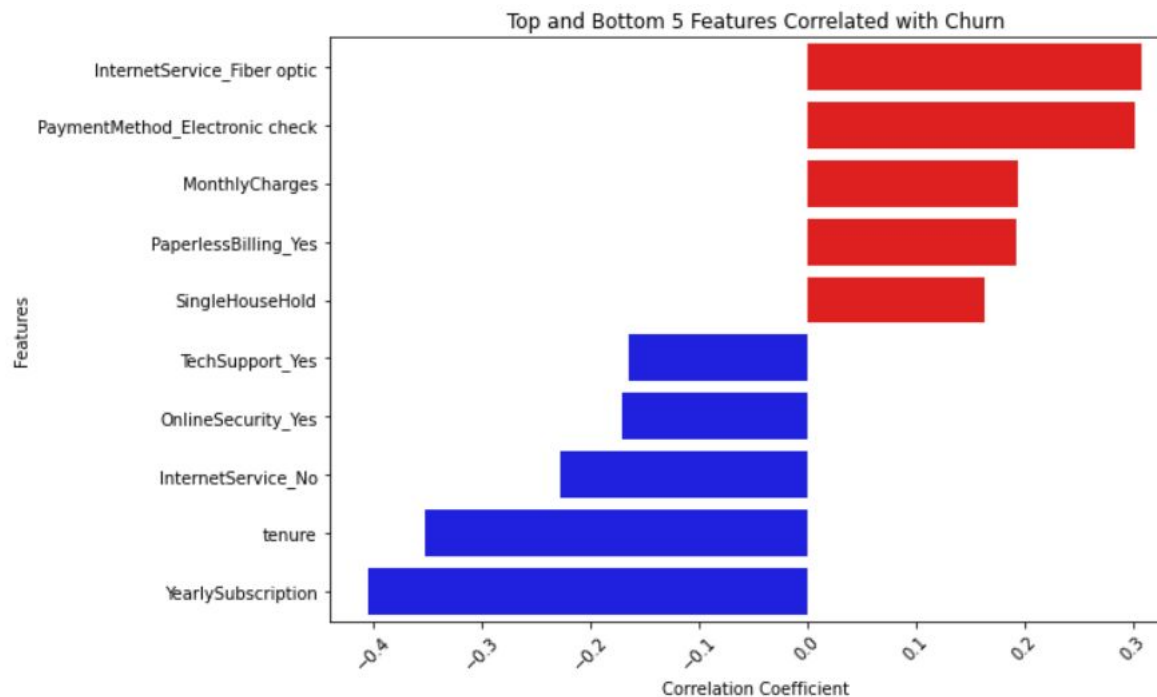
Conclusion

- Possible to predict most churn-events but with a lot of false positives
- Single households that use fiber optic internet service at high risk of churn
- Some factors need more investigation

Extra slides

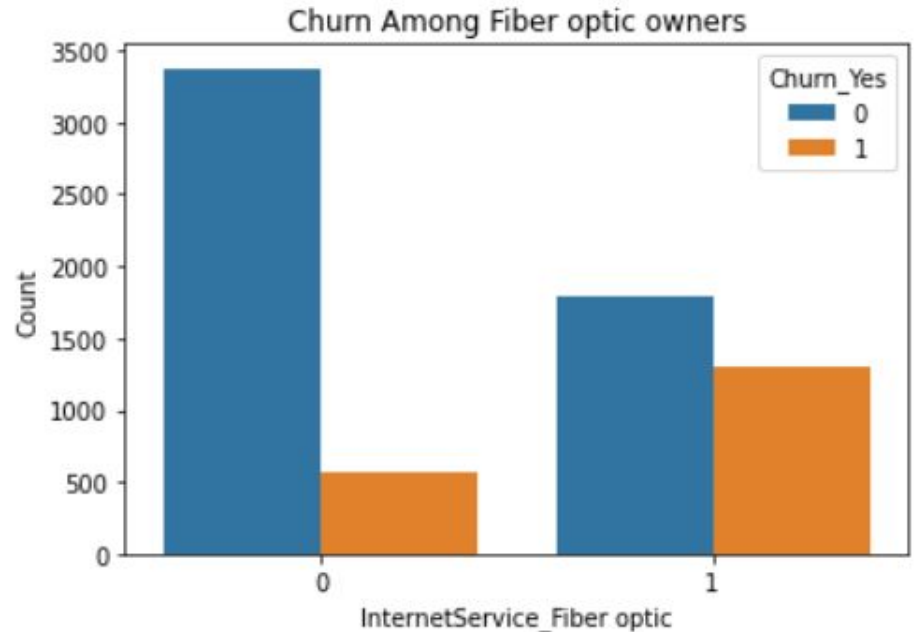
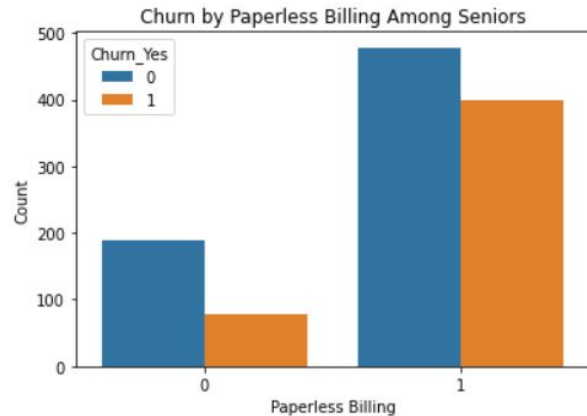


Extra slides



Extra slides

- Blue is correlated with churning, red with staying subscribed
- Yearly subscriptions a success
- InternetService_Fiber optic needs investigating



Most important factors at a glance

