

Capstone Project 1: Milestone Report

Problem Statement

Predict behavior to retain customers by using the data attributes. We are interested to know if the customer will churn or not and whether or not certain attributes contribute to it. We understand that this information will help the company catch potential customers earlier in the process and make an effort to retain them.

Data Wrangling

I looked at the shape and information stored in the data. I took a look at the data types and null values. On the whole this data was very clean and required minimal data cleaning steps. I took a look at the value counts of all the attributes and if there were any duplicates or anomalies but found none.

Looking at the data I saw that it contained customers which had 0 tenure i.e. they had just joined. I took the decision of removing such data because the total charges value was null and it is obviously not possible for them to churn when they just joined. I made a new dataset out of this filtered data for further investigation.

I have attached screenshots of some data cleaning steps.

```
[10]: #checks all the null values
print(data.isnull().sum())
```

```
customerID      0
gender          0
SeniorCitizen   0
Partner         0
Dependents      0
tenure          0
PhoneService    0
MultipleLines   0
InternetService 0
OnlineSecurity  0
OnlineBackup    0
DeviceProtection 0
TechSupport     0
StreamingTV     0
StreamingMovies 0
Contract        0
PaperlessBilling 0
PaymentMethod   0
MonthlyCharges  0
TotalCharges    0
Churn           0
```

```
# This is data where the customer has just joined and hasnt con
#These rows should not be included in the larger dataset
cust_new=data[data['TotalCharges']!=' ']
data=data[data['TotalCharges']==' ']
```

Data Story Findings

I see there are a few things that I would like to investigate. I see few variable that have a significant impact on the churn of a customer.

1. Payment Method that the user selects has some impact on this decision to churn
2. If the user has a month to month contract the user is more likely to leave and go to a competitor
3. Fiber optics internet has a higher churn rate than a DSL line due to competition.
4. Customers who opt for paperless billing have a higher chance of churning.
5. Customers who don't have a partner or dependents are more likely to churn than those who don't.

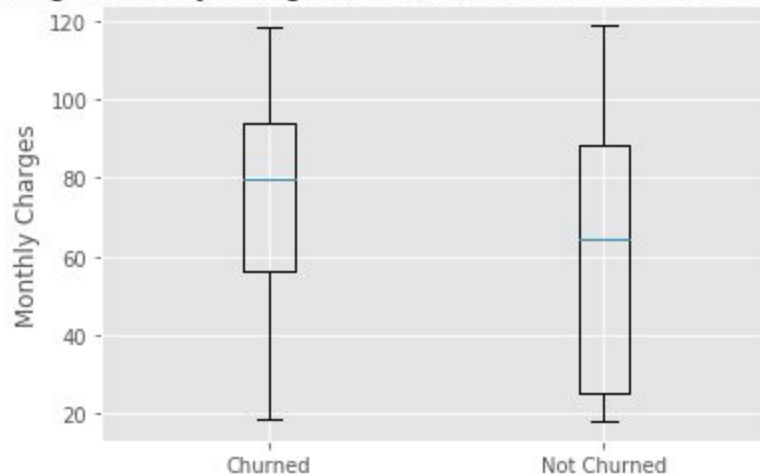
6. Customers who don't have tech support, online security , online backup services are more likely to churn.
7. The monthly charge has an effect on the churn
8. The tenure of a customer have a significant negative correlation

Each of these findings plays a part on the decision of the customer.

Inferential Report

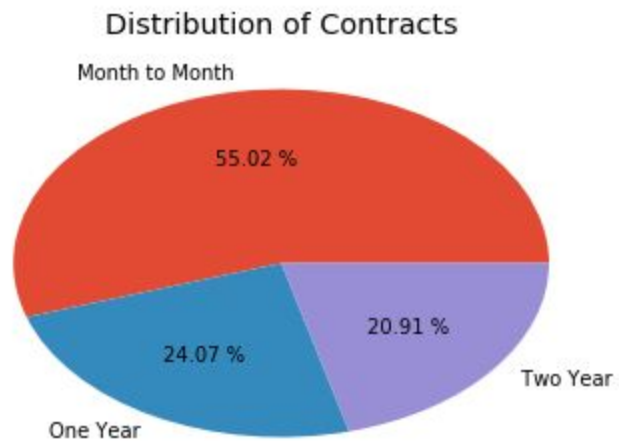
- Our Objective here is to find customer behavior to retain customers. The dataset provided gives us data about customers who churned within the last month, the services that each customer has signed up for, account details and demographic information as well.
- Using data exploration and inference we found certain patterns in the data that we feel will help us with predicting the people who churned. Each of those inferences are discussed further.
- Monthly Charges
 - As per our analysis for the customer who churned the average cost was 80 and the customer who did not churn the average cost was around 60. So If we have recently changed the price we should take this data into consideration and think about changing the price structure.

Average monthly charges for churned and Not churned customer



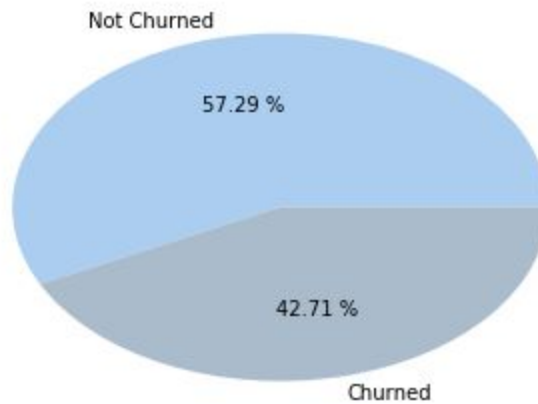
-
- Monthly Customer
 - When comparing monthly customers with our 1 year and 2 year customers. Monthly customers are highly likely to churn. This might be because they are just trying the service and don't like the commitment. 42% of our monthly customers have churned vs 11% of our 1 year customers and 3% of our 2 years customers.

There is definitely some strong correlation here. If the user has a month to month contract the user is more likely to leave and go to a competitor

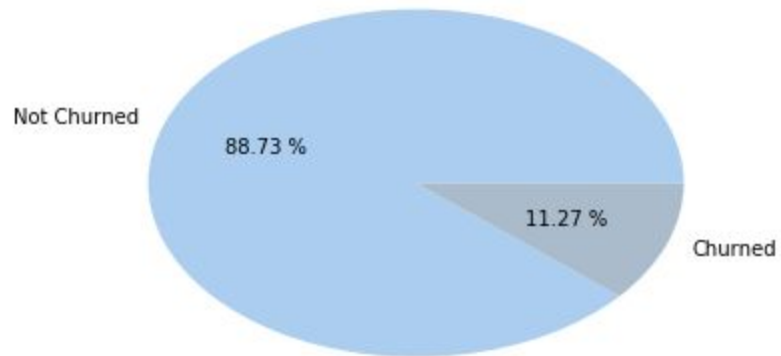


○

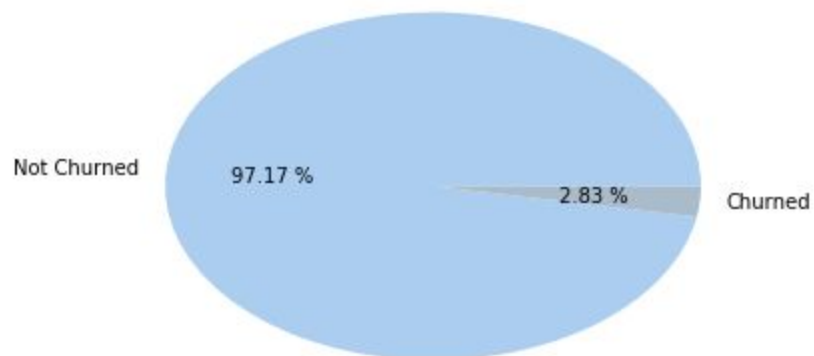
Monthly Customers Churned/Not Churned



One Year Customers Churned/Not Churned



Two Year Customers Churned/Not Churned



-
- Senior Citizen
 - There is a high correlation with a customer's likelihood to stay to the fact that he/she is a senior citizen. People who are senior citizen might be getting a better deal here and hence are likely to stay longer.

	SeniorCitizen	tenure	MonthlyCharges	Churn_1/0	PaperlessBilling_1/0	PhoneService_1/0
SeniorCitizen	1.000000	0.016567	0.220173	0.150889	0.156530	0.008576
tenure	0.016567	1.000000	0.247900	-0.352229	0.006152	0.008448
MonthlyCharges	0.220173	0.247900	1.000000	0.193356	0.352150	0.247398
Churn_1/0	0.150889	-0.352229	0.193356	1.000000	0.191825	0.011942
PaperlessBilling_1/0	0.156530	0.006152	0.352150	0.191825	1.000000	0.016505
PhoneService_1/0	0.008576	0.008448	0.247398	0.011942	0.016505	1.000000
Dependents_1/0	-0.211185	0.159712	-0.113890	-0.164221	-0.111377	-0.001762
Partner_1/0	0.016479	0.379697	0.096848	-0.150448	-0.014877	0.017706
gender_1/0	-0.001874	0.005106	-0.014569	-0.008612	-0.011754	-0.006488

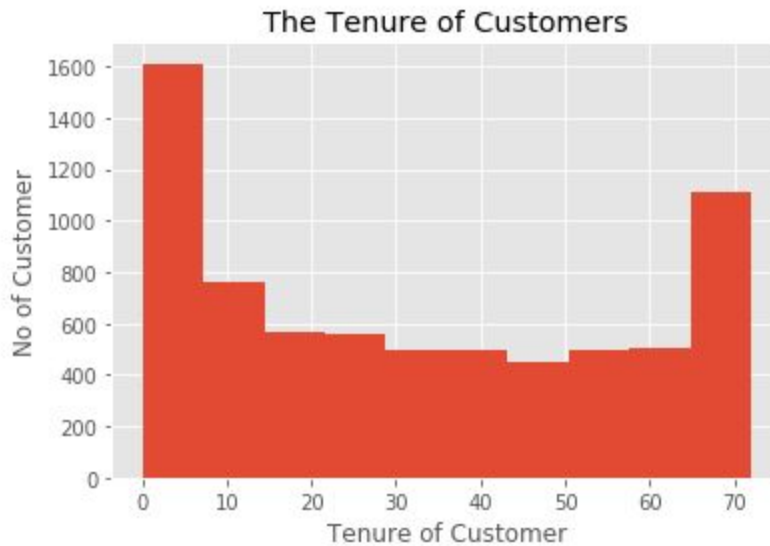
- Paperless Billing

- Having paperless billing has a higher likelihood of churning. This might hold true for a number of reasons such as more people read their email than the mails they get.

Churn	No	Yes
PaperlessBilling		
No	0.836699	0.163301
Yes	0.664349	0.335651
All	0.734630	0.265370

- Tenure

- The time the person has been loyal to the company has a high likelihood that he or she will continue that relationship. The .35 correlation suggests that it's a strong predictor of a person's ability to churn.



-
- Tech Support
 - If a person has not called tech support he or she is more likely to churn as compared to a person who has used it. Maybe that's the reason they churned in the first place. They found an issue they didn't call and directly churned. This leads us to the question as to how to reach such customers in time and promote our customer care support. Maybe use social media such as twitter to further increase our reach.
- Electronic Check
 - I assume echeck is not automated, hence people have to constantly do this manually and hence might sometime forget to do it. As .45 of the people who churned were doing it through electronics.

Churn	No	Yes
PaymentMethod		
Bank transfer (automatic)	0.832902	0.167098
Credit card (automatic)	0.847569	0.152431
Electronic check	0.547146	0.452854
Mailed check	0.808933	0.191067
All	0.734630	0.265370

-
- Fiber optics
 - People who have fiber optics are more likely to churn than people who use DSL. This might be due to the competition in Fiber optics and DSLs technology as well.

Churn	No	Yes
InternetService		
DSL	0.810409	0.189591
Fiber optic	0.581072	0.418928
No	0.925950	0.074050
All	0.734630	0.265370

-
- Partner or Dependents
 - People who do not have a partner or Dependents are more likely to churn and look for other products.
- Cross -selling
 - Customers who do not get any of our other services such as tech support, online security , online backup services are more likely to churn.