Telecom Users Project

Data Science Project

based on Communication Company Dataset

Overview

Any business wants to maximize the number of customers. To achieve this goal, it is important not only to try to attract new ones, but also to retain existing ones. Retaining a client will cost the company less than attracting a new one

The data contains information about almost six thousand users, their demographic characteristics, the services they use, the duration of using the operator's services, the method of payment, and the amount of payment.

X1	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity
1869	7010-BRBUU	Male	0	Yes	Yes	72	Yes	Yes	No	No internet service
4528	9688-YGXVR	Female	0	No	No	44	Yes	No	Fiber optic	No
6344	9286-DOJGF	Female	1	Yes	No	38	Yes	Yes	Fiber optic	No
6739	6994-KERXL	Male	0	No	No	4	Yes	No	DSL	No

OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
No internet service	No internet service	No internet service	No internet service	No internet service	Two year	No	Credit card (automatic)	24.10	1734.65	No
Yes	Yes	No	Yes	No	Month-to-month	Yes	Credit card (automatic)	88.15	3973.20	No
No	No	No	No	No	Month-to-month	Yes	Bank transfer (automatic)	74.95	2869.85	Yes
No	No	No	No	Yes	Month-to-month	Yes	Electronic check	55.90	238.50	No
No	Yes	No	No	No	Month-to-month	No	Electronic check	53.45	119.50	No

Rows : 5986

Columns : 22

	X1	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService
1	1869	7010-BRBUU	Male	0	Yes	Yes	72	Yes
2	4528	9688-YGXVR	Female	0	No	No	44	Yes

```
telecom <- read_csv("C:/Users/User/Downloads/telecom_users.csv")

View(telecom)

ctelecom <- select(telecom,-1)

View(ctelecom)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService
1	7010-BRBUU	Male	0	Yes	Yes	72	Yes
2	9688-YGXVR	Female	0	No	No	44	Yes

```
> dim(ctelecom)
[1] 5986   21
```

```
> summary(ctelecom$MonthlyCharges)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.25   35.65   70.40   64.80   89.90  118.75
```

Data Wrangling

1. Import The Dataset
2. Check The Dataset for Cleaning
3. Clean The Dataset
4. Remove (unnecessarily | Null) Values
5. View The Dataset
6. Get The Dataset Stat Information
7. Get The Dataset Summarization

Overview

filtering

```
seniorspills1 <- filter(ctelecom,SeniorCitizen==1,PaperlessBilling == "No" )
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod
1	7173-TETGO	Female	1	Yes	No	72	Yes	Yes	Fiber optic	Yes	No	No	No	No	No	Two year	No	Bank transfer (automatic)
2	4193-ORFCL	Female	1	No	No	1	Yes	No	DSL	No	No	No	No	No	No	Month-to-month	No	Mailed check
3	0529-ONKER	Male	1	No	No	15	Yes	Yes	Fiber optic	No	No	No	No	No	No	Month-to-month	No	Electronic check
4	2277-VWCNI	Female	1	No	No	4	Yes	Yes	DSL	No	No	No	No	No	No	Month-to-month	No	Mailed check
5	8075-GXIUB	Male	1	Yes	No	9	Yes	No	No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	One year	No	Credit card (automatic)
6	9274-CNFMO	Male	1	Yes	No	4	Yes	Yes	Fiber optic	No	No	No	No	No	No	Month-to-month	No	Electronic check
7	8250-ZNGGW	Female	1	No	No	5	Yes	No	DSL	Yes	Yes	No	Yes	No	No	Month-to-month	No	Credit card (automatic)

```
cp1 <- round(nrow(seniorspills1)/nrow(ctelecom),digits = 2)*100
```

```
> cp1
[1] 4
```

percentage

```
pay_methods <- levels(type.convert(ctelecom$PaymentMethod))
```

```
> pay_methods
[1] "Bank transfer (automati
c)" "Credit card (automatic)"
   "Electronic check"
  "Mailed check"
```

levels

mutate

```
total_monthly <- ctelecom %>% select(TotalCharges,MonthlyCharges)
yearly <- yearly_charges %>% select(yearly_charges)
periods <- months_period %>% select(months_period)

financial_Data <- data.frame(total_monthly,yearly,periods)
```

```
yearly_charges <- mutate(ctelecom,yearly_charges = MonthlyCharges *12 )
```

pipelines

	TotalCharges	MonthlyCharges	yearly_charges	months_period
1	1734.65	24.10	289.2	72
2	3973.20	88.15	1057.8	44
3	2869.85	74.95	899.4	38
4	238.50	55.90	670.8	4
5	119.50	53.45	641.4	2

```
months_period <- transmute(ctelecom,months_period = tenure )
```

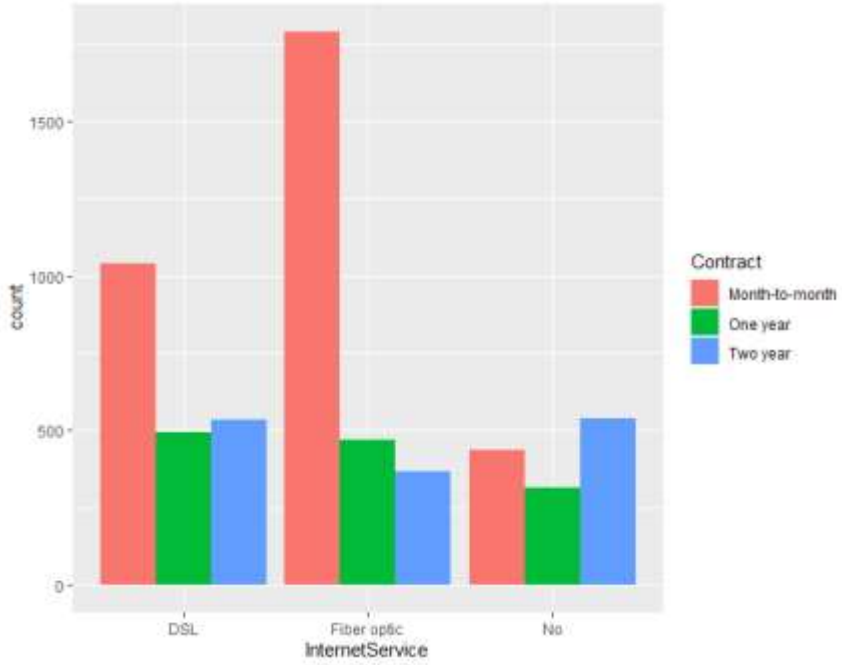
Financial Data Frame

transmute

Internet Services & Contract Types

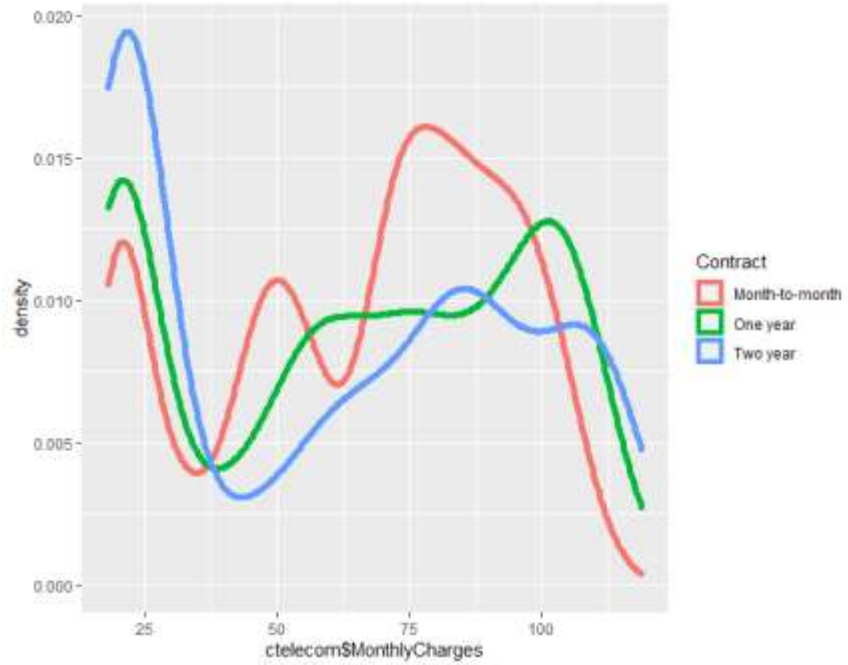
Insight

As shows in the chart, the most customers that use Fiber optic their contract is month to month and the most customers use DSL have One year contract but The customers with two years contract most of them they don't have internet service.



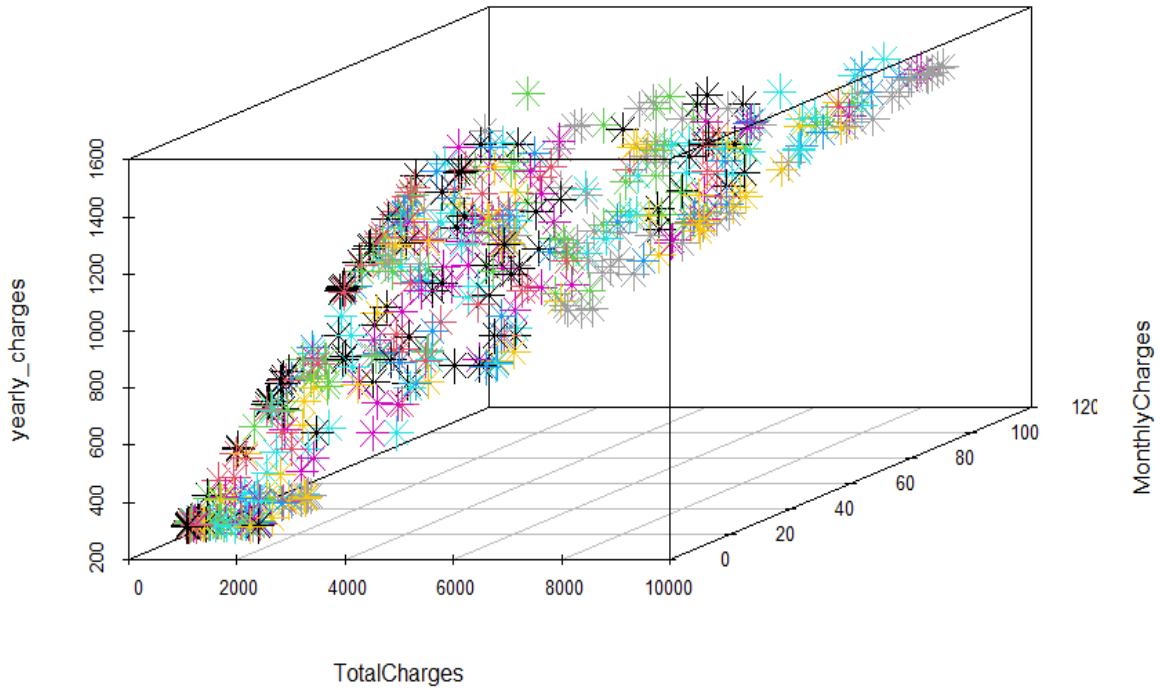
Insight

As shows in the chart, the less Monthly Charges Contract is the Two-year Contract and the expensive one is the month to month, the one-year contract in the middle between them.

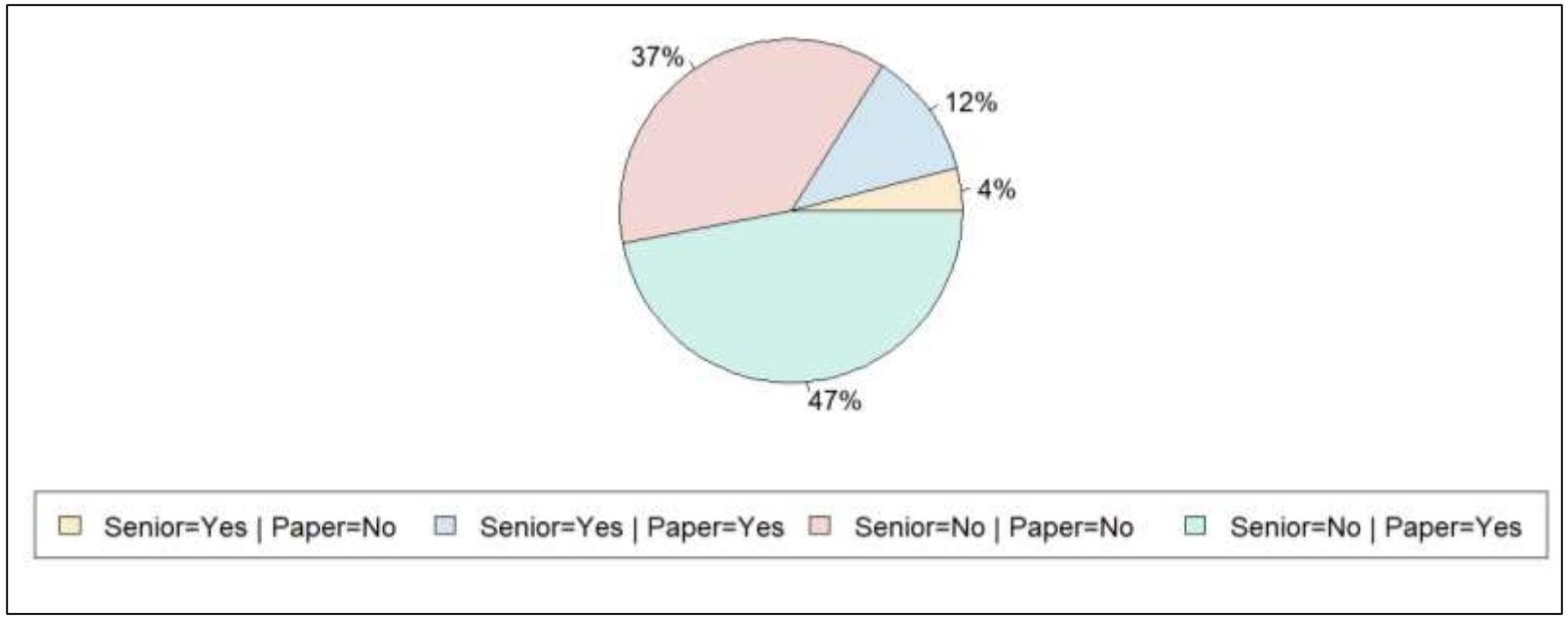


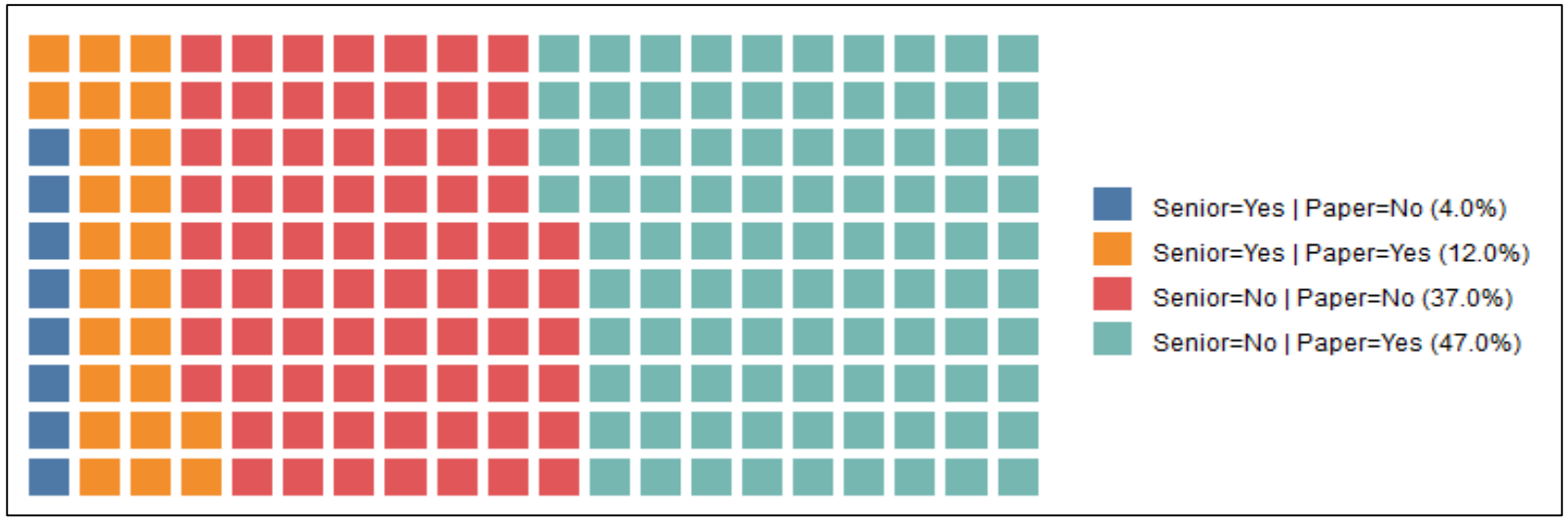
Total Charges

```
samp <- financial_Data[sample(nrow(financial_Data),500),]
scatterplot3d(samp, pch = 8)
```



-Strong -Positive -Linear





Insight

As shows in the charts, the biggest percentage are the young people who they still using the papers bills , and the smallest percentage are the old people who they don't use papers bills.

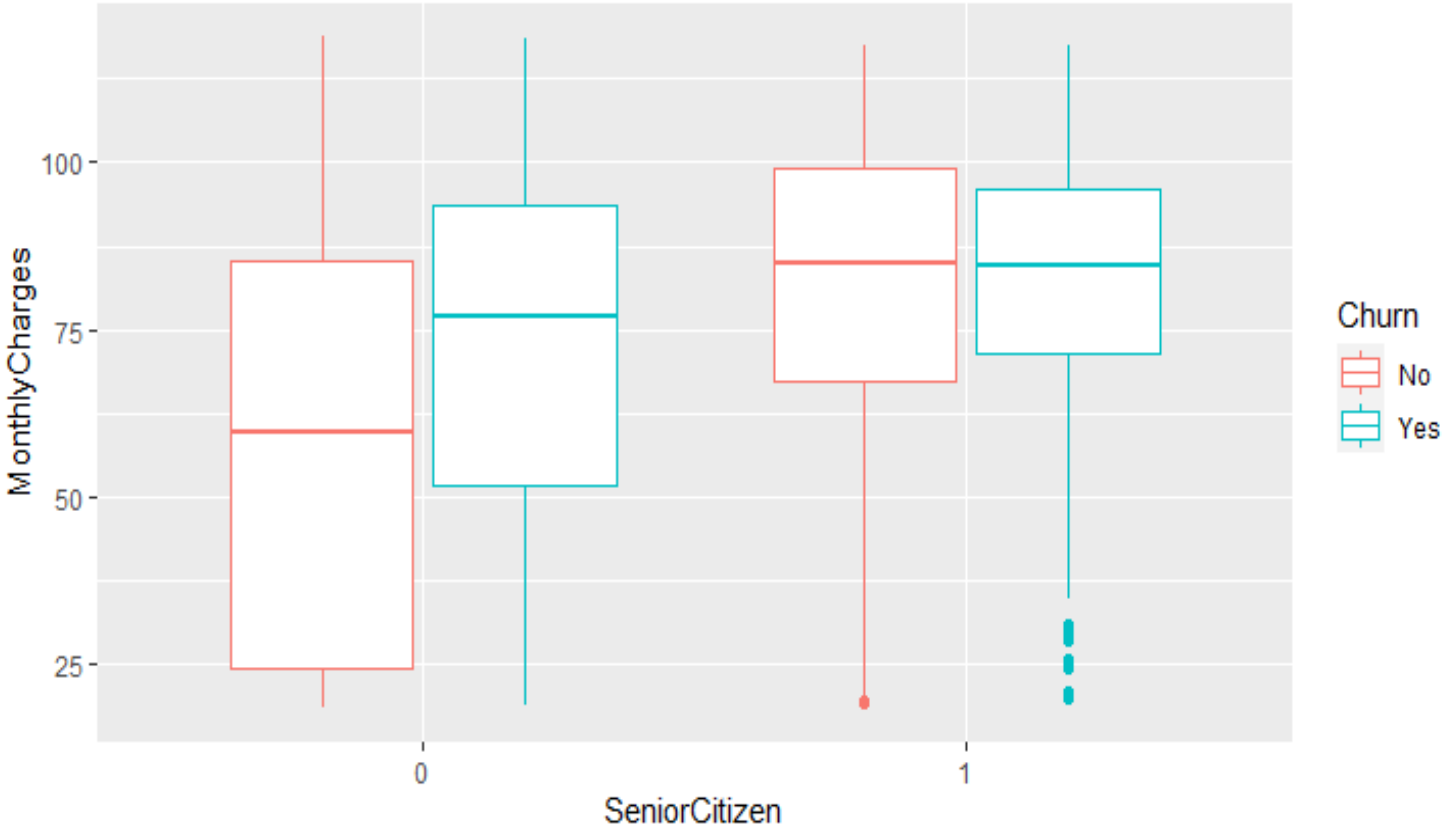
Explore Data Analysis

```
> str(df)
'data.frame':   5986 obs. of  22 variables:
 $ X               : int  1869 4528 6344 6739 432 2215 5260 6001 1480 5137 ...
 $ customerID      : chr  "7010-BRBUU" "9688-YGXVR" "9286-DOJGF" "6994-KERXL" ...
 $ gender          : chr  "Male" "Female" "Female" "Male" ...
 $ SeniorCitizen   : int  0 0 1 0 0 0 0 0 1 ...
 $ Partner         : chr  "Yes" "No" "Yes" "No" ...
 $ Dependents      : chr  "Yes" "No" "No" "No" ...
 $ tenure          : int  72 44 38 4 2 70 33 1 39 55 ...
 $ PhoneService    : chr  "Yes" "Yes" "Yes" "Yes" ...
 $ MultipleLines   : chr  "Yes" "No" "Yes" "No" ...
 $ InternetService : chr  "No" "Fiber optic" "Fiber optic" "DSL" ...
 $ OnlineSecurity  : chr  "No internet service" "No" "No" "No" ...
 $ OnlineBackup    : chr  "No internet service" "Yes" "No" "No" ...
 $ DeviceProtection: chr  "No internet service" "Yes" "No" "No" ...
 $ TechSupport     : chr  "No internet service" "No" "No" "No" ...
 $ StreamingTV     : chr  "No internet service" "Yes" "No" "No" ...
 $ StreamingMovies : chr  "No internet service" "No" "No" "Yes" ...
 $ Contract        : chr  "Two year" "Month-to-month" "Month-to-month" "Month-to-month" ...
 $ PaperlessBilling: chr  "No" "Yes" "Yes" "Yes" ...
 $ PaymentMethod   : chr  "Credit card (automatic)" "Credit card (automatic)" "Bank transfer (automatic)" "Electronic check" ...
 $ MonthlyCharges  : num  24.1 88.2 75 55.9 53.5 ...
 $ TotalCharges    : num  1735 3973 2870 238 120 ...
 $ Churn           : chr  "No" "No" "Yes" "No" ...
```

```r
# Change data type
df$SeniorCitizen <- sapply(df$SeniorCitizen, as.character)
```

How the Monthly charges distribution looks for Senior Citizen and Non-Senior Citizen? And how affect on Churn*?

```
df %>%
  ggplot(aes(y= MonthlyCharges, x= SeniorCitizen, color = Churn)) +
  geom_boxplot()
```



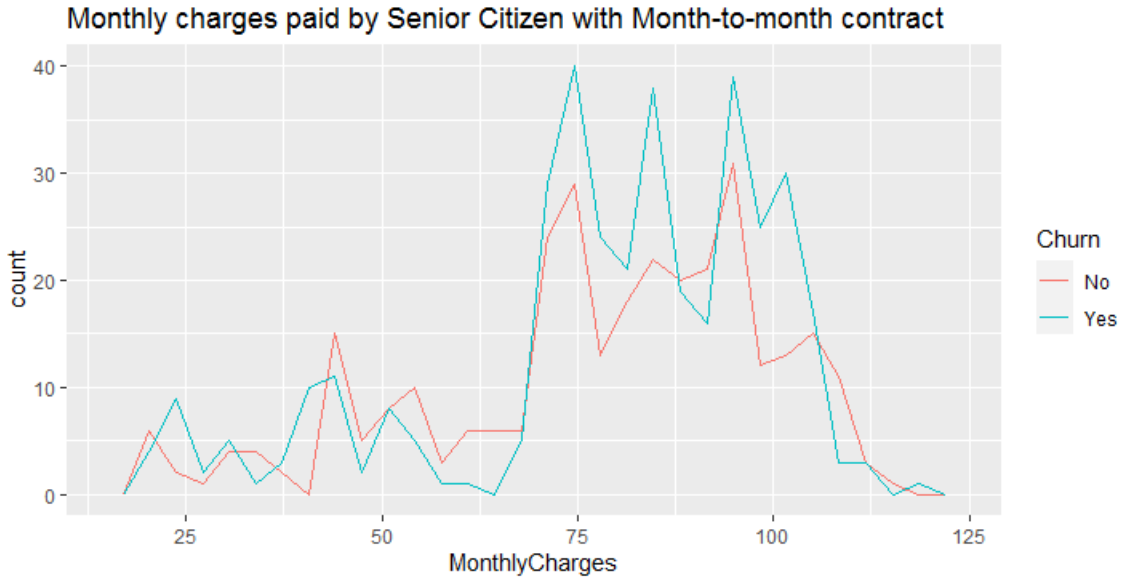
Insight

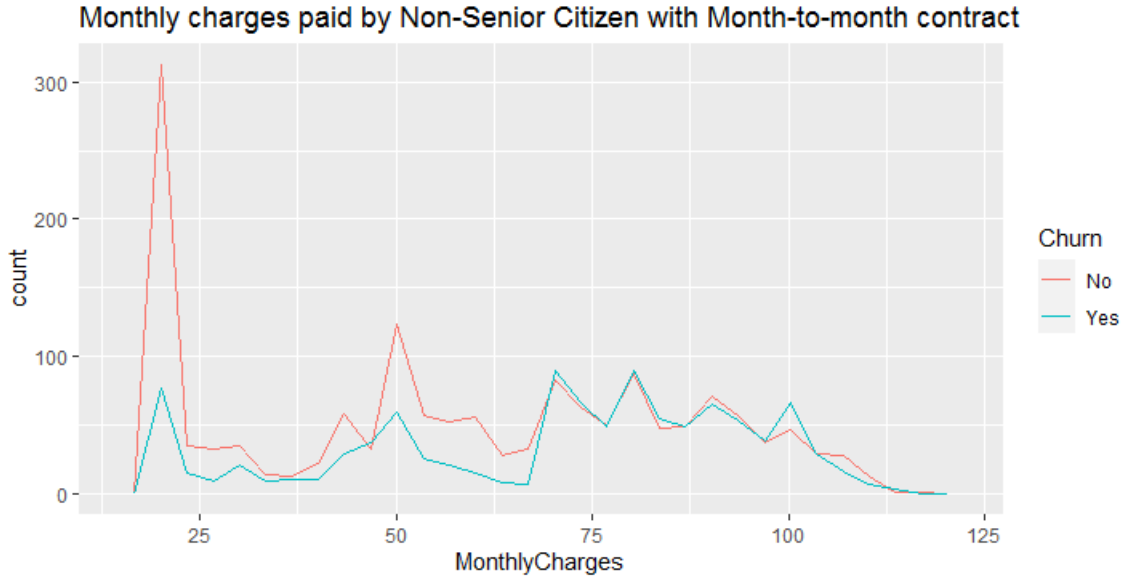
As shows in the chart, the Churn of Senior Citizen users is almost similar for the Monthly Charges. Churn of Non-Senior Citizen users is highest when Monthly Charges is highest.

What is the most Monthly charges paid by customers with Month-to-month contract?

```
df %>%
  filter(SeniorCitizen == "1" & Contract == "Month-to-month") %>%
  ggplot(aes(MonthlyCharges, color = Churn)) +
  ggtitle("Counts of monthly charges paid by customers based on churn Senior Citizen.") +
  geom_freqpoly()

df %>%
  filter(SeniorCitizen == "0" & Contract == "Month-to-month") %>%
  ggplot(aes(MonthlyCharges, color = Churn)) +
  ggtitle("Counts of monthly charges paid by customers based on churn Non-Senior Citizen") +
  geom_freqpoly()
```

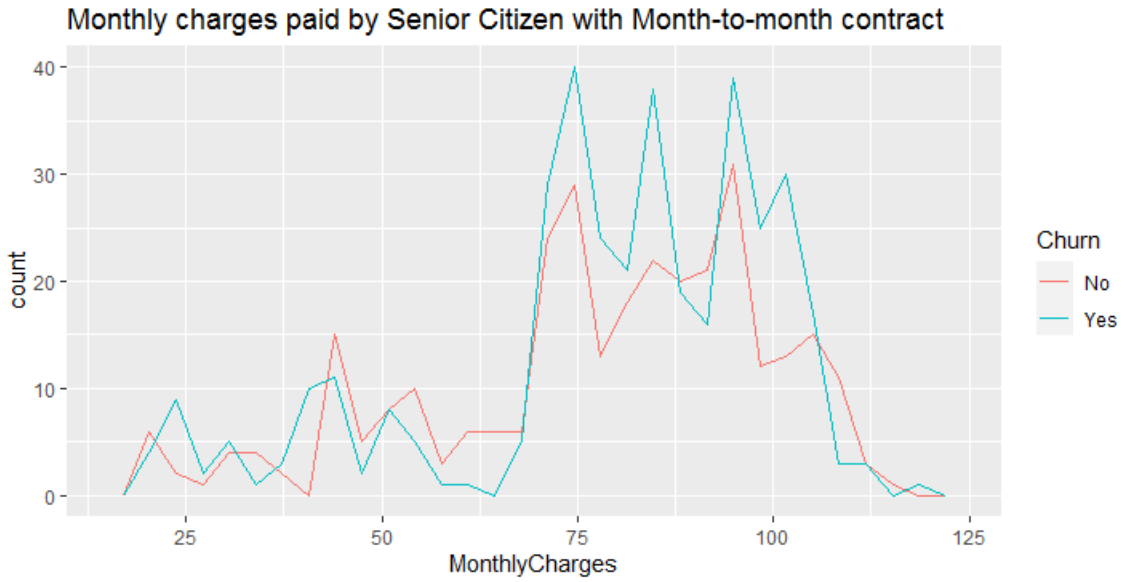


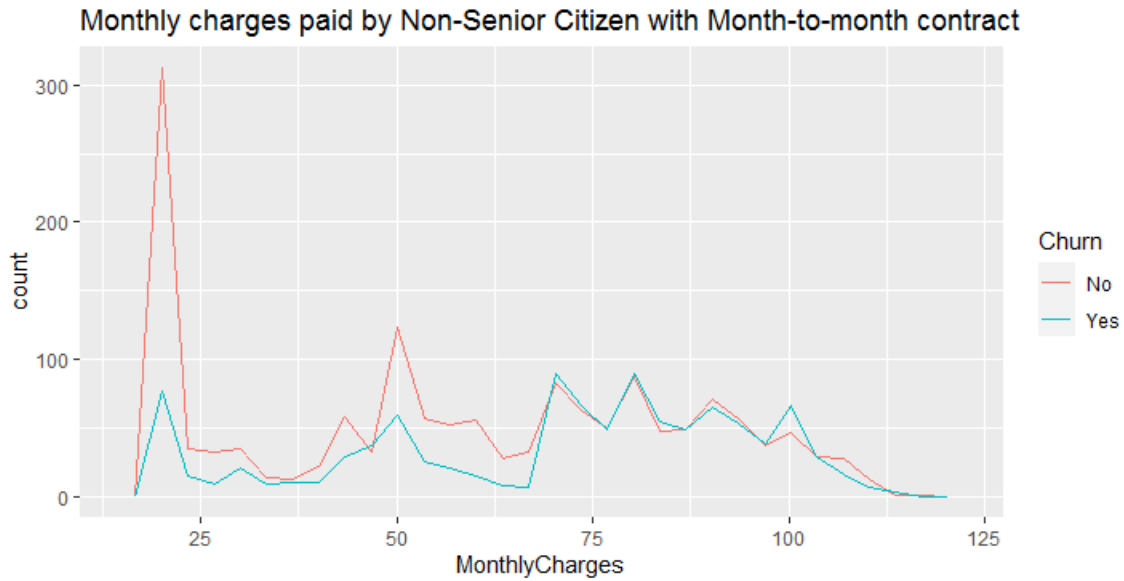


What is the most Monthly charges paid by customers with Month-to-month contract?

Insight

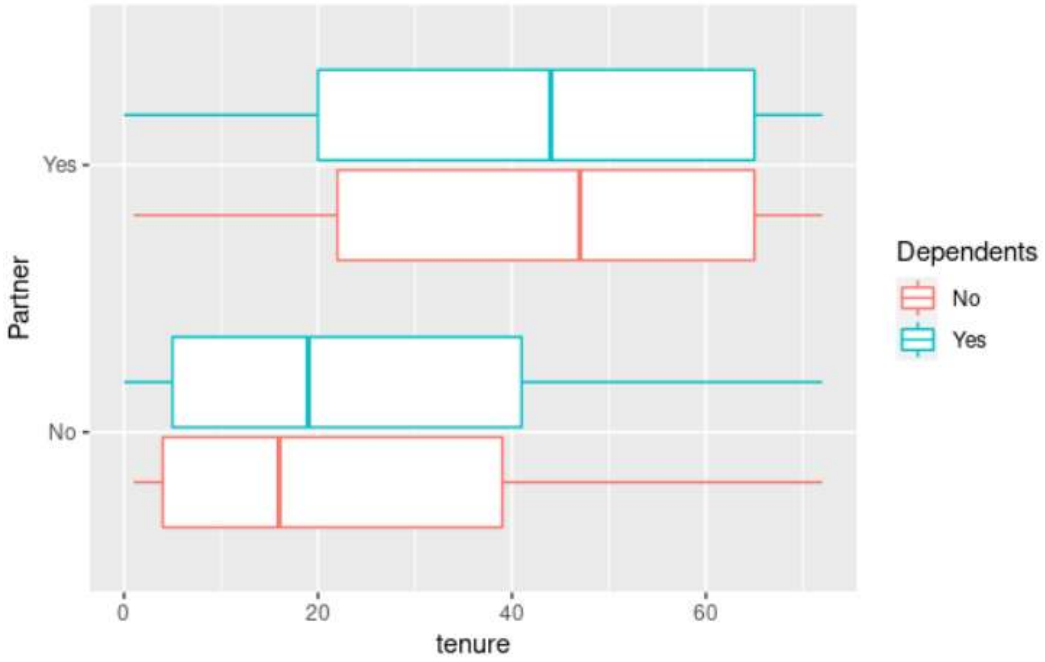
Most of Senior Citizen paid between 70 and 105 per month. For Non-Senior Citizen there is big peak when Monthly Charges less than 25. As charts illustrate, Non-Senior Citizen paid less and not churn the service. On otherwise, monthly charges for Senior Citizen a bit similar for both churn or not.





What's the period tenure for partners and are they dependents or independent?

```
ggplot(data = df,
mapping = aes(x = tenure, y = Partner, color =
Dependents)) + geom_boxplot()
```

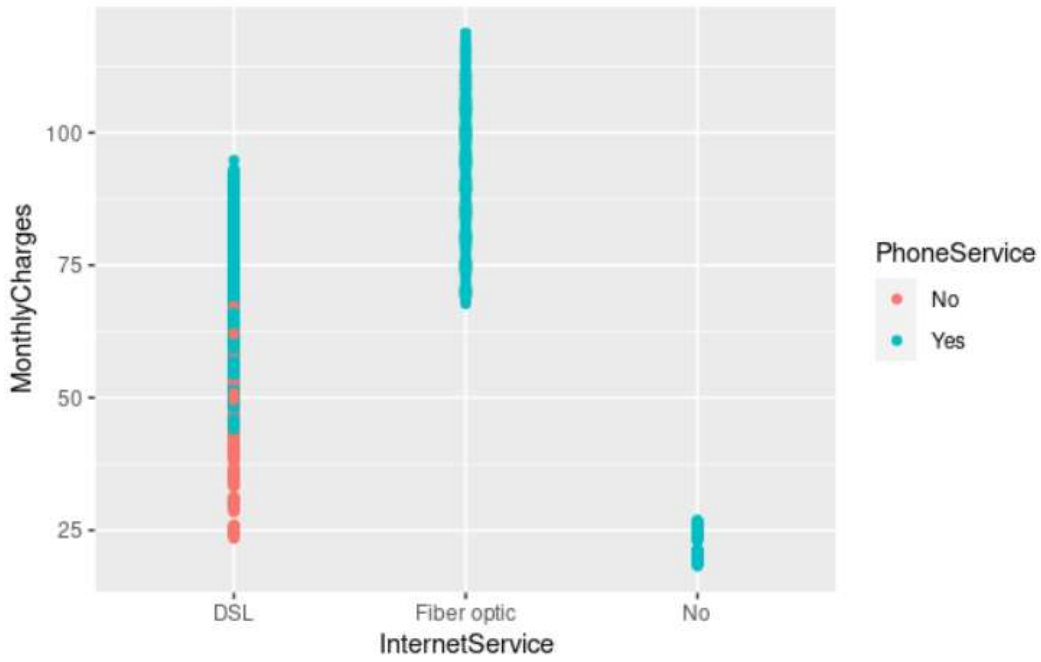


Insight

The period tenure for partners was between 0-75 and most partners are dependents

who have a monthly charges for internet service, have they also a phone service?

```
ggplot(data = df) +
geom_point(mapping = aes(x =InternetService,
y = MonthlyCharges, color = PhoneService))
```

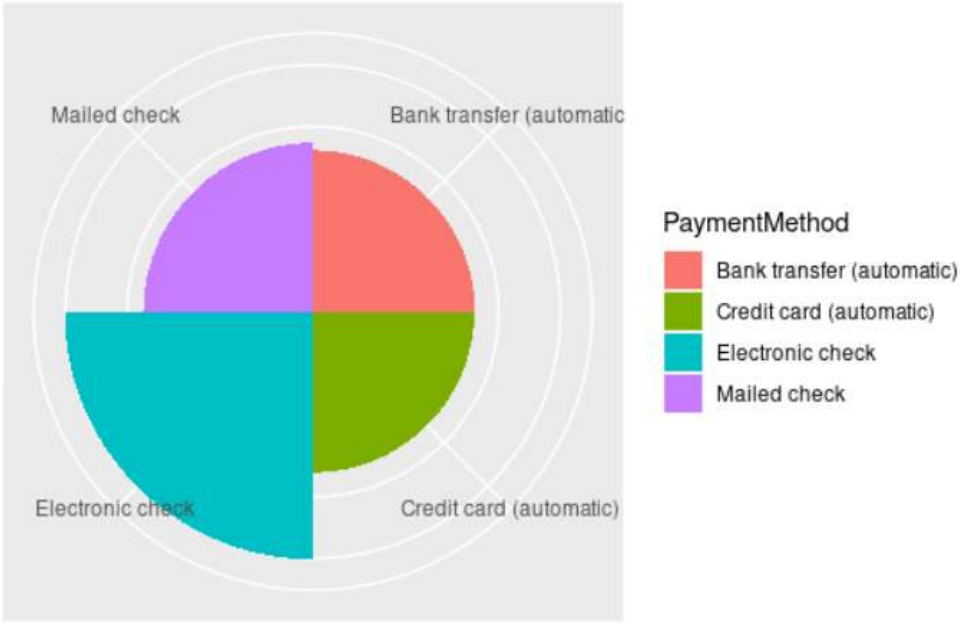


Insight

We can say yes, Since most of them have internet service and phone service, except who have a DSL internet service, we see in there are some of them haven't a phone service.

What the most used of Payment Method?

```
bar = ggplot(data = df) + geom_bar(mapping = aes(x = PaymentMethod, fill = PaymentMethod), show.legend = TRUE, width = 1) + theme(aspect.ratio = 1) +labs(x = NULL, y = NULL)bar + coord_polar()
```

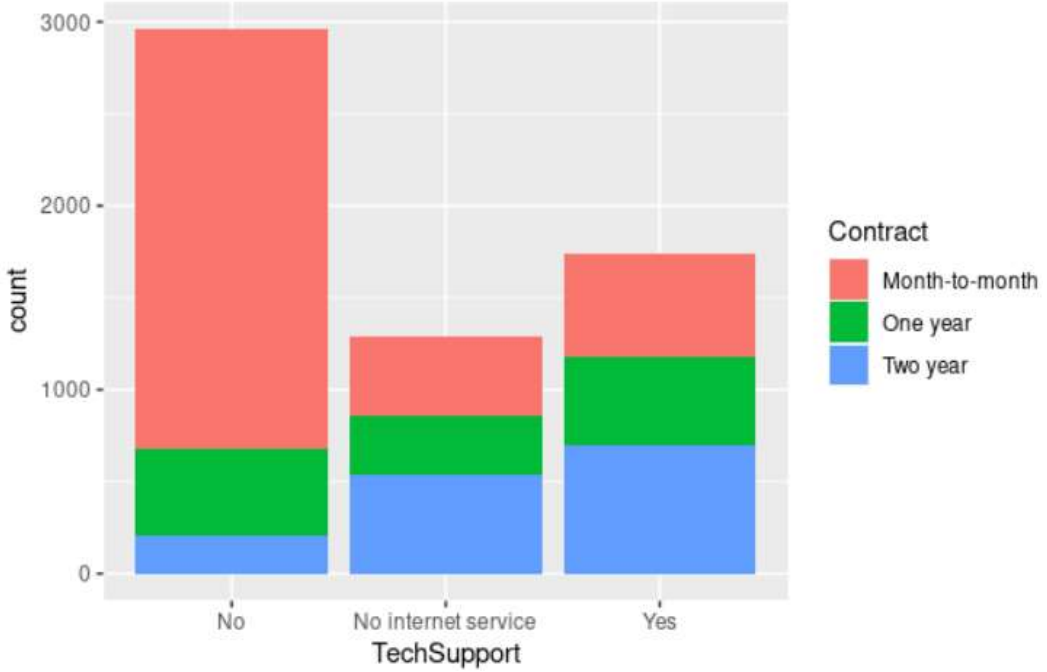


Insight

As shows in the chart, The most used of Payment Method is Electronic check

What is the most used technical support contract?

```
ggplot(data = df) + geom_bar(mapping = aes(x = TechSupport, fill = Contract))
```

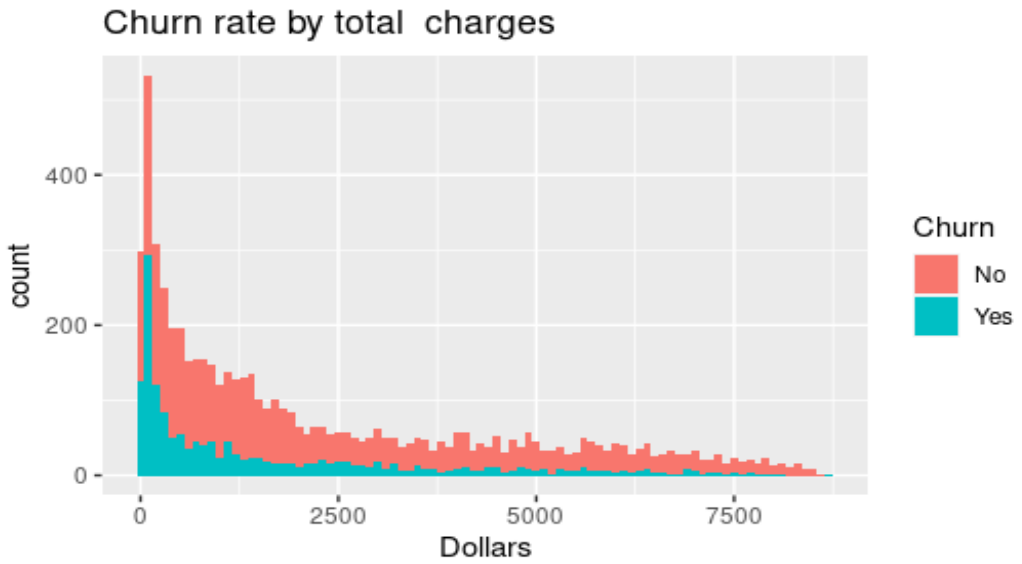


Insight

As shown in the chart, the most used technical support contracts are Two year

How does the total charges paid affect the users leave or not?

```
df%>%  ggplot(aes(x =TotalCharges , fill = Churn)) +
geom_histogram(binwidth = 100) +
labs(x = "Dollars", title = "Churn rate by total
charges")
df
```

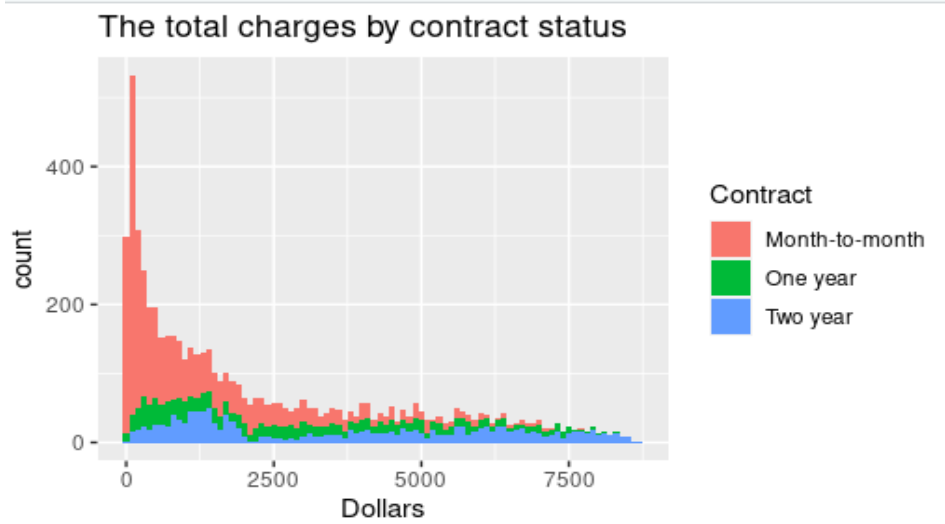


Insight

As shows in the chart, customers who have spent more with the company less likely to churn. That's mean customers who are more financially well off are less likely to leave.

How about the distribution for total charges split by contract status?

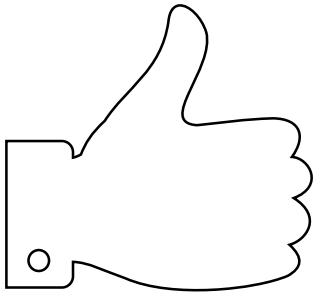
```
df%>%  ggplot(aes(x =TotalCharges , fill =Contract ))
+  geom_histogram(binwidth = 100) +  labs(x =
"Dollars",title = "The total charges by contract
status ")
df
```



Insight

As would be expected, the total charges of month-to-month contract customers is much higher than the longer contract customers. There is a large spike at 1 month, indicating that there are a large portion of customers that will leave the after just one month of service.

Thank You



Any questions?

Full code available here : The Code Source