Telecom Customers Churn

Final project

Aya Sabry

3/8/23

Supervised by: Eng. Asmaa

Table of Contents

Validate and clean data.

- The dataset consists of **47 Features** and **7043 records.** Note: Features are referred to as Variables or Columns, whereas Records are referred to as Instances or Rows.
- Go over the provided data for **duplicate features.**
- We notice duplicate **Customer ID in *customer satus sheet* and remove it.**

Customer ID	Quarter	Satisfaction Score	Customer Status	Churn Label	Churn Counter	Churn Score	CLTV	Churn Category	C
3606-TWKGI	Q3	1	Churned	Yes	1	84	2494	Competitor	C
4385-GZQXV	Q3	1	Churned	Yes	1	81	3810	Competitor	C
3488-PGMQJ	Q3	2	Churned	Yes	1	80	4345	Dissatisfaction	P

- Consider the data's **numerical and categorical Features** to determine whether formatting is necessary. We did the following processes:
    - Customer id -→ from General to text
    - Zip code→ text
    - Total Charges, Total Refunds, Total Extra Data Charges, Total Long-Distance Charges, Total Revenue, Avg Monthly Long-Distance Charges, Monthly Charge ---->Number

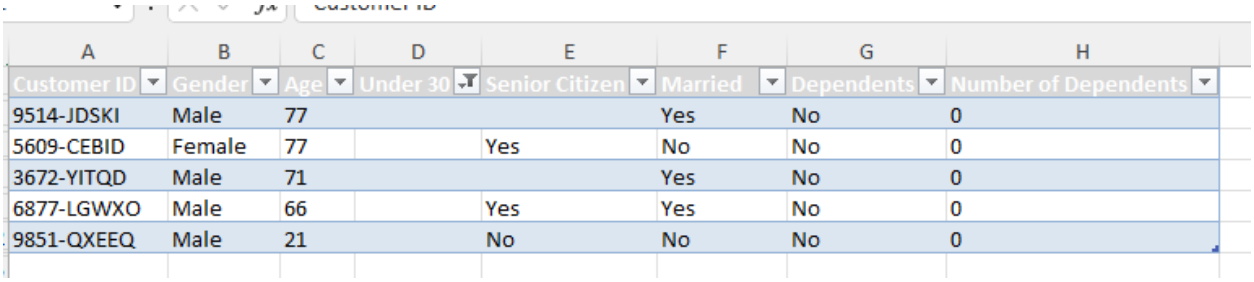
- **Check Blank Cells**

Customer ID	Quarter	Satisfaction Score	Customer Status	Churn Label	Churn Score	CLTV	Churn Category	Churn Reason
0265-EDXBD	Q3	2	Churned		69	4105	Attitude	Attitude of support person
4086-YQSNZ	Q3	2	Churned		89	3834	Attitude	Attitude of service provider
7216-EWTRS	Q3	1	Churned		65	3128	Attitude	Attitude of service provider
2504-DSHIH	Q3	3	Stayed		38	5788		
9705-IOVQQ	Q3	5	Stayed		20	5177		

    - *In Customer Status*
        - We Fill blank cells with "Yes" if Customer status = “Churned” else “No”

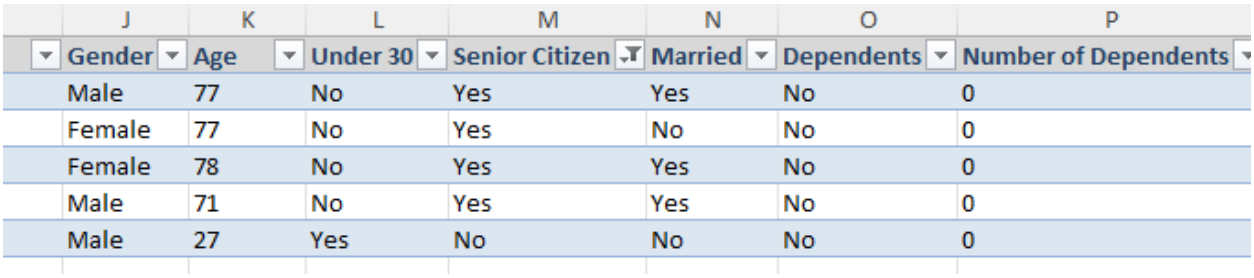
- *In Customer demographic*

  - Blank Cells in Senior Citizen and Under 30 Columns.



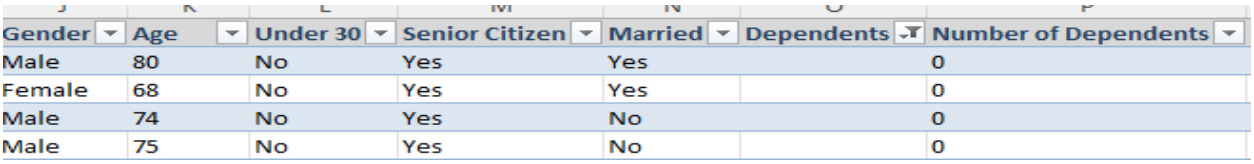
	A	B	C	D	E	F	G	H
	Customer ID	Gender	Age	Under 30	Senior Citizen	Married	Dependents	Number of Dependents
	9514-JDSKI	Male	77			Yes	No	0
	5609-CEBID	Female	77		Yes	No	No	0
	3672-YITQD	Male	71			Yes	No	0
	6877-LGWXO	Male	66		Yes	Yes	No	0
	9851-QXEEQ	Male	21		No	No	No	0

  - We Fill it according To Age Columns if Age < 30 or not and senior Citizen if >70 or not.



	J	K	L	M	N	O	P
	Gender	Age	Under 30	Senior Citizen	Married	Dependents	Number of Dependents
	Male	77	No	Yes	Yes	No	0
	Female	77	No	Yes	No	No	0
	Female	78	No	Yes	Yes	No	0
	Male	71	No	Yes	Yes	No	0
	Male	27	Yes	No	No	No	0

  - Also, we noticed Blank cells with Dependents Columns I choose to fill it with No as it is the Mode value with this column.



Gender	Age	Under 30	Senior Citizen	Married	Dependents	Number of Dependents
Male	80	No	Yes	Yes		0
Female	68	No	Yes	Yes		0
Male	74	No	Yes	No		0
Male	75	No	Yes	No		0

K	L	M	N	O	P
Age	Under 30	Senior Citizen	Married	Dependents	Number of Dependents
80	No	Yes	Yes	No	0
68	No	Yes	Yes	No	0
74	No	Yes	No	No	0
75	No	Yes	No	No	0

In customer Service sheet

- We noticed Blank Cells in Referred a Friend Column and Choose to fill it with “No” as it was the mode value of that column.

A	B	C	D	E	F	G	H
Customer ID	Quarter	Referred a Friend	Number of Referrals	Tenure in Months	Offer	Phone Service	Avg Monthly Long Distance Charges
3488-PGMQJ	Q3		0	8	None	Yes	1.78
0533-BNWKF	Q3		2	55	Offer B	Yes	42.82
6235-VDHOM	Q3		0	5	Offer E	No	0
7657-DYEPJ	Q3		0	38	Offer C	Yes	5.75
3871-IKPYH	Q3		0	1	None	Yes	16.19

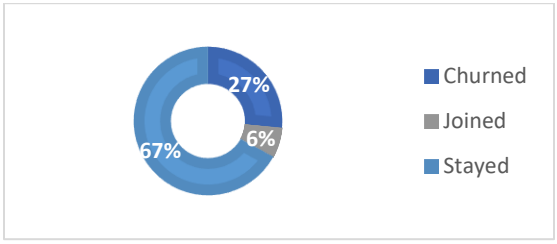
Manipulate data and communicate your insights.

1-*In Customer service sheet* we introduce "**Tenure in years"** column

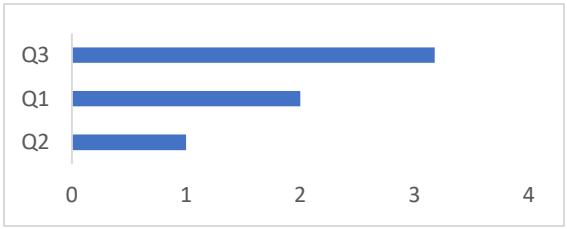
=ROUNDUP(W2/12,0)

My Insights from data

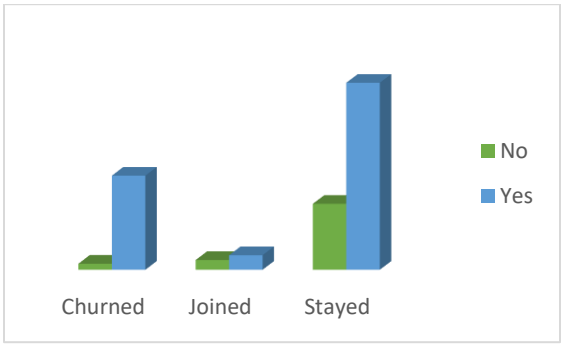
1-Customer Status



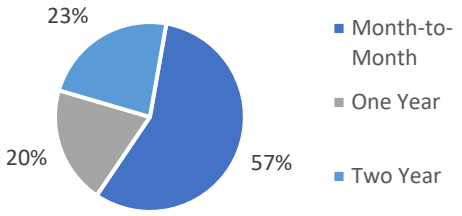
2 'Quarter': Q3 has noticeably higher 'Tenure in Years'.



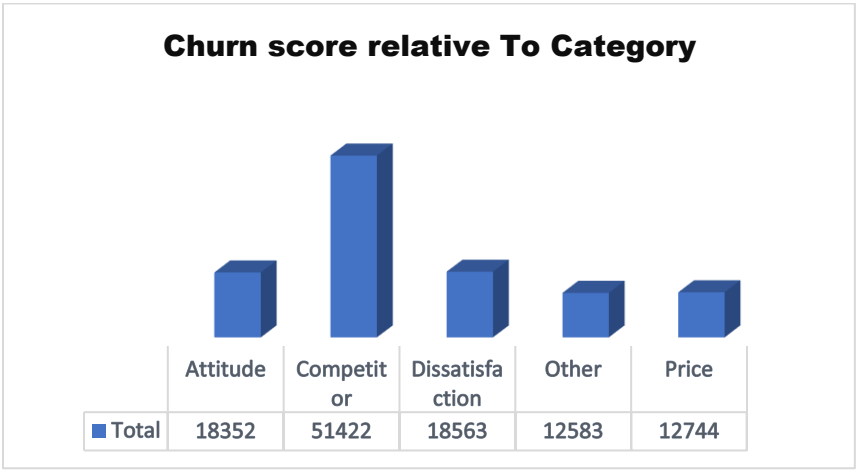
3- internet service Vs Customer Status.



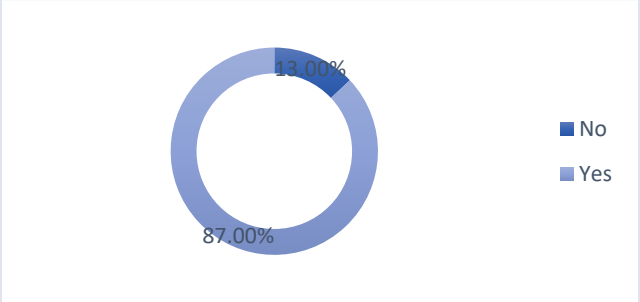
4- Average Churn score by Contract type



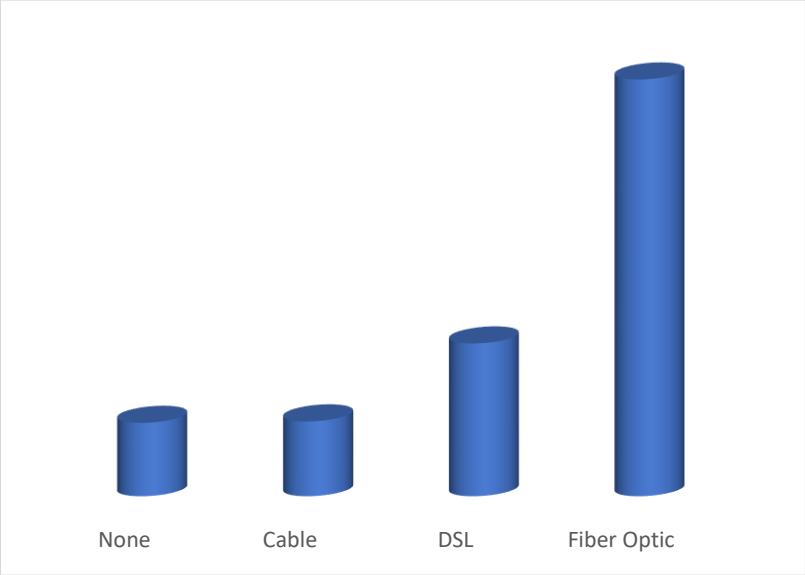
5-Churn score relative To Category reasons of churn



6-'Phone Service': Yes accounts for the majority of 'Total Extra Data Charges'.

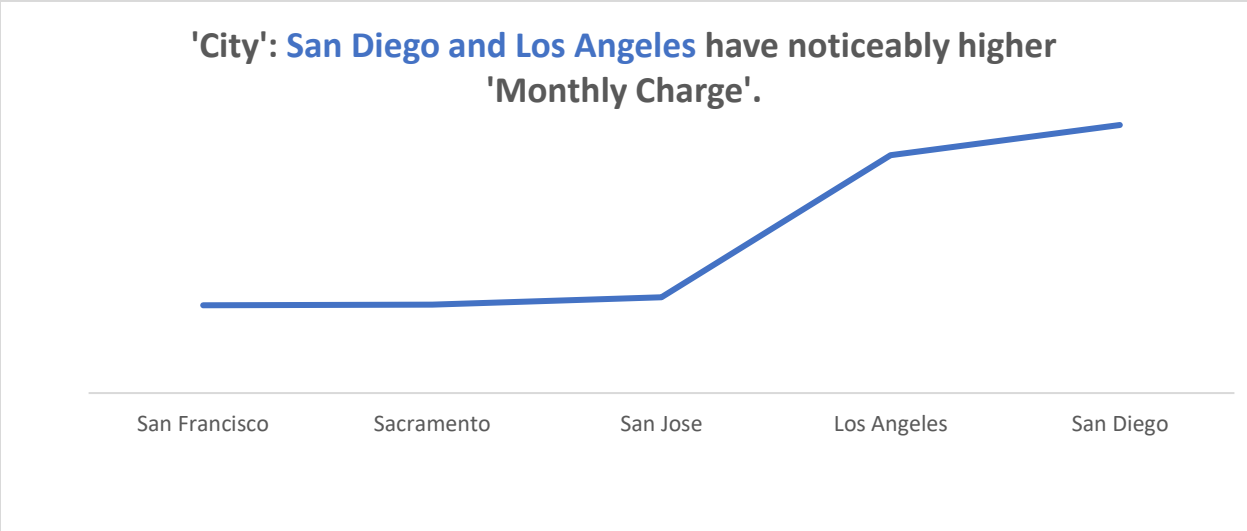


7- Percentage of Total Revenue by Internet Type (**Highest Revenue from Fiber Optic)**

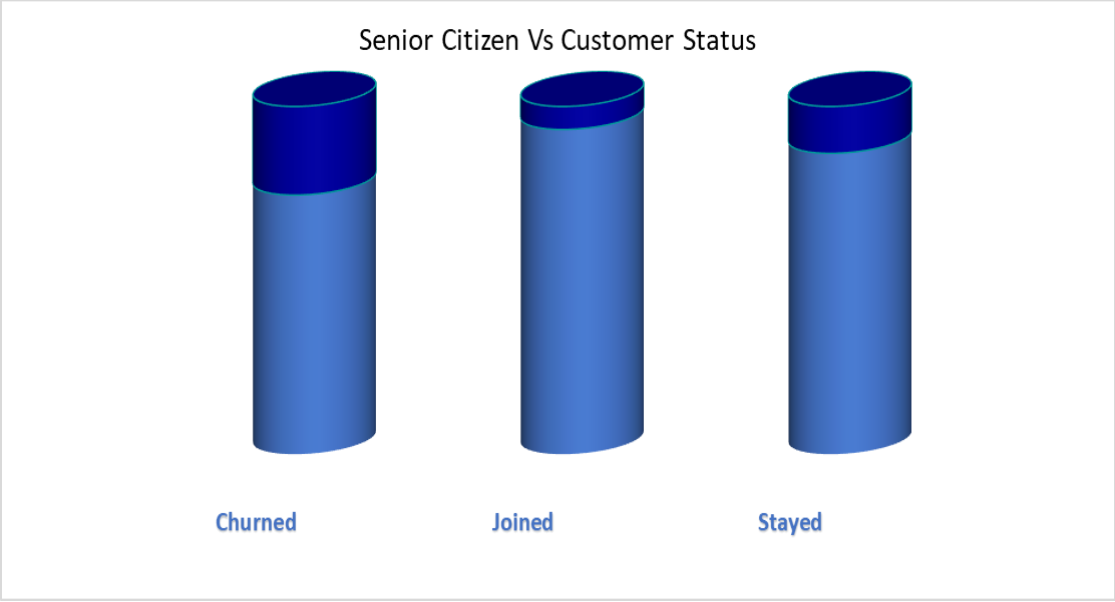


8-Top 5 cities Acc. To 'Monthly Charge'. (> 1% of total Revenues)

'City': San Diego and Los Angeles have noticeably higher 'Monthly Charge'.

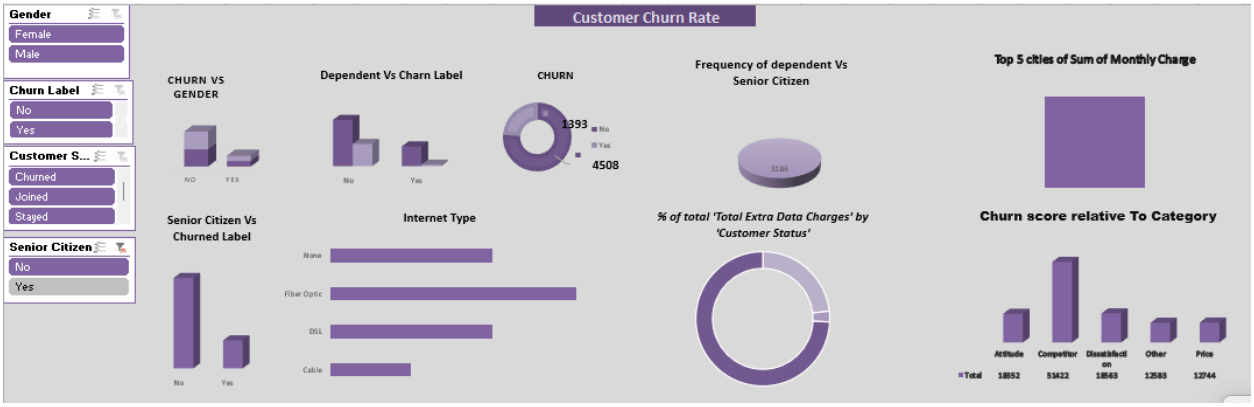


9- Senior Citizen Vs Customer Status

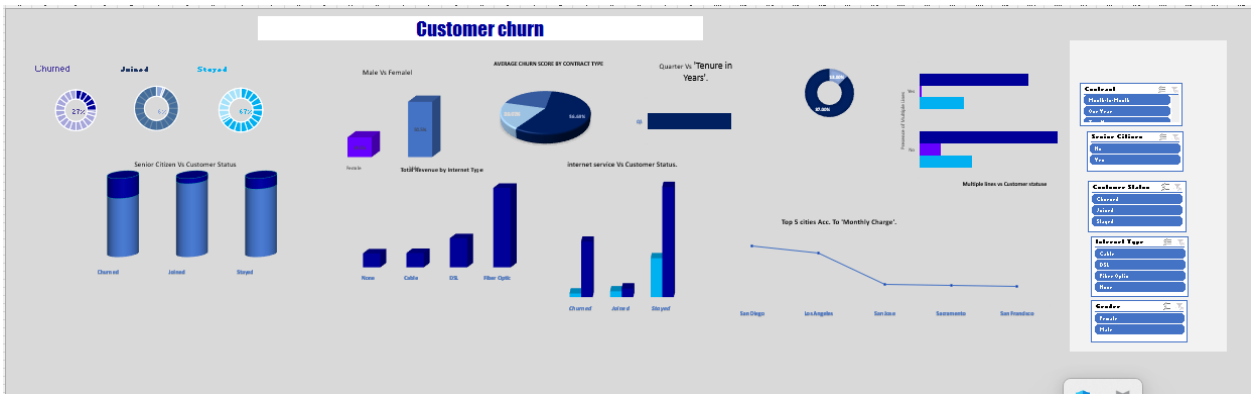


Design a Dashboard

First Version



Last version



What variables do you think significantly impact churn score? (Consider using categorical data encoding)

- **No Correlation** between **Gender and *Churn Score***
  With Anova Test No correlation acc. To high P-value.

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.285912	1	1.285912	0.002868844	0.95728598	3.842779869
Within Groups	3156013	7041	448.2337			
Total	3156015	7042				

- **Strong** Correlation between **Senior Citizen and *Churn Score***

With Anova Test significant correlation acc. To low P-value (<.05).

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	32801.22	1	32801.22	73.94735916	9.78E-18	3.842779869
Within Groups	3123213	7041	443.5753			
Total	3156015	7042				

- Also, **Strong Correlation** after Anova -test (<.05) between **and *Churn Score*** *and* ***'Internet Service',***

***'Contract ','*Streaming TV', 'Streaming Movies', 'Multiple lines', Online Security**

Online Backup.