ANALYZING TELECOM CUSTOMER CHURN WITH SAS AND SPSS: INSIGHTS AND STRATEGIES

Visionary Analysts Team

Content Layout

- Dataset Description
- Business Case
- Problem Statement
- Project Objectives
- Data Analysis Methods
- Analysis using SAS
- Analysis using SPSS
- Key Findings and Recommendations

Dataset Description: Telco Customer Churn

The raw data is in **csv Format** and contains :

- **7043** Observations (customers data(rows))
- **21** Variables (columns):
  - **Churn**: Customers who left within the last month.
  - **Services that each customer has signed up for** - phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
  - **Customer account information** – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
  - **Demographic info about customers** – gender, age range, and if they have partners and dependents

WA_Fn-UseC_-Telco-Customer-Churn.csv (977.5 kB)

Detail | Compact | Column | 21 of 21 columns

...Supp...	Streaming...	Streaming...	Contract	Paperless...	Payment...	MonthlyC...	TotalChar...	Churn
	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
	No	No	One year	No	Mailed check	56.95	1889.5	No
	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
	No	No	One year	No	Bank transfer (automatic)	42.3	1840.75	No
	No	No	Month-to-month	Yes	Electronic check	70.7	151.65	Yes
	Yes	Yes	Month-to-month	Yes	Electronic check	99.65	820.5	Yes

- **Dataset Source:**

kaggle

https://www.kaggle.com/code/danishmubashar/telco-customer-churn-99-acc/input

Dataset Description: Telco Customer Churn

Categorical Data	Numeric Data
Gender: Gender of the customer (Male/Female). **Senior Citizen**: Indicates if the customer is a senior citizen (0 for No, 1 for Yes). **Partner**: Whether the customer has a partner (Yes/No). **Dependents**: Whether the customer has dependents (Yes/No). **Phone Service**: Whether the customer has a phone service (Yes/No). **Multiple Lines**: If the customer has multiple lines (No, Yes, No phone service). **Internet Service**: Type of internet service used by the customer (DSL, Fiber optic, No). **Online Security**: If the customer has online security add-on (Yes, No, No internet service). **Online Backup**: If the customer has an online backup service (Yes, No, No internet service). **Device Protection**: If the customer has device protection (Yes, No, No internet service). **Tech Support**: If the customer has technical support (Yes, No, No internet service). **Streaming TV**: If the customer has streaming TV service (Yes, No, No internet service). **Streaming Movies**: If the customer has streaming movies service (Yes, No, No internet service). **Contract**: The contract type (Month-to-month, One year, Two year). **Paperless Billing**: If the customer has paperless billing (Yes/No). **Payment Method**: The method used for payment (e.g., Electronic check, Mailed check, Bank transfer, Credit card). **Churn**: Whether the customer churned or not (Yes/No).	**Tenure**: Number of months the customer has been with the company. **Monthly Charges**: Monthly charges for the customer. **Total Charges**: Total amount charged to the customer.

Business Case

In the highly competitive telecom industry, understanding and addressing customer churn is vital.

Losing customers results in:

Lost revenue,

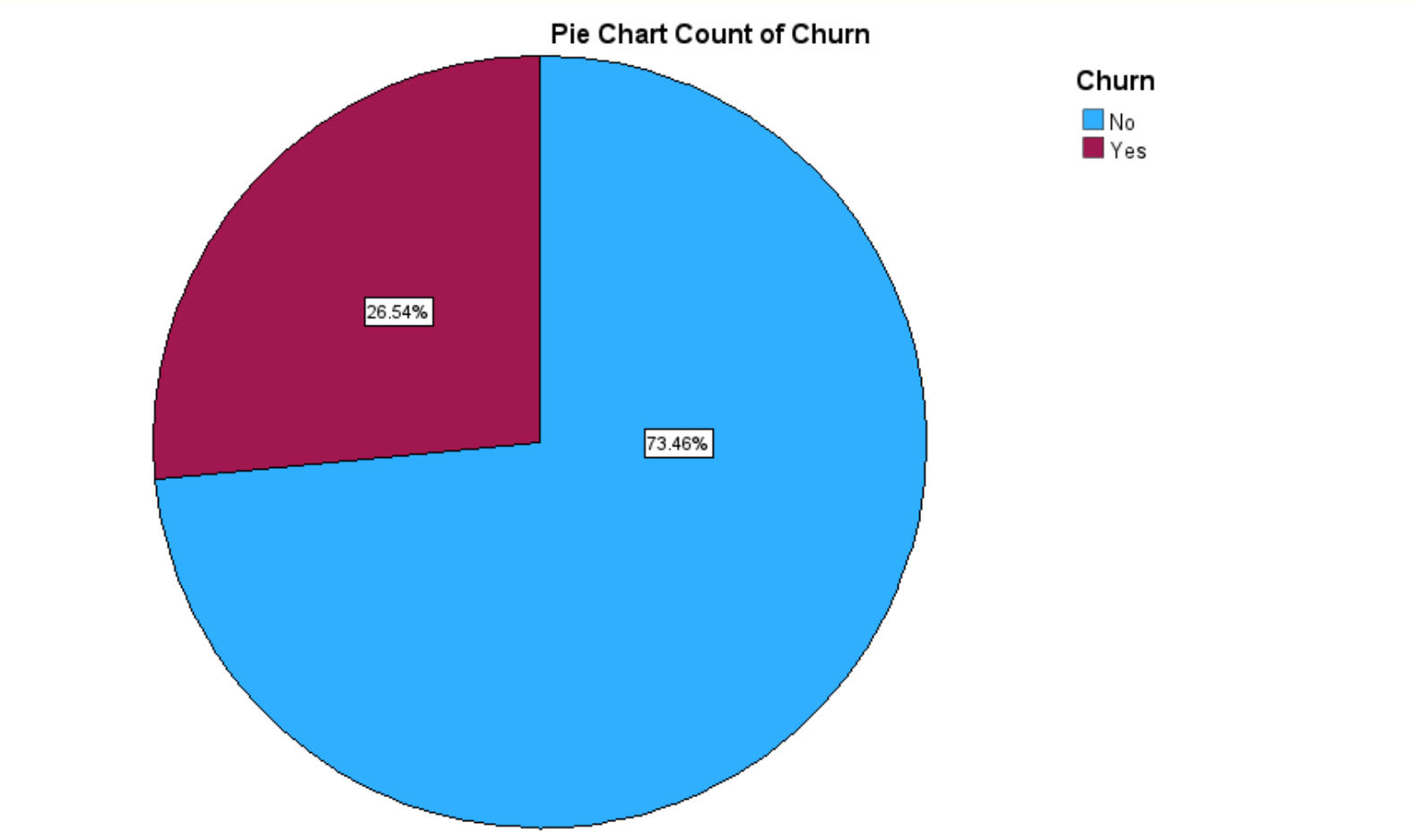
Higher acquisition costs, and

Can impact overall brand reputation.

This project aims to analyze factors that contribute to customer churn and provide data-driven strategies to enhance customer retention.

Problem Statement

Churn directly impacts revenue. Understanding the drivers of churn is crucial to implementing effective retention strategies.



Project Objectives

Objective1	Objective 2	Objective 3	Objective 4
Identify Drivers of Churn	Profile High-Risk Customers	Visualize Patterns and Trends	Recommend Retention Strategies

Data Analysis Methods

- Descriptive statistics (summary and frequency analysis) using SAS
- Visualizations like Pie and bar charts, boxplots, and histograms using SAS
- Chi-square tests for categorical variables (e.g., Contract and Churn) using SPSS
- Independent Samples t-test for Churn & Tenure variables using SPSS
- Regression analysis to assess numerical variable relationships (e.g., tenure and Total Charges) using SPSS

Data Analytics using SAS

Descriptive Statistics:

- Summary Statistics for Numerical Variables
- Frequency Counts for Categorical Variables

Data Visualization for:

- Categorical Variables (Contract Frequency& Payment Method)
- Histograms for Numerical Variables

Churn Rate Analysis by :

- Demographic and Service Variables (Gender, Senior Citizen, and Partner)
- Churn Rate by Contract and Service Types

Summary Statistics for Numerical Variables

Analysis Variable : tenure								
Mean	Std Dev	Variance	Minimum	Maximum	Mode	Range	N	N Miss
32.3711487	24.5594810	603.1681081	0	72.0000000	1.0000000	72.0000000	7043	0

Analysis Variable : MonthlyCharges						
Mean	Std Dev	Variance	Minimum	Maximum	N	N Miss
64.7616925	30.0900471	905.4109343	18.2500000	118.7500000	7043	0

Analysis Variable : TotalCharges								
Mean	Std Dev	Variance	Minimum	Maximum	Mode	Range	N	N Miss
2283.30	2266.77	5138252.41	18.8000000	8684.80	20.2000000	8666.00	7032	11

11 Missing values

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	custom	gender	SeniorC	Partner	Depend	tenure	PhoneS	Multiple	Interne	OnlineS	OnlineB	DeviceF	TechSu	Streami	Streami	Contrac	Paperle	Paymer	Monthl	TotalCh	Churn
490	4472-LVYC	Female	0	Yes	Yes	0	No	No phone	DSL	Yes	No	Yes	Yes	Yes	No	Two year	Yes	Bank trans	52.55		No
755	3115-CZM	Male	0	No	Yes	0	Yes	No	No	No interne	No interne	No interne	No interne	No interne	No interne	Two year	No	Mailed che	20.25		No
938	5709-LVOI	Female	0	Yes	Yes	0	Yes	No	DSL	Yes	Yes	Yes	No	Yes	Yes	Two year	No	Mailed che	80.85		No
1084	4367-NUY	Male	0	Yes	Yes	0	Yes	Yes	No	No interne	No interne	No interne	No interne	No interne	No interne	Two year	No	Mailed che	25.75		No
1342	1371-DWP	Female	0	Yes	Yes	0	No	No phone	DSL	Yes	Yes	Yes	Yes	Yes	No	Two year	No	Credit carc	56.05		No
3333	7644-OMV	Male	0	Yes	Yes	0	Yes	No	No	No interne	No interne	No interne	No interne	No interne	No interne	Two year	No	Mailed che	19.85		No
3828	3213-VVO	Male	0	Yes	Yes	0	Yes	Yes	No	No interne	No interne	No interne	No interne	No interne	No interne	Two year	No	Mailed che	25.35		No
4382	2520-SGTT	Female	0	Yes	Yes	0	Yes	No	No	No interne	No interne	No interne	No interne	No interne	No interne	Two year	No	Mailed che	20		No
5220	2923-ARZL	Male	0	Yes	Yes	0	Yes	No	No	No interne	No interne	No interne	No interne	No interne	No interne	One year	Yes	Mailed che	19.7		No
6672	4075-WKN	Female	0	Yes	Yes	0	Yes	Yes	DSL	No	Yes	Yes	Yes	Yes	No	Two year	No	Mailed che	73.35		No
6756	2775-SEFE	Male	0	No	Yes	0	Yes	Yes	DSL	Yes	Yes	No	Yes	No	No	Two year	Yes	Bank trans	61.9		No
7045																					

Summary Statistics for Numerical Variables after removing missing values

Analysis Variable : tenure								
Mean	Std Dev	Variance	Minimum	Maximum	Mode	Range	N	N Miss
32.4217861	24.5452597	602.4697742	1.0000000	72.0000000	1.0000000	71.0000000	7032	0

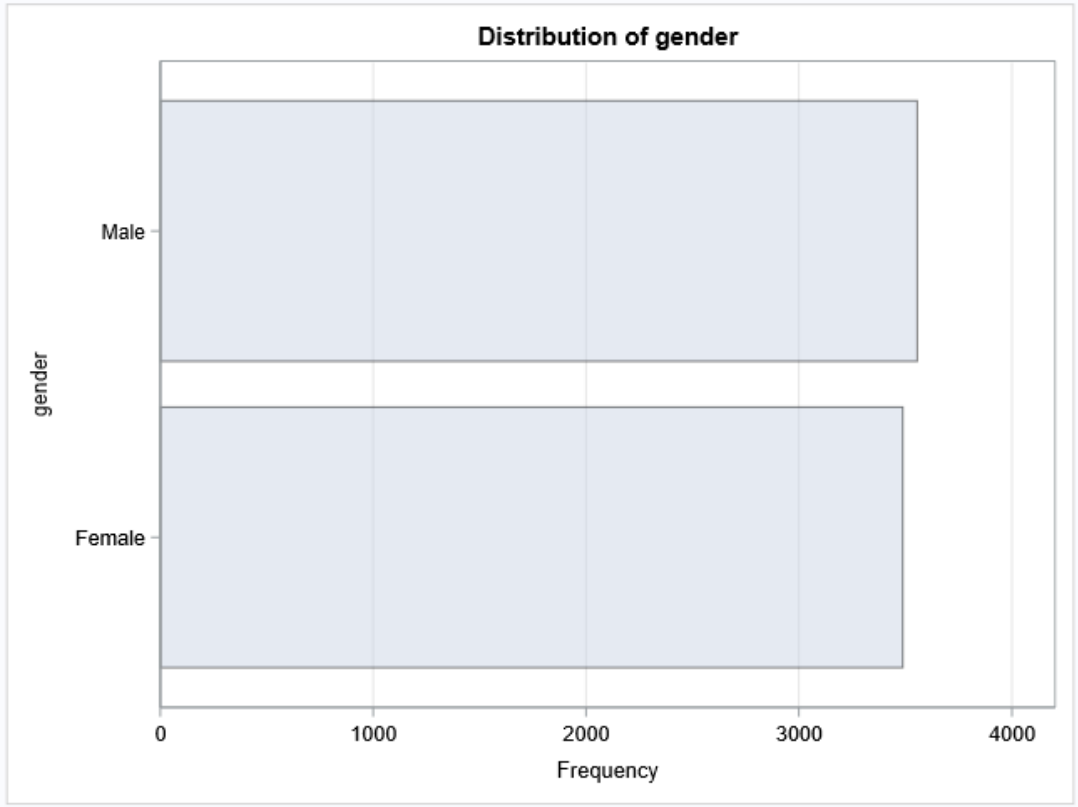
Analysis Variable : MonthlyCharges								
Mean	Std Dev	Variance	Minimum	Maximum	Mode	Range	N	N Miss
64.7982082	30.0859739	905.1658246	18.2500000	118.7500000	20.0500000	100.5000000	7032	0

Analysis Variable : TotalCharges								
Mean	Std Dev	Variance	Minimum	Maximum	Mode	Range	N	N Miss
2283.30	2266.77	5138252.41	18.8000000	8684.80	20.2000000	8666.00	7032	0

0 Missing values

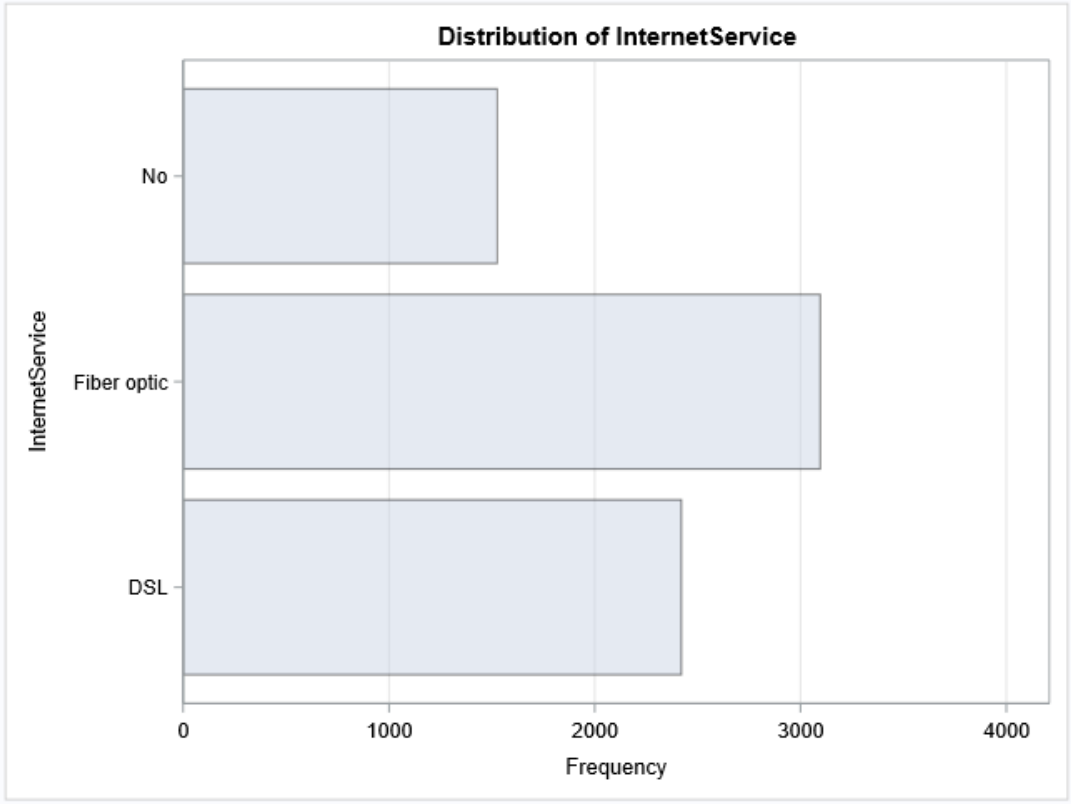
Frequency Counts for Categorical Variables: Gender and Internet Service

gender	Frequency	Percent
Female	3488	49.52
Male	3555	50.48



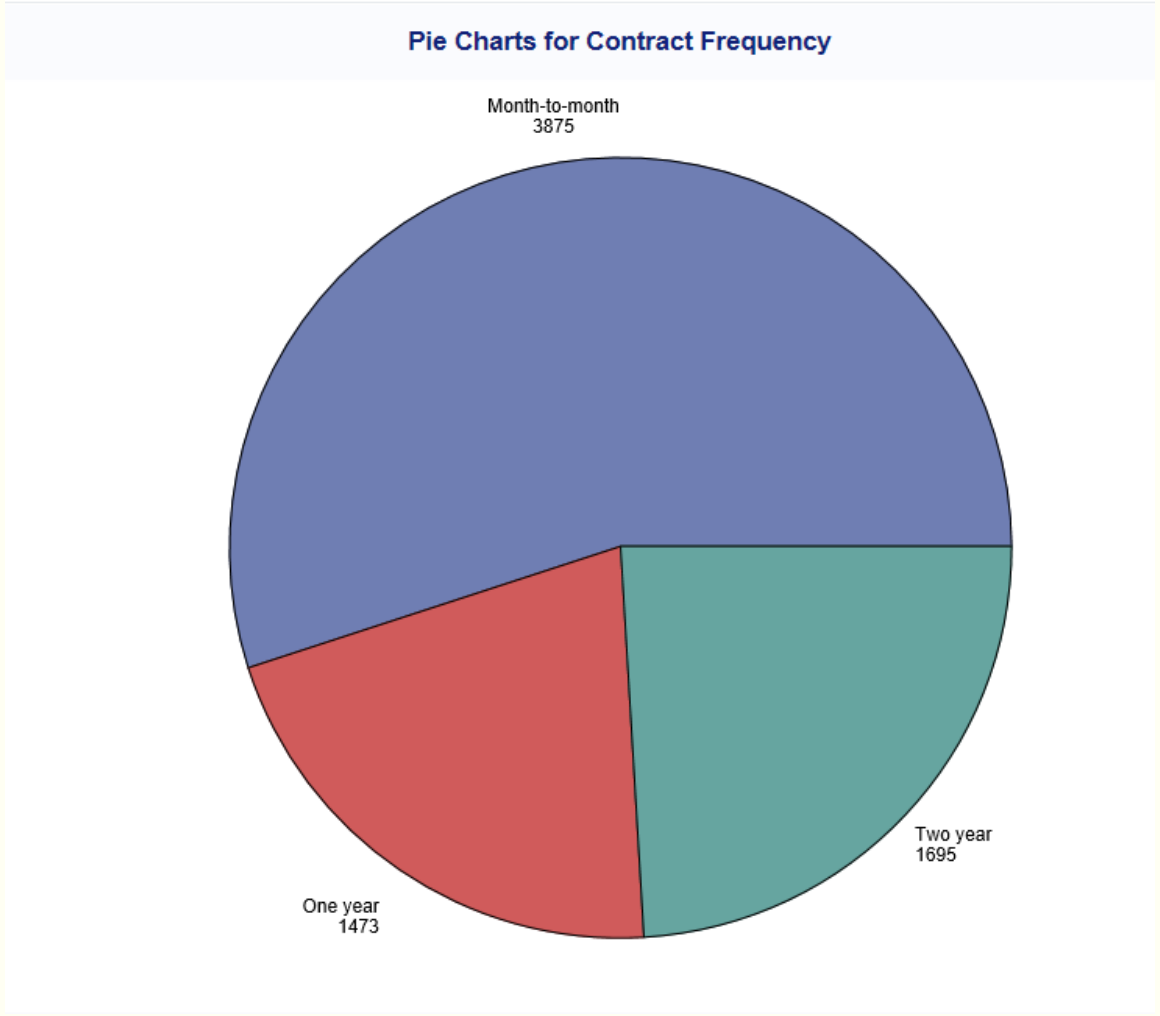
The dataset reveals an almost equal distribution of males and females, indicating a balanced representation of gender among customers.

InternetService	Frequency	Percent
DSL	2421	34.37
Fiber optic	3096	43.96
No	1526	21.67



Approximately 80% of customers in the dataset are subscribed to internet services, highlighting the significant role of internet offerings in the overall service portfolio.

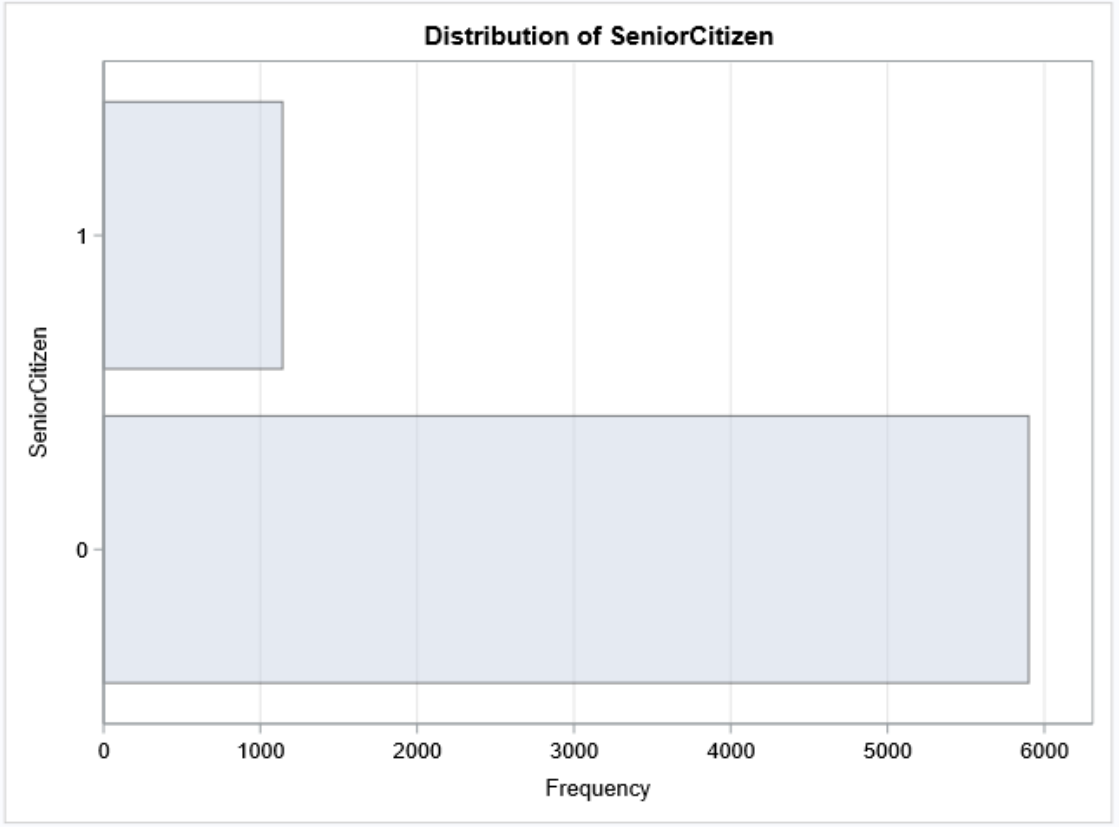
Frequency Counts for Categorical Variables: Contract Type and Senior Citizen



Over half of the customers (55%) are on month-to-month contracts, indicating a preference for flexible subscription options among the majority of the customer base.

0 for No, 1 for Yes

SeniorCitizen	Frequency	Percent
0	5901	83.79
1	1142	16.21

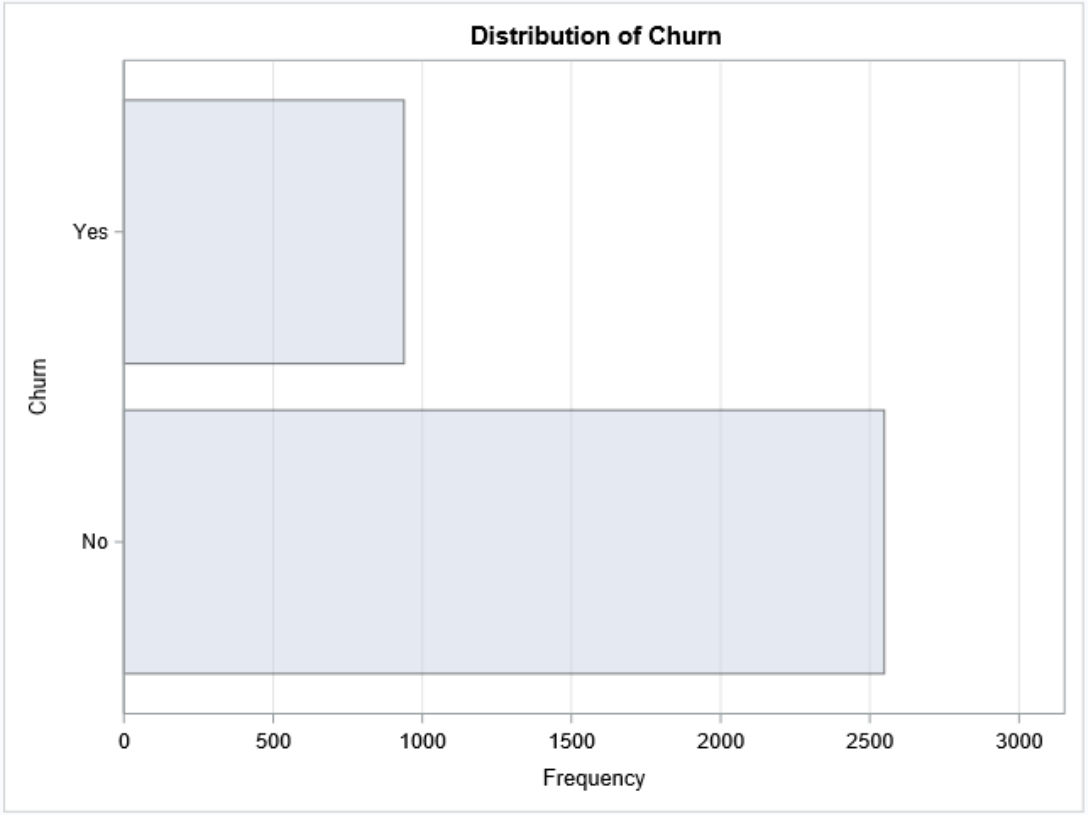


More than 80% of the customers are not senior citizens, indicating that the majority of the customer base falls within younger age groups.

Churn Rate by **Gender**

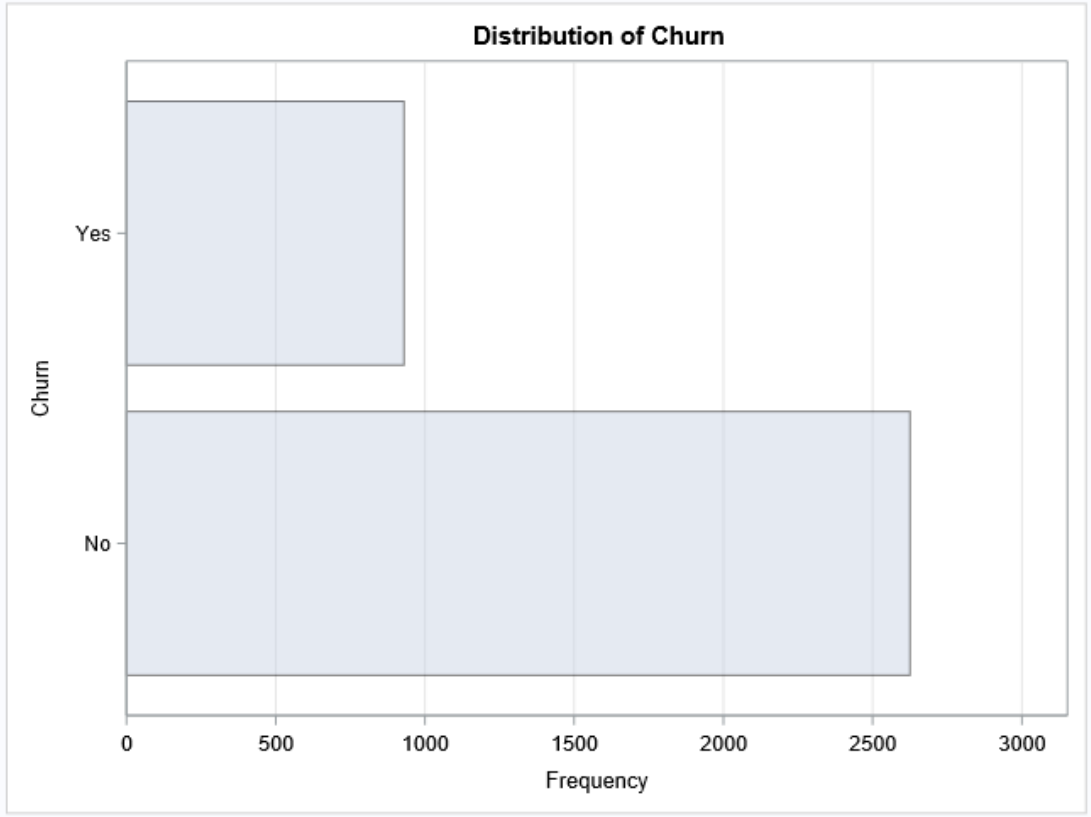
gender=Female

Churn	Frequency	Percent
No	2549	73.08
Yes	939	26.92



gender=Male

Churn	Frequency	Percent
No	2625	73.84
Yes	930	26.16



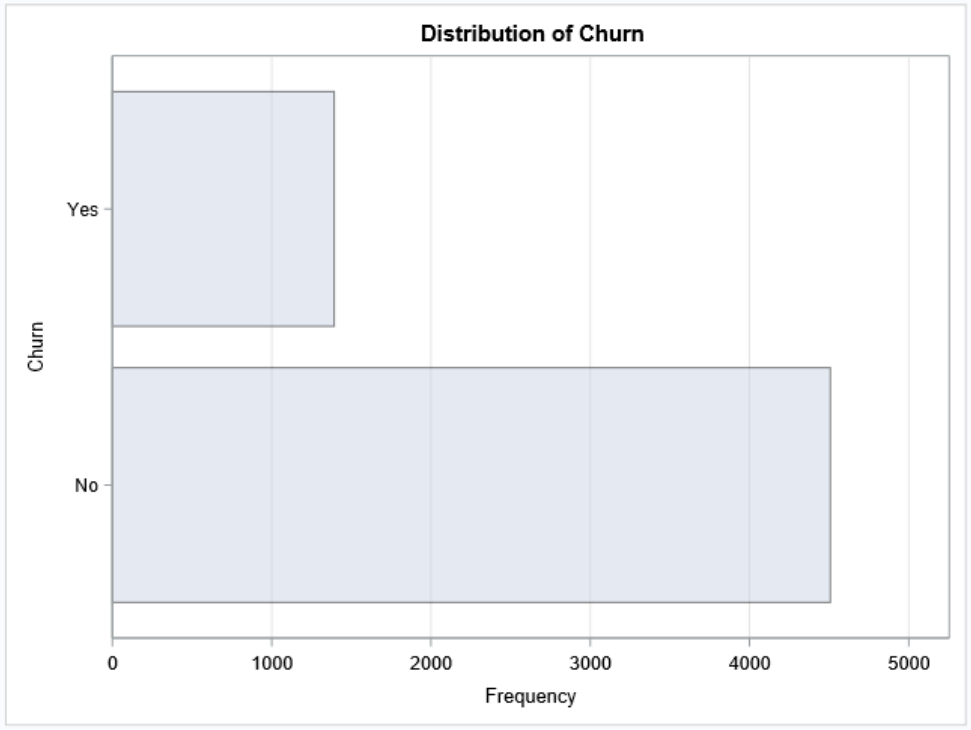
The churn rates for males and females are nearly identical at 26%, suggesting that gender may not have an impact on churn. This observation will be further validated through a Chi-Square test to confirm whether gender is truly independent of churn.

Churn Rate by **Senior Citizen**

SeniorCitizen: (0 = No , 1 = Yes).

SeniorCitizen=0

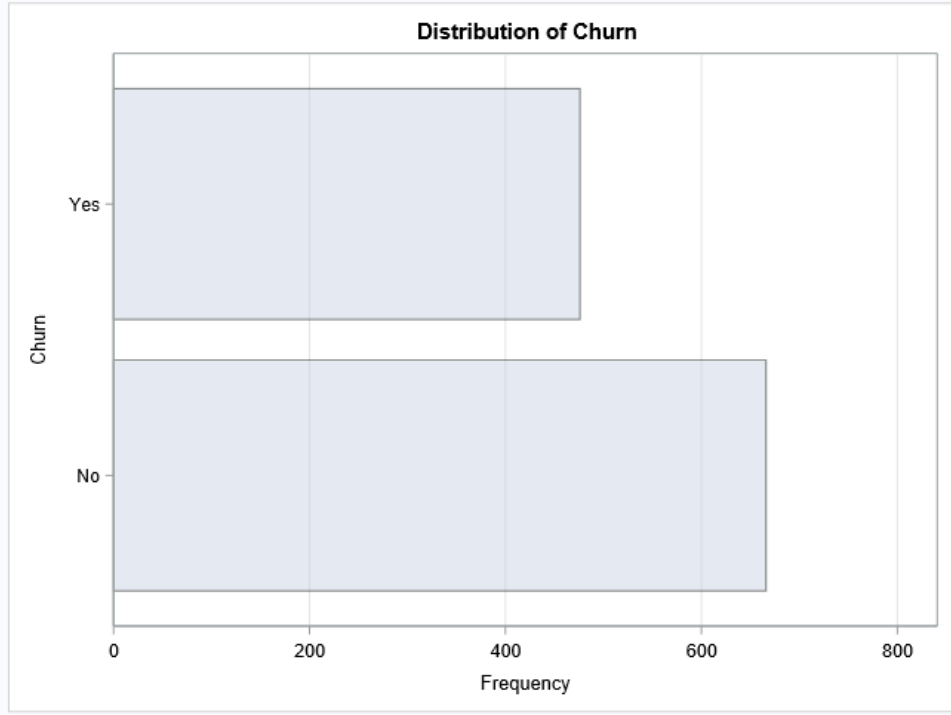
Churn	Frequency	Percent
No	4508	76.39
Yes	1393	23.61



SeniorCitizen: (0 = No , 1 = Yes).

SeniorCitizen=1

Churn	Frequency	Percent
No	666	58.32
Yes	476	41.68

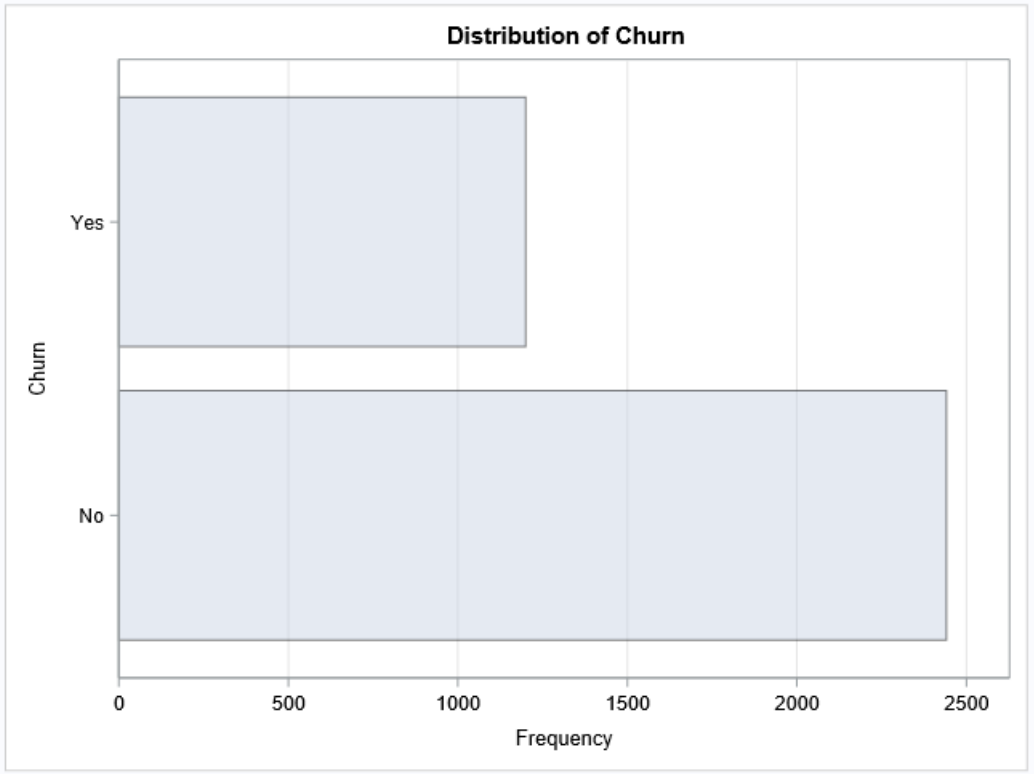


Senior citizens exhibit significantly higher churn rates compared to non-senior citizens. Specifically, 42% of senior citizens (476 out of1,142) have churned, compared to a 24% churn rate for non-senior citizens (1,393 out of 5,901).This disparity suggests that senior citizens may find the services either too expensive or challenging to use, contributing to their higher likelihood of leaving. Targeted interventions, such as simplified plans or tailored support, could help address this issue and improve retention among this demographic.

Churn Rate by having a **Partner**

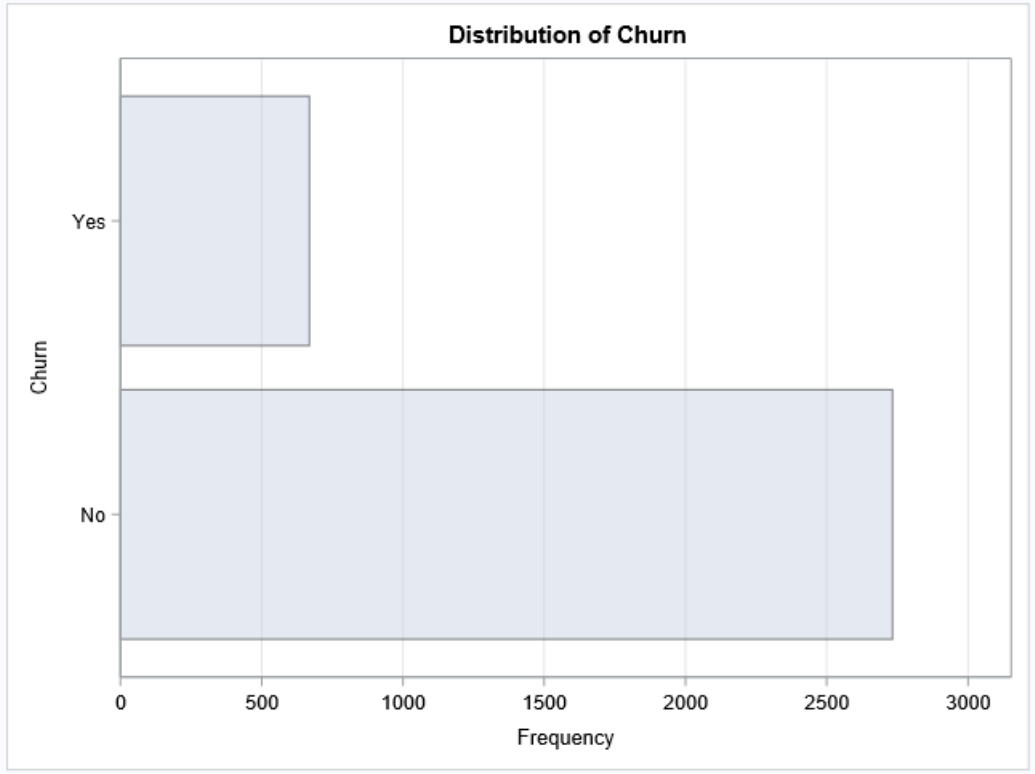
Partner=No

Churn	Frequency	Percent
No	2441	67.04
Yes	1200	32.96



Partner=Yes

Churn	Frequency	Percent
No	2733	80.34
Yes	669	19.66

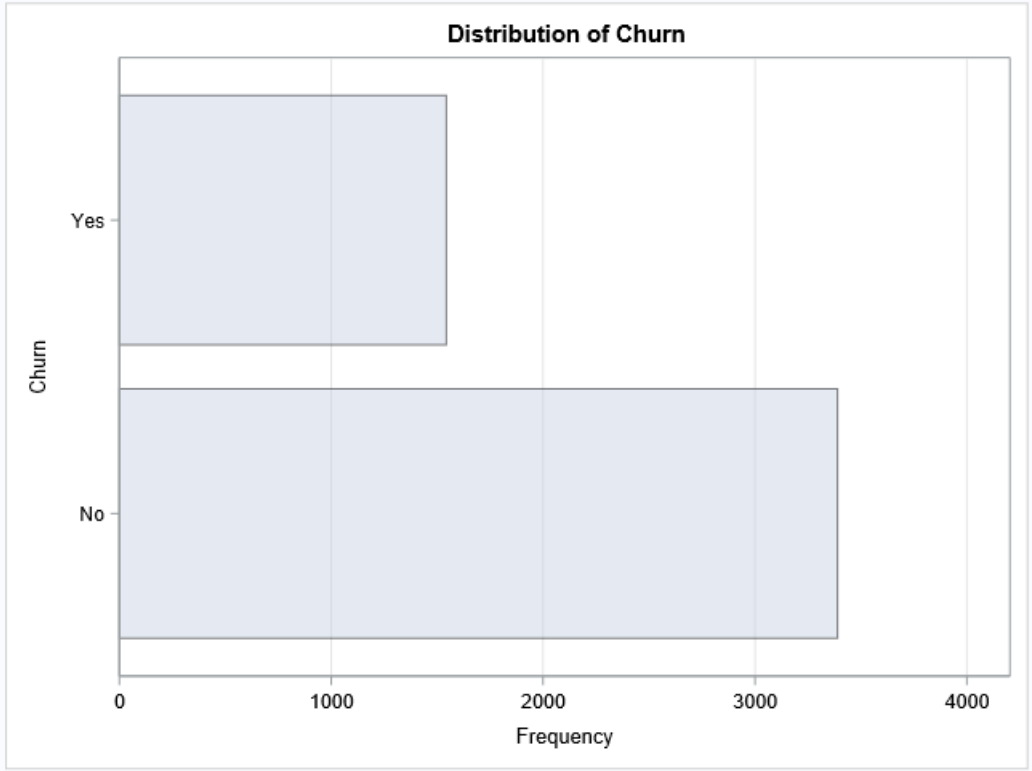


Customers with a partner have a lower churn rate of 20%, compared to 33% for customers without a partner.
This indicates that partnered customers may value stability and consistency in their services, making them less likely to leave. Conversely, customers without a partner may exhibit greater flexibility or be more price-sensitive, contributing to their higher likelihood of churning.

Churn Rate by **Dependents**

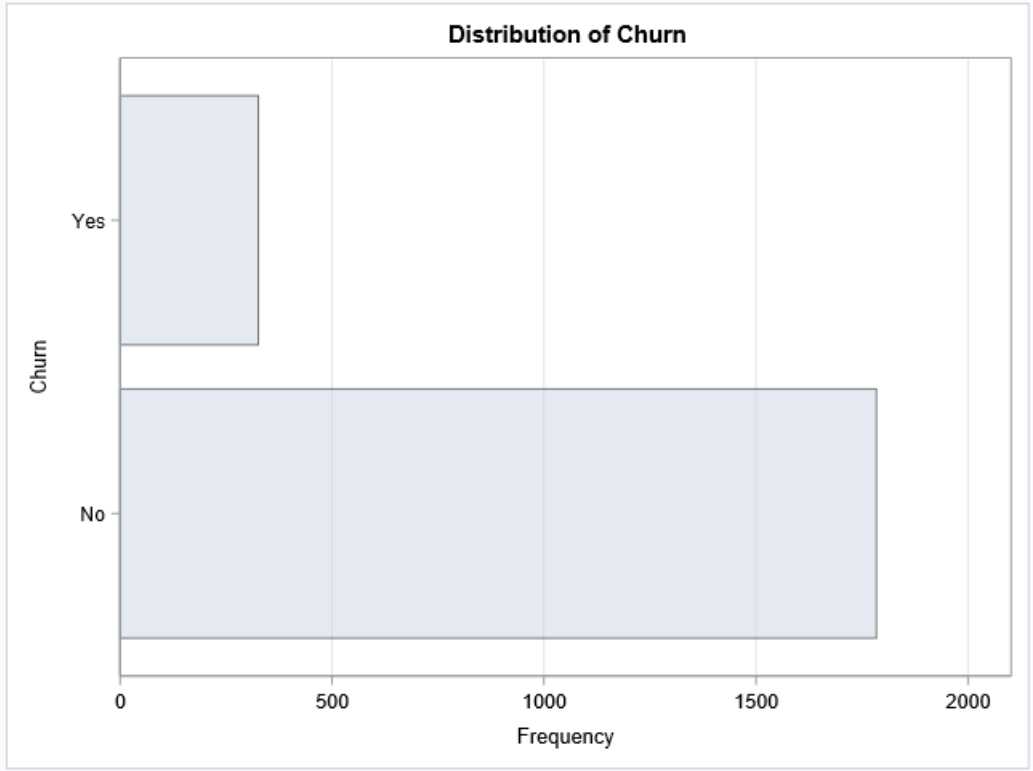
Dependents=No

Churn	Frequency	Percent
No	3390	68.72
Yes	1543	31.28



Dependents=Yes

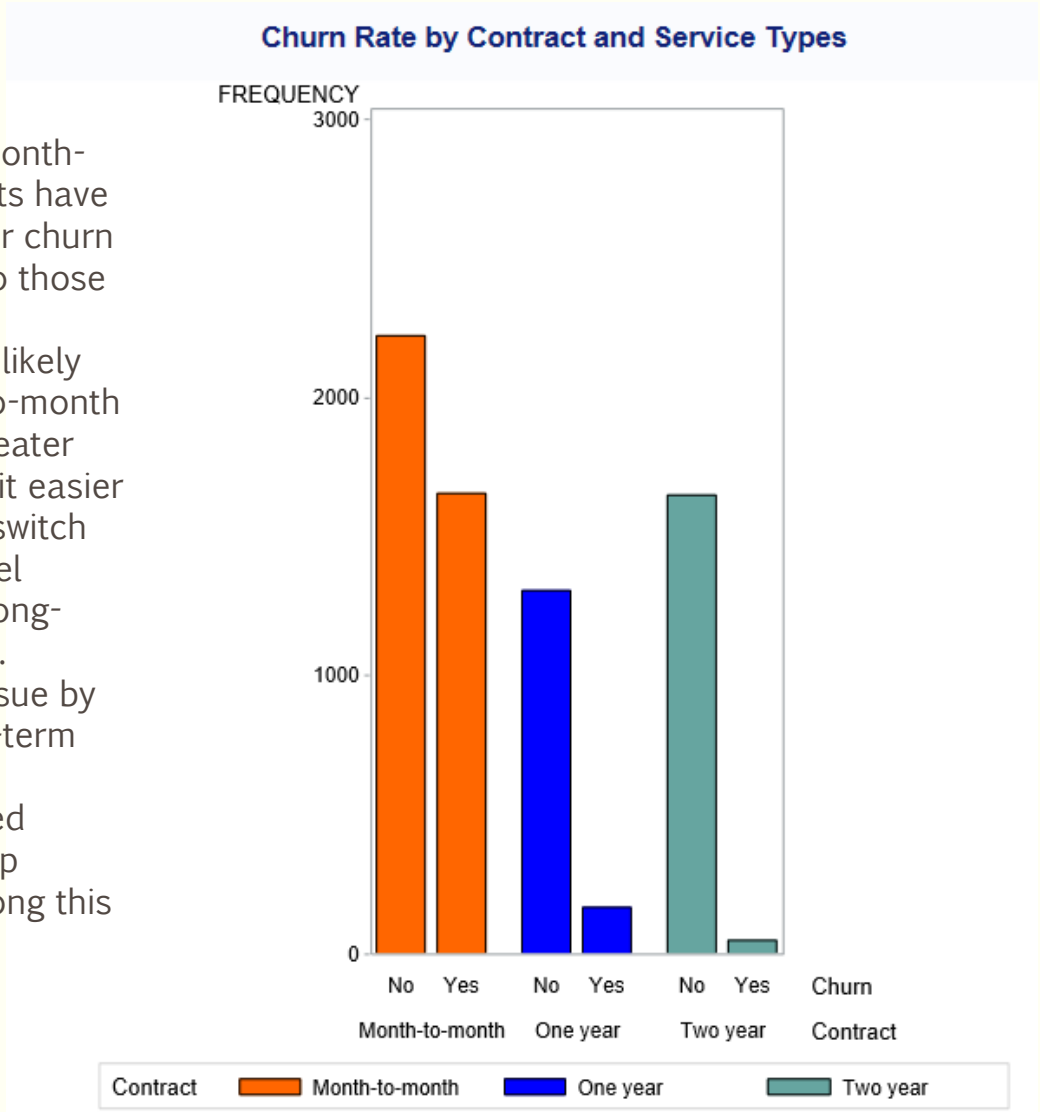
Churn	Frequency	Percent
No	1784	84.55
Yes	326	15.45



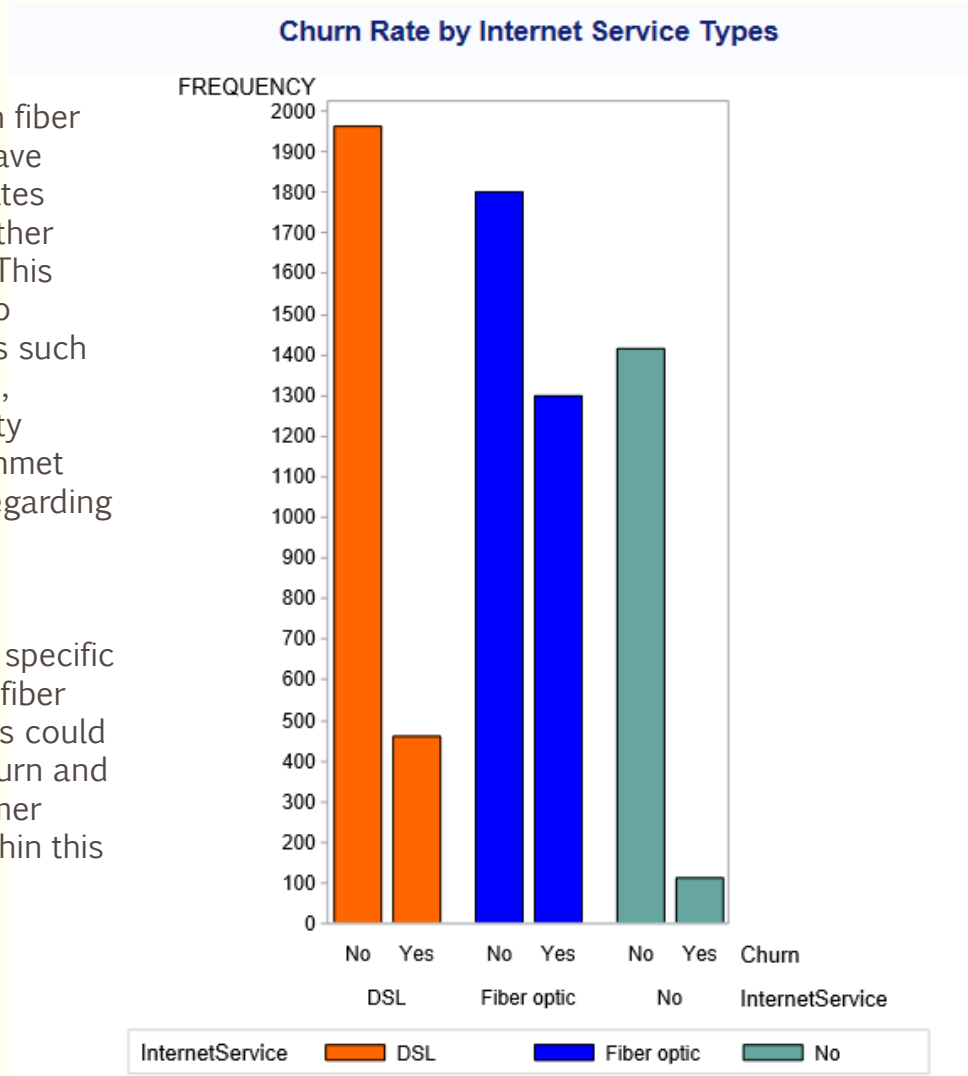
Customers with dependents have a significantly lower churn rate of **15%**, compared to **31%** for customers without dependents.
This suggests that customers with dependents may value stability and are less likely to switch services, potentially due to the importance of maintaining reliable services for their families. On the other hand, customers without dependents may be more flexible or price-sensitive, contributing to their higher churn rate.

Churn Rate by **Contract** and Service Types

Customers with month-to-month contracts have significantly higher churn rates compared to those with longer-term contracts. This is likely because month-to-month contracts offer greater flexibility, making it easier for customers to switch providers or cancel services without long-term commitment. Addressing this issue by incentivizing long-term contracts through discounts or added benefits could help reduce churn among this segment

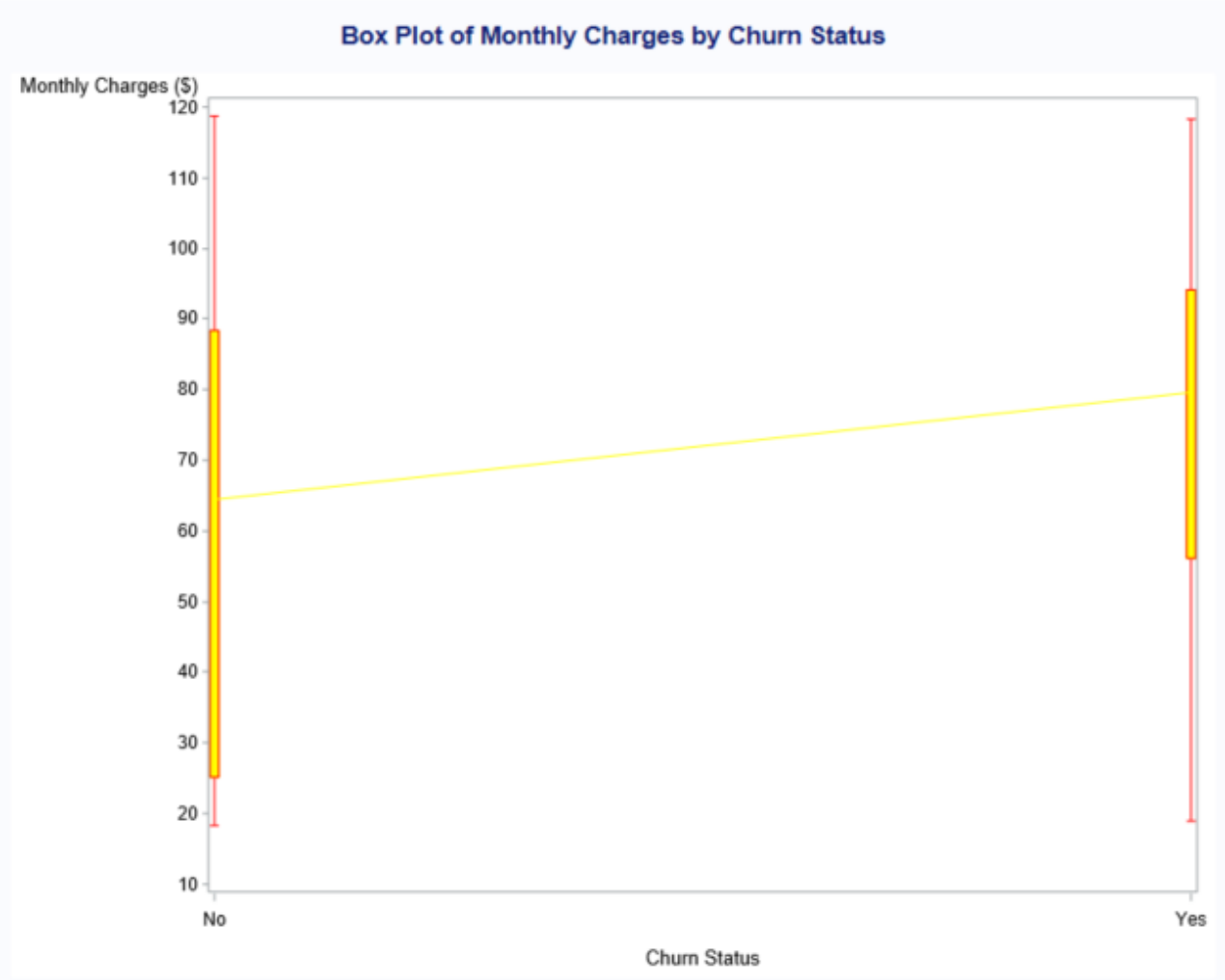


Customers with fiber optic service have higher churn rates compared to other service types. This could be due to potential issues such as higher costs, service reliability concerns, or unmet expectations regarding speed and performance. Identifying and addressing the specific pain points for fiber optic customers could help reduce churn and improve customer satisfaction within this segment.



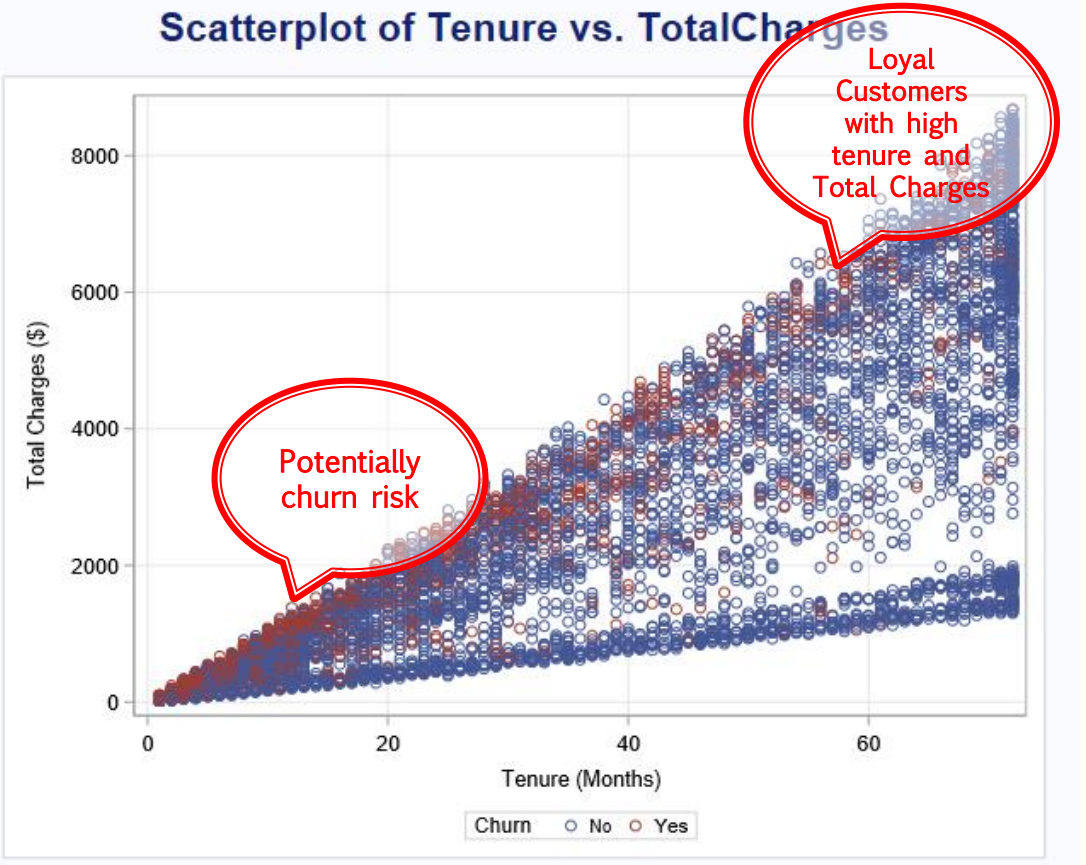
Churn Rate Rates by Monthly Charges

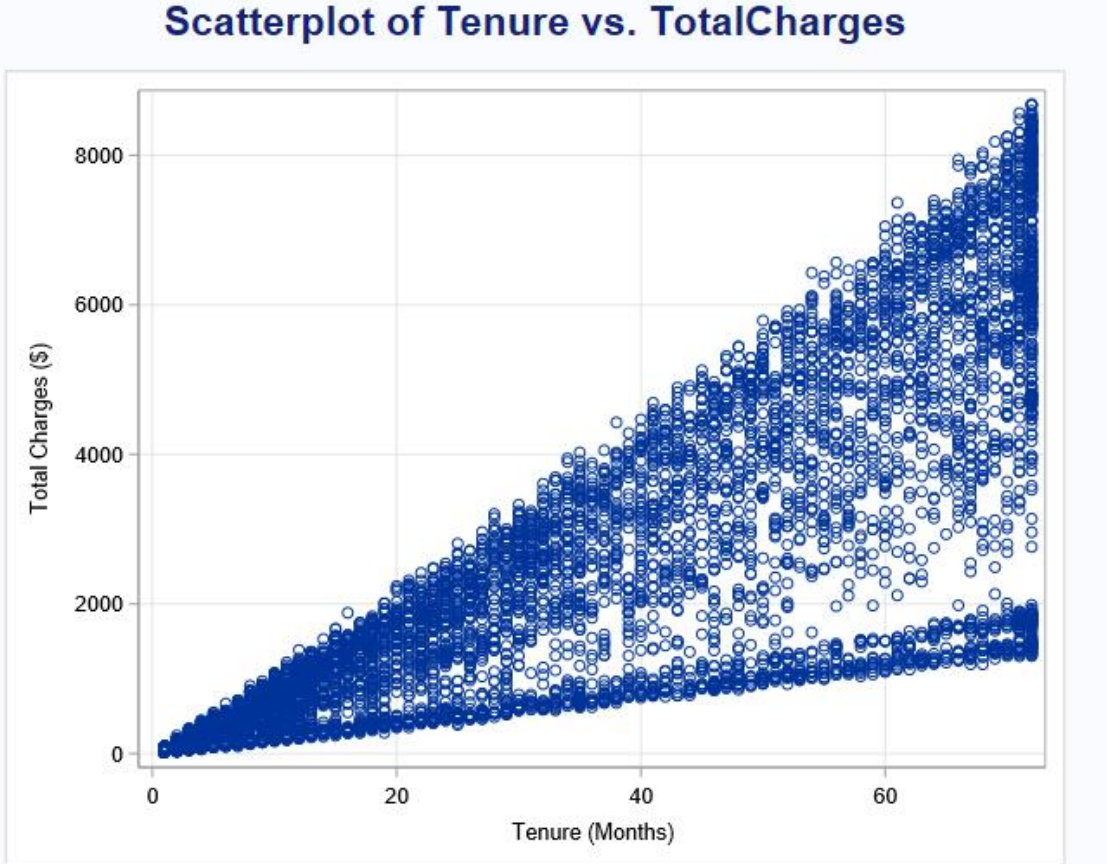
Churned customers (Yes) exhibit higher median and interquartile ranges for Monthly Charges compared to non-churned customers (No). This suggests that higher monthly charges may be a significant contributing factor to customer churn, potentially reflecting dissatisfaction with the perceived value or affordability of the service.
Addressing this by offering tailored pricing plans or added benefits for high-paying customers could help mitigate churn and enhance satisfaction.



Churn Rate Rates by Tenure vs. Total Charges

There is a clear positive trend indicating that Total Charges increase as Tenure grows, reflecting that long-term customers contribute more revenue over time. Additionally, the strong positive correlation between Tenure and Total Charges (Pearson correlation = 0.826) highlights the importance of retaining customers for longer periods, as their cumulative value significantly benefits the business.





Data Analytics using SPSS

Chi square test of independence for:

- Churn and Gender variables
- Churn and Senior Citizen variables
- Churn and Contract Type variables
- Churn and Partner variables

Compare Means and Proportions: Independent Samples t-test for:

- Gender & Tenure variables
- Churn & Tenure variables

Using regression, analyze the relationship between:

- Tenure and Monthly Charges
- Tenure and dependent Total Charges

Chi square test of independence for Churn and Gender variables

Research Question:
Does customer gender affect churn?

Hypotheses:

- **H_0 (Null Hypothesis):** Customer gender is independent of churn (i.e., there is no association between customer gender and churn).
- **H_1 (Alternative Hypothesis):** Customer gender is not independent of churn (i.e., customer gender affects churn).

Report:
We conducted a Chi-Square Test of Independence to determine whether the two categorical variables, customer gender and churn, are associated. Using a significance level of $\alpha = 0.05$, the test yielded a p-value of 0.470, which is greater than 0.05.

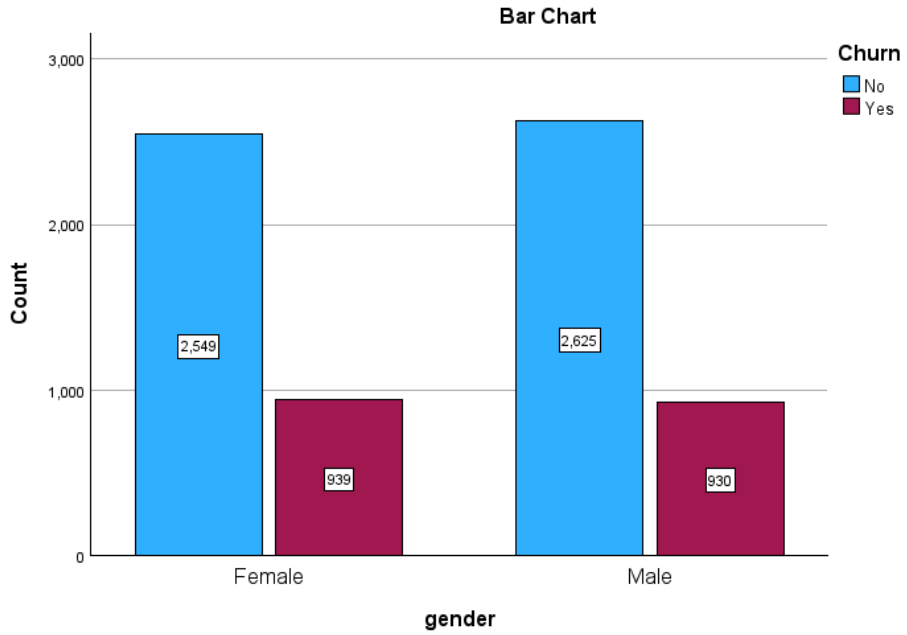
Conclusion:
Since the p-value exceeds the significance level, we fail to reject the null hypothesis. This means customer gender is independent of churn, and there is no significant association between the two variables. In other words, customer gender does not affect the likelihood of churn.
This finding suggests that churn-related factors are likely influenced by other variables, such as pricing, contract type, or service quality, rather than gender.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.522[a]	1	.470		
Continuity Correction[b]	.484	1	.487		
Likelihood Ratio	.522	1	.470		
Fisher's Exact Test				.483	.243
N of Valid Cases	7043				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 925.61.

b. Computed only for a 2x2 table



Chi square test of independence for Churn and Senior Citizen variables

Research Question:
Does being a senior citizen affect churn?

Hypotheses:

- **H_0 (Null Hypothesis):** Being a senior citizen is independent of churn (i.e., there is no association between being a senior citizen and churn).
- **H_1 (Alternative Hypothesis):** Being a senior citizen is not independent of churn (i.e., being a senior citizen affects churn).

Report:
We conducted a Chi-Square Test of Independence to determine whether there is an association between the two categorical variables, Senior Citizen status and churn. Using a significance level of $\alpha = 0.05$, the p-value obtained was **0.01**, which is less than 0.05.

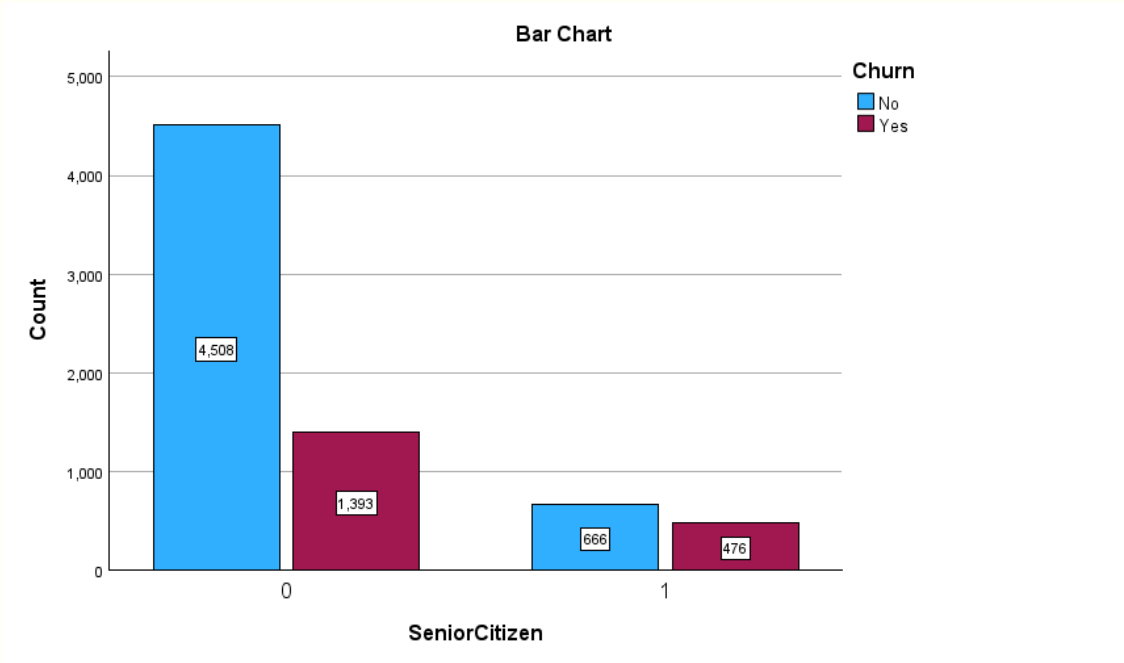
Conclusion:
Since the p-value is smaller than the significance level, we reject the null hypothesis. This indicates that the alternative hypothesis is true: being a senior citizen **is not independent of churn**. In other words, there is a significant association between senior citizen status and churn.
This finding suggests that senior citizens are more likely to churn, and addressing their specific needs (e.g., through tailored support or pricing plans) could help reduce churn within this group.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	160.352[a]	1	<.001		
Continuity Correction[b]	159.426	1	<.001		
Likelihood Ratio	148.991	1	<.001		
Fisher's Exact Test				<.001	<.001
N of Valid Cases	7043				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 303.05.

b. Computed only for a 2x2 table



Chi square test of independence for Churn & Contract Type variables

Research Question:
Does contract type affect churn?
Hypotheses:

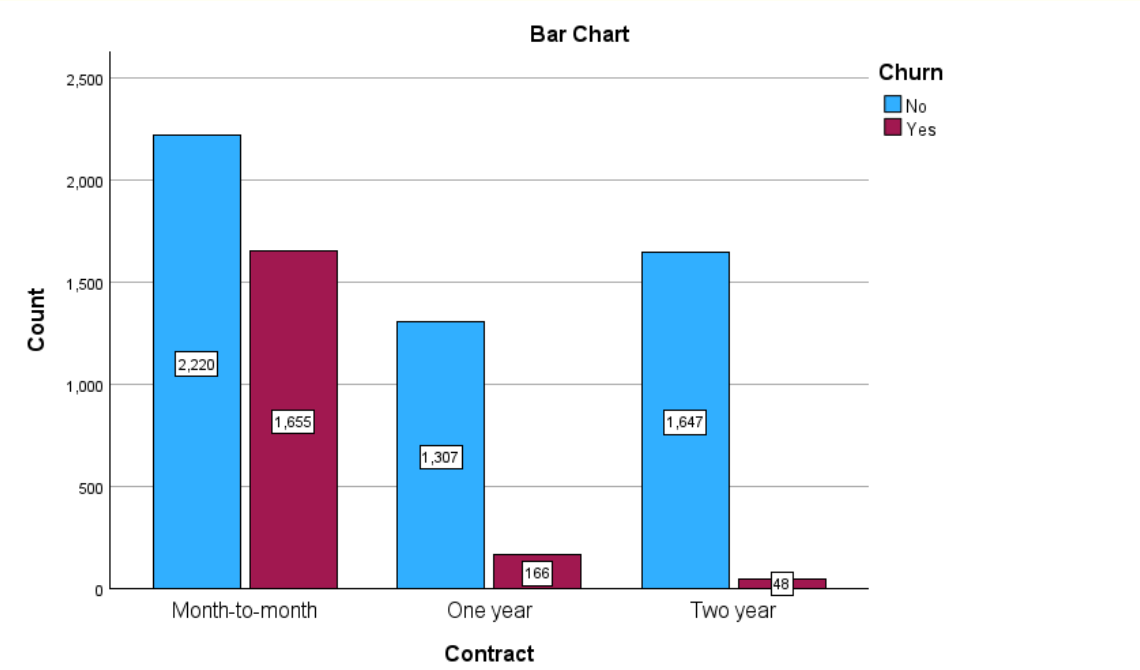
- **H_0 (Null Hypothesis):** Contract type is independent of churn (i.e., there is no association between contract type and churn).
- **H_1 (Alternative Hypothesis):** Contract type is not independent of churn (i.e., contract type affects churn).

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	1184.597[a]	2	<.001
Likelihood Ratio	1386.810	2	<.001
N of Valid Cases	7043		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 390.89.

Report:
We conducted a Chi-Square Test of Independence to assess whether there is an association between the two categorical variables, contract type and churn. Using a significance level of $\alpha = 0.05$, the p-value obtained was **0.01**, which is less than 0.05.
Conclusion:
Since the p-value is less than the significance level, we reject the null hypothesis. This indicates that the alternative hypothesis is true: **contract type is not independent of churn**. In other words, there is a significant association between contract type and churn. This finding suggests that customers with different contract types exhibit different churn behaviors, and addressing this through targeted strategies (e.g., offering incentives for long-term contracts) could help reduce churn.



Chi square test of independence for Churn and Partner variables

Research Question:
Does a customer's partner status (having a partner or not) affect churn?

Hypotheses:

- **H_0 (Null Hypothesis):** Partner status is independent of churn (i.e., there is no association between having a partner and churn).
- **H_1 (Alternative Hypothesis):** Partner status is not independent of churn (i.e., there is an association between having a partner and churn).

Report:
We conducted a Chi-Square Test of Independence to determine whether there is an association between the two categorical variables: partner status and churn. Using a significance level of **α = 0.05**, the test yielded a p-value of **0.01**, which is less than 0.05.

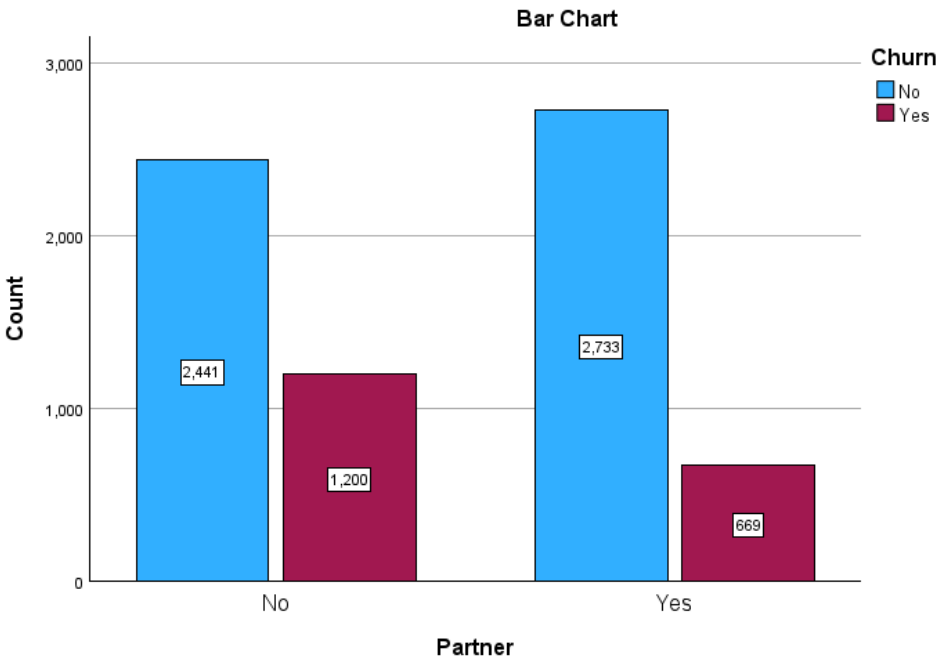
Conclusion:
Since the p-value is smaller than the significance level, we reject the null hypothesis and accept the alternative hypothesis. This indicates that **partner status is not independent of churn**, meaning there is a significant association between having a partner and the likelihood of churn.
This result suggests that customers with a partner may exhibit different churn behaviors compared to those without a partner, which can inform targeted retention strategies.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	159.415[a]	1	<.001		
Continuity Correction[b]	158.733	1	<.001		
Likelihood Ratio	161.336	1	<.001		
Fisher's Exact Test				<.001	<.001
N of Valid Cases	7043				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 902.79.

b. Computed only for a 2x2 table



Independent Samples t-test for Gender & Tenure variables

Research Question:
Is there a significant difference between the average tenure of males and females?
Hypotheses:
- **H_0 (Null Hypothesis):** There is no difference between the average tenure of males and females.
- **H_1 (Alternative Hypothesis):** There is a difference between the average tenure of males and females.

Variables:
- **Gender:** Categorical variable (Male/Female)
- **Tenure:** Scale variable (number of months or years with the company)

Preliminary Step:
Before conducting the t-test, we must first check whether the data meets the assumptions for the test. These assumptions include:
1. **Independence of observations**
2. **Normality** of the distribution of tenure within each gender group
3. **Homogeneity of variances** (i.e., similar variances in tenure for both males and females).

Once these assumptions are verified, we can proceed with the t-test to determine whether the difference in average tenure between males and females is statistically significant.

Checking Assumption 1: Independence Assumption

The assumption of independence states that the two groups being compared must be independent of each other. In our analysis, the independent variable is **Gender**, which consists of two distinct, independent categorical groups: **Male** and **Female**.

In the **Teleco Customer Churn.sav** dataset, gender serves as the independent variable, with the two groups (Male and Female) being independent of one another, ensuring that this assumption is met for the t-test.

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	gender	N	Percent	N	Percent	N	Percent
tenure	Female	3488	100.0%	0	0.0%	3488	100.0%
	Male	3555	100.0%	0	0.0%	3555	100.0%

Checking Assumption 2: Normality Assumption

Skewness and Kurtosis Z-values:

- Female:
  - Skewness Z-value = 0.244 / 0.041 = 5.89
  - Kurtosis Z-value = -1.381 / 0.083 = -16.66
- Male:
  - Skewness Z-value = 0.235 / 0.041 = 5.72
  - Kurtosis Z-value = -1.394 / 0.082 = -16.97

Since all four Z-values fall outside the range of -1.96 to +1.96, the data **does not meet the assumption of normality**.

Report:

As the data does not meet the **Assumption of Normality**, which states that each category of the dependent variable (Tenure) should be approximately normally distributed.

Since the normality assumption is not satisfied, we cannot proceed with the independent samples t-test. Instead, we will consider using a **nonparametric test**, such as the **Mann-Whitney U test**, which does not require the assumption of normality and can still determine if there is a significant difference between the two groups.

Descriptives

	gender			Statistic	Std. Error
tenure	Female	Mean		32.24	.414
		95% Confidence Interval for Mean	Lower Bound	31.43	
			Upper Bound	33.06	
		5% Trimmed Mean		31.77	
		Median		29.00	
		Variance		598.450	
		Std. Deviation		24.463	
		Minimum		0	
		Maximum		72	
		Range		72	
		Interquartile Range		46	
		Skewness		.244	.041
		Kurtosis		-1.381	.083
	Male	Mean		32.50	.414
		95% Confidence Interval for Mean	Lower Bound	31.68	
			Upper Bound	33.31	
		5% Trimmed Mean		32.05	
		Median		29.00	
		Variance		607.936	
		Std. Deviation		24.656	
		Minimum		0	
		Maximum		72	
		Range		72	
		Interquartile Range		47	
		Skewness		.235	.041
		Kurtosis		-1.394	.082

Nonparametric test :Gender & Tenure variables

Report:

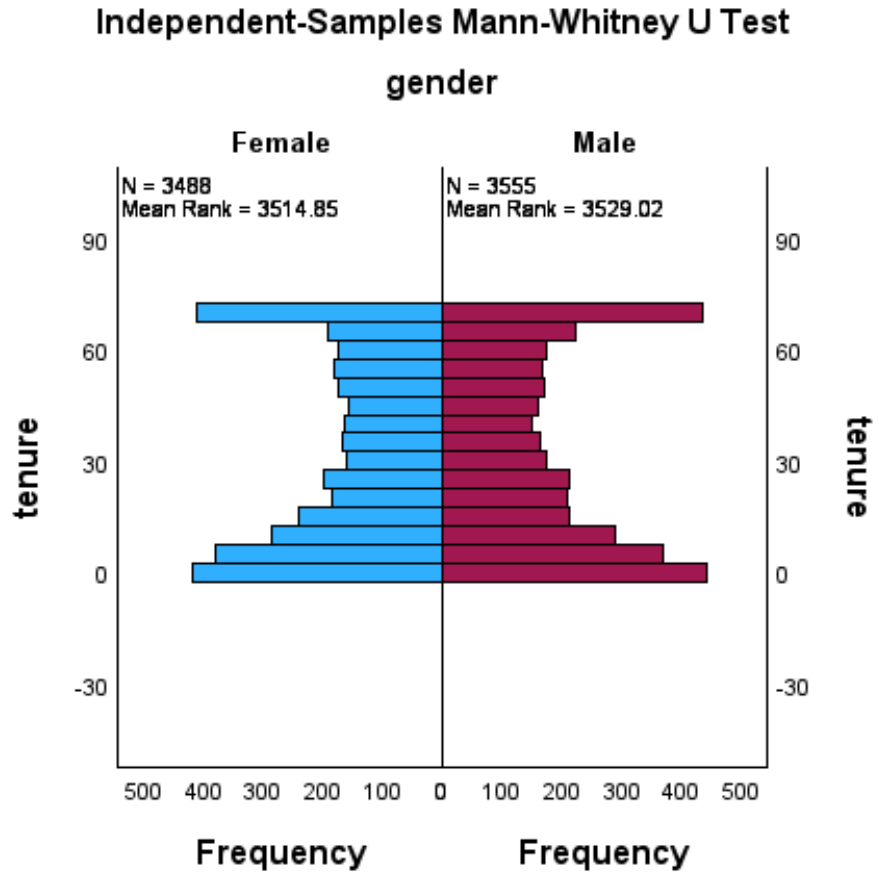
A nonparametric test was conducted to determine if there is a significant difference between the average tenure of males and females. In this analysis, we set the significance level (alpha) at 0.05. The resulting p-value from the nonparametric test was 0.770, which is greater than 0.05.

Conclusion:

Since the p-value exceeds the significance level, we fail to reject the null hypothesis. This indicates that there is no significant difference between the average tenure of males and females. Therefore, customer gender does not appear to have an impact on the average tenure.

Independent-Samples Mann-Whitney U Test Summary

Total N	7043
Mann-Whitney U	6224867.000
Wilcoxon W	12545657.000
Test Statistic	6224867.000
Standard Error	85272.591
Standardized Test Statistic	.293
Asymptotic Sig.(2-sided test)	.770



Independent Samples t-test for Churn & Tenure variables

Research Question:
Is there a significant difference between the average tenure of churned and non-churned customers?
Hypotheses:
- H_0 (Null Hypothesis): There is no difference between the average tenure of churned and non-churned customers.
- H_1 (Alternative Hypothesis): There is a difference between the average tenure of churned and non-churned customers.

Variables:
- Churn: Categorical variable (Churned vs. Non-churned)
- Tenure: Scale variable (Length of time as a customer)

This analysis aims to determine whether churn is associated with differences in the length of time customers have been with the company.

Preliminary Step:
Before conducting the t-test, we must first check whether the data meets the assumptions for the test. These assumptions include:
1. Independence of observations
2. Normality of the distribution of tenure within each gender group
3. Homogeneity of variances (i.e., similar variances in tenure for both males and females).

Once these assumptions are verified, we can proceed with the t-test to determine whether the difference in average tenure between males and females is statistically significant.

Checking Assumption 1: Independence Assumption

To meet the **Assumption of Independence**, we require two independent categorical groups that represent our independent variable (Churn).

In the **Teleco Customer Churn.sav** dataset, the independent variable is **Churn**, which consists of two independent categorical groups: **Yes** (churned) and **No** (not churned). These groups are independent of each other, fulfilling the assumption for conducting further analysis.

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	Churn	N	Percent	N	Percent	N	Percent
tenure	No	5174	100.0%	0	0.0%	5174	100.0%
	Yes	1869	100.0%	0	0.0%	1869	100.0%

Assumption 2: Assumption of normality

Skewness and Kurtosis Z-values Analysis:
For the **Un-churned** group (No):

- **Skewness Z-value** = -0.32 / 0.34 = -0.94
- **Kurtosis Z-value** = -1.414 / 0.068 = -20.79

For the **Churned** group (Yes):

- **Skewness Z-value** = 1.149 / 0.057 = 20.16
- **Kurtosis Z-value** = 0.196 / 0.113 = 1.73

Both the **skewness** and **kurtosis Z-values** fall outside the acceptable range of **-1.96 to +1.96**, indicating that the data **does not meet the assumption of normality**. This suggests that the distribution of tenure for both churned and non-churned customers is not normal.

Report:
the data does not meet the **Assumption of Normality**, which requires that each category of the dependent variable (Tenure) be approximately normally distributed.
Since the normality assumption is violated, we cannot apply the independent samples t-test. Instead, we will use a **nonparametric test**, such as the **Mann-Whitney U test**, which does not assume normality and can still assess whether there is a significant difference between the two groups.

Descriptives

	Churn			Statistic	Std. Error
tenure	No	Mean		37.57	.335
		95% Confidence Interval for Mean	Lower Bound	36.91	
			Upper Bound	38.23	
		5% Trimmed Mean		37.69	
		Median		38.00	
		Variance		581.474	
		Std. Deviation		24.114	
		Minimum		0	
		Maximum		72	
		Range		72	
		Interquartile Range		46	
		Skewness		-.032	.034
		Kurtosis		-1.414	.068
	Yes	Mean		17.98	.452
		95% Confidence Interval for Mean	Lower Bound	17.09	
			Upper Bound	18.87	
		5% Trimmed Mean		16.21	
		Median		10.00	
		Variance		381.465	
		Std. Deviation		19.531	
		Minimum		1	
		Maximum		72	
		Range		71	
		Interquartile Range		27	
		Skewness		1.149	.057
		Kurtosis		.196	.113

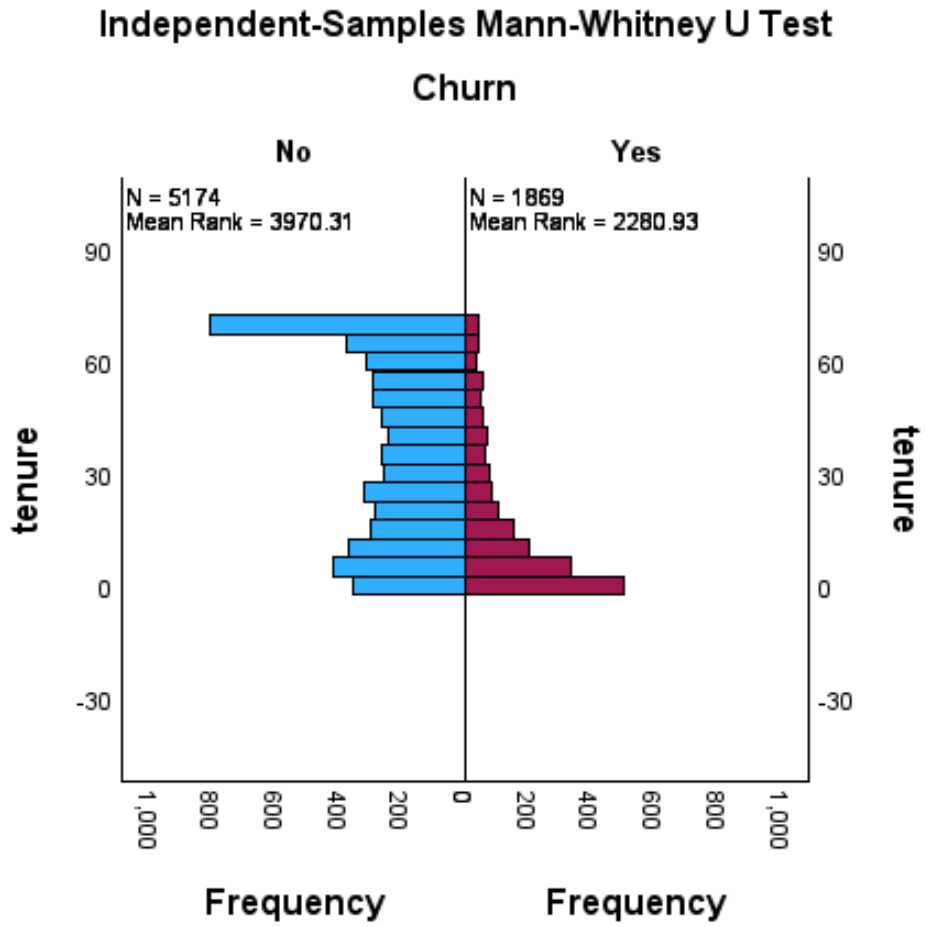
Nonparametric test : Churn and Tenure variables

Report:
A nonparametric test was conducted to determine if there is a significant difference between the average tenure of males and females. In this analysis, we set the significance level (alpha) at 0.05. The p-value obtained from the nonparametric test was 0.001, which is less than 0.05.

Conclusion:
Since the p-value is smaller than the significance level, we reject the null hypothesis and accept the alternative hypothesis. This indicates that there is a significant difference between the average tenure of churned and non-churned customers.

Independent-Samples Mann-Whitney U Test Summary

Total N	7043
Mann-Whitney U	2515538.000
Wilcoxon W	4263053.000
Test Statistic	2515538.000
Standard Error	75304.214
Standardized Test Statistic	-30.803
Asymptotic Sig.(2-sided test)	<.001



Correlation Between Numerical Variables

The results of Pearson Correlation (PC) indicate:

- Strong relationship between tenure and Total Charges as Pearson Correlation (PC) = 0.826 > 0.5 but < 0.9
- Strong relationship between Monthly Charges and Total Charges (PC= 0.651)
- Weak relationship between tenure and Monthly Charges as (PC = 0.248 < 0.3)

Correlations

		tenure	MonthlyCharges	TotalCharges
tenure	Pearson Correlation	1	.248**	.826**
	Sig. (2-tailed)		<.001	<.001
	N	7043	7043	7032
MonthlyCharges	Pearson Correlation	.248**	1	.651**
	Sig. (2-tailed)	<.001		<.001
	N	7043	7043	7032
TotalCharges	Pearson Correlation	.826**	.651**	1
	Sig. (2-tailed)	<.001	<.001	
	N	7032	7032	7032

**. Correlation is significant at the 0.01 level (2-tailed).

According to Correlation Interpretation: Cohen Table(1992)

Correlation Coefficient Value	Relationship
-0.3 to + 0.3	Weak
- 0.3 to - 0.5 or 0.3 to 0.5	Moderate
- 0.9 to – 0.5 or 0.5 to 0.9	Strong
-1.0 to – 0.9 or 0.9 to 1.0	Very Strong

Checking Regression Assumptions

Assumptions for Regression are:

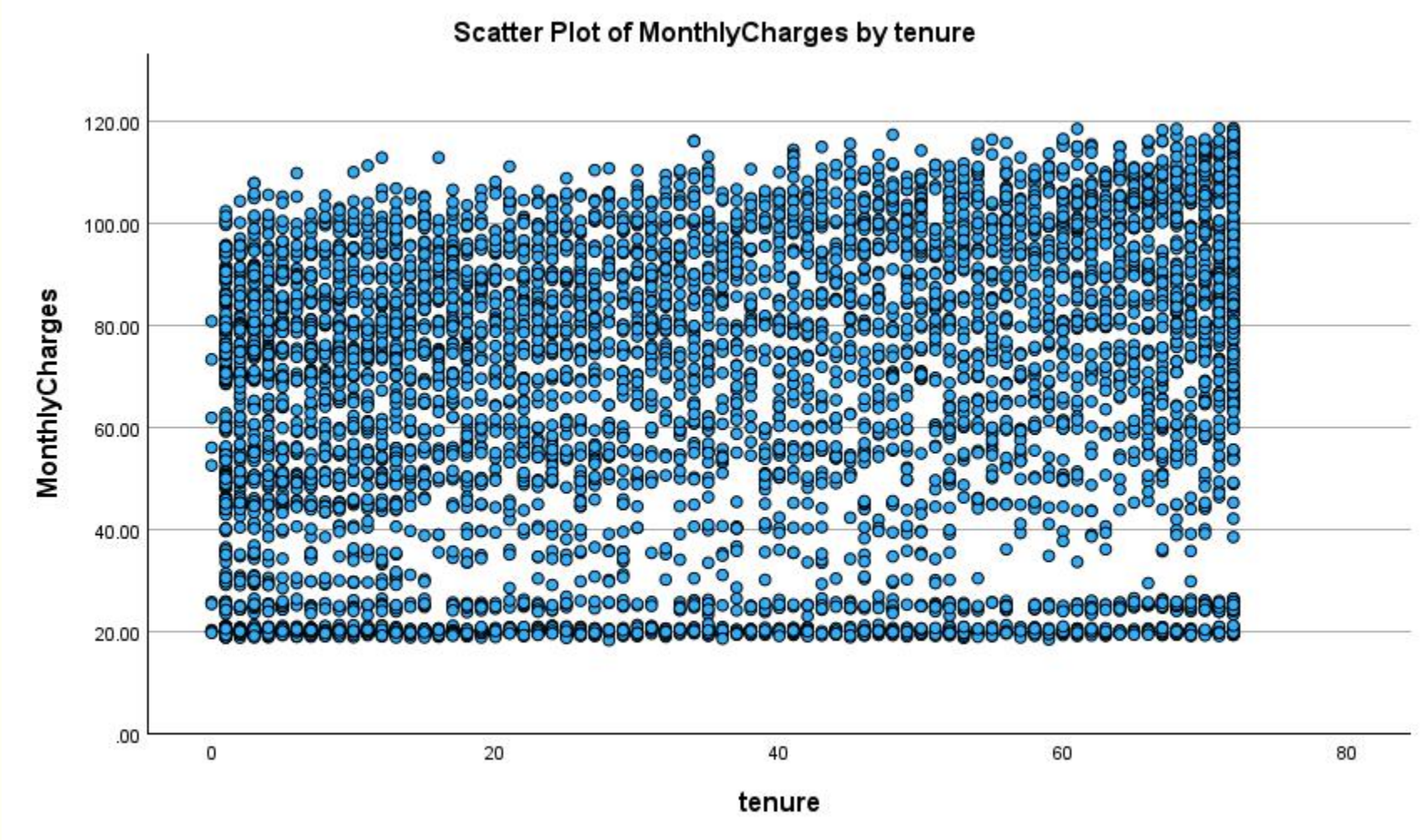
Assumption	Plot to check
The relationship between the independent and dependent variables is linear.	Original scatter plot of the independent and dependent variables
Homoscedasticity: when you plot the individual residual against the predicted value, the variance with respect to each predicted value should be constant.	Scatter plot of standardized predicted values and residuals
The residuals are independently normally distributed	Plot the residuals in a histogram

- independent variable: tenure
- dependent variables: Monthly Charges and Total Charges respectively

Assumption 1 for Regression: Linearity Between Tenure and Monthly Charges

The assumption of linearity states that there should be a linear relationship between the independent variable (Tenure) and the dependent variable (Monthly Charges). This can be assessed through a scatterplot.

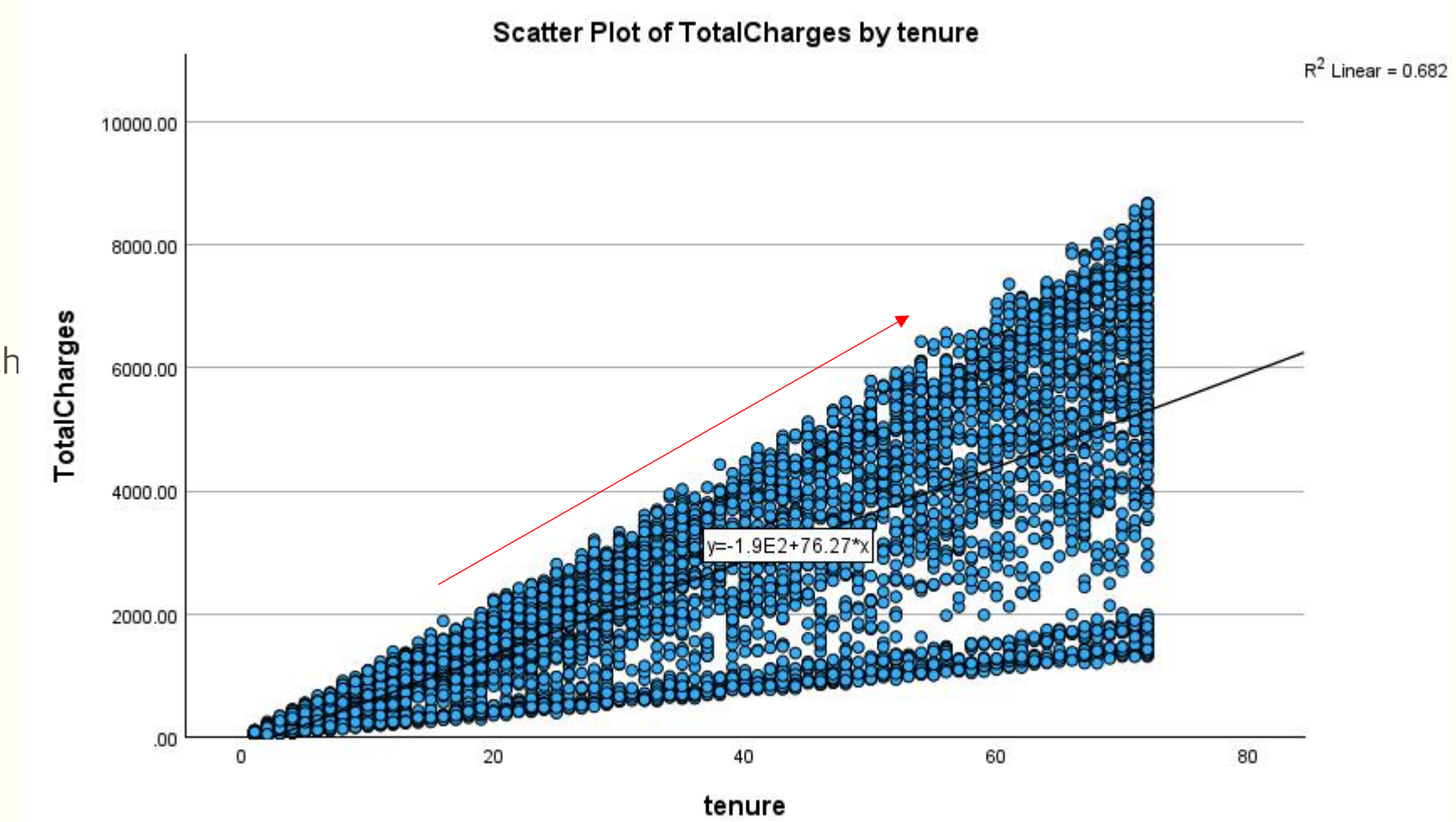
Report:
The scatterplot reveals a nonlinear relationship between Tenure and Monthly Charges, suggesting that the relationship is not linear. Therefore, the assumption of linearity is not met, and alternative approaches, such as data transformation or nonlinear regression, may be necessary to proceed with the analysis.



Assumption 1 for Regression: Linearity Between Tenure and Total Charges Variables

Checking Assumption 1 for Regression:
The relationship between the independent variable (Tenure) and the dependent variable (Total Charges) is linear."

Report:
The relationship between Tenure and Total Charges is **not linear**, as observed in the analysis. Therefore, the assumption of linearity is **not met**, indicating that a different approach, such as data **transformation or nonlinear regression**, may be needed for further analysis.



Key Findings and Recommendations

Project Objectives	Key Findings	Visualized Patterns and Trends	Recommended
Identify Drivers of Churn **And** Profile High-Risk Customers	• Customers with **month-to-month contracts** have significantly higher churn rates.	• **Churn rates are highest** for customers on flexible, short-term contracts.	• **Incentivize long-term contracts** by offering loyalty programs or discounts.
	• **Senior citizens** have a higher likelihood of churn.	• Senior citizens exhibit a **42% churn rate**, nearly **double** that of other groups.	• Provide **tailored services** for senior citizens, including simplified plans.
	• Customers with **higher monthly charges** are more likely to churn.	• **Strong correlation** between higher charges and churn (boxplots).	• Offer **discounts or value-added services** to high-paying customers.
	• Customers with **fiber optic service** show higher churn rates compared to other service types.	• Fiber optic users' churn patterns suggest **potential issues** (quality or pricing.)	• **Investigate** churn drivers for fiber optic services (e.g.cost or performance).
	• Customers without **partners** or **dependents** churn more frequently.	• **Churn rates**: 33% (no partner) vs. 20% (with partner); 31% (no dependents) vs. 15%.	• Design **targeted strategies** to address their specific **needs and concerns**(Discounts, Flexible contract, Value-added services…..)
	• Strong **nonlinear relationship** between Tenure and Total Charges	• Scatterplots show a **positive trend**: Total Charges increase with Tenure.	• Target low-tenure customers with **early intervention strategies**.

THANKS, YOU FOR YOUR LISTENING!

ANY QUESTIONS?