Customer Churn Analysis Deck

AI With Data Analytics Bootcamp
(Codecademy)

Marian Pattugalan

Problem Statement

Objective: Conduct an Exploratory Data Analysis (EDA) on a customer churn dataset to uncover insights.

Key Tasks:

1. Data Cleaning: Ensure the dataset is clean and ready for analysis.
2. Categorical Analysis: Analyze and visualize categorical variables to identify trends.
3. Numerical Analysis: Explore numerical variables and visualize their distributions.
4. Insights: Highlight key findings related to customer churn.

Data Overview

customerI	gender	SeniorCitiz	Partner	Dependen	tenure	PhoneSer	MultipleLi	InternetSe	OnlineSec	OnlineBac	DevicePro	TechSuppo	Streaming	Streaming	Contract	PaperlessB	PaymentM	MonthlyCl	TotalCharg	Churn
7590-VHV	Female	0	Yes	No	1	No	No phone	DSL	No	Yes	No	No	No	No	Month-to-	Yes	Electronic	29.85	29.85	No
5575-GNV	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed che	56.95	1889.5	No
3668-QPYI	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-	Yes	Mailed che	53.85	108.15	Yes
7795-CFO(	Male	0	No	No	45	No	No phone	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank trans	42.3	1840.75	No
9237-HQIT	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-	Yes	Electronic	70.7	151.65	Yes
9305-CDSI	Female	0	No	No	8	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-	Yes	Electronic	99.65	820.5	Yes
1452-KIOV	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	No	Month-to-	Yes	Credit carc	89.1	1949.4	No
6713-OKO	Female	0	No	No	10	No	No phone	DSL	Yes	No	No	No	No	No	Month-to-	No	Mailed che	29.75	301.9	No
7892-POO	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	No	Yes	Yes	Yes	Yes	Month-to-	Yes	Electronic	104.8	3046.05	Yes
6388-TABC	Male	0	No	Yes	62	Yes	No	DSL	Yes	Yes	No	No	No	No	One year	No	Bank trans	56.15	3487.95	No
9763-GRSI	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No	No	No	No	No	Month-to-	Yes	Mailed che	49.95	587.45	No
7469-LKBC	Male	0	No	No	16	Yes	No	No	No interne	No interne	No interne	No interne	No interne	No interne	Two year	No	Credit carc	18.95	326.8	No
8091-TTVA	Male	0	Yes	No	58	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	One year	No	Credit carc	100.35	5681.1	No
0280-XJGE	Male	0	No	No	49	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Yes	Month-to-	Yes	Bank trans	103.7	5036.3	Yes
5129-JLPIS	Male	0	No	No	25	Yes	No	Fiber optic	Yes	No	Yes	Yes	Yes	Yes	Month-to-	Yes	Electronic	105.5	2686.05	No
3655-SNQ	Female	0	Yes	Yes	69	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	Yes	Yes	Two year	No	Credit carc	113.25	7895.15	No
8191-XWS	Female	0	No	No	52	Yes	No	No	No interne	No interne	No interne	No interne	No interne	No interne	One year	No	Mailed che	20.65	1022.95	No
9959-WOF	Male	0	No	Yes	71	Yes	Yes	Fiber optic	Yes	No	Yes	No	Yes	Yes	Two year	No	Bank trans	106.7	7382.25	No
4190-MFL	Female	0	Yes	Yes	10	Yes	No	DSL	No	No	Yes	Yes	No	No	Month-to-	No	Credit carc	55.2	528.35	Yes
4183-MYF	Female	0	No	No	21	Yes	No	Fiber optic	No	Yes	Yes	No	No	Yes	Month-to-	Yes	Electronic	90.05	1862.9	No
8779-QRD	Male	1	No	No	1	No	No phone	DSL	No	No	Yes	No	No	Yes	Month-to-	Yes	Electronic	39.65	39.65	Yes
1680-VDC	Male	0	Yes	No	12	Yes	No	No	No interne	No interne	No interne	No interne	No interne	No interne	One year	No	Bank trans	19.8	202.25	No

This mixed-data type dataset contains 7,043 rows of customer information from a telecommunication company. It also has 20 columns covering demographics, services, and financials which will be important to compare against the churn column.

Approach

Phase 1: Understanding the Business Objective

- Define the objectives of the problem statement provided to ensure clarity with the end-goal of this project

Phase 2: Setting up the Data

- Load the .csv file into Google Collab, assign it to a dataframe, and understand the schema of the dataset

Phase 3: Data Cleaning

- Check the data types of each column, ensuring they are correct (e.g. customerID = object)
- Find any null values and impute the null values (e.g. with mean)
- Check for any replicated data
- Drop any irrelevant columns

Phase 4: Exploratory Data Analysis (EDA)

- Use EDA techniques to uncover patterns, trends, and relationships between columns of the dataset that may explain the churn rate occurring within the company.

Project Workflow

Understand business objectives

Define the objectives to ensure the goal of the project is clear

Understand data schema

Understand how each column interacts with one another

Perform data cleaning

Check data-types, any null values, duplicates, or errors

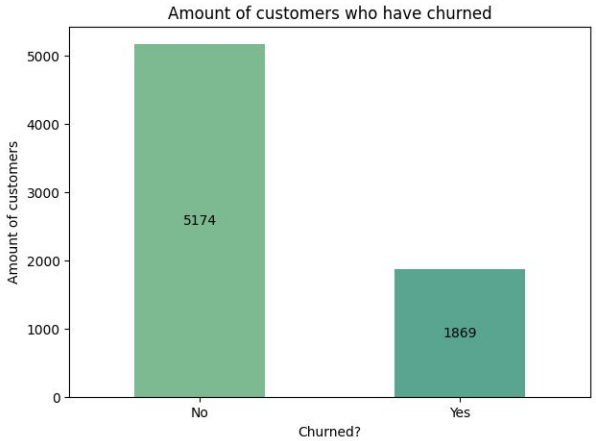
Exploratory Data Analysis (EDA)

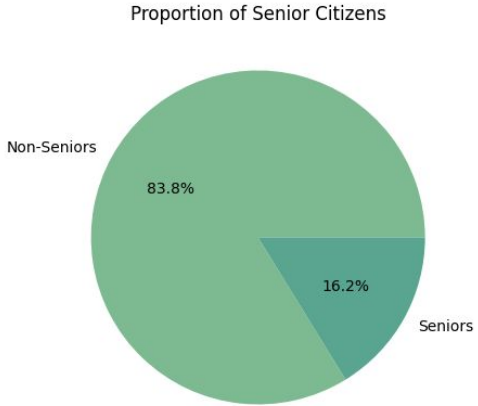
Perform EDA on the dataframe to extract insights and create visualisations for the presentation deck.

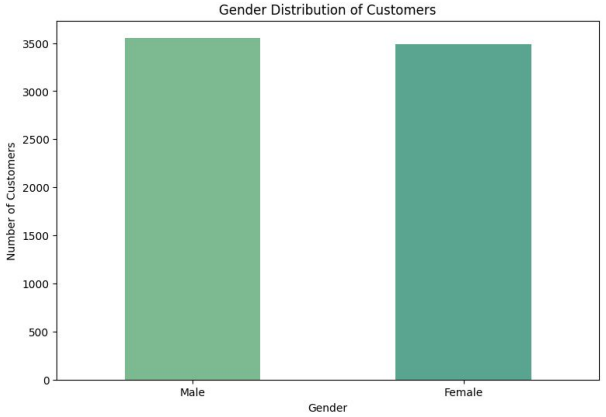
Univariate Analysis

FINDINGS:

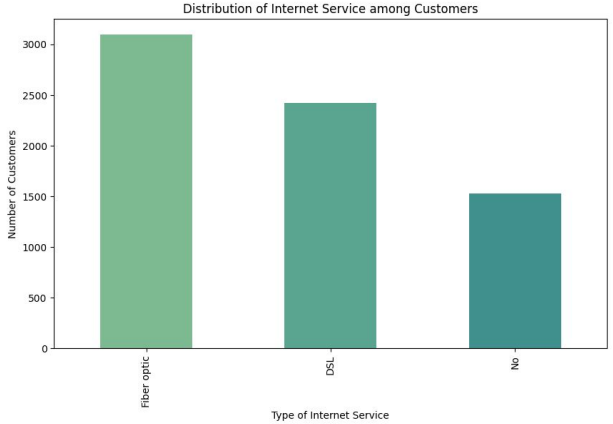
- The data shows that there is a 26.5% overall churn rate.
- Data is comprised of 5,901 (83.8%) non-senior citizen customers and 1,142 (16.2%) senior citizen customers.
- Gender distribution between customers is almost 1:1 ratio.

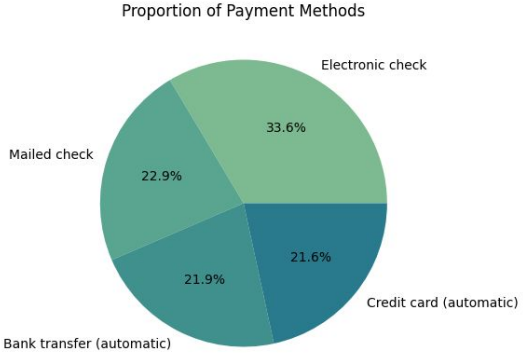






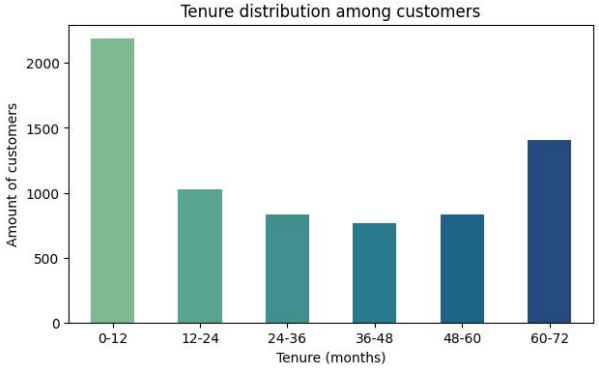
Univariate Analysis

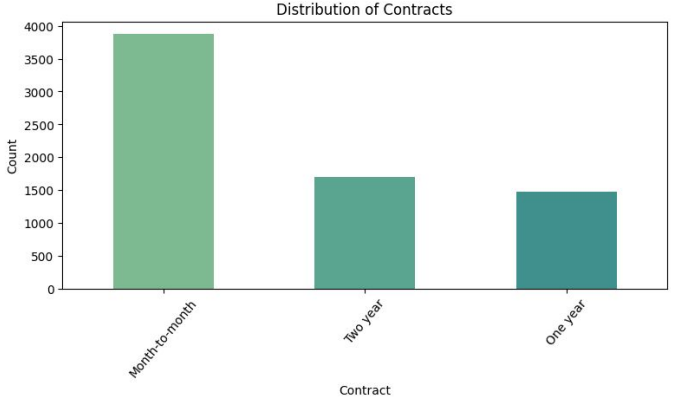




FINDINGS:

- Electronic checks are the most popular payment method.
- Fibre optic is the preferred internet service, but there is over 1,500 customers without any type of service.
- There is a bimodal distribution of customers based on their tenure: 0-12 months and 60-72 months are the peaks.
- Month-to-month contracts are favoured the most.



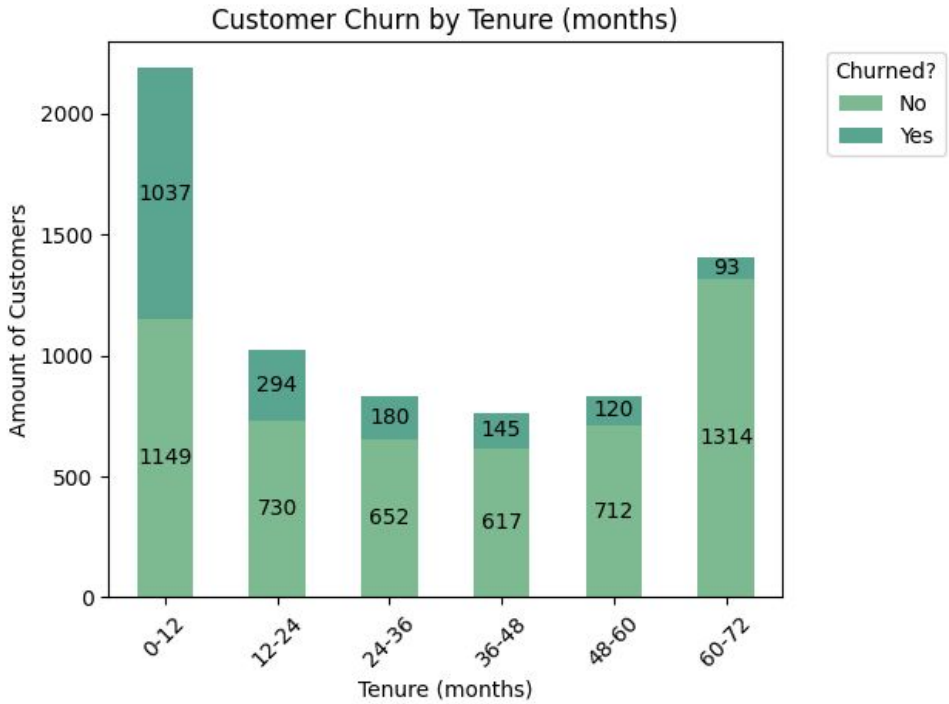


Most of these generated insights show the distribution of customers and what they have. However, these are not insightful enough to explain what could be contributing to the customer churn rate.

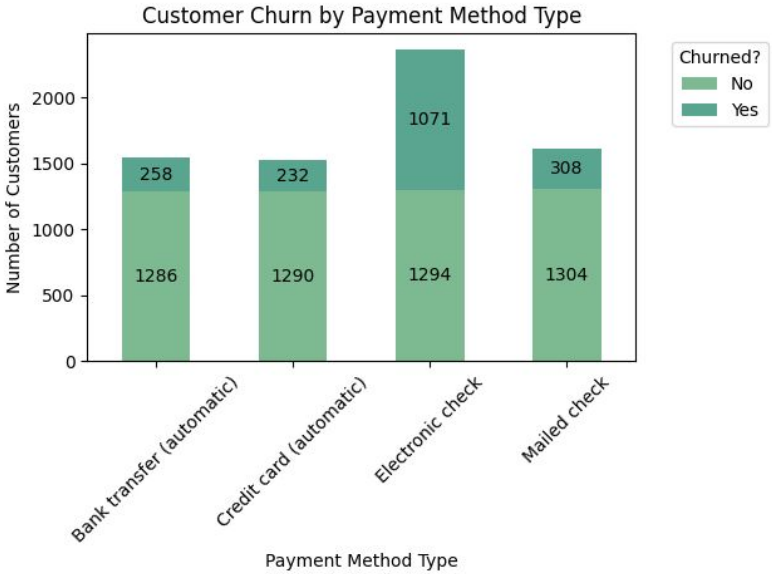
Bivariate Analysis

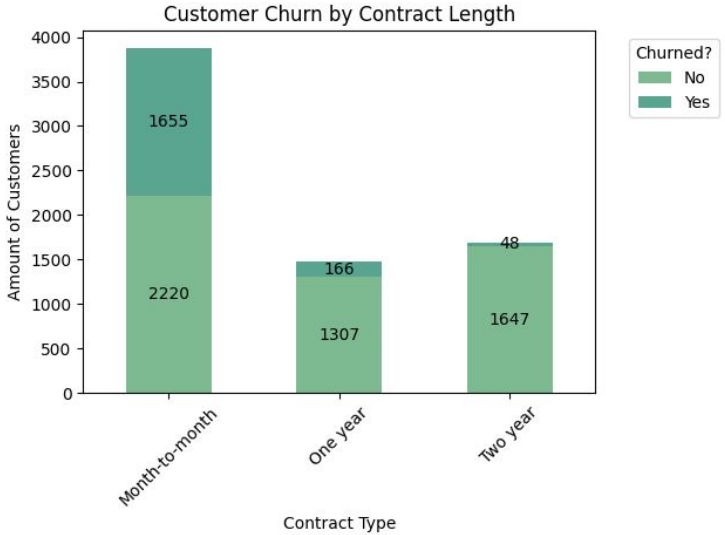
FINDINGS:

- Churn rates (from 47% to 6%) steadily declines the longer the customer's tenure is.
- Most customers churn within the first 12 months, indicating that new customers are at significant risk of leaving.
- Conversely, customers with the longer tenure are more likely to stay.



Bivariate Analysis





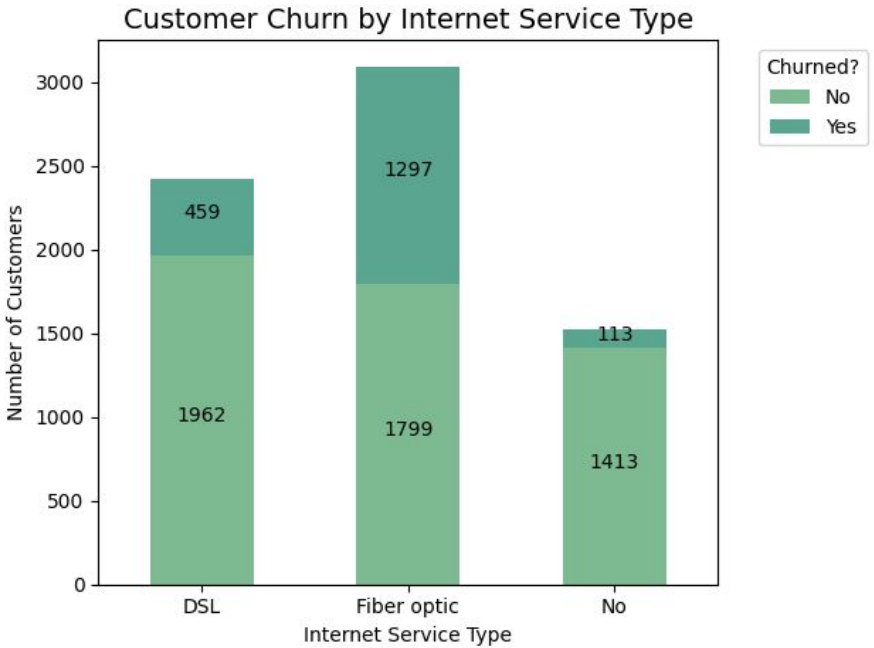
FINDINGS:

- Non-automated payment methods yield the highest churn rates: electronic check (45%) and mailed check (19%). Meanwhile, the automated methods only have a 15-16% churn rate.
- Monthly customers are significantly more likely to churn (42%) than customers who have paid for one-year (11%) and two-year (2%) contracts.

Bivariate Analysis

FINDINGS:

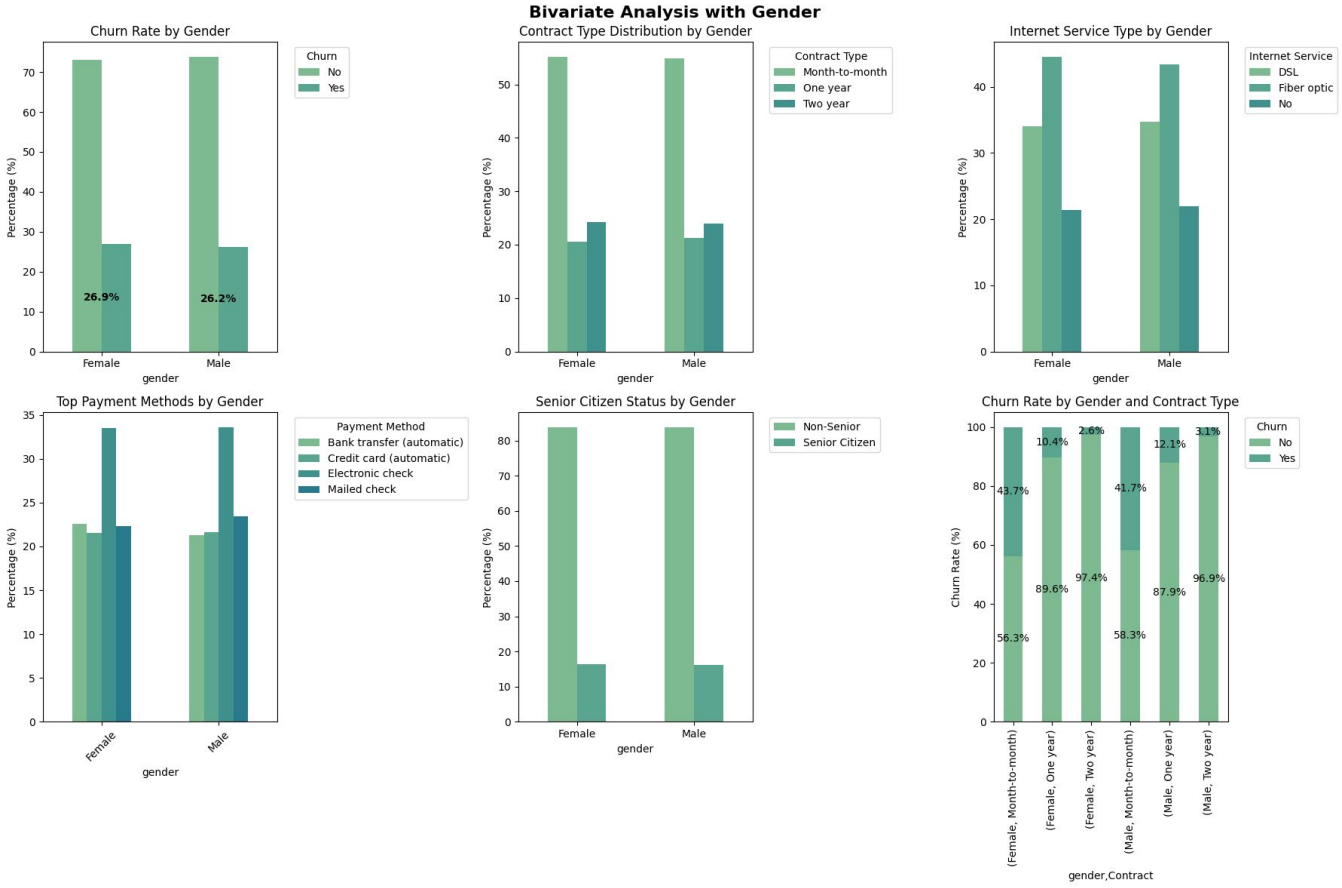
- Despite fiber optic being the most popular internet service, it also yields the highest churn rate (41)%).
- DSL is the least preferred internet service, but it has a lower churn rate (18%).
- Customers who do not receive their internet service from this telecommunication company have the lowest churn rate (7%).
- This suggests that there must be dissatisfaction with the service and/or price provided with the internet services which prompts customers to leave.



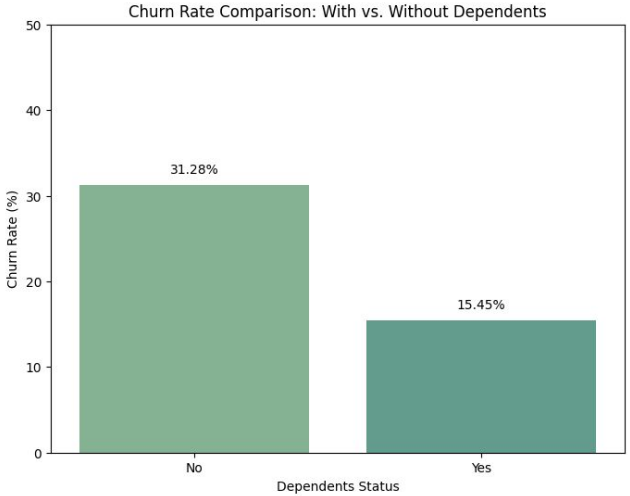
Bivariate Analysis

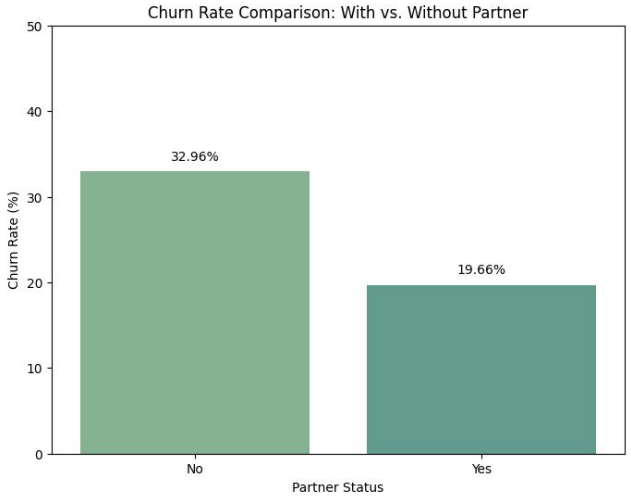
FINDINGS:

- Shows that the gender distribution among customers across different fields are equally distributed.
- Male churn rate (26.1%) Female churn rate (26.9%) are close to 1:1 ratio.
- Senior citizen distribution has an even distribution:
    - Males: 16.1%
    - Females: 16.3%
- Internet Service Type and Payment Method have an even distribution between genders
- The average monthly charges between males and females are comparable.
- Churn rates by gender and contract type are also similar between genders, showing minimal difference.



Bivariate Analysis

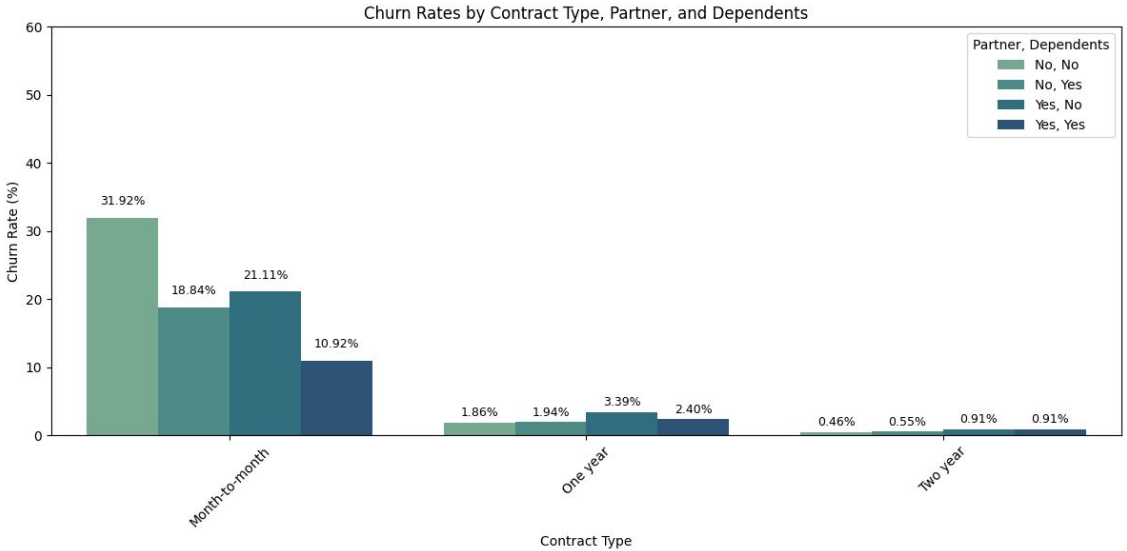




FINDINGS:

- Customers with dependents have a significantly lower churn rate compared to customers without any dependent.
- Similarly, customers with partners are less likely to leave compared to customers without partners. However, customers with dependents are more likely to stay over customers with partners.
- Lower churn rates may be due to compounded obligations to family and/or partners.

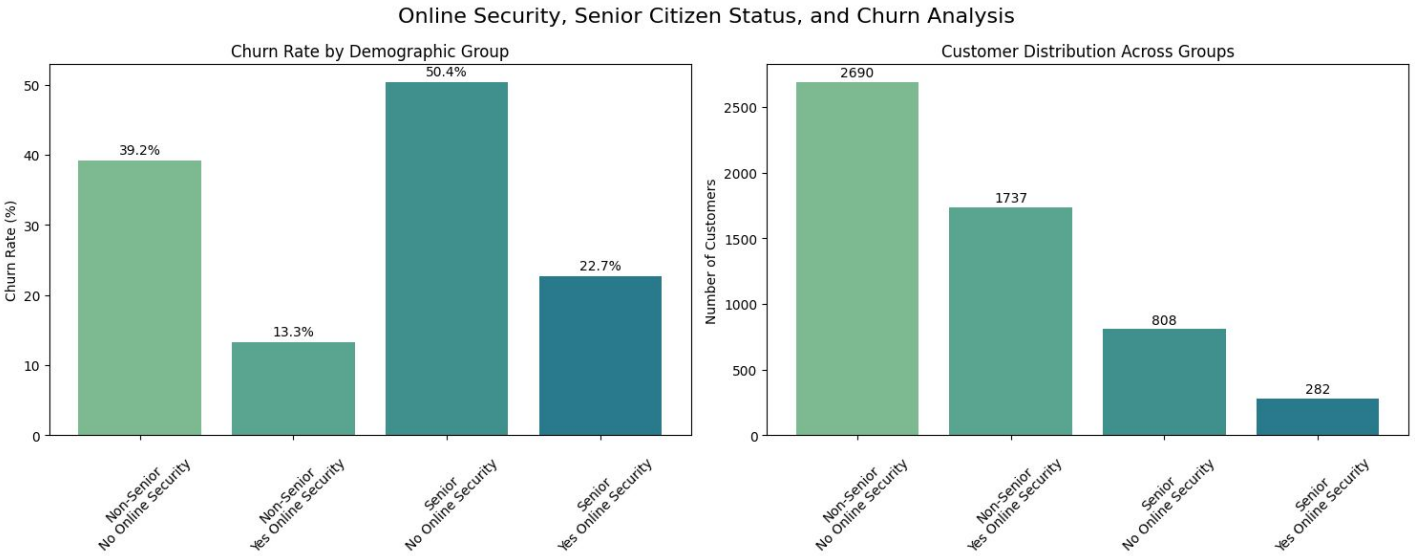
Bivariate Analysis



FINDINGS:

- Customers without partners and dependents (3,280) are more likely to choose the month-to-month flexible contract (69% of total) and leave within 0-12 months.
- Customers with both partners and dependents (1,749) are more likely to choose the two-year (41% of total) or month-to-month (33% of total) contract over the one-year plan.
- In all contract types, customers with partners and/or dependents are less likely to churn compared customers who are single without dependents.
- Churn rates significantly decreases for every type of customer (single, in relationship, with family) in the one-year and two-year plans.

Bivariate Analysis



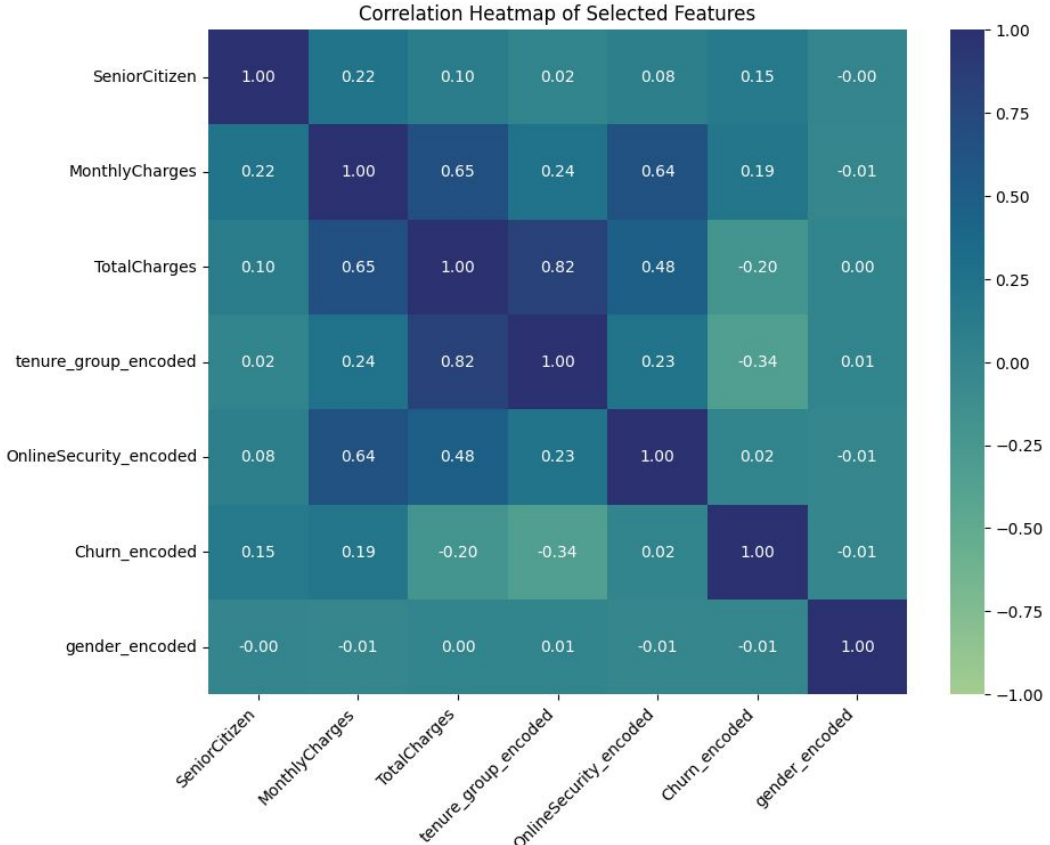
FINDINGS:

- Senior Citizens without online security have the highest churn rate (50%) and are more than twice more likely to churn compared to Senior Citizens *with* online security.
- Similarly, Non-Seniors without online security are also more likely to churn compared to Non-Seniors with online security, indicating that add-ons may be important for customer retention.

Multivariate Analysis

FINDINGS:

- Strong positive correlations between:
  - Total Charges and Tenure Group (0.82): as expected, the longer the customer stays with the company, the higher the total charges accumulated.
  - Monthly Charges and Online Security (0.64): customers who pay online security have higher monthly fees
- Weak positive correlation between:
  - Senior Citizen and Churn (0.15): senior citizens are slightly more likely to churn
- Negative correlation between:
  - Tenure Group and Churn (-0.34): customers with longer tenure are less likely to churn
- Weak negative correlation between:
  - Online Security and Churn (-0.01): having online security has minimal impact on reducing churn
- No correlation
  - Gender and other variables: gender has little to no impact on other variables, including churn



Key Insights

- Tenure and Contract Type are major predictors of churn as new customers with shorter contracts are more likely to churn. Customers with longer tenure and one-year or two-year contracts have remarkably lower churn rates.
- Gender is equally distributed across all aspects, including churn rate, and there is no significant difference between the two.
- Individuals using an electronic check are disproportionately represented in the churn category. Conversely, automated payment methods yields greater customer retention.
- There is only a small demographic of seniors in this sample, but their churn rate is impacted by the presence of online security, with senior citizens having online security being less likely to churn.
- Add-on services, such as online security, also enhance customer loyalty for the general demographic (39.2% churn rate without online security vs 13.3% churn rate *with* online security).
- Customers with the fiber optic internet service are more likely to churn compared to DSL and customers without any internet service, indicating dissatisfaction with the service given.
- Single, non-family oriented customers are more likely to opt for the flexible contract and are more likely to churn.
- Customers with partners and/or dependents are less likely to churn, indicating customer loyalty compounded by social obligations.

Recommendations

- Conduct an investigation into the internet services provided (reliability and customer service) and improve the quality provided to customers. Additionally, run promotions and match competitors' prices.
- Promote automated payment methods over electronic checks for seamless transactions.
- Market online security services during the onboarding process of customers, highlighting the importance of the service for safe online activity.
- Encourage customers to move to longer-term contracts by offering substantial discounts for one-year and two-year contracts and/or offering bundles with add-on services (e.g. internet service or online security).
- Additionally, launch targeted offers to customers with dependents but are on the month-to-month plan, offering family-oriented benefits with the longer term plans.
- Offer a customer satisfaction questionnaire to help improve overall customer service and quality of service, thus improving customer loyalty.