# Telecom users

## Overview of the dataset

The provided Excel file contains a comprehensive dataset capturing vital information about customers and their interactions with a specific service provider. Each row corresponds to a unique customer entry, while the columns encompass various attributes crucial for understanding customer behaviour and preferences.
The dataset begins by categorising customers based on gender and whether they are senior citizens. It delves into their relational status, indicating whether they have partners or dependents. The "tenure" column reveals the duration of the customer's subscription in months, providing insights into their loyalty.
Services such as phone service, internet connectivity, online security, backup, device protection, and tech support are meticulously documented. The dataset also delves into customers' entertainment preferences, detailing their streaming TV and movie services usage.
Crucial financial and contractual aspects are covered, including the type of contract, payment methods, monthly charges, and the cumulative total charges over the customer's engagement. Notably, the dataset incorporates a binary "Churn" indicator, signifying whether a customer has terminated their subscription.
This rich dataset is valuable for conducting diverse analyses, including customer segmentation, churn prediction, and understanding service utilisation patterns.

## Overview of the tool used

Python, a widely-used high-level programming language, emerges as a pivotal tool tailored for in-depth analysis of the mentioned dataset. Its robust libraries are indispensable, including NumPy for numerical operations, Pandas for seamless data manipulation, and Matplotlib plus Seaborn for compelling visualisations. Scikit-Learn empowers the creation of predictive models, while Jupyter Notebooks facilitate interactive exploration and documentation. Python's intuitive syntax accommodates diverse users, from novices to experts. It also encompasses tools like Requests and Beautiful Soup for web data extraction and libraries like TensorFlow and PyTorch for diving into intricate machine-learning aspects. Python's readability and active community offer an ideal environment for dissecting the dataset comprehensively, extracting insights, and predicting trends across various demographics and usage patterns.

## Uploading the Excel file

This code snippet utilises Google Colab and Pandas to streamline data analysis. After refining the Excel file by removing blank columns and adjusting the format, we uploaded the enhanced 'telecom_users_new.xlsx' file using `files.upload()`. With `pd.read_excel()`, we load the data into a Pandas DataFrame.

```
from google.colab import files
import pandas as pd

uploaded = files.upload()

df = pd.read_excel(uploaded['telecom_users_new.xlsx'])


print(df)
```

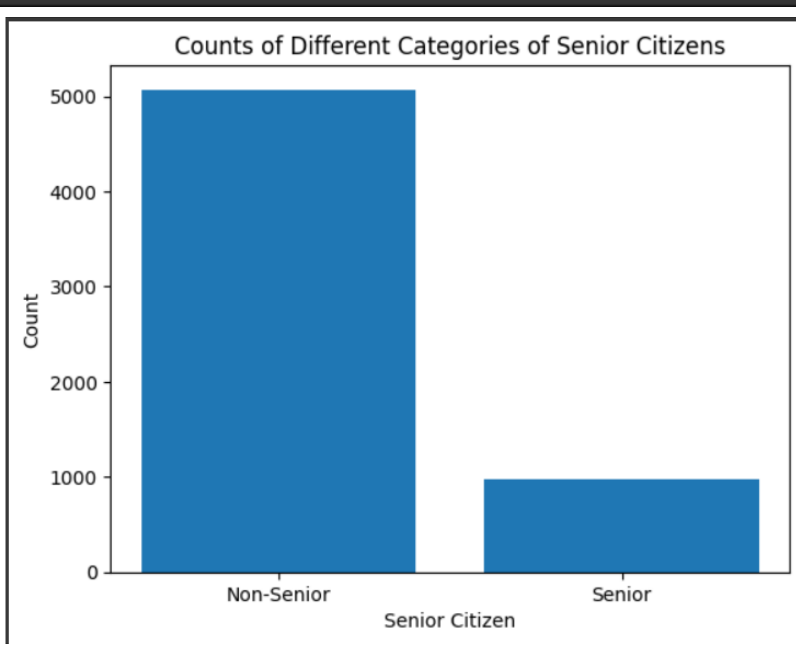
## Senior Citizens counts

```
SeniorCitizen_count = df['SeniorCitizen'].value_counts()
print(SeniorCitizen_count)
```

```
0    5069
1     981
Name: SeniorCitizen, dtype: int64
```

```
SeniorCitizen_count = df['SeniorCitizen'].replace({0: 'Non-Senior', 1: 'Senior'}).value_counts()
print(SeniorCitizen_count)
```

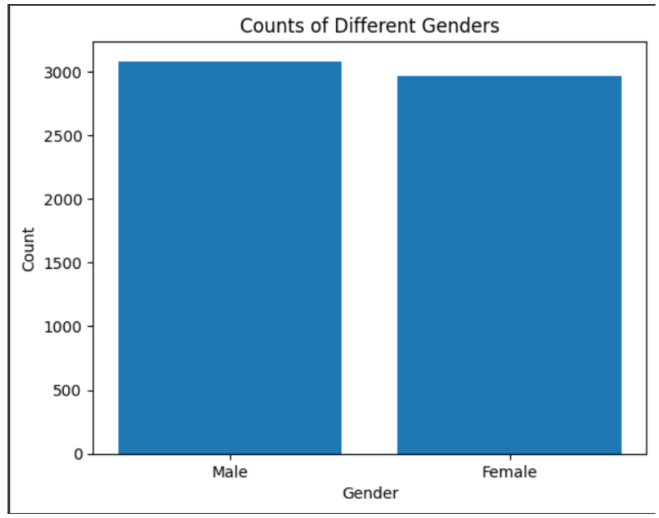
```
Non-Senior    5069
Senior         981
Name: SeniorCitizen, dtype: int64
```



## Gender counts

```
gender_count = df['gender'].value_counts()
print(gender_count)
```

```
Male      3081
Female    2969
Name: gender, dtype: int64
```




## Total count of Male and Female non Senior citizens

### a.Male

```
total_notseniorCitizen_male = df[(df['gender'] == 'Male') & (df['SeniorCitizen'] == 0)].shape[0]
print(total_notseniorCitizen_male)
```

```
2591
```

### b.Female

```
total_notseniorCitizen_Female = df[(df['gender'] == 'Female') & (df['SeniorCitizen'] == 0)].shape[0]
print(total_notseniorCitizen_Female)
```

```
2478
```

## Total Male Subscribers (not Senior Citizens) with the following

### a. Phone Service

```
total_notseniorCitizen_male_with_PhoneService = df[(df['gender'] == 'Male') &
                                                    (df['SeniorCitizen'] == 0) &
                                                     (df['PhoneService'] == 'Yes')].shape[0]
print(total_notseniorCitizen_male_with_PhoneService)
```

```
2336
```

b. Internet Service

```
total_notseniorCitizen_male_with_InternetService = df[(df['gender'] == 'Male') &
                                                    (df['SeniorCitizen'] == 0) &
                                                     (df['InternetService'] != 'No')].shape[0]
print(total_notseniorCitizen_male_with_InternetService)
```

```
1947
```

c. Device Protection

```
total_notseniorCitizen_male_with_DeviceProtection = df[(df['gender'] == 'Male') &
                                                    (df['SeniorCitizen'] == 0) &
                                                     (df['DeviceProtection'] != 'Yes')].shape[0]
print(total_notseniorCitizen_male_with_DeviceProtection)
```

```
1733
```

d. Streaming TV

```
total_notseniorCitizen_male_with_StreamingTV = df[(df['gender'] == 'Male') &
                                                    (df['SeniorCitizen'] == 0) &
                                                     (df['StreamingTV'] != 'Yes')].shape[0]
print(total_notseniorCitizen_male_with_StreamingTV)
```

```
1658
```

e. Paperless billing

```
total_notseniorCitizen_male_with_PaperlessBilling = df[(df['gender'] == 'Male') &
                                                    (df['SeniorCitizen'] == 0) &
                                                     (df['PaperlessBilling'] != 'Yes')].shape[0]
print(total_notseniorCitizen_male_with_PaperlessBilling)
```

```
1167
```

## Total Female Subscribers (who are not Senior Citizens)

a. Phone Service

```
total_notseniorCitizen_female_with_PhoneService = df[(df['gender'] == 'Female') &
                                                    (df['SeniorCitizen'] == 0) &
                                                     (df['PhoneService'] == 'Yes')].shape[0]
print(total_notseniorCitizen_female_with_PhoneService)
```

```
2224
```

b. Internet Service

```
total_notseniorCitizen_female_with_InternetService = df[(df['gender'] == 'Female') &
                                                       (df['SeniorCitizen'] == 0) &
                                                        (df['InternetService'] != 'No')].shape[0]
print(total_notseniorCitizen_female_with_InternetService)
```

```
1863
```

c. Device Protection

```
total_notseniorCitizen_female_with_DeviceProtection= df[(df['gender'] == 'Female') &
                                                       (df['SeniorCitizen'] == 0) &
                                                        (df['DeviceProtection'] != 'Yes')].shape[0]
print(total_notseniorCitizen_female_with_DeviceProtection)
```

```
1661
```

d. Streaming TV

```
total_notseniorCitizen_female_with_StreamingTV= df[(df['gender'] == 'Female') &
                                                  (df['SeniorCitizen'] == 0) &
                                                   (df['StreamingTV'] != 'Yes')].shape[0]
print(total_notseniorCitizen_female_with_StreamingTV)
```

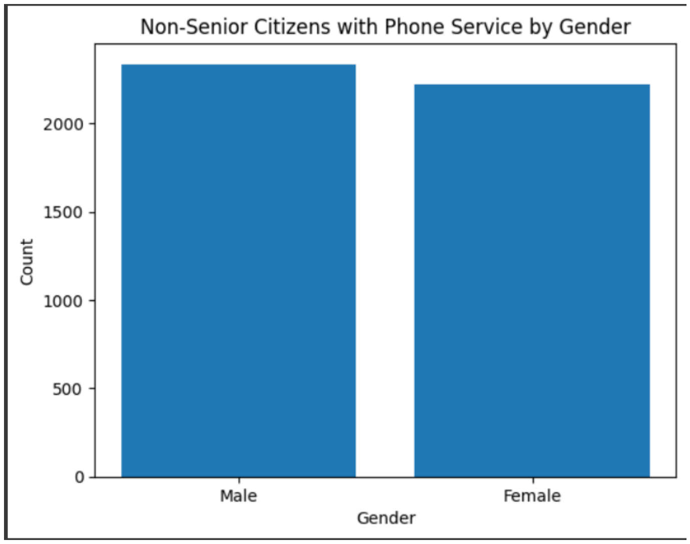
```
1571
```

e. Paperless billing

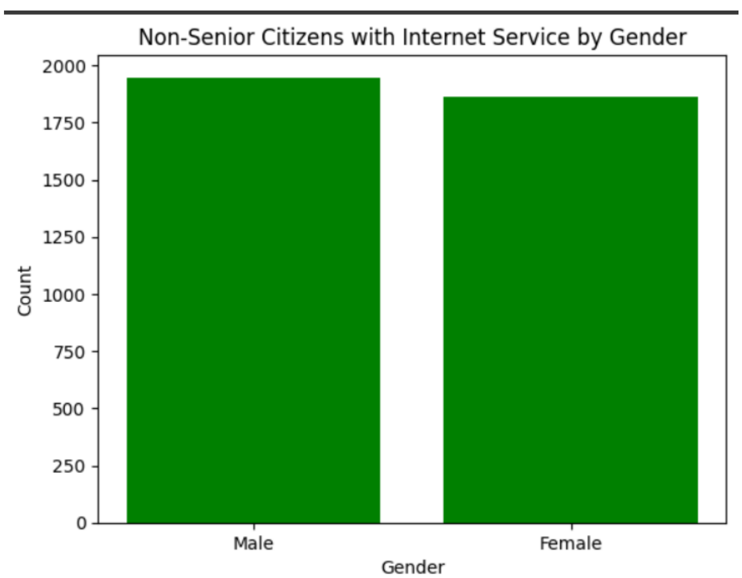
```
total_notseniorCitizen_female_with_PaperlessBilling= df[(df['gender'] == 'Female') &
                                                       (df['SeniorCitizen'] == 0) &
                                                        (df['PaperlessBilling'] != 'Yes')].shape[0]
print(total_notseniorCitizen_female_with_PaperlessBilling)
```

```
1077
```

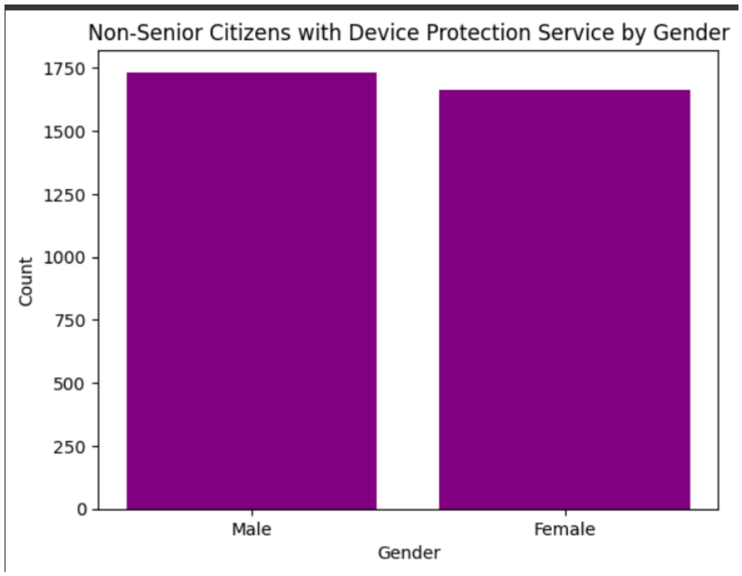
Non Senior citizens with Phone service by gender



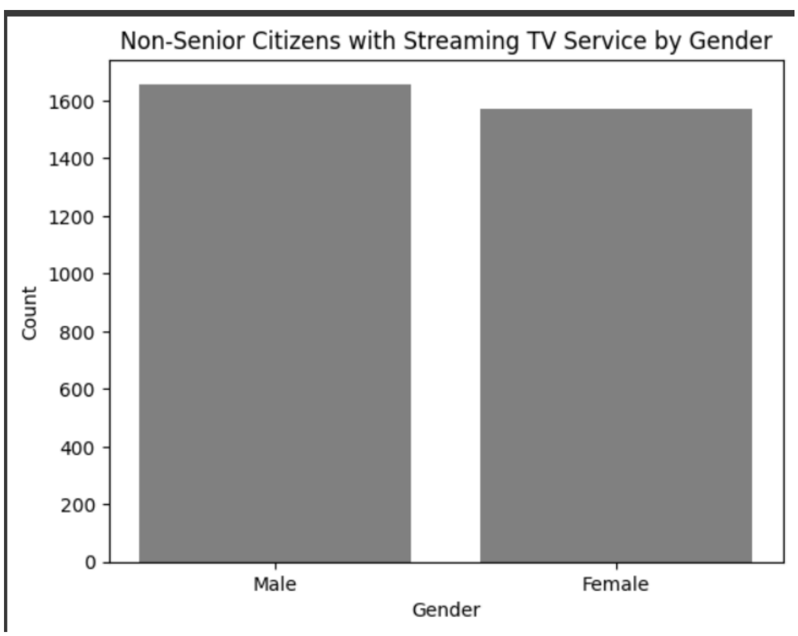

Non Senior citizens with Internet service by gender



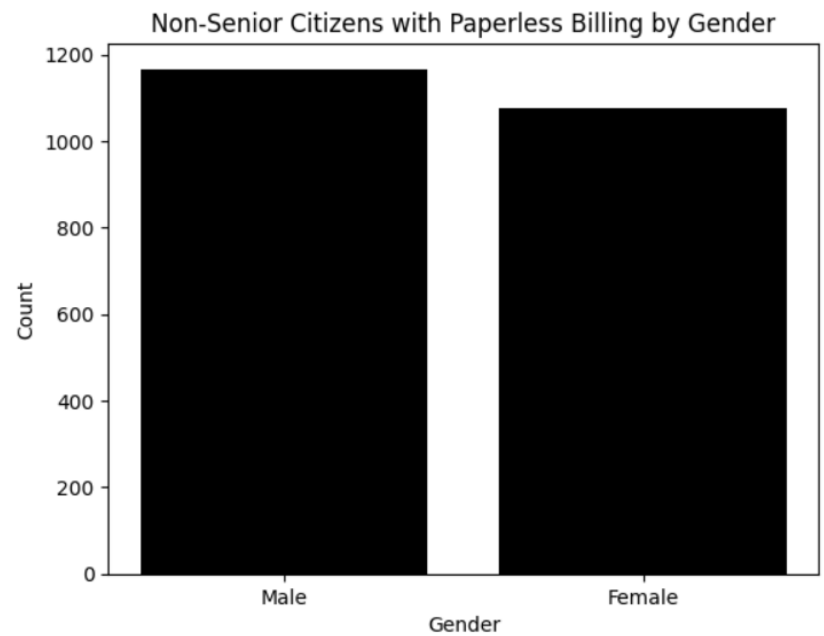
Non Senior citizens with Device protection by gender



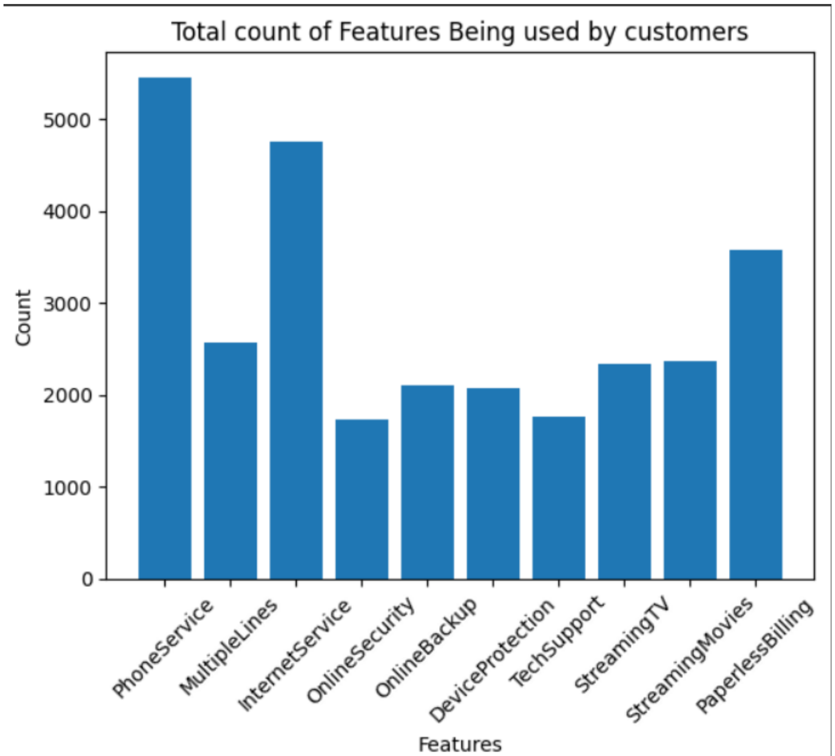

Non Senior citizens with Streaming TV by gender



Non Senior citizens with Paperless billing by gender

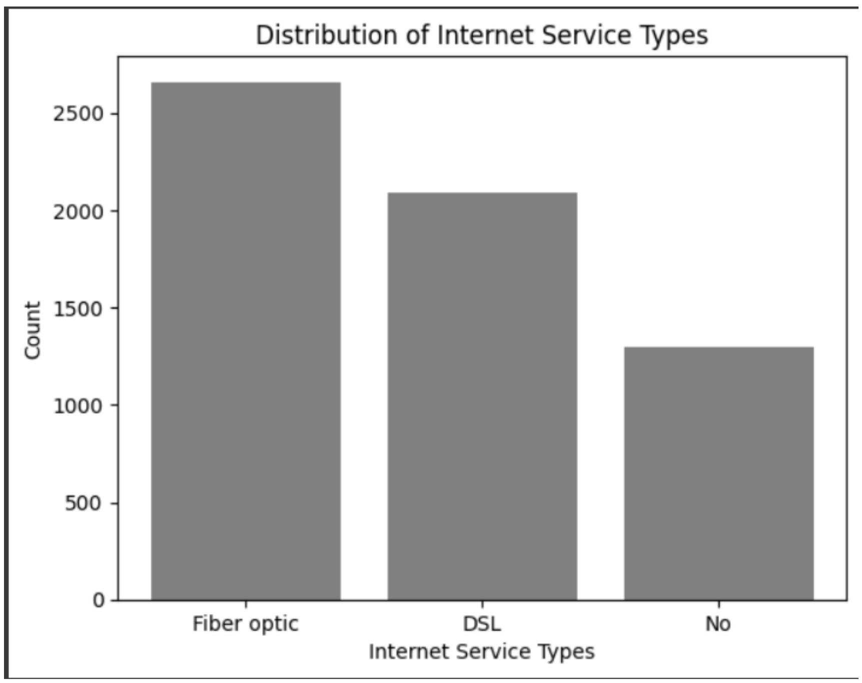



Total count of features being used by customers



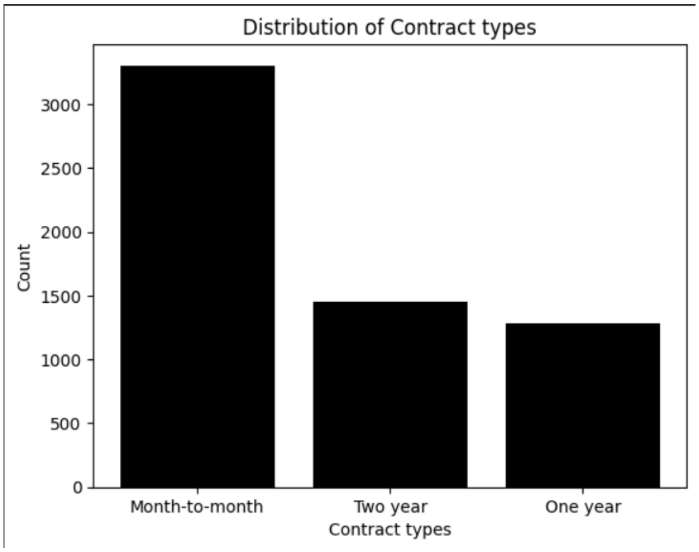
The bar graph reveals the distribution of feature usage among customers. The tallest bar corresponds to 'Phone Service', signifying its highest adoption rate, indicating that the majority of customers utilise this service. Following closely is 'Internet Service', with many customers accessing Internet services. In contrast, 'Online Security' and 'Device Protection' exhibit shorter bars, suggesting comparatively lower adoption. This insight underscores the varied popularity of these features, highlighting phone and internet services' prominence, while online security and device protection exhibit lower utilisation rates.

## Distribution of Internet service types



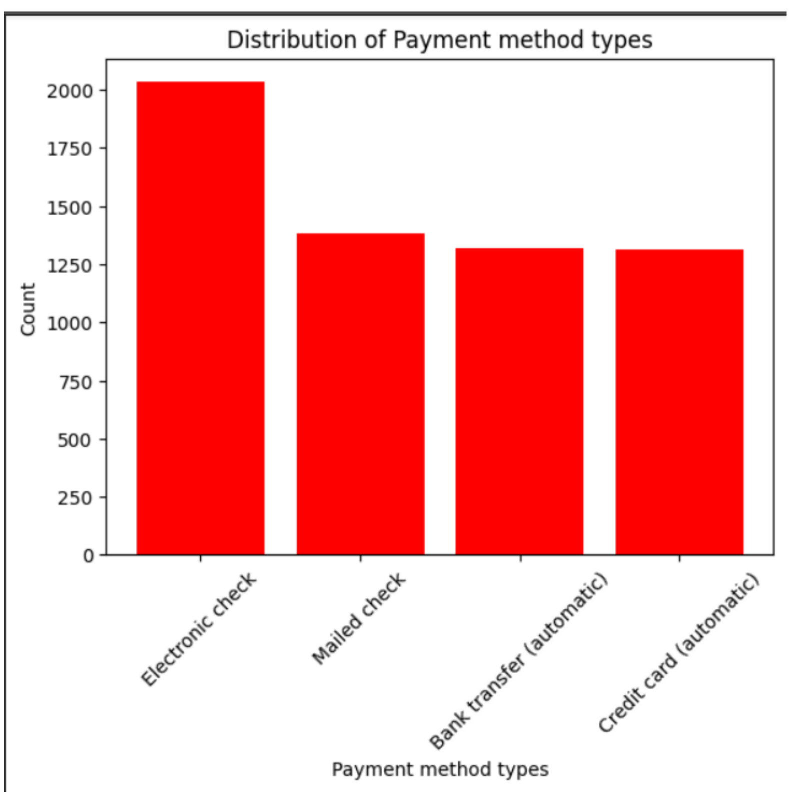

The bar graph illustrates the distribution of internet service types among customers. Among the three categories, Fiber optic exhibits the highest adoption rate, as indicated by the tallest bar. DSL follows closely in popularity. Conversely, the category of customers not using internet service registers the lowest count, represented by the shortest bar. This visualisation provides a clear overview of the usage patterns across different internet service types.

## Distribution of Contract types



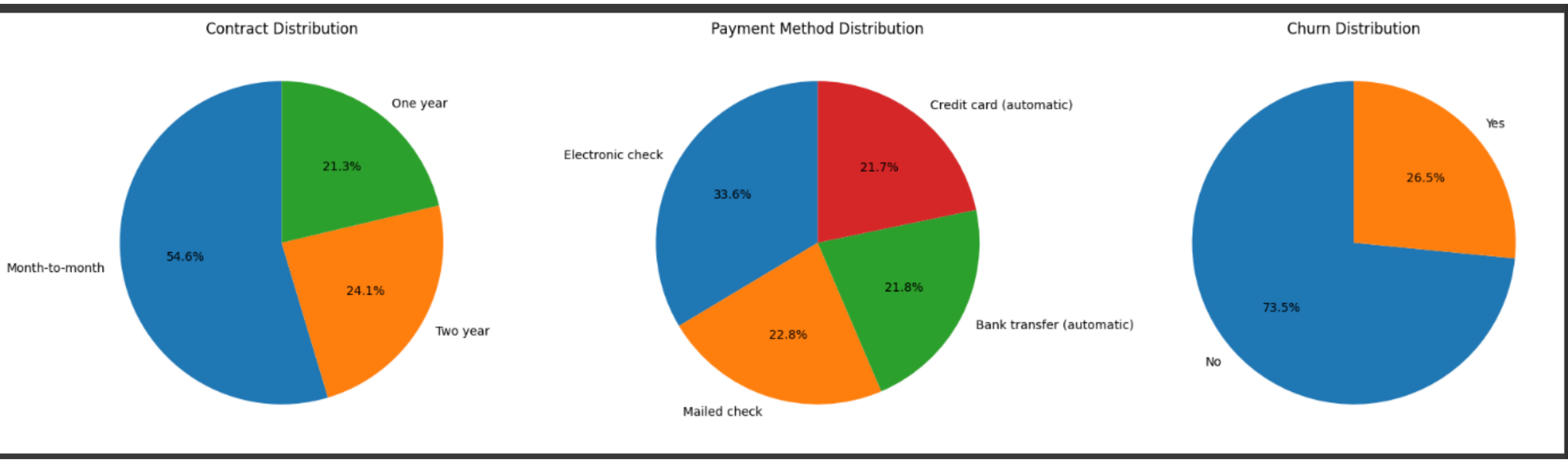

The bar graph depicting the distribution of contract types highlights three distinct categories: month-to-month, one-year, and two-year contracts. The most prevalent contract type is month-to-month, marked by the highest bar, signifying its popularity among customers. Conversely, the one-year contract exhibits lower customer adoption, whereas the two-year contract experiences a slightly higher level of customer participation. This visual representation succinctly conveys the varying preferences for contract durations, with month-to-month being favoured while the longer-term options garner relatively less.

## Distribution of Payment method types



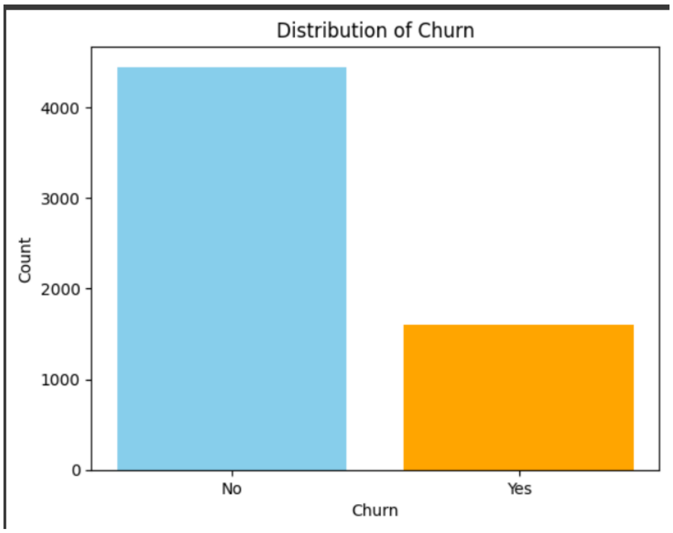
The bar graph outlines distinct payment methods, categorised into four types. Notably, the "Electronic check" method is the most favoured among customers, evidenced by the tallest bar. Conversely, the three categories exhibit comparable usage rates, albeit noticeably lower than the "Electronic check." This visualisation underscores the prevalence of the "Electronic check" payment method.

## Distribution of contract types and payment methods and tenure among those who either use phone or internet services



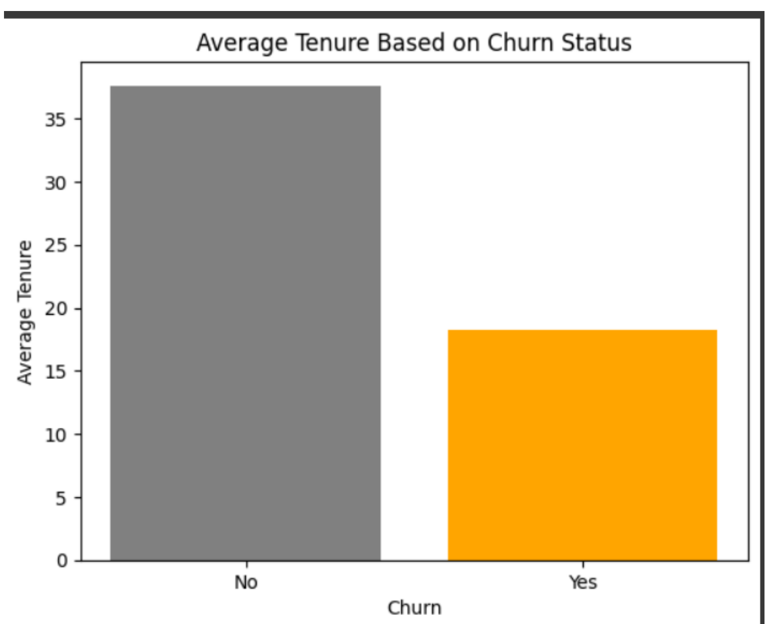

Due to the significant prevalence of Internet services and the high phone service usage among customers, our analysis is centred around these categories. Through the provided graphs, we observed that the most favoured contract type is month-to-month, indicating a preference for flexibility in commitment. Notably, the most prominent payment method is the "Electronic Check," highlighting its popularity among customers. Moreover, a substantial majority, accounting for 73.5% of cases, demonstrates a status of "No" for churn, implying strong retention of services.

## Distribution of Churn



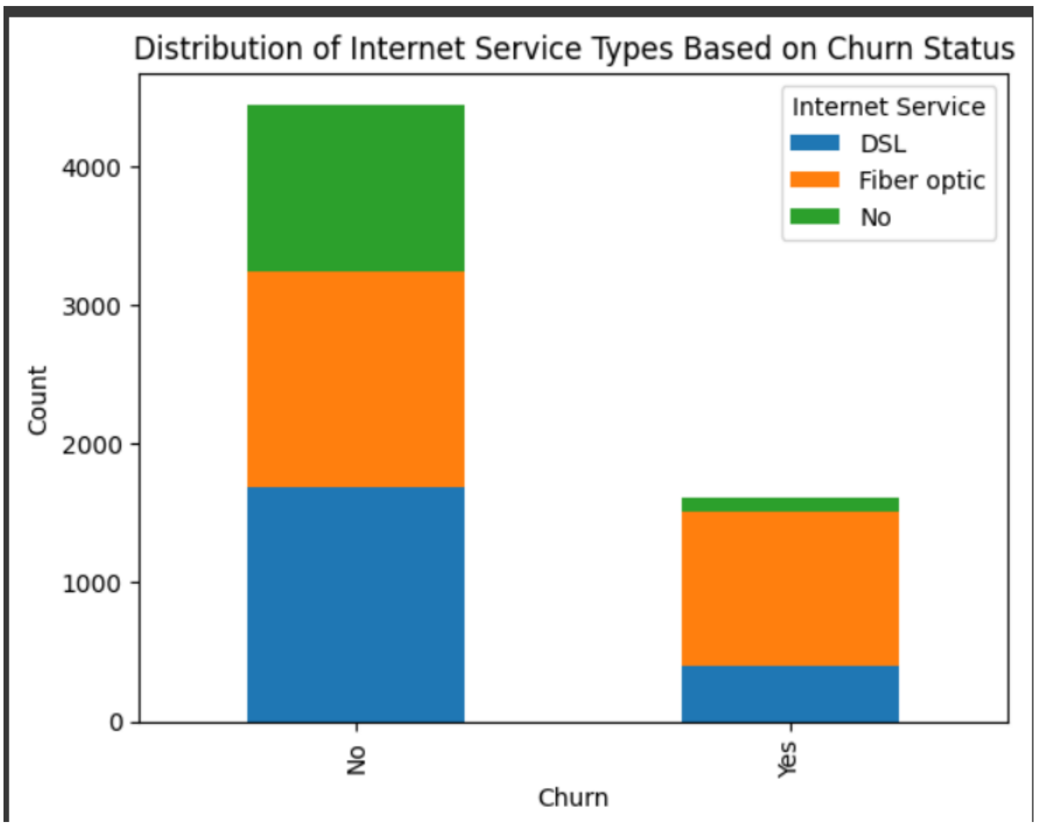

The bar graph vividly portrays the distribution of churn status among customers, categorised into two distinct outcomes: "Yes" and "No." This visual contrast underscores the prevailing trend of customers choosing to retain their services.

## Average Tenure based on Churn status



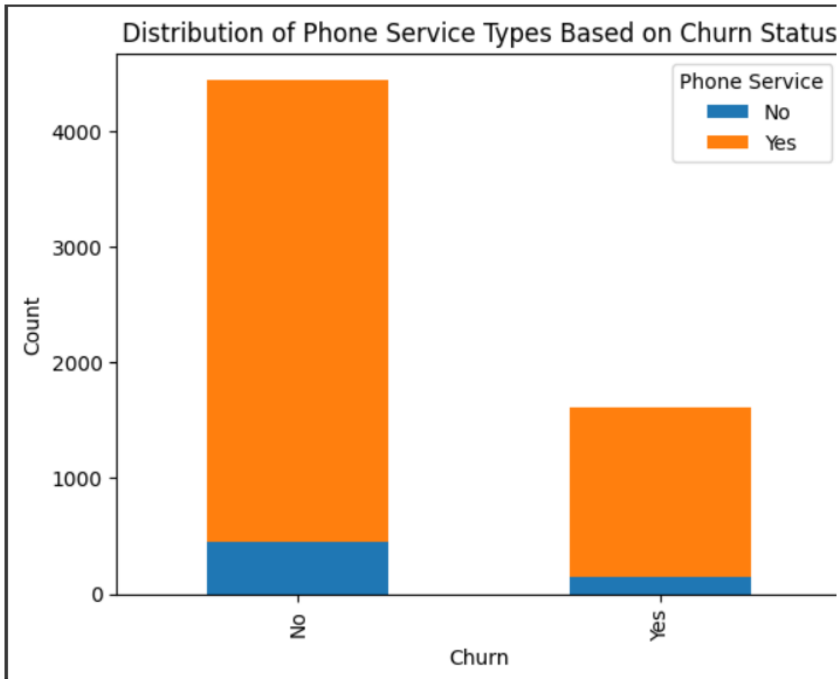
The bar graph provides a clear insight into the relationship between churn status and the average tenure of customers. Notably, customers with "No" churn status exhibit a significantly higher average tenure, surpassing 35 months. In stark contrast, customers with "Yes" churn status display a notably lower average tenure, ranging between 15 to 20 months. This visual representation effectively distinguishes the varying tenure dynamics between the two churn categories, highlighting the substantial difference in the average duration of customer relationships with the service.

## Distribution of Internet service types based on Churn status



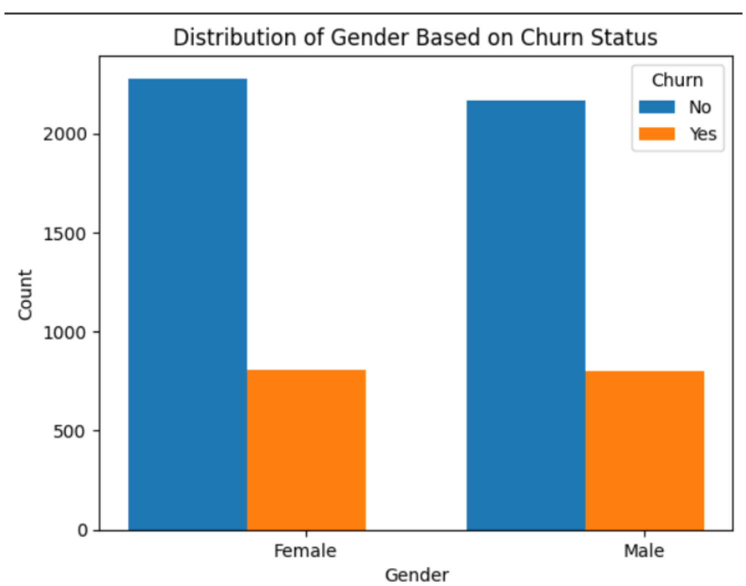

The stacked bar graph comprehensively shows internet service types distributed across different churn statuses. Within the "No" churn category, the subcategories exhibit a roughly balanced distribution, with those without internet service registering slightly lower. In contrast, the "Yes" churn category showcases a distinct pattern, prominently featuring "Fiber optic" as the dominant internet service type. This visualisation effectively contrasts the prevalence of internet service types within each churn status, illuminating the significant role of "Fiber optic" in the "Yes" churn cases.

## Distribution of Phone service types based on Churn status



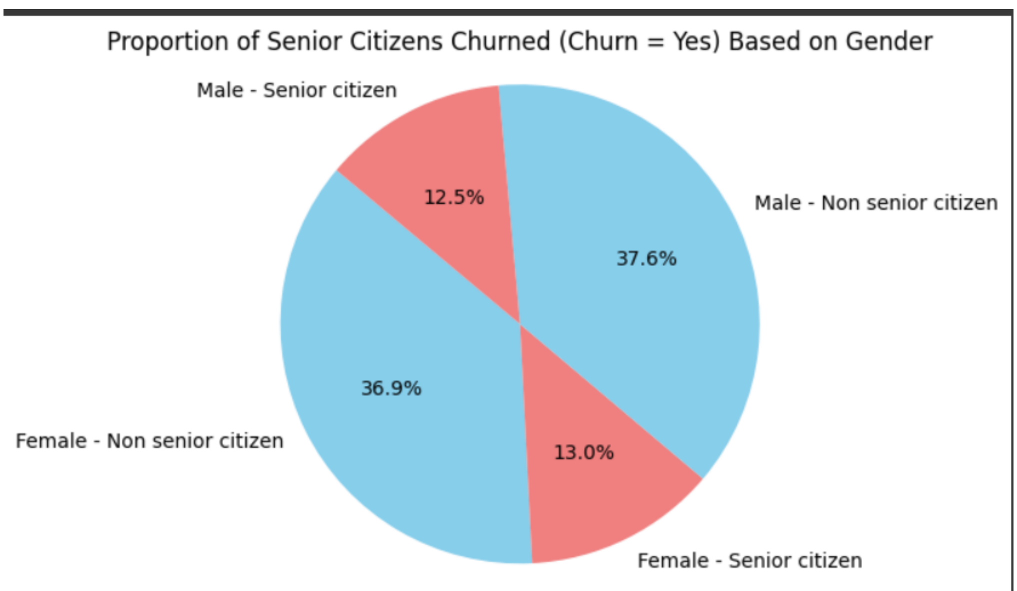

The stacked bar graph provides an insightful portrayal of phone service types distributed across churn statuses. Notably, both the "Yes" and "No" churn categories predominantly feature "Yes" as the prevailing phone service choice. This visual depiction underscores a shared inclination towards utilising phone service, irrespective of churn status.

## Distribution of gender Gender based on Churn status



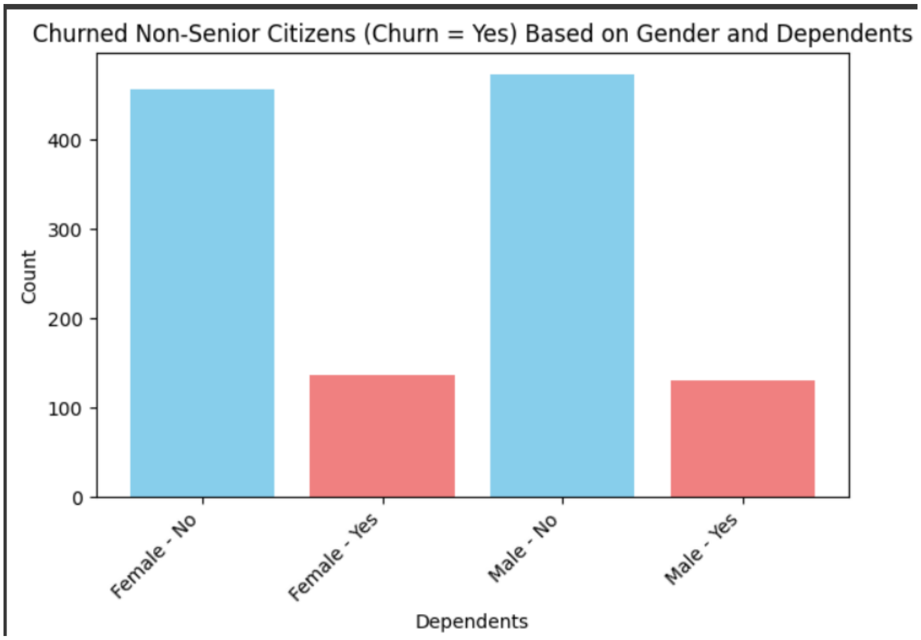
The grouped bar chart visually depicts the gender distribution across different churn statuses. Remarkably, the "No" churn category registers a higher count for both genders, indicating a prevailing trend of retaining services. Interestingly, while the "Yes" and "No" churn categories exhibit a slight predominance among males compared to females, this distinction remains inconsequential. This observation aligns with the overarching trend established in previous summaries, where the cumulative number of males significantly outweighs females.

## The proportion of Senior and non-senior citizens (churn = yes) based on gender



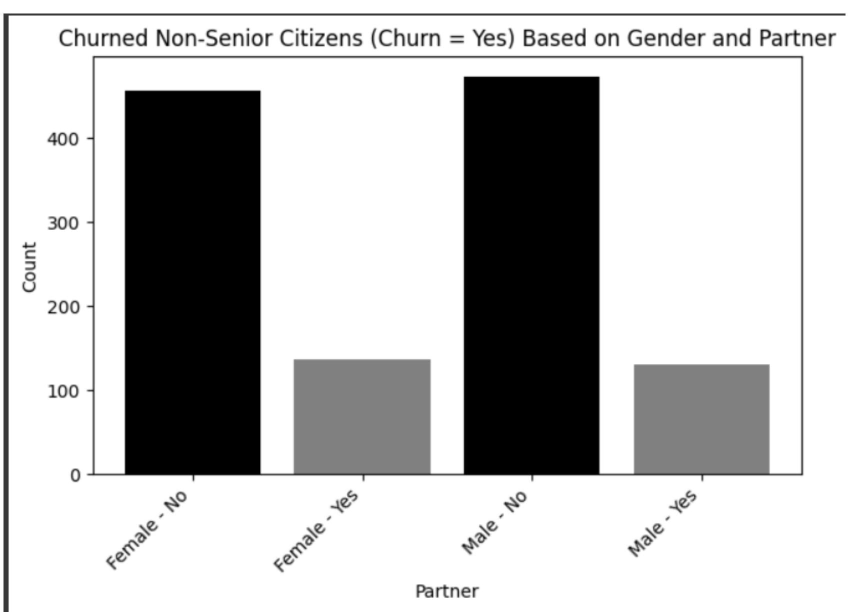

The pie chart illustrates the proportion of senior citizens who have churned (Churn = Yes) based on gender. The labels on the chart represent both gender and senior citizen status. Each label combines gender and senior citizen information to provide a comprehensive overview. Upon analysis, it is evident that the proportion of non-senior citizens who have churned is notably higher in both genders. This finding suggests that, regardless of gender, a more significant portion of non-senior citizens have churned compared to senior citizens.

## Non_Senior (Churn = Yes) based on Gender and Dependents



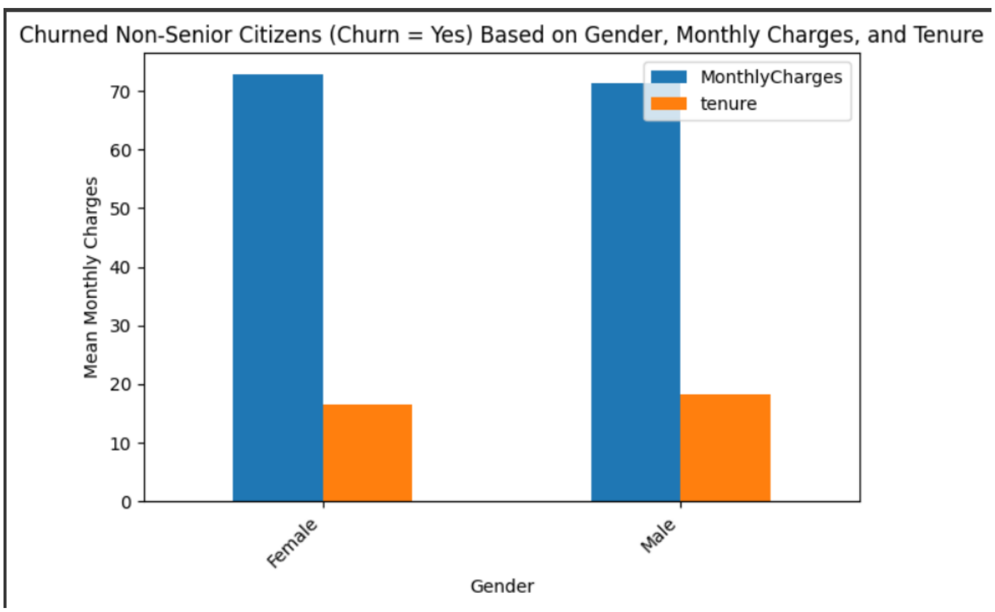

In the context of churned non-senior citizens (Churn = Yes), the bar graph provides insight into the distribution of dependents based on gender. Notably, among both genders, the 'Dependents = No' category has a higher count than 'Dependents = Yes.' Additionally, within the 'Dependents = No' category, the count is slightly higher for the female gender. Conversely, the 'Dependents = Yes' category shows a slightly higher count for the male gender. This visual representation underscores the gender-wise distribution of dependents among churned non-senior citizens, shedding light on potential trends in customer churn behaviour.

## Non_Senior (Churn = Yes) based on Gender and Partner



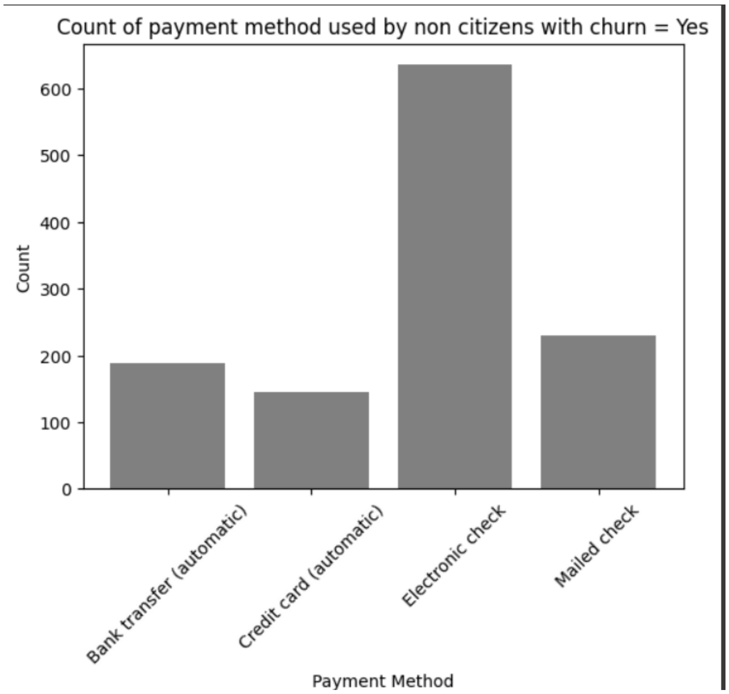
Examining churned non-senior citizens (Churn = Yes), the presented bar graph offers insights into the distribution of partnership status based on gender. Evidently, 'No Partner' holds a higher count than 'Yes Partner' for both genders. Further scrutiny reveals a nuanced trend: within the 'No Partner' category, the count is marginally higher for the male gender, indicating a slightly higher proportion of churned males with no partner. Conversely, the 'Yes Partner' category showcases a slight dominance in the female gender, suggesting a marginally higher proportion of churned females with a partner.

## Average of Monthly charges and Tenure of Non_Senior (Churn = Yes) based on Gender



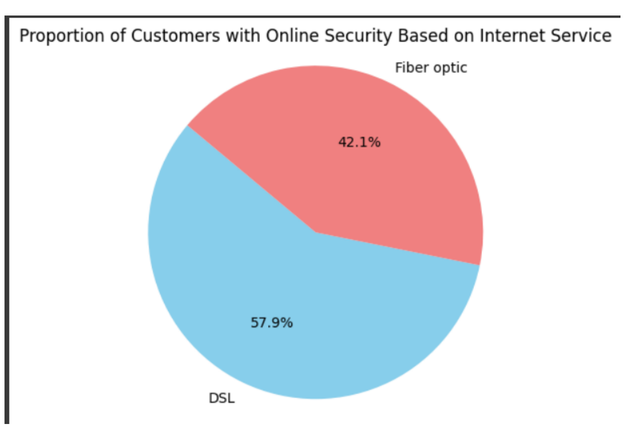

The bar graph compares mean monthly charges and tenure for churned non-senior citizens (Churn = Yes) based on gender. Upon analysis, it becomes evident that the average monthly charges for females are slightly higher than for males. Additionally, the average tenure for males is slightly greater than that of females. However, it is essential to note that these differences between genders in the analysed categories are relatively minor.

## Count of different payment methods used by non-Citizens (Churn = Yes)



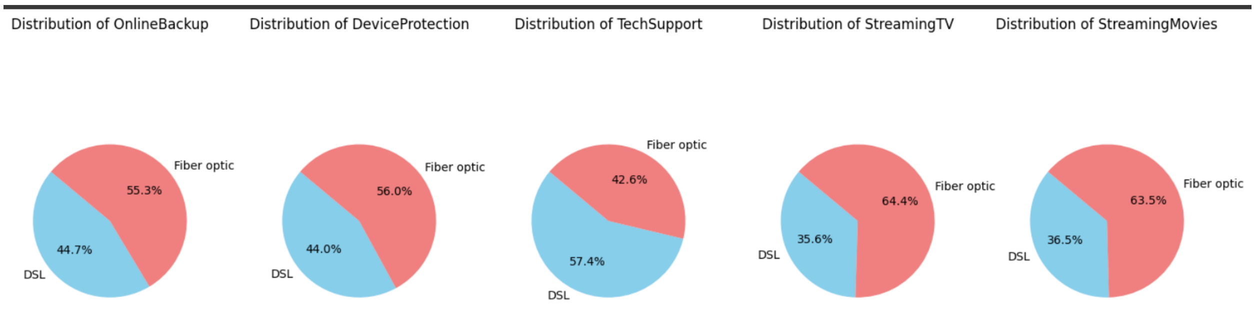

The bar graph illustrates the count of payment methods used by non-senior citizens with a churn status of "Yes." The graph provides valuable insights into the distribution of payment methods among this specific group of customers. Notably, the "Electronic Check" payment method is the most widely adopted choice, representing the highest usage rate among the options. Conversely, the "Credit Card (automatic)" payment method displays the lowest usage rate, indicating a lower prevalence than the other methods. It is worth mentioning that rate aligns with our earlier analysis of the overall distribution of payment methods across all customers. This consistency underscores the observed preference patterns even within the subset of churned non-senior citizens.

## The proportion of customers with online security based on Internet service



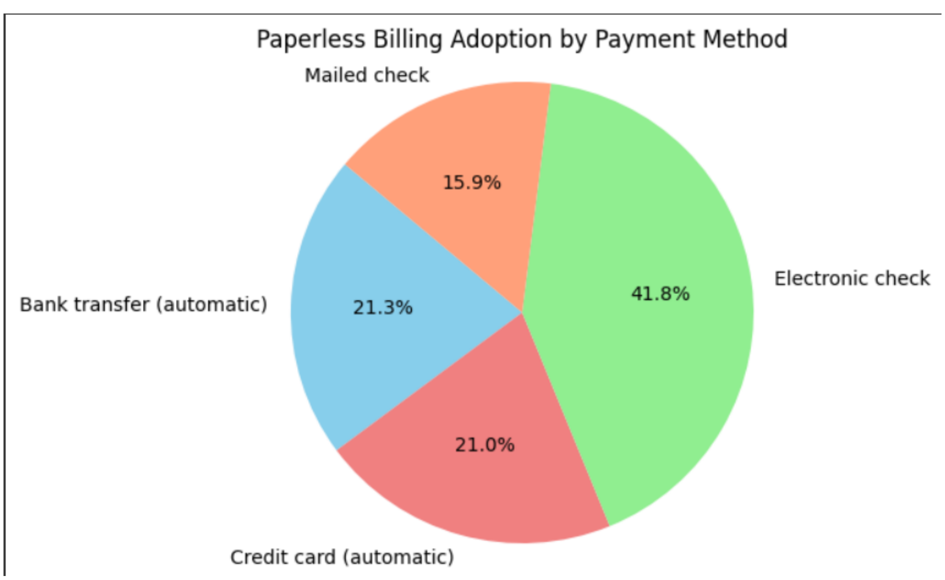
The pie chart visually represents the distribution of customers with Online Security based on different categories of Internet Service. Notably, the category corresponding to DSL Internet Service stands out with the highest proportion of customers who have Online Security enabled. This suggests that DSL users exhibit a relatively higher preference for Online Security among the available Internet Service options.

## Service Adoption Patterns and Preferences in Different Internet Service Types



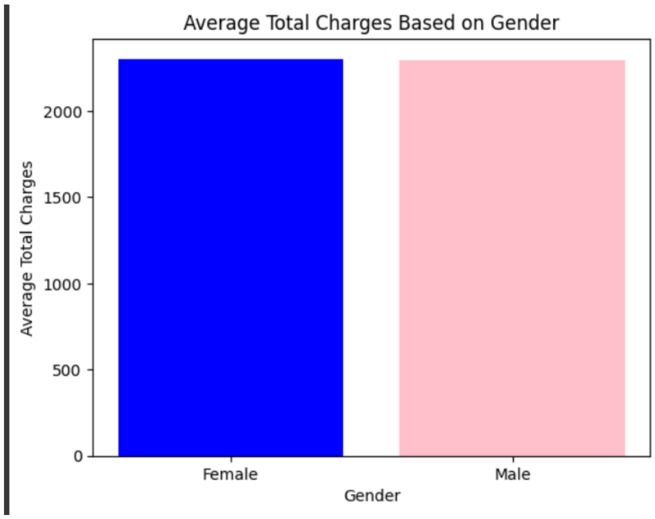

Based on the pie charts we created earlier for different features in relation to Internet Service, it's evident that DSL service users tend to utilise more 'TechSupport' and 'OnlineSecurity'. In contrast, Fiber optic service users have a higher preference for other services. This contrast could indicate varying user preferences and needs depending on the type of Internet Service they are subscribed to.

## Paperless Billing adoption by payment method



The presented pie chart offers an insightful depiction of paperless billing adoption across various payment methods. Notably, the "Electronic Check" method boasts the highest rate of paperless billing enrollment, occupying a substantial portion of the chart. Conversely, the "Credit Card (automatic)" method registers the lowest adoption rate for paperless billing, constituting a smaller portion of the overall distribution.

## Average total charges based on gender




The presented bar graph illustrates the distribution of average total charges based on gender. Notably, the graph showcases that the average total charges for females are slightly higher compared to males. This observation is intriguing given that the number of male registrants is relatively higher. Despite the higher male registrants, females exhibit slightly higher average total charges, suggesting potential variations in service utilization.