Project Report

Bachelor Of Computer Applications
2nd Semester

Exploratory Data Analysis Project

Insurance Dataset Based on Real-World Statistics

By

SATHVIK B R

2411021240027

Githublink: https://github.com/Sathvik0007/SDS-PROJECT-.git

Department Of Computer Application
Alliance University

Chandrapura - Anekal Main Road, Anekal
Bengaluru – 56210

Introduction

The dataset used in this analysis is a sales dataset that contains various attributes related to product sales, including 'Units Sold', 'Unit Price', and other relevant features. The primary objective of this analysis is to explore and visualize the distribution of key variables, specifically 'Units Sold' and 'Unit Price'. By employing statistical visualizations such as histograms and boxplots, we aim to gain insights into the sales performance, identify trends, and detect any anomalies or outliers in the data. This understanding can inform business decisions, optimize pricing strategies, and enhance inventory management.

Objectives

1. Analyze Sales Distribution: To visualize and understand the distribution of 'Units Sold' and 'Unit Price' to identify patterns, trends, and central tendencies in the sales data.

2. Identify Outliers: To detect any outliers in the 'Units Sold' and 'Unit Price' data using boxplots, which can indicate unusual sales behavior or pricing strategies that may need further investigation.

3. Understand Variability: To assess the variability and spread of the data, helping to understand the range of sales performance and pricing strategies across different products.

4. Support Data-Driven Decisions: To provide insights that can inform business decisions
related to inventory management, pricing strategies, and sales forecasting.

5. Enhance Reporting: To create visual representations of the data that can be used in reports and presentations, making it easier for stakeholders to grasp key insights quickly.

Libraries Used

1. **Pandas**: For data manipulation and analysis.
2. **Seaborn**: For creating visualizations like histograms and boxplots.
3. **Matplotlib**: For customizing and displaying plots.

These libraries facilitated effective analysis and visualization of the sales data.

Load the dataset

```
# Dataset of students
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df=pd.read_csv(r"/Users/sathvikbr/Documents/synthetic_insurance_data.csv")
df
```

```
      Age  Is_Senior Marital_Status  Married_Premium_Discount Prior_Insurance  \
0      47          0        Married                        86       1-5 years
1      37          0        Married                        86       1-5 years
2      49          0        Married                        86       1-5 years
3      62          1        Married                        86        >5 years
4      36          0         Single                         0        >5 years
...   ...        ...            ...                       ...             ...
9995   59          1         Single                         0       1-5 years
9996   18          0        Married                        86       1-5 years
9997   29          0        Married                        86         <1 year
9998   47          0         Single                         0         <1 year
9999   49          0       Divorced                         0       1-5 years

      Prior_Insurance_Premium_Adjustment  Claims_Frequency Claims_Severity  \
0                                     50                 0             Low
1                                     50                 0             Low
2                                     50                 1             Low
3                                      0                 1             Low
4                                      0                 2             Low
...                                  ...               ...             ...
9995                                  50                 0             Low
9996                                  50                 0          Medium
9997                                 100                 0             Low
9998                                 100                 0          Medium
9999                                  50                 0            High

      Claims_Adjustment     Policy_Type  ...  Time_Since_First_Contact  \
0                     0   Full Coverage  ...                        10
1                     0   Full Coverage  ...                        22
2                    50   Full Coverage  ...                        28
3                    50   Full Coverage  ...                         4
4                   100   Full Coverage  ...                        14
...                 ...             ...  ...                       ...
9995                  0   Full Coverage  ...                         6
9996                  0   Full Coverage  ...                         3
9997                  0   Full Coverage  ...                        29
9998                  0  Liability-Only  ...                         8
9999                  0  Liability-Only  ...                        11

      Conversion_Status  Website_Visits  Inquiries  Quotes_Requested  \
```

```
0                      0                5          1                  2
1                      0                5          1                  2
2                      0                4          4                  1
3                      1                6          2                  2
4                      1                8          4                  2
...                  ...              ...        ...                ...
9995                   1                4          3                  2
9996                   1                6          1                  3
9997                   1                3          4                  3
9998                   1                2          4                  1
9999                   1                5          0                  2

      Time_to_Conversion Credit_Score  Premium_Adjustment_Credit    Region  \
0                     99          704                        -50  Suburban
1                     99          726                        -50     Urban
2                     99          772                        -50     Urban
3                      2          809                        -50     Urban
4                     10          662                         50  Suburban
...                  ...          ...                        ...       ...
9995                   9          783                        -50     Urban
9996                   6          667                         50     Urban
9997                   3          637                         50     Urban
9998                  13          676                         50  Suburban
9999                   4          776                        -50  Suburban

      Premium_Adjustment_Region
0                            50
1                           100
2                           100
3                           100
4                            50
...                         ...
9995                        100
9996                        100
9997                        100
9998                         50
9999                         50

[10000 rows x 27 columns]
```

```
df.head (20)
```

```
    Age  Is_Senior Marital_Status  Married_Premium_Discount Prior_Insurance
\
0    47          0        Married                        86       1-5 years
1    37          0        Married                        86       1-5 years
2    49          0        Married                        86       1-5 years
3    62          1        Married                        86        >5 years
4    36          0         Single                         0        >5 years
5    36          0        Married                        86        >5 years
```

```
6    63          1       Married                          86       1-5 years
7    51          0        Single                           0         <1 year
8    32          0       Married                          86        >5 years
9    48          0        Single                           0        >5 years
10   33          0        Single                           0        >5 years
11   33          0       Married                          86         <1 year
12   43          0        Single                           0       1-5 years
13   18          0        Single                           0       1-5 years
14   18          0       Married                          86       1-5 years
15   31          0       Married                          86       1-5 years
16   24          0        Single                           0       1-5 years
17   44          0       Widowed                           0       1-5 years
18   26          0        Single                           0         <1 year
19   18          0       Married                          86       1-5 years

    Prior_Insurance_Premium_Adjustment  Claims_Frequency Claims_Severity  \
0                                   50                 0             Low
1                                   50                 0             Low
2                                   50                 1             Low
3                                    0                 1             Low
4                                    0                 2             Low
5                                    0                 0          Medium
6                                   50                 0             Low
7                                  100                 0             Low
8                                    0                 0             Low
9                                    0                 1            High
10                                   0                 1             Low
11                                 100                 1             Low
12                                  50                 1            High
13                                  50                 1             Low
14                                  50                 1             Low
15                                  50                 0             Low
16                                  50                 0             Low
17                                  50                 1             Low
18                                 100                 0             Low
19                                  50                 0             Low

    Claims_Adjustment     Policy_Type  ...  Time_Since_First_Contact  \
0                   0   Full Coverage  ...                        10
1                   0   Full Coverage  ...                        22
2                  50   Full Coverage  ...                        28
3                  50   Full Coverage  ...                         4
4                 100   Full Coverage  ...                        14
5                   0  Liability-Only  ...                        13
6                   0   Full Coverage  ...                         2
7                   0   Full Coverage  ...                         1
8                   0  Liability-Only  ...                        16
9                 200   Full Coverage  ...                        27
10                 50  Liability-Only  ...                        22
11                 50  Liability-Only  ...                        16
```

```
12               200  Liability-Only  ...                        29
13                50  Liability-Only  ...                        22
14                50   Full Coverage  ...                         7
15                 0   Full Coverage  ...                        23
16                 0   Full Coverage  ...                         8
17                50   Full Coverage  ...                         4
18                 0   Full Coverage  ...                        14
19                 0  Liability-Only  ...                        11

    Conversion_Status  Website_Visits  Inquiries  Quotes_Requested  \
0                   0               5          1                 2
1                   0               5          1                 2
2                   0               4          4                 1
3                   1               6          2                 2
4                   1               8          4                 2
5                   1               4          1                 1
6                   1               5          1                 2
7                   0               3          0                 2
8                   1               5          1                 3
9                   0               5          3                 2
10                  1               3          2                 1
11                  1               3          1                 3
12                  0               4          1                 3
13                  1               7          0                 3
14                  1               3          3                 2
15                  0               3          3                 2
16                  1               3          2                 3
17                  1               3          1                 3
18                  1               5          2                 1
19                  1               3          2                 1

    Time_to_Conversion Credit_Score  Premium_Adjustment_Credit    Region
0                   99          704                        -50  Suburban
1                   99          726                        -50     Urban
2                   99          772                        -50     Urban
3                    2          809                        -50     Urban
4                   10          662                         50  Suburban
5                    7          729                        -50     Rural
6                    1          795                        -50     Urban
7                   99          639                         50  Suburban
8                    3          724                        -50     Rural
9                   99          710                        -50     Urban
10                  13          688                         50  Suburban
11                   6          659                         50     Rural
12                  99          745                        -50     Urban
13                   9          777                        -50  Suburban
14                   8          668                         50  Suburban
15                  99          775                        -50     Urban
16                  10          695                         50  Suburban
17                   7          739                        -50     Urban
```

```
18                    8          750                           -50     Urban
19                   11          625                            50     Urban

    Premium_Adjustment_Region
0                          50
1                         100
2                         100
3                         100
4                          50
5                           0
6                         100
7                          50
8                           0
9                         100
10                         50
11                          0
12                        100
13                         50
14                         50
15                        100
16                         50
17                        100
18                        100
19                        100

[20 rows x 27 columns]
```

```
df.tail (20)
```

```
      Age  Is_Senior Marital_Status  Married_Premium_Discount Prior_Insurance  \
9980   33          0         Single                         0        >5 years
9981   32          0        Married                        86       1-5 years
9982   47          0        Married                        86       1-5 years
9983   43          0        Married                        86         <1 year
9984   22          0         Single                         0       1-5 years
9985   34          0       Divorced                         0       1-5 years
9986   34          0         Single                         0       1-5 years
9987   18          0        Married                        86       1-5 years
9988   25          0        Married                        86         <1 year
9989   28          0        Married                        86        >5 years
9990   61          1        Married                        86         <1 year
9991   42          0        Married                        86        >5 years
9992   49          0        Widowed                         0       1-5 years
9993   18          0         Single                         0       1-5 years
9994   57          1         Single                         0         <1 year
9995   59          1         Single                         0       1-5 years
9996   18          0        Married                        86       1-5 years
9997   29          0        Married                        86         <1 year
9998   47          0         Single                         0         <1 year
```

```
9999   49          0       Divorced                         0       1-5 years

      Prior_Insurance_Premium_Adjustment  Claims_Frequency Claims_Severity  \
9980                                   0                 0             Low
9981                                  50                 2             Low
9982                                  50                 0             Low
9983                                 100                 2            High
9984                                  50                 0             Low
9985                                  50                 0             Low
9986                                  50                 0             Low
9987                                  50                 0            High
9988                                 100                 0             Low
9989                                   0                 1             Low
9990                                 100                 1          Medium
9991                                   0                 0             Low
9992                                  50                 1             Low
9993                                  50                 0             Low
9994                                 100                 0             Low
9995                                  50                 0             Low
9996                                  50                 0          Medium
9997                                 100                 0             Low
9998                                 100                 0          Medium
9999                                  50                 0            High

      Claims_Adjustment     Policy_Type  ...  Time_Since_First_Contact  \
9980                  0   Full Coverage  ...                         4
9981                100  Liability-Only  ...                        26
9982                  0   Full Coverage  ...                        25
9983                400   Full Coverage  ...                         3
9984                  0  Liability-Only  ...                        27
9985                  0   Full Coverage  ...                         8
9986                  0   Full Coverage  ...                        26
9987                  0  Liability-Only  ...                         7
9988                  0  Liability-Only  ...                        24
9989                 50   Full Coverage  ...                        25
9990                100   Full Coverage  ...                        20
9991                  0   Full Coverage  ...                        16
9992                 50  Liability-Only  ...                        10
9993                  0  Liability-Only  ...                         5
9994                  0   Full Coverage  ...                        23
9995                  0   Full Coverage  ...                         6
9996                  0   Full Coverage  ...                         3
9997                  0   Full Coverage  ...                        29
9998                  0  Liability-Only  ...                         8
9999                  0  Liability-Only  ...                        11

      Conversion_Status  Website_Visits  Inquiries  Quotes_Requested  \
9980                  1               4          0                 1
9981                  1               5          5                 1
9982                  1               4          2                 1
```

9983	1	7	3	1
9984	1	5	0	2
9985	1	4	2	1
9986	1	4	2	3
9987	1	3	0	2
9988	1	7	3	2
9989	0	3	2	1
9990	1	4	3	3
9991	0	4	0	2
9992	1	4	0	3
9993	1	6	1	3
9994	1	3	3	1
9995	1	4	3	2
9996	1	6	1	3
9997	1	3	4	3
9998	1	2	4	1
9999	1	5	0	2

	Time_to_Conversion	Credit_Score	Premium_Adjustment_Credit	Region \
9980	5	817	-50	Suburban
9981	7	663	50	Urban
9982	5	729	-50	Rural
9983	2	714	-50	Urban
9984	9	626	50	Urban
9985	13	738	-50	Urban
9986	1	672	50	Urban
9987	2	710	-50	Rural
9988	12	691	50	Rural
9989	99	687	50	Urban
9990	6	646	50	Suburban
9991	99	662	50	Suburban
9992	10	828	-50	Suburban
9993	7	723	-50	Suburban
9994	5	709	-50	Urban
9995	9	783	-50	Urban
9996	6	667	50	Urban
9997	3	637	50	Urban
9998	13	676	50	Suburban
9999	4	776	-50	Suburban

	Premium_Adjustment_Region
9980	50
9981	100
9982	0
9983	100
9984	100
9985	100
9986	100
9987	0
9988	0

```
9989                        100
9990                         50
9991                         50
9992                         50
9993                         50
9994                        100
9995                        100
9996                        100
9997                        100
9998                         50
9999                         50

[20 rows x 27 columns]
```

```
df.describe()
```

```
                Age     Is_Senior  Married_Premium_Discount  \
count  10000.000000  10000.000000              10000.000000   
mean      39.991700      0.159300                 42.131400   
std       14.050358      0.365974                 42.993376   
min       18.000000      0.000000                  0.000000   
25%       29.000000      0.000000                  0.000000   
50%       39.000000      0.000000                  0.000000   
75%       50.000000      0.000000                 86.000000   
max       90.000000      1.000000                 86.000000   

       Prior_Insurance_Premium_Adjustment  Claims_Frequency  \
count                        10000.000000      10000.000000   
mean                            47.625000          0.497200   
std                             34.354438          0.716131   
min                              0.000000          0.000000   
25%                              0.000000          0.000000   
50%                             50.000000          0.000000   
75%                             50.000000          1.000000   
max                            100.000000          5.000000   

       Claims_Adjustment  Policy_Adjustment  Premium_Amount  \
count       10000.000000       10000.000000    10000.000000   
mean           36.780000         -79.860000     2219.571400   
std            65.910288          97.955806      148.521132   
min             0.000000        -200.000000     1800.000000   
25%             0.000000        -200.000000     2100.000000   
50%             0.000000           0.000000     2236.000000   
75%            50.000000           0.000000     2336.000000   
max           800.000000           0.000000     2936.000000   

       Safe_Driver_Discount  Multi_Policy_Discount  ...  Total_Discounts  \
count          10000.000000           10000.000000  ...     10000.000000   
mean               0.199900               0.305100  ...        30.110000   
std                0.399945               0.460473  ...        33.689782   
```

```
min               0.000000               0.000000  ...         0.000000
25%               0.000000               0.000000  ...         0.000000
50%               0.000000               0.000000  ...        50.000000
75%               0.000000               1.000000  ...        50.000000
max               1.000000               1.000000  ...       150.000000

       Time_Since_First_Contact  Conversion_Status  Website_Visits  \
count              10000.000000       10000.000000    10000.000000
mean                  15.478000           0.576700        5.022900
std                    8.677975           0.494107        2.238231
min                    1.000000           0.000000        0.000000
25%                    8.000000           0.000000        3.000000
50%                   16.000000           1.000000        5.000000
75%                   23.000000           1.000000        6.000000
max                   30.000000           1.000000       16.000000

          Inquiries  Quotes_Requested  Time_to_Conversion  Credit_Score  \
count  10000.000000      10000.000000         10000.00000  10000.000000
mean       1.996900          1.996900            46.07320    714.253400
std        1.415588          0.817409            45.44845     49.749487
min        0.000000          1.000000             1.00000    530.000000
25%        1.000000          1.000000             6.00000    681.000000
50%        2.000000          2.000000            12.00000    715.000000
75%        3.000000          3.000000            99.00000    748.000000
max        9.000000          3.000000            99.00000    850.000000

       Premium_Adjustment_Credit  Premium_Adjustment_Region
count               10000.000000               10000.000000
mean                  -11.320000                  64.325000
std                    48.704156                  39.232618
min                   -50.000000                   0.000000
25%                   -50.000000                  50.000000
50%                   -50.000000                  50.000000
75%                    50.000000                 100.000000
max                    50.000000                 100.000000

[8 rows x 21 columns]
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 27 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Age                             10000 non-null  int64
 1   Is_Senior                       10000 non-null  int64
 2   Marital_Status                  10000 non-null  object
 3   Married_Premium_Discount        10000 non-null  int64
 4   Prior_Insurance                 10000 non-null  object
```

```
 5   Prior_Insurance_Premium_Adjustment  10000 non-null  int64
 6   Claims_Frequency                    10000 non-null  int64
 7   Claims_Severity                     10000 non-null  object
 8   Claims_Adjustment                   10000 non-null  int64
 9   Policy_Type                         10000 non-null  object
 10  Policy_Adjustment                   10000 non-null  int64
 11  Premium_Amount                      10000 non-null  int64
 12  Safe_Driver_Discount                10000 non-null  int64
 13  Multi_Policy_Discount               10000 non-null  int64
 14  Bundling_Discount                   10000 non-null  int64
 15  Total_Discounts                     10000 non-null  int64
 16  Source_of_Lead                      10000 non-null  object
 17  Time_Since_First_Contact            10000 non-null  int64
 18  Conversion_Status                   10000 non-null  int64
 19  Website_Visits                      10000 non-null  int64
 20  Inquiries                           10000 non-null  int64
 21  Quotes_Requested                    10000 non-null  int64
 22  Time_to_Conversion                  10000 non-null  int64
 23  Credit_Score                        10000 non-null  int64
 24  Premium_Adjustment_Credit           10000 non-null  int64
 25  Region                              10000 non-null  object
 26  Premium_Adjustment_Region           10000 non-null  int64
dtypes: int64(21), object(6)
memory usage: 2.1+ MB
```

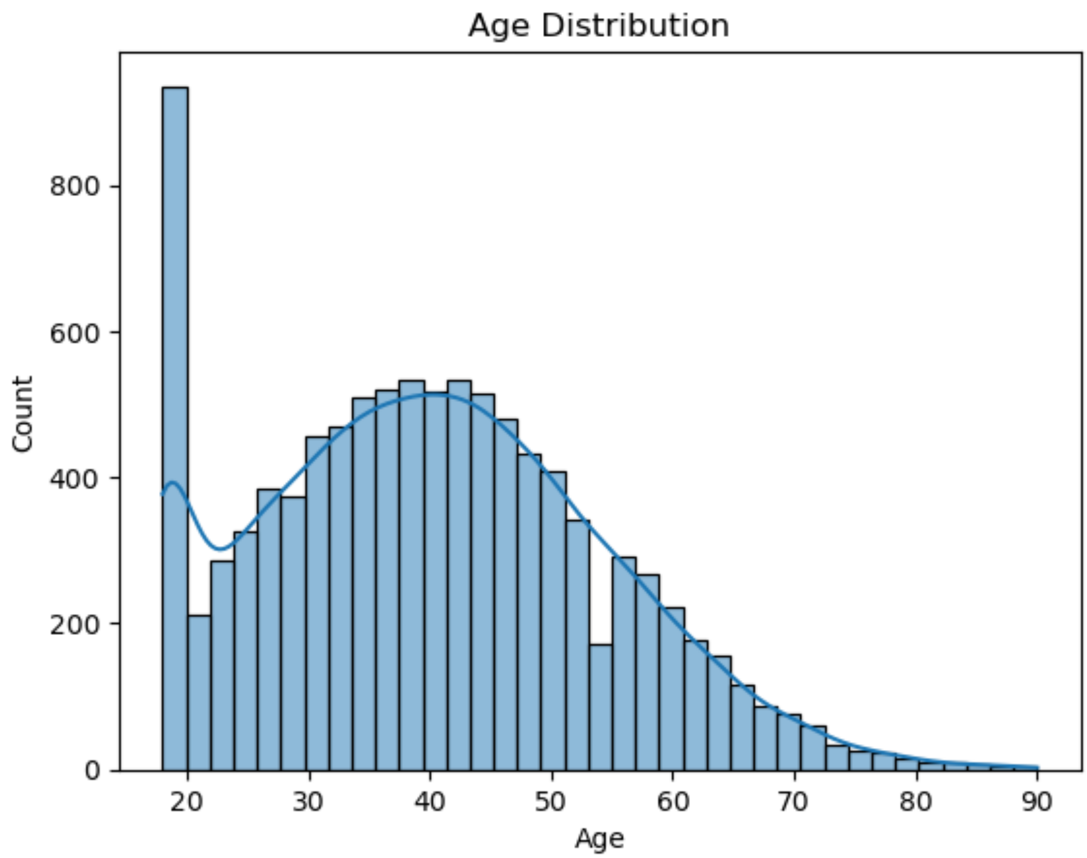
```
df.describe
```

```
<bound method NDFrame.describe of       Age  Is_Senior Marital_Status  
Married_Premium_Discount Prior_Insurance  \
0      47          0        Married                        86       1-5 years
1      37          0        Married                        86       1-5 years
2      49          0        Married                        86       1-5 years
3      62          1        Married                        86        >5 years
4      36          0         Single                         0        >5 years
...   ...        ...            ...                       ...             ...
9995   59          1         Single                         0       1-5 years
9996   18          0        Married                        86       1-5 years
9997   29          0        Married                        86         <1 year
9998   47          0         Single                         0         <1 year
9999   49          0       Divorced                         0       1-5 years

      Prior_Insurance_Premium_Adjustment  Claims_Frequency Claims_Severity  \
0                                     50                 0             Low
1                                     50                 0             Low
2                                     50                 1             Low
3                                      0                 1             Low
4                                      0                 2             Low
...                                  ...               ...             ...
9995                                  50                 0             Low
9996                                  50                 0          Medium
```

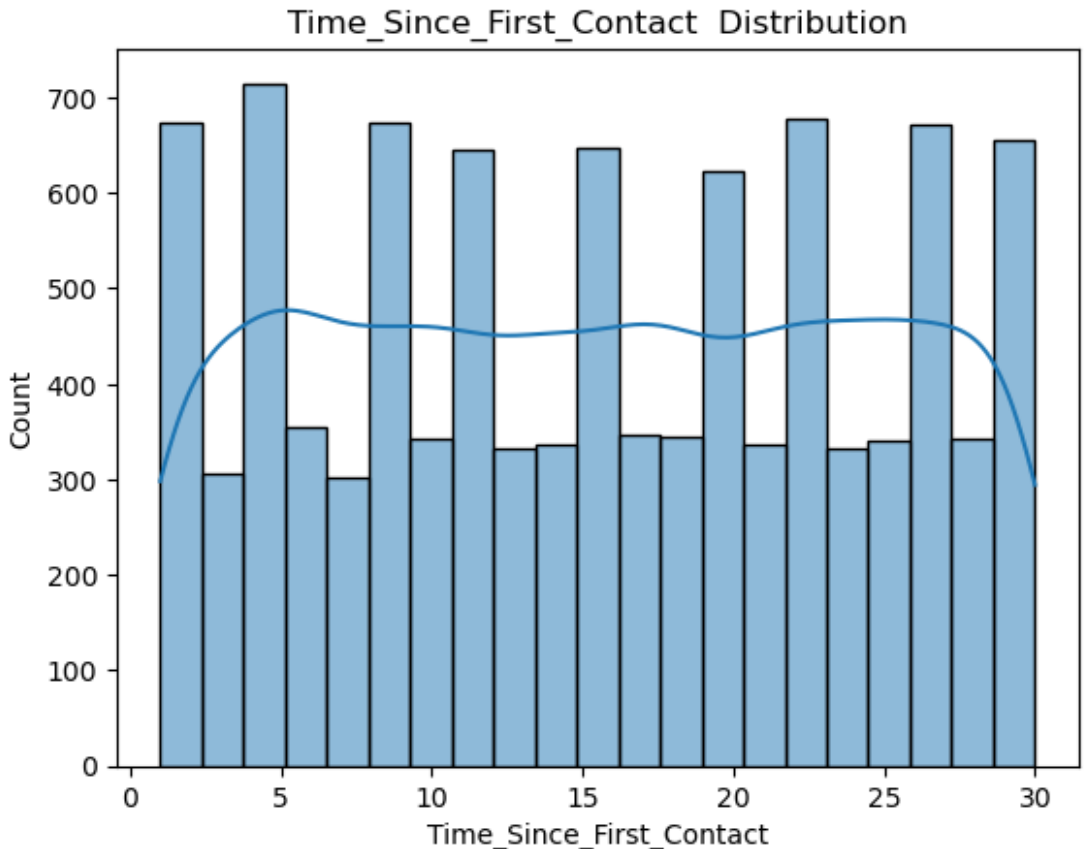
```
9997                               100                0           Low
9998                               100                0        Medium
9999                                50                0          High

      Claims_Adjustment     Policy_Type  ...  Time_Since_First_Contact  \
0                     0   Full Coverage  ...                        10
1                     0   Full Coverage  ...                        22
2                    50   Full Coverage  ...                        28
3                    50   Full Coverage  ...                         4
4                   100   Full Coverage  ...                        14
...                 ...             ...  ...                       ...
9995                  0   Full Coverage  ...                         6
9996                  0   Full Coverage  ...                         3
9997                  0   Full Coverage  ...                        29
9998                  0  Liability-Only  ...                         8
9999                  0  Liability-Only  ...                        11

      Conversion_Status  Website_Visits  Inquiries  Quotes_Requested  \
0                     0               5          1                 2
1                     0               5          1                 2
2                     0               4          4                 1
3                     1               6          2                 2
4                     1               8          4                 2
...                 ...             ...        ...               ...
9995                  1               4          3                 2
9996                  1               6          1                 3
9997                  1               3          4                 3
9998                  1               2          4                 1
9999                  1               5          0                 2

      Time_to_Conversion Credit_Score  Premium_Adjustment_Credit    Region  \
0                     99          704                        -50  Suburban
1                     99          726                        -50     Urban
2                     99          772                        -50     Urban
3                      2          809                        -50     Urban
4                     10          662                         50  Suburban
...                  ...          ...                        ...       ...
9995                   9          783                        -50     Urban
9996                   6          667                         50     Urban
9997                   3          637                         50     Urban
9998                  13          676                         50  Suburban
9999                   4          776                        -50  Suburban

      Premium_Adjustment_Region
0                            50
1                           100
2                           100
3                           100
4                            50
```

```
...                           ...
9995                          100
9996                          100
9997                          100
9998                           50
9999                           50

[10000 rows x 27 columns]>
```

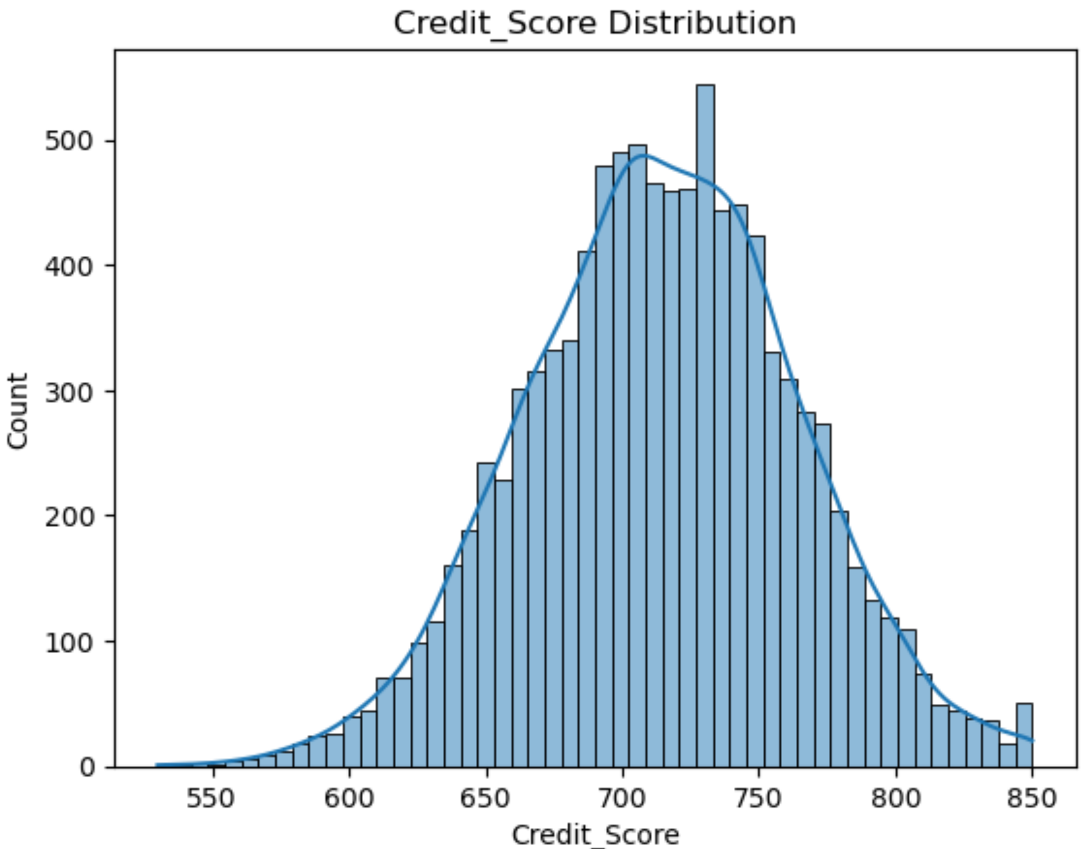
```
# Univariate Analysis: Numerical
sns.histplot(df['Age'], kde=True).set_title('Age Distribution')
plt.show()
```



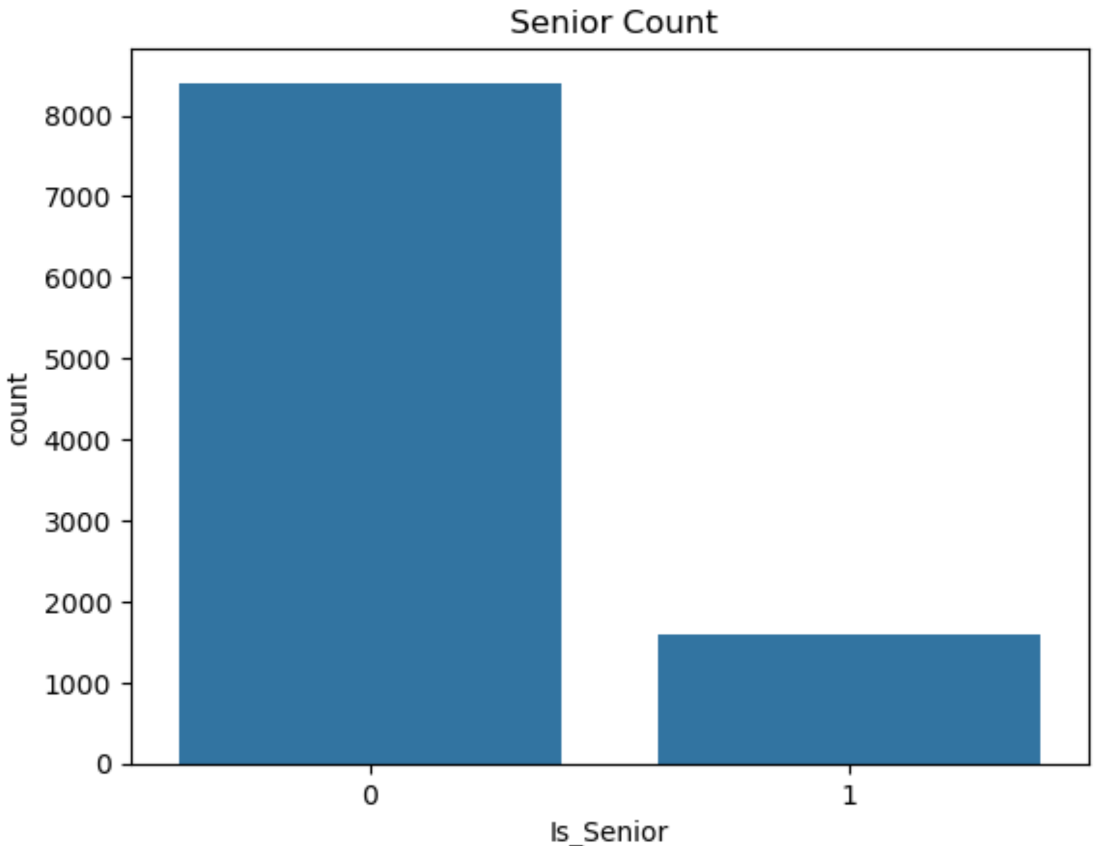

```
# Univariate Analysis: Numerical
sns.histplot(df['Time_Since_First_Contact'],
kde=True).set_title('Time_Since_First_Contact  Distribution')
plt.show()
```



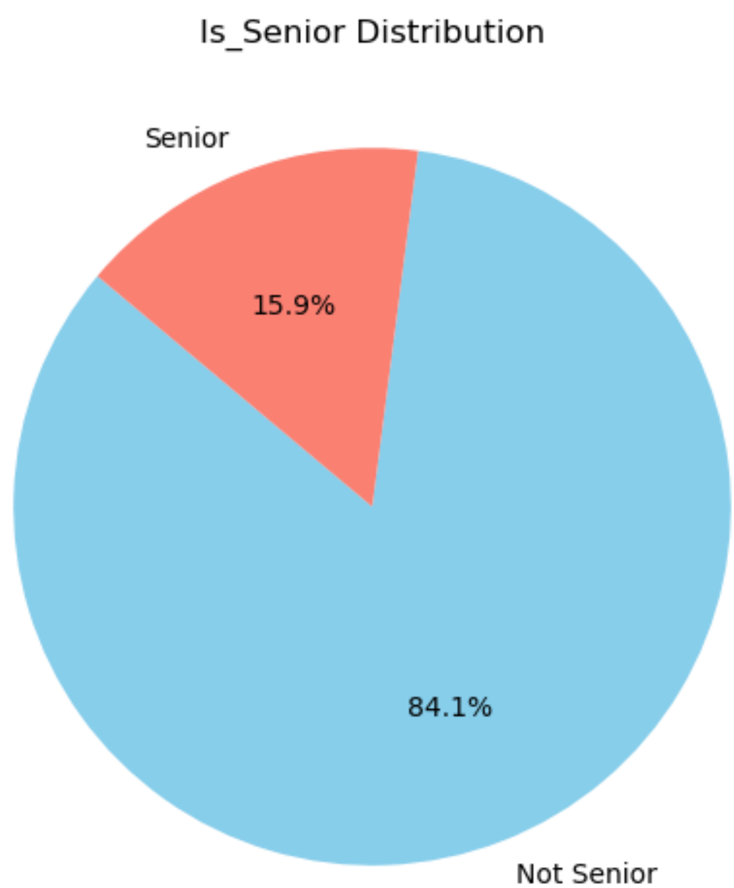

```
# Univariate Analysis: Numerical
sns.histplot(df['Credit_Score'], kde=True).set_title('Credit_Score
Distribution')
plt.show()
```



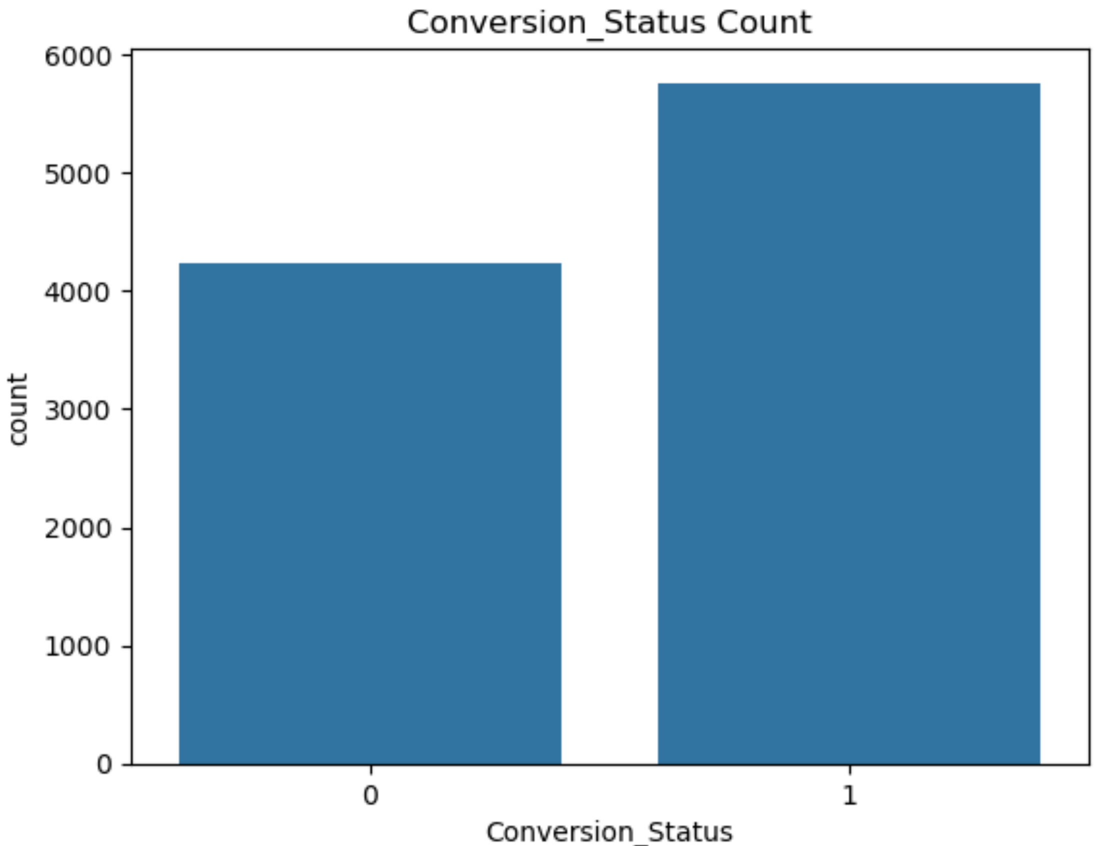

```
# Univariate Analysis: Categorical
sns.countplot(x='Is_Senior', data=df).set_title(' Senior Count')
plt.show()
```



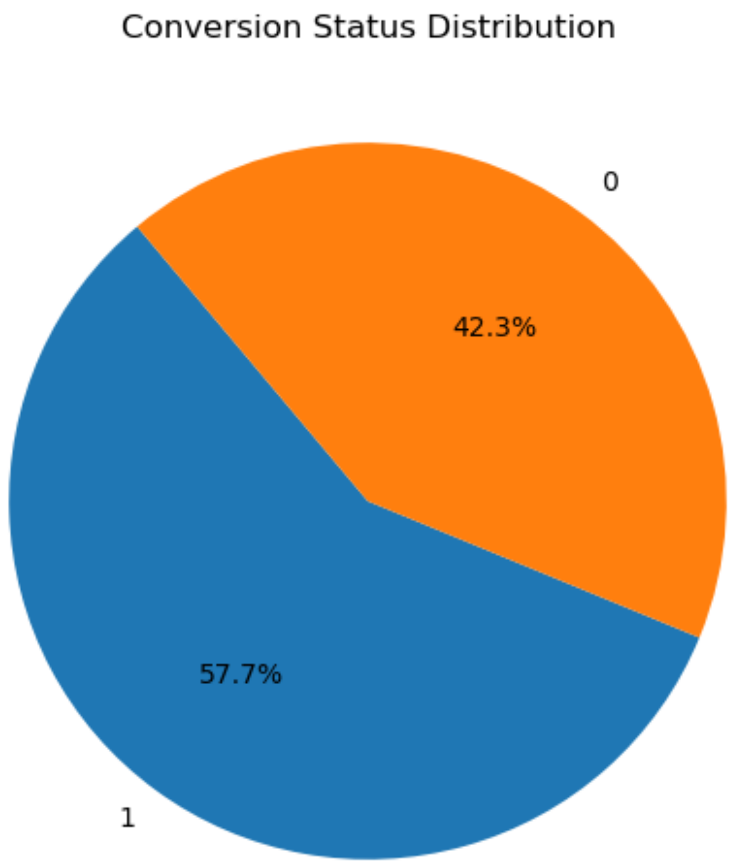

```
# Univariate Pie Chart: Senior distribution
depression_counts = df['Is_Senior'].value_counts()
plt.figure(figsize=(6, 6))
plt.pie(depression_counts, labels=['Not Senior', 'Senior'],
autopct='%1.1f%%', startangle=140, colors=['skyblue', 'salmon'])
plt.title('Is_Senior Distribution')
plt.show()
```



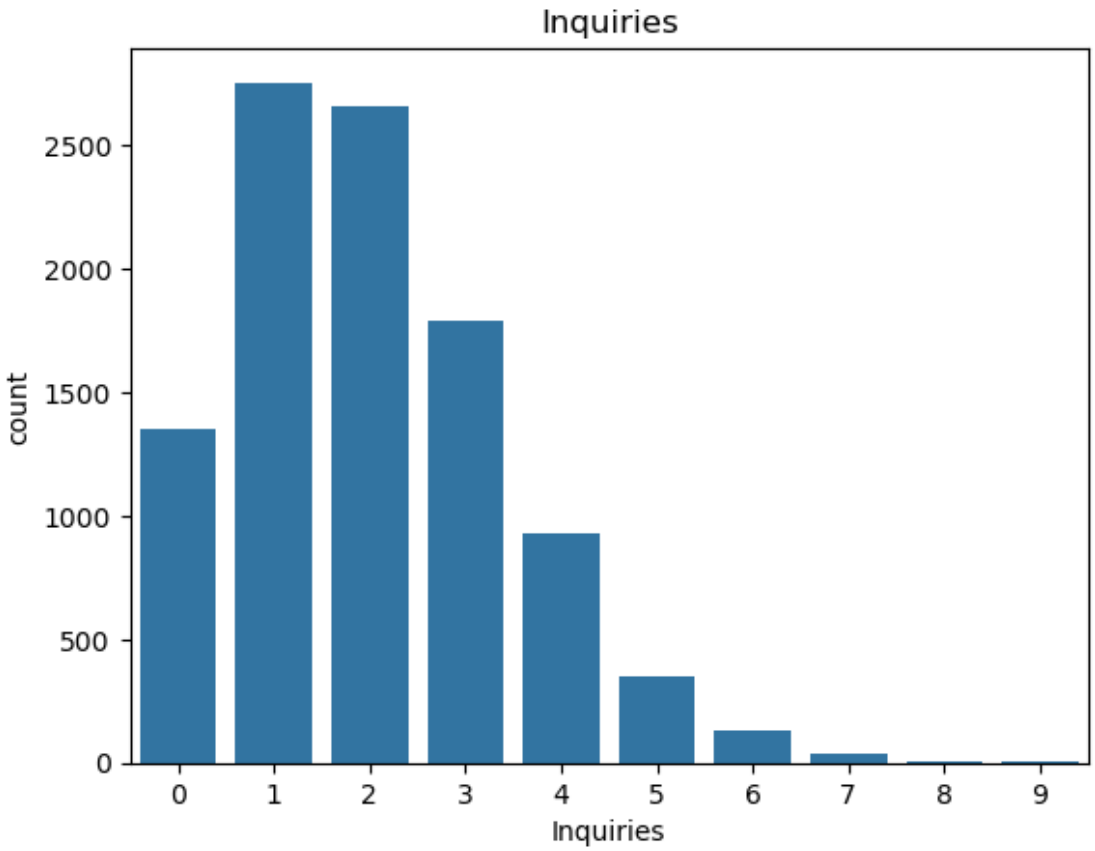

```
# Univariate Analysis: Categorical
sns.countplot(x='Conversion_Status', data=df).set_title('Conversion_Status
Count')
plt.show()
```



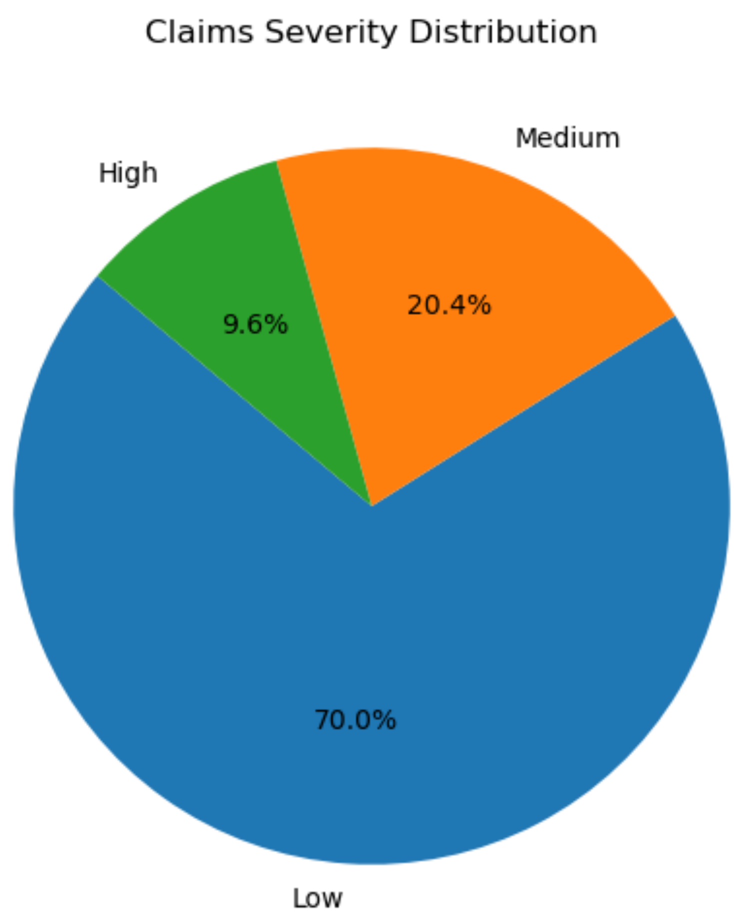

```
# Univariate Pie Chart: Conversion Status distribution
gender_counts = df['Conversion_Status'].value_counts()
plt.figure(figsize=(6, 6))
plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%',
startangle=130)
plt.title('Conversion Status Distribution')
plt.show()
```



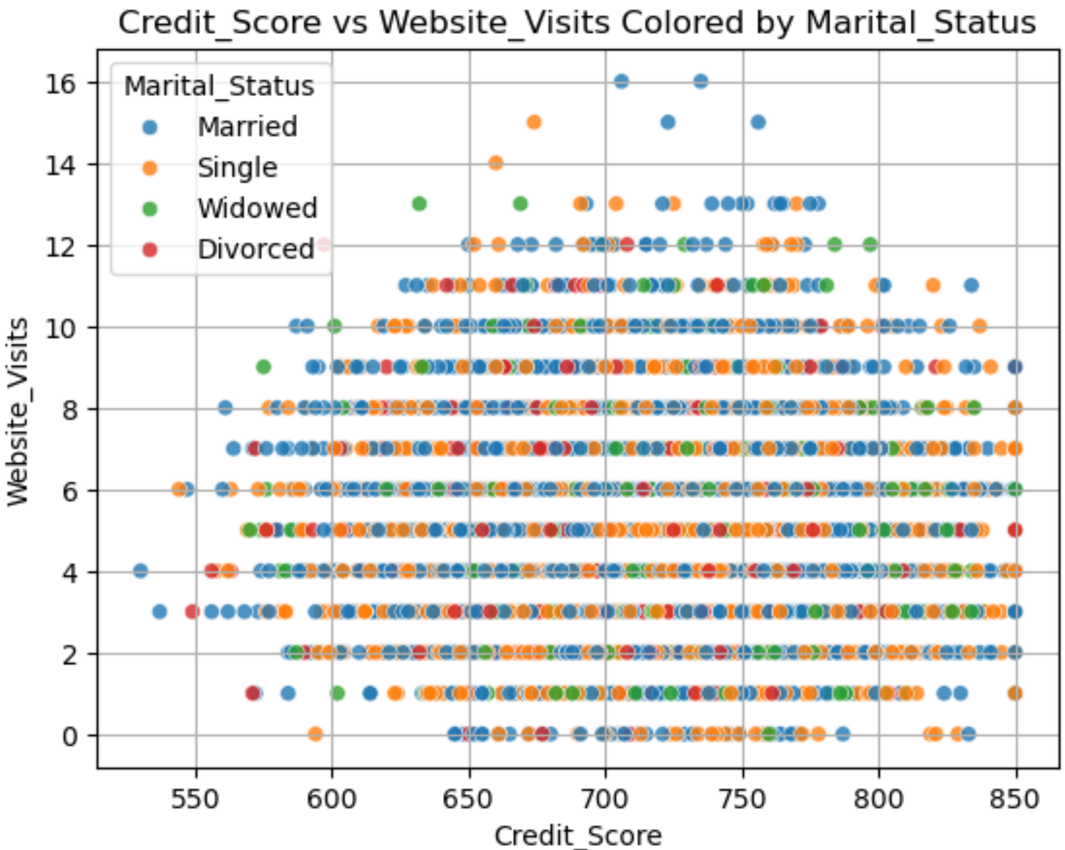

```
# Univariate Analysis: Categorical
sns.countplot(x='Inquiries', data=df).set_title('Inquiries')
plt.show()
```



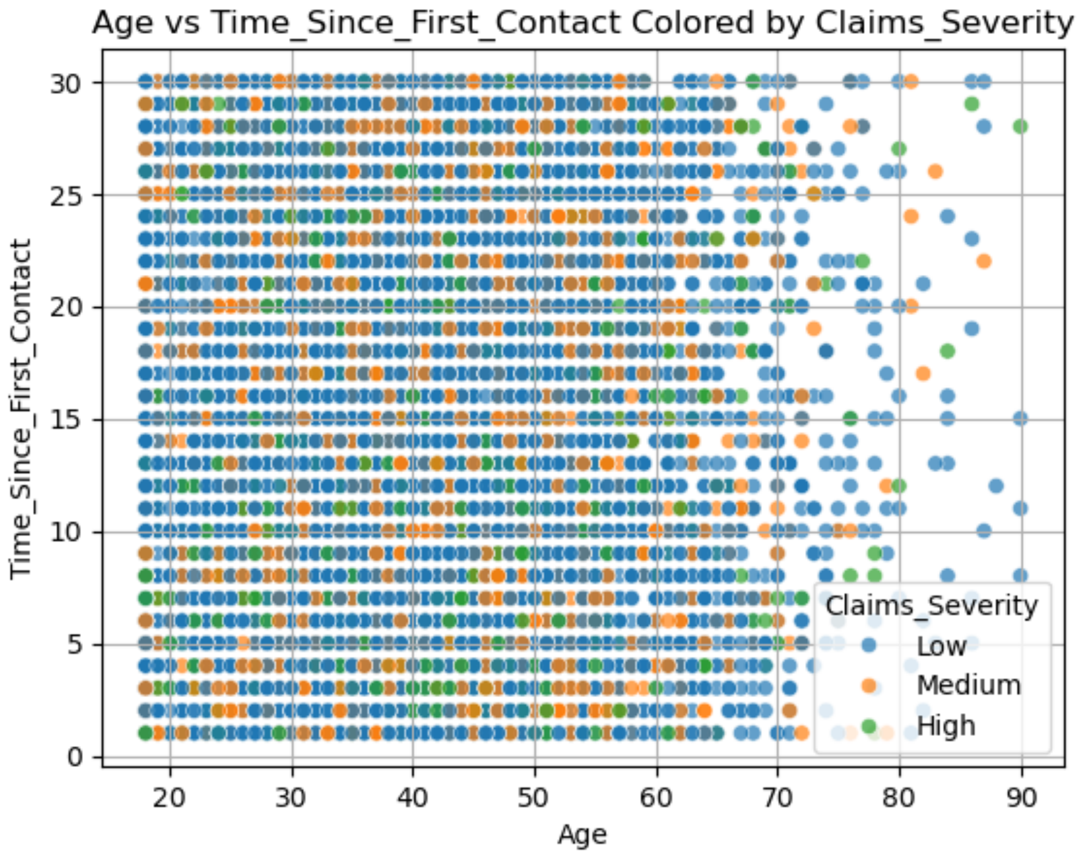

```
# Univariate Pie Chart: Claims_Severity
diet_counts = df['Claims_Severity'].value_counts()
plt.figure(figsize=(6, 6))
plt.pie(diet_counts, labels=diet_counts.index, autopct='%1.1f%%',
startangle=140)
plt.title('Claims Severity Distribution')
plt.show()
```



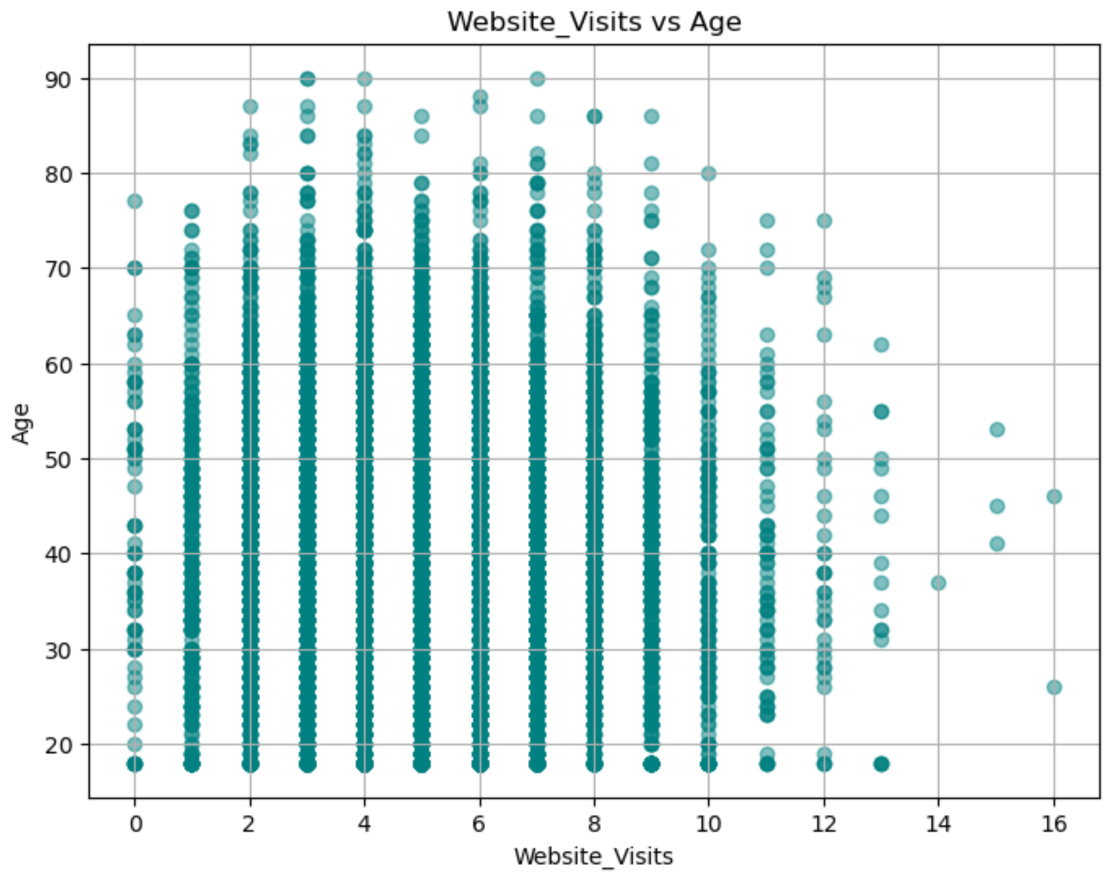

```
# Clean column names to remove any leading/trailing spaces
df.columns = df.columns.str.strip()

# Credit_Score vs  Website_Visits Colored by Marital_Status  scatter plot
sns.scatterplot(x='Credit_Score', y='Website_Visits', hue='Marital_Status',
data=df, alpha=0.8)
plt.title('Credit_Score vs Website_Visits Colored by Marital_Status')
plt.xlabel('Credit_Score')
plt.ylabel('Website_Visits')
plt.grid(True)
plt.show()
```



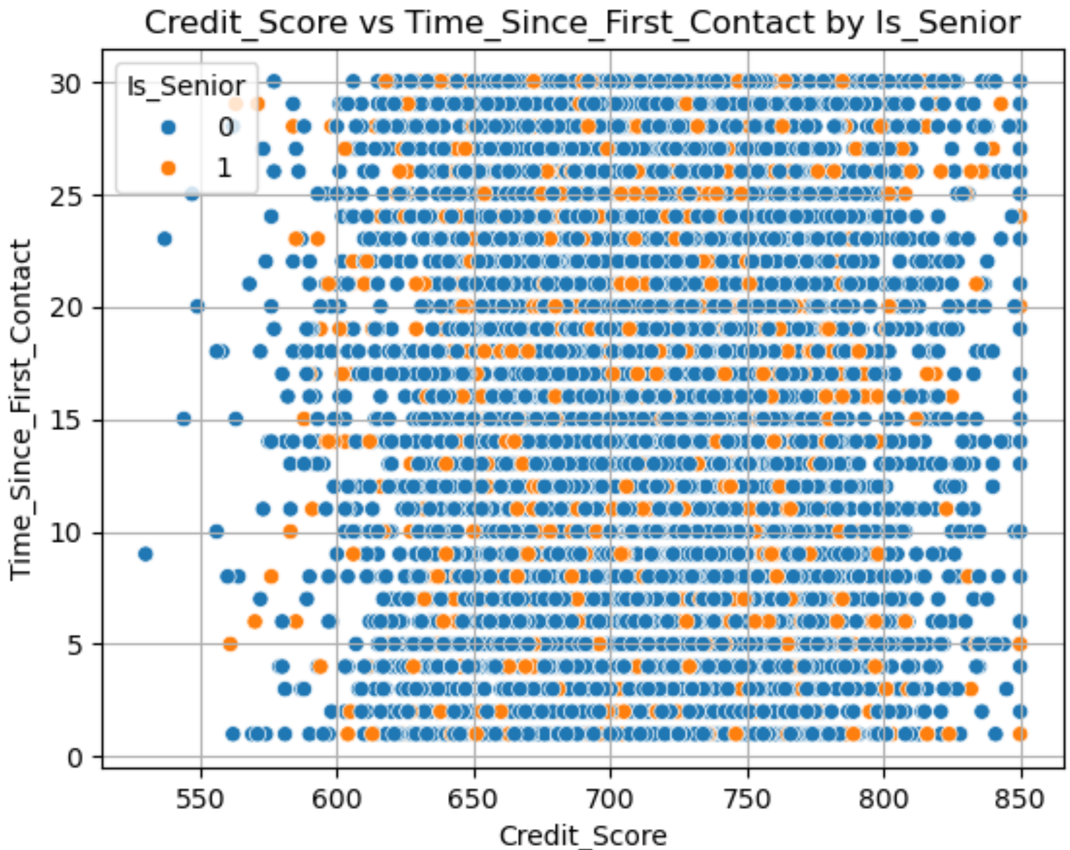

```
# Scatter plot colored by Claims_Severity
df.columns = df.columns.str.strip()
sns.scatterplot(x='Age', y='Time_Since_First_Contact', hue='Claims_Severity',
data=df, alpha=0.7)
plt.title('Age vs Time_Since_First_Contact Colored by Claims_Severity')
plt.xlabel('Age')
plt.ylabel('Time_Since_First_Contact')
plt.grid(True)
plt.show()
```



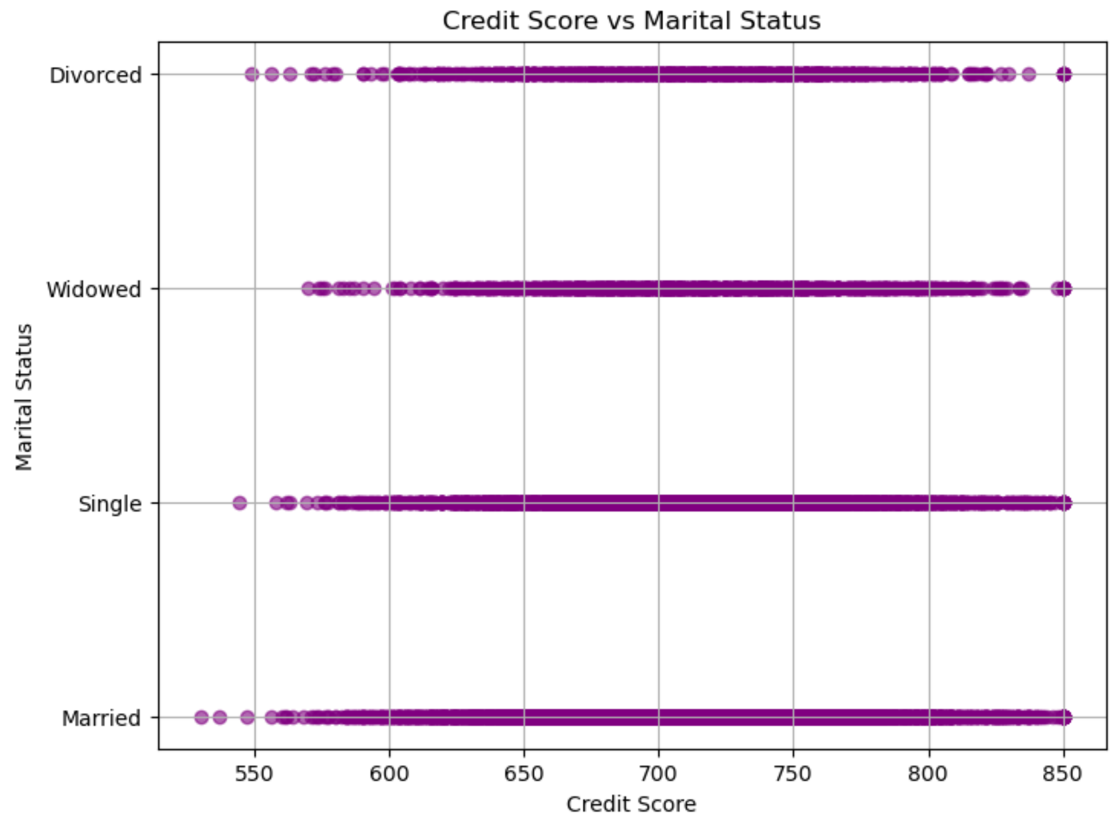

```
# Website_Visits vs Age scatter plot
plt.figure(figsize=(8, 6))
plt.scatter(df['Website_Visits'], df['Age'], color='teal', alpha=0.5)
plt.title('Website_Visits vs Age')
plt.xlabel('Website_Visits')
plt.ylabel('Age')
plt.grid(True)
plt.show()
```



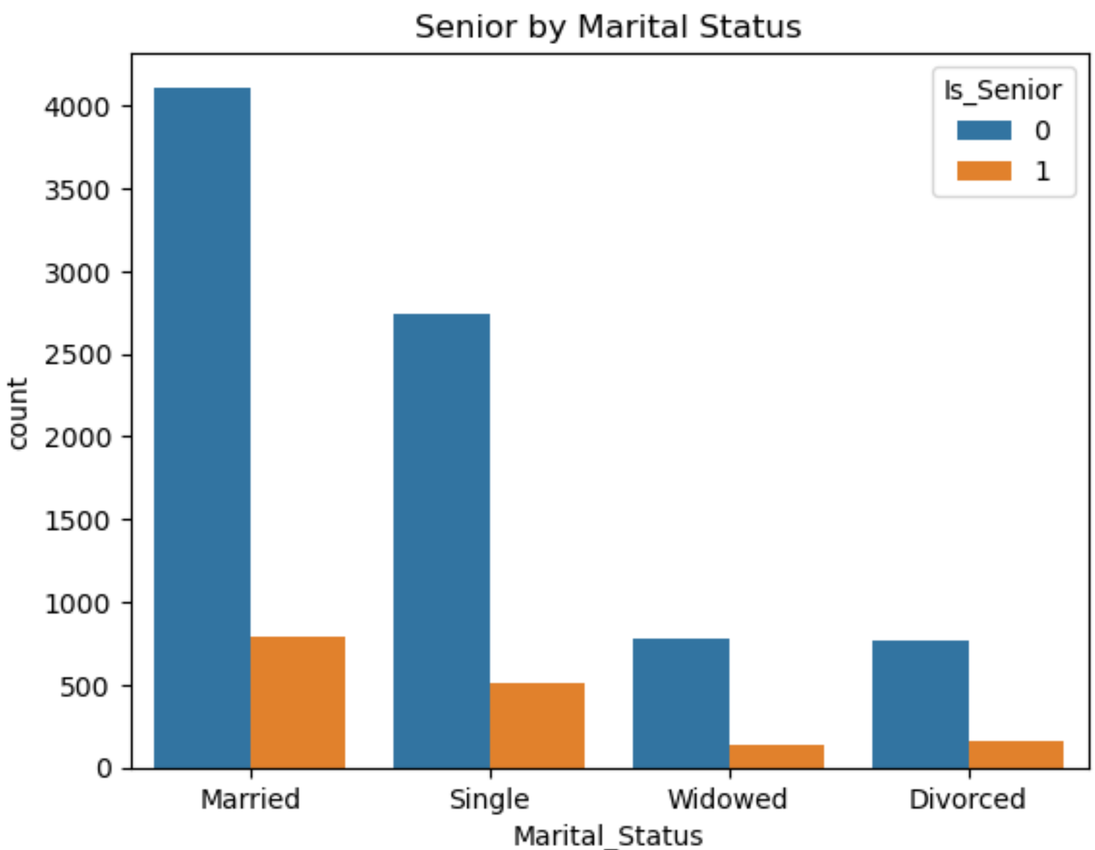

```
df.columns = df.columns.str.strip()

# Scatter plot with hue based on Conversion_Status
sns.scatterplot(x='Credit_Score', y='Time_Since_First_Contact',
hue='Is_Senior', data=df)
plt.title('Credit_Score vs Time_Since_First_Contact by Is_Senior')
plt.xlabel('Credit_Score')
plt.ylabel('Time_Since_First_Contact')
plt.grid(True)
plt.show()
```



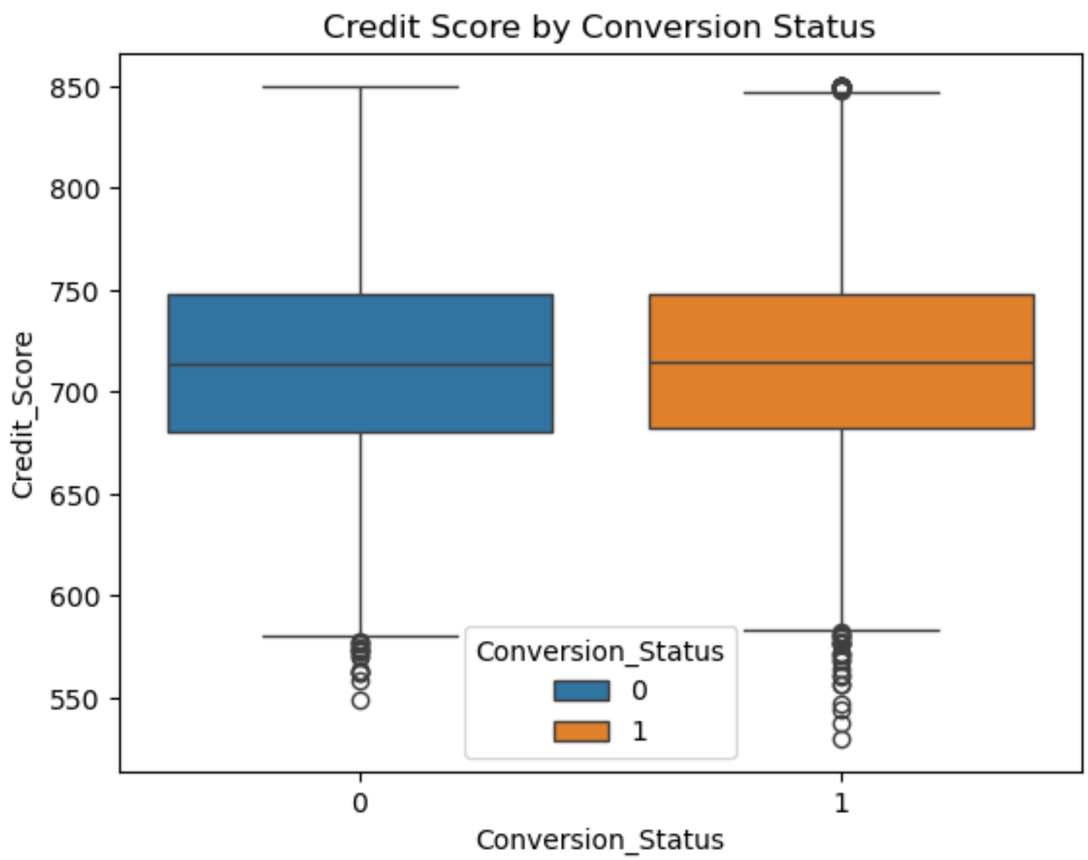

```python
# Scatter plot (Credit_Score vs Website_Visits)
plt.figure(figsize=(8, 6))
plt.scatter(df['Credit_Score'], df['Marital_Status'], alpha=0.6, c='purple')
plt.title('Credit Score vs Marital Status')
plt.xlabel('Credit Score')
plt.ylabel('Marital Status')
plt.grid(True)
plt.show()
```



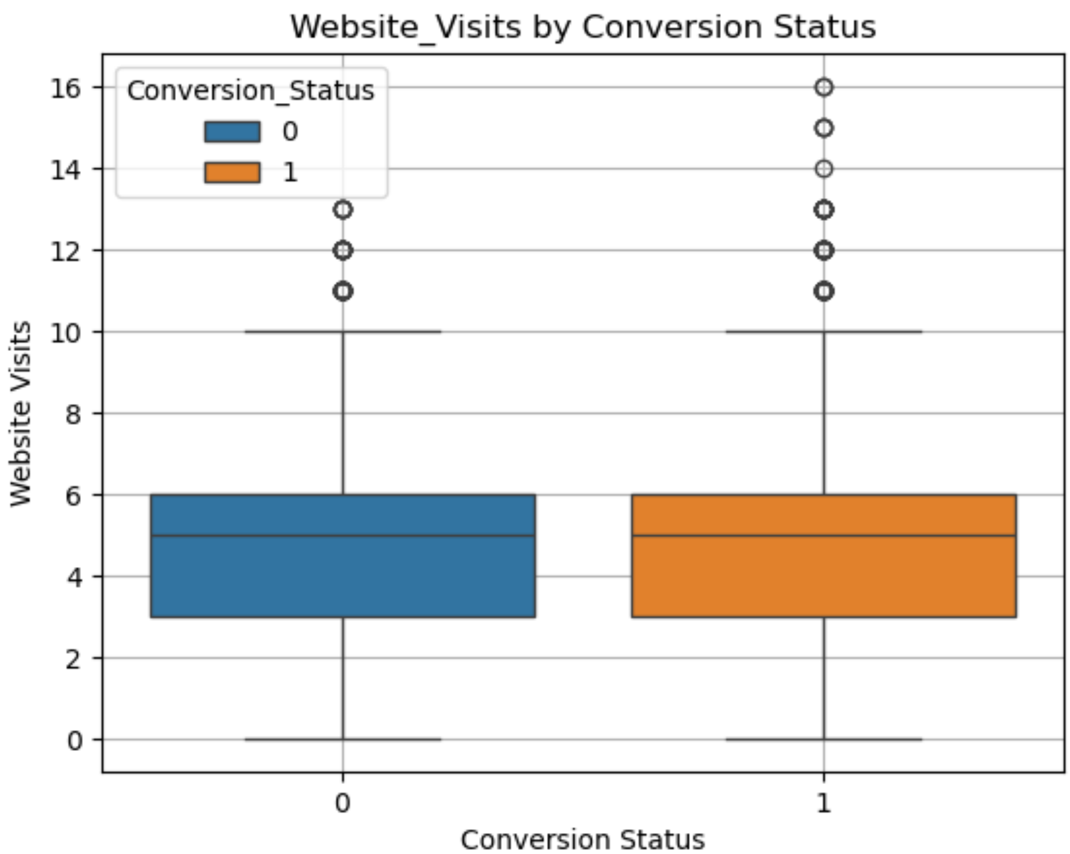

```
#  Bivariate Analysis Marital Status vs Senior
sns.countplot(x='Marital_Status', hue='Is_Senior', data=df)
plt.title('Senior by Marital Status')
plt.show()
```



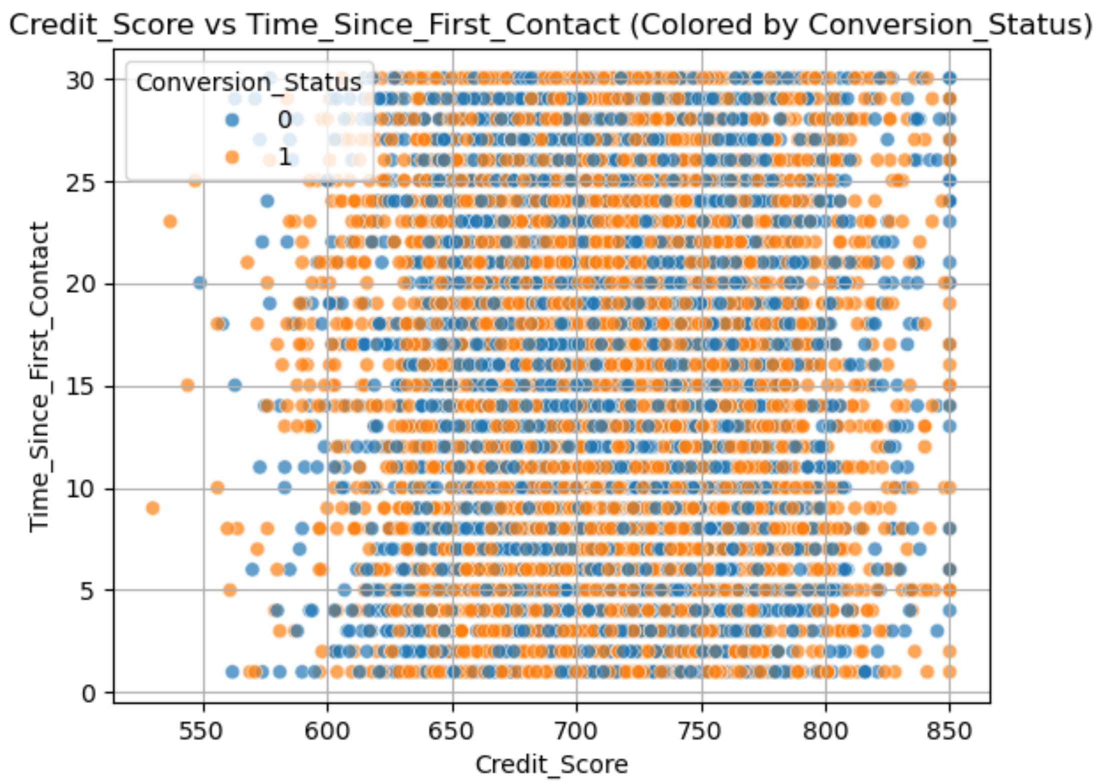

```
#  Bivariate Analysis Depression vs Credit_Score
sns.boxplot(x='Conversion_Status', y='Credit_Score',hue='Conversion_Status',
data=df)
plt.title('Credit Score by Conversion Status')
plt.show()
```




```
# Bivariate Analysis: Conversion_Status vs Website_Visits
sns.boxplot(x='Conversion_Status', y='Website_Visits',
hue='Conversion_Status', data=df)
plt.title('Website_Visits by Conversion Status')
plt.xlabel('Conversion Status')
plt.ylabel('Website Visits')
plt.grid(True)
plt.show()
```



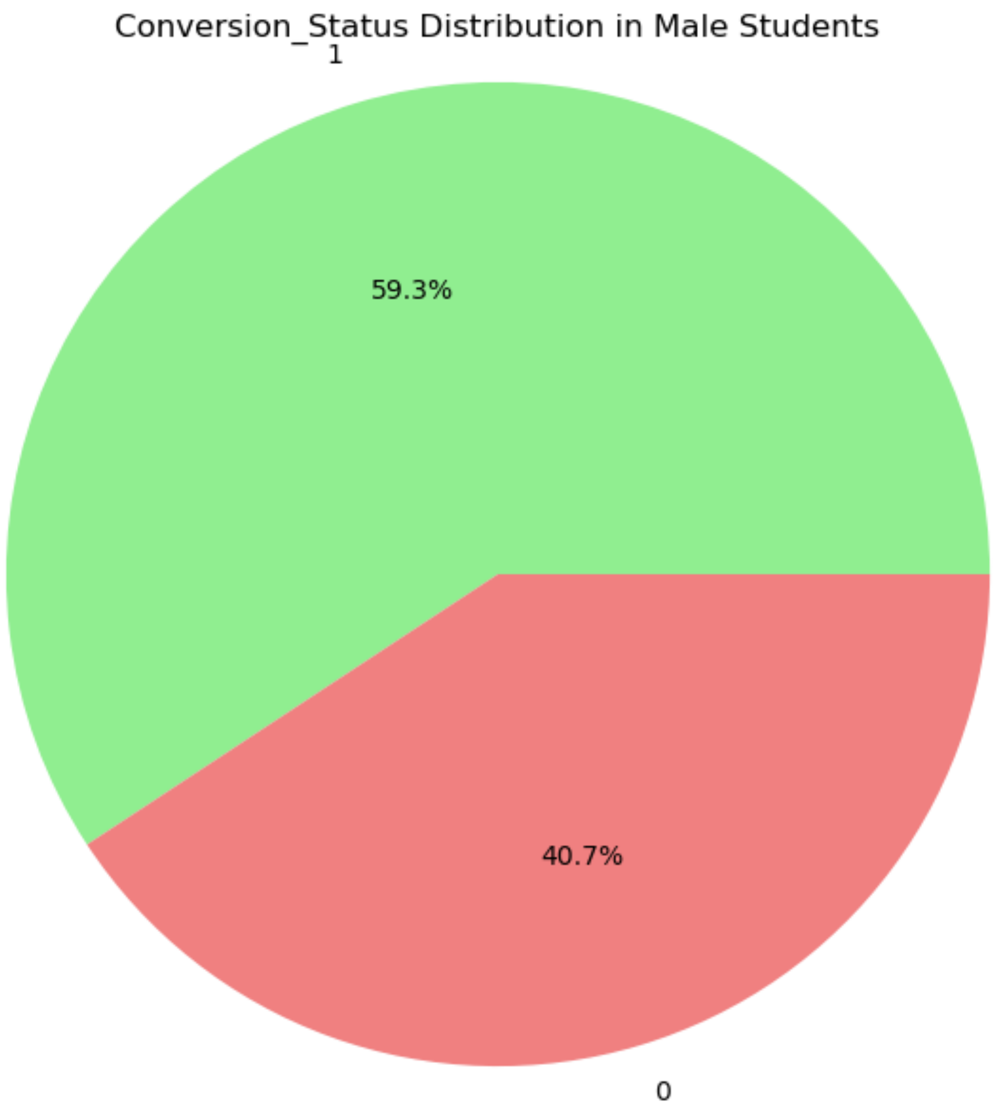

```
# Scatter Plot (Bivariate): Credit_Score vs Time_Since_First_Contact, colored
by Conversion_Status
sns.scatterplot(x='Credit_Score', y='Time_Since_First_Contact',
hue='Conversion_Status', data=df, alpha=0.7)
plt.title('Credit_Score vs Time_Since_First_Contact (Colored by
Conversion_Status)')
plt.xlabel('Credit_Score')
plt.ylabel('Time_Since_First_Contact')
plt.grid(True)
plt.show()
```

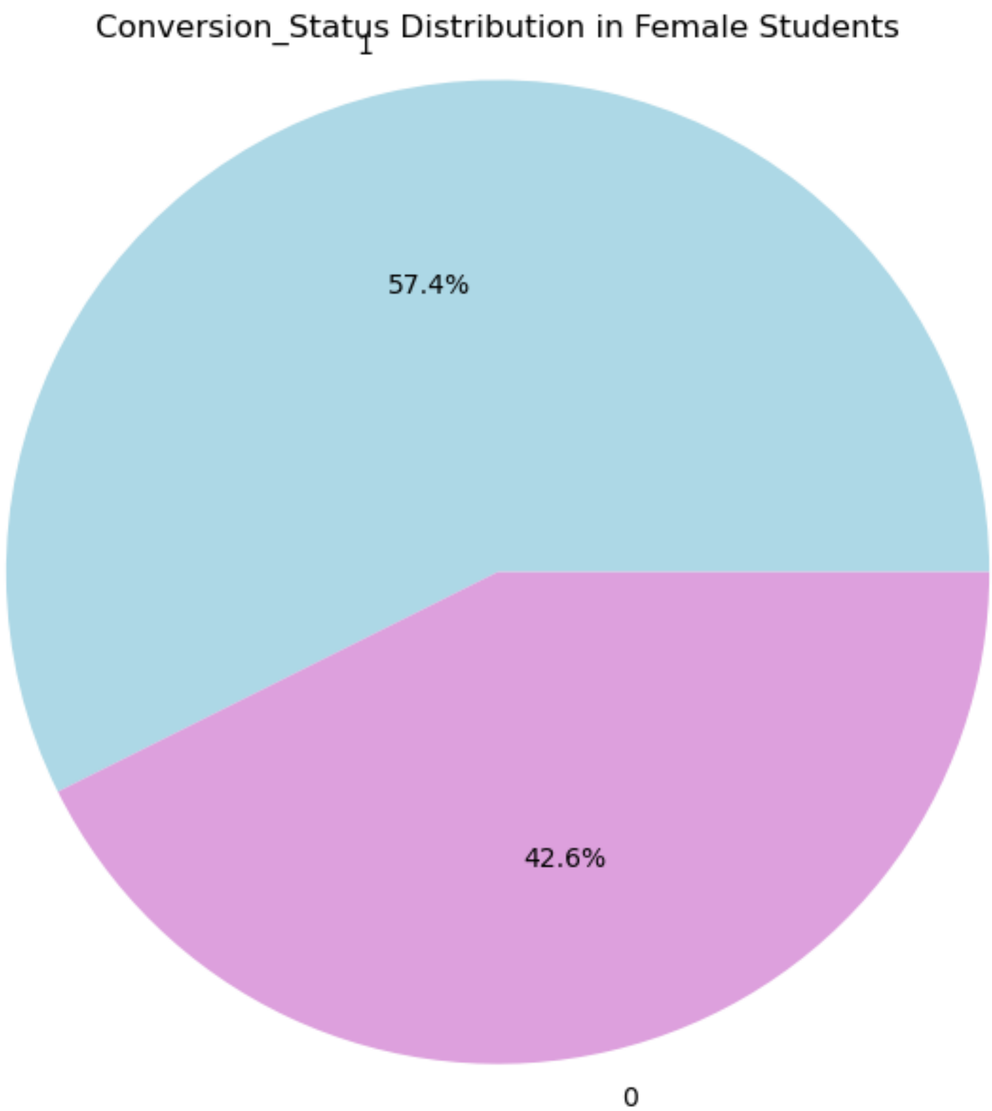



```
import matplotlib.pyplot as plt

# Clean column names (remove spaces)
df.columns = df.columns.str.strip()

# Show unique values for debugging
print("Is_Senior values:", df['Is_Senior'].unique())
print("Conversion_Status values:", df['Conversion_Status'].unique())

# Filter for Male (1) and Female (0) - check if they are int or str
male_df = df[df['Is_Senior'].astype(str) == '1']
female_df = df[df['Is_Senior'].astype(str) == '0']

# Get counts
male_counts = male_df['Conversion_Status'].value_counts()
female_counts = female_df['Conversion_Status'].value_counts()

# ----- Male Pie -----
if not male_counts.empty:
    plt.figure(figsize=(6, 6))
    plt.pie(male_counts, labels=male_counts.index, autopct='%1.1f%%',
colors=['lightgreen', 'lightcoral'])
    plt.title('Conversion_Status Distribution in Male Students')
    plt.axis('equal')  # Make pie chart round
    plt.tight_layout()
```

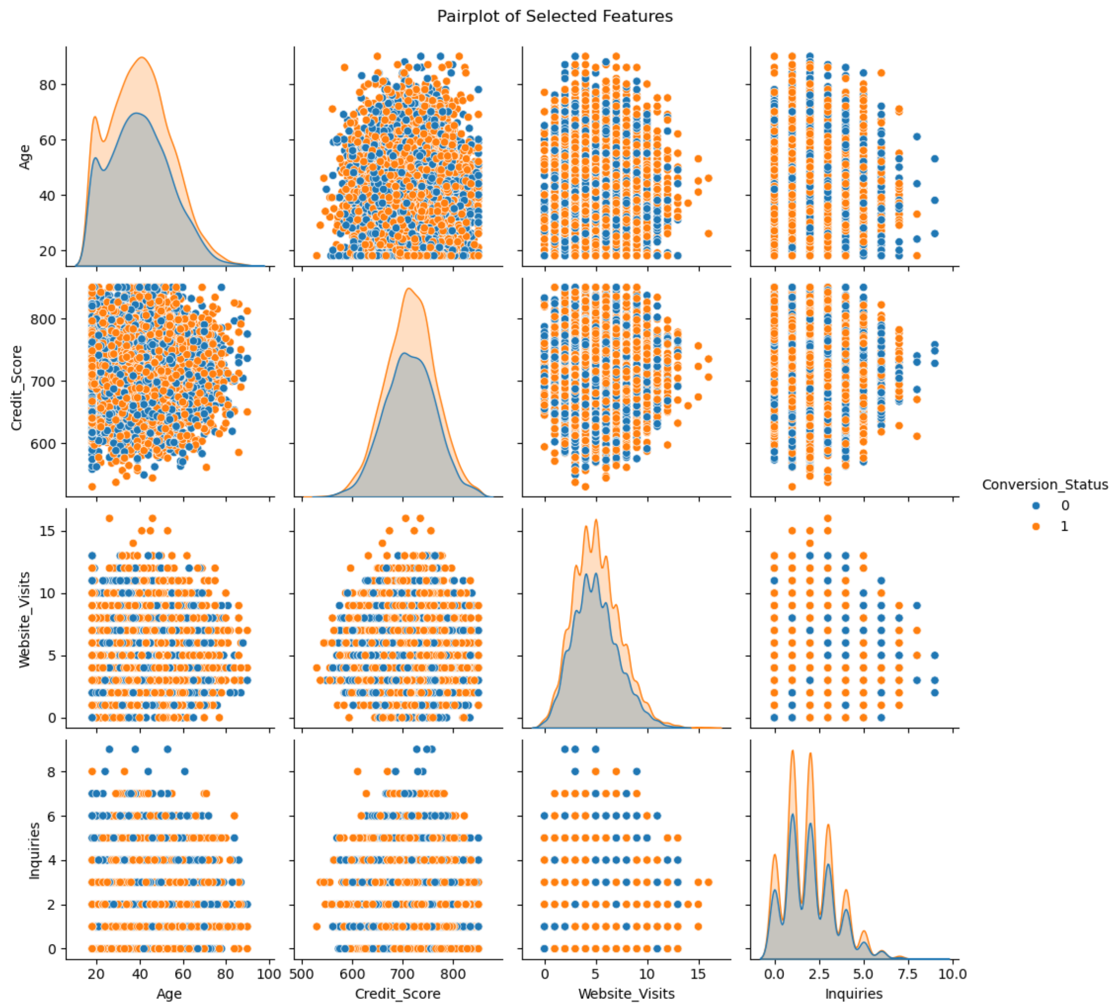
```
    plt.show()
else:
    print("⚠️ No Conversion_Status data found for Male students.")

# ----- Female Pie -----
if not female_counts.empty:
    plt.figure(figsize=(6, 6))
    plt.pie(female_counts, labels=female_counts.index, autopct='%1.1f%%', 
colors=['lightblue', 'plum'])
    plt.title('Conversion_Status Distribution in Female Students')
    plt.axis('equal')  # Make pie chart round
    plt.tight_layout()
    plt.show()
else:
    print("⚠️ No Conversion_Status data found for Female students.")
```

```
Is_Senior values: [0 1]
Conversion_Status values: [0 1]
```


Conversion_Status Distribution in Male Students
1
59.3%
40.7%
0

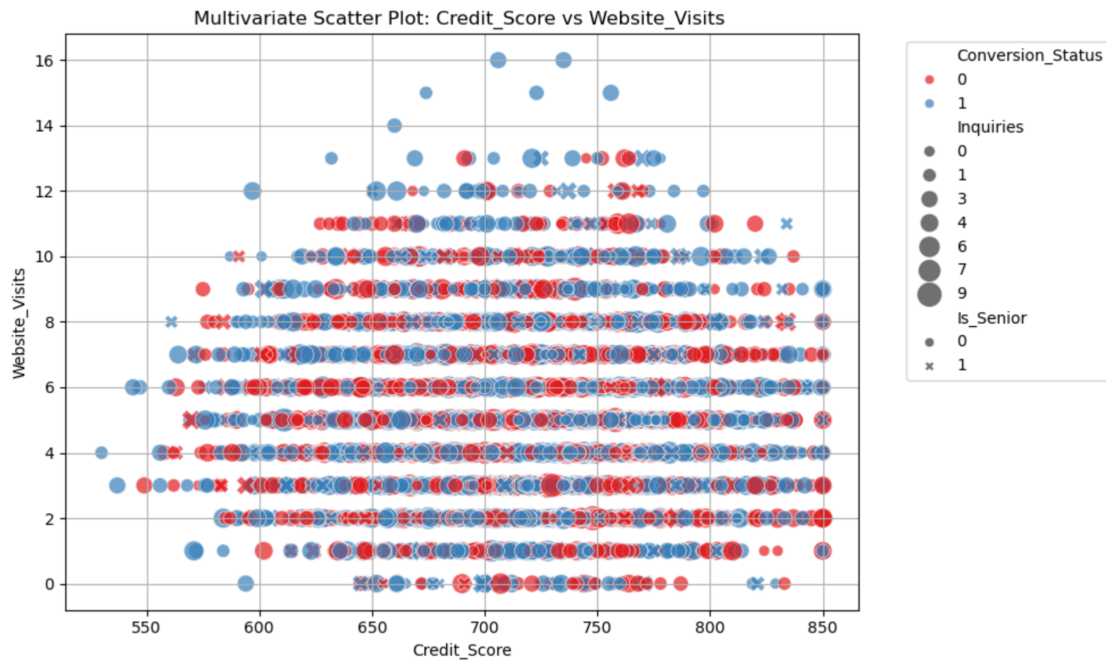



```
# Pairplot with Depression hue  Multivariate Analysis
selected_columns = ['Age', 'Credit_Score', 'Website_Visits', 'Inquiries',
'Conversion_Status']
sns.pairplot(df[selected_columns], hue='Conversion_Status')
plt.suptitle("Pairplot of Selected Features", y=1.02)
plt.show()
```

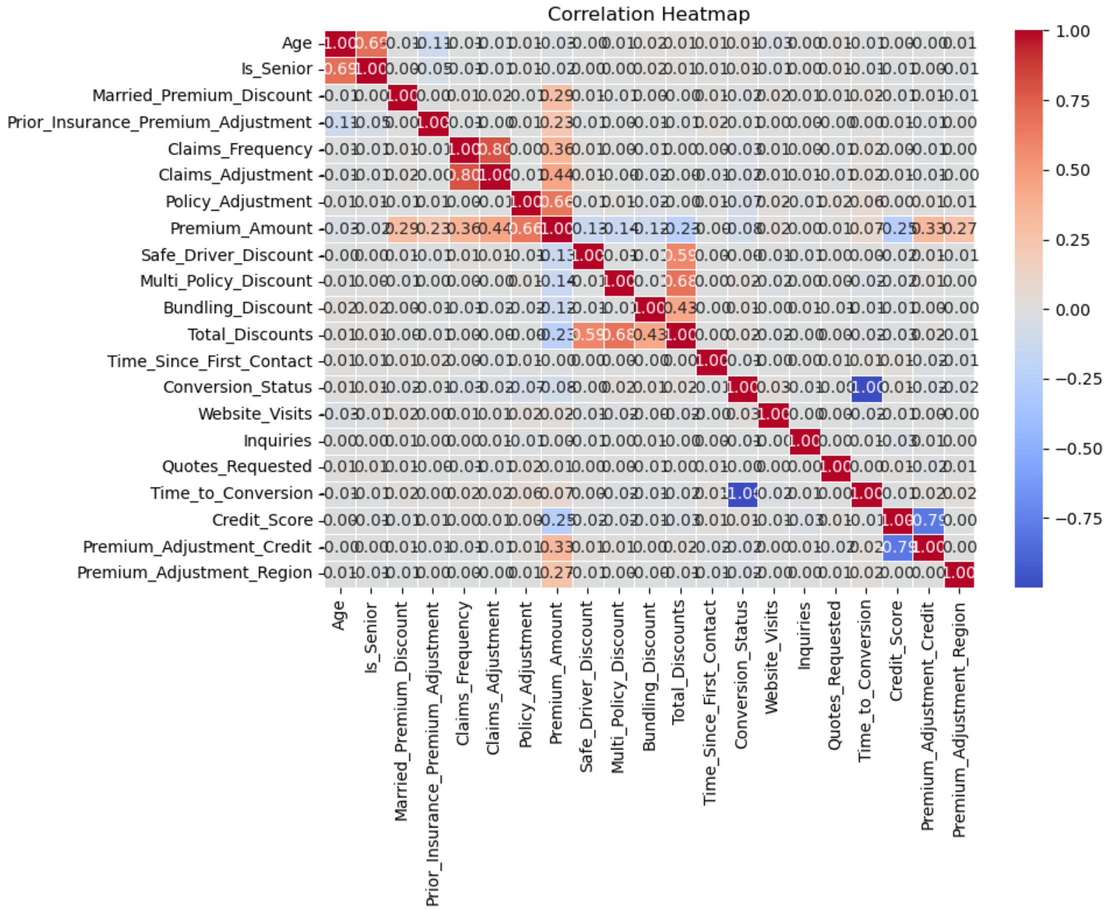



```python
# Multivariate Scatter Plot
plt.figure(figsize=(10, 6))
sns.scatterplot(
    x='Credit_Score',
    y='Website_Visits',
    hue='Conversion_Status',
    size='Inquiries',
    style='Is_Senior',
    data=df,
    palette='Set1',
    sizes=(50, 250),  # size range for points
    alpha=0.7
)

plt.title('Multivariate Scatter Plot: Credit_Score vs Website_Visits ')
plt.xlabel('Credit_Score')
plt.ylabel('Website_Visits ')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True)
```

```
plt.tight_layout()
plt.show()
```




```
# Correlation Heatmap
plt.figure(figsize=(10, 8))  # Optional: set figure size for better
readability
corr_matrix = df.corr(numeric_only=True)
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt=".2f",
linewidths=0.5)
plt.title('Correlation Heatmap')
plt.tight_layout()
plt.show()
```



Conclusion

In this analysis, we looked closely at a sales dataset by creating visual representations of important variables like 'Units Sold' and 'Unit Price' using histograms and boxplots. The histogram for 'Unit Price' showed us how prices are spread out across different products, helping us see where most prices fall and if there are any unusual pricing patterns. The boxplot for 'Unit Price' highlighted the average price, the range of prices, and any outliers that might indicate pricing issues. Similarly, examining 'Units Sold' gave us insights into which products are popular and helped us spot any unusual sales figures that might need further attention.
These visual tools not only helped us understand the data better but also provided a basis for making smart business decisions. By recognizing trends in sales and pricing, companies can adjust their strategies to meet customer needs, manage inventory more effectively, and improve overall sales performance.

Final Thoughts

Understanding how products are selling and how they are priced is essential for making good business choices. The insights from this analysis can guide businesses in managing their stock, changing prices, and boosting sales. For example, if we find that some products are selling well at low prices, it might be a good idea to raise those prices to increase profits. On the other hand, if some products are priced high but not selling well, they might need discounts or promotions to attract buyers.

Looking ahead, we could dive deeper into the data by exploring relationships between different factors, breaking down sales by product categories, or analyzing sales trends over time. Adding information about customer preferences and buying habits could also enhance our understanding and help create more effective marketing strategies.

Overall, using visual tools to analyze data has shown to be a powerful way to uncover important insights. By continually examining and interpreting sales data, businesses can stay flexible and responsive to changes in the market, ultimately leading to growth and success in a competitive environment.