jupyter Accident_Severity Last Checkpoint: 15 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [3]:
```
#import libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')

#Load and read the file
df=pd.read_csv("RTA Dataset.csv")
df.head()
```

Out[3]:

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experience	Type_of_vehicle	Owner_of_vehicle
0	17:02:00	Monday	18-30	Male	Above high school	Employee	1-2yr	Automobile	Owner
1	17:02:00	Monday	31-50	Male	Junior high school	Employee	Above 10yr	Public (> 45 seats)	Owner
2	17:02:00	Monday	18-30	Male	Junior high school	Employee	1-2yr	Lorry (41?100Q)	Owner
3	1:06:00	Sunday	18-30	Male	Junior high school	Employee	5-10yr	Public (> 45 seats)	Governmental
4	1:06:00	Sunday	18-30	Male	Junior high school	Employee	2-5yr	NaN	Owner

5 rows × 32 columns

In [4]: #shape/ size of the data

jupyter Accident_Severity Last Checkpoint: 15 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

In [4]:
```
#shape/ size of the data
df.shape
```

Out[4]: (12316, 32)

In [5]:
```
#checking the numerical statistics of the data
df.describe()
```

Out[5]:

	Number_of_vehicles_involved	Number_of_casualties
count	12316.000000	12316.000000
mean	2.040679	1.548149
std	0.688790	1.007179
min	1.000000	1.000000
25%	2.000000	1.000000
50%	2.000000	1.000000
75%	2.000000	2.000000
max	7.000000	8.000000

In [6]:
```
df.describe(include="all")
```

Out[6]:

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experience	Type_of_vehicle	Owner_of_veh
count	12316	12316	12316	12316	11575	11737	11487	11366	11
unique	1074	7	5	3	7	4	7	17	

8:43 AM 10/27/2023

jupyter Accident_Severity Last Checkpoint: 16 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

Out[6]:

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experience	Type_of_vehicle	Owner_of_veh
count	12316	12316	12316	12316	11575	11737	11487	11366	11
unique	1074	7	5	3	7	4	7	17	
top	15:30:00	Friday	18-30	Male	Junior high school	Employee	5-10yr	Automobile	Ow
freq	120	2041	4271	11437	7619	9627	3363	3205	10
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

11 rows × 32 columns

In [7]:
```
#checking data types of each columns
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12316 entries, 0 to 12315
Data columns (total 32 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Time                         12316 non-null  object
```

jupyter Accident_Severity Last Checkpoint: 16 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

```
 1   Day_of_week                  12316 non-null  object
 2   Age_band_of_driver           12316 non-null  object
 3   Sex_of_driver                12316 non-null  object
 4   Educational_level            11575 non-null  object
 5   Vehicle_driver_relation      11737 non-null  object
 6   Driving_experience           11487 non-null  object
 7   Type_of_vehicle              11366 non-null  object
 8   Owner_of_vehicle             11834 non-null  object
 9   Service_year_of_vehicle      8388 non-null   object
 10  Defect_of_vehicle            7889 non-null   object
 11  Area_accident_occured        12077 non-null  object
 12  Lanes_or_Medians             11931 non-null  object
 13  Road_allignment              12174 non-null  object
 14  Types_of_Junction            11429 non-null  object
 15  Road_surface_type            12144 non-null  object
 16  Road_surface_conditions      12316 non-null  object
 17  Light_conditions             12316 non-null  object
 18  Weather_conditions           12316 non-null  object
 19  Type_of_collision            12161 non-null  object
 20  Number_of_vehicles_involved  12316 non-null  int64
 21  Number_of_casualties         12316 non-null  int64
 22  Vehicle_movement             12008 non-null  object
 23  Casualty_class               12316 non-null  object
 24  Sex_of_casualty              12316 non-null  object
 25  Age_band_of_casualty         12316 non-null  object
 26  Casualty_severity            12316 non-null  object
 27  Work_of_casuality            9118 non-null   object
 28  Fitness_of_casuality         9681 non-null   object
 29  Pedestrian_movement          12316 non-null  object
 30  Cause_of_accident            12316 non-null  object
 31  Accident_severity            12316 non-null  object
dtypes: int64(2), object(30)
```

Data science ass2.pdf | Untitled Folder/ | Accident_Severity - Jupyter Note

localhost:8888/notebooks/Untitled%20Folder/Accident_Severity.ipynb

jupyter Accident_Severity Last Checkpoint: 16 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

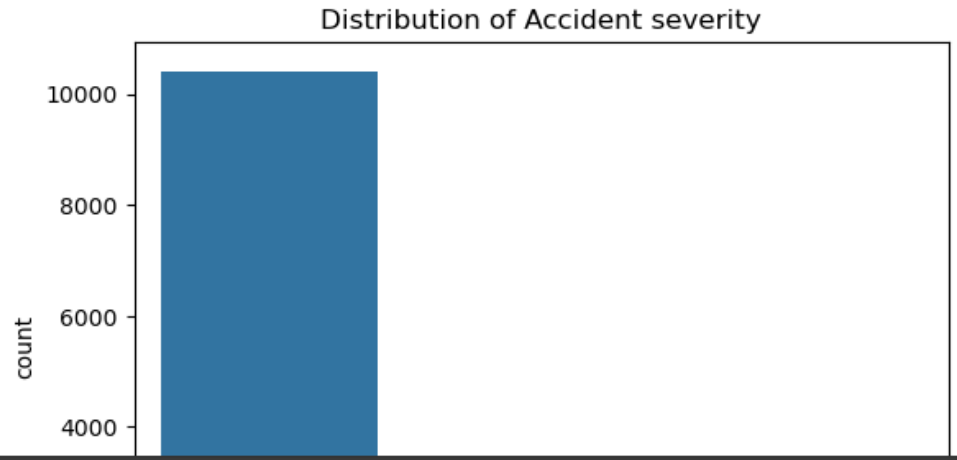
Run | Code

In [8]:
```
#Distribution of Accident severity
df['Accident_severity'].value_counts()
```

Out[8]:
```
Accident_severity
Slight Injury     10415
Serious Injury     1743
Fatal injury        158
Name: count, dtype: int64
```

In [9]:
```
#plotting the final class
sns.countplot(x = df['Accident_severity'])
plt.title('Distribution of Accident severity')
```

Out[9]:
```
Text(0.5, 1.0, 'Distribution of Accident severity')
```



8:43 AM
10/27/2023

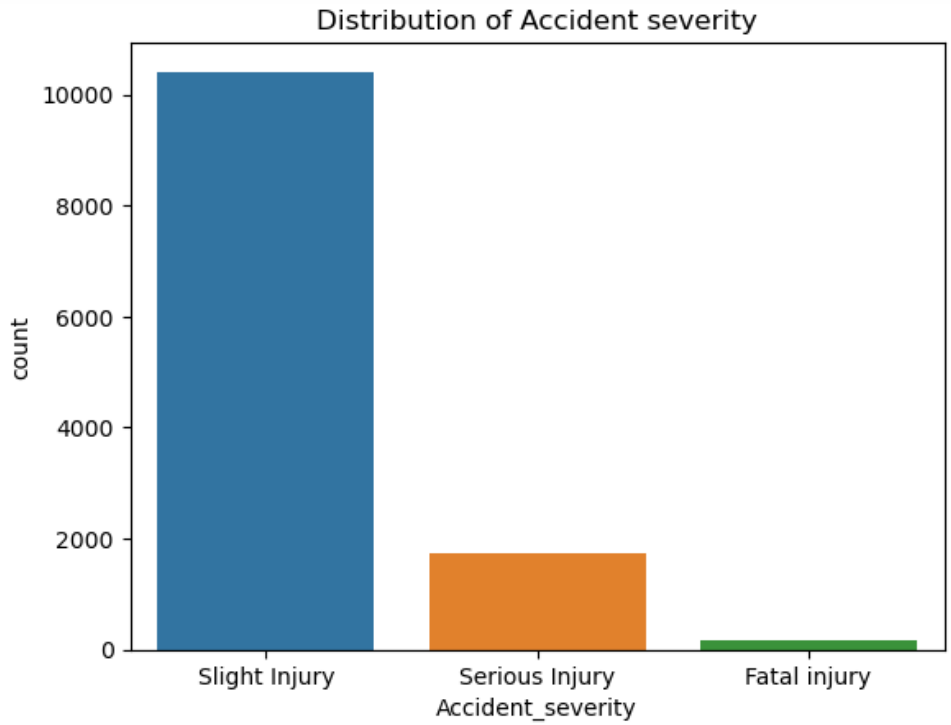
jupyter Accident_Severity Last Checkpoint: 16 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code



In [10]:
```
#plotting relationship between Number_of_casualties and Number_of_vehicles_involved
sns.scatterplot(x=df['Number_of_casualties'], y=df['Number_of_vehicles_involved'], hue=df['Accident_severity'])
```

Out[10]: <Axes: xlabel='Number_of_casualties', ylabel='Number_of_vehicles_involved'>

Data science ass2.pdf | Untitled Folder/ | Accident_Severity - Jupyter Note

localhost:8888/notebooks/Untitled%20Folder/Accident_Severity.ipynb

jupyter Accident_Severity Last Checkpoint: 16 minutes ago (autosaved)

Logout

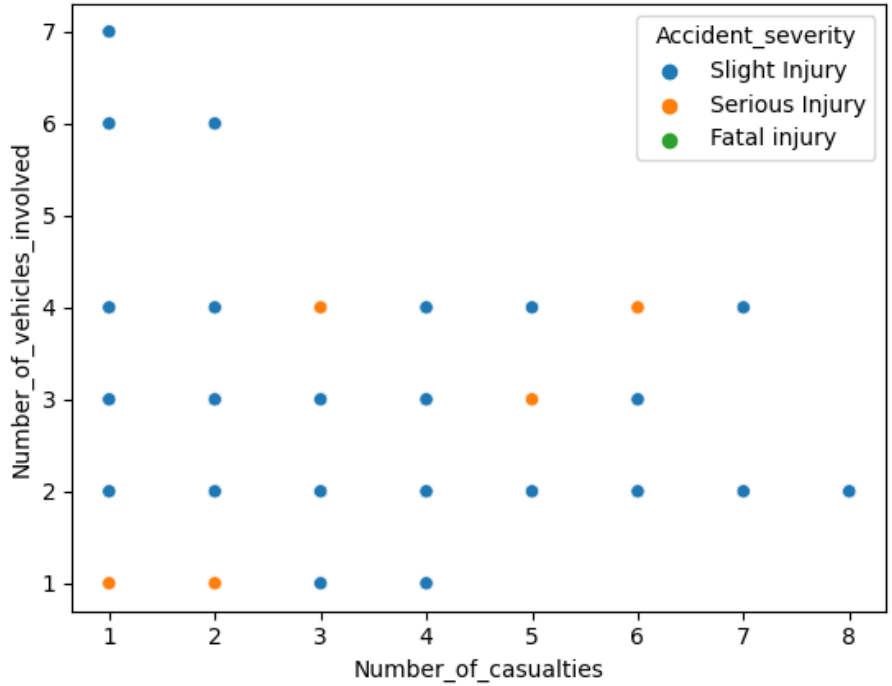
File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [10]:
```
#plotting relationship between Number_of_casualties and Number_of_vehicles_involved
sns.scatterplot(x=df['Number_of_casualties'], y=df['Number_of_vehicles_involved'], hue=df['Accident_severity'])
```

Out[10]: <Axes: xlabel='Number_of_casualties', ylabel='Number_of_vehicles_involved'>



8:44 AM
10/27/2023

Data science ass2.pdf | Untitled Folder/ | Accident_Severity - Jupyter Note

localhost:8888/notebooks/Untitled%20Folder/Accident_Severity.ipynb

jupyter Accident_Severity Last Checkpoint: 16 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

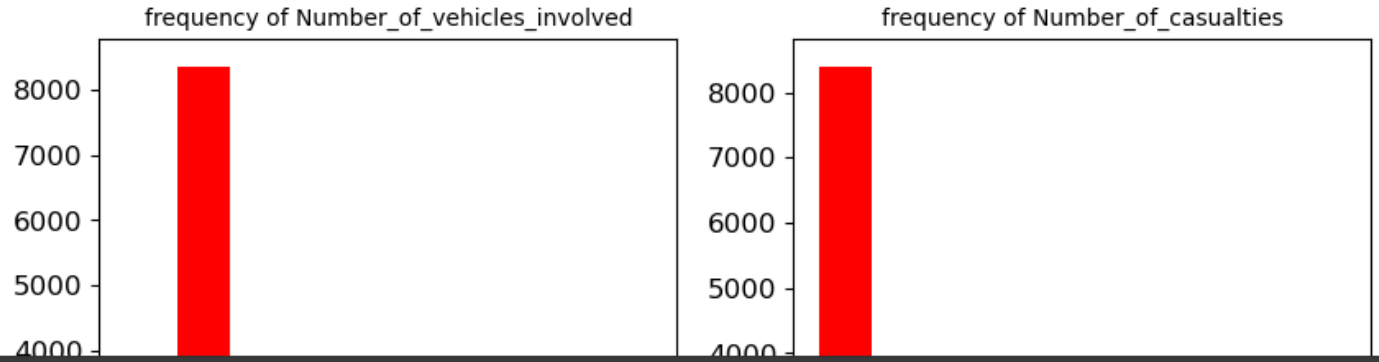
Trusted | Python 3 (ipykernel)

Run | Code

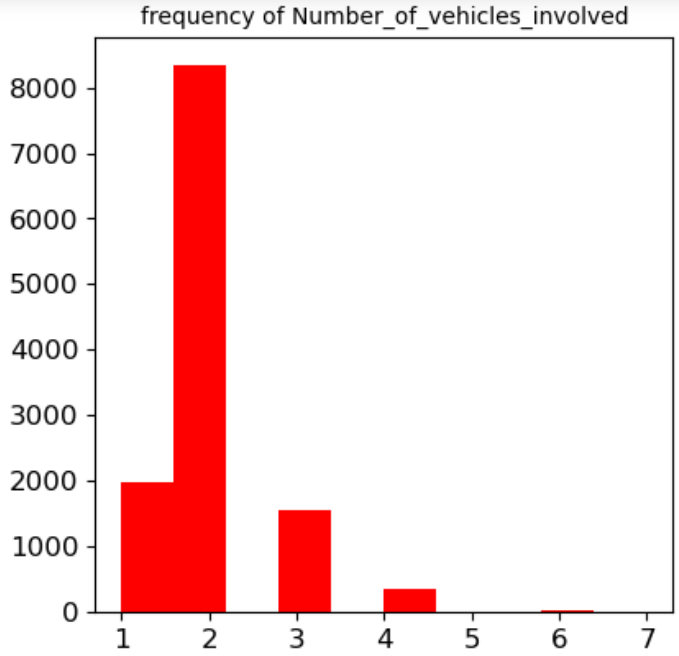
In [13]:
```
#storing numerical column names to a variable
numerical=[i for i in df.columns if df[i].dtype!='O']
print('The numerica variables are',numerical)
```

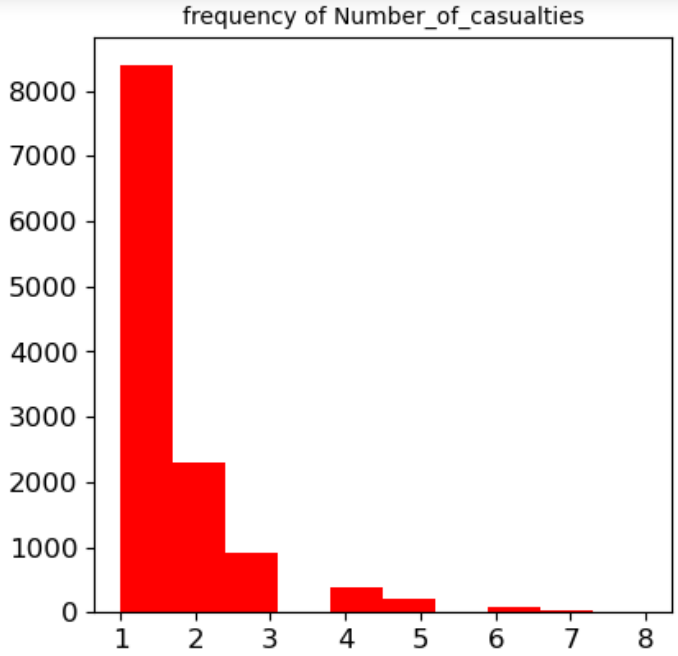
```
The numerica variables are ['Number_of_vehicles_involved', 'Number_of_casualties']
```

In [14]:
```
#distribution for numerical columns
plt.figure(figsize=(10,10))
plotnumber = 1
for i in numerical:
    if plotnumber <= df.shape[1]:
        ax1 = plt.subplot(2,2,plotnumber)
        plt.hist(df[i],color='red')
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title('frequency of '+i, fontsize=10)
    plotnumber +=1
```



8:44 AM 10/27/2023





In [17]:
```
#storing categorical column names to a new variable
categorical=[i for i in df.columns if df[i].dtype=='O']
print('The categorical variables are',categorical)
```

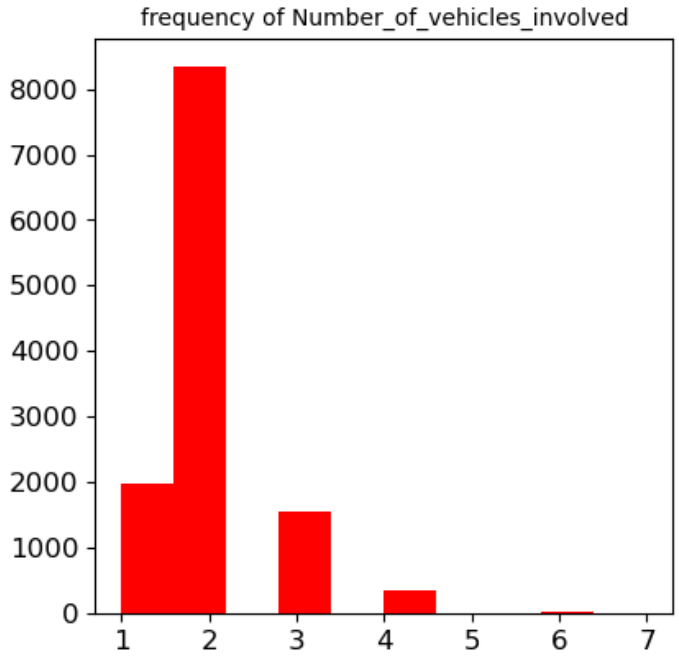
```
The categorical variables are ['Time', 'Day_of_week', 'Age_band_of_driver', 'Sex_of_driver', 'Educational_level', 'Vehicle_driver_relation', 'Driving_experience', 'Type_of_vehicle', 'Owner_of_vehicle', 'Service_year_of_vehicle', 'Defect_of_vehicle', 'Area accident occured', 'Lanes or Medians', 'Road allignment', 'Types of Junction', 'Road surface type', 'Road surface condition
```

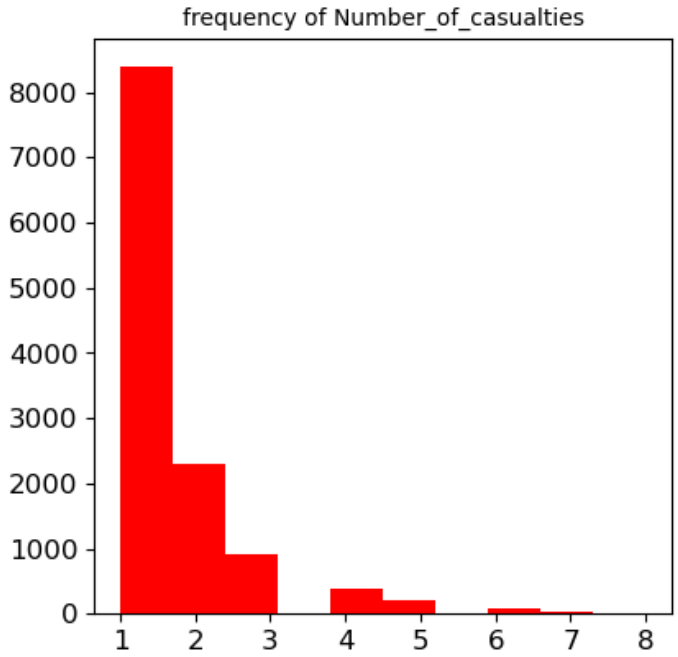
jupyter Accident_Severity Last Checkpoint: 17 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

```
        plt.hist(df[i],color='red')
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title('frequency of '+i, fontsize=10)
    plotnumber +=1
```





8:44 AM 10/27/2023

Data science ass2.pdf | Untitled Folder/ | Accident_Severity - Jupyter Note

localhost:8888/notebooks/Untitled%20Folder/Accident_Severity.ipynb

jupyter Accident_Severity Last Checkpoint: 17 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [17]:
```
#storing categorical column names to a new variable
categorical=[i for i in df.columns if df[i].dtype=='O']
print('The categorical variables are',categorical)
```

```
The categorical variables are ['Time', 'Day_of_week', 'Age_band_of_driver', 'Sex_of_driver', 'Educational_level', 'Vehicle_driver_relation', 'Driving_experience', 'Type_of_vehicle', 'Owner_of_vehicle', 'Service_year_of_vehicle', 'Defect_of_vehicle', 'Area_accident_occured', 'Lanes_or_Medians', 'Road_allignment', 'Types_of_Junction', 'Road_surface_type', 'Road_surface_conditions', 'Light_conditions', 'Weather_conditions', 'Type_of_collision', 'Vehicle_movement', 'Casualty_class', 'Sex_of_casualty', 'Age_band_of_casualty', 'Casualty_severity', 'Work_of_casuality', 'Fitness_of_casuality', 'Pedestrian_movement', 'Cause_of_accident', 'Accident_severity']
```

In [18]:
```
#count plot for categorical values
plt.figure(figsize=(10,200))
plotnumber = 1

for col in categorical:
    if plotnumber <= df.shape[1] and col!='Pedestrian_movement':
        ax1 = plt.subplot(28,1,plotnumber)
        sns.countplot(data=df, y=col, palette='muted')
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In[18], line 7
      5 for col in categorical:
```

8:44 AM 10/27/2023

Data science ass2.pdf | Untitled Folder/ | Accident_Severity - Jupyter Notebook

localhost:8888/notebooks/Untitled%20Folder/Accident_Severity.ipynb

jupyter Accident_Severity Last Checkpoint: 17 minutes ago (autosaved)

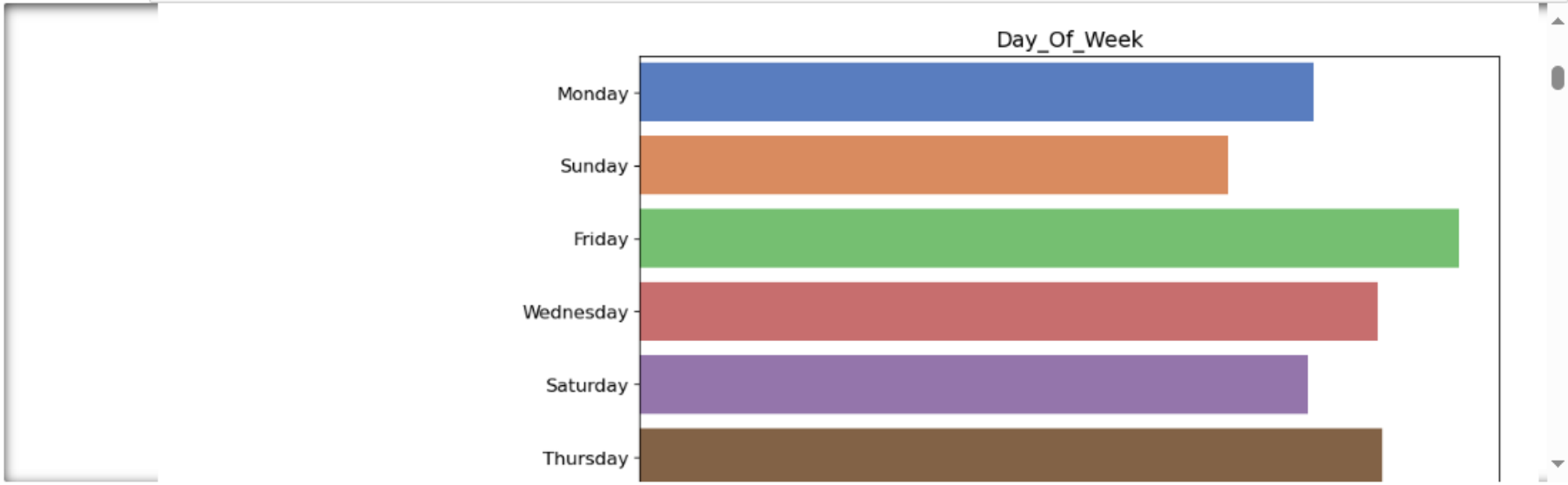
Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

```
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```



In [ ]:

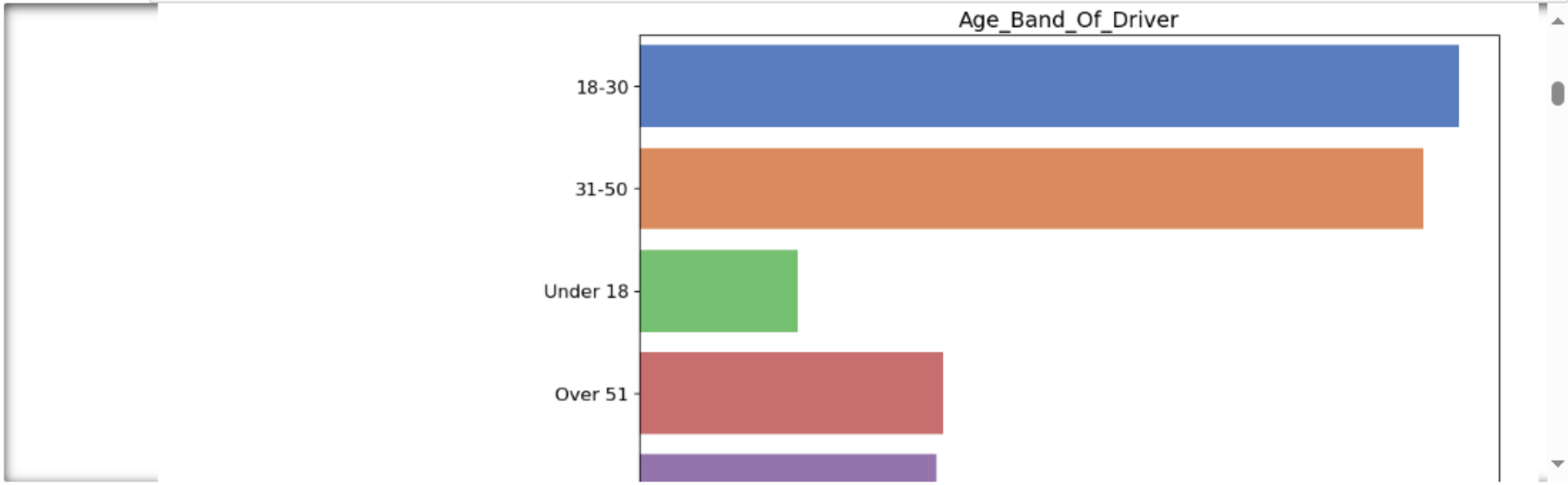
8:44 AM
10/27/2023

jupyter Accident_Severity Last Checkpoint: 17 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

```
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```



In [ ]:

Data science ass2.pdf | Untitled Folder/ | Accident_Severity - Jupyter Notebook

localhost:8888/notebooks/Untitled%20Folder/Accident_Severity.ipynb

jupyter Accident_Severity Last Checkpoint: 17 minutes ago (autosaved)

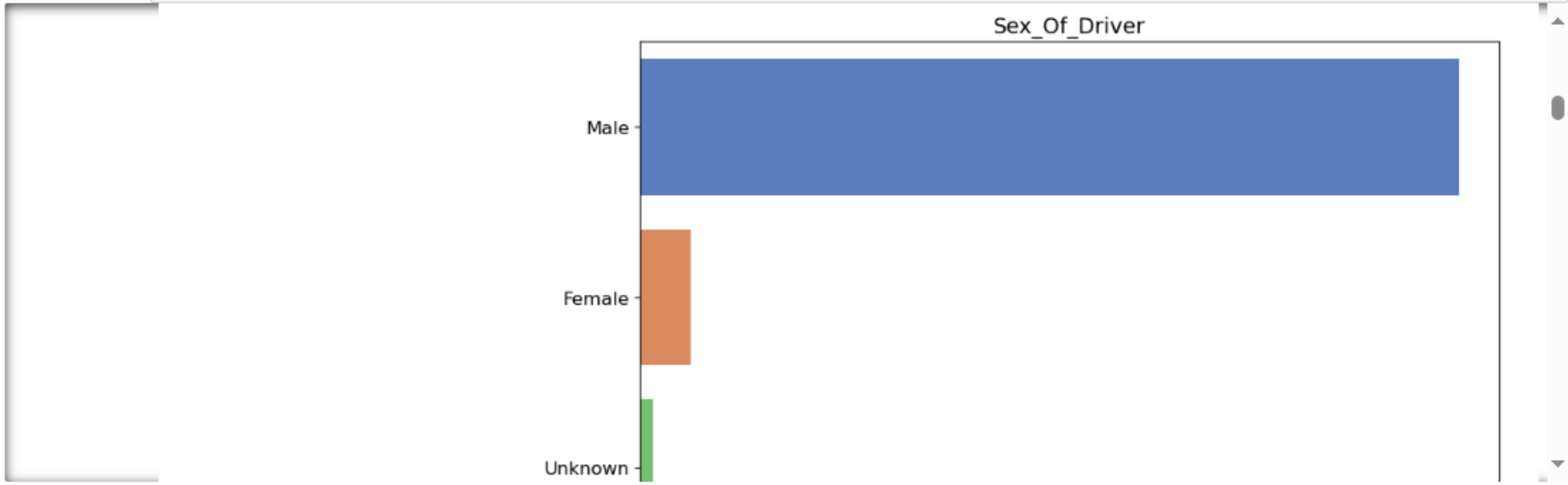
Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

```
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```



In [ ]:

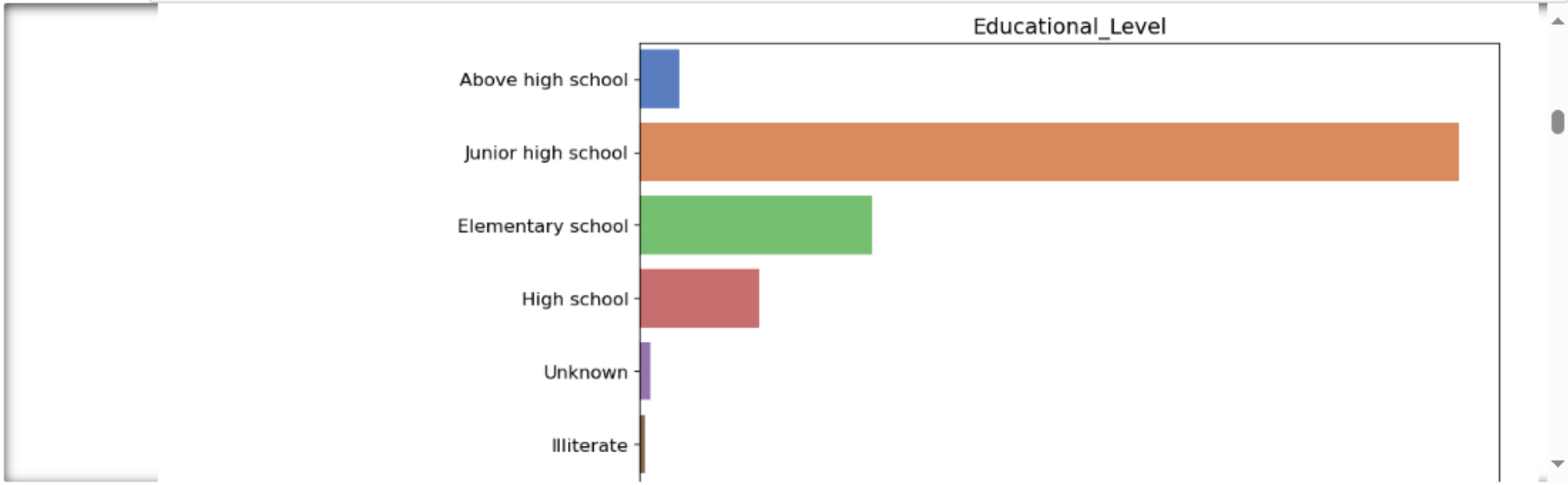
8:44 AM
10/27/2023

jupyter Accident_Severity Last Checkpoint: 17 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

```
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```



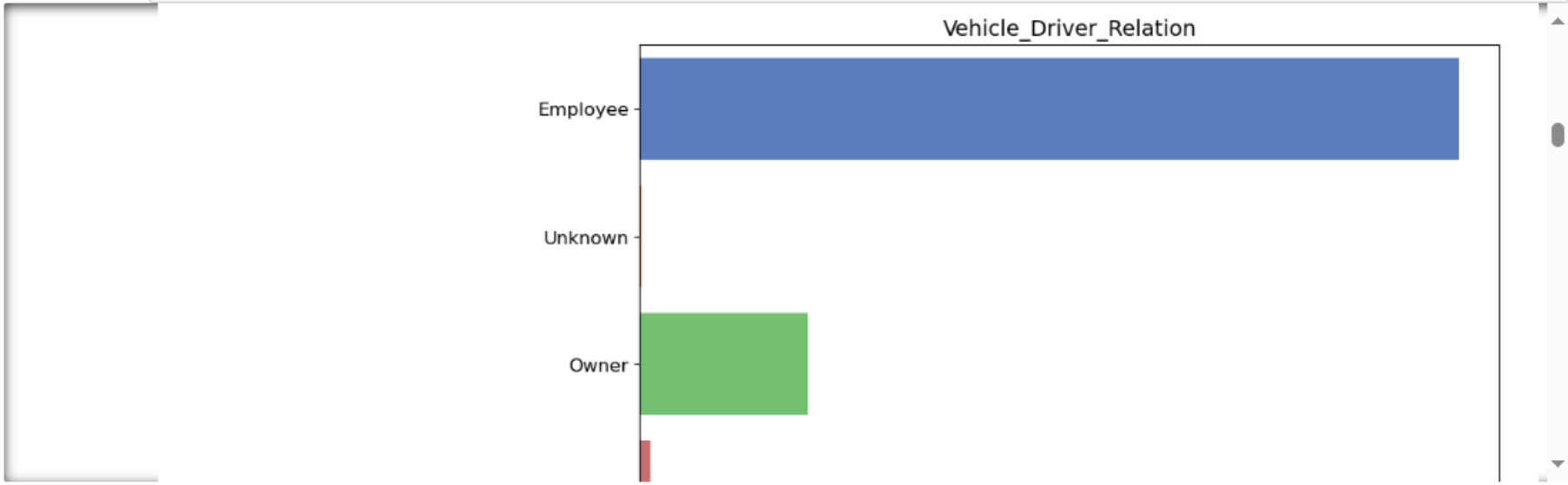
In [ ]:

jupyter Accident_Severity Last Checkpoint: 17 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

```
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```



In [ ]:

Data science ass2.pdf | Untitled Folder/ | Accident_Severity - Jupyter Notebook

localhost:8888/notebooks/Untitled%20Folder/Accident_Severity.ipynb

jupyter Accident_Severity Last Checkpoint: 17 minutes ago (autosaved)

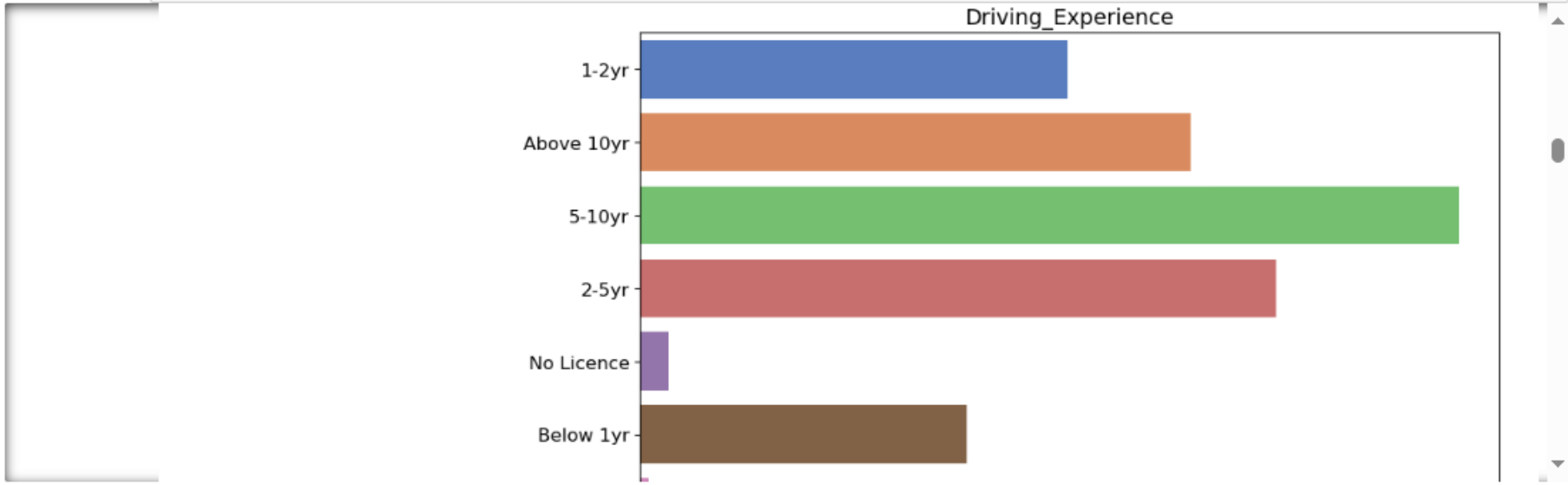
Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

```
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```



In [ ]:

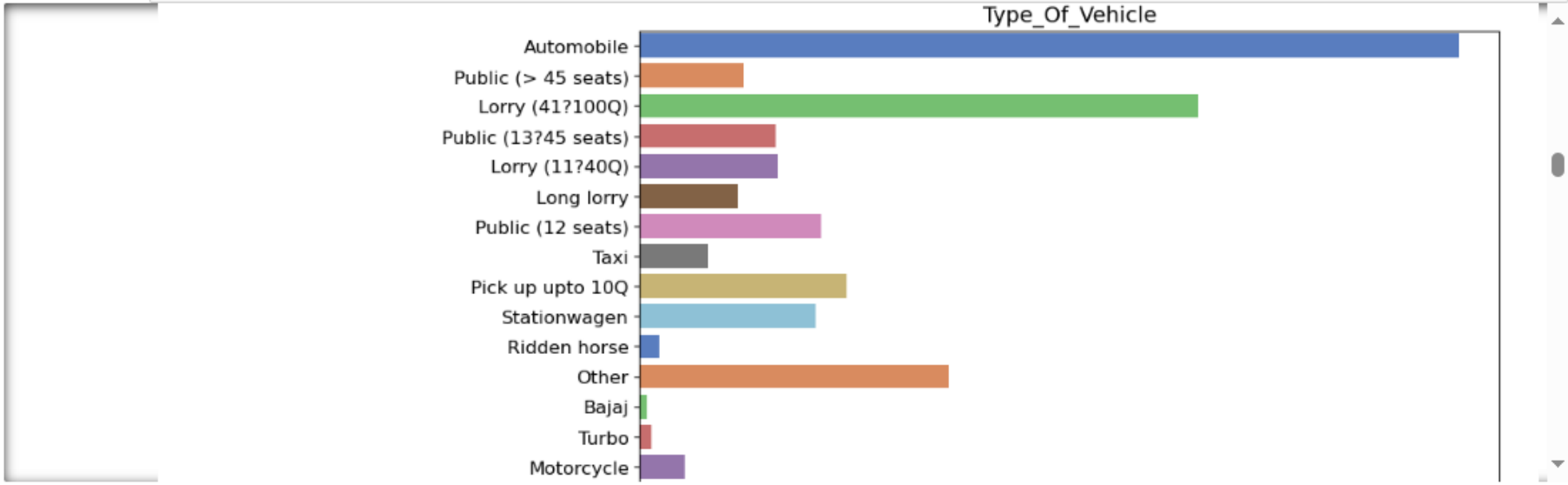
8:45 AM
10/27/2023

jupyter Accident_Severity Last Checkpoint: 17 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

```
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```



In [ ]:

8:45 AM 10/27/2023

Data science ass2.pdf | Untitled Folder/ | Accident_Severity - Jupyter Notebook

localhost:8888/notebooks/Untitled%20Folder/Accident_Severity.ipynb

jupyter Accident_Severity Last Checkpoint: 17 minutes ago (autosaved)

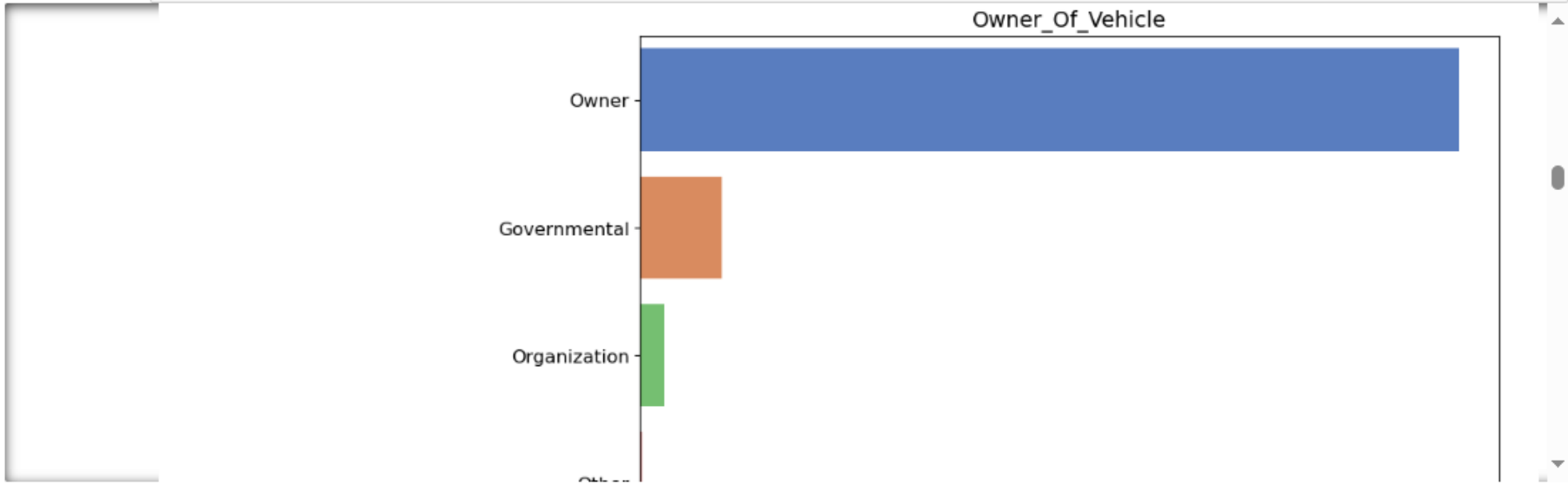
Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

```
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```



In [ ]:

8:45 AM
10/27/2023

Data science ass2.pdf | Untitled Folder/ | Accident_Severity - Jupyter Notebook

localhost:8888/notebooks/Untitled%20Folder/Accident_Severity.ipynb

jupyter Accident_Severity Last Checkpoint: 18 minutes ago (autosaved)

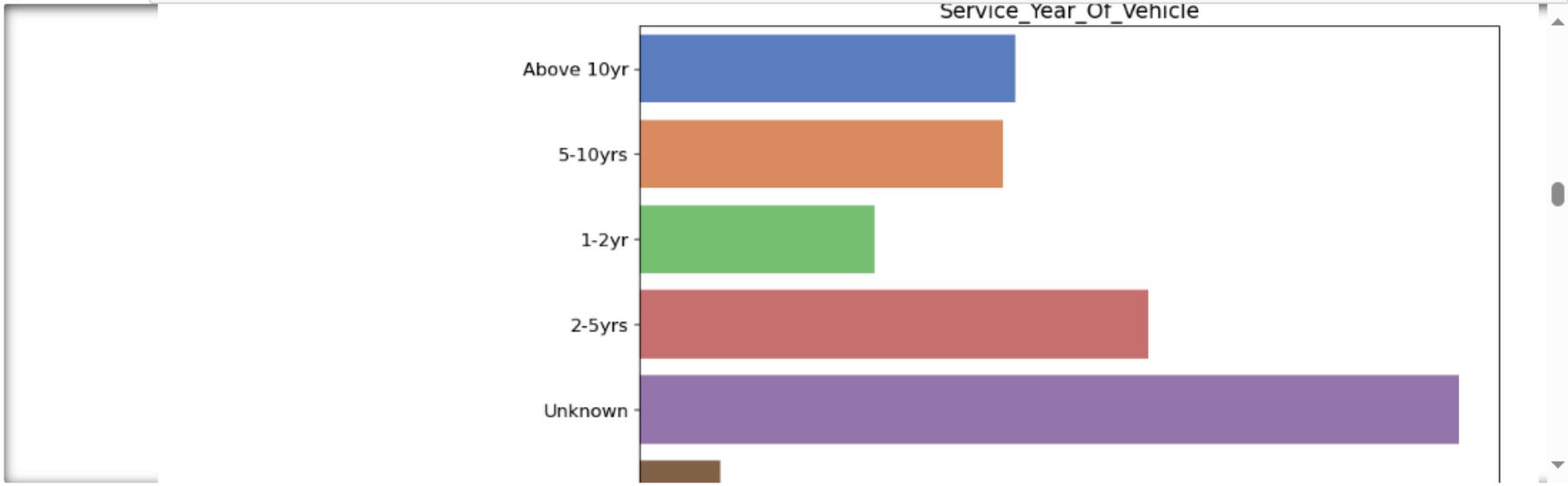
Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

```
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```



In [ ]:

8:45 AM
10/27/2023

Data science ass2.pdf | Untitled Folder/ | Accident_Severity - Jupyter Note

localhost:8888/notebooks/Untitled%20Folder/Accident_Severity.ipynb

jupyter Accident_Severity Last Checkpoint: 18 minutes ago (autosaved)

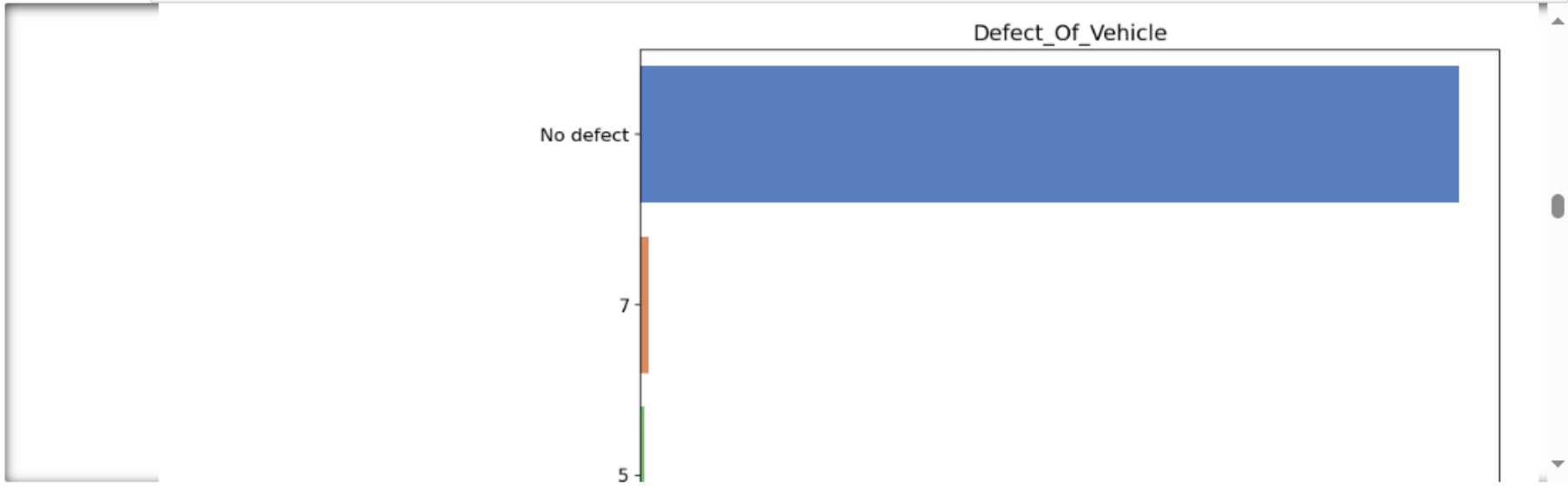
Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

```
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```



In [ ]:

8:45 AM
10/27/2023