

```
In [45]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [46]: df4=pd.read_csv("C:/Users/sujit/Downloads/accident_data.csv")
```

```
In [47]: df4
```

Date		Date	Time	Latitude	Longitude	Location Description	Weather Conditions	Road Conditions	Visibility	Accident Type	Injury Severity	Contributing Factors	Vehicle Type	No of Vehicles	Vehicle Speed	Day of Week	Month	Year
0	05-01-2022	15:30:00	40.7128	-74.0060	Highway	Foggy	Wet	Poor	Collision	Moderate	Speeding	Truck	1	45	Wednesday	January	2022	
1	10-02-2022	18:45:00	41.8781	-87.6298	Street	Rainy	Wet	Poor	Rollover	Severe	Drunk Driving	Motorcycle	1	40	Thursday	February	2022	
2	15-03-2022	12:00:00	37.7749	-122.4194	Intersection	Clear	Dry	Clear	Collision	Moderate	Weather	Car	2	25	Tuesday	March	2022	
3	20-04-2022	09:20:00	34.0522	-118.2437	Highway	Clear	Dry	Clear	Collision	Minor	Other	Car	2	35	Wednesday	April	2022	
4	25-05-2022	17:10:00	40.7128	-74.0060	Street	Rainy	Wet	Poor	Collision	Severe	Speeding	Car	1	50	Wednesday	May	2022	
142	10-08-2024	12:45:00	40.7128	-74.0060	Street	Rainy	Wet	Poor	Rollover	Severe	Drunk Driving	Motorcycle	1	30	Wednesday	August	2024	
143	15-09-2024	08:30:00	37.7749	-122.4194	Highway	Clear	Dry	Clear	Collision	Moderate	Weather	Car	2	35	Thursday	September	2024	
144	20-10-2024	16:20:00	34.0522	-118.2437	Street	Clear	Dry	Clear	Collision	Minor	Other	Car	2	45	Wednesday	October	2024	
145	25-11-2024	14:10:00	40.7128	-74.0060	Intersection	Rainy	Wet	Poor	Collision	Severe	Speeding	Car	1	55	Thursday	November	2024	
146	30-12-2024	10:00:00	41.8781	-87.6298	Highway	Clear	Dry	Clear	Rollover	Minor	Distracted Driving	Truck	1	30	Monday	December	2024	

147 rows × 17 columns

1= [20]:  
dfA.head(10)

2= [20]:

	Date	Time	Latitude	Longitude	Location Description	Weather Conditions	Road Conditions	Visibility	Accident Type	Injury Severity	Contributing Factors	Vehicle Type	No of Vehicles	Vehicle Speed	Day of Week	Month	Year
0	05-01-2022	15:30:00	-43.7128	-74.0060	Highway	Foggy	Wet	Poor	Collision	Moderate	Speeding	Truck	1	45	Wednesday	January	2022
1	10-02-2022	18:45:00	-41.8781	-87.6298	Street	Rainy	Wet	Poor	Rollover	Severe	Drunk Driving	Motorcycle	1	40	Thursday	February	2022
2	15-03-2022	12:00:00	37.7749	-122.4194	Intersection	Clear	Dry	Clear	Collision	Moderate	Weather	Car	2	25	Tuesday	March	2022
3	20-04-2022	09:20:00	34.0522	-118.2437	Highway	Clear	Dry	Clear	Collision	Minor	Other	Car	2	35	Wednesday	April	2022
4	25-05-2022	17:10:00	-43.7128	-74.0060	Street	Rainy	Wet	Poor	Collision	Severe	Speeding	Car	1	50	Wednesday	May	2022
5	30-06-2022	14:30:00	-41.8781	-87.6298	Intersection	Clear	Dry	Clear	Rollover	Minor	Distracted Driving	Truck	1	35	Thursday	June	2022
6	05-07-2022	20:15:00	34.0522	-118.2437	Street	Clear	Dry	Clear	Collision	Moderate	Speeding	Car	2	40	Monday	July	2022
7	10-08-2022	12:45:00	-43.7128	-74.0060	Highway	Rainy	Wet	Poor	Rollover	Severe	Drunk Driving	Motorcycle	1	30	Wednesday	August	2022
8	15-09-2022	08:30:00	37.7749	-122.4194	Intersection	Clear	Dry	Clear	Collision	Moderate	Weather	Car	2	35	Thursday	September	2022
9	20-10-2022	16:20:00	34.0522	-118.2437	Highway	Clear	Dry	Clear	Collision	Minor	Other	Car	2	45	Wednesday	October	2022

```
In [71]: df4.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 147 entries, 0 to 146
```

```
Data columns (total 17 columns):
```

#	Column	Non-Null Count	Dtype
0	Date	147 non-null	object
1	Time	147 non-null	object
2	Latitude	147 non-null	float64
3	Longitude	147 non-null	float64
4	Location Description	147 non-null	object
5	Weather Conditions	147 non-null	object
6	Road Conditions	147 non-null	object
7	Visibility	147 non-null	object
8	Accident Type	147 non-null	object
9	Injury Severity	147 non-null	object
10	Contributing Factors	147 non-null	object
11	Vehicle Type	147 non-null	object
12	No of Vehichles	147 non-null	int64
13	Vehicle Speed	147 non-null	int64
14	Day of Week	147 non-null	object
15	Month	147 non-null	object
16	Year	147 non-null	int64

```
dtypes: float64(2), int64(3), object(12)
```

```
memory usage: 19.7+ KB
```

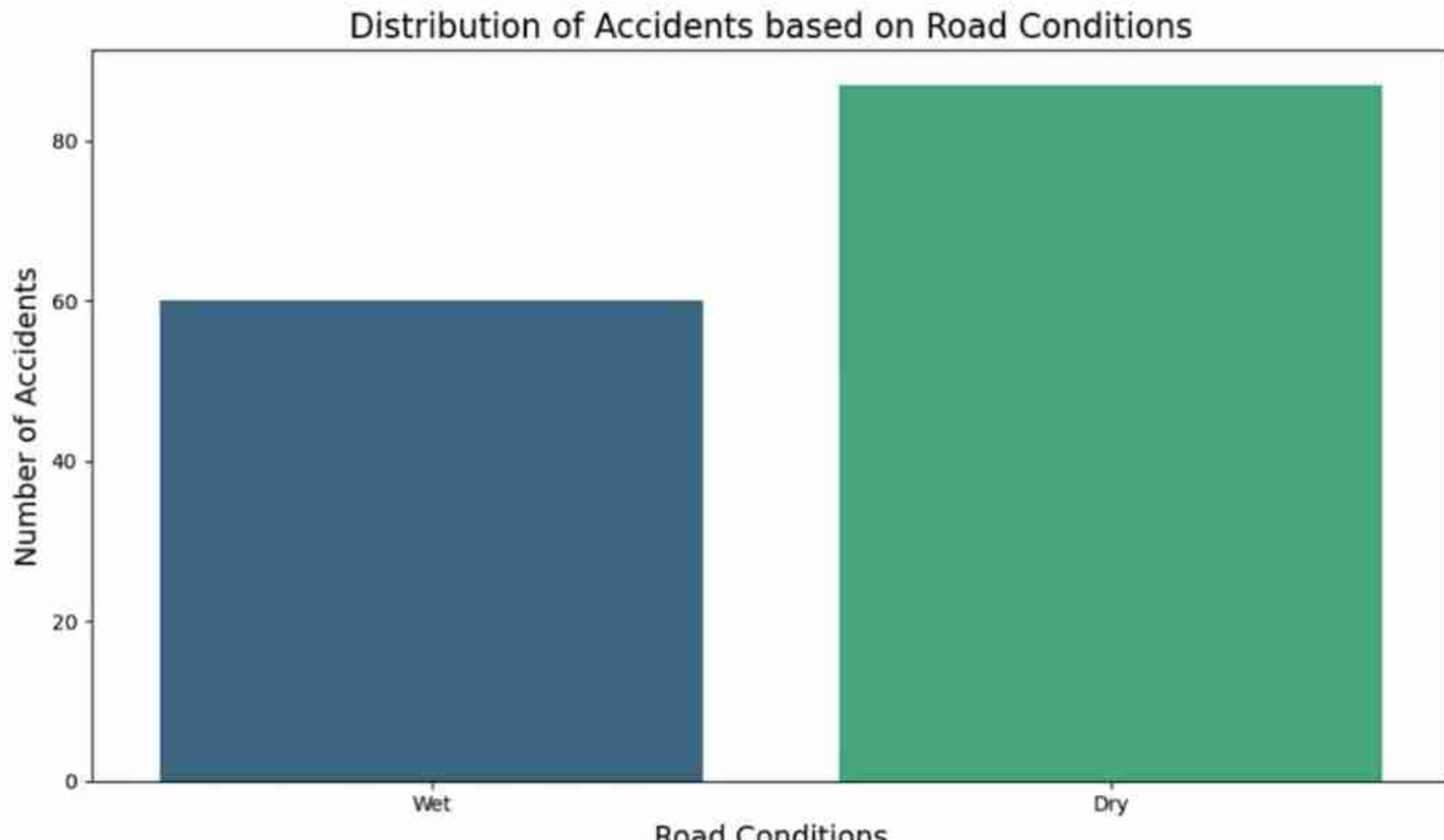
```
In [73]: df4.isnull().sum()
```

```
Out[73]: Date                0  
Time                0  
Latitude            0  
Longitude           0  
Location Description 0  
Weather Conditions  0  
Road Conditions     0  
Visibility          0  
Accident Type       0  
Injury Severity     0  
Contributing Factors 0  
Vehicle Type        0  
No of Vehichles     0  
Vehicle Speed       0  
Day of Week         0  
Month               0  
Year               0  
dtype: int64
```

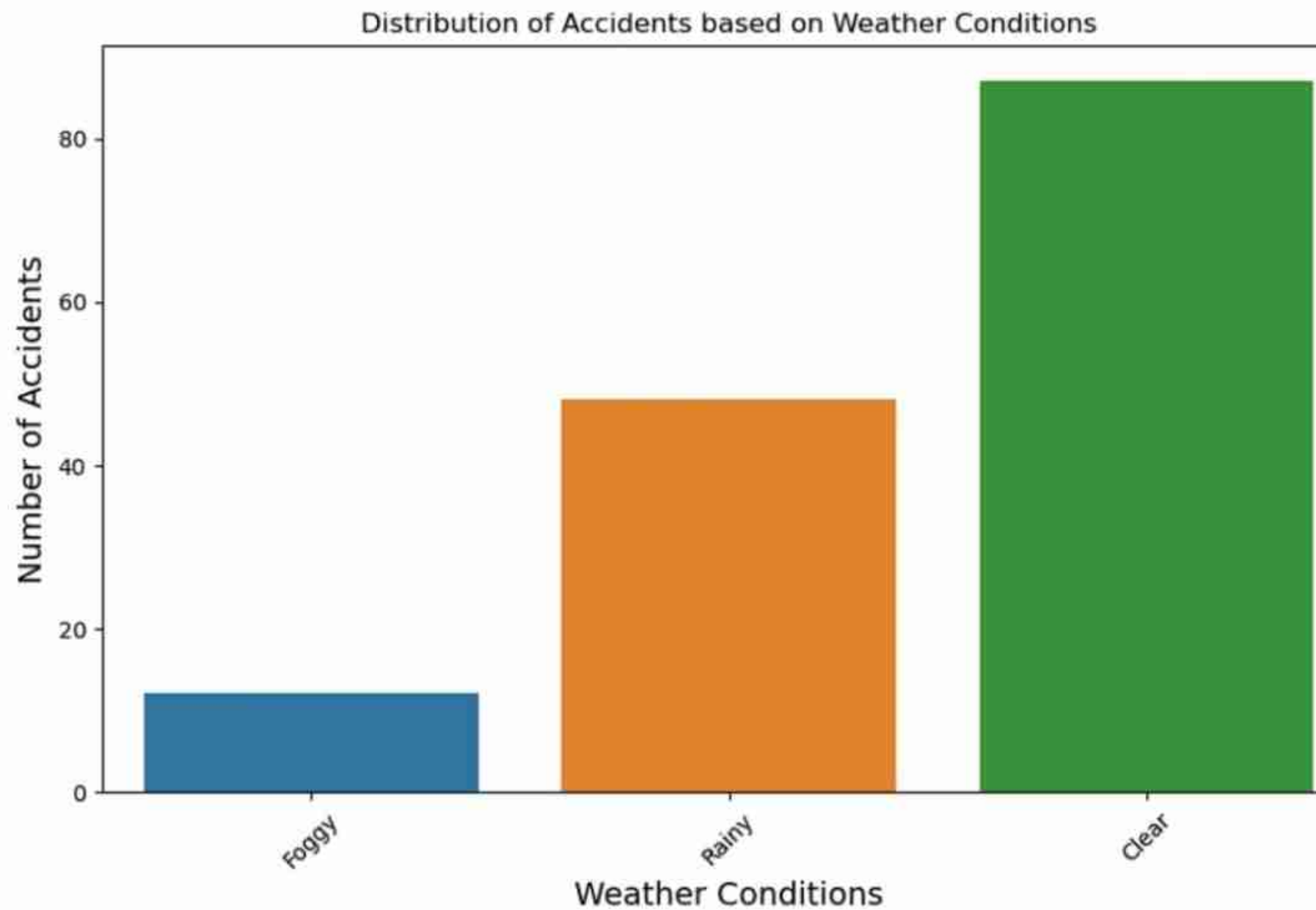
```
In [77]: df4['DateTime'] = pd.to_datetime(df4['Date'] + ' ' + df4['Time'], format='%d-%m-%Y %H:%M:%S')  
  
df4.set_index('DateTime', inplace=True)
```

```
2]: # Distribution of accidents based on road conditions
plt.figure(figsize=(10, 6))
sns.countplot(x='Road Conditions', data=df4, palette='viridis')
plt.title('Distribution of Accidents based on Road Conditions', fontsize=16)
plt.xlabel('Road Conditions', fontsize=14)
plt.ylabel('Number of Accidents', fontsize=14)

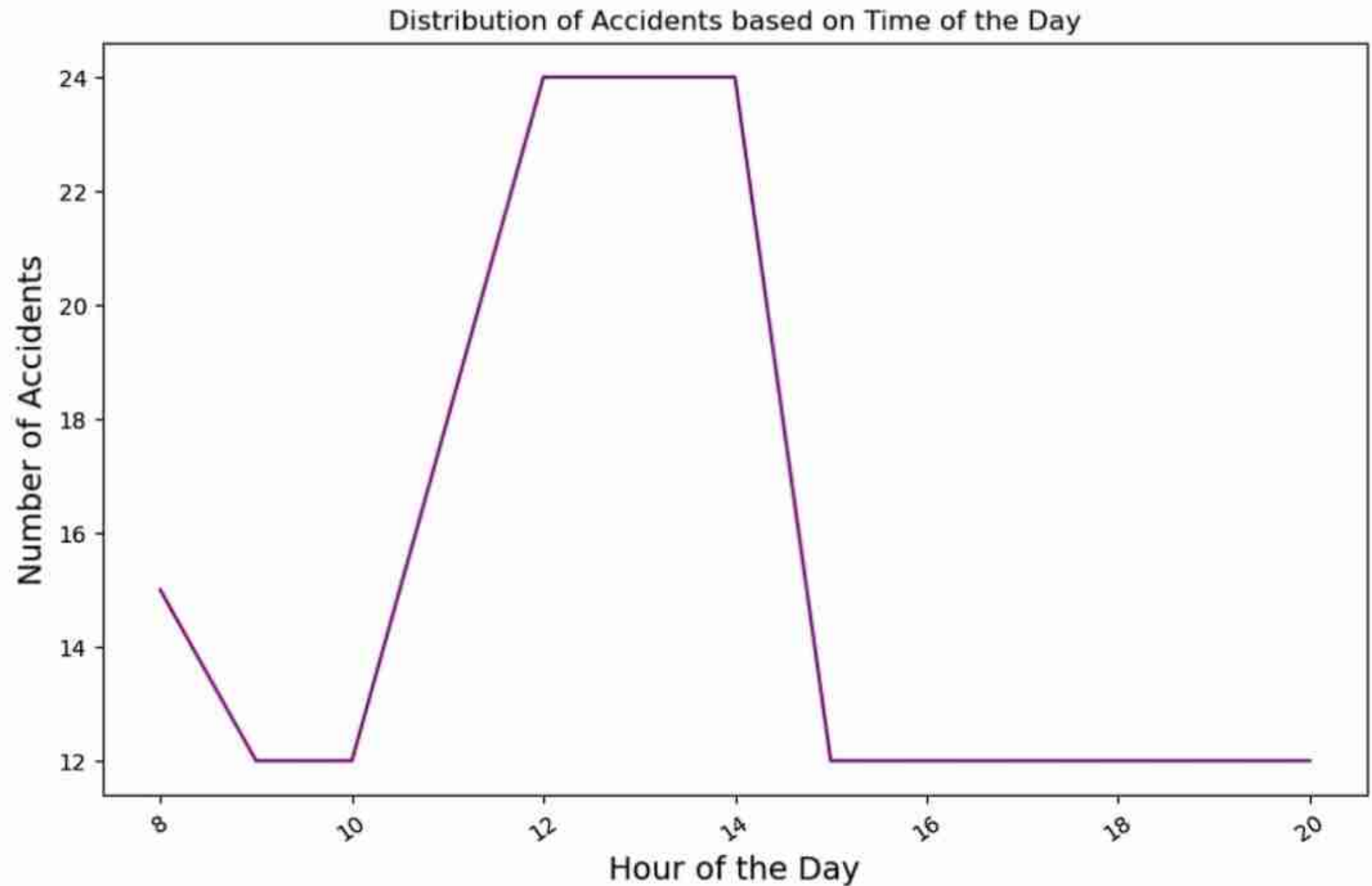
plt.tight_layout()
plt.show()
```



```
# Distribution of accidents based on weather conditions
plt.figure(figsize=(10, 6))
sns.countplot(x='Weather Conditions', data=df4)
plt.title('Distribution of Accidents based on Weather Conditions')
plt.xlabel('Weather Conditions', fontsize=14)
plt.ylabel('Number of Accidents', fontsize=14)
plt.xticks(rotation=45)
plt.show()
```

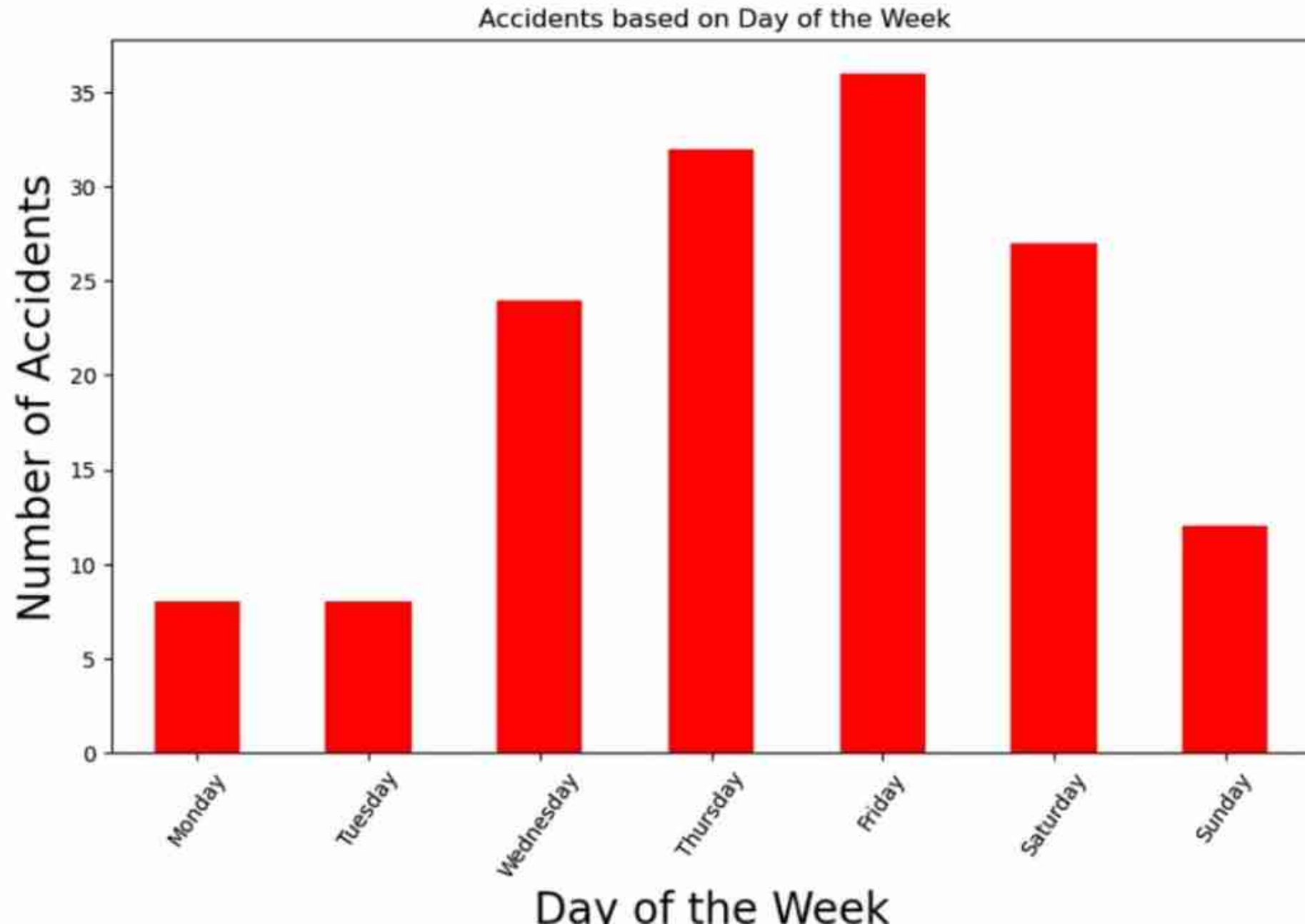


```
plt.figure(figsize=(10, 6))
df4.index.hour.value_counts().sort_index().plot(kind='line',color='purple')
plt.title('Distribution of Accidents based on Time of the Day')
plt.xlabel('Hour of the Day' , fontsize=14)
plt.ylabel('Number of Accidents' , fontsize=14)
plt.xticks(rotation=35)
plt.show()
```



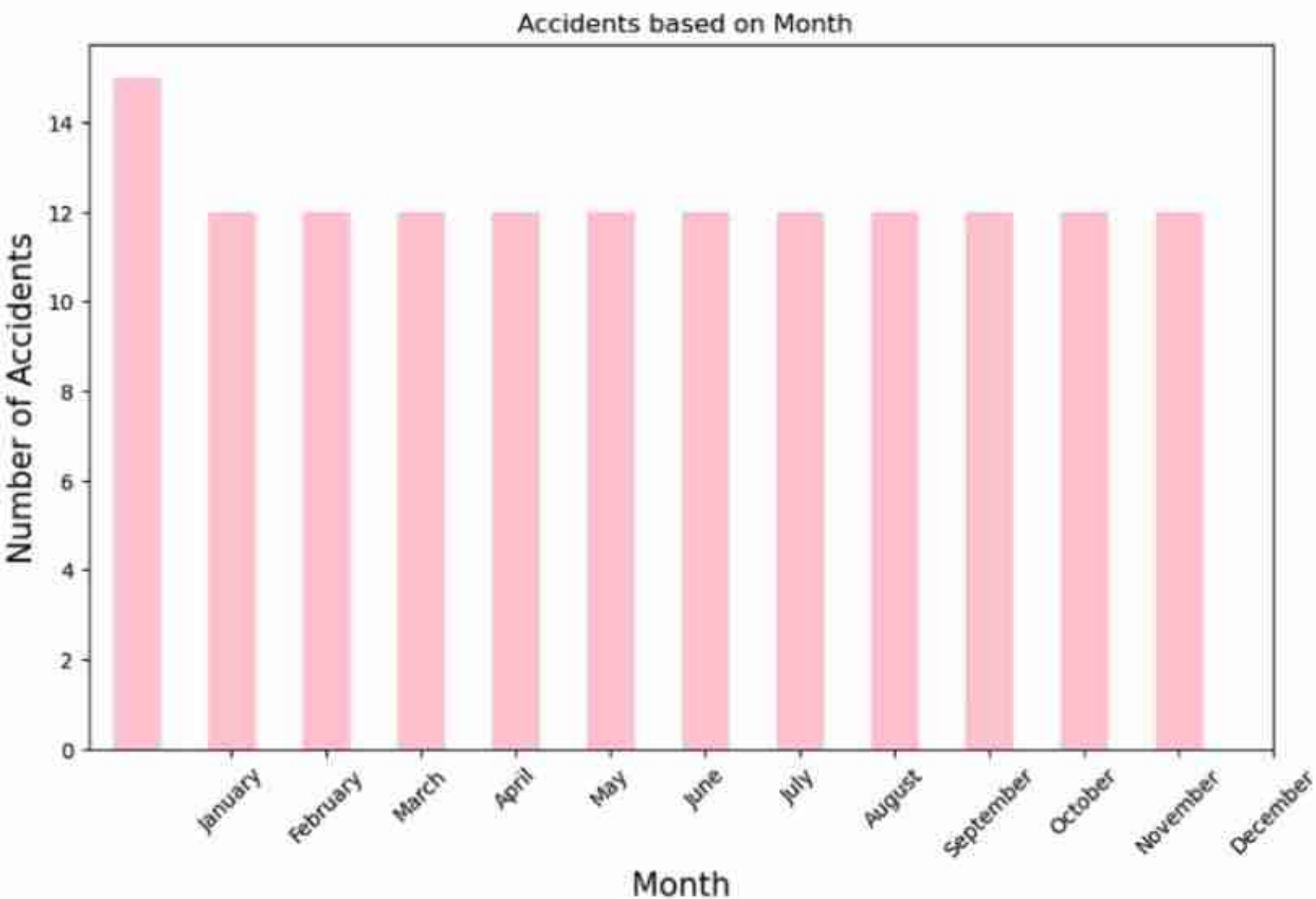


```
plt.figure(figsize=(10, 6))
df4.index.dayofweek.value_counts().sort_index().plot(kind='bar',color = "red")
plt.title(' Accidents based on Day of the Week')
plt.xlabel('Day of the Week' , fontsize=20)
plt.ylabel('Number of Accidents' , fontsize=20)
plt.xticks(range(7), ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday'], rotation=55)
plt.show()
```

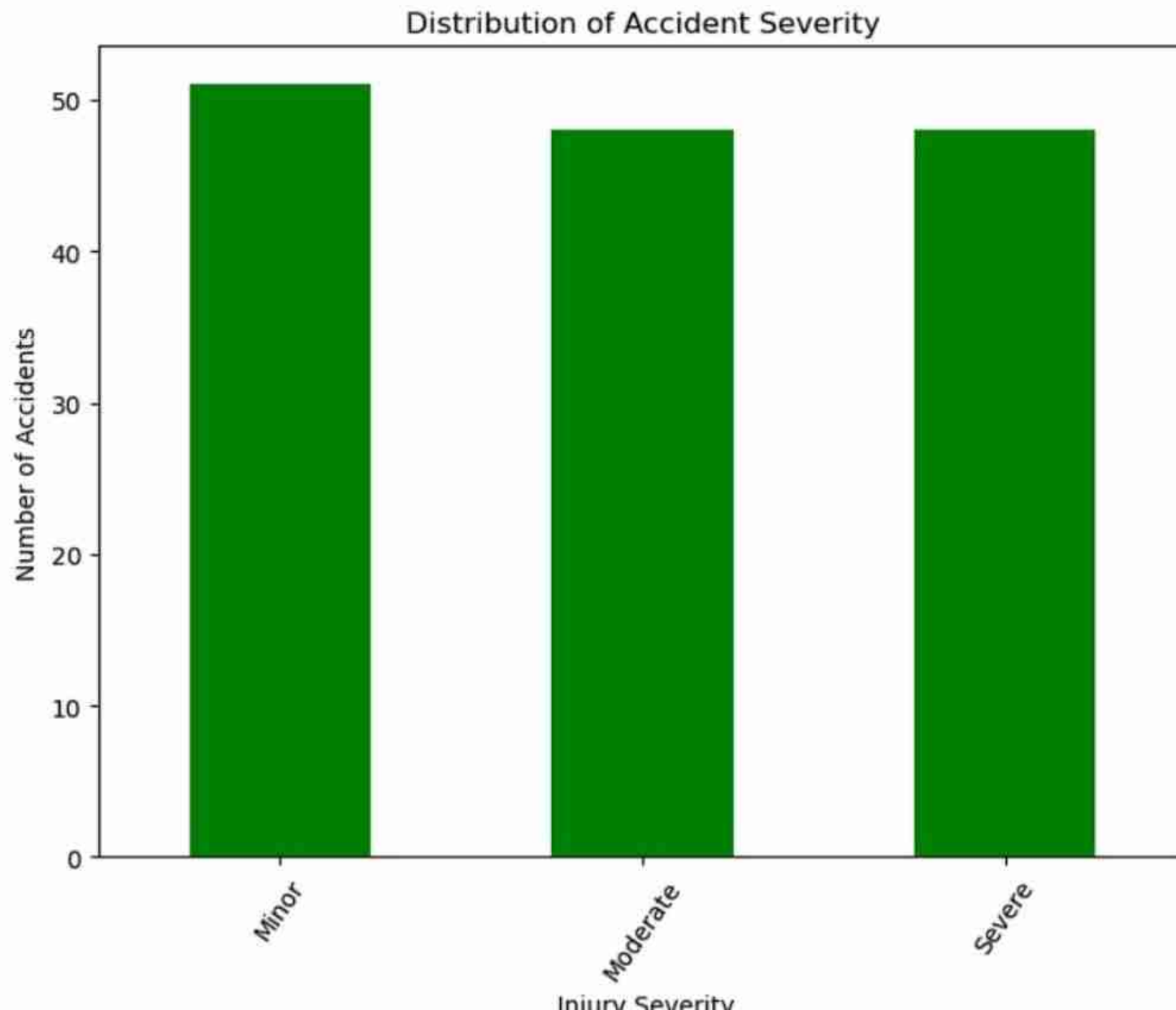




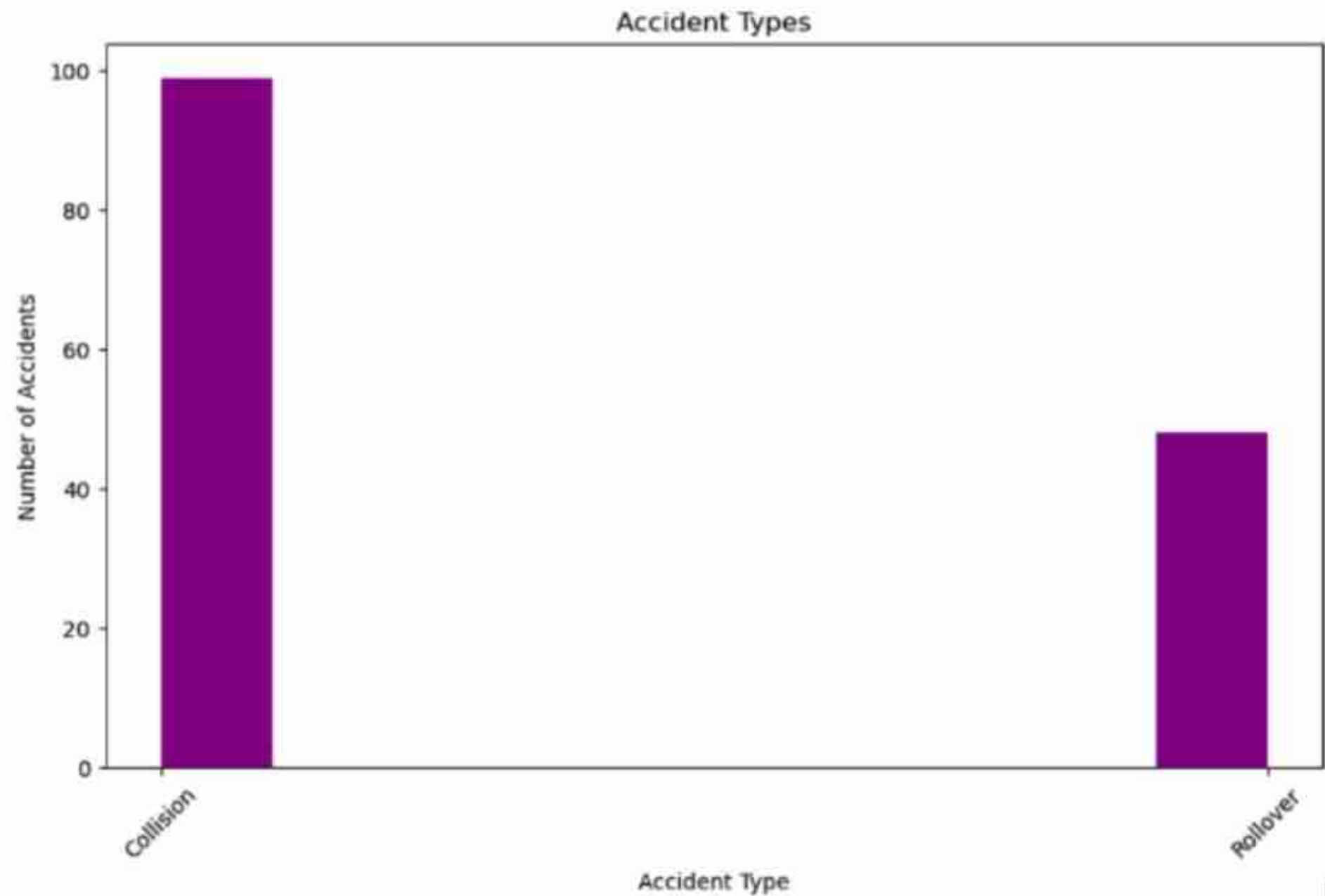
```
# Distribution of accidents based on month
plt.figure(figsize=(10, 6))
df4.index.month.value_counts().sort_index().plot(kind='bar',color="pink")
plt.title(' Accidents based on Month')
plt.xlabel('Month' , fontsize=15)
plt.ylabel('Number of Accidents',fontsize=15)
plt.xticks(range(1, 13), ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December'], rotations=45)
plt.show()
```



```
# Distribution of Accident Severity
plt.figure(figsize=(8, 6))
df4['Injury Severity'].value_counts().plot(kind='bar', color='green')
plt.title('Distribution of Accident Severity')
plt.xlabel(' Injury Severity', fontsize=10)
plt.ylabel('Number of Accidents', fontsize=10)
plt.xticks(rotation=55)
plt.show()
```



```
plt.figure(figsize=(10, 6))
plt.hist(df4['Accident Type'], bins=10, color='purple')
plt.title('Accident Types')
plt.xlabel('Accident Type')
plt.ylabel('Number of Accidents')
plt.xticks(rotation=55)
plt.show()
```



```
# Plot accidents over the years (curve/line plot)
plt.figure(figsize=(10, 6))
accidents_by_year.plot(color='skyblue', marker='o', linestyle=' ')
plt.title('Number of Accidents Over the Years')
plt.xlabel('Year')
plt.ylabel('Number of Accidents')
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```

