

In [1]:

```
1 # Importing necessary Libraries
2
3 import numpy as np
4 import pandas as pd
5 import plotly.express as px
6 import matplotlib.pyplot as plt
7 from matplotlib.ticker import NullFormatter
8 import matplotlib.ticker as ticker
9 import seaborn as sns
10 from sklearn import preprocessing
11 %matplotlib inline
```

In [2]:

```
1 # Importing Dataset
2
3 df = pd.read_csv("Data-Collisions.csv")
4 df.head()
```

G:\Python\Anaconda\lib\site-packages\IPython\core\interactiveshell.py:3058:
DtypeWarning: Columns (33) have mixed types. Specify dtype option on import
or set low_memory=False.
interactivity=interactivity, compiler=compiler, result=result)

Out[2]:

	SEVERITYCODE		X	Y	OBJECTID	INCKEY	COLDKEY	REPORTNO	ST
0	2	-122.323148	47.703140		1	1307	1307	3502005	Ma
1	1	-122.347294	47.647172		2	52200	52200	2607959	Ma
2	1	-122.334540	47.607871		3	26700	26700	1482393	Ma
3	1	-122.334803	47.604803		4	1144	1144	3503937	Ma
4	2	-122.306426	47.545739		5	17700	17700	1807429	Ma

5 rows × 38 columns

Data Preprocessing

In [3]:

```
1 # Checking all columns to remove irrelevant ones
2 df.isnull().sum()
```

Out[3]:

```
SEVERITYCODE      0
X                 5334
Y                 5334
OBJECTID          0
INCKEY            0
COLDETKEY         0
REPORTNO          0
STATUS            0
ADDRTYPE         1926
INTKEY           129603
LOCATION           2677
EXCEPTRSNCODE     109862
EXCEPTRSNDESC     189035
SEVERITYCODE.1    0
SEVERITYDESC      0
COLLISIONTYPE     4904
PERSONCOUNT      0
PEDCOUNT         0
PEDCYLCOUNT       0
VEHCOUNT          0
INCDATE           0
INCDTTM           0
JUNCTIONTYPE      6329
SDOT_COLCODE      0
SDOT_COLDESC      0
INATTENTIONIND    164868
UNDERINFL         4884
WEATHER           5081
ROADCOND          5012
LIGHTCOND         5170
PEDROWNOTGRNT     190006
SDOTCOLNUM        79737
SPEEDING          185340
ST_COLCODE        18
ST_COLDESC        4904
SEGLANEKEY        0
CROSSWALKKEY      0
HITPARKEDCAR      0
dtype: int64
```

In [4]:

```
1 # Dropping non-required columns
2 df.drop(["OBJECTID", "REPORTNO", "STATUS", "PEDROWNOTGRNT", "SDOTCOLNUM", "INATTENTIONIND"])
3
4 # Dropping all null values
5 df.dropna(axis = 0, inplace = True)
6
7 df.isnull().sum()
```

Out[4]:

```
SEVERITYCODE      0
X                 0
Y                 0
INCKEY            0
COLDETKEY         0
ADDRTYPE          0
INTKEY            0
LOCATION           0
EXCEPTRSNCODE     0
EXCEPTRSNDESC     0
SEVERITYCODE.1    0
SEVERITYDESC      0
COLLISIONTYPE     0
PERSONCOUNT      0
PEDCOUNT         0
PEDCYLCOUNT       0
VEHCOUNT          0
INCDATE           0
INCDTTM           0
JUNCTIONTYPE      0
SDOT_COLCODE      0
SDOT_COLDESC      0
UNDERINFL         0
WEATHER           0
ROADCOND          0
LIGHTCOND         0
SPEEDING          0
ST_COLCODE        0
ST_COLDESC        0
SEGLANEKEY        0
CROSSWALKKEY      0
HITPARKEDCAR      0
dtype: int64
```

In [5]:

```
1 df.head()
```

Out[5]:

	SEVERITYCODE		X	Y	INCKEY	COLDETKEY	ADDRTYPE	INTKEY	
147780	2	-122.334204	47.573700	205778	205938	Intersection	32334.0		
150966	2	-122.290875	47.704785	211254	211494	Intersection	36855.0		
151135	1	-122.344651	47.701430	213455	213715	Intersection	37301.0		

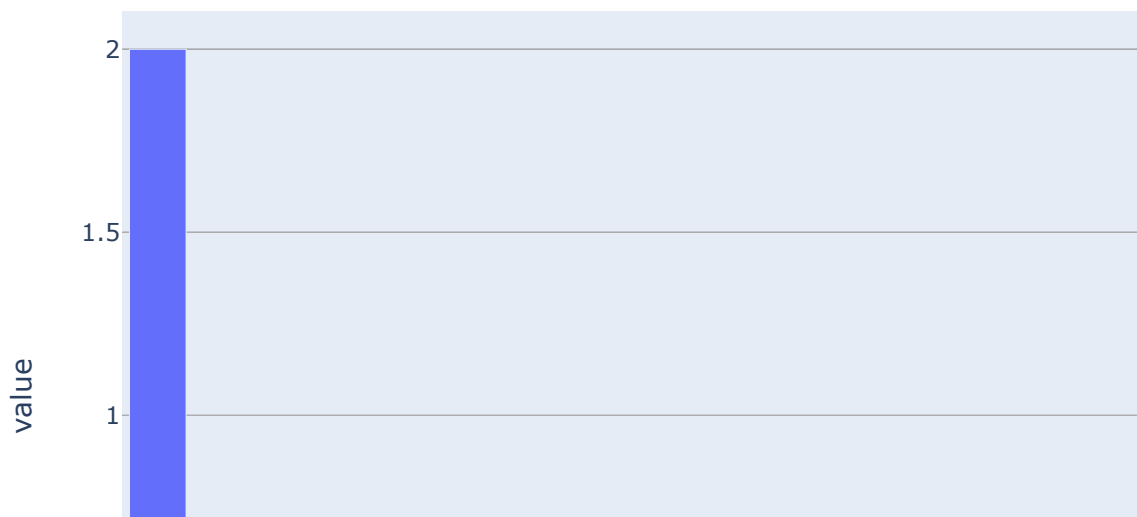
3 rows × 32 columns



Data Visualization

In [6]:

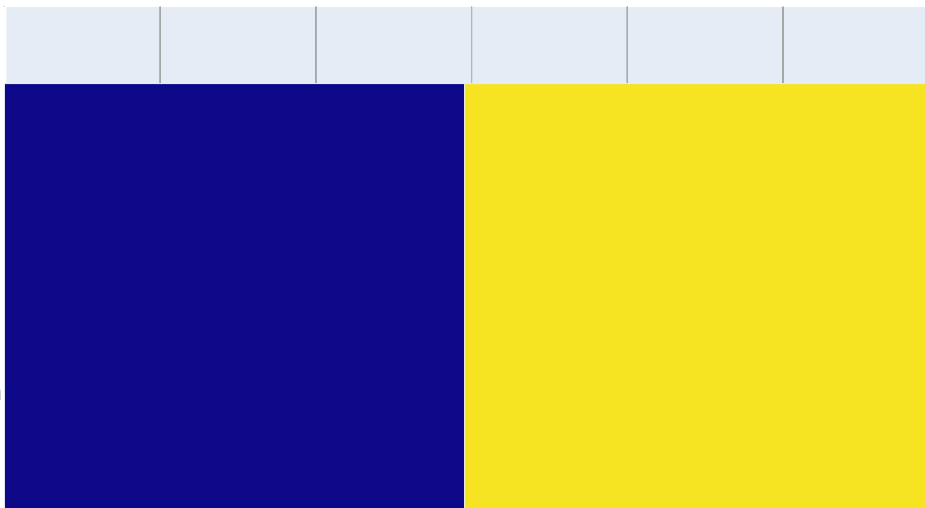
```
1 a = df["SEVERITYCODE"].value_counts()
2 x = a.index
3 y = a.values
4 fig = px.bar(df["SEVERITYCODE"])
5 fig.show()
```



In [7]:

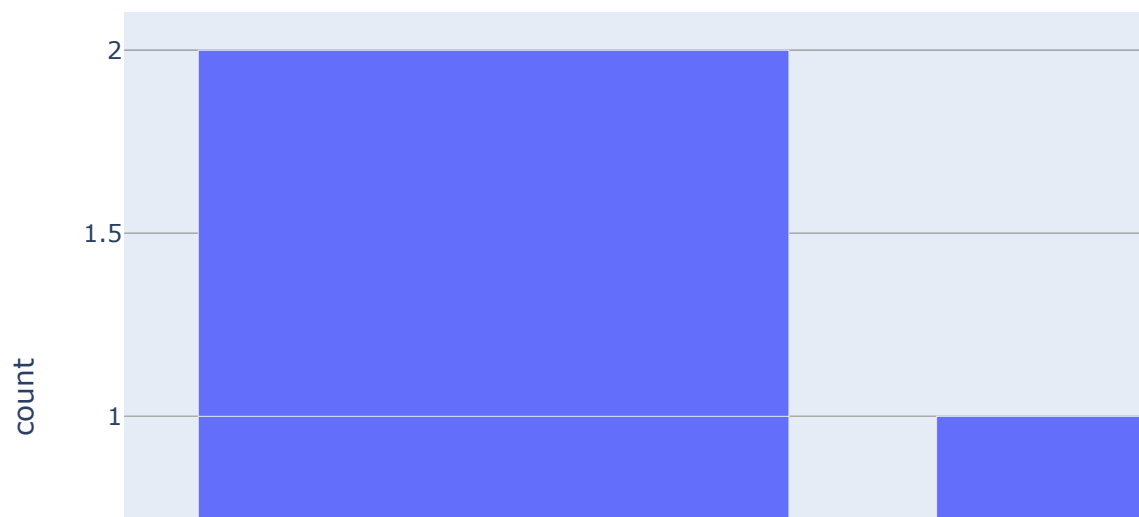
```
1 a = df["ADDRTYPE"]
2 xx = a.index
3 yy = a.values
4 fig = px.bar(a, x = xx, y = yy, color = xx)
5 fig.show()
```

> Intersection



In [8]:

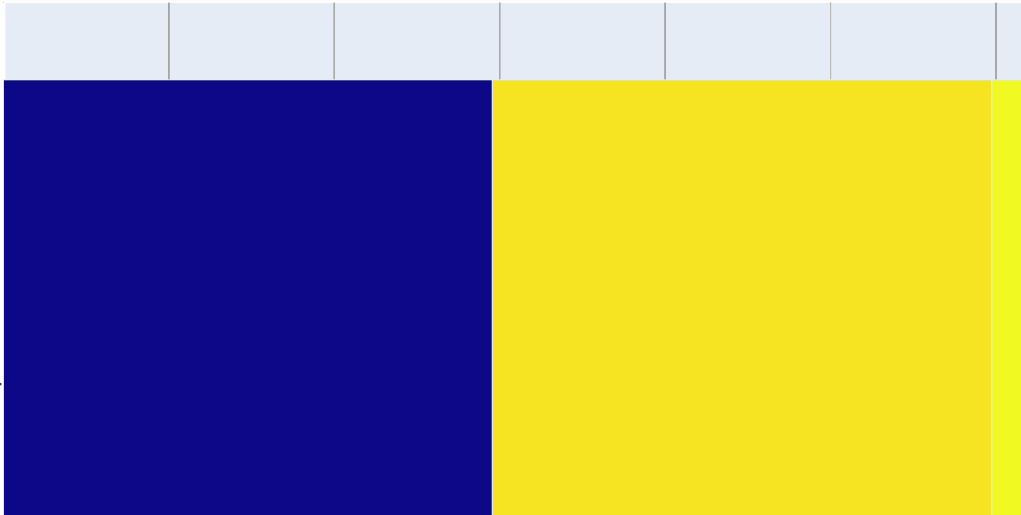
```
1 a = df["SEVERITYDESC"]
2 x = a.index
3 y = a.values
4 fig = px.bar(df["SEVERITYDESC"])
5 fig.show()
```



In [9]:

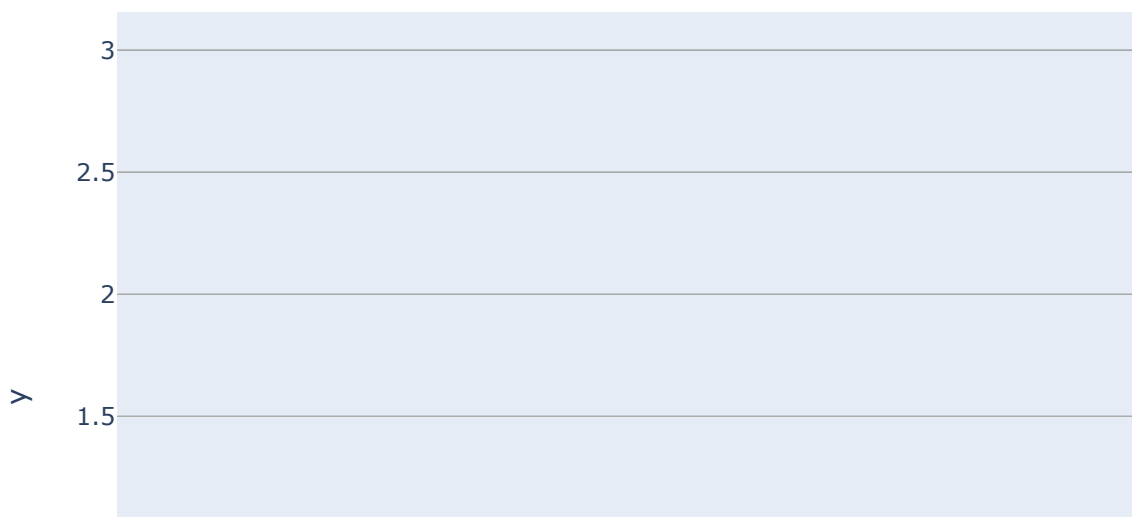
```
1 a = df["COLLISIONTYPE"]
2 b = df["COLLISIONTYPE"]
3 xx = a.index
4 yy = a.values
5 fig = px.bar(b, x = xx, y = yy, color = xx)
6 fig.show()
```

> Other



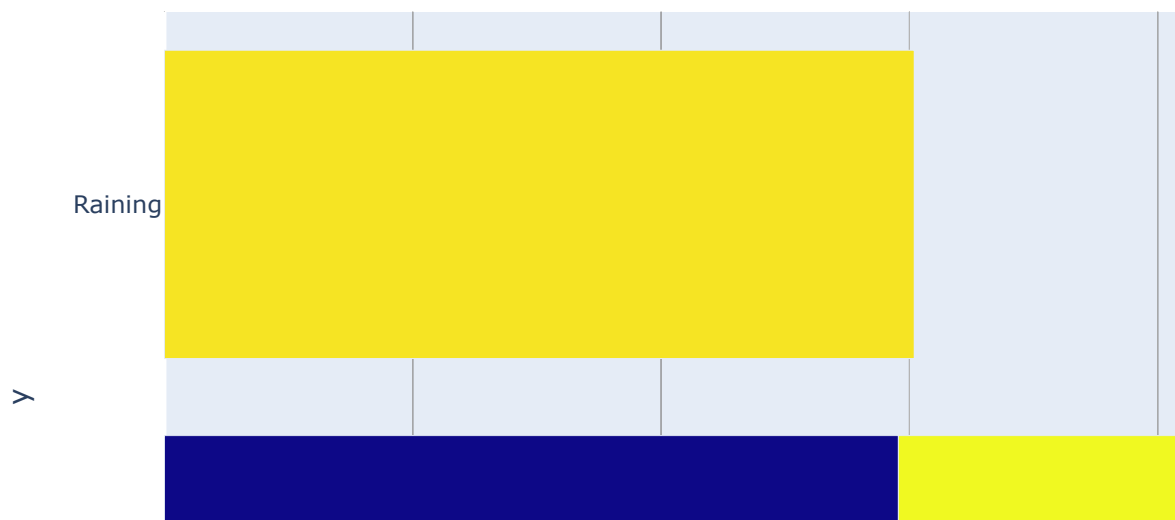
In [10]:

```
1 a = df["PERSONCOUNT"]
2 xx = a.index
3 yy = a.values
4 fig = px.bar(a, x = xx, y = yy, color = xx)
5 fig.show()
```



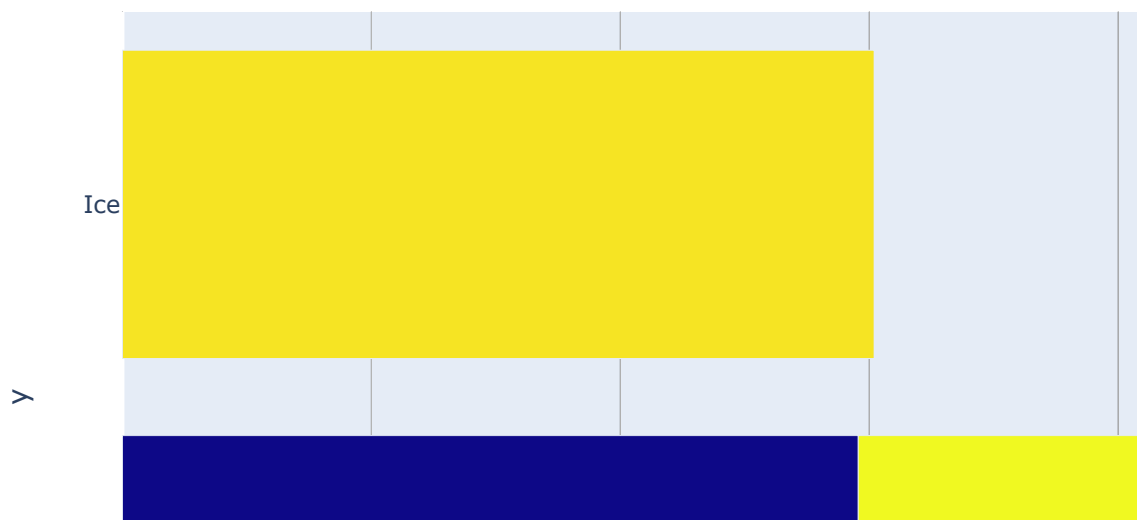
In [11]:

```
1 a = df["WEATHER"]
2 xx = a.index
3 yy = a.values
4 fig = px.bar(a, x = xx, y = yy, color = xx)
5 fig.show()
```



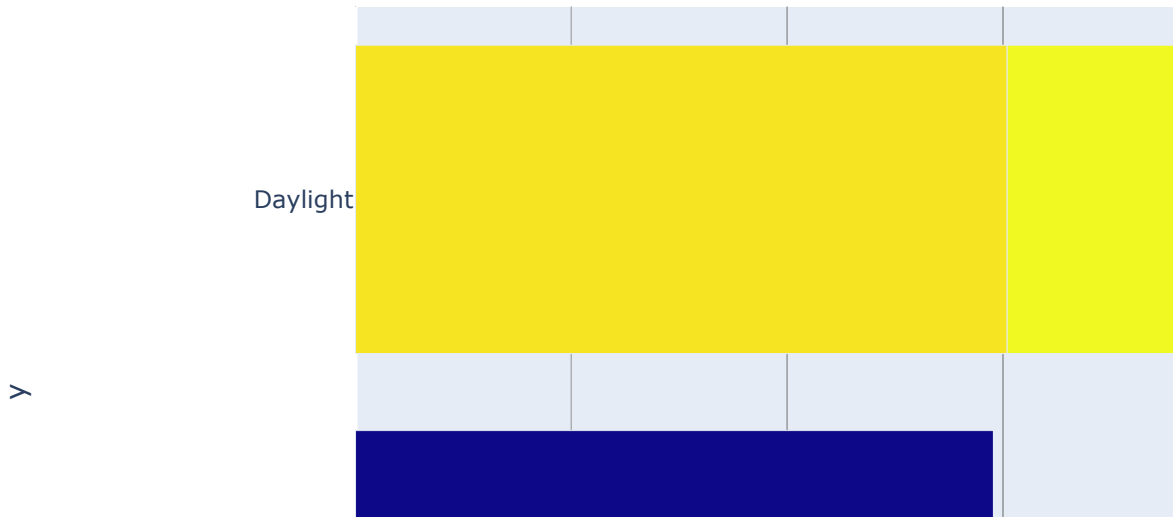
In [12]:

```
1 a = df["ROADCOND"]
2 xx = a.index
3 yy = a.values
4 fig = px.bar(a, x = xx, y = yy, color = xx)
5 fig.show()
```



In [13]:

```
1 a = df["LIGHTCOND"]
2 xx = a.index
3 yy = a.values
4 fig = px.bar(a, x = xx, y = yy, color = xx)
5 fig.show()
```



Conclusion

Number of accidents with Severity 1 is greater than that with Severity 2. More accidents occur at Blocks compared to that at Intersections. Majorly, accidents caused property damage to injuries. In most cases, only 1 person was involved in accidents while the maximum rose to 3. A very surprising thing noticed was accidents occurred more when:

1. Weather was clear
2. Road was dry
3. During Daylight

=> I have not supported my deductions with numbers as it is pretty clear from the graphs. Moreover, since my dataset is a sample and not census, therefore evaluating them in terms of numbers wouldn't be correct.

