In [3]:
```
import pandas as pd
import numpy as np
import csv

import scipy.stats as scs
import statsmodels.api as sm
import statsmodels.formula.api as sms
import scipy.stats as stats

from pltfunctions import hist_kde_plots
from haversine import haversine
from math import sqrt

from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.linear_model import LinearRegression
from sklearn.feature_selection import f_regression
import sklearn.metrics as metrics

import matplotlib.pyplot as plt
import seaborn as sns
```

Question 3

What vehicle defect is most common in reported crashes in t area?

In [12]:
```
df = pd.read_csv(r'data\Sample3.csv')
```

In [19]:
```
df_defect=df.groupby(df['VEHICLE_DEFECT']).count().sort_values(by='CRASH_DATE_
```

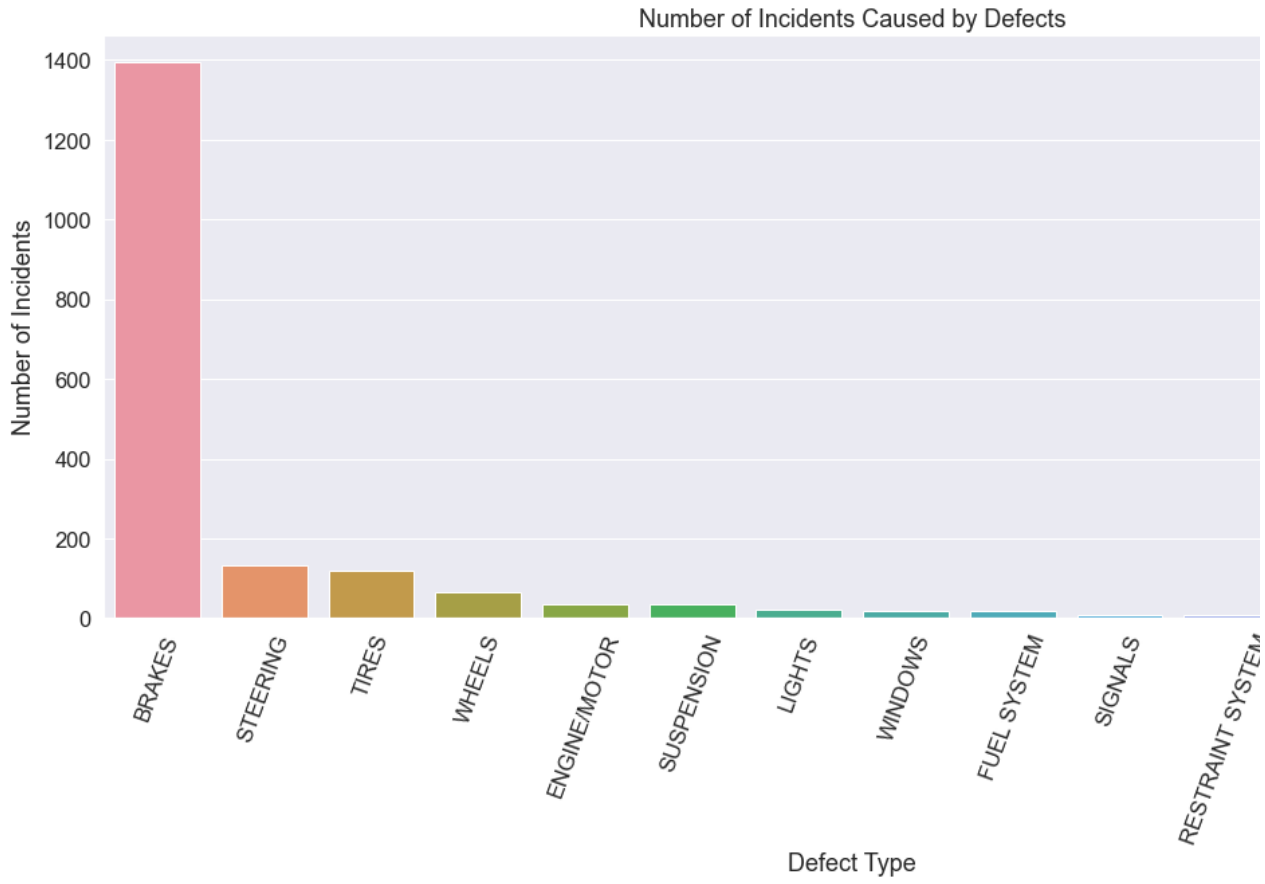
In [20]:
```
df_defect
```

	VEHICLE_DEFECT	Unnamed: 0	CRASH_DATE_x	UNIT_TYPE	MAKE	MODEL	VEHICLE
0	BRAKES	1394	1394	1394	1394	1394	1394
1	STEERING	132	132	132	132	132	132
2	TIRES	120	120	120	120	120	120
3	WHEELS	64	64	64	64	64	64
4	ENGINE/MOTOR	35	35	35	35	35	35
5	SUSPENSION	34	34	34	34	34	34
6	LIGHTS	23	23	23	23	23	23
7	WINDOWS	19	19	19	19	19	19
8	FUEL SYSTEM	17	17	17	17	17	17
9	SIGNALS	9	9	9	9	9	9
10	RESTRAINT SYSTEM	8	8	8	8	8	8
11	CARGO	4	4	4	4	4	4
12	EXHAUST	4	4	4	4	4	4
13	TRAILER COUPLING	1	1	1	1	1	1

14 rows × 49 columns

In [23]:

```
fig_dims = (20,8)
fig, ax = plt.subplots(figsize=fig_dims)
sns.set(font_scale=1.5)
sns.set_style("darkgrid")
palette = sns.set_palette("dark")
ax = sns.barplot(x=df_defect['VEHICLE_DEFECT'], y=df_defect['CRASH_DATE_x'])
ax.set(xlabel="Defect Type", ylabel="Number of Incidents")
plt.ticklabel_format(style='plain',axis='y')
plt.xticks(rotation=70)
plt.title('Number of Incidents Caused by Defects')
plt.show()
```



Question 3 Insights

There is no vehicle defect for the majority of accidents in the Chicago Area. H defect is listed, the majority of accidents are caused by malfunctioning brakes state registrations could be more demanding of individuals' vehicles.

In [ ]:

In [ ]:

In [ ]: