Introduction/Business Problem:

Using the collisions data provided by Coursera for the final capstone course, I wanted to look into the severity of car accidents that are due to bad weather conditions. How many car accidents are caused due to bad weather conditions? This data will bring awareness to people to drive extra cautiously during bad weather!

Data:

Using the data provided by Coursera on Collisions, I will investigate the connection between severity of car accidents and weather conditions. This data provides collisions from 2004 to the present in Seattle.

Methodology:

I used IBM Watson Studio to create the notebook and Github for the repository. The Python libraries that I used for data analysis includes Pandas,

First, I had to read the data from the provided csv file. Then, I used df.dtypes to see what type of data are the columns in the file. To investigate the connection between car accidents and weather conditions, I chose to focus on SEVERITYCODE (int), SEVERITYDESC (object), WEATHER (object), and ROADCOND (object).

```
import pandas as pd
df = pd.read_csv('https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv')
print(df.dtypes)
```

```
SEVERITYCODE        int64
X                 float64
Y                 float64
OBJECTID            int64
INCKEY              int64
COLDETKEY           int64
REPORTNO           object
STATUS             object
ADDRTYPE           object
INTKEY            float64
LOCATION           object
EXCEPTRSNCODE      object
EXCEPTRSNDESC      object
SEVERITYCODE.1      int64
SEVERITYDESC       object
COLLISIONTYPE      object
PERSONCOUNT         int64
PEDCOUNT            int64
PEDCYLCOUNT         int64
VEHCOUNT            int64
INCDATE            object
INCDTTM            object
JUNCTIONTYPE       object
SDOT_COLCODE        int64
SDOT_COLDESC       object
INATTENTIONIND     object
UNDERINFL          object
WEATHER            object
ROADCOND           object
LIGHTCOND          object
PEDROWNOTGRNT      object
SDOTCOLNUM        float64
SPEEDING           object
ST_COLCODE         object
ST_COLDESC         object
SEGLANEKEY          int64
CROSSWALKKEY        int64
HITPARKEDCAR       object
dtype: object
```

I ran a value count on WEATHER to see which weather condition had the most accidents. I also ran a value count on ROADCOND to see which type of roads had more accidents.

Results:

```
In [32]: df['WEATHER'].value_counts().to_frame()
```

Out[32]:

	WEATHER
Clear	111135
Raining	33145
Overcast	27714
Unknown	15091
Snowing	907
Other	832
Fog/Smog/Smoke	569
Sleet/Hail/Freezing Rain	113
Blowing Sand/Dirt	56
Severe Crosswind	25
Partly Cloudy	5

```
In [36]: df['ROADCOND'].value_counts().to_frame()
```

Out[36]:

	ROADCOND
Dry	124510
Wet	47474
Unknown	15078
Ice	1209
Snow/Slush	1004
Other	132
Standing Water	115
Sand/Mud/Dirt	75
Oil	64

Clear weather condition had the most incidents of collisions. Dry road conditions had the most incidents of collisions.
A new dataframe table was created just to show the focused columns.

Conclusion:

According to this data on collisions in Seattle from 2004 to the present, I do not see any relationship between bad weather conditions and wet road conditions that affected collisions. From the data, we see that there were a lot more collisions that happened on dry roads and also clear weather conditions. There are actually much less collisions that happen when weather and road conditions are not that great. This does not really surprise me since drivers tend to be more careful while driving when conditions are bad. The data shows that drivers are more likely to have a collision when weather conditions are good and roads are dry. This can be because the drivers are less careful in their driving during good conditions.