US Accidents Exploratory Data Analysis

January 14, 2023

Data Preparqation And Cleaning

Libraries that we will ue use here

[82]:
```
import pandas as pd
import numpy as np
import matplotlib as mt
```

ASK QUESTIONS AND ANSWERS

Talk about EDA Talk about the dataset (source . what it contains , how it will be useful) kaggle inform about the accidents can be useful to prevent the accidents

[42]:
```
Data = pd.read_csv(r"C:\Users\Deepak Kumar\Downloads\US_Accidents_Dec21_updated.
↪csv")
Data
```

[42]:

	ID	Severity	Start_Time	End_Time	\
0	A-1	3	2016-02-08 00:37:08	2016-02-08 06:37:08	
1	A-2	2	2016-02-08 05:56:20	2016-02-08 11:56:20	
2	A-3	2	2016-02-08 06:15:39	2016-02-08 12:15:39	
3	A-4	2	2016-02-08 06:51:45	2016-02-08 12:51:45	
4	A-5	3	2016-02-08 07:53:43	2016-02-08 13:53:43	
…	…	…	…	…	
2845337	A-2845338	2	2019-08-23 18:03:25	2019-08-23 18:32:01	
2845338	A-2845339	2	2019-08-23 19:11:30	2019-08-23 19:38:23	
2845339	A-2845340	2	2019-08-23 19:00:21	2019-08-23 19:28:49	
2845340	A-2845341	2	2019-08-23 19:00:21	2019-08-23 19:29:42	
2845341	A-2845342	2	2019-08-23 18:52:06	2019-08-23 19:21:31	

	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance(mi)	\
0	40.108910	-83.092860	40.112060	-83.031870	3.230	
1	39.865420	-84.062800	39.865010	-84.048730	0.747	
2	39.102660	-84.524680	39.102090	-84.523960	0.055	
3	41.062130	-81.537840	41.062170	-81.535470	0.123	
4	39.172393	-84.492792	39.170476	-84.501798	0.500	
…	…	…	…	…	…	
2845337	34.002480	-117.379360	33.998880	-117.370940	0.543	
2845338	32.766960	-117.148060	32.765550	-117.153630	0.338	

```
2845339  33.775450 -117.847790  33.777400 -117.857270         0.561
2845340  33.992460 -118.403020  33.983110 -118.395650         0.772
2845341  34.133930 -117.230920  34.137360 -117.239340         0.537

                                                 Description  …  Roundabout  \
0        Between Sawmill Rd/Exit 20 and OH-315/Olentang…  …       False
1                      At OH-4/OH-235/Exit 41 - Accident.  …       False
2                        At I-71/US-50/Exit 1 - Accident.  …       False
3                         At Dart Ave/Exit 21 - Accident.  …       False
4                      At Mitchell Ave/Exit 6 - Accident.  …       False
…                                                        …  …         …
2845337                          At Market St - Accident.  …       False
2845338    At Camino Del Rio/Mission Center Rd - Accident.  …       False
2845339  At Glassell St/Grand Ave - Accident. in the ri…  …       False
2845340     At CA-90/Marina Fwy/Jefferson Blvd - Accident.  …       False
2845341               At Highland Ave/Arden Ave - Accident.  …       False

         Station   Stop Traffic_Calming Traffic_Signal Turning_Loop  \
0          False  False           False          False        False
1          False  False           False          False        False
2          False  False           False          False        False
3          False  False           False          False        False
4          False  False           False          False        False
…          …    …              …              …          …
2845337    False  False           False          False        False
2845338    False  False           False          False        False
2845339    False  False           False          False        False
2845340    False  False           False          False        False
2845341    False  False           False          False        False

         Sunrise_Sunset Civil_Twilight Nautical_Twilight Astronomical_Twilight
0                 Night          Night             Night                 Night
1                 Night          Night             Night                 Night
2                 Night          Night             Night                   Day
3                 Night          Night               Day                   Day
4                   Day            Day               Day                   Day
…                   …            …                 …                   …
2845337             Day            Day               Day                   Day
2845338             Day            Day               Day                   Day
2845339             Day            Day               Day                   Day
2845340             Day            Day               Day                   Day
2845341             Day            Day               Day                   Day

[2845342 rows x 47 columns]
```

[3]:
```
Data.columns
```

[3]:
```
Index(['ID', 'Severity', 'Start_Time', 'End_Time', 'Start_Lat', 'Start_Lng',
       'End_Lat', 'End_Lng', 'Distance(mi)', 'Description', 'Number', 'Street',
       'Side', 'City', 'County', 'State', 'Zipcode', 'Country', 'Timezone',
       'Airport_Code', 'Weather_Timestamp', 'Temperature(F)', 'Wind_Chill(F)',
       'Humidity(%)', 'Pressure(in)', 'Visibility(mi)', 'Wind_Direction',
       'Wind_Speed(mph)', 'Precipitation(in)', 'Weather_Condition', 'Amenity',
       'Bump', 'Crossing', 'Give_Way', 'Junction', 'No_Exit', 'Railway',
       'Roundabout', 'Station', 'Stop', 'Traffic_Calming', 'Traffic_Signal',
       'Turning_Loop', 'Sunrise_Sunset', 'Civil_Twilight', 'Nautical_Twilight',
       'Astronomical_Twilight'],
      dtype='object')
```

Total rows and column in this data setset.

[4]:
```
Data.shape
```

[4]:
```
(2845342, 47)
```

[5]:
```
Data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2845342 entries, 0 to 2845341
Data columns (total 47 columns):
 #   Column                 Dtype
---  ------                 -----
 0   ID                     object
 1   Severity               int64
 2   Start_Time             object
 3   End_Time               object
 4   Start_Lat              float64
 5   Start_Lng              float64
 6   End_Lat                float64
 7   End_Lng                float64
 8   Distance(mi)           float64
 9   Description            object
 10  Number                 float64
 11  Street                 object
 12  Side                   object
 13  City                   object
 14  County                 object
 15  State                  object
 16  Zipcode                object
 17  Country                object
 18  Timezone               object
 19  Airport_Code           object
 20  Weather_Timestamp      object
 21  Temperature(F)         float64
 22  Wind_Chill(F)          float64
```

```
 23  Humidity(%)            float64
 24  Pressure(in)           float64
 25  Visibility(mi)         float64
 26  Wind_Direction         object
 27  Wind_Speed(mph)        float64
 28  Precipitation(in)      float64
 29  Weather_Condition      object
 30  Amenity                bool
 31  Bump                   bool
 32  Crossing               bool
 33  Give_Way               bool
 34  Junction               bool
 35  No_Exit                bool
 36  Railway                bool
 37  Roundabout             bool
 38  Station                bool
 39  Stop                   bool
 40  Traffic_Calming        bool
 41  Traffic_Signal         bool
 42  Turning_Loop           bool
 43  Sunrise_Sunset         object
 44  Civil_Twilight         object
 45  Nautical_Twilight      object
 46  Astronomical_Twilight  object
dtypes: bool(13), float64(13), int64(1), object(20)
memory usage: 773.4+ MB
```

Details about the Us accidents dataset.

[6]:
```
Data.describe()
```

[6]:

	Severity	Start_Lat	Start_Lng	End_Lat	End_Lng
count	2.845342e+06	2.845342e+06	2.845342e+06	2.845342e+06	2.845342e+06
mean	2.137572e+00	3.624520e+01	-9.711463e+01	3.624532e+01	-9.711439e+01
std	4.787216e-01	5.363797e+00	1.831782e+01	5.363873e+00	1.831763e+01
min	1.000000e+00	2.456603e+01	-1.245481e+02	2.456601e+01	-1.245457e+02
25%	2.000000e+00	3.344517e+01	-1.180331e+02	3.344628e+01	-1.180333e+02
50%	2.000000e+00	3.609861e+01	-9.241808e+01	3.609799e+01	-9.241772e+01
75%	2.000000e+00	4.016024e+01	-8.037243e+01	4.016105e+01	-8.037338e+01
max	4.000000e+00	4.900058e+01	-6.711317e+01	4.907500e+01	-6.710924e+01

	Distance(mi)	Number	Temperature(F)	Wind_Chill(F)
count	2.845342e+06	1.101431e+06	2.776068e+06	2.375699e+06
mean	7.026779e-01	8.089408e+03	6.179356e+01	5.965823e+01
std	1.560361e+00	1.836009e+04	1.862263e+01	2.116097e+01
min	0.000000e+00	0.000000e+00	-8.900000e+01	-8.900000e+01
25%	5.200000e-02	1.270000e+03	5.000000e+01	4.600000e+01
50%	2.440000e-01	4.007000e+03	6.400000e+01	6.300000e+01

```
75%    7.640000e-01  9.567000e+03    7.600000e+01   7.600000e+01
max    1.551860e+02  9.999997e+06    1.960000e+02   1.960000e+02

        Humidity(%)  Pressure(in)  Visibility(mi)  Wind_Speed(mph)  \
count  2.772250e+06  2.786142e+06    2.774796e+06     2.687398e+06
mean   6.436545e+01  2.947234e+01    9.099391e+00     7.395044e+00
std    2.287457e+01  1.045286e+00    2.717546e+00     5.527454e+00
min    1.000000e+00  0.000000e+00    0.000000e+00     0.000000e+00
25%    4.800000e+01  2.931000e+01    1.000000e+01     3.500000e+00
50%    6.700000e+01  2.982000e+01    1.000000e+01     7.000000e+00
75%    8.300000e+01  3.001000e+01    1.000000e+01     1.000000e+01
max    1.000000e+02  5.890000e+01    1.400000e+02     1.087000e+03

       Precipitation(in)
count       2.295884e+06
mean        7.016940e-03
std         9.348831e-02
min         0.000000e+00
25%         0.000000e+00
50%         0.000000e+00
75%         0.000000e+00
max         2.400000e+01
```

How many numeric column in this data?

[7]:
```
numerics = ['int16', 'int32', 'int64', 'float16', 'float32', 'float64']

new_df = Data.select_dtypes(include=numerics)
len(new_df.columns)
```

[7]:
```
14
```

How many missing values in dataset.

[8]:
```
Data.isnull()
```

[8]:
```
              ID  Severity  Start_Time  End_Time  Start_Lat  Start_Lng  End_Lat  \
0          False     False       False     False      False      False    False
1          False     False       False     False      False      False    False
2          False     False       False     False      False      False    False
3          False     False       False     False      False      False    False
4          False     False       False     False      False      False    False
…          …       …         …         …          …          …        …
2845337    False     False       False     False      False      False    False
2845338    False     False       False     False      False      False    False
2845339    False     False       False     False      False      False    False
2845340    False     False       False     False      False      False    False
2845341    False     False       False     False      False      False    False
```

```
         End_Lng  Distance(mi)  Description  …  Roundabout  Station   Stop  \
0          False         False        False  …       False    False  False
1          False         False        False  …       False    False  False
2          False         False        False  …       False    False  False
3          False         False        False  …       False    False  False
4          False         False        False  …       False    False  False
…            …             …            …  …         …        …      …
2845337    False         False        False  …       False    False  False
2845338    False         False        False  …       False    False  False
2845339    False         False        False  …       False    False  False
2845340    False         False        False  …       False    False  False
2845341    False         False        False  …       False    False  False

         Traffic_Calming  Traffic_Signal  Turning_Loop  Sunrise_Sunset  \
0                  False           False         False           False
1                  False           False         False           False
2                  False           False         False           False
3                  False           False         False           False
4                  False           False         False           False
…                    …               …             …               …
2845337            False           False         False           False
2845338            False           False         False           False
2845339            False           False         False           False
2845340            False           False         False           False
2845341            False           False         False           False

         Civil_Twilight  Nautical_Twilight  Astronomical_Twilight
0                 False              False                  False
1                 False              False                  False
2                 False              False                  False
3                 False              False                  False
4                 False              False                  False
…                   …                  …                      …
2845337           False              False                  False
2845338           False              False                  False
2845339           False              False                  False
2845340           False              False                  False
2845341           False              False                  False

[2845342 rows x 47 columns]
```

[9]:
```
Data.isnull().count()
```

[9]:
```
ID                     2845342
Severity               2845342
Start_Time             2845342
```

```
End_Time                2845342
Start_Lat               2845342
Start_Lng               2845342
End_Lat                 2845342
End_Lng                 2845342
Distance(mi)            2845342
Description             2845342
Number                  2845342
Street                  2845342
Side                    2845342
City                    2845342
County                  2845342
State                   2845342
Zipcode                 2845342
Country                 2845342
Timezone                2845342
Airport_Code            2845342
Weather_Timestamp       2845342
Temperature(F)          2845342
Wind_Chill(F)           2845342
Humidity(%)             2845342
Pressure(in)            2845342
Visibility(mi)          2845342
Wind_Direction          2845342
Wind_Speed(mph)         2845342
Precipitation(in)       2845342
Weather_Condition       2845342
Amenity                 2845342
Bump                    2845342
Crossing                2845342
Give_Way                2845342
Junction                2845342
No_Exit                 2845342
Railway                 2845342
Roundabout              2845342
Station                 2845342
Stop                    2845342
Traffic_Calming         2845342
Traffic_Signal          2845342
Turning_Loop            2845342
Sunrise_Sunset          2845342
Civil_Twilight          2845342
Nautical_Twilight       2845342
Astronomical_Twilight   2845342
dtype: int64
```

[10]:
```
Data.isnull().sum()
```

[10]:

ID	0
Severity	0
Start_Time	0
End_Time	0
Start_Lat	0
Start_Lng	0
End_Lat	0
End_Lng	0
Distance(mi)	0
Description	0
Number	1743911
Street	2
Side	0
City	137
County	0
State	0
Zipcode	1319
Country	0
Timezone	3659
Airport_Code	9549
Weather_Timestamp	50736
Temperature(F)	69274
Wind_Chill(F)	469643
Humidity(%)	73092
Pressure(in)	59200
Visibility(mi)	70546
Wind_Direction	73775
Wind_Speed(mph)	157944
Precipitation(in)	549458
Weather_Condition	70636
Amenity	0
Bump	0
Crossing	0
Give_Way	0
Junction	0
No_Exit	0
Railway	0
Roundabout	0
Station	0
Stop	0
Traffic_Calming	0
Traffic_Signal	0
Turning_Loop	0
Sunrise_Sunset	2867
Civil_Twilight	2867
Nautical_Twilight	2867
Astronomical_Twilight	2867

```
dtype: int64
```

Percentage of missing values

[11]:
```
Missing_percentage = Data.isnull().sum().sort_values(ascending = False)/
  ↪len(Data)
Missing_percentage
```

[11]:
```
Number                  6.129003e-01
Precipitation(in)       1.931079e-01
Wind_Chill(F)           1.650568e-01
Wind_Speed(mph)         5.550967e-02
Wind_Direction          2.592834e-02
Humidity(%)             2.568830e-02
Weather_Condition       2.482514e-02
Visibility(mi)          2.479350e-02
Temperature(F)          2.434646e-02
Pressure(in)            2.080593e-02
Weather_Timestamp       1.783125e-02
Airport_Code            3.356011e-03
Timezone                1.285961e-03
Nautical_Twilight       1.007612e-03
Civil_Twilight          1.007612e-03
Sunrise_Sunset          1.007612e-03
Astronomical_Twilight   1.007612e-03
Zipcode                 4.635647e-04
City                    4.814887e-05
Street                  7.029032e-07
Country                 0.000000e+00
Junction                0.000000e+00
Start_Time              0.000000e+00
End_Time                0.000000e+00
Start_Lat               0.000000e+00
Turning_Loop            0.000000e+00
Traffic_Signal          0.000000e+00
Traffic_Calming         0.000000e+00
Stop                    0.000000e+00
Station                 0.000000e+00
Roundabout              0.000000e+00
Railway                 0.000000e+00
No_Exit                 0.000000e+00
Crossing                0.000000e+00
Give_Way                0.000000e+00
Bump                    0.000000e+00
Amenity                 0.000000e+00
Start_Lng               0.000000e+00
End_Lat                 0.000000e+00
End_Lng                 0.000000e+00
```

```
Distance(mi)            0.000000e+00
Description             0.000000e+00
Severity                0.000000e+00
Side                    0.000000e+00
County                  0.000000e+00
State                   0.000000e+00
ID                      0.000000e+00
dtype: float64
```

[12]:
```
type(Missing_percentage)
```

[12]:
```
pandas.core.series.Series
```

[13]:
```
Missing_percentage != 0
```

[13]:
```
Number                   True
Precipitation(in)        True
Wind_Chill(F)            True
Wind_Speed(mph)          True
Wind_Direction           True
Humidity(%)              True
Weather_Condition        True
Visibility(mi)           True
Temperature(F)           True
Pressure(in)             True
Weather_Timestamp        True
Airport_Code             True
Timezone                 True
Nautical_Twilight        True
Civil_Twilight           True
Sunrise_Sunset           True
Astronomical_Twilight    True
Zipcode                  True
City                     True
Street                   True
Country                 False
Junction                False
Start_Time              False
End_Time                False
Start_Lat               False
Turning_Loop            False
Traffic_Signal          False
Traffic_Calming         False
Stop                    False
Station                 False
Roundabout              False
Railway                 False
```

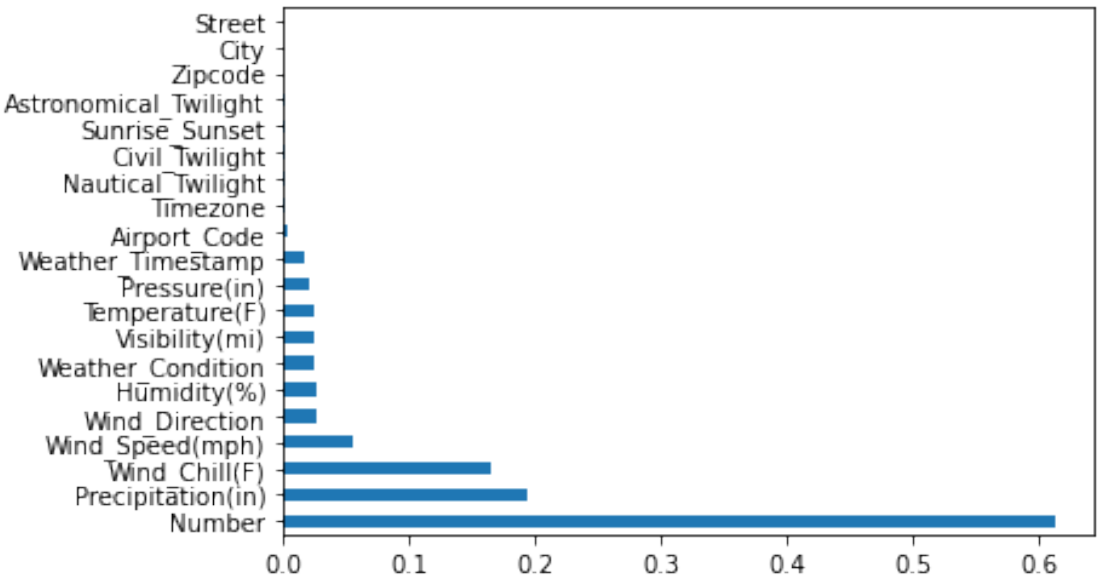
```
No_Exit                 False
Crossing                False
Give_Way                False
Bump                    False
Amenity                 False
Start_Lng               False
End_Lat                 False
End_Lng                 False
Distance(mi)            False
Description             False
Severity                False
Side                    False
County                  False
State                   False
ID                      False
dtype: bool
```

[14]:
```
Missing_percentage[Missing_percentage != 0]
```

[14]:
```
Number                  6.129003e-01
Precipitation(in)       1.931079e-01
Wind_Chill(F)           1.650568e-01
Wind_Speed(mph)         5.550967e-02
Wind_Direction          2.592834e-02
Humidity(%)             2.568830e-02
Weather_Condition       2.482514e-02
Visibility(mi)          2.479350e-02
Temperature(F)          2.434646e-02
Pressure(in)            2.080593e-02
Weather_Timestamp       1.783125e-02
Airport_Code            3.356011e-03
Timezone                1.285961e-03
Nautical_Twilight       1.007612e-03
Civil_Twilight          1.007612e-03
Sunrise_Sunset          1.007612e-03
Astronomical_Twilight   1.007612e-03
Zipcode                 4.635647e-04
City                    4.814887e-05
Street                  7.029032e-07
dtype: float64
```

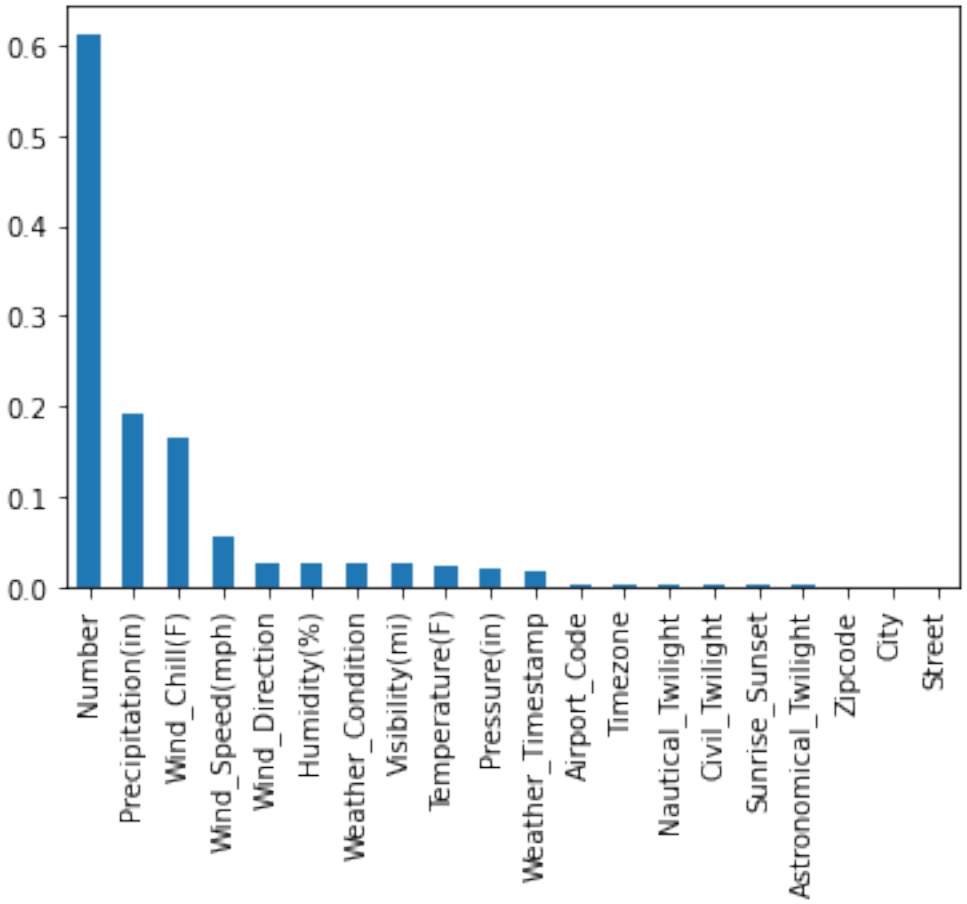
[15]:
```
Missing_percentage[Missing_percentage != 0].plot(kind = 'barh')
```

[15]:
```
<AxesSubplot:>
```



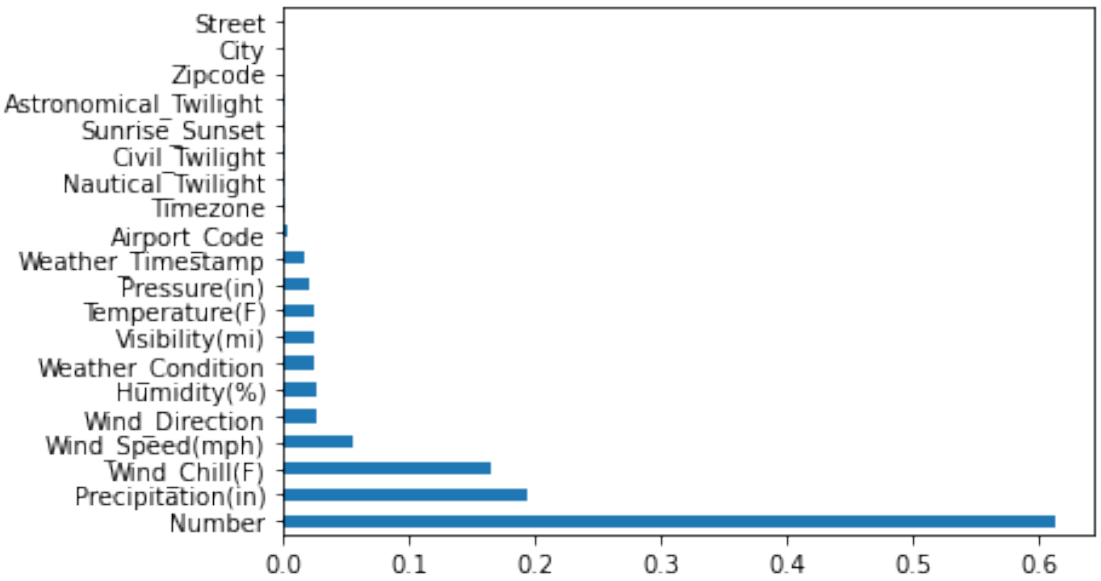
[16]:
```
Missing_percentage[Missing_percentage != 0].plot(kind = 'bar')
```

[16]: <AxesSubplot:>



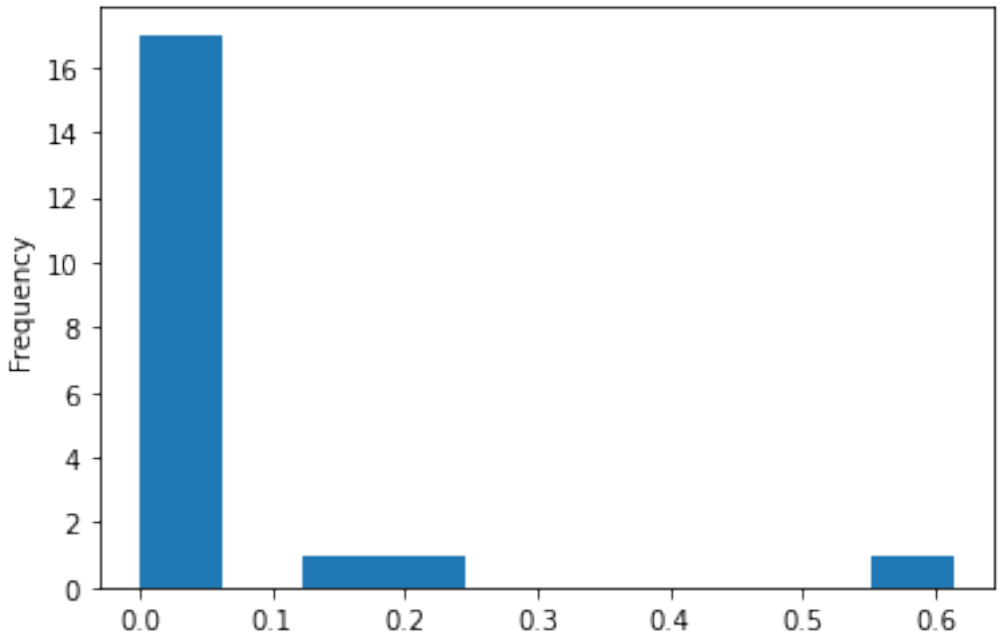
[17]:
```
Missing_percentage[Missing_percentage != 0].plot(kind = 'barh')
```

[17]: <AxesSubplot:>



[18]:
```
Missing_percentage[Missing_percentage != 0].plot(kind = 'hist')
```

[18]:
```
<AxesSubplot:ylabel='Frequency'>
```



Remove column that you don't want to use.

Column we will analyse.

City Start time Start lat, start lng Temperature Weather CX onition

[19]:
```
Data.City
```

[19]:
```
0               Dublin
1               Dayton
2           Cincinnati
3                Akron
4           Cincinnati
              …
2845337      Riverside
2845338      San Diego
2845339         Orange
2845340    Culver City
2845341       Highland
Name: City, Length: 2845342, dtype: object
```

[20]:
```
Data.City.count()
```

[20]:
```
2845205
```

[21]:
```
Cities = Data.City.unique()
len(Cities)
```

[21]:
```
11682
```

[22]:
```
Cities = Data.City.unique()
Cities[ :100]
```

[22]:
```
array(['Dublin', 'Dayton', 'Cincinnati', 'Akron', 'Williamsburg',
       'Cleveland', 'Lima', 'Westerville', 'Jamestown', 'Freeport',
       'Columbus', 'Toledo', 'Roanoke', 'Ft Mitchell', 'Edinburgh',
       'Fairborn', 'Shelbyville', 'Greensburg', 'Saint Paul',
       'Parkersburg', 'Indianapolis', 'Dundee', 'Jeffersonville',
       'Pittsburgh', 'Lewis Center', 'Dunkirk', 'Redkey', 'Milton',
       'Willshire', 'Straughn', 'Cambridge Springs', 'Fremont',
       'Louisville', 'South Charleston', 'Edinboro', 'Buckhannon',
       'Lockbourne', 'Painesville', 'Washington', 'Dunbar', 'Angola',
       'Edon', 'Medina', 'De Mossville', 'New Albany', 'Charleston',
       'Fort Wayne', 'Burnsville', 'Bedford', 'Clarksville', 'Lakewood',
       'Richfield', 'Sewickley', 'Independence', 'Westlake', 'Erlanger',
       'Grove City', 'Monroe', 'West Middlesex', 'Gaston', 'Economy',
       'Fairmount', 'Hagerstown', 'Walton', 'Crittenden', 'Coraopolis',
       'Holland', 'Greenfield', 'Anderson', 'Englewood', 'Knightstown',
       'Bentleyville', 'Memphis', 'Henryville', 'Kendallville', 'Avilla',
       'Ohio City', 'Van Wert', 'Rocky River', 'Sturgis', 'West Chester',
       'Orient', 'Madison', 'Deputy', 'Keystone', 'Mercer', 'Bryant',
       'Pennville', 'Kimbolton', 'Thornville', 'Wexford', 'Fishers',
```

```
       'Noblesville', 'Macedonia', 'Youngstown', 'Fairdale', 'Sutton',
       'Mount Sterling', 'Northwood', 'Huntington'], dtype=object)
```

[23]:
```
Cities_by_accident = Data.City.value_counts()
Cities_by_accident
```

[23]:
```
Miami                              106966
Los Angeles                         68956
Orlando                             54691
Dallas                              41979
Houston                             39448
                                    ...
Ridgedale                               1
Sekiu                                   1
Wooldridge                              1
Bullock                                 1
American Fork-Pleasant Grove            1
Name: City, Length: 11681, dtype: int64
```

Values exist or not

[24]:
```
'Los Angeles' in Data['City'].values
```

[24]:
```
True
```

[25]:
```
'New York' in Data['City'].values
```

[25]:
```
True
```

[26]:
```
'ABC' in Data['City'].values
```

[26]:
```
False
```

Total Accidents populataion in New York.

[27]:
```
'Los Angeles' in Data['City'].values
```

[27]:
```
True
```

[28]:
```
# total accident in las Angeles
Cities_by_accident['Los Angeles']
```

[28]:
```
68956
```

[29]:
```
Cities_by_accident['Miami']
```

[29]:
```
106966
```

[30]:
```
Cities_by_accident = Data.City.value_counts()
Cities_by_accident
```

[30]:
```
Miami                            106966
Los Angeles                       68956
Orlando                           54691
Dallas                            41979
Houston                           39448
                                  ...
Ridgedale                             1
Sekiu                                 1
Wooldridge                            1
Bullock                               1
American Fork-Pleasant Grove          1
Name: City, Length: 11681, dtype: int64
```

[31]:
```
Data.loc[Data['City']=='Los Angeles']
```

[31]:

	ID	Severity	Start_Time	End_Time	\
5235	A-5236	2	2016-03-22 19:36:44	2016-03-23 01:36:44	
5238	A-5239	2	2016-03-22 20:59:43	2016-03-23 02:59:43	
5253	A-5254	3	2016-03-23 07:59:47	2016-03-23 13:59:47	
5271	A-5272	2	2016-03-23 11:50:32	2016-03-23 17:50:32	
5273	A-5274	2	2016-03-23 12:16:45	2016-03-23 18:16:45	
…	…	…	…	…	
2844905	A-2844906	2	2019-08-22 17:07:14	2019-08-22 17:36:02	
2845305	A-2845306	3	2019-08-23 04:04:48	2019-08-23 04:33:53	
2845309	A-2845310	2	2019-08-23 12:52:31	2019-08-23 13:20:14	
2845312	A-2845313	2	2019-08-23 13:42:50	2019-08-23 14:10:06	
2845324	A-2845325	2	2019-08-23 15:45:43	2019-08-23 16:14:31	

	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance(mi)	\
5235	34.09256	-118.206220	34.092560	-118.206220	0.000	
5238	33.94819	-118.279730	33.946760	-118.279750	0.099	
5253	34.02330	-118.172880	34.021380	-118.173390	0.136	
5271	34.14470	-118.278650	34.141040	-118.277840	0.257	
5273	34.09914	-118.251853	34.099817	-118.251396	0.054	
…	…	…	…	…	…	
2844905	34.03693	-118.438770	34.025590	-118.429180	0.957	
2845305	34.07579	-118.276680	34.074310	-118.272250	0.273	
2845309	34.02379	-118.276390	34.025760	-118.275290	0.150	
2845312	34.07061	-118.263910	34.069740	-118.261550	0.148	
2845324	34.04365	-118.443730	34.049340	-118.448420	0.476	

	Description	…	Roundabout	\
5235	At Avenue 43 - Accident.	…	False	
5238	At Century Blvd - Accident.	…	False	

```
5253             At Whittier Blvd/Olympic Blvd - Accident.  …       False
5271                               At Colorado St - Accident.  …       False
5273     At I-5/Golden State Fwy - Accident. Left lane …  …       False
…                                                          …  …         …
2844905                 At I-10/Santa Monica Fwy - Accident.  …       False
2845305      At Benton Way/Rampart Blvd/Exit 5A - Accident.  …       False
2845309                                 At 28th St - Accident.  …       False
2845312               At Glendale Blvd/Union Ave - Accident.  …       False
2845324      At CA-2/Santa Monica Blvd/Exit 55A - Accident.  …       False

        Station   Stop Traffic_Calming Traffic_Signal Turning_Loop  \
5235      False   True           False          False        False
5238      False  False           False          False        False
5253      False  False           False          False        False
5271      False  False           False          False        False
5273      False  False           False          False        False
…         …    …              …              …          …
2844905   False  False           False          False        False
2845305   False  False           False          False        False
2845309   False  False           False          False        False
2845312   False  False           False          False        False
2845324   False  False           False          False        False

        Sunrise_Sunset Civil_Twilight Nautical_Twilight Astronomical_Twilight
5235             Night          Night               Day                   Day
5238             Night          Night             Night                 Night
5253               Day            Day               Day                   Day
5271               Day            Day               Day                   Day
5273               Day            Day               Day                   Day
…                  …              …                 …                     …
2844905            Day            Day               Day                   Day
2845305          Night          Night             Night                 Night
2845309            Day            Day               Day                   Day
2845312            Day            Day               Day                   Day
2845324            Day            Day               Day                   Day

[68956 rows x 47 columns]
```

[32]:
```
Data.columns
```

[32]:
```
Index(['ID', 'Severity', 'Start_Time', 'End_Time', 'Start_Lat', 'Start_Lng',
       'End_Lat', 'End_Lng', 'Distance(mi)', 'Description', 'Number', 'Street',
       'Side', 'City', 'County', 'State', 'Zipcode', 'Country', 'Timezone',
       'Airport_Code', 'Weather_Timestamp', 'Temperature(F)', 'Wind_Chill(F)',
       'Humidity(%)', 'Pressure(in)', 'Visibility(mi)', 'Wind_Direction',
       'Wind_Speed(mph)', 'Precipitation(in)', 'Weather_Condition', 'Amenity',
       'Bump', 'Crossing', 'Give_Way', 'Junction', 'No_Exit', 'Railway',
```

```
       'Roundabout', 'Station', 'Stop', 'Traffic_Calming', 'Traffic_Signal',
       'Turning_Loop', 'Sunrise_Sunset', 'Civil_Twilight', 'Nautical_Twilight',
       'Astronomical_Twilight'],
      dtype='object')
```

Ask and Answer the questions. 1.Are the more accidents in warmer sor colder areas? 2.Which states have the highest number of accidents?How about per capita? 3.Does new work show up in the data ?If yes , why is the count lower if this the most populated city.

[33]:
```
Data.describe()
```

[33]:

	Severity	Start_Lat	Start_Lng	End_Lat	End_Lng
count	2.845342e+06	2.845342e+06	2.845342e+06	2.845342e+06	2.845342e+06
mean	2.137572e+00	3.624520e+01	-9.711463e+01	3.624532e+01	-9.711439e+01
std	4.787216e-01	5.363797e+00	1.831782e+01	5.363873e+00	1.831763e+01
min	1.000000e+00	2.456603e+01	-1.245481e+02	2.456601e+01	-1.245457e+02
25%	2.000000e+00	3.344517e+01	-1.180331e+02	3.344628e+01	-1.180333e+02
50%	2.000000e+00	3.609861e+01	-9.241808e+01	3.609799e+01	-9.241772e+01
75%	2.000000e+00	4.016024e+01	-8.037243e+01	4.016105e+01	-8.037338e+01
max	4.000000e+00	4.900058e+01	-6.711317e+01	4.907500e+01	-6.710924e+01

	Distance(mi)	Number	Temperature(F)	Wind_Chill(F)
count	2.845342e+06	1.101431e+06	2.776068e+06	2.375699e+06
mean	7.026779e-01	8.089408e+03	6.179356e+01	5.965823e+01
std	1.560361e+00	1.836009e+04	1.862263e+01	2.116097e+01
min	0.000000e+00	0.000000e+00	-8.900000e+01	-8.900000e+01
25%	5.200000e-02	1.270000e+03	5.000000e+01	4.600000e+01
50%	2.440000e-01	4.007000e+03	6.400000e+01	6.300000e+01
75%	7.640000e-01	9.567000e+03	7.600000e+01	7.600000e+01
max	1.551860e+02	9.999997e+06	1.960000e+02	1.960000e+02

	Humidity(%)	Pressure(in)	Visibility(mi)	Wind_Speed(mph)
count	2.772250e+06	2.786142e+06	2.774796e+06	2.687398e+06
mean	6.436545e+01	2.947234e+01	9.099391e+00	7.395044e+00
std	2.287457e+01	1.045286e+00	2.717546e+00	5.527454e+00
min	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
25%	4.800000e+01	2.931000e+01	1.000000e+01	3.500000e+00
50%	6.700000e+01	2.982000e+01	1.000000e+01	7.000000e+00
75%	8.300000e+01	3.001000e+01	1.000000e+01	1.000000e+01
max	1.000000e+02	5.890000e+01	1.400000e+02	1.087000e+03

	Precipitation(in)
count	2.295884e+06
mean	7.016940e-03
std	9.348831e-02
min	0.000000e+00
25%	0.000000e+00
50%	0.000000e+00

```
75%          0.000000e+00
max          2.400000e+01
```

[53]:
```
Data['State'],['Temperature(F)']
```

[53]:
```
(0          OH
 1          OH
 2          OH
 3          OH
 4          OH
            ..
 2845337    CA
 2845338    CA
 2845339    CA
 2845340    CA
 2845341    CA
 Name: State, Length: 2845342, dtype: object,
 ['Temperature(F)'])
```

[62]:
```
Temp = Data['Temperature(F)'].unique().sum()
Temp
```

[62]:
```
nan
```

[75]:
```
Data.shape
```

[75]:
```
(2845342, 47)
```

[77]:
```
Temp1 = Data['Temperature(F)'].shape
```

[78]:
```
Temp2 = Data['Temperature(F)'].count()
```

[79]:
```
Temp3 = Data['Temperature(F)'].isnull().count()
```

[54]:
```
Data.head()
```

[54]:

	ID	Severity	Start_Time	End_Time	Start_Lat	\
0	A-1	3	2016-02-08 00:37:08	2016-02-08 06:37:08	40.108910	
1	A-2	2	2016-02-08 05:56:20	2016-02-08 11:56:20	39.865420	
2	A-3	2	2016-02-08 06:15:39	2016-02-08 12:15:39	39.102660	
3	A-4	2	2016-02-08 06:51:45	2016-02-08 12:51:45	41.062130	
4	A-5	3	2016-02-08 07:53:43	2016-02-08 13:53:43	39.172393	

	Start_Lng	End_Lat	End_Lng	Distance(mi)	\
0	-83.092860	40.112060	-83.031870	3.230	
1	-84.062800	39.865010	-84.048730	0.747	
2	-84.524680	39.102090	-84.523960	0.055	
3	-81.537840	41.062170	-81.535470	0.123	

```
4 -84.492792  39.170476 -84.501798         0.500

                                          Description  …  Roundabout Station  \
0  Between Sawmill Rd/Exit 20 and OH-315/Olentang…  …       False   False
1                 At OH-4/OH-235/Exit 41 - Accident.  …       False   False
2                   At I-71/US-50/Exit 1 - Accident.  …       False   False
3                    At Dart Ave/Exit 21 - Accident.  …       False   False
4                 At Mitchell Ave/Exit 6 - Accident.  …       False   False

    Stop Traffic_Calming Traffic_Signal Turning_Loop Sunrise_Sunset  \
0  False           False          False        False          Night
1  False           False          False        False          Night
2  False           False          False        False          Night
3  False           False          False        False          Night
4  False           False          False        False            Day

  Civil_Twilight Nautical_Twilight Astronomical_Twilight
0          Night             Night                 Night
1          Night             Night                 Night
2          Night             Night                   Day
3          Night               Day                   Day
4            Day               Day                   Day

[5 rows x 47 columns]
```

[ ]: