Data Statistics

In [1]:
```
# Packages loader

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

In [2]:
```
# Reading denver crime CSV file using pandas

df_denver_crime = pd.read_csv('denver_crime.csv')
```

In [3]:
```
# Displaying headers of dataset

df_denver_crime.head()
```

Out[3]:

	INCIDENT_ID	OFFENSE_ID	OFFENSE_CODE	OFFENSE_CODE_EXTENSION	OFFENSE_TYPE_ID	OFI
0	20206002576	20206002576230500	2305	0	theft-items-from-vehicle	the
1	20166006518	20166006518230500	2305	0	theft-items-from-vehicle	the
2	2021174815	2021174815299900	2999	0	criminal-mischief-other	
3	2017139511	2017139511549900	5499	0	traf-other	
4	202138020	202138020220300	2203	0	burglary-business-by-force	

In [4]:
```
# Learning about the missing information, data-types, non-null row count

df_denver_crime.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 536355 entries, 0 to 536354
Data columns (total 19 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   INCIDENT_ID             536355 non-null  int64
 1   OFFENSE_ID              536355 non-null  int64
 2   OFFENSE_CODE            536355 non-null  int64
 3   OFFENSE_CODE_EXTENSION  536355 non-null  int64
 4   OFFENSE_TYPE_ID         536355 non-null  object
 5   OFFENSE_CATEGORY_ID     536355 non-null  object
 6   FIRST_OCCURRENCE_DATE   536355 non-null  object
 7   LAST_OCCURRENCE_DATE    191519 non-null  object
 8   REPORTED_DATE           536355 non-null  object
 9   INCIDENT_ADDRESS        490754 non-null  object
```

```
 10  GEO_X                   531770 non-null  float64
 11  GEO_Y                   531770 non-null  float64
 12  GEO_LON                 531769 non-null  float64
 13  GEO_LAT                 531769 non-null  float64
 14  DISTRICT_ID             536354 non-null  float64
 15  PRECINCT_ID             536354 non-null  float64
 16  NEIGHBORHOOD_ID         536354 non-null  object
 17  IS_CRIME                536355 non-null  int64
 18  IS_TRAFFIC              536355 non-null  int64
dtypes: float64(6), int64(6), object(7)
memory usage: 77.7+ MB
```

In [5]:

```
# Calculating descriptive statistics

print("Descriptive Statistics For Denver Crime:\n")
df_denver_crime.describe()
```

```
Descriptive Statistics For Denver Crime:
```

Out[5]:

	INCIDENT_ID	OFFENSE_ID	OFFENSE_CODE	OFFENSE_CODE_EXTENSION	GEO_X	GE
count	5.363550e+05	5.363550e+05	536355.000000	536355.000000	5.317700e+05	5.317700e
mean	4.435913e+09	4.435913e+15	3647.774797	0.208929	3.163554e+06	1.695422e
std	1.376261e+10	1.376261e+16	1635.685588	0.578281	6.629676e+05	2.154744e
min	2.019600e+04	2.019652e+10	902.000000	0.000000	1.000000e+00	1.000000e
25%	2.017176e+09	2.017176e+15	2305.000000	0.000000	3.139137e+06	1.683189e
50%	2.018805e+09	2.018805e+15	2999.000000	0.000000	3.145961e+06	1.694794e
75%	2.020748e+09	2.020748e+15	5441.000000	0.000000	3.163800e+06	1.702120e
max	2.020890e+12	2.020890e+18	7399.000000	5.000000	4.067477e+07	1.046707e

In [6]:

```
# Calculating statistics for categorical columns

print("Categorical Statistics For Denver Crime:\n")
df_denver_crime.describe(include=['object'])
```

```
Categorical Statistics For Denver Crime:
```

Out[6]:

	OFFENSE_TYPE_ID	OFFENSE_CATEGORY_ID	FIRST_OCCURRENCE_DATE	LAST_OCCURRENCE_DATE
count	536355	536355	536355	191519
unique	200	15	356935	137003
top	traffic-accident	traffic-accident	1/1/2017 12:00:00 AM	5/29/2020 7:00:00 AM
freq	84847	126228	24	18

In [7]:
```
# Calculating statistics for all columns

print("All Statistics For Denver Crime:\n")
df_denver_crime.describe(include='all')
```

```
All Statistics For Denver Crime:
```

Out[7]:

	INCIDENT_ID	OFFENSE_ID	OFFENSE_CODE	OFFENSE_CODE_EXTENSION	OFFENSE_TYPE_ID	OFF
count	5.363550e+05	5.363550e+05	536355.000000	536355.000000	536355	
unique	NaN	NaN	NaN	NaN	200	
top	NaN	NaN	NaN	NaN	traffic-accident	
freq	NaN	NaN	NaN	NaN	84847	
mean	4.435913e+09	4.435913e+15	3647.774797	0.208929	NaN	
std	1.376261e+10	1.376261e+16	1635.685588	0.578281	NaN	
min	2.019600e+04	2.019652e+10	902.000000	0.000000	NaN	
25%	2.017176e+09	2.017176e+15	2305.000000	0.000000	NaN	
50%	2.018805e+09	2.018805e+15	2999.000000	0.000000	NaN	
75%	2.020748e+09	2.020748e+15	5441.000000	0.000000	NaN	
max	2.020890e+12	2.020890e+18	7399.000000	5.000000	NaN	

In [ ]:
```

```

Data Exploration

In [8]:
```
# Displaying all offense types

offense_types = df_denver_crime['OFFENSE_TYPE_ID'].sort_values().unique()
df_offense_types = pd.DataFrame(offense_types)
df_offense_types.rename(columns = {0:'Offense Types'}, inplace = True)
df_offense_types
```

Out[8]:

	Offense Types
0	accessory-conspiracy-to-crime
1	agg-aslt-police-weapon
2	aggravated-assault
3	aggravated-assault-dv
4	altering-vin-number
...	...

	Offense Types
195	weapon-poss-illegal-dangerous
196	weapon-unlawful-discharge-of
197	weapon-unlawful-sale
198	window-peeping
199	wiretapping

200 rows × 1 columns

In [9]:
```
# Displaying all offense types with count

offense_count = df_denver_crime.groupby(['OFFENSE_TYPE_ID'])['INCIDENT_ID'].nunique()
df_offense_count =  pd.DataFrame(offense_count)
df_offense_count.to_string()
```

Out[9]:
```
'                                  INCIDENT_ID\nOFFENSE_TYPE_ID
\naccessory-conspiracy-to-crime          118\nagg-aslt-police-weapon                3
90\naggravated-assault                  6671\naggravated-assault-dv
2932\naltering-vin-number                    2\nanimal-cruelty-to
168\nanimal-poss-of-dangerous               12\narson-business
97\narson-other                            280\narson-public-building
19\narson-residence                        151\narson-vehicle
213\naslt-agg-police-gun                    13\nassault-dv
8527\nassault-police-simple                826\nassault-simple
15300\nbigamy                                 1\nbomb-threat
183\nbribery                               38\nburg-auto-theft-busn-no-force
88\nburg-auto-theft-busn-w-force           228\nburg-auto-theft-resd-no-force
631\nburg-auto-theft-resd-w-force          189\nburglary-business-by-force
6753\nburglary-business-no-force           2203\nburglary-poss-of-tools
525\nburglary-residence-by-force           7119\nburglary-residence-no-force
9788\nburglary-safe                         110\nburglary-vending-machine
161\ncontraband-into-prison                388\ncontraband-possession
43\ncriminal-mischief-graffiti             2823\ncriminal-mischief-mtr-veh             1
9421\ncriminal-mischief-other               15923\ncriminal-trespassing
16993\ncurfew                               1203\ndisarming-a-peace-officer
38\ndisturbing-the-peace                  6925\ndrug-barbiturate-mfr
1\ndrug-barbiturate-possess                 25\ndrug-barbiturate-sell
12\ndrug-cocaine-possess                   2504\ndrug-cocaine-sell
1476\ndrug-forgery-to-obtain                 77\ndrug-fraud-to-obtain
150\ndrug-hallucinogen-mfr                   8\ndrug-hallucinogen-possess
141\ndrug-hallucinogen-sell                 67\ndrug-heroin-possess
2375\ndrug-heroin-sell                      633\ndrug-make-sell-other-drug
112\ndrug-marijuana-cultivation            220\ndrug-marijuana-possess
867\ndrug-marijuana-sell                   331\ndrug-methampetamine-possess
5830\ndrug-methampetamine-sell             1280\ndrug-methamphetamine-mfr
44\ndrug-opium-or-deriv-possess            168\ndrug-opium-or-deriv-sell
153\ndrug-pcs-other-drug                   1535\ndrug-poss-paraphernalia
4164\ndrug-synth-narcotic-possess           103\ndrug-synth-narcotic-sell
156\neavesdropping                           3\nescape
107\nescape-aiding                           3\nexplosive-incendiary-dev-pos
44\nexplosive-incendiary-dev-use            20\nexplosives-posses
1\nextortion                               175\nfailure-to-report-abuse
2\nfalse-imprisonment                      553\nfireworks-possession
172\nforgery-checks                         1041\nforgery-counterfeit-of-obj
```

```
   92\nforgery-other                          351\nforgery-poss-of-forged-ftd
   37\nforgery-poss-of-forged-inst            137\nforgery-posses-forge-device
   48\nfraud-by-telephone                    1056\nfraud-by-use-of-computer
 1229\nfraud-criminal-impersonation           647\nfraud-identity-theft
  482\nfraud-nsf-closed-account               160\ngambling-betting-wagering
    2\ngambling-device                          3\ngambling-gaming-operation
    9\nharassment                            1350\nharassment-dv
  531\nharassment-sexual-in-nature            440\nharassment-stalking-dv
  327\nhomicide-conspiracy                      2\nhomicide-family
   36\nhomicide-negligent                       1\nhomicide-other
  350\nhomicide-police-by-gun                  4\nillegal-dumping
  149\nimpersonation-of-police                39\nindecent-exposure
  872\nintimidation-of-a-witness              154\nkidnap-adult-victim
  235\nkidnap-dv                              228\nliquor-manufacturing
    2\nliquor-misrepresent-age-minor            2\nliquor-other-viol
    8\nliquor-possession                     4914\nliquor-sell
  497\nlittering                              114\nloitering
    7\nmenacing-felony-w-weap                4282\nmoney-laundering
    3\nobscene-material-mfr                    5\nobscene-material-possess
   39\nobstructing-govt-operation            152\nother-enviornment-animal-viol
  182\nparole-violation                        12\npawn-broker-viol
   63\npolice-disobey-lawful-order           425\npolice-false-information
 2926\npolice-interference                    1716\npolice-making-a-false-rpt
   51\npolice-obstruct-investigation          42\npolice-resisting-arrest
  740\nprobation-violation                     16\nproperty-crimes-other
  215\nprostitution-aiding                     6\nprostitution-engaging-in
  959\nprostitution-pimping                   17\npublic-fighting
  766\npublic-order-crimes-other             5807\npublic-peace-other
  788\nreckless-endangerment                 153\nriot
    1\nriot-incite                             2\nrobbery-bank
  190\nrobbery-business                      1870\nrobbery-car-jacking
  856\nrobbery-purse-snatch-w-force           285\nrobbery-residence
  437\nrobbery-street                        3482\nsex-aslt-fondle-adult-victim
  866\nsex-aslt-non-rape                      694\nsex-aslt-non-rape-pot
  172\nsex-aslt-rape                         2570\nsex-aslt-rape-pot
  190\nsex-aslt-w-object                       44\nsex-aslt-w-object-pot
   13\nsex-asslt-sodomy-man-strng-arm         36\nsex-off-fail-to-register
 1597\nsex-off-registration-viol              248\nstolen-property-buy-sell-rec
  219\ntheft-bicycle                        10179\ntheft-confidence-game
   30\ntheft-embezzle                        145\ntheft-fail-return-rent-veh
  655\ntheft-from-bldg                       6929\ntheft-from-mails
  576\ntheft-from-yards                         1\ntheft-gas-drive-off
   32\ntheft-items-from-vehicle             34920\ntheft-of-cable-services
    1\ntheft-of-motor-vehicle               39988\ntheft-of-rental-property
   53\ntheft-of-services                      805\ntheft-other
22459\ntheft-parts-from-vehicle              17722\ntheft-pick-pocket
  178\ntheft-purse-snatch-no-force           329\ntheft-shoplift
14453\ntheft-stln-veh-const-eqpt                53\ntheft-stln-vehicle-trailer
  816\ntheft-unauth-use-of-ftd               857\nthreats-to-injure
 5690\ntraf-habitual-offender               3336\ntraf-impound-vehicle
   11\ntraf-other                           26119\ntraf-vehicular-assault
  272\ntraf-vehicular-homicide                28\ntraffic-accident
84847\ntraffic-accident-dui-duid             3304\ntraffic-accident-hit-and-run
38077\nvehicular-eluding                     348\nvehicular-eluding-no-chase
 5214\nviolation-of-court-order              2787\nviolation-of-custody-order
   51\nviolation-of-restraining-order        3826\nweapon-altering-serial-number
   26\nweapon-by-prev-offender-powpo         1972\nweapon-carrying-concealed
  496\nweapon-carrying-prohibited            400\nweapon-fire-into-occ-bldg
  916\nweapon-fire-into-occ-veh              249\nweapon-flourishing
  384\nweapon-other-viol                     783\nweapon-poss-illegal-dangerous
```

```
798\nweapon-unlawful-discharge-of          5424\nweapon-unlawful-sale
4\nwindow-peeping                         106\nwiretapping
5'
```

In [10]:
```
# Count of each offense type in descending order

df_offense_count.sort_values(by=['INCIDENT_ID'], ascending=False)
```

Out[10]:

OFFENSE_TYPE_ID	INCIDENT_ID
traffic-accident	84847
theft-of-motor-vehicle	39988
traffic-accident-hit-and-run	38077
theft-items-from-vehicle	34920
traf-other	26119
...	...
bigamy	1
homicide-negligent	1
theft-from-yards	1
theft-of-cable-services	1
drug-barbiturate-mfr	1

200 rows × 1 columns

In [11]:
```
# Maximum offense type count from dataframe

print(max(df_offense_count['INCIDENT_ID']))


# Maximum offense type count from arrays

print(max(offense_count))
```

```
84847
84847
```

In [12]:
```
# Offense type count

offense_count = df_denver_crime.groupby(['OFFENSE_TYPE_ID'])['INCIDENT_ID'].aggregate([
df_offense_count =  pd.DataFrame(offense_count)
df_offense_count
```

Out[12]:

OFFENSE_TYPE_ID	count	min	max
accessory-conspiracy-to-crime	118	20217439	20218026296

OFFENSE_TYPE_ID	count	min	max
agg-aslt-police-weapon	390	20169656	20215003725
aggravated-assault	6671	201842	20218038293
aggravated-assault-dv	2932	2017320	20218030801
altering-vin-number	2	2018791389	2020529669
...	...	...	...
weapon-poss-illegal-dangerous	798	2021168	20215004275
weapon-unlawful-discharge-of	5424	20196	20215003394
weapon-unlawful-sale	4	2017685629	20195005045
window-peeping	106	201616483	20215003405
wiretapping	5	20183579	2019281208

200 rows × 3 columns

In [13]:
```
# Offense type count agrregate function

offense_count = df_denver_crime.groupby(['OFFENSE_TYPE_ID']).aggregate({'INCIDENT_ID':'
df_offense_count =  pd.DataFrame(offense_count)
df_offense_count
```

Out[13]:

OFFENSE_TYPE_ID	INCIDENT_ID
accessory-conspiracy-to-crime	118
agg-aslt-police-weapon	390
aggravated-assault	6671
aggravated-assault-dv	2932
altering-vin-number	2
...	...
weapon-poss-illegal-dangerous	798
weapon-unlawful-discharge-of	5424
weapon-unlawful-sale	4
window-peeping	106
wiretapping	5

200 rows × 1 columns

In [14]:
```
# Creating the traffic vs crime count matrix
```

```
traffic_crime = df_denver_crime.groupby(['IS_TRAFFIC', 'IS_CRIME'])['INCIDENT_ID'].coun
traffic_crime
```

Out[14]:
```
IS_TRAFFIC  IS_CRIME
0           1           409827
1           0           126228
            1              300
Name: INCIDENT_ID, dtype: int64
```

In [ ]:

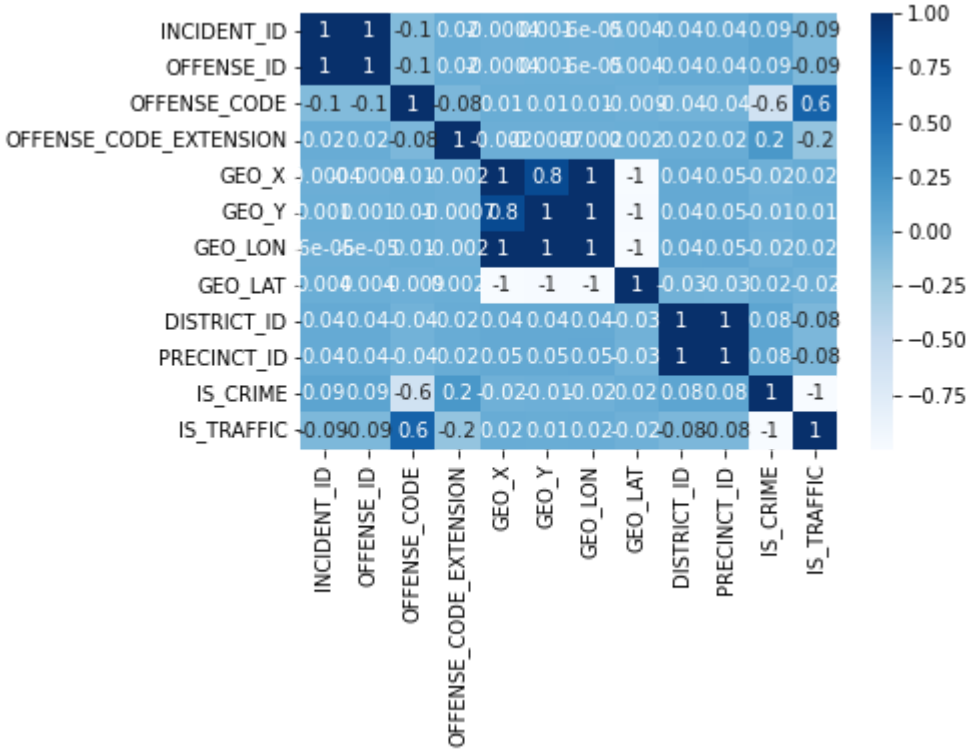
Visualization

In [15]:
```
# Correlation heatmap of the entire dataset

"""
https://seaborn.pydata.org/generated/seaborn.heatmap.html

seaborn.heatmap(data, *, vmin=None, vmax=None, cmap=None, center=None, robust=False, an
                linewidths=0, linecolor='white', cbar=True, cbar_kws=None, cbar_ax=None
                yticklabels='auto', mask=None, ax=None, **kwargs)
"""

heatmap = sns.heatmap(df_denver_crime.corr(), annot=True, cmap="Blues", fmt='.1g')
```



In [ ]:

Hypotheses

Hypothesis - The highest reported incidents are traffic related incidents.

Data

In [16]:
```
# To test this hypothesis, we can calculate the count of incidents in descending order.

df_offense_count.sort_values(by=['INCIDENT_ID'], ascending=False).head()
```

Out[16]:

OFFENSE_TYPE_ID	INCIDENT_ID
traffic-accident	84847
theft-of-motor-vehicle	39988
traffic-accident-hit-and-run	38077
theft-items-from-vehicle	34920
traf-other	26119

Conclusion - Since maximum count of incidents are related to traffic, we accept the above stated hypothesis.

Hypothesis - There are more incidents related to traffic than crime.

Data

In [17]:
```
# To test this hypothesis, we can calculate the count of crime and the count of traffic

pd.DataFrame(traffic_crime)
```

Out[17]:

IS_TRAFFIC	IS_CRIME	INCIDENT_ID
0	**1**	409827
1	**0**	126228
	1	300

Conclusion - Since, the count of crimes reported are more than the traffic incidents, we reject the above stated hypothesis.

In [ ]:
```

```