

\setcounter{secnumdepth}{0}

Stats 21 - HW 6 - Due 5/20/2023 by 11:59PM

Blair Lee - 005721089

Homework is generally an opportunity to practice coding and to train your problem solving and critical thinking skills. Putting Python to use is where learning happens.

Copying and pasting another's solutions takes away your learning opportunities. It is also academic dishonesty.

ChatGPT is always allowed in this class, but do remember, it is not foolproof and if your solution looks too much like another submission, I am required to file a report

Please use this document as your homework template and submit both the modified .ipynb file and a PDF export.

Introduction:

- Let's look at some large data on Crime
- This is crime data from the LA County Sheriff from 2019 and 2022.

```
In [ ]: ### Import the necessary libraries
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [ ]: ### Import the datasets 2019-PART_I_AND_II_CRIMES.csv
### Assign to a df called crime19 and use the 'LURN_SAK' as index
crime19 = pd.read_csv("2019-PART_I_AND_II_CRIMES.csv",
                      index_col='LURN_SAK',
                      dtype={'ZIP': str})

### create a date variable
crime19['date'] = pd.to_datetime(crime19['INCIDENT_DATE'],
                                format='%m/%d/%Y %I:%M:%S %p')

crime19.dtypes
```

```
Out[ ]: INCIDENT_DATE      object
INCIDENT_REPORTED_DATE    object
CATEGORY                  object
STAT                      int64
STAT_DESC                 object
ADDRESS                   object
STREET                    object
CITY                      object
ZIP                       object
INCIDENT_ID               object
REPORTING_DISTRICT        int64
SEQ                       int64
GANG_RELATED              object
UNIT_ID                   object
UNIT_NAME                 object
LONGITUDE                 float64
LATITUDE                  float64
PART_CATEGORY             int64
date                      datetime64[ns]
dtype: object
```

```
In [ ]: ### Import the datasets 2022-PART_I_AND_II_CRIMES.csv
### Assign to a df called crime22 and use the 'LURN_SAK' as index
crime22 = pd.read_csv("2022-PART_I_AND_II_CRIMES.csv",
                      index_col='LURN_SAK',
                      dtype={'ZIP': str})

### create a date variable
crime22['date'] = pd.to_datetime(crime22['INCIDENT_DATE'],
                                format='%m/%d/%Y %I:%M:%S %p')

crime22.dtypes
```

```
Out[ ]: INCIDENT_DATE      object
INCIDENT_REPORTED_DATE    object
CATEGORY                  object
STAT                      int64
STAT_DESC                 object
ADDRESS                   object
STREET                    object
CITY                      object
ZIP                       object
INCIDENT_ID               object
REPORTING_DISTRICT        int64
SEQ                       int64
GANG_RELATED              object
UNIT_ID                   object
UNIT_NAME                 object
LONGITUDE                 float64
LATITUDE                  float64
PART_CATEGORY             int64
date                      datetime64[ns]
dtype: object
```

```
In [ ]: # Problem 1. Please display the first 10 of the 2019 entries and the last 10
# Please sort the dataframes on 'date' before displaying the values.
crime19_sorted = crime19.sort_values(by = "date")
crime19_sorted.head(10)
```

Out[]:

	INCIDENT_DATE	INCIDENT_REPORTED_DATE	CATEGORY	STAT	STAT_DES
LURN_SAK					
18820256	01/01/1976 12:00:00 PM	01/25/2019	SEX OFFENSES FELONIES	120	S FELONIE Sex Crim Agair Childr (Unc
19040544	01/01/1976 12:05:45 PM	09/03/2019	SEX OFFENSES FELONIES	120	S FELONIE Sex Crim Agair Childr (Unc
18860114	01/01/1979 12:00:00 AM	03/08/2019	FORCIBLE RAPE	23	RAPE I FORC FEMA UNDER
18931381	07/02/1979 11:59:30 AM	05/12/2019	SEX OFFENSES FELONIES	120	S FELONIE Sex Crim Agair Childr (Unc
18899023	07/02/1979 12:00:00 PM	04/18/2019	FORCIBLE RAPE	27	RAPE I FORC MA UNDER
19062434	01/16/1980 03:16:07 PM	09/26/2019	SEX OFFENSES FELONIES	120	S FELONIE Sex Crim Agair Childr (Unc
18880046	07/02/1981 12:00:00 PM	03/29/2019	FORCIBLE RAPE	21	RAF FORCIBL Rape Forc Female YRS
19045032	08/16/1982 12:00:00 PM	09/07/2019	SEX OFFENSES FELONIES	120	S FELONIE Sex Crim Agair Childr (Unc
19141266	07/02/1985 11:59:30 AM	12/20/2019	SEX OFFENSES FELONIES	129	S FELONIE

	INCIDENT_DATE	INCIDENT_REPORTED_DATE	CATEGORY	STAT	STAT_DES
LURN_SAK					
					All Oth Sex Feloni
					S
			SEX		FELONIE
18859835	03/05/1987 12:00:00 PM	03/08/2019	OFFENSES FELONIES	120	Sex Crim Agair Childr

```
In [ ]: crime22_sorted = crime22.sort_values(by = "date")
        crime22_sorted.tail(10)
```

Out[]:

	INCIDENT_DATE	INCIDENT_REPORTED_DATE	CATEGORY	STAT	LURN_SAK
20110647	01/03/2023 02:00:00 PM	12/23/2022	GRAND THEFT AUTO	91	G VE Automobil
20100452	01/03/2023 02:23:00 PM	12/11/2022	GRAND THEFT AUTO	91	G VE Automobil
20116340	01/04/2023 11:20:00 AM	12/31/2022	GRAND THEFT AUTO	91	G VE Automobil
20116210	01/04/2023 10:17:00 PM	12/30/2022	GRAND THEFT AUTO	91	G VE Automobil
20114970	01/06/2023 07:30:00 PM	12/29/2022	GRAND THEFT AUTO	91	G VE Automobil
20116374	01/17/2023 02:00:00 PM	12/31/2022	GRAND THEFT AUTO	95	G VE
20100104	01/20/2023 05:30:00 PM	12/10/2022	GRAND THEFT AUTO	91	G VE Automobil
19967428	01/21/2023 02:30:00 AM	07/15/2022	SEX OFFENSES FELONIES	120	SEX FI C Ch
20046199	01/30/2023 04:00:00 PM	10/07/2022	GRAND THEFT AUTO	91	G VE Automobil
20105582	02/05/2023 04:30:00 PM	12/17/2022	GRAND THEFT AUTO	91	G VE Automobil

```

INCIDENT_DATE INCIDENT_REPORTED_DATE CATEGORY STAT
LURN_SAK

```

```

In [ ]: # Problem 2. How many rows and columns are in 2019
# and how many are in 2022?
print(crime19.shape)
print(crime22.shape)

```

```

(163438, 19)
(157248, 19)

```

```

In [ ]: # Problem 3. What crime category was number #1 (most frequently committed) i
# and what was it in 2022?)
crime19["CATEGORY"].value_counts()
# The most frequently committed crime in 2019 was larceny theft.

```

```

Out[ ]: LARCENY THEFT                31260
VEHICLE / BOATING LAWS             20506
NARCOTICS                          18364
NON-AGGRAVATED ASSAULTS            14213
VANDALISM                          10178
GRAND THEFT AUTO                   9929
BURGLARY                           9741
MISDEMEANORS MISCELLANEOUS         9579
AGGRAVATED ASSAULT                 6732
FRAUD AND NSF CHECKS               6716
WEAPON LAWS                        4344
ROBBERY                            4278
FELONIES MISCELLANEOUS             2291
DRUNK DRIVING VEHICLE / BOAT       2244
DRUNK / ALCOHOL / DRUGS            1920
FORGERY                            1778
SEX OFFENSES MISDEMEANORS          1708
OFFENSES AGAINST FAMILY            1462
LIQUOR LAWS                       1457
SEX OFFENSES FELONIES              1283
WARRANTS                           686
FORCIBLE RAPE                      581
DISORDERLY CONDUCT                 558
FEDERAL OFFENSES WITH MONEY        487
ARSON                              482
VAGRANCY                           321
RECEIVING STOLEN PROPERTY          186
CRIMINAL HOMICIDE                  120
GAMBLING                           27
FEDERAL OFFENSES W/O MONEY         7
Name: CATEGORY, dtype: int64

```

```

In [ ]: crime22["CATEGORY"].value_counts()
# The most committed crime in 2022 was larceny theft.

```

```
Out[ ]: LARCENY THEFT 32542
        GRAND THEFT AUTO 15701
        VEHICLE / BOATING LAWS 14750
        NON-AGGRAVATED ASSAULTS 13894
        NARCOTICS 11900
        BURGLARY 10473
        VANDALISM 10430
        FRAUD AND NSF CHECKS 10260
        AGGRAVATED ASSAULT 8299
        MISDEMEANORS MISCELLANEOUS 7213
        WEAPON LAWS 4758
        ROBBERY 4000
        FELONIES MISCELLANEOUS 2227
        OFFENSES AGAINST FAMILY 1906
        DRUNK DRIVING VEHICLE / BOAT 1689
        FORGERY 1260
        DRUNK / ALCOHOL / DRUGS 1063
        SEX OFFENSES FELONIES 995
        SEX OFFENSES MISDEMEANORS 904
        LIQUOR LAWS 557
        ARSON 549
        WARRANTS 526
        DISORDERLY CONDUCT 394
        FORCIBLE RAPE 388
        CRIMINAL HOMICIDE 182
        RECEIVING STOLEN PROPERTY 151
        FEDERAL OFFENSES WITH MONEY 139
        VAGRANCY 66
        GAMBLING 31
        FEDERAL OFFENSES W/O MONEY 1
        Name: CATEGORY, dtype: int64
```

```
In [ ]: # Problem 4. What the most frequent incident hour in both 2019 and 2022?
        # Also find the 3 hours in the day with the lowest number of crimes (safest)
        # in both 2019 and 2022
        crime19["Hour"] = crime19["date"].dt.round("H").dt.hour
        crime22["Hour"] = crime22["date"].dt.round("H").dt.hour
        crime19["Hour"].value_counts()
        # The most frequent incident hour in 2019 is at 12 PM.
```



```
Out[ ]: 12    11286
        0     10985
        16     9397
        18     9374
         2     8760
        20     8661
        14     8345
        17     7497
        22     7447
        10     7150
        19     7026
        15     6899
        13     6225
        21     6210
         1     6064
        11     5896
        23     5850
         8     5491
         4     5480
         3     5305
         9     4807
         6     3526
         7     2965
         5     2792
        Name: Hour, dtype: int64
```

```
In [ ]: crime22["Hour"].value_counts()
        # The most frequent incident hour at 2022 is 12 AM.
        # The three times with the least crime for both 2019 and 2022 are 5, 6, and
```

```
Out[ ]: 0     11416
        12    11193
         2     8537
        16     8530
        18     8361
        14     7864
        20     7800
        22     6928
        10     6888
        17     6784
        15     6243
        19     6087
         1     6033
        13     5928
         4     5867
        11     5559
         8     5553
        21     5469
        23     5350
         3     5210
         9     4422
         6     4352
         7     3575
         5     3299
        Name: Hour, dtype: int64
```

```
In [ ]: # Problem 5. Find the proportion (or percentage) of gang-related crimes by C
# and sort the values from the highest gang-related to the lowest
# for both 2019 and 2022
gang_related19 = crime19.groupby(by = "CATEGORY")["GANG_RELATED"].value_cour
gang_related19.sort_values(ascending = False)
```

```
Out[ ]: CATEGORY
CRIMINAL HOMICIDE          0.550000
WEAPON LAWS                0.130525
AGGRAVATED ASSAULT         0.087493
ROBBERY                    0.033427
RECEIVING STOLEN PROPERTY  0.026882
VANDALISM                  0.020927
FELONIES MISCELLANEOUS     0.014841
NARCOTICS                  0.009638
WARRANTS                   0.008746
FORGERY                    0.005062
NON-AGGRAVATED ASSAULTS    0.004292
OFFENSES AGAINST FAMILY    0.004104
GRAND THEFT AUTO           0.003928
MISDEMEANORS MISCELLANEOUS 0.003549
LIQUOR LAWS                0.003432
DRUNK / ALCOHOL / DRUGS    0.002083
DISORDERLY CONDUCT         0.001792
BURGLARY                   0.001745
FORCIBLE RAPE              0.001721
VEHICLE / BOATING LAWS     0.001414
FRAUD AND NSF CHECKS       0.001191
DRUNK DRIVING VEHICLE / BOAT 0.000891
LARCENY THEFT              0.000640
SEX OFFENSES MISDEMEANORS  0.000585
ARSON                      0.000000
GAMBLING                   0.000000
FEDERAL OFFENSES WITH MONEY 0.000000
SEX OFFENSES FELONIES      0.000000
VAGRANCY                   0.000000
FEDERAL OFFENSES W/O MONEY  0.000000
Name: YES, dtype: float64
```

```
In [ ]: gang_related22 = crime22.groupby(by = "CATEGORY")["GANG_RELATED"].value_cour
gang_related22.sort_values(ascending = False)
```

```
Out[ ]: CATEGORY
CRIMINAL HOMICIDE 0.587912
WEAPON LAWS 0.072509
AGGRAVATED ASSAULT 0.069406
GAMBLING 0.032258
ROBBERY 0.013250
FELONIES MISCELLANEOUS 0.007185
VANDALISM 0.006903
RECEIVING STOLEN PROPERTY 0.006623
NARCOTICS 0.003025
WARRANTS 0.001901
MISDEMEANORS MISCELLANEOUS 0.001802
NON-AGGRAVATED ASSAULTS 0.001439
GRAND THEFT AUTO 0.001401
SEX OFFENSES MISDEMEANORS 0.001106
VEHICLE / BOATING LAWS 0.000949
DRUNK / ALCOHOL / DRUGS 0.000941
FORGERY 0.000794
BURGLARY 0.000382
LARCENY THEFT 0.000246
FRAUD AND NSF CHECKS 0.000195
DISORDERLY CONDUCT 0.000000
OFFENSES AGAINST FAMILY 0.000000
ARSON 0.000000
SEX OFFENSES FELONIES 0.000000
FORCIBLE RAPE 0.000000
VAGRANCY 0.000000
FEDERAL OFFENSES WITH MONEY 0.000000
FEDERAL OFFENSES W/O MONEY 0.000000
DRUNK DRIVING VEHICLE / BOAT 0.000000
LIQUOR LAWS 0.000000
Name: YES, dtype: float64
```