Homework06

May 19, 2023

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

Haojie Liu 905943123

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.

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


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

```
[3]:  ### 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
```

```
[3]:  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
```

```python
[4]:  ### 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
```

```
[4]:  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
```

[5]: *# Problem 1. Please display the first 10 of the 2019 entries and the last 10*
    ↪entries of 2022.
    # Please sort the dataframes on 'date' before displaying the values.

    crime19.sort_values(by=['date']).iloc[0:10]

[5]:   INCIDENT_DATE INCIDENT_REPORTED_DATE
    LURN_SAK
    18820256 01/01/1976 12:00:00 PM   01/25/2019 \
    19040544 01/01/1976 12:05:45 PM   09/03/2019
    18860114 01/01/1979 12:00:00 AM   03/08/2019
    18931381 07/02/1979 11:59:30 AM   05/12/2019
    18899023 07/02/1979 12:00:00 PM   04/18/2019
    19062434 01/16/1980 03:16:07 PM   09/26/2019
    18880046 07/02/1981 12:00:00 PM   03/29/2019
    19045032 08/16/1982 12:00:00 PM   09/07/2019
    19141266 07/02/1985 11:59:30 AM   12/20/2019
    18859835 03/05/1987 12:00:00 PM   03/08/2019

                         CATEGORY STAT
    LURN_SAK
    18820256 SEX OFFENSES FELONIES   120 \
    19040544 SEX OFFENSES FELONIES   120
    18860114   FORCIBLE RAPE   23
    18931381 SEX OFFENSES FELONIES   120
    18899023   FORCIBLE RAPE   27
    19062434 SEX OFFENSES FELONIES   120
    18880046   FORCIBLE RAPE   21
    19045032 SEX OFFENSES FELONIES   120
    19141266 SEX OFFENSES FELONIES   129
    18859835 SEX OFFENSES FELONIES   120

                                              STAT_DESC
    LURN_SAK
    18820256 SEX FELONIES: Sex Crimes Against Children (Und… \
    19040544 SEX FELONIES: Sex Crimes Against Children (Und…
    18860114   RAPE BY FORCE, FEMALE UNDER 18
    18931381 SEX FELONIES: Sex Crimes Against Children (Und…
    18899023   RAPE BY FORCE, MALE UNDER 18
```

```
19062434  SEX FELONIES: Sex Crimes Against Children (Und…
18880046     RAPE, FORCIBLE: Rape by Force, Female 18 YRS +
19045032  SEX FELONIES: Sex Crimes Against Children (Und…
19141266            SEX FELONIES: All Other Sex Felonies
18859835  SEX FELONIES: Sex Crimes Against Children (Und…


                                        ADDRESS            STREET
LURN_SAK
18820256       19800 LORENCITA DR, COVINA, CA  91724     19800 LORENCITA DR   \
19040544                         UNK, LAKEWOOD, CA                      UNK
18860114       26300 ESHELMAN AVE, LOMITA, CA  90717     26300 ESHELMAN AVE
18931381       28800 SELFRIDGE DR, MALIBU, CA  90265     28800 SELFRIDGE DR
18899023            5200 ELIZABETH ST, CUDAHY, CA         5200 ELIZABETH ST
19062434             CORNELL ST, SAN FERNANDO, CA            CORNELL ST
18880046       28700 MADISON ST, VAL VERDE, CA  91384     28700 MADISON ST
19045032  19300 BARROSO ST, ROWLAND HEIGHTS, CA  91748    19300 BARROSO ST
19141266    12200 LA MIRADA BLVD, LA MIRADA, CA  90638  12200 LA MIRADA BLVD
18859835    11200 BYWAYS ST, SOUTH EL MONTE, CA  91733     11200 BYWAYS ST


                    CITY    ZIP    INCIDENT_ID  REPORTING_DISTRICT   SEQ
LURN_SAK
18820256          COVINA  91724  019-00651-2941               2941   651  \
19040544        LAKEWOOD    NaN  019-19248-1322               1322  19248
18860114          LOMITA  90717  019-00806-1713               1713   806
18931381          MALIBU  90265  019-02523-1013               1013  2523
18899023          CUDAHY    NaN  019-05536-0262                262  5536
19062434    SAN FERNANDO    NaN  019-00067-2035               2035    67
18880046       VAL VERDE  91384  019-04677-0662                662  4677
19045032  ROWLAND HEIGHTS  91748  019-07026-2932               2932  7026
19141266        LA MIRADA  90638  019-17888-0463                463  17888
18859835  SOUTH EL MONTE  91733  019-02912-0563                563  2912


          GANG_RELATED    UNIT_ID                     UNIT_NAME  LONGITUDE
LURN_SAK
18820256            NO  CA0190020                        WALNUT   -117.865  \
19040544            NO  CA0190013                      LAKEWOOD        NaN
18860114            NO  CA0190017                        LOMITA   -118.314
18931381            NO  CA0190022               MALIBU/LOST HILLS  -118.799
18899023            NO  CA0190002               EAST LOS ANGELES  -118.173
19062434            NO  CA0190000  LOS ANGELES COUNTY SHERIFF DEPT      NaN
18880046            NO  CA0190006          SANTA CLARITA VALLEY   -118.656
19045032            NO  CA0190020                        WALNUT   -117.882
19141266            NO  CA0190004                       NORWALK   -118.011
18859835            NO  CA0190005                        TEMPLE   -118.040


          LATITUDE   PART_CATEGORY              date
LURN_SAK
```

```
18820256    34.064                      2 1976-01-01 12:00:00
19040544       NaN                      2 1976-01-01 12:05:45
18860114    33.783                      1 1979-01-01 00:00:00
18931381    34.019                      2 1979-07-02 11:59:30
18899023    33.962                      1 1979-07-02 12:00:00
19062434       NaN                      2 1980-01-16 15:16:07
18880046    34.442                      1 1981-07-02 12:00:00
19045032    33.978                      2 1982-08-16 12:00:00
19141266    33.918                      2 1985-07-02 11:59:30
18859835    34.038                      2 1987-03-05 12:00:00
```

[6]: `crime22.sort_values(by=['date'],ascending=False).iloc[0:10]`

```
[6]:                 INCIDENT_DATE INCIDENT_REPORTED_DATE
     LURN_SAK
     20105582  02/05/2023 04:30:00 PM             12/17/2022  \
     20046199  01/30/2023 04:00:00 PM             10/07/2022
     19967428  01/21/2023 02:30:00 AM             07/15/2022
     20100104  01/20/2023 05:30:00 PM             12/10/2022
     20116374  01/17/2023 02:00:00 PM             12/31/2022
     20114970  01/06/2023 07:30:00 PM             12/29/2022
     20116210  01/04/2023 10:17:00 PM             12/30/2022
     20116340  01/04/2023 11:20:00 AM             12/31/2022
     20100452  01/03/2023 02:23:00 PM             12/11/2022
     20110647  01/03/2023 02:00:00 PM             12/23/2022


                        CATEGORY  STAT
     LURN_SAK
     20105582      GRAND THEFT AUTO    91  \
     20046199      GRAND THEFT AUTO    91
     19967428  SEX OFFENSES FELONIES   120
     20100104      GRAND THEFT AUTO    91
     20116374      GRAND THEFT AUTO    95
     20114970      GRAND THEFT AUTO    91
     20116210      GRAND THEFT AUTO    91
     20116340      GRAND THEFT AUTO    91
     20100452      GRAND THEFT AUTO    91
     20110647      GRAND THEFT AUTO    91


                                          STAT_DESC
     LURN_SAK
     20105582  GRAND THEFT VEHICLE (GTA): Automobile/Passenge…  \
     20046199  GRAND THEFT VEHICLE (GTA): Automobile/Passenge…
     19967428  SEX FELONIES: Sex Crimes Against Children (Und…
     20100104  GRAND THEFT VEHICLE (GTA): Automobile/Passenge…
     20116374          GRAND THEFT VEHICLE (GTA): Motorcycle
     20114970  GRAND THEFT VEHICLE (GTA): Automobile/Passenge…
```

```
20116210  GRAND THEFT VEHICLE (GTA): Automobile/Passenge…
20116340  GRAND THEFT VEHICLE (GTA): Automobile/Passenge…
20100452  GRAND THEFT VEHICLE (GTA): Automobile/Passenge…
20110647  GRAND THEFT VEHICLE (GTA): Automobile/Passenge…


                                                   ADDRESS
LURN_SAK
20105582          10TH ST W AND AVENUE H-11, LANCASTER, CA  \
20046199                     43500 17TH ST W, LANCASTER, CA
19967428            14300 MULBERRY DR, WHITTIER, CA  90604
20100104               400 W AVENUE S, PALMDALE, CA  93551
20116374         11400 LONG BEACH BLVD, LYNWOOD, CA  90262
20114970              3100 PAXTON AVE, PALMDALE, CA  93551
20116210             13000 LAMBERT RD, WHITTIER, CA  90606
20116340               900 W 132ND ST, COMPTON, CA  90220
20100452       MEYER RD AND HASTINGS DR, WHITTIER, CA  90604
20110647  CARSON ST AND NORWALK BLVD, HAWAIIAN GARDENS, …


                             STREET              CITY   ZIP    INCIDENT_ID
LURN_SAK
20105582   10TH ST W AND AVENUE H-11         LANCASTER   NaN  922-21424-1126  \
20046199              43500 17TH ST W        LANCASTER   NaN  922-17382-1132
19967428            14300 MULBERRY DR         WHITTIER  90604  022-08210-0491
20100104               400 W AVENUE S         PALMDALE  93551  922-16674-2602
20116374         11400 LONG BEACH BLVD         LYNWOOD  90262  922-19474-2112
20114970              3100 PAXTON AVE         PALMDALE  93551  922-17591-2612
20116210             13000 LAMBERT RD         WHITTIER  90606  922-07984-1513
20116340               900 W 132ND ST          COMPTON  90220  922-16790-2812
20100452     MEYER RD AND HASTINGS DR         WHITTIER  90604  022-15310-0494
20110647  CARSON ST AND NORWALK BLVD  HAWAIIAN GARDENS  90716  922-25063-1321


          REPORTING_DISTRICT    SEQ GANG_RELATED    UNIT_ID    UNIT_NAME
LURN_SAK
20105582                1126  21424           NO  CA0190024    LANCASTER  \
20046199                1132  17382           NO  CA0190024    LANCASTER
19967428                 491   8210           NO  CA0190004      NORWALK
20100104                2602  16674           NO  CA01900W9     PALMDALE
20116374                2112  19474           NO  CA01900V3      CENTURY
20114970                2612  17591           NO  CA01900W9     PALMDALE
20116210                1513   7984           NO  CA0190015  PICO RIVERA
20116340                2812  16790           NO  CA0190042      COMPTON
20100452                 494  15310          YES  CA0190004      NORWALK
20110647                1321  25063           NO  CA0190013     LAKEWOOD


          LONGITUDE  LATITUDE  PART_CATEGORY                 date
LURN_SAK
20105582   -118.148    34.667              1 2023-02-05 16:30:00
```

```
20046199   -118.160    34.673              1 2023-01-30 16:00:00
19967428        NaN       NaN              2 2023-01-21 02:30:00
20100104   -118.138    34.557              1 2023-01-20 17:30:00
20116374   -118.211    33.926              1 2023-01-17 14:00:00
20114970   -118.186    34.598              1 2023-01-06 19:30:00
20116210   -118.041    33.960              1 2023-01-04 22:17:00
20116340   -118.241    33.912              1 2023-01-04 11:20:00
20100452   -118.040    33.921              1 2023-01-03 14:23:00
20110647   -118.071    33.831              1 2023-01-03 14:00:00
```

[7]:
```python
# Problem 2. How many rows and columns are in 2019
# and how many are in 2022?
print(crime19.shape[0])
print(crime22.shape[0])
```

```
163438
157248
```

[8]:
```python
# Problem 3. What crime category was number #1 (most frequently committed) in
  ↪2019
# and what was it in 2022?)

print(crime19['CATEGORY'].value_counts()[0])
print(crime22['CATEGORY'].value_counts()[0])
crime19['INCIDENT_DATE']
```

```
31260
32542
```

[8]:
```
LURN_SAK
18928343    05/09/2019 04:05:00 AM
18868348    03/17/2019 10:05:00 AM
18830941    02/05/2019 07:57:00 PM
18959726    06/09/2019 11:15:00 PM
18974981    06/25/2019 07:00:00 PM
                   ...
18855003    03/03/2019 10:22:00 AM
18886362    04/05/2019 04:15:00 AM
18845349    02/20/2019 01:15:00 PM
18990369    07/12/2019 01:30:00 AM
18889903    04/09/2019 08:30:00 AM
Name: INCIDENT_DATE, Length: 163438, dtype: object
```

[9]:
```python
# 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
```

```
print('most frequent:')
print('crime19: 12: ',crime19['date'].dt.hour.value_counts()[0])
print('crime22: 0: ', crime22['date'].dt.hour.value_counts()[0])

print('least: ')
print('crime19:',crime19['date'].dt.hour.value_counts()[21:24])
print('crime22:', crime22['date'].dt.hour.value_counts()[21:24])
```

```
most frequent:
crime19: 12:  9301
crime22: 0:  9812
least:
crime19: date
4    4013
5    3212
6    3110
Name: count, dtype: int64
crime22: date
7    4272
6    4103
5    3607
Name: count, dtype: int64
```

[13]:
```
# Problem 5. Find the proportion (or percentage) of gang-related crimes by␣
  ↪Category
# and sort the values from the highest gang-related to the lowest
# for both 2019 and 2022

ef = crime22[crime22['GANG_RELATED']=='YES']['CATEGORY'].value_counts()/
  ↪crime22['CATEGORY'].value_counts()
ef = pd.DataFrame(ef)

df = crime19[crime19['GANG_RELATED']=='YES']['CATEGORY'].value_counts()/
  ↪crime19['CATEGORY'].value_counts()
df = pd.DataFrame(df)

print(df.sort_values(by='count',ascending=False))
print(ef.sort_values(by='count',ascending=False))
```

```
                              count
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                               NaN
FEDERAL OFFENSES W/O MONEY          NaN
FEDERAL OFFENSES WITH MONEY         NaN
GAMBLING                            NaN
SEX OFFENSES FELONIES               NaN
VAGRANCY                            NaN
                                  count
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
ARSON                               NaN
DISORDERLY CONDUCT                  NaN
```

```
DRUNK DRIVING VEHICLE / BOAT      NaN
FEDERAL OFFENSES W/O MONEY        NaN
FEDERAL OFFENSES WITH MONEY       NaN
FORCIBLE RAPE                     NaN
LIQUOR LAWS                       NaN
OFFENSES AGAINST FAMILY           NaN
SEX OFFENSES FELONIES             NaN
VAGRANCY                          NaN
```