task

January 6, 2022

[30]:
```
import pandas as pd
import numpy as np
```

0.0.1 The BTP Police Service are looking to understand how their stop and search policies have changed over time

The latest data is available through the public Police Data API https://data.police.uk/docs/method/stops-force/

1. Write a script in Python to programmatically pull down all stop and search data from the API for the BTP Police Service.

[31]:
```
LINK = 'https://data.police.uk/api/crimes-street-dates'


def pull_data(keyword):

    force_data = pd.read_json(LINK)

    force_data = force_data.explode('stop-and-search')

    force_data['date'] = pd.to_datetime(force_data['date'])

    res = force_data[force_data['stop-and-search'] == keyword]

    res['month'] = res['date'].dt.to_period('M')

    return res
```

[32]:
```
pull_data('btp')
```

```
/var/folders/84/p81hrkps12z5qktxknfn0r6m0000gq/T/ipykernel_58546/485637084.py:14
: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
```

```
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  res['month'] = res['date'].dt.to_period('M')
```

[32]:

	date	stop-and-search	month
0	2021-11-01	btp	2021-11
1	2021-10-01	btp	2021-10
2	2021-09-01	btp	2021-09
3	2021-08-01	btp	2021-08
4	2021-07-01	btp	2021-07
5	2021-06-01	btp	2021-06
6	2021-05-01	btp	2021-05
7	2021-04-01	btp	2021-04
8	2021-03-01	btp	2021-03
9	2021-02-01	btp	2021-02
10	2021-01-01	btp	2021-01
11	2020-12-01	btp	2020-12
12	2020-11-01	btp	2020-11
13	2020-10-01	btp	2020-10
14	2020-09-01	btp	2020-09
15	2020-08-01	btp	2020-08
16	2020-07-01	btp	2020-07
17	2020-06-01	btp	2020-06
18	2020-05-01	btp	2020-05
19	2020-04-01	btp	2020-04
20	2020-03-01	btp	2020-03
21	2020-02-01	btp	2020-02
22	2020-01-01	btp	2020-01
23	2019-12-01	btp	2019-12
24	2019-11-01	btp	2019-11
25	2019-10-01	btp	2019-10
26	2019-09-01	btp	2019-09
27	2019-08-01	btp	2019-08
28	2019-07-01	btp	2019-07
29	2019-06-01	btp	2019-06
30	2019-05-01	btp	2019-05
31	2019-04-01	btp	2019-04
32	2019-03-01	btp	2019-03
33	2019-02-01	btp	2019-02
34	2019-01-01	btp	2019-01
35	2018-12-01	btp	2018-12

[33]:
```
PATTERN = 'https://data.police.uk/api/stops-force?force={}&date={}'

def crawl_by_date(keyword, date):
    return pd.read_json(PATTERN.format(keyword, date))
```

[34]:
```
crawl_by_date('btp', '2021-10')
```

[34]:

	age_range	outcome	involved_person	\
0	18-24	A no further action disposal	True	
1	18-24	A no further action disposal	True	
2	over 34	A no further action disposal	True	
3	18-24	A no further action disposal	True	
4	10-17	A no further action disposal	True	
..	…	…	…	
924	None	A no further action disposal	True	
925	None	A no further action disposal	True	
926	18-24	A no further action disposal	True	
927	10-17	Arrest	True	
928	10-17	A no further action disposal	True	

	self_defined_ethnicity	gender	\
0	Other ethnic group - Not stated	Male	
1	White - English/Welsh/Scottish/Northern Irish/…	Male	
2	White - English/Welsh/Scottish/Northern Irish/…	Male	
3	Black/African/Caribbean/Black British - African	Male	
4	White - English/Welsh/Scottish/Northern Irish/…	Male	
..	…	…	
924	Other ethnic group - Not stated	Male	
925	Other ethnic group - Not stated	Male	
926	Black/African/Caribbean/Black British - Any ot…	None	
927	White - English/Welsh/Scottish/Northern Irish/…	Male	
928	Other ethnic group - Not stated	Male	

	legislation	\
0	Police and Criminal Evidence Act 1984 (section 1)	
1	Misuse of Drugs Act 1971 (section 23)	
2	Police and Criminal Evidence Act 1984 (section 1)	
3	Misuse of Drugs Act 1971 (section 23)	
4	Police and Criminal Evidence Act 1984 (section 1)	
..	…	
924	Misuse of Drugs Act 1971 (section 23)	
925	Misuse of Drugs Act 1971 (section 23)	
926	Misuse of Drugs Act 1971 (section 23)	
927	Police and Criminal Evidence Act 1984 (section 1)	
928	Police and Criminal Evidence Act 1984 (section 1)	

	outcome_linked_to_object_of_search	datetime	\
0	True	2021-10-04 17:00:00+00:00	
1	False	2021-10-01 00:30:00+00:00	
2	False	2021-10-01 08:13:00+00:00	
3	False	2021-10-01 13:45:00+00:00	
4	False	2021-10-01 16:05:00+00:00	
..	…	…	
924	False	2021-10-29 19:45:00+00:00	

```
925                                False 2021-10-29 19:45:00+00:00
926                                False 2021-10-29 15:25:00+00:00
927                                False 2021-10-29 14:05:00+00:00
928                                False 2021-10-11 13:15:00+00:00

     removal_of_more_than_outer_clothing  \
0                                   False
1                                   False
2                                   False
3                                   False
4                                   False
..                                    …
924                                 False
925                                 False
926                                 False
927                                 False
928                                 False

                                          outcome_object  \
0    {'id': 'bu-no-further-action', 'name': 'A no f…
1    {'id': 'bu-no-further-action', 'name': 'A no f…
2    {'id': 'bu-no-further-action', 'name': 'A no f…
3    {'id': 'bu-no-further-action', 'name': 'A no f…
4    {'id': 'bu-no-further-action', 'name': 'A no f…
..                                                 …
924  {'id': 'bu-no-further-action', 'name': 'A no f…
925  {'id': 'bu-no-further-action', 'name': 'A no f…
926  {'id': 'bu-no-further-action', 'name': 'A no f…
927              {'id': 'bu-arrest', 'name': 'Arrest'}
928  {'id': 'bu-no-further-action', 'name': 'A no f…

                                             location  operation  \
0    {'latitude': '51.519060', 'street': {'id': 149…        NaN
1    {'latitude': '51.143738', 'street': {'id': 149…        NaN
2    {'latitude': '51.459100', 'street': {'id': 148…        NaN
3    {'latitude': '51.485400', 'street': {'id': 148…        NaN
4    {'latitude': '51.532500', 'street': {'id': 148…        NaN
..                                                 …        …
924  {'latitude': '51.528000', 'street': {'id': 148…        NaN
925  {'latitude': '51.528000', 'street': {'id': 148…        NaN
926  {'latitude': '51.528000', 'street': {'id': 148…        NaN
927  {'latitude': '53.975900', 'street': {'id': 148…        NaN
928  {'latitude': '51.496200', 'street': {'id': 148…        NaN

    officer_defined_ethnicity           type  operation_name  \
0                       Black  Person search             NaN
1                       White  Person search             NaN
```

```
2                       White  Person search             NaN
3                       Black  Person search             NaN
4                       White  Person search             NaN
..                        …            …              …
924                     Black  Person search             NaN
925                     Black  Person search             NaN
926                     Black  Person search             NaN
927                     White  Person search             NaN
928                     White  Person search             NaN

                      object_of_search
0                         Stolen goods
1                      Controlled drugs
2                         Stolen goods
3                      Controlled drugs
4                         Stolen goods
..                                  …
924                    Controlled drugs
925                    Controlled drugs
926                    Controlled drugs
927           Article for use in theft
928  Evidence of offences under the Act

[929 rows x 16 columns]
```

2. Combine all data into a Pandas dataframe

[42]:
```
def combine_data(keyword):
    data = []
    asked_data = pull_data(keyword)

    for date in asked_data['month'].unique():
        data.append(crawl_by_date(keyword, date))

    df = pd.concat(data)

    return df
```

[44]:
```
df = combine_data('btp')
```

```
/var/folders/84/p81hrkps12z5qktxknfn0r6m0000gq/T/ipykernel_58546/485637084.py:14
: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
res['month'] = res['date'].dt.to_period('M')
```

[65]:
```
df.head()
```

[65]:
```
  age_range                        outcome  involved_person  \
0     10-17  A no further action disposal             True
1     10-17                        Arrest             True
2     18-24                        Arrest             True
3     25-34  A no further action disposal             True
4   over 34  A no further action disposal             True

                              self_defined_ethnicity  gender  \
0                    Other ethnic group - Not stated    Male
1  Black/African/Caribbean/Black British - Caribbean    Male
2                    Other ethnic group - Not stated    None
3                    Other ethnic group - Not stated  Female
4  White - English/Welsh/Scottish/Northern Irish/…    Male

                             legislation outcome_linked_to_object_of_search  \
0         Firearms Act 1968 (section 47)                               False
1  Misuse of Drugs Act 1971 (section 23)                                True
2  Misuse of Drugs Act 1971 (section 23)                                True
3  Misuse of Drugs Act 1971 (section 23)                               False
4  Misuse of Drugs Act 1971 (section 23)                               False

                   datetime removal_of_more_than_outer_clothing  \
0 2021-11-04 15:15:00+00:00                               False
1 2021-11-03 14:00:00+00:00                               False
2 2021-11-03 01:23:00+00:00                               False
3 2021-11-04 03:42:00+00:00                               False
4 2021-11-03 19:42:00+00:00                                True

                                    outcome_object  \
0  {'id': 'bu-no-further-action', 'name': 'A no f…
1              {'id': 'bu-arrest', 'name': 'Arrest'}
2              {'id': 'bu-arrest', 'name': 'Arrest'}
3  {'id': 'bu-no-further-action', 'name': 'A no f…
4  {'id': 'bu-no-further-action', 'name': 'A no f…

                                           location  operation  \
0                                              None        NaN
1  {'latitude': '50.798400', 'street': {'id': 148…        NaN
2  {'latitude': '50.798400', 'street': {'id': 148…        NaN
3  {'latitude': '51.532500', 'street': {'id': 148…        NaN
4  {'latitude': '51.528500', 'street': {'id': 148…        NaN

  officer_defined_ethnicity          type  operation_name  object_of_search
```

```
0                      None  Person search             NaN          Firearms
1                     Black  Person search             NaN  Controlled drugs
2                     Black  Person search             NaN  Controlled drugs
3                     Black  Person search             NaN  Controlled drugs
4                     White  Person search             NaN  Controlled drugs
```

3. Clean and format the data as you see appropriate.

[76]:
```
df = df.reset_index(drop=True)
```

[77]:
```
def clean_data():
    outcome_object = pd.json_normalize(df['outcome_object'])
    outcome_object.columns = ['outcome_object_id', 'outcome_object_name']

    df = pd.concat([df, outcome_object], axis=1)
    df = df.drop(columns=['outcome_object'])

    return df
```

[78]:
```
df.head()
```

[78]:
```
  age_range                        outcome  involved_person  \
0     10-17  A no further action disposal             True
1     10-17                        Arrest             True
2     18-24                        Arrest             True
3     25-34  A no further action disposal             True
4   over 34  A no further action disposal             True

                             self_defined_ethnicity  gender  \
0                  Other ethnic group - Not stated    Male
1  Black/African/Caribbean/Black British - Caribbean    Male
2                  Other ethnic group - Not stated    None
3                  Other ethnic group - Not stated  Female
4  White - English/Welsh/Scottish/Northern Irish/…    Male

                           legislation outcome_linked_to_object_of_search  \
0        Firearms Act 1968 (section 47)                              False
1  Misuse of Drugs Act 1971 (section 23)                               True
2  Misuse of Drugs Act 1971 (section 23)                               True
3  Misuse of Drugs Act 1971 (section 23)                              False
4  Misuse of Drugs Act 1971 (section 23)                              False

                   datetime removal_of_more_than_outer_clothing  \
0 2021-11-04 15:15:00+00:00                              False
1 2021-11-03 14:00:00+00:00                              False
2 2021-11-03 01:23:00+00:00                              False
3 2021-11-04 03:42:00+00:00                              False
```

```
4 2021-11-03 19:42:00+00:00                                      True

                                      outcome_object  \
0  {'id': 'bu-no-further-action', 'name': 'A no f…
1              {'id': 'bu-arrest', 'name': 'Arrest'}
2              {'id': 'bu-arrest', 'name': 'Arrest'}
3  {'id': 'bu-no-further-action', 'name': 'A no f…
4  {'id': 'bu-no-further-action', 'name': 'A no f…

                                           location  operation  \
0                                              None        NaN
1  {'latitude': '50.798400', 'street': {'id': 148…        NaN
2  {'latitude': '50.798400', 'street': {'id': 148…        NaN
3  {'latitude': '51.532500', 'street': {'id': 148…        NaN
4  {'latitude': '51.528500', 'street': {'id': 148…        NaN

  officer_defined_ethnicity           type  operation_name  object_of_search  \
0                      None  Person search             NaN          Firearms
1                     Black  Person search             NaN  Controlled drugs
2                     Black  Person search             NaN  Controlled drugs
3                     Black  Person search             NaN  Controlled drugs
4                     White  Person search             NaN  Controlled drugs

      outcome_object_id           outcome_object_name
0  bu-no-further-action  A no further action disposal
1             bu-arrest                        Arrest
2             bu-arrest                        Arrest
3  bu-no-further-action  A no further action disposal
4  bu-no-further-action  A no further action disposal
```

4. Write the data out to a csv file.

[79]:
```
df.to_csv('btp.csv')
```

5. Write a process that checks to see if new data has been added to the API at a given time each day - if so it should pull the new data and join it to the existing csv file.

[86]:
```
from datetime import datetime, timedelta
from threading import Timer
```

[92]:
```
x = datetime.today()
y = x.replace(day=x.day, hour=1, minute=0, second=0, microsecond=0) +␣
 ↪timedelta(days=1)

delta_t = y-x
secs = delta_t.total_seconds()
```

```
def job():
    org = pd.read_csv('btp.csv')
    new_data = combine_data('btp')

    if len(org) == len(new_data):
        pass
    else:
        new_data.to_csv('btp.csv')

    return

t = Timer(secs, hello_world)
t.start()
```

6. Prepare a short explanation of how you would store the data in a relational database if required think about the structure of the data and what schema you would apply need primarykeys ID - personal infor: age_range, involded_person, self_defined_ethnicity, gender, officer_defined_ethnicity. - action: outcome, object_of_search, outcome_linked_to_object_of_search, outcome_object_id, outcome_object_name, legislation, removal_of_more_than_outer_clothing. - location and date: location, datetime

7. Suppose entries were updated or deleted in the API how would your propose that your system should handle this? Answer: Comapre lengths, if they are equal length, do nothing. Else, overwrite the dataframe

8. Prepare an example of how you would create a view on the database to return an aggregate count of the number of each crime type by date

SQL server select distinct to check how to defind a crime

SELECT DISTINCT outcome
FROM df;

=> "Arrest" is a crime

select only date, not datetime (from "2021-11-04 15:15:00+00:00" to "2021-11-04")

SELECT
CONVERT(datetime, GETDATE()) date;

filter only "Arrest" , count legislation and group by date

CREATE VIEW crime(date,crime,crime_count)
AS SELECT date,legislation,COUNT(legislation)
FROM df
WHERE outcome='Arrest'
GROUP BY date;

9. If there are other technologies that you want to include in your solution, do feel free don't feel constrained by the brief This is your chance to show off all your data engineering skills! Should have IDs to check updated data.

[ ]: