```
IndentationError: unexpected indent
>>> exit()
^[[A(base) [mc7787@hlog-2 ls team3
ana_code  data_ingest  etl_code  profiling_code  screenshots
(base) [mc7787@hlog-2 ~]$ vi team3/profiling_code/police_short_profile.py
(base) [mc7787@hlog-2 ~]$ hdfs dfs -rm team3/profiling_code/police_short_profile.py
^[[A^[[A^[[A22/04/26 23:32:22 INFO fs.TrashPolicyDefault: Moved: 'hdfs://horton.hpc.nyu.edu:8020/user/mc7787/team3/profiling_code/police_short_profile.py' to trash at: hdfs://horton.hpc.nyu.edu:8020/user/mc7787/.Trash/Current/user/mc7787/team3/profiling_code/police_short_profile.py1651030342408
(base) [mc7787@hlog-2 ~]$ hdfs dfs -put team3/profiling_code/police_short_profile.py team3/profiling_code
(base) [mc7787@hlog-2 ~]$ PYTHONSTARTUP=team3/profiling_code/police_short_profile.py pyspark --deploy-mode client
Python 3.7.9 (default, Feb  5 2021, 09:29:06)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-39)] on linux
Type "help", "copyright", "credits" or "license" for more information.
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 2.4.0-cdh6.3.4
      /_/

Using Python version 3.7.9 (default, Feb  5 2021 09:29:06)
SparkSession available as 'spark'.
number of rows:  6000
number of rows with any nan:
0
null_percentage of rowss:  0.0
columns and their numbers of nans:
_c0             0
PdId            0
IncidntNum      0
Incident Code   0
Category        0
Descript        0
DayOfWeek       0
Date            0
PdDistrict      0
Address         0
X               0
Y               0
location        0
dtype: int64
_c0
0       1
1330    1
1342    1
5436    1
3387    1
       ..
4759    1
661     1
2708    1
4755    1
2047    1
Name: _c0, Length: 6000, dtype: int64
count    6000.000000
mean     2999.500000
std      1732.195139
min         0.000000
25%      1499.750000
50%      2999.500000
75%      4499.250000
```

```
661     1
2708    1
4755    1
2047    1
Name: _c0, Length: 6000, dtype: int64
count    6000.000000
mean     2999.500000
std      1732.195139
min         0.000000
25%      1499.750000
50%      2999.500000
75%      4499.250000
max      5999.000000
Name: _c0, dtype: float64
PdId
17628983306244    1
16061777513075    1
18024338802004    1
17626148306244    1
17074448704134    1
                 ..
17615492928150    1
17066877405014    1
17081774915161    1
16081727603074    1
17007759571012    1
Name: PdId, Length: 6000, dtype: int64
count    6.000000e+03
mean     1.676151e+13
std      7.875465e+11
min      3.114752e+12
25%      1.607136e+13
50%      1.701019e+13
75%      1.709571e+13
max      1.860713e+13
Name: PdId, dtype: float64
IncidntNum
170559220    3
160392701    3
160374050    3
170746336    2
170118036    2
            ..
160955070    1
171035324    1
160666299    1
170744506    1
170182655    1
Name: IncidntNum, Length: 5890, dtype: int64
count    6.000000e+03
mean     1.676151e+08
std      7.875465e+06
min      3.114752e+07
25%      1.607136e+08
50%      1.701019e+08
75%      1.709571e+08
max      1.860713e+08
Name: IncidntNum, dtype: float64
Incident Code
6244     843
28150    206
64020    201
4134     178
```

```
160392701    3
160374050    3
170746336    2
170118036    2
            ..
160955070    1
171035324    1
160666299    1
170744506    1
170182655    1
Name: IncidntNum, Length: 5890, dtype: int64
count    6.000000e+03
mean     1.676151e+08
std      7.875465e+06
min      3.114752e+07
25%      1.607136e+08
50%      1.701019e+08
75%      1.709571e+08
max      1.860713e+08
Name: IncidntNum, dtype: float64
Incident Code
6244     843
28150    206
64020    201
4134     178
71000    165
        ...
6131       1
6365       1
16626      1
16630      1
6133       1
Name: Incident Code, Length: 388, dtype: int64
count     6000.000000
mean     26350.090333
std      25981.068053
min       2004.000000
25%       6244.000000
50%       9320.000000
75%      62050.000000
max      75030.000000
Name: Incident Code, dtype: float64
Category
LARCENY/THEFT                  1819
OTHER OFFENSES                  760
NON-CRIMINAL                    689
ASSAULT                         481
VANDALISM                       387
VEHICLE THEFT                   244
WARRANTS                        239
BURGLARY                        234
SUSPICIOUS OCC                  227
DRUG/NARCOTIC                   141
ROBBERY                         129
MISSING PERSON                  124
FRAUD                           105
TRESPASS                         69
WEAPON LAWS                      68
SECONDARY CODES                  58
STOLEN PROPERTY                  38
RECOVERED VEHICLE                30
SEX OFFENSES, FORCIBLE           30
FORGERY/COUNTERFEITING           28
```

```
count     6000.000000
mean     26350.090333
std      25981.068053
min       2004.000000
25%       6244.000000
50%       9320.000000
75%      62050.000000
max      75030.000000
Name: Incident Code, dtype: float64
Category
LARCENY/THEFT                  1819
OTHER OFFENSES                  760
NON-CRIMINAL                    689
ASSAULT                         481
VANDALISM                       387
VEHICLE THEFT                   244
WARRANTS                        239
BURGLARY                        234
SUSPICIOUS OCC                  227
DRUG/NARCOTIC                   141
ROBBERY                         129
MISSING PERSON                  124
FRAUD                           105
TRESPASS                         69
WEAPON LAWS                      68
SECONDARY CODES                  58
STOLEN PROPERTY                  38
RECOVERED VEHICLE                30
SEX OFFENSES, FORCIBLE           30
FORGERY/COUNTERFEITING           28
PROSTITUTION                     21
DISORDERLY CONDUCT               18
DRUNKENNESS                      14
DRIVING UNDER THE INFLUENCE      13
ARSON                            10
KIDNAPPING                        6
BRIBERY                           4
EXTORTION                         4
SUICIDE                           3
EMBEZZLEMENT                      2
BAD CHECKS                        2
LIQUOR LAWS                       2
SEX OFFENSES, NON FORCIBLE        1
Name: Category, dtype: int64
count              6000
unique               33
top       LARCENY/THEFT
freq              1819
Name: Category, dtype: object
Descript
GRAND THEFT FROM LOCKED AUTO                    843
MALICIOUS MISCHIEF, VANDALISM                   206
AIDED CASE, MENTAL DISTURBED                    201
PETTY THEFT FROM LOCKED AUTO                    191
BATTERY                                         178
                                               ...
EMBEZZLEMENT, GRAND THEFT                         1
BURGLARY OF WAREHOUSE, ATTEMPTED FORCIBLE ENTRY   1
WIRETAPS, UNAUTHORIZED                            1
STOLEN CELLULAR PHONE, NON-CLONED, POSSESSION     1
SALE OF METH-AMPHETAMINE                          1
Name: Descript, Length: 367, dtype: int64
count                          6000
```

```
BURGLARY OF WAREHOUSE, ATTEMPTED FORCIBLE ENTRY      1
WIRETAPS, UNAUTHORIZED                               1
STOLEN CELLULAR PHONE, NON-CLONED, POSSESSION        1
SALE OF METH-AMPHETAMINE                             1
Name: Descript, Length: 367, dtype: int64
count                             6000
unique                             367
top       GRAND THEFT FROM LOCKED AUTO
freq                               843
Name: Descript, dtype: object
DayOfWeek
Friday       930
Saturday     893
Thursday     867
Tuesday      859
Wednesday    854
Monday       825
Sunday       772
Name: DayOfWeek, dtype: int64
count       6000
unique         7
top       Friday
freq         930
Name: DayOfWeek, dtype: object
Date
01/20/2016    29
12/15/2017    24
09/12/2017    24
10/30/2016    23
01/19/2016    21
              ..
02/04/2018     1
05/09/2018     1
08/01/2017     1
05/02/2018     1
05/29/2016     1
Name: Date, Length: 843, dtype: int64
count           6000
unique           843
top       01/20/2016
freq              29
Name: Date, dtype: object
PdDistrict
SOUTHERN      1117
NORTHERN       870
MISSION        837
CENTRAL        721
BAYVIEW        551
INGLESIDE      479
TARAVAL        409
RICHMOND       367
TENDERLOIN     359
PARK           290
Name: PdDistrict, dtype: int64
count         6000
unique          10
top       SOUTHERN
freq          1117
Name: PdDistrict, dtype: object
Address
800 Block of BRYANT ST          158
800 Block of MARKET ST           57
0 Block of UNITEDNATIONS PZ      23
```

```
Name: PdDistrict, dtype: int64
count         6000
unique          10
top       SOUTHERN
freq          1117
Name: PdDistrict, dtype: object
Address
800 Block of BRYANT ST          158
800 Block of MARKET ST           57
0 Block of UNITEDNATIONS PZ      23
1000 Block of POTRERO AV         22
500 Block of JOHNFKENNEDY DR     21
                               ...
1600 Block of VISITACION AV       1
BEACH ST / MASON ST               1
TURK ST / LARKIN ST               1
3400 Block of 3RD ST              1
1200 Block of THOMAS AV           1
Name: Address, Length: 3532, dtype: int64
count                       6000
unique                      3532
top       800 Block of BRYANT ST
freq                         158
Name: Address, dtype: object
X
-122.403405    158
-122.407634     26
-122.406521     24
-122.414318     23
-122.419672     22
               ...
-122.421489      1
-122.431310      1
-122.407114      1
-122.437011      1
-122.410981      1
Name: X, Length: 3675, dtype: int64
count    6000.000000
mean     -122.423050
std         0.035783
min      -122.513642
25%      -122.433633
50%      -122.417336
75%      -122.406669
max      -120.500000
Name: X, dtype: float64
Y
37.775421    158
37.784189     26
37.785063     24
37.779944     23
37.765050     22
             ...
37.724615      1
37.725282      1
37.776447      1
37.776470      1
37.781007      1
Name: Y, Length: 3675, dtype: int64
count    6000.000000
mean       37.778117
std         0.674703
min        37.708200
```

```
X
-122.403405    158
-122.407634     26
-122.406521     24
-122.414318     23
-122.419672     22
               ...
-122.421489      1
-122.431310      1
-122.407114      1
-122.437011      1
-122.410981      1
Name: X, Length: 3675, dtype: int64
count    6000.000000
mean     -122.423050
std         0.035783
min      -122.513642
25%      -122.433633
50%      -122.417336
75%      -122.406669
max      -120.500000
Name: X, dtype: float64
Y
37.775421    158
37.784189     26
37.785063     24
37.779944     23
37.765050     22
             ...
37.724615      1
37.725282      1
37.776447      1
37.776470      1
37.781007      1
Name: Y, Length: 3675, dtype: int64
count    6000.000000
mean       37.778117
std         0.674703
min        37.708200
25%        37.757101
50%        37.775421
75%        37.785074
max        90.000000
Name: Y, dtype: float64
location
POINT (-122.40340479147905 37.775420706711)       158
POINT (-122.4076335207421 37.78418935014246)       26
POINT (-122.4065209871443 37.785062942166064)      24
POINT (-122.41431785788087 37.779944405204645)     23
POINT (-122.41967178029562 37.76505012146682)      22
                                                 ...
POINT (-122.40635183429443 37.786031157636174)      1
POINT (-122.39837433248645 37.78341960171633)       1
POINT (-122.40942036456 37.781615026578585)         1
POINT (-122.41081097310051 37.75109141772154)       1
POINT (-122.4185339723666 37.80499773501531)        1
Name: location, Length: 3675, dtype: int64
count                                          6000
unique                                         3675
top       POINT (-122.40340479147905 37.775420706711)
freq                                            158
Name: location, dtype: object
>>>
```