## Cleaning Data

### August 28, 2020

```
[1]: import numpy as np
    import pandas as pd
    import math
[2]: data = pd.read_csv('Arrest_Data_from_2020_to_Present.csv')
[3]:
    data
[3]:
           Report ID Report Type
                                                Arrest Date
                                                                Time
                                                                       Area ID
           202111570
                               RFC
                                    07/09/2020 12:00:00 AM
                                                              1830.0
    0
                                                                            21
                          BOOKING
                                    03/25/2020 12:00:00 AM
    1
             5906894
                                                              2045.0
                                                                             6
    2
           201511771
                               RFC
                                    06/23/2020 12:00:00 AM
                                                               930.0
                                                                            15
                                                              2322.0
    3
           200911637
                               RFC
                                    06/24/2020 12:00:00 AM
                                                                             9
                                    06/23/2020 12:00:00 AM
    4
           200612567
                               RFC
                                                              1210.0
                                                                             6
                                                                 . . .
    40018
             5955800
                          BOOKING
                                    06/17/2020 12:00:00 AM
                                                               122.0
                                                                            15
    40019
             5957296
                          BOOKING
                                    06/19/2020 12:00:00 AM
                                                              1130.0
                                                                            21
    40020
             5957111
                          BOOKING
                                    06/19/2020 12:00:00 AM
                                                               100.0
                                                                             7
                                    06/17/2020 12:00:00 AM
    40021
             5955822
                          BOOKING
                                                               200.0
                                                                            16
    40022
                          BOOKING
                                    06/16/2020 12:00:00 AM
                                                               300.0
                                                                            12
             5955230
             Area Name
                         Reporting District
                                               Age Sex Code Descent Code
    0
                                                29
                Topanga
                                         2125
    1
             Hollywood
                                         645
                                                24
                                                                            . . .
    2
           N Hollywood
                                        1581
                                                60
                                                           F
                                                                         0
    3
              Van Nuys
                                         901
                                                46
                                                           Μ
                                                                         Η
    4
             Hollywood
                                         642
                                                27
                                                           М
                                                                         В
    40018
           N Hollywood
                                        1591
                                                21
                                                           F
                                                                         W
    40019
                                        2156
                                                28
               Topanga
    40020
               Wilshire
                                         758
                                                25
    40021
              Foothill
                                        1698
                                                39
                                                           М
                                                                         W
    40022
           77th Street
                                        1258
                                                14
           Census Tracts
                                              Location Precinct Boundaries \
    0
                    281.0
                           POINT (-118.6031 34.2129)
                                                                       887.0
                           POINT (-118.3434 34.1015)
    1
                    414.0
                                                                      1303.0
    2
                           POINT (-118.4071 34.1478)
                                                                      1505.0
```

```
3
               220.0 POINT (-118.4662 34.2085)
                                                                1323.0
4
               2105.0 POINT (-118.3593 34.0944)
                                                                   NaN
. . .
                  . . .
                                                                   . . .
                352.0 POINT (-118.4097 34.1447)
40018
                                                                1236.0
40019
                297.0 POINT (-118.6059 34.1905)
                                                                 434.0
40020
                648.0 POINT (-118.3308 34.0482)
                                                                1345.0
40021
                167.0
                      POINT (-118.365 34.2203)
                                                                1374.0
40022
               792.0 POINT (-118.2739 33.9718)
                                                                 998.0
                  Booking Date LA Specific Plans Booking Time
0
                           NaN
                                              NaN
                                                            NaN
1
       03/25/2020 12:00:00 AM
                                              NaN
                                                         2324.0
2
                           NaN
                                              NaN
                                                            NaN
3
                           NaN
                                              NaN
                                                            NaN
4
                           NaN
                                              NaN
                                                            NaN
                                                            . . .
40018 06/17/2020 12:00:00 AM
                                              6.0
                                                          347.0
40019
      06/19/2020 12:00:00 AM
                                              2.0
                                                         1617.0
40020 06/19/2020 12:00:00 AM
                                              NaN
                                                          406.0
40021 06/17/2020 12:00:00 AM
                                              NaN
                                                          512.0
40022 06/16/2020 12:00:00 AM
                                              7.0
                                                          434.0
      Council Districts
                               Booking Location \
0
                     4.0
                                             NaN
1
                     8.0
                                       HOLLYWOOD
2
                     5.0
                                             NaN
                                             NaN
3
                     3.0
4
                     NaN
                                             NaN
. . .
40018
                     5.0 METRO - JAIL DIVISION
40019
                     4.0
                              VALLEY - JAIL DIV
40020
                    12.0 METRO - JAIL DIVISION
                          METRO - JAIL DIVISION
40021
                    5.0
40022
                    13.0
                                         77TH ST
       Neighborhood Councils (Certified) Booking Location Code
0
                                      37.0
                                                               NaN
1
                                      34.0
                                                            4206.0
2
                                      84.0
                                                               NaN
3
                                      18.0
                                                               NaN
4
                                                               NaN
                                       NaN
. . .
                                       . . .
                                                               . . .
40018
                                      84.0
                                                            4273.0
40019
                                      49.0
                                                            4279.0
40020
                                      52.0
                                                            4273.0
                                      21.0
40021
                                                            4273.0
40022
                                      46.0
                                                            4212.0
```

#### [40023 rows x 31 columns]

```
[4]: data.dtypes
[4]: Report ID
                                             int64
    Report Type
                                            object
    Arrest Date
                                            object
    Time
                                           float64
    Area ID
                                             int64
    Area Name
                                            object
    Reporting District
                                             int64
                                             int64
    Age
    Sex Code
                                            object
    Descent Code
                                            object
    Charge Group Code
                                           float64
    Charge Group Description
                                            object
    Arrest Type Code
                                            object
    Charge
                                            object
    Charge Description
                                            object
    Disposition Description
                                            object
    Address
                                            object
    Cross Street
                                            object
    LAT
                                           float64
    Zip Codes
                                           float64
    LON
                                           float64
    Census Tracts
                                           float64
    Location
                                            object
    Precinct Boundaries
                                           float64
    Booking Date
                                            object
    LA Specific Plans
                                           float64
    Booking Time
                                           float64
    Council Districts
                                           float64
    Booking Location
                                            object
    Neighborhood Councils (Certified)
                                           float64
    Booking Location Code
                                           float64
    dtype: object
[5]: data['Descent Code'] = data['Descent Code'].map({'H': 'Hispanic/Latin/Mexican',
                                'A': 'Other Asian',
                                'B': 'Black',
                                'C' : 'Chinese',
                                'D' : 'Cambodian',
                                'F' : 'Filipino',
                                'G' : 'Guamanian',
                                'I': 'American Indian/Alaskan Native',
                                'J' : 'Japanese',
                                'K' : 'Korean',
```

```
'L' : 'Laotian',
                                '0' : 'Other',
                                'P' : 'Pacific Islander',
                                'S' : 'Samoan',
                                'U' : 'Hawaiian',
                                'V' : 'Vietnamese',
                                'W' : 'White',
                                'X' : 'Unknown',
                                'Z' : 'Asian Indian'})
 [6]: data['Report Type'] = data['Report Type'].map({
                              'RFC': 'RFC',
                              'BOOKING': 'BOOKED',
                               })
 [7]: data['Arrest Date'] = pd.to_datetime(data['Arrest Date'])
 [8]: data.head(2)
        Report ID Report Type Arrest Date
                                              Time Area ID
                                                             Area Name \
 [8]:
     0 202111570
                          RFC 2020-07-09 1830.0
                                                         21
                                                               Topanga
          5906894
                       BOOKED 2020-03-25 2045.0
     1
                                                          6 Hollywood
        Reporting District Age Sex Code
                                                     Descent Code ... \
     0
                      2125
                             29
                                       M Hispanic/Latin/Mexican ...
                       645
                                                            White ...
     1
                             24
                                       М
                                        Location Precinct Boundaries \
        Census Tracts
                281.0 POINT (-118.6031 34.2129)
     0
                                                                887.0
                414.0 POINT (-118.3434 34.1015)
     1
                                                               1303.0
                  Booking Date LA Specific Plans Booking Time Council Districts \
                           NaN
                                                           NaN
                                                                             4.0
     0
                                              {\tt NaN}
     1 03/25/2020 12:00:00 AM
                                              NaN
                                                        2324.0
                                                                             8.0
      Booking Location Neighborhood Councils (Certified) Booking Location Code
     0
                    NaN
                                                       37.0
                                                                                NaN
             HOLLYWOOD
                                                       34.0
     1
                                                                            4206.0
     [2 rows x 31 columns]
[51]: #set(data['Time'])
[10]: sum(data['Time'].isnull())
[10]: 2
[11]: data['Time'] = data['Time'].fillna(0.0)
[12]: sum(data['Time'].isnull())
[12]: 0
```

```
[13]: data['Time'] = data['Time'].apply(lambda x: int(x))
[14]: data['Time'] = data['Time'].apply(lambda x: 0 if x == 2400 else x)
[15]: max(data['Time'])
[15]: 2359
[16]: def time_clean(miltime):
         num = miltime
         if num < 60:
             return pd.to_datetime(num, format='%M').strftime('%H:%M')
         if num > 60 and num < 1000:
             return pd.to_datetime('0' + str(num), format='%H%M').strftime('%H:%M')
         if num >= 1000:
             return pd.to_datetime(num, format='%H%M').strftime('%H:%M')
         else:
             return
[17]: newtime = data['Time'].apply(time_clean)
[18]: data['Time'] = newtime
[19]: data.dtypes
[19]: Report ID
                                                     int64
     Report Type
                                                    object
     Arrest Date
                                           datetime64[ns]
     Time
                                                    object
     Area ID
                                                     int64
     Area Name
                                                    object
     Reporting District
                                                     int64
     Age
                                                     int64
     Sex Code
                                                    object
     Descent Code
                                                    object
     Charge Group Code
                                                   float64
     Charge Group Description
                                                    object
                                                    object
     Arrest Type Code
     Charge
                                                    object
     Charge Description
                                                    object
     Disposition Description
                                                    object
     Address
                                                    object
     Cross Street
                                                    object
    LAT
                                                   float64
     Zip Codes
                                                   float64
    LON
                                                   float64
     Census Tracts
                                                   float64
     Location
                                                    object
     Precinct Boundaries
                                                   float64
     Booking Date
                                                    object
     LA Specific Plans
                                                   float64
```

Booking Time float64
Council Districts float64
Booking Location object
Neighborhood Councils (Certified) float64
Booking Location Code float64

dtype: object

[20]:	data											
[20]:		Report ID	Repor	t Type	Arre	st Dat	te	Time	Area ID	Area Name	· \	
	0	202111570		RFC	202	0-07-0	9	18:30	21	Topanga	ı	
	1	5906894		BOOKED	202	0-03-2	25	20:45	6	Hollywood	l	
	2	201511771		RFC	202	2020-06-23 2020-06-24 2020-06-23		09:30 15		N Hollywood		
	3	200911637		RFC	202			23:22 12:10	9	Van Nuys	3	
	4	200612567		RFC	202				6	Hollywood	l	
	40018	5955800		BOOKED	202	0-06-3	17	01:22	15	N Hollywood	l	
	40019	5957296		BOOKED	202	2020-06-19 2020-06-19		11:30 01:00	21	Topanga	ι	
	40020	5957111		BOOKED	202				7	Wilshire	)	
	40021	5955822		BOOKED	202	0-06-3	17	02:00	16	Foothill	-	
	40022	5955230		BOOKED	202	0-06-3	16	03:00	12	77th Street	;	
		Reporting			_	x Code	Э			t Code	\	
	0		2	2125	29		ľ	Hispani	c/Latin/M			
	1			645	24	1				White		
	2		1	.581	60	I				Other		
	3			901	46	1		Hispani	c/Latin/M			
	4			642	27	1	ľ			Black		
					• •							
	40018			.591	21	I			/	White		
	40019		2	2156	28	1		-	c/Latin/M			
	40020		4	758	25	1		Hispanio	c/Latin/M			
	40021			.698	39		4		/	White		
	40022		1	.258	14	I	•	ніspani	c/Latin/M	lexican		
	Census Tracts Location Precinct Boundaries							\				
	0			POINT	(-118	.6031		1.2129)		887.0		
	1							1.1015)		1303.0		
	2							1.1478)		1505.0		
	3			POINT	(-118	.4662	34	1.2085)		1323.0		
	4							1.0944)		NaN		
	40018	35	52.0	POINT	(-118	.4097	34	1.1447)		1236.0		
	40019	29	97.0	POINT	(-118	.6059	34	1.1905)		434.0		
	40020	64	18.0	POINT	(-118	.3308	34	1.0482)		1345.0		
	40021	16	37.0	POINT	(-11	8.365	34	1.2203)		1374.0		
	40022	79	92.0	POINT	(-118	.2739	33	3.9718)		998.0		

```
Booking Date LA Specific Plans Booking Time \
     0
                                 NaN
                                                    NaN
                                                                  NaN
     1
            03/25/2020 12:00:00 AM
                                                    NaN
                                                               2324.0
     2
                                 NaN
                                                    NaN
                                                                  NaN
     3
                                 NaN
                                                    NaN
                                                                  NaN
     4
                                 NaN
                                                    NaN
                                                                  NaN
            06/17/2020 12:00:00 AM
     40018
                                                    6.0
                                                                347.0
     40019
            06/19/2020 12:00:00 AM
                                                    2.0
                                                               1617.0
     40020
            06/19/2020 12:00:00 AM
                                                                406.0
                                                    NaN
     40021
            06/17/2020 12:00:00 AM
                                                                512.0
                                                    NaN
     40022 06/16/2020 12:00:00 AM
                                                    7.0
                                                                434.0
           Council Districts
                                     Booking Location \
     0
                          4.0
                                                   NaN
                                             HOLLYWOOD
                          8.0
     1
     2
                          5.0
                                                   NaN
     3
                          3.0
                                                   NaN
     4
                          NaN
                                                   NaN
                           . . .
     . . .
     40018
                          5.0
                                METRO - JAIL DIVISION
     40019
                          4.0
                                    VALLEY - JAIL DIV
     40020
                         12.0
                              METRO - JAIL DIVISION
     40021
                               METRO - JAIL DIVISION
                          5.0
     40022
                         13.0
                                               77TH ST
            Neighborhood Councils (Certified)
                                                  Booking Location Code
     0
                                            37.0
                                                                      NaN
                                                                  4206.0
     1
                                            34.0
     2
                                            84.0
                                                                      NaN
     3
                                            18.0
                                                                      NaN
     4
                                             NaN
                                                                      NaN
                                             . . .
                                                                      . . .
     40018
                                            84.0
                                                                  4273.0
     40019
                                            49.0
                                                                  4279.0
     40020
                                            52.0
                                                                  4273.0
     40021
                                            21.0
                                                                  4273.0
     40022
                                            46.0
                                                                  4212.0
     [40023 rows x 31 columns]
[21]: data['Hour'] = pd.to_datetime(data['Time'],format= '%H:%M').dt.hour
[22]: data['Arrest Date'].dt.month
[22]: 0
              7
              3
     1
     2
              6
```

```
3 6
4 6
...
40018 6
40019 6
40020 6
40021 6
40022 6
```

Name: Arrest Date, Length: 40023, dtype: int64

# [23]: data['Month'] = data['Arrest Date'].dt.month

[24]: data.dtypes

[24]:	Report ID	int64
	Report Type	object
	Arrest Date	datetime64[ns]
	Time	object
	Area ID	int64
	Area Name	object
	Reporting District	int64
	Age	int64
	Sex Code	object
	Descent Code	object
	Charge Group Code	float64
	Charge Group Description	object
	Arrest Type Code	object
	Charge	object
	Charge Description	object
	Disposition Description	object
	Address	object
	Cross Street	object
	LAT	float64
	Zip Codes	float64
	LON	float64
	Census Tracts	float64
	Location	object
	Precinct Boundaries	float64
	Booking Date	object
	LA Specific Plans	float64
	Booking Time	float64
	Council Districts	float64
	Booking Location	object
	Neighborhood Councils (Certified)	float64
	Booking Location Code	float64
	Hour	int64
	Month	int64

dtype: object

```
[25]: data
[25]:
             Report ID Report Type Arrest Date
                                                     Time
                                                            Area ID
                                                                        Area Name
     0
             202111570
                                 RFC
                                       2020-07-09
                                                    18:30
                                                                  21
                                                                          Topanga
     1
               5906894
                             BOOKED
                                       2020-03-25
                                                    20:45
                                                                   6
                                                                        Hollywood
     2
             201511771
                                 RFC
                                       2020-06-23
                                                    09:30
                                                                  15
                                                                      N Hollywood
     3
                                                                   9
             200911637
                                 RFC
                                       2020-06-24
                                                    23:22
                                                                         Van Nuys
             200612567
                                 RFC
                                                                   6
                                                                        Hollywood
                                       2020-06-23
                                                    12:10
     . . .
                    . . .
                                 . . .
                                                       . . .
                                                                 . . .
     40018
               5955800
                             BOOKED
                                       2020-06-17
                                                    01:22
                                                                      N Hollywood
                                                                  15
     40019
               5957296
                              BOOKED
                                                    11:30
                                                                  21
                                       2020-06-19
                                                                          Topanga
                                                                   7
     40020
               5957111
                              BOOKED
                                       2020-06-19
                                                    01:00
                                                                         Wilshire
     40021
               5955822
                              BOOKED
                                       2020-06-17
                                                    02:00
                                                                  16
                                                                         Foothill
                                                                  12
     40022
               5955230
                              BOOKED
                                       2020-06-16
                                                    03:00
                                                                      77th Street
                                   Age Sex Code
             Reporting District
                                                              Descent Code
     0
                             2125
                                    29
                                               М
                                                   Hispanic/Latin/Mexican
     1
                              645
                                    24
                                               Μ
                                                                      White
     2
                                               F
                             1581
                                    60
                                                                      Other
     3
                              901
                                               М
                                    46
                                                   Hispanic/Latin/Mexican
     4
                              642
                                    27
                                               М
                                                                      Black
                              . . .
     40018
                                               F
                             1591
                                    21
                                                                      White
     40019
                             2156
                                    28
                                                   Hispanic/Latin/Mexican
     40020
                              758
                                    25
                                                   Hispanic/Latin/Mexican
     40021
                             1698
                                    39
                                               М
                                                                      White
     40022
                             1258
                                    14
                                                   Hispanic/Latin/Mexican
             Precinct Boundaries
                                               Booking Date LA Specific Plans
     0
                             887.0
                                                          NaN
                                                                              NaN
     1
                            1303.0
                                    03/25/2020 12:00:00 AM
                                                                              NaN
     2
                            1505.0
                                                          NaN
                                                                              NaN
     3
                            1323.0
                                                          NaN
                                                                              NaN
     4
                               NaN
                                                          NaN
                                                                              NaN
     . . .
                                    06/17/2020 12:00:00 AM
     40018
                            1236.0
                                                                              6.0
     40019
                                    06/19/2020 12:00:00 AM
                             434.0
                                                                              2.0
     40020
                            1345.0
                                    06/19/2020 12:00:00 AM
                                                                              NaN
     40021
                            1374.0
                                    06/17/2020 12:00:00 AM
                                                                              NaN
                                    06/16/2020 12:00:00 AM
                                                                              7.0
     40022
                             998.0
            Booking Time Council Districts
                                                     Booking Location
     0
                      NaN
                                          4.0
                                                                    NaN
     1
                   2324.0
                                          8.0
                                                             HOLLYWOOD
     2
                                          5.0
                      NaN
                                                                    NaN
     3
                      NaN
                                          3.0
                                                                    NaN
     4
                      NaN
                                          NaN
                                                                    NaN
```

```
347.0
     40018
                                        5.0
                                             METRO - JAIL DIVISION
     40019
                  1617.0
                                        4.0
                                                 VALLEY - JAIL DIV
     40020
                  406.0
                                       12.0
                                             METRO - JAIL DIVISION
     40021
                  512.0
                                        5.0
                                             METRO - JAIL DIVISION
     40022
                  434.0
                                       13.0
                                                            77TH ST
           Neighborhood Councils (Certified) Booking Location Code
                                                                        Hour
                                                                             Month
     0
                                                                                  7
                                          37.0
                                                                  NaN
                                                                          18
     1
                                          34.0
                                                               4206.0
                                                                          20
                                                                                  3
     2
                                                                           9
                                          84.0
                                                                  NaN
                                                                                  6
     3
                                          18.0
                                                                  NaN
                                                                          23
                                                                                  6
     4
                                           NaN
                                                                  NaN
                                                                          12
                                                                                  6
     40018
                                          84.0
                                                               4273.0
                                                                          1
                                                                                  6
     40019
                                          49.0
                                                               4279.0
                                                                          11
                                                                                  6
     40020
                                          52.0
                                                               4273.0
                                                                                  6
                                                                           1
     40021
                                                               4273.0
                                                                           2
                                                                                  6
                                          21.0
     40022
                                          46.0
                                                               4212.0
     [40023 rows x 33 columns]
[26]: set(data['Sex Code'])
[26]: {'F', 'M'}
[27]: bureau_dict = {
       "Valley": [21, 17, 19, 16, 10, 9, 15],
       "West": [8, 6, 7, 20, 14],
       "Central": [11, 2, 1, 13, 4],
         "South" : [3, 12, 18, 5]
     }
[28]: bureau names = []
     for area_id in data['Area ID']:
         for key in bureau_dict:
             ids = bureau_dict[key]
             if area_id in ids:
                  bureau_names.append(key)
[52]: #bureau_names
[30]: data['Bureau'] = bureau_names
[31]: data.head(10)
[31]:
        Report ID Report Type Arrest Date
                                              Time
                                                     Area ID
                                                                Area Name
     0 202111570
                           RFC
                                2020-07-09
                                             18:30
                                                          21
                                                                  Topanga
          5906894
                        BOOKED
                                2020-03-25
     1
                                             20:45
                                                           6
                                                                Hollywood
     2 201511771
                           RFC
                                2020-06-23
                                             09:30
                                                          15
                                                             N Hollywood
```

```
3
   200911637
                       RFC
                             2020-06-24
                                          23:22
                                                         9
                                                               Van Nuys
                       RFC
                             2020-06-23
                                                         6
4
   200612567
                                          12:10
                                                              Hollywood
   200612602
                       RFC
                             2020-07-10
                                          14:30
                                                         6
                                                              Hollywood
                                                         6
6
   200611612
                       RFC
                             2020-06-17
                                          18:45
                                                              Hollywood
7
   201512657
                       RFC
                             2020-07-13
                                          09:50
                                                       15
                                                            N Hollywood
                   BOOKED
8
     5956479
                             2020-06-18
                                          09:50
                                                       14
                                                                Pacific
9
   202012596
                       RFC
                             2020-07-24
                                                       20
                                                                Olympic
                                          12:50
                         Age Sex Code
   Reporting District
                                                    Descent Code
0
                  2125
                          29
                                         Hispanic/Latin/Mexican
1
                   645
                          24
                                     М
                                                            White
                                                                    . . .
2
                   1581
                          60
                                     F
                                                            Other
3
                   901
                          46
                                     М
                                         Hispanic/Latin/Mexican
4
                   642
                          27
                                     Μ
                                                            Black
5
                   666
                          73
                                     М
                                                            Black
6
                   646
                          31
                                     Μ
                                                            Black
7
                                     F
                   1505
                          46
                                                            White
8
                   1431
                          44
                                     F
                                                            White
9
                  2004
                          24
                                         Hispanic/Latin/Mexican
              Booking Date LA Specific Plans Booking Time Council Districts
0
                        NaN
                                            NaN
                                                           NaN
                                                                               4.0
1
   03/25/2020 12:00:00 AM
                                            NaN
                                                       2324.0
                                                                               8.0
2
                        NaN
                                            NaN
                                                           NaN
                                                                               5.0
3
                        NaN
                                            NaN
                                                           NaN
                                                                               3.0
4
                        NaN
                                            NaN
                                                           NaN
                                                                               NaN
5
                        NaN
                                            NaN
                                                           NaN
                                                                               8.0
6
                        NaN
                                            NaN
                                                           NaN
                                                                               8.0
7
                        NaN
                                            NaN
                                                           NaN
                                                                               5.0
   06/18/2020 12:00:00 AM
                                           10.0
                                                       1108.0
                                                                              10.0
8
9
                        NaN
                                            NaN
                                                           NaN
                                                                               7.0
  Booking Location Neighborhood Councils (Certified) Booking Location Code
0
                                                     37.0
                                                                               NaN
                NaN
         HOLLYWOOD
                                                     34.0
                                                                            4206.0
1
2
                NaN
                                                     84.0
                                                                               NaN
                                                     18.0
3
                NaN
                                                                               NaN
4
                NaN
                                                      NaN
                                                                               NaN
                                                     34.0
5
                NaN
                                                                               NaN
6
                NaN
                                                     82.0
                                                                               NaN
7
                NaN
                                                     21.0
                                                                               NaN
            PACIFIC
                                                                            4214.0
8
                                                     30.0
9
                NaN
                                                     89.0
                                                                               NaN
        Month
                Bureau
  Hour
             7
0
    18
                Valley
             3
1
    20
                  West
```

Valley	6	9	2
Valley	6	23	3
West	6	12	4
West	7	14	5
West	6	18	6
Valley	7	9	7
West	6	9	8
West	7	12	9

### [10 rows x 34 columns]

## [32]: data.dtypes

[32]:	Report ID	int64		
	Report Type	object		
	Arrest Date	datetime64[ns]		
	Time	object		
	Area ID	int64		
	Area Name	object		
	Reporting District	int64		
	Age	int64		
	Sex Code	object		
	Descent Code	object		
	Charge Group Code	float64		
	Charge Group Description	object		
	Arrest Type Code	object		
	Charge	object		
	Charge Description	object		
	Disposition Description	object		
	Address	object		
	Cross Street	object		
	LAT	float64		
	Zip Codes	float64		
	LON	float64		
	Census Tracts	float64		
	Location	object		
	Precinct Boundaries	float64		
	Booking Date	object		
	LA Specific Plans	float64		
	Booking Time	float64		
	Council Districts	float64		
	Booking Location	object		
	Neighborhood Councils (Certified)	float64		
	Booking Location Code	float64		
	Hour	int64		
	Month	int64		
	Bureau	object		

dtype: object

```
[37]: data = data.rename(columns = {'Descent Code': 'Ethnicity'})
[38]:
     data
[38]:
             Report ID Report Type Arrest Date
                                                     Time
                                                            Area ID
                                                                        Area Name
     0
             202111570
                                 RFC
                                       2020-07-09
                                                    18:30
                                                                 21
                                                                          Topanga
               5906894
                             BOOKED
                                                    20:45
                                                                  6
     1
                                       2020-03-25
                                                                        Hollywood
     2
             201511771
                                 RFC
                                                    09:30
                                                                 15
                                                                      N Hollywood
                                       2020-06-23
                                                                  9
     3
             200911637
                                 RFC
                                       2020-06-24
                                                    23:22
                                                                         Van Nuys
                                       2020-06-23
                                                                  6
             200612567
                                 RFC
                                                    12:10
                                                                        Hollywood
                                                      . . .
                    . . .
                                 . . .
                                                                . . .
     40018
               5955800
                             BOOKED
                                       2020-06-17
                                                    01:22
                                                                 15
                                                                      N Hollywood
     40019
               5957296
                             BOOKED
                                      2020-06-19
                                                    11:30
                                                                 21
                                                                          Topanga
     40020
               5957111
                             BOOKED
                                      2020-06-19
                                                    01:00
                                                                  7
                                                                         Wilshire
     40021
               5955822
                             BOOKED
                                      2020-06-17
                                                    02:00
                                                                 16
                                                                         Foothill
     40022
               5955230
                             BOOKED
                                      2020-06-16
                                                    03:00
                                                                     77th Street
                                                                 12
                                   Age Sex Code
                                                                 Ethnicity
             Reporting District
     0
                                    29
                             2125
                                               Μ
                                                   Hispanic/Latin/Mexican
     1
                              645
                                    24
                                               M
                                                                      White
                                               F
     2
                             1581
                                    60
                                                                      Other
     3
                             901
                                    46
                                               M
                                                   Hispanic/Latin/Mexican
     4
                             642
                                    27
                                               M
                                                                      Black
                                               F
     40018
                             1591
                                    21
                                                                      White
     40019
                                    28
                             2156
                                               Μ
                                                   Hispanic/Latin/Mexican
     40020
                              758
                                    25
                                                   Hispanic/Latin/Mexican
     40021
                            1698
                                    39
                                                                      White
     40022
                            1258
                                                   Hispanic/Latin/Mexican
                                    14
                        Booking Date LA Specific Plans Booking Time
     0
                                  NaN
                                                      NaN
                                                                     NaN
     1
             03/25/2020 12:00:00 AM
                                                      NaN
                                                                 2324.0
     2
                                  NaN
                                                      NaN
                                                                     NaN
     3
                                  NaN
                                                      NaN
                                                                     NaN
     4
                                  NaN
                                                      NaN
                                                                    NaN
                                                      . . .
                                                                     . . .
             06/17/2020 12:00:00 AM
     40018
                                                      6.0
                                                                  347.0
     40019
             06/19/2020 12:00:00 AM
                                                      2.0
                                                                 1617.0
     40020
             06/19/2020 12:00:00 AM
                                                      NaN
                                                                  406.0
     40021
             06/17/2020 12:00:00 AM
                                                      NaN
                                                                  512.0
             06/16/2020 12:00:00 AM
     40022
                                                      7.0
                                                                  434.0
            Council Districts
                                      Booking Location
     0
                           4.0
                                                     NaN
     1
                           8.0
                                              HOLLYWOOD
     2
                           5.0
                                                     NaN
     3
                           3.0
                                                     NaN
```

```
4
                           NaN
                                                    NaN
                           . . .
     40018
                           5.0 METRO - JAIL DIVISION
     40019
                           4.0
                                     VALLEY - JAIL DIV
     40020
                          12.0 METRO - JAIL DIVISION
     40021
                           5.0 METRO - JAIL DIVISION
     40022
                          13.0
                                                77TH ST
           Neighborhood Councils (Certified) Booking Location Code Hour
                                                                               Month \
     0
                                           37.0
                                                                     NaN
                                                                           18
                                                                                    7
     1
                                           34.0
                                                                 4206.0
                                                                           20
                                                                                    3
     2
                                           84.0
                                                                     NaN
                                                                            9
                                                                                    6
     3
                                           18.0
                                                                    NaN
                                                                           23
                                                                                    6
     4
                                            NaN
                                                                    {\tt NaN}
                                                                           12
                                                                                    6
                                            . . .
                                                                     . . .
     . . .
                                                                           . . .
                                                                 4273.0
                                                                                    6
     40018
                                           84.0
                                                                            1
     40019
                                           49.0
                                                                 4279.0
                                                                                    6
                                                                           11
     40020
                                           52.0
                                                                 4273.0
                                                                                    6
                                                                            1
     40021
                                                                 4273.0
                                                                                    6
                                           21.0
     40022
                                           46.0
                                                                 4212.0
                                                                                    6
            Bureau
     0
            Valley
     1
               West
     2
            Valley
     3
            Valley
               West
                . . .
     . . .
     40018
            Valley
     40019
            Valley
     40020
               West
     40021
            Valley
     40022
              South
     [40023 rows x 34 columns]
[40]: #area names --> zip code of police station
     list(set(data['Area Name']))
[40]: ['Pacific',
      'N Hollywood',
      'Southwest',
      'Newton',
      'West LA',
      'Hollenbeck',
      'Mission',
      'Olympic',
      'Harbor',
```

```
'Wilshire',
      'Topanga',
      'West Valley',
      '77th Street',
      'Devonshire',
      'Rampart',
      'Hollywood',
      'Southeast',
      'Central',
      'Van Nuys',
      'Foothill',
      'Northeast']
[41]: areazip_dict = {'N Hollywood': 91601,
      'Rampart': 90017,
      'Southwest': 90062,
      'West Valley': 91335,
      'Wilshire': 90019,
      'Olympic': 90006,
      'Central': 90014,
      'Devonshire': 91325,
      'Southeast': 90061,
      'Hollywood': 90028,
      'Hollenbeck': 90033,
      'Mission': 91345,
      'Northeast': 90065,
      'Pacific': 90066,
      'West LA': 90025,
      'Topanga': 91304,
      'Van Nuys': 91401,
      'Harbor': 90731,
      'Newton': 90011,
      'Foothill': 91331,
      '77th Street': 90003}
[42]: zipincome_dict = \{91601 : 52607,
      90017: 31423,
      90062: 44835,
      91335 : 60413,
      90019: 52625,
      90006: 35411,
      90014: 35857,
      91325 : 68642,
      90061: 39739,
      90028: 43998,
      90033 : 35548,
      91345 : 75969,
      90065 : 63147,
```

```
90066: 79088,
      90025: 83655,
      91304: 68282,
      91401 : 54058,
      90731 : 51161,
      90011: 36954,
      91331: 60360,
      90003 : 37226
     }
[45]: data['Area Zip'] = data['Area Name'].map(areazip_dict)
[48]: data['Zip Median Income'] = data['Area Zip'].map(zipincome_dict)
[49]: data.head(10)
[49]:
        Report ID Report Type Arrest Date
                                               Time
                                                      Area ID
                                                                  Area Name
                                              18:30
        202111570
                            RFC
                                 2020-07-09
                                                           21
                                                                    Topanga
     1
          5906894
                        BOOKED
                                 2020-03-25
                                              20:45
                                                            6
                                                                  Hollywood
     2
        201511771
                            RFC
                                              09:30
                                                           15
                                 2020-06-23
                                                                N Hollywood
     3
        200911637
                            RFC
                                 2020-06-24
                                              23:22
                                                            9
                                                                   Van Nuys
     4
        200612567
                            RFC
                                 2020-06-23
                                              12:10
                                                            6
                                                                  Hollywood
     5
        200612602
                            RFC
                                 2020-07-10
                                              14:30
                                                            6
                                                                  Hollywood
        200611612
                            RFC
                                 2020-06-17
                                              18:45
                                                            6
                                                                  Hollywood
     7
        201512657
                            RFC
                                 2020-07-13
                                              09:50
                                                           15
                                                                N Hollywood
                        BOOKED
     8
          5956479
                                 2020-06-18
                                              09:50
                                                           14
                                                                    Pacific
     9
        202012596
                            RFC
                                 2020-07-24
                                              12:50
                                                           20
                                                                    Olympic
        Reporting District
                             Age Sex Code
                                                           Ethnicity
     0
                               29
                       2125
                                             Hispanic/Latin/Mexican
     1
                        645
                               24
                                          М
                                                                White
                                                                       . . .
                                          F
     2
                       1581
                               60
                                                                Other
     3
                        901
                               46
                                          М
                                             Hispanic/Latin/Mexican
     4
                        642
                               27
                                          М
                                                                Black
     5
                        666
                               73
                                          Μ
                                                                Black
     6
                        646
                               31
                                          Μ
                                                                Black
     7
                       1505
                               46
                                          F
                                                                White
     8
                                          F
                       1431
                               44
                                                                White
     9
                       2004
                               24
                                             Hispanic/Latin/Mexican
        Booking Time Council Districts Booking Location
     0
                  NaN
                                      4.0
                                                        NaN
     1
               2324.0
                                      8.0
                                                 HOLLYWOOD
     2
                  NaN
                                      5.0
                                                        NaN
     3
                  NaN
                                      3.0
                                                        NaN
     4
                  NaN
                                                        NaN
                                     NaN
     5
                  NaN
                                      8.0
                                                        NaN
                                      8.0
                                                        NaN
     6
                  NaN
     7
                  NaN
                                      5.0
                                                        NaN
```

	8	1108.0	10.0 PACIFIC										
	9	NaN	7.0		NaN								
		Neighborhood Councils	(Certified)	Booking	Location	Code	Hour	Month	Bureau	\			
	0	O	37.0	Ü		NaN	18	7	Valley				
	1		34.0		42	206.0	20	3	West				
	2		84.0			NaN	9	6	Valley				
	3		18.0			NaN	23	6	Valley				
	4		NaN			NaN	12	6	West				
	5		34.0			NaN	14	7	West				
	6		82.0			NaN	18	6	West				
	7		21.0			NaN	9	7	Valley				
	8		30.0		42	214.0	9	6	West				
	9		89.0			NaN	12	7	West				
		Area Zip Zip Median											
	0	91304	68282										
	1	90028	43998										
	2	91601	52607										
	3	91401	54058										
	4	90028	43998										
	5	90028	43998										
	6	90028	43998										
	7	91601	52607										
	8	90066 90006	79088 35411										
	9	90000	33411										
	Г1	[10 rows x 36 columns]											
F 7													
[47]:	da	ta.dtypes											
[47]:	Re	port ID			int64								
	Re	port Type			object								
	Ar	rest Date		datetime64[ns]									
	Ti	me		object									
	Ar	ea ID	int64 object										
		ea Name											
		porting District	int64										
	Ag		int64 object										
		x Code											
		hnicity			object								
		arge Group Code			float64								
		arge Group Description			object								
		rest Type Code			object								
		arge			object								
		arge Description			object								
	1/1	5005 L. 1011 DESCETULION		object.									

object object

Charge Description
Disposition Description
Address

```
Cross Street
                                                   object
    LAT
                                                  float64
    Zip Codes
                                                  float64
    LON
                                                  float64
    Census Tracts
                                                  float64
    Location
                                                   object
    Precinct Boundaries
                                                  float64
    Booking Date
                                                   object
    LA Specific Plans
                                                  float64
    Booking Time
                                                  float64
    Council Districts
                                                  float64
    Booking Location
                                                   object
                                                  float64
    Neighborhood Councils (Certified)
    Booking Location Code
                                                  float64
    Hour
                                                    int64
    Month
                                                    int64
    Bureau
                                                   object
     Area Zip
                                                    int64
     dtype: object
[50]: data.to_csv('Clean_Arrest_Digital_Humanities.csv')
```

[]: