# No Show Code

December 11, 2020

[1]:
```
# imports
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
%matplotlib inline
```

[2]:
```
#grab and peep my data

noshow = pd.read_csv('noshowappointments-kagglev2-may-2016 (1).csv')
```

[3]:
```
#peep data
noshow.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110527 entries, 0 to 110526
Data columns (total 14 columns):
 #   Column          Non-Null Count   Dtype  
---  ------          --------------   -----  
 0   PatientId       110527 non-null  float64
 1   AppointmentID   110527 non-null  int64  
 2   Gender          110527 non-null  object 
 3   ScheduledDay    110527 non-null  object 
 4   AppointmentDay  110527 non-null  object 
 5   Age             110527 non-null  int64  
 6   Neighbourhood   110527 non-null  object 
 7   Scholarship     110527 non-null  int64  
 8   Hipertension    110527 non-null  int64  
 9   Diabetes        110527 non-null  int64  
 10  Alcoholism      110527 non-null  int64  
 11  Handcap         110527 non-null  int64  
 12  SMS_received    110527 non-null  int64  
 13  No-show         110527 non-null  object 
dtypes: float64(1), int64(8), object(5)
memory usage: 11.8+ MB
```

[4]:
```
dupes =noshow.duplicated()
```

[5]:
```
dupes.sum()
```

[5]:
```
0
```

[6]:
```
noshow.isnull().sum()
```

[6]:
```
PatientId         0
AppointmentID     0
Gender            0
ScheduledDay      0
AppointmentDay    0
Age               0
Neighbourhood     0
Scholarship       0
Hipertension      0
Diabetes          0
Alcoholism        0
Handcap           0
SMS_received      0
No-show           0
dtype: int64
```

[7]:
```
# No dupes  no show all good
```

[8]:
```
#Are people who are reciving welfare more or less likely to show up for␣
↪appointments?
```

[9]:
```
noshow['Neighbourhood'].value_counts()
```

[9]:
```
JARDIM CAMBURI                 7717
MARIA ORTIZ                    5805
RESISTÊNCIA                    4431
JARDIM DA PENHA                3877
ITARARÉ                        3514
                               …
ILHA DO BOI                      35
ILHA DO FRADE                    10
AEROPORTO                         8
ILHAS OCEÂNICAS DE TRINDADE       2
PARQUE INDUSTRIAL                 1
Name: Neighbourhood, Length: 81, dtype: int64
```

[10]:
```
noshow['SMS_received'].value_counts(normalize = True)* 100
```

[10]:
```
0    67.897437
1    32.102563
Name: SMS_received, dtype: float64
```

[11]:
```
noshow['No-show'].value_counts()
```

[11]:
```
No     88208
Yes    22319
Name: No-show, dtype: int64
```

[12]:
```
# Are people who recieve wellfare more or less likely to show up?
```

[13]:
```
noshow = noshow.replace({'No-show': {'Yes': True, 'No': False}})
```

[14]:
```
noshow = noshow.replace({'No-show': {True: '0', False: 1}})
```

[15]:
```
noshow.head()
# No-show 'no'means they came, 'yes'means they didn't show
```

[15]:

| | PatientId | AppointmentID | Gender | ScheduledDay | \ |
|---|---|---|---|---|---|
| 0 | 2.987250e+13 | 5642903 | F | 2016-04-29T18:38:08Z | |
| 1 | 5.589978e+14 | 5642503 | M | 2016-04-29T16:08:27Z | |
| 2 | 4.262962e+12 | 5642549 | F | 2016-04-29T16:19:04Z | |
| 3 | 8.679512e+11 | 5642828 | F | 2016-04-29T17:29:31Z | |
| 4 | 8.841186e+12 | 5642494 | F | 2016-04-29T16:07:23Z | |

| | AppointmentDay | Age | Neighbourhood | Scholarship | Hipertension | \ |
|---|---|---|---|---|---|---|
| 0 | 2016-04-29T00:00:00Z | 62 | JARDIM DA PENHA | 0 | 1 | |
| 1 | 2016-04-29T00:00:00Z | 56 | JARDIM DA PENHA | 0 | 0 | |
| 2 | 2016-04-29T00:00:00Z | 62 | MATA DA PRAIA | 0 | 0 | |
| 3 | 2016-04-29T00:00:00Z | 8 | PONTAL DE CAMBURI | 0 | 0 | |
| 4 | 2016-04-29T00:00:00Z | 56 | JARDIM DA PENHA | 0 | 1 | |

| | Diabetes | Alcoholism | Handcap | SMS_received | No-show |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 0 | 0 | 1 |
| 2 | 0 | 0 | 0 | 0 | 1 |
| 3 | 0 | 0 | 0 | 0 | 1 |
| 4 | 1 | 0 | 0 | 0 | 1 |

We have unneeded columns that can be removed, such as Neighbourhood, Appointment Day, ect

[16]:
```
noshow.drop(['PatientId', 'AppointmentID'], axis = 1, inplace = True)
```

This leaves us with age, if the person is on wellfare or not, and a few conditions they may have, weather they showed up for the appointment or no

[17]:
```
show = noshow.groupby('No-show')
```

[18]:
```
show.head()
```

[18]:
```
    Gender          ScheduledDay        AppointmentDay  Age      Neighbourhood  \
0        F  2016-04-29T18:38:08Z  2016-04-29T00:00:00Z   62    JARDIM DA PENHA
1        M  2016-04-29T16:08:27Z  2016-04-29T00:00:00Z   56    JARDIM DA PENHA
2        F  2016-04-29T16:19:04Z  2016-04-29T00:00:00Z   62      MATA DA PRAIA
3        F  2016-04-29T17:29:31Z  2016-04-29T00:00:00Z    8  PONTAL DE CAMBURI
4        F  2016-04-29T16:07:23Z  2016-04-29T00:00:00Z   56    JARDIM DA PENHA
6        F  2016-04-27T15:05:12Z  2016-04-29T00:00:00Z   23         GOIABEIRAS
7        F  2016-04-27T15:39:58Z  2016-04-29T00:00:00Z   39         GOIABEIRAS
11       M  2016-04-26T08:44:12Z  2016-04-29T00:00:00Z   29     NOVA PALESTINA
17       F  2016-04-28T09:28:57Z  2016-04-29T00:00:00Z   40          CONQUISTA
20       F  2016-04-27T07:51:14Z  2016-04-29T00:00:00Z   30     NOVA PALESTINA

    Scholarship  Hipertension  Diabetes  Alcoholism  Handcap  SMS_received  \
0             0             1         0           0        0             0
1             0             0         0           0        0             0
2             0             0         0           0        0             0
3             0             0         0           0        0             0
4             0             1         1           0        0             0
6             0             0         0           0        0             0
7             0             0         0           0        0             0
11            0             0         0           0        0             1
17            1             0         0           0        0             0
20            0             0         0           0        0             0

    No-show
0         1
1         1
2         1
3         1
4         1
6         0
7         0
11        0
17        0
20        0
```

[19]:
```
show.tail()
```

[19]:
```
        Gender          ScheduledDay        AppointmentDay  Age  \
110484       F  2016-06-03T14:43:56Z  2016-06-07T00:00:00Z   45
110492       M  2016-06-08T08:50:19Z  2016-06-08T00:00:00Z   33
110496       F  2016-06-06T17:35:38Z  2016-06-08T00:00:00Z   37
110515       M  2016-06-06T15:58:05Z  2016-06-08T00:00:00Z   33
110516       F  2016-06-07T07:45:16Z  2016-06-08T00:00:00Z   37
110522       F  2016-05-03T09:15:35Z  2016-06-07T00:00:00Z   56
110523       F  2016-05-03T07:27:33Z  2016-06-07T00:00:00Z   51
110524       F  2016-04-27T16:03:52Z  2016-06-07T00:00:00Z   21
```

```
110525      F  2016-04-27T15:09:23Z  2016-06-07T00:00:00Z   38
110526      F  2016-04-27T13:30:56Z  2016-06-07T00:00:00Z   54

         Neighbourhood  Scholarship  Hipertension  Diabetes  Alcoholism  \
110484  BARRO VERMELHO            0             0         0           0
110492     MARIA ORTIZ            0             1         0           0
110496     MARIA ORTIZ            0             1         0           0
110515     MARIA ORTIZ            0             1         0           0
110516     MARIA ORTIZ            0             0         0           0
110522     MARIA ORTIZ            0             0         0           0
110523     MARIA ORTIZ            0             0         0           0
110524     MARIA ORTIZ            0             0         0           0
110525     MARIA ORTIZ            0             0         0           0
110526     MARIA ORTIZ            0             0         0           0

        Handcap  SMS_received  No-show
110484        0             0        0
110492        0             0        0
110496        0             0        0
110515        0             0        0
110516        0             0        0
110522        0             1        1
110523        0             1        1
110524        0             1        1
110525        0             1        1
110526        0             1        1
```

[ ]:

[20]:
```
#Only people who didn't show
no_show = noshow[noshow['No-show'] == '1']
```

[21]:
```
#only people who showed up
show = noshow[noshow['No-show'] == '0']
```

[22]:
```
no_show.describe()
```

[22]:
```
       Age  Scholarship  Hipertension  Diabetes  Alcoholism  Handcap  \
count  0.0          0.0           0.0       0.0         0.0      0.0
mean   NaN          NaN           NaN       NaN         NaN      NaN
std    NaN          NaN           NaN       NaN         NaN      NaN
min    NaN          NaN           NaN       NaN         NaN      NaN
25%    NaN          NaN           NaN       NaN         NaN      NaN
50%    NaN          NaN           NaN       NaN         NaN      NaN
75%    NaN          NaN           NaN       NaN         NaN      NaN
max    NaN          NaN           NaN       NaN         NaN      NaN
```

```
       SMS_received  No-show
count           0.0      0.0
mean            NaN      NaN
std             NaN      NaN
min             NaN      NaN
25%             NaN      NaN
50%             NaN      NaN
75%             NaN      NaN
max             NaN      NaN
```

[23]:
```
gender_count = no_show.groupby('Gender').count()
gender_count
```

[23]:
```
Empty DataFrame
Columns: [ScheduledDay, AppointmentDay, Age, Neighbourhood, Scholarship,
Hipertension, Diabetes, Alcoholism, Handcap, SMS_received, No-show]
Index: []
```

[24]:
```
show.describe()
```

[24]:
```
        Age  Scholarship  Hipertension  Diabetes  Alcoholism  Handcap  \
count   0.0          0.0           0.0       0.0         0.0      0.0
mean    NaN          NaN           NaN       NaN         NaN      NaN
std     NaN          NaN           NaN       NaN         NaN      NaN
min     NaN          NaN           NaN       NaN         NaN      NaN
25%     NaN          NaN           NaN       NaN         NaN      NaN
50%     NaN          NaN           NaN       NaN         NaN      NaN
75%     NaN          NaN           NaN       NaN         NaN      NaN
max     NaN          NaN           NaN       NaN         NaN      NaN

       SMS_received  No-show
count           0.0      0.0
mean            NaN      NaN
std             NaN      NaN
min             NaN      NaN
25%             NaN      NaN
50%             NaN      NaN
75%             NaN      NaN
max             NaN      NaN
```

[25]:
```
#All the ladies
no_showF = noshow[noshow['Gender'] == 'F']
```

[26]:
```
#All the gents
no_showM = noshow[noshow['Gender'] == 'M']
```

[27]:
```
MvF = noshow.groupby(['Gender', 'No-show']).mean()
```

[28]:
```
MvF
```

[28]:
```
                      Age  Scholarship  Hipertension  Diabetes  Alcoholism  \
Gender No-show
F      0        36.162190     0.144306      0.182061  0.069686    0.021105
       1        39.590417     0.117860      0.221535  0.080163    0.015984
M      0        30.833010     0.061100      0.144337  0.053463    0.047767
       1        34.461372     0.049609      0.172696  0.062141    0.057102

                   Handcap  SMS_received
Gender No-show
F      0         0.018569      0.460463
       1         0.019792      0.305384
M      0         0.023560      0.396634
       1         0.028196      0.265358
```

[29]:
```
gender_count = noshow.groupby(['Gender', 'No-show']).count()
gender_count
```

[29]:
```
                ScheduledDay  AppointmentDay    Age  Neighbourhood  \
Gender No-show
F      0               14594           14594  14594          14594
       1               57246           57246  57246          57246
M      0                7725            7725   7725           7725
       1               30962           30962  30962          30962

                Scholarship  Hipertension  Diabetes  Alcoholism  Handcap  \
Gender No-show
F      0              14594         14594     14594       14594    14594
       1              57246         57246     57246       57246    57246
M      0               7725          7725      7725        7725     7725
       1              30962         30962     30962       30962    30962

                SMS_received
Gender No-show
F      0               14594
       1               57246
M      0                7725
       1               30962
```

[30]:
```
gender_total = noshow.groupby('Gender').count()
gender_total
```

[30]:
```
        ScheduledDay  AppointmentDay    Age  Neighbourhood  Scholarship  \
Gender
F              71840           71840  71840          71840        71840
M              38687           38687  38687          38687        38687
```

```
        Hipertension  Diabetes  Alcoholism  Handcap  SMS_received  No-show
Gender
F              71840     71840       71840    71840         71840    71840
M              38687     38687       38687    38687         38687    38687
```

[31]:
```
proportion = gender_count/gender_total
proportion
```

[31]:
```
                     Age  Alcoholism  AppointmentDay  Diabetes   Handcap  \
Gender No-show
F      0        0.203146    0.203146        0.203146  0.203146  0.203146
       1        0.796854    0.796854        0.796854  0.796854  0.796854
M      0        0.199679    0.199679        0.199679  0.199679  0.199679
       1        0.800321    0.800321        0.800321  0.800321  0.800321

                Hipertension  Neighbourhood  No-show  SMS_received  \
Gender No-show
F      0            0.203146       0.203146      NaN      0.203146
       1            0.796854       0.796854      NaN      0.796854
M      0            0.199679       0.199679      NaN      0.199679
       1            0.800321       0.800321      NaN      0.800321

                ScheduledDay  Scholarship
Gender No-show
F      0            0.203146     0.203146
       1            0.796854     0.796854
M      0            0.199679     0.199679
       1            0.800321     0.800321
```

[ ]:
```
Patients with cancer: 0.105
Patients without cancer: 0.895
Patients with cancer who tested positive: 0.905
Patients with cancer who tested negative: 0.095
Patients without cancer who tested positive: 0.204
Patients without cancer who tested negative: 0.796
```