call centre

February 24, 2024

[1]:
```
import pandas as pd
```

[2]:
```
call_centre = pd.read_excel(r"D:\Data\01 Call-Center-Dataset.xlsx")
```

[3]:
```
call_centre
```

[3]:
```
     Call Id    Agent       Date      Time             Topic Answered (Y/N)  \
0     ID0001    Diane  2021-01-01  09:12:58   Contract related              Y
1     ID0002    Becky  2021-01-01  09:12:58  Technical Support              Y
2     ID0003  Stewart  2021-01-01  09:47:31   Contract related              Y
3     ID0004     Greg  2021-01-01  09:47:31   Contract related              Y
4     ID0005    Becky  2021-01-01  10:00:29    Payment related              Y
…      …      …         …       …                 …            …
4995  ID4996      Jim  2021-03-31  16:37:55    Payment related              Y
4996  ID4997    Diane  2021-03-31  16:45:07    Payment related              Y
4997  ID4998    Diane  2021-03-31  16:53:46    Payment related              Y
4998  ID4999      Jim  2021-03-31  17:02:24          Streaming              Y
4999  ID5000    Diane  2021-03-31  17:39:50   Contract related              N

     Resolved  Speed of answer in seconds AvgTalkDuration  Satisfaction rating
0           Y                       109.0        00:02:23                  3.0
1           N                        70.0        00:04:02                  3.0
2           Y                        10.0        00:02:11                  3.0
3           Y                        53.0        00:00:37                  2.0
4           Y                        95.0        00:01:00                  3.0
…       …                          …             …                  …
4995        Y                        22.0        00:05:40                  1.0
4996        Y                       100.0        00:03:16                  3.0
4997        Y                        84.0        00:01:49                  4.0
4998        Y                        98.0        00:00:58                  5.0
4999        N                         NaN             NaN                  NaN

[5000 rows x 10 columns]
```

[4]:
```
call_centre =call_centre.drop('Call Id', axis = 1)
```

[5]:
```
call_centre
```

[5]:

```
        Agent        Date      Time             Topic Answered (Y/N)  \
0       Diane  2021-01-01  09:12:58   Contract related              Y
1       Becky  2021-01-01  09:12:58  Technical Support              Y
2     Stewart  2021-01-01  09:47:31   Contract related              Y
3        Greg  2021-01-01  09:47:31   Contract related              Y
4       Becky  2021-01-01  10:00:29    Payment related              Y
…       …         …       …                …            …
4995      Jim  2021-03-31  16:37:55    Payment related              Y
4996    Diane  2021-03-31  16:45:07    Payment related              Y
4997    Diane  2021-03-31  16:53:46    Payment related              Y
4998      Jim  2021-03-31  17:02:24          Streaming              Y
4999    Diane  2021-03-31  17:39:50   Contract related              N

     Resolved  Speed of answer in seconds AvgTalkDuration  Satisfaction rating
0           Y                       109.0        00:02:23                  3.0
1           N                        70.0        00:04:02                  3.0
2           Y                        10.0        00:02:11                  3.0
3           Y                        53.0        00:00:37                  2.0
4           Y                        95.0        00:01:00                  3.0
…       …                         …             …                  …
4995        Y                        22.0        00:05:40                  1.0
4996        Y                       100.0        00:03:16                  3.0
4997        Y                        84.0        00:01:49                  4.0
4998        Y                        98.0        00:00:58                  5.0
4999        N                         NaN             NaN                  NaN

[5000 rows x 9 columns]
```

[6]:

```
call_centre.dtypes
```

[6]:

```
Agent                          object
Date                           object
Time                           object
Topic                          object
Answered (Y/N)                 object
Resolved                       object
Speed of answer in seconds    float64
AvgTalkDuration                object
Satisfaction rating           float64
dtype: object
```

[7]:

```
call_centre.describe(include= 'all')
```

[7]:

```
       Agent         Date      Time      Topic Answered (Y/N) Resolved  \
count   5000         5000      5000       5000           5000     5000
unique     8           90       375          5              2        2
top      Jim   2021-01-11  11:55:41  Streaming              Y        Y
```

```
freq     666          84        30       1022           4054     3646
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

        Speed of answer in seconds AvgTalkDuration  Satisfaction rating
count                  4054.000000            4054          4054.000000
unique                         NaN             391                  NaN
top                            NaN        00:04:43                  NaN
freq                           NaN              22                  NaN
mean                     67.520720             NaN             3.403552
std                      33.592872             NaN             1.212220
min                      10.000000             NaN             1.000000
25%                      39.000000             NaN             3.000000
50%                      68.000000             NaN             3.000000
75%                      97.000000             NaN             4.000000
max                     125.000000             NaN             5.000000
```

```
[8]: call_centre.isnull().sum()
```

```
[8]: Agent                           0
     Date                            0
     Time                            0
     Topic                           0
     Answered (Y/N)                  0
     Resolved                        0
     Speed of answer in seconds    946
     AvgTalkDuration               946
     Satisfaction rating           946
     dtype: int64
```

```
[9]: call_centre.groupby(['Agent','Answered (Y/N)'])['Answered (Y/N)'].count()
```

```
[9]: Agent    Answered (Y/N)
     Becky    N                 114
              Y                 517
     Dan      N                 110
              Y                 523
     Diane    N                 132
              Y                 501
     Greg     N                 122
              Y                 502
     Jim      N                 130
```

```
         Y                  536
Joe      N                  109
         Y                  484
Martha   N                  124
         Y                  514
Stewart  N                  105
         Y                  477
Name: Answered (Y/N), dtype: int64
```

[10]:

```
call_centre.groupby(['Agent','Answered (Y/N)','Resolved'])[['Resolved']].count()
```

[10]:

```
                                 Resolved
Agent   Answered (Y/N) Resolved
Becky   N              N              114
        Y              N               55
                       Y              462
Dan     N              N              110
        Y              N               52
                       Y              471
Diane   N              N              132
        Y              N               49
                       Y              452
Greg    N              N              122
        Y              N               47
                       Y              455
Jim     N              N              130
        Y              N               51
                       Y              485
Joe     N              N              109
        Y              N               48
                       Y              436
Martha  N              N              124
        Y              N               53
                       Y              461
Stewart N              N              105
        Y              N               53
                       Y              424
```

[11]:

```
call_centre['Answered (Y/N)'].value_counts()
```

[11]:

```
Answered (Y/N)
Y    4054
N     946
Name: count, dtype: int64
```

[12]:

```
call_centre['Resolved'].value_counts()
```

```
[12]: Resolved
      Y    3646
      N    1354
      Name: count, dtype: int64
```

```
[13]: call_centre['Agent'].value_counts()
```

```
[13]: Agent
      Jim        666
      Martha     638
      Diane      633
      Dan        633
      Becky      631
      Greg       624
      Joe        593
      Stewart    582
      Name: count, dtype: int64
```

```
[14]: call_centre['Speed of answer in seconds'].sum()
```

```
[14]: 273729.0
```

```
[15]: call_centre['Speed of answer in seconds'].median()
```

```
[15]: 68.0
```

```
[16]: call_centre.groupby(['Agent'])[['Resolved']].count()
```

```
[16]:          Resolved
      Agent
      Becky         631
      Dan           633
      Diane         633
      Greg          624
      Jim           666
      Joe           593
      Martha        638
      Stewart       582
```

```
[17]: call_centre.groupby('Agent')[['Speed of answer in seconds']].sum()
```

```
[17]:          Speed of answer in seconds
      Agent
      Becky                       33776.0
      Dan                         35189.0
      Diane                       33200.0
      Greg                        34359.0
      Jim                         35560.0
```

```
Joe                         34358.0
Martha                      35717.0
Stewart                     31570.0
```

```
[18]: call_centre.groupby('Agent')[['Speed of answer in seconds']].median()
```

```
[18]:          Speed of answer in seconds
      Agent
      Becky                          64.0
      Dan                            67.0
      Diane                          64.0
      Greg                           70.0
      Jim                            65.0
      Joe                            72.0
      Martha                         70.0
      Stewart                        65.0
```

```
[19]: call_centre.groupby('Agent')[['Satisfaction rating']].median()
```

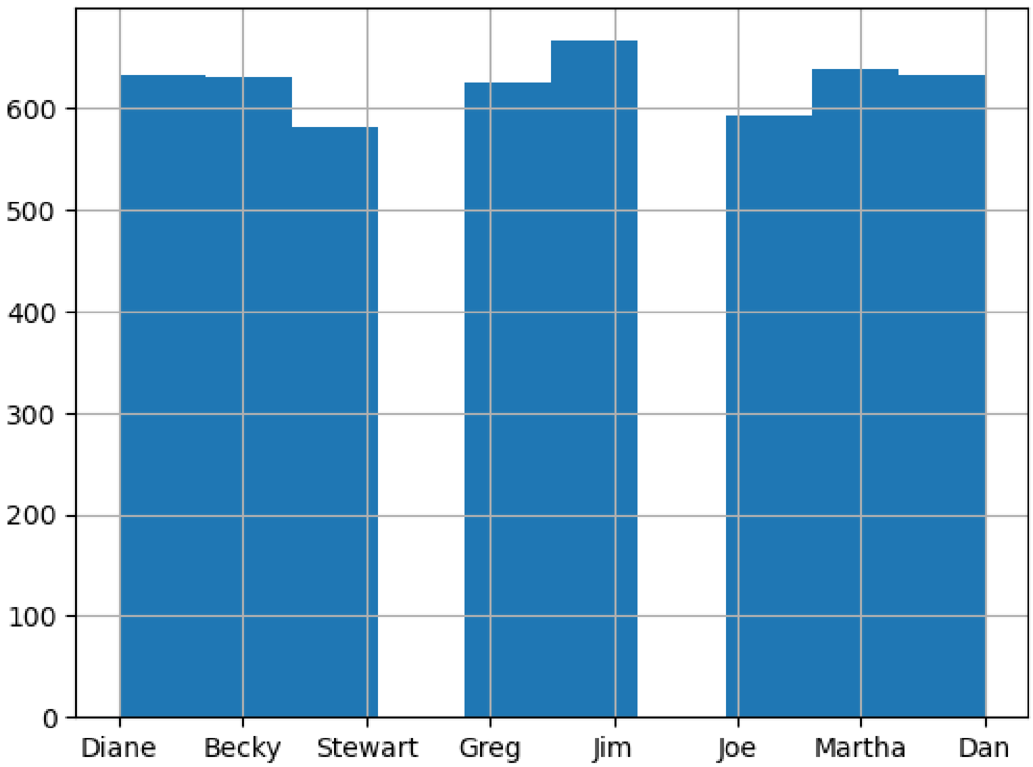
```
[19]:          Satisfaction rating
      Agent
      Becky                     4.0
      Dan                       3.0
      Diane                     3.0
      Greg                      3.0
      Jim                       3.5
      Joe                       3.0
      Martha                    4.0
      Stewart                   4.0
```

```
[20]: call_centre.groupby('Agent')[['Satisfaction rating']].median()
```

```
[20]:          Satisfaction rating
      Agent
      Becky                     4.0
      Dan                       3.0
      Diane                     3.0
      Greg                      3.0
      Jim                       3.5
      Joe                       3.0
      Martha                    4.0
      Stewart                   4.0
```

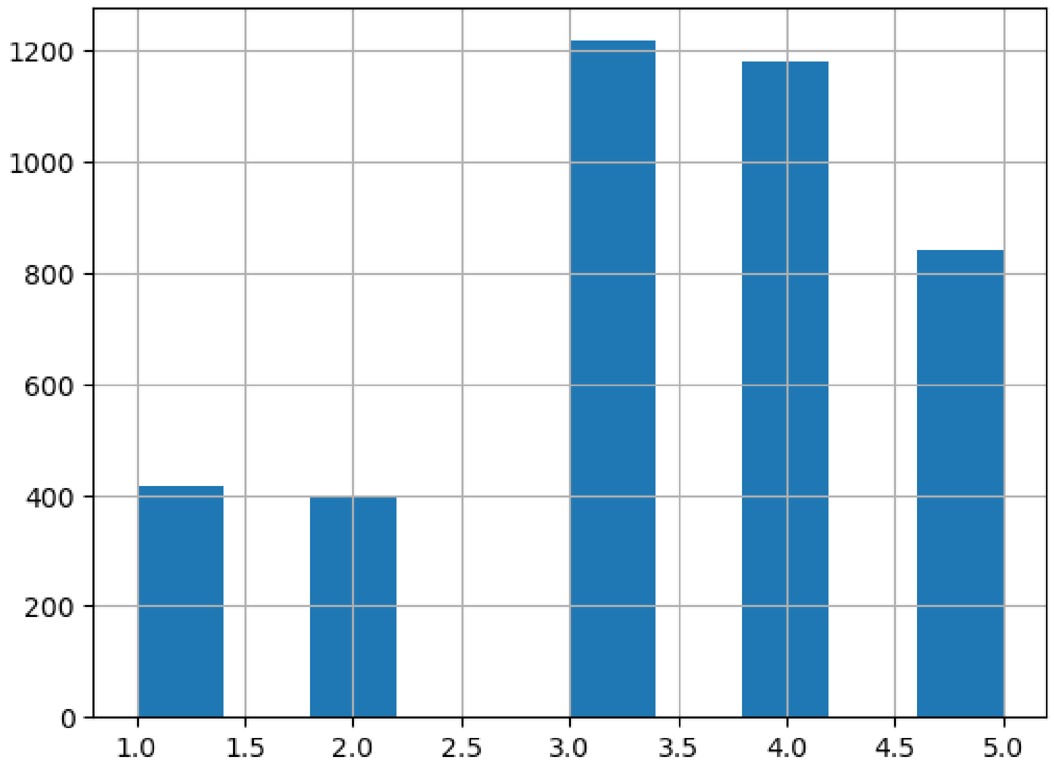
```
[21]: call_centre['Agent'].hist()
```

```
[21]: <Axes: >
```



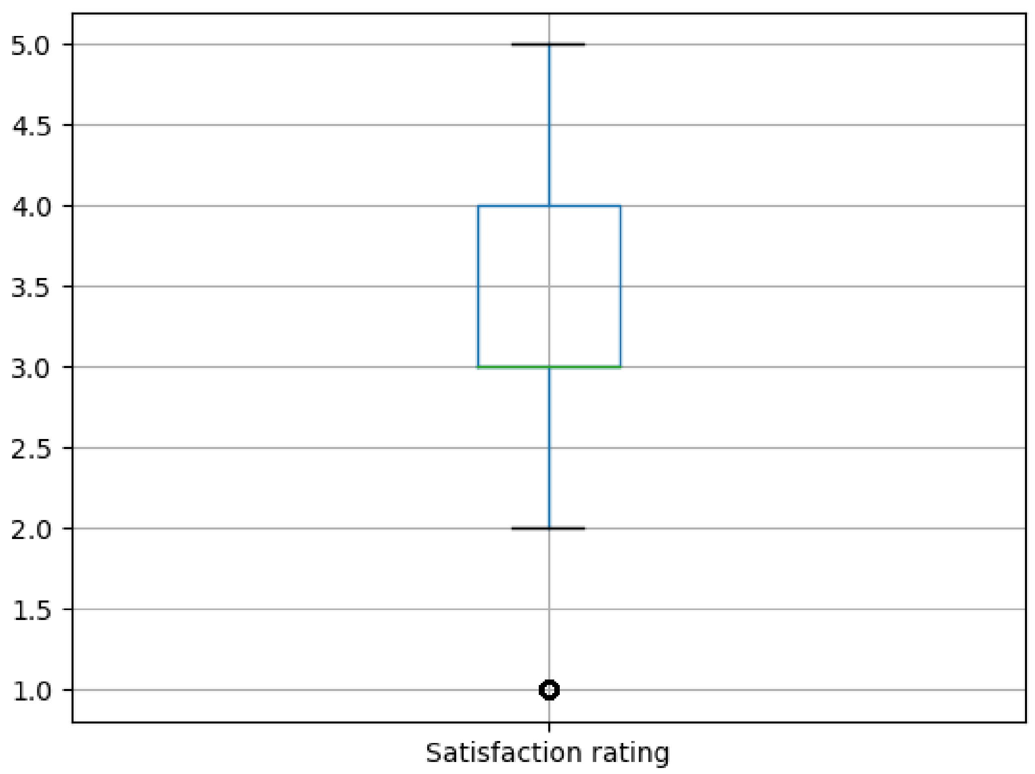
```
[22]: call_centre['Satisfaction rating'].hist()
```

```
[22]: <Axes: >
```



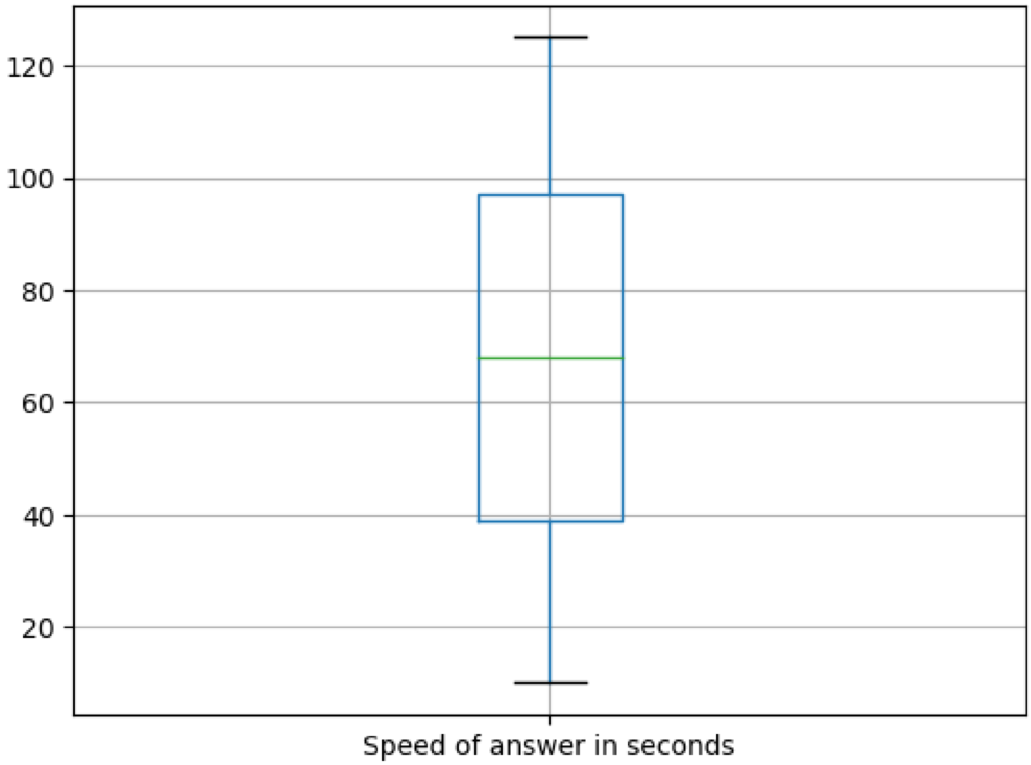
```
[23]: call_centre.boxplot('Satisfaction rating')
```

```
[23]: <Axes: >
```



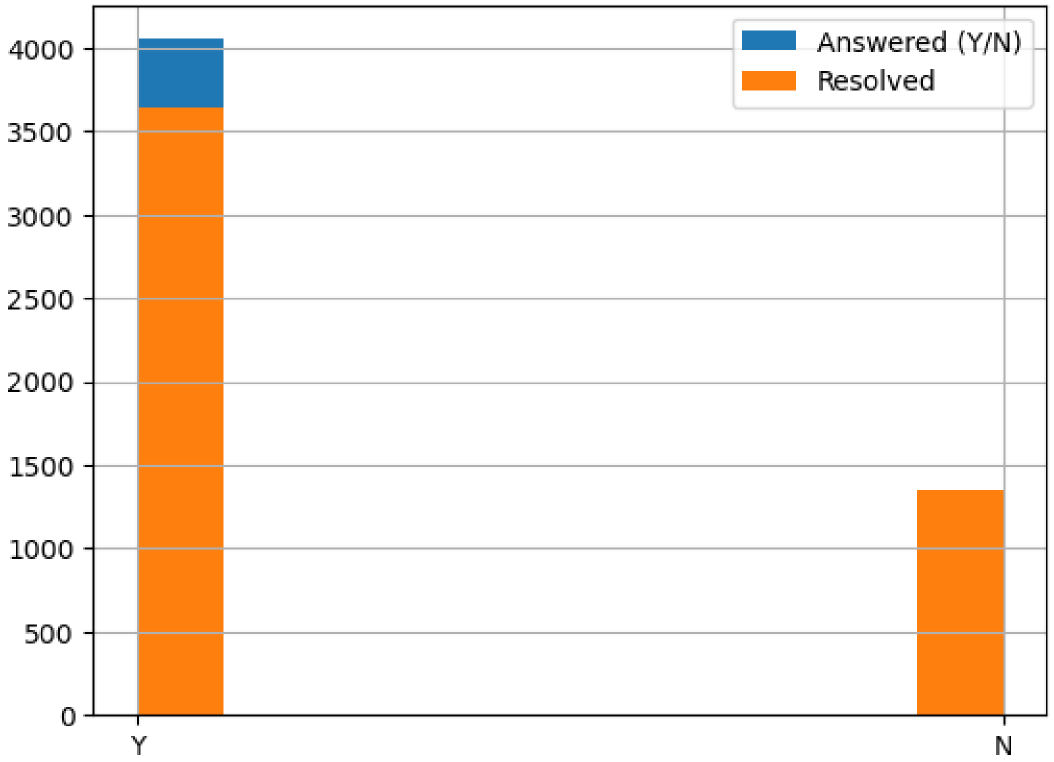
```
[24]: call_centre.boxplot('Speed of answer in seconds')
```

```
[24]: <Axes: >
```



```
[25]: call_centre['Answered (Y/N)'].hist(legend = True)
      call_centre['Resolved'].hist(legend = True)
```

```
[25]: <Axes: >
```



```
[26]: call_centre['Time'] = pd.to_datetime(call_centre['Time'],format='%H:%M:%S')
```

```
[27]: call_centre['Time'].median()
```

```
[27]: Timestamp('1900-01-01 13:29:17')
```

```
[28]: call_centre['AvgTalkDuration'] = pd.
       ↪to_datetime(call_centre['AvgTalkDuration'],format='%H:%M:%S')
```

```
[29]: call_centre['AvgTalkDuration'].median()
```

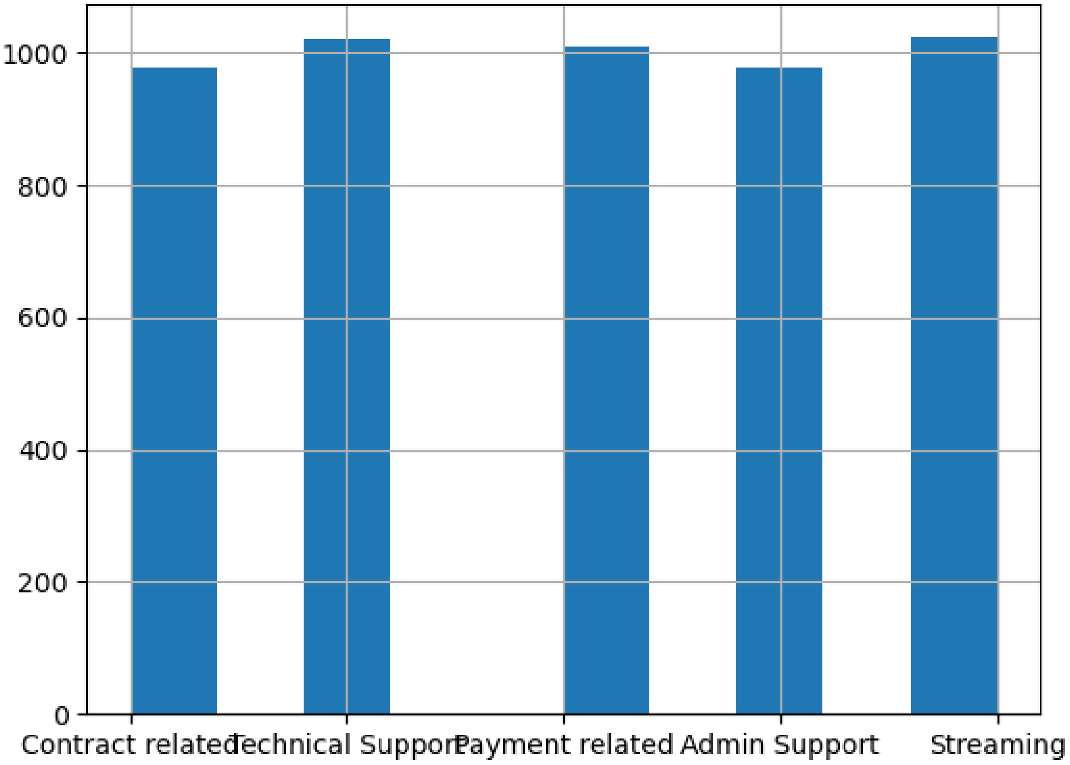
```
[29]: Timestamp('1900-01-01 00:03:46')
```

```
[30]: call_centre['Topic'].value_counts()
```

```
[30]: Topic
      Streaming            1022
      Technical Support    1019
      Payment related      1007
      Contract related      976
      Admin Support         976
      Name: count, dtype: int64
```

```
[31]: call_centre['Topic'].hist()
```

```
[31]: <Axes: >
```



```
[32]: call_centre[call_centre['Satisfaction rating']== 5 ]
```

```
[32]:
       Agent        Date                Time              Topic  \
16      Greg  2021-01-01 1900-01-01 11:57:07  Technical Support
20       Jim  2021-01-01 1900-01-01 12:02:53  Technical Support
21       Dan  2021-01-01 1900-01-01 12:02:53      Admin Support
34     Becky  2021-01-01 1900-01-01 13:42:14          Streaming
40       Jim  2021-01-01 1900-01-01 14:47:02  Technical Support
…      …         …                 …                …
4971     Joe  2021-03-30 1900-01-01 17:25:26   Contract related
4974     Dan  2021-03-31 1900-01-01 09:10:05  Technical Support
4979   Becky  2021-03-31 1900-01-01 10:49:26      Admin Support
4988  Martha  2021-03-31 1900-01-01 13:22:05   Contract related
4998     Jim  2021-03-31 1900-01-01 17:02:24          Streaming

     Answered (Y/N) Resolved  Speed of answer in seconds     AvgTalkDuration  \
16                Y        Y                        45.0 1900-01-01 00:05:32
20                Y        Y                        74.0 1900-01-01 00:05:22
```

```
21                Y        Y                          89.0 1900-01-01 00:05:50
34                Y        Y                         119.0 1900-01-01 00:01:31
40                Y        Y                          27.0 1900-01-01 00:02:09
…             …      …                         …                 …
4971              Y        Y                          42.0 1900-01-01 00:00:35
4974              Y        Y                         102.0 1900-01-01 00:02:26
4979              Y        Y                          37.0 1900-01-01 00:04:19
4988              Y        Y                          62.0 1900-01-01 00:05:49
4998              Y        Y                          98.0 1900-01-01 00:00:58

      Satisfaction rating
16                    5.0
20                    5.0
21                    5.0
34                    5.0
40                    5.0
…                   …
4971                  5.0
4974                  5.0
4979                  5.0
4988                  5.0
4998                  5.0

[843 rows x 9 columns]
```

```
[33]: call_centre.groupby(['Topic','Answered (Y/N)'])[['Answered (Y/N)']].count()
```

```
[33]:                               Answered (Y/N)
      Topic             Answered (Y/N)
      Admin Support     N                          181
                        Y                          795
      Contract related  N                          187
                        Y                          789
      Payment related   N                          189
                        Y                          818
      Streaming         N                          175
                        Y                          847
      Technical Support N                          214
                        Y                          805
```

```
[34]: call_centre.groupby(['Topic','Answered (Y/N)','Resolved'])[['Answered (Y/N)']].
        ↪count()
```

```
[34]:                                        Answered (Y/N)
      Topic             Answered (Y/N) Resolved
      Admin Support     N              N                    181
                        Y              N                     72
```

```
                              Y                     723
Contract related  N           N                     187
                  Y           N                      80
                              Y                     709
Payment related   N           N                     189
                  Y           N                      89
                              Y                     729
Streaming         N           N                     175
                  Y           N                      98
                              Y                     749
Technical Support N           N                     214
                  Y           N                      69
                              Y                     736
```

[35]:
```
call_centre[call_centre['Resolved']== 'Y'][['Agent']]
```

[35]:
```
        Agent
0       Diane
2     Stewart
3        Greg
4       Becky
6       Diane
…         …
4990  Stewart
4995      Jim
4996    Diane
4997    Diane
4998      Jim

[3646 rows x 1 columns]
```

[36]:
```
m = call_centre[call_centre['Satisfaction rating']== 5 ]
```

[37]:
```
m.groupby('Agent')[['Satisfaction rating']].count()
```

[37]:
```
         Satisfaction rating
Agent
Becky                    101
Dan                      118
Diane                    107
Greg                     105
Jim                      111
Joe                       89
Martha                   112
Stewart                  100
```

```
[38]: import matplotlib.pyplot as plt
      import seaborn as sns
```

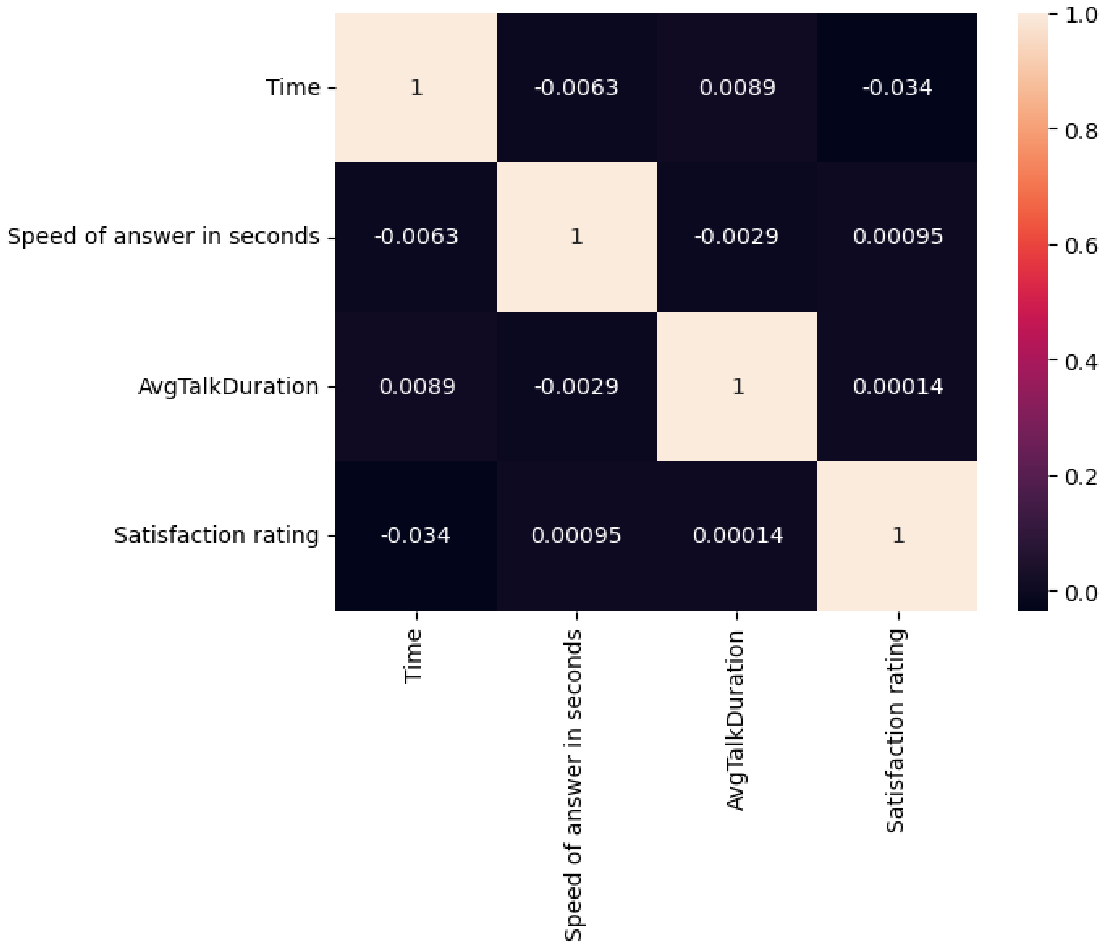
```
[43]: call_centre[['Time','Speed of answer in␣
       ↪seconds','AvgTalkDuration','Satisfaction rating']].corr()
```

```
[43]:                                Time  Speed of answer in seconds  \
      Time                        1.000000                   -0.006302
      Speed of answer in seconds -0.006302                    1.000000
      AvgTalkDuration             0.008891                   -0.002898
      Satisfaction rating        -0.033811                    0.000952

                                  AvgTalkDuration  Satisfaction rating
      Time                               0.008891            -0.033811
      Speed of answer in seconds        -0.002898             0.000952
      AvgTalkDuration                    1.000000             0.000143
      Satisfaction rating                0.000143             1.000000
```

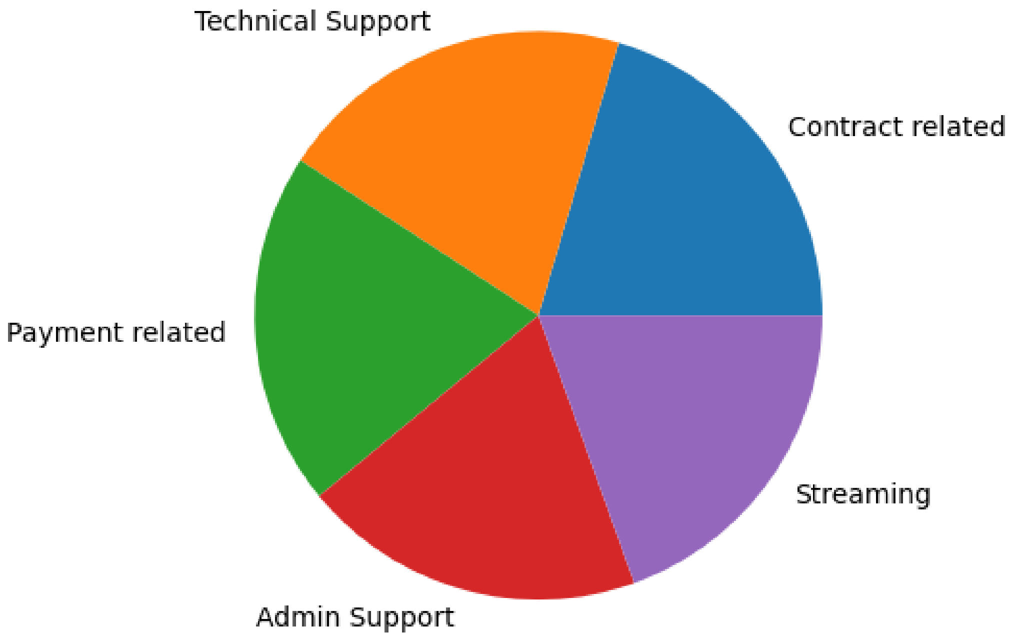
```
[45]: sns.heatmap(call_centre[['Time','Speed of answer in␣
       ↪seconds','AvgTalkDuration','Satisfaction rating']].corr(), annot= True)
```

```
[45]: <Axes: >
```



```
[58]: plt.pie(call_centre['Topic'].value_counts(), labels = call_centre['Topic'].
       ↪unique())
```

```
[58]: ([<matplotlib.patches.Wedge at 0x18b3e7dae90>,
        <matplotlib.patches.Wedge at 0x18b3de02550>,
        <matplotlib.patches.Wedge at 0x18b3e8024d0>,
        <matplotlib.patches.Wedge at 0x18b3e831690>,
        <matplotlib.patches.Wedge at 0x18b3e848590>],
       [Text(0.8808964967586289, 0.6588029766162073, 'Contract related'),
        Text(-0.38105295546897955, 1.0318908106618432, 'Technical Support'),
        Text(-1.098280553382501, -0.0614802900271908, 'Payment related'),
        Text(-0.2922598624495957, -1.0604641308412763, 'Admin Support'),
        Text(0.8995670446667726, -0.6330711904276555, 'Streaming')])
```


Technical Support
Contract related
Payment related
Streaming
Admin Support

[ ]: