Untitled11

February 7, 2024

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

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

[3]:
```
cs
```

[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]:
```
cs.info
```

[4]:
```
<bound method DataFrame.info of      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]>
```

[5]:
```
cs['Answered (Y/N)'].value_counts()
```

[5]:
```
Answered (Y/N)
Y    4054
N     946
Name: count, dtype: int64
```

[6]:
```
cs.dtypes
```

[6]:
```
Call Id                        object
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]: cs['Date'] = pd.to_datetime(cs['Date'], format='%Y-%m-%d')
```

```
[8]: cs.dtypes
```

```
[8]: Call Id                                    object
     Agent                                      object
     Date                               datetime64[ns]
     Time                                       object
     Topic                                      object
     Answered (Y/N)                             object
     Resolved                                   object
     Speed of answer in seconds                float64
     AvgTalkDuration                            object
     Satisfaction rating                       float64
     dtype: object
```

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

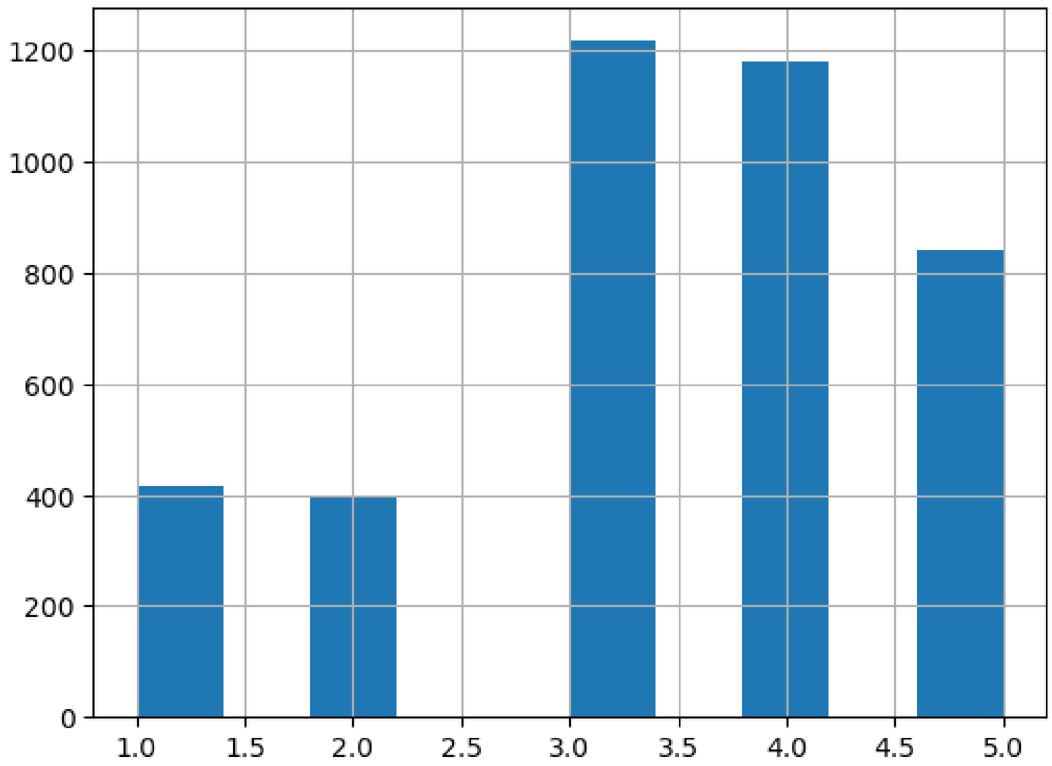
```
[10]: cs.dtypes
```

```
[10]: Call Id                                    object
      Agent                                      object
      Date                               datetime64[ns]
      Time                               datetime64[ns]
      Topic                                      object
      Answered (Y/N)                             object
      Resolved                                   object
      Speed of answer in seconds                float64
      AvgTalkDuration                            object
      Satisfaction rating                       float64
      dtype: object
```

1 detecting the outliers

```
[11]: cs['Satisfaction rating'].hist()
```

```
[11]: <Axes: >
```



[12]:

```
mean = cs['Satisfaction rating'].mean()
```

[13]:

```
std = cs['Satisfaction rating'].std()
```

[14]:

```
lower_limit = cs['Satisfaction rating'].mean() - 1 * std
upper_limit = cs['Satisfaction rating'].mean() + 1 * std
print(lower_limit ,upper_limit)
print()
print(((cs['Satisfaction rating']>= lower_limit )&(cs['Satisfaction rating']<=␣
  ↪upper_limit)).mean())
```

```
2.191332328392151 4.615771766329112

0.4796
```

[15]:

```
lower_limit = cs['Satisfaction rating'].mean() - 2 * std
upper_limit = cs['Satisfaction rating'].mean() + 2 * std
print(lower_limit ,upper_limit)
print()
print(((cs['Satisfaction rating']>= lower_limit )&(cs['Satisfaction rating']<=␣
  ↪upper_limit)).mean())
```

```
0.9791126094236708 5.827991485297591
```

```
0.8108
```

```
[16]: lower_limit = cs['Satisfaction rating'].mean() - 3 * std
      upper_limit = cs['Satisfaction rating'].mean() + 3 * std
      print(lower_limit ,upper_limit)
      print()
      print(((cs['Satisfaction rating']>= lower_limit )&(cs['Satisfaction rating']<=␣
       ↪upper_limit)).mean())
```

```
-0.23310710954480918 7.040211204266072

0.8108
```

```
[17]: from scipy import stats
```

```
[18]: cs['z_score'] = stats.zscore(cs['Satisfaction rating'])
```

```
[19]: cs
```

```
[19]:      Call Id    Agent       Date                Time              Topic  \
      0     ID0001    Diane 2021-01-01 1900-01-01 09:12:58   Contract related
      1     ID0002    Becky 2021-01-01 1900-01-01 09:12:58  Technical Support
      2     ID0003  Stewart 2021-01-01 1900-01-01 09:47:31   Contract related
      3     ID0004     Greg 2021-01-01 1900-01-01 09:47:31   Contract related
      4     ID0005    Becky 2021-01-01 1900-01-01 10:00:29    Payment related
      …        …       …         …                   …                 …
      4995  ID4996      Jim 2021-03-31 1900-01-01 16:37:55    Payment related
      4996  ID4997    Diane 2021-03-31 1900-01-01 16:45:07    Payment related
      4997  ID4998    Diane 2021-03-31 1900-01-01 16:53:46    Payment related
      4998  ID4999      Jim 2021-03-31 1900-01-01 17:02:24          Streaming
      4999  ID5000    Diane 2021-03-31 1900-01-01 17:39:50   Contract related

           Answered (Y/N) Resolved  Speed of answer in seconds AvgTalkDuration  \
      0                 Y        Y                       109.0        00:02:23
      1                 Y        N                        70.0        00:04:02
      2                 Y        Y                        10.0        00:02:11
      3                 Y        Y                        53.0        00:00:37
      4                 Y        Y                        95.0        00:01:00
      …              …       …                          …              …
      4995              Y        Y                        22.0        00:05:40
      4996              Y        Y                       100.0        00:03:16
      4997              Y        Y                        84.0        00:01:49
      4998              Y        Y                        98.0        00:00:58
      4999              N        N                         NaN             NaN

            Satisfaction rating  z_score
      0                     3.0      NaN
```

```
1                      3.0      NaN
2                      3.0      NaN
3                      2.0      NaN
4                      3.0      NaN
…                    …      …
4995                   1.0      NaN
4996                   3.0      NaN
4997                   4.0      NaN
4998                   5.0      NaN
4999                   NaN      NaN

[5000 rows x 11 columns]
```

[20]:
```
cs[(cs['z_score']> 3) | (cs['z_score'] < -3)]
```

[20]:
```
Empty DataFrame
Columns: [Call Id, Agent, Date, Time, Topic, Answered (Y/N), Resolved, Speed of
answer in seconds, AvgTalkDuration, Satisfaction rating, z_score]
Index: []
```

[21]:
```
cs.isna().sum()
```

[21]:
```
Call Id                        0
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
z_score                     5000
dtype: int64
```

[22]:
```
cs.drop_duplicates()
```

[22]:
```
     Call Id    Agent       Date                Time                Topic  \
0     ID0001    Diane 2021-01-01 1900-01-01 09:12:58    Contract related
1     ID0002    Becky 2021-01-01 1900-01-01 09:12:58   Technical Support
2     ID0003  Stewart 2021-01-01 1900-01-01 09:47:31    Contract related
3     ID0004     Greg 2021-01-01 1900-01-01 09:47:31    Contract related
4     ID0005    Becky 2021-01-01 1900-01-01 10:00:29     Payment related
…        …      …          …                   …                  …
4995  ID4996      Jim 2021-03-31 1900-01-01 16:37:55     Payment related
4996  ID4997    Diane 2021-03-31 1900-01-01 16:45:07     Payment related
4997  ID4998    Diane 2021-03-31 1900-01-01 16:53:46     Payment related
```

```
4998  ID4999      Jim 2021-03-31 1900-01-01 17:02:24          Streaming
4999  ID5000    Diane 2021-03-31 1900-01-01 17:39:50   Contract related

     Answered (Y/N) Resolved  Speed of answer in seconds AvgTalkDuration  \
0                 Y        Y                       109.0        00:02:23
1                 Y        N                        70.0        00:04:02
2                 Y        Y                        10.0        00:02:11
3                 Y        Y                        53.0        00:00:37
4                 Y        Y                        95.0        00:01:00
…             …      …                         …             …
4995              Y        Y                        22.0        00:05:40
4996              Y        Y                       100.0        00:03:16
4997              Y        Y                        84.0        00:01:49
4998              Y        Y                        98.0        00:00:58
4999              N        N                         NaN             NaN

      Satisfaction rating  z_score
0                     3.0      NaN
1                     3.0      NaN
2                     3.0      NaN
3                     2.0      NaN
4                     3.0      NaN
…                   …      …
4995                  1.0      NaN
4996                  3.0      NaN
4997                  4.0      NaN
4998                  5.0      NaN
4999                  NaN      NaN

[5000 rows x 11 columns]
```

[23]:
```
cs['Satisfaction rating'].value_counts()
```

[23]:
```
Satisfaction rating
3.0    1218
4.0    1180
5.0     843
1.0     417
2.0     396
Name: count, dtype: int64
```

[24]:
```
jim = cs[cs['Agent']=='Jim'].sample(n=20 , random_state= 8765, replace=True)
```

[25]:
```
jim
```

[25]:
```
     Call Id Agent       Date                Time              Topic  \
387   ID0388   Jim 2021-01-07 1900-01-01 16:40:48          Streaming
```

```
210   ID0211   Jim 2021-01-04 1900-01-01 14:28:19  Technical Support
991   ID0992   Jim 2021-01-18 1900-01-01 12:00:00    Payment related
644   ID0645   Jim 2021-01-11 1900-01-01 15:59:02  Technical Support
2143  ID2144   Jim 2021-02-06 1900-01-01 17:31:12  Technical Support
397   ID0398   Jim 2021-01-07 1900-01-01 17:55:41    Payment related
2467  ID2468   Jim 2021-02-12 1900-01-01 13:07:41      Admin Support
3123  ID3124   Jim 2021-02-24 1900-01-01 15:21:36          Streaming
1334  ID1335   Jim 2021-01-24 1900-01-01 17:25:26    Payment related
1585  ID1586   Jim 2021-01-29 1900-01-01 10:29:17    Payment related
4189  ID4190   Jim 2021-03-15 1900-01-01 12:10:05    Payment related
3365  ID3366   Jim 2021-02-28 1900-01-01 13:23:31      Admin Support
4138  ID4139   Jim 2021-03-14 1900-01-01 11:52:48   Contract related
3809  ID3810   Jim 2021-03-08 1900-01-01 10:36:29      Admin Support
4984  ID4985   Jim 2021-03-31 1900-01-01 12:50:24    Payment related
1805  ID1806   Jim 2021-02-01 1900-01-01 13:36:29          Streaming
2110  ID2111   Jim 2021-02-06 1900-01-01 11:25:26          Streaming
4952  ID4953   Jim 2021-03-30 1900-01-01 11:24:00          Streaming
1436  ID1437   Jim 2021-01-26 1900-01-01 14:18:14          Streaming
1989  ID1990   Jim 2021-02-04 1900-01-01 12:38:53    Payment related

     Answered (Y/N) Resolved  Speed of answer in seconds AvgTalkDuration  \
387               Y        Y                       119.0        00:03:03
210               N        N                         NaN             NaN
991               Y        Y                       114.0        00:03:10
644               Y        Y                        53.0        00:04:37
2143              Y        Y                        68.0        00:05:42
397               Y        Y                        32.0        00:03:23
2467              N        N                         NaN             NaN
3123              Y        Y                        61.0        00:04:25
1334              Y        Y                        17.0        00:02:35
1585              Y        Y                        94.0        00:01:12
4189              N        N                         NaN             NaN
3365              N        N                         NaN             NaN
4138              Y        Y                        11.0        00:06:28
3809              Y        Y                        34.0        00:05:17
4984              N        N                         NaN             NaN
1805              Y        Y                        37.0        00:03:02
2110              Y        Y                        78.0        00:03:42
4952              Y        Y                        50.0        00:05:21
1436              Y        Y                        77.0        00:04:26
1989              Y        Y                        83.0        00:05:17

      Satisfaction rating  z_score
387                   4.0      NaN
210                   NaN      NaN
991                   4.0      NaN
644                   1.0      NaN
```

```
2143                      5.0       NaN
397                       4.0       NaN
2467                      NaN       NaN
3123                      5.0       NaN
1334                      1.0       NaN
1585                      3.0       NaN
4189                      NaN       NaN
3365                      NaN       NaN
4138                      3.0       NaN
3809                      2.0       NaN
4984                      NaN       NaN
1805                      4.0       NaN
2110                      2.0       NaN
4952                      3.0       NaN
1436                      2.0       NaN
1989                      4.0       NaN
```

[26]:
```
Diane = cs[cs['Agent']=='Diane'].sample(n=20 , random_state= 8765, replace=True)
```

[27]:
```
jimm= (jim['Satisfaction rating']) .mean()
```

[28]:
```
Dianem= (Diane['Satisfaction rating']) .mean()
```

[29]:
```
(Dianem, jimm)
```

[29]:
```
(3.0833333333333335, 3.1333333333333333)
```

[30]:
```
jimm - Dianem
```

[30]:
```
0.04999999999999982
```

[31]:
```
stats.ttest_ind(a = Dianem , b = jimm , equal_var=False)
```

```
C:\Users\User\anaconda3\Lib\site-packages\scipy\stats\_stats_py.py:1103:
RuntimeWarning: divide by zero encountered in divide
  var *= np.divide(n, n-ddof)  # to avoid error on division by zero
C:\Users\User\anaconda3\Lib\site-packages\scipy\stats\_stats_py.py:1103:
RuntimeWarning: invalid value encountered in scalar multiply
  var *= np.divide(n, n-ddof)  # to avoid error on division by zero
```

[31]:
```
TtestResult(statistic=nan, pvalue=nan, df=1.0)
```

[32]:
```
cs.groupby('Agent').agg({'Agent':'count', 'Satisfaction rating':'mean'})
```

[32]:
```
         Agent  Satisfaction rating
Agent
Becky      631             3.371373
Dan        633             3.447419
```

```
Diane      633             3.405190
Greg       624             3.404382
Jim        666             3.393657
Joe        593             3.330579
Martha     638             3.470817
Stewart    582             3.400419
```

```
[33]: Jim = cs[cs['Agent']=='Jim'].sample(n=33, random_state=674657, replace=True)
```

```
[34]: Jim
```

```
[34]:
     Call Id Agent       Date                Time              Topic  \
268   ID0269   Jim 2021-01-05 1900-01-01 15:44:38      Admin Support
2622  ID2623   Jim 2021-02-15 1900-01-01 13:26:24   Contract related
4219  ID4220   Jim 2021-03-15 1900-01-01 17:47:02  Technical Support
2577  ID2578   Jim 2021-02-14 1900-01-01 16:37:55      Admin Support
1821  ID1822   Jim 2021-02-01 1900-01-01 15:28:48      Admin Support
4117  ID4118   Jim 2021-03-13 1900-01-01 17:24:00  Technical Support
2727  ID2728   Jim 2021-02-17 1900-01-01 11:24:00   Contract related
3083  ID3084   Jim 2021-02-23 1900-01-01 15:21:36  Technical Support
2545  ID2546   Jim 2021-02-14 1900-01-01 10:30:43    Payment related
2886  ID2887   Jim 2021-02-20 1900-01-01 11:00:58      Admin Support
3947  ID3948   Jim 2021-03-10 1900-01-01 17:35:31  Technical Support
4316  ID4317   Jim 2021-03-17 1900-01-01 16:52:19  Technical Support
1163  ID1164   Jim 2021-01-21 1900-01-01 16:20:38   Contract related
4422  ID4423   Jim 2021-03-20 1900-01-01 09:57:36  Technical Support
2477  ID2478   Jim 2021-02-12 1900-01-01 17:52:48          Streaming
1847  ID1848   Jim 2021-02-02 1900-01-01 09:44:38   Contract related
2629  ID2630   Jim 2021-02-15 1900-01-01 14:28:19  Technical Support
2691  ID2692   Jim 2021-02-16 1900-01-01 16:14:53          Streaming
4965  ID4966   Jim 2021-03-30 1900-01-01 16:04:48          Streaming
257   ID0258   Jim 2021-01-05 1900-01-01 14:42:43          Streaming
2814  ID2815   Jim 2021-02-18 1900-01-01 14:52:48    Payment related
214   ID0215   Jim 2021-01-04 1900-01-01 14:49:55   Contract related
2334  ID2335   Jim 2021-02-10 1900-01-01 13:04:48    Payment related
1138  ID1139   Jim 2021-01-21 1900-01-01 10:14:53   Contract related
1585  ID1586   Jim 2021-01-29 1900-01-01 10:29:17    Payment related
3974  ID3975   Jim 2021-03-11 1900-01-01 12:02:53      Admin Support
2023  ID2024   Jim 2021-02-04 1900-01-01 17:47:02          Streaming
184   ID0185   Jim 2021-01-04 1900-01-01 10:27:50          Streaming
4440  ID4441   Jim 2021-03-20 1900-01-01 12:24:29          Streaming
1049  ID1050   Jim 2021-01-19 1900-01-01 11:25:26      Admin Support
90    ID0091   Jim 2021-01-02 1900-01-01 13:03:22   Contract related
2384  ID2385   Jim 2021-02-11 1900-01-01 11:55:41    Payment related
4909  ID4910   Jim 2021-03-29 1900-01-01 12:00:00    Payment related

     Answered (Y/N) Resolved  Speed of answer in seconds AvgTalkDuration   \
268               Y        Y                       95.0        00:01:47
2622              N        N                        NaN             NaN
4219              N        N                        NaN             NaN
2577              Y        Y                       40.0        00:06:39
1821              Y        Y                       63.0        00:05:50
4117              N        N                        NaN             NaN
2727              Y        Y                       70.0        00:00:49
3083              N        N                        NaN             NaN
2545              Y        Y                       54.0        00:02:17
2886              N        N                        NaN             NaN
3947              Y        Y                       99.0        00:02:47
4316              N        N                        NaN             NaN
1163              Y        Y                       79.0        00:00:47
4422              Y        Y                       80.0        00:03:07
2477              Y        N                      103.0        00:03:57
1847              Y        Y                       39.0        00:06:59
2629              Y        Y                       14.0        00:01:13
2691              Y        Y                       65.0        00:04:48
4965              N        N                        NaN             NaN
257               Y        N                      110.0        00:02:38
2814              N        N                        NaN             NaN
214               Y        Y                       45.0        00:03:02
2334              Y        Y                       30.0        00:02:21
1138              Y        Y                       73.0        00:03:13
1585              Y        Y                       94.0        00:01:12
3974              Y        Y                       92.0        00:02:30
2023              Y        Y                       89.0        00:06:24
184               Y        Y                       90.0        00:04:17
4440              N        N                        NaN             NaN
1049              Y        Y                       96.0        00:05:39
90                Y        Y                       46.0        00:05:41
2384              Y        Y                       31.0        00:00:47
4909              Y        Y                       92.0        00:01:02

      Satisfaction rating  z_score
268                   5.0      NaN
2622                  NaN      NaN
4219                  NaN      NaN
2577                  5.0      NaN
1821                  1.0      NaN
4117                  NaN      NaN
2727                  4.0      NaN
3083                  NaN      NaN
2545                  2.0      NaN
2886                  NaN      NaN
3947                  3.0      NaN
4316                  NaN      NaN
```

```
1163                    3.0      NaN
4422                    5.0      NaN
2477                    3.0      NaN
1847                    4.0      NaN
2629                    5.0      NaN
2691                    1.0      NaN
4965                    NaN      NaN
257                     3.0      NaN
2814                    NaN      NaN
214                     3.0      NaN
2334                    4.0      NaN
1138                    1.0      NaN
1585                    3.0      NaN
3974                    1.0      NaN
2023                    4.0      NaN
184                     3.0      NaN
4440                    NaN      NaN
1049                    5.0      NaN
90                      1.0      NaN
2384                    4.0      NaN
4909                    3.0      NaN
```

[35]:

```
(Jim['Satisfaction rating']).mean()
```

[35]:

```
3.1666666666666665
```

[36]:

```
(Jim['Satisfaction rating']).std()
```

[36]:

```
1.403928236326068
```

[37]:

```
import numpy as np
confedence_interval = 0.95
z_score = 1.65
standard_error = (cs['Satisfaction rating']==3.0).std() / np.sqrt(cs.shape[0])
margin_of_error = standard_error * z_score
print(standard_error , margin_of_error)
```

```
0.006071176149311366 0.010017440646363753
```

[38]:

```
lower = (cs['Satisfaction rating']==3.0).mean() - margin_of_error
```

[39]:

```
upper = (cs['Satisfaction rating']==3.0).mean() + margin_of_error
```

[40]:

```
print(lower, upper)
```

```
0.23358255935363625 0.25361744064636377
```

[41]:

```
stats.norm.interval(confidence=confedence_interval,loc = (cs['Satisfaction␣
 ↪rating']==3.0).mean() , scale=standard_error)
```

```
[41]: (0.23170071340355117, 0.2554992865964489)
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
```

```
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
[ ]:
```