RATING PREDICTION PROJECT

PROBLEM STATEMENT:

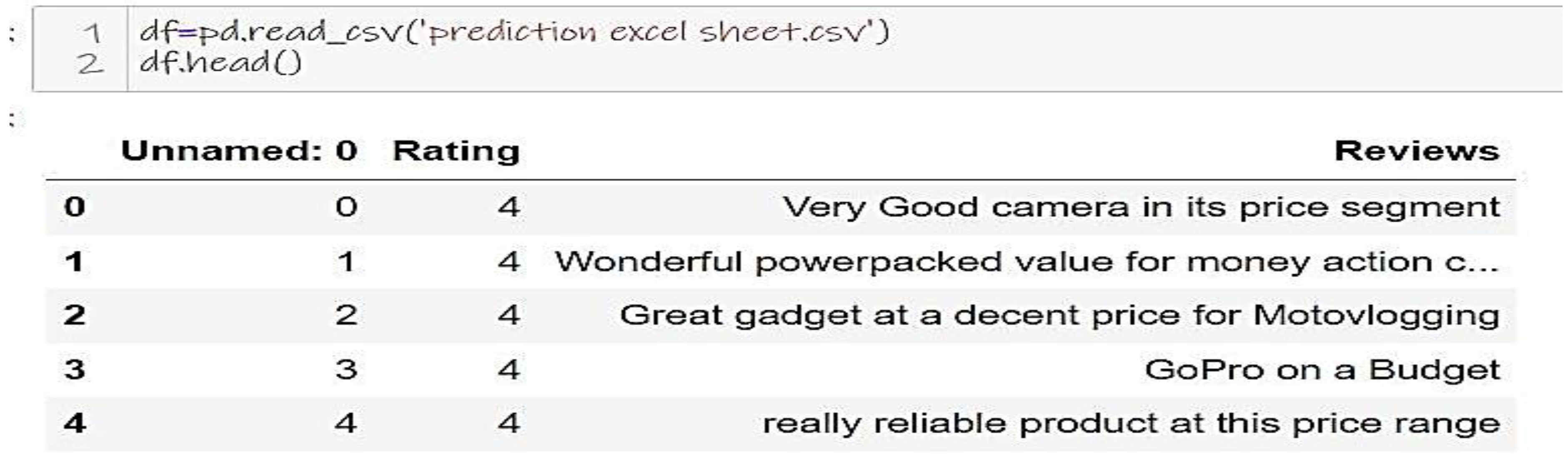
We have a client who has a website where people write different reviews for technical products. Now they are adding a new feature to their website i.e. The reviewer will have to add stars(rating) as well with the review. The rating is out 5 stars and it only has 5 options available 1 star, 2 stars, 3 stars, 4 stars, 5 stars. Now they want to predict ratings for the reviews which were written in the past and they don’t have a rating. So, we have to build an application which can predict the rating by seeing the review.

UNDERSTANDING:

We all know ratings and reviews are important. While they’ve only been around for about two decades, it’s hard to imagine shopping without them. According to consumer research we conducted of 30,000+ global shopping, the majority (**88%**) of shopping use reviews to discover and evaluate products and on based on these the customers easy to find things. The ability to successfully decide whether a review will be helpful to other customers and thus give the product more exposure is vital to companies that support these reviews, e-commercial sites like Amazon, Flipkart, Snapdeal, etc. In this research project, show a new approach to enhance the accuracy of the rating prediction by using machine learning methods the training performance of our model changes as we change the training method, the dataset used for training and the features used in the model. It help to build an application which can predict the rating by seeing the review.

EDA STEPS AND VISUALIZATIONS:

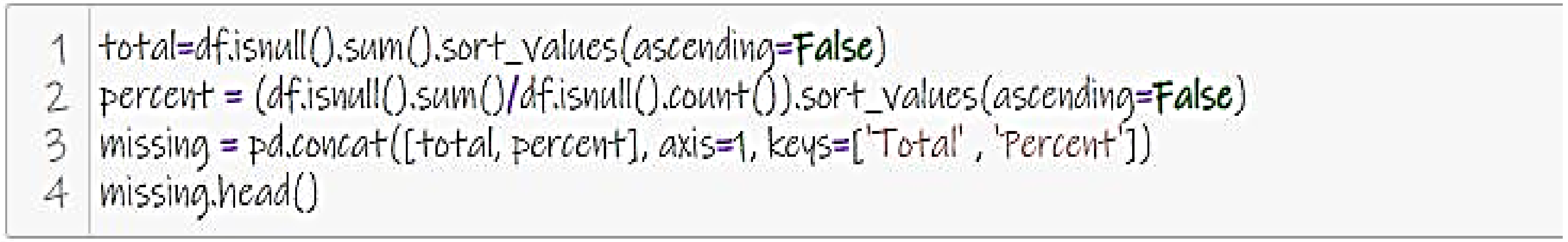
1. Data Collection



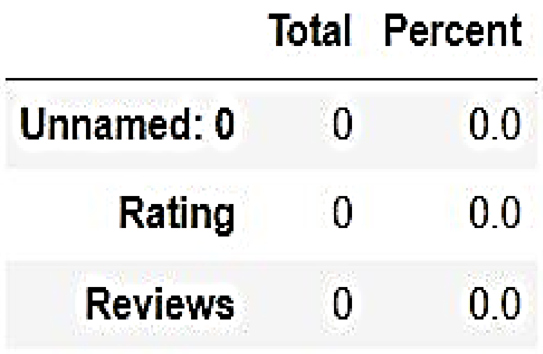
```
df=pd.read_csv('prediction excel sheet.csv')
df.head()
```

```
   Unnamed: 0  Rating                                            Reviews
0           0       4          Very Good camera in its price segment
1           1       4  Wonderful powerpacked value for money action c...
2           2       4    Great gadget at a decent price for Motovlogging
3           3       4                             GoPro on a Budget
4           4       4          really reliable product at this price range
```

2. Data Cleaning

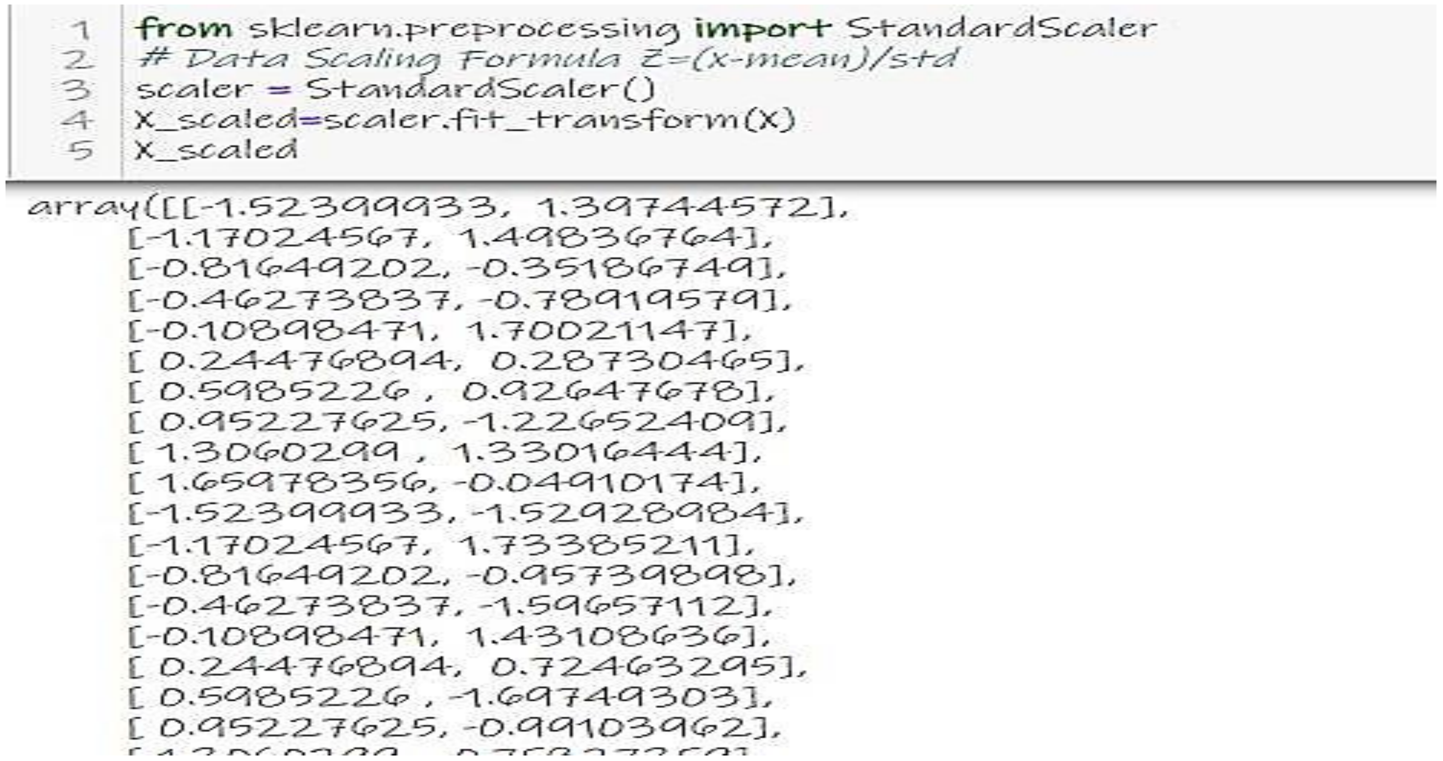


```
total=df.isnull().sum().sort_values(ascending=False)
percent = (df.isnull().sum()/df.isnull().count()).sort_values(ascending=False)
missing = pd.concat([total, percent], axis=1, keys=['Total' , 'Percent'])
missing.head()
```



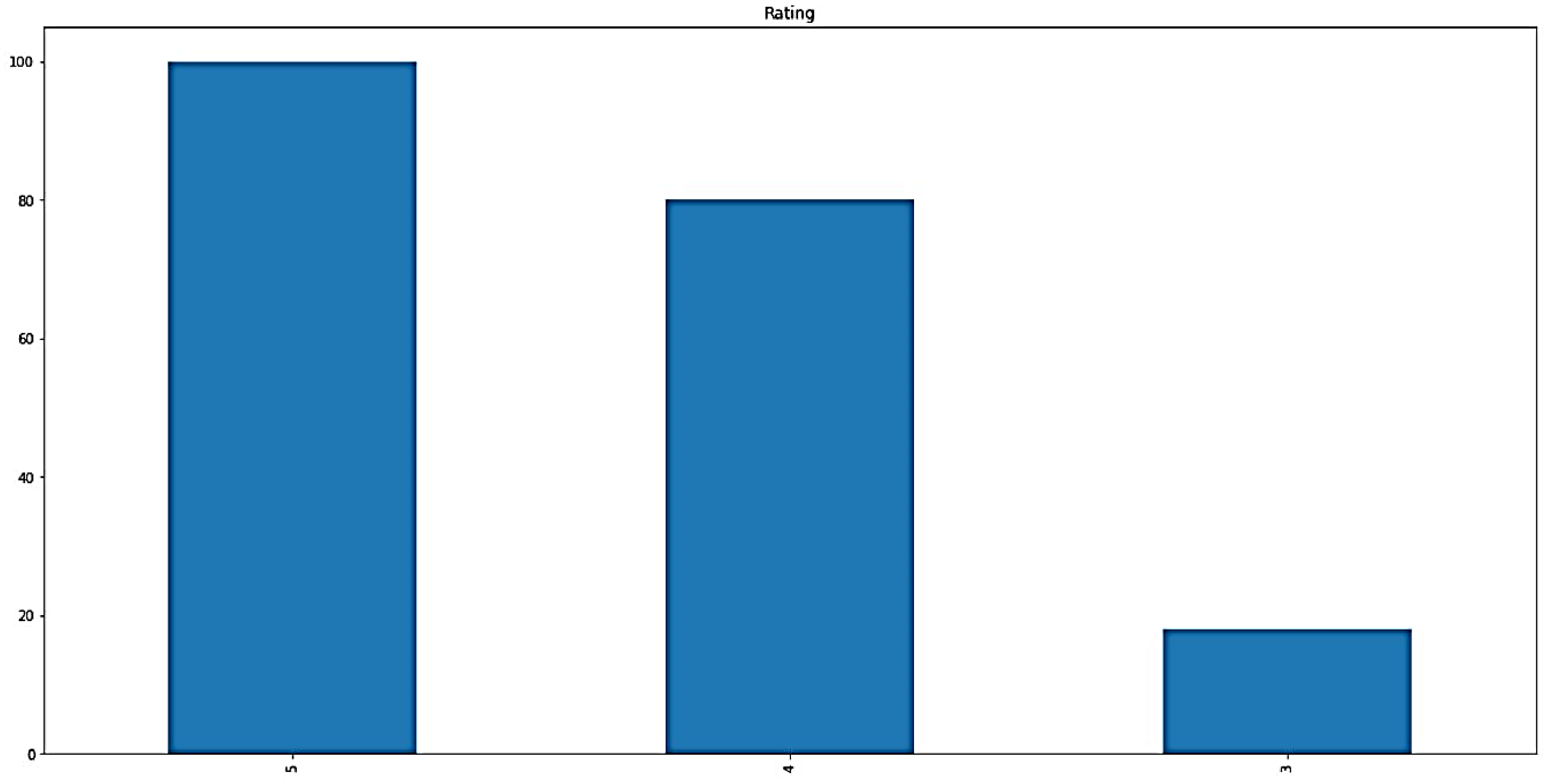
	Total	Percent
Unnamed: 0	0	0.0
Rating	0	0.0
Reviews	0	0.0

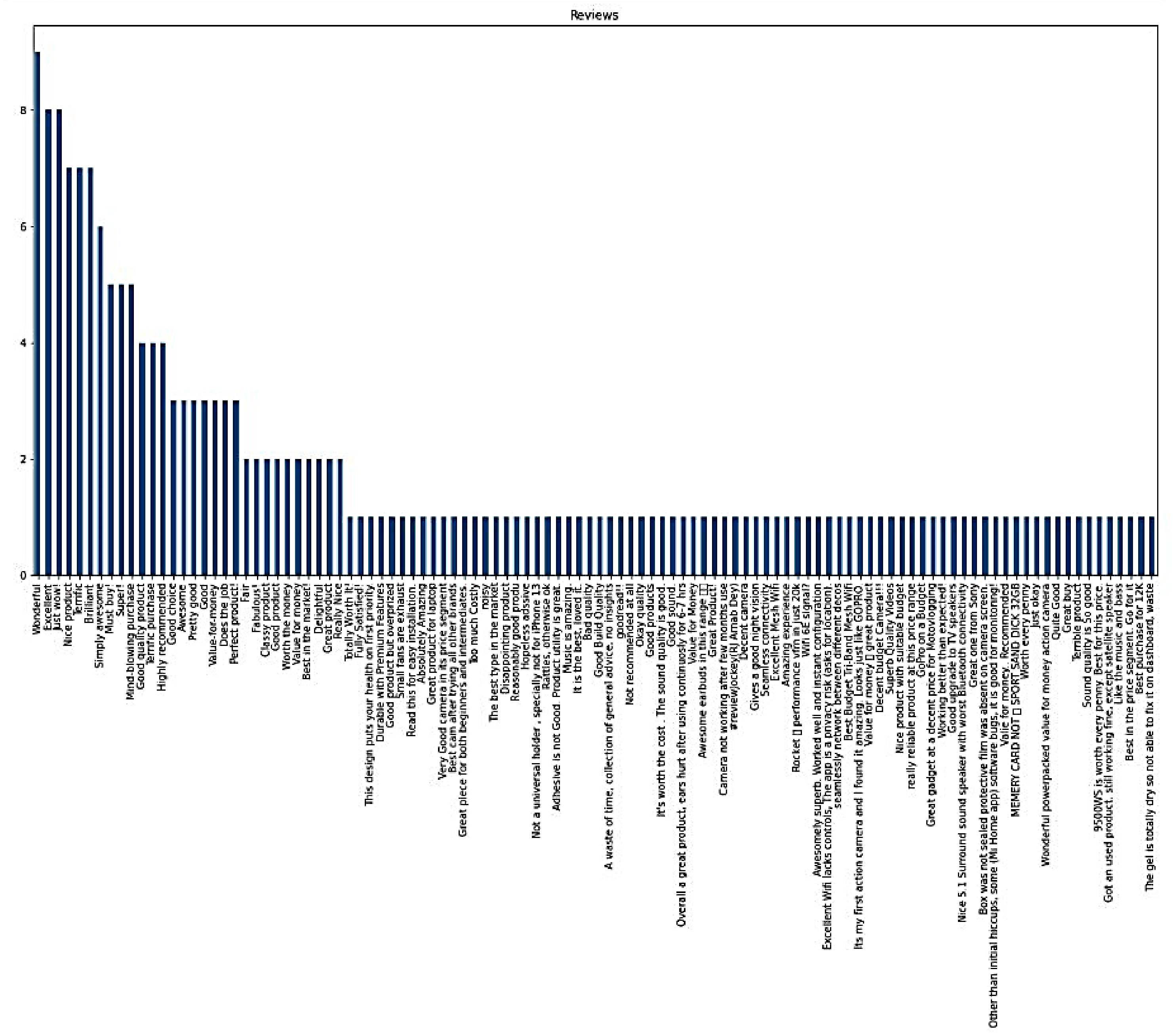
3. Univariate Analysis



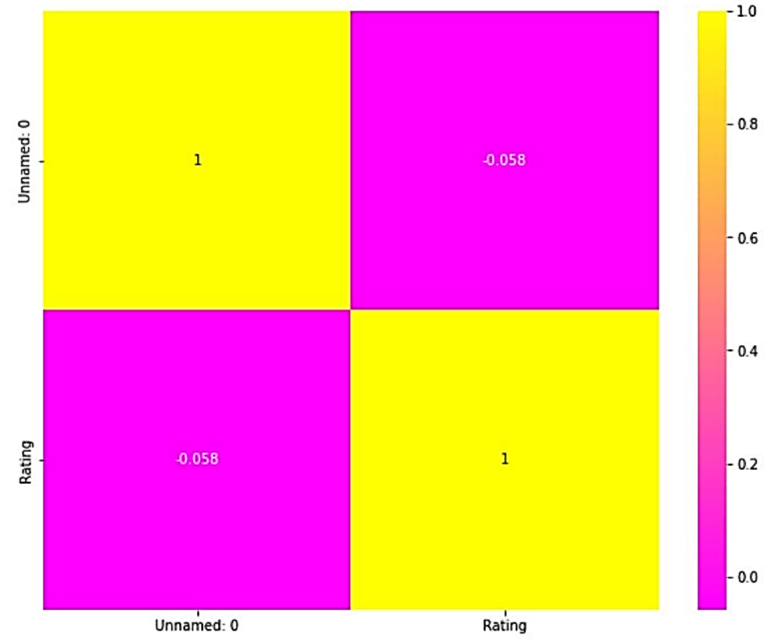
```
from sklearn.preprocessing import StandardScaler
# Data Scaling Formula Z=(x-mean)/std
scaler = StandardScaler()
X_scaled=scaler.fit_transform(X)
X_scaled
```

```
array([[-1.52399933,  1.39744572],
       [-1.17024567,  1.49836764],
       [-0.81649202, -0.35186749],
       [-0.46273837, -0.78919579],
       [-0.10898471,  1.70021147],
       [ 0.24476894,  0.28730465],
       [ 0.5985226 ,  0.92647678],
       [ 0.95227625, -1.22652409],
       [ 1.3060299 ,  1.33016444],
       [ 1.65978356, -0.04910174],
       [-1.52399933, -1.52928984],
       [-1.17024567,  1.73385211],
       [-0.81649202, -0.95739898],
       [-0.46273837, -1.59657112],
       [-0.10898471,  1.43108636],
       [ 0.24476894,  0.72463295],
       [ 0.5985226 , -1.69749303],
       [ 0.95227625, -0.99103962],
```

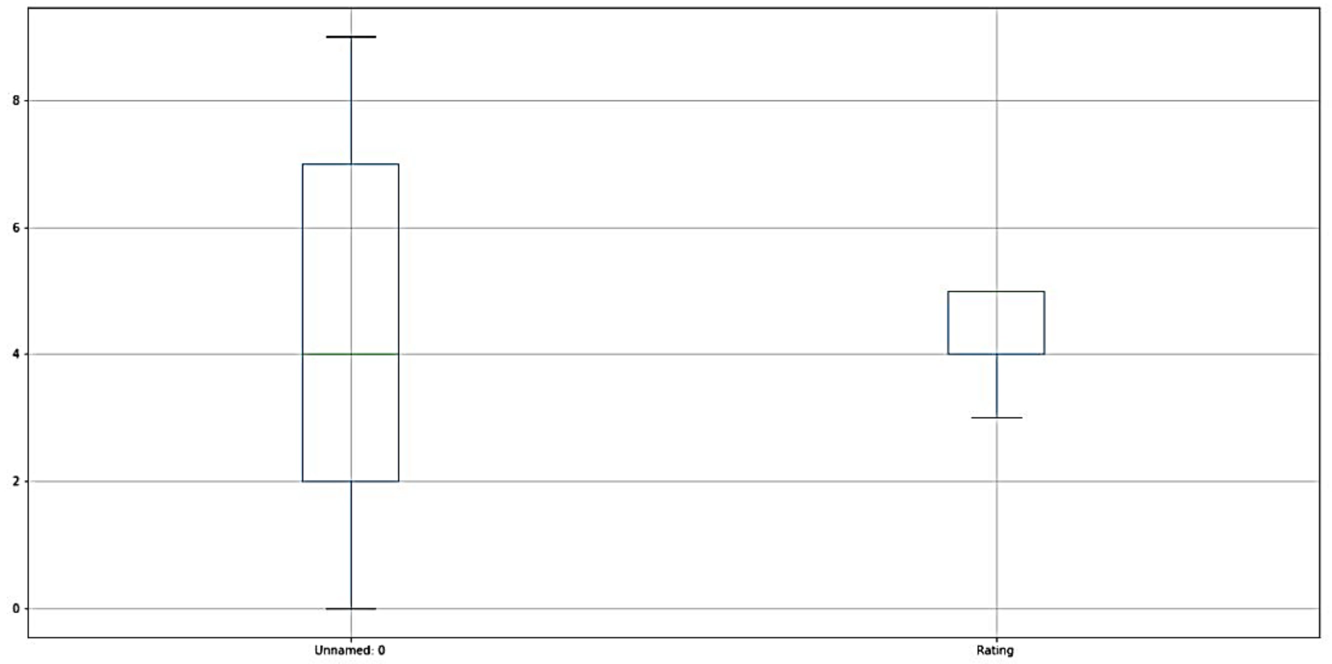

Rating
100
80
60
40
20
0
5
4
3

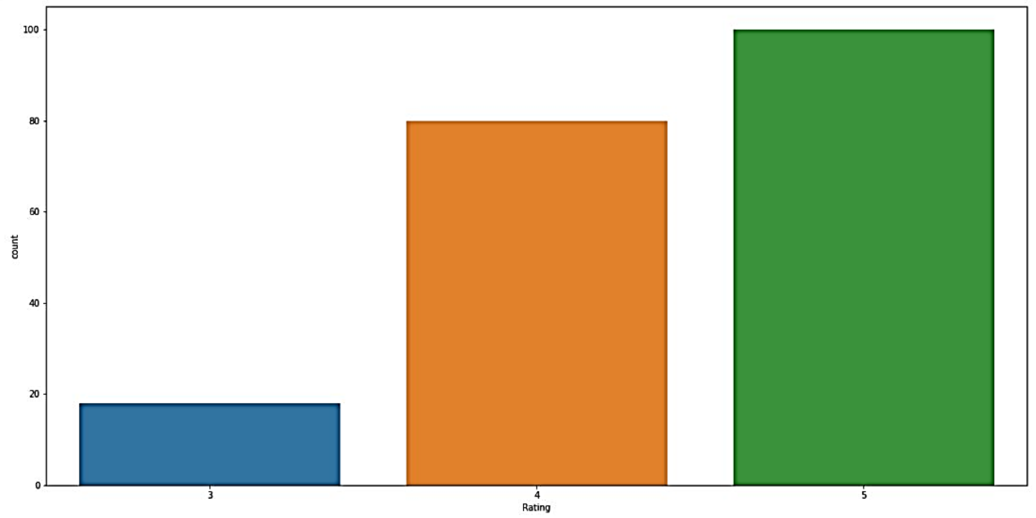

Reviews
0
2
4
6
8
Wonderful
Excellent
Just wow!
Nice product
Terrific
Brilliant
Simply awesome
Must buy!
Super!
Mind-blowing purchase
Good quality product
Terrific purchase
Highly recommended
Good choice
Awesome
Pretty good
Good
Value-for-money
Does the job
Perfect product!
Fair
Fabulous!
Classy product
Good product
Worth the money
Value for money
Best in the market!
Delightful
Great product
Really Nice
Totally Worth It!
Fully Satisfied!
This design puts your health on first priority
Durable with Premium Features
Good product but overprized
Small fans are exhaust
Read this for easy installation.
Absolutely Amazing
Great product for laptop
Very Good camera in its price segment
Best cam after trying all other brands
Great piece for both beginners and intermediates.
Too much Costly
noisy
The best type in the market
Disappointing product
Reasonably good produ
Hopeless adssive
Not a universal holder , specially not for iPhone 13
Rattles, otherwise ok
Adhesive is not Good. Product utility is great.
Music is amazing.
It is the best, loved it.
Bad quality
Good Build Quality
A waste of time, collection of general advice, no insights
goodread!!
Not recommended at all
Okay quality
Good products
It's worth the cost . The sound quality is good.
Good sound.
Overall a great product, ears hurt after using continuosly for 6-7 hrs
Value for Money
Awesome earbuds in this range ▯▯
Great Product!
Camera not working after few months use
#reviewjockey(RJ Arnab Dey)
Decent camera
Gives a good night vision
Seamless connectivity
Excellent Mesh Wifi
Amazing experience
Rocket ▯ performance vfm in just 20k
Wifi 6E signal?
Awesomely superb. Worked well and instant configuration.
Excellent Wifi lacks controls, The app is a privacy risk (asks for location).
seamlessly network between different decos
Best Budget Tri-Band Mesh Wifi
Its my first action camera and I found it amazing. Looks just like GOPRO
Value for money ▯ great product
Decent budget Camera!!!
Superb Quality Videos
Nice product with suitable budget
really reliable product at this price range
GoPro on a Budget
Great gadget at a decent price for Motovlogging
Working better than expected!
Good upgrade to TV speakers
Nice 5.1 Surround sound speaker with worst Bluetooth connectivity
Great one from Sony
Box was not sealed protective film was absent on camera screen.
Other than initial hiccups, some (Mi Home app) software bugs, it is good for monitoring!
Valie for money.. Recommended
MEMERY CARD NOT ▯ SPORT SAND DICK 32GB
Worth every penny
Just okay
Wonderful powerpacked value for money action camera
Quite Good
Great buy
Terrible product
Sound quality is So good
9500WS is worth every penny. Best for this price.
Got an used product, still working fine, except satellite speaker
Like the music and bass
Best in the price segment. Go for it
Best purchase for 12K
The gel is totally dry so not able to fix it on dashboard, waste

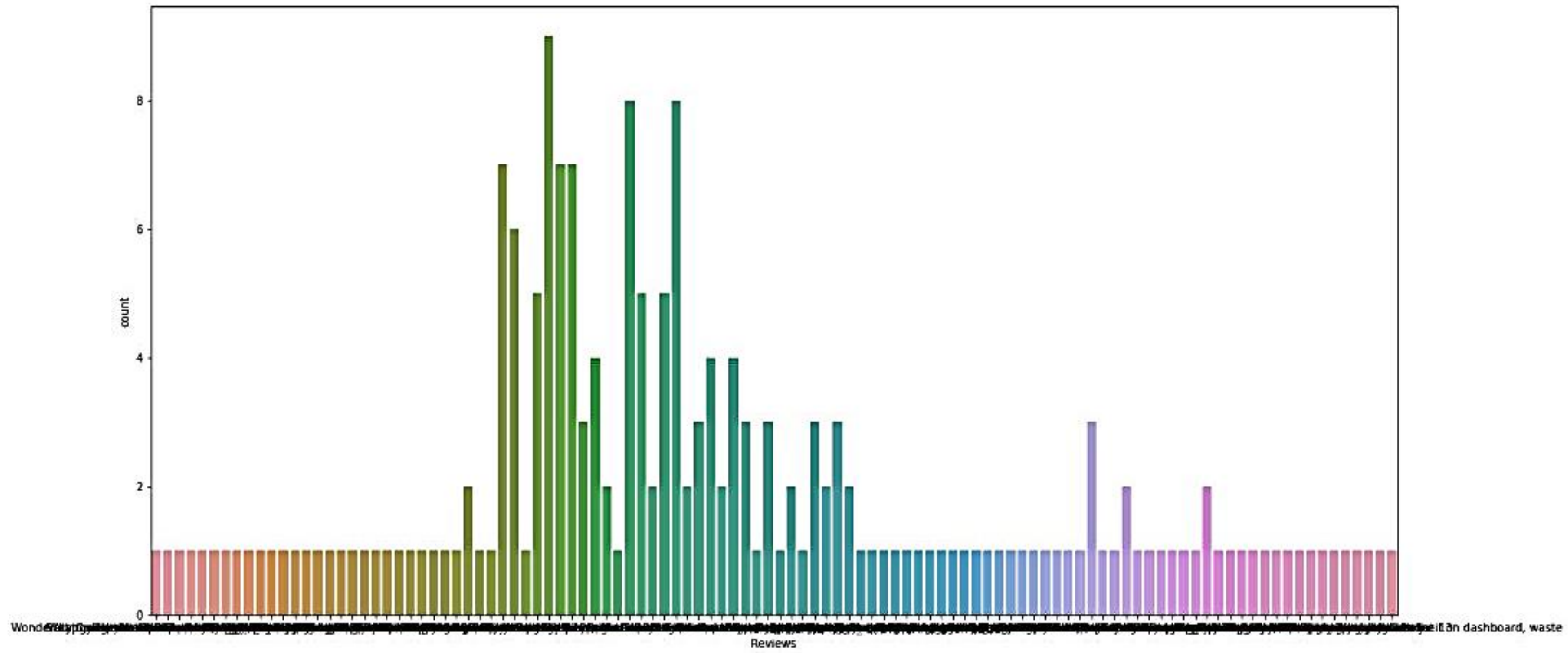
```
plt.figure(figsize=(10,8))
sns.heatmap(corr,annot=True,cmap='spring')
plt.show()
```

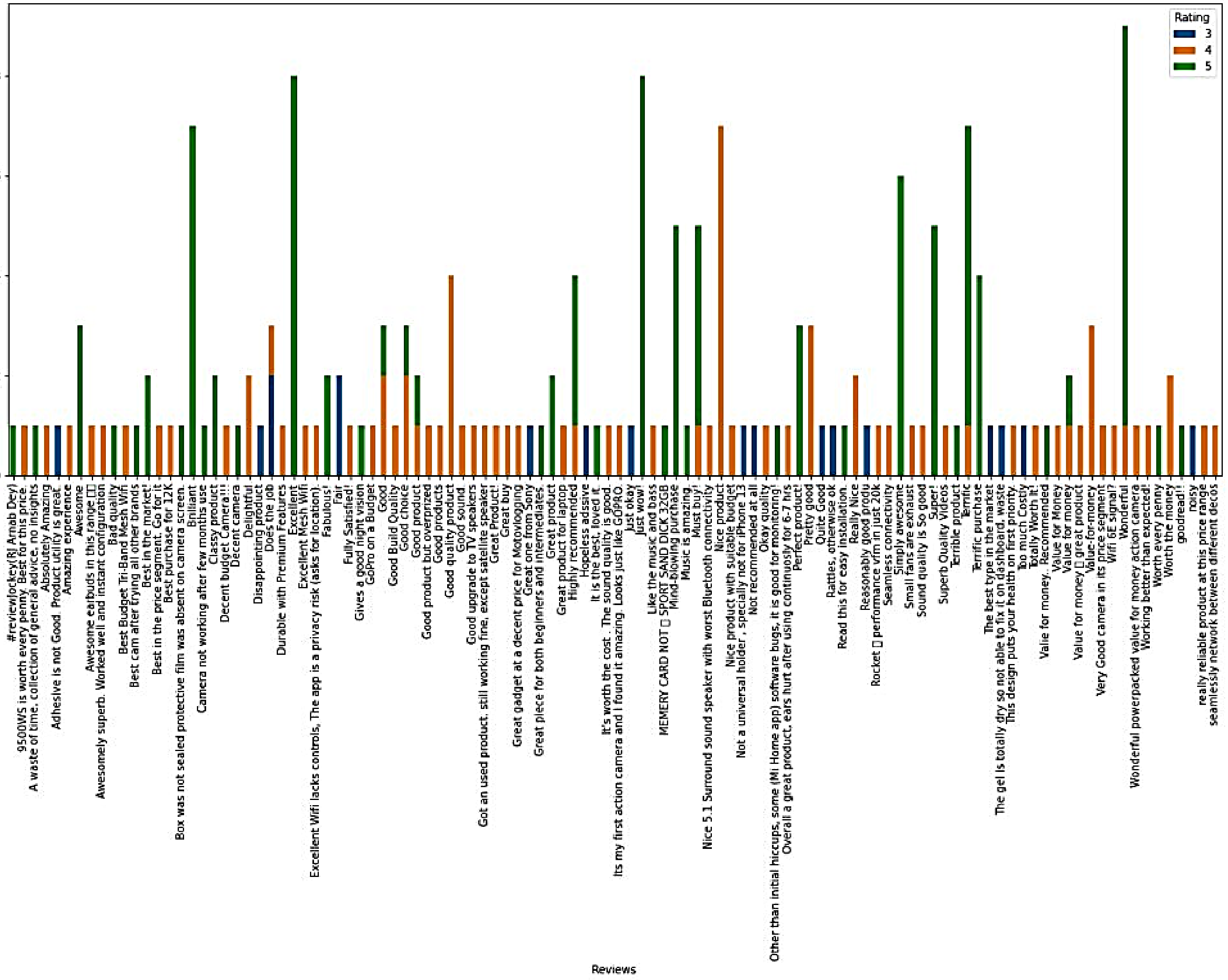


```
plt.figure(figsize=(20,10))
df.boxplot();
```

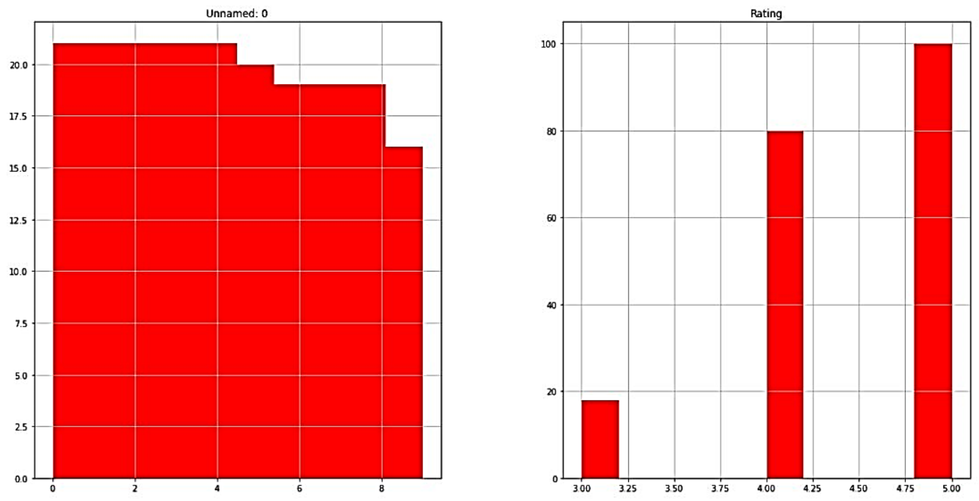



count
Rating
0
20
40
60
80
100
3
4
5

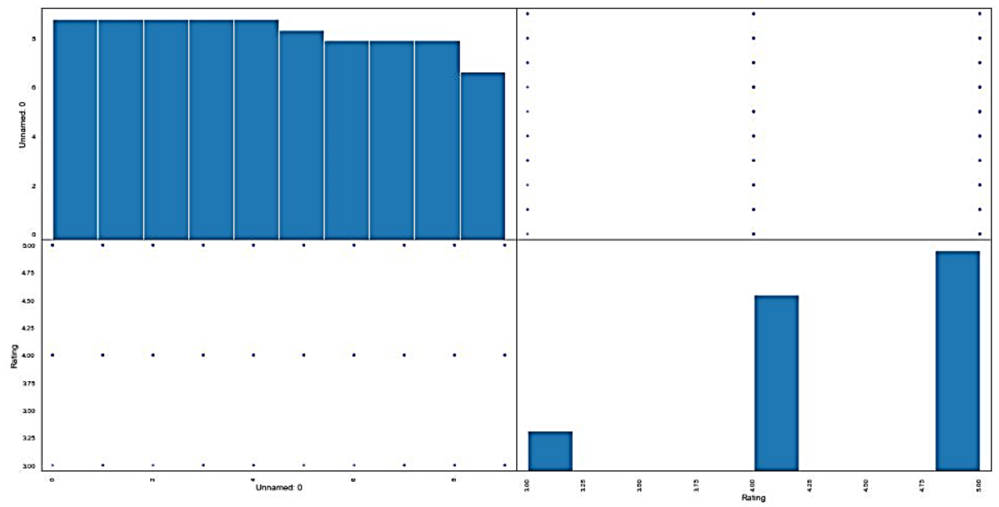

count
Reviews
0
2
4
6
8
Wonderful
an dashboard, waste


Rating
3
4
5
0
2
4
6
8
#reviewJockey[RJ Arnab Dey]
9500WS is worth every penny. Best for this price.
A waste of time, collection of general advice, no insights
Absolutely Amazing
Adhesive is not Good. Product utility is great.
Amazing experience
Awesome
Awesome earbuds in this range □□
Awesomely superb. Worked well and instant configuration
Bad quality
Best Budget Tri-Band Mesh Wifi
Best cam after trying all other brands
Best in the market!
Best in the price segment. Go for it
Best purchase for 12K
Box was not sealed protective film was absent on camera screen.
Brilliant
Camera not working after few months use
Classy product
Decent budget Camera!!!
Decent camera
Delightful
Disappointing product
Does the job
Durable with Premium Features
Excellent
Excellent Mesh Wifi
Excellent Wifi lacks controls, The app is a privacy risk (asks for location).
Fabulous!
Fair
Fully Satisfied!
Gives a good night vision
GoPro on a Budget
Good
Good Build Quality
Good choice
Good product
Good product but overprized
Good products
Good quality product
Good sound
Good upgrade to TV speakers
Got an used product. still working fine, except satellite speaker
Great Product!
Great buy
Great gadget at a decent price for Motovlogging
Great one from Sony
Great piece for both beginners and intermediates
Great product
Great product for laptop
Highly recommended
Hopeless adssive
It is the best, loved it
It's worth the cost . The sound quality is good.
Its my first action camera and I found it amazing. Looks just like GOPRO
Just okay
Just wow!
Like the music and bass
MEMERY CARD NOT □ SPORT SAND DICK 32GB
Mind-blowing purchase
Music is amazing
Must buy!
Nice 5.1 Surround sound speaker with worst Bluetooth connectivity
Nice product
Nice product with suitable budget
Not a universal holder , specially not for iPhone 13
Not recommended at all
Okay quality
Other than initial hiccups, some (Mi Home app) software bugs, it is good for monitoring
Overall a great product, ears hurt after using continuosly for 6-7 hrs
Perfect product!
Pretty good
Quite Good
Rattles, otherwise ok
Read this for easy installation.
Really Nice
Reasonably good produ
Rocket □ performance vfm in just 20k
Seamless connectivity
Simply awesome
Small fans are exhaust
Sound quality is So good
Super!
Superb Quality Videos
Terrible product
Terrific
Terrific purchase
The best type in the market
The gel is totally dry so not able to fix it on dashboard. waste
This design puts your health on first priority.
Too much Costly
Totally Worth it!
Value for money.. Recommended
Value for Money
Value for money
Value for money □ great product
Value-for-money
Very Good camera in its price segment
Wifi 6E signal?
Wonderful
Wonderful powerpacked value for money action camera
Working better than expected!
Worth every penny
Worth the money
goodread!!
noisy
really reliable product at this price range
seamlessly network between different decos
Reviews

```python
df.hist(color='r',figsize=(20,10))
plt.show()
```

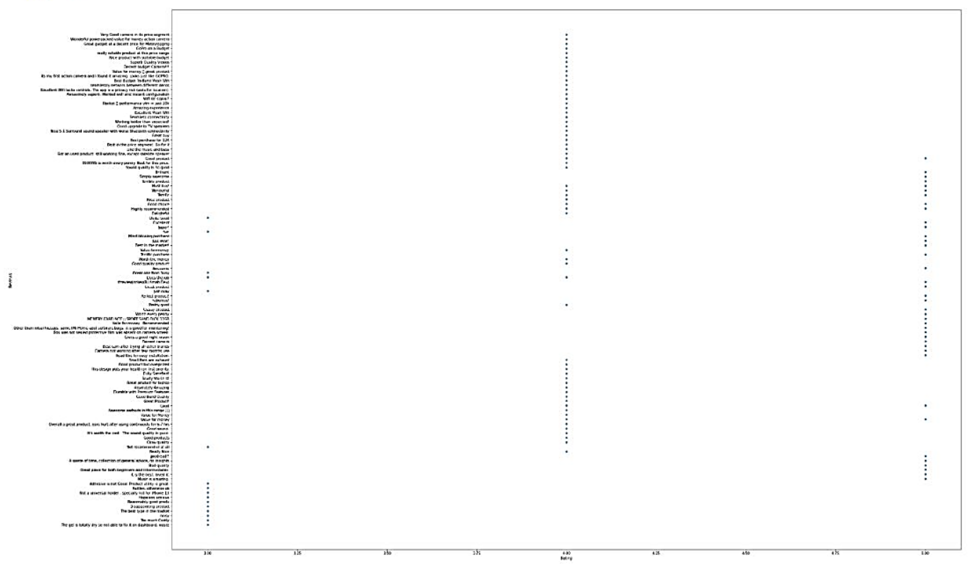


```python
from pandas.plotting import scatter_matrix
scatter_matrix(df,figsize=(20,10),color='b')
plt.show()
```

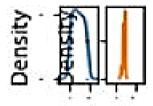


```python
plt.figure(figsize=(40,28))
sns.scatterplot(df['Rating'],df['Reviews'])
print(df[['Rating','Reviews']].corr())
plt.show()
```

```
        Rating
Rating     1.0
```

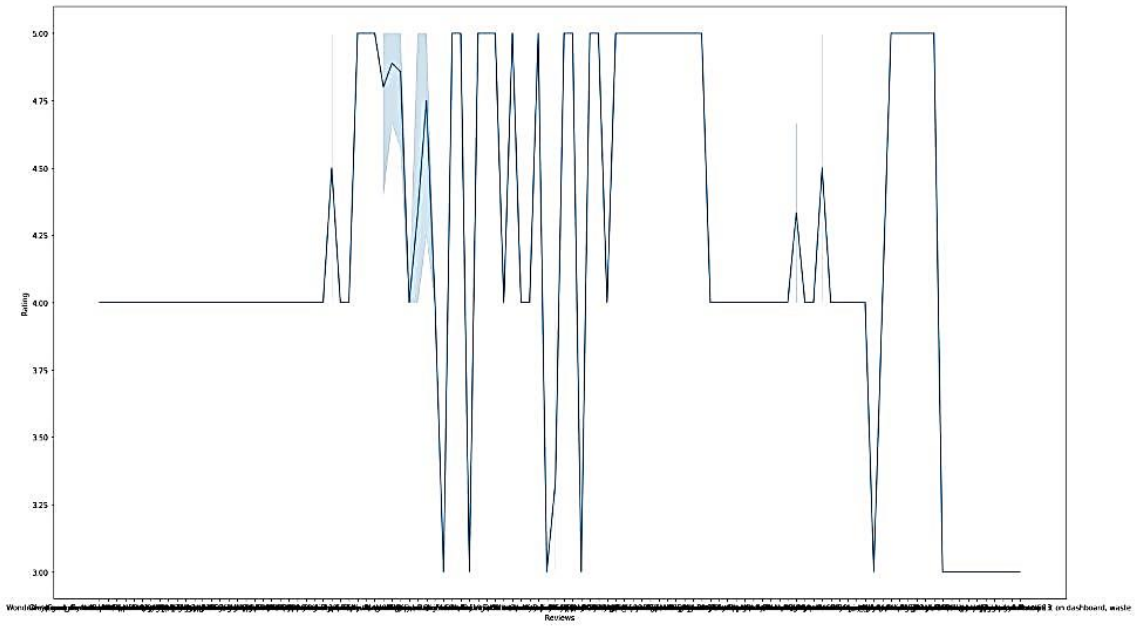


```python
df.plot(kind='density',subplots=True,layout=(5,15),legend=False,fontsize=2);
```



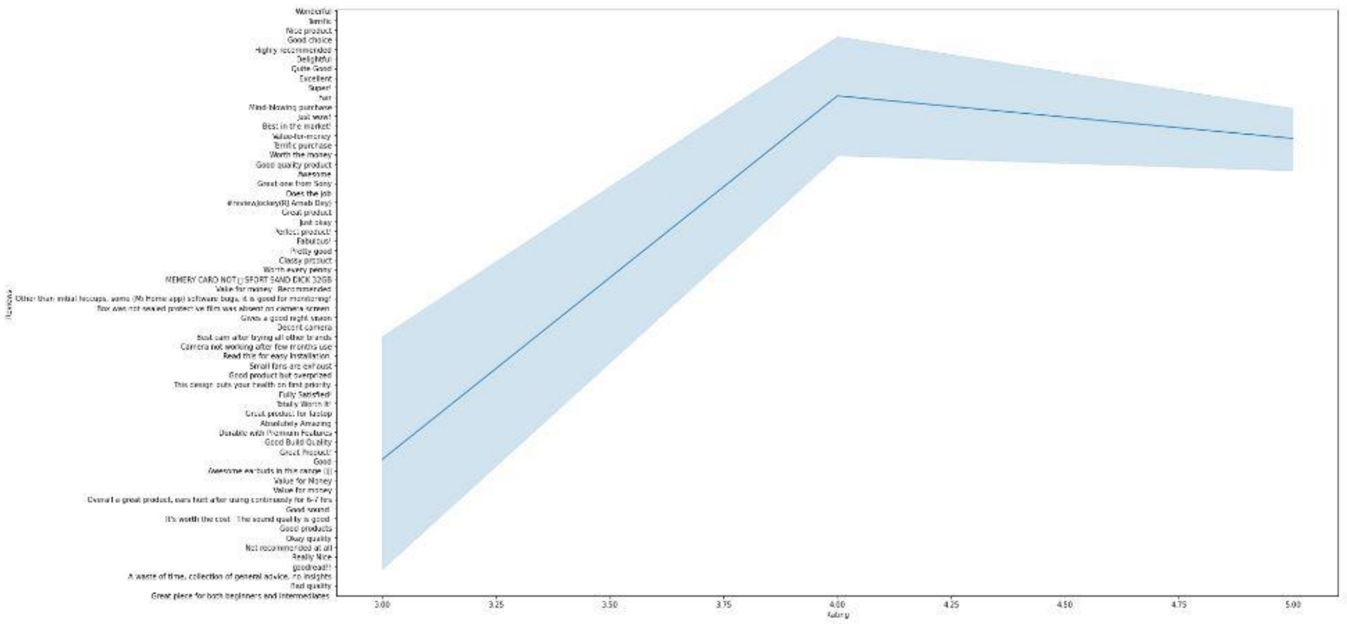
```
plt.figure(figsize=(25,15))
sns.lineplot(x='Reviews',y='Rating',data=df)
```

```
<AxesSubplot:xlabel='Reviews', ylabel='Rating'>
```



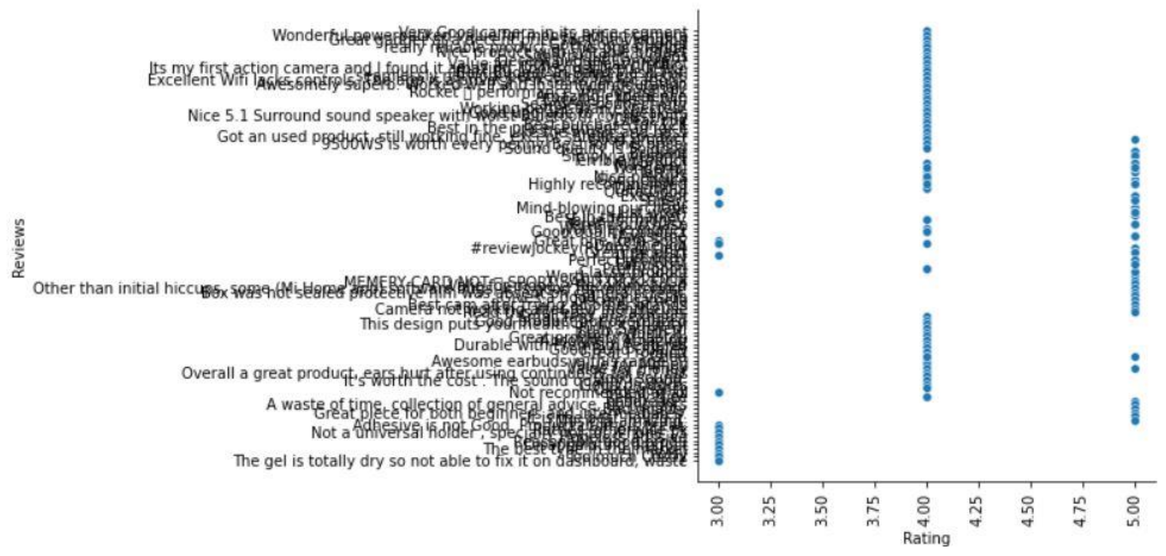
```
plt.figure(figsize=(25,15))
sns.lineplot(x='Rating',y='Reviews',data=df)
```

```
<AxesSubplot:xlabel='Rating', ylabel='Reviews'>
```



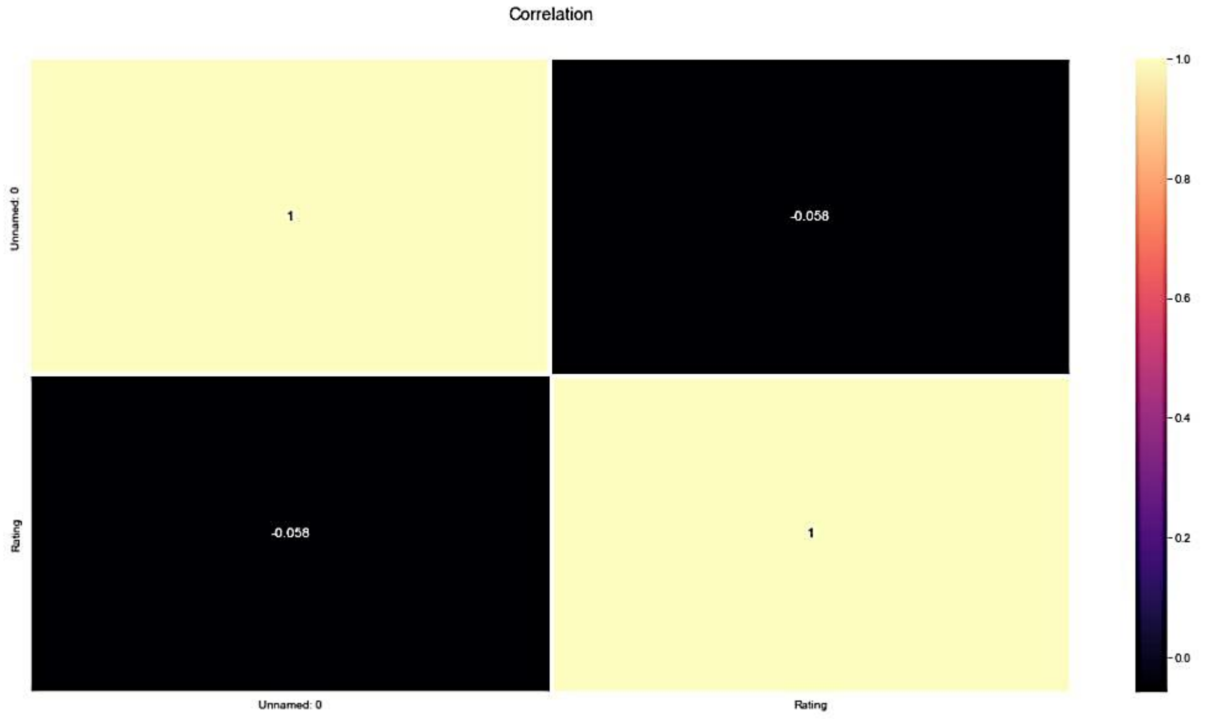
```python
plt.figure(figsize=(40,28))
sns.relplot(x='Rating', y='Reviews',data=df)
plt.xticks(rotation=90);
```

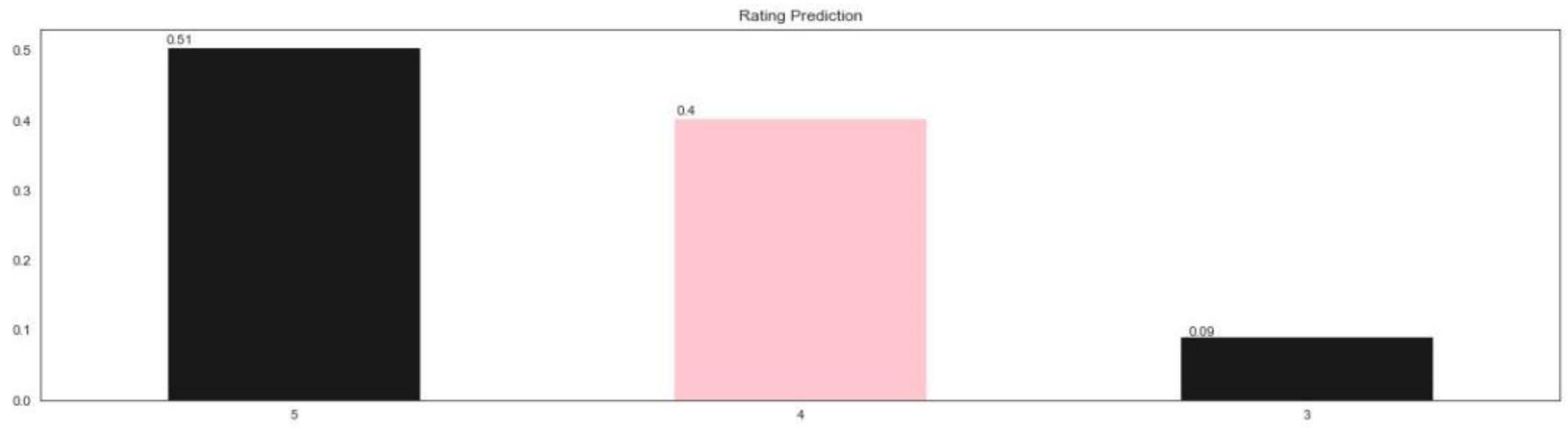
```
<Figure size 2880x2016 with 0 Axes>
```

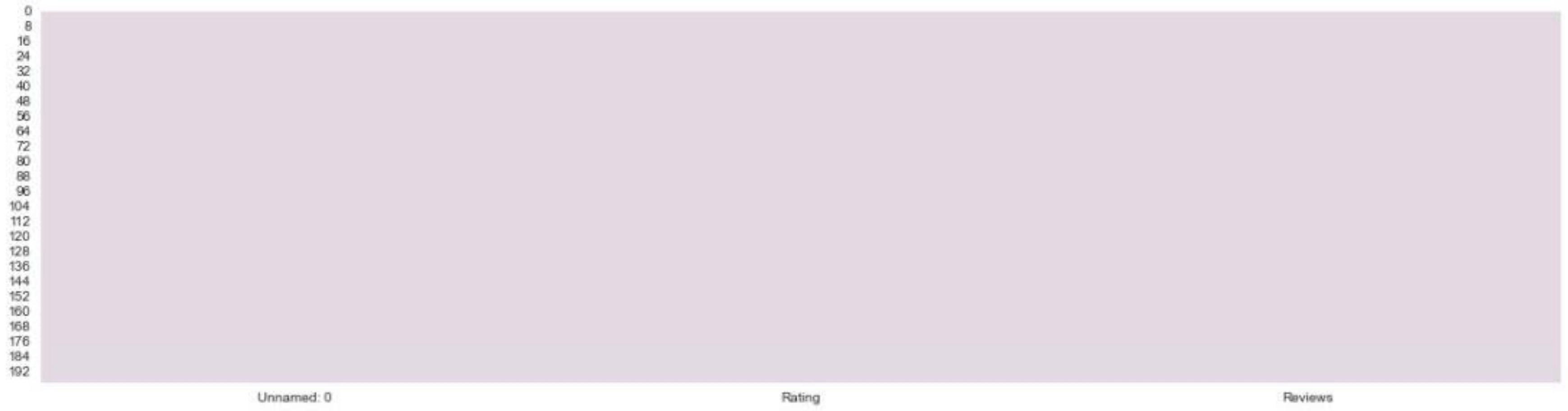


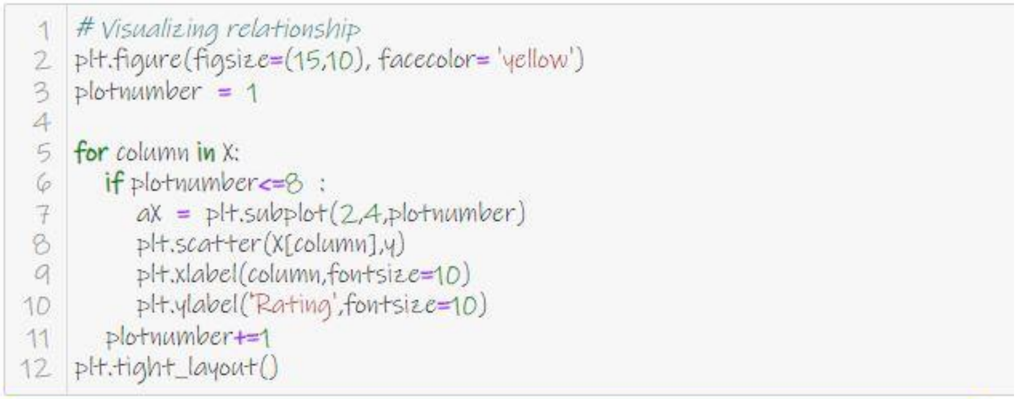
```python
def correlation_heatmap(df):
    ax=plt.subplots(figsize=(20,10))
    colormap=sns.diverging_palette(220, 10, as_cmap = True)
    ax=sns.heatmap(df.corr(),cmap="magma",annot=True,linewidths=0.1,vmax=1.0,linecolor='white',annot_kws={'fontsize':12})
    plt.title('Correlation',y=1.05,size=15)

correlation_heatmap(df)
```



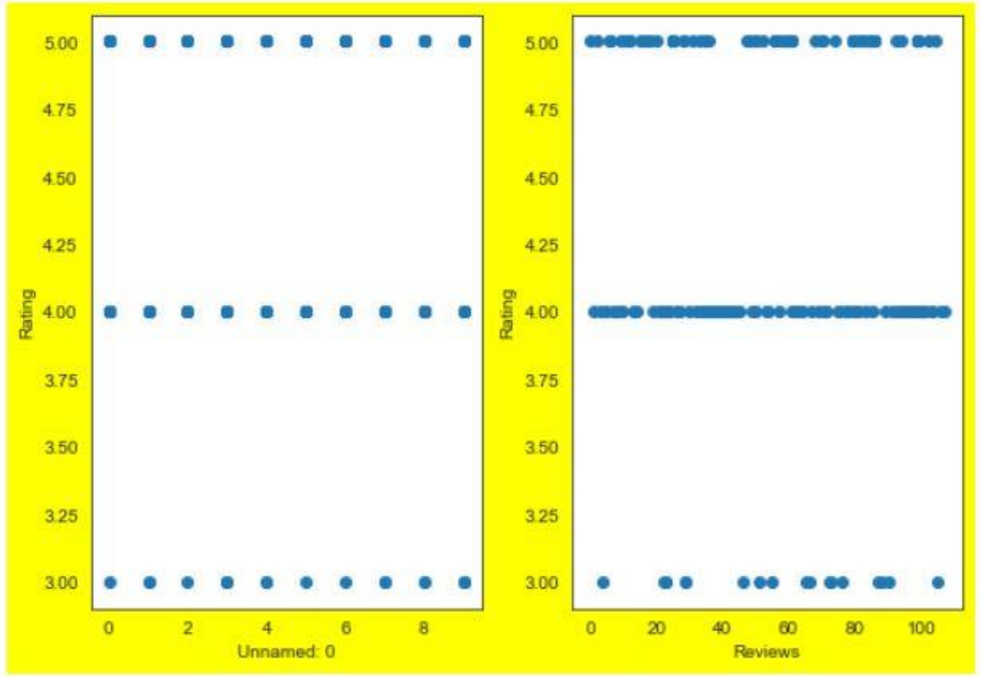

Rating Prediction
0.51
0.4
0.09
0.5
0.4
0.3
0.2
0.1
0.0
5
4
3


0
8
16
24
32
40
48
56
64
72
80
88
96
104
112
120
128
136
144
152
160
168
176
184
192
Unnamed: 0
Rating
Reviews

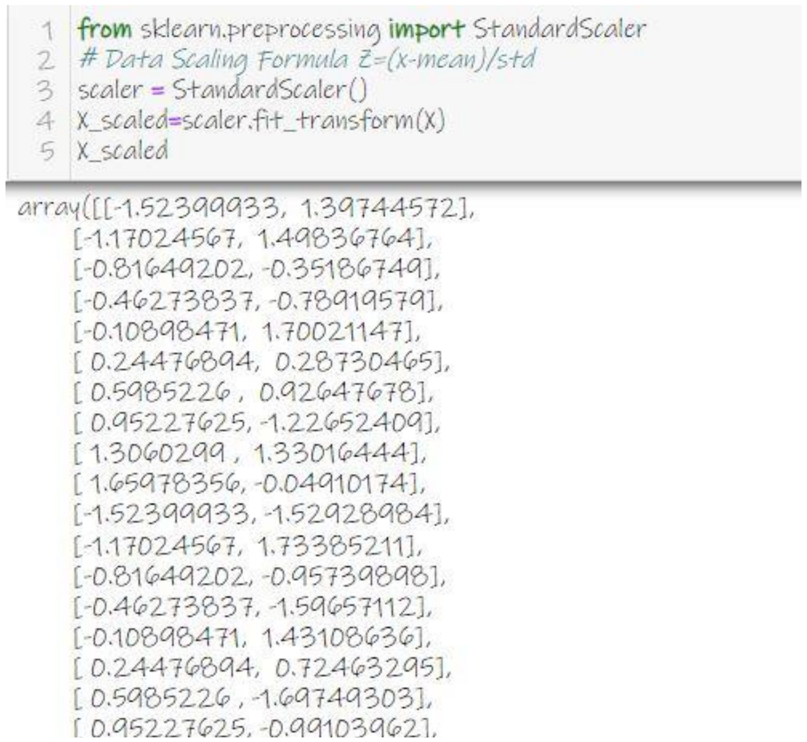


```python
# Visualizing relationship
plt.figure(figsize=(15,10), facecolor= 'yellow')
plotnumber  = 1

for column in X:
    if plotnumber<=8 :
        aX = plt.subplot(2,4,plotnumber)
        plt.scatter(X[column],y)
        plt.xlabel(column,fontsize=10)
        plt.ylabel('Rating',fontsize=10)
    plotnumber+=1
plt.tight_layout()
```

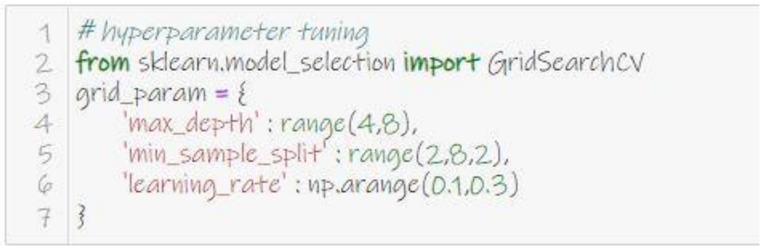


VISUALIZING THE RELATIONSHIP WITH RATING



```python
from sklearn.preprocessing import StandardScaler
# Data Scaling Formula Z=(x-mean)/std
scaler = StandardScaler()
X_scaled=scaler.fit_transform(X)
X_scaled
```

```
array([[-1.52399933,  1.39744572],
       [-1.17024567,  1.49836764],
       [-0.81649202, -0.35186749],
       [-0.46273837, -0.78919579],
       [-0.10898471,  1.70021147],
       [ 0.24476894,  0.28730465],
       [ 0.5985226 ,  0.92647678],
       [ 0.95227625, -1.22652409],
       [ 1.3060299 ,  1.33016444],
       [ 1.65978356, -0.04910174],
       [-1.52399933, -1.52928984],
       [-1.17024567,  1.73385211],
       [-0.81649202, -0.95739898],
       [-0.46273837, -1.59657112],
       [-0.10898471,  1.43108636],
       [ 0.24476894,  0.72463295],
       [ 0.5985226 , -1.69749303],
       [ 0.95227625, -0.99103962],
```



```python
# hyperparameter tuning
from sklearn.model_selection import GridSearchCV
grid_param = {
    'max_depth' : range(4,8),
    'min_sample_split' : range(2,8,2),
    'learning_rate' : np.arange(0.1,0.3)
}
```

STEPS AND ASSUMPTIONS TO COMPLETE THE PROJECT:

1. The purpose of this case is to understand the rise of e-commerce which brought a significant rise in the importance of customer reviews and evaluate used to predict rating and to develop a strategy that utilizes data mining techniques towards ratings prediction.

2. There are two main methods to approach this problem. The first one is based on review text content analysis and uses the principles of natural language process (the NLP method). This method lacks the insights that can be drawn from the relationship between costumers and items.

3. The model of the independent variables and dependent variables are exactly vary with the variables.

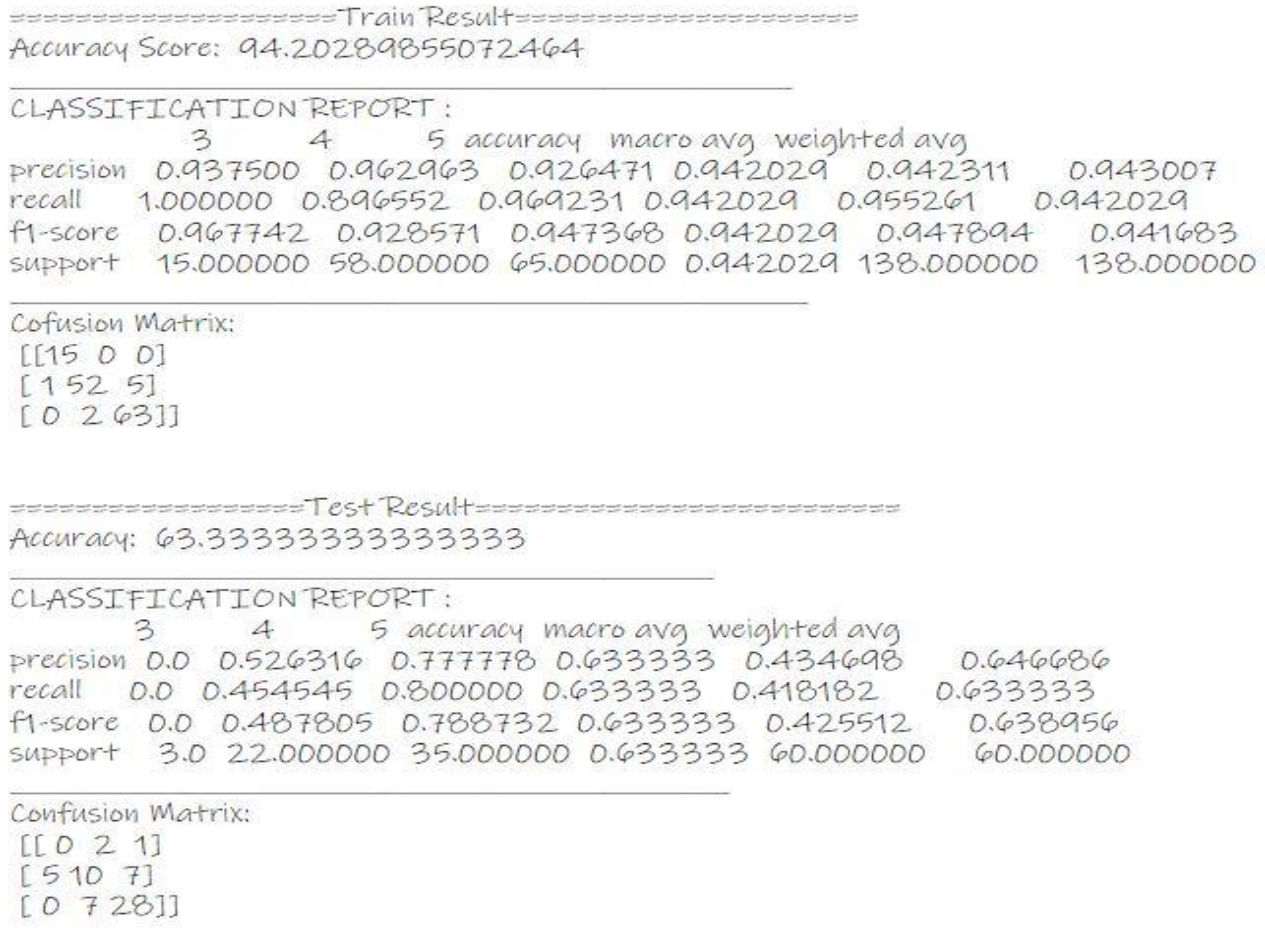
4. It can accordingly manipulating the strategy of the areas that will yield high returns as it make easier for the clients.

5. By visualizations there are many things to be noted when it will according to work each other

6. By preprocessing the data it means that from the help of label encoder helps the dataset column to transform to fit another column in to it.

FINALIZED MODEL:

Accuracy score:
Train Result : 94.2%
Test Result : 63.3%



```
====================Train Result====================
Accuracy Score: 94.20289855072464
_______________________________________________
CLASSIFICATION REPORT :
                  3          4          5  accuracy  macro avg  weighted avg
precision  0.937500   0.962963   0.926471  0.942029   0.942311      0.943007
recall     1.000000   0.896552   0.969231  0.942029   0.955261      0.942029
f1-score   0.967742   0.928571   0.947368  0.942029   0.947894      0.941683
support   15.000000  58.000000  65.000000  0.942029 138.000000    138.000000
_______________________________________________
Cofusion Matrix:
 [[15  0  0]
 [ 1 52  5]
 [ 0  2 63]]
```

```
=================Test Result========================
Accuracy: 63.33333333333333
_______________________________________________
CLASSIFICATION REPORT :
             3          4          5  accuracy  macro avg  weighted avg
precision  0.0   0.526316   0.777778  0.633333   0.434698      0.646686
recall     0.0   0.454545   0.800000  0.633333   0.418182      0.633333
f1-score   0.0   0.487805   0.788732  0.633333   0.425512      0.638956
support    3.0  22.000000  35.000000  0.633333  60.000000     60.000000
_______________________________________________
Confusion Matrix:
 [[ 0  2  1]
 [ 5 10  7]
 [ 0  7 28]]
```

CONCLUSION

In conclusion, combining the formerly know data about each user' similarity to other users with the sentiment analysis of the rating and reviews itself, does help to improve the model prediction of rate the user's review, will get the purpose.

THANK YOU