

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels as sm
import seaborn as sns
import plotly.express as px
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')

nike= pd.read_csv(r'/content/drive/MyDrive/collab/nike_data_2022_09.csv')
nike
```

	url	name	sub_title	brand	model	
0	https://www.nike.com/t/dri-fit-team-minnesota-...	Nike Dri-FIT Team (MLB Minnesota Twins)	Men's Long-Sleeve T-Shirt	Nike	14226571	
1	https://www.nike.com/t/club-américa-womens-dri...	Club América	Women's Nike Dri-FIT Soccer Jersey Dress	Nike	13814665	Black
2	https://www.nike.com/t/sportswear-swoosh-mens-...	Nike Sportswear Swoosh	Men's Overalls	Nike	13015648	Black

```
nike.nunique()

url          112
name          96
sub_title     80
brand         1
model        112
color         76
price         63
currency       1
availability   2
description   112
raw_description 112
avg_rating     7
review_count  11
images       108
available_sizes 35
uniq_id       112
scraped_at     21
dtype: int64

mens-polo-... College B... mens polo... NIKE 14040700 Sign
nike.drop(['url', 'brand', 'model', 'currency', 'description', 'raw_description', 'images', 'unic
Big Kids'

nike.head()
```

	name	sub_title	color	price	availability	avg_rating	review_count
--	------	-----------	-------	-------	--------------	------------	--------------

nike.tail()

	name	sub_title	color	price	availability	avg_rating	review_count
107	Air Jordan XXXV Low DS PF	Basketball Shoes	Hyper Jade/Bright Crimson/Blue Orbit/Metallic ...	165.0	InStock	NaN	N
108	Nike College Dri-FIT (UCLA)	Men's Polo	Signal Blue	50.0	InStock	NaN	N
109	Jordan	Big Kids' (Boys') Full-Zip Jacket	Black	75.0	OutOfStock	NaN	N

nike.isnull().sum()

name	0
sub_title	0
color	2
price	0
availability	4
avg_rating	89
review_count	89
available_sizes	56
scraped_at	0
dtype: int64	

nike.dtypes

name	object
sub_title	object
color	object
price	float64
availability	object
avg_rating	float64
review_count	float64
available_sizes	object
scraped_at	object
dtype: object	

nike.describe()

	price	avg_rating	review_count
count	112.000000	23.000000	23.000000
mean	61.506250	4.773913	31.391304
std	39.887738	0.336043	72.942530
min	9.970000	3.800000	1.000000
25%	31.720000	4.750000	1.000000

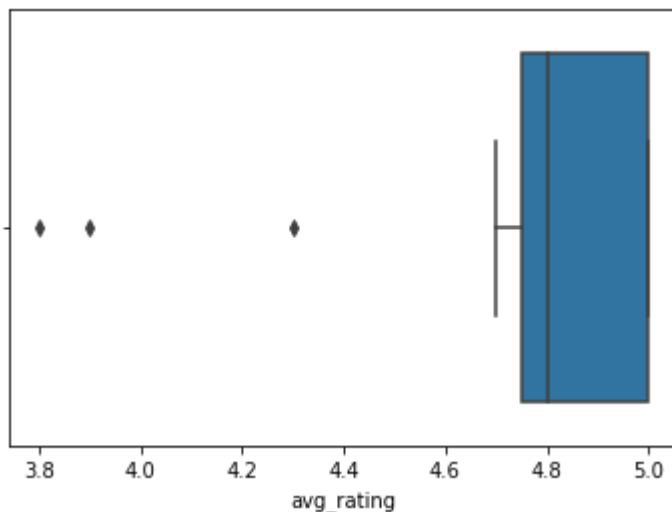
```
nike.describe(include='object')
```

	name	sub_title	color	availability	available_sizes	scrape
count	112	112	110	108	56	
unique	96	80	76	2	35	
top	Nike Sportswear	Men's Shoes	Black	InStock	S M L XL 2XL	20-09-2022
freq	6	5	15	67	12	

Outliers Detection

```
sns.boxplot(nike['avg_rating'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f2a30881850>
```



```
Q1= nike.avg_rating.quantile(0.25)
```

```
Q3= nike.avg_rating.quantile(0.75)
```

```
Q1,Q3
```

```
(4.75, 5.0)
```

```
IQR=Q3-Q1
```

```
IQR
```

0.25

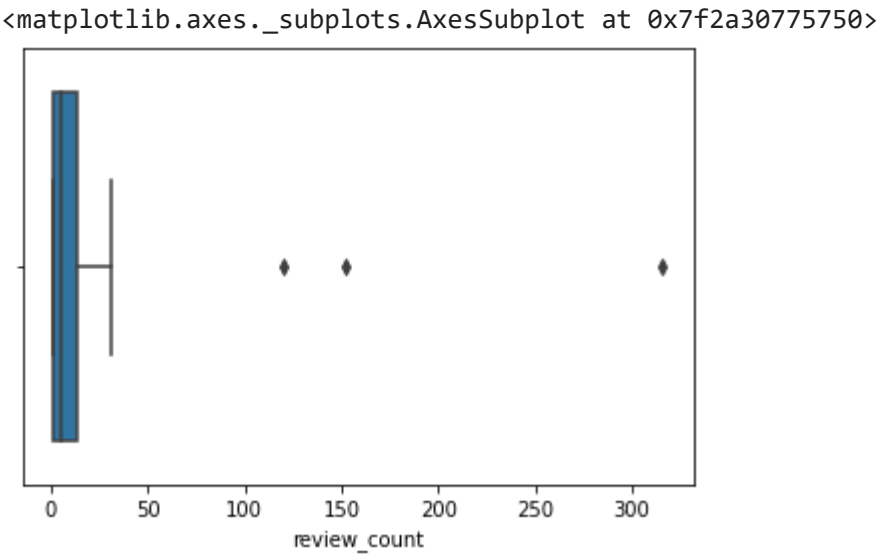
```
l1= Q1-1.5*IQR
u1= Q3+1.5*IQR
l1,u1
```

(4.375, 5.375)

```
nike[(nike.avg_rating<l1)|(nike.avg_rating>u1)]
```

	name	sub_title	color	price	availability	avg_rating	review_count
48	Nike Pro	Men's Tights	White/Black	24.97	OutOfStock	4.3	152.0
49	Nike SFB Gen	Tactical Boot	Black/Black/Black	140.00	InStock	3.8	28.0

```
sns.boxplot(nike['review_count'])
```



```
Q1= nike.review_count.quantile(0.25)
Q3=nike.review_count.quantile(0.75)
Q1,Q3
```

(1.0, 13.0)

```
IQR=Q3-Q1
IQR
```

12.0

```
l1= Q1-1.5*IQR
```

```
u1= Q3+1.5*IQR
l1,u1
```

```
(-17.0, 31.0)
```

```
nike[(nike.review_count<l1)|(nike.review_count>u1)]
```

	name	sub_title	color	price	availability	avg_rating	review_count
23	Air Jordan 1 Zoom CMFT	Shoes	Fossil Stone/Fossil Rose/Psychic Purple/White	150.00	InStock	4.8	12
48	Nike Pro	Men's Tights	White/Black	24.97	OutOfStock	4.3	15
52	Nike Bell Bump		Black/Black/White	12.00	InStock	4.7	21

Handling Null Values

```
nike['avg_rating'] = nike['avg_rating'].fillna(nike['avg_rating'].median(),axis=0)
nike['review_count'] = nike['review_count'].fillna(nike['review_count'].median(),axis=0)
```

```
# Since both avg rating and review count had outliers we have used median to handle null v
```

```
nike.describe()
```

	price	avg_rating	review_count
count	112.000000	112.000000	112.000000
mean	61.506250	4.794643	10.419643
std	39.887738	0.149979	34.193840
min	9.970000	3.800000	1.000000
25%	31.720000	4.800000	5.000000
50%	45.000000	4.800000	5.000000
75%	85.242500	4.800000	5.000000
max	175.000000	5.000000	316.000000

```
nike['color'] = nike['color'].fillna(nike['color'].mode()[0])
nike['availability'] = nike['availability'].fillna(nike['availability'].mode()[0])
nike['available_sizes'] = nike['available_sizes'].fillna(nike['available_sizes'].mode()[0])
```

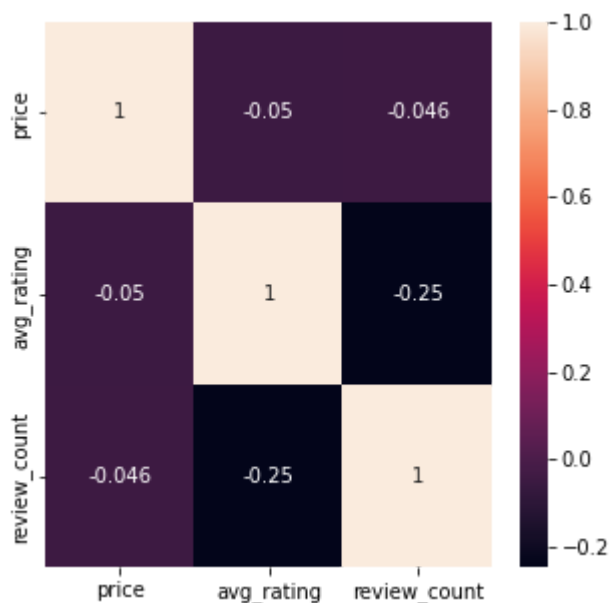
```
# Since all the three values were categorical which is obj type we have used mode to handl
```

```
nike.describe(include='object')
```

	name	sub_title	color	availability	available_sizes	scrape
count	112	112	112	112	112	
unique	96	80	76	2	35	
top	Nike Sportswear	Men's Shoes	Black	InStock	S M L XL 2XL	20-09-2022
freq	6	5	17	71	68	

Correlation

```
plt.figure(figsize=(5,5))
sns.heatmap(nike.corr(),annot=True)
plt.show()
nike.corr()
```



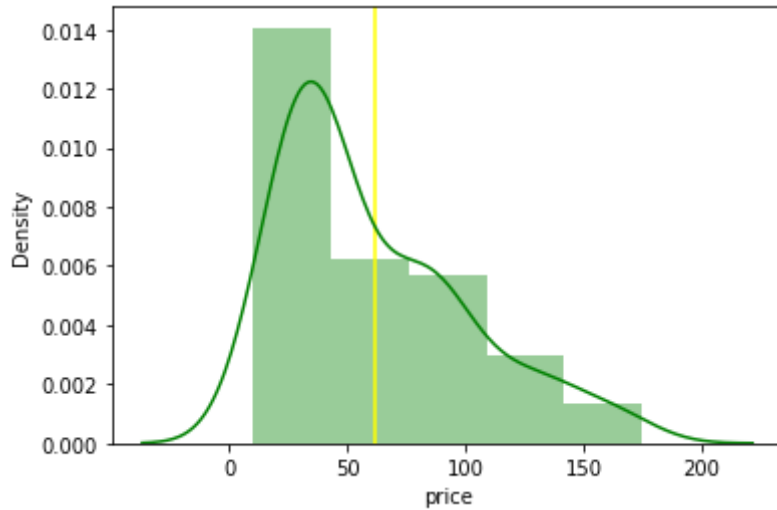
	price	avg_rating	review_count
price	1.000000	-0.049675	-0.045953
avg_rating	-0.049675	1.000000	-0.245673
review_count	-0.045953	-0.245673	1.000000

#Through the correlation chart we could easily identify that all the values are negatively

Visualization

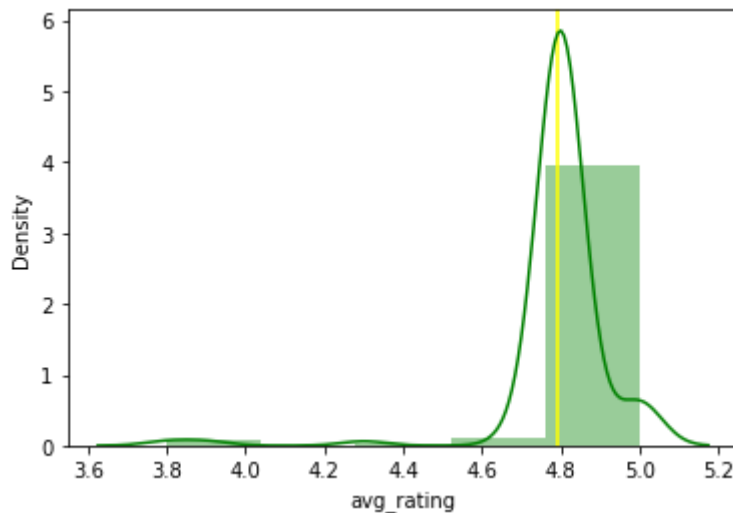
```
plt.axvline(nike['price'].mean(), 0,1, color = 'yellow')
sns.distplot(nike['price'], color='green',bins = 5, kde = True)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f2a302a15d0>

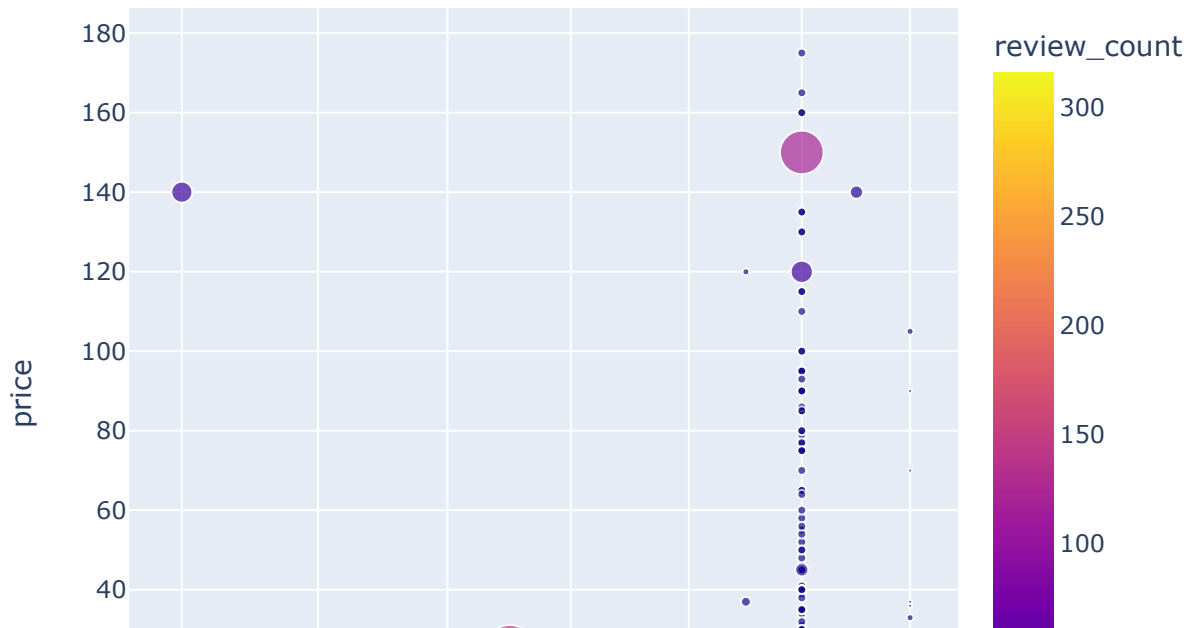


```
plt.axvline(nike['avg_rating'].mean(), 0,1, color = 'yellow')
sns.distplot(nike['avg_rating'], color='green',bins = 5, kde = True)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f2a30775f50>



```
px.scatter(nike,x="avg_rating", y="price", size="review_count", color="review_count",hover
size_max=25)
```

```
sns.countplot(nike['availability'],palette='husl')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f2a2d8ac610>



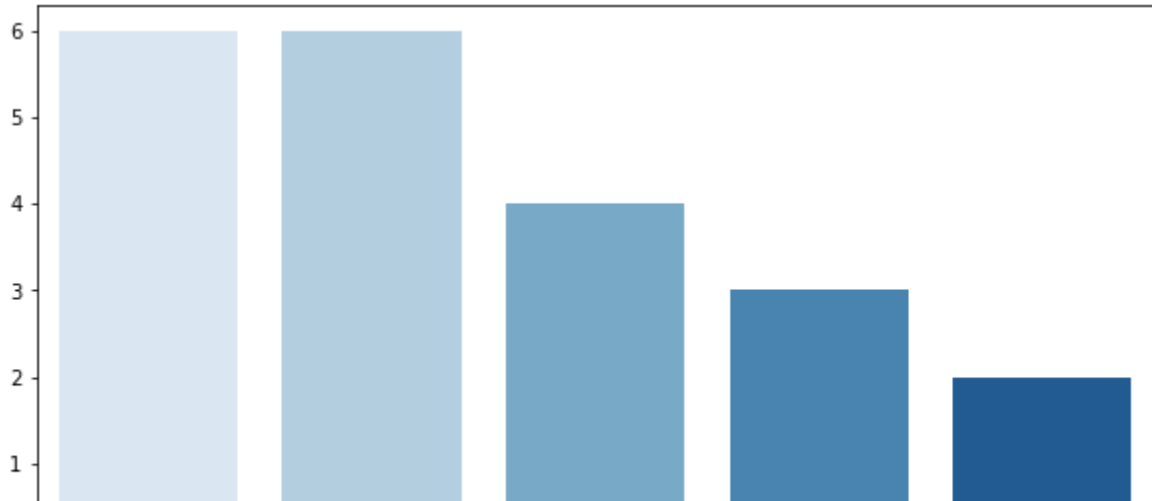
```
try:
```

```
    name= nike['name'].value_counts()[:5]
    plt.figure(figsize=(10,5))
    ax = sns.barplot(x=list(name.keys()),y=list(name.values),palette='Blues')
    for container in ax.containers:
        ax.bar_label(container)
```

```
except AttributeError:
    print('AttributeError Spotted')
```

```
finally:
    print('Top 5 Nike Names')
```

AttributeError Spotted
Top 5 Nike Names

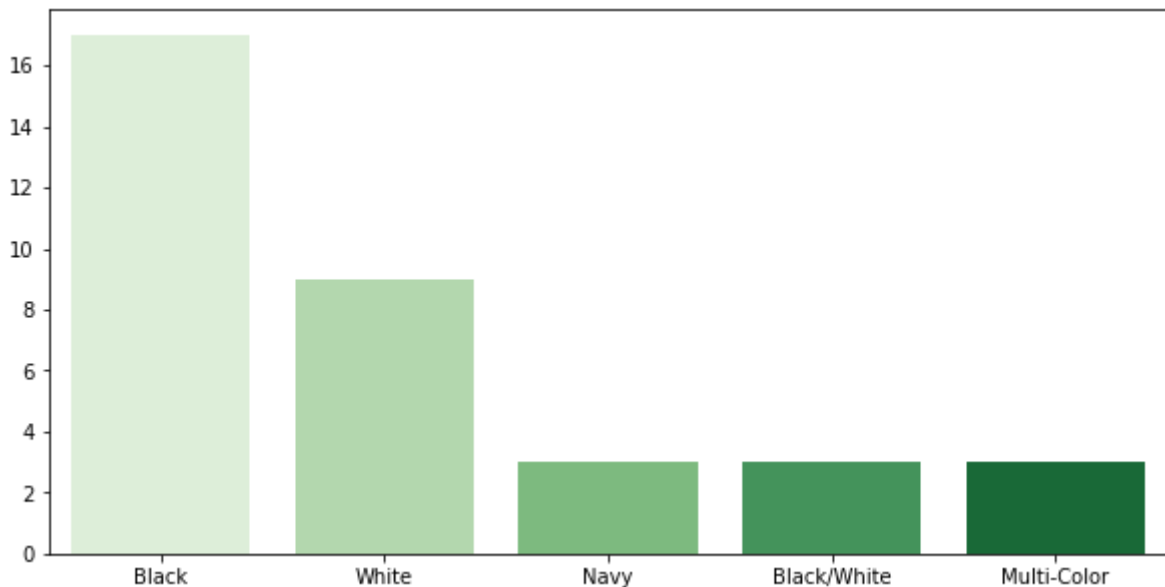


```
try:
    color = nike['color'].value_counts()[:5]
    plt.figure(figsize=(10,5))
    ax = sns.barplot(x=list(color.keys()),y=list(color.values),palette='Greens')
    for container in ax.containers:
        ax.bar_label(container)
```

```
except AttributeError:
    print('AttributeError Spotted')
```

```
finally:
    print('Top 5 Nike Colors')
```

AttributeError Spotted
Top 5 Nike Colors

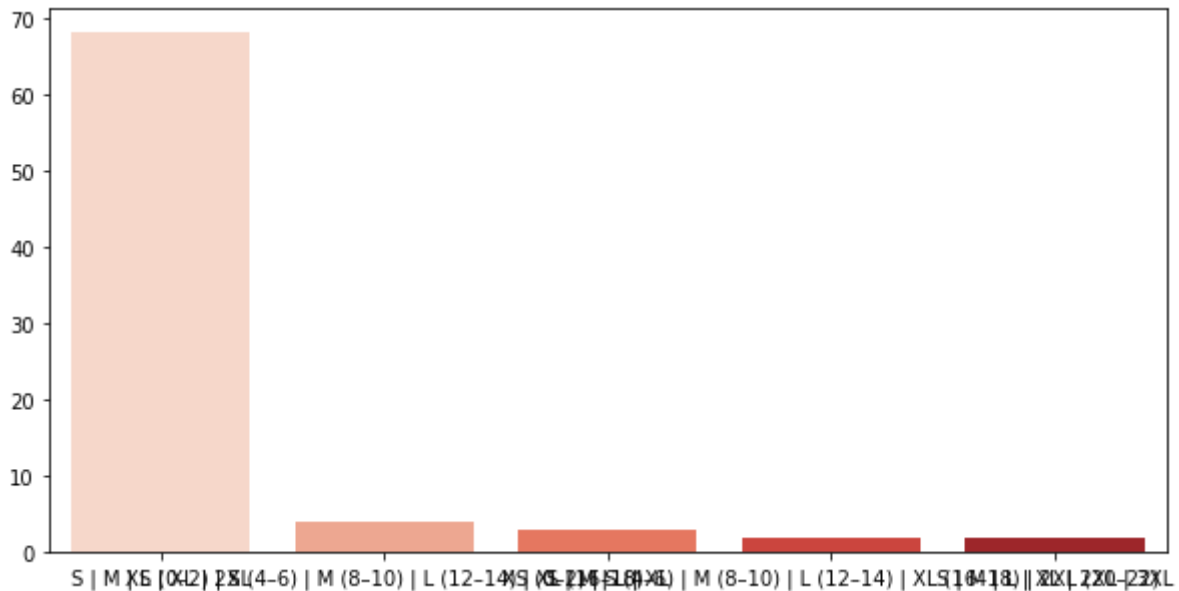


```
try:
    available = nike['available_sizes'].value_counts()[:5]
    plt.figure(figsize=(10,5))
    ax = sns.barplot(x=list(available.keys()),y=list(available.values),palette='Reds')
    for container in ax.containers:
        ax.bar_label(container)
```

```
except AttributeError:
    print('AttributeError Spotted')
```

```
finally:
    print('Top 5 Available Sizes')
```

AttributeError Spotted
Top 5 Available Sizes



Analysis

Most feasible price was around 40-50 USD

Average rating of Nike was of 4.5

Avg rating of 4.8 has received numerous review count of 5 within various price range

Maximum review count at avg rate 4.8 is of 120 which is Air Jordan 1 Zoom CMFT of price 150 USD

Nike Essentials within the price range of 12 USD has the avg rating of 4.7 and contains highest review count of 316

Top 5 Purchase from Nike has been of Nike sportswear being the highest followed by Nike, Nike dri FIT, Jordan and Club America

Top 5 Colors in purchase was of Black being the highest followed by White, Navy, Black/white and Multicolor

Size which is most available is S|M|L|XL|2XL

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 3:28 PM

