

```
In [1]: # Import the libraries that are important for our analysis
import pandas as pd
import numpy as nm
import seaborn as sns
import matplotlib.pyplot as plt
import re
```

```
In [2]: df = pd.read_csv('shoe sales.csv', low_memory=False) #import the data
```

```
In [3]: df.info() #information of our data
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1006 entries, 0 to 1005
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Brand           1006 non-null   object
1   Model           1006 non-null   object
2   Type            1006 non-null   object
3   Gender          1006 non-null   object
4   Size            1006 non-null   object
5   Color           1006 non-null   object
6   Material        1006 non-null   object
7   Price (USD)     1006 non-null   object
dtypes: object(8)
memory usage: 63.0+ KB
```

```
In [4]: df.shape #using this command we can know that how much columns and rows our data has.
```

```
Out[4]: (1006, 8)
```

```
In [5]: df.head() #showing our top 5 rows of our data
```

```
Out[5]:
```

	Brand	Model	Type	Gender	Size	Color	Material	Price (USD)
0	Nike	Air Jordan 1	Basketball	Men	US 10	Red/Black	Leather	\$170.00
1	Adidas	Ultra Boost 21	Running	Men	US 9.5	Black	Primeknit	\$180.00
2	Reebok	Classic Leather	Casual	Men	US 11	White	Leather	\$75.00
3	Converse	Chuck Taylor	Casual	Women	US 8	Navy	Canvas	\$55.00
4	Puma	Future Rider	Lifestyle	Women	US 7.5	Pink	Mesh	\$80.00

```
In [6]: #Converting the Column "Price (USD)" into the Integer format
df["Price (USD)"] = df["Price (USD)"].apply(lambda x: int(re.search(r'\d+', x).group()))
df.head()
```

```
Out[6]:
```

	Brand	Model	Type	Gender	Size	Color	Material	Price (USD)
0	Nike	Air Jordan 1	Basketball	Men	US 10	Red/Black	Leather	170
1	Adidas	Ultra Boost 21	Running	Men	US 9.5	Black	Primeknit	180
2	Reebok	Classic Leather	Casual	Men	US 11	White	Leather	75
3	Converse	Chuck Taylor	Casual	Women	US 8	Navy	Canvas	55
4	Puma	Future Rider	Lifestyle	Women	US 7.5	Pink	Mesh	80

```
In [7]: df.describe()
```

Out[7]: **Price (USD)**

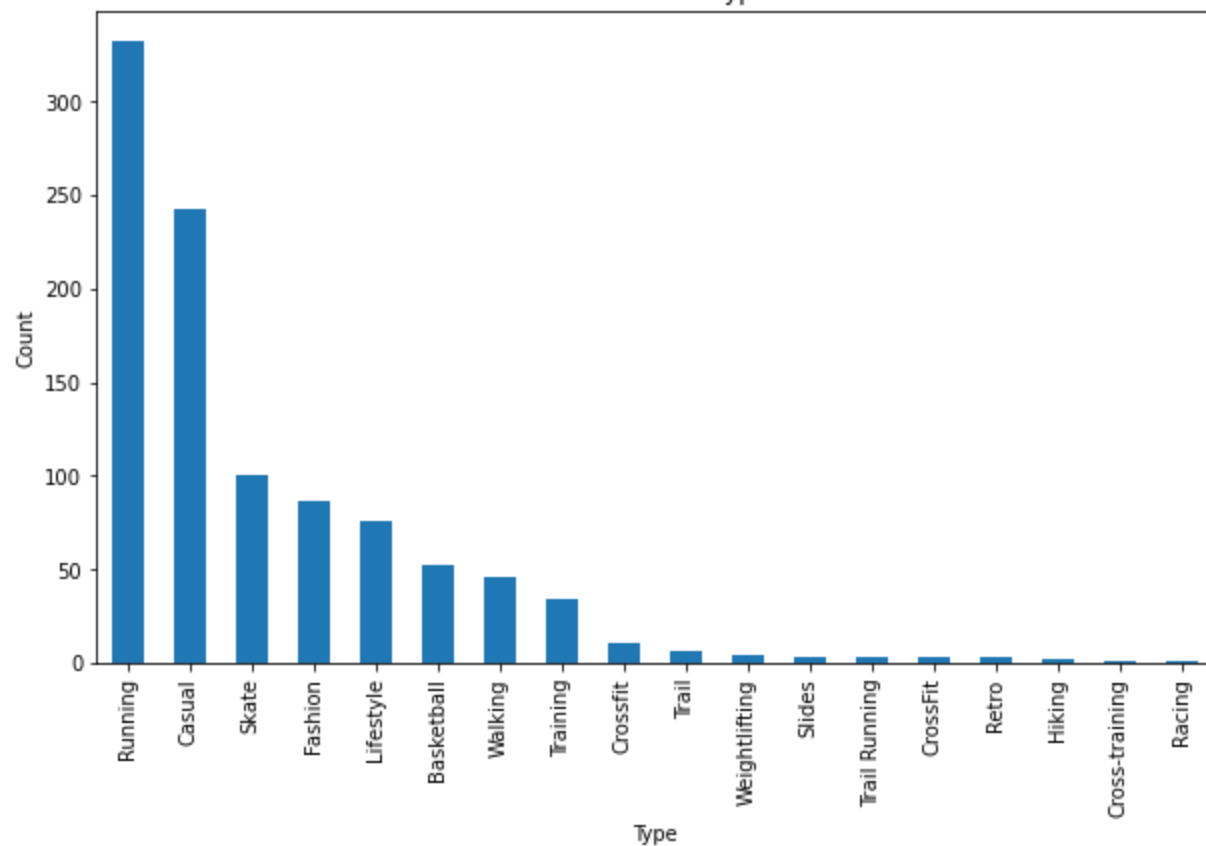
count	1006.000000
mean	101.304175
std	39.216923
min	25.000000
25%	70.000000
50%	90.000000
75%	130.000000
max	250.000000

```
In [8]: # Checking that Which type of shoes are being sold by the companies
df['Type'].value_counts()
```

```
Out[8]: Running          332
Casual          243
Skate           100
Fashion          86
Lifestyle        76
Basketball       52
Walking          46
Training         34
Crossfit         11
Trail            6
Weightlifting     4
Slides           3
Trail Running     3
CrossFit          3
Retro            3
Hiking           2
Cross-training    1
Racing           1
Name: Type, dtype: int64
```

```
In [9]: #Graphical interpretetion of Number of sales of Different type of shoes
plt.figure(figsize=(10, 6))
shoe_type = df['Type'].value_counts()
shoe_type.plot(kind='bar')
plt.title('No.of Sales of different Type of Shoes')
plt.xlabel('Type')
plt.ylabel('Count')
plt.show()
```

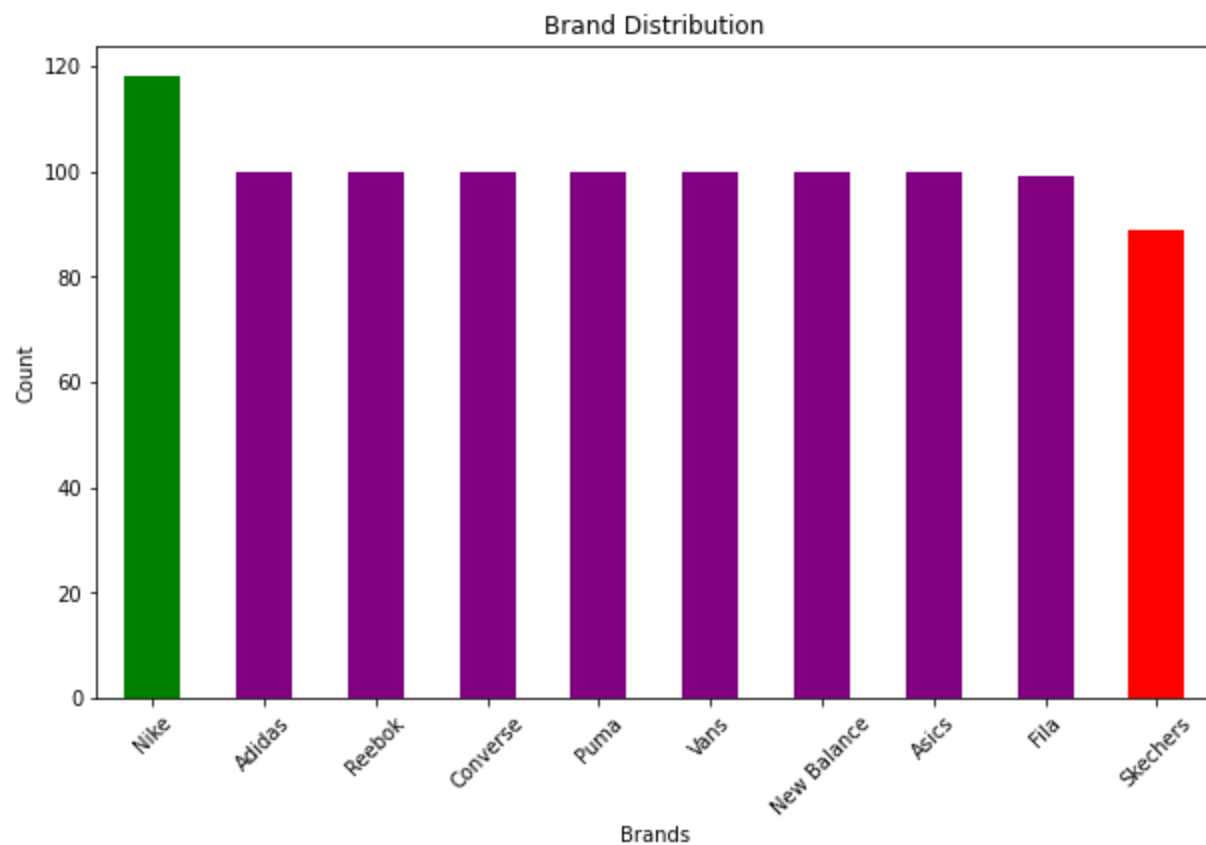
No.of Sales of different Type of Shoes



```
In [10]: #Checking the Popularity of the brand
high_sales = df['Brand'].value_counts()
high_sales
```

```
Out[10]: Nike          118
Adidas          100
Reebok          100
Converse        100
Puma            100
Vans            100
New Balance     100
Asics           100
Fila            99
Skechers        89
Name: Brand, dtype: int64
```

```
In [11]: # Graphical Interpretatiopn that which brand was most famous
plt.figure(figsize=(10, 6))
Brand_color = ['Green', 'Purple', 'Purple', 'Purple', 'Purple', 'Purple', 'Purple', 'Purple', 'P
high_sales.plot(kind='bar', color=Brand_color)
plt.title('Brand Distribution')
plt.xlabel('Brands')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```

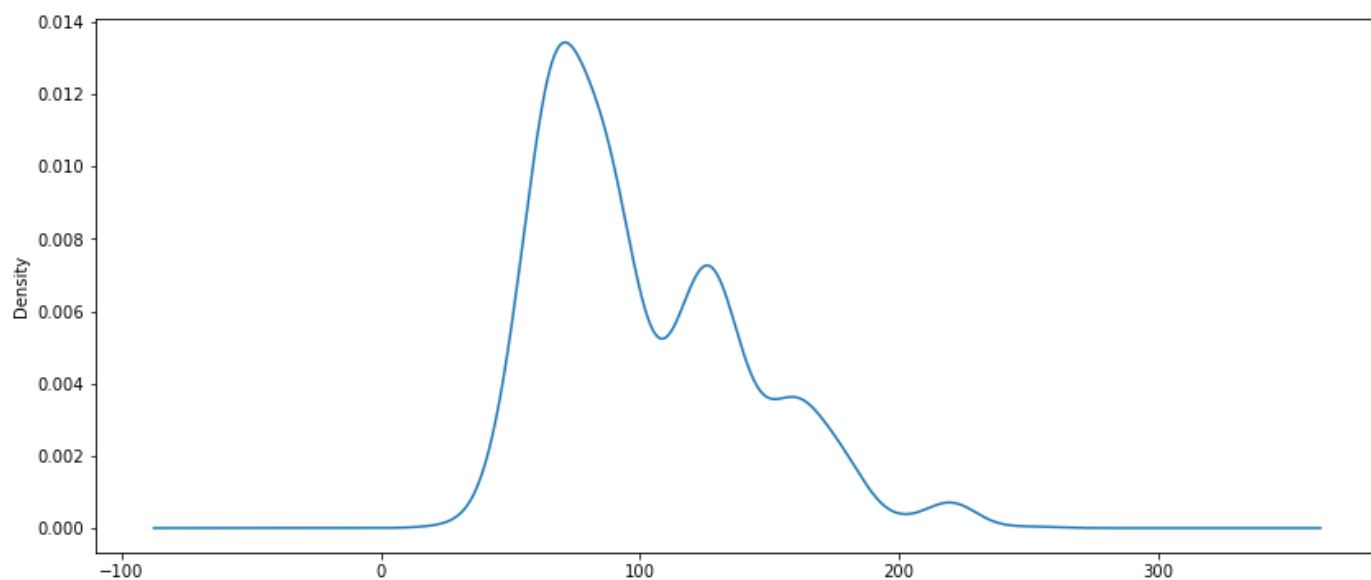


```
In [12]: df['Price (USD)'].mean()
```

```
Out[12]: 101.3041749502982
```

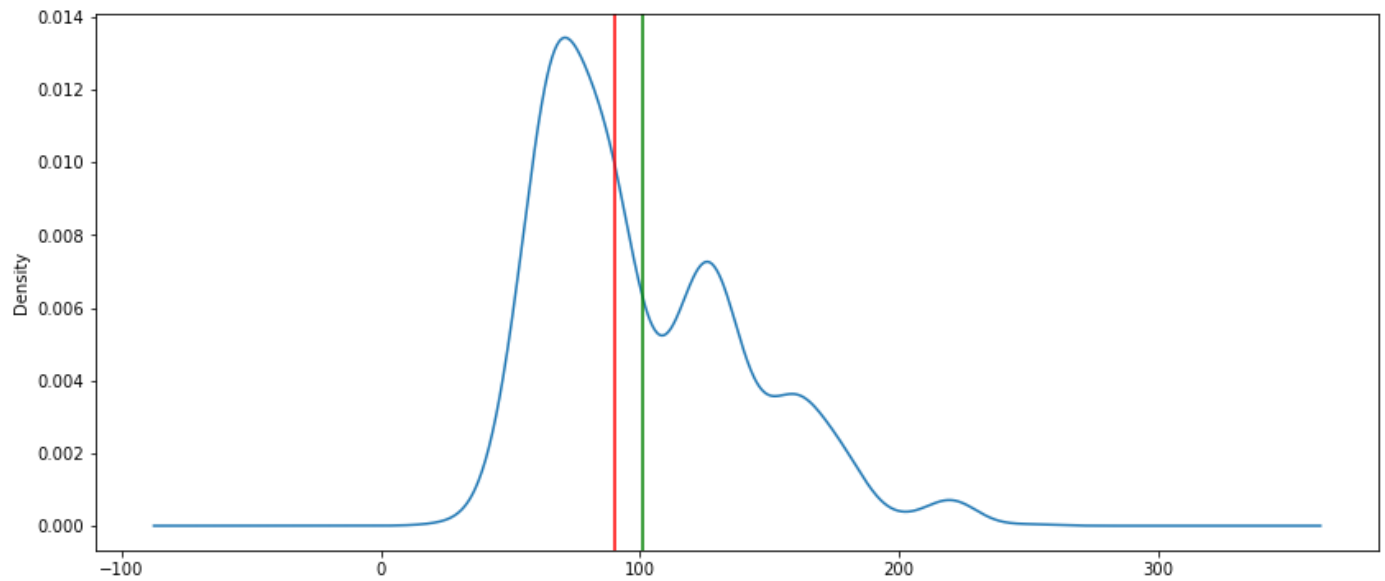
```
In [13]: # Plot for Kernal density estimation
df['Price (USD)'].plot(kind='kde', figsize=(14,6))
```

```
Out[13]: <AxesSubplot:ylabel='Density'>
```



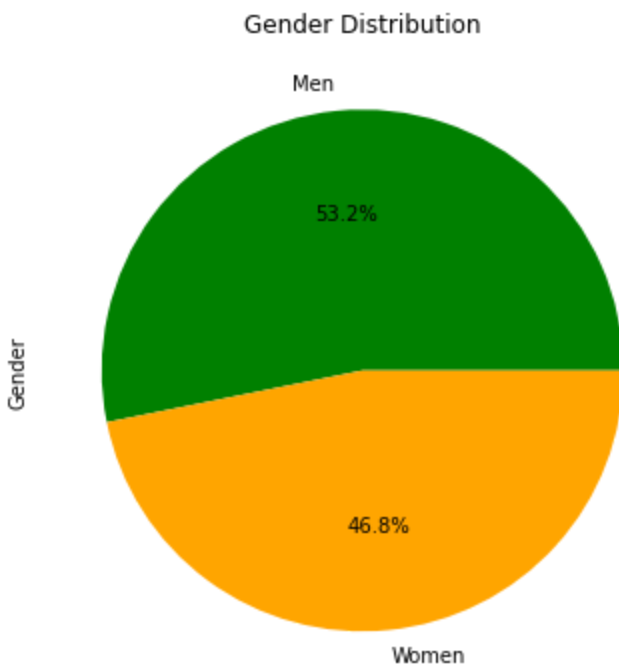
```
In [14]: ax = df['Price (USD)'].plot(kind='kde', figsize=(14,6))
ax.axvline(df['Price (USD)'].mean(),color='Green')
ax.axvline(df['Price (USD)'].median(),color='Red')
```

```
Out[14]: <matplotlib.lines.Line2D at 0x1ea2d6f2e80>
```



```
In [15]: #Checking that Which gender has bought more shoes
gender_dist = df['Gender'].value_counts()
plt.figure(figsize=(10, 6))
gender_dist.plot(kind='pie', autopct='%1.1f%%', colors=['Green', 'Orange'])
plt.title('Gender Distribution')
```

```
Out[15]: Text(0.5, 1.0, 'Gender Distribution')
```



```
In [16]: #Checking that which Model was more famous from the Nike Company
count_condition = df.loc[df['Brand'] == 'Nike', 'Model'].value_counts()

print(count_condition)
```

Air Force 1	31
Blazer Mid '77	10
Air Zoom Pegasus 38	10
React Infinity Run Flyknit 2	8
Air Force 1 '07	7
React Infinity Run Flyknit	7
Air Max 2090	6
Air Force 1 Low	4
Air Max 97	4
Air Max 90	4

React Infinity Run	4
Zoom Pegasus 38	3
Zoom Fly 3	2
Air Max 95	2
Blazer Low	2
ZoomX Vaporfly Next%	1
SB Dunk Low	1
ZoomX Invincible	1
Air Jordan 1	1
KD13	1
Blazer Mid	1
Free RN Flyknit	1
Flyknit Trainer	1
Air Max 270	1
ZoomX Invincible Run	1
React Infinity Run 2	1
Air Zoom Tempo NEXT% FlyEase	1
Free RN 5.0	1
Epic React Flyknit	1

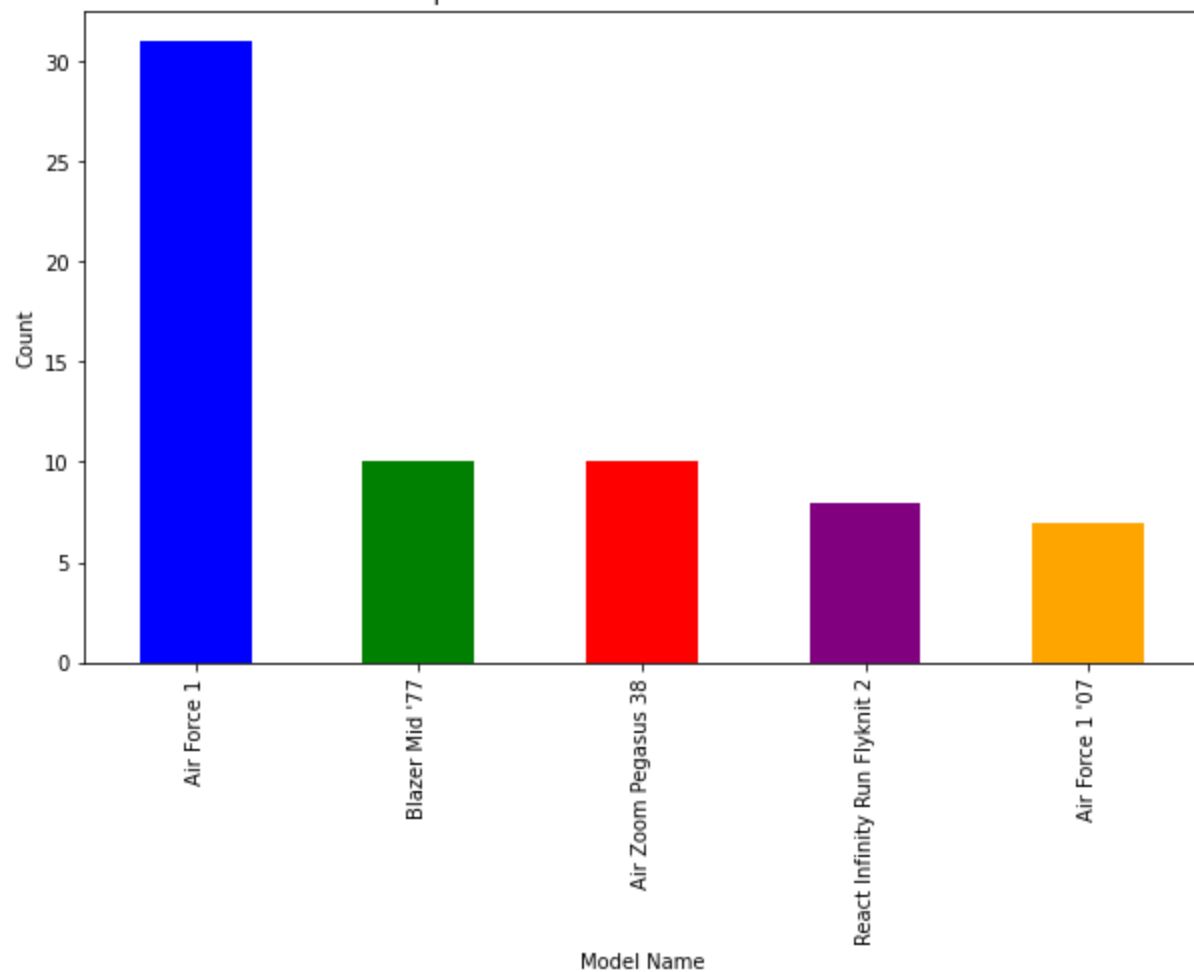
Name: Model, dtype: int64

```
In [17]: Top_models = count_condition.head(5)
Top_models
```

```
Out[17]: Air Force 1          31
Blazer Mid '77          10
Air Zoom Pegasus 38     10
React Infinity Run Flyknit 2    8
Air Force 1 '07          7
Name: Model, dtype: int64
```

```
In [18]: plt.figure(figsize=(10,6))
model_colors = ['blue', 'green', 'red', 'purple', 'orange']
Top_models.plot(kind='bar',color=model_colors)
plt.title('Top 5 Models of Nike in terms of SALES')
plt.xlabel('Model Name')
plt.ylabel('Count')
plt.show()
```

Top 5 Models of Nike in terms of SALES



```
In [19]: #Checking that which brand was more famous in Adidas.
Top_adidas = df.loc[df['Brand'] == 'Adidas', 'Model'].value_counts().head(5)
Top_adidas
```

```
Out[19]: NMD_R1          39
Superstar          14
Yeezy Boost 350 V2  12
Ultraboost DNA      8
Yeezy Boost 350      5
Name: Model, dtype: int64
```

```
In [21]: plt.figure(figsize=(10,6))
Top_adidas.plot(kind='bar')
plt.title('Top 5 saling models of Adidas')
plt.xlabel('Model Name')
plt.ylabel('Count')
plt.show()
```

Top 5 saling models of Adidas

