Python Basic Data Analyst

In [214…
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Files supported by pandas including (csv, excel, sql, json, parquet, ...)

In [215…
```
data = pd.read_csv("/Users/user/Downloads/Flavors.csv")
```

In [216…
```
print(data)
```

```
                  Flavor Base Flavor Liked  Flavor Rating  Texture Rating  \
0     Mint Chocolate Chip     Vanilla   Yes           10.0             8.0
1               Chocolate   Chocolate   Yes            8.8             7.6
2                 Vanilla     Vanilla    No            4.7             5.0
3            Cookie Dough     Vanilla   Yes            6.9             6.5
4              Rocky Road   Chocolate   Yes            8.2             7.0
5               Pistachio     Vanilla    No            2.3             3.4
6             Cake Batter     Vanilla   Yes            6.5             6.0
7              Neapolitan     Vanilla    No            3.8             5.0
8  Chocolte Fudge Brownie   Chocolate   Yes            8.2             7.1

   Total Rating
0          18.0
1          16.6
2           9.7
3          13.4
4          15.2
5           5.7
6          12.5
7           8.8
8          15.3
```

Exploring the data.

View the first five rows.

In [217…
```
print(data.head())
```

```
               Flavor Base Flavor Liked  Flavor Rating  Texture Rating  \
0  Mint Chocolate Chip     Vanilla   Yes           10.0             8.0
1            Chocolate   Chocolate   Yes            8.8             7.6
2              Vanilla     Vanilla    No            4.7             5.0
3         Cookie Dough     Vanilla   Yes            6.9             6.5
4           Rocky Road   Chocolate   Yes            8.2             7.0

   Total Rating
0          18.0
1          16.6
2           9.7
3          13.4
4          15.2
```

View the last five rows.

In [218…
```
print(data.tail())
```

```
                   Flavor Base Flavor Liked  Flavor Rating  Texture Rating  \
4              Rocky Road   Chocolate   Yes            8.2             7.0
5               Pistachio     Vanilla    No            2.3             3.4
6             Cake Batter     Vanilla   Yes            6.5             6.0
7              Neapolitan     Vanilla    No            3.8             5.0
8  Chocolte Fudge Brownie   Chocolate   Yes            8.2             7.1

   Total Rating
4          15.2
5           5.7
6          12.5
7           8.8
8          15.3
```

Get summary statistics.

In [219…
```
print(data.describe())
```

```
       Flavor Rating  Texture Rating  Total Rating
count         9.0000        9.000000      9.000000
mean          6.6000        6.177778     12.800000
std           2.5387        1.478832      4.030509
min           2.3000        3.400000      5.700000
25%           4.7000        5.000000      9.700000
50%           6.9000        6.500000     13.400000
75%           8.2000        7.100000     15.300000
max          10.0000        8.000000     18.000000
```

Check the data types and missing values:

In [220…
```
print(data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9 entries, 0 to 8
Data columns (total 6 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Flavor          9 non-null      object 
 1   Base Flavor     9 non-null      object 
 2   Liked           9 non-null      object 
 3   Flavor Rating   9 non-null      float64
 4   Texture Rating  9 non-null      float64
 5   Total Rating    9 non-null      float64
dtypes: float64(3), object(3)
memory usage: 564.0+ bytes
None
```

Data Cleansing.

Handle missing values.

In [221…
```
print(data.isnull())
```

	Flavor	Base Flavor	Liked	Flavor Rating	Texture Rating	Total Rating
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
5	False	False	False	False	False	False
6	False	False	False	False	False	False
7	False	False	False	False	False	False
8	False	False	False	False	False	False

In [222…
```
print(data.isnull().sum())
```

```
Flavor            0
Base Flavor       0
Liked             0
Flavor Rating     0
Texture Rating    0
Total Rating      0
dtype: int64
```

Filling or dropping empty cell.

In [223…
```
print(data.fillna("value"))
```

```
                  Flavor Base Flavor Liked  Flavor Rating  Texture Rating  \
0     Mint Chocolate Chip     Vanilla   Yes           10.0             8.0
1               Chocolate   Chocolate   Yes            8.8             7.6
2                 Vanilla     Vanilla    No            4.7             5.0
3            Cookie Dough     Vanilla   Yes            6.9             6.5
4              Rocky Road   Chocolate   Yes            8.2             7.0
5               Pistachio     Vanilla    No            2.3             3.4
6             Cake Batter     Vanilla   Yes            6.5             6.0
7              Neapolitan     Vanilla    No            3.8             5.0
8  Chocolte Fudge Brownie   Chocolate   Yes            8.2             7.1

   Total Rating
0          18.0
1          16.6
2           9.7
3          13.4
4          15.2
5           5.7
6          12.5
7           8.8
8          15.3
```

Filling specific column.

In [224…

```
print(data["Flavor Rating"].fillna("value"))
```

```
0    10.0
1     8.8
2     4.7
3     6.9
4     8.2
5     2.3
6     6.5
7     3.8
8     8.2
Name: Flavor Rating, dtype: float64
```

Droppign null value columns.

In [225…

```
print(data.dropna())
```

```
                  Flavor Base Flavor Liked  Flavor Rating  Texture Rating  \
0     Mint Chocolate Chip     Vanilla   Yes           10.0             8.0
1               Chocolate   Chocolate   Yes            8.8             7.6
2                 Vanilla     Vanilla    No            4.7             5.0
3            Cookie Dough     Vanilla   Yes            6.9             6.5
4              Rocky Road   Chocolate   Yes            8.2             7.0
5               Pistachio     Vanilla    No            2.3             3.4
6             Cake Batter     Vanilla   Yes            6.5             6.0
7              Neapolitan     Vanilla    No            3.8             5.0
8  Chocolte Fudge Brownie   Chocolate   Yes            8.2             7.1

   Total Rating
0          18.0
1          16.6
2           9.7
3          13.4
4          15.2
5           5.7
6          12.5
7           8.8
8          15.3
```

Dropping specific column null values.

In [226…

```
data.dropna(subset=["Flavor Rating"], inplace = True)
```

Returns True for every row that is a duplicate.

In [227…

```
print(data.duplicated())
```

```
0    False
1    False
2    False
3    False
4    False
5    False
6    False
7    False
8    False
dtype: bool
```

View Columns data type.

In [228…

```
data.dtypes
```

Out[228]:
```
Flavor            object
Base Flavor       object
Liked             object
Flavor Rating     float64
Texture Rating    float64
Total Rating      float64
dtype: object
```

View your data Columns.

In [229…
```
print(data.columns)
```

```
Index(['Flavor', 'Base Flavor', 'Liked', 'Flavor Rating', 'Texture Rating',
       'Total Rating'],
      dtype='object')
```

Convert data types.

In [230…
```
data['Flavor Rating'] = data['Flavor Rating'].astype(int)
```

Data analysis and manipulation.

Select multiple columns.

In [231…
```
print(data[['Flavor','Base Flavor']])
```

```
                  Flavor Base Flavor
0     Mint Chocolate Chip     Vanilla
1               Chocolate   Chocolate
2                 Vanilla     Vanilla
3            Cookie Dough     Vanilla
4              Rocky Road   Chocolate
5               Pistachio     Vanilla
6             Cake Batter     Vanilla
7              Neapolitan     Vanilla
8  Chocolte Fudge Brownie   Chocolate
```

Select rows by index.

In [232…
```
print(data.loc[1])
```

```
Flavor            Chocolate
Base Flavor       Chocolate
Liked                   Yes
Flavor Rating             8
Texture Rating          7.6
Total Rating           16.6
Name: 1, dtype: object
```

Sorting by Column.

In [233…
```
print(data.sort_values('Flavor'))
```

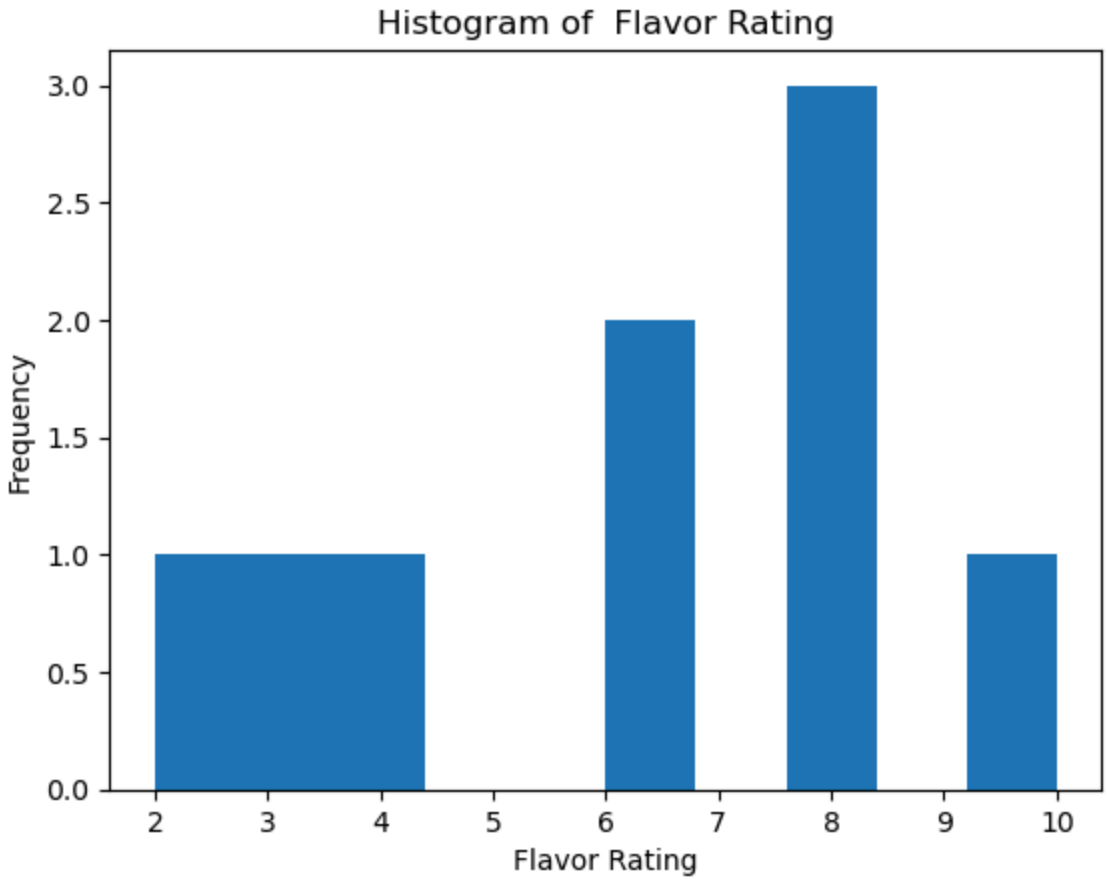
```
                  Flavor Base Flavor Liked  Flavor Rating  Texture Rating  \
6            Cake Batter     Vanilla   Yes              6             6.0
1              Chocolate   Chocolate   Yes              8             7.6
8  Chocolte Fudge Brownie   Chocolate   Yes              8             7.1
3           Cookie Dough     Vanilla   Yes              6             6.5
0    Mint Chocolate Chip     Vanilla   Yes             10             8.0
7             Neapolitan     Vanilla    No              3             5.0
5              Pistachio     Vanilla    No              2             3.4
4             Rocky Road   Chocolate   Yes              8             7.0
2                Vanilla     Vanilla    No              4             5.0

   Total Rating
6          12.5
1          16.6
8          15.3
3          13.4
0          18.0
7           8.8
5           5.7
4          15.2
2           9.7
```

Visualization.

Histogram.

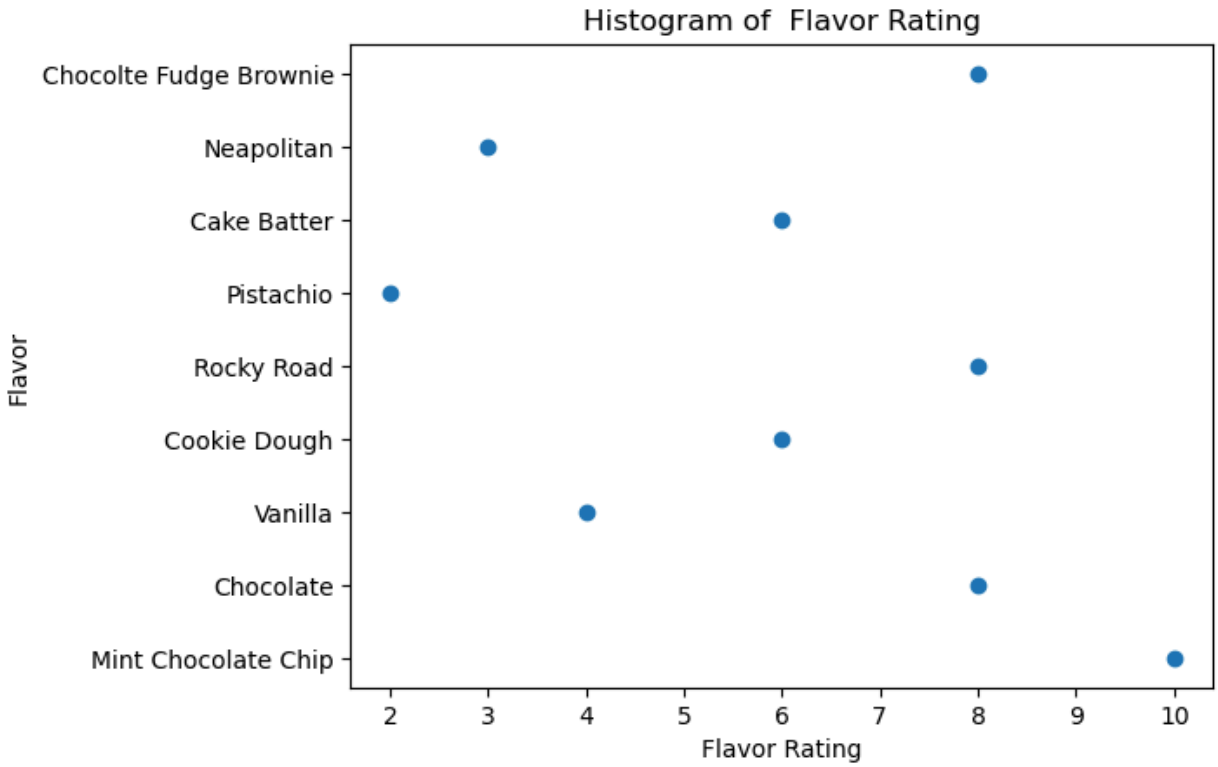
In [234…
```
plt.hist(data['Flavor Rating'])
plt.xlabel('Flavor Rating')
plt.ylabel('Frequency')
plt.title('Histogram of  ' + 'Flavor Rating')
plt.show()
```



Scatter.

In [235…

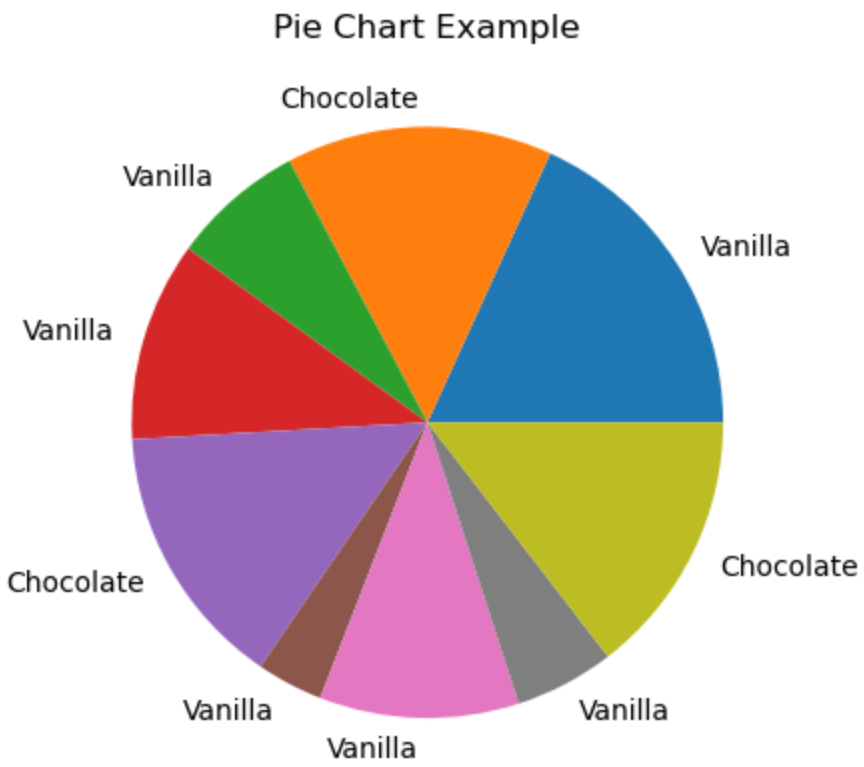
```
plt.scatter(data['Flavor Rating'],data['Flavor'])
plt.xlabel('Flavor Rating')
plt.ylabel('Flavor')
plt.title('Histogram of  ' + 'Flavor Rating')
plt.show()
```



Pie Chart.

In [236…

```
labels = data['Base Flavor']
values = data['Flavor Rating']
plt.pie(values, labels=labels)
plt.title("Pie Chart Example")
plt.show()
```


Pie Chart Example
Chocolate
Vanilla
Vanilla
Vanilla
Chocolate
Chocolate
Vanilla
Vanilla
Vanilla

In [ ]: