WEEKLY PROJECT 17-SANJAY ANAND V

Data Visualisation on Sales Dataset

1)Importing libraries and dataset

```python
In [37]: import numpy as np
         import pandas as pd
         import datetime
         import matplotlib.pyplot as plt
         import seaborn as sns
```

```python
In [38]: df=pd.read_csv('sales.csv')
         df.head()
```

Out[38]:

	order_id	product	quantity	price	date	address	month
0	236670	Wired Headphones	2	11.99	08/31/19 22:21	359 Spruce St, Seattle, WA 98101	August
1	236671	Bose SoundSport Headphones	1	99.99	08/15/19 15:11	492 Ridge St, Dallas, TX 75001	August
2	236672	iPhone	1	700.0	08/06/19 14:40	149 7th St, Portland, OR 97035	August
3	236673	AA Batteries (4-pack)	2	3.84	08/29/19 20:59	631 2nd St, Los Angeles, CA 90001	August
4	236674	AA Batteries (4-pack)	2	3.84	08/15/19 19:53	736 14th St, New York City, NY 10001	August

2)Dealing with null values

```python
In [39]: df.isnull().sum()
```

```
Out[39]: order_id    0
         product     0
         quantity    0
         price       0
         date        0
         address     0
         month       0
         dtype: int64
```

Preprocessing date column

```python
In [40]: Date=[]
         for i in df['date']:
             Date.append(i.split(" ")[0])
         df['Date']=Date
         time=[]
         for i in df['date']:
             time.append(i.split(" ")[1])
         df['time']=time

         del df['date']


         df.head()
```

Out[40]:

	order_id	product	quantity	price	address	month	Date	time
0	236670	Wired Headphones	2	11.99	359 Spruce St, Seattle, WA 98101	August	08/31/19	22:21
1	236671	Bose SoundSport Headphones	1	99.99	492 Ridge St, Dallas, TX 75001	August	08/15/19	15:11
2	236672	iPhone	1	700.0	149 7th St, Portland, OR 97035	August	08/06/19	14:40
3	236673	AA Batteries (4-pack)	2	3.84	631 2nd St, Los Angeles, CA 90001	August	08/29/19	20:59
4	236674	AA Batteries (4-pack)	2	3.84	736 14th St, New York City, NY 10001	August	08/15/19	19:53

3)Number of unique orders placed

```python
In [41]: c=len((df['order_id']).unique())
         print("The number of unique orders placed so far are",c)
```

```
The number of unique orders placed so far are 178438
```

4)Number of unique products in the dataset

```python
In [42]: l=len((df['product']).unique())
         print("The number of unique products in the dataset are",l)
```

```
The number of unique products in the dataset are 20
```

5) The Most Expensive product

```python
In [43]: df['price'].unique()
```

```
Out[43]: array(['11.99', '99.99', '700.0', '3.84', '379.99', '109.99', '1700.0',
                '600.0', '149.99', '14.95', '150.0', '2.99', '11.95', '389.99',
                '999.99', '300.0', '400.0', 'Price Each', '600', '150', '1700',
                '300', '400', '700'], dtype=object)
```

```python
In [44]: c=len(df['order_id'])
         for i in range(c):
             if 'Price' in df['price'][i]:
                 df['price'][i]=0
             # print(df['price'][i])
             #if(df['price'][i]=='Price Each'):
                 #df.replac(e(df['price'][i],'0')
         df['price'].unique()
```

```
Out[44]: array(['11.99', '99.99', '700.0', '3.84', '379.99', '109.99', '1700.0',
                '600.0', '149.99', '14.95', '150.0', '2.99', '11.95', '389.99',
                '999.99', '300.0', '400.0', 0, '600', '150', '1700', '300', '400',
                '700'], dtype=object)
```

```python
In [45]: q=len(df['order_id'])
         v=[]
         for i in range(q):
             v.append(int(float(df['price'][i])))
         w=max(v)
         df['price']=v
         for j in range(q):
             if(df['price'][j]==w):
                 k=df['product'][j]
         print("The most expensive product on the list is",k,"and its price is",w,".")
```

```
The most expensive product on the list is Macbook Pro Laptop and its price is 1700 .
```

6)Address from which the most number of orders are placed

```python
In [46]: q=len(df['order_id'])
         for i in range(q):
             if (df['address'][i]=='Purchase Address'):
                 df=df.drop(i)
         most_orders_address = df['address'].value_counts().idxmax()
         print("The address with the most orders is: ", most_orders_address)
```
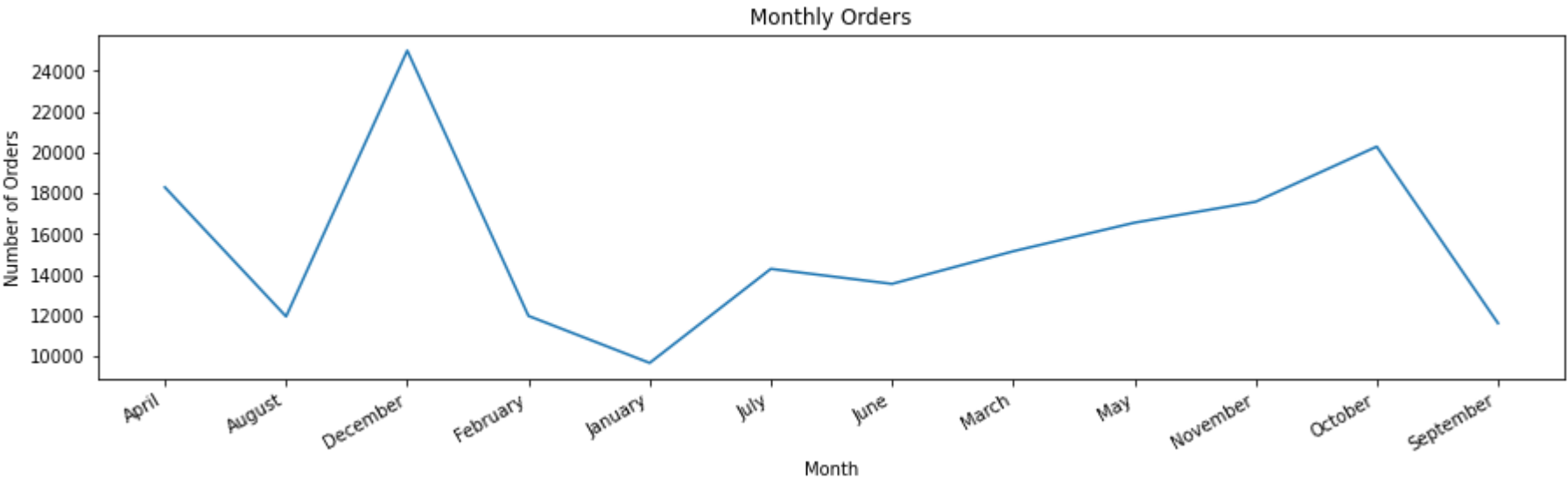
```
The address with the most orders is:  193 Forest St, San Francisco, CA 94016
```

7)Number of orders in each month - Line graph

```python
In [47]: monthly_orders = df.groupby('month')['order_id'].count().reset_index()
         fig, axs = plt.subplots(figsize = (15,4))

         sns.lineplot(x='month', y='order_id', data=monthly_orders)

         fig.autofmt_xdate()
         plt.xlabel('Month')
         plt.ylabel('Number of Orders')
         plt.title('Monthly Orders')
         plt.show()
```



```
In [ ]:
```