In [32]:
```
import pandas as pd
import os
```

In [55]:
```
annual_sale=pd.read_csv("https://raw.githubusercontent.com/svkarthik86/Assignment/main/Sales_Data/Sales_April_2019.csv")
annual_sale.head(5)
```

Out[55]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st St, Dallas, TX 75001
1	NaN	NaN	NaN	NaN	NaN	NaN
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	176560	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001

In [49]:
```
print(annual_sale.info())
annual_sale.isna().sum().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18383 entries, 0 to 18382
Data columns (total 6 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   Order ID          18324 non-null  object
 1   Product           18324 non-null  object
 2   Quantity Ordered  18324 non-null  object
 3   Price Each        18324 non-null  object
 4   Order Date        18324 non-null  object
 5   Purchase Address  18324 non-null  object
dtypes: object(6)
memory usage: 861.8+ KB
None
```

Out[49]:
```
354
```

In [35]:
```
annual_sale.Product.memory_usage()
```

Out[35]:
```
147192
```

In [36]:
```
annual_sale.Product.value_counts()[:1]
```

Out[36]:
```
Lightning Charging Cable    2201
Name: Product, dtype: int64
```

In [37]:
```
annual_sale.columns=[i.replace(" ","_") for i in annual_sale.columns]
```

In [38]:
```
annual_sale.dropna(inplace=True)
```

In [39]:
```
annual_sale.Product.memory_usage()
```

Out[39]:
```
293184
```

In [40]:
```
annual_sale.Product=annual_sale.Product.astype("category")
annual_sale.Product.memory_usage()
```

Out[40]:
```
165632
```

In [41]:
```
annual_sale['month']=annual_sale.Order_Date.str[:2]
```

In [46]:
```
annual_sale.Product.value_counts()[:1]
```

Out[46]:
```
Lightning Charging Cable    2201
Name: Product, dtype: int64
```

In [54]:
```
annual_sale[annual_sale.Price_Each.astype(float)>200].loc[:,["Product",]]
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
Input In [54], in <cell line: 1>()
----> 1 annual_sale[annual_sale.Price_Each.astype(float)>200].loc[:,["Product",]]

File C:\anconda\lib\site-packages\pandas\core\generic.py:5575, in NDFrame.__getattr__(self, name)
   5568 if (
   5569     name not in self._internal_names_set
   5570     and name not in self._metadata
   5571     and name not in self._accessors
   5572     and self._info_axis._can_hold_identifiers_and_holds_name(name)
   5573 ):
   5574     return self[name]
-> 5575 return object.__getattribute__(self, name)

AttributeError: 'DataFrame' object has no attribute 'Price_Each'
```

In [ ]: