PROGRAM 8: PIVOTING USING PANDAS IN PYTHON

DESCRIPTION:

Mr. Jeff Bezos, the CEO of Amazon wants an immediate report of sales of products from Amazon for the fiscal year 2019-2020 (only for the USA). He is looking forward towards you, the emerging data analysts to help him in analy zing the data and come up with potential observations that can help him in improving his business. Do lend a hel ping hand so that your report helps him in better decision making.

NOTE: Use the concept of pivot_table(), query() and aggregation mandatorily to accomplish the task.

Importing required library & data.!

In [2]:
```
import calendar
import numpy as np
import pandas as pd
import seaborn as sns
import plotly.io as pio
import plotly.express as px
import matplotlib.pyplot as plt
import plotly.graph_objects as go
```

In [3]:
```
Data = pd.read_excel(open('test.xlsx','rb'),sheet_name='Sheet1')
Original = Data
Data.head(7)
```

Out[3]:

	CustomerName	Segment	City	Code	State	Category	SubCategory	Sales_in_M	Qty	Month
0	Claire Gute	Consumer	Henderson	KY	Kentucky	Furniture	Bookcases	261.9600	11685	7
1	Claire Gute	Consumer	Henderson	KY	Kentucky	Furniture	Chairs	731.9400	18572	5
2	Darrin Van Huff	Corporate	Los Angeles	CA	California	Office Supplies	Labels	14.6200	13235	9
3	Sean O'Donnell	Consumer	Fort Lauderdale	FL	Florida	Furniture	Tables	957.5775	10416	4
4	Sean O'Donnell	Consumer	Fort Lauderdale	FL	Florida	Office Supplies	Storage	22.3680	4064	9
5	Brosina Hoffman	Consumer	Los Angeles	CA	California	Furniture	Furnishings	48.8600	1442	8
6	Brosina Hoffman	Consumer	Los Angeles	CA	California	Office Supplies	Art	7.2800	9729	8

Data Cleaning,Data Transformation,Combining Data

In [4]:
```
# Distinct

DistinctCount_Data = Data.nunique(axis=0)
print(DistinctCount_Data)
```

```
CustomerName     793
Segment            3
City             529
Code              48
State             49
Category           3
SubCategory       17
Sales_in_M      5757
Qty             7712
Month             12
dtype: int64
```

In [5]:
```
# Checking NaN

Checking_NaN = Data[Data.isna().any(axis=1)]
Checking_NaN.head()
```

Out[5]:

CustomerName	Segment	City	Code	State	Category	SubCategory	Sales_in_M	Qty	Month

In [6]:
```
# Converting sales amount, Calculating Unit price, Month format and Adding Location info.

Data["Sales"] = round(Data.Sales_in_M * 1000000)
Data["Unit_Price"] = round(Data.Sales/Data.Qty,2)
Data["loc"] = Data.City+'('+Data.Code+')'
Data.drop(Data.columns[[2]], axis = 1, inplace = True)
Data['Month'] = Data['Month'].apply(lambda x: calendar.month_abbr[x])
Data.head(5)
```

Out[6]:

	CustomerName	Segment	Code	State	Category	SubCategory	Sales_in_M	Qty	Month	Sales	Unit_Price	loc
0	Claire Gute	Consumer	KY	Kentucky	Furniture	Bookcases	261.9600	11685	Jul	261960000.0	22418.49	Henderson(KY)
1	Claire Gute	Consumer	KY	Kentucky	Furniture	Chairs	731.9400	18572	May	731940000.0	39410.94	Henderson(KY)
2	Darrin Van Huff	Corporate	CA	California	Office Supplies	Labels	14.6200	13235	Sep	14620000.0	1104.65	Los Angeles(CA)
3	Sean O'Donnell	Consumer	FL	Florida	Furniture	Tables	957.5775	10416	Apr	957577500.0	91933.32	Fort Lauderdale(FL)
4	Sean O'Donnell	Consumer	FL	Florida	Office Supplies	Storage	22.3680	4064	Sep	22368000.0	5503.94	Fort Lauderdale(FL)

In [7]:
```
Data.drop(Data.columns[[9]], axis = 1, inplace = True)
Original = Data
Data.keys()
```

Out[7]:
```
Index(['CustomerName', 'Segment', 'Code', 'State', 'Category', 'SubCategory',
       'Sales_in_M', 'Qty', 'Month', 'Unit_Price', 'loc'],
      dtype='object')
```

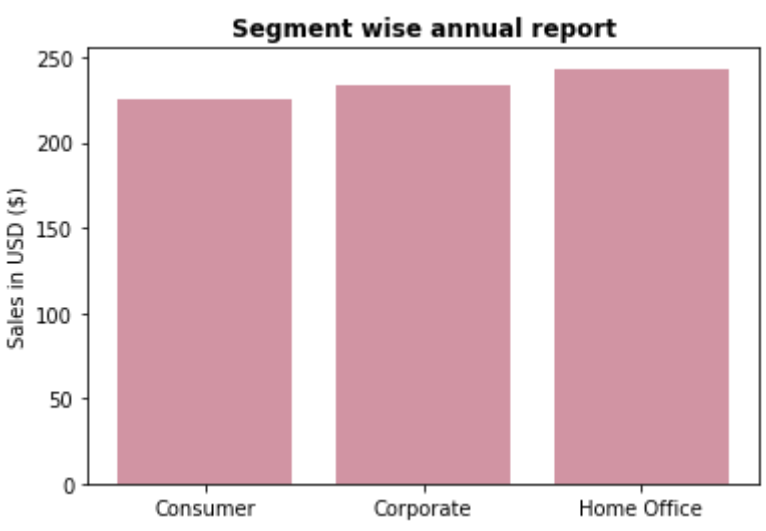
Data Exploration, Analysis & Presentation

In [8]:
```
# Segment

report_a = pd.pivot_table(Data,index=["Segment"])
report_a
```

Out[8]:

Segment	Qty	Sales_in_M	Unit_Price
Consumer	9906.075083	225.065777	132449.210696
Corporate	9997.974263	233.150720	77249.566285
Home Office	9832.293814	243.403309	62924.670624

In [9]:
```
height = report_a["Sales_in_M"]
bars = ("Consumer","Corporate","Home Office")
y_pos = np.arange(len(bars))

plt.bar(y_pos, height, color=(0.7, 0.3, 0.4, 0.6))
plt.xticks(y_pos, bars)
plt.title("Segment wise annual report",fontweight="bold")
plt.ylabel("Sales in USD ($)")
plt.show()
```



Interpretation for Segment wise annual report:

```
Highest Turnover    - Home Office
Average Turnover    - Corporate
Lowest  Turnover    - Consumer
```

In [10]:
```
# Sales Mean value by double indexing Segment & Category.

report_b = pd.pivot_table(Data,
                          index=["Segment","Category"],
                          values=["Sales_in_M"],
                          aggfunc=[np.sum,np.mean,np.std]
                         )
report_b
```

Out[10]:

		sum	mean	std
		Sales_in_M	Sales_in_M	Sales_in_M
Segment	Category			
Consumer	Furniture	387696.2580	354.708379	526.119649
	Office Supplies	359352.6080	116.976760	420.559840
	Technology	401011.6650	428.431266	942.456757
Corporate	Furniture	220321.7018	350.830735	469.294323
	Office Supplies	224130.5360	125.704170	347.746356
	Technology	244041.8370	450.261692	1100.079021
Home Office	Furniture	120640.6159	337.928896	479.687283
	Office Supplies	121939.1900	115.691831	325.144577
	Technology	182402.3710	544.484690	1519.861125

In [11]:
```
report_b = report_b.reset_index()
report_b.columns=["Segment","Category","Sales_Sum","Sales_mean","Sales_std"]
fig = px.line(report_b,
              x="Segment", y="Sales_mean",
              height=350, color='Category',
              title='Sales Mean value'
             )
fig.show()
```

Interpretation for Sales Mean-Values

Notably products under:-

Technology Category under every segments proving higher revenue in sales.
Where,
Office Supplies left idle and Furniture sales remains consistance.

Comparision wise Furniture is falling down but Office supplies remains constant.

In [12]:
```
# Segment & Category in details

report_c = pd.pivot_table(Data,
                          index=["Segment","Category"],
                          values=["Qty"],
                          aggfunc=[np.sum,np.max,np.min]
                         )
report_c
```

Out[12]:

		sum	amax	amin
		Qty	Qty	Qty
Segment	**Category**			
Consumer	Furniture	10938885	19984	27
	Office Supplies	30322583	19959	1
	Technology	9269421	19993	10
Corporate	Furniture	6270396	19942	37
	Office Supplies	17756633	19998	17
	Technology	5496989	19996	39
Home Office	Furniture	3605910	19959	78
	Office Supplies	10359103	19999	27
	Technology	3202172	19971	183

In [13]:
```
report_c = report_c.reset_index()
report_c.columns=["Segment","Category","Total_Quantity","Max_Quantity","Min_Quantity"]

fig = px.bar(report_c, x="Segment", y="Total_Quantity",
             color='Category', barmode='group',
             title="Segment-wise Category Report", height=350)
fig.show()
```

Insight form products and their sales quantity

As per previous figure interpretation, we concluded that Technology lead on company revenue.
But here, Technology pulled to bottom.

It's clearly shoes that's Quantity wise Technology based products sold very less comparatively.
But Technology was been expensive then other category.

In [14]:
```
# State wise, Category profit aggregation

report_d = pd.pivot_table(Data,
                index=["Segment","State"],
                values=["Sales_in_M"],
                columns=["Category"],
                aggfunc=[np.sum, np.max, np.min],fill_value=0)
report_d
```

Out[14]:

		sum			amax			amin		
		Sales_in_M			Sales_in_M			Sales_in_M		
	Category	Furniture	Office Supplies	Technology	Furniture	Office Supplies	Technology	Furniture	Office Supplies	Technology
Segment	State									
Consumer	Alabama	4038.620	1799.180	1699.740	1819.860	900.080	979.950	8.960	7.270	29.000
	Arizona	5424.698	4461.324	6529.056	933.536	946.764	1879.960	4.272	2.025	23.200
	Arkansas	2413.370	3551.770	2836.870	1067.940	1793.980	881.930	22.200	6.630	26.850
	California	74931.726	69215.106	78272.218	3610.848	8187.650	4476.800	3.480	1.440	5.560
	Colorado	4870.625	4483.395	6173.952	662.880	1332.496	1983.968	4.224	1.080	46.688
...	...	...	...	...	...	...	...	...	...	...
	Virginia	4542.530	2118.020	791.060	2244.480	489.920	187.980	12.420	3.760	21.800
	Washington	12448.152	3913.502	8645.246	2348.820	359.320	2395.200	4.180	2.496	7.992
Home Office	West Virginia	0.000	536.480	0.000	0.000	447.840	0.000	0.000	6.240	0.000
	Wisconsin	1788.040	2122.730	1575.850	687.400	629.100	999.960	47.400	3.600	1.980
	Wyoming	1603.136	0.000	0.000	1603.136	0.000	0.000	1603.136	0.000	0.000

140 rows × 9 columns

In [15]:
```
report_d.query('Segment == ["Corporate"] & State == ["New Jersey","New Mexico","New York"]')
```

Out[15]:

		sum			amax			amin		
		Sales_in_M			Sales_in_M			Sales_in_M		
	Category	Furniture	Office Supplies	Technology	Furniture	Office Supplies	Technology	Furniture	Office Supplies	Technology
Segment	State									
Corporate	New Jersey	2417.650	5025.480	10816.120	854.940	1793.98	9099.93	8.28	5.76	90.00
	New Mexico	0.000	447.746	822.030	0.000	255.85	595.00	0.00	4.17	67.04
	New York	24604.817	23841.070	26382.464	1579.746	2504.74	5199.96	9.82	3.52	2.97

In [16]:
```
report_d.query('State == ["California","Georgia","Texas","Washington"]')
```

Out[16]:

		sum			amax			amin		
		Sales_in_M			Sales_in_M			Sales_in_M		
	Category	Furniture	Office Supplies	Technology	Furniture	Office Supplies	Technology	Furniture	Office Supplies	Technology
Segment	State									
Consumer	California	74931.7260	69215.106	78272.218	3610.8480	8187.650	4476.800	3.480	1.440	5.560
	Georgia	4605.1900	11892.450	7619.150	1628.8200	6354.950	2939.930	18.840	1.780	67.800
	Texas	35640.5260	28754.842	30965.362	2453.4300	3930.072	8159.952	1.988	0.444	7.992
	Washington	21925.3400	21218.740	30722.440	2665.6200	3266.376	13999.960	6.240	1.344	11.540
Corporate	California	49091.3405	51539.578	44100.492	2887.0560	5083.960	4799.984	6.280	2.940	7.920
	Georgia	1638.1100	9489.880	4568.170	595.3800	2690.970	2999.950	20.240	3.760	5.950
	Texas	17245.5018	10702.858	25408.944	1227.9984	1554.936	3059.982	2.328	0.556	5.544
	Washington	10252.9800	14911.426	11169.024	1367.8400	2793.528	3149.930	7.400	3.592	27.600
Home Office	California	28193.4690	18651.064	32311.470	2676.6720	1856.190	4158.912	4.950	1.810	0.990
	Georgia	2078.1800	5015.450	1312.530	723.9200	1270.990	484.830	7.040	9.450	71.980
	Texas	6747.6164	4824.916	8281.966	981.3720	727.296	1718.400	5.344	1.408	7.992
	Washington	12448.1520	3913.502	8645.246	2348.8200	359.320	2395.200	4.180	2.496	7.992

Insights:

Segment & State-wise indexing,
Over Categories-wise SUM, MAX, MIN sales report.

Insight will be used to analysis data through diffferent perspective

In [17]:
```
# Sub-Category level

report_e = pd.pivot_table(Data,
                          index=["Category","SubCategory"],
                          values=["Sales_in_M","Qty"],
                          aggfunc=[np.sum]
                         )
report_e
```

Out[17]:

		sum	
		Qty	Sales_in_M
Category	SubCategory		
Furniture	Bookcases	2270344	113813.1987
	Chairs	6174348	322822.7310
	Furnishings	9336425	89212.0180
	Tables	3034074	202810.6280
Office Supplies	Appliances	4514831	104618.4030
	Art	7848858	26705.4100
	Binders	14633213	200028.7850
	Envelopes	2359969	16128.0460
	Fasteners	2223968	3001.9600
	Labels	3350727	12347.7260
	Paper	13535206	76828.3040
	Storage	8109223	219343.3920
	Supplies	1862324	46420.3080
Technology	Accessories	7067251	164186.7000
	Copiers	697285	146248.0940
	Machines	1207632	189238.6310
	Phones	8996414	327782.4480

In [18]:
```
report_e = report_e.reset_index()
report_e.columns=["Category","SubCategory","TotalQuantity","TotalSales"]
fig = px.bar(report_e,
             x="Category", y="TotalQuantity",
             color="SubCategory",
             title="Category"
            )
fig.show()
```

Interpretation for Category and sub category wise annual report:

Furnishings from Furniture & Binders, Paper from Office Supplies sold higher by product units.
Accessories and Phones from Technology category shows higher in product sales.

So considering future earth and eco-friendly environment,
we promote valueable #GoGreen methods among every dealers and customers.
Tech gaint Apple Inc. announced their revised policy.

In [19]:
```
# Customer geographics analysis

report_f = pd.pivot_table(Data,
                          index=["State","Code"],
                          values=["CustomerName"],
                          aggfunc=[np.count_nonzero]
                         )
report_f
```

Out[19]:

		count_nonzero
		CustomerName
State	Code	
Alabama	AL	61
Arizona	AZ	223
Arkansas	AR	60
California	CA	1946
Colorado	CO	179
Connecticut	CT	82
Delaware	DE	93
District of Columbia	WA	10
Florida	FL	373
Georgia	GA	177

In [20]:
```
report_f = report_f.reset_index()
report_f.columns=["State","Code","TotalCustomers"]
report_f.sort_values(by=['TotalCustomers'], inplace=True, ascending=False)
df = report_f
report_f.head(7)
```

Out[20]:

	State	Code	TotalCustomers
3	California	CA	1946
30	New York	NY	1097
41	Texas	TX	973
36	Pennsylvania	PA	582
45	Washington	WA	504
11	Illinois	IL	483
33	Ohio	OH	454

In [21]:
```
fig = go.Figure(data=go.Choropleth(
locations=df['Code'],                    # Spatial coordinates
z = df['TotalCustomers'].astype(float), # Data to be color-coded
locationmode = 'USA-states',             # set of locations match entries in `locations`
colorscale = 'sunset',
text=df['State'] ,                       # hover text
colorbar_title = "No. of Customer's",
))
fig.update_layout(
    title_text = 'Total number of Customers by State',
    geo_scope='usa', # limite map scope to USA
)
fig.show()
```

In [ ]:
```
```

In [22]:
```
# Geographical based Quantity analysis

report_g = pd.pivot_table(Data,
                          index=["State","Code"],
                          values=["Qty"],
                          aggfunc=[np.sum]
                         )
report_g
```

Out[22]:

		sum
		Qty
State	Code	
Alabama	AL	663260
Arizona	AZ	2201305
Arkansas	AR	538459
California	CA	19153871
Colorado	CO	1756075
Connecticut	CT	796323
Delaware	DE	878086
District of Columbia	WA	96635
Florida	FL	3770161
Georgia	GA	1837449

In [23]:
```
report_g = report_g.reset_index()
report_g.columns=["State","Code","Qty"]
report_g.sort_values(by=['Qty'], inplace=True, ascending=False)
df = report_g
report_g
```

Out[23]:

	State	Code	Qty
3	California	CA	19153871
30	New York	NY	10856452
41	Texas	TX	9514200
36	Pennsylvania	PA	5873237
11	Illinois	IL	4916954
45	Washington	WA	4860676
33	Ohio	OH	4532872
8	Florida	FL	3770161
20	Michigan	MI	2505570
31	North Carolina	NC	2484240
1	Arizona	AZ	2201305
44	Virginia	VA	2055080

In [24]:
```
fig = go.Figure(data=go.Choropleth(
    locations=df["Code"],           # Spatial coordinates
    z = df['Qty'].astype(float),    # Data to be color-coded
    locationmode = 'USA-states',    # set of locations match entries in `locations`
    colorscale = 'ylorrd',
    #text=df['loc'] ,                # hover text
    colorbar_title = "Product-wise Quantity in Millions",
))
fig.update_layout(
    title_text = 'Total number of products sold by State',
    geo_scope='usa',                # limite map scope to USA
)
fig.show()
```

In [25]:
```
# Geographical based Sales analysis

report_h = pd.pivot_table(Data,
                          index=["State","Code"],
                          values=["Sales_in_M"],
                          aggfunc=[np.sum]
                         )
report_h.head(10)
```

Out[25]:

		sum
		Sales_in_M
State	Code	
Alabama	AL	19510.6400
Arizona	AZ	35272.6570
Arkansas	AR	11678.1300
California	CA	446306.4635
Colorado	CO	31841.5980
Connecticut	CT	13384.3570
Delaware	DE	27322.9990
District of Columbia	WA	2865.0200
Florida	FL	88436.5320
Georgia	GA	48219.1100

In [26]:
```
report_h = report_h.reset_index()
report_h.columns=["State","Code","Sales_in_M"]
report_h.sort_values(by=["Sales_in_M"], inplace=True, ascending=False)
df = report_h
report_h.head()
```

Out[26]:

	State	Code	Sales_in_M
3	California	CA	446306.4635
30	New York	NY	306361.1470
41	Texas	TX	168572.5322
45	Washington	WA	135206.8500
36	Pennsylvania	PA	116276.6500

In [27]:
```
fig = go.Figure(data=go.Choropleth(
    locations=df["Code"],           # Spatial coordinates
    z = df['Sales_in_M'].astype(float),    # Data to be color-coded
    locationmode = 'USA-states',    # set of locations match entries in `locations`
    colorscale = 'plotly3',
    #text=df['loc'] ,                # hover text
    colorbar_title = "Product-wise Quantity in Millions",
))
fig.update_layout(
    title_text = 'Sales by State',
    geo_scope='usa',                # limite map scope to USA
)
fig.show()
```

Interpretation on Location (State & sales)

```
Total number of Customers by State
           Highest: California(CA)-1946
           Lowest : Wyoming(WY)-1

Total number of products sold by State
           Highest: California(CA)-19153871
           Lowest: Wyoming(WY)-16608

Sales by State
           Highest sales:California    (CA)-446306.4635 M
           Lowest sales: North Dakota (ND)-919.9100 M
```

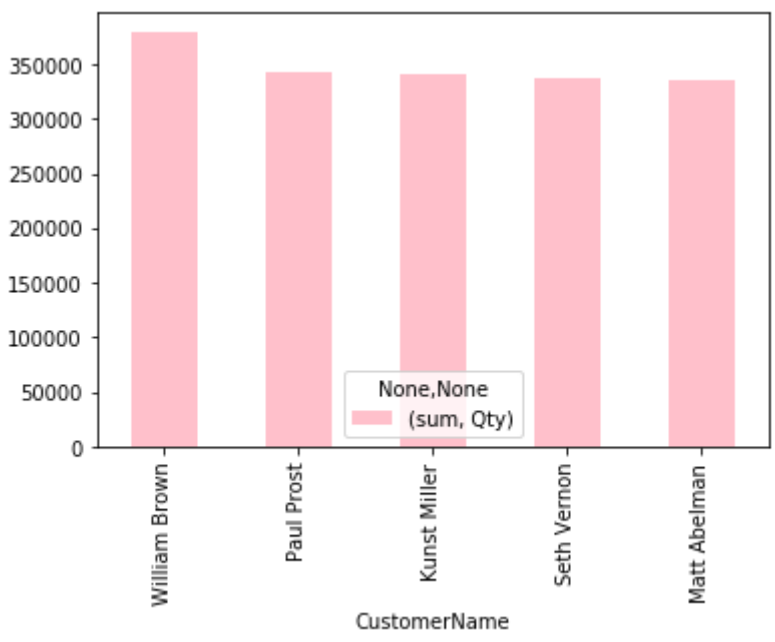
In [28]:
```
report_i = pd.pivot_table(Data,
                          index=["CustomerName"],
                          values=["Qty"],
                          aggfunc=[np.sum]
                         )
report_i.head(10)
```

Out[28]:

	sum
	Qty
CustomerName	
Aaron Bergman	40249
Aaron Hawkins	96381
Aaron Smayling	106609
Adam Bellavance	243182
Adam Hart	207372
Adam Shillingsburg	277108
Adrian Barton	219534
Adrian Hane	152523
Adrian Shami	39160
Aimee Bixby	96193

In [29]:
```
pd.pivot_table(Data,index=['CustomerName'],values=['Qty'],aggfunc=[np.sum]).sort_values([('sum','Qty')],ascending=False)
```

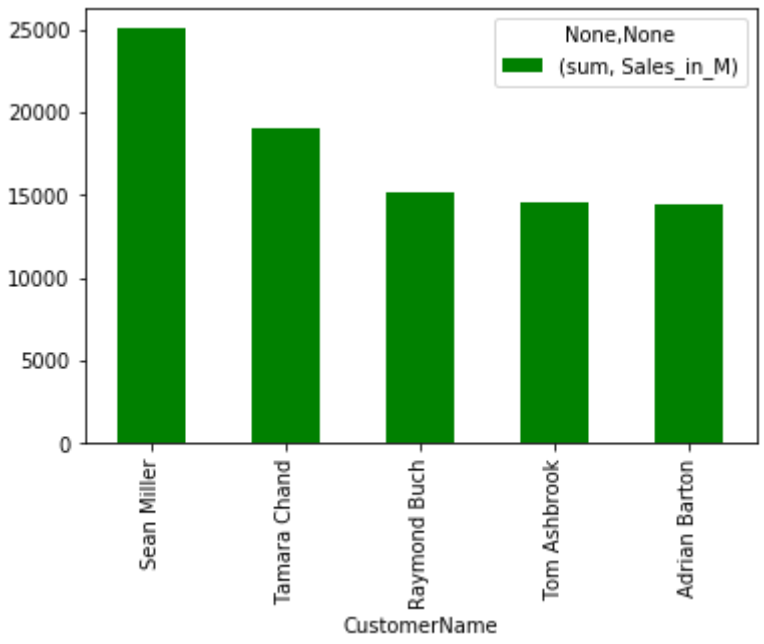
Out[29]: <matplotlib.axes._subplots.AxesSubplot at 0x1453f37ef88>



In [30]:
```
### Sales_in_M
```

In [31]:
```
pd.pivot_table(Data,index=['CustomerName'],values=['Sales_in_M'],aggfunc=[np.sum]).sort_values([('sum','Sales_in_M')],as
```

Out[31]: <matplotlib.axes._subplots.AxesSubplot at 0x1453f3658c8>



Interpretation on Customer

Hightest amount of purchase of products was done by William Brown.
More number of products were purchased by Sean Miller

In [ ]:

REPORT FOR FISCAL YEAR 2019-20 FINANCIAL ANALYTICS

Only for the USA

There are 4 main categories to analyse sales of product :

- segments
- categories of product
- location(state)
- customer

Insights based on Segment:

```
Highest Turnover    - Home Office
Average Turnover    - Corporate
Lowest  Turnover    - Consumer

Company can focus more on Consumer segment for increasing turnover.
```

Insights based on Categories of product:

```
*Furnishings from Furniture & Binders, Paper from Office Supplies sold higher by product units.
Accessories and Phones from Technology category shows higher in product sales.

So considering future earth and eco-friendly environment,
we promote valueable #GoGreen methods among every dealers and customers.
Tech gaint Apple Inc. announced their revised policy.

*Technology Category under every segments proving higher revenue in sales.
Where,
Office Supplies left idle and Furniture sales remains consistance.

Individually Furniture, Office Supplies and technology are idle for consumer and corporate segment.
But from corporate to home office segment,
there is falling down for furniture and increase for technology while furniture stays constant.

*It's clearly shoes that's Quantity wise Technology based products sold very less comparatively.
But Technology was been expensive then other category.
```

Insights based on location (state wise):

```
Sales by State
      Highest sales:California    (CA)-446306.4635 M
      Lowest sales: North Dakota (ND)-919.9100 M

Total number of products sold by State
      Highest: California(CA)-19153871
      Lowest: Wyoming(WY)-16608

Total number of Customers by State
      Highest: California(CA)-1946
      Lowest : Wyoming(WY)-1
```

Insights based on customer:

```
Sales in Million : William Brown
Qty : Sean Miller
```

Decisions that can be taken

In the **Category** the company should focus on increasing the technological equipments ordered quantity by more advertising and creating a demand in the market and also the focus can be on the increasing of sales in terms of money of office supplies by increasing the profit margin so that the company can grow as whole

In the **State** company should focus on Wyoming to sale more in terms of quantity and the company should sale more and more in the state like North Dakota

In the **City** company should focus on Cedar Rapids to sale more in terms of quantity and the company should sale more and more in the City like Abilene

In the **Segment** company should focus on Home Office products to sale more in terms of quantity and the company should sale more and more in the segment like Home Office only

---

Color set for graph:

'aggrnyl', 'agsunset', 'algae', 'amp', 'armyrose', 'balance', 'blackbody', 'bluered', 'blues', 'blugrn', 'bluyl', 'brbg', 'brwnyl', 'bugn', 'bupu', 'burg', 'burgyl', 'cividis', 'curl', 'darkmint', 'deep', 'delta', 'dense', 'earth', 'edge', 'electric', 'emrld', 'fall', 'geyser', 'gnbu', 'gray', 'greens', 'greys', 'haline', 'hot', 'hsv', 'ice', 'icefire', 'inferno', 'jet', 'magenta', 'magma', 'matter', 'mint', 'mrybm', 'mygbm', 'oranges', 'orrd', 'oryel', 'peach', 'phase', 'picnic', 'pinkyl', 'piyg', 'plasma', 'plotly3', 'portland', 'prgn', 'pubu', 'pubugn', 'puor', 'purd', 'purp', 'purples', 'purpor', 'rainbow', 'rdbu', 'rdgy', 'rdpu', 'rdylbu', 'rdylgn', 'redor', 'reds', 'solar', 'spectral', 'speed', 'sunset', 'sunsetdark', 'teal', 'tealgrn', 'tealrose', 'tempo', 'temps', 'thermal', 'tropic', 'turbid', 'twilight', 'viridis', 'ylgn', 'ylgnbu', 'ylorbr', 'ylorrd'