jupyter Untitled11 Last Checkpoint: 40 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[5]:
```
startups = pd.read_excel('Downloads/startup-expansion.xlsx')
startups
```

[5]:

	Store ID	City	State	Sales Region	New Expansion	Marketing Spend	Revenue
0	1	Peoria	Arizona	Region 2	Old	2601	48610
1	2	Midland	Texas	Region 2	Old	2727	45689
2	3	Spokane	Washington	Region 2	Old	2768	49554
3	4	Denton	Texas	Region 2	Old	2759	38284
4	5	Overland Park	Kansas	Region 2	Old	2869	59887
...	...	...	...	...	...	...	...
145	146	Paterson	New Jersey	Region 1	New	2251	34603
146	147	Brownsville	Texas	Region 2	New	3675	63148

150 rows × 7 columns

[9]:
```
startups.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   Store ID         150 non-null    int64 
 1   City             150 non-null    object
 2   State            150 non-null    object
 3   Sales Region     150 non-null    object
 4   New Expansion    150 non-null    object
 5   Marketing Spend  150 non-null    int64 
 6   Revenue          150 non-null    int64 
dtypes: int64(3), object(4)
memory usage: 8.3+ KB
```

[13]:
```
startups[['Marketing Spend','Revenue']].describe().round(2)
```

[13]:

	Marketing Spend	Revenue
min	1811.00	15362.00
25%	2662.25	21113.50
50%	2898.00	42993.00
75%	3111.50	51145.50
max	3984.00	68828.00

[17]:
```
startups['City'].nunique()
```

[17]:
```
149
```

[21]:
```
startups['City'].unique()
```

[21]:
```
array(['Peoria', 'Midland', 'Spokane', 'Denton', 'Overland Park',
       'Yonkers', 'Birmingham', 'Antioch', 'Worcester', 'Rochester',
       'Rialto', 'Santa Maria', 'Las Cruces', 'Jackson', 'Hillsboro',
       'Temecula', 'Tallahassee', 'Fontana', 'Kent', 'Broken Arrow',
       'Concord', 'Modesto', 'Montgomery', 'Burbank', 'Elk Grove',
       'Port St. Lucie', 'Elizabeth', 'Salt Lake City', 'Waco', 'Edison',
       'Boulder', 'Grand Rapids', 'Tyler', 'Charleston', 'Huntsville',
       'Pearland', 'Inglewood', 'Oxnard', 'Miramar', 'Cape Coral',
       'Lakeland', 'Centennial', 'Lowell', 'Ontario', 'Palm Bay',
       'Murfreesboro', 'Vancouver', 'Topeka', 'West Valley City',
       'New Haven', 'Pueblo', 'Costa Mesa', 'Garden Grove',
       'Fort Lauderdale', 'North Charleston', 'Cambridge', 'Greeley',
       'Gresham', 'Amarillo', 'High Point', 'Vista', 'Tacoma', 'Mesquite',
       'Augusta', 'Elgin', 'Aurora', 'Gainesville', 'Dayton',
       'Wichita Falls', 'Naperville', 'Clovis', 'Billings', 'Surprise',
       'Coral Springs', 'Visalia', 'Killeen', 'Orange', 'Richardson',
       'South Bend', 'Fayetteville', 'Sioux Falls', 'Grand Prairie',
       'Stamford', 'West Palm Beach', 'Knoxville', 'Renton', 'McAllen',
       'Woodbridge', 'Shreveport', 'Bellevue', 'Huntington Beach',
       'Santa Clarita', 'Sterling Heights', 'Mobile', 'Bridgeport',
       'Daly City', 'Sandy Springs', 'Cedar Rapids', 'Columbus',
       'Moreno Valley', 'Pompano Beach', 'Savannah', 'West Jordan',
```

```
       'Des Moines', 'Green Bay', 'Santa Rosa', 'San Mateo', 'Warren',
       'Norwalk', 'Lafayette', 'Providence', 'Chattanooga', 'Tempe',
       'Joliet', 'Rancho Cucamonga', 'Glendale', 'Paterson',
       'Brownsville', 'Rockford', 'College Station', 'Thousand Oaks'],
      dtype=object)
```

[25]:
```
startups['City'].value_counts()
```

[25]:
```
City
Arron           1
Fullerton       1
Manchester      1
Everett         1
Thousand Oaks   1
Name: count, Length: 149, dtype: int64
```

[29]:
```
startups['State'].unique()
```

[29]:
```
array(['Arizona', 'Texas', 'Washington', 'Kansas', 'New York', 'Alabama',
       'California', 'Massachusetts', 'New Mexico', 'Mississippi',
       'Oregon', 'Florida', 'Oklahoma', 'New Jersey', 'Utah', 'Colorado',
       'Michigan', 'South Carolina', 'Virginia', 'Ohio', 'New Hampshire',
       'Connecticut', 'Iowa', 'Arkansas', 'Tennessee', 'North Carolina',
       'Georgia', 'Illinois', 'Montana', 'Indiana', 'South Dakota',
       'Louisiana', 'Minnesota', 'Wisconsin', 'Rhode Island'],
      dtype=object)
```

[37]:
```
startups['State'].value_counts()
```

[37]:
```
State
California        40
Texas             17
Connecticut        4
New Jersey         4
Arizona            3
Tennessee          3
Iowa               3
Michigan           3
South Carolina     3
Utah               3
Massachusetts      3
Kansas             3
New York           3
Louisiana          2
North Carolina     2
Ohio               2
Virginia           2
Oregon             2
Mississippi        1
New Mexico         1
Arkansas           1
New Hampshire      1
Oklahoma           1
Montana            1
Indiana            1
South Dakota       1
```

[41]:
```
array(['Region 2', 'Region 1'], dtype=object)
```

[45]:
```
startups['Sales Region'].value_counts()
```

[45]:
```
Sales Region
Region 2    86
Region 1    64
Name: count, dtype: int64
```

[49]:
```
startups['New Expansion'].unique()
```

[49]:
```
array(['Old', 'New'], dtype=object)
```

[53]:
```
startups['New Expansion'].value_counts()
```

[53]:
```
New Expansion
Old    140
New     10
Name: count, dtype: int64
```

[57]:
```
startups.isna().sum()
```

```
Revenue           0
dtype: int64
```

[61]:
```
startups.duplicated().sum()
```

[61]:
```
0
```

[65]:
```
startups.shape
```

[65]:
```
(150, 7)
```

[69]:
```
len(startups['Store ID'].value_counts())
```

[69]:
```
150
```

[73]:
```
startups.head(10)
```

[73]:

	Store ID	City	State	Sales Region	New Expansion	Marketing Spend	Revenue
0	1	Peoria	Arizona	Region 2	Old	2601	48610
1	2	Midland	Texas	Region 2	Old	2727	45689
2	3	Spokane	Washington	Region 2	Old	2768	49554
6	7	Birmingham	Alabama	Region 1	Old	3110	60338
7	8	Antioch	California	Region 2	Old	2593	19569
8	9	Worcester	Massachusetts	Region 1	Old	2675	59840
9	10	Rochester	New York	Region 1	Old	2984	64906

[77]:
```
startups.sample(10)
```

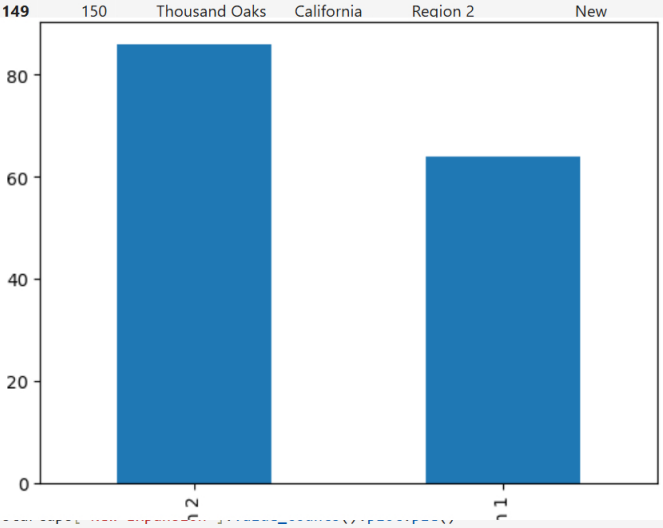
[77]:

	Store ID	City	State	Sales Region	New Expansion	Marketing Spend	Revenue
76	77	West Valley City	Utah	Region 2	Old	2555	49609
40	41	Syracuse	New York	Region 1	Old	2373	41313
74	75	Vancouver	Washington	Region 2	Old	3329	58951
26	27	Elizabeth	New Jersey	Region 1	Old	2557	45017
18	19	Kent	Washington	Region 2	Old	2894	43183
87	88	High Point	North Carolina	Region 1	Old	3262	19752
104	105	Orange	California	Region 2	Old	2830	36821

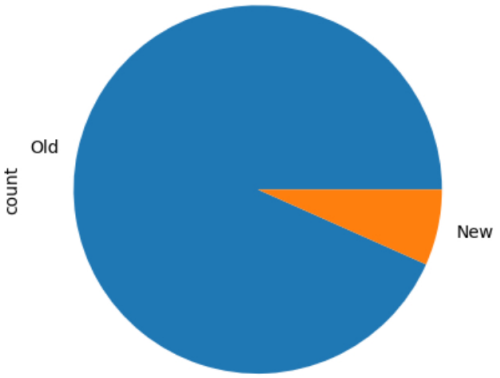
[81]:
```
startups.tail(10)
```

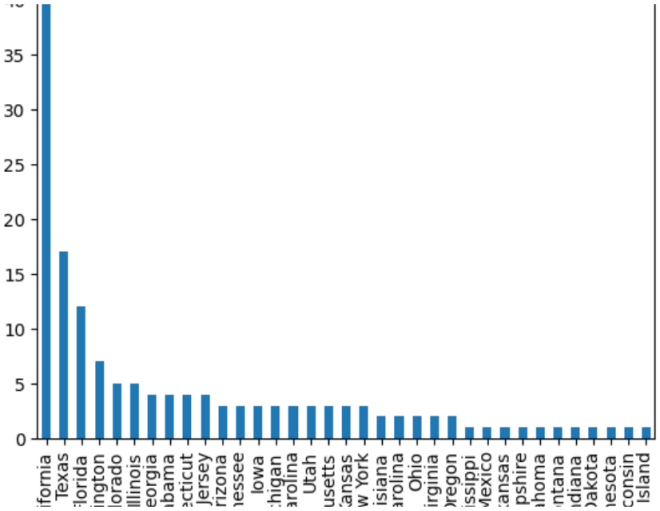
[81]:

	Store ID	City	State	Sales Region	New Expansion	Marketing Spend	Revenue
140	141	Chattanooga	Tennessee	Region 2	New	3587	55357
141	142	Tempe	Arizona	Region 2	New	2911	48954
142	143	Joliet	Illinois	Region 1	New	3279	48315
143	144	Rancho Cucamonga	California	Region 2	New	2945	52366
144	145	Glendale	California	Region 2	New	2363	49376
145	146	Paterson	New Jersey	Region 1	New	2251	34603
146	147	Brownsville	Texas	Region 2	New	3675	63148
147	148	Rockford	Illinois	Region 1	New	2648	43377
148	149	College Station	Texas	Region 2	New	2994	22457
149	150	Thousand Oaks	California	Region 2	New	2431	40141



[89]:
```
<Axes: ylabel='count'>
```





[97]:
```
startups.groupby('New Expansion')
```

[97]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x00000244538EED20>
```

[101]:
```
startups[startups['New Expansion'] == 'Old'].groupby(['State']).sum()['Revenue'].nlargest(10)
```

[101]:
```
State
California     1362468
Texas           554964
Florida         479023
Washington      298013
Alabama         221025
New York        160046
Connecticut     158511
Georgia         157656
Colorado        156495
Michigan        147759
Name: Revenue, dtype: int64
```

[105]:
```
startups[startups['New Expansion'] == 'New'].groupby(['State']).sum()['Revenue'].nlargest(10)
```

[105]:
```
State
California    141883
```

[109]:
```
startups['Profit'] = startups['Revenue'] - startups['Marketing Spend']
```

[113]:
```
startups
```

[113]:

	Store ID	City	State	Sales Region	New Expansion	Marketing Spend	Revenue	Profit
0	1	Peoria	Arizona	Region 2	Old	2601	48610	46009
1	2	Midland	Texas	Region 2	Old	2727	45689	42962
2	3	Spokane	Washington	Region 2	Old	2768	49554	46786
3	4	Denton	Texas	Region 2	Old	2759	38284	35525

4	5	Overland Park	Kansas	Region 2	Old	2869	59887	57018
...	...	...	...	...	...	...	...	...
145	146	Paterson	New Jersey	Region 1	New	2251	34603	32352
146	147	Brownsville	Texas	Region 2	New	3675	63148	59473
147	148	Rockford	Illinois	Region 1	New	2648	43377	40729

[117]:
```
startups['ROMS'] = round((startups['Profit'] / startups['Marketing Spend']) * 100, 2)
```

[121]:
```
startups['ROMS%'] = startups['ROMS'] / 100
```

[123]:
```
startups
```

[123]:

	Store ID	City	State	Sales Region	New Expansion	Marketing Spend	Revenue	Profit	ROMS	ROMS%
0	1	Peoria	Arizona	Region 2	Old	2601	48610	46009	1768.90	17.6890
3	4	Denton	Texas	Region 2	Old	2759	38284	35525	1287.60	12.8760
4	5	Overland Park	Kansas	Region 2	Old	2869	59887	57018	1987.38	19.8738
...	...	...	...	...	...	...	...	...	...	...
145	146	Paterson	New Jersey	Region 1	New	2251	34603	32352	1437.23	14.3723
146	147	Brownsville	Texas	Region 2	New	3675	63148	59473	1618.31	16.1831
147	148	Rockford	Illinois	Region 1	New	2648	43377	40729	1538.10	15.3810
148	149	College Station	Texas	Region 2	New	2994	22457	19463	650.07	6.5007
149	150	Thousand Oaks	California	Region 2	New	2431	40141	37710	1551.21	15.5121

150 rows × 10 columns

[131]:
```
startups.to_csv('startups-expansion-Modified.csv')
```

[ ]:
```
```