Project 1

November 5, 2023

[ ]:
```
#conda install psycopg2
```

TOPICS: 1. Predict the future sales volume and price of brands in different locations(states), in order to provide stocking and pricing strategies. 2. Seasonality of the store sales in different state saledate, supply chain management, in different type of product 3. Return and purchase rate of different products 4. Inflation trends through years in stores by different products

[1]:
```
import psycopg2
import pandas as pd

# Connection parameters
host = "pg.analytics.northwestern.edu"
port = "5432"
database = "everything2023"
user = ""
password = ""

# Establish a connection to the database
conn = psycopg2.connect(
    host=host,
    port=port,
    database=database,
    user=user,
    password=password
)
```

[2]:
```
cursor = conn.cursor()
sql_query = "SELECT * FROM group_13.deptinfo;"
cursor.execute(sql_query)
deptinfo = pd.read_sql_query(sql_query, conn)

cursor = conn.cursor()
sql_query2 = "SELECT * FROM group_13.trnsact LIMIT 1000000;"
cursor.execute(sql_query2)
trnsact = pd.read_sql_query(sql_query2, conn)

cursor = conn.cursor()
sql_query3 = "SELECT * FROM group_13.skstinfo;"
```

```
cursor.execute(sql_query3)
skstinfo = pd.read_sql_query(sql_query3, conn)
skstinfo.head()

cursor = conn.cursor()
sql_query4 = "SELECT * FROM group_13.strinfo;"
cursor.execute(sql_query4)
strinfo = pd.read_sql_query(sql_query4, conn)

cursor = conn.cursor()
sql_query5 = "SELECT * FROM group_13.skuinfo2;"
cursor.execute(sql_query5)
skuinfo = pd.read_sql_query(sql_query5, conn)

conn.close()
```

[3]:
```
skstinfo
```

[3]:

	SKU	STORE	COST	RETAIL	unknown
0	2560675	5603	10.5	29.0	0
1	2560675	5604	10.5	29.0	0
2	2560675	5704	10.5	29.0	0
3	2560675	5802	10.5	29.0	0
4	2560675	5803	10.5	29.0	0
…	…	…	…	…	…
39230141	2560675	5404	10.5	29.0	0
39230142	2560675	5502	10.5	29.0	0
39230143	2560675	5503	10.5	29.0	0
39230144	2560675	5504	10.5	29.0	0
39230145	2560675	5602	10.5	29.0	0

[39230146 rows x 5 columns]

[4]:
```
skuinfo.head()
```

[4]:

	SKU	DEPT	CLASSID	UPC	STYLE	COLOR	SIZE	\
0	0	1	2	3	4	5	6	
1	3	6505	113	400000003000	00	F55KT2	WHISPERWHITE	P8EA
2	4	8101	002	400000004000	22	615CZ4	SPEARMI	S
3	5	7307	003	400000005000	7LBS	245-01	34 SILVER	KING
4	8	3404	00B	400000008000	622	F05H84	MORNING MI	2T

	PACKSIZE	VENDOR	BRAND
0	7	8	9
1	1	5119207	TURNBURY
2	1	3311144	C A SPOR
3	1	5510554	BEAU IDE

```
4         1  2912827  HARTSTRI
```

[5]:
```
deptinfo.head()
```

[5]:
```
   DEPT  DEPTDESC  Unknow
0   800  CLINIQUE       0
1   801  LESLIE         0
2  1100  GARY F         0
3  1107  JACQUES        0
4  1202  CABERN         0
```

[6]:
```
trnsact
```

[6]:
```
             SKU  STORE  REGISTER  TRANNUM        SEQ    SALEDATE STYPE  \
0        1383360   5403       520     1800          0  2005-06-03     P
1        1383360   5403       520     2000          0  2005-06-03     P
2        1383360   5403       520     2000          0  2005-06-03     R
3        1383360   5503        50     1700  746705895  2005-03-12     P
4        1383360   5503        50     2100  295808931  2005-04-06     P
…            …    …         …       …           …          …   …
999995   1483552   7407       910     3100          0  2005-03-13     P
999996   1483552   7407       910     3200          0  2005-05-02     P
999997   1483552   7502       720     1800          0  2004-08-24     P
999998   1483552   7502       750     4100  440609046  2005-01-24     P
999999   1483552   7502       770     1400          0  2005-07-12     P

         QUANTITY  ORGPRICE  SPRICE   AMT    INTERID  MIC  Unknow
0               1      10.0    10.0  10.0  797500015  107       0
1               1      10.0    10.0  10.0  797400015  107       0
2               1      10.0    10.0  10.0  797600015  107       0
3               1      10.0    10.0  10.0  773500027  107       0
4               1      10.0    10.0  10.0  138100014  107       0
…           …        …     …    …         …  …     …
999995          1      16.0    12.0  12.0  528500012   15       0
999996          1      16.0    12.0  12.0  256900012   15       0
999997          1      16.0    16.0  16.0  651400049   15       0
999998          1      16.0     8.0   8.0  751400012   15       0
999999          1      16.0    16.0  16.0  724200013   15       0

[1000000 rows x 14 columns]
```

[7]:
```
strinfo.head()
```

[7]:
```
   store                 city state    zip  x
0      2  ST. PETERSBURG          FL  33710  0
1      3  ST. LOUIS               MO  63126  0
2      4  LITTLE ROCK             AR  72201  0
```

```
3      7  FORT WORTH              TX  76137  0
4      9  TEMPE                   AZ  85281  0
```

0.1 Clean Data

Drop the last column:

[8]:
```
# Drop unknow column (the last column):
deptinfo.drop(columns=["Unknow"],inplace=True)
deptinfo.head()
```

[8]:
```
   DEPT  DEPTDESC
0   800  CLINIQUE
1   801  LESLIE
2  1100  GARY F
3  1107  JACQUES
4  1202  CABERN
```

[9]:
```
# Drop the last unknown column:
trnsact.drop(columns=["Unknow"],inplace=True)
trnsact
```

[9]:
```
            SKU  STORE  REGISTER  TRANNUM        SEQ    SALEDATE STYPE  \
0       1383360   5403       520     1800          0  2005-06-03     P
1       1383360   5403       520     2000          0  2005-06-03     P
2       1383360   5403       520     2000          0  2005-06-03     R
3       1383360   5503        50     1700  746705895  2005-03-12     P
4       1383360   5503        50     2100  295808931  2005-04-06     P
…           …    …          …        …          …          …    …
999995  1483552   7407       910     3100          0  2005-03-13     P
999996  1483552   7407       910     3200          0  2005-05-02     P
999997  1483552   7502       720     1800          0  2004-08-24     P
999998  1483552   7502       750     4100  440609046  2005-01-24     P
999999  1483552   7502       770     1400          0  2005-07-12     P

        QUANTITY  ORGPRICE  SPRICE   AMT    INTERID  MIC
0              1      10.0    10.0  10.0  797500015  107
1              1      10.0    10.0  10.0  797400015  107
2              1      10.0    10.0  10.0  797600015  107
3              1      10.0    10.0  10.0  773500027  107
4              1      10.0    10.0  10.0  138100014  107
…            …        …       …     …          …  …
999995         1      16.0    12.0  12.0  528500012   15
999996         1      16.0    12.0  12.0  256900012   15
999997         1      16.0    16.0  16.0  651400049   15
999998         1      16.0     8.0   8.0  751400012   15
999999         1      16.0    16.0  16.0  724200013   15
```

```
[1000000 rows x 13 columns]
```

[10]:
```
# Drop the last unknown column:
skstinfo.drop(columns=["unknown"],inplace=True)
skstinfo
```

[10]:

	SKU	STORE	COST	RETAIL
0	2560675	5603	10.5	29.0
1	2560675	5604	10.5	29.0
2	2560675	5704	10.5	29.0
3	2560675	5802	10.5	29.0
4	2560675	5803	10.5	29.0
…	…	…	…	…
39230141	2560675	5404	10.5	29.0
39230142	2560675	5502	10.5	29.0
39230143	2560675	5503	10.5	29.0
39230144	2560675	5504	10.5	29.0
39230145	2560675	5602	10.5	29.0

```
[39230146 rows x 4 columns]
```

[11]:
```
# Drop the last unknown column:
strinfo.drop(columns=['x'],inplace=True)
strinfo
```

[11]:

	store	city	state	zip
0	2	ST. PETERSBURG	FL	33710
1	3	ST. LOUIS	MO	63126
2	4	LITTLE ROCK	AR	72201
3	7	FORT WORTH	TX	76137
4	9	TEMPE	AZ	85281
..	…	…	…	…
448	9808	GILBERT	AZ	85233
449	9812	METAIRIE	LA	70006
450	9900	LITTLE ROCK	AR	72201
451	9906	LITTLE ROCK	AR	72201
452	9909	CHEYENNE	WY	82009

```
[453 rows x 4 columns]
```

[12]:
```
strinfo.columns = ['STORE','CITY','STATE','ZIP']
strinfo
```

[12]:

	STORE	CITY	STATE	ZIP
0	2	ST. PETERSBURG	FL	33710
1	3	ST. LOUIS	MO	63126
2	4	LITTLE ROCK	AR	72201

```
3        7  FORT WORTH              TX  76137
4        9  TEMPE                   AZ  85281
..     ...                   ...   ...    ...
448   9808  GILBERT                 AZ  85233
449   9812  METAIRIE                LA  70006
450   9900  LITTLE ROCK             AR  72201
451   9906  LITTLE ROCK             AR  72201
452   9909  CHEYENNE                WY  82009

[453 rows x 4 columns]
```

Check Missing Value:

[13]:
```
strinfo.isna().sum()
```

[13]:
```
STORE    0
CITY     0
STATE    0
ZIP      0
dtype: int64
```

[14]:
```
deptinfo.isna().sum()
```

[14]:
```
DEPT        0
DEPTDESC    0
dtype: int64
```

[15]:
```
trnsact.isna().sum()
```

[15]:
```
SKU         0
STORE       0
REGISTER    0
TRANNUM     0
SEQ         0
SALEDATE    0
STYPE       0
QUANTITY    0
ORGPRICE    0
SPRICE      0
AMT         0
INTERID     0
MIC         0
dtype: int64
```

[16]:
```
skstinfo.isna().sum()
```

[16]:
```
SKU       0
STORE     0
```

```
COST      0
RETAIL    0
dtype: int64
```

[17]:
```
skuinfo.isna().sum()
```

[17]:
```
SKU         0
DEPT        0
CLASSID     0
UPC         0
STYLE       0
COLOR       0
SIZE        0
PACKSIZE    0
VENDOR      0
BRAND       0
dtype: int64
```

[18]:
```
skuinfo
```

[18]:

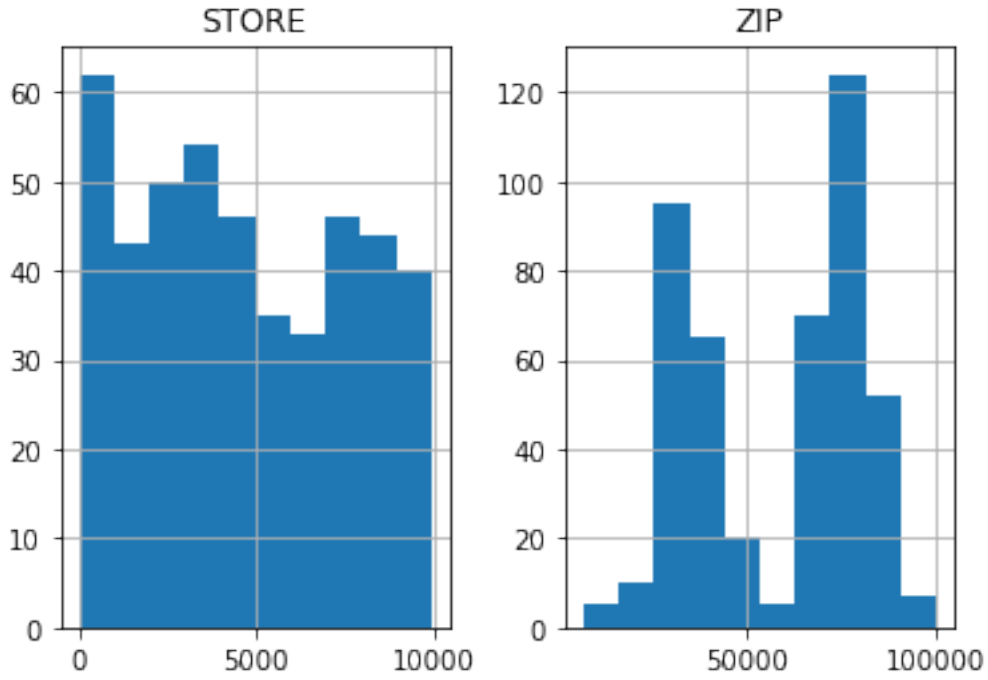
	SKU	DEPT	CLASSID	UPC	STYLE	COLOR	SIZE	PACKSIZE	VENDOR	BRAND
0	0	1	2	3	4	5	6	7	8	9
1	3	6505	113	400000003000	F55KT2	WHISPERWHITE	P8EA	1	5119207	TURNBURY
2	4	8101	002	400000004000	615CZ4	SPEARMI	S	1	3311144	C A SPOR
3	5	7307	003	400000005000	245-01	34 SILVER	KING	1	5510554	BEAU IDE
4	8	3404	00B	400000008000	F05H84	MORNING MI	2T	1	2912827	HARTSTRI
…	…	…	…	…	…	…	…	…	…	…
1556026	9999973	3103	009	400009973999	S3JAYV	STONE	4	1	6813115	POLO JEA
1556027	9999974	9801	726	400009974999	G50171	NAVY MULTI	10	1	9212766	GABAR IN
1556028	9999991	2301	004	400009991999	MDU201	618RED ROSE	8	1	23272	JONES/LA
1556029	9999992	1202	402	400009992999	F52UN1	PALE JADE	L	1	1446212	CABERNET
1556030	9999997	2503	111	400009997999	1XKBGO	210CHOPNK	7	1	8515392	RAMPAGE

[1556031 rows x 10 columns]

0.1.1 strinfo:

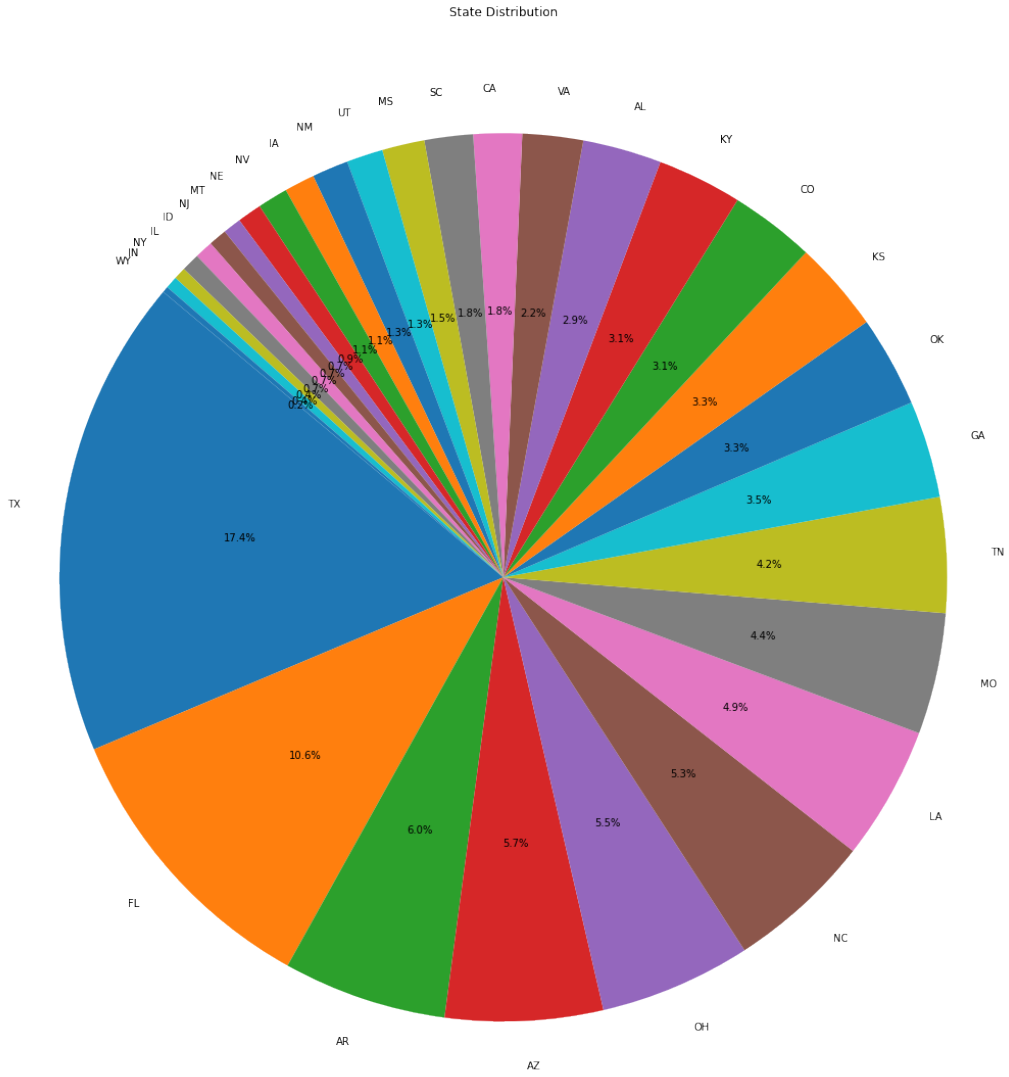
[19]:
```
strinfo.hist()
```

[19]:
```
array([[<AxesSubplot:title={'center':'STORE'}>,
        <AxesSubplot:title={'center':'ZIP'}>]], dtype=object)
```



[20]:
```
import matplotlib.pyplot as plt

state_counts = strinfo['STATE'].value_counts()
plt.figure(figsize=(20, 20))  # Optional: Set the figure size
plt.pie(state_counts, labels=state_counts.index, autopct='%1.1f%%',␣
 ↪startangle=140)
plt.title('State Distribution')
```

[20]:
```
Text(0.5, 1.0, 'State Distribution')
```



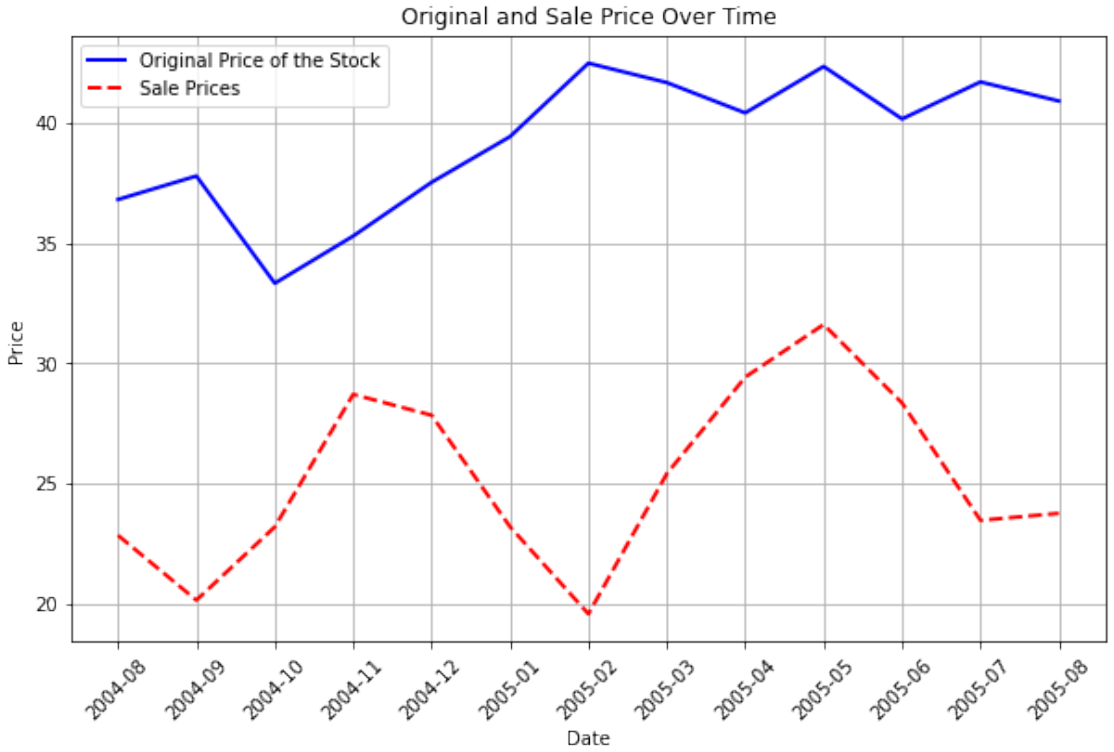
0.1.2 trnsact:

[21]:
```
trnsact.dtypes
```

[21]:
```
SKU           int64
STORE         int64
REGISTER      int64
TRANNUM       int64
SEQ           int64
SALEDATE     object
STYPE        object
QUANTITY      int64
```

```
ORGPRICE    float64
SPRICE      float64
AMT         float64
INTERID       int64
MIC           int64
dtype: object
```

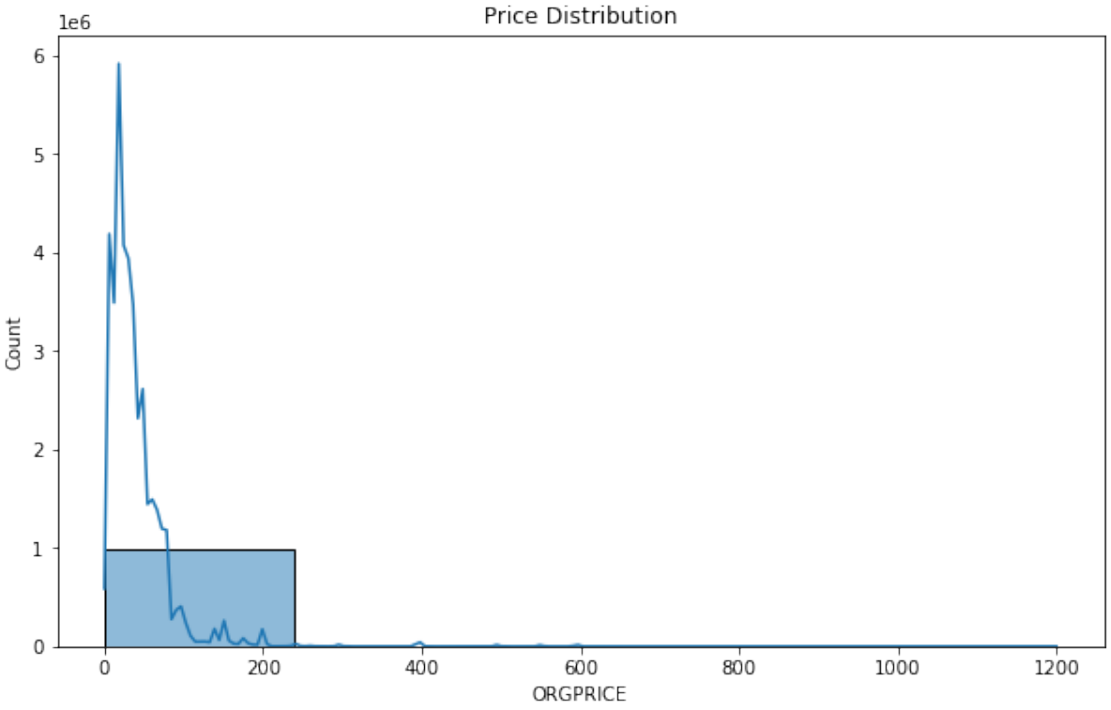
[22]:
```
# Assuming 'SALEDATE' is in a datetime format
trnsact['SALEDATE'] = pd.to_datetime(trnsact['SALEDATE'])

# Extract year and month from 'SALEDATE'
trnsact['Year'] = trnsact['SALEDATE'].dt.year
trnsact['Month'] = trnsact['SALEDATE'].dt.month

# Group by year and month and calculate the mean
trnsact_group_price = trnsact.groupby(['Year', 'Month']).mean()

plt.figure(figsize=(10, 6))  # Optional: Set the figure size

# Assuming 'Year' and 'Month' are now separate columns
date_labels = [f"{year}-{month:02d}" for year, month in zip(trnsact_group_price.
 ↪index.get_level_values('Year'), trnsact_group_price.index.
 ↪get_level_values('Month'))]

plt.plot(date_labels, trnsact_group_price['ORGPRICE'], label='Original Price of␣
 ↪the Stock', color='blue', linestyle='-', linewidth=2)
plt.plot(date_labels, trnsact_group_price['SPRICE'], label='Sale Prices',␣
 ↪color='red', linestyle='--', linewidth=2)

# Add labels and a legend
plt.xlabel('Date')
plt.ylabel('Price')
plt.title('Original and Sale Price Over Time')
plt.legend()

# Display the line chart
plt.grid(True)  # Optional: Display grid lines
plt.xticks(rotation=45)  # Optional: Rotate x-axis labels for better readability
plt.show()
```

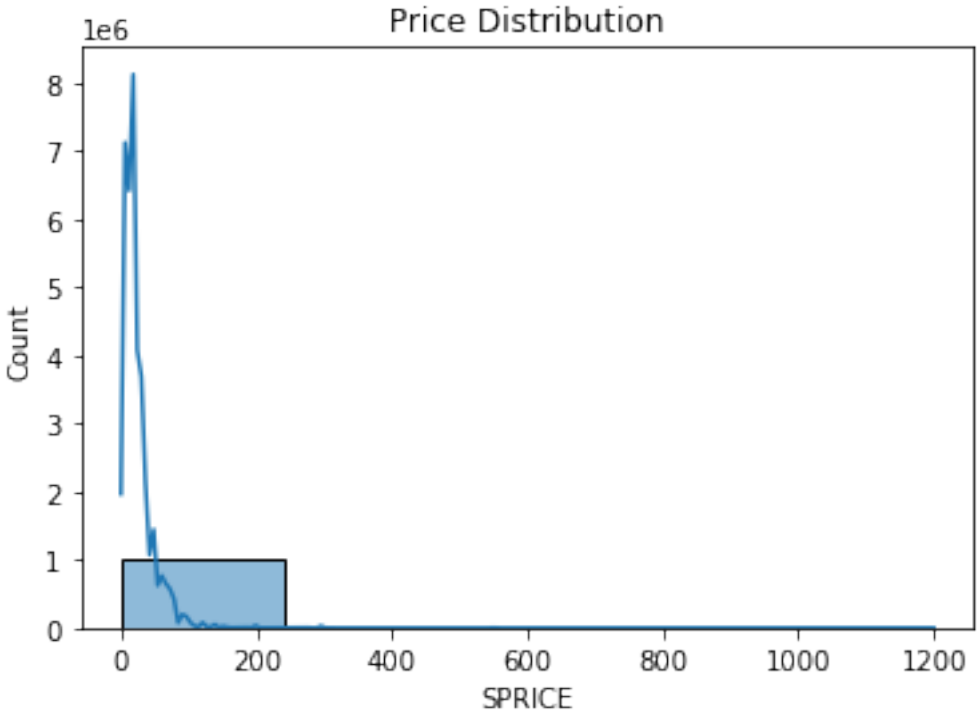


[23]:
```
trnsact.columns
```

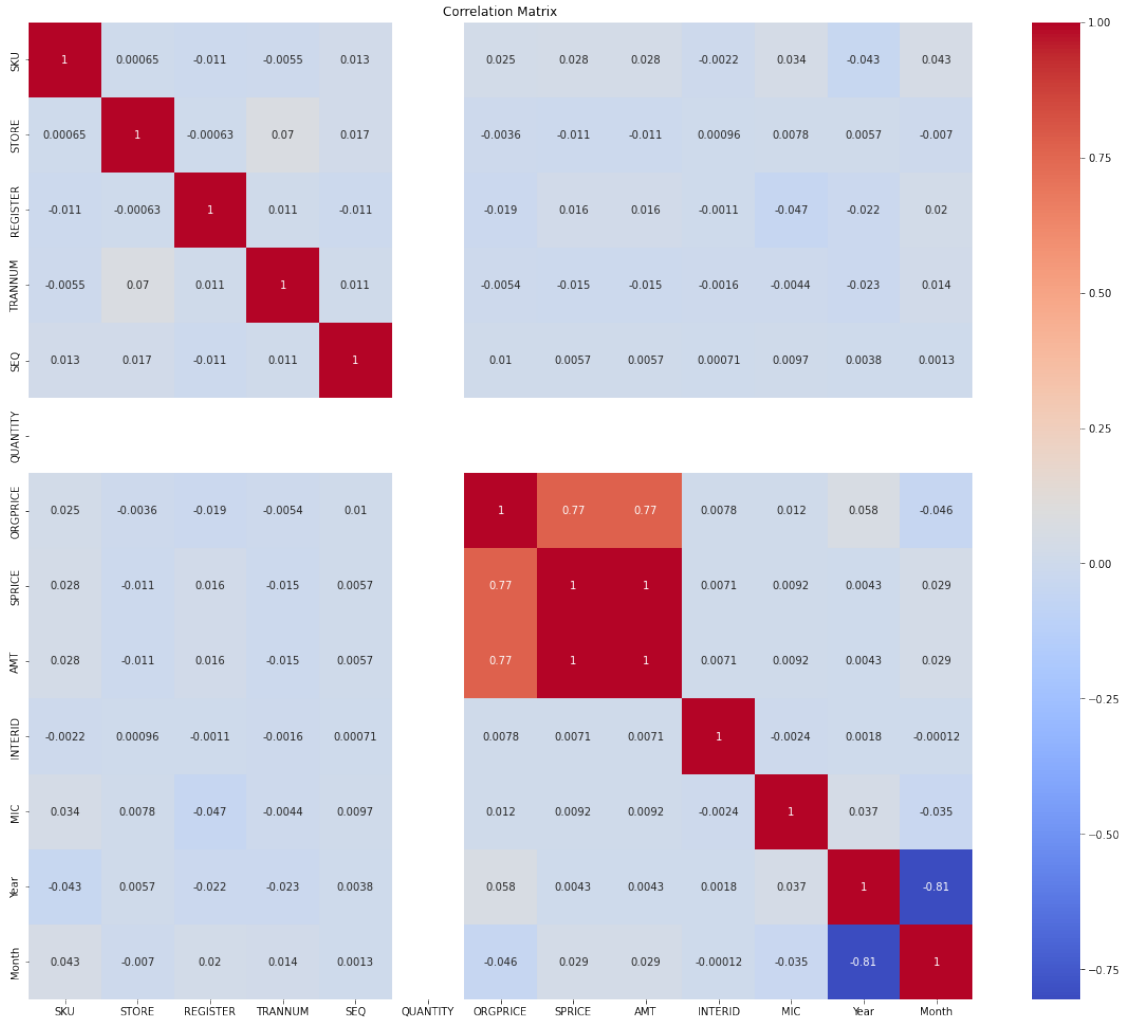
[23]:
```
Index(['SKU', 'STORE', 'REGISTER', 'TRANNUM', 'SEQ', 'SALEDATE', 'STYPE',
       'QUANTITY', 'ORGPRICE', 'SPRICE', 'AMT', 'INTERID', 'MIC', 'Year',
       'Month'],
      dtype='object')
```

[24]:
```
import seaborn as sns
# Data Distribution and Visualization
plt.figure(figsize=(10, 6))
sns.histplot(trnsact['ORGPRICE'], bins=5, kde=True)
plt.title('Price Distribution')
plt.show()
sns.histplot(trnsact['SPRICE'], bins=5, kde=True)
plt.title('Price Distribution')
plt.show()
```


Price Distribution
1e6
6
5
4
3
2
1
0
Count
0
200
400
600
800
1000
1200
ORGPRICE


Price Distribution
1e6
8
7
6
5
4
3
2
1
0
Count
0
200
400
600
800
1000
1200
SPRICE

[25]:
```
# Correlation Analysis
correlation_matrix = trnsact.corr()
plt.figure(figsize=(20, 17))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```



[ ]:

0.1.3 Merge Datasets Based on Topic 1:

[26]:
```
skuinfo['SKU'] = skuinfo['SKU'].astype(int)
skuinfo['DEPT'] = skuinfo['DEPT'].astype(int)
skuinfo['UPC'] = skuinfo['UPC'].astype(int)
skuinfo['PACKSIZE'] = skuinfo['PACKSIZE'].astype(int)
```

```
skuinfo['VENDOR'] = skuinfo['VENDOR'].astype(int)

skuinfo.dtypes
```

[26]:
```
SKU          int64
DEPT         int64
CLASSID     object
UPC          int64
STYLE       object
COLOR       object
SIZE        object
PACKSIZE     int64
VENDOR       int64
BRAND       object
dtype: object
```

[28]:
```
merge_table = pd.merge(trnsact, skuinfo, on='SKU', how='inner')
merge_table = pd.merge(merge_table, skstinfo, on=['SKU', 'STORE'], how='inner')
merge_table
```

[28]:

	SKU	STORE	REGISTER	TRANNUM	SEQ	SALEDATE	STYPE	\
0	1383398	6703	580	3000	77002396	2004-09-23	P	
1	1383415	9002	220	5600	14509342	2005-08-25	P	
2	1383415	9002	400	2400	0	2004-09-10	P	
3	1383462	3502	351	3900	825605363	2005-03-04	P	
4	1383462	7907	30	3600	0	2004-08-10	P	
…	…	…	…	…	…	…	…	
520728	1483552	7407	910	3100	0	2005-03-13	P	
520729	1483552	7407	910	3200	0	2005-05-02	P	
520730	1483552	7502	720	1800	0	2004-08-24	P	
520731	1483552	7502	750	4100	440609046	2005-01-24	P	
520732	1483552	7502	770	1400	0	2005-07-12	P	

	QUANTITY	ORGPRICE	SPRICE	…	CLASSID	UPC		STYLE	\
0	1	40.0	20.00	…	009	400003398138	-04	EGG939	
1	1	69.0	12.07	…	214	400003415138	Y439	LIBERT	
2	1	69.0	17.25	…	214	400003415138	Y439	LIBERT	
3	1	39.0	9.75	…	312	400003462138	3	508224	
4	1	39.0	39.00	…	312	400003462138	3	508224	
…	…	…	…	…	…	…	…	…	
520728	1	16.0	12.00	…	501	400003552148	55	JWRU29	
520729	1	16.0	12.00	…	501	400003552148	55	JWRU29	
520730	1	16.0	16.00	…	501	400003552148	55	JWRU29	
520731	1	16.0	8.00	…	501	400003552148	55	JWRU29	
520732	1	16.0	16.00	…	501	400003552148	55	JWRU29	

	COLOR	SIZE	PACKSIZE	VENDOR	BRAND	COST	RETAIL

```
0       MULTI          ALL                1  6916222  MARY FRA   16.0  20.00
1       DINAVMU2 Q     060M               1  5010255  ENZO ANG   27.5  17.25
2       DINAVMU2 Q     060M               1  5010255  ENZO ANG   27.5  17.25
3       NAVY           095M               1  9036489  STRIDE R   18.0   9.75
4       NAVY           095M               1  9036489  STRIDE R   18.0   9.75
…              …          …       …       …        …    …     …
520728  BLUE           ALL                1    13396  LIZ CLAI    6.4  16.00
520729  BLUE           ALL                1    13396  LIZ CLAI    6.4  16.00
520730  BLUE           ALL                1    13396  LIZ CLAI    6.4  16.00
520731  BLUE           ALL                1    13396  LIZ CLAI    6.4  16.00
520732  BLUE           ALL                1    13396  LIZ CLAI    6.4  16.00

[520733 rows x 26 columns]
```

[29]:
```
merge_table = pd.merge(merge_table, deptinfo, on = 'DEPT', how='inner')
merge_table = pd.merge(merge_table, strinfo, on = 'STORE', how='inner')
merge_table
```

[29]:
```
             SKU  STORE  REGISTER  TRANNUM        SEQ   SALEDATE STYPE  \
0        1383398   6703       580     3000   77002396 2004-09-23     P
1        1428516   6703       570     3800  547606993 2005-08-25     P
2        1428516   6703       580     4700          0 2005-07-15     P
3        1446407   6703       550     6500          0 2005-02-23     P
4        1446407   6703       560     1500          0 2005-01-30     P
…            …      …         …        …          …        …    …
520728   1464816    404       490      600   38604273 2005-01-06     P
520729   1464816    404       490      600   38604273 2005-01-06     P
520730   1464816    404       490      800  462908445 2005-02-09     P
520731   1468625    404       490      500          0 2005-05-01     P
520732   1478625    404       450     1100          0 2005-03-18     P

         QUANTITY  ORGPRICE  SPRICE  …        SIZE  PACKSIZE   VENDOR  \
0               1      40.0   20.00  …  ALL                1  6916222
1               1      20.0   20.00  …  ALL                1  7619403
2               1      20.0   20.00  …  ALL                1  7619403
3               1      12.0    3.00  …  ALL                1  1411309
4               1      12.0    5.62  …  ALL                1  1411309
…             …        …      …     …          …        …       …
520728          1     200.0   50.00  …  30                 1  9011646
520729          1     200.0   50.00  …  30                 1  9011646
520730          1     200.0   50.00  …  30                 1  9011646
520731          1      50.0   25.00  …  TOT                1  9729207
520732          1     150.0   75.00  …  22                 1  9729207

             BRAND   COST  RETAIL  DEPTDESC                  CITY STATE    ZIP
0        MARY FRA  16.00    20.0  P&Y       NORTH OLMSTED           OH  44070
1        PRESTON    5.82    20.0  P&Y       NORTH OLMSTED           OH  44070
```

```
2       PRESTON     5.82    20.0  P&Y       NORTH OLMSTED           OH  44070
3       COLLECTI    3.00     6.0  P&Y       NORTH OLMSTED           OH  44070
4       COLLECTI    3.00     6.0  P&Y       NORTH OLMSTED           OH  44070
…            …     …      …        …                          …   …     …
520728  DELSEY     40.00    50.0  H SIERR   PINE BLUFF              AR  71601
520729  DELSEY     40.00    50.0  H SIERR   PINE BLUFF              AR  71601
520730  DELSEY     40.00    50.0  H SIERR   PINE BLUFF              AR  71601
520731  MURANO     11.47    12.5  H SIERR   PINE BLUFF              AR  71601
520732  MURANO     34.43    37.5  H SIERR   PINE BLUFF              AR  71601

[520733 rows x 30 columns]
```

[ ]:
```
# Export DataFrame to a CSV file (which can be saved with a .css extension)
merge_table.to_csv('merge_table.csv', index=False)
```

[116]:
```
import pandas as pd
merge_table = pd.read_csv("merge_table.csv")
```

[23]:
```
# # Calculate profit (SPRICE - RETAIL) for each group of 'STATE' and 'STORE'
# merge_table['PROFIT'] = merge_table['SPRICE'] - merge_table['RETAIL']

# # Group the data by 'STATE' and 'STORE', and calculate the mean profit for␣
 ↪each group
# grouped_state = merge_table.groupby(['STATE', 'STORE'])['PROFIT'].sum().
 ↪reset_index()

# # Find the group with the highest mean profit within each state
# highest_mean_profit_per_state = grouped_state.groupby('STATE').apply(lambda x:
 ↪ x.loc[x['PROFIT'].idxmax()]).reset_index(drop=True)

# highest_mean_profit_per_state.sort_values(by='PROFIT', ascending=False)
```

Analysis of Highest Profit and Discount per Brand within Each State and Store:

[117]:
```
# Calculate profit (SPRICE - RETAIL) for each group of 'STATE', 'STORE', and␣
 ↪'BRAND'
merge_table['PROFIT'] = merge_table['SPRICE'] - merge_table['RETAIL']
merge_table['discount'] = merge_table['ORGPRICE'] - merge_table['SPRICE']

# Group the data by 'STATE', 'STORE', and 'BRAND', and calculate the total␣
 ↪profit for each group
grouped_state_store_brand = merge_table.groupby(['STATE', 'STORE',␣
 ↪'BRAND'])[['PROFIT','QUANTITY','discount']].sum().reset_index()

# Find the group with the highest profit within each 'BRAND' and each 'STATE'␣
 ↪and 'STORE' combination
```

```
highest_profit_per_brand_state_store = grouped_state_store_brand.
 ↪groupby(['BRAND', 'STATE', 'STORE']).apply(
    lambda x: x.loc[x['PROFIT'].idxmax()]
).reset_index(drop=True)

highest_profit_per_brand_state_store = highest_profit_per_brand_state_store.
 ↪sort_values(by=['STATE', 'PROFIT'], ascending=[True, False])
highest_profit_per_brand_state_store
```

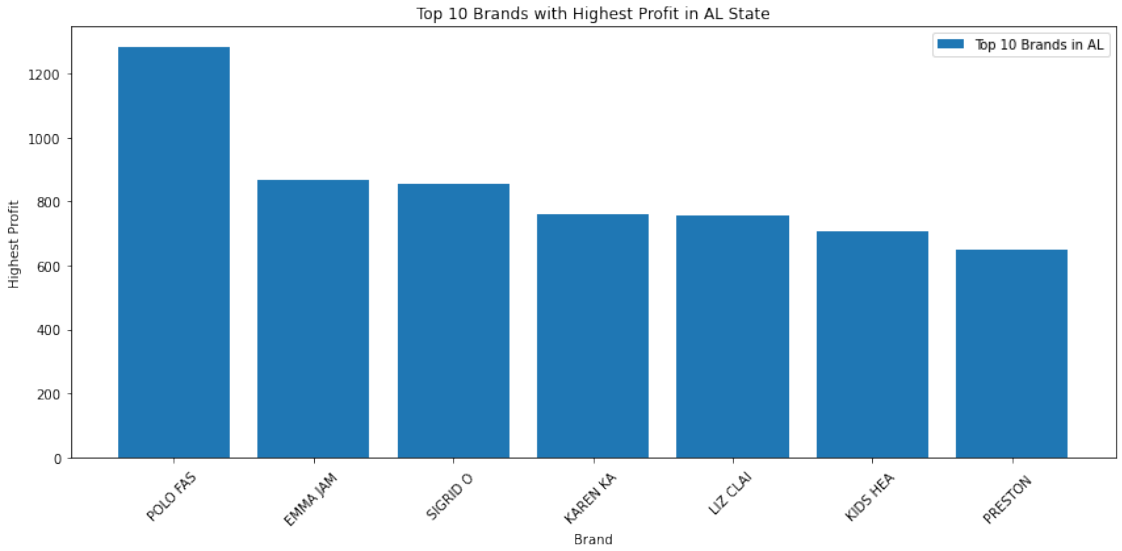
[117]:

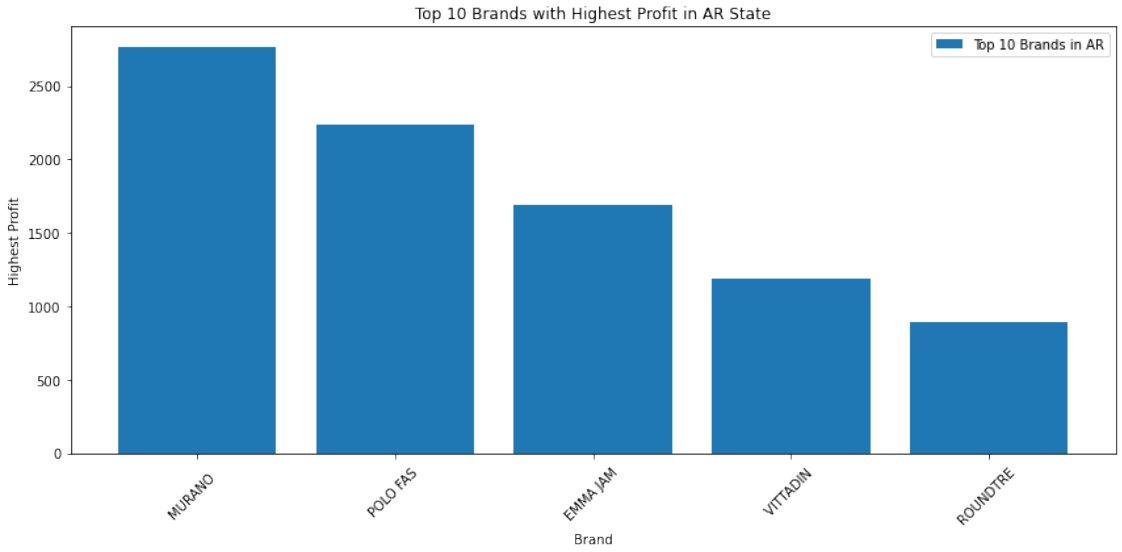
	STATE	STORE	BRAND	PROFIT	QUANTITY	discount
28588	AL	7002	POLO FAS	1282.63	40	2759.07
28587	AL	6004	POLO FAS	1241.16	61	2691.29
9510	AL	7402	EMMA JAM	865.85	44	649.57
33829	AL	4102	SIGRID O	855.52	30	1469.85
33837	AL	7302	SIGRID O	796.25	25	1188.92
…	…	…	…	…	…	…
11385	WY	9909	FRANCISC	-18.06	8	18.06
23318	WY	9909	MILCO IN	-19.58	36	19.58
4559	WY	9909	CABERNET	-41.51	48	438.54
13665	WY	9909	H.H. BRO	-66.84	9	467.34
28904	WY	9909	POLO FAS	-128.18	20	1956.40

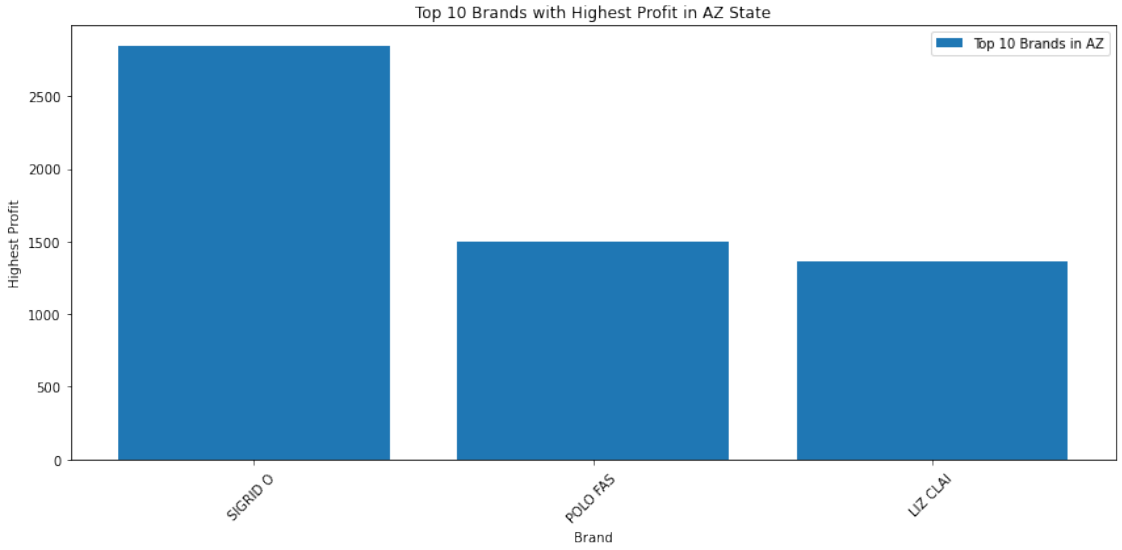
[40393 rows x 6 columns]

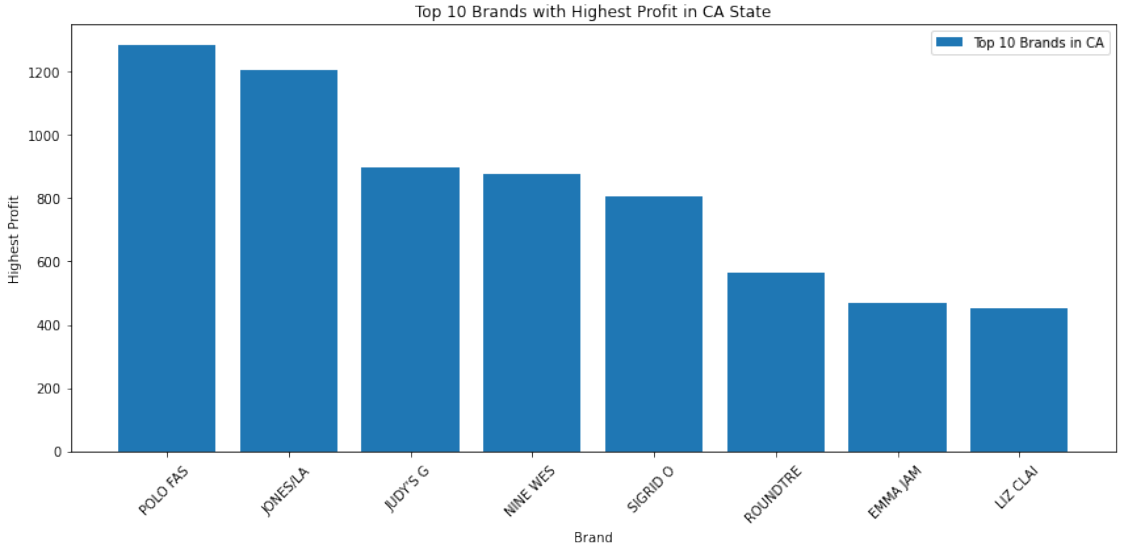
[118]:
```
import matplotlib.pyplot as plt

# Top brands with highest profit in each state

# Get a list of unique states
states = highest_profit_per_brand_state_store['STATE'].unique()

# Create a bar chart for each state
for state in states:
    state_data =␣
 ↪highest_profit_per_brand_state_store[highest_profit_per_brand_state_store['STATE']␣
 ↪== state]

    # Select the top 10 brands with the highest profit within the state
    top_brands = state_data.nlargest(10, 'PROFIT')

    # Create a figure and axis
    fig, ax = plt.subplots(figsize=(12, 6))

    # Plot the highest profit for each of the top 10 brands within the state
    ax.bar(top_brands['BRAND'], top_brands['PROFIT'], label=f'Top {10} Brands␣
 ↪in {state}')
```

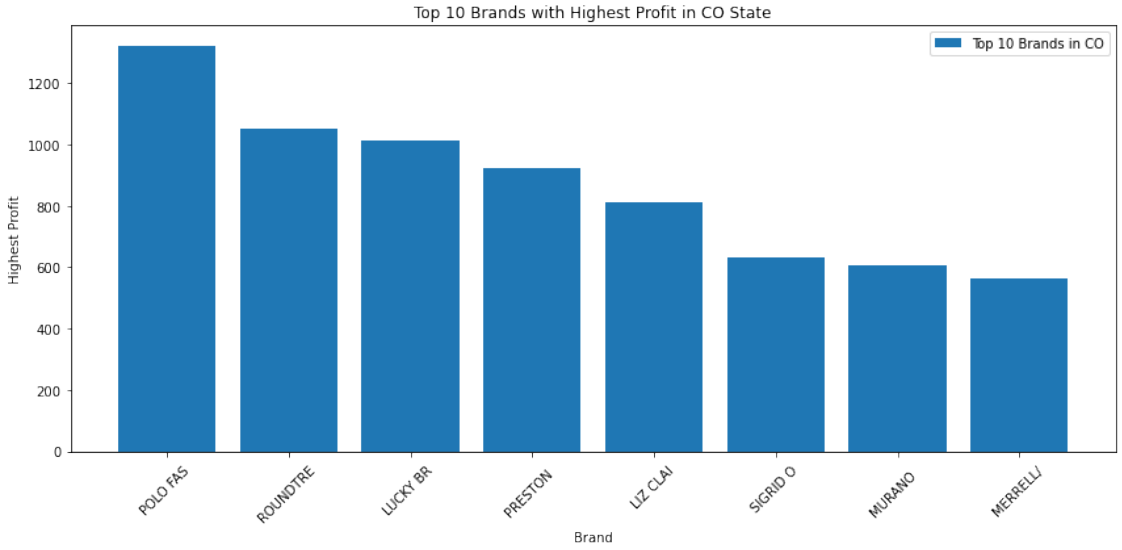
```
    # Set labels and title
    ax.set_xlabel('Brand')
    ax.set_ylabel('Highest Profit')
    ax.set_title(f'Top {10} Brands with Highest Profit in {state} State')

    # Rotate x-axis labels for better readability
    plt.xticks(rotation=45)

    # Add a legend
    ax.legend()

    # Show the plot for each state
    plt.tight_layout()
    plt.show()
```

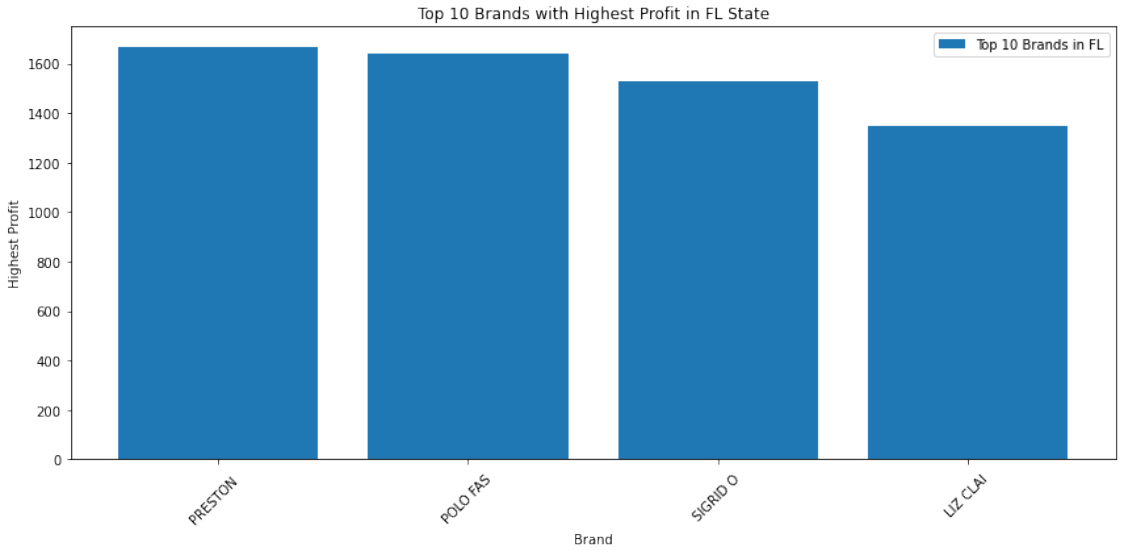


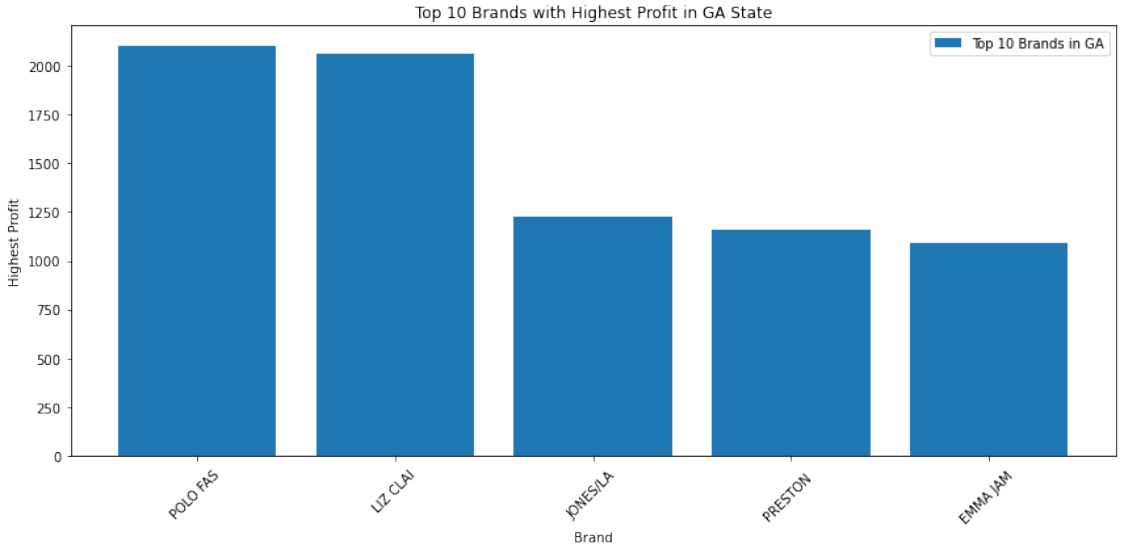

Top 10 Brands with Highest Profit in AR State
Top 10 Brands in AR
Highest Profit
2500
2000
1500
1000
500
0
MURANO
POLO FAS
EMMA JAM
VITTADIN
ROUNDTRE
Brand

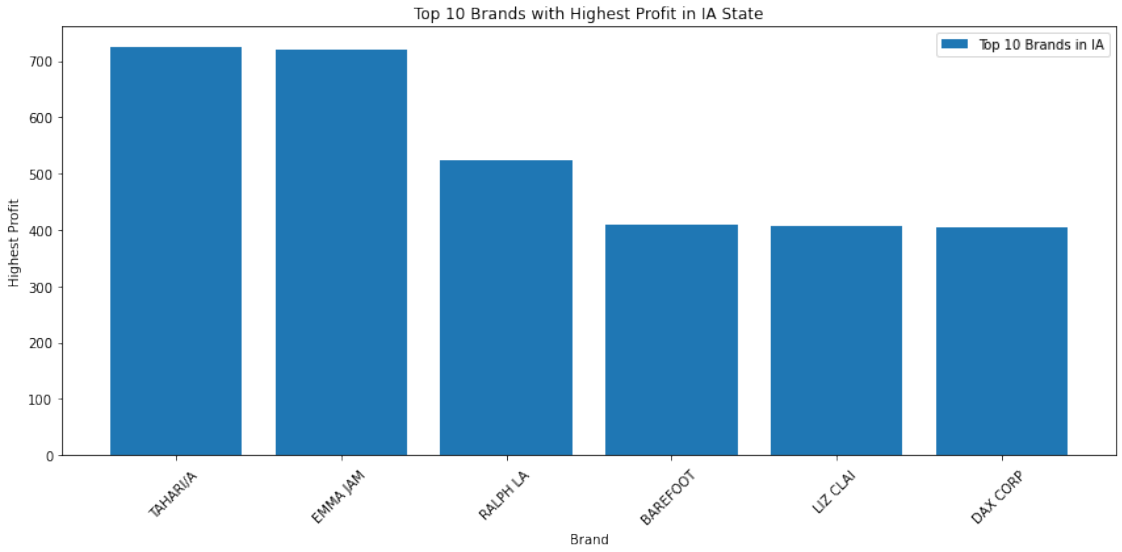

Top 10 Brands with Highest Profit in AZ State
Top 10 Brands in AZ
Highest Profit
2500
2000
1500
1000
500
0
SIGRID O
POLO FAS
LIZ CLAI
Brand

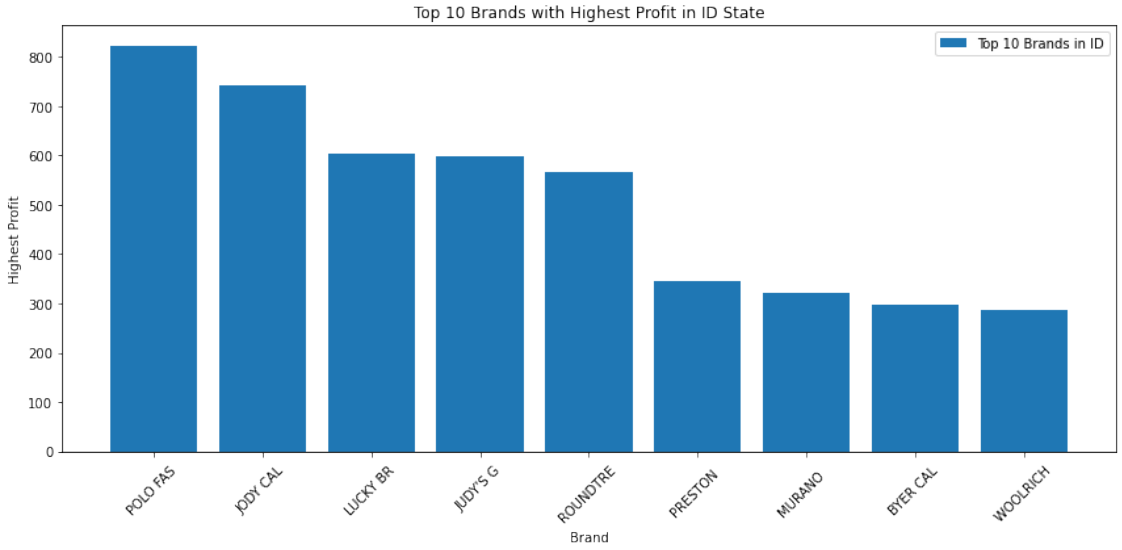

Top 10 Brands with Highest Profit in CA State
Top 10 Brands in CA
Highest Profit
0
200
400
600
800
1000
1200
POLO FAS
JONES/LA
JUDY'S G
NINE WES
SIGRID O
ROUNDTRE
EMMA JAM
LIZ CLAI
Brand

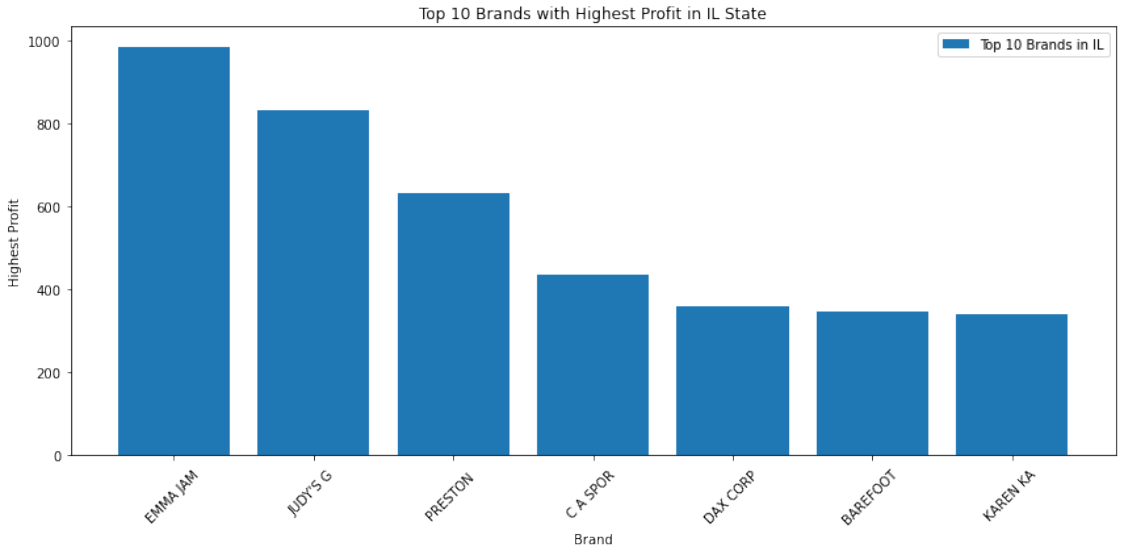

Top 10 Brands with Highest Profit in CO State
Top 10 Brands in CO
Highest Profit
0
200
400
600
800
1000
1200
POLO FAS
ROUNDTRE
LUCKY BR
PRESTON
LIZ CLAI
SIGRID O
MURANO
MERRELL/
Brand

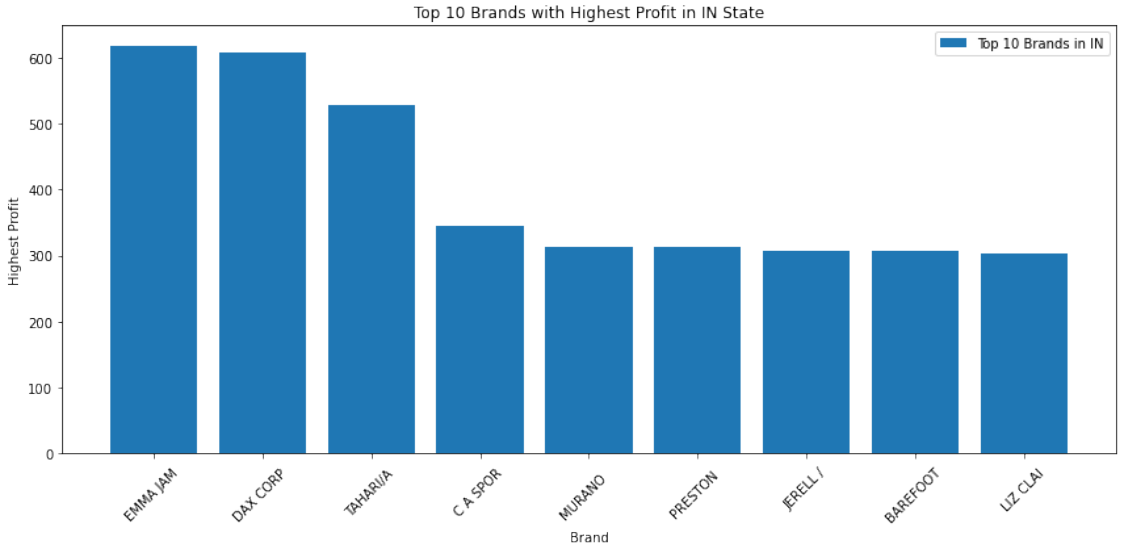

Top 10 Brands with Highest Profit in FL State
Top 10 Brands in FL
Highest Profit
0
200
400
600
800
1000
1200
1400
1600
PRESTON
POLO FAS
SIGRID O
LIZ CLAI
Brand

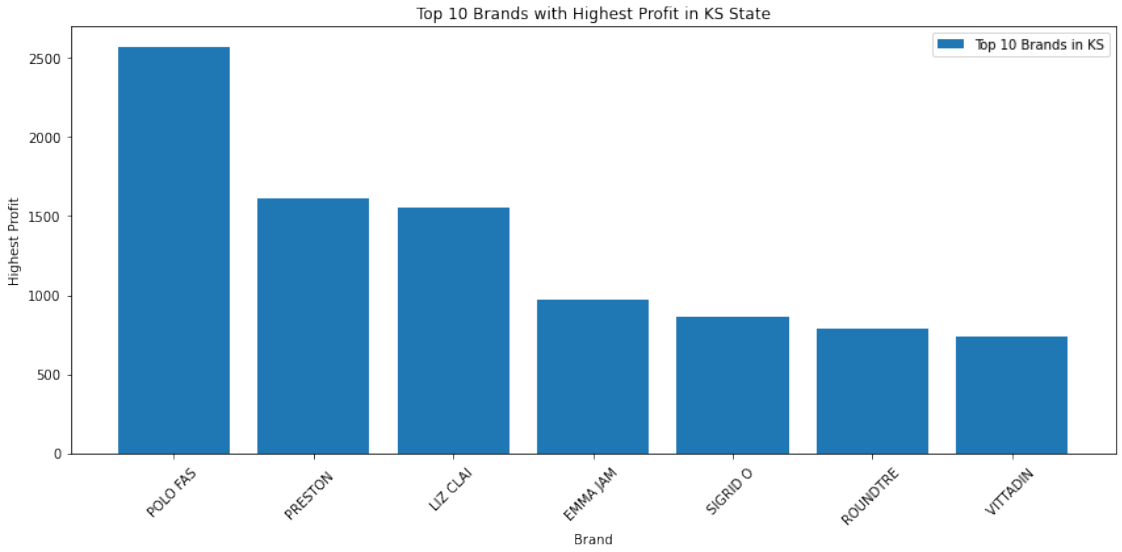

Top 10 Brands with Highest Profit in GA State
Top 10 Brands in GA
Highest Profit
0
250
500
750
1000
1250
1500
1750
2000
POLO FAS
LIZ CLAI
JONES/LA
PRESTON
EMMA JAM
Brand

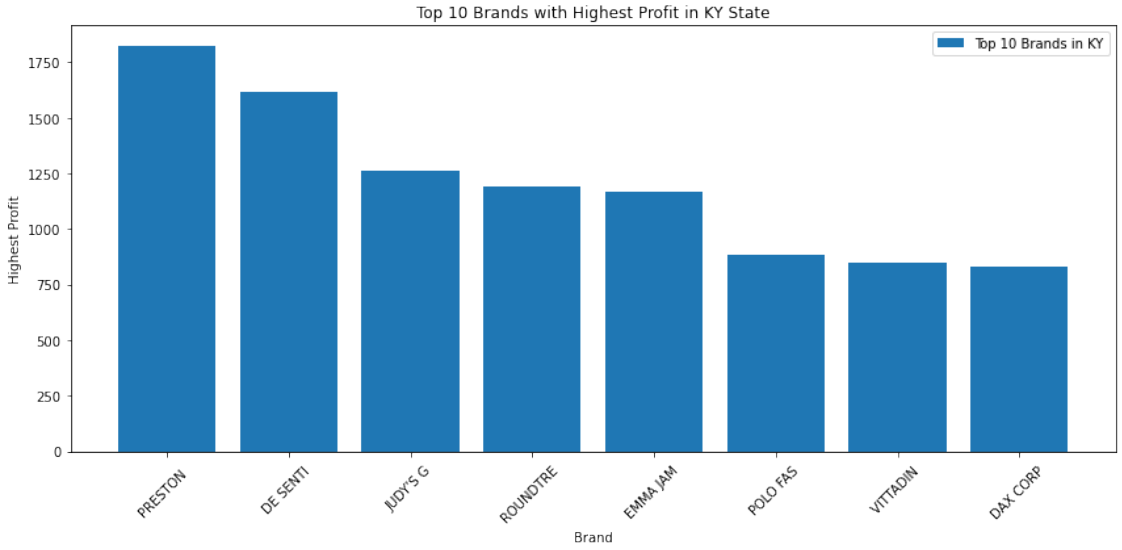

Top 10 Brands with Highest Profit in IA State
Top 10 Brands in IA
Highest Profit
0
100
200
300
400
500
600
700
TAHARI/A
EMMA JAM
RALPH LA
BAREFOOT
LIZ CLAI
DAX CORP
Brand

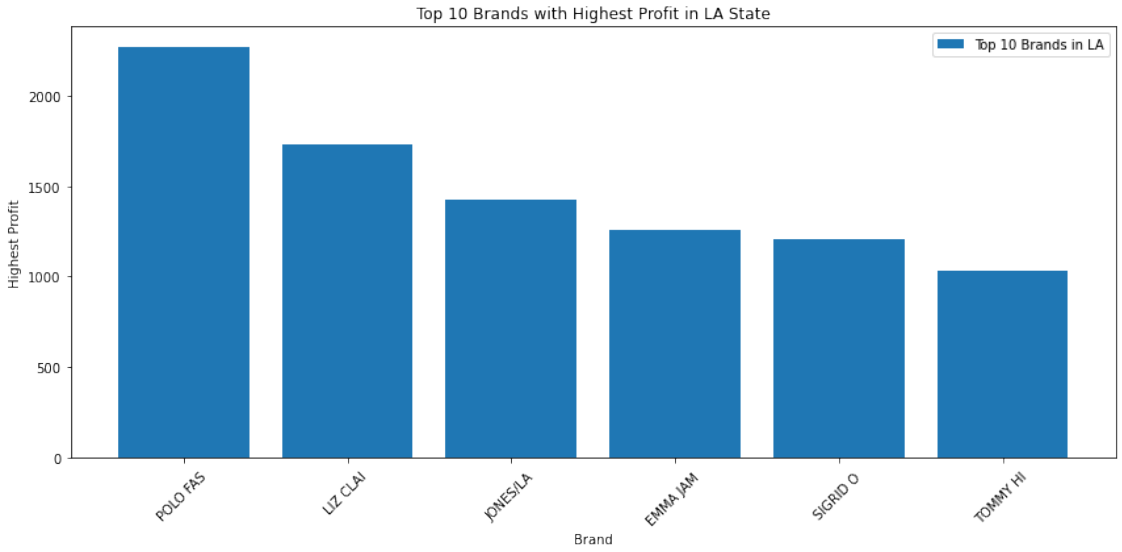

Top 10 Brands with Highest Profit in ID State
Top 10 Brands in ID
Highest Profit
0
100
200
300
400
500
600
700
800
POLO FAS
JODY CAL
LUCKY BR
JUDY'S G
ROUNDTRE
PRESTON
MURANO
BYER CAL
WOOLRICH
Brand

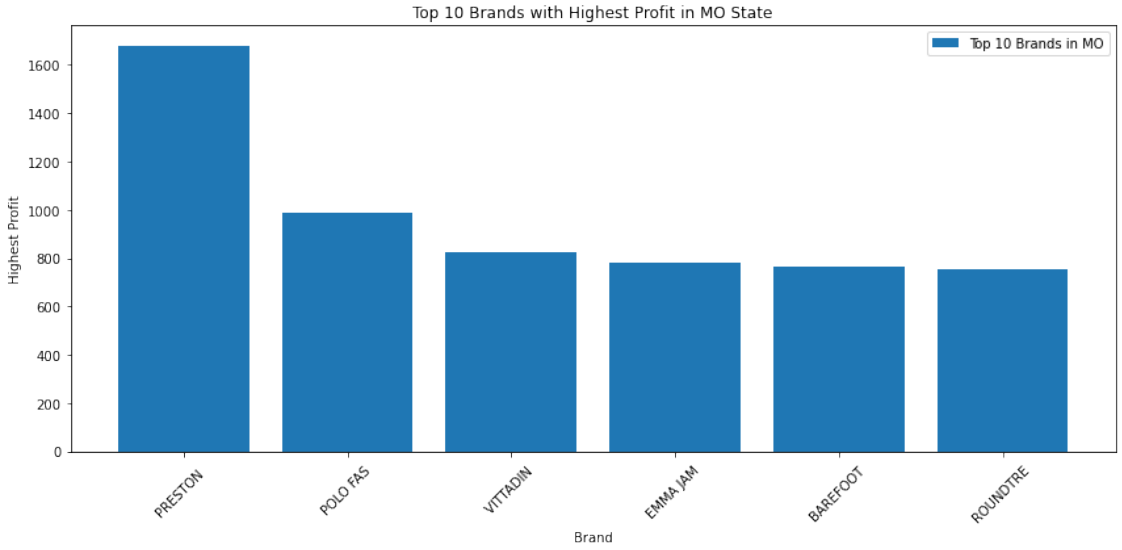

Top 10 Brands with Highest Profit in IL State
Top 10 Brands in IL
Highest Profit
1000
800
600
400
200
0
EMMA JAM
JUDY'S G
PRESTON
C A SPOR
DAX CORP
BAREFOOT
KAREN KA
Brand

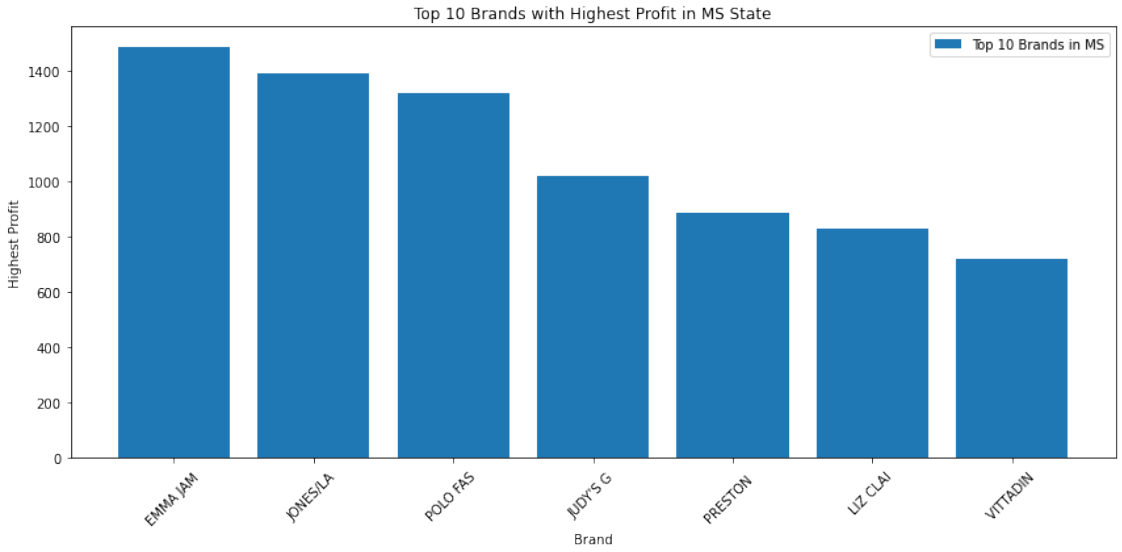

Top 10 Brands with Highest Profit in IN State
Top 10 Brands in IN
Highest Profit
600
500
400
300
200
100
0
EMMA JAM
DAX CORP
TAHARI/A
C A SPOR
MURANO
PRESTON
JERELL /
BAREFOOT
LIZ CLAI
Brand

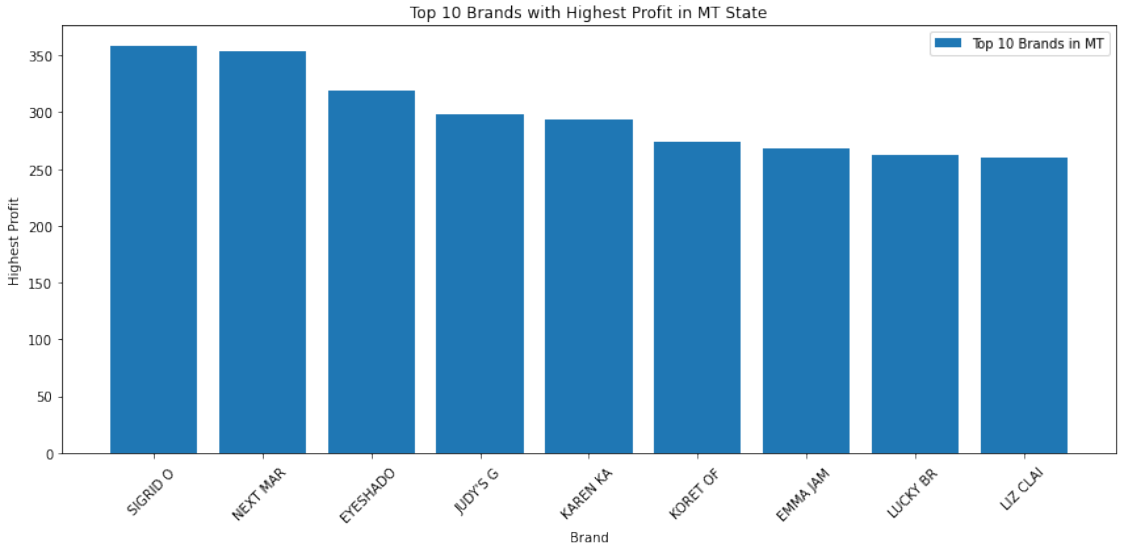

Top 10 Brands with Highest Profit in KS State
Top 10 Brands in KS
Highest Profit
2500
2000
1500
1000
500
0
POLO FAS
PRESTON
LIZ CLAI
EMMA JAM
SIGRID O
ROUNDTRE
VITTADIN
Brand

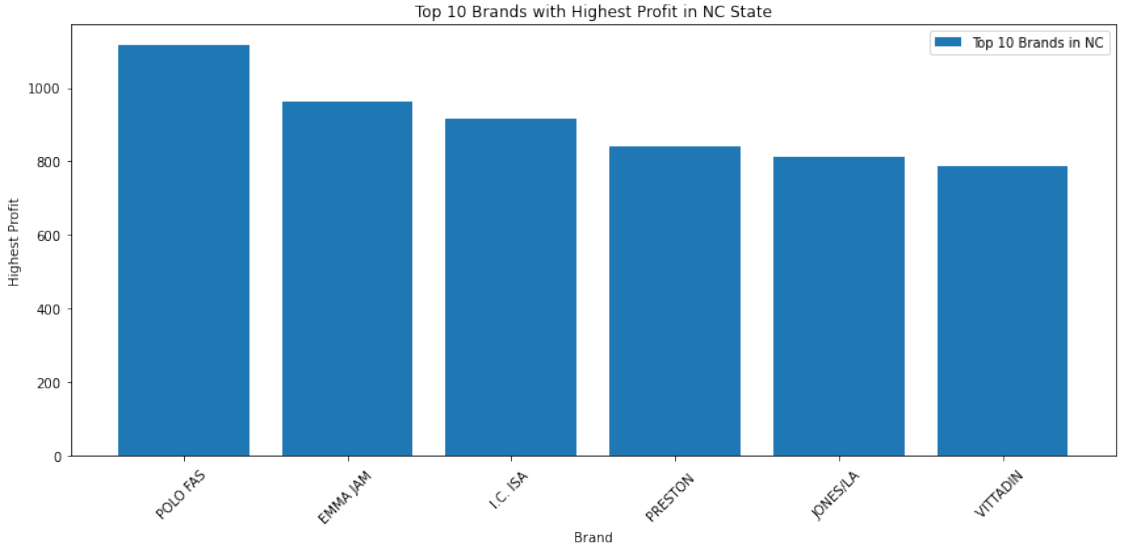

Top 10 Brands with Highest Profit in KY State
Top 10 Brands in KY
Highest Profit
1750
1500
1250
1000
750
500
250
0
PRESTON
DE SENTI
JUDY'S G
ROUNDTRE
EMMA JAM
POLO FAS
VITTADIN
DAX CORP
Brand

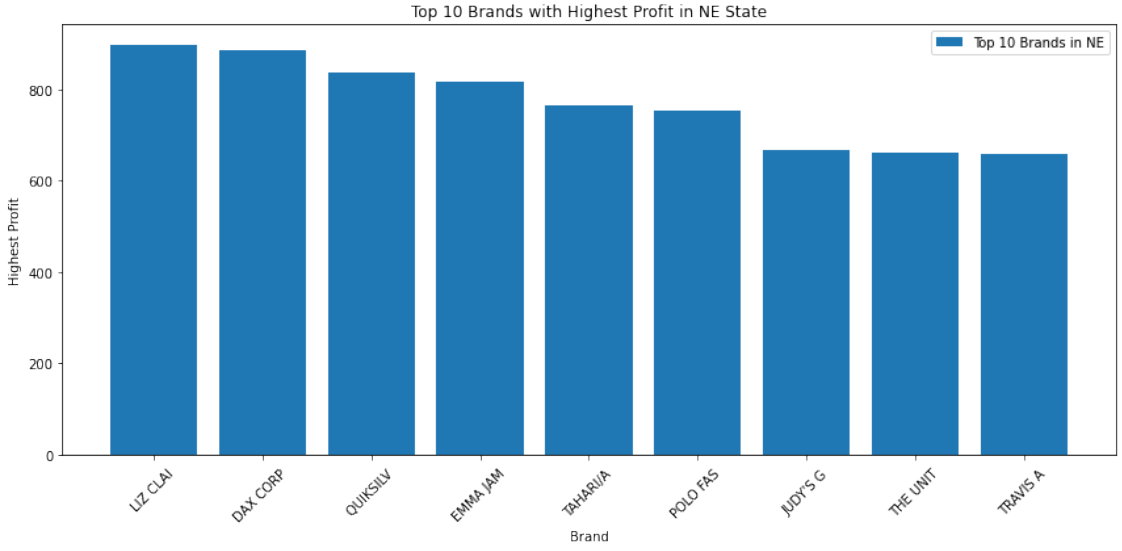

Top 10 Brands with Highest Profit in LA State
Top 10 Brands in LA
Highest Profit
2000
1500
1000
500
0
POLO FAS
LIZ CLAI
JONES/LA
EMMA JAM
SIGRID O
TOMMY HI
Brand

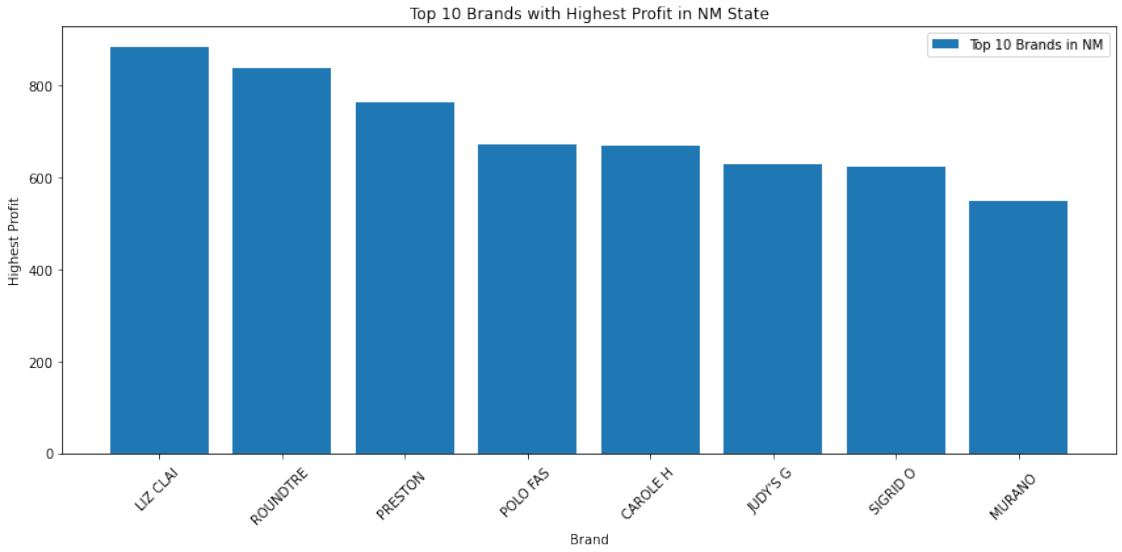

Top 10 Brands with Highest Profit in MO State
Top 10 Brands in MO
Highest Profit
1600
1400
1200
1000
800
600
400
200
0
PRESTON
POLO FAS
VITTADIN
EMMA JAM
BAREFOOT
ROUNDTRE
Brand

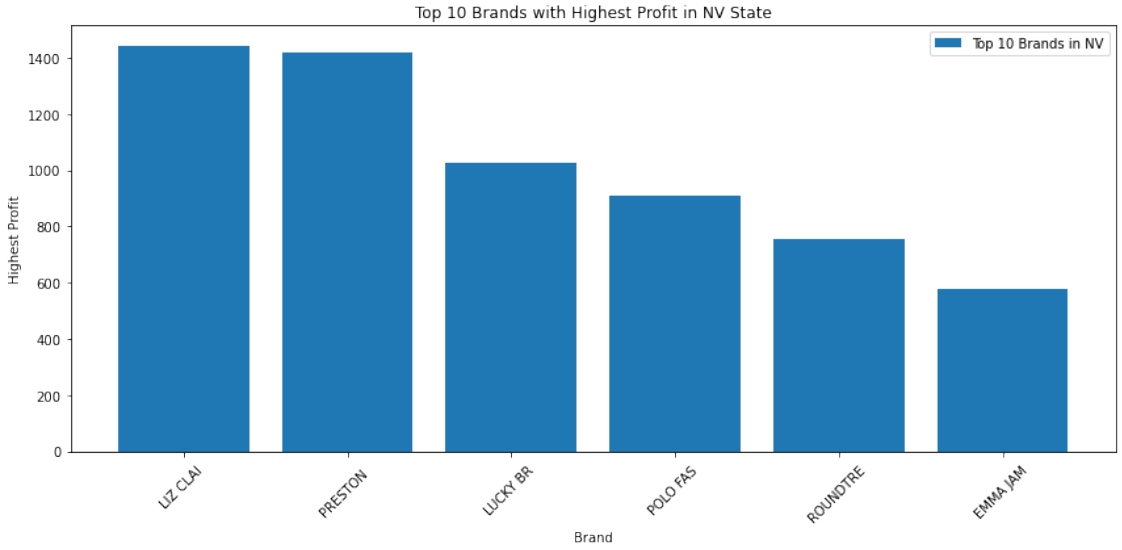

Top 10 Brands with Highest Profit in MS State
Top 10 Brands in MS
Highest Profit
0
200
400
600
800
1000
1200
1400
EMMA JAM
JONES/LA
POLO FAS
JUDY'S G
PRESTON
LIZ CLAI
VITTADIN
Brand

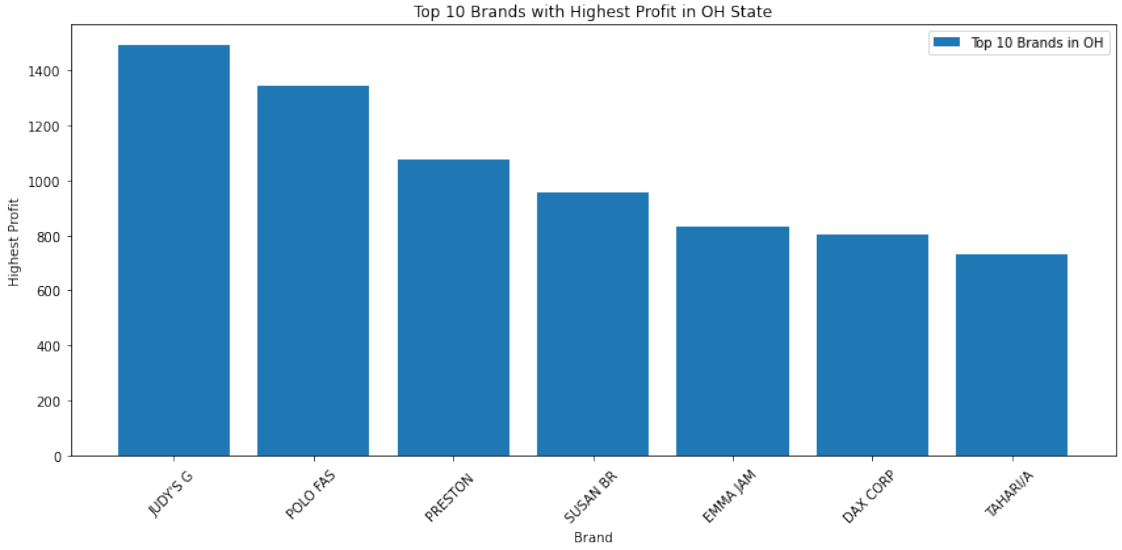

Top 10 Brands with Highest Profit in MT State
Top 10 Brands in MT
Highest Profit
0
50
100
150
200
250
300
350
SIGRID O
NEXT MAR
EYESHADO
JUDY'S G
KAREN KA
KORET OF
EMMA JAM
LUCKY BR
LIZ CLAI
Brand

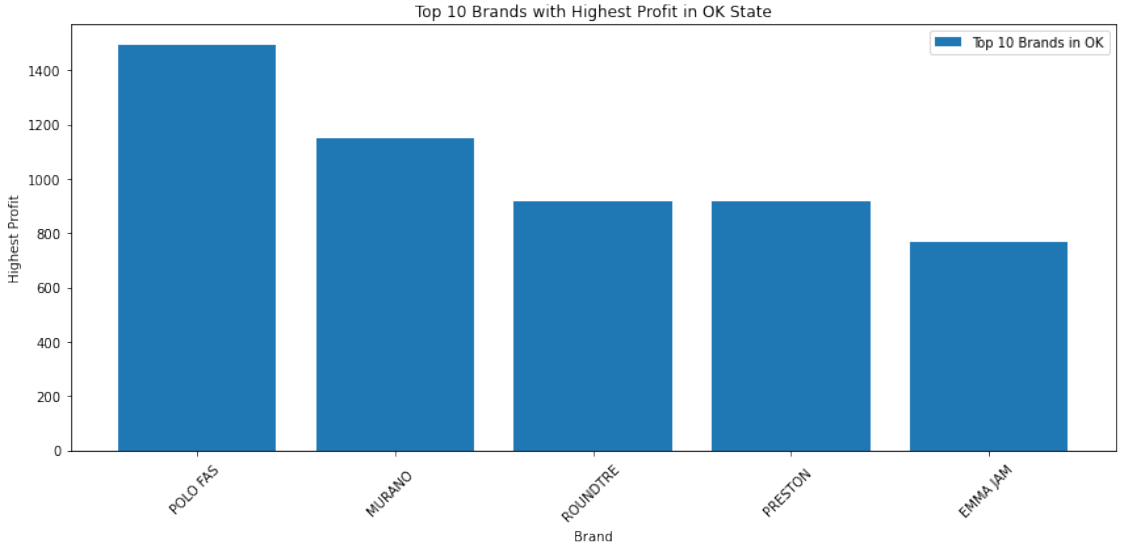

Top 10 Brands with Highest Profit in NC State
Top 10 Brands in NC
Highest Profit
0
200
400
600
800
1000
POLO FAS
EMMA JAM
I.C. ISA
PRESTON
JONES/LA
VITTADIN
Brand

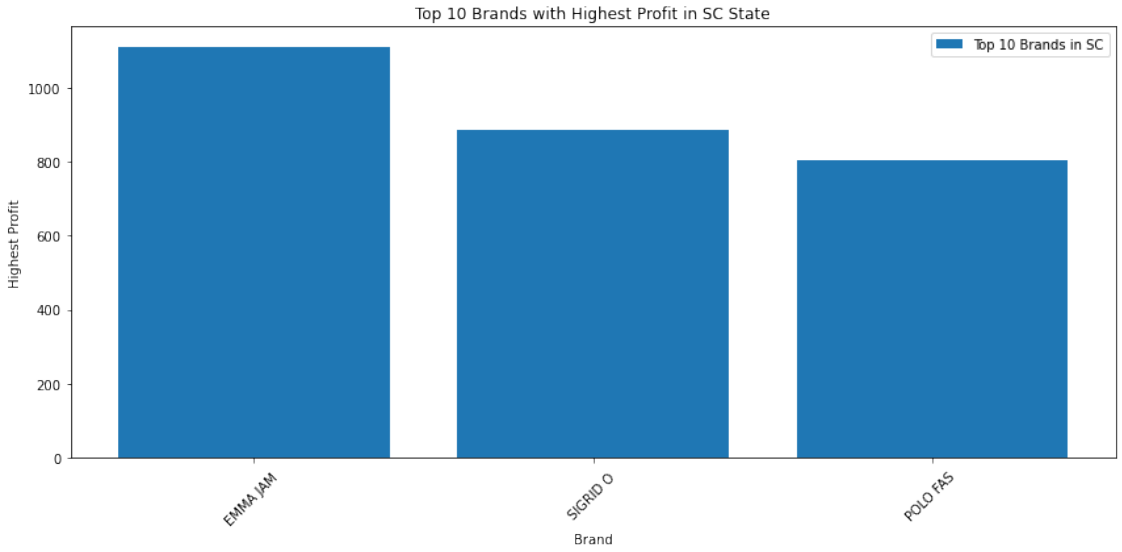

Top 10 Brands with Highest Profit in NE State
Top 10 Brands in NE
Highest Profit
0
200
400
600
800
LIZ CLAI
DAX CORP
QUIKSILV
EMMA JAM
TAHARI/A
POLO FAS
JUDY'S G
THE UNIT
TRAVIS A
Brand

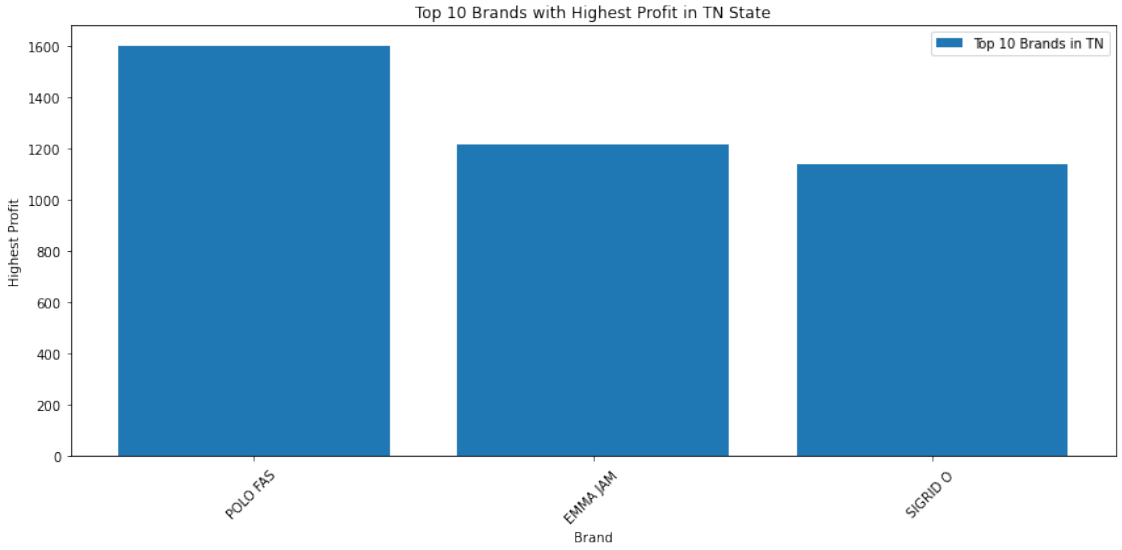

Top 10 Brands with Highest Profit in NM State
Top 10 Brands in NM
Highest Profit
0
200
400
600
800
LIZ CLAI
ROUNDTRE
PRESTON
POLO FAS
CAROLE H
JUDY'S G
SIGRID O
MURANO
Brand

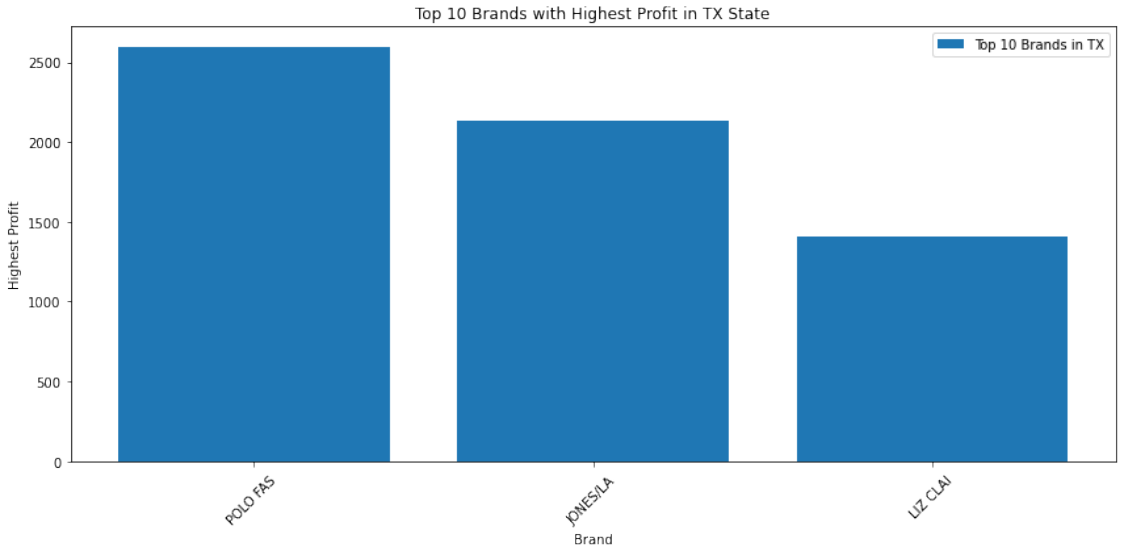

Top 10 Brands with Highest Profit in NV State
Top 10 Brands in NV
Highest Profit
0
200
400
600
800
1000
1200
1400
LIZ CLAI
PRESTON
LUCKY BR
POLO FAS
ROUNDTRE
EMMA JAM
Brand

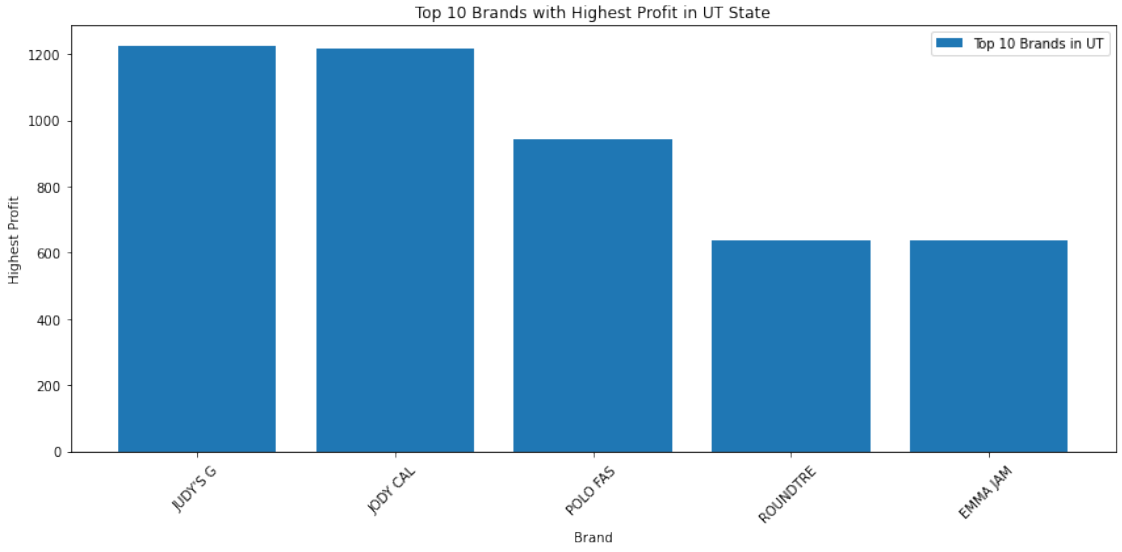

Top 10 Brands with Highest Profit in OH State
Top 10 Brands in OH
Highest Profit
0
200
400
600
800
1000
1200
1400
JUDY'S G
POLO FAS
PRESTON
SUSAN BR
EMMA JAM
DAX CORP
TAHARI/A
Brand

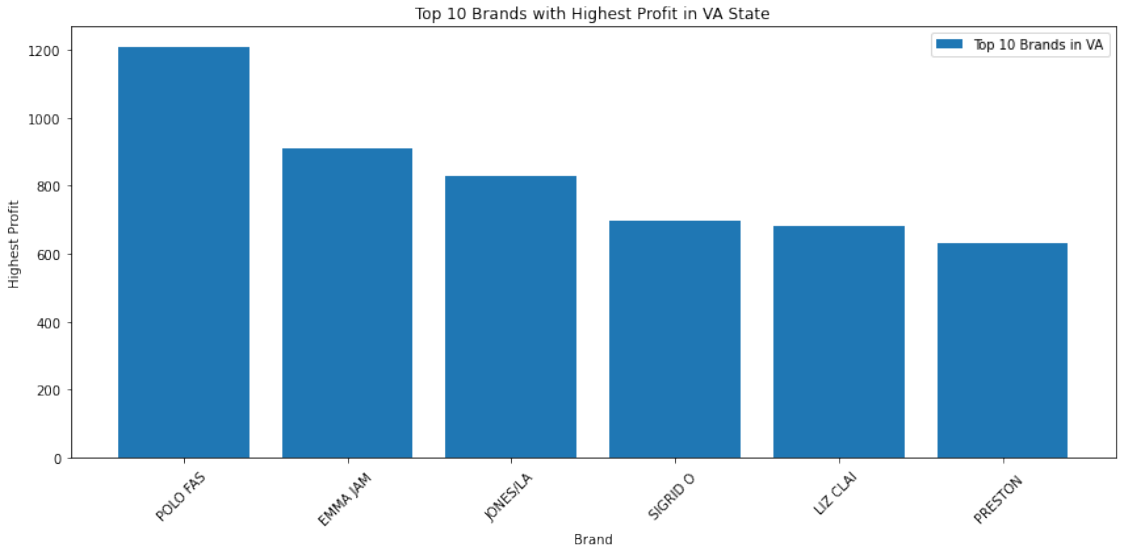

Top 10 Brands with Highest Profit in OK State
Top 10 Brands in OK
Highest Profit
0
200
400
600
800
1000
1200
1400
POLO FAS
MURANO
ROUNDTRE
PRESTON
EMMA JAM
Brand

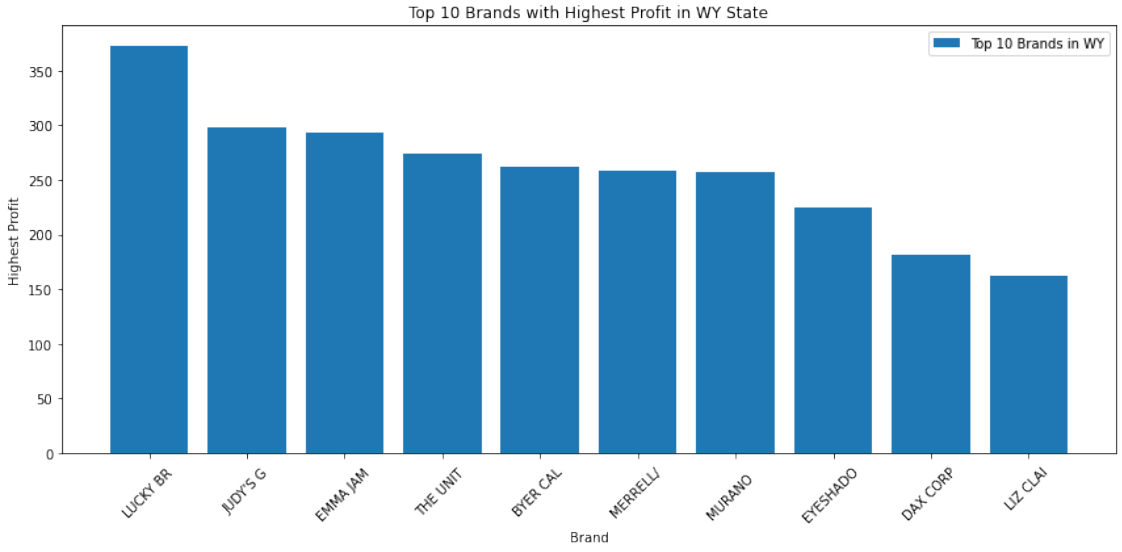

Top 10 Brands with Highest Profit in SC State
Top 10 Brands in SC
Highest Profit
0
200
400
600
800
1000
EMMA JAM
SIGRID O
POLO FAS
Brand


Top 10 Brands with Highest Profit in TN State
Top 10 Brands in TN
Highest Profit
0
200
400
600
800
1000
1200
1400
1600
POLO FAS
EMMA JAM
SIGRID O
Brand


Top 10 Brands with Highest Profit in TX State
Top 10 Brands in TX
Highest Profit
2500
2000
1500
1000
500
0
POLO FAS
JONESLA
LIZ CLAI
Brand


Top 10 Brands with Highest Profit in UT State
Top 10 Brands in UT
Highest Profit
1200
1000
800
600
400
200
0
JUDY'S G
JODY CAL
POLO FAS
ROUNDTRE
EMMA JAM
Brand





[121]:
```
from collections import Counter

# Create a dictionary to store the top 10 brands with the highest profit for␣
 ↪each state
state_top_brands = {}

for state in states:
    state_data =␣
 ↪highest_profit_per_brand_state_store[highest_profit_per_brand_state_store['STATE']␣
 ↪== state]
```

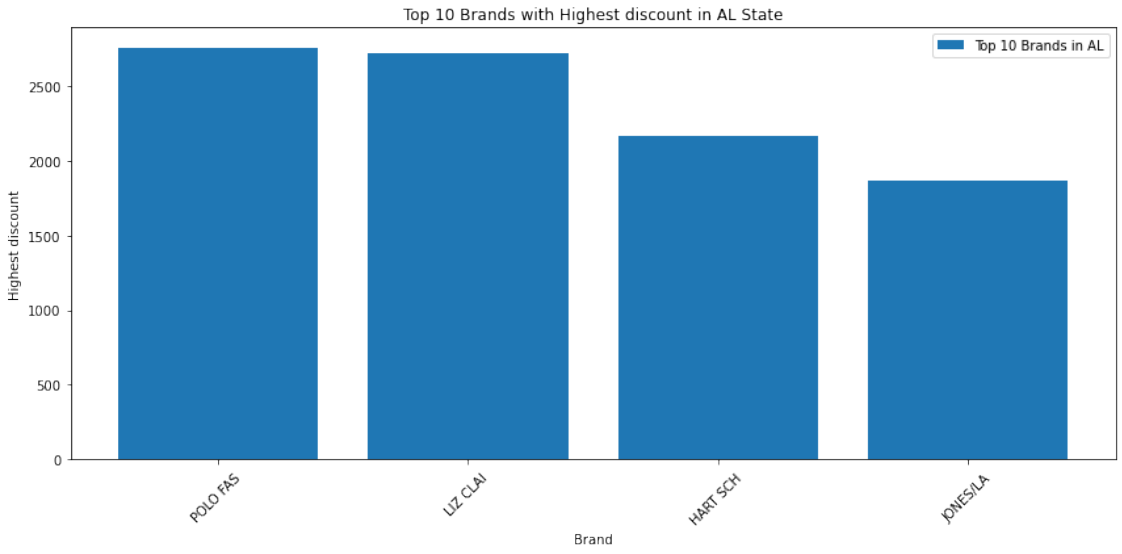
```
    top_brands = list(state_data.nlargest(10, 'PROFIT')['BRAND'])
    state_top_brands[state] = top_brands

# Find the common brands among all states
common_brands = set(state_top_brands[states[0]])  # Initialize with the brands␣
↪from the first state

# Iterate through the states and find the common brands
for state in states:
    common_brands = common_brands.intersection(state_top_brands[state])

# Count the occurrences of each brand in the common brands set
brand_counts = Counter(brand for state in states for brand in␣
↪state_top_brands[state])

# Find the most common brands
most_common_brands = [brand for brand, _ in brand_counts.most_common(5)]

# Print the first five most common brands
print("Most Common Brands That Have High Profit:")
for rank, brand in enumerate(most_common_brands, start=1):
    print(f"{rank} - {brand}")
```

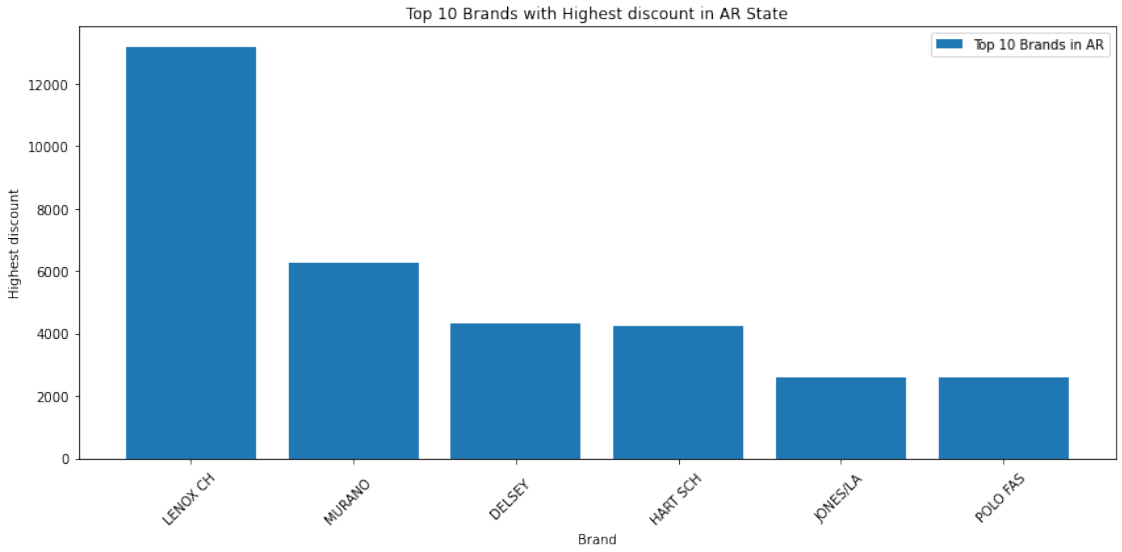
```
Most Common Brands That Have High Profit:
1 - POLO FAS
2 - EMMA JAM
3 - LIZ CLAI
4 - PRESTON
5 - JUDY'S G
```

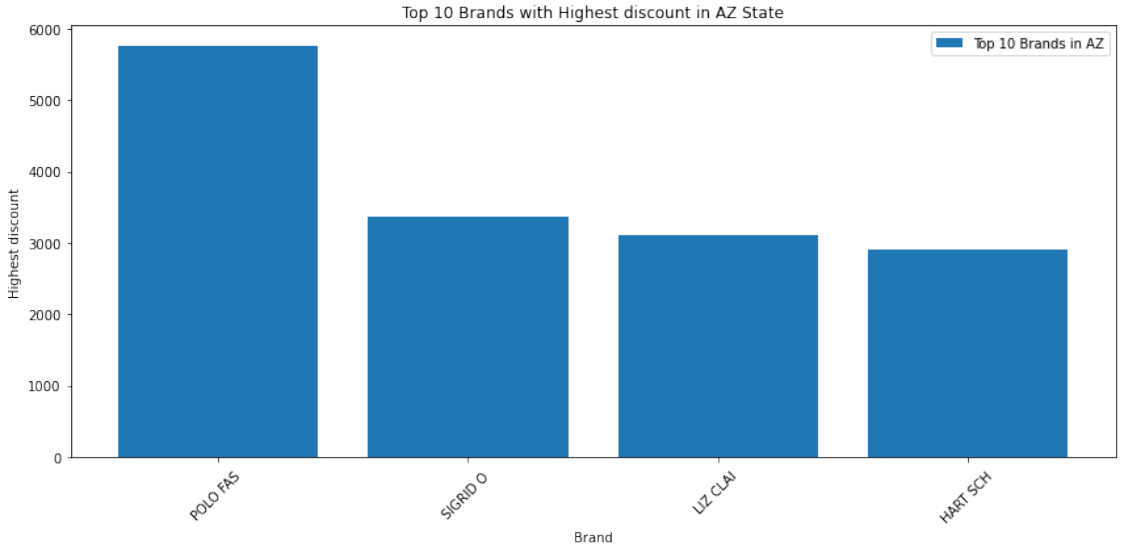
In AL, POLO FAS, EMMA JAM, SIGRID O, KAREN KA, LIZ CLAI, KIDS HEA, and PRESTON are the brands that have top profit. In AR, MURANO, POLO FAS, EMMA JAM, VITTADIN, and ROUNDTRE have high profits compare to other brands. In AZ, SIGRID O, POLO FAS, and LI CLAI have high profits. In CA, POLO FAS, JONES/LA, JUDY's G, NINE WES, SIGRID O, ROUNDTRE, EMMA JAM, and LIZ CLAI have high profits. … The first five common brands that have high profit are POLO FAS, EMMA JAM, LIZ CLAI, PRESTON, and JUDY'S G.

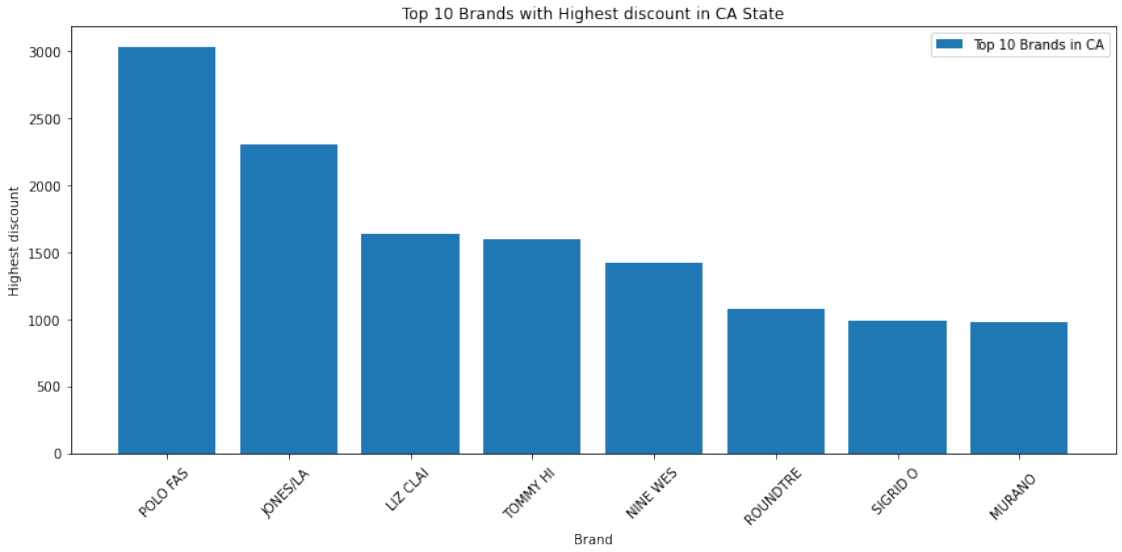
[122]:
```
# Top brands with highest discount in each state

# Create a bar chart for each state
for state in states:
    state_data =␣
↪highest_profit_per_brand_state_store[highest_profit_per_brand_state_store['STATE']␣
↪== state]

    # Select the top 10 brands with the highest profit within the state
    top_brands = state_data.nlargest(10, 'discount')
```

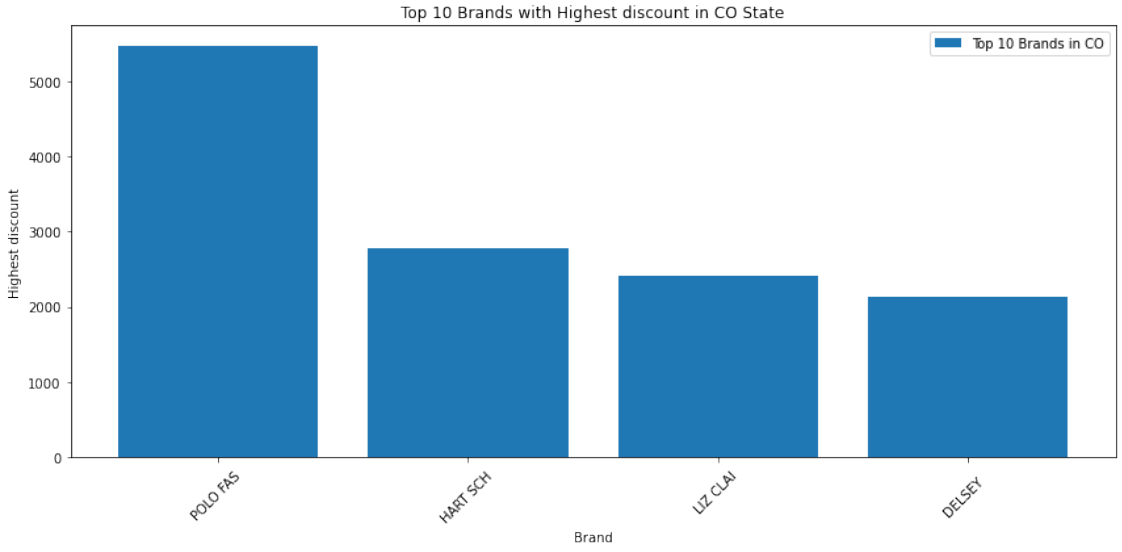
```
    # Create a figure and axis
    fig, ax = plt.subplots(figsize=(12, 6))

    # Plot the highest profit for each of the top 10 brands within the state
    ax.bar(top_brands['BRAND'], top_brands['discount'], label=f'Top {10} Brands␣
↪in {state}')

    # Set labels and title
    ax.set_xlabel('Brand')
    ax.set_ylabel('Highest discount')
    ax.set_title(f'Top {10} Brands with Highest discount in {state} State')

    # Rotate x-axis labels for better readability
    plt.xticks(rotation=45)

    # Add a legend
    ax.legend()

    # Show the plot for each state
    plt.tight_layout()
    plt.show()
```

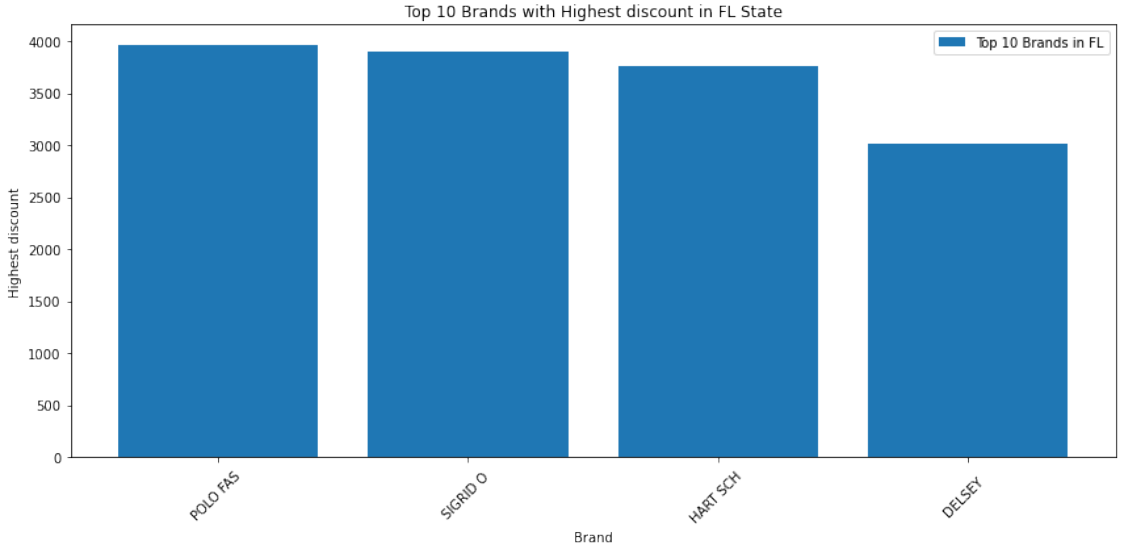


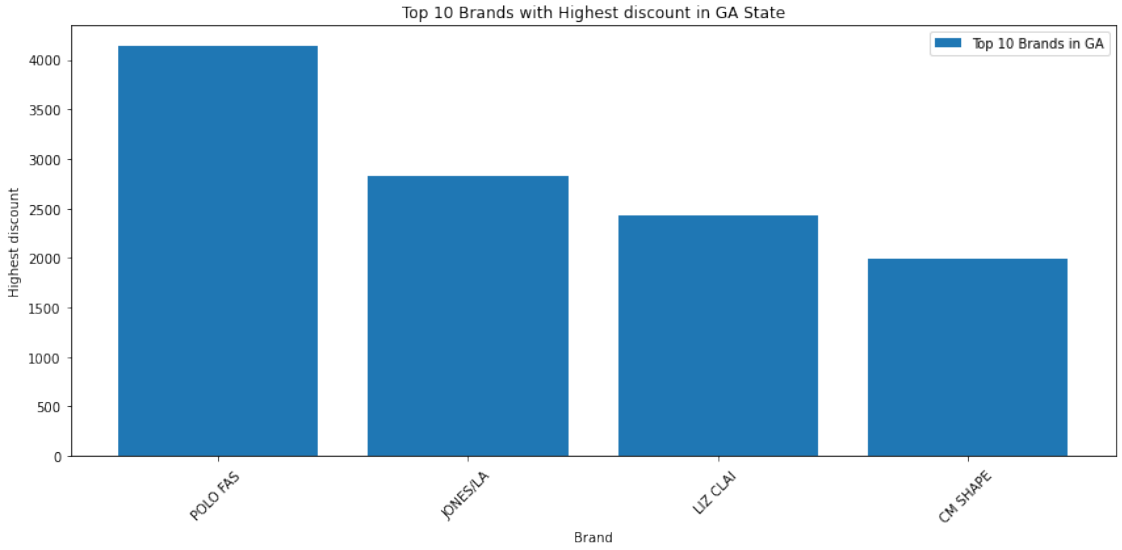

Top 10 Brands with Highest discount in AR State
Top 10 Brands in AR
Highest discount
0
2000
4000
6000
8000
10000
12000
LENOX CH
MURANO
DELSEY
HART SCH
JONES/LA
POLO FAS
Brand

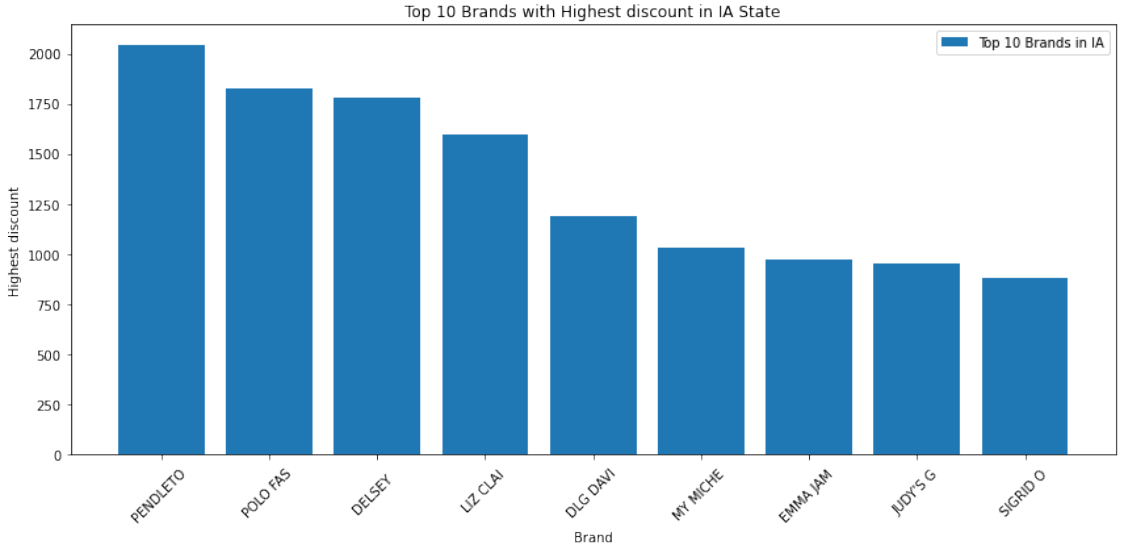

Top 10 Brands with Highest discount in AZ State
Top 10 Brands in AZ
Highest discount
0
1000
2000
3000
4000
5000
6000
POLO FAS
SIGRID O
LIZ CLAI
HART SCH
Brand

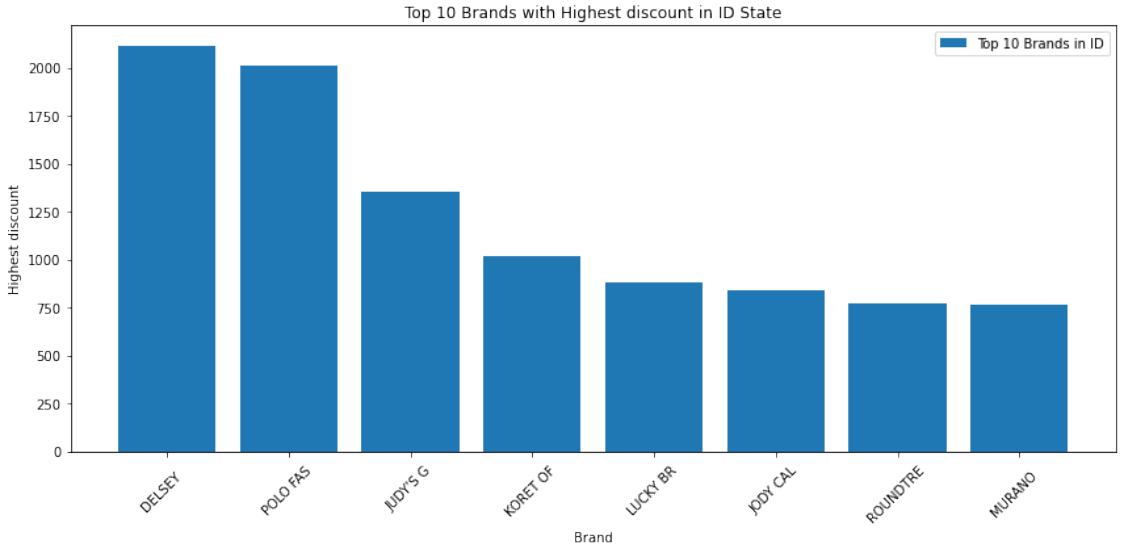

Top 10 Brands with Highest discount in CA State
Top 10 Brands in CA
Highest discount
3000
2500
2000
1500
1000
500
0
POLO FAS
JONES/LA
LIZ CLAI
TOMMY HI
NINE WES
ROUNDTRE
SIGRID O
MURANO
Brand

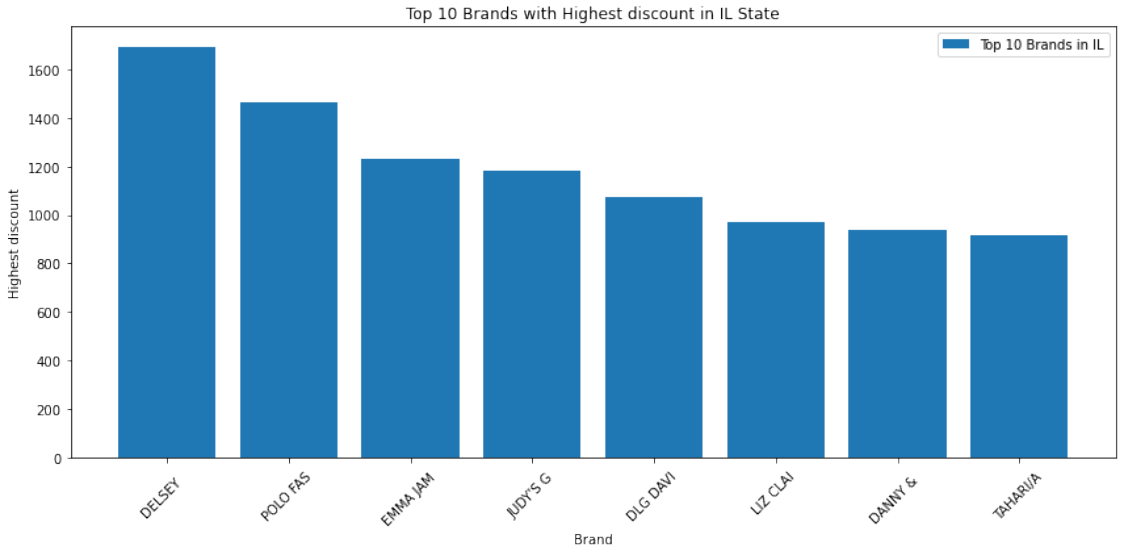

Top 10 Brands with Highest discount in CO State
Top 10 Brands in CO
Highest discount
5000
4000
3000
2000
1000
0
POLO FAS
HART SCH
LIZ CLAI
DELSEY
Brand

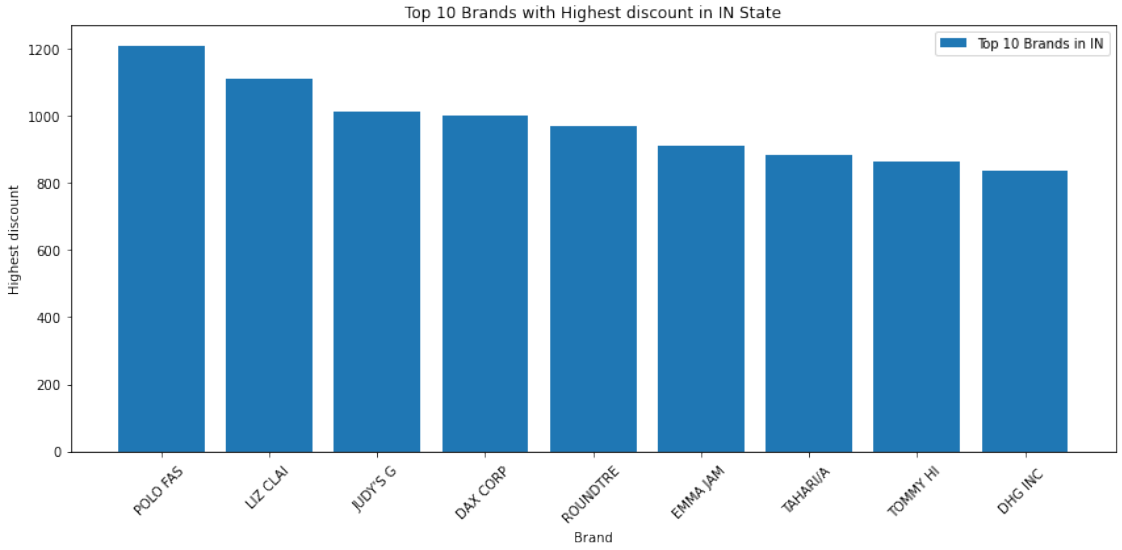

Top 10 Brands with Highest discount in FL State
Top 10 Brands in FL
Highest discount
0
500
1000
1500
2000
2500
3000
3500
4000
POLO FAS
SIGRID O
HART SCH
DELSEY
Brand

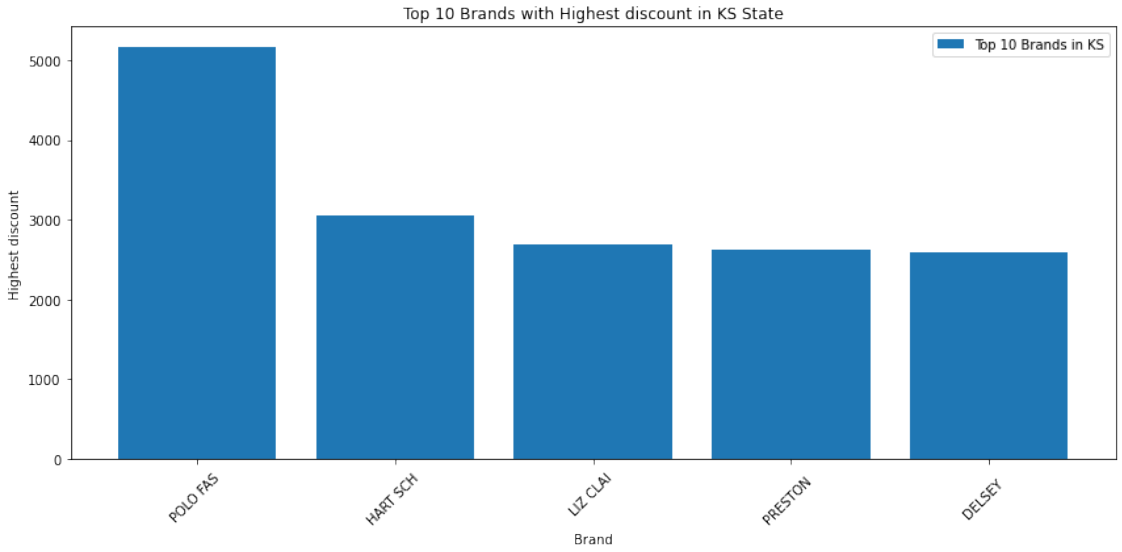

Top 10 Brands with Highest discount in GA State
Top 10 Brands in GA
Highest discount
0
500
1000
1500
2000
2500
3000
3500
4000
POLO FAS
JONES/LA
LIZ CLAI
CM SHAPE
Brand

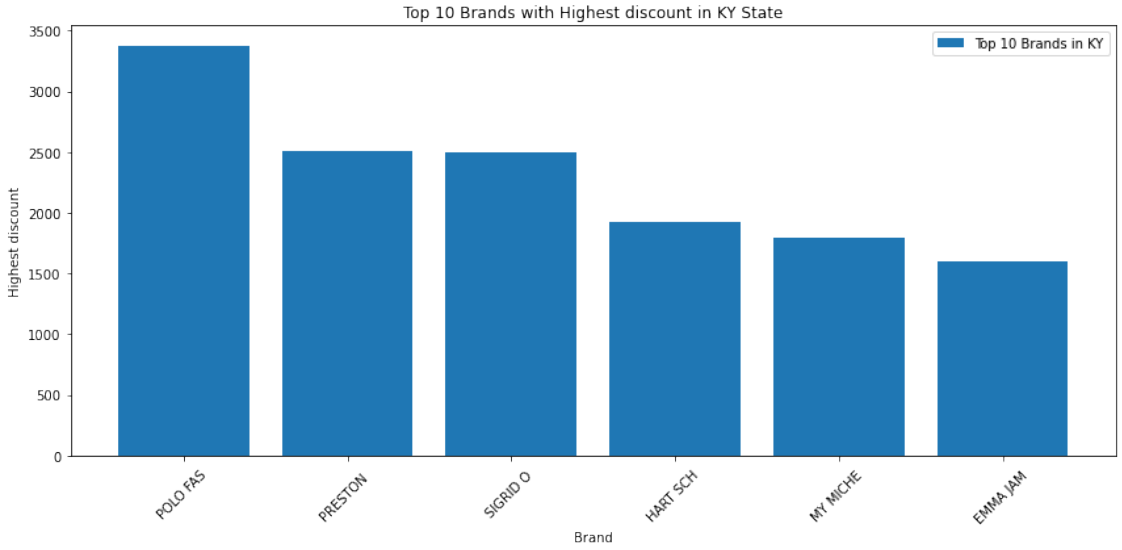

Top 10 Brands with Highest discount in IA State
Top 10 Brands in IA
Highest discount
2000
1750
1500
1250
1000
750
500
250
0
PENDLETO
POLO FAS
DELSEY
LIZ CLAI
DLG DAVI
MY MICHE
EMMA JAM
JUDY'S G
SIGRID O
Brand

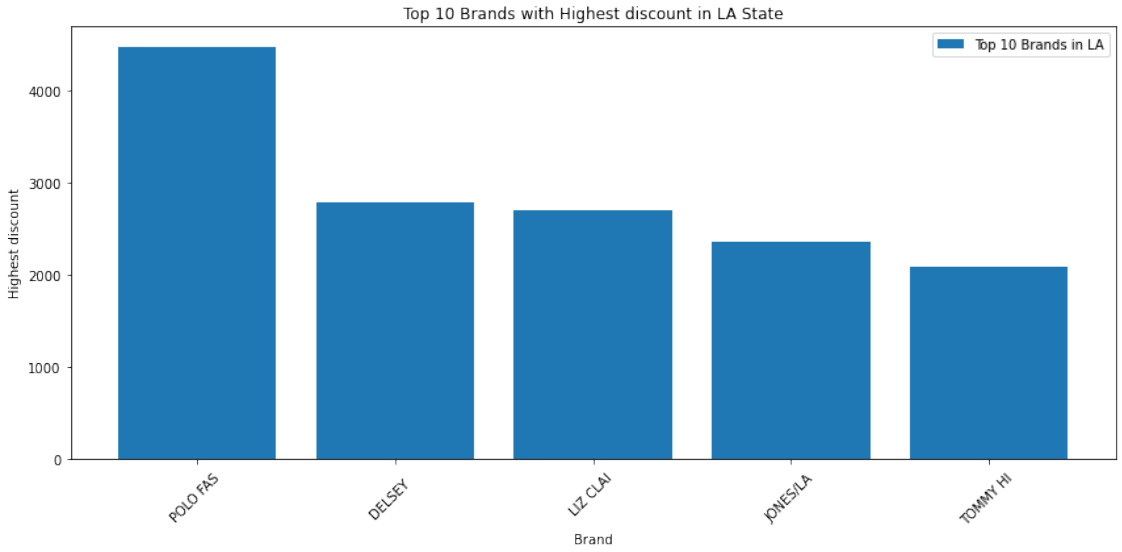

Top 10 Brands with Highest discount in ID State
Top 10 Brands in ID
Highest discount
2000
1750
1500
1250
1000
750
500
250
0
DELSEY
POLO FAS
JUDY'S G
KORET OF
LUCKY BR
JODY CAL
ROUNDTRE
MURANO
Brand

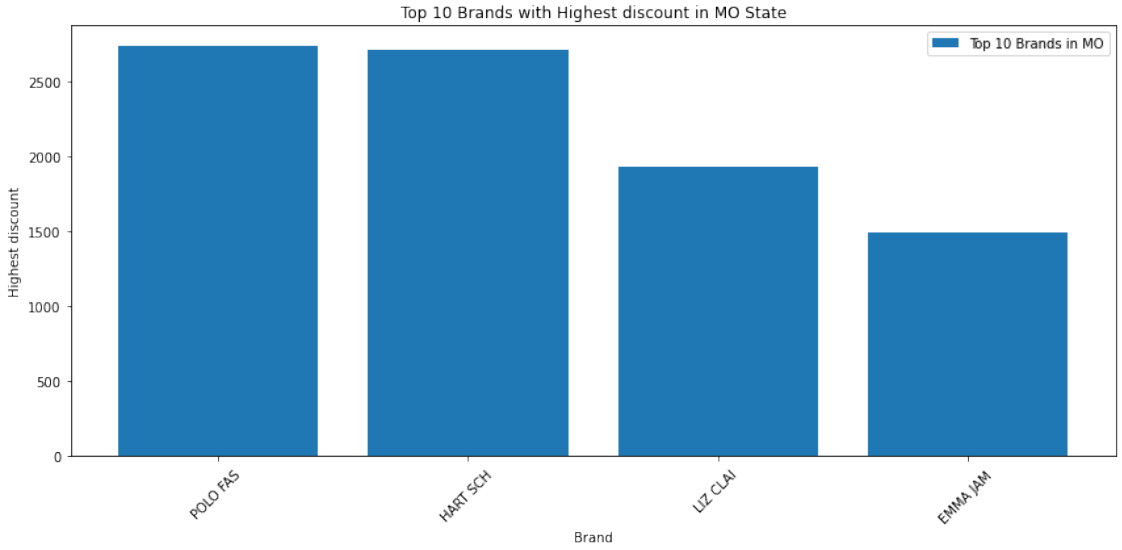

Top 10 Brands with Highest discount in IL State
Top 10 Brands in IL
Highest discount
0
200
400
600
800
1000
1200
1400
1600
DELSEY
POLO FAS
EMMA JAM
JUDY'S G
DLG DAVI
LIZ CLAI
DANNY &
TAHARI/A
Brand

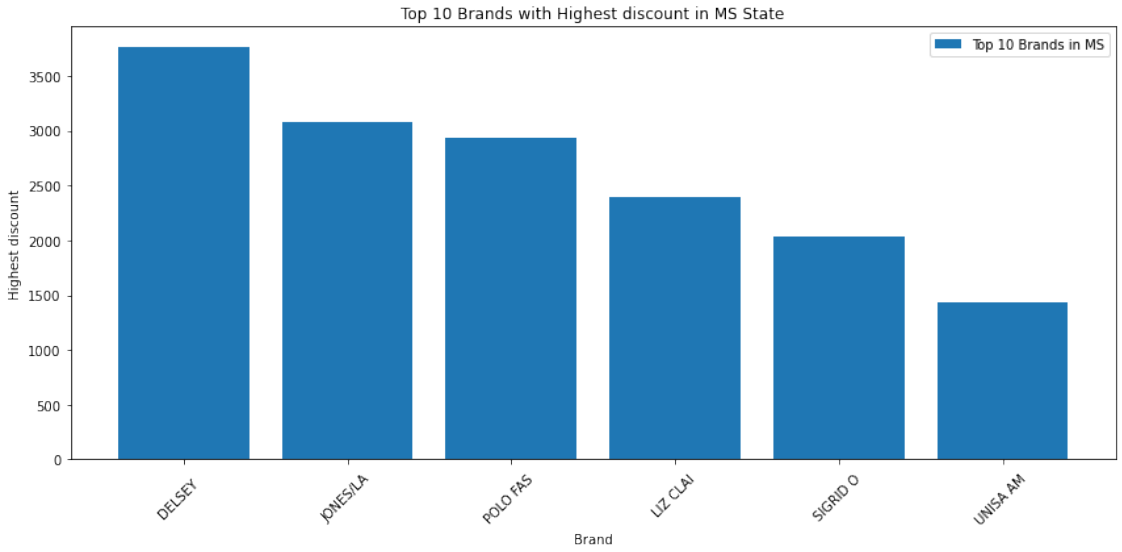

Top 10 Brands with Highest discount in IN State
Top 10 Brands in IN
Highest discount
0
200
400
600
800
1000
1200
POLO FAS
LIZ CLAI
JUDY'S G
DAX CORP
ROUNDTRE
EMMA JAM
TAHARI/A
TOMMY HI
DHG INC
Brand

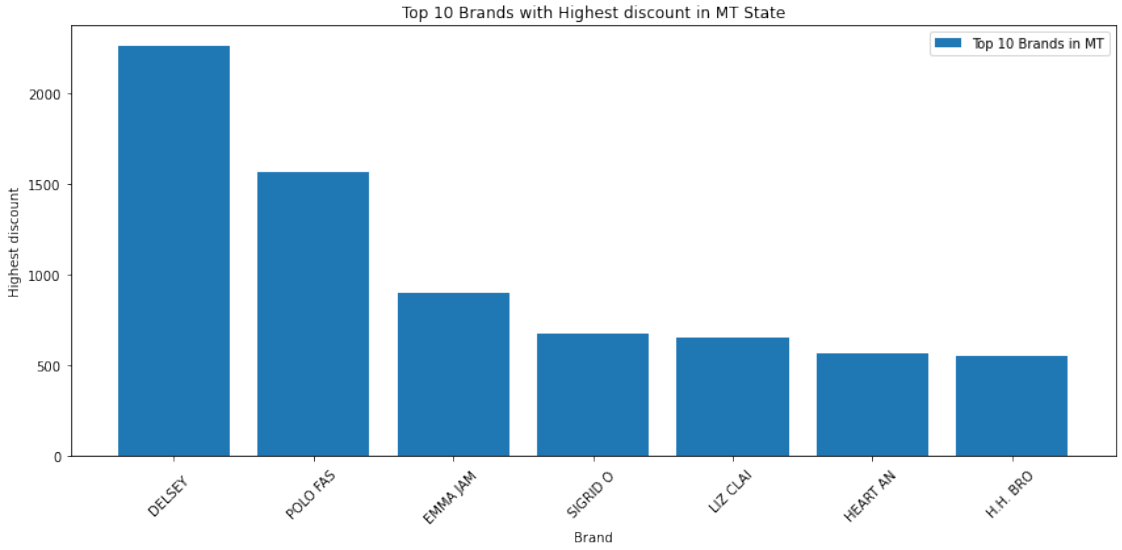

Top 10 Brands with Highest discount in KS State
Top 10 Brands in KS
Highest discount
0
1000
2000
3000
4000
5000
POLO FAS
HART SCH
LIZ CLAI
PRESTON
DELSEY
Brand

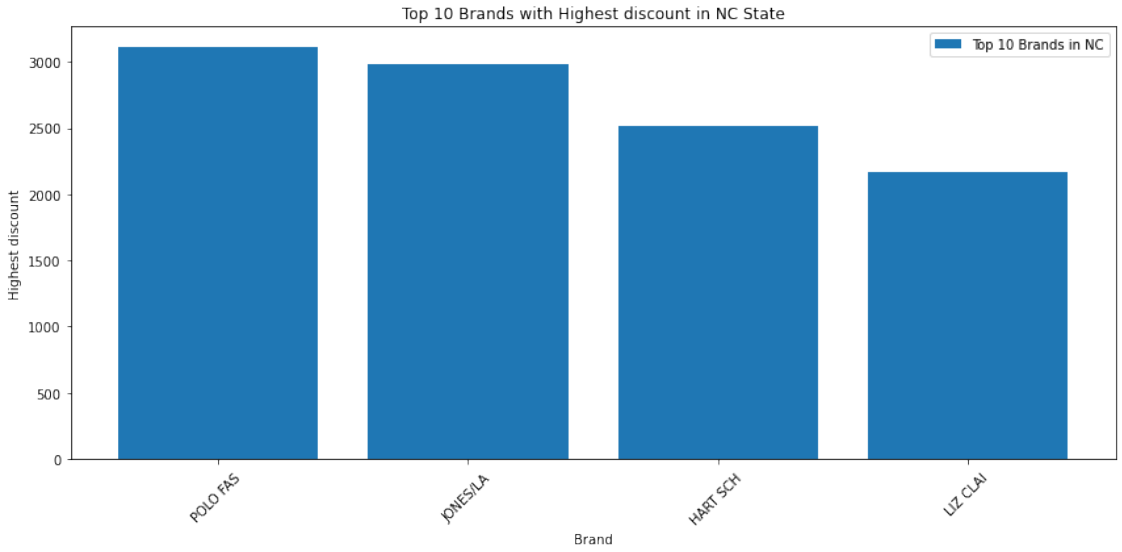

Top 10 Brands with Highest discount in KY State
Top 10 Brands in KY
Highest discount
0
500
1000
1500
2000
2500
3000
3500
POLO FAS
PRESTON
SIGRID O
HART SCH
MY MICHE
EMMA JAM
Brand

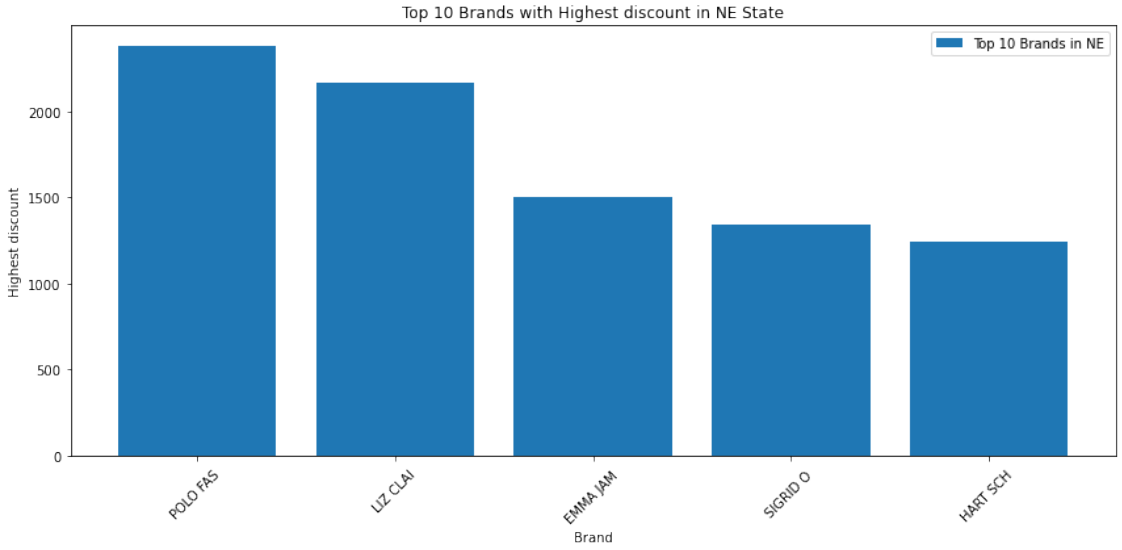

Top 10 Brands with Highest discount in LA State
Top 10 Brands in LA
Highest discount
0
1000
2000
3000
4000
POLO FAS
DELSEY
LIZ CLAI
JONES/LA
TOMMY HI
Brand

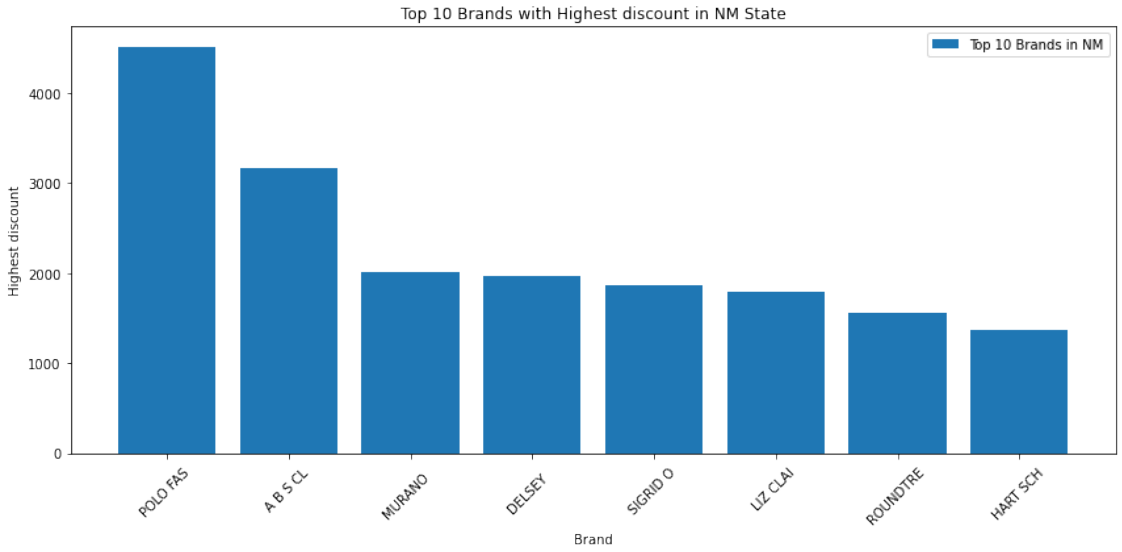

Top 10 Brands with Highest discount in MO State
Top 10 Brands in MO
Highest discount
0
500
1000
1500
2000
2500
POLO FAS
HART SCH
LIZ CLAI
EMMA JAM
Brand

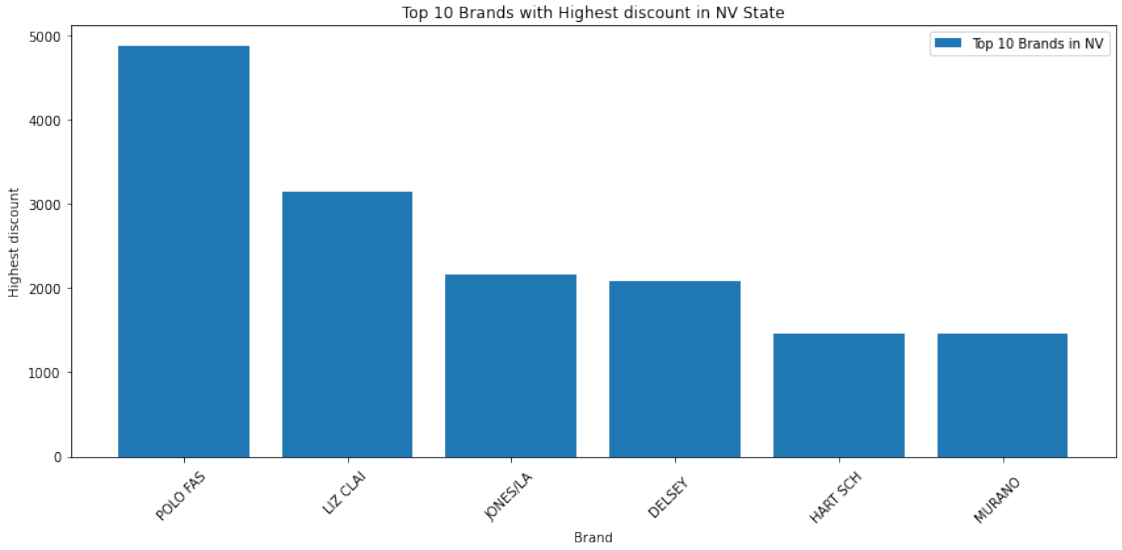

Top 10 Brands with Highest discount in MS State
Top 10 Brands in MS
Highest discount
0
500
1000
1500
2000
2500
3000
3500
DELSEY
JONES/LA
POLO FAS
LIZ CLAI
SIGRID O
UNISA AM
Brand

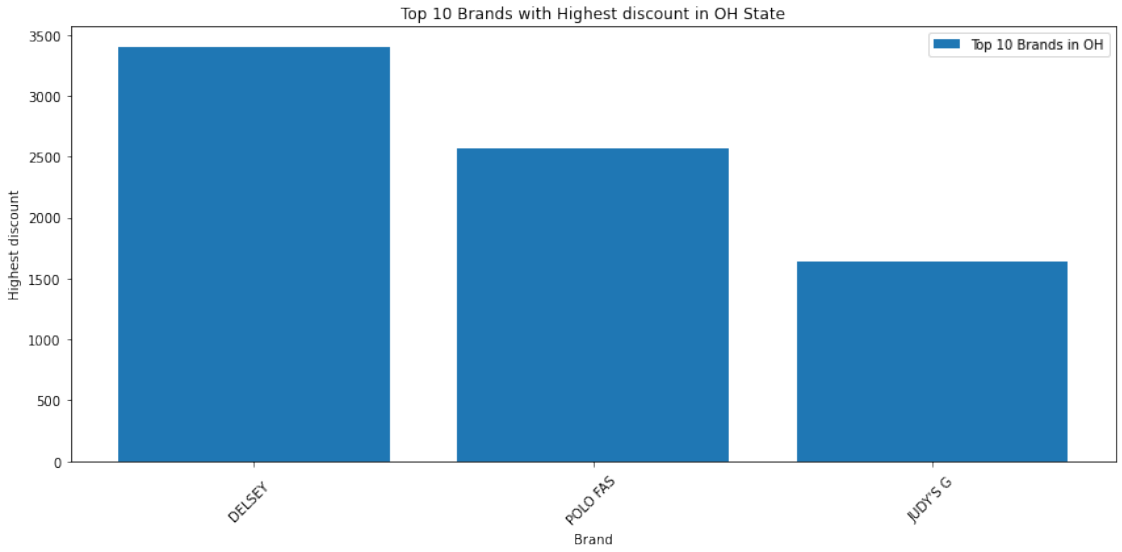

Top 10 Brands with Highest discount in MT State
Top 10 Brands in MT
Highest discount
0
500
1000
1500
2000
DELSEY
POLO FAS
EMMA JAM
SIGRID O
LIZ CLAI
HEART AN
H.H. BRO
Brand

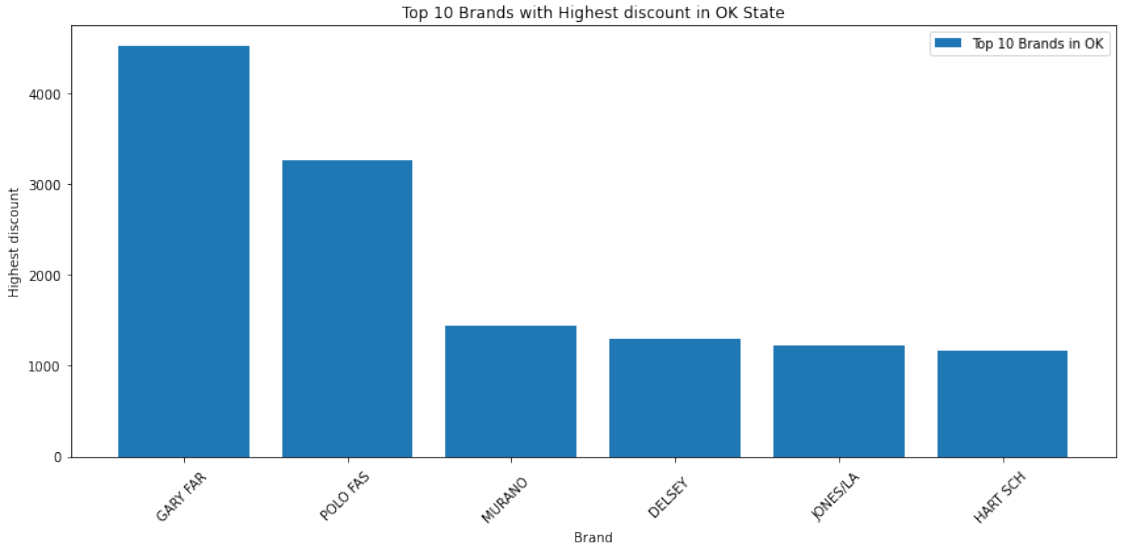

Top 10 Brands with Highest discount in NC State
Top 10 Brands in NC
Highest discount
3000
2500
2000
1500
1000
500
0
POLO FAS
JONES/LA
HART SCH
LIZ CLAI
Brand

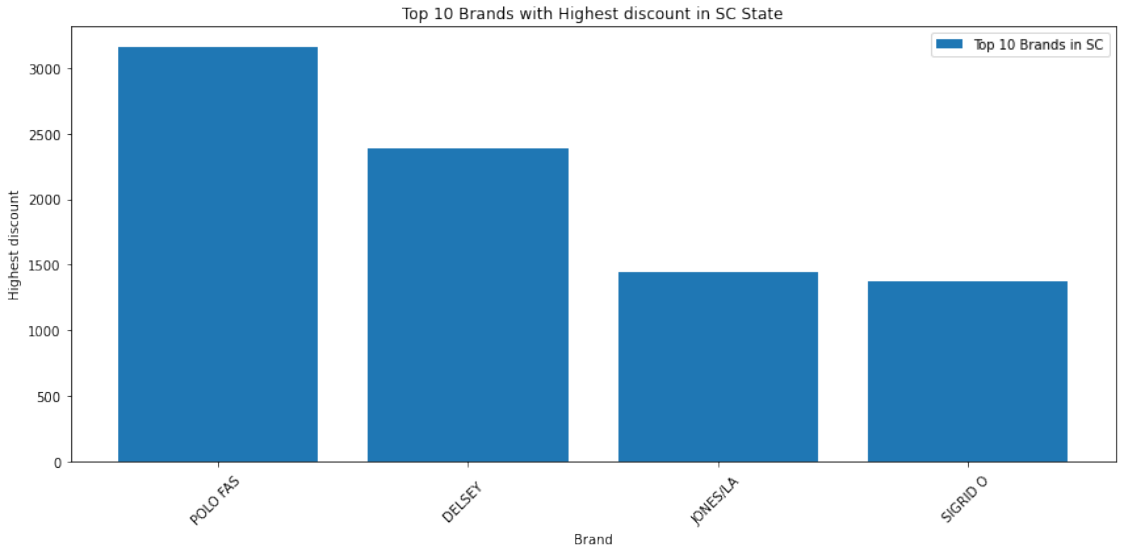

Top 10 Brands with Highest discount in NE State
Top 10 Brands in NE
Highest discount
2000
1500
1000
500
0
POLO FAS
LIZ CLAI
EMMA JAM
SIGRID O
HART SCH
Brand

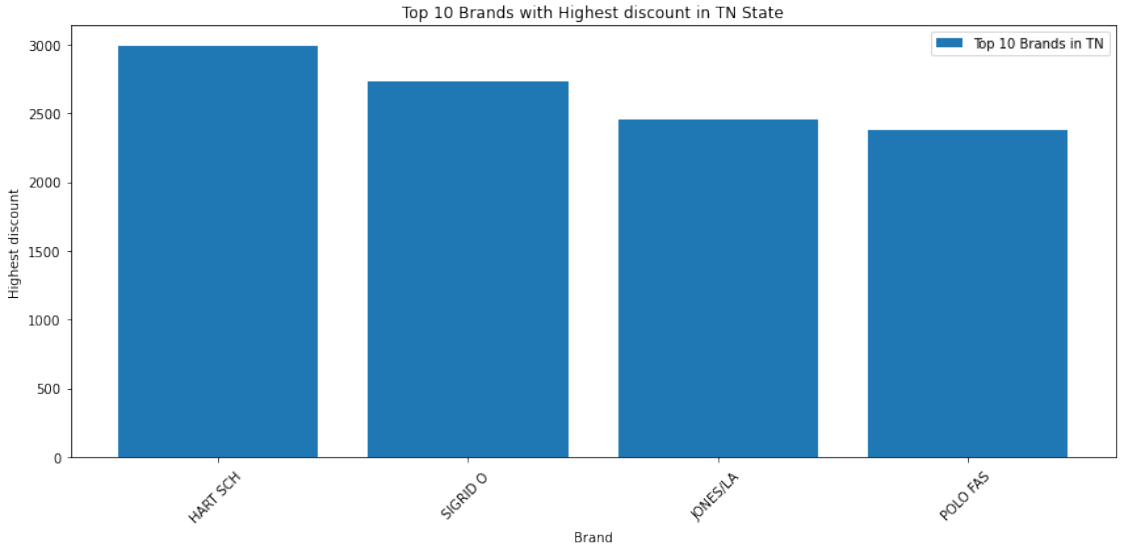

Top 10 Brands with Highest discount in NM State
Top 10 Brands in NM
Highest discount
0
1000
2000
3000
4000
POLO FAS
A B S CL
MURANO
DELSEY
SIGRID O
LIZ CLAI
ROUNDTRE
HART SCH
Brand

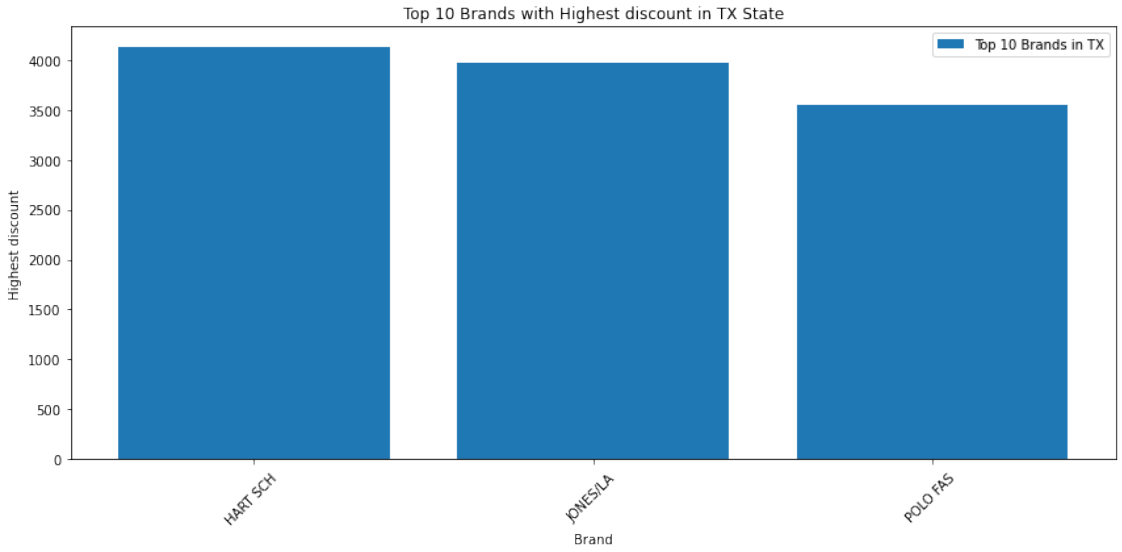

Top 10 Brands with Highest discount in NV State
Top 10 Brands in NV
Highest discount
0
1000
2000
3000
4000
5000
POLO FAS
LIZ CLAI
JONES/LA
DELSEY
HART SCH
MURANO
Brand

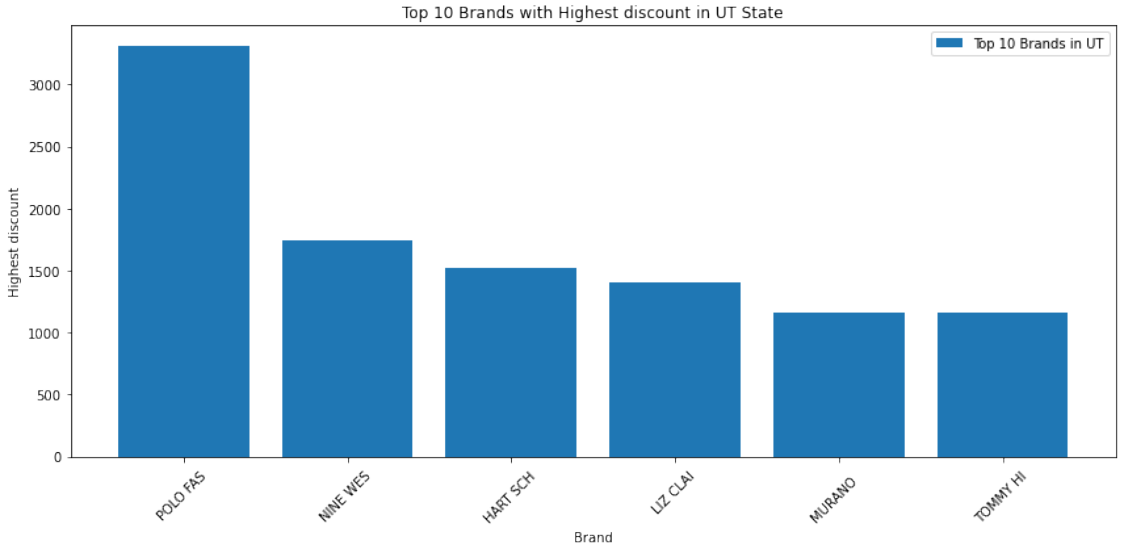

Top 10 Brands with Highest discount in OH State
Top 10 Brands in OH
Highest discount
3500
3000
2500
2000
1500
1000
500
0
DELSEY
POLO FAS
JUDY'S G
Brand

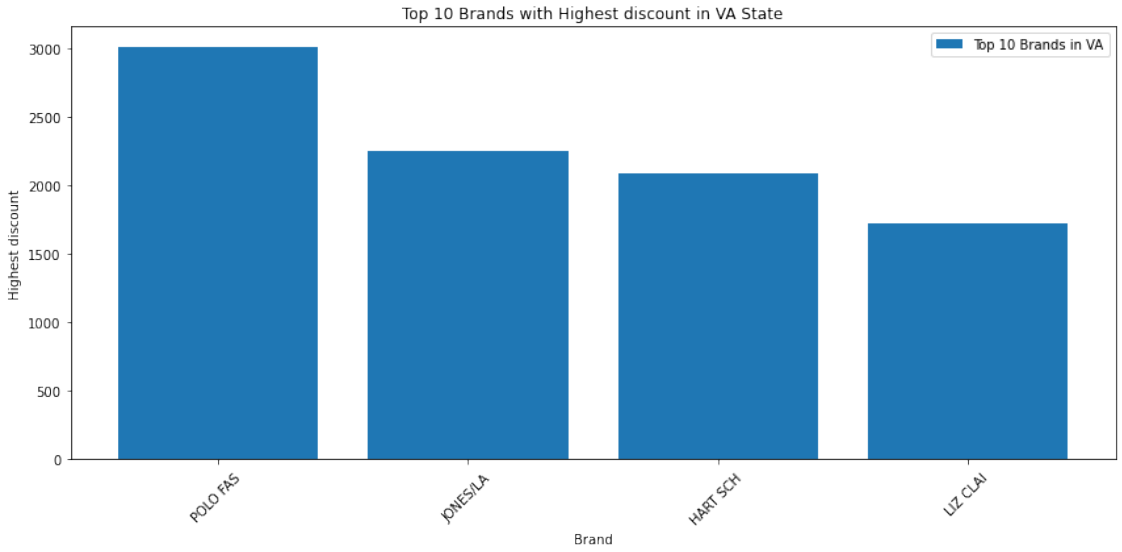

Top 10 Brands with Highest discount in OK State
Top 10 Brands in OK
Highest discount
4000
3000
2000
1000
0
GARY FAR
POLO FAS
MURANO
DELSEY
JONES/LA
HART SCH
Brand

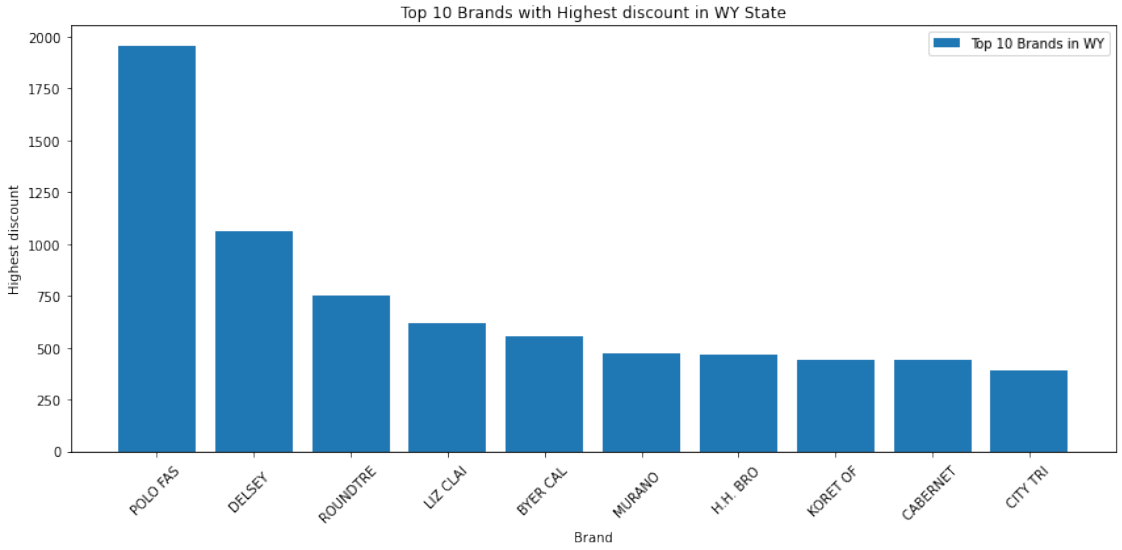

Top 10 Brands with Highest discount in SC State
Top 10 Brands in SC
Highest discount
3000
2500
2000
1500
1000
500
0
POLO FAS
DELSEY
JONES/LA
SIGRID O
Brand


Top 10 Brands with Highest discount in TN State
Top 10 Brands in TN
Highest discount
3000
2500
2000
1500
1000
500
0
HART SCH
SIGRID O
JONES/LA
POLO FAS
Brand


Top 10 Brands with Highest discount in TX State
Top 10 Brands in TX
Highest discount
0
500
1000
1500
2000
2500
3000
3500
4000
HART SCH
JONES/LA
POLO FAS
Brand


Top 10 Brands with Highest discount in UT State
Top 10 Brands in UT
Highest discount
0
500
1000
1500
2000
2500
3000
POLO FAS
NINE WES
HART SCH
LIZ CLAI
MURANO
TOMMY HI
Brand





[123]:
```
# Create a dictionary to store the top 10 brands with the highest discount for␣
↪each state
state_top_brands = {}

for state in states:
    state_data =␣
↪highest_profit_per_brand_state_store[highest_profit_per_brand_state_store['STATE']␣
↪== state]
    top_brands = list(state_data.nlargest(10, 'discount')['BRAND'])
    state_top_brands[state] = top_brands
```

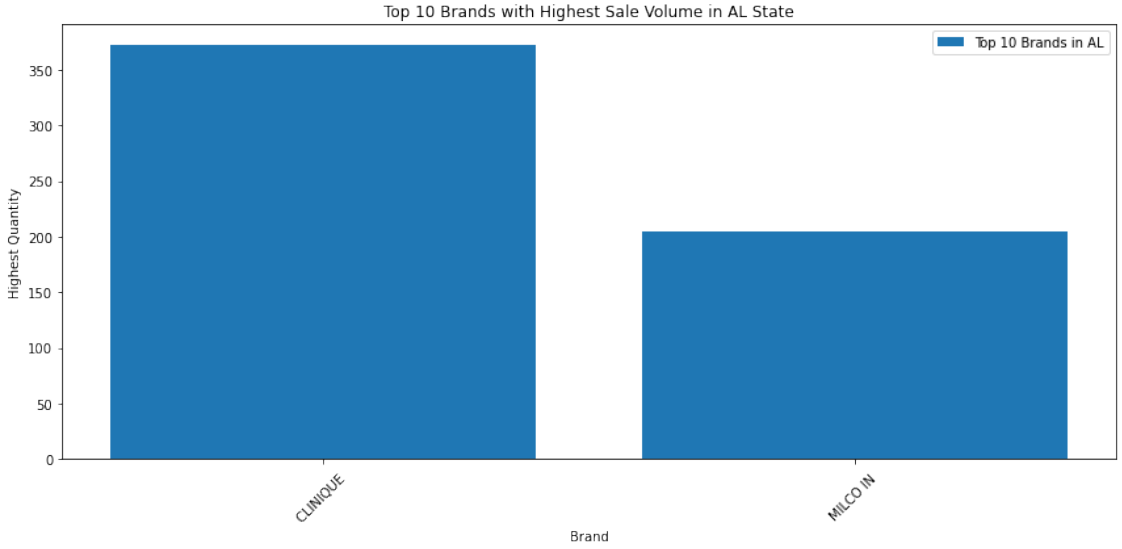
```
# Find the common brands among all states
common_brands = set(state_top_brands[states[0]])  # Initialize with the brands␣
↪from the first state

# Iterate through the states and find the common brands
for state in states:
    common_brands = common_brands.intersection(state_top_brands[state])

# Count the occurrences of each brand in the common brands set
brand_counts = Counter(brand for state in states for brand in␣
↪state_top_brands[state])

# Find the most common brands
most_common_brands = [brand for brand, _ in brand_counts.most_common(5)]

# Print the first five most common brands
print("Most Common Brands That Have High Discount:")
for rank, brand in enumerate(most_common_brands, start=1):
    print(f"{rank} - {brand}")
```

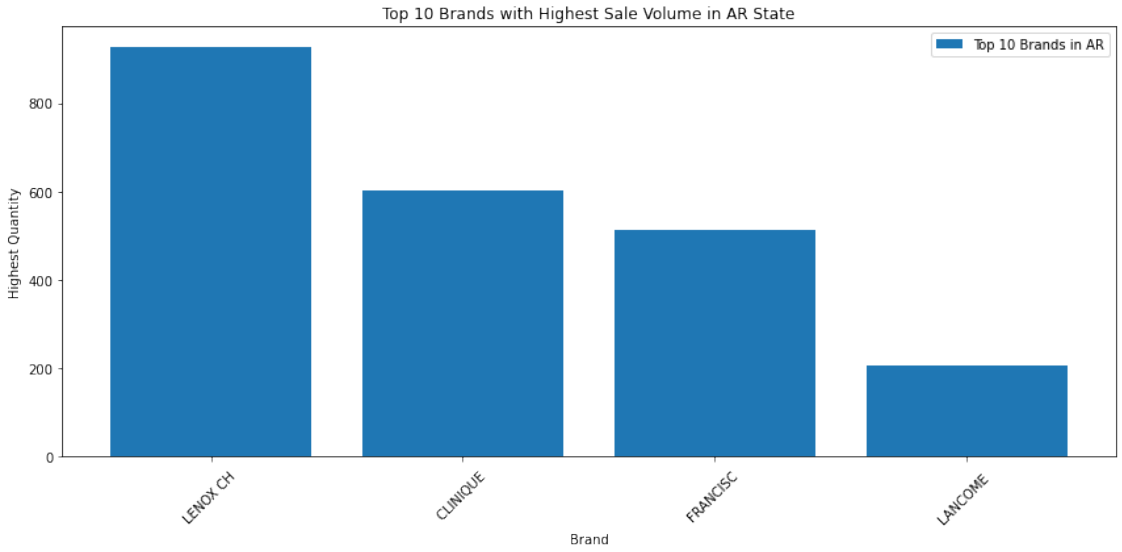
```
Most Common Brands That Have High Discount:
1 - POLO FAS
2 - LIZ CLAI
3 - DELSEY
4 - HART SCH
5 - JONES/LA
```

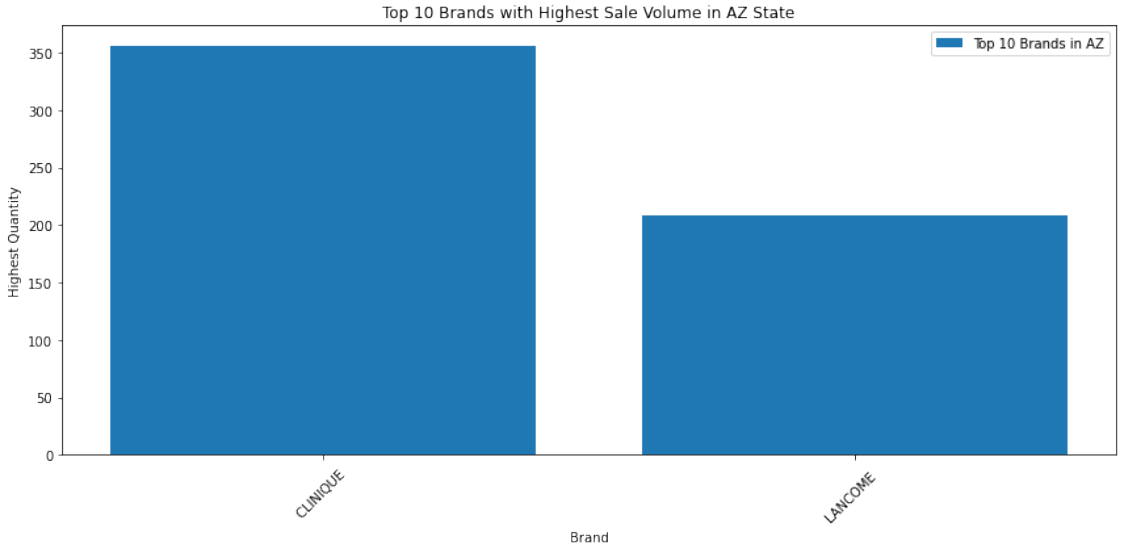
In AL, POLO FAS, LIZ CLAI, HART SCH, and JONES/LA are the brands that have highest discount. In AR, LENOX CH, MURANO, DELSEY, HART SCH, JONES/LA, and POLO FAS have high discounts compare to other brands. In AZ, POLO FAS, SIGRID O, LIZ CLAI, and HART SCH have high discounts. In CA, POLO FAS, JONES/LA, LIZ CLAI, TOMMY HI, NINE WES, ROUNDTRE, SIGRID O, and MURANO have high discounts. … The first five most common brands that have high discount are POLO FAS, LIZ CLAI, DELSEY, HART SCH, and JONES/LA.

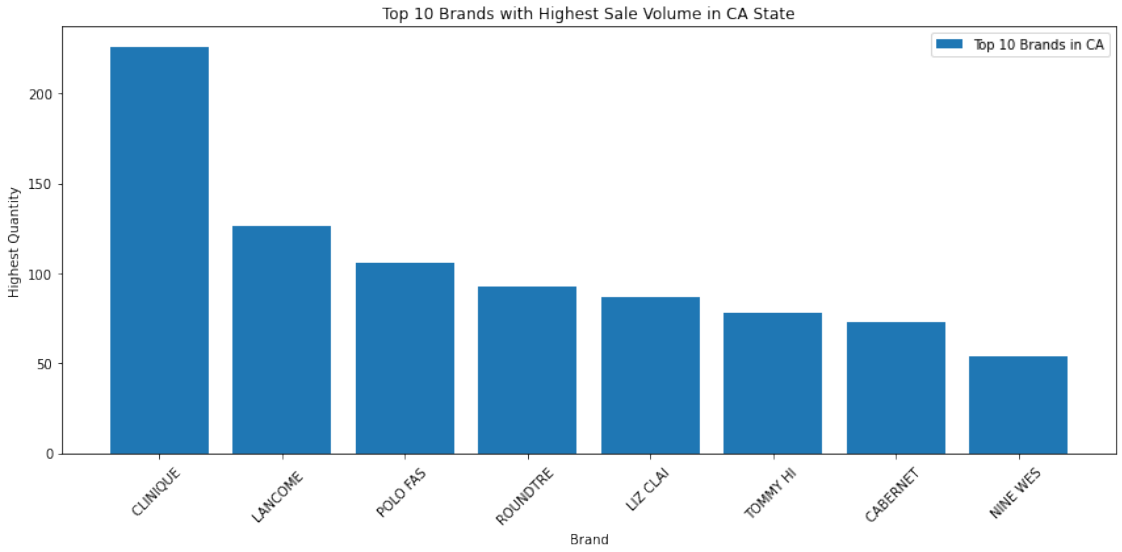
[124]:
```
# Top brands with highest sale volume in each state

# Create a bar chart for each state
for state in states:
    state_data =␣
↪highest_profit_per_brand_state_store[highest_profit_per_brand_state_store['STATE']␣
↪== state]

    # Select the top 10 brands with the highest profit within the state
    top_brands = state_data.nlargest(10, 'QUANTITY')

    # Create a figure and axis
    fig, ax = plt.subplots(figsize=(12, 6))
```

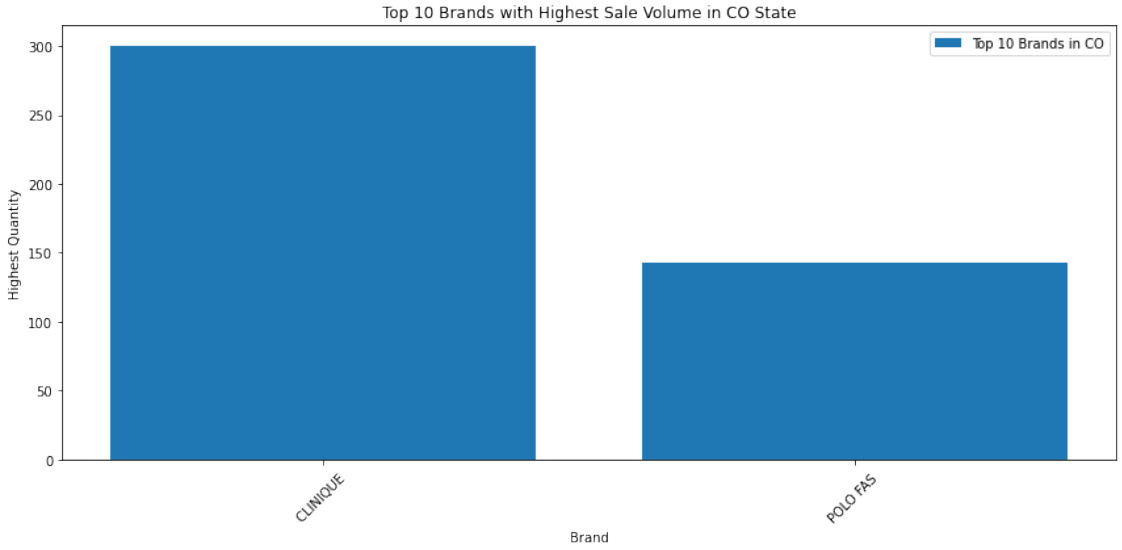
```
    # Plot the highest profit for each of the top 10 brands within the state
    ax.bar(top_brands['BRAND'], top_brands['QUANTITY'], label=f'Top {10} Brands␣
↪in {state}')

    # Set labels and title
    ax.set_xlabel('Brand')
    ax.set_ylabel('Highest Quantity')
    ax.set_title(f'Top {10} Brands with Highest Sale Volume in {state} State')

    # Rotate x-axis labels for better readability
    plt.xticks(rotation=45)

    # Add a legend
    ax.legend()

    # Show the plot for each state
    plt.tight_layout()
    plt.show()
```

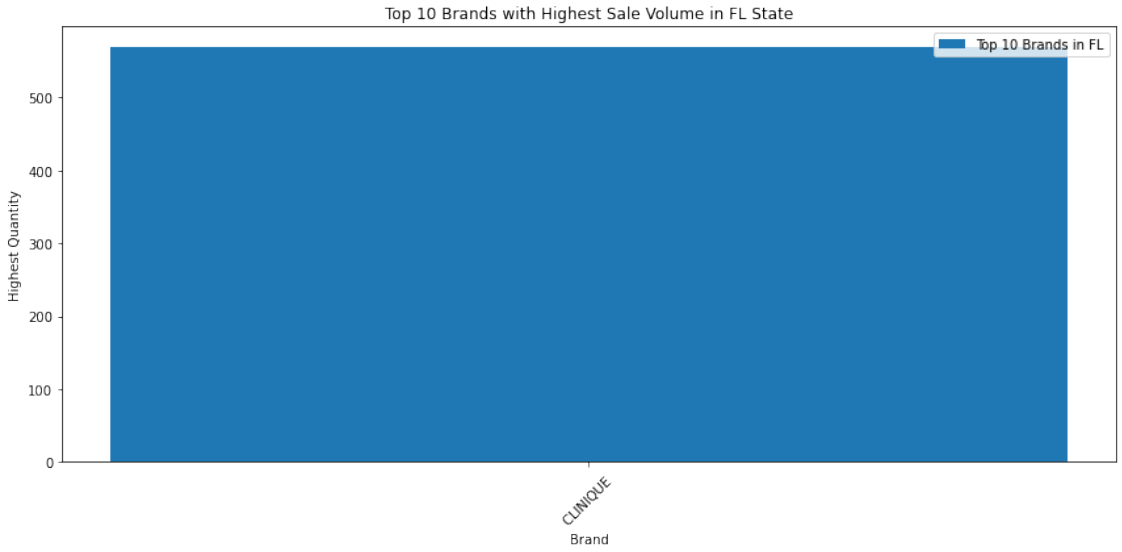


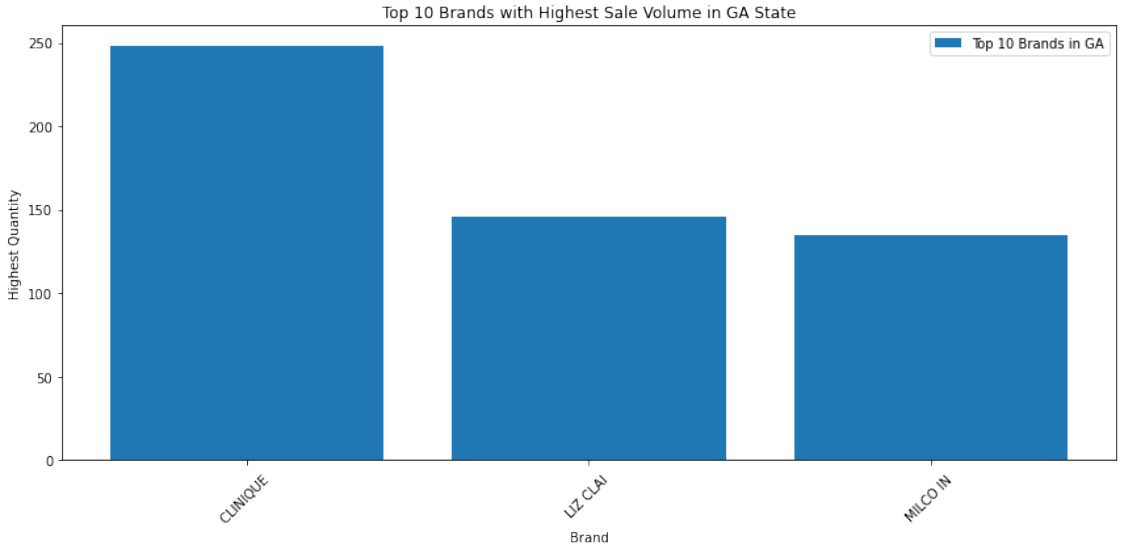

Top 10 Brands with Highest Sale Volume in AR State
Top 10 Brands in AR
Highest Quantity
0
200
400
600
800
LENOX CH
CLINIQUE
FRANCISC
LANCOME
Brand

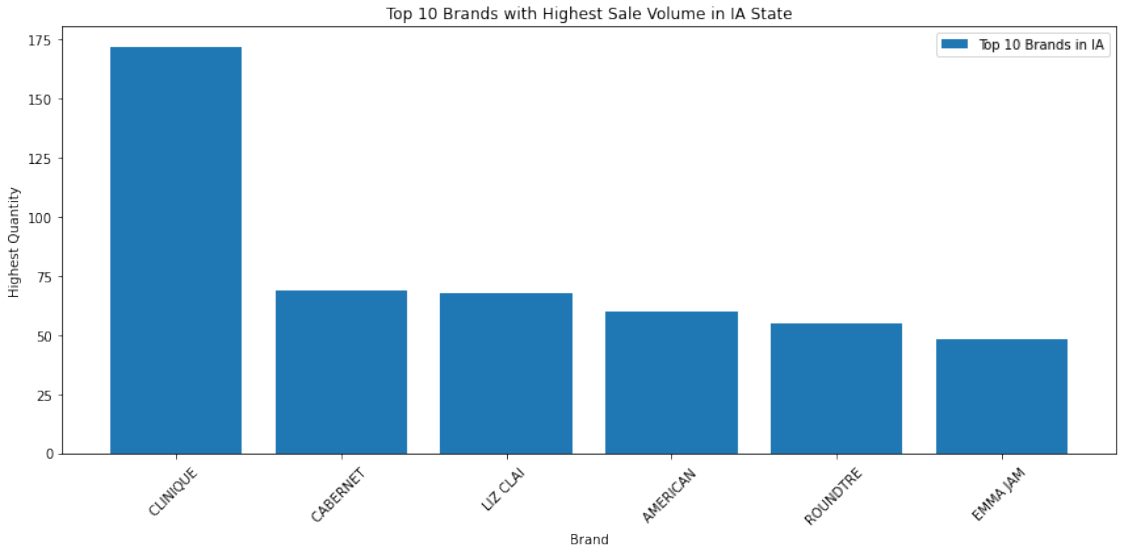

Top 10 Brands with Highest Sale Volume in AZ State
Top 10 Brands in AZ
Highest Quantity
0
50
100
150
200
250
300
350
CLINIQUE
LANCOME
Brand

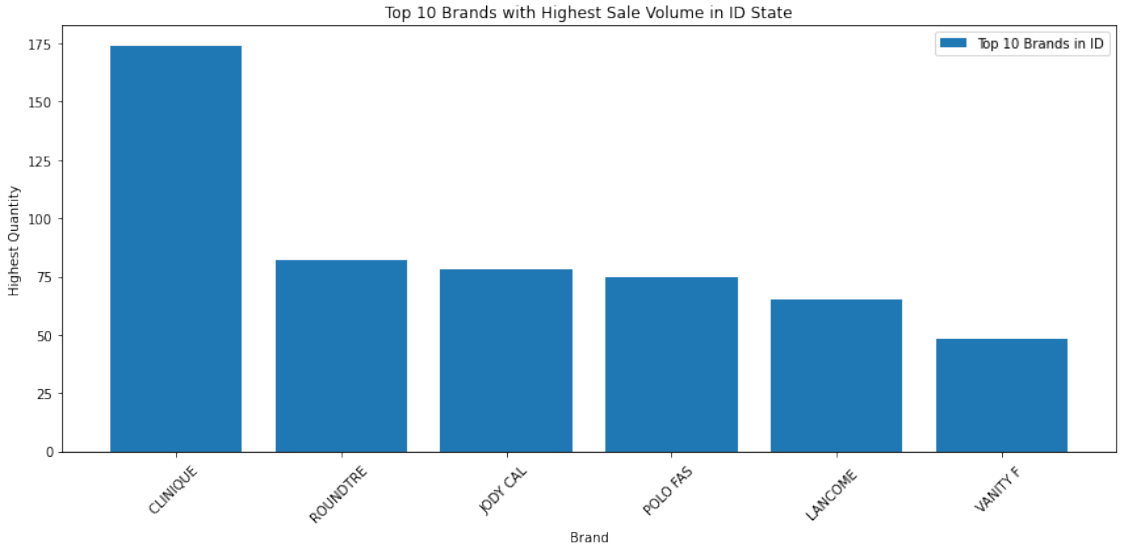

Top 10 Brands with Highest Sale Volume in CA State
Top 10 Brands in CA
Highest Quantity
0
50
100
150
200
CLINIQUE
LANCOME
POLO FAS
ROUNDTRE
LIZ CLAI
TOMMY HI
CABERNET
NINE WES
Brand

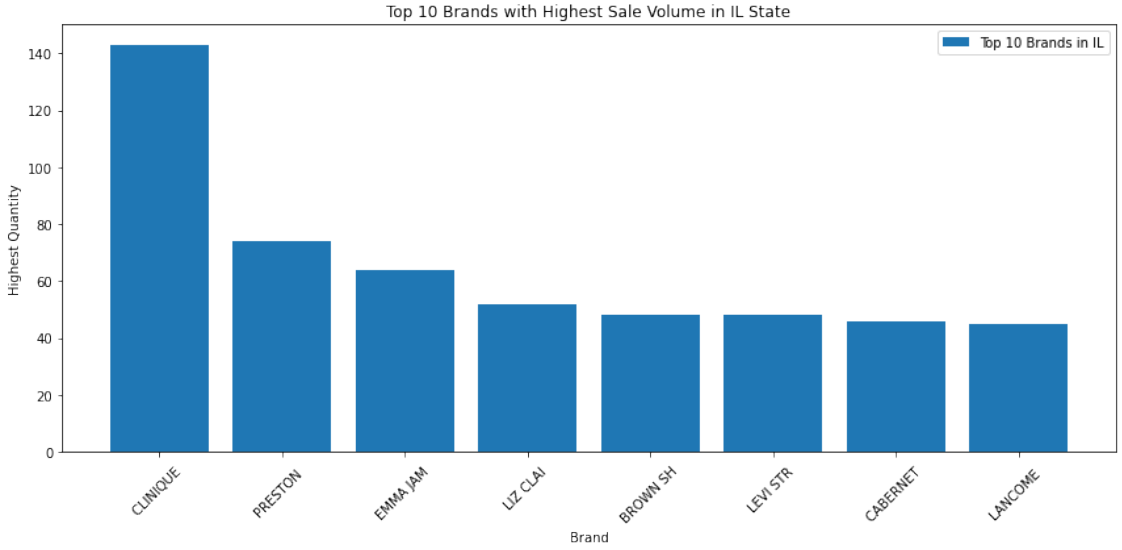

Top 10 Brands with Highest Sale Volume in CO State
Top 10 Brands in CO
Highest Quantity
0
50
100
150
200
250
300
CLINIQUE
POLO FAS
Brand

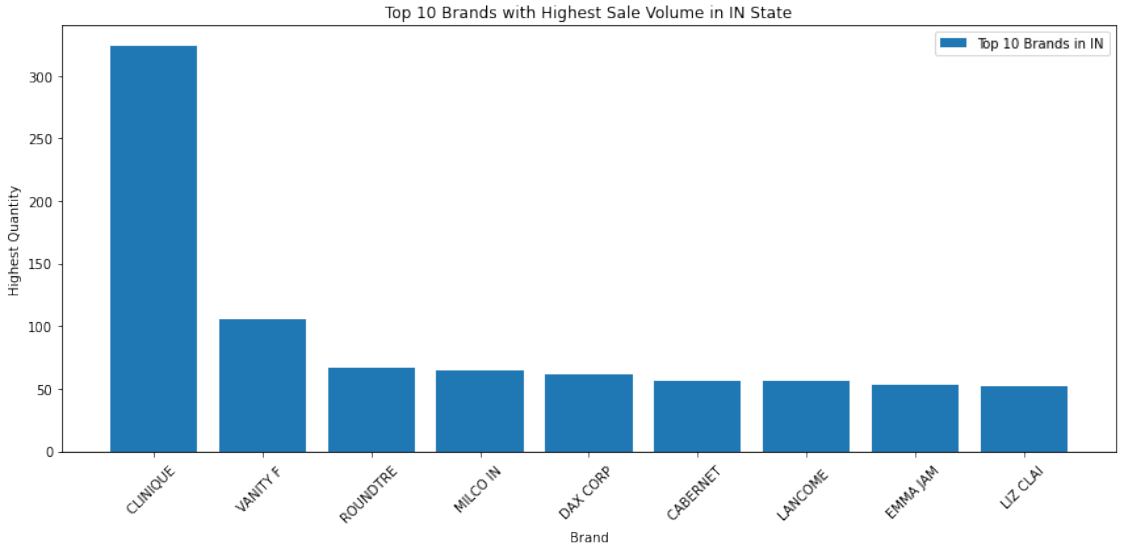

Top 10 Brands with Highest Sale Volume in FL State
Top 10 Brands in FL
Highest Quantity
0
100
200
300
400
500
CLINIQUE
Brand

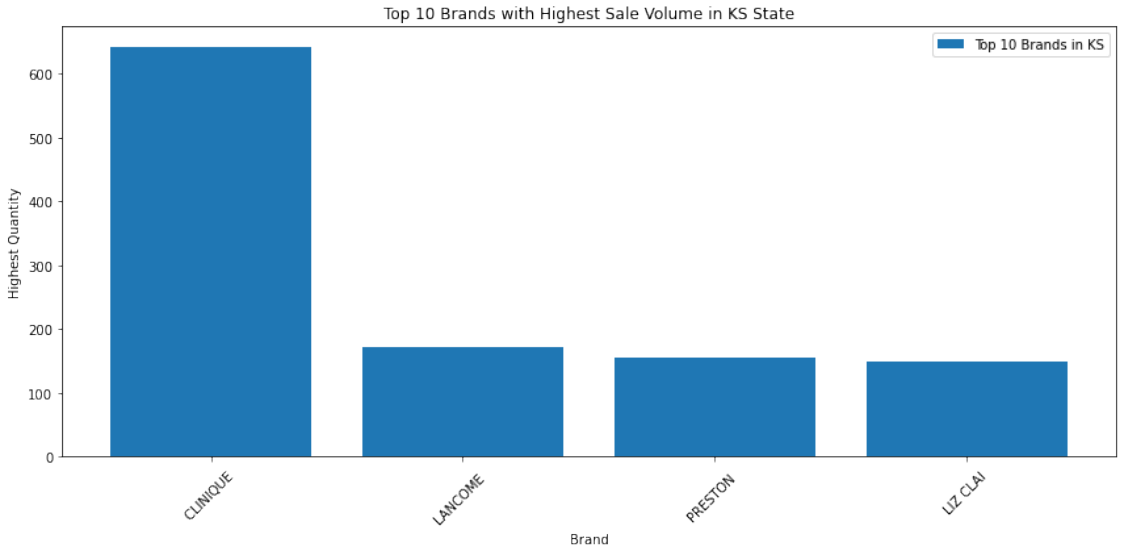

Top 10 Brands with Highest Sale Volume in GA State
Top 10 Brands in GA
Highest Quantity
0
50
100
150
200
250
CLINIQUE
LIZ CLAI
MILCO IN
Brand

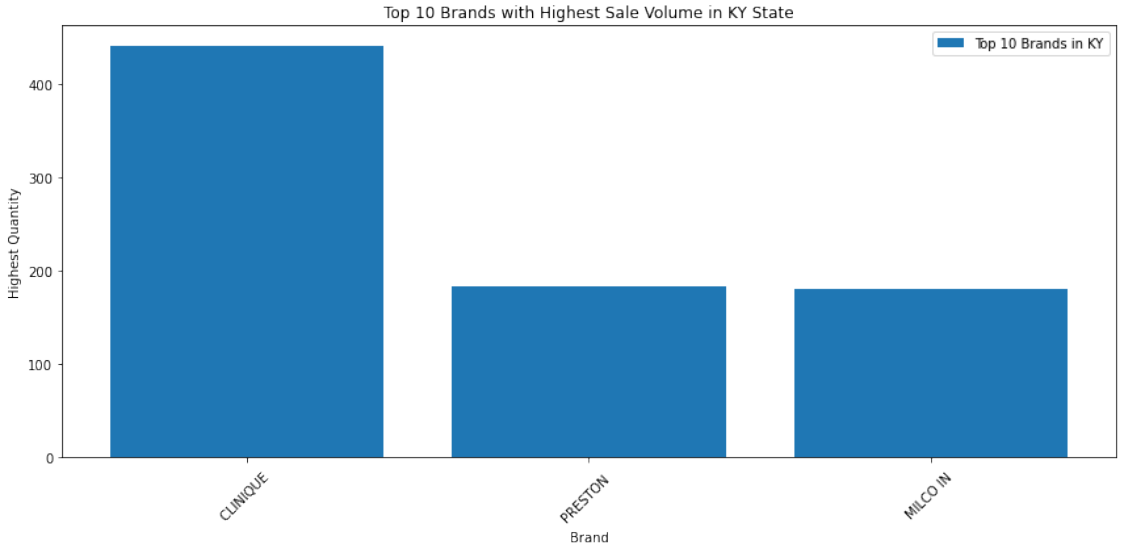

Top 10 Brands with Highest Sale Volume in IA State
Top 10 Brands in IA
Highest Quantity
0
25
50
75
100
125
150
175
CLINIQUE
CABERNET
LIZ CLAI
AMERICAN
ROUNDTRE
EMMA JAM
Brand

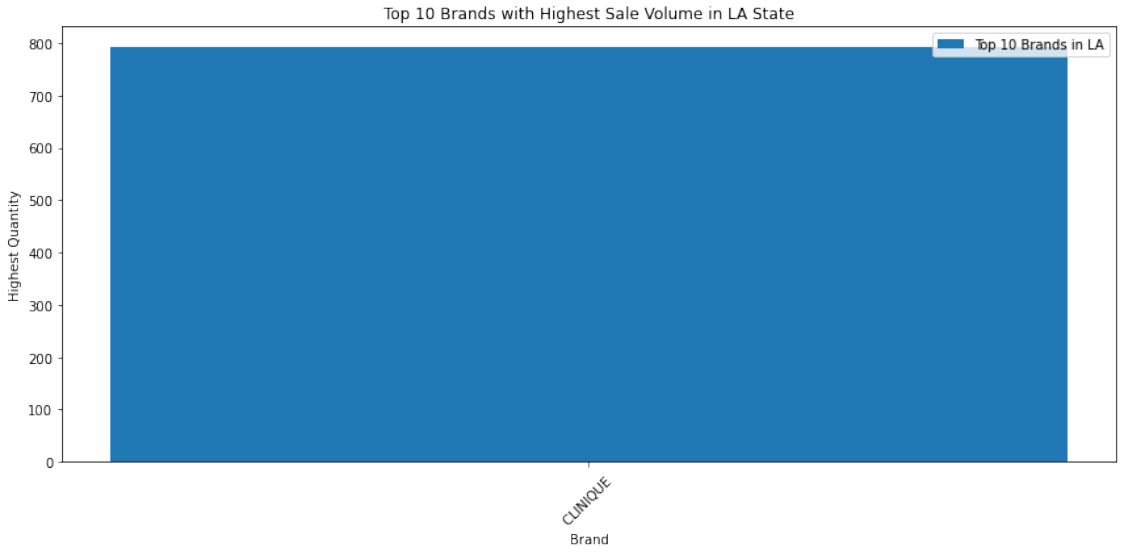

Top 10 Brands with Highest Sale Volume in ID State
Top 10 Brands in ID
Highest Quantity
0
25
50
75
100
125
150
175
CLINIQUE
ROUNDTRE
JODY CAL
POLO FAS
LANCOME
VANITY F
Brand

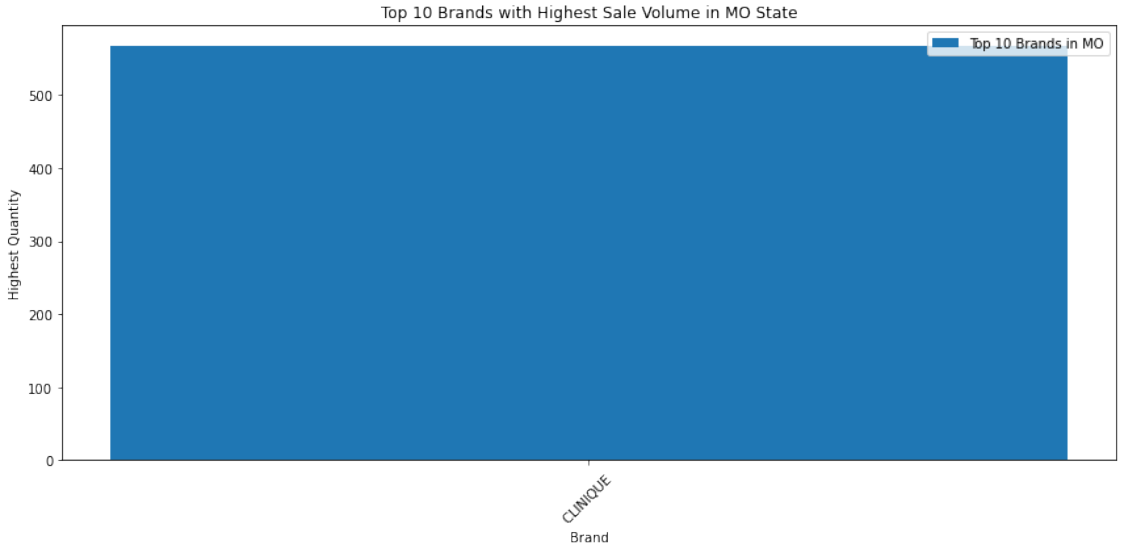

Top 10 Brands with Highest Sale Volume in IL State
Top 10 Brands in IL
Highest Quantity
0
20
40
60
80
100
120
140
CLINIQUE
PRESTON
EMMA JAM
LIZ CLAI
BROWN SH
LEVI STR
CABERNET
LANCOME
Brand

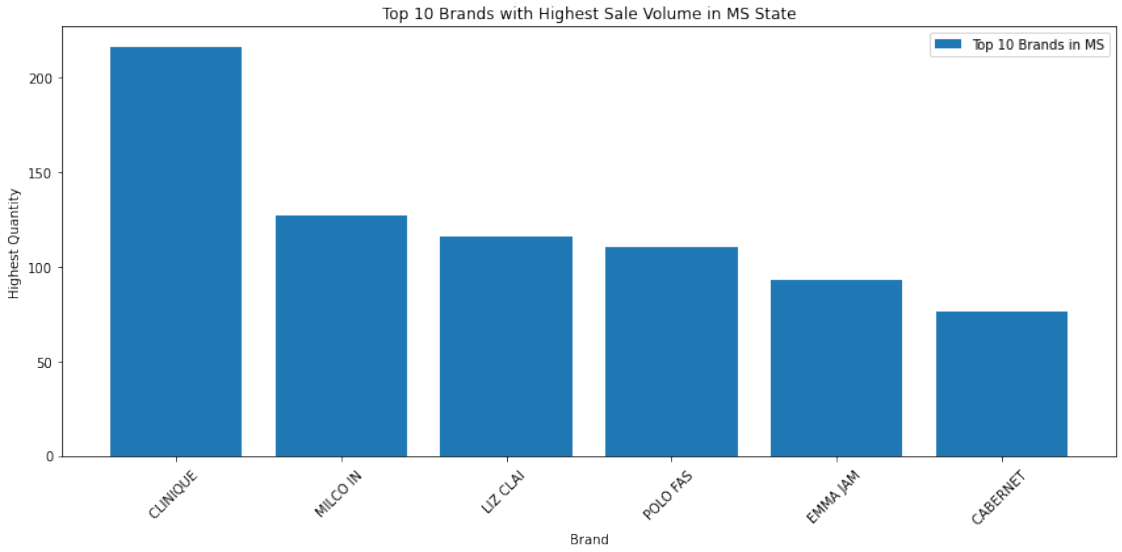

Top 10 Brands with Highest Sale Volume in IN State
Top 10 Brands in IN
Highest Quantity
0
50
100
150
200
250
300
CLINIQUE
VANITY F
ROUNDTRE
MILCO IN
DAX CORP
CABERNET
LANCOME
EMMA JAM
LIZ CLAI
Brand

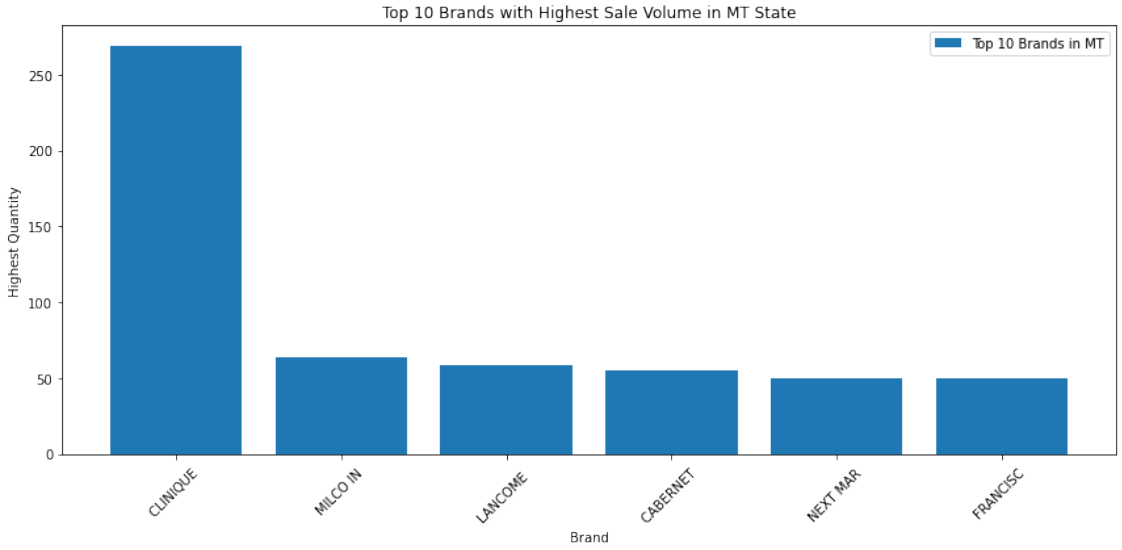

Top 10 Brands with Highest Sale Volume in KS State
Top 10 Brands in KS
Highest Quantity
0
100
200
300
400
500
600
CLINIQUE
LANCOME
PRESTON
LIZ CLAI
Brand

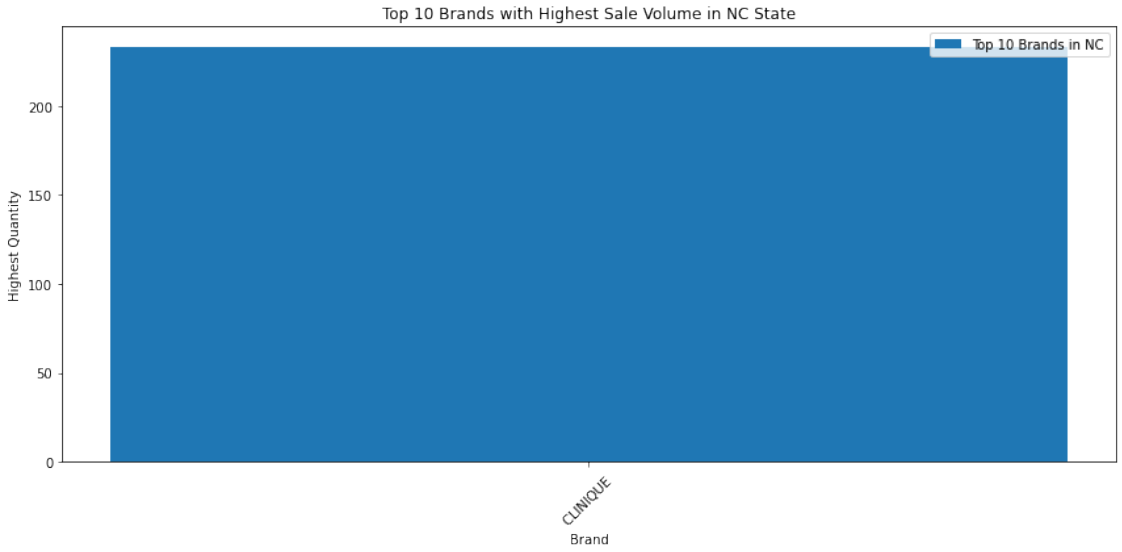

Top 10 Brands with Highest Sale Volume in KY State
Top 10 Brands in KY
Highest Quantity
0
100
200
300
400
CLINIQUE
PRESTON
MILCO IN
Brand

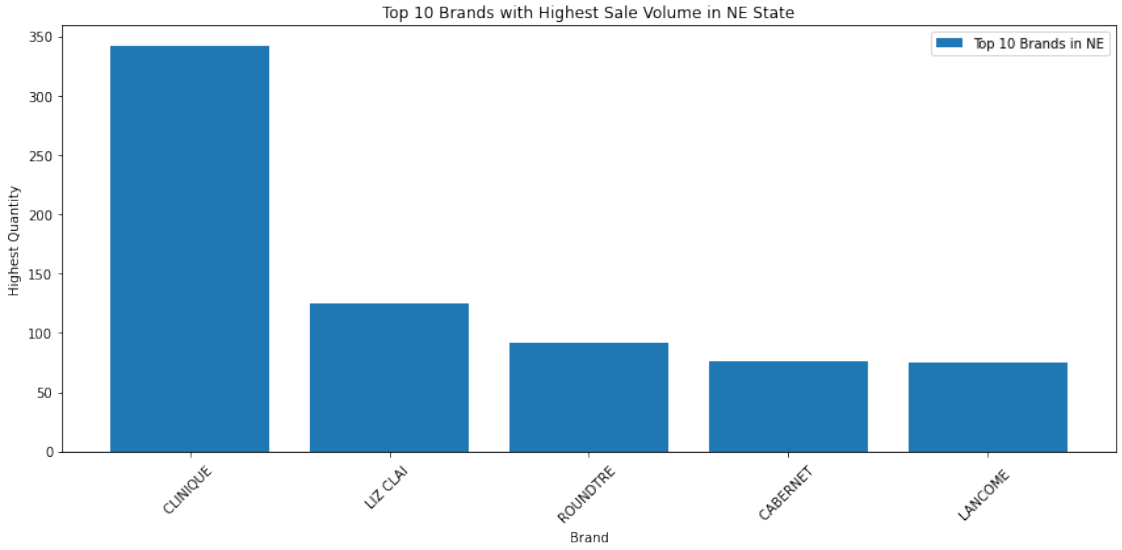

Top 10 Brands with Highest Sale Volume in LA State
Top 10 Brands in LA
Highest Quantity
800
700
600
500
400
300
200
100
0
CLINIQUE
Brand

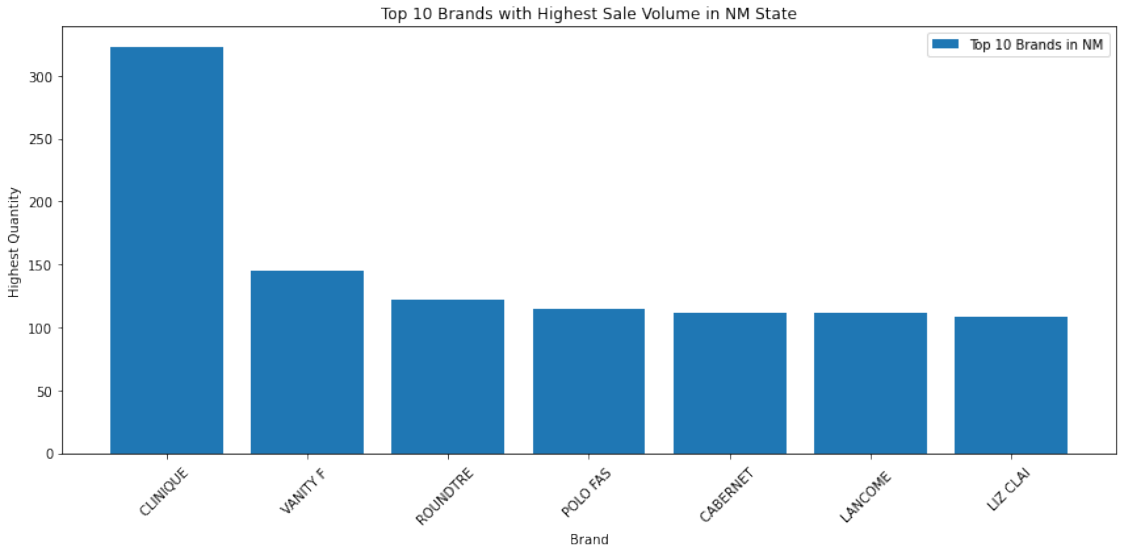

Top 10 Brands with Highest Sale Volume in MO State
Top 10 Brands in MO
Highest Quantity
500
400
300
200
100
0
CLINIQUE
Brand

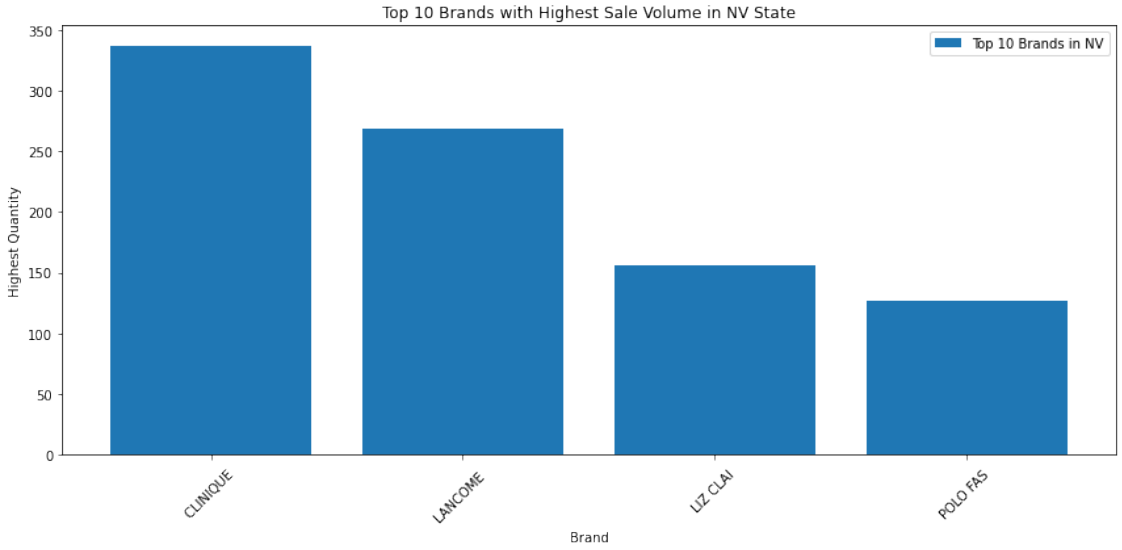

Top 10 Brands with Highest Sale Volume in MS State
Top 10 Brands in MS
Highest Quantity
0
50
100
150
200
CLINIQUE
MILCO IN
LIZ CLAI
POLO FAS
EMMA JAM
CABERNET
Brand

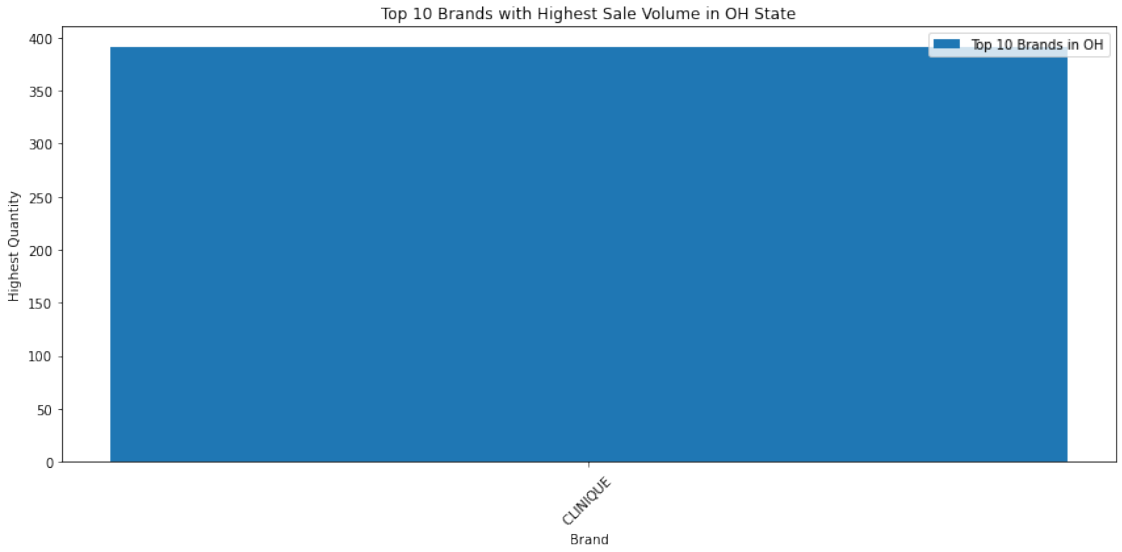

Top 10 Brands with Highest Sale Volume in MT State
Top 10 Brands in MT
Highest Quantity
0
50
100
150
200
250
CLINIQUE
MILCO IN
LANCOME
CABERNET
NEXT MAR
FRANCISC
Brand

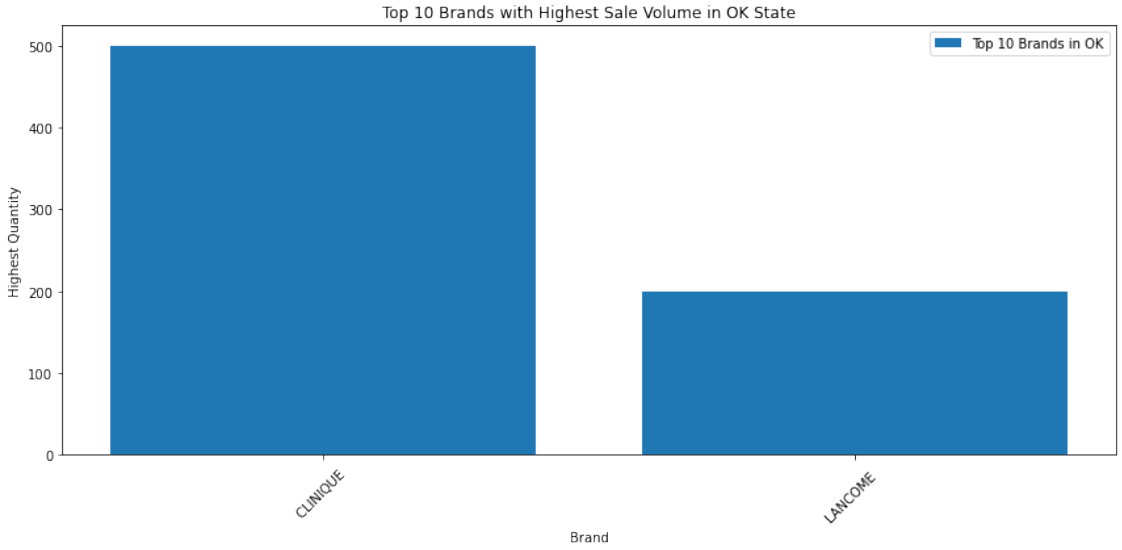

Top 10 Brands with Highest Sale Volume in NC State
Top 10 Brands in NC
Highest Quantity
0
50
100
150
200
CLINIQUE
Brand

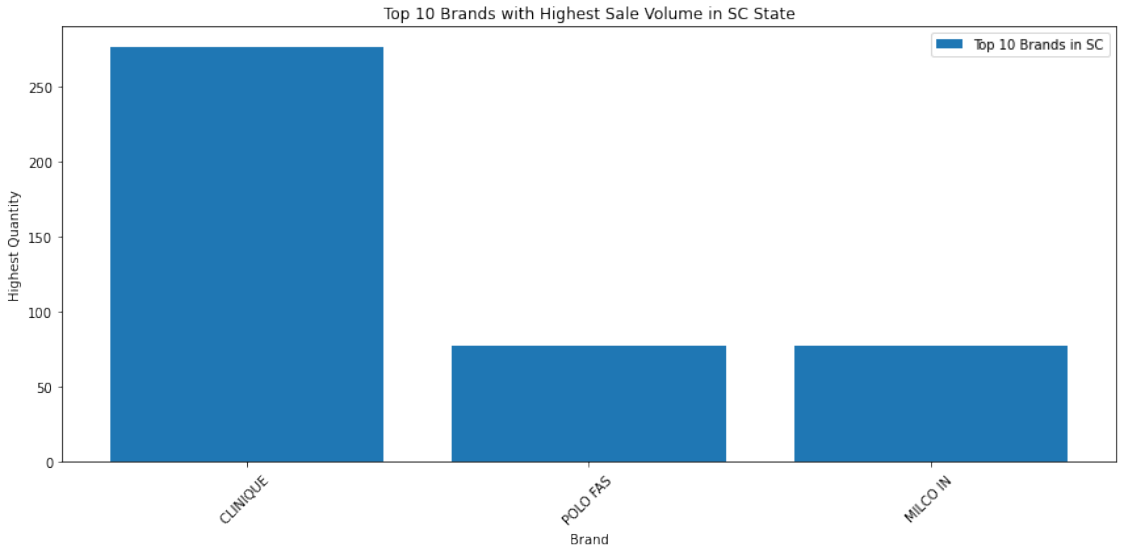

Top 10 Brands with Highest Sale Volume in NE State
Top 10 Brands in NE
Highest Quantity
0
50
100
150
200
250
300
350
CLINIQUE
LIZ CLAI
ROUNDTRE
CABERNET
LANCOME
Brand

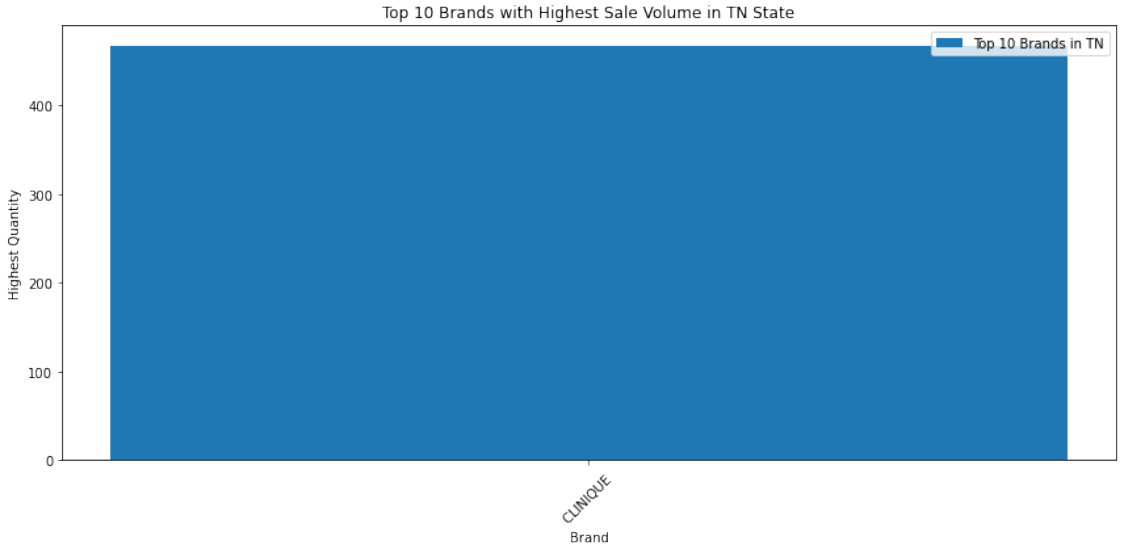

Top 10 Brands with Highest Sale Volume in NM State
Top 10 Brands in NM
Highest Quantity
0
50
100
150
200
250
300
CLINIQUE
VANITY F
ROUNDTRE
POLO FAS
CABERNET
LANCOME
LIZ CLAI
Brand

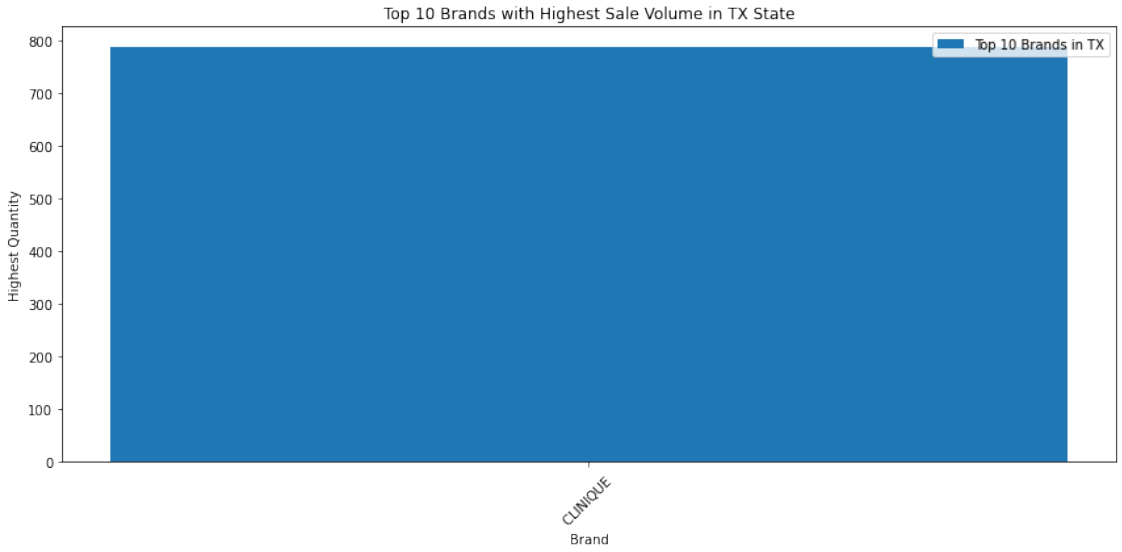

Top 10 Brands with Highest Sale Volume in NV State
Top 10 Brands in NV
Highest Quantity
0
50
100
150
200
250
300
350
CLINIQUE
LANCOME
LIZ CLAI
POLO FAS
Brand

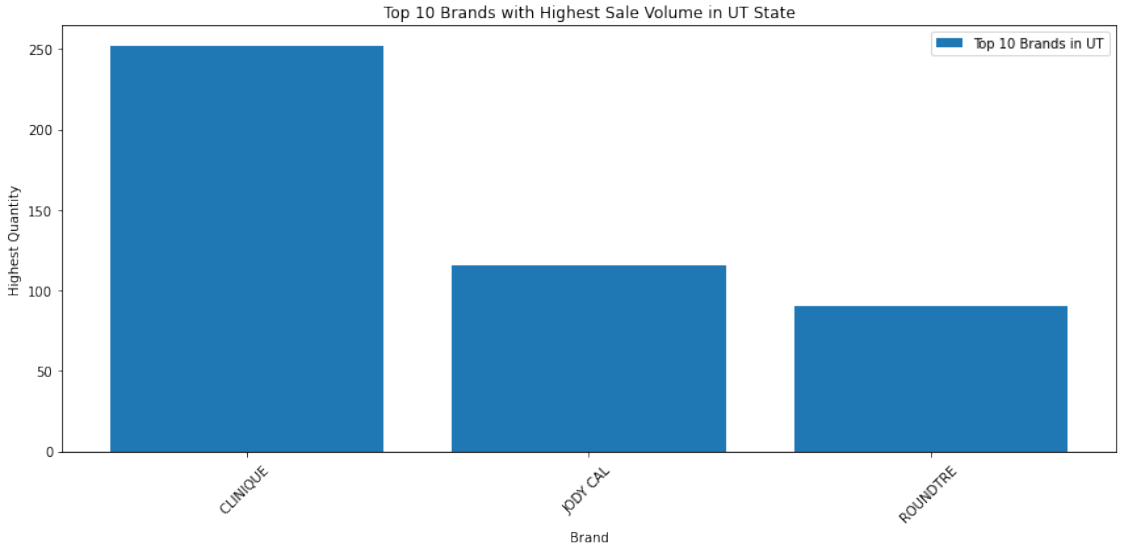

Top 10 Brands with Highest Sale Volume in OH State
Top 10 Brands in OH
Highest Quantity
0
50
100
150
200
250
300
350
400
CLINIQUE
Brand

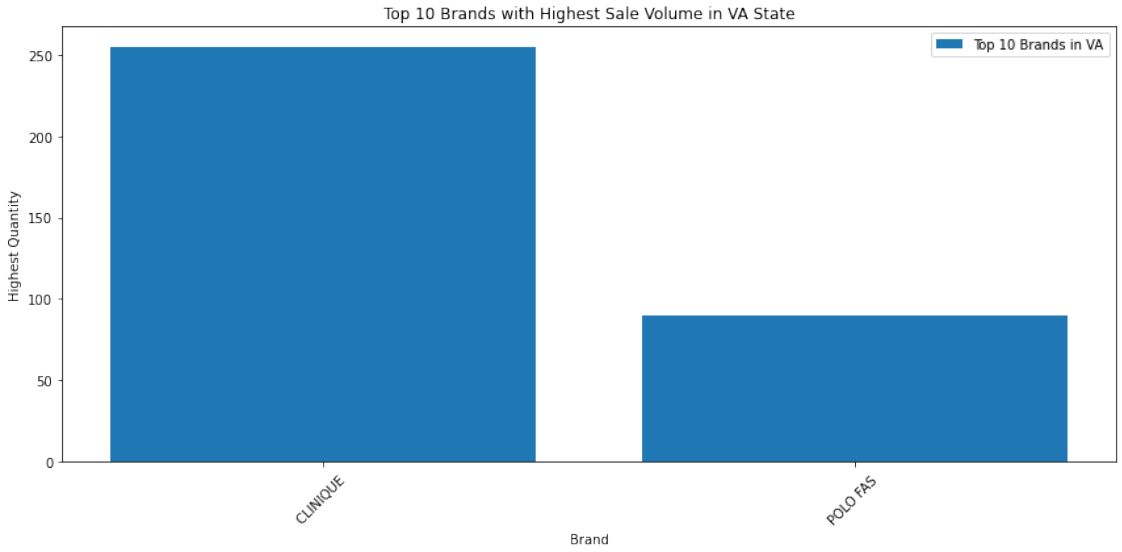

Top 10 Brands with Highest Sale Volume in OK State
Top 10 Brands in OK
Highest Quantity
0
100
200
300
400
500
CLINIQUE
LANCOME
Brand

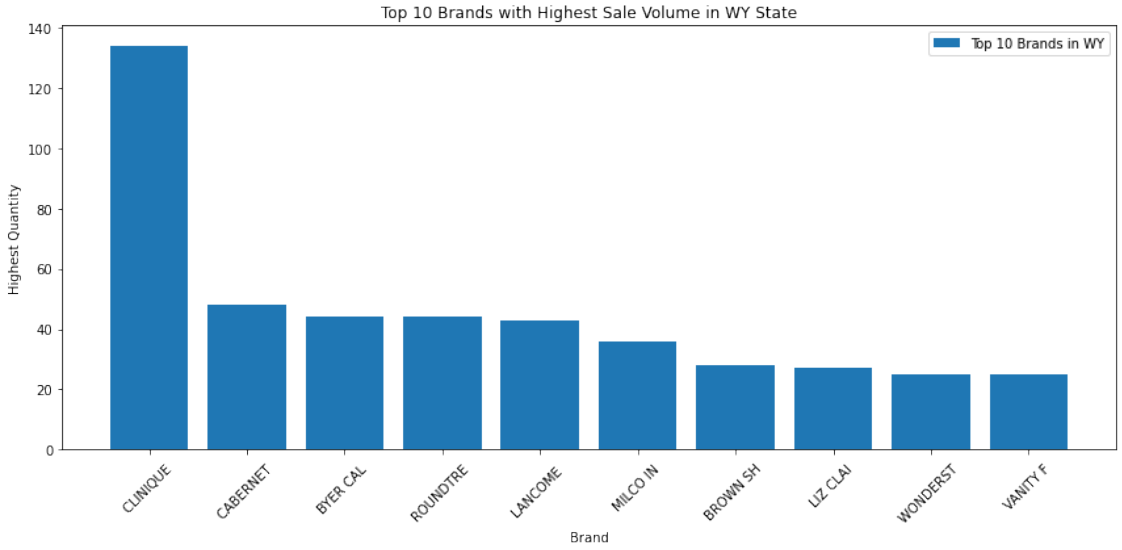

Top 10 Brands with Highest Sale Volume in SC State
Top 10 Brands in SC
Highest Quantity
0
50
100
150
200
250
CLINIQUE
POLO FAS
MILCO IN
Brand


Top 10 Brands with Highest Sale Volume in TN State
Top 10 Brands in TN
Highest Quantity
0
100
200
300
400
CLINIQUE
Brand


Top 10 Brands with Highest Sale Volume in TX State
Top 10 Brands in TX
Highest Quantity
800
700
600
500
400
300
200
100
0
CLINIQUE
Brand


Top 10 Brands with Highest Sale Volume in UT State
Top 10 Brands in UT
Highest Quantity
250
200
150
100
50
0
CLINIQUE
JODY CAL
ROUNDTRE
Brand





[125]:
```
# Create a dictionary to store the top 20 brands with the highest sale volume␣
↪for each state
state_top_brands = {}

for state in states:
    state_data =␣
↪highest_profit_per_brand_state_store[highest_profit_per_brand_state_store['STATE']␣
↪== state]
    top_brands = list(state_data.nlargest(10, 'QUANTITY')['BRAND'])
    state_top_brands[state] = top_brands
```

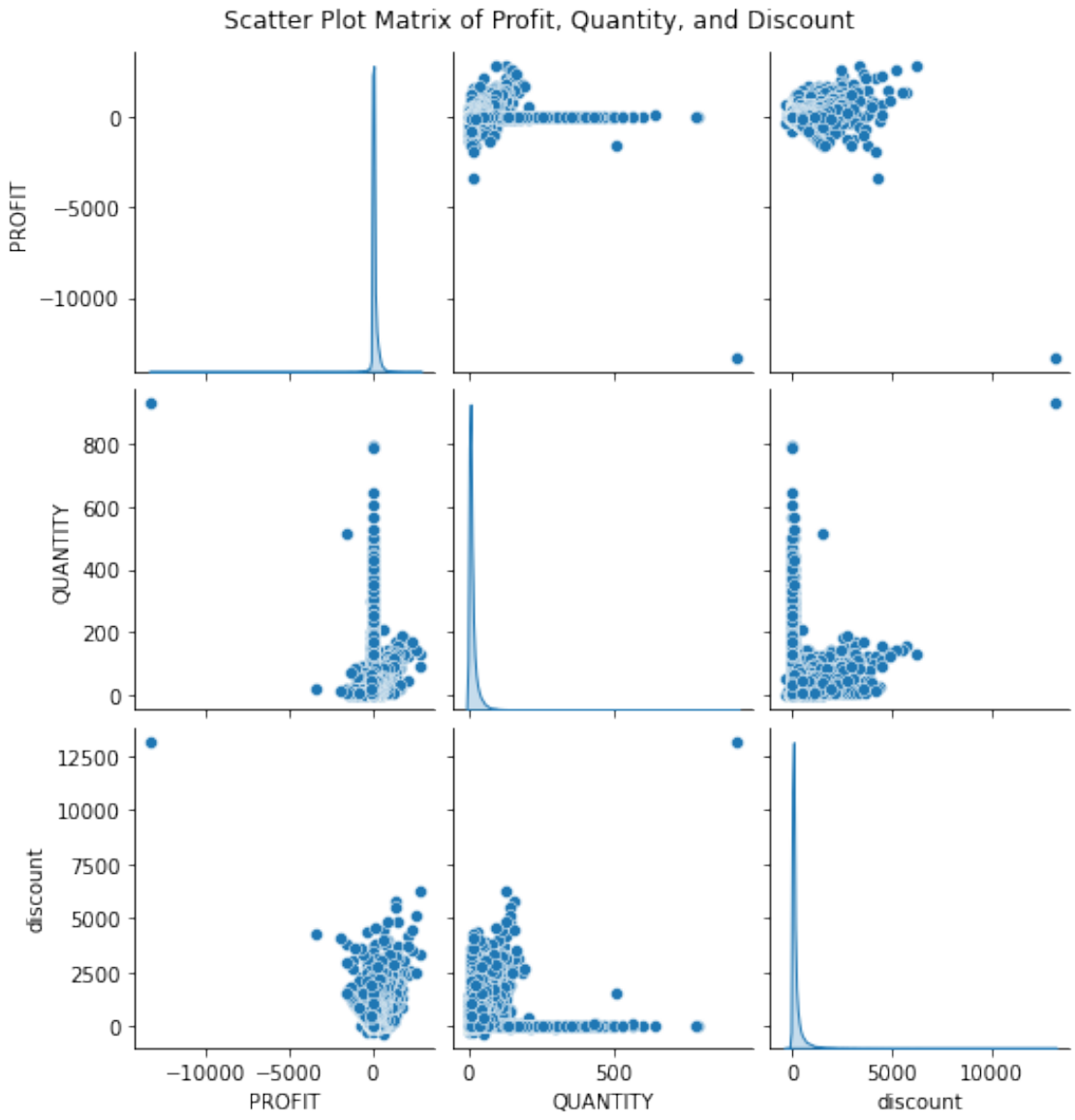
```
# Find the common brands among all states
common_brands = set(state_top_brands[states[0]])  # Initialize with the brands␣
↪from the first state

# Iterate through the states and find the common brands
for state in states:
    common_brands = common_brands.intersection(state_top_brands[state])

# Count the occurrences of each brand in the common brands set
brand_counts = Counter(brand for state in states for brand in␣
↪state_top_brands[state])

# Find the most common brands
most_common_brands = [brand for brand, _ in brand_counts.most_common(5)]

# Print the first five most common brands
print("Most Common Brands That Have High Sale Volume:")
for rank, brand in enumerate(most_common_brands, start=1):
    print(f"{rank} - {brand}")
```

```
Most Common Brands That Have High Sale Volume:
1 - CLINIQUE
2 - LANCOME
3 - LIZ CLAI
4 - MILCO IN
5 - CABERNET
```

The first five most common brands that have high sale volumes are CLINIQUE, LANCOME, LIZ CLAI, MILCO IN, and CABERNET

Scatter Plot Matrix:

[126]:
```
import seaborn as sns

# Select the relevant columns from your DataFrame
data_to_plot = highest_profit_per_brand_state_store[['PROFIT', 'QUANTITY',␣
↪'discount']]

# Create a Scatter Plot Matrix using Seaborn
scatter_plot_matrix = sns.pairplot(data_to_plot, diag_kind='kde', markers='o')

# Add a title to the Scatter Plot Matrix
scatter_plot_matrix.fig.suptitle("Scatter Plot Matrix of Profit, Quantity, and␣
↪Discount", y=1.02)

# Display the plot
plt.show()
```



This seems they don't have a clear relationship among quantity, discount, and profit.

[ ]: