

Create a sale data analysis of any commercial store.

importing Libraries

In [134...]

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)
```

importing data set

In [6]: df = pd.read_csv(r"C:\Users\Meharaj\Downloads\archive\Top 50 Fast-Food Chains in

In [8]: df

Out[8]:

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
0	Arby's	4462	1309	2293	1116	3409	40
1	Baskin-Robbins	686	296	2317	0	2317	102
2	Bojangles	1485	1924	496	277	773	15
3	Burger King	10033	1470	7054	51	7105	24
4	Carl's Jr.	1560	1400	1011	47	1058	-21
5	Checkers/Rally's	931	1145	568	266	834	-13
6	Chick-fil-A	16700	6100	2650	82	2732	155
7	Chipotle	7547	2641	0	2966	2966	198
8	Church's Chicken	776	870	731	161	892	-13
9	Culver's	2489	3099	831	6	837	55
10	Dairy Queen	4494	1036	4337	2	4339	-22
11	Del Taco	931	1551	306	294	600	4
12	Domino's	8641	1317	6185	375	6560	205
13	Dunkin'	10416	1127	9244	0	9244	161
14	El Pollo Loco	973	2000	292	189	481	1
15	Firehouse Subs	1044	909	1101	39	1140	9
16	Five Guys	2093	3172	911	479	1390	8
17	Freddy's Frozen Custard & Steakburgers	759	1842	391	29	420	32
18	Hardee's	2100	1117	1536	198	1734	-32
19	In-N-Out Burger	1175	3200	0	370	370	5
20	Jack in the Box	4077	1843	2055	163	2218	-23
21	Jersey Mike's	2203	1145	2087	13	2100	246
22	Jimmy John's	2301	866	2616	41	2657	48
23	KFC	5100	1408	3906	47	3953	10
24	Krispy Kreme	996	4000	51	307	358	6
25	Little Caesars	4185	1000	3601	580	4181	-28
26	Marco's Pizza	899	934	957	45	1002	48
27	McAlister's Deli	869	1866	472	33	505	24

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
28	McDonald's	45960	3420	12775	663	13438	244
29	Moe's Southwest Grill	661	1073	658	1	659	-22
30	Panda Express	4452	2173	147	2187	2334	150
31	Panera Bread	5650	2700	1130	950	2080	-25
32	Papa Johns	3486	1147	2564	600	3164	30
33	Papa Murphy's	809	643	1213	27	1240	-53
34	Pizza Hut	5500	1022	6526	22	6548	-13
35	Popeyes Louisiana Kitchen	4775	1867	2716	41	2754	146
36	QDOBA	835	1006	406	333	739	2
37	Raising Cane's	2377	4893	23	544	567	58
38	Shake Shack	777	3679	25	218	243	38
39	Sonic Drive-In	5835	1643	3232	320	3552	26
40	Starbucks	24300	1200	6497	8953	15450	113
41	Subway	9350	438	21147	0	21147	-1043
42	Taco Bell	12600	1823	6540	462	7002	203
43	Tim Hortons	687	1200	637	0	637	4
44	Tropical Smoothie Cafe	948	1009	1038	1	1039	125
45	Wendy's	11111	1895	5535	403	5938	57
46	Whataburger	3089	3640	131	742	873	29
47	White Castle	615	1749	0	349	349	-6
48	Wingstop	2278	1592	1498	36	1534	175
49	Zaxby's	2233	2484	761	147	908	3

Data Description

In [11]: `df.head() #head: This shows the first 5 rows by default.`

Out[11]:

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
0	Arby's	4462	1309	2293	1116	3409	40
1	Baskin-Robbins	686	296	2317	0	2317	102
2	Bojangles	1485	1924	496	277	773	15
3	Burger King	10033	1470	7054	51	7105	24
4	Carl's Jr.	1560	1400	1011	47	1058	-21

In [13]: `df.info() #info : This method provides an overview of the dataset, including col`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Fast-Food Chains      50 non-null    object 
 1   U.S. Systemwide Sales (Millions - U.S Dollars) 50 non-null    int64  
 2   Average Sales per Unit (Thousands - U.S Dollars) 50 non-null    int64  
 3   Franchised Stores     50 non-null    int64  
 4   Company Stores        50 non-null    int64  
 5   2021 Total Units      50 non-null    int64  
 6   Total Change in Units from 2020 50 non-null    int64  
dtypes: int64(6), object(1)
memory usage: 2.9+ KB
```

In [15]: `df.describe() #describe :This function computes statistical summaries like mean,`

Out[15]:

	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
count	50.000000	50.000000	50.00000	50.00000	50.000000	50.000000
mean	4965.060000	1857.660000	2663.96000	503.50000	3167.400000	29.700000
std	7531.439444	1163.266418	3801.49754	1332.16928	4124.173105	173.769875
min	615.000000	296.000000	0.00000	0.00000	243.000000	-1043.000000
25%	935.250000	1084.000000	478.00000	33.75000	788.250000	-4.250000
50%	2289.500000	1510.500000	1115.50000	176.00000	1634.000000	24.000000
75%	5400.000000	2129.750000	3103.00000	396.00000	3516.250000	91.000000
max	45960.000000	6100.000000	21147.00000	8953.00000	21147.000000	246.000000

```
In [17]: df.shape #shape: attribute in pandas is used to check the dimensions of a DataFrame
```

```
Out[17]: (50, 7)
```

```
In [19]: df.tail() #tail : method in pandas is used to display the last 5 rows of a DataFrame
```

```
Out[19]:
```

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
45	Wendy's	11111	1895	5535	403	5938	57
46	Whataburger	3089	3640	131	742	873	29
47	White Castle	615	1749	0	349	349	-6
48	Wingstop	2278	1592	1498	36	1534	175
49	Zaxby's	2233	2484	761	147	908	3

```
In [21]: df.dtypes #dtypes = Incorrect data types can cause issues during analysis.
```

```
Out[21]: Fast-Food Chains                         object  
U.S. Systemwide Sales (Millions - U.S Dollars)    int64  
Average Sales per Unit (Thousands - U.S Dollars)    int64  
Franchised Stores                           int64  
Company Stores                             int64  
2021 Total Units                           int64  
Total Change in Units from 2020          int64  
dtype: object
```

```
In [23]: df.columns #columns : attribute in pandas is used to display the column Labels (
```

```
Out[23]: Index(['Fast-Food Chains', 'U.S. Systemwide Sales (Millions - U.S Dollars)',  
               'Average Sales per Unit (Thousands - U.S Dollars)', 'Franchised Stores',  
               'Company Stores', '2021 Total Units',  
               'Total Change in Units from 2020'],  
               dtype='object')
```

Data Cleaning

```
In [26]: df.isnull().sum() #isnull().sum():method is used to identify missing values in
```

```
Out[26]: Fast-Food Chains                         0  
U.S. Systemwide Sales (Millions - U.S Dollars)    0  
Average Sales per Unit (Thousands - U.S Dollars)    0  
Franchised Stores                           0  
Company Stores                             0  
2021 Total Units                           0  
Total Change in Units from 2020          0  
dtype: int64
```

From the given dataset we conclude that their is No Missing Values.

Duplicates

```
In [36]: df.duplicated().sum()
```

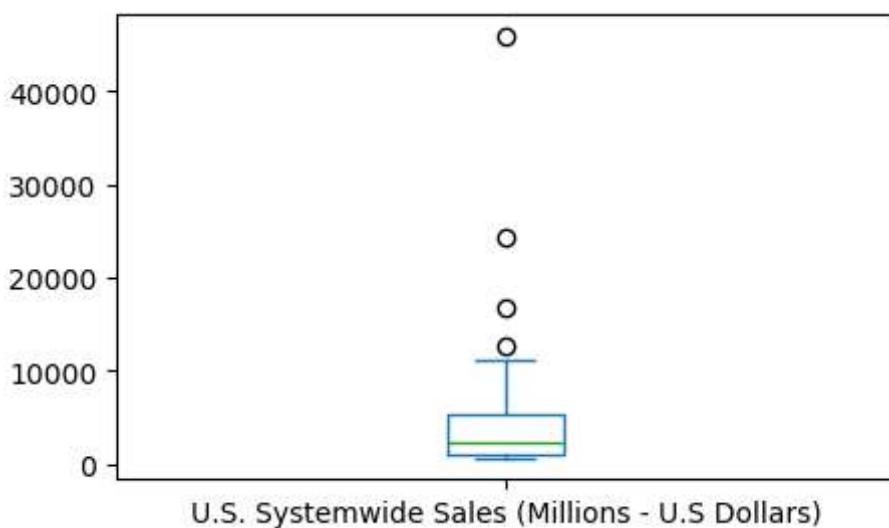
```
Out[36]: 0
```

There are No Duplicates From the given Dataset.

Outliers

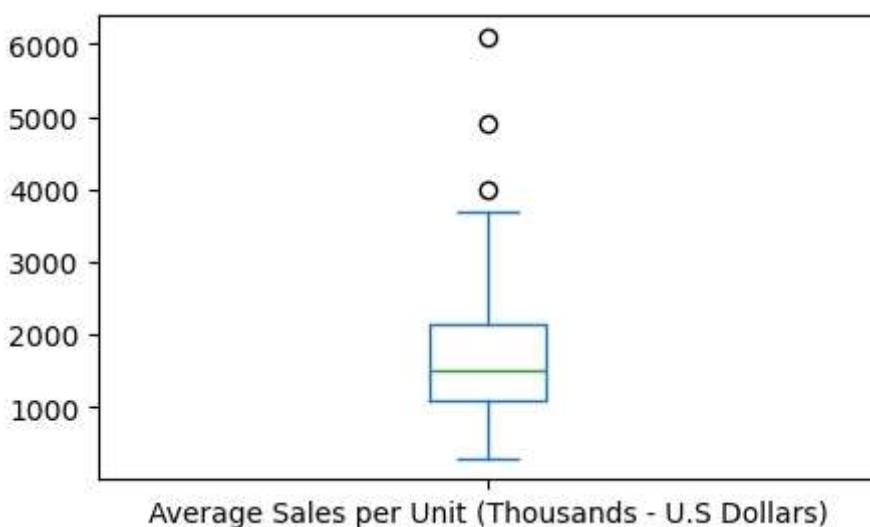
```
In [40]: plt.figure(figsize = (5,3))
df['U.S. Systemwide Sales (Millions - U.S Dollars)'].plot(kind='box')
```

```
Out[40]: <Axes: >
```



```
In [42]: plt.figure(figsize = (5,3))
df['Average Sales per Unit (Thousands - U.S Dollars)'].plot(kind='box')
```

```
Out[42]: <Axes: >
```



```
In [44]: plt.figure(figsize = (5,3))
df['Franchised Stores'].plot(kind='box')
```

```
Out[44]: <Axes: >
```



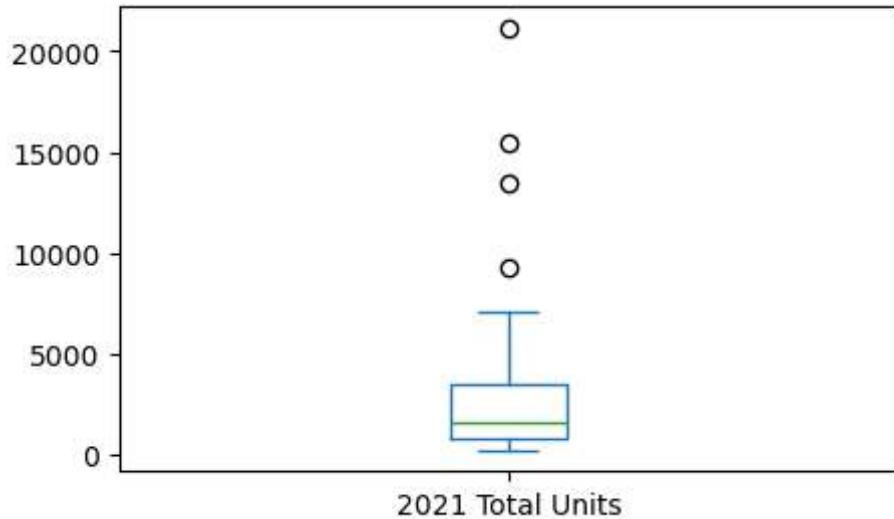
```
In [46]: plt.figure(figsize = (5,3))
df['Company Stores'].plot(kind='box')
```

```
Out[46]: <Axes: >
```



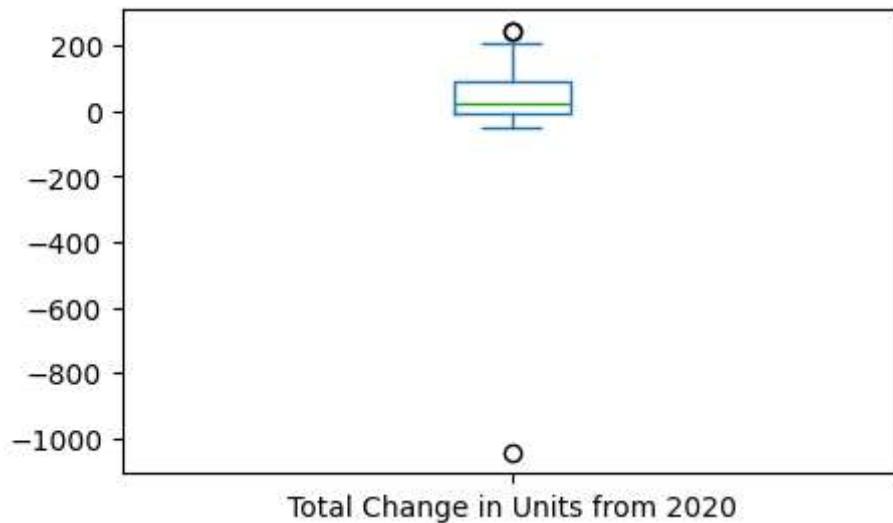
```
In [48]: plt.figure(figsize = (5,3))
df['2021 Total Units'].plot(kind='box')
```

```
Out[48]: <Axes: >
```



```
In [50]: plt.figure(figsize = (5,3))
df['Total Change in Units from 2020'].plot(kind='box')
```

```
Out[50]: <Axes: >
```



There are Outliers from the given dataset

Type Casting

```
In [188...]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 7 columns):
 #   Column                                         Non-Null Count  Dtype  
--- 
 0   Fast-Food Chains                           50 non-null    object  
 1   U.S. Systemwide Sales (Millions - U.S Dollars) 50 non-null    int64   
 2   Average Sales per Unit (Thousands - U.S Dollars) 50 non-null    int64   
 3   Franchised Stores                         50 non-null    int64   
 4   Company Stores                            50 non-null    int64   
 5   2021 Total Units                          50 non-null    int64   
 6   Total Change in Units from 2020          50 non-null    int64  
dtypes: int64(6), object(1)
memory usage: 2.9+ KB

```

In [190...]: df.head()

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
0	Arby's	4462	1309	2293	1116	3409	40
1	Baskin-Robbins	686	296	2317	0	2317	102
2	Bojangles	1485	1924	496	277	773	15
3	Burger King	10033	1470	7054	51	7105	24
4	Carl's Jr.	1560	1400	1011	47	1058	-21

In [194...]: df.columns

```

Out[194...]: Index(['Fast-Food Chains', 'U.S. Systemwide Sales (Millions - U.S Dollars)', 'Average Sales per Unit (Thousands - U.S Dollars)', 'Franchised Stores', 'Company Stores', '2021 Total Units', 'Total Change in Units from 2020'],
dtype='object')

```

UniVariate Analysis

In [87]: df.head()

Out[87]:

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
0	Arby's	4462	1309	2293	1116	3409	40
1	Baskin-Robbins	686	296	2317	0	2317	102
2	Bojangles	1485	1924	496	277	773	15
3	Burger King	10033	1470	7054	51	7105	24
4	Carl's Jr.	1560	1400	1011	47	1058	-21

HISTOGRAM PLOT = A histogram plot is a graphical representation of the distribution of numerical data. It groups the data into bins (intervals) and displays the frequency of data points in each bin as rectangular bars.

In [89]: `df["U.S. Systemwide Sales (Millions - U.S Dollars)"].min()`

Out[89]: 615

In [91]: `df["U.S. Systemwide Sales (Millions - U.S Dollars)"].max()`

Out[91]: 45960

In [93]: `df["U.S. Systemwide Sales (Millions - U.S Dollars)"].mean()`

Out[93]: 4965.06

In [99]: `df[df["U.S. Systemwide Sales (Millions - U.S Dollars)"]==615]`

Out[99]:

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
47	White Castle	615	1749	0	349	349	-6

In [102...]: `df[df["U.S. Systemwide Sales (Millions - U.S Dollars)"]==45960]`

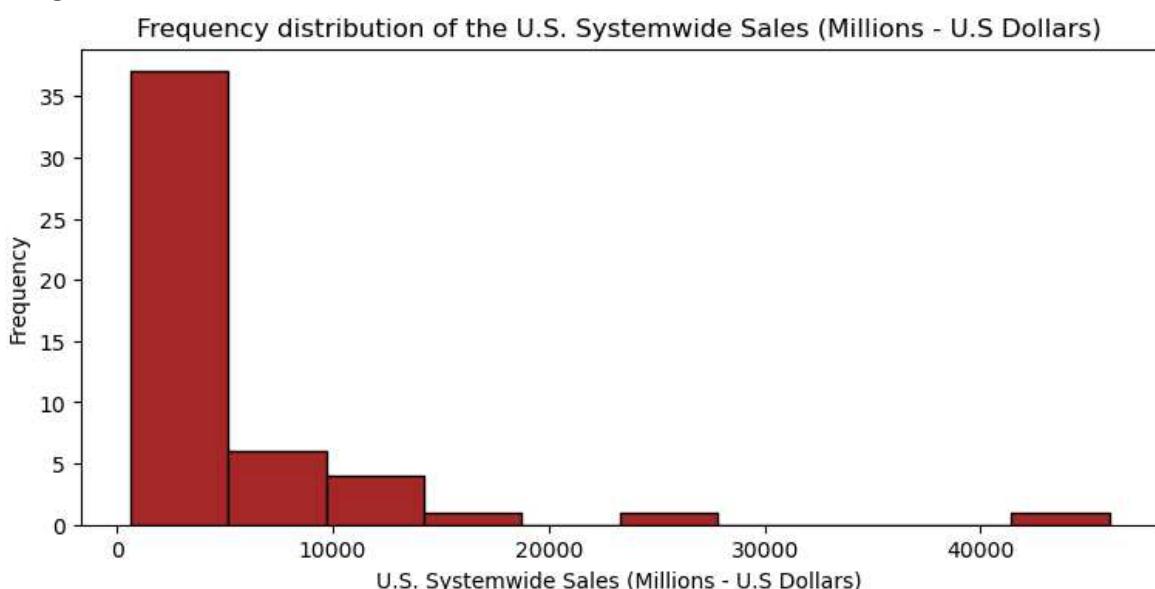
Out[102...]

Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
28 McDonald's	45960	3420	12775	663	13438	244

In [128...]

```
plt.figure(figsize = (6,3))
colors = ['brown', 'green']
plt.figure(figsize = (9,4))
df['U.S. Systemwide Sales (Millions - U.S Dollars)'].plot(kind= 'hist', edgecolor= colors)
plt.title('Frequency distribution of the U.S. Systemwide Sales (Millions - U.S Dollars)')
plt.xlabel("U.S. Systemwide Sales (Millions - U.S Dollars)")
plt.show()
```

<Figure size 600x300 with 0 Axes>



There are the maximum of 615 U.S. Systemwide Sales (Millions - U.S Dollars) for White Castle and minimum of 45960 U.S. Systemwide Sales (Millions - U.S Dollars) for McDonald's

In [108...]

```
df["Average Sales per Unit (Thousands - U.S Dollars)"].min()
```

Out[108...]

296

In [110...]

```
df["Average Sales per Unit (Thousands - U.S Dollars)"].max()
```

Out[110...]

6100

In [112...]

```
df["Average Sales per Unit (Thousands - U.S Dollars)"].mean()
```

Out[112...]

1857.66

In [114...]

```
df[df["Average Sales per Unit (Thousands - U.S Dollars)"] == 296]
```

Out[114...]

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
1	Baskin-Robbins	686	296	2317	0	2317	102

In [116...]

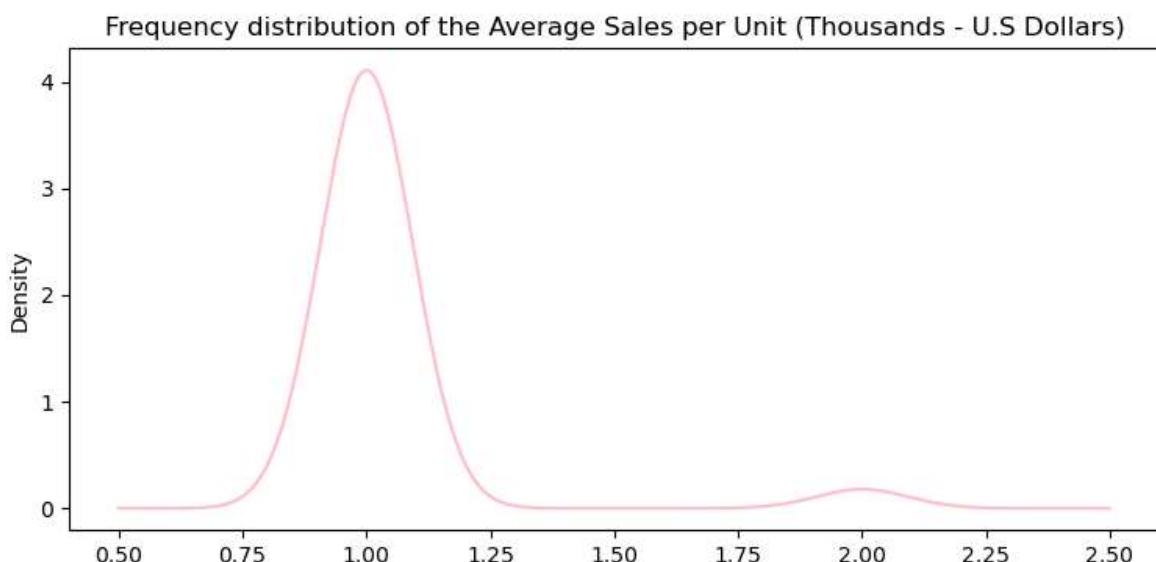
```
df[df["Average Sales per Unit (Thousands - U.S Dollars)"] == 6100]
```

Out[116...]

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
6	Chick-fil-A	16700	6100	2650	82	2732	155

In [184...]

```
plt.figure(figsize = (9,4))
df['Average Sales per Unit (Thousands - U.S Dollars)'].value_counts().plot(kind='density')
plt.title('Frequency distribution of the Average Sales per Unit (Thousands - U.S Dollars)')
plt.show()
```



There are minimum 296 Average Sales per Unit (Thousands - U.S Dollars) for Baskin-Robbins and maximum of 6100 Average Sales per Unit (Thousands - U.S Dollars) for Chick-fil-A

In [118...]

```
df["Franchised Stores"].min()
```

Out[118...]

```
0
```

In [120...]

```
df["Franchised Stores"].max()
```

Out[120...]

```
21147
```

```
In [122... df["Franchised Stores"].mean()
```

```
Out[122... 2663.96
```

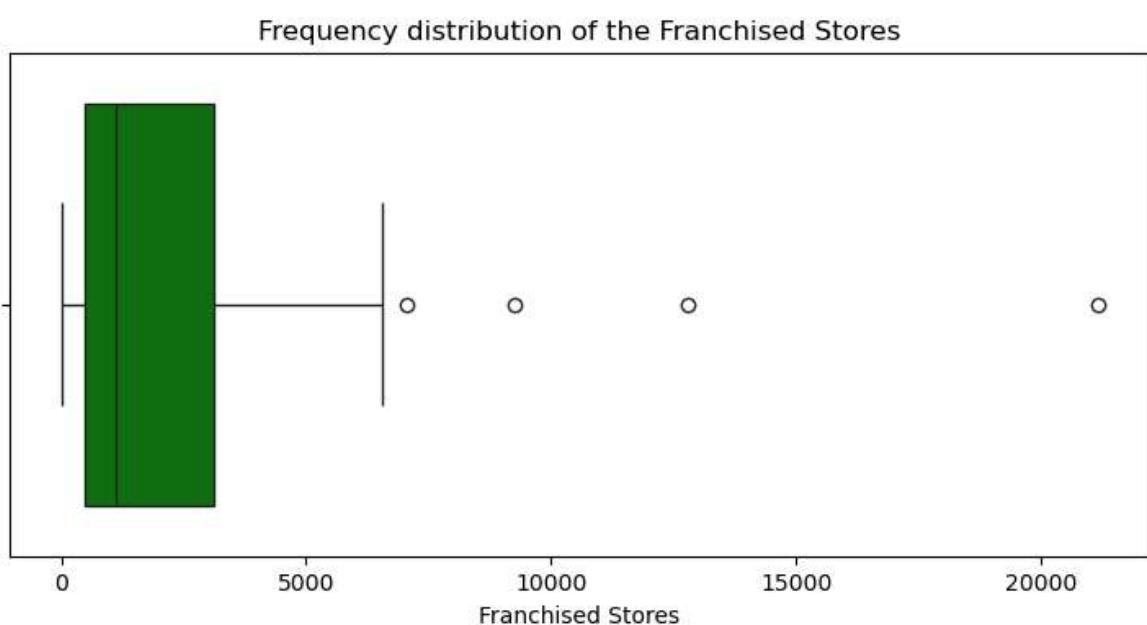
```
In [124... df[df["Franchised Stores"]==0]
```

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
7	Chipotle	7547	2641	0	2966	2966	198
19	In-N-Out Burger	1175	3200	0	370	370	5
47	White Castle	615	1749	0	349	349	-6

```
In [126... df[df["Franchised Stores"]==21147]
```

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
41	Subway	9350	438	21147	0	21147	-1043

```
In [182... plt.figure(figsize = (9,4))
sns.boxplot(data = df, x = "Franchised Stores", color = "green")
plt.title('Frequency distribution of the Franchised Stores')
plt.show()
```



There are 0 Franchised Stores for Chipotle,In-N-Out Burger,White Castle

There are 21147 Franchised Stores for Subway

```
In [140...]: df["Company Stores"].min()
```

```
Out[140...]: 0
```

```
In [142...]: df["Company Stores"].max()
```

```
Out[142...]: 8953
```

```
In [144...]: df["Company Stores"].mean()
```

```
Out[144...]: 503.5
```

```
In [146...]: df[df["Company Stores"] == 0]
```

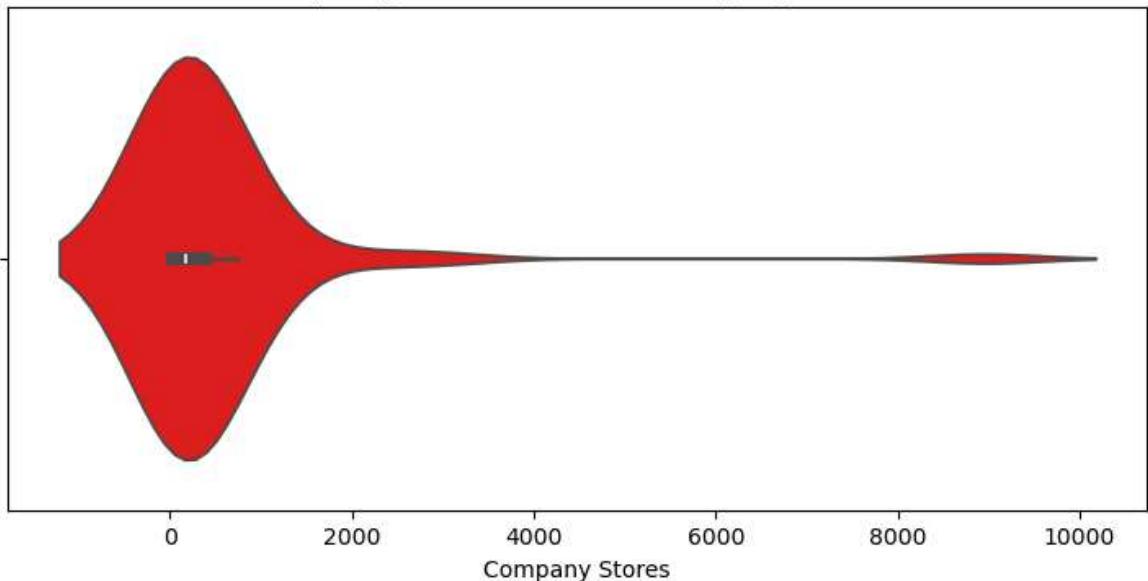
	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
1	Baskin-Robbins	686	296	2317	0	2317	102
13	Dunkin'	10416	1127	9244	0	9244	161
41	Subway	9350	438	21147	0	21147	-1043
43	Tim Hortons	687	1200	637	0	637	4

```
In [148...]: df[df["Company Stores"] == 8953]
```

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
40	Starbucks	24300	1200	6497	8953	15450	113

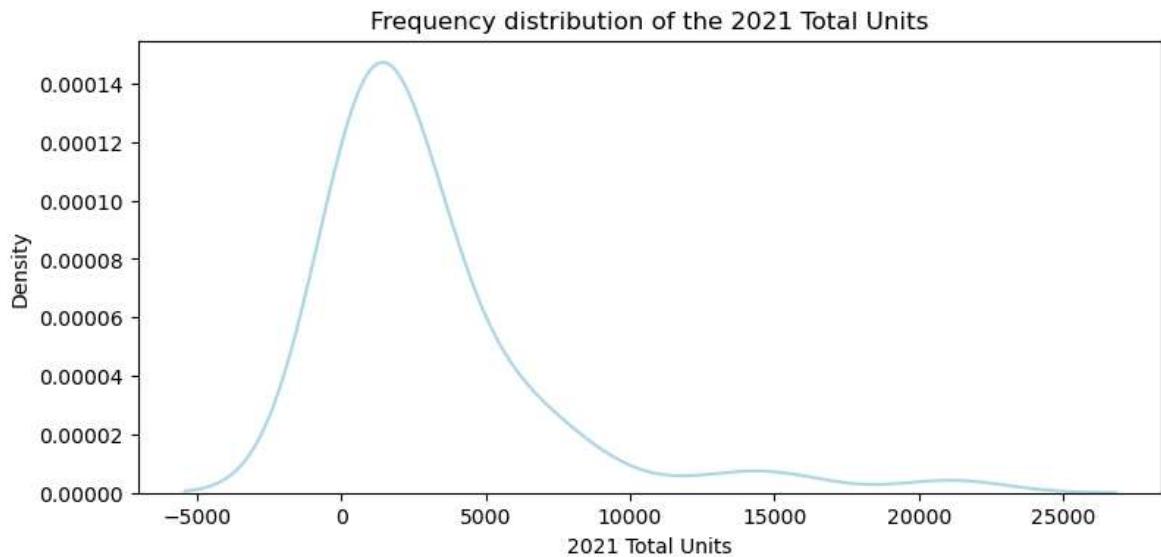
```
In [180...]: plt.figure(figsize = (9,4))
sns.violinplot(data = df, x = "Company Stores", color = "red")
plt.title('Frequency distribution of the Company Stores ')
plt.show()
```

Frequency distribution of the Company Stores



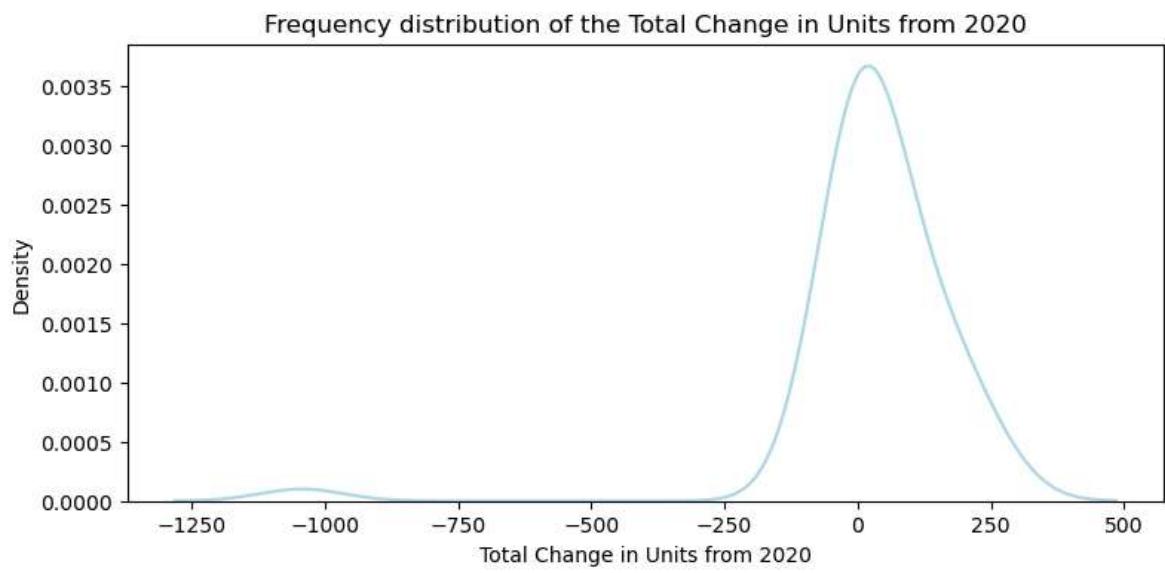
In []:

```
plt.figure(figsize = (9,4))
sns.kdeplot(data = df, x = "2021 Total Units",color = "lightblue")
plt.title('Frequency distribution of the 2021 Total Units')
plt.show()
```



In [176...]

```
plt.figure(figsize = (9,4))
sns.kdeplot(data = df, x = "Total Change in Units from 2020",color = "lightblue")
plt.title('Frequency distribution of the Total Change in Units from 2020')
plt.show()
```



bivariate analysis

In [198...]

df

Out[198...]

	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
0	Arby's	4462	1309	2293	1116	3409	40
1	Baskin-Robbins	686	296	2317	0	2317	102
2	Bojangles	1485	1924	496	277	773	15
3	Burger King	10033	1470	7054	51	7105	24
4	Carl's Jr.	1560	1400	1011	47	1058	-21
5	Checkers/Rally's	931	1145	568	266	834	-13
6	Chick-fil-A	16700	6100	2650	82	2732	155
7	Chipotle	7547	2641	0	2966	2966	198
8	Church's Chicken	776	870	731	161	892	-13
9	Culver's	2489	3099	831	6	837	55
10	Dairy Queen	4494	1036	4337	2	4339	-22
11	Del Taco	931	1551	306	294	600	4
12	Domino's	8641	1317	6185	375	6560	205
13	Dunkin'	10416	1127	9244	0	9244	161
14	El Pollo Loco	973	2000	292	189	481	1
15	Firehouse Subs	1044	909	1101	39	1140	9
16	Five Guys	2093	3172	911	479	1390	8
17	Freddy's Frozen Custard & Steakburgers	759	1842	391	29	420	32
18	Hardee's	2100	1117	1536	198	1734	-32
19	In-N-Out Burger	1175	3200	0	370	370	5
20	Jack in the Box	4077	1843	2055	163	2218	-23
21	Jersey Mike's	2203	1145	2087	13	2100	246
22	Jimmy John's	2301	866	2616	41	2657	48
23	KFC	5100	1408	3906	47	3953	10
24	Krispy Kreme	996	4000	51	307	358	6
25	Little Caesars	4185	1000	3601	580	4181	-28
26	Marco's Pizza	899	934	957	45	1002	48
27	McAlister's Deli	869	1866	472	33	505	24

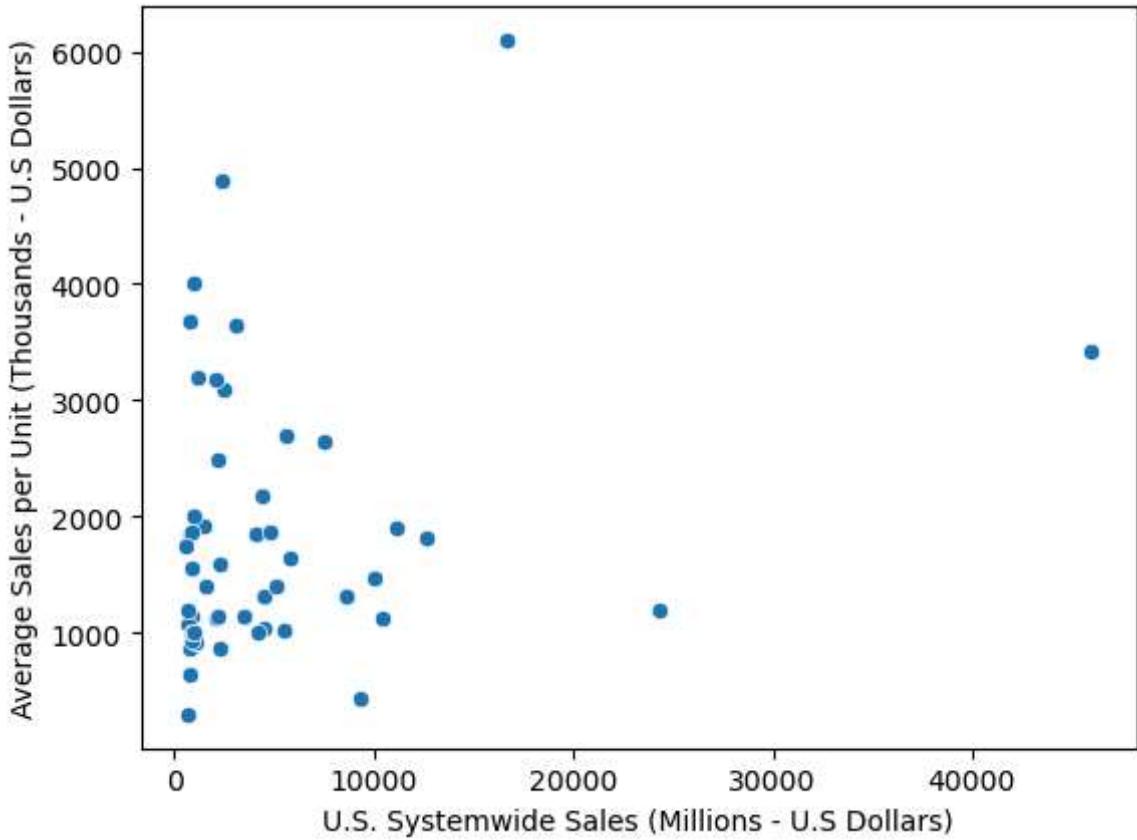
	Fast-Food Chains	U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
28	McDonald's	45960	3420	12775	663	13438	244
29	Moe's Southwest Grill	661	1073	658	1	659	-22
30	Panda Express	4452	2173	147	2187	2334	150
31	Panera Bread	5650	2700	1130	950	2080	-25
32	Papa Johns	3486	1147	2564	600	3164	30
33	Papa Murphy's	809	643	1213	27	1240	-53
34	Pizza Hut	5500	1022	6526	22	6548	-13
35	Popeyes Louisiana Kitchen	4775	1867	2716	41	2754	146
36	QDOBA	835	1006	406	333	739	2
37	Raising Cane's	2377	4893	23	544	567	58
38	Shake Shack	777	3679	25	218	243	38
39	Sonic Drive-In	5835	1643	3232	320	3552	26
40	Starbucks	24300	1200	6497	8953	15450	113
41	Subway	9350	438	21147	0	21147	-1043
42	Taco Bell	12600	1823	6540	462	7002	203
43	Tim Hortons	687	1200	637	0	637	4
44	Tropical Smoothie Cafe	948	1009	1038	1	1039	125
45	Wendy's	11111	1895	5535	403	5938	57
46	Whataburger	3089	3640	131	742	873	29
47	White Castle	615	1749	0	349	349	-6
48	Wingstop	2278	1592	1498	36	1534	175
49	Zaxby's	2233	2484	761	147	908	3

Numerical vs. Numerical

In [206]:

```
sns.scatterplot(x=df['U.S. Systemwide Sales (Millions - U.S Dollars)'], y=df['Av
```

```
plt.show()
```



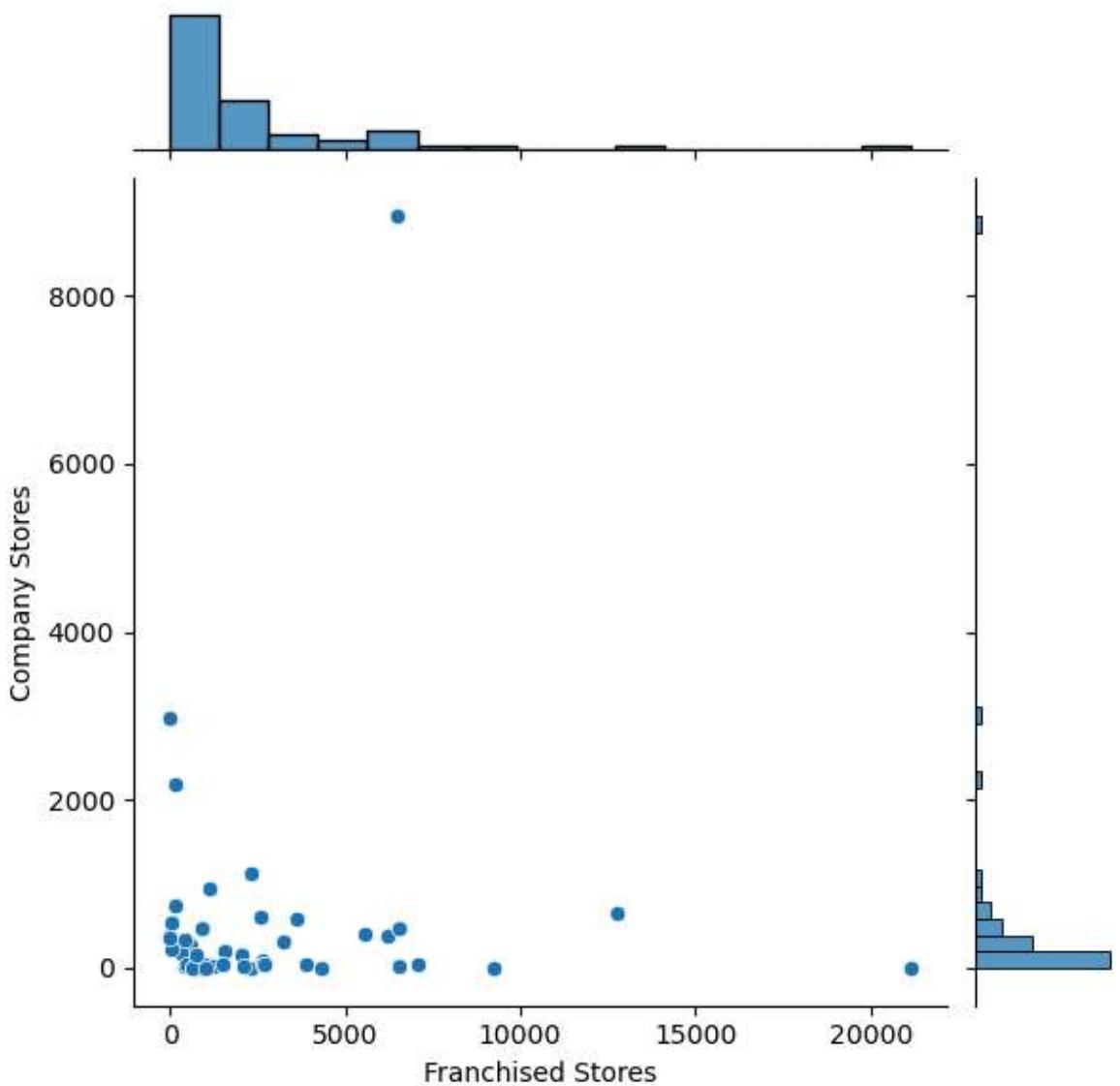
```
In [218]: correlation_matrix = df.select_dtypes(include=['number']).corr()  
correlation_matrix
```

Out[218...]

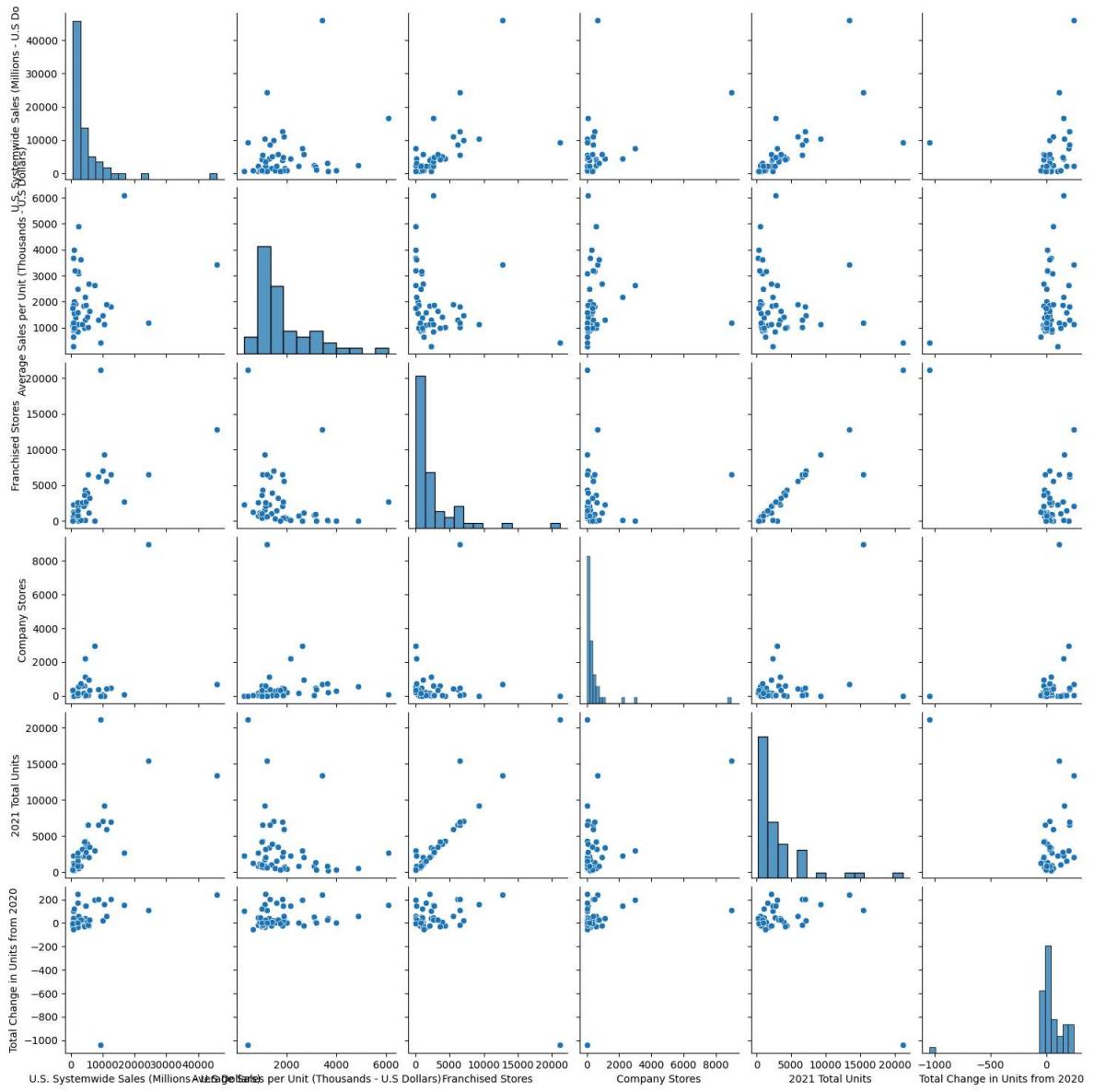
U.S. Systemwide Sales (Millions - U.S Dollars)	Average Sales per Unit (Thousands - U.S Dollars)	Franchised Stores	Company Stores	2021 Total Units	Total Change in Units from 2020
U.S. Systemwide Sales (Millions - U.S Dollars)	1.000000	0.231458	0.628534	0.400637	0.708770
Average Sales per Unit (Thousands - U.S Dollars)	0.231458	1.000000	-0.211528	0.021543	-0.188019
Franchised Stores	0.628534	-0.211528	1.000000	0.077277	0.946721
Company Stores	0.400637	0.021543	0.077277	1.000000	0.394251
2021 Total Units	0.708770	-0.188019	0.946721	0.394251	1.000000
Total Change in Units from 2020	0.167489	0.235000	-0.481116	0.146101	-0.396291

In [220...]

```
sns.jointplot(x=df['Franchised Stores'], y=df['Company Stores'], kind='scatter')
plt.show()
```

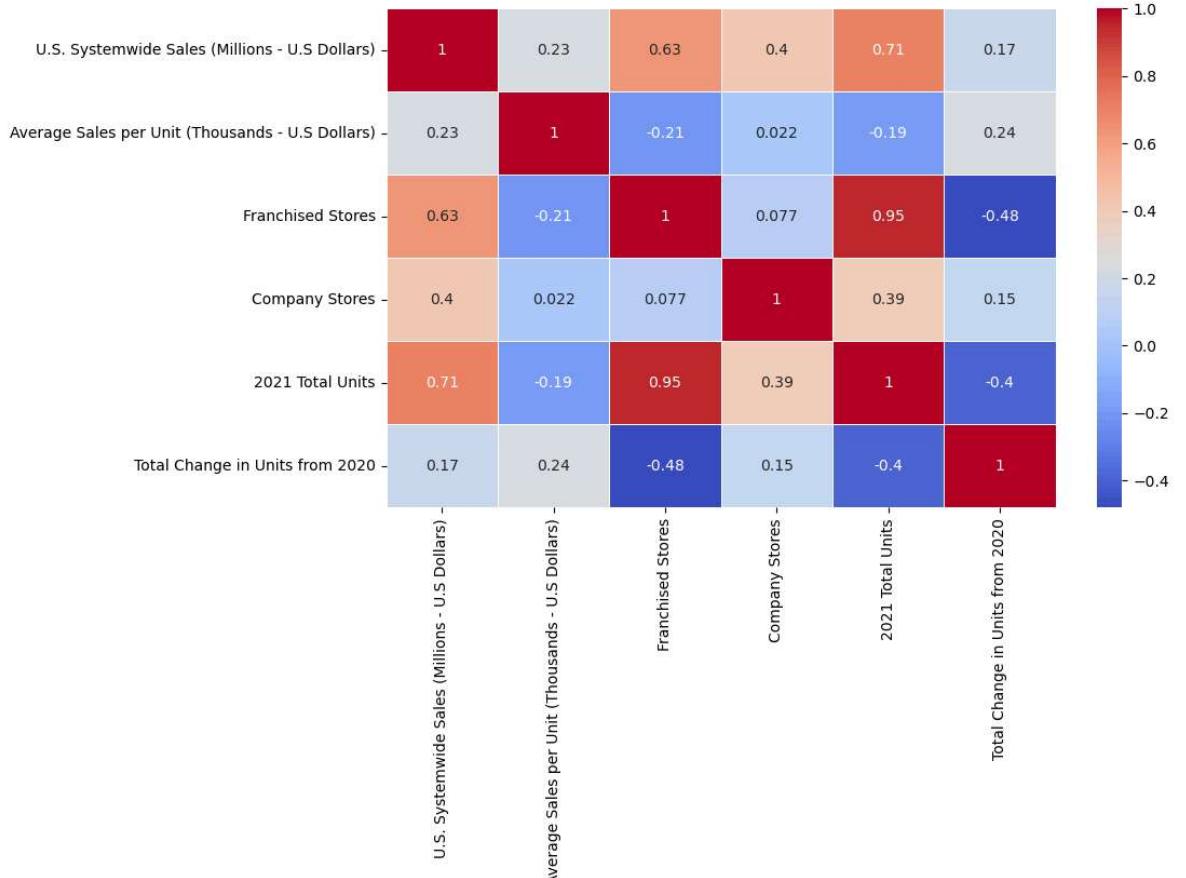


```
In [222]: sns.pairplot(df)
plt.show()
```



multivariate analysis

```
In [227]: plt.figure(figsize=(10,6))
sns.heatmap( df.select_dtypes(include=['number']).corr(), annot=True, cmap='cool'
```



In []: