Caffeine and Calorie Content of Drinks

Ally Blitch, Cagan Abney, Fabi Estrada

Inspiration

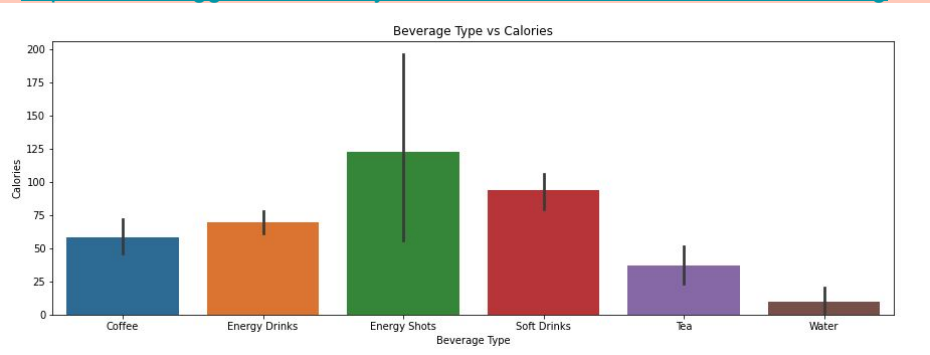
We chose the Caffeine Content in Drinks Dataset because we are familiar with caffeinated drinks, and it is important to know the calorie and caffeine content of beverages that we often consume. The dataset has a high usability rating and has had other visualization projects completed off of it.

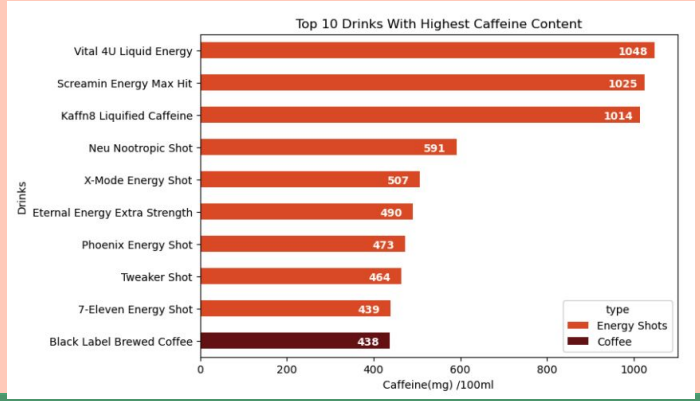
Similar EDAs:

https://www.kaggle.com/code/darshitapangam/caffeinated-drinks-eda-content-based-recommender

https://www.kaggle.com/code/bngphmhu/analysis-of-data-on-caffeine-content-in-beverage

https://www.kaggle.com/code/yadhua/caffeine-content-eda-and-clustering

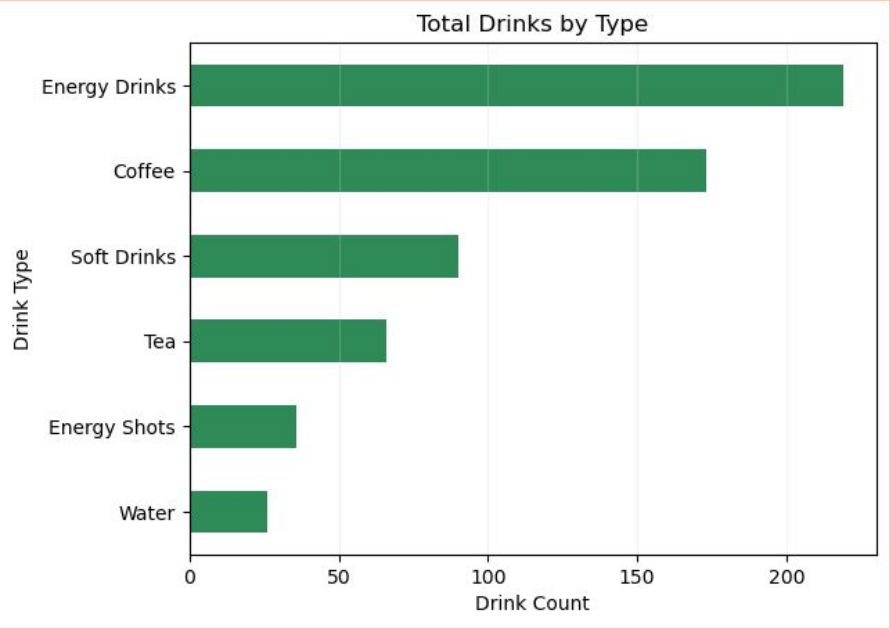




Dataset

[Caffeine Content of Drinks]

This dataset is 29.21 kB which consists of 5 columns and 611 rows. The data includes volume of drinks (mL), calorie content, caffeine content (mg), drink name and type.



GitHub Link

https://github.com/Cagan124/SMU-Project-1

Research Questions

1. Does coffee/tea have a lower volume to caffeine ratio than other drink categories?
2. Does soda have a higher calorie to caffeine ratio than other drink categories?
3. Which drink type has the most caffeine per 100 mL?

Data Engineering

- Cleaned up column titles
- Added columns to show the calories and caffeine per 100 mL to get a better comparison

Initial Dataset

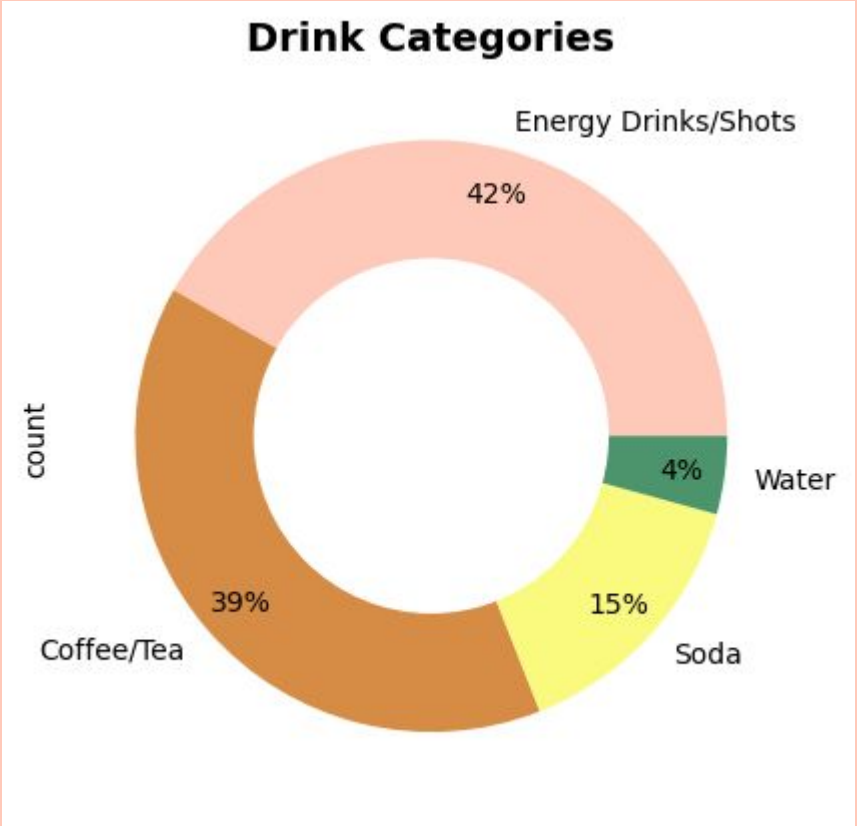
	drink	Volume (ml)	Calories	Caffeine (mg)	type
0	Costa Coffee	256.993715	0	277	Coffee
1	Coffee Friend Brewed Coffee	250.191810	0	145	Coffee
2	Hell Energy Coffee	250.191810	150	100	Coffee
3	Killer Coffee (AU)	250.191810	0	430	Coffee
4	Nescafe Gold	250.191810	0	66	Coffee

Final Dataset

	drink	volume	calories	caffeine	drink type	category	caffeine_100ml	calories_100ml
0	Costa Coffee	256.993715	0	277	Coffee	Coffee/Tea	107.784737	0.000000
1	Coffee Friend Brewed Coffee	250.191810	0	145	Coffee	Coffee/Tea	57.955534	0.000000
2	Hell Energy Coffee	250.191810	150	100	Coffee	Coffee/Tea	39.969334	59.954001
3	Killer Coffee (AU)	250.191810	0	430	Coffee	Coffee/Tea	171.868136	0.000000
4	Nescafe Gold	250.191810	0	66	Coffee	Coffee/Tea	26.379760	0.000000

Data Engineering Cont.

- Created categories for the drink types
  - Wanted to combine the 3 smaller drink types with the 3 larger ones to create 3 parent categories
  - We decided to separate them back into 4 categories because the data for soda and water ran bimodally in our violin plots



Data Engineering Cont.

- We noticed we had some outliers (max values) when running our .describe
- We then ran each column in descending order to check where the outliers were coming from (see next slide)

	volume	calories	caffeine
count	610.000000	610.000000	610.000000
mean	346.543630	75.527869	134.693443
std	143.747738	94.799919	155.362861
min	7.393375	0.000000	0.000000
25%	236.588000	0.000000	50.000000
50%	354.882000	25.000000	100.000000
75%	473.176000	140.000000	160.000000
max	1419.528000	830.000000	1555.000000

Data Engineering Cont.

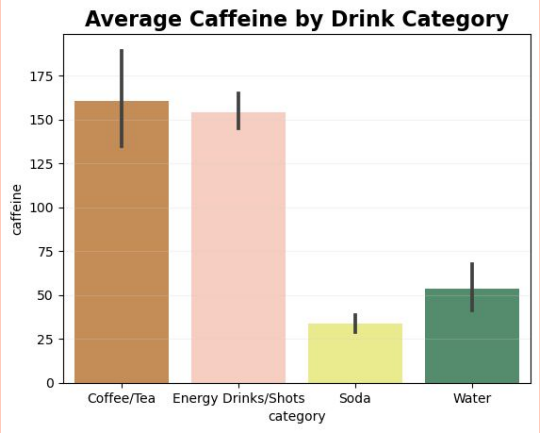
	drink	volume	calories	caffeine	drink type
85	Black Label Brewed Coffee	354.882	0	1555	Coffee
102	Very Strong Coffee	354.882	0	1350	Coffee
92	Devils Brew Extreme Caffeine Coffee	354.882	0	1325	Coffee
101	Taft Coffee (EU)	354.882	0	1182	Coffee
95	High Voltage Coffee (AU)	354.882	0	1150	Coffee

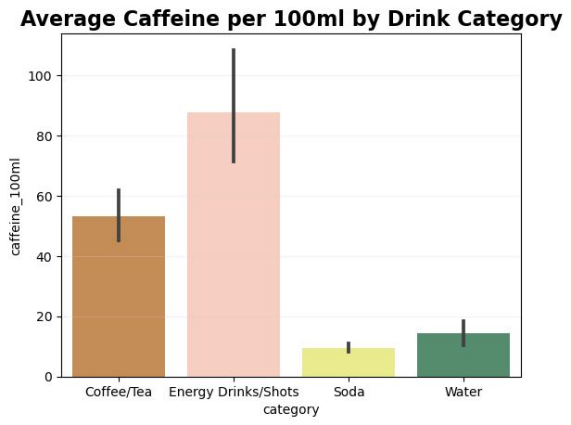
	drink	volume	calories	caffeine	drink type
29	Starbucks Bottled Iced Coffee	1419.528000	240	640	Coffee
530	McDonalds Sweet Tea	946.352000	160	100	Tea
214	Monster Hydro	751.166900	150	188	Energy Drinks
532	Fuze Iced Tea	709.764000	160	24	Tea
30	Baskin Robbins Cappuccino Blast	709.764000	470	234	Coffee

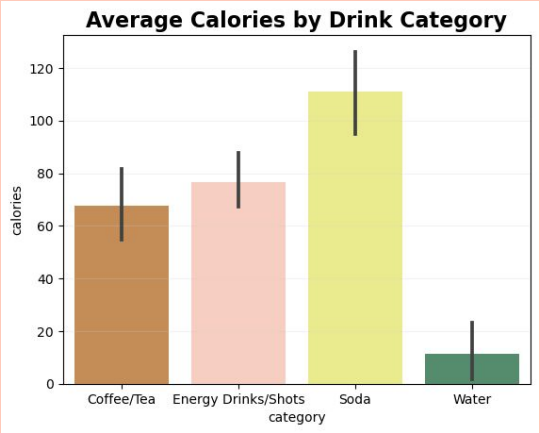
	drink	volume	calories	caffeine	drink type
37	Arby's Jamocha Shake	473.176000	830	12	Coffee
30	Baskin Robbins Cappuccino Blast	709.764000	470	234	Coffee
13	Dare Iced Coffee	500.087885	429	160	Coffee
81	Big Train Java Chip Ice Coffee	354.882000	410	49	Coffee
52	Peet's Caffe Mocha	473.176000	390	165	Coffee

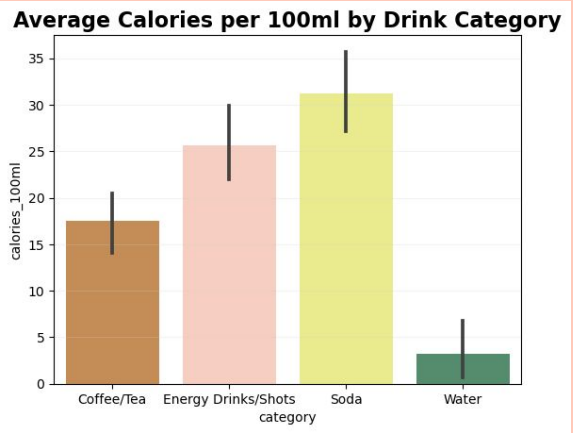
- After looking at these outliers we decided they were still representative of the variations in drink options on the market and chose to include them in our analysis

- Coffee/Tea's average caffeine dropped significantly in comparison to the other categories when looking at average caffeine per 100mL.
- Soda is statistically significant in terms of calories compared to the other categories in both average calories and average calories per 100ml.









Checking for Correlations:

- Coffee/Tea:
  - Calories & Volume: stronger correlation than original dataset
  - Caffeine & Volume: stronger correlation than original dataset
  - Calories & Caffeine: slightly weaker correlation than original dataset

- Energy Drinks/Shots:
  - Calories & Volume: slightly weaker correlation than original dataset
  - Caffeine & Volume: about equal correlation as original dataset
  - Calories & Caffeine: stronger correlation than original dataset

Original Dataset:

	volume	calories	caffeine
volume	1.000000	0.341998	0.110770
calories	0.341998	1.000000	-0.126021
caffeine	0.110770	-0.126021	1.000000

Coffee/Tea

	volume	calories	caffeine	caffeine_100ml	calories_100ml
volume	1.000000	0.409093	0.162447	-0.197330	0.181752
calories	0.409093	1.000000	-0.102982	-0.213935	0.917602
caffeine	0.162447	-0.102982	1.000000	0.862237	-0.139445
caffeine_100ml	-0.197330	-0.213935	0.862237	1.000000	-0.145618
calories_100ml	0.181752	0.917602	-0.139445	-0.145618	1.000000

Energy Drinks/Shots

	volume	calories	caffeine	caffeine_100ml	calories_100ml
volume	1.000000	0.326900	0.110534	-0.646098	-0.218562
calories	0.326900	1.000000	-0.158162	-0.276000	0.518352
caffeine	0.110534	-0.158162	1.000000	0.287678	-0.213583
caffeine_100ml	-0.646098	-0.276000	0.287678	1.000000	0.259719
calories_100ml	-0.218562	0.518352	-0.213583	0.259719	1.000000

Checking for correlations:

- Soda:
  - Calories & Volume: weaker correlation than original dataset
  - Caffeine & Volume: weaker correlation than original dataset
  - Calories & Caffeine: slightly stronger correlation than original dataset

- Water:
  - Calories & Volume: much weaker correlation than original dataset
  - Caffeine & Volume: weaker correlation than original dataset
  - Calories & Caffeine: slightly stronger correlation than original dataset

Original Dataset:

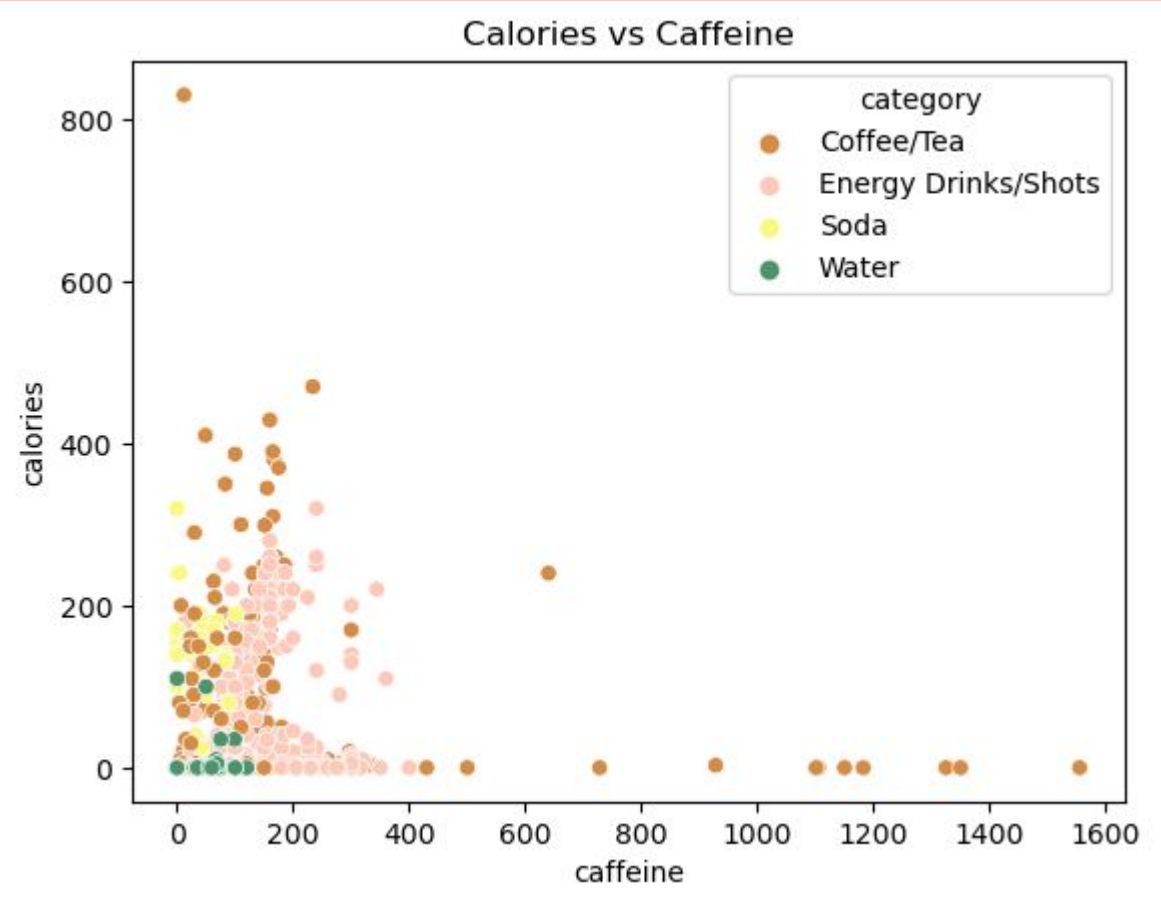
	volume	calories	caffeine
volume	1.000000	0.341998	0.110770
calories	0.341998	1.000000	-0.126021
caffeine	0.110770	-0.126021	1.000000

Soda

	volume	calories	caffeine	caffeine_100ml	calories_100ml
volume	1.000000	0.277084	0.073379	-0.033439	0.031704
calories	0.277084	1.000000	-0.156792	-0.182037	0.960212
caffeine	0.073379	-0.156792	1.000000	0.984448	-0.168080
caffeine_100ml	-0.033439	-0.182037	0.984448	1.000000	-0.178579
calories_100ml	0.031704	0.960212	-0.168080	-0.178579	1.000000

Water

	volume	calories	caffeine	caffeine_100ml	calories_100ml
volume	1.000000	0.043064	0.038770	-0.299305	-0.130185
calories	0.043064	1.000000	-0.147638	-0.128281	0.946101
caffeine	0.038770	-0.147638	1.000000	0.925890	-0.162300
caffeine_100ml	-0.299305	-0.128281	0.925890	1.000000	-0.089144
calories_100ml	-0.130185	0.946101	-0.162300	-0.089144	1.000000


Calories vs Caffeine
category
Coffee/Tea
Energy Drinks/Shots
Soda
Water
calories
caffeine
0
200
400
600
800
1000
1200
1400
1600

Original Dataset:

	volume	calories	caffeine
count	610.000000	610.000000	610.000000
mean	346.543630	75.527869	134.693443
std	143.747738	94.799919	155.362861
min	7.393375	0.000000	0.000000
25%	236.588000	0.000000	50.000000
50%	354.882000	25.000000	100.000000
75%	473.176000	140.000000	160.000000
max	1419.528000	830.000000	1555.000000

Coffee/Tea

	volume	calories	caffeine	caffeine_100ml	calories_100ml
count	239.000000	239.000000	239.000000	239.000000	239.000000
mean	342.665051	67.769874	160.623431	53.132138	17.480515
std	161.783104	112.777805	221.709081	67.446041	26.878013
min	12.716605	0.000000	0.000000	0.000000	0.000000
25%	236.588000	0.000000	50.000000	19.020407	0.000000
50%	325.308500	5.000000	120.000000	34.870746	1.127135
75%	473.176000	115.000000	165.000000	58.565067	30.163675
max	1419.528000	830.000000	1555.000000	438.173816	175.410418

- Coffee/Tea is lower in calories and higher in caffeine on average
  - Skews right in calories and caffeine
- Energy Drinks/Shots is about equal in calories and higher in caffeine on average
  - Skews right in calories and caffeine

Energy Drinks/Shots

	volume	calories	caffeine	caffeine_100ml	calories_100ml
count	255.000000	255.000000	255.000000	255.000000	255.000000
mean	342.209465	76.764706	154.298039	87.688772	25.686244
std	152.118572	81.350362	78.599437	148.911333	33.626471
min	7.393375	0.000000	0.000000	0.000000	0.000000
25%	250.043943	5.000000	90.000000	31.975467	2.057105
50%	354.882000	35.000000	142.000000	33.814057	19.020407
75%	473.176000	131.500000	200.000000	61.536610	44.573302
max	751.166900	320.000000	400.000000	1048.235752	281.783804

Original Dataset:

	volume	calories	caffeine
count	610.000000	610.000000	610.000000
mean	346.543630	75.527869	134.693443
std	143.747738	94.799919	155.362861
min	7.393375	0.000000	0.000000
25%	236.588000	0.000000	50.000000
50%	354.882000	25.000000	100.000000
75%	473.176000	140.000000	160.000000
max	1419.528000	830.000000	1555.000000

- Soda is higher in calories and lower in caffeine on average
  - Skews left in calories and caffeine
- Water is lower in calories and caffeine on average
  - Skews right in calories and slightly left in caffeine

Soda

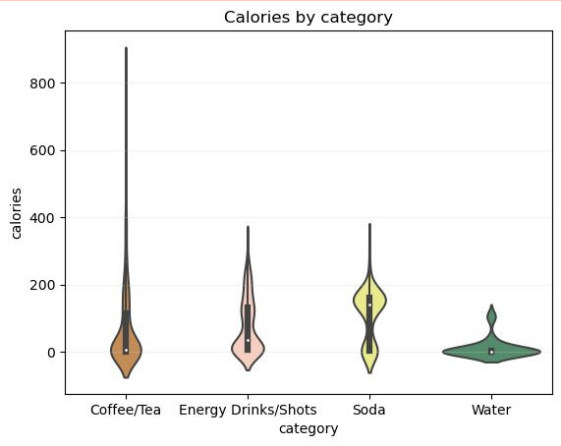
	volume	calories	caffeine	caffeine_100ml	calories_100ml
count	90.000000	90.000000	90.000000	90.000000	90.000000
mean	355.243454	111.111111	33.677778	9.506661	31.199914
std	41.509635	75.598354	24.915961	6.849069	21.166006
min	236.588000	0.000000	0.000000	0.000000	0.000000
25%	354.882000	2.500000	9.250000	2.606500	0.704460
50%	354.882000	140.000000	37.000000	10.707785	40.365530
75%	354.882000	160.000000	47.750000	13.455177	45.085409
max	591.470000	320.000000	102.000000	25.148447	101.442170

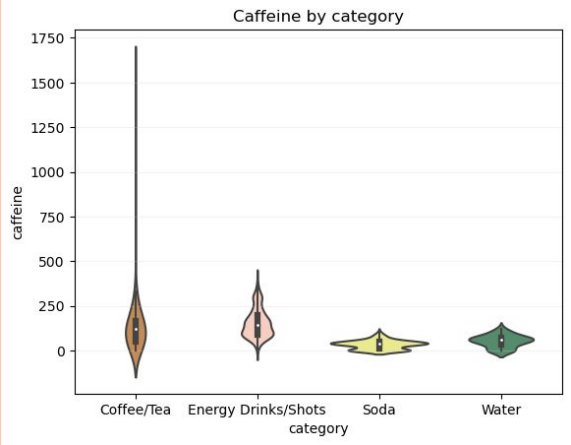
Water

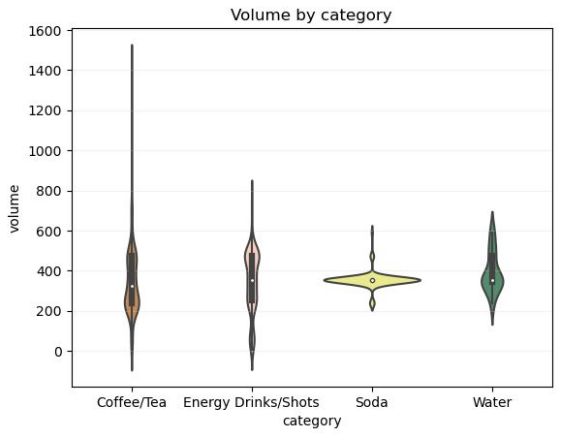
	volume	calories	caffeine	caffeine_100ml	calories_100ml
count	26.000000	26.000000	26.000000	26.000000	26.000000
mean	394.590111	11.538462	53.730769	14.406834	3.186064
std	99.702740	29.146843	34.060602	10.863597	8.280316
min	236.588000	0.000000	0.000000	0.000000	0.000000
25%	343.791937	0.000000	35.000000	8.453514	0.000000
50%	354.882000	0.000000	60.000000	12.342631	0.000000
75%	473.176000	3.750000	73.750000	22.052646	0.792517
max	591.470000	110.000000	120.000000	39.569641	37.195462

Data Analysis

- We created violin plots to visually compare calories, caffeine, and volume across the different categories







T-test Results

- The p-value comparing the caffeine of soda and water is below 0.05
  - This result shows the caffeine values are statistically significantly different

```
4]:  1 grp0 = df.loc[df.category == "Coffee/Tea","caffeine"]
     2 grp1 = df.loc[df.category == "Energy Drinks/Shots","caffeine"]
     3 grp2 = df.loc[df.category == "Soda","caffeine"]
     4 grp3 = df.loc[df.category == "Water","caffeine"]
```

```
5]:  1 st.ttest_ind(grp0,grp1)
5]: TtestResult(statistic=0.4278307582724591, pvalue=0.6689615437595988, df=492.0)
```

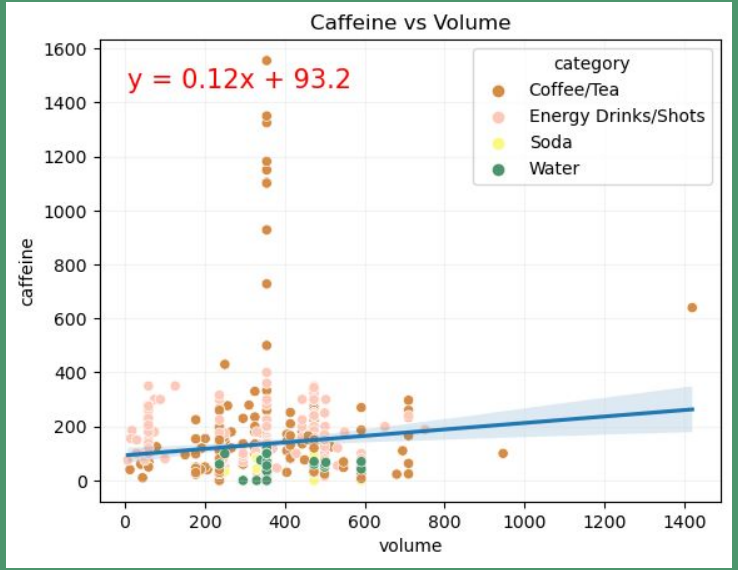
```
]:  1 st.ttest_ind(grp1,grp3)
]: TtestResult(statistic=6.4543005696776605, pvalue=4.79774554259269e-10, df=279.0)
```

```
1 st.ttest_ind(grp2,grp3)
TtestResult(statistic=-3.3129377635121506, pvalue=0.0012376061281805858, df=114.0)
```

Research Question 1:

Does Coffee/Tea have a lower volume to caffeine ratio than other drink categories?

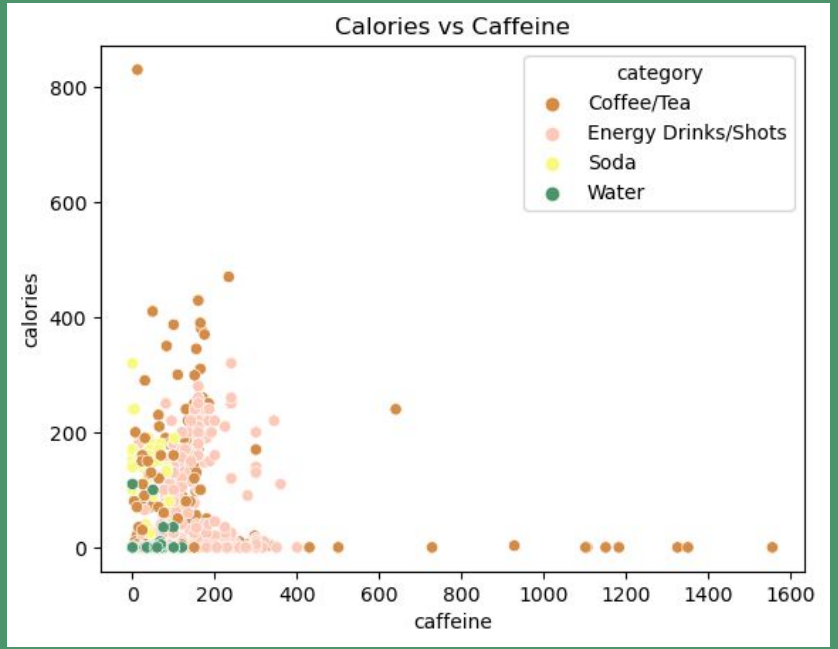
According to the data, coffee/tea does have a lower volume to caffeine ratio than other drink types.



Research Question 2:

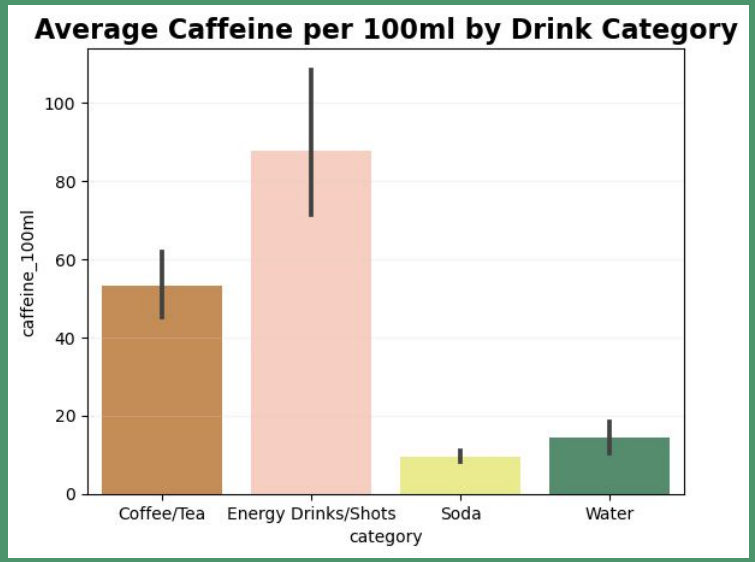
Does Soda have a higher calorie to caffeine ratio than other drink categories?

According to the data, soda does have a higher calorie to caffeine ratio than other drink types.



Research Question 3:

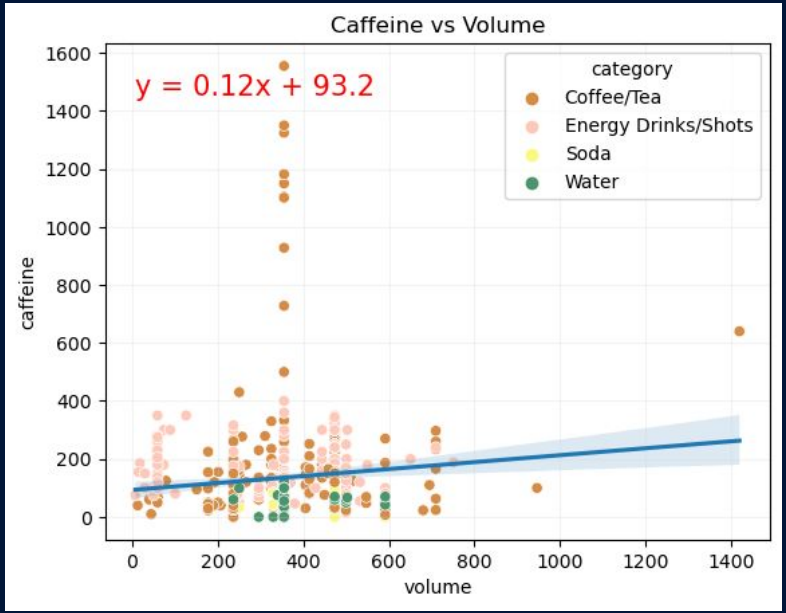
Which drink type has the most caffeine per 100 mL?

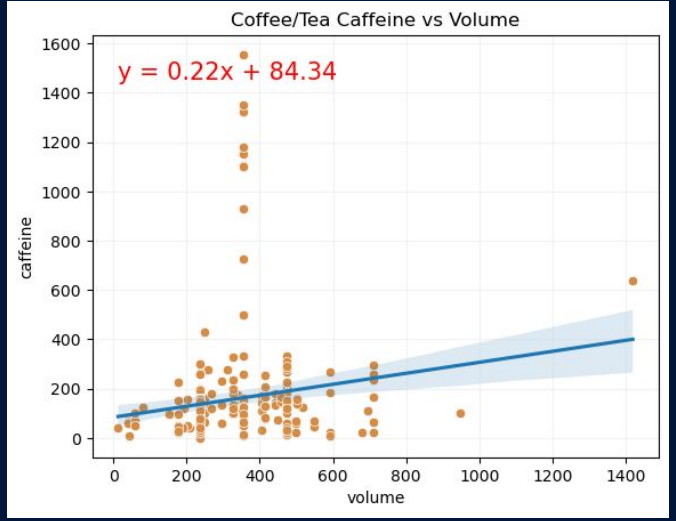


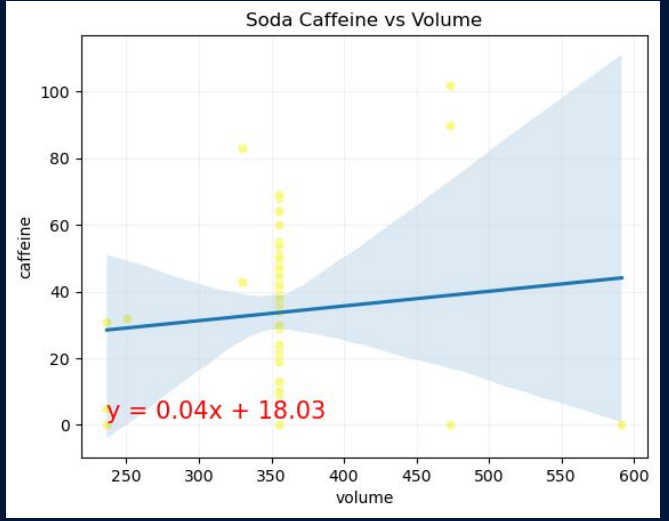
According to the graph, “Energy Drinks/Shots” has the most caffeine per 100mL.

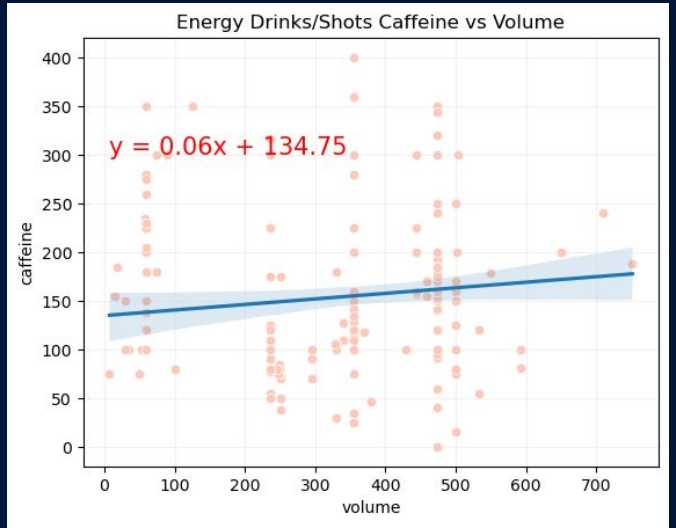
Regression

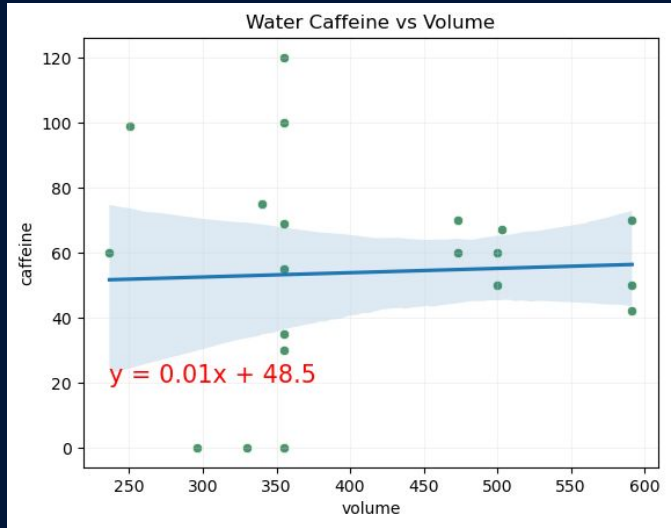
- The relationship between caffeine and volume is not highly significant because the slope is only 0.12
- This is caused by energy drinks/shots, soda, and water all having weak correlations between caffeine and volume




Coffee/Tea Caffeine vs Volume
y = 0.22x + 84.34
caffeine
volume
0
200
400
600
800
1000
1200
1400
1600


Soda Caffeine vs Volume
y = 0.04x + 18.03
caffeine
volume
0
20
40
60
80
100
250
300
350
400
450
500
550
600


Energy Drinks/Shots Caffeine vs Volume
y = 0.06x + 134.75
caffeine
volume
0
50
100
150
200
250
300
350
400
0
100
200
300
400
500
600
700


Water Caffeine vs Volume
y = 0.01x + 48.5
caffeine
volume
0
20
40
60
80
100
120
250
300
350
400
450
500
550
600

Conclusions

- Coffee/tea drinks are typically the most highly caffeinated, but energy drinks/shots have more caffeine per 100mL
- Soda has significantly lower caffeine levels compared to calories on average.
- Water is low in calories and caffeine on average

Future Work

- A future study may be more effective comparing only personal sized/canned beverages or only beverages found in restaurants and coffee shops.
    - This data includes bulk sizes sold at grocery stores that are not expected to be consumed entirely in one sitting, therefore caffeine and calorie count may be more useful to view at a serving size level
- Further work could involve grouping together brands to compare within a certain brand or between brands
- Another future study could include prices to further compare value of drink options

Questions?