Future 50 Restaurants
Exploratory Analysis

Aaron Bone, abone@bellarmine.edu

I. INTRODUCTION

This dataset gives a prediction for the future top 50 restaurants based off of sales in 2019. This is only for growing, newer restaurants, so no Subway, McDonalds, etc.

II. DATA SET DESCRIPTION

This dataset originally comes with 9 variables, but after I split the column "Location" into "City" and "State" it has 10 columns with 50 rows. Here is the table with all data types and percentage missing. This dataset was luckily not missing a single piece of data.

Here are descriptions of each variable:

Rank: Rank in which the restaurant appears on the top 50.
Restaurant: Name of the Restaurant.
Sales: Total sales expressed as ($000,000), meaning Sales = 24, Sales = $240,000.
YOY_Sales: Year Over Year Sales, shows a percent difference between this year and the previous, the slope essentially.
Units: Number of restaurants.
YOY_Units: Year Over Year Units, shows a percent difference between this year and the previous growth in units.
Unit_Volume: Average Unit Volume in ($000)
Franchising: Yes or No boolean on whether the restaurant is a franchise.
City: City of original unit.
State: State of original unit.

Table 1: Data Types and Missing Data

Variable Name	*Data Type*	*Missing Data (%)*
Rank	Int64	0%
Restaurant	Object	0%
Sales	Object	0%
YOY_Sales (%)	float64	0%
Units	int64	0%
YOY_Units (%)	float64	0%
Unit_Volume	int64	0%
Franchising	Object	0%
City	Object	0%
State	Object	0%

III. Data Set Summary Statistics

The statistical variables are Sales, YOY_Sales, Units, YOY_Units, and Unit_Volume.

Table 2: Summary Statistics for Top 50

Variable Name	Count	Mean	Standard Deviation	Min	25th	50th	75th	Max
Sales	50	33.78	9.59653	20.0000	24.2500	34.5000	42.0000	49.0000
YOY_Sales	50	33.69800	24.54920	14.4000	20.9000	25.5000	33.8250	130.500
Units	50	34.7000	34.70000	7.00000	16.0000	27.0000	45.5000	105.000
YOY_Units	50	27.4460	23.23105	4.00000	14.3000	19.9000	32.6750	116.700
Unit_Volume	50	1592.600	994.6767	465.000	867.500	1260.00	2020.00	4300.00

Table 3: Proportions for City

City	Frequency	Proportion (%)
New York City	8	.16
Columbus	3	.06
Charlotte	2	.04
San Francisco	2	.04
Yorba Linda	1	.02
Douglas	1	.02
Irvine	1	.02
Huntington Beach	1	.02
Frisco	1	.02
Fairfax	1	.02
Fairburn	1	.02
Denver	1	.02
Doral	1	.02
Los Angeles	1	.02
Conway	1	.02
Blue Bell	1	.02
Belmar	1	.02
Atlanta	1	.02
Anaheim	1	.02
Kettering	1	.02
McAllen	1	.02
Louisville	1	.02
Wilmington	1	.02
Washington	1	.02
Wall Township	1	.02
Spartanburg	1	.02
Seattle	1	.02
Scottsdale	1	.02
San Roman	1	.02

Plano	1	.02
Pasadena	1	.02
Orlando	1	.02
Orange Park	1	.02
Omaha	1	.02
Olivette	1	.02
Memphis	1	.02
Medford	1	.02
Mechanicsburg	1	.02
Agoura Hills	1	.02

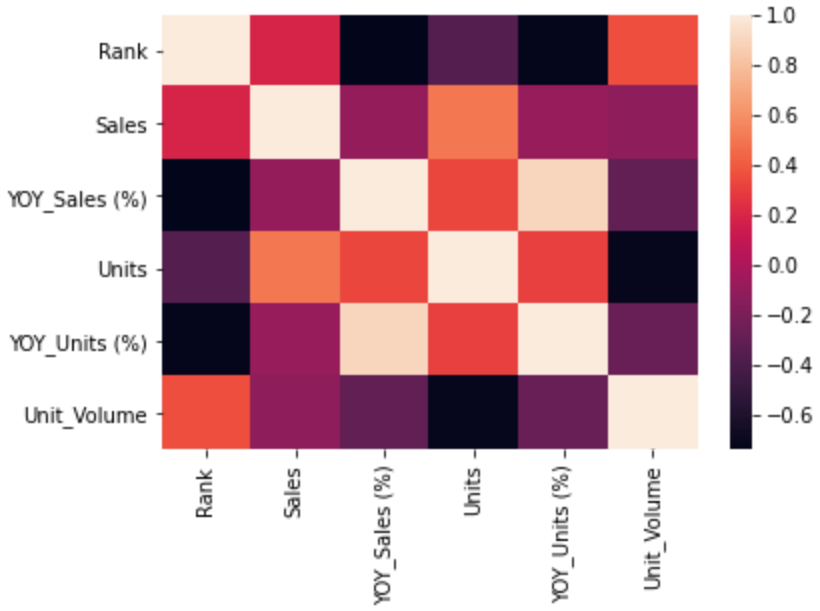
State	**Frequency**	**Proportion (%)**
California	10	.20
New York	8	.16
Ohio	4	.08
Texas	3	.06
Florida	3	.06
North Carolina	3	.06
Georgia	3	.06
New Jersey	2	.04
Pennsylvania	2	.04
Kentucky	1	.02
Arkansas	1	.02
Colorado	1	.02
D.C.	1	.02
Washington	1	.02
Montana	1	.02
Virgina	1	.02
Nebraska	1	.02
Oregon	1	.02
South Carolina	1	.02
Tennessee	1	.02

Franchising	**Frequency**	**Proportion**
Yes	29	.58
No	21	.42

Table 4: Correlation Table

	Rank	**Sales**	**YOY_Sales**	**Units**	**YOY_Units**	**Unit_Volume**
Rank		0.187535	-0.739141	-0.358137	-0.723734	0.359759
Sales	0.187535		-0.092597	0.504153	-0.085280	-0.117048
YOY_Sales	-0.739141	-0.092597		0.331898	0.902171	-0.301664
Units	-0.358137	0.504153	0.331898		0.306347	-0.712994
YOY_Units	-0.723734	-0.085280	0.902171	0.306347		-0.276040
Unit_Volume	0.359759	-0.117048\	-0.301664	-0.712994	-0.276040	

Heatmap:

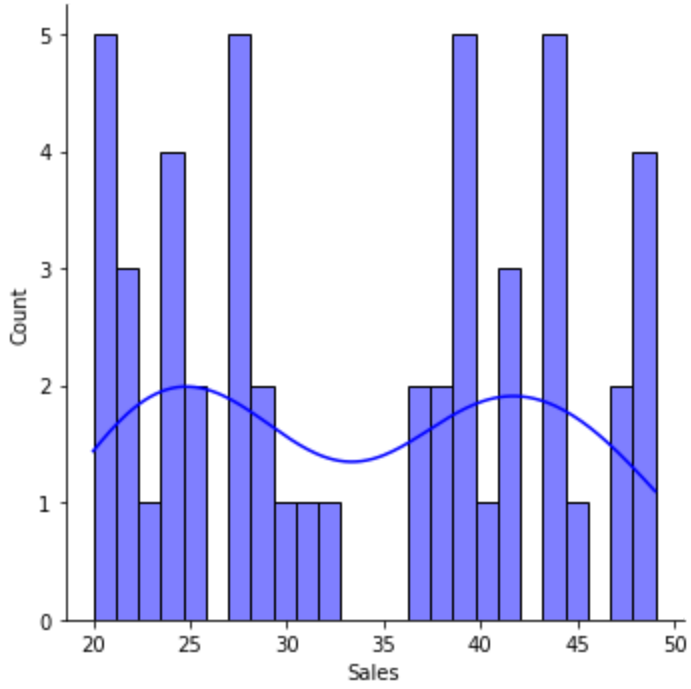


IV. DATA SET GRAPHICAL EXPLORATION

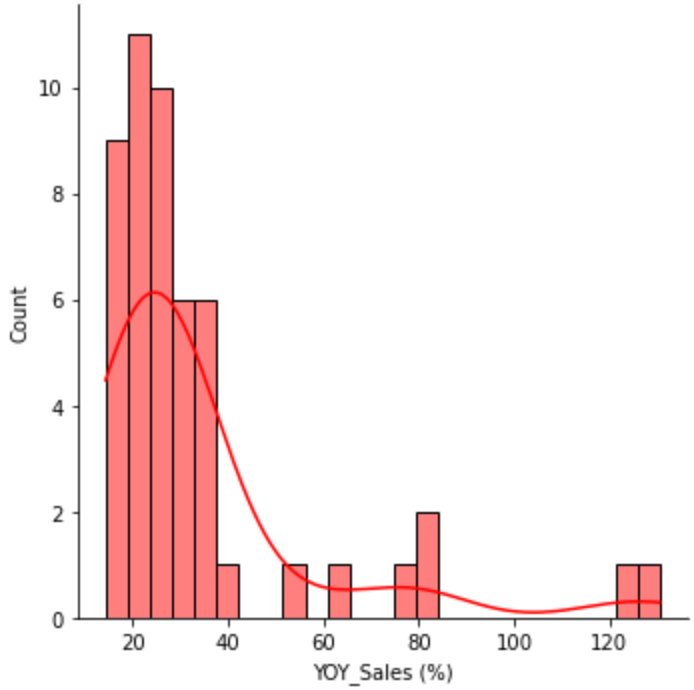
Narrative introduction to the section. In each section below, indicate any interesting distributions, anomalies, imbalance, etc. that you notice.

A. Distributions

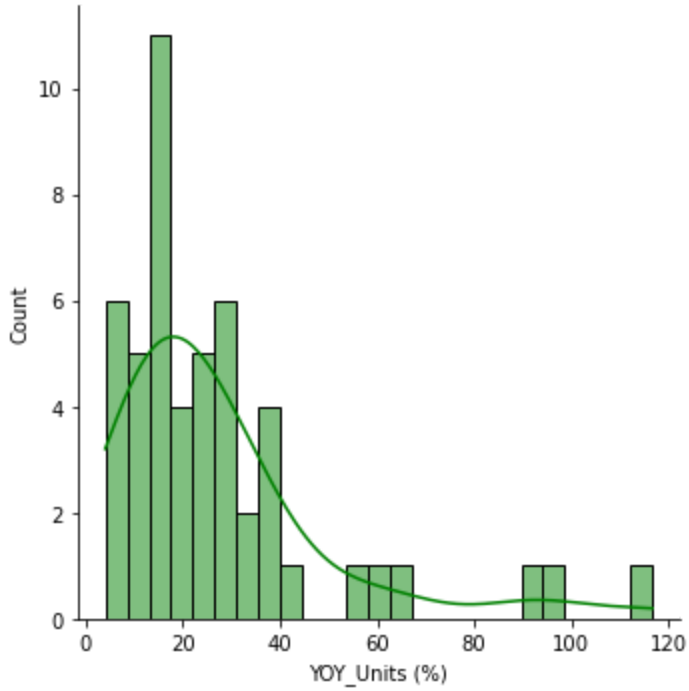
Sales Distribution:



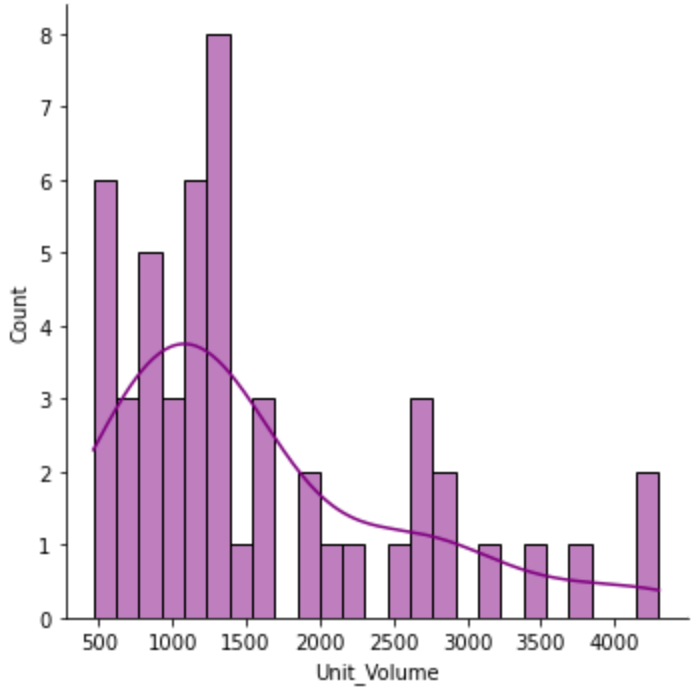
YOY_Sales Distribution:



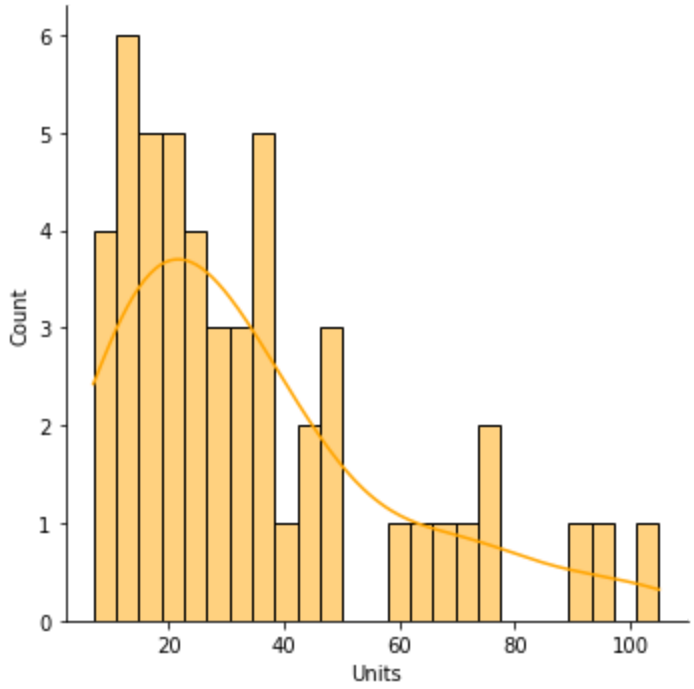
YOY_Unit Distribution:



Unit Volume Distribution:

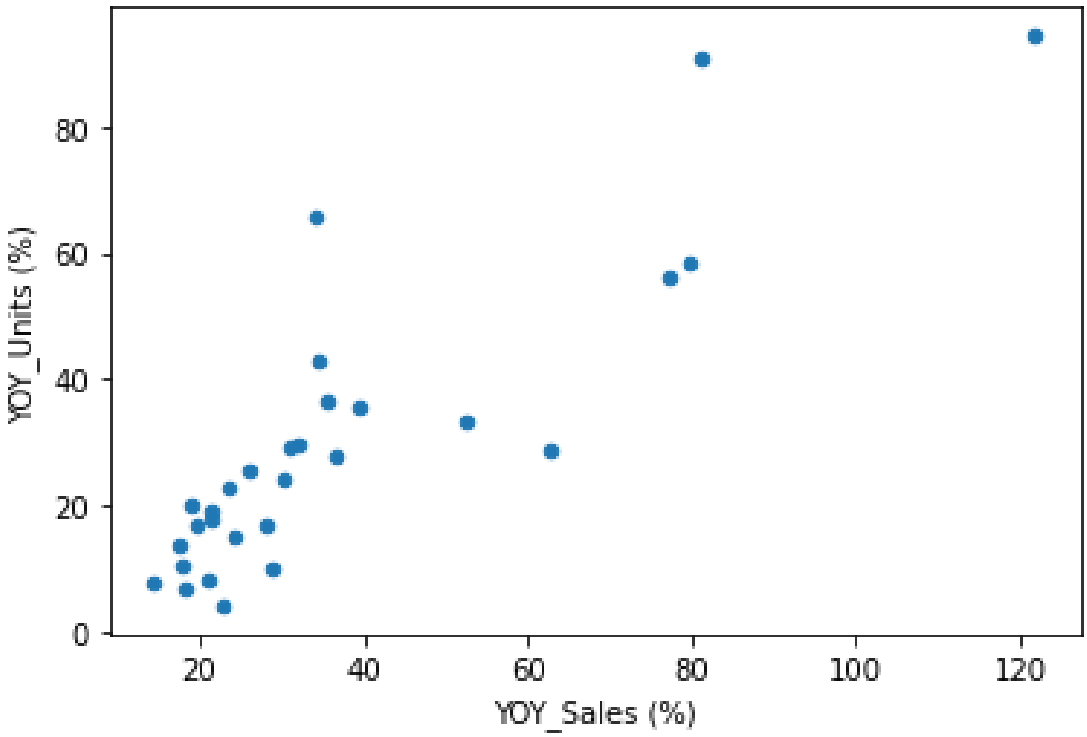


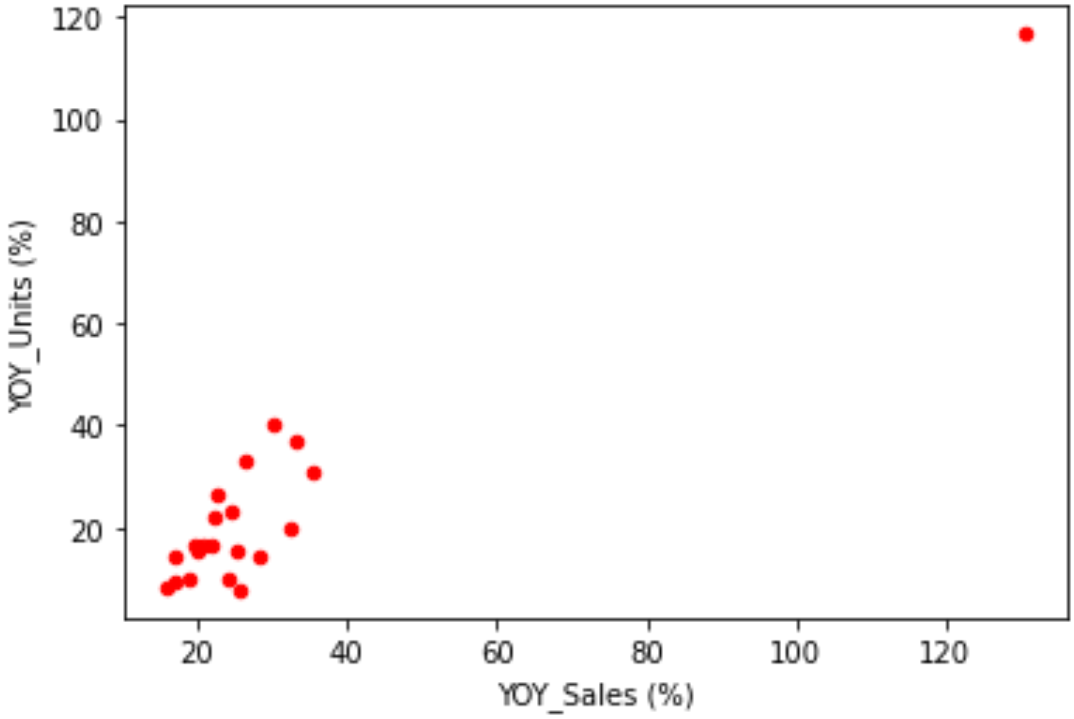
Unit Distribution:

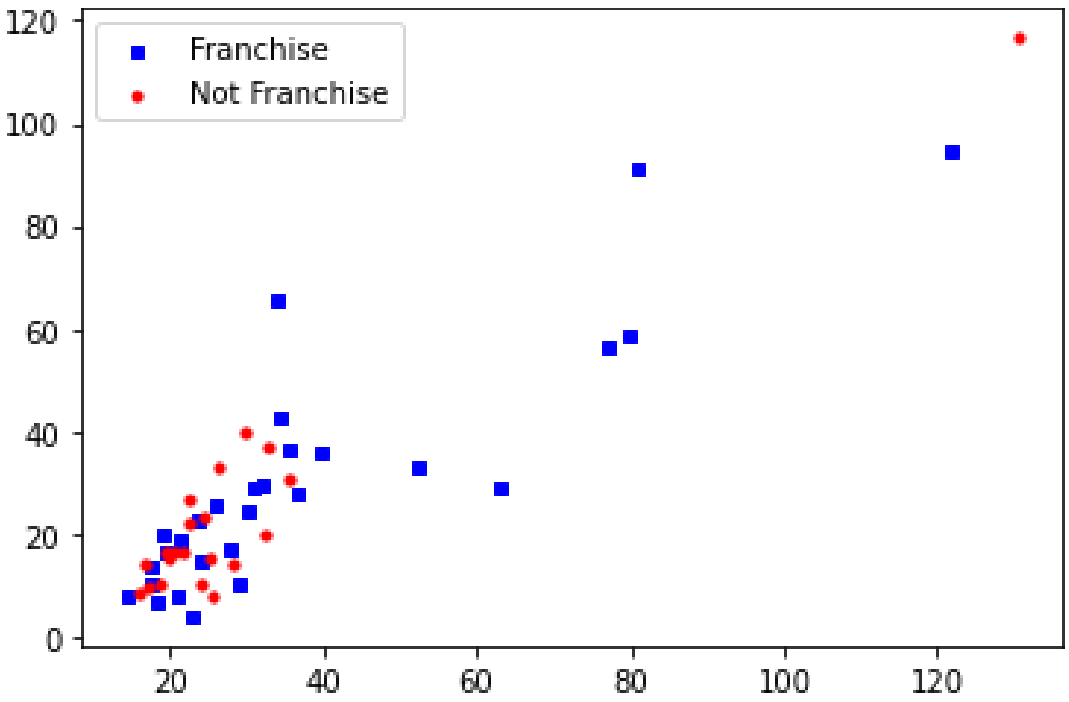


B. Scatter Plots

YOY_Sales and YOY_Units correlate heavily, so let's align those in a scatter plot.

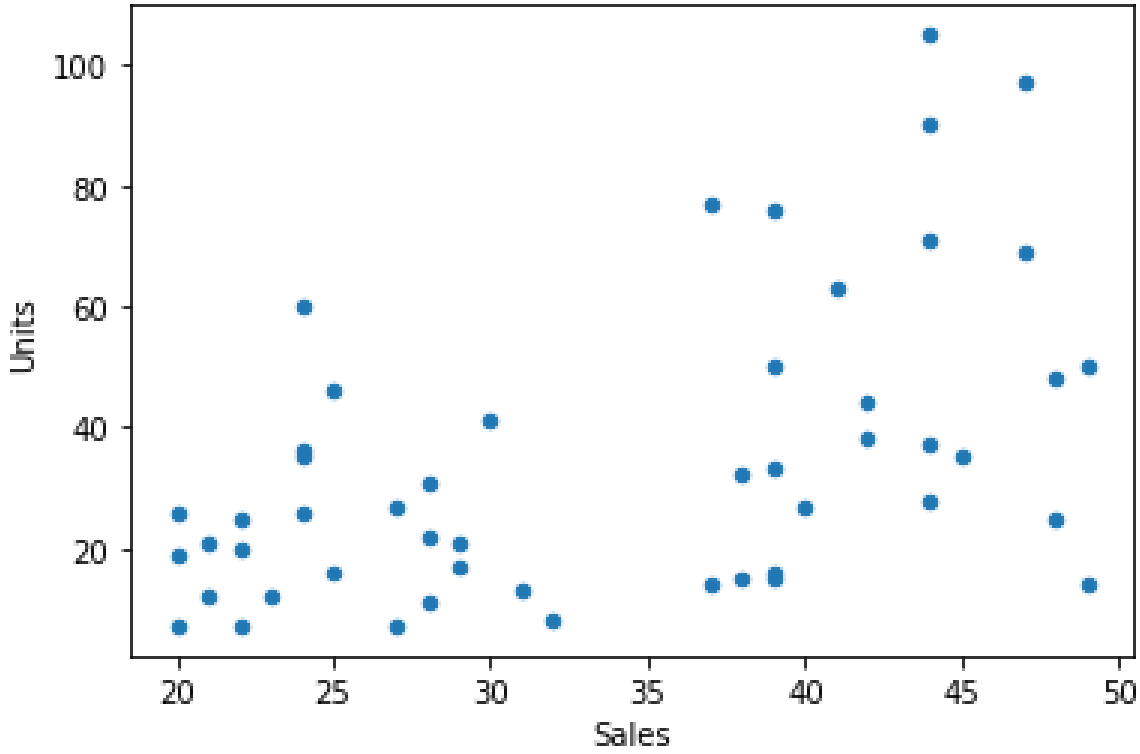






This dataset prioritizes rate of growth over size, so the top right corner, Rank 1 "Evergreens," while having fewer locations and not being a franchise, outperforms nearly everyone by a large margin. On average though, the **not** franchises seem to struggle more in YOY, which makes sense especially when considering the number of stores.

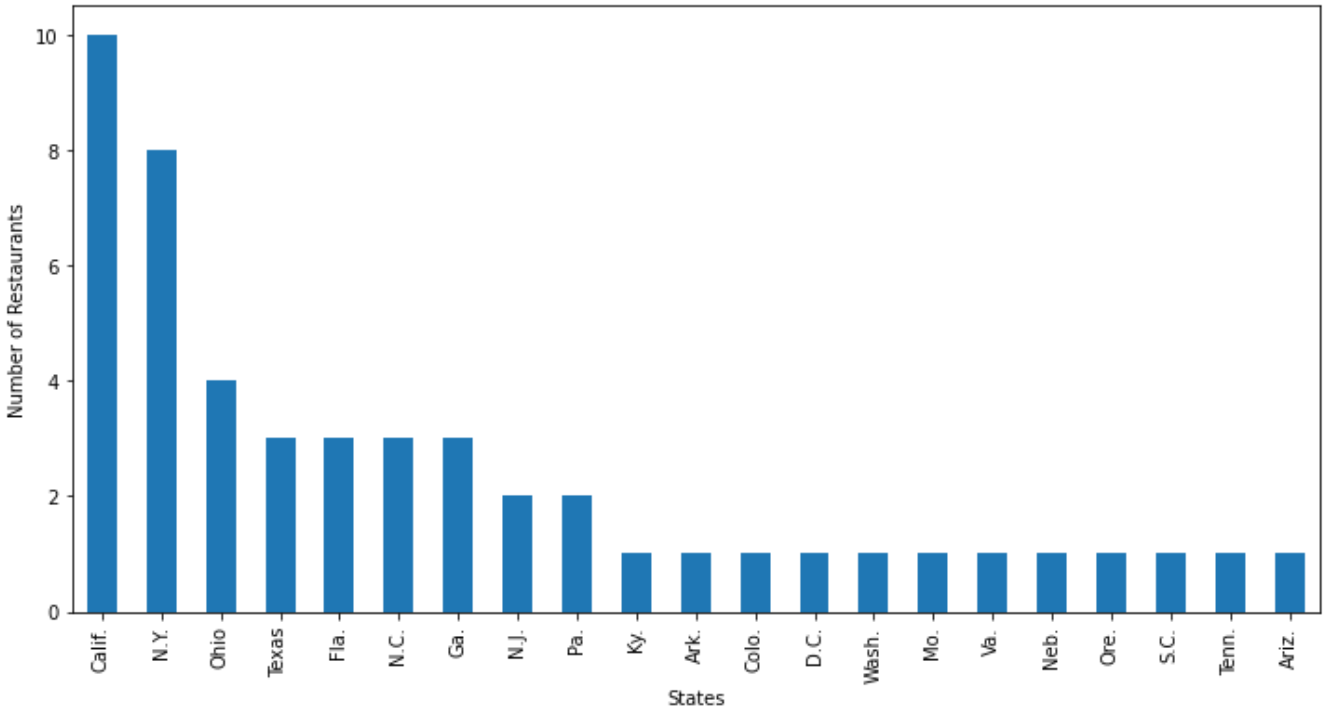
The next highest correlation is Sales and Units:



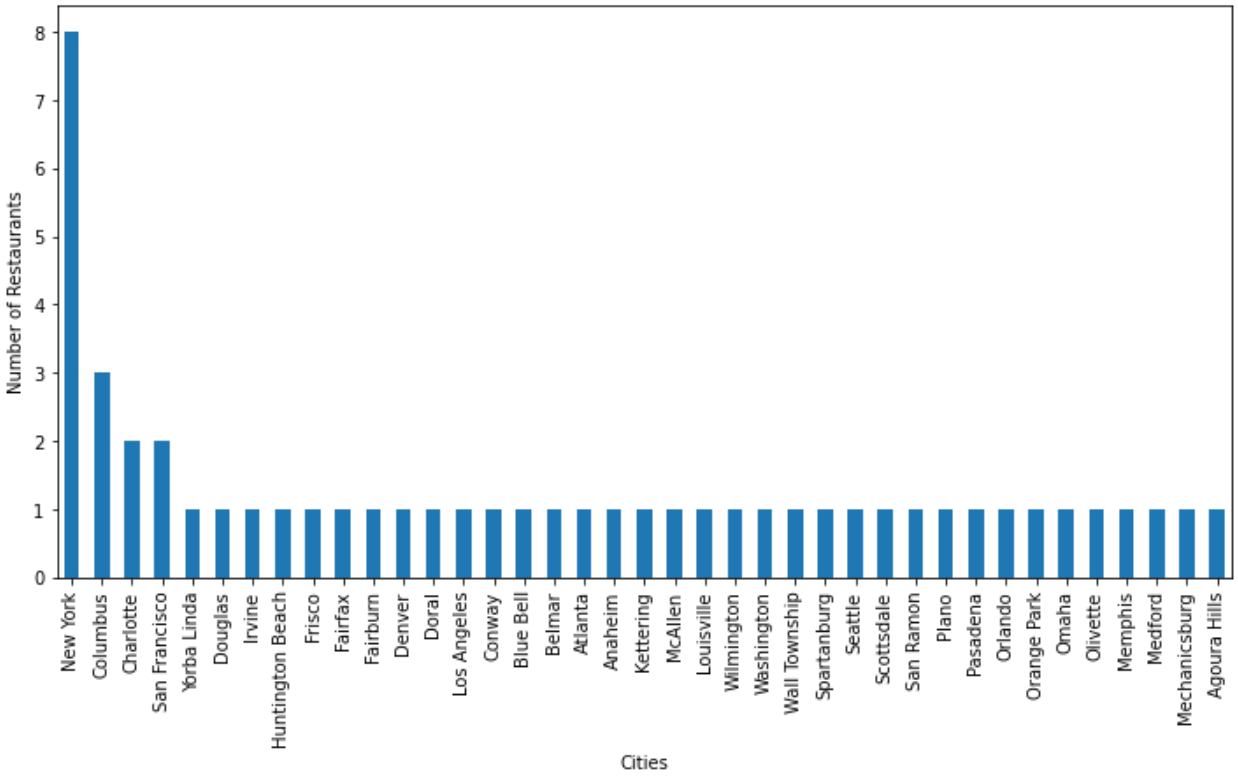
This graph tells you that typically having more restaurants means more sales. Not much of a mystery there, but the outliers here can be pretty interesting. Take the bottom right corner, one of the lowest number of units, but the actual sales are skyrocketing. Or conversely, take the top leftmost point, a fairly high number of units, but still on the lower end of sales.

C. Barcharts

Here's a barchart showing the states with the most restaurants.

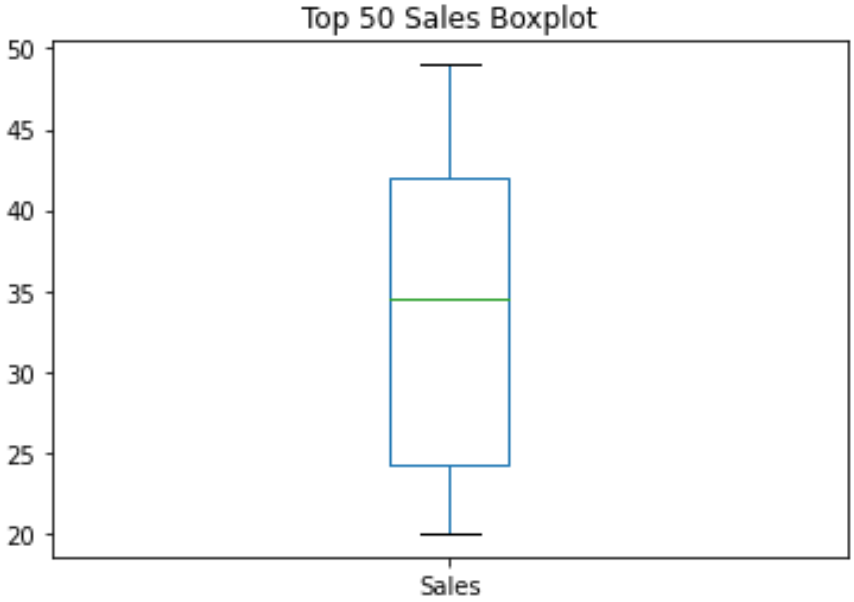


Here's a version for cities:

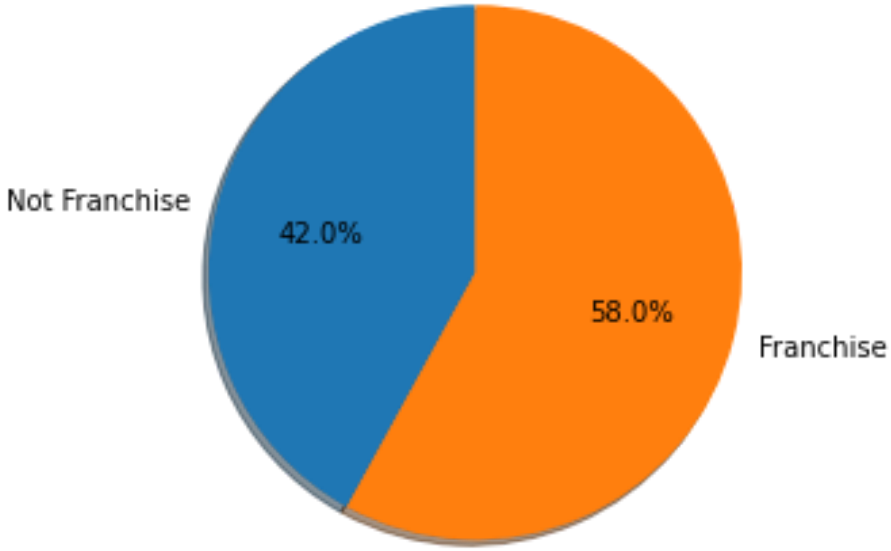


D. Other Plots

Here's a boxplot of Sales:



Pie chart of Franchise vs Not Franchise:



V. SUMMARY OF FINDINGS

This dataset focused on growth above size, which allowed it to highlight restaurants that might be attempting new strategies in the restaurant meta. The highest growing restaurant, Evergreens, is on the smaller side of units, but their YOY_Sales is increasing rapidly, 130%. California is the hot seat for growth in restaurants, dominating with 20% of this list being located within the state. Though New York City does win it out when taking a city by city look, having 16% of this list. Noticeably, other major cities in the US such as Chicago, Seattle, any of the major Texan cities, are all lacking in the same ferocity as New York and California. It's also known that most restaurants on this list are franchises, especially those which have the highest sales.