coffee shop sales

PROJECT OBJECTIVES:

The objectives of this project are to see some insight into the popularity of coffee shops, including what they sell the most of and when they have the most business.

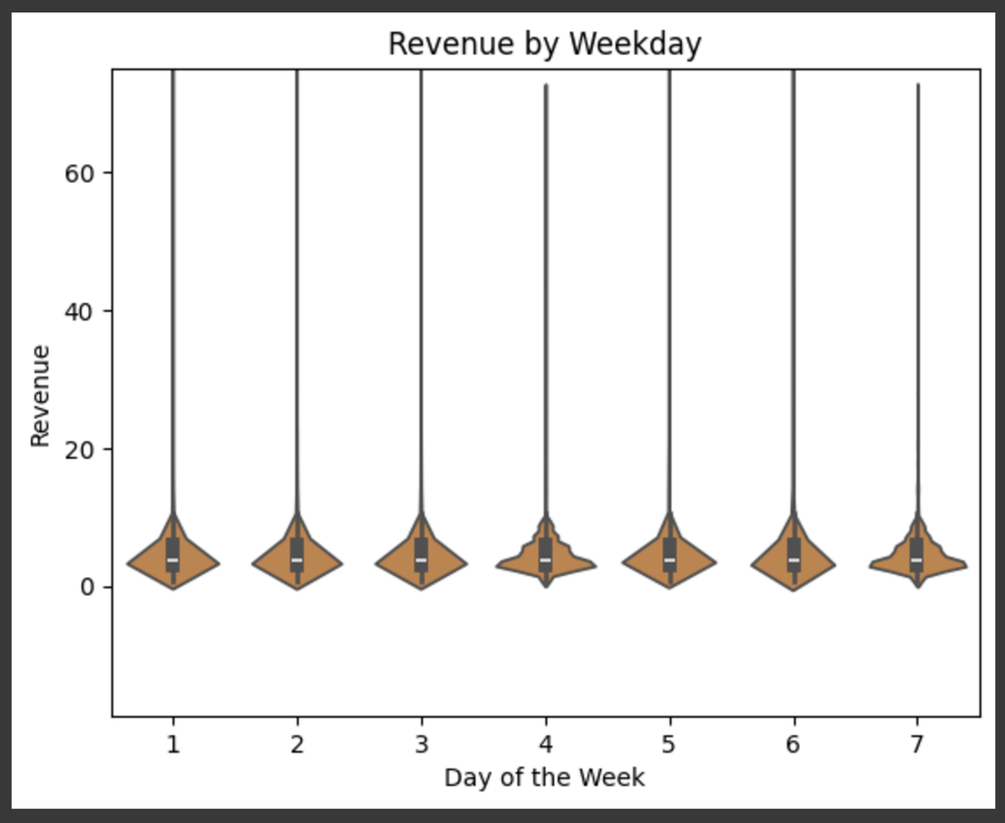
DATASET METADATA:

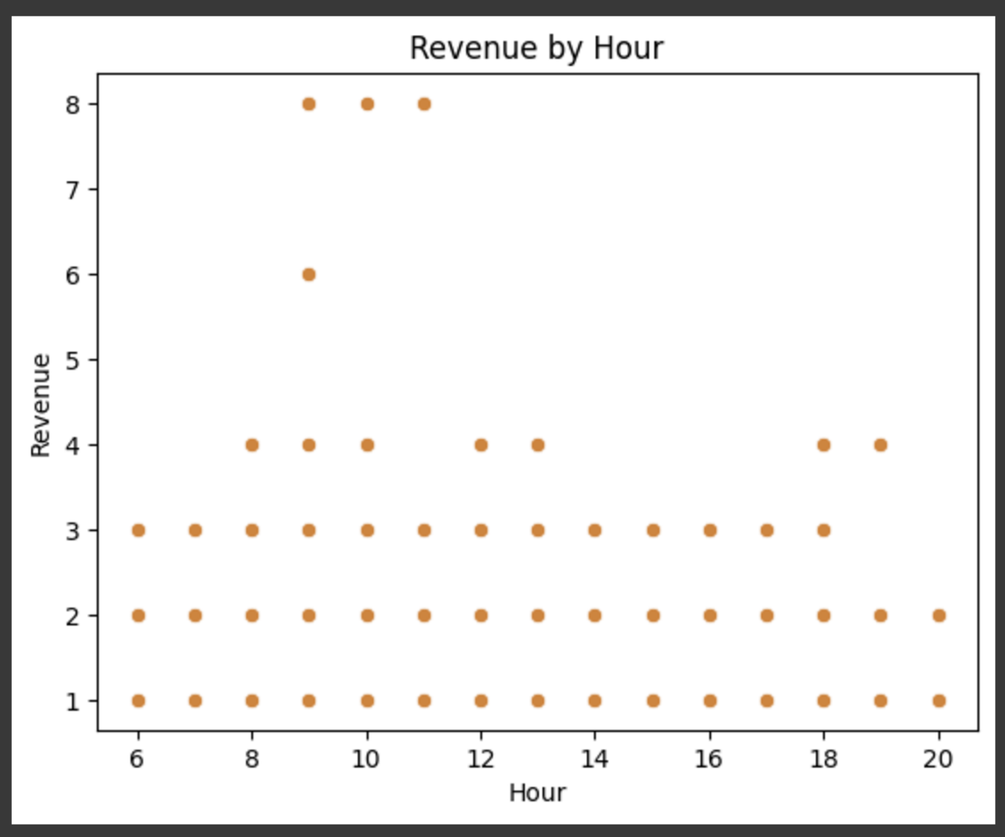
	transaction_id	transaction_qty	store_id	product_id	unit_price	Revenue	Month	Weekday	Hour
count	149116.000000	149116.000000	149116.000000	149116.000000	149116.000000	149116.000000	149116.000000	149116.000000	149116.000000
mean	74737.371872	1.438276	5.342063	47.918607	3.382219	4.686367	3.988881	3.982336	11.735790
std	43153.600016	0.542509	2.074241	17.930020	2.658723	4.227099	1.673091	1.996650	3.764662
min	1.000000	1.000000	3.000000	1.000000	0.800000	0.800000	1.000000	1.000000	6.000000
25%	37335.750000	1.000000	3.000000	33.000000	2.500000	3.000000	3.000000	2.000000	9.000000
50%	74727.500000	1.000000	5.000000	47.000000	3.000000	3.750000	4.000000	4.000000	11.000000
75%	112094.250000	2.000000	8.000000	60.000000	3.750000	6.000000	5.000000	6.000000	15.000000
max	149456.000000	8.000000	8.000000	87.000000	45.000000	360.000000	6.000000	7.000000	20.000000

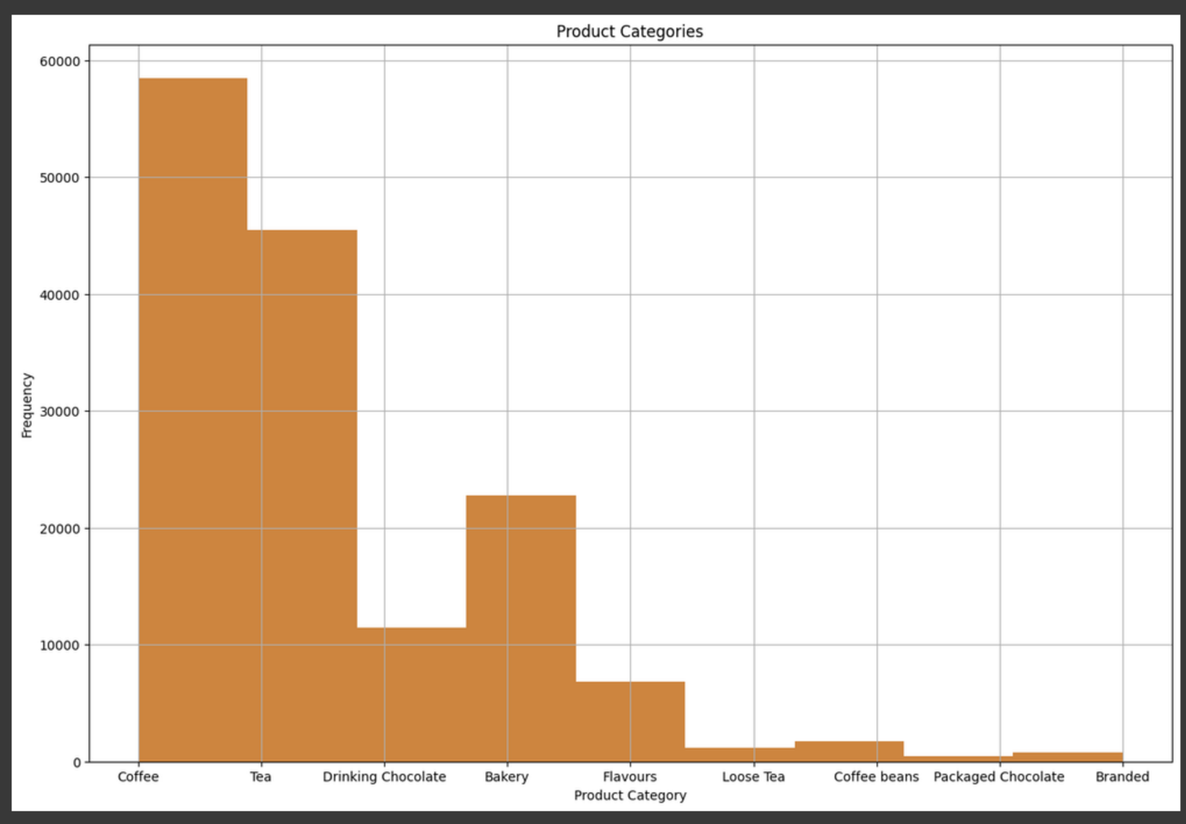
INFORMATION FOUND:

I found that while coffee is the most frequent purchase, only 39% of total purchases are coffee drinks, and typically no one buys more than two drinks at a time.

Revenue made stays consistent among hours of the day and also days of the week, which I was not expecting to see. While there are peaks during business hours, I had assumed there would be drops during off hours which there aren't.







TOOLS USED:

I downloaded the dataset as a csv file from kaggle and used google colab to run the python commands needed to analyze the data.

I read the csv file using pandas, and the .describe() command to find the metadata. I used a combination of pandas, seaborn, and matplotlib.pyplot to visualize the dataset in the form of a histogram, scatterplot, and violin plot.