Taco CSS Case Study

Introduction

The goal of this case study is to analyze the dataset for the Taco CSS and identify what CSS is doing that benefits Taco CSS's operations along with discrete gaps in Taco CSS's performance. This analysis will serve to improve CSS's ability to improve Taco CSS's operational performance in order to continue driving revenue growth and retaining their business.

Attached to this case study will be all code (Python and SQL) used to to clean and analyze the given dataset.

General Insights

Doordash dominates UberEats in both revenue and order volume:

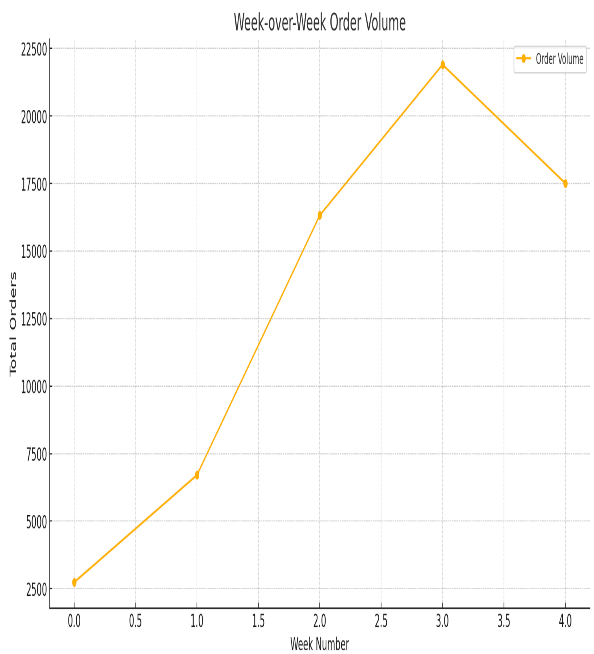
- Nearly $85 billion vs $357 million in revenue.
- $4 billion orders vs ~15.36 million orders.

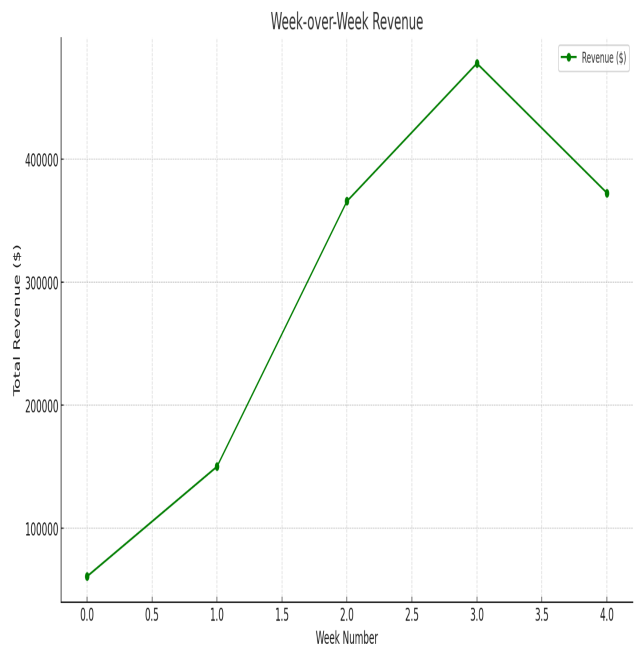
In general it is safe to assume that specifically for Taco CSS, more customers are ordering through Doordash than UberEats.

	delivery platform	total_revenue	total_orders
1	doordash	84881384679.76399	"3898503844"
2	ubereats	357385914.8999891	"15256836"

Over both platforms customer satisfaction averages around a 4.95, indicating customers are generally happy with deliveries from Taco CSS. This highlights an overall success in both the quality of the orders and the service from drivers regardless of the platform they ordered on.

Weekly Performance

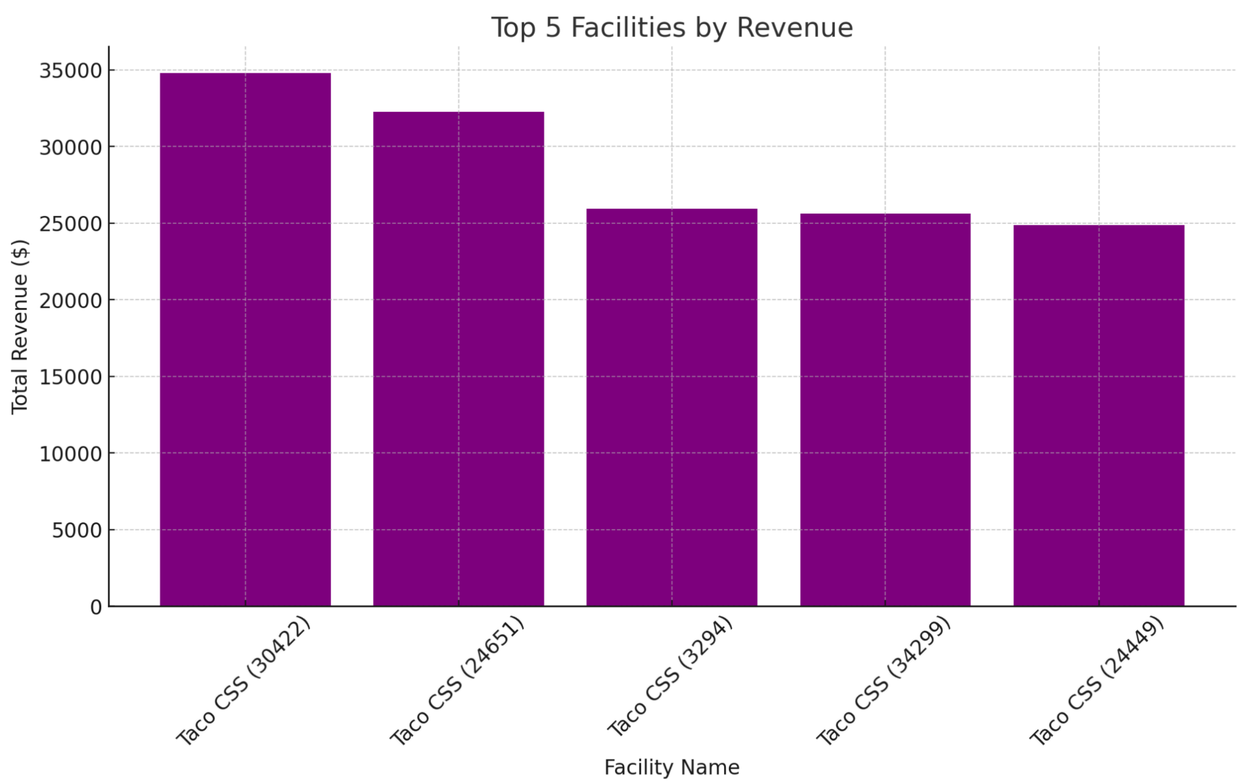




The weekly performance is based on a 4 week stretch over March 2023. In this analysis it is assumed this growth is close to the implementation of Otter for Taco CSS. It is clear positive trend over the first 3 weeks that seems to begin to stabilize around the same level as the week 2 level. However, it is important to note while we do have a lot of data, the time range can be much larger allowing us to more accurately analyze week over week performance. While we have clear evidence of growth here, we can't attribute the success purely based on Otter's implementation, there are mare many known variables that could be attributed to the success of Taco CSS.

Top Performing Locations

	facility_name	facility_timezone	total_orders	total_revenue	avg_rating
1	Taco CSS (30422)	America/Chicago	1612	34788.77000000006	4.929280397022333
2	Taco CSS (24651)	America/New_York	1402	32257.32000000006	4.90228245363766
3	Taco CSS (3294)	America/New_York	1185	25945.199999999983	4.89957805907173
4	Taco CSS (34299)	America/New_York	1278	25642.320000000025	4.944444444444445
5	Taco CSS (24449)	America/New_York	1173	24878.970000000038	4.898550724637682

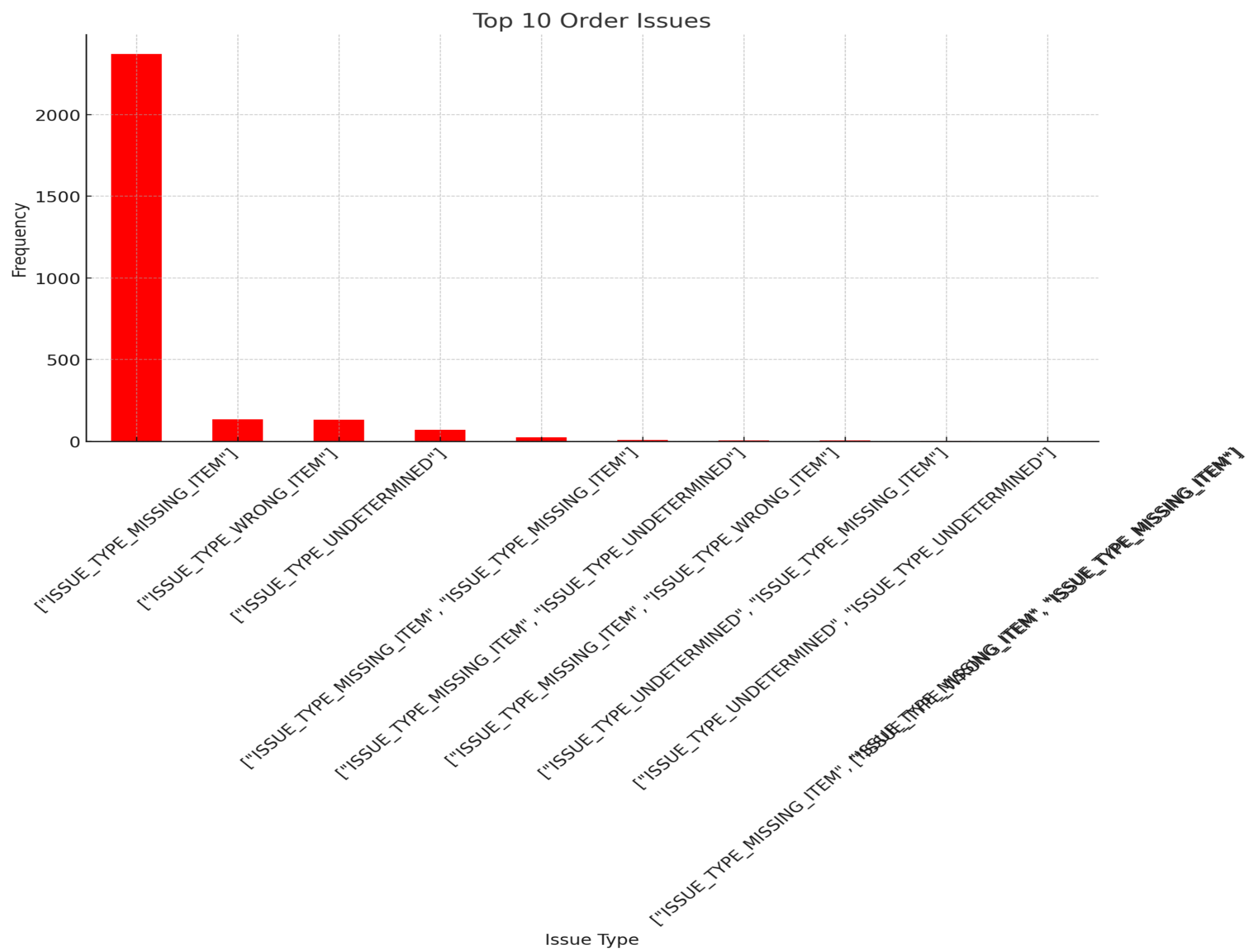


These are the top performing locations by Revenue and while we have an overall average rating of 4.5 it is clear that the top performing facilities have at least a 4.9 rating. It is also important to note that these facilities probably are generating more revenue due to the density of the city they are located in. I specifically included timezone to narrow down which coast performed best as well and it is clear the east coast, which is comprised of many denser cities is the clear winner. However overall, Chicago seems to perform better than all of those but it is important to remember Chicago is the biggest city in the midwest and there are not many competing cities in the Central time zone that compete with Eastern otherwise.

Order Issues

I initially started with a general overview of issues across all facilities. This will show a clear common issue across the board. It is also important to determine which locations are the most problematic. I wanted to see if just like the top performing locations, the locations with the most issues would correlate with the locations that received the most orders as because the issues were higher, customers could request a conditional refund.

	order_issue_types	issue_count
1	["ISSUE_TYPE_MISSING_ITEM"]	2372
2	["ISSUE_TYPE_WRONG_ITEM"]	135
3	["ISSUE_TYPE_UNDETERMINED"]	132
4	["ISSUE_TYPE_MISSING_ITEM", "ISSUE_TYPE_MISSING_ITEM"]	71
5	["ISSUE_TYPE_MISSING_ITEM", "ISSUE_TYPE_UNDETERMINED"]	24
6	["ISSUE_TYPE_MISSING_ITEM", "ISSUE_TYPE_WRONG_ITEM"]	9
7	["ISSUE_TYPE_UNDETERMINED", "ISSUE_TYPE_MISSING_ITEM"]	5
8	["ISSUE_TYPE_UNDETERMINED", "ISSUE_TYPE_UNDETERMINED"]	5
9	["ISSUE_TYPE_MISSING_ITEM", "ISSUE_TYPE_MISSING_ITEM", "ISSUE_TYPE_MISSING_ITEM"]	3
10	["ISSUE_TYPE_WRONG_ITEM", "ISSUE_TYPE_MISSING_ITEM"]	1



Based on the data across all locations it is clear that customers not receiving an item (indicated by “MISSING_ITEM”) is the clear dominator across the board of all issues. The data presents 10 issues but upon further inspection there are really 3 issues and some happen together. These issues are missing item, wrong item, and undetermined. It makes sense that individual issues happen more than grouped issues so the data doesn’t jump out too much. However a missing item vs a wrong item are 2 vastly different issues that I will cover resolution strategies on in the following section.

When grouping the issues by each facility we see a MOSTLY expected correlation between issues and orders a facility receives.

	facility_name	order_issue_types	facility_timezone	issue_count
1	Taco CSS (24651)	["ISSUE_TYPE_MISSIN...	America/New_York	113
2	Taco CSS (3294)	["ISSUE_TYPE_MISSIN...	America/New_York	82
3	Taco CSS (1938)	["ISSUE_TYPE_MISSIN...	America/New_York	77
4	Taco CSS (24449)	["ISSUE_TYPE_MISSIN...	America/New_York	71
5	Taco CSS (34299)	["ISSUE_TYPE_MISSIN...	America/New_York	68

However it is interesting to note that Chicago (facility 30422) doesn't make the top 5 list at all. There are many theories that can be made out of this but one conclusion we can almost certainly say is that while there is a correlation between order volume and issues, there isn't a strong one.

Another problem to note is high order cancellations at some facilities.

	facility_name	total_orders	cancellations	cancel_rate
1	Taco CSS (30409)	967	"100"	10.341261633919338
2	Taco CSS (26084)	656	"52"	7.926829268292683
3	Taco CSS (2260)	593	"44"	7.419898819561552
4	Taco CSS (27361)	434	"30"	6.912442396313364
5	Taco CSS (24449)	1173	"73"	6.223358908780904

These are our top 5 facilities where cancellations are being recorded. Based on the average delivery times it seems the average order acceptance to delivery start time is low so its unapparent what exactly the issue may be here.

	facility_name	total_orders	cancellations	avg_delivery_time	cancel_rate
1	Taco CSS (30409)	944	"96"	3.4311440677966103	10.169491525423728
2	Taco CSS (26084)	639	"50"	3.023474178403756	7.82472613458529
3	Taco CSS (27361)	425	"28"	3.364705882352941	6.588235294117647
4	Taco CSS (2260)	555	"35"	2.6342342342342344	6.306306306306307
5	Taco CSS (24449)	1130	"71"	2.9123893805309735	6.283185840707965

Suggestions

First and foremost based on the issues extracted from the dataset, it is clear that there are way more “Missing Items” than “Wrong Items”. If the inverser were true this would be an easy problem to solve as its always the restaurant’s fault. However a missing item can indicate that

- A wrong item was given so the correct item may be indicated as missing also.
- The restaurant didn’t give a wrong item, only a missing item.
- Since the item was “missing” the customer can’t prove that with a picture so a customer can be dishonest about a missing item.
- The driver could have taken the item so the customer never received it.
-

Clearly it is very easy to get an item marked as missing whether its by genuine fault, or dishonesty by either the courier or the customer. In order to solve this there are multiple measures that can be taken.

1. Verify to the customer each item is in the bag prior to the driver receiving it
2. Seal the bags in such a way that once

Doordash has a much higher order rate than UberEats. Why is there such a heavy gap there? While Doordash does have a higher usage rate, to have more UberEats customers order from TacoCSS we can provide UberEats specific discounts. This will bring in customers looking at other restaurants on UberEats to order from Taco CSS and also trickle users from other platforms (including Doordash). This is helpful because while it provides an overall increase in orders, it also allows us to collect more data on how UberEats service and courier perform when scaling order volume.

Regarding the cancellations I believe we need this dataset to be a little more complete so we can determine multiple things specific to this customer. Are there specific, consistent items missing, is the facility in a far location from most orders? These are just simple questions but more data can help us have a better understanding of what we need to do at facilities with over a 5% cancellation rate.

Code

GitHub

- GitHub
    - Contains all Python Code

SQL:

- All SQL here: SQL Queries
- Cleaned Dataset which is referred to as the “sales” table is cleaned_taco_final.csv