DoorDash – Delivery Data Analysis and Business Insights

1. Project Objective

This project aims to analyze delivery data at Grocery, Convenience, Alcohol stores on DoorDash platform and then generate business insights along with recommendations for potential improvements in marketing and supply chain management.

2. Data Summary - Data Cleaning and Preparation

The dataset in this analysis contains 60,583 records at order item level of one-month' worth of delivery data for the New Business Verticals (Grocery, Convenience, Alcohol, and DashMart) at DoorDash. The glossary can be found here.

Data Summary

Submarket location	Cincinnati, Ohio		
Time period	09/15/2022 - 10/14/2022		
Number of stores	4 stores (1 DashMart and 3 Grocery stores)		
	All	**Without Dasher ID**	
		Canceled Deliveries	Non-Canceled
Total order items	60,583	669	340
Total order deliveries	13,085	205	145
Total order value	$301,507	$3,361	$1,474

Data Cleaning

Among 13,085 unique deliveries made, there are 350 deliveries not having Dasher ID information:

- 205 deliveries ($3,361 in order value) are canceled orders
- 145 deliveries ($1,474 in order value) are non-canceled, could be pick-up orders

As the goal of this analysis is to understand deliveries made by dashers and identify opportunities to improve the operation, these 350 orders will be excluded from the analysis. This should not significantly impact the analysis results since the exclusion only accounts for roughly 1.6% of total order value.

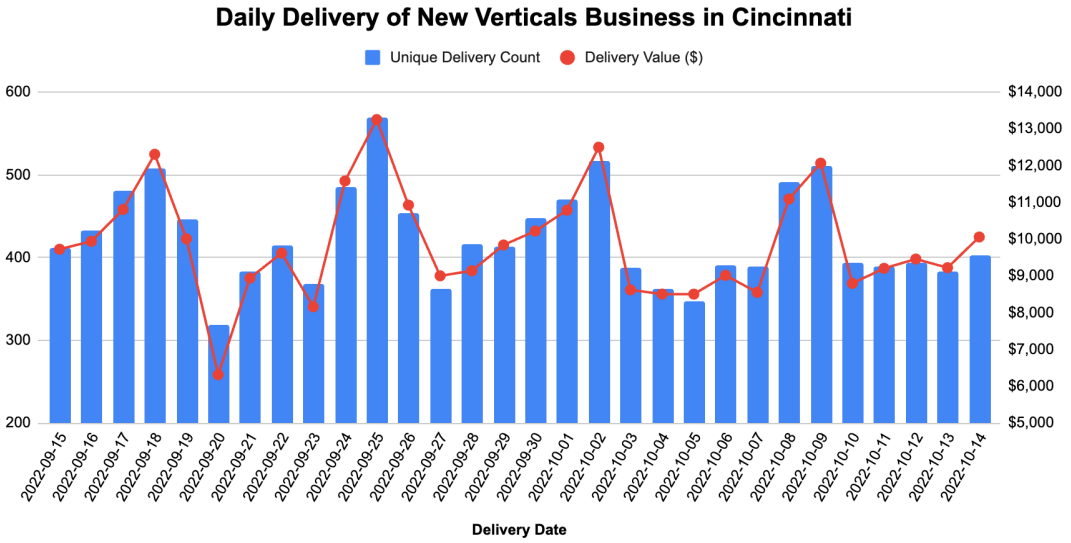
(The process of cleaning data and loading data to MySQL is executed in Python.)

3. Analysis Details

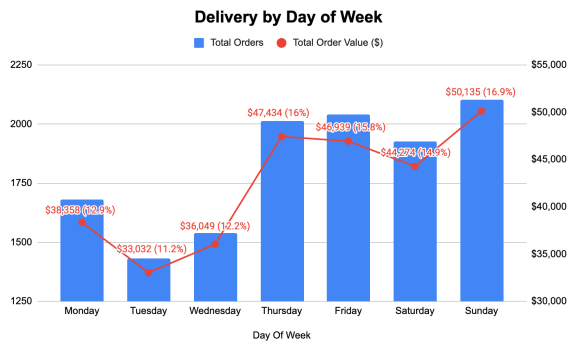
Daily Delivery Frequencies

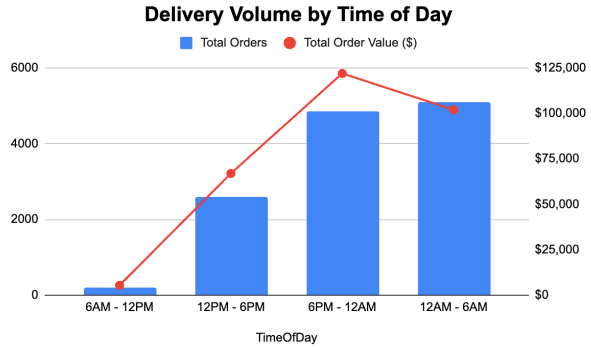
Insights

- Overall, there is a cyclical pattern in daily deliveries throughout the course of the month, with the amount of deliveries always peaked on Sundays.



- High order volumes happened on Thursdays to Sundays, especially at night time (after 6PM). The total order value made on Sundays accounted for 17% (~$50K) of the total of this 30-day period, the most out of all days of week.





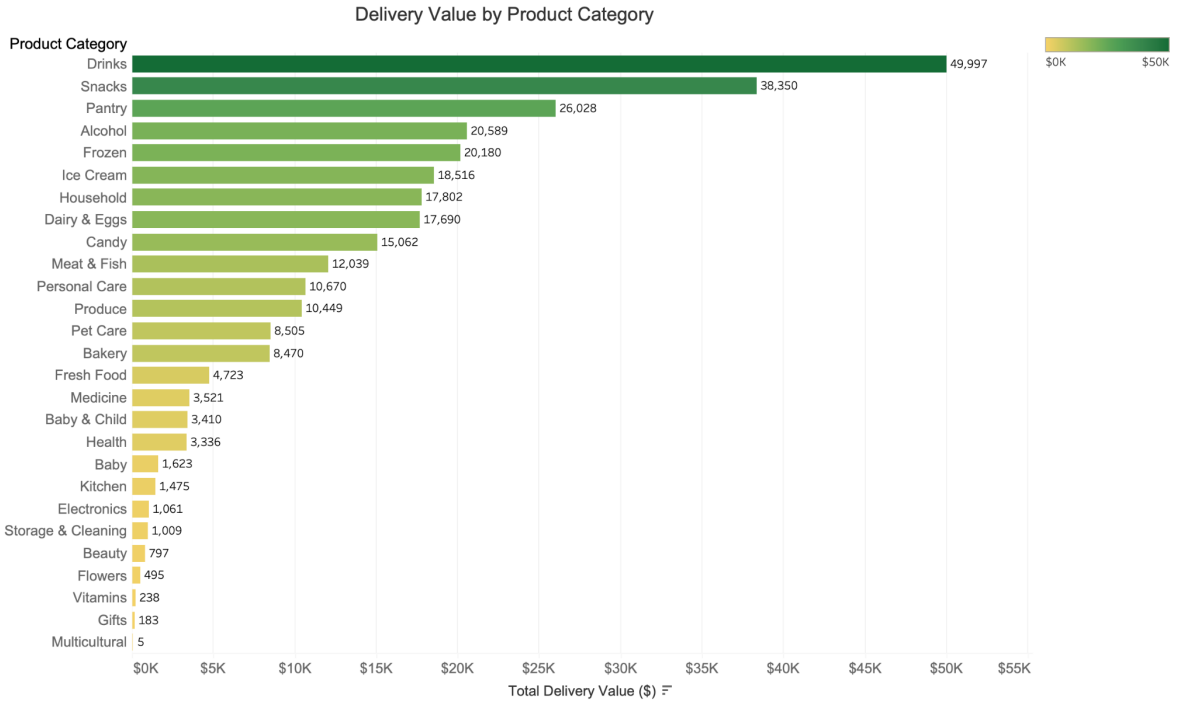
Recommendation:

- Incentivize delivery orders on Monday - Wednesday by offering deals and/or lowering delivery costs for earlier in the day through emails or push notifications
- Work closely with stores to optimize staffing and inventory allocation accordingly to ensure meeting demand during high-demand times (nighttime, Thursday-Sunday)

Most frequently ordered products

Insights

- The top 5 categories are Drinks, Snacks, Pantry, Alcohol, and Frozen products.



- While Snacks and Pantry were mostly ordered on Sundays, Alcohol and Drinks were the 2 popular products on Fridays.

	Delivery Day						
Product ..	Monday	Tuesday	Wednes..	Thursday	Friday	Saturday	Sunday
Drinks	$6,305	$5,547	$6,159	$8,010	$8,368	$7,454	$8,154
Snacks	$4,784	$4,344	$4,822	$5,975	$6,138	$5,822	$6,467
Pantry	$3,467	$2,850	$2,977	$4,038	$3,549	$3,986	$5,162
Alcohol	$2,294	$2,056	$2,070	$3,051	$4,211	$3,589	$3,318
Frozen	$2,484	$2,423	$2,358	$3,517	$3,048	$2,911	$3,440

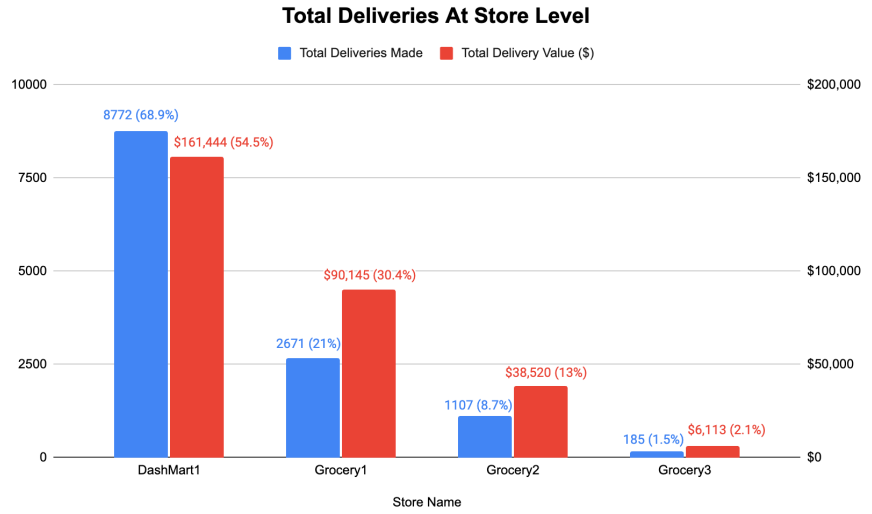
Recommendation:

- Focusing marketing efforts through promotional strategies on these product categories. Offer pantry/snacks promotions on Sunday and notify the promotion on the DoorDash app some time after 6PM (peak time) to attract new customers or to remind returning customers of their shopping behavior.
- Ensure sufficient stocks for these popular products and schedule restocks to arrive before peak days. Prioritize the availability of alcohol/drinks products on Friday and of pantry/snacks on Sunday. Maintain a safety stock level for these product categories, especially on peak days and evening/night time.

Store Sales and Operation Performance

Insights

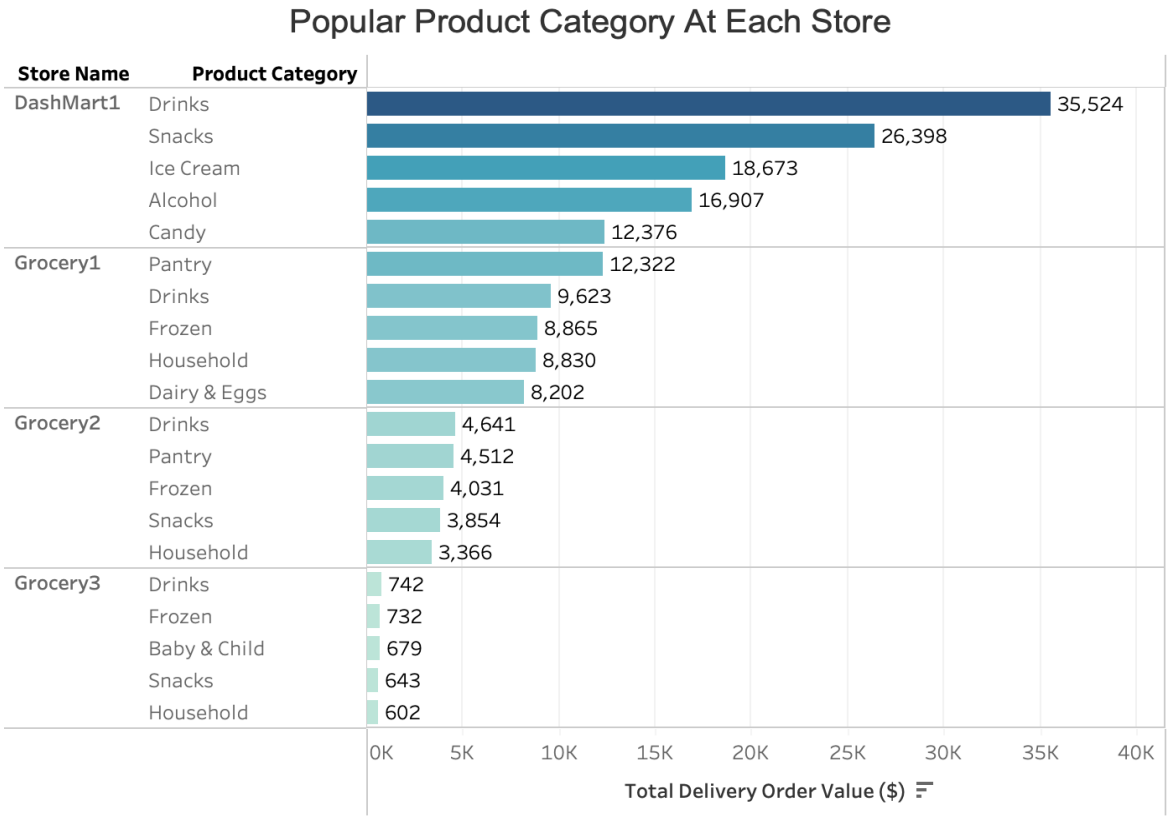
- DashMart accounts for about 69% of total deliveries made within the time period, therefore driving most of sales in Cincinnati in this time period.
- Grocery1 and Grocery2 take up about 21% and 8% of the month's total deliveries, respectively. Grocery3 is a low-performing store with the least amount of deliveries made (2%). The store could be a new business on DoorDash's platform with low customer awareness.



- DashMart1 possesses great operation excellence given high fulfillment rate (99.8%), low missing rate (0.2%), and high substitution matching rate (100%). However, the rate of having a substitution for missing items during the month was only 50%. There is an opportunity to increase this number to better fulfill orders.
- Other grocery stores maintained a similar operating level in fulfilling order (85%+) and substituting missing items (60%+). These metrics can be improved to enhance operation excellence at these grocery stores and bring customer satisfaction.

Store Name	Item Fulfilled Rate	Item Missing Rate	Substitution Rate	Substitution Match Original Category
Grocery1	85.4%	14.6%	70.8%	94.3%
Grocery2	87.4%	12.6%	70.0%	90.1%
Grocery3	87.4%	12.6%	62.3%	98.7%
DashMart1	99.8%	0.2%	50.7%	100%

- Even though the top 5 product categories are slightly different at each store, Drinks are usually the most bought product. Most alcohol deliveries were ordered at store DashMart1.



Recommendation:

- DashMart1 has the most deliveries in Cincinnati during the month. The store also served customers' demand well with a very low rate of missing items, indicating meeting demand ability and high level of customer satisfaction. We recommend driving traffic to DashMart by prioritizing DashMart results in app interface or search/recommendation engine and focusing on advertising/promotion for customer acquisition and retention. Leveraging its strength in alcohol deliveries, DashMart1 can focus on expanding the product range to cater to customer preferences. DoorDash should continue to prioritize and invest in this vertical to maintain and expand its market dominance.
- Recognizing the low performance of Grocery3, DoorDash can implement marketing and promotional strategies to increase visibility and attract local customers to this store.

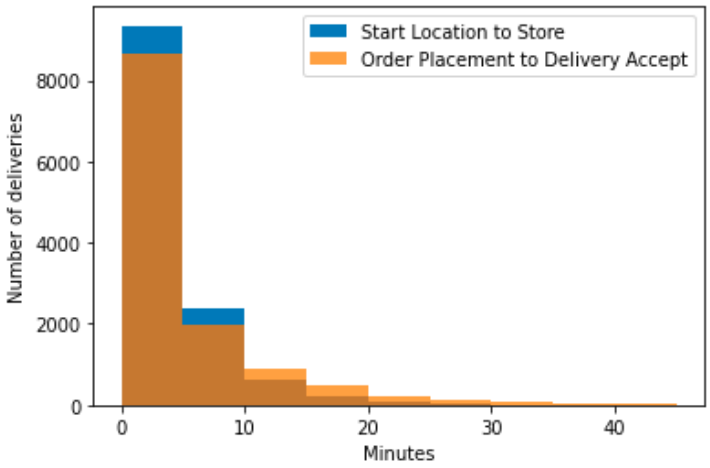
Delivery performance of dashers

A dasher after accepting a delivery request needs to travel from their starting location to the store to pick up the order, and then drive to the delivery location.

Insights:

- Customers reported an order with missing or incorrect items 1.63% of the time (~207 deliveries) out of the total of 12735 deliveries made in the month. There is also 4.7% of the time (~598 deliveries) dashers were late by 20 minutes in delivering the items.

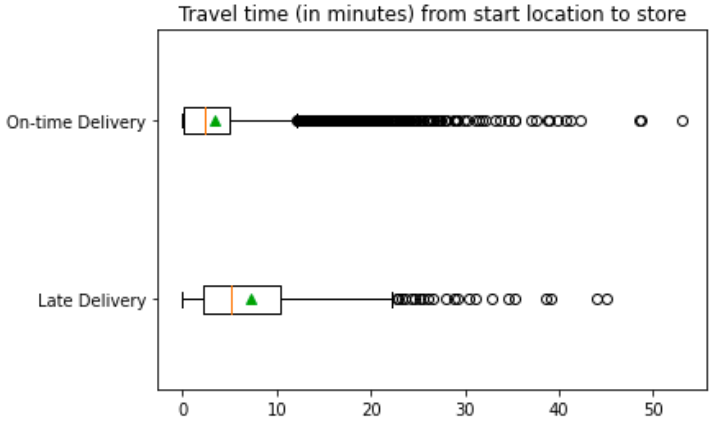
- From observing the deliveries made, it usually took a dasher less than 10 minutes to accept a delivery and less than 10 minutes to travel from their start location to the store. However, there are still outlier observations where it took more than 30 minutes (or even more than an hour) for a dasher to accept a delivery or to travel to the store.



Time from start location to store:
Median: 2.5 minutes
75% percentile: 5.2 minutes

Time for a delivery be accepted:
Median: 1.9 minutes
75% percentile: 6.5 minutes

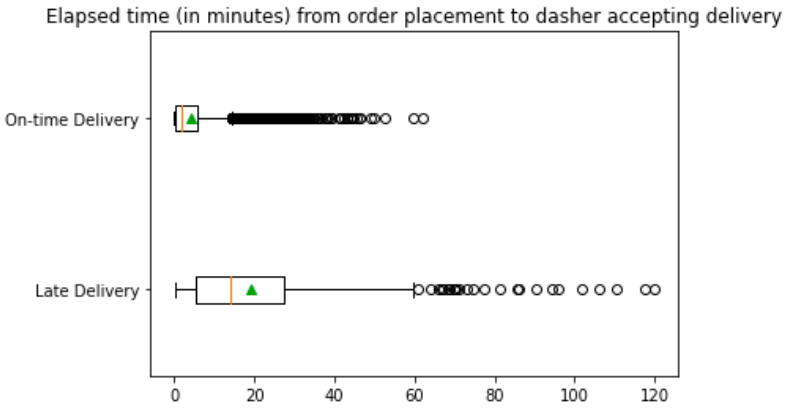
- Longer travel time of dashers from their start location to the store could result in delivery orders being late by 20 minutes. In late deliveries, travel time of the dasher to store location is 2 times higher than in on-time deliveries.



For on-time delivery:
Median: 2.4 minutes
75% percentile: 5 minutes

For late delivery (late > 20min):
Median: 5.2 minutes
75% percentile: 10 minutes

- Longer elapsed time from order placement time to dasher accepting the delivery could also result in late deliveries.



For on-time delivery:
Median: 1.8 minutes
75% percentile: 5.9 minutes

For late delivery (late > 20min):
Median: 14 minutes
75% percentile: 27.5 minutes

Recommendation:

- To reduce the number of late deliveries, it is worthwhile to take effort to reduce the total time from when customers place a delivery to when they receive it, and to better predict the estimated delivery time for an overall better customer experience with the New Vertical businesses and with DoorDash's delivery service in general.
- Additional data that would be helpful to further understand dashers' performance: travel time from store to delivery location, delivery completed timestamp. With this data, we would be able to understand how long it would take for drivers to complete the delivery.