MGT 6203 Group Project Proposal

TEAM INFORMATION

Team #: 102

Team Members:

1. Kimberley Griffith
    a. Professional Background: Worked as a Data Analyst an insurance company from 2013 – 2022; developed and presented data insights against large data sets to solve questions. Performed regression analysis on claims and administered the company's database and software systems. Currently working as Data Analyst role at pharmaceuticals company; developing and improving reports for stakeholders to understand team performance and Rx sales.
    b. Education Background:
        i. 2006 – 2009: BASc in Computer Science and Accounting, University of the West Indies
        ii. 2020 – Present: Msc in Data Analytics, Georgia Tech
    c. Previous Analytics: Professional experience gained through Data Analyst roles. Also taken the following courses in the program: CSE 6040, CSE 6242, ISYSE 6501, ISYE 6402.
2. Amit Ubhi
    a. Professional Background: Working as a Senior Product Engineer at 3M Canada. For the first two years was a manufacturing engineer before moving into a lab role as a product engineer (last four years). Working at a manufacturing site there is a lot of data generated for our products which is why the certificate at Georgia Tech caught my interest.
    b. Education Background:
        i. 2008 – 2013: BASc in Chemical Engineering, University of Ottawa
        ii. 2008 – 2013: BSc in Computing Technology, University of Ottawa
        iii. 2014 – 2015: M.Eng in Chemical Engineering, University of Ottawa
        iv. 2021 – present: Graduate Certificate in Data Science for the Chemical Industry, Georgia Tech
    c. I have taken two courses in Graduate Certificate so far. Course and project titles are below:
        i. ChBE 6745 (Fall 2021) - "Prediction of the Frequency Distribution of Algae based on the Measured Chemical Concentration, Season, River Size and River Flow Velocity"
        ii. ChBE 6746 (Spring 2022) - "Data-driven multiobjective optimization (Reproducing of scientific paper: https://www.science.org/doi/10.1126/sciadv.abf7435)"

3. Marissa Karl
    a. Professional Background: I’ve worked as an Engineer and Program Manager at Buckeye Partners, LP since 2017. Buckeye Partners is part of the energy sector as a midstream oil and gas company. I managed several programs for Buckeye ranging from $350,000 – $1.8 million annually, as well as managing the maintenance and troubleshooting of emissions control equipment.
    b. Educational Background:
        i. Bachelors of Science in Mechanical Engineering, Lehigh University
        ii. Bachelors of Science in Business, Lehigh university
    c. Previous Analytics: I’ve taken one Georgia Tech course (CSE 6040) so far, and am currently taking ISYE 6501 as well as MGT 6203. For several years at Buckeye, I created, developed, and reported out on key performance indicators regarding equipment operating time, preventive maintenance, greenhouse gas emissions, etc.
4. Arch Wilson
    a. Professional Background: I am a submarine office in the U.S. Navy. I currently work at the United States Special Operations Command in the acquisitions department for maritime equipment. Most of my professional experience is related to nuclear engineering and dynamic operational planning.
    b. Educational Background:
        i. Bachelors of Science in Mechanical Engineering, University of Pittsburgh
        ii. Prospective Nuclear Engineering Officer Course, U.S. Navy
    c. Previous Analytics: No professional experience, but I have taken the following courses in the program.
        i. CSE 6040
        ii. CSE 6501
        iii. ISYE 6644
        iv. MGT 8803
5. Brandi Rolston
    a. Professional Background: I am a Procurement DFX Project Manager for Cooper Lighting. My primary role is managing a portfolio of DFX (Design for Excellence) projects focused on bill of material cost savings. Most of my professional experience has been continuous improvement within manufacturing.
    b. Educational Background:
        i. Bachelor' of Science in Mechanical Engineering, Kettering University (formerly GMI)
        ii. Six Sigma Black Belt
    c. Previous Analytics: The only analytics role that I’ve had thus far has been Six Sigma statistical analysis. Within the OMSA, I’ve taken:
        i. ISYE 6501
        ii. MGT 8803
        iii. MGT 8823

OBJECTIVE/PROBLEM

Project Title: Data Driven Food Delivery Decisions (DDFDD)

Background Information on chosen project topic:

In today's world, food delivery makes up a significant portion of the overall restaurant market share. Globally, food delivery has been estimated to be worth more than $150 billion (Kabir Ahuja, 2021). Since the COVID-19 pandemic, the food delivery market has more than doubled, with more people opting to skip eating in, and instead eat in the comfort of their own home.

Delivery is crucial to a restaurant's ability to survive and thrive in a market with increasingly shrinking margins, and any insights on a restaurants ability to maintain or increase customer satisfaction and sales revenue, while thoughtfully reducing operating costs is a competitive advantage. The data provided will allow us to gain these insights and make actionable recommendations or conclusions for a restaurant to have a more successful delivery business.

Problem Statement:

The purpose of our analysis is to provide the necessary information to decision makers to improve their vendor rating and sales through the most impactful factors related to delivery service.

Primary Research Question (RQ):

How do delivery times affect the sales volume and overall rating of a restaurant?

Supporting Research Questions:

1. What other factor(s) have a major impact on either sales volume/overall rating?
2. Does weather have an impact on volume of food delivery orders?
3. How are sales volumes and restaurant ratings related?
4. Which type of restaurant has the most impact on restaurant rating?

Business Justification:

This problem can help restaurant owners focus on what specifically needs to be focused on to help them be successful in food delivery. If their delivery times are too long, they can hire new drivers, or develop a more efficient preparation process. If people are more likely to get delivery if it's raining, staff can be re-assigned to focus on delivery operations. If higher fees have no effect on customer patterns, owners can charge more for delivery.

The success of a restaurant is rarely dependent on just the few factors that we are analyzing here. However, given how data like this is increasingly available, analyzing it gives the decision makers a little more insight that could lead to a competitive edge over their business rivals.

DATASET/PLAN FOR DATA

Data Sources

Data derived from:

- https://www.kaggle.com/datasets/mrmorj/restaurant-recommendation-challenge?select=orders.csv
- https://www.visualcrossing.com/weather/weather-data-services

Data Description:

Both the Orders and Vendors datasets come from an app-based delivery service, Akeed, which was started in 2019 in Oman.

Dataset 1: Orders

The orders dataset contains roughly 135,000 observations. There are several ID columns, including customer ID and vendor ID, which can be used to merge and analyse more in depth details regarding the resturant and the customer making the order.

The dataset contains 26 attributes, with brief exerpts shown below.

akeed_order_id	customer_id	item_count	grand_total	payment_mode	vendor_discount_amount	is_rated	vendor_rating	driver_rating	deliverydistance	preparationtime	delivery_time	order_accepted_time
307997	M7BU6GY	1	8.4	1	0	Yes	4	5	5.69	40	12/23/2019 17:00	12/23/2019 17:57
308018	EVBECSV	1	9.5	1	0	Yes	5	5	3.49	35	12/23/2019 17:30	12/23/2019 18:11
308070	NA9NTQN	1	9.8	1	0	No		0	7.34	40	12/23/2019 17:10	12/23/2019 18:41
308105	Q2DPNU0	1	11.1	1	0	No		0	3.11	40	12/23/2019 17:35	12/23/2019 19:11
308297	11MG3HH	2	27	1	0	Yes	5	5	14.47	55	12/23/2019 20:00	12/23/2019 20:42
308314	GC1ZAJE	2	7.1	1	0	No		0	6.85	50	12/23/2019 19:16	12/23/2019 20:47

driver_accepted_time	ready_for_pickup_time	picked_up_time	delivered_time	vendor_id	created_at	LOCATION_NUMBER	LOCATION_TYPE	CID X LOC_NUM X VENDOR
12/23/2019 18:06	12/23/2019 18:00	12/23/2019 18:28	12/23/2019 19:01	86	12/23/2019 17:57	3	Work	M7BU6GY X 3 X 86
12/23/2019 18:41	12/23/2019 18:46	12/23/2019 18:51	12/23/2019 19:00	113	12/23/2019 18:10	1	Home	EVBECSV X 1 X 113
12/23/2019 18:46	12/23/2019 18:46	12/23/2019 18:59	12/23/2019 19:12	113	12/23/2019 18:41	0	Work	NA9NTQN X 0 X 113
12/23/2019 19:13	12/23/2019 19:22	12/23/2019 19:28	12/23/2019 19:40	386	12/23/2019 19:06	1	Other	Q2DPNU0 X 1 X 386
12/23/2019 20:46	12/23/2019 20:51	12/23/2019 21:07	12/23/2019 21:26	356	12/23/2019 20:40	1	Other	11MG3HH X 1 X 356
12/23/2019 20:52	12/23/2019 20:57	12/23/2019 21:01	12/23/2019 21:31	76	12/23/2019 20:47	1	Home	GC1ZAJE X 1 X 76

Factors that will be extracted from this dataset are:

1. Item Count – the total number of items included in the order
2. Total Cost
3. Delivery Distance – how far between the restaurant and delivery location?
4. Location Type – was this order delivered to home, work, other?
5. Picked Up Time – when was the order picked up
6. Delivered Time – when was the order delivered?

Dataset 2: Vendors

The Vendors dataset contains 100 unique vendors, and associated details specific to each vendor. The vendor ID field will be used to merge the orders dataset with the vendor data.

The dataset contains 59 attributes of varying usefulness.

id	authentic	latitude	longitude	vendor_category_en	delivery_	serving_d	is_open	OpeningTime	OpeningTime2	prepratio	commission	discount_	status	verified	rank	language	vendor_rating
4	118597	-0.5886	0.754434	Restaurants	0	6	1	11:00AM-11:30PM	-	15	0	0	1	1	11	EN	4.4
13	118608	-0.47165	0.74447	Restaurants	0.7	5	1	08:30AM-10:30PM	-	14	0	0	1	1	11	EN	4.7
20	118616	-0.40753	0.643681	Restaurants	0	8	1	08:00AM-10:45PM	-	19	0	0	1	1	1	EN	4.5
23	118619	-0.58538	0.753811	Restaurants	0	5	1	10:59AM-10:30PM	-	16	0	0	1	1	11	EN	4.5
28	118624	0.480602	0.55285	Restaurants	0.7	15	1	11:00AM-11:45PM	-	10	0	0	1	1	11	EN	4.4
33	118629	-0.49465	0.743318	Restaurants	0.7	6	1	11:00AM-10:30PM	-	17	0	0	1	1	11	EN	4.6
43	118639	-0.11501	0.545973	Restaurants	0.7	15	1	11:00AM-11:45PM	-	15	0	0	1	1	11	EN	4.3
44	118640	-0.93656	0.081933	Restaurants	0.7	15	1	11:00AM-11:45PM	-	14	0	0	1	1	11	EN	4.3
55	118651	-1.17015	0.103477	Restaurants	0.7	10	1	09:00AM-11:30PM	-	19	0	0	1	1	11	EN	4.5

Factors that will be extracted from this dataset are:

1. Delivery Charge – does the vendor have a delivery charge?
2. Serving Distance – How far away will the vendor deliver to?
3. Preparation Time – How long does the food take to prepare?
4. Vendor Rating – What is the rating of the restaurant between 0 -5?
5. Vendor Tag Name – Tags describing the type of vendor – i.e., Pizza, Burgers, Asian
6. Hours Open – Used to calculate number of hours open per week

Dataset 3: Weather

NOTE: The data sourced for this project comes from a delivery company Akeed, which was founded in 2019 in Muscat, Oman. We will assume the data is from Muscat based on this information.

The dataset includes the daily minimum, maximum and average temperature in Muscat. The data also includes the real feel, precipitation, and cloud cover. We will be extracting average temperature as well as precipitation for analysis.

An excerpt can be found below:

name	datetime	tempmax	tempmin	temp	feelsliken	feelsliken	feelslike	dew	humidity	precip	precippro	precipcov
muscat	5/1/2019	96.8	84.2	89.6	93.7	82.5	88.1	54.3	31.5	0	0	0
muscat	5/2/2019	100.4	78.8	89	96.5	78.8	87.4	50.5	28.8	0	0	0
muscat	5/3/2019	96.8	77	88.4	95.3	77	88.4	58.7	39.2	0	0	0
muscat	5/4/2019	95	78.8	89.1	102.6	78.8	92.4	64.4	47.5	0	0	0
muscat	5/5/2019	98.6	83.8	91.4	98.4	85.4	92.9	58.7	39.8	0	0	0
muscat	5/6/2019	105.8	84.2	94.4	100.2	82	92.6	51.5	25.7	0	0	0
muscat	5/7/2019	102.2	84.2	92.7	102.6	82.2	93.8	58.9	35.1	0	0	0
muscat	5/8/2019	96.8	82.4	87.3	107.8	88.3	98.6	76	70.6	0	0	0
muscat	5/9/2019	90	84.2	87.2	108.1	91.7	101.1	78.3	76.3	0	0	0
muscat	5/10/2019	93.2	79.2	87.8	107.8	79.2	99.1	75.4	68.3	0	0	0
muscat	5/11/2019	105.8	87.4	96	102.1	90.3	96.5	59.6	32.5	0	0	0
muscat	5/12/2019	104	82.4	95.9	100.5	80.5	92.9	48	20.6	0	0	0

datetime	preciptyp	snow	snowdept	windgust	windspee	winddir	sealevelp	cloudcove	visibility	solarradia	solarener	uvindex	severerisk	sunrise	sunset
5/1/2019					17.2	241.8	1005.4	0	6.2	342.8	29.5	10		2019-05-0	2019-05-0
5/2/2019					15	226.4	1005.2	0	6.2	342.5	29.5	10		2019-05-0	2019-05-0
5/3/2019					11.4	172.6	1006.2	0	6.2	342.3	29.6	10		2019-05-0	2019-05-0
5/4/2019					12.8	123.8	1007.7	0	6.1	340.1	29.5	10		2019-05-0	2019-05-0
5/5/2019					11.3	184.2	1009.3	4.6	6.2	330	28.6	10		2019-05-0	2019-05-0
5/6/2019					16.1	218.4	1007	0	5.7	324.7	28	10		2019-05-0	2019-05-0
5/7/2019					11.4	147.6	1004.1	0	5.3	336.6	29.2	10		2019-05-0	2019-05-0
5/8/2019					13.3	137.7	1004.2	0	5.1	327	28.3	10		2019-05-0	2019-05-0
5/9/2019					15	69.9	1004.6	9.1	5.3	320.3	27.8	10		2019-05-0	2019-05-0
5/10/2019					11.2	104.9	1004.8	8.7	5.3	331.1	28.7	10		2019-05-1	2019-05-1
5/11/2019					15	209.7	1005.8	0	6.2	316.8	27.3	10		2019-05-1	2019-05-1
5/12/2019					16.1	235.5	1005.8	0	5.5	330	28.5	10		2019-05-1	2019-05-1

Key Variables:

Independent variables:

- Delivery Charge
- Delivery Distance
- Daily Precipitation
- Daily Temperature
- Location Delivered To – Home, Work, Other

Variables that will be created – Independent:

- Average item cost – total cost divided by number of items
- Total Delivery Time – Difference of delivery time vs pick up time
- Hours Open Per Week – Calculated from open/close time per day per vendor

Variables that will be created – Dependent:

- Total Sales Volume for each vendor
- Number of Sales for each vendor

Dependent variables to be analyzed:

- Vendor Rating

We hypothesize that delivery time will have the most impact on vendor rating, while hours open and precipitation will have the most impact on sales volume.

APPROACH/METHODOLOGY

Planned Approach:

Merge Datasets

The order table would be considered the primary dataset and contains key information such as the order total, creation date and delivery times. It will need to be merged with the vendor table using the Vendor ID field to capture data such as the vendor's rating. This is key as we aim to determine how delivery times affect the rating of the restaurant. The Weather data will be merged as well for exploration; we'll analyze the correlation with the day's temperatures and precipitation.

Data Transformation

Some records do not contain relevant information such as the Picked Up and Delivered Time and may have to be discarded as the data is cleaned. The time it takes to deliver an order can be calculated in minutes using the difference between the Picked Up Time and the Delivered Time.

We will perform a check for outliers using Cook's Distance to determine whether any records need to be excluded.

Data Splitting

Data should be sliced into two sets for determining the best model using the standard percentages

(i) Training/Validation set—a subset to train and validate model — 75%
(ii) Test set—a subset to test the trained model — 25%.

Modeling

We will perform regression analysis using the restaurant's overall rating as the dependent variable and delivery times as the independent variable, as well as other factors for further exploration. This will allow us to answer the question 'How do delivery times affect the sales volume and overall rating of a restaurant?'.

In addition, we perform a second regression analysis on the data using total sales volume and rating at the outcome variables and the other available fields as dependent variables. We will not make any assumptions about the attributes which contribute to a high sales volume. Instead, we will apply Stepwise Regression to train and optimize factors which should be included. Then, the actual model would be completed using just the key fields in our data.

We will analyze R-Squared values when applying regression to our training data to ensure that we have the best model and that it captures the variability of the data.

Anticipated Conclusions/Hypothesis

We expect that shorter delivery times are correlated to a high overall restaurant rating and will rely on the modelling to determine whether this is true or not.

Some other initial results we expect to see is whether the delivery location (ie – home, work, etc) has any impact on the rating and total sales. We anticipate the data will point to the customer mostly ordering to their home. Another interesting result to confirm is whether the closer the customer lives, the lower they will spend versus the further the customer lives, the more they will spend to get the most out of a long delivery order.

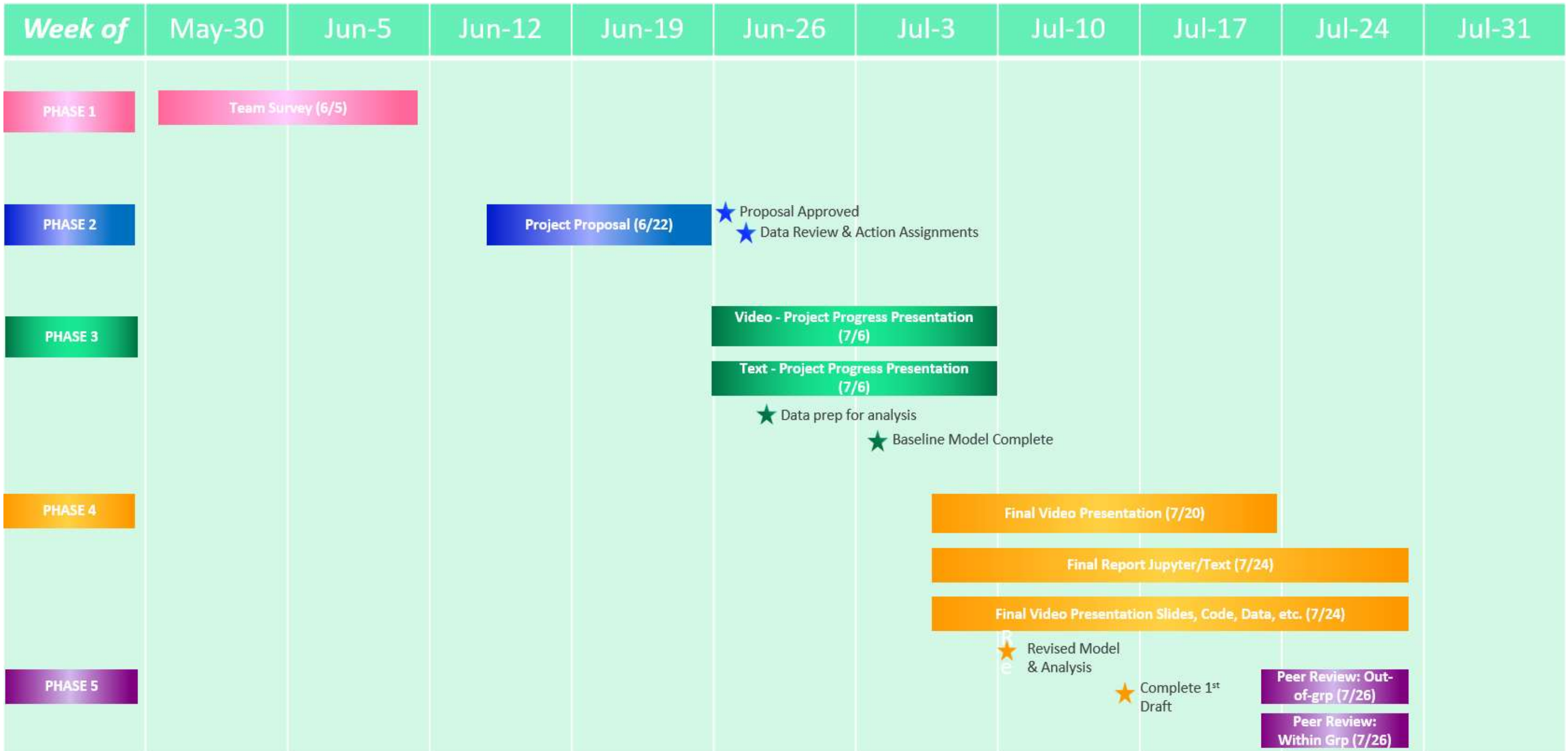
We might also expect the vendor's delivery charges and opening times to have a correlation with ratings and sales. High charges may result in lower ratings and longer opening times may result in greater sales.

Business Decisions Which Will be Impacted by the Results of Analysis and Possible Benefits

A business decision that could be impacted by the results of this analysis is that if the data indicates there are days of the week with minimal delivery orders, it could result in reduced working hours for delivery drivers. It is possible restaurant owners may notice they do not generate a lot of income during select hours of the day and may decide to keep delivery open only for certain time slots to maximize income. A business could also consider limiting their delivery radius if the data indicates most ordering customers live nearby (X distance) or even not delivering to customers if the estimated time exceeds a certain number of minutes (from data analysis).

However, if the data shows that there are certain days (i.e., Friday, Saturdays, etc.) that are the busiest, owners could leverage this to modify their menu based on feedback to ensure they are getting the most out of their delivery driver (working full hours on one day than having fewer hours spread throughout the week). The data could also help restaurant owners understand if hiring more staff on rainier days could result in more income due to quick delivery service because of the extra preparation. Also, if higher fees have no effect on customer patterns on those busy days, owners can charge more for delivery.

PROJECT TIMELINE/PLANNING



APPENDIX

References

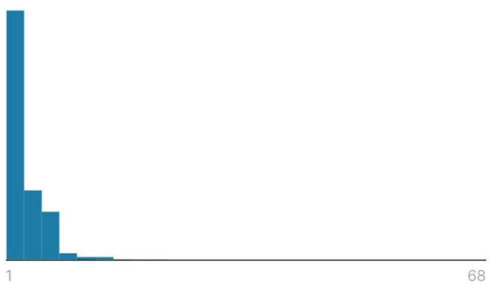
Kabir Ahuja, V. C. (2021, September 22). *Ordering in: The rapid evolution of food delivery*. Retrieved from https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ordering-in-the-rapid-evolution-of-food-delivery

High Level Data Details

Order Details

item_count

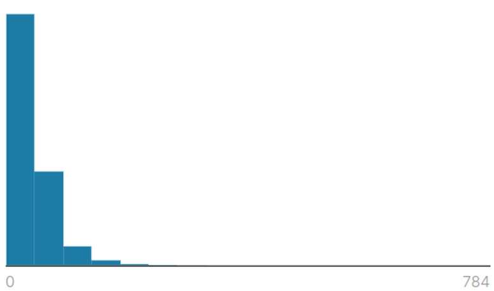
how many items were in the order



Valid	128k	95%
Mismatched	0	0%
Missing	6925	5%
Mean	2.41	
Std. Deviation	1.65	
Quantiles	1	Min
	1	25%
	2	50%
	3	75%
	68	Max

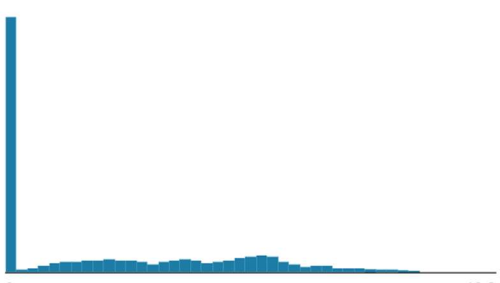
grand_total

total cost



Valid	135k	100%
Mismatched	0	0%
Missing	0	0%
Mean	15.4	
Std. Deviation	12.6	
Quantiles	0	Min
	8.2	25%
	11.7	50%
	18.5	75%
	784	Max

Order details



Valid	135k	100%
Mismatched	0	0%
Missing	0	0%
Mean	4.1	
Std. Deviation	4.36	
Quantiles	0	Min
	0	25%
	2.9	50%
	7.92	75%
	19.8	Max

picked_up_time

Order details

[null]	38%
2019-11-12 21:33:51	0%
Other (83862)	62%

Valid	83.9k	62%
Mismatched	0	0%
Missing	51.4k	38%
Unique	83.6k	
Most Common	2019-11-12 ...	0%

delivered_time

Order details

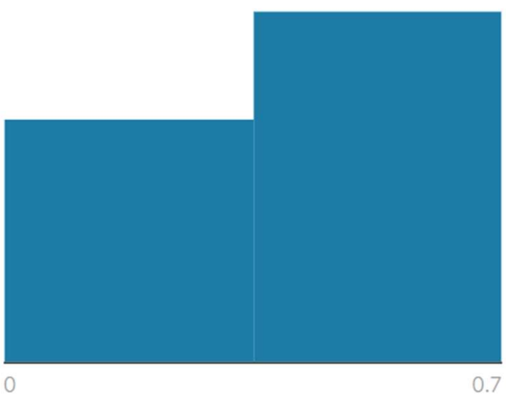
[null]	37%
2019-10-26 18:05:30	0%
Other (85738)	63%

Valid	85.7k	63%
Mismatched	0	0%
Missing	49.6k	37%
Unique	85.5k	
Most Common	2019-10-26 ...	0%

Vendor Details

delivery_charge

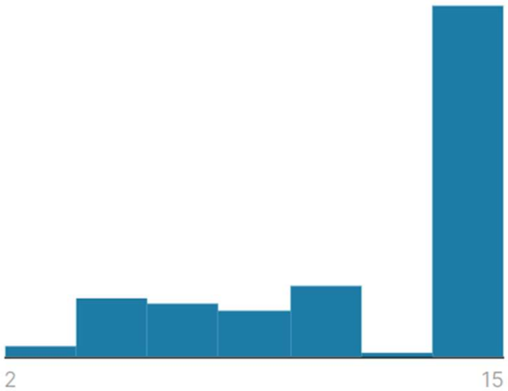
self-explanatory



Valid	100	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.41	
Std. Deviation	0.34	
Quantiles	0	Min
	0	25%
	0.7	50%
	0.7	75%
	0.7	Max

serving_distance

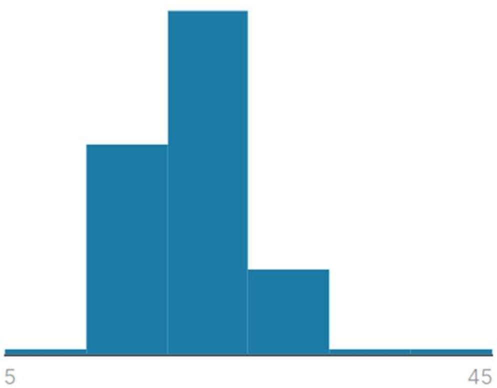
self-explanatory



Valid	100	100%
Mismatched	0	0%
Missing	0	0%
Mean	11.8	
Std. Deviation	4.09	
Quantiles	2	Min
	8	25%
	15	50%
	15	75%
	15	Max

prepration_time

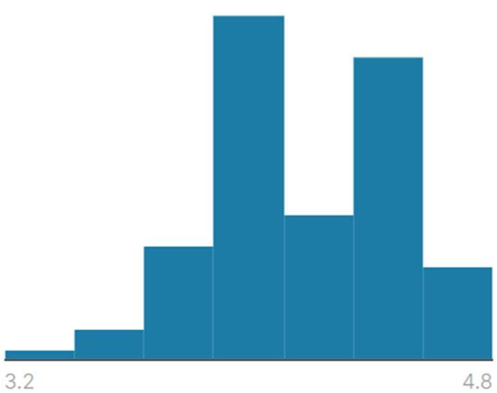
self-explanatory



Valid	100	100%
Mismatched	0	0%
Missing	0	0%
Mean	14	
Std. Deviation	4.3	
Quantiles	5	Min
	10	25%
	15	50%
	15	75%
	45	Max

vendor_rating

self-explanatory



Valid	100	100%
Mismatched	0	0%
Missing	0	0%
Mean	4.35	
Std. Deviation	0.25	
Quantiles	3.2	Min
	4.2	25%
	4.4	50%
	4.5	75%
	4.8	Max