



Come up with new important metrics

Create dashboards to solve vital questions

Create dashboards to keep track of KPI

End

# 1st Goal: clean, restructure and deal with missing data / come up with important metrics that are not being tracked

- Let's first clean and restructure the data set in a way we can import this data set to any program without any difficulty.
- To be able to restructure the dates and change the type of field for each column. I used python, tableau and excel.
- My first recommendation is to make sure that dates are stored in a recognizable format.

01 03:00:25



1/1/1900 3:00:25 AM

- Since the complete date is missing, we will preset month=1 of year=1900
- Now that the data is restructured let's focus on N/A(missing values).

## MISSING VALUES

We want to check which data field has missing values

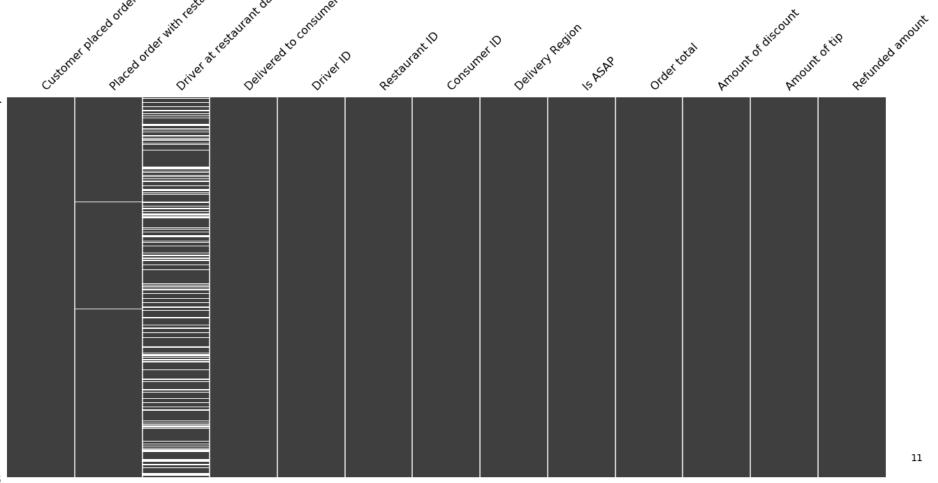
This graph provides a better understanding on where the missing data is located. The white lines in the rows represent each null for that specific column.

Most of the missing

data is at the column

"Driver at restaurant

time datetime"

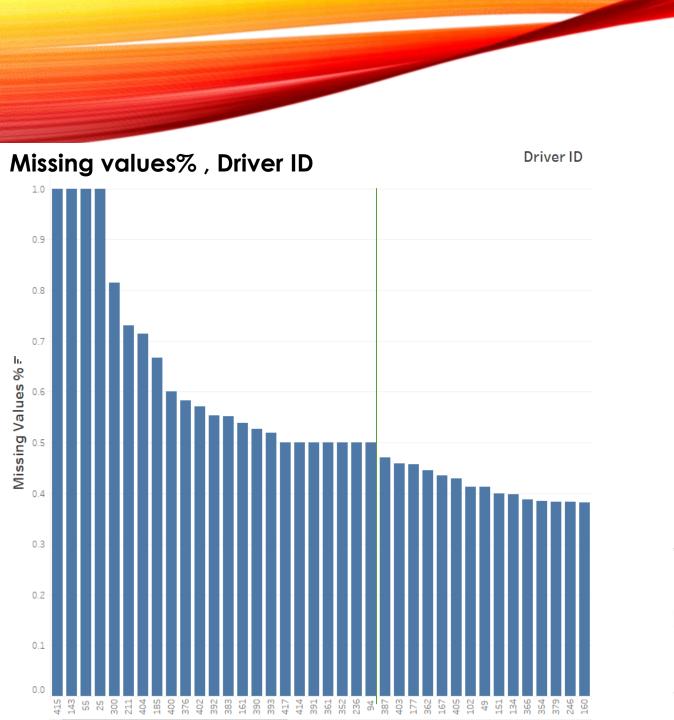


18078

- From the last graph we can identify that there are 2 columns with missing values.
  - "Placed order with restaurant datetime". Since the missing value count is very low there is not rush to fix this column
  - "Driver at restaurant datetime': this field is missing a significant amount of data. Let's try to recognize why and how to fix it.
    - I don't understand why Doordash is missing so much of this metric. I would assume that the driver needs to click manually that he has arrived. I would recommend that it automatically saves the arrival time.
    - More important metrics could also be tracked, for example: zip code, traffic in area, miles, etc.
    - "Driver at restaurant datetime" is a crucial field since without it we can't determine if our drivers are spending to much time waiting for the food.
      - Also, to be able to figure out the time spent waiting at the restaurant, Doordash would need to create a new column that keeps track of the time when the driver leaves the restaurant and when the food is ready, which I would also recommend keeping track off.
    - Since Doordash has not implemented this into the app we have to assume that it might never happen. So, we need to figure out a way to identify which drivers are not saving the arrival time.

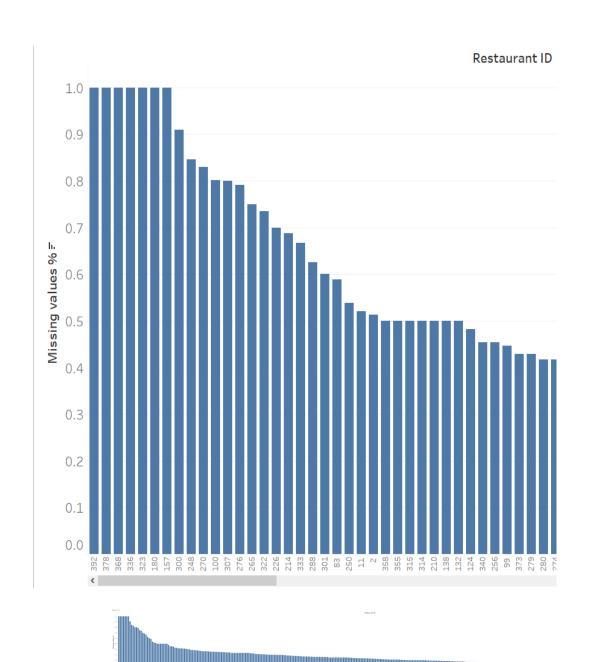
 Since Doordash has not implemented the automatic retrieval of the time when the driver gets to the restaurant, we want to find a way to find out which drivers or restaurants are causing this issue.





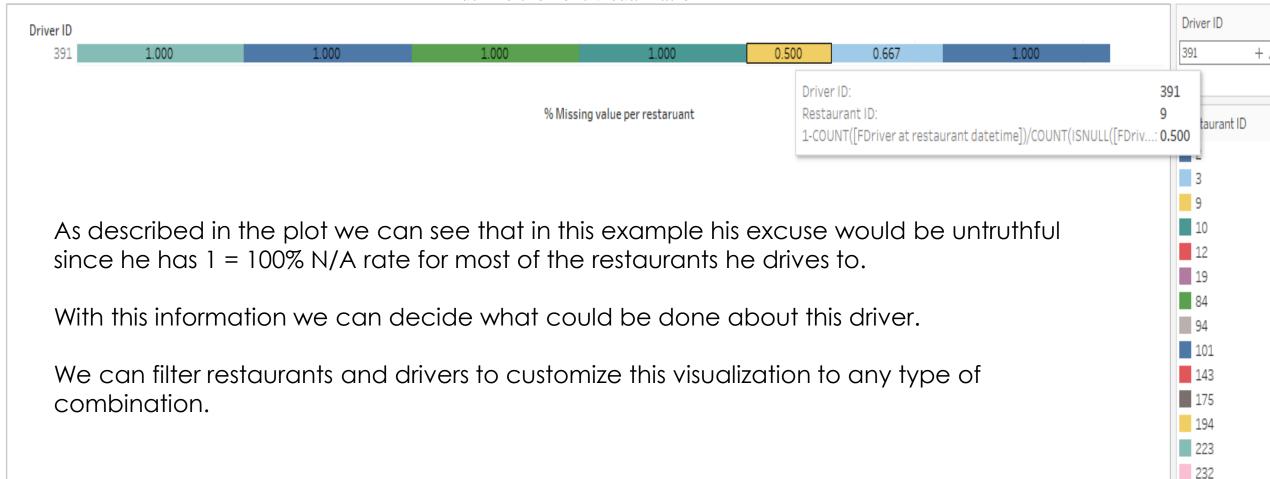
- Let's start and check if there is a specific driver that does not save the arrival time:
  - Using the two graphs in this PPT we can see that there is a high increase of N/A rate after 0.5(50%)
  - Contact those drivers and request to mark their arrival time and ask why they haven't been doing so.
  - Try to create incentives or an easier interface to encourage drivers to mark their arrival time.

- We will do the same procedure to get restaurants with high N/A ratio
  - We can now contact restaurant with the top % missing values and check what is happening.
    - Maybe drivers know that the restaurant in request takes longer to make the food and if they mark their arrival time Doordash expect the drivers to take at most X minutes picking up the food. However, for this specific restaurant the usual wait is around x+5 and these drivers might fear getting reprimanded for the restaurant fault.



- Now that we know which drivers could be causing the missing values % Doordash should contact them. When contacting the drivers two scenarios two could happen:
  - The drivers didn't realize this was a requirement and change their behavior
    - Best scenario
  - The drivers blame on location of the restaurant, waiting time, parking, walking distance as the reason why he or she does not want to save their arrival time.
    - For this situation I have created a dashboard that we can utilize to determine if is indeed something is going on with the restaurant or if the driver is providing us with false information
    - Let's see an example in the next PPT:

• Example: Doordash calls driver 391since its N/A rate is 50%, according to the first graph discussed in ppt #7. Driver says he/she always saves his/her arrival rate but when he/she goes to the restaurant X he/she doesn't. He/she claim that the parking is too far away from the restaurant or its always busy. So, if the driver did save the time, he/she would be afraid that the customer or Doordash might think it's the driver's fault the food is taking longer than expected and lower their tip or something similar. To try to decide if the driver is being trithfull we can utilize the next visualization



• I was going to cerate another dashboard just like the last one but for restaurants. However, for that graph to be useful Doordash would need to start retrieving the time when the food is ready "food ready metric datetime" and departure time of the restaurant. Which is why I would highlight that those metric should be recorded



#### • Recommendation, part 1:

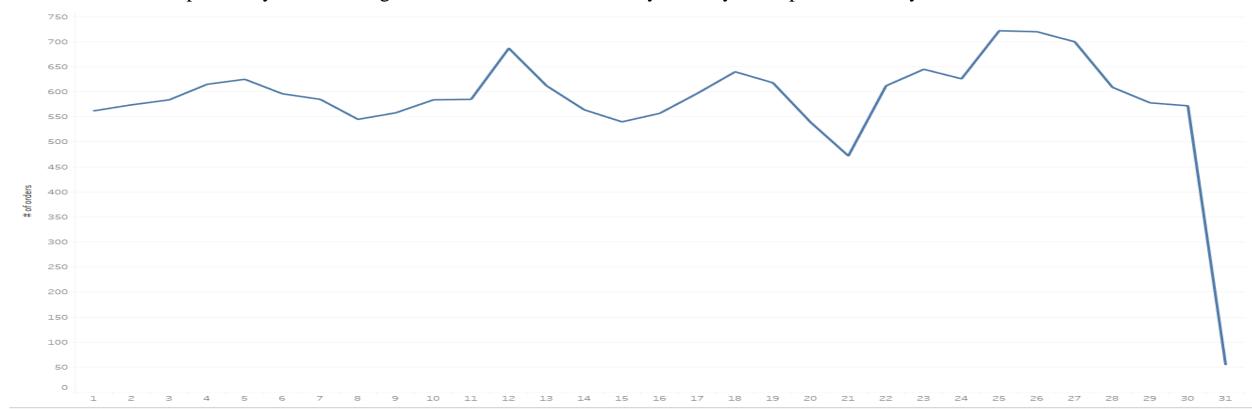
- Recommendation:
  - Program the app so that it records the time the driver arrives to the shop automatically
    - Until this happens use dashboard to try to decrease missing values by identifying and contacting specific drivers and restaurants.
  - Create a new field that provides the times when the food is ready. Crucial field to determine if drivers are waiting to long for the food. Also, useful to determine if driver takes longer than usual to pick up the food once he arrives.
  - Other extra fields to keep track of zip code, traffic in area, miles, etc.

• Now that we have solved the data structure and lack of information issue lets move on

## # OF ORDERS

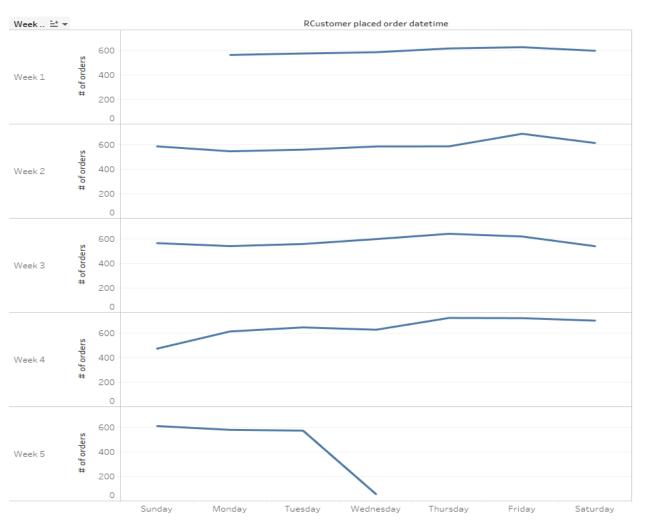
# 2nd Goal: Recognize patterns in the data to schedule the correct amount of drivers for those times

- Lets first try to find trends within our data.
- Since we only have 1 month, we can't study how month affects the amount of orders, but we can do some research on how it fluctuates within the month
  - Assumption: day 31 is missing some information since it very unlikely to drop that suddenly

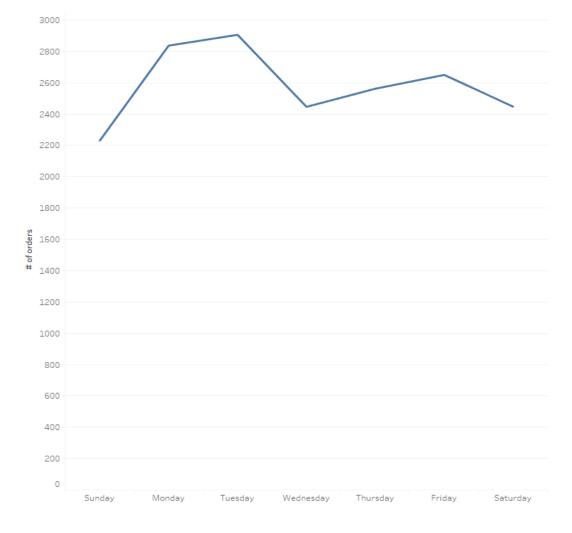


## WEEKLY/WEEKDAY/ TREND

### Aggregated sum of # of orders by week



#### Aggregated sum of # of orders by weekday

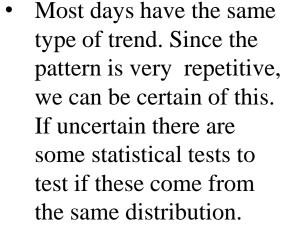


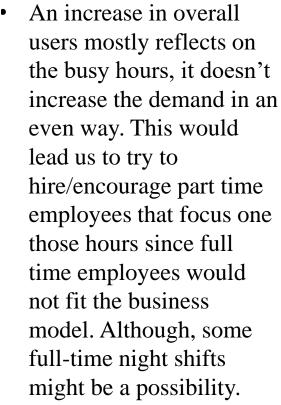
Next ppt for insights

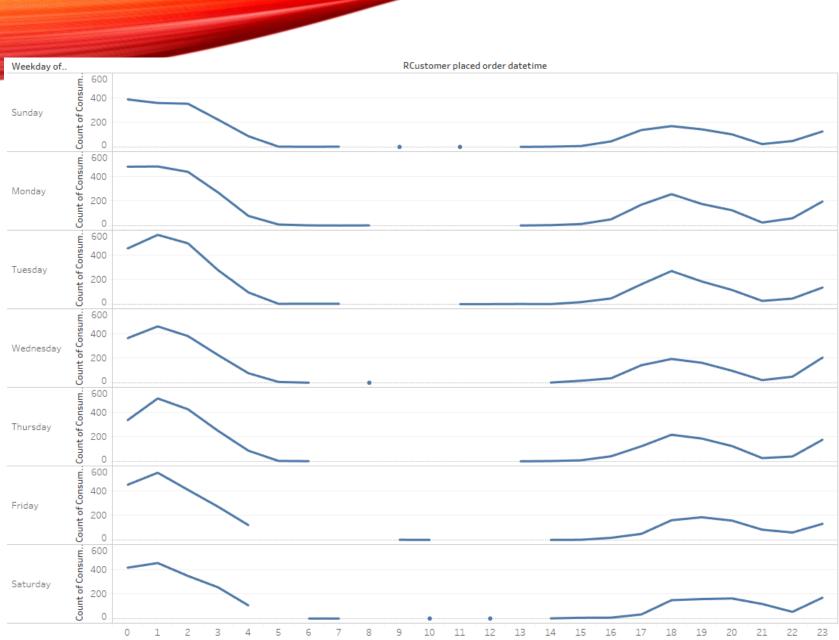


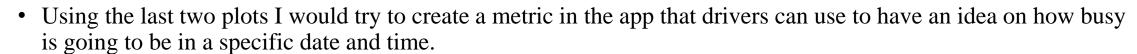
- Both graph shows a peek on the day of Friday with an number of orders at 620-650. Followed by a slight constant decrease until Sunday/Monday that reaches 600-550.
- After Monday it hovers around 600 and around Thursday starts increasing to repeat the cycle.
- Week one compared to week 4 have not that much in difference. This can tells us the % of retained users has not changed
- It would be good idea to create a metric that tells us newly acquire users and lost users, in that way we could keep track of the type of growth of the service.

## HOUR TREND









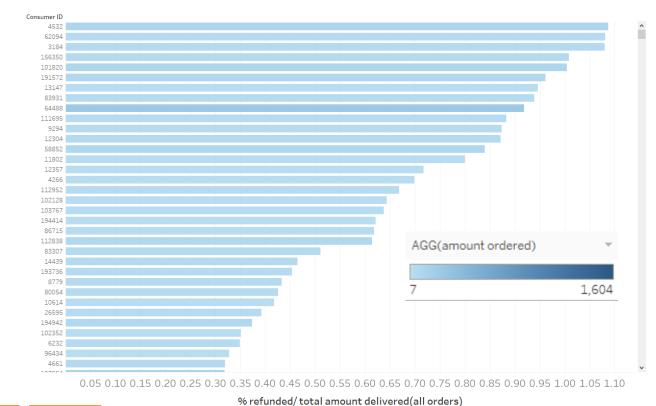
- The reason for this is that if there are too many drivers and not that many orders then drivers would try to find other ways to make money since Doordash wouldn't be reliable. If this scenario was inverted, more orders than drivers, then some customers won't be able to find drivers to take their order and they would look for a different app. Both scenarios are harmful for Doordash that's why a good balance would be ideal
- This gold ratio would be when customers can always order, maybe must pay extra or wait longer in peek hours but never too much to be discourage of using Doordash.
- And what is that golden ratio?
  - To be able to determine that ratio further analysis need to be done since distance, traffic, #orders, zip code, location, and other dimensions should be analyzed in combination of #orders and #drivers

## **SECURITY**

- There are always ways to trick the delivery app and try to scam the company. This could be done by drivers and customers.
  - This is a major issue. To deal with this matter I decided to create a visualization that represents the % of money refunded per driver over the total amount of \$ delivered by driver
    - This metric will help us identify if an individual has found a bug in the app or a loophole in the regulations.
  - Also, I created another visualization per customer that represents the % of money refunded over the total amount of \$ spent during the start of his account.

## **CUSTOMERS**

- If we see a dark blue color with a high %(a number close to 1 or higher) it means that that customer is a regular, and it is getting a high value of \$ refunded. Therefore, should create a note to further research that customer.
- Example: In the graph on the right, we can see that customer 4532 has a high return% (red flag) but since its color its light then he hasn't ordered that much. So, if we do more research on the orders of that customer, we can see that:

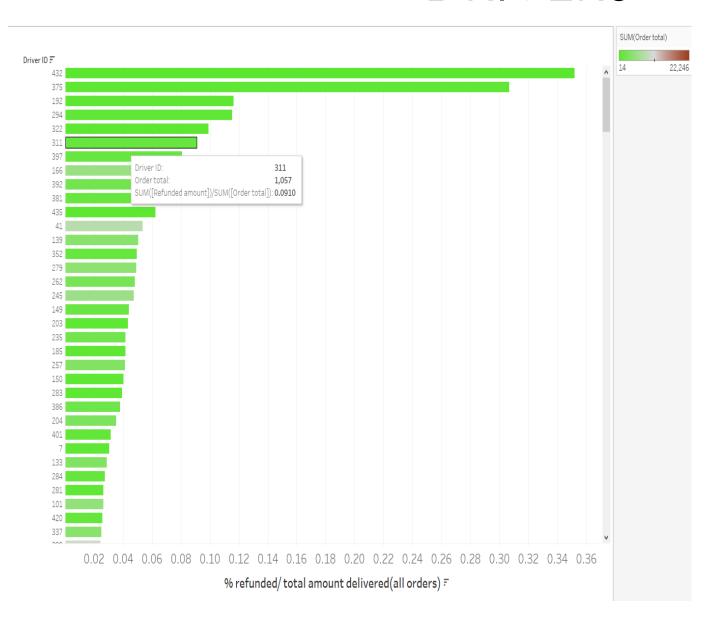




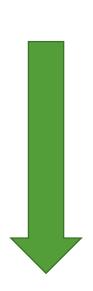
He only ordered once and got a refund something might have happened, not a red flag, although the amount of the order is quite large so maybe dig a little deeper  A red flags would be a dark color with a high % of refunded \$.Since nothing like that shows up in the last graph, we can stay calm for now.

## **DRIVERS**

Same reasoning as the last visualization if we find a red value on the top then explore that specific driver



- In this graphs we were able to find a red flag
  - Driver 248 has a total amount for that month of 17293 and has almost 2% refund rate which is not that high.
  - Since the amount of \$ he has received is high it would be better to check what's going on





.00 0.00 0.10 0.12 0.14 0.10 0.10 0.20 0.22 0.24 0.20 0.20 0.30 0.32 0.34

% refunded/ total amount delivered(all orders) =



- Nothing seems to suspicious.
- I would have created a reusable dashboard to try to check if a driver is indeed doing something odd(lack of time). For this time I did it manually in excel and nothing seemed out of normal.

Driver ID   ■ Resta	aurant ID → Cons	umer ID  Delivery Region	Is ASAP ▼	Order total ▼ Amou	ınt of discount 🔻 Amoun	t of tip 🔻 Refu	nded amount 🔻 time real 🔻
248	98	8534 Mountain View	TRUE	29.87	0	0	0
248	77	2696 Mountain View	TRUE	40.64	0	1.59	0
248	30	2408 Mountain View	TRUE	32.64	0	1.63	0
248	43	7702 Mountain View	TRUE	15.73	0	0.44	0
248	39	1856 Mountain View	TRUE	26.66	0	0.95	0
248	43	75490 Mountain View	TRUE	16.28	6	1	0
248	217	6474 Mountain View	TRUE	25.58	0	0.9	19.57
248	87	75533 Mountain View	TRUE	58.64	6	4.84	0
248	44	9466 Mountain View	TRUE	19.54	0	0.98	0
248	217	1361 Mountain View	FALSE	29.87	0	1.09	0
248	44	74371 Mountain View	TRUE	169.74	0	7.25	0
248	41	76288 Mountain View	TRUE	24.98	6	0	0
248	47	6817 Palo Alto	TRUE	60.86	0	3.04	0
248	98	74937 Mountain View	FALSE	89.52	0	4	0
248	112	2938 Mountain View	TRUE	60.27	0	4.99	0
248	64	7451 Mountain View	TRUE	66.19	0	5.53	0
248	95	4124 Mountain View	TRUE	39.6	0	1.98	0
248	25	16859 Mountain View	TRUE	43.41	0	2.17	0
248	13	65616 Mountain View	TRUE	85.39	6	4.27	0
248	215	10125 Mountain View	TRUE	26.55	0	0.94	0
248	266	21164 Mountain View	FALSE	17.42	0	0.52	0
248	77	4093 Mountain View	FALSE	231.62	0	10.09	12
248	43	12108 Mountain View	FALSE	20.66	0	2.07	0
248	18	35250 Mountain View	TRUE	39.66	0	1.98	0
248	306	1074 Mountain View	FALSE	63.2	0	2.63	0
248	77	1719 Mountain View	TRUE	143.7	0	6.05	0
248	77	1719 Mountain View	TRUE	32.59	0	1.22	0
248	216	6422 Mountain View	FALSE	36.83	0	1.41	0
248	76	31093 Mountain View	TRUE	50.26	0	2.51	0
248	39	10114 Mountain View	TRUE	20.68	0	2.02	0
240	210	2051 Mountain View	TDITE	11 00	n	n 27	0

### • Recommendation, part 2:

- Promote part time employees during high peek hours and if possible, contract full time night shift employees
- Utilize the weekly, weekday, day and hour visualization to project and assess the number of drives that will be needed in the future.
  - With more time and data try to create a forecasting algorithm to try to estimate the trend better.
  - Try to perfect the ratio of # orders to # of drivers so that everyone benefits out of it. More information and time needed to try to come up with this complex ratio.
- Utilize customers and drivers plots to try to verify there is no hole in the regulations, policy or coding of the app.
  - If a user or customer seems to be getting high % refund start a more in-depth investigation