Eminence Case Study

April 27, 2020

1 OVERVIEW AND METHODOLOGY

Welcome to Rob Guzman's case study! In this case study, we begin with data from a CSV file containing Uber and Lyft data. At base, there are two primary goals:

1. Understand the data and clean it as best as we can.
2. Perform some basic analysis and provide leads to future steps.

A few additional points:

- Part of understanding the data is determining which columns are useful, which ones are largely useless, and which ones can be transformed into new columns. We will do all of this. At the end of the process, I will also say a few things about loading and storing this dataset in a relational database (if a file like this were loaded on a routine basis).
- When we perform our cleaning of the data, not everything will likely be entirely pristine. Capturing every detail would take time and much testing, and future files might contain anomalies that we do not see here. For the sake of this case study, we will pursue the "80-20 rule" and try to get the biggest bang for our buck.
- This notebook is written in "real-time." What this means is that as you venture down you will see my progression through the data as I did originally. Thus, many statements are run to see results and explore the data. Data cleanining and analysis is an *inductive* process, so we must look at the way the data actually are in order to gain an idea as to what we should do with it (or to it).
- Finally, when we get to the end part of the notebook and discuss data anlysis, I will provide several leads to possible future projects that one could pursue with this data. This is also not complete, but is meant to stimulate discussion.

With that, imagine yourself in the same position that a customer in one of these records: sit back, and enjoy the ride!

2 PART 1: EXPLORATORY ANALYSIS AND DATA CLEANING

We begin with imports and load the raw data into a Pandas dataframe. We will make extensive use of the Pandas library. We also run the info() function to get an idea as to what the data look like. Note that we receive a warning about mixed types: this is to be expected, since the data are likely filthy.

[1]:

```
import pandas as pd
import csv
import datetime
import numpy as np
import re

fileName = "C:/Users/rg255/Downloads/Data_Rideshare/Data_RideShare.csv"
rawData = pd.read_csv(fileName)

rawData.info()
```

```
C:\Users\rg255\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.8_qbz5n2
kfra8p0\LocalCache\local-packages\Python38\site-
packages\IPython\core\interactiveshell.py:3062: DtypeWarning: Columns (19) have
mixed types.Specify dtype option on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1877060 entries, 0 to 1877059
Data columns (total 33 columns):
 #   Column                     Dtype
---  ------                     -----
 0   merchant_name              object
 1   user_id                    object
 2   order_number               object
 3   order_time                 object
 4   email_time                 object
 5   insert_time                object
 6   update_time                object
 7   order_total_amount         float64
 8   order_points               float64
 9   order_shipping             float64
 10  order_tax                  float64
 11  order_subtotal             float64
 12  order_total_qty            float64
 13  product_description        object
 14  product_subtitle           object
 15  item_quantity              float64
 16  item_price                 float64
 17  digital_transaction        int64
 18  checksum                   object
 19  product_reseller           object
 20  product_category           object
 21  order_discount             float64
 22  SKU                        float64
 23  item_id                    float64
 24  order_pickup               int64
 25  from_domain                object
 26  email_subject             object
 27  delivery_date             object
 28  start_source_folder_date  object
 29  end_source_folder_date    object
 30  file_id                   float64
 31  source_dttimestamp        object
 32  dttimestamp               object
dtypes: float64(12), int64(2), object(19)
memory usage: 472.6+ MB
```

One thing to check immediately is whether or not there are any exact duplicates in the data. In rare cases, one might expect such duplicates, but it is generally a bada practice.

```
[2]: rawData.duplicated().value_counts()
```

```
[2]: False    1877060
     dtype: int64
```

We are fortunate that there are no exact duplicates. I am sure we will find other reasons to remove rows, though.

Now that we have loaded the data in its raw form, we want to do some exploratory data analysis on the various columns to ensure that we do not have any bad data. As part of our larger goal of understanding the data, we will need to attach significance and an interpretation to each column in this dataset. At the end, I will summarize our findings here. If you are interested in the progression, read on.

2.1 MERCHANT NAME

We begin with the merchant name, which we expect to have two values: Uber and Lyft. Let's verify this with the value_counts() function.

```
[3]: rawData["merchant_name"].value_counts()
```

```
[3]: Uber    1292235
     Lyft     584825
     Name: merchant_name, dtype: int64
```

As expected, everything is tagged as either Uber or Lyft. Since merchant_name is a string, we would likely want to create our own table for merchants in our relational implementation of this data. That way, we can save space by storing an integer id instead of the full name. Not to mention, we can catch errors in spelling and new vendors with this approach that might come about in the future.

2.2 USER_ID

The next field to look at is user_id. For this value, we are interested in a few questions, such as:

1) what is the intended type?
2) what is the cardinality?

3) are there any that are missing?

We can proceed to answer these questions with the following commands. The first command will check the size of the row's contents (which I suspect we will see regularity in, since the field appears to be a GUID of some variety). The second will look for uniqueness across the rows (I do not expect this to hold, since one customer should be able to have multiple orders in the dataset). The third will look for particular values that have a high cardinality, which might indicate some problem. The fourth will look for missing data.

```
[4]: rawData["user_id"].apply(len).value_counts()
```

```
[4]: 64    1877060
     Name: user_id, dtype: int64
```

```
[5]: rawData.groupby("user_id")["merchant_name"].nunique()
```

```
[5]: user_id
     00014dfd77b41f5fa092a161f01ac5e63db0cbdc7bf3db9457b4748755003f03    2
     000226075b0be56c5a97290d4fbb03726e05b60d6c291cbc35a54edd9776013a    1
     000389577a74818063cf0cd729df90cdd29b7aec12870ee9d03ad3ae6d9600b7    2
     00079f4d49e4a26390863a9f7095a33c65d1ae056ba7f04fc49e62cf7454dba4    1
     000a460deaef4601c9c628fdb672bf9dd370a24fe4381b295cd5577a073c61dd    1
                                                                        ..
     fff90c2e3e918635517c38edfd6fcb7ea31f79a55a31908d2238795f6451538d    1
     fff91ead0bcdd9e9a92226e8c8f0c0ebf829ddc72559ac01583432f04a81de86    2
     fff952d9b26c067b5f7d3b072fed02c1769c80fe3c6db5223131bc3c61804ff9    1
     fffa8b12ea52b9e228c1ba4bd6b3ff2f0780a6cb6388b49f6a075c55575eed9e    1
     fffcfd1f4826db24e7f72bcd40f387dc3257d89ccd99e3ca52ecaa752784fb21    1
     Name: merchant_name, Length: 29291, dtype: int64
```

```
[6]: rawData["user_id"].value_counts()
```

```
[6]: 5a2ddafb9b4bc4dc0e506e0cc74ba53469295908e586a8e63bf3930a7b77ff9f    3137
     bcf5acdc42396cd2c5e86689627f42f096e101dd56afb74f54baae39c2d4f74d    2463
     1de4cda1ee476a68486e3f5755099343e971523826ed258e9b76dce8193b9435    2341
     bfe91c67be3a0b2bf89b2a124f640f85e96253b0242d62788a16f7aac1660598    2332
     057c3839aede4b0b466b9b0203cfc82824d4c5e3d9c7b41e763b9b550f2a1f84    2285
                                                                         …
     30689ff7867a81b62bad7294689602634d659be6efc906c5ab3794b0e0c2eb9c       1
     d4787b589c40f1629e073f3254071ddfcb9d78f1cc2357fafc3e97e9a19b8209       1
     479d24a7379caf0e3257cf0f3fe9b77e677c7a2722f830a0577d03828990b006       1
     ce0386c2c21dd8365936a50f02ff1fb985ba251735898f5f32787a5ab249df59       1
     c6ec6e7f6d1abd0a4b785631d0f1944abd8f22d96bd9ab388e8e6fa0eae0c5ae       1
     Name: user_id, Length: 29291, dtype: int64
```

```
[7]: rawData["user_id"].isnull().value_counts()
```

```
[7]: False    1877060
     Name: user_id, dtype: int64
```

As we can see from the results above:

1) There no missing values in the data.
2) Everything is a 64 character long string.
3) That there are cases where a single user_id has many orders.
4) Some user_ids are used by both Lyft and Uber data.

These findings raise some interesting questions.

First, what is going on with those customers that have hundreds or thousands of orders? Are these really individuals making thousands of trips? This seems to me to be unlikely; I think it is more likely that this is a shared account, perhaps corporate, that allows for many individuals to take trips, perhaps simultaneously. One way to validate this would be to look for instances where a single user_id has more than one simultaneous trip. This is something we should check later after we have cleaned the date and time data. So let's put a pin in that one (we will revisit it).

Another question is: is the user_id field universal for this file? Did our vendor consolidate all of their user data into a single table and then assign unique Ids to each of them, so that a user_id that is shared by both Uber and Lyft are really the same person? If so, it means that we could potentially be able to see whether individuals *switch vendors*. That would be something!

I think it is likely true, given the nature of the field. Assume for the moment that these user_ids were assigned by the vendors themselves. What are the odds that both Uber and Lyft would choose to use the same mechanism for tracking customers (64 character long strings) **and** that they would have a collision with such a gargantuan number of possibilities for assignment of these ids. The number of possible ids is approximately 36ˆ64 = 4.01 x 10ˆ99 give or take a few, since I am sure there are pathological cases (all 1's or all 0's) removed.

We can do some basic analysis assuming it is correct and abandon those results if it turns out to be false. This would be something that we would ask the vendor before we went to the bank with it.

2.3 ORDER_NUMBER

Anyways, on to the order_number field. We can ask the same questions as we did for user_id and note that in theory, this field should be particularly cleaned. The reason is that I would expect that **order_number makes sense as a primary key** for this dataset. If there is more than one row for each order_number, that would raise some serious questions about duplication.

Below, we run a few commands to check for missing values, whether or not the same order_number is used by different merchants, and then, if we find misssing values or dupes, we will try to gain some basic information about some of those rows.

```
[8]: rawData["order_number"].isnull().value_counts()
```

```
[8]: False    1837187
     True       39873
     Name: order_number, dtype: int64
```

```
[9]: rawData[rawData["order_number"].isnull() == True]["email_subject"].
      ↪value_counts()
```

```
[9]: Confirmation: You just got paid                       14841
     ****** getting paid                                   11811
     Lyft Tip Increase Receipt                              4248
     Your Ride Pass is ready to roll                        2536
     You're getting paid                                     481
                                                           …
     *****'s last 44 minutes riding with Uber **               1
     ********* last 4**** minutes riding with Uber **          1
     *******'s last 169 minutes riding with Uber **            1
     *******'s last 285 minutes riding with Uber **            1
     ******'s last 64 minutes riding with Uber **              1
     Name: email_subject, Length: 537, dtype: int64
```

There are missing values (that is, null order_numbers) in this dataset. If we are interested in quality data, a record with no order_number is probably not good data. In this case, it is a relatively small amount of data (roughly 2% of the data). If we take a cursory look at the the emails associated with the data that are missing order_numbers, we can see that they do not appear to be passenger pickups. Rather, they refer to getting paid, or buying a ride pass, or getting a tip increase.

To simplify the analysis, we can remove these rows from the data set and continue with only those rows that have an order_number. We can always do a separate analysis in the future for the data that do not have an order_number. So let’s remove those cases. Let’s move on to cardinality.

```
[10]: #Fix for missing order_number
      rawData = rawData[rawData['order_number'].notna()]
```

```
[11]: multipleOrders = rawData[rawData.groupby('order_number')['merchant_name'].
       ↪transform('nunique') > 1]
      multipleOrders
```

```
[11]: Empty DataFrame
      Columns: [merchant_name, user_id, order_number, order_time, email_time,
      insert_time, update_time, order_total_amount, order_points, order_shipping,
      order_tax, order_subtotal, order_total_qty, product_description,
      product_subtitle, item_quantity, item_price, digital_transaction, checksum,
      product_reseller, product_category, order_discount, SKU, item_id, order_pickup,
      from_domain, email_subject, delivery_date, start_source_folder_date,
      end_source_folder_date, file_id, source_dttimestamp, dttimestamp]
      Index: []

      [0 rows x 33 columns]
```

```
[12]: rawData[rawData["order_number"] == 'f53f4d13-e355-442a-af0e-385bbab1e20d']
```

[12]:

```
        merchant_name                                            user_id  \
642787           Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…
1635799          Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…

                                 order_number               order_time  \
642787   f53f4d13-e355-442a-af0e-385bbab1e20d  2017-08-11 00:00:00.000
1635799  f53f4d13-e355-442a-af0e-385bbab1e20d  2017-08-11 00:00:00.000

                      email_time              insert_time  \
642787   2017-08-11 09:19:17.000  2017-08-11 09:21:39.000
1635799  2017-08-11 09:19:17.000  2017-08-11 09:20:52.000

                     update_time  order_total_amount  order_points  \
642787   2019-02-03 09:29:29.000                8.21           NaN
1635799  2019-02-03 10:05:16.000                8.21           NaN

         order_shipping  …  item_id  order_pickup       from_domain  \
642787              NaN  …      NaN             1  uber.us@uber.com
1635799             NaN  …      NaN             1  uber.us@uber.com

                                email_subject            delivery_date  \
642787   Your Thursday evening trip with Uber  2020-02-04 00:00:00.000
1635799    ****** took an Uber trip on Aug 11  2020-02-04 00:00:00.000

         start_source_folder_date  end_source_folder_date  file_id  \
642787                 2020-02-05              2020-02-05      NaN
1635799                2020-02-05              2020-02-05      NaN

              source_dttimestamp              dttimestamp
642787   2020-02-06 04:58:04.987  2020-02-06 21:58:57.453
1635799  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453

[2 rows x 33 columns]
```

[13]:

```
rawData[rawData["order_number"] == 'eba603be-154a-41fa-9108-3d79437dda0a']
```

[13]:

```
        merchant_name                                            user_id  \
685409           Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…
1637444          Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…

                                 order_number               order_time  \
685409   eba603be-154a-41fa-9108-3d79437dda0a  2019-05-09 00:00:00.000
1637444  eba603be-154a-41fa-9108-3d79437dda0a  2019-05-09 00:00:00.000

                      email_time              insert_time  \
685409   2019-05-10 01:53:52.000  2019-05-10 01:54:14.000
1637444  2019-05-10 01:53:55.000  2019-05-10 01:54:13.000
```

```
                     update_time  order_total_amount  order_points  \
685409   2019-05-10 01:54:14.000               10.33           NaN
1637444  2019-05-10 01:54:13.000               10.33           NaN

         order_shipping  …  item_id  order_pickup       from_domain  \
685409              NaN  …      NaN             1  uber.us@uber.com
1637444             NaN  …      NaN             1  uber.us@uber.com

                               email_subject            delivery_date  \
685409   Your Thursday evening trip with Uber  2020-02-04 00:00:00.000
1637444  Your Thursday evening trip with Uber  2020-02-04 00:00:00.000

         start_source_folder_date  end_source_folder_date  file_id  \
685409                 2020-02-05              2020-02-05      NaN
1637444                2020-02-05              2020-02-05      NaN

              source_dttimestamp              dttimestamp
685409   2020-02-06 04:58:04.987  2020-02-06 21:58:57.453
1637444  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453

[2 rows x 33 columns]
```

```
[14]: rawData[rawData["order_number"] == 'c5e4242c-fd0f-40fb-83b3-b7c6da3d5aaa']
```

```
[14]:         merchant_name                                            user_id  \
      1826464          Uber  49db41a62acde6bf83031945f020d4fcc62fb31d4b4bd3…
      1827339          Uber  f1a1a64fa4a03a39c0388410c28418010c09d5eca78fff…

                                       order_number               order_time  \
      1826464  c5e4242c-fd0f-40fb-83b3-b7c6da3d5aaa  2018-10-04 00:00:00.000
      1827339  c5e4242c-fd0f-40fb-83b3-b7c6da3d5aaa  2018-10-04 00:00:00.000

                            email_time              insert_time  \
      1826464  2018-10-04 20:37:36.000  2019-09-13 06:29:09.000
      1827339  2018-10-04 20:37:39.000  2019-09-19 01:35:28.000

                           update_time  order_total_amount  order_points  \
      1826464  2019-09-13 06:29:09.000                 5.0           NaN
      1827339  2019-09-19 01:35:28.000                 5.0           NaN

               order_shipping  …  item_id  order_pickup       from_domain  \
      1826464             NaN  …      NaN             1  uber.us@uber.com
      1827339             NaN  …      NaN             1  uber.us@uber.com

                                                 email_subject  \
      1826464  Receipt for canceled trip on Thursday afternoon
```

```
1827339  Receipt for canceled trip on Thursday afternoon

                   delivery_date  start_source_folder_date  \
1826464  2020-02-04 00:00:00.000                2020-02-05
1827339  2020-02-04 00:00:00.000                2020-02-05

         end_source_folder_date  file_id       source_dttimestamp  \
1826464              2020-02-05      NaN  2020-02-06 04:58:04.987
1827339              2020-02-05      NaN  2020-02-06 04:58:04.987

                     dttimestamp
1826464  2020-02-06 21:58:57.453
1827339  2020-02-06 21:58:57.453

[2 rows x 33 columns]
```

[15]:
```
rawData[rawData["order_number"] == '54f01ff4-47f2-4b90-b156-4ad5f6fa3466']
```

[15]:
```
        merchant_name                                            user_id  \
989232           Uber  49db41a62acde6bf83031945f020d4fcc62fb31d4b4bd3…
1323144          Uber  f1a1a64fa4a03a39c0388410c28418010c09d5eca78fff…

                                 order_number               order_time  \
989232   54f01ff4-47f2-4b90-b156-4ad5f6fa3466  2017-10-02 00:00:00.000
1323144  54f01ff4-47f2-4b90-b156-4ad5f6fa3466  2017-10-02 00:00:00.000

                      email_time              insert_time  \
989232   2017-10-03 02:33:20.000  2019-09-13 05:02:02.000
1323144  2017-10-03 02:33:21.000  2019-09-18 23:33:30.000

                     update_time  order_total_amount  order_points  \
989232   2019-09-13 05:02:02.000               27.85           NaN
1323144  2019-09-18 23:33:30.000               27.85           NaN

         order_shipping  …  item_id  order_pickup  \
989232              NaN  …      NaN             1
1323144             NaN  …      NaN             1

                           from_domain                       email_subject  \
989232   receipts.nyc.suburbs@uber.com  Your Monday evening trip with Uber
1323144  receipts.nyc.suburbs@uber.com  Your Monday evening trip with Uber

                   delivery_date  start_source_folder_date  \
989232   2020-02-04 00:00:00.000                2020-02-05
1323144  2020-02-04 00:00:00.000                2020-02-05

         end_source_folder_date  file_id       source_dttimestamp  \
989232               2020-02-05      NaN  2020-02-06 04:58:04.987
1323144              2020-02-05      NaN  2020-02-06 04:58:04.987

                     dttimestamp
989232   2020-02-06 21:58:57.453
1323144  2020-02-06 21:58:57.453

[2 rows x 33 columns]
```

Another thing that we notice is that there are also cases where the order_number is shared by multiple rows, and they do seem to be the same transaction. We can tell by observing a few examples that nearly everything is the same for these rows, with minor differences. Sometimes, the user_id is different. Sometimes, the user_id is the same and there is a small difference in the insert and update times (which also seem to always be the same, making the columns redundant). We can kill several birds with one stone and look into those cases where there is more than one row for the same order_number and determine how many occur that are of this latter sort (i.e. they have the same data, in essence, but with a minor difference in their datetime of insertion / update). In that case, we can select one of the two and remove the other. But for the first case (i.e. they have different user_ids), this might indicate a potentially duplicate user_id.

In either case, let's start by checking the assumption that insert_time and update_time are the same.

[16]:
```
rawData[rawData.insert_time != rawData.update_time]
```

[16]:
```
        merchant_name                                                  user_id  \
15               Lyft  2b1abda962224acdfacdb7d774cfd444463f4a4cd4438a…
20               Uber  ff9f60463b4677e745a0f3953bc63446c41d6895eee2e0…
21               Uber  d34e4f4290e3a873c42269b218a6f334a546ac931daf88…
22               Uber  ff9f60463b4677e745a0f3953bc63446c41d6895eee2e0…
24               Lyft  d34e4f4290e3a873c42269b218a6f334a546ac931daf88…
…               …                                                 …
1877052          Uber  815c1d46eda38992bf976aa45e10fd81c871fa77b64e6b…
1877053          Uber  263f69452a0c2d46968effe172ba5f4f20785dc0640f85…
1877054          Uber  415b8e1517a45479a2cbabbad5afc6a6cd5f12bc2aa8dd…
1877055          Uber  d3862ea64b55b9107d80ed5ff4eaa8f400ec7c2574562e…
1877056          Uber  7de81565b4aa4ca57f1e30553c163783353006f4039ca2…

                                 order_number               order_time  \
15                        1102508880002627982  2018-02-18 17:00:00.000
20       75df2365-a71e-447a-bed4-0d9ab87e490d  2018-03-08 00:00:00.000
21       d4596113-443b-425b-b5dd-a016ba3292d6  2018-08-11 00:00:00.000
22       eeefb0f8-9228-4e7f-8543-e2043ed0462e  2019-09-02 00:00:00.000
24                        1065615207807636404  2017-11-11 09:00:00.000
…                                       …                      …
1877052  e9d1e1af-f8c2-45f9-9241-65040369cf75  2018-03-31 00:00:00.000
1877053  6e11bab3-3b10-40c2-970e-66bdf17effdb  2017-08-15 00:00:00.000
1877054  7cc351eb-ab18-4b32-b474-c5cc8d233252  2017-07-26 00:00:00.000
```

```
1877055  1d15fd3c-487f-4847-806c-63b5723e2bff                      NaN
1877056  664a6bea-6ab3-40e1-97fc-d309e4d46a35                      NaN

                      email_time             insert_time  \
15       2018-02-20 01:16:14.000  2018-02-20 01:16:22.000
20       2018-03-08 12:41:23.000  2018-11-07 02:43:53.000
21       2018-08-12 00:41:09.000  2018-08-12 00:59:36.000
22       2019-09-03 03:49:19.000  2019-09-03 04:04:34.000
24       2017-11-12 15:22:52.000  2017-11-12 15:23:18.000
…                              …                        …
1877052  2018-04-01 01:32:50.000  2018-04-01 01:32:52.000
1877053  2017-08-15 23:24:49.000  2017-08-15 23:25:14.000
1877054  2017-07-27 03:50:16.000  2017-07-27 03:50:35.000
1877055  2018-02-07 01:55:34.000  2020-02-01 12:12:08.000
1877056  2018-06-08 04:34:48.000  2018-06-08 04:37:19.000

                     update_time  order_total_amount  order_points  \
15       2020-02-19 15:44:28.000               12.69           NaN
20       2020-02-06 17:48:29.000               40.41           NaN
21       2020-02-06 13:48:02.000                9.19           NaN
22       2020-02-06 18:28:40.000               14.75           NaN
24       2020-02-06 13:23:07.000               18.17           NaN
…                              …                 …            …
1877052  2018-09-21 04:52:25.000                5.00           NaN
1877053  2018-09-21 05:16:10.000                5.00           NaN
1877054  2018-09-21 06:25:06.000                5.00           NaN
1877055  2020-02-01 12:12:32.000                0.00           NaN
1877056  2019-03-18 17:56:25.000                0.00           NaN

         order_shipping  …  item_id  order_pickup            from_domain  \
15                  NaN  …      NaN             1  no-reply@lyftmail.com
20                  NaN  …      NaN             1       uber.us@uber.com
21                  NaN  …      NaN             1       uber.us@uber.com
22                  NaN  …      NaN             1       uber.us@uber.com
24                  NaN  …      NaN             1  no-reply@lyftmail.com
…                     …  …        …           …                      …
1877052             NaN  …      NaN             1       uber.us@uber.com
1877053             NaN  …      NaN             1       uber.us@uber.com
1877054             NaN  …      NaN             1       uber.us@uber.com
1877055             NaN  …      NaN             3       uber.us@uber.com
1877056             NaN  …      NaN             3       uber.us@uber.com

                                    email_subject  \
15                   Your ride with ****** on February 18
20       Thanks for *******! ***** updated your Thursda…
21                   Your Saturday evening trip with Uber
22           [Business] Your Monday evening trip with Uber
```

```
24                    Your ride with ****** on November 11
…                                                    …
1877052      Receipt for canceled trip on Saturday evening
1877053     Receipt for canceled trip on Tuesday afternoon
1877054     Receipt for canceled trip on Wednesday evening
1877055       We corrected the fare for your ride on Feb 6
1877056       We corrected the fare for your ride on Jun 8

                   delivery_date  start_source_folder_date  \
15       2020-02-20 00:00:00.000                2020-02-20
20       2020-02-07 00:00:00.000                2020-02-07
21       2020-02-07 00:00:00.000                2020-02-07
22       2020-02-07 00:00:00.000                2020-02-07
24       2020-02-07 00:00:00.000                2020-02-07
…                            …                       …
1877052  2020-02-04 00:00:00.000                2020-02-05
1877053  2020-02-04 00:00:00.000                2020-02-05
1877054  2020-02-04 00:00:00.000                2020-02-05
1877055  2020-02-04 00:00:00.000                2020-02-05
1877056  2020-02-04 00:00:00.000                2020-02-05

         end_source_folder_date  file_id       source_dttimestamp  \
15                   2020-02-20      NaN  2020-02-21 01:12:01.170
20                   2020-02-07      NaN  2020-02-08 01:13:23.440
21                   2020-02-07      NaN  2020-02-08 01:13:23.440
22                   2020-02-07      NaN  2020-02-08 01:13:23.440
24                   2020-02-07      NaN  2020-02-08 01:13:23.440
…                            …       …                       …
1877052              2020-02-05      NaN  2020-02-06 04:58:04.987
1877053              2020-02-05      NaN  2020-02-06 04:58:04.987
1877054              2020-02-05      NaN  2020-02-06 04:58:04.987
1877055              2020-02-05      NaN  2020-02-06 04:58:04.987
1877056              2020-02-05      NaN  2020-02-06 04:58:04.987

                     dttimestamp
15       2020-02-21 01:44:32.740
20       2020-02-08 01:42:23.217
21       2020-02-08 01:42:23.217
22       2020-02-08 01:42:23.217
24       2020-02-08 01:42:23.217
…                            …
1877052  2020-02-06 21:58:57.453
1877053  2020-02-06 21:58:57.453
1877054  2020-02-06 21:58:57.453
1877055  2020-02-06 21:58:57.453
1877056  2020-02-06 21:58:57.453
```

```
[1152312 rows x 33 columns]
```

It looks like many rows have an appreciable difference between the two times, so we should keep both. If we are going to consolidate or dedupe the rows where we have the same user_id on two copies of the same order_number, the story is going to be more complicated. Because the important data (such as the quantitative fields) appear to be the same and the volume of such cases is low, it would make sense to preliminarily delete one of the two rows in these cases and then consult the vendor as to why this would be the case.

We will count the number of these cases, show an example of one of them, remove the dupes and then check that they have been successfully removed.

[17]:
```
#get the cases where the same order_number shows up more than one, with the␣
 ↪same user_id attached.
r = rawData.groupby(['order_number','user_id'])
r.filter(lambda x: len(x) > 1)
```

[17]:
```
        merchant_name                                           user_id  \
941              Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…
1253             Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…
4227             Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…
4932             Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…
6215             Uber  2bf147359455550128c01affb30131c96caebc94e00d59…
…                 …                                                 …
1870540          Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…
1874113          Uber  5c3bb245c6cb9def713d87bc8463b1629c18014a992c59…
1874889          Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…
1875005          Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…
1875191          Uber  5c3bb245c6cb9def713d87bc8463b1629c18014a992c59…

                                 order_number               order_time  \
941      e7f516e9-faeb-48da-81b4-54ba7817c4ba  2017-10-21 00:00:00.000
1253     1f42a518-9114-4b9f-ab80-9f870bee4ff3  2019-05-12 00:00:00.000
4227     34b86217-7309-4446-b913-6142c4f2c7d5  2017-07-23 00:00:00.000
4932     2cb7535e-9169-4954-b25f-ad28c3e32911  2018-08-20 00:00:00.000
6215     de3616b0-3e09-4e2e-b035-ade881ee5518  2018-12-05 00:00:00.000
…                                           …                      …
1870540  f1ae9806-7681-4a55-af40-b9b569b04b21  2017-08-03 00:00:00.000
1874113  70e6f4ef-fc49-4fda-995a-f7a857ec745b  2019-10-29 00:00:00.000
1874889  92280897-4d62-4b24-b7b8-230977180259  2018-05-25 00:00:00.000
1875005  98e672cf-c357-47d6-9ba8-409a3aff60e2  2018-07-04 00:00:00.000
1875191  d6cbc033-731c-46cd-9821-7d4ebed58b79  2018-12-07 00:00:00.000

                      email_time              insert_time  \
941      2017-10-21 07:38:00.000  2017-10-21 07:39:08.000
1253     2019-05-13 04:30:50.000  2019-05-13 04:31:08.000
4227     2017-07-24 00:10:33.000  2017-07-24 08:34:13.000
4932     2018-08-20 16:21:16.000  2018-08-20 16:51:52.000
```

```
6215     2018-12-05 19:53:59.000  2018-12-05 20:28:17.000
…                          …                      …
1870540  2017-08-03 23:02:42.000  2017-08-03 23:02:58.000
1874113  2019-10-29 23:32:12.000  2019-10-29 23:47:45.000
1874889  2018-05-26 00:58:32.000  2018-05-26 00:58:48.000
1875005  2018-07-05 01:38:50.000  2018-07-05 01:51:14.000
1875191  2018-12-07 20:45:24.000  2018-12-07 21:28:30.000

                     update_time  order_total_amount  order_points  \
941      2019-02-03 09:29:38.000               11.85           NaN
1253     2019-05-13 04:31:08.000               14.93           NaN
4227     2019-02-03 10:05:18.000               27.39           NaN
4932     2019-02-03 09:30:34.000                5.82           NaN
6215     2019-11-25 23:52:17.000               13.76           NaN
…                          …                 …           …
1870540  2019-02-03 09:29:30.000               69.64           NaN
1874113  2019-10-29 23:47:45.000                8.00           NaN
1874889  2018-05-26 00:58:48.000               33.64           NaN
1875005  2019-02-03 10:06:01.000               16.03           NaN
1875191  2019-07-22 12:29:41.000                7.72           NaN

         order_shipping  …  item_id  order_pickup  \
941                 NaN  …      NaN             1
1253                NaN  …      NaN             1
4227                NaN  …      NaN             1
4932                NaN  …      NaN             1
6215                NaN  …      NaN             1
…                 …  …      …           …
1870540             NaN  …      NaN             1
1874113             NaN  …      NaN             1
1874889             NaN  …      NaN             1
1875005             NaN  …      NaN             1
1875191             NaN  …      NaN             1

                           from_domain  \
941                   uber.us@uber.com
1253                  uber.us@uber.com
4227                  uber.us@uber.com
4932                  uber.us@uber.com
6215                  uber.us@uber.com
…                                …
1870540               uber.us@uber.com
1874113  receipts.nyc.suburbs@uber.com
1874889               uber.us@uber.com
1875005               uber.us@uber.com
1875191  receipts.nyc.suburbs@uber.com
```

```
                                             email_subject  \
941                   Your Saturday evening trip with Uber
1253                    Your Sunday evening trip with Uber
4227                    ****** took an Uber trip on Jul 23
4932                    Your Monday morning trip with Uber
6215     [Business] Your Wednesday afternoon trip with …
…                                                    …
1870540               Your Thursday evening trip with Uber
1874113                Your Tuesday evening trip with Uber
1874889                 Your Friday evening trip with Uber
1875005              Your Wednesday evening trip with Uber
1875191               Your Friday afternoon trip with Uber

                   delivery_date  start_source_folder_date  \
941      2020-02-04 00:00:00.000                2020-02-05
1253     2020-02-04 00:00:00.000                2020-02-05
4227     2020-02-04 00:00:00.000                2020-02-05
4932     2020-02-04 00:00:00.000                2020-02-05
6215     2020-02-04 00:00:00.000                2020-02-05
…                          …                       …
1870540  2020-02-04 00:00:00.000                2020-02-05
1874113  2020-02-04 00:00:00.000                2020-02-05
1874889  2020-02-04 00:00:00.000                2020-02-05
1875005  2020-02-04 00:00:00.000                2020-02-05
1875191  2020-02-04 00:00:00.000                2020-02-05

         end_source_folder_date  file_id       source_dttimestamp  \
941                  2020-02-05      NaN  2020-02-06 04:58:04.987
1253                 2020-02-05      NaN  2020-02-06 04:58:04.987
4227                 2020-02-05      NaN  2020-02-06 04:58:04.987
4932                 2020-02-05      NaN  2020-02-06 04:58:04.987
6215                 2020-02-05      NaN  2020-02-06 04:58:04.987
…                         …      …                      …
1870540              2020-02-05      NaN  2020-02-06 04:58:04.987
1874113              2020-02-05      NaN  2020-02-06 04:58:04.987
1874889              2020-02-05      NaN  2020-02-06 04:58:04.987
1875005              2020-02-05      NaN  2020-02-06 04:58:04.987
1875191              2020-02-05      NaN  2020-02-06 04:58:04.987

                     dttimestamp
941      2020-02-06 21:58:57.453
1253     2020-02-06 21:58:57.453
4227     2020-02-06 21:58:57.453
4932     2020-02-06 21:58:57.453
6215     2020-02-06 21:58:57.453
…                          …
1870540  2020-02-06 21:58:57.453
```

```
1874113  2020-02-06 21:58:57.453
1874889  2020-02-06 21:58:57.453
1875005  2020-02-06 21:58:57.453
1875191  2020-02-06 21:58:57.453

[2000 rows x 33 columns]
```

[18]:

```
#an example of one of the dupes queried in the cell above.
rawData[rawData['order_number'] == 'e7f516e9-faeb-48da-81b4-54ba7817c4ba']
```

[18]:

```
       merchant_name                                            user_id  \
941             Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…
730095          Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…

                                order_number               order_time  \
941     e7f516e9-faeb-48da-81b4-54ba7817c4ba  2017-10-21 00:00:00.000
730095  e7f516e9-faeb-48da-81b4-54ba7817c4ba  2017-10-21 00:00:00.000

                     email_time              insert_time  \
941     2017-10-21 07:38:00.000  2017-10-21 07:39:08.000
730095  2017-10-21 07:38:02.000  2017-10-21 07:39:59.000

                    update_time  order_total_amount  order_points  \
941     2019-02-03 09:29:38.000               11.85           NaN
730095  2019-02-03 10:05:27.000               11.85           NaN

        order_shipping  …  item_id  order_pickup       from_domain  \
941                NaN  …      NaN             1  uber.us@uber.com
730095             NaN  …      NaN             1  uber.us@uber.com

                               email_subject            delivery_date  \
941     Your Saturday evening trip with Uber  2020-02-04 00:00:00.000
730095  Your Saturday evening trip with Uber  2020-02-04 00:00:00.000

        start_source_folder_date  end_source_folder_date  file_id  \
941                   2020-02-05              2020-02-05      NaN
730095                2020-02-05              2020-02-05      NaN

             source_dttimestamp              dttimestamp
941     2020-02-06 04:58:04.987  2020-02-06 21:58:57.453
730095  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453

[2 rows x 33 columns]
```

There are 2000 such rows, so there should be 1000 rows removed by our dedupe effort. To do this, we will need to define a new rank column which will assign the rows we want to delete a value larger than 1. That way, we can remove them and then remove the column. To select, we will keep the later of the two based on which one has the larger (that is, more recent) update_time.

[19]:
```
rawData['order_rank'] = rawData.sort_values(["update_time"], ascending =␣
↪[False]).groupby(
    ['order_number','user_id']).cumcount() + 1
rawData = rawData[rawData['order_rank'] == 1]
```

[20]:
```
#now returns one row, as it should.
rawData[rawData['order_number'] == 'e7f516e9-faeb-48da-81b4-54ba7817c4ba']
```

[20]:
```
       merchant_name                                            user_id  \
730095          Uber  0a476432e254265e760aaef7a7ad76c76cd456b8ae4ef4…

                                order_number               order_time  \
730095  e7f516e9-faeb-48da-81b4-54ba7817c4ba  2017-10-21 00:00:00.000

                     email_time              insert_time  \
730095  2017-10-21 07:38:02.000  2017-10-21 07:39:59.000

                    update_time  order_total_amount  order_points  \
730095  2019-02-03 10:05:27.000               11.85           NaN

        order_shipping  …  order_pickup       from_domain  \
730095             NaN  …             1  uber.us@uber.com

                               email_subject            delivery_date  \
730095  Your Saturday evening trip with Uber  2020-02-04 00:00:00.000

       start_source_folder_date  end_source_folder_date  file_id  \
730095               2020-02-05              2020-02-05      NaN

             source_dttimestamp              dttimestamp order_rank
730095  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453          1

[1 rows x 34 columns]
```

As we can see, the dupes where the same user_id has been selected have been removed. Now all we are left with is the case where order_numbers are duplicated with more than one user_id in the picture. The question to ask here is: are these necessarily bad rows?

One way to do this is to examine which user_ids are duplicated. When we run the below code, we can see that there are only 29 user_ids that account for the 1126 duplicated order_numbers (queries to validate these figures are below).

A question that we might want to ask the vendor is whether or not it is possible for two user_ids to legitimately be attached to the same order_number. If that is possible, then user_id would have to be a part of our primary key. Otherwise, we would want to find the causes of the duplication and remove it.

For now, the more likely scenario is that these are merely duplicates and we should be able to simply dedupe them by taking, uniformly, the smaller value. Because there are not many rows that are affected by this, it seems like a good idea for now. I would reiterate though: it is important that we consult the vendor as to whether or not it would be possible for more than one user_id to be connected to the same order.

We do this below.

```
[21]: #this will validatae that there are only 29 user_ids that are responsible for␣
       ↪the duplication.
      s = rawData[rawData.groupby('order_number')['user_id'].transform('nunique') >␣
       ↪1]["order_number"]
      len(rawData[rawData['order_number'].isin(list(s))]['user_id'].unique())
```

```
[21]: 29
```

```
[22]: #this will validate that there are 1126 orders which are affected.
      rawData[rawData.groupby('order_number')['user_id'].transform('nunique') >␣
       ↪1]["order_number"]
```

```
[22]: 2598       a0e69d1e-04a7-4159-883a-2d13ca979625
      3879       9010d559-010e-4d83-b51f-69a47d9724b6
      4091       da27a1bd-c598-4a1b-a2b3-92cc45a59c84
      7026       d614df14-7956-45ec-a253-08e2f9438741
      8506       dce8f194-a9ea-4382-ac52-aad7b7a4e147
                                 …
      1871310    8709c6dd-d8b1-4ae0-925b-07878016ad26
      1872414    d88ca05a-ade3-4238-a801-3458fcafe071
      1873741    b51e8599-7a11-4025-9124-8a5c55d9c220
      1874972    1e785c03-3b64-4e88-b50f-fe1107942e93
      1875612    9b829a98-c08d-43c6-ad2d-108876790b03
      Name: order_number, Length: 1126, dtype: object
```

```
[23]: #remove the remaining dupes
      rawData['order_rank2'] = rawData.sort_values(["user_id"], ascending = [True]).
       ↪groupby(
          ['order_number']).cumcount() + 1
      rawData = rawData[rawData['order_rank2'] == 1]
```

```
<ipython-input-23-1abd51b8eb42>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  rawData['order_rank2'] = rawData.sort_values(["user_id"], ascending =
[True]).groupby(
```

[24]:
```
#prove that this is fixed
s = rawData[rawData.groupby('order_number')['user_id'].transform('nunique') >␣
 ↪1]["order_number"]
len(rawData[rawData['order_number'].isin(list(s))]['user_id'].unique())
```

[24]:
```
0
```

We are good on missing order_numbers and duplication. A final issue that we need to resolve with regards to the order_numbers is their data type. Is there any regularity to this column? Let's check this last and then we will have a pristine order_number column that we can use as a primary key!

[25]:
```
rawData["order_number"].apply(len).value_counts()
```

[25]:
```
36    1283310
19     552235
13         63
18         12
9           3
8           1
Name: order_number, dtype: int64
```

As we can see from the above, the majority of the data are either 36 characters long or 19 characters long, possibly corresponding to the two different formats that we get from Uber and from Lyft. We will validate this below.

The others, though, appear to be missing digits. Let's validate the assumption that the two largest categories correspond to Uber and Lyft and then let's look at the remaining cases.

[26]:
```
#given that it is 36 characters long, which vendor is it?
rawData[rawData["order_number"].apply(lambda x: len(x) == 36)]['merchant_name'].
 ↪value_counts()
```

[26]:
```
Uber    1283310
Name: merchant_name, dtype: int64
```

[27]:
```
#given that it is 19 characters long, which vendor is it?
rawData[rawData["order_number"].apply(lambda x: len(x) == 19)]['merchant_name'].
 ↪value_counts()
```

[27]:
```
Lyft    552235
Name: merchant_name, dtype: int64
```

[28]:
```
#given that it is 13 characters long,which vendor is it?
rawData[rawData["order_number"].apply(lambda x: len(x) == 13)]['merchant_name'].
 ↪value_counts()
```

```
[28]: Uber    63
      Name: merchant_name, dtype: int64
```

```
[29]: #given that it is 18 characters long, which vendor is it?
      rawData[rawData["order_number"].apply(lambda x: len(x) == 18)]['merchant_name'].
       ↪value_counts()
```

```
[29]: Lyft    12
      Name: merchant_name, dtype: int64
```

```
[30]: #given that it is 9 characters long, which vendor it is?
      rawData[rawData["order_number"].apply(lambda x: len(x) == 9)]['merchant_name'].
       ↪value_counts()
```

```
[30]: Lyft    3
      Name: merchant_name, dtype: int64
```

```
[31]: #given that it is 8 characters long, which vendor is it?
      rawData[rawData["order_number"].apply(lambda x: len(x) == 8)]['merchant_name'].
       ↪value_counts()
```

```
[31]: Lyft    1
      Name: merchant_name, dtype: int64
```

Looks like our assumption was correct: **Uber order_numbers typically come in with 36 characters and Lyft order_numbers typically come in with 19 characters.** The others are likely deficient or corrupted order_numbers.

We would need to inquire as to why these came in corrupted. It makes sense for us to remove these bad rows from the data set and continue on, with a pristine order_number column which can be used as a primary key for the row.

```
[32]: #make sure we keep the good order_numbers.
      rawData = rawData[rawData["order_number"].apply(lambda x: len(x) >= 19)]
```

```
[33]: rawData["order_number"].apply(len).value_counts()
```

```
[33]: 36    1283310
      19     552235
      Name: order_number, dtype: int64
```

That concludes our look at the order_number field. This field should be useable as a primary key, and we will return to it in the later part of the notebook when we build our the final table.

2.4 DATE FIELDS

The next few fields can be treated together, as they are all dates and times for the row. The order_time seems to be the most impotant, since it specifies the date and time that the transaction occurred. Email_time can probably be handy, since I interpret that to be the time that the customer receives an email regarding the order. The next two columns, insert_time and update_time, are likely only used by the data provider to track when the row got inserted and when it was last updated. **These are less useful, I think, because they are likely bound to the data provider's systema and not the source data from Uber and Lyft.**

While we are on the subject of fields that are bound to the data vendor and likely have little bearing on the analysis, I would also include the following fields in that set:

- **checksum:** This has the same structure as a standard MD5 hash sum and is not something we need to worry about.
- **delivery_date:** This seems to be the date that the data was delivered from the vendor.
- **start_source_folder_date:** This likely corresponds to a date on the backend.
- **end_source_folder_date:** This is similar to start_source_folder_date.
- **file_id:** This is likely the ID of the file on the vendor's side.
- **source_dttimestamp:** This is likely a timestamp indicating when the file was last modified by the source.
- **dttimestamp:** This may be a field that the reseller of the data uses. It is not immediately clear on how this relates to source_dttimestamp, but it is likely another "bookkeeping" column.

These fields might be important if we were loading this data into a relational database or archiving it in some sort of data warehouse and would likely be loaded just for completeness, but for the purposes of this exercise, they are ignored since they do not provide any immediate insights.

We will try to work with order_time, and if for some reason that value is missing, I think that we should use the email_time in its place.

For order_time, we want to identify if there are any missing values, and to see if they are all the same type (that is, a datetime). Potentially, we will want to split this field into dates and times to faciltate analysis. So let's do that first.

[34]:
```
#Extract the order date.
rawData['date_of_order'] = rawData["order_time"].apply(
    lambda x: str(x).replace(".000","")).apply(
        lambda y: y[0:11])
```

[35]:
```
#Extract the order time.
rawData['time_of_order'] = rawData["order_time"].apply(
    lambda x: str(x).replace(".000","")).apply(
        lambda y: y[12:21])
```

After we have done this, we next need to look into which ones are not valid dates. To replace these, we will want to use the email date and time, since these should be reasonably close to the actual date of the transaction.

[36]:
```
#Get the cases where the time is empty. This is a clear indication that the␣
 ↪original was NaN.
rawData[rawData["time_of_order"] == ""]
```

```
[36]:
        merchant_name                                            user_id  \
60               Uber  60aeb143258e1aa9a9c11268b385a548b67d1b0a3a8db2…
61               Uber  60aeb143258e1aa9a9c11268b385a548b67d1b0a3a8db2…
102              Uber  60f694086ec28ff67feff1656580b0534c107088b122e9…
103              Uber  bf013a62bd037a392a42aa587ef4d5ae228e666bc701df…
1083             Uber  8fec8dcc2e63a92d0d8adf513861a4720657e6f78fb808…
…               …                                                …
1876272          Uber  0e4c8c48e1fcdd880e32dce19ba8c9b6913bfeb6168fa8…
1876336          Uber  5b3243ecf544f9e1f1d3351cc19aabfce6a58906cd2eba…
1877055          Uber  d3862ea64b55b9107d80ed5ff4eaa8f400ec7c2574562e…
1877056          Uber  7de81565b4aa4ca57f1e30553c163783353006f4039ca2…
1877059          Lyft  bdaeb8a007ac5004b9e439e896fa380828ad0236ade1aa…

                                 order_number order_time  \
60       030afba5-c085-47a5-9aa9-2e879dfb82d2        NaN
61       9c7cd4b9-9daf-4887-95f9-116947f5b2cb        NaN
102      19d175f5-ecc4-44eb-80ba-964987db8e95        NaN
103      f0e39333-05b0-4325-875e-db449034d668        NaN
1083     74265ee7-ab4e-4cfc-a3b6-1546bbd00320        NaN
…                                       …        …
1876272  18550dfd-5f39-48ec-b7e7-a720c6583c0a        NaN
1876336  61083f81-15a8-5362-82f2-4db3ee0ebd1c        NaN
1877055  1d15fd3c-487f-4847-806c-63b5723e2bff        NaN
1877056  664a6bea-6ab3-40e1-97fc-d309e4d46a35        NaN
1877059                   1309397943170611458        NaN

                      email_time              insert_time  \
60       2018-12-04 08:07:48.000  2018-12-04 08:11:14.000
61       2018-05-07 19:12:47.000  2018-05-07 19:29:48.000
102      2018-02-04 08:22:20.000  2018-11-15 08:50:00.000
103      2018-06-19 14:51:32.000  2018-06-19 13:47:25.000
1083     2018-01-06 21:33:53.000  2018-01-06 21:34:26.000
…                          …                      …
1876272  2018-02-24 13:21:44.000  2018-02-24 13:21:53.000
1876336  2019-12-23 01:40:29.000  2019-12-23 01:40:47.000
1877055  2018-02-07 01:55:34.000  2020-02-01 12:12:08.000
1877056  2018-06-08 04:34:48.000  2018-06-08 04:37:19.000
1877059  2019-08-30 14:16:51.000  2019-08-30 15:20:08.000

                     update_time  order_total_amount  order_points  \
60       2020-02-08 12:54:00.000               34.02           NaN
61       2020-02-08 12:30:19.000                8.14           NaN
102      2019-03-18 21:23:04.000               11.02           NaN
103      2019-11-25 23:21:21.000                0.00           NaN
1083     2019-03-18 20:51:42.000                1.00           NaN
…                          …                 …           …
1876272  2019-03-18 17:34:28.000                0.00           NaN
```

```
1876336  2019-12-23 01:40:47.000               32.54           NaN
1877055  2020-02-01 12:12:32.000                0.00           NaN
1877056  2019-03-18 17:56:25.000                0.00           NaN
1877059  2019-08-30 15:20:08.000               43.44           NaN

         order_shipping  …            delivery_date  \
60                  NaN  …  2020-02-09 00:00:00.000
61                  NaN  …  2020-02-09 00:00:00.000
102                 NaN  …  2020-02-04 00:00:00.000
103                 NaN  …  2020-02-04 00:00:00.000
1083                NaN  …  2020-02-04 00:00:00.000
…                 …  …                      …
1876272             NaN  …  2020-02-04 00:00:00.000
1876336             NaN  …  2020-02-04 00:00:00.000
1877055             NaN  …  2020-02-04 00:00:00.000
1877056             NaN  …  2020-02-04 00:00:00.000
1877059             NaN  …  2020-02-04 00:00:00.000

         start_source_folder_date  end_source_folder_date file_id  \
60                     2020-02-09              2020-02-09     NaN
61                     2020-02-09              2020-02-09     NaN
102                    2020-02-05              2020-02-05     NaN
103                    2020-02-05              2020-02-05     NaN
1083                   2020-02-05              2020-02-05     NaN
…                           …                      …     …
1876272                2020-02-05              2020-02-05     NaN
1876336                2020-02-05              2020-02-05     NaN
1877055                2020-02-05              2020-02-05     NaN
1877056                2020-02-05              2020-02-05     NaN
1877059                2020-02-05              2020-02-05     NaN

              source_dttimestamp              dttimestamp  order_rank  \
60       2020-02-10 01:15:12.837  2020-02-10 01:26:23.327           1
61       2020-02-10 01:15:12.837  2020-02-10 01:26:23.327           1
102      2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
103      2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1083     2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
…                          …                      …         …
1876272  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1876336  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877055  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877056  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877059  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1

         order_rank2 date_of_order time_of_order
60                 1           nan
61                 1           nan
```

```
102                  1           nan
103                  1           nan
1083                 1           nan
…              …           …           …
1876272              1           nan
1876336              1           nan
1877055              1           nan
1877056              1           nan
1877059              1           nan

[7915 rows x 37 columns]
```

[37]:

```
#Another way to produce the set above.
rawData[rawData["date_of_order"] == 'nan']
```

[37]:

```
        merchant_name                                           user_id  \
60               Uber  60aeb143258e1aa9a9c11268b385a548b67d1b0a3a8db2…
61               Uber  60aeb143258e1aa9a9c11268b385a548b67d1b0a3a8db2…
102              Uber  60f694086ec28ff67feff1656580b0534c107088b122e9…
103              Uber  bf013a62bd037a392a42aa587ef4d5ae228e666bc701df…
1083             Uber  8fec8dcc2e63a92d0d8adf513861a4720657e6f78fb808…
…                …                                                 …
1876272          Uber  0e4c8c48e1fcdd880e32dce19ba8c9b6913bfeb6168fa8…
1876336          Uber  5b3243ecf544f9e1f1d3351cc19aabfce6a58906cd2eba…
1877055          Uber  d3862ea64b55b9107d80ed5ff4eaa8f400ec7c2574562e…
1877056          Uber  7de81565b4aa4ca57f1e30553c163783353006f4039ca2…
1877059          Lyft  bdaeb8a007ac5004b9e439e896fa380828ad0236ade1aa…

                                 order_number order_time  \
60       030afba5-c085-47a5-9aa9-2e879dfb82d2        NaN
61       9c7cd4b9-9daf-4887-95f9-116947f5b2cb        NaN
102      19d175f5-ecc4-44eb-80ba-964987db8e95        NaN
103      f0e39333-05b0-4325-875e-db449034d668        NaN
1083     74265ee7-ab4e-4cfc-a3b6-1546bbd00320        NaN
…                                       …        …
1876272  18550dfd-5f39-48ec-b7e7-a720c6583c0a        NaN
1876336  61083f81-15a8-5362-82f2-4db3ee0ebd1c        NaN
1877055  1d15fd3c-487f-4847-806c-63b5723e2bff        NaN
1877056  664a6bea-6ab3-40e1-97fc-d309e4d46a35        NaN
1877059                   1309397943170611458        NaN

                      email_time              insert_time  \
60       2018-12-04 08:07:48.000  2018-12-04 08:11:14.000
61       2018-05-07 19:12:47.000  2018-05-07 19:29:48.000
102      2018-02-04 08:22:20.000  2018-11-15 08:50:00.000
103      2018-06-19 14:51:32.000  2018-06-19 13:47:25.000
1083     2018-01-06 21:33:53.000  2018-01-06 21:34:26.000
```

```
…                          …                      …
1876272  2018-02-24 13:21:44.000  2018-02-24 13:21:53.000
1876336  2019-12-23 01:40:29.000  2019-12-23 01:40:47.000
1877055  2018-02-07 01:55:34.000  2020-02-01 12:12:08.000
1877056  2018-06-08 04:34:48.000  2018-06-08 04:37:19.000
1877059  2019-08-30 14:16:51.000  2019-08-30 15:20:08.000

                     update_time  order_total_amount  order_points  \
60       2020-02-08 12:54:00.000               34.02           NaN
61       2020-02-08 12:30:19.000                8.14           NaN
102      2019-03-18 21:23:04.000               11.02           NaN
103      2019-11-25 23:21:21.000                0.00           NaN
1083     2019-03-18 20:51:42.000                1.00           NaN
…                          …                 …           …
1876272  2019-03-18 17:34:28.000                0.00           NaN
1876336  2019-12-23 01:40:47.000               32.54           NaN
1877055  2020-02-01 12:12:32.000                0.00           NaN
1877056  2019-03-18 17:56:25.000                0.00           NaN
1877059  2019-08-30 15:20:08.000               43.44           NaN

         order_shipping  …            delivery_date  \
60                  NaN  …  2020-02-09 00:00:00.000
61                  NaN  …  2020-02-09 00:00:00.000
102                 NaN  …  2020-02-04 00:00:00.000
103                 NaN  …  2020-02-04 00:00:00.000
1083                NaN  …  2020-02-04 00:00:00.000
…                 …  …                      …
1876272             NaN  …  2020-02-04 00:00:00.000
1876336             NaN  …  2020-02-04 00:00:00.000
1877055             NaN  …  2020-02-04 00:00:00.000
1877056             NaN  …  2020-02-04 00:00:00.000
1877059             NaN  …  2020-02-04 00:00:00.000

         start_source_folder_date  end_source_folder_date file_id  \
60                     2020-02-09              2020-02-09     NaN
61                     2020-02-09              2020-02-09     NaN
102                    2020-02-05              2020-02-05     NaN
103                    2020-02-05              2020-02-05     NaN
1083                   2020-02-05              2020-02-05     NaN
…                           …                     …     …
1876272                2020-02-05              2020-02-05     NaN
1876336                2020-02-05              2020-02-05     NaN
1877055                2020-02-05              2020-02-05     NaN
1877056                2020-02-05              2020-02-05     NaN
1877059                2020-02-05              2020-02-05     NaN

              source_dttimestamp              dttimestamp  order_rank  \
60       2020-02-10 01:15:12.837  2020-02-10 01:26:23.327           1
61       2020-02-10 01:15:12.837  2020-02-10 01:26:23.327           1
102      2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
103      2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1083     2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
…                          …                       …         …
1876272  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1876336  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877055  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877056  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877059  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1

         order_rank2 date_of_order time_of_order
60                 1          nan
61                 1          nan
102                1          nan
103                1          nan
1083               1          nan
…             …           …           …
1876272            1          nan
1876336            1          nan
1877055            1          nan
1877056            1          nan
1877059            1          nan

[7915 rows x 37 columns]
```

[38]:

```
#For those cases where the order time is empty, use the time from the␣
 ↪email_time field.
rawData.loc[rawData.time_of_order == "","time_of_order"] =␣
 ↪rawData["email_time"].apply(
    lambda x: str(x).replace(".000","")).apply(
        lambda y: y[12:21])
```

[39]:

```
#Validate that there is an order time for all rows now.
rawData[rawData["time_of_order"] == ""]
```

[39]:

```
Empty DataFrame
Columns: [merchant_name, user_id, order_number, order_time, email_time,
insert_time, update_time, order_total_amount, order_points, order_shipping,
order_tax, order_subtotal, order_total_qty, product_description,
product_subtitle, item_quantity, item_price, digital_transaction, checksum,
product_reseller, product_category, order_discount, SKU, item_id, order_pickup,
from_domain, email_subject, delivery_date, start_source_folder_date,
end_source_folder_date, file_id, source_dttimestamp, dttimestamp, order_rank,
order_rank2, date_of_order, time_of_order]
Index: []
```

```
[0 rows x 37 columns]
```

[40]:

```
#For those cases where the order date is empty, use the date from the␣
↪email_time field.
rawData.loc[rawData.date_of_order == "nan","date_of_order"] =␣
↪rawData["email_time"].apply(
    lambda x: str(x).replace(".000","")).apply(
        lambda y: y[0:11])
```

[41]:

```
#Validate that there is an order date for all rows now.
rawData[rawData["date_of_order"] == 'nan']
```

[41]:

```
Empty DataFrame
Columns: [merchant_name, user_id, order_number, order_time, email_time,
insert_time, update_time, order_total_amount, order_points, order_shipping,
order_tax, order_subtotal, order_total_qty, product_description,
product_subtitle, item_quantity, item_price, digital_transaction, checksum,
product_reseller, product_category, order_discount, SKU, item_id, order_pickup,
from_domain, email_subject, delivery_date, start_source_folder_date,
end_source_folder_date, file_id, source_dttimestamp, dttimestamp, order_rank,
order_rank2, date_of_order, time_of_order]
Index: []

[0 rows x 37 columns]
```

That concludes our treatment of the date and time fields.

2.5 QUANTITATIVE TRANSACTION FIELDS

The next fields that we will treat which can be considered together are the quantitative fields associated with the transaction. Those are:

- **order_total_amount**: Assumed to be a dollar amount.
- **order_points**: Assumed to be an integer.
- **order_shipping**: Assumed to be a dollar amount.
- **order_tax**: Assumed to be a dollar amount.
- **order_subtotal**: Assumed to be a dollar amount.
- **order_total_qty**: Assumed to be an integer.
- **item_quantity**: Assumed to be an integer.
- **item_price**: Assumed to be a dollar amount.
- **order_discount**: Assumed to be a dollar amount.

For each of these, we would expect them all to be numeric. All the dollar values should be floats, and the ones that I would expect to be integral are tagged as such above. Apart from their numerical integrity, we will want to flag any cases where these values are negative and what that would mean. Finally, we will likely want to check that the data are internally consistent. What this means is that there is some equation that holds for the row. Here are a few contenders to check:

order_total_amount = order_subtotal + order_tax + order_shipping - order_discount
item_price * item_quantity = order_subtotal

We will run all of these checks below.

```
[42]: #check for negative order_total_amount.
      rawData[rawData["order_total_amount"] < 0]
```

```
[42]: Empty DataFrame
      Columns: [merchant_name, user_id, order_number, order_time, email_time,
      insert_time, update_time, order_total_amount, order_points, order_shipping,
      order_tax, order_subtotal, order_total_qty, product_description,
      product_subtitle, item_quantity, item_price, digital_transaction, checksum,
      product_reseller, product_category, order_discount, SKU, item_id, order_pickup,
      from_domain, email_subject, delivery_date, start_source_folder_date,
      end_source_folder_date, file_id, source_dttimestamp, dttimestamp, order_rank,
      order_rank2, date_of_order, time_of_order]
      Index: []

      [0 rows x 37 columns]
```

```
[43]: #check for null order_total_amount.
      rawData[rawData.order_total_amount.isnull()]
```

```
[43]:         merchant_name                                            user_id  \
      1360             Lyft  5b69ede3701835a7fb11fee5d42d07c5b80703f05257ea…
      5806             Lyft  1cdb134f3f668022210ac5323672434316f160c21d22bd…
      13499            Lyft  744dc32b48c58b67cabb71739eb6a7958df01178ba33a7…
      14382            Lyft  c8556e625c5f6f9884ef6322d3975b55abed50c3e5ba5a…
      17694            Lyft  4e12961bf71a4282498c57d55b581fbe6508d21fef7471…
      …               …                                                 …
      1822029          Lyft  d3862ea64b55b9107d80ed5ff4eaa8f400ec7c2574562e…
      1836213          Lyft  4d8c4d961238a49b5de02ed9906a5e9535f920146437e8…
      1846284          Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
      1848645          Lyft  848dd50e1001c7d43026f8a9451e8cb4bfdc52bcf23fe6…
      1873504          Lyft  c8556e625c5f6f9884ef6322d3975b55abed50c3e5ba5a…

                      order_number order_time               email_time  \
      1360     1218012321017954872        NaN  2018-12-28 07:49:30.000
      5806     1294076628908439242        NaN  2019-07-21 06:52:42.000
      13499    1251958412834542698        NaN  2019-03-29 16:00:10.000
      14382    1006654894788916950        NaN  2017-06-06 17:47:12.000
      17694    1151098900292235936        NaN  2018-06-30 23:48:27.000
      …                          …        …                       …
      1822029  1176581819408577772        NaN  2018-09-07 06:42:19.000
      1836213  1258037178350048182        NaN  2019-04-14 04:08:18.000
      1846284  1181388796506241698        NaN  2018-09-19 22:27:24.000
      1848645  1096591544868441510        NaN  2018-02-04 02:37:27.000
```

```
1873504  1035972902790946176        NaN  2017-08-24 17:43:18.000

                     insert_time              update_time  order_total_amount  \
1360     2019-12-14 00:47:12.000  2019-12-14 00:47:12.000                 NaN
5806     2019-07-26 10:52:27.000  2019-07-26 10:52:27.000                 NaN
13499    2019-11-30 06:34:47.000  2019-11-30 06:34:47.000                 NaN
14382    2019-12-05 00:58:29.000  2019-12-05 00:58:29.000                 NaN
17694    2019-09-23 14:26:35.000  2019-09-23 14:26:35.000                 NaN
…                          …                      …                 …
1822029  2020-02-01 12:53:11.000  2020-02-01 12:53:11.000                 NaN
1836213  2019-09-17 12:34:13.000  2019-09-17 12:34:13.000                 NaN
1846284  2020-02-17 15:29:31.000  2020-02-17 15:29:31.000                 NaN
1848645  2019-09-27 21:53:24.000  2019-09-27 21:53:24.000                 NaN
1873504  2019-12-05 01:02:16.000  2019-12-05 01:02:16.000                 NaN

         order_points  order_shipping  …            delivery_date  \
1360              NaN             NaN  …  2020-02-04 00:00:00.000
5806              NaN             NaN  …  2020-02-04 00:00:00.000
13499             NaN             NaN  …  2020-02-04 00:00:00.000
14382             NaN             NaN  …  2020-02-04 00:00:00.000
17694             NaN             NaN  …  2020-02-04 00:00:00.000
…               …             …  …                      …
1822029           NaN             NaN  …  2020-02-04 00:00:00.000
1836213           NaN             NaN  …  2020-02-04 00:00:00.000
1846284           NaN             NaN  …  2020-02-18 00:00:00.000
1848645           NaN             NaN  …  2020-02-04 00:00:00.000
1873504           NaN             NaN  …  2020-02-04 00:00:00.000

         start_source_folder_date  end_source_folder_date file_id  \
1360                   2020-02-05              2020-02-05     NaN
5806                   2020-02-05              2020-02-05     NaN
13499                  2020-02-05              2020-02-05     NaN
14382                  2020-02-05              2020-02-05     NaN
17694                  2020-02-05              2020-02-05     NaN
…                           …                     …     …
1822029                2020-02-05              2020-02-05     NaN
1836213                2020-02-05              2020-02-05     NaN
1846284                2020-02-18              2020-02-18     NaN
1848645                2020-02-05              2020-02-05     NaN
1873504                2020-02-05              2020-02-05     NaN

              source_dttimestamp              dttimestamp  order_rank  \
1360     2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
5806     2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
13499    2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
14382    2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
17694    2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
```

```
…                          …                      …         …
1822029  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1836213  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1846284  2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
1848645  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1873504  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1

         order_rank2 date_of_order time_of_order
1360               1   2018-12-28        7:49:30
5806               1   2019-07-21        6:52:42
13499              1   2019-03-29        6:00:10
14382              1   2017-06-06        7:47:12
17694              1   2018-06-30        3:48:27
…              …           …           …
1822029            1   2018-09-07        6:42:19
1836213            1   2019-04-14        4:08:18
1846284            1   2018-09-19        2:27:24
1848645            1   2018-02-04        2:37:27
1873504            1   2017-08-24        7:43:18

[282 rows x 37 columns]
```

[44]:
```
#check for non-null order_points
rawData[rawData.order_points.notnull()]
```

[44]:
```
Empty DataFrame
Columns: [merchant_name, user_id, order_number, order_time, email_time,
insert_time, update_time, order_total_amount, order_points, order_shipping,
order_tax, order_subtotal, order_total_qty, product_description,
product_subtitle, item_quantity, item_price, digital_transaction, checksum,
product_reseller, product_category, order_discount, SKU, item_id, order_pickup,
from_domain, email_subject, delivery_date, start_source_folder_date,
end_source_folder_date, file_id, source_dttimestamp, dttimestamp, order_rank,
order_rank2, date_of_order, time_of_order]
Index: []

[0 rows x 37 columns]
```

[45]:
```
#check for non-null order_shipping.
rawData[rawData.order_shipping.notnull()]
```

[45]:
```
Empty DataFrame
Columns: [merchant_name, user_id, order_number, order_time, email_time,
insert_time, update_time, order_total_amount, order_points, order_shipping,
order_tax, order_subtotal, order_total_qty, product_description,
product_subtitle, item_quantity, item_price, digital_transaction, checksum,
product_reseller, product_category, order_discount, SKU, item_id, order_pickup,
from_domain, email_subject, delivery_date, start_source_folder_date,
end_source_folder_date, file_id, source_dttimestamp, dttimestamp, order_rank,
order_rank2, date_of_order, time_of_order]
Index: []

[0 rows x 37 columns]
```

[46]:

```python
#check for non-null order_tax.
rawData[rawData.order_tax.notnull()]
```

[46]:

```
        merchant_name                                            user_id  \
17               Uber  b870e9c8674a6adb4871b0d3d98d8f2b170b458958172f…
20               Uber  ff9f60463b4677e745a0f3953bc63446c41d6895eee2e0…
22               Uber  ff9f60463b4677e745a0f3953bc63446c41d6895eee2e0…
26               Uber  ce59f0fecac89f2bd9dc1ea891ba12218bfe316e4d1a3f…
50               Uber  ba1dcbf325454d2d1be9dd49c1dbb11d651a6b4a32bf95…
…               …                                                …
1876376          Lyft  bb6b430dc1a9aec71a609a947bd38c23af4132559d9bbe…
1876377          Lyft  4c3ae6e60bee8dba46426a2c8f241d8d57d2a635c835ba…
1876378          Lyft  4d714b5bf2b1b2c0711712b42856a7e8726092cfc98c26…
1876395          Lyft  b73ef1b7eda98285b6854b19c9d64aed54597b6b7332fb…
1877059          Lyft  bdaeb8a007ac5004b9e439e896fa380828ad0236ade1aa…

                                 order_number               order_time  \
17       b62d01ea-2d55-414c-98bd-036e58e1d935  2019-10-12 00:00:00.000
20       75df2365-a71e-447a-bed4-0d9ab87e490d  2018-03-08 00:00:00.000
22       eeefb0f8-9228-4e7f-8543-e2043ed0462e  2019-09-02 00:00:00.000
26       77b157d3-80f0-4361-a700-7d70cfbdf4ef  2019-10-09 00:00:00.000
50       7c7a40ff-640d-4bbc-9c83-a5df24724fe3  2017-06-09 00:00:00.000
…                                       …                      …
1876376                   1043318510370527388  2017-09-12 08:51:00.000
1876377                   1273066725668988002  2019-05-24 11:53:00.000
1876378                   1047598251241535332  2017-09-23 21:49:00.000
1876395                   1145589905645970196  2018-06-14 23:22:00.000
1877059                   1309397943170611458                      NaN

                      email_time              insert_time  \
17       2019-10-13 03:12:25.000  2020-02-06 13:28:04.000
20       2018-03-08 12:41:23.000  2018-11-07 02:43:53.000
22       2019-09-03 03:49:19.000  2019-09-03 04:04:34.000
26       2019-10-09 21:30:52.000  2020-02-07 06:47:25.000
50       2017-06-09 06:35:50.000  2020-02-08 04:36:59.000
…                          …                      …
1876376  2017-09-12 23:46:01.000  2017-09-12 23:46:22.000
1876377  2019-05-25 17:06:30.000  2019-05-25 17:21:26.000
1876378  2017-09-25 01:58:50.000  2017-09-25 01:59:28.000
1876395  2018-06-15 21:38:35.000  2018-06-15 21:56:54.000
```

```
1877059  2019-08-30 14:16:51.000  2019-08-30 15:20:08.000

                     update_time  order_total_amount  order_points  \
17       2020-02-06 13:28:04.000               15.37           NaN
20       2020-02-06 17:48:29.000               40.41           NaN
22       2020-02-06 18:28:40.000               14.75           NaN
26       2020-02-07 06:47:26.000               24.56           NaN
50       2020-02-08 04:36:59.000                6.89           NaN
…                          …                 …           …
1876376  2019-07-20 15:53:03.000                4.16           NaN
1876377  2019-07-20 02:10:53.000                4.29           NaN
1876378  2019-07-22 01:27:25.000                4.12           NaN
1876395  2019-07-21 15:25:03.000               18.61           NaN
1877059  2019-08-30 15:20:08.000               43.44           NaN

         order_shipping  …            delivery_date  \
17                  NaN  …  2020-02-07 00:00:00.000
20                  NaN  …  2020-02-07 00:00:00.000
22                  NaN  …  2020-02-07 00:00:00.000
26                  NaN  …  2020-02-07 00:00:00.000
50                  NaN  …  2020-02-08 00:00:00.000
…                  …  …                      …
1876376             NaN  …  2020-02-04 00:00:00.000
1876377             NaN  …  2020-02-04 00:00:00.000
1876378             NaN  …  2020-02-04 00:00:00.000
1876395             NaN  …  2020-02-04 00:00:00.000
1877059             NaN  …  2020-02-04 00:00:00.000

         start_source_folder_date  end_source_folder_date file_id  \
17                     2020-02-07              2020-02-07     NaN
20                     2020-02-07              2020-02-07     NaN
22                     2020-02-07              2020-02-07     NaN
26                     2020-02-07              2020-02-07     NaN
50                     2020-02-08              2020-02-08     NaN
…                           …                     …     …
1876376                2020-02-05              2020-02-05     NaN
1876377                2020-02-05              2020-02-05     NaN
1876378                2020-02-05              2020-02-05     NaN
1876395                2020-02-05              2020-02-05     NaN
1877059                2020-02-05              2020-02-05     NaN

              source_dttimestamp              dttimestamp  order_rank  \
17       2020-02-08 01:13:23.440  2020-02-08 01:42:23.217           1
20       2020-02-08 01:13:23.440  2020-02-08 01:42:23.217           1
22       2020-02-08 01:13:23.440  2020-02-08 01:42:23.217           1
26       2020-02-08 01:13:23.440  2020-02-08 01:42:23.217           1
50       2020-02-09 01:13:22.257  2020-02-09 01:23:23.383           1
```

```
…                          …                      …         …
1876376  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1876377  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1876378  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1876395  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877059  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1

         order_rank2 date_of_order time_of_order
17                 1   2019-10-12        0:00:00
20                 1   2018-03-08        0:00:00
22                 1   2019-09-02        0:00:00
26                 1   2019-10-09        0:00:00
50                 1   2017-06-09        0:00:00
…              …           …           …
1876376            1   2017-09-12        8:51:00
1876377            1   2019-05-24        1:53:00
1876378            1   2017-09-23        1:49:00
1876395            1   2018-06-14        3:22:00
1877059            1   2019-08-30        4:16:51

[274384 rows x 37 columns]
```

[47]:
```
#check for negative order_tax.
rawData[rawData.order_tax.notnull()]['order_tax'].apply(lambda x: x < 0).
 ↪value_counts()
```

[47]:
```
False    274384
Name: order_tax, dtype: int64
```

[48]:
```
#check for non-null order_subtotal.
rawData[rawData.order_subtotal.notnull()]
```

[48]:
```
        merchant_name                                            user_id  \
0                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
1                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
2                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
3                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
4                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
…               …                                                …
1877022          Uber  43e6b22f5ba9ee3ecb69a010fe4953d259a49c48d6a586…
1877023          Uber  9e6c5389d4a77b5b4759327123cf15f8e5b88aed90c215…
1877042          Lyft  d53e69b902a46edb3bbd7ae5c123ca6640e3d98016f9b0…
1877043          Lyft  2cf747a6f2457994d7edf8dc78e89f75e10cb623eb534e…
1877044          Lyft  64769f38f439e5fb09f98269c387ad55e3c20036a5fe4a…

                                 order_number               order_time  \
0                         1183621593749890808  2018-09-25 09:00:00.000
```

```
1                        1155413807788118982  2018-07-11 08:38:00.000
2                        1311366563734805982  2019-09-04 14:54:00.000
3                        1202358710117776328  2018-11-14 19:56:00.000
4                        1234240944632141798  2019-02-08 17:52:00.000
…                                          …                        …
1877022  09b48aaf-c3c3-44e1-95d8-c4a54b79765b  2017-12-02 00:00:00.000
1877023  0476d11c-fcbf-46d2-88a5-b94caf00bf73  2019-04-27 00:00:00.000
1877042                  1016508429902029614  2017-07-02 02:57:00.000
1877043                  1336921758364990548  2019-11-12 09:45:00.000
1877044                  1034207524058760544  2017-08-18 19:38:00.000

                      email_time              insert_time  \
0        2018-09-25 17:02:20.000  2020-02-17 15:41:57.000
1        2018-07-12 14:11:27.000  2020-02-17 12:28:35.000
2        2019-09-04 21:22:29.000  2020-02-18 04:51:14.000
3        2018-11-15 03:28:19.000  2020-02-17 16:40:01.000
4        2019-02-09 04:42:52.000  2020-02-17 20:18:30.000
…                              …                        …
1877022  2017-12-03 02:52:05.000  2019-08-01 20:41:31.000
1877023  2019-04-28 03:27:31.000  2019-04-28 04:16:33.000
1877042  2017-07-02 07:21:18.000  2017-07-03 08:41:39.000
1877043  2019-11-13 18:34:10.000  2019-11-13 19:12:43.000
1877044  2017-08-19 01:35:15.000  2017-08-19 01:36:56.000

                     update_time  order_total_amount  order_points  \
0        2020-02-17 15:41:57.000                9.65           NaN
1        2020-02-17 12:28:35.000                8.65           NaN
2        2020-02-18 04:51:14.000               11.39           NaN
3        2020-02-17 16:40:01.000                6.32           NaN
4        2020-02-17 20:18:30.000               10.90           NaN
…                              …                  …            …
1877022  2019-09-04 20:43:23.000                8.76           NaN
1877023  2019-07-23 03:21:46.000                8.18           NaN
1877042  2019-07-21 23:05:56.000               11.74           NaN
1877043  2019-11-13 19:12:43.000               23.86           NaN
1877044  2019-07-22 02:02:08.000                4.09           NaN

         order_shipping  …            delivery_date  \
0                   NaN  …  2020-02-18 00:00:00.000
1                   NaN  …  2020-02-18 00:00:00.000
2                   NaN  …  2020-02-18 00:00:00.000
3                   NaN  …  2020-02-18 00:00:00.000
4                   NaN  …  2020-02-18 00:00:00.000
…                     …  …                        …
1877022             NaN  …  2020-02-04 00:00:00.000
1877023             NaN  …  2020-02-04 00:00:00.000
1877042             NaN  …  2020-02-04 00:00:00.000
```

```
1877043             NaN  …  2020-02-04 00:00:00.000
1877044             NaN  …  2020-02-04 00:00:00.000

         start_source_folder_date  end_source_folder_date file_id  \
0                      2020-02-18              2020-02-18     NaN
1                      2020-02-18              2020-02-18     NaN
2                      2020-02-18              2020-02-18     NaN
3                      2020-02-18              2020-02-18     NaN
4                      2020-02-18              2020-02-18     NaN
…                           …                     …     …
1877022                2020-02-05              2020-02-05     NaN
1877023                2020-02-05              2020-02-05     NaN
1877042                2020-02-05              2020-02-05     NaN
1877043                2020-02-05              2020-02-05     NaN
1877044                2020-02-05              2020-02-05     NaN

              source_dttimestamp              dttimestamp  order_rank   \
0        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
1        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
2        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
3        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
4        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
…                          …                      …         …
1877022  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877023  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877042  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877043  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877044  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1

         order_rank2 date_of_order time_of_order
0                  1   2018-09-25        9:00:00
1                  1   2018-07-11        8:38:00
2                  1   2019-09-04        4:54:00
3                  1   2018-11-14        9:56:00
4                  1   2019-02-08        7:52:00
…              …           …           …
1877022            1   2017-12-02        0:00:00
1877023            1   2019-04-27        0:00:00
1877042            1   2017-07-02        2:57:00
1877043            1   2019-11-12        9:45:00
1877044            1   2017-08-18        9:38:00

[1766122 rows x 37 columns]
```

[49]:

```
#check for negative order_subtotals.
rawData[rawData.order_subtotal.notnull()]['order_subtotal'].apply(lambda x: x < 0).value_counts()
```

[49]:

```
False    1766122
Name: order_subtotal, dtype: int64
```

[50]:

```
#check for non-null order_total_qty.
rawData[rawData.order_total_qty.notnull()]
```

[50]:

```
        merchant_name                                             user_id  \
0                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
1                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
2                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
3                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
4                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
…               …                                                …
1877053          Uber  263f69452a0c2d46968effe172ba5f4f20785dc0640f85…
1877054          Uber  415b8e1517a45479a2cbabbad5afc6a6cd5f12bc2aa8dd…
1877055          Uber  d3862ea64b55b9107d80ed5ff4eaa8f400ec7c2574562e…
1877056          Uber  7de81565b4aa4ca57f1e30553c163783353006f4039ca2…
1877059          Lyft  bdaeb8a007ac5004b9e439e896fa380828ad0236ade1aa…

                                 order_number               order_time  \
0                         1183621593749890808  2018-09-25 09:00:00.000
1                         1155413807788118982  2018-07-11 08:38:00.000
2                         1311366563734805982  2019-09-04 14:54:00.000
3                         1202358710117776328  2018-11-14 19:56:00.000
4                         1234240944632141798  2019-02-08 17:52:00.000
…                                       …                      …
1877053  6e11bab3-3b10-40c2-970e-66bdf17effdb  2017-08-15 00:00:00.000
1877054  7cc351eb-ab18-4b32-b474-c5cc8d233252  2017-07-26 00:00:00.000
1877055  1d15fd3c-487f-4847-806c-63b5723e2bff                      NaN
1877056  664a6bea-6ab3-40e1-97fc-d309e4d46a35                      NaN
1877059                   1309397943170611458                      NaN

                      email_time              insert_time  \
0        2018-09-25 17:02:20.000  2020-02-17 15:41:57.000
1        2018-07-12 14:11:27.000  2020-02-17 12:28:35.000
2        2019-09-04 21:22:29.000  2020-02-18 04:51:14.000
3        2018-11-15 03:28:19.000  2020-02-17 16:40:01.000
4        2019-02-09 04:42:52.000  2020-02-17 20:18:30.000
…                          …                      …
1877053  2017-08-15 23:24:49.000  2017-08-15 23:25:14.000
1877054  2017-07-27 03:50:16.000  2017-07-27 03:50:35.000
1877055  2018-02-07 01:55:34.000  2020-02-01 12:12:08.000
1877056  2018-06-08 04:34:48.000  2018-06-08 04:37:19.000
1877059  2019-08-30 14:16:51.000  2019-08-30 15:20:08.000

                     update_time  order_total_amount  order_points  \
0        2020-02-17 15:41:57.000                9.65           NaN
```

```
1        2020-02-17 12:28:35.000                8.65           NaN
2        2020-02-18 04:51:14.000               11.39           NaN
3        2020-02-17 16:40:01.000                6.32           NaN
4        2020-02-17 20:18:30.000               10.90           NaN
…                          …                 …           …
1877053  2018-09-21 05:16:10.000                5.00           NaN
1877054  2018-09-21 06:25:06.000                5.00           NaN
1877055  2020-02-01 12:12:32.000                0.00           NaN
1877056  2019-03-18 17:56:25.000                0.00           NaN
1877059  2019-08-30 15:20:08.000               43.44           NaN

         order_shipping  …            delivery_date  \
0                   NaN  …  2020-02-18 00:00:00.000
1                   NaN  …  2020-02-18 00:00:00.000
2                   NaN  …  2020-02-18 00:00:00.000
3                   NaN  …  2020-02-18 00:00:00.000
4                   NaN  …  2020-02-18 00:00:00.000
…                 …  …                      …
1877053             NaN  …  2020-02-04 00:00:00.000
1877054             NaN  …  2020-02-04 00:00:00.000
1877055             NaN  …  2020-02-04 00:00:00.000
1877056             NaN  …  2020-02-04 00:00:00.000
1877059             NaN  …  2020-02-04 00:00:00.000

         start_source_folder_date  end_source_folder_date file_id  \
0                      2020-02-18              2020-02-18     NaN
1                      2020-02-18              2020-02-18     NaN
2                      2020-02-18              2020-02-18     NaN
3                      2020-02-18              2020-02-18     NaN
4                      2020-02-18              2020-02-18     NaN
…                           …                     …     …
1877053                2020-02-05              2020-02-05     NaN
1877054                2020-02-05              2020-02-05     NaN
1877055                2020-02-05              2020-02-05     NaN
1877056                2020-02-05              2020-02-05     NaN
1877059                2020-02-05              2020-02-05     NaN

              source_dttimestamp              dttimestamp  order_rank  \
0        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
1        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
2        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
3        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
4        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
…                          …                      …         …
1877053  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877054  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877055  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
```

```
1877056  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877059  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1

         order_rank2 date_of_order time_of_order
0                  1    2018-09-25       9:00:00
1                  1    2018-07-11       8:38:00
2                  1    2019-09-04       4:54:00
3                  1    2018-11-14       9:56:00
4                  1    2019-02-08       7:52:00
…              …           …           …
1877053            1    2017-08-15       0:00:00
1877054            1    2017-07-26       0:00:00
1877055            1    2018-02-07       1:55:34
1877056            1    2018-06-08       4:34:48
1877059            1    2019-08-30       4:16:51

[1835545 rows x 37 columns]
```

[51]:
```
#check for negative order_total_qty
rawData[rawData.order_total_qty.notnull()]['order_total_qty'].apply(lambda x: x␣
 ↪< 0).value_counts()
```

[51]:
```
False    1835545
Name: order_total_qty, dtype: int64
```

[52]:
```
#check distribution for order_total_qty.
rawData.order_total_qty.value_counts()
```

[52]:
```
1.0    1835545
Name: order_total_qty, dtype: int64
```

[53]:
```
#check for non-null item_quantity.
rawData[rawData.item_quantity.notnull()]
```

[53]:
```
        merchant_name                                           user_id  \
0                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
1                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
2                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
3                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
4                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
…               …                                                …
1877053          Uber  263f69452a0c2d46968effe172ba5f4f20785dc0640f85…
1877054          Uber  415b8e1517a45479a2cbabbad5afc6a6cd5f12bc2aa8dd…
1877055          Uber  d3862ea64b55b9107d80ed5ff4eaa8f400ec7c2574562e…
1877056          Uber  7de81565b4aa4ca57f1e30553c163783353006f4039ca2…
1877059          Lyft  bdaeb8a007ac5004b9e439e896fa380828ad0236ade1aa…
```

```
                                 order_number               order_time  \
0                         1183621593749890808  2018-09-25 09:00:00.000
1                         1155413807788118982  2018-07-11 08:38:00.000
2                         1311366563734805982  2019-09-04 14:54:00.000
3                         1202358710117776328  2018-11-14 19:56:00.000
4                         1234240944632141798  2019-02-08 17:52:00.000
…                                       …                       …
1877053  6e11bab3-3b10-40c2-970e-66bdf17effdb  2017-08-15 00:00:00.000
1877054  7cc351eb-ab18-4b32-b474-c5cc8d233252  2017-07-26 00:00:00.000
1877055  1d15fd3c-487f-4847-806c-63b5723e2bff                      NaN
1877056  664a6bea-6ab3-40e1-97fc-d309e4d46a35                      NaN
1877059                   1309397943170611458                      NaN

                      email_time              insert_time  \
0        2018-09-25 17:02:20.000  2020-02-17 15:41:57.000
1        2018-07-12 14:11:27.000  2020-02-17 12:28:35.000
2        2019-09-04 21:22:29.000  2020-02-18 04:51:14.000
3        2018-11-15 03:28:19.000  2020-02-17 16:40:01.000
4        2019-02-09 04:42:52.000  2020-02-17 20:18:30.000
…                          …                      …
1877053  2017-08-15 23:24:49.000  2017-08-15 23:25:14.000
1877054  2017-07-27 03:50:16.000  2017-07-27 03:50:35.000
1877055  2018-02-07 01:55:34.000  2020-02-01 12:12:08.000
1877056  2018-06-08 04:34:48.000  2018-06-08 04:37:19.000
1877059  2019-08-30 14:16:51.000  2019-08-30 15:20:08.000

                     update_time  order_total_amount  order_points  \
0        2020-02-17 15:41:57.000                9.65           NaN
1        2020-02-17 12:28:35.000                8.65           NaN
2        2020-02-18 04:51:14.000               11.39           NaN
3        2020-02-17 16:40:01.000                6.32           NaN
4        2020-02-17 20:18:30.000               10.90           NaN
…                          …                 …           …
1877053  2018-09-21 05:16:10.000                5.00           NaN
1877054  2018-09-21 06:25:06.000                5.00           NaN
1877055  2020-02-01 12:12:32.000                0.00           NaN
1877056  2019-03-18 17:56:25.000                0.00           NaN
1877059  2019-08-30 15:20:08.000               43.44           NaN

         order_shipping  …            delivery_date  \
0                   NaN  …  2020-02-18 00:00:00.000
1                   NaN  …  2020-02-18 00:00:00.000
2                   NaN  …  2020-02-18 00:00:00.000
3                   NaN  …  2020-02-18 00:00:00.000
4                   NaN  …  2020-02-18 00:00:00.000
…                 …  …                      …
1877053             NaN  …  2020-02-04 00:00:00.000
```

```
1877054             NaN  …  2020-02-04 00:00:00.000
1877055             NaN  …  2020-02-04 00:00:00.000
1877056             NaN  …  2020-02-04 00:00:00.000
1877059             NaN  …  2020-02-04 00:00:00.000

         start_source_folder_date  end_source_folder_date file_id  \
0                      2020-02-18              2020-02-18     NaN
1                      2020-02-18              2020-02-18     NaN
2                      2020-02-18              2020-02-18     NaN
3                      2020-02-18              2020-02-18     NaN
4                      2020-02-18              2020-02-18     NaN
…                           …                     …     …
1877053                2020-02-05              2020-02-05     NaN
1877054                2020-02-05              2020-02-05     NaN
1877055                2020-02-05              2020-02-05     NaN
1877056                2020-02-05              2020-02-05     NaN
1877059                2020-02-05              2020-02-05     NaN

              source_dttimestamp              dttimestamp  order_rank  \
0        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
1        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
2        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
3        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
4        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
…                          …                      …         …
1877053  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877054  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877055  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877056  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877059  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1

         order_rank2 date_of_order time_of_order
0                  1   2018-09-25        9:00:00
1                  1   2018-07-11        8:38:00
2                  1   2019-09-04        4:54:00
3                  1   2018-11-14        9:56:00
4                  1   2019-02-08        7:52:00
…              …           …           …
1877053            1   2017-08-15        0:00:00
1877054            1   2017-07-26        0:00:00
1877055            1   2018-02-07        1:55:34
1877056            1   2018-06-08        4:34:48
1877059            1   2019-08-30        4:16:51

[1835545 rows x 37 columns]
```

[54]:
```
#check for negative item_quantity.
rawData[rawData.item_quantity.notnull()]['item_quantity'].apply(lambda x: x <␣
 ↪0).value_counts()
```

[54]:
```
False    1835545
Name: item_quantity, dtype: int64
```

[55]:
```
#check item_quantity distribution.
rawData.item_quantity.value_counts()
```

[55]:
```
1.0    1835545
Name: item_quantity, dtype: int64
```

[56]:
```
#check for non-null item_price
rawData[rawData.item_price.notnull()]
```

[56]:
```
        merchant_name                                            user_id  \
0                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
1                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
2                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
3                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
4                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
…                 …                                                …
1877053          Uber  263f69452a0c2d46968effe172ba5f4f20785dc0640f85…
1877054          Uber  415b8e1517a45479a2cbabbad5afc6a6cd5f12bc2aa8dd…
1877055          Uber  d3862ea64b55b9107d80ed5ff4eaa8f400ec7c2574562e…
1877056          Uber  7de81565b4aa4ca57f1e30553c163783353006f4039ca2…
1877059          Lyft  bdaeb8a007ac5004b9e439e896fa380828ad0236ade1aa…

                                 order_number               order_time  \
0                         1183621593749890808  2018-09-25 09:00:00.000
1                         1155413807788118982  2018-07-11 08:38:00.000
2                         1311366563734805982  2019-09-04 14:54:00.000
3                         1202358710117776328  2018-11-14 19:56:00.000
4                         1234240944632141798  2019-02-08 17:52:00.000
…                                           …                        …
1877053  6e11bab3-3b10-40c2-970e-66bdf17effdb  2017-08-15 00:00:00.000
1877054  7cc351eb-ab18-4b32-b474-c5cc8d233252  2017-07-26 00:00:00.000
1877055  1d15fd3c-487f-4847-806c-63b5723e2bff                      NaN
1877056  664a6bea-6ab3-40e1-97fc-d309e4d46a35                      NaN
1877059                   1309397943170611458                      NaN

                      email_time              insert_time  \
0        2018-09-25 17:02:20.000  2020-02-17 15:41:57.000
1        2018-07-12 14:11:27.000  2020-02-17 12:28:35.000
2        2019-09-04 21:22:29.000  2020-02-18 04:51:14.000
3        2018-11-15 03:28:19.000  2020-02-17 16:40:01.000
```

```
4        2019-02-09 04:42:52.000  2020-02-17 20:18:30.000
…                           …                      …
1877053  2017-08-15 23:24:49.000  2017-08-15 23:25:14.000
1877054  2017-07-27 03:50:16.000  2017-07-27 03:50:35.000
1877055  2018-02-07 01:55:34.000  2020-02-01 12:12:08.000
1877056  2018-06-08 04:34:48.000  2018-06-08 04:37:19.000
1877059  2019-08-30 14:16:51.000  2019-08-30 15:20:08.000

                     update_time  order_total_amount  order_points  \
0        2020-02-17 15:41:57.000                9.65           NaN
1        2020-02-17 12:28:35.000                8.65           NaN
2        2020-02-18 04:51:14.000               11.39           NaN
3        2020-02-17 16:40:01.000                6.32           NaN
4        2020-02-17 20:18:30.000               10.90           NaN
…                           …                …            …
1877053  2018-09-21 05:16:10.000                5.00           NaN
1877054  2018-09-21 06:25:06.000                5.00           NaN
1877055  2020-02-01 12:12:32.000                0.00           NaN
1877056  2019-03-18 17:56:25.000                0.00           NaN
1877059  2019-08-30 15:20:08.000               43.44           NaN

         order_shipping  …            delivery_date  \
0                   NaN  …  2020-02-18 00:00:00.000
1                   NaN  …  2020-02-18 00:00:00.000
2                   NaN  …  2020-02-18 00:00:00.000
3                   NaN  …  2020-02-18 00:00:00.000
4                   NaN  …  2020-02-18 00:00:00.000
…                 …  …                      …
1877053             NaN  …  2020-02-04 00:00:00.000
1877054             NaN  …  2020-02-04 00:00:00.000
1877055             NaN  …  2020-02-04 00:00:00.000
1877056             NaN  …  2020-02-04 00:00:00.000
1877059             NaN  …  2020-02-04 00:00:00.000

         start_source_folder_date  end_source_folder_date file_id  \
0                      2020-02-18              2020-02-18     NaN
1                      2020-02-18              2020-02-18     NaN
2                      2020-02-18              2020-02-18     NaN
3                      2020-02-18              2020-02-18     NaN
4                      2020-02-18              2020-02-18     NaN
…                           …                     …     …
1877053                2020-02-05              2020-02-05     NaN
1877054                2020-02-05              2020-02-05     NaN
1877055                2020-02-05              2020-02-05     NaN
1877056                2020-02-05              2020-02-05     NaN
1877059                2020-02-05              2020-02-05     NaN
```

```
              source_dttimestamp             dttimestamp  order_rank  \
0        2020-02-19 01:03:42.093 2020-02-19 03:00:52.077           1
1        2020-02-19 01:03:42.093 2020-02-19 03:00:52.077           1
2        2020-02-19 01:03:42.093 2020-02-19 03:00:52.077           1
3        2020-02-19 01:03:42.093 2020-02-19 03:00:52.077           1
4        2020-02-19 01:03:42.093 2020-02-19 03:00:52.077           1
…                            …                       …         …
1877053  2020-02-06 04:58:04.987 2020-02-06 21:58:57.453           1
1877054  2020-02-06 04:58:04.987 2020-02-06 21:58:57.453           1
1877055  2020-02-06 04:58:04.987 2020-02-06 21:58:57.453           1
1877056  2020-02-06 04:58:04.987 2020-02-06 21:58:57.453           1
1877059  2020-02-06 04:58:04.987 2020-02-06 21:58:57.453           1

         order_rank2 date_of_order time_of_order
0                  1    2018-09-25       9:00:00
1                  1    2018-07-11       8:38:00
2                  1    2019-09-04       4:54:00
3                  1    2018-11-14       9:56:00
4                  1    2019-02-08       7:52:00
…              …           …           …
1877053            1    2017-08-15       0:00:00
1877054            1    2017-07-26       0:00:00
1877055            1    2018-02-07       1:55:34
1877056            1    2018-06-08       4:34:48
1877059            1    2019-08-30       4:16:51

[1835545 rows x 37 columns]
```

[57]:

```
#check for negative item_price
rawData[rawData.item_price.notnull()]['item_price'].apply(lambda x: x < 0).
 ↪value_counts()
```

[57]:

```
False    1835545
Name: item_price, dtype: int64
```

[58]:

```
#check for distribution of item_price.
rawData.item_price.value_counts()
```

[58]:

```
5.00      40792
5.80       8268
7.19       7969
4.00       7081
7.00       6499
          …
131.09        1
96.29         1
237.19        1
```

```
171.37        1
127.20        1
Name: item_price, Length: 17483, dtype: int64
```

[59]:

```
#check for non-null order_discount.
rawData[rawData.order_discount.notnull()]
```

[59]:

```
        merchant_name                                            user_id  \
3                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
9                Lyft  3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
21               Uber  d34e4f4290e3a873c42269b218a6f334a546ac931daf88…
22               Uber  ff9f60463b4677e745a0f3953bc63446c41d6895eee2e0…
33               Uber  23bd1ab2d2da7dadb08d3132c60c5739f199bc1d470642…
…               …                                                …
1876653          Lyft  5aaf780a857ec06a8eddfbef3accd8e83446a7a671fbf5…
1876654          Lyft  5c43f9277d8385055cf459508b511c695b1959062e7b59…
1877042          Lyft  d53e69b902a46edb3bbd7ae5c123ca6640e3d98016f9b0…
1877043          Lyft  2cf747a6f2457994d7edf8dc78e89f75e10cb623eb534e…
1877044          Lyft  64769f38f439e5fb09f98269c387ad55e3c20036a5fe4a…

                                 order_number               order_time  \
3                         1202358710117776328  2018-11-14 19:56:00.000
9                         1167033116179181344  2018-08-11 16:05:00.000
21       d4596113-443b-425b-b5dd-a016ba3292d6  2018-08-11 00:00:00.000
22       eeefb0f8-9228-4e7f-8543-e2043ed0462e  2019-09-02 00:00:00.000
33       2e751818-b5f7-43f4-966b-bc2101f2d512  2018-01-25 00:00:00.000
…                                       …                      …
1876653                   1016686096294065404  2017-07-02 11:23:00.000
1876654                   1155149323623239730  2018-07-10 17:31:00.000
1877042                   1016508429902029614  2017-07-02 02:57:00.000
1877043                   1336921758364990548  2019-11-12 09:45:00.000
1877044                   1034207524058760544  2017-08-18 19:38:00.000

                      email_time              insert_time  \
3        2018-11-15 03:28:19.000  2020-02-17 16:40:01.000
9        2018-08-11 22:36:23.000  2020-02-17 14:01:57.000
21       2018-08-12 00:41:09.000  2018-08-12 00:59:36.000
22       2019-09-03 03:49:19.000  2019-09-03 04:04:34.000
33       2018-01-26 12:51:31.000  2020-02-07 06:55:24.000
…                          …                      …
1876653  2017-07-03 05:31:26.000  2018-03-24 15:03:07.000
1876654  2018-07-11 04:01:31.000  2018-07-11 04:13:03.000
1877042  2017-07-02 07:21:18.000  2017-07-03 08:41:39.000
1877043  2019-11-13 18:34:10.000  2019-11-13 19:12:43.000
1877044  2017-08-19 01:35:15.000  2017-08-19 01:36:56.000

                     update_time  order_total_amount  order_points  \
3        2020-02-17 16:40:01.000                6.32           NaN
9        2020-02-17 14:01:57.000                6.64           NaN
21       2020-02-06 13:48:02.000                9.19           NaN
22       2020-02-06 18:28:40.000               14.75           NaN
33       2020-02-07 06:55:36.000               34.70           NaN
…                          …                 …           …
1876653  2019-07-20 12:18:53.000                4.03           NaN
1876654  2019-07-20 11:37:10.000               16.53           NaN
1877042  2019-07-21 23:05:56.000               11.74           NaN
1877043  2019-11-13 19:12:43.000               23.86           NaN
1877044  2019-07-22 02:02:08.000                4.09           NaN

         order_shipping  …            delivery_date  \
3                   NaN  …  2020-02-18 00:00:00.000
9                   NaN  …  2020-02-18 00:00:00.000
21                  NaN  …  2020-02-07 00:00:00.000
22                  NaN  …  2020-02-07 00:00:00.000
33                  NaN  …  2020-02-07 00:00:00.000
…                 …  …                      …
1876653             NaN  …  2020-02-04 00:00:00.000
1876654             NaN  …  2020-02-04 00:00:00.000
1877042             NaN  …  2020-02-04 00:00:00.000
1877043             NaN  …  2020-02-04 00:00:00.000
1877044             NaN  …  2020-02-04 00:00:00.000

         start_source_folder_date  end_source_folder_date file_id  \
3                      2020-02-18              2020-02-18     NaN
9                      2020-02-18              2020-02-18     NaN
21                     2020-02-07              2020-02-07     NaN
22                     2020-02-07              2020-02-07     NaN
33                     2020-02-07              2020-02-07     NaN
…                           …                     …     …
1876653                2020-02-05              2020-02-05     NaN
1876654                2020-02-05              2020-02-05     NaN
1877042                2020-02-05              2020-02-05     NaN
1877043                2020-02-05              2020-02-05     NaN
1877044                2020-02-05              2020-02-05     NaN

              source_dttimestamp              dttimestamp  order_rank  \
3        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
9        2020-02-19 01:03:42.093  2020-02-19 03:00:52.077           1
21       2020-02-08 01:13:23.440  2020-02-08 01:42:23.217           1
22       2020-02-08 01:13:23.440  2020-02-08 01:42:23.217           1
33       2020-02-08 01:13:23.440  2020-02-08 01:42:23.217           1
…                          …                      …         …
1876653  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1876654  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
```

```
1877042  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877043  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1
1877044  2020-02-06 04:58:04.987  2020-02-06 21:58:57.453           1

         order_rank2 date_of_order time_of_order
3                  1    2018-11-14       9:56:00
9                  1    2018-08-11       6:05:00
21                 1    2018-08-11       0:00:00
22                 1    2019-09-02       0:00:00
33                 1    2018-01-25       0:00:00
…                …            …             …
1876653            1    2017-07-02       1:23:00
1876654            1    2018-07-10       7:31:00
1877042            1    2017-07-02       2:57:00
1877043            1    2019-11-12       9:45:00
1877044            1    2017-08-18       9:38:00

[289627 rows x 37 columns]
```

[60]:

```
#check for negative order_discount.
rawData[rawData.order_discount.notnull()]['order_discount'].apply(lambda x: x <␣
 ↪0).value_counts()
```

[60]:

```
False    289627
Name: order_discount, dtype: int64
```

[61]:

```
#Check order_discount distribution.
rawData.order_discount.value_counts()
```

[61]:

```
6.00      19197
2.00      12210
4.00       6931
10.00      5548
3.00       4093
          …
90.32         1
25.83         1
142.28        1
43.59         1
28.53         1
Name: order_discount, Length: 3679, dtype: int64
```

Well, I am pleasantly surprised that the numeric integrity checks look pretty solid! Here is a column-by-column summary for the numerical transaction fields:

- **order_total_amount:** This field is never negative and appears to be mostly correct as a dollar amount, with only a meager 282 rows that are NaN. Action for this field is to simply remove those rows that have a NaN for their order_total_amount (these are likely bad rows).

- **order_points, order_shipping:** These are ALWAYS NaN and as such, we can dismiss them. Be gone, children of the void!
- **order_tax:** This field is never negative, but does have a value of NaN most of the time. The fact that it does have non-zero and well-defined numbers the rest of the time means that we will want to include it, and set the cases where it is NaN to be 0.
- **order_subtotal:** This field is never negative, but does have some NaNs. For these cases, we should set the subtotal to 0 and see what the order_total_amount gives us. We may also be able to manually calculate it based on the other values (item_quantity and item_price).
- **order_total_qty, item_quantity:** These fields are always present, and always equal to the same value: 1. As a result, we can safely ignore them, and they greatly simplify the formulae I wrote in the earlier blocks.
- **item_price:** This is never negative nor is it NaN. We can use this for calculation checks in the next section.
- **order_discount:** This field is never negative, and has a numeric value some of the time. Other times, it is NaN. For these cases, we will handle it like order_tax and set those equal to 0.

In the below blocks, we implement the changes I mentioned above.

```
[62]: rawData = rawData[rawData.order_total_amount.notnull()]
      rawData.loc[rawData.order_tax.isnull(),"order_tax"] = 0
      rawData.loc[rawData.order_discount.isnull(),"order_discount"] = 0
      rawData.loc[rawData.order_subtotal.isnull(),"order_subtotal"] = 0
```

Now we want to implement our formulae to do some integrity checks. The formulae that we were considering before were:

order_total_amount = order_subtotal + order_tax + order_shipping - order_discount
item_price * item_quantity = order_subtotal

Note that the first formula simplifies to:

order_total_amount = order_subtotal + order_tax - order_discount

And the second simplifies to:

item_price = order_subtotal

Let's check these out.

```
[63]: #Check formula 1
      (rawData.order_total_amount == rawData.order_subtotal + rawData.order_tax -␣
       ↪rawData.order_discount).value_counts()
```

```
[63]: False    1234460
      True      600803
      dtype: int64
```

```
[64]: #Check formula 2
      (rawData.item_price == rawData.order_subtotal).value_counts()
```

```
[64]: True     1768104
      False      67159
      dtype: int64
```

As we can see, formula 2 looks pretty good; most of the rows satisfy it. As fara as formula 1 goes, it looks as though something is definitely missing. One possibility is that the discount should not be subtracted. Let's look at this possibility:

```
[65]: #Modified formula 1
      (rawData.order_total_amount == rawData.order_subtotal + rawData.order_tax).
       ↪value_counts()
```

```
[65]: False    1291794
      True      543469
      dtype: int64
```

This does even worse! Let's try one more:

```
[66]: #Modified formula 1, one more time
      (rawData.order_total_amount == rawData.order_subtotal).value_counts()
```

```
[66]: False    1205975
      True      629288
      dtype: int64
```

Marginally better than the first, but still nothing to write home about.

Given the uncertain nature of these checks and the huge number of rows effected, I would say that we should not implement these integrity checks without further consultation from the vendor as far as removing rows. However, a follow-up to this case study should inquire as to why these formula checks are not passing. The question will be noted; for now, let's move on.

2.6 PRODUCT FIELDS

The next round of fields that we need to investigate are those concerened with products. These are the following:

- product_description
- product_subtitle
- product_reseller
- product_category
- SKU
- item_id

For these fields, the question is: what are their values? Are they informative? Are they redundant? Let's take a look.

[67]:
```
#What does this field show us?
rawData["product_description"]
```

[67]:
```
0                          Lyft fare (1.63mi, 12m 23s)
1                           Lyft fare (1.73mi, 7m 55s)
2                           Lyft fare (1.54mi, 7m 33s)
3                         Shared fare (2.56mi, 14m 2s)
4                           Lyft fare (1.54mi, 7m 38s)
                              …
1877053                                            NaN
1877054                                            NaN
1877055    0.46 miles   00:02:58 Trip time   uberX Car
1877056     0.70 miles   00:04:41 Trip time   Pool Car
1877059                                            NaN
Name: product_description, Length: 1835263, dtype: object
```

As the above snippet of the results indicates, this field is **rich** with information about the trip. In particular, we have access to the following information:

- trip duration, in hours, minutes and seconds.
- trip distance, in miles
- whether it was a shared fare (or pool), a standard fare, or an Uberx, as well as sometimes what sort of vehicle was used

We need to first identify the formats for each of these (their formats appear to differ from merchant to merchant) and separate out the data that is contained in these fields with some string manipulation.

There is a lot of potential here for us to further decompose this data. We can use a battery of regex expressions to identify the various patterns and store the data in their own columns. I will demonstrate this with a few regex expressions below for both Uber and Lyft.

[68]:
```
#Let's use some code to identify the regexes that we want for Uber.
uber = rawData[rawData.merchant_name == 'Uber']
[x for x in uber.product_description.astype(str) if 'miles' in x]

#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberX Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberX VIP Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   VIP Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   Express POOL Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberXL Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   Select Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberPOOL Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   LUX Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   BLACK CAR Car

#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberX Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberX VIP Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   VIP Car
```

```
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   Express POOL Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberXL Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   Select Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberPOOL Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   LUX Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   BLACK CAR Car
```

```
[68]: ['5.27 miles   00:09:12 Trip time   uberX Car',
       '2.92 miles   00:22:05 Trip time   SUV Car',
       '2.26 miles   00:09:56 Trip time   UberX Car',
       '3.65 miles   00:12:50 Trip time   uberX Car',
       '6.82 miles   00:19:32 Trip time   uberX Car',
       '2.92 miles   00:23:57 Trip time   POOL Car',
       '3.24 miles   00:16:35 Trip time   POOL Car',
       '2.56 miles   00:08:42 Trip time   uberX Car',
       '9.29 miles   00:15:00 Trip time   uberX Car',
       '5.50 miles   00:23:23 Trip time   Pool Car',
       '0.87 miles   00:12:31 Trip time   Pool Car',
       '0.56 miles   00:05:48 Trip time   uberX Car',
       '7.82 miles   00:26:18 Trip time   uberX Car',
       '0.60 miles   00:06:11 Trip time   UberX VIP Car',
       '1.02 miles   00:03:58 Trip time   uberX Car',
       '1.13 miles   00:04:31 Trip time   uberX Car',
       '4.65 miles   00:11:01 Trip time   uberX Car',
       '4.51 miles   00:10:04 Trip time   uberX Car',
       '3.46 miles   00:14:06 Trip time   uberX Car',
       '2.70 miles   00:08:48 Trip time   uberX Car',
       '4.13 miles   00:14:25 Trip time   uberX Car',
       '4.19 miles   00:08:40 Trip time   PREMIUM Car',
       '1.25 miles   00:06:11 Trip time   uberX Car',
       '1.38 miles   00:05:10 Trip time   UberX Car',
       '1.58 miles   00:08:22 Trip time   uberX Car',
       '2.08 miles   00:10:26 Trip time   Express POOL Car',
       '3.41 miles   00:11:39 Trip time   uberX Car',
       '2.70 miles   00:09:29 Trip time   uberX Car',
       '5.96 miles   00:19:49 Trip time   uberX Car',
       '11.52 miles   00:20:28 Trip time   VIP Car',
       '2.86 miles   00:07:56 Trip time   uberX Car',
       '3.67 miles   00:12:53 Trip time   UberX Car',
       '2.98 miles   00:16:07 Trip time   uberX Car',
       '4.05 miles   00:16:17 Trip time   uberX Car',
       '1.67 miles   00:08:54 Trip time   uberX Car',
       '3.44 miles   00:08:45 Trip time   uberX Car',
       '13.31 miles   00:35:36 Trip time   UberXL Car',
       '36.07 miles   01:18:45 Trip time   uberX Car',
       '3.29 miles   00:11:28 Trip time   uberX Car',
       '4.76 miles   00:15:23 Trip time   uberX Car',
       '6.64 miles   00:15:23 Trip time   uberX Car',
       '2.59 miles   00:09:33 Trip time   UberX Car',
       '3.28 miles   00:08:18 Trip time   UberX Car',
       '9.14 miles   00:20:15 Trip time   Express Pool Car',
       '12.79 miles   00:24:23 Trip time   uberX Car',
       '0.87 miles   00:05:54 Trip time   uberX Car',
       '5.86 miles   00:11:46 Trip time   uberX Car',
       '3.50 miles   00:16:35 Trip time   uberX Car',
       '1.25 miles   00:04:27 Trip time   UberX Car',
       '4.21 miles   00:12:59 Trip time   uberX Car',
       '2.89 miles   00:09:46 Trip time   uberX Car',
       '5.59 miles   00:16:32 Trip time   UberX Car',
       '2.03 miles   00:18:50 Trip time   Express POOL Car',
       '3.91 miles   00:10:05 Trip time   uberX Car',
       '3.97 miles   00:10:21 Trip time   Select Car',
       '11.49 miles   00:38:02 Trip time   UberX Car',
       '1.70 miles   00:04:48 Trip time   uberX Car',
       '6.42 miles   00:13:11 Trip time   uberX VIP Car',
       '17.66 miles   00:39:37 Trip time   UberX Car',
       '3.90 miles   00:11:53 Trip time   uberXL Car',
       '2.48 miles   00:07:24 Trip time   uberX Car',
       '1.97 miles   00:17:49 Trip time   uberX Car',
       '13.20 miles   00:20:40 Trip time   uberX Car',
       '1.03 miles   00:04:32 Trip time   uberX Car',
       '3.82 miles   00:10:54 Trip time   UberXL Car',
       '1.32 miles   00:09:32 Trip time   Black SUV Car',
       '8.32 miles   00:27:45 Trip time   uberX Car',
       '3.29 miles   00:08:23 Trip time   uberX Car',
       '0.95 miles   00:04:37 Trip time   uberX Car',
       '5.28 miles   00:11:21 Trip time   UberX Car',
       '11.62 miles   00:19:03 Trip time   uberX Car',
       '1.78 miles   00:10:29 Trip time   uberX Car',
       '2.90 miles   00:16:29 Trip time   uberX Car',
       '2.14 miles   00:10:57 Trip time   uberX Car',
       '1.06 miles   00:04:25 Trip time   Express Pool Car',
       '6.43 miles   00:42:03 Trip time   Pool Car',
       '2.71 miles   00:16:27 Trip time   uberX Car',
       '2.12 miles   00:04:54 Trip time   uberX Car',
       '7.81 miles   00:18:21 Trip time   uberX Car',
       '4.85 miles   00:12:45 Trip time   uberX Car',
       '4.05 miles   00:36:42 Trip time   uberXL Car',
       '10.50 miles   00:21:32 Trip time   uberX Car',
       '8.65 miles   00:35:22 Trip time   POOL Car',
       '9.82 miles   00:13:10 Trip time   uberX Car',
       '14.08 miles   00:35:35 Trip time   uberX Car',
       '4.84 miles   00:13:01 Trip time   Pool Car',
       '2.88 miles   00:07:46 Trip time   uberX Car',
       '1.39 miles   00:09:47 Trip time   uberX Car',
       '3.13 miles   00:17:55 Trip time   Pool Car',
       '0.96 miles   00:05:43 Trip time   uberX Car',
       '0.66 miles   00:03:55 Trip time   uberX Car',
       '1.18 miles   00:06:34 Trip time   UberX Car',
       '20.47 miles   00:22:33 Trip time   UberX Car',
       '4.42 miles   00:37:51 Trip time   Express Pool Car',
       '11.04 miles   00:50:04 Trip time   uberX Car',
       '6.44 miles   00:23:45 Trip time   POOL Car',
       '0.85 miles   00:02:39 Trip time   UberX Car',
       '1.84 miles   00:10:43 Trip time   uberXL Car',
       '2.38 miles   00:15:39 Trip time   POOL Car',
       '1.59 miles   00:10:40 Trip time   uberX Car',
       '17.15 miles   00:23:41 Trip time   uberXL Car',
       '2.83 miles   00:08:58 Trip time   uberX Car',
       '27.21 miles   00:40:54 Trip time   uberX Car',
       '2.73 miles   00:07:11 Trip time   uberX Car',
       '6.68 miles   00:31:13 Trip time   uberX Car',
       '1.14 miles   00:06:17 Trip time   uberX Car',
       '17.38 miles   00:21:44 Trip time   uberX Car',
       '6.26 miles   00:23:40 Trip time   uberX Car',
       '10.98 miles   00:34:22 Trip time   Express POOL Car',
       '1.30 miles   00:10:39 Trip time   Express Pool Car',
       '3.37 miles   00:12:26 Trip time   uberX Car',
       '25.97 miles   00:32:57 Trip time   POOL Car',
       '8.21 miles   00:17:34 Trip time   uberX Car',
       '2.80 miles   00:13:38 Trip time   uberX Car',
       '5.47 miles   00:36:05 Trip time   POOL Car',
       '12.42 miles   00:28:15 Trip time   uberXL Car',
       '2.86 miles   00:10:28 Trip time   uberPOOL Car',
       '11.82 miles   00:30:58 Trip time   POOL Car',
       '3.27 miles   00:12:51 Trip time   uberX Car',
       '0.49 miles   00:03:19 Trip time   uberX Car',
       '0.96 miles   00:04:38 Trip time   UberX Car',
       '0.97 miles   00:04:39 Trip time   UberX Car',
       '8.70 miles   00:17:47 Trip time   uberX Car',
       '3.67 miles   00:15:34 Trip time   uberPOOL Car',
       '19.57 miles   00:28:12 Trip time   uberX Car',
       '2.00 miles   00:07:29 Trip time   uberX Car',
       '1.46 miles   00:07:44 Trip time   uberX Car',
       '5.97 miles   00:15:50 Trip time   uberX Car',
       '3.74 miles   00:09:05 Trip time   uberX Car',
       '2.29 miles   00:11:54 Trip time   uberX Car',
       '1.90 miles   00:15:56 Trip time   UberX Car',
       '8.86 miles   00:20:58 Trip time   uberX Car',
       '1.75 miles   00:12:13 Trip time   uberXL Car',
       '0.85 miles   00:07:34 Trip time   BLACK Car',
       '5.30 miles   00:15:48 Trip time   UberX Car',
       '1.24 miles   00:09:01 Trip time   POOL Car',
       '2.44 miles   00:10:09 Trip time   uberCAB Car',
       '1.16 miles   00:04:54 Trip time   uberX Car',
       '5.20 miles   00:13:58 Trip time   VIP Car',
       '1.22 miles   00:11:30 Trip time   Express POOL Car',
       '1.99 miles   00:13:00 Trip time   Express POOL Car',
       '4.83 miles   00:11:32 Trip time   uberX Car',
       '1.94 miles   00:10:13 Trip time   uberX Car',
       '2.46 miles   00:10:15 Trip time   uberX Car',
       '1.45 miles   00:04:17 Trip time   uberX Car',
       '25.75 miles   00:27:45 Trip time   uberX Car',
       '1.91 miles   00:08:19 Trip time   uberXL Car',
       '2.43 miles   00:11:58 Trip time   uberX Car',
       '33.61 miles   00:49:10 Trip time   uberX Car',
       '46.93 miles   01:18:20 Trip time   uberX Car',
       '1.48 miles   00:06:28 Trip time   uberX Car',
       '0.68 miles   00:05:35 Trip time   Express POOL Car',
       '0.83 miles   00:13:44 Trip time   POOL Car',
       '27.08 miles   00:34:52 Trip time   uberX Car',
       '2.62 miles   00:13:06 Trip time   POOL Car',
       '1.32 miles   00:10:40 Trip time   Pool Car',
       '10.99 miles   00:15:00 Trip time   uberX Car',
       '2.88 miles   00:07:59 Trip time   uberX Car',
       '1.94 miles   00:05:07 Trip time   uberX Car',
       '1.81 miles   00:10:51 Trip time   UberRUSH Car',
       '28.92 miles   00:42:07 Trip time   UberX Car',
       '6.03 miles   00:13:03 Trip time   POOL Car',
       '4.32 miles   00:10:59 Trip time   uberX Car',
       '2.81 miles   00:19:36 Trip time   uberX Car',
       '7.66 miles   00:10:41 Trip time   UberX Car',
       '20.15 miles   00:35:00 Trip time   uberX Car',
       '3.11 miles   00:06:06 Trip time   uberX Car',
       '8.16 miles   00:12:10 Trip time   uberX Car',
       '1.75 miles   00:05:36 Trip time   uberX Car',
       '3.32 miles   00:15:58 Trip time   uberX Car',
       '22.39 miles   00:27:43 Trip time   uberX Car',
       '6.26 miles   00:25:21 Trip time   Pool Car',
       '17.66 miles   00:23:12 Trip time   UberX Car',
       '3.61 miles   00:34:19 Trip time   Pool Car',
       '3.76 miles   00:16:03 Trip time   POOL Car',
       '8.37 miles   00:35:16 Trip time   uberX Car',
       '1.55 miles   00:06:31 Trip time   uberX Car',
       '0.72 miles   00:02:29 Trip time   POOL Car',
       '6.46 miles   00:13:59 Trip time   uberX Car',
       '4.43 miles   00:15:01 Trip time   UberX Car',
       '5.89 miles   00:10:05 Trip time   uberX Car',
       '0.50 miles   00:08:10 Trip time   UberBLACK Car',
       '2.64 miles   00:06:55 Trip time   UberX Car',
       '3.09 miles   00:32:05 Trip time   Express POOL Car',
       '8.12 miles   00:10:50 Trip time   uberX Car',
       '22.25 miles   00:29:03 Trip time   UberX Car',
       '0.70 miles   00:10:27 Trip time   POOL Car',
       '24.86 miles   00:38:42 Trip time   UberX Car',
       '1.26 miles   00:05:24 Trip time   uberX Car',
       '12.60 miles   00:21:20 Trip time   uberX Car',
       '0.74 miles   00:15:03 Trip time   POOL Car',
       '7.86 miles   00:17:52 Trip time   uberX Car',
       '0.47 miles   00:03:51 Trip time   LUX Car',
       '30.65 miles   00:45:36 Trip time   uberX Car',
       '10.25 miles   00:16:45 Trip time   uberX Car',
       '5.93 miles   00:18:23 Trip time   uberX Car',
       '1.85 miles   00:10:17 Trip time   UberX Car',
       '0.66 miles   00:04:39 Trip time   uberX Car',
       '7.43 miles   00:42:35 Trip time   POOL Car',
       '3.61 miles   00:20:35 Trip time   Express Pool Car',
       '2.08 miles   00:05:49 Trip time   POOL Car',
       '5.37 miles   00:11:34 Trip time   UberX Car',
       '3.78 miles   00:12:47 Trip time   uberX Car',
       '2.84 miles   00:22:14 Trip time   uberX Car',
       '4.16 miles   00:14:34 Trip time   uberX Car',
       '1.15 miles   00:03:25 Trip time   uberX Car',
       '4.80 miles   00:12:41 Trip time   uberX Car',
       '1.80 miles   00:07:47 Trip time   uberX Car',
       '4.18 miles   00:11:56 Trip time   Pool Car',
       '5.98 miles   00:23:04 Trip time   uberX Car',
       '0.84 miles   00:04:51 Trip time   BLACK Car',
       '9.17 miles   00:15:41 Trip time   VIP Car',
       '2.48 miles   00:14:05 Trip time   UberX Car',
       '6.87 miles   00:14:59 Trip time   uberXL Car',
       '1.39 miles   00:04:21 Trip time   uberX Car',
       '3.72 miles   00:17:43 Trip time   uberX Car',
       '1.11 miles   00:06:03 Trip time   uberX Car',
       '1.76 miles   00:18:27 Trip time   uberX VIP Car',
       '9.52 miles   01:02:58 Trip time   BLACK CAR Car',
       '2.98 miles   00:28:22 Trip time   POOL Car',
       '1.83 miles   00:11:51 Trip time   uberX Car',
       '23.65 miles   00:27:28 Trip time   UberX Car',
       '6.60 miles   00:18:27 Trip time   uberX Car',
       '3.63 miles   00:31:35 Trip time   EXPRESS POOL Car',
       '4.37 miles   00:09:50 Trip time   uberX Car',
       '8.15 miles   00:23:48 Trip time   uberX Car',
       '2.89 miles   00:11:57 Trip time   UberX Car',
       '10.04 miles   00:35:42 Trip time   uberX Car',
       '5.41 miles   00:27:50 Trip time   POOL Car',
       '0.94 miles   00:10:02 Trip time   uberPOOL Car',
       '9.14 miles   00:21:56 Trip time   uberX Car',
       '0.20 miles   00:02:42 Trip time   UberX Car',
       '2.55 miles   00:18:31 Trip time   uberX Car',
       '14.78 miles   00:24:18 Trip time   uberX VIP Car',
       '12.07 miles   01:28:36 Trip time   POOL Car',
       '10.71 miles   00:42:14 Trip time   POOL Car',
       '4.39 miles   00:11:24 Trip time   uberX Car',
       '6.81 miles   00:23:43 Trip time   uberX Car',
       '4.46 miles   00:25:15 Trip time   uberX VIP Car',
       '13.70 miles   00:22:51 Trip time   uberX VIP Car',
       '2.75 miles   00:12:11 Trip time   UberX Car',
       '1.12 miles   00:07:52 Trip time   uberX Car',
       '12.10 miles   00:20:28 Trip time   uberX Car',
       '2.88 miles   00:21:43 Trip time   POOL Car',
       '8.90 miles   00:18:49 Trip time   Pool Car',
       '4.67 miles   00:19:25 Trip time   UberX Car',
       '1.81 miles   00:11:32 Trip time   UberX VIP Car',
       '1.55 miles   00:13:24 Trip time   POOL Car',
       '5.25 miles   00:14:35 Trip time   uberX Car',
       '3.04 miles   00:48:31 Trip time   uberX Car',
       '8.85 miles   00:15:34 Trip time   uberX VIP Car',
       '1.55 miles   00:13:58 Trip time   UberX VIP Car',
       '5.92 miles   00:33:12 Trip time   UberX Car',
       '0.70 miles   00:09:04 Trip time   uberPOOL Car',
       '2.19 miles   00:05:53 Trip time   uberX Car',
       '35.16 miles   00:37:59 Trip time   UberX Car',
       '6.82 miles   00:17:14 Trip time   UberX Car',
       '6.62 miles   00:31:31 Trip time   POOL Car',
       '11.70 miles   00:28:53 Trip time   UberX Car',
       '1.58 miles   00:07:16 Trip time   POOL Car',
       '0.63 miles   00:07:06 Trip time   UberX Car',
       '2.42 miles   00:19:49 Trip time   UberX Car',
       '9.71 miles   00:25:51 Trip time   UberX Car',
       '3.05 miles   00:23:07 Trip time   uberX Car',
       '18.05 miles   00:29:51 Trip time   BLACK VIP Car',
       '13.50 miles   00:23:54 Trip time   uberX Car',
       '2.38 miles   00:14:56 Trip time   Pool Car',
       '0.33 miles   00:03:12 Trip time   POOL Car',
       '1.30 miles   00:03:38 Trip time   VIP Car',
       '1.73 miles   00:04:34 Trip time   UberX Car',
       '2.20 miles   00:13:57 Trip time   uberX Car',
       '5.62 miles   00:10:57 Trip time   Pool Car',
       '3.40 miles   00:13:54 Trip time   uberX Car',
       '2.09 miles   00:07:05 Trip time   uberX Car',
       '0.87 miles   00:04:22 Trip time   Pool Car',
       '12.99 miles   00:27:34 Trip time   uberX Car',
       '1.52 miles   00:06:13 Trip time   uberX Car',
       '24.67 miles   00:33:21 Trip time   uberX Car',
       '3.33 miles   00:17:53 Trip time   uberX Car',
       '3.36 miles   00:33:35 Trip time   POOL Car',
       '24.81 miles   00:55:11 Trip time   uberX Car',
       '0.82 miles   00:09:14 Trip time   uberX Car',
       '4.50 miles   00:20:14 Trip time   UberX Car',
       '1.35 miles   00:07:47 Trip time   uberX Car',
       '13.39 miles   00:32:06 Trip time   UberX Car',
       '1.72 miles   00:11:09 Trip time   POOL Car',
       '3.00 miles   00:06:43 Trip time   uberX Car',
       '4.80 miles   00:35:48 Trip time   uberX Car',
       '31.99 miles   00:43:38 Trip time   uberX Car',
       '7.38 miles   00:15:40 Trip time   UberX Car',
       '0.89 miles   00:03:59 Trip time   uberX Car',
       '0.98 miles   00:04:41 Trip time   uberX Car',
       '1.28 miles   00:05:31 Trip time   uberX Car',
       '1.67 miles   00:08:19 Trip time   Pool Car',
       '2.42 miles   00:13:21 Trip time   uberX VIP Car',
       '4.48 miles   00:19:29 Trip time   UberX Car',
       '3.48 miles   00:13:56 Trip time   POOL Car',
       '1.40 miles   00:28:21 Trip time   UberX VIP Car',
       '1.36 miles   00:04:51 Trip time   uberX Car',
       '0.73 miles   00:08:21 Trip time   POOL Car',
       '7.47 miles   00:19:54 Trip time   uberX Car',
       '11.77 miles   00:36:32 Trip time   uberX Car',
       '14.71 miles   00:45:09 Trip time   Pool Car',
       '1.31 miles   00:12:10 Trip time   Pool Car',
       '2.57 miles   00:07:57 Trip time   uberX Car',
       '1.30 miles   00:05:15 Trip time   Pool Car',
       '3.58 miles   00:09:17 Trip time   UberX Car',
       '23.51 miles   00:23:46 Trip time   uberX Car',
       '29.36 miles   00:57:03 Trip time   uberX Car',
       '18.21 miles   00:59:27 Trip time   uberX Car',
       '8.78 miles   00:30:44 Trip time   UberX Car',
       '9.50 miles   00:23:16 Trip time   UberX Car',
       '6.81 miles   00:23:40 Trip time   POOL Car',
       '5.76 miles   00:26:04 Trip time   uberX Car',
       '5.24 miles   00:15:36 Trip time   uberX Car',
       '2.12 miles   00:11:24 Trip time   uberX Car',
       '4.49 miles   00:26:08 Trip time   EXPRESS POOL Car',
       '1.17 miles   00:06:23 Trip time   uberX Car',
       '4.81 miles   00:10:54 Trip time   uberX Car',
       '4.61 miles   00:27:02 Trip time   UberX Car',
       '0.92 miles   00:07:44 Trip time   POOL Car',
       '7.55 miles   00:17:36 Trip time   uberX Car',
       '0.68 miles   00:03:24 Trip time   uberX Car',
       '1.15 miles   00:06:44 Trip time   uberX Car',
       '8.21 miles   00:30:41 Trip time   uberX Car',
       '4.05 miles   00:17:38 Trip time   Pool Car',
       '35.72 miles   00:47:32 Trip time   uberX Car',
       '4.44 miles   00:12:42 Trip time   uberX Car',
       '1.17 miles   00:06:58 Trip time   uberX Car',
       '1.54 miles   00:08:52 Trip time   uberX Car',
       '7.01 miles   00:19:04 Trip time   uberX Car',
       '3.98 miles   00:12:01 Trip time   POOL Car',
       '1.02 miles   00:04:23 Trip time   uberX Car',
       '4.86 miles   00:45:37 Trip time   uberX Car',
       '2.39 miles   00:14:59 Trip time   uberX VIP Car',
       '4.63 miles   00:34:45 Trip time   UberX VIP Car',
       '3.17 miles   00:17:51 Trip time   uberX Car',
       '1.29 miles   00:07:21 Trip time   uberX Car',
       '6.71 miles   00:30:30 Trip time   POOL Car',
       '5.03 miles   00:17:33 Trip time   POOL Car',
       '1.71 miles   00:07:12 Trip time   UberX VIP Car',
       '0.88 miles   00:08:10 Trip time   UberX Car',
       '3.76 miles   00:15:04 Trip time   POOL Car',
       '1.65 miles   00:11:38 Trip time   uberX Car',
       '3.43 miles   00:18:10 Trip time   UberX VIP Car',
       '1.00 miles   00:07:49 Trip time   uberX Car',
       '5.48 miles   00:48:05 Trip time   POOL Car',
       '17.14 miles   00:48:46 Trip time   Black VIP Car',
       '3.33 miles   00:35:24 Trip time   POOL Car',
       '9.55 miles   00:35:06 Trip time   Pool Car',
       '4.75 miles   00:36:55 Trip time   uberX VIP Car',
       '4.29 miles   00:20:10 Trip time   uberX Car',
       '0.91 miles   00:11:29 Trip time   POOL Car',
       '43.87 miles   00:51:23 Trip time   VIP Car',
       '2.20 miles   00:08:20 Trip time   uberX Car',
       '3.85 miles   00:09:01 Trip time   Pool Car',
       '3.34 miles   00:31:45 Trip time   POOL Car',
       '17.48 miles   00:35:36 Trip time   uberX Car',
       '11.17 miles   00:28:31 Trip time   POOL Car',
       '3.81 miles   00:18:28 Trip time   BLACK Car',
       '1.33 miles   00:13:41 Trip time   uberPOOL Car',
       '1.23 miles   00:12:18 Trip time   uberX Car',
       '10.78 miles   00:17:16 Trip time   uberX Car',
       '7.91 miles   00:25:35 Trip time   Select Car',
       '4.37 miles   00:35:02 Trip time   uberPOOL Car',
       '15.48 miles   00:26:39 Trip time   UberX Car',
       '0.67 miles   00:02:56 Trip time   uberX Car',
       '3.42 miles   00:09:11 Trip time   Pool Car',
       '0.82 miles   00:05:44 Trip time   uberX VIP Car',
       '2.81 miles   00:10:27 Trip time   uberX Car',
       '1.40 miles   00:04:52 Trip time   uberX Car',
       '6.09 miles   00:20:27 Trip time   uberX Car',
       '6.55 miles   00:22:57 Trip time   uberX Car',
       '1.66 miles   00:08:08 Trip time   UberX Car',
       '4.59 miles   00:27:48 Trip time   uberX Car',
       '3.03 miles   00:07:27 Trip time   UberXL Car',
       '4.21 miles   00:10:13 Trip time   UberX Car',
       '5.09 miles   00:11:05 Trip time   uberPOOL Car',
       '1.30 miles   00:04:28 Trip time   UberX Car',
       '3.09 miles   00:15:48 Trip time   Pool Car',
       '1.91 miles   00:06:51 Trip time   uberX Car',
       '0.79 miles   00:09:23 Trip time   uberX Car',
       '6.16 miles   00:10:44 Trip time   uberX Car',
       '2.04 miles   00:06:13 Trip time   UberX Car',
       '6.01 miles   00:15:29 Trip time   uberX Car',
       '3.82 miles   00:09:42 Trip time   uberX Car',
       '1.67 miles   00:05:28 Trip time   uberX Car',
       '10.45 miles   00:19:45 Trip time   uberX Car',
       '0.99 miles   00:04:09 Trip time   uberX Car',
       '15.65 miles   00:23:37 Trip time   uberX Car',
       '24.01 miles   00:26:42 Trip time   uberX Car',
       '3.68 miles   00:09:21 Trip time   UberX Car',
       '2.10 miles   00:12:44 Trip time   uberX Car',
       '5.89 miles   00:15:50 Trip time   UberX Car',
       '1.97 miles   00:08:24 Trip time   uberX Car',
       '9.56 miles   00:14:59 Trip time   uberX Car',
       '1.78 miles   00:09:12 Trip time   uberX Car',
       '10.07 miles   00:18:06 Trip time   Pool Car',
       '0.84 miles   00:04:09 Trip time   uberX Car',
       '1.51 miles   00:06:37 Trip time   POOL Car',
       '4.72 miles   00:18:09 Trip time   uberX Car',
       '1.72 miles   00:05:08 Trip time   uberX Car',
       '9.09 miles   00:13:49 Trip time   SUV Car',
       '6.51 miles   00:16:57 Trip time   Pool Car',
       '16.16 miles   00:21:03 Trip time   UberXL Car',
       '4.38 miles   00:14:54 Trip time   uberX Car',
       '2.89 miles   00:14:11 Trip time   POOL Car',
       '2.66 miles   00:12:40 Trip time   UberX Car',
       '1.72 miles   00:13:04 Trip time   uberX VIP Car',
       '11.25 miles   00:28:03 Trip time   uberX Car',
       '3.17 miles   00:10:16 Trip time   uberX Car',
       '11.93 miles   00:23:38 Trip time   UberX Car',
       '8.12 miles   00:23:03 Trip time   uberX Car',
       '1.62 miles   00:10:36 Trip time   uberX Car',
       '3.83 miles   00:13:43 Trip time   uberX Car',
       '1.94 miles   00:16:53 Trip time   uberX Car',
       '7.35 miles   00:13:35 Trip time   uberX Car',
       '1.96 miles   00:07:26 Trip time   uberX Car',
       '1.87 miles   00:11:34 Trip time   uberX Car',
       '10.32 miles   00:14:28 Trip time   UberX Car',
       '8.17 miles   00:16:50 Trip time   UberX Car',
       '3.51 miles   00:10:52 Trip time   Pool Car',
       '4.11 miles   00:14:23 Trip time   Pool Car',
       '2.94 miles   00:20:45 Trip time   uberX Car',
       '7.40 miles   00:29:02 Trip time   uberX Car',
       '11.02 miles   00:20:59 Trip time   uberX VIP Car',
       '1.10 miles   00:04:44 Trip time   POOL Car',
       '6.48 miles   00:25:11 Trip time   POOL Car',
       '2.38 miles   00:06:14 Trip time   uberX Car',
       '13.88 miles   00:34:28 Trip time   UberX Car',
       '6.12 miles   00:30:53 Trip time   uberX VIP Car',
       '2.11 miles   00:17:16 Trip time   UberX Car',
       '0.20 miles   00:07:22 Trip time   POOL Car',
       '1.77 miles   00:11:40 Trip time   uberX Car',
       '6.39 miles   00:23:13 Trip time   POOL Car',
       '3.42 miles   00:29:55 Trip time   POOL Car',
       '2.41 miles   00:16:50 Trip time   uberX Car',
       '0.47 miles   00:07:11 Trip time   POOL Car',
       '5.57 miles   00:21:53 Trip time   uberX Car',
       '1.00 miles   00:07:41 Trip time   uberX VIP Car',
       '3.13 miles   00:09:52 Trip time   UberX Car',
       '1.43 miles   00:10:21 Trip time   uberX VIP Car',
       '2.24 miles   00:14:11 Trip time   uberX VIP Car',
       '1.18 miles   00:04:38 Trip time   uberX Car',
       '2.10 miles   00:22:37 Trip time   uberX Car',
       '5.92 miles   00:44:50 Trip time   UberX Car',
       '4.67 miles   00:23:39 Trip time   uberX Car',
       '5.69 miles   00:51:21 Trip time   uberX Car',
       '1.80 miles   00:09:21 Trip time   uberX Car',
       '1.09 miles   00:05:59 Trip time   POOL Car',
       '2.39 miles   00:12:30 Trip time   uberX Car',
       '3.32 miles   00:21:26 Trip time   Pool Car',
       '2.45 miles   00:15:43 Trip time   uberX Car',
       '1.41 miles   00:07:37 Trip time   POOL Car',
       '1.96 miles   00:10:27 Trip time   uberX Car',
       '4.27 miles   00:26:07 Trip time   Pool Car',
       '0.58 miles   00:03:14 Trip time   POOL Car',
       '4.53 miles   00:25:59 Trip time   POOL Car',
       '0.81 miles   00:08:27 Trip time   UberX Car',
       '1.60 miles   00:13:55 Trip time   BLACK CAR Car',
       '3.13 miles   00:14:18 Trip time   UberX VIP Car',
       '1.88 miles   00:15:12 Trip time   uberX Car',
       '1.95 miles   00:23:19 Trip time   uberX Car',
       '4.46 miles   00:18:19 Trip time   UberX Car',
       '34.76 miles   00:50:23 Trip time   uberX Car',
       '7.38 miles   00:18:53 Trip time   UberX Car',
       '11.31 miles   00:25:55 Trip time   POOL Car',
       '3.04 miles   00:10:25 Trip time   UberX Car',
       '4.34 miles   00:11:54 Trip time   uberX Car',
       '0.84 miles   00:07:25 Trip time   UberX Car',
       '5.72 miles   00:19:28 Trip time   UberXL Car',
       '1.46 miles   00:10:24 Trip time   UberX Car',
       '6.21 miles   00:11:29 Trip time   uberX Car',
       '8.95 miles   00:23:50 Trip time   uberX Car',
       '2.01 miles   00:05:36 Trip time   uberX Car',
       '3.04 miles   00:15:11 Trip time   uberX Car',
       '0.80 miles   00:05:10 Trip time   Pool Car',
       '1.87 miles   00:07:02 Trip time   uberX Car',
       '3.01 miles   00:08:17 Trip time   UberX Car',
       '0.93 miles   00:12:16 Trip time   SUV Car',
       '3.16 miles   00:09:50 Trip time   uberX Car',
       '2.38 miles   00:11:59 Trip time   uberX Car',
       '15.64 miles   00:18:42 Trip time   UberX Car',
       '7.08 miles   00:16:45 Trip time   UberXL Car',
       '9.16 miles   00:12:39 Trip time   UberX Car',
       '6.35 miles   00:28:10 Trip time   uberX Car',
       '18.06 miles   00:24:12 Trip time   uberX Car',
       '27.45 miles   00:35:27 Trip time   uberX Car',
       '13.97 miles   00:22:42 Trip time   uberX Car',
       '5.74 miles   00:18:33 Trip time   uberX Car',
       '7.67 miles   00:12:02 Trip time   UberX Car',
       '7.46 miles   00:17:43 Trip time   UberX Car',
       '6.07 miles   00:12:51 Trip time   uberX Car',
       '2.00 miles   00:09:01 Trip time   UberXL Car',
       '8.47 miles   00:20:06 Trip time   uberX Car',
       '10.77 miles   00:51:27 Trip time   Pool Car',
       '4.45 miles   00:08:16 Trip time   uberX Car',
       '5.85 miles   00:11:48 Trip time   VIP Car',
       '8.06 miles   00:14:29 Trip time   uberX Car',
       '1.52 miles   00:10:22 Trip time   uberX Car',
       '33.07 miles   00:47:17 Trip time   UberX Car',
       '2.47 miles   00:09:18 Trip time   Pool Car',
       '7.77 miles   00:22:49 Trip time   POOL Car',
       '0.86 miles   00:04:47 Trip time   uberX Car',
       '0.79 miles   00:09:48 Trip time   UberX Car',
       '16.33 miles   00:33:30 Trip time   uberX Car',
       '1.14 miles   00:04:50 Trip time   uberX Car',
       '1.64 miles   00:07:29 Trip time   uberX Car',
       '5.68 miles   00:16:18 Trip time   uberX Car',
       '1.72 miles   00:07:27 Trip time   POOL Car',
       '0.77 miles   00:04:24 Trip time   uberX Car',
       '4.66 miles   00:22:41 Trip time   uberX Car',
       '17.27 miles   00:41:08 Trip time   uberX Car',
       '3.63 miles   00:15:33 Trip time   uberX Car',
       '0.58 miles   00:05:16 Trip time   UberX Car',
       '0.95 miles   00:06:47 Trip time   UberX Car',
       '15.55 miles   00:20:14 Trip time   uberX Car',
       '1.45 miles   00:04:49 Trip time   uberX Car',
       '14.04 miles   00:34:12 Trip time   POOL Car',
       '35.36 miles   00:40:33 Trip time   uberX Car',
       '4.12 miles   00:15:08 Trip time   uberX Car',
       '17.40 miles   00:26:22 Trip time   uberPOOL Car',
       '6.56 miles   00:15:45 Trip time   UberX Car',
       '3.16 miles   00:12:30 Trip time   uberX Car',
       '1.65 miles   00:06:34 Trip time   POOL Car',
       '30.48 miles   00:46:08 Trip time   uberX Car',
       '1.84 miles   00:14:11 Trip time   EXPRESS POOL Car',
       '8.61 miles   00:11:22 Trip time   UberX Car',
       '4.49 miles   00:26:22 Trip time   uberX Car',
       '8.67 miles   00:29:28 Trip time   Express Pool Car',
       '6.01 miles   00:22:42 Trip time   uberX Car',
       '1.81 miles   00:08:17 Trip time   uberX Car',
       '20.31 miles   00:55:11 Trip time   uberX Car',
       '8.21 miles   00:34:22 Trip time   uberPOOL Car',
       '5.20 miles   00:14:13 Trip time   uberX Car',
       '8.35 miles   00:22:34 Trip time   POOL Car',
       '2.84 miles   00:07:34 Trip time   Pool Car',
       '12.39 miles   00:41:54 Trip time   Pool Car',
       '2.77 miles   00:13:24 Trip time   UberX Car',
       '23.26 miles   00:27:24 Trip time   uberX Car',
       '13.49 miles   00:15:58 Trip time   uberX Car',
       '16.54 miles   00:27:22 Trip time   POOL Car',
       '0.88 miles   00:05:50 Trip time   uberX Car',
       '0.13 miles   00:01:27 Trip time   uberX Car',
       '1.32 miles   00:07:18 Trip time   UberX Car',
       '2.09 miles   00:12:02 Trip time   uberX Car',
       '0.78 miles   00:03:50 Trip time   uberX Car',
       '0.94 miles   00:05:22 Trip time   Express Pool Car',
       '21.60 miles   00:35:47 Trip time   uberX Car',
       '7.25 miles   00:13:22 Trip time   uberX Car',
       '4.27 miles   00:11:44 Trip time   UberX Car',
       '2.09 miles   00:09:16 Trip time   uberX Car',
       '3.58 miles   00:13:02 Trip time   uberX Car',
       '2.17 miles   00:07:33 Trip time   uberX Car',
       '0.48 miles   00:03:02 Trip time   uberX Car',
       '1.69 miles   00:17:43 Trip time   UberX Car',
       '0.96 miles   00:10:42 Trip time   uberPOOL Car',
       '6.09 miles   00:29:40 Trip time   Pool Car',
       '1.66 miles   00:10:26 Trip time   uberX Car',
       '14.00 miles   00:22:34 Trip time   POOL Car',
       '1.01 miles   00:09:38 Trip time   POOL Car',
       '1.13 miles   00:13:33 Trip time   POOL Car',
       '2.22 miles   00:24:03 Trip time   uberX Car',
       '18.44 miles   00:53:17 Trip time   uberX Car',
       '2.15 miles   00:05:41 Trip time   uberX Car',
       '7.48 miles   00:14:52 Trip time   uberX Car',
       '5.54 miles   00:07:48 Trip time   uberX Car',
       '1.42 miles   00:05:50 Trip time   uberX Car',
       '3.01 miles   00:08:18 Trip time   uberX Car',
       '7.32 miles   00:15:45 Trip time   UberX Car',
       '5.98 miles   00:24:04 Trip time   Express POOL Car',
       '8.06 miles   00:10:59 Trip time   uberX Car',
       '0.72 miles   00:04:31 Trip time   uberX Car',
       '1.13 miles   00:15:22 Trip time   POOL Car',
       '2.56 miles   00:09:25 Trip time   Pool Car',
       '10.05 miles   00:20:24 Trip time   uberX Car',
       '23.47 miles   00:28:27 Trip time   uberX Car',
       '2.61 miles   00:07:24 Trip time   uberX Car',
       '2.24 miles   00:08:57 Trip time   POOL Car',
       '6.27 miles   00:16:01 Trip time   uberX Car',
       '38.01 miles   00:42:49 Trip time   uberX Car',
       '16.21 miles   00:28:43 Trip time   UberX VIP Car',
       '6.16 miles   00:21:41 Trip time   uberX Car',
       '12.76 miles   00:20:07 Trip time   POOL Car',
       '1.38 miles   00:07:28 Trip time   uberX Car',
       '27.35 miles   00:38:27 Trip time   UberX Car',
       '4.51 miles   00:11:06 Trip time   POOL Car',
       '5.66 miles   00:15:30 Trip time   Pool Car',
       '13.70 miles   00:21:28 Trip time   uberX Car',
       '0.47 miles   00:02:25 Trip time   VIP Car',
       '12.83 miles   00:17:00 Trip time   uberX Car',
       '6.56 miles   00:24:57 Trip time   Pool Car',
       '2.09 miles   00:07:57 Trip time   uberX Car',
       '3.10 miles   00:06:52 Trip time   UberX Car',
       '11.50 miles   00:19:46 Trip time   uberX Car',
       '2.92 miles   00:15:46 Trip time   uberX Car',
       '8.81 miles   00:12:00 Trip time   uberX Car',
       '0.35 miles   00:02:36 Trip time   uberX Car',
       '2.46 miles   00:06:23 Trip time   uberX Car',
       '4.92 miles   00:12:47 Trip time   uberX Car',
       '2.90 miles   00:06:24 Trip time   uberX Car',
       '12.18 miles   00:27:22 Trip time   uberX Car',
       '1.14 miles   00:06:29 Trip time   uberX Car',
       '19.55 miles   00:30:49 Trip time   uberX Car',
       '3.41 miles   00:12:37 Trip time   uberX Car',
       '6.54 miles   00:32:32 Trip time   uberX Car',
       '1.77 miles   00:15:58 Trip time   POOL Car',
       '5.36 miles   00:09:29 Trip time   uberX Car',
       '1.40 miles   00:08:30 Trip time   uberX Car',
       '1.39 miles   00:12:00 Trip time   uberX Car',
       '2.28 miles   00:13:06 Trip time   POOL Car',
       '3.93 miles   00:12:29 Trip time   Express Pool Car',
       '1.96 miles   00:14:24 Trip time   Express POOL Car',
       '6.86 miles   00:19:41 Trip time   uberX Car',
       '6.29 miles   00:18:18 Trip time   uberX Car',
       '0.78 miles   00:03:26 Trip time   uberX Car',
       '11.32 miles   00:17:21 Trip time   POOL Car',
       '5.35 miles   00:06:37 Trip time   uberX Car',
       '3.46 miles   00:12:23 Trip time   uberX Car',
       '7.66 miles   00:13:27 Trip time   uberX Car',
       '3.20 miles   00:15:36 Trip time   UberX Car',
       '13.67 miles   00:37:30 Trip time   uberX Car',
       '2.13 miles   00:16:33 Trip time   uberX Car',
       '1.53 miles   00:06:10 Trip time   UberXL Car',
       '0.75 miles   00:05:49 Trip time   uberX Car',
       '4.19 miles   00:06:58 Trip time   Pool Car',
       '2.42 miles   00:08:04 Trip time   uberPOOL Car',
       '3.38 miles   00:11:28 Trip time   Pool Car',
       '6.46 miles   00:20:36 Trip time   uberX Car',
       '3.80 miles   00:22:19 Trip time   POOL Car',
       '16.64 miles   00:45:31 Trip time   Express POOL Car',
       '12.61 miles   00:19:08 Trip time   uberX Car',
       '1.92 miles   00:04:17 Trip time   uberX Car',
       '4.28 miles   00:10:29 Trip time   uberX Car',
       '0.70 miles   00:01:59 Trip time   uberX Car',
       '2.79 miles   00:07:16 Trip time   uberX Car',
       '8.05 miles   00:18:44 Trip time   POOL Car',
       '1.11 miles   00:12:13 Trip time   POOL Car',
       '4.96 miles   00:13:04 Trip time   uberX Car',
       '14.26 miles   00:46:42 Trip time   uberX Car',
       '3.33 miles   00:15:50 Trip time   uberX Car',
       '12.95 miles   00:18:42 Trip time   uberX Car',
       '15.12 miles   00:22:25 Trip time   uberXL Car',
       '2.22 miles   00:12:25 Trip time   uberX Car',
       '1.70 miles   00:14:58 Trip time   uberX Car',
       '2.17 miles   00:13:39 Trip time   uberX VIP Car',
       '4.68 miles   00:30:49 Trip time   Pool Car',
       '3.59 miles   00:10:45 Trip time   uberXL Car',
       '2.71 miles   00:08:29 Trip time   uberX Car',
       '11.12 miles   00:15:50 Trip time   uberX Car',
       '13.36 miles   00:37:30 Trip time   UberX Car',
       '4.14 miles   00:15:57 Trip time   Pool Car',
       '3.33 miles   00:13:16 Trip time   UberX Car',
       '1.90 miles   00:09:09 Trip time   uberX Car',
       '2.15 miles   00:10:16 Trip time   uberX Car',
       '6.76 miles   00:16:16 Trip time   UberX Car',
       '2.20 miles   00:11:43 Trip time   UberX Car',
       '2.18 miles   00:20:04 Trip time   Pool Car',
       '2.41 miles   00:14:09 Trip time   UberX Car',
       '3.26 miles   00:09:50 Trip time   uberX Car',
       '2.33 miles   00:11:09 Trip time   uberX Car',
       '4.57 miles   00:10:20 Trip time   UberX Car',
       '2.09 miles   00:06:53 Trip time   uberX Car',
       '1.21 miles   00:08:39 Trip time   POOL Car',
       '5.05 miles   00:22:48 Trip time   uberX Car',
       '5.37 miles   00:11:59 Trip time   uberX Car',
       '1.92 miles   00:12:45 Trip time   Pool Car',
       '9.25 miles   00:15:49 Trip time   uberX Car',
       '1.74 miles   00:07:22 Trip time   VIP Car',
       '7.12 miles   00:15:37 Trip time   uberX Car',
       '3.19 miles   00:23:54 Trip time   POOL Car',
       '7.28 miles   00:34:55 Trip time   SUV Car',
       '4.67 miles   00:16:11 Trip time   UberX Car',
       '1.67 miles   00:06:30 Trip time   uberPOOL Car',
       '1.08 miles   00:03:45 Trip time   uberX Car',
       '21.49 miles   00:34:40 Trip time   uberX Car',
       '1.27 miles   00:07:36 Trip time   POOL Car',
       '22.09 miles   00:25:36 Trip time   UberX Car',
       '2.30 miles   00:10:28 Trip time   UberX Car',
       '1.21 miles   00:10:01 Trip time   uberX Car',
       '7.07 miles   00:43:17 Trip time   Express POOL Car',
       '2.92 miles   00:22:00 Trip time   POOL Car',
       '0.80 miles   00:09:04 Trip time   uberX Car',
       '7.55 miles   00:18:47 Trip time   uberX Car',
       '10.02 miles   00:27:47 Trip time   uberX Car',
       '2.23 miles   00:17:05 Trip time   uberXL Car',
       '1.88 miles   00:07:45 Trip time   uberX Car',
       '13.29 miles   00:23:32 Trip time   POOL Car',
       '10.48 miles   00:22:02 Trip time   uberX Car',
       '5.17 miles   00:27:37 Trip time   Express POOL Car',
       '11.17 miles   00:48:42 Trip time   UberX Car',
       '15.49 miles   00:18:51 Trip time   uberX Car',
       '1.46 miles   00:07:23 Trip time   uberX Car',
       '1.86 miles   00:06:05 Trip time   uberX Car',
       '4.83 miles   00:31:49 Trip time   POOL Car',
       '0.78 miles   00:08:49 Trip time   uberX Car',
       '10.18 miles   00:39:37 Trip time   UberX Car',
       '18.79 miles   00:53:50 Trip time   UberX VIP Car',
       '2.76 miles   00:13:54 Trip time   uberX Car',
       '0.74 miles   00:08:42 Trip time   UberX Car',
       '25.28 miles   00:36:22 Trip time   uberX Car',
       '17.18 miles   00:23:41 Trip time   UberX Car',
       '3.75 miles   00:16:29 Trip time   uberX Car',
       '0.53 miles   00:03:57 Trip time   uberX Car',
       '0.91 miles   00:04:17 Trip time   uberX Car',
       '1.01 miles   00:04:09 Trip time   Pool Car',
       '2.96 miles   00:09:19 Trip time   UberX Car',
       '25.95 miles   00:28:54 Trip time   uberX Car',
       '3.72 miles   00:10:21 Trip time   Pool Car',
       '11.28 miles   00:33:39 Trip time   uberX Car',
       '4.17 miles   00:07:40 Trip time   UberX Car',
       '8.92 miles   00:17:46 Trip time   uberX Car',
       '4.21 miles   00:14:52 Trip time   uberX Car',
       '2.34 miles   00:11:04 Trip time   Express Pool Car',
       '14.22 miles   00:31:28 Trip time   uberX Car',
       '5.95 miles   00:12:25 Trip time   POOL Car',
       '7.53 miles   00:21:51 Trip time   uberX Car',
       '2.27 miles   00:12:00 Trip time   uberX Car',
       '15.01 miles   00:47:59 Trip time   uberX Car',
       '14.28 miles   00:23:00 Trip time   UberX Car',
       '4.88 miles   00:12:06 Trip time   UberX Car',
       '3.39 miles   00:18:42 Trip time   uberX Car',
       '5.45 miles   00:15:23 Trip time   Select Car',
       '1.39 miles   00:15:34 Trip time   Select Car',
       '8.14 miles   00:20:13 Trip time   uberX Car',
       '0.82 miles   00:02:38 Trip time   uberX Car',
       '0.76 miles   00:11:15 Trip time   POOL Car',
       '38.93 miles   00:51:46 Trip time   uberX Car',
       '1.04 miles   00:11:36 Trip time   POOL Car',
       '4.49 miles   00:15:32 Trip time   UberSUV Car',
       '2.95 miles   00:12:49 Trip time   uberX Car',
       '3.82 miles   00:24:26 Trip time   Pool Car',
       '2.67 miles   00:16:56 Trip time   uberPOOL Car',
       '3.18 miles   00:17:49 Trip time   uberX Car',
       '0.93 miles   00:05:29 Trip time   POOL Car',
       '0.73 miles   00:03:18 Trip time   uberX Car',
       '0.58 miles   00:06:06 Trip time   Express Pool Car',
       '3.58 miles   00:09:49 Trip time   POOL Car',
       '19.06 miles   00:43:36 Trip time   UberX Car',
       '1.72 miles   00:06:42 Trip time   uberX Car',
       '6.25 miles   00:18:01 Trip time   uberX Car',
       '22.27 miles   00:31:12 Trip time   uberX Car',
       '1.19 miles   00:03:24 Trip time   uberX Car',
       '2.28 miles   00:11:10 Trip time   uberX Car',
       '3.44 miles   00:09:06 Trip time   uberX Car',
       '2.89 miles   00:08:52 Trip time   uberX Car',
       '4.34 miles   00:13:11 Trip time   uberX Car',
       '3.20 miles   00:12:31 Trip time   VIP Car',
       '5.42 miles   00:12:58 Trip time   uberX Car',
       '6.81 miles   00:14:52 Trip time   uberX Car',
       '11.13 miles   00:16:17 Trip time   UberX Car',
       '10.58 miles   00:15:30 Trip time   uberX Car',
       '5.80 miles   00:11:05 Trip time   POOL Car',
       '1.48 miles   00:04:02 Trip time   uberXL Car',
       '2.08 miles   00:13:27 Trip time   uberX Car',
       '0.69 miles   00:04:52 Trip time   uberX Car',
       '3.66 miles   00:09:54 Trip time   uberX Car',
       '12.58 miles   00:34:55 Trip time   UberXL Car',
       '7.87 miles   00:21:37 Trip time   uberX Car',
       '23.57 miles   00:41:34 Trip time   uberX Car',
       '2.14 miles   00:11:36 Trip time   uberX VIP Car',
       '0.81 miles   00:04:52 Trip time   uberX VIP Car',
       '1.38 miles   00:12:18 Trip time   uberX Car',
       '2.84 miles   00:18:04 Trip time   uberX VIP Car',
       '2.92 miles   00:19:55 Trip time   UberX Car',
       '1.89 miles   00:09:24 Trip time   POOL Car',
       '3.75 miles   00:23:36 Trip time   uberX Car',
       '6.34 miles   00:24:25 Trip time   uberX VIP Car',
       '3.04 miles   00:24:45 Trip time   uberX VIP Car',
       '1.15 miles   00:05:41 Trip time   uberX VIP Car',
       '3.21 miles   00:21:55 Trip time   UberX Car',
       '4.42 miles   00:27:05 Trip time   Pool Car',
       '1.44 miles   00:23:35 Trip time   uberX VIP Car',
       '4.25 miles   00:28:06 Trip time   Pool Car',
       '9.25 miles   00:27:10 Trip time   uberX Car',
       '3.82 miles   00:22:00 Trip time   uberX VIP Car',
       '9.47 miles   00:24:01 Trip time   uberX Car',
       '1.77 miles   00:16:43 Trip time   uberX VIP Car',
       '6.58 miles   00:20:11 Trip time   uberX Car',
       '11.55 miles   00:31:14 Trip time   POOL Car',
       '16.20 miles   00:38:54 Trip time   UberX Car',
       '2.14 miles   00:08:39 Trip time   uberX Car',
       '2.44 miles   00:17:48 Trip time   uberX Car',
       '6.09 miles   00:19:24 Trip time   uberX VIP Car',
       '0.82 miles   00:03:28 Trip time   Pool Car',
       '2.47 miles   00:14:00 Trip time   uberX Car',
       '3.24 miles   00:15:51 Trip time   POOL Car',
       '6.85 miles   00:46:43 Trip time   Pool Car',
       '2.32 miles   00:12:18 Trip time   uberX VIP Car',
       '2.18 miles   00:18:40 Trip time   POOL Car',
       '0.08 miles   00:00:37 Trip time   Pool Car',
       '4.46 miles   00:22:35 Trip time   Pool Car',
       '2.94 miles   00:27:44 Trip time   POOL Car',
       '2.18 miles   00:14:28 Trip time   Pool Car',
       '4.80 miles   00:20:20 Trip time   uberXL Car',
       '15.22 miles   00:33:22 Trip time   uberX Car',
       '7.03 miles   00:17:35 Trip time   uberX Car',
       '2.15 miles   00:05:57 Trip time   UberX Car',
       '35.76 miles   01:00:33 Trip time   UberX Car',
       '12.42 miles   00:23:13 Trip time   uberX Car',
       '6.35 miles   00:37:36 Trip time   UberX Car',
       '9.00 miles   00:30:52 Trip time   UberX Car',
       '89.67 miles   01:40:17 Trip time   uberX VIP Car',
       '14.66 miles   00:38:41 Trip time   UberX Car',
       '13.84 miles   00:26:03 Trip time   uberX Car',
       '0.86 miles   00:03:47 Trip time   UberX Car',
       '4.84 miles   00:19:18 Trip time   POOL Car',
       '0.67 miles   00:06:27 Trip time   Express Pool Car',
       '13.71 miles   00:31:02 Trip time   POOL Car',
       '0.70 miles   00:02:47 Trip time   uberX Car',
       '3.89 miles   00:18:11 Trip time   UberX Car',
       '2.83 miles   00:12:06 Trip time   uberX Car',
       '1.66 miles   00:14:44 Trip time   Pool Car',
       '3.97 miles   00:08:56 Trip time   UberX Car',
       '9.63 miles   00:15:03 Trip time   uberX Car',
       '9.17 miles   00:18:20 Trip time   uberX Car',
       '1.29 miles   00:07:31 Trip time   Black SUV Car',
       '1.36 miles   00:10:57 Trip time   UberX Car',
       '13.21 miles   00:17:53 Trip time   uberX Car',
       '6.28 miles   00:13:51 Trip time   uberX Car',
       '6.99 miles   00:20:12 Trip time   POOL Car',
       '5.93 miles   00:16:36 Trip time   uberX Car',
       '0.96 miles   00:06:51 Trip time   POOL Car',
       '4.37 miles   00:07:27 Trip time   UberX Car',
       '2.92 miles   00:09:43 Trip time   uberX Car',
       '4.92 miles   00:13:25 Trip time   uberX Car',
       '1.03 miles   00:05:39 Trip time   uberX Car',
       '1.45 miles   00:12:24 Trip time   UberX Car',
       '16.01 miles   00:19:43 Trip time   UberX Car',
       '12.40 miles   00:24:00 Trip time   UberX Car',
       '21.23 miles   00:33:17 Trip time   uberX Car',
       '5.10 miles   00:22:10 Trip time   uberX Car',
       '14.93 miles   00:26:38 Trip time   uberX Car',
       '0.61 miles   00:02:49 Trip time   uberX Car',
       '6.61 miles   00:16:32 Trip time   uberX Car',
       '1.60 miles   00:07:14 Trip time   uberX Car',
       '4.22 miles   00:18:03 Trip time   uberX Car',
       '2.63 miles   00:10:01 Trip time   Pool Car',
       '4.56 miles   00:07:27 Trip time   uberX Car',
       '14.66 miles   00:22:05 Trip time   uberX Car',
       '1.13 miles   00:09:05 Trip time   uberX Car',
       '0.80 miles   00:02:26 Trip time   POOL Car',
       '8.83 miles   00:21:08 Trip time   uberX Car',
       '1.02 miles   00:04:24 Trip time   uberX Car',
       '3.15 miles   00:08:19 Trip time   POOL Car',
       '3.87 miles   00:09:26 Trip time   uberX Car',
       '0.90 miles   00:03:34 Trip time   uberX Car',
       '13.60 miles   00:31:20 Trip time   UberX Car',
       '0.82 miles   00:04:36 Trip time   uberX Car',
       '5.42 miles   00:14:24 Trip time   uberX Car',
       '4.01 miles   00:19:55 Trip time   uberX Car',
       '2.07 miles   00:09:34 Trip time   uberX Car',
       '5.04 miles   00:17:31 Trip time   uberX Car',
       '1.11 miles   00:04:46 Trip time   UberX Car',
       '2.16 miles   00:07:01 Trip time   uberX Car',
       '2.33 miles   00:12:04 Trip time   uberX Car',
       '3.60 miles   00:19:35 Trip time   POOL Car',
       '10.87 miles   00:21:09 Trip time   uberX Car',
       '8.20 miles   00:15:00 Trip time   uberX Car',
       '12.99 miles   00:27:32 Trip time   uberX Car',
       '2.44 miles   00:10:16 Trip time   uberX Car',
       '3.61 miles   00:10:33 Trip time   uberX Car',
       '10.28 miles   00:24:08 Trip time   BLACK Car',
       '3.17 miles   00:12:40 Trip time   uberX Car',
       '11.20 miles   00:24:51 Trip time   UberX Car',
       '0.95 miles   00:08:06 Trip time   Express POOL Car',
       '2.30 miles   00:10:10 Trip time   uberX Car',
       '13.53 miles   00:19:29 Trip time   uberX Car',
       '2.62 miles   00:08:36 Trip time   UberX Car',
       '8.73 miles   00:13:54 Trip time   uberX Car',
       '2.11 miles   00:12:44 Trip time   UberX Car',
       '1.40 miles   00:08:51 Trip time   uberX Car',
       '11.78 miles   00:18:24 Trip time   uberXL Car',
       '2.81 miles   00:15:26 Trip time   uberX Car',
       '0.92 miles   00:19:44 Trip time   Pool Car',
       '24.39 miles   00:43:28 Trip time   uberX Car',
       '10.36 miles   00:25:59 Trip time   uberX Car',
       '9.14 miles   00:36:11 Trip time   POOL Car',
       '1.77 miles   00:06:42 Trip time   uberX Car',
       '5.90 miles   00:11:24 Trip time   uberX Car',
       '0.41 miles   00:02:56 Trip time   UberX Car',
       '1.71 miles   00:08:00 Trip time   POOL Car',
       '2.10 miles   00:05:46 Trip time   uberX Car',
       '1.33 miles   00:15:44 Trip time   uberXL Car',
       '18.19 miles   00:25:16 Trip time   XL Car',
       '5.81 miles   00:12:21 Trip time   uberX Car',
       '6.50 miles   00:32:00 Trip time   uberX Car',
       '1.52 miles   00:06:33 Trip time   UberX Car',
       '3.58 miles   00:12:06 Trip time   POOL Car',
       '19.22 miles   01:03:32 Trip time   uberX Car',
       '11.82 miles   00:44:02 Trip time   uberX Car',
       '1.78 miles   00:07:32 Trip time   uberX Car',
       '17.48 miles   00:21:01 Trip time   uberX Car',
       '1.58 miles   00:08:20 Trip time   UberX Car',
       '7.04 miles   00:12:07 Trip time   UberX Car',
       '0.54 miles   00:04:50 Trip time   uberX Car',
       '3.59 miles   00:11:37 Trip time   uberX Car',
       '1.81 miles   00:09:35 Trip time   POOL Car',
       '23.74 miles   00:35:47 Trip time   Select Car',
       '3.88 miles   00:09:22 Trip time   uberX Car',
       '1.16 miles   00:05:18 Trip time   uberX Car',
       '7.79 miles   00:21:01 Trip time   UberX Car',
       '2.94 miles   00:06:58 Trip time   UberX Car',
       '1.21 miles   00:04:29 Trip time   uberX Car',
       '1.51 miles   00:07:08 Trip time   uberX Car',
       '3.80 miles   00:28:04 Trip time   POOL Car',
       '6.28 miles   00:16:43 Trip time   POOL Car',
       '7.01 miles   00:18:55 Trip time   uberX Car',
       '2.03 miles   00:10:56 Trip time   POOL Car',
       '9.40 miles   00:13:28 Trip time   uberX Car',
       '3.65 miles   00:12:09 Trip time   uberX Car',
       '8.85 miles   00:18:25 Trip time   POOL Car',
       '8.27 miles   00:14:10 Trip time   uberX Car',
       '1.79 miles   00:09:15 Trip time   uberX Car',
       '10.29 miles   00:30:13 Trip time   uberX Car',
       '17.39 miles   00:36:22 Trip time   uberX Car',
       '12.99 miles   00:18:58 Trip time   uberX Car',
       '9.69 miles   00:18:50 Trip time   uberX Car',
       '0.52 miles   00:04:48 Trip time   uberXL Car',
       '1.12 miles   00:10:25 Trip time   uberX Car',
       '3.35 miles   00:13:29 Trip time   uberX Car',
       '1.18 miles   00:09:13 Trip time   uberX Car',
       '4.61 miles   00:09:40 Trip time   XL Car',
       '49.26 miles   00:56:19 Trip time   uberX Car',
       '7.11 miles   00:14:47 Trip time   uberX Car',
       '9.55 miles   00:32:41 Trip time   uberX Car',
       '2.74 miles   00:16:56 Trip time   Express POOL Car',
       '2.48 miles   00:12:41 Trip time   uberX Car',
       '10.61 miles   00:14:09 Trip time   uberX Car',
       '17.22 miles   00:18:17 Trip time   uberX Car',
       '15.87 miles   00:45:15 Trip time   UberX Car',
       '25.70 miles   00:49:13 Trip time   POOL Car',
       '4.25 miles   00:26:24 Trip time   UberRUSH Car',
       '3.93 miles   00:18:27 Trip time   uberX Car',
       '0.63 miles   00:03:20 Trip time   uberX Car',
       '4.05 miles   00:13:01 Trip time   POOL Car',
       '2.11 miles   00:05:05 Trip time   UberX Car',
       '3.51 miles   00:14:45 Trip time   uberX Car',
       '17.08 miles   00:22:54 Trip time   uberX Car',
       '2.44 miles   00:10:37 Trip time   uberX Car',
       '21.57 miles   00:24:34 Trip time   BLACK Car',
       '1.55 miles   00:10:47 Trip time   uberX Car',
       '1.07 miles   00:13:03 Trip time   Pool Car',
       '2.64 miles   00:11:37 Trip time   uberXL Car',
       '3.24 miles   00:06:27 Trip time   uberX Car',
       '9.87 miles   00:13:23 Trip time   uberX Car',
       '8.14 miles   00:12:07 Trip time   uberX Car',
       '0.91 miles   00:04:08 Trip time   UberX Car',
       '2.06 miles   00:07:36 Trip time   uberX Car',
       '1.71 miles   00:25:20 Trip time   uberPOOL Car',
       '2.71 miles   00:09:29 Trip time   uberX Car',
       '1.16 miles   00:04:31 Trip time   uberX Car',
       '7.56 miles   00:15:58 Trip time   uberX Car',
       '7.27 miles   00:15:02 Trip time   uberX Car',
       '2.87 miles   00:07:20 Trip time   uberX Car',
       '3.50 miles   00:16:26 Trip time   uberX Car',
       '2.87 miles   00:09:50 Trip time   uberX Car',
       '1.16 miles   00:04:51 Trip time   uberX Car',
       '8.47 miles   00:18:34 Trip time   UberX Car',
       '14.52 miles   00:27:51 Trip time   uberX Car',
       '11.12 miles   00:29:03 Trip time   uberX Car',
       '4.49 miles   00:31:52 Trip time   Pool Car',
       '3.02 miles   00:06:40 Trip time   uberX Car',
       '2.77 miles   00:11:11 Trip time   uberX Car',
       '14.18 miles   00:27:35 Trip time   uberX Car',
       '2.52 miles   00:07:56 Trip time   uberX Car',
       '6.64 miles   00:27:38 Trip time   UberX Car',
       '8.02 miles   00:42:52 Trip time   uberXL Car',
       '12.51 miles   00:19:58 Trip time   UberX Car',
       '4.86 miles   00:15:50 Trip time   uberX Car',
       '1.42 miles   00:05:42 Trip time   uberX VIP Car',
       '1.13 miles   00:08:31 Trip time   uberX VIP Car',
       '3.09 miles   00:31:56 Trip time   POOL Car',
       '2.49 miles   00:12:41 Trip time   POOL Car',
       '10.04 miles   00:32:15 Trip time   uberX Car',
       '18.13 miles   00:35:17 Trip time   uberX Car',
       '1.99 miles   00:11:56 Trip time   POOL Car',
       '4.50 miles   00:17:49 Trip time   POOL Car',
       '7.22 miles   00:26:29 Trip time   POOL Car',
       '2.88 miles   00:12:57 Trip time   POOL Car',
       '1.95 miles   00:07:11 Trip time   uberX Car',
       '3.55 miles   00:19:35 Trip time   POOL Car',
       '3.45 miles   00:06:45 Trip time   Pool Car',
       '3.53 miles   00:11:46 Trip time   uberX Car',
       '13.31 miles   00:18:53 Trip time   POOL Car',
       '0.86 miles   00:05:34 Trip time   Express Pool Car',
       '5.15 miles   00:13:07 Trip time   UberX Car',
       '5.46 miles   00:12:52 Trip time   uberX Car',
       '5.11 miles   00:14:49 Trip time   uberX Car',
       '2.52 miles   00:10:33 Trip time   UberXL Car',
       '1.49 miles   00:09:30 Trip time   uberX Car',
       '3.33 miles   00:31:40 Trip time   Express POOL Car',
       '3.97 miles   00:23:32 Trip time   Express Pool Car',
       '3.32 miles   00:08:26 Trip time   VIP Car',
       '5.04 miles   00:34:23 Trip time   POOL Car',
       '6.03 miles   00:18:06 Trip time   uberX Car',
       '2.23 miles   00:13:40 Trip time   uberX Car',
       '1.37 miles   00:05:01 Trip time   uberX Car',
       '3.04 miles   00:09:46 Trip time   uberX Car',
       '18.83 miles   00:24:41 Trip time   POOL Car',
       …]
```

[69]:

```
#some other formats for the Uber cases.
[x for x in uber.product_description.astype(str) if 'miles' not in x]

#regex: UberX {d}.{d}{d} mi | {d}{d} min
#regex: UberX {d}.{d}{d} mi | {d} min
#regex: UberX {d}{d}.{d}{d} mi | {d}{d} min
#regex: UberX {d}{d}.{d}{d} mi | {d} min
```

[69]:

```
['UberX 0.00 mi | 2 min',
 'UberX 8.37 mi | 12 min',
 'UberX 3.75 mi | 12 min',
 'nan',
 'UberX 2.79 mi | 19 min',
 'UberX 1.97 mi | 5 min',
 'nan',
 'UberX 1.27 mi | 5 min',
 'UberX 0.46 mi | 2 min',
 'UberX 1.81 mi | 11 min',
 'UberX 1.38 mi | 10 min',
 'UberX 33.68 mi | 49 min',
 'UberX 23.18 mi | 27 min',
 'UberX 1.65 mi | 12 min',
 'UberX 0.95 mi | 6 min',
 'UberX 22.08 mi | 25 min',
 'UberX 1.47 mi | 4 min',
 'Express Pool 15.79 mi | 21 min',
 'UberX 15.83 mi | 20 min',
 'UberX 24.65 mi | 38 min',
 'nan',
 'UberX 2.44 mi | 9 min',
 'UberX 6.21 mi | 15 min',
 'UberX 1.79 mi | 7 min',
 'UberX 13.79 mi | 27 min',
 'nan',
 'nan',
 'nan',
 'UberX Ride cancelled',
 'UberX Ride cancelled',
 'UberX VIP Ride cancelled',
 'UberX Ride cancelled',
 'UberX Ride cancelled',
 'Black 28.57 mi | 58 min',
 'UberX Ride cancelled',
 'Express Pool 21.46 mi | 32 min',
 'Pool Ride cancelled',
 'UberX 7.61 mi | 16 min',
 'UberX 8.19 mi | 14 min',
 '10.27 kilometers   00:15:48 Trip time   uberX Car',
 'UberX 12.69 km | 16 min',
 'UberX 11.35 km | 23 min',
 'UberX 6.02 mi | 18 min',
 '1.27 kilometers   00:05:35 Trip time   UberX Car',
 'UberX 23.63 mi | 1 h 21 min',
 'UberX 17.38 mi | 33 min',
 'UberX 6.70 km | 21 min',
 '14.95 kilometers   01:23:01 Trip time   uberX Car',
 'UberX 8.31 mi | 23 min',
 'UberX 2.80 mi | 12 min',
 'UberX 0.89 mi | 3 min',
 'UberX 9.32 mi | 24 min',
 'UberX 18.98 mi | 28 min',
 'UberX 1.07 mi | 5 min',
 'Black 13.44 mi | 22 min',
 'UberX 1.70 mi | 5 min',
 'Express Pool 8.61 mi | 32 min',
 'UberX 4.52 mi | 11 min',
 'UberX 1.99 mi | 8 min',
 'WAV 4.18 mi | 1 h 2 min',
 'Comfort 5.71 mi | 31 min',
 'UberX 2.49 mi | 9 min',
 'UberX 0.96 mi | 4 min',
 'UberX 1.71 mi | 6 min',
 'UberX 1.40 mi | 15 min',
 'UberX Diamond 15.55 mi | 33 min',
 'UberX 5.98 mi | 11 min',
 'UberX 3.13 mi | 12 min',
 'UberX Diamond 13.69 mi | 27 min',
 'UberX Diamond 19.33 mi | 41 min',
 'UberX 6.65 mi | 15 min',
 'UberX 2.42 mi | 16 min',
 'UberX 24.94 mi | 34 min',
 'UberX 1.52 mi | 7 min',
 'UberX 3.15 mi | 9 min',
 'UberX 65.44 mi | 1 h 11 min',
 'UberX 9.12 mi | 26 min',
 'UberX 4.65 mi | 13 min',
 'UberX 6.78 mi | 24 min',
 'UberX 9.68 mi | 30 min',
 'UberX 13.09 mi | 29 min',
 'Select 39.08 mi | 47 min',
 'UberX 2.00 mi | 8 min',
 'UberX 13.31 mi | 26 min',
 'UberX 29.06 mi | 36 min',
 'Select 27.39 mi | 1 h 42 min',
 'UberX 13.42 mi | 20 min',
 'UberX 3.42 mi | 12 min',
 'UberX 3.33 mi | 13 min',
 'Express Pool 16.63 mi | 46 min',
 'UberX 13.02 mi | 17 min',
 'UberX 5.01 mi | 11 min',
 'UberX 5.80 mi | 9 min',
 'UberX Ride cancelled',
 'UberX 1.47 mi | 5 min',
 'UberX 6.74 mi | 19 min',
 'UberX 6.12 mi | 15 min',
 'UberX 25.49 mi | 51 min',
 'UberX 1.34 mi | 4 min',
 'Comfort 9.19 mi | 17 min',
 'UberX 18.20 mi | 27 min',
 'Express Pool 0.71 mi | 5 min',
 'UberX 1.07 mi | 13 min',
 'UberX 9.27 mi | 16 min',
 'Pool 22.55 mi | 34 min',
 'UberX 38.43 mi | 50 min',
 'UberX Ride cancelled',
 'UberXL 4.68 mi | 9 min',
 'UberX Ride cancelled',
 'UberX 1.74 mi | 6 min',
 'UberX 6.32 mi | 15 min',
 'uberX 1.01 mi | 5 min',
 'UberX 38.60 mi | 57 min',
 'UberX 4.08 mi | 15 min',
 'Black 0.55 mi | 4 min',
 'UberX 19.31 mi | 44 min',
 'Black 4.23 mi | 25 min',
 'UberX 2.23 mi | 6 min',
 'UberX 6.12 mi | 22 min',
 'UberX 5.75 mi | 33 min',
 'UberX 2.43 mi | 17 min',
 'UberX 3.15 mi | 10 min',
 'UberX 16.39 mi | 38 min',
 'Comfort 44.53 mi | 1 h 15 min',
 'UberXL 1.03 mi | 4 min',
 'UberX 1.08 mi | 4 min',
 'UberX 6.19 mi | 23 min',
 'UberX 2.80 mi | 9 min',
 'UberX 0.33 mi | 2 min',
 'UberX 3.00 mi | 8 min',
 'UberX 2.93 mi | 7 min',
 'UberXL 4.97 mi | 12 min',
 'UberXL 0.94 mi | 6 min',
 'UberX 27.40 mi | 46 min',
 'UberX 2.29 mi | 6 min',
 'UberX 12.67 mi | 34 min',
 'UberX 10.77 mi | 19 min',
 'UberX 2.71 mi | 11 min',
 'UberX 1.99 mi | 9 min',
 'Black 4.03 mi | 13 min',
 'Express Pool 10.81 mi | 31 min',
 'UberX 4.66 mi | 14 min',
 'UberX 19.21 mi | 22 min',
 'UberX 11.26 mi | 22 min',
 'UberX 0.64 mi | 2 min',
 'UberX 8.32 mi | 22 min',
 'UberX 3.89 mi | 11 min',
 'UberX 2.52 mi | 19 min',
 'UberX 3.33 mi | 13 min',
 'UberX 1.21 mi | 5 min',
 'UberX 21.62 mi | 27 min',
 'UberX 1.03 mi | 9 min',
 'Express Pool 0.80 mi | 3 min',
 'UberX 1.22 mi | 3 min',
 'UberX 5.81 mi | 10 min',
 'UberX 8.14 mi | 16 min',
 'UberX 5.45 mi | 12 min',
 'UberX 3.12 mi | 7 min',
 'UberX 3.25 mi | 5 min',
 'UberX 3.47 mi | 8 min',
 'UberX 2.17 mi | 11 min',
 'UberX 0.78 mi | 4 min',
 'UberX 14.03 mi | 24 min',
 'UberX 1.69 mi | 8 min',
 'UberX 1.22 mi | 6 min',
 'UberX 7.58 mi | 23 min',
 'UberX 3.46 mi | 12 min',
 'UberX 0.63 mi | 6 min',
 'UberX 3.04 mi | 11 min',
 'UberX 10.21 mi | 20 min',
 'UberX 4.85 mi | 19 min',
 'UberX 3.22 mi | 9 min',
 'UberX 2.56 mi | 8 min',
 'UberX 18.75 mi | 24 min',
 'UberX 0.81 mi | 2 min',
 'UberX 6.56 mi | 20 min',
 'UberX 2.06 mi | 10 min',
 'UberX 25.77 mi | 32 min',
 'Express Pool 2.21 mi | 22 min',
 'Express Pool 2.51 mi | 10 min',
 'UberX 12.97 mi | 22 min',
 'UberX 22.80 mi | 27 min',
 'UberX 6.56 mi | 26 min',
 'UberX 10.94 mi | 20 min',
 'UberX 9.56 mi | 19 min',
 'UberX 6.47 mi | 11 min',
 'UberX 5.90 mi | 12 min',
 'WAV 1.75 mi | 7 min',
 'UberX 14.05 mi | 19 min',
 'nan',
 'nan',
 'nan',
 'nan',
 'UberX Ride cancelled',
 'Lux Ride cancelled',
 'Car Seat 2.67 mi | 19 min',
 'UberX 10.87 mi | 21 min',
 'UberX 1.65 mi | 6 min',
 'UberX 4.82 mi | 23 min',
 'UberXL 7.15 mi | 8 min',
 'Pool 3.43 mi | 9 min',
 'UberX 11.03 mi | 25 min',
 'Select 0.83 mi | 3 min',
 'UberX 8.18 mi | 18 min',
 'Black 36.45 mi | 55 min',
 'UberX 28.49 mi | 35 min',
 'UberX 6.44 mi | 15 min',
 'UberX 13.24 mi | 18 min',
 'UberX 0.60 mi | 4 min',
 'UberX 5.66 mi | 15 min',
 'UberX 2.20 mi | 5 min',
 'UberX 1.38 mi | 4 min',
 'UberX 6.93 mi | 18 min',
 'UberX 7.13 mi | 25 min',
 'UberX 13.23 mi | 22 min',
 'UberX 15.39 mi | 20 min',
 'UberX 1.50 mi | 15 min',
 'UberX 6.44 mi | 15 min',
 'UberX 2.53 mi | 10 min',
 'UberXL 15.44 mi | 25 min',
 'UberXL 3.09 mi | 17 min',
 'UberX 2.85 mi | 12 min',
 'UberX 2.51 mi | 11 min',
 'UberX 1.37 mi | 7 min',
 'UberX 8.96 mi | 11 min',
 'Comfort 1.97 mi | 8 min',
 'UberX 1.45 mi | 3 min',
 'UberX 2.45 mi | 10 min',
 'UberX 4.38 mi | 9 min',
 'UberX 5.89 mi | 12 min',
 'UberX 3.61 mi | 16 min',
 'UberX 2.11 mi | 7 min',
 'UberX 1.17 mi | 3 min',
 'UberX 3.43 mi | 9 min',
 'UberX 5.65 mi | 14 min',
 'Black 1.96 mi | 15 min',
 'UberX 1.93 mi | 6 min',
 'UberX 1.41 mi | 7 min',
 'Base Fare $0.00 + $0.15 / minute  Duration Fare 00:01:38',
 'UberX 0.74 mi | 4 min',
 'UberX VIP 1.60 mi | 7 min',
 'UberX 2.50 mi | 20 min',
 'UberX 15.64 mi | 24 min',
 'UberXL 0.71 mi | 14 min',
 'UberX 8.63 mi | 21 min',
 'UberX VIP 2.54 mi | 14 min',
 'UberX 3.00 mi | 6 min',
 'UberX 20.51 mi | 32 min',
 'UberX 2.00 mi | 5 min',
 'UberX 14.58 mi | 20 min',
 'UberXL 1.03 mi | 8 min',
 'UberX 0.95 mi | 5 min',
 'Express Pool 10.15 mi | 42 min',
 'UberXL 0.60 mi | 3 min',
 'Express Pool 10.01 mi | 18 min',
 'UberX 5.32 mi | 16 min',
 'Pool 5.52 mi | 15 min',
 'UberX 3.70 mi | 13 min',
 'UberXL 0.24 mi | 2 min',
 'UberX 2.27 mi | 12 min',
 'UberX 2.91 mi | 10 min',
 'UberX 1.41 mi | 5 min',
 'UberX 8.19 mi | 17 min',
 'UberX 14.21 mi | 33 min',
 'UberX 2.06 mi | 5 min',
 'Pool 4.50 mi | 27 min',
 'UberX 0.74 mi | 4 min',
 'Pool 5.06 mi | 12 min',
 'Pool 9.07 mi | 31 min',
 'UberXL 19.42 mi | 32 min',
 'UberX 3.34 mi | 16 min',
 'UberX 3.66 km | 9 min',
 '3.04 kilometers   00:09:56 Trip time   uberX Car   Toyota Make',
 'UberX 9.16 km | 15 min',
 'UberX 2.83 mi | 10 min',
 'Base Fare $1.00 + $0.26 / minute  Duration Fare 00:23:20',
 'UberXL 6.54 mi | 31 min',
 'Black VIP 2.72 mi | 11 min',
 'UberX 4.29 mi | 15 min',
 'UberX 7.04 mi | 22 min',
 'UberX 1.45 mi | 8 min',
 'UberX 1.20 mi | 11 min',
 'Comfort 2.31 mi | 14 min',
 'UberX 0.83 mi | 5 min',
 'UberXL 8.74 mi | 22 min',
 'UberX 2.55 mi | 11 min',
 'UberX 9.44 mi | 33 min',
 'UberX 0.04 mi | 3 min',
 'UberX VIP 3.21 mi | 19 min',
 'Pool 3.64 mi | 9 min',
 'UberX 4.11 mi | 16 min',
 'UberX 29.19 mi | 28 min',
 'UberX 15.58 mi | 33 min',
 'UberX 7.44 mi | 12 min',
 'UberX 1.73 mi | 6 min',
 'UberX 4.83 mi | 12 min',
 'UberX 9.48 mi | 19 min',
 'UberX 1.41 mi | 7 min',
 'Pool 4.23 mi | 18 min',
 'UberX 1.34 mi | 7 min',
 'UberX 1.41 mi | 8 min',
 'UberX 18.67 mi | 1 h 20 min',
 'UberX 0.89 mi | 7 min',
 'UberX 2.77 mi | 13 min',
 'UberX VIP 2.06 mi | 13 min',
 'UberX 11.87 mi | 21 min',
 'UberX 2.65 mi | 11 min',
 'UberX 1.42 mi | 6 min',
 'UberX 5.50 mi | 14 min',
 'UberX 1.29 mi | 7 min',
 'UberX 15.70 mi | 26 min',
 'UberX 1.44 mi | 6 min',
 'Express Pool 1.75 mi | 10 min',
 'UberX 9.85 mi | 15 min',
 'UberX 10.89 mi | 13 min',
 'Lux 6.30 mi | 23 min',
 'UberX 13.30 mi | 19 min',
 'UberX 1.49 mi | 20 min',
 'UberX 2.88 mi | 5 min',
 'UberX 49.23 mi | 1 h 33 min',
 'UberX 3.39 mi | 13 min',
 'UberX 17.37 mi | 25 min',
 'UberX 4.26 mi | 17 min',
 'Express Pool 8.67 mi | 41 min',
 'UberX 3.19 mi | 19 min',
 'Pool 1.62 mi | 5 min',
 'UberX 2.47 mi | 18 min',
 'UberX 14.40 mi | 17 min',
 'UberX 0.48 mi | 5 min',
 'UberX 24.95 mi | 33 min',
 'UberX 5.89 mi | 48 min',
 'UberX Diamond 16.55 mi | 40 min',
 'UberX 2.03 mi | 17 min',
 'UberX 3.21 mi | 12 min',
 'UberX 4.13 mi | 20 min',
 'Lux 10.63 mi | 21 min',
 'UberX 1.48 mi | 15 min',
 'UberX 3.43 mi | 10 min',
 'UberX 0.58 mi | 3 min',
 'UberX 16.69 mi | 25 min',
 'UberX 7.54 mi | 17 min',
 'UberX 9.74 mi | 55 min',
 'UberX 1.10 mi | 10 min',
 'nan',
 'Base Fare $1.00 for 15 minutes + $0.07 / minute  Duration Fare 00:57:38',
 'UberX 1.54 mi | 7 min',
 'UberXL 0.94 mi | 9 min',
 'UberX 4.28 mi | 11 min',
 'UberX 0.88 mi | 5 min',
 'UberX VIP 16.86 mi | 1 h 35 min',
 'UberX 5.07 mi | 25 min',
 'Pool 7.93 mi | 23 min',
 'UberX 0.60 mi | 4 min',
 'UberX 1.41 mi | 20 min',
 'UberX 1.72 mi | 8 min',
 'UberXL 4.81 mi | 17 min',
 '5.25 kilometers   00:12:10 Trip time   uberX Car',
 'UberX 25.51 mi | 34 min',
 'UberX 2.66 mi | 23 min',
 'UberX 10.80 mi | 14 min',
 'UberX 14.55 mi | 29 min',
 'UberX 2.66 mi | 16 min',
 'UberX 4.91 mi | 18 min',
 'UberX 3.25 mi | 17 min',
 'UberXL 1.19 mi | 6 min',
 'UberX 4.99 mi | 10 min',
 'UberX 6.57 mi | 31 min',
 'Express Pool 27.52 mi | 38 min',
 'UberX 3.61 mi | 14 min',
 'UberX 40.63 mi | 1 h 12 min',
 'UberX 1.88 mi | 7 min',
 'UberX 1.74 mi | 4 min',
 'UberX 0.72 mi | 2 min',
 'UberX 8.54 mi | 18 min',
 'VIP 4.14 mi | 16 min',
 'UberXL 1.52 mi | 15 min',
 'UberX 11.43 mi | 20 min',
 'UberX 4.44 mi | 10 min',
 'Express Pool 9.66 mi | 42 min',
 'UberX 1.60 mi | 7 min',
 'UberX 5.15 mi | 16 min',
 'UberX 1.20 mi | 3 min',
 'Express Pool 8.93 mi | 24 min',
 'UberX 1.67 mi | 8 min',
 'UberX 0.86 mi | 7 min',
 'UberX 10.38 mi | 30 min',
 'UberX 3.09 mi | 11 min',
 'UberX 10.03 mi | 13 min',
 'UberX 3.67 mi | 11 min',
 'UberX 18.32 mi | 21 min',
 'UberX 1.43 mi | 5 min',
 'UberX 6.19 mi | 17 min',
 'Pool 1.80 mi | 11 min',
 'UberX 5.95 mi | 11 min',
 'UberX 11.56 mi | 16 min',
 'UberX 0.88 mi | 2 min',
 'nan',
 'UberX 63.16 mi | 1 h 49 min',
 'Black SUV 1.39 mi | 17 min',
 'Base Fare $3.00 for 20 minutes + $0.15 / minute  Duration Fare 00:13:00',
 'UberX 1.69 mi | 11 min',
 'UberX 1.42 mi | 15 min',
 'UberX 2.29 mi | 13 min',
 'UberX 26.10 mi | 1 h 0 min',
 'UberX 1.64 mi | 7 min',
 'UberX 0.95 mi | 4 min',
 'UberX 1.92 mi | 5 min',
 'Select 9.80 mi | 24 min',
 'UberX 1.57 mi | 7 min',
 'UberX 3.27 mi | 8 min',
 'UberX 11.17 mi | 21 min',
 'UberX 9.26 mi | 17 min',
 'UberX 3.27 mi | 11 min',
 'UberX 5.22 mi | 9 min',
 'UberX 13.01 mi | 31 min',
 'UberX 2.07 mi | 17 min',
 'UberX 8.19 mi | 16 min',
 'UberX 1.94 mi | 13 min',
 'Express Pool 2.41 mi | 11 min',
 'UberX 5.87 mi | 17 min',
 'UberX 4.76 mi | 16 min',
 'UberX 25.27 mi | 36 min',
 'UberX 3.25 mi | 20 min',
 'Select 1.72 mi | 15 min',
 'UberX 4.87 mi | 11 min',
 'UberX 2.37 mi | 8 min',
 'UberX 1.19 mi | 5 min',
 'UberX 2.81 mi | 9 min',
 'UberX 3.69 mi | 13 min',
 'UberX 3.63 mi | 9 min',
 'UberX 10.74 mi | 28 min',
 'UberX 2.74 mi | 8 min',
 'UberX 4.77 mi | 13 min',
 'UberX 5.93 mi | 18 min',
 'UberX 17.30 mi | 21 min',
 'UberX 36.52 mi | 50 min',
 'UberX 2.56 mi | 11 min',
 'UberX 1.30 mi | 15 min',
 'UberX 1.50 mi | 5 min',
 'UberX 2.91 mi | 6 min',
 'UberX 11.46 mi | 51 min',
 'Express Pool 3.77 mi | 18 min',
 'UberX 9.74 mi | 28 min',
 'UberX 28.70 mi | 37 min',
 'UberX 4.40 mi | 18 min',
 'UberX 3.85 mi | 10 min',
 'Express Pool 2.65 mi | 19 min',
 'UberX 1.44 mi | 12 min',
 'UberX 6.35 mi | 11 min',
 'UberX 2.14 mi | 7 min',
 'UberX 9.23 mi | 18 min',
 'UberX 2.28 mi | 5 min',
 'UberX 0.69 mi | 3 min',
 'UberX 16.84 mi | 26 min',
 'UberX 3.97 mi | 12 min',
 'Express Pool 8.13 mi | 32 min',
 'Pool 0.30 mi | 1 min',
 'UberX 76.07 mi | 1 h 18 min',
 'UberX 4.17 mi | 8 min',
 'Pool 2.85 mi | 12 min',
 'Express Pool 1.51 mi | 5 min',
 'UberX 3.28 mi | 7 min',
 'UberX 1.94 mi | 9 min',
 'UberX 2.20 mi | 6 min',
 'UberX 2.81 mi | 8 min',
 'UberX 1.82 mi | 6 min',
 'UberX 12.49 mi | 20 min',
 'Comfort 2.21 mi | 7 min',
 'UberX 1.69 mi | 6 min',
 'UberX 4.40 mi | 19 min',
 'Pool 3.37 mi | 7 min',
 'UberX 1.84 mi | 15 min',
 'UberX 0.94 mi | 6 min',
 'nan',
 'nan',
 'nan',
 'nan',
 'nan',
 'nan',
 'UberX 2.86 mi | 8 min',
 'UberX 7.02 mi | 19 min',
 'UberX 3.61 mi | 16 min',
 'UberX 1.47 mi | 8 min',
 'UberXL 4.22 mi | 12 min',
 'Pool 1.18 mi | 4 min',
 'UberX 5.81 mi | 16 min',
 'UberX 33.32 mi | 47 min',
 'UberX 1.14 mi | 5 min',
 'UberX 2.79 mi | 13 min',
 'Express Pool 6.34 mi | 11 min',
 'Express Pool 0.94 mi | 2 min',
 'UberX 1.88 mi | 6 min',
 'UberXL 8.25 mi | 23 min',
 'UberX 3.78 mi | 10 min',
 'UberX 14.98 mi | 23 min',
 'Express Pool 3.72 mi | 15 min',
 'UberX Diamond 10.03 mi | 25 min',
 'Express Pool 2.39 mi | 11 min',
 'Select 3.86 mi | 13 min',
 'UberX 4.20 mi | 14 min',
 'UberXL 22.56 mi | 32 min',
 'UberX 0.34 mi | 1 min',
 'nan',
 'UberX 14.79 mi | 25 min',
 'UberX 11.90 mi | 17 min',
 'Base Fare $0.00 + $0.30 / minute  Duration Fare 00:03:30',
 'Pool 0.91 mi | 7 min',
 'UberX 56.60 mi | 53 min',
 'UberX 3.36 mi | 11 min',
 'Pool 14.64 mi | 45 min',
 'UberX 6.10 mi | 15 min',
 'UberX 3.47 mi | 23 min',
 'Base Fare $1.00 + $0.26 / minute  Duration Fare 00:04:02',
 'Black VIP 3.84 mi | 27 min',
 'UberX 14.79 mi | 24 min',
 'UberX 2.21 mi | 8 min',
 'UberX VIP 1.01 mi | 6 min',
 'Pool 6.02 mi | 24 min',
 'Pool 0.74 mi | 4 min',
 'Base Fare $0.00  Duration Fare $3.30',
 'UberX 17.60 mi | 1 h 14 min',
 'Black 1.59 mi | 17 min',
 'UberX 0.58 mi | 11 min',
 'UberX 6.57 mi | 15 min',
 'UberX 5.82 mi | 13 min',
 'UberX VIP 25.86 mi | 39 min',
 'UberX 1.12 mi | 10 min',
 'UberX 1.17 mi | 8 min',
 'UberX 8.59 mi | 42 min',
 'Base Fare $0.00 + $0.15 / minute  Duration Fare 00:06:13',
 'UberX 2.44 mi | 22 min',
 'Pool 2.41 mi | 11 min',
 'UberX VIP 13.71 mi | 28 min',
 'UberX 1.52 mi | 13 min',
 'UberX 5.12 mi | 24 min',
 'UberX 11.40 mi | 31 min',
 'UberX 18.55 mi | 1 h 14 min',
 'UberX VIP 1.03 mi | 7 min',
 'UberX 1.49 mi | 12 min',
 'UberX 2.33 mi | 16 min',
 'UberX 4.31 mi | 20 min',
 'UberX 2.68 mi | 22 min',
 'Pool 2.66 mi | 21 min',
 'UberX Ride cancelled',
 'UberX 3.97 mi | 12 min',
 'Black 6.03 km | 13 min',
 'UberX 10.48 mi | 17 min',
 '20.00 kilometers   00:17:10 Trip time   UberFLASH Car',
 'Select 1.08 km | 1 min',
 'UberX 5.26 mi | 11 min',
 'UberX 1.74 mi | 11 min',
 'UberX 2.83 mi | 9 min',
 'UberX 31.85 mi | 34 min',
 'UberX 24.73 mi | 37 min',
 'UberX Ride cancelled',
 'UberX 0.88 mi | 6 min',
 'UberX 11.19 mi | 18 min',
 'UberX 2.22 mi | 5 min',
 'UberX 10.81 mi | 26 min',
 'UberX Ride cancelled',
 'Pool 3.50 mi | 8 min',
 'UberX 2.72 mi | 9 min',
 'Black SUV Ride cancelled',
 'UberX Ride cancelled',
 'UberX 4.55 mi | 10 min',
 'UberX 13.74 mi | 19 min',
 'UberX 8.81 mi | 33 min',
 'Select 24.98 mi | 33 min',
 'UberX 2.28 mi | 7 min',
 'Express Pool 13.32 mi | 33 min',
 'UberXL Ride cancelled',
 'UberX 2.26 mi | 8 min',
 'Premier SUV 0.70 mi | 5 min',
 'UberX 5.31 mi | 13 min',
 'UberX 4.40 mi | 19 min',
 'UberX 18.87 mi | 46 min',
 'UberX 3.79 mi | 21 min',
 'Express Pool Ride cancelled',
 'UberX 5.30 mi | 13 min',
 'UberX 13.83 mi | 18 min',
 'UberX 26.68 mi | 40 min',
 'UberX 3.84 mi | 11 min',
 'Express Pool 3.25 mi | 12 min',
 'UberX 3.55 mi | 20 min',
 'Pool 6.30 mi | 16 min',
 'UberX 29.90 mi | 38 min',
 'UberX 4.27 mi | 19 min',
 'Black SUV 18.02 mi | 58 min',
 'UberX 10.48 mi | 22 min',
 'UberX 0.33 mi | 3 min',
 'UberX 1.15 mi | 4 min',
 'UberX 8.06 mi | 19 min',
 'UberX 1.41 mi | 5 min',
 'UberX 1.79 mi | 7 min',
 'UberX 1.08 mi | 5 min',
 'UberX Diamond 4.93 mi | 15 min',
 'UberX 1.03 mi | 6 min',
 'UberX 6.10 mi | 10 min',
 'UberX 16.17 mi | 22 min',
 'UberX 2.32 mi | 11 min',
 'UberX 3.47 mi | 8 min',
 'UberX 3.27 mi | 10 min',
 'UberX 18.24 mi | 40 min',
 'UberX 2.19 mi | 6 min',
 'UberX 7.81 mi | 16 min',
 'Lux 4.11 mi | 23 min',
 'UberX 15.56 mi | 38 min',
 'UberX 10.47 mi | 15 min',
 'UberX Ride cancelled',
 'UberX 6.21 mi | 18 min',
 'UberX 6.01 mi | 15 min',
 'UberX 3.09 mi | 7 min',
 'UberX 4.02 mi | 11 min',
 'UberX 4.97 mi | 11 min',
 'UberX 1.69 mi | 8 min',
 'UberX 2.09 mi | 14 min',
 'UberX 1.11 mi | 13 min',
 'UberX 1.32 mi | 5 min',
 'UberX 18.26 mi | 25 min',
 'Express Pool 4.56 mi | 31 min',
 'UberX 20.61 mi | 36 min',
 'Black SUV 1.67 mi | 10 min',
 'UberX Ride cancelled',
 'UberX 12.49 mi | 29 min',
 'UberX 22.61 mi | 31 min',
 'UberX Diamond 1.03 mi | 7 min',
 'UberX 2.06 mi | 6 min',
 'UberX 3.23 mi | 8 min',
 'Select 1.93 mi | 6 min',
 'UberX 13.18 mi | 40 min',
 'UberX 2.42 mi | 9 min',
 'UberX 7.49 mi | 38 min',
 'UberX 3.12 mi | 10 min',
 'Select 7.27 mi | 17 min',
 'UberX Ride cancelled',
 'UberX 2.25 mi | 6 min',
 'UberXL 4.23 mi | 12 min',
 'UberX 4.39 mi | 8 min',
 'UberX 1.70 mi | 5 min',
 'UberX 1.73 mi | 4 min',
 'UberX 1.55 mi | 10 min',
 'Black 18.52 mi | 1 h 10 min',
 'UberX 2.99 mi | 18 min',
 'UberX 9.18 mi | 14 min',
 'UberX 28.54 mi | 34 min',
 'UberX 7.54 mi | 15 min',
 'Comfort 21.82 mi | 38 min',
 'UberXL 0.47 mi | 3 min',
 'UberX 9.44 mi | 28 min',
 'Pool 2.26 mi | 7 min',
 'UberX 3.14 mi | 25 min',
 'UberX 12.69 mi | 23 min',
 'UberX 16.52 mi | 24 min',
 'UberX 0.84 mi | 5 min',
 'UberX 1.54 mi | 9 min',
 'UberX 10.28 mi | 20 min',
 'UberX 20.80 mi | 24 min',
 'UberX 2.50 mi | 26 min',
 'Pool 12.70 mi | 32 min',
 'Express Pool 17.05 mi | 22 min',
 'UberX 0.63 mi | 4 min',
 'UberX 1.86 mi | 6 min',
 'UberX 0.71 mi | 4 min',
 'UberX 2.73 mi | 8 min',
 'UberX 13.17 mi | 36 min',
 'UberX 19.54 mi | 24 min',
 'UberX 13.22 mi | 19 min',
 'UberX 4.14 mi | 33 min',
 'UberX 2.24 mi | 5 min',
 'Express Pool 3.14 mi | 26 min',
 'uberX 7.41 mi | 16 min',
 'Select 0.91 mi | 4 min',
 'UberX 28.69 mi | 33 min',
 'Lux 30.20 mi | 38 min',
 'UberX 2.89 mi | 12 min',
 'UberX 1.36 mi | 6 min',
 'UberX 2.00 mi | 6 min',
 'UberX 2.49 mi | 8 min',
 'UberXL 4.78 mi | 12 min',
 'UberX 10.94 mi | 25 min',
 'UberX 5.82 mi | 11 min',
 'Select 8.24 mi | 16 min',
 'UberX 6.10 mi | 14 min',
 'UberX 2.43 mi | 8 min',
 'Comfort 13.09 mi | 16 min',
 'UberX 9.70 mi | 21 min',
 'UberX 1.50 mi | 4 min',
 'UberX 4.54 mi | 13 min',
 'UberX 2.58 mi | 6 min',
 'Black 1.98 mi | 10 min',
 'UberX 1.71 mi | 5 min',
 'UberX 3.63 mi | 9 min',
 'UberX 3.30 mi | 8 min',
 'UberX 5.87 mi | 11 min',
 'UberX 16.65 mi | 18 min',
 'Pool 9.67 mi | 13 min',
 'UberX 20.60 mi | 20 min',
 'UberX 2.33 mi | 4 min',
 'UberX 3.33 mi | 11 min',
 'Express Pool 4.28 mi | 14 min',
 'UberX 1.33 mi | 6 min',
 'UberX 3.75 mi | 22 min',
 'UberX 7.10 mi | 12 min',
 'UberX 16.36 mi | 30 min',
 'UberX 16.64 mi | 24 min',
 'UberX 1.33 mi | 8 min',
 'UberX 14.00 mi | 18 min',
 'Express Pool 7.99 mi | 12 min',
 'UberX 9.29 mi | 16 min',
 'UberX 1.38 mi | 9 min',
 'UberX 1.68 mi | 10 min',
 'Pool 20.06 mi | 54 min',
 'UberX 11.39 mi | 22 min',
 'UberX 1.80 mi | 11 min',
 'UberX 4.49 mi | 14 min',
 'UberX 7.33 mi | 21 min',
 'UberX 3.32 mi | 11 min',
 'UberX 0.20 mi | 5 min',
 'UberX 6.95 mi | 13 min',
 'UberX 7.92 mi | 22 min',
 'UberX 1.31 mi | 6 min',
 'Pool 1.32 mi | 8 min',
 'Express Pool 24.01 mi | 37 min',
 'UberX 2.96 mi | 24 min',
 'UberX 1.44 mi | 5 min',
 'UberX 1.34 mi | 5 min',
 'UberX 1.60 mi | 9 min',
 'Black 5.73 mi | 15 min',
 'UberX 2.40 mi | 12 min',
 'UberX 2.81 mi | 9 min',
 'UberX 1.44 mi | 7 min',
 'UberX 2.98 mi | 7 min',
 'UberX 0.45 mi | 5 min',
 'UberX Diamond 5.10 mi | 13 min',
 'Black 12.09 mi | 27 min',
 'UberX 6.14 mi | 12 min',
 'UberX 3.58 mi | 7 min',
 'UberX 18.06 mi | 28 min',
 'UberX 2.57 mi | 7 min',
 'UberX 0.75 mi | 5 min',
 'UberX 4.74 mi | 13 min',
 'Pool Ride cancelled',
 'UberX Ride cancelled',
 'UberX Ride cancelled',
 'Black VIP Ride cancelled',
 'nan',
 'nan',
 'nan',
 'nan',
 'nan',
 'nan',
 'nan',
 'nan',
 'nan',
 'nan',
 'nan',
 'nan',
 'UberX VIP 2.15 mi | 9 min',
 'UberX Diamond 7.66 mi | 21 min',
 'UberX 0.91 mi | 8 min',
 'UberX 1.07 mi | 6 min',
 'Pool 7.80 mi | 36 min',
 'UberX 1.02 mi | 9 min',
 'UberX 1.46 mi | 10 min',
 'UberX 1.86 mi | 9 min',
 'UberX 1.77 mi | 7 min',
 'Pool 1.67 mi | 12 min',
 'Pool 1.26 mi | 9 min',
 'UberX 2.47 mi | 10 min',
 'UberX Diamond 2.10 mi | 12 min',
 'UberX Diamond 2.13 mi | 27 min',
 'UberX 3.06 mi | 17 min',
 'UberX 7.98 mi | 38 min',
 'UberX 5.98 mi | 16 min',
 'UberX 15.93 mi | 23 min',
 'UberX 2.07 mi | 6 min',
 'UberX 24.99 mi | 47 min',
 'Pool 3.75 mi | 16 min(s)',
 'UberX 2.30 mi | 16 min',
 'Express Pool 25.43 mi | 34 min',
 'UberX 7.23 mi | 18 min',
 'UberX 16.09 mi | 52 min',
 'UberX 1.50 mi | 15 min',
 'UberX 3.61 mi | 11 min',
 'UberX 3.30 mi | 7 min',
 'UberX 2.11 mi | 20 min',
 'UberX 0.53 mi | 2 min',
 'UberX 29.88 mi | 46 min',
 'UberX 2.85 mi | 19 min',
 'Pool 1.49 mi | 7 min',
 'UberX 2.31 mi | 15 min',
 'UberX 26.86 mi | 40 min',
 'UberX 28.05 mi | 1 h 35 min',
 'UberX 8.94 mi | 19 min',
 'UberX 12.68 mi | 35 min',
 'UberX 5.53 mi | 13 min',
 'UberX 27.79 mi | 33 min',
 'UberX 7.56 mi | 14 min',
 'UberXL 3.24 mi | 8 min',
 'UberX 4.21 mi | 12 min',
 'UberX 6.32 mi | 14 min',
 'UberX 1.38 mi | 4 min',
 'UberX 1.65 mi | 13 min',
 'UberX 21.35 mi | 24 min',
 'UberX 1.26 mi | 9 min',
 'UberX 2.14 mi | 11 min',
 'UberX 2.81 mi | 8 min',
 'nan',
 'UberX 1.43 mi | 7 min',
 'UberXL Ride cancelled',
 'UberX 0.47 mi | 4 min',
 'UberX 10.63 mi | 29 min',
 'UberX 42.37 mi | 58 min',
 'UberX 10.92 mi | 27 min',
 'Express Pool 1.73 mi | 15 min',
 'UberX 10.87 mi | 20 min',
 'UberX 1.84 mi | 16 min',
 'Express Pool 2.27 mi | 11 min',
 'UberX 1.18 mi | 10 min',
 'Express Pool 1.77 mi | 11 min',
 'Express Pool 10.82 mi | 23 min',
 'UberX 5.96 mi | 16 min',
 'Black SUV 8.18 mi | 22 min',
 'UberX 2.65 mi | 10 min',
 'UberX Ride cancelled',
 'UberX 1.81 mi | 8 min',
 'Express Pool 11.18 mi | 22 min',
 'Express Pool 13.59 mi | 28 min',
 'Express Pool 5.75 mi | 21 min',
 'UberX 20.12 mi | 22 min',
 'Pool 6.19 mi | 27 min',
 'UberX 7.73 mi | 20 min',
 'Pool 0.40 mi | 2 min',
 'Express Pool 10.90 mi | 17 min',
 'Express Pool 11.83 mi | 24 min',
 'UberXL 15.09 mi | 48 min',
 'UberX 4.26 mi | 19 min',
 'UberX 5.86 mi | 16 min',
 'UberX 0.97 mi | 6 min',
 'UberX 1.48 mi | 14 min',
 'UberX 20.12 mi | 44 min',
 'UberXL Ride cancelled',
 'UberX 1.25 mi | 18 min',
 'UberX 5.64 mi | 17 min',
 'nan',
 'UberX 14.59 mi | 23 min',
 'Base Fare $0.00  Duration Fare $27.30',
 'UberX 1.77 mi | 8 min',
 'Pool 0.98 mi | 4 min',
 'UberX VIP 3.19 mi | 17 min',
 'UberX 2.83 mi | 7 min',
 'UberX 5.11 mi | 25 min',
 'Black 3.21 mi | 20 min',
 'UberXL 1.51 mi | 11 min',
 'UberX 1.47 mi | 12 min',
 'Black VIP 1.99 mi | 20 min',
 'UberX 7.14 km | 24 min',
 'UberX 6.36 mi | 14 min',
 'UberX 1.11 mi | 5 min',
 'UberX 13.14 mi | 50 min',
 'UberX 4.73 mi | 17 min',
 'UberX 6.32 mi | 19 min',
 'UberX 7.66 mi | 20 min',
 'Express Pool 3.78 mi | 18 min',
 'UberX 18.45 mi | 27 min',
 'Black SUV 4.34 mi | 16 min',
 'UberX 7.53 mi | 15 min',
 'Comfort 7.41 mi | 19 min',
 'UberX Ride cancelled',
 'UberXL 3.11 mi | 8 min',
 'UberX 1.52 mi | 11 min',
 'UberX 11.77 mi | 45 min',
 'Select 1.11 mi | 5 min',
 'Pool 1.92 mi | 9 min',
 'UberX 4.27 mi | 14 min',
 'UberX 30.53 mi | 43 min',
 'UberX 7.46 mi | 23 min',
 'UberX 7.09 mi | 22 min',
 'UberX 12.23 mi | 26 min',
 'UberX 2.57 mi | 8 min',
 'UberX 10.35 mi | 26 min',
 'UberX 9.96 mi | 19 min',
 'UberX 3.51 mi | 16 min',
 'UberX 10.28 mi | 24 min',
 'Comfort 20.38 mi | 24 min',
 'UberX 2.50 mi | 7 min',
 'UberX 4.91 mi | 12 min',
 'UberX 1.72 mi | 8 min',
 'Express Pool 1.29 mi | 6 min',
 'UberX 6.47 mi | 25 min',
 'UberX 3.25 mi | 7 min',
 'Pool 1.43 mi | 4 min',
 'UberX 13.67 mi | 22 min',
 'UberX 18.47 mi | 24 min',
 'UberX 9.06 mi | 36 min',
 'UberX 5.73 mi | 21 min',
 'UberX 9.09 mi | 15 min',
 'UberX 13.35 mi | 29 min',
 'UberX 2.76 mi | 7 min',
 'UberX 8.30 mi | 20 min',
 'UberX 2.96 mi | 10 min',
 'UberX 5.93 mi | 12 min',
 'UberX 27.06 mi | 38 min',
 'UberX 2.35 mi | 9 min',
 'UberX 1.39 mi | 6 min',
 'UberX 1.41 mi | 15 min',
 'UberX 24.46 mi | 32 min',
 'UberX 7.57 mi | 22 min',
 'UberX 0.83 mi | 3 min',
 'UberX 11.44 mi | 15 min',
 'UberX 1.94 mi | 8 min',
 'UberX 3.73 mi | 9 min',
 'UberX 0.62 mi | 3 min',
 'UberX 2.79 mi | 8 min',
 'UberX 1.09 mi | 5 min',
 'UberX 3.47 mi | 8 min',
 'UberX 3.32 mi | 11 min',
 'UberX 2.98 mi | 16 min',
 'nan',
 'nan',
 'nan',
 'Car Seat 10.31 mi | 33 min',
 'UberX 0.94 mi | 4 min',
 'UberX 3.11 mi | 27 min',
 'UberX 3.79 mi | 18 min',
 'UberX 2.19 mi | 12 min',
 'UberX 2.53 mi | 5 min',
 'UberX 7.42 mi | 15 min',
 'UberX 3.30 mi | 13 min',
 'UberX 1.34 mi | 6 min',
 'Pool 5.06 mi | 27 min',
 'UberX VIP 4.36 mi | 18 min',
 'UberX 1.90 mi | 8 min',
 'Express Pool 16.62 mi | 22 min',
 'UberXL 11.06 mi | 22 min',
 'UberX 10.34 mi | 16 min',
 'UberX 15.45 mi | 26 min',
 'UberX 14.16 mi | 23 min',
 'UberX 8.79 mi | 35 min',
 'UberX 1.34 mi | 4 min',
 'UberX 8.42 mi | 16 min',
 'UberX 2.27 km | 10 min',
 'nan',
 'nan',
 'UberXL 37.59 mi | 51 min',
 'UberX 6.87 mi | 23 min',
 'UberX 10.85 mi | 1 h 12 min',
 'UberXL 2.76 mi | 25 min',
 'UberX 2.22 mi | 11 min',
 'Black 0.76 mi | 7 min',
 'UberX 1.28 mi | 6 min',
 'UberX 1.89 mi | 10 min',
 'UberX VIP 2.34 mi | 21 min',
 'UberX 0.77 mi | 8 min',
 'UberX 2.18 mi | 16 min',
 'Base Fare $0.00  Duration Fare $4.29',
 'UberX 1.98 mi | 19 min',
 'Pool 6.34 mi | 16 min',
 'Pool 1.58 mi | 11 min',
 'UberX 0.85 mi | 8 min',
 'UberX 4.13 mi | 25 min',
 'UberX 6.65 mi | 19 min',
 'Pool 6.66 mi | 26 min',
 'UberX 12.74 mi | 24 min',
 'UberX 30.68 mi | 41 min',
 'UberX 2.60 mi | 15 min',
 'Pool 0.58 mi | 7 min',
 'UberX VIP 1.88 mi | 26 min',
 'UberX 1.78 mi | 13 min',
 'UberX 1.54 mi | 7 min',
 'UberX 3.52 mi | 14 min',
 'UberX VIP 4.73 mi | 24 min',
 'UberX 3.04 mi | 15 min',
 'UberX 1.36 mi | 12 min',
 'Pool 1.21 mi | 9 min',
 'UberX 4.89 mi | 19 min',
 'Pool 16.03 mi | 48 min',
 'UberX 1.48 mi | 5 min',
 'UberX 8.22 mi | 14 min',
 'Black 2.52 km | 8 min',
 'UberX 10.03 km | 16 min',
 'UberX 3.16 mi | 12 min',
 'UberX 3.72 mi | 7 min',
 'UberX 4.03 mi | 8 min',
 '4.73 kilometers   00:10:00 Trip time   uberXL Car',
 'UberX 3.00 mi | 9 min',
 'Black 2.05 mi | 14 min',
 'UberX Ride cancelled',
 'Comfort 13.24 mi | 15 min',
 'UberX 1.00 mi | 5 min',
 'UberX 0.94 mi | 5 min',
 'UberX 26.80 mi | 32 min',
 'UberX 2.18 mi | 5 min',
 'UberX 14.58 mi | 26 min',
 'UberX 1.84 mi | 5 min',
 'UberX 2.11 mi | 14 min',
 'UberX 14.74 mi | 1 h 6 min',
 'UberX 3.29 mi | 10 min',
 'UberX 1.50 mi | 7 min',
 'UberX 24.50 mi | 40 min',
 …]
```

Looks promising from a regex perspective. Let's define some new columns and extract the data from this field.

[70]:
```
rawData.loc[:,"Distance_Traveled"] = 'NaN'
rawData.loc[:,"Trip_Time"] = 'NaN'
rawData.loc[:,"Product_Type"] = 'NaN'
```

Now we use some of our regex magic to parse the data into these fields for Uber.

[71]:
```
rawData.loc[rawData.product_description.str.contains('^\d\.\d\d miles',na =␣
 ↪False) == True,
             "Distance_Traveled"] = rawData[rawData.product_description.str.
 ↪contains(
    '^\d\.\d\d miles',na = False)].product_description.apply(lambda x: x[0:4])

rawData.loc[rawData.product_description.str.contains('^\d\d\.\d\d miles',na =␣
 ↪False) == True,
             "Distance_Traveled"] = rawData[rawData.product_description.str.
 ↪contains(
    '^\d\d\.\d\d miles',na = False)].product_description.apply(lambda x: x[0:5])

rawData.loc[rawData.product_description.str.contains('^UberX \d\.\d\d',na =␣
 ↪False) == True,
             "Distance_Traveled"] = rawData[rawData.product_description.str.
 ↪contains(
    '^UberX \d\.\d\d',na = False)].product_description.apply(lambda x: x[6:10])

rawData.loc[rawData.product_description.str.contains('^UberX \d\d\.\d\d',na =␣
 ↪False) == True,
             "Distance_Traveled"] = rawData[rawData.product_description.str.
 ↪contains(
    '^UberX \d\d\.\d\d',na = False)].product_description.apply(lambda x: x[6:
 ↪11])

rawData.loc[rawData.product_description.str.contains('^\d\.\d\d miles',na =␣
 ↪False) == True,
             "Trip_Time"] = rawData[rawData.product_description.str.contains(
    '^\d\.\d\d miles',na = False)].product_description.apply(lambda x: x[13:21])

rawData.loc[rawData.product_description.str.contains('^\d\d\.\d\d miles',na =␣
 ↪False) == True,
             "Trip_Time"] = rawData[rawData.product_description.str.contains(
    '^\d\d\.\d\d miles',na = False)].product_description.apply(lambda x: x[14:
 ↪22])

rawData.loc[rawData.product_description.str.contains('^UberX \d\.\d\d mi \|␣
 ↪\d\d min',na = False) == True,
             "Trip_Time"] = rawData[rawData.product_description.str.contains(
  '^UberX \d\.\d\d mi \| \d\d min',na = False)].product_description.
 ↪apply(lambda x: x[16:18])

rawData.loc[rawData.product_description.str.contains('^UberX \d\.\d\d mi \| \d␣
 ↪min',na = False) == True,
            "Trip_Time"] = rawData[rawData.product_description.str.contains(
  '^UberX \d\.\d\d mi \| \d min',na = False)].product_description.
 ↪apply(lambda x: x[16:17])

rawData.loc[rawData.product_description.str.contains('^\d\.\d\d miles',na =␣
 ↪False) == True,
            "Product_Type"] = rawData[rawData.product_description.str.contains(
  '^\d\.\d\d miles',na = False)].product_description.apply(lambda x: x[31:])

rawData.loc[rawData.product_description.str.contains('^\d\d\.\d\d miles',na =␣
 ↪False) == True,
            "Product_Type"] = rawData[rawData.product_description.str.contains(
  '^\d\d\.\d\d miles',na = False)].product_description.apply(lambda x: x[32:])

rawData.loc[rawData.product_description.str.contains('^UberX \d\.\d\d mi',na =␣
 ↪False) == True,
            "Product_Type"] = rawData[rawData.product_description.str.contains(
  '^UberX \d\.\d\d mi',na = False)].product_description.apply(lambda x:␣
 ↪"UberX")

rawData.loc[rawData.product_description.str.contains('^UberX \d\.\d\d mi',na =␣
 ↪False) == True,
            "Product_Type"] = rawData[rawData.product_description.str.contains(
  '^UberX \d\d\.\d\d mi',na = False)].product_description.apply(lambda x:␣
 ↪"UberX")

#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberX Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberX VIP Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   VIP Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   Express POOL Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberXL Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   Select Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberPOOL Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   LUX Car
#regex: {d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   BLACK CAR Car

#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberX Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberX VIP Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   VIP Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   Express POOL Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberXL Car
```

```
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   Select Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   uberPOOL Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   LUX Car
#regex: {d}{d}.{d}{d} miles   {d}{d}:{d}{d}:{d}{d} Trip time   BLACK CAR Car

#regex: UberX {d}.{d}{d} mi | {d}{d} min
#regex: UberX {d}.{d}{d} mi | {d} min
#regex: UberX {d}{d}.{d}{d} mi | {d}{d} min
#regex: UberX {d}{d}.{d}{d} mi | {d} min
```

[72]:

```
#Now we do the same thing for Lyft
lyft = rawData[rawData.merchant_name == 'Lyft']
lyft["product_description"]

#regex: Lyft Fare (d.ddmi, ddm, dds)
```

[72]:

```
0                             Lyft fare (1.63mi, 12m 23s)
1                              Lyft fare (1.73mi, 7m 55s)
2                              Lyft fare (1.54mi, 7m 33s)
3                            Shared fare (2.56mi, 14m 2s)
4                              Lyft fare (1.54mi, 7m 38s)
                               …
1877035    Base fare $2.20 22m 19s $8.70 14.24 mi $12.96
1877042                       Lyft fare (9.03mi, 18m 20s)
1877043                      Lyft fare (12.60mi, 47m 43s)
1877044                       Line fare (1.82mi, 19m 35s)
1877059                                               NaN
Name: product_description, Length: 551954, dtype: object
```

[73]:

```
rawData.loc[rawData.product_description.str.contains(
    'Lyft fare',na = False) == True,"Distance_Traveled"] = rawData[rawData.
 ↪product_description.str.contains(
    'Lyft fare',na=False)].product_description.apply(
    lambda x: x[x.find('('):x.find(')')+1]).apply(
    lambda y: y[y.find('(')+1:y.find(',')-2])

rawData.loc[rawData.product_description.str.contains(
    'Lyft fare',na = False) == True,"Trip_Time"] = rawData[rawData.
 ↪product_description.str.contains(
    'Lyft fare',na=False)].product_description.apply(
    lambda x: x[x.find('('):x.find(')')+1]).apply(
    lambda y: y[y.find(',')+1:y.find('s')]).apply(
    lambda z: ('00:' + z[0:z.find('m')] + ':' + z[z.find('m') + 2:]).replace('␣
 ↪',''')).apply(
    lambda w: w[0:3] + '0' + w[3:] if w[4] == ':' else w).apply(
    lambda q: q[0:6] + '0' + q[6:] if len(q) == 7 else q)
```

```
rawData.loc[rawData.product_description.str.contains(
    'Lyft fare',na = False) == True,"Product_Type"] =rawData[rawData.
 ↪product_description.str.contains(
    'Lyft fare',na=False)].product_description.apply(
    lambda x: x[0:x.find('(')])
```

[74]:
```
#distribution for product_type.
rawData.Product_Type.value_counts()
```

[74]:
```
NaN                                                              460564
Lyft fare                                                        396818
   uberX Car                                                     326027
   UberX Car                                                      90184
   POOL Car                                                       64442
                                                                  …
   Altamonte Car                                                      2
   UberX Diamond Car                                                  2
   property.description.8bae3ac5-90dc-50b4-8f02-8d96fa1a4c03 Car      2
   Black SUV Car Seat Car                                             1
   PSTA-Patricia/Main Car                                             1
Name: Product_Type, Length: 83, dtype: int64
```

[75]:
```
#distribution for product_subtitle
rawData.product_subtitle.unique()
```

[75]:
```
array([nan], dtype=object)
```

[76]:
```
#distribution for product_reseller.
rawData.product_reseller.value_counts()
```

[76]:
```
Tee Pee Mexican Food                              4
Shake Shack (Rice Village)                        3
La Madeleine                                      2
Sabor a Perú                                      2
Moreno Bakery                                     2
Sea Siam                                          2
Wingstop (5222 De Zavala Rd Ste 324)              2
Waffle House (1910 N Collins St)                  2
El Chorro                                         2
Shake Shack (Mary Brickell Village)               2
Cristino's Coal Oven Pizza                        2
Shake Shack (Dallas Uptown)                       1
KFC                                               1
Taco Bell(3101 East Thomas Road )                 1
Tarbell's The Tavern                              1
Shipley Donuts - 510 E ABRAM ST                   1
North Italia (North Arcadia 4925 N 40th St)       1
```

```
Taco Bell (640 NE 125th St)                     1
Pluckers-Dallas                                 1
Yummy House China Bistro (Hillsborough Ave)     1
McDonald's® (Louis Henna)                       1
Golden China (Lighthouse Point)                 1
Pomo                                            1
Pizza Hut (6141 Windhaven Pkwy Ste 135)         1
Taco Bell (390 W 29th St)                       1
Taco Bell (8658 Highway 121)                    1
Sacks Sandwiches                                1
Le Peep - 9807 KATY FWY                         1
Pizza Hut (8400 W Bellfort St)                  1
Whataburger (9394 Lyndon B Johnson Fwy)         1
McDonald's® - Brooklyn (880 Coney Island)       1
369 Oriental Bistro                             1
Top China (Midtown, Grand Central District)     1
Thai House Restaurant                           1
McDonald's® (Blue Island & Loomis)              1
TacoBell - 3550 Ulmerton Rd                     1
Angela's Cafe (Inwood Road)                     1
Avocado Grill (WPB)                             1
Eggsellent Cafe - Frisco                        1
Pho An Hoa                                      1
Hao                                             1
Grill Chill Pita                                1
C.R. Chicks (Greenwood Plaza)                   1
Name: product_reseller, dtype: int64
```

[77]:
```
#product_category distribution.
rawData.product_category.value_counts()
```

[77]:
```
Travel    1835263
Name: product_category, dtype: int64
```

[78]:
```
#distribution for SKU.
rawData.SKU.value_counts()
```

[78]:
```
Series([], Name: SKU, dtype: int64)
```

[79]:
```
#item_id distribution.
rawData.item_id.value_counts()
```

[79]:
```
Series([], Name: item_id, dtype: int64)
```

So as we can see from all of the above, the only truly useful field is product_description, which allows us to infer several fields: travel time, distance traveled, and details about the ride itself (such as which car was used, whether it was VIP or pool, etc.).

2.7 MISCELLANEOUS FIELDS

This leaves us with the last bit of analysis on the the following miscellaneous columns:

- digital_transaction
- order_pickup
- from_domain
- email_subject

We analyze the data in these fields below to see if there is anything useful that we can get out of them.

```
[80]: #digital_transaction distribution.
      rawData.digital_transaction.value_counts()
```

```
[80]: 0    1835263
      Name: digital_transaction, dtype: int64
```

```
[81]: #order_pickup distribution.
      rawData.order_pickup.value_counts()
```

```
[81]: 1    1814578
      3      12543
      0       8142
      Name: order_pickup, dtype: int64
```

```
[82]: #what happens when order_pickup is not 1?
      rawData[rawData.order_pickup != 1]["email_subject"]
```

```
[82]: 7                 We corrected the fare for your ride on Aug 3
      23                              You've got Uber Cash at 5% off
      34                              You've got Uber Cash at 5% off
      94                We corrected the fare for your ride on May 6
      95               We corrected the fare for your ride on Nov 20
                                                   …
      1877039           We corrected the fare for your ride on Jun 6
      1877040    [Personal] We corrected the fare for your ride…
      1877041          We corrected the fare for your ride on Aug 23
      1877055           We corrected the fare for your ride on Feb 6
      1877056           We corrected the fare for your ride on Jun 8
      Name: email_subject, Length: 20685, dtype: object
```

```
[83]: #from_domain distribution
      rawData.from_domain.value_counts()
```

```
[83]: uber.us@uber.com                 1246780
      no-reply@lyftmail.com             546450
      receipts.nyc.suburbs@uber.com      17129
      receipts.upstate.ny@uber.com        4897
```

```
noreply@uber.com                    3026
                                  …
receipts.gatineau@uber.com             1
uber.uae@uber.com                      1
receipts.madinah@uber.com              1
receipts.texarkana@uber.com            1
uber.israel@uber.com                   1
Name: from_domain, Length: 71, dtype: int64
```

As the above code shows, order_pickup seems to indicate whether or not the transaction was a real pick-up or whether or not it was some sort of informational message, such as a correction. It makes sense on this basis, given the relatively small number of rows, to remove those values for order_pickup that are not 1.

Digital_Transaction is a useless field since it is always 0. The from_domain seems to have some regional data for the receipts that come from upstate or other countries (such as Israel). Other than that, I cannot see much use for it immediately for this analysis.

As such, we will reject rows with the condition that order_pickup is not 1, and select only those rows that we care about into a new dataframe, which we can use to perform analysis.

```
[84]: rawData = rawData[rawData.order_pickup == 1]
```

2.8 FINAL TABLE CREATION

We have gone through all of the columns and ascertained that some are more useful than others. We have also derived some new fields based on the columns that we had. The below code takes the useful fields and loads it into a final table, which also has an index on order_number. Since it is unique, we can now rest happily knowing that our set has a primary key.

```
[85]: finalData =␣
       ↪rawData[['merchant_name','user_id','order_number','order_total_amount','date_of_order',
                                   ␣
       ↪'time_of_order','Distance_Traveled','Trip_Time','Product_Type']].copy()
```

```
[86]: finalData.set_index('order_number')
```

```
[86]:                                      merchant_name  \
      order_number
      1183621593749890808                           Lyft
      1155413807788118982                           Lyft
      1311366563734805982                           Lyft
      1202358710117776328                           Lyft
      1234240944632141798                           Lyft
      …                                            …
      f247ae1a-b81b-43ad-82cf-b35ee9b9667d          Uber
      e9d1e1af-f8c2-45f9-9241-65040369cf75          Uber
      6e11bab3-3b10-40c2-970e-66bdf17effdb          Uber
```

```
7cc351eb-ab18-4b32-b474-c5cc8d233252          Uber
1309397943170611458                           Lyft

user_id  \
order_number
1183621593749890808
3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
1155413807788118982
3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
1311366563734805982
3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
1202358710117776328
3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
1234240944632141798
3a0a2eca8cce05eaf4b8213040f836e122ff210a1d7e88…
…
…
f247ae1a-b81b-43ad-82cf-b35ee9b9667d
1d0ed9a5725e5b71411e3796b084358fea628f259628c2…
e9d1e1af-f8c2-45f9-9241-65040369cf75
815c1d46eda38992bf976aa45e10fd81c871fa77b64e6b…
6e11bab3-3b10-40c2-970e-66bdf17effdb
263f69452a0c2d46968effe172ba5f4f20785dc0640f85…
7cc351eb-ab18-4b32-b474-c5cc8d233252
415b8e1517a45479a2cbabbad5afc6a6cd5f12bc2aa8dd…
1309397943170611458
bdaeb8a007ac5004b9e439e896fa380828ad0236ade1aa…

                                      order_total_amount date_of_order  \
order_number
1183621593749890808                                 9.65   2018-09-25
1155413807788118982                                 8.65   2018-07-11
1311366563734805982                                11.39   2019-09-04
1202358710117776328                                 6.32   2018-11-14
1234240944632141798                                10.90   2019-02-08
…                                                  …           …
f247ae1a-b81b-43ad-82cf-b35ee9b9667d                5.00   2017-07-05
e9d1e1af-f8c2-45f9-9241-65040369cf75                5.00   2018-03-31
6e11bab3-3b10-40c2-970e-66bdf17effdb                5.00   2017-08-15
7cc351eb-ab18-4b32-b474-c5cc8d233252                5.00   2017-07-26
1309397943170611458                                43.44   2019-08-30

                                     time_of_order Distance_Traveled  \
order_number
1183621593749890808                        9:00:00              1.63
1155413807788118982                        8:38:00              1.73
1311366563734805982                        4:54:00              1.54
```

```
1202358710117776328                         9:56:00            NaN
1234240944632141798                         7:52:00           1.54
…                                             …              …
f247ae1a-b81b-43ad-82cf-b35ee9b9667d        0:00:00            NaN
e9d1e1af-f8c2-45f9-9241-65040369cf75        0:00:00            NaN
6e11bab3-3b10-40c2-970e-66bdf17effdb        0:00:00            NaN
7cc351eb-ab18-4b32-b474-c5cc8d233252        0:00:00            NaN
1309397943170611458                         4:16:51            NaN

                                     Trip_Time Product_Type
order_number
1183621593749890808                   00:12:23   Lyft fare
1155413807788118982                   00:07:55   Lyft fare
1311366563734805982                   00:07:33   Lyft fare
1202358710117776328                        NaN         NaN
1234240944632141798                   00:07:38   Lyft fare
…                                         …          …
f247ae1a-b81b-43ad-82cf-b35ee9b9667d       NaN         NaN
e9d1e1af-f8c2-45f9-9241-65040369cf75       NaN         NaN
6e11bab3-3b10-40c2-970e-66bdf17effdb       NaN         NaN
7cc351eb-ab18-4b32-b474-c5cc8d233252       NaN         NaN
1309397943170611458                        NaN         NaN

[1814578 rows x 8 columns]
```

[87]:
```
finalData.index.is_unique
```

[87]:
```
True
```

3 PART 2: DATA ANALYSIS IDEAS

In the last part of this case study, we explore myriad ideas for doing something with this new dataset now that we have it. Here are some ideas for analysis that we could potentially pursue with this data:

- **Order Date and Time Analysis:** We could look into when orders are placed throughout the day for each merchant. Are they getting more business early in the morning? Late at night? What are the slow periods? We could also look into the *calendar dates* for orders. For each merchant, what are the slow seasons? What are the busy times? How do weekends compare to holidays and weekdays? How do Fridays look compared to Thursdays?
- **Traffic Patterns and Speed:** Since we have derived the distance that each trip took and how long each trip took, we can derive the *average speed* for the trips. We could potentially correlate this with the time of day that the trip occurred to infer traffic patterns on that day, in the region that they traveled!
- **Market Cap Analysis:** We could look at the sales for each merchant. If this dataset is representative, what is the relative difference between Uber and Lyft in terms of market cap

(assuming that these two firms account for nearly 100% of the total rideshare market)? What if we partition the data by season; does that change?

- **Customer Retention Analysis:** We could look into the distribution of customers for both merchants. Which has more customers? For the customers in each firm's network, what is the average/median for the number of transactions per customer? What is the average/median spend per customer?

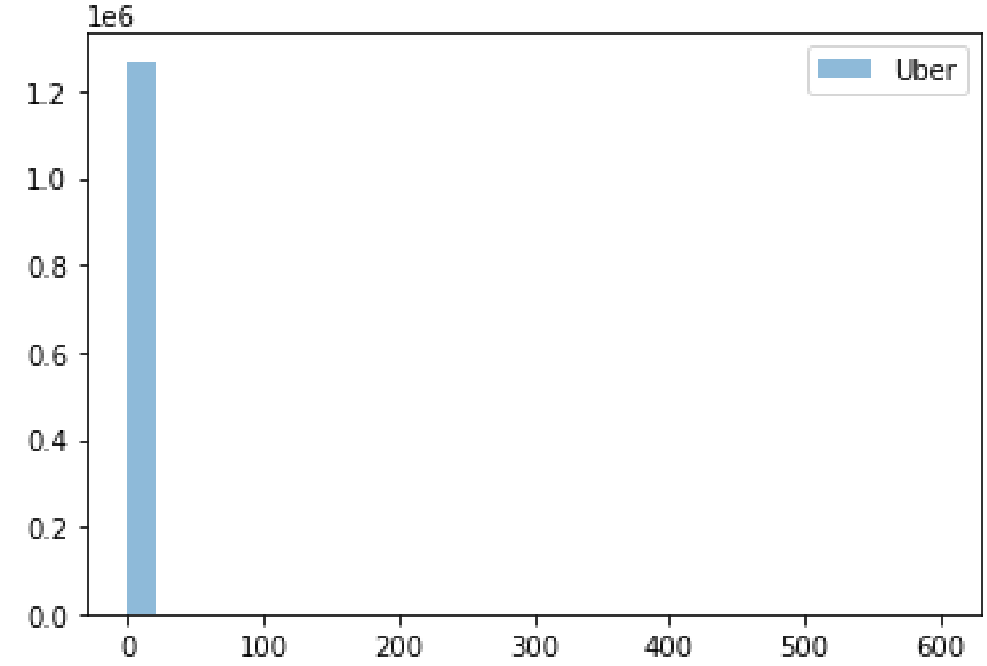
There are no doubt other possible cases to study, but in the interest of time and space, we will restrict ourselves to the first two. We leave the other two for future extensions of the case study.

3.1 Order Date and Time Analysis

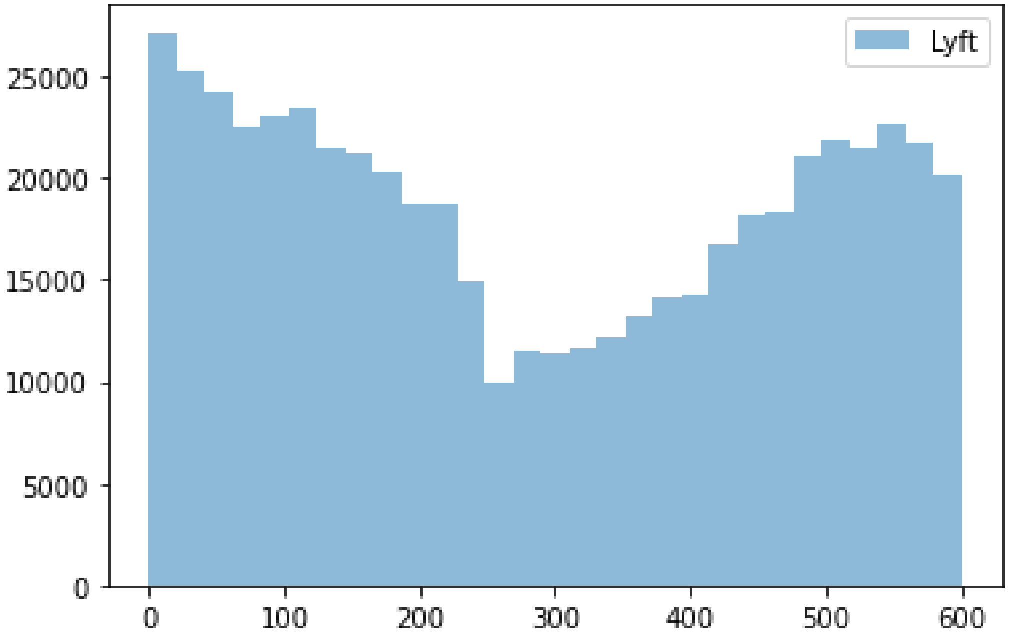
For this, we will want to partition each merchant and examine a plot of the order time for each. As a basic first step, we can look at histograms for each firm. The code below generates these histograms for Uber and Lyft.

[88]:

```
from matplotlib import pyplot as plt

uberTimes = finalData[finalData.merchant_name == 'Uber'].time_of_order
lyftTimes = finalData[finalData.merchant_name == 'Lyft'].time_of_order
uberTimes.dropna()
lyftTimes.dropna()

uberTimeInMinutes = uberTimes.apply(
    lambda x: 60*int(x[0:x.find(':')]) + int(x[x.find(':')+1:x.find(':')+3]))

lyftTimeInMinutes = lyftTimes.apply(
    lambda x: 60*int(x[0:x.find(':')]) + int(x[x.find(':')+1:x.find(':')+3]))

b = np.linspace(0,600,30)

plt.hist(uberTimeInMinutes, b, alpha=0.5,label = 'Uber')
plt.legend(loc = 'upper right')
plt.show()
```



[89]:
```
plt.hist(lyftTimeInMinutes ,b, alpha=0.5,label = 'Lyft')
plt.legend(loc = 'upper right')
plt.show()
```

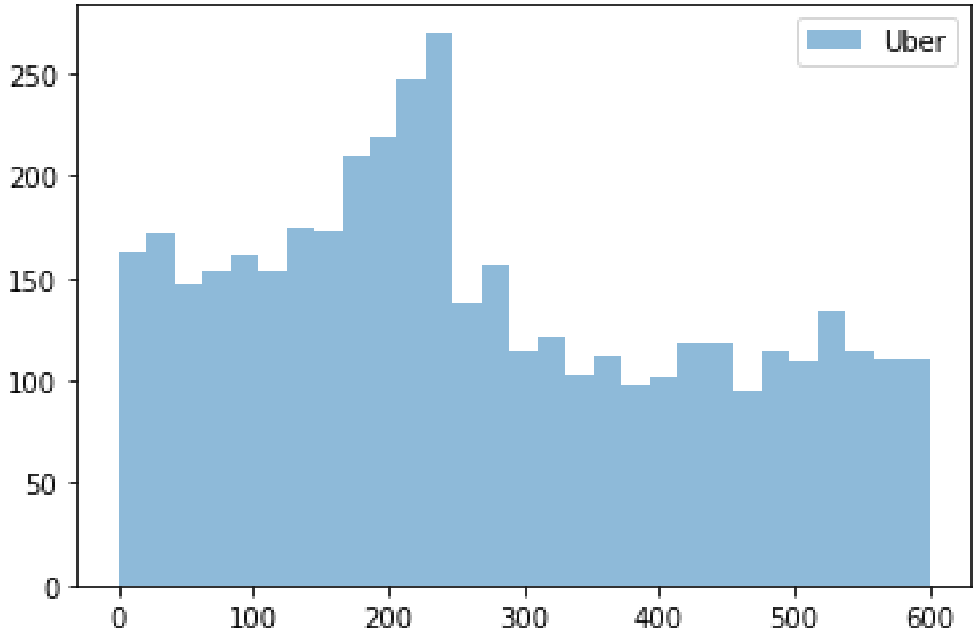


As we can see from the above, Uber and Lyft have some serious differences in their data. Lyft appears to have a valid time component, which shows high usage in the wee hours of the morning (near 0, which is midnight), and then it gradually sinks in the afternoon, only to pick up again in the evening (just in time for people's evening commute).

As for Uber, most of the data are clustered around 0, which could either mean that the orders are overwhelmingly being placed close to midnight or (perhaps more likely) the time data for Uber are deficient for some reason. This is something that we could further investigate. As a test, let's see what it looks like if we exclude 0.

[90]:

```
uberTimeInMinutes = uberTimeInMinutes[uberTimeInMinutes > 0]
plt.hist(uberTimeInMinutes, b, alpha=0.5,label = 'Uber')
plt.legend(loc = 'upper right')
plt.show()
```



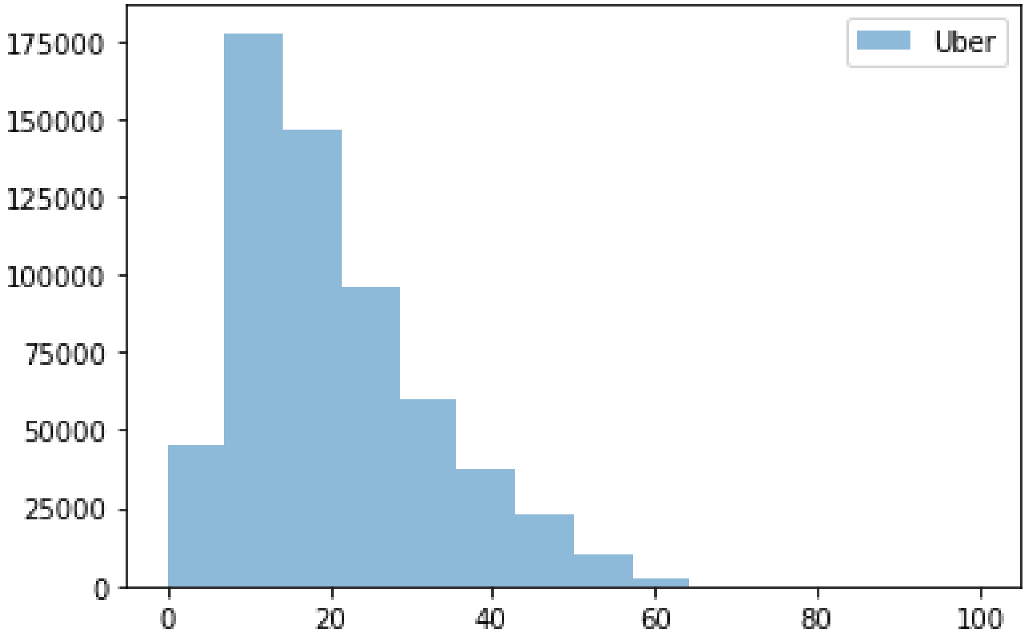
As we can see, there are some nonzero elements here. This indicates that perhaps the time field is not entirely corrupt, but only that our logic for parsing it for Uber rows is insufficient. This would merit more investigation. We could use similar tools to investigate the distribution of dates as well, but we leave that as a future consideration and move on to speed and traffic patterns.

3.2 Traffic Patterns and Speed

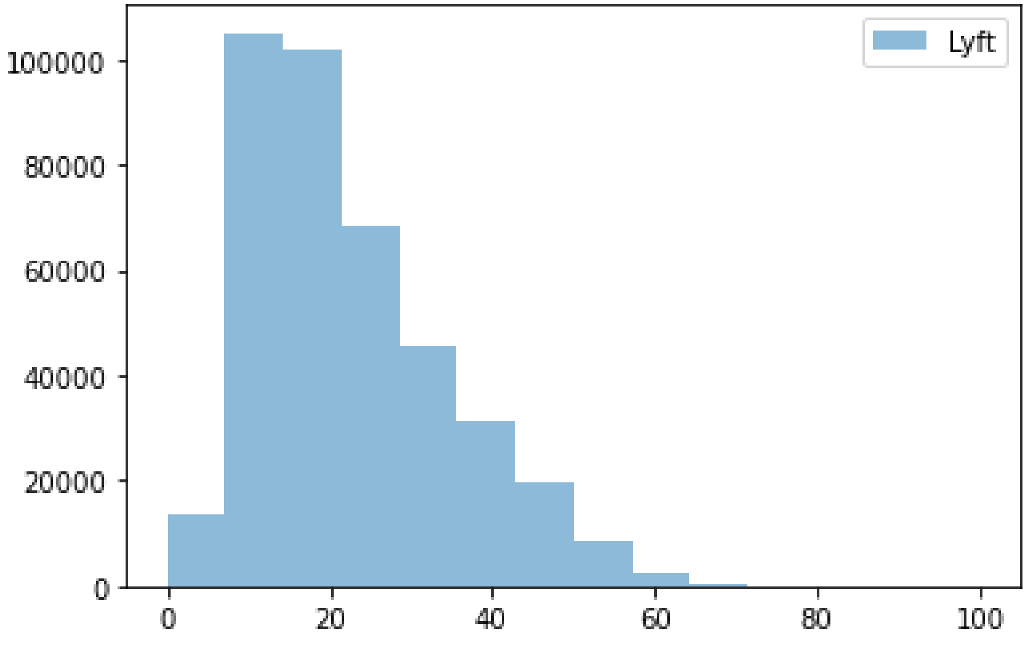
Since we have access to the total trip time and total travel time fields, we can determine the average and median speeds for each of Uber and Lyft. The below code handles this.

[120]:

```
uberSpeed = finalData[(finalData.merchant_name == 'Uber') & (finalData.
 ↪Distance_Traveled.notnull()) & (
    finalData.Distance_Traveled.apply(lambda x: x != "NaN")) & (finalData.
 ↪Trip_Time.notnull()) & (
    finalData.Trip_Time.apply(lambda x: x != "NaN"))]
uberSpeed = uberSpeed[['Distance_Traveled','Trip_Time']]
uberSpeed["Hours"] = uberSpeed.Trip_Time.apply(
    lambda x: int(x[0:2]) if x.find(":") > 0 else 0)
uberSpeed['Minutes'] = uberSpeed.Trip_Time.apply(
    lambda x: int(x[3:5])/60.00 if x.count(':') == 2 else 0)
uberSpeed = uberSpeed[uberSpeed.Hours + uberSpeed.Minutes > 0]
uberSpeed['Average_Speed'] = uberSpeed.Distance_Traveled.astype(float) /␣
 ↪(uberSpeed.Hours.astype(float) + uberSpeed.Minutes.astype(float))
uberSpeed

b1 = np.linspace(0,100,15)

plt.hist(uberSpeed.Average_Speed,b1,alpha=0.5,label='Uber')
plt.legend(loc = 'upper right')
plt.show()
```



[156]:
```
lyftSpeed = finalData[(finalData.merchant_name == 'Lyft') & (finalData.
 ↪Distance_Traveled.notnull()) & (
    finalData.Distance_Traveled.apply(lambda x: x != "NaN")) & (finalData.
 ↪Trip_Time.notnull()) & (
    finalData.Trip_Time.apply(lambda x: x != "NaN"))]
lyftSpeed = lyftSpeed[['Distance_Traveled','Trip_Time']]

lyftSpeed["Hours"] = lyftSpeed.Trip_Time.apply(
    lambda x: int(x[0:2]) if x.find(":") > 0 else 0)

lyftSpeed['Minutes'] = lyftSpeed.Trip_Time.apply(
    lambda x: x[3:5])

lyftSpeed = lyftSpeed[lyftSpeed.Minutes.str.contains(":") == False]

lyftSpeed = lyftSpeed[lyftSpeed.Hours.astype(int) + lyftSpeed.Minutes.
 ↪astype(int).apply(
lambda x: x/60.00) > 0]

lyftSpeed['Average_Speed'] = lyftSpeed.Distance_Traveled.astype(float) /␣
 ↪(lyftSpeed.Hours.astype(float) + lyftSpeed.Minutes.astype(float).apply(
lambda x: x/60.00))


plt.hist(lyftSpeed.Average_Speed,b1,alpha=0.5,label='Lyft')
plt.legend(loc = 'upper right')
plt.show()
```



As we can see from the above, the average speed for both firms appears similar.

4 FINAL THOUGHTS: EXTENSIONS AND QUESTIONS

4.1 Questions for the Vendor

Now that we have reached the end of the case study, I want to take this space to collect some of the questions that I would ask of the data provider, if given the opportunity. These come from the previous cells of analysis:

1. Are User_Ids universally assigned?
2. Why would a row ever be misssing an order_number?
3. Why would there be 2 rows with the same order_numbere and the same user_id?
4. Is it possible for there to be 2 or more user_ids connected to the same order number?
5. Typically, it appears as though order_numbers for Uber are 36 character-long strings and those for Lyft are 19 character-long strings. Why do we get some few that do not conform to these standards?
6. Why is it the case that my order_total integrity checks failed? What is the relationship between order total, subtotal, tax, and discount?

4.2 Relational Database Modeling

Another point to make in ending the case study is that there are at least three different, normalized tables that one would create for this dataset. The following are the names of the tables and the fields from the file that I would associate with each:

- **Merchant**: This would have one row per merchant in the dataset. In this case, we have two: Uber and Lyft. The fields that I would associate with this table include: merchant_name.

- **User**: This would have one row per customer in the dataset. The fields would include: user_id.

- **Product**: This would have one row per product.

- **Order**: This would be our primary fact table. This table would have foreign keys to the Merchant, User, and Product tables. The fields that I would also include in this table from the dataset include: order_number, time_of_order, date_of_order, order_total_amount, order_tax, order_subtotal, order_total_qty, trip_time, distance_traveled, item_quantity, item_price, order_discount

Altogether, this would constitute a star schema. We could add additional fields to these tables as well, pending answers to some of our questions from the vendor.