Chicago Taxi Trips

Davide Bartoletti
Jesus Currel

Dataset

Taxi trips reported to the City of Chicago, the dataset is heavy and characterised by 29 columns, the following reported are the ones used for our analysis

Features:

- *Trip ID*: unique identifier for the trip
- *Taxi ID*: unique identifier for the taxi
- *Trip start Timestamp*: when the trip started
- *Fare* : fare for the trip
- *Tips* : tip for the trip
- Extra: extra charges for the trip
- *Payment method*: type of payment for the trip

Data transformation

All data were filtered thanks the usage of Pandas, due to the restriction on the maximum uploading dimension on Observable (50MB). Before the filtering operations , we removed empty and null values. Then the useless columns were removed and data were grouped according to the purpose of the visualization. After on observable the data were mapped to specific objects in order to get the different plots.

Questions:

- Is there any correlation between the extra charges and fare/tips? do they differ based on payment methods?
- How do people tip when compared to fare charged and payment method?
- How has the amount of taxi taken changed during the year 2021, immediately after the Covid-19 restrictions?

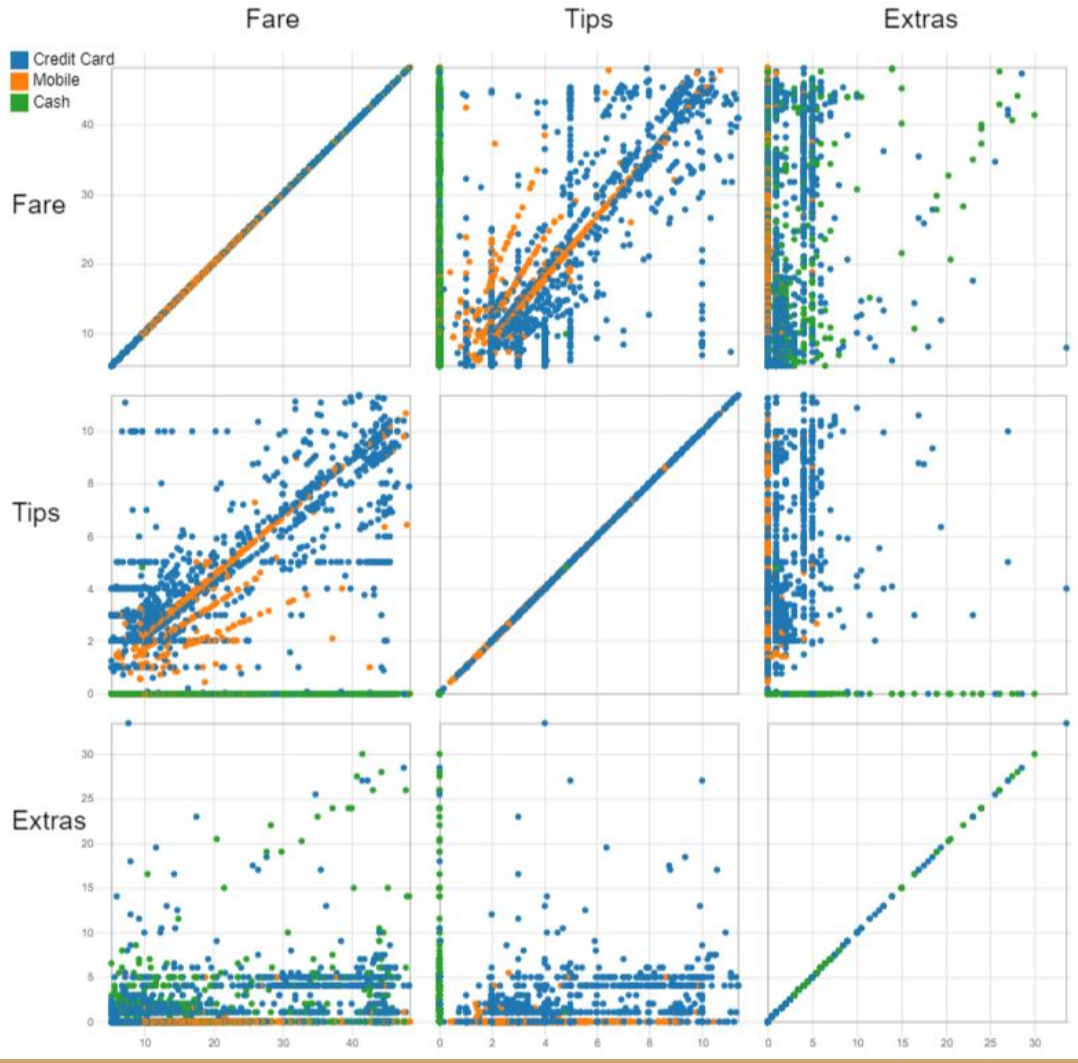
Matrix scatterplot

The matrix scatterplot shows the behaviour of the fare and tips and extra charges amounts.

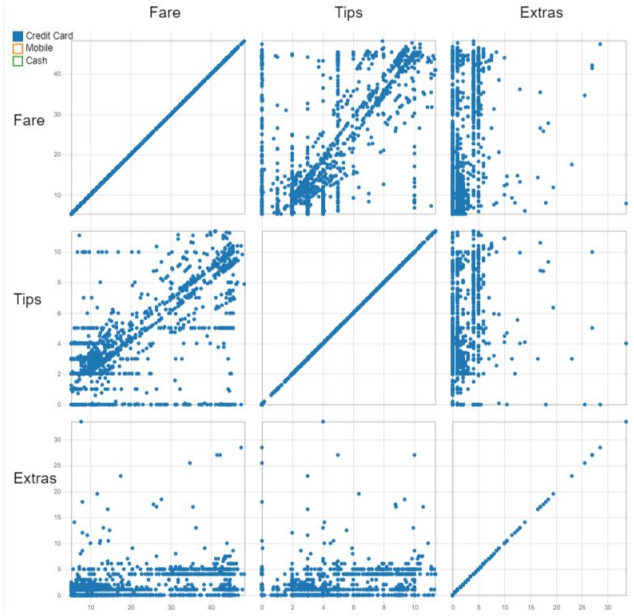
The color encodes the different type of payments: Credit Card, Mobile and Cash

The graph can also be separated different payments to help with over plotting

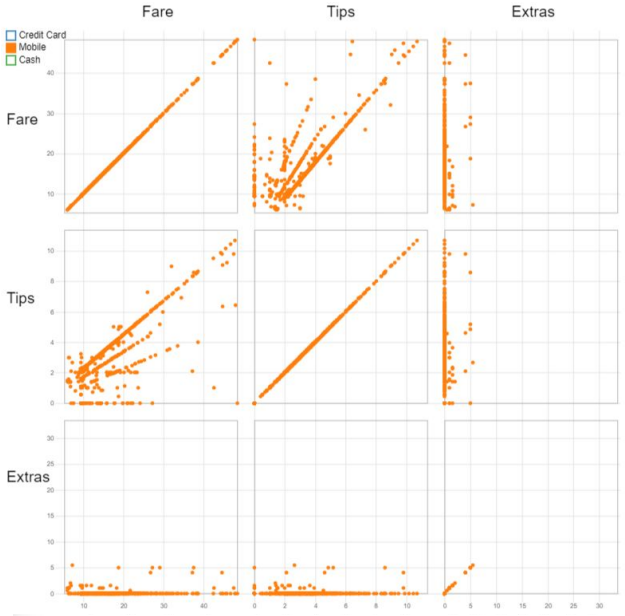
The remaining kind of payments were removed because they only included fare and no tips or extra charges



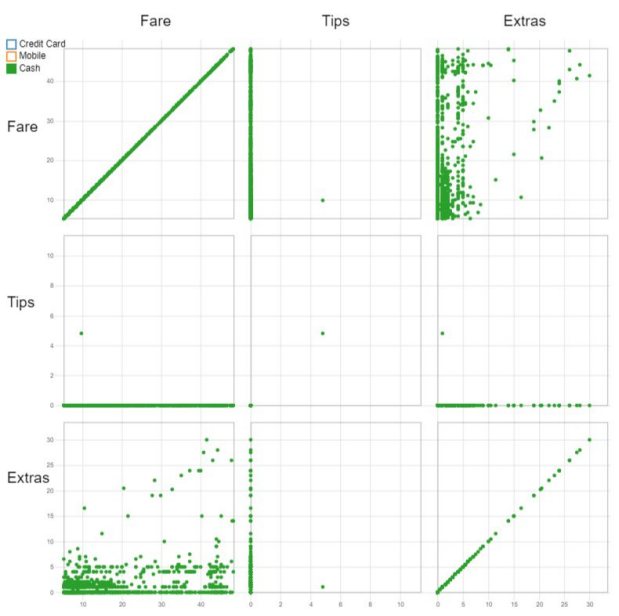
Payment methods



- The preferred way of payment is the credit card type, as shown from the majority of blue dots
- Horizontal rows show fixed amounts for the tips, like 1,2,3 or 5 $, independently from the fare amount



- For mobile payments we can see a linear behaviour for the tips, probably due to the possibility on choosing the % of tips according to the fair (18%, 20%, 25%)



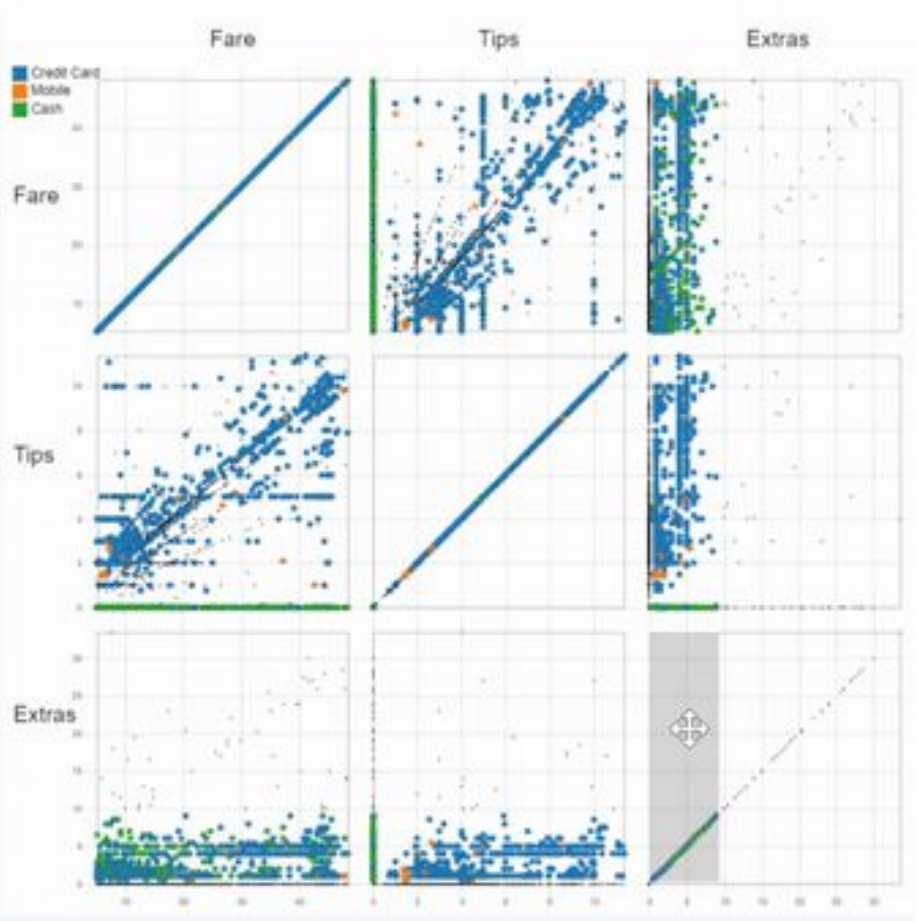
- For the Cash payments the tips is always zero, probably because there is no track of these.

Data selection

Thanks the usage of a brushing operation is possible to select specific range of data. It will result with a window on which the dots are colored with respect to the remaining ones that will be filled with a blackcolor.

Here we see a sliding brush example when looking at increasing extra charges

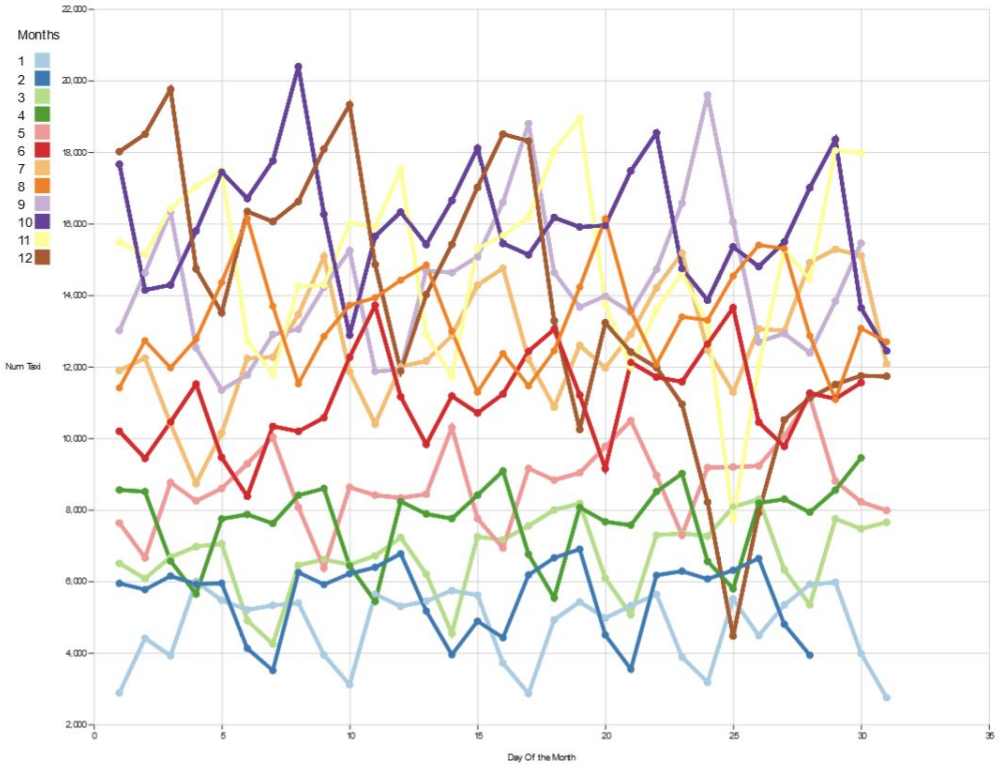
What we see is that there is no correlation between higher extra charges and the amount of tips/fare is charged.
What we do see is that Mobile payments have the smallest amounts of extra charges.

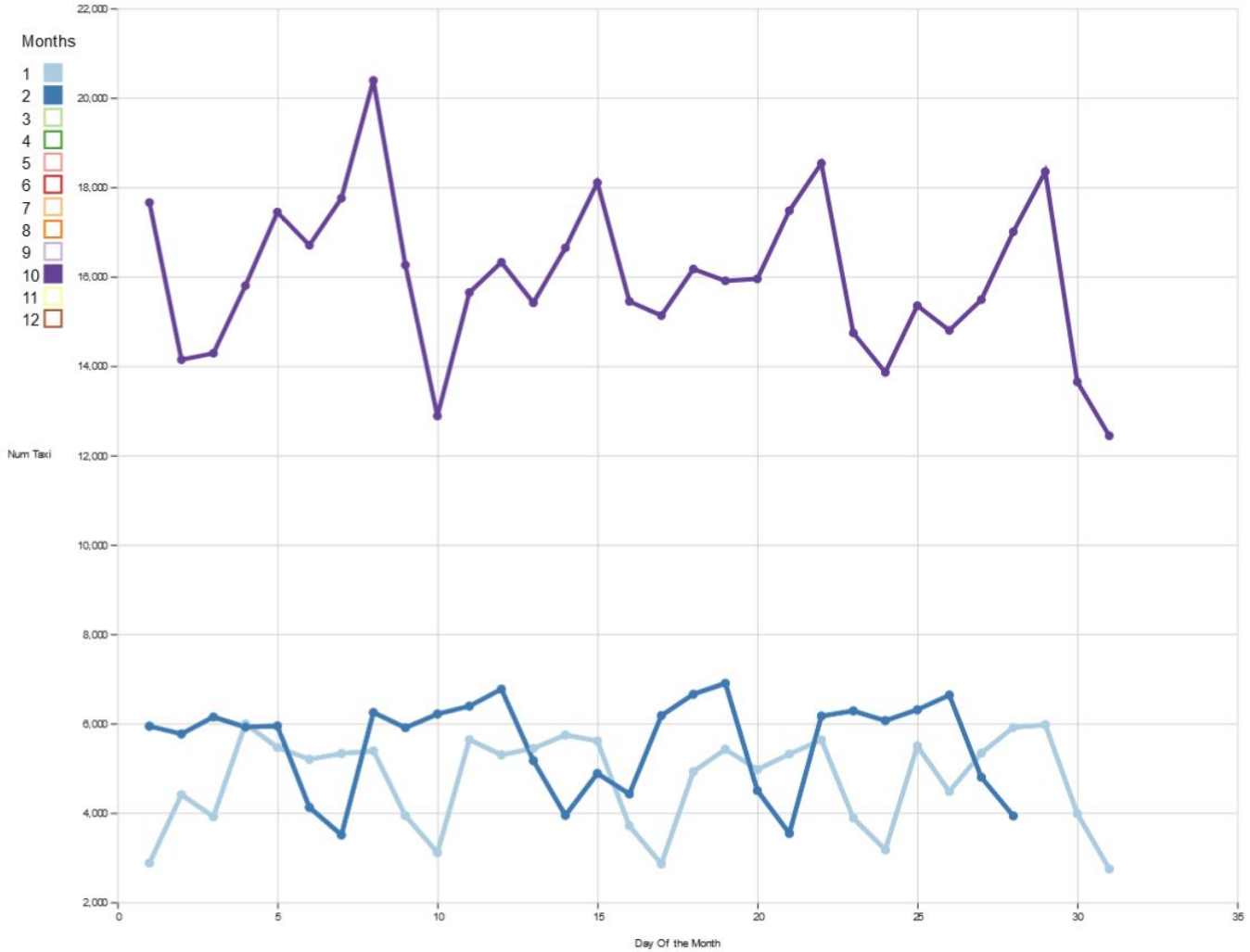


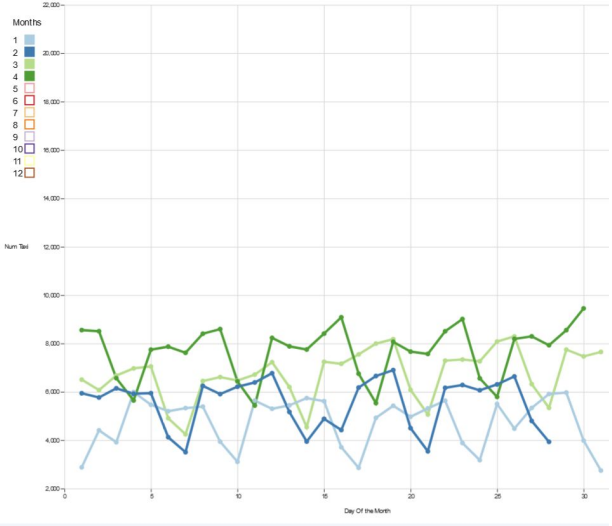
Month behaviour in year 2021

The plot used to analyse this kind of behaviour is a multiline plot in which each colored line encodes a specific month of the year. On the y axis is reported the number of rides, while on the x axis, the different days of the month. This kind of representation fits perfectly for timeseries and quantity values as the one described before.

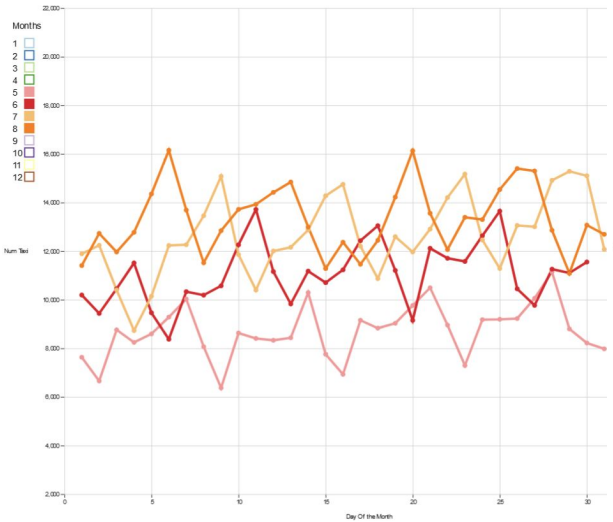
Thanks the usage of the colored squares on the left, is possible to select or uncheck a specific month to make it appear in the visualization.



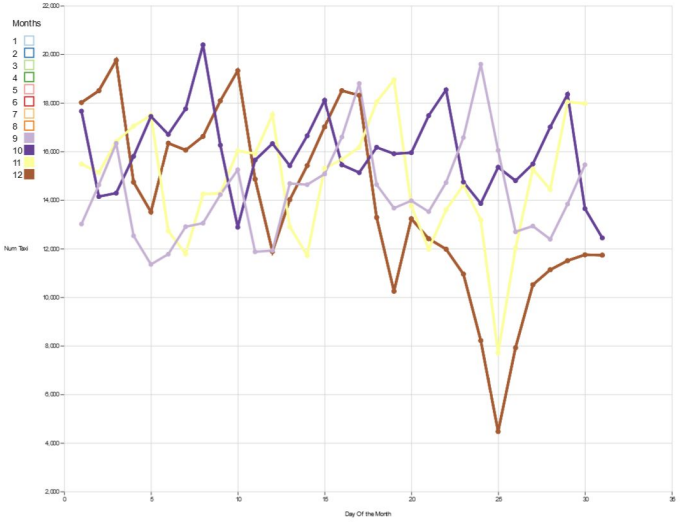

Months
1
2
3
4
5
6
7
8
9
10
11
12
Num Taxi
22,000
20,000
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
5
10
15
20
25
30
35
Day Of the Month



For the first four months the number of taxi rides was low, probably due to the restrictions, so taxi were taken only by workers in this period.



Overall increasing in the number of taxi rides in the second part of the year, probably to the easing of restrictions.



Increasing in the number of taxi, except for two quick decrease on Christmas day and the day after the Thanksgiving.

Conclusions

Thanks to the visualizations proposed we were able to answer to the questions introduced at the beginning.

With the first plot we were able to see a majority of payments done through credit card and the different kind of tips related to the amount of money the method used to pay. We also saw the lack of correlation between extra charges, fare, and tips.

The connected scatterplot in the second case allow us to track the flow of taxi rides during an entire year. We are able to analyse different months together looking for interesting points characterised by big increase or decrease in the amount of rides provided during 2021.