**SpekIt Technical Challenge**

**Business Problem**

Pull the following from this link <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

* Taxi trip records from 4 sources
* Associated data dictionaries
* Taxi zone lookups

Answer the following questions about the effects of the Corona Virus on taxi transportation in NYC:

* What has been the economic impact of the pandemic on NYC taxi revenue over time?
* What was the most expensive trip? Between what zones? Conversely, what was the cheapest trip? What is the difference between these before and during the peak impact of the pandemic?
* What's the most popular payment method? Did this change because of the pandemic?
* What’s the most expensive day of the week to travel on? Did this change because of the pandemic?

Submit your answers in the form of a GitHub repository, with additional instructions on how to reproduce them. Store the data into an AWS Postgres instance (free tier) or equivalent and provide access to us to review.

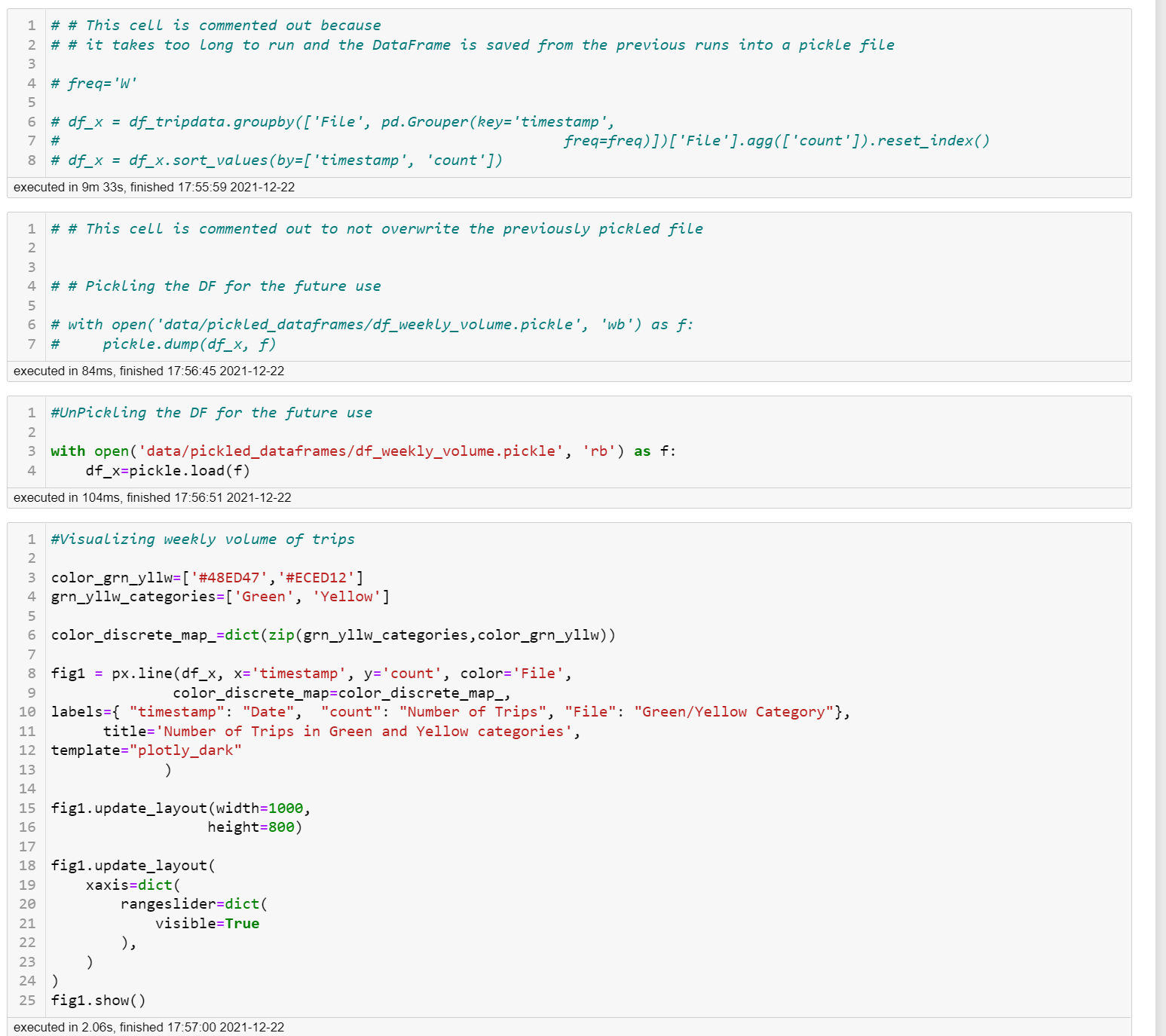
**Pre-processing of the original data**

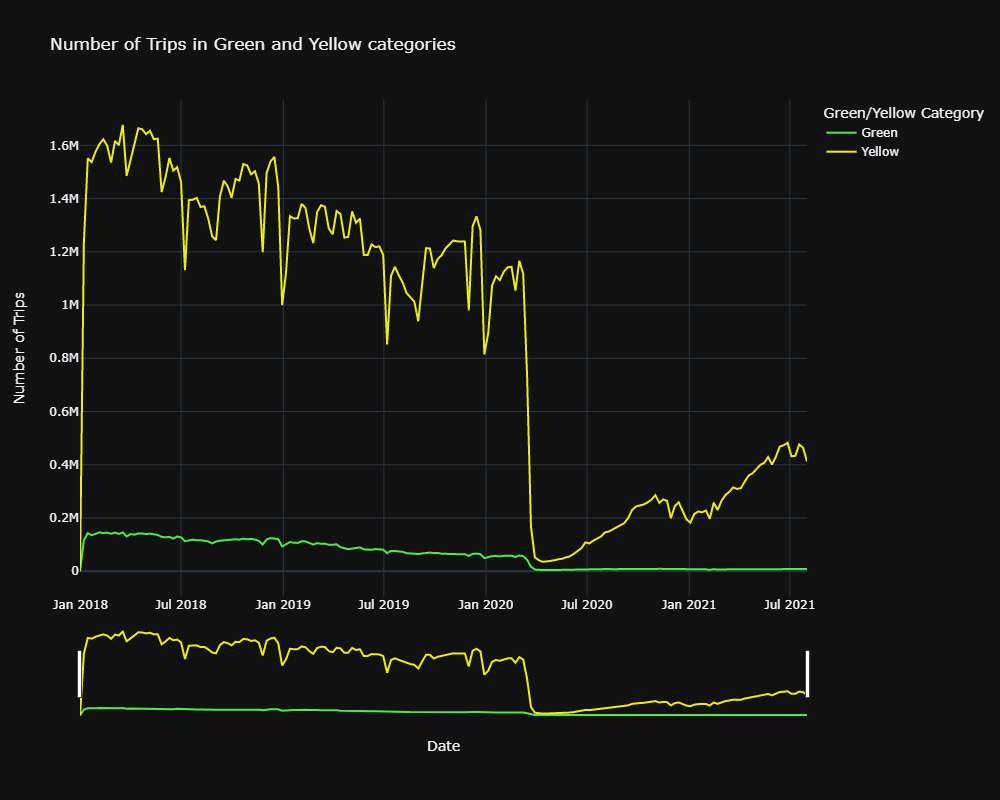
* Only files from 2018, 2019, 2020 and 2021 were used
* Only filed with the Yellow and the Green tripdata were used
* Rows with VendorID equal NULL were dropped
* Only 75% of the records chosen randomly were included in the final analysis
* Duplicate rows were dropped
* Rows with timestamps with errors we dropped
* The final sqlite Database has two tables

**Most of the modifications to the original data were done to reduce the volume of the records unmanageable on my home machine. The usage of AWS PostgreSQL instance was cost prohibitive.**

**Visualization:**

Number of trips between 2018 and 2021 (July). The plot is interactive in Jupyter notebook but GitHub does not accept plotly images





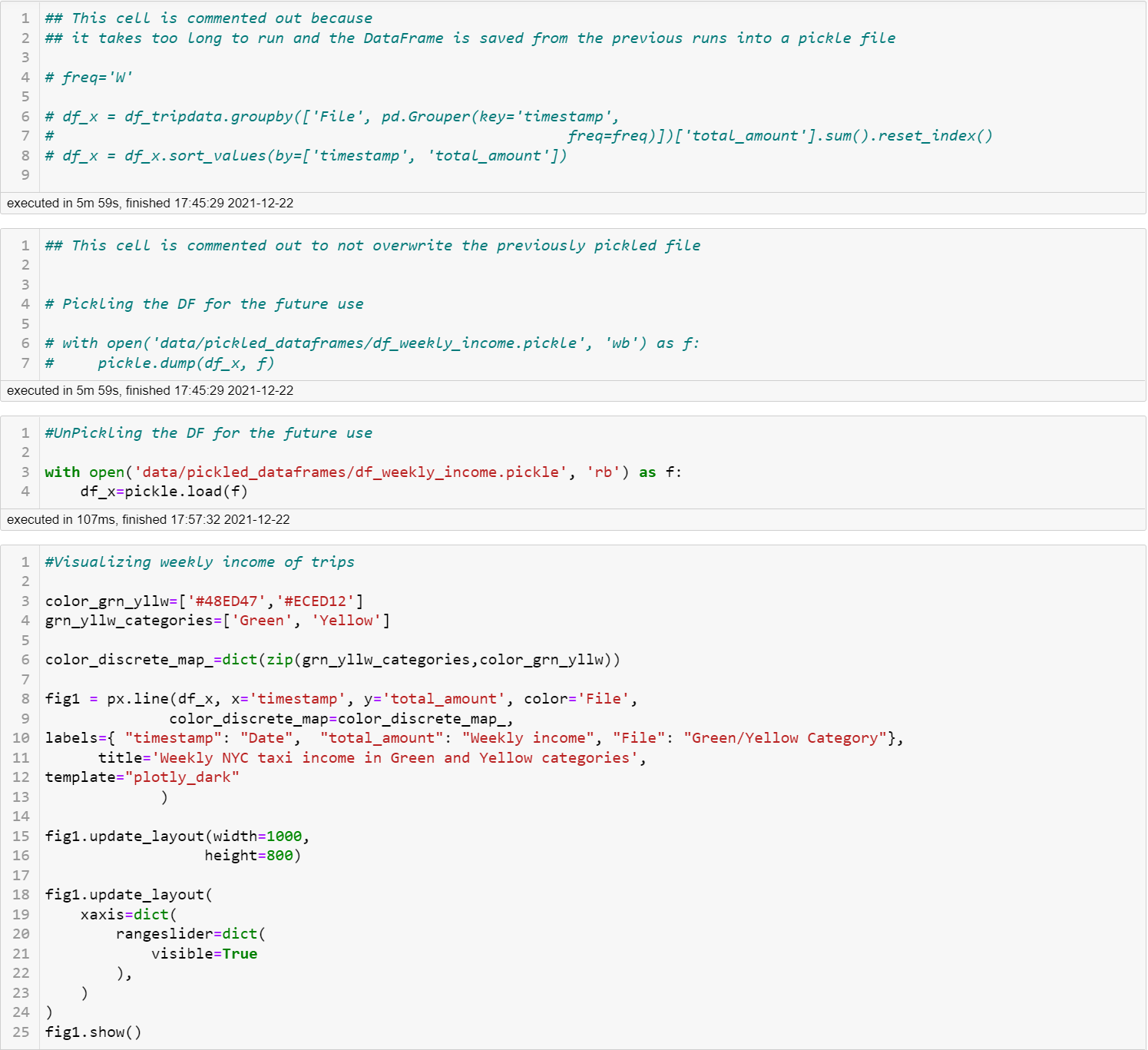
**This visualization shows a pretty dramatic impact of the pandemic on the weekly trip volume of NYC taxi services both in Yellow and in Green categories**.

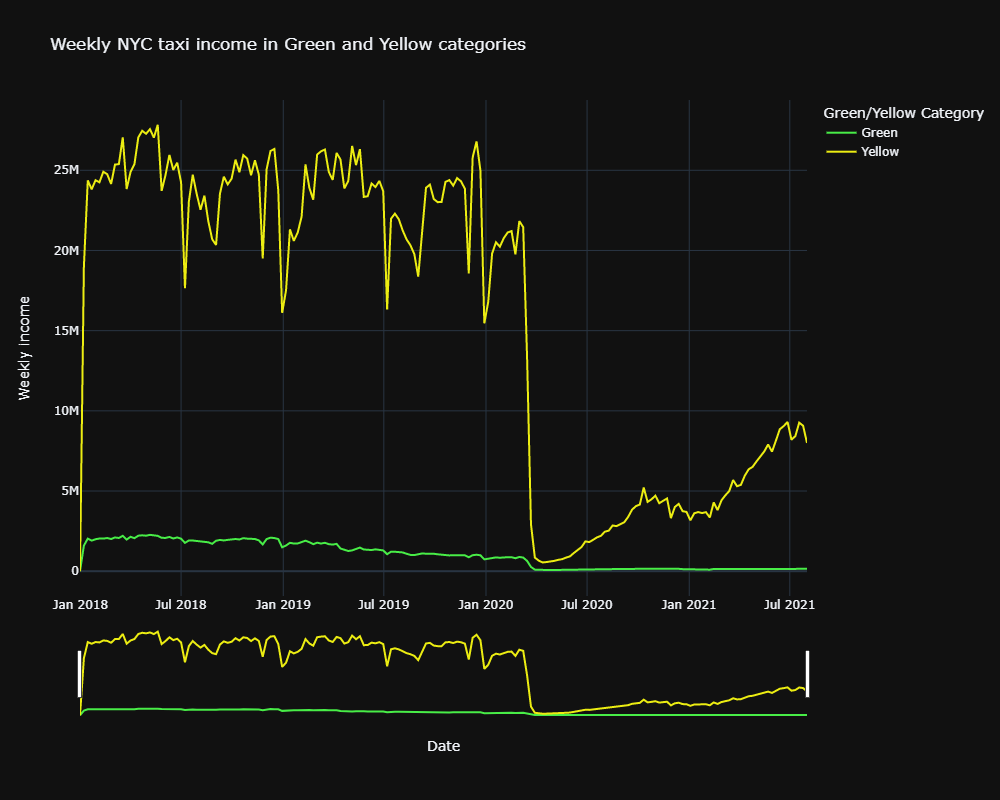
* While the volume of the trips is significantly more significant in the Yellow category, the trend in both categories is the same. Dramatic drop at the beginning of the pandemic in March 2020 with a slow recovery in Yellow category up to about 35-40% of the pre-pandemic volume. The Green category remains almost flat at probably 10% of the pre-pandemic level.

It is also worth mentioning that both categories were experiencing a downward trend over the last two pre-pandemic years. Possibly due to the competition from Uber, Lyft, and alike

* .

Weekly income of NYC taxi over the period of 2018 (January) and 2021 (July)





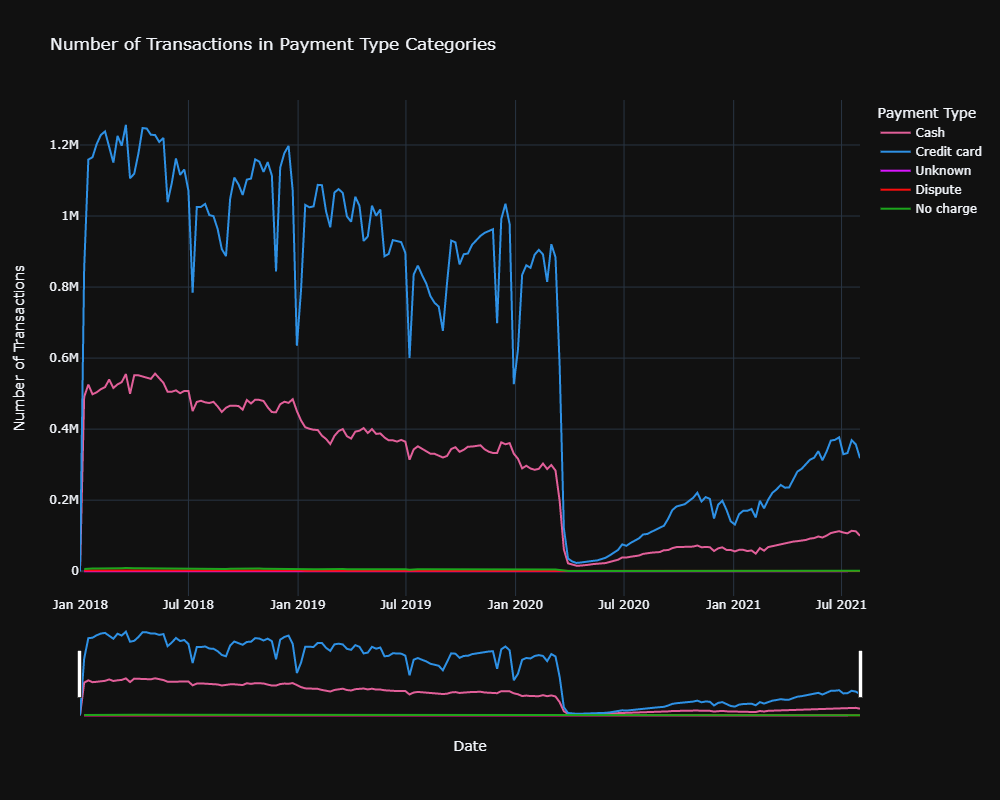
**This visualization shows a dramatic impact of the pandemic on the gross earning of NYC taxi services in the Yellow and Green categories. The trend follows the previous visualization of the volume of trips**.

* While the earnings are significantly more significant in the Yellow category, the trend in both categories is the same. Dramatic drop at the beginning of the pandemic in March 2020 with a slow recovery in Yellow category up to about 30% of the pre-pandemic volume. The Green category remains almost flat at probably about 20% of the pre-pandemic level.

* It is also worth mentioning that both categories were experiencing a downward trend over the last two pre-pandemic years. Possibly due to the competition from Uber, Lyft, and alike. However, the Yellow category felt the competing market's financial impact less than the Green category. Possibly their strategy to remain competitive working better.

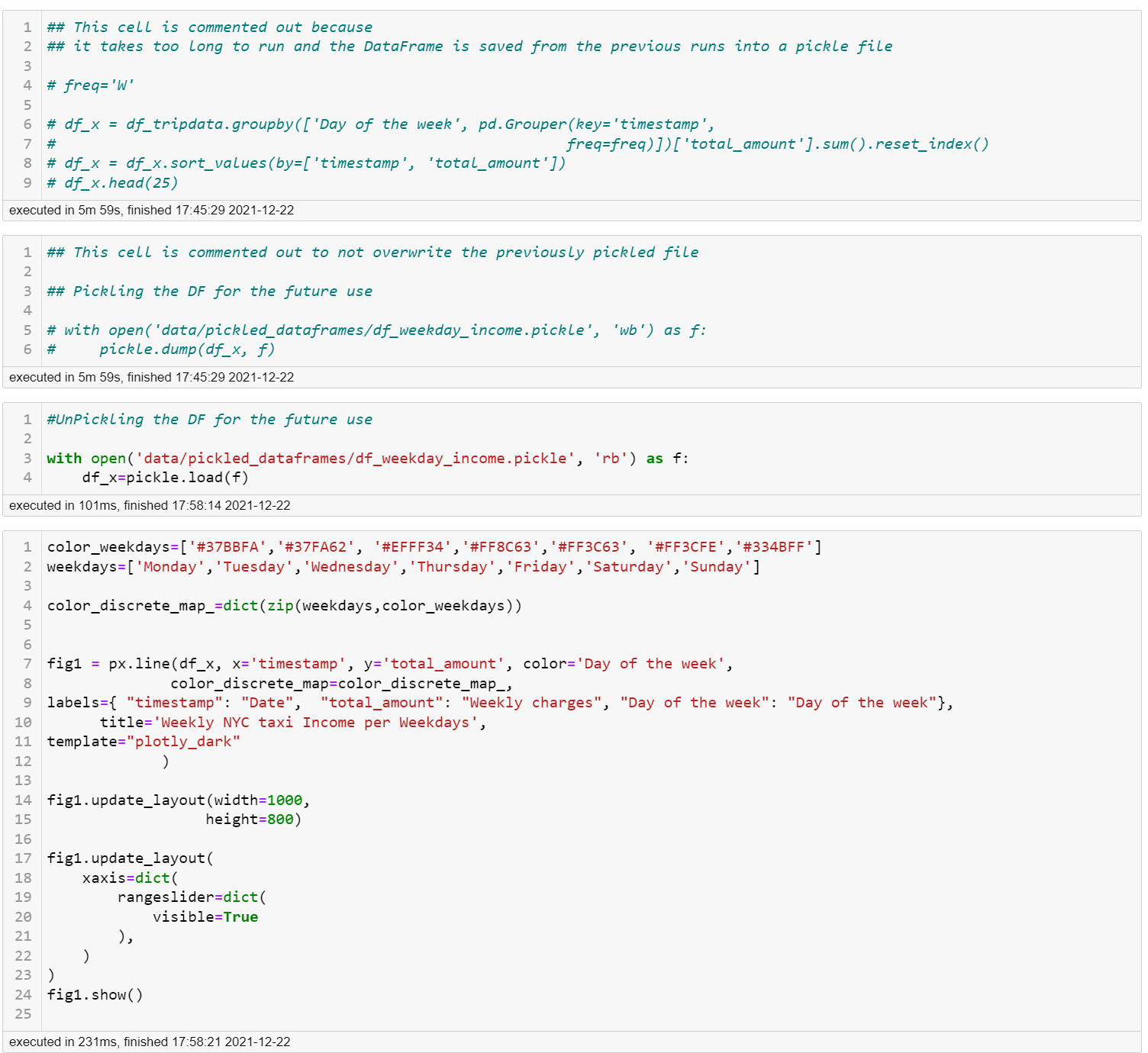
Number of transactions in each payment category between 2018 (January) and 2021 (July)

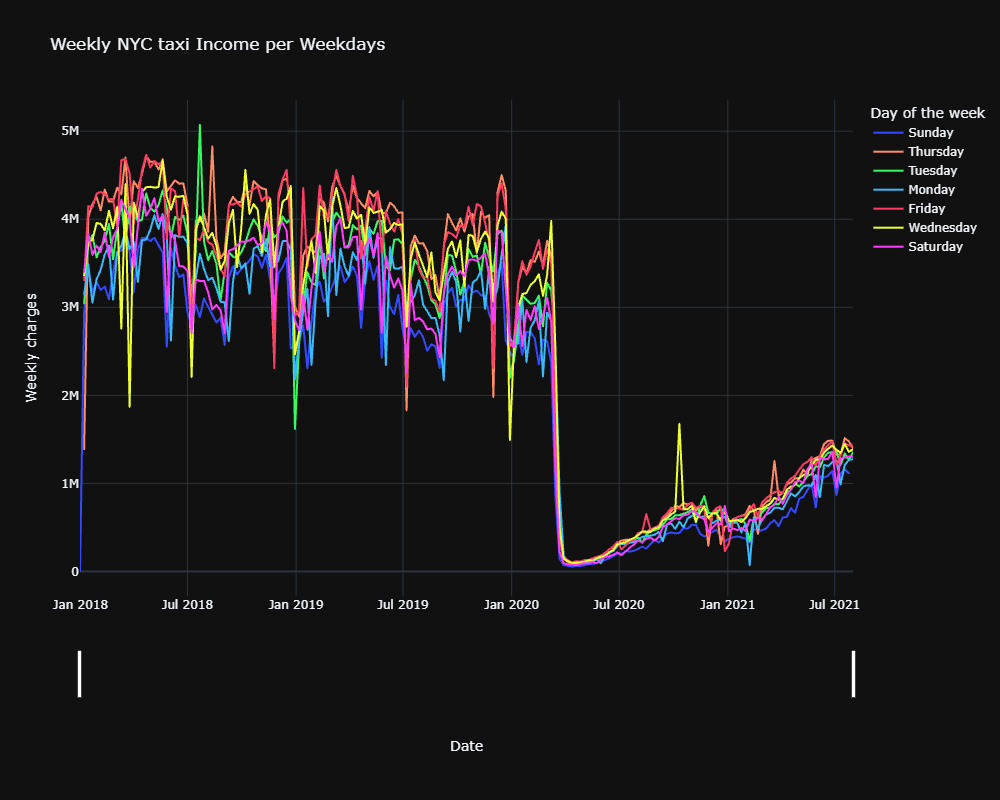




**The most transactions are made with credit cards followed by cash transactions. This tendency remained for after the pandemic onset with number of credit card transactions growing faster than the number of cash transactions**

Weekly NYC taxi income per days of the week:





**The visualization is too close to make any conclusions. It seems that Friday and Thursday are the days with the most income for NYC taxi days, and Sunday rides bring the least income. However, I will address this issue in the Numeric analysis section.**

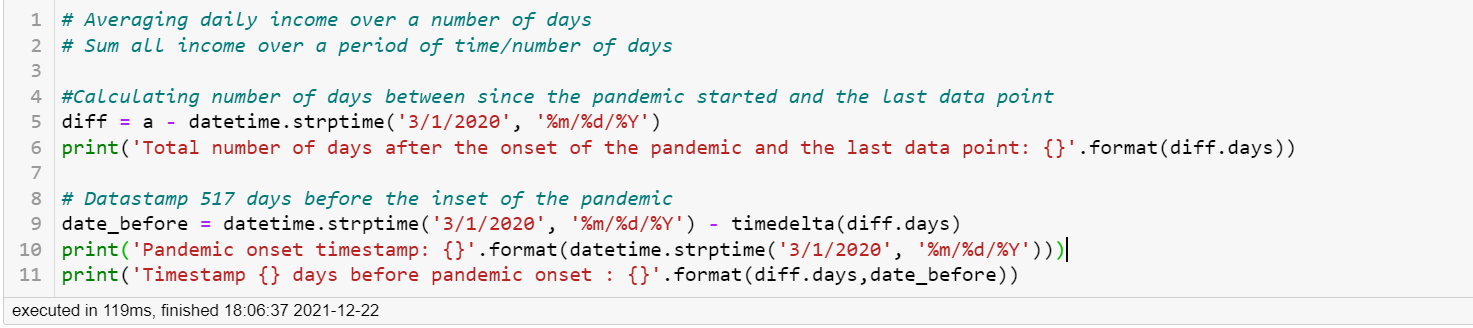
**Numeric Analysis:**

### What has been the economic impact of the pandemic on NYC taxi income over time?

*Averaging daily income in all categories over two periods:*

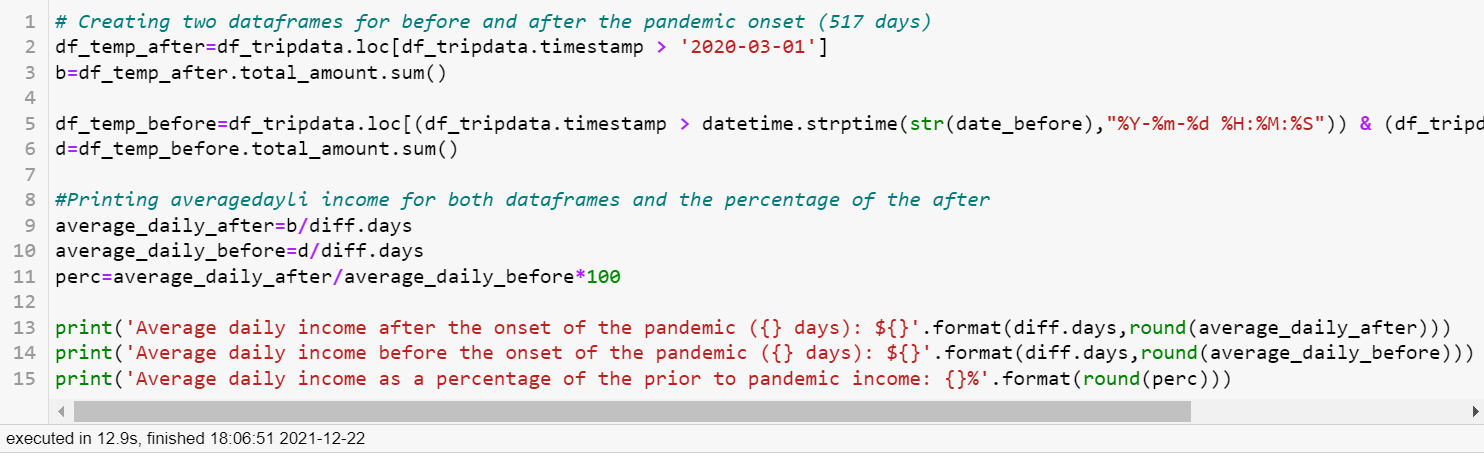
* It is impossible to calculate the revenue for this problem because we don't know the budget of NYC taxi; therefore I am calculating the income.
* the number of days since the onset of the pandemic measures on March 1, 2020 (approximate) and the last day in the data files provided
* the same number of days before March 1, 2020

**The total number of days after the onset of the pandemic and the last data point: 517. That is the number of days I will be using to average income and volume of transactions**

Total number of days after the onset of the pandemic and the last data point: 517

Pandemic onset timestamp: 2020-03-01 00:00:00

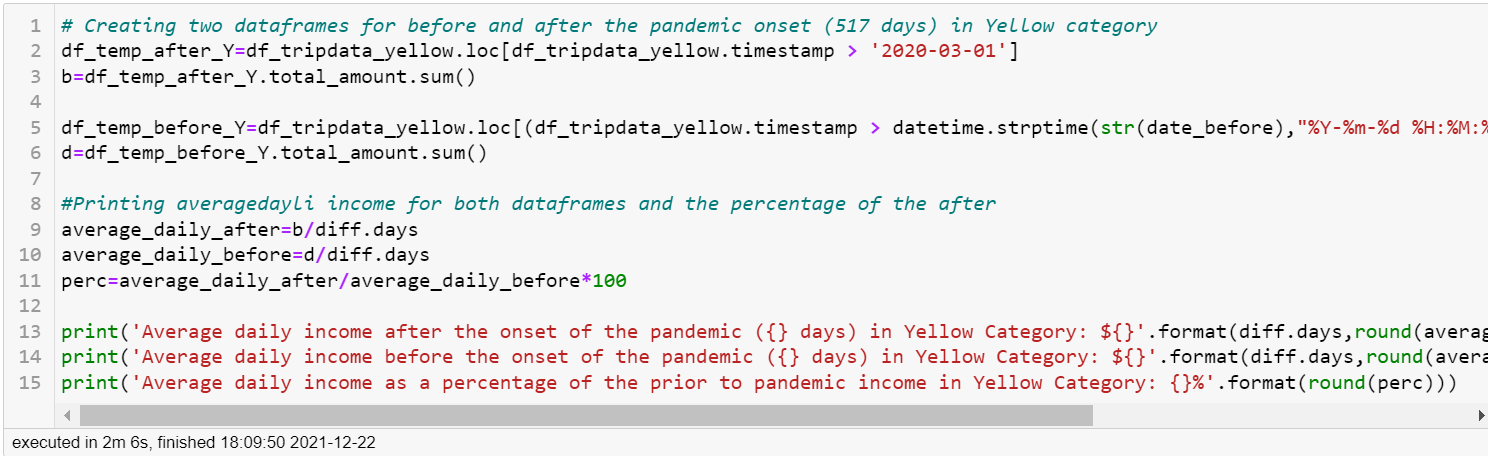
Timestamp 517 days before pandemic onset: 2018-10-01 00:00:00

Average daily income after the onset of the pandemic (517 days): $675301

Average daily income before the onset of the pandemic (517 days): $3474778

Average daily income as a percentage of the prior to pandemic income: 19%

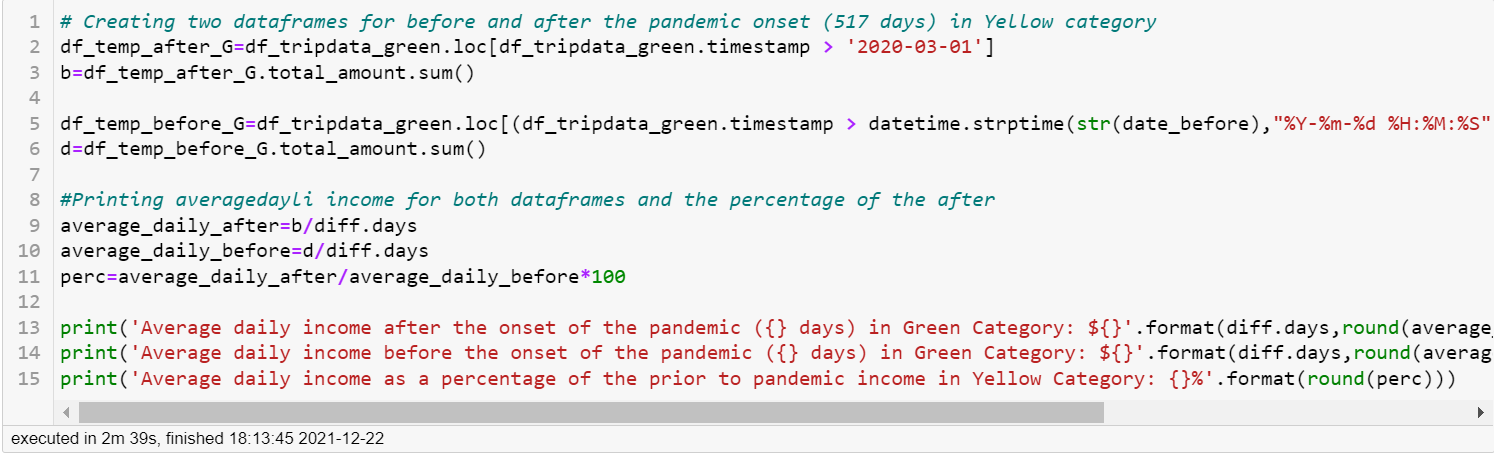
#### Pandemic impact on the daily income in green and yellow categories



Average daily income after the onset of the pandemic (517 days) in Yellow Category: $653354

Average daily income before the onset of the pandemic (517 days) in Yellow Category: $3280259

Average daily income as a percentage of the prior to pandemic income in Yellow Category: 20%



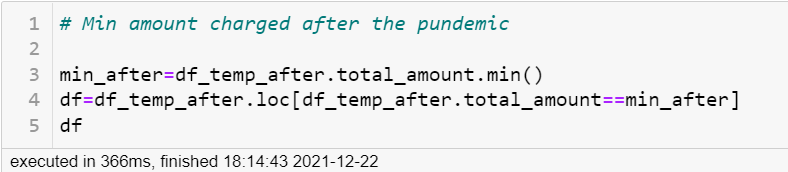
Average daily income after the onset of the pandemic (517 days) in Green Category: $21947

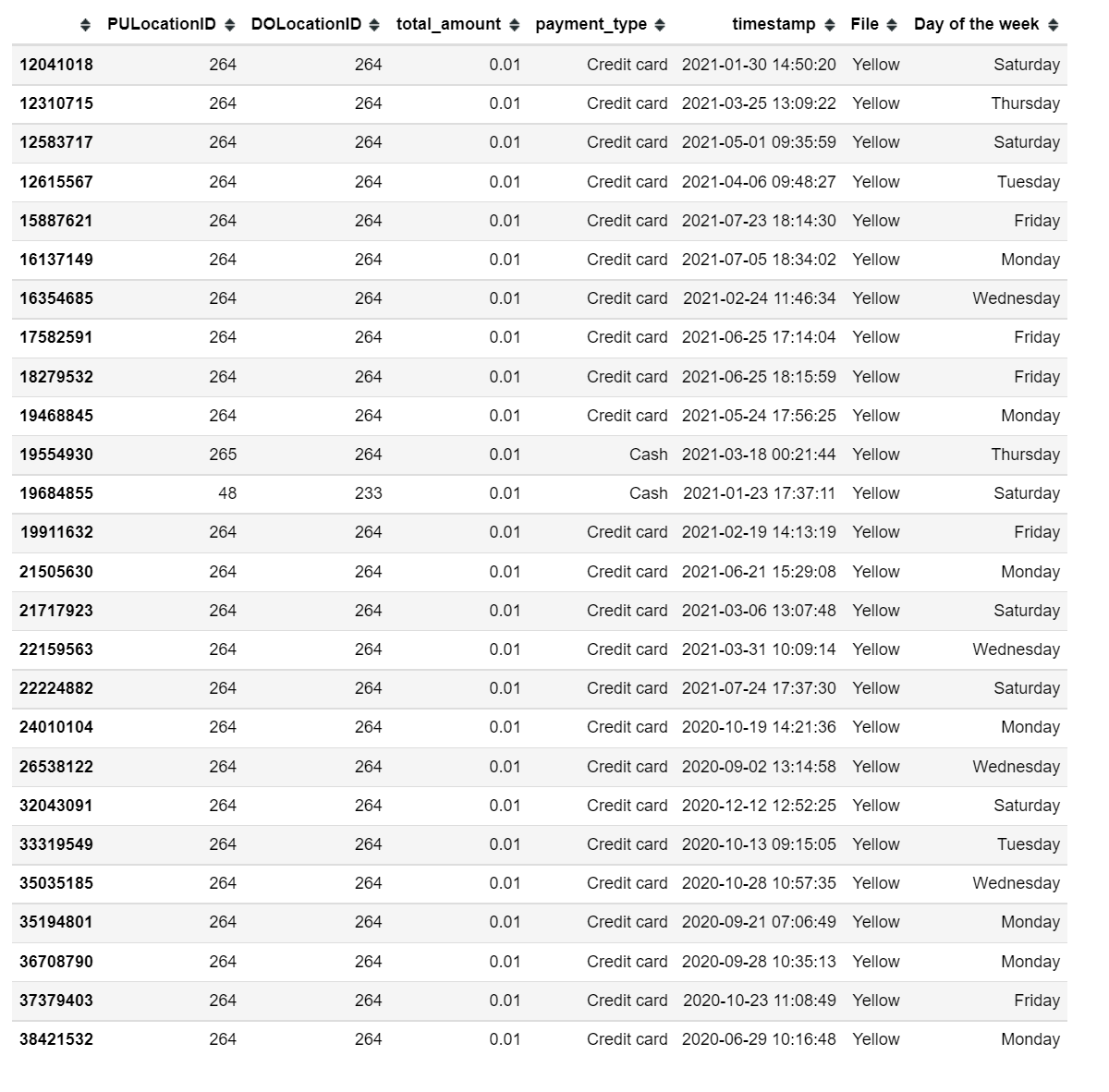
Average daily income before the onset of the pandemic (517 days) in Green Category: $194519

Average daily income as a percentage of the prior to pandemic income in Yellow Category: 11%

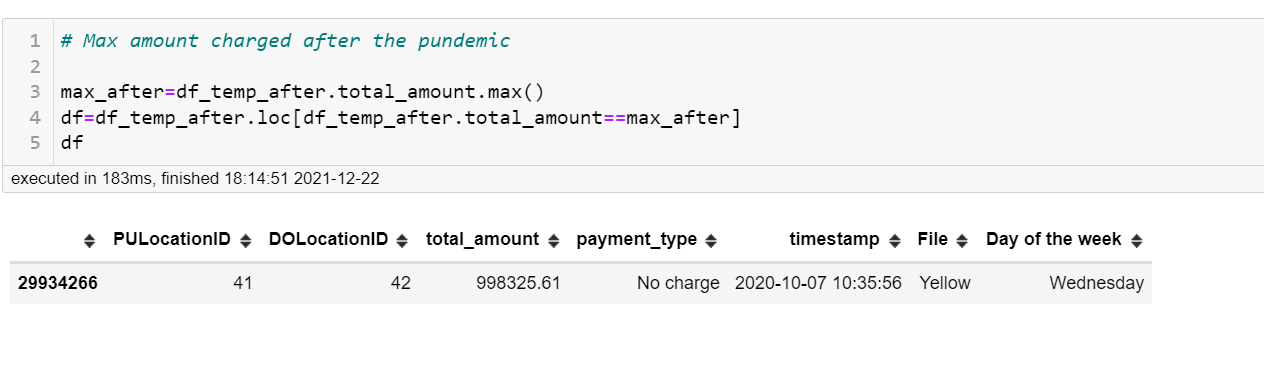
### Maximum and minimum charges before and after the pandemic?

**AFTER the onset of the pandemic**

****

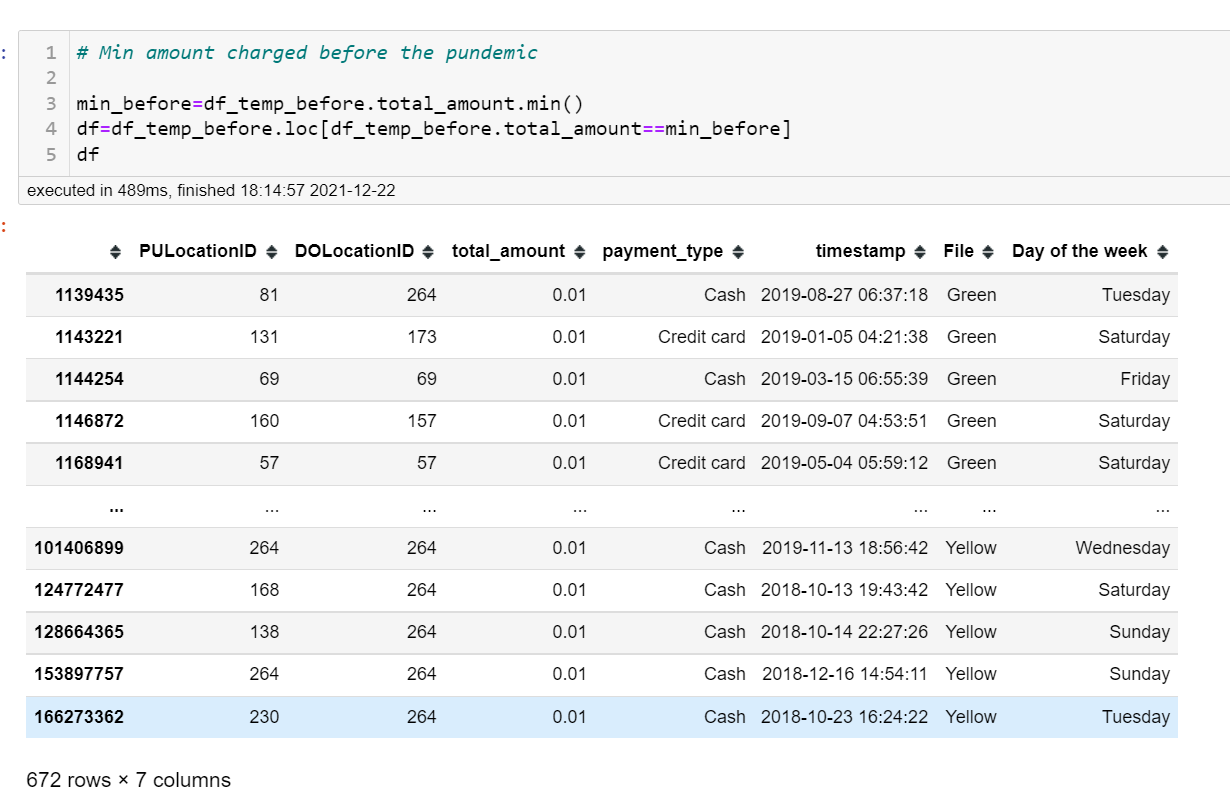
****

**There are 26 trips after the onset of the pandemic with $0.01 charges; most of them are between 264-264 zones (the same zone), with one standing out, it's between zones 48 and 233.**

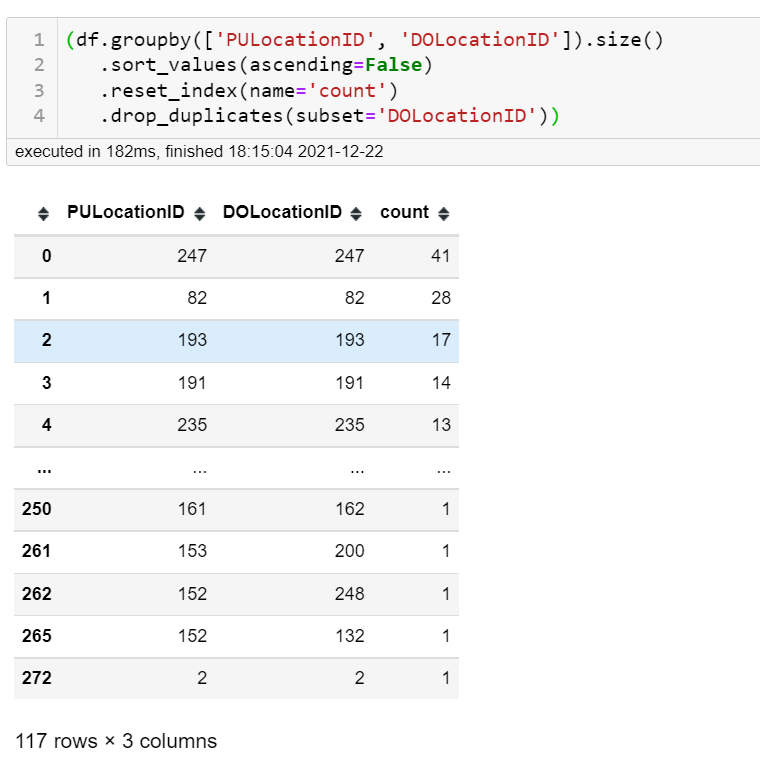
****

**There is just one trip after the onset of the pandemic with a charge of $998325.61; it's between zones 41 and 42. It is not clear why this insane amount has been charged. Possible to look at the original data for this date.**

**BEFORE the onset of the pandemic**

****

**There are 672 trips before the onset of the pandemic with $0.01 charges, and therefore a combination of a variety of zones.**

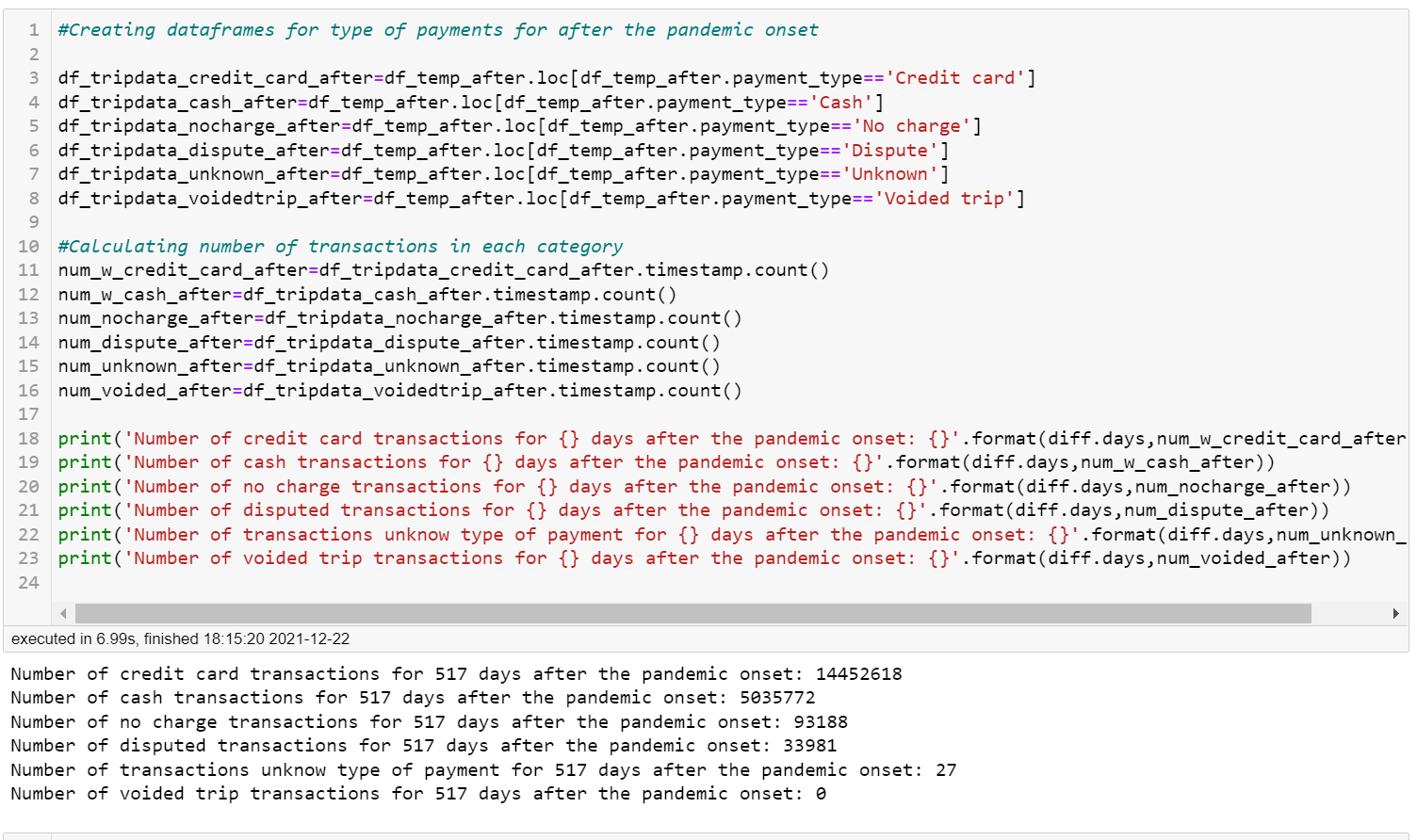
****

**There are 117 combinations of zones before the onset of the pandemic with $0.01 charges.**

**There is just one trip after the onset of the pandemic with a charge of $671124.94; it's between zones 138 and 181. It is not clear why this insane amount has been charged. Possible to look at the original data for this date.**

### What's the most popular payment method? Did this change because of the pandemic?

**AFTER the onset of the pandemic**



Number of credit card transactions for 517 days after the pandemic onset: 14452618

Number of cash transactions for 517 days after the pandemic onset: 5035772

Number of no charge transactions for 517 days after the pandemic onset: 93188

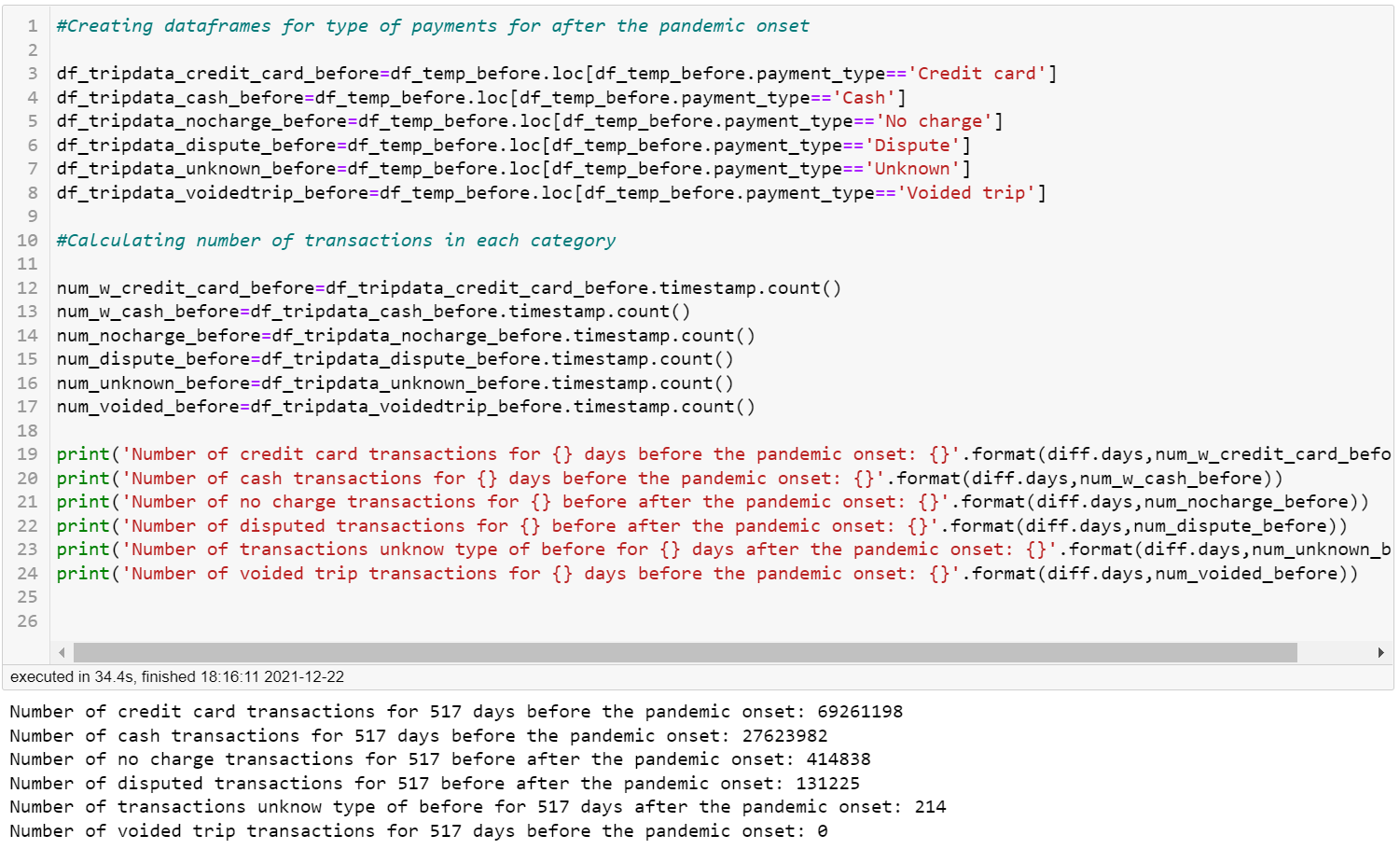
Number of disputed transactions for 517 days after the pandemic onset: 33981

Number of transactions unknown type of payment for 517 days after the pandemic onset: 27

Number of voided trip transactions for 517 days after the pandemic onset: 0

**Credit card is the most popular type of payment followed by cash payments after the pandemic onset.**

**BEFORE the onset of the pandemic**

Number of credit card transactions for 517 days before the pandemic onset: 69261198

Number of cash transactions for 517 days before the pandemic onset: 27623982

Number of no charge transactions for 517 before after the pandemic onset: 414838

Number of disputed transactions for 517 before after the pandemic onset: 131225

Number of transactions unknow type of before for 517 days after the pandemic onset: 214

Number of voided trip transactions for 517 days before the pandemic onset: 0

**Credit card is the most popular type of payment followed by cash payments after the pandemic onset**

### What’s the most expensive day of the week to travel on? Did this change because of the pandemic?

**It was unclear what it means “most expensive day of the week. Therefore, I calculated both the total amount of charges per day of the week for both periods**

**AND**

**An average trip cost per day of the week for both periods**

**1.**

**AFTER the onset of the pandemic**

****

* Total amount of charges on Monday over 517 days after the pandemic onset: $46851069.18
* Total amount of charges on Tuesday over 517 days after the pandemic onset: $51740456.04
* Total amount of charges on Wednesday over 517 days after the pandemic onset: $55562527.58
* Total amount of charges on Thursday over 517 days after the pandemic onset: $55111901.04
* Total amount of charges on Friday over 517 days after the pandemic onset: $54907415.91
* Total amount of charges on Saturday over 517 days after the pandemic onset: $46423669.32
* Total amount of charges on Sunday over 517 days after the pandemic onset: $38533511.46

**Over the 517 days since the onset of the pandemic the income per each day of the week is:**

* Monday: $46851069
* Tuesday: $51740456
* Wednesday: $55562528
* Thursday: $55111901
* Friday: $54907416
* Saturday: $46423669
* Sunday: $38533511

**The most expensive (bringing the most money) day of the week is Wednesday and the least expensive is Sunday**

**BEFORE the onset of the pandemic**

* ****Total amount of charges on Monday over 517 days before the pandemic onset: $238247633.69
* Total amount of charges on Tuesday over 517 days before the pandemic onset: $257295459.35
* Total amount of charges on Wednesday over 517 days before the pandemic onset: $271477547.33
* Total amount of charges on Thurday over 517 days before the pandemic onset: $283509485.42
* Total amount of charges on Friday over 517 days before the pandemic onset: $280996827.36
* Total amount of charges on Saturday over 517 days before the pandemic onset: $245110746.74
* Total amount of charges on Sunday over 517 days before the pandemic onset: $219822658.5

**Over the 517 days before the onset of the pandemic the income per each day of the week is:**

* Monday: $238247633
* Tuesday: $257295459
* Wednesday: $271477547
* Thursday: $283509485
* Friday: $280996827
* Saturday: $245110747
* Sunday: $219822659

**The most expensive (bringing the most money) day of the week is Thursday and the least expensive is Sunday**

**2.**

**Average trip cost**

**AFTER the onset of the pandemic**

****

Average trip cost on Monday over 517 days after the pandemic onset: $17.7

Average trip cost on Tuesday over 517 days after the pandemic onset: $17.48

Average trip cost on Wednesday over 517 days after the pandemic onset: $17.99

Average trip cost on Thursday over 517 days after the pandemic onset: $17.86

Average trip cost on Friday over 517 days after the pandemic onset: $17.82

Average trip cost on Saturday over 517 days after the pandemic onset: $17.51

Average trip cost on Sunday over 517 days after the pandemic onset: $18.34

**The most expensive day of the week after the pandemic onset is Wednesday and the least expensive is Tuesday. H the difference is negligible.**

**BEFORE the onset of the pandemic**

****

Average trip cost on Monday over 517 days before the pandemic onset: $18.7

Average trip cost on Tuesday over 517 days before the pandemic onset: $18.38

Average trip cost on Wednesday over 517 days before the pandemic onset: $18.67

Average trip cost on Thursday over 517 days before the pandemic onset: $18.92

Average trip cost on Friday over 517 days before the pandemic onset: $18.69

Average trip cost on Saturday over 517 days before the pandemic onset: $17.35

Average trip cost on Sunday over 517 days before the pandemic onset: $18.31

**The most expensive day of the week prior to pandemic is Thursday and the least expensive is Saturday. However, the difference is negligible.**