

data\_prepare.ipynb

In data\_prepare.ipynb, we randomly chose 240 thousand of trips information from the whole 2015 yellow cab dataset, which has more than 100 million rows.

definite\_region.ipynb

I assigned different taxi trips to the nearest citibike station and joined the table of citibike and taxi data according to the citibike station id. I assigned each taxi trip data a citibike station id according to its nearest citibike station.

statistics on region.ipynb

I just do some statistics on the region data. For example, the average cost per mile for taxi trip on different months within a year and different days within a week.

Visualization Tools:

1. cartoDB
2. plotly