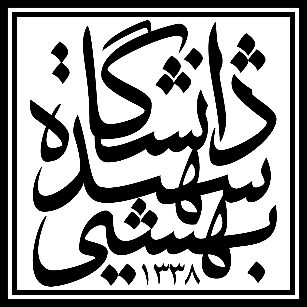
****

**گزارش تمرین 1**

96222096

سید حسین محمدی

\* Note: some extra histplots have been placed in “EX1\_x/working\_data/histplot”

----------------------------------------------------------------------------------

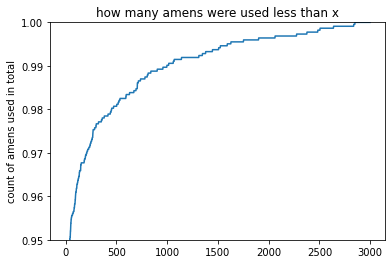
Part A: AirBnB dataset

After wrangling with the dataset, I made 4 files containing the following:

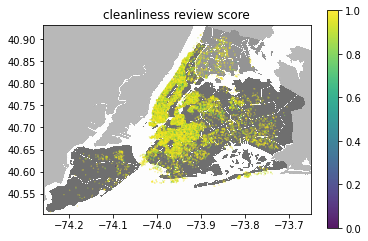
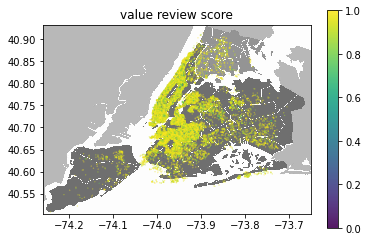
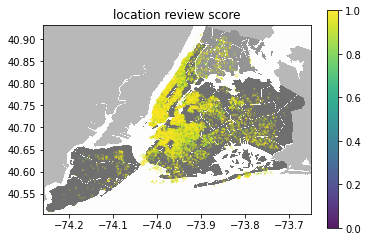
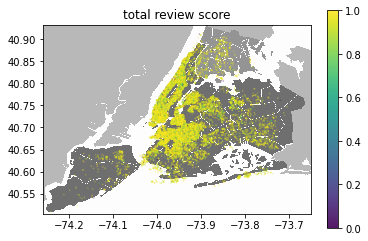
* separated\_cols.csv: removed unnecessary cols. reformatted some categorial cols into numeric.
* unique\_counts.csv: count of each unique value in separated\_cols.csv. some entries in some cols and some of their entries were removed for being too rare. Some numerical data were rounded. Some cols containing arrays, were unzipped.
* mean\_chart.csv: 10000 rows of sampled means and std.
* normality\_chart.csv: D’Agostino & Pearson test results.

The following graphs were made:

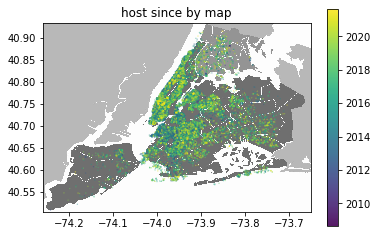
* This shows, 95% of amens types listed, appear less than 50 times and thus can be deleted

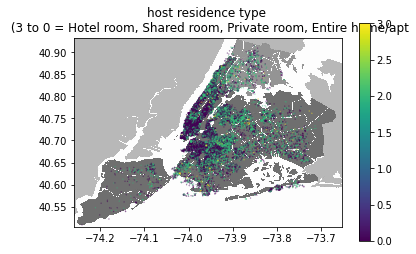


* Review score for each category by location on map

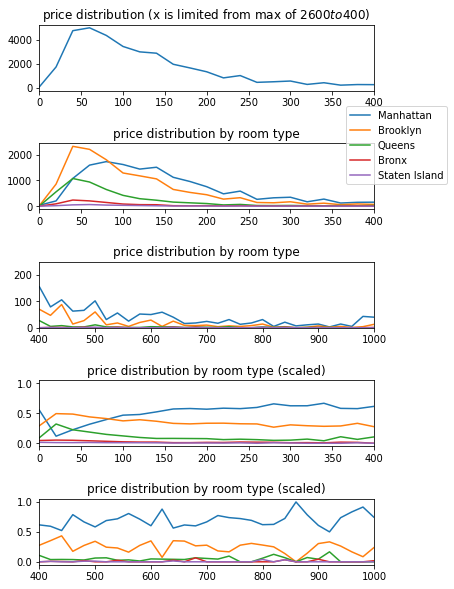
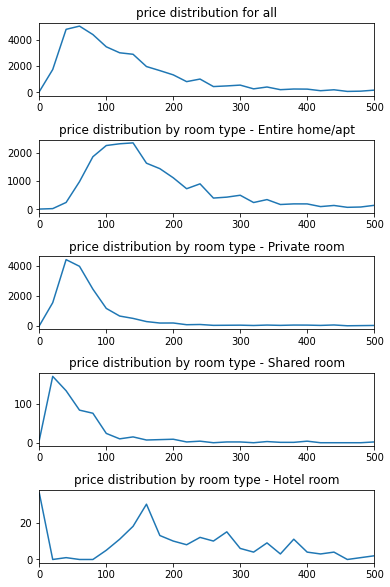
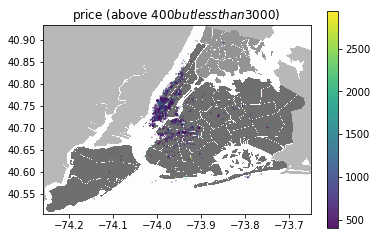
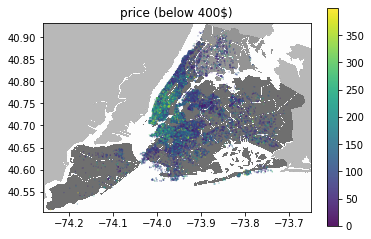


* Time of registering on map

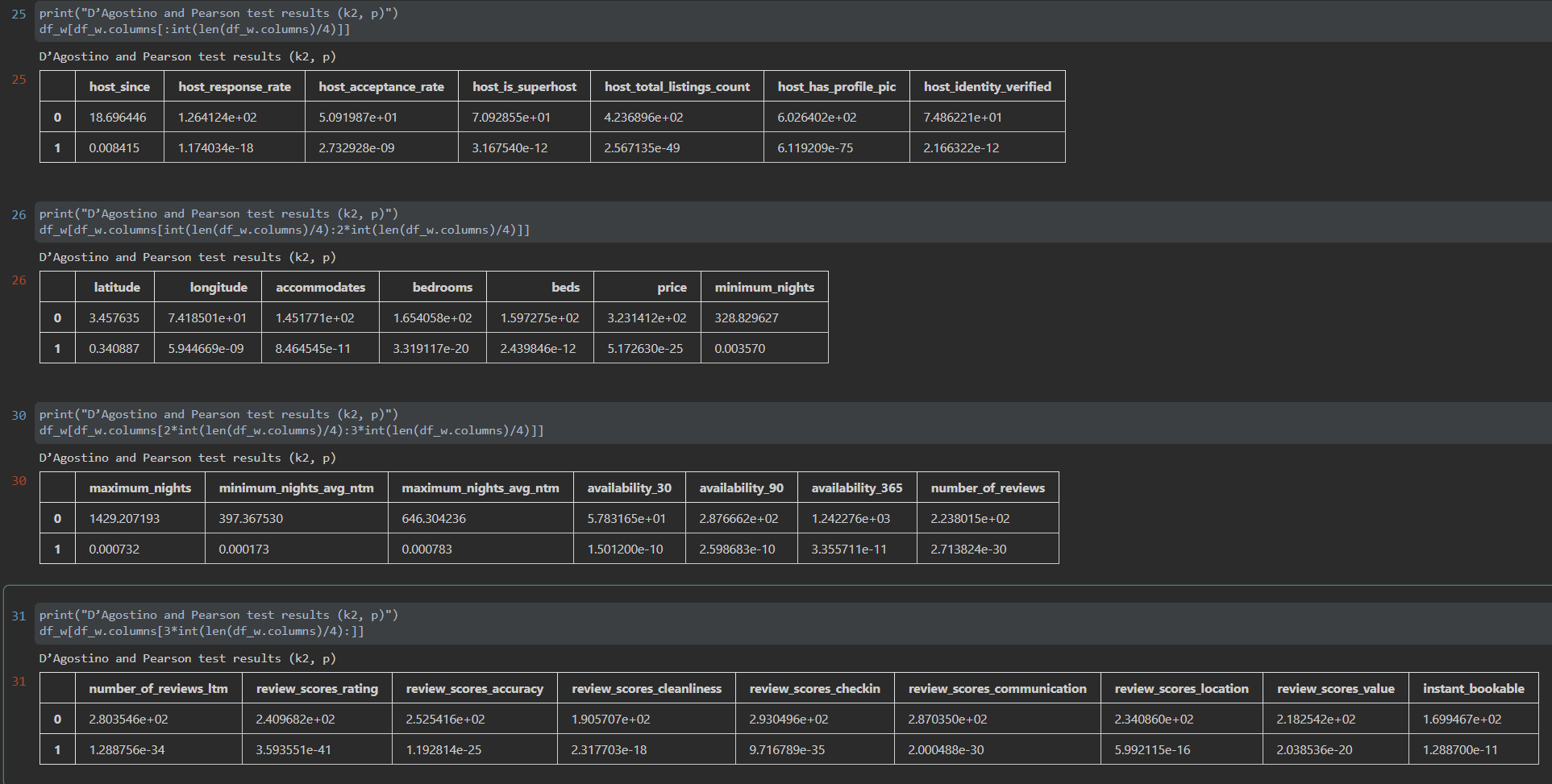


Residence type on map 

* prices on map



* D’Agostino & Pearson test results



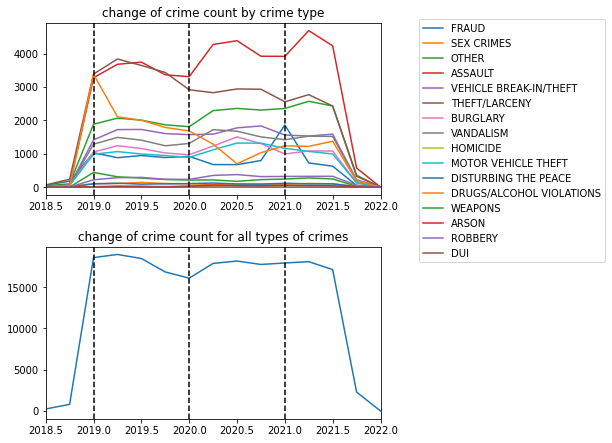
Part B: Crime Dataset

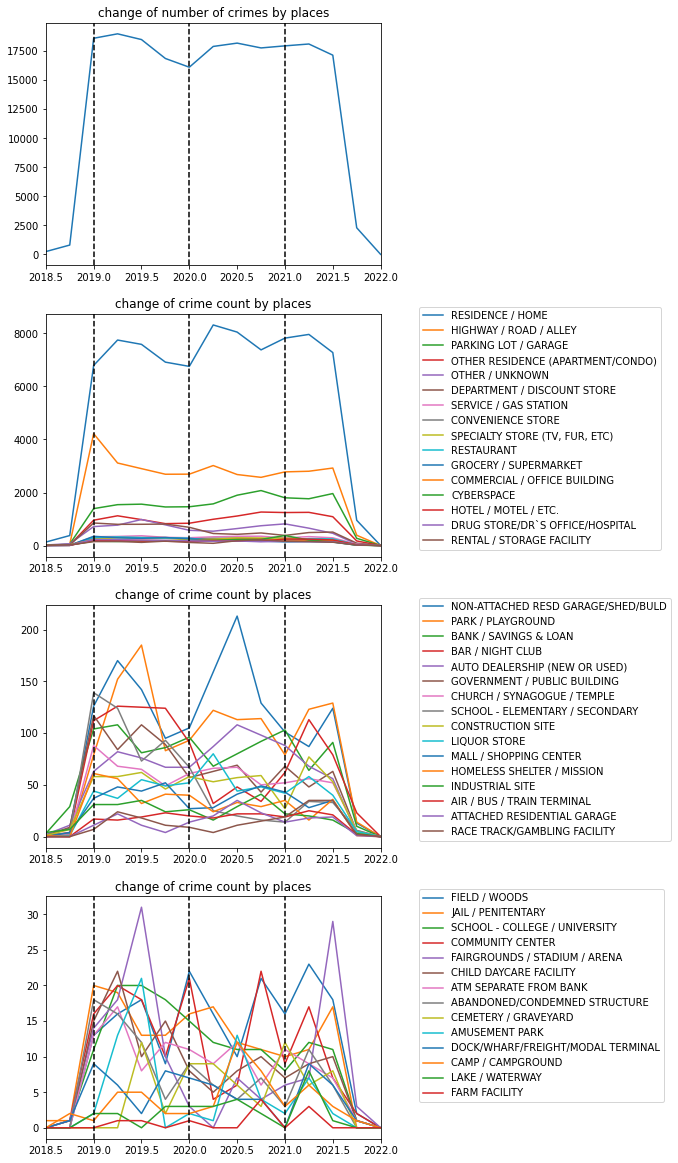
After wrangling with the dataset, I made 4 files containing the following:

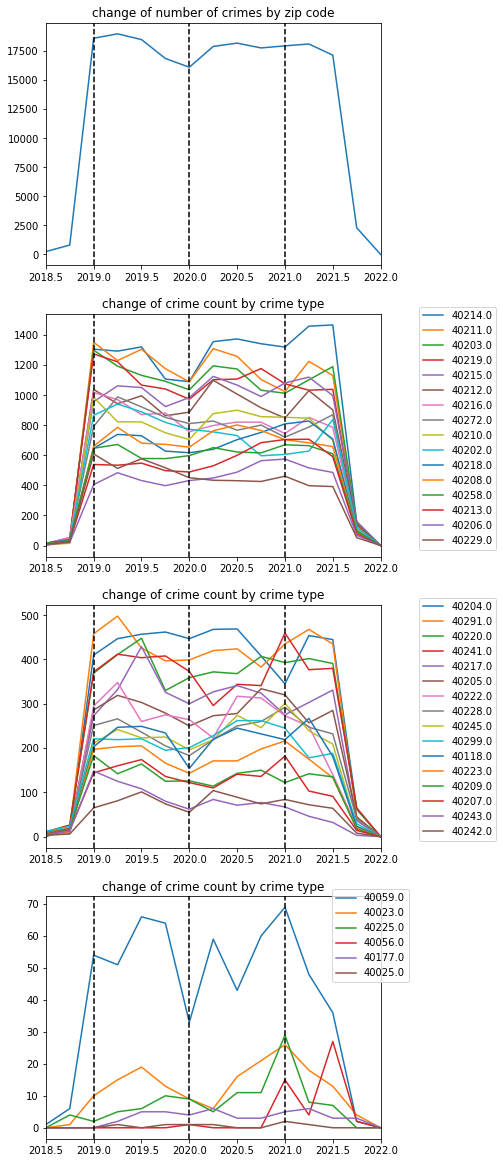
* separated\_cols.csv: removed unnecessary cols. reformatted some categorial cols into numeric.
* unique\_counts.csv: count of each unique value in separated\_cols.csv. some entries in some cols were removed for being too rare. Some numerical data were rounded
* mean\_chart.csv: 10000 rows of sampled means and std.
* normality\_chart.csv: D’Agostino & Pearson test results.

The following graphs were made:

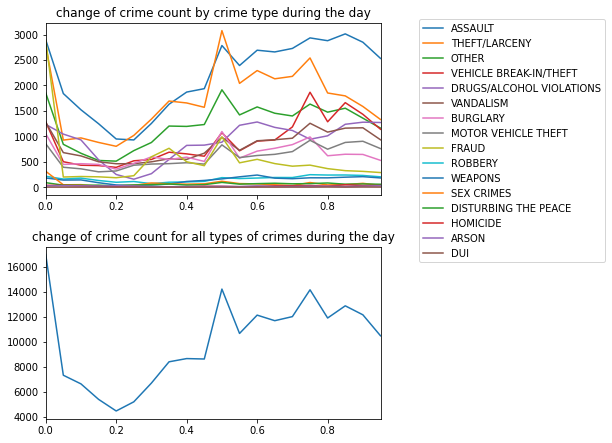
* change of the number of crime occurrences during the past few years:

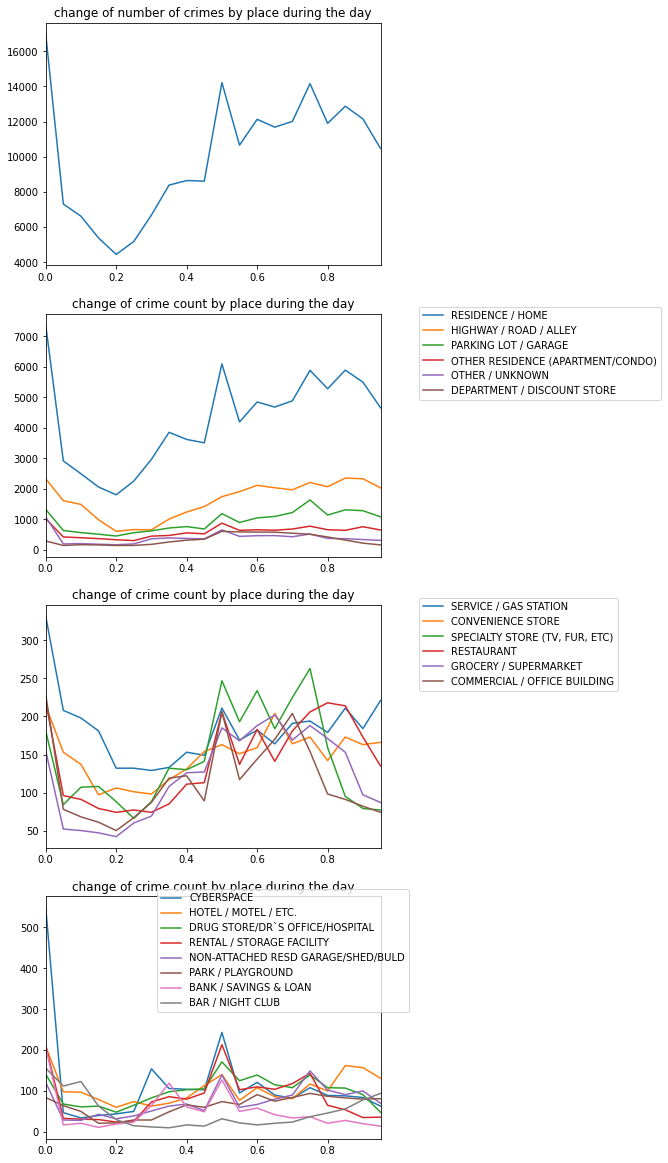


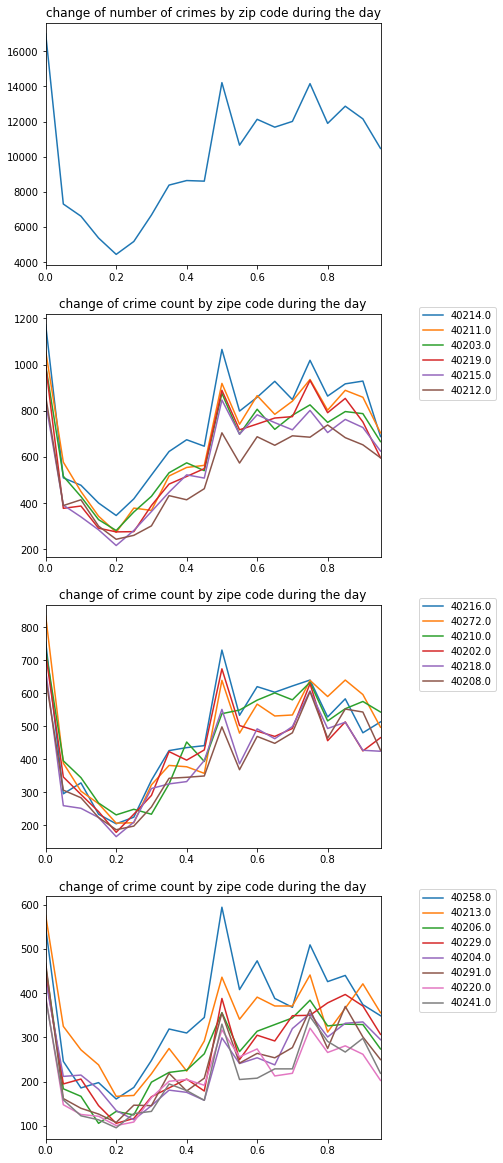




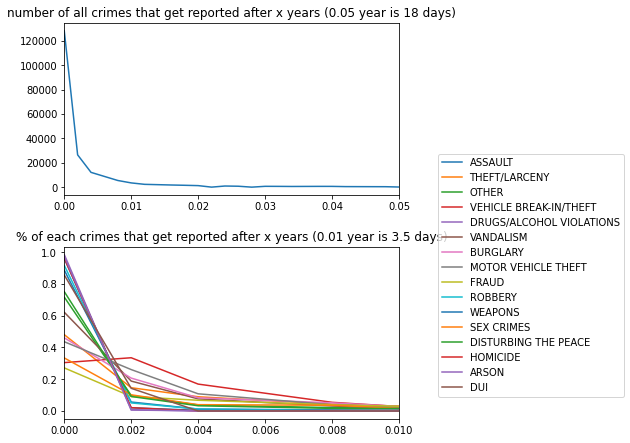
* change of the number of crimes occurrences during a day cycle:



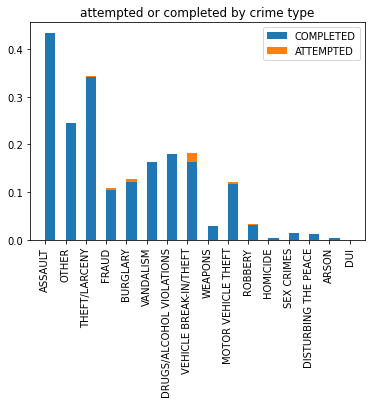


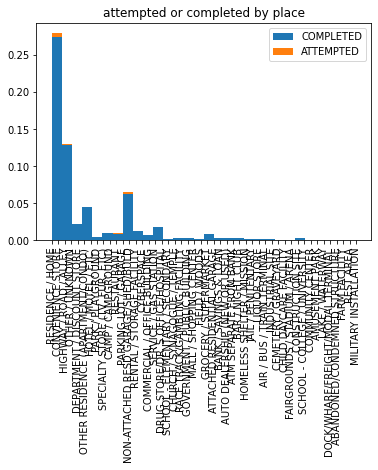


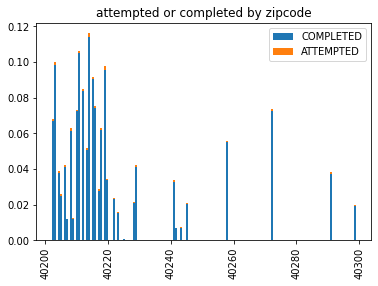
* how fast crimes get reported after occurring:

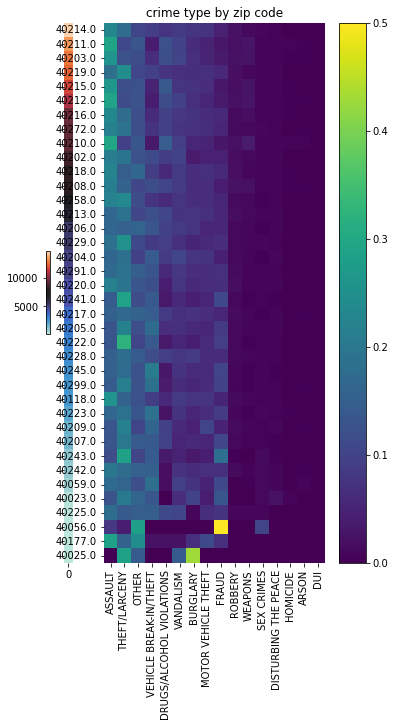
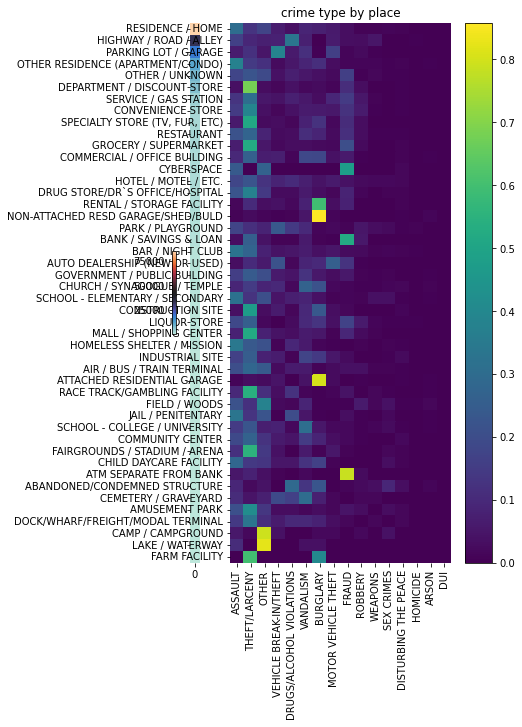


* attempted or completed cases:







* each row indicates the % of crimes in that zip code (the line on the left is the total number of crimes):  
   
* each row indicates the % of crimes in that place (the line on the left is the total number of crimes):  
  
* each row indicates the % of crimes in places in that zipcode (the line on the left is the total number of crimes):

