گزارش تمرین 1

96222096

سید حسین محمدی

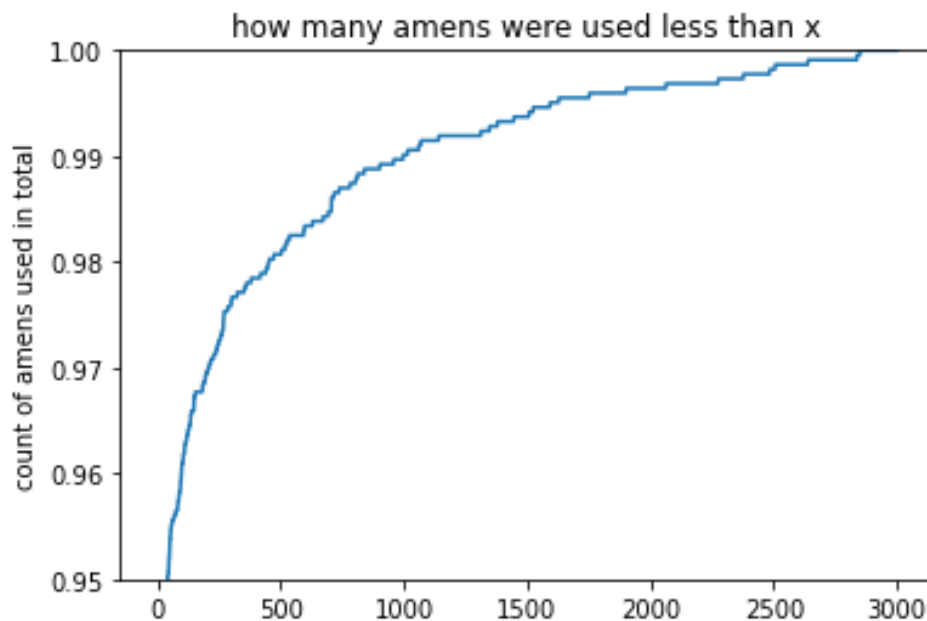
Part A: AirBnB dataset

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.
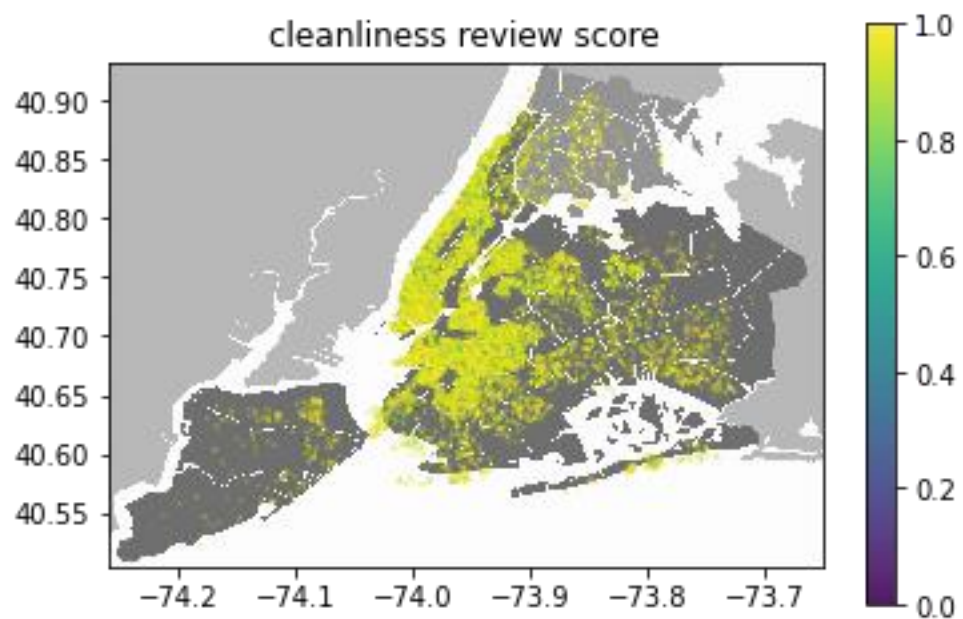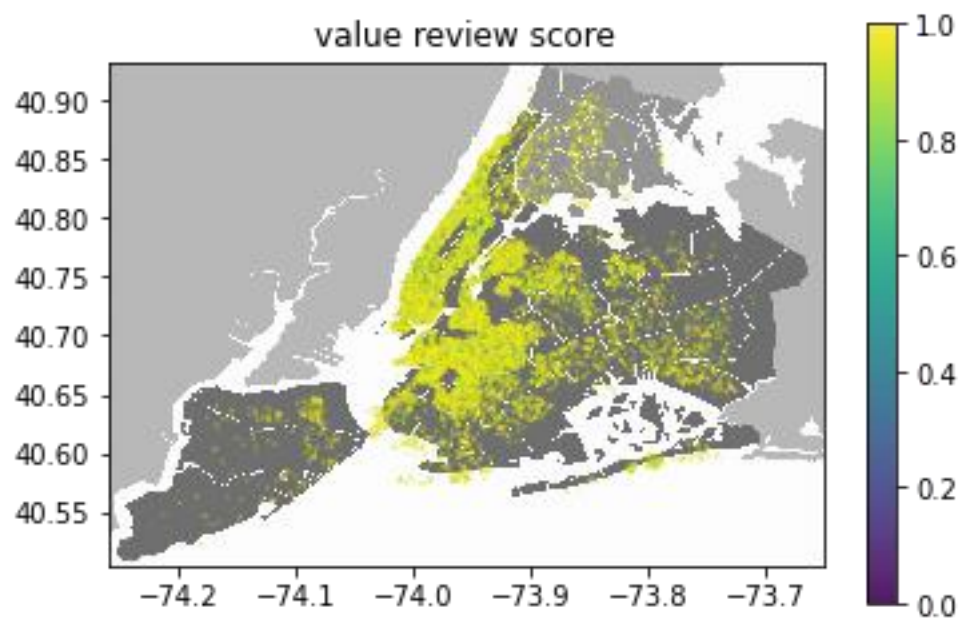- normality_chart.csv: D'Agostino & Pearson test results.

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

total review score

location review score

value review score

cleanliness review score
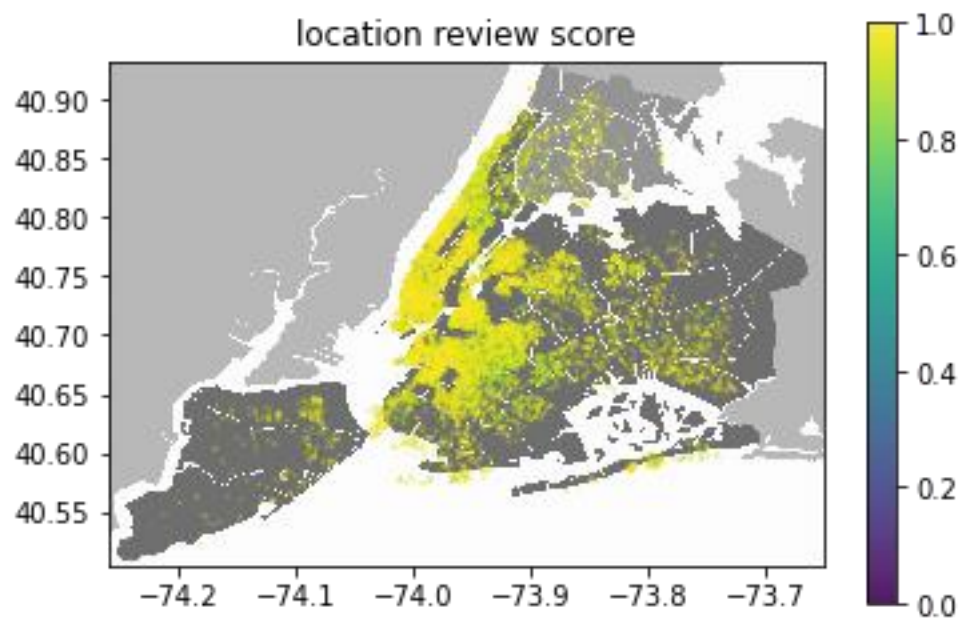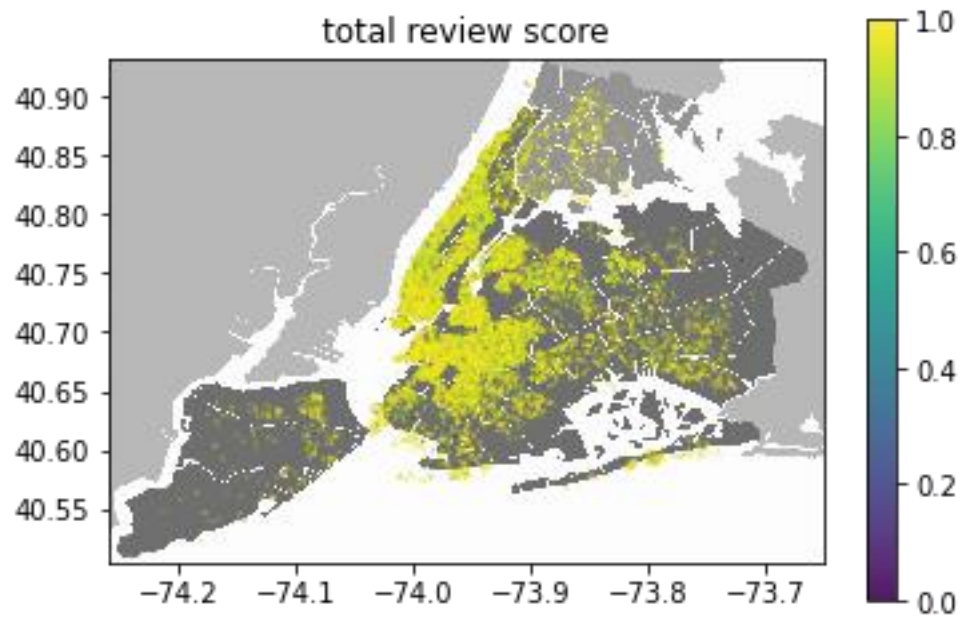
- Time of registering on map

host since by map

Residence type on map

host residence type
(3 to 0 = Hotel room, Shared room, Private room, Entire home/apt

- prices on map

price (below 400$)

price (above 400$ but less than 3000)
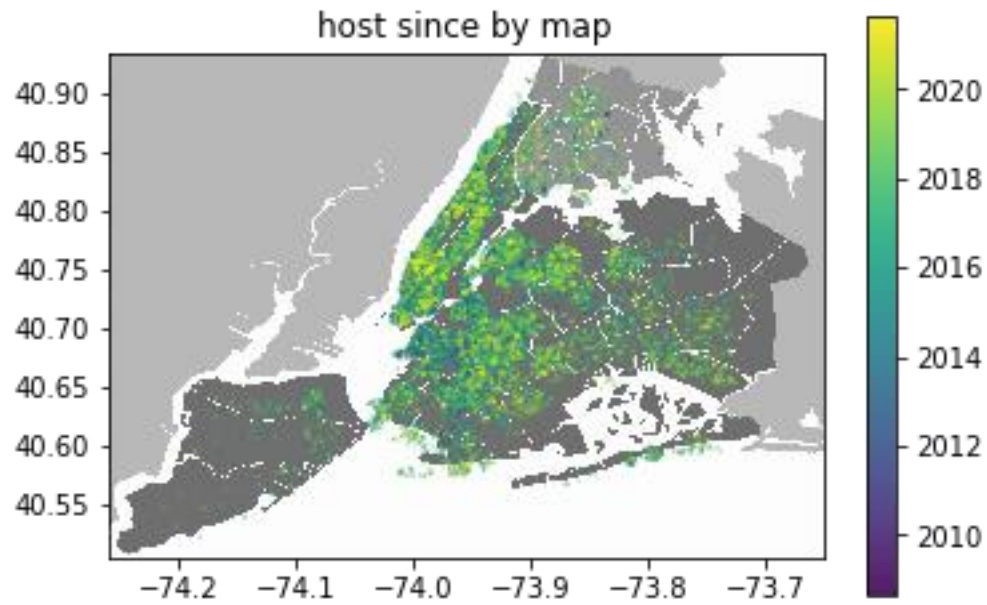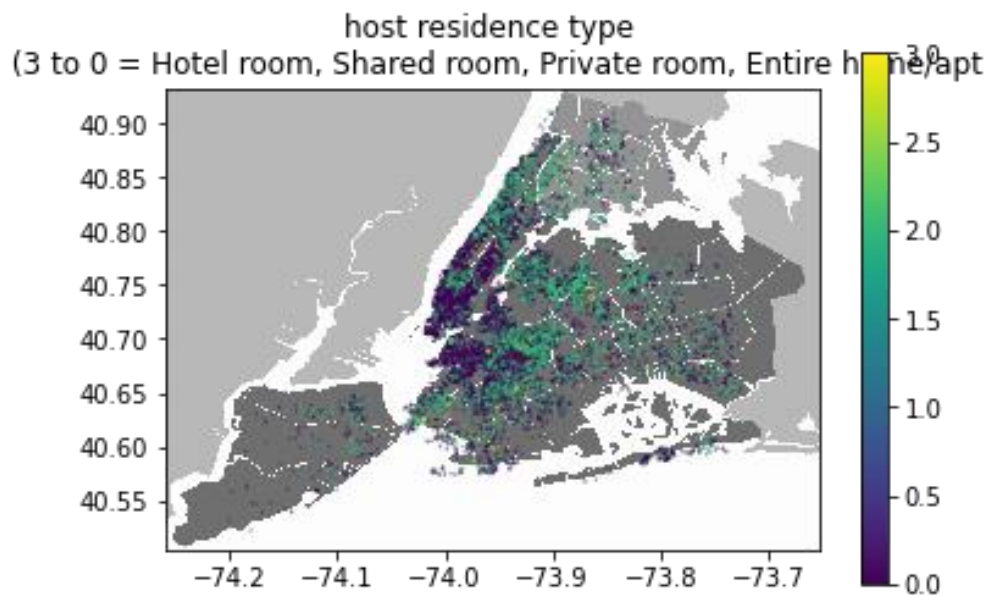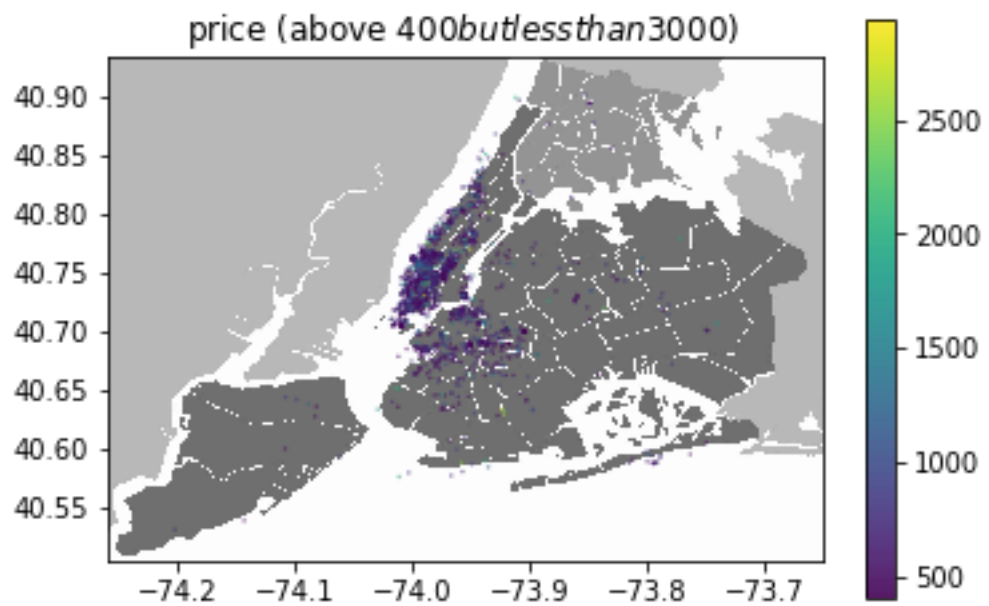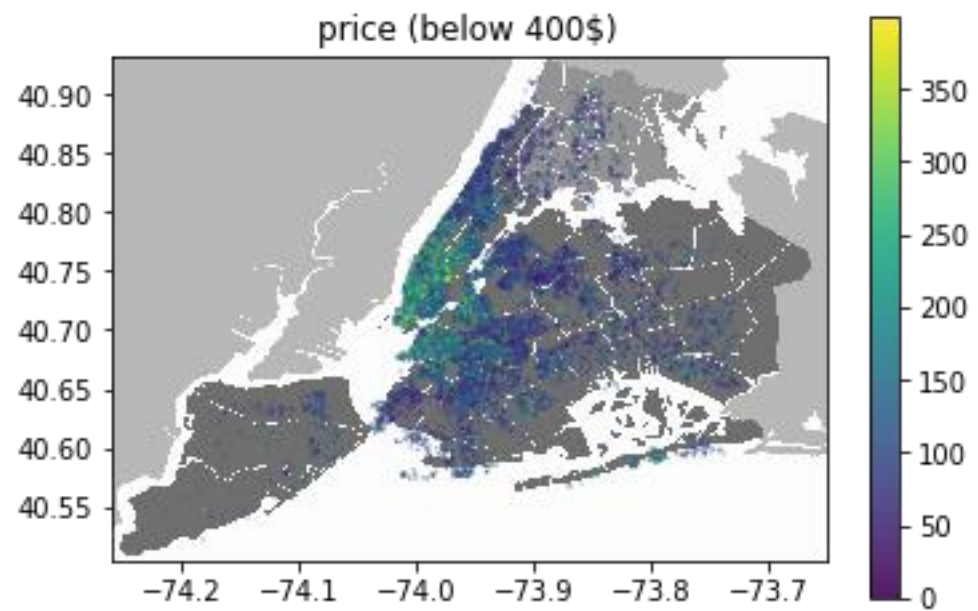
price distribution for all
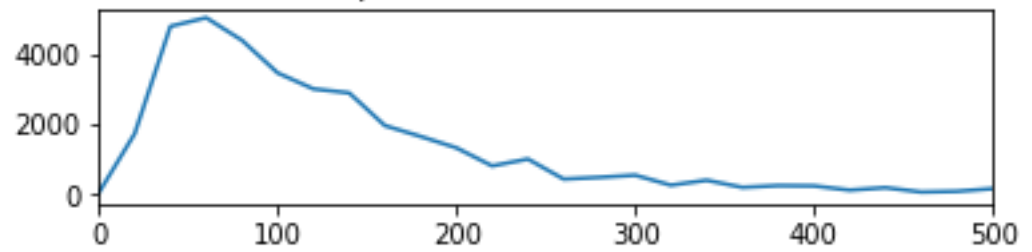
price distribution by room type - Entire home/apt
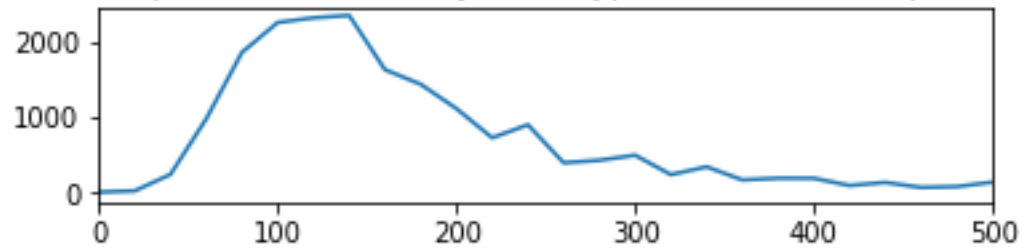
price distribution by room type - Private room
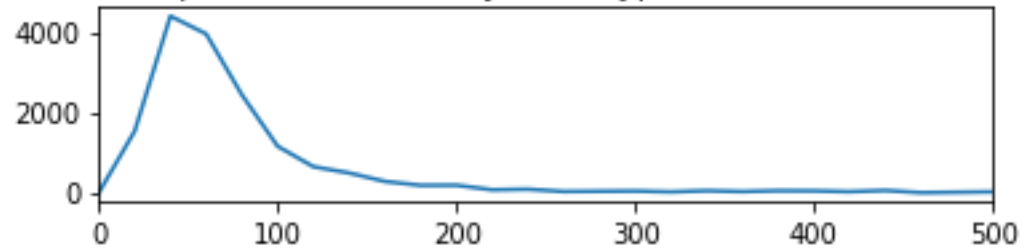
price distribution by room type - Shared room

price distribution by room type - Hotel room

price distribution (x is limited from max of 2600 to 400)

price distribution by room type

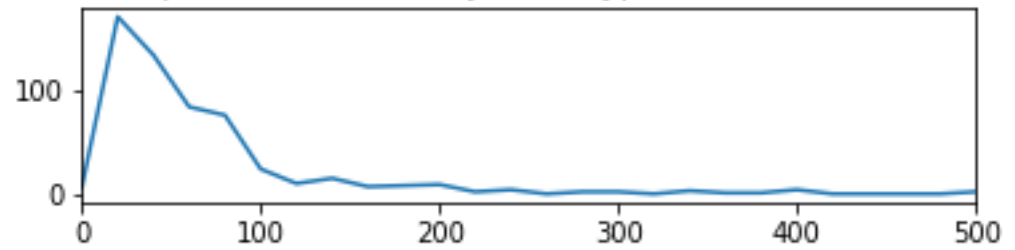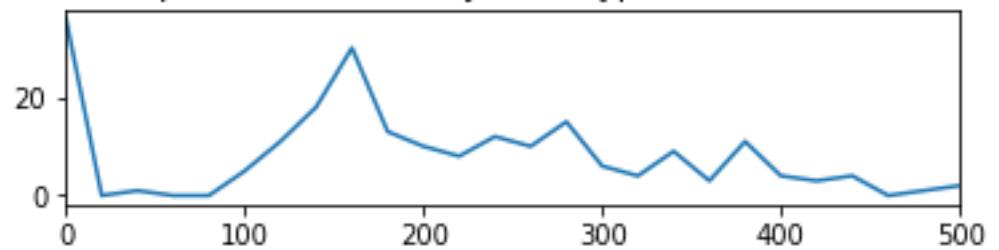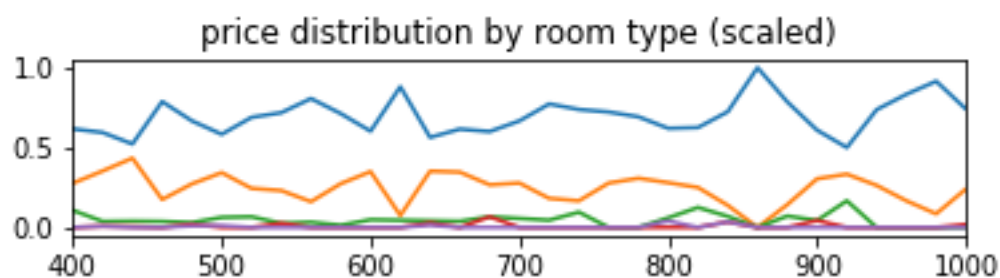## price distribution by room type

price distribution by room type (scaled)

price distribution by room type (scaled)

Manhattan
Brooklyn
Queens
Bronx
Staten Island
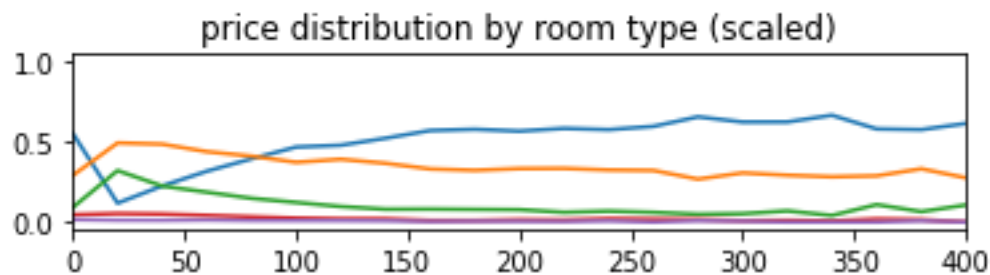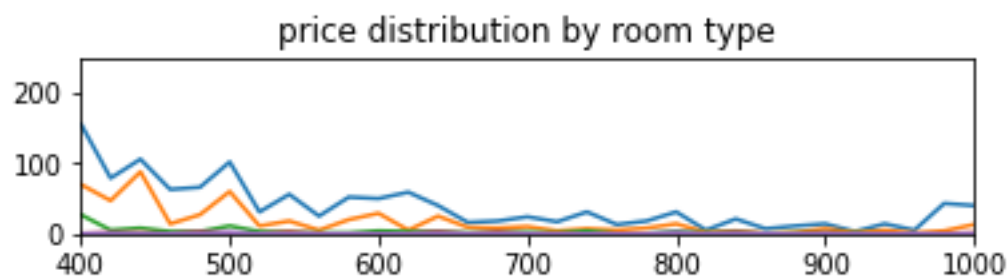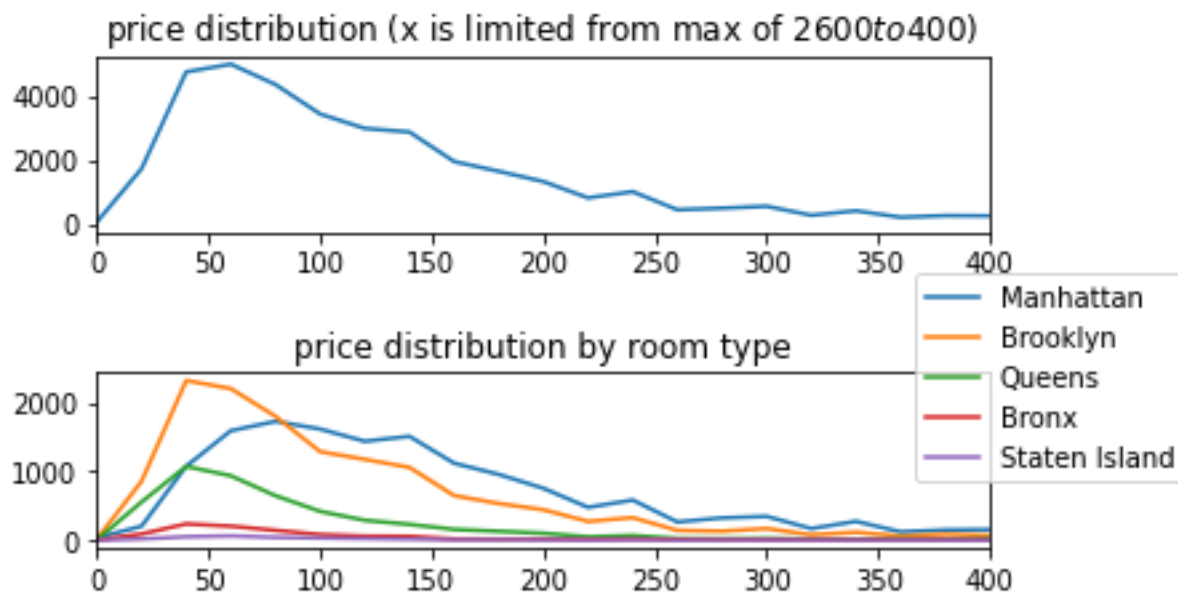
- D'Agostino & Pearson test results

```
25 print("D'Agostino and Pearson test results (k2, p)")
   df_w[df_w.columns[:int(len(df_w.columns)/4)]]
```

D'Agostino and Pearson test results (k2, p)

25		host_since	host_response_rate	host_acceptance_rate	host_is_superhost	host_total_listings_count	host_has_profile_pic	host_identity_verified
	0	18.696446	1.264124e+02	5.091987e+01	7.092855e+01	4.236896e+02	6.026402e+02	7.486221e+01
	1	0.008415	1.174034e-18	2.732928e-09	3.167540e-12	2.567135e-49	6.119209e-75	2.166322e-12

```
26 print("D'Agostino and Pearson test results (k2, p)")
   df_w[df_w.columns[int(len(df_w.columns)/4):2*int(len(df_w.columns)/4)]]
```

D'Agostino and Pearson test results (k2, p)

26		latitude	longitude	accommodates	bedrooms	beds	price	minimum_nights
	0	3.457635	7.418501e+01	1.451771e+02	1.654058e+02	1.597275e+02	3.231412e+02	328.829627
	1	0.340887	5.944669e-09	8.464545e-11	3.319117e-20	2.439846e-12	5.172630e-25	0.003570

```
30 print("D'Agostino and Pearson test results (k2, p)")
   df_w[df_w.columns[2*int(len(df_w.columns)/4):3*int(len(df_w.columns)/4)]]
```

D'Agostino and Pearson test results (k2, p)

30		maximum_nights	minimum_nights_avg_ntm	maximum_nights_avg_ntm	availability_30	availability_90	availability_365	number_of_reviews
	0	1429.207193	397.367530	646.304236	5.783165e+01	2.876662e+02	1.242276e+03	2.238015e+02
	1	0.000732	0.000173	0.000783	1.501200e-10	2.598683e-10	3.355711e-11	2.713824e-30

```
31 print("D'Agostino and Pearson test results (k2, p)")
   df_w[df_w.columns[3*int(len(df_w.columns)/4):]]
```

D'Agostino and Pearson test results (k2, p)

| 31		number_of_reviews_ltm	review_scores_rating	review_scores_accuracy	review_scores_cleanliness	review_scores_checkin	review_scores_communication	review_scores_location	review_scores_value	instant_bookable
	0	2.803546e+02	2.409682e+02	2.525416e+02	1.905707e+02	2.930496e+02	2.870350e+02	2.340860e+02	2.182542e+02	1.699467e+02
	1	1.288756e-34	3.593551e-41	1.192814e-25	2.317703e-18	9.716789e-35	2.000488e-30	5.992115e-16	2.038536e-20	1.288700e-11

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:
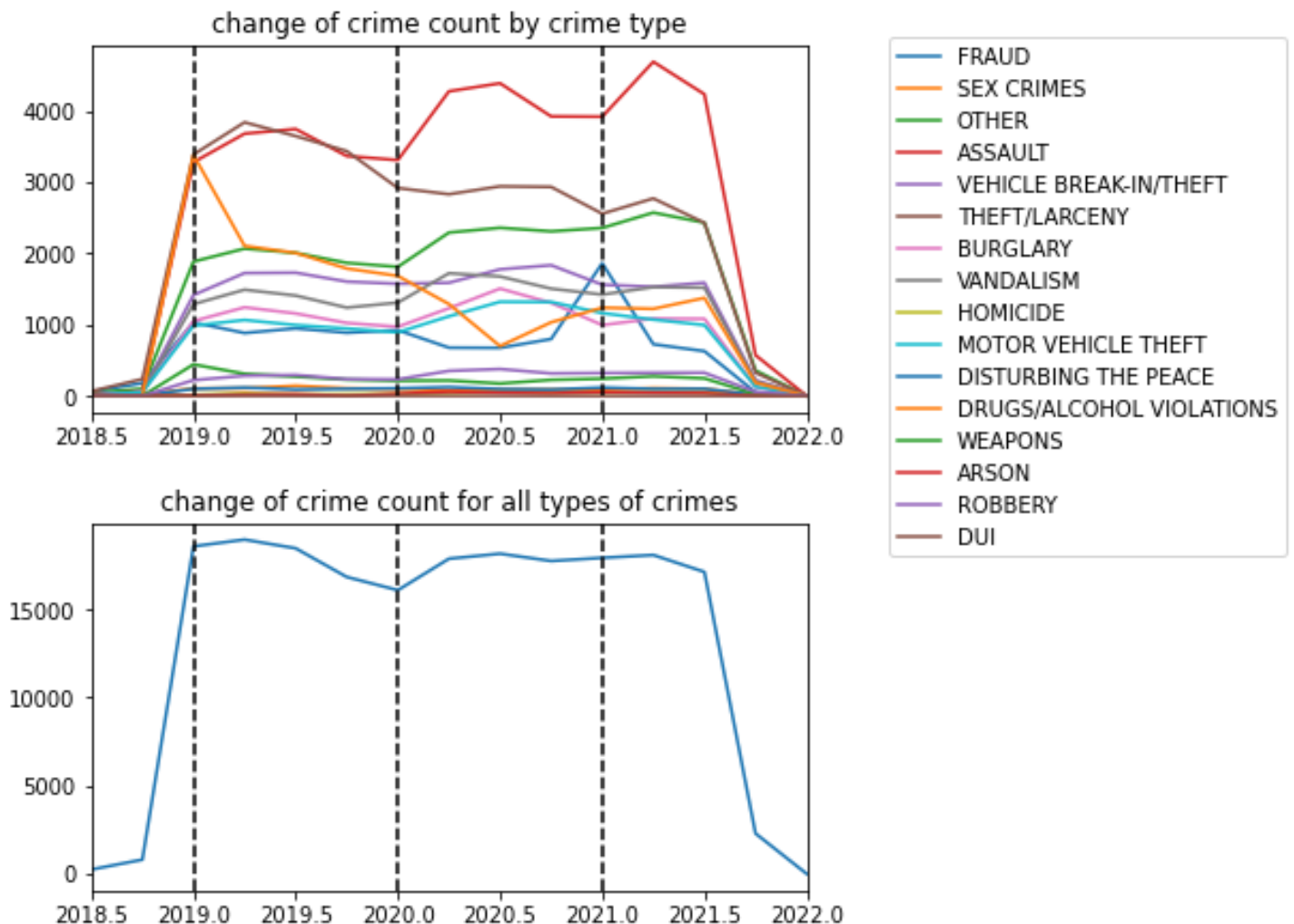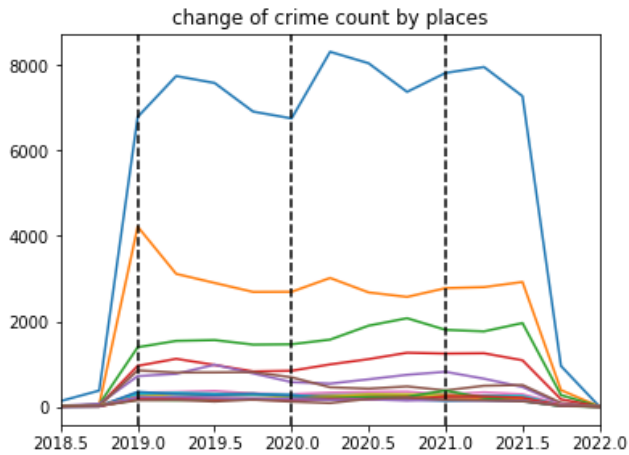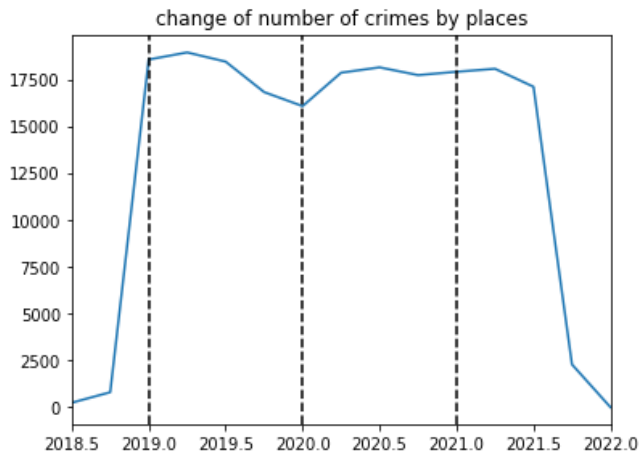
change of number of crimes by places

## change of crime count by places

- RESIDENCE / HOME
- HIGHWAY / ROAD / ALLEY
- PARKING LOT / GARAGE
- OTHER RESIDENCE (APARTMENT/CONDO)
- OTHER / UNKNOWN
- DEPARTMENT / DISCOUNT STORE
- SERVICE / GAS STATION
- CONVENIENCE STORE
- SPECIALTY STORE (TV, FUR, ETC)
- RESTAURANT
- GROCERY / SUPERMARKET
- COMMERCIAL / OFFICE BUILDING
- CYBERSPACE
- HOTEL / MOTEL / ETC.
- DRUG STORE/DR`S OFFICE/HOSPITAL
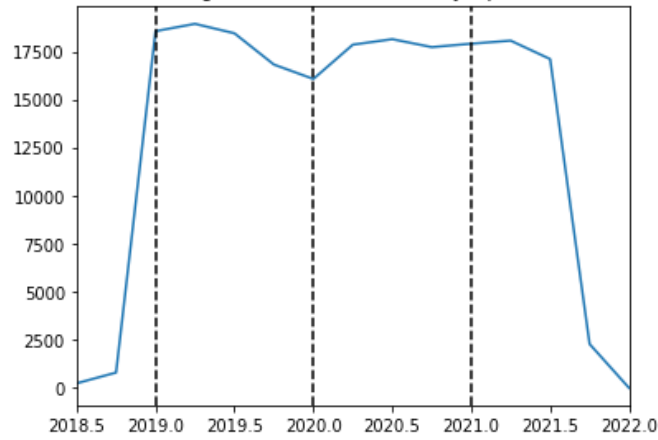- RENTAL / STORAGE FACILITY

change of crime count by places
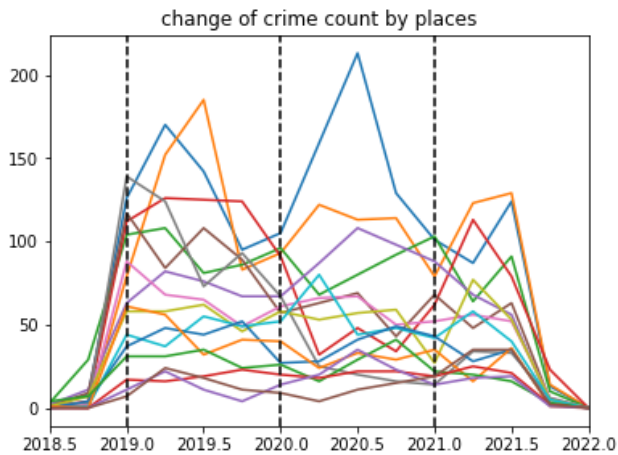
- NON-ATTACHED RESD GARAGE/SHED/BULD
- PARK / PLAYGROUND
- BANK / SAVINGS & LOAN
- BAR / NIGHT CLUB
- AUTO DEALERSHIP (NEW OR USED)
- GOVERNMENT / PUBLIC BUILDING
- CHURCH / SYNAGOGUE / TEMPLE
- SCHOOL - ELEMENTARY / SECONDARY
- CONSTRUCTION SITE
- LIQUOR STORE
- MALL / SHOPPING CENTER
- HOMELESS SHELTER / MISSION
- INDUSTRIAL SITE
- AIR / BUS / TRAIN TERMINAL
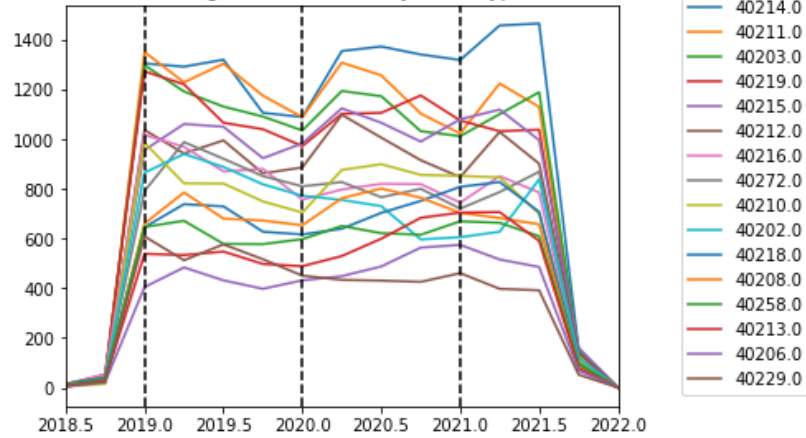- ATTACHED RESIDENTIAL GARAGE
- RACE TRACK/GAMBLING FACILITY

change of crime count by places
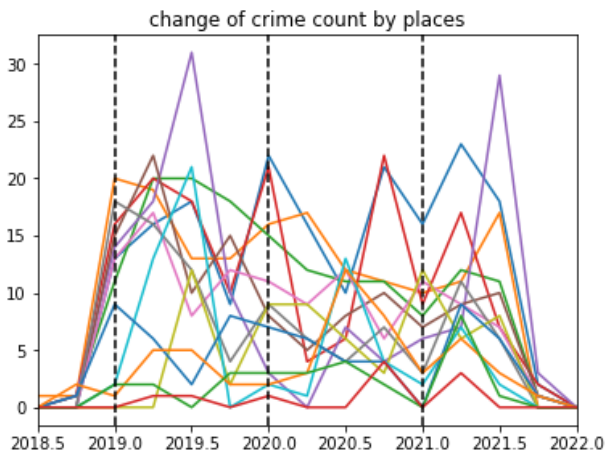
- FIELD / WOODS
- JAIL / PENITENTARY
- SCHOOL - COLLEGE / UNIVERSITY
- COMMUNITY CENTER
- FAIRGROUNDS / STADIUM / ARENA
- CHILD DAYCARE FACILITY
- ATM SEPARATE FROM BANK
- ABANDONED/CONDEMNED STRUCTURE
- CEMETERY / GRAVEYARD
- AMUSEMENT PARK
- DOCK/WHARF/FREIGHT/MODAL TERMINAL
- CAMP / CAMPGROUND
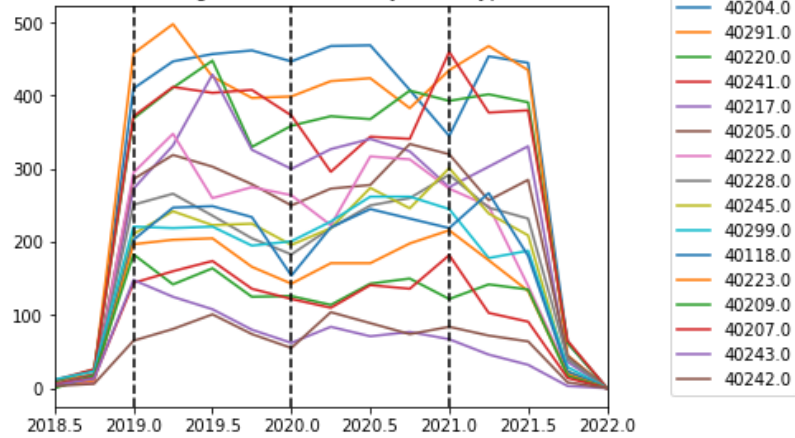- LAKE / WATERWAY
- FARM FACILITY
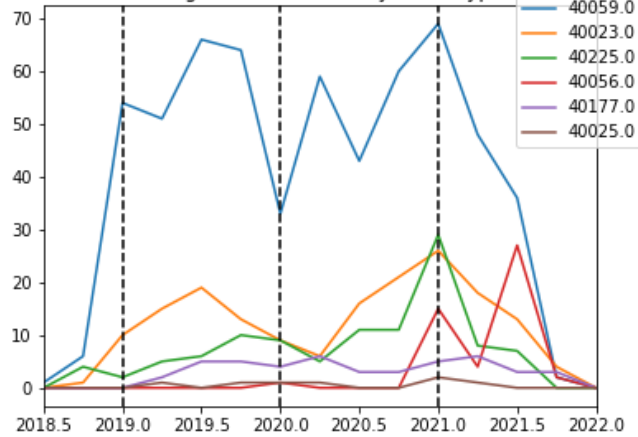
change of number of crimes by zip code
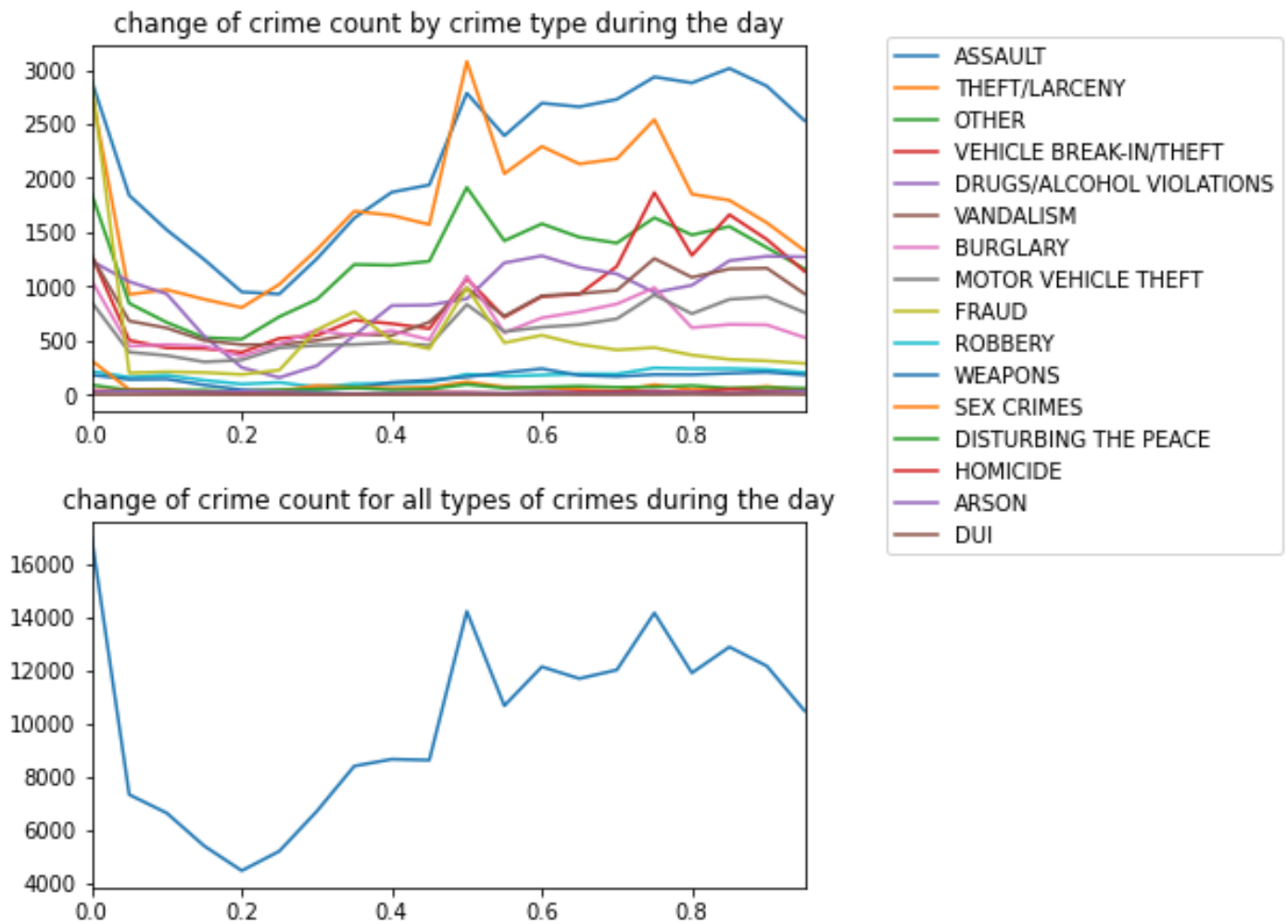
change of crime count by crime type

| 40214.0 |
| 40211.0 |
| 40203.0 |
| 40219.0 |
| 40215.0 |
| 40212.0 |
| 40216.0 |
| 40272.0 |
| 40210.0 |
| 40202.0 |
| 40218.0 |
| 40208.0 |
| 40258.0 |
| 40213.0 |
| 40206.0 |
| 40229.0 |

change of crime count by crime type

| 40204.0 |
| 40291.0 |
| 40220.0 |
| 40241.0 |
| 40217.0 |
| 40205.0 |
| 40222.0 |
| 40228.0 |
| 40245.0 |
| 40299.0 |
| 40118.0 |
| 40223.0 |
| 40209.0 |
| 40207.0 |
| 40243.0 |
| 40242.0 |

change of crime count by crime type

| 40059.0 |
| 40023.0 |
| 40225.0 |
| 40056.0 |
| 40177.0 |
| 40025.0 |

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

change of crime count by crime type during the day

Legend:
- ASSAULT
- THEFT/LARCENY
- OTHER
- VEHICLE BREAK-IN/THEFT
- DRUGS/ALCOHOL VIOLATIONS
- VANDALISM
- BURGLARY
- MOTOR VEHICLE THEFT
- FRAUD
- ROBBERY
- WEAPONS
- SEX CRIMES
- DISTURBING THE PEACE
- HOMICIDE
- ARSON
- DUI

change of crime count for all types of crimes during the day

change of number of crimes by place during the day
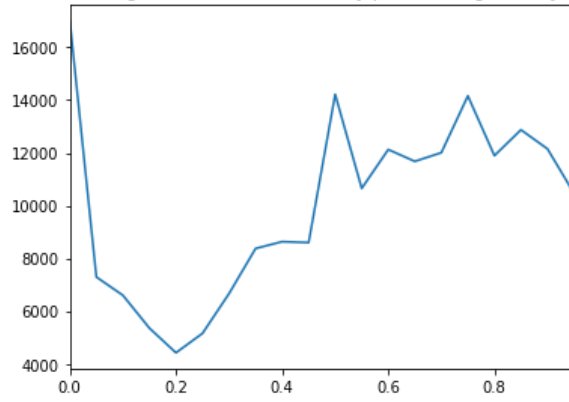
change of crime count by place during the day
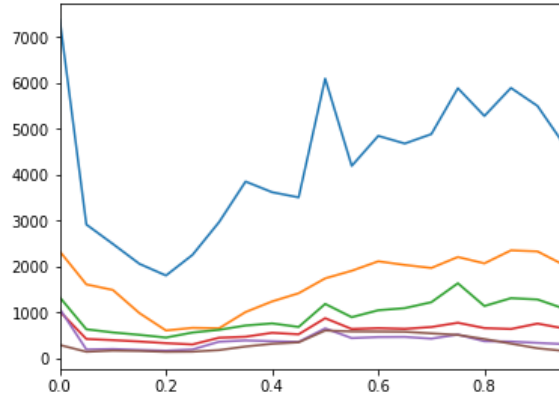
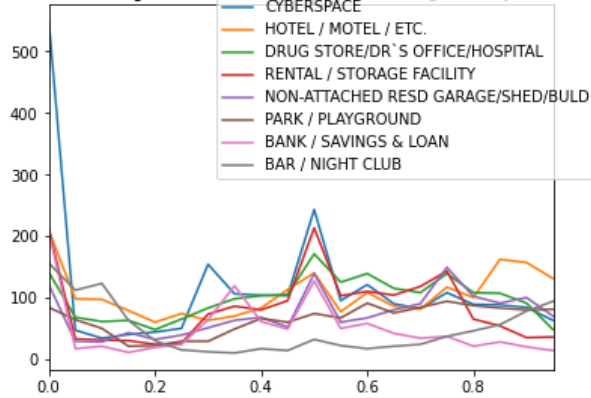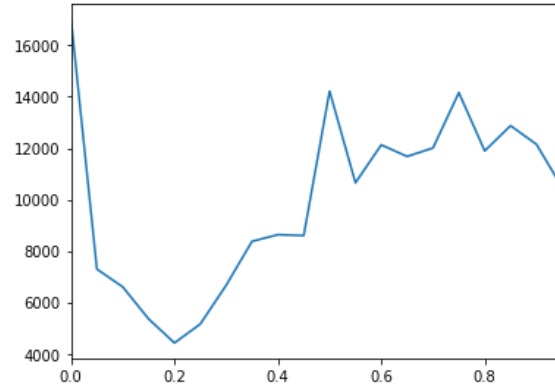|	
—	RESIDENCE / HOME
—	HIGHWAY / ROAD / ALLEY
—	PARKING LOT / GARAGE
—	OTHER RESIDENCE (APARTMENT/CONDO)
—	OTHER / UNKNOWN
—	DEPARTMENT / DISCOUNT STORE

change of crime count by place during the day
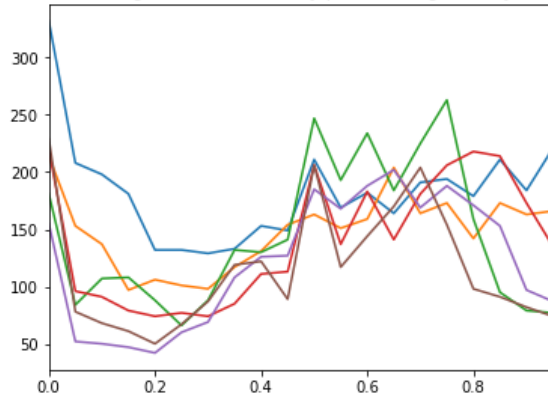
—	SERVICE / GAS STATION
—	CONVENIENCE STORE
—	SPECIALTY STORE (TV, FUR, ETC)
—	RESTAURANT
—	GROCERY / SUPERMARKET
—	COMMERCIAL / OFFICE BUILDING

change of crime count by place during the day

—	CYBERSPACE
—	HOTEL / MOTEL / ETC.
—	DRUG STORE/DR`S OFFICE/HOSPITAL
—	RENTAL / STORAGE FACILITY
—	NON-ATTACHED RESD GARAGE/SHED/BULD
—	PARK / PLAYGROUND
—	BANK / SAVINGS & LOAN
—	BAR / NIGHT CLUB

change of number of crimes by zip code during the day

change of crime count by zipe code during the day

| 40214.0 |
| 40211.0 |
| 40203.0 |
| 40219.0 |
| 40215.0 |
| 40212.0 |

change of crime count by zipe code during the day

| 40216.0 |
| 40272.0 |
| 40210.0 |
| 40202.0 |
| 40218.0 |
| 40208.0 |

change of crime count by zipe code during the day

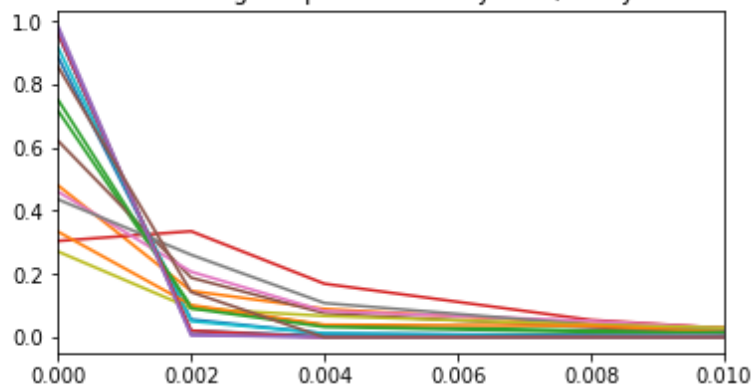| 40258.0 |
| 40213.0 |
| 40206.0 |
| 40229.0 |
| 40204.0 |
| 40291.0 |
| 40220.0 |
| 40241.0 |
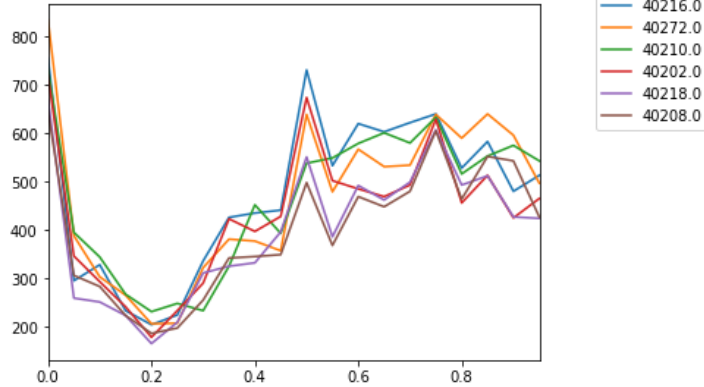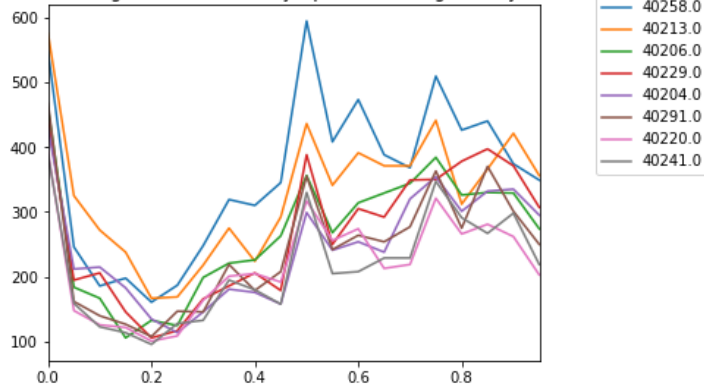
- how fast crimes get reported after occurring:

number of all crimes that get reported after x years (0.05 year is 18 days)
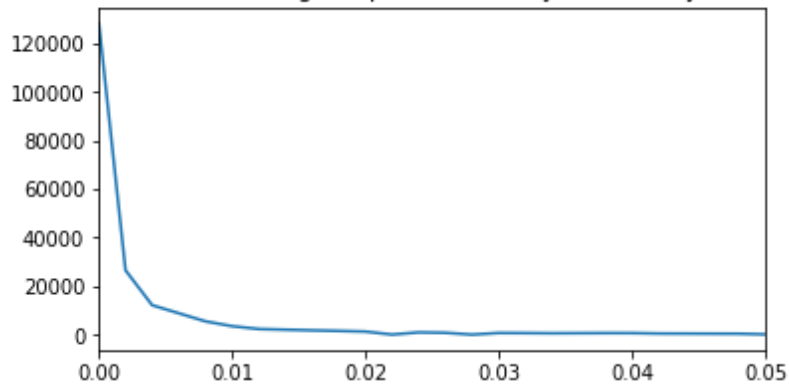
% of each crimes that get reported after x years (0.01 year is 3.5 days)

Legend:
- ASSAULT
- THEFT/LARCENY
- OTHER
- VEHICLE BREAK-IN/THEFT
- DRUGS/ALCOHOL VIOLATIONS
- VANDALISM
- BURGLARY
- MOTOR VEHICLE THEFT
- FRAUD
- ROBBERY
- WEAPONS
- SEX CRIMES
- DISTURBING THE PEACE
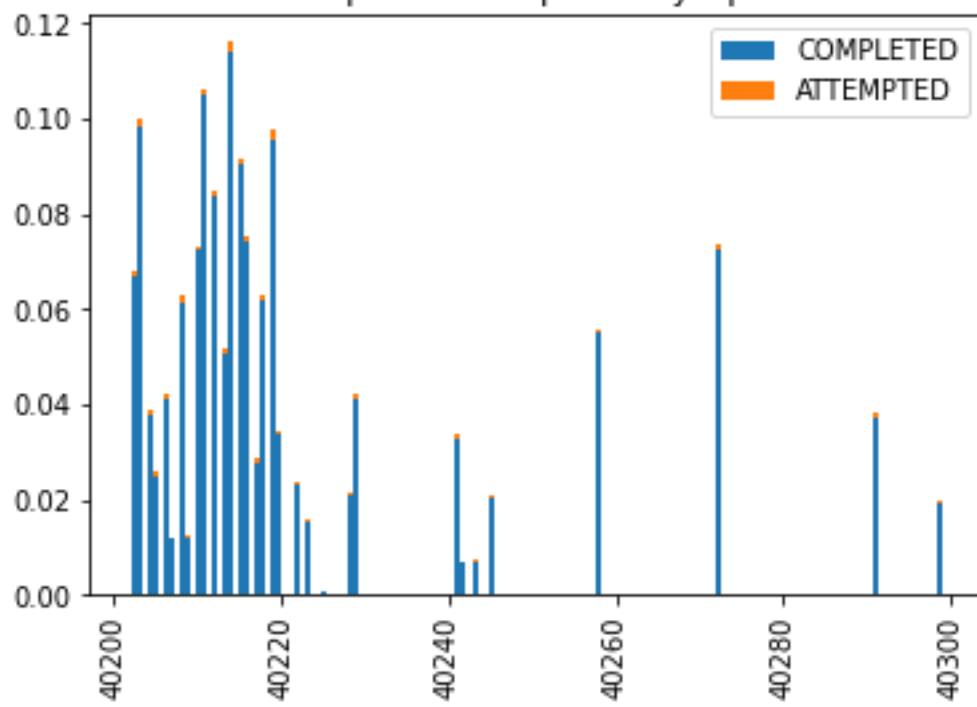- HOMICIDE
- ARSON
- DUI

- attempted or completed cases:

attempted or completed by crime type
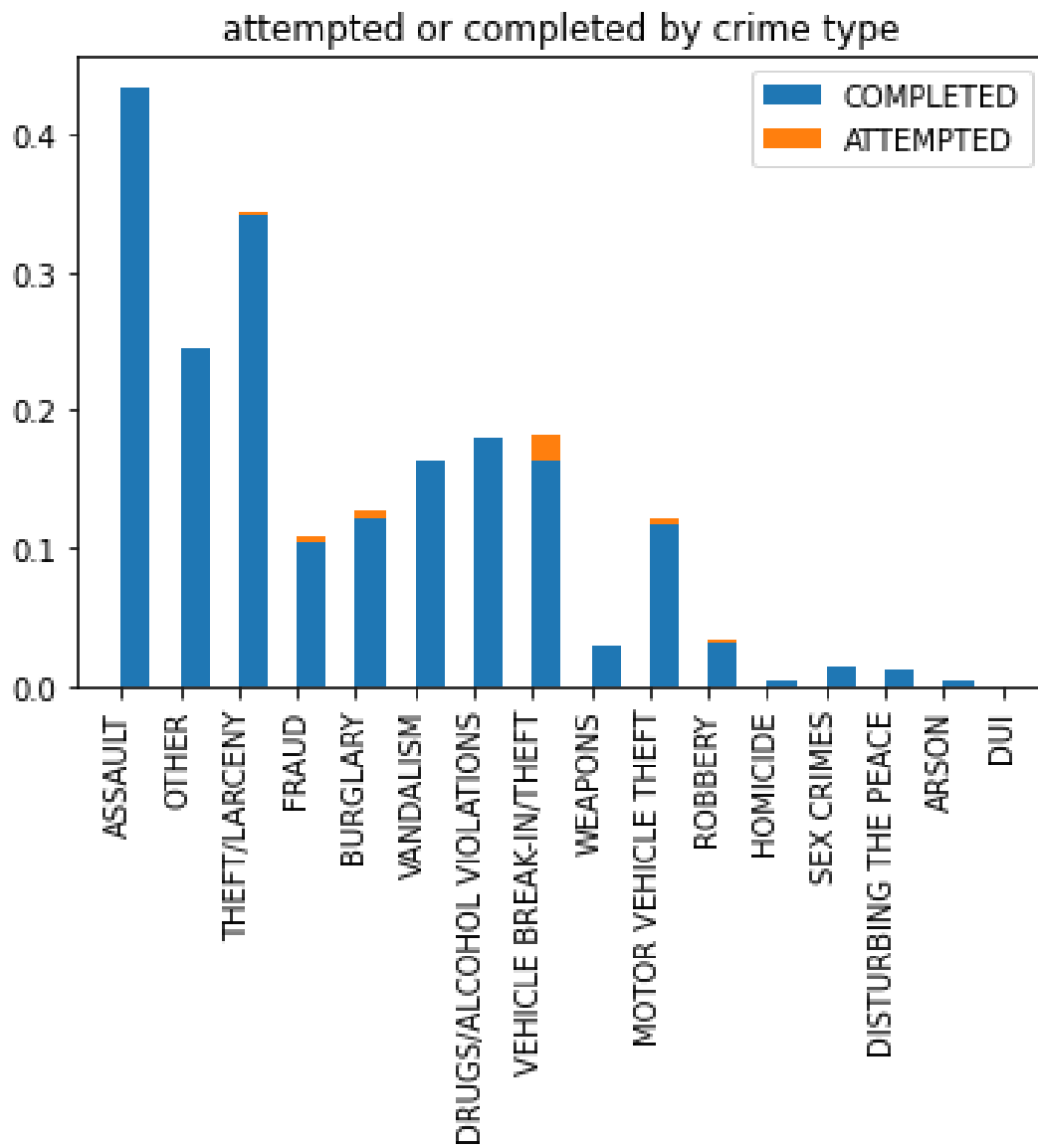
attempted or completed by place
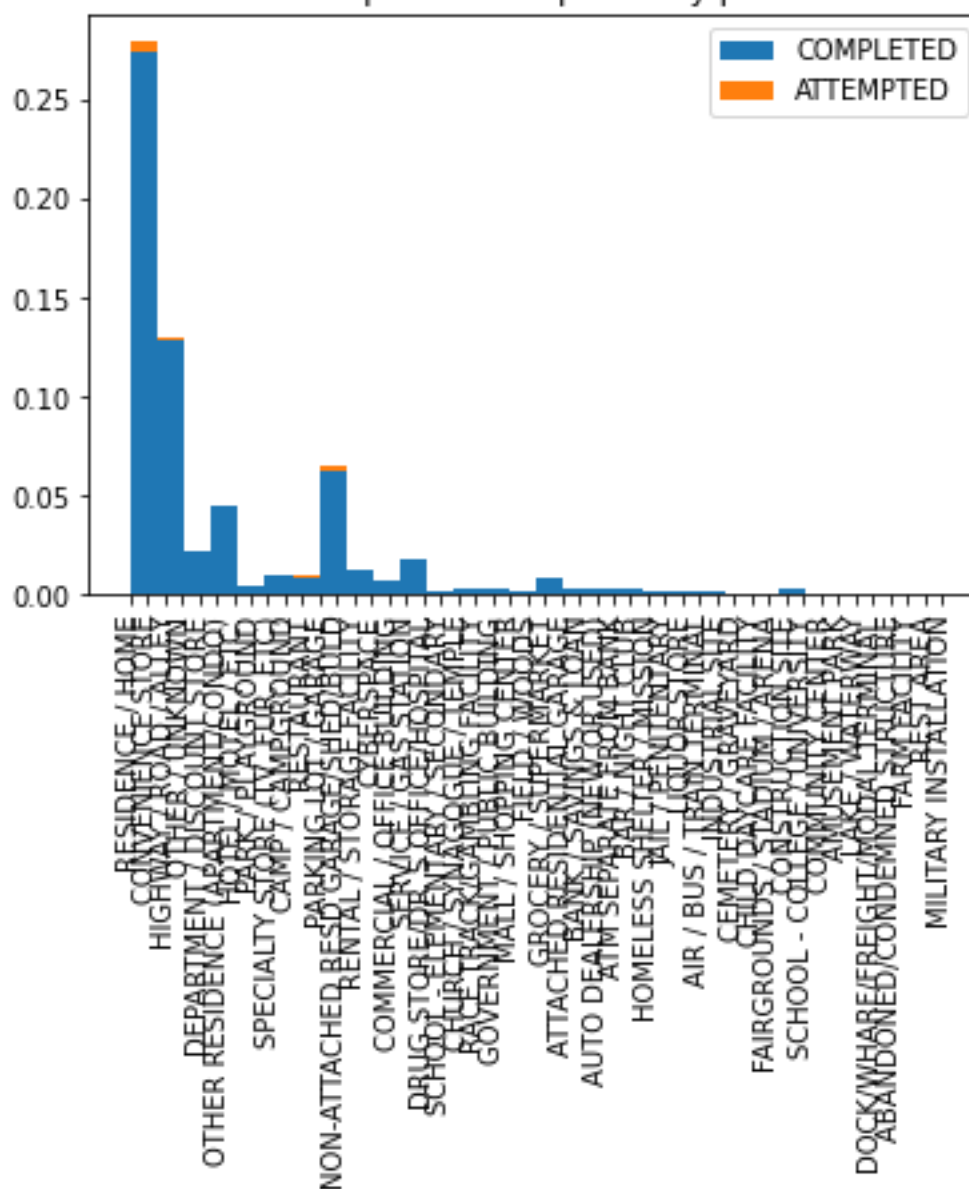
attempted or completed by zipcode

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

crime type by zip code

- each row indicates the % of crimes in that place (the line on the left is the total number of crimes):

crime type by place

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

zip code by place
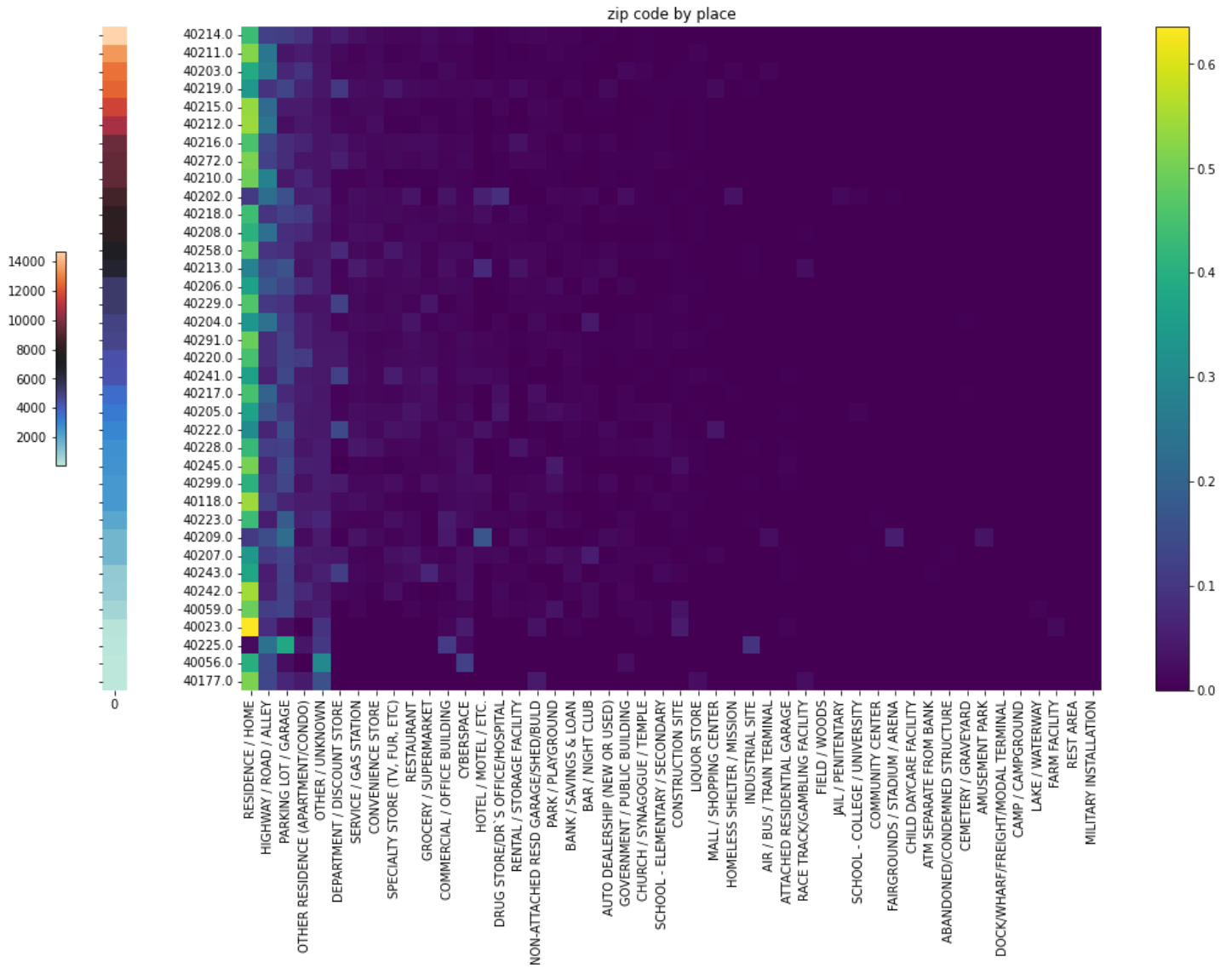
D'Agostino and Pearson test results (k2, p)

	DATE_REPORTED	DATE_OCCURED	ZIP_CODE	Time_reported	Time_occurred
0	1.631308e+02	157.454611	9.430088e+01	57.011350	54.878254
1	2.585490e-09	0.012716	1.840919e-08	0.000003	0.000005