Alyssa Cox
Machine Learning: HW1
Link to my GitHub repo: https://github.com/coxas/CAPP30254
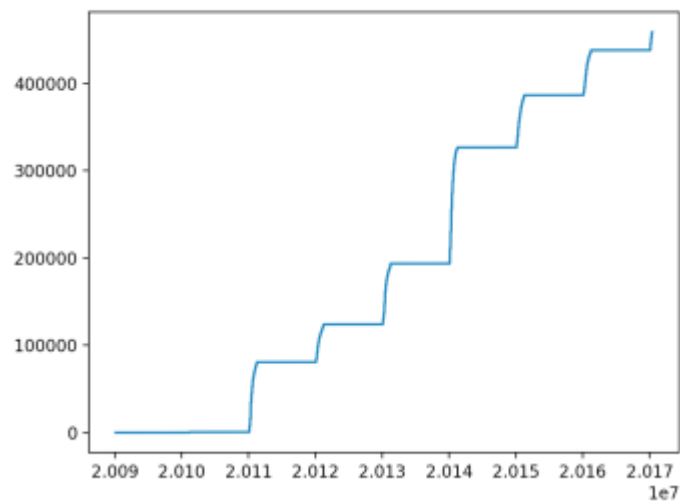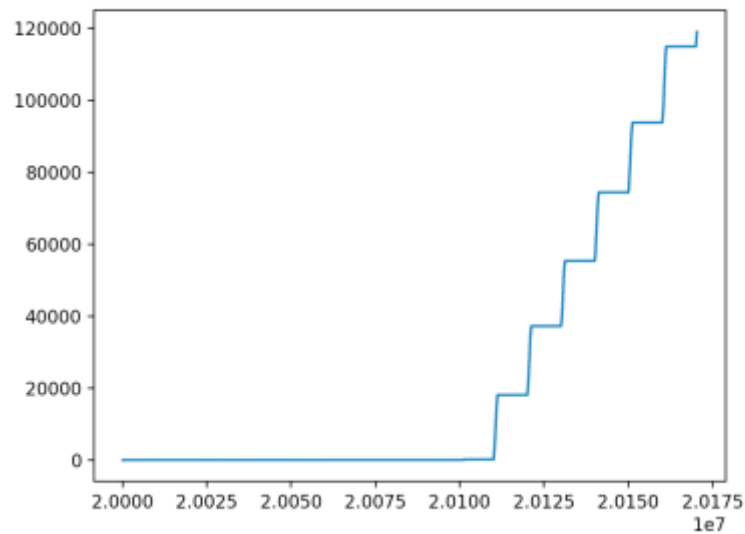
Problem 1:

Requests Over Time

Graffiti Requests Over Time (for some reason there are "Creation Dates" marked as in the 1920s, don't know what that's about):
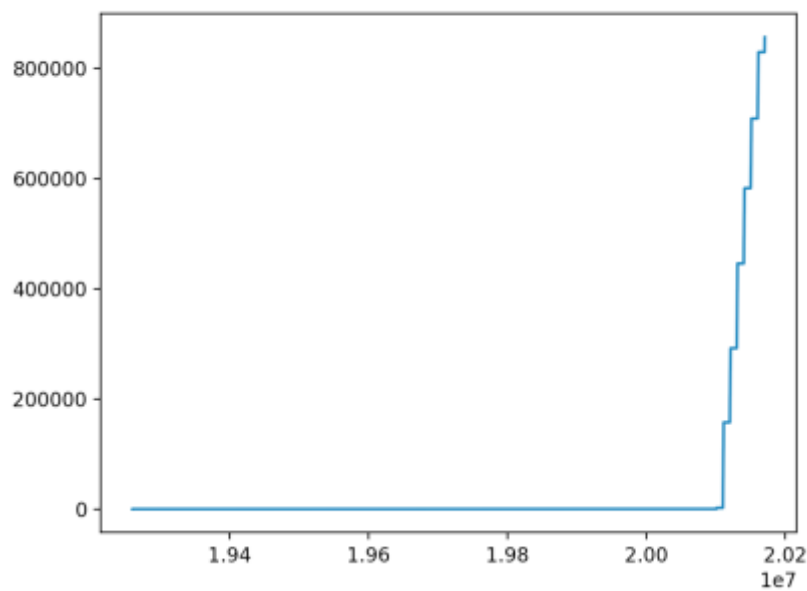
Pothole Requests Over Time:

Sanitation Requests Over Time:

When I tried to find the requests over time for the Abandoned Buildings Dataset, it said that some of the columns in the dataset had mixed types and wouldn't print the graph.

There are a lot of subgroups for each group, and putting each requests-over-time graph here would be excessive. I've included a graph of one subgroup for each of the three large groups that have subgroups. They all seem to mirror the plots for their main groups, respectively.

Graffiti on Metal Requests Over Time:
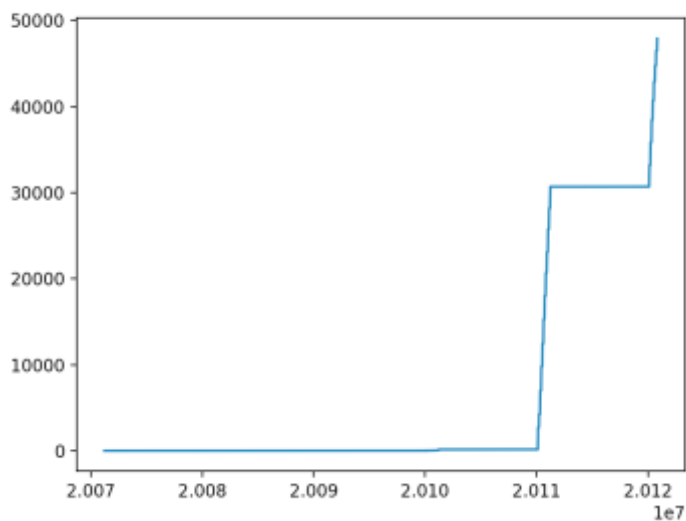
Sanitation Requests About "Garbage in Yard" over time:
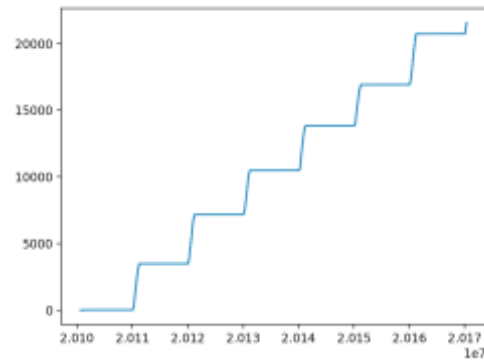
Again, there were no specific subtypes for potholes.

Average Response time:

Graffiti	8.05 days
Sanitation	9.55 days
Potholes	29.81 days

There was only a creation date, but not a completion date, for Abandoned Buildings, so I wasn't able to calculate response time.

Amazing to see how long it takes the city to answer pothole requests. Must be why there are still so many potholes in Hyde Park!

Requests by Subtype:

Graffiti:

Metal - Unpainted	15744
Asphalt	691
Limestone	3150
Stucco	2837
Other/Unknown Surface	6565
Aluminum Siding	42900
Tree	726
Vinyl Siding	14401
Glass	12146
Metal - Painted	142177
Wood - Unpainted	10799
Wood - Painted	48044
Marble/Granite	1056
Other / Unknown Surface	87059
Metal	47903
Brick - Painted	166052
Cement (Sidewalk, Alley, Wall, Curb)	50475
Brick - Unpainted	198701

Potholes: There weren't any specific subtypes for potholes.

Sanitation:

Graffiti Commercial Vehicle	88
Overflowing carts	8944
Dumpster not being emptied	9223
Construction Site Cleanliness/Fence	4194
Garbage in yard	21564
Dog feces in yard	6361
Standing water	1158

Abandoned Buildings:

Building is Boarded Up	33
Boarded	6347
Building is Open / Unsecure	258
Open	44739

Areas with Most Requests:

Type of Request	Neighborhood Name	Ward	Zip Code
Graffiti	South Lawndale	14	60632
Potholes	Austin	41	60629
Sanitation	Austin	6	60620
Abandoned Buildings	South Englewood	17	60636

Requests by Neighborhood (Ward):

Ward, Potholes		Ward, Graffiti		Ward, Sanitation		Ward, Buildings	
0.0	1311	0.0	271	0.0	3	0.0	10
1.0	8626	1.0	49078	1.0	2638	1.0	386
2.0	11718	2.0	14492	2.0	1962	2.0	689
3.0	6500	3.0	4016	3.0	1883	3.0	1409
4.0	5691	4.0	2339	4.0	1048	4.0	417
5.0	8426	5.0	2073	5.0	1606	5.0	880
6.0	9619	6.0	1099	6.0	4843	6.0	3124
7.0	6697	7.0	2187	7.0	3937	7.0	2927
8.0	12495	8.0	1625	8.0	3015	8.0	2349
9.0	8420	9.0	1910	9.0	3139	9.0	2739
10.0	9872	10.0	18804	10.0	1642	10.0	1615
11.0	8360	11.0	34359	11.0	2563	11.0	575
12.0	7891	12.0	47890	12.0	1353	12.0	517
13.0	12746	13.0	18465	13.0	1629	13.0	454
14.0	9824	14.0	64920	14.0	2683	14.0	903
15.0	6278	15.0	14014	15.0	2716	15.0	3486
16.0	6379	16.0	15957	16.0	2533	16.0	4046
17.0	7441	17.0	2907	17.0	3972	17.0	4863
18.0	11036	18.0	8952	18.0	3653	18.0	1115
19.0	13534	19.0	1436	19.0	1357	19.0	491
20.0	6239	20.0	5693	20.0	2682	20.0	2590
21.0	11883	21.0	1054	21.0	4688	21.0	2397
22.0	5503	22.0	35734	22.0	1457	22.0	538
23.0	14129	23.0	30808	23.0	1885	23.0	379
24.0	7663	24.0	6880	24.0	3021	24.0	2763
25.0	8296	25.0	48708	25.0	1888	25.0	455
26.0	6868	26.0	21462	26.0	2797	26.0	664
27.0	10433	27.0	20223	27.0	3147	27.0	1696
28.0	8692	28.0	11972	28.0	3648	28.0	3226
29.0	9679	29.0	8253	29.0	2018	29.0	931
30.0	6928	30.0	27715	30.0	1983	30.0	527
31.0	7503	31.0	29628	31.0	2099	31.0	437
32.0	14169	32.0	28526	32.0	4448	32.0	374
33.0	6478	33.0	36700	33.0	2206	33.0	264
34.0	11303	34.0	1171	34.0	3237	34.0	4053
35.0	8653	35.0	38698	35.0	2930	35.0	489
36.0	11654	36.0	9546	36.0	1810	36.0	454
37.0	7581	37.0	7304	37.0	2722	37.0	1615
38.0	10239	38.0	14638	38.0	1786	38.0	252
39.0	11853	39.0	15068	39.0	2041	39.0	189
40.0	9700	40.0	16532	40.0	1782	40.0	172
41.0	16623	41.0	1856	41.0	1474	41.0	161
42.0	11738	42.0	17213	42.0	1079	42.0	51
43.0	6859	43.0	12148	43.0	3148	43.0	117
44.0	5392	44.0	13875	44.0	1554	44.0	87
45.0	13212	45.0	12760	45.0	1981	45.0	239
46.0	3482	46.0	8779	46.0	1519	46.0	89
47.0	7655	47.0	32000	47.0	1610	47.0	192
48.0	5747	48.0	9456	48.0	837	48.0	84
49.0	8452	49.0	11171	49.0	1256	49.0	80
50.0	10855	50.0	13299	50.0	2041	50.0	155

Five interesting things I learned about 311 requests:
- There are a lot of different ways that people identify with their community. There are the 77 recognized "community areas", which is what I usually understand as the "neighborhood". For example, I live in Hyde Park, but my favorite community area is Andersonville. Actually, that's not even a Community Area, it's a neighborhood withing the Community Area of Edgewater! However, citizens also identify with their wards, because each ward is tied together politically, via the Aldermen who help run the city. There are also zip codes, which generally identify an area for mail delivery.
- The community area with the most graffiti calls was South Lawndale. For the most abandoned/vacant buildings reported it was West Englewood. For both potholes and sanitation, it was Austin.
- These neighborhoods are all on the South or West side are some of the poorest and most dangerous neighborhoods in Chicago.
- On average, there are longer completion times for community areas on the South and West sides of the City.
- There wasn't much use of the 311 line until around the year 2000, when calls dramatically spiked for all four types of calls.

Problem 2:

You will notice I don't really have much for this section. I ran out of time. I had to present my BA thesis last week and spent all of my time working on that instead of this homework. I also don't know where Professor Rayid got the idea that we worked with APIs in 121-122. I haven't worked with an API before and I simply did not have time to figure it out. I'm not a bad or slacker student; my transcript will attest to that. I just simply did not have adequate time to dedicate to learning something from scratch.

Ideally, though, in theory, I would use the Census data to create a new dataframe using pandas and join it on the ones that I already created in part 1, by using latitude/longitude as a join condition or by using the latitude and longitude to determine census tract, and then categorize each address as belonging to a certain census tract. I would then use the functions I created in problem 1 to pull out characteristics related to demographics for, say, the ten zip codes with the most requests.

Problem 3:

A: There were six total calls from 7500 S Wolcott Ave. Five were about potholes and one was about sanitation. The probability of an incoming call being about potholes is .8333, and the probability of it being about sanitation is 0.1667. Therefore, it is most likely that an incoming call from 7500 S Wolcott Ave will concern potholes.

B: Lawndale is in Community Area 29. 4584 requests about Graffiti Removal came from Lawndale. There were 11914 requests about Graffiti Removal from Uptown, which is Community Area 3. There were

C: There were 260 total calls about Graffiti Removal. 100 came from Englewood, giving a probability of (100 / 260) = 0.3846. 160 came from Uptown, giving a probability of 0.6154. Therefore, it is about 23% more likely that the call about Graffiti Removal came from Uptown.