DATATHON

Saransh Arora
Samarth Modi
Pradeep Prabhakar

Seattle Crime Dataset

Dataset: https://data.seattle.gov/Public-Safety/Call-Data/33kz-ixgy | View the Dataset Schema Here

Task Category: Data Visualization, Data Analysis, and Machine Learning Modelling

- This data represents crime reported to the Seattle Police Department (SPD). Each row contains the record of a unique event where at least one criminal offense was reported by a member of the community or detected by an officer in the field.
- These data contain offenses and offense categorization coded to simulate the standard reported to the FBI under the National Incident-Based Reporting System (NIBRS) and used to generate a Uniform Crime Report (UCR) summary statistics.

Questions to be answered

- What formula would you use for a metric to rank the safety of each neighborhood for a student, a family with a young child below the age of 15, and an elderly person? What would be the ranking result when this formula is used for each category of people?
- To what extent and accuracy can we predict the level of change in crime for a neighborhood?
- What unique statistics and patterns were you able to identify in Seattle's crime?

About the data

- The Seattle crimes dataset that we used has 523k tuples and 11 features. For the sake of simplicity, we worked only on crimes reported in year 2017 and 2018. This gave us roughly 100k rows.
- The dataset provided the following features:
  - Incident occurrence date and time
  - Incident reporting date and time
  - Crime description
  - Neighborhood of crime occurrence

Table Preview

View Data | Create Visualization

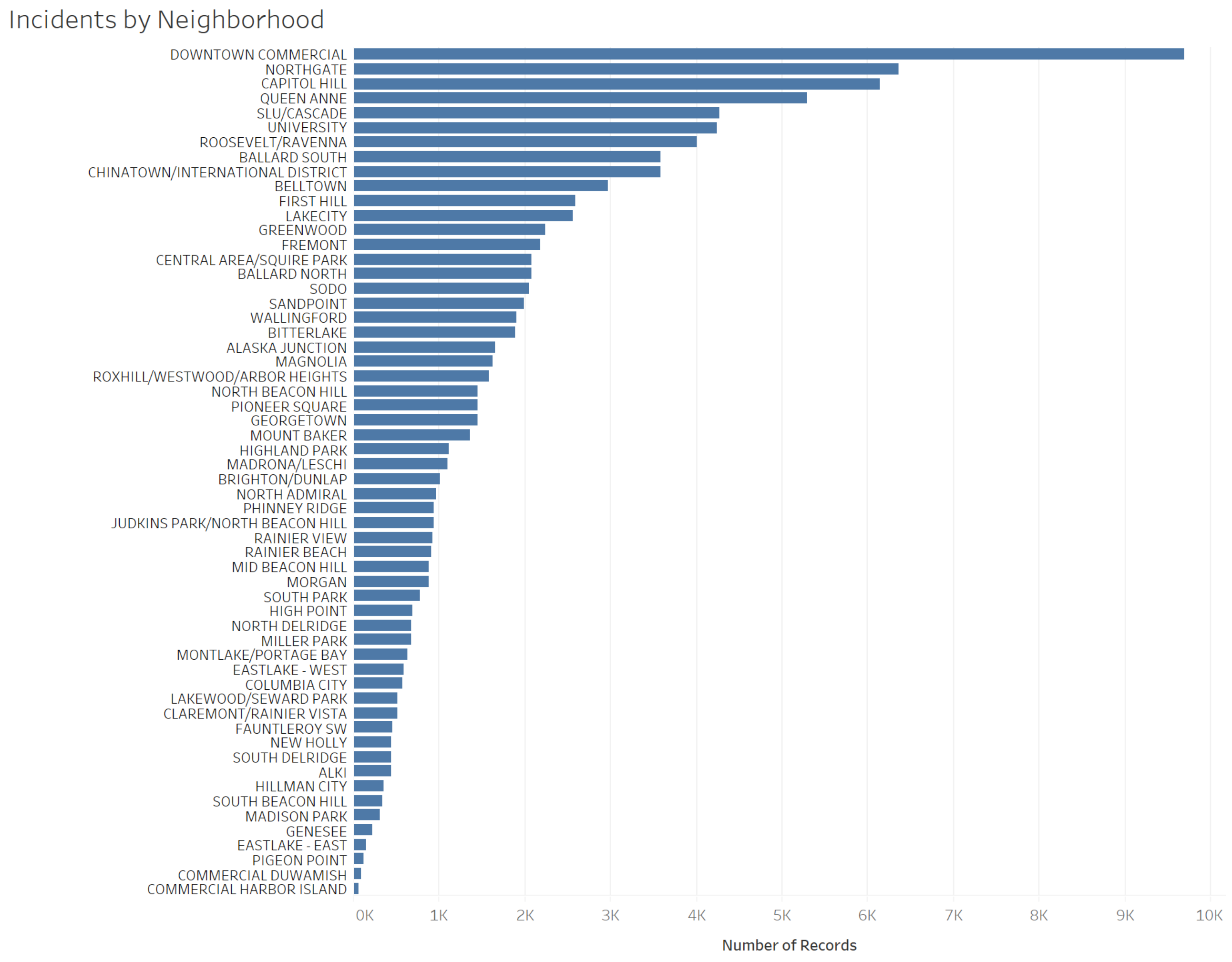
Repor ↑	Occurr...	Occurr...	Report...	Report...	Crime ...	Primar...	Precinct	Sector	Beat	Neighb...
197500007...	12/16/1975	900	12/16/1975	1500	BURGLARY-...	BURGLARY-...	SOUTH	R	R3	LAKEWOO...
197600006...	01/01/1976	1	01/31/1976	2359	SEX OFFEN...	SEXOFF-IN...	UNKNOWN			UNKNOWN
197900004...	01/28/1979	1600	02/09/1979	1430	CAR PROWL	THEFT-CAR...	EAST	G	G2	CENTRAL A...
198100003...	08/22/1981	2029	08/22/1981	2030	HOMICIDE	HOMICIDE-...	SOUTH	S	S2	BRIGHTON/...
198100007...	02/14/1981	2000	02/15/1981	435	BURGLARY-...	BURGLARY-...	SOUTHWEST	W	W3	ROXHILL/W...
198800005...	09/29/1988	155	09/29/1988	155	MOTOR VE...	VEH-THEFT-...	WEST	M	M2	SLU/CASCA...
199300004...	10/08/1993	2213	10/08/1993	2213	HOMICIDE	HOMICIDE-...	SOUTH	R	R2	CLAREMON...
199400002...	06/08/1994	0	06/12/1994	844	THEFT-ALL ...	THEFT-OTH	SOUTHWEST	F	F1	HIGH POINT
199600005...	12/08/1996	1130	12/08/1996	1700	CAR PROWL	THEFT-CAR...	SOUTH	O	O1	SODO
199900003...			01/01/1999		THEFT-SHO...	THEFT-SHO...	UNKNOWN			UNKNOWN
200000002...	05/12/2000	2330	05/14/2000	1055	CAR PROWL	THEFT-CAR...	WEST	Q	Q3	SLU/CASCA...
200100002...	01/15/2001	2310	01/15/2001	2310	HOMICIDE	HOMICIDE-...	WEST	K	K3	DOWNTOW...
200200001...	03/27/2002	1307	03/27/2002	1307	DUI	DUI-LIQUOR	EAST	G	G2	CENTRAL A...
200200005...	12/08/2002	1520	12/08/2002	1608	CAR PROWL	THEFT-CAR...	EAST	G	G1	FIRST HILL

< Previous   Next >   Showing A unique crime report record. s 1-14 out of 523,591

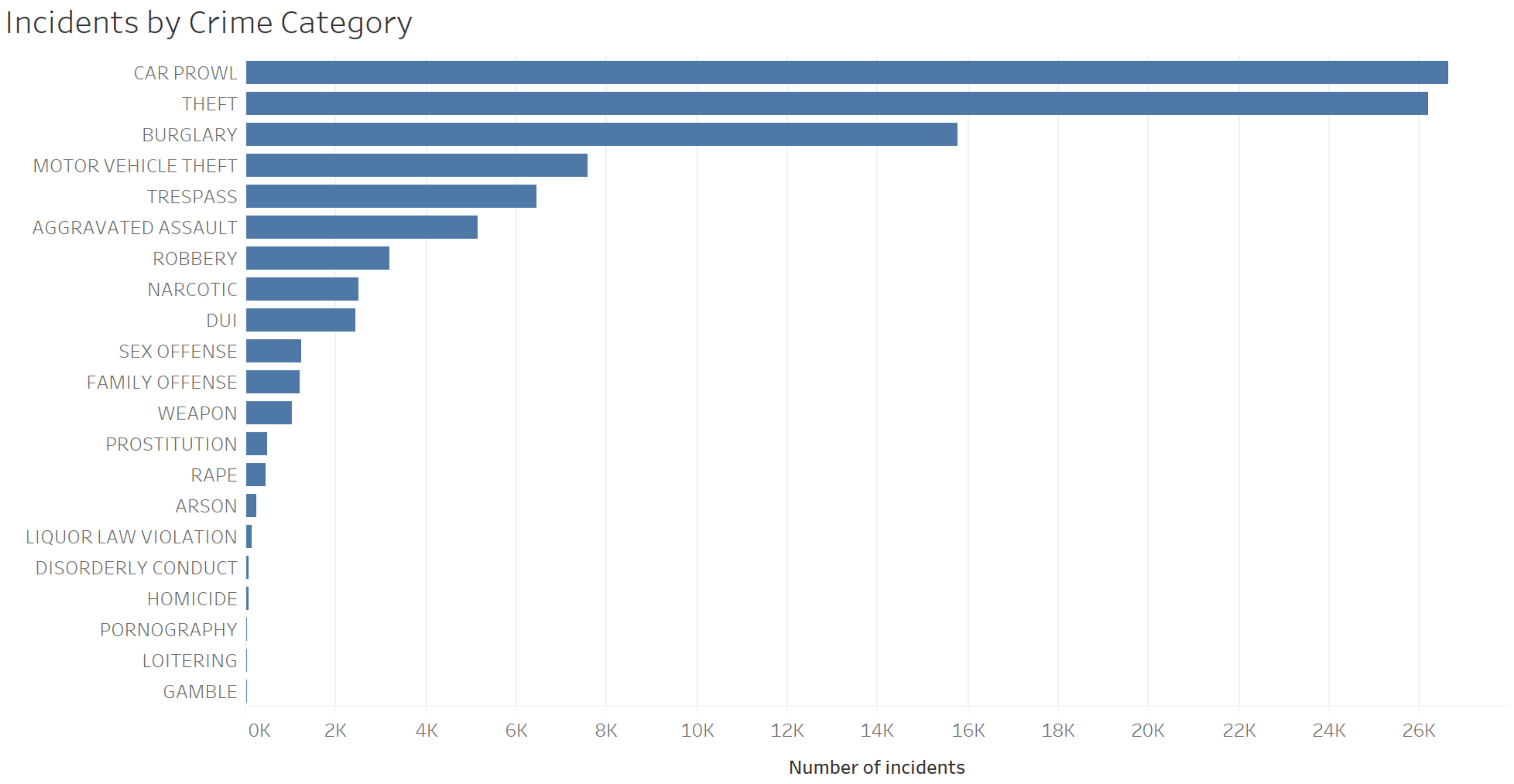
Data preparation

- The data was not clean, and we had to do the following:
  - Dropped the records where the neighborhood is unknown
  - Formatted the time occurred/reported variables
  - A lot of events occurred at around 00:00 and our guess is that since most of the events are not reported immediately, by default they are captured as 00:00 in the dataset

Number of Incidents per Neighborhood

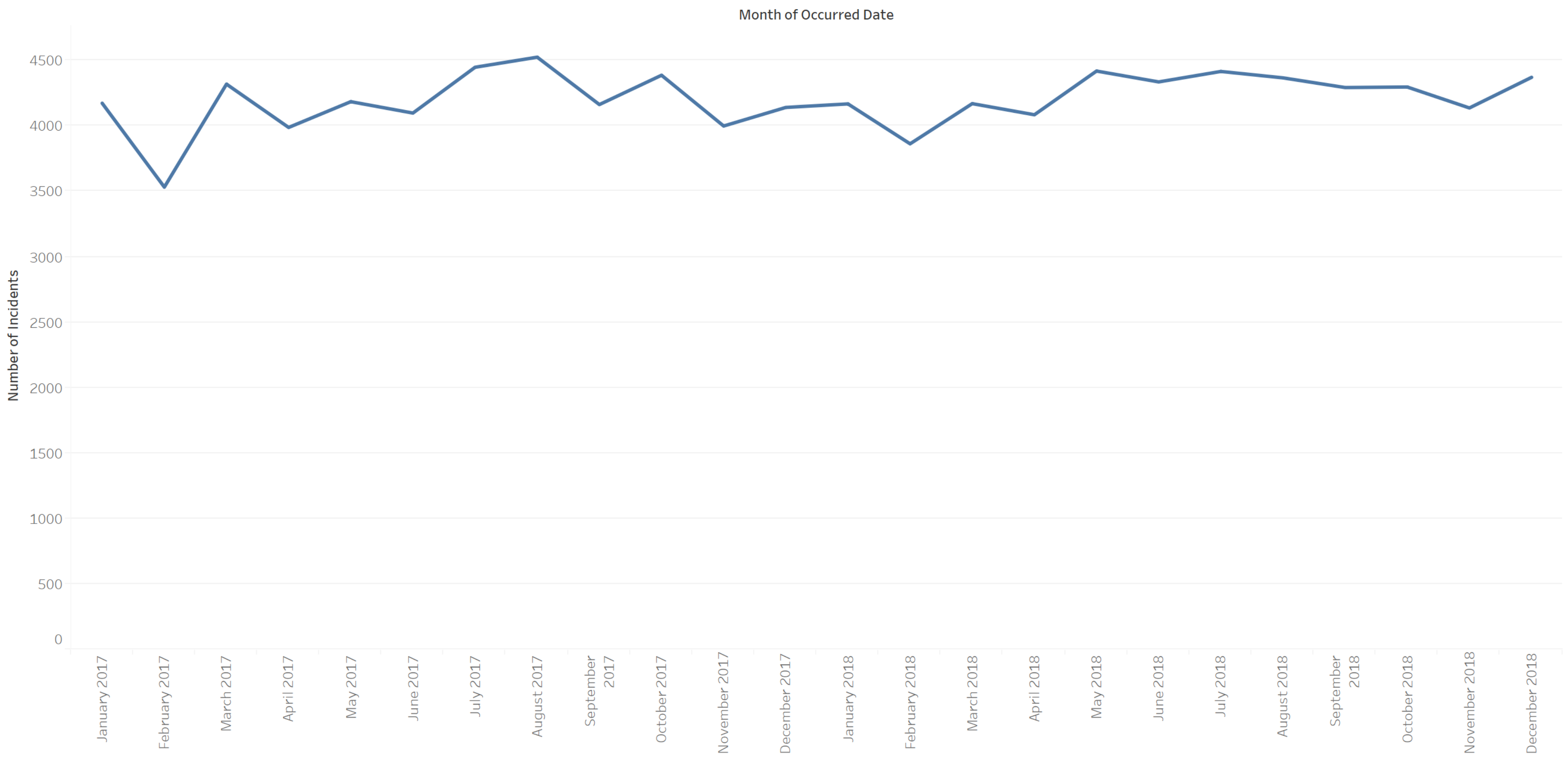


Number of Incidents by Crime Category



Number of Incidents vs Month of Incident

Sheet 1



Number of Incidents occurred by hour vs month

Incidents occured by hour vs month

	Occurred Date											
Hour of Occ..	January	Februa..	March	April	May	June	July	August	Septe..	October	Novem..	Decem..
0	473	482	473	480	570	531	576	594	539	564	523	595
1	291	230	265	235	282	279	283	276	255	300	232	273
2	240	189	221	203	202	211	254	244	232	243	204	229
3	170	139	186	193	205	174	216	213	198	180	162	182
4	156	137	153	151	156	163	177	198	170	146	156	163
5	128	122	140	123	155	113	149	146	157	157	140	144
6	171	134	169	173	139	151	185	189	163	162	155	145
7	194	212	217	215	208	188	211	200	197	210	198	208
8	266	262	298	274	277	261	321	288	272	291	274	271
9	266	281	263	254	260	272	283	315	277	309	261	292
10	316	269	301	298	330	338	305	306	327	294	284	301
11	350	297	343	297	313	322	336	336	340	310	348	313
12	473	435	490	443	468	465	525	493	464	515	449	517
13	361	336	362	366	379	367	358	367	404	376	390	380
14	370	353	419	366	399	357	425	408	367	409	390	401
15	433	369	431	408	451	420	399	426	408	452	432	421
16	444	409	459	430	437	450	433	506	413	472	436	480
17	522	413	496	515	518	501	499	504	508	510	471	550
18	556	453	553	495	504	542	477	514	510	493	510	507
19	485	410	505	470	444	453	460	497	464	481	492	463
20	488	409	495	492	533	488	520	523	482	513	443	489
21	442	384	436	421	452	439	471	470	485	448	407	439
22	387	385	429	407	472	505	553	450	471	483	431	405
23	359	285	385	364	450	444	448	429	353	366	348	344

Incidents reported by hour vs month

	Reported Date											
	January	Februa..	March	April	May	June	July	August	Septe..	October	Novem..	Decem..
0	185	167	169	176	237	201	218	198	199	196	171	192
1	154	133	167	151	199	161	167	193	164	171	136	160
2	149	129	145	123	152	138	176	149	138	173	124	164
3	93	84	110	113	135	122	131	131	101	119	105	91
4	130	105	108	113	117	118	127	126	127	98	111	104
5	122	100	111	104	124	113	117	137	115	135	100	126
6	158	143	141	167	191	167	197	187	158	155	144	163
7	282	268	292	286	351	287	307	314	292	277	249	309
8	397	346	452	404	405	380	445	431	410	426	413	403
9	500	431	477	423	484	477	547	530	525	538	509	465
10	531	441	544	482	508	516	536	503	498	546	508	517
11	479	408	467	444	432	437	494	532	521	492	507	505
12	528	448	566	489	498	490	563	568	500	557	493	546
13	537	504	509	510	520	542	498	544	526	602	565	596
14	506	466	513	436	502	481	508	531	463	506	494	528
15	555	466	498	517	526	526	489	539	518	545	488	484
16	531	450	502	512	486	484	524	509	485	545	491	465
17	478	394	474	464	473	447	470	502	495	500	467	472
18	477	401	468	458	457	466	445	469	436	420	459	439
19	399	325	404	380	419	372	411	414	388	363	339	437
20	387	358	409	398	423	434	451	419	407	413	383	407
21	383	349	395	337	397	429	412	400	397	358	376	384
22	301	294	287	336	347	327	361	300	307	313	263	310
23	239	186	223	244	264	289	272	246	223	262	220	211

Calculating how unsafe a neighborhood is for a particular group

- Given the data we have, we calculate how unsafe a neighborhood is for a particular group. We roughly need answers to the following questions:
  - What crimes are committed in each neighborhood?
  - Who was (or can be) affected by the committed crime?
  - How frequently is crime committed in each neighborhood?

  Based on the above information we tried to derive a metric to assess how unsafe a neighborhood is. We calculated a value in range [0,1] with 0 being extremely safe and 1 being extremely unsafe. This can be plotted as a heatmap, which could lucidly give a view of which area should be avoided by what group?

Metrics used

- The ratio of each type of crime committed in every neighborhood and the the total number of crimes committed in that neighborhood.
- How much is a group affected by each crime in a neighborhood
- The frequency by which crimes are committed in every neighborhood.

Methodology

1. Prepared a matrix of Group affected vs Type of crime. Each cell represents whether the group is affected by the type of crime or not (A logical matrix) [3 x 131].
2. Prepared a matrix of Type of Crime vs Neighborhood. Each cell contains the ratio of number of incidents of that particular crime in that neighborhood to the total number of crimes committed in that neighborhood. [131 x 58]
3. Calculated a matrix of Group vs Neighbor by performing matrix multiplication of the above two matrices. [3 x 58]
4. Calculated frequency of crime in each neighborhood, by taking a ratio of number of days of a year when a crime was committed and total number of days in a year (365). [1x58]
5. Finally, performed element wise multiplication for each group. Thus we finally calculated a [3x58] matrix.

Type of Crime vs Neighborhood

C9 0.00804828973843058

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Primary Offense Description	ASKA JUNCTIC	ALKI	ALLARD NORT	ALLARD SOUT	BELLTOWN	BITTERLAKE	IGHTON/DUNL	CAPITOL HILL	L AREA/SQUIR	/INTERNATION	MONT/RAINIER	COLUMBIA CIT	IERCIAL DUWA	RCIAL HARBOR	ITOWN COMME	ASTLAKE - EAS	ASTLAKE - WES	AUNTLEROY S	FIRST HILL
2	ADULT-VULNERABLE-FINANCIAL	0.002115954	0.000804829	0.001080205	0.00025623	0.000260247	0.00180036	0.001131542	0.000195217	0.000900385	0.000194793	0.002792182	0.001181684	0.003144654	0	6.1297E-05	0	0	0.000421053	0.000643271
3	ADULT-VULNERABLE-NEGLECT	0.001410636	0.001609658	0.00063012	0.000128115	6.50618E-05	0.00020004	0.000424328	3.25362E-05	0.000491119	6.49308E-05	0.000398883	0.001181684	0	0	8.17294E-05	0	0	0.000421053	0.000214424
4	ADULT-VULNERABLE-PHYSICAL ABUSE	0.000282127	0	0.000270051	0.000192172	0	0.00060012	0.000282885	3.25362E-05	0.000327413	6.49308E-05	0.001196649	0.000295421	0	0	2.04323E-05	0	0	0.000842105	0.000500322
5	ARSON-BUSINESS	0.000282127	0	9.00171E-05	0.000448402	0.000260247	0	0.000282885	0.000390434	0.000245559	0.000129862	0	0.000590842	0	0	0.000204323	0	0	0.000421053	0.000357373
6	ARSON-OTHER	0.000564254	0.001207243	0.001620308	0.000768689	0.000325309	0.00040008	0.000848656	0.000845941	0.000654825	0.000454516	0.000797766	0.001181684	0	0	0.00034735	0.001162791	0.000577201	0.000842105	0.000714745
7	ARSON-RESIDENCE	0.000282127	0	0.000540103	0.000192172	0.000195185	0.00090018	0.000424328	0.000455507	0.000491119	0.000129862	0.000398883	0.000295421	0	0	6.1297E-05	0	0.0002886	0.000842105	0.000571796
8	ARSON-VEHICLE	0.000282127	0	0.000180034	6.40574E-05	0.000130124	0.00060012	0.00155587	0.000130145	0.000409266	0.000194793	0.001196649	0.000886263	0.003144654	0	0.000122594	0	0.0002886	0	0.000285898
9	ASSLT-AGG-BODYFORCE	0.005360418	0.00804829	0.004950941	0.006341682	0.014834092	0.00520104	0.008910891	0.012754189	0.005320455	0.006493085	0.011966494	0.007090103	0.006289308	0.011299435	0.009419313	0.003488372	0.001731602	0.002526316	0.010578229
10	ASSLT-AGG-CHILD-BODYFORCE	0.000423191	0	0.000360068	0.000128115	0	0.0005001	0.001414427	6.50724E-05	0.000572972	0.000194793	0.001994416	0.000295421	0	0	0	0	0	0	7.14745E-05
11	ASSLT-AGG-DV-BODYFORCE	0.007194245	0.005633803	0.003510667	0.003907501	0.003903709	0.00830166	0.014144272	0.004424923	0.009904232	0.002986819	0.012365377	0.010044313	0.012578616	0	0.001818479	0.004651163	0.003751804	0.004631579	0.007075977
12	ASSLT-AGG-DV-GUN	0.000564254	0.000402414	0.000450086	0.000192172	0.000325309	0.0005001	0.002263083	9.76086E-05	0.000736678	6.49308E-05	0.000398883	0.002658789	0	0	0.000122594	0	0	0.000842105	0.000214424
13	ASSLT-AGG-DV-WEAPON	0.006065736	0.010865191	0.006031146	0.004355903	0.003708523	0.01070214	0.02446959	0.005791443	0.011868708	0.003571197	0.020343039	0.020384047	0.009433962	0.005649718	0.002165829	0.004651163	0.003174603	0.007578947	0.009434637
14	ASSLT-AGG-GUN	0.0015517	0.005633803	0.001980376	0.001153033	0.002797658	0.00210042	0.013719943	0.002342606	0.007039371	0.003311473	0.009174312	0.005317578	0.012578616	0.005649718	0.001614156	0.001162791	0.000577201	0.000842105	0.0036452
15	ASSLT-AGG-POLICE-BODYFORCE	0.000282127	0	0	0.000192172	0.000325309	0	0.000282885	0.000325362	0.000654825	0.000194793	0	0	0	0	0.000183891	0	0.0002886	0.000421053	0.000214424
16	ASSLT-AGG-POLICE-GUN	0	0	0	0	0	0	0.000282885	9.76086E-05	8.18532E-05	0	0	0	0	0	4.08647E-05	0	0	0	0
17	ASSLT-AGG-POLICE-WEAPON	0.000282127	0.002012072	0.000360068	0.000448402	0.000650618	0.00060012	0.000707214	0.001138767	0.000572972	0.000519447	0.000797766	0.001181684	0.003144654	0.005649718	0.000633403	0	0.000577201	0	0.001000643
18	ASSLT-AGG-WEAPON	0.012413599	0.014084507	0.012062292	0.018768817	0.022251139	0.017503501	0.027722772	0.023035627	0.017598428	0.023180313	0.027124053	0.018020679	0.018867925	0.028248588	0.023885415	0.008139535	0.010678211	0.009263158	0.028160961
19	BURGLARY-FORCE-NONRES	0.035548032	0.018108652	0.029075524	0.043302799	0.021730644	0.030606121	0.020367751	0.020497804	0.020790701	0.026686579	0.0331073	0.033973412	0.078616352	0.129943503	0.011462548	0.012790698	0.04992785	0.018105263	0.018154528
20	BURGLARY-FORCE-RES	0.055438003	0.063983903	0.091007291	0.03049132	0.007742355	0.075915183	0.117680339	0.023979177	0.080870918	0.005129537	0.109293977	0.098375185	0	0	0.001205509	0.053488372	0.041847042	0.097684211	0.025659352
21	BURGLARY-NOFORCE-NONRES	0.01636338	0.01167002	0.011882258	0.019793735	0.015354587	0.013402681	0.00523338	0.01848056	0.011541295	0.014674372	0.009174312	0.015361891	0.022012579	0.084745763	0.025356544	0.015116279	0.02049062	0.007578947	0.016725038
22	BURGLARY-NOFORCE-RES	0.03780505	0.061971831	0.057880997	0.032541157	0.01509434	0.062912583	0.044695898	0.03276395	0.057706475	0.005713915	0.047467092	0.049039882	0	0.005649718	0.003228311	0.05	0.05021645	0.071578947	0.030019298
23	BURGLARY-OTHER	0.000705318	0.000402414	0.000540103	0.000384344	0.000390371	0.00060012	0.000282885	0.000195217	0.000245559	6.49308E-05	0.000797766	0.000295421	0	0	4.08647E-05	0	0.000577201	0.000421053	0.000571796
24	BURGLARY-SECURE PARKING-NONRES	0.002115954	0.001207243	0.000450086	0.001985779	0.003513338	0.00260052	0.000282885	0.002733041	0.001064091	0.000973963	0	0.000886263	0.003144654	0.005649718	0.002615341	0.001162791	0.006060606	0	0.005432063
25	BURGLARY-SECURE PARKING-RES	0.032867823	0.026156942	0.0068413	0.028633656	0.023357189	0.025405081	0.003960396	0.036928583	0.01244168	0.003636128	0.000398883	0.008862629	0	0	0.002635773	0.05	0.063203463	0.000842105	0.032806804
26	CHILD-ABANDON	0.000282127	0.000402414	0.000270051	0.000128115	0	0.00040008	0.000707214	0.000553115	0.000572972	0.000194793	0.001196649	0.000590842	0	0	4.08647E-05	0	0	0.000421053	0.005360589
27	CHILD-ABUSED-NOFORCE	0.000423191	0.000402414	0.000270051	0.000128115	0	0.00030006	0.001131542	9.76086E-05	0.000654825	6.49308E-05	0	0.000886263	0	0	6.1297E-05	0	0	0.000842105	0.000428847
28	CHILD-ENDANGERMENT	0.000846382	0.000402414	0.00063012	0.000384344	0.000650618	0.00120024	0.002121641	0.000845941	0.002046329	0.000389585	0.001196649	0.000886263	0	0	0.000796862	0	0	0.001684211	0.002144236
29	CHILD-HARBOR MINOR	0	0	0	0	0	0.00010002	0	0	8.18532E-05	0	0	0	0	0	4.08647E-05	0	0	0	0
30	CHILD-NEGLECT	0	0.000804829	0.000270051	0.000384344	0.000130124	0.00080016	0.001414427	0.000325362	0.000736678	0	0.001196649	0.001477105	0	0	0.000224756	0	0	0.000842105	0.001215067
31	CHILD-OTHER	0.006912117	0.012072435	0.008731659	0.003907501	0.001821731	0.01060212	0.016831683	0.003156011	0.012114267	0.002986819	0.016354208	0.012703102	0	0	0.005904949	0.001162791	0.003751804	0.001263158	0.018869273
32	DISORDERLY CONDUCT	0.000705318	0.000402414	9.00171E-05	0.000320287	0.000780742	0.00020004	0.000424328	0.000911014	0.000572972	0.00077917	0.000398883	0.000590842	0	0	0.001634588	0	0	0	0.000142949
33	DUI-DRUGS	0.001269573	0.000804829	0.001890359	0.001985779	0.001301236	0.00330066	0.002121641	0.004392387	0.004583777	0.001947925	0.001994416	0.002658789	0.01572327	0.005649718	0.001123779	0.002325581	0.001154401	0.002526316	0.003430777
34	DUI-LIQUOR	0.027507406	0.033802817	0.027275182	0.039715585	0.024398178	0.019503901	0.028712871	0.026451928	0.024064828	0.012856308	0.023534105	0.009453471	0.194968553	0.124293785	0.011074333	0.01744186	0.011544012	0.019789474	0.020227289
35	ENDANGERMENT	0.000141064	0.000804829	0.000180034	0.000192172	0.000325309	0.00020004	0.000848656	0.000585652	0.000409266	0.000129862	0	0	0.003144654	0	0.000326918	0	0	0.000421053	0.000285898
36	GAMBLE-BETTING	0	0	0	0	0	0	0.000141443	0	0.000163706	6.49308E-05	0	0	0	0	6.1297E-05	0	0	0	0
37	GAMBLE-OPERATE	0	0	0	0	0	0	0.000141443	0		0	0	0	0	0	0	0	0	0	0
38	HARBOR - BOATING UNDER INFLUENCE	0.000282127	0	0	0.000320287	0.000325309	0	0.000707214	9.76086E-05	5	6.49308E-05	0	0.001181684	0	0	8.17294E-05	0	0	0	7.14745E-05
39	HOMICIDE-NEG-MANS-BODYFORCE	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
40	HOMICIDE-NEG-MANS-GUN	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
41	HOMICIDE-NEG-MANS-VEHICLE	0	0	0	0	6.50618E-05	0	0	0		0	0	0	0	0	0	0	0	0	0

Group affected vs Type of crime

Primary Offense Description	Student	Family	Elderly
ASSLT-AGG-GUN	1	1	1
ASSLT-AGG-DV-WEAPON	1	1	1
ASSLT-AGG-WEAPON	1	1	1
ASSLT-AGG-DV-BODYFORCE	1	1	1
ASSLT-AGG-BODYFORCE	1	1	1
ASSLT-AGG-POLICE-WEAPON	1	1	1
ASSLT-AGG-DV-GUN	1	1	1
ASSLT-AGG-CHILD-BODYFORCE	0	1	0
ASSLT-AGG-POLICE-BODYFORCE	1	1	1
ASSLT-AGG-POLICE-GUN	1	1	1
ARSON-OTHER	1	1	1
ARSON-VEHICLE	1	1	1
ARSON-RESIDENCE	1	1	1
ARSON-BUSINESS	0	0	0
BURGLARY-FORCE-RES	1	1	1
BURGLARY-FORCE-NONRES	1	1	1
BURGLARY-SECURE PARKING-RES	1	1	1
BURGLARY-NOFORCE-RES	1	1	1
BURGLARY-SECURE PARKING-NONRES	1	0	0
BURGLARY-NOFORCE-NONRES	1	1	1
THEFT-CARPROWL	0	1	0
THEFT-LICENSE PLATE	0	1	0
THEFT-AUTOACC	0	1	0
THEFT-AUTO PARTS	0	1	0
DISORDERLY CONDUCT	1	1	1
DUI-LIQUOR	1	0	0
DUI-DRUGS	1	0	0
HARBOR - BOATING UNDER INFLUENCE	1	0	0
CHILD-OTHER	0	1	0
CHILD-NEGLECT	0	1	0
ADULT-VULNERABLE-FINANCIAL	0	0	1
ENDANGERMENT	0	1	1
CHILD-ENDANGERMENT	0	1	0
CHILD-ABANDON	0	1	0
INTERFERE WITH REPORT-DV	0	1	0
ADULT-VULNERABLE-PHYSICAL ABUSE	0	0	1
ADULT-VULNERABLE-NEGLECT	0	0	1
CHILD-ABUSED-NOFORCE	0	1	0
CHILD-HARBOR MINOR	0	1	0
GAMBLE-BETTING	1	0	0

Future Work

(which we couldn’t do given the time constraint)

- Geospatial analysis:
  - How criminal activities move around in the Seattle Neighborhood. Do beats adjacent to crime prone areas also become unsafe over time?
- Metric addition:
  - What neighborhoods become unsafe during which part of the year (A periodic trend).