## What are the streets with the most crime? What type of crime and when do they happen?

Removed all records with missing location data because the impact of these missing locations was so significant that all the results ended in missing data.

To find the location with the most crime, by type of crime, hour, and day of the week a Simple K Means algorithm was used. Multiple runs were made using different numbers of clusters. The result using 3 clusters was that Washington St had the most crime for all UCR parts. Surprisingly most incidents happen during daylight between 11 am and 2 pm and between Sunday, Tuesday, and Wednesday. Part 1 crimes are the most serious type of crime happening Tuesdays around 1 pm.

💙 Weka Explorer		- 🗆 X								
Preprocess Classify Cluster Associate Select attri	butes Visualize									
Clusterer										
SimpleKMeans -Init U -max-candidates 100 -	periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 4 -A "weka.core.EuclideanDistance -R first-last"	-1 500 -num-slots								
Cluster mode	Clusterer output									
<ul> <li>Use training set</li> </ul>	(lustaving model (full training oct)									
O Supplied test set Set	Clustering model (luli training set)									
Percentage split     % 66										
Classes to clusters evaluation	kMeans									
(Nom) STREET										
Store clusters for visualization	Number of iterations: 4									
	Within cluster sum of squared errors: 130352.45027549368									
Ignore attributes	Initial starting points (random):									
	Cluster 0: Wednesday, 11, 'Part Three', 'PARK PLZ'									
Start Stop	Cluster 1: Sunday, 3, 'Part One', 'HUNTINGTON AVE'									
Result list (right-click for options)	Cluster 2: Tuesday,8,'Part One','BLACKWOOD ST'									
12:23:55 - SimpleKMeans	Missing values globally replaced with mean/mode									
12:24:46 - SimpleKMeans										
12:27:03 - SimpleKMeans	Final cluster centroids:									
12:27:17 - SimpleKMeans	Attribute Full Data 0 1	2								
12:27:53 - SimpleKMeans	(69260.0) (33845.0) (19826.0) (15589	·.0)								
12:45:51 - SimpleKMeans	DAY OF WFFK Friday Wednesday Sunday Tues	sdav								
12:47:17 - SimpleKMeans	HOUR 13.0837 14.1563 11.314 13.0	056								
12:47:59 - SimpleKMeans	UCR_PART Part Three Part Three Part Two Part	One								
12:52:43 - SimpleKMeans	STREET WASHINGTON ST WASHINGTON ST WASHINGTON ST WASHINGTON	I ST								
12:55:11 - SimpleKMeans										
12:55:35 - SimpleKMeans										
	Time taken to build model (full training data) : 0.13 seconds	ime taken to build model (full training data) : 0.13 seconds								
	=== Model and evaluation on training set ===									
	Clustered Instances									
	0 33845 (49%)									
	1 19826 (29%)									
	2 15589 (23%)									
Status										
OK		Log x								

After running the algorithm with 8 clusters, the results were somewhat similar since Washington St appeared in 6 out of the 8 clusters with the highest number of records. Again, most crimes seem to happen during daylight and on weekdays.

Weka Explorer										-	o ×
Preprocess Classify Cluster Associate Select att	tributes Visualize										
Clusterer											
Choose SimpleKMeans -init 0 -max-candidates 100	-periodic-pruning 10000 -min-density 2.0 -t1	-1.25 -12 -1.0 -N 4 -A *\	weka.core.EuclideanD	istance -R first-last" -I 5i	00 -num-slots 1 -S 10						
Cluster mode	Clusterer output										
Use training set	interal stateing points (random	J •									4
O Supplied test set Set	Cluster 0: Wednesday, 11, 'Part T Cluster 1: Sunday, 3, 'Part One',	ustes 0: Wednesday,11, 'Wart Turee', 'MARK FL2' instes 1: Sunday, 'Yart C'm', 'WATTNEYCON AVE'									
O Percentage split % 66	Cluster 2: Tuesday,8, 'Part One'	, 'BLACKWOOD ST'									
<ul> <li>Classes to clusters evaluation</li> </ul>	• Lanke Associe Sentidius II Versit           • Lanke										
(Nom) STREET	Cluster 4: Inursday, 13, 'Part On Cluster 5: Thursday, 2, 'Part Two	", 'E EIGHTH ST'	AVL-								
Store clusters for visualization	Cluster 6: Sunday, 19, 'Part Thre	e','NEWBURY ST'									
	Cluster 7: Tuesday, 11, 'Part Two	*, COMMONWEALTH AV	VE *								
Ignore attributes	Missing values globally replace	d with mean/mode									
Start Stop	Final cluster centroids:										
Result list (right-click for options)	LINE CENTER CENTER		Cluster#								
	Attribute	Full Data	0	1	2	3	4	5	6	7	
12:23:55 - SimpleKMeans		(69260.0)	(1/210.0)	(4073.0)	(4258.0)	(8016.0)	(5875.0)	(7559.0)	(15908.0)	(6361.0)	
12-24-40 - SimplekMeans	DAY_OF_WEEK	Friday	Wednesday	Sunday	Tuesday	Sunday	Thursday	Thursday	Sunday	Tuesday	
12:27:17 - SimpleKMeans	HOUR	13.0837	9.736	8.7955	13.7003	16.7651	16.6148	8.1003	16.9143	12.9159	
12:27:53 - SimpleKMeans	STREET	WASHINGTON ST	WASHINGTON ST	HUNTINGTON AVE	WASHINGTON ST	WASHINGTON ST MASSI	CHUSETTS AVE	WASHINGTON ST	WASHINGTON ST	WASHINGTON ST	
12:45:51 - SimpleKMeans											
12:47:17 - SimpleKMeans											
12:47:35 - SimpleKMeans											
12.47.59 - SimpleKMeans	Time taken to build model (full	training data) :	0.34 seconds								
12:52:43 - Simplekileans	www.Model and evaluation on tra	ining set and									
12:55:35 - SimpleKMeans		anang oco									
	Clustered Instances										
	0 17210 ( 25%)										
	1 4073 ( 6%)										
	2 4258 ( 6%)										
	4 5875 ( 8%)										
	5 7559 (11%)										
	6 15908 (23%) 7 6361 ( 5%)										
											7
Status											
ок										Log	-@* ×0

## When and where do crimes involving shootings happen?

\*\*\*After removing all records with No shooting, the dataset was reduced to 210. This was done to only focus our attention on records where there was a shooting and find out the day, hour, month, and street where the shootings occurred. By keeping the data where there was no shooting, all clusters would result in no shooting clusters.

Clusters for the shooting data with 5 clusters were distributed thru 5 different streets. Most shootings occurred during the months of March to June, the shootings happened mostly during daylight and were distributed between weekdays and weekends. It is of no surprise that the shootings mostly happened during months of warmer weather, rather than during the winter.

🖉 Weka Explorer							_	×
Preprocess Classify Cluster Associate Select att	ibutes Visualize							
Clusterer								
Choose SimpleKMeans - init 0 - max-candidates 100 -	periodic-pruning 1000(	) -min-density 2.0 -t1 -1.25 -t2 -1.0	) -N 5 -A "weka.core	e.EuclideanDistanc	e -R first-last" -l	500 -num-slots 1 -	S 10	
Cluster mode	Clusterer output							
Use training set								
O Supplied test set Set	=== Clustering	model (full training set)						Â
O Percentage split % 66								
Classes to clusters evaluation	kMeans							
(Nom) STREET								
Store clusters for visualization	Number of itera	tions: 5						
	Within cluster	sum of squared errors: 272	.0919101796333					
Ignore attributes	Initial startin	g points (random):						
Start Stop	Cluster 0: 5,Fr	iday,5,'WAYLAND ST'						
Result list (right-click for options)	Cluster 1: 9,5u Cluster 2: 6.Fr	iday, 3, 'W NEWION SI'						
	Cluster 3: 7,Mo	nday,21,'MASSACHUSETTS AVE	2					
12:23:55 - SimpleKMeans	Cluster 4: 9,Th	ursday,17,'QUINCY ST'						
12:24:46 - SimpleKMeans	Missing Values	clobally replaced with mea	n (mode					
12:27:03 - SimpleKMeans	missing values	giobally replaced with mea	n/mode					
12:27:52 SimpleKMeans	Final cluster c	entroids:						
12:45:51 SimpleKMeans			Cluster#					
12:43:31 - SimpleKMeans	Attribute	Full Data	0	1	2	3	4	
12:47:35 - SimpleKMeans		(209.0)	(50.0)	(24.0)	(45.0)	(64.0)	(26.0)	
12:47:59 - SimpleKMeans	MONTH	5.4258	6.14	4.9167	4.9333	6.3594	3.0769	
12:52:43 - SimpleKMeans	DAY_OF_WEEK	Saturday	Saturday	Sunday	Friday	Monday	Thursday	
12:55:11 - SimpleKMeans	HOUR	14.2488	8.44	11.4583	16.4667	17.2656	16.7308	
12:55:35 - SimpleKMeans	STREET	WASHINGTON ST	BRACKETT ST	MILLET ST WAS	SHINGTON ST	NORFOLK ST	WINTER ST	
13:07:15 - SimpleKMeans								
13:10:21 - SimpleKMeans								
13:10:37 - SimpleKMeans	Time taken to b	uild model (full training	data) : 0 secon	nds				
	=== Model and e	valuation on training set						
	Clustered Insta	nces						
	0 50 ( 24	8)						
	1 24 (11	s) 8)						
	2 45 (22	s)						
	3 64 (31	8)						
	4 26 (12	8)						
	-			)				 •
( <u> </u>								
Status								

## Classifying shooting data

Using hour, month, day of the week, and street to classify the possibility of a shooting showed a 75% correct classification but the rules were very hard to obtain because of the large number of Street names. The tree is difficult to read.

🥥 Weka Explorer											-		×
Preprocess Classify Cluster Associate	Select attributes Vis	sualize											
Classifier	1												
Choose J48 -C 0.1 -M 2													
Test options	Classifier output												
<ul> <li>Use training set</li> </ul>	STREET = MCNULTY	CT: Sund	lay (2.05/	0.05)									
O Supplied test set	STREET = BORDER	ST: Frida	y (5.12/0	.12)									
Cross-validation Folds 10	Number of Leaves	umber of Leaves : 110											
O Percentage split % 66	Size of the tree	ize of the tree : 123											
More options													
	Time taken to bu	ild model	: 0 secon	ıds									
(Nom) HOUR	Stratified -	******	dation :	_									
	=== Summary ===	1055-Vall	uation ==	-									
Start Stop	b analyz j												
Result list (right_click for options)	Correctly Classi	fied Inst	ances	158		75.5981	ę						
	Incorrectly Clas	sified In	stances	51		24.4019	<b>e</b>						
13:16:48 - trees.J48	Mean absolute er	ror		0.70	13								
13:17:07 - trees.J48	Root mean square	d error		0.21	.84								
13:17:24 - trees.J48	Relative absolut	e error		36.01	.05 %								
13:17:34 - trees.J48	Root relative sq	uared err	or	62.98	8 %								
13:17:52 - trees.J48	Total Number of	Instances		209									
13:18:21 - trees.J48	Detailed Acc	unacu Bu	Class										
13:19:10 - trees.J48	Decailed Acc	uracy by	C1455										
13:19:22 - trees.J48		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class			
13:19:28 - trees.J48		0.958	0.255	0.529	0.958	0.681	0.600	0.931	0.816	Saturday			
13:19:40 - trees.J48		0.750	0.012	0.931	0.750	0.831	0.807	0.902	0.816	Wednesday			
13:19:47 - trees.J48		0.591	0.005	0.929	0.591	0.722	0.719	0.933	0.670	Sunday			
13:19:54 - trees.J48		0.935	0.000	0.935	0.935	0.935	0.924	0.986	0.917	Friday			
13:20:08 - trees.J48		0.657	0.011	0.920	0.657	0.767	0.743	0.911	0.774	Monday			
13:24:10 - trees.J48		0.632	0.011	0.857	0.632	0.727	0.714	0.930	0.755	Thursday			
13:24:21 - trees.J48	Weighted Avg.	0.756	0.066	0.827	0.756	0.760	0.731	0.929	0.778				
13:26:51 - trees.J48	Confusion Ma	triv											
13:26:55 - trees.J48	CONTUSION Na	CIIX											
13:26:58 - trees.J48	a b c d e	fg <	classi	fied as									
	46 0 0 0 0	0 2	a = Satur	day									
	5 27 1 0 2	1 0	b = Wedne	sday									
	9 0 13 0 0	0 0 1	c = Sunda	ly Iatr									
	2 0 0 0 29	0 0 1	e = Frida	la y									
	10 1 0 1 0	23 0	f = Monda	iy i									
	5 1 0 0 0	1 12	g = Thurs	day									
												_	•
Status													
											Log		A x 0



Results without the Street data and using only hour, day of the week, and month, resulted in 58% correct classification, therefore these are not reliable.

😮 Weka Explorer										-		×
Preprocess Classify Cluster Associate Select attributes	Visualize											
Classifier												
Choose J48 -C 0.1 -M 2												
Test options Classifier output												
◯ Use training set												
O Supplied test set Set Number of L	Number of Leaves : 110											
Cross-validation Folds 10 Size of the	Size of the tree : 122											
O Percentage split % 66												
More options Time taken	Time taken to build model: 0 seconds											
=== Stratif	ied cross-vali	dation ==	=									
(Nom) HOUR												
Correctly C	lassified Inst	ances	123		58.8517	8						
Statt Stop Incorrectly	Classified Ir	stances	86	0.4	41.1483	8						
Result list (right-click for options)	te error		0.12	:37								
13:16:48 - trees_148 Root mean s	quared error		0.28	48								
13:17:07 - trees 148 Relative ab	solute error		51.43	34 %								
13:17:24 - trees 148 Root relation	/e squared eri	or	82.13	76 %								
13:17:34 trees 148	r of Instances	3	209									
13:17:52 - trees 148 === Detaile	Accuracy By	Class ===										
13:18:21 - trees 148	a moouracy by	01400										
13:10:10 - trees 148	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class			
12:10:22 trace 149	0.688	0.106	0.660	0.688	0.673	0.574	0.834	0.661	Saturday			
13:13:22 - 0005.340	0.583	0.075	0.618	0.583	0.600	0.520	0.897	0.647	Wednesday			
13:19:20 - uees.340	0.455	0.080	0.400	0.455	0.426	0.354	0.796	0.409	Sunday			
13:19:40 - 0005.340	0.333	0.031	0.500	0.333	0.400	0.364	0.869	0.472	Tuesday			
13.19.47 - trees.348	0.710	0.052	0.595	0.710	0.647	0.561	0.944	0.767	Monday			
13:19:54 - trees.J48	0.579	0.058	0.500	0.579	0.537	0.488	0.851	0.507	Thursday			
13:20:08 - trees.J48 Weighted Av	g. 0.589	0.075	0.592	0.589	0.587	0.515	0.862	0.620	-			
13:24:10 - trees.J48												
13:24:21 - trees.J48 === Confusi-	on Matrix ===											
13:26:51 - trees.J48			et - 4									
13:26:55-trees.J48 a D C	1 e I g < 1 5 1 2 1	Classi	iled as dav									
13:26:58 - trees.J48 4 21 1	0 2 7 1 1	b = Wedne	sdav									
4 2 10	30121	c = Sunda	y									
2 2 5	52011	d = Tuesd	ay									
3 0 0 3	2 22 0 4	e = Frida	У									
3 6 3	L 1 20 1	f = Monda	Y									
	5 0 11	g = Thurs	day									
			7									
Status												
ок										Log	-05	». х0

To run the Apriori algorithm the Month attribute was changed from Numeric to Nominal using the NumerictoNominal filter in Weka.