Atlanta Crime Analysis

Group # 3
Group members:
1. Caroline Larry
2. Harsha Vinoy
3. Tianyue Yang
4. Viera Garcia
5. Chithra Sundaram
6. Jennifer Gaddie
7. Ricardo Benitez

Description:
Analyze Atlanta crime data from 2009 to 2019 and 2020 to 2021 to identify crime patterns or trends based on different locations in Atlanta. Use Machine learning to predict different situations for crime occurrence in Atlanta from 2009 to 2019. Provide visualizations for crime occurrences from 2020 to 2021.

Datasets:
https://www.atlantapd.org/i-want-to/crime-data-downloads
- Atlanta - Crime Dataset 2020 (June - Dec) - https://www.atlantapd.org/home/showpublisheddocument/3587/637418951324270000

offense_id	rpt_date	occur_date	occur_time	poss_date	poss_time	beat	apartment	apartment	location	watch	location_t	UC2_Litera	UCR_Num	neighborh	npu	lat	long
200021209	1,02,2020	1,02,2020	15:21	1,02,2020	15:35	511	NA	NA	12 BROAD	Evening W		21	HOMICIDE	110	Downtown	M	33.75352	-84.39189
200061707	1,06,2020	1,06,2020	0:00	1,06,2020	18:25	109	NA	NA	785 PEYTO	Evening W	NA		HOMICIDE	110	Center Hill	J	33.77655	-84.4677
200100013	1,10,2020	1,09,2020	23:45	1,10,2020	0:00	110	NA	NA	1977 JONE	Morning W		20	HOMICIDE	110	West High	G	33.78738	-84.45208
200151818	1,16,2020	1,15,2020	17:15	1,15,2020	17:40	401	NA	NA	549 JOSEPH	Evening W		99	HOMICIDE	110	West End	T	33.74001	-84.41782
200180440	1,18,2020	1,18,2020	3:31	1,18,2020	3:31	206	NA	NA	28 IRBY AV	Morning W		13	HOMICIDE	110	South Tuxe	B	33.84102	-84.38018
200331738	2,03,2020	2,02,2020	19:25	2,03,2020	3:34	407	APT	1103	751 FAIRBU	NA	NA		HOMICIDE	110	Fairburn M	H	33.73454	-84.51059
200340072	2,03,2020	2,03,2020	0:15	2,03,2020	0:45	112	NA	NA	3196 DONA	Morning W		10	HOMICIDE	110	Collier Hei	I	33.78575	-84.49181
200350345	2,04,2020	2,04,2020	2:00	2,04,2020	6:00	408	NA	NA	1700 STAN	Morning W	NA		HOMICIDE	110	NA	R	33.70907	-84.44364
200461126	2,15,2020	2,15,2020	13:00	2,15,2020	13:05	207	APT	2104	1000 PEACI	Evening W		26	HOMICIDE	110	NA	E	33.81353	-84.38943
200491623	2,18,2020	2,18,2020	16:15	2,18,2020	16:24	105	NA	NA	1156 JOSEF	Evening W		13	HOMICIDE	110	Hunter Hill	K	33.76322	-84.4253
200571409	2,26,2020	2,22,2020	22:00	2,26,2020	15:00	308	NA	NA	2520 JONES	Evening W	NA		HOMICIDE	110	Leila Valle	Z	33.6849	-84.36412
200582180	2,27,2020	2,27,2020	20:20	2,27,2020	20:22	308	NA	NA	1047 REDFC	Evening W	NA		HOMICIDE	110	Rebel Vall	Z	33.68746	-84.3566
200610909	3,01,2020	3,01,2020	11:49	3,01,2020	11:49	303	NA	NA	725 GARIB	NA	NA		HOMICIDE	110	Mechanics	V	33.73521	-84.3994
200681491	3,08,2020	3,08,2020	18:39	3,08,2020	18:39	210	NA	NA	3393 PEAC	Evening W	NA		HOMICIDE	110	Lenox	B	33.84676	-84.36212
200731916	3,13,2020	3,13,2020	18:51	3,13,2020	18:52	614	NA	NA	836 GLEND	Evening W	NA		HOMICIDE	110	Midtown	E	33.77741	-84.37671
200741761	3,14,2020	3,14,2020	19:00	3,14,2020	19:15	302	NA	NA	273 FLETCH	Evening W		18	HOMICIDE	110	Pittsburgh	V	33.72585	-84.39695
200761557	3,17,2020	3,16,2020	19:00	3,16,2020	21:00	307	NA	NA	211 THORN	Day Watch	NA		HOMICIDE	110	Joyland	Y	33.71239	-84.39527

       Relevant columns:
- Report data
- Occurrence date
- Location
- UC2 Literal
- Lat
- Lng

- Atlanta - Crime Data 2021 (Jan - Aug) - https://www.atlantapd.org/home/showpublisheddocument/4230

offense id	rpt_date	occur_date	occur_day	occur_day	occur_time	poss_date	poss_time	beat	zone	location	ibr_code	UC2_Litera	neighborh	npu	lat	long
21140096	5,20,2021	5,19,2021	Wednesda	4	15:20	5,19,2021	15:25	103		1 1720	23F	LARCENY-F	Hills Park	D	33.801693	-84.43645
21282015	8,06,2021	8,06,2021	Friday	6	22:23	8,06,2021	22:24	508		5 300	23F	LARCENY-F	Downtowr	M	33.762549	-84.38787
21320347	2,01,2021	1,31,2021	Sunday	1	0:00	2,01,2021	6:53	509		5 229	23F	LARCENY-F	Downtowr	M	33.760759	-84.38377
21791668	6,28,2021	6,27,2021	Sunday	1	21:17	6,28,2021	20:00	102		1 870	23H	LARCENY-N	Vine City	L	33.755768	-84.41663
61583880	2,11,2021	2,11,2021	Thursday	5	19:15	2,11,2021	20:29	401		4 850 OAK	23F	LARCENY-F	West End	T	33.740594	-84.4158
200017887	6,15,2021	6,13,2021	Sunday	1	13:00	6,13,2020	13:30	603		6 595	23C	LARCENY-N	Old Fourth	M	33.769813	-84.38127
200041082	1,05,2021	1,04,2021	Monday	2	16:00	1,04,2021	17:00	605		6 275	23C	LARCENY-N	Grant Park	W	33.746503	-84.37854
203371412	1,27,2021	#########	Wednesda	4	18:00	#########	18:33	509		5 610		240	AUTO THEF	Downtowr	M	33.770654	-84.38947
203471614	1,04,2021	#########	Saturday	7	19:00	#########	19:30	503		5 950 W	23F	LARCENY-F	Midtown	E	33.780458	-84.38881
203510133	1,02,2021	#########	Wednesda	4	1:50	#########	7:00	409		4 1856	13A	AGG ASSAI	Southwest	R	33.704542	-84.46834
203510899	3,16,2021	#########	Wednesda	4	10:00	#########	23:00	508		5 30 IVAN	23F	LARCENY-F	Downtowr	M	33.764315	-84.38824
203510961	2,11,2021	#########	Tuesday	3	21:50	#########	22:00	304		3 12	13A	AGG ASSAI	Peoplesto	V	33.730455	-84.38757
203560702	3,21,2021	#########	Monday	2	12:00	#########	12:00	410		4 2067 BENT	23H	LARCENY-N	Campbellt	R	33.701046	-84.46646
203570108	1,04,2021	#########	Tuesday	3	1:34	#########	2:01	210		2 3450		220	BURGLARY	Buckhead	B	33.848247	-84.35465
203571590	2,08,2021	#########	Tuesday	3	6:46	#########	6:46	409		4 3113		240	AUTO THEF	Southwest	R	33.692592	-84.49101

Relevant columns:
- Report data
- Occurrence date
- Location
- UC2 Literal
- Lat
- Lng

- Atlanta crime data 2009 - 2019 -
  https://www.atlantapd.org/home/showpublisheddocument/3051/637141450644130000

Report Nu	Report Dat	Occur Date	Occur Time	Possible D	Possible Ti	Beat	Apartment	Apartment	Location	Shift Occur	Location Ty	UCR Literal	UCR #	IBR Code	Neighborh	NPU	Latitude	Longitude
90010930	1,01,2009	1,01,2009	1145	1,01,2009	1148	411			2841 GREEI	Day Watch	8	LARCENY-N	630	2303	Greenbriar	R	33.68845	-84.49328
90011083	1,01,2009	1,01,2009	1330	1,01,2009	1330	511			12 BROAD	Day Watch	9	LARCENY-N	630	2303	Downtowr	M	33.7532	-84.39201
90011208	1,01,2009	1,01,2009	1500	1,01,2009	1520	407			3500 MART	Unknown	8	LARCENY-N	630	2303	Adamsville	H	33.75735	-84.50282
90011218	1,01,2009	1,01,2009	1450	1,01,2009	1510	210			3393 PEAC	Evening W	8	LARCENY-N	630	2303	Lenox	B	33.84676	-84.36212
90011289	1,01,2009	1,01,2009	1600	1,01,2009	1700	411			2841 GREEI	Unknown	8	LARCENY-N	630	2303	Greenbriar	R	33.68677	-84.49773
90011327	1,01,2009	1,01,2009	1645	1,01,2009	1645	609			1217 CARO	Day Watch	24	LARCENY-N	630	2303	Edgewood	O	33.75786	-84.34875
90011450	1,01,2009	1,01,2009	1740	1,01,2009	1815	408			2685 METR	Evening W	12	LARCENY-N	630	2303	Venetian H	S	33.70827	-84.45385
90011608	1,01,2009	1,01,2009	1627	1,01,2009	2110	210			3393 PEAC	Evening W	8	LARCENY-N	630	2303	Lenox	B	33.84676	-84.36212
90010721	1,01,2009	1,01,2009	815	1,01,2009	820	411			3030 HEAD	Day Watch	12	LARCENY-N	630	2303	Greenbriar	R	33.68841	-84.48742
90010108	1,01,2009	#########	2000	1,01,2009	100	507			360 MARIE	Morning W	18	LARCENY-F	640	2305	Downtowr	M	33.76269	-84.39702
90010247	1,01,2009	#########	2300	1,01,2009	135	609			195 ARIZOI	Morning W	13	LARCENY-F	640	2305	Edgewood	O	33.75902	-84.3312
90010267	1,01,2009	#########	2230	1,01,2009	130	610			58 SAUNDE	Morning W	20	LARCENY-F	640	2305	Kirkwood	O	33.75441	-84.32465
90010730	1,01,2009	1,01,2009	100	1,01,2009	900	403			1457 VENE	Day Watch	20	LARCENY-F	640	2305	Venetian H	S	33.71464	-84.43612
90010745	1,01,2009	#########	2030	1,01,2009	800	208		313	3558 PIEDN	Day Watch	18	LARCENY-F	640	2305	Buckhead	B	33.85115	-84.37944
90010774	1,01,2009	1,01,2009	130	1,01,2009	923	608			456 CANDL	Day Watch	13	LARCENY-F	640	2305	Candler Pa	N	33.76675	-84.34595
90010805	1,01,2009	1,01,2009	900	1,01,2009	945	604		1305	626 DEKALI	Day Watch	18	LARCENY-F	640	2305	Old Fourth	M	33.75283	-84.36687
90010917	1,01,2009	#########	1915	1,01,2009	1137	503			422 9TH ST	Day Watch	13	LARCENY-F	640	2305	Midtown	E	33.7808	-84.37334
90011020	1,01,2009	#########	1900	1,01,2009	1230	509			590 W PEA	Day Watch	18	LARCENY-F	640	2305	Downtowr	M	33.77023	-84.38815
90011049	1,01,2009	#########	1700	#########	2200	502			188 14TH S	Day Watch	18	LARCENY-F	640	2305	Midtown	E	33.78663	-84.38139
90011054	1,01,2009	#########	1700	#########	700	614			897 MONR	Day Watch	13	LARCENY-F	640	2305	Midtown	E	33.77917	-84.36851
90011104	1,01,2009	1,01,2009	100	1,01,2009	1200	407			4076 BREW	Evening Watch		LARCENY-F	640	2305	Wilson Mil	H	33.75174	-84.52101
90011207	1,01,2009	#########	100	#########	1200	113		1234	162O HOLL	Evening W	18	LARCENY-F	640	2305	Scotts Cros	G	33.79988	-84.47023

- Need housing data for Atlanta - zip code level
- Need Demographic/population data - zip code level
- Need zipcode based on lat/lng

Proposal:
Create a database that cross references crimes, demographics, and housing cost locations in Georgia and use an ML model to predict the probability of crimes in a particular area at a particular time of day.

We will present the details in a website using HTML/CSS/Javascript, and include the interactive ML model.

Data Visualization ideas:
1. Count of crime
2. Atlanta Map - zip code level

3. Rate of crime by the hour
4. Alcohol consumption and total crime
5. Count of crime category by year

Machine Learning ideas/options: Atlanta Crime Data 2009 - 2019
- Time vs occurrence of crime - Correlation

Structure/Tools:
1. Data cleaning
   a. Jupyter notebook/Pandas
2. Database
   a. MongoDB
3. Flask App
   a. Json link
4. HTML/CSS/Bootstrap
   a. Webpage
5. Visualizations:
   a. Plotly - JS
   b. Tableau
6. Machine Learning
   a. Collab
7. Heroku:
   a. Launch website - connect everything

Process:
1. Define the roles for the team
2. Pull the data
3. Clean the data
   a. Filter out months
   b. Drop unnecessary columns
   c. Format data(dates)
4. Website will include:
   a. Homepage
   b. Page to house the ML model
   c. Map
   d. About the team page
5. Machine learning:
   a. Receive input from user
   b. Predict situation