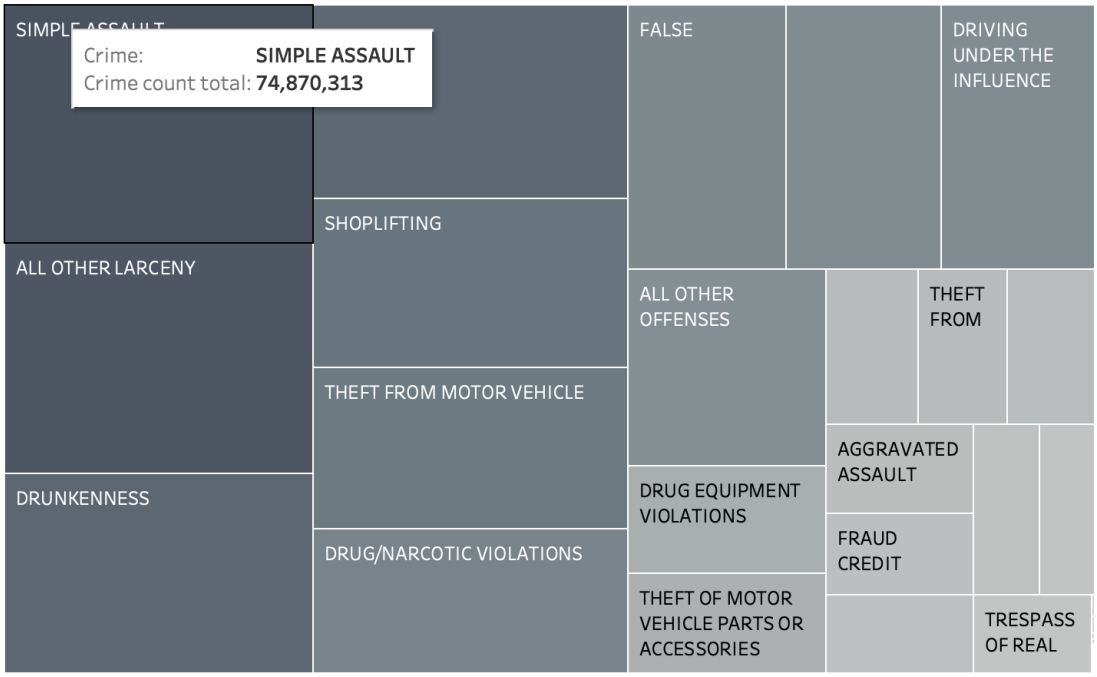
DENTON CRIME STATISTICS

Exploratory Data Analysis

April 03, 2019

HOW SAFE IS DENTON COUNTY?

Law enforcement agencies publish yearly crime records in open data portals which can be analyzed to answer basic questions like the chance of victimization, which regions are riskier than the other regions and safety related other questions. In this paper I am trying to study the Denton county crime data to answer these questions.



Basic Research Questions

- How safe is Denton?
- Which are the most dangerous streets in the Denton county?
- How Denton crime rates changed over years?
- Is there a prime time for the crimes?
- Is there any correlation between Crime rate and Population growth?
- What are your chances of being a victim?

The initial question intended to be answered from this study was regarding "How safe is Denton?" But in the course of this study other questions were added iteratively through exploratory data analysis. But before detailing the EDA and how the study progressed, a word about the data set.

The Data Set

The initial data set was obtained from City of Denton open data portal and was collected by various law enforcement agencies in the Denton county. The data set contained single entry for each crime reported detailing the type of offence, timestamp, suppressed address, type of location and the details on agency which collected the information. More than 84000 records where present in the data set from year 2009 to 2019.

_id	crime	locname	incidentdatetime	publicadress	agency	accuracy
1	DRUG/NARCOTIC VIOLATIONS	COMMERCIAL/OFFICE BUILDING	2009-12-31T00:26:00	30XX COLORADO BLVD	Denton Police Department	Address
2	ALL OTHER OFFENSES	HIGHWAY/ROAD/ALLEY	2009-12-31T00:59:00	39XX E MCKINNEY ST	Denton Police Department	Address
3	DRIVING UNDER THE INFLUENCE	HIGHWAY/ROAD/ALLEY	2009-12-31T02:08:14	LINDSEY ST // FORT WORTH DR	Denton Police Department	Intersection
4	DRUNKENNESS	BAR/NIGHT CLUB	2009-12-31T02:51:17	10XX S AVE C	Denton Police Department	Address
5	DRUNKENNESS	HIGHWAY/ROAD/ALLEY	2009-12-31T02:44:49	13XX W HICKORY ST	Denton Police Department	Address
6	DRUNKENNESS	HIGHWAY/ROAD/ALLEY	2009-12-31T03:32:31	1XX W OAK ST	Denton Police Department	Address
7	ALL OTHER OFFENSES	RESIDENCE/HOME	2009-12-31T04:24:00	14XX BERNARD ST	Denton Police Department	Address
8	DRUG/NARCOTIC VIOLATIONS	RESIDENCE/HOME	2009-12-31T04:24:00	14XX BERNARD ST	Denton Police Department	Address

Figure 1:Initial data Set

Exploratory Data Analysis

Initial analysis of data revealed fields in the data set that are of very limited use in data analysis. The field 'agency' just had two distinct entries for the 84000 records as 'Denton Police Department' and 'TX0610200'. Inference was made that the entry recorded the agency which entered the particular record in the data set. Denton Police Department was easy to understand but which agency was indicated using TX0610200 was not clear as the data dictionary or metadata failed to describe it. Hence this field was not used in the data analysis.

The second field that was identified to be problematic in the initial analysis was the field 'accuracy'. This filed also had just 2 distinct values for 84000 records and they were 'Address' and 'Intersection'. Since this field also had only limited power in a data analysis, this filed also was omitted from further analysis.

The next focus was the address field. To safeguard the privacy of victims the addresses were partially suppressed. Even though exact address was not available to facilitate location-based studies, street level addresses were obtainable. Excel built in functions were used to trim the address to street address and remove the suppressed part of the address. To aid in geocoding, city Denton and state Texas was appended using excel built in function.

The field incidentdatetime was separated out into three fields time, date and year.

A frequency analysis of the fields gave the following results.

Field	Count
Crime	97
TypeofLocation	44
Address	2800
Year	10

Figure 2: Count of Distinct fields

For year 2009, only 36 records were available and for the date Dec 31st. Hence these records were not considered for further studies. For year 2019 data was available for only 2 months and this was also omitted from further analysis.

Data Transformations

Geocoding

To facilitate location-based studies, it was decided to add geo-mapping for the street addresses. Macros were used to geocode street addresses and obtain latitude and longitude.

Crime Categorization as crime level

Based on the severity of crimes, the field 'crime level' was added which categorized crimes into high, medium and low. Felonies were classified high, misdemeanors medium and infractions as low.

Adding Population Statistics to crime data

To facilitate comparison of population to crime and calculate crime rates, yearly population of Denton county was added to the dataset.

Data Discrepancies

Initial visualization of location-based data brought to light some data discrepancies.

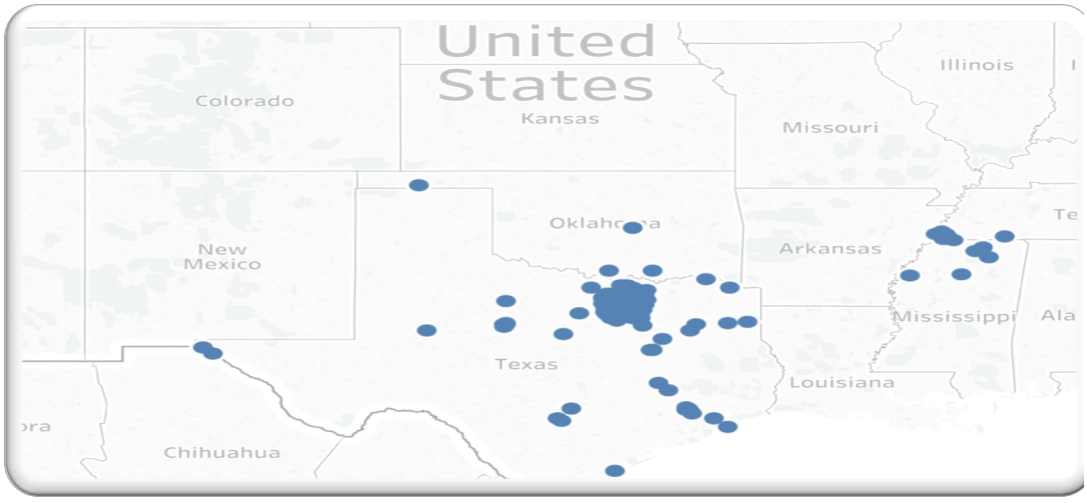


Figure 3:Data Discrepancy-Geocoded locations outside Texas

Manual geocoding was done to correct the issues aroused due to automatic geocoding.

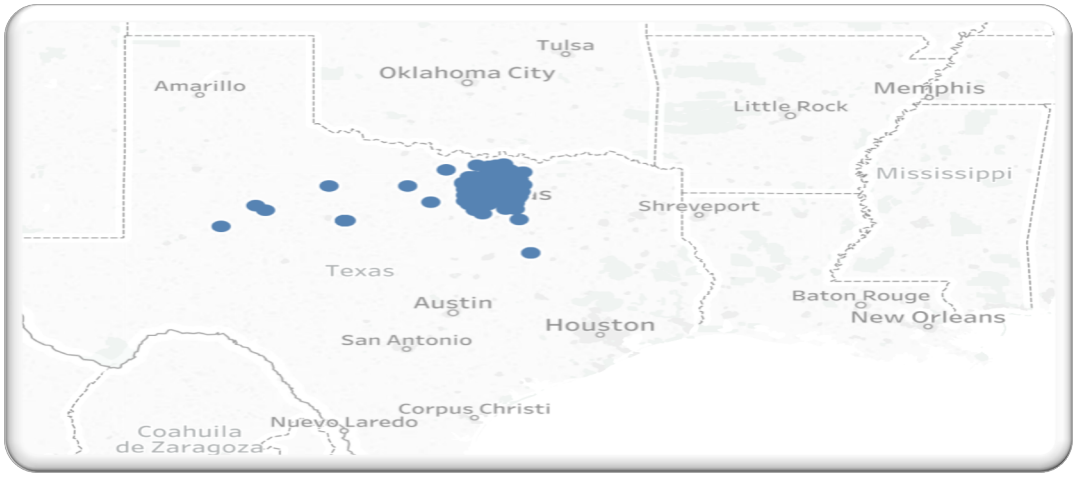


Figure 4: Manually corrected geocoded locations

Data Anomaly

But still few records showed locations outside Denton county which had to be removed as the address where strictly from inside Denton county and those records were considered anomalies.

The clean data set given below was obtained after the above-mentioned EDA.

	A	B	C	G	H	I	L	M	N
1	crime	crimelevel	TypeofLocation	Year	incidentdate	Time	Address	Latitude	Longitud
2	SIMPLE ASSAULT	Low	RESIDENCE/HOME	2010	1/25/10	18:20:00	SPRINGFIELD ,Denton,Tx	31.6620407	-96.49
3	DRIVING UNDER THE INFLUENCE	Medium	RESIDENCE/HOME	2011	10/18/11	23:40:27	SPRINGFIELD ,Denton,Tx	31.6620407	-96.49
4	LIQUOR LAW VIOLATIONS	Medium	HIGHWAY/ROAD/ALLEY	2016	4/2/16	2:53:46	FANNIN MAP 511C ,Denton,Tx	31.6710475	-96.49
5	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	CONVENIENCE STORE	2017	9/21/17	23:38:00	US-87 ,Denton,Tx	32.2630303	-101.4
6	DRIVING UNDER THE INFLUENCE	Medium	HIGHWAY/ROAD/ALLEY	2017	9/16/17	17:11:29	ORIOLE ST ,Denton,Tx	32.39953	-99.4
7	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	OTHER/UNKNOWN	2018	3/9/18	14:55:00	FOREST CEDARS DR ,Denton,Tx	32.40455	-99.5
8	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	COMMERCIAL/OFFICE BUILDING	2016	2/12/16	14:20:00	FLINT HILL ST ,Denton,Tx	32.405003	-99.5
9	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	DEPARTMENT/DISCOUNT STORE	2015	12/3/15	0:01:00	ELAM RD ,Denton,Tx	32.406728	-99.4
10	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	BANK/SAVINGS AND LOAN	2017	8/25/17	15:40:00	FM 813 W. ,Denton,Tx	32.42916	-96.4
11	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	OTHER/UNKNOWN	2018	11/20/18	18:03:00	JOHN HENRY DR ,Denton,Tx	32.5579692	-97.2
12	FALSE PRETENSE/SWINDLE	Low	BANK/SAVINGS AND LOAN	2014	3/17/14	13:33:00	S HULEN RD ,Denton,Tx	32.6315232	-97.38
13	ALL OTHER OFFENSES	Low	SCHOOL/COLLEGE	2012	5/24/12	13:20:00	W WILLIS AVE ,Denton,Tx	32.635972	-100.2
14	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	AUTO DEALERSHIP NEW/USED	2017	1/8/17	17:25:00	I20 HWY ,Denton,Tx	32.6567616	-96.74
15	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	RESIDENCE/HOME	2014	10/1/14	16:10:00	E CAMP WISDOM RD ,Denton,Tx	32.66197	-96.8

Figure 5: Clean data set after EDA

Once the cleaner data set started taking shape, it was evident that this data set is a very good candidate for location-based studies as well as time series

analysis. This realization paved way for refining the initial question of "How safe is Denton" to "How Denton crime rates changed over years?" and "Is there a prime time for the crimes?". Also "Which are the safest and riskier parts of the county" was another question that was identified at this stage.

But before diving into answering these questions a series of data frequency analysis was done to assess the distribution of data across each year.

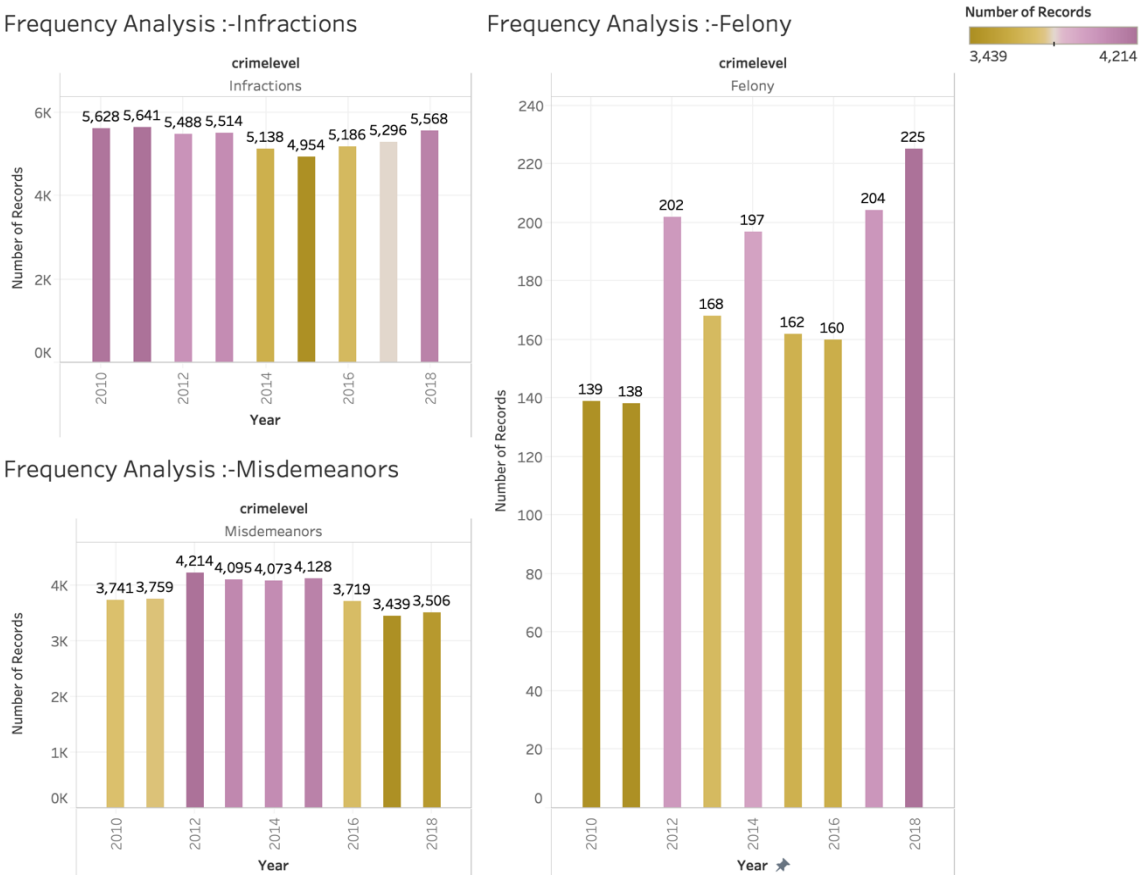


Figure 6:Crime across years segregated on crime level

This exploratory data analysis of frequency of crimes across years indicated a dip in crime rates during year 2015 except for misdemeanors.

An interesting fact from frequency analysis of location type to crime count is that all locations has count of felonies less than count of misdemeanors less than count of infractions except for Highway/road/alleys.

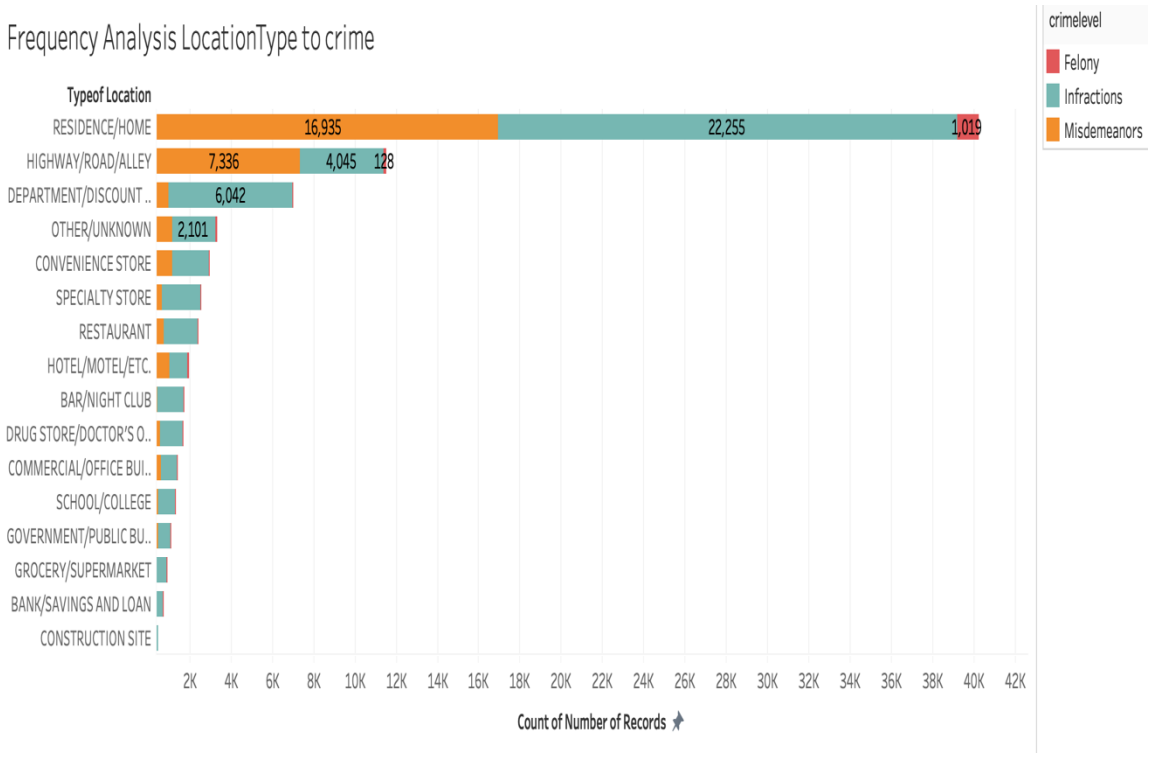


Figure 7LocationType Crime concentration

DATA ANALYSIS

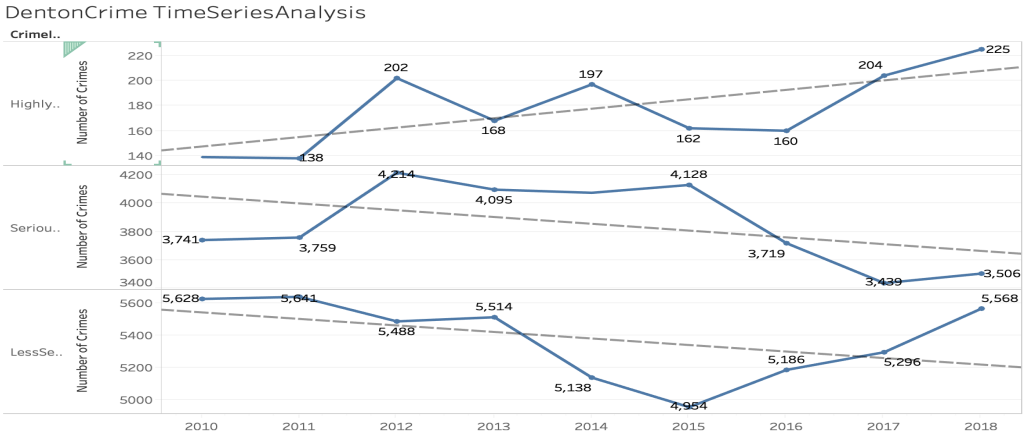


Figure 8:How number of crimes changes over years

To answer, "How Denton crime changed over years?" crimes at different crime levels were plotted against year and the above graph was obtained. From this time series analysis, we can infer that less serious crimes(infractions) and serious crimes(misdemeanors) showed a decrease over time but felonies are on rise.

To answer, "Is there a prime time for the crimes?" the crimes at different crime levels were plotted against hour of the day and the below graph was obtained.

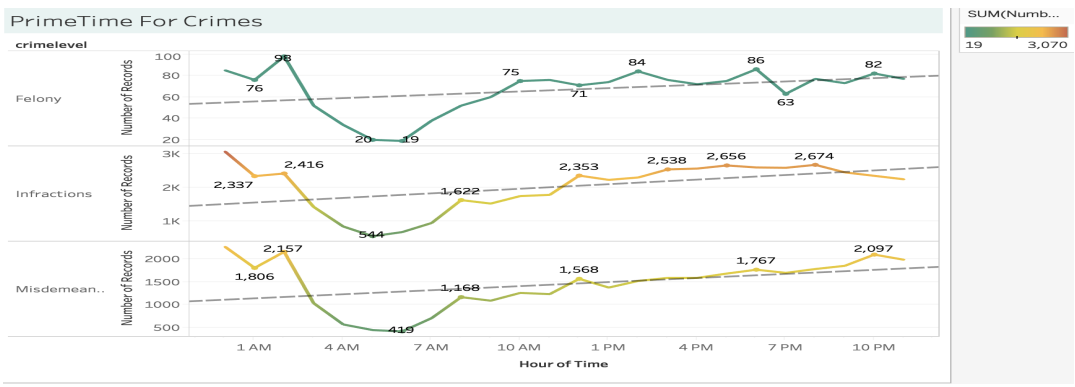


Figure 9:Crime count at different time of the day

From the graph, we can see that highest number of crimes takes place at around 2am in the morning and then there is a comparatively safe time till 6am. From there the crime count increases throughout the day.

Going back to Figure6, there is a decline in count of crimes around 2015. To understand the reason behind it, population change across years was also taken into consideration.

Year	Population	Crime count total	Felonies	Misdemeanors	Infractions
2018	136,268	9299	225	3506	5568
2017	136,268	8939	204	3439	5296
2016	134,385	9065	160	3719	5186
2015	131,276	9244	162	4128	4954
2014	128,403	9408	197	4073	5138
2013	125,184	9777	168	4095	5514
2012	123,660	9904	202	4214	5488
2011	120,198	9538	138	3759	5641
2010	117,052	9508	139	3741	5628

Figure 10: Population to crime

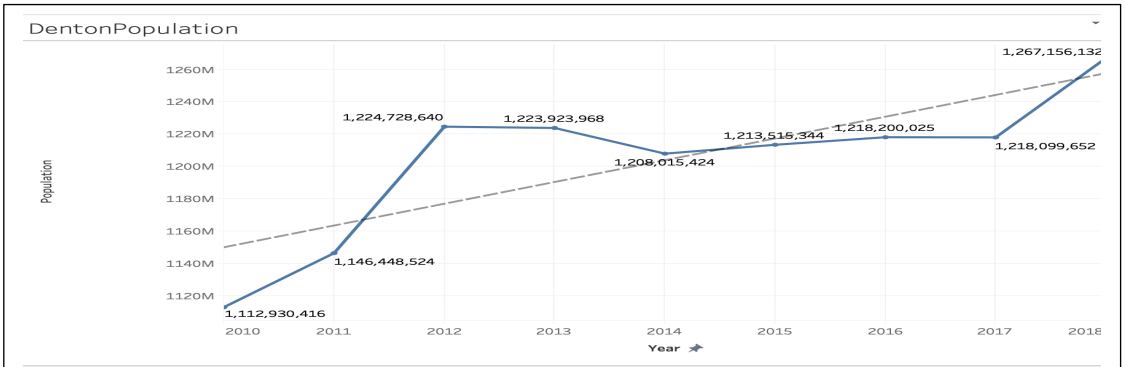


Figure 11:Denton Population growth

From figure 9, we can see a flat region of where the population of Denton was in a phase of limited growth from 2014 till 2016. The reason for declining crime rates can be attributed to this decline in population growth. This is an area which needs further studies.

When we factor in population to the crime data the next important question emerged. What are your chances of becoming a victim?

Now we come to another question about chance of being a victim in Denton county. The following table answers this question.

Year	Chance of being victim	Chance of being victim to felony	Chance of being victim to Misdemeanor	Chance of being victim to Infractions
2018	14.65404882	605.6355556	38.867085	24.47341954
2017	15.24421076	667.9803922	39.62430939	25.73036254
2016	14.82460011	839.90625	36.13471363	25.91303509
2015	14.2012116	810.345679	31.80135659	26.49899071
2014	13.64827806	651.7918782	31.52541124	24.99085247
2013	12.80392759	745.1428571	30.56996337	22.70293798
2012	12.4858643	612.1782178	29.34504034	22.53279883
2011	12.602013	871	31.97605746	21.30792413
2010	12.31089609	842.1007194	31.28896017	20.7981521

Figure 12:One in N chance of victimization

From the table it is clear that you have 1 in 14 chance of becoming a victim to crime, 1 in 605 chance of becoming a victim to felony, 1 in 38 chance of becoming a victim to misdemeanor and 1 in 24 chance of becoming a victim to infractions in Denton county according to 2018 data. Chance is calculated using crime count with respect to population.

This table put forward questions on crime rate change over time.

Crime rate is calculated by dividing the number of reported **crimes** by the total population; the result is multiplied by 100,000. This crime rate is plotted against year in the following graph.

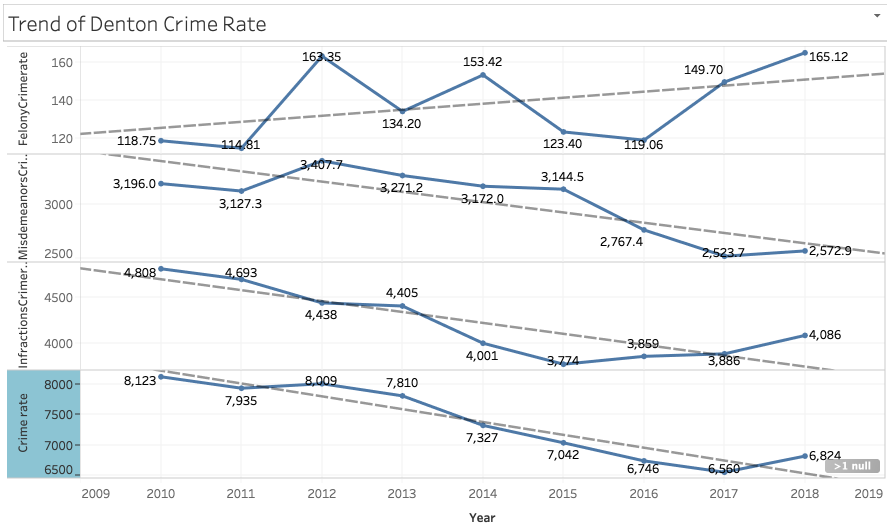


Figure 13:Trend of Denton crime rate

This graph shows a decreasing trend in all crimes except felonies which is on the rise. This is a concerning finding and needs closer look by the law enforcement agencies. Also, we need to analyze the statistical significance of this trend which is outside the scope of this paper.

Location based Studies

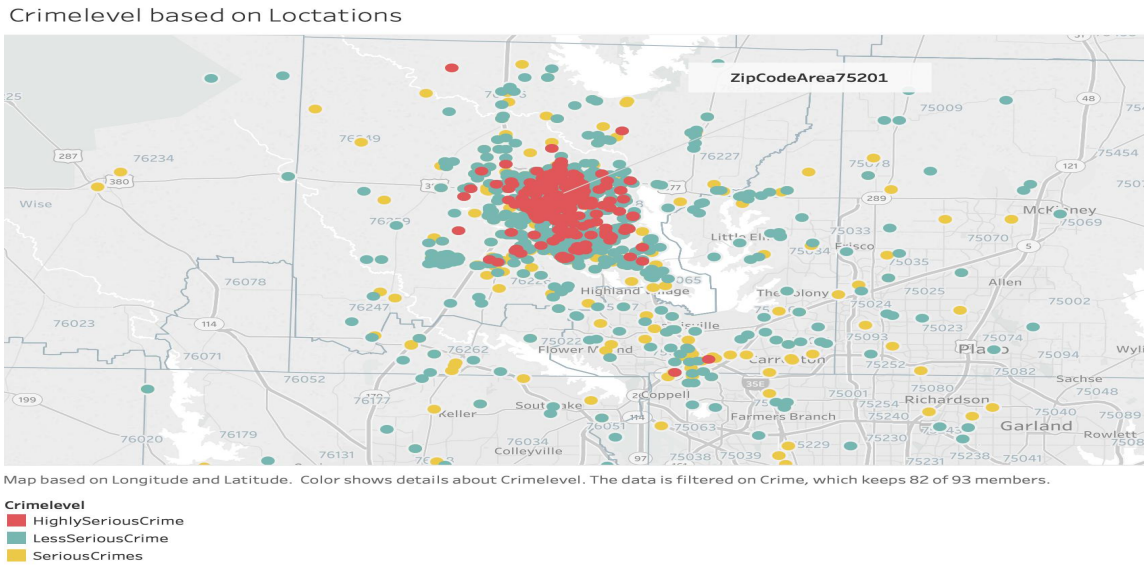


Figure 14:Prime Location of crimes

Moving on to location-based studies, in an attempt to see which region of Denton county has risk of witnessing a greater number of crimes, the below map was used. It plots the location of crimes and color code it into three different categories.

It can be seen that highly serious crimes(felonies) are concentrated in the downtown area (zip code 75201) whereas infractions and misdemeanors are spread across the entire county.

At this point a new question regarding location of crimes was identified. “What type of location is prone to crimes?” To drill down deep and identify locations/street that are at high risk, a street level view of map was plotted, and count of crimes were plotted using size of points as in the below graph.

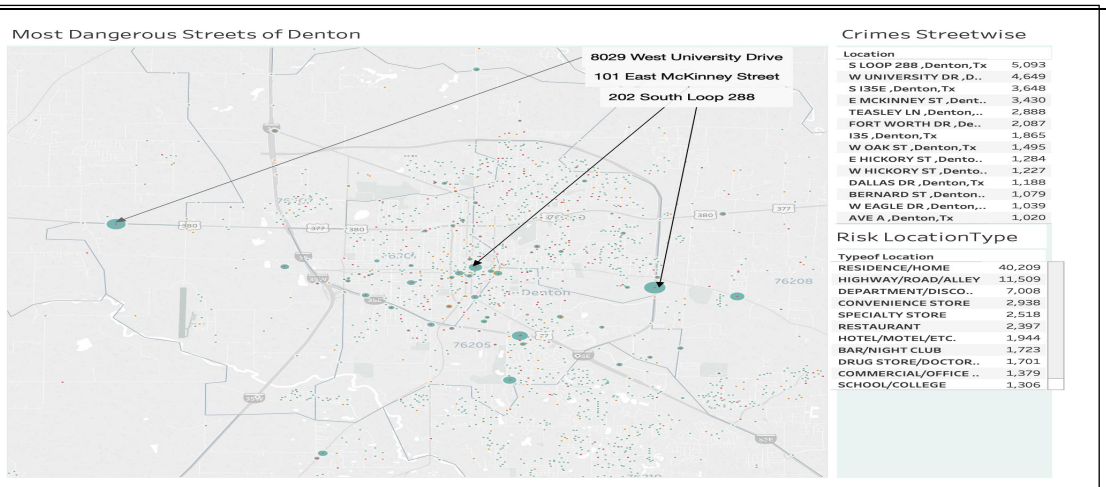


Figure 15:Street level view of crime locations

Three locations were identified to be at high risk and they are labeled in the map. Also, if you look closely at the Crimes streetwise table S 135E is the third in regard to number of crimes but is not showing up on the map. This is because this is a long road with crime locations scattered all along it. A close inspection of Risk location type table reveal highways/roads/alleys stands second in number of crimes, first being residences.

But since the number of felonies are less compared to misdemeanors and infractions a more drilled down visualization was needed to identify streetwise distribution of rimes at different levels. The following visualization helped in inferring details.

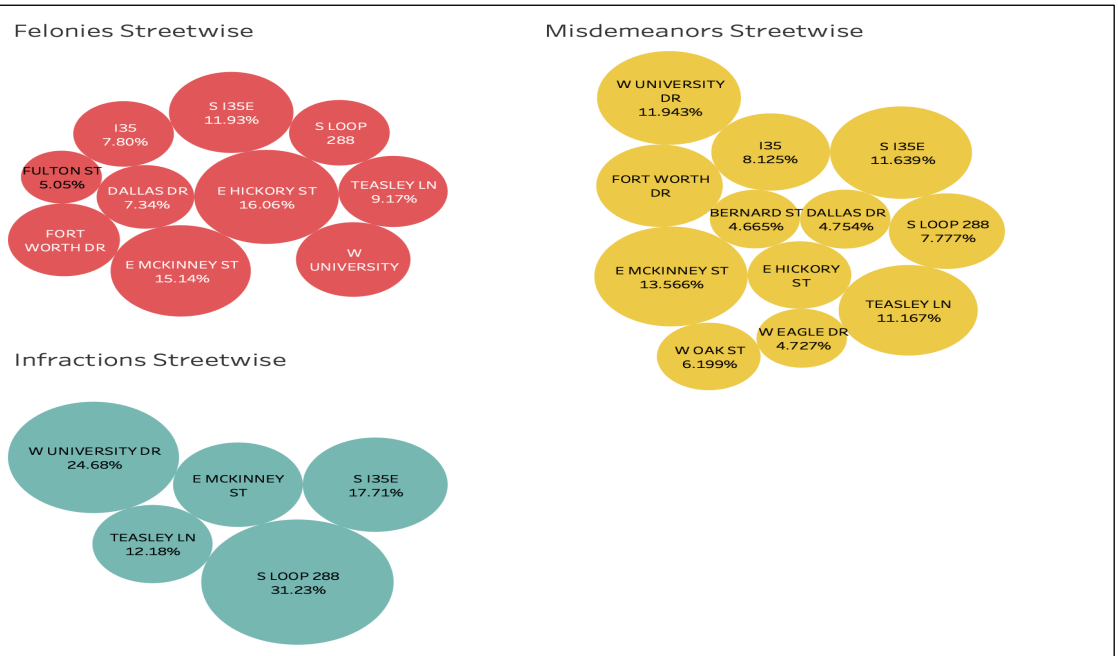


Figure 16:High Risk streets categorized into crime levels

The street identified to be riskier from previous visualization emerged as riskier in this one also. But an interesting development in this visualization is that in felonies street wise the E Hickory St emerged as the street where the highest number of felonies takes place.

Future Studies

Leading indicator analysis is the future extrapolation of this paper. This method studies time lagged incidents that can predict occurrence of crimes, location and time. Also, statistical significance analysis of observed trend needs to be performed.

Tools Used

- **Excel**
- **Tableau**