Things to Consider in Toronto Neighborhood

Dong Yi Kim | Elizabeth Salas martinez | Saif Gorges | Saloni Gupta | Sooyeon Kim

Motivation & Summary

Where to live or open a business?

Crime Rate

Income

House Prices

Questions & Data

Questions

- Which areas in Toronto have high average income/crime rates?
- Are there more red light cameras/speeding count in lower income regions?
- Is there a higher crime rate in lower income regions?
- Are higher income regions less likely to commit a serious crime?
- Is there a correlation between house prices and crime rates?
- Are there more Starbucks locations in higher income neighborhoods?

Data

- Toronto Neighbourhood Data - Toronto City Open Data
- Toronto Neighbourhood Income - Toronto City Open Data
- Toronto Crime Data - Toronto City Open Data
- Toronto House Price - Toronto City Open Data
- Red Light Cameras - Kaggle
- Speeding Count - ArcGIS API
- Starbucks Locations - Kaggle

Data Cleanup & Exploration

Solution
GeoPandas

Starbucks Locations & Red Light Cameras

```
nb = os.path.join('Resources', 'Neighbourhoods', "Neighbourhoods.shp")
regions = gpd.read_file(nb)
regions['neighbourhood'] = regions['FIELD_7'].str.replace(' \(.+\)', '').str.lower()
regions.sample(5)
```

Original Data

```
starbucks_df = pd.read_csv('Resour
starbucks_df.head()
```

	ID	Lat	Long
0	75921-104040	43.086574	-79.059356
1	3997-146205	43.077276	-79.082792
2	75525-35359	43.078906	-79.081879
3	75633-86381	43.079703	-79.082204
4	75790-96396	43.082189	-79.082469

```
regions[['neighbourhood', 'geometry']]
```

	neighbourhood	geometry
0	wychwood	POLYGON ((-79.43592 43.68015, -79.43492 43.680...
1	yonge-eglinton	POLYGON ((-79.41096 43.70408, -79.409

Used geometry (polygon) of Each neighbourhood

```
for lng_lat in lng_lats:
    point = Point(lng_lat[0], lng_lat[1])

    for i in np.arange(len(regions)):
        poly = regions.loc[i, 'geometry']

        if point.within(poly):
            neighbourhood_id.append(regions.loc[i, 'FIELD_6'])
            neighbourhood_name.append(regions.loc[i, 'neighbourhood'])
```

```
clean_starbucks
```

	Lat	Long	Hood ID	Neighbourhood
0	43.797471	-79.148805	131	rouge
1	43.743753	-79.216334	139	scarborough village
2	43.770374	-79.186485	136	west hill
3	43.816306	-79.293317	130	milliken

groupby & value_counts() methods

Main Task
Each location's latitude/longitude

Assign the locations in Toronto to each Neighbourhood

Problem with using postal codes

Final Starbucks Locations DataFrame

```
starbucks_final
```

	Hood_ID	Neighbourhood	Number of Stores
0	76	bay street corridor	26
1	77	waterfront communities-the island	19
2	75	church-yonge corridor	8

Data Cleanup & Exploration

Toronto Neighbourhood Crime Rates Data

■ Original Data Set

	_id	OBJECTID	Neighbourhood	Hood_ID	Population	Assault_2014	Assault_2015	Assault_2016	Assault_2017	Assault_2018	...	T
0	1	16	South Parkdale	85	21849	202	226	231	229	220	...	
1	2	17	South Riverdale	70	27876	215	207	236	243	304	...	
2	3	18	St.Andrew-Windfields	40	17812	53	41	48	45	55	...	
3	4	19	Taylor-Massey	61	15683	127	92	97	107	123	...	
4	5	20	Humber Summit	21	12416	76	89	118	116	109	...	

■ Final Data Frame

Neighbourhood	Hood_ID	Assault Rate	Auto Theft Rate	Break&Enter Rate	Homicide Rate	Robberey Rate	Theft Over Rate	Total Average Rate
South Parkdale	85	1148.8	91.5	407.3	4.6	151.0	100.7	317.32
South Riverdale	70	936.3	143.5	477.1	0.0	125.6	75.3	292.97
St.Andrew-Windfields	40	325.6	196.5	466.0	0.0	67.4	33.7	181.53
Taylor-Massey	61	777.9	76.5	401.7	6.4	82.9	19.1	227.42
Humber Summit	21	950.4	1087.3	459.1	24.2	225.5	177.2	487.28
...	...	...	...	...	...	...	...	...
Humewood-Cedarvale	106	320.2	111.4	181.0	0.0	69.6	27.8	118.33
Islington-City Centre West	14	504.9	391.2	282.0	2.3	79.6	84.2	224.03
Danforth	66	786.3	72.4	413.8	0.0	186.2	41.4	250.02
Rustic	28	593.5	372.2	140.8	10.1	30.2	10.1	192.82
Scarborough Village	139	1046.4	107.6	239.2	0.0	167.4	12.0	262.10

■ Original Data set

- Extract the necessary columns (Focus on 2019 Crime Data)
- Calculate the Total Average Crime Rate
  ▶️ Rate of each of crimes for 2019 per 100,000 population
- Rename the columns
- Remove the outliers

Data Cleanup & Exploration

Speeding Counts Data

Goal	Challenge	Solution
Count number of speeding tickets regionally in the GTA	Data available as geojson through API	Used pandas to parse data and append to list and create dataframe

1. JSON File

```
url = "https://services.arcgis.com/S9th0jAJ7bqgIRjw/arcgis/rest/se

response = requests.get(url).json()
pprint(response)
```


```
{'exceededTransferLimit': True,
 'features': [{'attributes': {'ACCLASS': 'Non-Fatal Injury',
                              'ACCLOC': 'Intersection Related',
                              'ACCNUM': '893184',
                              'AG_DRIV': 'Yes',
                              'ALCOHOL': 'Yes',
                              'AUTOMOBILE': 'Yes',
                              'CYCACT': None,
                              'CYCCOND': None,
                              'CYCLIST': None,
                              'CYCLISTYPE': None,
                              'DATE': 1136091600000,
                              'DISABILITY': None,
                              'DRIVACT': None,
                              'DRIVCOND': None,
                              'District': 'Toronto and East York',
                              'Division': 54,
                              'EMERG_VEH': None,
                              'FATAL_NO': None,
```


2. Parse and append

```
neighbourhood = []
speeding = []

response['features'][0]
for x in range(len(response['features'])):
    neighbourhood.append(response['features'][x]["attributes"]['Neighbourhood'])
    speeding.append(response['features'][x]['attributes']["SPEEDING"])

speeding_df = pd.DataFrame(zip(neighbourhood,speeding))
speeding_df = speeding_df.rename(columns = {0:"Neighbourhood",1:"Speeding"})
speeding_df["Neighbourhood"] = speeding_df["Neighbourhood"].str.replace(r' \(.*\)','').str.replace('\d+', '')
speeding_df.head()
```

3. Zipped table

	Neighbourhood	Speeding
0	Woodbine-Lumsden	Yes
1	Woodbine-Lumsden	Yes
2	Woodbine-Lumsden	Yes
3	Woodbine-Lumsden	Yes
4		es

```
speeding_count_df = speeding_df.groupby("Neighbourhood")["Speeding"].count()
speeding_count_df = pd.DataFrame(speeding_count_df)
speeding_count_df
```

4. Final Output

Neighbourhood	Speeding
Agincourt North	21
Agincourt South-Malvern West	23
Alderwood	14
Annex	15
Banbury-Don Mills	30

Data Cleanup & Exploration

Toronto Neighborhood Income Data

```
income_df = pd.read_excel('Resources/neighbourhood-income-d
income_df.head()
```

	Category	Topic	Attribute	City of Toronto	Agincourt North	Agincourt South-Malvern West	Alderwood
0	Population	Population	Population, 2011	2615060.0	30279.0	21988.0	11904.0
1	Population	Population	Population, 2006	2503281.0	30156.0	21562.0	11656.0

```
# Did transpose of all cities columns to 1 single column
transposed_df =income_df.loc[:,'Agincourt North':'Yorkdale-Glen Park']
cities_data = transposed_df.transpose().reset_index()
cities_data.columns = ["Neighborhood","Average_Income","Income Group"]
cities_data
```

	Neighbourhood	Average_Income	Income Group
0	South Parkdale	46896.0	25-50K
1	South Riverdale	83491.0	75-100K
2	St.Andrew-Windfields	189629.0	175-200K
3	Taylor-Massey	51242.0	50-75K
4	Humber Summit	68431.0	50-75K
...	...	...	...
135	Humewood-Cedarvale	92623.0	75-100K
136	Islington-City Centre West	89289.0	75-100K
137	Danforth	85379.0	75-100K
138	Rustic	56844.0	50-75K
139	Scarborough Village	62141.0	50-75K

140 rows × 3 columns

Steps

1) Filtered data on **Category**= 'Income', **Topic** = 'Income of households' and **Attribute** = 'Average household total income $' to narrow down the dataset.
2) Selected all the neighborhoods of Toronto and transposed the dataset such that the cities moved from columns to row.
3) Created bins and range to display average income in the form of range.

Data Cleanup & Exploration

Final : Merging all the cleaned dataframes

Crime Rate Data

Income Data

House Prices Data

Red Lights Camera / Speeding Count Data

Starbucks Locations Data

Merge on:
“Neighbourhood Name”
or
“Neighbourhood ID”

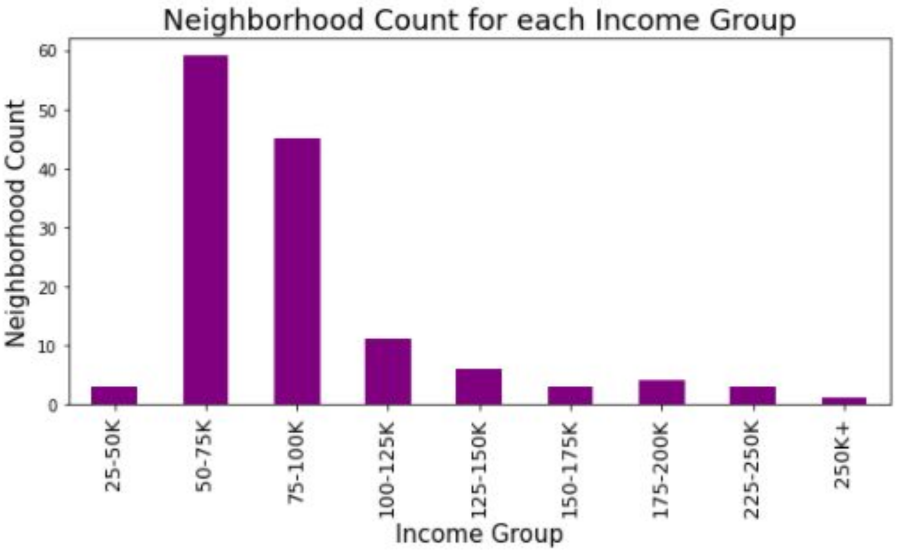
Categorized **Income Ranges** using **bin**

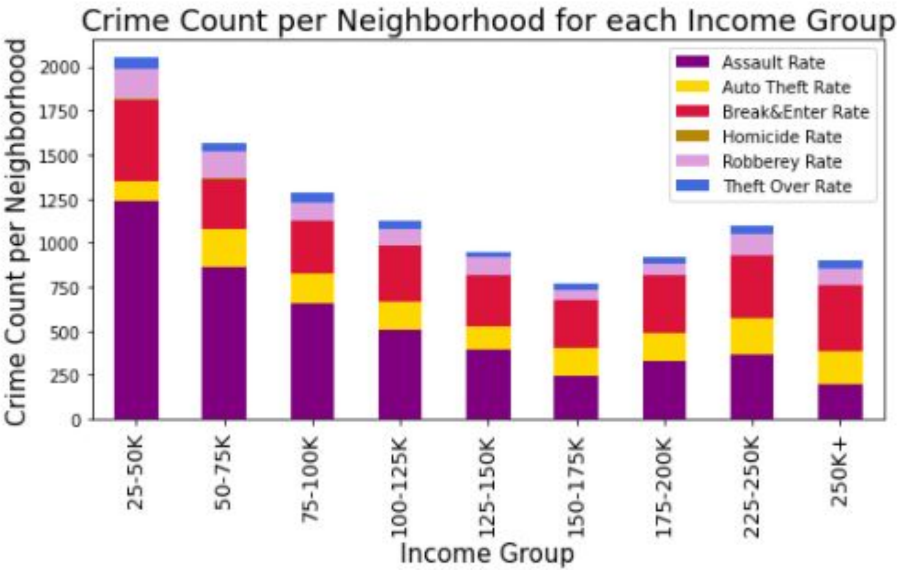
Final Merged Data

Neighbourhood	Hood_ID	Total Average Rate	Average_Income	Number of Starbucks Stores	Number of Red Lights	Home Prices	Speeding	Income Group
South Parkdale	85	317.32	46896.0	1	2	500054.0	26	25-50K
South Riverdale	70	292.97	83491.0	1	2	537087.0	20	75-100K
St.Andrew-Windfields	40	181.53	189629.0	1	1	1363202.0	3	175-200K
Taylor-Massey	61	227.42	51242.0	0	0	254151.0	4	50-75K
Humber Summit	21	487.28	68431.0	0	3	354381.0	23	50-75K
...	...	...	...	...	...	...	...	...
Humewood-Cedarvale	106	118.33	92623.0	0	0	800442.0	8	75-100K

Data Analysis - Crime vs Income Range

How does crime rate vary for neighborhoods in different Income Range?

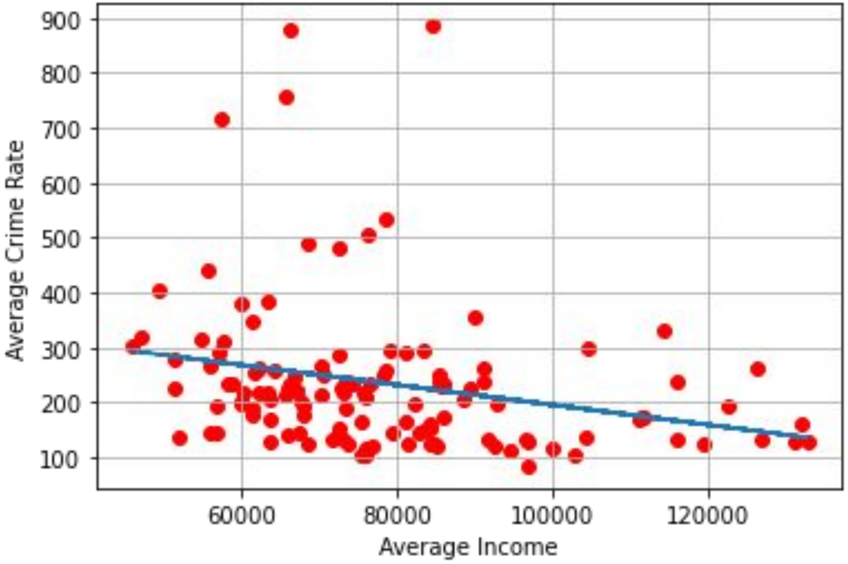




Answer: **The graph shows Crime Count per neighborhood for each income group which depicts lower Income regions have higher crime rate per neighborhood. We can also see that Assault Rate Crime is the most prominent crime across Toronto followed by Break and Enter Crime.**

Data Analysis

Is there a higher crime rate in lower income regions?

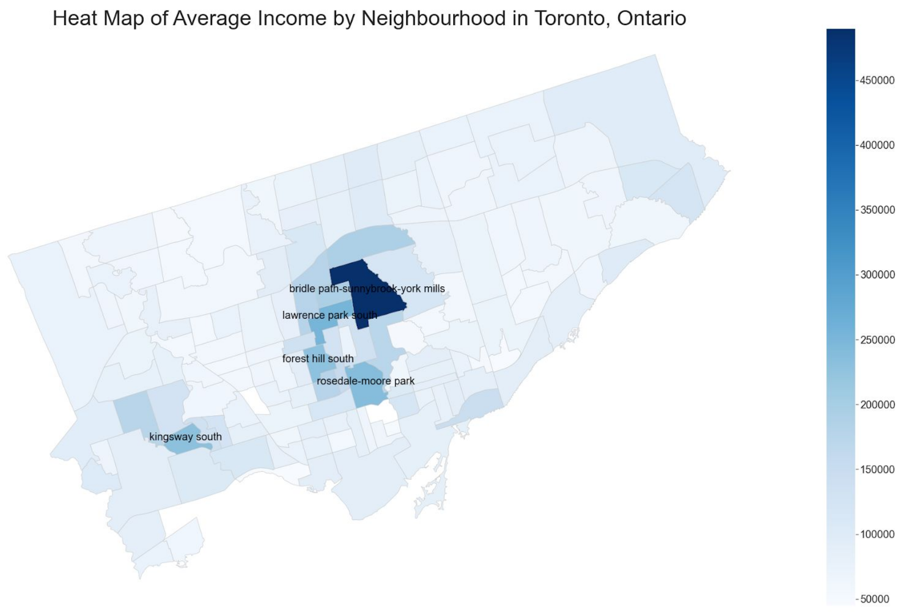


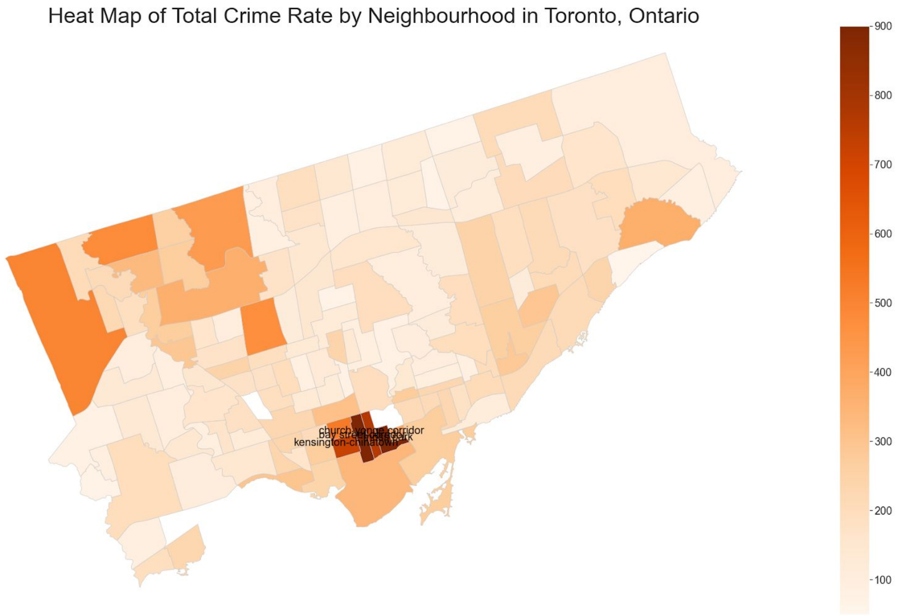
The correlation coefficient is -0.25714
p value is 0.004.

Answer: There's a negative correlation between crime rate and income rate.
Low Income area has higher crime rate.

Data Analysis

Which areas in Toronto have high average income/crime rates?



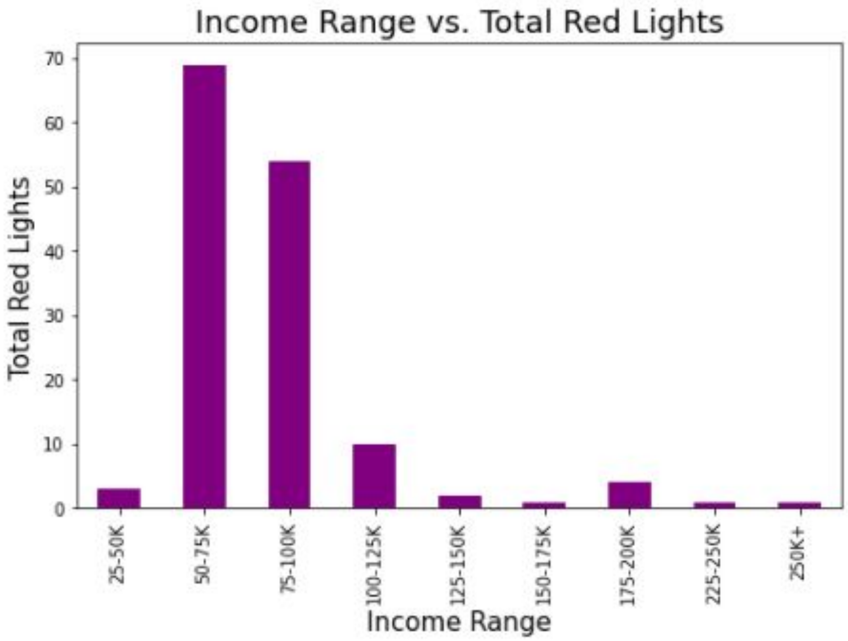


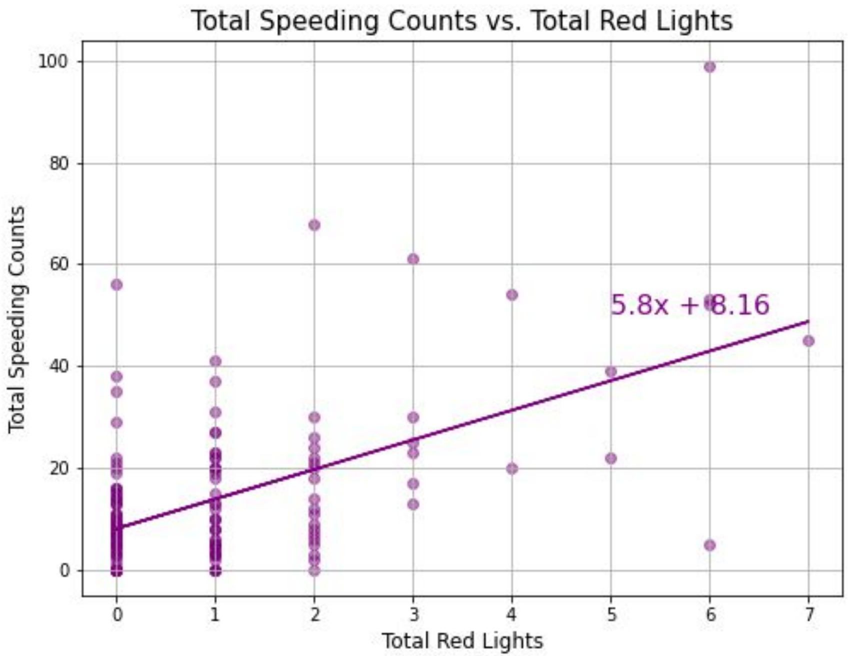
Midtown Neighbourhoods
i.e. Bridle Path-Sunnybrook-York Mills,
Lawrence Park South

Downtown & Northwest
i.e. Bay Street Corridor ,
Church-Yonge Corridor, Kensington-Chinatown

Data Analysis

Are there more red light cameras in lower income regions?

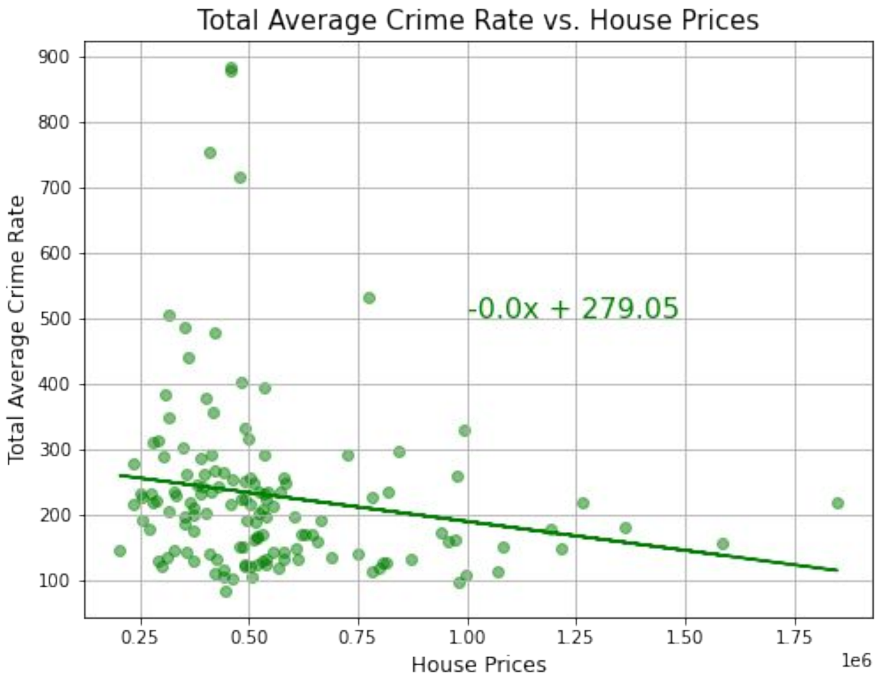




Answer: No. Red light cameras are mostly implemented in areas where people tend to speed more often.

Data Analysis

Is there a correlation between house prices and crime rates?

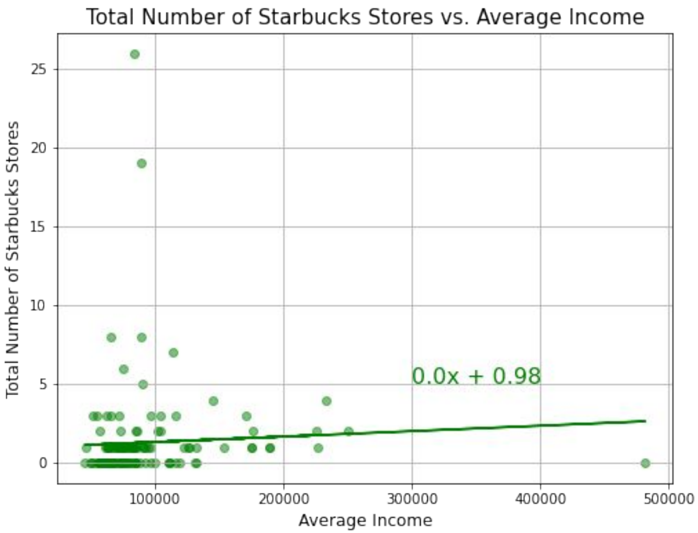


The correlation coefficient is -0.17803.
p value is 0.03601.
r square value is 0.03169.

Answer: It seems that there is a moderate negative correlation between crime rate and house prices.

Data Analysis

Are there more Starbucks locations in higher income neighborhoods?



The correlation coefficient is 0.0584.
p value is 0.49946.
r square value is 0.00341.

Answer: **NO Correlation between stores location and income**

Stores are significantly concentrated in downtown area

Foot Traffic seems to be a more important factor i.e. offices, businesses

Discussion & Post Mortem

1 As we discussed in the previous Data Analysis section, some assumptions were as what we expected, but some were not.

2 For Crime vs Income, dataset was disbalanced such that we had much greater number of neighborhoods for income range of 50-75K and 75-100K as compared to other ranges. To rectify this problem , we modified our approach and analysed the crime count per neighborhood which confirmed our initial hypothesis that on average the lower income regions have higher crime rate.

3 Red light cameras were shown to be highly prominent in areas where there were a lot of speeding tickets.

4 The criteria for categorizing Neighbourhoods in Toronto varied between different datasets which made it difficult to merge them. However, we were able to solve the problem using GeoPandas, Polygon, latitude, and longitude.

5 If we had more time, we would research about foot traffic as we assume it is an important factor in business locations strategy. We would like to find high foot traffic area of Toronto.

Q & A

Thank you.