IBM Developer
SKILLS NETWORK

The Socioeconomic Indicators and Crime Rate in Chicago Analysis with Data Science

June 16, 2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Introduction

In this study, we wanted to analyze the living conditions in the Chicago based on socioeconomic factors, the public school system, and the crime occurring in the city.

We wanted to understand the poverty level in Chicago and what factors are contributing to the poor living conditions for the families in the Chicago area

Section 1

Methodology

Methodology

Executive Summary

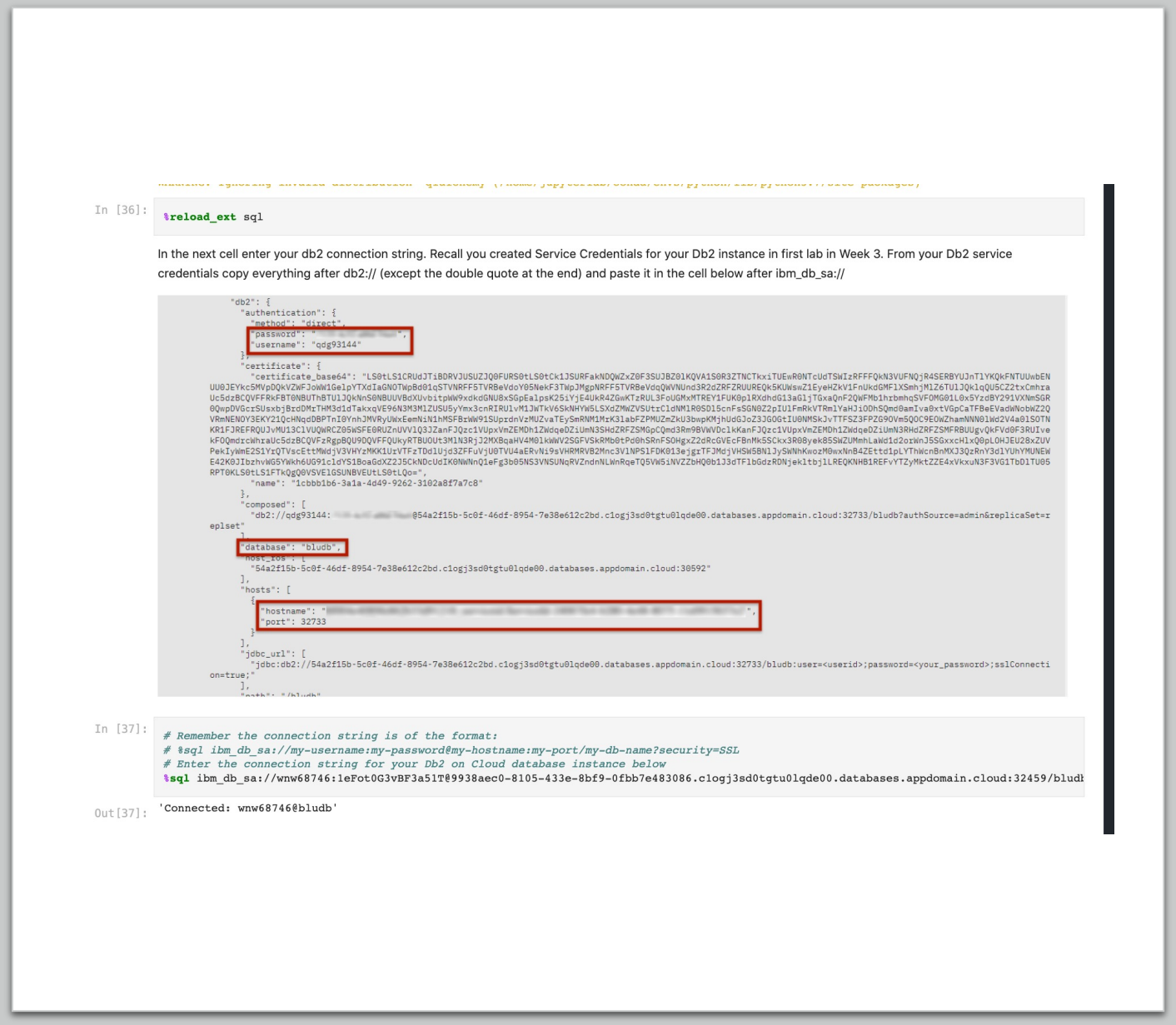
Data collection methodology:

- I collected the data through extracting tables from three different datasets: Chicago Census Data, Chicago Public Schools, and Chicago Crime Data.
- Once I organized the data into tables, I used SQL to query through the table and extract the inform we wanted.

Performed exploratory data analysis (EDA) using visualization and SQL

Data Collection

- I collected the data through extracting tables from three different datasets: Chicago Census Data, Chicago Public Schools, and Chicago Crime Data through the IBM Db2 Cloud web services. I then connected the database to my Jupyter notebook page and loaded the sql extension.



In [36]:
```
%reload_ext sql
```

In the next cell enter your db2 connection string. Recall you created Service Credentials for your Db2 instance in first lab in Week 3. From your Db2 service credentials copy everything after db2:// (except the double quote at the end) and paste it in the cell below after ibm_db_sa://


```
"db2": {
  "authentication": {
    "method": "direct",
    "password": "[redacted]",
    "username": "qdg93144"
  },
  "certificate": {
    "certificate_base64": "LS0tLS1CRUdJTiBDRVJUSUZJQ0FURS0tLS0tCk1JSURFakNDQWZxZ0F3SUJBZ0lKQVA1S0R3ZTNCTkxiTUEwR0NTcUdTSWIzRFFFQkN3VUFNQjR4SERBYUJnTlVKQkFNTUUwbEN
UU0JEYkc5MVpDQkVZWFJoWW1GelpYTXdIaGNOTWpBd01qSTVNRFF5TVRBeVdoY05NekF3TWpJMgpNRFF5TVRBeVdqQWVNUnd3R2dZRFZRUUREQk5KUWswZ1EyeHZkV1FnUkdGMFlXSmhjM1Z6TUlJQklqQU5CZ2txCmhra
Uc5dzBCQVFFRkFBT0NBUThBTUlJQkNnS0NBUUVBdXUvbitpWW9xdkdNU0xSGpEalpsK25iYjE4UkR4ZGwKTzRUL3FoUGMxMTREY1FUK0plRXdhdG13alJTaXaQnF2QWFMb1hrbmhqSVFOMG01L0x5YzdBY291VXNmSGR
0QwpDVGcrSUsxbjBrdDMrTHM3d1dTakxqVE96N3M3M1ZUSU5yYmx3cnRIRU1vM1JWTkV6SkNHYW5LSXdZMWZVSUtrClNNM1R0SD15cnFsSGN0Z2pIU1FmRkVTRmlYaHJiODhSQmd0amIva0xtVGpCaTFBeEVadWNobWZ2Q
VRmNENOY3EKY21QcHNqdDBPTnI0YnhJMVRyUWxEemNiN1hMSFBtWW91SUprdnVzMUZvaTEySmRNM1MrK3labFZPMUZmZkU3bwpKMjhUdGJoZ3JGOGtIU0NMSkJvTTFSZ3FPZG9OVm5QOC9EOWZhamNNN0lWd2V4a0lSOTN
KR1FJREFRQUJvMU13ClVUQWRCZ05WSFE0RUZnUVV1Q3JZanFJQzc1VUpxVmZEMDh1ZWdqeDZiUmN3SHdZRFZSMGpCQmd3Rm9BVWVDclkKanFJQzc1VUpxVmZEMDh1ZWdqeDZiUmN3RHdZRFZSMFRBUUgvQkFVd0F3RUIve
kF0QmdrcWhraUc5dzBCQVFzRgpBQU9DQVFFQUkyRTBUOUt3M1N3RjJ2MXBqaHV4M01kWWV2SGFVSkRMb0tPd0hSRnFSOHgxZ2dRcGVEcFBnMk5SCkx3R08yek8SSWZUMmhLaWd1d2orWnJ5SGxxcHlxQ0pLOHJEU28xZUV
PekIyWmE2S1YrQTVscEttMWdjV3VHYzMKK1UrVTFzTDd1Ujd3ZFFuVjU0TVU4aERvNi9sVHRMRVB2Mnc3V1NPS1FDK013ejgxTFJMdjVHSW5BNlJySWNhKwozM0wxNnB4ZEttd1pLYThWcnBnMXJ3QzRnY3d1YUhYMUNEW
E42K0JIbzhvWG5YWkh6UG91cldYS1BoaGdXZ2J5CkNDcUdIK0NWNnQ1eFg3b05NS3VNSUNqRVZndnNLWnRqeTQ5VW5iNVZZbHQ0b1J3dTFlbGdzRDNjekltbjlLREQKNHB1REFvYTZyMktZZE4xVkxuN3F3VG1TbDlTU05
RPT0KLS0tLS1FTkQgQ0VSVElGSUNBVEUtLS0tLQo=",
    "name": "1cbbb1b6-3a1a-4d49-9262-3102a8f7a7c8"
  },
  "composed": [
    "db2://qdg93144:[redacted]@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb?authSource=admin&replicaSet=r
eplset"
  ],
  "database": "bludb",
  "host_ips": [
    "54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30592"
  ],
  "hosts": [
    {
      "hostname": "[redacted]",
      "port": 32733
    }
  ],
  "jdbc_url": [
    "jdbc:db2://54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb:user=<userid>;password=<your_password>;sslConnecti
on=true;"
  ],
  "path": "/bludb"
```


In [37]:
```
# Remember the connection string is of the format:
# %sql ibm_db_sa://my-username:my-password@my-hostname:my-port/my-db-name?security=SSL
# Enter the connection string for your Db2 on Cloud database instance below
%sql ibm_db_sa://wnw68746:leFot0G3vBF3a51T@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bludb
```

Out[37]:
```
'Connected: wnw68746@bludb'
```

Data Collection/Data Wrangling

10.9

WNW68746.CENSUS_DATA

Back

Export to CSV

COMMUNITY_AREA_NUMBER	COMMUNITY_AREA_NAME	PERCENT_OF_HOUSING_CROWDED	PERCENT_HOUSEHOLDS_BELOW_POVERTY	PERCENT_AGED_16__UNEMPLOYED
1	Rogers Park	7.7	23.6	8.7
2	West Ridge	7.8	17.2	8.8
3	Uptown	3.8	24.0	8.9
4	Lincoln Square	3.4	10.9	8.2
5	North Center	0.3	7.5	5.2
6	Lake View	1.1	11.4	4.7
7	Lincoln Park	0.8	12.3	5.1
8	Near North Side	1.9	12.9	7.0
9	Edison Park	1.1	3.3	6.5

Items per page: 50 1–50 items 1 page 1

Data Collection/Data Wrangling

WNW68746.CHICAGO_CRIME_DATA

Back

Export to CSV

ID	CASE_NUMBER	DATE	BLOCK	IUCR	PRIMARY_TYPE	DESCRIPTION	LOCATION_DESCRIPTION	ARREST	DOMESTIC	BEAT
21149	HW519443	2013-11-03	044XX S RICHMOND ST	110	HOMICIDE	FIRST DEGREE MURDER	HOUSE	TRUE	TRUE	922
23469	JA359626	2017-07-23	015XX E 82ND ST	110	HOMICIDE	FIRST DEGREE MURDER	STREET	FALSE	FALSE	411
1326195	G021609	2001-01-11	087XX S ESCANABA AV	9901	DOMESTIC VIOLEN	DOMESTIC VIOLENCE	APARTMENT	TRUE	TRUE	423
1340847	G040244	2001-01-19	063XX N NAGLE AV	820	THEFT	$500 AND UNDER	GROCERY FOOD STORE	TRUE	FALSE	1611
1353618	G056330	2001-01-27	078XX S SAWYER AV	460	BATTERY	SIMPLE	RESIDENCE PORCH/HALLWAY	TRUE	TRUE	835
1363954	G070193	2001-02-03	004XX W WRIGHTWOOD AV	460	BATTERY	SIMPLE	RESIDENCE	FALSE	FALSE	2333

Items per page: 50 | 1–50 items | 1 page 1

Data Collection/Data Wrangling

WNW68746.CHICAGO_PUBLIC_SCHOOLS

Back

Export to CSV

SCHOOL_ID	NAME_OF_SCHOOL	Elementary, Middle, or High School	STREET_ADDRESS	CITY	STATE	ZIP_CODE	PHONE_NUMBER	LINK	NET
400018	Austin Business and Entrepreneurship Academy High School	HS	231 N Pine Ave	Chicago	IL	60644	(773) 534-6316		Wes Netv
609674	Chicago Vocational Career Academy High School	HS	2100 E 87th St	Chicago	IL	60617	(773) 535-6100	http://schoolreports.cps.edu/SchoolProgressReport_Eng/Spring2011Eng_609674.pdf	Sou Netv
609676	Paul Laurence Dunbar Career Academy High School	HS	3000 S King Dr	Chicago	IL	60616	(773) 534-9000	http://schoolreports.cps.edu/SchoolProgressReport_Eng/Spr	Sou Netv

Items per page: 50 1–50 items 1 page 1

EDA with SQL

Problems

Now write and execute SQL queries to solve assignment problems

Problem 1

Find the total number of crimes recorded in the CRIME table.

In [38]:
```
%sql SELECT COUNT(*) AS TOTAL_CRIMES FROM CHICAGO_CRIME_DATA
```

```
 * ibm_db_sa://wnw68746:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bludb
Done.
```

Out[38]:

total_crimes
533

Problem 2

List community areas with per capita income less than 11000.

In [41]:
```
%sql select COMMUNITY_AREA_NAME from CENSUS_DATA where PER_CAPITA_INCOME < 11000
```

```
 * ibm_db_sa://wnw68746:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bludb
Done.
```

Out[41]:

community_area_name
West Garfield Park
South Lawndale
Fuller Park
Riverdale

Problem 3

List all case numbers for crimes involving minors?(children are not considered minors for the purposes of crime analysis)

In [43]:
```
%sql select CASE_NUMBER from CHICAGO_CRIME_DATA where lcase(primary_type) = 'offense involving children' or lcase(description) like '%minc
```

```
 * ibm_db_sa://wnw68746:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bludb
Done.
```

Out[43]:

case_number
HL266884
HK238408

Problem 4

List all kidnapping crimes involving a child?

In [44]:
```
%sql select CASE_NUMBER from CHICAGO_CRIME_DATA where primary_type='KIDNAPPING'
```

```
 * ibm_db_sa://wnw68746:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bludb
Done.
```

Out[44]:

case_number
HN144152

Problem 5

What kinds of crimes were recorded at schools?

In [47]:
```
%sql select PRIMARY_TYPE as crime_recorded_atschool from CHICAGO_CRIME_DATA where LOCATION_DESCRIPTION like '%SCHOOL%'
```

```
 * ibm_db_sa://wnw68746:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bludb
Done.
```

Out[47]:

crime_recorded_atschool
BATTERY
BATTERY
BATTERY
BATTERY
BATTERY
CRIMINAL DAMAGE
NARCOTICS
NARCOTICS
ASSAULT
CRIMINAL TRESPA
PUBLIC PEACE VI
PUBLIC PEACE VI

Problem 6

List the average safety score for each type of school.

In [29]:
```
%sql select avg(safety_score) as average_safety_score from CHICAGO_PUBLIC_SCHOOLS
```

```
 * ibm_db_sa://wnw68746:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bludb
(ibm_db_dbi.ProgrammingError) ibm_db_dbi::ProgrammingError: SQLNumResultCols failed: [IBM][CLI Driver][DB2/LINUXX8664] SQL0206N  "SAFETY_S
CORE" is not valid in the context where it is used.  SQLSTATE=42703 SQLCODE=-206
[SQL: select avg(safety_score) as average_safety_score from CHICAGO_PUBLIC_SCHOOLS]
(Background on this error at: http://sqlalche.me/e/13/f405)
```

EDA with SQL

EDA with SQL

Problem 7

List 5 community areas with highest % of households below poverty line

In [49]:
```
%sql select COMMUNITY_AREA_NUMBER, COMMUNITY_AREA_NAME from CENSUS_DATA order by percent_households_below_poverty desc limit 5;
```

```
 * ibm_db_sa://wnw68746:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu01qde00.databases.appdomain.cloud:32459/bludb
Done.
```

Out[49]:

community_area_number	community_area_name
54	Riverdale
37	Fuller Park
68	Englewood
29	North Lawndale
27	East Garfield Park

Problem 8

Which community area is most crime prone?

In [50]:
```
%sql select COMMUNITY_AREA_NUMBER,count(case_number) as no_of_cases from CHICAGO_CRIME_DATA group by COMMUNITY_AREA_NUMBER order by no_of
```

```
 * ibm_db_sa://wnw68746:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu01qde00.databases.appdomain.cloud:32459/bludb
Done.
```

Out[50]:

community_area_number	no_of_cases
25	43

Double-click **here** for a hint

Problem 9

Use a sub-query to find the name of the community area with highest hardship index

In [51]:
```
%sql select COMMUNITY_AREA_NAME from CENSUS_DATA where hardship_index=(select max(hardship_index) from CENSUS_DATA)
```

```
 * ibm_db_sa://wnw68746:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu01qde00.databases.appdomain.cloud:32459/bludb
Done.
```

Out[51]:

community_area_name
Riverdale

Problem 10

Use a sub-query to determine the Community Area Name with most number of crimes?

In [ ]:
```
%sql select COMMUNITY_AREA_NAME from CENSUS_DATA where COMMUNITY_AREA_NUMBER = (select community_area_number from (select * from (select (
```

Github URL

- https://github.com/sniyas2002/Python-Project-for-Data-Science/blob/main/DB0201EN-PeerAssign-v5.ipynb

Predictive Analysis (Classification)

- From this analysis, we discovered that the total number of crimes recorded, community areas with per capita income less than 11000, case numbers for crimes involving minors, kidnapping crimes involving a child, types of crimes recorded at schools, average safety score for each type of school, 5 community areas with highest % of households below poverty line, community area that is most crime prone, the name of the community area with highest hardship index, and the Community Area Name with most number of crimes.

Results

- We know that there were 533 crimes recorded from the Chicago crime data.
- We know West Garfield Park, South Lawndale, Fuller Park, and Riverdale have a per capita income less than 11000.
- We know HL266884 and HK238408 are the case numbers involving minors.
- We know HN144152 is the case number for a kidnapping involving a child.
- We know crimes recorded at school were BATTERY, CRIMINAL DAMAGE, NARCOTICS, ASSAULT, CRIMINAL TRESPASSING, and PUBLIC PEACE VI.

Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations

<Folium Map Screenshot 1>

- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites’ location markers on a global map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 2>

- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Conclusions

- In conclusion, we can determine that socioeconomic factors have influence on the crime rate in the areas and public schools. Areas with low per capita incomes and high hardship index had the most crime in the area and had students from public schools from that area who committed crime.
- We need to address the poverty issue in Chicago and improve the living conditions for people, so they don’t need to rely on crimes to live.

Innovative Insights

- Some innovative insights to take away is to look at every detail in the table because you can find some surprising takeaways such as the type of crime done at school and which community the student charged was from.