Analysis of Chicago public schools neighbourhoods

Keerthi Bandarapally

Project Introduction/Business Problem:

- The main purpose of this project is to analyze Chicago public schools data and suggest parents of school children better locality and area to look out for housing. Analysis performed includes getting the geocodes of the addresses already present in the dataset using foursquare API. Results from this project helps parents of school children to find better housing area with higher safety and affordable within their requirements.

Data:

- Dataset with the name Chicago Public schools has got columns like school name, type of school(Elementary, Middle, or High School), street address, city, state and zipcode of schools, school certificate(healthy or not),Safety icon,safety score, average performance of the students, misconduct rate of students, student to teacher ratio, community area name, police distrcit and many other details. I will be not using all the columns for analysis however, only the required columns are retained and unwanted columns will be dropped. During the analysis, address columns are used to generate geocodes using foursquare API and other columns like safety score, healthy_school_certified,misconduct rate of students will be primely considered in order to suggest the best school to the parents. This particular dataset is available online. For convenience, I downloaded the dataset and below is the dataset I imported.

Dataset:

```
os.getcwd()

df=pd.read_csv('Chicago_Public_Schools.csv')
```

```
df.head()
```

	School ID	NAME_OF_SCHOOL	Elementary, Middle, or High School	Street Address	City	State	ZIP Code	Network Manager	Collaborative Name	Adequate Yearly Progress Made?
0	610038	Abraham Lincoln Elementary School	ES	615 W Kemper Pl	Chicago	IL	60614	Fullerton Elementary Network	NORTH-NORTHWEST SIDE COLLABORATIVE	No
1	610281	Adam Clayton Powell Paideia Community Academy ...	ES	7511 S South Shore Dr	Chicago	IL	60649	Skyway Elementary Network	SOUTH SIDE COLLABORATIVE	No
2	610185	Adlai E Stevenson Elementary School	ES	8010 S Kostner Ave	Chicago	IL	60652	Midway Elementary Network	SOUTHWEST SIDE COLLABORATIVE	No
3	609993	Agustin Lara Elementary Academy	ES	4619 S Wolcott Ave	Chicago	IL	60609	Pershing Elementary Network	SOUTHWEST SIDE COLLABORATIVE	No
4	610513	Air Force Academy High School	HS	3630 S Wells St	Chicago	IL	60609	Southwest Side High School Network	SOUTHWEST SIDE COLLABORATIVE	NDA

```
df.columns
```

```
Index(['School ID', 'NAME_OF_SCHOOL', 'Elementary, Middle, or High School',
       'Street Address', 'City', 'State', 'ZIP Code', 'Network Manager',
       'Collaborative Name', 'Adequate Yearly Progress Made? ',
       'Track Schedule', 'CPS Performance Policy Status',
       'CPS Performance Policy Level', 'HEALTHY_SCHOOL_CERTIFIED',
       'Safety Icon ', 'SAFETY_SCORE', 'Family Involvement Icon',
       'Family Involvement Score', 'Environment Icon ', 'Environment Score',
       'Instruction Icon ', 'Instruction Score', 'Leaders Icon ',
       'Leaders Score ', 'Teachers Icon ', 'Teachers Score',
       'Parent Engagement Icon ', 'Parent Engagement Score',
       'Parent Environment Icon', 'Parent Environment Score',
       'AVERAGE_STUDENT_ATTENDANCE', 'Rate of Misconducts (per 100 students) ',
       'Average Teacher Attendance',
       'Individualized Education Program Compliance Rate ', 'Pk-2 Literacy %',
       'Pk-2 Math %', 'Gr3-5 Grade Level Math %', 'Gr3-5 Grade Level Read % ',
       'Gr3-5 Keep Pace Read %', 'Gr3-5 Keep Pace Math %',
       'Gr6-8 Grade Level Math %', 'Gr6-8 Grade Level Read %',
       'Gr6-8 Keep Pace Math%', 'Gr6-8 Keep Pace Read %',
       'Gr-8 Explore Math %', 'Gr-8 Explore Read %', 'ISAT Exceeding Math %',
       'ISAT Exceeding Reading % ', 'ISAT Value Add Math',
       'ISAT Value Add Read', 'ISAT Value Add Color Math',
       'ISAT Value Add Color Read', 'Students Taking  Algebra %',
       'Students Passing  Algebra %', '9th Grade EXPLORE (2009) ',
       '9th Grade EXPLORE (2010) ', '10th Grade PLAN (2009) ',
       '10th Grade PLAN (2010) ', 'Net Change EXPLORE and PLAN',
       '11th Grade Average ACT (2011) ', 'Net Change PLAN and ACT',
       'College Eligibility %', 'Graduation Rate %',
       'College Enrollment Rate %', 'COLLEGE_ENROLLMENT',
       'General Services Route ', 'Freshman on Track Rate %', 'X_COORDINATE',
       'Y_COORDINATE', 'Latitude', 'Longitude', 'COMMUNITY_AREA_NUMBER',
       'COMMUNITY_AREA_NAME', 'Ward', 'Police District', 'Location'],
      dtype='object')
```

Exploratory Data Analysis:

In [28]:
```
med = df1['SAFETY_SCORE'].median()
med
```

Out[28]:
```
48.0
```

In [29]:
```
mod=df1['SAFETY_SCORE'].mode()
mod
```

Out[29]:
```
0    99.0
dtype: float64
```

In [30]:
```
df1.corr()
```

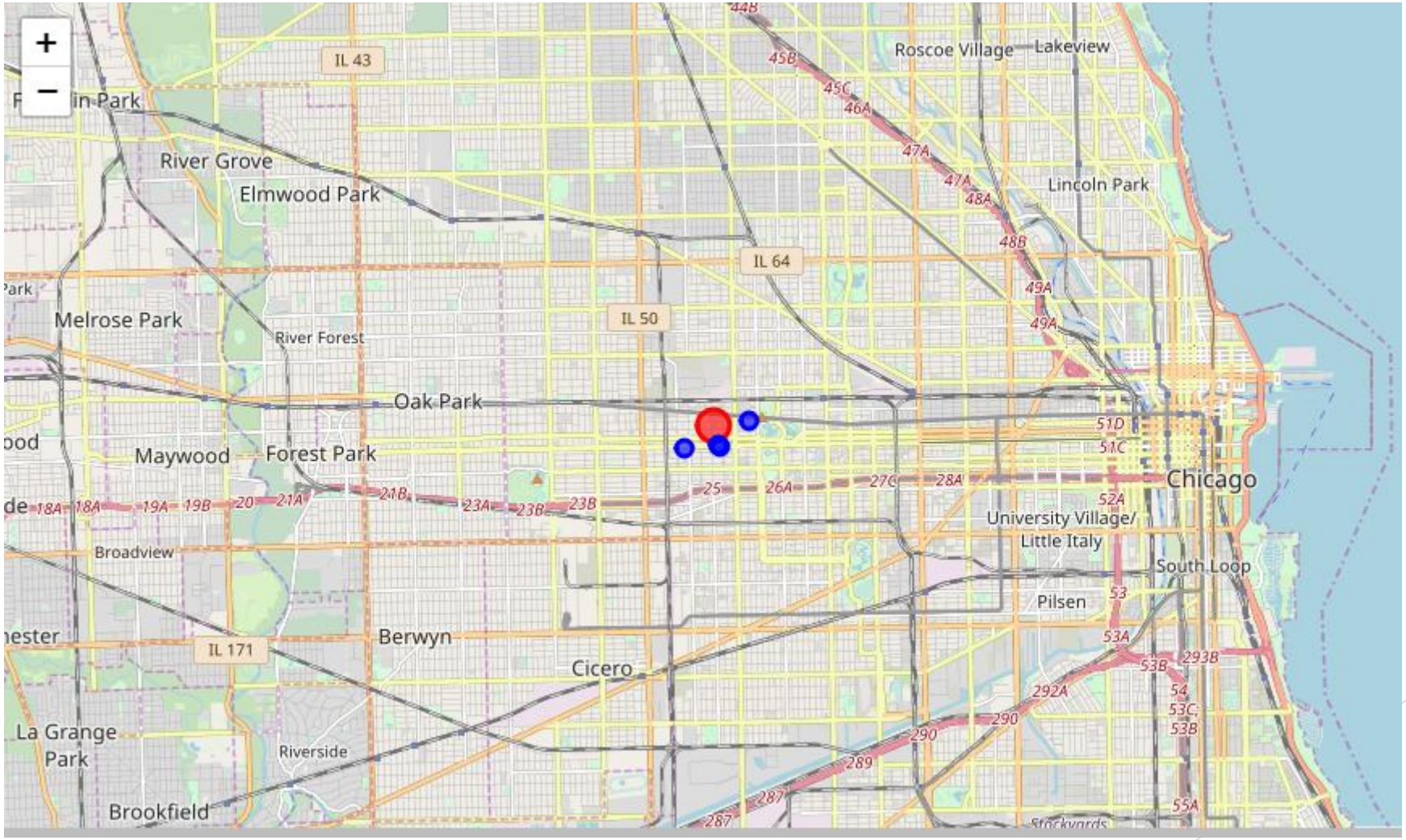
Out[30]:

	School ID	ZIP Code	SAFETY_SCORE	Rate of Misconducts (per 100 students)	COLLEGE_ENROLLMENT	COMMUN
School ID	1.000000	-0.030672	0.025852	0.014200	0.021080	0.024711
ZIP Code	-0.030672	1.000000	-0.016577	0.036830	0.021061	0.038242
SAFETY_SCORE	0.025852	-0.016577	1.000000	-0.401740	0.089813	-0.375084
Rate of Misconducts (per 100 students)	0.014200	0.036830	-0.401740	1.000000	-0.213106	0.165289
COLLEGE_ENROLLMENT	0.021080	0.021061	0.089813	-0.213106	1.000000	-0.093115
COMMUNITY_AREA_NUMBER	0.024711	0.038242	-0.375084	0.165289	-0.093115	1.000000
Ward	-0.018890	0.151968	0.292289	-0.178519	0.174740	-0.529770
Police District	-0.021886	0.162293	0.306982	-0.181752	0.199743	-0.469796

RESULT:

Hence from the exploratory data analysis and other analysis conducted,some of the zipcodes of the ares which can be suggested to the parents of school children are 60617 60621 60649 60617 60628 60636 60647 60636 60629 60628 60628 60629 60619 60637 60637 60620 60624 60636 60615

Using foursquare API, restaurants around one of the addresses of one of the zipcode mentioned was plotted. Three restaurants were plotted as shown below.



Conclusion:

Using Foursquare API, we could find the venues around the addresses of schools having higher safety score. Zipcodes and addresses of the good schools are identified using Data Analysis methods