

Timely and effective care

An analysis of the public dataset from CMS

Pranav Kumar Mishra

Wednesday, May 29, 2024

Table of contents

Import Data	1
Parameters	1
Data Processing	2
Filtering for Zipcode:	2
Variable exploration: Score	2
Generate Figures and Tables	4
Save the run dataset	5

The computer’s name is pminformatics.rush.edu The current directory has been set to /home/pranav/work/pranavmishra90/courses/by-mishra/research-data-management

Import Data

Parameters

Caution: The run directory already exists for Run # 002 A notebook directory is already present: /home/pranav/work/pranavmishra90/courses/by-mishra/research-data-management/data/analysis/timely_effective/runs/run_002/notebooks/

Data Processing

Data Sanitization

```
Facility ID          object
Facility Name        object
City/Town            object
State                object
ZIP Code              int64
County/Parish         object
Measure ID           object
Measure Name          object
Compared to National object
Score                float64
Footnote              object
Start Date            datetime64[ns]
End Date              datetime64[ns]
dtype: object
```

Subset the data

Filtering for Zipcode:

The raw dataset `care_raw_df` has 4,405 zip codes with 173,304 facilities.

After filtering, 158 zip codes with 6,264 facilities remain in the dataset `care_df`

Outlier Analysis

Variable exploration: Score

	Score
count	3877.000000
mean	3385.893854
std	17857.281487
min	0.000000
25%	0.269000
50%	1.346000

	Score
75%	11.840000
max	329722.000000

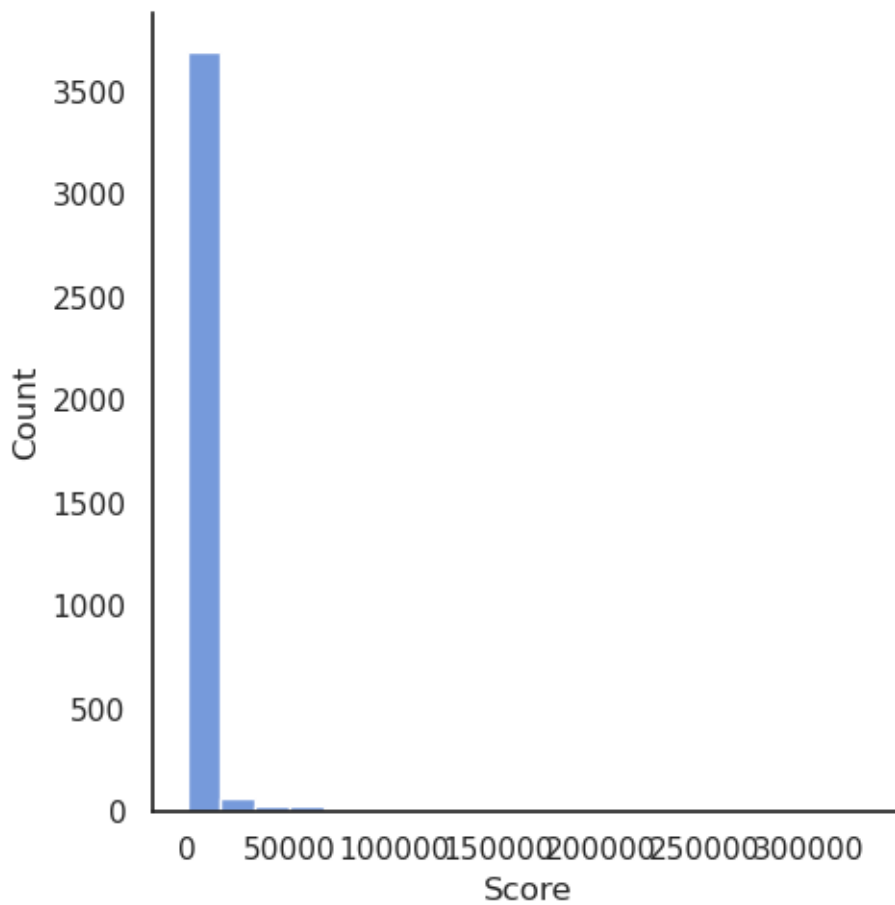
Extremes of data

Lowest Values:

[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]

Highest Values:

[181596.0, 194569.0, 201314.0, 203229.0, 215520.0, 226034.0, 229620.0, 249798.0, 294771.0, 329722.0]



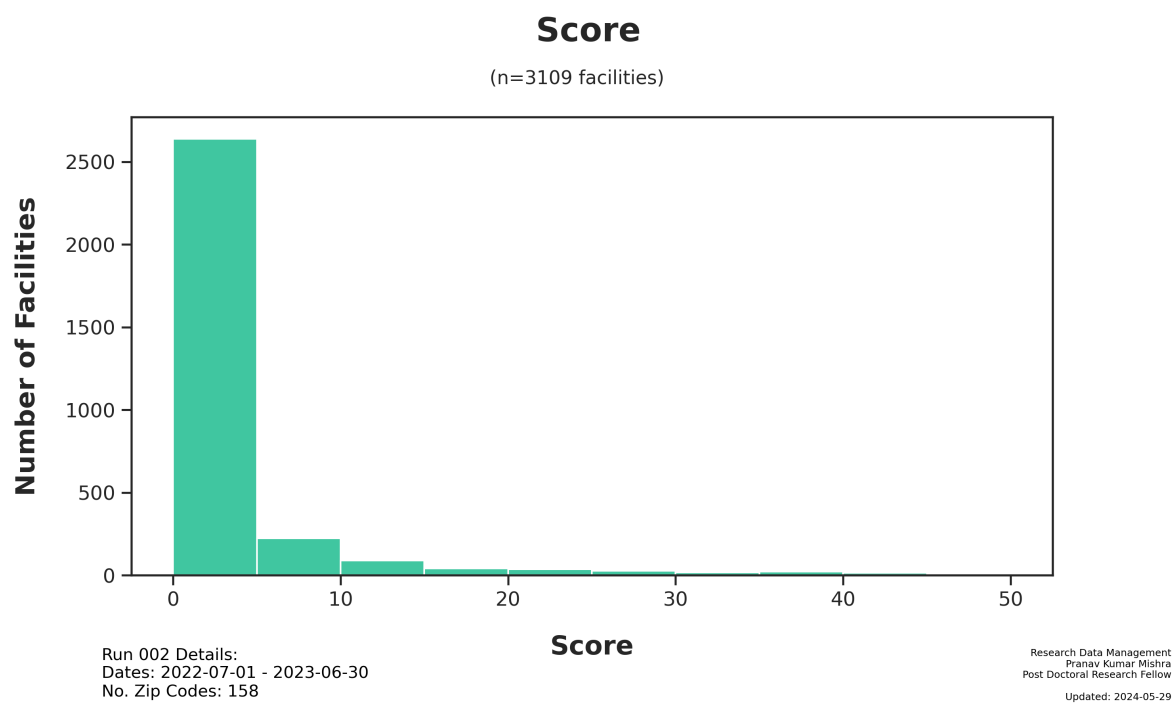
$3 \times IQR$ Extremes for Score

- Num cases before outlier filter: 6,264
- IQR: 11.57 (11.84 - 0.27)
- Outliers: Score < -34.44 OR Score > 46.55

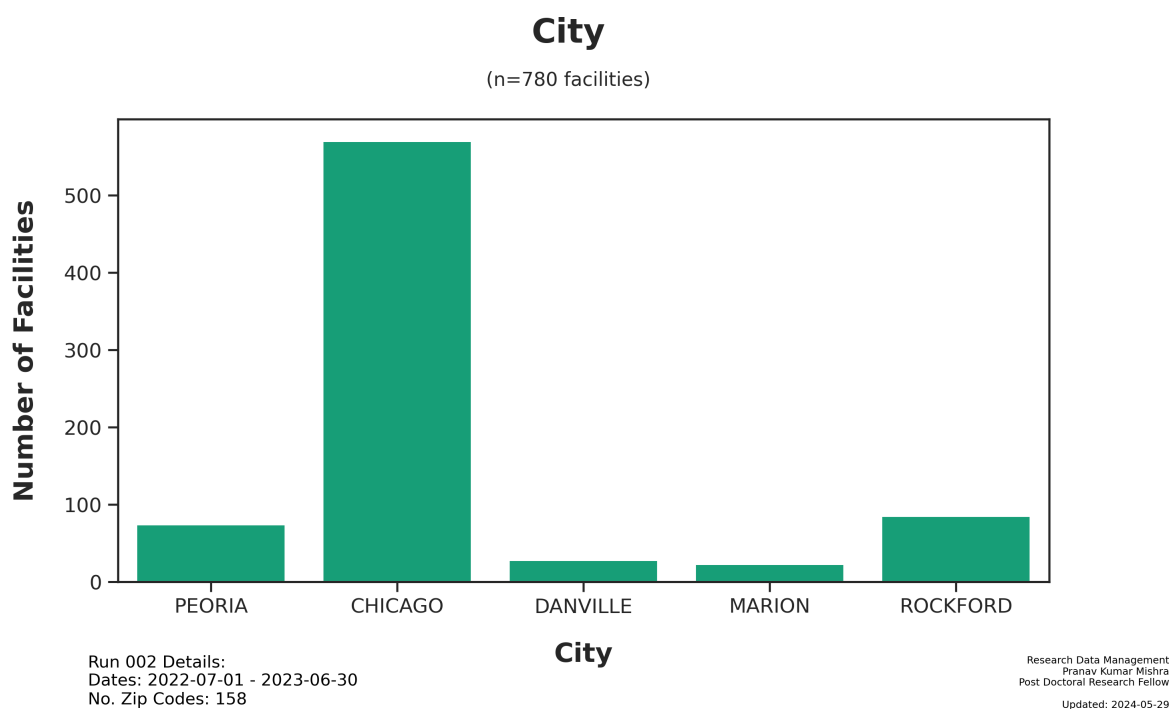
Cases after the outlier filter: 5,496 (Removed 768)

Generate Figures and Tables

Run 002 Details: Dates: 2022-07-01 - 2023-06-30 No. Zip Codes: 158



'Figure saved to /home/pranav/work/pranavmishra90/courses/by-mishra/research-data-management



Save the run dataset