



ANALYSIS REPORT

Library visit in Chicago

JULY 4, 2024

SUBROTO KARMOKAR(23361280)
Methods of Advanced Data Engineering
Department of Artificial Intelligence

1. Introduction

The objective of this project is to integrate and analyze visitor data from Chicago's libraries and local weather data to assess how conducive the city's environment is for library visitors. By examining the visitor counts from the Chicago Public Library dataset for the year 2018 and correlating this with the city's weather patterns, we aim to determine if Chicago's infrastructure and climate support a favorable environment for library patrons. This analysis will benefit library enthusiasts and urban planners by providing insights into the optimal conditions for library visits and areas for infrastructural improvements.

2. Used Data and Transformation

The analysis utilizes two primary datasets:

- **Library Visitor Data:** This dataset is sourced from the [Data.gov](#) portal, detailing the monthly visitor counts for various branches of the Chicago Public Library in 2018. The data is aggregated to provide total monthly visitor counts across all branches. The library data was transformed to aggregate monthly visitor counts.
- **Weather Data:** This dataset is sourced from [Meteostat](#) and includes hourly weather observations for Chicago city for the year 2018. Key variables include **average temperature** (tavg), **snowfall** (snow), **precipitation** (prcp), and **wind speed** (wspd). The weather data was processed to calculate monthly averages for tavg, snow, prcp, and wspd.

	id	month	visitors
	Filter	Filter	Filter
1	1	Jan	628850
2	2	Feb	588219
3	3	Mar	792718
4	4	Apr	715800
5	5	May	681893
6	6	Jun	719255
7	7	Jul	752548
8	8	Aug	794686
9	9	Sep	700404
10	10	Oct	804583
11	11	Nov	709926
12	12	Dec	572282

	id	month	tavg	snow	prcp	wspd
	Filter	Filter	Filter	Filter	Filter	Filter
1	1	Jan	-8.28	69.67	0.04	18.21
2	2	Feb	-5.22	71.1	0.19	16.86
3	3	Mar	-4.45	60.35	0.05	17.23
4	4	Apr	-3.41	54.96	0.06	16.23
5	5	May	11.43	63.05	0.13	14.2
6	6	Jun	16.51	71.22	0.28	14.48
7	7	Jul	17.1	63.82	0.07	13.4
8	8	Aug	18.36	68.78	0.11	13.12
9	9	Sep	14.37	67.69	0.12	14.43
10	10	Oct	5.66	67.43	0.23	15.31
11	11	Nov	-2.81	68.71	0.09	16.26
12	12	Dec	-3.07	75.08	0.1	14.01

Figure 1: Pipeline output of Library visitor and Weather.

Both transformed datasets were stored in a SQLite database for ease of querying and analysis. The database schema includes separate tables for library visitors and weather data.

2.1. License

The Chicago Library visitor's dataset is covered under an open data license, enabling reuse and redistribution for different uses. Meteostat provides weather data licensed under non-commercial terms (CC BY-NC 4.0), allowing for sharing and non-commercial use with proper credit.

3. Analysis

Methodology: This is following steps for the analysis-

- **Data Preprocessing:** The data processing involved summing all visitor counts from different locations of Chicago City's libraries to obtain total monthly visitors. Additionally, only weather data from 2018 was filtered, and any unnecessary information was removed.
- **Data Integration:** Both datasets were combined on a monthly basis.

- **Visualization:** Graphical representations were created to visualize trends and correlations.
- **Seasonal Analysis:** Visitor data was categorized by season to analyze seasonal variations.

4. Results

4.1. Monthly Trends in Library Visits and Weather Variables

- **Visitor Counts:**



Figure 2: Library visitors line graph.

The graph shows the Chicago city's monthly library visitors for a year, peaking in October at above 800,000 and hitting a low in December with about 550,000 visitors. The visitor count fluctuates significantly, generally ranging between 500,000 and 800,000.

- **Weather Variables:**

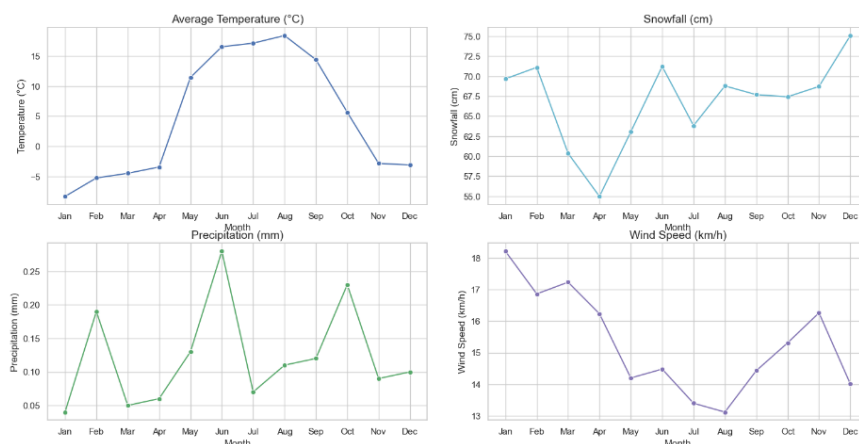


Figure 3: Average Temperature, Precipitation, Snowfall and Wind speed line graph.

The four graphs illustrate monthly weather data, showing clear seasonal trends. **Average temperature** rises from below -5°C in January to around 20°C in July, then falls back to -2°C by December. **Snowfall** peaks in December at about 75 cm. **Precipitation** is highest in June at 0.27 mm, with lower amounts in other months. **Wind speed** is highest in January at around 18 km/h, and lowest in August at about 14 km/h.

4.2. Correlation Analysis

- **Temperature and Visitors:**

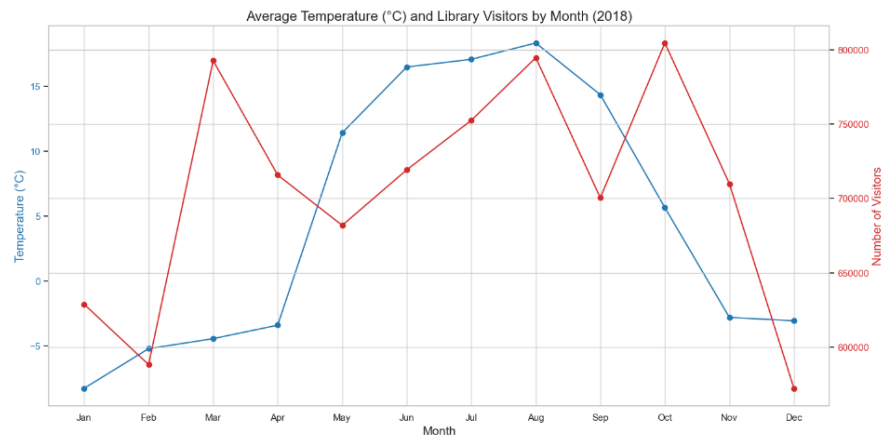


Figure 4: Average Temperature and Library visitors line graph.

The graph displays the relationship between average monthly temperature (°C) and the number of library visitors in 2018. Temperature (blue line) rises from below -5°C in January to a peak around 20°C in July, then falls back to -2°C by December. The number of visitors (red line) peaks in March, August and October at around 800,000, with a significant drop in December to about 600,000. The trends suggest that library visits are highest in moderate to cool months, while extreme temperatures see fewer visitors.

- **Wind Speed and Visitors:**

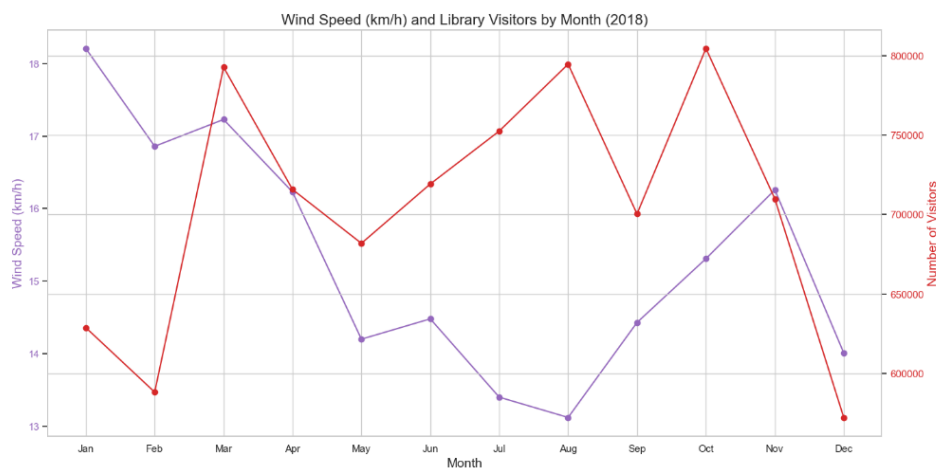


Figure 5: Wind speed and Library visitors line graph.

The graph shows the relationship between wind speed (in km/h) and the number of library visitors per month for the year 2018. Wind speed is plotted on the left y-axis in purple, while the number of visitors is plotted on the right y-axis in red. The data indicates a fluctuation in both wind speed and visitor numbers throughout the year, with wind speed peaking in January and visitor numbers peaking in March, August and October. A general inverse correlation between wind speed and the number of visitors can be observed, especially in the summer months.

4.3. Seasonal Analysis

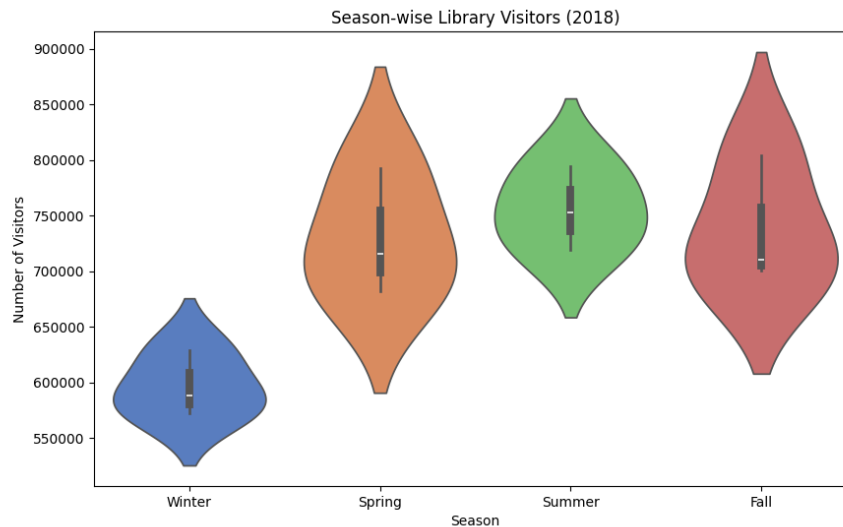


Figure 6: Season-wise visitor distribution violin graph.

The violin plot illustrates the distribution of library visitors across different seasons in 2018. The number of visitors is plotted on the y-axis, with the seasons (Winter, Spring, Summer, Fall) on the x-axis. The plot shows that visitor numbers are lowest in Winter and highest in Fall, with Spring and Summer having similar, moderate distributions. Each season's distribution displays the density of visitors, with the wider sections indicating higher frequencies of visitor counts within that range.

5. Conclusions

This analysis explored the relationship between library visitor counts and weather conditions in Chicago for the year 2018. The key findings are as follows:

- **Visitor Trends:** Library visits exhibit seasonal patterns, with variations likely influenced by weather conditions.
- **Temperature Influence:** There is a noticeable correlation between average temperatures and library visits, with higher visits observed during moderate temperatures.
- **Wind Speed Influence:** Wind speed appears to have a less direct, but still notable, impact on visitor counts.

6. Limitation

While the analysis provided valuable insights, there are limitations. The data is restricted to one year, and additional years could provide a more comprehensive understanding. Furthermore, other factors like special library events or socio-economic variables were not considered. Future studies should incorporate these elements for a more holistic analysis.