Safe at Home? Exploring the Relationship Between Housing Environment and Children's Lead Exposure

Team members: Nicole Sosa and Seoeun Hong

Topic and why this is important:

Our project aims to investigate the correlation between children's elevated blood lead levels and their housing environment, taking into account the potential economic factor of whether children with lower income are more likely to have elevated blood lead levels when living in certain housing conditions. By analyzing data on blood lead levels and housing conditions, we hope to provide insights into the impact of housing conditions on children's health and development and raise awareness of the importance of safe and healthy housing for all children. Exposure to lead from sources such as old paint, water pipelines, and lead dust can have severe health consequences, including delayed growth and development, brain and nervous system damage, and other serious health effects. By considering the potential economic factor, we hope to better understand the intersection of housing conditions and socioeconomic status and identify whether children from lower-income households face a greater risk of elevated blood lead levels due to housing conditions.

Goodness

1. Since all the datasets were obtained from government websites - DATA.NY.GOV - and provided by the New York state department of health and New York state department of labour, we can ensure the trustworthiness of both our dataset and results.
2. Extensive Profiling of data to identify any anomalies, incomplete data, and any outliers. We used summary statistics and examined the unique values of columns to identify any potential issues in the data.
3. After profiling, we try to correct these errors and inconsistencies of data by removing rows with null values, removing outliers, and only extracting data that is necessary.
4. Moreover, we used these papers as reference (Policies to Reduce Lead Exposure: Lessons from Buffalo and Rochester and Relation Between Housing Age, Housing Value, and Childhood Blood Lead Levels in Children in Jefferson County, Ky), which provided valuable insights into the housing conditions that influence children's blood lead levels. We used this information to identify the necessary fields and data needed for a more accurate analysis.

All Data Sources

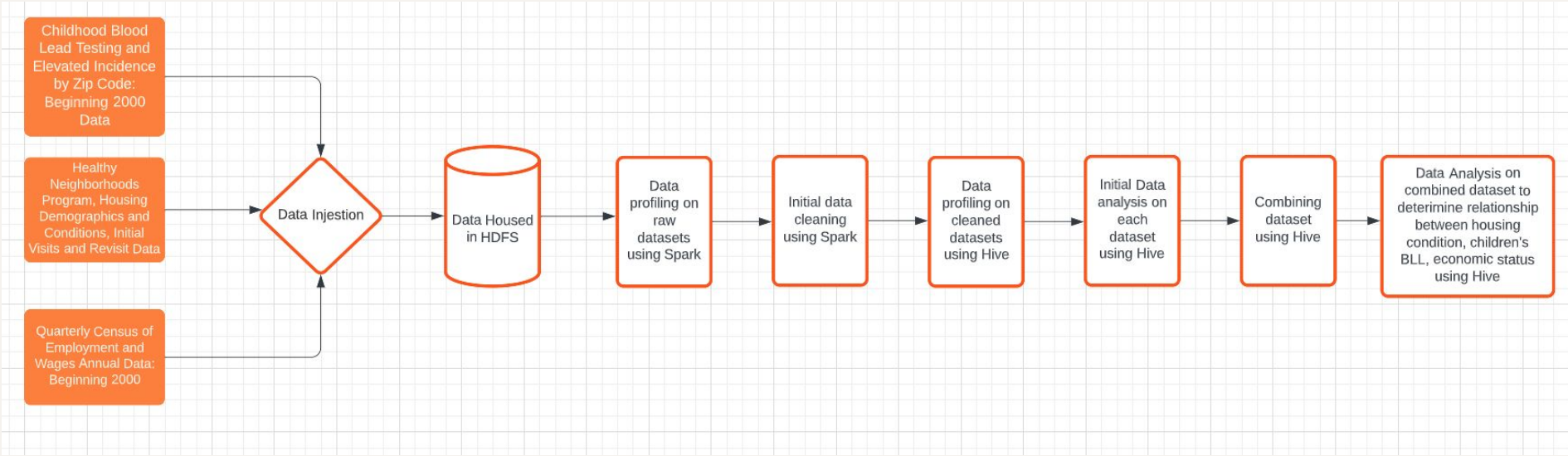
Name: Childhood Blood Lead Testing and Elevated Incidence by Zip Code: Beginning 2000
Description: This dataset provides information on lead testing and exposure among children in New York State (excluding New York City). It includes the number and rate of children tested for lead, and the number of children identified with elevated concentrations of lead in their blood for the first time during a specified time period.
Size of data: 32.4 K rows
Link to data:
https://health.data.ny.gov/Health/Childhood-Blood-Lead-Testing-and-Elevated-Incidenc/d54z-enu8

Name: Healthy Neighborhoods Program, Housing Demographics and Conditions, Initial Visits and Revisits
Description: The NYS Healthy Neighborhoods Program (HNP) aims to decrease housing-related illnesses and injuries through its healthy homes program. The dataset contains details about building features, demographic information of the primary respondent, and whether or not 34 specific housing conditions were present during each county program's visit to the dwellings.
Size of data: 12 K
Link to data:
https://health.data.ny.gov/Health/Healthy-Neighborhoods-Program-Housing-Demographics/jbwf-vnai

Continuation of Data Sources …

Name: Quarterly Census of Employment and Wages Annual Data: Beginning 2000
Description: This dataset provides information on employment and wage data from most nonfarm employers in New York State, covering private-sector and government employees insured under the state's unemployment insurance law, while excluding certain worker categories such as agricultural workers.
Size of data: 59K rows
Link to data:
https://data.ny.gov/Economic-Development/Quarterly-Census-of-Employment-and-Wages-Annual-Da/shc7-xcbw

Design Diagram



Code Challenge

During the analysis of a combined dataset, it was necessary to identify which county had the highest or lowest values for each column. Additionally, we wanted to compare the counties with the highest and lowest values in one column to those in the other columns.

To address this issue, we used window functions and the RANK() function to create a table with a unique rank assigned to each county based on their values for each column. This table allowed us to easily identify each county's position relative to other counties in each column. Furthermore, this approach eliminated the need for multiple queries and manual comparison of query results, making the process more efficient and accurate.

Code Challenge cont …

```
0: jdbc:hive2://localhost:10000> SELECT county,
. . . . . .>     less_than_5,
. . . . . .>     ROW_NUMBER() OVER (ORDER BY less_than_5 DESC) AS less_than_5_rank,
. . . . . .>     5_to_10,
. . . . . .>     ROW_NUMBER() OVER (ORDER BY 5_to_10 DESC) AS 5_to_10_rank,
. . . . . .>     num_dwellings,
. . . . . .>     ROW_NUMBER() OVER (ORDER BY num_dwellings DESC) AS num_dwellings_rank,
. . . . . .>     avg_econ_status,
. . . . . .>     ROW_NUMBER() OVER (ORDER BY avg_econ_status DESC) AS avg_econ_status_rank
. . . . . .> FROM county_econ_percentage;
```

county	less_than_5	less_than_5_rank	5_to_10	5_to_10_rank	num_dwellings	num_dwellings_rank	avg_econ_status	avg_econ_status_rank
WESTCHESTER	93.86	6	6.14	13	12113	8	67695.01	1
ROCKLAND	95.64	3	4.36	16	14950	7	51318.27	2
ALBANY	89.13	13	10.87	6	7922	11	48974.82	3
MONROE	91.62	10	8.38	9	23807	5	47075.37	4
ERIE	87.84	16	12.16	3	48812	1	46174.89	5
ONONDAGA	88.35	15	11.65	4	31946	3	45063.94	6
RENSSELAER	88.57	14	11.43	5	7663	12	43082.39	7
SCHENECTADY	90.71	11	9.29	8	10491	10	42755.79	8
ORANGE	94.02	5	5.98	14	25050	4	42516.58	9
TOMPKINS	96.34	2	3.66	17	10833	9	38986.29	10
NIAGARA	92.23	8	7.77	11	34329	2	38647.69	11
COLUMBIA	87.35	17	12.65	2	1231	18	38441.44	12
ONEIDA	83.42	18	16.58	1	7456	13	38392.48	13
BROOME	92.73	7	7.27	12	4523	14	37966.6	14
CLINTON	98.89	1	1.11	18	16382	6	35436.77	15
CAYUGA	91.74	9	8.26	10	3986	15	35148.21	16
CORTLAND	89.69	12	10.31	7	3559	16	34207.18	17
TIOGA	95.6	4	4.4	15	2371	17	32003.25	18

```
18 rows selected (13.891 seconds)
```

Insights

- Our observations revealed that the top 5 counties with the highest number of children testing between 5 to 10 for blood lead levels, as well as the lowest 5 counties, were generally the same as those counties with the highest number of housing complaints related to paint issues, chemical smell, and the age of the house.

- We also observed that the top 5 counties with the highest number of children testing below 5 for blood lead levels were generally the same counties with the least number of housing complaints.

- We were unable to find any significant relationship between housing conditions or children's blood lead levels with **economic status**. This may indicate that there is no relationship, or it may suggest that our dataset is flawed.

Acknowledgements

We would like to sincerely thank DATA.NY.GOV for providing us with all three of our datasets for this project.

Also thank you Professor Malavet for helping us with your guidance on our project.

References

Kim, D. Y., Staley, F., Curtis, G., & Buchanan, S. (2002). Relation between housing age, housing value, and childhood blood lead levels in children in Jefferson County, Ky. American journal of public health, 92(5), 769–772. https://doi.org/10.2105/ajph.92.5.769

Magavern S. (2018). Policies to Reduce Lead Exposure: Lessons from Buffalo and Rochester. International journal of environmental research and public health, 15(10), 2197. https://doi.org/10.3390/ijerph15102197

New York State Department of Health. (2022). Childhood Blood Lead Testing and Elevated Incidence by Zip Code: Beginning 2000 [Data set]. https://health.data.ny.gov/Health/Childhood-Blood-Lead-Testing-and-Elevated-Incidenc/d54z-enu8

New York State Department of Health. (2019). Healthy Neighborhoods Program, Housing Demographics and Conditions, Initial Visits and Revisits: Beginning 2006 - 2009 Funding Cycle [Data set]. https://health.data.ny.gov/Health/Healthy-Neighborhoods-Program-Housing-Demographics/jbwf-vnai

New York State Department of Labor. (2022). Quarterly Census of Employment and Wages Annual Data: Beginning 2000 [Data set]. State of New York. https://data.ny.gov/Economic-Development/Quarterly-Census-of-Employment-and-Wages-Annual-Da/shc7-xcbw