# Resident Impact Data Project Technical Report – Allentown, PA

February 2025

## Table of Contents

# Background

Allentown, Pennsylvania, is grappling with an eviction crisis that affects the entire Lehigh Valley region. The city faces an alarming rate of 18 evictions per day, ranking third highest in the state[1]. This crisis has far-reaching consequences, impacting various aspects of community life. The high rate of evictions leads to increased transience, which affects school attendance and contributes to truancy issues. Moreover, evicted families often find themselves in substandard living conditions, facing overcrowding or homelessness[2]. The city administration is working to identify the root causes and correlations of evictions to implement effective policies and programs to address the issue. To combat this problem, Allentown officials are working with internal and external stakeholders, such as the county court system, to gather eviction data and collaborate with advocacy nonprofits like NorthPenn Legal to develop targeted interventions and support systems for affected residents.

# Goals and Challenges

## Problem Statement and Data-Driven Questions

The RIDP team began the analysis by developing a clear problem statement to guide the project's direction, scope, and analysis. The problem statement read:

"Lehigh Valley faces an eviction crisis, with 18 renters facing eviction every day; the county has the third highest rate within the state. Some of the highest zip codes for eviction are within Allentown. Allentown does not know what leads to or is correlated with evictions and thus cannot necessarily direct the right policy or programmatic interventions to address them."

## Potential Challenges

At the beginning of this project, we identified three main challenges. First, the city does not collect data on evictions or eviction filings. This data is gathered, processed, and managed by the state court system and published at the county level (Lehigh County). Additionally, the city lacks the legal and jurisdictional authority to intervene to prevent evictions when they occur. Finally, city officials lack familiarity with the legal intricacies of the eviction process. Although there is a willingness to collaborate on this issue, the city officials cannot do it alone; they must

---

[1] Housing Alliance of Pennsylvania. (2022). Revealing Opportunities and Challenges: An Analysis of Eviction Filings in Pennsylvania. https://housingalliancepa.org/wp-content/uploads/Statewide-Eviction-Analysis_Housing-Alliance-of-PA_Feb-2022.pdf

[2] Ashton, A. R. (2021, February 25). The American eviction crisis explained. The Appeal. https://theappeal.org/the-lab/explainers/the-american-eviction-crisis-explained/

engage several internal and external stakeholders to understand, analyze, and collaboratively develop solutions to tackle this issue.

# Data Description

## Datasets Description

For this project, we utilized two different datasets. The primary dataset consists of eviction filing records from Lehigh County, covering June 2022 to July 2023. The data from the Housing Alliance of Pennsylvania is organized in a long format, with each row representing a judgment component linked to an eviction case. Data aggregation is essential for case-level analysis since a single eviction case may involve multiple judgment components and several defendants. Key variables include eviction case identifiers (`docket_number`), details about the defendant and plaintiff, case disposition outcomes, financial judgments, and filing dates. Geographic information, such as the defendant's city, zip code of residence, and judicial assignments for each case, are also included.

The second dataset is a map layer of voting precincts in Lehigh County from the county's ArcGIS REST repository. This dataset contains multiple polygons representing all the voting precincts and information about the magisterial district to which each precinct belongs. By aggregating these precincts, we will create a polygon that encompasses the various magisterial districts within Lehigh County. Additionally, using data from Pennsylvania's court system, we can link the court districts to the judges assigned to them between 2022 and 2023.

## Data Quality and Completeness Assessment

### Missing Data

Some variables show missing data, particularly `judgment_component_type` (10.67% missing), `net_judgement` (10.67% missing), and `attorney_name` (83.36% missing). Variables like `defendant_state` and `plaintiff_state` also have minor missingness (<0.03%). Not understanding why missing values in important variables exist, especially `case_disposition`, could bias the results of any analysis done with this dataset and hide correlations between outcomes and other demographic or geographical variables.

### Descriptive Statistics

The dataset contains 38,301 observations across 27 variables. It is structured in a long format, each row representing a judgment component. Appendix 1 contains a detailed description of each variable and descriptive statistics, but here are some relevant examples:

- The `docket_number` variable is unique per eviction case.
- The `defendant_city` field contains multiple spellings of the same locations. For instance, the dataset contains 12 different ways to spell Allentown, such as "Alllentown" or "Alletnown".
- The `plaintiff_name` variable contains 1,665 unique observations. However, after an initial review, we identified several similar plaintiff names with different spellings due to errors in how this variable was collected. For instance, in some cases, the `plaintiff_name` was recorded as *1013 West Linden Street LLC,* while in others, it was noted as *1013 W Linden Street LLC*. Because of these discrepancies in how the same plaintiff's name was recorded in the original database, this variable must be reviewed and cleaned to draw meaningful conclusions.
- `filing_date`, `case_disposition_date`, and `order_for_possession_issued_date` cover July 2022 to August 2023.
- The average monthly rent is approximately $1,259.91 (SD: $553.28), while claim amounts average $2,700.04 (SD: $2,458.21)—additionally, the `judgment_component_amount` variable exhibits significant variance, indicating diverse financial impacts of judgments.

## Duplications and Unique Identifiers

Eviction cases can have multiple judgment components, leading to multiple entries per case. We must apply a de-duplication process to extract unique eviction cases by `docket number`. Cleaning efforts must also be conducted for `defendant_city` and `plaintiff_name`, resolving inconsistencies in spelling variations of locations and plaintiffs. The potential unique identifiers are:

- `docket_number`: Serves as a primary identifier for eviction cases.
- `defendant_name`: While helpful, multiple appearances of the same name across different cases limit its uniqueness.
- `plaintiff_name`: Identifies landlords or property management companies but is subject to variations in spelling and formatting.

## Data Caveats

- Long Format: Each eviction case may have multiple judgment components, necessitating aggregation for case-level analysis.
- Inconsistent Spellings: Before analysis, variations in spelling for `defendant_city` (e.g., "Alllentown," "Alletnown") and `plaintiff_name` need to be resolved.
- Missing Case Disposition Data: Some cases lack disposition details, impacting assessments of eviction outcomes.
- High Missingness in Attorney Data: The `attorney_name` field is absent in over 80% of records, limiting its analytical usefulness.

- Extreme Values in Financial Variables: Monthly rent and claim amounts have significant outliers (e.g., the maximum rent is $18,500), requiring further validation.

# Analytical methodology

## Software and Infrastructure

The analysis used R version 4.3.2, employing libraries such as `tidyverse` for data cleaning, `skimr` for summary statistics, and `ggplot2` for visualization. We performed geospatial mapping in ArcGIS Online to visualize how magisterial districts map to Allentown. We managed geographic data processing using `GeoJSON` and the `SF` package, allowing the integration of eviction records with jurisdictional boundaries.

## Data Cleaning

The primary variable that needed to be cleaned and standardized before analysis was the `plaintiff_name`. As mentioned earlier, during our review of this variable, we discovered duplicate plaintiff names corresponding to the same entity, although their spellings varied. To address these differences and compile a comprehensive list of plaintiffs, we developed a step-by-step methodology that first standardized the format of the names and then sought to identify similar names to determine if they referred to the same entity. Here are the steps we followed to achieve this:

1. **Data Deduplication**
    - We removed duplicate records by grouping them by docket number and plaintiff name.
    - Before cleaning, there were 1,665 unique plaintiffs.
2. **Text Standardization**
    - We converted all names to lowercase and removed special characters (e.g., ".", ",", "/").
    - We standardized common abbreviations:
        i. `mgmt → management,` `street → st.`
        ii. `west → w,` `east → e,` `north → n,` `south → s`
    - This standardization reduced the unique plaintiff count to 1,567.
3. **Levenshtein Distance Matching**
    - We applied the **Levenshtein (LV) distance** metric to detect similar plaintiff names by calculating the minimum number of single-character edits needed to transform one name into another.

    - To implement this, we used the `stringdist` package[3] in R to create an LV matrix with all the pairs of similar names across the entire list of plaintiffs.
    - A similarity threshold of 3 was set to identify potential matches.
4. **Manual Review and ChatGPT-Aided Matching**
    - We manually reviewed matches suggested by the Levenshtein method.
    - We resolved additional inconsistencies using ChatGPT to suggest similar entity names.
    - We manually corrected standardized names based on specific rules:
        i. Adding missing street suffixes (e.g., `"1039 mechanic llc" → "1039 mechanic st llc"`)
        ii. Resolving missing or incorrect cardinal directions (e.g., `"440 tilghman st llc" → "440 w tilghman st llc"`)
        iii. Matching plaintiffs with extra or missing spaces and abbreviations (e.g., `"520 hamilton oplp" → "520 hamilton op lp"`)
        iv. Consolidating alternative spellings or common name variations (e.g., `"capece dave s", "capece david a", "capece david s" → "capece david"`)
    - We did not match some names due to insufficient evidence.
5. **Final Unique Plaintiffs and Serial Evictors Identification**
    - After manual corrections, the final number of unique plaintiffs was **1,343**.

The final cleaned list more accurately represents the entities that filed the most eviction cases in Lehigh County. Appendix 2 provides a more detailed description of this methodology.

## Data Merging with the Magisterial Districts Shapefile

We then decided to merge the eviction court filings dataset with the magisterial district area where the eviction occurred, providing a clearer understanding of where evictions were happening in Allentown. We had to use the magisterial district as an approximation because the exact address of the property for which the defendant was being evicted is not available in the original dataset. The only variable we could use for a geographical assessment of evictions was `judge_assigned`, as judges are assigned to cases in the districts over which they have jurisdiction.

To merge this data, we started by importing the voting precinct shapefile from Lehigh County's ArcGIS repository, which contained magistrate district IDs in the format "XX-X-XX." Due to an issue with the GeoJSON parser interpreting these IDs as dates, we manually corrected the format by converting them to a character string, removing the first two digits, and reformatting

[3] A more detailed description of the LV distance method can be found in Appendix 2 and in the R package documentation, available here.

them to restore the original structure. This ensured that the IDs were preserved accurately for further processing.

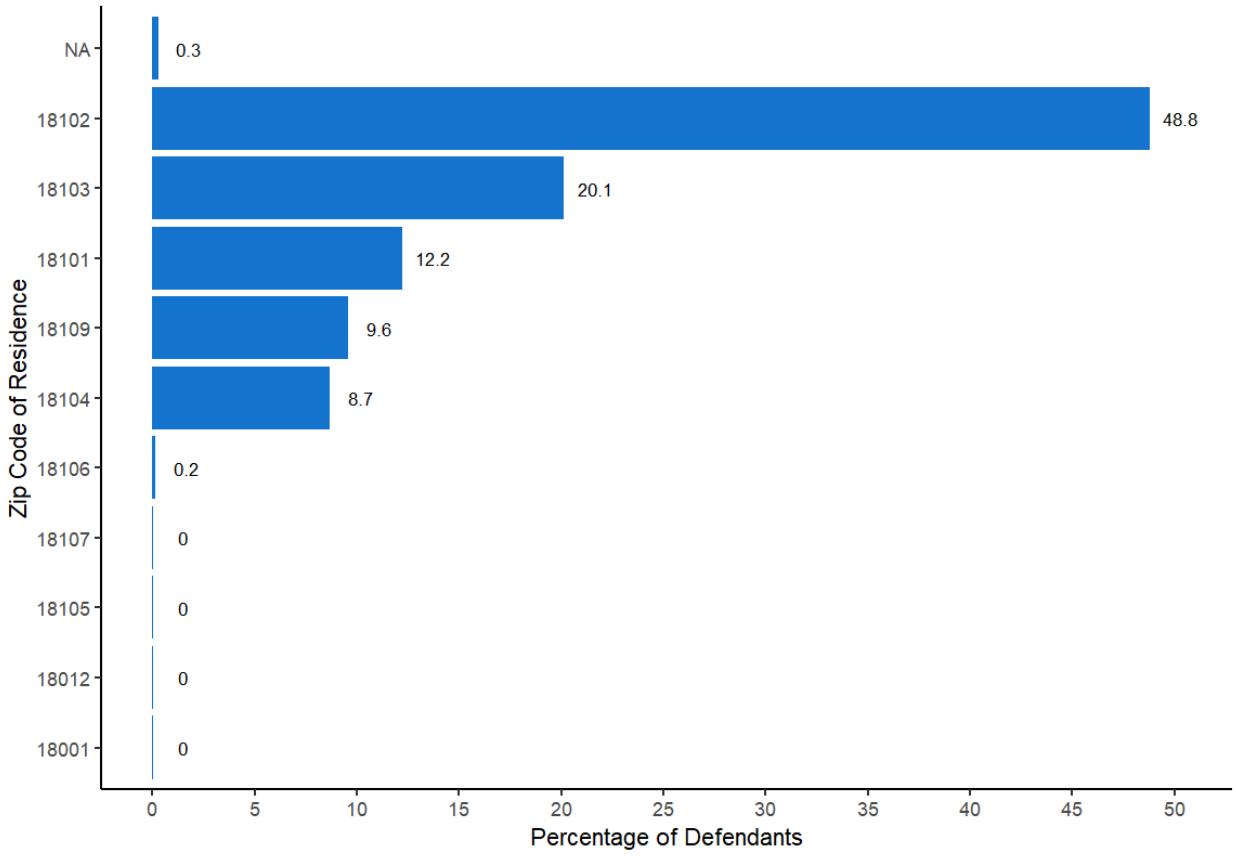
Next, we aggregated the voting precinct data to the magistrate district level by merging polygons and refining the geometries to eliminate minor boundaries. We then assigned the judge's name presiding over each magistrate district, including exceptions for newly created or recently modified districts. We combined this spatial data with a table of eviction filing statistics by judge, renaming the columns for clarity. We utilized the 2024 voting precinct shapefile, including the newly established “31-1-09” magisterial district. This district was absent in the 2022-2023 file and had no eviction cases. Finally, we saved the processed data as a shapefile, creating a spatial dataset that links magistrate districts, judges, and eviction filing data for analysis or visualization. More details about this merging can be found in Appendix 1.

## Exploratory Data Analysis

After cleaning and standardizing the data, we obtained all the unique cases in Allentown using the defendant’s reported city. The analysis identified **7,136 unique defendants**[4] in Allentown involved in **5,328 eviction cases**. Nearly half of these defendants (48.) lived in **zip code 18102**, indicating that evictions were heavily concentrated in this area, as illustrated in Figure 1.

[4] We considered each defendant and case as a “unique case defendant”. It could be the case that the same name appears in multiple cases, however each appearance is considered a unique defendant.

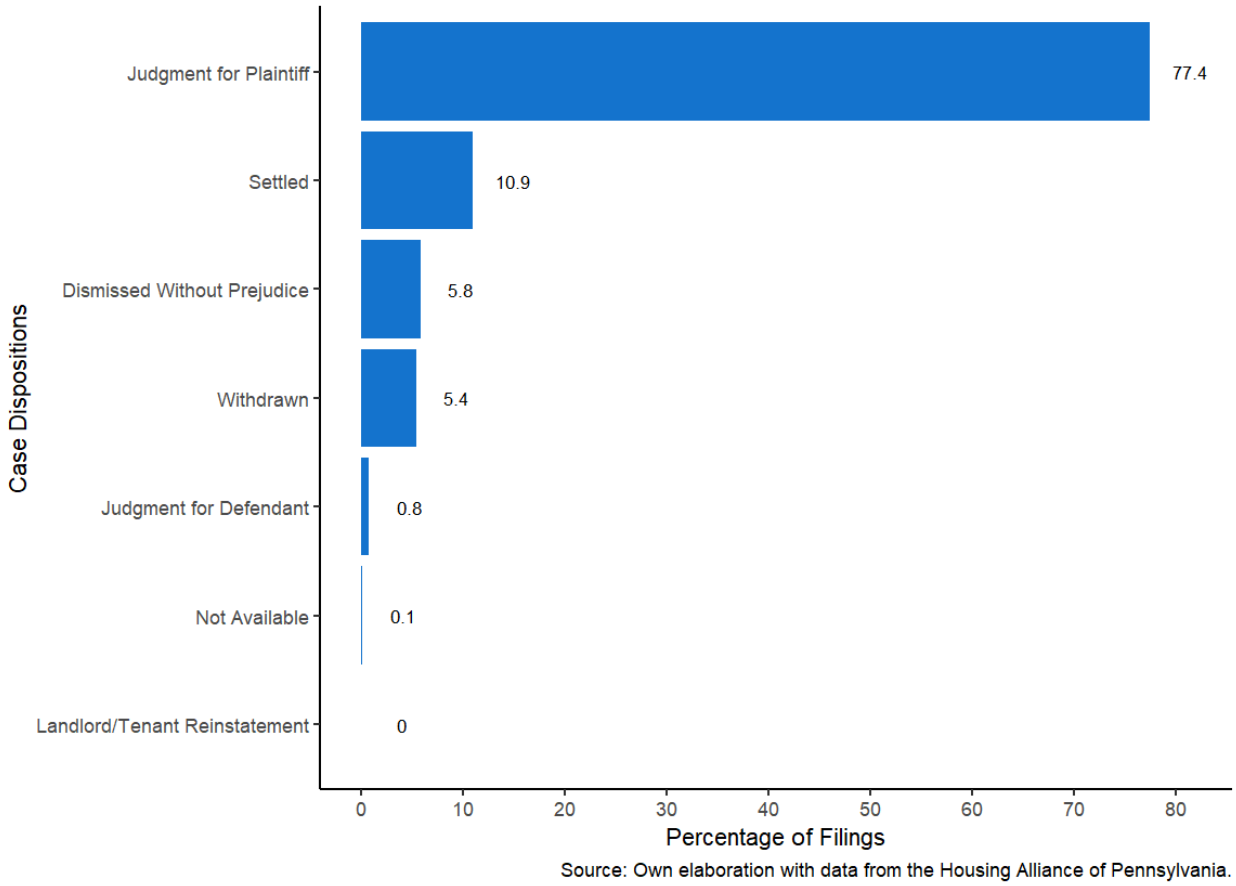
Figure 1. Percentage of Defendants by Zip Code of Residence




Source: Own elaboration with data from the Housing Alliance of Pennsylvania.

Additionally, the analysis revealed that eviction cases overwhelmingly favored landlords. More than 77% of judgments were ruled in favor of the plaintiff, while only 10.9% were settled, and just 0.77% favored the defendant, as shown in Figure 2. Financially, the average monthly rent in eviction cases was $1,216.58, with a median of $1,180. Furthermore, 75% of all eviction filings involved properties with rents under $1,400, although some extreme cases reached as high as $11,000 per month. The claims filed by landlords mirrored this pattern, with an average claim amount of $2,299.51 and a median of $1,715.60. However, some claims were considerably higher, resulting in a skewed distribution, as presented in Appendix 1.

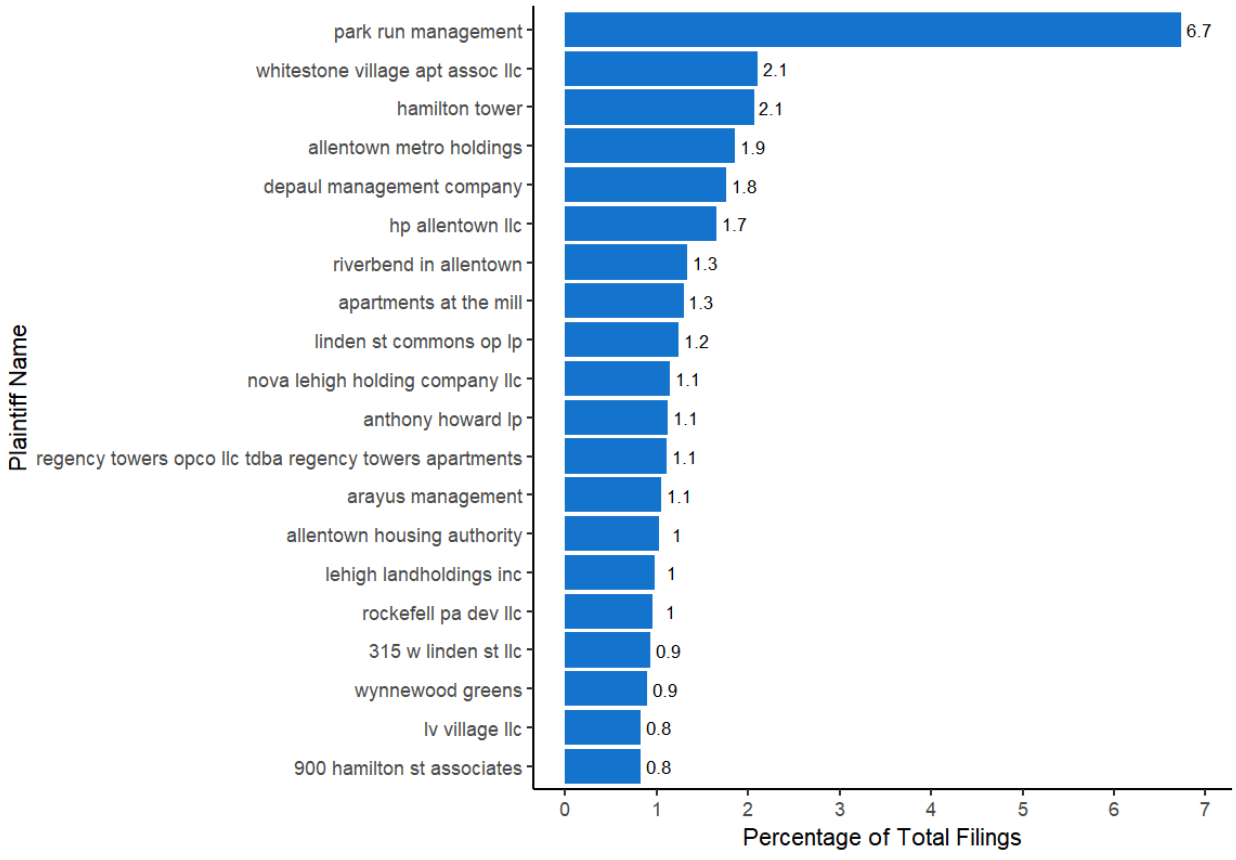
Figure 2. Case Dispositions in Filings in Allentown




Source: Own elaboration with data from the Housing Alliance of Pennsylvania.

The analysis also examined whether certain landlords, or "serial evictors," were responsible for a disproportionate number of cases. Initially, “1521 Julian LLC” appeared to be the most frequent filer, with 12 eviction cases. However, we found several inconsistencies in how plaintiff names were recorded and implemented a data-cleaning process described above. This effort **reduced the number of unique plaintiffs from 1,665 to 1,343**. It revealed that **“Park Run Management” was responsible for 359 eviction filings**, making it the most active evictor in Allentown, accounting for 6.7% of cases. Figure 3 presents the most common evictors in Allentown in 2022 and 2023.

Figure 3. Plaintiffs with Most Eviction Filings in Allentown




Source: Own elaboration with data from the Housing Alliance of Pennsylvania.

Another significant finding was the role of a small number of judges in overseeing eviction cases. Six judges handled over 93% of all eviction filings in Lehigh County, with Judge David M. Howells Jr. presiding over 21.36% of cases, followed by Judges Karen C. Devine and Rashid Santiago.

While these findings provide valuable insights into eviction patterns in Allentown, further analysis is necessary to refine the results and address data limitations. Relying on the defendant's city to determine the eviction location may also lead to undercounts, as some tenants could have already moved out of Allentown before their cases were filed. Additionally, more work is needed to identify and track serial evictors by mapping LLCs to their owners or management companies.

Future research should explore the financial burden placed on tenants, including court fees and the amount of money needed to avoid eviction. Examining variations in claim amounts by zip code and judgment component could offer further insights into eviction drivers. Moreover, historical comparisons with previous reports from the Housing Alliance of Pennsylvania could help determine eviction trends in Allentown over time.

## Mapping Evictions by Magisterial District

Finally, we can map the total number of evictions by magisterial district onto a map of Allentown. Figure 4 shows the percentage of eviction filings by magisterial district and judge out of the total eviction filings in Lehigh County. This map reveals that downtown and south Allentown account for most of the eviction filings in the entire county. An interactive version of this map is available here.

# Key Findings

1. **Overview of Eviction Filings in Allentown**
    a. Allentown faces 18 evictions per day, making it the third-highest eviction rate in Pennsylvania.
    b. From 2022 to 2023, there were **7,136 unique defendants involved in 5,328 eviction cases.**
    c. **48% of eviction cases in Lehigh County are concentrated in zip code 18102,** indicating a geographic hotspot for housing instability.
    d. **77% of cases resulted in rulings favoring landlords**, while only 0.77% favored tenants.
2. **Financial Burden on Tenants**
    a. The **average monthly rent in eviction cases is $1,216.58** (median: $1,180).
    b. **75% of evictions involve properties with rents under $1,400**.
    c. The average claim amount by landlords is $2,299.51 (median: $1,715.60).
    d. Some extreme cases had rents as high as $11,000 per month.
3. **Patterns of Eviction Filings**
    a. A small number of landlords were identified during the analysis.
        i. **Park Run Management Co. filed 359 cases, making it the top evictor,** accounting for 6.7% of all cases in the county.
    b. Six judges handled over 93% of eviction filings in Lehigh County.
        i. Judge David M. Howells Jr. presided over 21.36% of cases.
4. **Data Quality Issues**
    a. Before analysis, several data inconsistencies in city names (e.g., 12 variations of "Allentown") should be accounted for.
    b. There are duplicate plaintiff names that could cause a misrepresentation of eviction patterns if not analyzed and standardized.
    c. Some data on judgment types (10.67%) and attorney names (83.36%) are missing, which could limit further analysis.
5. **Geographic Trends**
    a. **Downtown and South Allentown** have the highest eviction rates.

# Final Recommendations

1. **Improve Data Collection and Access**
    a. Continue collaborating with the county court system to improve data quality and obtain timely, case-level eviction data.
    b. Explore the possibility of collecting more data to conduct address-level geocoding to better analyze eviction hotspots.
2. **Strengthen Tenant Protections**
    a. Work with nonprofits (e.g., NorthPenn Legal) to increase and make available resources for tenants facing eviction cases. A multilingual hub with different resources for tenants who might face eviction could increase awareness of tenants' rights and provide tenants with information and tools to face this issue.
3. **Monitor Serial Evictors**
    a. Continue to analyze the serial evictors list to understand their rental practices, which buildings they manage, and how they use evictions as part of their rental practices.
    b. Consider investigating ownership of the LLCs and management companies shown in the list using the local business registry or other county or state databases.
4. **Address Geographic Disparities**
    a. Focus on zip code 18102 and South Allentown, where evictions are most concentrated.
    b. Develop tenant outreach programs in high-eviction districts to provide legal aid and financial counseling.
5. **Conduct Further Analysis**
    a. Study the long-term impact of eviction filings on families, including school truancy and homelessness.
    b. Compare historical eviction trends to assess the effectiveness of past interventions.
    c. Evaluate court fee structures and their impact on eviction case outcomes.

By implementing these recommendations, Allentown can improve eviction prevention strategies, support at-risk tenants, and develop policies based on data-driven insights.

# Resources

- Housing Alliance of Pennsylvania. (2022). Revealing Opportunities and Challenges: An Analysis of Eviction Filings in Pennsylvania. https://housingalliancepa.org/wp-content/uploads/Statewide-Eviction-Analysis_Housing-Alliance-of-PA_Feb-2022.pdf
- Ashton, A. R. (2021, February 25). The American eviction crisis explained. The Appeal. https://theappeal.org/the-lab/explainers/the-american-eviction-crisis-explained/
- Allentown and Lehigh County Evictions by Magisterial District Map
- Voting Precincts in Lehigh County
- RIDP Exploratory Analysis Slides (12/02/2024)
- Allentown RIDP - Housing Questions
- Eviction Prevention Resource Library - Housing Alliance of Pennsylvania

# Appendices

- Appendix 1: RIDP Exploratory Analysis
- Appendix 2: RIDP Serial Evictors Data Cleaning and Analysis