
AUTOMATING PROCESSES IN DATA REPORTING

Alain Lopez

Data Analytics & e-Publishing

 [in/al-the-analyst](https://www.linkedin.com/in/al-the-analyst)

 [www.BlueSageInsights](http://www.BlueSageInsights.com)



This project showcases the use of task automation in Excel and Word to improve data reporting accuracy and turnaround time.

The Case

The client, Career Resources, Inc., publishes a yearly data report with a descriptive analysis of the reentry population in their state as a whole and within ten regions to help identify needs and address gaps in services for individuals returning home from incarceration.

The population is broken down into two subgroups and the data is a day-in-time census with information about each subgroup's demographics, system involvement, and assessment scores (which are used for treatment and program planning in the correctional system).

For a nonprofit organization with a small team like theirs, such a detailed data report comes with challenges.

- Firstly, each data section (one statewide and ten regional) is comprehensive — containing 28 or more data visualizations.
- Additionally, the design and layout need to remain consistent throughout so everything reads as one cohesive report once compiled.
- Lastly, for the report to be timely and useful, it needs to be published by a certain deadline.

Solution Overview

My role in the team is to create the data visualizations and deliver a compiled report as a Word document and as a PDF using the clients layout and design specifications.

I have broken down the project into three phases (Data Preparation, Creating the Visualizations, and Creating the Report) and used different features:

- ✓ Power Query helps streamline the preparation phase.
- ✓ Excel formulas containing named ranges are used to pull data dynamically and then referenced in the visualizations.
- ✓ For the report, Microsoft 365's object pasting feature makes it possible to link Excel visualizations to Word templates.

The sections that follow provide details and visuals on how each phase was streamlined.

Preparing the Data

Two different groups are represented in the report, so the client provides two tables, each in a separate Excel file. I consider the process in this phase as a mini-ETL (extract-transform-load) process.

Duplicates of the tables get brought into an Excel workbook that is just for preparing the data. Some preliminary cleaning is done. The tables are then loaded into Power Query, where the rest of the cleaning and transformation has been automated. At the end, the tables are appended to a dataset containing data from previous years.

Column1	Column2
Region	Bridgeport
Year	6/1/2023
Subgroup	CS
Total Count	392
Age	
18-24	9
25-34	97
35-44	132
45-54	105
55-64	37
65+	12

Column1	Column2
Region	Bridgeport
Year	6/1/2023
Subgroup	EOS
Total Count	86
Age	
18-24	6
25-34	29
35-44	31
45-54	10
55-64	8
65+	2

The two tables for 2023 are brought into Power Query

Subgroup	Region	Age	18-24
CS	Bridgeport	Age	24
EOS	Bridgeport	Age	3
CS	Bridgeport	Age	9
EOS	Bridgeport	Age	6

The automated steps prepare, combine, and add the 2023 data to a dataset containing data from all the previous years.

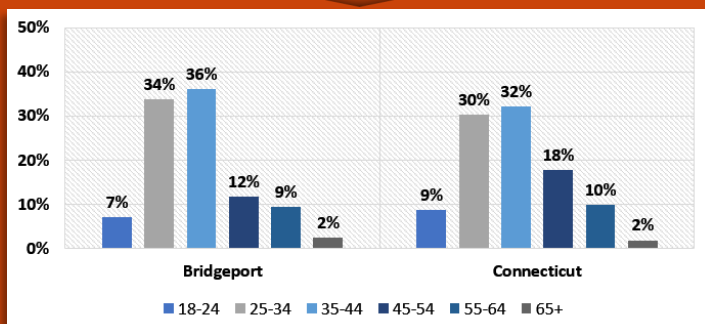
Creating the Data Visualizations

Year	Subgroup	Region	18-24	25-34	35-44	45-54	55-64	65+
2023	EOS	Bridgeport	6	29	31	10	8	2

Prepared dataset loaded from Power Query

Age		
18-24	6	7%
25-34	29	34%
35-44	31	36%
45-54	10	12%
55-64	8	9%
65+	2	2%

Range of filtered data for referencing



The data visualization process happens in a separate Excel workbook. One sheet (Data Import) contains the prepared dataset from Power Query. With the data in tabular form within one large dataset, it is easier to search for specific information.

In a second sheet (Data Analysis), data for different census categories are filtered into cell ranges based on the region and year. The formulas are made dynamic by using the Filter() function and two named cells — "Region" and "censusYear". When the named cells' values are changed, all the cells update. Since there are two subgroups, there is a set of ranges for each.

Each set is referenced in a corresponding data visualization sheet, so the data visualizations update automatically whenever the Data Analysis sheet updates.

Creating the Report

The three main sections of the report are the analysis, the statewide data, and the regional data. The components are prepared in separate Word documents which have been formatted to have the same layout and design.

The first document (the Report template) contains the introductory pages, the descriptive analysis and two blank pages as placeholders for the data visualization sections. In the Statewide and Regional templates, most of the data visualizations have been added as linked objects from the Visualizations Excel file. Updating the templates each year mainly involves selecting and refreshing the objects.

What We Are Learning Through Data

This is the third consecutive year that the State of Reentry Report is providing data to our legislators, community organizations, policy and academic research, government agencies, and other community stakeholders on the barriers faced by those released from incarceration and who are seeking to successfully reenter their communities. As Connecticut State Representative Robyn A. Puffer states,

"By identifying these barriers and developing strategies to overcome them, Connecticut can help returning citizens gain employment, stay crime-free, and contribute positively to their communities. It's great to see such important issues being addressed in a collaborative and data-driven way!"

We do not know all the policy and strategic planning uses where this report's data is being used, but we do have examples:

Connecticut Issues for Children used the State of Reentry (SOR) data in their 2023 work, *Reeling from the Impact: Rethinking Reentry, Wage and Policy Barriers*.¹ Connecticut's Commission on Economic and Industry? A major Connecticut regional nonprofit organization is using SOR data as one tool in their toolbox to access future community needs and their impact goals. Connecticut's SOR data is also being used nationally by academic researchers in their growing work on addressing conditions of reentry.

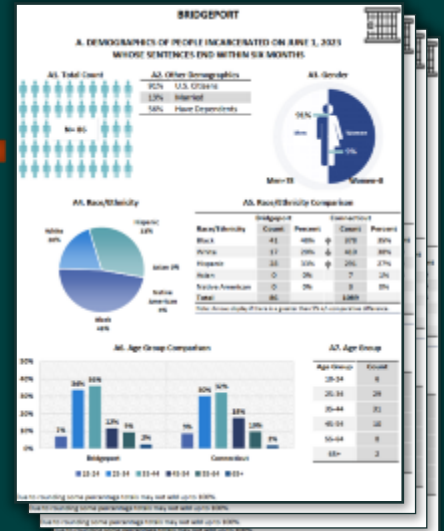
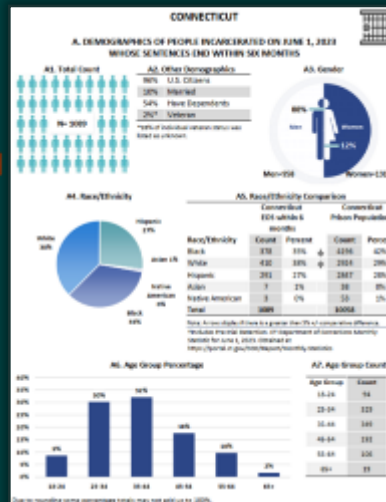
Key barriers we are measuring and finding persistent are high rates of substance abuse, mental health issues, and housing instability among our returning citizens. These factors are combined with low levels of education and paid employment history within our returning citizens numbers. All these conditions and barriers to employment and successful contribution to Connecticut's growth as a State require data-driven state and regional strategies. We see this report as one data-driven tool to help shape those strategies.

THE DATA

To better understand the unique needs of the reentry population in Connecticut and share a light on the magnitude of the challenges they face, the data presented in this report represents a "snapshot of a day-in-the-life" of people under custody of the Connecticut Department of Correction. Two different subgroups are represented in this report: 1) individuals who were incarcerated on June 1, 2023, with an end of sentence date within six months, and 2) individuals under community supervision on June 1, 2023. The

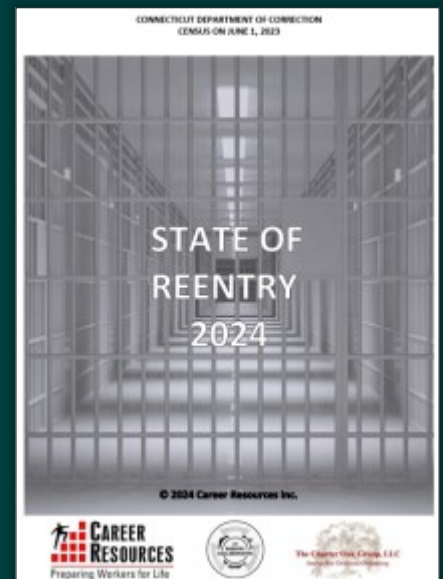
¹Pages 8, 132-6, December. <https://www.childrensissues.org/wp-content/uploads/2023/11/Reeling-from-the-Impact-Rethinking-Reentry-Wage-and-Policy-Barriers-2023-11-16.pdf>

²Paul L. Jaganat, 2023. *Using Reentry Data to Address Reentry Barriers and Policy Barriers*. Connecticut's Commission on Economic and Industry. Connecticut Issues for Children. Obtained on December 12, 2023 at <https://www.childrensissues.org/wp-content/uploads/2023/11/Reeling-from-the-Impact-Rethinking-Reentry-Wage-and-Policy-Barriers-2023-11-16.pdf>



To create data visualizations for each region, the filter in the "Visualizations" Excel file is used to select the regions one-by-one, and a copy of the Regional template is saved each time. Once the three sections have been made, they get combined into one Word document, the Table of Contents is updated and a PDF is created.

By having several steps in each phase automated, accuracy has improved. As a result, the project takes less time to complete not only because the visualizations and the compilation are done quicker but also because less revisions are needed.



The full report is available [on Career Resources' website](#).