PSEO COALITION

Using PSEO Data in Practice

Tools, Insights, and Key Considerations

July 29, 2025

Presenter and Host

Christopher Peña, PhD
Assistant Professor of the Practice, University of Denver
Intern Supervisor, PSEO Coalition

Gina Johnson, PhD
Principal & Cofounder, Data EmpowerED Consulting
Director, PSEO Coalition

Introduction

Agenda

- PSEO Data in Context
- Accessing the Data
- Preparing the Data
- General Considerations
- Technical Considerations
- Questions and Discussion

Learning Objectives

By the end of this webinar, participants will be able to:

- Explain what PSEO data are and how they differ from other federal datasets.
- Identify key elements of the data and important data nuances that impact analysis.
- Access and navigate data using the PSEO Explorer and related tools.
- Apply data in common analytical workflows using Python, Tableau, or PowerBI.
- Interpret results responsibly, understanding limitations and best practices.

PSEO Data in Context

Public Data Sources

- U.S. Department of Education (ED)
  - College Scorecard
    - Median earnings
    - Percent of postsecondary graduates earning more than high school graduates
  - National Center for Education Statistics (NCES)
    - Annual earnings by educational attainment
    - Employment and unemployment rates by educational attainment
    - Employment outcomes of bachelor's degree holders
- U.S. Census Bureau
  - American Community Survey (ACS)
    - Median earnings by educational attainment
    - Median earnings by geography and demographics

Accessing the Data

Downloading the Excel File
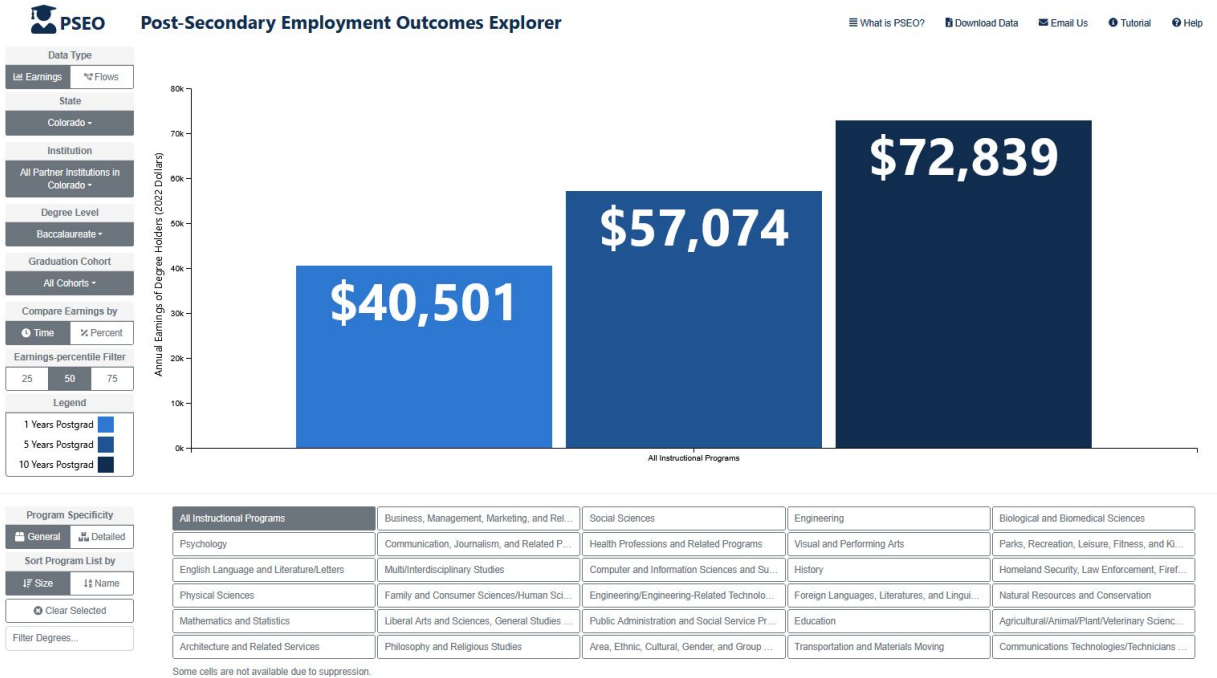
Downloading the Excel File

ADVANTAGES

- File includes qualitative labels.
- File is smaller due to being limited to one state.
- File includes both Earnings and Flows data as separate sheets.

DISADVANTAGES

- National and multi-state analysis requires downloading multiple files.
- File includes label formatting, such as carriage returns.

Downloading the CSV File

// Census.gov › Business & Industry › Center for Economic Studies › Longitudinal Employer-Household Dynamics › Data

Longitudinal Employer-Household Dynamics

| Main | Applications | Data | Workshops/Webinars | Training | Research | State Partners | LED in Action |

Applications

- J2J Explorer
- LED Extraction Tool
- OnTheMap
- OnTheMap for Emergency Management
- PSEO Explorer
- QWI Explorer
- VEO Explorer

Useful Links

- Center for Economic Studies
- J2J Data
- LODES Data
- PSEO Data
- QWI Data
- VEO Data
- LED Workshop

Contact Information

Email us:

General

LODES/OnTheMap

QWI/QWI Explorer

J2J/J2J Explorer

Post-Secondary Employment Outcomes (PSEO)

Post-Secondary Employment Outcomes (PSEO) are experimental tabulations developed by researchers at the U.S. Census Bureau. PSEO data provide earnings and employment outcomes for college and university graduates by degree level, degree major, post-secondary institution, and state of institution. These statistics are generated by matching university transcript data with a national database of jobs, using state-of-the-art confidentiality protection mechanisms to protect the underlying data.

The PSEO are made possible through data sharing partnerships between universities, university systems, State Departments of Education, State Labor Market Information offices, and the U.S. Census Bureau. PSEO data are available for post-secondary institutions whose transcript data have been made available to the Census Bureau through a data-sharing agreement.

Download Public-Use Data

We release two classes of files for each of the tabulations, Graduate Earnings and Employment Flows:

- Comprehensive dataset, which includes all institutions and crossings
- State datasets, which include all institutions in a state and are a subset of the above release

Data files are provided in zipped CSV and XLS formats and can be downloaded below. The XLS files have variable labels attached, but do not include all the possible rows from Employment Flows, due to size constraints.

State/Territory: [All ⌄]

[View Files]   Close

Data Schema

pseo_all_institutions.csv	18 Jun 2025 15:36	42 KB
pseo_all_partners.txt	25 Jun 2025 14:57	4 KB
pseoe_all.csv.gz	18 Jun 2025 15:36	11 MB
pseof_all.csv.gz	18 Jun 2025 15:36	161 MB
version_pseo.txt	18 Jun 2025 15:36	4 KB

PSEO data can also be accessed via the PSEO Explorer visualization tool. This interactive tool allows for comparisons of employment outcomes through dynamic grouped bar charts and employment flows through Sankey diagrams. To browse PSEO data files, go to: lehd.ces.census.gov/data/pseo/. Column definitions and other important information can be found in the schema files located in: lehd.ces.census.gov/data/schema/latest/lehd_public_use_schema.html.

Downloading the CSV File

ADVANTAGES

- File includes data for all participating institutions or for a single state.
- File includes additional aggregations not available in the Excel file.
- Earnings and Flows data are provided in separate files.

DISADVANTAGES

- File does not include qualitative labels.
- File size is larger due for national data to delivering a larger data set.

Automating the Process

Step 2: Download, unzip, and read in the raw PSEO earnings data.

```
[513]:  url = "https://lehd.ces.census.gov/data/pseo/latest_release/all/pseoe_all.csv.gz"  # Link to the compressed data file.
        gz_filename = "pseoe_all.csv.gz"  # Name of the compressed data file.
        unzipped_filename = "pseoe_all.csv"  # Name of the data file within the compressed data file.

        try:
            # Step 1: Download the file.
            print(f"Downloading file from {url}...")
            response = requests.get(url, stream=True)
            response.raise_for_status()  # Raise an HTTPError for bad responses (4xx and 5xx)

            # Write the downloaded content to a .gz file.
            with open(gz_filename, "wb") as gz_file:
                gz_file.write(response.content)
            print(f"File downloaded successfully as {gz_filename}.")

            # Step 2: Read the data directly from the .gz file into a data frame.
            print(f"Reading data from {gz_filename} into a data frame ...")
            with gzip.open(gz_filename, "rt") as gz_file:
                pseoe = pd.read_csv(gz_file, dtype={'institution':'string', 'degree_level':'string', 'cipcode':'string'}, low_memory=False)
            print("Data read successfully into a data frame.")

            # Step 3: Clean up - delete the .gz file and the unzipped file if it exists.
            print(f"Deleting {gz_filename}...")
            os.remove(gz_filename)
            if os.path.exists(unzipped_filename):
                print(f"Deleting {unzipped_filename}...")
                os.remove(unzipped_filename)
```

Selecting an Option

When selecting which type of file to download, consider:

- What is the scope of data you will use?
- Where do you want to do calculations and transformations?
- How much time do you have to prepare and develop data products?
- What tools will you use to prepare, analyze, and communicate the data?
- What is your level of technical expertise with the tools?

Preparing the Data

Metadata and Documentation

Post-Secondary Employment Outcomes (PSEO)

Post-Secondary Employment Outcomes (PSEO) are experimental tabulations developed by researchers at the U.S. Census Bureau. PSEO data provide earnings and employment outcomes for college and university graduates by degree level, degree major, post-secondary institution, and state of institution. These statistics are generated by matching university transcript data with a national database of jobs, using state-of-the-art confidentiality protection mechanisms to protect the underlying data.

The PSEO are made possible through data sharing partnerships between universities, university systems, State Departments of Education, State Labor Market Information offices, and the U.S. Census Bureau. PSEO data are available for post-secondary institutions whose transcript data have been made available to the Census Bureau through a data-sharing agreement.

Download Public-Use Data

We release two classes of files for each of the tabulations, Graduate Earnings and Employment Flows:

- Comprehensive dataset, which includes all institutions and crossings
- State datasets, which include all institutions in a state and are a subset of the above release

Data files are provided in zipped CSV and XLS formats and can be downloaded below. The XLS files have variable labels attached, but do not include all the possible rows from Employment Flows, due to size constraints.

State/Territory: Alabama ⌄

[View Files] [Close]

Data Schema

| pseo_al.xlsx | | 18 Jun 2025 15:36 | 3 MB |
pseo_al_institutions.csv		18 Jun 2025 15:36	188 B
pseo_al_partners.txt		18 Jun 2025 15:36	95 B
pseoe_al.csv.gz		18 Jun 2025 15:36	93 KB
pseof_al.csv.gz		18 Jun 2025 15:36	1 MB
version_pseo.txt		18 Jun 2025 15:36	122 B

PSEO data can also be accessed via the PSEO Explorer visualization tool. This interactive tool allows for comparisons of employment outcomes through dynamic grouped bar charts and employment flows through Sankey diagrams. To browse PSEO data files, go to: lehd.ces.census.gov/data/pseo/. Column definitions and other important information can be found in the schema files located in: lehd.ces.census.gov/data/schema/latest/lehd_public_use_schema.html.

PSEO Help

Learn more about PSEO by choosing one of the links below.

- PSEO Methodology and Data Sources
- PSEO Data Notices (427 KB)
- PSEO Data Schema for Most Recent Release
- PSEO Technical Documentation (213 KB)
- Technical Appendix for PSEO Protection System (225 KB)
- PSEO Partnership SOPs

EXPERIMENTAL DATA
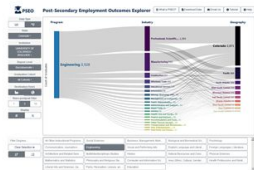
Metadata and Documentation

5.15.2. CIP Codes

(label_cipcode.csv)

CIP codes are sourced from the National Center for Education Statistics (NCES), Integrated Postsecondary Education Data System (IPI Data are reported using 2020 CIP codes, for all years.

cipcode	label	cip_level	CIPFamily	CIPDefinition
00	All Instructional Programs	A	00	All CIP Codes
01	Agricultural/Animal/ Plant/Veterinary Science and Related Fields	2	01	Instructional programs that focus on agriculture, animal, plant, veterinary, and related sciences and that prepares individuals to apply specific knowledge, methods, and techniques to the management and performance of agricultural and veterinary operations.
01.00	Agriculture, General	4	01	Instructional content is defined in code 01.0000.
01.0000	Agriculture, General	6	01	A program that focuses on the general principles and practice of agricultural research and production and that may prepare individuals to apply this knowledge to the solution of practical agricultural problems. Includes instruction in basic animal, plant, and soil science; animal husbandry and plant cultivation; soil conservation; and agricultural operations such as farming, ranching, and agricultural business.

File Structure

The PSEO Earnings and Flows data sets use a wide format with measures captured as columns rather than values in rows.

ADVANTAGES
- Math across measures is simplified, such as *percent working in state.*
- Merging related data sources is simplified.

DISADVANTAGES
- Tableau and Power BI prefer tall data formats.

Adding Data Labels

BASE PSEO DATA

- PSEO aggregate levels
- Institution names
- Institution level names
- Degree level names
- CIP code level names
- CIP code names
- Graduation cohort names
- Geography names
- Industry names

5.14. Degree Level

(label_degree_level.csv)

The degree levels are sourced from the National Center for Education Statistics (NCES), Integrated Postsecondary Education Data System (IPEDS).

degree_level	label
00	All Degree Levels
01	Certificate < 1 year
02	Certificate 1-2 years
03	Associates
04	Certificate 2-4 years
05	Baccalaureate
06	Post-Bacc Certificate
07	Masters
08	Post-Masters Certificate
17	Doctoral - Research/Scholarship
18	Doctoral - Professional Practice

Bringing Data Together

Data can be joined when columns are shared between files/sources:

SOLUTIONS

Excel	Use the XLOOKUP or INDEX + MATCH functions.
Python	Read data into a data frame and merge using Pandas.
Tableau	Blend or join data on the Data Sources tab.
Power BI	Merge or relate data using Power Query.

Recommendations for Automation

- Determine the scope of the file you are creating.
- Drop unused columns and write to CSV to minimize file size.
- Be mindful of memory limitations on your computer.
- Use left joins to retain records when related data might be missing.
- Use original naming conventions to facilitate consistency and joins.
- Drop duplicated join columns where appropriate.
- Define the data types for string columns that look like numbers:
  - Institution (OPEID number)
  - CIP code (2-digit and 4-digit)
  - Degree level

General Considerations

Cohort Model

Microdata are aggregated into cohorts based on the degree level:
- All Cohorts
- Baccalaureate – 3 years
- All other degree levels – 5 years

Be mindful of when to use the All Cohorts level to avoid incorrectly summarizing the aggregate data!

Record Aggregations (Sample)

	agg_level_pseo	grad_char	firm_char	cip_level	inst_level	geo_level	ind_level	by_grad_c	by_degree	pseoe	pseof
1	agg_level_pseo	grad_char	firm_char	cip_level	inst_level	geo_level	ind_level	by_grad_c	by_degree	pseoe	pseof
2	26	Degree Level * State of Institution		A	S	N	A	0	1	1	1
3	28	Degree Level * CIP 2-digit * State of Institution		2	S	N	A	0	1	1	1
4	30	Degree Level * CIP 4-digit * State of Institution		4	S	N	A	0	1	1	0
5	32	Degree Level * Start Year for Graduation Cohort * State of Institution		A	S	N	A	1	1	1	1
6	34	Degree Level * CIP 2-digit * Start Year for Graduation Cohort * State of Institution		2	S	N	A	1	1	1	1
7	36	Degree Level * CIP 4-digit * Start Year for Graduation Cohort * State of Institution		4	S	N	A	1	1	1	0
8	38	Degree Level * Institution ID		A	I	N	A	0	1	1	1
9	40	Degree Level * CIP 2-digit * Institution ID		2	I	N	A	0	1	1	1
10	42	Degree Level * CIP 4-digit * Institution ID		4	I	N	A	0	1	1	0
11	44	Degree Level * Start Year for Graduation Cohort * Institution ID		A	I	N	A	1	1	1	1
12	46	Degree Level * CIP 2-digit * Start Year for Graduation Cohort * Institution ID		2	I	N	A	1	1	1	1
13	48	Degree Level * CIP 4-digit * Start Year for Graduation Cohort * Institution ID		4	I	N	A	1	1	1	0
14	74	Degree Level * State of Institution	NAICS Sector	A	S	N	S	0	1	0	1
15	76	Degree Level * CIP 2-digit * State of Institution	NAICS Sector	2	S	N	S	0	1	0	1
16	78	Degree Level * CIP 4-digit * State of Institution	NAICS Sector	4	S	N	S	0	1	0	0
17	80	Degree Level * Start Year for Graduation Cohort * State of Institution	NAICS Sector	A	S	N	S	1	1	0	1
18	82	Degree Level * CIP 2-digit * Start Year for Graduation Cohort * State of Institution	NAICS Sector	2	S	N	S	1	1	0	1
19	84	Degree Level * CIP 4-digit * Start Year for Graduation Cohort * State of Institution	NAICS Sector	4	S	N	S	1	1	0	0
20	86	Degree Level * Institution ID	NAICS Sector	A	I	N	S	0	1	0	1
21	88	Degree Level * CIP 2-digit * Institution ID	NAICS Sector	2	I	N	S	0	1	0	1
22	90	Degree Level * CIP 4-digit * Institution ID	NAICS Sector	4	I	N	S	0	1	0	0
23	92	Degree Level * Start Year for Graduation Cohort * Institution ID	NAICS Sector	A	I	N	S	1	1	0	1
24	94	Degree Level * CIP 2-digit * Start Year for Graduation Cohort * Institution ID	NAICS Sector	2	I	N	S	1	1	0	1
25	96	Degree Level * CIP 4-digit * Start Year for Graduation Cohort * Institution ID	NAICS Sector	4	I	N	S	1	1	0	0
26	122	Degree Level * State of Institution	Census Division	A	S	D	A	0	1	0	1
27	124	Degree Level * CIP 2-digit * State of Institution	Census Division	2	S	D	A	0	1	0	1
28	126	Degree Level * CIP 4-digit * State of Institution	Census Division	4	S	D	A	0	1	0	0
29	128	Degree Level * Start Year for Graduation Cohort * State of Institution	Census Division	A	S	D	A	1	1	0	1
30	130	Degree Level * CIP 2-digit * Start Year for Graduation Cohort * State of Institution	Census Division	2	S	D	A	1	1	0	1
31	132	Degree Level * CIP 4-digit * Start Year for Graduation Cohort * State of Institution	Census Division	4	S	D	A	1	1	0	0
32	134	Degree Level * Institution ID	Census Division	A	I	D	A	0	1	0	1
33	136	Degree Level * CIP 2-digit * Institution ID	Census Division	2	I	D	A	0	1	0	1
34	138	Degree Level * CIP 4-digit * Institution ID	Census Division	4	I	D	A	0	1	0	0
35	140	Degree Level * Start Year for Graduation Cohort * Institution ID	Census Division	A	I	D	A	1	1	0	1
36	142	Degree Level * CIP 2-digit * Start Year for Graduation Cohort * Institution ID	Census Division	2	I	D	A	1	1	0	1
37	144	Degree Level * CIP 4-digit * Start Year for Graduation Cohort * Institution ID	Census Division	4	I	D	A	1	1	0	0

Column Aggregations

The following aggregations are available within the data:

Institution: All Institutions in [State]

CIP Level: All Degree Fields

CIP Code: All Instructional Programs

Graduation Cohort: All Cohorts

Geography: National (50 states + DC)

Industry: All NAICS Sectors

Technical Considerations

Removing Carriage Returns

Certain qualitative labels in the Excel file include carriage returns (/r and /n).

SOLUTIONS

Excel	Use the CLEAN function.
Python	Use the str.replace method.
Tableau	Use the REPLACE function.
Power BI	Use the SUBSTITUTE function.

Deriving the Graduation Cohort Label

The graduation cohort label must be derived from the graduation cohort to display the start and end years:

Step 15: Add grad cohort labels to the data frame.

```
[539]:  pseoe['label_grad_cohort'] = pseoe.apply(
            lambda row: 'All Cohorts' if row['grad_cohort'] == 0
                        else f"{row['grad_cohort']}-{row['grad_cohort'] + row['grad_cohort_years'] - 1}",
            axis=1
        )

        print("Grad cohort labels added to the data frame.")
```

Treat grad_cohort values as numbers to enable addition and subtraction.

Identifying Institutions

Institutions in the PSEO data are identified using the OPEID.
- Out of approximately 900 distinct institutions in the PSEO data, 63 do not exist in the IPEDS universe.
- While institution name can be used to join PSEO and IPEDS data for most colleges and universities, it is not always a perfect match.

	A	B	C	D	E	F	G
1	institution	label_institution_pseo	ipeds_id	label_institution_ipeds	opeid	institution_level	label_institution_level
2	00156900	Georgia Institute of Technology	139755	Georgia Institute of Technology-Main Campus	00156900	1	4YR
3	03170300	Texas A & M University - Texarkana	224545	Texas A&M University-Texarkana	03170300	1	4YR
4	00108300	University of Arizona (The)	104179	University of Arizona	00108300	1	4YR

Adding State for State-Level Aggregations

Institutions in [State/Division] records do not have a corresponding state code.

SOLUTION

Join the statefips field to a dictionary with state codes.

5.13.2. Institution

(label_institution.csv)

Institution identifiers are sourced from the U.S. Department of Education, Federal Student Aid office. This list has been supplemented with records for regional groupings of institutions (may be used in future PSEO tabulations).

institution	label	city	institution_state	inst_level	statefips
1	Institutions in New England Division			D	
2	Institutions in Middle Atlantic Division			D	
3	Institutions in East North Central Division			D	
…					
00137000	University of Colorado Boulder	Boulder	CO	I	08
00309000	Ohio State University	Columbus	OH	I	39
00332900	Pennsylvania State University (The)	University Park	PA	I	42
00389500	University of Wisconsin - Madison	Madison	WI	I	55

Visualizing Geography

You can use publicly available shape files to visualize geography flows:
- PSEO Geography column = U.S. Census Bureau Division column.

Both Tableau and Power BI can easily accommodate geographic data:
- Consider whether you can connect to an ESRI (ArcGIS) server.
- For Power BI, you may need to connect to a TopoJSON file.

Next Up in Our Webinar Series...

Advanced PSEO Data Use: APIs and Data Integration for Deeper Analysis

**Thursday, August 28, 2025 | 3 – 4 p.m. ET**

In this advanced session, learn how to use the Postsecondary Employment Outcomes (PSEO) data programmatically using the API and how to enrich your analyses by supplementing PSEO with other key datasets, including IPEDS survey data, U.S. Census earnings data, and cost of living indicators. This webinar is ideal for users looking to build custom workflows and perform more comprehensive, context-rich analyses using PSEO data.

Register

PSEO
Resources

PSEO Resources
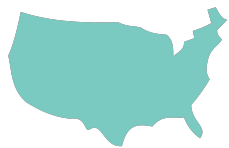
PSEO Coalition
pseocoalition.org

Resource Library
pseocoalition.org/resource-library/

PSEO Explorer
lehd.ces.census.gov/applications/pseo

PSEO Datasets
lehd.ces.census.gov/data/pseo_experimental.html