

Datasheets for Dataset 7: time series dataset

Motivation for Dataset Creation

Why was the dataset created?

Dataset 7 comes from Labor Force Statistics from the Current Population Survey. The Current Population Survey (CPS) is a monthly survey of households conducted by the Bureau of Census for the Bureau of Labor Statistics. It provides a comprehensive body of data on the labor force, employment, unemployment, persons not in the labor force, hours of work, earnings, and other demographic and labor force characteristics.

What (other) tasks could the dataset be used for?

Yes.

Has the dataset been used for any tasks already?

Yes.

CPS is one of the main labor dataset collected by the US government. This dataset has been used extensively in academic literature for decades.

Dataset Composition

The proportion of people working in different occupations is broken down by race and gender as follows.

Occupation, race, and Hispanic or Latino ethnicity	Total		Men		Women	
	2019	2020	2019	2020	2019	2020
TOTAL						
Total, 16 years and over (in thousands)	157,538	147,705	83,460	78,560	74,078	69,234
Percent of total employed	100.0	100.0	100.0	100.0	100.0	100.0
Management, professional, and related occupations	40.8	43.1	37.1	39.1	44.9	47.5
Management, business, and financial operations occupations	17.1	18.4	18.1	19.1	16.1	17.5
Professional and related occupations	23.6	24.7	19.8	20.6	28.8	30.0
Service occupations	12.1	15.5	12.7	12.5	21.0	18.8
Sales and office occupations	21.2	20.1	15.8	14.6	27.3	26.3
Sales and related occupations	9.9	9.6	9.6	9.2	10.3	10.0
Office and administrative support occupations	11.3	10.5	6.2	5.4	17.0	16.3
Natural resources, construction, and maintenance occupations	9.1	9.0	16.3	16.0	1.0	1.1
Farming, fishing, and forestry occupations	0.7	0.7	1.0	1.0	0.4	0.4
Construction and extraction occupations	5.3	5.2	9.6	9.4	0.4	0.4
Installation, maintenance, and repair occupations	3.1	3.1	5.6	5.6	0.3	0.3
Production, transportation, and material moving occupations	11.8	12.3	17.2	17.7	5.8	6.2
Production occupations	5.4	5.1	7.3	6.9	3.3	3.1
Transportation and material moving occupations	6.4	7.2	9.9	10.8	2.5	3.1
White						
Total, 16 years and over (in thousands)	122,441	115,341	66,109	62,446	56,331	52,895
Percent of total employed	100.0	100.0	100.0	100.0	100.0	100.0
Management, professional, and related occupations	41.4	43.4	37.5	39.3	46.1	48.4
Management, business, and financial operations occupations	18.1	19.2	19.2	20.2	16.7	18.1

Data Collection Process

The CPS is administered by the Census Bureau using a probability-selected sample of about 60,000 occupied households. The fieldwork is conducted during the calendar week that includes the 19th of the month. The questions refer to activities during the prior week; that is, the week that includes the 12th of the month. Households from all 50 states and the District of Columbia are in the survey for 4 consecutive months, out for 8, and then return for another 4 months before leaving the sample permanently. This design ensures a high degree of continuity from one month to the next (as well as over the year). The 4-8-4 sampling scheme has the added benefit of allowing the constant replenishment of the sample without excessive burden to respondents.

The CPS questionnaire is a completely computerized document that is administered by Census Bureau field representatives across the country through both personal and telephone interviews. Additional telephone interviewing is conducted from the Census Bureau's two centralized collection facilities in Jeffersonville, Indiana and Tucson, Arizona.

In general, the BLS publishes labor force data only for people aged 16 and over, since those under 16 are limited in their labor market activities by compulsory schooling and child labor laws. No upper age limit is used, and full-time students are treated the same as nonstudents. One person generally responds for all eligible members of the household. The person who responds is called the "reference person" and usually is the person who either owns or rents the housing unit. If the reference person is

not knowledgeable about the employment status of the others in the household, attempts are made to contact those individuals directly.

Dataset Preprocessing

What preprocessing/cleaning was done?

No preprocessing/cleaning was done.

Dataset Distribution

How is the dataset distributed?

The dataset is distributed through the [BLS website](#).

Dataset Maintenance

Who is supporting/hosting/maintaining the dataset?

Bureau of Labor Statistics.

Will the dataset be updated?

Yes. It is part of the regular survey that the Bureau of Labor Statistics will conduct every year.

Legal & Ethical Considerations

If the dataset relates to people (e.g., their attributes) or was generated by people, were they informed about the data collection?

Yes. Responses are obtained by mail, Internet or other electronic means, email, telephone, or personal visit.

If it relates to people, were they told what the dataset would be used for and did they consent? What community norms exist for data collected from human communications?

Yes. As mentioned in the data collection part, this data set is collected by telephone, email, personal visits with consent.

If it relates to people, could this dataset expose people to harm or legal action?

No. This dataset only contains the proportion of people working in a certain industry, and it is not directly related to any individual.

If it relates to people, does it unfairly advantage or disadvantage a particular social group?

No. This dataset is collected to show the distribution of people working in different occupations and different social groups.

If it relates to people, were they provided with privacy guarantees?

Yes. This dataset is not personally-identifiable.

Does the dataset comply with the EU General Data Protection Regulation (GDPR)?

Yes. This is a dataset that complies with the GDPR.

Does the dataset contain information that might be considered sensitive or confidential?

No. This dataset is public information published on the BLS website.

Does the dataset contain information that might be considered inappropriate or offensive?

No. This dataset is meant to show the distribution of people working in different occupations and provide a guide for designing public policies.