

Evan Reid

University of Wisconsin – Stout

MSCS 335 – Machine Learning

### Current Population Survey Dataset Description

The Current Population Survey is one of the oldest, largest, and most recognized surveys in the United States. This survey provides information about individuals in society such as work, earnings, and our education. The CPS is used to collect data for a variety of other studies that keep the nation informed of the economic and social well-being of its people. This survey is taken monthly, however the Census Bureau website does not provide the raw data of survey responses from individuals unless you have the access rights. However, a few months of the survey have been published online. The August 2016 CPS dataset can be found in a raw .CSV format on Kaggle.com via the following link:

<https://www.kaggle.com/census/current-population-survey>

I wanted to create a dataset that is much like the UCI Adult dataset (from 1995):

<http://mlr.cs.umass.edu/ml/datasets/Adult>

The 2016 CPS dataset has many columns and many incomplete survey answers, so I edited the to capture the most relevant data. It is important to note that I have stripped the data of any multiple-job holding individuals, so the below data only reflects individuals with a single occupation. This updated version of the 2016 CPS dataset is very similar to the UCI adult dataset with some added columns.

The Current Population Survey (CPS) is administered, processed, researched and disseminated by the U.S. Census Bureau on behalf of the Bureau of Labor Statistics (BLS).

### Column Headers

**1. PRTAGE** (Numerical) – Persons age in years

Valid Entries: 0 – 99

**2. PESEX** (Categorical) – Sex (Cate

Valid Entries:

1 Male

2 Female

**3. GESTFIPS** (Categorical) – State Code

Valid Entries:

01	AL	30	MT
02	AK	31	NE
04	AZ	32	NV
05	AR	33	NH
06	CA	34	NJ
08	CO	35	NM
09	CT	36	NY
10	DE	37	NC

11	DC	38	ND
12	FL	39	OH
13	GA	40	OK
15	HI	41	OR
16	ID	42	PA
17	IL	44	RI
18	IN	45	SC
19	IA	46	SD
20	KS	47	TN
21	KY	48	TX
22	LA	49	UT
23	ME	50	VT
24	MD	51	VA
25	MA	53	WA
26	MI	54	WV
27	MN	55	WI
28	MS	56	WY
29	MO		

**4. PRCITSHP (Categorical) – Citizenship status**

Valid Entries:

- 1 NATIVE, BORN IN THE UNITED STATES
- 2 NATIVE, BORN IN PUERTO RICO OR OTHER U.S. ISLAND AREAS
- 3 NATIVE, BORN ABROAD OF AMERICAN PARENT OR PARENTS
- 4 FOREIGN BORN, U.S. CITIZEN BY NATURALIZATION
- 5 FOREIGN BORN, NOT A CITIZEN OF THE UNITED STATES

**5. PTDRACE (Categorical) – Race**

Valid Entries:

- 1 White Only
- 2 Black Only
- 3 American Indian, Alaskan Native Only
- 4 Asian Only
- 5 Hawaiian/Pacific Islander Only
- 6 White-Black
- 7 White-AI
- 8 White-Asian
- 9 White-HP
- 10 Black-AI
- 11 Black-Asian
- 12 Black-HP
- 13 AI-Asian
- 14 AI-HP
- 15 Asian-HP

**6. PEMARITL (Categorical) – Marital status**

Valid Entries:

- 1 MARRIED - SPOUSE PRESENT
- 2 MARRIED - SPOUSE ABSENT
- 3 WIDOWED
- 4 DIVORCED
- 5 SEPARATED
- 6 NEVER MARRIED

**7. HRNUMHOU** (Numerical) – Total number of persons living in household

Valid Entries: 0-16

**8. PRNMCHLD** (Numerical) – Number of own children under 18 years of age

Valid Entries: 0 - 99

**9. PEEDUCA** (Numerical/Categorical) – Highest level of school completed, or degree received

Valid Entries:

- 31 LESS THAN 1ST GRADE
- 32 1ST, 2ND, 3RD OR 4TH GRADE
- 33 5TH OR 6TH GRADE
- 34 7TH OR 8TH GRADE
- 35 9TH GRADE
- 36 10TH GRADE
- 37 11TH GRADE
- 38 12TH GRADE NO DIPLOMA
- 39 HIGH SCHOOL GRAD-DIPLOMA OR EQUIV (GED)
- 40 SOME COLLEGE BUT NO DEGREE
- 41 ASSOCIATE DEGREE-OCCUPATIONAL/VOCATIONAL
- 42 ASSOCIATE DEGREE-ACADEMIC PROGRAM
- 43 BACHELOR'S DEGREE (EX: BA, AB, BS)
- 44 MASTER'S DEGREE (EX: MA, MS, MEng, MEd, MSW)
- 45 PROFESSIONAL SCHOOL DEG (EX: MD, DDS, DVM)
- 46 DOCTORATE DEGREE (EX: PhD, EdD)

**10. PRCOW1** (Categorical) – Class of worker

Valid Entries:

- 1 FEDERAL GOVT
- 2 STATE GOVT
- 3 LOCAL GOVT
- 4 PRIVATE (INCL. SELF-EMPLOYED INCORP.)
- 5 SELF-EMPLOYED, UNINCORP.
- 6 WITHOUT PAY

**11. PRDTIND1** (Categorical) – Industry of worker

Valid Entries:

- 1 Agriculture
- 2 Forestry, logging, fishing, hunting, and trapping
- 3 Mining
- 4 Construction
- 5 Nonmetallic mineral product manufacturing
- 6 Primary metals and fabricated metal products
- 7 Machinery manufacturing
- 8 Computer and electronic product manufacturing
- 9 Electrical equipment, appliance manufacturing
- 10 Transportation equipment manufacturing
- 11 Wood products
- 12 Furniture and fixtures manufacturing
- 13 Miscellaneous and not specified manufacturing
- 14 Food manufacturing
- 15 Beverage and tobacco products

- 16 Textile, apparel, and leather manufacturing
- 17 Paper and printing
- 18 Petroleum and coal products manufacturing
- 19 Chemical manufacturing
- 20 Plastics and rubber products
- 21 Wholesale trade
- 22 Retail trade
- 23 Transportation and warehousing
- 24 Utilities
- 25 Publishing industries (except internet)
- 26 Motion picture and sound recording industries
- 27 Broadcasting (except internet)
- 28 Internet publishing and broadcasting
- 29 Telecommunications
- 30 Internet service providers and data processing services
- 31 Other information services
- 32 Finance
- 33 Insurance
- 34 Real estate
- 35 Rental and leasing services
- 36 Professional and technical services
- 37 Management of companies and enterprises
- 38 Administrative and support services
- 39 Waste management and remediation services
- 40 Educational services
- 41 Hospitals
- 42 Health care services, except hospitals
- 43 Social assistance
- 44 Arts, entertainment, and recreation
- 45 Accommodation
- 46 Food services and drinking places
- 47 Repair and maintenance
- 48 Personal and laundry services
- 49 Membership associations and organizations
- 50 Private households
- 51 Public administration
- 52 Armed forces

**12. PRDTOCC1 (Categorical) – Occupation of Worker**

Valid Entries:

- 1 Management occupations
- 2 Business and financial operations occupations
- 3 Computer and mathematical science occupations
- 4 Architecture and engineering occupations
- 5 Life, physical, and social science occupations
- 6 Community and social service occupations
- 7 Legal occupations
- 8 Education, training, and library occupations
- 9 Arts, design, entertainment, sports, and media occupations
- 10 Healthcare practitioner and technical occupations
- 11 Healthcare support occupations
- 12 Protective service occupations
- 13 Food preparation and serving related occupations

- 14 Building and grounds cleaning and maintenance occupations
- 15 Personal care and service occupations
- 16 Sales and related occupations
- 17 Office and administrative support occupations
- 18 Farming, fishing, and forestry occupations
- 19 Construction and extraction occupations
- 20 Installation, maintenance, and repair occupations
- 21 Production occupations
- 22 Transportation and material moving occupations
- 23 Armed Forces

**13. PRHRUSL** (Numerical/Categorical) – Usual hours worked weekly

Valid Entries:

- 1 0-20 HRS
- 2 21-34 HRS
- 3 35-39 HRS
- 4 40 HRS
- 5 41-49 HRS
- 6 50 OR MORE HRS

**14. HEFAMINC** (Numerical/Categorical) – Family Income expressed as an integer representing income groups

Valid Entries:

- 1 LESS THAN \$5,000
- 2 5,000 TO 7,499
- 3 7,500 TO 9,999
- 4 10,000 TO 12,499
- 5 12,500 TO 14,999
- 6 15,000 TO 19,999
- 7 20,000 TO 24,999
- 8 25,000 TO 29,999
- 9 30,000 TO 34,999
- 10 35,000 TO 39,999
- 11 40,000 TO 49,999
- 12 50,000 TO 59,999
- 13 60,000 TO 74,999
- 14 75,000 TO 99,999
- 15 100,000 TO 149,999
- 16 150,000 OR MORE

**15. INCOVER50** (Categorical) – Value representing if the family income meets 50,000 a year.

Valid Entries: 0 and 1, 1 meaning income is 50,000 or greater

**16. ISMARRIED** (Categorical) – Value representing if the individual is married.

Valid Entries: 0 and 1, 1 meaning the individual is married