Future Insights: Trends in US Computer, Engineering, & Science Occupations

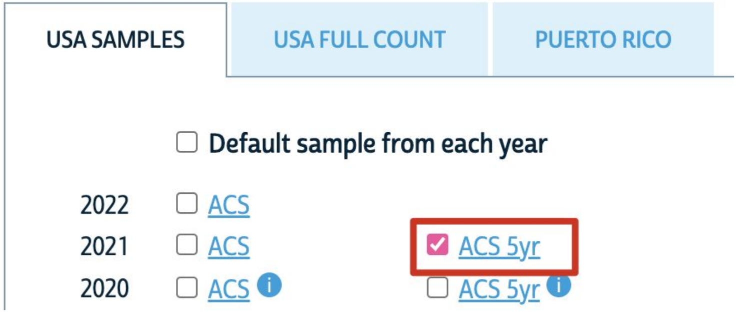
Group 3:
Yining Wang
Keyi Wang
Parisima Abdali
Jiaan Cao

Dataset Source

IPUMS USA

https://usa.ipums.org/usa/index.shtml – **Customize** the dataset

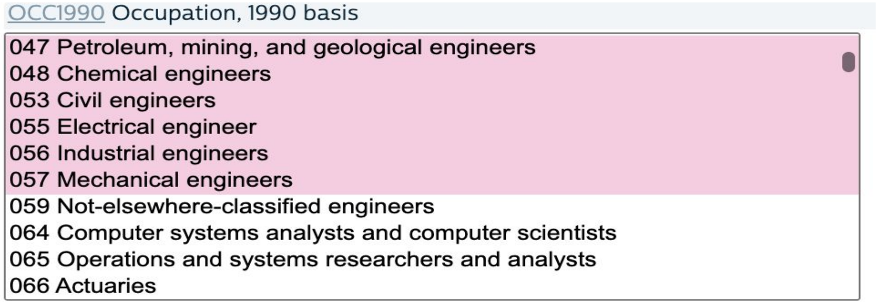
- Select year: **5-year** (2017-2021)



□ Default sample from each year

2022	□ ACS	
2021	□ ACS	☑ ACS 5yr
2020	□ ACS	□ ACS 5yr

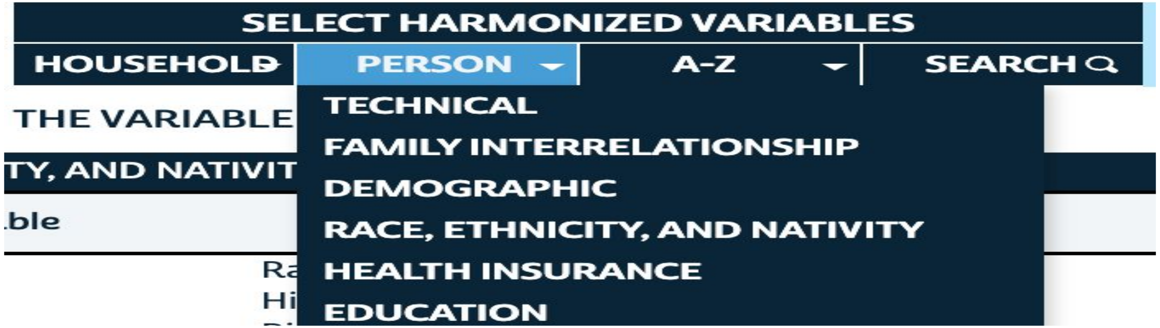
- Select **computer, engineering, & science** occupations



OCC1990 Occupation, 1990 basis

047 Petroleum, mining, and geological engineers
048 Chemical engineers
053 Civil engineers
055 Electrical engineer
056 Industrial engineers
057 Mechanical engineers
059 Not-elsewhere-classified engineers
064 Computer systems analysts and computer scientists
065 Operations and systems researchers and analysts
066 Actuaries

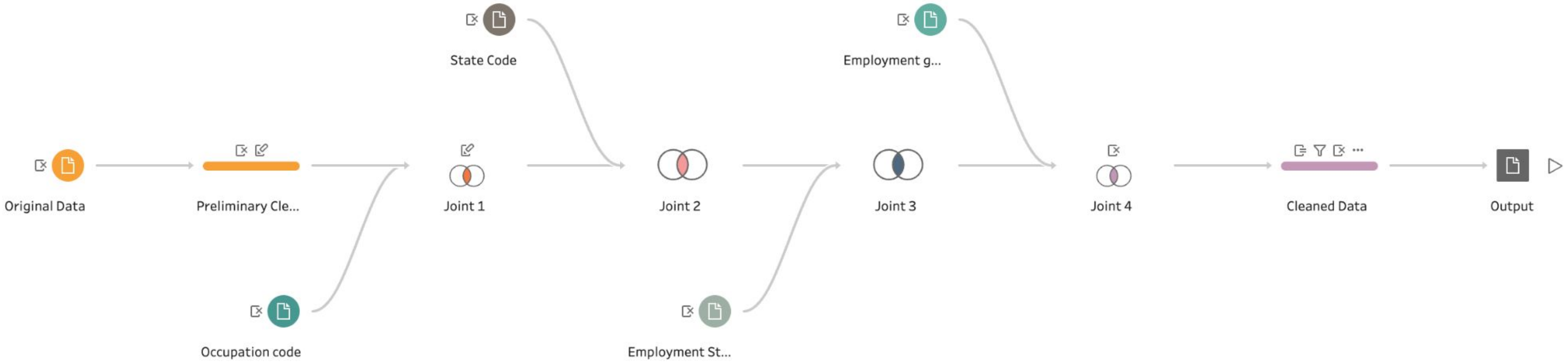
- Select **variables**



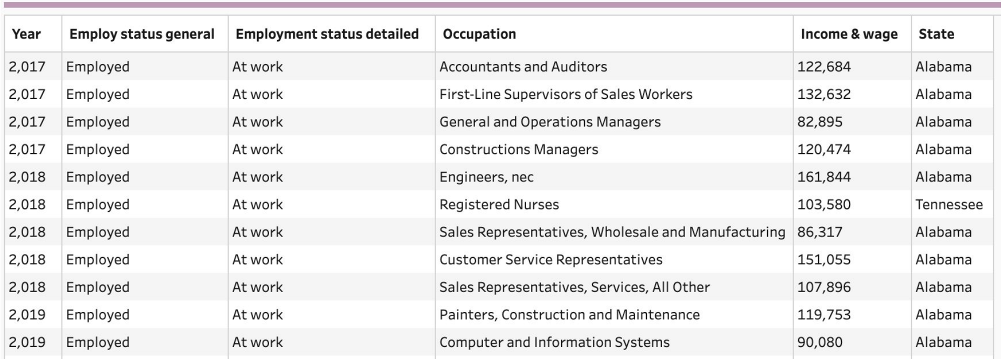
Summary Statistics

Employment Trends	Diversity correlation	Education & skills		Industry & sector	
Employment Status	**Year**	**Year**	Year of college	**Year**	Class of worker
Occupation	Gender	Gender	Degree level	Sample	Worked last year
Income & wage	**State**	Age	**State**	Serial	Occupation
State	Race	Birth year	Field of degree (General)	Pernum	Industry category
Year	Occupation	Race	Field of degree (Detailed)	Employment status label	Industry subcategory
	Age	Ethnicity	Occupation	Labor force label	**State**

Data Cleaning and Preparation Operations - Employment Trend



Final Result →



Year	Employ status general	Employment status detailed	Occupation	Income & wage	State
2,017	Employed	At work	Accountants and Auditors	122,684	Alabama
2,017	Employed	At work	First-Line Supervisors of Sales Workers	132,632	Alabama
2,017	Employed	At work	General and Operations Managers	82,895	Alabama
2,017	Employed	At work	Constructions Managers	120,474	Alabama
2,018	Employed	At work	Engineers, nec	161,844	Alabama
2,018	Employed	At work	Registered Nurses	103,580	Tennessee
2,018	Employed	At work	Sales Representatives, Wholesale and Manufacturing	86,317	Alabama
2,018	Employed	At work	Customer Service Representatives	151,055	Alabama
2,018	Employed	At work	Sales Representatives, Services, All Other	107,896	Alabama
2,019	Employed	At work	Painters, Construction and Maintenance	119,753	Alabama
2,019	Employed	At work	Computer and Information Systems	90,080	Alabama

Data Cleaning and Preparation Operations - Employment Trend

1) Original Data

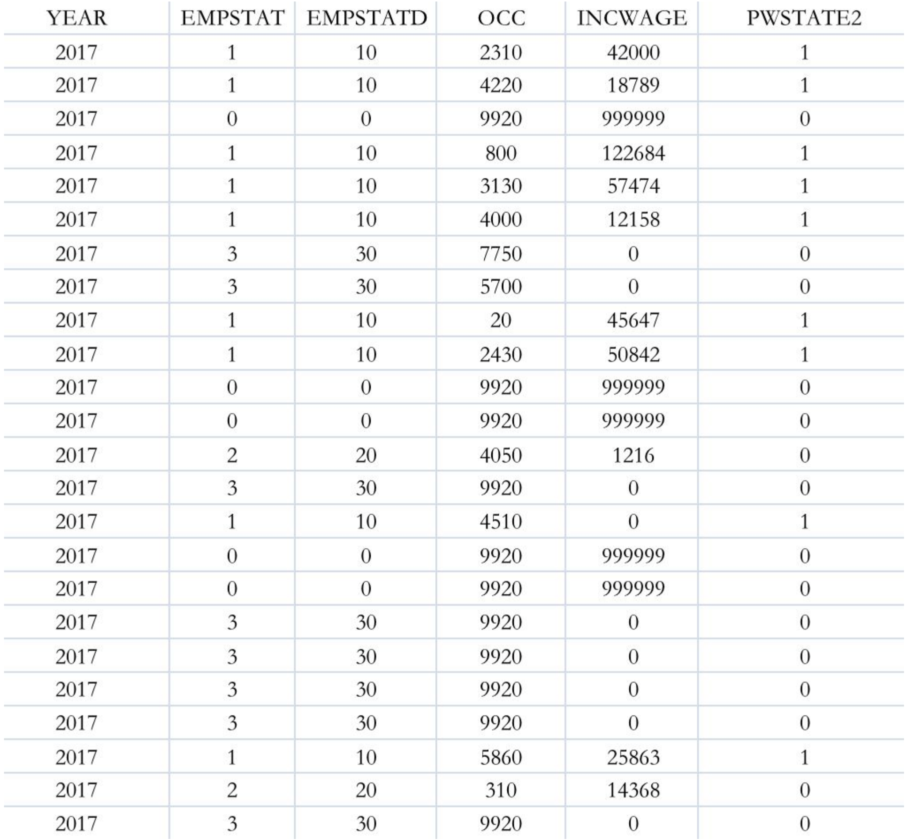
Year: 2017-2021

EMPSTAT = Employment Status:
whether the respondent was a part of the labor force,-- working or seeking work -- and unemployed.

OCC = Occupation:
a harmonized occupation coding scheme based on the Census Bureau's ACS occupation classification scheme.

INCWAGW = Income & wage:
each respondent's total pre-tax wage and salary income

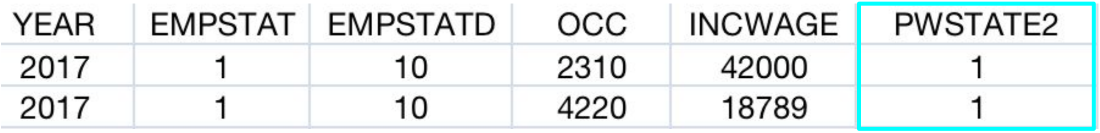
PWSTATE2 = **State**:
the state in which the respondent's primary workplace was located.



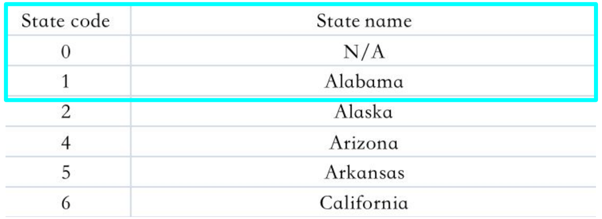
YEAR	EMPSTAT	EMPSTATD	OCC	INCWAGE	PWSTATE2
2017	1	10	2310	42000	1
2017	1	10	4220	18789	1
2017	0	0	9920	999999	0
2017	1	10	800	122684	1
2017	1	10	3130	57474	1
2017	1	10	4000	12158	1
2017	3	30	7750	0	0
2017	3	30	5700	0	0
2017	1	10	20	45647	1
2017	1	10	2430	50842	1
2017	0	0	9920	999999	0
2017	0	0	9920	999999	0
2017	2	20	4050	1216	0
2017	3	30	9920	0	0
2017	1	10	4510	0	1
2017	0	0	9920	999999	0
2017	0	0	9920	999999	0
2017	3	30	9920	0	0
2017	3	30	9920	0	0
2017	3	30	9920	0	0
2017	3	30	9920	0	0
2017	1	10	5860	25863	1
2017	2	20	310	14368	0
2017	3	30	9920	0	0

Data Cleaning and Preparation Operations - Employment Trend

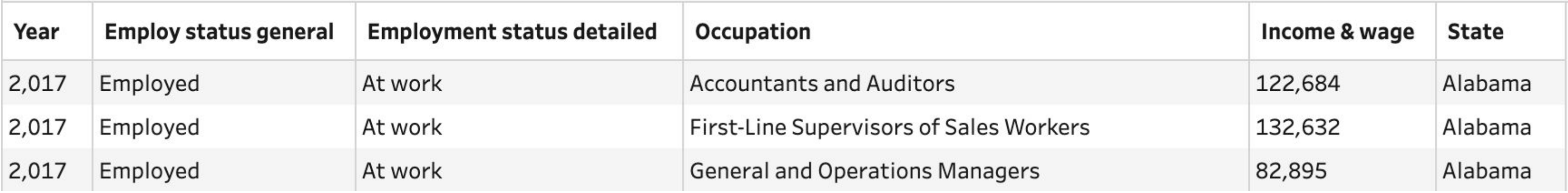
2) Code-to-label



YEAR	EMPSTAT	EMPSTATD	OCC	INCWAGE	PWSTATE2
2017	1	10	2310	42000	1
2017	1	10	4220	18789	1



State code	State name
0	N/A
1	Alabama
2	Alaska
4	Arizona
5	Arkansas
6	California

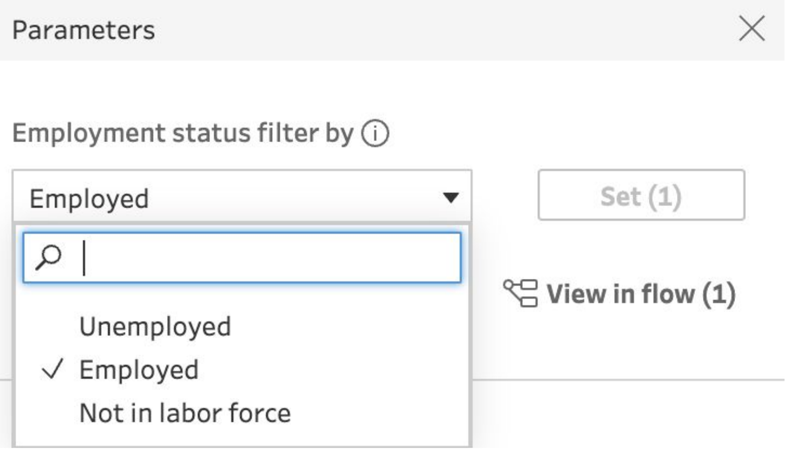


Year	Employ status general	Employment status detailed	Occupation	Income & wage	State
2,017	Employed	At work	Accountants and Auditors	122,684	Alabama
2,017	Employed	At work	First-Line Supervisors of Sales Workers	132,632	Alabama
2,017	Employed	At work	General and Operations Managers	82,895	Alabama

- Using **join** to connect the code file with original data to transfer the number to text
- Remove the N/A value

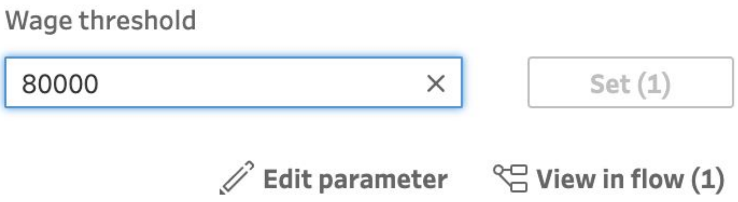
Data Cleaning and Preparation Operations - Employment Trend

3) Create Parameters



Parameters

Employment status filter by: Employed

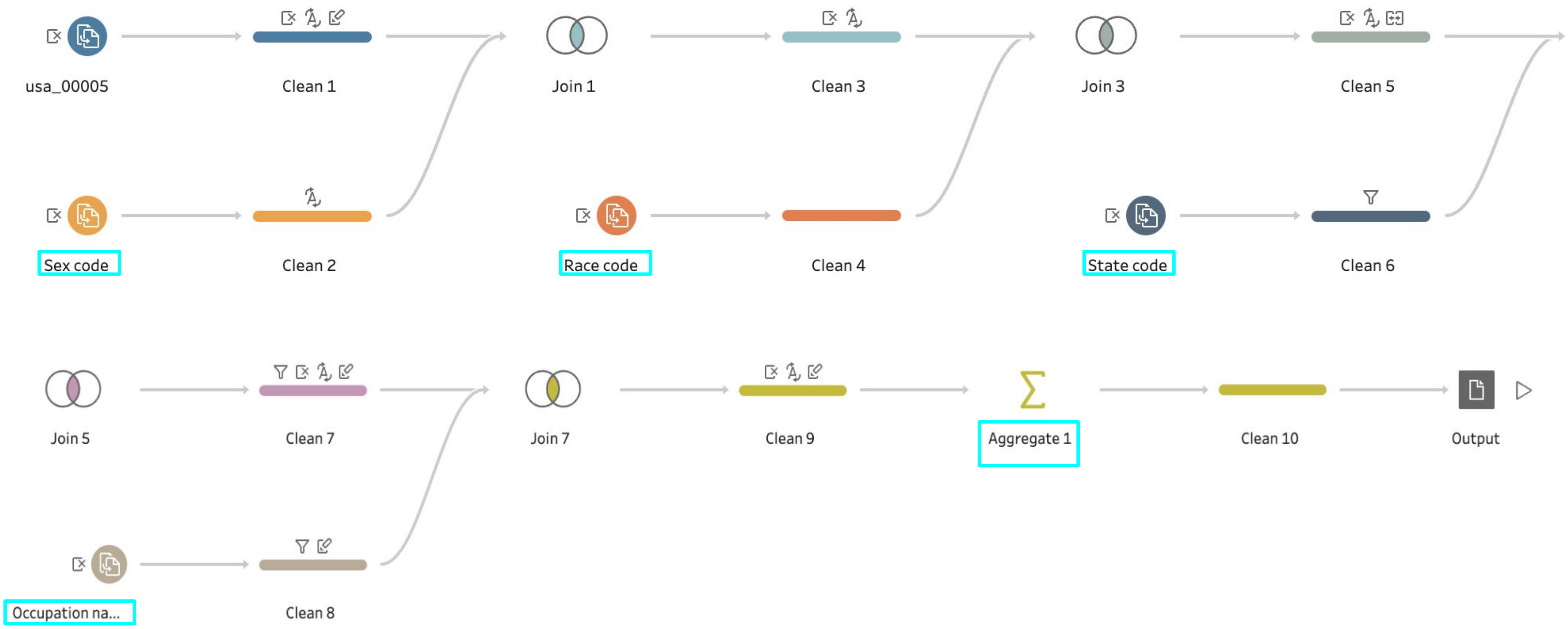


Wage threshold: 80000

- Filter the employment status: Employed

- Filter the wage range

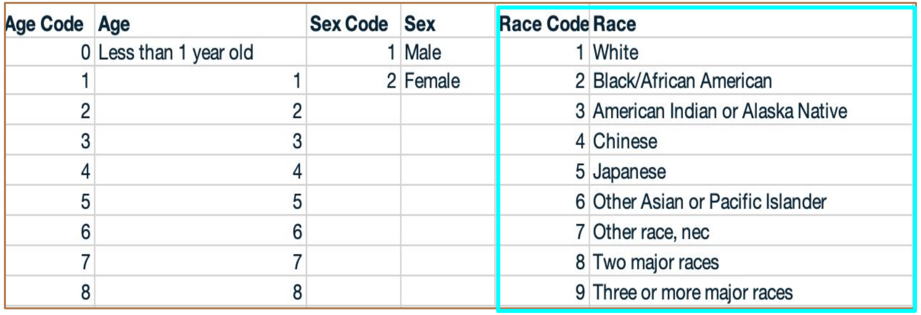
Data Cleaning and Preparation Operations - Diversity and Inclusion



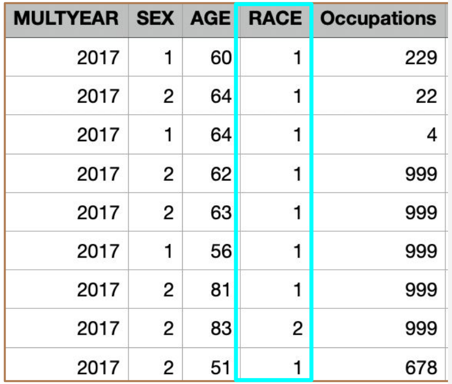
Data Cleaning and Preparation Operations - Diversity and Inclusion

1) Clean and Prepare the Data

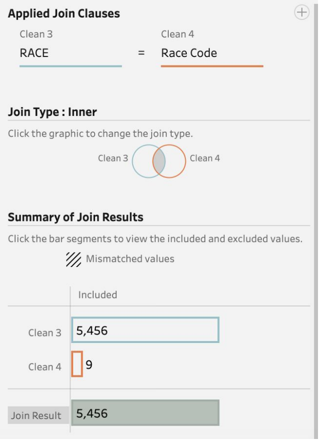
- Delete null values, filtering out irrelevant data, correcting data types
- Use **Join** to turn all the numbers to words



Age Code	Age	Sex Code	Sex	Race Code	Race
0	Less than 1 year old	1	Male	1	White
1	1	2	Female	2	Black/African American
2	2			3	American Indian or Alaska Native
3	3			4	Chinese
4	4			5	Japanese
5	5			6	Other Asian or Pacific Islander
6	6			7	Other race, nec
7	7			8	Two major races
8	8			9	Three or more major races



MULTYEAR	SEX	AGE	RACE	Occupations
2017	1	60	1	229
2017	2	64	1	22
2017	1	64	1	4
2017	2	62	1	999
2017	2	63	1	999
2017	1	56	1	999
2017	2	81	1	999
2017	2	83	2	999
2017	2	51	1	678



Applied Join Clauses

Clean 3		Clean 4
RACE	=	Race Code

Join Type : Inner

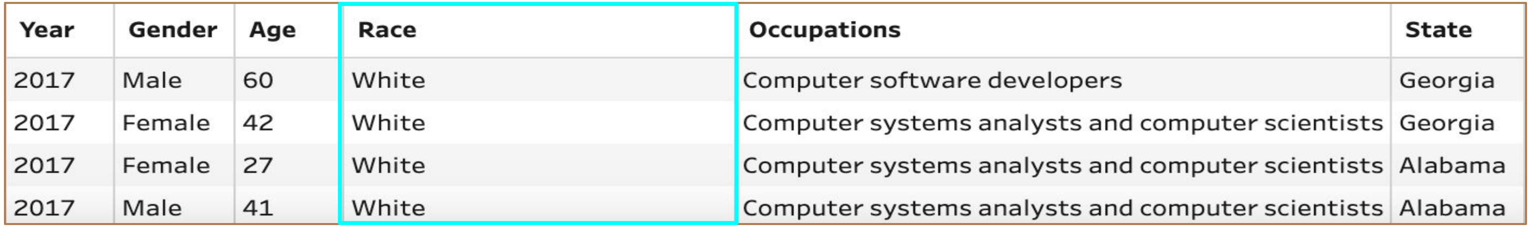
Click the graphic to change the join type.

Summary of Join Results

Click the bar segments to view the included and excluded values.

▨ Mismatched values

	Included
Clean 3	5,456
Clean 4	9
Join Result	5,456



Year	Gender	Age	Race	Occupations	State
2017	Male	60	White	Computer software developers	Georgia
2017	Female	42	White	Computer systems analysts and computer scientists	Georgia
2017	Female	27	White	Computer systems analysts and computer scientists	Alabama
2017	Male	41	White	Computer systems analysts and computer scientists	Alabama

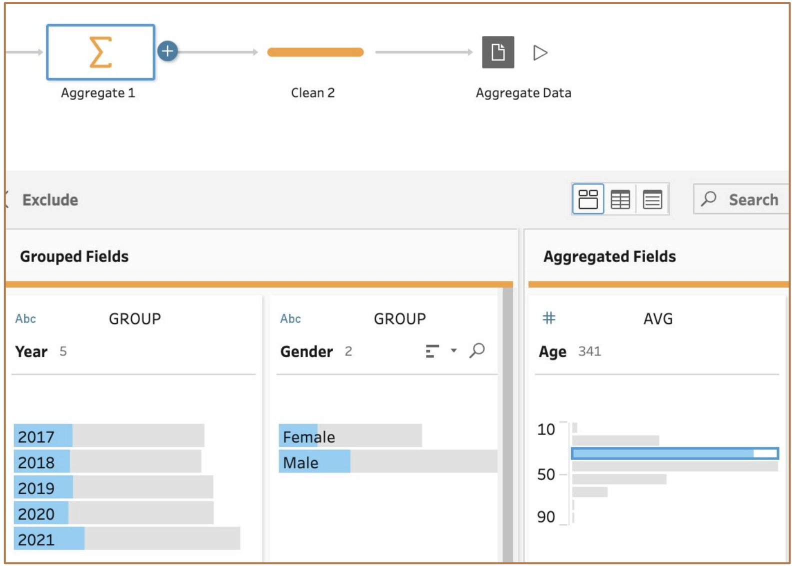
Data Cleaning and Preparation Operations - Diversity and Inclusion

2) Create an Aggregation Step

- Find the **average age** for each year

3) Final result

- Output data and run the flow



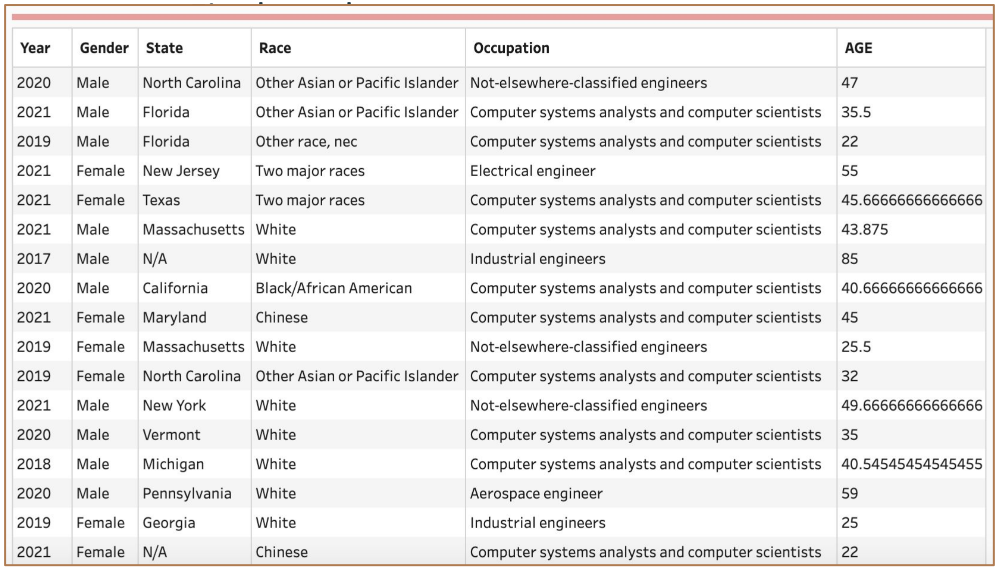
Grouped Fields

GROUP Year 5

GROUP Gender 2

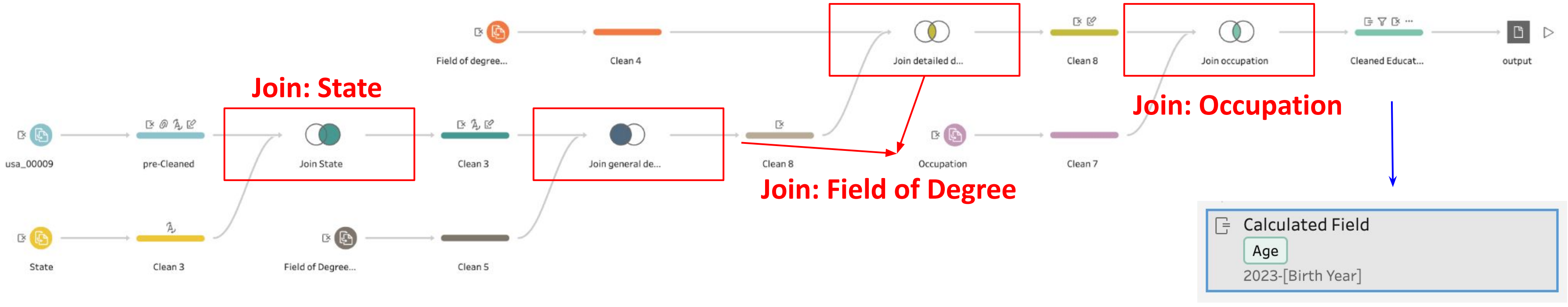
Aggregated Fields

AVG Age 341



Year	Gender	State	Race	Occupation	AGE
2020	Male	North Carolina	Other Asian or Pacific Islander	Not-elsewhere-classified engineers	47
2021	Male	Florida	Other Asian or Pacific Islander	Computer systems analysts and computer scientists	35.5
2019	Male	Florida	Other race, nec	Computer systems analysts and computer scientists	22
2021	Female	New Jersey	Two major races	Electrical engineer	55
2021	Female	Texas	Two major races	Computer systems analysts and computer scientists	45.66666666666666
2021	Male	Massachusetts	White	Computer systems analysts and computer scientists	43.875
2017	Male	N/A	White	Industrial engineers	85
2020	Male	California	Black/African American	Computer systems analysts and computer scientists	40.66666666666666
2021	Female	Maryland	Chinese	Computer systems analysts and computer scientists	45
2019	Female	Massachusetts	White	Not-elsewhere-classified engineers	25.5
2019	Female	North Carolina	Other Asian or Pacific Islander	Computer systems analysts and computer scientists	32
2021	Male	New York	White	Not-elsewhere-classified engineers	49.66666666666666
2020	Male	Vermont	White	Computer systems analysts and computer scientists	35
2018	Male	Michigan	White	Computer systems analysts and computer scientists	40.54545454545455
2020	Male	Pennsylvania	White	Aerospace engineer	59
2019	Female	Georgia	White	Industrial engineers	25
2021	Female	N/A	Chinese	Computer systems analysts and computer scientists	22

Data Cleaning and Preparation Operations - Education and Skill



Join: State

Join: Field of Degree

Join: Occupation

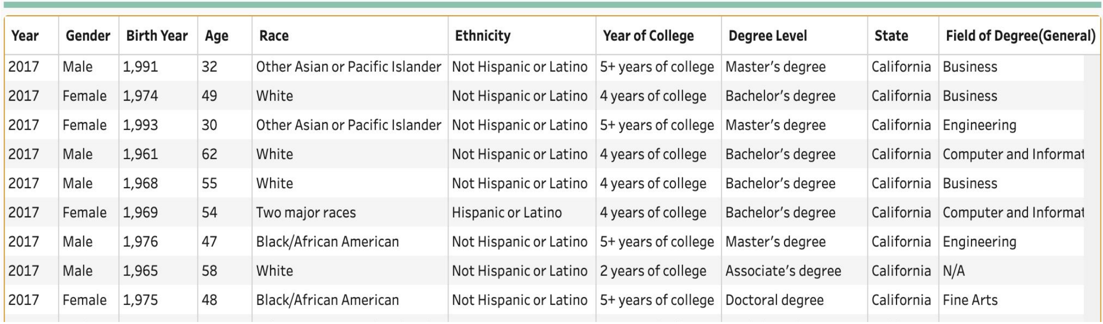
Calculated Field

Age

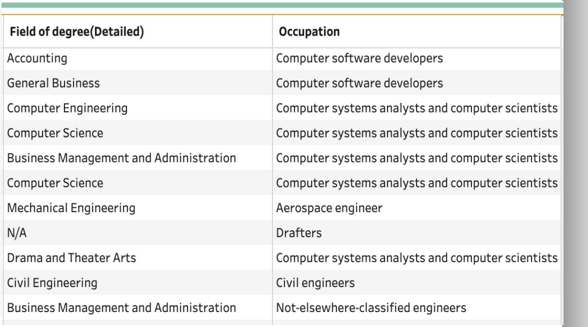
2023-[Birth Year]

Calculate: Age

Final Result

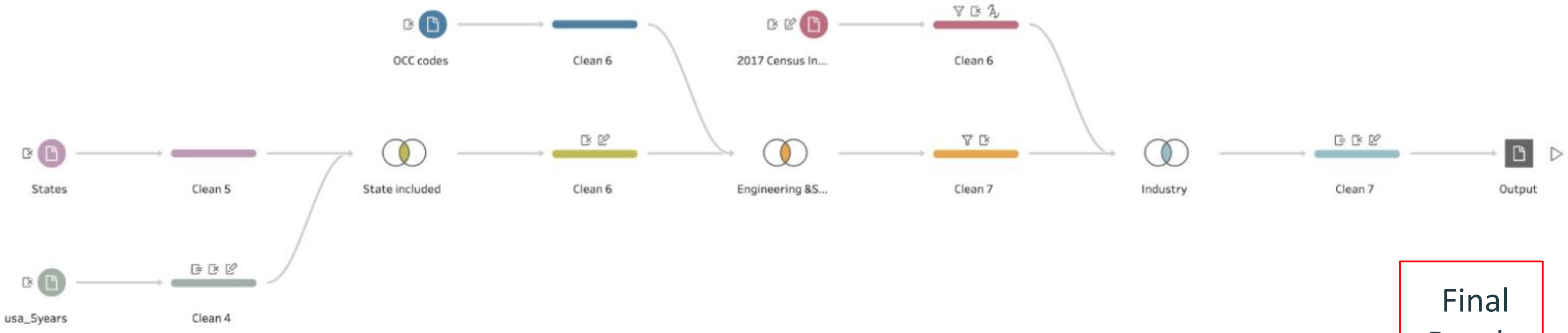


Year	Gender	Birth Year	Age	Race	Ethnicity	Year of College	Degree Level	State	Field of Degree(General)
2017	Male	1,991	32	Other Asian or Pacific Islander	Not Hispanic or Latino	5+ years of college	Master's degree	California	Business
2017	Female	1,974	49	White	Not Hispanic or Latino	4 years of college	Bachelor's degree	California	Business
2017	Female	1,993	30	Other Asian or Pacific Islander	Not Hispanic or Latino	5+ years of college	Master's degree	California	Engineering
2017	Male	1,961	62	White	Not Hispanic or Latino	4 years of college	Bachelor's degree	California	Computer and Informat
2017	Male	1,968	55	White	Not Hispanic or Latino	4 years of college	Bachelor's degree	California	Business
2017	Female	1,969	54	Two major races	Hispanic or Latino	4 years of college	Bachelor's degree	California	Computer and Informat
2017	Male	1,976	47	Black/African American	Not Hispanic or Latino	5+ years of college	Master's degree	California	Engineering
2017	Male	1,965	58	White	Not Hispanic or Latino	2 years of college	Associate's degree	California	N/A
2017	Female	1,975	48	Black/African American	Not Hispanic or Latino	5+ years of college	Doctoral degree	California	Fine Arts

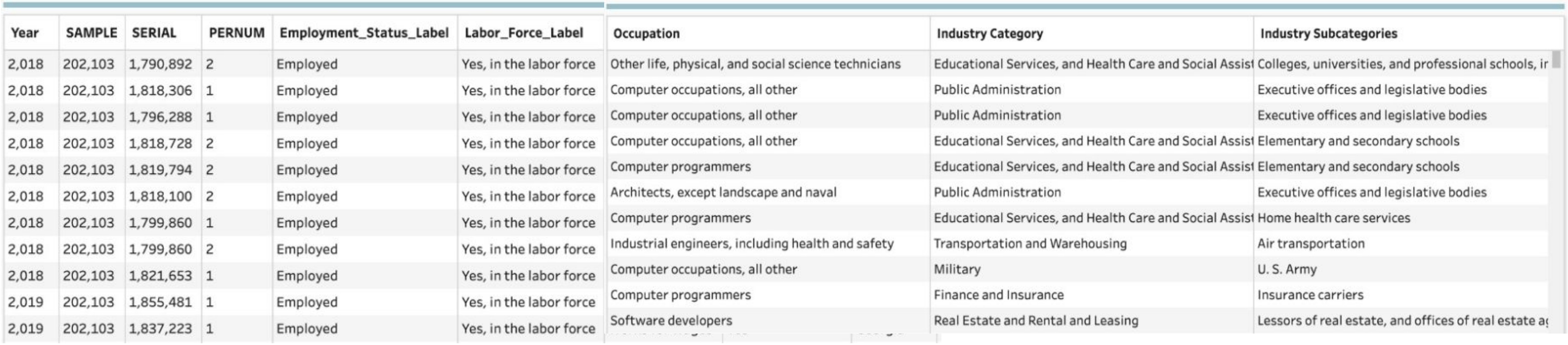


Field of degree(Detailed)	Occupation
Accounting	Computer software developers
General Business	Computer software developers
Computer Engineering	Computer systems analysts and computer scientists
Computer Science	Computer systems analysts and computer scientists
Business Management and Administration	Computer systems analysts and computer scientists
Computer Science	Computer systems analysts and computer scientists
Mechanical Engineering	Aerospace engineer
N/A	Drafters
Drama and Theater Arts	Computer systems analysts and computer scientists
Civil Engineering	Civil engineers
Business Management and Administration	Not-elsewhere-classified engineers

Data Cleaning and Preparation Operations - Industry and Sector



Final Result

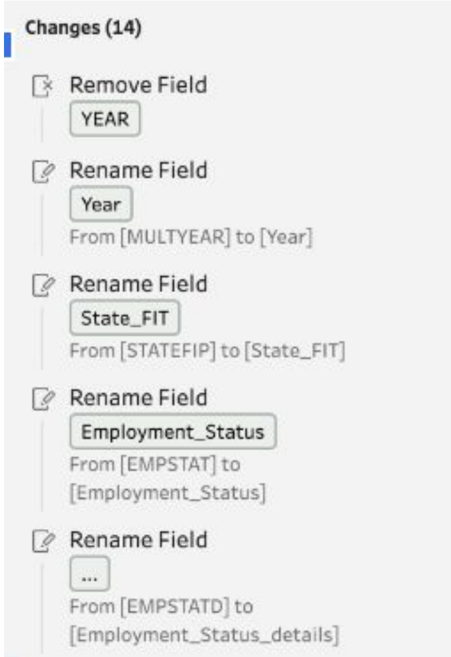


Year	SAMPLE	SERIAL	PERNUM	Employment_Status_Label	Labor_Force_Label	Occupation	Industry Category	Industry Subcategories
2,018	202,103	1,790,892	2	Employed	Yes, in the labor force	Other life, physical, and social science technicians	Educational Services, and Health Care and Social Assist	Colleges, universities, and professional schools, ir
2,018	202,103	1,818,306	1	Employed	Yes, in the labor force	Computer occupations, all other	Public Administration	Executive offices and legislative bodies
2,018	202,103	1,796,288	1	Employed	Yes, in the labor force	Computer occupations, all other	Public Administration	Executive offices and legislative bodies
2,018	202,103	1,818,728	2	Employed	Yes, in the labor force	Computer occupations, all other	Educational Services, and Health Care and Social Assist	Elementary and secondary schools
2,018	202,103	1,819,794	2	Employed	Yes, in the labor force	Computer programmers	Educational Services, and Health Care and Social Assist	Elementary and secondary schools
2,018	202,103	1,818,100	2	Employed	Yes, in the labor force	Architects, except landscape and naval	Public Administration	Executive offices and legislative bodies
2,018	202,103	1,799,860	1	Employed	Yes, in the labor force	Computer programmers	Educational Services, and Health Care and Social Assist	Home health care services
2,018	202,103	1,799,860	2	Employed	Yes, in the labor force	Industrial engineers, including health and safety	Transportation and Warehousing	Air transportation
2,018	202,103	1,821,653	1	Employed	Yes, in the labor force	Computer occupations, all other	Military	U. S. Army
2,019	202,103	1,855,481	1	Employed	Yes, in the labor force	Computer programmers	Finance and Insurance	Insurance carriers
2,019	202,103	1,837,223	1	Employed	Yes, in the labor force	Software developers	Real Estate and Rental and Leasing	Lessors of real estate, and offices of real estate a

Data Cleaning and Preparation Operations - Industry and Sector

1) Remove unnecessary columns

2) Sampling data



Changes (14)

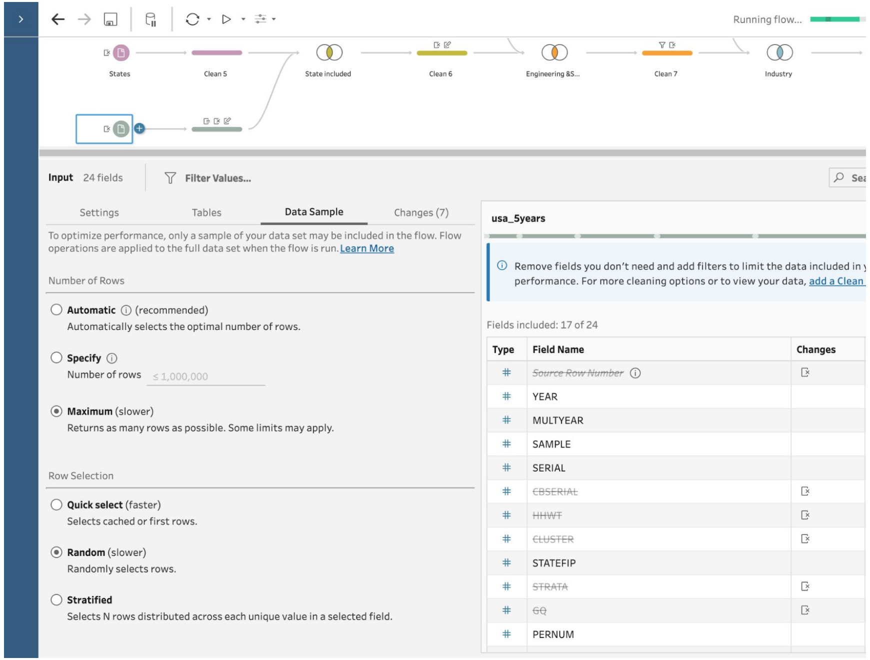
Remove Field
YEAR

Rename Field
Year
From [MULTYEAR] to [Year]

Rename Field
State_FIT
From [STATEFIP] to [State_FIT]

Rename Field
Employment_Status
From [EMPSTAT] to [Employment_Status]

Rename Field
...
From [EMPSTATD] to [Employment_Status_details]



Input 24 fields | Filter Values... | Sea

Settings | Tables | **Data Sample** | Changes (7)

To optimize performance, only a sample of your data set may be included in the flow. Flow operations are applied to the full data set when the flow is run. Learn More

Number of Rows

○ **Automatic** ⓘ (recommended)
Automatically selects the optimal number of rows.

○ **Specify** ⓘ
Number of rows ≤ 1,000,000

● **Maximum** (slower)
Returns as many rows as possible. Some limits may apply.

Row Selection

○ **Quick select** (faster)
Selects cached or first rows.

● **Random** (slower)
Randomly selects rows.

○ **Stratified**
Selects N rows distributed across each unique value in a selected field.

usa_5years

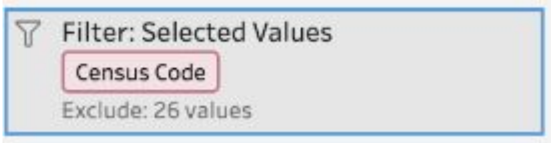
ⓘ Remove fields you don't need and add filters to limit the data included in y performance. For more cleaning options or to view your data, add a Clean

Fields included: 17 of 24

Type	Field Name	Changes
#	~~Source Row Number~~ ⓘ	☒
#	YEAR	
#	MULTYEAR	
#	SAMPLE	
#	SERIAL	
#	~~CBSERIAL~~	☒
#	~~HHWT~~	☒
#	~~CLUSTER~~	☒
#	STATEFIP	
#	~~STRATA~~	☒
#	~~GQ~~	☒
#	PERNUM	

Data Cleaning and Preparation Operations - Industry and Sector

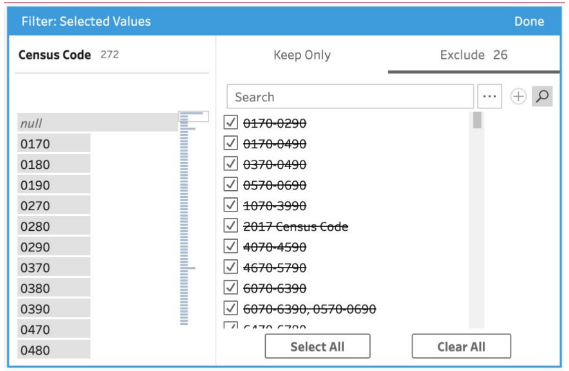
3) Filtering rows



Filter: Selected Values

Census Code

Exclude: 26 values

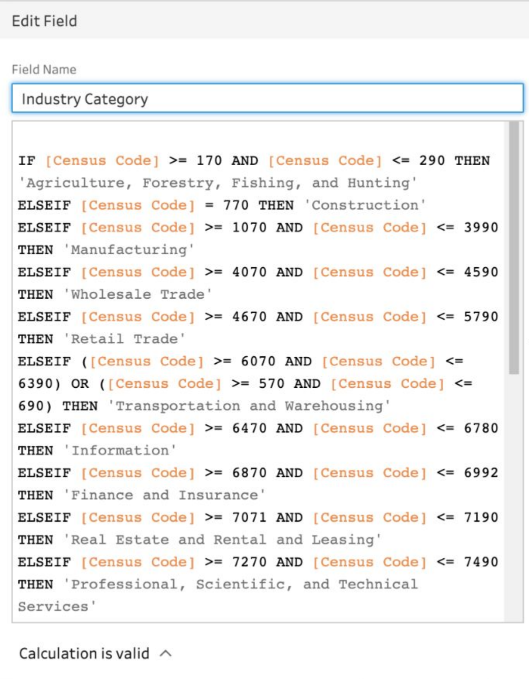


Census Code 272

null
0170
0180
0190
0270
0280
0290
0370
0380
0390
0470
0480

☑ ~~0170-0290~~
☑ ~~0170-0490~~
☑ ~~0370-0490~~
☑ ~~0570-0690~~
☑ ~~1070-3990~~
☑ ~~2017 Census Code~~
☑ ~~4070-4590~~
☑ ~~4670-5790~~
☑ ~~6070-6390~~
☑ ~~6070-6390, 0570-0690~~

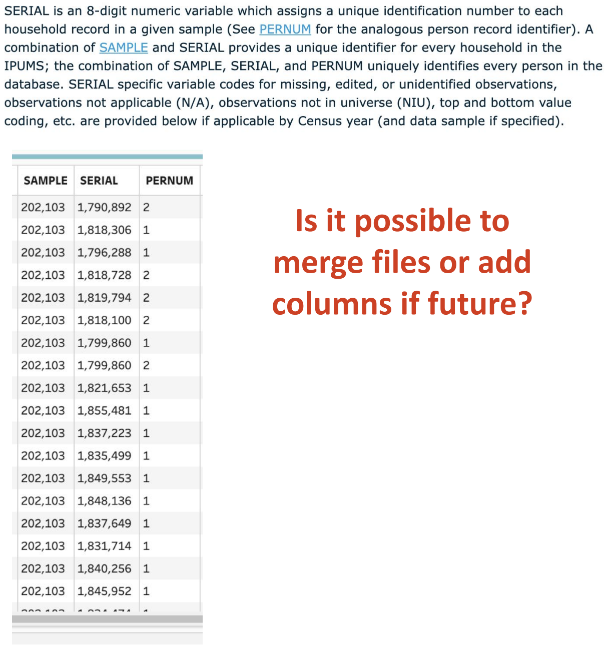
4) Create a Calculated field "Industry Category" to replace codes with labels



Field Name: Industry Category

```
IF [Census Code] >= 170 AND [Census Code] <= 290 THEN
'Agriculture, Forestry, Fishing, and Hunting'
ELSEIF [Census Code] = 770 THEN 'Construction'
ELSEIF [Census Code] >= 1070 AND [Census Code] <= 3990
THEN 'Manufacturing'
ELSEIF [Census Code] >= 4070 AND [Census Code] <= 4590
THEN 'Wholesale Trade'
ELSEIF [Census Code] >= 4670 AND [Census Code] <= 5790
THEN 'Retail Trade'
ELSEIF ([Census Code] >= 6070 AND [Census Code] <=
6390) OR ([Census Code] >= 570 AND [Census Code] <=
690) THEN 'Transportation and Warehousing'
ELSEIF [Census Code] >= 6470 AND [Census Code] <= 6780
THEN 'Information'
ELSEIF [Census Code] >= 6870 AND [Census Code] <= 6992
THEN 'Finance and Insurance'
ELSEIF [Census Code] >= 7071 AND [Census Code] <= 7190
THEN 'Real Estate and Rental and Leasing'
ELSEIF [Census Code] >= 7270 AND [Census Code] <= 7490
THEN 'Professional, Scientific, and Technical
Services'
```

Calculation is valid



SERIAL is an 8-digit numeric variable which assigns a unique identification number to each household record in a given sample (See PERNUM for the analogous person record identifier). A combination of SAMPLE and SERIAL provides a unique identifier for every household in the IPUMS; the combination of SAMPLE, SERIAL, and PERNUM uniquely identifies every person in the database. SERIAL specific variable codes for missing, edited, or unidentified observations, observations not applicable (N/A), observations not in universe (NIU), top and bottom value coding, etc. are provided below if applicable by Census year (and data sample if specified).

SAMPLE	SERIAL	PERNUM
202,103	1,790,892	2
202,103	1,818,306	1
202,103	1,796,288	1
202,103	1,818,728	2
202,103	1,819,794	2
202,103	1,818,100	2
202,103	1,799,860	1
202,103	1,799,860	2
202,103	1,821,653	1
202,103	1,855,481	1
202,103	1,837,223	1
202,103	1,835,499	1
202,103	1,849,553	1
202,103	1,848,136	1
202,103	1,837,649	1
202,103	1,831,714	1
202,103	1,840,256	1
202,103	1,845,952	1

Is it possible to merge files or add columns if future?

Challenges

1. Data Selection
    a. Inflexible government data that are cleaned and well-structured
    b. Detailed data needed for in-depth analysis

2. Data Cleaning
    a. Large data- Filtering and sampling
    b. Data with confusing code
        i. Join
        ii. define parameters
    c. Files with different sample size

Insights from data

Employment

- Trend Over Time
- Geographical Analysis
- Wage Distribution

Industry

- Occupation Distribution
- Demands Over Time

Diversity

- Gender & Age
- Race & Ethnicity

Education

- Common majors
- Education Levels

THANK YOU :)