

## Assignment 02

- 1) Read the adult.csv file available in the **data** folder on the KNIME Hub. The data are provided by the [UCI Machine Learning Repository](#).
- 2) Calculate the average age and count for each one of the 4 groups defined by sex and income values
- 3) Join the two aggregated values to the original table

### Step 1: Read the adult.csv file

The screenshot shows a KNIME workflow titled "Local - Assignment 2". The workflow consists of three main nodes: a "CSV Reader" node, a "GroupBy" node, and a "Joiner" node. The "CSV Reader" node is connected to the "GroupBy" node, which is then connected to the "Joiner" node. A tooltip for the "GroupBy" node indicates it is used for grouping data. The "Joiner" node has a tooltip stating "This node dialog is not supported here." Below the nodes is a preview table titled "1: File Table" showing the first 10 rows of the dataset. The columns listed are: #, RowID, age, workclass, fnlwgt, education, education., marital-status, occupation, relations., race, and sex. The data rows show various characteristics of the adult population, such as age ranging from 16 to 90, and marital status including Never-married, Married-civ-spouse, Divorced, etc.

#	RowID	age	workclass	fnlwgt	education	education.	marital-status	occupation	relations.	race	sex
1	Row0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male
2	Row1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male
3	Row2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaner	Not-in-family	White	Male
4	Row3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaner	Husband	Black	Male
5	Row4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female
6	Row5	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female
7	Row6	49	Private	160187	9th	5	Married-spouse	Other-service	Not-in-family	Black	Female
8	Row7	52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male
9	Row8	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female
10	Row9	42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male

## Power BI and KNIME

**Step 2:** Calculate the average age and count for each one of the 4 groups defined by sex and income values

This node dialog is not supported here.

Open dialog

#	RowID	Sex	income	Mean(age)	Count(age)
1	Row0	Female	<=50K	36.211	9592
2	Row1	Female	>50K	42.126	1179
3	Row2	Male	<=50K	37.147	15128
4	Row3	Male	>50K	44.626	6662

**Step 3:** Join the two aggregated values to the original value

#	RowID	sex	capital-g...	capital-lo...	hours-per...	native-co...	income	sex (Right)	income (...	Mean(age)	Count(a...
1	Male	2174	0	40	United-States	<=50K	Female	<=50K	36.211	9592	
2	Male	0	0	13	United-States	<=50K	Female	>50K	42.126	1179	
3	Male	0	0	40	United-States	<=50K	Male	<=50K	37.147	15128	
4	Male	0	0	40	United-States	<=50K	Male	>50K	44.626	6662	