

## 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 displays the KNIME software interface. On the left, the 'CSV Reader' node is selected, showing its configuration panel. The 'Autodetect format' button is highlighted. The 'Advanced Settings' tab is active, showing options for 'Limit data rows scanned' and 'Support changing file schemas'. A note indicates that if the node can't read the file, the 'File Reader' node should be used. The 'Parallel reading' section is also visible.

The main workspace shows a workflow diagram with the following nodes: 'CSV Reader' (labeled 'Add comment'), 'GroupBy', and 'Joiner'. The 'CSV Reader' node is connected to the 'GroupBy' node, which is then connected to the 'Joiner' node.

On the right, the 'CSV Reader' node dialog is open, displaying the message: 'This node dialog is not supported here. Open dialog'.

Below the workflow, the 'Table' view shows the data loaded from the CSV file. The table has 15 columns and 32561 rows. The columns are: #, RowID, age, workclass, fnlwgt, education, education..., marital-st..., occupation, relations..., race, and st. The first 10 rows of data are displayed:

#	RowID	age	workclass	fnlwgt	education	education...	marital-st...	occupation	relations...	race	st
1	Row0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male
2	Row1	50	Self-emp-not-in	83311	Bachelors	13	Married-civ-spo	Exec-manageris	Husband	White	Male
3	Row2	38	Private	215646	HS-grad	9	Divorced	Handlers-clean	Not-in-family	White	Male
4	Row3	53	Private	234721	11th	7	Married-civ-spo	Handlers-clean	Husband	Black	Male
5	Row4	28	Private	338409	Bachelors	13	Married-civ-spo	Prof-specialty	Wife	Black	Female
6	Row5	37	Private	284582	Masters	14	Married-civ-spo	Exec-manageris	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-in	209642	HS-grad	9	Married-civ-spo	Exec-manageris	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-spo	Exec-manageris	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

The screenshot shows the KNIME software interface. On the left, the 'Info' panel for the 'GroupBy' node is visible, explaining its function: it groups rows of a table by unique values in selected group columns and aggregates the remaining columns. The main workspace shows a workflow starting with a 'CSV Reader' node, followed by a 'GroupBy' node, and then a 'Joiner' node. The 'GroupBy' node's configuration dialog is open on the right, showing a message: 'This node dialog is not supported here. Open dialog'. Below the workflow, the output of the 'GroupBy' node is displayed as a table with 4 rows and 6 columns.

#	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

The screenshot shows the KNIME software interface with the 'Joiner' node configuration dialog open. The dialog shows 'Matching Criteria' set to 'All of the following' and 'Compare values in join columns by' set to 'Value and type'. The main workspace shows a workflow starting with a 'CSV Reader' node, followed by a 'GroupBy' node, and then a 'Joiner' node. The 'Joiner' node's configuration dialog is open on the right, showing the matching criteria. Below the workflow, the output of the 'Joiner' node is displayed as a table with 4 rows and 19 columns.

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