

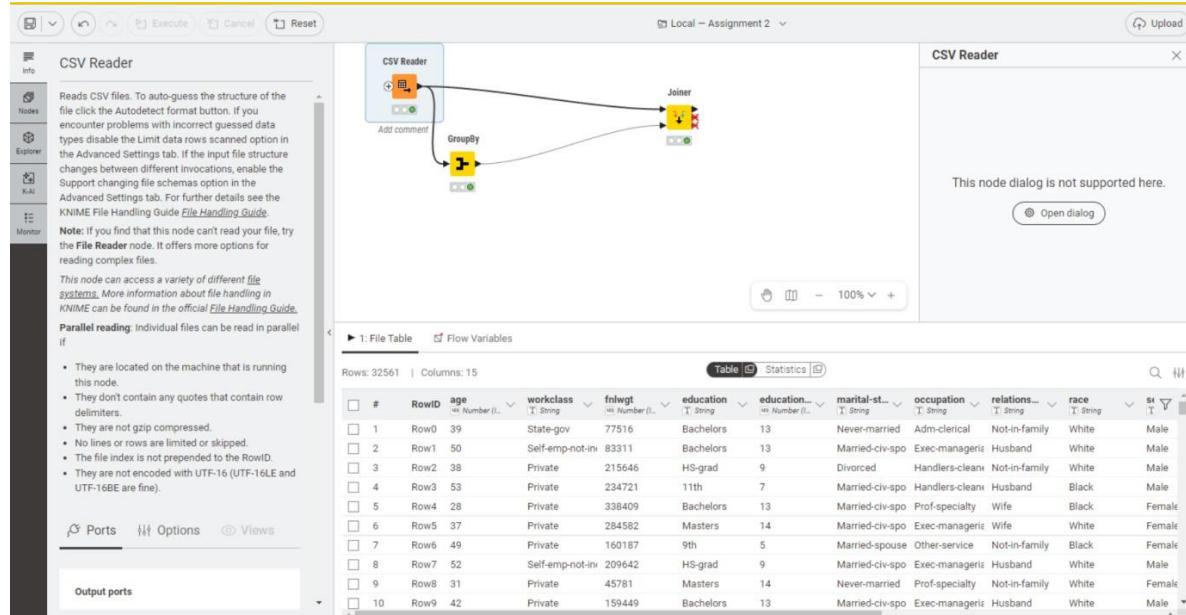
# Assignment 2

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



## 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 interface with the following components:

- GroupBy Node Info:** Describes the node's function: "Groups the rows of a table by the unique values in the selected group columns. A row is created for each unique set of values of the selected group column. The remaining columns are aggregated based on the specified aggregation settings. The output table contains one row for each unique value combination of the selected group columns." It also details aggregation methods for other columns.
- Process Flow:** A CSV Reader node feeds into a GroupBy node, which then feeds into a Joiner node.
- Joiner Node Info:** States that the node dialog is not supported here. An "Open dialog" button is present.
- Table View:** Shows the aggregated data with 4 rows and 4 columns. The columns are RowID, #, RowID, sex, income, Mean(age), and Count(age).

#	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 interface with the following components:

- Joiner Node Info:** Describes the node's function: "Combines two tables similar to a join in a database. It combines each row from the top input port with each row from the bottom input port that has identical values in selected columns. Rows that remain unmatched can also be output." It includes a link to KNIME E-Learning Course documentation.
- Process Flow:** A CSV Reader node feeds into a GroupBy node, which then feeds into a Joiner node.
- Joiner Node Info:** Shows matching criteria set to "All of the following".
- Table View:** Shows the joined data with 4 rows and 19 columns. The columns include sex, capital-g..., capital-lo..., hours-per..., native-co..., income, sex (Right), income (~), Mean(age), and Count(age).

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