

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

1) Read the adult.csv file

The screenshot shows the KNIME interface with the CSV Reader node selected. The left pane displays the node's documentation, including a note about parallel reading. The main workspace shows a flow diagram with a CSV Reader node connected to a Joiner node. The right pane shows the CSV Reader node dialog, which is not supported here. Below the flow diagram, the data table is displayed with 15 columns and 32561 rows.

#	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-managerial	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-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-in	209642	HS-grad	9	Married-civ-spo	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-spo	Exec-managerial	Husband	White	Male

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 GroupBy node selected. The left pane displays the node's documentation, including a note about the 'Manual Aggregation' tab. The main workspace shows a flow diagram with a CSV Reader node connected to a Joiner node. The right pane shows the GroupBy node dialog, which is not supported here. Below the flow diagram, the aggregated data table is displayed with 4 columns and 4 rows.

#	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

3) Join the two aggregated values to the original value

The screenshot shows the KNIME software interface. On the left, the 'Joiner' node is selected, displaying its description and input/output ports. The main workspace shows a workflow with a 'CSV Reader' node connected to a 'GroupBy' node, which is then connected to a 'Joiner' node. The 'Joiner' node's configuration panel on the right shows 'Matching Criteria' set to 'All of the following' and 'Compare values in join columns by' set to 'Value and type'. Below the configuration, the output table is displayed with 4 rows and 19 columns.

id	sex	capital-g...	capital-lo...	hours-per...	native-co...	income	sex (Right)	income (...	Mean(age)	Count(a...
ing	T: String	ve Number (L...	ve Number (L...	ve Number (L...	T: String	T: String	T: String	T: String	ve Number (L...	ve Number (L...
te	Male	2174	0	40	United-States	<=50K	Female	<=50K	36.211	9592
te	Male	0	0	13	United-States	<=50K	Female	>50K	42.126	1179
te	Male	0	0	40	United-States	<=50K	Male	<=50K	37.147	15128
k	Male	0	0	40	United-States	<=50K	Male	>50K	44.626	6662