

K.R. MANGALAM UNIVERSITY, GURUGRAM-122103

SCHOOL OF ENGENIERRING AND TECHNOLOGY

ASSIGNMENT 2

Data Analysis with Power BI s KNIME

ETMMML174



Department: SOET	Session: 2025-27
Program: MCA (AI & ML)	Semester: 1
Course Code: ETMMML174	College Roll no: 2501940069
Course Name: Data Analysis with Power BI & KNIME	
Submitted by: Tanya Bhatia	Faculty: Mr. Mohammad Ajaz

- 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 configured. The left pane displays the node's settings, including a list of supported file formats and a warning about parallel reading. The main workspace shows a workflow with a CSV Reader node connected to a Joiner node. The right pane shows the CSV Reader node's dialog, which is not supported in this view. Below the workflow, a data preview table is visible, showing the first 10 rows of the adult.csv file.

#	RowID	Age	workclass	fnlwgt	education	education...	marital st...	occupation	relations...	race	sex
1	Row1	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male
2	Row1	50	Self-emp-not-inc	93311	Bachelors	13	Married-civ-sp	Exec-managerial	Husband	White	Male
3	Row2	36	Private	319946	HS-grad	9	Divorced	Handlers-cleaner	Not-in-family	White	Male
4	Row2	53	Private	234721	11th	7	Married-civ-sp	Handlers-cleaner	Husband	Black	Male
5	Row4	38	Private	238409	Bachelors	13	Married-civ-sp	Prof-specialty	Wife	Black	Female
6	Row5	37	Private	254362	Masters	14	Married-civ-sp	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	209942	HS-grad	9	Married-civ-sp	Exec-managerial	Husband	White	Male
9	Row8	31	Private	45791	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female
10	Row9	42	Private	153444	Bachelors	13	Married-civ-sp	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

GroupBy

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 columns. 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.

The columns to aggregate can be either defined by selecting the columns directly, by name based on a search pattern or based on the data type. Input columns are handled in this order and only considered once e.g. columns that are added directly on the 'Manual Aggregation' tab are ignored even if their name matches a search pattern on the 'Pattern Based Aggregation' tab or their type matches a defined type on the 'Type Based Aggregation' tab. The same holds for columns that are added based on a search pattern. They are ignored even if they match a criterion that has been defined in the 'Type Based Aggregation' tab.

The 'Manual Aggregation' tab allows you to change the aggregation method of more than one column. In order to do so select the columns to change, open the context menu with a right mouse click and select the aggregation method to use.

In the 'Pattern Based Aggregation' tab you can assign aggregation methods to columns based on a search pattern. The pattern can be either a string with wildcards or a regular expression. Columns where the name matches the pattern but where the data type is not compatible with the selected aggregation method are ignored. Only columns that have not been selected as group columns or that have not been selected as aggregation columns on the 'Manual Aggregation' tab are considered.

1. Group table | **Flow variables**

	RowID	sex	income	Mean(age)	Count(age)
1	Row0	Female	<=50K	36.211	8992
2	Row1	Female	>50K	42.126	1179
3	Row2	Male	<=50K	37.147	15128
4	Row3	Male	>50K	44.826	6662

3) Join the two aggregated values to the original value

Joiner

This node combines two tables similar to a join in a database. It combines each row from the top input part with each row from the bottom input part that has identical values in selected columns. Rows that remain unmatched can also be output.

External resources:

→ [KDD99 Learning Course](#): Join inner join, right outer join, left outer join, full outer join

Parts | **Options** | **Visuals**

Input parts

- Type: Left table
- Type: Right table
- Type: Right input table

Output parts

- Type: Join result

Joiner

Matching Criteria

Match

All of the following | any of the following

Add matching criterion

Compare values in join columns by

Value and type

Discard | Apply and Execute | Apply

1. Join result | **2. Left unmatched rows** | **3. Right unmatched rows** | **Flow variables**

rowid	sex	capital-g.	capital-la.	hours-per.	native-us.	income	sex (Right)	income (.	Mean(age)	Count(a.
0	Male	2176	0	60	United States	<=50K	Female	<=50K	36.211	8992
1	Male	0	0	73	United States	<=50K	Female	<=50K	42.126	1179
2	Male	0	0	40	United States	<=50K	Male	<=50K	37.147	15128
3	Male	0	0	40	United States	<=50K	Male	<=50K	44.826	6662