

# **K.R. MANGALAM UNIVERSITY, GURUGRAM-122103**

## **SCHOOL OF ENGENIERRING AND TECHNOLOGY**

### **ASSIGNMENT 2**

#### **Data Analysis with Power BI & KNIME**

**ETMMML174**

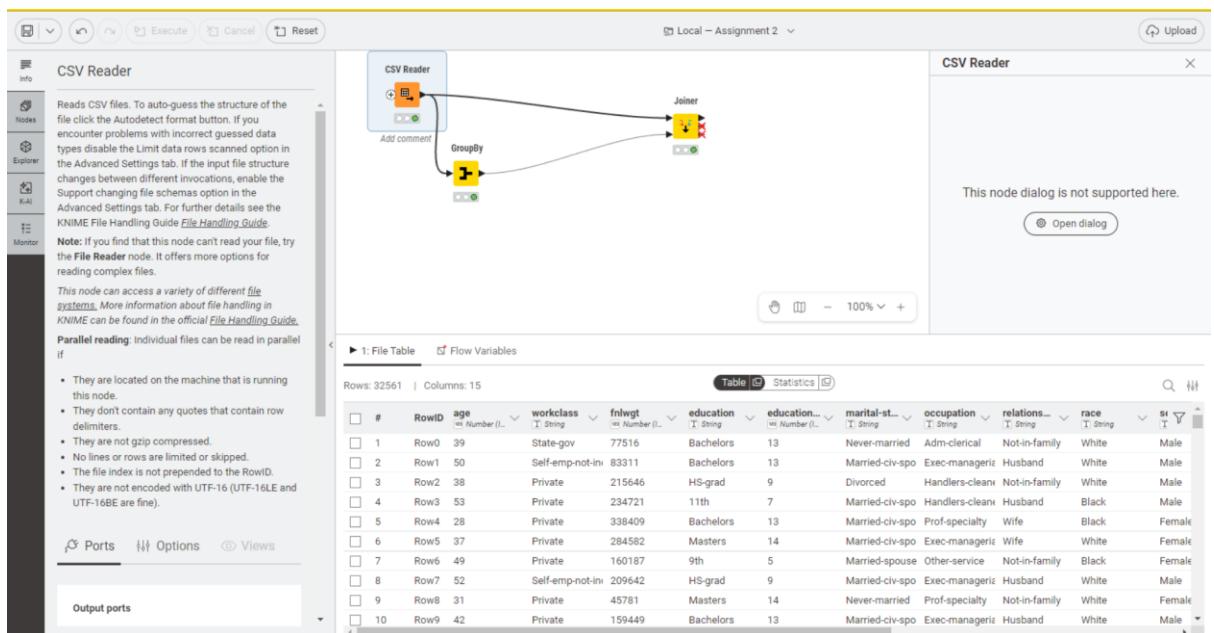


<b>Department:</b> SOET	<b>Session:</b> 2025-27
<b>Program:</b> MCA (AI & ML)	<b>Semester:</b> 1
<b>Course Code:</b> ETMMML174	<b>College Roll no:</b> 2501940074
<b>Course Name:</b> Data Analysis with Power BI & KNIME	
<b>Submitted by:</b> Khushi	<b>Faculty:</b> Mr. Mohammad Ajaz

## 2 Power BI 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

### 1) Read the adult.csv file



- 2) Calculate the average age and count for each one of the 4 groups defined by sex and income values

### 3 Power BI Assignment 2

The screenshot shows the KNIME interface with the following components:

- GroupBy Node Dialog:** On the left, the "Info" tab is selected, providing detailed documentation on how the GroupBy node groups rows by unique values in specified columns and aggregates other columns based on aggregation settings.
- Process Flow:** A flow diagram where a "CSV Reader" node connects to a "Joiner" node. The "Joiner" node has two input ports: one from a "GroupBy" node and another from the "CSV Reader".
- Result View:** On the right, a table titled "Group table" displays the results of the join operation. The table has four rows and five 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

### 3) Join the two aggregated values to the original value

The screenshot shows the KNIME interface with the following components:

- Joiner Node Dialog:** On the left, the "Info" tab is selected, explaining that the Joiner node combines two tables similar to a join in a database, matching rows from both inputs based on specified criteria.
- Process Flow:** A flow diagram where a "CSV Reader" node connects to a "Joiner" node. The "Joiner" node has two input ports: one from a "GroupBy" node and another from the "CSV Reader".
- Result View:** On the right, a table titled "Join result" displays the joined dataset. The table has four rows and 19 columns, combining the original data with the aggregated mean age and count of age groups:
 

	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