

# **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> 2501940069
<b>Course Name:</b> Data Analysis with Power BI & KNIME	
<b>Submitted by:</b> Tanya Bhatia	<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

The screenshot shows a KNIME workflow titled "Local - Assignment 2". It consists of three main nodes: a "CSV Reader" node on the left, a "CCW" node in the center, and another "CSV Reader" node on the right. A connection from the left "CSV Reader" to the "CCW" node is labeled "empty". A connection from the "CCW" node to the right "CSV Reader" is labeled "CCW". The right "CSV Reader" has a tooltip: "This node dialog is not supported here." Below the nodes is a table view showing the first 10 rows of the dataset.

#	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex
1	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male
2	50	Self-emp-not-inc	33211	Bachelors	13	Married-spouse-absent	Exe-manager	Husband	White	Male
3	38	Private	316446	HS-grad	9	Divorced	Handlers-cleaner	Not-in-family	White	Male
4	42	Private	234721	HS-grad	7	Married-spouse-absent	Handlers-cleaner	Husband	Black	Male
5	38	Private	33489	Bachelors	13	Married-spouse-absent	Prof-specialty	Wife	Black	Female
6	53	Private	28482	Masters	14	Married-spouse-absent	Exe-manager	Wife	White	Female
7	40	Private	160187	HS-grad	8	Married-spouse-absent	Other-service	Not-in-family	Black	Female
8	54	Self-emp-not-inc	309442	HS-grad	9	Married-spouse-absent	Exe-manager	Husband	White	Male
9	46	Private	40781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female
10	40	Private	151444	Bachelors	13	Married-spouse-absent	Exe-manager	Husband	White	Male

### 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 two main components:

- GroupBy Node:** A yellow node with three input ports (left, right, bottom) and one output port (join). It has a configuration panel with detailed documentation about grouping rows by selected group columns and aggregating remaining columns.
- Joiner Node:** A yellow node with two input ports (left, right) and one output port (join). It has a configuration panel for matching criteria.

The data table below the nodes shows the results of the aggregation and joining process:

	Sex	Age	Income	MeanAge
1	Male	21-30	>50K	36.211
2	Female	21-30	<=50K	40.125
3	Male	31-40	<=50K	42.176
4	Male	31-40	>50K	37.147
5	Male	41-50	<=50K	44.428
6	Male	41-50	>50K	46.612

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

The screenshot shows the KNIME interface with a Joiner node and a data table:

- Joiner Node:** A yellow node with two input ports (left, right) and one output port (join). It has a configuration panel for matching criteria.
- Data Table:** A table titled "1. Join result" showing the final joined dataset.

The data table shows the original data with additional columns for mean age and count per sex and age group, joined with the aggregated data from the previous step:

	Sex	Age	Capital-g...	Capital-l...	Avg Age...	Native-ct...	Income_<=...	Income_>=...	MeanAge[...]	CountP[...]
1	Male	21-30	2174	0	41	United States	<=50K	>50K	36.211	9982
2	Male	21-30	0	0	33	United States	<=50K	<=50K	40.125	1179
3	Male	31-40	0	0	41	United States	<=50K	<=50K	42.176	18128
4	Male	31-40	0	0	40	United States	<=50K	>50K	37.147	44428