

K.R. MANGALAM UNIVERSITY, GURUGRAM-122103

SCHOOL OF ENGENIERING AND TECHNOLOGY

ASSIGNMENT 1

Data Analysis with Power BI & KNIME

ETMMML174



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

1 Power BI Assignment 1

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 count and average age of women with income >50K

3) Calculate the averages of all numerical columns for each one of the 4 groups defined by sex and income values

4) Calculate

- the number of missing values in the occupation column
- the number of non-missing rows in the occupation column
- the number of rows in the occupation column
- the number of rows in the marital-status column

Notice that the last two aggregations should provide the same numbers!

1) Read the adult.csv file

#	RowID	age	workclass	fnwgt	education	education...	marital-st...	occupation	relations...	race	sex	capital-g...	capital-lo...	hours-per...
1	Row0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40
2	Row1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13
3	Row2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaner	Not-in-family	White	Male	0	0	40
4	Row3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaner	Husband	Black	Male	0	0	40
5	Row4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40
6	Row5	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40
7	Row6	49	Private	160187	9th	5	Married-spouse	Other-service	Not-in-family	Black	Female	0	0	16
8	Row7	52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	45
9	Row8	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084	0	50
10	Row9	42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	5178	0	40

2) A) Filter Female and Income >50k using Row Filter

2 Power BI Assignment 1

The screenshot shows the KNIME Data Wrangler interface. A workflow is built with the following nodes:

- CSV Reader**: Loads the adult dataset.
- Row Filter**: Filters rows where sex is Female.
- GroupBy**: Groups by income level (`fnlgt`).
- GroupBy**: Groups by education level (`education`).
- GroupBy**: Groups by marital status (`marital-status`).

The resulting table has 1179 rows and 15 columns. The data includes columns like RowID, age, workclass, fnlgt, education, marital-status, occupation, relations, race, sex, capital-gain, capital-loss, and hours-per-week.

Row Filter configuration:

- Match row if matched by: All criteria
- Filter column: sex
- Operator: Equals

- 2) B) Calculate the Count and Average age of women with income >50k

The screenshot shows the KNIME Data Wrangler interface. A workflow is built with the following nodes:

- CSV Reader**: Loads the adult dataset.
- Row Filter**: Filters rows where sex is Female.
- GroupBy**: Groups by income level (`fnlgt`).
- GroupBy**: Groups by education level (`education`).
- GroupBy**: Groups by marital status (`marital-status`).

The resulting table has 1 row and 2 columns, labeled "Group table". The data includes columns RowID and Count*(age). The count is 1179 and the mean age is 42.126.

GroupBy configuration (not supported here):

- This node dialog is not supported here.

- 3) Calculate the averages of all numerical columns for each one of the 4 groups defined by sex and income value

3 Power BI Assignment 1

The screenshot shows the KNIME Data Wrangler interface with a workflow titled "Local - tutorial". The workflow consists of a "CSV Reader" node connected to three "GroupBy" nodes. The first "GroupBy" node has a "Row Filter" node preceding it. The second and third "GroupBy" nodes have arrows pointing back to the first one. The output of the first "GroupBy" node is a "Table" with the following data:

#	RowID	sex	income	Mean(age)	Mean(education)	Mean(capital-g)	Mean(capital-L)	Mean(hours-pe)
1	Row0	Female	<=50K	36.211	9.82	121.986	47.364	35.917
2	Row1	Female	>50K	42.126	11.787	4,200.389	173.649	40.427
3	Row2	Male	<=50K	37.147	9.452	165.724	56.807	40.694
4	Row3	Male	>50K	44.626	11.581	3,971.766	198.78	46.366

4) Calculate:

- the number of **missing values** in the *occupation* column
- the number of **non-missing rows** in the *occupation* column
- the **number of rows** in the *occupation* column
- the **number of rows** in the *marital-status* column

The screenshot shows the KNIME Data Wrangler interface with a workflow titled "Local - tutorial". The workflow consists of a "CSV Reader" node connected to three "GroupBy" nodes. The first "GroupBy" node has a "Row Filter" node preceding it. The second and third "GroupBy" nodes have arrows pointing back to the first one. The output of the first "GroupBy" node is a "Table" with the following data:

#	RowID	Missing value count(occupation)	Count(occupation)	Count(marital-status)
1	Row0	0	32561	32561