Power BI and KNIME 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

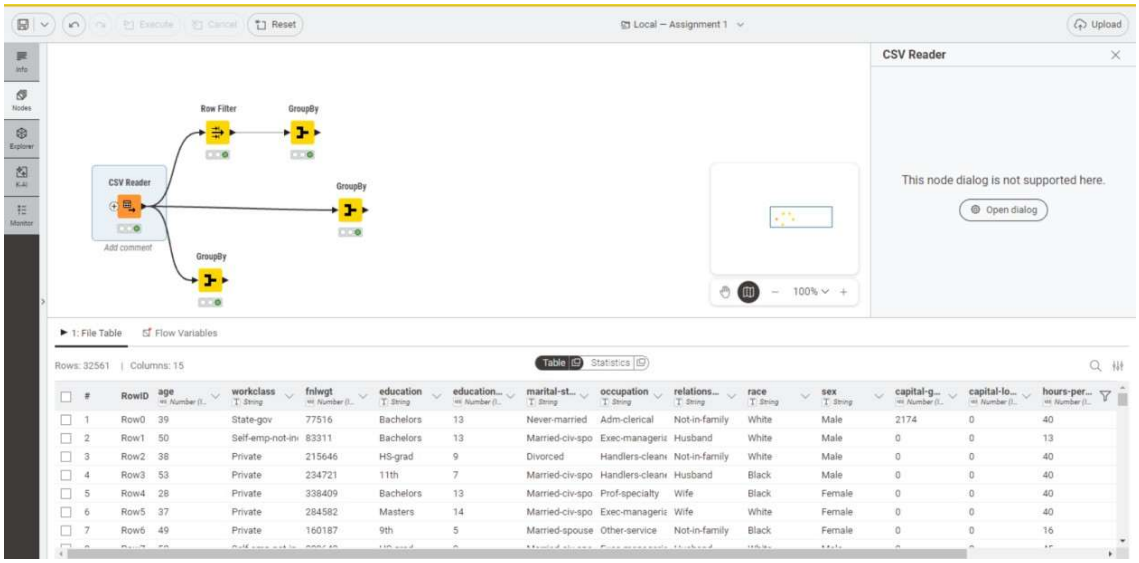
3) Calculate the averages of all numerical columns for each one of the 4 groups defined by sex and income values

4) Calculate

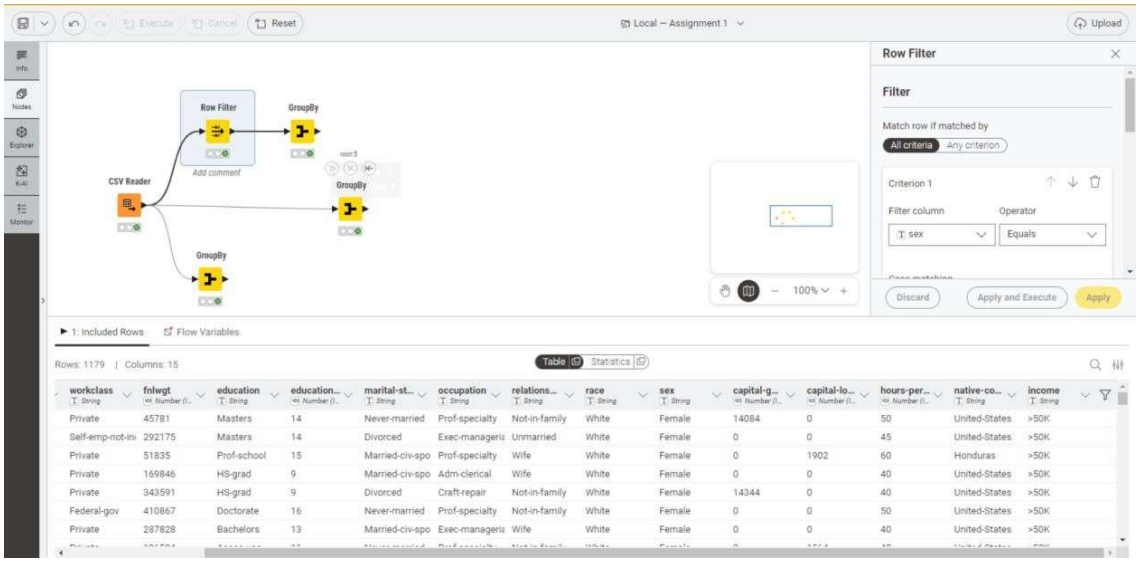
- the number of missing values in the occupation column
- the number of non-missing rows in the occupation column
- the number of rows in the occupation column
- the number of rows in the marital-status column

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

Step 1: Read CSV File “adult.csv”

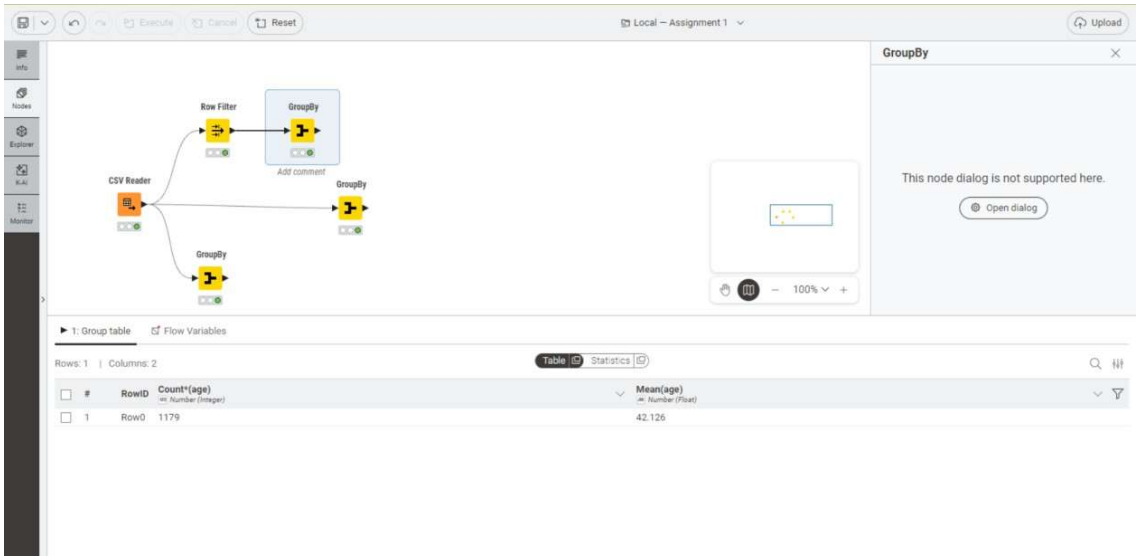


Step 2: Filter Row for Women with income >50K



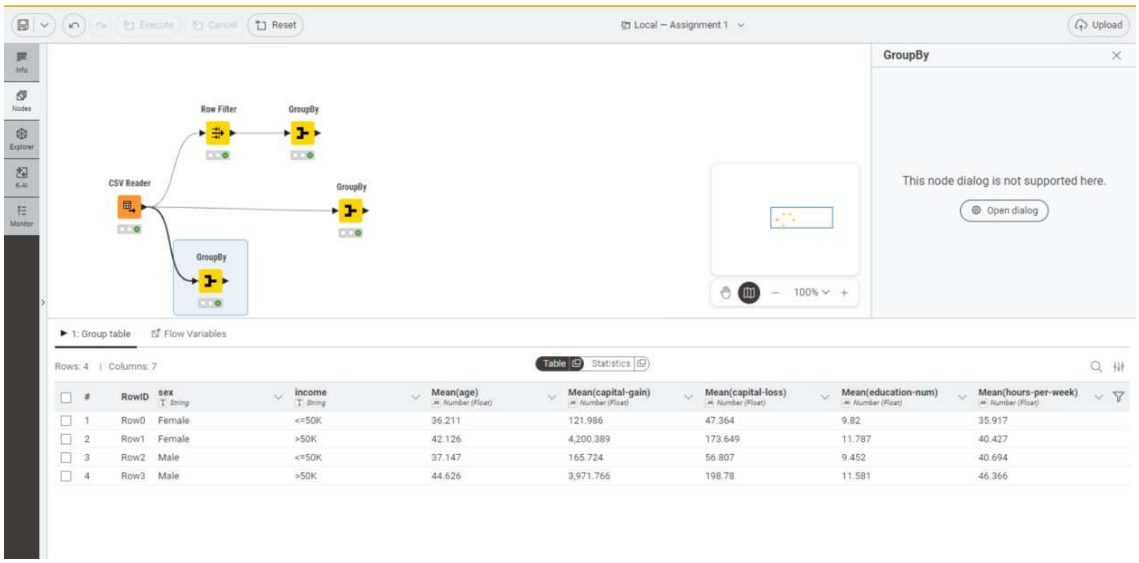
workclass	fnlwgt	education	education...	marital-st...	occupation	relations...	race	sex	capital-g...	capital-lo...	hours-per...	native-co...	income
Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084	0	50	United-States	>50K
Self-emp-not-in	292175	Masters	14	Divorced	Exec-managerial	Unmarried	White	Female	0	0	45	United-States	>50K
Private	51835	Prof-school	15	Married-civ-spo	Prof-specialty	Wife	White	Female	0	1902	60	Honduras	>50K
Private	169846	HS-grad	9	Married-civ-spo	Adm-clerical	Wife	White	Female	0	0	40	United-States	>50K
Private	343591	HS-grad	9	Divorced	Craft-repair	Not-in-family	White	Female	14344	0	40	United-States	>50K
Federal-gov	410867	Doctorate	16	Never-married	Prof-specialty	Not-in-family	White	Female	0	0	50	United-States	>50K
Private	287828	Bachelors	13	Married-civ-spo	Exec-managerial	Wife	White	Female	0	0	40	United-States	>50K

Step 3: Use GroupBy node to calculate the count and average age of women with income >50K



#	RowID	Count*(age)	Mean(age)
1	Row0	1179	42.126

Step 4: Use GroupBy node to calculate the average of all numerical column for each of the 4group defined by sex and income value



Step 5: Use GroupBy node to calculate Missing value count for occupation, non-missing value count for occupation, no of rows in occupation column, no of rows in martial-status