K.R. Mangalam University
School of Engineering & Technology
Department of Computer Science

Power BI LAB RECORD

Course Code: MCA (AI&ML)
Course Title: Data Analysis with Power Bi & Knime

Submitted by:
Name: Ayush Raj
Program: MCA
Roll No.: 2501940005

Submitted to:
Faculty Name: Aijaz Mohammad

Experiment 1

About KNIME:-

KNIME is an open-source data analytics platform where you build workflows visually using nodes. It helps with everything from cleaning data to training machine-learning models, all without forcing you to sit there writing code till your soul evaporates.

What it can do / real-life uses

1. **Data cleaning and preparation**
   Companies use KNIME to fix messy datasets, combine files, handle missing values, and shape data before analysis.
2. **Machine learning tasks**
   It's used to build models for predictions like customer churn, fraud detection, sales forecasting, and medical risk analysis.
3. **Automation of repetitive work**
   Businesses set up workflows that run on a schedule to generate reports, update dashboards, refresh databases, or send alerts.
4. **Integration across systems**
   KNIME connects different tools and databases so organizations can pull data from multiple places without manual effort.
5. **Decision support**
   Analysts and teams use it to create simple, visual pipelines that help them make evidence-based decisions quickly.

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

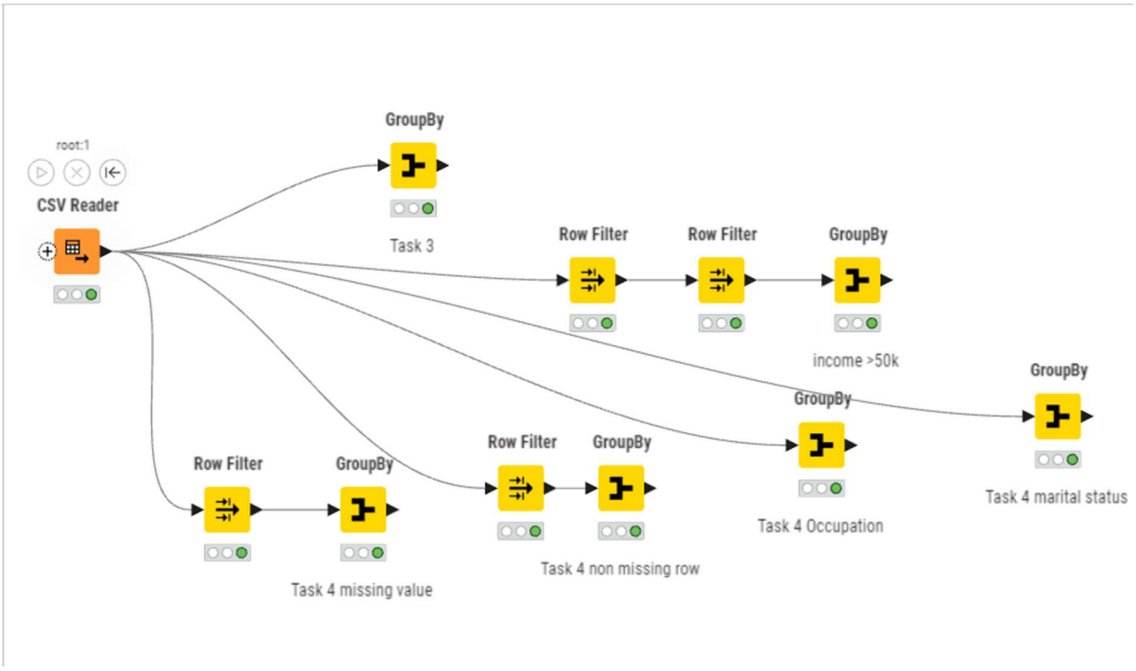
3) Calculate the averages of all numerical columns for each one of the 4 groups defined by sex and income values

4) Calculate

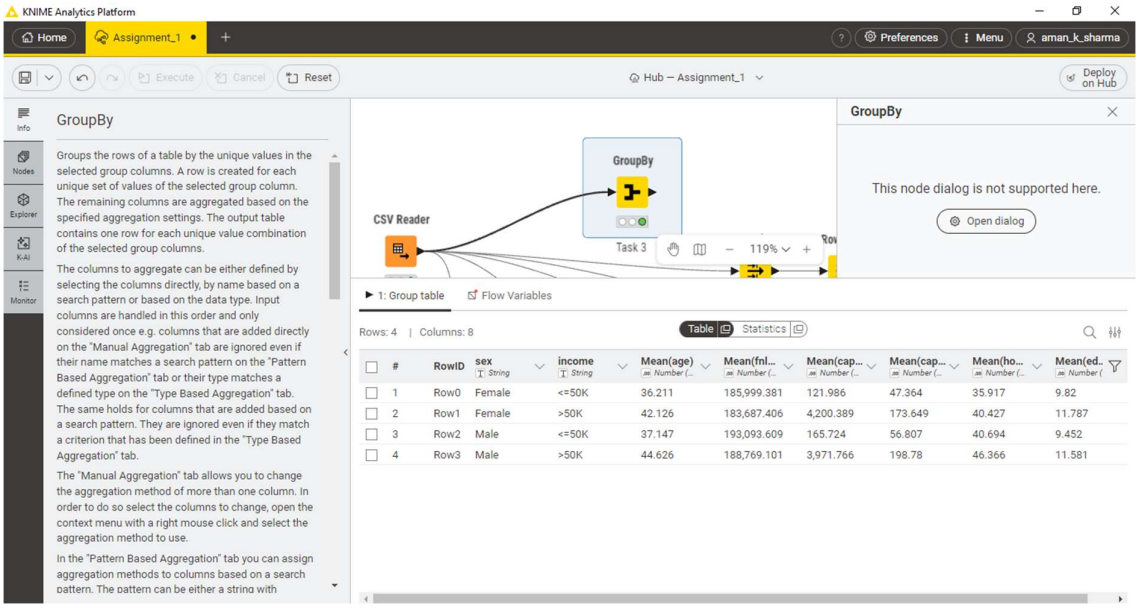
- the number of missing values in the occupation column
- the number of non-missing rows in the occupation column
- the number of rows in the occupation column
- the number of rows in the marital-status column

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

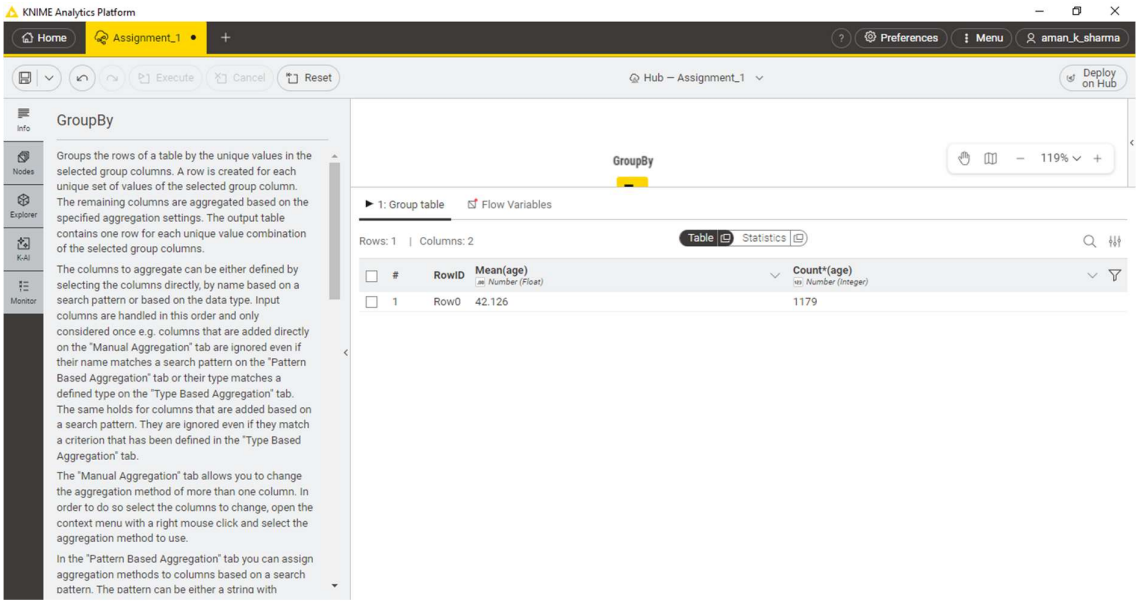
1)


root:1
CSV Reader
GroupBy
Task 3
Row Filter
Row Filter
GroupBy
income >50k
GroupBy
Task 4 marital status
GroupBy
Task 4 Occupation
Row Filter
GroupBy
Task 4 missing value
Row Filter
GroupBy
Task 4 non missing row

2)

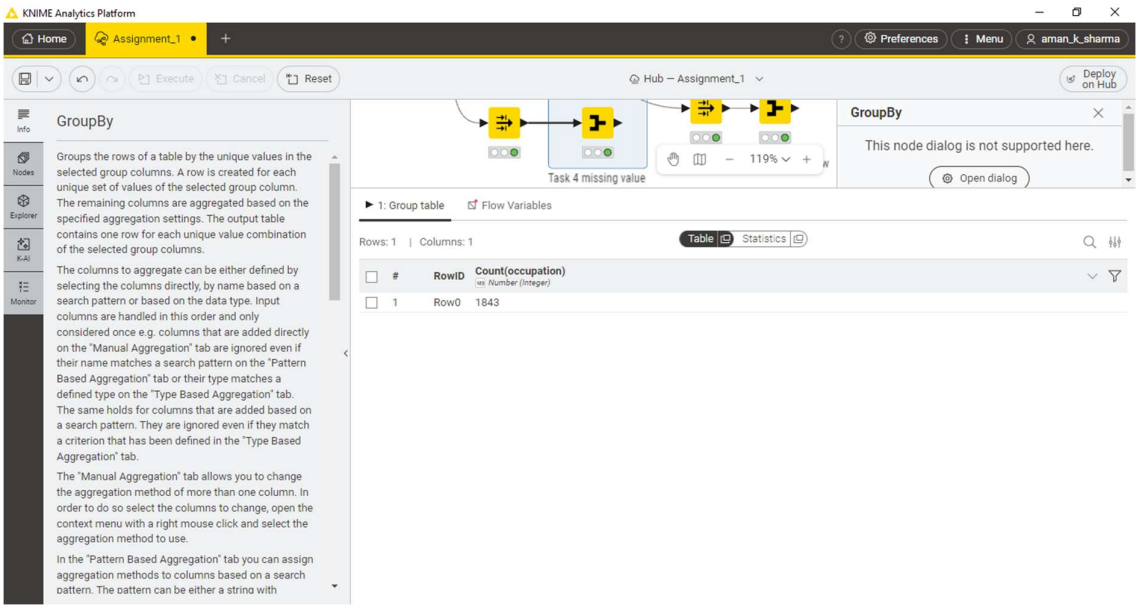

KNIME Analytics Platform
Home
Assignment_1
Preferences
Menu
aman_k_sharma
Execute
Cancel
Reset
Hub – Assignment_1
Deploy on Hub
GroupBy
Groups the rows of a table by the unique values in the selected group columns. A row is created for each unique set of values of the selected group column. The remaining 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
CSV Reader
GroupBy
Task 3
119%
GroupBy
This node dialog is not supported here.
Open dialog
1: Group table
Flow Variables
Rows: 4 | Columns: 8
Table
Statistics
RowID sex income Mean(age) Mean(fnl... Mean(cap... Mean(cap... Mean(ho... Mean(ed..
1 Row0 Female <=50K 36.211 185,999.381 121.986 47.364 35.917 9.82
2 Row1 Female >50K 42.126 183,687.406 4,200.389 173.649 40.427 11.787
3 Row2 Male <=50K 37.147 193,093.609 165.724 56.807 40.694 9.452
4 Row3 Male >50K 44.626 188,769.101 3,971.766 198.78 46.366 11.581

3)

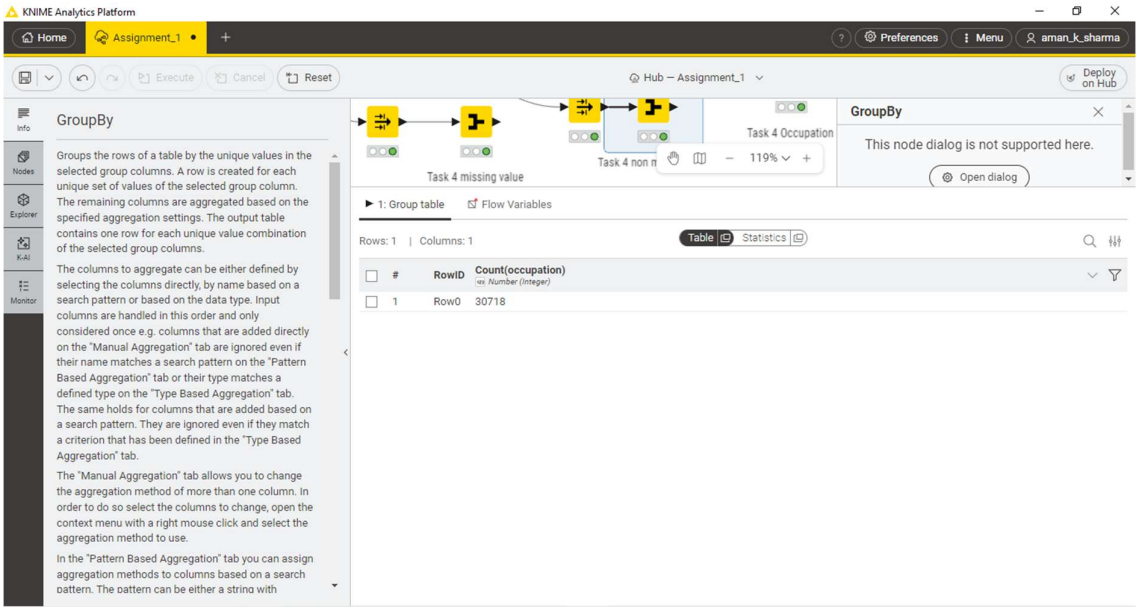

KNIME Analytics Platform
Home
Assignment_1
Preferences
Menu
aman_k_sharma
Execute
Cancel
Reset
Hub – Assignment_1
Deploy on Hub
GroupBy
Groups the rows of a table by the unique values in the selected group columns. A row is created for each unique set of values of the selected group column. The remaining 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
GroupBy
119%
1: Group table
Flow Variables
Rows: 1 | Columns: 2
Table
Statistics
RowID Mean(age) Number (Float) Count*(age) Number (Integer)
1 Row0 42.126 1179

4)

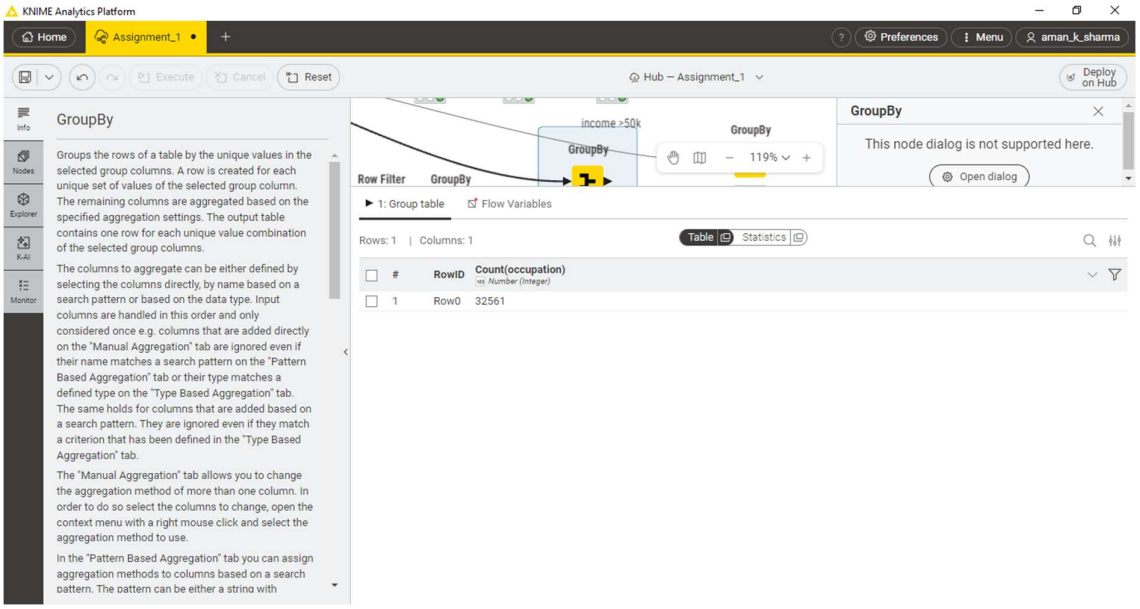
i)


KNIME Analytics Platform
Home
Assignment_1
Preferences
Menu
aman_k_sharma
Execute
Cancel
Reset
Hub – Assignment_1
Deploy on Hub
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 column. 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
Task 4 missing value
119%
GroupBy
This node dialog is not supported here.
Open dialog
1: Group table
Flow Variables
Rows: 1 | Columns: 1
Table
Statistics
RowID Count(occupation) Number (Integer)
1 Row0 1843

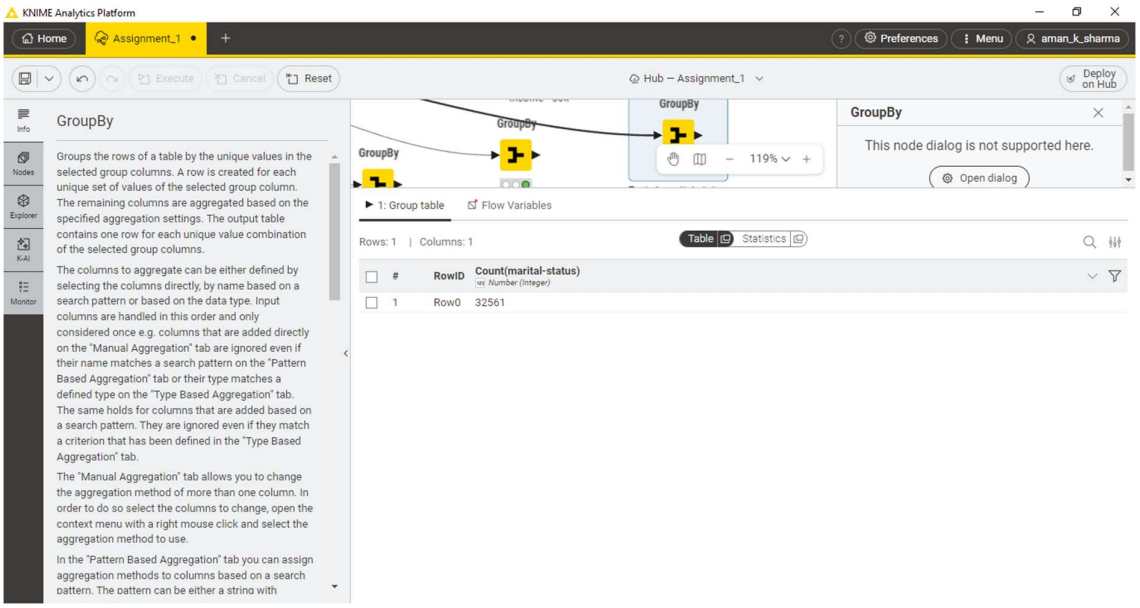
ii)


KNIME Analytics Platform
Home
Assignment_1
Preferences
Menu
aman_k_sharma
Execute
Cancel
Reset
Hub – Assignment_1
Deploy on Hub
Info
Nodes
Explorer
K-AI
Monitor
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 column. 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
Task 4 missing value
Task 4 Occupation
119%
GroupBy
This node dialog is not supported here.
Open dialog
1: Group table
Flow Variables
Rows: 1 | Columns: 1
Table
Statistics
RowID Count(occupation)
Number (Integer)
1 Row0 30718

iii)


KNIME Analytics Platform
Home
Assignment_1
Preferences
Menu
aman_k_sharma
Execute
Cancel
Reset
Hub – Assignment_1
Deploy on Hub
Info
Nodes
Explorer
K-AI
Monitor
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 column. 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
income >50k
GroupBy
Row Filter
GroupBy
119%
GroupBy
This node dialog is not supported here.
Open dialog
1: Group table
Flow Variables
Rows: 1 | Columns: 1
Table
Statistics
RowID Count(occupation)
Number (Integer)
1 Row0 32561

iv)


KNIME Analytics Platform
Home
Assignment_1
Preferences
Menu
aman_k_sharma
Execute
Cancel
Reset
Hub – Assignment_1
Deploy on Hub
Info
Nodes
Explorer
K-AI
Monitor
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 column. 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
GroupBy
GroupBy
GroupBy
119%
GroupBy
This node dialog is not supported here.
Open dialog
1: Group table
Flow Variables
Rows: 1 | Columns: 1
Table
Statistics
RowID Count(marital-status)
Number (Integer)
1 Row0 32561

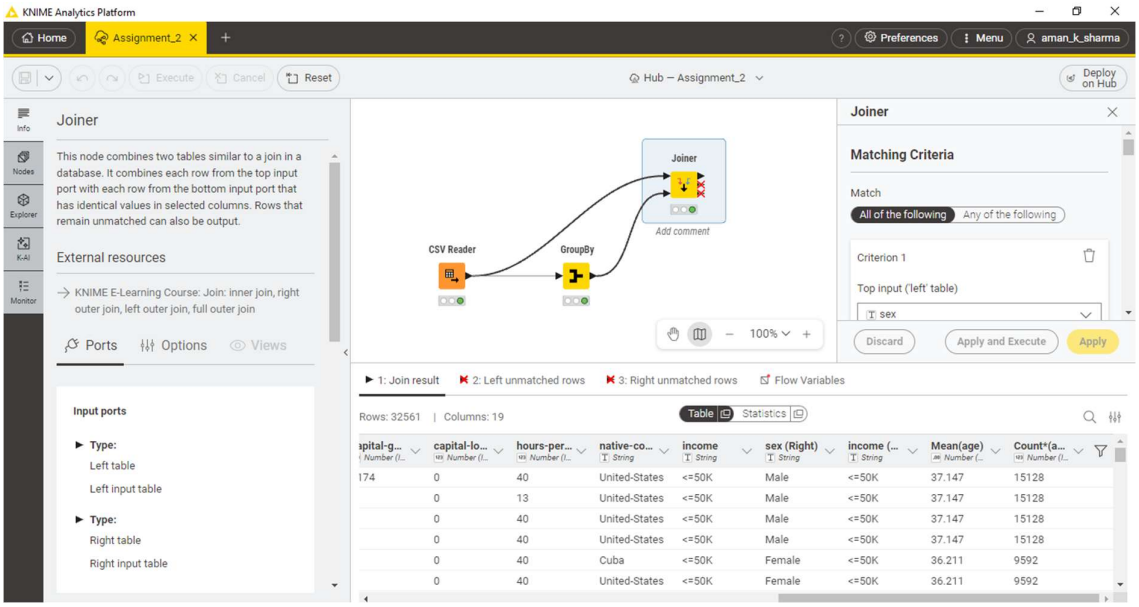
Experiment 2

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.

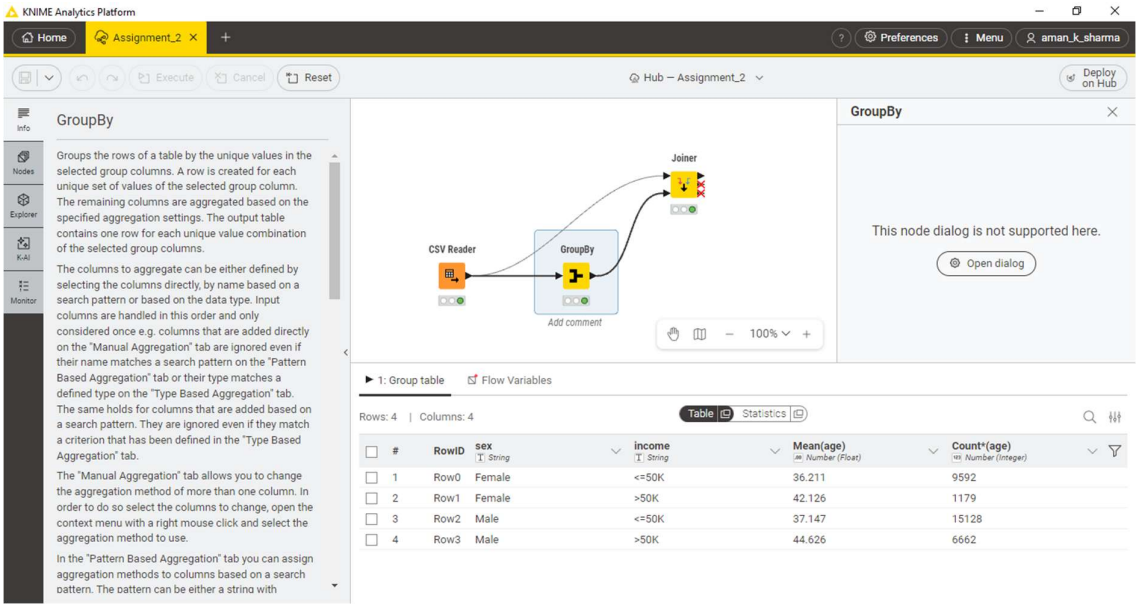
2) Calculate the average age and count for each one of the 4 groups defined by sex and income values

3) Join the two aggregated values to the original table

1)



2)


KNIME Analytics Platform
Home
Assignment_2
Preferences
Menu
aman_k_sharma
Execute
Cancel
Reset
Hub – Assignment_2
Deploy on Hub
Info
Nodes
Explorer
K-AI
Monitor
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 column. 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
Joiner
CSV Reader
GroupBy
Add comment
100%
GroupBy
This node dialog is not supported here.
Open dialog
1: Group table
Flow Variables
Rows: 4 | Columns: 4
Table
Statistics
RowID sex String income String Mean(age) Number (Float) Count*(age) Number (Integer)
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

Experiment 3

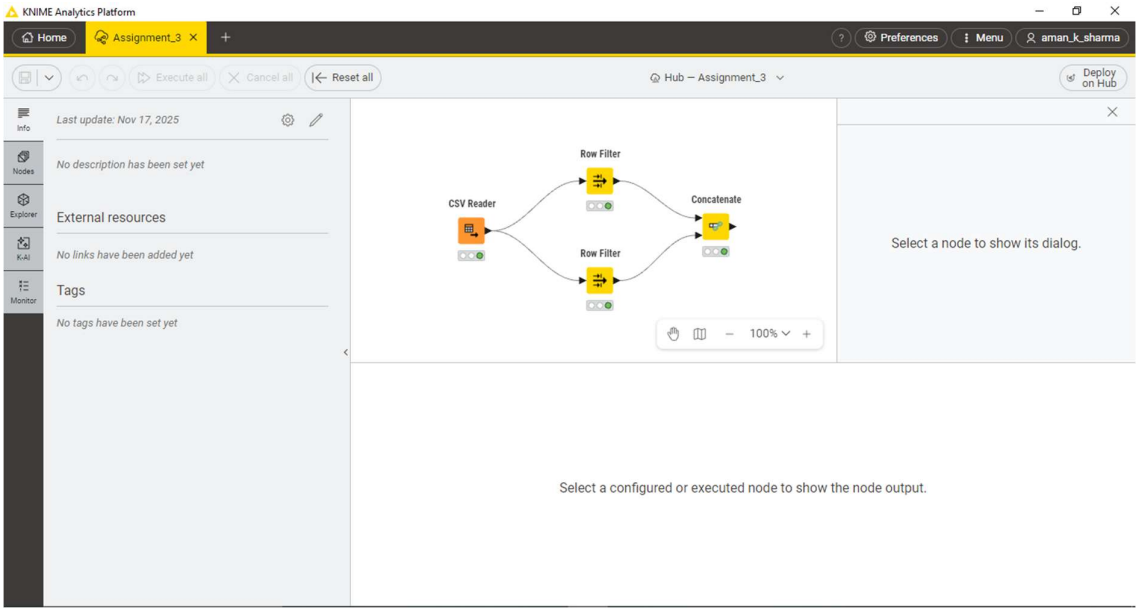
1) Read the adult.csv file available in the data folder on the KNIME Hub. The data are provided by the UCI Machine Learning Repository.

2) Extract people with age between 20 and 40 (both included) and working in a workclass starting with "S"

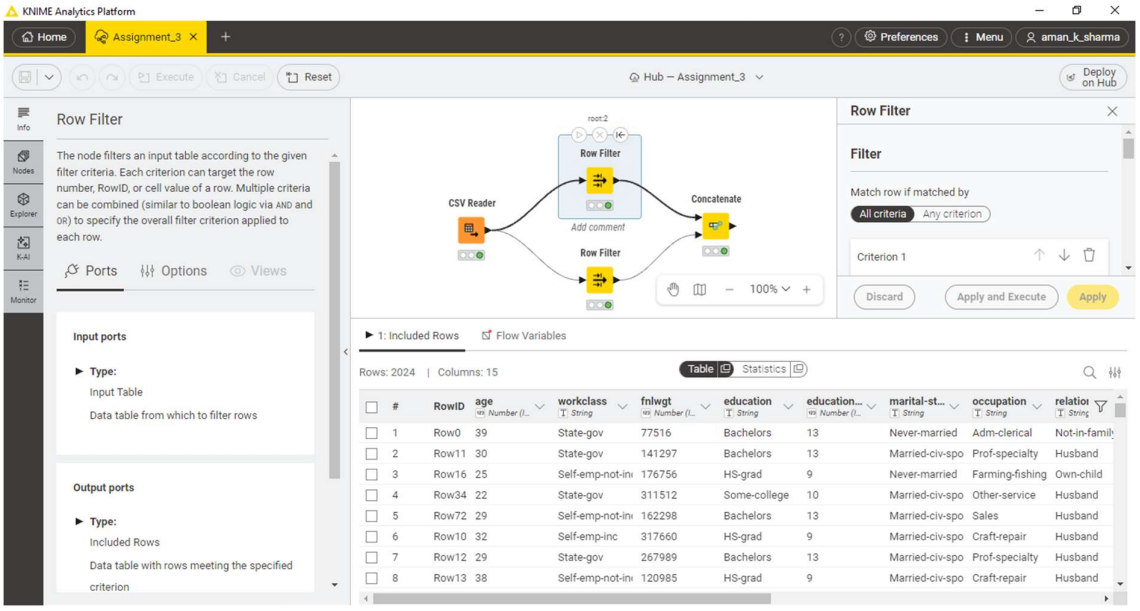
3) Extract people with age between 40 and 60 (both included) and working in a workclass starting with "P"

4) Concatenate both subsets into a single data table

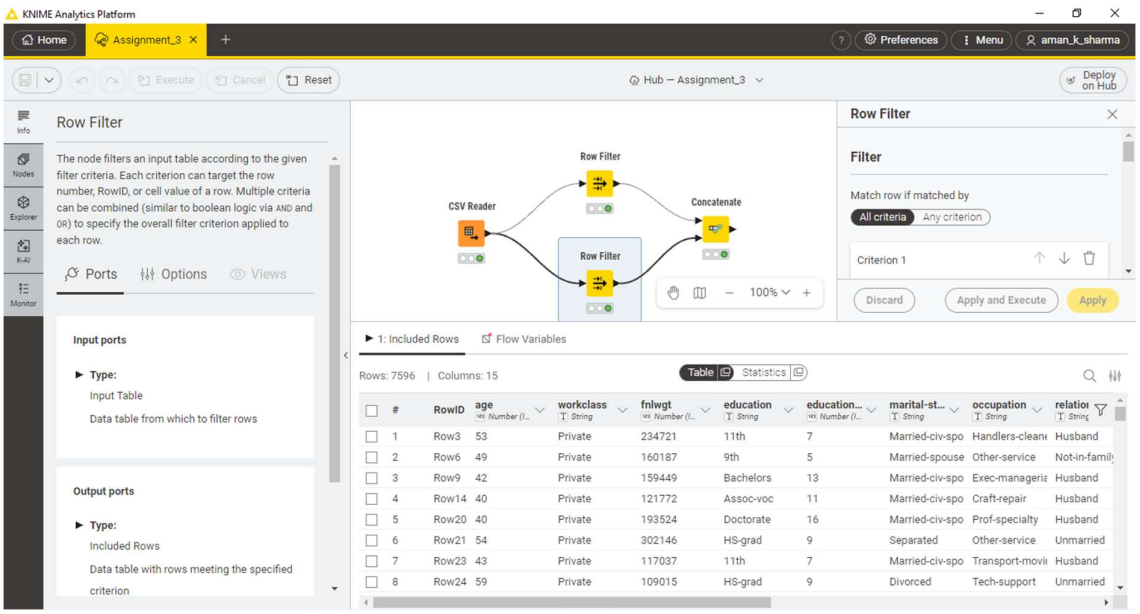
1)



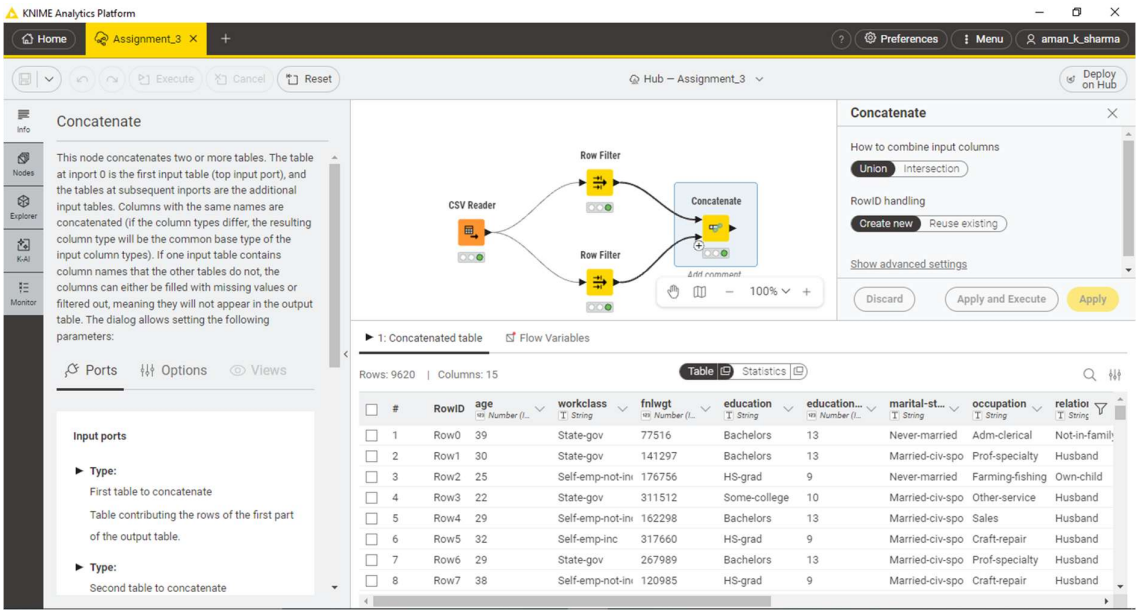
2)


KNIME Analytics Platform
Home
Assignment_3
Preferences
Menu
aman_k_sharma
Execute
Cancel
Reset
Hub – Assignment_3
Deploy on Hub
Row Filter
The node filters an input table according to the given filter criteria. Each criterion can target the row number, RowID, or cell value of a row. Multiple criteria can be combined (similar to boolean logic via AND and OR) to specify the overall filter criterion applied to each row.
Ports
Options
Views
Input ports
Type:
Input Table
Data table from which to filter rows
Output ports
Type:
Included Rows
Data table with rows meeting the specified criterion
CSV Reader
Row Filter
Add comment
Row Filter
Concatenate
100%
Row Filter
Filter
Match row if matched by
All criteria
Any criterion
Criterion 1
Discard
Apply and Execute
Apply
1: Included Rows
Flow Variables
Rows: 2024 | Columns: 15
Table
Statistics
RowID age workclass fnlwgt education education... marital-st... occupation relatio
1 Row0 39 State-gov 77516 Bachelors 13 Never-married Adm-clerical Not-in-family
2 Row11 30 State-gov 141297 Bachelors 13 Married-civ-spo Prof-specialty Husband
3 Row16 25 Self-emp-not-in 176756 HS-grad 9 Never-married Farming-fishing Own-child
4 Row34 22 State-gov 311512 Some-college 10 Married-civ-spo Other-service Husband
5 Row72 29 Self-emp-not-in 162298 Bachelors 13 Married-civ-spo Sales Husband
6 Row10 32 Self-emp-inc 317660 HS-grad 9 Married-civ-spo Craft-repair Husband
7 Row12 29 State-gov 267989 Bachelors 13 Married-civ-spo Prof-specialty Husband
8 Row13 38 Self-emp-not-in 120985 HS-grad 9 Married-civ-spo Craft-repair Husband

3)


KNIME Analytics Platform
Home
Assignment_3
Preferences
Menu
aman_k_sharma
Execute
Cancel
Reset
Hub – Assignment_3
Deploy on Hub
Row Filter
The node filters an input table according to the given filter criteria. Each criterion can target the row number, RowID, or cell value of a row. Multiple criteria can be combined (similar to boolean logic via AND and OR) to specify the overall filter criterion applied to each row.
Ports
Options
Views
Input ports
Type:
Input Table
Data table from which to filter rows
Output ports
Type:
Included Rows
Data table with rows meeting the specified criterion
Row Filter
CSV Reader
Row Filter
Concatenate
100%
Row Filter
Filter
Match row if matched by
All criteria
Any criterion
Criterion 1
Discard
Apply and Execute
Apply
1: Included Rows
Flow Variables
Rows: 7596 | Columns: 15
Table
Statistics
RowID age workclass fnlwgt education education... marital-st... occupation relatio
1 Row3 53 Private 234721 11th 7 Married-civ-spo Handlers-cleane Husband
2 Row6 49 Private 160187 9th 5 Married-spouse Other-service Not-in-family
3 Row9 42 Private 159449 Bachelors 13 Married-civ-spo Exec-managerial Husband
4 Row14 40 Private 121772 Assoc-voc 11 Married-civ-spo Craft-repair Husband
5 Row20 40 Private 193524 Doctorate 16 Married-civ-spo Prof-specialty Husband
6 Row21 54 Private 302146 HS-grad 9 Separated Other-service Unmarried
7 Row23 43 Private 117037 11th 7 Married-civ-spo Transport-movin Husband
8 Row24 59 Private 109015 HS-grad 9 Divorced Tech-support Unmarried

4)


KNIME Analytics Platform
Home
Assignment_3
Preferences
Menu
aman_k_sharma
Execute
Cancel
Reset
Hub – Assignment_3
Deploy on Hub
Info
Nodes
Explorer
K-AI
Monitor
Concatenate
This node concatenates two or more tables. The table at inport 0 is the first input table (top input port), and the tables at subsequent inports are the additional input tables. Columns with the same names are concatenated (if the column types differ, the resulting column type will be the common base type of the input column types). If one input table contains column names that the other tables do not, the columns can either be filled with missing values or filtered out, meaning they will not appear in the output table. The dialog allows setting the following parameters:
Ports
Options
Views
Input ports
Type:
First table to concatenate
Table contributing the rows of the first part of the output table.
Type:
Second table to concatenate
CSV Reader
Row Filter
Row Filter
Concatenate
100%
Concatenate
How to combine input columns
Union
Intersection
RowID handling
Create new
Reuse existing
Show advanced settings
Discard
Apply and Execute
Apply
1: Concatenated table
Flow Variables
Rows: 9620 | Columns: 15
Table
Statistics
RowID age workclass fnlwgt education education... marital-st... occupation relatio
1 Row0 39 State-gov 77516 Bachelors 13 Never-married Adm-clerical Not-in-famil
2 Row1 30 State-gov 141297 Bachelors 13 Married-civ-spo Prof-specialty Husband
3 Row2 25 Self-emp-not-in 176756 HS-grad 9 Never-married Farming-fishing Own-child
4 Row3 22 State-gov 311512 Some-college 10 Married-civ-spo Other-service Husband
5 Row4 29 Self-emp-not-in 162298 Bachelors 13 Married-civ-spo Sales Husband
6 Row5 32 Self-emp-inc 317660 HS-grad 9 Married-civ-spo Craft-repair Husband
7 Row6 29 State-gov 267989 Bachelors 13 Married-civ-spo Prof-specialty Husband
8 Row7 38 Self-emp-not-in 120985 HS-grad 9 Married-civ-spo Craft-repair Husband