

## Assignment 03

- 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

### Step 1: Read the adult.csv file

The image shows the KNIME software interface. On the left, a workflow is visible with a 'CSV Reader' node, followed by two 'Row Filter' nodes, and a 'Concatenate' node. The 'CSV Reader' node has a comment 'Add comment'. The 'Row Filter' nodes are connected to the 'Concatenate' node. On the right, a 'CSV Reader' dialog box is open, displaying the message 'This node dialog is not supported here.' and an 'Open dialog' button. Below the workflow, a table view shows the first 10 rows of the data. The table has 15 columns: #, RowID, age, workclass, fnlwgt, education, education-num, marital-st, occupation, relations, race, sex, capital-g, capital-lo, and hours-per. The data is as follows:

#	RowID	age	workclass	fnlwgt	education	education-num	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-in	83311	Bachelors	13	Married-civ-spo	Exec-manageriz	Husband	White	Male	0	0	13
3	Row2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleanz	Not-in-family	White	Male	0	0	40
4	Row3	53	Private	234721	11th	7	Married-civ-spo	Handlers-cleanz	Husband	Black	Male	0	0	40
5	Row4	28	Private	338409	Bachelors	13	Married-civ-spo	Prof-specialty	Wife	Black	Female	0	0	40
6	Row5	37	Private	284582	Masters	14	Married-civ-spo	Exec-manageriz	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-in	209642	HS-grad	9	Married-civ-spo	Exec-manageriz	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-spo	Exec-manageriz	Husband	White	Male	5178	0	40

## Power BI and KNIME

**Step 2:** Extract people with age between 20 and 40 (both included) and working in a work class starting with “S”

The KNIME workflow for Step 2 consists of the following components:

- CSV Reader:** Reads the input data.
- Row Filter (Top):** Filters rows where the 'age' column is greater than or equal to 20.
- Row Filter (Bottom):** Filters rows where the 'workclass' column starts with 'S'.
- Concatenate:** Combines the results of the two Row Filter nodes.

The **Row Filter** configuration panel shows the following settings:

- Filter:** Match row if matched by **All criteria**.
- Criterion 1:** Filter column: **age**, Operator: **Greater than or equal to**, Value: **20**.

The resulting table shows 10 rows of data:

#	RowID	age	workclass	fnlwgt	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	Row11	30	State-gov	141297	Bachelors	13	Married-civ-spo	Prof-specialty	Husband	Asian-Pac-Islan	Male	0	0	40
3	Row16	25	Self-emp-not-inc	176756	HS-grad	9	Never-married	Farming-fishing	Own-child	White	Male	0	0	35
4	Row34	22	State-gov	311512	Some-college	10	Married-civ-spo	Other-service	Husband	Black	Male	0	0	15
5	Row72	29	Self-emp-not-inc	162298	Bachelors	13	Married-civ-spo	Sales	Husband	White	Male	0	0	70
6	Row10	32	Self-emp-inc	317660	HS-grad	9	Married-civ-spo	Craft-repair	Husband	White	Male	7688	0	40
7	Row12	29	State-gov	267989	Bachelors	13	Married-civ-spo	Prof-specialty	Husband	White	Male	0	0	50
8	Row13	38	Self-emp-not-inc	120985	HS-grad	9	Married-civ-spo	Craft-repair	Husband	White	Male	4386	0	35
9	Row17	28	State-gov	175325	HS-grad	9	Married-civ-spo	Protective-serv	Husband	White	Male	0	0	40
10	Row17	28	State-gov	149624	Bachelors	13	Married-civ-spo	Prof-specialty	Husband	White	Male	0	0	40

**Step 3:** Extract People with age between 40 and 60 (both included) and working in a work class starting with “P”

The KNIME workflow for Step 3 consists of the following components:

- CSV Reader:** Reads the input data.
- Row Filter (Top):** Filters rows where the 'age' column is greater than or equal to 40.
- Row Filter (Bottom):** Filters rows where the 'workclass' column starts with 'P'.
- Concatenate:** Combines the results of the two Row Filter nodes.

The **Row Filter** configuration panel shows the following settings:

- Filter:** Match row if matched by **All criteria**.
- Criterion 1:** Filter column: **age**, Operator: **Greater than or equal to**, Value: **40**.

The resulting table shows 10 rows of data:

#	RowID	age	workclass	fnlwgt	education	education--	marital-st--	occupation	relations--	race	sex	capital-g--	capital-lo--	hours-per--
1	Row3	53	Private	234721	11th	7	Married-civ-spo	Handlers-clean	Husband	Black	Male	0	0	40
2	Row6	49	Private	160187	9th	5	Married-spouse	Other-service	Not-in-family	Black	Female	0	0	16
3	Row9	42	Private	159449	Bachelors	13	Married-civ-spo	Exec-manageris	Husband	White	Male	5178	0	40
4	Row21	54	Private	302146	HS-grad	9	Separated	Other-service	Unmarried	Black	Female	0	0	20
5	Row23	43	Private	117037	11th	7	Married-civ-spo	Transport-movis	Husband	White	Male	0	2042	40
6	Row24	59	Private	109015	HS-grad	9	Divorced	Tech-support	Unmarried	White	Female	0	0	40
7	Row29	49	Private	193366	HS-grad	9	Married-civ-spo	Craft-repair	Husband	White	Male	0	0	40
8	Row32	45	Private	386940	Bachelors	13	Divorced	Exec-manageris	Own-child	White	Male	0	1408	40
9	Row35	48	Private	242406	11th	7	Never-married	Machine-op-ins	Unmarried	White	Male	0	0	40
10	Row43	49	Private	94638	HS-grad	9	Separated	Adm-clerical	Unmarried	White	Female	0	0	40

# Power BI and KNIME

## Step 4: Concatenate both subsets into a single data

The workflow diagram shows a CSV Reader node connected to two Row Filter nodes. Both Row Filter nodes are connected to a Concatenate node. The Concatenate node has a settings panel on the right with the following options:

- How to combine input columns: **Union** (selected), Intersection
- RowID handling: **Create new** (selected), Reuse existing
- Show advanced settings

The bottom of the interface shows a table with 15 columns and 10 rows displayed. The table is titled '1: Concatenated table' and has 8717 rows and 15 columns.

#	RowID	age	workclass	fnlwgt	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	30	State-gov	141297	Bachelors	13	Married-civ-spo	Prof-specialty	Husband	Asian-Pac-Isian	Male	0	0	40
3	Row2	25	Self-emp-not-in	176756	HS-grad	9	Never-married	Farming-fishing	Own-child	White	Male	0	0	35
4	Row3	22	State-gov	311512	Some-college	10	Married-civ-spo	Other-service	Husband	Black	Male	0	0	15
5	Row4	29	Self-emp-not-in	162298	Bachelors	13	Married-civ-spo	Sales	Husband	White	Male	0	0	70
6	Row5	32	Self-emp-inc	317660	HS-grad	9	Married-civ-spo	Craft-repair	Husband	White	Male	7688	0	40
7	Row6	29	State-gov	267989	Bachelors	13	Married-civ-spo	Prof-specialty	Husband	White	Male	0	0	50
8	Row7	38	Self-emp-not-in	120985	HS-grad	9	Married-civ-spo	Craft-repair	Husband	White	Male	4386	0	35
9	Row8	28	State-gov	175325	HS-grad	9	Married-civ-spo	Protective-serv	Husband	White	Male	0	0	40
10	Row9	28	State-gov	149624	Bachelors	13	Married-civ-spo	Prof-specialty	Husband	White	Male	0	0	40