

# Assignment 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

## Step 1: Read the adult.csv file

The screenshot shows a KNIME workflow titled "Local – Assignment 3". The workflow consists of the following steps:

- A "CSV Reader" node is connected to a "Row Filter" node.
- The output of the "Row Filter" node is connected to a "Concatenate" node.
- The output of the "Concatenate" node is connected to another "Row Filter" node.
- The output of the second "Row Filter" node is also connected to the "Concatenate" node.

The "CSV Reader" node has an "Add comment" field. The "Concatenate" node has a "Comment" field. A tooltip on the "Concatenate" node says: "This node dialog is not supported here." There is a "Open dialog" button next to it.

Below the workflow, there is a "File Table" view showing the first 10 rows of the dataset:

#	RowID	age	workclass	fnlwgt	education	education-leve...	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

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

Uday Garg  
2501940027  
Power Bi Assignment – 3

The screenshot shows a data processing workflow. It starts with a 'CSV Reader' node, followed by a 'Row Filter' node with the condition 'age > 40'. This is followed by a 'Concatenate' node, which then leads to another 'Row Filter' node with the condition 'workclass starts with "P"'. The final output is a table with 1857 rows.

#	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 screenshot shows a data processing workflow similar to the previous one, but with different filter conditions. It starts with a 'CSV Reader' node, followed by a 'Row Filter' node with the condition 'age >= 40 AND age <= 60'. This is followed by a 'Concatenate' node, which then leads to another 'Row Filter' node with the condition 'workclass starts with "P"'. The final output is a table with 6860 rows.

#	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-moving	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

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

Screenshot of a data processing tool interface showing a workflow for concatenating CSV files.

**Workflow Diagram:**

```

graph LR
    CSVReader[CSV Reader] --> RowFilter1[Row Filter]
    RowFilter1 --> Concatenate[Concatenate]
    CSVReader --> RowFilter2[Row Filter]
    RowFilter2 --> Concatenate
    Concatenate --> Output[Add comment]
  
```

**Concatenate Settings:**

- How to combine input columns: Union
- RowID handling: Create new
- Show advanced settings

**Data Preview:**

Rows: 8717 | Columns: 15

#	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-Islan	Male	0	0	40
3	Row2	25	Self-emp-not-inv	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-inv	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-inv	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