IoT Domain Analyst

Lab Record — Lab 8

<div align="right">- 28 May 2021</div>

Programme	:	B.Tech(CSE)	Semester	:	Winter 2020–21
Course Title	:	IoT Domain Analyst – Lab	Code	:	ECE3502
			Slot	:	L5+L6
Name	:	**Aadhitya Swarnesh**	Registration. No	:	▨▨▨▨▨▨
Faculty (s)	:	▨▨▨▨▨	Expt. No	:	**8**

Experiment 1 :

Data Science — Pre-processing and Analysis of Datasets.

Aim :

To perform the required pre-processing of a dataset and use the "KNIME" program to create interactive visualisations demonstrating the features of the dataset and the relation between their parameters.

Description :

✦ Import the Dataset into KNIME workflow

✦ Perform the required pre-processing steps like choosing the required and necessary columns, rows, etc. We also handle the NULL values here.

✦ We create interactive visual representations of the data in the dataset to better understand the relationships between the parameters of the dataset

Flow Diagrams :

Stage 1 : Retrieve the Dataset

In this step, we download the dataset , or prepare it, or collect the data and form a dataset, or perform any of the processes to prepare a dataset. Here we download a dataset of the sales data of users. The dataset looks as follows :

| product	country	date	quantity	amount	card	Cust_ID
prod_4	unknown	2008-12-12	1	3		Cust_8
prod_3	China	2009-04-10	2	160	N	Cust_2
prod_3	China	2009-04-10	2	160	Y	Cust_5
prod_3	China	2009-05-10	2	160		Cust_2
prod_3	USA	2009-05-20	20	1600		Cust_3
prod_3	Brazil	2009-06-08	15	1200		Cust_7
prod_1	USA	2009-07-04	2	70	Y	Cust_3

Stage 2 : Import the Dataset

In this stage, we create a new KNIME workflow and import the CSV file into the workflow, either by dragging it into the workflow, or by choosing a CSV node, and configuring the path of the file leading to our dataset. The configuration of this node looks as follows :
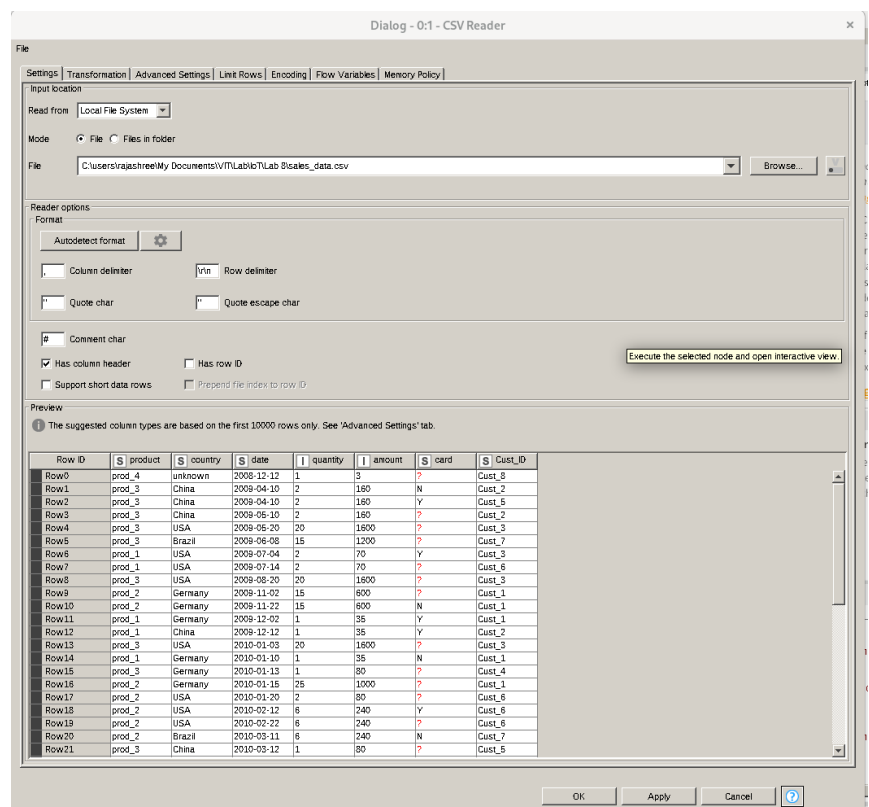
Stage 3 : Filter out the not so necessary Columns

In this stage, we choose the columns that are necessary for visualisation, and remove the remaining columns like "date", "card_id", "customer ID", etc which are unique for each record and will thus not serve any purpose for an analysis project. The configuration of this node is as follows :

Column Filter

Node 2
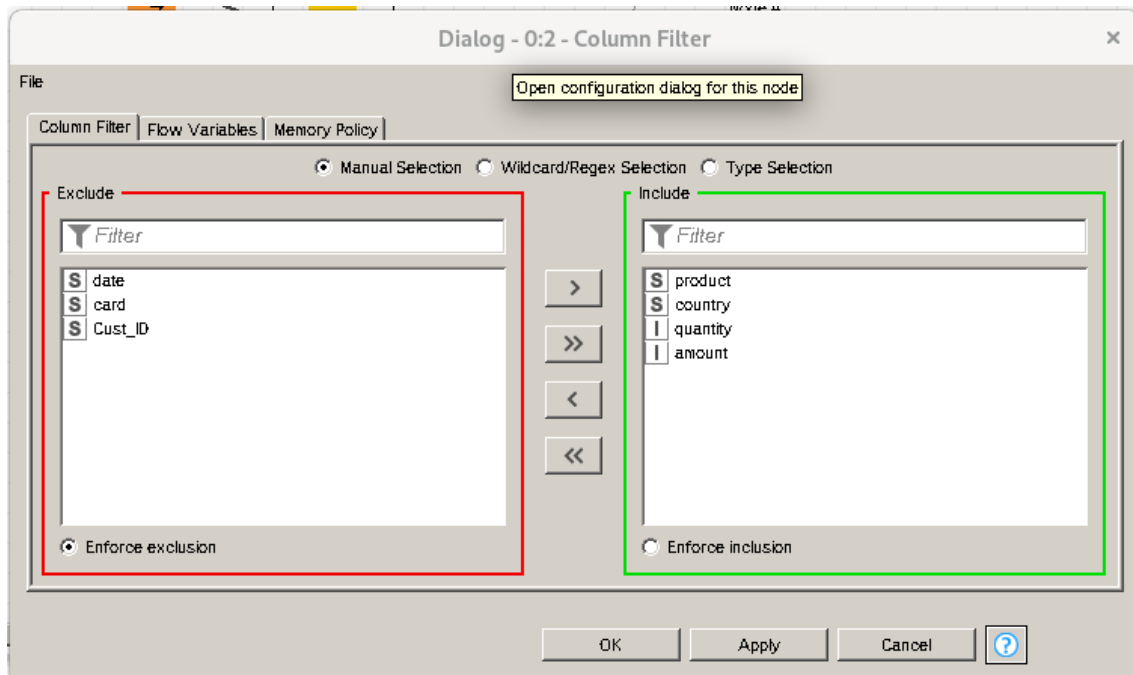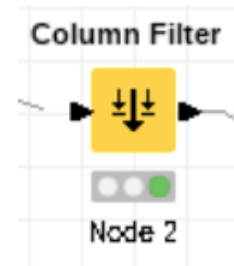
Stage 4 : Filter the Rows
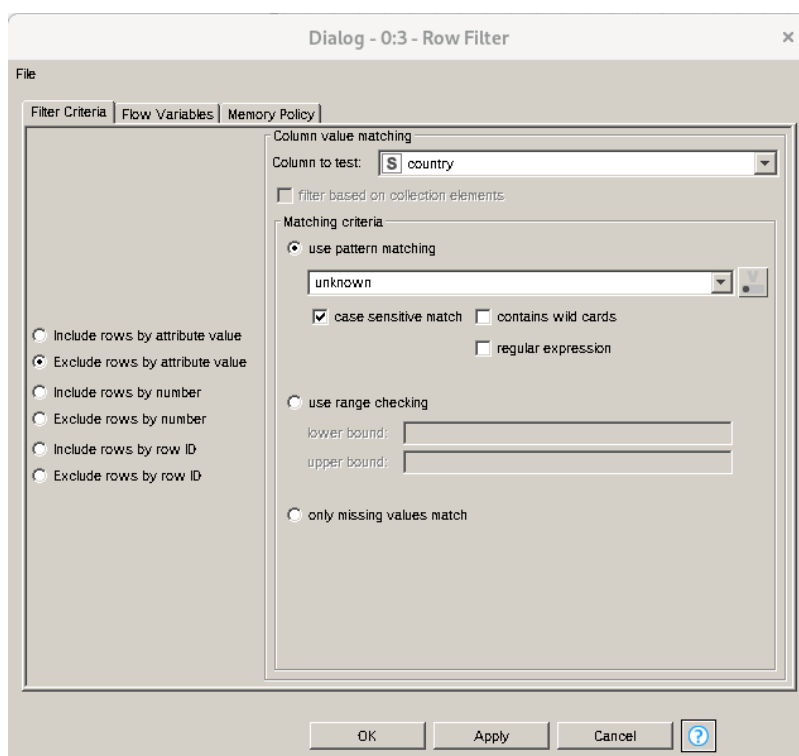
In this stage, we remove the rows with "NULL" values, as these would just interfere with the analysis. The configuration of this node is as follows :

Row Filter

Node 3
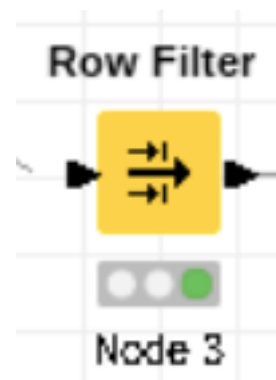
Stage 5 : Visualisation — Bar Chart

In this stage, we have finished pre-processing the dataset, and are now going to visualise the dataset using interactive charts and diagrams. Here we have implemented a Bar chart of the amount and quantity across the regions. The **amount** is the one in **orange**, and the **quantity** is in **blue**. The configuration of this node looks as follows :
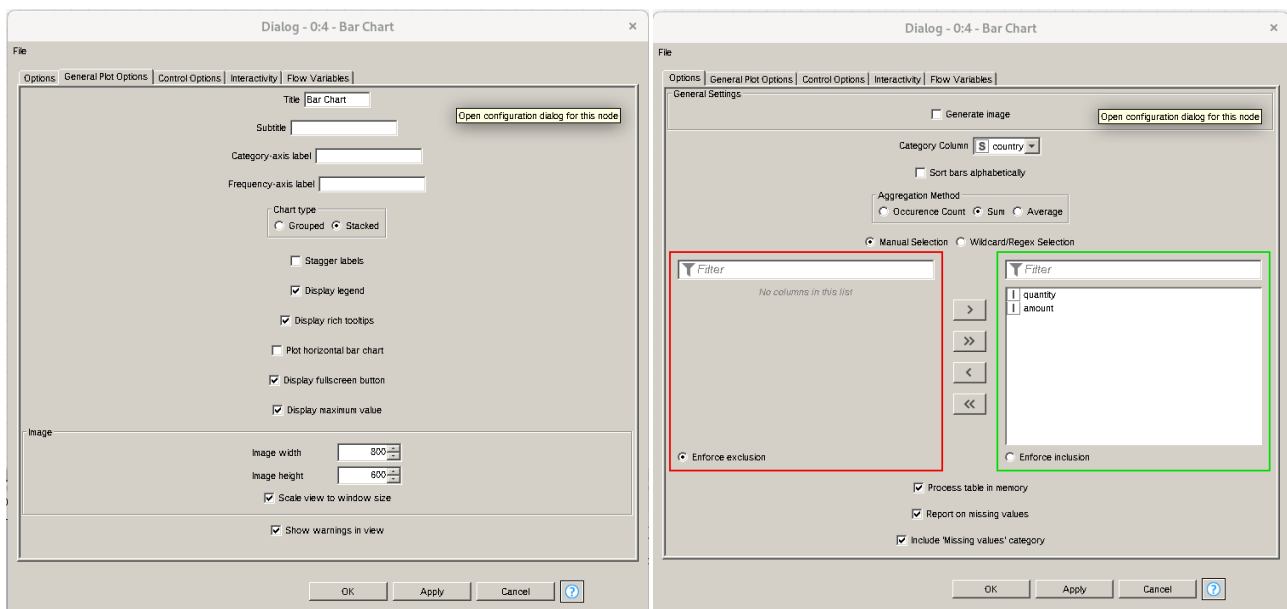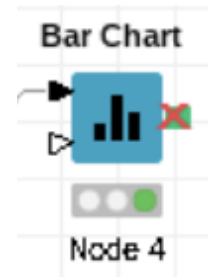
Stage 6 : Visualisation — Pie Chart

As done in the previous part, we have a plotted a pie chart denoting the share of amount and quantity across countries. The configuration of this node looks as follows :
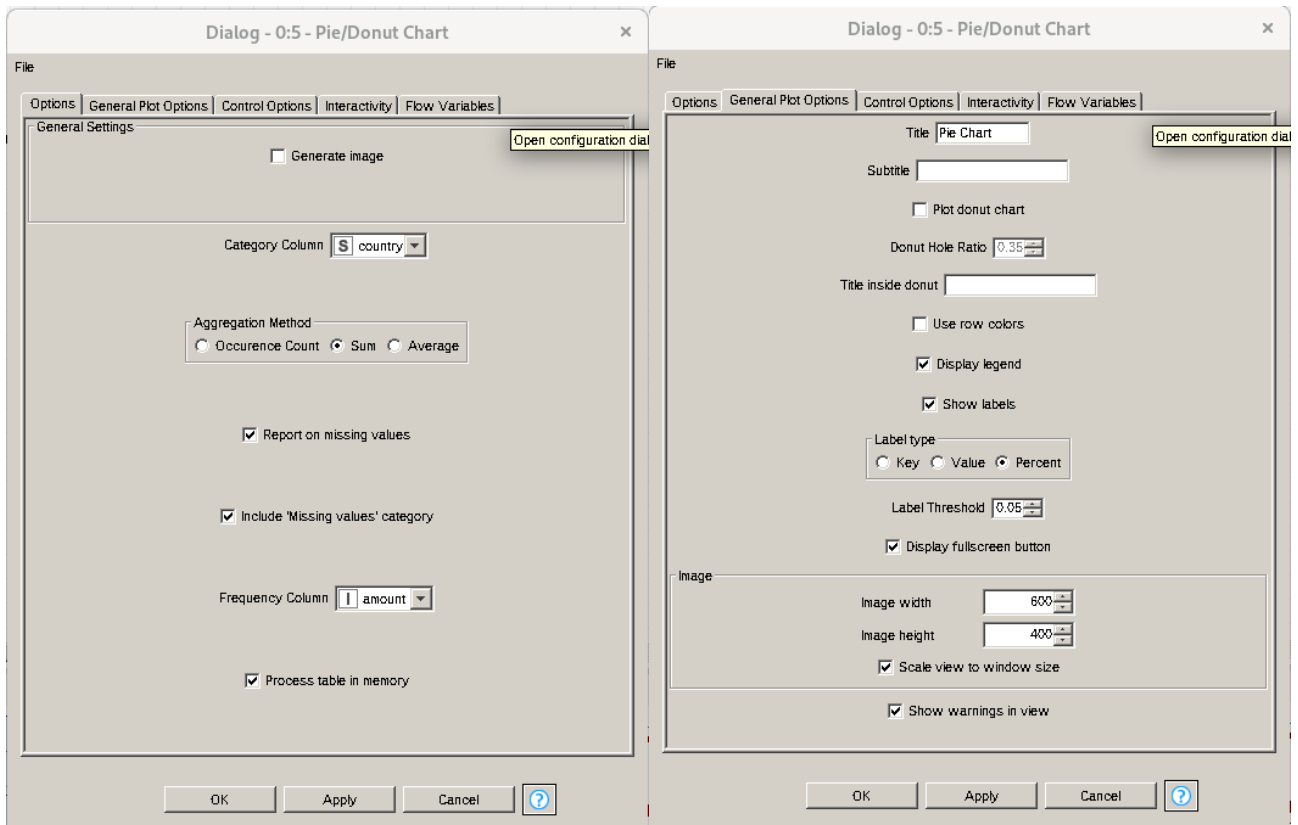
Stage 7 : Complete Workflow

This diagram represents the complete flow with all the nodes and the connections between them which denotes the flow of data in the workflow and across the various nodes used.
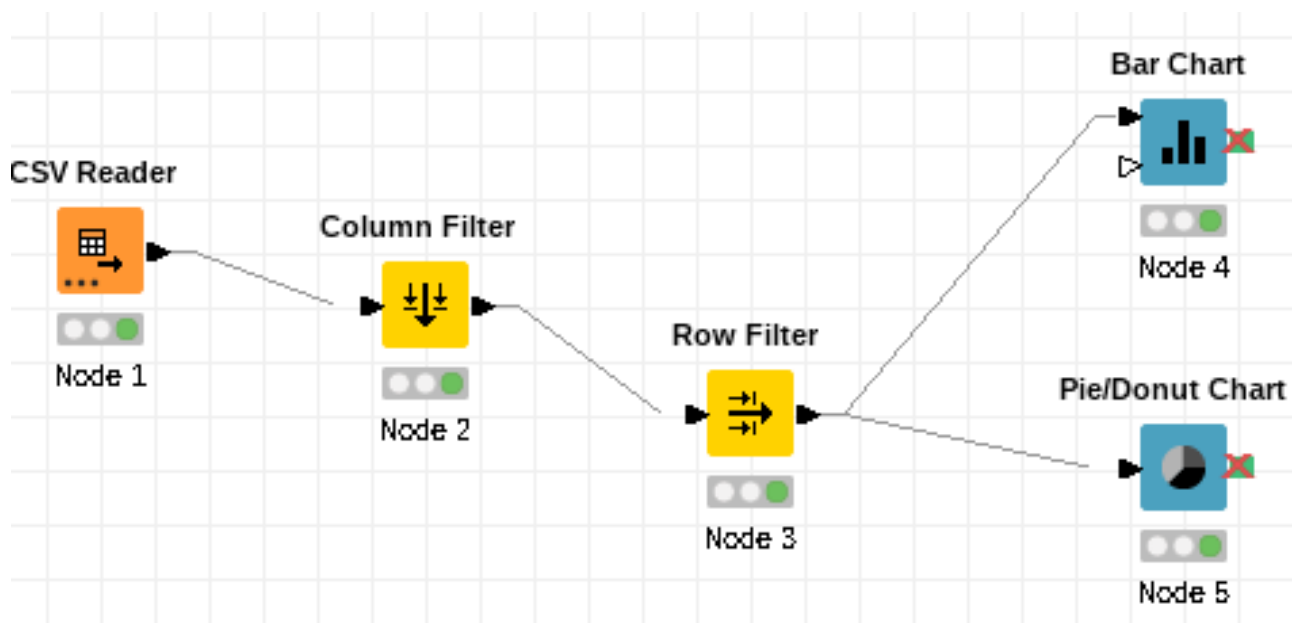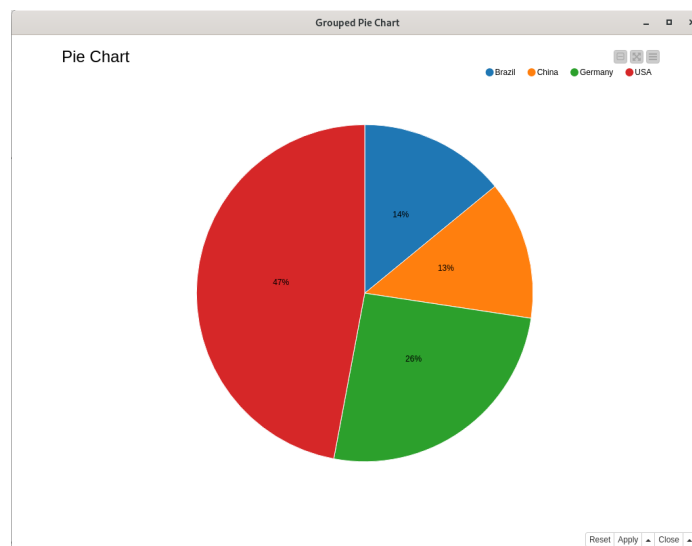
Result :

We have completed the procedures mentioned in the previous sections and have obtained the following results.

This is the BAR Chart that we have obtained which the amount and quantity across the regions. The **amount** is the one in **orange**, and the **quantity** is in **blue**.
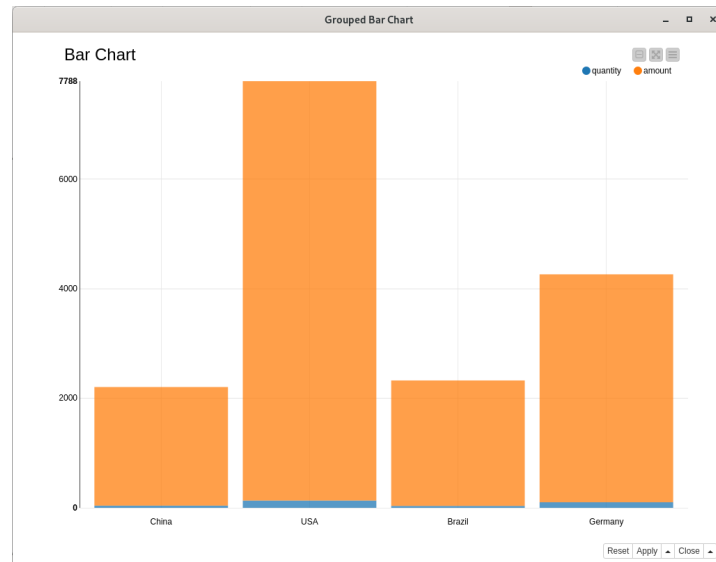
This is the Pie chart that we have obtained which denotes the share of amount and quantity across countries.

We have thus produced successful results and have completed a Data Science lab experiment using the KNIME Software and have plotted our results here using the concepts of IoT and Data Science.