Knime Steps

The Steps to Knime

1. **Added Read CSV Node to Diagram.**

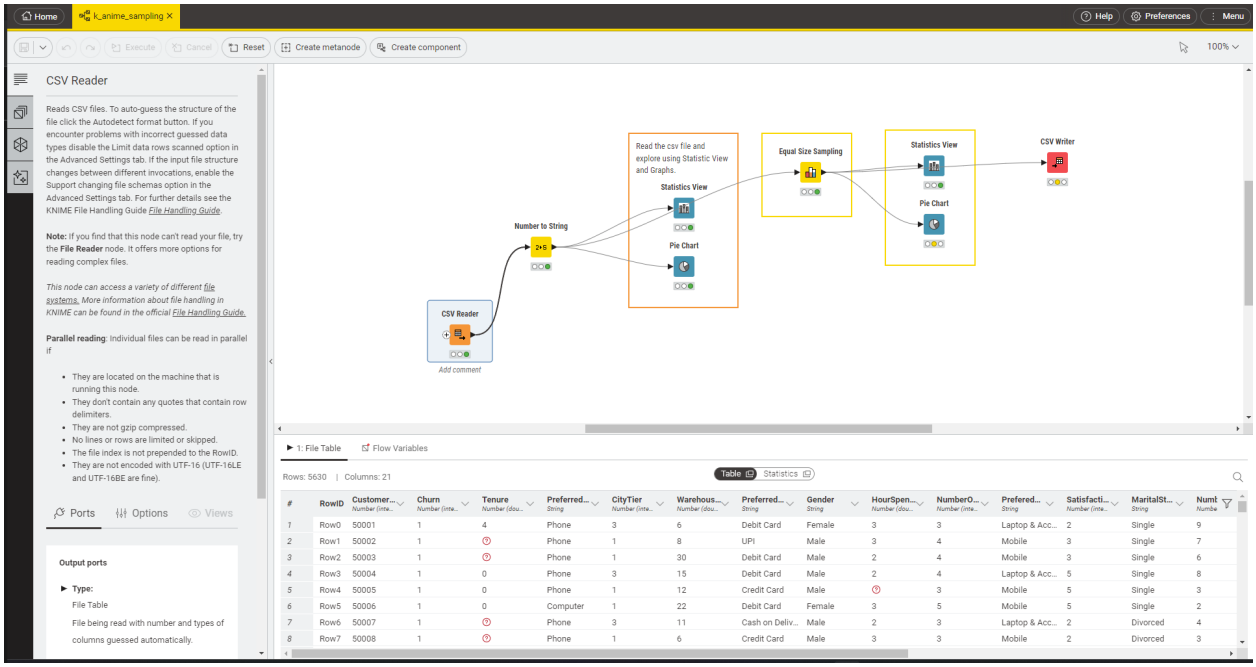
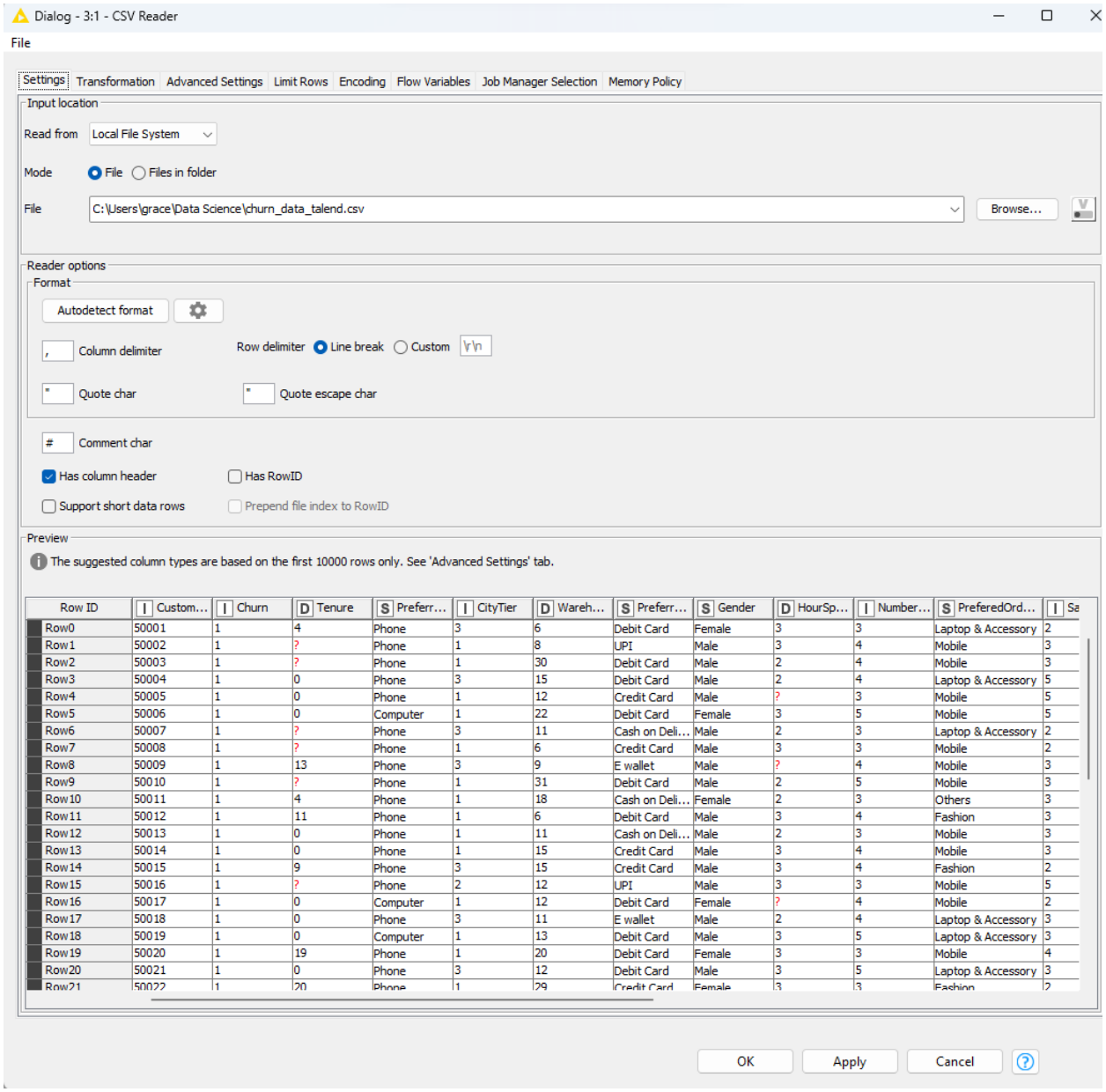


Figure 1: Overview of Diagram



Row ID	Custom...	Churn	Tenure	Preferr...	CityTier	Wareh...	Preferr...	Gender	HourSp...	Number...	PreferedOrd...	Sa
Row0	50001	1	4	Phone	3	6	Debit Card	Female	3	3	Laptop & Accessory	2
Row1	50002	1	?	Phone	1	8	UPI	Male	3	4	Mobile	3
Row2	50003	1	?	Phone	1	30	Debit Card	Male	2	4	Mobile	3
Row3	50004	1	0	Phone	3	15	Debit Card	Male	2	4	Laptop & Accessory	5
Row4	50005	1	0	Phone	1	12	Credit Card	Male	?	3	Mobile	5
Row5	50006	1	0	Computer	1	22	Debit Card	Female	3	5	Mobile	5
Row6	50007	1	?	Phone	3	11	Cash on Deli...	Male	2	3	Laptop & Accessory	2
Row7	50008	1	?	Phone	1	6	Credit Card	Male	3	3	Mobile	2
Row8	50009	1	13	Phone	3	9	E wallet	Male	?	4	Mobile	3
Row9	50010	1	?	Phone	1	31	Debit Card	Male	2	5	Mobile	3
Row10	50011	1	4	Phone	1	18	Cash on Deli...	Female	2	3	Others	3
Row11	50012	1	11	Phone	1	6	Debit Card	Male	3	4	Fashion	3
Row12	50013	1	0	Phone	1	11	Cash on Deli...	Male	2	3	Mobile	3
Row13	50014	1	0	Phone	1	15	Credit Card	Male	3	4	Mobile	3
Row14	50015	1	9	Phone	3	15	Credit Card	Male	3	4	Fashion	2
Row15	50016	1	?	Phone	2	12	UPI	Male	3	3	Mobile	5
Row16	50017	1	0	Computer	1	12	Debit Card	Female	?	4	Mobile	2
Row17	50018	1	0	Phone	3	11	E wallet	Male	2	4	Laptop & Accessory	3
Row18	50019	1	0	Computer	1	13	Debit Card	Male	3	5	Laptop & Accessory	3
Row19	50020	1	19	Phone	1	20	Debit Card	Female	3	3	Mobile	4
Row20	50021	1	0	Phone	3	12	Debit Card	Male	3	5	Laptop & Accessory	3
Row21	50022	1	20	Phone	1	29	Credit Card	Female	3	3	Fashion	2

Figure 2: Imported the dataset and configured the settings.

2. **Added Statistical Views and Pie Chart to Diagram as shown in figure 1.**

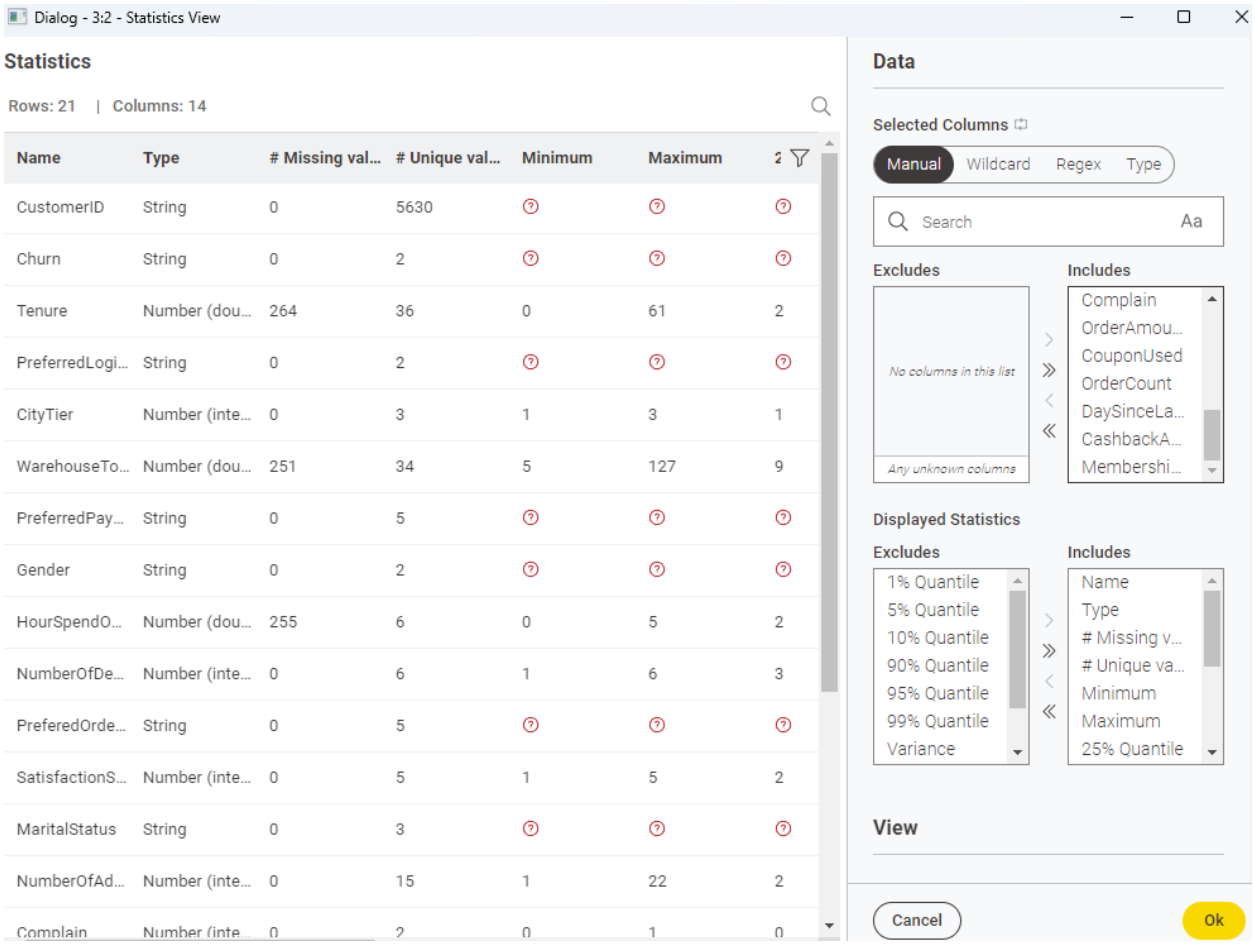


Figure 3: Configured Statistical Views Setting On What Statistics to Display

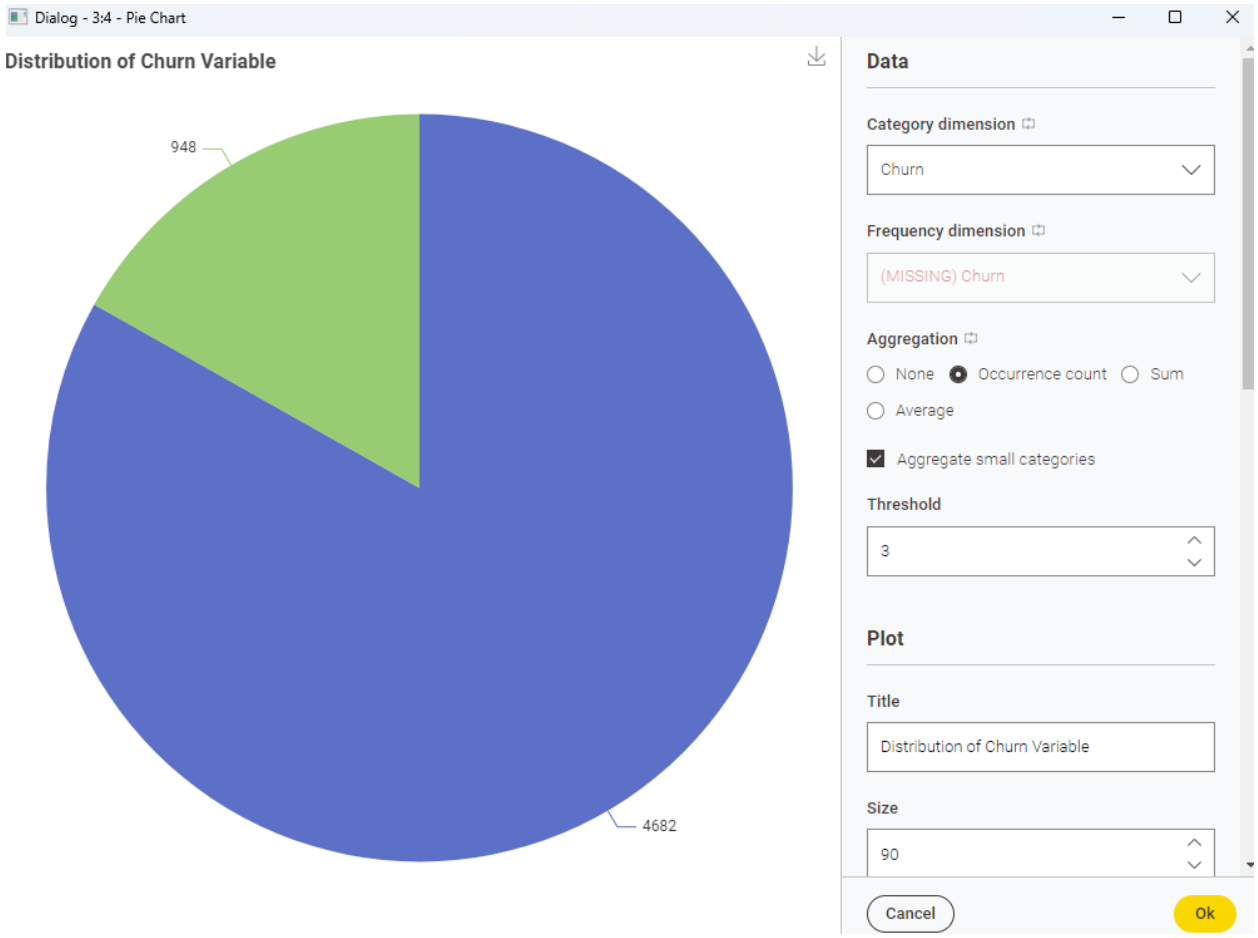


Figure 4: Pie Chart Setting Configuration

3. After determining that the data is imbalanced, Equal Size Sampling Node was added.

Because I needed the Churn Variable to be balanced but with the Churn Variable being an integer, it couldn't be used as the reference category for balancing.

4. Added Number to String Node as shown in figure 1.

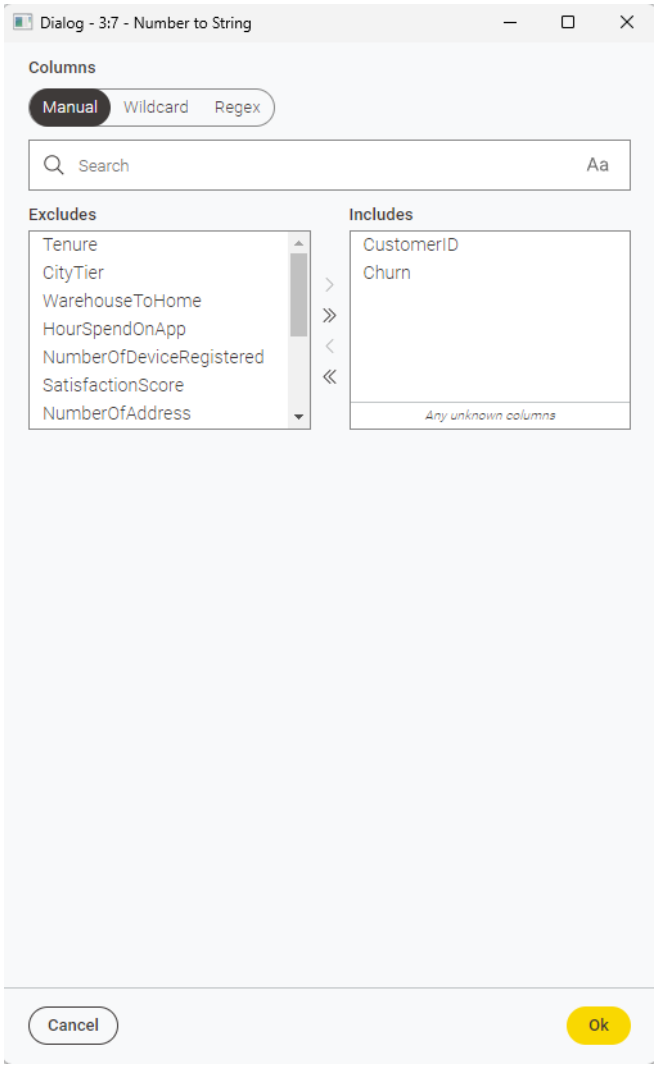


Figure 5: Setting of Number to String Node.

5. Re-configure Settings of Equal Size Sampling Node as shown in figure 6.

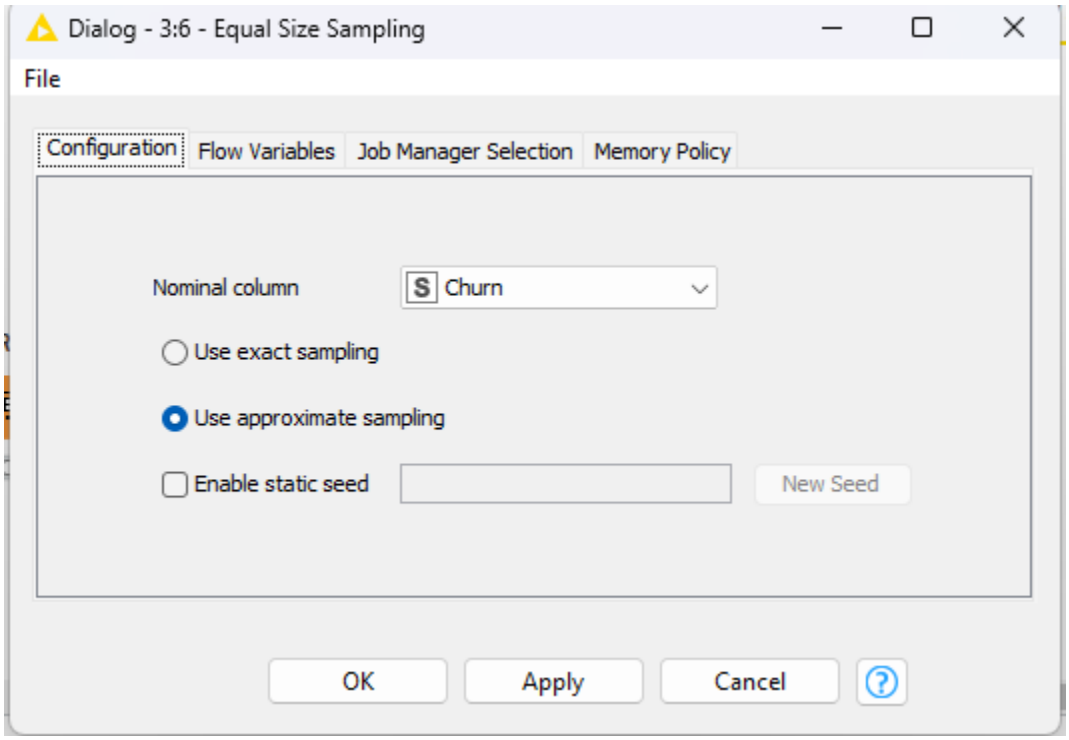


Figure 6: Configuration of Equal Size Sampling

6. Added Statistics View and Pie Chart Node

- Set Configuration exactly the same as the first two Statistics View and Pie Chart to confirm the proportion and distribution after sampling is similar to the population.

7. Added CSV Writer Node

As shown in figure 1, a CSV writer node was added to export the sample file so that it can be imported to SAS e-Miner.