

Babu Banarasi Das University



Case Study

on

Predictive Analysis on COVID-19 data Using SPSS

SUBMITTED TO:

Mr. Robin tyagi sir

SUBMITTED BY:

Name- Gautam kumar yadav

Name- Dhruv Gupta

Agenda/Definition:

This project focuses on predicting future covid-19 case trends using IBM SPSS modeler's data mining and predictive analytics tools.

By leveraging historical data (such as confirmed, recovered, and deaths case), the project applies statistical and machine learning techniques to identify trends and patterns in the spread of the virus.

Outcomes/Learning:

The project successfully predicts future covid -19 case trends using historical data, helping in early planning, healthcare resource management, and preventive decision-making.

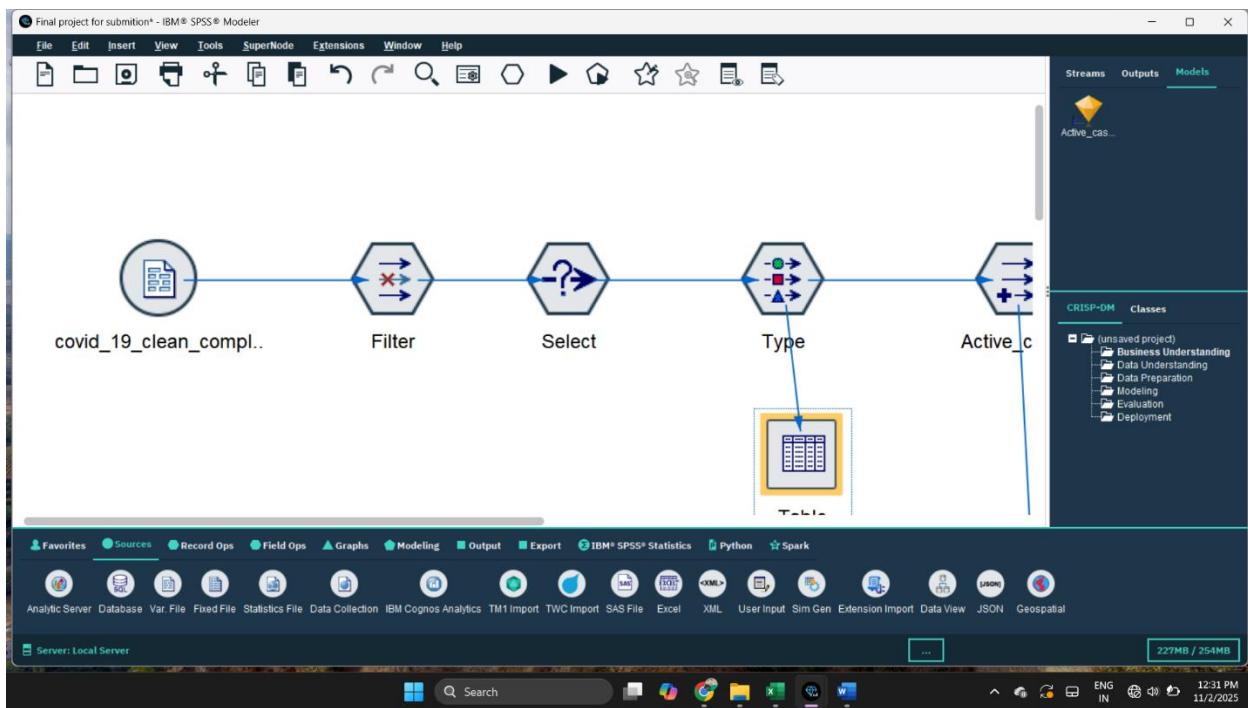
Required Tool: IBM SPSS modeler software

Working:

This project collects historical COVID -19 data, processes it in IBM SPSS modeler using data preparation and modeling nodes, and applies predictive algorithms to forecast future case trends.

Step 1:

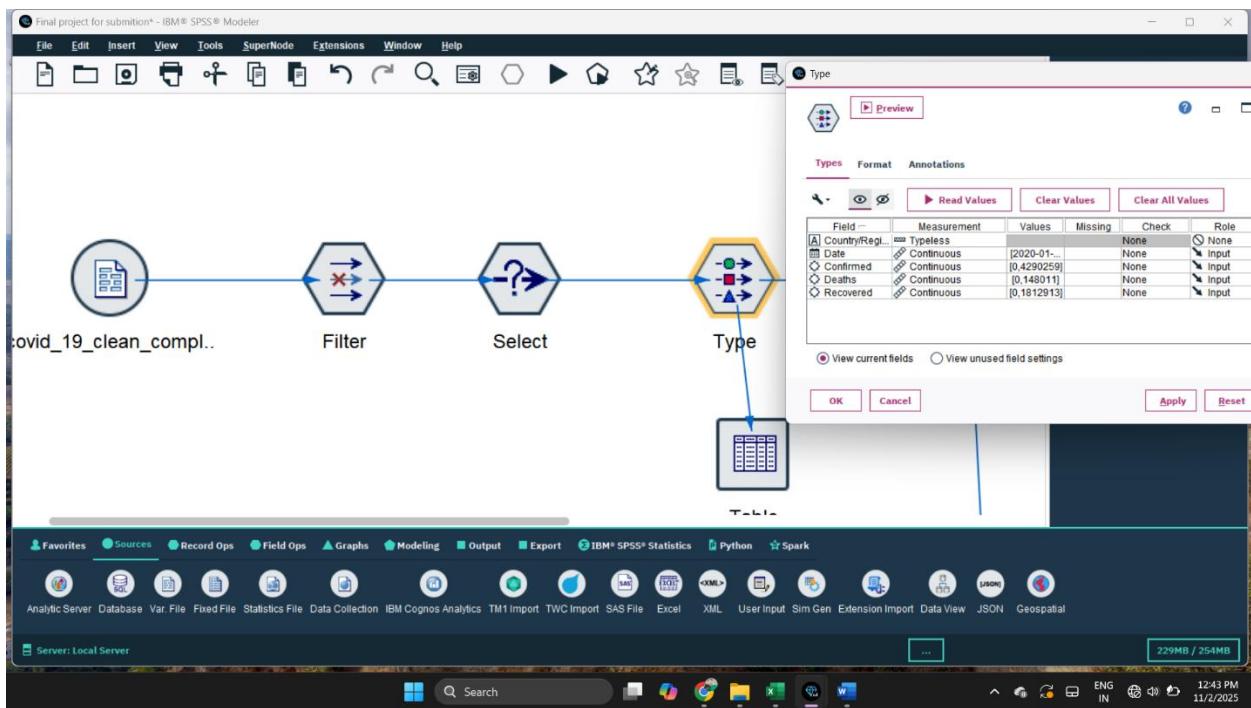
- Importing data



- First of all, you have to take a var file to import covid-19.csv data.
- After that you have to take filter node and connect from var file to filter fields(like Country/regions, Recovered, deaths, Date, Conformed).
- Then take select node to choose relevant columns(like – Date, Confirmed, Deaths)

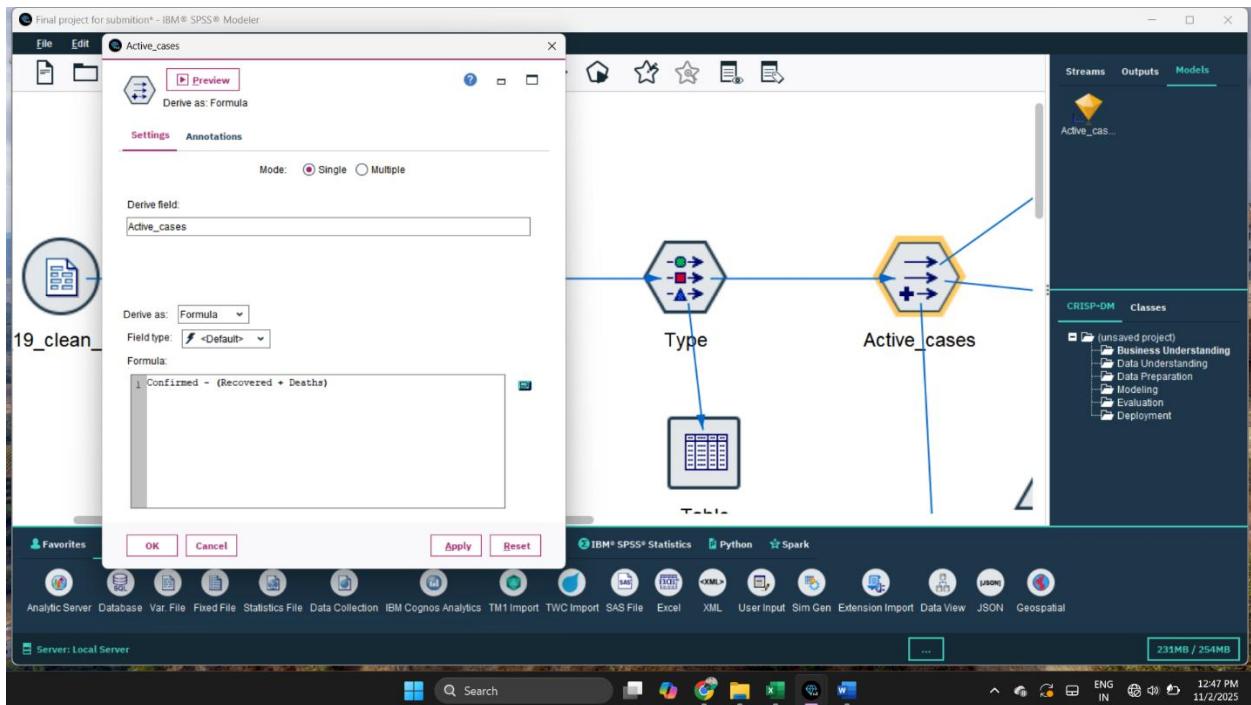
Step 2:

- **Data preparation**



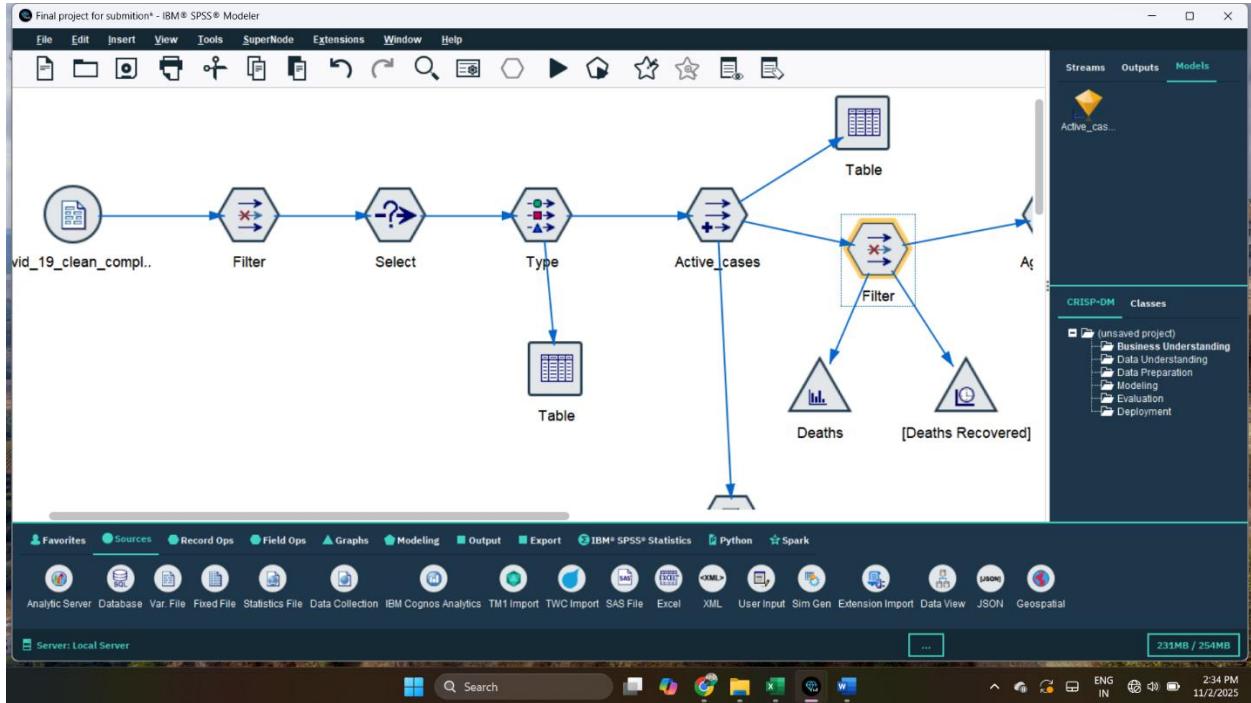
- Then take type node and connect from select node to define measurement levels(continuous, categorical, etc.)

Step 3:



- Then take derive node to create new columns like “Active cases = Conformed – (Recovered + Deaths)”.
- Connect a table to view new columns.

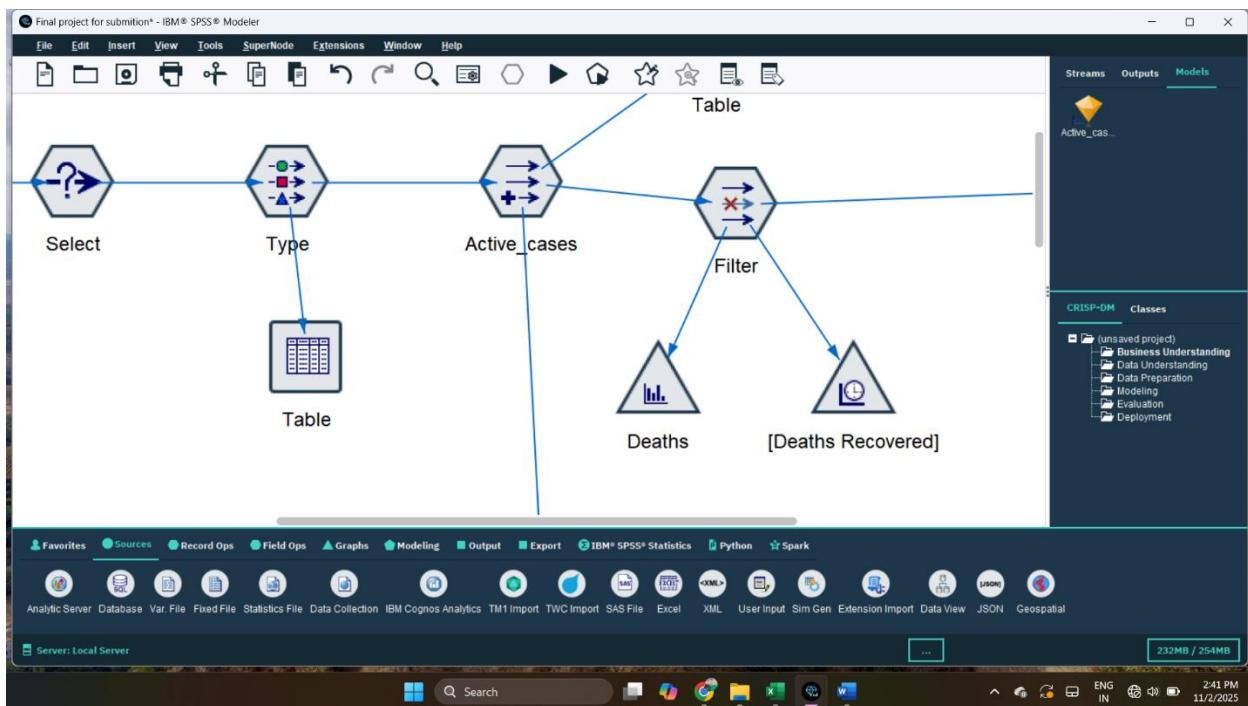
Step 4:



- Connect filter node to remove missing or invalid records.

Step 5:

- **Data Exploration**



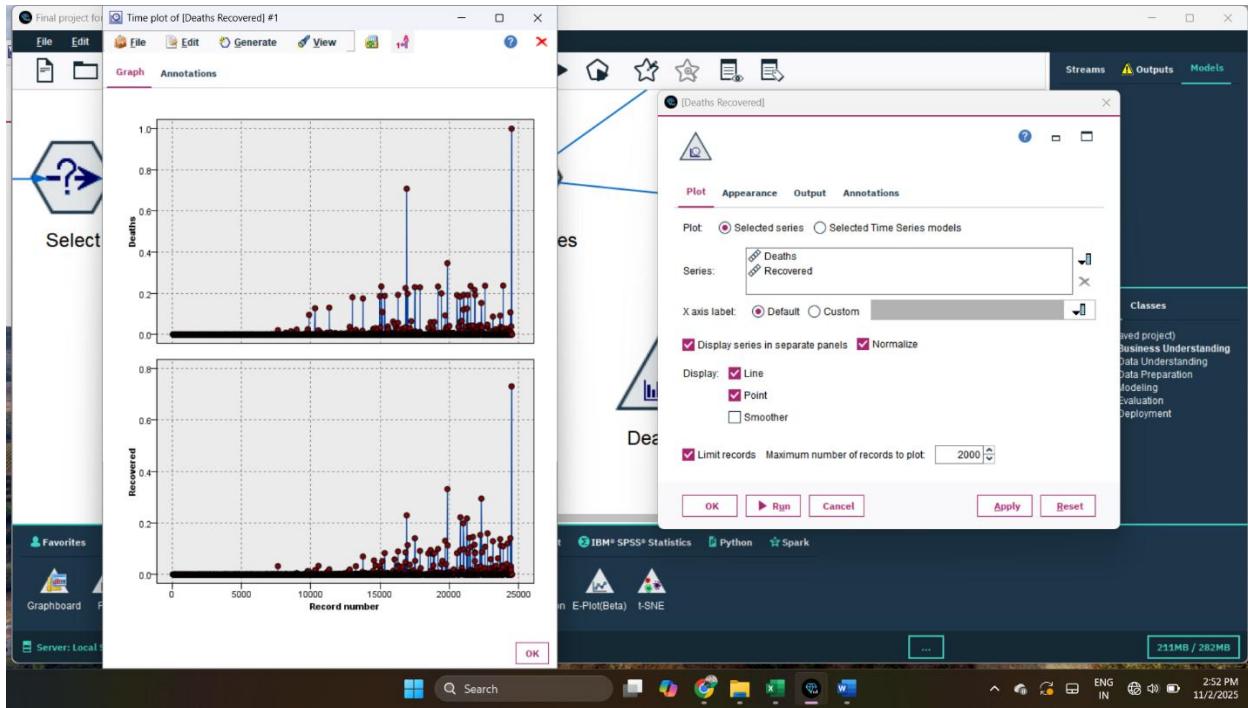
- Here you have to take to charts histogram and time plot

Step 6:

In the fields section, you have to take Deaths because you are going to visualize trends in deaths over time in histogram

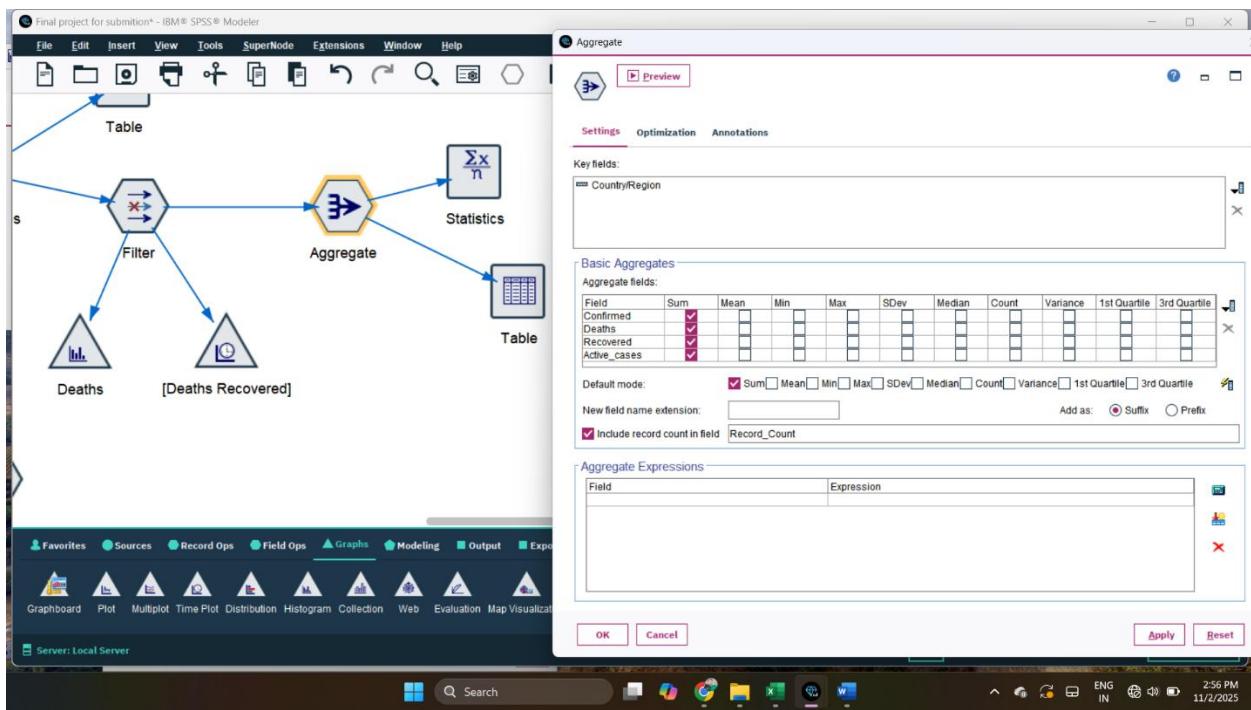
- In the fields section, you have to take Deaths because you are going to visualize trends in deaths over time in histogram

Step 7:

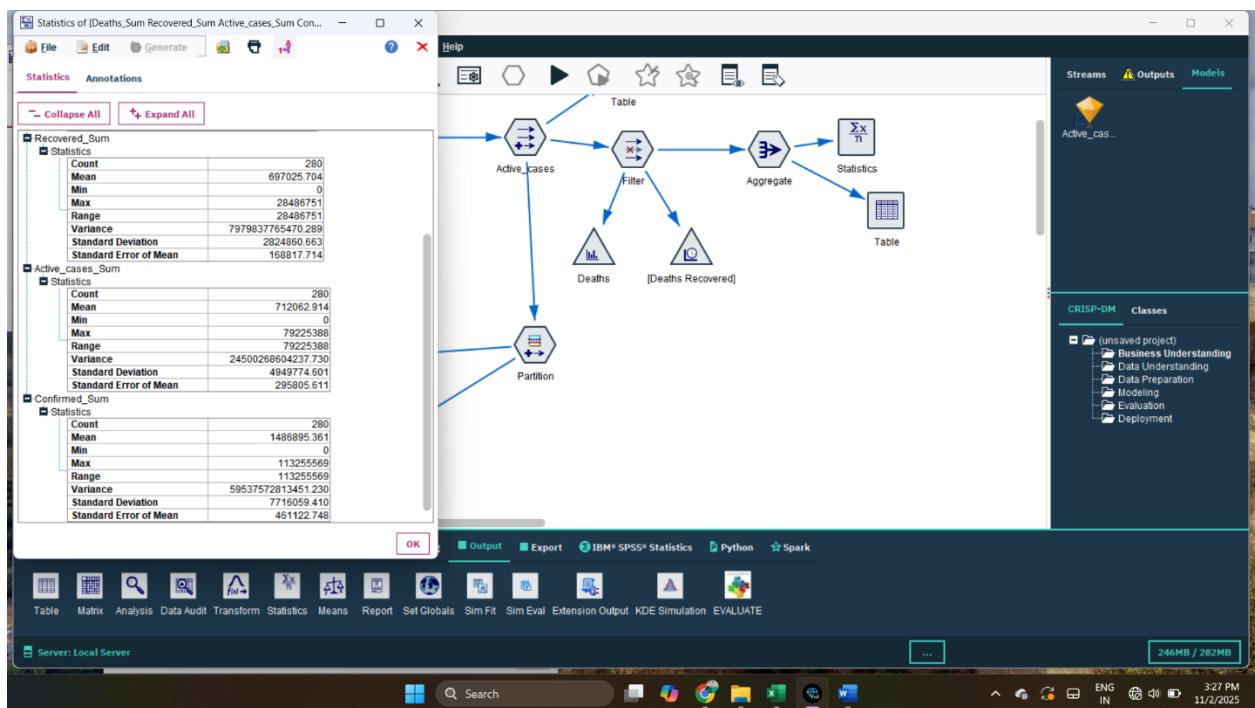


- In the time plot you have to take Deaths and Recovered in series to view trends

Step 8:

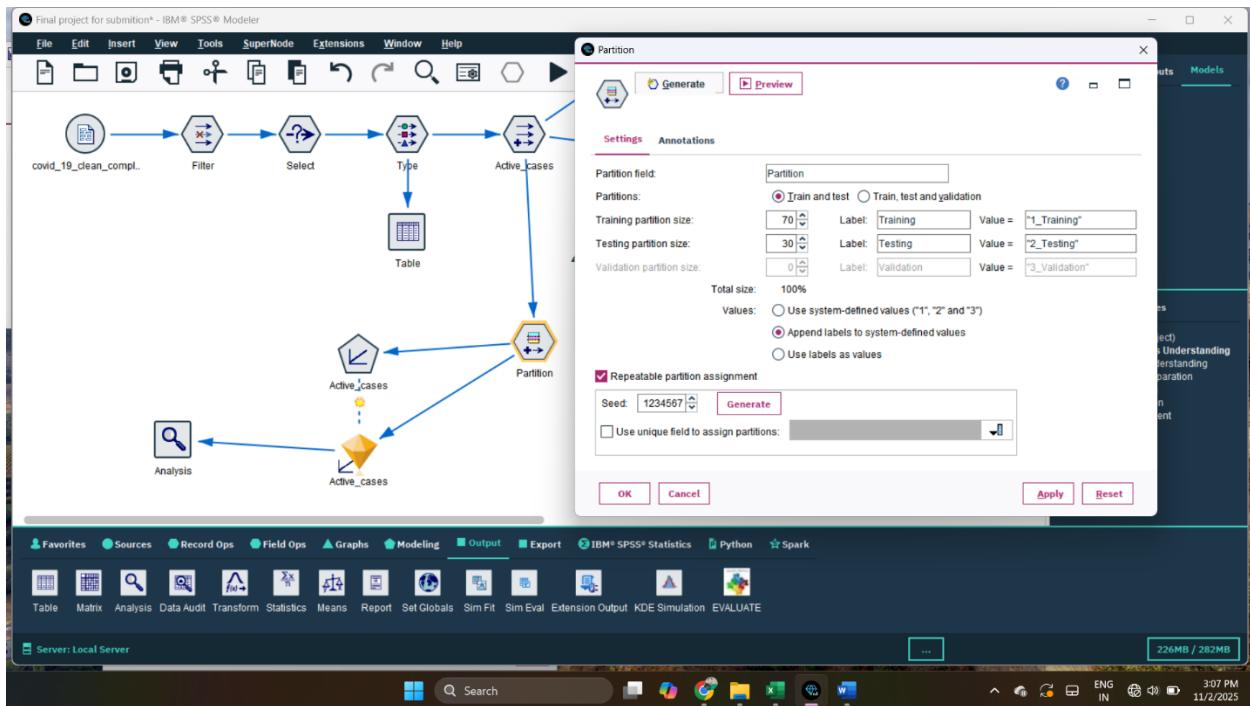


- Connect aggregate node from filter node to check country wise totals like this.
- And connect statistics node to get mean, max, min values.

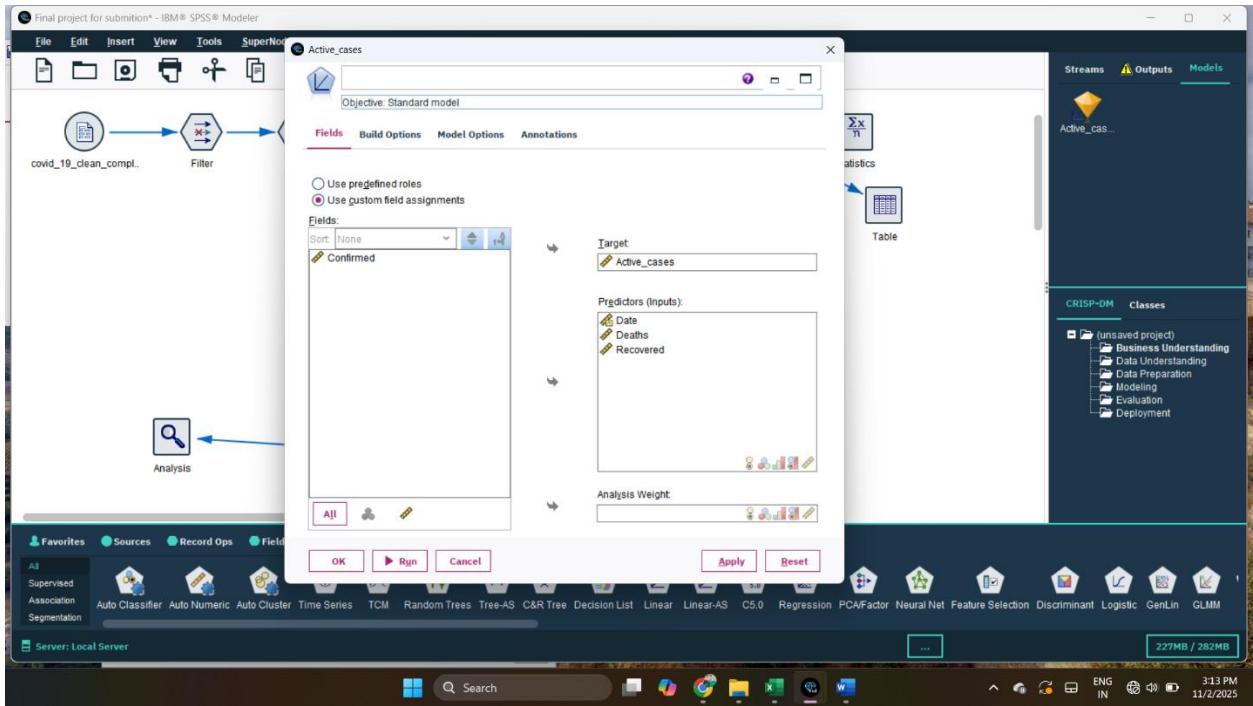


Step 9:

Predictive modeling



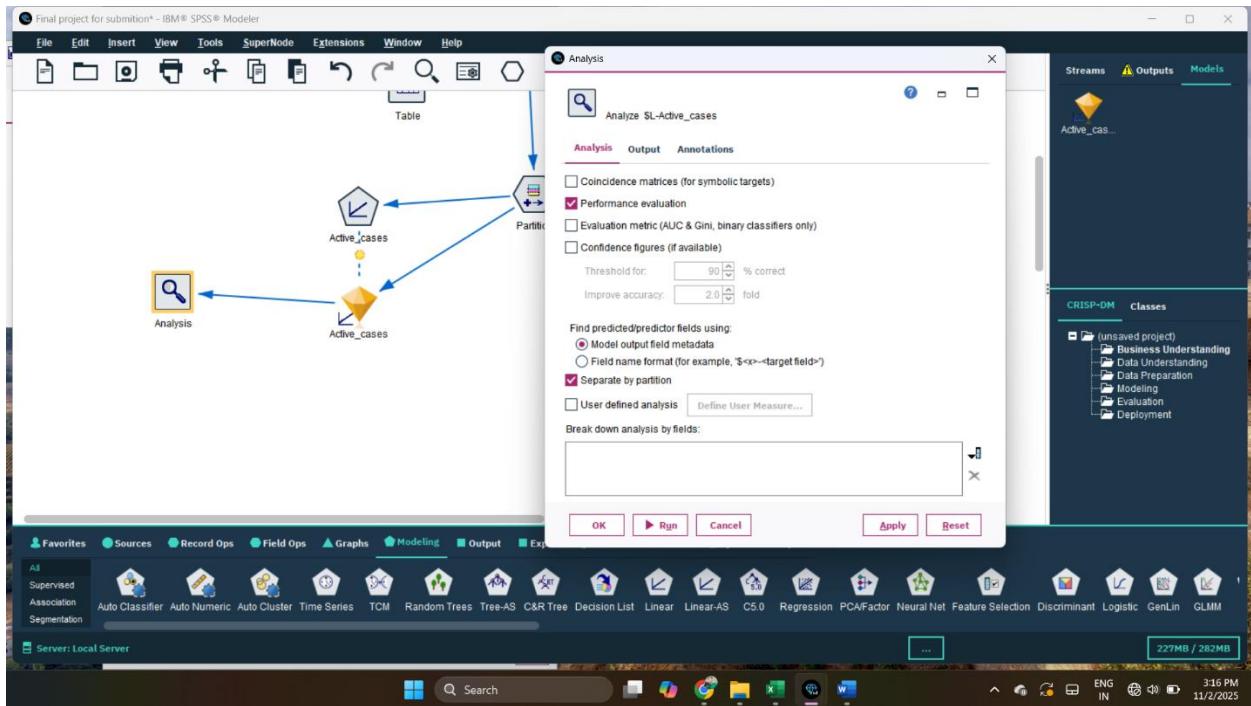
- Take partition node connect from aggregate node to split data set (e.g., 70% for training, 30% for testing) as given above.
- Then take linear Regression model for the modeling of the data set.



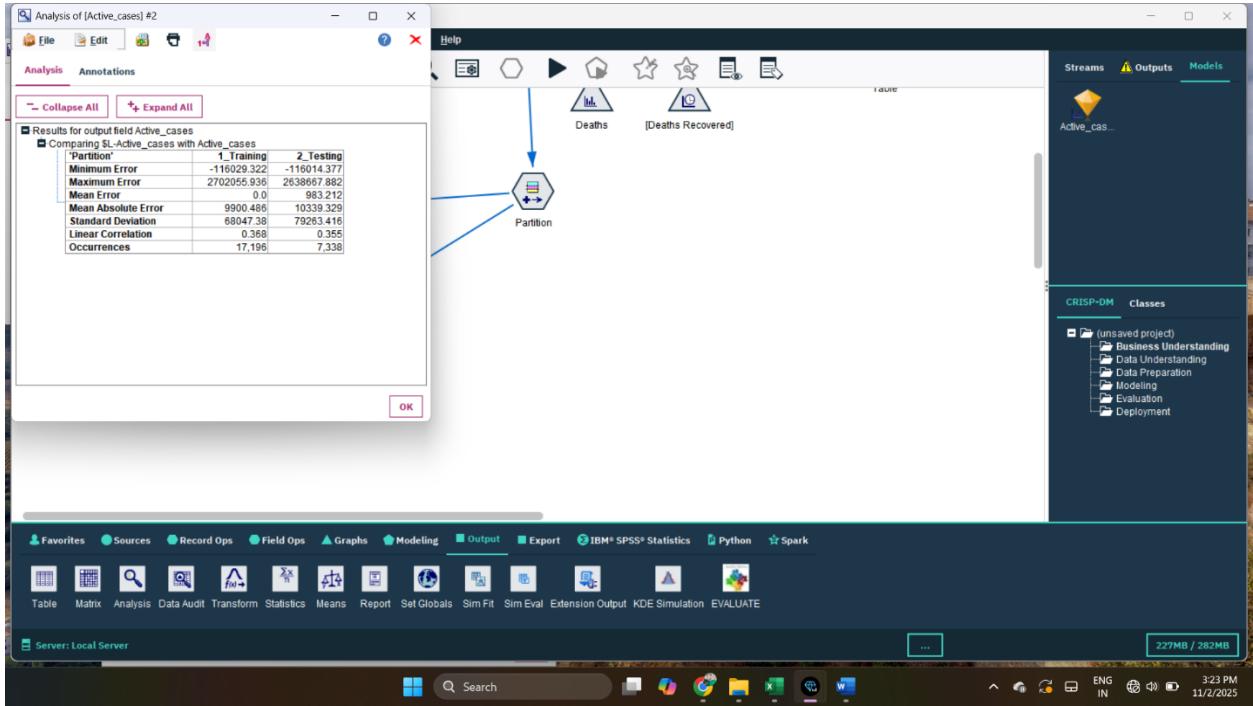
- And set like this target should be Active cases .
- Then apply and run

Step 10:

- Model Evaluation



- After modeling you have to take analysis node from output section.
- And only check and on performance evaluation and separate by partition
- Apply and run



- You can see the output above.

Conclusion:

the model helps to forecast upcoming COVID -19 case trends, assisting in planning and prevention strategies