**Sec/Dept: BSCS 6-A**

**Course: DATA MINING**

**Group Members: NISHA RAMZAN (19122089)**

**AFIFA EJAZ (19122123)**

**Submitted to: Dr.Uzma**

# **Objective:**

In this study, we are using methods Time series and Regression analysis to learn more about earthquakes in Pakistan. We aim to figure out if there are certain patterns or reasons why earthquakes happen, especially over time. By doing this, we hope to help with things like predicting earthquakes and understanding the risks better. Our main goals are to find out if there are any trends in earthquake data, understand how different factors affect earthquakes, and improve our knowledge to better prepare for and handle earthquakes in Pakistan.

# **Introduction and Background:**

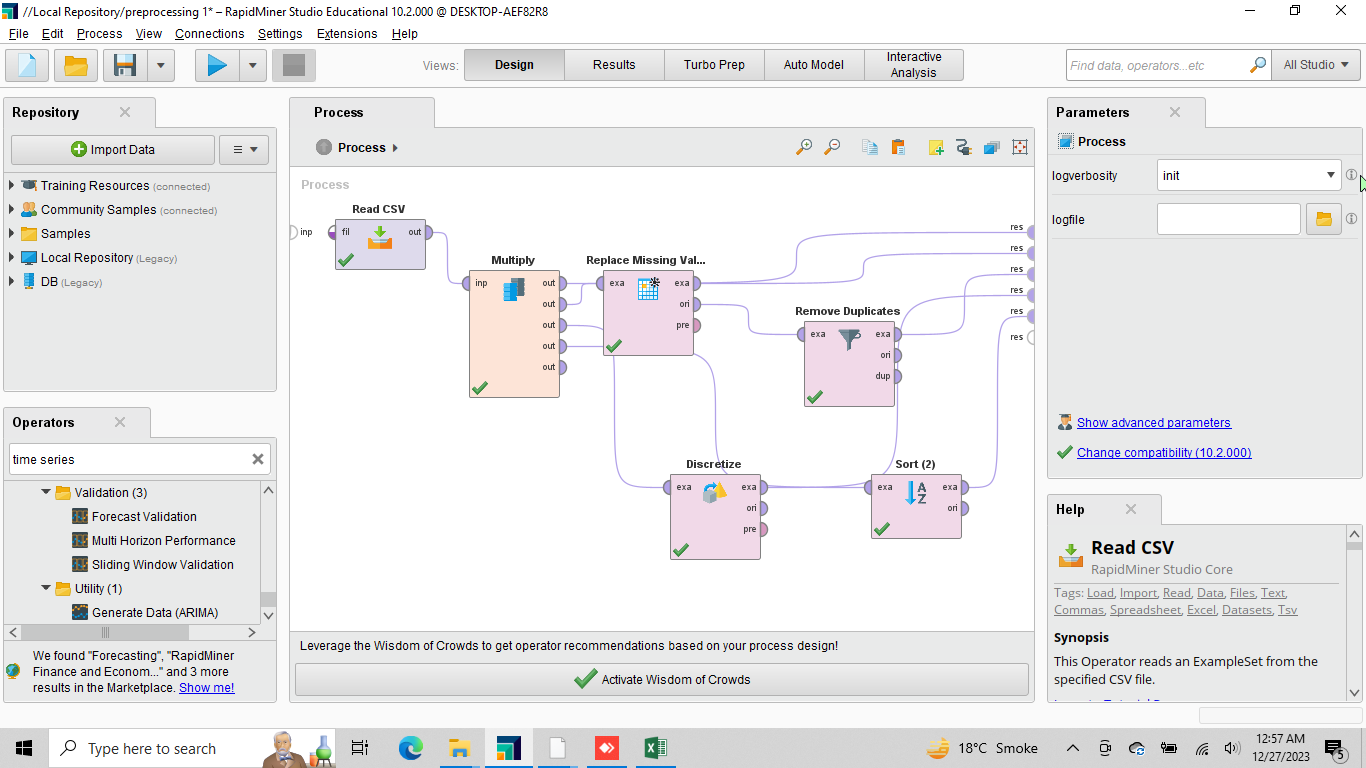
Earthquakes are natural events that can have significant impacts on communities and infrastructure. In the context of Pakistan, understanding the patterns and reasons behind earthquakes is crucial for better preparation and risk management. This study focuses on exploring earthquake data using special techniques, like time series and regression analysis, to uncover insights into when and why earthquakes happen. Pakistan, located in a seismically active region, faces unique challenges related to earthquake risk. By delving into the data, we aim to contribute valuable knowledge that can enhance our ability to predict and respond to seismic events, ultimately promoting safer communities in the face of natural disasters.

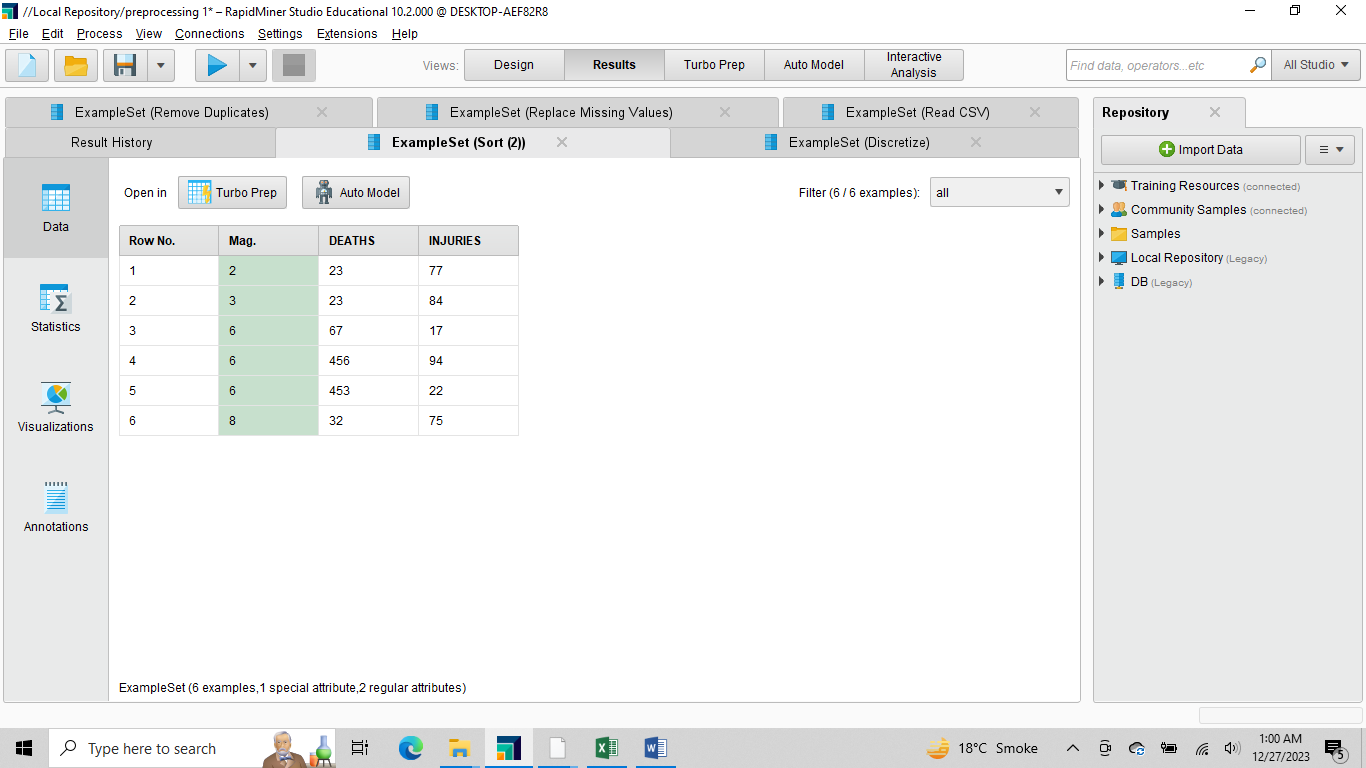
# **Data Collection:**

This data is collected from the website [www.kaggle.com/](http://www.kaggle.com/) , the data collected is attached named (Earthquakes in Pakistan ORIGINAL)

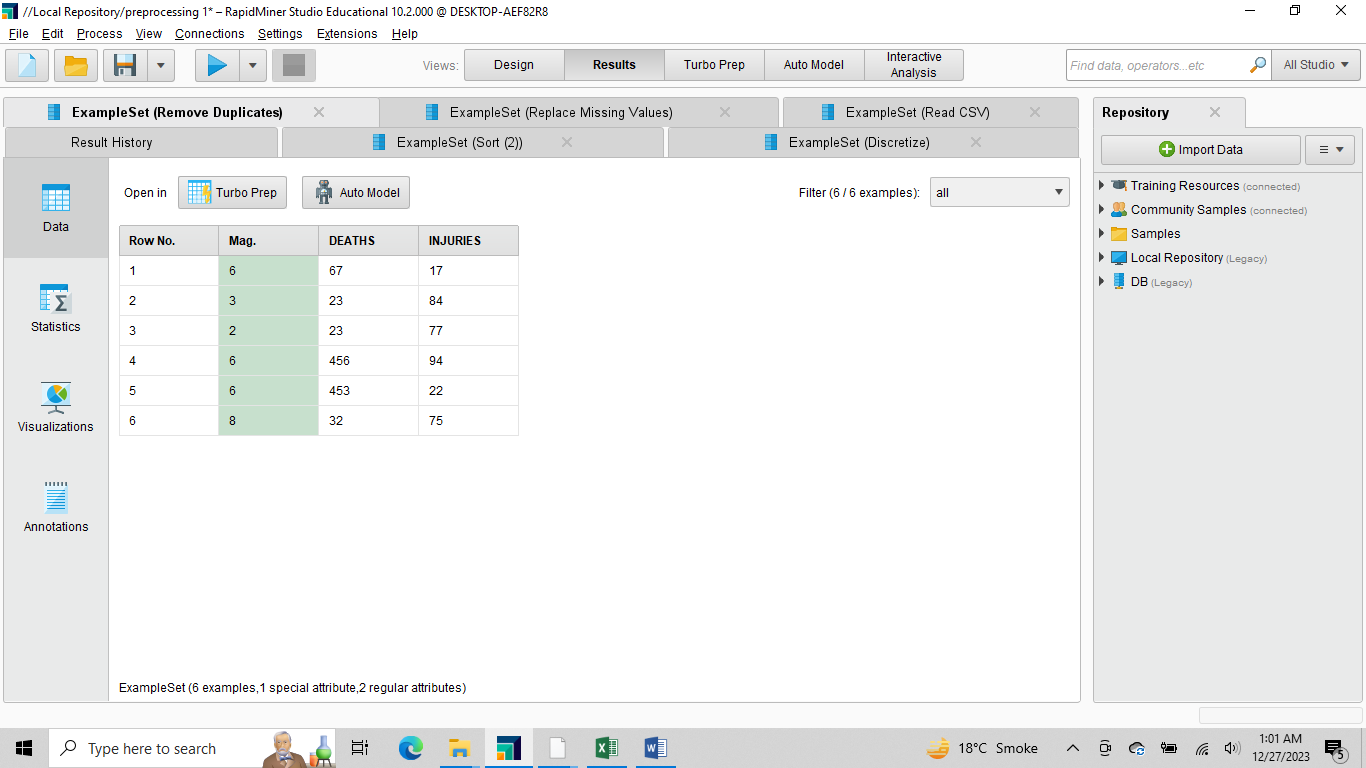
# **Data Preprocessing:**

In preparing our earthquake data for analysis, we undertook several crucial steps to ensure the accuracy and reliability of our results.( Earthquakes in Pakistan UNCLEANED)

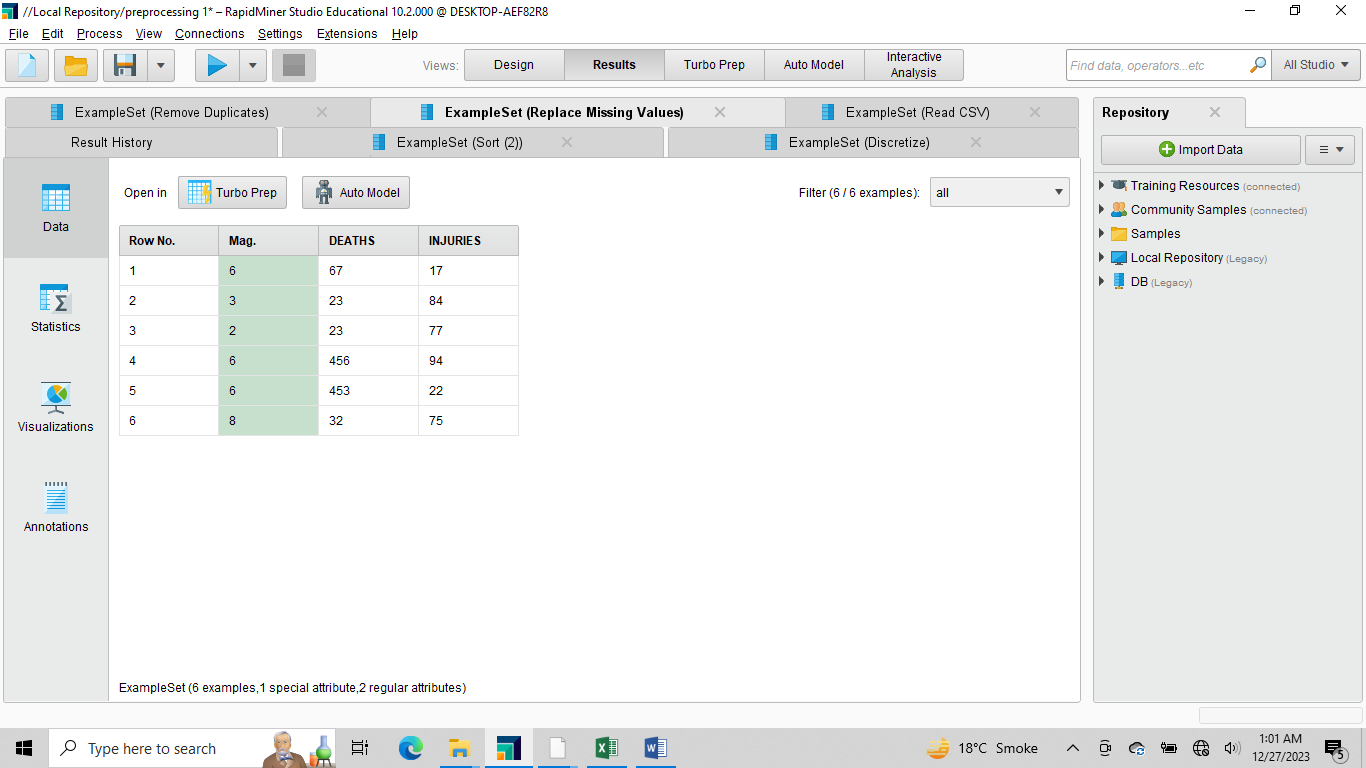




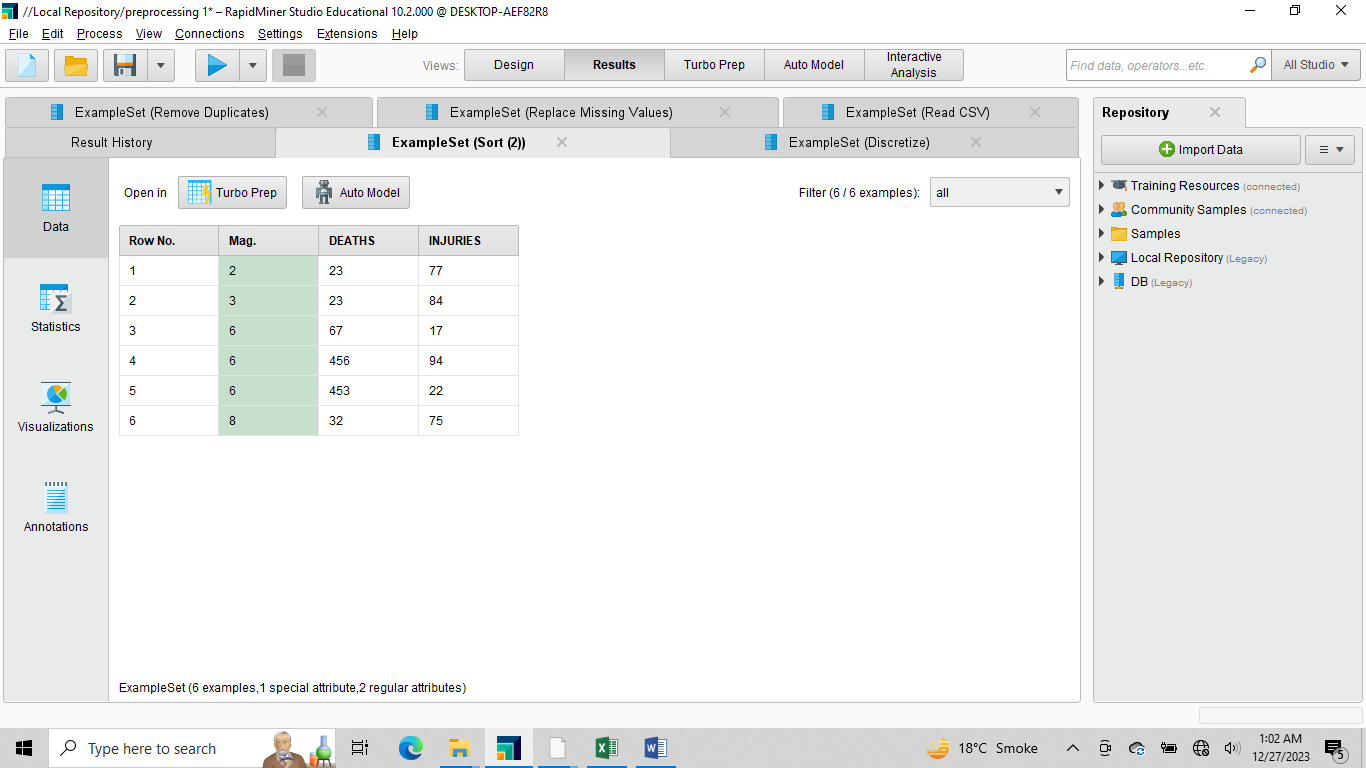
## **Removing duplicates:** Duplicate entries were identified and removed from the dataset. This step is crucial to prevent repetition and maintain the integrity of our analysis.



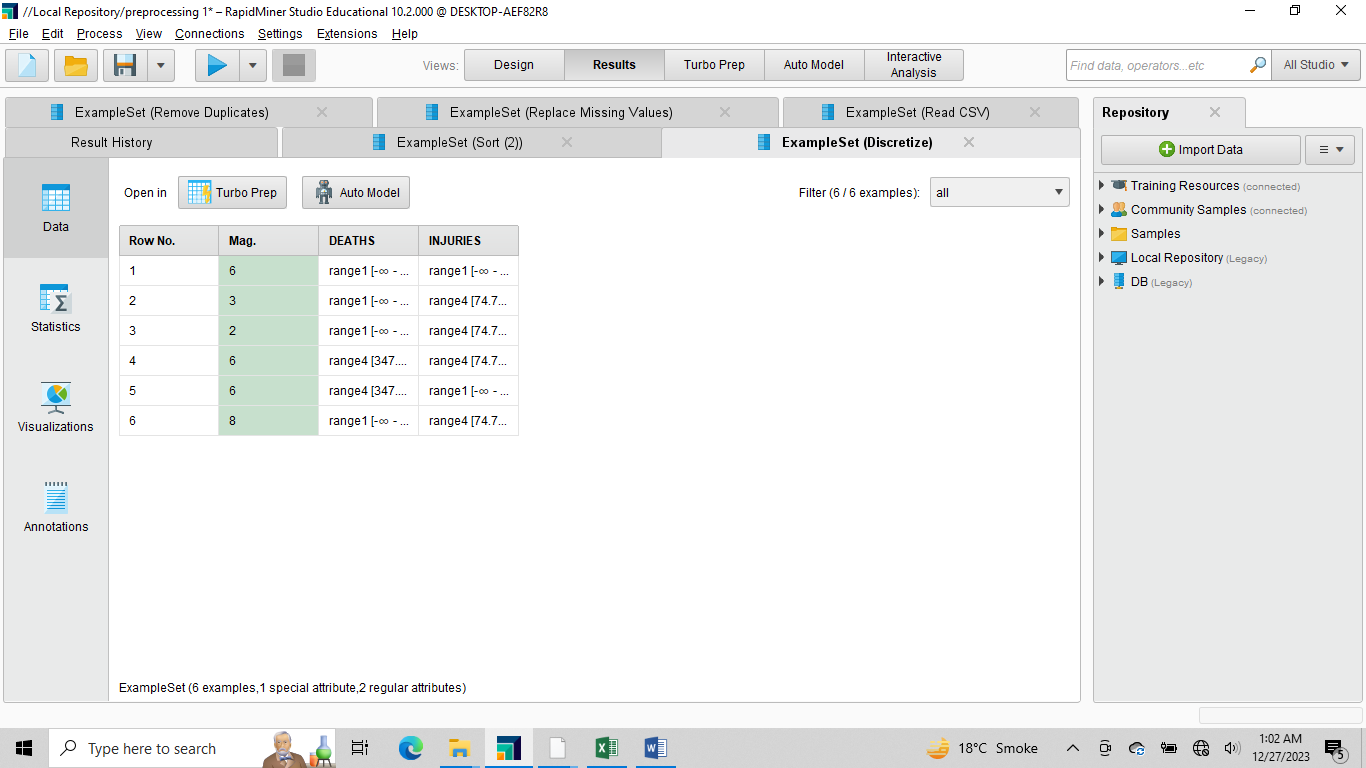
## **Replacing missing values:** We made sure that our data didn't have any missing pieces. Any gaps in the information were either filled in or, if needed, we decided how to best deal with them.



## **Sorting by Magnitude:** To get a clearer picture of our earthquake data, we sorted it based on magnitude values. This helps us see how earthquakes of different sizes are distributed and understand their impact more effectively.



## **Discretization:** For some aspects of our study, we grouped continuous data into categories. In our case, we did this with certain variables, like earthquake magnitudes, to make it easier to analyze patterns and trends.

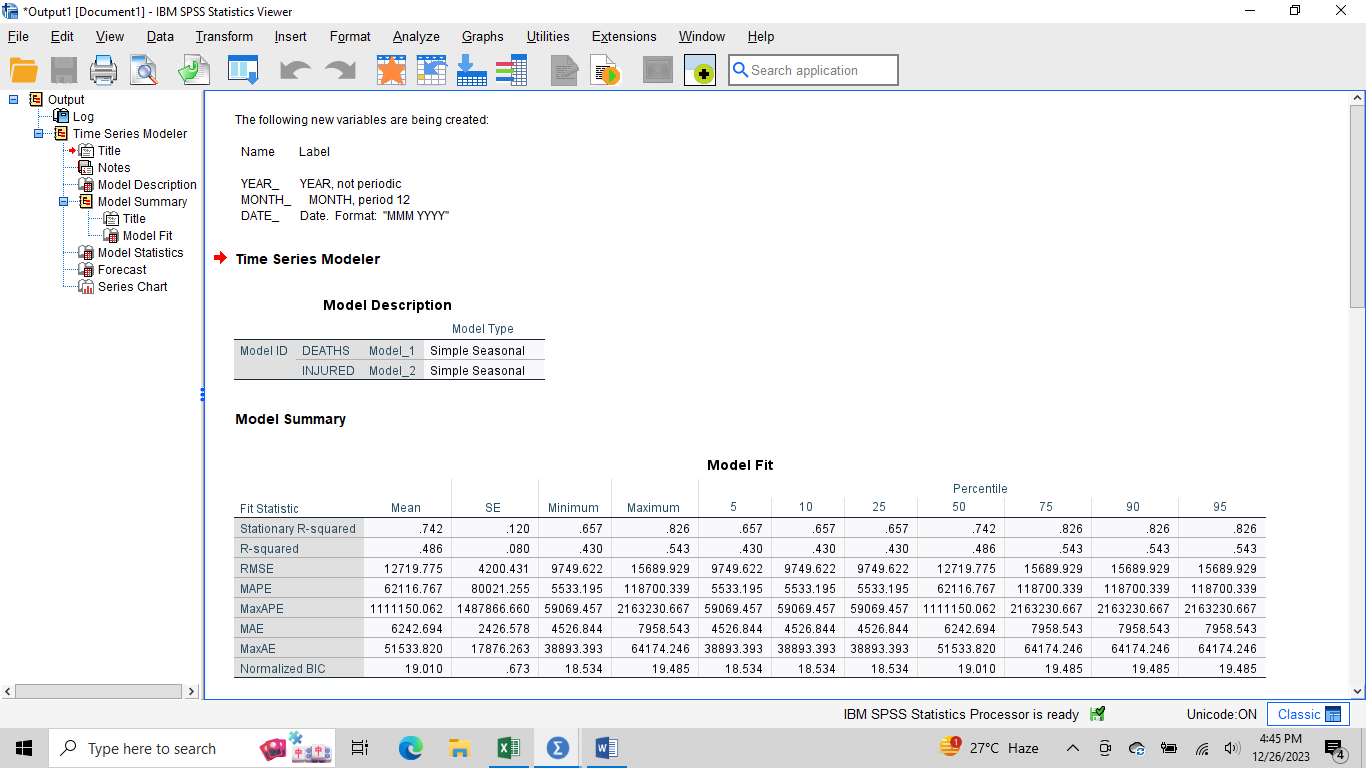


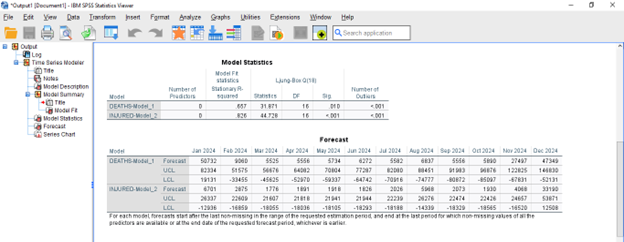
# **Modelling And Evaluation:**

(Earthquakes in Pakistan CLEANED)

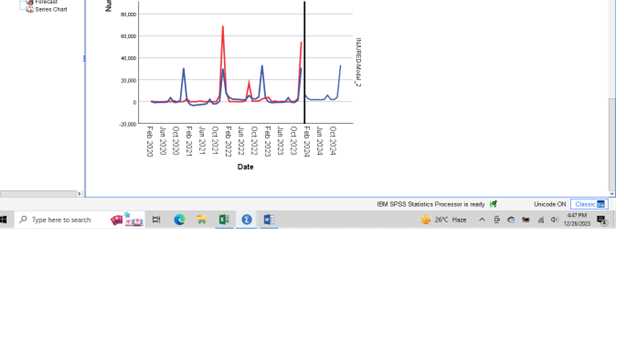
## **Time Series Analysis:**

In our time series analysis, where we utilized a seasonal model with date formats represented as (month, year), we focused on predicting the occurrences of injuries and deaths due to earthquakes. We used a method called model extract, and the specific model type employed was a simple seasonal model.





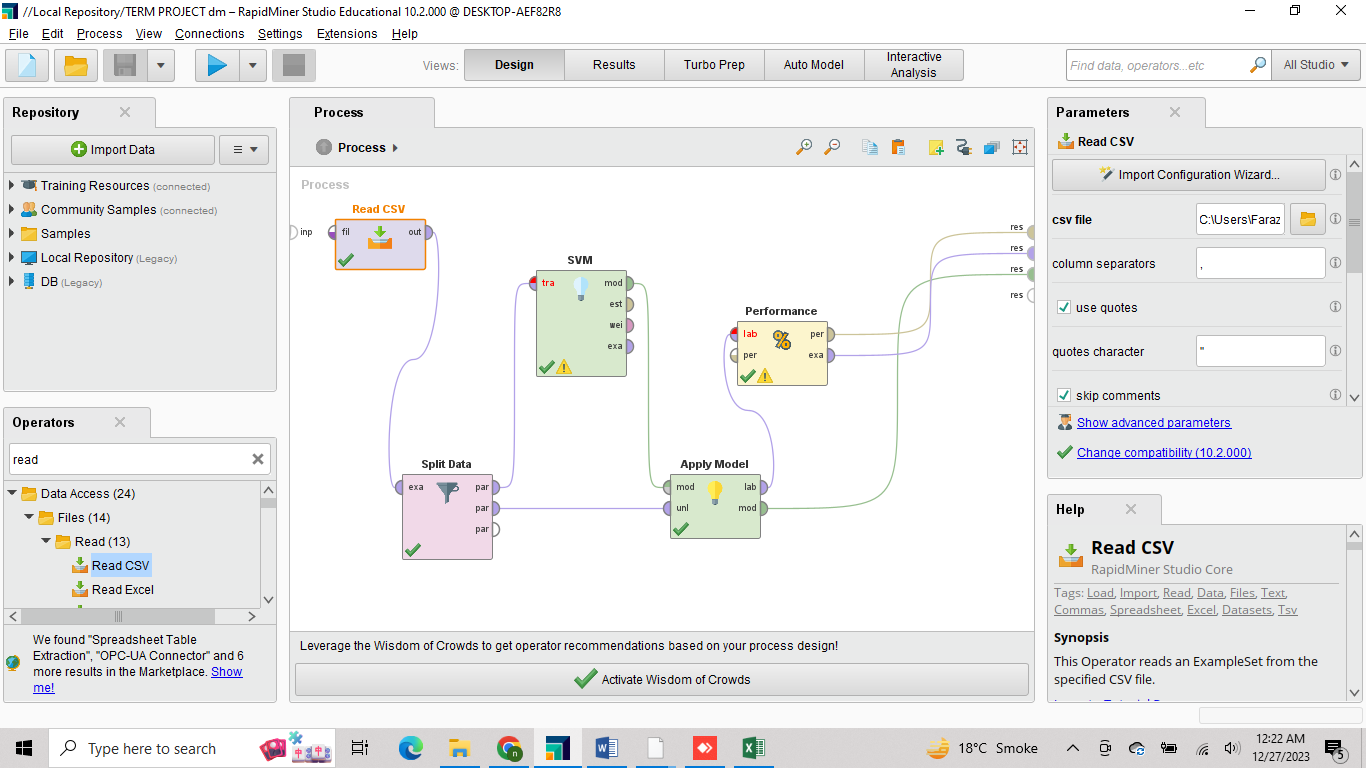


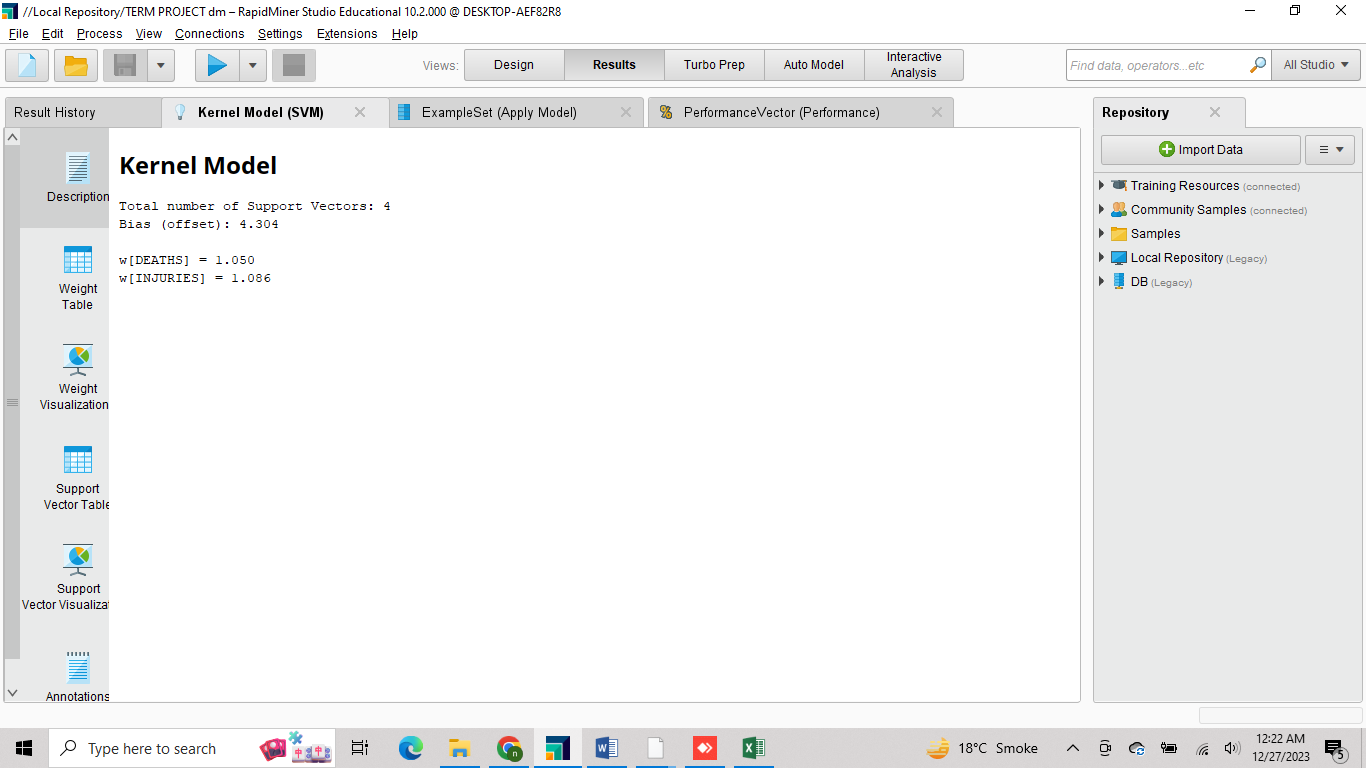


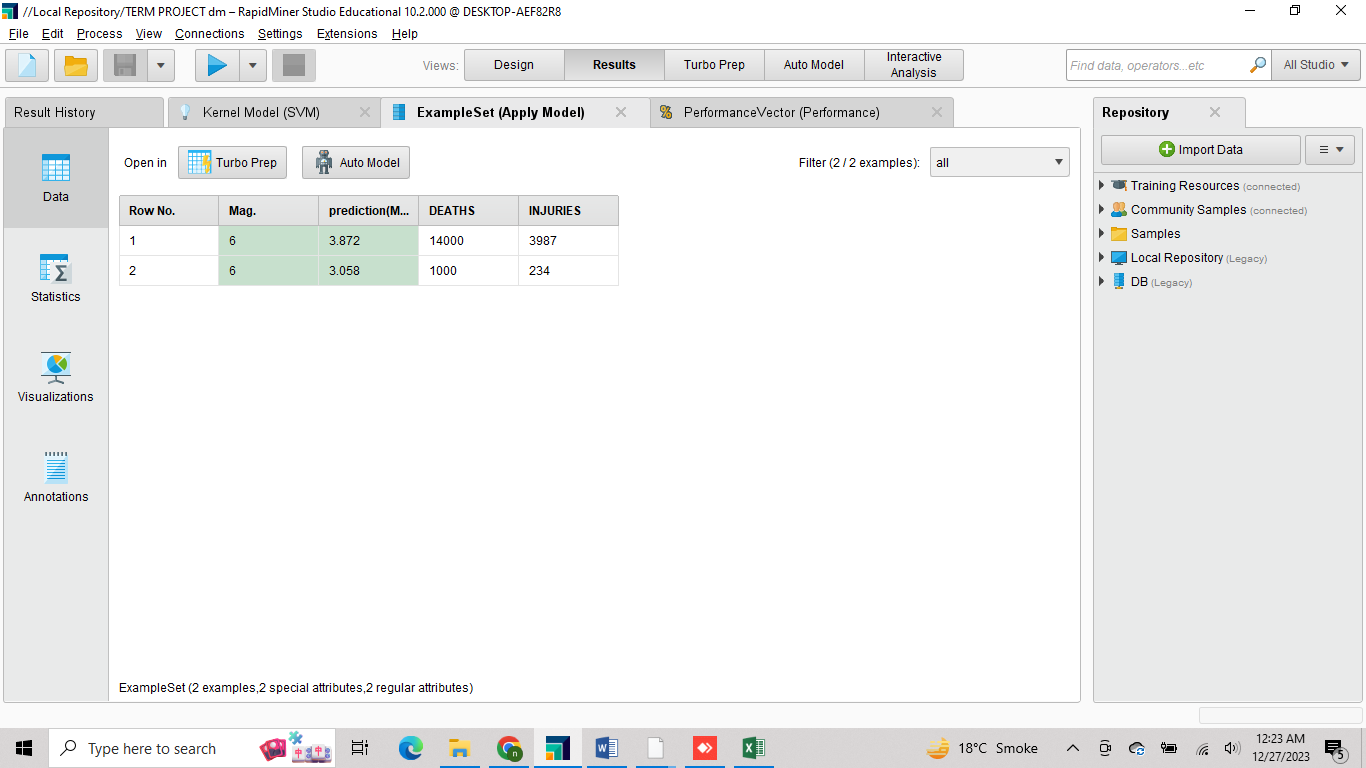
## **Regression Analysis with SVM :**

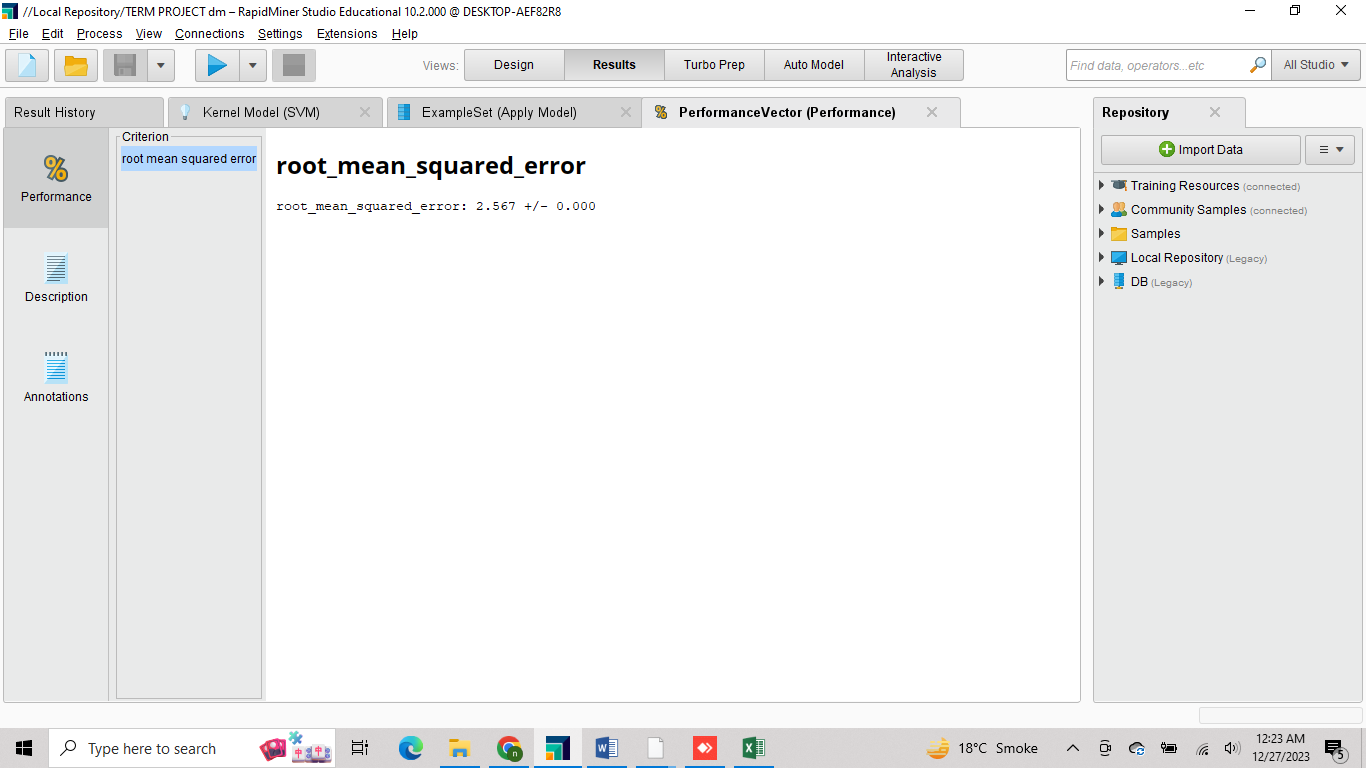
In our study, we used Support Vector Machines (SVM) for regression analysis to predict earthquake magnitudes. Specifically, we focused on understanding how the number of deaths and injuries related to earthquakes might give us clues about their size in Pakistan.

SVM acted like a smart detective, helping us uncover connections between earthquake magnitudes and the human impact they caused. Our findings shed light on the relationship between the labeled magnitudes and the recorded incidents of deaths and injuries.









## **Results:**

In our Support Vector Machines (SVM) regression analysis, we achieved a Root Mean Squared Error (RMSE) of 2.567. The RMSE is a measure of how well our SVM model predicted earthquake magnitudes based on the recorded incidents of deaths and injuries.

For Time series the results revealed clear and meaningful patterns in the occurrence of injuries and deaths over different months and years. The seasonal model effectively captured the recurring nature of these incidents, highlighting specific times when they tend to increase or decrease.

## **Conclusion:**

In our exploration of earthquakes in Pakistan, we sourced our data from Kaggle, a platform known for hosting diverse datasets. This dataset encompassed crucial information regarding earthquake occurrences, magnitudes, and human impacts. To ensure the reliability of our findings, we initiated a thorough preprocessing phase, addressing missing values, sorting the data by magnitude, removing duplicates, and discretizing certain variables. These steps laid the groundwork for our subsequent analyses.

In the realm of regression analysis, the implementation of Support Vector Machines (SVM) unveiled a noteworthy connection between earthquake magnitudes and human impact factors, achieving a Root Mean Squared Error (RMSE) of 2.567. Transitioning to time series analysis, a simple seasonal model illuminated recurring patterns in injuries and deaths over different time periods.

In a nutshell, our study uncovered the secrets of earthquakes in Pakistan, blending how they affect people and understanding their patterns over time. This journey opens the door for more exploration, helping us prepare better for the unpredictable dance between nature and our communities.