TITLE: "Rain Forecasting for June-July Using December Rainfall Data"

SUPERVISOR: T. V. Bharat

BY: PRANAY KUMAR PARATE, 210104070

Introduction:

Objective of the Study:

- The primary goal of this project is to develop a predictive model for forecasting rainfall during the months of June and July, using historical data from the previous December. Accurate rain forecasting is crucial for agriculture, urban planning, and disaster management.

Significance:

- Early predictions help in making informed decisions regarding water resource management and preparation for potential flooding events.

Overview of Methodology:

- Data sources include rainfall measurements, wind data, humidity levels, atmospheric pressure, and sea surface temperature. A regression model was applied to predict rainfall, but the results did not meet expectations.

Data Collection:

Data Used:

- **Rainfall Data:** Monthly rainfall data (in mm) for December.
- **Wind Data:** Wind speed and direction, which influence weather patterns.
- **Humidity:** Humidity levels, a key factor in cloud formation and precipitation.
- **Atmospheric Pressure:** Pressure changes that can indicate upcoming weather conditions.
- **Sea Surface Temperature (SST):** Ocean temperatures that affect climate conditions.

Data Source: **IBM Data**

- I worked with data provided by my PhD advisor from IBM, which offered insights into broader climatic patterns.

Methodology:

- **Model Selection:**
    - Regression models, including linear regression and possibly more advanced forms like polynomial or multiple regression, were chosen due to their ability to predict continuous outcomes such as rainfall.
- **Why Regression?**
    - Regression models are widely used for weather prediction due to their ability to establish relationships between multiple variables and forecast future outcomes.
- **Implementation:**
    - The model was trained using December's data as the input features, with the target variable being the rainfall in the upcoming June-July period.

Results and Challenges:

SUMMARY OUTPUT

Regression Statistics	
Multiple	0.27364
R Square	0.07488
Adjusted	-0.2615
Standard	5.02264
Observat	16

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regressic	4	22.4613	5.61532	0.22259	0.92018
Residual	11	277.496	25.2269		
Total	15	299.958			

	Coefficient	*andard Err*	*t Stat*	*P-value*	*.ower 95%*	*Jpper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	9.24826	1507.36	0.00614	0.99521	-3308.4	3326.92	-3308.4	3326.92
X Variabl	0.00139	1.50275	0.00093	0.99928	-3.3061	3.30893	-3.3061	3.30893
X Variabl	-0.0649	0.23029	-0.2817	0.78337	-0.5717	0.44198	-0.5717	0.44198
X Variabl	-0.4795	0.68507	-0.7	0.49847	-1.9874	1.02829	-1.9874	1.02829
X Variabl	-0.0247	0.3747	-0.066	0.94856	-0.8494	0.79998	-0.8494	0.79998

Dependent variables: X1=sea level pressure, X2=relative humidity, X3=wind speed, X4=rainfall in December month

Independent variable: rainfall during July month

- **Model Results:**
    - The regression model produced predictions, but the results were not as accurate as expected. Metrics like Root Mean Squared Error (RMSE) and R-squared indicated room for improvement.
- **Challenges:**
    - One major challenge was overfitting, where the model performed well on the training data but poorly on new data. Additionally, some key variables may not have been fully captured by the model.
- **Unexpected Findings:**
    - The complexity of weather systems and potential missing interactions between variables could have contributed to the model's underperformance.