

# **PM Accelerator Data Analysis Report**

## **1. Introduction**

The purpose of this report is to analyze data related to the PM Accelerator mission, including data cleaning, exploratory data analysis (EDA), forecasting models, and advanced analyses. The goal is to derive insights that align with the initiative's objectives.

## **2. Data Cleaning & Preprocessing**

* **Handle Missing Values:** Applied imputation techniques (mean, median, or mode) or removed missing values where necessary.
* **Outlier Detection:** Used statistical methods and visualization techniques to identify and handle outliers.
* **Data Normalization:** Scaled data to ensure uniformity and improved model performance.

## **3. Exploratory Data Analysis (EDA)**

* **Summary Statistics:** Analyzed key statistics like mean, median, standard deviation, and distributions.
* **Data Visualization:** Used histograms, box plots, and scatter plots to identify trends and relationships.
* **Correlation Analysis:** Identified relationships between key variables using heatmaps and correlation coefficients.
* **Temperature & Precipitation Analysis:** Generated visualizations to understand patterns in climate data.

## **4. Model Building**

* **Time Series Forecasting:** Built a forecasting model using ARIMA, Prophet, and LSTM.
* **Evaluation Metrics:** Assessed performance using RMSE, MAE, and R-squared.
* **Use of Lastupdated Feature:** Integrated the 'lastupdated' feature for accurate time series analysis.

## **5. Advanced Assessment**

### **Advanced EDA**

* **Anomaly Detection:** Implemented techniques to identify and analyze outliers in the dataset.

### **Forecasting with Multiple Models**

* **Model Comparison:** Built and compared multiple forecasting models, including ARIMA, LSTM, and Random Forest.
* **Ensemble Model:** Created an ensemble approach to improve forecast accuracy.

### **Unique Analyses**

* **Climate Analysis:** Studied long-term climate patterns and variations across different regions.
* **Environmental Impact:** Analyzed air quality and its correlation with various weather parameters.
* **Feature Importance:** Applied techniques like SHAP values and permutation importance to assess feature significance.
* **Spatial Analysis:** Visualized geographical patterns and trends in the dataset.
* **Geographical Patterns:** Explored how weather conditions vary across countries and continents.

## **6. Insights & Recommendations**

* **Key Findings:** Summarized insights from the analyses, including trends and significant variables.
* **Actionable Recommendations:** Provided data-driven suggestions to improve decision-making.
* **Future Work:** Identified areas for further research and improvements in data collection.

## **7. Visualizations**

* **Graphs & Charts:** Included relevant charts to illustrate findings (e.g., bar charts, line graphs, and pie charts).
* **Interactive Dashboards:** Suggested use of Tableau/Power BI for stakeholder engagement.

[google colab](https://colab.research.google.com/drive/1eJEacMlSgg6CLPq9zzsCVLmBPAey80Yf?usp=sharing)

<https://colab.research.google.com/drive/1eJEacMlSgg6CLPq9zzsCVLmBPAey80Yf?usp=sharing>

## **8. Conclusion**

This report provides a comprehensive analysis of the data, helping to align with the PM Accelerator mission. The findings can be leveraged for strategic planning and impact assessment.

## **9. Repository Submission**

The full report, datasets, and code are available in the GitHub Repository. The README.md provides detailed documentation of the methodology and results.

*Prepared by:* [Bonthu.sujana]  
*Date:* [15-03-2025]