\*\*Introduction\*\*

In the epoch of environmental consciousness, the significance of renewable energy in bolstering and enriching our lifestyle is paramount. As we navigate towards a sustainable future, comprehending the intricacies of power generation becomes indispensable. This article delves into the Global Power Plant Database, a comprehensive project encapsulating the world's energy infrastructure, and explores its potential to scrutinize and forecast energy production from both renewable and non-renewable sources.

\*\*Problem Definition\*\*

The dataset is a comprehensive compilation of power plant data from India, encompassing details such as the energy source type, geographical location, capacity, and more. The objective is to comprehend, scrutinize, and forecast energy production, with an emphasis on renewables, to facilitate superior planning and decision-making.

Given the dataset's exclusive focus on Indian power plants, attributes like 'country' and 'country\_long' become superfluous as they remain constant. Hence, these were excluded from the analysis. Similarly, the 'url' attribute, which may not contribute significant insights for our specific analytical goals, was also disregarded. This selective attribute inclusion process enhances our analysis's efficiency by concentrating on the most pertinent data.

\*\*Data Analysis\*\*

The dataset is vast, covering multiple years and regions, necessitating meticulous cleansing, exploration, and visualization. Analysing temporal trends, regional disparities, and correlations between different variables provides a profound understanding of global energy dynamics.

Our data analysis uncovers a current excessive dependence on non-renewable energy sources compared to renewables. As we envision the future, it's imperative that we explore strategies to augment our reliance on renewable energy sources. This transition is not only environmentally prudent but also essential to cater to the power demands of an expanding population. Consequently, we must prioritize the investigation of sustainable energy alternatives in the forthcoming years.

Our analysis discloses a substantial reliance on non-renewable energy sources, particularly coal. However, an encouraging trend is on the horizon, with renewable energy sources collectively approaching 58% of total energy production.

It's evident that coal holds a dominant position in the energy landscape when juxtaposed with other fuels. This heavy reliance on coal is further highlighted when we scrutinize the top 20 companies with the highest megawatt production, which predominantly employ coal for energy generation. This observation underscores the pressing need for a more sustainable energy strategy.

\*\*EDA Concluding Remarks\*\*

Our dataset contains numerous null values, which necessitates appropriate handling. One approach is to replace these nulls with the mode of the data, ensuring a more robust dataset for analysis.

Additionally, we've opted to exclude several columns from our dataset. This decision stems from the understanding that these columns could introduce unnecessary variance or bias. Such bias could potentially lead to over fitting or under fitting of our predictive models, thereby adversely affecting their performance and the validity of our conclusions. This step is crucial in maintaining the integrity of our analysis and ensuring accurate outcomes.

\*\*Pre-processing\*\*

The dataset undergoes pre-processing steps, including handling missing values, scaling features, and encoding categorical variables. This ensures the data is ready for modelling.

\*\*Building Models\*\*

By employing machine learning algorithms like Random Forests, Gradient Boosting, and Support Vector Machines, we predict energy production for features like plant type, capacity. These models provide valuable insights into the future of energy production and assist policymakers in making informed decisions.

\*\*Concluding Remarks\*\*

The transition to renewable energy sources is vital for a sustainable future. The Global Power Plant Database Project, with its comprehensive dataset and sophisticated modelling techniques, plays a crucial role in this transition. By harnessing its insights, we can optimize energy production, reduce carbon emissions, and create a more sustainable world for future generations.