Data Analytics Group Project

Comprehensive Analysis of Energy Project Development

Project By:-

Siddhesh Otari

Introduction

- The dataset includes a variety of wind energy projects, ranging in size from small scale establishments like Walmart canada Corp wind turbine which has a single turbine producing 0.02MW to bigger projects like Kent Breeze Corp wind farms which has 8 turbines producing 80MW.

- The study of the impact of location specific variables on the approval and execution of wind energy products is made possible by this geographical variability.

- The diversity of proponent types makes it possible to investigate the wage in which various Stakeholders approach and maneuver to the regulatory processes pertaining to wind energy projects.

- This diversity offers insights into the various complexities and factors related to each project type.
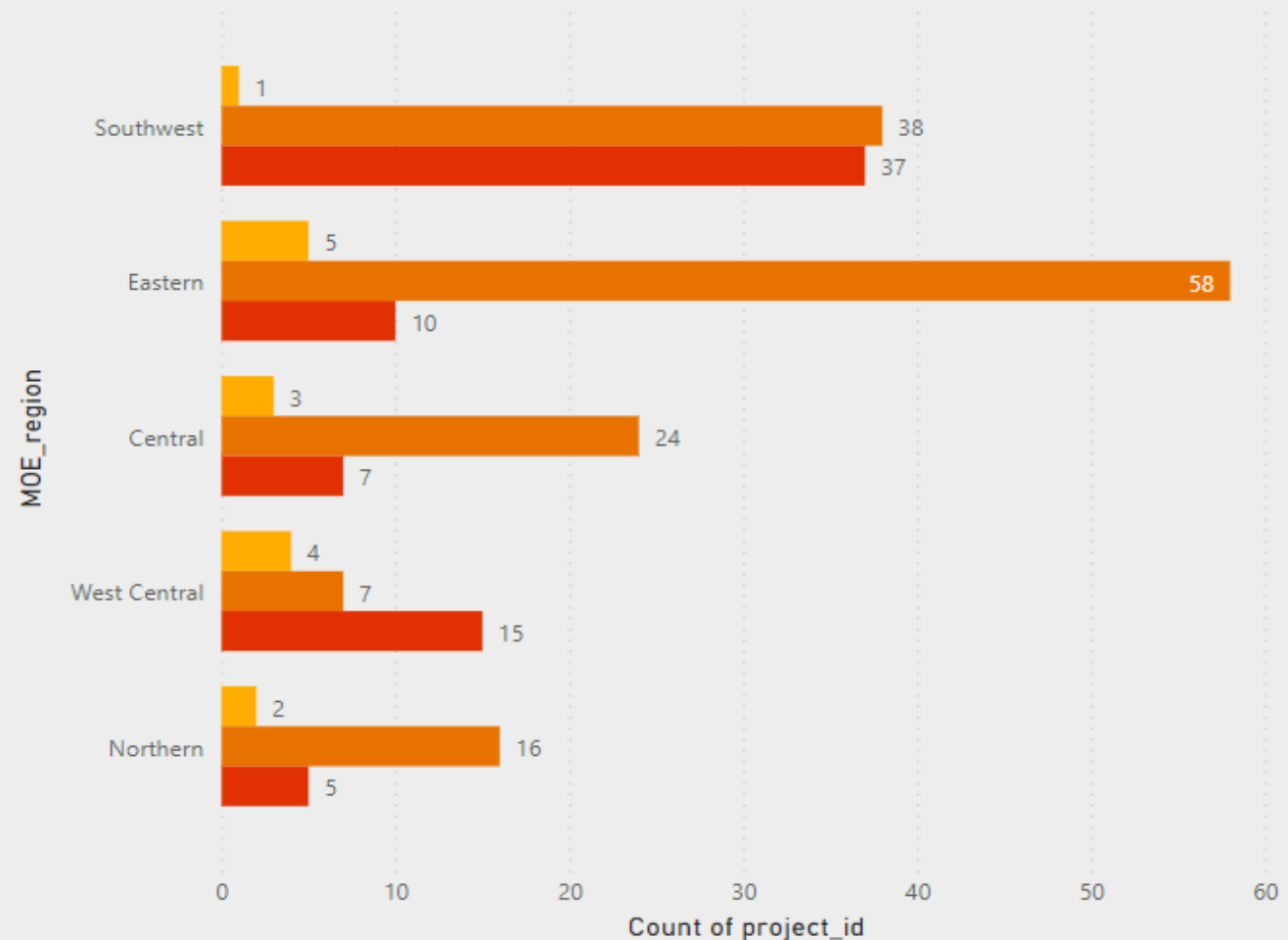
Data Cleaning/RDBMS

- Import the necessary libraries: pandas for data manipulation and numpy for numerical operations.

- Read an Excel file into a Pandas DataFrame.

- Remove leading and trailing whitespaces from column names.

- Filter out all the rows if any null value is present in any of the cells.

- Replace empty strings with NaN (Not a Number) in the DataFrame.

- Fill NaN values with the previous non-null value using forward fill (fill).

- Save the cleaned DataFrame back to an Excel file.

Insight 1

Count of project_id by MOE_region and Project_type

- The dataset indicates a higher degree of project activity in the Southwest region, where there is a concentration of energy projects.
- The two most common types of projects are solar and wind, with solar projects being more common, particularly in the Southwest and Eastern regions, demonstrating a concentration on sustainable energy sources.
- The Northern region is notable for its focus on solar projects, indicating a particular approach or interest in utilizing the region's sun energy resources.
- Additionally, bioenergy projects can be found in areas like Eastern and Central, suggesting research into a variety of non-conventional energy production channels.

Insight 2

Count of project_id by Project_type

- 15 (6.47%)
- 74 (31.9%)
- 143 (61.64%)

Project_type
- ● Bioenergy
- ● Solar
- ● Wind
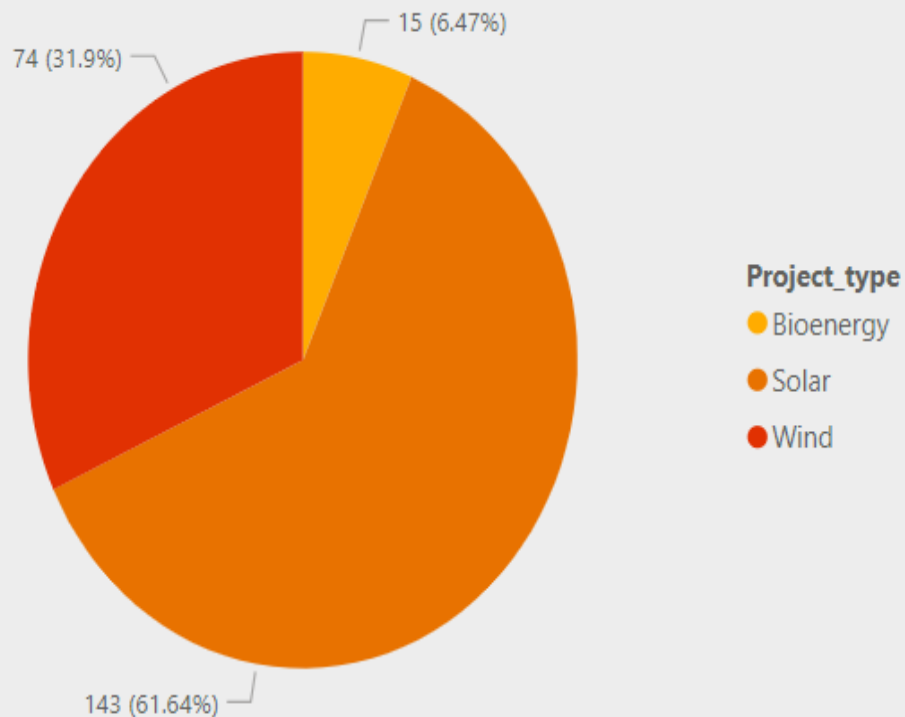
- Solar and wind energy projects make up the majority of the dataset, with solar projects being the most prevalent.
- There is a dedication to using wind energy resources as evidenced by the expansion of wind projects across different locations.
- There are also bioenergy projects, demonstrating a curiosity in various energy generation techniques.
- Geographic preferences are suggested by the differences in the distribution of wind and solar projects by region. Project evolution trends may become apparent through additional analysis over time.
- Personalized plans for areas according to popular project kinds might help promote a varied and sustainable energy portfolio.

Insight 3

Count of project_id by Application_status and Project_type
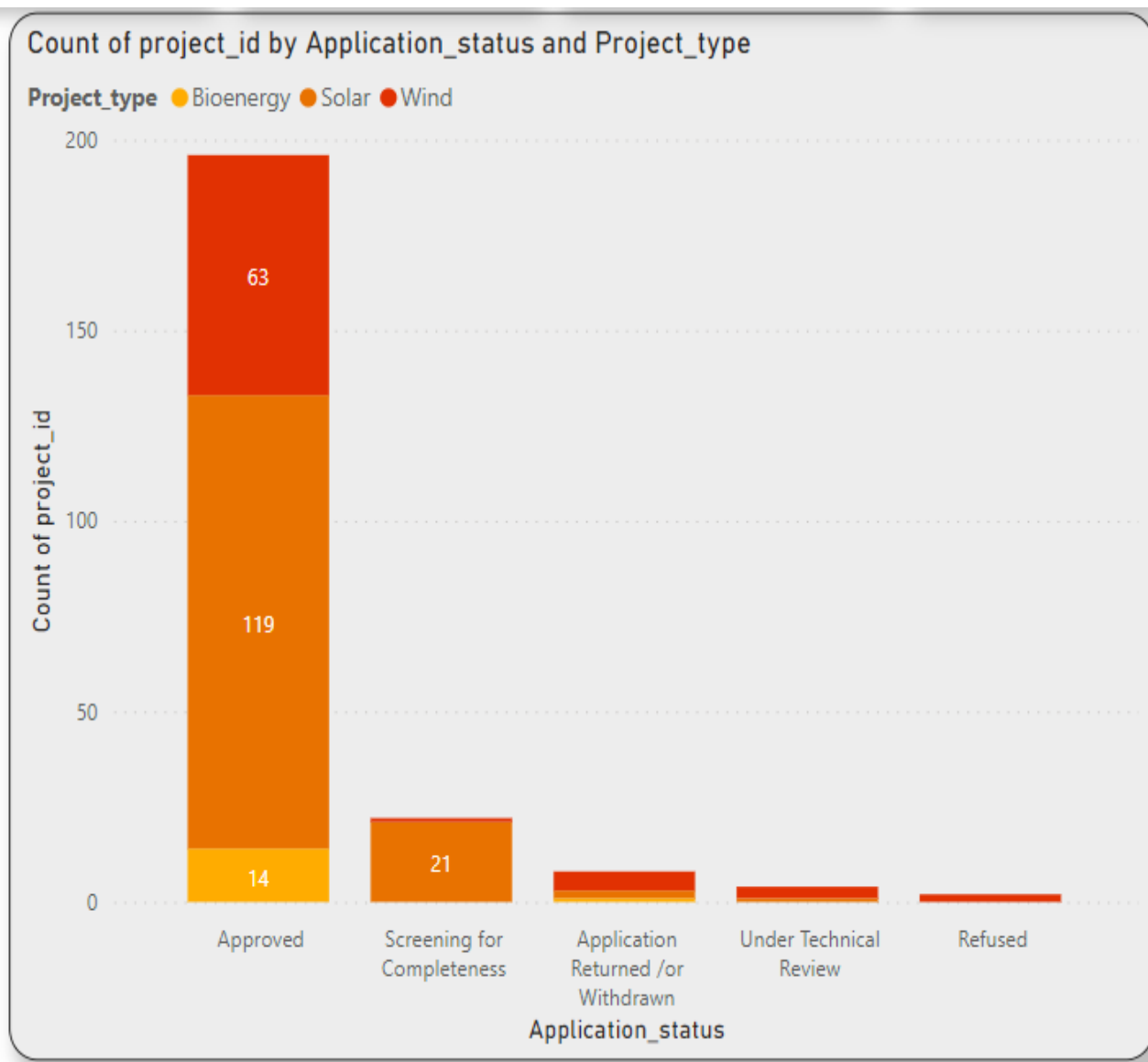
- According to the dataset, most energy projects—mostly of the solar and wind varieties—are approved.
- While some projects are undergoing technical assessment, others may experience rejections or application returns. Lesser bioenergy initiatives are either approved or pending review.
- The dataset provides insights into the various stages of energy projects, highlighting the importance of innovations linked to solar and wind energy and the necessity for more research into the causes of application-related outcomes.

Further Analysis

- <u>Feature Engineering</u>: Enhance model predictiveness by creating new features from existing ones in your machine learning dataset.

- <u>Advanced Visualization Techniques</u>: Elevate data comprehension with advanced visualization methods like interactive dashboards or 3D visualizations.

- <u>Model Building and Evaluation</u>: Progress to the model-building stage, tuning algorithms, and assessing performance using metrics such as accuracy, precision, recall, or F1 score.

- <u>Predictive Modeling</u>: Apply machine learning algorithms to make predictions or classifications based on your cleaned and visualized dataset.

- <u>Ensemble Learning</u>: Improve accuracy by exploring ensemble learning, combining predictions from multiple models.

- <u>Deployment Considerations</u>: Strategize the deployment of your machine learning model, accounting for scalability, integration ease, and ongoing maintenance in real-world settings.

Thank you…