**CO2 Emissions Clustering Analysis**

**Project Brief**

**Executive Summary**

This project analyzes global carbon emissions data to identify natural groupings of countries based on their emissions profiles, energy usage patterns, and economic factors. By applying clustering techniques to comprehensive climate data, we aim to reveal insights about which countries share similar characteristics and challenges, providing a foundation for more targeted and effective climate policy approaches.

**Business Context**

Climate change mitigation requires understanding the diverse circumstances of different countries. A one-size-fits-all approach to emissions reduction is ineffective due to vast differences in economic development, energy infrastructure, and resource availability. This clustering analysis provides a data-driven framework for developing differentiated strategies aligned with countries' unique emissions profiles.

**Project Scope**

This analysis represents the initial foundation of a larger climate analytics initiative. The current scope focuses on:

1. Identifying natural country groupings based on emissions characteristics
2. Creating descriptive profiles for each country cluster
3. Visualizing global distribution of emissions clusters
4. Providing actionable insights for differentiated policy approaches

**Methodology**

**Data Sources**

The analysis uses a comprehensive global carbon emissions dataset containing 79 variables related to:

* CO2 emissions (overall and by source)
* Energy consumption metrics
* Population and GDP figures
* Emissions intensity measurements
* Historical trends and growth rates

**Analysis Approach**

1. **Data Preparation (Day 1)**
   * Clean and normalize dataset
   * Engineer relevant features (emissions mix percentages, growth rates)
   * Create focused dataset with key emissions, energy, and economic indicators
2. **Clustering Analysis (Day 2)**
   * Apply k-means clustering algorithm
   * Determine optimal cluster count using statistical methods
   * Create descriptive profiles for each cluster
   * Identify outliers and special cases
3. **Visualization & Insights (Day 3)**
   * Develop interactive Tableau storyboard
   * Create geographic representations of clusters
   * Design comparative visualizations of cluster characteristics
   * Generate actionable insights for each cluster

**Deliverables**

1. Python notebooks documenting data preparation and clustering methodology
2. Processed dataset with cluster assignments
3. Interactive Tableau storyboard with 4-5 pages of visualizations
4. Written analysis of cluster characteristics and policy implications
5. Recommendations for differentiated approaches by cluster

**Timeline**

* **Day 1:** Data cleaning, preparation, and feature engineering
* **Day 2:** Clustering implementation, algorithm optimization, and cluster profiling
* **Day 3:** Tableau visualization development and insights documentation

**Success Criteria**

1. Clear identification of 4-6 distinct, interpretable country clusters
2. Interactive visualizations effectively communicating cluster characteristics
3. Actionable insights for policy differentiation by cluster
4. Foundation established for future predictive modeling work

**Future Extensions**

This initial clustering analysis sets the foundation for more advanced analytical work:

* Temporal analysis of cluster transitions over time
* Predictive modeling of future emissions by cluster
* Classification models to identify key determinants of cluster membership
* Policy simulation to model intervention impacts

**Resource Requirements**

* Python 3.9+ with pandas, scikit-learn, and related libraries
* Tableau Desktop or Public for visualization
* Access to the complete emissions dataset
* 2-3 days of dedicated analysis time