Project Proposal

Project Title: Analysis of Global Infrastructure

Group ID (Group 4 on Canvas): **George -** 016844490, **Brendan -** 017372459

Members: George Wilfert, Brendan Dishion

Project Summary: Our goal is to look at a multitude of economic results of large-scale infrastructure projects to see how the infrastructure interacts with its immediate environment. While understanding the impacts, we want to gain a better understanding of the criteria that are used to justify and source locations and projects. Using data from governmental and private sources, we will gain deep insights into the complexity of projects and try to quantify the impact in a visually appealing manner. There are short term features such as capital investment, job creation, and regional economic growth, but we also want to explore long term financial sustainability. Metrics such as GDP, employment rate, and other economic indicators are crucial to distilling complex concepts into easily manageable insights. We will be using technical tools such as clustering models in order to find similarities and differences in our data. Additionally, using regression analysis and machine learning methods we want to be able to predict outcomes of infrastructure projects and sourcing of locations.

Objective: Our goal is to develop a structured methodology to assess infrastructure impacts by analyzing the interconnected economic, environmental, social, and logistical factors that shape large-scale projects. Through data-driven analysis, we aim to provide insights into how infrastructure can be designed and implemented more effectively, optimizing for long-term sustainability, efficiency, and maximum public benefit.

Broader Impacts: Infrastructure is the lifeline of our society. Everything we do from driving to drinking water is provided through infrastructure. However, trying to optimize these large-scale projects to provide the most benefits to society tends to be increasingly difficult as complexity increases. Creating a methodology to understand the impacts of infrastructure would allow people to gain better insights into how to best provide for the most amount of people. The audience for our outcome report includes infrastructure project planners along with other business people who deal with the expenditures of infrastructure budgeting/planning. By presenting our findings to project managers, they should be pleased to see our results in a visually appealing manner, presenting the most efficient budget allocation and planning techniques.

Data Sources: We will use multiple datasets from the sources below, which is an accumulation of infrastructure, climate, and other economic datasets.

1st data source: https://maps.worldbank.org/datasets

- https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators (WDI). The World Development Indicators is a massively compiled dataset that presents the most current and accurate global development data available. We will extract data such as 'Public-private partnerships investment in water and sanitation (current US$)', 'Quality of port infrastructure', 'Logistics performance index: Quality of trade and transport-related infrastructure', and other related data that captures infrastructure costs and more.

2nd data source: https://data.gov/

- https://catalog.data.gov/dataset/nhcci (National Highway Construction Cost Index)
- https://catalog.data.gov/dataset/california-power-plants-b4c40 (California Energy Commission)
- https://catalog.data.gov/dataset/transportation-public-financial-statistics-tpfs (Transportation Public Financial Statistics)

3rd data source: https://www.eia.gov/opendata/

- Using APIs to get energy usage by region and type

Expected Major Findings:

- What is the relationship between infrastructure spending and economic growth in different regions?
    - **Hypothesis:** Higher infrastructure investment will correlate with increased economic growth, measured by several economic factors like GDP, employment rate, and productivity improvements.
- How do infrastructure costs vary across different regions, and what factors contribute to these differences?
    - **Hypothesis:** Infrastructure costs will be more expensive in urban areas due to several factors, such as acquiring land permits, acquisition fees, and of course, labor fees. Rural areas will likely have a higher cost of transportation due to the distance traveled for material shipment.

Techniques:

- Regression Analysis (Economic Growth Insights):
    - Analyzing how infrastructure spending affects varying economic aspects, such as employment rate, GDP, and business growth over time.
    - E.g: If highway development investments increased in a particular region/country, how much did the GDP increase compared to developing regions?
- Cost Benefit Analysis of Infrastructure Projects:
    - Comparing project costs vs. long-term economic benefits, essentially analyzing which types of infrastructure projects yield the greatest economic return.

- Cluster Modeling
    - Clustering based on inherent differences in complex projects could give us additional information into a multitude of indicators of future success or prior inputs that cause infrastructure to be developed.
    - E.g: Clustering on metrics such as regulations, region, or local economy may lead to direct insights that can be used for future features in other models.