MIS548

PROJECT MILESTONE REPORT

MIS 548 – DATA WRANGLING

TEAM 1:
ARSLAN AHMAD
CAROLY CORONADO-VARGAS
CHENRUI NIU
JOSEPH D HERST
LEX BRIAN

Dataset

We have decided to use the "World Inequality Dataset," available at https://wid.world/data/. The reason for using this particular dataset is that it is pervasive and has information available regarding the demographical, social, economic, and environmental factors of 331 different geographic entities (countries, regions, states, and grouped regions). The number of available parameters differs from one entity to the other, but on average, 18 parameters are available for each (with a minimum of 1 and a maximum of 184). The available data period is from 1807 to 2017, but it is not the same for all the parameters. On average, 49 years of data is available for each parameter.

Another reason for selecting this dataset is that it is not in a very clean format, which gives us more room to practice the data-wrangling skills we are learning in this course.

We are particularly interested in countries' economic health and total carbon emissions footprints. Key parameters include Gross Domestic Product (GDP), Net Income, Total Exports, and Population. Each country is categorized into geographical regions for comprehensive analysis.

Business Problem Overview

Client: Policy makers, environmentally conscientious businesses, environmental advocates

Problem: Addressing the urgent consequence of climate change has become increasingly vital. It is essential that we all assume responsibility for protecting our environment, starting from the leading cause of the issue - major corporations and governmental entities. To create impactful change and reduce the speed of climate change, we must obtain carbon neutrality in industries through sustainable practices. Collaborating across sectors and government initiatives is crucial in forging a healthier and more sustainable future for our planet.

- *"Carbon dioxide concentrations have increased substantially since the beginning of the industrial era, rising from an annual average of 280 ppm in the late 1700s to 414 ppm in 2021 – a 48 percent increase. Almost all of this increase is due to human activities."* **[1]**

- *"Climate change can also impact human health by worsening air and water quality, increasing the spread of certain diseases, and altering the frequency or intensity of extreme weather events."* **[2]**

- *"Increases in the frequency and intensity of extreme weather events, such as heat waves, droughts, and floods, can increase losses to property, cause costly disruptions to society, and reduce the affordability of insurance."* **[2]**

Purpose of Project: Find a relationship between countries' economic health and carbon emissions footprints using data exploration tools on various economic health parameters. Forecast future trends by adjusting parameters.

Future Plans

Roles and responsibilities:

Role	Name
Group Leader/Project Manager	Brian Lex
Data Engineer	Caroly Coronado-Vargas Arslan Ahmad
Business analyst	Chenrui Niu
Marketer	Joseph Herst

Milestones:

Milestone	Target End	Status	Comments
Dataset Selection	12/15	Complete	See the dataset section above
Data Quality Assessment	12/22	In Progress	
Data Cleaning and Transformation	12/22	Not Started	
Data Integration	12/29	Not Started	
Handling Missing Data	12/29	Not Started	
Exploratory Data Analysis	1/5	Not Started	
Documentation and Presentation	1/11	Not Started	

References

[1] U. E. P. Agency, "Climate Change Indicators: Atmospheric Concentrations of Greenhouse Gases," 18 12 2023. [Online]. Available: https://www.epa.gov/climate-indicators/climate-change-indicators-atmospheric-concentrations-greenhouse-gases.

[2] U. E. P. Agency, "Impacts of Climate Change," 18 12 2023. [Online]. Available: https://www.epa.gov/climatechange-science/impacts-climate-change.