**Design Document: GlobaLens**

------------------------------------------------------------------------------------------------------------------------------------------

Project Title: GDP GlobaLens Design Document

Eric Galanter CISC 4900

Stakeholders:

- Eric Galanter, [ericissacgalanter@gmail.com](mailto:ericissacgalanter@gmail.com)

- Haad Azher,

- Tanmin Ahmed,

- Supervisor: Zack DeSario, Data Science instructor, CUNY Tech Prep, me@zackdesario.com

Elevator Pitch or Abstract:

Our project aims to develop an interactive predictive model that estimates the Gross Domestic Product (GDP) of different countries based on various global factors. These factors include global finance data, supply chain data, world import/export data, happiness index, and ease of doing business. We intend to identify the factors that most significantly impact GDP and visualize these insights through an intuitive dashboard. Policymakers, businesses, and researchers can leverage this tool to understand and predict economic trends.

Designs and Plans for Technical Implementation:

- Data Collection: Obtain data from diverse sources on specified factors. Using Kaggle as a basis, find varying datasets relating to global statistics (happiness index, agriculture, import/export, etc.)Assign each team member to specific data aspects.

- Data Preprocessing and Exploration: Address data cleaning, missing values, and normalization. Standardize and scale the data for implementation. Merge datasets together. Find relationships in the dataset and visualize how they relate to one another and GDP. Conduct feature engineering for the model development cycle.

- Model Development: Use a linear regression model for GDP prediction. The choice of model is subject to change based on data engineering and exploration. Evaluate feature importance and fine-tune the model.

- Model Evaluation: Assess model performance using metrics like MAE and R-squared. The goal is to increase the proficiency of the model. Employ cross-validation for robustness and better accuracy.

- Visualization: Develop an interactive dashboard with visualizations. Highlight GDP factor impacts.

- Documentation: Maintain detailed project documentation to ensure reproducibility.

Role breakdown:

Eric: data visualization

Haad: linear regression

Tanmin: front end

Challenge: Modeling the complex relationships between multiple factors and GDP accurately can be challenging.

Solution: Start with simpler models and gradually increase complexity as needed. Consider ensemble methods, deep learning, or time series analysis depending on the data and project goals.

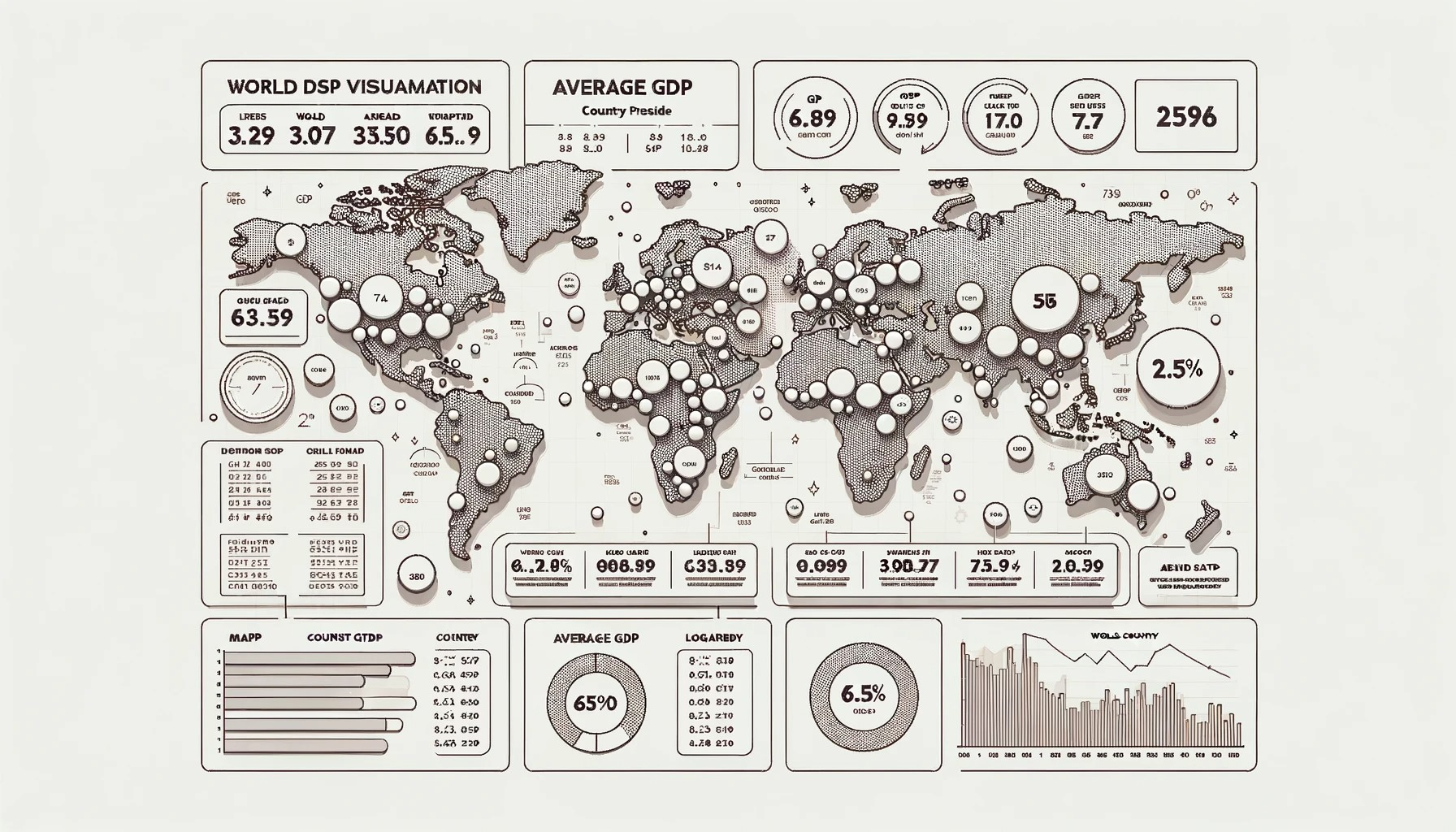
Challenge: The initial solution and implementation might be a little simplistic seeing that a simple linear regression might be enough to find the feature importance in relation to GDP.

Solution: Iteratively improve the model by introducing different visualizations and other model/algo solutions. We might have to put more emphasis on the front end as well to be able to illustrate how well our model operates.

Challenge: Bias in training data can lead to biased predictions and inaccurate insights.

Solution: Carefully analyze training data for biases related to gender, race, geography, or other factors. Mitigate bias through preprocessing and fairness-aware algorithms.

Dashboard Sketch:



Project Flowchart:

A diagram of data processing

Description automatically generated

Tentative Schedule:

- Weeks 1-2: Data Collection and Preprocessing

- Weeks 3-4: Model Development and Training

- Weeks 5-6: Model Evaluation and Refinement

- Week 7: Visualization and Dashboard Development

- Week 8: Documentation and Final Reporting

Descriptions of Task Goals with Time Estimates:

- Task 1: Data Collection, evaluation, and Preprocessing (2 weeks)

- Task 2: Model Development, Training, evaluation, and refinement (3 weeks)

- Task 3: Visualization for data findings (2 weeks)

- Task 4: Dashboard Development (1 week)

Data Sources:

- Global Finance Data: <https://www.kaggle.com/datasets/yusufglcan/country-data/data>

- Supply Chain Data: <https://www.kaggle.com/datasets/dorothyjoel/us-regional-sales>

- World Import/Export Data: <https://www.kaggle.com/datasets/muhammadtalhaawan/world-export-and-import-dataset>

Sample Use Cases for Technical Implementation:

- Policymakers: Understand policy changes that may impact GDP.

- Businesses: Make informed market-entry decisions.

- Researchers: Study GDP correlations and utilize the model for academic research.