Project Proposal: Data Analysis and Visualization Course

# Project Title: RAG-Oriented Simulation of US Presidential Debates

# Objective:

The project aims to develop a Retrieval-Augmented Generation (RAG) system that simulates debates between US presidential nominees. The system will analyze and compare the performance of RAG-generated responses against non-RAG responses using a dataset compiled from previous presidential speeches and the real debate between Trump and Biden.

# Data Collection:

- Data Source: We will scrape and collect data from publicly available transcripts of speeches made by US presidential nominees, as well as the actual Trump-Biden debate.

- Data Processing: The collected data will be preprocessed to build a comprehensive knowledge base that can be utilized by the RAG model.

# Methodology:

1. RAG Model Implementation: Implement a RAG model using the scraped dataset, simulating responses in a debate format.

2. Comparison: Analyze the generated responses and compare them with non-RAG generated responses to evaluate performance and realism.

# Why This Project Is Interesting:

- Innovation in AI Debating: The project explores how RAG models can enhance AI-driven debates, pushing the boundaries of natural language generation.

- Evaluation of Model Effectiveness: It offers a unique opportunity to assess the effectiveness of RAG in generating contextually relevant and convincing arguments.

# Applications Beyond Debates:

The RAG model can be used to create realistic simulations of historical or fictional debates in virtual worlds. This could be applied to video games, educational simulations, or virtual reality experiences where characters engage in dynamic, context-aware conversations