

Group 40 Proposal

We have chosen the Nobel Prize dataset, which contains comprehensive details about Nobel Prize recipients across various categories. This dataset includes 18 columns, with key features of the year of the award, the category, and the recipients' birth and death information, as well as their organizational affiliations. Utilizing this extensive dataset, we plan to uncover insights into the global distribution of intellectual and creative accomplishment. Our analysis will focus on the geographical and temporal distribution of Nobel Prize recipients, identifying trends and examining the relationship between societal, economic, and educational factors in different countries. This analysis is important because it could serve as a benchmark for assessing the effectiveness of the systems that promote intellectual achievement or highlight a country's prioritization in specific fields.

Our group decided to use the data source from Kaggle. We choose the Nobel Prize dataset, which records the geographical and temporal distribution of Nobel Prize recipients from 1901 to 2023. It is a scientific website (<https://www.kaggle.com/datasets/shayalvaghasiya/nobel-prize-data>).

Using our dataset, we will employ statistical methods to analyze trends in Nobel Prizes, including demographic studies that highlight patterns such as the rise of the U.S. Economics Prize winners and network analysis that reveals collaborative networks. We also intended to use the concepts in our course to create Predictive models that will predict future winners by analyzing data trends and sentiment analysis, and literature and peace citations will explore thematic shifts from optimism to contemporary issues. This multifaceted approach aims to illuminate the dynamics of the Nobel Prize selection, combining historical insights with forward-looking predictions.

Using demographic study, we may reveal an increasing diversity among Nobel laureates over time, including more winners from non-Western countries, and a gradual increase in female laureates across all categories. The research could also highlight trends in the age of laureates at the time of award. Additionally, predictive modeling could identify key factors that increase the likelihood of winning a Nobel Prize, such as publication in high-impact journals, association with prestigious institutions, or involvement in globally recognized research projects. The model might be able to predict potential future laureates based on current trends in research excellence, citation networks, and emerging fields of study.

