About:

The importance of the ability to accurately forecast outcomes of key issues in science and technology should be obvious to the majority of individuals. What is not as obvious is how one would accomplish this feat. SciCast was designed for this purpose. SciCast is a research project, which according to its website, is "run by George Mason University and sponsored by the U.S. Government to forecast the outcomes of key issues in science and technology." The concept of SciCast is built around the notion that information collected from many informed individuals is often better at forecasting an outcome than that same information collected from a handful of experts in the subject. This concept is often referred to as crowdsourcing. Users who wish to participate must first register with scicast.org, after which they can begin making forecasts. Once an individual has registered for SciCast, he or she is then free to search through questions on the site and make forecasts on those questions as desired.

Objectives:

The objective for this project was to publish the SciCast data and documents to an open data repository, harmonizing with other prediction market data archives. Plan for data to be available for 10+ years. Design, pre-register and carry out an analysis of this data to replicate some previous results and test a new idea. Work with the Replication Markets team on their Round 1-4 data.

The user will be able to access, download, and analyze the data freely using a user-friendly interface. In addition, the platform will provide visualization tools to user for better understanding the data published.

Data Quality Assessment:

- Uniqueness: The data set provided by SciCast was unique because it was the result of forecast surveys that they conducted. Their data was unique as it was able to explore the relationships between questions and the users.
- Accuracy: The data set was accurate in the sense that it correctly represented the users' forecasts. However, the users did not forecast with total accuracy and their predictions were not always found to be true.
- Atomicity: As there were no transactions and all records were kept consistently, atomicity was not a concern in the data set.
- Conformity: The data was formatted consistently and following a pattern of size, type, and format.