Problem Statement: How can historical tornado track and impact data be leveraged to model the potential effects of a hypothetical tornado in terms of both human toll and property damage?

Context: Tornadoes are one of nature's most powerful and potentially destructive forces and are a part of life for much of the United States. Even with vast improvements in detection time and warning methods, they represent a significant threat to both life and property. Can historical data on tornadoes tell us anything about our levels of future risk? To understand the threat tornadoes might pose in the future, we first need to understand how these storm events and their impacts have changed over time.

The major questions we want to try to investigate to create this understanding are as follows:

1. How have improvements in early detection and warning propagation for tornadoes affected the resulting human toll (i.e. the number of injuries/fatalities)?
2. How has property damage trended over time?
3. Has the region in which tornadoes are the most common shifted or expanded over time?
4. Is the number of tornadoes per year /strength of tornadoes increasing over time?
5. Is there a "tornado season", and has its length changed over time?
6. What is the impact of climate on tornadoes? Do uncharacteristically hot or wet years produce more or stronger tornadoes?
7. Is there a time of day in which tornadoes are most common? Is there a significant difference in human toll for tornadoes that touch down in the middle of the night?
8. Is there a strong correlation between the strength of a tornado and its duration/distance traveled on the ground? Has either measure changed at all over time?

It is our assertion that leveraging data to find answers to the above questions will provide a stable base for predicting the impact of a tornado with a defined set of parameters.

Criteria for success: This project's success will hinge on the creation of a predictive model that can take relevant variables as inputs and output the predicted level of injuries, deaths, and property damage, as well as the relative likelihood of that tornado's occurrence. This will be provided via a github repository containing notebooks covering every step of the analytical process, a written final report, and an executive slide deck.

Scope of solution space: Our analysis will be limited to recorded tornado touchdowns in the US & Territories between 1950 and 2019.

Constraints:

- Data fidelity: The base dataset from NOAA is not structured for expedient tornado-by-tornado analysis. Tornadoes do not have unique identifiers, and more impactful tornadoes exist across multiple rows in the dataset. There are also null values to account for.
- Need for additional datasets: For our predictive model to be most effective, it is crucial that we surround the impact of each tornado with as much context as possible. This means adding available geographically linked data to expand the list of potential explanatory variables.
- Data Granularity: The supplemental datasets will necessarily be linked to the base dataset by the smallest geographic unit available, which in this case is the county. Any variation in a target variable *within* a county will be lost.

Potential Stakeholders:

- People or businesses considering relocation within a tornado-prone area
- Climatologists and other weather researchers
- First-responders

Data Sources:

- Tornado Dataset NOAA's Storm Events Database. Contains descriptive statistics for all recorded tornado events between 1950 and 2019 in the US & Territories.
- Population Density by County from Census.gov
- Land Area by County from Census.gov
- Median and Total Home Age by County from Census.gov
- Historical US climate data by county from NOAA CDO