# Data Quality Form

# 👉 Business Question: Where have the most tornadoes occurred in Texas over the past 5 years? What $ impact has it had? Fatalities?

https://www.kaggle.com/datasets/michaelbryantds/tornadoes?select=1950-2021\_all\_tornadoes.csv

For each item below, give a score ranging from 1 to 5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** |
| Poor | Below Average | Average | Good | Excellent |

# 1. Relevance

* + **DESCRIPTION**: How relevant is this data to the business question you are trying to answer? Are all of the fields useful as a part of the analysis?
  + **SCORE**: 2

# 2. Timeliness

* + **DESCRIPTION**: How up to date is the data? Is it recent enough and covering a long enough range of time to be useful?
  + **SCORE**: 5

# 3. Historical Usage

* + **DESCRIPTION**: Has this data been utilized in other analyses? If there is not much usage of the data source, is there a reason for this? Could other users be avoiding this data set for a particular reason?
  + **SCORE**: 1

# 4. Cleanliness

* + **DESCRIPTION**: How clean is the data set? Are there a large amount of null values, duplicated rows, or inconsistent formatting of the data? Will it take a lot of work to get it in usable form?
  + **SCORE**: 3

# 5. Credibility

* + **DESCRIPTION**: What source is this data set coming from? Is this considered a trustworthy source? Are there any potential biases that this data source may hold that could skew their presentation of the data?
  + **SCORE**: 4  
      
       
      
    TOTAL SCORE: 15/25