

2023 Data Derby Data Description and Rules



Welcome to Data Derby 2023! This year's Data Derby is in-person! The event date is April 8th. The location is the Hennepin Tech Community College at Eden Prairie. For more information, please check with <https://mnstateitcoe.org/data-derby-analytics/>

Overview of the data

According to the United States Department of Agriculture (USDA), in 2023, all food prices are predicted to increase between 3.5 and 4.5 percent, food-at-home prices are predicted to increase between 3.0 and 4.0 percent, and food-away-from-home prices are predicted to increase between 4.0 and 5.0 percent (source: <https://www.ers.usda.gov>). This year's Data Derby will be investigating the food cost increase and predicting future food prices.

Data Derby provides you 15 datasets on food costs from 1980's to 2022. Those datasets offered the food prices averaged in U.S. cities by month at each year. The prices listed are the price per pound in U.S. dollars. The data covers prices of fruits (bananas, oranges, strawberries, and tomatoes), animal products (beef chunks, beef steaks, pork, chicken and eggs), grains (bread, flour, and rice) and dairy products (American cheese, cheddar cheese, and milk).

Data Derby also offered datasets of the consumer price index (CPI) change, CPI historical forecast, producer price index (PPI) change, PPI historical forecast, and inflation rate from 1936 to 2022. You may not need all of those CPI and PPI related data, but those may help with your analysis.

General information and rules of the competition

Participating teams will be working on the data provided from March 1st to April 7th. All the data and questions will be sent to the teams and faculty advisor by email. Additional information may be sent through email between March 1st and April 8th. Each team is encouraged to look for additional related data to answer the questions of their level.

There are two divisions in this year's competition: novice level and advanced level. The novice level is for undergraduate students with completing at most two data science/statistics related courses. The advanced level is for undergraduate with more than two statistics or data science related courses completed or graduate students. Faculty advisor will determine your team level.

Each team will be making a presentation with answering all the questions of their level, explaining any patterns, and providing insights of the data. The judges will be judging the presentations based on rubrics (will be shared at the end of March). Awards will be given to the first, second and third places of each level.

Each team in the Novice level has 10 minutes to present their findings with an additional 5 minutes for Q&A. Each team in the Advanced level has 15 minutes to present their findings with an additional 5 minutes for Q&A. **During the presentation, please do NOT release your institution affiliation or your faculty advisor name.** Team(s) will lose 25% of the points if any release of those information during the presentation. During the presentation, please clearly state your findings, insights, and your data processing and cleaning.

Each team is required to send the draft version of their presentation to boern.vang@metrostate.edu by March 31st. The final version can be somewhat different from the draft version. **Each team should include the information of their level of participation (Novice or Advanced level) in the draft.** Any team who answer the advanced level questions will be automatically placed in the advanced level.

Any questions related to the competition, please contact boern.vang@metrostate.edu