

MATH113/CAIS105: Intro to Data Science

Fall 2023

Mini-Project 02

Mini-Project is due at the end of class on the day indicated on the course schedule.

Learning Objectives:

- Analyze a topic of interest with one data table

Part 1

Work in groups of 2-3. Pick from one of the following topics for your analysis:

- Avocados: <https://www.kaggle.com/datasets/neuromusic/avocado-prices>
- Olympics: <https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results>
- Cereal: <https://www.kaggle.com/datasets/crawford/80-cereals>
- Squirrels: <https://www.thesquirrelcensus.com/>

Your analysis should consist of at least 3 questions that you can answer with your dataset (like we've done with examples in class). At a minimum, analysis must include:

- Adding a new column to your dataset
- Filtering your dataset to a subset, and answering a question about that subset
- Aggregating and summarizing your data by groups

You should complete your analysis in an R Markdown file. In the file, label the code chunk where you completed each of the required tasks above.

You will give a 5 minute presentation of your findings. Your presentation must include:

- The questions you asked and their answers
- At least one visualization

Submission

Save your R Markdown as a PDF and submit as a group through PLATO.