

Walmart Sales Project

Description: In this project, you will clean a dataset containing Walmart sales data and analyze sales, unemployment rate, CPI, fuel price, and other fields.

Difficulty: Medium

Data download: <https://www.kaggle.com/datasets/mikhail1681/walmart-sales>

Data Cleaning Practice:

Using Python, R, Excel, or another data cleaning tool of your choice, clean the data to meet the following criteria:

- ☐ Data is sorted first by store number (ascending) and second by date (ascending)
- ☐ Date is in the format MM-DD-YYYY
- ☐ Weekly Sales is rounded to the nearest 2 decimal places
- ☐ Temperature is rounded to the nearest whole number
- ☐ Fuel Price is rounded to the nearest 2 decimal places
- ☐ CPI is rounded to the nearest 3 decimal places
- ☐ Unemployment is rounded to the nearest 3 decimal places
- ☐ Ensure that there is no missing data

Business Questions:

- Which holidays affect weekly sales the most?
- Which stores in the dataset have the lowest and highest unemployment rate? What factors do you think are impacting the unemployment rate?
- Is there any correlation between CPI and Weekly Sales? How does the correlation differ when the Holiday Flag is 0 versus when the Holiday Flag is 1?
- Why do you think Fuel Price is included in this dataset? What conclusions can be made about Fuel Price compared to any of the other fields?