Walmart Sales Data Analysis Questions in R

Martin Van Waerebeke

July 19, 2024

Statistics & Visualization

1. **Data Inspection:**
   - How many rows and columns does the dataset 'walmart_data' have? Use `dim()` or `nrow()` and `ncol()`.
   - What are the data types of the columns in 'walmart_data'? Use `str()` or `class()`.
   - Put the Date column in the Date format rather than char, use `as.Date()`
   - Are there any missing values in 'walmart_data'? If yes, in which columns? Use `is.na()` and `colSums()`.
2. **Statistical Analysis:**
   - What is the average, median, and range of 'Weekly_Sales' in 'walmart_data'? Use `mean()`, `median()`, and `range()`.
   - Which store in 'walmart_data' has the highest average weekly sales? Use `aggregate()` or `tapply()`. Then, use `dataframe[order(dataframe$column_name), ]`.
3. **Data Visualization:**
   - Create a bar chart showing average weekly sales for one store in 'walmart_data' using `barplot()`.
   - Create a line graph depicting weekly sales over time for a specific store in 'walmart_data' using `plot()`.

Advanced Analysis

3. **Exploratory Analysis:**
   - Which store performed the largest weekly sales / CMI value and at what date ?
   - Identify any noticeable trends or patterns in weekly sales across different stores in 'walmart_data'. Use `plot()`.
   - Find correlations between 'Weekly_Sales' and other variables like 'Temperature', 'Fuel_Price', 'CPI', and 'Unemployment' in 'walmart_data'. Use `cor()` or `pairs()`.

Open question

7. **Advanced Data Manipulation:**
   - How would you handle missing data in 'walmart_data'?