Babson CODE
February, 2023

Data Analysis Hackathon

Scenario:

Jason's runs a homestyle bakery with five locations in the suburbs of Chicago. It specializes in muffins and the only products it sells are three types of muffins. Each store has multiple managers, only one of which works on a given day. Managers are paid $36 an hour. Each store has various general staff, all of whom make $18 per hour. All employees work eight hours per day. If an employee works more than forty hours a week, they are paid 1.5x more. The stores each have different rents. Jason has had his friend take note of certain important information about the store operations, but this data is formatted in a non-standard way, so analysis of it is not currently possible. He wants it to be reformatted and analyzed. That is your task.

Data format:
All data is on one .txt file
Row format – Date, Worker, Worker, Worker, Worker, Manager, Total Daily Sales, Blueberry Muffin Sales, Chocolate Muffin Sales, Banana Muffin Sales
Example row – January 05, 2015, Evelyn, Robert, Enris, Cole, $179.0, 29, 22, 20
Number of staff will vary per store and per day, manager name will always be last
Stores will be separated by an indentation and "LOCATION: Hinsdale" (view file for clarification)
Rent cost is mentioned below store location

Deliverables, Performance Criteria, and Rules are on the back side of this page.

Deliverables:

1. Find if the sales of muffins vary independently in each store, present correlation value
2. Are sales higher when particular employees are there? What item?
3. Have some employees been stealing? Who? How much?
4. Create graphs to visualize the following
    a. Sales over time by store (single graph)
    b. Item sales over time by store (graph per store)
    c. Bar chart with revenue, cost for each store (graph per month)
5. Give the monthly income statement for each store with COGS for each product, cost of management, cost of general staff, revenue from each product, total revenue, and profitability. (a separate .txt file for each statement named store_month_year.txt is preferred)

Performance criteria:

1. If you cannot complete deliverables which require repetitive action (i.e. creating multiple CSV files) you should focus on one (i.e. create one CSV file) – go narrow and deep rather than shallow and broad
2. No comments in code or well named variables are necessary, accuracy of output is the main concern (it is best practice though if you have time)
3. There will be three random challenges yelled out during the contest, first correct answer gets $5

Rules:

1. You can use the internet whenever and however you see fit.
2. You can collaborate with your peers only if you are not trying to compete to win. Feel free to talk casually though, just not about the topic.
3. Ask clarifying questions to the organizers at any time