

Hello (stakeholder)!

Hope your day is going well! I just wrapped up my investigation into the JSON files you sent over and wanted to share my preliminary thoughts and findings:

Initial Findings

I went ahead and used one of the questions you sent over with the files to help me investigate quality issues and workflows for the current data structure. I took a look at the top 5 brands by number of receipts for January and February, but I ended up with a 13-way tie for top brand for both months. I'll come back to this problem when I explain the current state of the data.

January	February
Betty Crocker	Betty Crocker
Bisquick	Bisquick
Bugles	Bugles
Cascadian Farm	Cascadian Farm
Cheerios	Cheerios
Chex	Chex
Cinnamon Toast Crunch	Cinnamon Toast Crunch
Fiber One	Fiber One
Garden of Eatin'	Garden of Eatin'
Gold Medal	Gold Medal
Good Earth	Good Earth
Green Giant	Green Giant
Haagen-Dazs	Haagen-Dazs

Considering the same 13 companies tied for top brand both months, there's something going on with the data itself. **I would not recommend communicating these findings to others since they're likely incorrect at this stage.**

Initial Questions

In order to provide effective recommendations, I have some initial questions:

- How is the data being collected? Using APIs? User input? Something else?
 - There are a lot of missing values in important locations. Understanding where this data is being collected and how will go along way in making future analysis more efficient and straightforward. I'm most interested in the methods for collecting receipt data.
- Are you open to discussing building a data warehouse in order to streamline investigations?
 - The type of files originally sent over are not easy to work with for data analysis. Warehousing the data will free up time for analysts, save space, and we can ensure consistent data manipulation for use.

I expect these questions will lead to further discussions about the data you're collecting. Your answers will help me determine what concerns we may face for future performance and scaling considerations for the solutions we may select from going forward. If you need further clarification on anything I've mentioned here, we can schedule a quick call to talk through everything!

Thanks,

Deanna