ASK:

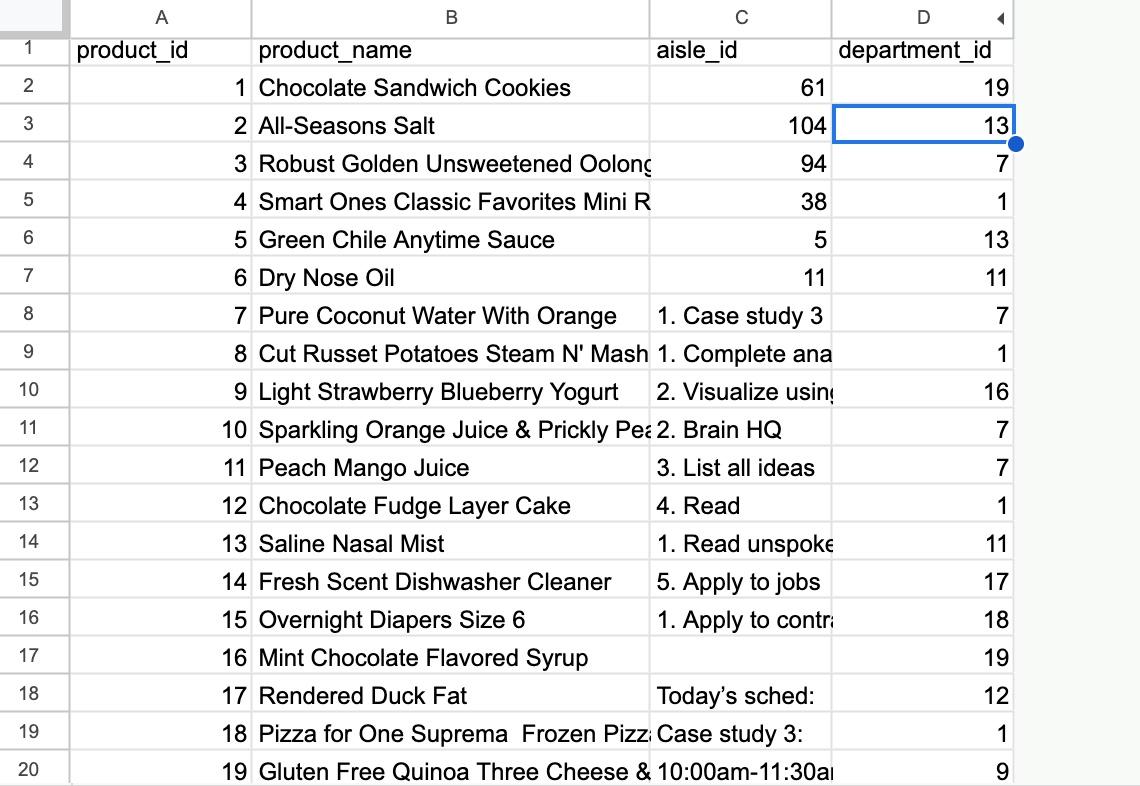
Problem:

An employer from a company who focuses on low sugar products has developed a new sweetener and wants to determine which product category possesses the most consumer demand for natural sweeteners so they know which business would be their ideal customers. They want you to find a dataset containing information on consumer purchases and use that dataset to determine which food product category has the most demand to produce items that are low in sugar.

1. What sweet goods are currently the most popular among consumers
2. What sweetened goods drive the most grocery sales
3. What platform is a good representative of consumer sales

Prepare:

1. The dataset used is one that contains data points pertaining to consumer purchases of online grocery store items from instacart. It is divided into six csv files organized in the form of rows and columns. These files include:
   1. Product.csv
   2. Aisle.csv
   3. Department.csv
   4. Orders.csv
2. Before being processed, the product dataset was uploaded onto excel and it’s contents were filtered to only include items that fall in sweetened goods categories. This was done by first filtering the aisle column and formatting the other columns to organize themselves with respect to the filter.





Process:

SELECT DISTINCT product\_name, product\_id FROM temp-424017.GroceryStoreConsumerPurchases\_Data.Product\_Data

#determines if there are any duplicates for product\_name and product\_id which could lead to misrepresentation of values in the data

GROUP BY

product\_id,

product\_name;

#determines whether there are any null values that would need to be filtered to make sure the data is clean

SELECT product\_id is NULL OR product\_name is NULL OR aisle\_id is NULL OR department\_id is NULL

FROM temp-424017.GroceryStoreConsumerPurchases\_Data.Product\_Data;

Analysis:

#Creating a new data table that contains columns from both the aisle dataset and the product dataset so that the aisle id's, aisle names, product id and product names are included in one table

WITH Product\_Aisle\_Table AS(

SELECT Products.product\_name, Aisle.aisle, Products.product\_id, Aisle.aisle\_id

FROM temp-424017.GroceryStoreConsumerPurchases\_Data.Product\_Data AS Products

INNER JOIN temp-424017.GroceryStoreConsumerPurchases\_Data.Aisles\_Data #Using inner join to join the aisle names and aisle id's

AS Aisle

ON Products.aisle\_id= Aisle.aisle\_id) #matching up the primary key aisle id in the aisles table with the foreign key in the products table to join both tables

SELECT Product\_Aisle.aisle, AVG(Orders.add\_to\_cart\_order) #selecting the two columns needed for comparison from each table

FROM Product\_Aisle\_Table AS Product\_Aisle

INNER JOIN

temp-424017.GroceryStoreConsumerPurchases\_Data.order\_products\_train\_data AS Orders #joining the data from the orders table with the newly created Product\_Aisle table

ON

Product\_Aisle.product\_id=Orders.product\_id #joining the two tables by matching up the primary key product id in the product aisle table with the foreign key product id in the orders table to join both tables

GROUP BY

Product\_Aisle.aisle

ORDER BY

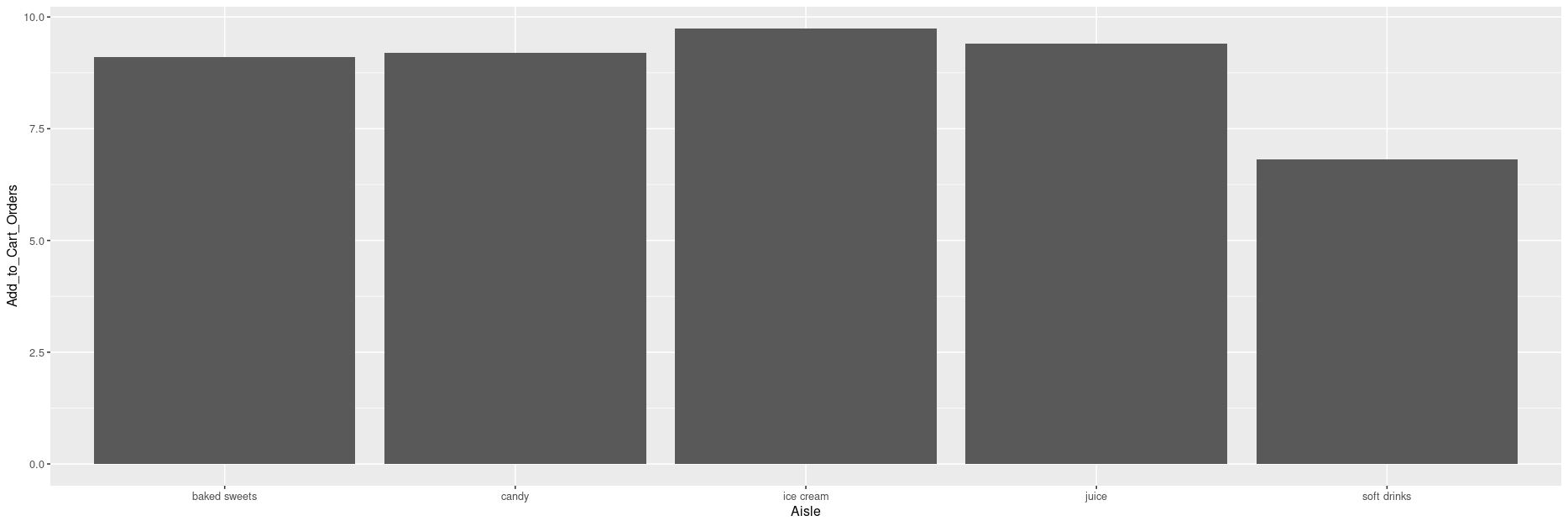
AVG(Orders.add\_to\_cart\_order) DESC;

Initial results:

| Aisle | Average\_Add\_to\_Cart\_Orders |
| --- | --- |
| ice cream toppings | 10.82738095 |
| frozen juice | 10.54081633 |
| honeys syrups nectars | 10.13582402 |
| breakfast bars pastries | 9.700699746 |
| baking supplies decor | 9.549360146 |
| cookies cakes | 9.505711423 |
| air fresheners candles | 9.475164011 |
| baking ingredients | 9.365678484 |
| frozen dessert | 9.274403471 |
| refrigerated pudding desserts | 9.220358589 |
| ice cream ice | 9.128770506 |
| candy chocolate | 8.9320702 |
| cocoa drink mixes | 8.808851224 |
| breakfast bakery | 8.444523399 |
| juice nectars | 8.11358885 |
| bakery desserts | 7.519653564 |
| soft drinks | 6.808710609 |

The results from this table were then consolidated to include fewer categories as aisle names that clearly housed the same category of sweetened product were combined together and the average of both values were recorded (ex: ice cream toppings + ice cream, air freshener candies + candy chocolate). The final results are recorded in the table below:

| Aisle | Average\_Add\_to\_Cart\_Orders |
| --- | --- |
| ice cream | 9.74351831 |
| juice | 9.399770106 |
| baked sweets | 9.106918187 |
| candy | 9.203617106 |
| soft drinks | 6.808710609 |



Conclusion:

The results of the analysis as indicated by the graph show that the most popular sweetened goods that consumers purchase pertain to the ice cream and frozen desserts category followed by juice, candy, baked sweets and soft drinks.

My recommendations:

1.The manufacturer as a result should focus on developing their business by first selling their product to popular and innovative ice cream brands and observe if they are able to generate a sizable revenue and profit from such a transaction before moving their efforts onto the second most popular, the third and so on and so forth.

2. The manufacturer should also expect the summer, fall and winter seasons to be their most profitable as the weather and seasonal holidays drive ice cream, juice, and candy (The top three products that are in most demand by consumers). As a result they should direct most of their marketing and inventory storage efforts to consumers and businesses during these times.

3. The business owner should also take into account that two income streams can be generated from both selling to businesses and to consumers. If products are targeted towards consumers than