Anupama Bhatta    
Professor Michael Hess

SI 564    
November 29, 2023

Final Project: Fruits and Vegetable Prices in USA

A diagram of fruits and vegetables

Description automatically generatedFrom: Lawrence Summerset

To: DBA team

Dear DBA team,

I trust this message finds you all well! Our research group has recently acquired a massive dataset containing information on fruits and vegetable prices in the USA. Unfortunately, I find myself overwhelmed with the intricacies of this dataset, and your support in deciphering it would be immensely beneficial.

We are interested in a handful of data points from the *fruits\_n\_veggies* data.

1. How many items in the dataset have a *priceUnit* of “pound”?  
   -> There are a total of 94 items in the dataset that have a *priceUnit* of “pound”.

**Queries:**  
i. USE fruits\_n\_veggies;  
ii. SHOW tables;  
iii. SELECT COUNT(1) AS pound\_item\_count FROM prices WHERE priceUnit LIKE '%pound%';

**Screenshots:**A black screen with white text

Description automatically generated

1. Can you list all distinct forms and the count of items associated with each form?  
   -> All the distinct forms and the count of items associated with each form are listed in the screenshot below.  
   **Queries:**  
   i. SELECT f.formName, COUNT(DISTINCT i.itemID) AS item\_count FROM forms f LEFT JOIN prices p ON f.formID = p.formID LEFT JOIN items i ON p.itemID = i.itemID GROUP BY f.formName;  
   **Screenshots:**  
   A screenshot of a computer

   Description automatically generated
2. What are the top 5 most expensive items, including their prices and the forms they belong to?  
   -> The top five most expensive items, including their prices and the forms they belong to are listed in the screenshot below.  
   **Queries:**  
   i. SELECT i.itemName, ROUND(p.price, 2) as price, f.formName FROM prices p JOIN items i ON p.itemID= i.itemID JOIN forms f ON p.formID = f.formID ORDER BY p.price DESC LIMIT 5;  
   **Screenshots:**A screenshot of a computer

   Description automatically generated
3. Calculate the average price for each *priceUnit* across all items.  
   -> The average price for each *priceUnit* across all items is listed in the screenshot below.  
   **Queries:**  
   i. SELECT priceUnit, ROUND(AVG(price), 2) AS avg\_price FROM prices GROUP BY priceUnit;  
   **Screenshots:**A black screen with white arrows

   Description automatically generated
4. What is the average price of items for each form, displayed in descending order of average price?  
   -> The average price of items for each form, displayed in descending order of average price is listed in the screenshot below.  
   **Queries:**i. SELECT f.formName, ROUND(AVG(p.price), 2) AS avg\_price FROM prices p JOIN forms f ON p.formID = f.formID GROUP BY f.formName ORDER BY avg\_price DESC;  
   **Screenshots:**  
   A screenshot of a computer

   Description automatically generated
5. What is the total amount spent on items for each form?  
   -> The total amount spent on items for each form is listed in the screenshot below.  
   **Queries:**i. SELECT f.formName, SUM(p.price) AS total\_price FROM forms f LEFT JOIN prices p ON f.formID = p.formID GROUP BY f.formName; **Screenshots:**  
   **A screenshot of a computer

   Description automatically generated**
6. What are the forms with the highest price variation among their items?  
   -> The forms with the highest price variation among their items are listed as below,  
   **Queries:**i. SELECT f.formName, ROUND(MAX(p.price) - MIN(p.price), 2) AS price\_variation FROM prices p JOIN forms f ON p.formID = f.formID GROUP BY f.formName ORDER BY price\_variation DESC;  
   **Screenshots:**A screenshot of a computer

   Description automatically generated

Lastly, can you please create an outline of why you made the database design choices you made? Examples of database design choices could be explaining what the primary table in the database is and why it's an important table (maybe it's the table that connects all other tables together in your database). You could also discuss other table relationships, or if you chose not to include specific fields/tables that were in your original data source explaining why you made those choices.  
-> The *fruits\_n\_veggies* database showcases a deliberate design with three core tables: *items* for fruits and vegetables, *forms* for different varieties or forms of said fruits and vegetables, and *prices* for linking *items* and *forms* with pricing information. The *prices* table is my central table as it establishes crucial relationships between *items* and *forms* and connects various facets of the database. It employs auto-incrementing primary keys for unique identifiers and maintains referential integrity through foreign keys, which have been established for both *itemID* and *formID* across tables. In the *prices* table, the *itemID* field is a foreign key that references the *itemID* field in the *items* table. This relationship ensures that every entry in the *prices* table, representing a specific pricing instance, is associated with a valid item from the *items* table. Similarly, the *formID* field in the *prices* table is a foreign key that references the *formID* field in the *forms* table, establishing a connection between pricing information and various forms or varieties in the database. These foreign key constraints enforce referential integrity, preventing the creation of "orphaned" records and maintaining consistency in the relationships between items, forms, and prices.

To summarize, the structure reflects a normalized design, minimizing redundancy and optimizing data integrity. The choice of separate tables for distinct entities (*items*, *forms*, and *prices*) supports flexibility, allowing seamless additions without necessitating substantial schema alterations.

**Database Export**The database export file has been [linked here as fruits\_n\_veggies.sql](fruits_n_veggies.sql) and also submitted separately to Canvas. **Queries:**i. mysqldump --set-gtid-purged=OFF -h 34.71.12.223 --port 10940 -u axbhatta-rw -p fruits\_n\_veggies > fruits\_n\_veggies.sql

Thank you for your help on this matter! I’m looking forward to seeing your results.

Dr Lawrence Summerset

Head of Research