Coffee Shop Analysis

Calvin Gerard

Table of Contents

Business Questions

01

The details of the case

Process

02

The steps I took to answer the business questions

Findings

03

Information gathered from the case

Improvements

04

What improvements could be made for future analysis

01
Business Questions

The details of the case

Business Questions

This case study is based off of <u>this</u> dataset from Kaggle, which I used to answer the following questions:

- Total Orders: What is the total number of customer orders?
- Total Sales: What is the total revenue generated?
- Total Items: What items are being sold and how many?
- Average Order Value: How much revenue is made per order?
- Sales by Category: How much revenue is made per each category of item sold?
- Top Selling Items: What are the most popular items being sold?
- Orders by Hour: How many orders are made throughout the day?
- Sales by Hour: How much revenue is made throughout the day?
- Orders In or Out: What percentage of orders are dine-in or takeout?

Business Questions (cont.)

- Total Quantity by Ingredient: How much of each ingredient is in storage?
- Total Cost of Ingredients: How much does each ingredient cost to stock?
- Calculate Cost of Coffee: How much does each coffee item cost to make?
- Percentage Stock Remaining by Ingredients: What percentage of each ingredient remains in stock?
- List of Ingredients to Re-order: What ingredients does the shop need to restock?
- Total Staff Cost: How much does the shop spend on staff salary?
- Total Hours Worked: How many hours do the shop's employees work?
- Hours Worked by Staff Member: How many hours do each of the employees work?
- Cost per Staff Member: What is the hourly salary of each staff member?

02

Process

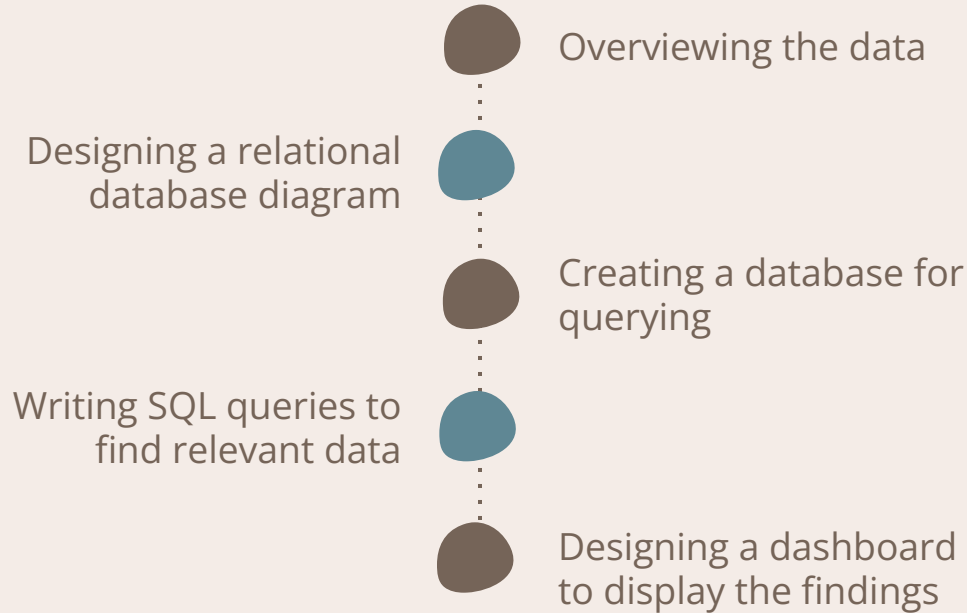The steps I took to answer the business questions

Process

Overviewing the data

Designing a relational database diagram

Creating a database for querying

Writing SQL queries to find relevant data

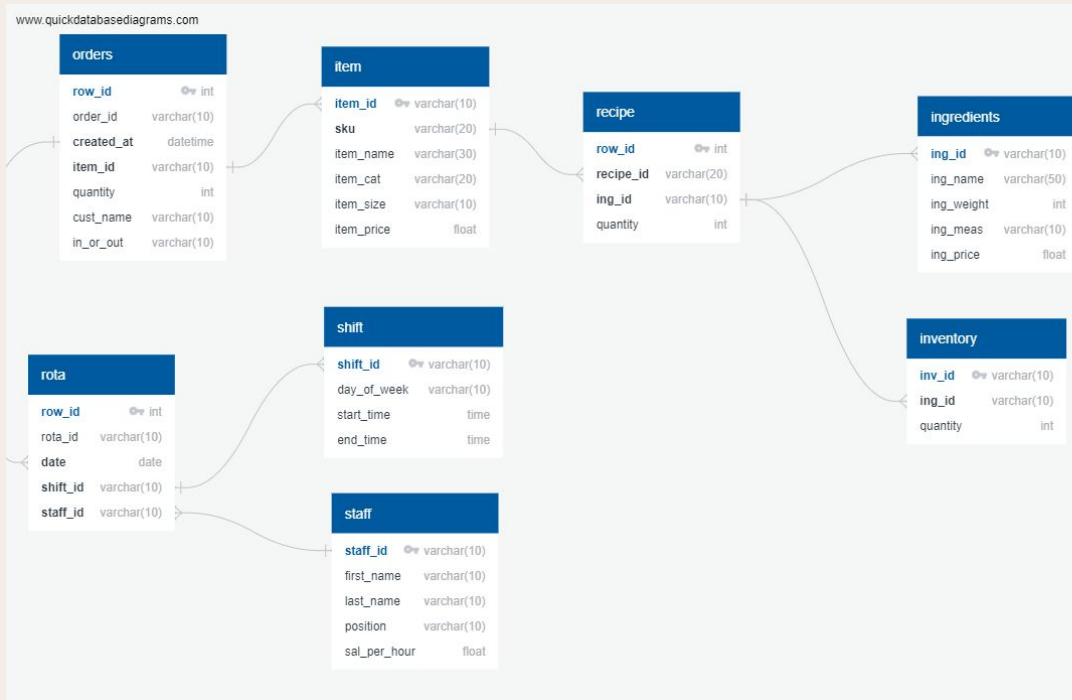Designing a dashboard to display the findings

Overviewing the Data

At the beginning of the project I looked over the CSV files on Kaggle to see each table individually and identify what primary and foreign keys I would use to create my relational database diagram for later use. I also kept an eye out for null values and other issues with the data. At this point in time, the only major issues with the data were null values in the "in_or_out" column of the orders table and a small number of mistyped order IDs from the same table. Where applicable, both of these null or erroneous data types were excluded from the data analysis and the overall findings are still representative of the business as per the data provided and the scope of the project.

Designing a Relational Database Diagram

Using Quick Database Diagrams, I made this chart that also allowed me to export the MySQL file for database I used. The connections between each table that denote primary and foreign keys carry over to database creation when you export the code that is used to create the diagram.

www.quickdatabasediagrams.com

orders

row_id	int
order_id	varchar(10)
created_at	datetime
item_id	varchar(10)
quantity	int
cust_name	varchar(10)
in_or_out	varchar(10)

item

item_id	varchar(10)
sku	varchar(20)
item_name	varchar(30)
item_cat	varchar(20)
item_size	varchar(10)
item_price	float

recipe

row_id	int
recipe_id	varchar(20)
ing_id	varchar(10)
quantity	int

ingredients

ing_id	varchar(10)
ing_name	varchar(50)
ing_weight	int
ing_meas	varchar(10)
ing_price	float

inventory

inv_id	varchar(10)
ing_id	varchar(10)
quantity	int

shift

shift_id	varchar(10)
day_of_week	varchar(10)
start_time	time
end_time	time

rota

row_id	int
rota_id	varchar(10)
date	date
shift_id	varchar(10)
staff_id	varchar(10)

staff

staff_id	varchar(10)
first_name	varchar(10)
last_name	varchar(10)
position	varchar(10)
sal_per_hour	float

Creating a Database for Querying

After making my diagram, I used Navicat for MySQL as my database tool to import the structure for the database that I made in the previous slide. Once the database was created and had the tables ready to fill, I began importing CSV files for each of the tables.

Writing SQL Queries to Find Relevant Data

With the database filled and having defined relations between tables, I then started writing my queries, which are available on my Github page for this project for clearer viewing.

Designing a Dashboard to Display the Findings

Finally, after querying the data to find answers to our business questions, I used Looker Studio to make dashboards that answer the business questions with interactive elements to sort and highlight data. While this presentation will explain the findings, you can find the actual report here, or as a pdf on the Github page linked on the last slide.

03
Findings

Information gathered from the case

Findings

Now we will cover the findings by each bullet point from the business questions slides:
- Total Orders: 461
- Total Sales: $1857.45
- Total Items: 466
- Average Order Value: $4.03
- Sales by Category: 58.2% Hot Drinks, 31.3% Cold Drinks, 10.4% Snacks
- Top Selling Items: The three most popular items being sold are Cold Mocha, White Mocha, and Caramel Macchiato
- Orders by Hour: Orders are highest between 7AM-9AM and 12PM-1PM
- Sales by Hour: Sales by hour follow orders by hour, with the aforementioned time period showing higher sales relative to orders, indicating that sales made at those times include more multi-item or more expensive orders
- Orders In or Out: 53% orders are dine-in, while 47% of orders are takeout (excluding the 17% of null values in the original orders table)

Findings (cont.)

- Total Quantity by Ingredient: Whole Milk, Barista caramel sauce, and Lemons are all especially low in stock, indicating a possible need to keep higher stock of those ingredients based on what items are selling well
- Total Cost of Ingredients: Total cost for the ingredients is broken down in the report along with the percent remaining and total quality of each ingredient
- Calculate Cost of Coffee: Each coffee item costs between $0.1 and $0.98 to make
- Percentage Stock Remaining by Ingredients: Refer to the previously mentioned table for the clearest representation of this data
- List of Ingredients to Re-order: Whole Milk, Barista caramel sauce, Lemons, and potentially other ingredients depending on the store's stocking practices
- Total Staff Cost: $960.00
- Total Hours Worked: 96
- Hours Worked by Staff Member: Refer to the table on the Staff Data page to see how many hours each employee worked
- Cost per Staff Member: $10 per hour

04 Improvements

What improvements could be made for future analysis

Improvements

If this were a real-world project and I was able to get ahold of the coffee shop to clarify some of the data, I would try to get some of the null and erroneous values clarified for better analysis. From my analysis, I found that orders had item codes attributed to them that did not exist in the items table (It025, It026, It027, It028). This combined with other data issues (such as one value being entered as It0010 instead of It001 or It010) and null "in_or_out" values rendered some portions of the data unusable depending on what question I was trying to answer, particularly the orders with missing item number values having to be excluded because I could not track the revenue on them.

Thanks