Group 2 Members:
Carreon, Ma. Addine Anne T.

Lucas, Sydney Anne V.

Peralta, Nathalya O.

Quipit, Dale Russel

Dataset: The analysis adopts a comprehensive systematic approach by utilizing the “Coffee Shop Sales/Inventory/Staff" from Kaggle, authored by Vira Mthck. (https://www.kaggle.com/datasets/viramatv/coffee-shop-data)

ASND's Coffee Shop: Knowing the Crowd's Favorites for Profitable Perks

The investigation of assessing how menu items impact customers' experience of being pleased and satisfied, which correlates to the profitability earnings, is analyzed to address the significance of orders and items to the coffee shop’s success.

Variable	Description
order_id	Customer order records
item_id	Menu item information
item_name	Specific menu items ordered by the customer
item_cat	Menu item category of the item name
item_size	Menu item size of the item name
item_price	Menu item prize of the item name
ing_id	Ingredient records
ing_name	Specific ingredient name
recipe_id	Recipe information records

Case/Scenario: A group of four friends, namely ASND, hired a data scientist for their newly opened coffee shop with the aim of investigating and determining the most popular order and its associated profitable item.

Question: What menu items keep our consumers pleased and our earnings high that can help us to improve the menu order list and the coffee shop’s progress?

Process

The end-to-end project provides an overview of coffee shop dataset findings regarding the most popular order and its associated profitable item, aiming to enhance menu items and progress the success of the coffee shop. In executing the group's desired objective, it follows a step-by-step process.

I. Data Finding

The group first initialized to seek a data source that will cover the criteria to allow normalization, creating an entity relationship diagram, RegEx rules, synthetic data generalization, SQLite3 database, and various table ingestion. The used dataset is a coffee shop data that offers a comprehensive overview of coffee shop operations, covering orders, items, recipes, ingredients, inventory, staff, shifts, and rota. It provides valuable insights for optimizing coffee shop operations and is a realistic simulation for data analysis and business optimization.

II. Data Story

With the provided dataset, the group discusses the potential case or scenario that will align to their business question. The group assumes to be data scientists that will examine the data given from the newly opened coffee shop. They aim to discover what makes customers happy and what brings in the most money.

III. Creating a problem statement

In order for the group to attain proper needed visualization, they created a problem statement that focuses on understanding what makes their coffee shop successful by looking at their sales, inventory, and recipe data. They want to see what makes customers satisfied and what helps them make more money, through analyzing how people order, which items are popular, and what ingredients they use the most.

IV. Data Preprocessing

The group uses tools such as pandas and SQL for analyzing and discussing the structure and in the preparation of the dataset. In data processing, pandas was used for reading, cleaning, and manipulating data from Excel and CSV files, including tasks like selecting relevant columns, renaming, ensuring consistency, handling missing values, transforming data types, and merging tables. While the utilization of SQL queries are to create normalized tables in a SQLite database, defining schemas, establishing relationships, ingesting preprocessed data, and extracting, filtering, and aggregating data directly from the database.

The pandas data preprocessing shows multiple tables that includes order details, item details, ingredient inventory, and recipe ingredient details. These tables display as a function to

gather information that is needed in data exploratory analysis. Following that is the SQL's data preprocessing that reveals the ERD of the coffee shop. The group created an ERD that structures the relationship of the database acquiring a total of five tables namely: order, item, inventory, recipe, and ingredient. Following the ERD is defining the RegEx rules. The utilization of RegEx is employed to verify each row in the data frame and provides a list of tuples indicating if the row passes validation.

V. Data Exploration

The group explores the ordering patterns, identifies popular items, and assess their profitability. In the data exploratory analysis, customer ordering patterns, item popularity, and profitability were analyzed applying both Pandas and SQL. For panda's data exploration, data visualization of item categories distribution, customer distribution by order type, top selling item orders, and number of orders per item. This panda's data visualization helps to assess the objective of the group in understanding what makes customers happy and what brings in the most money.

The SQL's data exploration uses the SQL database that is employed in the data preprocessing. This includes data visualization of Top 10 most popular orders, Top 10 most profitable orders, average order value over time, inventory status, and ingredient price distribution. Through this visualization, the highest popular orders are obtained to respond to the group's objective.

VI. RND & Conclusion

After the data preprocessing and data exploration of both pandas and SQL, the group presented the key findings regarding the most popular order and its associated profitable item, along with discussions on their implications for menu optimization and business growth. The results and discussion shows the detailed assessment and understanding of the group in the provided data preprocessing and data exploration. In which allows them to conduct a conclusion based on the results and discussion. The group's conclusion summarizes the analysis and provides actionable recommendations for improving the menu order list and maximizing the coffee shop's progress.