**IFT 533: Data Visualization & Reporting for IT**

Alok Sinha, Divyansh Rajesh Gupta, Jahnavi Singh, Sanskruti Suryabhan Jalte

Project Phase 1

Planning

04/07/2024

# Section 1: Dataset Description

The dataset is regarding the coffee shops of a certain franchise based over different locations with multiple stores. The dataset provides a valuable insight regarding as to how the store is doing in business and gives much more information regarding its day-to-day business.

## Columns in the dataset

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Category** | **Domain** |
| Transaction\_id | Nominal | 149,116 Distinct values |
| Transaction\_date | Interval | Date range– From 1/1/2023 to 6/30/2023 |
| Transaction\_time | Interval | Time ranges from 6am until 9pm |
| Store\_id | Nominal | 3,5,8 |
| Total\_Bill | Ratio | 0.8, 1.6, 2, 2.1, 2.2, 2.4, 2.45, 2.5, 2.55, 2.65, 3, 3.1, 3.2, 3.25, 3.5, 3.75, 4, 4.06, 4.2, 4.25, 4.38, 4.4, 4.5, 4.69, 4.75, 4.9, 5, 5.1, 5.3, 5.63, 6, 6.2, 6.3, 6.4, 6.5, 6.6, 7, 7.35, 7.5, 7.6, 7.65, 7.95, 8, 8.5, 8.95, 9, 9.25, 9.3, 9.5, 10, 10.5, 10.95, 11.25, 12, 12.75, 13.33, 13.5, 14, 14.25, 14.75, 15, 17, 18, 19.75, 20.45, 21, 22.5, 23, 24, 28, 36, 45, 56, 72, 360 |
| product\_category | Nominal | Tea, Coffee, Bakery, Branded, Coffee Beans, Drinking Chocolate, Flavours, Loose Tea, Packaged Chocolate |
| product\_type | Nominal | Brewed Herbal Tea, Barista Espresso, Biscotti, Black Tea, Brewed Black Tea, Brewed Chai Tea, Brewed Green Tea, Chai tea, Clothing, Drinking Choclate, Drip Chocolate, Espresso Beans, Gourmet Beans, Gourmet Brewed Coffee, Green beans, Green tea, Herbal Tea, Hot Chocolate, House blend Beans, Housewares, Organic Beans, Organic Brewed coffee, Organic chocolate, Patry, Premium Beans, Premium brewed coffee, Regular syrup, Scone, Sugar free syrup. |
| product\_detail | Nominal | Peppermint, Almond Croissant, Brazilian, Brazilian – Organic, Cappuccino, Carmel syrup, Chili Mayan, Chocolate Chip Biscotti, Chocolate Croissant, Chocolate syrup, Civet Cat, Columbian Medium Roast, Cranberry Scone, Croissant, Dark Chocolate, Earl Grey, English Breakfast, Espresso Roast, Espresso shot, Ethiopia, Ginger Biscotti, Ginger Scone, Guatemalan Sustainably Grown, Hazelnut syrup, I Need My Bean! Diner mug, I Need My Bean! Latte cup, I Need My Bean! T-shirt, Jamacian Coffee River, Jumbo Savory Scone, Latte, Lemon Grass, Morning Sunrise Chai, Oatmeal Scone, Organic Decaf Blend, Our Old Time Doner Blend, Ouro Brasileiro shot, Peppermint, Primo Espresso Roast, Scottish Cream Scone, Serenity Green Tea, Spicy Eye Opener Chai, Sugar Free Vanilla syrup, Sustainably Grown Organic, Traditional Blend Chai |
| Size | Nominal | Small, Medium, Large |
| Month Name | Nominal | January, February, March, April, May, June, July, August, September, October, November, December |
| Day Name | Nominal | Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday |
| Hour | Interval | 24-hour format |
| Month | Ordinal | Range from 1 to 12 |
| Day of Week | Ordinal | Range from 1 to 7 with 1 being Sunday and 7 being Saturday |
| Store\_location | Nominal | All the unique store locations present in the dataset (e.g., "Astoria", "Hell’s Kitchen", " Lower Manhattan", etc.) |
| Product\_id | Nominal | Positive integers (e.g., 1, 2, 3, ...) |
| Transaction\_qty | Ratio | Positive integers (e.g., 1, 2, 3, ...) |
| Unit\_price | Ratio | Positive real numbers (e.g., 2.99, 3.75, 4.50, ...) |

* **Transaction\_id:** A unique identifier for each individual customer purchase.
* **Transaction\_date:** The date on which the transaction took place.
* **Transaction\_time:** The time of day at which the transaction took place.
* **Store\_id:** A unique identifier for the coffee shop location where the transaction occurred.
* **Total\_Bill:** This column likely represents a quantitative measurement that can be compared, has a true zero point, and arithmetic operations like addition and subtraction are meaningful.
* **product\_category:** The general category to which the product belongs, which aids in categorizing products into separate segments for analysis.
* **product\_type:** A more particular type or classification of the product, allowing for more detailed categorization inside the larger category.
* **product\_detail:** A detailed description or specification of the product, providing more information about its specific features or attributes.
* **Size**: Indicates the dimension or quantity of coffee orders, reflecting variations such as small, medium, or large.
* **Month Name**: Denotes the specific name of the month in which the coffee order was placed, facilitating chronological organization and analysis.
* **Day Name**: Specifies the name of the day of the week corresponding to when the coffee order was made, aiding in understanding weekly patterns and trends.
* **Hour**: Signifies the specific hour of the day at which the coffee order was placed, providing insights into peak hours of activity and customer behavior.
* **Month**: Represents the numerical value assigned to each month, offering a standardized and structured approach to analyzing temporal data.
* **Store\_location :** This column represents the physical location or address of the coffee shop store where the transaction took place. It could be a descriptive name (e.g., "Hell's Kitchen," "Downtown," etc.) or an actual address. This information is useful for analyzing sales performance across various locations and identifying high-performing or underperforming stores.
* **Product\_id** : This column contains a unique identifier or code assigned to each product sold at the coffee shop. It serves to distinguish between various products without relying on product names, which may be lengthy or inconsistent. This column is crucial for analyzing sales and performance at the individual product level.
* **Transaction\_qty** : This column represents the quantity or number of units of a particular product purchased in a single transaction. For example, if a customer buys three cups of coffee, the Transaction\_qty for that transaction would be 3. This information is essential for calculating total sales, revenue, and understanding product demand and popularity.
* **Unit price**: The Unit\_price column specifies the price charged for a single unit of a particular product. It is typically represented as a numerical value, likely in the currency used by the coffee shop (e.g., $3.00 for a large coffee). This column, combined with the Transaction\_qty, allows for calculating the total bill or revenue generated from each transaction.

# Section 2: Prospective Dashboard Users

1. **Future new owner of the franchise:** Apotential new ownercan use the dashboard to see customer trends (popular drinks, peak hours), analyze sales (revenue by time, product performance), and optimize operations (staffing, inventory) - all leading to informed business decisions.
2. **Coffee Shop Owners/Managers**: They can use the dashboard to analyze sales performance, identify popular products, monitor inventory levels, optimize staffing requirements, and make data-driven decisions to improve operations and profitability.
3. **Business Analysts**: They can use the dashboard to conduct in-depth analysis of sales data, identify patterns and correlations, and generate insights to support strategic decision-making.
4. **Marketing Team**: They can use the dashboard to assess the effectiveness of marketing campaigns, identify potential areas for promotion, and analyze customer preferences.
5. **Financial analysts**: Interested in the revenue aspects, analyzing total bills, and the relationship between product types/categories and sales performance.
6. **Supply chain specialists**: To monitor and optimize the flow of products from different categories and types, ensuring that the supply meets the demand.

# Section 3: List of User Requirements & Potential Questions

1. How do total sales fluctuate throughout the week for each store location?
2. Analyse the pattern in total sales over a day for each day of the week.
3. What is the average amount per order
4. Check on the sale of different product categories for each location.
5. What is the top-selling product types in each product category?
6. What is the trend of total bill over time during the day?
7. Can we identify any seasonal patterns in sales for different product categories?
8. How do product details affect the total bill in different product categories?
9. How does the presence of discounts or promotions affect sales volume?
10. Can we identify any correlation between transaction time and sales performance?
11. What is the distribution of transaction quantities across different product categories?
12. How does customer loyalty, measured by repeat purchases, vary across different product types?
13. Are there any significant differences in sales performance between online and in-store transactions?
14. Can we identify any outliers or anomalies in sales data that may require further investigation?
15. How do sales volumes fluctuate throughout the day, week, month, or year? Are there any notable patterns or seasonality?
16. What are the peak hours and days for sales, and how does this information vary across different store locations?
17. Which store locations are the most profitable, and which ones may require operational improvements?
18. Can we identify any opportunities for cross-selling or product bundling based on customer purchasing patterns?
19. How does the inventory turnover rate vary across different products or locations?

# Section 4: References

Mural-<https://app.mural.co/t/datainferno5748/m/datainferno5748/1712259109140/ff0818ea885e7263f6f069800b5df742a0216c9a?sender=uf2bf15f167bea60cf8003234>

Dataset- <https://www.kaggle.com/datasets/divu2001/coffee-shop-sales-analysis>