About the Dataset :

Transaction records for Coffee House, a fictitious coffee shop operating out of three Outlets (Hell's Kitchen, Astoria, Lower Manhattan) in NYC Location. This Dataset includes the transaction date, timestamp and location, along with product-level details.

Data Description :

- `transaction_id` : Unique sequential ID representing an individual transaction
- `transaction_date` : Date of the transaction (MM/DD/YY)
- `transaction_time` : Timestamp of the transaction (HH:MM:SS)
- `transaction_qty` : Quantity of items sold
- `store_id` : Unique ID of the coffee shop where the transaction took place
- `store_location` : Location of the coffee shop where the transaction took place
- `product_id` : Unique ID of the product sold
- `unit_price` : Retail price of the product sold
- `product_category` : Description of the product category
- `product_type` : Description of the product type
- `product_detail` : Description of the product detail

In Mind Questions:

- **General Questions related to the existence of**

  - missing values?

  - wrong datatypes for columns?

  - complete duplicates in the data?

  - outliers in each column?

- **Recommended Analysis**

  - How have Coffee House's sales trended over time?

  - Which days of the week tend to be busiest, and why do you think that's the case?

  - Which products are sold most and least often? Which drive the most revenue for the business?

Conclusion

Insights :

- Year, Sales exclusively from 2023.
- `Month`, Most sales within a six-month period
- `Hour`, Most sales were from 8 am to 11 pm
- `Transaction quantity`, Most sales were singles followed by 2 lots
- `Stores`, The sales rates of the three stores were similar, with the most significant difference observed at the `Hell's Kitchen` store
- `Unit price`, Price distribution is `positively skewed`, We understand from this that most Coffee have lower prices, while a few Coffee have higher prices
- **Category :**
  - it's evident that the three stores offer a wide range of services.
  - `Coffee` stands as the best-selling product.
  - Tea also exhibits a high sales rate.
- **Product Type :**
  - In the `Bakery` category, the most sold item is the scone.
  - In the `Branded` category, the best-selling item is Housewares.
  - In the `Coffee` category, the best-selling item is Gourmet brewed coffee.
  - In the `Coffee beans` category, the best-selling item is Organic Beans.
  - The `Drinking Chocolate` category solely offers hot chocolate.
  - `Regular syrup` are the top-selling items in the Flavours category.
  - In the `Loose Tea` category, the best-selling item is Chai tea.
  - In the `Packaged Chocolate` category, the best-selling item is Drinking Chocolate.
  - In the `Tea` category, the best-selling item is Brewed Chai tea.

Bivariate Analysis & Visualizations

`Stores VS product categories`

- **Astoria Store :**

  - The highest selling product category is **Coffee** with a count of **20,025** units

  - Following Coffee, **Tea** stands as the second most popular category with a count of **16,260** units

  - Other categories like Bakery, Drinking Chocolate, Flavours, Coffee beans, Loose Tea, Branded, and Packaged Chocolate have varying counts, ranging from **7289** to **110** units, in **descending order**

- **Hell's Kitchen Store :**

  - Similar to Astoria, **Coffee** leads in sales, amounting to **20,187** units

  - **Tea** follows closely as the second most popular category with **15,277** units

  - Other categories, such as Bakery, Drinking Chocolate, Flavours, Coffee beans, Loose Tea, Packaged Chocolate, Branded, and Packaged Chocolate, have counts ranging from **7617** to **119** units

- **Lower Manhattan Store :**

  - **Coffee** is again the top-selling category with **18,204** units

  - **Tea** follows with **13,912** units

  - Bakery, Drinking Chocolate, Flavours, Coffee beans, Loose Tea, Branded, Packaged Chocolate have counts ranging from **7890** to **180** units, in **descending order**

Stores VS Most transaction_qty

- **Astoria Store :**

  - The most frequent transaction quantity observed is **1**, with a count of **30,207** instances

  - The second most frequent transaction quantity is **2**, noted **20,392** times

- **Hell's Kitchen Store:**

  - The most frequent transaction quantity observed is **1**, noted **29,793** times

  - The second most frequent transaction quantity is **2**, observed **20,932** times

  - Additionally, there are **8** transactions with a quantity of **8**

- **Lower Manhattan Store:**

  - The most frequent transaction quantity observed is **1**, recorded **27,159** times.

  - The second most frequent transaction quantity is **2**, seen **17,318** times.

  - There are occurrences of transaction quantities **3**, **4**, and **6**, but they are less frequent compared to quantities **1** and **2**

```
Stores VS Hour sales
```

- **Astoria Store :**

  - Peak hours are from **9 AM** to **10 AM**, and from **4 PM** to **5 PM**, with counts ranging between **5,083** and **5,291** during these periods

  - The least busy hours are from **7 AM** to **8 AM**, with counts ranging between **4,181** and **4,966**

- **Hell's Kitchen Store :**

  - Busiest hours are from **8 AM** to **10 AM**, with counts ranging between **6,767** and **6,909**

  - The least busy hours are from **6 AM** to **7 AM** and from **6 PM** to **8 PM**

- **Lower Manhattan Store :**

  - Peak hours are from **9 AM** to **10 AM**, with a count of **6,297**, and from **2 PM** to **3 PM**, with a count of **3,051**

  - The least busy hours are from **7 PM** to **8 PM**, with counts dropping to **75** and **125**

Category VS Total revenue

- **Top Revenue Product Types :**

  - **Barista Espresso** with a revenue of **91,406.20**

  - **Brewed Chai tea** with a revenue of **77,081.95**

  - **Hot chocolate** with a revenue of **72,416.00**

  - **Gourmet brewed coffee** with a revenue of **70,034.60**

  - **Brewed Black tea** with a revenue of **47,932.00**

- **Lowest Revenue Product Types :**

  - **Organic Chocolate** with a revenue of **1,679.60**

  - **Green tea** with a revenue of **1,470.75**

  - **Green beans** with a revenue of **1,340.00**

- **Other Observations :**

  - Various products contribute to revenue across different ranges, with a mix of beverages (like **teas, coffees**) and snacks (**biscotti, pastries**).

  - **Coffee-related products**, such as various brews and espresso variants, hold prominent positions in revenue generation.

  - The lowest revenue-generating products are predominantly related to **tea** and specific beverage variants.

```
Month VS revenue
```

- **Jun**e had the highest total revenue, reaching **166,485.88**, making it the most profitable month.

  - **May** followed closely behind with a total revenue of **156,727.76**, indicating strong performance.

  - **Apri**l and **March** also performed well, generating revenues of **118,941.08** and **98,834.68**, respectively.

  - **January** and **February** had comparatively lower revenue figures at **81,677.74** and **76,145.19**, respectively.