Name: Aqsa Younus
ERP: 18571

Dataset: Dashi Foods Dataset

Domain Knowledge

E-commerce data is the data that is constantly expanding, creating many possibilities for data analysts to see trends, and understand how the business model is working and what we could do to improve business models.

This dataset, in particular, is unique. This dataset belongs to the company Dashi Foods. It is a local Pakistani Company that is dedicated to innovate and produce new food products for convenience with taste. This dataset comprises various sheets which show a database like structure because of the various ID columns connecting all the sheets together.

The sheets are:

- **Customer Ledger:** This sheet has all the information regarding the balances of each customer. It indicates how the transactions are being ensued and through what means.
- **Customer:** A purely dimensional data, best for machine learning purposes since it has various boolean valued columns.
- **CustomerWise:** A sheet rich in KPIs ideal for e-commerce analysis. The sheet can provide answers to potential analysis questions regarding customer buying behavior.
- **Item**: Similar to the customer sheet, this sheet was also rich in dimensions by lacking in KPIs. However, it had a variety of boolean columns perfect for machine learning algorithms.
- **Purchase**: Contains data regarding the purchase of each item, the category and quantity of the product ordered by the customers.
- **DataDump**: Though named as datadump, this sheet was the most ideal. It had the right balance of KPIs and dimensions. It also has customer and product data together, which would aid in product/Customer behavior analysis.

Problem Statement

- To understand the trends E-commerce/Commerce Trends and Transactions of Dashi Foods.
- To illustrate customer buying preferences and behavior w.r.t E-commerce of Dashi Foods
- To explore in depth the customer and business relationship of Dashi Foods

Definitions and Key Terms:

- **Customer Ledger:** A customer ledger is a record of all of a company's accounts receivable. A customer ledger is a specific part of a company's general ledger devoted entirely to the company's transactions with its customers.

Questions for Data Understanding:

These questions are personal questions asked by me as a data analyst in order to understand the dataset better. I was having issues with the customer ledger dataset so I added this section for the Dashi Dataset

- Is credit good for a company?
    - Suppliers often look at your business credit score before offering terms, and having good credit makes it easier to negotiate favorable terms with them.

- What is a negative balance?
    - A negative balance occurs when the ending balance in an accounting record is the reverse of the expected normal balance. This expectation is based on an account's classification within the chart of accounts. A negative balance should arise relatively rarely. For example, if an asset account has a credit balance, rather than its normal debit balance, then it is said to have a negative balance.

Possible Dashboard Audience

- Marketing team of Dashi so that they could create marketing strategies accordingly
- CEO of Dashi so that they get an insight into how their business is running

Data Transformation/Wrangling/Cleaning

Because the data is large. I did extensive data wrangling through Python because it makes it much easier to understand the data and its nuances. I added the ipython notebook. The in-depth details are written on the markdown (objectives/assumptions/reasoning). Below, I summarize what I have done to transform the data.

- Divided the wrangling into multiple steps
- Checked for missing values
- Created msno matrix as well
- Changed data type of columns where needed
- Created a heat map to check for correlation in the columns
- Deleted columns that barely had any correlation
- Checked for unique values and count of potential dimensions to eliminate incorrect inputs
- Created histograms and box plots for anomaly detection
- Checked if any potential KPI needed imputation on missing values
- Removed commas from the string of numerics before converting them to float

(further details in the ipython notebook)

While using PowerBI, I had to do a little more wrangling since unexpected values that were not null, but were incorrect inputs. I had to fix those as well.

Also for anomalies, I always use box plots and histograms because they are not only easy to read but also can be manipulated by limiting the data created to make box plots and histograms. We can also remove anomalies to check how the distribution of the dataset is.

Final Dimensions and KPIs

(potential ones are in the sheet)

"The KPI acronym stands for key performance indicator—it's a metric that measures how projects, individuals, departments or businesses perform in terms of strategic goals and objectives. KPIs are a way for stakeholders to see if they're making progress or if the business is on track. A dimension is a structure that categorizes facts and measures in order to enable users to answer business questions."

Dimensions

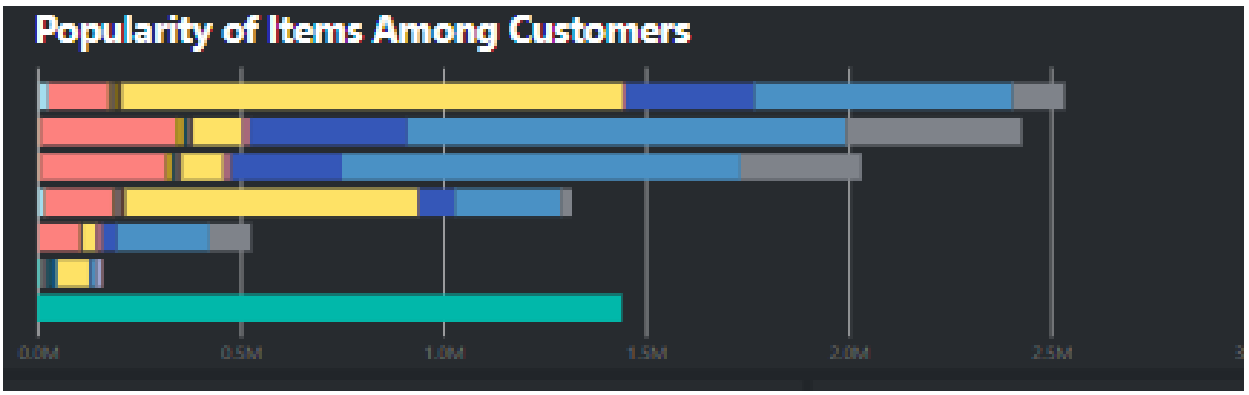
- Item
- Customer
- City
- Transaction Date
- Transaction Type
- Transaction ID
- Category
- Unit

KPIs

- Gross amounts
- Net Amount
- Amount (balances from customer ledger)
- Quantity
- packets delivered

Charts and Explanation

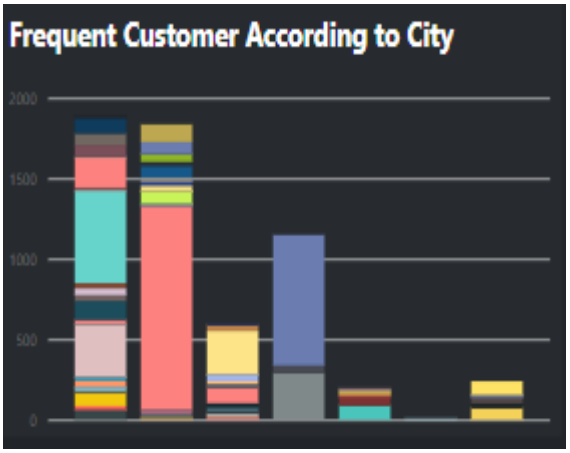
(the pixel might be a little bad because I zoomed these charts)

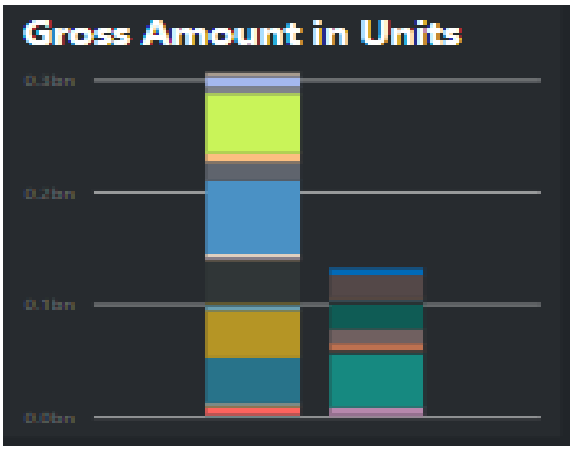


Quantity Items Bought w.r.t to Customer

The insight it provides:

- Comparison between popularity of items among customers
- Quantity of each item bought by a specific customer
- Showers trends of customer behavior when it comes to products of Dashi
- Explains customer preference for products
- Helps with creating personalized marketing strategies for each customer such as discounts, etc.





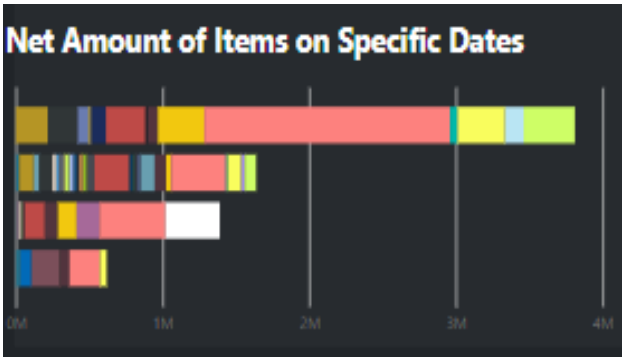
Quantity Items Bought from a Particular City w.r.t to Customer (left)

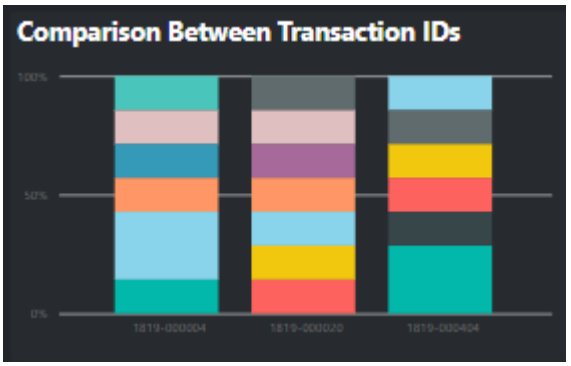
- Comparison of customer base of Dashi Foods in every city that they deliver products
- Most frequent customer in each of the city
- Comparison of customer buying trend according to city
- Quantity of products bought by each customer in the past three years.

Gross Amount w.r.t Item and Unit

- Understanding which is the preferred unit of buying among customers
- Which item contributes to the gross amount for Dashi
- How much does each item contribute to the gross amount
- Answers question *"Are things bought in pieces also bought in kilograms? If yes, then which one?"*

- Helps company understand which mode of packaging (units/kgs) should be more focused on.





Net Amount Accumulated by Item w.r.t Transaction Date (left)

(this chart personally would have looked better on tableau)

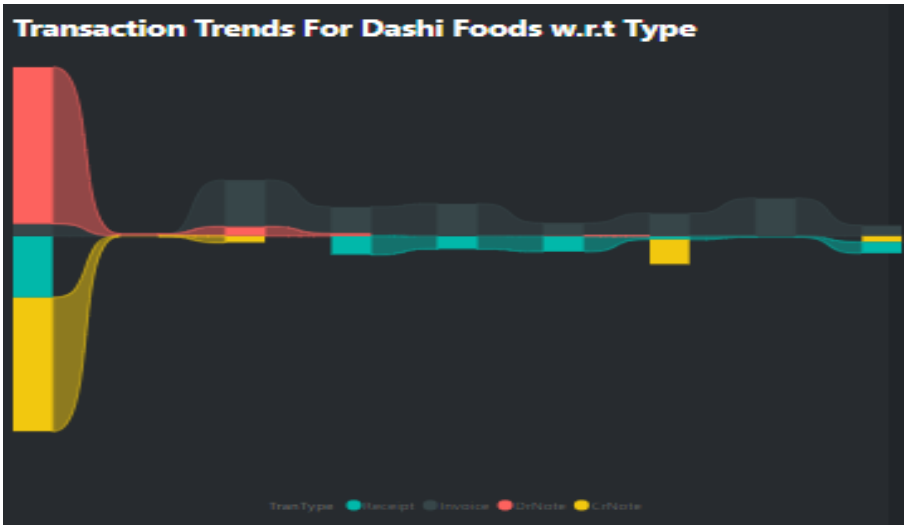
The insight it provides:

- Trends according to each date, which date has most/least net amount accumulated
- Understanding on why this trend occurs
- Explains which item contributes the most/least on specific days.

Gross Amount w.r.t to Transaction ID and Quantity Bought

The insight it provides:

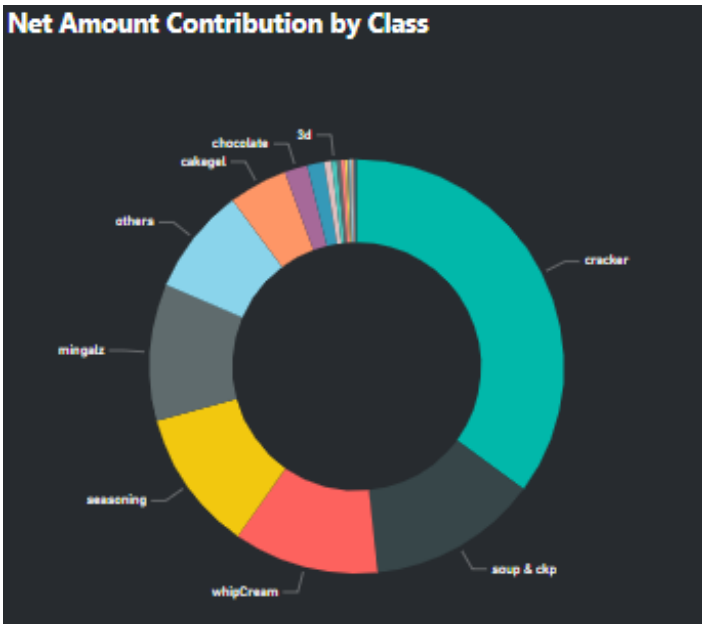
- Understanding how each transaction ID contributes to gross amount
- How much the quantity affects that gross amount bought
- Answers question *"does bigger quantity equate higher gross amounts?"*



Debit/Credit w.r.t Transaction Time Across Transaction Dates

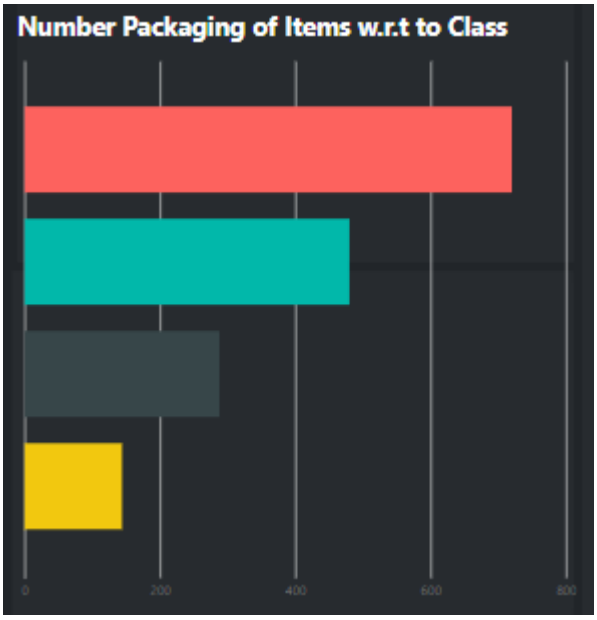
- Shows debit/credit transactions from the customer ledger in the past three years

- Subcategorized by different transaction methods to understand what mode of payment are customers comfortable with the most
- Negative part of the waterfall explains the negative balances in the customer ledger which can be linked to why negative/positive balances are occurring in certain types of payment methods



Net Amount w.r.t Class

- Provides comparison of different classes according to net worth
- Along with other charts, it explains how class and product yield different net worth.
- Zooming out from items to umbrella of items.



Packets Delivered w,r.t Item Name and Class

- Compares different classes and how many packets were delivered from each class
- Provides comparison packets delivered between classes as well
- Provides a zoomed in version of the gross amount in units to see the number of packets ordered/delivered customer buying preference inclined towards buying in pieces.

Potential Analysis

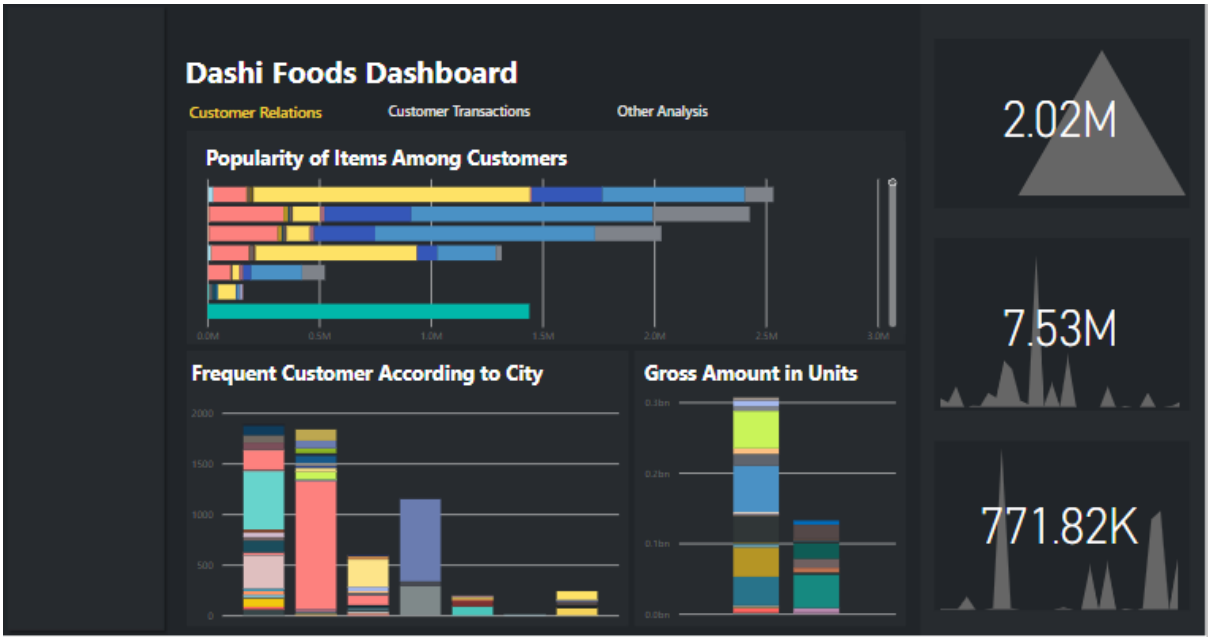
- Which item is the most popular among customers? Which customer has bought the most of that item?
    - *The seasoning chicken power 18 grams is the most popular among the Dashi customers. It was mostly ordered by Anus Jamsheed Road, qty 1237440.*
- From which city does Dashi have the most customers? Which customer is a frequent buyer from dashi?
    - *Dashi has a strong customer base in Karachi.*
    - *A customer named 786 (possible incorrect value input) from FCL is the most frequent buyer from Dashi.*
- Which item has been delivering the most gross amount to the company?
    - *Mingalz 12g, Tiki yellow has been delivering the most gross amount to the company with a total of 65548861 rupees from 2018 to 2022,*
- Which mode of unit delivery is producing the most gross amount for the company?
    - *Customers of Dashi prefer buying the product in pieces rather than in overall grams, even though each piece has grams associated with it. However, is not taken into consideration.*
- Explain Transaction trend of customers in the last 10 Days of December 2018 with respect to DrNote Method of Payment
    - *We can see a dwindling nature at the last ten days of transaction of December 2018, starting with the first drastic drop of 98.78% and then slowly halting to a zero by the end of December.*
- On 1/1/2022, which item produced the most net amount for the company? Is it consistent every first day of the year?
    - *It can be noted that chicken seasoning powder has been producing the most net amount on the first day of 2022, and that has been remarkably consistent since 2020, which shows that the chicken seasoning is a very successful product.*

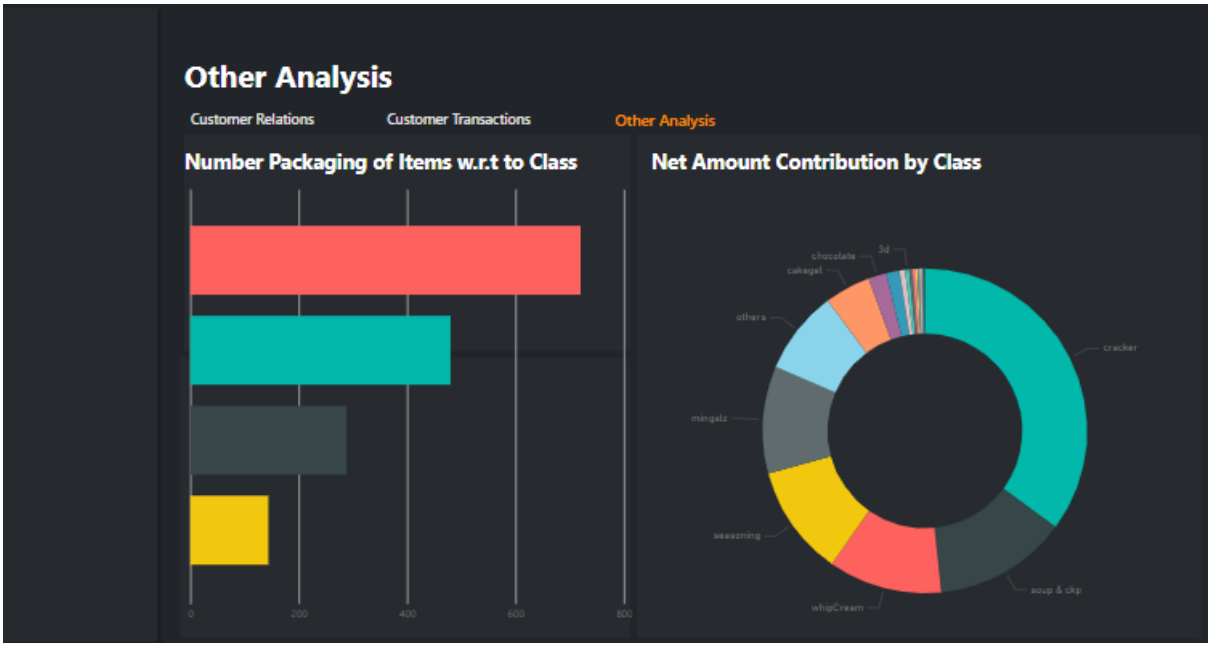
However, in 2019, white crackers and chicken seasoning were almost equally contributing to the net amount. However, that changed in the coming years.

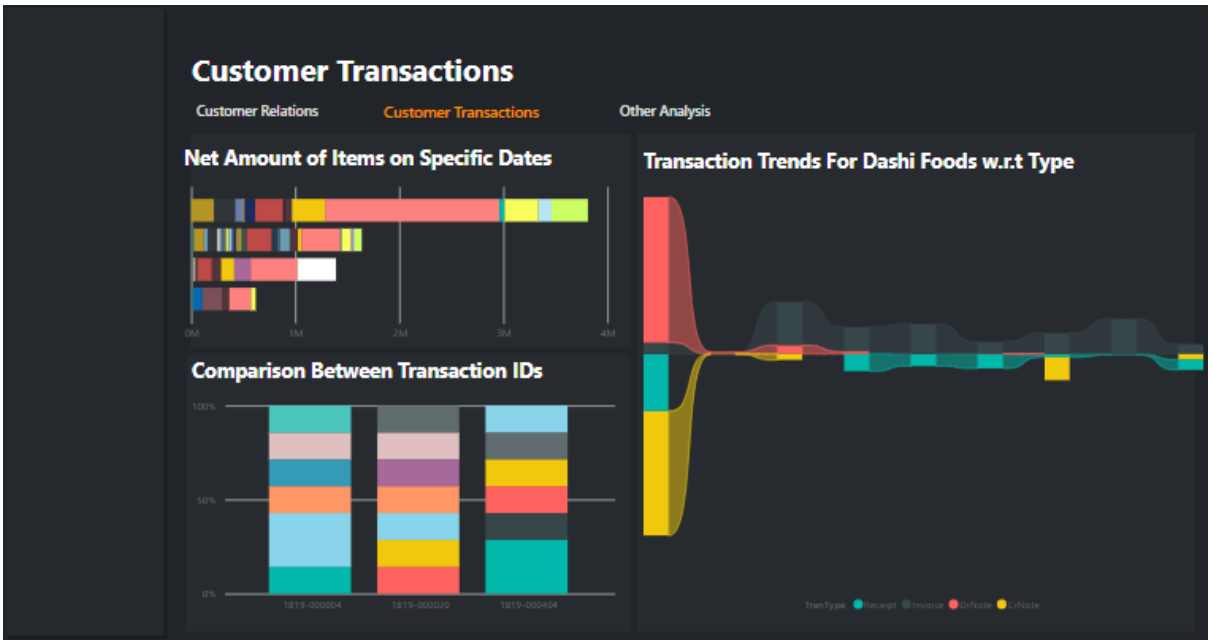
- Explain the contribution of gross amount in number of items bought from transaction ID 1819-00004
    - *The transaction ID has brought over 5700 products from Dashi. Interestingly, 144 Items contribute to more than 28% of the gross amount of what they have bought from Dashi.*

- Which class of items produces the most net amount for the company?
    - *Crackers have been producing the most net amount for Dashi, followed by soup and whip cream.*

- How many packets of green raita from the class raita were overall sold by Dashi in 3 years?
    - *Dashi sold 720 packets of green raita from the class raita in three years.*







Further/Dashboard Insights

Explanation for Dashboard:

- The dashboard is oriented according to the problem statements. The first dashboard contains charts related to customer relations with Dashi Foods and their behavior towards the product.
- Some KPI cards were added in the corner of the dashboard to explain the trends of each KPI from 2018 to 2022.
- The second dashboard is dedicated to charts that provide explanations regarding customer transactions. This includes trends in transactions and how the transactions are impacting the overall business of Dashi Foods.

- Because the first two dashboards were full of complex charts, I reserved the last dashboard (other analysis) for simpler charts that would assist in smaller and quick analysis of certain KPIs.
- If there were more KPIs, I think I would have been able to make stronger charts and add charts such as line, area chart, e.t.c.
- It should be noted that many charts are interlinked in the dashboards. I will explain that further under the next heading below.
- Through the dashboard, the company can get ideas and insights about the customer behavior, and look into each customer's behavior and trends that they have been showing for the past three years. This way they can also provide personalized discounts, e.t.c to their customers.
- Many of the charts in the screenshot are filtered versions of the chart in order to provide neatness and to answer queries. The filters can be manipulated according to the BI query required.
- The KPI cards reflect:
    - Currency against gross amount (to see gross trends in different currency)
    - Category against gross amount (to see gross trends in different categories)
    - Class against quantity (to see quantity ordered trends in different categories)

Further Insights:

The best thing about PowerBI is the interactivity of charts that they provide, which makes it easy to understand the links between all the charts.

- For instance, by interacting with the bars of popularity and gross amount charts, we can see that the most popular item is not the one that makes the most gross amount for the company. Dashi Foods can look into this and then create selling marketing strategies accordingly.
- We can also notice from the graph that just because Karachi has a large customer base for Dashi, does not mean that the most frequent buyer is also from Karachi. Therefore, we can notice that some cities have a diverse set of customers, while some have a limited but loyal set of customers.
- We can also notice that just because an item is popular does not mean that one specific customer buys that item the most. For instance, we can see that even though chicken seasoning is a popular Dashi Food item, Asia Food traders have bought yellow tikki more than chicken seasoning.
- Dividing the gross amount according to the unit helps us understand which mode of unit is preferred by the customer. The company can look into why it occurs to understand the behavior of their customers further.
- Similarly, the waterfall chart explains the transactions, the popularity of the mode of transactions and the debit/credit change of customers. Dashi can look into which mode of transaction is more popular and focus on marketing, such as discounts on those transactions.
- Another interesting thing to note about the data set, is that through the donut chart, we see that the crackers class produces the most net amount whereas seasoning class

comes at 4th in terms of net amount which is remarkable considering chicken seasoning is a lot more popular than white crackers.
- Lastly, looking into net amounts on specific dates helps us analyze trends according to each day would offer insights about seasonal trends of Dashi, for instance, on what time of the year does the sales go higher and when do they not? e.t.c
-

Final Thought/Opinions

About the Data:

- The data had a lot of dimensions by hardly any KPIs, a lot of the KPIs were repeated across multiple sheets, making most of the data redundant
- The data was clean, though, for null values. Missing values required little imputation, especially for the KPIs, which is why the analysis is strong for them.
- The data was incomplete. I believe more sheets would have given proper insights to the dynamics of the business, therefore would have given even better analysis.
- There was a slight confusion whether the data was entirely e-commerce. I believe the ledger report and godown column make it seem like the data partly alludes to physical transactions, so I kept that in mind during wrangling and analysis.
- There were a lot of boolean columns, especially for customers and items, making it an ideal dataset for some machine learning algorithms as well.
- There were a lot of incorrect inputs for customer names. I think I could only fix it if I had personally talked to someone who maintained the data because categorical data is best fixed when talked to the owner of the dataset.

Lastly, I understand that the dataset has limitations because it is obvious that they extracted the data from a database. We could observe relations through IDs in each of the sheets. However, from better analysis, I think having different datasets for each item would provide wonderful analysis. Unfortunately, that would be highly time-consuming to extract data and orient it entirely on items. Perhaps if I had done the course of data warehousing, I would have been able to brainstorm better about such a scenario.

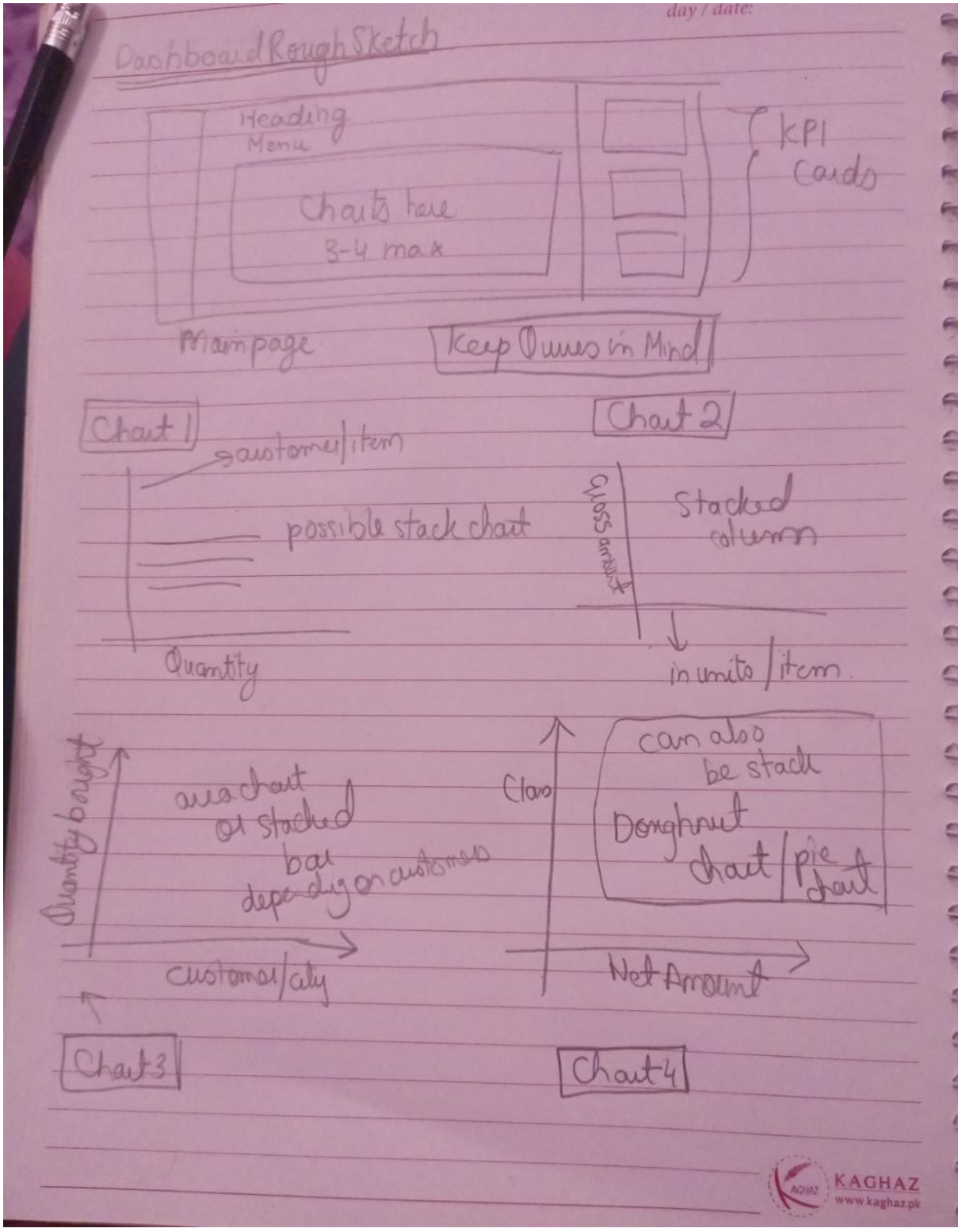
About the Final Project:

I personally would prefer python and Tableau over PowerBI. I felt restricted by the limited options that PowerBI provides relative to Python and it is more effort to create charts on PowerBI compared to Tableau. The data itself should have had more columns that one could analyze in order to get further insights into the E-commerce data of Dashi Foods.

Other than that, the excel sheet includes all the relevant data I could gather. In the Data information sheet, I added all the information necessary to explain the Dataset. The columns which have no information in regards to potential queries is because these columns were not a part of the analysis, and columns that were not part of the analysis were either dropped or ignored.

Rough Charts/Notes

The rough notes consist of possible dashboard design, which I took an idea from the previous PowerBI assignment, and brainstorming about potential charts and the targeted queries. I also wrote down my brainstorming thought process at the end.



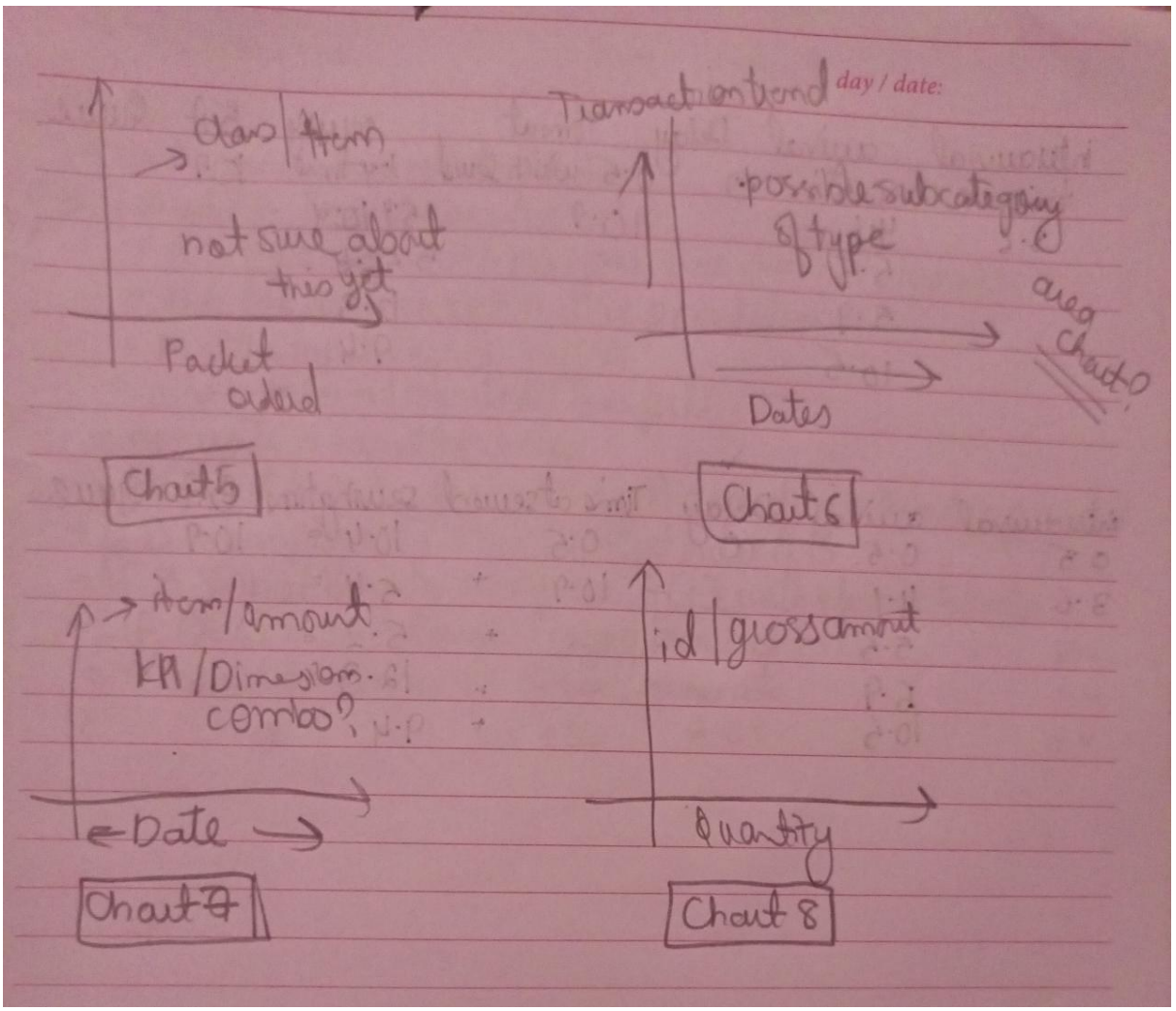

class / item
not sure about this yet
Packet added
Chart5
Transaction trend
day / date:
possible subcategory of type
Dates
area chart?
Chart6
item / amount
KPI / Dimensions combos?
Date
Chart7
id / grossamount
Quantity
Chart 8

Chart ①

→ Will answer queries of popularity of an item and possibly will explain which customer buys which item the most

Chart ③

→ will answer possible queries regarding which customer is more frequent, where do they belong from, and which city has most customer contributions from.

Chart ②

→ will answer queries that are related to orders (i.e which mode of ordering is popular pieces or kgs?) which item contributes to most gross amount.

Chart ④

→ This will answer simple queries such as which class has most/least net amount

Chart ⑤

→ multiD, will give answers regarding packets ordered w.r.t to class & item.

Chart ⑥

→ trend across time on transactions.

day / date:

Chart (7)

→ time series graph for item/amount qnts

→ Chart (8)

Data related to customers and their contribution to gross amount wrt to quantity they ordered

Things to keep in mind

1. There is a possibility not all charts would be made
2. Dashi Foods is not entirely e-commerce
3. Possible that not all Dimensions would be used
4. charts may not be what is in your mind
5. Trends are important!
6. Avoid Unnecessary charts
7. Wrangling impacts charts.

Note of Acknowledgement:

I would like to thank Dr. Tariq Mahmood for being a wonderful instructor throughout the semester. Not only did I learn various analytic skills and tools, but after three years of studying computer science, I finally realized what domain I can pursue at the start of my career. Thank you for introducing me to data analytics and business intelligence. Hopefully, I will be able to continue to hone my skills in the future. And if there are any opporunities where I am able to put my skills to practice, please let me know. JazakAllah.

Reference(s):

http://www.dashifoods.com/about-us/
https://agrihunt.com/articles/food-industry/various-food-industries-in-pakistan/
https://fundbox.com/resources/guides/business-credit-score/#:~:text=and%20very%20important.-,Establishing%20good%20credit%20scores%20is%20important%20because%3A,for%20establishing%20lines%20of%20credit
https://www.accountingtools.com/articles/what-is-a-negative-balance.html