

# School of InfoComm Technology

**Data Exploration & Analysis Assignment**

Diploma in DS

April 2022 Semester

**ASSIGNMENT 2**

(40% of DEA Module)

11th July 2022 – 7th August 2022

**Submission Deadline:**

**Presentation: 7th August 2022 (Sunday), 11:59PM**

**Report and files: 7th August 2022 (Sunday), 11:59PM**

|  |  |  |
| --- | --- | --- |
| **Tutorial Group** | **:** | **T01 / T02 / T03** |
| **Student Name** | **:** | Ang Jing Yi Clairer |
| **Student Number** | **:** | S10223223 |

**Penalty for late submission:**

10% of the marks will be deducted every calendar day after the deadline.

**NO** submission will be accepted after 14th August 2022 (Sunday), 11:59PM.

Contents

[School of InfoComm Technology 1](file:///C:\Users\angcl\Documents\ASG2\DEA\ASG2.docx#_Toc110849607)

[Abstract/Overview 2](#_Toc110849608)

[Creating Dashboard to answer business question 2](#_Toc110849609)

[Further Analysis 6](#_Toc110849610)

[Summary 7](#_Toc110849611)

# 

# Abstract/Overview

This report’s purpose is to understand the FoodFresh dataset as well as answering the business questions that is given as well as my own question. Later on after answering the questions, I did further analysis on the trends that I had found while creating the dashboard to better improve the business transactions.

# Creating Dashboard to answer business question

Before starting the creation of dashboards, I went to first transform the dataset to include what I need like ‘Sales’, ‘Revenue’ and ‘Age (groups)’. During the creation for each dashboard, I also excluded data that has blanks as due to the column of CustomerID under the Trans table I have realised some transactions does not have CustomerID which resulted in blanks as relationship between the Trans Table and Customer Table was supposed to be the primary key ‘CustomerID’.

For the first Dashboard, I created it to answer the first business question ‘Which period has the most Sales?’. I first created the filters for the month and year where people can interact with the filters to find out the data they want to see. Secondly, I created ‘Monthly Sales in each Year’ where the line chart can view each of the month’s sales. I disable the month filter for it as it is based on year. Thirdly, I created ‘Sales by Quarter in each Year’ where people can view the bar graph for quarters of the Year (e.g., Jan-Mar – Qtr. 1). After that, I created the ‘Daily Sales by Each Month of the Year’ where people can view the line chart for each month’s daily sales to see the highest and lowest for each month. Then I created the ‘Sales by Product Category for Each Month of the Year’ where people can view the bar graph for each product category sales so that they can understand which products earns the most sales for each month or year to allow them to continue advertising these products for better sales. Lastly, I created a card for ‘Highest Daily Sales’ for each month to allow people to better understand how these days have high sales.

Graphical user interface

Description automatically generated

For the second dashboard it is answer the question ‘Who are the Top 10 High Spending customers?’. I first created the filters for the month, year and the brand name where people can interact with the filters to find out the data they want to see. Then I created ‘Top 10 Customers’ by using bar graph to present the top 10 spending customers. After that, I created a scatter chart to show the distribution of ‘Sales by Gender and Age’ to allow the company to see how different gender in each age group spends. Then I created a pie chart to see the ‘Top 10 Age group by Revenue’ to allow company to understand where most of the revenues are from which age groups. Lastly, I created a card to present the Total Revenue to allow company to see the overall revenue they have earned.

Graphical user interface, chart, application, Excel

Description automatically generated

For the third question ‘What are the Top & Bottom 5 Selling items/brands?’. I first created the filters for the year and the brand name where people can interact with the filters to find out the data they want to see. I first created the ‘Revenue by ProductCategory’ by using a pie chart to show percentage of revenues each product category contributes. Then I created 2 cards one for ‘Top 5 Selling items’ and another for ‘Bottom 5 selling items’ to show which product contributes the most revenue which will allow company to focus more on the most sales product while for bottom selling products company can take note of it and produce lesser stocks of it to reduce their costs.

Chart, pie chart

Description automatically generated

For the last question, I created the question ‘What are the Top Selling items in each Region?’. This question is to figure out the distribution of sales of top products so that company can focus these products for different region by promoting it. Since these products contributes the most revenue for the company it is natural that more of such products are to be manufactured and be able to satisfy the customers in the region.

I first created the filters for the month, year and the brand name where people can interact with the filters to find out the data they want to see. Then I created map ‘Top 5 most Sold Product Category in each Region’ where company can see the top 5 product category in the pie charts of each region and allow them to compare and adjust their products accordingly to the demands. Then I created a multi-row card to present each product category and the revenue they contribute. Lastly, I created ‘Top 5 Product Category by Revenue and Brand Nam’' by using a bar chart to show what are the top 5 product categories and in each of the top 5 categories which brand customers buys from the most.

Map

Description automatically generated

Finally, from all this created boards, I picked some visuals from each of them and combined them to make the final dashboard that allows company to view the required data they need to answer business questions.

I first created the filters for the month, year and the brand name where people can interact with the filters to find out the data they want to see. Then I created 2 cards, one for the Total Revenue and the other for Total Sales to allow company to view the overall revenue and sales. Then I created a pie chart ‘Sales by BranchRegion’ to show the different stores located and how much they contribute within the different regions. I also created ‘Sales by Region and BrandName’ bar chart to show how much each brandName contributes to the revenue of each region. After that, I took the map from question four which shows ‘Top 5 Most Sold Product Category in each Region’ to answer question four. Then I took the bar chart ‘Top 10 Customer’ to show the top 10 spending customer which will answer to the second question. I then took the pie chart ‘Revenue by ProductCategory’ and I also took the Top and Bottom 5 selling items and convert them into a bar chart to answer question 3. Lastly, I took the ‘Monthly Sales in each Year’ and ‘Daily Sales by each Month of the Year’ to answer question one.

Graphical user interface, application

Description automatically generated

# Further Analysis

After doing the questions and the overall dashboard, I realised a trend and decided to analyse it further through creating another dashboard to explore how it contributes the trend. The trend I saw was the West produced more Revenue among the 5 Regions. I first created a bar chart ‘Sales by region and Age (groups) which will show me how different age group spends in different regions which I can see, those who are aged 71 and above spends the most amongst the other age groups which I can infer that they have more spending power considering they are most likely retired and has the most disposable income compared to the rest. Also, I created the filter for the different Brand Names to see what products are usually bought. From filtering to ‘KopiFresh’, I can tell that noodles from ‘KopiFresh’ are frequently bought by consumers which can allow company to focus on the noodle products more to allow consumers to continue buying them as well as attracting new consumers to buy their noodles from ‘KopiFresh’. Lastly, I created ‘Count of Name by Region’ bar chart to see if the sum of people in each region contributes to how each region sales are. From this bar chart I can infer that it is true as West holds the most consumers, a total of 565, unlike the other regions who holds less than 500 people. This makes West Region to have the most advantage over the other Regions as there are more people spending compared to other Regions.

Graphical user interface, chart, application, sunburst chart

Description automatically generated

# Summary

In summary, to answer the four business questions, the final dashboard I have made can answer the questions.

To answer question 1, May 2020 has the most Sales compared to all the Months of the current data.

For question 2, ‘Koh Eng Mei’, Chong Kuan Choon’, ‘Goh Beng Yee’, ‘Yong Chun lay’, ‘Ng Gim Yue’, ‘Khoo Chun Lian’, ‘Leong Hock Chun’, ‘Chen Poh Tee’, ‘Chang Lian Yew’ and ‘Chen Lay Kuan’ are the top 10 highest spending customers.

For question 3, Top 5 selling items are ‘Black Forest’, ‘Peanut Butter Addiction’, ‘Strawberry Shortcake’, ‘Orea Cheese’ and ‘Coffee Indulgence’ while the Bottom 5 selling items are ‘Fire Floss’, ‘Curry Dozo’, ‘Road Hog’, ‘Smart Alack’, ‘Peanut Butter Thick Toast’.

For question 4, the top 5 most sold product in each region are Appetiser, Cake, Desserts, Noodles and Rice.

For trends of why West contributes the most Sales is due it holding the most amount of people compared to the other regions who only has less than 500 people.