



# FRESH OUT OF THE OVEN

Bakery Business Report

Prepared For  
**DR ZAINURA BINTI IDRUS**

Prepared By

# PROBAKE

Hot Guys Who Bake and Smile  
for Passion

# TEAM MEMBERS



## MUHAMMAD AMIR FAHMY BIN MUHALITH @ MUHALIS

2021149601

Data Engineer



## MUHAMAD FAIDI AKIF BIN MD ALI

2021196337

Data Scientist



## MUHAMMAD SYAZWAN FIKRI BIN SAHRAN

2021100233

Data Engineer



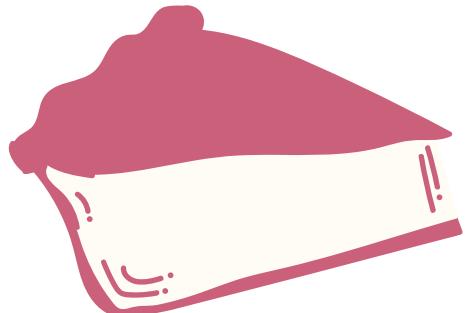
## ZUBLI QUZAINI BIN ZUBLI

2021119947

Project Manager

# CONTENTS

 **PLANNING**



 **METHODOLOGY**

 **DESIGN & DEVELOPMENT**

 **ANALYSIS & FINDINGS**

 **CONCLUSION**



# **CHAPTER 1**

# **PLANNING**

**This chapter will discuss the background of the research**

# Introduction

Small-medium enterprise is among the main contributor to the local and global economy.

**78% of the SMEs are consist of micro-companies**

There are constant demand for mouth-savouring desserts and loaves of bread opens more opportunities for small business to enter the market.

However, there is a lack of sustainability factors for small companies to expand or grow their businesses.

Therefore, strategic planning plays an important role as a key factor in decision-making and critical analysis.

The advancement of technology in Big Data and Data Visualisation allows gaining insights and hidden pattern in data that is useful for the formulation of strategic planning in the context small businesses such as bakery store.



# Problem Statement



**There is a rise in demand among small businesses to understand the preferences of the customers**

The problem statement of the research describes the current state of polemic among small businesses due to the difficulties in understanding the preferences of customers.

The business may have a large amount of data from the sales transaction to inventory management. However, the lack of analysis technique used causes the data to become a silo of information instead of knowledge.

In addition, it is known to believe that the application of statistical and machine learning techniques requires a high complexity and incurs a large amount of cost for further analytical execution, which became a huge barrier for small businesses. This research aims to open a new paradigm of effective data analysis and visualisation for small businesses.

# Aim

To develop a dashboard that allow exploration on data analytics and strategic business planning for small business

# Objective

- To collect data related to the bakery's daily sales transaction
- To design and develop a business analytical dashboard for bakery sales
- To extract information and customers' buying pattern based on the sales data

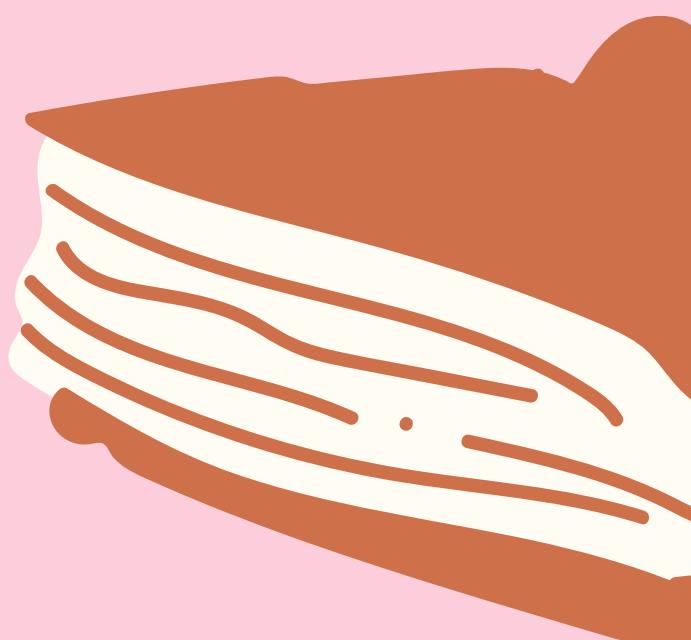
# Scope

The dataset used in this research is about Bakery Sales which contains data about bakery delivery services in Korea. The data is acquired from Kaggle containing 27 attributes and 2421 rows.

# **Significant**

The research will help the business owner to identify the main factor that affects the demand of bakery products

The business owner also can gain insight into the business performance in real-time

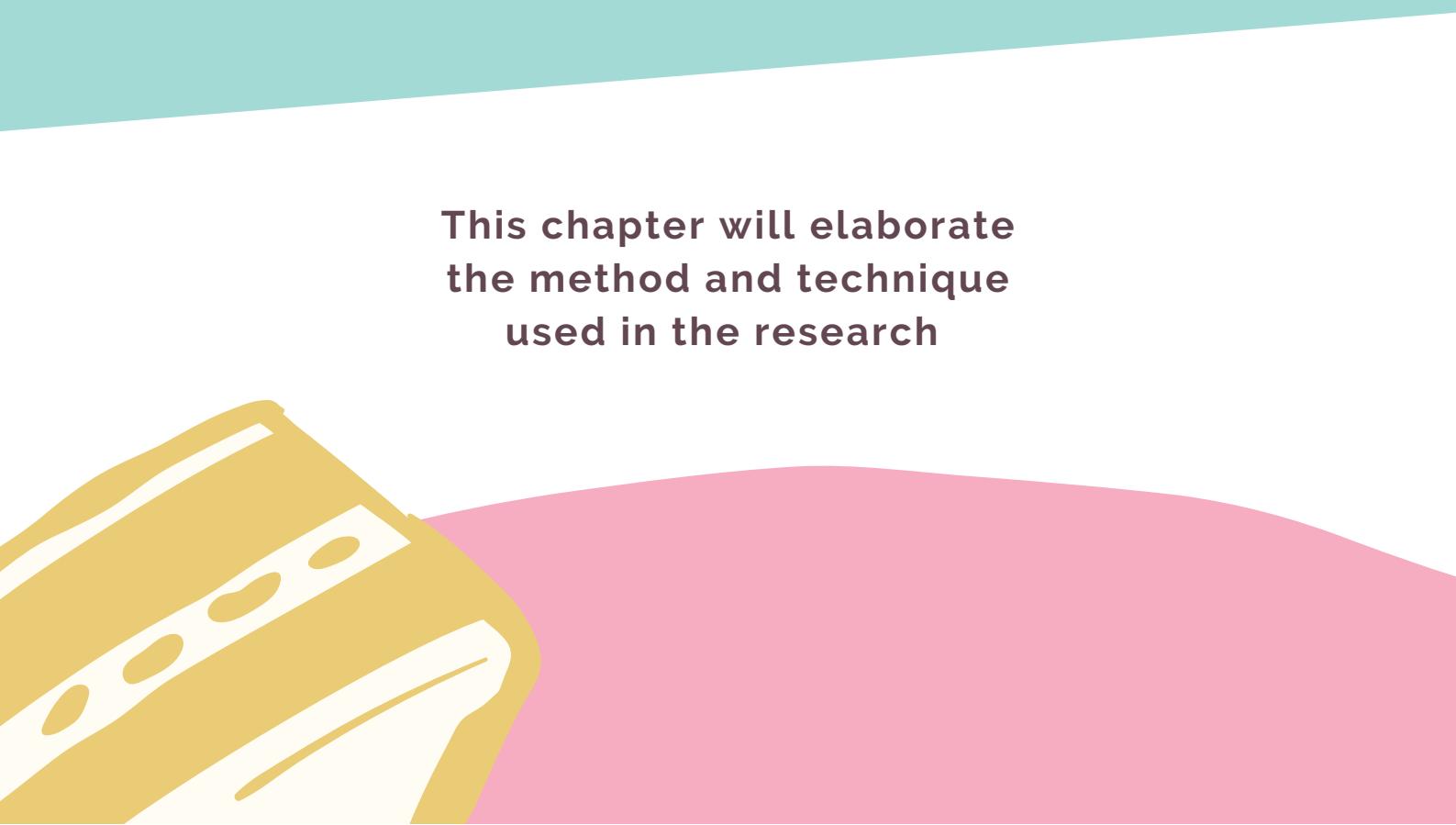


**Therefore, businesses can make critical decisions making to expand and identify new opportunity**



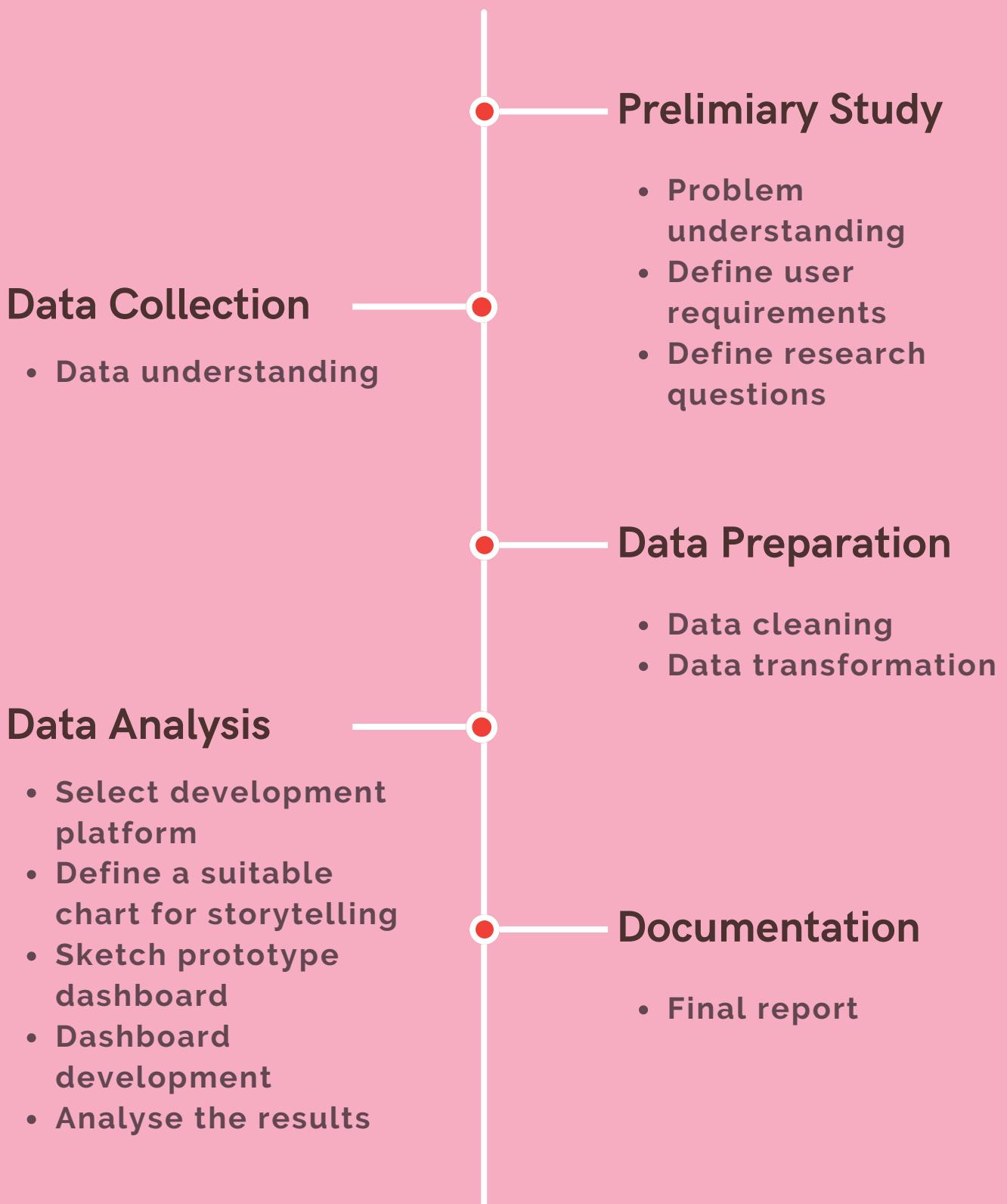
# **CHAPTER 2**

# **METHODOLOGY**



**This chapter will elaborate  
the method and technique  
used in the research**

# Research Flow



# User Requirements

The dashboard's purpose is to assist small bakeries in making analytical from daily sales data. The visualisation must be easy to understand, eye-catching and placed in an interactive manner to ensure compelling storytelling

## Functional Requirements

- Users can filter the dashboard based on the products
- Users can click on the visualisation to view the details
- Users able to make interpretations from the visualisation

## Non-Functional Requirements

- Users only have read access to the dashboard
- Only admin and owner can make changes
- The dashboard must able to update if there is any changes in the data

# Research Questions

The research questions describe the direction of the research toward business strategic planning



**What is the difference between orders during weekend and weekdays?**



**What is the total number of customers per month?**



**What is the association of sales between baked goods and drinks?**



**What is the distribution of orders based on region?**



**What is the total gross sales per month?**

# Data Collection

The dataset is about Bakery Sales, which contains the data about the delivery services for a small bakery in Korea

The screenshot shows a Kaggle dataset page for 'Bakery Sales'. At the top, there's a navigation bar with a user icon, the name 'HOSUB JEONG - UPDATED A YEAR AGO', a notebook icon labeled '37', a 'New Notebook' button, a download icon labeled 'Download (28 KB)', and a three-dot menu icon. Below the navigation is the title 'Bakery Sales' in bold, followed by a subtitle 'Our Bakery Sales(Delivery)'. To the right of the title is a thumbnail image of several croissants. Underneath the title, there are four tabs: 'Data' (which is underlined), 'Code (7)', 'Discussion (0)', and 'Metadata'. In the 'About Dataset' section, there are two columns. The left column contains 'Context' (describing work at a bakery in Korea starting in July 2019) and 'Content' (noting it's basket data with 27 columns and listing attributes like datetime, day of week, total amount, place, and angbutter). The right column contains 'Usability' (9.71), 'License' (Data files © Original Authors), and 'Expected update frequency' (Weekly). The entire page has a light pink background.

**Figure 2.1 Bakery Sales Dataset**

(Source: <https://www.kaggle.com/datasets/hosubjeong/bakery-sales>)

Figure 2.1 shows the Bakery Sales Dataset. The dataset contains 27 attributes and 2421 rows.

# Data Description

The data description section shows the details of the attributes

ATTRIBUTES	DESCRIPTION	ATTRIBUTES	DESCRIPTION
datetime	Order time	milk_tea	Tea
day_of_week	Day of the week	gateau_chocolate	Piece of chocolate cake
total_amount	Total amount	pandoro	Italian sweet bread
place	Customer's place	cheese_cake	Cheese cake
angbutter	Pretzel with red beans	lemon_aade	Lemonade
plain_bread	Bread	orange_pound	Orange pound cake
jam	Peach jam	wiener	Sausage bread
mad_garlic	Garlic cuisine	vanilla_latte	Vanilla Coffee
americano	Coffee	berry_aade	Berries soda
tiramisu_croissant	Croissants with tiramisu cream	tiramisu	Tiramisu cake
cacao_deep	Croissant with Valrhona chocolate	meringue_cookies	Cookies
pain_au_chocolate	Pain au chocolate	caffé_latte	Coffee with milk
almond_croissant	Croissant with almond cream	croque_monsieur	Hot sandwich

# Data Dictionary

The data dictionary for the Bakery Sales Dataset

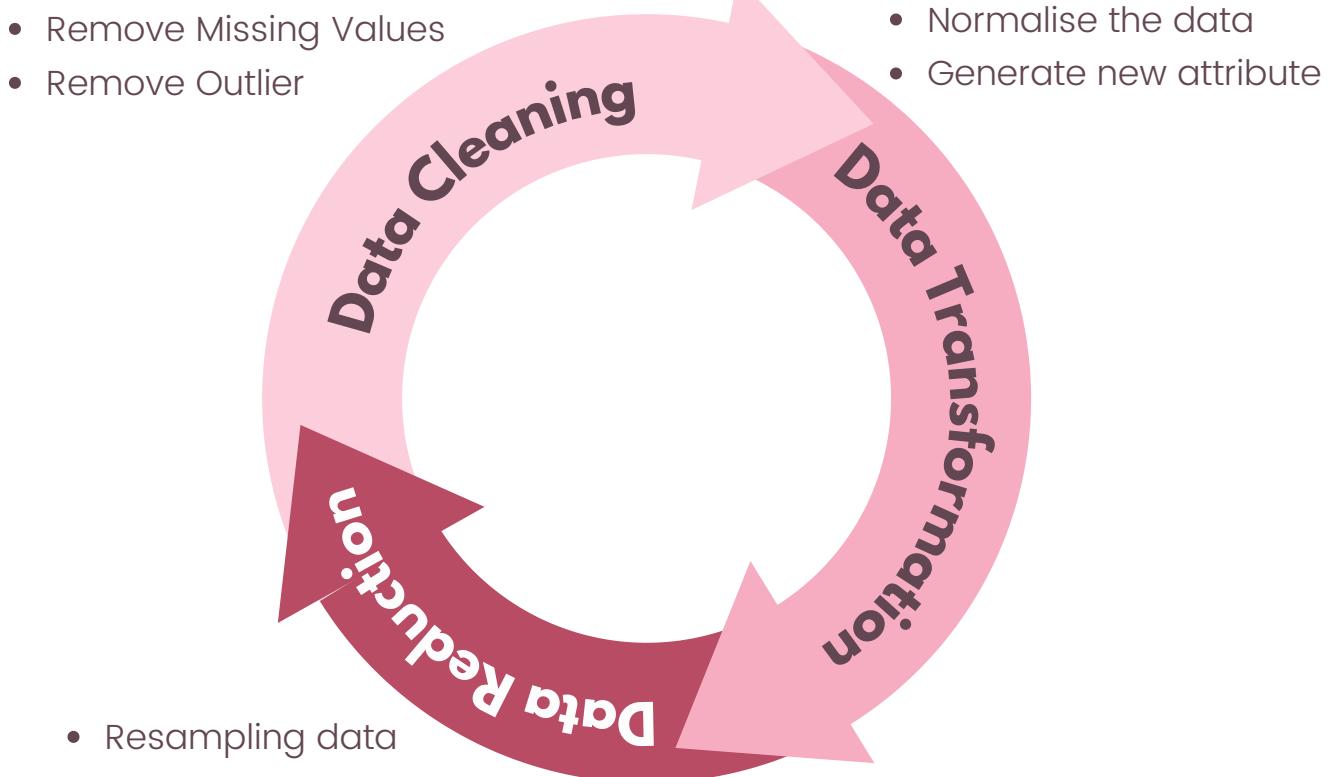
ATTRIBUTES	DATA TYPE	FIELD SIZE	DESCRIPTION
ID	Numeric	5	Unique identification for each order
Order Time	Date & Time	10	Time of order made
Day of Week	String	10	Day of the week
Total Amount	Numeric	8	Total amount
Customer's Location	String	15	Customer's place
Pretzel Filled with Red Beans and Gourmet Butter	Numeric	5	Pretzel filled with red beans and gourmet butter
Plain Bread	Numeric	5	Plain bread
Peach Jam	Numeric	5	Peach jam
Ice Americano	Numeric	5	Americano with ice
Croissant	Numeric	5	Croissant
Ice Cafe Latte	Numeric	5	Caffe latte
Tiramisu Croissant	Numeric	5	Croissant filled with tiramisu cream and fruit

# Data Dictionary

The data dictionary for the Bakery Sales Dataset

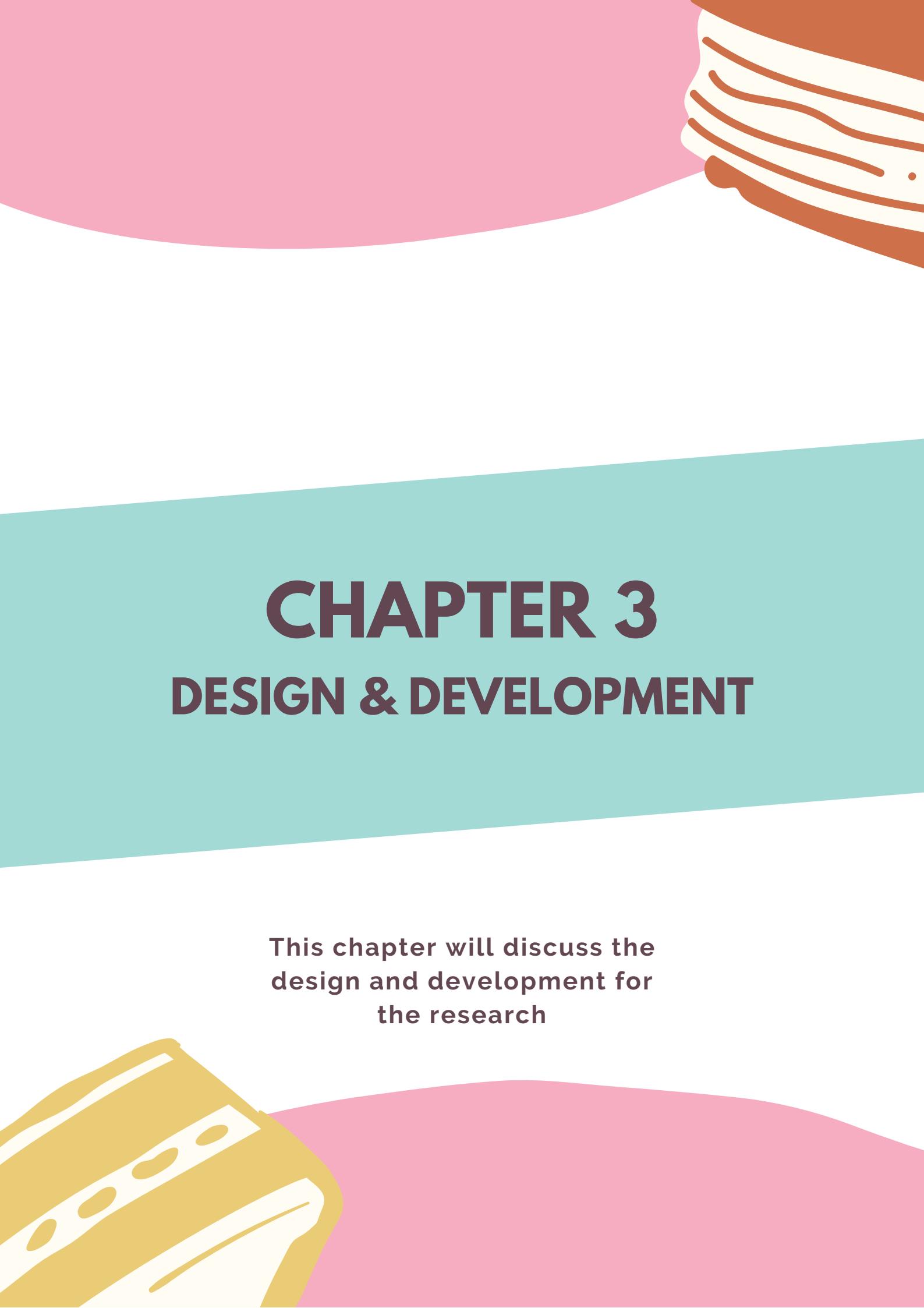
ATTRIBUTES	DATA TYPE	FIELD SIZE	DESCRIPTION
Cacao Deep	Numeric	5	Croissant covered in Valrhona chocolate
Pain Au Chocolat	Numeric	5	Pain au chocolat
Almond Croissant	Numeric	5	Croissant filled with almond cream
Croque Monsieur	Numeric	5	Croque monsieur
Mad Garlic	Numeric	5	Mad garlic
Milk Tea	Numeric	5	Mariage Frères milk tea
Gatuea Au Chocolat	Numeric	5	Piece of chocolate cake
Pandoro	Numeric	5	Italian le pain
Cheese Cake	Numeric	5	Cheese cake
Lemon Ade	Numeric	5	Lemon ade
Orange Pound	Numeric	5	Orange pound cake
Wiener	Numeric	5	Sausage bread
Vanilla Latte	Numeric	5	Vanilla latte brewed with Madagascar vanilla bean

# Data Preparation



The Data Preparation process consists of Data Cleaning, Data Transformation and Data Reduction. The data cleaning phase includes removing missing values and outliers.

In addition, the data transformation process covers normalises the data and generates new attributes such as sales and Pearson Correlation Coefficient values using Power BI Dax



# **CHAPTER 3**

## **DESIGN & DEVELOPMENT**

**This chapter will discuss the  
design and development for  
the research**

# Dashboard Design

The design phase begins with developing prototype dashboard. The prototype dashboard will gives an idea of the overall of the project.

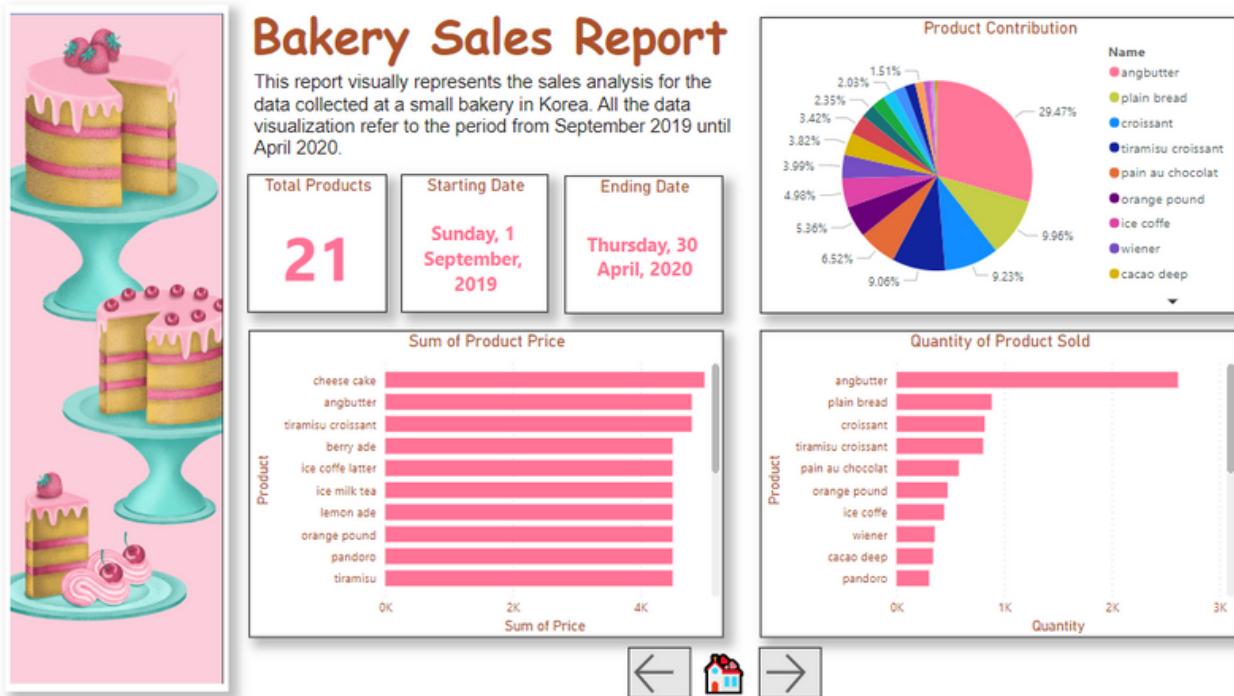


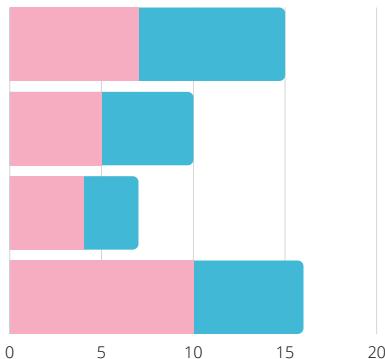
Figure shows the prototype dashboard consisting of several charts such as pie chart, bar chart and card design.

In addition, the dashboard also have navigation button at the bottom of the dashboard. The selection of colours for the dashboard are based on our theme which is Hot Pink and Pastel Pink.



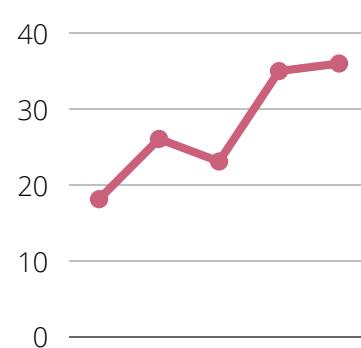
# Chart Type

The selection of chart type are critical to ensure the data can be represented in concise manner and able to tell a comprehensive story telling.



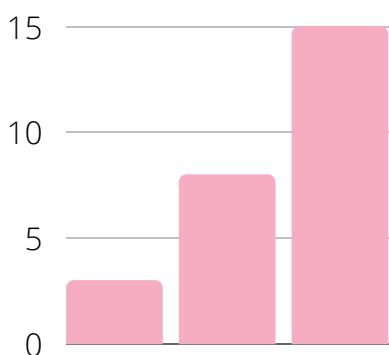
**Stacked Bar Chart**

- Suitable to compared multiple elements in the data
- Easily understandable for storytelling



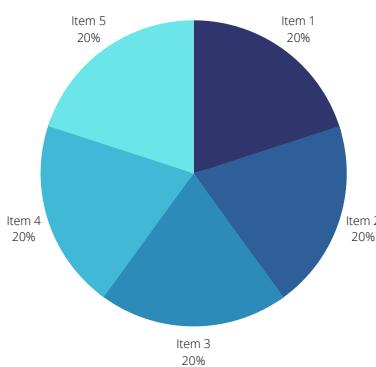
**Line Chart**

- Suitable for continuous data
- Show trends for temporal data



**Bar Chart**

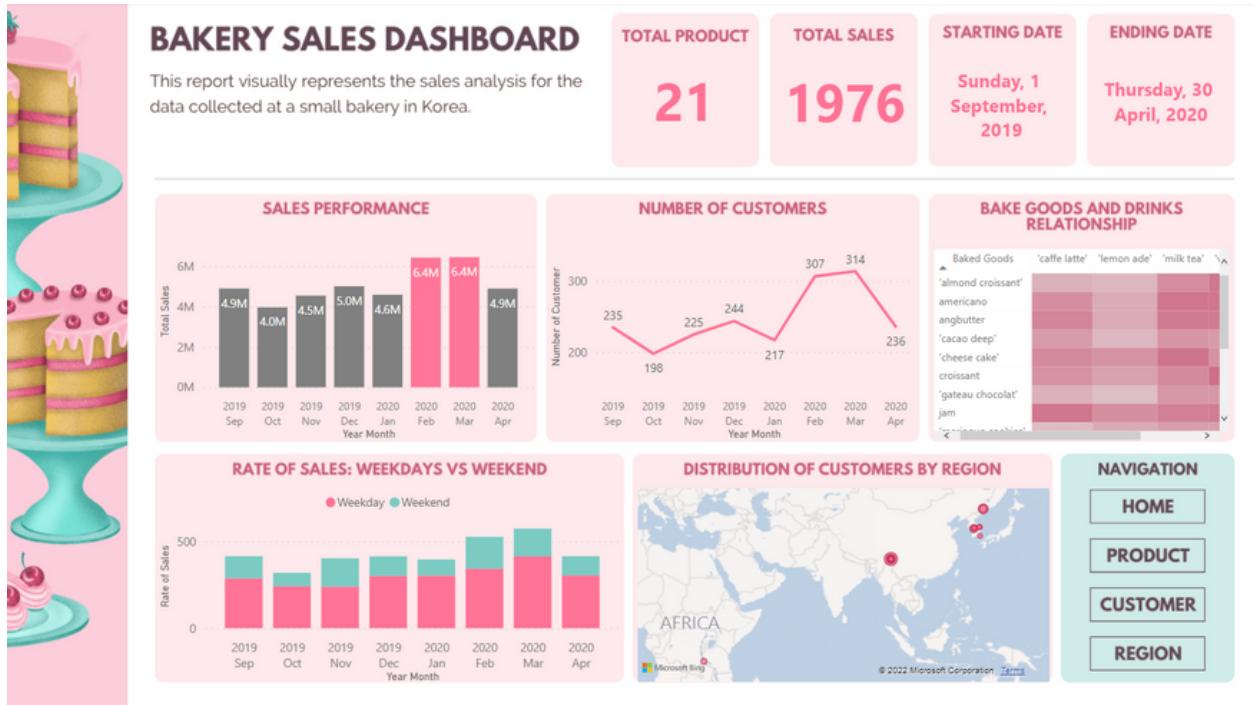
- Suitable for categorical data
- Shows trends based on each type



**Pie Chart**

- Suitable to show overall data
- Each slice can be represented by the category

# Final Dashboard



The figure shows the finalize dashboard consisting of multiple charts such as bar chart, stack bar chart, line chart and matrix table.

The arrangement of the dashboard is based on the research questions stated in Chapter 2. The dashboard also have navigation button for the user to view in details based on each category such as Product, Customer and Region

The setup of the dashboard are based on International Business Communication Standard, that emphasize on data driven that able to supply user understanding and insights.

The goal of the dashboard is to allow users capture the overall performance of the business and produce a critical decision making based on the information extracted.



# **CHAPTER 4**

## **ANALYSIS & FINDINGS**

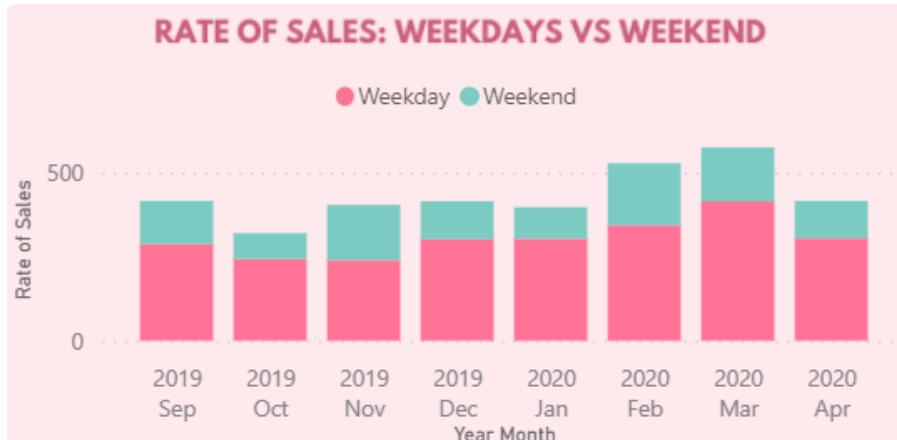


**This chapter will elaborate  
the analysis and findings  
obtained from the research**

# Analysis and Findings



**What is the difference between orders during weekend and weekdays?**



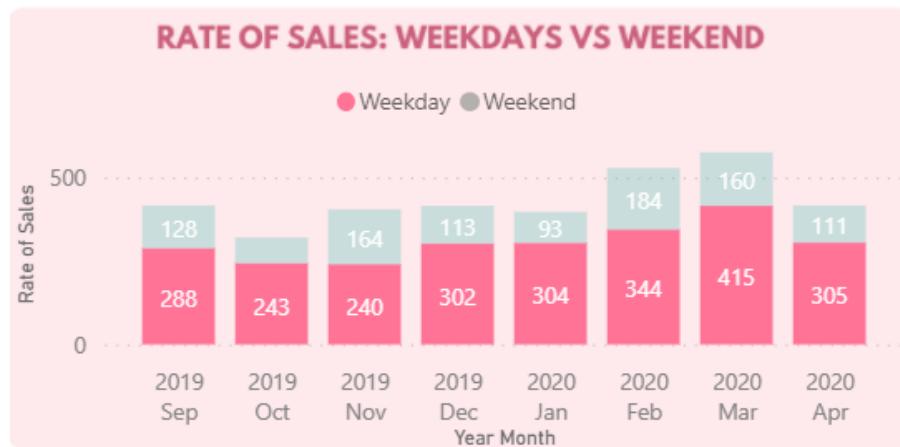
Stacked Area Chart: Rate of Sales between Weekend and Weekdays

The chart above shows the rate of sales between weekend and weekdays for each month from September 2019 to April 2020. The pink bar represents the total sales on weekdays and the turquoise bar indicate the total sales on weekend.

Based on the chart, the amount of sales on Weekdays are higher than sales on Weekend. The highest sales on weekday is 415 and the lowers sales on weekday 240. Besides that, the highest sales on weekend is 184 and the lowest sales on weekend is 77. The trends of sales on weekdays increases from October 2019 to March 2020, whereas the trends of sales on weekends fluctuate for each of the months. The range of sales on weekdays is 175 while it is 107 on the weekend.

The margin of sales between weekend and weekdays are 231. Therefore, it is shown that total sales on weekdays are triple the number of sales made on weekend.

# Analysis and Findings



Stacked Area Chart: Overview of The Sales Rate On Weekdays

The hypothesis can be made from the chart is most of the customers are people that are working near to the bakery. Therefore, the customers may have their lunch or enjoying tea time with the client at the bakery.

On the other hand, the sales on weekend may indicate orders from celebration such as birthday parties or events. Thus, the volume of sales are lower than the sales made in the weekdays.

The spike of orders in March 2020 can be hypothesized due to Jeju Fire Festival, to welcome a new year for harvesting calendar. As people gather to enjoy mouth-watering savoury during the festival.

Based on the hypothesis, the bakery can boost the production of breads and cakes on weekdays as the rate of sales are higher from Monday to Friday. The workers nearby to the bakery may enjoy breads sold at the bakery during lunch hours such as angbutter, plain bread and croissant because to the size and the carbohydrate content of breads that suitable to people on diet.

# Analysis and Findings



**What is the total number of customers per month?**



Line Chart : Number of Customer Per Month

From the line graph, we can see the trends in the number of customers that buy the products for the month starting from September 2019 until April 2020.

From the graph, we can interpret that in March 2020, the highest number of customers achieve which is a total of 314 customers. The graph trends look dramatically fall which loss around 78 customers from March 2020 and only obtain 236 customers in April 2020. The significantly drop really affected the sales performance of the bakery. It seems affected with the Movement Control Order(MCO) which announce on the 18 March 2020 because of the spreading of the COVID-19 Virus. From that scenario , we can see not only this bakery company affected , but all the economic around the world influenced by the pandemic.

In addition, between January 2020 and February 2020, the number of customers increase significantly from 217 to 307, which increased by around 90 customers and there is a great sales corresponding to the increasing trends.

# Analysis and Findings



Line Chart: Increase of Customers from January to February 2020

Other than that, there is a slightly increased number of customers from October 2019 until December 2019 which is on October 2019 , there is 198 customers, in November 2019 there is 225 customers and in December 2019, a total of 224 customers. So the number of customers started to grow gently from October to December 2019.

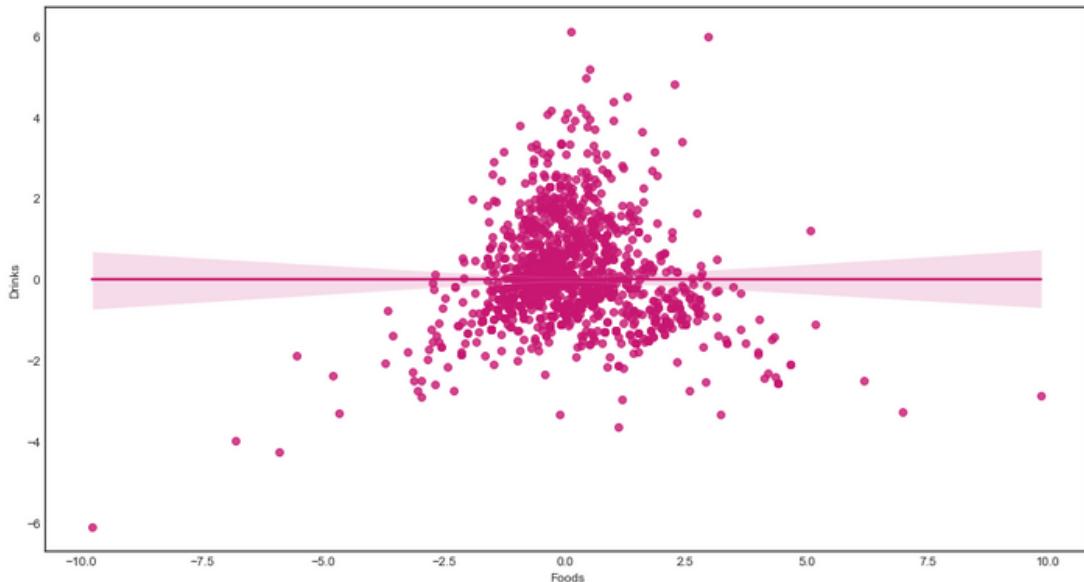
In addition, in the starting two months of the operations, it seems like the number of customers drops negligibly from 235 to 198 which we can see that the bakery still in the phase of planning the best strategy to make their sales performance growth better.

To sum up, we can conclude that the number of customer trends achieves by each month from September 2019 to April 2020 does not consistent which will push some of the bakeries to find the best solution to increase their customers every month.

# Analysis and Findings



**Is there a relationship between the drinks and the baked goods?**



Scatter plot: Drinks vs Foods

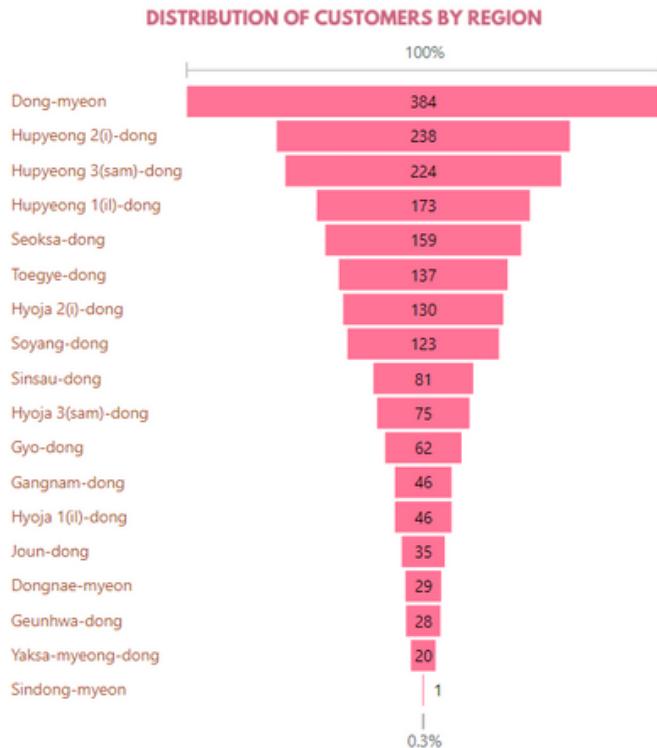
Another finding that have been discovered is the relationship between the drinks and the baked goods, or lack thereof. The scatter plot shows the plotting of the variables of drinks against baked goods. A scatter plot is used because scatter plots are great for visualizing the correlation between two quantitative variables.

By looking at the clusters of the data points, it is hard to observe a definite trend as the data points are not consistently close to the line of best fit. Concerning the line of best fit, the line here is almost perfectly horizontal. This can be interpreted as the two variables having no correlation between each other whatsoever.

# Analysis and Findings



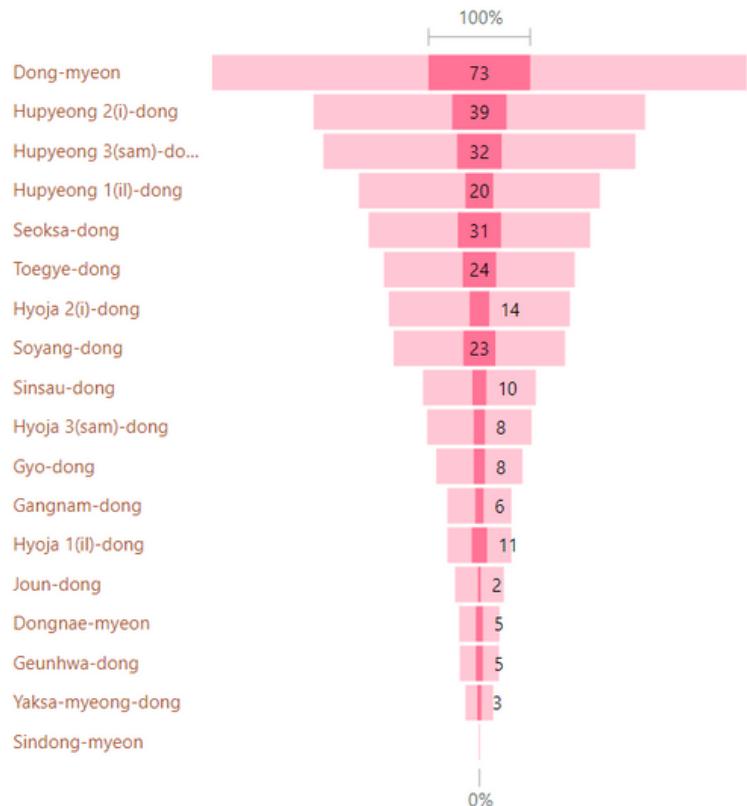
## What is the distribution of orders based on region?



Funnel Chart: Distribution of Customers By Region

The distribution of customers can be seen in the above display in a funnel visualization, and geographical bubble map through a region segmentation. Funnel visualization display the highest number of customers on a region at the top and decreasing until it reaches the smallest number of customers of a region at the bottom, forming a funnel. that displays the total number of customers from September 2019 to April 2020. There are a total of 18 regions which orders have been delivered to.

# Analysis and Findings



Funnel Chart: Distribution of Customers By Region in March 2020

Dong-myeon has the highest out of all regions with a total of 384 number of customers. Contributing the most for every month except for October 2019 and January 2020. In contrast, Sindong-myeon has only 1 customer through 8 months of business. Meanwhile, Hupyeong 2(i)-dong and Hupyeong 3(sam)-dong are relatively close to one another with only 14 number of customers as a difference.

It is also seen that majority of the region had a drastic increase from January 2020 to February 2020 , a total of 217 to 307 customers, a difference of 90. Gyo-dong was one of region that suddenly had an increase of 13 customers while others such as Dong-myeon almost doubles from 35 to 62 and Hupyeong 2(i)-dong more than double the number of customers from 21 to 51. A trend shows a steady increase in number of customers from 307 to 314, while Dong-myeon still being the biggest contributor. However, the trend plummets to 236 in April 2020 because they lost some of their biggest contributor which are Hupyeong 2(i)-dong, Hupyeong 1(il)-dong, Hupyeong 3(sam)-dong.

# Analysis and Findings



Funnel Chart: Distribution of Customers By Region in April 2020

Hupyeong 2(i)-dong and Hupyeong 3(sam)-dong are relatively close to one another in terms of total number of customer indicates that they have a similar delivery price that might be expensive compared to dong-myeon which causes people to have second thoughts about getting their delivery. Some hypotheses can be inferred as the number of customer keeps decreasing for each region. Which means that dong-myeon have the cheapest delivery price rate out of all locations.

However, since this a local bakery, this might not be the case, as hupyeong i-dong and hupyeong sam-dong have similar distance with seoksa-dong and are relatively close to each other, around 7 minute travel time estimation by transit, while dong-myeon and hupyeong has 38 minute transit , on other hand, place with lower customers such as Gangnam-dong has similar transit time about 40 minutes, but yet has lower number of customers compared to Dong-myeon.

This could also be due to lack of advertising in the area, word of mouth, or advertising in mobile application Bae Mi.

The dramatic rise in January 2020 to February 2020 might have been the cause of food tasting better, or promotions relating to special events in South Korea. The lost their contributor are relatively close to each other in term of distance are relatively close to each other, hence there might have been a new competitor new this area.

# **Analysis and Findings**

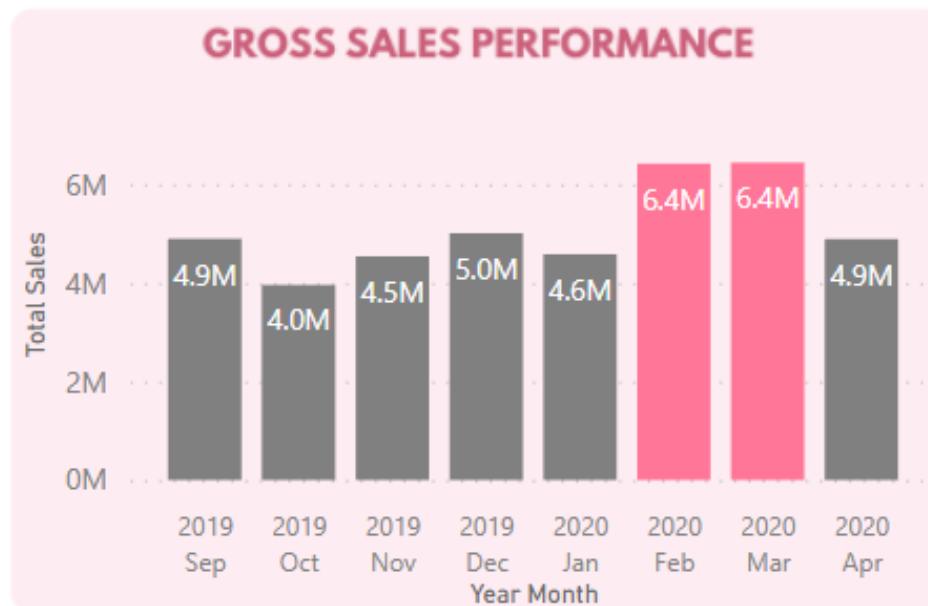
Continuing the report from previous page, it is still unclear if those competitor exist or not can maintain their reputation in the first month, an increase might be expected especially from this 3 area if this bakery does a special promotion to attract customers.

More promotions regarding delivery fee should be considered to expand the business regions and for people who live in Dong-myeon should be given an exclusive offer since it is the biggest asset that contributes towards the business. An increase in number of customers from all majority of the region should be expected

# Analysis and Findings



What is the total gross sales per month?



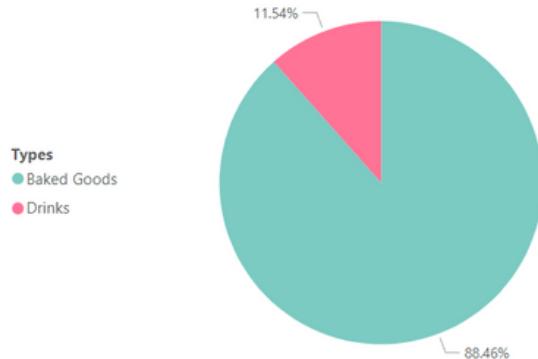
Bar Chart: Monthly Gross Sales Performance

The bar graph above shows the total gross sales per month starting from September 2019 to April 2020

February 2020 and March 2020 have the highest total sales per month, more than 6 million south Korean won, equivalent to RM20000, While October have the lowest below 4 million south Korean won, equivalent to RM14000. The bakery make about an average of 4 million South Korean won per month.

# Analysis and Findings

SALES CONTRIBUTION BY PRODUCTS



Pie Chart: Sales Contribution by Products

Let's talk about what product contribute to these sales. Based on the pie chart, It can be seen that, baked goods contribute an insanely huge portion of 88.46% by selling 7841 quantities of goods that is coloured green, while drinks only contribute 11.54%, having sold a total quantities of 1023 drink which is coloured pink.

For baked goods, Angbutter sold a total of 2.6 thousand and contributed 12.5 million south Korean won in total, this is due to the food being a popular Korean delicacy. Similarly to other bread products as angbutter, such as tiramisu croissant, plain bread and croissant are the top contributors. While pandoro and almond croissant falls behind. Despite tiramisu croissant falls behind on the fourth place as number of product sold, it contributes the second highest, and despite that tiramisu croissant is in the top sellers, tiramisu, which is a cake, was not sold at all in the span of 8 months.

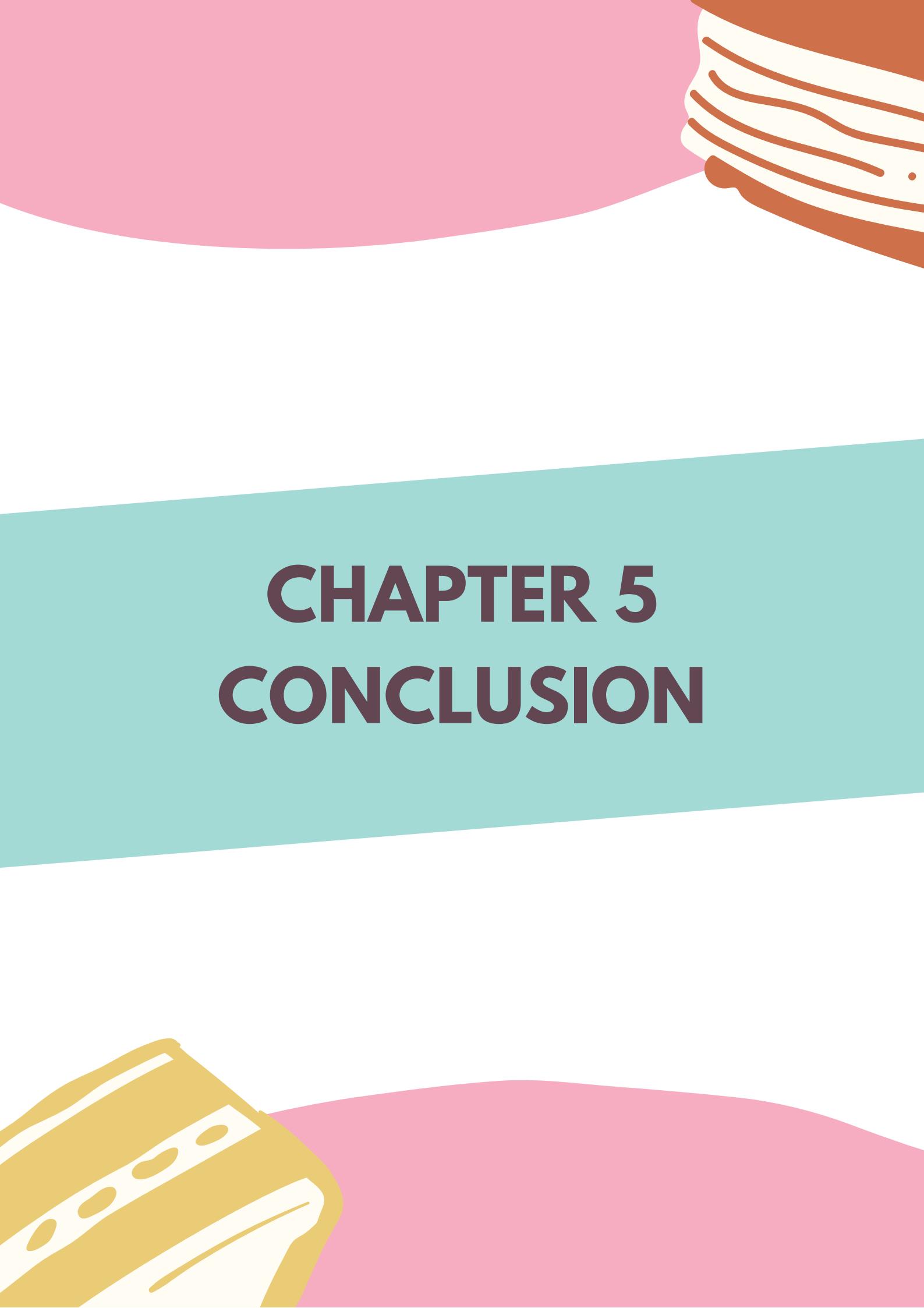
# Analysis and Findings

Continuing the report, as for drinks, coffees such as ice coffee, vanilla latte and ice coffee latter contribute 80.25% in number of orders compared to other type of drink which are tea and beverages. ice coffee contributed 441 number of orders and contribute to 1.76 million south Korean won in sales, while other coffees vanilla latte and ice coffee latte contributed 9 million and 8.1 million in sales respectively. Beverages are sold the lowest coming at 40 and 28 for berry ade and lemonade respectively.

Angbutter is a very popular dish among Koreans, hence why it has contributed so much to the monthly sales. Meanwhile, lemon ade, and berry ade that did not contribute too much could be because these beverages are popular menu during summer which ends in September, hence why it is not sold a lot for the rest of the time frame which starts in Sempember 2019.

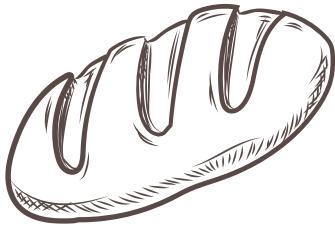
Angbutter has a big potential to attract more customer. However it overshadows other menus, therefore, a promotion around angbutter and another combination of food should be made as a combo menu that acts as a way to encourage customer to try other menu.

In conclusion, the bakery will be able to stay on top of the market, if the bakery focuses on angbutter since that is the most popular item they have ever sold.



# **CHAPTER 5**

# **CONCLUSION**



## CONCLUSION

The purpose of this study is to develop a dashboard that allows the exploration of data analytics and strategic business planning for small businesses.

So developing a visualization dashboard, really assist the business owner to identify the main factor that affects the demand for bakery products and making sure that the small business can boost their sales performance from the analysis.

Moreover, it also gives some advantages to the small business owner which they can gain insight into the business performance in real-time. Hence, we can say that, with the benefit of data visualization, all the objective from this project was successfully achieved.



# REFERENCES

- Choudhury, A. (2022). *What Are Data Types and Why Are They Important?*  
<https://amplitude.com/blog/data-types>
- Holtz, Y. (2018). *Network chart with R and igraph from any type of input – the R Graph Gallery.*  
<https://r-graph-gallery.com/257-input-formats-for-network-charts.html>
- Monday.com. (2021). *What's a timeline chart? | monday.com Blog.*  
<https://monday.com/blog/remote-work/timeline-chart/>
- Munzner, T. (2015). *Visualization analysis & design* (E. Maguire, Ed.; 1st edition) [Book]. CRC Press.
- Peters, K. (2021). *Line Chart Definition.* <https://www.investopedia.com/terms/l/linechart.asp>

# KAGGLE LINK



## Bakery Sales

Our Bakery Sales(Delivery)

 [kaggle.com](https://www.kaggle.com)

[https://www.kaggle.com/datasets/hosubjeong/bakery-sales?  
select=Bakery+Sales.csv](https://www.kaggle.com/datasets/hosubjeong/bakery-sales?select=Bakery+Sales.csv)

# POWER BI LINK

<https://bit.ly/probake-dashboard>

# **THANK YOU**

<https://bit.ly/probake-dashboard>