Cardiff
Metropolitan
University

Assignment

CIS7026 - Business Process and Data Analysis

BY

Mohammad Aziz
(St20192233)

Table of content

Part - 1

Task 1 - Introduction

Greggs plc is a bakery food based british company, headquarter located in Newcastle upon tyne, England, United Kingdom (wikipedia, n.d.). It is mainly known for its special fresh products like bakes, sandwiches, doughnuts, sausage rolls and vanilla slices. According to the Food-To-Go(FTG) market it has kept 10.7% shares of occasions and that makes Greggs a dominating company over others (kamcity, 2022). John Robson Greggs was a British businessman who founded this company in 1939 since then there are over 2000 Greggs, nine regional bakeries such as Scotch Pie and around 25,200 employees working all over the UK. It sells unsold stock, rejected products by factory through the small number of outlet shops which are in the wales, northern and central England (wikipedia, n.d.).

There are several competitors in the field of bakery for Greggs including A Preta, Finsbury, BIMBO group, Subway, Costa, Starbucks, Cafe nero, and BP. Ups and downs have come in terms of revenue, supply chain and staff during the omicron variant. Business of this firm rose 0.8% in the last variant of coronavirus compared to 2019. The price of products became high due to a hike in ingredients and wages of staff (Neville, 2022). In 2021 the revenue generated was £811.3 million however, £1.23 Billion revenue was made in 2022(first quarterly basis) by Greggs Plc and this result shows that this firm has come out of the pandemic. Gregg has started recovering economically and set its plans to become a multi channel business.

The primary aim of this report is to determine, analyze and comment on the shortage of the Greggs's businesses during omicron and justify and evaluate using methods and technical tools. There are five types of activities going to be discussed in this report. Those are strategy analysis, investigating the situation, considering perspective, needs and requirement analysis. Each analysis is done in a separate section using several tools and techniques. The analysis of this report helps Greggs's businesses to overcome issues and challenges.

Task 2 - Justification of Tools and Techniques

To perform strategy analysis on Greggs Plc, The model of Michael Porter that is Porter's Five force model is being used rather than using Pastel analysis because it is popularly used in the private sector firms to analyze effective external environments (Vining, 2011). It helps firms in their strength, weaknesses and position of the firm in the competitive industries. Moreover, It helps in their business profit, finding the nature of the competitors and analyzing threats to customers and suppliers. However, Pestel analysis is based on external factors such as political, economical, social, technological, environmental and legal. It helps firms to make decisions for their businesses on the basis of external factors. Resource audits help Greggs to understand the strength and weaknesses of the resources.

Tools such as rich pictures and data flow diagrams are used to investigate the situation. Rich picture is the summary of the business situation. It contains icons, text, arrows, pictures and symbols and they are altogether connected with the arrows and lines used to graphically show the Greggs business situation. It shows the relationship with the stakeholders, competitors, different departments, customers, staff etc. Dfd is the graphical representation of information flow through a system without decision in a business information system because it shows how businesses of Greggs data move sequentially. UML diagrams show the relationship and interaction between the actors and the objects. It helps Greggs to organize its documentation and workplace.
CATWOE is used to analyze the business perspective and understand various stakeholders perspectives. This analysis is used in this report to examine and solve the business issues involved with multiple stakeholders. Power interest grid is a matrix tool used to manage and categorize stakeholders and influence them towards selecting the firm.
UML use case diagram is a tool to interact between system and customers and also to show the top level of the business system.
UML diagrams are the visual representation. It allows staff or team members to see how business is or will work graphically.
These above mentioned tools will definitely help the Greggs Plc to analyze and evaluate the issues and challenges easily.

Task 3 - Analysis

In this segment, There are several tools and techniques used to get insight into the business of Greggs Plc.

3.1 Greggs's Strategy Analysis: Porter's five model

It is the process of designing and formulating strategic plans to make appropriate decisions for businesses in order for a firm to work smoothly. It needs various deep research to deal with several elements in this report. It helps to set goals and objectives of a particular firm. This analysis helps to find out what area of the firm requires improvement.
The selected firm for strategy analysis is Greggs Plc and tools that are used is the Porter's five force model given in the figure 1 where 5 competitive points examine the firm's strength and weaknesses and tries to keep its stand in the competitive industries.

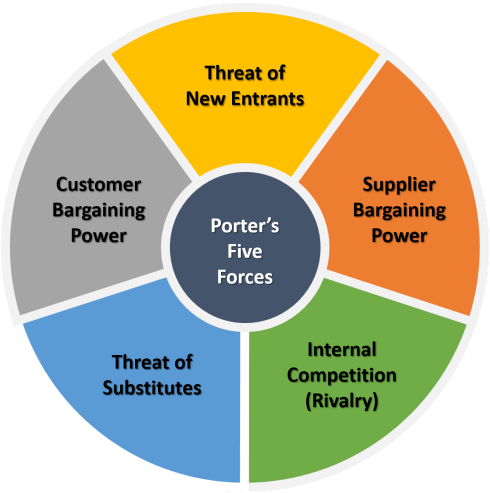

Figure 1 Michael porter's five model

Porter's five force model is not just for analyzing contemporaneous competition in the industries but also to make strategic decisions during pandemic (omicron). This model concentrates on how Greggs Plc constructed its competitive name in the field of retail industry. It is also used by managers of Greggs Plc to explore profit in the customer service sector.

3.1.1 Threat of New Entrants

Threats or barriers for new entrants in bakery industries are comparatively lower than other industries. New entrants creates opportunities, challenges, innovation, new ideas, and pressures to reducing Greggs's products prices or upgrade in products and services and providing high quality products to customers. Greggs is a well established firm, strong economics and very hard

to achieve for any new entrants. Moreover, there are barriers for the new entrants like capital requirements, strict rules of license and strict COVID-19 regulation.

3.1.2 Bargaining power of suppliers

Greggs is strong in terms of suppliers and supplies high quality products and services to its customers and always makes customers satisfied. It has its own way to keep this efficient because there are various suppliers run under Greggs who always keep their eyes to keep improving its supply chain and its efficiency. Recently, Greggs has decreased its supplier during the pandemic(COVID-19) and started focusing on developing more shops(Hall, 2020). Therefore, bargaining power for suppliers is low shown in the table 1.

Threat of New Entrants	Low	Proper distribution channels and Economies of scale	Strict licensing regulations and capital Requirements
Bargaining Power Of Suppliers	Low	A strong supply Chain	Several centres Of excellence
Bargaining Power Of Buyers	Moderate	Sensitivity of Consumers to Prices	Economic volatility
Threat of Substitute	High	Globalisation	Great product Differentiation In the market
Existing Rivalry	High	In-store bakeries And craft bakeries	Unique approaches to Business

Table 1 Michael five porter's model

3.1.3 Bargaining power of Buyers

Bargaining power of buyers is moderate as Greggs is an established, reputed and big firm serving millions of customers a week and its main aim is to provide high quality products to every customer in the UK(Greggs, 2022). Expectations of buyers are more on high quality products with minimum rate. Greggs Suppliers are more than the number of firms that's why there are not many options for buyers. However, the competitors are Finsbury Food group, Grupo Bimbo and Pret a manger who kept eyes on the Greggs strategies, products and prices. Trust, honesty and

good relationship with buyers or customers makes Greggs business grow more and forces customers not go to its competitors.

3.1.4 Threat of Substitute Products or services

There are so many substitute products available in the industries that Greggs operates and that creates a big challenge for Greggs because customers can buy similar products from other retailers too. For example, PoundBakery sells similar pies, pasties and sausage rolls as Greggs. There are japanese and chinese that provide food on go services same as Greggs. Therefore, from above, it can be concluded that the threats of substitute products or services are high for Greggs Plc given in the table 1.

3.1.5 Existing Rivalry

Greggs Plc. is the British leading and one of the largest bakery firms in the United Kingdom. It is a dynamic firm with changing customer demand and taste. This firm has generated revenue of about £1.23 billion (52% increased from last year) in the year 2021 and kept its reputation on the top. However, Total net income is £117 million and Loss was just £13 million(uk.investing, 2022). There are few critical competitors in these industries that always use their business strategies and plans. For instance, Pret a manger provides similar types of product, services, quantity and quality to the customers as Greggs and has started opening its shops in northern England to spread its businesses where Greggs Plc has already established (Hancock, 2021). Other competitors such as Grupo Bimbo and Finsbury food group are also in the queue. From the above paragraph, it can be drawn that Existing rivalry is high and it can be seen in the table 1.

3.1.6 Resource audit

It is the technique of strategic analysis used to understand the resources and competencies of the organization. It is an internal audit. This assists to understand what the organization needs and what things need to improve in order to keep running the businesses. There are mainly 4 types of resource audits: physical resources, human resources, financial resources and intangible (accountlearning, 2020).

3.1.6.1 Physical resources

- Operating 3,000 shops in the UK, 2100 branches alone in Newcastle upon tyne (Barrie, 2021).
- The 30,715 sq ft. headquarter's building with tennis court, football pitch and 500 trees are planted at Q9, Quorum, Newcastle upon tyne (Ford, 2015).
- Bakery equipment such as traywash, traveling ovens, roll plant, rack oven, Pie machine and Doughnut fryer (foodprocessing, 2007).
- EPOS hardware used for touchscreen terminals in over 300 stores (foodprocessing, 2007).

3.1.6.2 Human resources

- 25,000 employees all over the UK (Ford, 2022).
- All employees have at least a high school pass.

3.1.6.3 Financial resources

- £1.23 billion revenue generated by Greggs Plc all over the UK (Holzhäuser, 2022).
- 50% shares of the company controlled by top 17 shareholders (St Wall, 2022).

3.1.6.4 Intangible

- 3rd rank in number of total outlets (statista, 2016).
- Biggest bakery chain in the UK (Croce, 2019).
- The Retail week award(Ford, 2020)

3.2 Investigate Situation

3.2.1 Rich picture diagram of Greggs Plc.

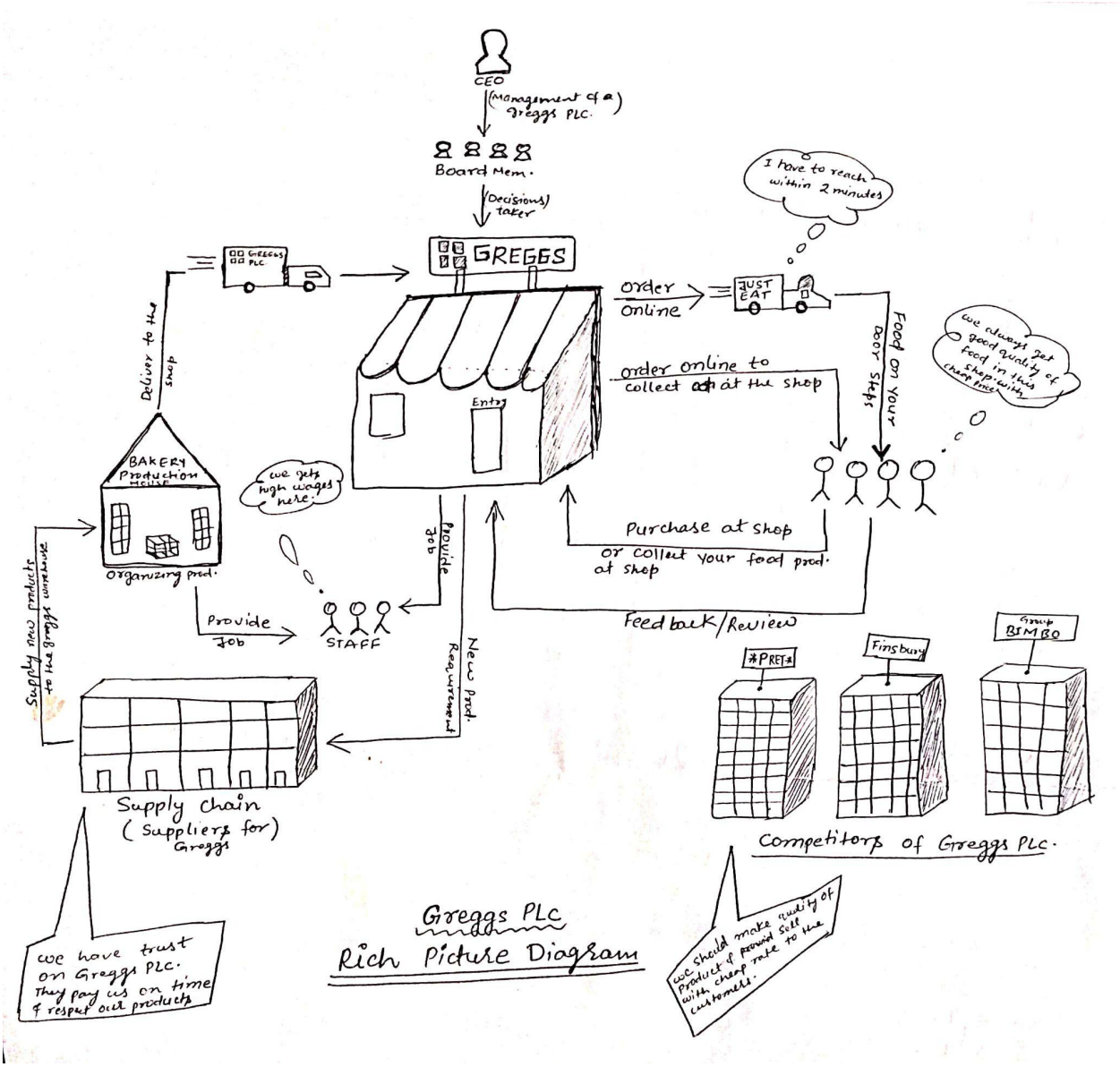

Figure 2 Rich picture diagram of Greggs Plc.

A rich picture diagram is the drawing overview of an organization. It helps to explain about the organization's vision and its business plan easily by illustration. It is used to help the business analyst in analyzing any business plan of an organization. Furthermore, it also helps in communication especially when there is language or cultural difference (Gates, 2016).
From the above rich picture diagram, it can be seen the overall strategies and business plan of Greggs Plc. Roger Whiteside, an chief executive officer has the responsibility to manage all the operations including structure, strategies, profitability, agendas and connections with the main boards. Board members take decisions before passing any project for an organization. There are CEOs, board members, customers, staff, and suppliers acting as stakeholders and they are responsible for the well working of the organization. Here, main competitors of the Greggs Plc

are Pret a manger, Finsbury, Group BIMBO and it is seen in the above figure 2. There is almost no covid nowadays but still Greggs kept rules and regulations from the UK government such as sensitization, mask over the nose and mouth, and keeping distance to retain the customers safe from light variations of COVID (Omicron). Online services like order your food on the door within 2 minutes and order the food & pick order on chosen time from the nearest Greggs store.

3.2.2 Dfd context diagram of Greggs online food ordering system

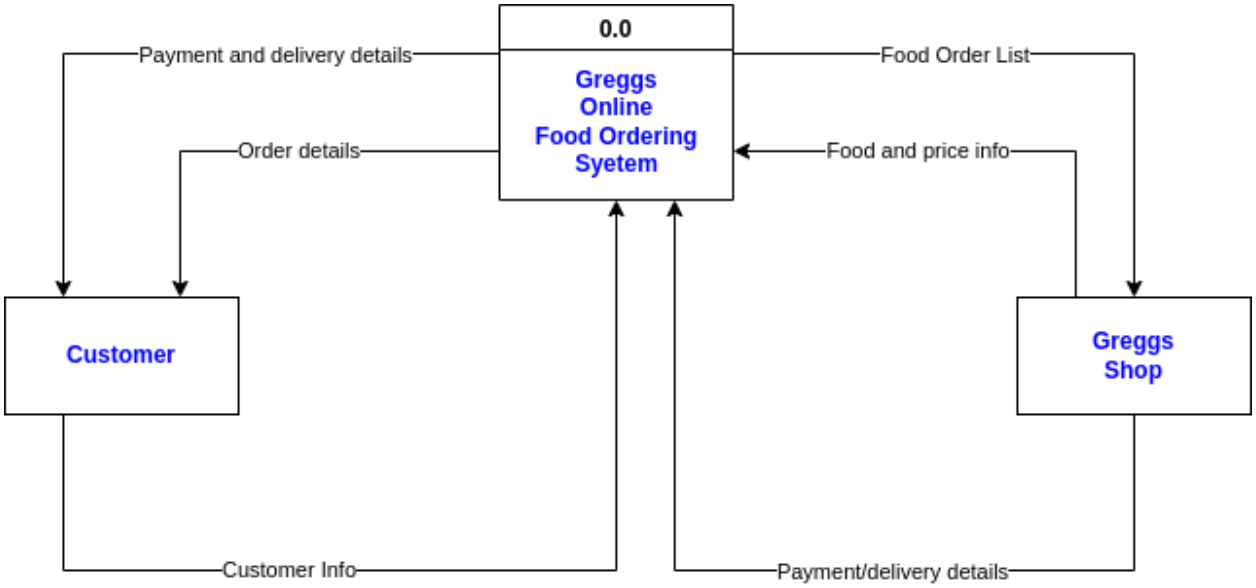

Figure 3 Level 0 Dfd or context diagram of Greggs food ordering system online

Dfd context diagram is also known as data flow diagram or 0 level or top level flow diagram where data or information used to flow in business information systems using various box and arrow diagrams. It is used to understand the process, discover potential problems and improvement in efficiency in the business process (Chi, 2021). The 0 level dfd diagram is also an input output system. Customer's order acts as input however output is food order to kitchen, receipt, management and report to the shop manager mentioned in the figure 3.

3.3 Consider Perspective

3.3.1 Catwoe Analysis of Greggs Plc.

CATWOE is the abbreviation of customers, actors, transformation process, worldview, owners and environmental constraints (pestleanalysis, 2015). It was developed by David Smith in 1975 (Kukhnavets, 2017). CATWOE is the technique in the business to analyze what an organization wants to attain, understanding stakeholder perspective and how many areas that an organization has (Business Change Academy, 2020).

Customer	Customers are the consumers, employees, products and services. These customers will be affected by the changes in Greggs Plc. These changes are mainly handled by the Greggs's staff.
Actors	Actors referred to as the staff members who will be working at the Greggs's shop. Therefore, Customers will be more important than anything else. They have to serve food to the customers. That means staff have responsibility for the reputation of the shop.
Transformation process	Transformation process referred to as board members to provide good and quality of products to the customers and get profit in returns.
World-view	To provide services very quickly, best quality food, fascinate investors, attract more and more customers, spread businesses world wide. Keeping customers satisfied with the quality, quantity with affordable cost is the bigger picture. This needs to be improved. And this will impact on overall Greggs business, revenue and profits.
Owners	Chief executive officer and board members are the owners of Greggs who have responsibility to take decisions and management for the growth in businesses of Greggs.
Environmental Constraints	Environmental constraints in the case of Greggs Plc. are its policy, regulations, privacy laws, laws of the countries and counties such

	as Wales, England, Scotland and Northern Ireland have bit different laws. New employees are provided training for betterment by making sure it will not affect the company's businesses.

Table 2 CATOE analysis of Greggs Plc.

3.3.2 Power/Interest Grid of Greggs Stakeholders

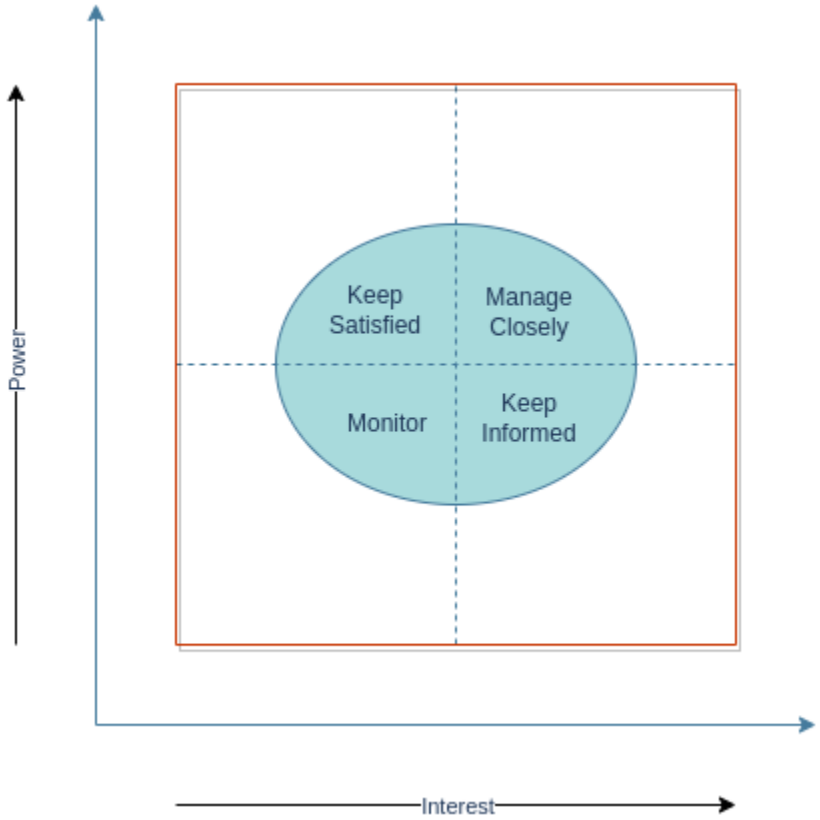

Figure 4 Power/interest grid of Greggs Plc.

The main stakeholders are the Chief executive officer, board members, staff, customers, shareholders and they are categorized accordingly on the basis of power and interest of Greggs Plc. The below are the description of the above power/interest grid image.

Keep Satisfied - High Power and Medium Interest

Stakeholders such as shareholders have high power but medium interest in the Greggs Plc. There are 17 shareholders which have 50% shares of Greggs controlled by them. it means 2.9% have shares for each shareholder (Massachusetts News, 2022). Government has impacted on the

business of Greggs by putting its laws and regulations such as recently, they have put COVID-19 regulations on their import and export of goods and services which had an impact so much on the Greggs Plc. It should be satisfying.

Monitor - Low Power and Low Interest
Suppliers have low power and low interest in the Greggs because recently, Greggs has put a bit more cost on the ingredients and shortage of staff during pandemics (Kollewe, 2021). It should be monitored.

Manage Closely - High Power and High Interest
There are stakeholders like board members which are responsible for decision taking and management and have high power and high interest such as Royal London Asset Management Limited hold the highest percentage around 6.1% shares outstanding followed by Massachusetts Financial Services Company and BlackRock, Inc. which have equal shares of 5.0% (simplywall.st, 2022). They are the most important shareholders for Greggs Plc and must be managed closely in all ways.

Keep Informed - Low Power and High Interest
Customers and employees come under the stakeholders who have low power but have high interest in the Greggs Plc. They should keep informed.

Stakeholders	Power	Interest
Government, shareholders	High	Medium
Suppliers	Low	Low
Board members	High	High
Customers, Employees	Low	High

Table 3 Power/interest grid of Greggs stakeholders

3.4 Analyze Needs:

3.4.1 Decision Tree for food ordering system online

Decision tree for customer

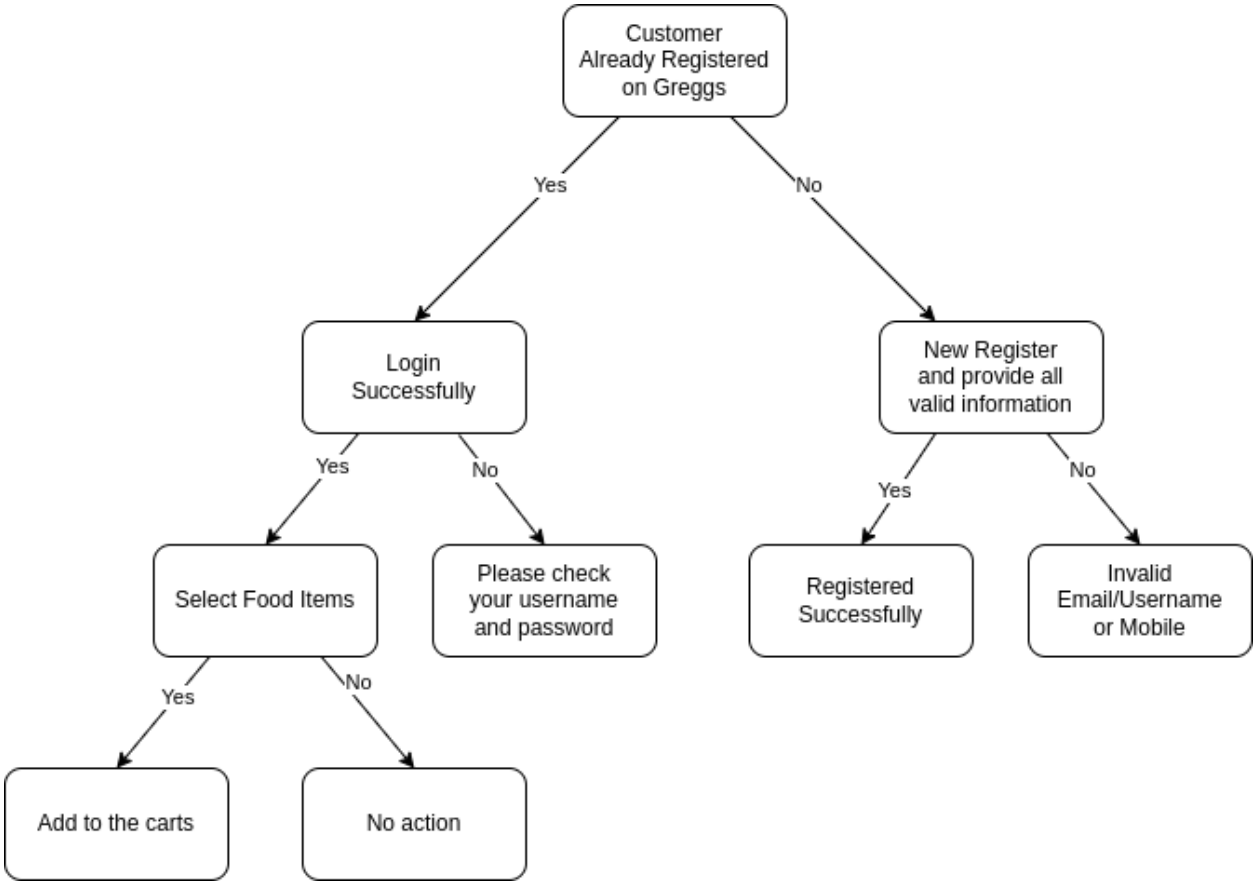

Figure 5 decision tree for customer

Decision Tree for payment

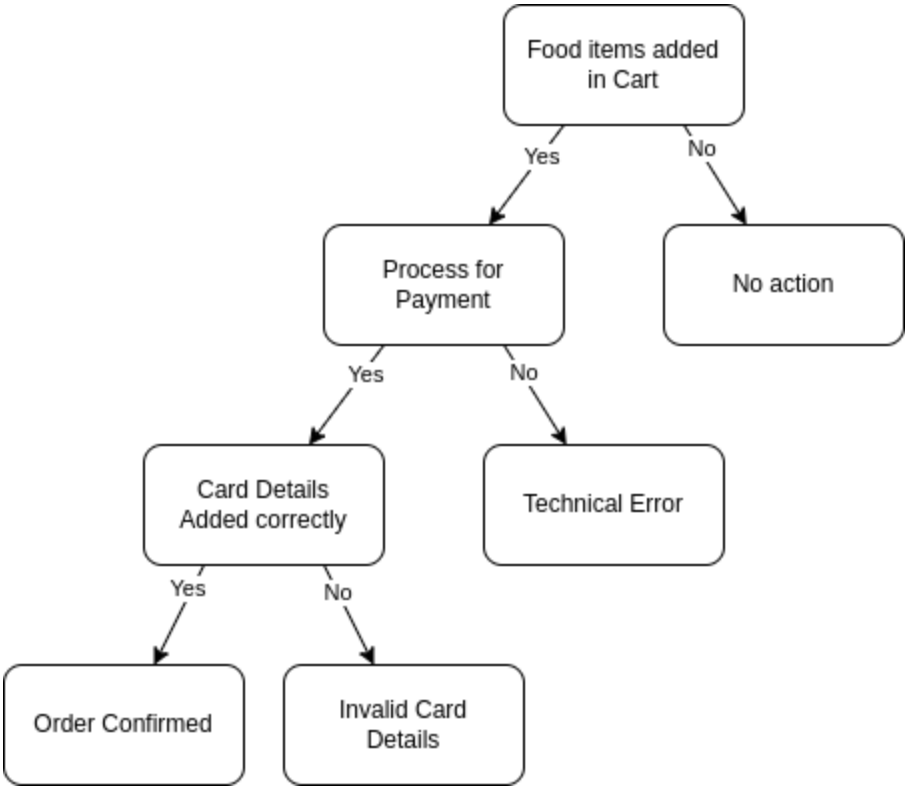

Figure 6 decision tree for payment

3.4.2 UML use case diagram for Greggs food ordering system

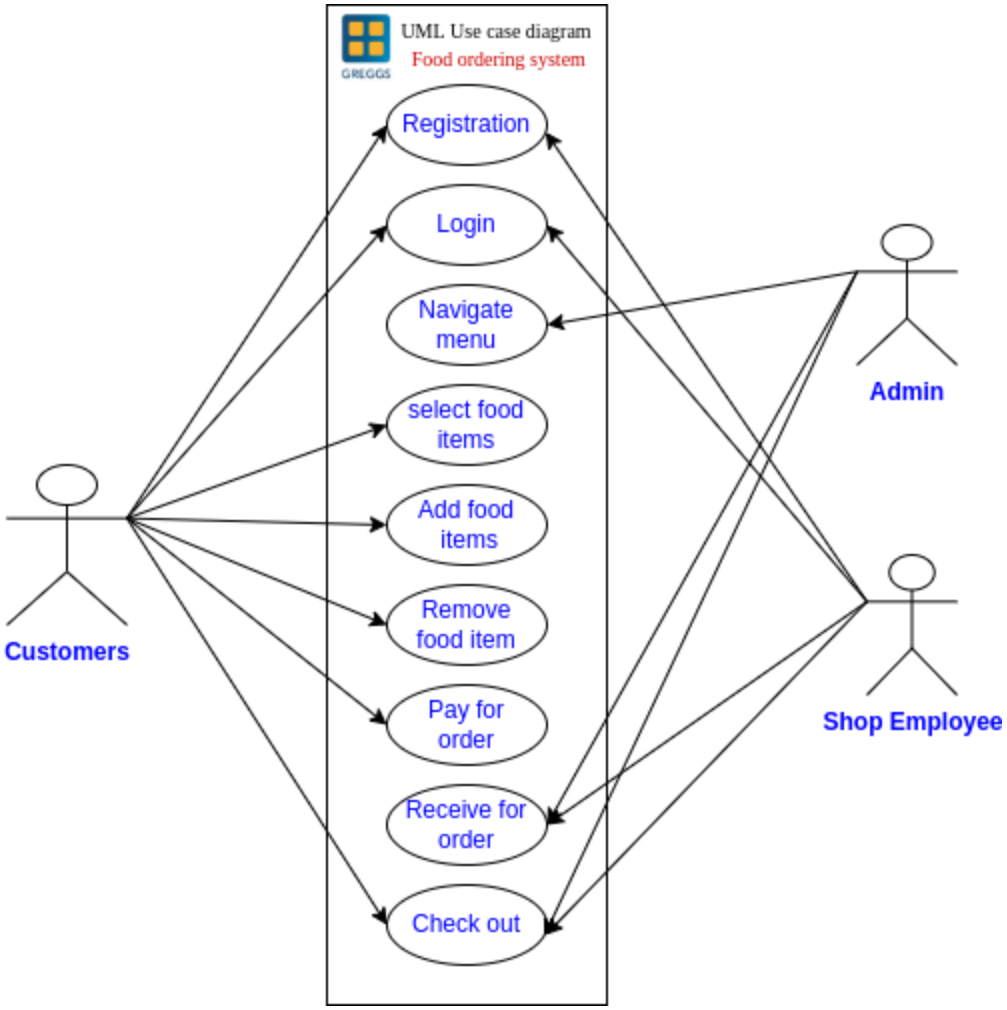

Figure 7 UML use case diagram for Greggs food ordering system online

The above UML use case diagram is drawn for Greggs Plc can be seen in the figure 7 where 9 use case (including registration, login, navigation menu, select food items, add food items, remove food items, pay for order, receive for order, check out) and 3 actors (Customers, admin, and shop employee) are mentioned. All the use cases are covered with a long box boundary also known as system boundary. Use cases are strongly connected with actors by using a solid link. Customer as an actor is connected with every use case except “navigation menu” and “receive for the order” however, admin and shop employees are connected strongly with “registration”, “login”, “navigation menu”, “receive for order” and “check out”. Customers used to order their food from Greggs shop using an online ordering system specially during Covid 19 lockdown to keep themselves safe but nowadays a very light version of coronavirus (omicron) is left around the UK. It has been seen the selling of Greggs food products continuously growing using online delivery and click and collect (Skeldon, 2021). Therefore, Greggs has started focusing on online order and started increasing its shops in around every small cities, towns as well as villages in the UK.

3.5 Define requirements:

3.5.1 Level 0 Dfd diagrams of Greggs Plc.

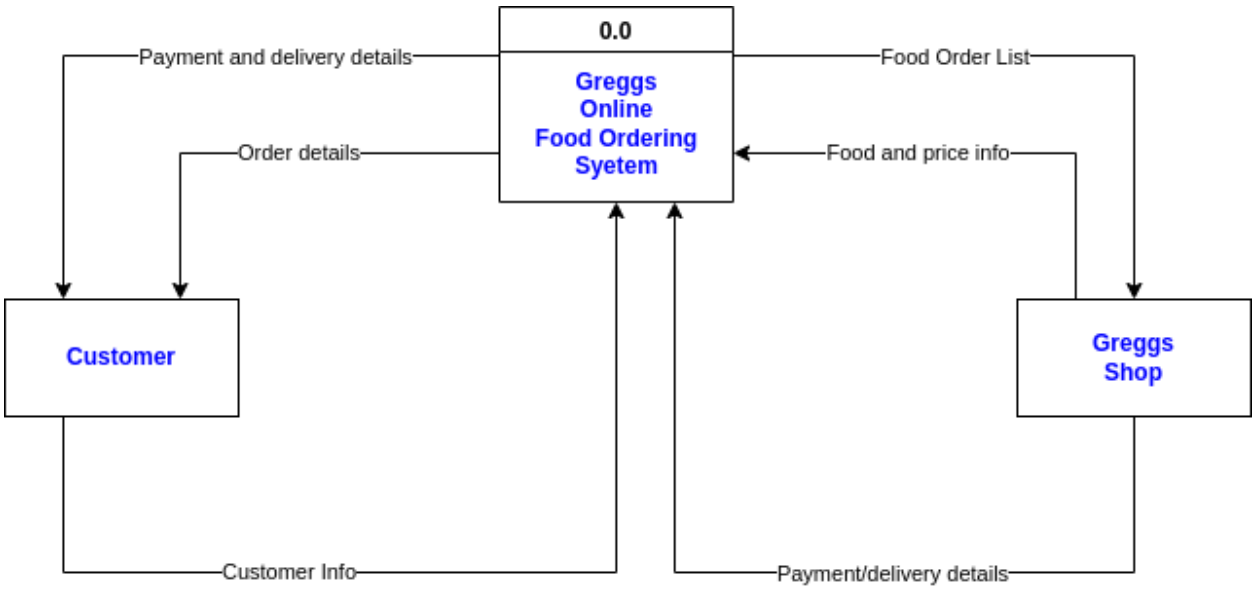

Figure 8 Level 0 dfd diagram of Greggs Plc.

Dfd level 0 is an acronym of data flow diagram level 0 which is also known as context diagram. It is an abstraction view and the system acts as a single process and has a relationship with the external entities (GeeksforGeeks, 2019). It basically shows the whole system as a single bubble where input and output data shows through the solid labeled arrow links. The above level 0 Dfd is shown for the online food order in Greggs Plc in figure 8.

3.5.2 Level 1 Dfd diagrams of Greggs Plc.

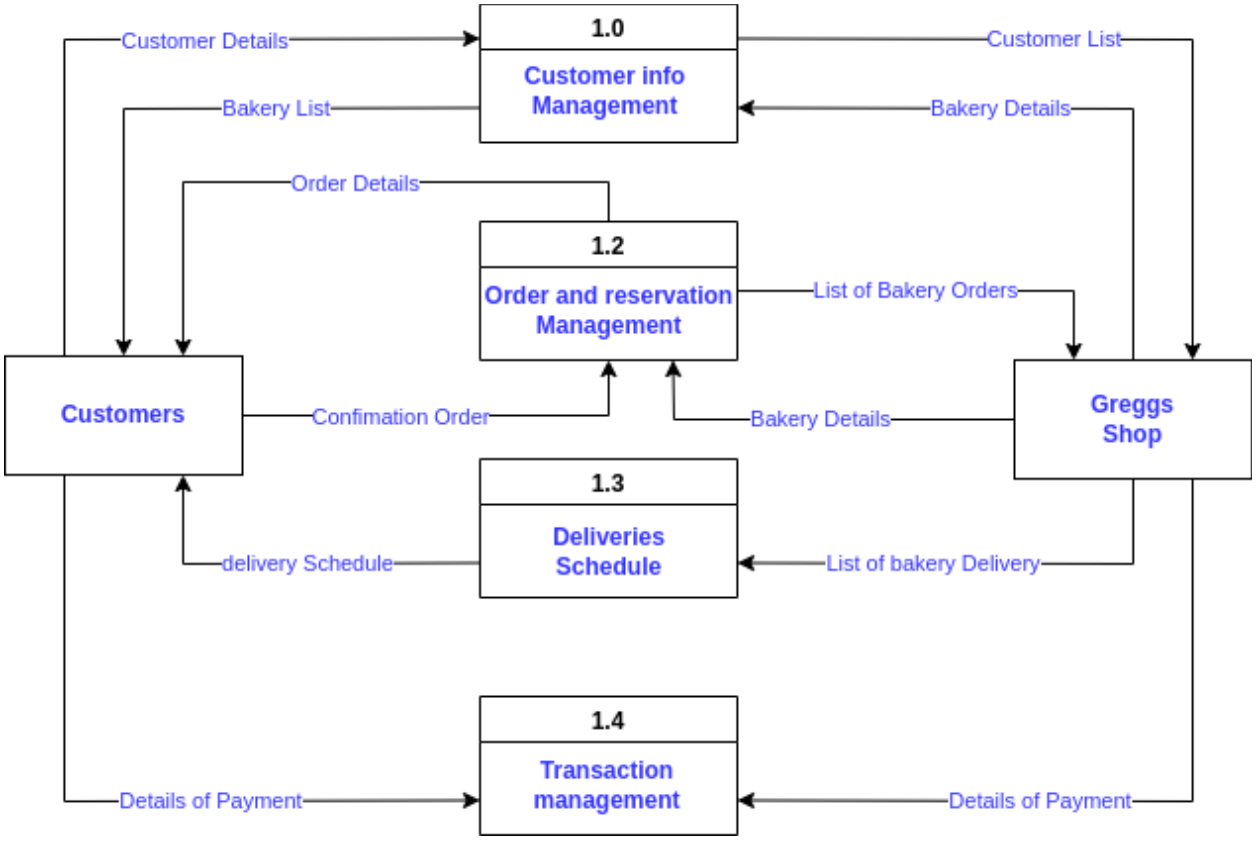

Figure 9 Level 1 dfd diagram for Greggs Plc

Level 1 Dfd is the decomposition of Level 0 Dfd or context diagram. So basically Level 1 Dfd is broken down into multiple bubbles and this provides more depth details about the system and processes. Level 0 Dfd process is broken down into subprocesses and its system breakdown into subsystems (GeeksforGeeks, 2019). The above Dfd level 1 is drawn for the Greggs Plc online food ordering system and can be seen in the figure 9.

3.5.3 UML Class diagram

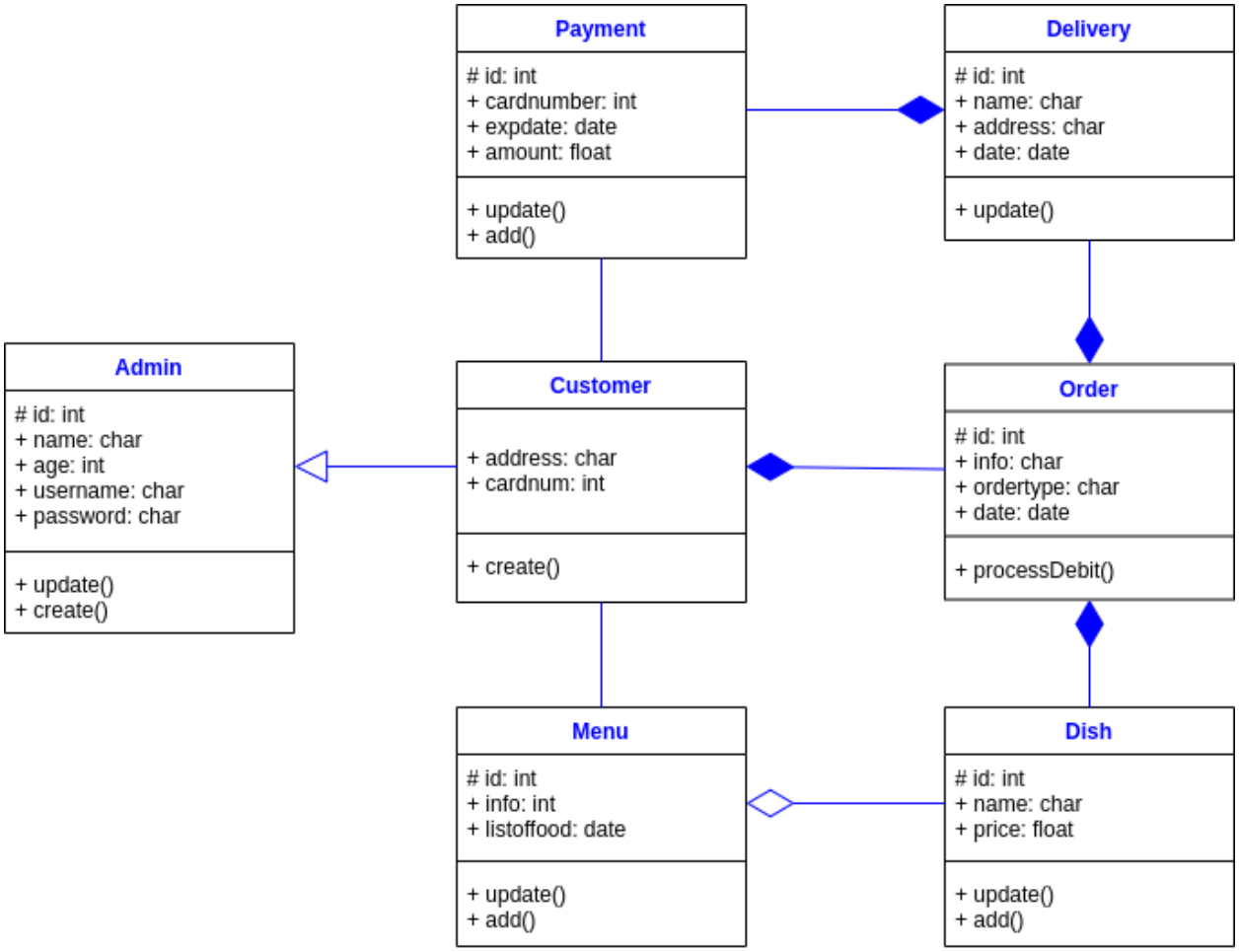

Figure 10 UML class diagram for Greggs Plc

The above UML class diagram of Greggs online ordering system shows how the data and information is handled in the structure. It gives an overview of a software system (Walker, 2019). The class has information about data and every class has their own attributes such as customer class has address and cardnum as attributes. And each class has methods as well such as menu class that have update() and add() methods shown in the figure 10. The structure of the class has a box containing 3 sections where the first section is the name of the class, middle section is the attributes and third section is the methods. Link represents the relationship between the classes. The above class diagram has classes named payment, delivery, customer, order, menu, dish and admin. And they are connected with each other by a solid line with some relation.

3.5.4 UML Sequence diagram

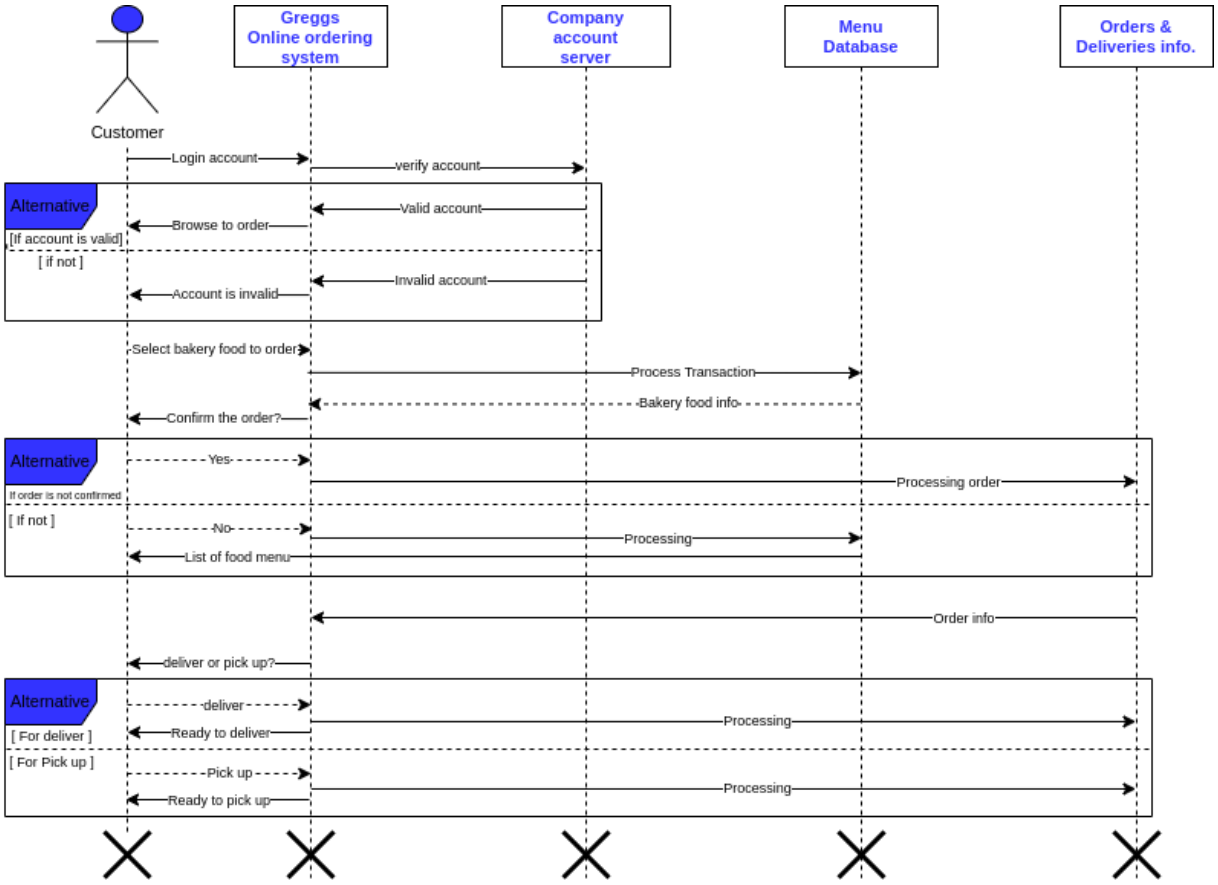

Figure 11 Sequence diagram for Greggs Plc

Sequence diagram is one of the UML diagrams where it shows how a system works and how each object in the diagram interacts with each other in time sequence (gleek.io, n.d.). In the above sequence diagram of Greggs online food ordering system, customers and classes are represented by stick man and objects respectively given in the figure 11. synchronous messages are passed from one object to another through solid lines however, asynchronous messages are sent through dashed lines.

Task - 4

4.1 Comment and observation:

From the above different analysis of Greggs Plc where several different techniques have been used specially five Michael Porter's model and tools such as rich picture and data flow diagram, a good observation is made. Porter's model is used to analyze the Greggs firm and help in their strategic decision. Moreover, it helped firms in the evaluation of profitability (Hart, 2021). Greggs Plc needs to build more customer attraction by opening shops and supermarkets and providing the preferable bakery food to them in every corner of the cities in Europe because this is the customers who purchase the bakery food from Greggs by trusting on the brand name. Therefore, a huge attraction of customers especially in times of disappearing pandemic(Omicron) is not that difficult. Greggs can compare its bakery food with its competitors by looking at their cheapest and most purchasable products and applying vertical integration so that it can help in the efficiency of products, food quality and can be sold at a cheaper rate. Another strategy is packing and design needs not to be so fancy that make food products more expensive (Laja, 2019). And keep the customers attracted toward Greggs Shop by eliminating the threats and new entrants.

There are various diagrams used in this analysis such as rich diagrams where it helps to see overall the organization elements in a single nutshell. Different Gregg's competitors, Gregg's shops, customers, etc are illustrated and what they are thinking is visualized in a single page. Data flow diagram is made to see how the Gregg's food bakery data flows such as how payment data flows when a customer orders food online from Greggs bakery. UML use case diagram is drawn for seeing how many use cases and actors are used in the whole system. A decision tree has been made for payment and order for the system to take decisions on click. Greggs should focus more on the online food ordering system and partner with popular online ordering services because sometimes customers are not interested in walking over the shop instead they order online through their mobile.

4.2 Evaluation

To make the judgment on the basis of above analysis, Dfd, rich diagram, decision tree, etc are used. Dfd level 0 or context diagram has shown the general overview of the online ordering system where dfd level 1 breakdown the main processes of level 0 into sub processes to see a deep level (Valcheva, n.d.). Rich diagrams have shown the links between all the objects graphically such as how shops have linked with staff, customers, warehouse, suppliers and competitors. Moreover what each and every one thinks about the Greggs. This helps Greggs keep an eye on everything in the business. Catwoe analysis provided to solve the business problem by defining six elements (Kukhnavets, 2017).

Sequence diagram is an interaction diagram which captures interaction between the objects (Felici, 2004). Decision tree is used to visually represent decision and decisions making (Gupta, 2017). Power/interest grid is used for categorizing stakeholders when any projects change take place and allow them to manage effectively (improvementservice, 2020). Stakeholders are

plotted in the grid plot having interest on the x axis and power on the y axis. These diagrams help the readers to understand Greggs as a trusted organization in a better way.

Part 2

Introduction

Market Basket Analysis is a data mining technique which is used to reveal frequently bought items from store or retails and it mainly works with association rule (GeeksforGeeks, 2022). MBA is the popular example of frequent itemset mining (Kadlaskar, 2021). It recognizes and is used to analyze the pattern of products which are purchased by customers. MBA helps companies or organizations in their promotion of products or deals, services, offers, sales and achieve highest turnover however, the technique helps to achieve the analysis conveniently. MBA is the finest application for deep learning and machine learning in retail stores. Moreover, it finds the purchasing behavior of the customers from using historical data and what items are likely to be bought together by customers. For example, If a customer is purchasing a t-shirt then what is the probability that the same customer will frequently purchase pants(what color) as well together. Retailing companies used this technique for various decision making processes such as analyzing customer behavior, cross marketing and design of catalogs (Kadlaskar, 2021).

Association rules in the market basket analysis are the prediction of suggested products that are purchased together by customers. It is also known as the {If} - > {Then} rule. It counts the frequencies of the purchased products with the suggested one (teachtarget, 2019). Moreover, transactional data of retail stores or supermarkets are analyzed by association rules. There are three parameters of association rule are given in the following

1. **Support:** It calculates how often the frequent itemsets are appearing in all the transactions. For example, itemset $A = \{t\text{-}shirt\}$ and itemset $B = \{laptop\}$ and there are several transactions that have t-shirts rather than laptop. Now it is easy to guess the next transaction correctly. Therefore, there is higher support for t-shirts than laptops.

$$Support\ (\{A\} \rightarrow \{B\}) = support\ \{A \cup B\}$$
$$or$$
$$Both\ transaction\ A\ and\ B\ /\ total\ number\ of\ transactions$$
$$Support\{A \rightarrow B\} = f(A, B)/S$$

Where f denotes frequency and S denotes total number of transactions.

2. **Confidence:** Confidence is the chance that a customer will buy item X will also buy item Y. Therefore, if the confidence is high then the rule will also be strong. The ratio of the chance of having favorable events is to the chance of having antecedents (Select Statistical Consultants, 2017).

$$Confidence\{X \rightarrow Y\} = Support\{X \rightarrow Y\} / Support\{X\}$$

$$Confidence(\{X\} \rightarrow \{Y\}) = \text{Both } X \text{ and } Y \text{ transaction} / X \text{ transaction}$$

3. **Lift:** Lift is the division of percentage of confidence to percentage of support (GeeksforGeeks, 2022). Lift is equal to 1 means there is no correlation amongst the items, lift is greater than 1 it means there is positive correlation between itemset and if the lift is less than 1 means there is negative correlation (Chauhan, 2019).

$$\text{Lift} = \text{confidencepercentage/supportpercentage}$$

$$Lift(\{X\} \rightarrow \{Y\}) = (X \text{ and } Y \text{ transaction} / (X \text{ transaction})) / \text{fraction of transaction } Y$$

Application

The application is based on the big dataset of an online retail store. In this dataset, Originally there are a total of 5,41,910 rows and 8 columns using a predefined function called *df.info()*. Transactions have been made from 38 countries and it is seen by using code *len(df["Country"].value_counts())* given in the figure 12. Spyder, a software used for analyzing the data and writing programming code in python.

```
import pandas as pd
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules

df = pd.read_excel('/home/gambler/OnlineRetail.xlsx')
print(df.head())
print(df.info)
print(len(df["Country"].value_counts()))
```

Figure 12 source code of counting countries

Index	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	2010-12-01 08:26:00	7.65	17850	United Kingdom
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	2010-12-01 08:26:00	4.25	17850	United Kingdom
7	536366	22633	HAND WARMER UNION JACK	6	2010-12-01 08:28:00	1.85	17850	United Kingdom
8	536366	22632	HAND WARMER RED POLKA DOT	6	2010-12-01 08:28:00	1.85	17850	United Kingdom
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	2010-12-01 08:34:00	1.69	13047	United Kingdom
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	2010-12-01 08:34:00	2.1	13047	United Kingdom
11	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	2010-12-01 08:34:00	2.1	13047	United Kingdom

Table 4 dataframe of online retail store

The dataset is too big, that is why it is only taken for Australia. The python code is written to select only the rows that have a country as Australia and it can be seen in the below image.

```
basket = (df[df['Country'] =="Australia"]
          .groupby(['InvoiceNo', 'Description'])['Quantity']
          .sum().unstack().reset_index().fillna(0)
          .set_index('InvoiceNo'))
print(basket)
```

Figure 13 Source code for selecting the country Australia

InvoiceNo	)box Ipod C	Office Mirrc	UR SPACE	ED PARTY	PEGS IN W	USE PAIN	EGGS HA	E PEG PL/	ARDS WI	IALL TUBE	LL TUBE F	S SMALL T
537813	0	0	0	0	0	0	0	0	0	0	0	0
537815	0	0	0	0	0	0	0	0	0	0	0	0
537816	0	0	0	0	0	0	0	0	0	0	0	0
537817	0	0	24	0	0	0	0	0	0	0	0	0
537819	0	0	0	0	0	0	0	0	0	0	0	0
537822	0	0	24	0	0	0	0	0	0	0	0	0
537823	0	0	0	0	0	0	0	0	1	5	3	2
537825	0	0	0	0	0	0	0	0	0	0	0	0

Table 5 dataframe of country Australia

There are zeros and other numbers are seen but most of them are zeros. And it has to make 1 to the numbers(positive numbers only) other than zero. Now here Hot encoding technique is used to convert 0 or less than 0 to 0 and the number which is greater than 0 converts to 1 and then finally used map function to map it into the Australia dataset. The hot-encoding function source code can be seen in the figure down below.

```
def encode_units(x):
    if x <= 0:
        return 0
    if x >= 1:
        return 1
basket_sets = basket.applymap(encode_units)
```

Figure 14 Source code for hot encoding

Now, the dataset is ready to apply the algorithm known as apriori algorithm which is used to mining frequent itemset and generate association rules (Educative, n.d.). The parameters of the apriori algorithm are support, confidence and lift. The below code image will show and generate itemsets having at least support of 8%. Minimum support can be taken any number but here it is taken 8%. The total transaction for the frequent itemsets are 92 and a sample of 6 transactions is shown below in the image.

```
frequent_itemsets = apriori(basket_sets, min_support=0.08, use_colnames=True)
```

Figure 15 python code for apriori algorithm

Index	support	itemsets
0	0.0877193	frozenset({'4 TRADITIONAL SPINNING TOPS'})
1	0.105263	frozenset({'ALARM CLOCK BAKELIKE GREEN'})
2	0.105263	frozenset({'ALARM CLOCK BAKELIKE RED'})
3	0.122807	frozenset({'BAKING SET 9 PIECE RETROSPOT'})
4	0.140351	frozenset({'BAKING SET SPACEBOY DESIGN'})
5	0.0877193	frozenset({'BLACK/BLUE POLKADOT UMBRELLA'})

Table 6 for support and itemsets

Association Rules i.e lift is taken greater than or equal to 6 and confidence is greater than or equal to 0.8 as it is standard values.

```
aa=rules[ (rules['lift'] >= 6) &
       (rules['confidence'] >= 0.8) ]
```

Figure 16 apriori algorithm for the country Australia

Index	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	frozenset({'ALARM CLOCK BAKELIKE RED'})	frozenset({'ALARM CLOCK BAKELIKE GREEN'})	0.105263	0.105263	0.105263	1	9.5
1	frozenset({'ALARM CLOCK BAKELIKE GREEN'})	frozenset({'ALARM CLOCK BAKELIKE RED'})	0.105263	0.105263	0.105263	1	9.5
6	frozenset({'BLUE HAPPY BIRTHDAY BUNTING'})	frozenset({'PINK HAPPY BIRTHDAY BUNTING'})	0.0877193	0.122807	0.0877193	1	8.14286
8	frozenset({'DOLLY GIRL LUNCH BOX'})	frozenset({'CIRCUS PARADE LUNCH BOX'})	0.105263	0.0877193	0.0877193	0.833333	9.5
9	frozenset({'CIRCUS PARADE LUNCH BOX'})	frozenset({'DOLLY GIRL LUNCH BOX'})	0.0877193	0.105263	0.0877193	1	9.5
10	frozenset({'SPACEBOY LUNCH BOX'})	frozenset({'CIRCUS PARADE LUNCH BOX'})	0.105263	0.0877193	0.0877193	0.833333	9.5
11	frozenset({'CIRCUS PARADE LUNCH BOX'})	frozenset({'SPACEBOY LUNCH BOX'})	0.0877193	0.105263	0.0877193	1	9.5
12	frozenset({'DOLLY GIRL LUNCH BOX'})	frozenset({'REGENCY CAKESTAND 3 TIER'})	0.105263	0.105263	0.0877193	0.833333	7.91667

Table 7 support, confidence and lift for Australia

From the above table shown in the image, it can be clearly seen that confidence is 83.33% for circus parade and dolly girl. It means generally a customer in Australia purchases both the items (circus parade and dolly girl) together however lift is 9.5 shows the positive correlations between both the itemsets. The above data indicates that circus parade(consequents) highly depend on dolly girl(antecedents). This analysis shows that whenever a customer purchases a dolly girl they are likely to purchase a circus parade because without a circus parade there is no use of a dolly girl in toy itemset. The support does not give the guarantee 100% that if someone is buying a circus parade will surely buy a dolly girl because at the index 12 there is a dolly girl in antecedents but there is no circus parade in consequents even though support is 8%. Only 8% chances that both the items could be purchased together.

```
germany_aa=rules[ (rules['lift'] >= 4) &
       (rules['confidence'] >= 0.5) ]
```

Figure 17 Apriori algorithm for the country Germany

Index	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	frozenset({'PLASTERS IN TIN CIRCUS PARADE'})	frozenset({'PLASTERS IN TIN WOODLAND ANIMALS'})	0.115974	0.137856	0.0678337	0.584906	4.24289
7	frozenset({'PLASTERS IN TIN SPACEBOY'})	frozenset({'PLASTERS IN TIN WOODLAND ANIMALS'})	0.107221	0.137856	0.0612691	0.571429	4.14512
11	frozenset({'RED RETROSPOT CHARLOTTE BAG'})	frozenset({'WOODLAND CHARLOTTE BAG'})	0.0700219	0.126915	0.059081	0.84375	6.64817

Table 8 support, confidence and lift for Germany

Second analysis has been done on the country Germany can be seen above two pictures. Lift for Germany is taken 4 and 0.5 for confidence. The table shows that there is 57% confidence for spaceboy and woodland animals. It means in Germany, People who purchase spaceboy likely to purchase woodland animals together however support is 6.

Conclusion and recommendation

From analyzing both the countries, the recommendation has been prepared in the following:
It is seen during analyzing the itemsets of Australia that toys containing circus parade, dolly girl, spaceboy are more purchasable together rather than others therefore, it is recommended to keep mentioned toys in the same section inside the supermarket or shops in order to increase sales, profits and revenue.
On the other hand, during the analysis of Germany it is seen that customers are more interested in purchasing spaceboy and woodland animals together therefore it is recommended to keep these toys on the toy floor specially in separate sections of supermarkets so that parents in Germany could find these toys easily for their kids.
From the above analysis, it is concluded that assertion rule is most important in market basket analysis. It does not help only in buying the frequent itemsets for a seller but also the correlation for the prices of itemsets. This helps companies in their growth of sales, offers, advertisement, and recognizes customers behaviors in terms of purchasing items.

Reference

Neville, S. (2022). *Greggs boss reveals price rises and Covid issues 'as bad as pingdemic'*. [online] Available at: https://www.independent.co.uk/news/uk/roger-whiteside-covid-omicron-b1987744.html [Accessed 20 May 2022].

kamcity, (2022). *Greggs Dominating Food-To-Go Market - KamCity*. [online] Available at: https://www.kamcity.com/namnews/uk-and-ireland/foodservice/greggs-dominating-food-to-go-market/#:~:text=An%20update%20to%20the%20Lumina.

Wikipedia, (2020). *Greggs*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Greggs.

Disruptivnation.com. (2022). [online] Available at: https://disruptivnation.com/wp-content/uploads/2019/10/porters-five-forces.png?w=1024 [Accessed 21 May 2022].

Investing.com UK. (n.d.). *Greggs PLC (GRG) Financial Report Summary*. [online] Available at: https://uk.investing.com/equities/greggs-financial-summary.

Contributor, Pestle. (2015). *How CATWOE helps in Business Analysis and Decision Making*. [online] pestleanalysis.com. Available at: https://pestleanalysis.com/catwoe/.

Business Change Academy. (2020). *What Is CATWOE Analysis?* [online] Available at: https://businesschange.academy/learn/stakeholder/what-is-catwoe-analysis/#:~:text=CATWOE%20Analysis%20is%20a%20technique,direction%20of%20the%20business%20change..

Kukhnavets, P. (2017). *Why Do We Need CATWOE Analysis for Problem Solving?* [online] Available at: https://blog.ganttpro.com/en/catwoe-analysis-soft-systems-methodology-and-root-definitions/#:~:text=CATWOE%20technique%20was%20developed%20by [Accessed 23 May 2022].

Massachusetts News. (2022). *Here's What Greggs Plc (LON:GRG) Shareholder Ownership Structure Looks Like*. [online] Available at: https://darik.news/massachusetts/heres-what-greggs-plc-longrg-shareholder-ownership-structure-looks-like/202202514262.html [Accessed 24 May 2022].

Kollewe, J. (2021). *Greggs warns of rising costs and shortages as it reports bigger sales*. [online] Available at: https://www.theguardian.com/business/2021/oct/05/greggs-costs-shortages-sales-prices.

simplywall.st. (n.d.). *Here's What Greggs plc's (LON:GRG) Shareholder Ownership Structure Looks Like*. [online] Available at: https://simplywall.st/stocks/gb/consumer-services/lse-grg/greggs-shares/news/heres-what-greggs-plcs-longrg-shareholder-ownership-structur [Accessed 24 May 2022].

Skeldon, P. (2021). *Greggs plans digital future as it records full year loss*. [online] Available at: https://internetretailing.net/industry/industry/greggs-plans-digital-future-as-it-records-full-year-loss-22879.

GeeksforGeeks. (2019). *Levels in Data Flow Diagrams (DFD) - GeeksforGeeks*. [online] Available at: https://www.geeksforgeeks.org/levels-in-data-flow-diagrams-dfd/.

Walker, A. (2019). *UML Class Diagram Tutorial with Examples*. [online] Available at: https://www.guru99.com/uml-class-diagram.html.

www.gleek.io. (n.d.). *Diagram maker for developers*. [online] Available at: https://www.gleek.io/sequence.html [Accessed 25 May 2022].

GeeksforGeeks. (2022). *Market Basket Analysis in Data Mining*. [online] Available at: https://www.geeksforgeeks.org/market-basket-analysis-in-data-mining/.

Analytics Vidhya. (2021). *Market basket Analysis | Guide on Market Basket Analysis*. [online] Available at: https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-market-basket-analysis/.

Select Statistical Consultants. (2017). *Market Basket Analysis: Understanding Customer Behaviour*. [online] Available at: https://select-statistics.co.uk/blog/market-basket-analysis-understanding-customer-behaviour/.

KDnuggets. (n.d.). *Market Basket Analysis: A Tutorial*. [online] Available at: https://www.kdnuggets.com/2019/12/market-basket-analysis.html.

Educative. (n.d.). *What is the Apriori algorithm?* [online] Available at: https://www.educative.io/edpresso/what-is-the-apriori-algorithm.

Hart, D. (2021). *Porter's Five Forces: Increase The Profitability of Your Business*. [online] Available at: https://www.thepowermba.com/en/blog/porters-five-forces.

Valcheva, S. (2018). *Data Flow Diagram: Examples (Context & Level 1), Explanation, Tutorial*. [online] Available at: https://www.intellspot.com/data-flow-diagram-examples/.

Kukhnavets, P. (2017). *How to Solve Business Issues With the Help of CATWOE Analysis?* [online] Gantt Chart GanttPRO Blog. Available at: https://blog.ganttpro.com/en/catwoe-analysis-soft-systems-methodology-and-root-definitions/.

Felici, M. (2004). *Sequence Diagrams*. [online] Available at: https://www.inf.ed.ac.uk/teaching/courses/seoc/2011_2012/notes/SEOC08_notes.pdf.

Gupta, P. (2017). *Decision Trees in Machine Learning*. [online] Towards Data Science. Available at: https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052.

improvementservice, I. (2020). *Power/Interest Grid*. [online] Improvement Service. Available at: https://www.improvementservice.org.uk/business-analysis-framework/consider-perspectives/powerinterest-grid#:~:text=The%20power%2Finterest%20grid%20is.