

# **APAN 5310 SQL & Relational Databases**

## **Revolutionizing Decision-Making System at ABC Foodmart**

**Group 1:**

**Name: Fangran Liu   UNI: fl2662**

**Name: Jui-Hung Chen   UNI: jc6156**

**Name: Joanne Lin   UNI: yl5504**

**Name: Ranqui Xu   UNI: rx2235**

**Name: Shanshan Wang   UNI: sw3858**

**Name: Tse An Chen   UNI: tc3336**

## Client scenario

Our team's goal is to improve ABC Foodmart's decision making system, which was reported to be occasionally sluggish and errors. Furthermore, with the aim of expanding its presence by opening three additional branches in Brooklyn, a more efficient data management system, seeking real-time dashboards to manage and monitor vital aspects of the business data, is recognized as the critical need. Therefore, this report will describe a comprehensive overview of the improved system, detailing analytical steps and providing supplementary support files.

The improved system's plan is to streamline operations, improve efficiency to ensure scalability when the foodmart expands continuously. The process involves a detailed review of existing business processes, designing a core enterprise resource planning system, implementations in different branches and the establishment of KPIs. The basic steps are listed below:

- Reviewing existing business processes
  - Conduct a thorough analysis of current processes, identifying pain points and inefficiencies. This analysis will explore beyond the surface-level, which aims to not only find the errors, but also understand the bottlenecks and how to avoid them in the future business.
  - Assess the need for ERP modules focused on sales, store locations, staff management, deliveries, gross margin and so on.
  - Understand how these processes can be interconnected and manage them through a combination of interviews, workflow observations and data analysis, helping the team gain a comprehensive understanding of ABC foodmart's operational landscape.
- Designing core ERP system modules
  - Based on the assessment, categorize and design core ERP modules to address the identified needs. They will be tailored to meet the specific needs in particular phases. For example, various modules focused on sales, store locations, staff management, deliveries and other parts of the business will be built.
  - Prepare for data migration from spreadsheets and paper binders to the new system with each module and phases being crafted meticulously.
- Implementations in different branches

- It can be implemented in various branches of ABC Foodmart, and prioritized modules based on their importance to each branch's need. For example, sales and staff management may be prioritized in specific branches.
- Design technical support strategies to facilitate employees and managers in understanding and utilizing the system. Besides, we will provide robust technical support strategies, including the provision of a comprehensive training module and resources to ensure the smooth operation of the new system.
- Set key metrics and design analytical models
  - Establish key metrics and design analytical models to create visible and measurable KPIs. By collaborating closely with stakeholders, particular KPIs aligning with different branches and strategies will be defined.
  - Ensure the design is scalable, anticipating future expansion of ABC Foodmart. The system is able to accommodate the future expansion without significant reconfiguration.

### **Stakeholder responsibilities**

#### *Tse-An Chen*

- Project scenario brainstorming`
- Conducted team contract & stakeholder responsibilities
- Conducted ETL process
- Conducted analytical showcase examples with queries

#### *Jui-Hung Chen*

- Project scenario brainstorming
- Conducted normalization plan & ER diagram
- Conducted ETL process

#### *Joanne Lin*

- Project scenario brainstorming

- Conducted normalization plan & ER diagram
- Conducted ETL process

*Fangran Liu*

- Project scenario brainstorming
- Conducted consultant/client scenario and problem statement
- Conducted dataset tables outline

*Shanshan Wang*

- Project scenario brainstorming
- Conducted problem statement
- Conducted dataset tables outline
- Conducted analytical showcase examples with queries

*Ranqiu Xu*

- Project scenario brainstorming
- Conducted brief outline of project goal
- Conducted database interaction plan & dashboard showcase

## **Dataset Description**

This data contains a broader range of metrics that can provide a comprehensive overview of FOODmart Ltd, which is one of the biggest supermarkets in Australia. This dataset does not only include both operational and financial figures like wage, sales, trading hours, and number of employees but also contains more specific details regarding the store, such as the type of location and the state of store location.

The dataset was selected for its feature to provide a multi-dimensional analysis by incorporating operational, financial, and human resources across different stores. With this dataset, our team is able to develop a comprehensive management system that supports not only current operations but also optimizes the information system and strategy for expanding the Brooklyn stores. The

transition from paperwork to real-time dashboards with this dataset allows ABC Food Mart to gain immediate access and insight. Based on the optimization, ABC Food Mart will be able to facilitate the store's system by facilitating accurate decision-making and reducing errors during the expansion.

StoreN	Sales \$m	Wages \$m	NoStaff	Location	Loc'n (Num)	State	State (Num)	AgeYrs	GrossProfit	Adv_1000	Competitors	HrsTrading	Sundays
1	12.5	2.3	60	Mall	2	NSW	1	10	0.712	171	3	110	Sun:No
2	14.5	2.7	69	Mall	2	Vic	2	8	0.091	213	5	134	Sun:No
3	19	3.1	79	Country	3	Vic	2	7	1.075	255	2	98	Sun:Yes
4	18.2	2.6	66	Mall	2	Qld	3	7	1.372	287	1	85	Sun:Yes
5	7.6	2	51	Strip	1	Qld	3	15	2.148	112	0	72	Sun:Yes
6	18.5	2.7	62	Country	3	Vic	2	6	2.019	238	0	77	Sun:Yes
7	13.1	2.4	61	Country	3	NSW	1	7	0.662	124	2	100	Sun:Yes
8	14.9	2.5	59	Mall	2	NSW	1	6	0.7	214	2	95	Sun:No
9	17.1	2.7	65	Country	3	WA	5	8	0.937	215	3	112	Sun:No
10	9.2	2.1	55	Strip	1	Vic	2	16	0.065	154	5	135	Sun:Yes
11	10.3	2.2	65	Mall	2	Vic	2	10	2.144	97	2	100	Sun:No
12	19.3	3.1	74	Mall	2	ACT	8	7	0.248	301	2	96	Sun:No
13	8.1	1.8	43	Strip	1	Vic	2	23	1.607	123	1	72	Sun:Yes

<https://www.kaggle.com/datasets/dermisfit/foodmart-dataset>

## Normalization Plan & ER Diagram

*Normalization:*

[https://docs.google.com/spreadsheets/d/1-iQEOqiDt6HhRyv\\_kY2uasYnqxWcVHPmPmASRmk1zE/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1-iQEOqiDt6HhRyv_kY2uasYnqxWcVHPmPmASRmk1zE/edit?usp=sharing)

*SQL code for normalized tables(only for lion mail):*

<https://drive.google.com/drive/folders/1ArdBYB-LN5Nmc9hwRo-0klMID1K03X9W?usp=sharing>

*ER Diagram Lucidchart*

[https://lucid.app/lucidchart/5af80994-de42-4d6e-bb2f-2e9b35dcbb78/edit?viewport\\_loc=-661%2C401%2C2723%2C1442%2C0\\_0&invitationId=inv\\_b65a7a58-0f34-400f-8d6b-0711fe552f15](https://lucid.app/lucidchart/5af80994-de42-4d6e-bb2f-2e9b35dcbb78/edit?viewport_loc=-661%2C401%2C2723%2C1442%2C0_0&invitationId=inv_b65a7a58-0f34-400f-8d6b-0711fe552f15)

**ETL process in detail.**

*GitHub repo:*

<https://github.com/ThisisJoanne/PostgreSQL-with-Python/tree/main>

## Normalization - 3NF

Since our dataset does not have any duplicate attributes, 1NF remains the same as the original dataset. For the 2NF, we make sure all the attributes in the table rely on a primary key, and we divide the 1NF to store info, financial analysis, revenues, and employee tables. The separation of the tables is based on the correlation between each attribute. For instance, storeinfo contains AgeYrs, State, Location, NoStaff, Competitors, and HrsTrading.

In the final step, achieving the Third Normal Form (3NF), we eliminate attributes that do not have a direct relationship with the primary key. This involves distributing the data across multiple tables, each designed to focus on a specific aspect of the information. By doing so, we ensure that each table only contains data that directly pertains to its primary key, enhancing clarity and reducing redundancy in the database structure.

## ETL Process

For the extract phase, the scripts start by reading the dataset that we obtained from Kaggle. Second, we connected the PostgreSQL by using psycopg[`binary`] in order to insert the SQL code in the Python environment.

For the transform phase, in order to properly arrange data into tables, establishing a structured schema is essential. The script defines the SQL commands to create different tables, including store details, service, revenue, hours, financial analysis, employees, cost, storeinfo, manager, and manager\_details.

For the load phase, we import the data into the tables that we created in the transform phase by executing SQL commands.

## Analytical procedures

### *1. Which state has the highest sales revenue in million?*

The analysis shows that New South Wales (NSW) is the top-performing state, while the Northern Territory (NT) is the lowest performer. From a business standpoint, the higher performance in NSW could be attributed to its larger population, although this also suggests a more competitive market for foodmarts. Conversely, the Northern Territory,

being less competitive, presents a significant opportunity for market expansion.

	state character varying (50) 🔒	total_revenue numeric 🔒
1	NSW	461.80
2	Vic	348.60
3	Qld	294.20
4	SA	233.00
5	WA	217.60
6	ACT	76.10
7	Tas	62.40
8	NT	55.20

***2. Which state is the most competitive in the food market industry, how many competitors?***

The previous query indicates that New South Wales (NSW) generates the most revenue and now we can know it has more competitors. Compared with NSW, the Australian Capital Territory (ACT), Northern Territory (NT), and Tasmania (Tas) have fewer competitors, and have more opportunities to expand foodmart in these areas.

	state character varying (50) 🔒	total_competitors bigint 🔒
1	NSW	95
2	Vic	69
3	Qld	60
4	SA	59
5	WA	34
6	ACT	23
7	NT	20
8	Tas	16

***3. How many supermarkets in each state have home delivery service?***

The result indicates that supermarkets offering home delivery services do not necessarily report higher profits. For instance, while South Australia (SA) ranks fourth in terms of performance, only two stores in the state offer home delivery. This suggests that home delivery does not play a crucial role in influencing revenue.

	state character varying (50) 🔒	home_deliver bigint 🔒
1	NSW	11
2	Vic	11
3	Qld	9
4	WA	4
5	ACT	3
6	SA	3
7	NT	2
8	Tas	2

**4. Do stores open on sunday have higher gross margin than those do not open on sunday?**

Based on gross margin, 13 out of the 18 supermarkets that are open on Sunday rank highly. It is also worth to mention that the top six supermarkets with the highest gross margins all operate on Sundays. This indicates that extended operating hours, such as opening on Sundays, could be a significant factor in achieving higher revenues.

	storeno integer 🔒	gross_margin numeric (10,2) 🔒	sundays boolean 🔒
1	86	34.10	true
2	52	33.00	true
3	5	28.30	true
4	137	24.70	true
5	91	21.40	true
6	47	21.00	true
7	11	20.80	false
8	129	20.20	false
9	13	19.80	true



	storeno integer 🔒	gross_margin numeric (10,2) 🔒	sundays boolean 🔒
10	54	19.30	false
11	148	19.30	false
12	58	19.10	true
13	133	19.00	true
14	92	18.80	true
15	145	18.50	false
16	51	18.30	true
17	85	18.30	true
18	74	18.00	true

## 5. Do older managers generate higher gross margin than younger managers?

Prior to evaluating the analysis, it was presumed that older managers might contribute to higher gross margins compared to younger managers. However, the findings reveal that younger managers are equally capable of producing high gross margins. Interestingly, in the top five states with the highest gross margins, the managers are approximately 40 years old.

	state character varying (50) 🔒	average_age numeric 🔒	average_gross_margin numeric 🔒
1	Qld	38.777777777777778	11.1851851851851852
2	Vic	34.8000000000000000	10.2200000000000000
3	WA	40.6875000000000000	10.1312500000000000
4	NSW	39.4000000000000000	9.0700000000000000
5	SA	40.0454545454545455	8.0000000000000000
6	ACT	36.5000000000000000	7.1166666666666667
7	Tas	42.4000000000000000	6.0800000000000000
8	NT	33.2500000000000000	3.8500000000000000

## 6. Which Location(Country, Strip, Mall,etc) contribute the most sales?

This output provides a business insight on location selection for future expanding strategic decisions, by knowing if there are location-related reasons contributing to the high sales, such as being in a high-traffic area or an economically prosperous region.

storeno integer	sales_m numeric (10,2)	location character varying (255)	state character varying (50)
65	23.50	Country	WA
115	21.00	Country	NT
122	20.40	Country	Qld
17	19.50	Strip	Qld
12	19.30	Mall	ACT

The top-performing stores are located in 'Country' areas. This could indicate that stores in less urbanized or rural areas have the potential to generate significant sales. It might suggest a lack of competition in the country areas or a market demand has been fulfilled by the store.

## 7. Is advertising a crucial factor of high revenue?

storeno integer	sales_m numeric (10,2)	adv_1000 numeric (10,2)	adv_m numeric	adv_percentage_of_sales numeric
65	23.50	348.00	0.34800000000000000000	1.48085106382978723400
115	21.00	296.00	0.29600000000000000000	1.40952380952380952400
122	20.40	284.00	0.28400000000000000000	1.39215686274509803900
17	19.50	369.00	0.36900000000000000000	1.89230769230769230800
12	19.30	301.00	0.30100000000000000000	1.55958549222797927500

This table showed a percentage of advertising cost in sales revenue. It shows a significant portion of the store's sales is being reinvested into advertising. Investing advertisement increases the potential customer base and retains current customers. And its effect went back to sales. By comparing these percentages between stores, companies are able to evaluate the effectiveness of advertising. In order to optimize advertisement or switch to another promotion plan.

## 8. Does a higher percentage of unionized employees (UnionPerc) correlate with different cost structures or sales figures?

unionperc_range text	average_wages numeric	average_sales numeric
0-25%	2.6000000000000000	10.133333333333333
26-50%	2.3721088435374150	11.6904761904761905

According to the data analysis, sales performance increases with a higher percentage of

unionized workers. This could suggest that stores with a moderate unionization percentage (between 26 and 50 percent) may be gaining benefits from unionization organizations like better training, stronger staff morale, or more experienced employees, all of which can boost sales.

**9. Is there a relationship between the number of staff and store performance metrics like sales or profit margins ?**

storeno integer	no_staff integer	sales_m numeric (10,2)	grossprofit numeric (10,2)	profit_margin_percentage numeric
44	117	15.30	0.10	0.65359477124183006500
30	102	16.30	0.08	0.49079754601226993900
17	99	19.50	1.29	6.61538461538461538500
34	98	16.10	0.97	6.02484472049689441000
120	96	11.40	0.83	7.28070175438596491200
46	91	14.80	1.97	13.31081081081081100
126	89	18.30	0.71	3.87978142076502732200

It's noteworthy to note that there isn't a direct, obvious correlation between staff number and profit margins or sales. There doesn't seem to be a consistent trend in the profit margin percentages with the number of staff. Compared to Store No. 34, which has 98 people and has a lower profit margin percentage of 6.02%, Store No. 120 has 96 employees and a higher profit margin percentage of 7.28%. Higher profit margins at stores with fewer staff members could indicate better operational efficiency or higher productivity per employee. With the fewest employees, store No. 28 has a comparatively high profit margin of 17.47%. Its business plan is deemed effective and serves as a model for other retailers.

**10. Is there a relationship between the number of training managers receive (Mng\_Train) and store performance in terms of sales or profitability?**

mng_train character varying (100) 🔒	average_sales numeric 🔒	average_profit numeric 🔒
1	14.5200000000000000	1.002000000000000000
2	11.8423076923076923	1.01461538461538461538
3	10.9489795918367347	1.01122448979591836735
4	9.9444444444444444	1.2577777777777778
5	13.9250000000000000	0.69083333333333333333
6	8.1750000000000000	0.822500000000000000

Surprisingly, there is no obvious pattern in the data indicating that higher training sessions always correlate with higher sales or profit margins. There may be diminishing returns on more training, as managers with six training sessions have the lowest average sales and profit whereas managers with five training sessions have higher sales but lower average profits. Based on this data, there seems to be an optimal training session at the number four. By doing this, this may indicate an increase in average profit gain.

## Database Interaction

To interact with the designed database system, our approach includes specific tools and methodologies tailored to different user groups within the organization, specifically analysts and C-level executives.

### *For Analysts:*

Analysts are provided with direct access to the database through SQL interfaces such as pgAdmin, enabling them to perform both complex and routine queries. Additionally, they will use an IPython Notebook equipped with pre-built scripts for analysis, data manipulation, and visualization, all coded in Python. This setup not only facilitates sophisticated data handling but also enhances the efficiency of executing predefined analytical processes.

### *For C-Level Executives:*

A customized dashboard is designed for C-level executives, integrating Excel for data sourcing and including interactive elements via Jupyter Notebook widgets. This allows executives to modify parameters and see updated results instantaneously, providing a dynamic and engaging data interaction experience. Furthermore, the IPython Notebook supports the generation of visual reports in formats such as PDF and HTML, specifically catered for executive review, ensuring clarity and depth in reporting.

### ***General Benefits:***

The integration of Python with tools like Jupyter Notebooks and SQL interfaces significantly streamlines data integration and automation processes, catering to both technical and non-technical users. The inclusion of interactive dashboards and widgets makes it possible for all users, including those without technical expertise, to effectively engage with and derive insights from the data, promoting an inclusive data-driven decision-making culture within the organization.

### **Redundancy and Performance**

Incremental Backups will be good for the future of ABC Foodmart', These backups only save data that has changed since the last backup. They are less resource-intensive than full backups and can be performed more frequently. However, restoration requires the last full backup and all subsequent incremental backups, making the process more complex. For a chain store that needs to develop, more reliable redundancy and performance, we must not use Full Backups just because of convenience

To ensure redundancy and performance for the retail chain, implementing an incremental backup strategy is both efficient and cost-effective. Initially, a comprehensive backup of the entire system is necessary to establish a complete data set. Subsequently, incremental backups should be scheduled based on business needs, such as daily or hourly, to capture only the data that has changed since the last backup. This approach significantly reduces storage demands and speeds up the backup process. Selecting a cloud service platform that supports incremental backups, like AWS's Amazon RDS or Google Cloud Platform's Cloud SQL, is crucial as these services offer robust data backup and recovery functionalities.

For ongoing maintenance and optimization, it's essential to regularly test and verify the integrity and effectiveness of the backups. This ensures that the business can rapidly resume operations in the event of data loss. Monitoring backup activities to ensure they execute as planned and without errors is also critical. Over time, as the business grows and data evolves, the backup strategy should be adjusted to optimize resource use and cost. Incremental backups not only reduce the backup window but also enhance the speed of data recovery, thus improving data management efficiency and strengthening business continuity and customer trust.

As a highly scalable, flexible and cost-effective business, cloud hosting is the superior choice due to its scalability, redundancy, performance, and cost-efficiency.

**Scalability:** The cloud provides elastic scalability which is essential for businesses with variable demand across multiple locations. It allows for easy adjustment of resources in line with business needs without the upfront cost of physical infrastructure.

**Redundancy and Reliability:** Cloud services offer robust disaster recovery and redundancy plans. Data can be backed up in multiple locations automatically, enhancing business continuity compared to on-premises solutions where setting up such redundancy is often more complex and costly.

**Performance:** Modern cloud services can deliver high performance that rivals or exceeds on-premises solutions, with the added benefit of geographic distribution of data centers reducing latency for different regional markets within Australia.

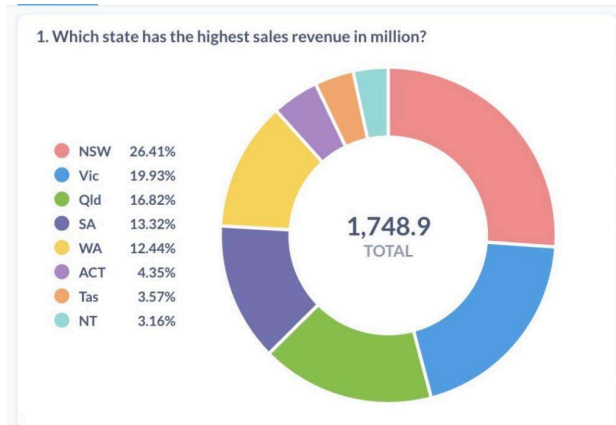
**Cost and Management Efficiency:** Cloud solutions reduce the need for ongoing capital investment in hardware and physical infrastructure management. They offer a pay-as-you-go model that can be more cost-effective, particularly for businesses needing to manage diverse and geographically spread operations.

Cloud hosting provides them with a powerful platform for high redundancy and excellent performance. By choosing cloud services, chain stores can take advantage of the automation and management tools provided by the cloud to not only ensure high availability of data and applications, but also to quickly adjust resources based on real-time changes in customer flow and sales data. This kind of flexibility and scalability is hard to match with local hosting, especially when you need to support stores dispersed across multiple states.

## **Dashboard showcase**

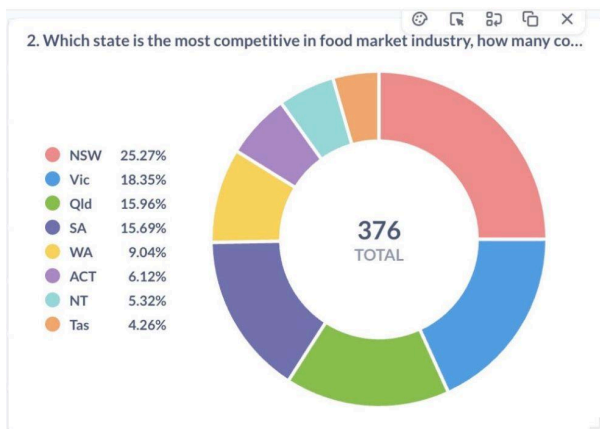
1

This dashboard displays revenue by state, which can reflect a variety of information such as the market saturation, market share, rankings and trends. It can help C- level offer quickly capture key business metrics and compare performance across states



2

This dashboard displays competitors by state, which can reflect a variety of information such as competition density, market potential indicators. It can also be linked with the previous dashboard to show the relationship between activity revenue and the number of competitors. It can help C-level offer understand the market structure, assess the competitive environment, and identify potential regions for business expansion



3

This dashboard displays home delivery by state, which allows users to understand which states are more active in this field and which states may have market space. It can help C-level offer identify geographic areas for service optimization or expansion, analyze market demand, and assess the potential for service delivery



4

This dashboard displays gross margin and business hours by state, which can help business leaders quickly gain insight into the profitability of each store and make data-based business decisions, such as adjusting business hours to improve profitability. In addition, it can also be used as a tool to evaluate the impact of Sunday operations on gross margin



## Conclusion

In conclusion, our primary goal was to enhance ABC Foodmart's decision-making system in order to facilitate the firm's expansion. This included the implementation of a robust relational database management system, efficient extract, transform and load process, along with various comprehensive data analysis techniques. By accomplishing these, we aimed to provide the firm with real-time insights and streamlined operations to enhance its efficiency and scalability, and better prepare for future expansion and businesses.

The implementation of the RMDS enabled the firm to centralize its data and streamline its management across all branches. This centralized approach facilitated easier access to massive business information. Moreover, the ETL processes ensured that data from disparate sources could be seamlessly integrated into the RMDS, providing a comprehensive view of the company's operations.



Through delving into data analysis, we were able to uncover valuable insights into various aspects of ABC Foodmarts's businesses, and provided 10 examples of analytical procedures. For example, by comparing different sales in various stores, we can build different strategies for different stores and different prioritized modules and phases. Another cross-functional example is that we can analyze if home delivery or parking spot influences the sales of the specific store, to see if the strategy and the decision making system could be adjusted to improve in order to increase the revenue.

Our RDMS implementation offered numerous benefits, including improved data management, enhanced decision-making capabilities, scalability, redundancy, and efficiency gains. The comprehensive insights gained from data analysis empowered ABC Foodmart to make informed decisions, optimize resource allocation, and drive business growth. Overall, our efforts contributed to ABC Foodmart's success by providing a centralized platform for efficient data management and enabling data-driven decision-making in a dynamic and competitive market landscape.