



SQL Based Business Performance Analysis

A Data Analysis Approach with SQL: Data-Driven Insights for Northwind Traders Company Operations

Monali Patil

Project Overview

The company, Northwind Traders, specializes in food imports and exports across the globe. All trading data is stored in the Postgres database called "Northwind." This project aims to help various departments like Logistics, HR, Pricing, and Sales gain valuable insights from the company's operational data through SQL queries. These insights are used to assess performance and guide decision-making for the upcoming year.

Business Problem

This project focuses on analysing the performance of Northwind Traders, a company that deals with the import and export of specialized foods globally. The project utilizes a PostgreSQL database to store and analyse the company's trading and transactional data, which consists of 15 datasets from various departments, including Product, Logistics, Human Resources, Pricing, and Sales. The project aims to extract insights to inform strategic decisions across departments by querying the data using SQL.

Business Inquiries

Key departments raised specific business inquiries:

- **Product Team:** Assess product pricing strategy based on products priced between \$10 and \$50.
- **Logistics Team:** Retrospective analysis of shipping performance for the years 1996–1997.
- **HR Team:** Determine the age of employees when they joined the company and their current managers.
- **Pricing Team:** Identify product price changes and compare products against the category's average and median prices.
- **Sales Team:** Build KPIs to measure employee performance based on sales amount, discounts, and total orders

Data Understanding

The organisation stores and maintains all its trading and transactional data in the Postgres database named “Northwind” consisting of 15 datasets for various respective divisions’ details. The below table gives high-level information about the database and its tables.

Database Name: Northwind			
Sr.No.	Table Names	Count of Records	Description
1.	categories	8	Master table of food categories.
2.	products	77	Master table containing products details.
3.	suppliers	29	Master table of food suppliers.
4.	order_details	2155	Secondary table of food orders information on orders placed by customers.
5.	orders	830	Master table tracking all orders placed by customers.
6.	employees	9	Master tables of the company’s employee details.
7.	employeeterritories	49	Secondary table of employees & territories.
8.	territories	53	Master table of employee’s territories.
9.	region	4	Master table of regional employee allocation and related data.
10.	customers	91	Master table of customers information for all orders.
11.	customercustomerdemo	0	-
12.	customerdemographics	0	-
13.	shippers	6	Master table of order shippers.
14.	shippers_tmp	6	-
15.	usstates	51	-
Total Datasets:		15	

Table A: Database and its tables details.

The master tables include all the distinct records of a specific entity whereas secondary tables hold all the records of the entity’s doings. For instance, an order table consists of unique order information set by customers. Similarly, order_details contains all information and features of each and every particular order which specifies (order -> order_details) one to many relation between order and order_details table.

An Entity Relationship Diagram (ERD) below visually represents the relationships between these datasets, providing a comprehensive view of the database schema. It displays the relational integrity, relationship between entities, and different attributes of the datasets. And these  symbols indicate one -> many relation between the tables.

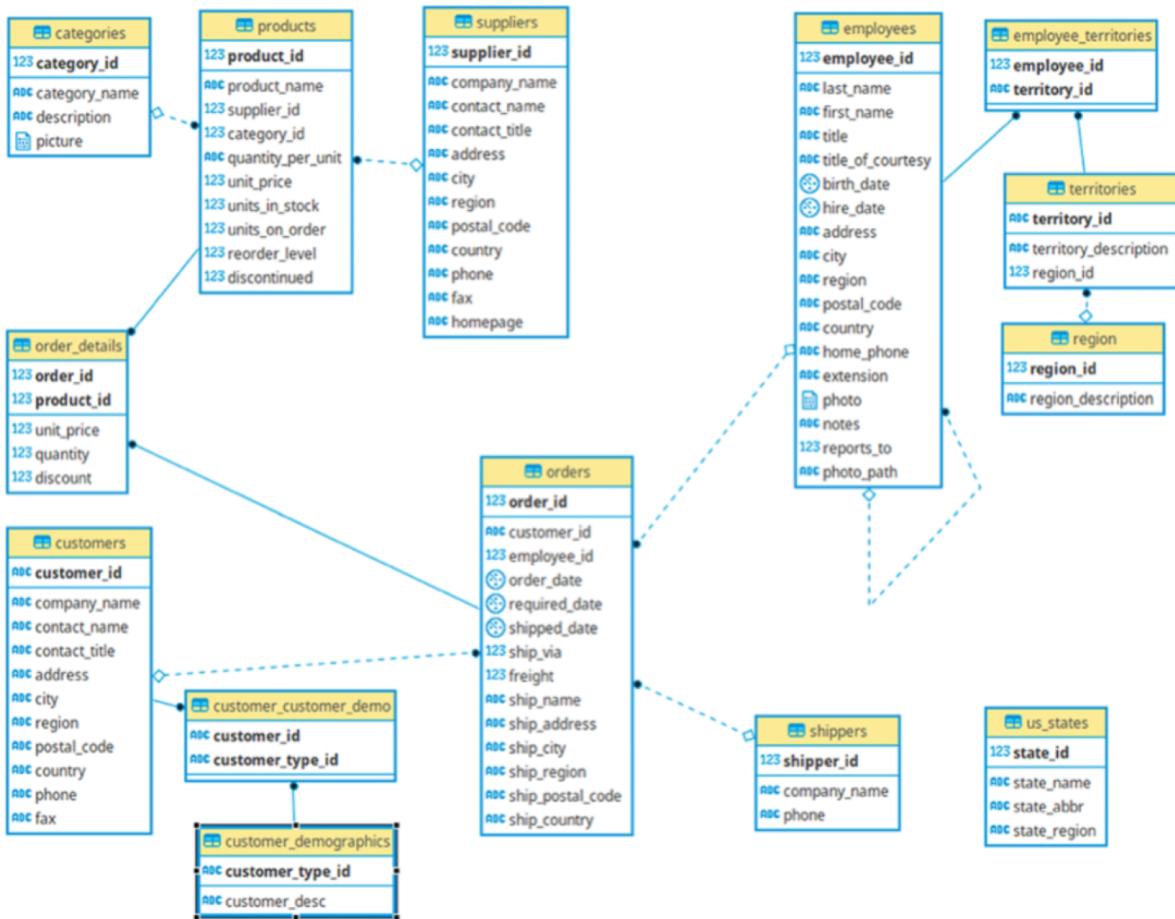


Figure A: Entity Relationship Diagram (ERD) for the dataset.

Various departments of the company essentially Product, Logistics, Human Resource, Pricing, and Sales needed to analyse the performance of their business relations and thus to facilitate this objective, data from the below highlighted datasets provided important knowledge which would assist them to take further strategic actions and decisions.

Sr.No.	Table Names	Count of Rows
1.	categories	8
2.	products	77
3.	suppliers	29
4.	order_details	2155
5.	orders	830
6.	employees	9
7.	employeeterritories	49
8.	territories	53
9.	region	4
10.	customers	91
11.	customercustomerdemo	0
12.	customerdemographics	0
13.	shippers	6
14.	shippers_tmp	6
15.	usstates	51

Table B: Highlighting datasets that help acknowledge business purposes.

Database Setup: Instructions to Create and Connect to the Postgres Database.

Below are the steps to position the open-source Postgres database environment.

1. Download and install the Postgres database as per the desired OS environment.
For instance, PostgreSQL version 11.17
2. Then download and install any integrated development environment (IDE) tool that would facilitate communication with the Postgres database.
For instance, download and install the DBeaver database tool which will act as a database manager and SQL client to interact with the Postgres database.

Below are the steps to Create and Connect to the Database (Northwind) in the Postgres environment.

1. For the creation of the database, we need to execute a command in the DBeaver script console:
Syntax: `createdb <database name>;`
For Example: `createdb Northwind;`
Note: If the database is successfully created then there will be no response/error message else a relevant message will be prompted, and we need to debug and fix it.
2. Then create the schema/user under which all the tables/datasets will reside/be created. If the schema is not created, tables will be part of the default schema named "public".
3. Create a table. Following is the syntax.

Syntax:

`CREATE TABLE <table name> (`

```

<"1st attribute name"> <data type>,
<"2nd attribute name"> <data type>,
<"3rd attribute name"> <data type>
);

```

For Example:

Creating table: categories with the attributes: category_id (with datatype: smallint), category_name (character), description (text) and picture (bytea)

```

CREATE TABLE categories (
  "category_id" smallint NOT NULL, <- NOT NULL ensure value is left NULL
  "category_name" character varying(15) NOT NULL,
  "description" text,
  "picture" bytea
);

```

4. Insert data/records in the tables. Following is the syntax.

Syntax:

```

INSERT INTO <table name> VALUES (<value for 1st attribute>, <value
for 2nd attribute>, <value for 3rd attribute>);

```

For Example:

Inserting records in the table.

```

INSERT INTO categories VALUES (1, 'Beverages', 'Soft drinks,
coffees, teas, beers, and ales', '\x');
INSERT INTO categories VALUES (2, 'Condiments', 'Sweet and savory
sauces, relishes, spreads, and seasonings', '\x');

```

5. In the DBeaver tool which operates like both the Postgres database manager and SQL client to connect with the database, it will look like the below after the successful formation of all tables and records.

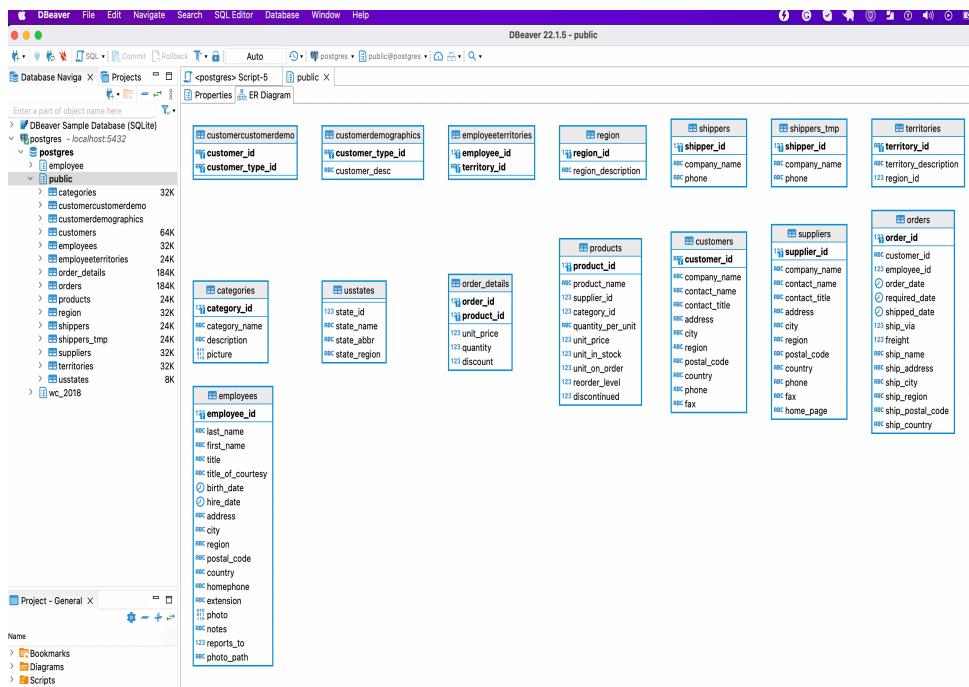


Figure B: After the creation of all the tables and their attributed view in the DBeaver database tool.

6. In addition, connecting and accessing the data from the Postgres database through programming languages like Python requires the use of the psycopg2 package which assists in establishing a connection and interacting with the Postgres database. Following are the steps.

- a. Import psycopg2 package.

```
import psycopg2
```

- b. Create connection variables.

```
host: <Hostname on which Northwind Postgres database resides>
port: <Port number of the database listens to requests>
user: <User name to connect to the database>
password: <"Password for the user to connect to the database">
```

For instance:

```
host: localhost
port: 5432
user: admin
password: "*****"
```

- c. Create a connection with the help of the above connection variables and the package.

```
connection = psycopg2.connect(
    user=user,
    password=password,
    host=host,
    port=port,
    database=Northwind
)
```

- d. Check and verify the established connection.

```
connection.get_dsn_parameters()
connection.status
```

Note: Output 1 signifies a successful connection.

Answers to Business Inquiries.

1] For their annual review of the company pricing strategy, the Product Team wants to look at the products that are currently being offered for a specific price range (\$10 to \$50).

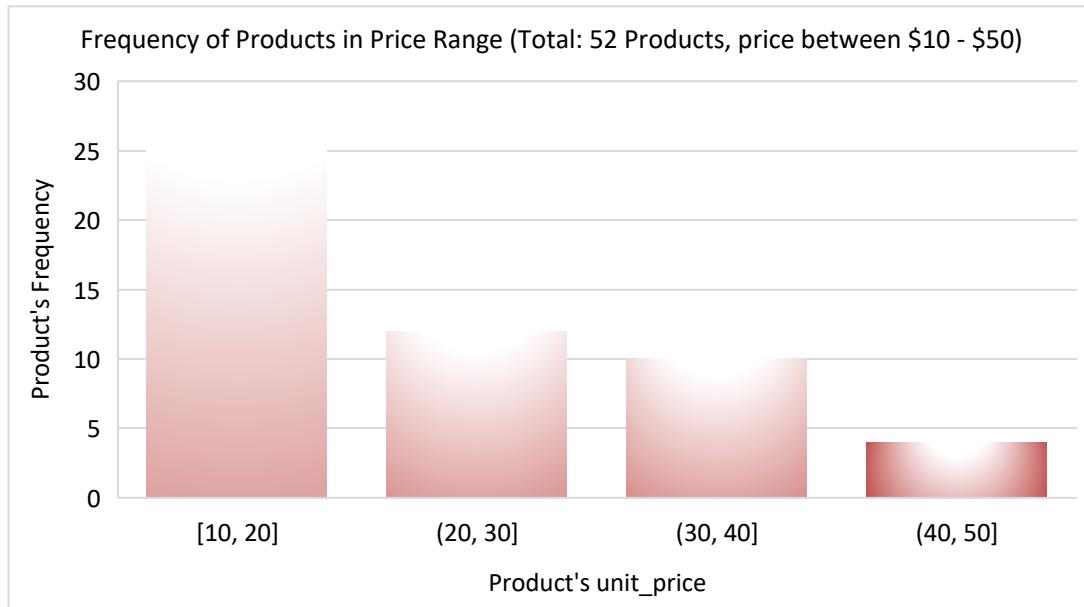


Figure 1: Number of products in the various specific price range between \$10 and \$50.



Figure 2: Product names with a price range between \$10 and \$50.

The above figure 1, illustrates the number of products in the 4 different price groups ranging from \$10 to \$50 which the Product team wants to consider for the annual review of the company pricing strategy. Below are some related highlights.

- Total: 52 food products that are currently available/not discontinued in this requested price range.
- The majority of the products 26 (i.e. 50% of the total) are within the price group [10 – 20], followed by 12 and 10 products from the price group [20 – 30] and [30 – 40] respectively.
- The name of 4 products from the price group [40 – 50] are: Tarte au sucre, Ipoh Coffee, Schoggi Schokolade and Vegie-spread.

- There are only 3 products for \$10 namely: Aniseed Syrup, Longlife Tofu and Sir Rodney's Scones.
- Overall half of the available food products (26) are above \$20 and the rest (26) are below \$20 signifying it's a 50:50 ratio.

2] The Logistics Team wants to do a retrospection of their performances for the year 1997, in order to identify for which countries they didn't perform well.

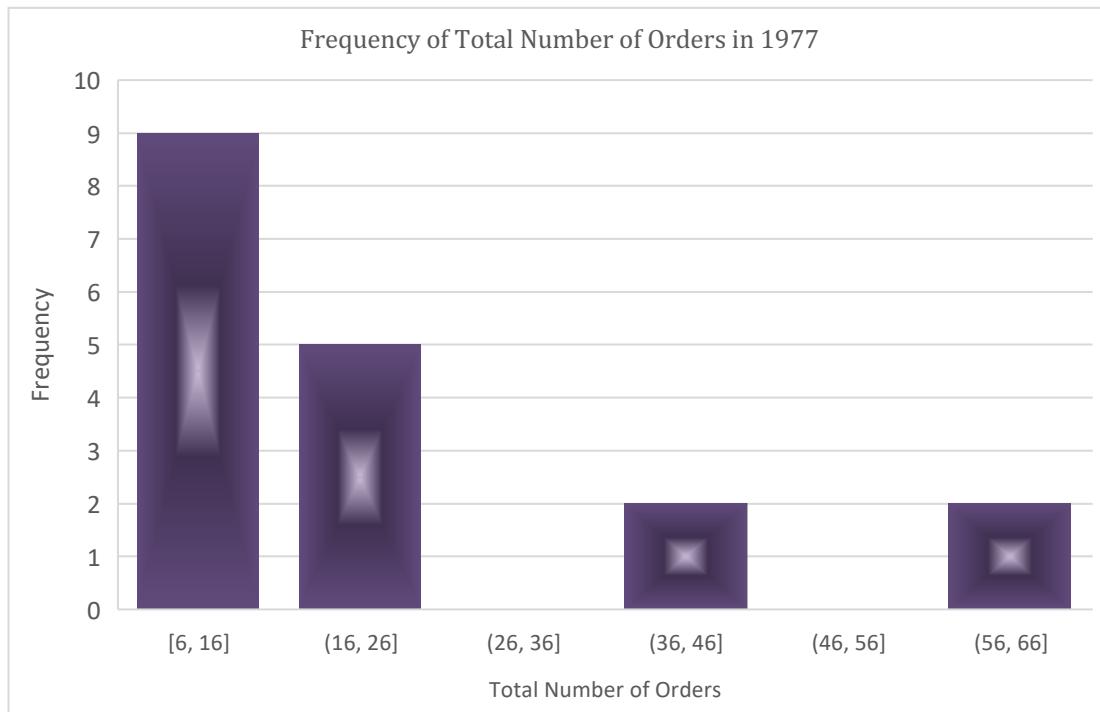


Figure 3: Frequency of the total number of orders (greater than 5) in the year of order date 1977.

Figure 3 informs the frequency of the total number of orders and the Logistic team wishes to retrospect their performance of the year 1977 and for the total number of orders greater than 5.

- In all, 9 out of 18 orders are from the total volume of the order's group [6 – 16], 5 from [16 – 26] and just 2 each from the total volume of the order's group [36 -46] and [56 – 66].
- There are no orders within the total orders group [26 – 36] and [46 – 56].
- The total number of all the orders is less than 66.



Figure 4: Shipping country's names in the year of order data 1997, with their total number of orders (> 5) and average days between the order shipping (≥ 3 and < 20 days).

The above figure 4 shows the names of the shipping countries which took more than or equal to 3 but less than 20 average days from the date order was placed until shipping was completed. And countries from which more than 5 total orders were booked.

- Almost all the countries took less than 10 average days to ship and deliver the order from its ordered date except the 4 countries.
- Only 4 countries set more than 10 average days, namely the UK - 10.42 average days for 26 total orders, the USA - 10.91 average days for the highest 62 total orders, Portugal - 11.25 average days for 6 and Ireland 11.45 average days for 11 orders.
- The UK with 62 and Germany with 60, top the rank for the maximum number of orders to be confirmed, followed by Brazil and France with 39 and 38 total orders respectively. Rest all countries have order scores less than 26.

3] The HR Team wants to know for each employee what was their age on the date they joined the company and whom they currently report to.

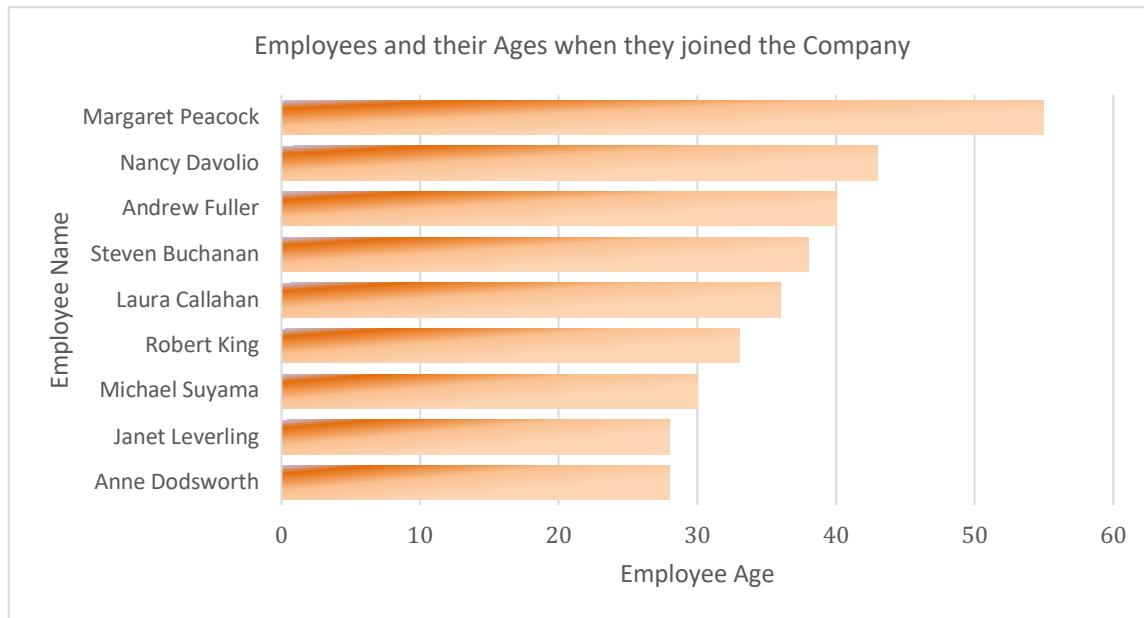


Figure 5: Employees and their ages when they joined the company.

Figure 5 displays the age of all the employees, the HR team likes to know at the time when they started working with the company.

- Only one worker named Margaret Peacock was above 50 years when he/she joined the company.
- Rest all employees were below 45 years of age when they commenced working with the company.
- Overall the organisation was mixed of employees of age groups ranging from 28 to 55 years.

Employee Full Name	Employee's Job Title	Employee's Age	Employee's Tenure	Manager Full Name	Manager's Title
Anne Dodsworth	Sales Representative	28	27	Steven Buchanan	Sales Manager
Janet Leverling	Sales Representative	28	30	Andrew Fuller	Vice President, Sales
Michael Suyama	Sales Representative	30	28	Steven Buchanan	Sales Manager
Robert King	Sales Representative	33	28	Steven Buchanan	Sales Manager
Laura Callahan	Inside Sales Coordinator	36	28	Andrew Fuller	Vice President, Sales
Steven Buchanan	Sales Manager	38	28	Andrew Fuller	Vice President, Sales
Andrew Fuller	Vice President, Sales	40	30		
Nancy Davolio	Sales Representative	43	30	Andrew Fuller	Vice President, Sales
Margaret Peacock	Sales Representative	55	29	Andrew Fuller	Vice President, Sales

Figure 6: Employee's and their manager's information along with their age when they joined the company and tenure.

The above figure 6 explains the HR team with requested details such as the names of employees and their current manager with the job title along with tenure and age when they joined the company.

- All the employees currently either report to Andrew Fuller who is the Vice President, of Sales or Steven Buchanan, the Sales Manager.
- 6 out of 9 workers are Sales Representatives and 1 is an Inside Sales Coordinator.
- Andrew Fuller, Janet Leverling and Nancy Davolio are working with the company for the longest term over 30 years. However, the rest of the employees are with the company for over 27 years.

4] The Logistics Team wants to do a retrospection of their global performances over 1996-1997, in order to identify for which month they perform well.

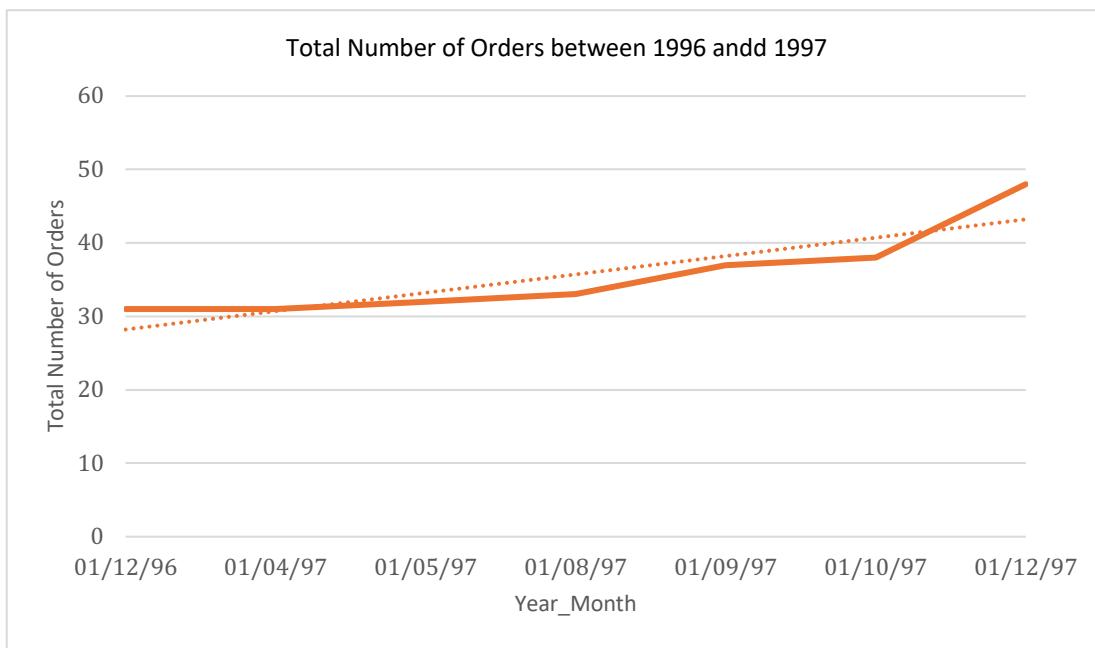


Figure 7: Total number of orders greater than 20 between 1996 and 1997 year.

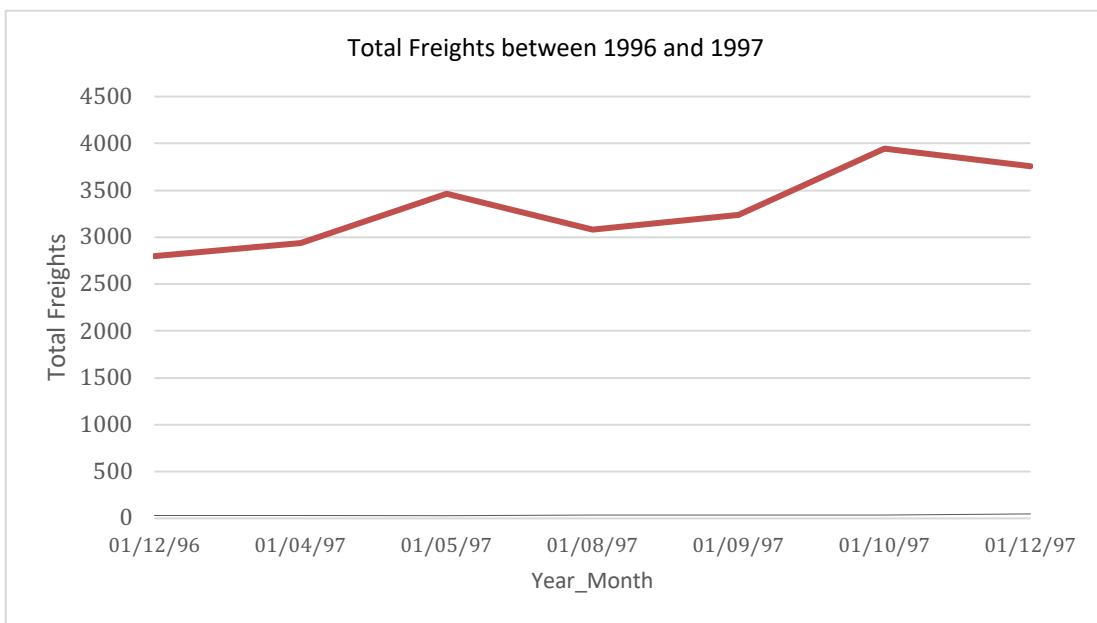


Figure 8: Total freights greater than 2500 between 1996 and 1997 year.

The above figures 7 and 8, show the total orders greater than 20 and total freights greater than 2500 for months between 1996 – 1997 for which the Logistics team wants to assess their performance.

- Over the period of 1 year, the total orders (which are more than 20), increased from an initial 31 in December 1996 to 48 by December 1997.
- For the initial period of 5 months, the order total remained the same, thereby increasing steadily for the rest of the year and scored the highest 48 by end of 1997.
- Total freights started at over 2500 in December 1996 and recorded slightly more than 3500 by the end of 1997.
- The total freight showed consistent progress with some exceptions over the timeframe of 1 year.
- Overall both the total orders (more than 20) and total freights (more than 2500) endure improved scores and performance between 1996 - 1997.

5] The Pricing Team wants to know which products had an unit price increase and the percentage increase was not between 10% and 30%.

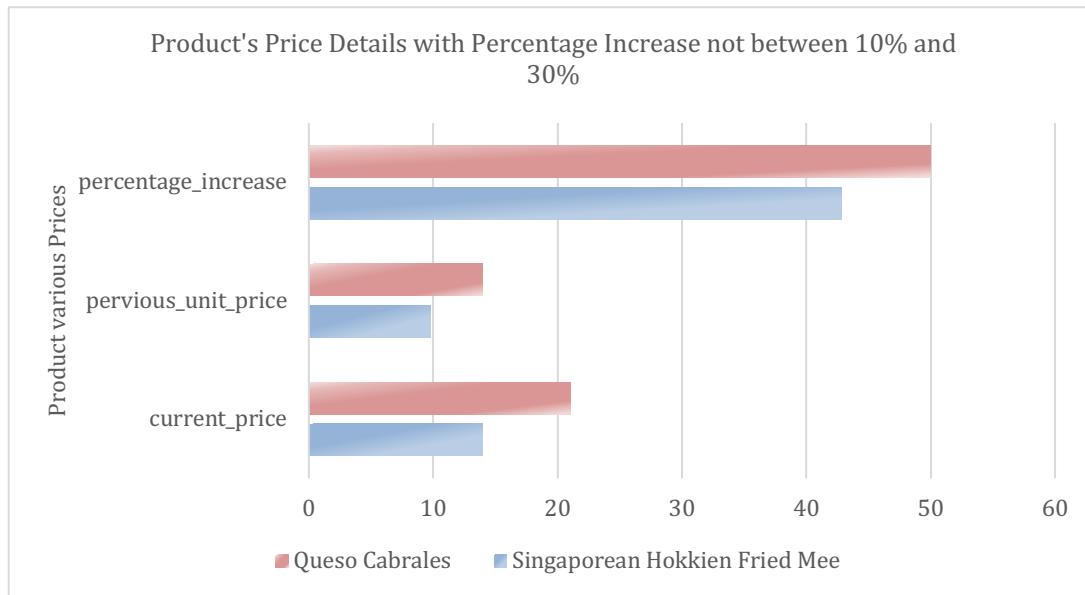


Figure 9: Product's price details with percentage increase not between 10% and 30%.

Product Name	Current Unit_Price	Pervious Unit_Price	Percentage Increase
Singaporean Hokkien Fried Mee	14	9.8	42.8571
Queso Cabrales	21	14	50

Figure 10: Product's current and previous prices with their percentage increase not between 10% and 30%.

Figures 9 and 10 display, the names of the products with their current price and previous price based on the date the order was placed and accordingly their percentage increased over this period, not between 10% and 30% which the Product team is interested to examine.

- Queso Cabrales item's earlier and latest price was greater than Singaporean Hokkien Fried Mee in both instances and thus its percentage increase is more than the other.
- The price of both products has increased significantly between their initial and current order bookings. Singaporean Hokkien Fried Mee product recorded over 42% increase and Queso Cabrales, doubled to a 50% increase.
- These are only two food items for which a per cent increase is not between 10% and 30%.
- However, over the period of the initial and current order, both these items' unit price is still below \$25.

6] The Pricing Team wants to know how each category performs according to their below price range.

1. Below \$10
2. \$10 - \$20
3. \$20 - \$50
4. Over \$50

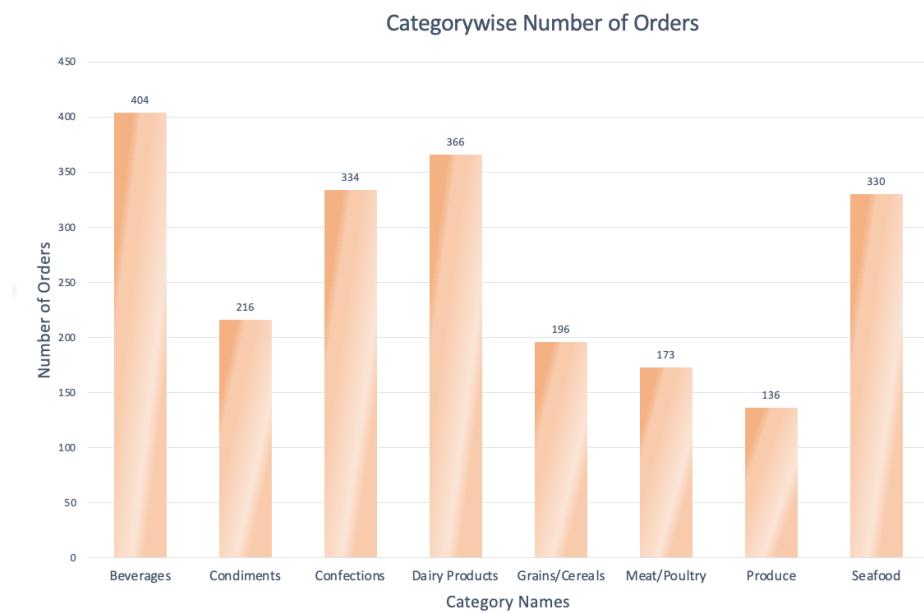


Figure 11: Category-wise total number of orders.

The above figure 11 shows the total orders for all the categories of food products.

- The highest performance amount of orders totalling 404 was confirmed for the beverages category, followed by 366 for dairy products, 330 for seafood, and 334 for confections.
- The smallest volume of orders 136 was placed for the produce category of food items.
- For the rest category of the products namely meat/poultry, grains/cereals and condiments the total orders ranged from 135 to 220 which informed average performance in contrast to others.

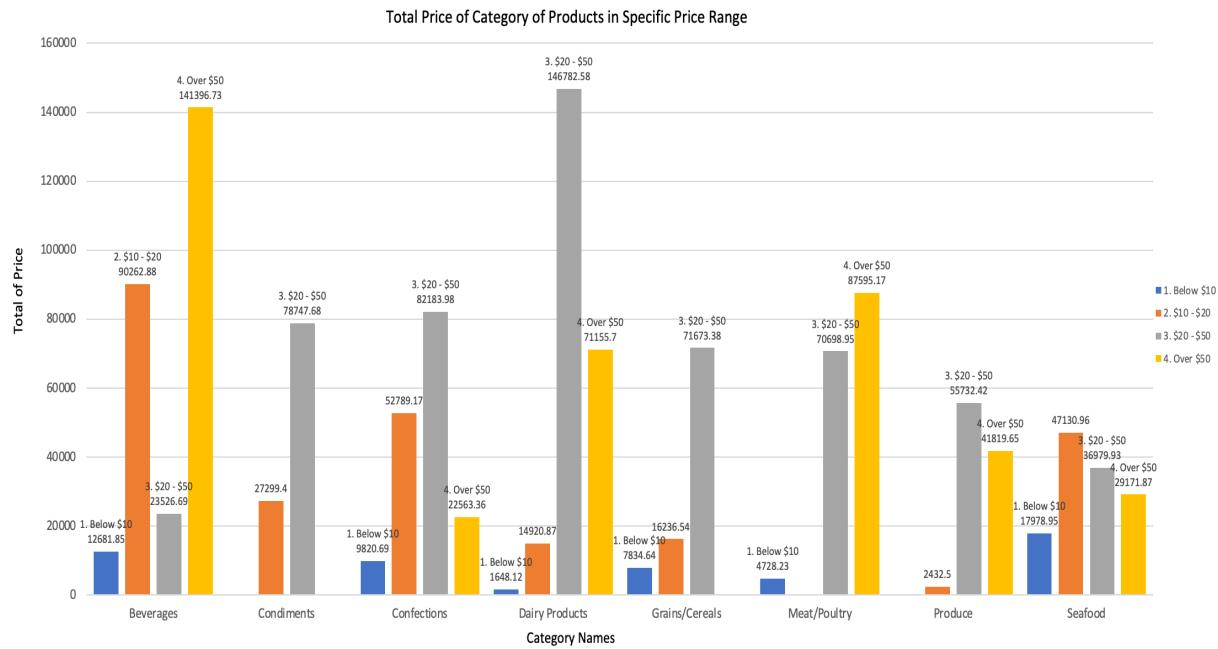


Figure 12: Total price of the category of products in a specific price range.



Figure 13: Total amount of categories including discount.

Figures 12 and 13 present the total price of the categories in the specified price range and the total amount of sales of categories including the discount offered for which the Pricing team need to conduct their assessment.

- The two price ranges in ranking - first, \$20 - \$50 in the beverages category and second, over \$50 in dairy products achieved the highest results as compared to other price ranges and categories.
- The lowest was noted for meat/poultry for \$20 - \$50 price range with zero, kept on with a slight increase for dairy products in Below \$10 and produce in \$10 -\$20 range.
- Similarly, beverages and dairy products top the list of the total amount of sales including discounts given.
- Confections, meat/poultry and seafood categories showed a marginally higher volume of sales compared to the remaining categories' condiments, grains/cereals and produce.

7] The Logistics Team wants to know what is the current state of our regional suppliers' stocks for each category of product. And their supplier region" as:

1. "America"
2. "Europe"
3. "Asia"
4. "Oceania"

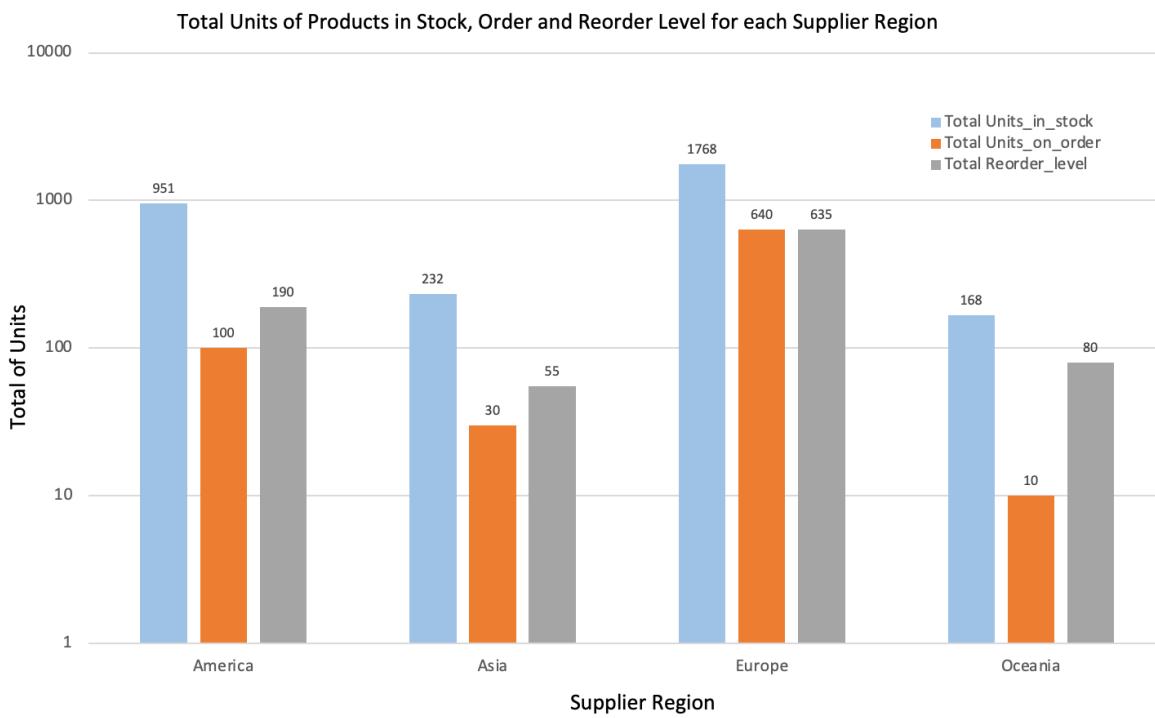


Figure 14: Total units of products in stock, on order and reorder level for each supplier region.

The above figure 14 displays the total units of products for 3 groups i) stock, ii) on order and reorder level for supplier regions requested by the Logistics team.

- Across all the regions total units of the products in stock are higher than on-order and reorder levels.

- For the Europe region, almost the same units of products were recorded on order and reorder level. Whereas for the rest 3 regions, America, Asia and Oceania, remarkably higher units were reordered compared to on-order product units.
- In all, the highest units of products were maintained in stock and for on-order and reorder levels in the Europe region which informs its better state than other regions.

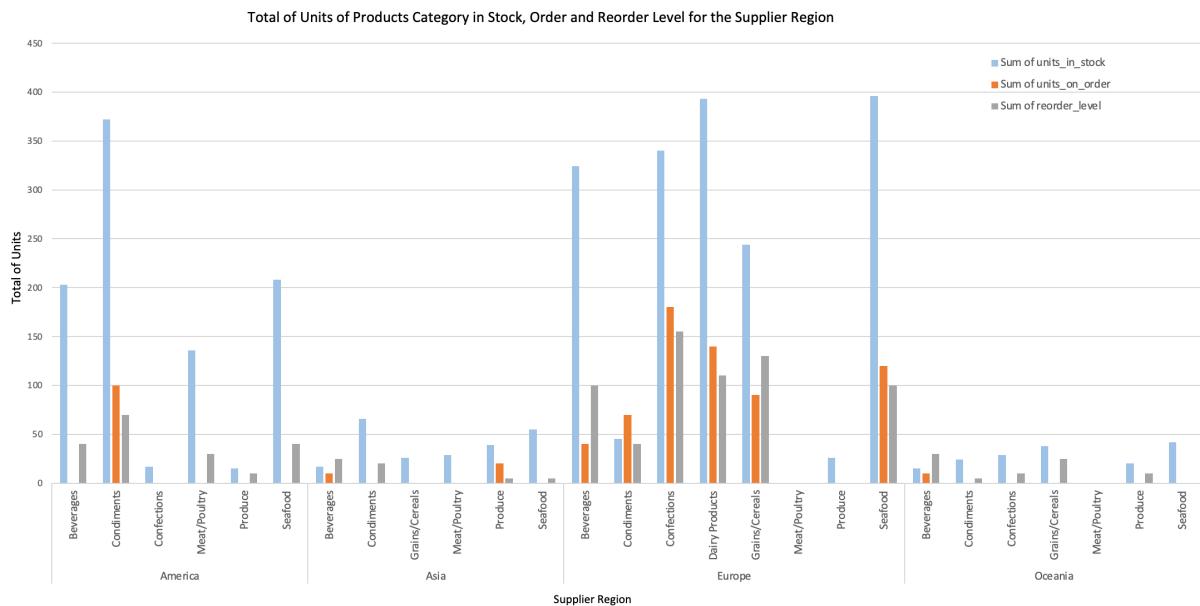


Figure 15: Total of units of products category in stock, on-order, and reorder level for the supplier region.

Figure 15 emphasises the total units of categories of products for 3 groups i) stock, ii) on order and reorder level for specific supplier regions given by the Logistics team in order to determine their state.

- Oceania had minimal units of categories of products recording less than 50 total units for all groups – in stock, on-order and reorder level as compared to other regions.
- Europe listed the greatest units of products for all the groups except for a few categories namely meat/poultry with no units and few units for produce just in stock.
- For the condiments category of products, America reported noticeably better performance than Europe. And the produce category recorded the lowest for all regions.
- In all, Europe and America stated a stronger position than Asia and Oceania.

8] The Pricing Team wants to know for each currently offered product how their unit price compares against their categories average and median unit price. And their position against the category average and median unit price as:

1. "Below Average"
2. "Average"
3. "Over Average"

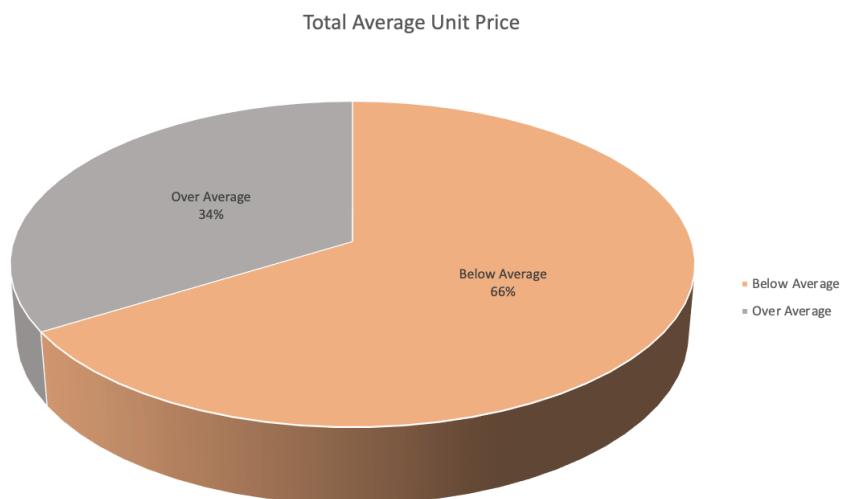


Figure 16: Total average price.

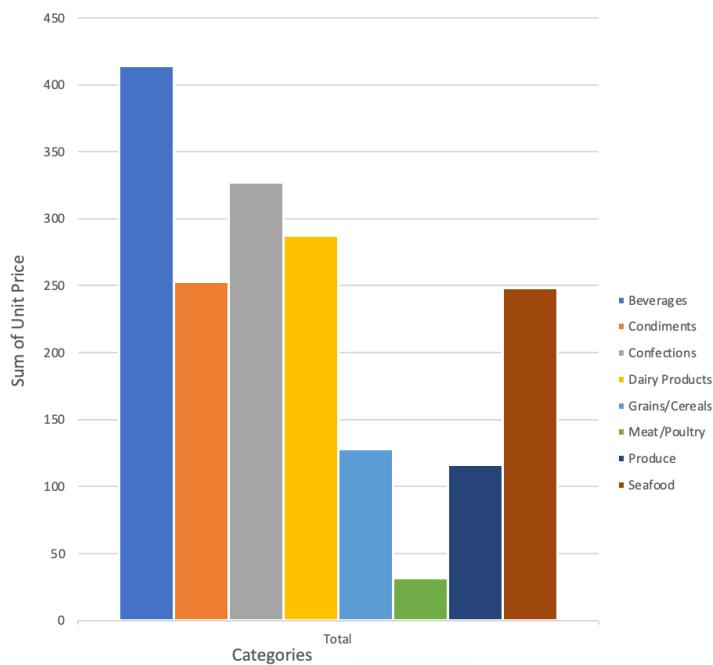


Figure 17: Categories and their sum of unit price.

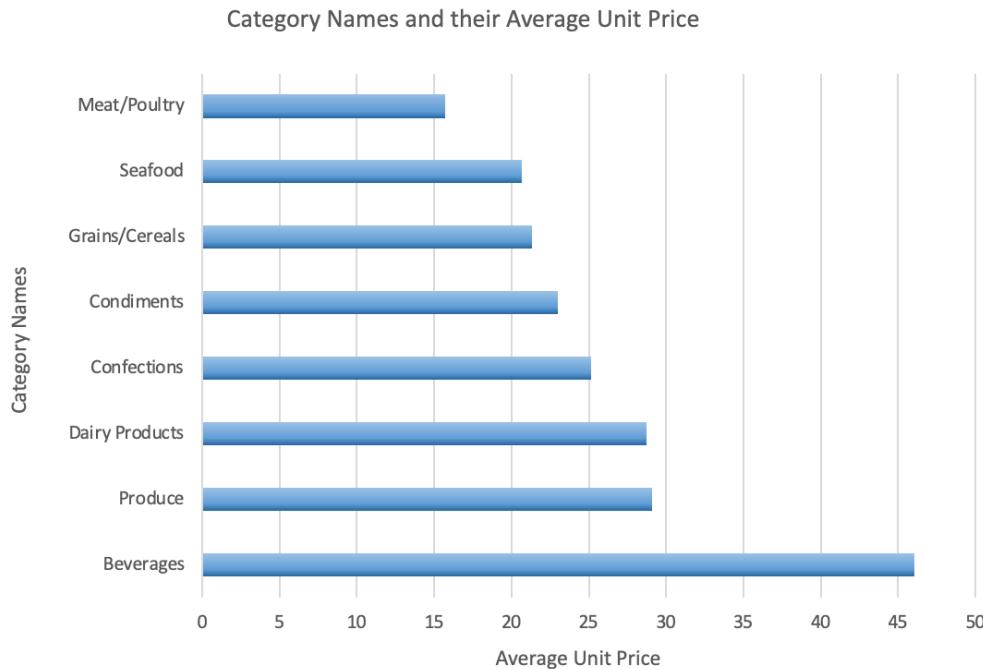


Figure 18: Category names and their average unit price.

The above figures display categories of products and it's the total average price and other aggregation details of all the currently available product categories for the Pricing team to access.

- Of the currently available products, 34% of product categories are over average and 66% below average.
- The total unit price of the beverages category shows the highest amount as compared to others. Whereas, meat/poultry is at the lowest level.
- Also, similar, beverages and meat/poultry are at an extreme level on the scale of average unit price.
- The produce and dairy products category shows the almost same average price and the remaining category's average price is below 25.

9] The Sales Team wants to build a list of KPIs to measure employees' performances.

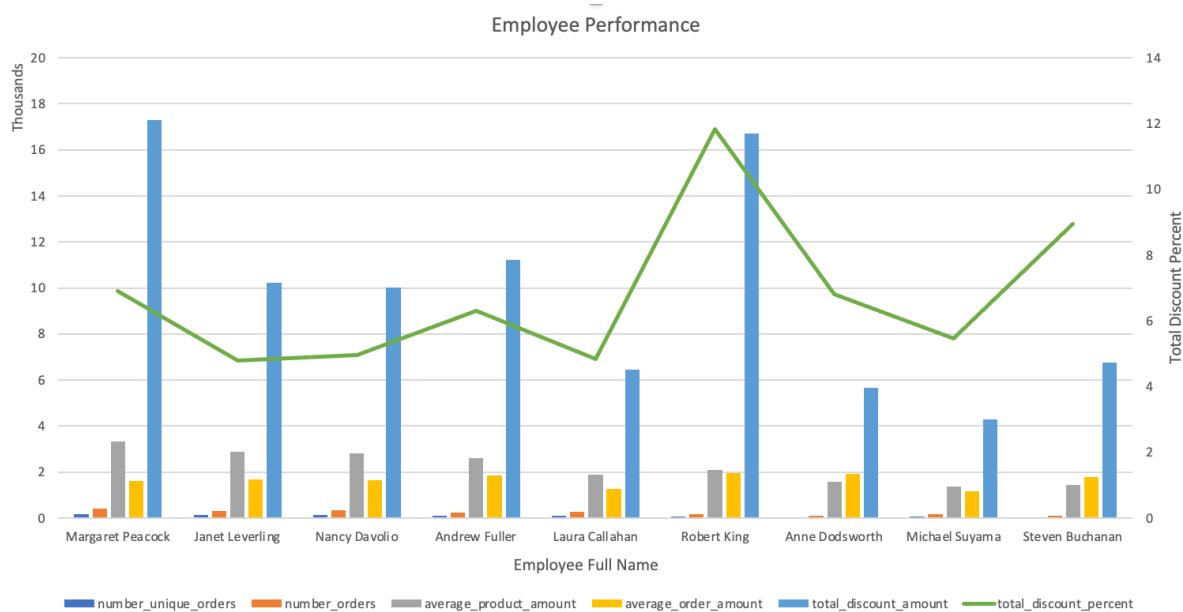


Figure 19: Employee's performance.

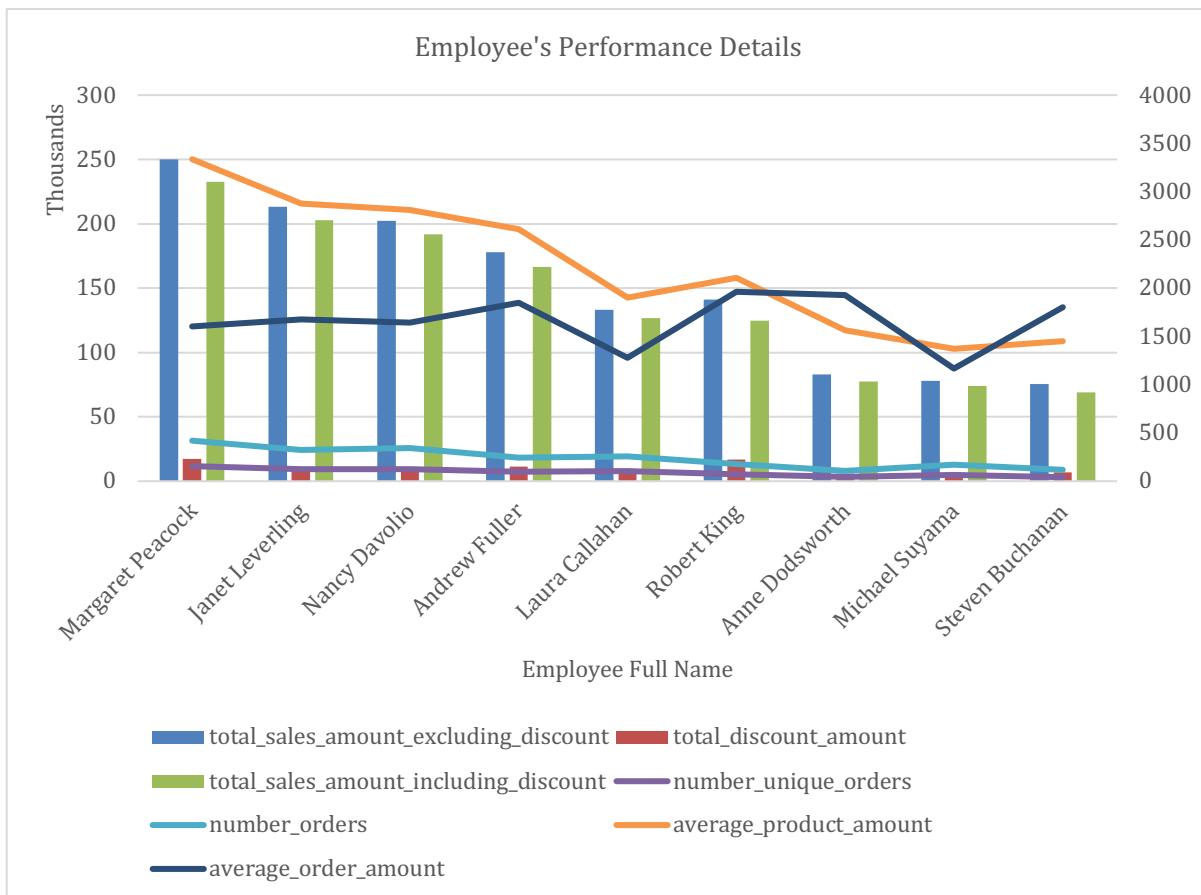


Figure 20: Employee's performance details.

The above figure illustrates the employee's performance that the Sales team want to access. It features various performance indicators.

- Irrespective of the discount, employee Margaret Peacock tops the rank and has outstanding performance with over 230 thousand in total sales amount. The next in rank are Janet Leverling in 2nd position and Nancy Davolio in 3rd with both over 200 thousand in total sales amount.
- The least performers are Steven Buchanan, Michael Suyama and Anne Dodsworth, all with total sales amount in between 68 and 77 thousand.
- The performance indicator with respect to i) the total number of unique orders, ii) the total number of orders and iii) the average product amount directly corresponds to the total sales amount including/excluding discount indicator throughout the employees.
- For the total discount amount, two employees Margaret Peacock and Robert King outperform others by offering over 16 thousand total discounts.

Employee Full Name	Number Unique Orders	Total Discount Amount	Total Sales Amount Including Discount	Total Discount Percent
Margaret Peacock	156	17296.61	232890.84	6.91
Janet Leverling	127	10238.45	202812.84	4.8
Nancy Davolio	123	10036.11	192107.6	4.96
Andrew Fuller	96	11211.51	166537.75	6.3
Laura Callahan	104	6438.76	126862.27	4.83
Robert King	72	16727.76	124568.23	11.83
Anne Dodsworth	43	5655.93	77308.06	6.81
Michael Suyama	67	4284.97	73913.12	5.47
Steven Buchanan	42	6775.47	68792.28	8.96

Figure 21: Additional employee performance details.

The above figure provides information about employee performance so the Sales team can evaluate workers on additional attributes.

- In terms of the number of distinct orders handled, the top 3 employees are Margaret Peacock (156), Janet Leverling (127) and Nancy Davolio (123).
- Robert King addressed 72 distinct orders, on the other hand, he rendered the highest total discount of 11.83%.
- Though Andrew Fuller and Laura Callahan noted fewer orders compared to the top 3 performers in distinct orders, their total sales are noticeably higher over 120 thousand.

10] The Sales Team wants to build another list of KPIs to measure employees' performances across each category.

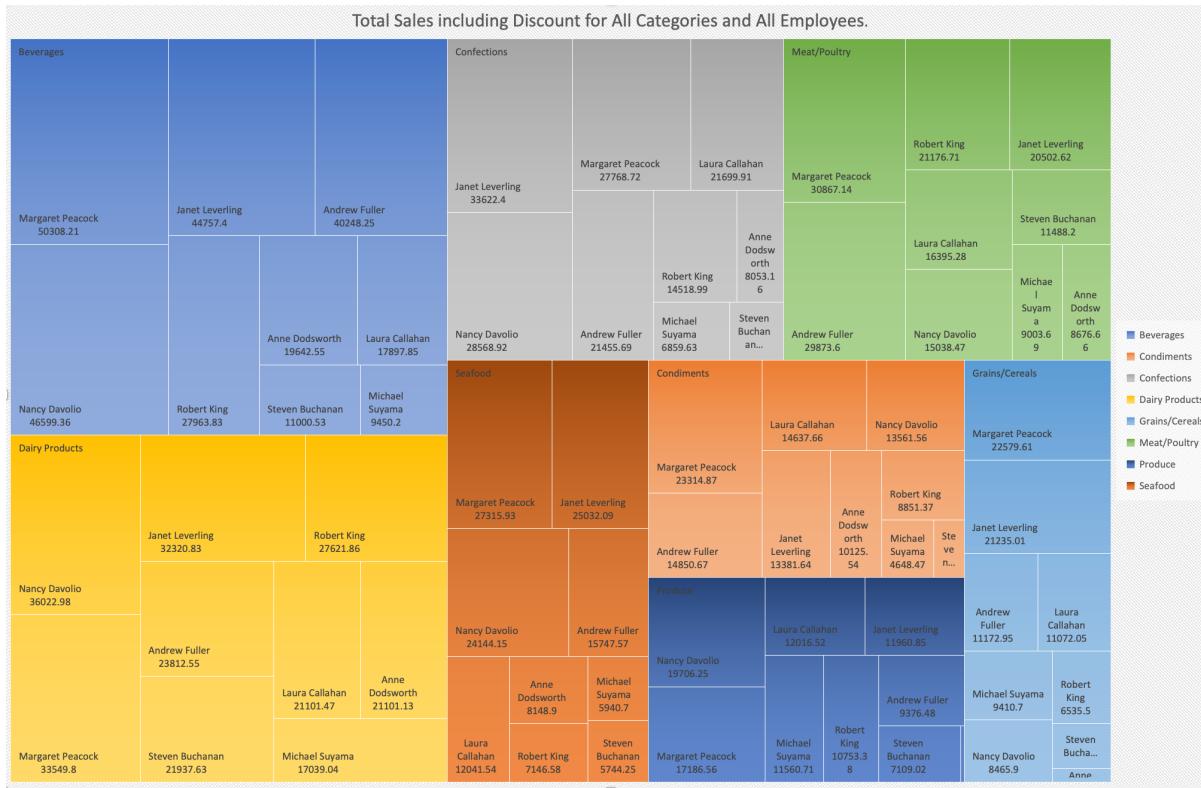


Figure 22: Total sales including a discount for all categories and all employees.

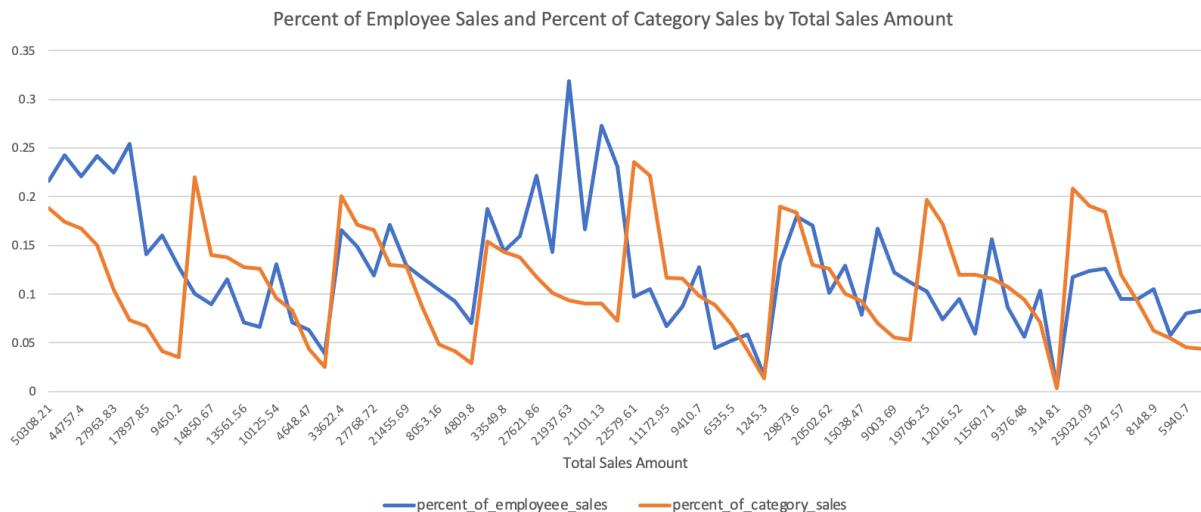


Figure 23: Percent of employee sales and present of category sales by total sales amount.

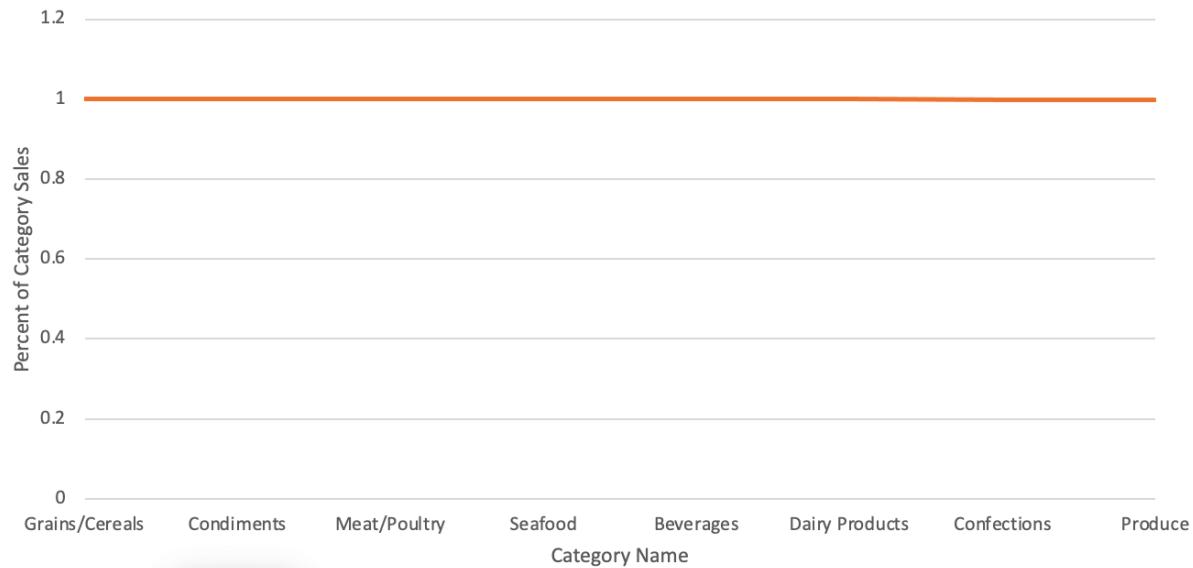


Figure 24: Sum of the percent of category sales across categories.

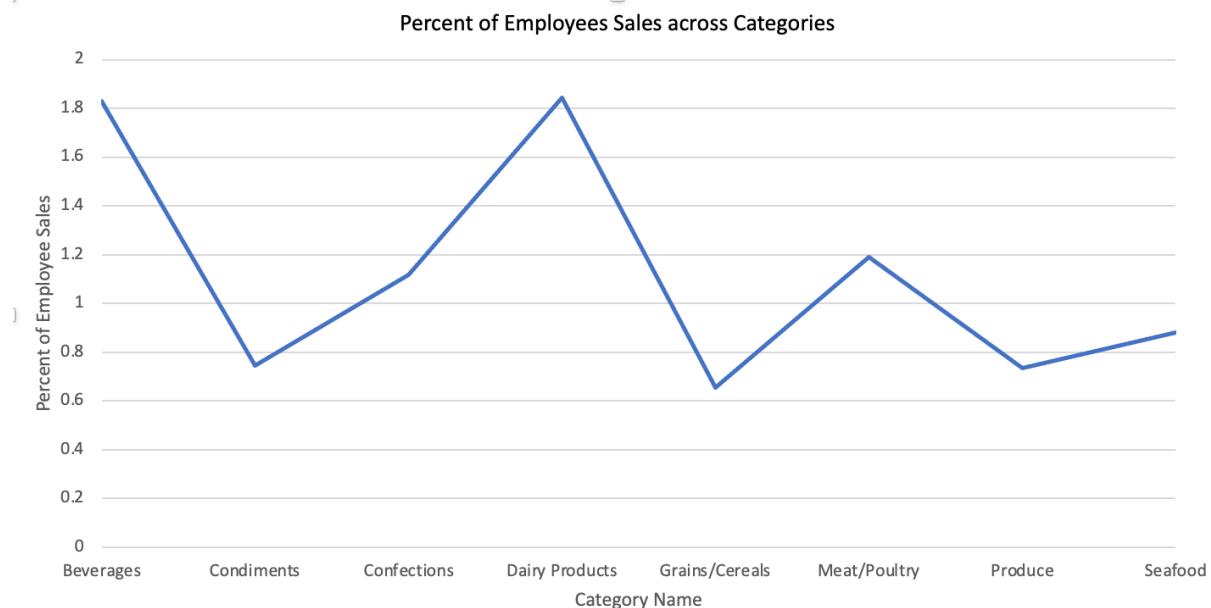


Figure 25: Sum of the percent of employees' sales across categories.

The above figures illustrate total sales including a discount for all categories and all employees, the percent of employee sales and category sales by total sales amount and the sum of the percent of the category, and employee sales across categories as the Sales team needed to prepare a list of employee performance across categories.

- Margaret Peacock ranks at the top of the total sales amount considering discounts in the majority of categories including beverages, condiments, grains/cereals, meat/poultry and seafood (i.e. 5 out of 8 categories).
- For beverages in 2nd position it's Nancy Davolio and in 3rd place Andrew Fuller. Similarly, for condiments, it's Andrew Fuller and Laura Callahan.
- Nancy Davolio, Margaret Peacock and Laura Callahan are the top 3 performers for the category confections.
- In the grains/cereals category, Janet Leverling is in 2nd position and Andrew Fuller is in 3rd.
- For dairy products, the top 3 rankers are Nancy Davolio, Janet Leverling and Robert King.
- Robert King is in 2nd place and Janet Leverling at 3rd in the meat/poultry category.
- The produce category has Nancy Davolio, Margaret Peacock and Laura Callahan as the top total sales workers.
- For seafood, Janet Leverling is in 2nd position and Nancy Davolio is in 3rd.
- The sum of the percent of category sales is relatively even across the categories. However, the sum of the percent of employee sales across categories shows relevant fluctuations.