FINAL REPORT OF NORTHWIND DATABASE

Simon Lim

24661225
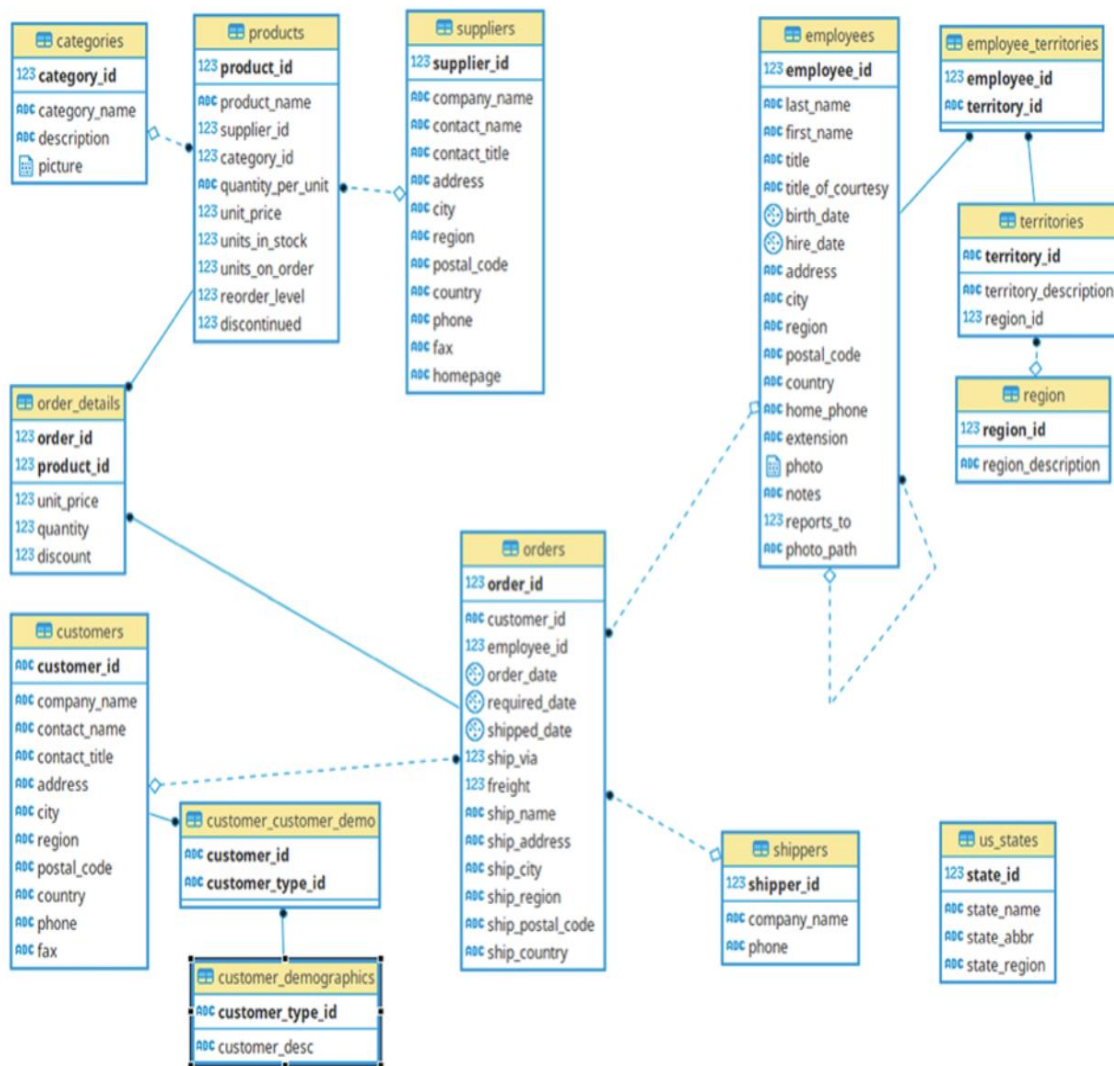
Data Science Practice Assignment 2

1 PRESENTATION OF DATA

The dataset "Northwind" contains 14 tables, including categories, customers, customer_customer_demo, customer_demographics, employees, employee_territories, orders, order_details, products, region, shippers, suppliers, territories, us_states.

- Categories: list of product categories
- Customers: buyers of Northwind products
- Customer_customer_demo & customer demographics: customers detail, including customer id
- Employee_territories: the territory employee is working in and employee id
- Employees: people who work for Northwind
- Orders: information about each order such as customer who placed the orders, employee who handled the order, and shipment date etc.
- Order details: transaction details of each order such as products, unit price and quantity, discount applied
- Products: the products Northwind buys and sells
- Region: information about region
- Suppliers: who supply products to Northwind
- Shippers: companies who ship products from Northwind to customers
- Us_states: details of USA states

1.1. The diagram of the relationships between tables

categories
- 123 **category_id**
- ABC category_name
- ABC description
- picture

products
- 123 **product_id**
- ABC product_name
- 123 supplier_id
- 123 category_id
- ABC quantity_per_unit
- 123 unit_price
- 123 units_in_stock
- 123 units_on_order
- 123 reorder_level
- 123 discontinued

suppliers
- 123 **supplier_id**
- ABC company_name
- ABC contact_name
- ABC contact_title
- ABC address
- ABC city
- ABC region
- ABC postal_code
- ABC country
- ABC phone
- ABC fax
- ABC homepage

employees
- 123 **employee_id**
- ABC last_name
- ABC first_name
- ABC title
- ABC title_of_courtesy
- birth_date
- hire_date
- ABC address
- ABC city
- ABC region
- ABC postal_code
- ABC country
- ABC home_phone
- ABC extension
- photo
- ABC notes
- 123 reports_to
- ABC photo_path

employee_territories
- 123 **employee_id**
- ABC **territory_id**

territories
- ABC **territory_id**
- ABC territory_description
- 123 region_id

region
- 123 **region_id**
- ABC region_description

order_details
- 123 **order_id**
- 123 **product_id**
- 123 unit_price
- 123 quantity
- 123 discount

customers
- ABC **customer_id**
- ABC company_name
- ABC contact_name
- ABC contact_title
- ABC address
- ABC city
- ABC region
- ABC postal_code
- ABC country
- ABC phone
- ABC fax

customer_customer_demo
- ABC **customer_id**
- ABC **customer_type_id**

customer_demographics
- ABC **customer_type_id**
- ABC customer_desc

orders
- 123 **order_id**
- ABC customer_id
- 123 employee_id
- order_date
- required_date
- shipped_date
- 123 ship_via
- 123 freight
- ABC ship_name
- ABC ship_address
- ABC ship_city
- ABC ship_region
- ABC ship_postal_code
- ABC ship_country

shippers
- 123 **shipper_id**
- ABC company_name
- ABC phone

us_states
- 123 **state_id**
- ABC state_name
- ABC state_abbr
- ABC state_region

1.2 Table for dimensions and number of samples

| Tables	Dimensions	Number of samples
Categories	- Category_id - Category_name - Description - Picture	8
Customer_customer_demo	- Customer_id - Customer_type_id	-
Customer_demographics	- Customer_type_id - Customer_desc	-
Customers	- Customer_id - Company_name - Contact_name - Contact_title - Address - City - Region - Postal_code - Country - Phone - fax	91
Employees	- Employee_id - Last_name - First_name - Title - Title_of_courtesy - Birth_date - Hire_date - Address - City - Region - Postal_code - Country - Homephone - Extension - Photo - Notes - Reports_to - Photo_path	9

Employee_territories	- Employee_id - Territory_id	49
Orders	- Order_id - Customer_id - Employee_id - Order_date - Required_date - Shipped_date - Ship_via - Freight - Ship_name - Ship_address - Ship_city - Ship_region - Ship_postal_code - Ship_country	830
Order_details	- Order_id - Product_id - Unit_price - Quantity - Discount	2155
Products	- Product_id - Product_name - Supplier_id - Category_id - Quantity_per_unit - Unit_price - Unit_in_stock - Unit_on_order - Reorder_level - Discontinued	77
Region	- Region_id - Region_description	4
Shippers	- Shipper_id - Company_name - Phone	6
Suppliers	- Supplier_id - Company_name - Contact_name - Contact_title - Address - City - Region - Postal_code - Country - Phone - Fax	29

	- Home_page	
Territories	- Territory_id - Territory_description - Region_id	53
Us_states	- State_id - State_name - State_abbr - State_region	51

2 INSTRUCTIONS TO CONNECT TO DATABASE

1. Download the Northwind dataset, including tables and sample data on Canvas (I downloaded Northwind dataset, tables and samples on https://docs.yugabyte.com/preview/sample-data/northwind/).
2. In DBeaver, create to postgresSQL database, with localhost and port (5432).
3. Create Schemas named Northwind.
4. Create 14 tables with given columns and set a primary key and foreign keys for each table (shown in diagram).
5. Insert sample data for each table (make sure to insert reference tables first and then insert tables, which include foreign keys, otherwise will make an error).
6. Start running queries to answer business questions.

3 ANSWERING BUSINESS QUESTIONS

1. There are 52 products that are currently being offered for a price range between $10 to $50 with the lowest price of $10 for "Sir Rodney's Scones" and the highest price of $49.3 for "Tarte au sucre".

Products that are offered between $10 and $50
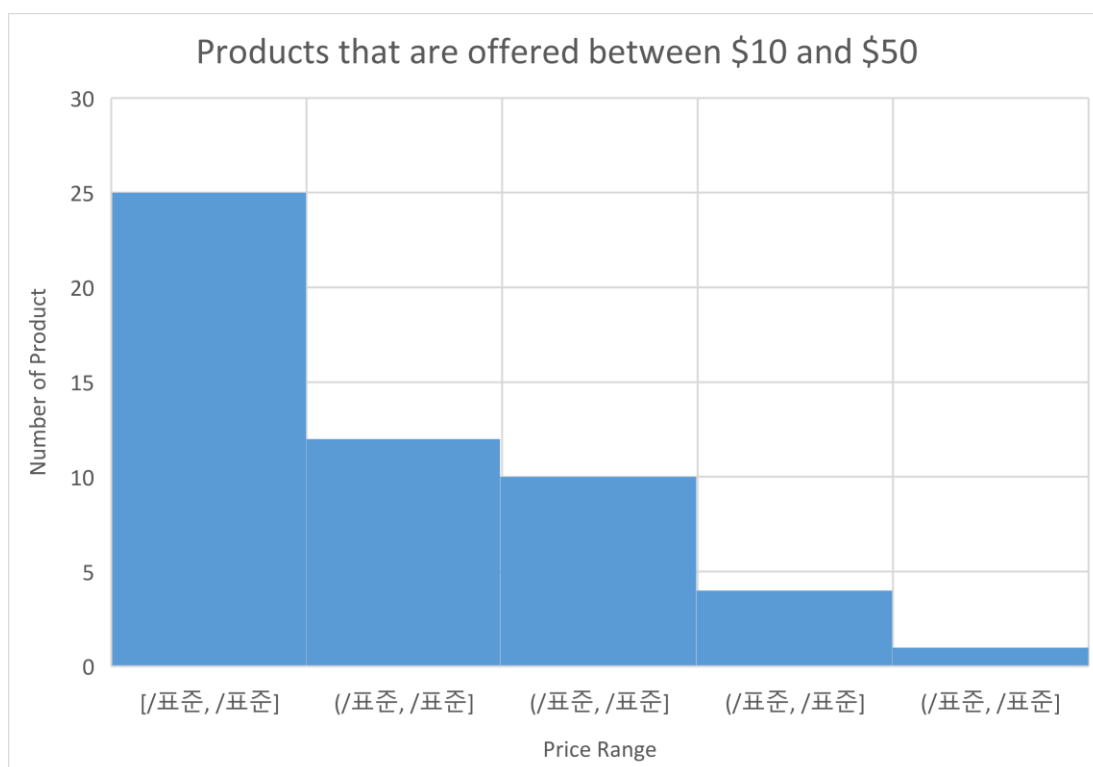
2. The Logistics Team wants to do a retrospection of their performances for the year 1997, in order to identify for which countries they didn't perform well. There are 18

countries, in which Northwind products were ordered and shipped in 1997. Accordingly, the lowest number of orders in 1997 is Argentina, which had only 6 orders, with average 5.67 days between the order date and the shipping date.

Number of orders and average days in 1997
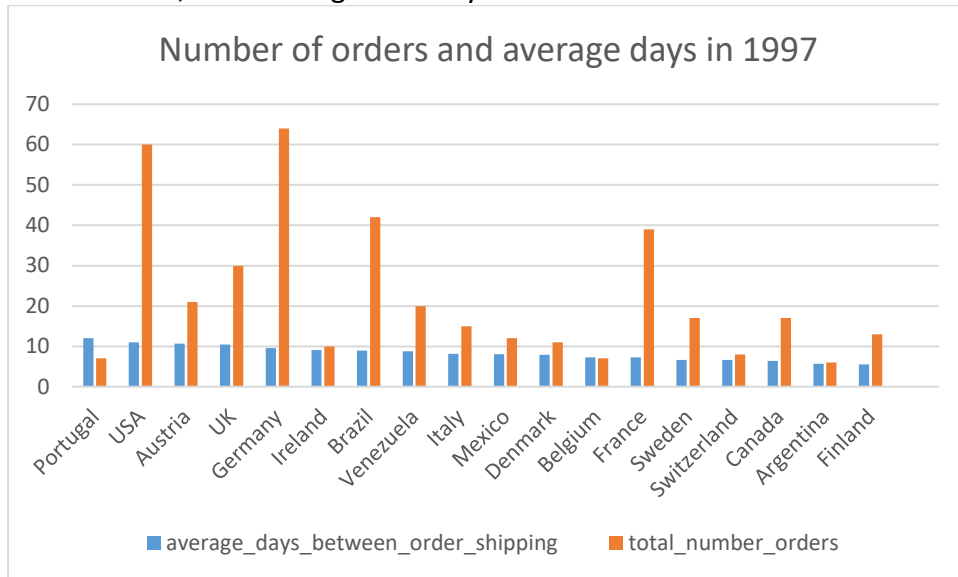
3. The HR Team wants to know for each employee what was their age on the date they joined the company and who they currently report to. There are 9 employees in Northwind and here is a table for each employee with the age they were hired and the name of their manager.

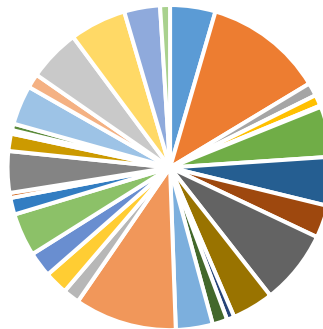| Employee name	The age employee was hired	Name of the manager
Anne Dodsworth	29	Steven Buchanan
Janet Leverling	29	Andrew Fuller
Michael Suyama	30	Steven Buchanan
Robert King	34	Steven Buchanan
Laura Callahan	36	Andrew Fuller
Steven Buchanan	39	Andrew Fuller
Andrew Fuller	40	None
Nancy Davolio	43	Andrew Fuller
Margaret Peacock	56	Andrew Fuller

4. The Logistics Team wants to do a retrospection of their global performance over 1996-1997, in order to identify which month they perform well. There were 7 months over 1996-1997, which are October 1997, December 1997, May 1997, September 1997, August 1997, April 1997 and December 1996. The highest number of orders was in December 1997.

total_number_orders

(Line chart — Y-axis: "Number of orders" ranging 0 to 60; X-axis: "Year-month" from 1/12/1996 through 1/12/1997)

5. The Pricing Team wants to know which products had an unit price increase and the percentage increase was not between 10% and 30%. Out of 80 products, there are only two products, "Queso Cabrales" (50% increase) and "Singaporean Hokkien Fried Mee" (42.86% increase), which had a unit price increase greater than 30%. Exceptionally, "Mozzarella di Giovanni" product had a unit price decrease about -20%. Other than that, all other products have no price change.
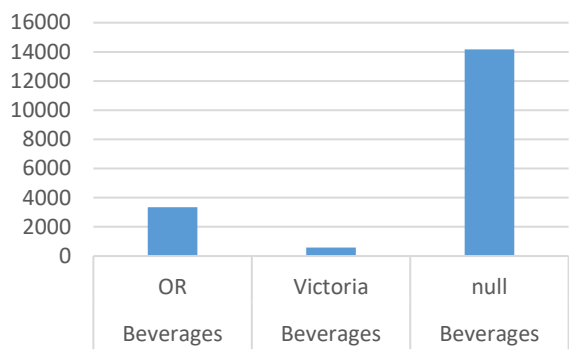
6. The Pricing Team wants to know how each category performs according to their price range. Here is a pie chart, describing total number of orders for different categories in different price range.

Total number of orders for each category in different price range

- Beverages 1. BELOW $10
- Beverages 2. $10 - $20
- Beverages 3. $20 - $50
- Beverages 4. OVER $50
- Condiments 1. BELOW $10
- Condiments 2. $10 - $20
- Condiments 3. $20 - $50
- Confections 1. BELOW $10
- Confections 2. $10 - $20
- Confections 3. $20 - $50
- Confections 4. OVER $50
- Dairy Products 1. BELOW $10
- Dairy Products 2. $10 - $20
- Dairy Products 3. $20 - $50
- Dairy Products 4. OVER $50
- Grains/Cereals 1. BELOW $10
- Grains/Cereals 2. $10 - $20
- Grains/Cereals 3. $20 - $50
- Meat/Poultry 1. BELOW $10
- Meat/Poultry 2. $10 - $20
- Meat/Poultry 3. $20 - $50
- Meat/Poultry 4. OVER $50
- Produce 1. BELOW $10
- Produce 2. $10 - $20
- Produce 3. $20 - $50
- Produce 4. OVER $50
- Seafood 1. BELOW $10
- Seafood 2. $10 - $20
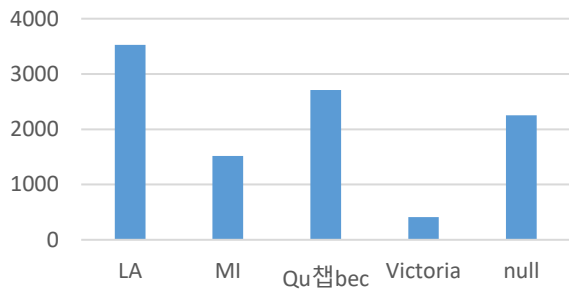- Seafood 3. $20 - $50
- Seafood 4. OVER $50

7. The Logistics Team wants to know what the current state of our regional suppliers' stocks for each category of product is. There are 8 categories, including beverages, condiments, confections, dairy products, grains/cereals, meat/poultry, produce and seafood. Here are the graphs for each category stocks in different state.
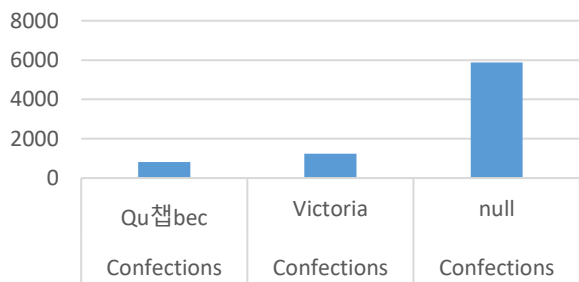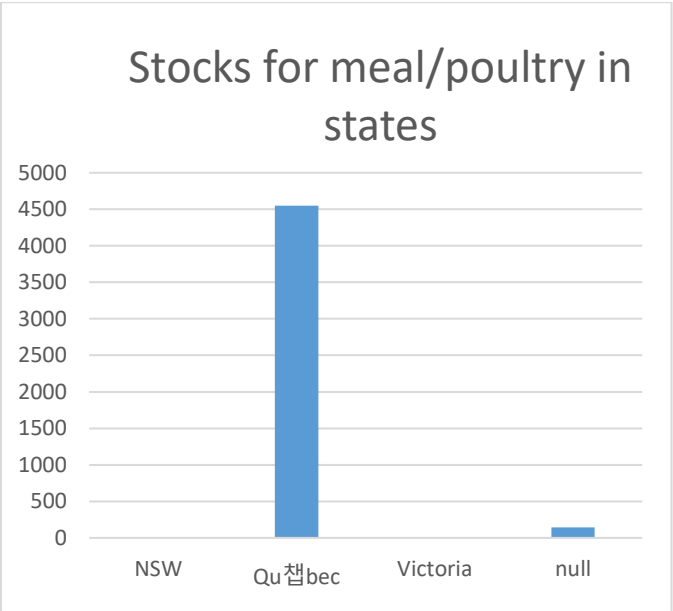
Stocks for beverages in states

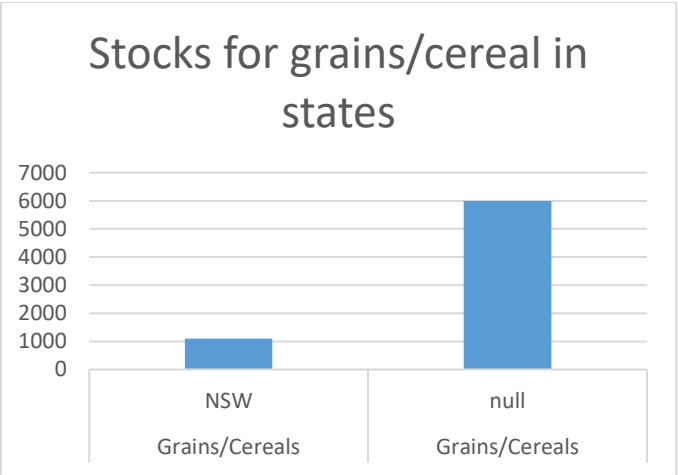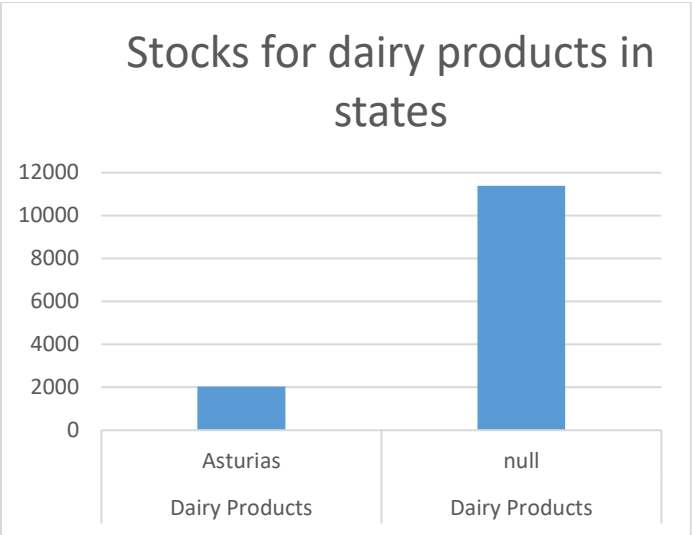|		
OR Beverages	Victoria Beverages	null Beverages

Y-axis: 0, 2000, 4000, 6000, 8000, 10000, 12000, 14000, 16000

Stocks for condiments in states

LA	MI	Qu챕bec	Victoria	null

Y-axis: 0, 1000, 2000, 3000, 4000

Stocks for confections in states

Qu챕bec Confections	Victoria Confections	null Confections

Y-axis: 0, 2000, 4000, 6000, 8000

Stocks for dairy products in states

	Asturias Dairy Products	null Dairy Products
12000		
10000		
8000		
6000		
4000		
2000	▮	
0		

Stocks for grains/cereal in states

	NSW Grains/Cereals	null Grains/Cereals
7000		
6000		
5000		
4000		
3000		
2000		
1000		
0		

Stocks for meal/poultry in states

	NSW	Qu챕bec	Victoria	null
5000				
4500				
4000				
3500				
3000				
2500				
2000				
1500				
1000				
500				
0				

Stocks for produce in states

	MI Produce	NSW Produce	null Produce

Stocks for produce in states

	MA Seafood	Victoria Seafood	null Seafood

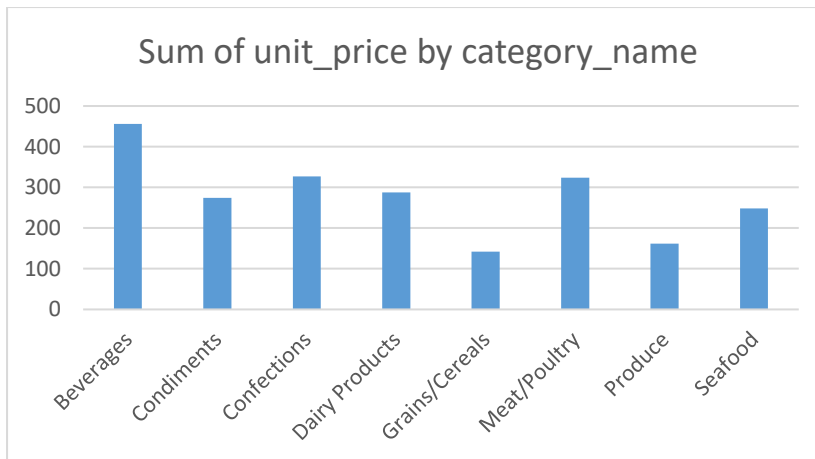8. The Pricing Team wants to know for each currently offered product how their unit price compares against their categories average and median unit price.

Sum of unit_price by category_name

Beverages: The average unit price is 37.98 and the median unit price is 18.

Condiments: The average unit price is 22.85 and the median unit price is 21.05.

Confections: The average unit price is 25.16 and the median unit price is 16.25.

Dairy Products: The average unit price is 28.73 and the median unit price is 32.

Grains/Cereals: The average unit price is 20.25 and the median unit price is 19.5.

Meat/Poultry: The average unit price is 54.01 and the median unit price is 32.8.

Produce: The average unit price is 32.37 and the median unit price is 30.

Seafood: The average unit price is 20.68 and the median unit price is 15.

Sum of median_per_category by category_name
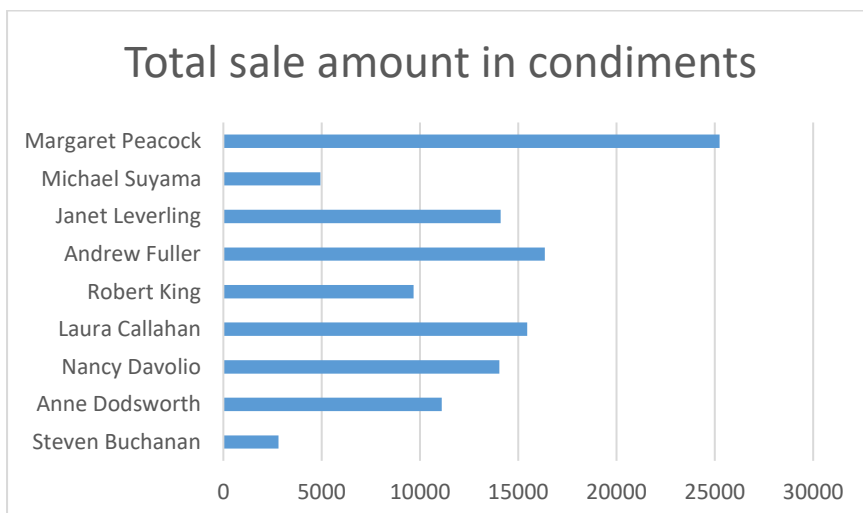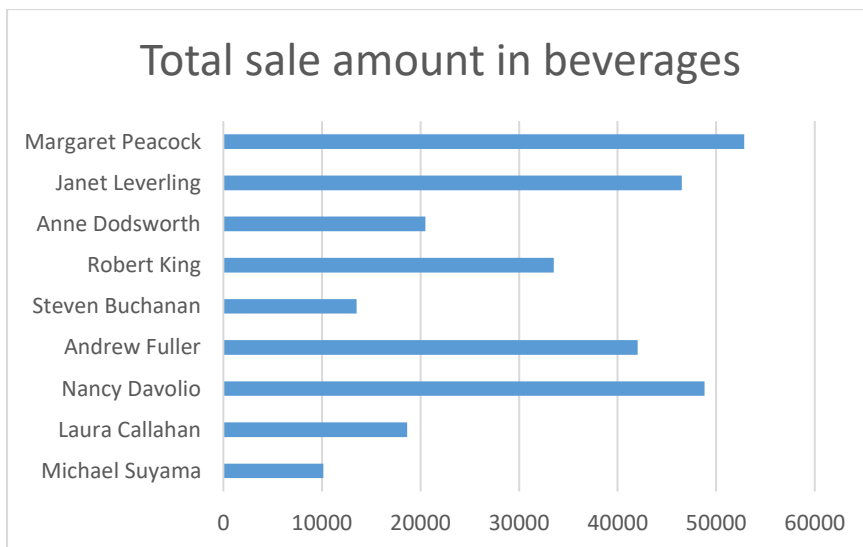
Average_unit_price per unit price
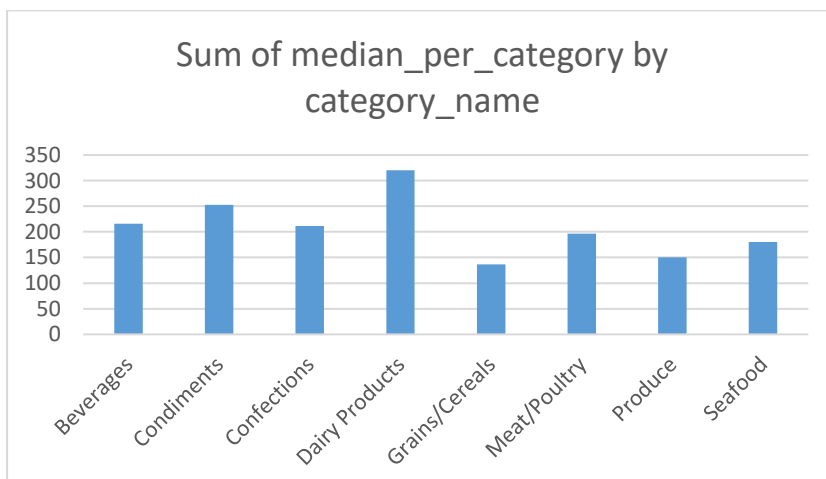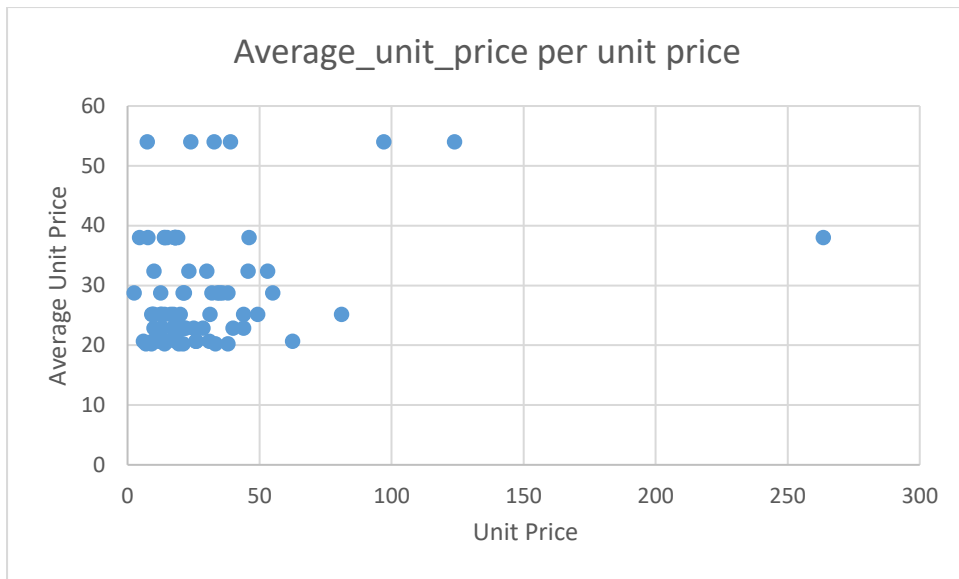
9. The Sale Teams wants to build a list of KPIs to measure employees' performances. There are 9 employees in total and the graphs show total sale amount excluding discount and average order amount.

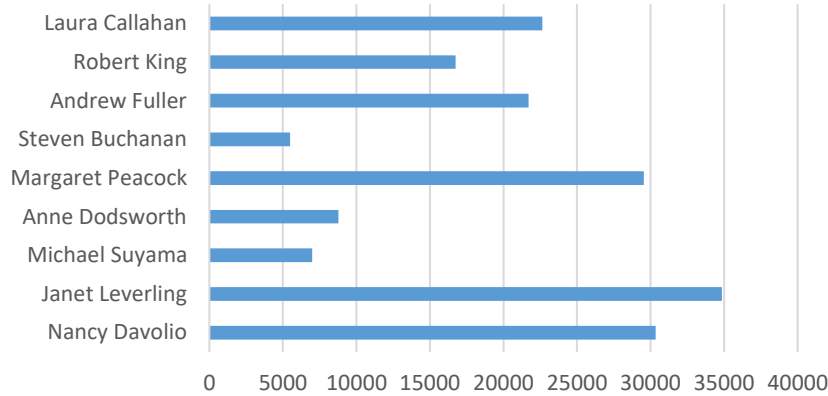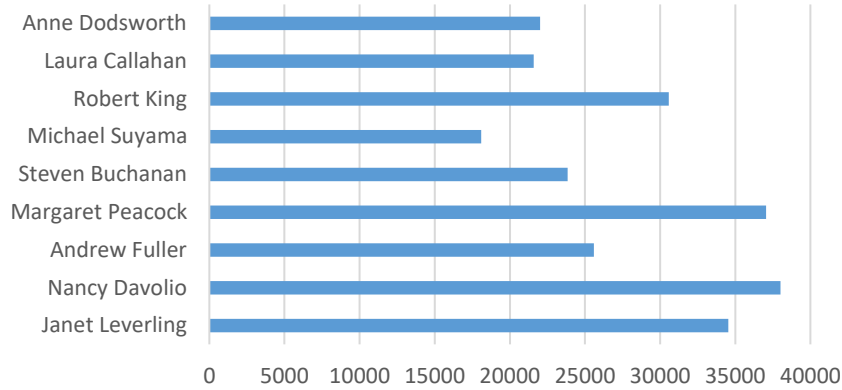total_sale_amount_excluding_discount

average_order_amount
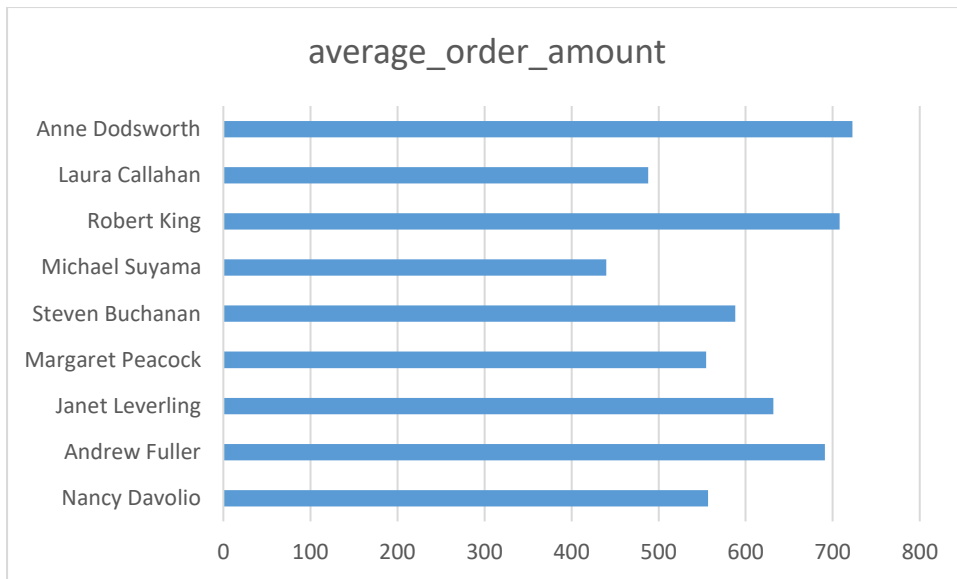
10. The Sales Team wants to build another list of KPIs to measure employees' performance across each category. Here are the graphs of employee's total sale amount of each category.
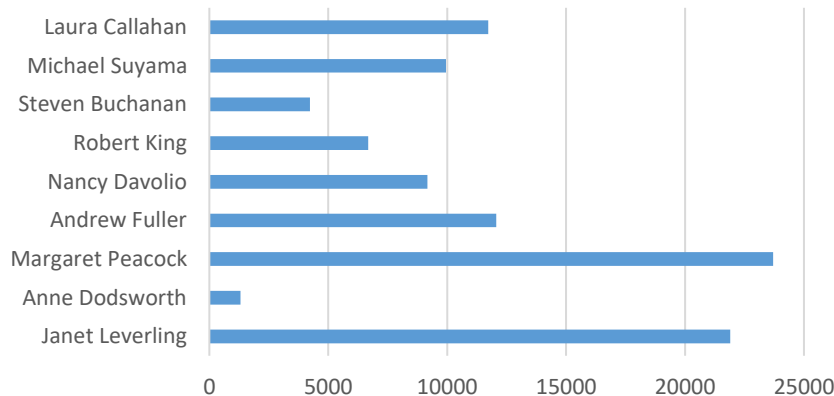

Total sale amount in beverages

Total sale amount in condiments

Total sale amount in confections

Salesperson	
Laura Callahan	~23000
Robert King	~16500
Andrew Fuller	~22000
Steven Buchanan	~4500
Margaret Peacock	~29000
Anne Dodsworth	~9000
Michael Suyama	~7500
Janet Leverling	~35000
Nancy Davolio	~30000

(x-axis: 0, 5000, 10000, 15000, 20000, 25000, 30000, 35000, 40000)

Total sale amount in dairy product

Salesperson	
Anne Dodsworth	~23000
Laura Callahan	~22000
Robert King	~30000
Michael Suyama	~18000
Steven Buchanan	~24000
Margaret Peacock	~38000
Andrew Fuller	~26000
Nancy Davolio	~38500
Janet Leverling	~34000

(x-axis: 0, 5000, 10000, 15000, 20000, 25000, 30000, 35000, 40000)

Total sale amount in grains/cereal

Salesperson	
Laura Callahan	~12000
Michael Suyama	~10000
Steven Buchanan	~4000
Robert King	~6500
Nancy Davolio	~9500
Andrew Fuller	~12000
Margaret Peacock	~23500
Anne Dodsworth	~1500
Janet Leverling	~21500

(x-axis: 0, 5000, 10000, 15000, 20000, 25000)

Total sale amount in meat/poultry

Employee	Amount
Anne Dodsworth	~10000
Laura Callahan	~17000
Andrew Fuller	~32500
Robert King	~25000
Michael Suyama	~9000
Nancy Davolio	~16000
Janet Leverling	~20500
Steven Buchanan	~12000
Margaret Peacock	~33500

(x-axis: 0, 5000, 10000, 15000, 20000, 25000, 30000, 35000, 40000)
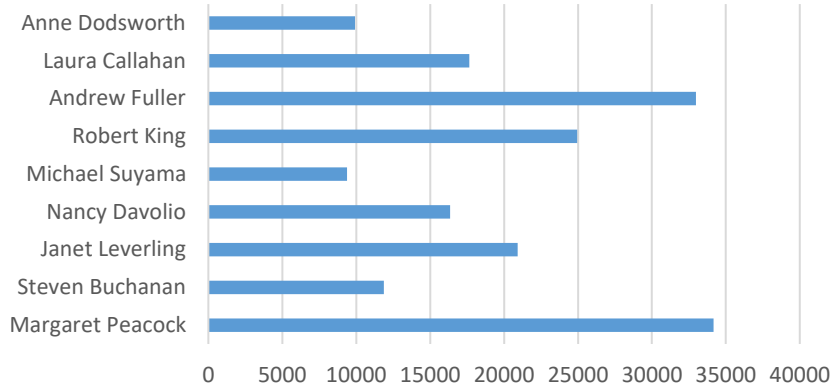
Total sale amount in produce

Employee	Amount
Robert King	~11500
Anne Dodsworth	~300
Janet Leverling	~13000
Laura Callahan	~12000
Steven Buchanan	~7000
Margaret Peacock	~18000
Michael Suyama	~12500
Andrew Fuller	~9000
Nancy Davolio	~20500

(x-axis: 0, 5000, 10000, 15000, 20000, 25000)
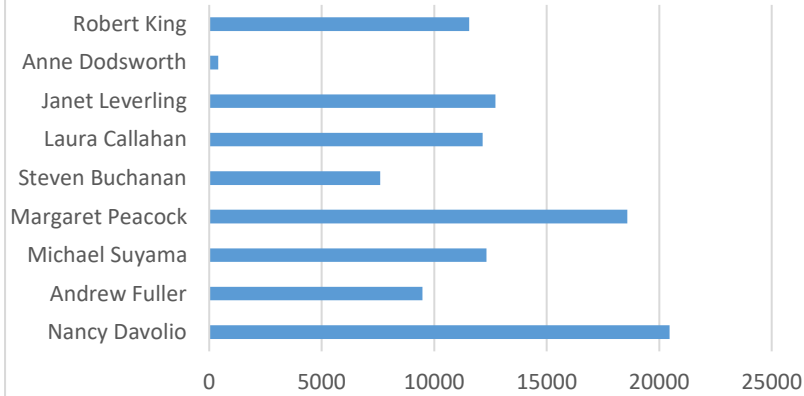
Total sale amount in seafood

Employee	Amount
Robert King	~7500
Janet Leverling	~28000
Anne Dodsworth	~9000
Nancy Davolio	~25000
Laura Callahan	~14000
Margaret Peacock	~29500
Michael Suyama	~6000
Steven Buchanan	~6000
Andrew Fuller	~17500

(x-axis: 0, 5000, 10000, 15000, 20000, 25000, 30000, 35000)

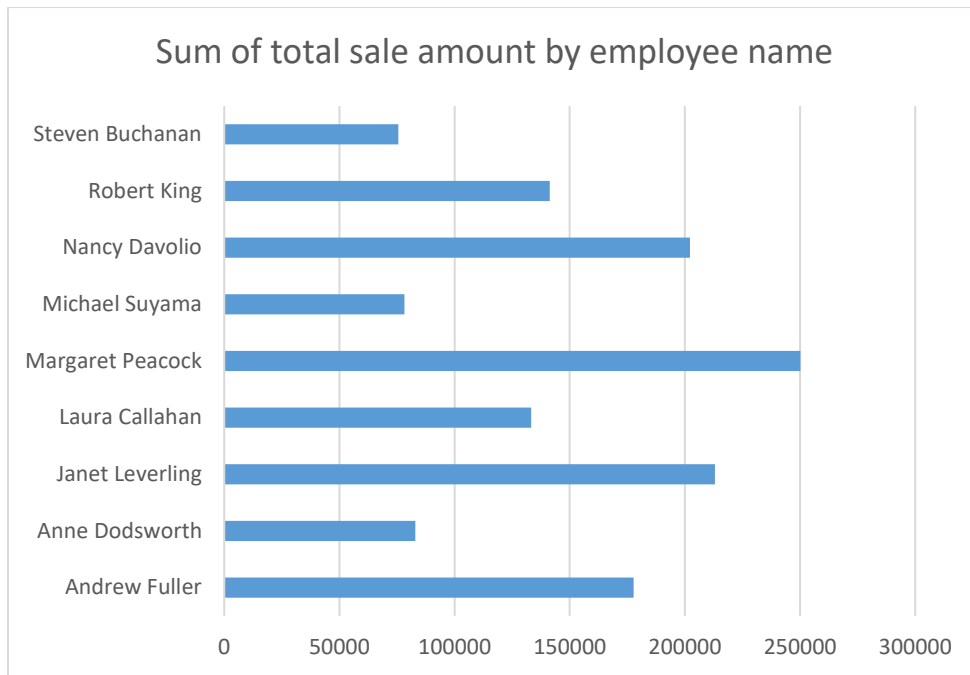Sum of total sale amount by employee name
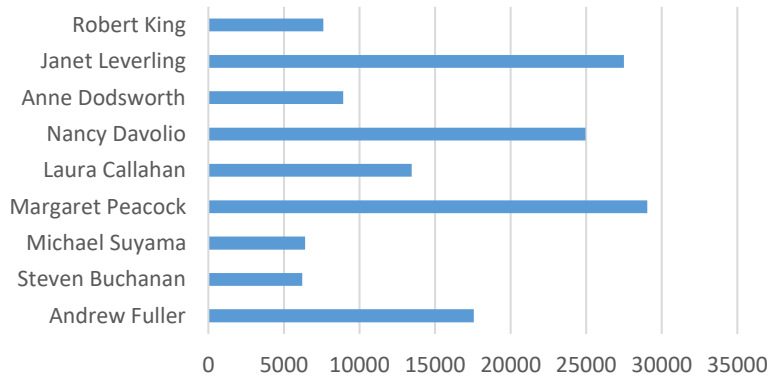
4 REFERENCE

Northwind sample database. Documentation. (n.d.). Retrieved October 8, 2022, from
https://docs.yugabyte.com/preview/sample-data/northwind/