Instructions to Connect to Northwind Database

To access the Northwind database, download PostgreSQL packages or installers from https://www.postgresql.org/download/ for the user's operating system, following the prompts to finish setup to completion. Similarly, complete this for DBeaver from https://dbeaver.io/download/. Configure a Postgres user with administrator privileges by creating a user and a corresponding password on server "localhost" (i.e. user's computer), server 5432 during setup and use this information to create a new connection for Postgres in DBeaver.

Once this is established, copy the whole text from the northwind.postgres_renamed.sql file into a script and execute it. This will load the data into DBeaver and connect the user to the Northwind Database.

All raw SQL output generated for the requested tasks are formatted as csv files attached in same zip. These are numbered corresponding to their task number.

Task 1

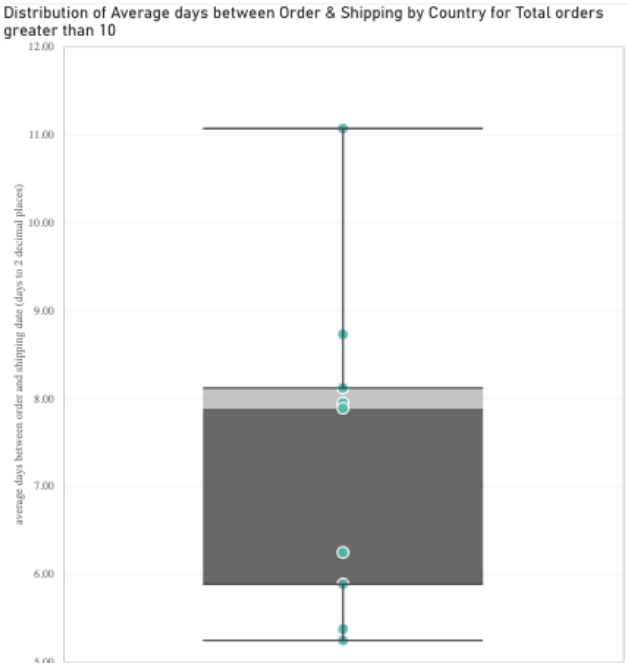
Product team task requirement for reviewing pricing strategy was to display the name and price for products between the price of $20 and $50 that are not discontinued, showing the higher priced items first. All required information was obtained from the products table filtered and ordered on the above criteria. The minimum unit price is $20.00, and the maximum unit price is $49.30. The mean price is $31.13, median price is $31.00 and the middle 50 percent of prices are between $23.63 and $37.00, showing a skew towards the lower end of the requested price range.

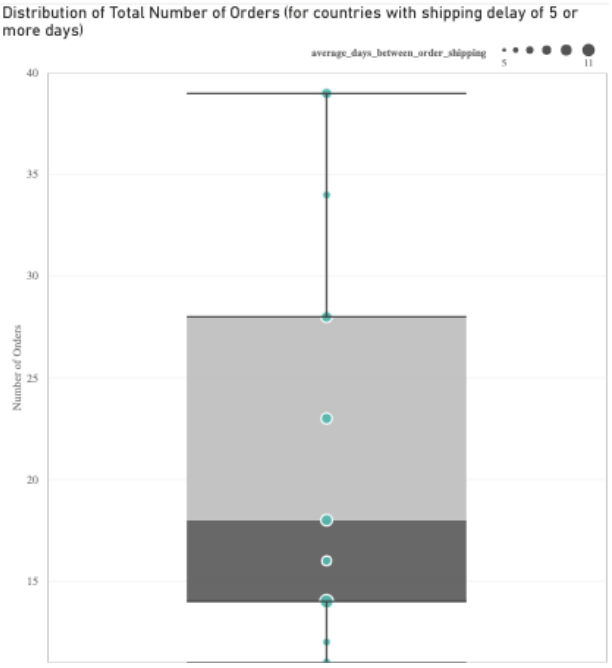
Task 2

The task goal was to identify countries in which the logistics team's performance was poor in 1998 from orders data. To achieve this, countries were returned alphabetically with average days between order and shipping date greater than or equal to 5 days and total orders more than 10. Ship country was extracted from orders along with average days between order and shipping as the averaged difference between shipped and order date as days rounded to 2 decimal places. Total number orders were calculated from the count of ship country. The poorest performance was for customers in Sweden at 11.07 days whereas the best performance was for Spanish customers at 5.3 days, for countries with total orders greater than 10.

The average days (across all countries) between order and shipping is 7.39 days. The median of average days between order and shipping date across the examined countries was 7.89 days, with the middle 50 percent of countries having between 5.89 and 8.12 days. The average number of orders for the countries explored was 22 with the middle 50 percent of number of orders being between 14 and 28

orders. The maximum and minimum total number of orders were 11 and 39 orders.



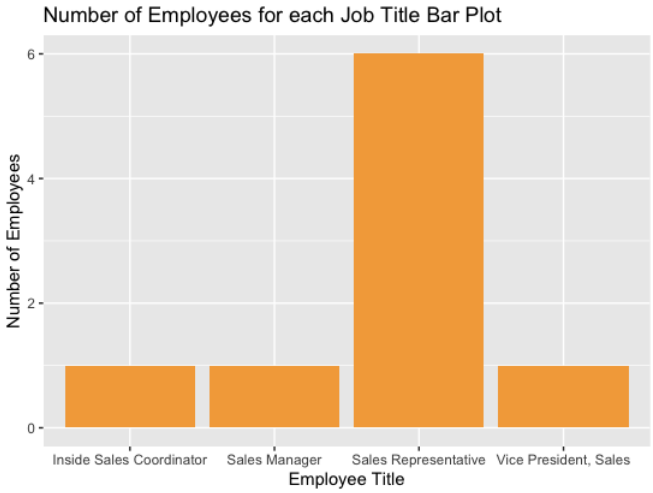


Task 3

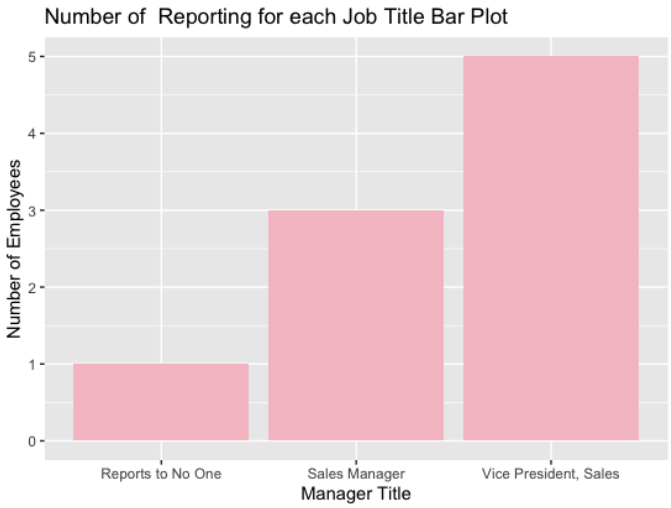
This HR task objective was to observe the employee reporting structure of the organization with employees' age. The requested output included employees' full names, job title, age at time of hire, their manager's full name and title from employees data: these should then be ordered by youngest age and employee name alphabetically. Employees full names were generated from the combination of first and last names, employee title from title and employee age at hire as the difference between birth dates and hire dates. Information for managers were generated from the self full join of employees to align employee reports to with manager id. This allowed additional manager full names to be generated from the combination of first and last names and titles from title. These were limited to entries to only those with employee full name entries.

The youngest hire was Janet Leverling at 28 years 7 months and 2 days and the oldest was Margaret Peacock at 55 years 7 months and 14 days.

There were the most sales representatives and an equal number of other job titles.

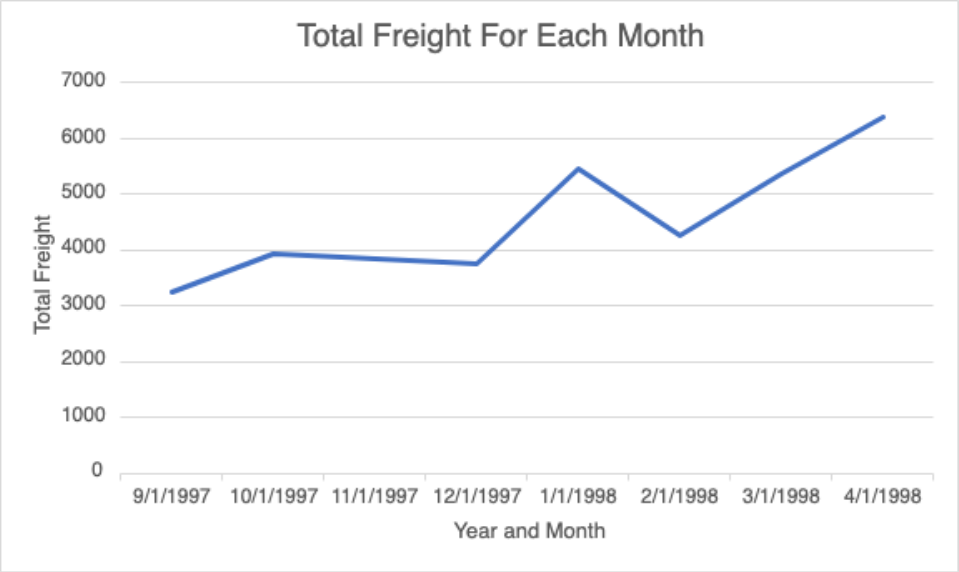


Most employees reported to the Vice President, followed by the Sales Manager. One person, Andrew Fuller the Vice President, did not have a manager to which he reports (for which the manager full name and manager title fields are blank). The young hire age was 28 years 7 mons 2 days, whereas the oldest was 55 years 7 mons 14 days.

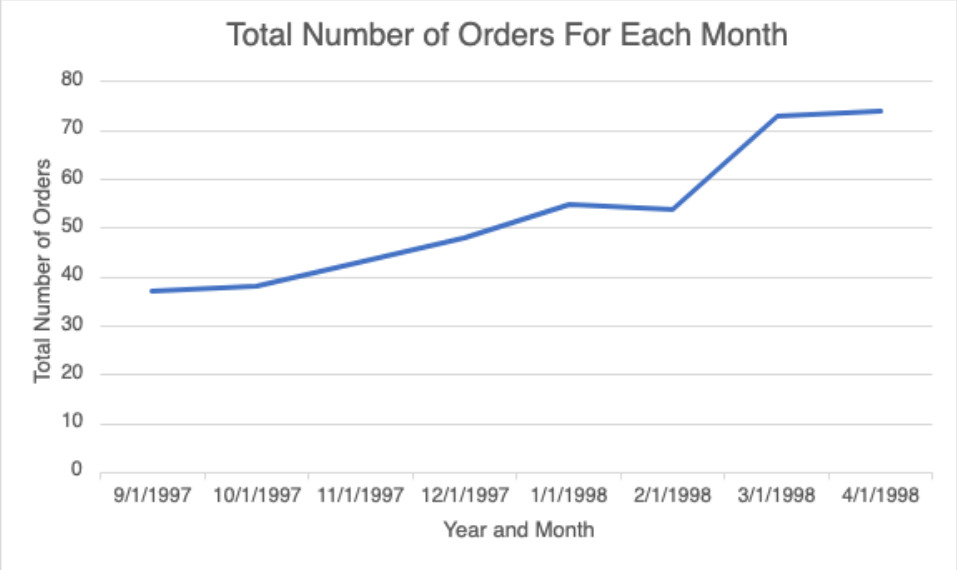

Number of Reporting for each Job Title Bar Plot
Number of Employees
0
1
2
3
4
5
Reports to No One
Sales Manager
Vice President, Sales
Manager Title

Task 4

The logistics' team required a review of performance between 1997 and 1998 by identifying the high order and freight performance months from orders data. Only months with total orders greater than 35 were returned for the specified years, ordering by the highest freight. Year month was generated by conversion of order_date to a varchar (text/string), starting at the first character with a length of 8 to drop the day. To this 01 was appended as per requested format. Total number orders was the orders count of year months. Total freight was calculated from the sum of freight for year months as a integer. Total freight for each month was graphed over the months between September 1997 and April 1998 by numbering their sequential month (September 1997 as 1 and April 1998 as 7), mostly increasing over time with the highest total freight month as April 1998. This was also seen as the case for the total number of orders.

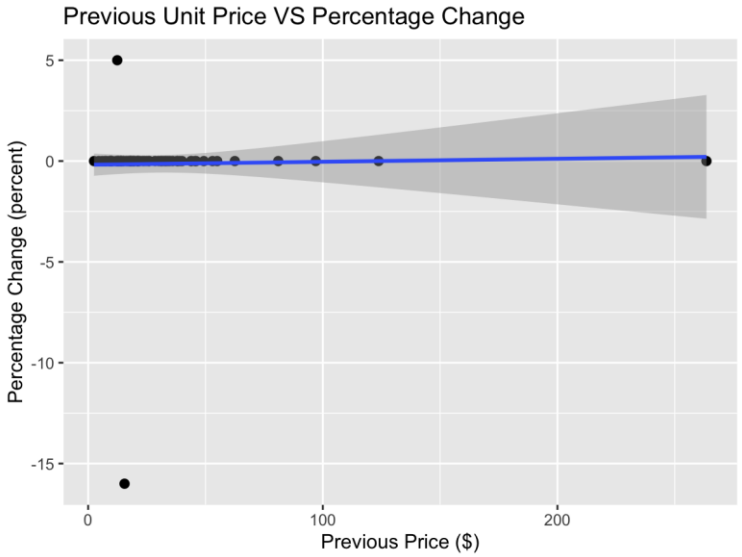


Total number of orders also increased at a mostly positive rate, more so than total freight.

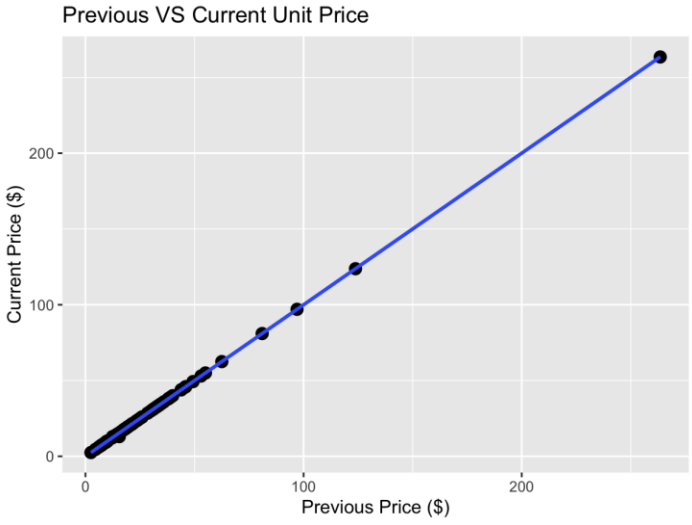


Task 5

The pricing team requested product pricing information regarding increases not between 20 percent and 30 percent with total number of orders greater than 10 orders. This was to be shown as product name, current and initial price, and percentage increase, ordered by lowest percentage increases first. Product name was listed alongside, current price as the products unit price rounded to 2 decimal places; previous unit price as orders unit price to 2 decimal places; and the percentage increase of these prices as the difference of these prices over previous unit price, multiplied by 100 as an integer. All prices were not seen to have a price change under the specified criteria except for Genen Shouyu, having 2 price changes of 16 and –5 percent.

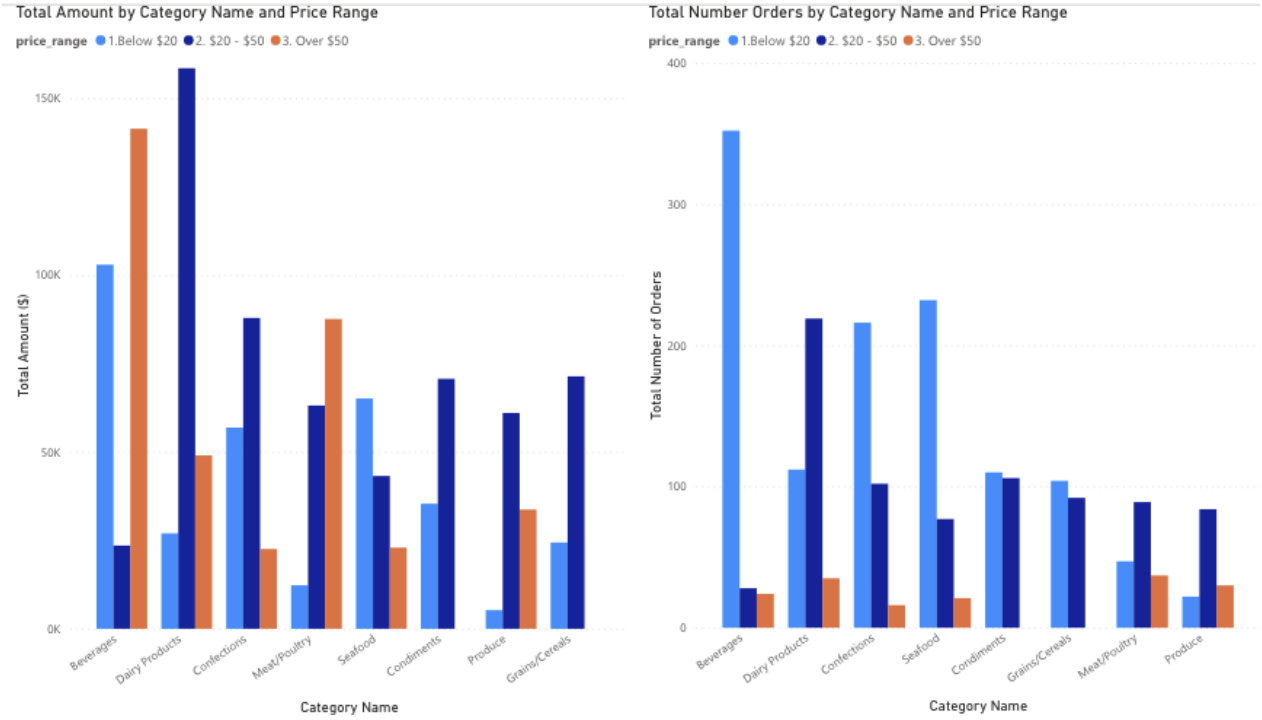


All information was obtained from the products table except for initial price being obtained from the order details data by a full outer join on products product id with orders product id, then, filtered and ordered for the specified criteria.



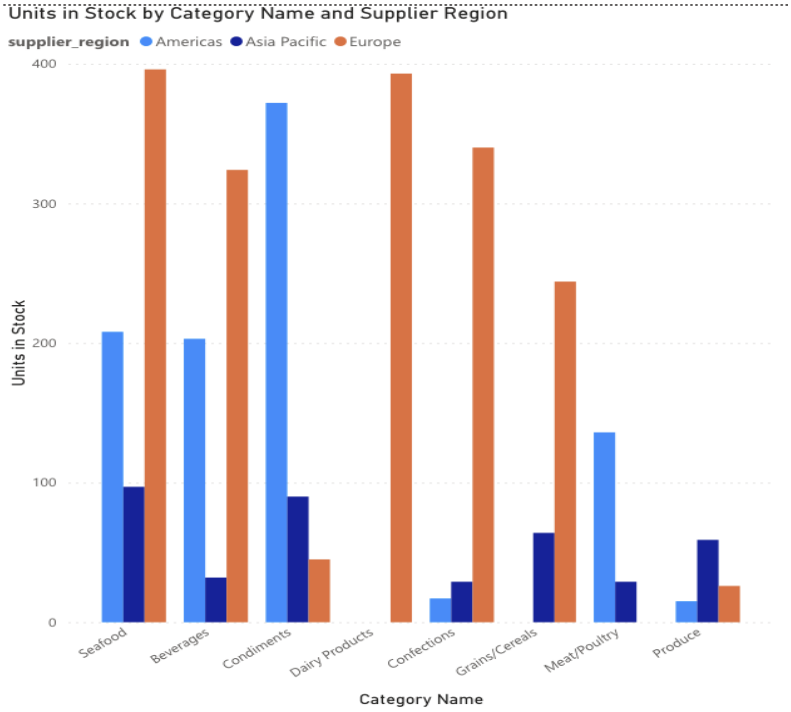
Task 6

This task for the pricing team focuses on the categories' performance dependent on price range in the following ranges: below 20$, 20-50$ and over 50$. Category name was obtained from products via a join with order details on product id on both for category id, then a another join for the names in categories. Price range was calculated as 1.Below $20 for unit price less than 20, 2. $20 - $50 for greater than or equal to 20, and 3. Over $50 for greater than 50; and total amount as sum of unit price multiplied by quantity and the discount rate subtracted from 1 as an integer over category name and price range. Total number order was the count of rows by category name and price range. The total amount of sales including discount and total number of orders in each category and price range combinations were also requested, ordered by category alphabetically and price range from the lowest range. This was achieved by obtaining order and price information from order details; product name from products; and category names from categories data. In terms of sales and order numbers for the below 20$ price range, Beverages was the highest and Produce was the lowest performing category. For 20-50$, Dairy Products was the highest whereas Beverages was the lowest performer for both sales and orders. Finally, for over 50$ categories, Beverages was the highest sales and Meat/Poultry was the highest order performer. The weakest performer in the over 50$ was Confectionery for both sales and orders.

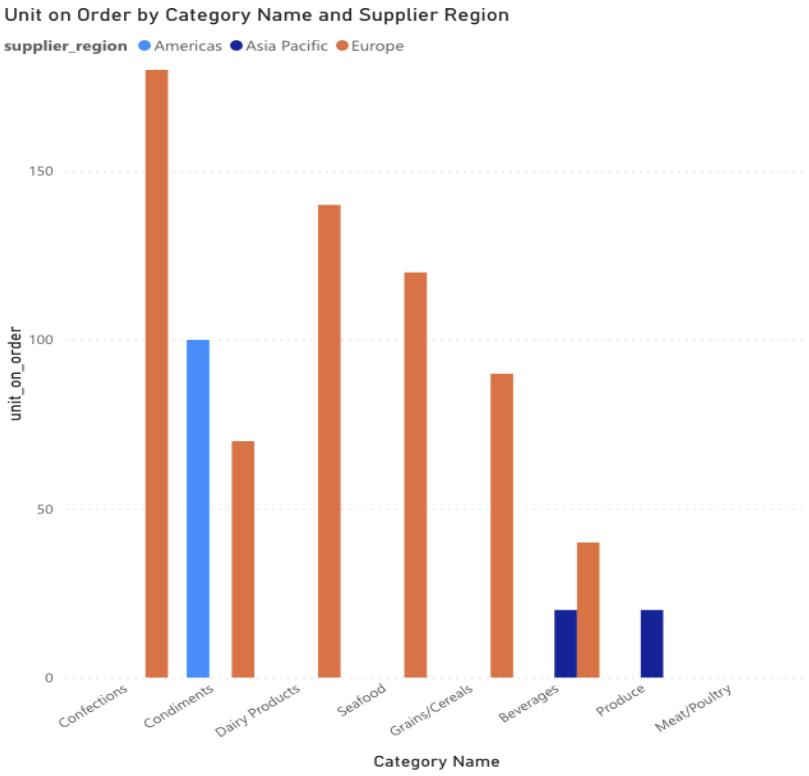


Task 7

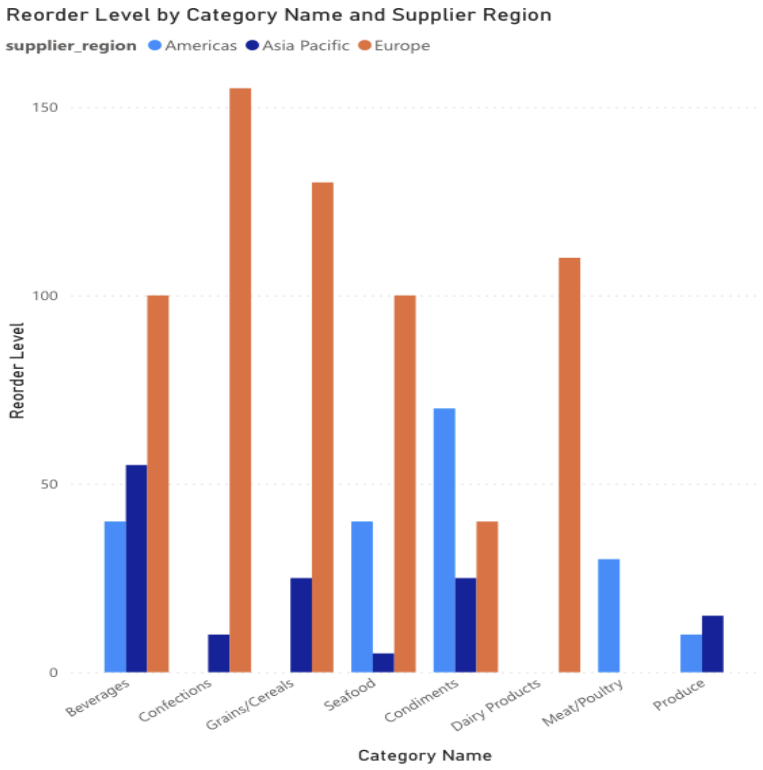
To assist with the logistics team's task of understanding current regional suppliers' stocks for each category of product; the supplier regions; category names; total units in stock, order, and reorder level were returned. Category name was found from categories and supplier region by grouping the relevant countries to the regions Americas, Asia Pacific and Europe with a case when statements. These were used to partition rows for the sum of unit in stock, unit on order and reorder level as new columns labelled similarly. This was achieved by two joins to obtain unit and reorder information from products to be combined with supplier country information for supplier region. Category id was also returned to join this with the same column in products. These were ordered by category name, supplier region and reorder level (all ascending). This was done through the combination of suppliers (for country) with products (remaining data except for category names), then this new combination with categories for category names. The returned out plotted below reveals the overall higher units in stock for Europe, especially compared to the lowest stocked Asia Pacific.



For Units on Order, Europe again has the greatest overall order, with Asia Pacific only having Produce and Beverages on order, and the Americas only having condiments on order.



Europe, again, had the highest level of reorder overall, followed by the Americas.



Task 8

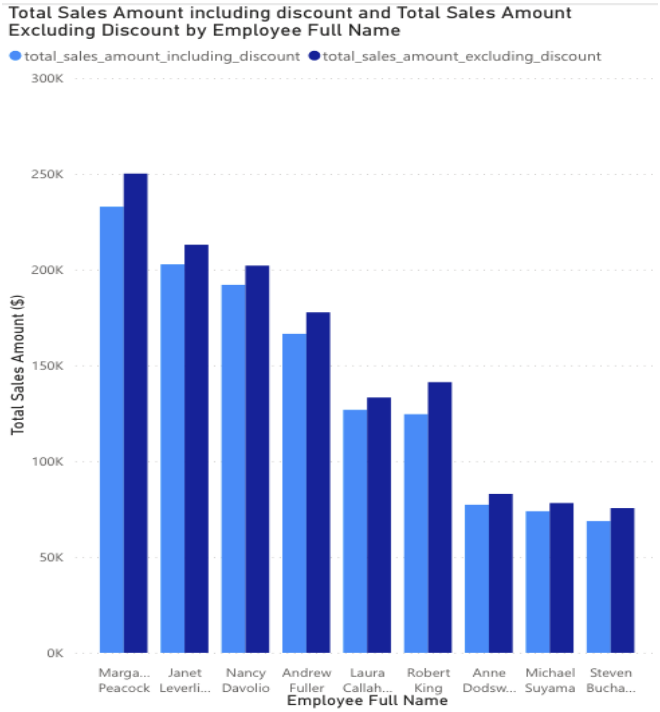
The task assigned by the pricing team to compare products' unit prices with its categories median and average unit price requires the data from products and categories. Product list of categories' names, product name, unit price, category average unit price (and its price position within the category), and category median unit price (and its price position within the category) were returned on items that were not discontinued. Only category name was required from categories. Product name and unit price was obtained from products, with the average of unit price for each category to 2 decimal places as average unit price. Median unit price was category id ascending and descending ordered unit prices binned in 2 bins; then using this to find the two middle rows by halving the sum of maximum of the ascending prices and minimum of descending prices; and finally rounding to 2 decimal places. Generated average and median unit price were used to make average and median unit price position: for unit price less than average/median unit price "Below _" was generated. This was repeated for unit price equal as well as greater than average/median unit price for the average/median unit price positions. These results were then ordered by category name and product name, both in alphabetical order. The output was then populated into the tables below. For every category, each product's average or median unit price within the price position groups were summed. Average unit price over each category was either below or over average, whereas median unit price over categories had all categories within below and over median and beverages, condiments and confections for equal median.

median_unit_price_position	Beverages	Condiments	Confections	Dairy Products	Grains/Cereals	Meat/Poultry	Produce	Seafood	**Total**
Below Median	72.00	105.25	97.50	165.00	60.75	15.73	53.26	100.20	**669.69**
Equal Median	54.00	21.05	16.25						**91.30**
Over Median	36.00	105.25	97.50	165.00	60.75	15.73	53.26	100.20	**633.69**
Total	**162.00**	**231.55**	**211.25**	**330.00**	**121.50**	**31.46**	**106.52**	**200.40**	**1,394.68**

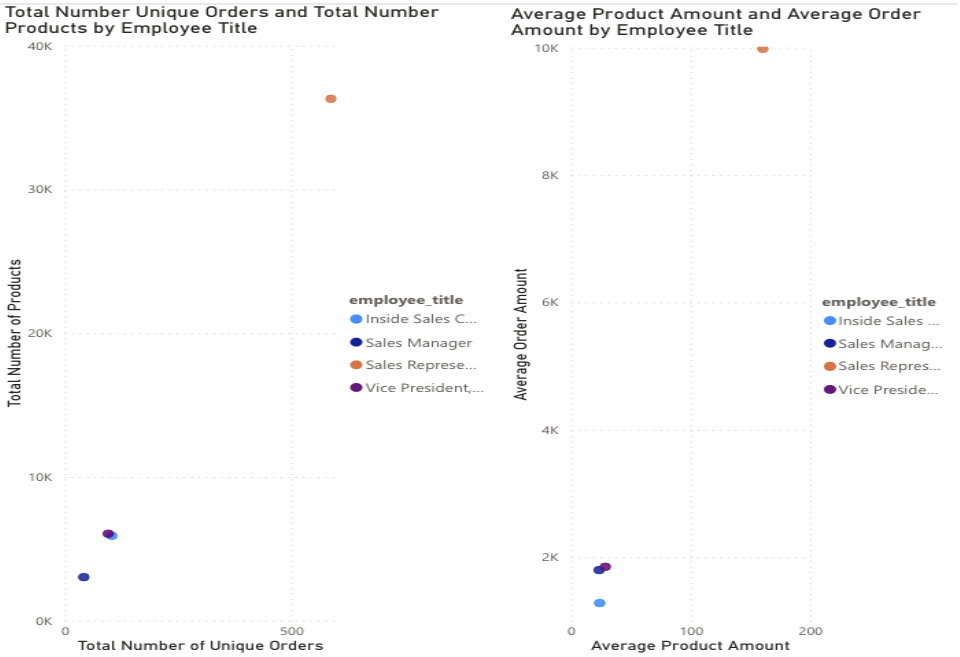
average_unit_price_position	Beverages	Condiments	Confections	Dairy Products	Grains/Cereals	Meat/Poultry	Produce	Seafood	**Total**
Below Average	368.24	160.93	226.44	114.92	85.16	15.72	58.12	165.44	**1,194.97**
Over Average	46.03	91.96	100.64	172.38	42.58	15.72	58.12	82.72	**610.15**
Total	**414.27**	**252.89**	**327.08**	**287.30**	**127.74**	**31.44**	**116.24**	**248.16**	**1,805.12**

Task 9

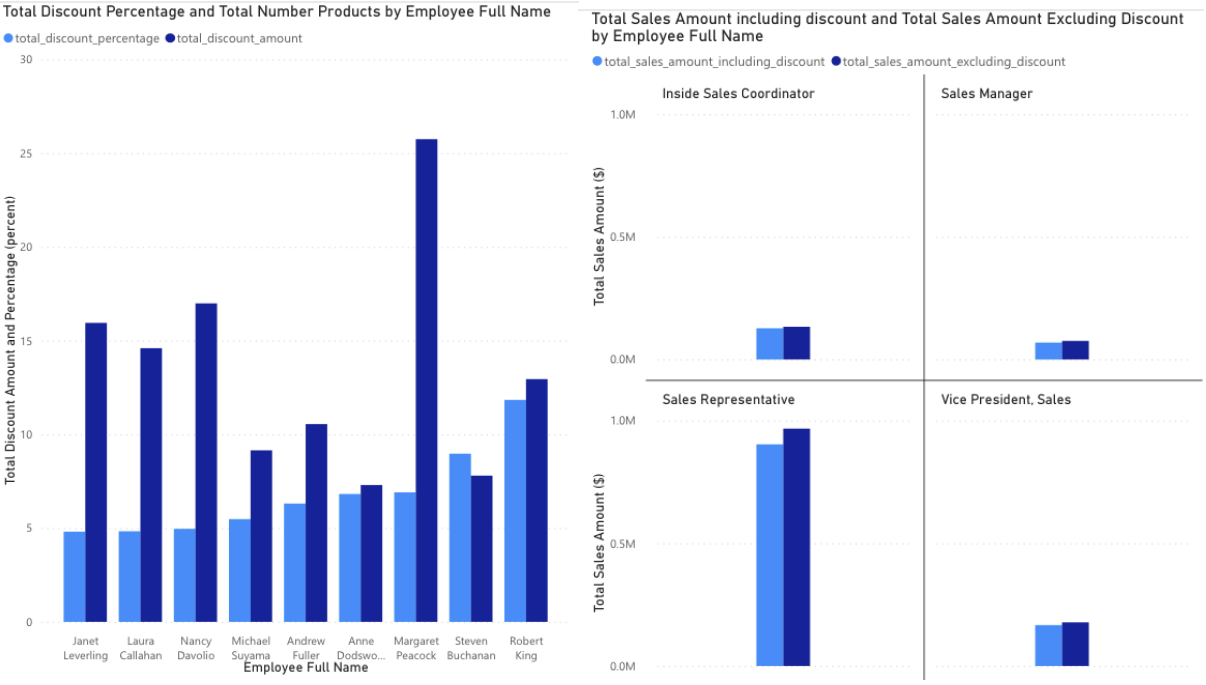
The task from the sales team is to measure employees' KPIs requested the following sales information: names, job titles, total sales (with and without discount), total number of unique orders, total number of products, average product amount (excluding discount), average order amount (excluding discount), and total discount amount and percentage. Employee full name was generated from the first and last name combined with a space, and title as employee title. Total sales amount excluding discount was calculated from multiplied unit price and quantity for each employees rounded to 2 decimal places. Total number unique orders were calculated from one less than the sum of ascending and descending rank of order id of each employee. Total number of products was calculated from the sum of quantity for each employee. Average product amount was the average unit price for each employee as a decimal to 2 places. Similar to this, total discount amount was calculated but with sum discount rather than average unit price. Total sales amount including discount was calculated from the sum of multiplication of unit price, quantity and 1 minus the discount rate for each employee to 2 decimal places. All of this was then ordered by total sales amount including discount from the highest value. The information for this output was generated from employee name and title from employees, order sales information from orders and order details by two joins between employees and orders as well as order details. The output was used to plot the following graphs. Margaret Peacock had the highest and Steven Buchanan had the lowest total sales (both including and excluding discounts).



For both highest unique orders vs number of products and average product vs average order amount, sales representatives had the highest figures and product diversity.

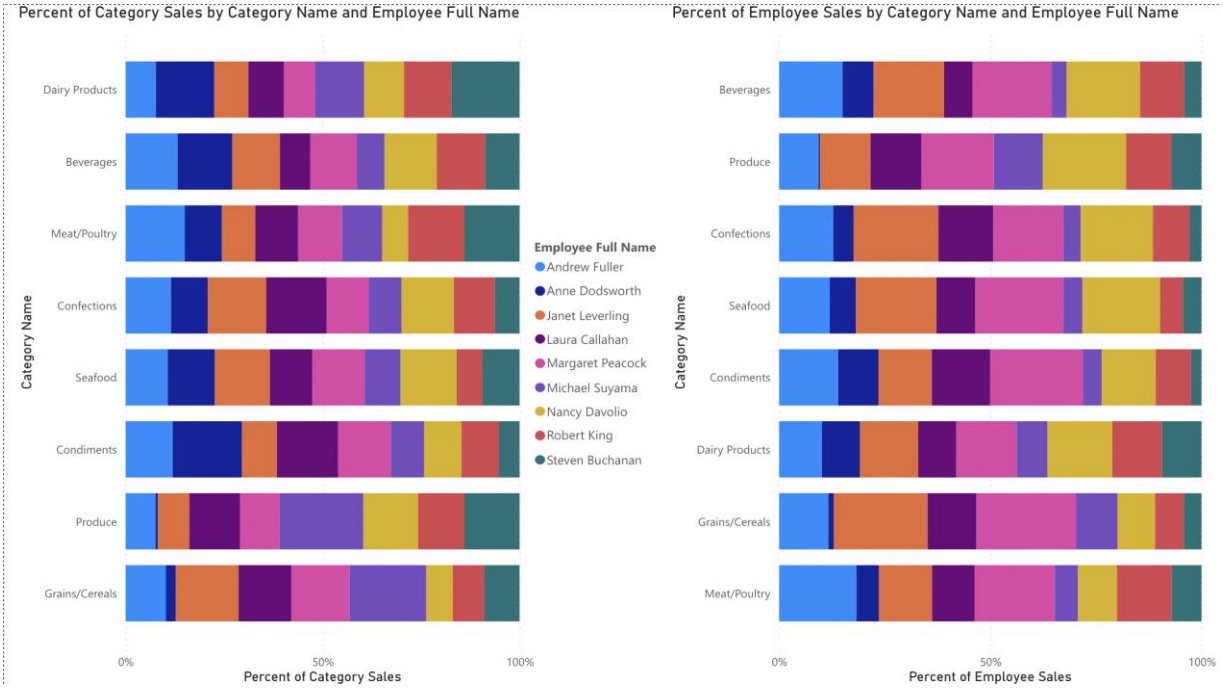


Margaret Peacock also had the highest total discount percentage where Steven Buchanan had the lowest, with Robert King having the highest total discount amount where Janet Leverling had the lowest. Total sales (with and without a discount) were highest for the sales representatives and lowest for the sales manager.



Task 10

This task's objective from the sales team was to measure employee performance across each category. To do so, category name, employee full name, total sales amount including discount, employees' sales as percentage of total sales amount including discount, and employee sales as a percentage of category sales. This was achieved through combining initially the orders and employees data to obtain the orders completed by each employee then adding order details, products and categories to gain insight into these respective areas. This was then ordered by category name alphabetically, highest total sales amount and plotted in the graphs below. Category name was obtained from categories and employee full name from first and last name combination. Total sale amount was calculated from sum of order price of each category and employee rounded to 2 decimal places. Percentage of employee sales was the total sale amount over the sum of total sale amount multiplied by 100 for each category to 2 decimal places, a formula repeated for percent of category sales for employees instead of categories. From the percentage of category sales, apart from slight underperformance from Anne Dodsworth (especially for produce and grains/cereals) there were no significant under or over performers. Performance was seen to vary from category to category.



Performance calculated by percentage of total sales reveals more significant differences: Janet Leverling, Margaret Peacock and Nancy Davolio were seen to dominate sales performance whilst Anne Dodsworth, Michael Suyama and Steven Buchanan were the poorest performers. A similar result was seen for analysis of total sales amounts.

Total Sale Amount by Category Name and Employee Full Name