3.2 dplyr Exercises

Business Examples using the Superstore Dataset

1. Filter the Superstore dataset for "Furniture" sales in the "West" region.

2. Select the "Customer Name", "State", and "Sales" columns and arrange them in descending order of sales.

3. Calculate the average profit per sub-category and arrange the results in ascending order. 🔗

Here, we are considering the entire Superstore dataset, not just the "Furniture" category or "West" region:

4. Identify Highest Selling Products in Furniture Category

Question: What are the top 5 highest selling products in the Furniture category?

5. Yearly Sales Trend for Technology in the East Region

Question: What is the yearly sales trend for Technology products in the East region?

To solve this question we need to use `lubridate` package.

`lubridate` is an R package that makes it easier to work with dates and times. It provides functions to parse various date-time formats, extract parts of dates or times, and perform calculations with date-time objects. With `lubridate`, you can add or subtract time periods from dates, handle time zones, and efficiently compare dates. Its functions are designed to simplify common use cases for date-time data and integrate well with the `tidyverse` set of packages. It is a go-to choice in R for date-time manipulation due to its user-friendly and intuitive syntax.

6. Average Discount Impact on Profits Across Segments

Question: How does the average discount given affect profits across different customer segments?

7. State-wise Profitability for Office Supplies

Question: Which states are the most and least profitable for Office Supplies?

8. Create a new column that represents the sales as a percentage of total sales within each category.

For this, we will first calculate the total sales within the "Furniture" category and then compute the percentage for each sale.

The code adds a column that shows sales as a percentage of the total sales within each category using `dplyr` from the `tidyverse`.

When you use `mutate()` on a grouped data frame in `dplyr`, it will preserve the grouping. However, this does not mean the data is permanently grouped within the resultant data frame. Instead, the grouping metadata is kept so that further `dplyr` operations can take advantage of it.

The `ungroup()` function is used to explicitly remove this grouping metadata from the data frame. While it's not always necessary to include `ungroup()` at the end of your operation, it can be a good practice when you want to ensure that the returned data frame doesn't carry any group structure, especially before performing operations that should not be influenced by the previous grouping.

If you don't call `ungroup()`, the returned data frame will still contain the grouping information, which might affect how subsequent functions work with the data frame. For example, if you pass this data frame to another `summarise()` call without removing the grouping, it will perform the summarisation by the existing groups.

However, when you just print the data frame or write it to a file, it will appear ungrouped. The grouping only affects how `dplyr` handles the data for further operations. If no further grouped operations are intended, then whether you ungroup the data or not may not have any visible impact on the output, but it's generally a good practice to ungroup a data frame if you've finished with group-specific operations to avoid unexpected results in subsequent analyses.

9. Correlation between Sales and Profit in Sub-Categories

Question: Is there a correlation between sales and profit within each sub-category?

```
# Calculate correlation between Sales and Profit for each Sub-Category
```

Before running the above examples, make sure to have the necessary libraries loaded (`dplyr`, `lubridate` for date operations) and that your data is properly formatted (e.g., `Order Date` is in the correct date format). The examples provided assume that the data types and formats are correct and that the Superstore dataset is named `superstore` in your R environment.

More Business Questions

Presented below are additional business-oriented questions that require a combination of data manipulation and analysis skills.

These questions require that you apply `dplyr` verbs creatively and think critically about business strategies and outcomes.

1- Customer Segmentation Analysis

Question: How do sales and profits differ among different customer segments (Consumer, Corporate, Home Office)?

Which segment is the most profitable, and why might that be the case?

Solution and Explanation:

To analyze sales and profits across different customer segments, we group the data by the `Segment` column and then summarize it to calculate total sales, total profit, and average profit for each segment. This provides insights into which customer segment is most profitable and hints at possible reasons behind the differences in profitability.

```
# Analyze sales and profits by customer segment


# View the results
```

2- Product Performance Evaluation

Question: Identify the top 5 and bottom 5 products based on profit margins. Discuss potential reasons for their performance and suggest strategies for improving sales of the less profitable products.

Solution and Explanation:

To identify products by profitability, we first calculate the profit margin for each product. Then, we rank these products to find the top 5 and bottom 5 based on their profit margins. This analysis can reveal which products are performing well and which are not, providing a basis for strategic adjustments.

```
# Calculate profit margin and identify top and bottom 5 products



# View the results
```

3- Regional Sales Analysis

Question: Compare the total sales and profits across different regions. Are there regions that underperform? What strategies could be implemented to boost sales in those regions?

Solution and Explanation:

This analysis involves grouping the data by region and summarizing it to calculate total sales and profits for each region. By comparing these figures, we can identify which regions might be underperforming and consider strategies to improve their performance.

```
# Analyze sales and profits by region




# View the results
```

4- Time Series Analysis

Question: How have sales and profits trended over time? Is there seasonality in sales? Identify the peak and low sales periods throughout the year.

Solution and Explanation:

For this analysis, we first need to convert the `Order Date` to a Date type. We then group the data by month and year to analyze trends and seasonality in sales and profits over time.

```
# Convert Order Date to Date type and perform time series analysis



# View the results
```

5- Discount Impact Analysis

Question: Analyze the impact of discounts on sales volume and profitability. Is there a discount threshold that maximizes profit without significantly hurting sales?

Solution and Explanation:

To explore the relationship between discounts, sales, and profitability, we group the data by discount level. Then, we calculate the average sales and profit for each discount level to understand how discounts affect profitability and sales volume.

```
# Analyze the impact of discounts on sales and profit



# View the results
```

6- Shipping Mode Analysis

Question: Examine how different shipping modes (Standard Class, Second Class, First Class, Same Day) affect sales and profits. Is there a correlation between ship mode and product category or segment?

Solution and Explanation:

This analysis looks at sales and profits across different shipping modes and explores potential correlations with product categories or customer segments. Understanding these relationships can help in optimizing shipping strategies for better profitability.

```
# Analyze sales and profits by shipping mode



# View the results
```

7- Product Category Analysis

Question: Within each product category (Furniture, Office Supplies, Technology), identify the most and least profitable sub-categories. Discuss possible reasons and suggest adjustments to product offerings or marketing strategies.

Solution and Explanation:

By breaking down sales and profits within each product category to the sub-category level, we can identify high and low performers. This detailed view provides insights for targeted strategies to enhance product offerings and marketing efforts.

```
# Analyze profitability by product category and sub-category



# View the results
```

8- Customer Loyalty and Order Size

Question: Who are the top 10 most loyal customers by number of orders and by total sales? Analyze their purchasing patterns to determine what drives their loyalty.

Solution and Explanation:

Identifying and analyzing the top 10 most loyal customers by both order frequency and sales volume offers valuable insights into customer loyalty drivers. This understanding can inform strategies to enhance customer retention and increase order size.

```
# Identify top 10 loyal customers by order frequency and total sales



# View the results
```

9- Market Penetration and Expansion

Question: Identify cities and states with low sales volume. What could be potential reasons for low penetration in these areas? Propose strategies for market expansion.

Solution and Explanation:

Focusing on areas with low sales volume helps in pinpointing under-served markets. Analyzing these areas can reveal opportunities for market penetration and expansion strategies.

```
# Identify areas with low sales volume



# View the results
```

10- Profit Efficiency by State or Region

Question: Which states or regions have the highest and lowest profit margins? Analyze factors that might contribute to these disparities and suggest ways to increase efficiency.

Solution and Explanation:

This analysis ranks states or regions by profit margin to highlight areas of operational efficiency and inefficiency. Understanding these disparities can guide strategic decisions to improve profit margins across different markets.

```
# Calculate and compare profit efficiency by state or region


# View the results
```