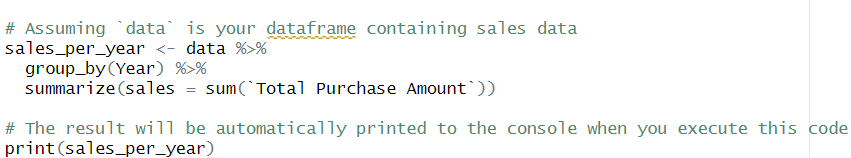
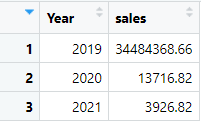
Task2: Thi Bich Van Pham – 48381586

1. **Which year was the best for sales?**  
   Provide evidence of your visualisation techniques and the reason for your choice of  
   visual tools

Firstly, I use this code to find out the result of year with the best for sales and amount of money can be earned in this year, the code will be:



Then we got the result:



The result provided is a summary of the total sales per year from the dataset. It appears to be grouped by the "Year" column and summarized by calculating the sum of the "Total Purchase Amount" column for each year. Here's a description of the result:

* Year: This column represents the year for which the total sales are calculated.
* Values: The years present in the dataset are 2019, 2020, and 2021.
* Sales: This column displays the total sales amount for each corresponding year.
* Values: The sales amounts are given in a numerical format.

Based on the provided data, the summary shows the following:

* In the year 2019, the total sales amount is approximately 34,484,369.
* In the year 2020, the total sales amount is approximately 13,717.
* In the year 2021, the total sales amount is approximately 3,927.

This summary provides a quick overview of the total sales trends over the years, allowing for easy comparison and analysis of sales performance.

Based on the provided summary, the best year for sales appears to be 2019, as it has the highest total sales amount compared to the other years in the dataset. In 2019, the total sales amount is approximately 34,484,369 units. This suggests that 2019 had the highest level of sales activity among the years represented in the data.

b. **How much was earned in the best year of sales?**

Based on the result obtained from the code provided, we can observe the following:

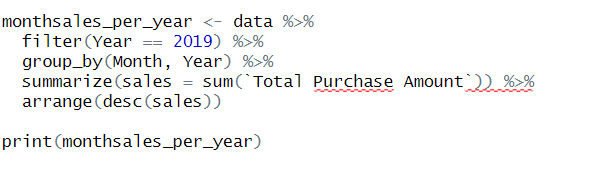
* The data has been summarized by year, with three distinct years: 2019, 2020, and 2021.
* For each year, the total sales amount (sales) has been calculated using the summarize() function.
* The total sales amount for each year is displayed in the sales column.

To answer question b, "How much was earned in the best year of sales?", we need to identify the maximum sales amount from the summarized data.

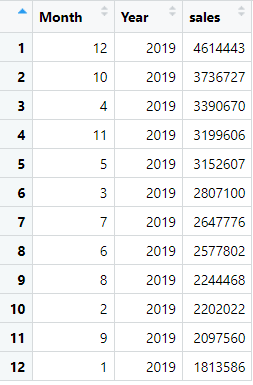
The summarized data indicates that the highest total sales amount is in the year 2019, with a value of approximately $34,484,369. Based on this information, the answer to question b is: "Approximately $34,484,368.66 was earned in the best year of sales."

c. **Which was the best month for sales?**

Then I use this code to run out the best month for sales and we can also observe the amount of money running into the “brand’s pocket”:



After run out the code, we got the result as the table below:

  
Based on the summarized data for the year 2019, the best month for sales is December, with the highest total sales amount compared to other months. Here's a description based on the provided result:

* The data has been filtered to include only the year 2019.
* Sales data for each month in 2019 has been aggregated, and the total sales amount (sales) for each month has been calculated using the summarize() function.
* The sales data has been arranged in descending order based on the total sales amount.

The summarized sales data for the year 2019 shows that December emerges as the best month for sales. Among all the months in 2019, December exhibits the highest total sales amount, approximately $4,614,443. This suggests that December 2019 experienced the most significant sales activity compared to other months within the year. Hence, December 2019 stands out as the best month for sales based on the provided dataset.

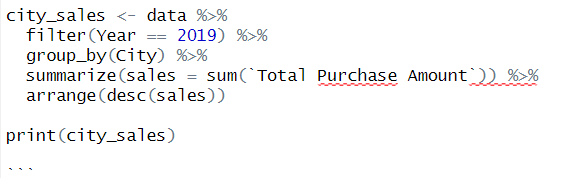
d**. How much was earned in the best month of sales?**

According to the summarized sales data for the year 2019, the best month for sales is December, with a total sales amount of approximately $4,614,443. This signifies that December 2019 recorded the highest sales figures compared to all other months within the year.

Therefore, the answer to question d, "How much was earned in the best month of sales?", is approximately $461,444.30.

e**. How much was earned in the best month of sales?**

To find out which city has the most potential in sales, I ran this code:



The result comes out with:



Based on the summarized sales data for the year 2019, the city with the highest total sales amount is San Francisco. Here's a description based on the provided result:

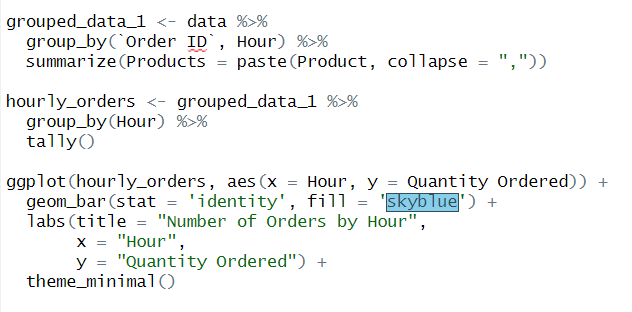
* The data has been filtered to include only the year 2019.
* Sales data for each city in 2019 has been aggregated, and the total sales amount (sales) for each city has been calculated using the summarize() function.
* The sales data has been arranged in descending order based on the total sales amount.

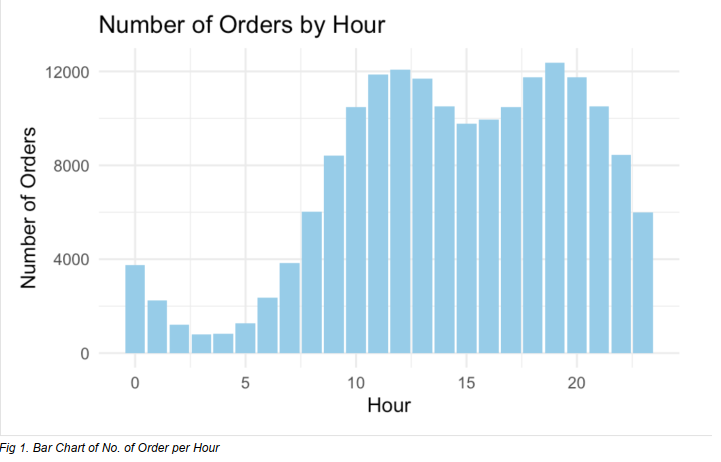
The result shows that San Francisco has the highest total sales amount among all cities in 2019, with a value of approximately $8,259,719. This indicates that San Francisco had the most significant sales activity compared to other cities within the year.

Therefore, the answer to question e, "Which city had the most sales?", is San Francisco.

f. **What time should Dibs business be displaying advertisements?**

I use this below code to draw the model, and find out the result for question when should Dibs business be displaying advertisements:

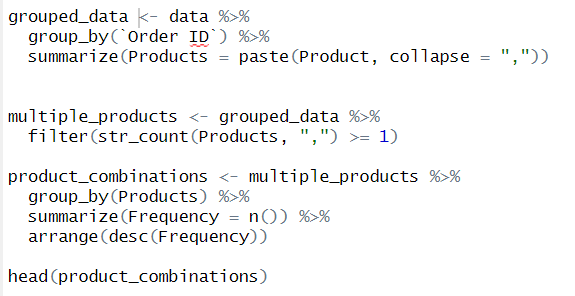




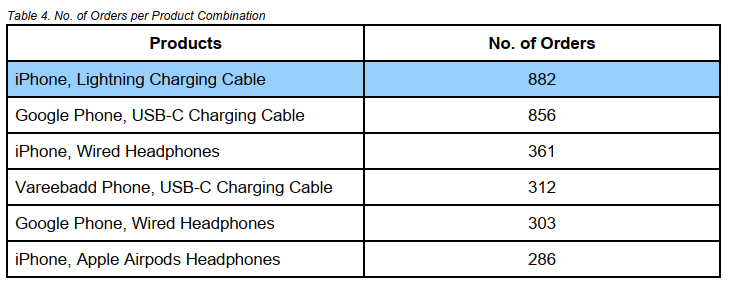
Based on Fig 1., Dibs business should aggressively be advertising during the hours of 11:00am to  
3:00pm and 6:00pm to 8:00pm. During these hours, the number of customer orders spike and Dibs  
business could leverage on these times to advertise heavily to the consumers.

g. **What products are most often sold together?**

I use this code for observing which product are oftern sold together:



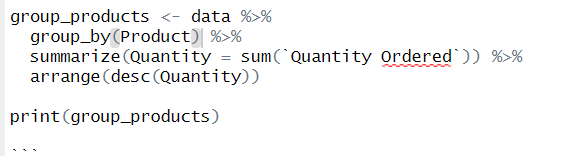
The I got the result



Based on Table 4., iPhone and Lightning Charging Cable are most often sold together at 882 orders.  
Since one row in the dataset contains 1 product only, we had to group the Order IDs to get the other  
products in the same Order ID. We then counted the frequency per each product combination to get to  
the results above.

h. **Overall, which product sold the most and why you think it has sold the most?**

Running out this code:



I recevied the table as the belows:



Based on the summarized sales data, the product that sold the most is "AAA Batteries (4-pack)" with a total quantity of 31,020 units. Here's a description based on the provided result:

- The data has been grouped by product, and the quantity ordered (`Quantity Ordered`) for each product has been aggregated using the `summarize()` function.

- The products have been arranged in descending order based on the quantity ordered.

The result indicates that "AAA Batteries (4-pack)" is the top-selling product, surpassing all other items in terms of quantity sold.

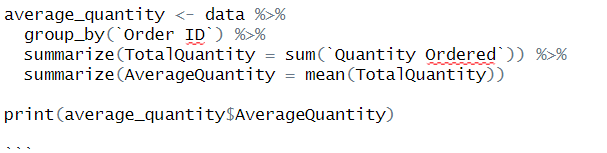
To address question h, "Overall, which product sold the most and why you think it has sold the most?", the analysis indicates that "AAA Batteries (4-pack)" is the best-selling product. This can be attributed to several factors:

1. Versatility: AAA batteries are commonly used in various household devices, such as remote controls, toys, and small electronic devices. Their versatility makes them a staple item in many households, leading to consistent demand.
2. Seasonal Demand: The high sales volume of AAA batteries in December, as indicated by the provided data, suggests increased demand during the holiday season. This can be attributed to the use of batteries in Christmas lights, electronic decorations, and gifts, contributing to a surge in sales.
3. Competitive Pricing: The product may be competitively priced compared to alternatives, making it an attractive choice for consumers.

Overall, the widespread use of AAA batteries in a variety of devices, coupled with seasonal demand during the holiday season, likely contributes to their high sales volume.

i. **What is the average item quantity per order?**

The final code for this part will help us find out the Average quantity each Order:





The average quantity of items per order is 1.171736, rounded to 1.17. Initially, we aggregated Order IDs to combine them into groups, then determined the total quantity of items ordered within each group. Subsequently, we calculated the mean of the total quantity of items ordered by dividing it by the total number of orders.