

Analytics Report

Looking back, moving forward



Unit of Study

BUSA8000

Presented by

Group 9

Date

May 31, 2024

Group 9

Authors

Khuat Son Tra Nguyen

Pham Thi Bich Van

Quoc Trung Nguyen

Duc Huy Tran

Student ID

48144134

48381586

48299405

48477192

Section

Task 1, 2, 5

Task 2

Task 2, 3

Task 4

Group Issue Addressing:

It is with regret that we must disclose an issue regarding the performance and contributions of one of our team members, Pham Thi Bich Van [48381586], who was assigned Task 2 in our project. Despite multiple reminders and extensions, Pham Thi Bich Van failed to complete her assigned task in a timely manner, ultimately submitting her work several days after the internal deadline. Furthermore, concerns were raised regarding the quality and originality of her submission, as it did not address the correct set of questions and appeared to be copied from another source. As a result, other team members had to take over her portions to ensure the project's progress and maintain academic integrity. This situation has caused significant delays and placed additional burdens on the remaining group members. We have documented the issues and sought guidance from the course instructors to address this matter formally and ensure fairness in the evaluation process.

Table of Contents

Problem Statement	3
Section 1: Data Pre-Processing	3
Section 2: Descriptive Statistics.....	5
a. Dibs' worst year and its total sales	5
b. Dibs' earnings in the best year of sales	5
c. Top sales month in the best year of sales	5
d. Top sales city in Dibs' best year of sales	6
e. Optimal advertisement timing for maximizing sales	6
f. Frequently purchased product pairs	7
g. Identifying the Best-Selling Item and Reasons Behind Its Success:	8
h. Least sold product in the best year of sales.....	8
Section 3: Data Visualization.....	9
a. Monthly Sales Trend vs Average Monthly Sales.....	9
b. Total Sales by State.....	11
c. Top 10 Products in the Best Year of Sales.....	11
d. Monthly Order vs Monthly Average.....	14
e. Daily Order vs Daily Average Order	15
f. Hourly Order vs Hourly Average Order	16
Section 4: Future in US market	17
1/ Methodology	17
2/ Modelling	18
3/ Conclusion.....	21
Section 5: Sales & Marketing Recommendations	22
▸ Pain Point #1: Leveraging Current Database to Drive Sales	22
▸ Pain Point #2: Customer Loyalty	23
References (APA7).....	24
Appendix	24

Problem Statement

▸ Dibs Company

Due to **its rapid growth and struggles to increase sales (pain point #1) and customer loyalty (pain point #2) despite having customer purchase history data**, Dibs needs to **analyze customer data to understand customers better, identify patterns and trends in their behavior** in order to improve its sales and marketing strategies, develop targeted campaigns, and increase sales and customer loyalty by the time aligning with the company's objectives

Section 1: Data Pre-Processing

▸ Sales Database - Years 2019, 2020, and 2021

The concept of “garbage in, garbage out” highlights that poor-quality data entry leads to unreliable data output (Kilkenny & Robinson, 2018). Moreover, effective sales strategies rely on accurate and timely sales forecasting, a process greatly enhanced by a well-prepared database (Pawar, 2023).

▸ **Objective:** To prepare for the subsequent stages, we aim to *develop a comprehensive monthly sales database* from raw sales data files for the years 2019, 2020, and 2021.

The database will be meticulously cleaned to ensure it is error-free, consistent, and complete. The process involves a detailed examination of the raw data for each variable (as summarized below) to understand its characteristics. Subsequently, we will identify and address any issues associated with each variable. The table below summarizes the identified issues and the solutions implemented to resolve them for each variable in the dataset.

#	Column	Issue(s)	Implementation(s)
1	Order ID	None	None
2	Product	Incorrect Data Entries Noises, Missing Values	Replace with imputed values Remove noises and missing values
3	Quantity Ordered	Values = 0 Formatting Issues	Remove any null values Change formatting to integer
4	Price Each	Incorrect Data Entries Formatting Issues	Replace with imputed values Change formatting to integer
5	Order Date	Incorrect Data Entries Not Ideal for Data Manipulation Noises, Missing Values	Replace with imputed values Create 5 new columns: 'Day', 'Month', 'Year', 'Hour' and 'Minute' Remove noise and missing values (if any)
6	Purchase Address	Not Ideal for Data Manipulation Incorrect Data Entries	Create 2 new columns: 'City' and 'State' Replace with imputed values

Table 1. Summary of Data Transformation.

Furthermore, we create a new column for the **`Total Purchase Amount`**, which is calculated by multiplying the **`Quantity Ordered`** by the **`Price Each`**. This is intended to serve the following tasks. Finally, we will export a cleaned CSV file, **`sales_clean.csv`**, to proceed with the subsequent stages.

Section 2: Descriptive Statistics

▸ Descriptive Summary - Years 2019, 2020, and 2021

▸ **Objectives:** In this section, our objective is to analyze the data to identify relationships between various variables and extract key insights relevant to Dibs.

▸ 8 Key Insights

a. Dibs' worst year and its total sales

Based on the provided data, the worst year of sales for Dibs **was 2021**, with a total sales amount of **\$3,927**.

Year	Total Sales (\$)
2019	34,484,368
2020	13,716
2021	3,926

b. Dibs' earnings in the best year of sales

Based on the provided data, the best year of sales for Dibs was **2019**, with a total sales amount of **\$34,484,369**. This figure reflects the peak of the company's financial performance.

c. Top sales month in the best year of sales

In the **best year of sales 2019**, the following table illustrates the monthly sales figures.

Month	Total Monthly Sales (\$)
Dec	4,614,443
Oct	3,199,606
Apr	3,390,670

The best sales month in 2019 is **December**, with total sales amounting to **\$4,614,443**. This figure significantly surpasses the sales of other months.

d. Top sales city in Dibs’ best year of sales

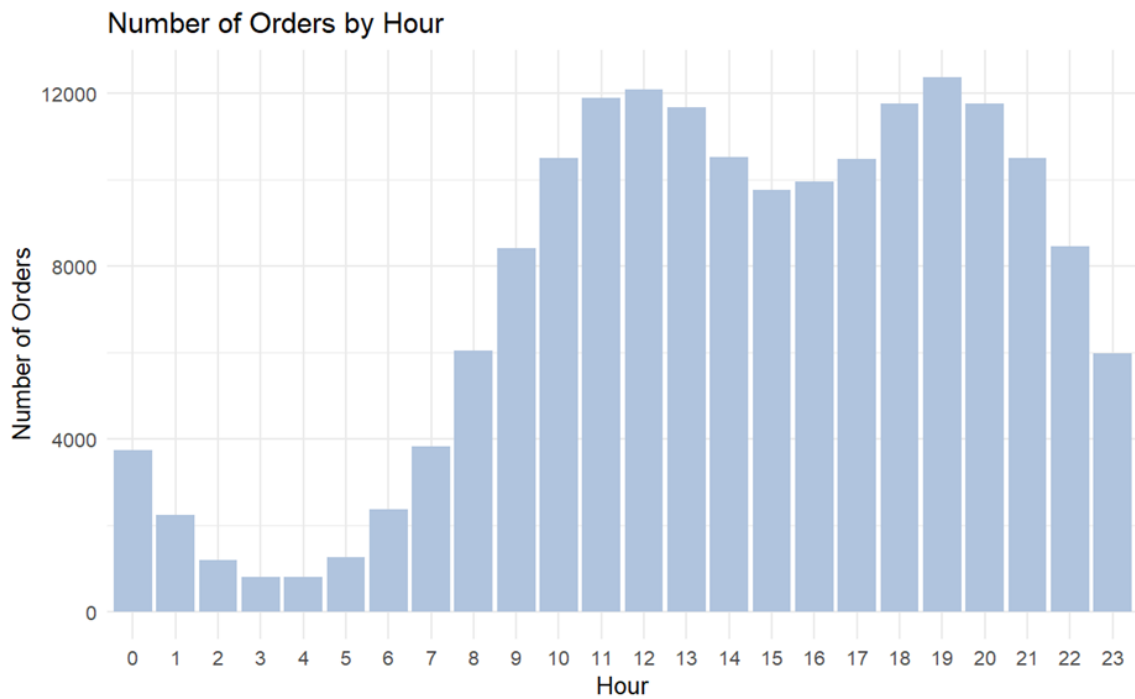
City	Total Regional Sales (in \$)
San Francisco	8,259,719
Los Angeles	5,451,571
New York	4,662,977
Boston	3,661,315
Atlanta	2,795,499
Dallas	2,766,263
Seattle	2,747,605
Portland	2,320,338
Austin	1,819,082

The following table illustrates the total sales in each city in 2019. **San Francisco** stands out as the top-performing city with total sales reaching \$8,259,719. This figure significantly exceeds those of other cities.

e. Optimal advertisement timing for maximizing sales

A "golden time" for customer purchases spans from **11 AM to 7 PM**. This period consistently shows the highest volume of orders, indicating peak customer activity. This insight suggests that advertising efforts would be most effective if concentrated within these hours.

The following bar chart illustrates Dibs' order patterns during their best sales year (2019).



f. Frequently purchased product pairs

Pairs of products	Number of Orders
iPhone, Lightning Charging Cable	882
Google Phone, USB-C Charging Cable	856
iPhone, Wired Headphones	361
Vareebadd Phone, USB-C Charging Cable	312
Google Phone, Wired headphones	303
iPhone, Apple Airpods Headphones	286

The **iPhone and Lightning Charging Cable** is the most frequently purchased combination, with a total of 882 orders. This pairing likely reflects a common need among consumers to have compatible charging options for their Apple devices.

g. Identifying the Best-Selling Item and Reasons Behind Its Success:

Product	Quantity Ordered
AAA Batteries (4-pack)	31,017
AA Batteries (4-pack)	27,635
USB-C Charging Cable	23,975

The top-selling product is **AAA Batteries (4-pack)** with 31,017 orders. This high volume of sales can likely be attributed to the universal need for batteries in various household and personal devices. Batteries are a repeat purchase item, essential for powering a wide range of products.

h. Least sold product in the best year of sales

With a total order of **646 orders**, the **LG Dryer** had the lowest number of orders in Dibs' best year. This low figure could be due to several factors such as the high price point of dryers compared to smaller, more frequently purchased items, or alternative products.

Section 3: Data Visualization

▸ Descriptive Analysis - Year 2019, 2020, 2021 in US market

▸ **Objectives:** This section presents visualizations that offer a comprehensive overview of sales performance. It explores key metrics:

1. Monthly Sales Trends vs Average Monthly Sales
2. Total Sales by State
3. Top 10 Sales Products
4. Monthly Order versus Monthly Average
5. Daily Order versus Daily Average
6. Hourly Order versus Hourly Average

a. Monthly Sales Trend vs Average Monthly Sales

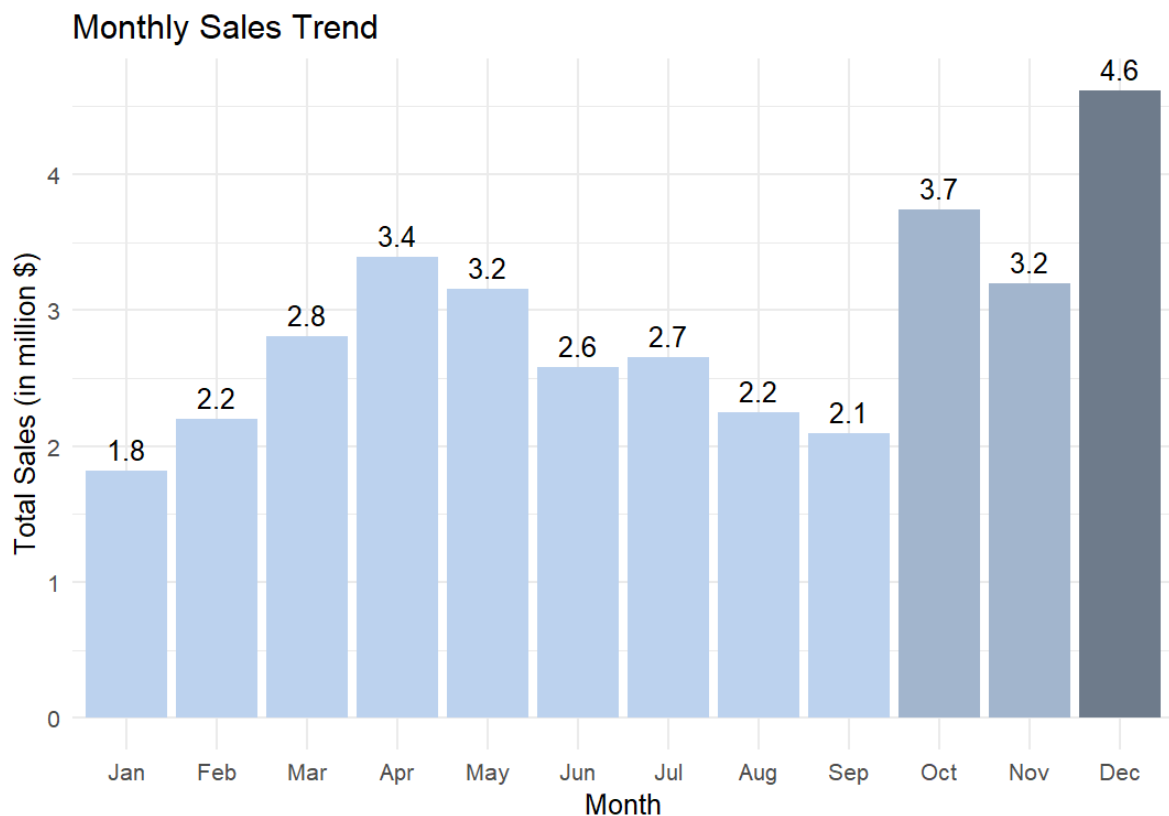


Figure 1. Monthly Sales Bar Chart in the years 2019, 2020, 2021.

- ▶ The peak in sales during November and December can be attributed to the holiday shopping season, with consumers making purchases for Christmas and end-of-year celebrations.
- ▶ The exceptionally high sales in December likely correspond to the peak shopping period leading up to Christmas.

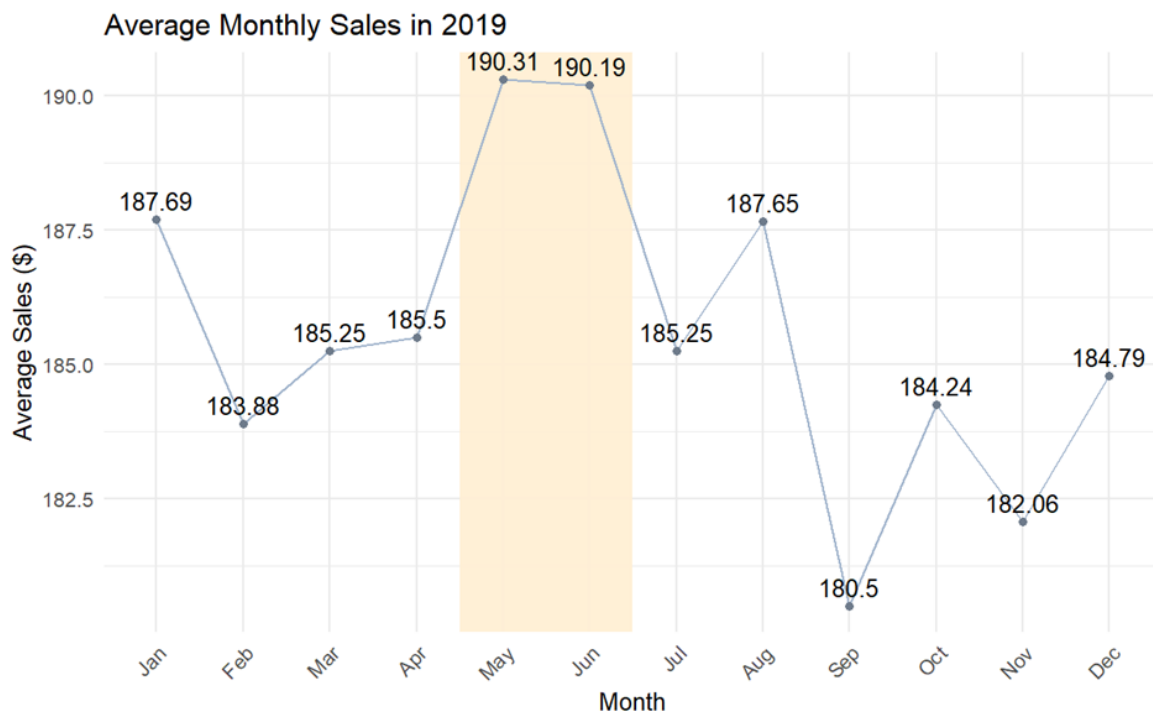


Figure 2. Average Monthly Sales in years 2019, 2020, 2021.

- ▶ The noticeable decline in average sales during September and October could be related to the end of the fiscal year for many businesses in the US, which typically falls on September 30th.
- ▶ The slight recovery towards the end of the year aligns with the holiday shopping season, as observed in the Monthly Sales Trend chart.

Insights: The holiday season, particularly Christmas and Boxing Day, is a major driver for increased consumer spending and higher sales towards the end of the year. This can inform Dibs in planning their sales and marketing campaign timelines.

b. Total Sales by State

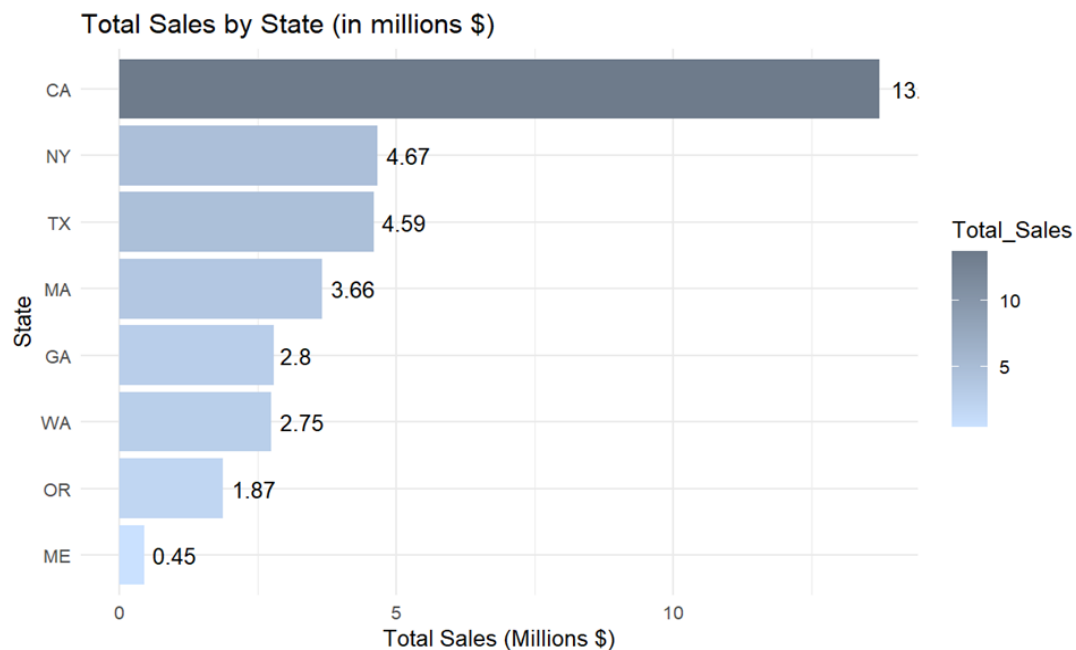


Figure 3. Total Sales by State in the years 2019, 2020, 2021.

► The chart of total sales by state shows that California leads with the highest sales figures, nearly tripling those of New York, which is in second place. Maine has the lowest sales.

Insights: This disparity can be attributed to significant differences in population, economic activity, and disposable income between these states. These insights can guide Dibs in **market segmentation** and **prioritizing sales and marketing efforts**.

c. Top 10 Products in the Best Year of Sales

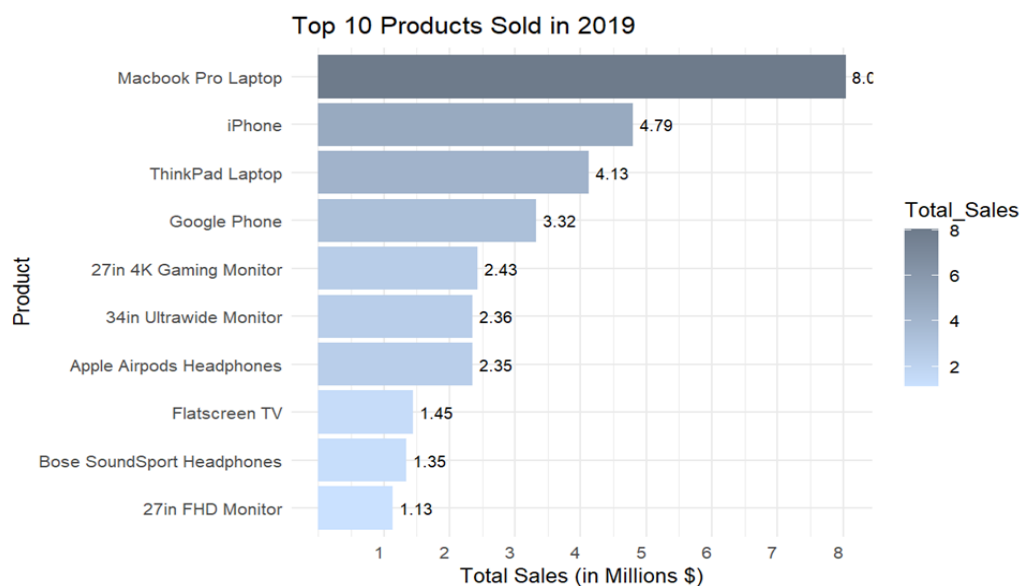


Figure 4. Top 10 Products in year 2019.

► In the banner sales year of 2019, the top 10 products were mainly mobile devices and electronics from brands like Apple, Google, and Microsoft. The MacBook Pro, the highest-selling item, generated around \$8 million in sales, highlighting consumer preference for premium devices and Apple's brand dominance.

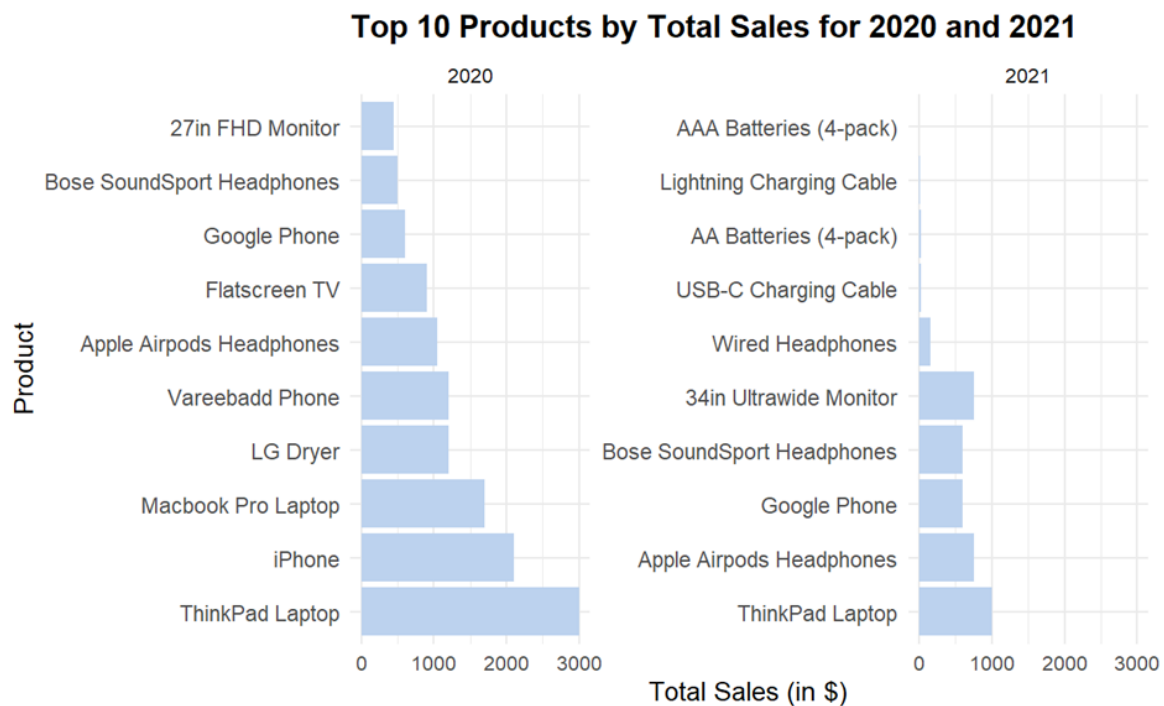


Figure 5. Top 10 Products in the years 2020 and 2021.

► In 2020 and 2021, Lenovo's ThinkPad Laptop replaced Apple's Macbook Pro as the dominant product due to disruptions in Apple's China-based supply chain caused by COVID-19 (Tom Coughlin, 2020). Production shortages led consumers to alternatives like ThinkPad to meet surging demand for electronics amid remote work and online activities (Boston, 2021), potentially resulting in temporary market share loss for Apple.

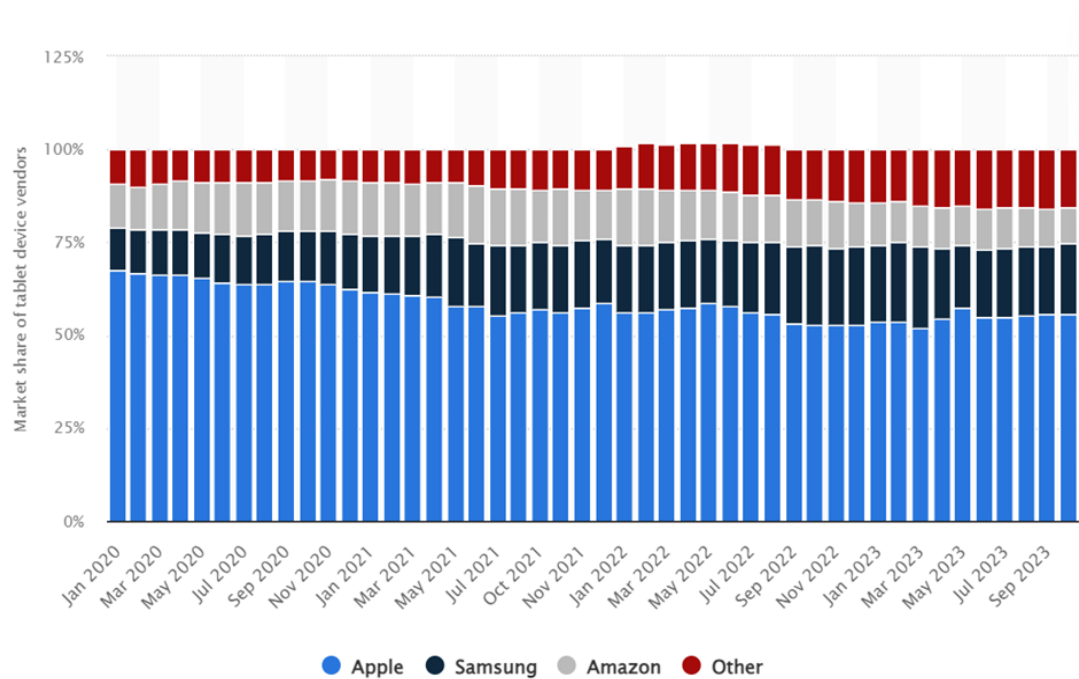


Figure 6. Market share of leading tablet device vendors in the United States from January 2020 to October 2023 (Laricchia, 2023)

Insights: Regarding Dibs' sales strategies, this insight can guide the adjustment of product offerings to better align with customer preferences. Emphasizing premium mobile devices and electronics from leading brands like Apple, Google, and Microsoft can capitalize on demonstrated consumer demand.

d. Monthly Order vs Monthly Average

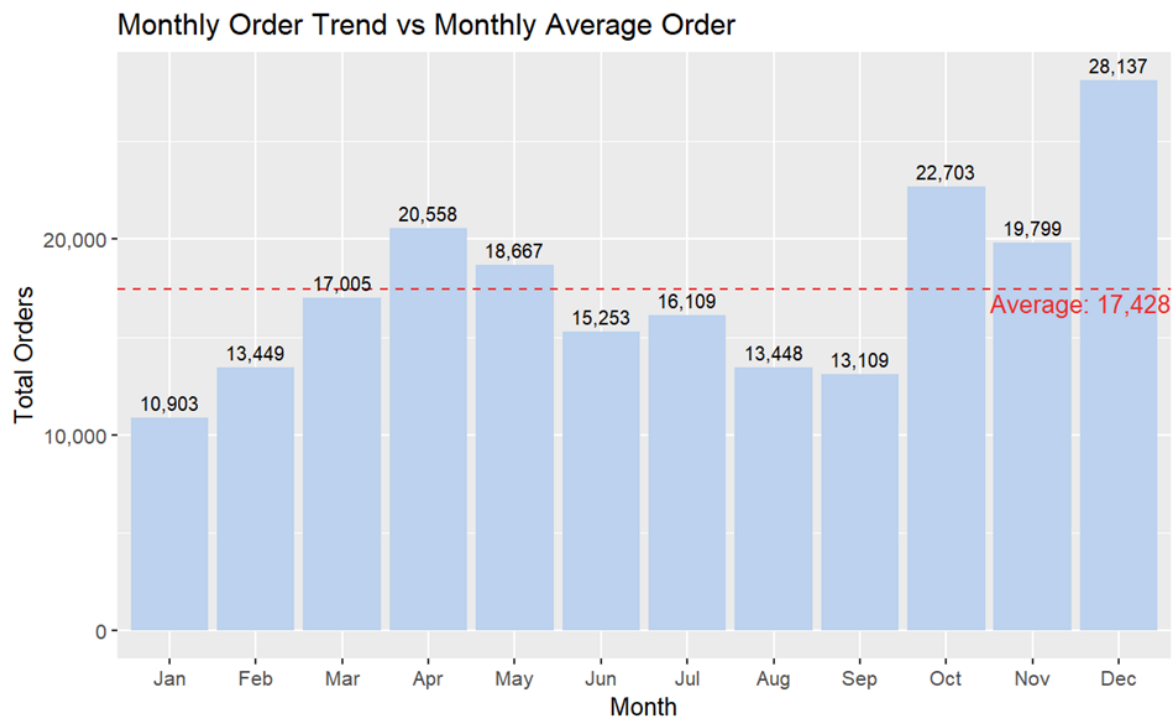


Figure 7. Monthly Order Trend vs Monthly Average Trend in the years 2019, 2020, 2021.

► The monthly order trend shows a seasonal pattern with December as the peak period (28,137 orders), likely driven by holiday shopping, while January had the lowest volume (10,003 units). Spikes above the average (17,428 orders) were observed in April, May, October, November, and December, potentially due to new product launches, promotions, and year-end shopping.

Insights: This can provide Dibs with valuable information regarding the most opportune seasons for their marketing campaigns.

e. Daily Order vs Daily Average Order

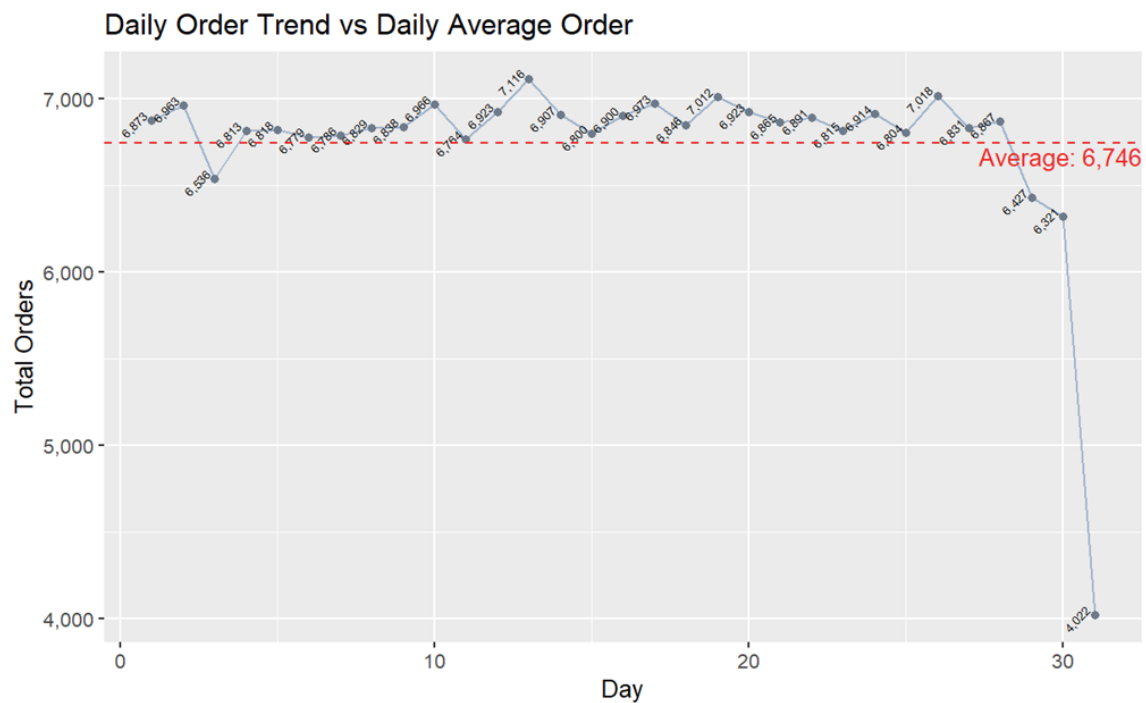


Figure 8. Daily Order Trend vs Daily Average Order in the years 2019, 2020, 2021.

► The daily order trend reveals a drop towards the end of the month, possibly attributed to the conclusion of monthly promotions and consumer spending habits. This observation acknowledges that not every month has 31 days.

Insights: Customers tend to shop frequently throughout the month, diminishing the significance of daily trends. Therefore, Dibs can safely disregard daily considerations when planning product launches or marketing campaigns.

f. Hourly Order vs Hourly Average Order

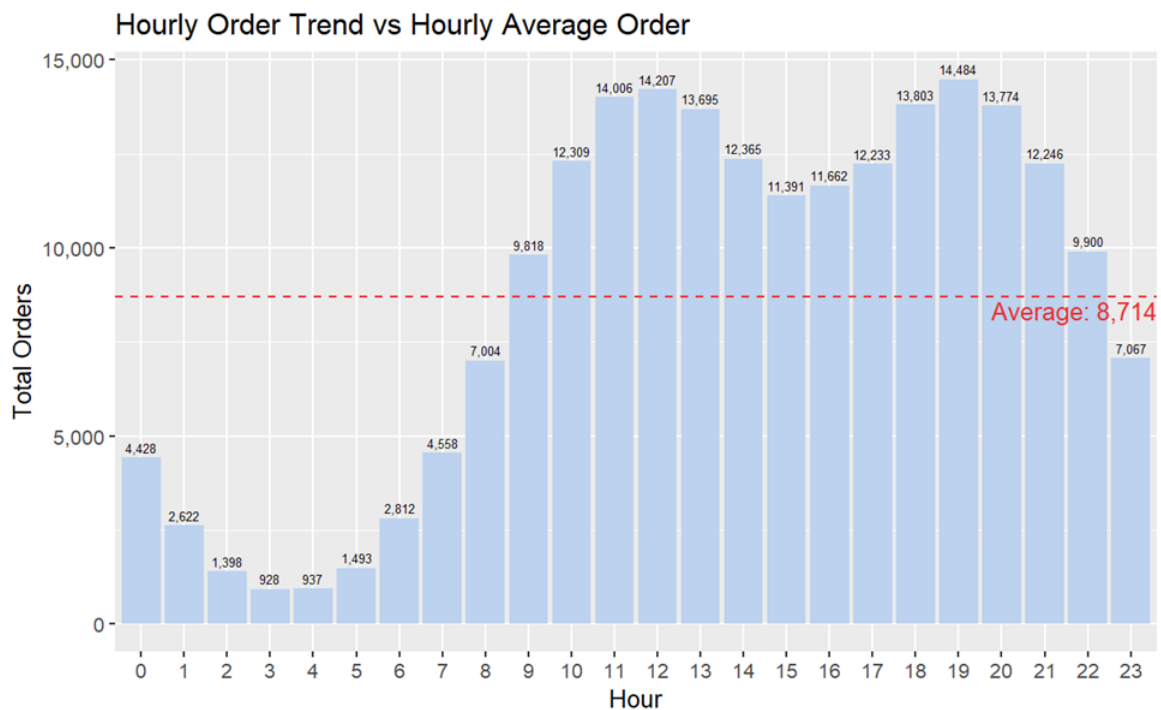


Figure 9. Hourly Order vs Hourly Average Order in the years 2019, 2020, 2021.

► The hourly order trend peaked at 7 PM (around 14,500 units), coinciding with post-work shopping, while the lowest volumes were recorded between 3 AM and 4 AM when most consumers are asleep. The average hourly order volume was 8,714 units.

Insights: The factor of hourly analysis holds significant importance as it facilitates the identification of notable sales increases during specific time slots. The company should effectively utilize this information to strategically run ads or publish content within these windows of time.

Section 4: Future in US market

► Predictive Analysis - For Dibs' Sales Strategies

1/ Methodology

◆ Data Preparation

To begin, even after the extensive data cleaning in Task 1, the data preparation steps in Task 4 are still necessary due to:

- **Date Filtering:** Filtering out rows where the Order.Date year is NA or 2020 ensures the data is consistent and relevant.
- **Calculating Total Sales:** Adding a Total.Sales column ($\text{Quantity.Ordered} * \text{Price.Each}$) is crucial for sales forecasting.
- **Aggregating Sales:** Data is aggregated to daily and weekly levels, essential for time series analysis and forecasting.
- **Time Series Conversion:** Converting weekly sales data into a time series format is required for applying models like the linear trend and Holt-Winters models.

◆ **Data Splitting:** Split the dataset into 80% training and 20% testing sets. The training set is for building the model, and the testing set evaluates the model's forecasting accuracy to inform sales and marketing strategies.

2/ Modelling

◆ Time series analysis

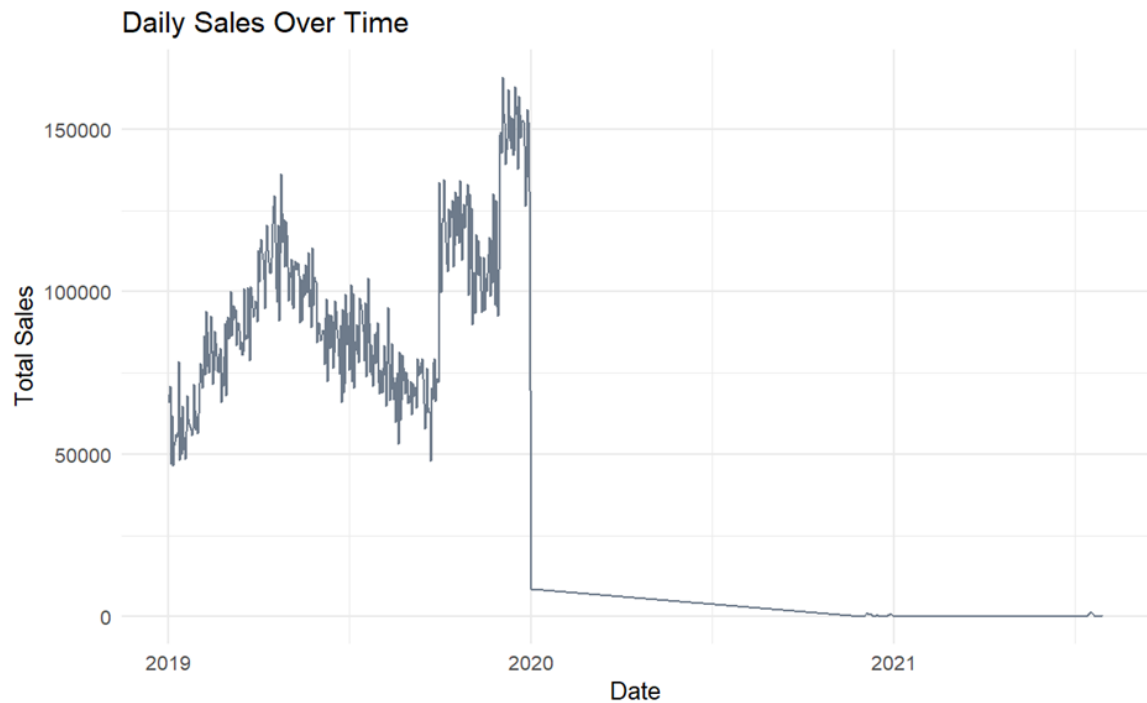


Figure 10. Daily Total Sales Over Time.

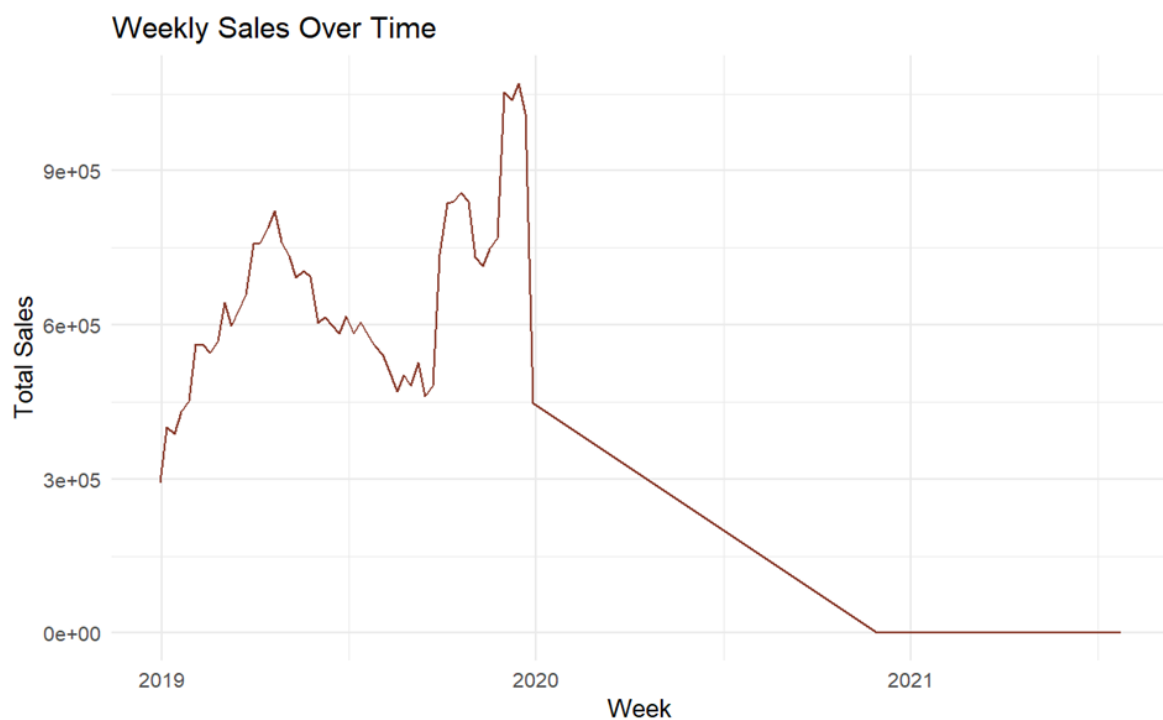


Figure 11. Weekly Total Sales Over Time.

Based on the time series analysis of daily and weekly sales data visualized in the provided plots, there appears to be a significant drop in sales around the beginning of 2020, followed by a prolonged period of low sales. This sudden change could be attributed to an outlier event or potential data quality issues that need to be addressed before selecting an appropriate forecasting model.

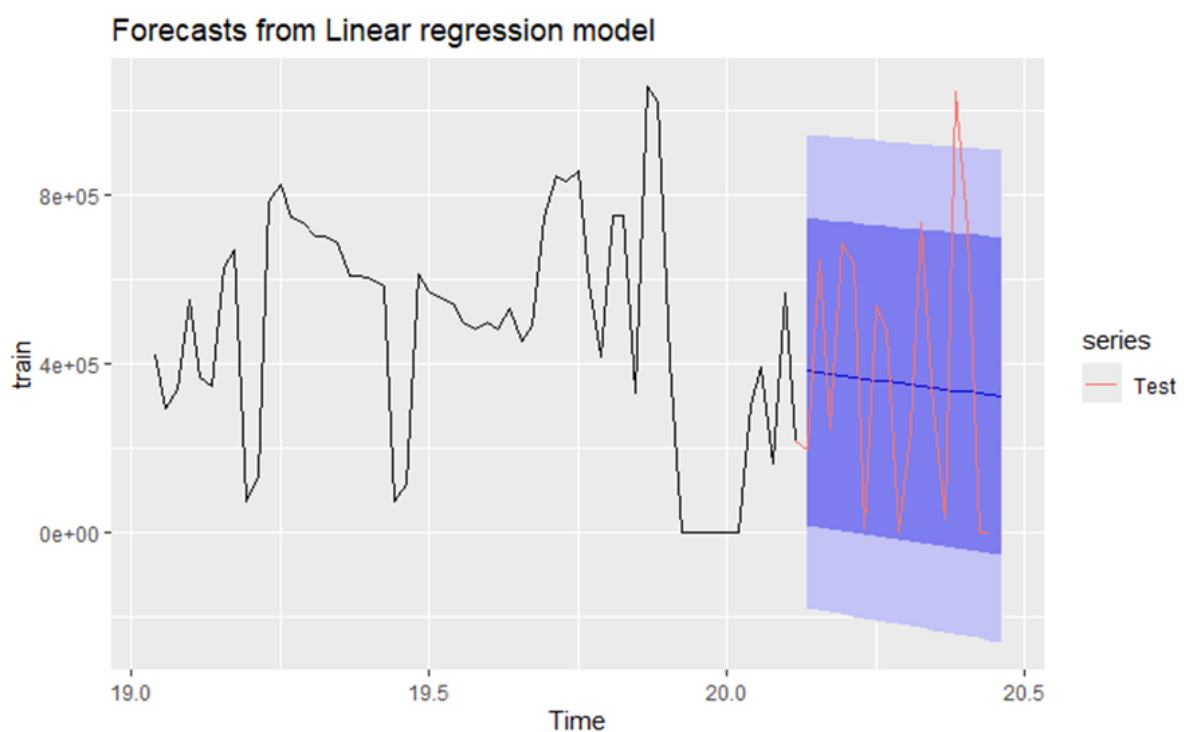
To build predictive models that can inform the company's sales and marketing strategies, the following techniques will be employed:

1. **Linear Regression Model:** A linear regression model will also be developed due to its simplicity and interpretability. This model assumes a linear relationship between the target variable (sales) and time, making it suitable for capturing overall trends in the data.
2. **Holt-Winters Exponential Smoothing:** This method is suitable for the given data as it can capture both trend and seasonality components present in the sales time series. The Holt-Winters model will be trained on the historical sales data to forecast future sales trends and seasonal patterns.

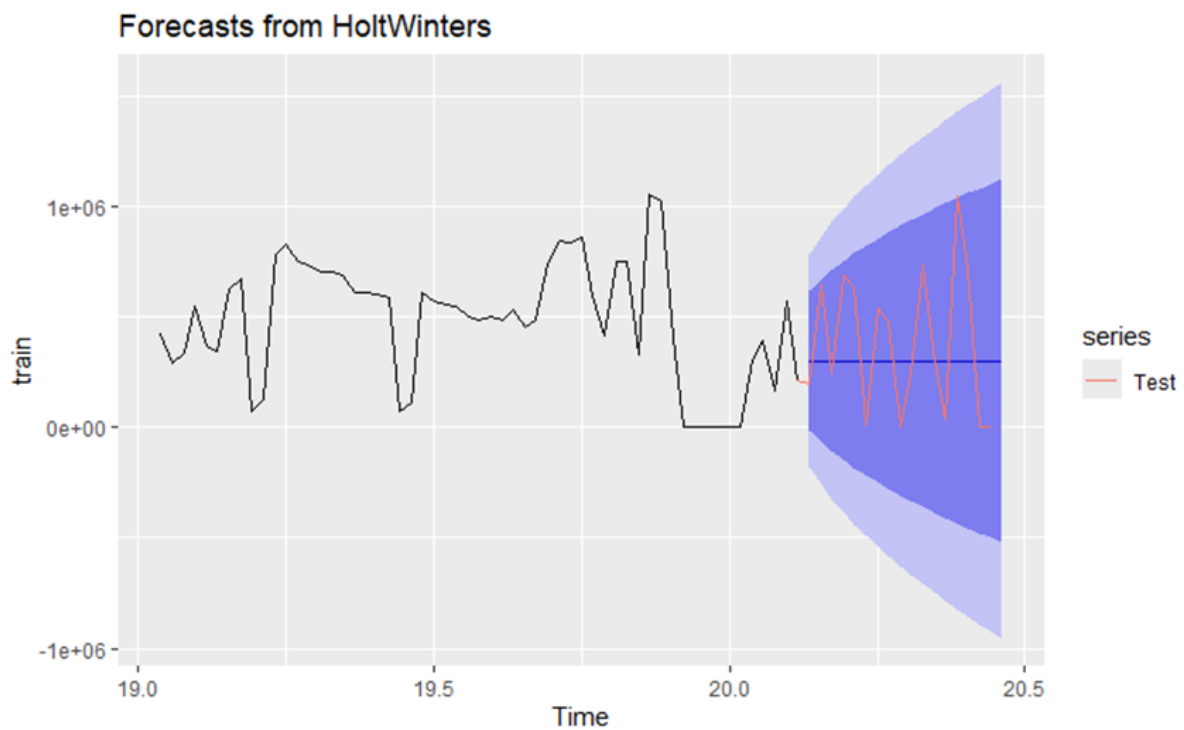
By utilizing these two techniques, the company can generate forecasts for future sales, enabling them to make informed decisions regarding sales and marketing strategies. The Holt-Winters model will provide insights into **expected seasonal fluctuations**, while the linear regression model can help identify **overall sales trajectories**.

◆ Forecasting

► Linear Regression Model:



► **Holt-Winters Exponential Smoothing**



► **Accuracy Table**

Model <chr>	ME <dbl>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>	ACF1 <dbl>
Linear Trend	-1.669077e-11	270104.6	219637.9	-11545.158	11574.51	0.532476
Holtwinters	1.300695e+04	247106.4	176782.0	-5418.738	5450.00	0.134347

Table 2. Accuracy Scores on selected Predictive models.

Evaluating the accuracy metrics, the Holt-Winters model performed better than Linear Regression, suggesting it is more suitable for predicting sales in this case. However, the Holt-Winters model assumes the data exhibits trend and seasonality components, which may be violated due to the lack of previous years' sales data and the impact of COVID-19 pandemic on supply chains and consumer behavior.

While the Holt-Winters forecast provides a reasonable starting point, its reliability is limited by the incomplete data and extraordinary events that disrupted sales patterns. To inform robust sales and marketing strategies, a deeper analysis of customer behavior and market trends is crucial, as these factors significantly influence future sales. The predictive models should be complemented with insights into consumer preferences, market dynamics, and external factors shaping sales patterns for a comprehensive

strategic approach. To study further about company's data and provide a further analysis, future sale-marketing strategies, we suggest studying about customer behavior and market-trend.

3/ Conclusion

In summary, the Holt-Winters model, accounting for trend and seasonality, performed better but had limited reliability due to the lack of previous years' data and the impact of COVID-19 on consumer behavior.

The plummeting sales in 2020 coincided with lockdowns and a rapid shift towards online shopping, challenging companies like Dib to adapt quickly. To develop robust strategies, the predictive models should be complemented with an analysis of evolving customer behaviors, online shopping patterns, and market dynamics.

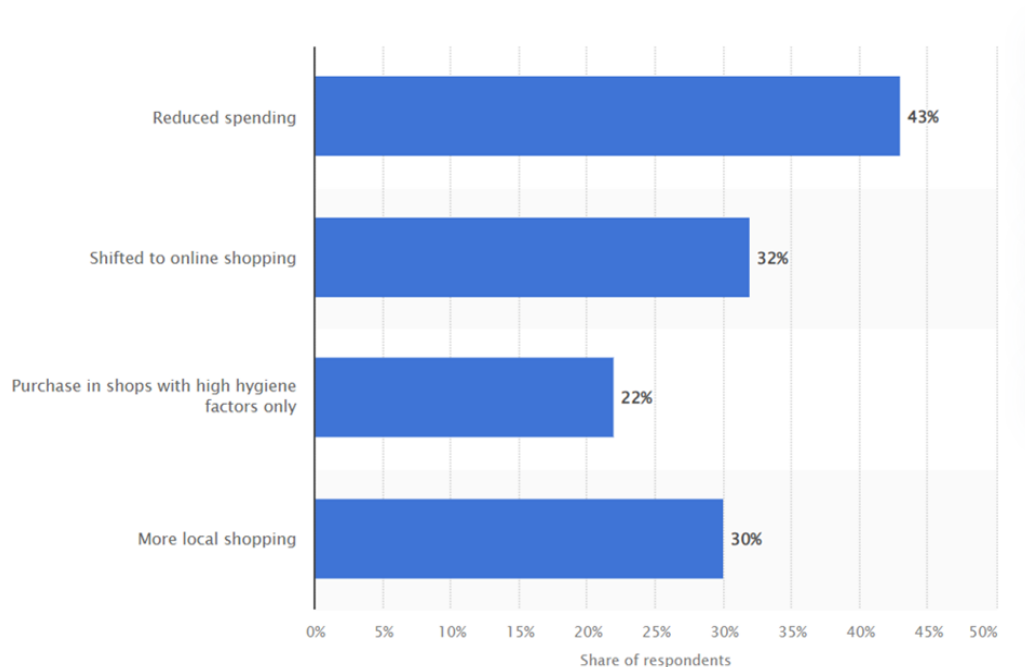


Figure 12. Changes in Customer Behaviors. Source: (Statista Research Department, 2023)

Thus, by combining forecasting techniques with insights into customer preferences and market trends, the company can enhance its online presence, digital marketing, and alignment with e-commerce demand, enabling informed sales and marketing strategies tailored to changing consumer needs.

Section 5: Sales & Marketing Recommendations

► For Dibs' Strategies

► Pain Point #1: Leveraging Current Database to Drive Sales

To help Dibs leverage their customer data for increased sales, we recommend adopting a customer-centric approach using the SAVE (Solution, Access, Value, Education) model. This method aligns closely with customer needs and behaviors (Wani, 2013), providing better insights than the traditional no-longer-relevant product-centric 4Ps marketing mix (Fou, 2021). Our analysis using the SAVE framework has shown its effectiveness in understanding customers and informing strategies to boost sales and loyalty.

✦ Solution

- **Data-driven Product Offerings.** Given the dominance of Apple's iPad in the tablet market, it would be a strategic move for Dibs to incorporate iPads into their product lineup, catering to the strong consumer preference for Apple's products and tapping into their loyal customer base.
- **Prioritize E-commerce Platform.** With the surge in online shopping and electronic payments during the pandemic, Dibs must prioritize their e-commerce platform and online shopping experience, including a user-friendly website, seamless checkout process, and secure payment gateways.
- **Address Non-response Bias.** Increase market penetration in underserved areas to mitigate potential non-response bias in the data.

✦ Access

- **Expanding and Enhancing Distribution Channels.** In light of the increased adoption of online shopping and electronic payments during the pandemic, Dibs must prioritize their e-commerce platform and online shopping experience. This includes ensuring a user-friendly website, seamless checkout process, and secure payment gateways to facilitate online transactions and cater to the growing demand for digital shopping convenience.
- **Address Non-response Bias.** Increase market penetration in underserved areas to mitigate potential non-response bias in the data.

<p>✦ Value</p> <ul style="list-style-type: none"> ► Capitalize on Stimulus Check Spending. Monitor and align marketing campaigns and promotional strategies with the timing of stimulus payments, offering targeted deals and discounts to capture increased spending during these periods. ► Adapt to Economic Conditions. Observe economic factors impacting customer spending habits and offer competitive pricing, bundles, and discounts accordingly. ► Leverage Customer Feedback. Utilize customer reviews, testimonials, and industry ratings to gain valuable insights and address customer needs effectively. ► Enhance Customer Experience. Provide extended warranties, support services, and trade-in programs to add value and improve customer satisfaction. 	<p>✦ Education</p> <ul style="list-style-type: none"> ► Content Sharing. Utilize insights on customer spending habits to customize content on social media and other video platforms, effectively reaching and engaging the target audience. ► Strategic Partnerships. Collaborate with industry experts, influencers, and thought leaders to expand the company's customer base through targeted marketing strategies.
--	--

► **Pain Point #2: Customer Loyalty**

Stage 1	► Customer Database	Implement a comprehensive customer database with unique customer IDs, enabling effective tracking of purchase histories and facilitating the development and expansion of loyalty programs.
Stage 2	► Monitor Consumer Behavior	Regularly analyze customer data, conduct market research, and monitor industry reports to stay ahead of evolving consumer behaviors and preferences in the post-pandemic landscape, ensuring relevance and adaptability.

By implementing these recommendations, Dibs can effectively navigate the post-COVID-19 market landscape, cater to evolving customer needs, and position themselves as a customer-centric and adaptive brand, poised for success in the dynamic retail environment.

References (APA7)

1. Federal Reserve Bank of Boston. (2021, December 17). *Has COVID changed consumer payment behavior?* Federal Reserve Bank of Boston.
<https://www.bostonfed.org/publications/research-department-working-paper/2021/has-covid-changed-consumer-payment-behavior.aspx>
2. Fou, D. A. (2021, January 25). *4P's Are Dead—Because They're Academic, Not Practical And More Irrelevant Than Ever*. Forbes.
<https://www.forbes.com/sites/augustinefou/2021/01/25/4ps-are-deadbecause-theyre-academic-not-practical/?sh=4ae0acc36a27>
3. Kilkenny, M. F., & Robinson, K. M. (2018). Data quality: “Garbage in – garbage out.” *Health Information Management Journal*, 47(3), 103–105.
<https://doi.org/10.1177/1833358318774357>
4. Laricchia, F. (2023, October 27). *Tablet vendors market share U.S. 2020*. Statista.
<https://www.statista.com/statistics/1120402/market-share-tablet-device-vendors-us/>
5. Nayak, J., Mishra, M., Naik, B., Swapnarekha, H., Cengiz, K., & Shanmuganathan, V. (2021). An Impact Study of COVID-19 on Six Different Industries: Automobile, Energy and Power, Agriculture, Education, Travel and Tourism and Consumer Electronics. *Expert Systems*, 39(3). NCBI. <https://doi.org/10.1111/exsy.12677>
6. None Shrikant Pawar, None Aiman Sheikh, & None A. R. Sonule. (2023). Sales analysis using data mining. *International Journal of Advanced Research in Science, Communication and Technology*, 304–309. <https://doi.org/10.48175/ijarsct-9419>
7. Statista Research Department. (2023, January 3). *Australia: Consumer shopping behavior changes as a result of coronavirus pandemic 2020*. Statista.
<https://www.statista.com/statistics/1192200/australia-consumer-shopping-behavior-changes-as-a-result-of-coronavirus-pandemic/>
8. Wani, T. A. (2013). From 4Ps to SAVE: A Theoretical Analysis of Various Marketing Mix Models. *SSRN Electronic Journal*, 1(1). <https://doi.org/10.2139/ssrn.2288578>