## Chapter 1: Data Cleaning

## To provide informed recommendations for Dibs regarding their sales strategies, the initial and crucial stage is data preprocessing. 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** | **Issues** | **Implementations (in the following order)** |
| Order ID | None | None |
| Product | Incorrect Data Entries Noises, Missing Values | Replace with imputed values Remove noises and missing values |
| Quantity Ordered | Values = 0Formatting Issues | Remove any null valuesChange formatting to integer |
| Price Each | Incorrect Data EntriesFormatting Issues | Replace with imputed valuesChange formatting to integer |
| 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) |
| Purchase Address | Not Ideal for Data ManipulationIncorrect Data Entries | Create 2 new columns: ‘City’ and ‘State’Replace with imputed values |

Finally, we will export a cleaned CSV file, "sales\_clean.csv," for other team members to proceed with the subsequent stages.

## Chapter 2: Descriptive Analysis

**Objectives:**

In this chapter, we aim to find Dibbs' worst and best sales years, the top sales month and city in the best year, optimal ad times, frequently bought product pairs, and the best and least sold products.

**Findings:**

a. Dibb’s worst year and its total sales:

|  |  |
| --- | --- |
| **Year** | **Total Sales (in $)** |
| 2019 | 34,484,368 |
| 2020 | 13,716 |
| 2021 | 3,926 |

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

b. Dibb’s earnings in best year of sales:

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

c. Top sales month in best year of sales:

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

|  |  |
| --- | --- |
| **Month** | **Total Sales (in $)** |
| 1 | 1,813,586 |
| 2 | 2,202,022 |
| 3 | 2,807,100 |
| 4 | 3,390,670 |
| 5 | 3,152,607 |
| 6 | 2,577,802 |
| 7 | 2,647,776 |
| 8 | 2,244,468 |
| 9 | 2,907,560 |
| 10 | 3,736,727 |
| 11 | 3,199,606 |
| 12 | 4,614,443 |

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, indicating a peak in consumer spending and heightened retail activity during the holiday season.

d. Top sales city in Dibb’s best year of sales:

The following table illustrates total sales in each city in 2019.

|  |  |
| --- | --- |
| **City** | **Total 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 |

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:

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

A graph of order by hour

Description automatically generated

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.

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, driving up the frequency of these items being bought together.

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

|  |  |
| --- | --- |
| **Product** | **Quantity Ordered** |
| AAA Batteries (4-pack) | 31,017 |
| AA Batteries (4-pack) | 27,635 |
| USB-C Charging Cable | 23,975 |
| Lightning Charging Cable | 23,217 |
| Wired Headphones | 20,557 |
| Apple Airpods Headphones | 15,661 |
| Bose SoundSport Headphones | 13,457 |
| 27in FHD Monitor | 7,550 |
| iPhone | 6,849 |
| 27in 4K Gaming Monitor | 6,244 |

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 best year of sales:

With a total order of **646 orders**, the **LG Dryer** had the lowest number of orders in Dibbs' 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, less frequent replacement cycles, or possible market competition offering more appealing alternatives.

## Chapter 3: Data Visualization

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

1. Monthly Sales Trends versus 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:

A graph of a number of months

Description automatically generated with medium confidence

**Monthly Sales Trend:**

* 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.

A graph with numbers and lines

Description automatically generated

**Average Monthly Sales:**

* 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.
* Businesses may be more focused on closing out their fiscal year and finalizing budgets during this period, potentially leading to lower sales activity.
* The slight recovery towards the end of the year aligns with the holiday shopping season, as observed in the Monthly Sales Trend chart.

Possible underlying reasons behind the sales patterns:

* 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.
* The end of the fiscal year for US businesses may contribute to a temporary lull in sales during September and October as companies prioritize year-end activities.

b. Total Sales by State:

A graph with numbers and bars

Description automatically generated

The total sales by state analysis reveals that California dominates with the highest sales figures, while Maine records the lowest sales. This disparity could be attributed to California's larger population, higher economic activity, and greater market reach, which provides a broader customer base for Dibbs. In contrast, Maine's smaller population and less robust economic environment likely contributed to its lower sales.

c. Top 10 Products in the best year of Sales:

A graph with numbers and text

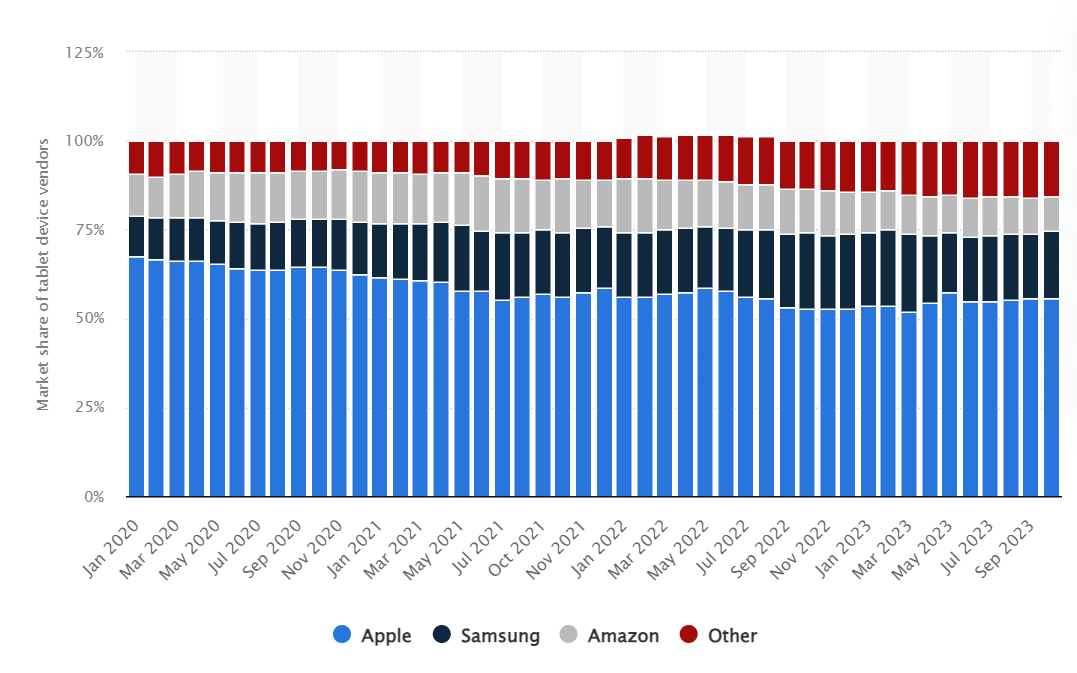
Description automatically generated

In the banner sales year of 2019, the top 10 products sold were predominantly mobile devices and electronics from leading brands like Apple, Google, and Microsoft. The MacBook Pro laptop emerged as the highest-selling item, generating around $8 million in sales, showcasing consumers' strong preference for premium mobile devices and Apple's brand dominance in driving Dibbs' success. This trend underscores the importance of stocking in-demand, popular products to fuel sales growth.

A graph of sales and sales

Description automatically generated with medium confidence

However, in 2020 and 2021, Apple's MacBook Pro was no longer among the top 10 products, being replaced by the ThinkPad laptop from Lenovo. This shift can be attributed to the disruptions caused by the COVID-19 pandemic on Apple's supply chain, which heavily relies on manufacturing in China (Tom Coughlin, 2020). As a result of production shortages and supply constraints, consumers likely turned to alternative laptop brands like ThinkPad to meet the surging demand for electronics driven by the transition to remote work and online activities during the pandemic (Boston, 2021). This led to a temporary loss of market share for Apple during this period of heightened consumer demand.



**STATISTICS #1:** Market share of leading tablet device vendors in the United States from January 2020 to October 2023

d. Monthly Order vs Monthly Average:

A graph of a number of blue bars

Description automatically generated with medium confidence

The monthly order trend analysis shows a clear seasonal pattern with fluctuations around the average of 17,428 orders. December stands out as the peak period with 28,137 orders, suggesting a significant surge in consumer demand during the holiday season. Conversely, January recorded the lowest order volume at 10,003 units. Notable spikes above the average are observed in April, May, October, November, and December, implying an increase in orders during specific periods, potentially driven by factors such as new product launches, promotional events, and the year-end holiday shopping frenzy.

e. Daily Order vs Daily Average Order:

A graph of a trend

Description automatically generated with medium confidence

The daily order trend graph shows that the average daily order volume is around 6,746 units. From days 1 to 27, orders slightly fluctuated around this average. However, a significant drop in orders was observed after day 27, with the lowest volume recorded on the 31st day. This decline could be attributed to the end of monthly promotional cycles or consumer spending habits, where purchases taper off as consumers await new month's promotions or paycheck arrivals.

A graph of a number of blue and white bars

Description automatically generated with medium confidencef. Hourly Order vs Hourly Average Order:

The hourly order trend reveals an average of 8,714 orders per hour, with significant variations throughout the day. Order volumes exceeded this average from 10 AM to 10 PM, peaking at 7 PM with approximately 14,500 units. This peak coincides with consumers' post-work shopping activities, suggesting that evenings are prime time for online shopping. The lowest order volumes were recorded between 3 AM and 4 AM when most consumers are asleep.

# Task 4: Building a predictive model for Dibs organization to be able to predict future sales.

1. **Data Preparation**

To begin, even after the extensive data cleaning and preprocessing performed in Task 1, the additional data preparation steps taken in Task 4 are still redone as necessary due to the following:

* 1. **Date Filtering:** The code filters out rows where the Order.Date year is NA or 2020. This step is important because it ensures that the data used for model building and forecasting is consistent and relevant to the problem at hand.
  2. **Calculating Total Sales:** The code introduces a new column **Total.Sales**, which is calculated as **Quantity.Ordered \* Price.Each**. This column represents the total sales amount for each transaction, which is a crucial variable for sales forecasting models.
  3. **Aggregating Daily and Weekly Sales:** Task 4 aggregates the data to daily and weekly levels, which are more suitable for time series analysis and forecasting. This step is essential for capturing temporal patterns and trends in the data.
  4. **Time Series Conversion**: The code converts the aggregated weekly sales data into a time series format, which is a prerequisite for applying time series forecasting models like the linear trend model and Holt-Winters model.

1. **Data Splitting**

The dataset should be split into training (80%) and testing (20%) sets, where the training set is used to build and refine the predictive model, and the testing set is used to evaluate the model's forecasting accuracy for informing sales and marketing strategies.

1. **Time series analysis**

To begin, let’s look at sales overtime (daily and weekly).

A graph showing a line graph

Description automatically generated

**A graph showing a line graph

Description automatically generated**

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**.

The forecasting models will be evaluated using appropriate accuracy metrics, and the best-performing model will be selected for implementation. The forecasted sales figures can then be used to optimize marketing campaigns, inventory management, resource allocation, and other strategic initiatives to align with anticipated sales patterns and trends.

1. **Forecasting**

To build predictive models that can inform the company's sales and marketing strategies, the techniques of Linear Regression and Holt-Winters Exponential Smoothing were employed.

Linear Regression is suitable when the sales data exhibits a linear trend over time and when explanatory variables like advertising spend or economic indicators can be included to improve accuracy. It offers interpretability by quantifying the effects of predictor variables. However, it assumes linearity, normality of residuals, and constant variance.

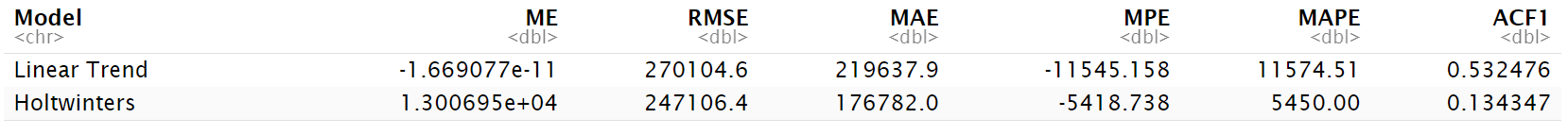
Holt-Winters' method is advantageous when sales data shows both trend and seasonal patterns, which are common in many businesses. It can handle additive or multiplicative seasonality and adapt to changes in level, trend, or seasonality. Unlike ARIMA models, it doesn't require data transformations, making it more straightforward.

A graph showing a graph of a graph

Description automatically generated with medium confidence

A graph showing a graph showing a number of different types of data

Description automatically generated with medium confidence

Accuracy Table:

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.

1. **Connected Story**

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.A graph with blue rectangles

Description automatically generated

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.

# Task 5: Sales and Marketing Recommendations

Our recommendations for Dibs are based on the SAVE (Solution, Access, Value, Education) framework, which aims to be comprehensive and well-aligned with the insights better compared to tradition 4Ps. To some extent, SAVE is considered to be more customer-centric () than the tradition product-centric 4Ps ().

|  |  |
| --- | --- |
| Solution:   * Positioning offerings as solutions for remote work, online learning, and digital lifestyles by curating product bundles/packages * Tailoring recommendations based on customer segments (professionals, students, families) * Given the dominance of Apple's iPad in the tablet market, with a market share of around 55.9% in the United States as of October 2023, it would be a strategic move for Dibs to incorporate iPads into their product lineup. By offering iPads, Dibs can cater to the strong consumer preference for Apple's products and tap into the brand's loyal customer base. | Access:   * Expanding distribution channels through online marketplaces * Addressing non-response bias by increasing market penetration in underserved areas * Enhancing online sales and e-commerce capabilities. * In light of the increased adoption of online shopping and electronic payments during the pandemic, it is crucial for Dibs to 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. |
| Value:   * Communicating long-term value, durability, and future compatibility * Offering competitive pricing, bundles, and discounts * Leveraging customer reviews, testimonials, and industry ratings * Providing extended warranties, support services, and trade-in programs | Education:   * Developing educational content (buying guides, product comparisons, tutorials) * Leveraging social media, blogs, and video platforms for content sharing * Offering virtual consultations and in-store demonstrations * Collaborating with industry experts, influencers, and thought leaders |

Meanwhile, Dibs can also consider:

**Capitalize on Stimulus Check Spending:** Given the observed boost in consumer spending following stimulus check issuance, Dibs should closely monitor and align their marketing campaigns and promotional strategies with the timing of stimulus payments. This could include targeted offers, discounts, or limited-time deals to capture the increased spending potential during these periods.

**Monitor and Adapt to Evolving Consumer Behaviors:** As consumer behaviors continue to evolve in the post-pandemic landscape, Dibs should remain agile and responsive to emerging trends. Regularly analyzing customer data, conducting market research, and monitoring industry reports will enable Dibs to stay ahead of the curve and adapt their strategies accordingly, ensuring they remain relevant and aligned with the ever-changing consumer preferences and behaviors.

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

References:

Boston, F. R. B. of. (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

Tom Coughlin, T. (2020). *Impact of COVID-19 on the Consumer Electronics Market*.