Case Study 7

# Lip Products in Indonesia

A Comprehensive Study

| | |
|---|---|
| Arka Mukhopadhyay | B23120 |
| Pranab Ray | B23169 |
| Kamal Yadav | B23209 |
| Arani Ghosh | B23119 |
| Ayuj Aryan | B23198 |
| Kunal Sharma | B23079 |

**March 28, 2024**

# Contents

# List of Tables

# List of Figures

# 1 Introduction

## 1.1 Background

Lips are often the unsung heroes of our skincare routine. With thinner skin, fewer oil glands, and no natural protection from the elements, lips are prone to dryness, chapping, and cracking. To keep them looking and feeling their best, it's essential to incorporate lip care into your daily routine. Lip products are one of the most important parts of makeup. It is even considered as the most used beauty product in the world.

> *"There are no right or wrong guidelines when it comes to lip color"* – Clarissa Luna, a celebrity makeup artist in New York.

With a wide array of lip beauty products to choose from, finding the perfect one for you can be quite challenging. Dermatologists advise protecting your lips from the sun with lipsticks with at least SPF 15.

There are a lot of Lip products and Brands available all over the world. We would like to analyze the specification of the Lip products that are available in Indonesia's shade, brands, type and the prices that they are offering.

## 1.2 Objective

This research aims to determine several indicators of lip products in Indonesia based on population and sample gathered from secondary data which are:

1. To identify the nature of data provided for lip products.
2. To identify the mean, maximum and minimum price of a popular lip products in Indonesia.
3. To identify the popular brands of lip products in Indonesia.
4. To identify the mean shade diversity available for lip products.
5. To identify the common type of Lip Product in Indonesia.
6. To represent and visualize the data in a more understandable way.

# 2 Data and Statistics

## 2.1 What is Data?

Merriam Webster describes **Data** as factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation.

Before a problem is analyzed, all the information available must be converted into data. **Measurement** in the systemic process of assigning numbers to objects and their properties to facilitate the use of mathematics in studying and describing objects and their relationships.

## 2.2 Types of Data

### 2.2.1 Based on Source of Data

1. **Primary Data**: This type of data is collected firsthand by the researcher or investigator directly from the source. It involves gathering data through methods like surveys, interviews, observations, experiments, etc. Primary data is original and specific to the research or study at hand.

2. **Secondary Data**: Secondary data refers to data that has already been collected by someone else for a different purpose. This data is obtained from sources such as books, journals, government publications, websites, databases, etc. Secondary data analysis involves using existing data to derive insights or conclusions.

3. **Tertiary Data**: Tertiary data is derived from primary and secondary sources. It involves the aggregation, compilation, and analysis of primary and secondary data to create new datasets or information. Tertiary data is often used for market research, trend analysis, and decision-making processes.

**There are several benefits of using secondary data**:

- It is cost-effective being readily available and accessible at lower costs/for free.
- Using existing data eliminates the need for conducting new research, allowing researchers to analyze data immediately.
- It contributes to a better understanding of the problem.
- It serves as a foundation for comparing the data gathered by the researcher across different time periods and geographies; thereby facilitating trend analysis and benchmarking.

**However, there are also disadvantages of using secondary data**:

- Secondary data rarely fits within the framework of marketing research factors since researchers have limited control over methods of collection, processing and categorization of data.
- The quality, precision and reliability of secondary data is unknown.
- Data may be out of date.

### 2.2.2 Based on Levels of Measurement

**Quantitative data** consists of numerical or measurable values. It is typically collected through structured methods such as surveys, experiments, or measurements. Quantitative data can be analyzed statistically to identify patterns, trends, and relationships. [1]

Subcategories of Quantitative Data include:

1. **Discrete Data**: It comprises distinct, separate values that can be counted individually. Examples include the number of students in a class, the number of cars in a parking lot, etc.
2. **Continuous Data**: Continuous data represents measurements that can take any value within a range. It is typically obtained through instruments like scales, thermometers, or rulers. Examples include height, weight, temperature, and time.
3. **Interval Data**: Data that can be added or subtracted but not multiplied/divided. They do not have a true zero point. For example: temperature, year, etc.
4. **Ratio Data**: Data that can be added, subtracted, multiplied or divided. They have a true zero point. For example: height, weight, age and so on.

Data need not be inherently numeric to be useful in an analysis. For instance, male and female both are commonly used in almost any statistics report involving population but there is nothing numeric about these categories. This category of data is known as **Qualitative Data**.

Statisticians commonly distinguish two types of Qualitative Data:

1. **Nominal Data**: Categorical data without any inherent order or hierarchy. The categories are purely distinct labels or names. For example: types of fruits, colors, types of transportation, etc.
2. **Ordinal Data**: Categorical data with a natural order or hierarchy. While the categories have a meaningful sequence, the differences between them may not be uniform. Examples include education levels (e.g., high school, bachelor's degree, master's degree) or ratings.

### 2.3 What is Statistics?

Statistics is the science of data. This involves collecting, classifying, summarizing, organizing, analyzing, and interpreting data. It involves methods for designing experiments and surveys, gathering data, and drawing conclusions from that data. Statistics is widely used in various fields such as science, business, economics, engineering, social sciences, and many others. It helps in making informed decisions, predicting outcomes, testing hypotheses, and understanding patterns and trends in data.
There are two kinds of Statistics:

1. **Inferential Statistics**: Statistical inference is the science of characterizing or making decisions about a population by using information from a sample drawn from that population. This includes hypothesis testing, confidence intervals, and regression analysis.

2. **Descriptive Statistics**: Descriptive statistics uses data that provides a description of the population either through numerical calculated graphs or tables. It provides a graphical summary of data. It includes:

    - Measures of Central Tendency (Mean, Median and Mode)
    - Measures of Variability (Range, Variance, Dispersion, and so on)

# 3 Sampling

### 3.1 Key Terminologies

1. **Population**: The population refers to the entire group of individuals, objects, or events who represent a characteristic. A *census* study involves the entire population.

2. **Sample**: A sample is a subset of the population selected for observation or measurement.

3. **Sampling**: Sampling is the process of selecting a *sample* from the *population* to make statistical inferences and estimate population characteristics.

4. **Sampling Frame**: A sampling frame is a list of all the individuals, objects, or events in the population from which the sample will be selected.

### 3.2 Sampling Schemes

A good sample must reflect all the characteristics (of importance) of the population. A sample that accurately reflects its population characteristics is called a *representative sample.* A sample that is not representative of the population characteristics

is called a *biased sample.* The reliability or accuracy of conclusions drawn concerning a population depends on whether or not the sample is properly chosen so as to represent the population sufficiently well. [2]
The selection of a sampling method depends on factors such as the nature of the investigation, the availability of sampling frames (lists of population members), financial resources, desired accuracy level, and data collection method (e.g., questionnaires or interviews).
Common sampling techniques include:

1. **Simple Random Sample**: A sample selected in such a way that every element of the population has an equal chance of being chosen is called a simple random sample.

2. **Systematic Sampling**: A systematic sample is a sample in which every $k^{th}$ element in the sampling frame is selected after a suitable random start for the first element with the population listed in some defined order.

3. **Stratified Sample**: Here, a sample obtained by stratifying (dividing into non-overlapping groups) the sampling frame based on some factor(s) and then selecting some elements from each of the strata. A population with N elements is first divided into 's' sub-populations, then a sample is drawn from each subpopulation independently.

4. **Cluster or Area Sampling**: In cluster sampling, the sampling unit contains naturally existing groups of elements called clusters instead of individual elements of the population. A cluster is an intact group naturally available in the field.

## 3.3 Bias in Sampling

Sampling bias refers to the systematic error introduced into a sample as a result of the sampling method. It occurs when some members of a population are systematically more likely to be selected in a sample than others, leading to inaccurate or misleading conclusions and limits the generalizability of the findings. [1] [3]
The following are some common types of sampling biases:

1. **Selection Bias**: Selection bias exists if some potential subjects are more likely than others to be selected for the study sample; usually due to the sampling process.

2. **Volunteer Bias**: Volunteer bias refers to the fact that people who volunteer to be in studies are usually not representative of the population as a whole. For this reason, results from entirely volunteer samples might be considerably different from those who do not volunteer.

3. **Non-Response Bias**: Non-response bias occurs when individuals selected for the sample do not respond to the survey or study. This can lead to underrepresentation of certain groups in the sample, skewing the results.

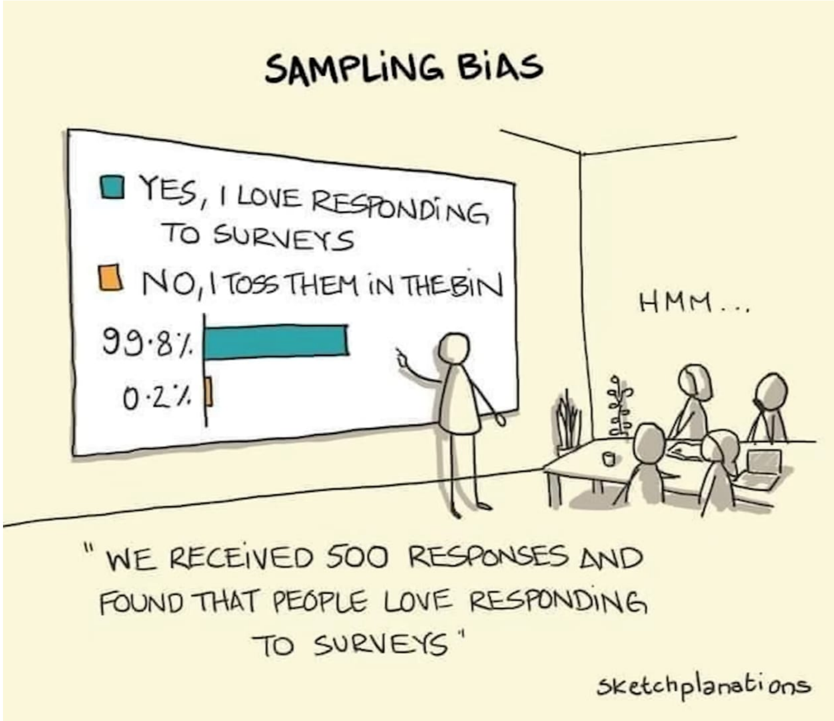

SAMPLING BIAS
🟦 YES, I LOVE RESPONDING TO SURVEYS
🟧 NO, I TOSS THEM IN THE BIN
99.8%
0.2%
HMM...
"WE RECEIVED 500 RESPONSES AND FOUND THAT PEOPLE LOVE RESPONDING TO SURVEYS"
sketchplanations

Figure 1: Example of Sampling Bias

# 4 Methodology

## 4.1 Data Collection

Considering the gigantic dataset pertaining to lip products in Indonesia, secondary data is used in our study in order to minimize time spent collecting the same otherwise.

The initial data for this study was obtained from the sample of 50 lip products given in a prior report [4]. The data given was sourced from 2022 reports, and hence, outdated. We used latest datasets to update 2 parameters - number of shades available in the market and the price of the lipsticks.

To find the latest and credible data, we have used the maximum retail price (MRP) of the product as mentioned on the company's official Indonesian Website. Note that there are some companies which don't have country specific websites (or don't have an operational official website at all). In such cases, we have used Lazada/Shoppee to collect the data. Why these websites? These two websites are leading e-commerce websites in southeast Asia and Taiwan and hence are most credible source that could have been used.

## 4.2 Data Analysis

This study utilizes stratified sampling to determine the sample by dividing the population into subgroups based on factors like brands and product types. Also, the data that is used in this study is nominal in nature since it has categories without necessarily implying mathematical order. Libraries such as matplotlib, numpy, seaborn, and pandas provide easy means to analyze and visualize datasets.

The composition of lip products by type (solid, crayon, stick, and liquid) is examined, with details displayed on bar and donut charts. Brand distribution is analyzed using the 'Brand' column, showing the count and percentage of products per brand on a bar plot.

Statistical measures such as mean, median, mode, and quartiles are calculated for prices and the number of shades. Price distribution is visualized through a histogram and box plot, aiding in identifying price ranges and distribution. A bar graph displays the distribution of products based on the number of shades they offer.

Kernel Density Estimation (KDE) is used on histograms to create smooth curves representing the data distribution. This enhances visualization by revealing underlying patterns and providing a refined understanding of price and shade distribution across the dataset.

Color mapping is applied to visualize frequency levels, with lower frequencies represented in blue and higher frequencies in red. Lastly, data is grouped by companies to determine the total number of shades offered by each company, providing insights into shade diversity among different brands.

# 5 Raw Data

Table 1: Raw Data for Study (2023)

| Lip Product | Brand | Type of Lip Product | Shades | Price |
|---|---|---|---|---|
| NIVEA LIP BALM SOOTHE & PRTECT | Beiersdorf | Stick | 2 | 50000 |
| Extra lip tint | Bobbi Brown | Stick | 10 | 711636 |
| Perfect Matte Lip Coat | Dear Me Beauty | Liquid | 6 | 129000 |
| Creamytint | Emina | Liquid | 5 | 46000 |
| magic potion lip tint | Emina | Liquid | 5 | 50000 |
| Squeeze me up Lip Matte | Emina | Liquid | 4 | 58000 |
| Smoochies Lip balm | Emina | Solid | 1 | 32000 |
| Matte Lip Liquid | ESQA | Liquid | 7 | 165000 |
| Dear Darling Water gel tint | Etude House | Liquid | 3 | 55000 |
| Organic lip balm | Eucalie | Stick | 1 | 79000 |
| lip and cheek dual use liquid | Focallure | Liquid | 10 | 38000 |
| Melted Matte Lip | Goban Cosmetics | Liquid | 6 | 130000 |
| Sheen. Tinted lip balm + UV filter | HALE. | Stick | 4 | 98000 |
| Urban Lip Cream Matte | Implora | Liquid | 20 | 25000 |
| Beauty Lip & Cheeck Crayon | Indoganic | Crayon | 2 | 129000 |
| Vivid oil tint | Innisfree | Liquid | 4 | 104000 |
| Metallic Lip Cream | Inul Beauty | Liquid | 5 | 89000 |
| Infalible Pro Matte Lip Liquid | L'oreal | Liquid | 9 | 150000 |
| Rouge Signature Liquid Matte Lipstick | L'oreal | Liquid | 14 | 151376 |
| Color Riche Matte | L'oreal | Stick | 8 | 354267 |
| Intense Matte Lip Cream | Liquid | Liquid | 12 | 119000 |
| Longlasting Matte Lip Cream Metalic | LT Pro | Liquid | 3 | 109900 |
| Ultra Light Lip Stain | Luxcrime | Liquid | 8 | 79000 |
| Airy lip mousse | Luxcrime | Liquid | 8 | 109000 |
| Dew tinted 6hr lip moisturizer | Mad for Makeup | Stick | 6 | 109000 |
| magnifique lip tint | Madame Gie | Liquid | 6 | 33000 |
| Brilliant Glaze Lip Liquide | Madame Gie | Liquid | 6 | 35000 |
| Moist Velvet & Smooth Lip Liquide | Madame Gie | Liquid | 6 | 15765 |
| Hydrastay lip whip | Makeover | Liquid | 12 | 119000 |
| Powestay Transfer Proof Matte Lip Cream | Makeover | Liquid | 12 | 135000 |
| Sensational Liquid Matte | Maybelline | Liquid | 19 | 66,023 |
| color sensational lip tint | Maybelline | Liquid | 19 | 45000 |
| Super Stay Matte Ink | Maybelline | Liquid | 19 | 239571 |
| Color sensational the powder mattes | Maybelline | Stick | 24 | 88900 |
| Hydra Lip Cheek Tint | Mineral Botanica | Liquid | 4 | 51900 |
| the one A-Z lip balm SPF 25 | Oriflame | Stick | 2 | 149000 |
| Lip Cream | PIXY | Liquid | 16 | 55000 |
| 2 in 1 color tint | Purbasari | Liquid | 3 | 51900 |
| Lip Cream Series | Raiku | Liquid | 13 | 118000 |
| SUEDED! Lip & Cheek Cream | Rollover Reaction | Liquid | 12 | 109000 |
| Juicy Lip Balm | Rose All day | Stick | 3 | 119000 |
| Lip Color | Runa Beauty | Stick | 5 | 138000 |
| Lip Care | Sensatia Botanica | Liquid | 5 | 80000 |
| Coconut lip sleeping balm | Tiff Body | Liquid | 1 | 88000 |
| delight tony tint | Tony Moly | Liquid | 3 | 49000 |
| Exclusive Matte Lip Cream | Wardah | Liquid | 24 | 66500 |
| Colorfit Velvet Matte Lip Mousse | Wardah | Liquid | 14 | 79000 |
| Everyday Moisture Lip nutrition | Wardah | Stick | 2 | 28500 |
| Color Fit Ultralight Matte | Wardah | Stick | 5 | 47500 |
| The Simplicity Love You tint | Y.O.U | Liquid | 4 | 45100 |

# 6 Sampled Data

## 6.1 On Basis of Brands

Table 2: Products grouped by Brand

| Brand | Frequency | Percentage | Cumulative Percentage |
|---|---|---|---|
| Beiersdorf | 1 | 2.0 | 2.0 |
| Bobbi Brown | 1 | 2.0 | 4.0 |
| Dear Me Beauty | 1 | 2.0 | 6.0 |
| Emina | 4 | 8.0 | 14.0 |
| ESQA | 1 | 2.0 | 16.0 |
| Etude House | 1 | 2.0 | 18.0 |
| Eucalie | 1 | 2.0 | 20.0 |
| Focallure | 1 | 2.0 | 22.0 |
| Goban Cosmetics | 1 | 2.0 | 24.0 |
| HALE. | 1 | 2.0 | 26.0 |
| Implora | 1 | 2.0 | 28.0 |
| Indoganic | 1 | 2.0 | 30.0 |
| Innisfree | 1 | 2.0 | 32.0 |
| Inul Beauty | 1 | 2.0 | 34.0 |
| L'oreal | 3 | 6.0 | 40.0 |
| Liquid | 1 | 2.0 | 42.0 |
| LT Pro | 1 | 2.0 | 44.0 |
| Luxcrime | 2 | 4.0 | 48.0 |
| Mad for Makeup | 1 | 2.0 | 50.0 |
| Madame Gie | 3 | 6.0 | 56.0 |
| Makeover | 2 | 4.0 | 60.0 |
| Maybelline | 4 | 8.0 | 68.0 |
| Mineral Botanica | 1 | 2.0 | 70.0 |
| Oriflame | 1 | 2.0 | 72.0 |
| PIXY | 1 | 2.0 | 74.0 |
| Purbasari | 1 | 2.0 | 76.0 |
| Raiku | 1 | 2.0 | 78.0 |
| Rollover Reaction | 1 | 2.0 | 80.0 |
| Rose All day | 1 | 2.0 | 82.0 |
| Runa Beauty | 1 | 2.0 | 84.0 |
| Sensatia Botanica | 1 | 2.0 | 86.0 |
| Tiff Body | 1 | 2.0 | 88.0 |
| Tony Moly | 1 | 2.0 | 90.0 |
| Wardah | 4 | 8.0 | 98.0 |
| Y.O.U | 1 | 2.0 | 100.0 |

### 6.2 On Basis of Type

Table 3: Products grouped by Type

| **Type** | **Frequency** | **Percentage** | **Cumulative Percentage** |
|---|---|---|---|
| Stick | 12 | 24.0 | 24.0 |
| Liquid | 36 | 72.0 | 96.0 |
| Solid | 1 | 2.0 | 98.0 |
| Crayon | 1 | 2.0 | 100.0 |

### 6.3 On Basis of Shades

Table 4: Products grouped by Shades

| **Shades** | **Frequency** | **Percentage** | **Cumulative Percentage** |
|---|---|---|---|
| 1 | 3 | 6.0 | 6.0 |
| 2 | 4 | 8.0 | 14.0 |
| 3 | 5 | 10.0 | 24.0 |
| 4 | 5 | 10.0 | 34.0 |
| 5 | 6 | 12.0 | 46.0 |
| 6 | 6 | 12.0 | 58.0 |
| 7 | 1 | 2.0 | 60.0 |
| 8 | 3 | 6.0 | 66.0 |
| 9 | 1 | 2.0 | 68.0 |
| 10 | 2 | 4.0 | 72.0 |
| 12 | 4 | 8.0 | 80.0 |
| 13 | 1 | 2.0 | 82.0 |
| 14 | 2 | 4.0 | 86.0 |
| 16 | 1 | 2.0 | 88.0 |
| 19 | 3 | 6.0 | 94.0 |
| 20 | 1 | 2.0 | 96.0 |
| 24 | 2 | 4.0 | 100.0 |

# 7 Graphs and Stats

Table 5: Descriptive Statistics of Shades and Prices

| **Attribute** | **Shades** | **Price** |
|---|---|---|
| Count | 50 | 50 |
| Mean | 8.04 | 104456.76 |
| Median | 6 | 84000 |
| Mode | 5, 6 | 79000, 109000, 119000 |
| Std Dev | 6.11 | 105480.23 |
| Variance | 37.3 | 11126079540.23 |
| Minimum | 1 | 15765 |
| Maximum | 24 | 711636 |
| **Percentiles:** | | |
| - 0th | 1 | 15765 |
| - 25th | 4 | 50000 |
| - 50th | 6 | 84000 |
| - 75th | 12 | 119000 |
| - 100th | 24 | 711636 |

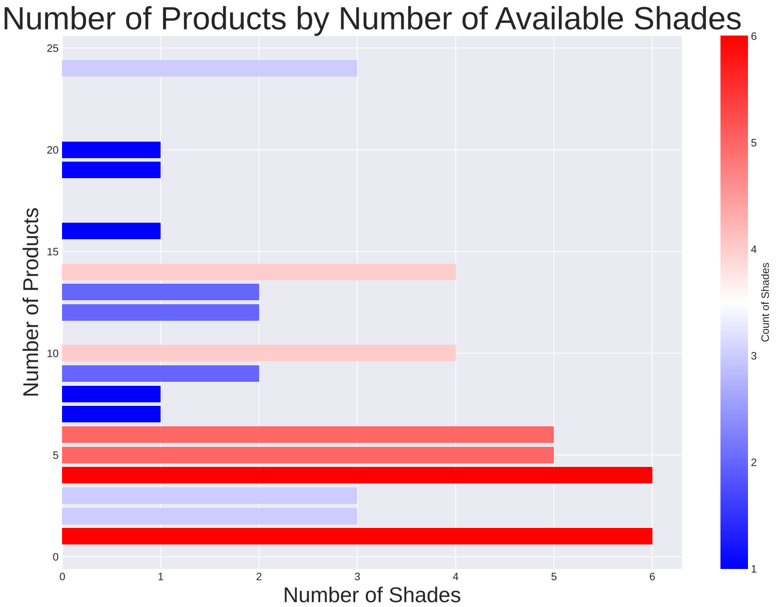


Figure 2: Total Products Grouped by Shades

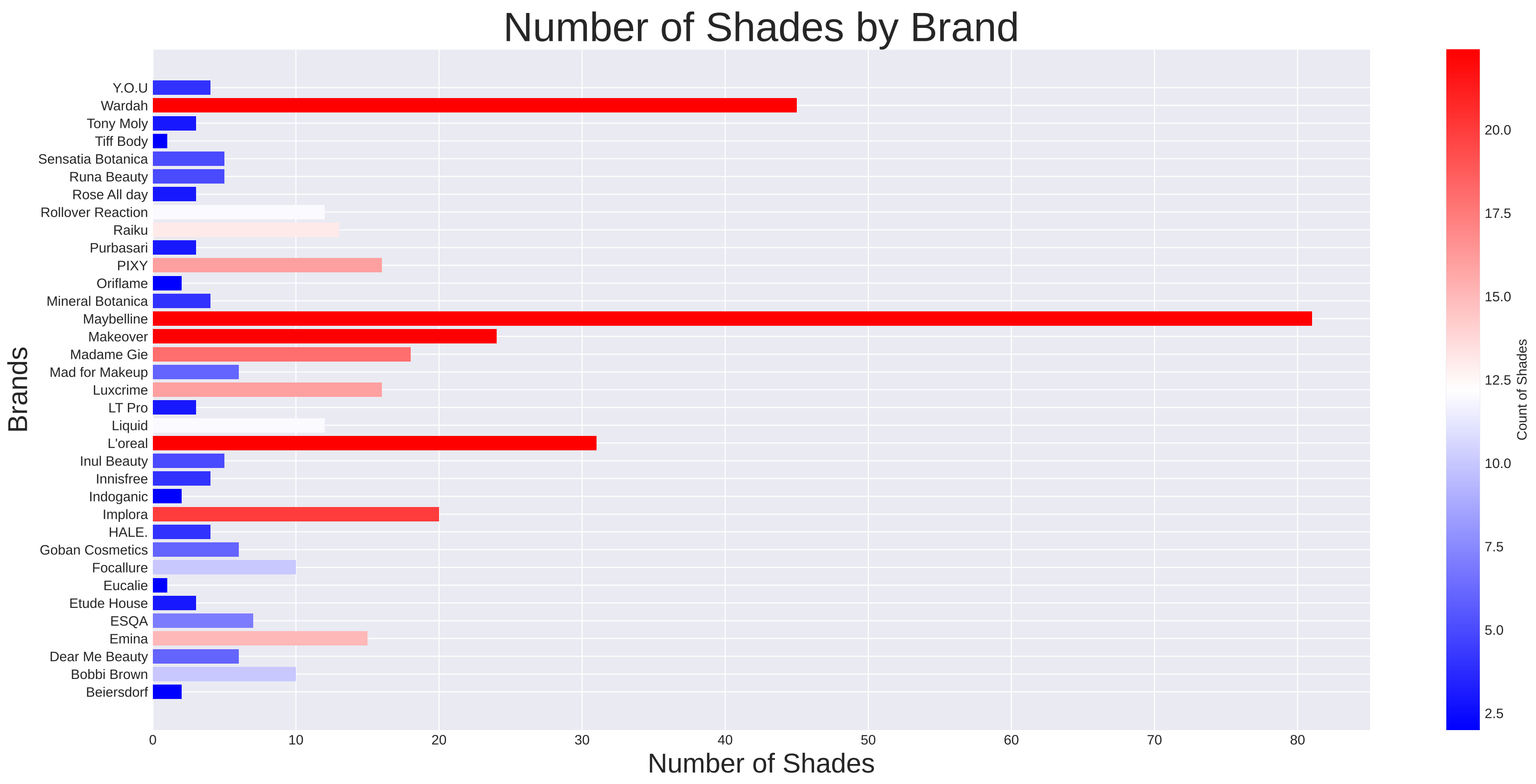


Figure 3: Total Shades Grouped by Brands

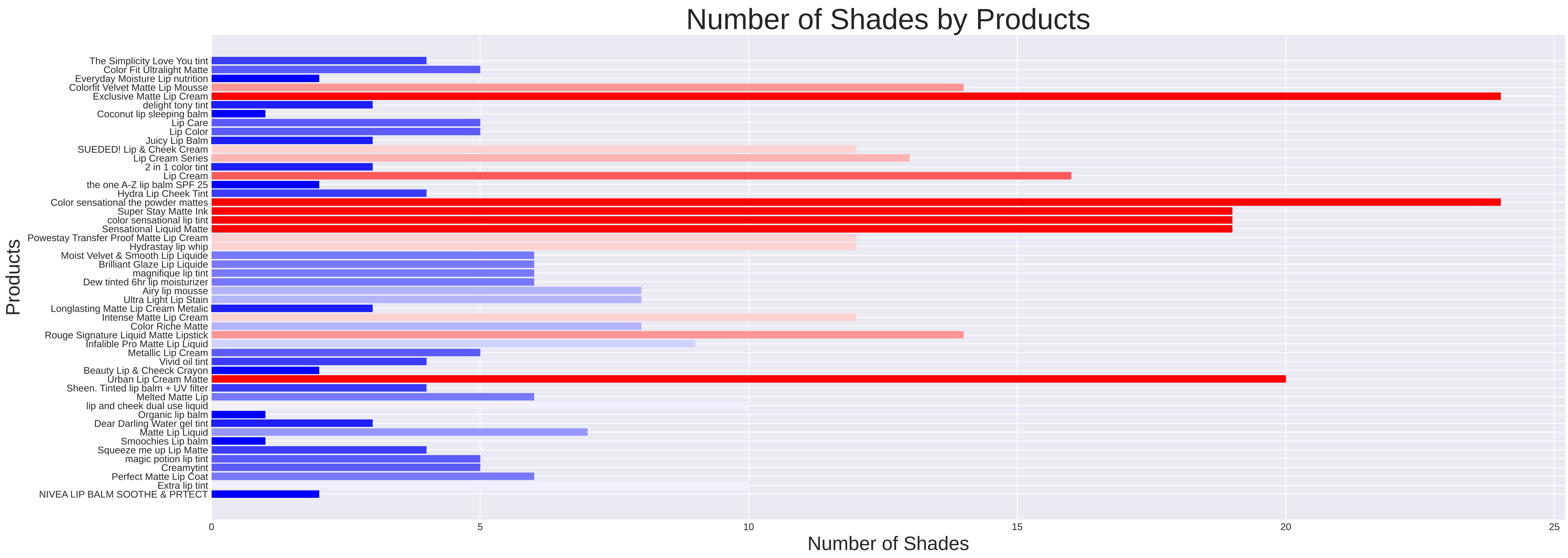
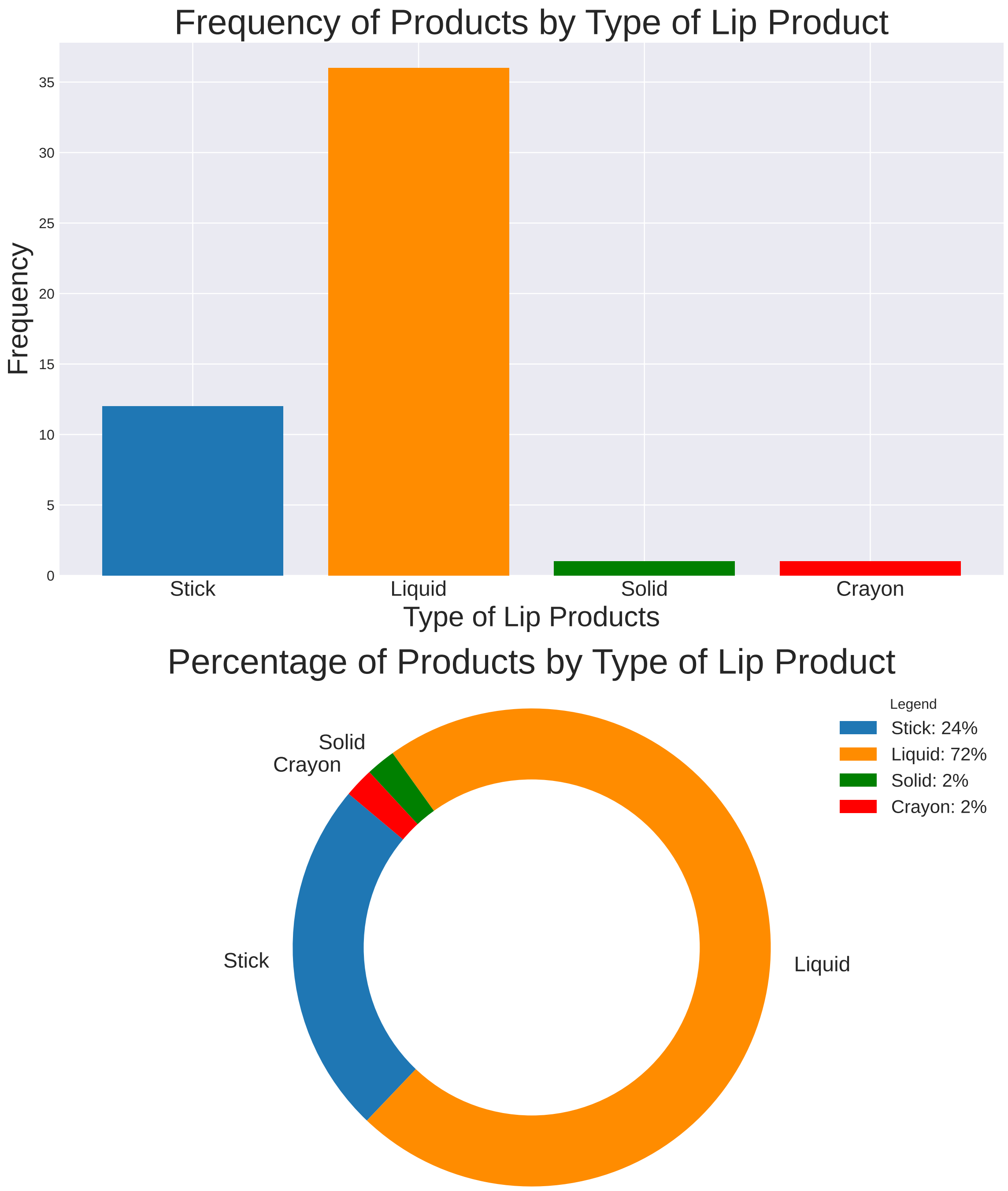


Figure 4: Total Shades Grouped by Product



Frequency of Products by Type of Lip Product

Percentage of Products by Type of Lip Product

Figure 5: Total Products Grouped by Type

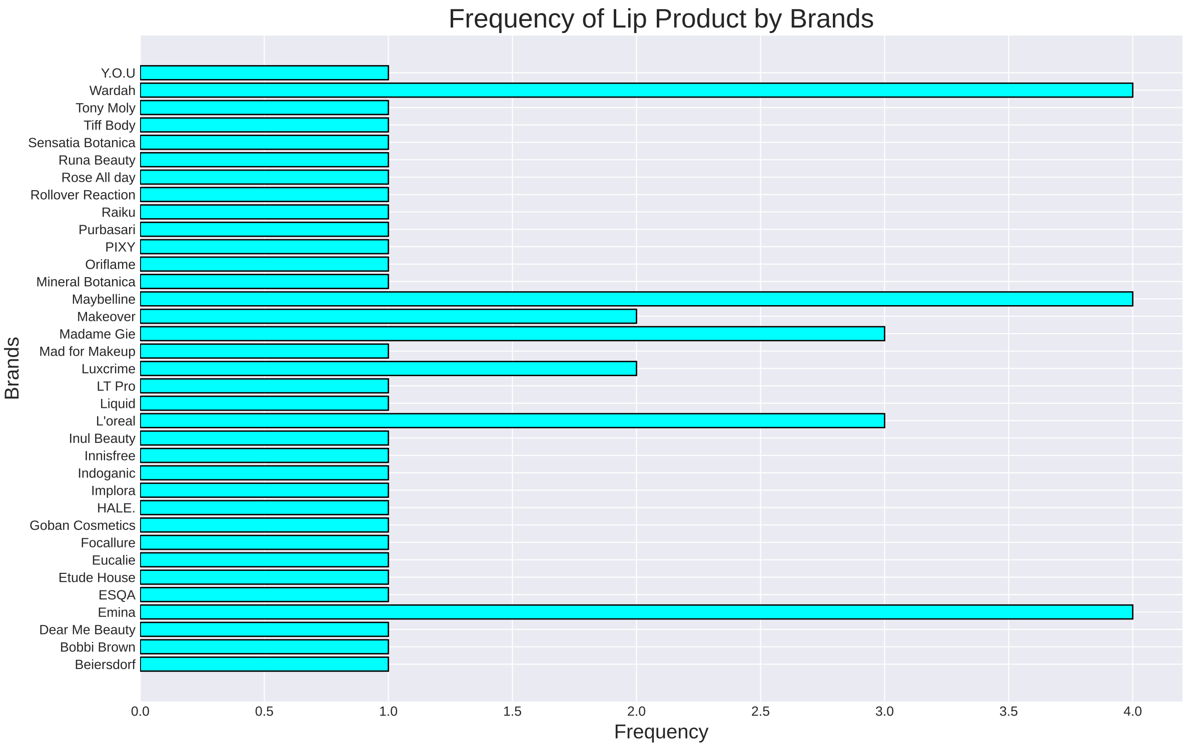


Figure 6: Frequency of Products by Brands

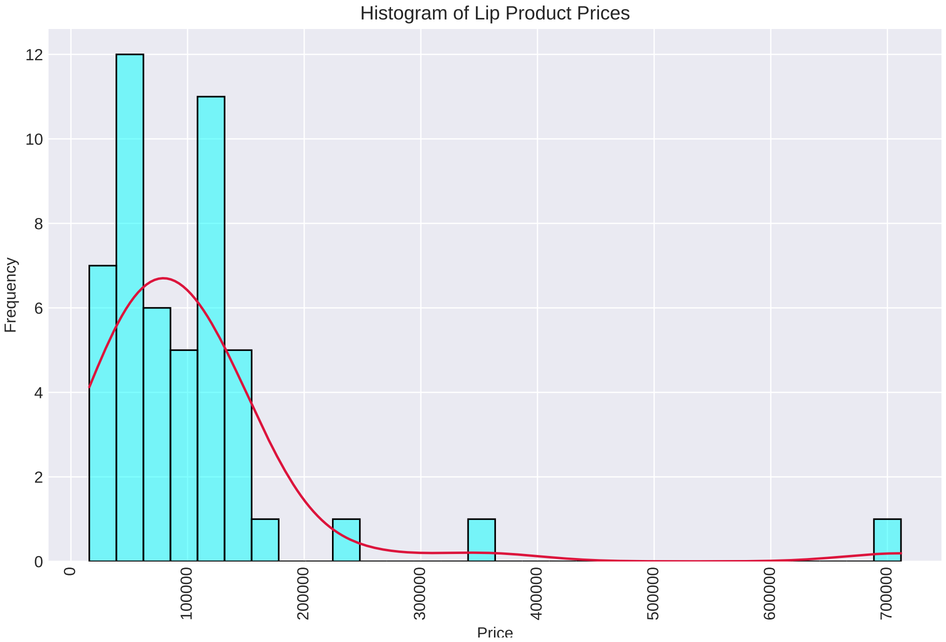


Figure 7: Product - Price Distribution - with KDE

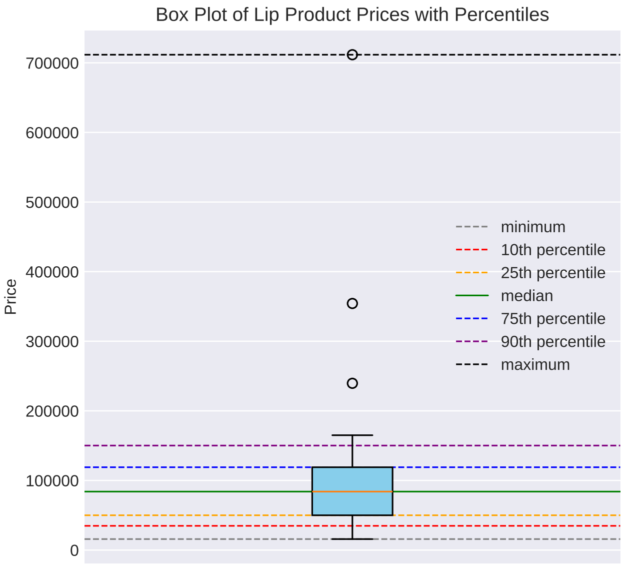


Figure 8: Product - Price Distribution - BoxPlot

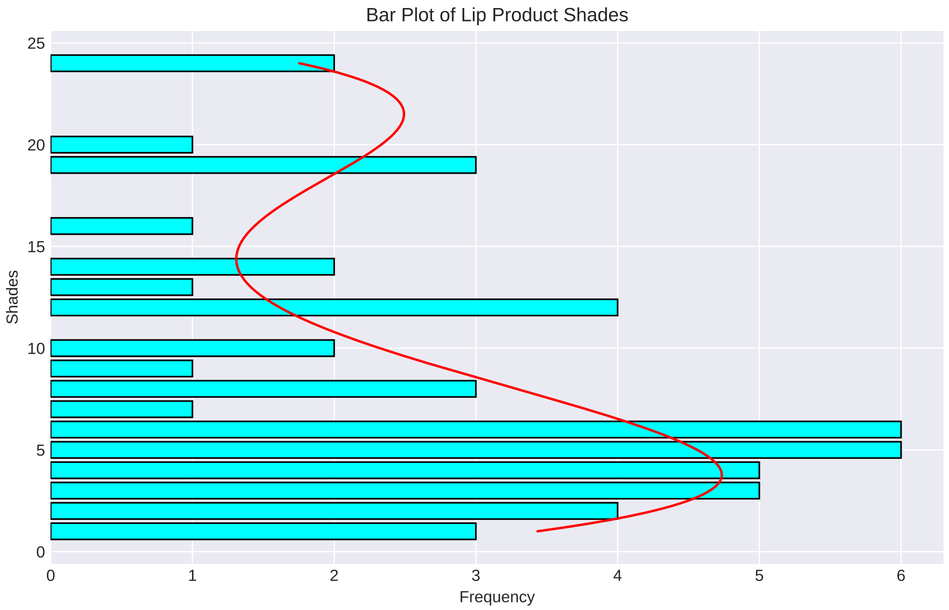


Figure 9: Product - Shades Distribution - with KDE

# 8 Inference

Most of the products are present in the price range of 15000 to 175000. Some of the products lie outside this range and can be seen as occasional peaks in the bar plot. The maximum price of any product is 711636 and the minimum is 15765. The mean price of all products is 104456 and the median is 84000. The data also shows a standard deviation of around 105480.

The number of shades per product varies from 1 to 24, with the mean being 8.04 and the median being 6. Also, companies like Wardah, Maybelline, Makeover, L'oreal and Implora boast a large number of shades in their products while other companies do not have as much variety.

Both variables have relatively high standard deviations compared to their means, indicating wide variability when it comes to prices and shade counts around their respective averages.

Outliers are present in both variables, especially considering the large difference between the 75th percentile and the maximum values for both price and shades.

Most of the lip products produced by these brands fall under the 'liquid' category, as evident from the donut chart. It accounts for 72 percent of the total products. From the remaining, 24 percent are 'stick' type and solid and crayon account for 2 percent each.

Also, most companies only have one product present in the data, with some exceptions being Emina, L'oreal, Luxcrime, Madame Gie, Makeover, Maybelline and Wardah which have 2 or more products each. Emina, Wardah and Maybelline form the most popular brands in Indonesia closely followed by Madame Gie and L'oreal.

# References

[1] Sarah Boslaugh. *Statistics in a Nutshell.* O'Reilly Media, 2nd edition, 2012.

[2] Chris P. Tsokos and Kandethody M. Ramachandran. *Mathematical Statistics with Applications.* Academic Press, 1st edition, March 2009.

[3] Pritha Bhandari. Sampling bias and how to avoid it | types & examples, 2020.

[4] Dety Nurfadilah and Yulita F Susanti. *Case Study for Descriptive Statistics.* Ipmi Press, 2022.

# Contributions

Table 6: Contributions of the authors

| Name | Roll No. | Contribution in Report Writing | Contribution in Analysis | Details of use of web resources/Codes/ AI tools, etc. | Overall Contribution to work done |
|---|---|---|---|---|---|
| Arka Mukhopadhyay | B23120 | | | | |
| Pranab Ray | B23169 | | | | |
| Kamal Yadav | B23209 | | | | |
| Arani Ghosh | B23119 | | | | |
| Ayuj Aryan | B23198 | | | | |
| Kunal Sharma | B23079 | | | | |