**TABLE OF CONTENT**

1. What is Statistics ?
2. Summary
3. Introduction
4. Analysis
5. Visualization
6. Interpretation
7. Reference

**1) WHAT IS STATISTICS?**

**A basic area of mathematics called statistics is concerned with gathering, analysing, and interpreting data. It is crucial in many domains, including business, medicine, social sciences, and engineering, as it facilitates the discovery of patterns, trends, and connections in data. We may efficiently summarize data, make predictions, and direct decision-making processes by utilizing statistical tools.**

**Early techniques for doing activities like taking censuses and maintaining records may be found in the early civilizations that gave rise to statistics. Originating from the Latin word "status," the name "statistics" refers to its original function in state government and administration. With the analysis of mortality statistics and the development of core principles by pioneers like as John Graunt, the science started to take shape in the 17th century.**

**Mathematicians like Blaise Pascal and Pierre-Simon Laplace made significant contributions to the subject of probability theory development. Significant contributions were made in the 18th and 19th centuries by individuals like as Sir Francis Galton, who introduced ideas like regression and correlation, and Carl Friedrich Gauss, who invented the least squares approach. Many of the statistical approaches utilized today have their roots in these achievements. The development of digital technology and computers over the 20th century and beyond resulted in a significant transformation. These developments allowed for more intricate analyses and the management of enormous volumes of data, which completely changed statistical methods.**

**These days, statistics and data science are closely intertwined, with statistics driving innovations and offering insights across a wide range of areas. With its historical foundations and contemporary uses, statistics is now a vital tool for comprehending and using data in many different domains.**

**2) SUMMARY**

This study looked at personal care product sales information from many Indian cities. Conclusion is concluded by calculating:

**1) CENTRAL TENDENCY**

* **MEAN**
* **MEDIAN**
* **QUARTILE**
* **5 NUMBER SUMMARY**

**2) MEASURES OF DISPERSION**

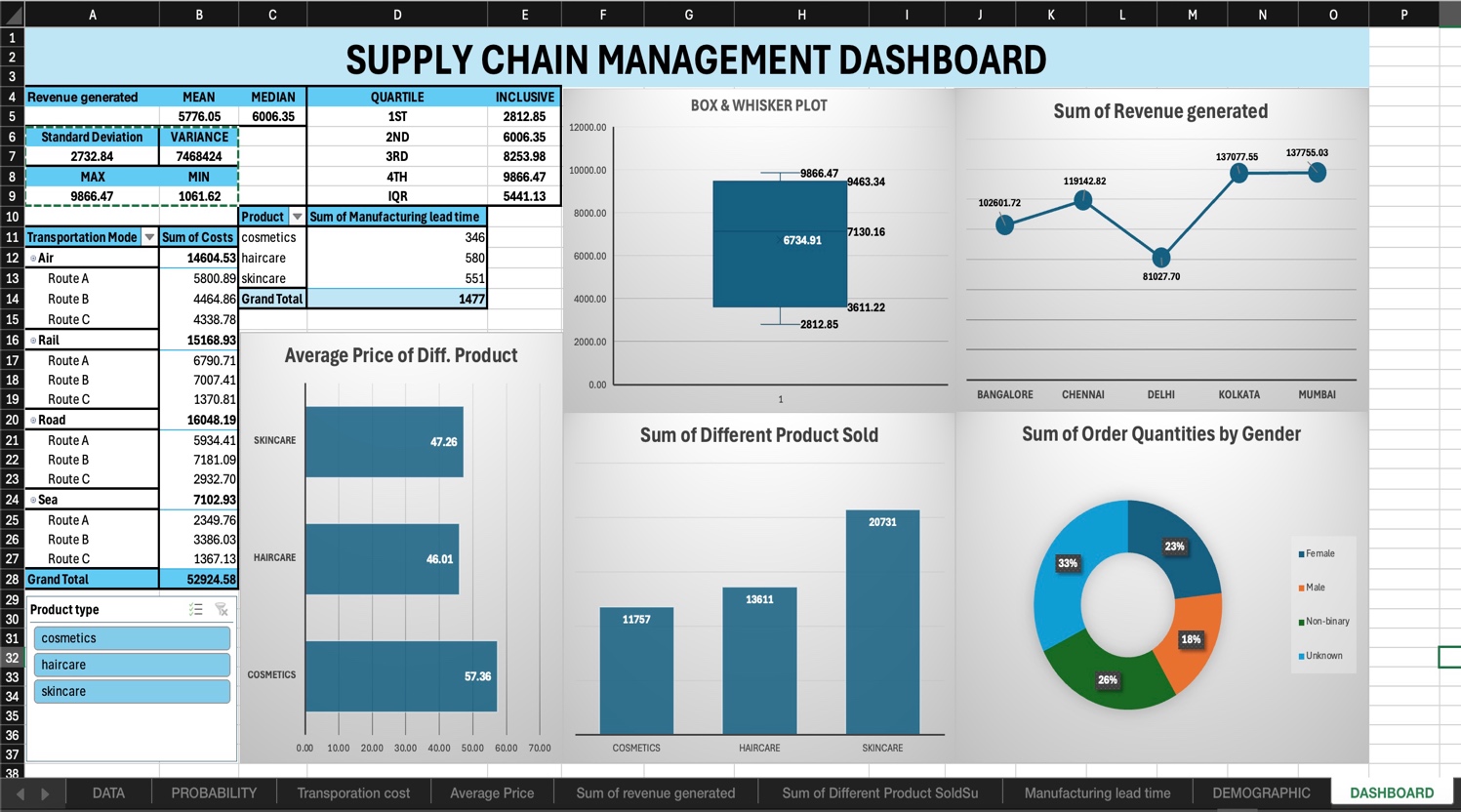
* **INTERQUARTILE RANGE**
* **STANDARD DEVIATION**
* **VARIANCE**

**3) CHARTS/ GRAPHS**

* **BAR CHARTS**
* **LINE CHARTS**
* **PIE CHARTS**
* **BOX WHISKER GRAPH**

**4) PROBABILITY**

**3) INTRODUCTION:**

The sales statistics of personal care products—cosmetics, haircare, and skincare—across different routes and places in India are analysed in this research. Understanding customer preferences, sales performance, and regional market dynamics is important since it may help with inventory management and corporate strategy. A thorough picture of the market environment is provided by the dataset, which contains complete data on the quantity of items sold, product categories, pricing, availability, and client demographics and routes of transportation

**4) ANALYSIS**

The following are the primary characteristics of the dataset:

**Product Type:** Cosmetics, skincare, and haircare are among the categories.

**Number of Products Sold:** The total amount of money sold at various places for each sort of product.

**Price:** The amount each thing is sold for.

**Availability:** Every product's stock levels.

**Location:** Various cities, including Delhi, Bangalore, Chennai, Kolkata, and Mumbai.  
Statistical Synopses

**Total Items Sold via Each Route:**

Route A: 19,778

Route B: 18,370

Route C: 7,951

Grand total: 46,099

**Best-Selling Areas:**

Kolkata: 7,522 products sold

Delhi: 3,020 products sold

Mumbai: 3,390 products sold

**5) VISUALIZATION:**

Bar Chart: This graph emphasizes the dominance of sale of skincare product by showing the total quantity of each product sold (CHART 1)

Line Graph With Markers : This graph emphasizes on the total sum revenue generated by each city.(Chart 2)

Bar Chart : This graphs shows the average cost of each product ( cosmetics, skincare , haircare) (CHART 3)

Whisker Plot Graph : This graphs shows the 5 number summary which includes quartile , quartile 2 , quartile 3 , quartile 4 and maximum and minimum value of each product category.

(CHART 4)

Pie Chart : This graph emphasis on the purchase of quantities of goods ordered by each gender in overall supply chain. (CHART 5)

**6) Interpretation:**

There are some important insights from the given data:

**Mean (Average):**

Interpretation: It is also possible to culminate the idea of the mean with reference to the total amount of all the revenue indicators that have been collected, divided by the number of such indicators.

Purpose: It provides the simple mathematical average of the values of the revenues, but like all averages it can be very much distorted by one or two values which are much higher or lower than the normal and therefore not necessarily representing an average revenue.

**Median:**

Interpretation: The median simply refers to the middle point when all the piece of data in revenue has been arranged. If the number of observation is odd then the median is the observation in the middle and if the number of observation is even then the median is the contention and if the number of observations is even, then the median is the mean of the two middle observations.

Purpose: Median is again more suitable measure of central tendency than mean in case the data set contains odd number or it is skewed or uses extreme values; hence provides the real picture of ‘average’ revenues.

|  |  |  |
| --- | --- | --- |
| **Revenue generated** | **MEAN** | **MEDIAN** |
|  | **5776.05** | **6006.35** |

**Quartiles:**

Interpretation: This splits data into four equal parts called quartiles where the data points are placed with relation to its ranking. That is the first quartile, second quartile or median while the third quartile is Q3 at the 75 percent or 3rd quartile.

Purpose: For their part, quartiles assist in also establishing the dispersion of data or information. Q1 inform the revenue down to which are 25% of data located and Q3 provide the rate down to which are 75% of data located. Studying Q3 and Q1 the measure known as IQR = Q3 – Q1 offers the spread of the middle half of the data set and brings out the absence or actual presence of outliers in a simpler way.

|  |  |
| --- | --- |
| **QUARTILE** | **INCLUSIVE** |
| **1ST** | **2812.85** |
| **2ND** | **6006.35** |
| **3RD** | **8253.98** |
| **4TH** | **9866.47** |
| **IQR** | **5441.13** |

**Standard Deviation:**

Interpretation: Standard Deviation is used to indicate just how much deviation there is from the mean. These are: It is important underlining that a small value of the standard deviation indicates that the values are close to its average, in opposition, a high value indicates that the values are spread out.

Purpose: Knowledge in standard deviation is also useful in establishing the stability or otherwise of any number of revenues. A large standard deviation suggests significant revenue fluctuations, potentially indicating business instability or seasonal variations.

**Variance:**

Interpretation: Variance is the square of standard deviation, providing a measure of how much revenue values spread from the mean. Though variance is less interpretable in terms of units, it’s fundamental in measuring data spread.

Purpose: Variance helps understand overall data dispersion, useful in statistical models and calculations that describe variability.

|  |  |
| --- | --- |
| **Standard Deviation** | **VARIANCE** |
| **2732.84** | **7468424** |
| **MAX** | **MIN** |
| **9866.47** | **1061.62** |

**Why These Measures Are Important: ( Central Tendency and Measure of Dispersion)**

Decision Making: These statistics summarize revenue data, aiding stakeholders in understanding overall performance and revenue variability, crucial for budgeting, forecasting, and trend identification.

Risk Assessment: Variability measures, like standard deviation and variance, are essential for assessing financial risk. High revenue variability may indicate potential financial instability.

Identifying Outliers: Quartiles and the median help detect outliers or extreme values that could influence overall revenue trends, important for quality control or anomaly detection.

Understanding Data Distribution: Together, these measures provide a comprehensive view of revenue distribution, enabling more informed analysis and decision-making.

**Chart 1 ( Bar Chart )**

The bar charts help us to identify few basic insights:

1. Category Comparison : It help us to identify which product is performing better in sales volume. Skincare performed the best as the number of product sold in skincare category is the maximum (20731)

SKINCARE > HAIRCARE > COSMETICS

1. Sales Distribution : It helps in determining if the company's sales are more evenly distributed throughout several categories or if it is relying on just a few of important categories.

**Chart 2 (Line Graph )**

The line graph emphasis on:

1. Revenue trends across cities : It helps in identifying how revenue varies across different cities. Each marker on line graph represents revenue generated in each city.

Mumbai has generated the highest revenue (Rs. 1337755.03)

1. Comparative Analysis : By reviewing revenue of each city we can provide relative actions to improve the performance of that city.

Delhi has generated lowest revenue (81027.70) adequate measures can be used to improve the performance of Delhi.

**Chart 3 (Bar Chart)**

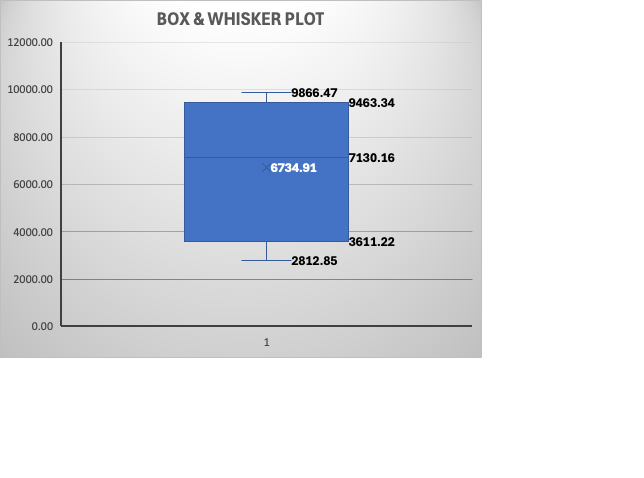
The chart emphasis on :

1. Price Strategy : The variations in average prices may throw light on the various categories' pricing strategies. Different pricing structures, like premium pricing for some categories and competitive price for others, may be indicated by significant variance.

The average price of cosmetic product is relatively high ( Rs. 57.36) which indicate it as quite premium product range.

1. Market Positioning : Average price of each category shows the positioning of product in market. Category of product with lower average value (Haircare Rs. 46.01) might be consider as every day and afforadable product.

**Chart 4 ( Whisker Plot Graph)**



**The graph emphasis on :**

The big 5 summary which shows the Q1 , Q2 , Q3 , Max and Min value in the plot of the data. It shows the outlier i.e. ( 9866.47) and the middle plot area represent the median value .

Quartile 1 represent 25% of the data i.e. 2812.85

Quartile 2 represents 50% of the data i.e. 6006.35

Quartile 3 represents 75% of the data i.e. 8253.98

Interquartile Range = Q3-Q1

IQR = 5441.13

**Chart 5 ( Pie Chart)**

The graph emphasised on the distribution of goods sold to each gender.

This helps in identifying the potential buyers as the data shows 33% of the buyers are unknown but it also shows 23% of female are the potential buyers . So, that we can identify the potential target audience.

**References:** https://www.kaggle.com/