

# PROJECT SUPPLY CHAIN ANALYSIS

DIGITAL EGYPT PIONEERS INITIATIVE OF THE MINISTRY OF  
COMMUNICATIONS AND INFORMATION TECHNOLOGY - DEPI



# OUR TEAM



Hassan  
Nasr



Mohamed  
Omar



Ahmed  
Abdel-Aziz



Mohamed  
Samy



Mohamed  
Ahmed



Amr  
Taher



# CONTENT

- 01** INTRODUCTION
- 02** DATA STAGES
- 03** ANALYSIS QUESTIONS PHASE USING PYTHON
- 04** ANALYSIS QUESTIONS PHASE USING SQL
- 05** VISUALIZATION DASHBOARD

# INTRODUCTION



**Supplier Performance Evaluation:** To identify the best and worst-performing suppliers based on the quality of materials they provide. This evaluation will allow us to make informed decisions regarding supplier selection and retention, ensuring that only high-quality materials are sourced to maintain production standards.



**Plant Efficiency in Defect Detection:** To determine which manufacturing plants are most effective at identifying and rejecting defective materials, thereby minimizing downtime. By pinpointing plants with strong defect detection processes, we can enhance overall production efficiency and reduce disruptions caused by poor-quality materials.

This project aims to analyze supply chain data to optimize product availability, sales performance, and operational efficiency. Key focus areas include pricing, stock levels, shipping costs, and defect rates to drive better decision-making and improve overall supply chain management.

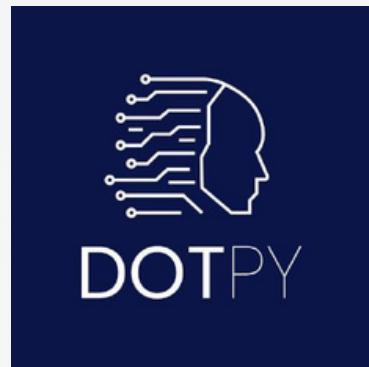


وزارة الاتصالات  
وتقنيهedia المعلومات



# DATA STAGES

*Data cleaning involves collecting raw data, identifying inconsistencies, and correcting errors such as duplicates or missing values. It also includes standardizing formats and transforming data to fit the desired structure for analysis. Finally, the cleaned data is validated to ensure it meets quality standards.*



# DATA STAGES

## Data Cleaning and Preprocessing

***Before proceeding with analysis, I conducted an initial assessment of the dataset to identify any potential issues, such as missing values, duplicates, and data inconsistencies. Below are the steps I took, along with the Python functions used for each task.***

### Data Overview

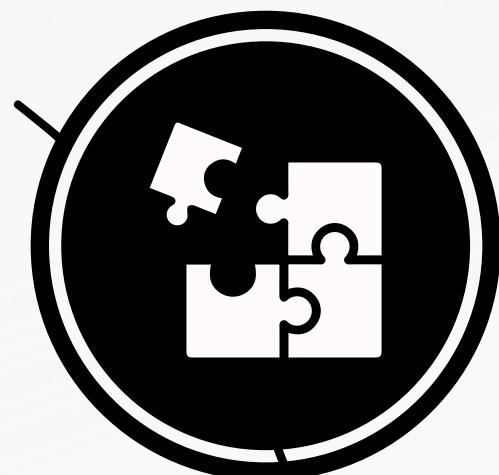
#### Checking for Null or Empty Values



/



#### Detecting Duplicate Records



/

# 1- DATA OVERVIEW

*To understand the structure of the dataset, I used the info() function in Python, which provided a summary of the dataset including:*

- Number of rows and columns
- Data types of each column (e.g., integers, floats, objects)
- The number of non-null values in each column

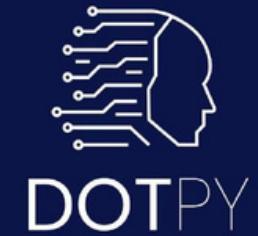
```
# To read data
data_file = pd.read_csv(r"F:\PROJECT\supply_chain_data.csv")
#Look at the sample of data ex. 5 rows
print(data_file.head())
```

			Product type	SKU	Price	...	Total Cost	Operational Profit	Margin
0	haircare	SKU0	69.808006	...	236.988527	8425.008266	97.264043		
1	skincare	SKU1	14.843523	...	546.398923	6914.501143	92.676501		
2	haircare	SKU2	11.319683	...	180.662780	9397.086845	98.113724		
3	skincare	SKU3	61.163343	...	292.130469	7474.705957	96.238746		
4	skincare	SKU4	4.805496	...	1019.396340	1667.108811	62.054927		
..	...	...	...	...	...	...	...	...	
95	haircare	SKU95	77.903927	...	846.385266	6539.978678	88.541246		
96	cosmetics	SKU96	24.423131	...	211.898776	7486.525990	97.247505		
97	haircare	SKU97	3.526111	...	613.802424	3757.114156	85.957123		
98	skincare	SKU98	19.754605	...	889.213356	7636.739204	89.570510		
99	haircare	SKU99	68.517833	...	250.126931	8935.058898	97.276844		

[100 rows x 27 columns]>



## 2-CHECKING FOR NULL OR EMPTY VALUES



To ensure the dataset was complete, I created a function to check for any missing or null values across all columns. Detecting null values early is critical for avoiding issues later in the analysis, such as inaccurate computations or visualizations.

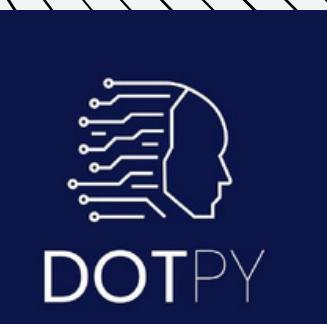
```
#Look at the missing of values
missing_data = data_file.isnull().sum()
print(missing_data)
```

```
[5 rows x 24 columns]
Product type      0
SKU               0
Price              0
Availability       0
Number of products sold 0
Revenue generated 0
Customer demographics 0
Stock levels       0
Lead times         0
Order quantities   0
Shipping times     0
Shipping carriers   0
Shipping costs      0
Supplier name       0
Location            0
Lead time           0
Production volumes 0
Manufacturing lead time 0
Manufacturing costs 0
Inspection results 0
Defect rates        0
Transportation modes 0
Routes              0
Costs                0
dtype: int64
```

**Result:** there's not any missing value in the data.



## 3-DETECTING DUPLICATE RECORDS



*Duplicate records can distort analysis results, so I wrote a function to check for duplicates. This ensured that each row in the dataset was unique and represented a distinct data point.*

```
#TO remove duplicates  
data_file.drop_duplicates(inplace=True)
```

*Result: Total duplicate rows: 0*



# ANALYSIS QUESTIONS PHASE USING PYTHON

*In the Analysis Questions phase using Python, data is explored by asking key questions and using Python libraries like Pandas and NumPy to extract insights. Visualizations and statistical methods help uncover trends and patterns. This phase drives decision-making by focusing on specific business or research objectives.*



# ANALYSIS QUESTIONS PHASE USING PYTHON

## 1- What is the percentage of defect rates for each product type using python?

**Tools Used:** Python libraries:

- Pandas Used for data manipulation and analysis.
- Matplotlib Used to create visualizations, including bar charts to represent sales across different regions

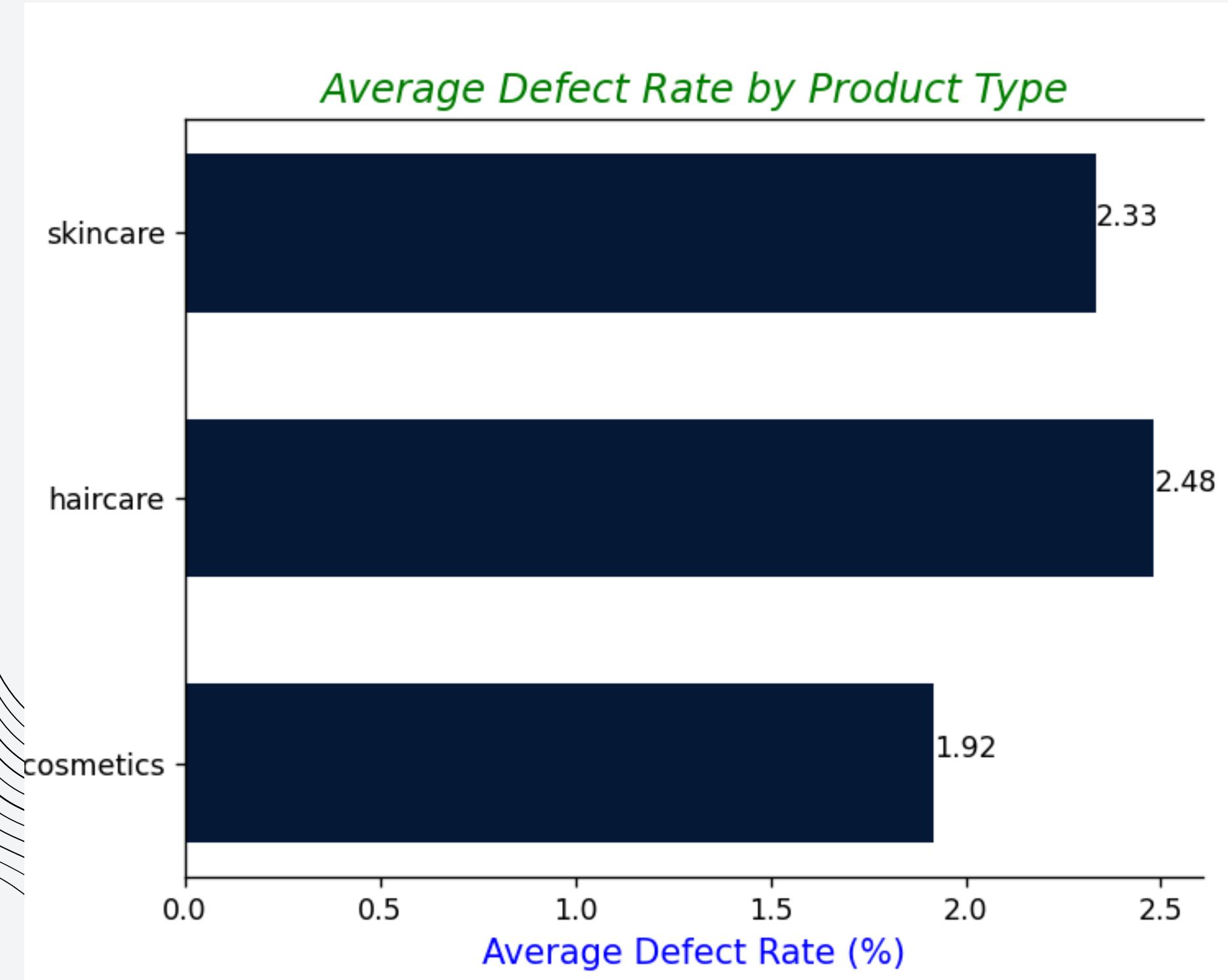
```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 # Read the CSV file
5 df = pd.read_csv(r"F:\PROJECT\supply_chain_data.csv")
6
7 # Select columns
8 loc = df[['Product type', 'Defect rates']]
9
10 # Group by 'Product type' and calculate the mean defect rates
11 calc = loc.groupby('Product type')['Defect rates'].mean()
12
13 # Create a barh chart with control in name title and size
14 plt.barh(calc.index,calc.values,color = '#091B39',height=0.6)
15 plt.title('Average Defect Rate by Product Type',fontsize=14, color = 'g',style = "italic")
16 plt.xlabel('Average Defect Rate (%)', fontsize=12, color='blue')
17
18 # Add text labels with defect rates formatted to 2 decimal places
19 for i , s in enumerate(calc):
20     plt.text(s, i ,f'{s:.2f}' , ha ='left' , va= 'bottom')
21
22 # To remove right gradline in chart So that the number does not interfere with the line
23 plt.gca().spines['right'].set_visible(False)
24
25 # Display the chart
26 plt.show()
```

### Steps in the Code :

- **Data Loading and Preparation:** The code imports necessary libraries and reads a CSV file containing product types and defect rates, selecting relevant columns and grouping them by product type to calculate the average defect rate for each product.
- **Chart Creation:** A horizontal bar chart is generated using matplotlib, with customized colors, reduced bar height for closer spacing, and gridlines added to improve readability.
- **Customization and Formatting:** The code adds a suitable title, "Average Defect Rate by Product Type," formats axis labels, ensures text labels show only two decimal places, and removes chart borders (spines) for a clean, modern look.

# ANALYSIS QUESTIONS PHASE USING PYTHON

**Chart :**



## **Results Interpretation:**

The chart effectively communicates which product types have higher or lower defect rates, enabling stakeholders to identify areas for improvement in quality control or production processes.

By visually comparing the lengths of the bars, users can easily assess the relative performance of each product category at a glance.

The clear formatting and labeling allow for quick insights, which is essential for data-driven decision-making in a supply chain context.



# ANALYSIS QUESTIONS PHASE USING PYTHON



## 2- How sales differ across various regions? Using Python

**Tools Used:** *Python Libraries:*

- Pandas: Used for data manipulation and analysis.
- Matplotlib: Used to create visualizations, including bar charts to represent sales across different regions.

### Steps and Process

#### 1- Data Loading:

The sales data was imported from an csv file into Python using Pandas.

```
5  
4 df = pd.read_csv(r"D:\DEPI\project\supply_chain_data_updated.csv")  
5
```

#### 2- Select column:

The relevant columns for analysis were Location (Region) and Number of Products Sold.

```
5  
6 # Selecting the column of the region and number of sales  
7 location = df[['Location', 'Number of products sold']]  
8
```

## 3- Analysis Questions Phase:

The data was grouped by Location (Region), and the total number of products sold in each region was summed using Pandas' groupby() and sum() functions.

```
9 # Group by 'Location' and sum the 'Number of products sold'  
10 region = location.groupby('Location')['Number of products sold'].sum()
```

## 4- Data Visualization:

A bar chart was created using Matplotlib to visualize how the number of products sold varies across different regions. This visualization helps identify which regions have the highest and lowest sales figures.

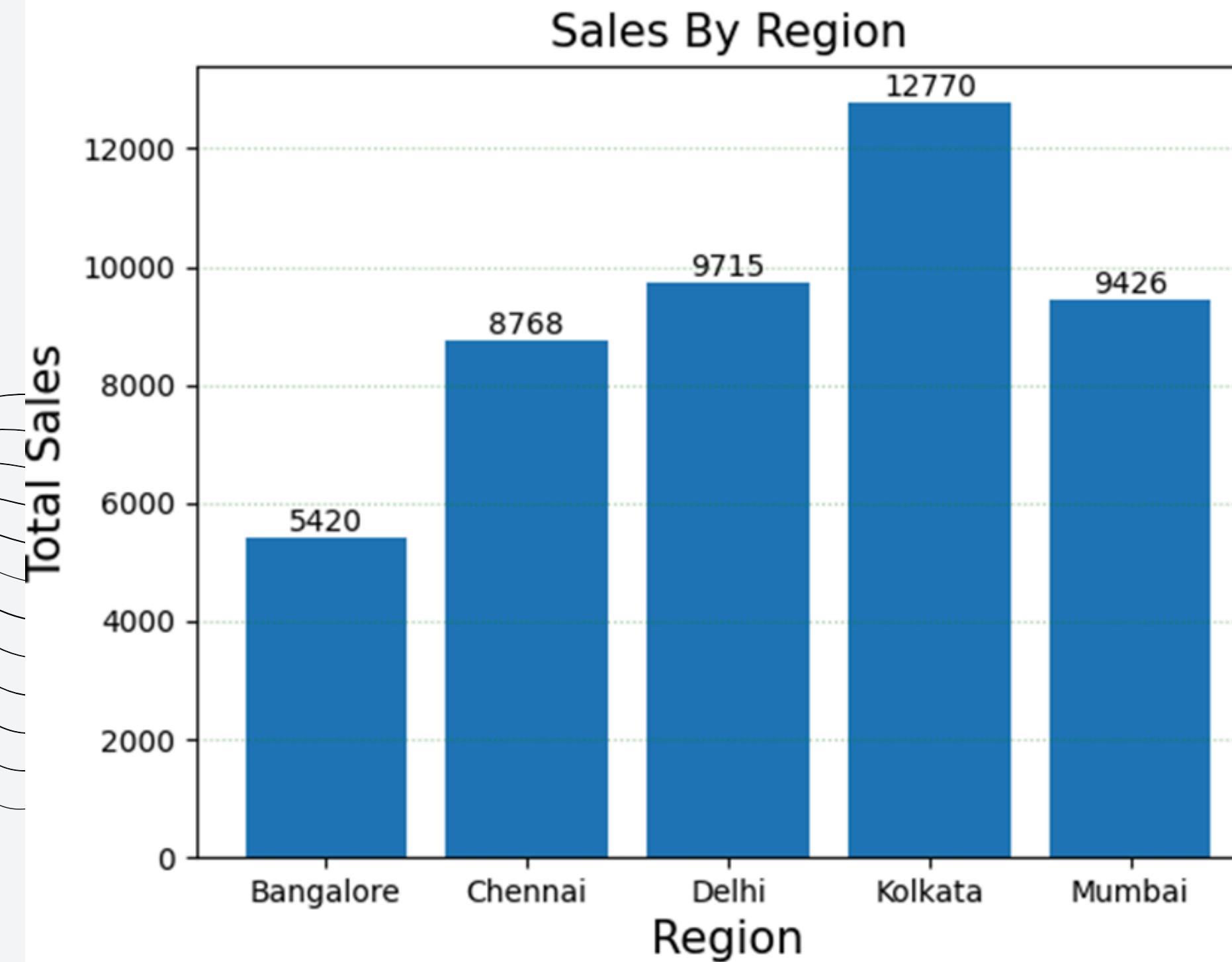
```
12 # Create bar chart to visualize the result  
13 plt.bar(region.index,region.values)  
14 plt.grid(axis='y',color='g',alpha = 0.35, linestyle = ':')  
15 plt.title('Sales By Region',fontsize=15)  
16 plt.xlabel('Region',fontsize=15)  
17 plt.ylabel('Total Sales',fontsize=15)  
18 for ind,vl in enumerate(region):  
19     plt.text(ind,vl,vl,ha='center',va='bottom')  
20  
21 plt.show()  
22
```



# ANALYSIS QUESTIONS PHASE USING PYTHON



## Findings from the Bar Chart



- **Insight:**

The bar chart provided a clear representation of how product sales differed across regions.

- **Observations:**

Kolkata recorded the highest number of product sales with (12770).

Bangalore had significantly lower sales with (5420), suggesting potential opportunities for improvement or further investigation.

Other regions showed moderate sales, with some variability that may warrant deeper analysis.



# ANALYSIS QUESTIONS PHASE USING SQL



*How sales differ across various regions? Using SQL*

```
SQLQuery2.sql - D...JQMGHT\DELL (62)*  X  SQLQuery1.sql - D...JQMGHT\DELL (77))  
use Supply_Chain  
  
select Location ,sum(Number_of_products_sold)  
from supply_chain_data  
group by Location
```

100 % ◀

Results Messages

	Location	(No column name)
1	Bangalore	5420
2	Chennai	8768
3	Delhi	9715
4	Kolkata	12770
5	Mumbai	9426

Query executed successfully.

***SQL is a powerful language used for querying and managing data in relational databases. By using SQL, analysts can extract, filter, and aggregate data, enabling detailed analysis of patterns, trends, and relationships within datasets. It helps transform raw data into meaningful insights for decision-making and reporting.***



# ANALYSIS QUESTIONS PHASE USING PYTHON



*How sales differ across various regions? Using Tableau*





# ANALYSIS QUESTIONS PHASE USING SQL

*In the Analysis Questions phase using SQL, data is queried to extract relevant insights through filtering, aggregation, and joining tables. SQL functions help identify trends, patterns, and relationships in the dataset. This phase focuses on answering specific business or research questions for informed decision-making.*





# ANALYSIS QUESTIONS PHASE USING SQL



## *First Question: Total Of Revenue by Revenue*

-----Total of Revenue by products-----

```
-select Product_type,sum(Revenue_generated) as 'Total Revenue'  
from supply_chain_data  
group by Product_type  
order by 'Total Revenue'
```

	Product_type	Total Revenue
1	cosmetics	161521.265999483
2	haircare	174455.390605462
3	skincare	241628.162133063

- The Result show that the most profitable is Skincare Products***
- And the least profitable is Cosmetics Products***



# ANALYSIS QUESTIONS PHASE USING SQL



## Second Question: Effect Of Gender on Products

-----Total of Product by Gender-----

```
select Customer_demographics ,count(Product_type) as 'Total Number Of products'  
from supply_chain_data  
group by Customer_demographics  
order by 'Total Number Of products'
```

	Customer_demographics	Total Number Of products
1	Male	21
2	Non-binary	23
3	Female	25
4	Unknown	31

- *The Female is the most using of these products than Male, And the Unknown Customer is most than both.*

## Third Question: Products Across Locations

-----Products Accross Locations -----

```
select Location ,Product_type,count(Product_type) as 'Total Number Of products'
from supply_chain_data
group by rollup (Location,Product_type)
order by 'Total Number Of products'
```

	Location	Product_type	Total Number Of products
6	Chennai	haircare	6
7	Bangalore	skincare	6
8	Delhi	cosmetics	6
9	Mumbai	haircare	7
10	Mumbai	skincare	7
11	Mumbai	cosmetics	8
12	Kolkata	haircare	8
13	Bangalore	haircare	9
14	Chennai	skincare	9
15	Kolkata	skincare	13
16	Delhi	NULL	15
17	Bangalore	NULL	18
18	Chennai	NULL	20
19	Mumbai	NULL	22
20	Kolkata	NULL	25
21	NULL	NULL	100

- **The Kolkata is the most city of using products,**
- **The Delhi is the least city of using products**



# ANALYSIS QUESTIONS PHASE USING SQL



## Fourth Question: Status Of Products

```
----- Status Of Products -----  
select Product_type,Inspection_results,count(Inspection_results) as 'Number Of Product Status'  
from supply_chain_data  
group by Product_type,Inspection_results  
order by 'Number Of Product Status'
```

	Product_type	Inspection_results	Number Of Product Status
1	cosmetics	Pass	6
2	haircare	Pass	6
3	cosmetics	Fail	10
4	cosmetics	Pending	10
5	skincare	Pass	11
6	haircare	Fail	13
7	skincare	Fail	13
8	haircare	Pending	15
9	skincare	Pending	16

- **The skincare is the most Pending Product, the skincare and haircare are the most Fail Products and skincare is the most Pass Products**



# ANALYSIS QUESTIONS PHASE USING SQL



## Fifth Question: Products Across Transportation Modes

```
----- Number Of Products Across Transportation Modes -----
select Product_type, Transportation_modes,COUNT(*) as 'Number Of Products'
from supply_chain_data
group by Product_type,Transportation_modes
```

	Product_type	Transportation_modes	Number Of Products
1	cosmetics	Air	5
2	haircare	Air	8
3	skincare	Air	13
4	cosmetics	Rail	9
5	haircare	Rail	10
6	skincare	Rail	9
7	cosmetics	Road	6
8	haircare	Road	12
9	skincare	Road	11
10	cosmetics	Sea	6
11	haircare	Sea	4
12	skincare	Sea	7

- **Air Mode: the Skincare Product is the most transported.**
- **Rail Mode: the Haircare Product is the most transported.**
- **Road Mode: the Haircare Product is the most transported.**
- **Sea Mode: the Skincare Product is the most transported.**
- **The Cosmetics Product is more portable during the Rail.**
- **The Haircare Product is more Road during the Rail.**
- **The Skincare Product is more portable during the Air.**

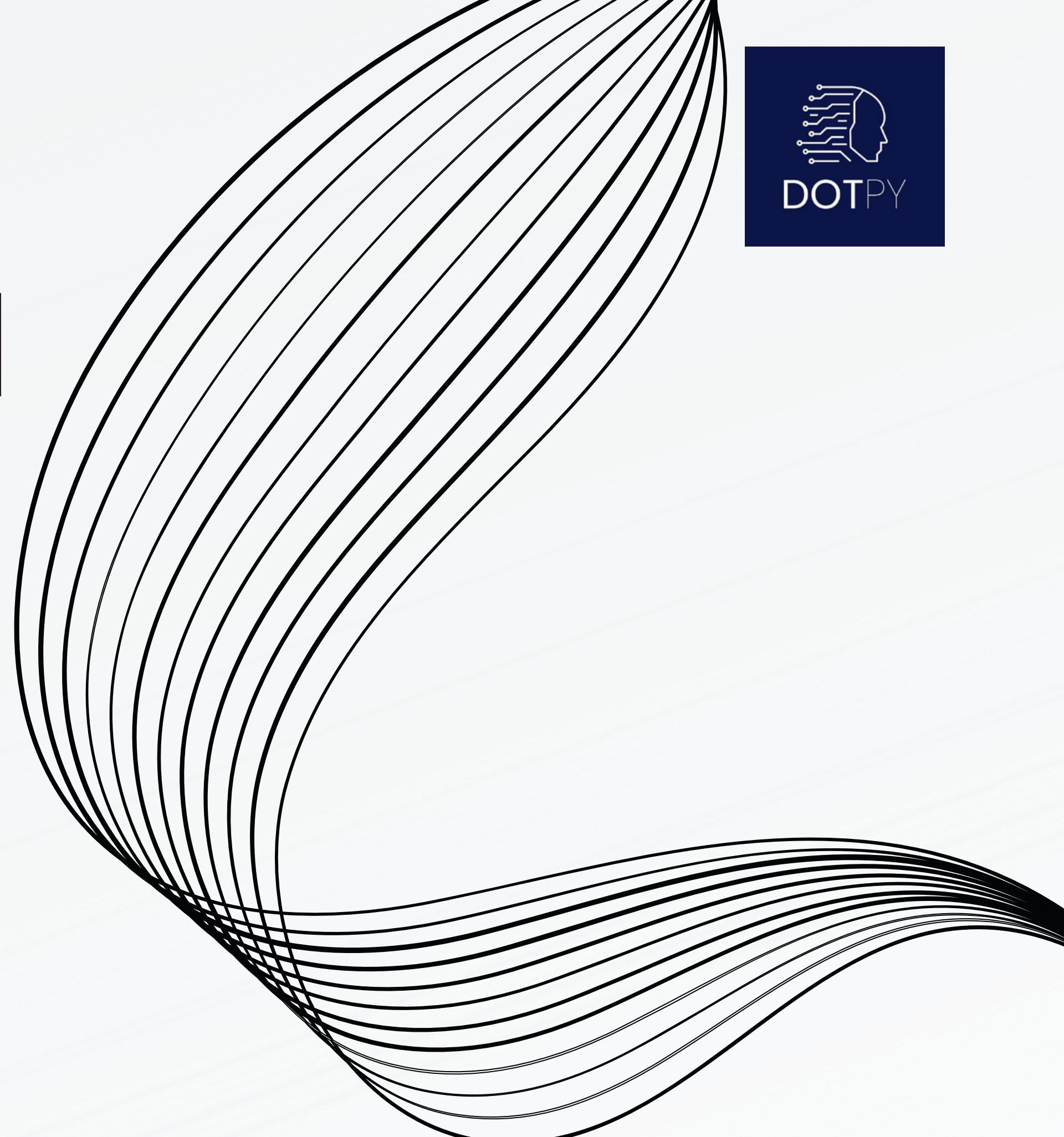


**REPORTING PHASE IS OVER  
AND THE NEXT STEP IS DASHBOARDS.**



# VISUALIZATION DASHBOARD

*The Visualization Dashboard phase involves creating interactive and informative displays to represent key insights from data. Charts, graphs, and tables are used to highlight trends, patterns, and comparisons. These visualizations support data-driven decision-making and make complex information easier to understand.*

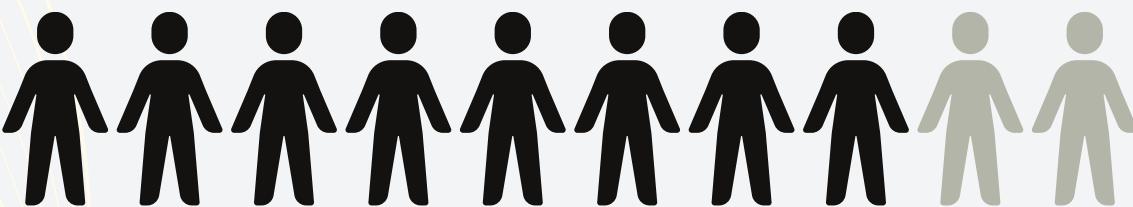


# TABLEAU



Tableau is a powerful data visualization tool used to create interactive and shareable dashboards. It enables users to analyze large datasets with ease, providing insights through various charts, graphs, and maps. Tableau's intuitive interface makes it accessible for both technical and non-technical users to visualize and understand data effectively.

80%





Dashboard 1  
Overview

## Dashboard 1 | OVERVIEW

D1

Most Customer Gender Sales			
Female	Male	Non-binary	Unknown
12,801	7,507	10,580	15,211

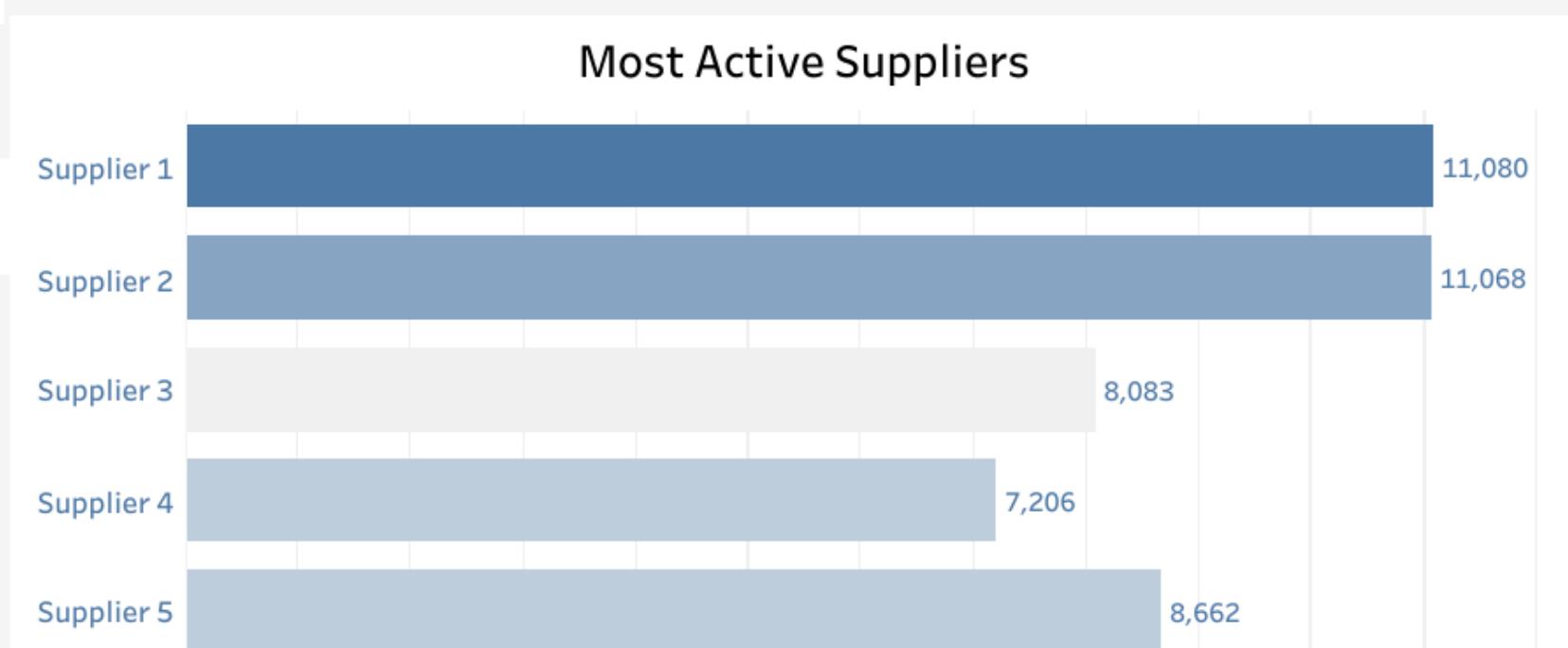


D2

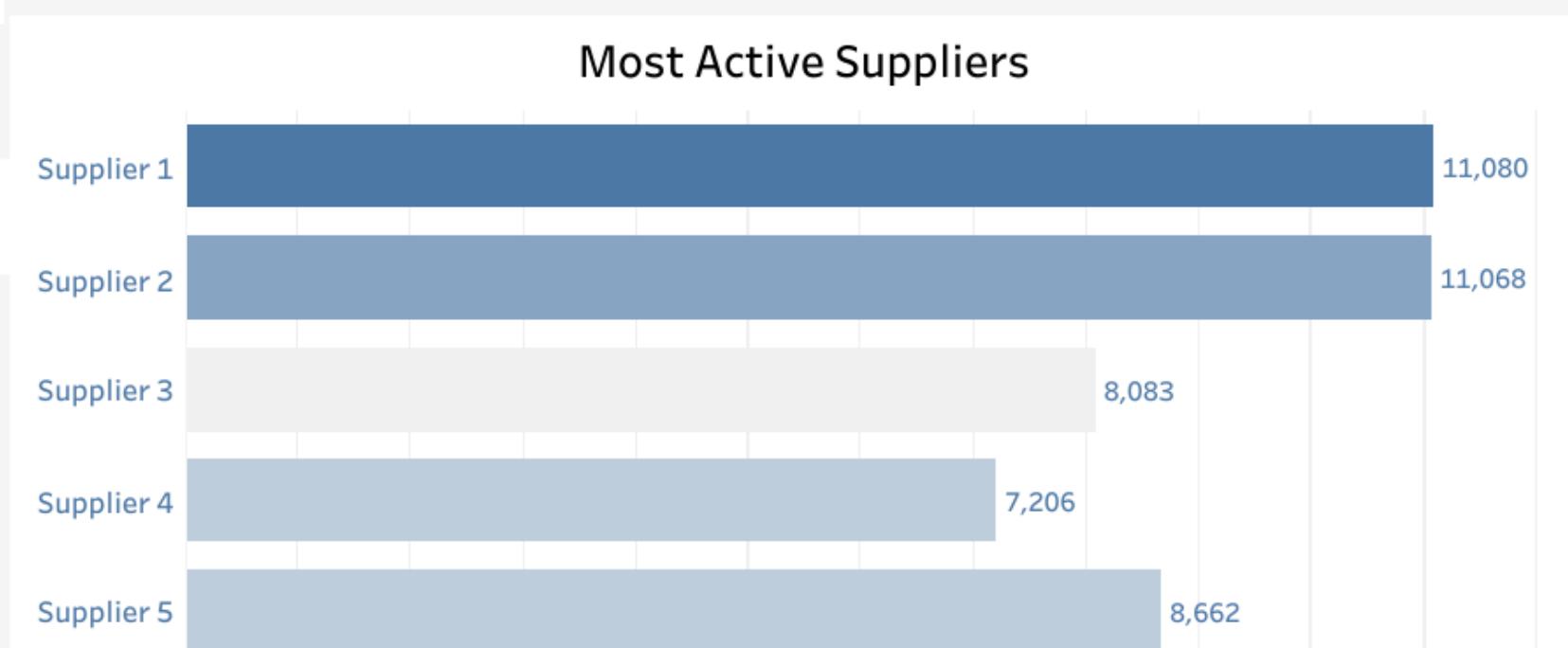
Profit Ratio By Location			
	cosmetics	haircare	skincare
Bangalore	0.1660	0.3630	0.2890
Chennai	0.2491	0.2183	0.1617
Delhi	0.1255	0.1299	0.1377
Kolkata	0.2749	0.1921	0.1409
Mumbai	0.2274	0.1973	0.1975



D3



D4



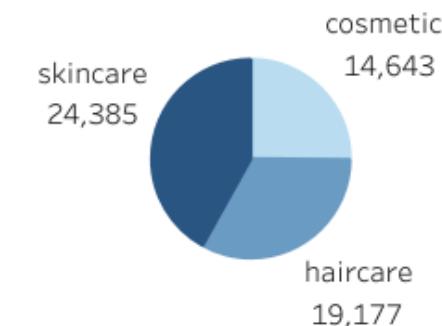
D5



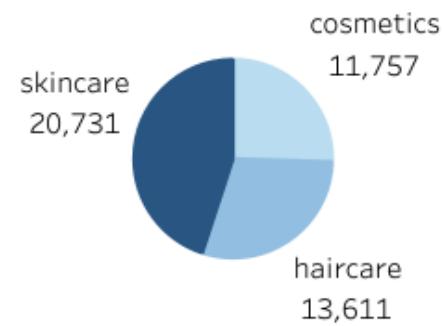


## DASHBOARD 2 | Product Type Performance

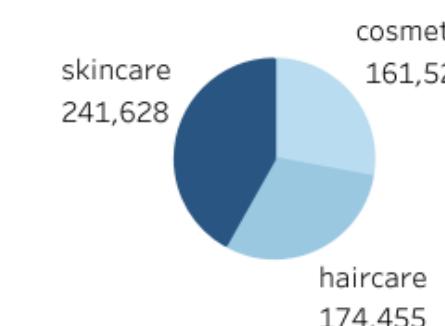
**Cost Breakdown by Product Type**



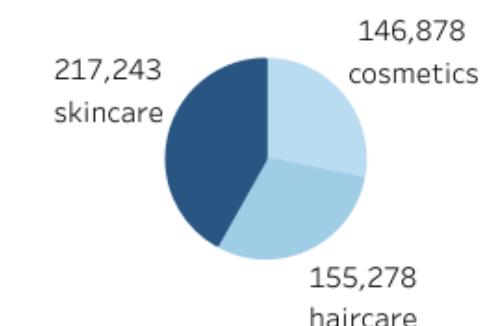
**No. Product Sold**



**Revenue of product**



**Profit by Product Type**



D1



D2



D3



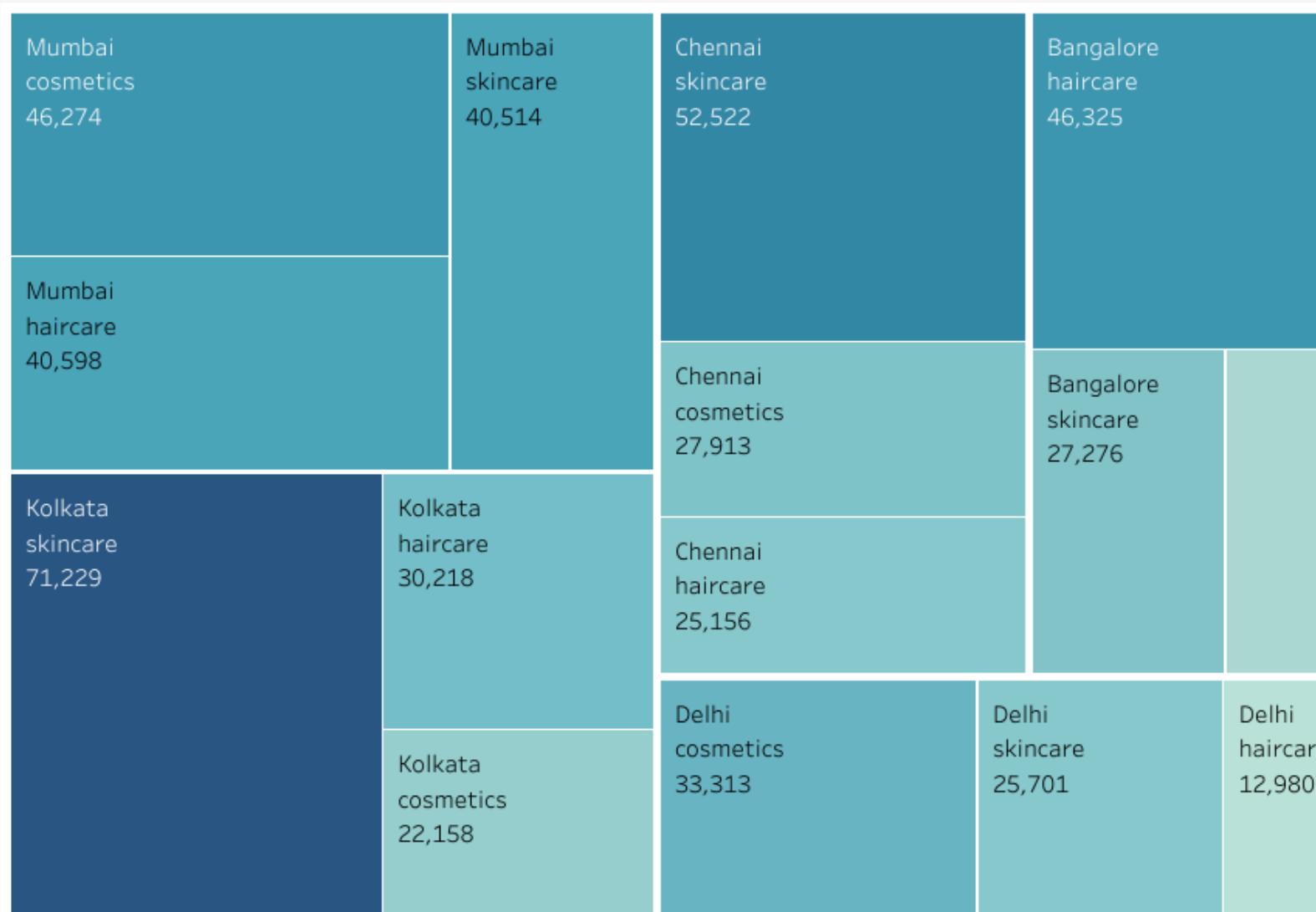
D4



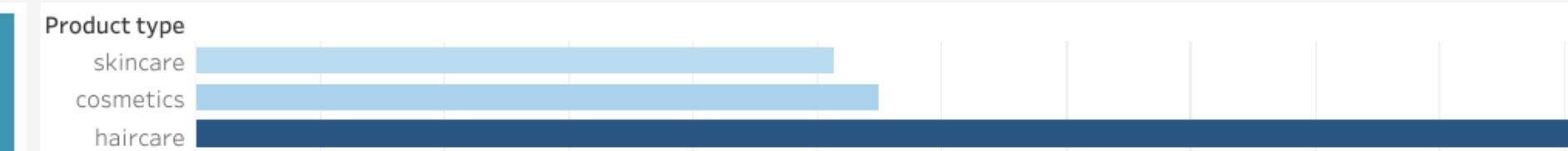
D5



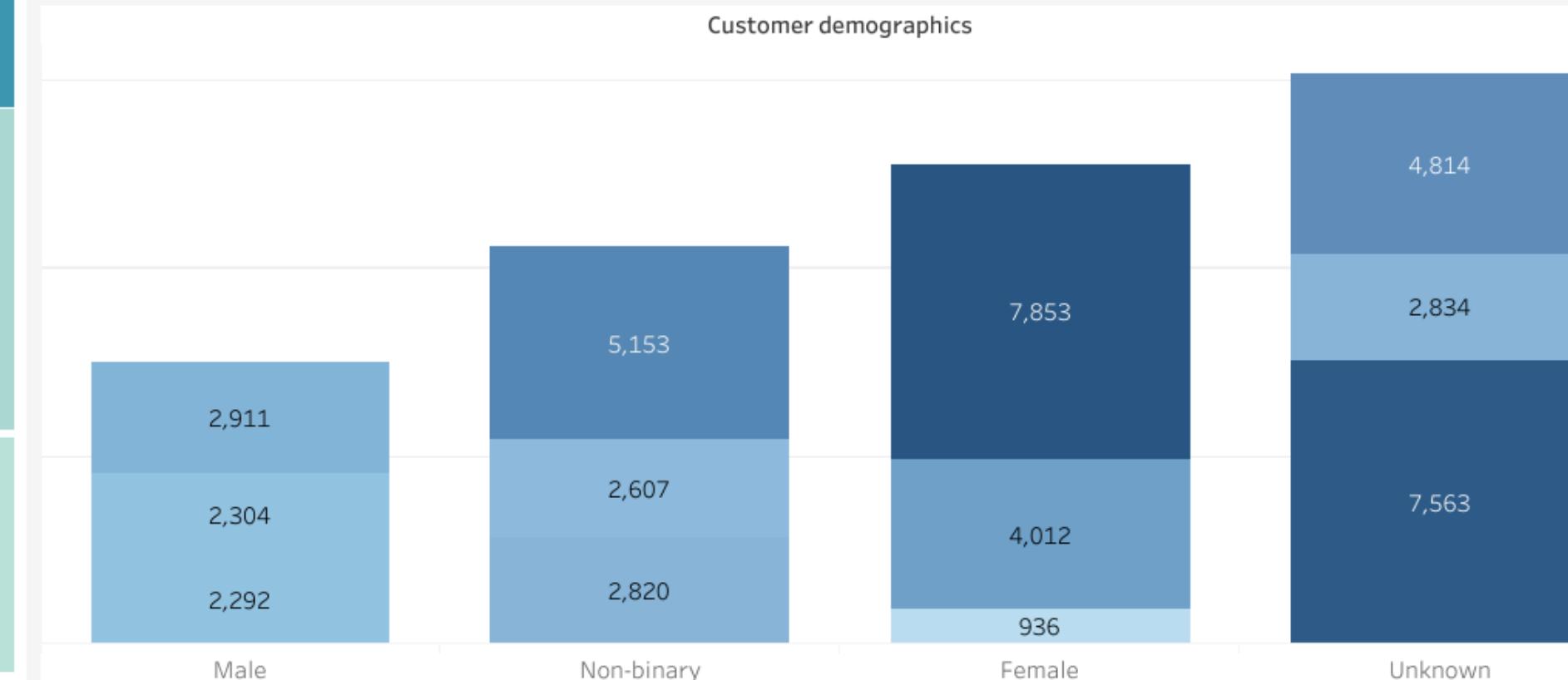
**Profit by Location**



**Stock Efficiency**



**No. Product Sold Vs Customer demographics**

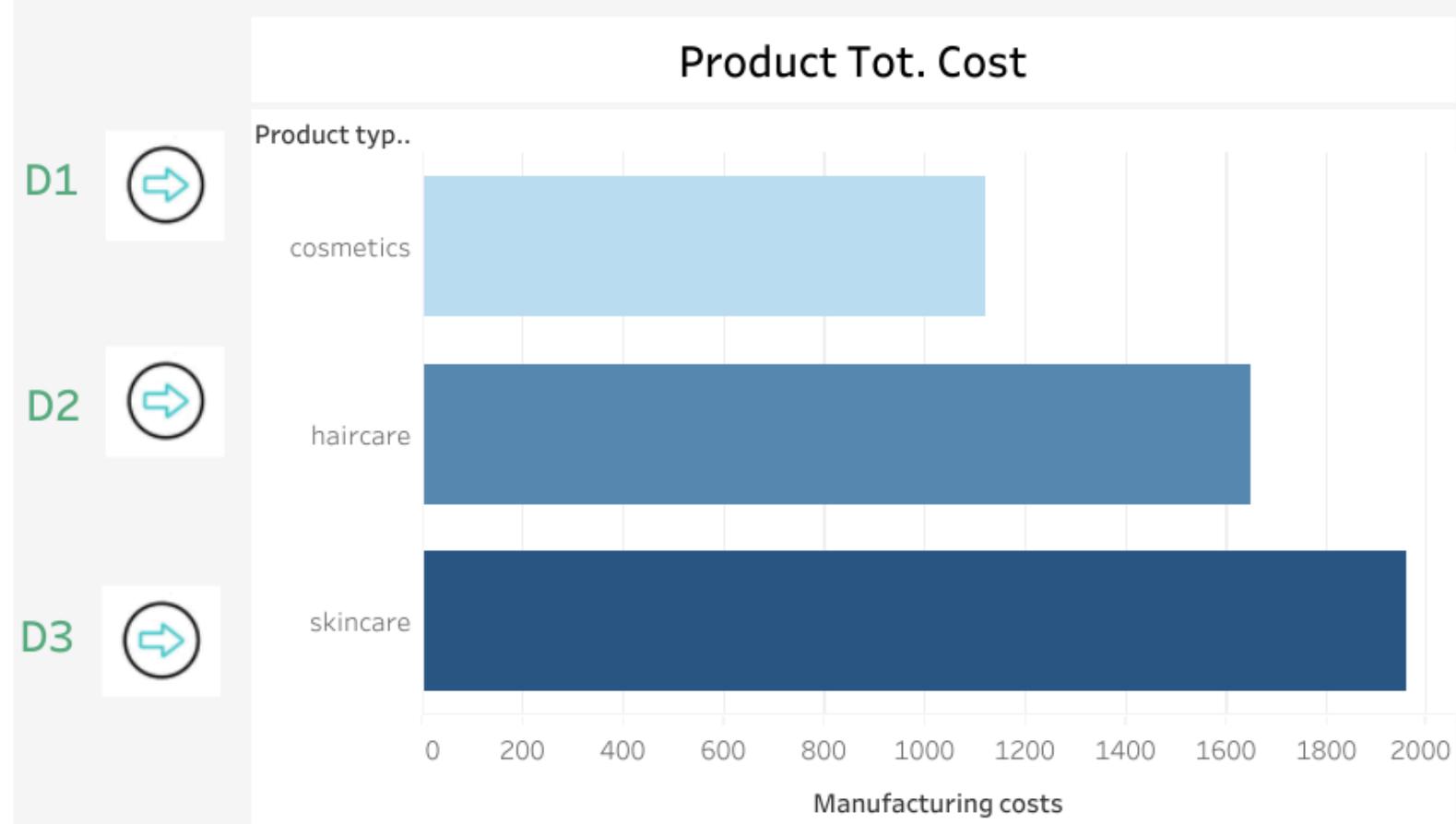




المركز العربي للبحوث العلمية  
Arab Center for Research and Training

## DASHBOARD 3 | PRODUCTS

Product Tot. Cost



Tot. Products Sold



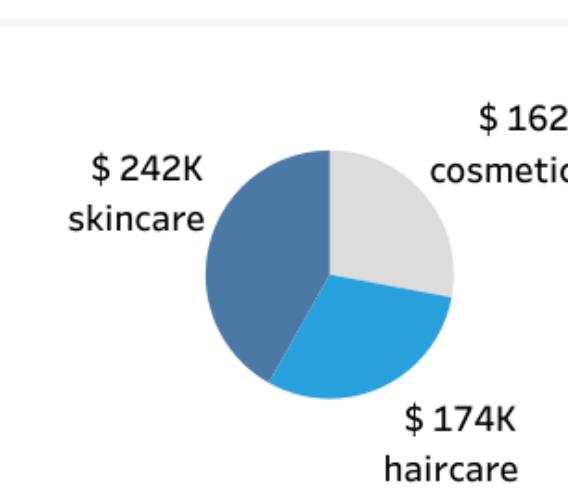
AVG Defect Rate

2.28%

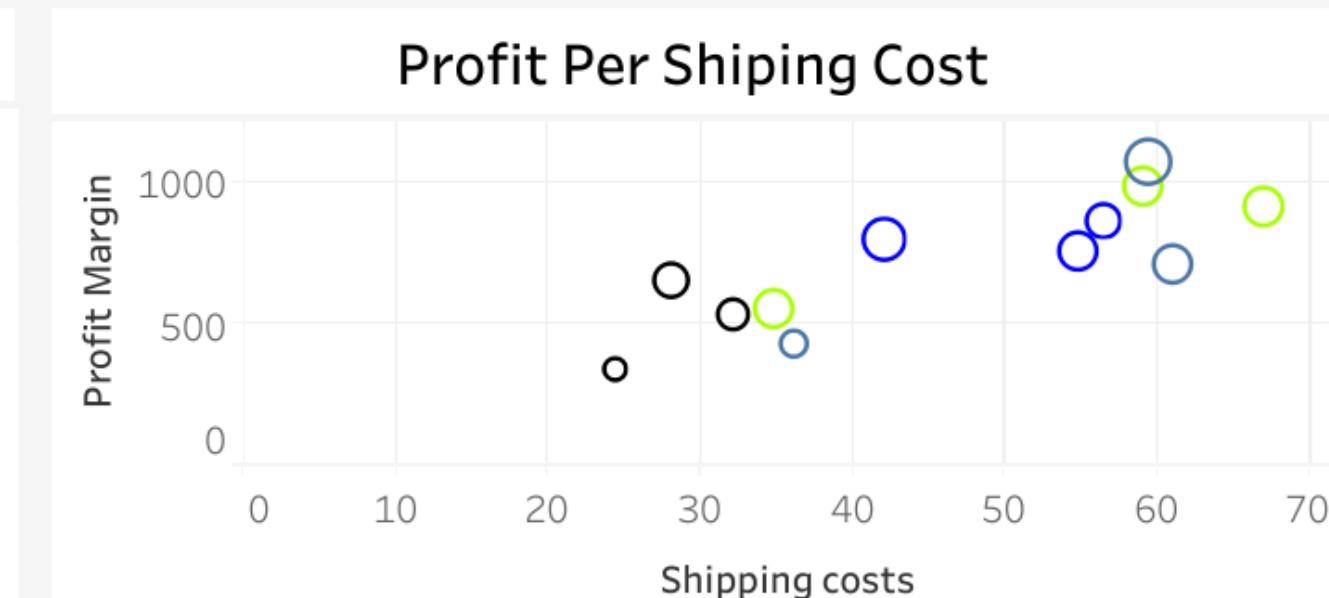
Stock Level

4,777

Revenue Per Product



Profit Per Shipping Cost



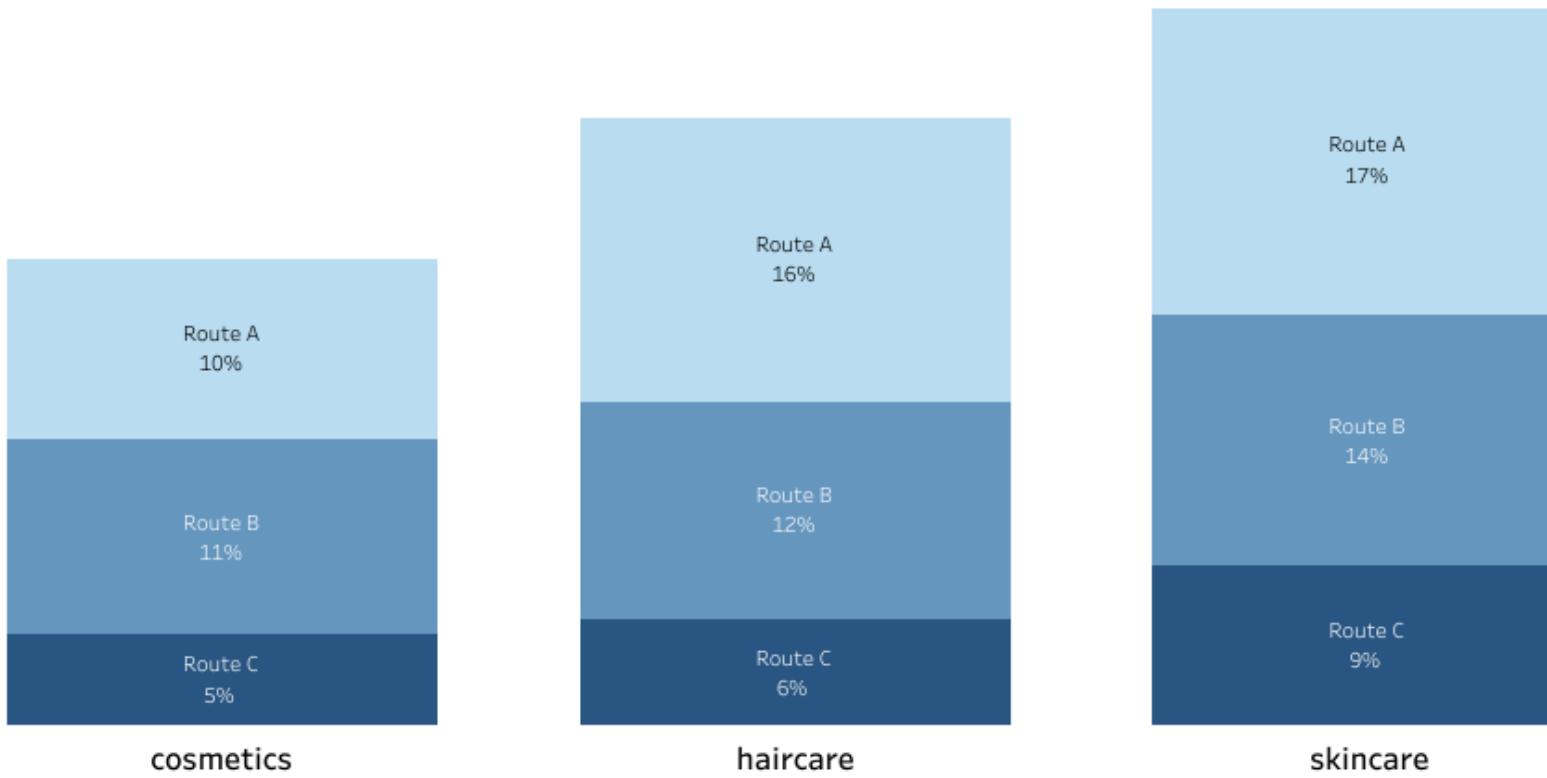


## DASHBOARD 4 | Insights Of Products

D1



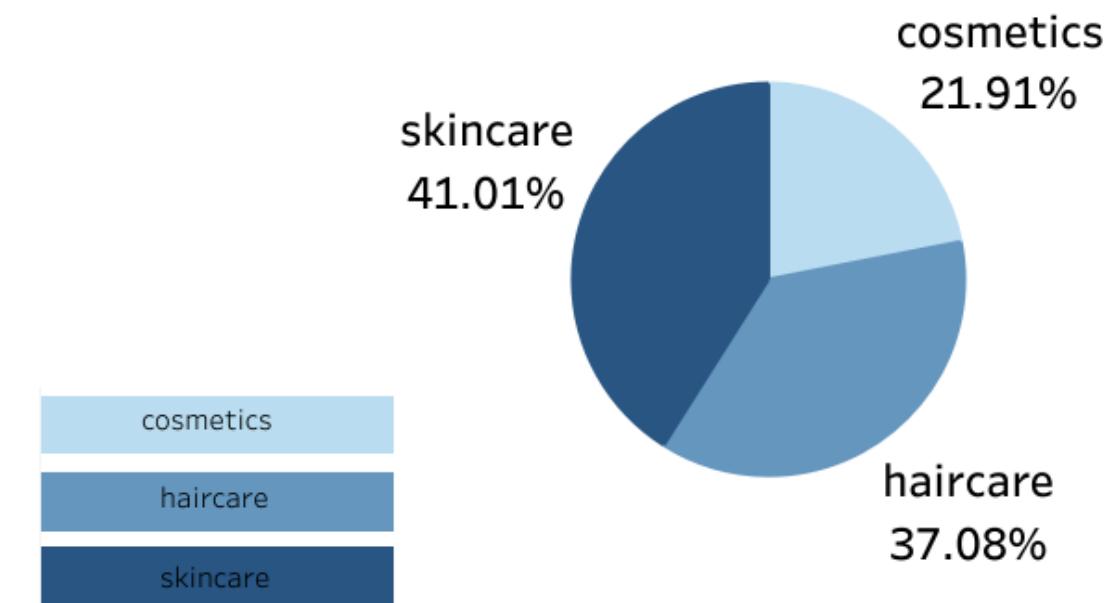
Percentage Of Routes Across Products



D2



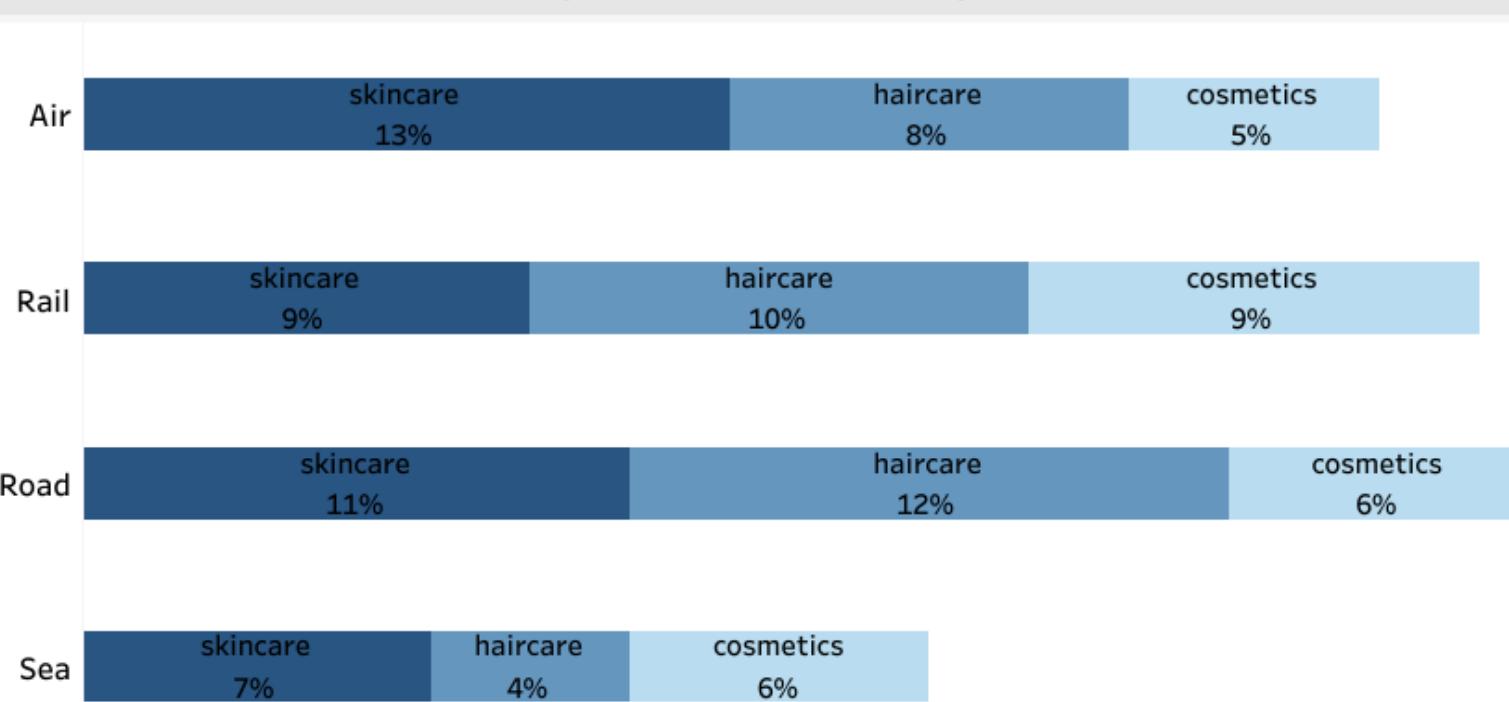
Percentage Of Defect Rates



D3



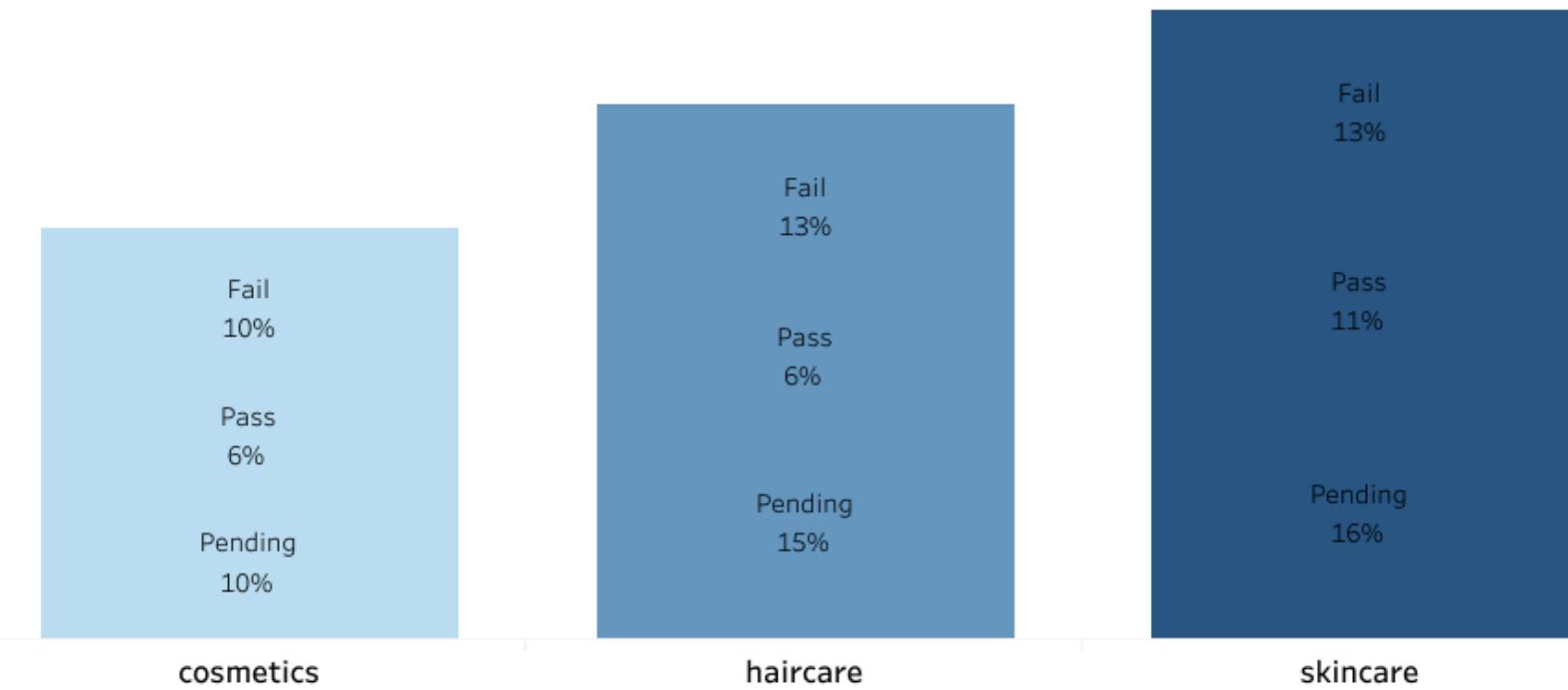
Percentage Of Products Across Transportation



D4



Percentage Of Inspections Across Products



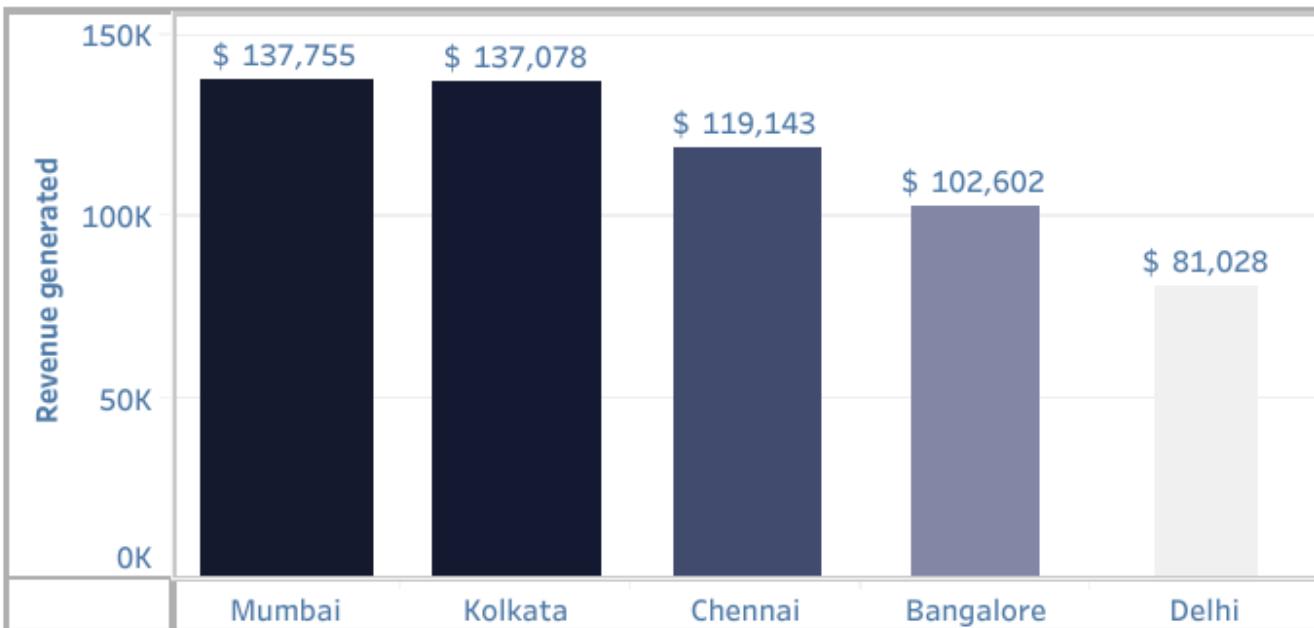


## DASHBOARD 5 | SALES

Total Profit

8,607

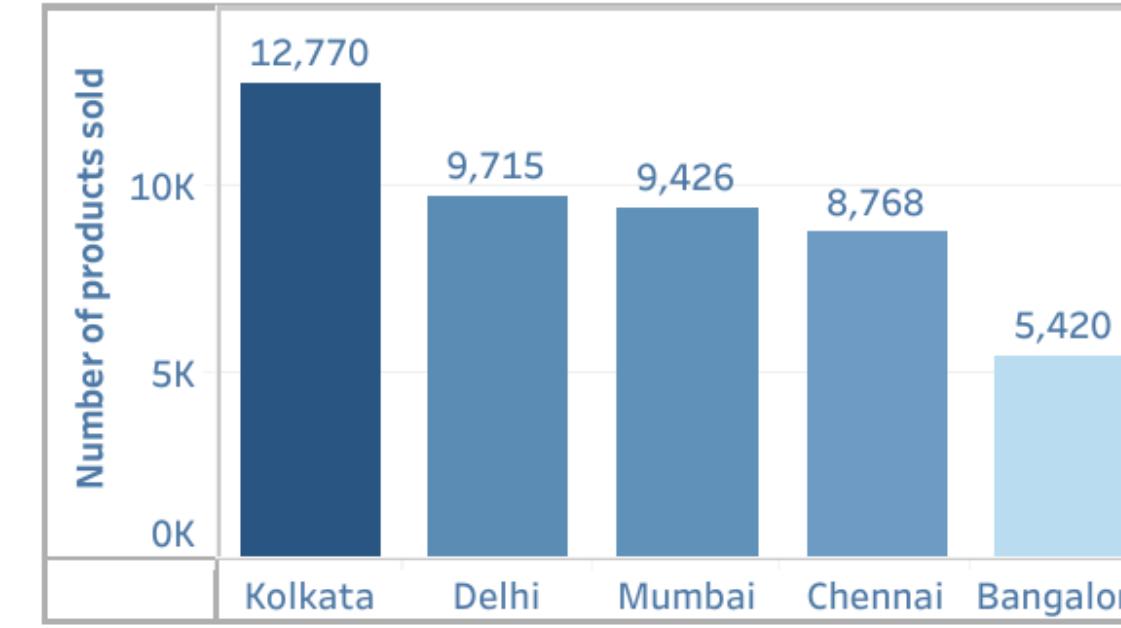
revenue generated by city



Total Sales

46,099

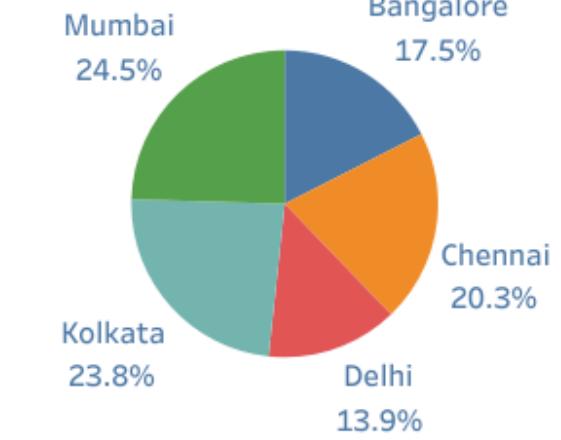
Sales By City



Total Costs

58,206

operational profit



transportation modes by city

Transportation modes1	Bangalore	Chennai	Delhi	Kolkata	Mumbai
Air	3	6	4	6	7
Rail	5	5	8	6	4
Road	8	5	1	10	5
Sea	2	4	2	3	6

defect rates in products by city

skincare Kolkata 12.261%	skincare Mumbai 7.608%	skincare Chennai 10.184%	haircare Kolkata 9.507%	haircare Mumbai 6.834%	cosmetics Mumbai 6.063%	cosmetics Delhi 5.785%
			haircare Chennai 8.747%			
			skincare Delhi 4.502%		cosmetics Chennai 4.241%	
			haircare Bangalore 7.595%		cosmetics Kolkata	





رواد مصر الرقمية  
هيئة الاتصالات  
والتكنولوجيا المعلومات



# THANK YOU