



DATA ANALYSIS PROJECT

CONTENT

- 01** ABOUT ME
- 02** ABOUT MY PROJECT
- 03** DATA CLEANING AND UNIVARIANT ANALYSIS
- 04** FEATURES ENGINEERING
- 05** THE QUESTIONS I WANT ANSWERED
- 06** PROJECT ANALYSIS
- 07** DATA ANALYSIS SUMMARY

ABOUT ME



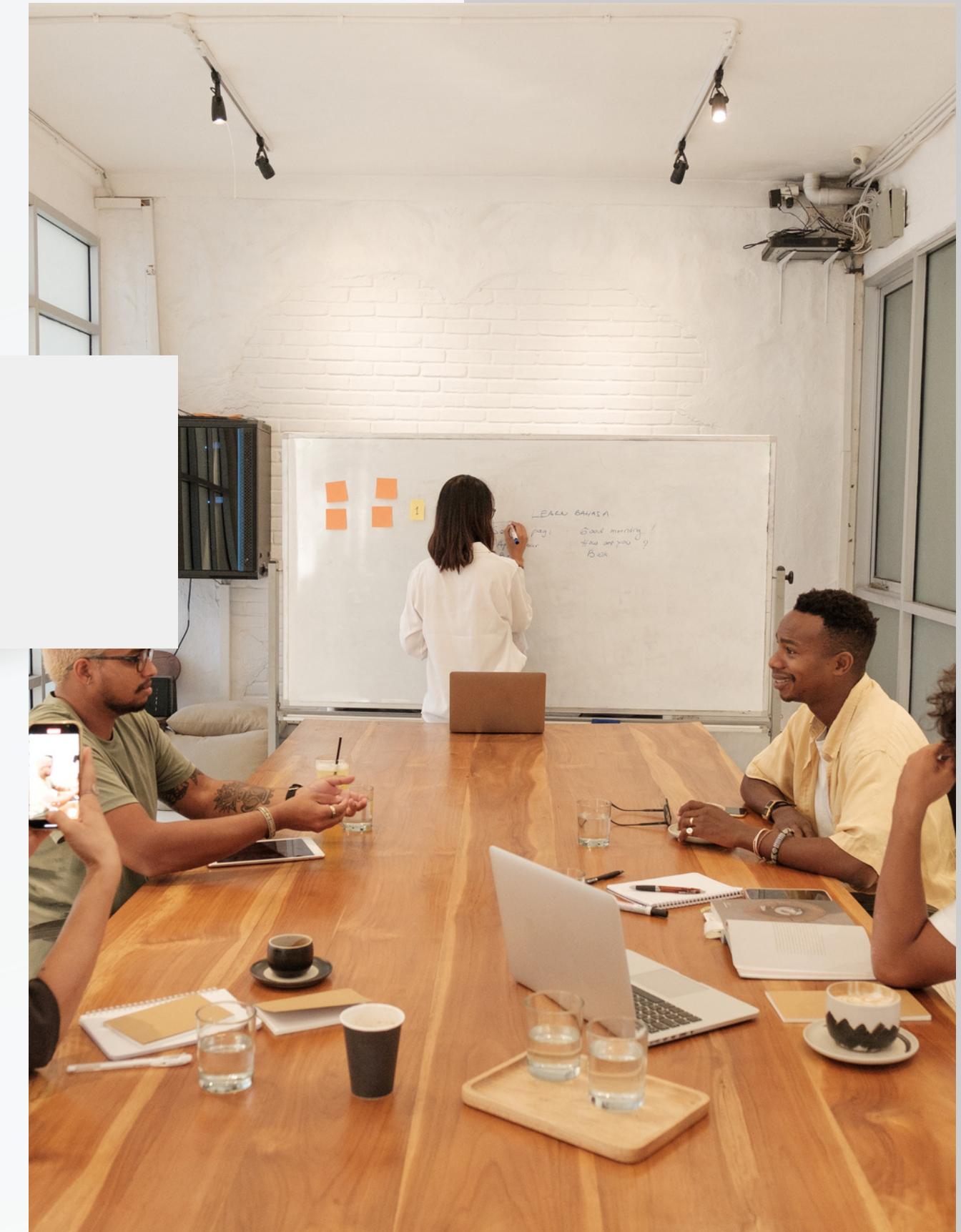
I am Islam Omar, an accountant at Ceramica Cleopatra Group. I have a passion for data analysis, especially in data related to sales. I have experience in Excel, Power BI, Tableau, Google Studio, SQL and Python. A graduate of the Faculty of Commerce, Ain Shams University, holding a data analysis diploma and a data science diploma.



ABOUT MY PROJECT



The project is an analysis of four years' sales data from 2019 to 2022 in more than one continent and country

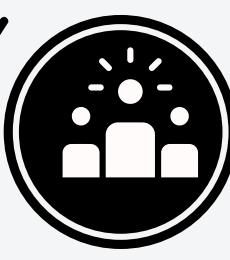
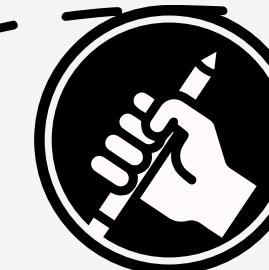


DATA CLEANING AND UNIVARIANT ANALYSIS

Data cleaning

I cleaned the data from the missing and incorrect cells, deleted the unimportant columns, modified the date and the data type

0	Date	15288 non-null	datetime64[ns]
1	Invoice Type	15288 non-null	object
2	Customer ID	15288 non-null	int64
3	Item Code	15288 non-null	int64
4	Qty	15288 non-null	int64
5	Discount	15288 non-null	float64
6	Item Category	15288 non-null	object
7	Item Type	15288 non-null	object
8	Item Name	15288 non-null	object
9	Unit	15288 non-null	object
10	Price	15288 non-null	int64
11	Cost	15288 non-null	int64
12	Customer Name	15288 non-null	object
13	Continent Name	15288 non-null	object
14	Country	15288 non-null	object
15	Region	15288 non-null	object
16	Sales Rep.	15288 non-null	object
17	Channel	15288 non-null	object
dtypes: datetime64[ns](1), float64(1), int64(5), object(11)			
memory usage: 2.2+ MB			

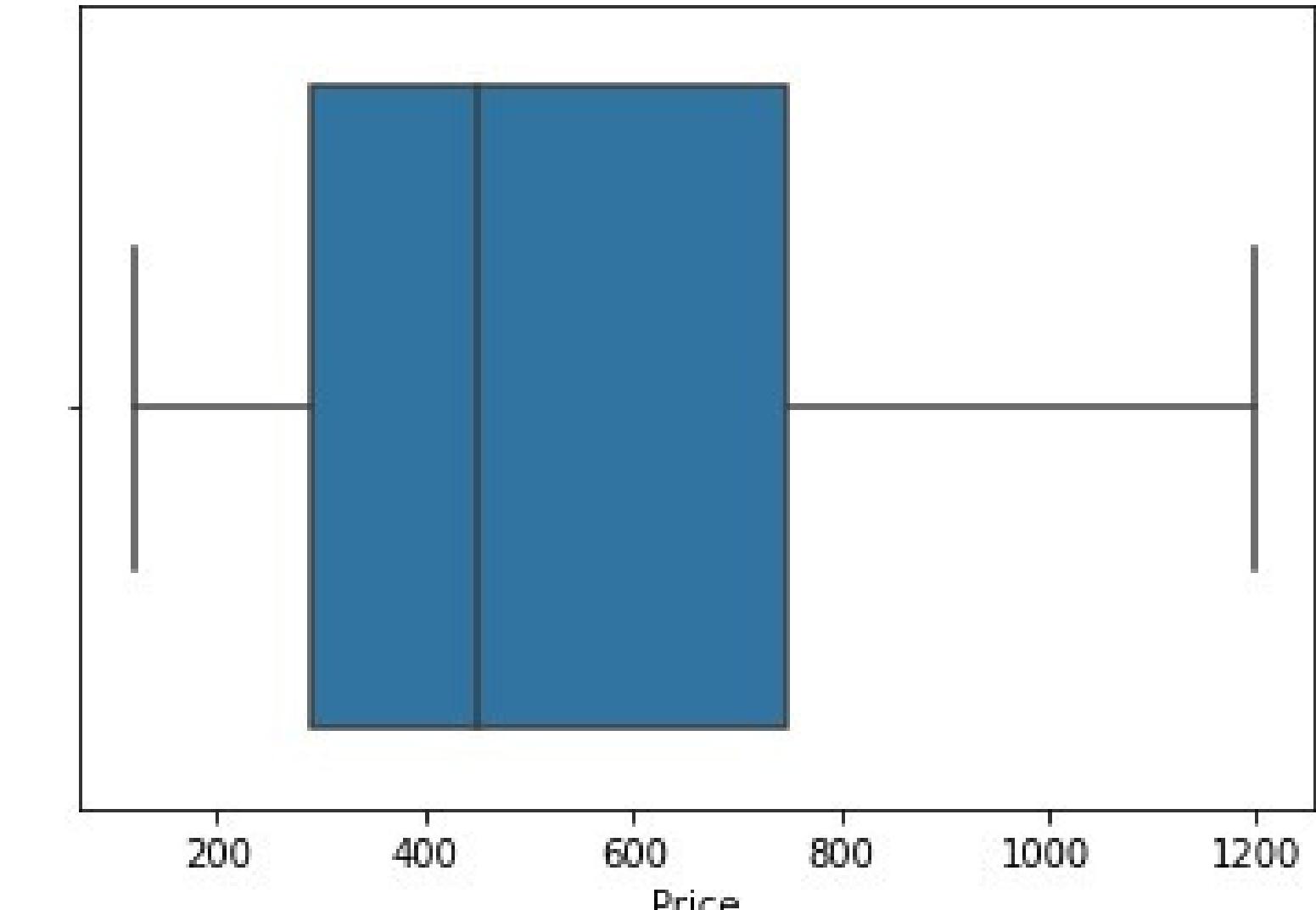


check outliers

I confirmed the outlier values, but I discovered that the values are normal

```
2]: sns.boxplot(data= allrevn, x= 'Price')
```

```
2]: <AxesSubplot:xlabel='Price'>
```



FEATURES ENGINEERING



Using the price, cost, and quantity column, you extracted the selling price, profit, total costs, growth rate, and profit rate

N°1



I extracted the name of the month, its number, the name of the day and the year

N°2



I extracted the four seasons using apply

N°3

STRATEGIES



```
allrevn["Year"] = allrevn['Date'].dt.year  
allrevn["Month"] = allrevn['Date'].dt.month  
allrevn["Month Name"] = allrevn['Date'].dt.month_name()  
allrevn['Day'] = allrevn['Date'].dt.day  
  
def map_months(x):  
    if x in [12, 1, 2]:  
        return 'Winter'  
    elif x in [3, 4, 5]:  
        return 'Spring'  
    elif x in [6, 7, 8]:  
        return 'Summer'  
    elif x in [9, 10, 11]:  
        return 'Autumn'  
  
allrevn['Season'] = allrevn['Month'].apply(map_months)  
  
allrevn["total_cost"] = allrevn["Qty"] * allrevn["Cost"]  
allrevn["total_price"] = allrevn["Qty"] * allrevn["Price"]  
allrevn["total_discount"] = allrevn["total_price"] * allrevn["Discount"]  
allrevn["net_price"] = allrevn["total_price"] - allrevn["total_discount"]  
allrevn["profit"] = allrevn["net_price"] - allrevn["total_cost"]
```

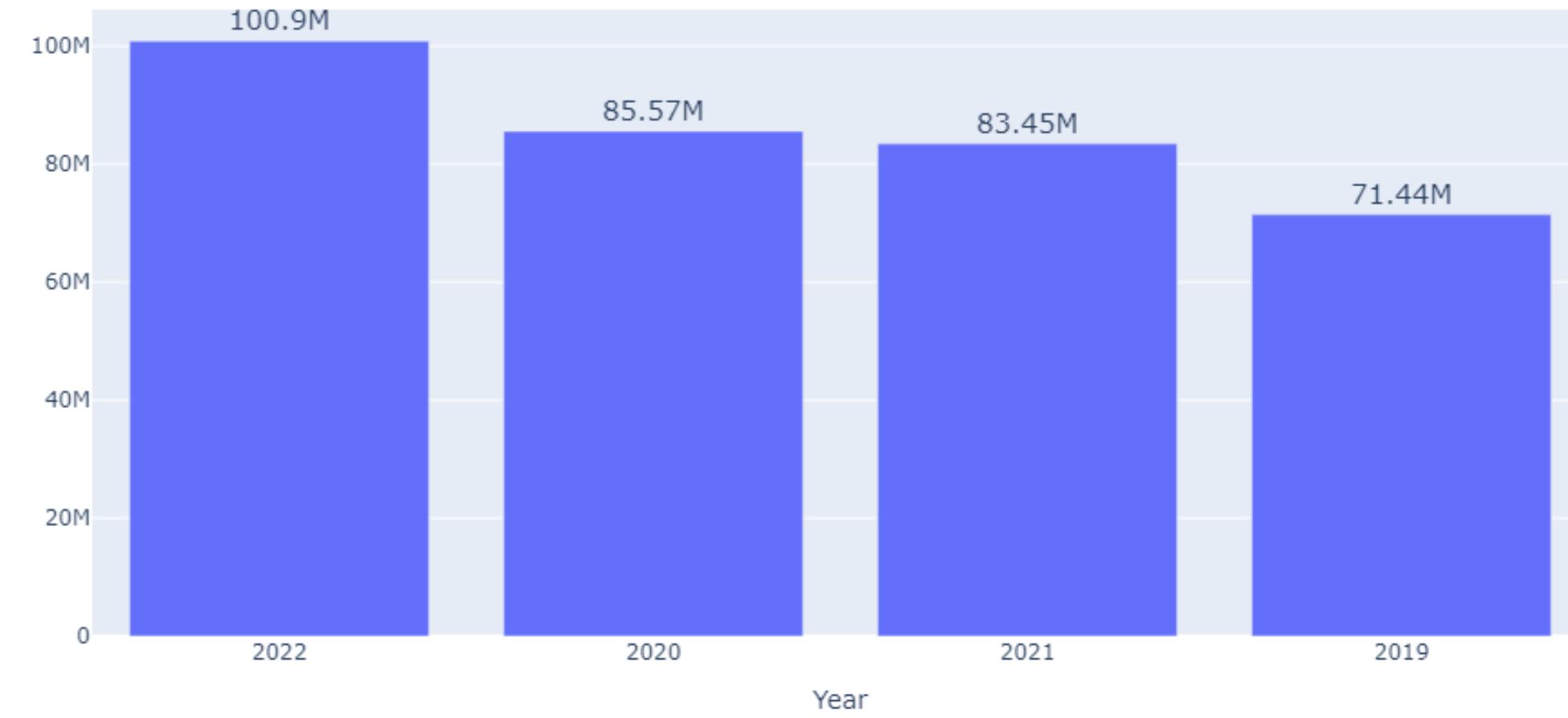
THE QUESTIONS I WANT ANSWERED

- What is the start and end date of the data
- What is the total sales?
- What is the total cost of the discounts?
- What are the total costs
- What are the total profits?
- What are the total quantities?
- Indicate the best years and indicate the rate of growth
- Explain the best selling months
- Describe the best selling seasons of the year
- Describe your best selling men
- Describe your best customers
- Explain the best category in terms of sales
- Explain the percentage of profit in each category
- Indicate the continents and countries in which you are located

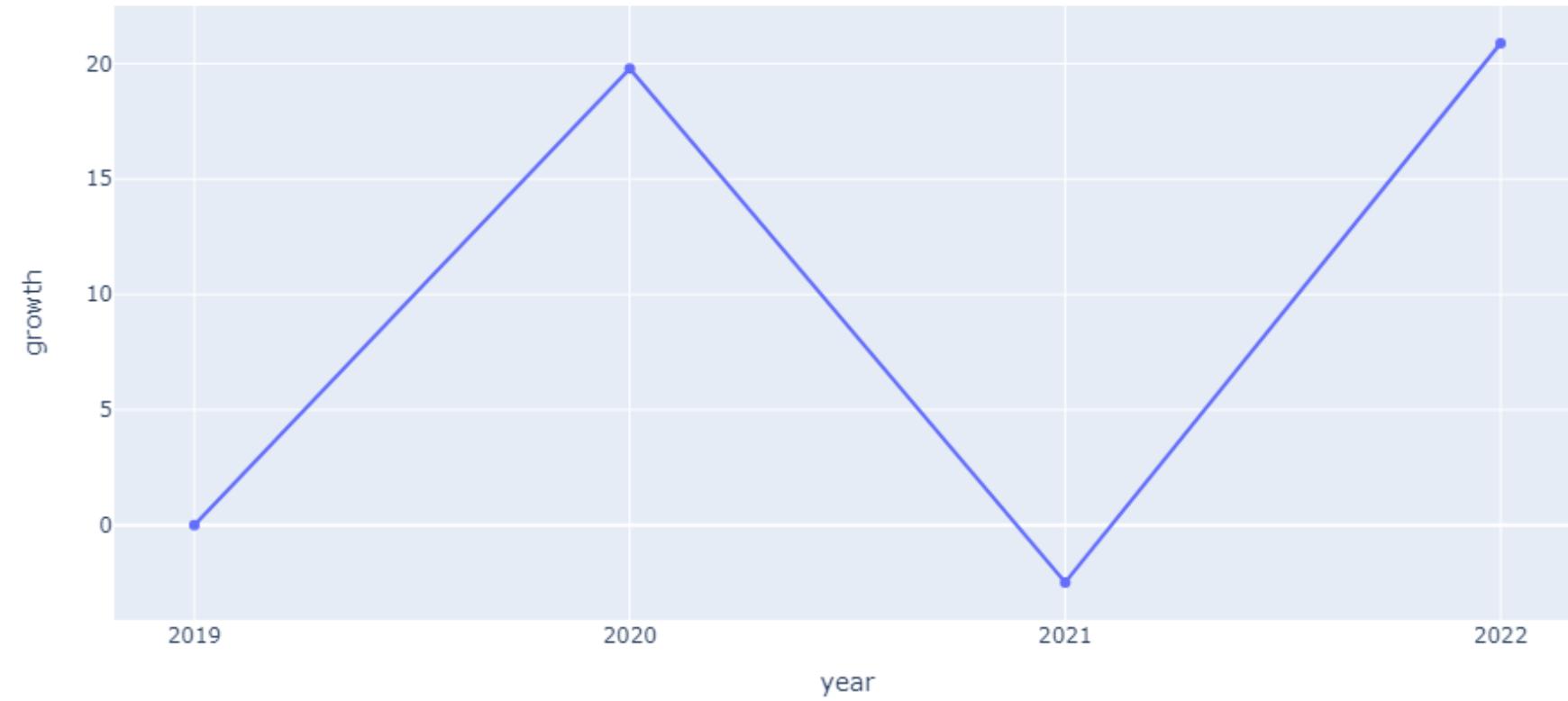


INDICATE THE BEST YEARS AND INDICATE THE RATE OF GROWTH

Total revenue by year

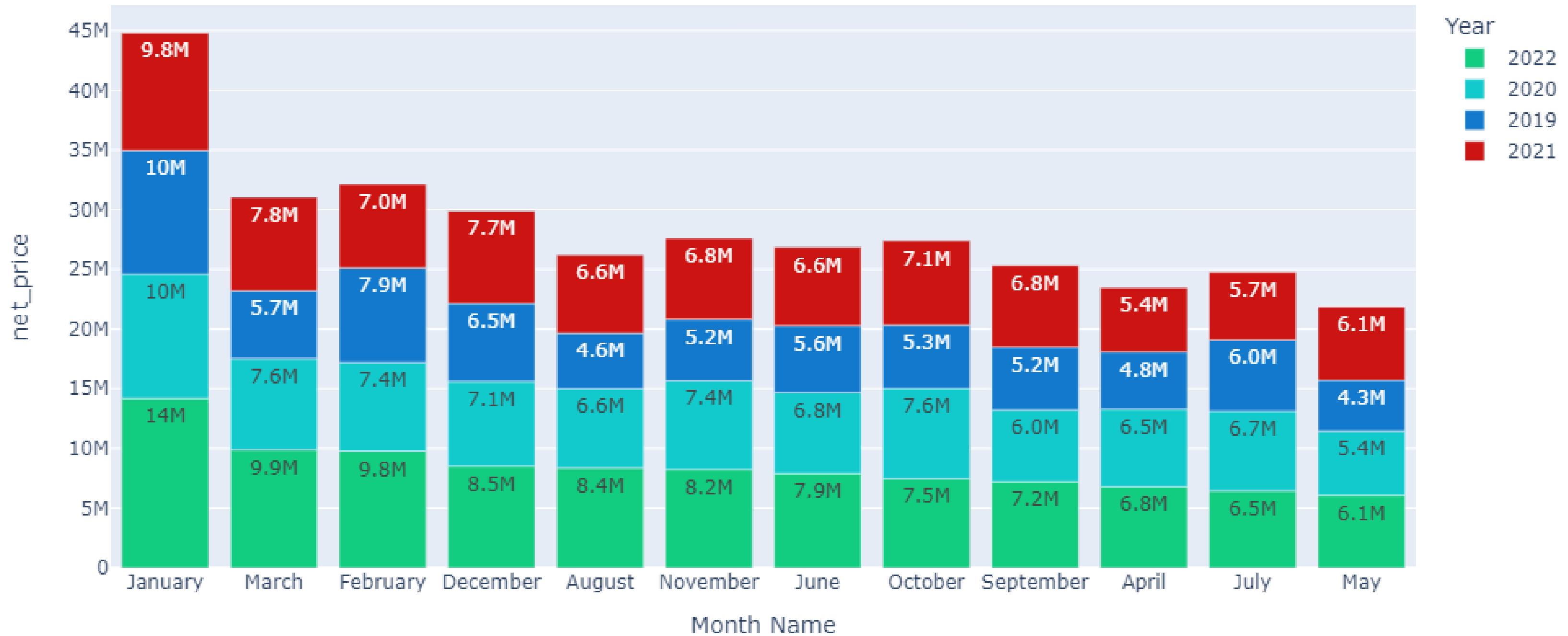


Growth Rate

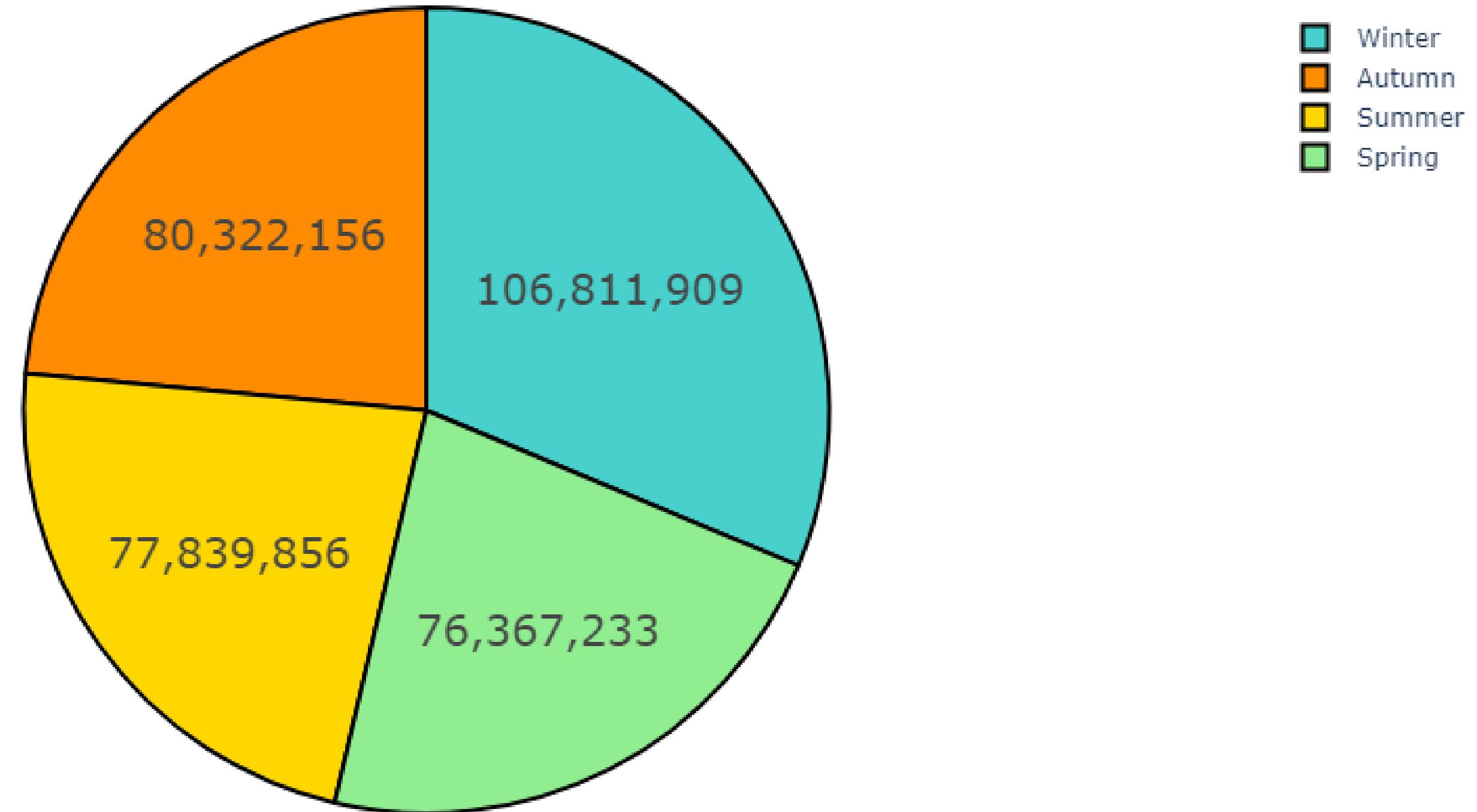


EXPLAIN THE BEST SELLING MONTHS

revneue by month each year

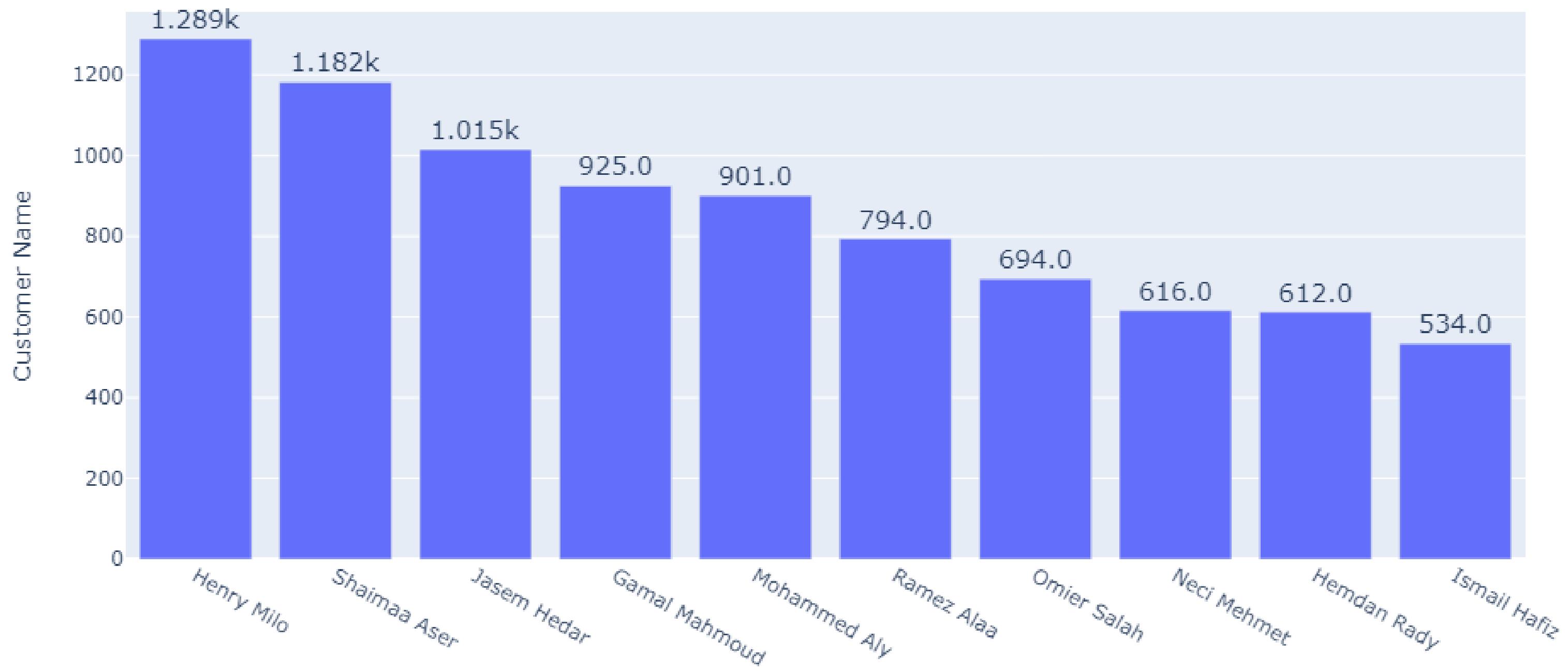


DESCRIBE THE BEST SELLING SEASONS OF THE YEAR

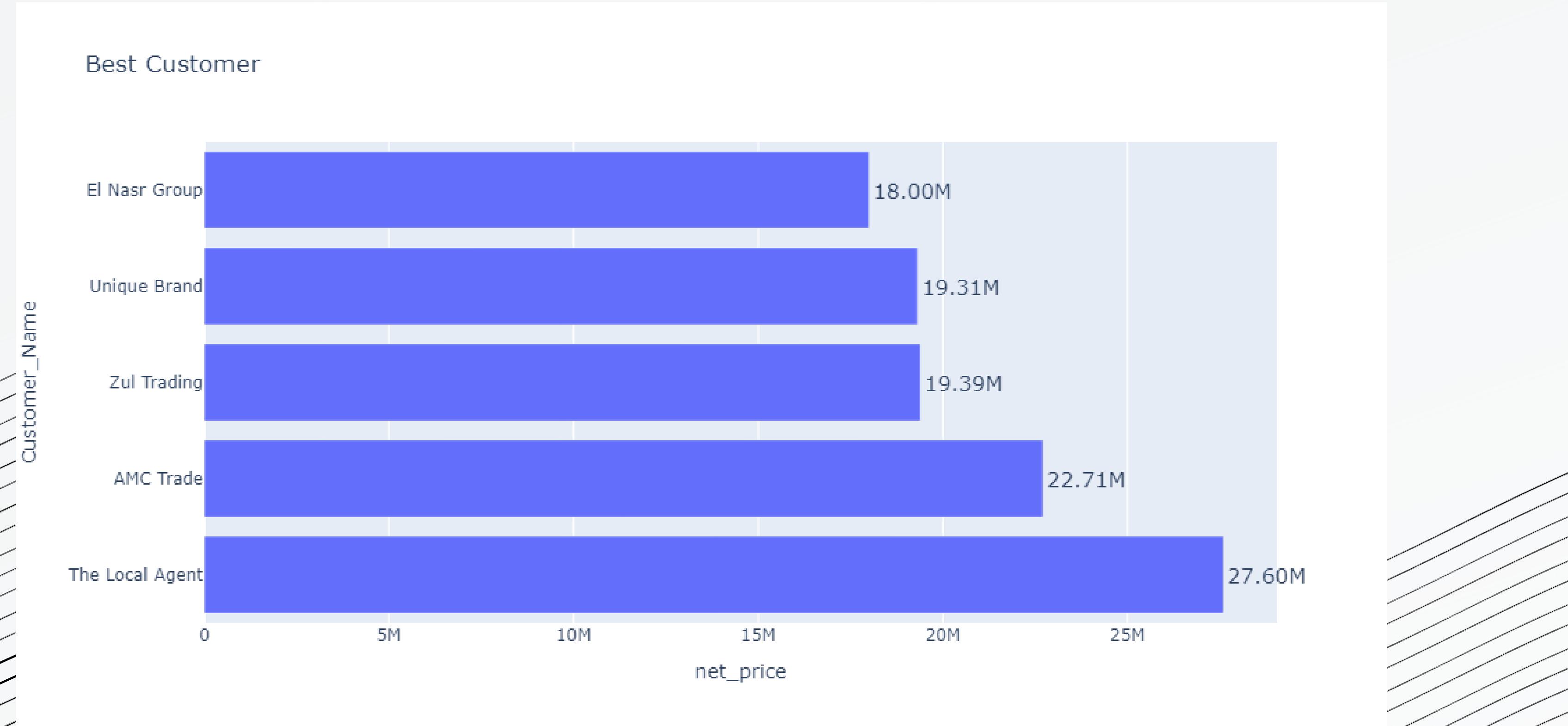


BEST SALES

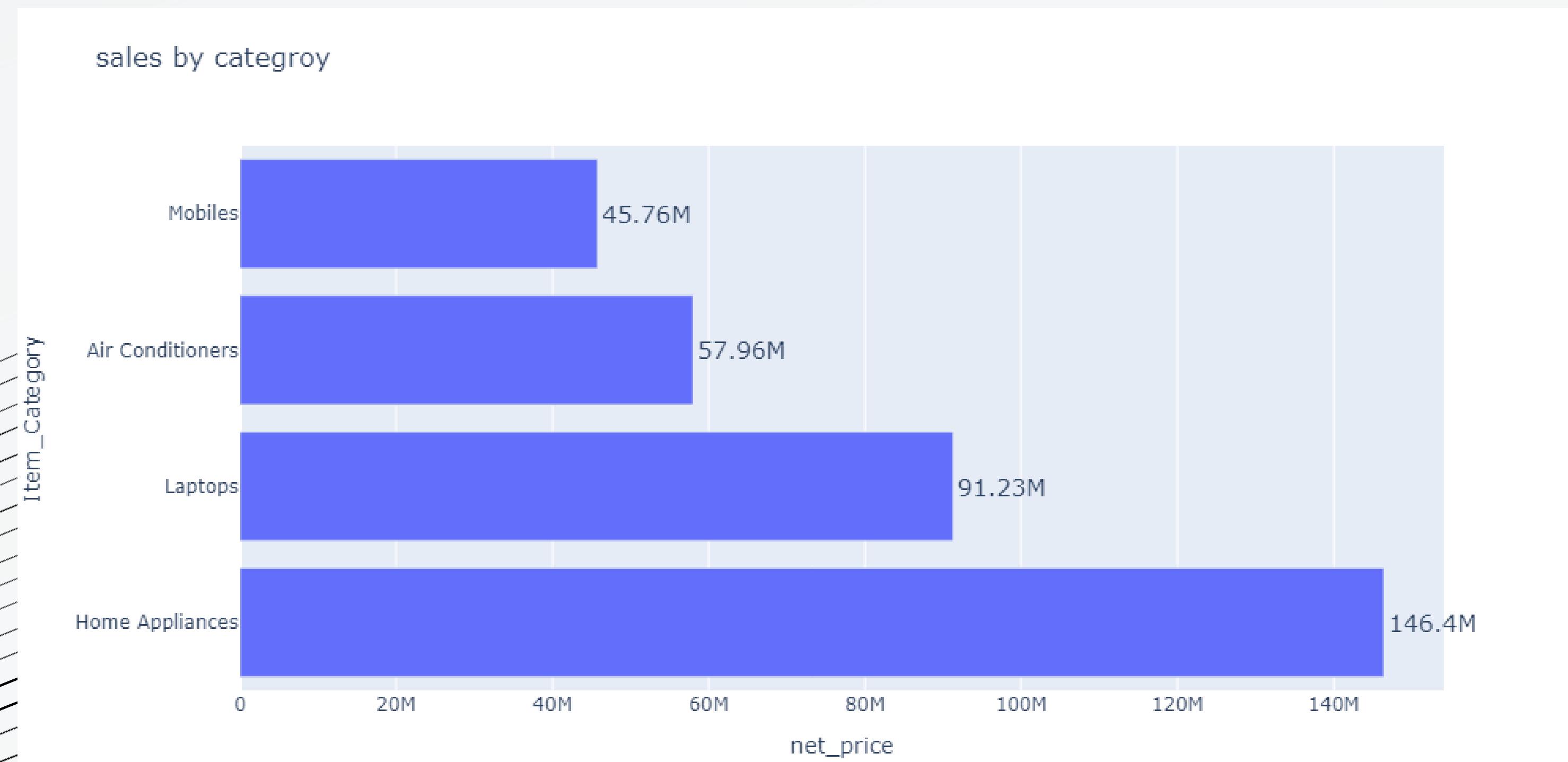
Best sales



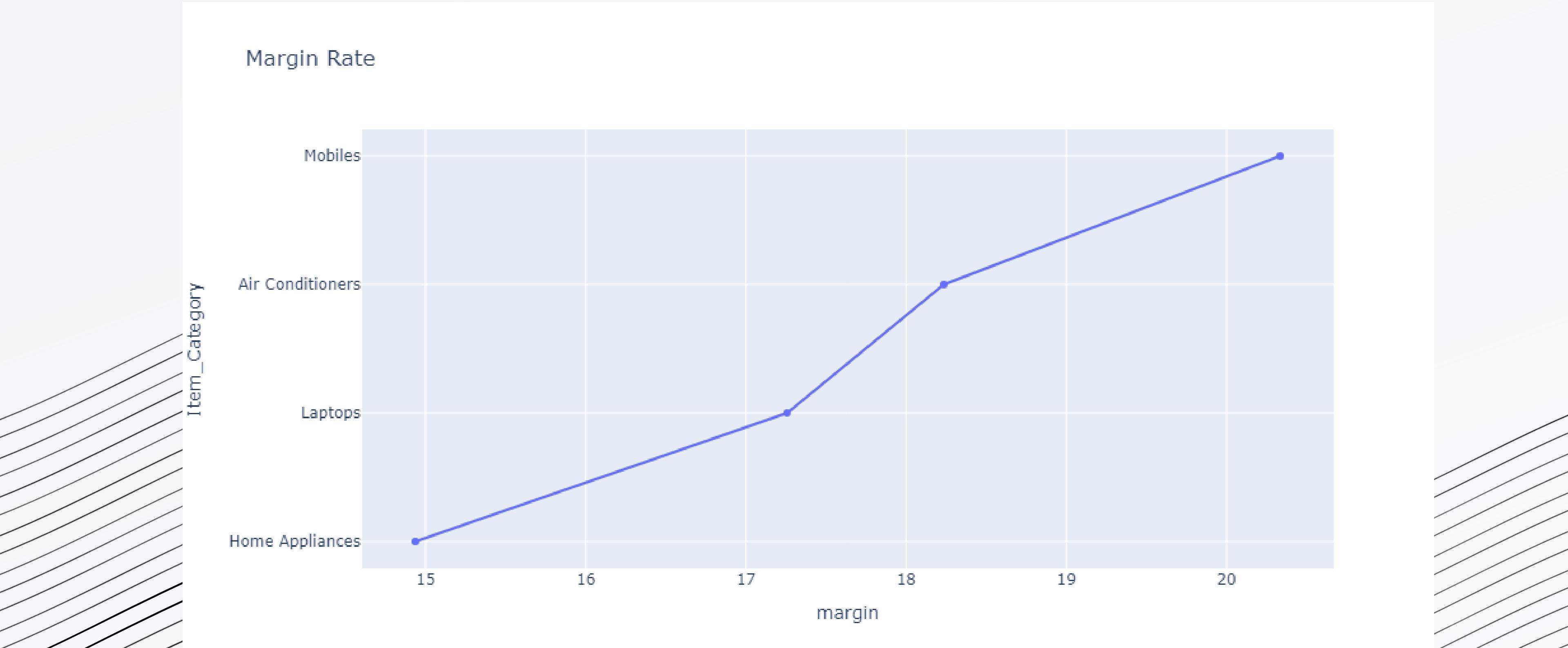
BEST CUSTOMER



SALES BY CATE GROY



MARGIN RATE



INDICATE THE CONTINENTS AND COUNTRIES IN WHICH YOU ARE LOCATED



Continent Name

- Africa
- Asia
- Europe



Year=2022

2022

2020

2021

2019

COST BY CATE GROY

