

## TASK 1

# 1.DATA ANALYTICS ON POC SALES DATA

This is the complete data after joining all the tables.

	id	customer_id	date_id	product_sku	units_sold	category	group	product_name	price	date	year	month	country_id	country_name
0	1	815	12	P018	4	Snow	Apparel	M'S Loft Parka	490	2022-01-12	2022	1	5	Sweden
1	5404	815	481	P005	3	Cycling	Apparel	Essential Road Sock	20	2023-04-26	2023	4	5	Sweden
2	9116	815	651	P015	1	Snow	Goggles	Fovea	200	2023-10-13	2023	10	5	Sweden
3	236	676	12	P011	1	Snow	Helmets	Obex Bc Mips	270	2022-01-12	2022	1	5	Sweden
4	1338	676	124	P007	2	Cycling	Armor	Vpd System Torso	260	2022-05-04	2022	5	5	Sweden
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9995	3195	347	259	P014	2	Snow	Goggles	Nexal	230	2022-09-16	2022	9	3	Austria
9996	5465	347	467	P009	2	Cycling	Armor	Joint Vpd 2.0 Elbow	100	2023-04-12	2023	4	3	Austria
9997	987	347	99	P004	1	Cycling	Apparel	M'S Essential Road	120	2022-04-09	2022	4	3	Austria
9998	2331	347	189	P006	3	Cycling	Apparel	M'S Reform Enduro Jersey	90	2022-07-08	2022	7	3	Austria
9999	2740	347	226	P008	4	Cycling	Armor	Joint Vpd 2.0 Knee	120	2022-08-14	2022	8	3	Austria

customer_id	date_id	product_sku	units_sold	category	group	product_name	price	date	year	month	country_id	country_name	order_revenue
815	12	P018	4	Snow	Apparel	M'S Loft Parka	490	2022-01-12	2022	january	5	Sweden	1960
815	481	P005	3	Cycling	Apparel	Essential Road Sock	20	2023-04-26	2023	april	5	Sweden	60
815	651	P015	1	Snow	Goggles	Fovea	200	2023-10-13	2023	october	5	Sweden	200
676	12	P011	1	Snow	Helmets	Obex Bc Mips	270	2022-01-12	2022	january	5	Sweden	270
676	124	P007	2	Cycling	Armor	Vpd System Torso	260	2022-05-04	2022	may	5	Sweden	520
...	...	...	...	...	...	...	...	...	...	...	...	...	...
347	259	P014	2	Snow	Goggles	Nexal	230	2022-09-16	2022	september	3	Austria	460
347	467	P009	2	Cycling	Armor	Joint Vpd 2.0 Elbow	100	2023-04-12	2023	april	3	Austria	200
347	99	P004	1	Cycling	Apparel	M'S Essential Road	120	2022-04-09	2022	april	3	Austria	120
347	189	P006	3	Cycling	Apparel	M'S Reform Enduro Jersey	90	2022-07-08	2022	july	3	Austria	270
347	226	P008	4	Cycling	Armor	Joint Vpd 2.0 Knee	120	2022-08-14	2022	august	3	Austria	480

We calculated the value of each order and stored it in another column known as 'order\_revenue,' which we calculated as: `df['order_revenue'] = df['price'] * df['units_sold']`.

## 1. Year-Wise Revenue Analysis

Total Revenue generated till date is: 4931560

```
Percentage of revenue by year:  
year  
2023    60.24%  
2022    39.76%  
Name: order_revenue, dtype: object
```

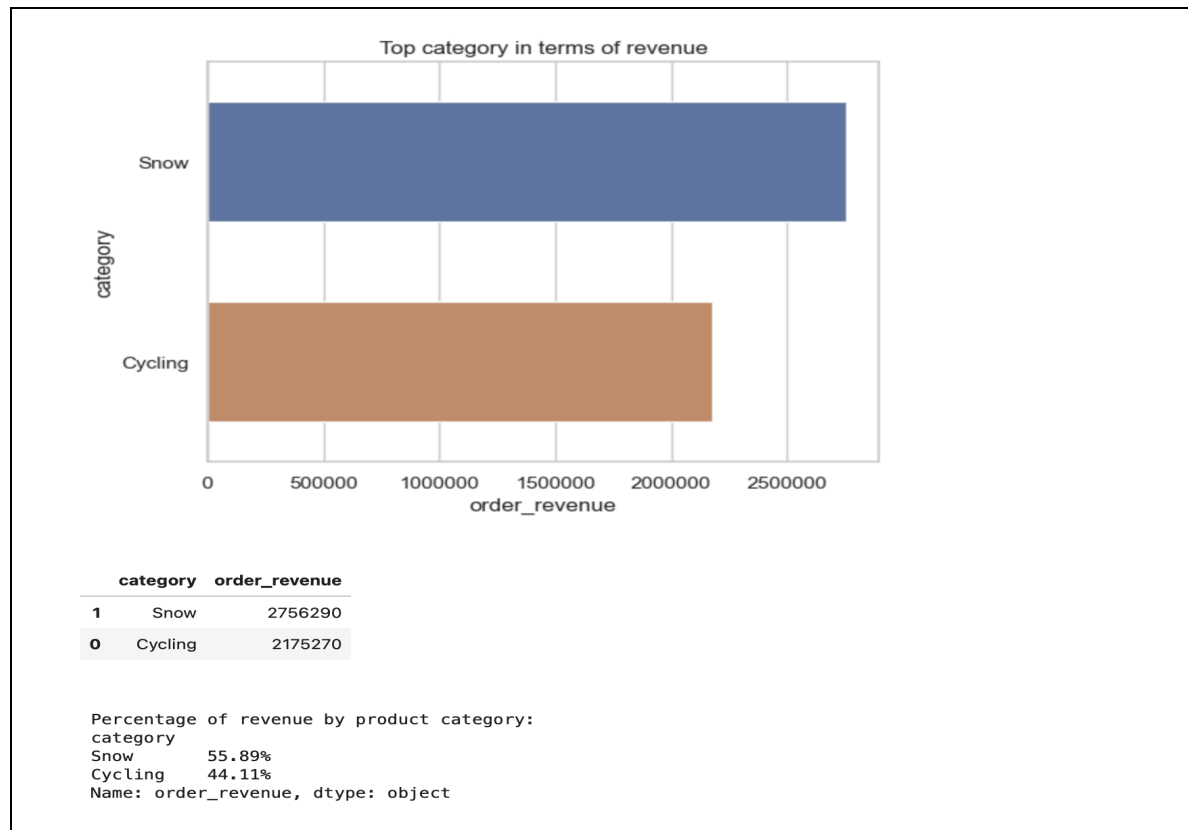
```
year  
2022    1,960,960  
2023    2,970,600  
Name: order_revenue, dtype: object
```

Percentage Increase in Revenue from 2022 to 2023: 51.49%

Now we will observe the trends category wise.

	category	units_sold
0	Cycling	15811
1	Snow	9206

```
Percentage of sales by product category:  
category  
Cycling    63.20%  
Snow       36.80%  
Name: units_sold, dtype: object
```



We can observe that the snow category has fewer sales but still contributes more to total revenue than the cycling category. This provides us with an early indication that the snow category has products with a higher average price and, therefore, a higher Average Order Value (AOV).

We see similar trends in the data for 2023 alone, where the patterns are consistent with the overall trends.

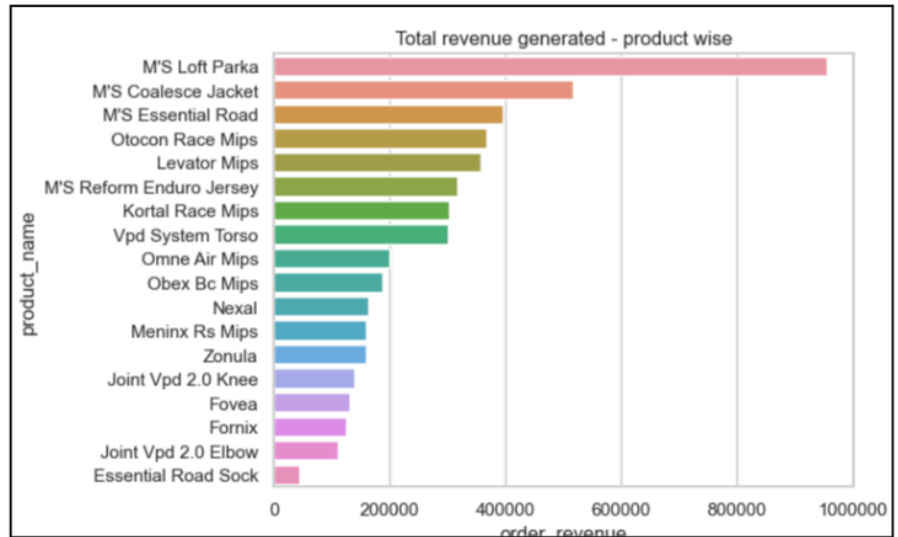
## 2. Product-Wise Revenue Analysis

	product_name	category	order_revenue
0	M'S Loft Parka	Snow	954,520
1	M'S Coalesce Jacket	Snow	517,800
2	M'S Essential Road	Cycling	396,360
3	Otocon Race Mips	Cycling	367,950
4	Levator Mips	Snow	358,050
5	M'S Reform Enduro Jersey	Cycling	316,260
6	Kortal Race Mips	Cycling	302,400
7	Vpd System Torso	Cycling	300,040
8	Omne Air Mips	Cycling	198,900
9	Obex Bc Mips	Snow	188,190
10	Nexal	Snow	164,220
11	Meninx Rs Mips	Snow	158,760
12	Zonula	Snow	158,470
13	Joint Vpd 2.0 Knee	Cycling	138,240
14	Fovea	Snow	131,000
15	Fornix	Snow	125,280
16	Joint Vpd 2.0 Elbow	Cycling	110,200
17	Essential Road Sock	Cycling	44,920

	product_name	category	order_revenue	percent
0	M'S Loft Parka	Snow	954520	19.36
1	M'S Coalesce Jacket	Snow	517800	10.50
2	M'S Essential Road	Cycling	396360	8.04
3	Otocon Race Mips	Cycling	367950	7.46
4	Levator Mips	Snow	358050	7.26
5	M'S Reform Enduro Jersey	Cycling	316260	6.41
6	Kortal Race Mips	Cycling	302400	6.13
7	Vpd System Torso	Cycling	300040	6.08
8	Omne Air Mips	Cycling	198900	4.03
9	Obex Bc Mips	Snow	188190	3.82
10	Nexal	Snow	164220	3.33
11	Meninx Rs Mips	Snow	158760	3.22
12	Zonula	Snow	158470	3.21
13	Joint Vpd 2.0 Knee	Cycling	138240	2.80
14	Fovea	Snow	131000	2.66
15	Fornix	Snow	125280	2.54
16	Joint Vpd 2.0 Elbow	Cycling	110200	2.23
17	Essential Road Sock	Cycling	44920	0.91

revneue genrated by top 3 products alone in %: 37.89

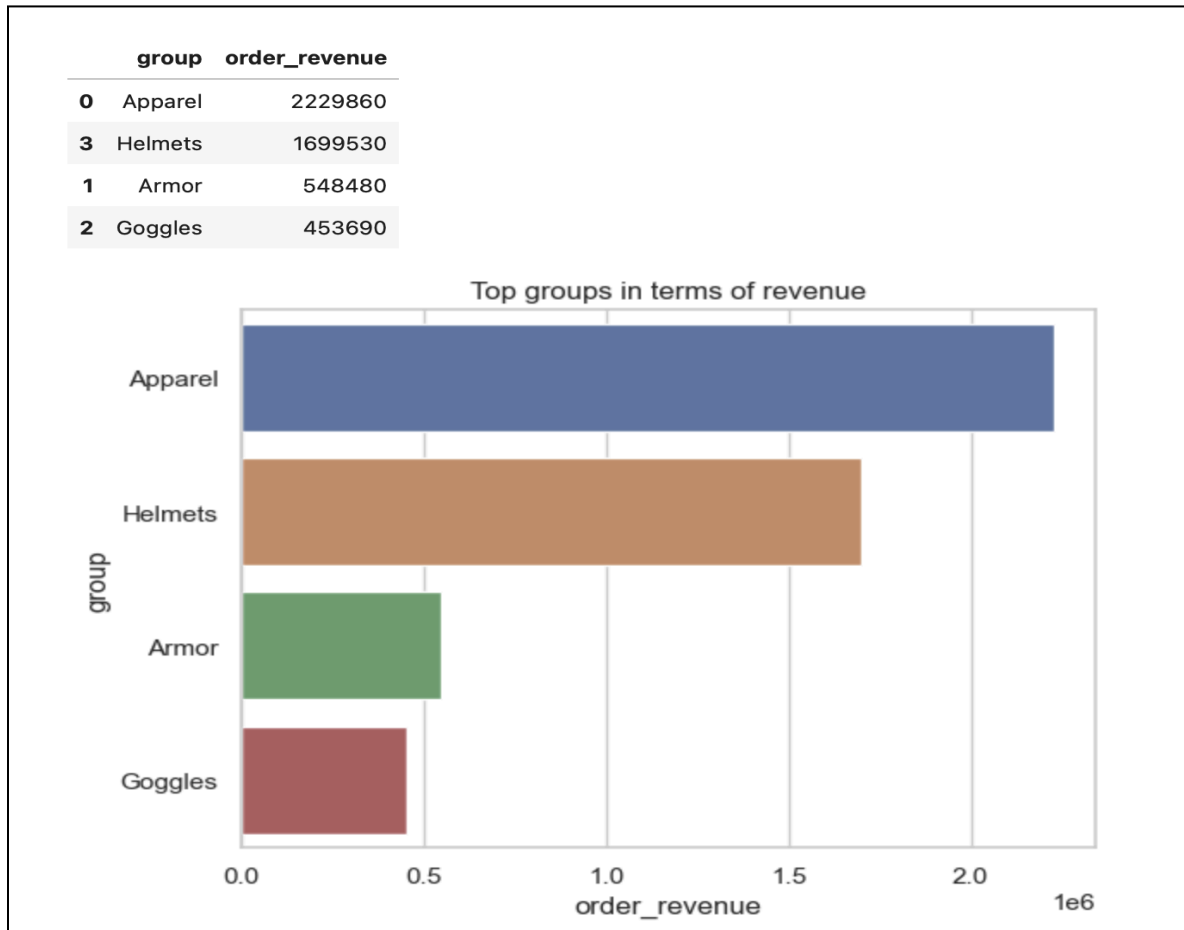


We can observe that the product 'Loft Parka (men),' out of all the 18 products, is our 'hero product' and contributes almost 20% of the total revenue.

The top 3 products contribute to almost 37% of the revenue, which indicates a highly skewed revenue distribution among the products and can be a huge blow to revenue if we run out of stock for these products; we have to maintain our inventory accordingly.

Two products from the snow category alone contribute 30% to revenue, showing dependency on the snow category. We can distribute it better toward cycling.

### 3. Revenue Analysis by Groups

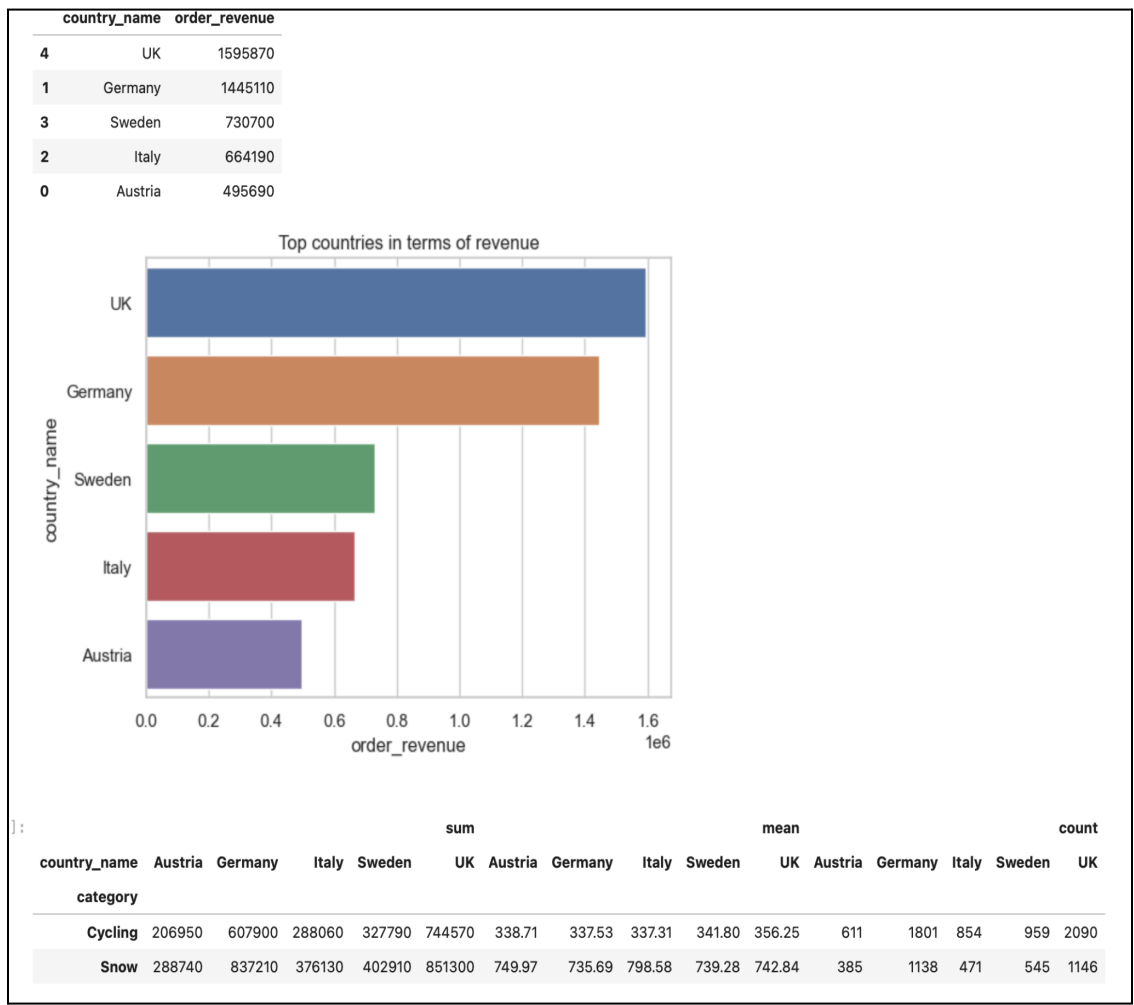


Apparel and helmets tend to generate almost 80% of the revenue.

# 4. Revenue Analysis by Country

We can see that the UK and Germany contribute the most to the revenue.

Therefore, we can plan to optimize pricing and delivery costs for these locations to increase profits. We can also develop marketing and sales strategies targeting customers in these countries.



Most popular product category in each country:

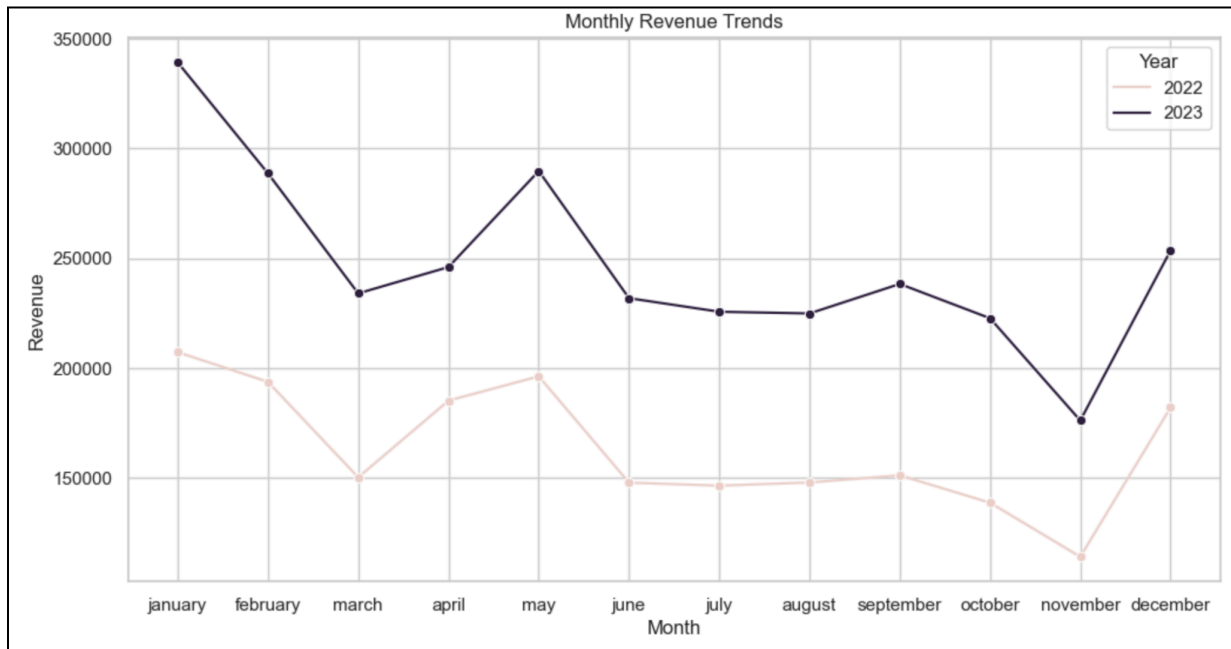
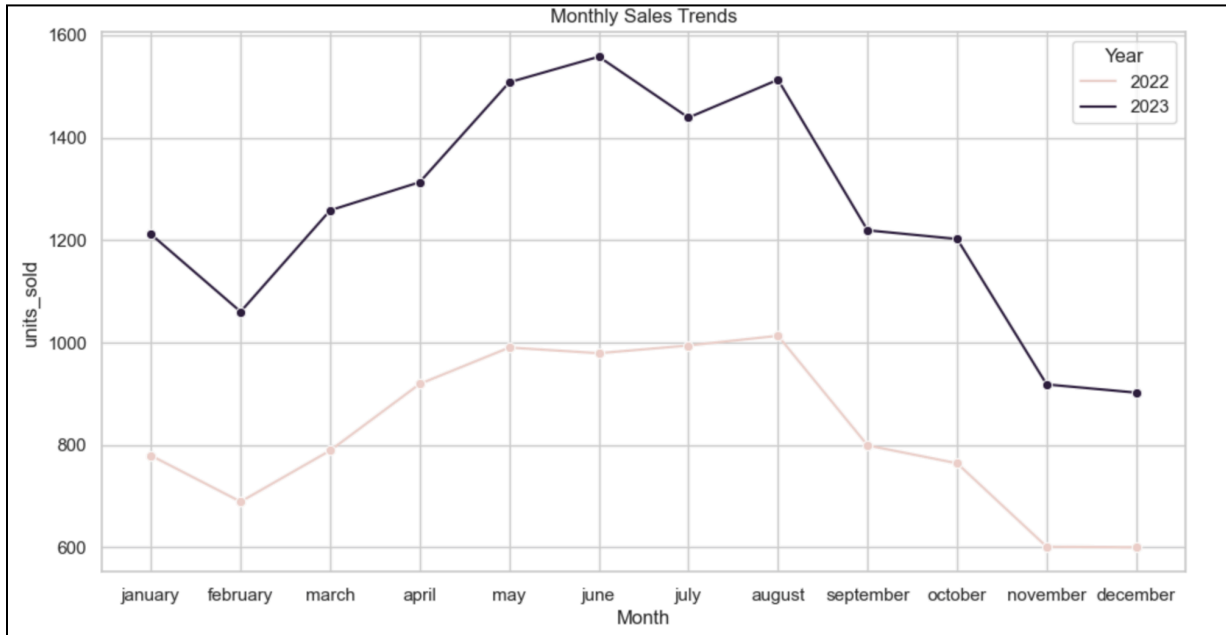
	country_name	category	units_sold
0	Austria	Cycling	1531
2	Germany	Cycling	4469
4	Italy	Cycling	2132
6	Sweden	Cycling	2374
8	UK	Cycling	5305

Most popular product in each country:

	country_name	product_name	units_sold
17	Austria	Zonula	64
35	Germany	Zonula	198
53	Italy	Zonula	82
71	Sweden	Zonula	67
89	UK	Zonula	278

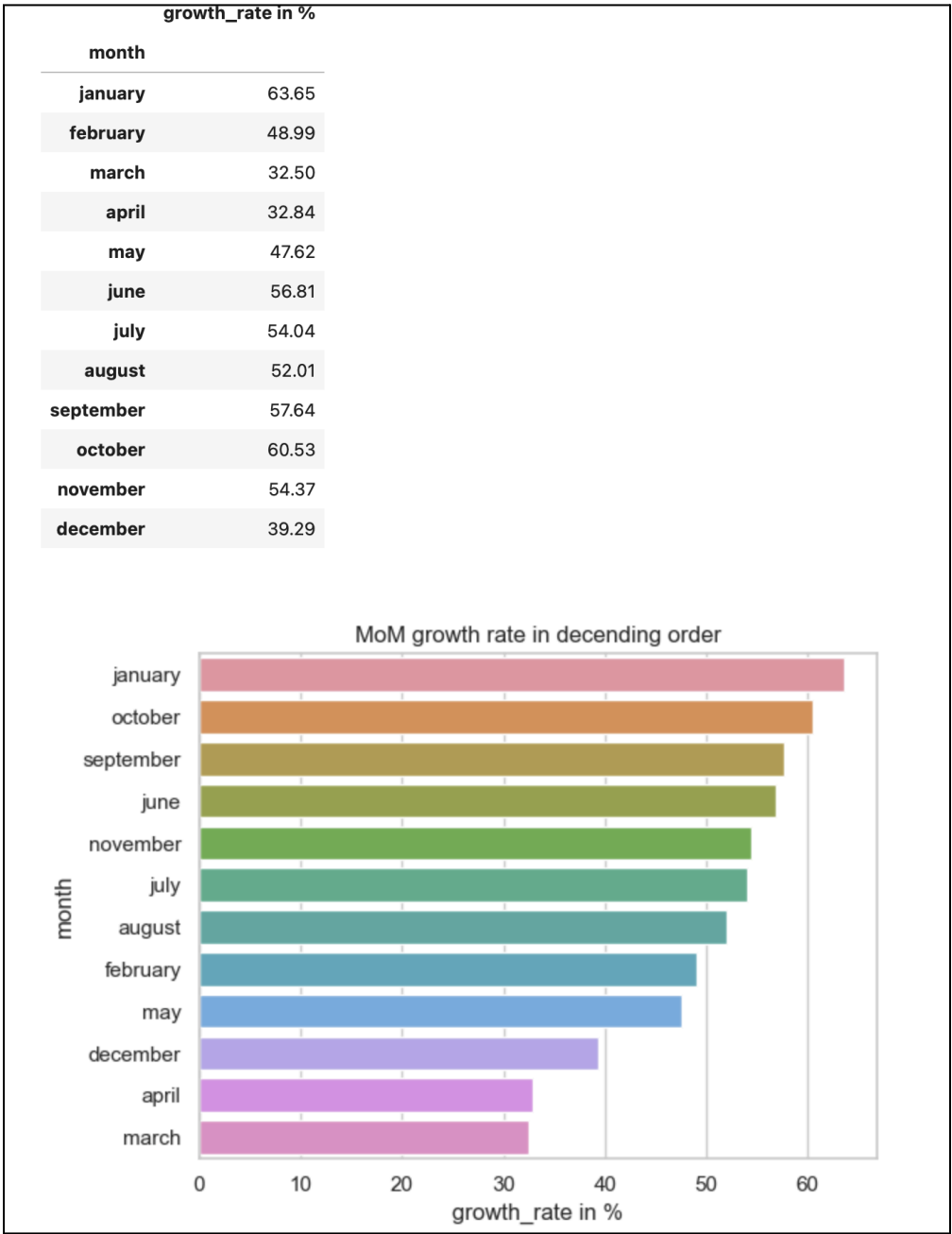
The popular products and categories in the respective countries.

## 5. Revenue and Sales Analysis by Month



Winter months tend to generate more revenue, indicating a clear seasonal trend. This is largely driven by the sales of snow category products. While summer months, associated with the cycling category, also see significant sales, they do not contribute to the revenue as much as the winter months do.

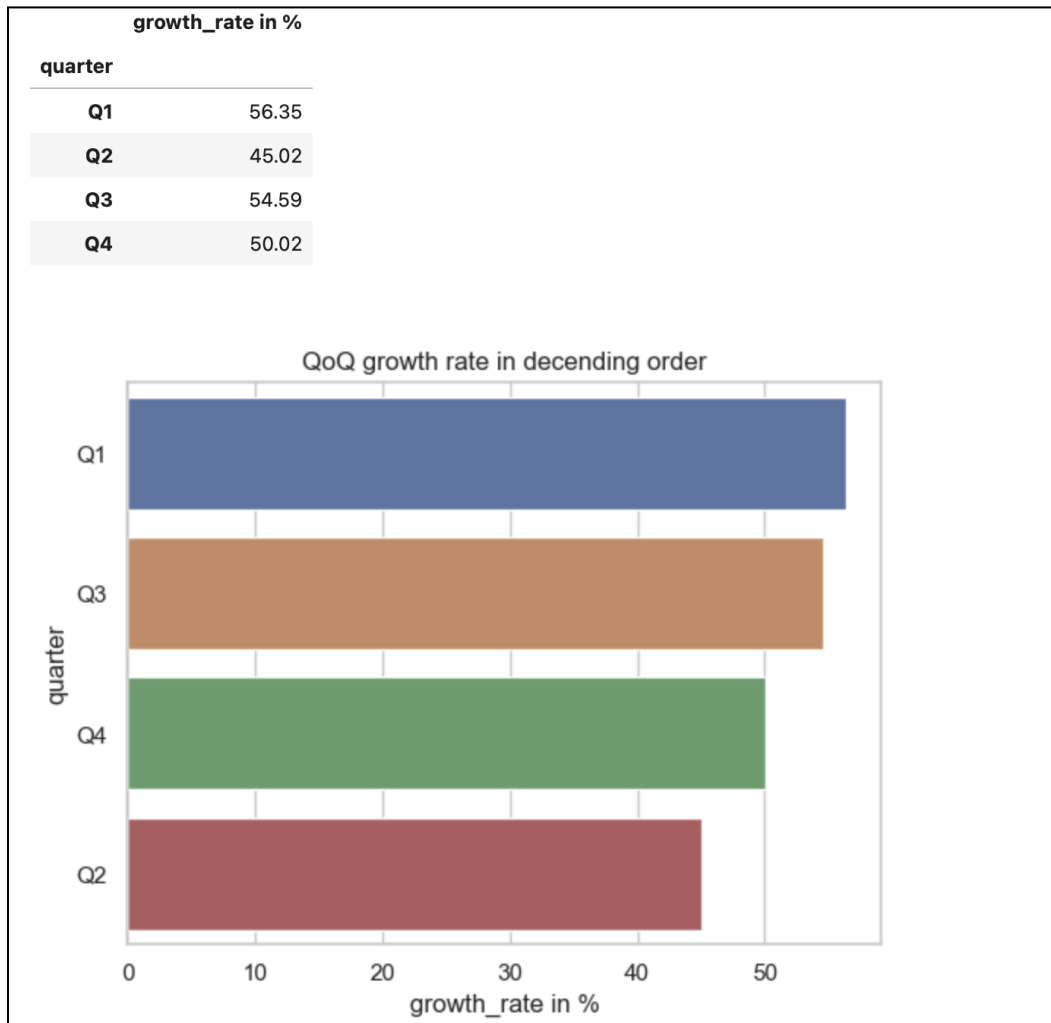
## 6. A) Growth Rate Analysis Month-on-Month



A higher growth rate in the winter months is a good sign, as we are experiencing significant growth during our revenue-generating season.



## 6. B) Growth Rate Analysis Quarter-on-Quarter



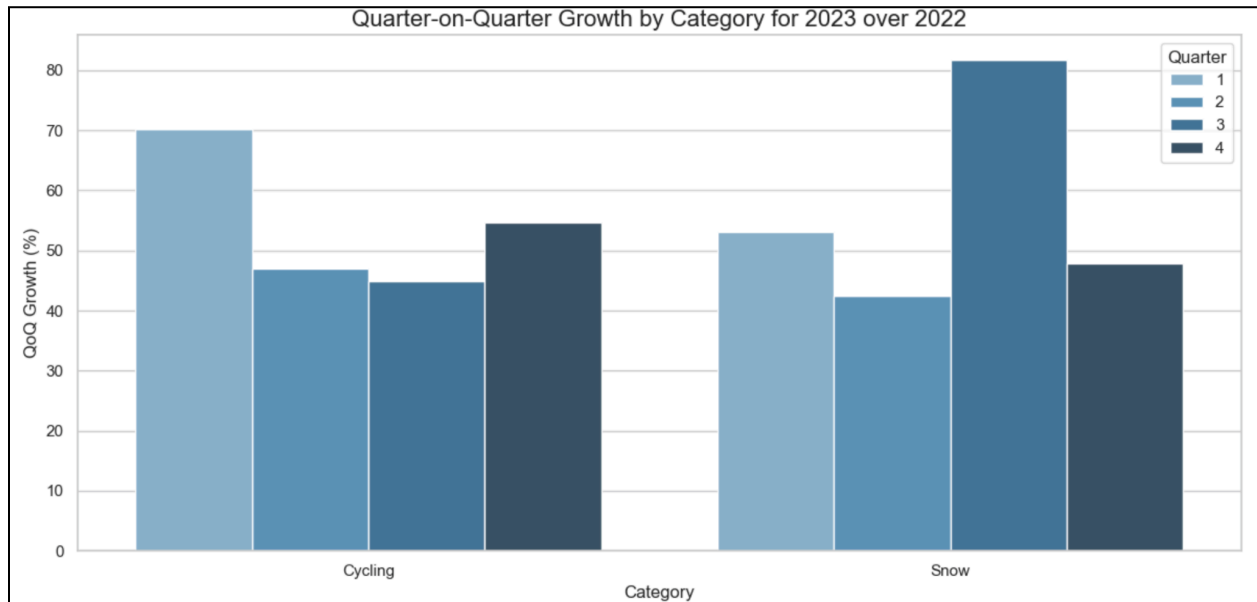
Q2 is the only quarter where the growth rate is less than 50%, showing that our revenue growth is a little lower in the second quarter.

## 7. Category-Wise (QoQ) Growth Rate Analysis

	category	quarter	order_revenue_2022	order_revenue_2023	QoQ_growth
0	Cycling	1	102890	175080	70.16
1	Cycling	2	301980	443940	47.01
2	Cycling	3	328550	476110	44.91
3	Cycling	4	136190	210530	54.59
4	Snow	1	448340	686750	53.18
5	Snow	2	227310	323620	42.37
6	Snow	3	117050	212740	81.75
7	Snow	4	298650	441830	47.94

Cycling in Q1: People tend to buy cycling products before the spring and summer seasons begin.

Snow in Q3: People tend to purchase snow category gear before the skiing season starts.



## 8. Analysis of Average Order Value

```
Overall AOV 493.156

Overall AOV by country  country_name  AOV
0      Austria 497.68
1      Germany 491.70
2      Italy 501.28
3      Sweden 485.84
4      UK 493.16

Overall AOV by category  category  AOV
0  Cycling 344.46
1    Snow 747.98

Overall AOV 2022 490.24

Overall AOV 2023 495.1
```

```
AOV by country for 2022  country_name  AOV
0      Austria 511.61
1      Germany 488.73
2      Italy 480.43
3      Sweden 486.85
4      UK 490.86

AOV by country for 2023  country_name  AOV
0      Austria 488.90
1      Germany 493.81
2      Italy 514.53
3      Sweden 485.12
4      UK 494.62

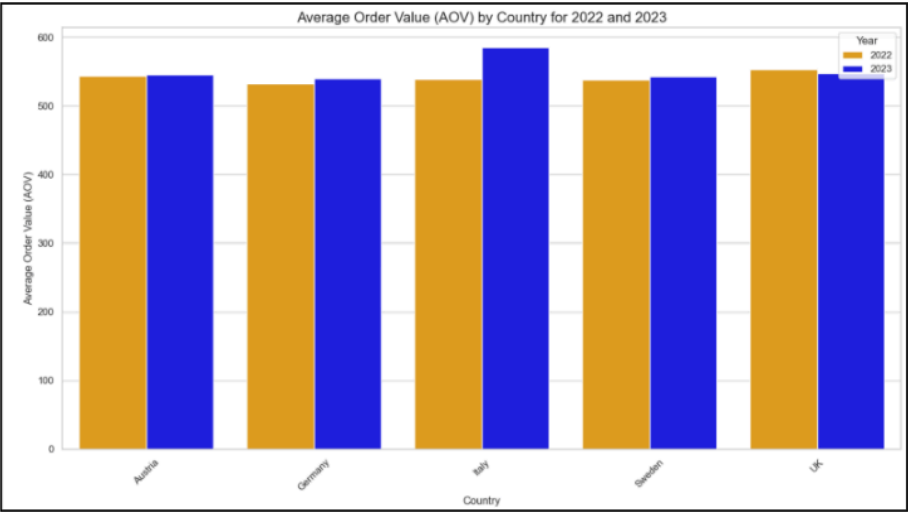
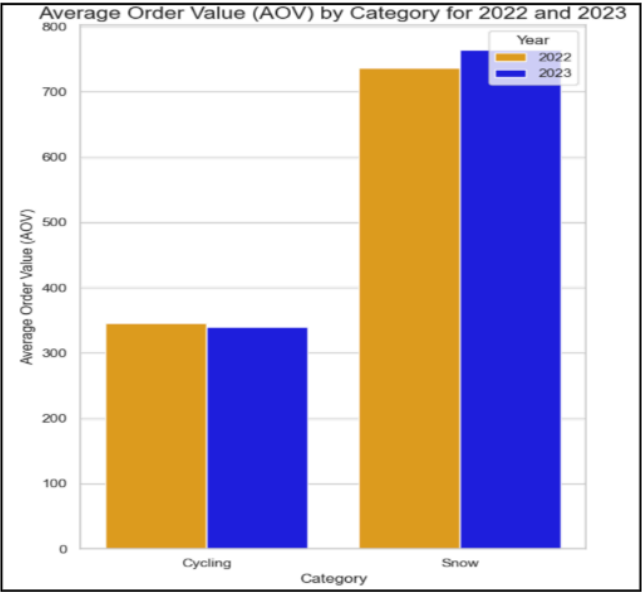
AOV by country for 2022  category  AOV
0  Cycling 344.81
1    Snow 738.40

AOV by country for 2023  category  AOV
0  Cycling 344.23
1    Snow 754.39

percentage of orders greater than AOV: 35.34
percentage of orders greater than AOV 2023: 35.65

Percentage of Orders Greater Than AOV by Country:
country_name  percentage_greater_than_aov
0      Austria 35.24
1      Germany 35.35
2      Italy 35.77
3      Sweden 34.51
4      UK 35.57

Percentage of Orders Greater Than AOV by Category:
category  percentage_greater_than_aov
0  Cycling 19.64
1    Snow 62.25
```



There has not been much of an increase in the average order value from 2022 to 2023 across all countries and categories. This suggests that revenue per order is stagnant and that the increased revenue is solely due to the rise in the number of orders.

To achieve sustainable growth and improve engagement rates, we should focus on increasing the average order value through cross-selling products using various machine learning methods.

# 9. Retention and Churn Rate Analysis

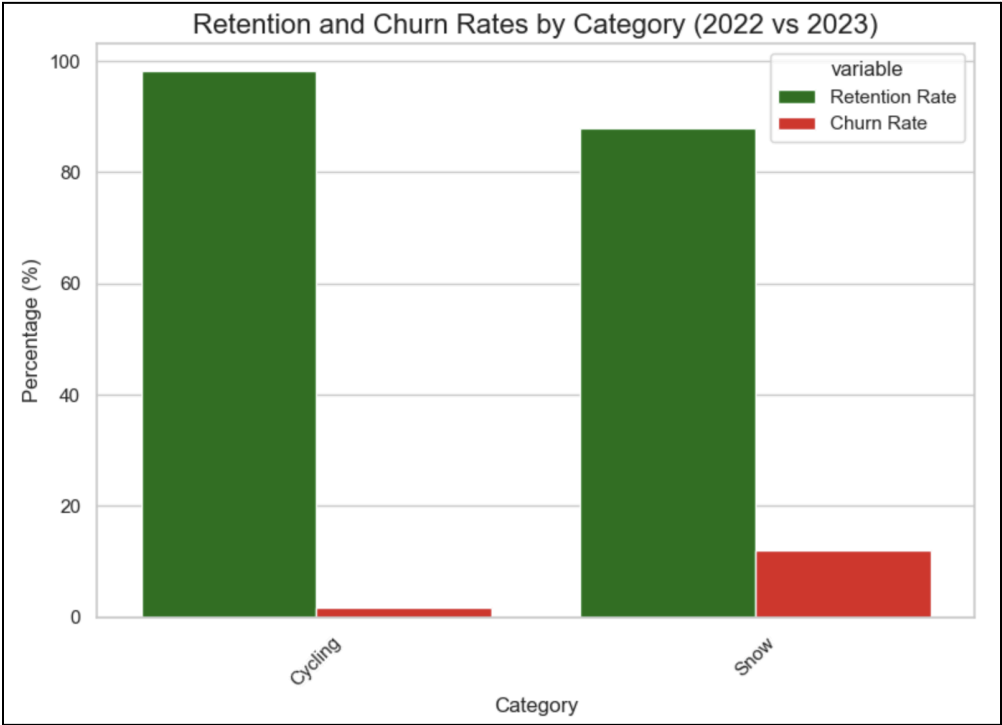
Retention Rate: 99.80%  
Churn Rate: 0.20%

```
retention_by_category in % is:
{'Cycling': 98.16414686825054, 'Snow': 87.85166240409207}

churn_by_category in % is:
{'Cycling': 1.8358531317494602, 'Snow': 12.148337595907927}

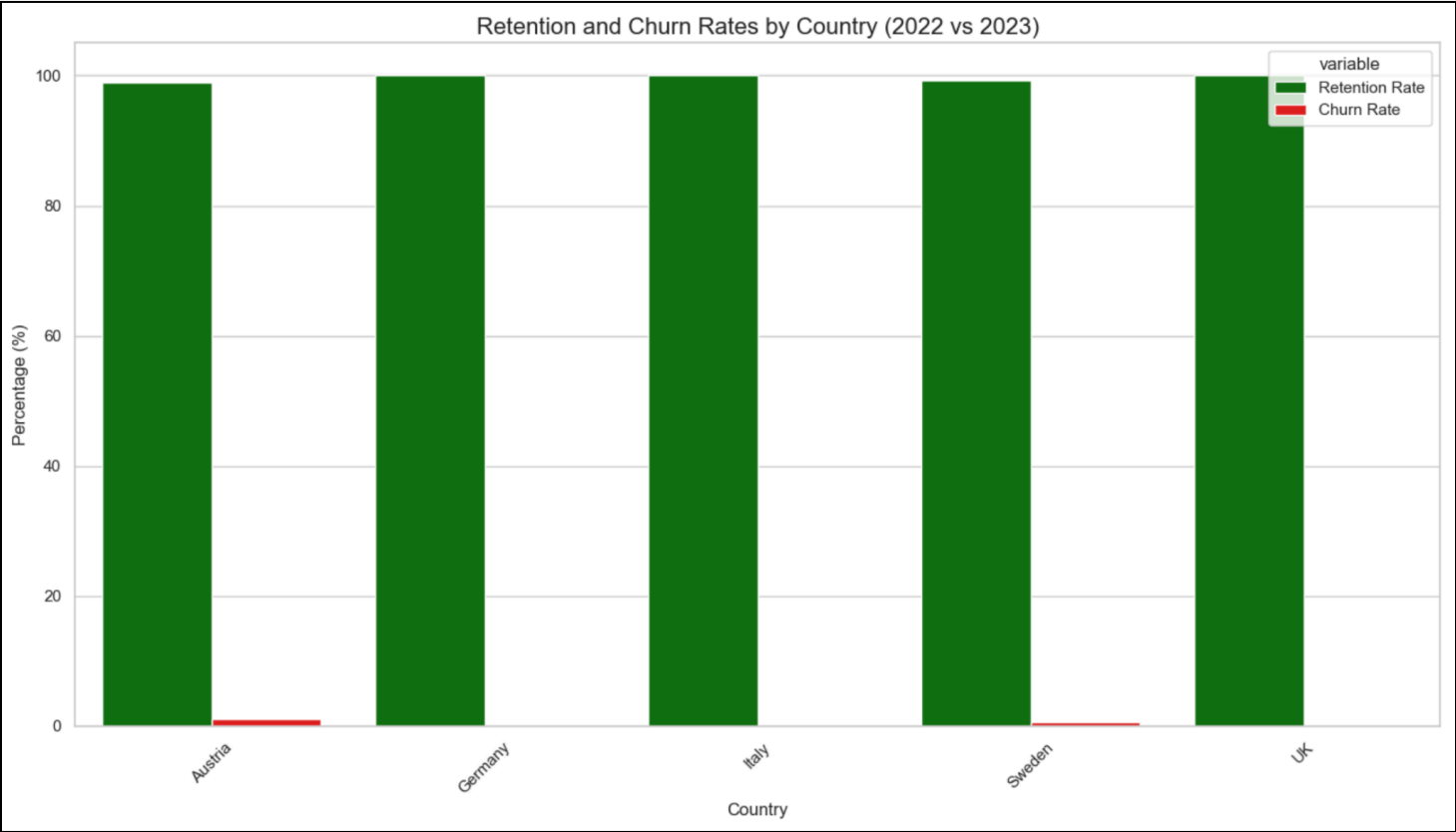
retention_by_country in % is:
{'Austria': 98.9795918367347, 'Germany': 100.0, 'Italy': 100.0, 'Sweden': 99.33774834437085, 'UK': 100.0}

churn_by_country in % is:
{'Austria': 1.0204081632653061, 'Germany': 0.0, 'Italy': 0.0, 'Sweden': 0.6622516556291391, 'UK': 0.0}
```



We observe a churn rate of 12% in the snow category, which is concerning because our revenue is heavily driven by snow category products, as shown above.

The churn rate by country is not significant, which is a positive sign, indicating that we are not losing customers from any particular region.



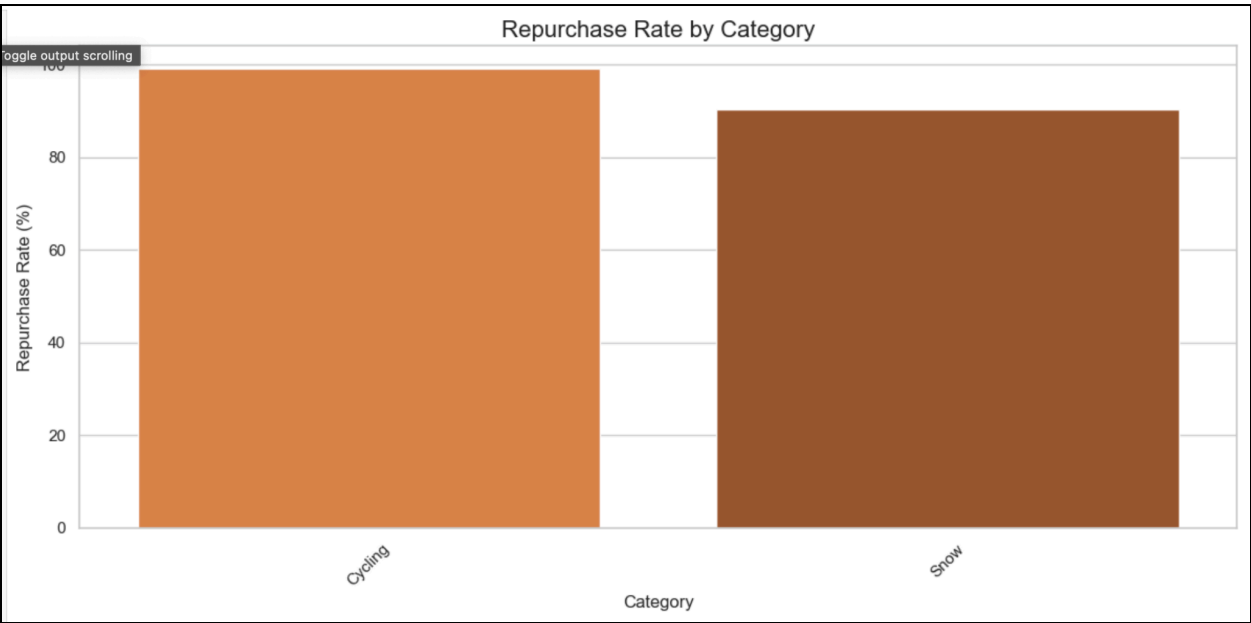
# 10. Repurchase Rate Analysis

Re-Purchase rate:  
100.0

Repurchase rate category wise

	category	repeat_customers	total_customers	repurchase_rate
0	Cycling	989	998	99.10
1	Snow	881	974	90.45

A decrease in the repurchase rate from the snow category can be alarming. We should encourage customers to revisit snow category products and drive sales through improved personalized recommendations, increasing cross-sell opportunities. Additionally, targeting snow category customers with promotional offers and marketing techniques can help boost engagement.



## 11. Customer Lifespan Value Analysis

```
#1 Average Order Value (AOV)
AOV = order_df['order_revenue'].mean() ##calculated the AOV

#2 Calculate Purchase Frequency (Total)
unique_customers = order_df['customer_id'].nunique()
purchase_frequency = total_orders / unique_customers ## calculating the purchase frequency of customers

#3 Average Customer Lifespan (assuming 2 years)
average_customer_lifespan = 2 # CLV for a life span of 2 years

#4 CLV (Total)
CLV_total = AOV * purchase_frequency * average_customer_lifespan
print("Average Customer Lifespan value is:", CLV_total)
print("Average purchase frequency is",purchase_frequency)
print("AOV is",AOV)

Average Customer Lifespan value is: 9863.12
Average purchase frequency is 10.0
AOV is 493.156
```

We have a strong Customer Lifetime Value (CLV), primarily due to a higher average purchase frequency, which is a positive sign. However, we can further improve the Average Order Value (AOV) to generate more revenue from each customer over their entire lifespan.

A higher AOV drives more profit, as delivery and marketing costs are fixed, whereas a higher purchase frequency increases revenue but does not contribute to profits as much as a higher AOV does.



## 2. PROFITABILITY CALCULATION

**Revenue Calculation:** Revenue can be calculated directly from the dataset using the column 'order\_revenue', which represents the total amount earned from each order.

**Total Revenue = Sum of all values in the 'order\_revenue' column.**

To calculate profitability accurately, we need to account for the following cost factors:

1. **Cost of Goods Sold (COGS):**
  - Represents the direct cost of producing the goods that were sold.
  - Assume a fixed percentage of the price for each product to calculate COGS, which could be 40-50% of the product price for the E-commerce industry.
  - $\text{COGS} = \text{units\_sold} * (\text{cost per unit of product})$ .
2. **Shipping and Fulfillment Costs:**
  - These costs include expenses for packaging, shipping, handling, and logistics.
  - Assume a fixed cost per order or a percentage of the order value. For instance, we can assume shipping costs are 5-10% of the order\_revenue.
3. **Marketing and Advertising Costs:**
  - The amount spent on marketing campaigns to drive traffic to the website and promote products.
  - Assume a fixed percentage of revenue (e.g., 10-15%) for marketing expenses.
4. **Transaction Fees:**
  - Includes credit card fees, payment gateway charges, and other transaction-related fees.
  - Typically around 2-3% of the total order value.
5. **Operating Costs:**
  - Fixed costs associated with running the business (e.g., salaries, rent, software, technology costs).
  - Assume a monthly fixed cost that can be spread evenly across the orders in the dataset (for example 1000 euros per month).
6. **Returns and Refunds:**
  - Some products might be returned or refunded, which directly impacts profitability.
  - Assume a return rate and subtract the cost of returned items from the revenue.
7. **Customer Acquisition Cost (CAC):**
  - This is the cost associated with acquiring a new customer, including expenses on marketing, sales, promotions, and any other related activities.
  - In our case, we can assume a fixed cost or a percentage of the marketing expenses that goes towards acquiring new customers.
  - **Assumptions for CAC:**
    - Assume that 40% of the total marketing costs are used for customer acquisition.
    - CAC can be calculated as:
      - $\text{CAC} = 0.4 * \text{Marketing Costs}$
8. **Tax rate:**

- This is the cost associated with the taxes to be paid on the gross profit earned on each order.
- Assume that the tax rate is 20% on profit.
  - $\text{Tax} = 0.2 * \text{Gross Profit}$

**Gross Profit = Total Revenue-(COGS+Shipping Costs+Marketing Costs+Transaction Fees+Operating Costs+Return Costs+CAC)**

**Net Profit = Gross Profit - Tax**

```
# Assumptions for cost calculations
COGS_percentage = 0.4 # Cost of Goods Sold as 60% of the price
shipping_percentage = 0.08 # Shipping costs as 8% of order revenue
marketing_percentage = 0.12 # Marketing costs as 12% of order revenue
transaction_fee_percentage = 0.025 # Transaction fees as 2.5% of order revenue
return_rate_percentage = 0.05 # Return costs as 5% of order revenue
monthly_operating_cost = 10000 # Fixed monthly operating costs
CAC_percentage = 0.4 # Customer Acquisition Cost as 40% of marketing costs
Tax_percentage = 0.2
```

	order_revenue	COGS	Shipping_Costs	Marketing_Costs	Transaction_Fees	\
0	1960	784.00	156.80	235.20	49.00	
1	60	24.00	4.80	7.20	1.50	
2	200	80.00	16.00	24.00	5.00	
3	270	108.00	21.60	32.40	6.75	
4	520	208.00	41.60	62.40	13.00	
	Operating_Costs	Return_Costs	CAC	Gross_Profit	Tax	Net_Profit
0	12.50	98.00	94.08	530.42	106.08	424.34
1	11.11	3.00	2.88	5.51	1.10	4.41
2	12.50	10.00	9.60	42.90	8.58	34.32
3	12.50	13.50	12.96	62.29	12.46	49.83
4	10.00	26.00	24.96	134.04	26.81	107.23

# 3. DATA SCIENCE AND MACHINE LEARNING TECHNIQUES

## 1. Forecasting

- Problem being solved: Predict future sales, revenue, or demand.
- Approach: Use time-series forecasting models like ARIMA, Prophet.
- How it can help:
  - It can improve inventory management, ensures products are in stock based on **forecasted demand, reducing lost sales or revenue due to stockouts.**
  - Forecasting trends could optimize the timing of promotions to increase order size during high demand which ultimately can **improve the AOV.**

## 2. Product Recommendations

- Problem being solved: Suggest products to customers based on their past behavior or purchases.
- Approach: Use models like Collaborative Filtering, Content-Based Filtering, or Matrix Factorization (SVD) to recommend products.
- How it can help:
  - Cross-sell/Up-sell: By recommending related products, you increase the customer's basket size and recommending complementary products will lead to **higher AOV and hence revenue.**
  - Improves customer experience: Personalized product recommendations keep customers engaged, increasing loyalty and hence can **increase repurchase rate.**

## 3. Customer Segmentation

- Problem being solved: Group customers based on their behaviors, preferences, or demographics.
- Approach: Use K-Means Clustering, Hierarchical Clustering, or DBSCAN to identify segments.
- How it helps:
  - Personalized Marketing: Segmentation allows for more targeted campaigns, targeted efforts will lead to higher engagement, retention and low churn, which improve **conversion rates.**
  - Loyalty programs: Tailoring loyalty offers to high-value customers can increase retention and **hence can increase CLV for them.**

#### 4. Price Optimization

- Problem being solved: Find the optimal price point that maximizes profit without reducing sales.
- Approach: Use Regression models or algorithms like gradient descent for dynamic pricing, considering factors like demand, competition, seasonality, and customer behavior.
- How it helps:
  - Maximizes profit: Adjusting pricing based on demand and elasticity ensures **better margins** and hence **higher AOV and profits**.
  - Increases sales volume: Dynamic pricing can help attract customers when demand is low and hence boost total revenue and profits.

#### 5. Churn Prediction & Optimization

- Problem being solved: Identify customers at risk of leaving the platform and take preventive actions.
- Approach: Use Logistic Regression, Random Forest, or XGBoost for churn prediction.
- How it helps:
  - Prevents customer loss: Early identification of churn enables you to retain customers through targeted campaigns or incentives. As **reducing churn** directly **improves retention**, retaining customers for a longer period **increases their lifetime value** and **increases repeat purchases**.
  - **Lowers acquisition costs**: Retaining existing customers is more cost-effective than acquiring new ones.

#### 6. Customer Lifetime Value (CLV) Prediction

- Problem being solved: Estimate the future value of a customer based on their past behavior and transactions.
- Approach: Use Cohort Analysis, RFM (Recency, Frequency, Monetary) or models like CLV prediction models using Gradient Boosting or Neural networks.
- How it helps:
  - Improves customer targeting: Focus on high-CLV customers for loyalty programs or exclusive offers as improving CLV by understanding and retaining high-value customers **boosts long-term profitability**. Also, Identifying low-CLV customers helps in creating targeted win-back campaigns.
  - Better resource allocation: Helps in deciding how much to spend on retaining specific customer segments.