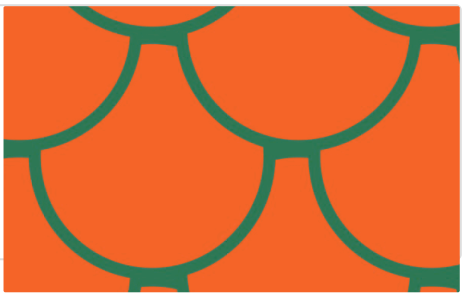
DG Task

Data Source

Adventure Works

Kaggle is the world's largest data science community with powerful tools and resources to help you achieve your data science goals.

https://www.kaggle.com/datasets/ukveteran/adventure-works/data



Fact and Dimension tables

Fact Tables

1. sales (stored as yearly tables: sales_2015, sales_2016, sales_2017)
2. returns

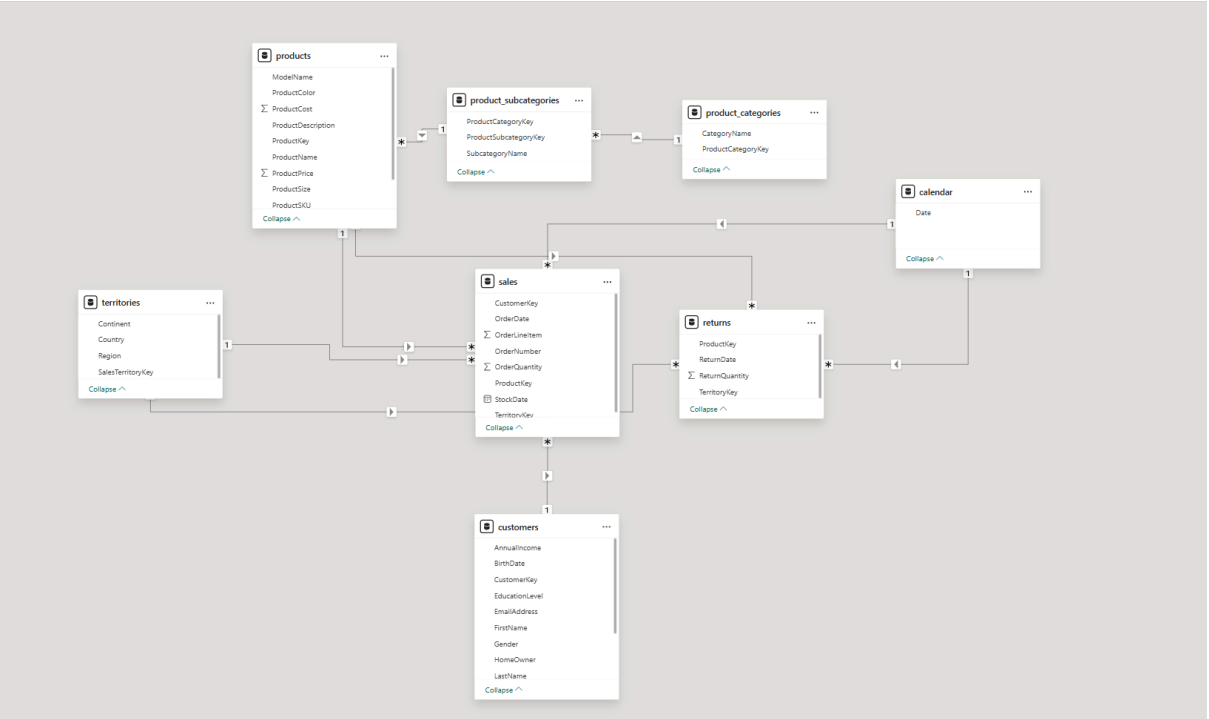
Dimension Tables

1. customers
2. products
3. product_categories
4. product_subcategories
5. calendar
6. territories

Connection of columns across tables

sales_*.ProductKey → product.ProductKey
sales_*.CustomerKey → customers.CustomerKey
sales_*.TerritoryKey → territories.SalesTerritoryKey

returns.TerritoryKey → territories.SalesTerritoryKey
returns.ProductKey → products.ProductKey



From: table (column)	Relationship	To: table (column)	Status
product_subcategories (Produ...	* — 1	product_categories (ProductC...	Active
products (ProductSubcategory...	* — 1	product_subcategories (Produ...	Active
returns (ProductKey)	* — 1	products (ProductKey)	Active
returns (ReturnDate)	* — 1	calendar (Date)	Active
returns (TerritoryKey)	* — 1	territories (SalesTerritoryKey)	Active
sales (CustomerKey)	* — 1	customers (CustomerKey)	Active
sales (OrderDate)	* — 1	calendar (Date)	Active
sales (ProductKey)	* — 1	products (ProductKey)	Active
sales (TerritoryKey)	* — 1	territories (SalesTerritoryKey)	Active

Tasks

4.1 Part - A Data Model & SQL Analysis

1) Data Model Understanding (short write-up)

Main Tables

1. sales_2015
2. sales_2016
3. sales_2017
4. returns
5. customers
6. products
7. product_categories
8. product_subcategories
9. territories

Data Model and Table Relationships

Sales Fact Tables

- Sales data is stored across multiple yearly tables (`sales_2015`, `sales_2016`, `sales_2017`) that share the same schema and together form the central **Sales fact table**.
- Each sales record is uniquely identified by `OrderNumber`, which serves as the **primary key**.
- The Sales tables capture transactional details such as `OrderDate`, `StockDate`, `OrderQuantity`, and `OrderLineItem`.
- `ProductKey`, `CustomerKey`, and `TerritoryKey` act as **foreign keys**, linking sales transactions to the `products`, `customers`, and `territories` dimension tables.
- The `OrderDate` column can be logically aligned with the Calendar table to support **time-based analysis**, even though no explicit foreign-key relationship is enforced.

Product Dimensions and Hierarchy

- The `products` table is identified by `ProductKey` as its **primary key** and contains detailed product attributes such as name, cost, price, and model information.
- Products are linked to `product_subcategories` using `ProductSubcategoryKey`.
- `product_subcategories` are linked to `product_categories` using `ProductCategoryKey`.
- This hierarchy enables analysis at multiple product levels, such as category and subcategory.

Customer, Territory, and Returns Tables

- The `customers` dimension uses `CustomerKey` as its **primary key** and contains demographic and socioeconomic attributes.
- The `territories` dimension is identified by `SalesTerritoryKey` as its **primary key** and provides geographic context such as region, country, and continent.
- The `returns` table acts as a separate **fact table** that captures returned quantities.

- `returns` data is linked to the Products and Territories dimensions using `ProductKey` and `TerritoryKey`, allowing return trends to be analyzed alongside sales.

---

2) SQL for Business Questions

1. **Which are the top 10 products by total revenue generated across all years (2015–2017)?**

```
select products.ProductName, sum(sales.OrderQuantity * products.Product
Price) as Revenue
from sales
left join products
on products.ProductKey = sales.ProductKey
group by products.ProductName
order by Revenue desc
limit 10;
```

Result:

	CategoryName	ProductName	Revenue
▶	Bikes	Mountain-200 Black, 46	1241753.509199993
	Bikes	Mountain-200 Black, 42	1233557.1163999934
	Bikes	Mountain-200 Silver, 38	1213851.8856000006
	Bikes	Mountain-200 Silver, 46	1182780.591600002
	Bikes	Mountain-200 Black, 38	1165936.875799997
	Bikes	Mountain-200 Silver, 42	1133066.5212000043
	Bikes	Road-250 Black, 52	689373.75
	Bikes	Road-250 Red, 58	661013.4375
	Bikes	Road-250 Black, 48	641379.375
	Bikes	Road-150 Red, 48	640510.3300000016

1. Bikes are the top sellers

2. **How much total revenue was generated in each country (United States, Canada, France, Germany, Australia, and United Kingdom)?**

```
select t.Country, sum(s.OrderQuantity * pr.ProductPrice) as Revenue
from sales as s
left join products as pr
on pr.ProductKey = s.ProductKey
left join territories as t
on t.SalesTerritoryKey = s.TerritoryKey
group by t.Country
order by Revenue desc;
```

Result:

Country	Revenue
United States	7938999.417702192
Australia	7416456.200101364
United Kingdom	2902562.0949000977
Germany	2524679.9728000253
France	2362643.3193
Canada	1769245.8144999298

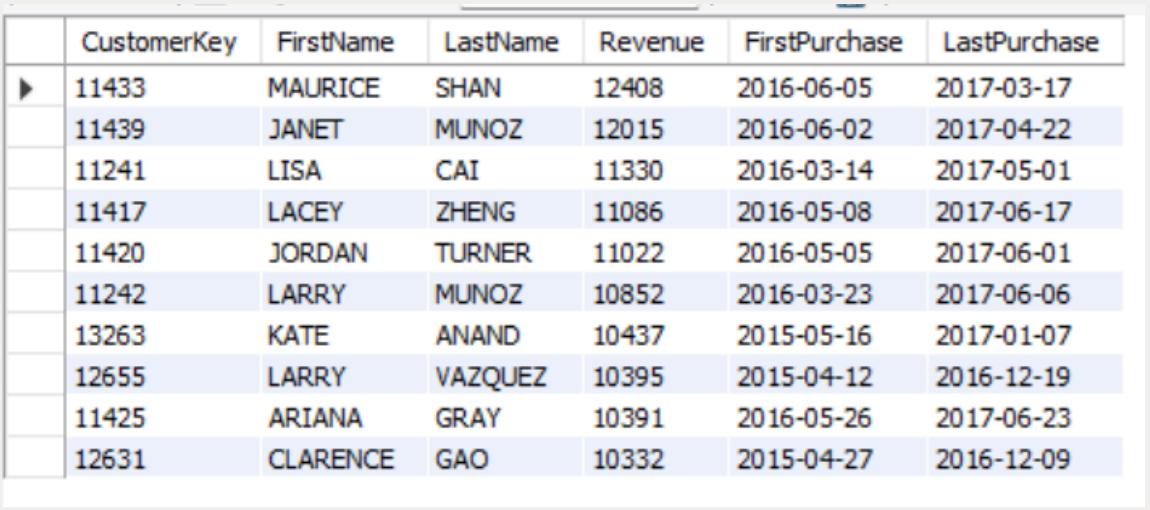
1. The US generated high revenue followed by Australia and UK
2. There is a huge difference of revenue between the first two countries and the remaining countries.

3. **Who are the top 10 most valuable customers based on total revenue generated?**

```
select c.CustomerKey, c.FirstName, c.LastName, round(sum(s.OrderQuantit
y * p.ProductPrice)) as Revenue, min(s.OrderDate) as FirstPurchase, max(s.
OrderDate) as LastPurchase
from sales as s
left join customers as c
```

```
on c.CustomerKey = s.CustomerKey
left join products as p
on p.ProductKey = s.ProductKey
group by c.CustomerKey, c.FirstName, c.LastName
order by Revenue desc
limit 10;
```

Result:



CustomerKey	FirstName	LastName	Revenue	FirstPurchase	LastPurchase
11433	MAURICE	SHAN	12408	2016-06-05	2017-03-17
11439	JANET	MUNOZ	12015	2016-06-02	2017-04-22
11241	LISA	CAI	11330	2016-03-14	2017-05-01
11417	LACEY	ZHENG	11086	2016-05-08	2017-06-17
11420	JORDAN	TURNER	11022	2016-05-05	2017-06-01
11242	LARRY	MUNOZ	10852	2016-03-23	2017-06-06
13263	KATE	ANAND	10437	2015-05-16	2017-01-07
12655	LARRY	VAZQUEZ	10395	2015-04-12	2016-12-19
11425	ARIANA	GRAY	10391	2016-05-26	2017-06-23
12631	CLARENCE	GAO	10332	2015-04-27	2016-12-09

1. The result shows the top 10 customers by revenue with their time on market

4. **Which Region has the highest total return quantity?**

```
select t.Region, sum(r.ReturnQuantity) as Returns
from returns as r
left join territories as t
on r.TerritoryKey = t.SalesTerritoryKey
group by t.Region
order by Returns desc;
```

Result:

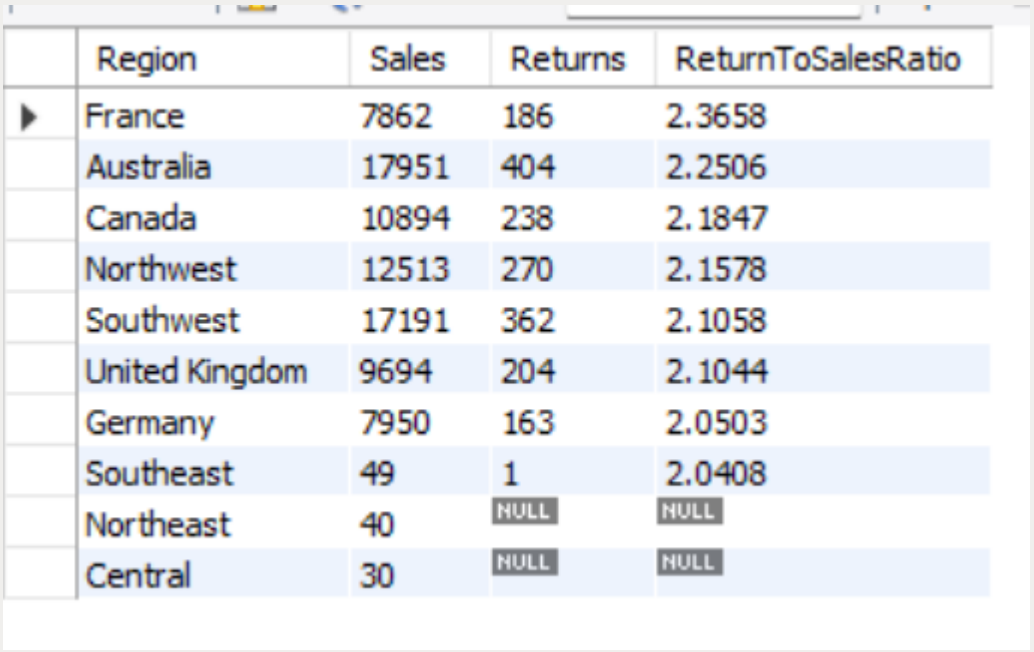
	Region	Returns
▶	Australia	404
	Southwest	362
	Northwest	270
	Canada	238
	United Kingdom	204
	France	186
	Germany	163
	Southeast	1

1. Australia has the highest return quantities over all years followed by Southwest and Northwest regions.
2. Southeast region has the lowest return quantities.

Follow up:

Is the high returns in Australia and Southeast indicates any decline in product quality ?

```
select s.Region, s.Sales, r.Returns, r.returns / s.Sales * 100 as ReturnTo
SalesRatio
from vw_sales_by_region as s
left join vw_returns_by_region as r
on s.Region = r.Region
order by ReturnToSalesRatio desc;
```



Region	Sales	Returns	ReturnToSalesRatio
France	7862	186	2.3658
Australia	17951	404	2.2506
Canada	10894	238	2.1847
Northwest	12513	270	2.1578
Southwest	17191	362	2.1058
United Kingdom	9694	204	2.1044
Germany	7950	163	2.0503
Southeast	49	1	2.0408
Northeast	40	NULL	NULL
Central	30	NULL	NULL

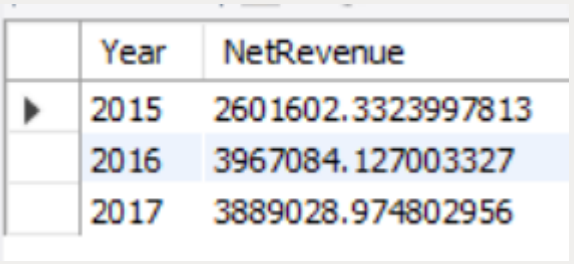
1. The high returns of Australia is simply due to high sales
2. All the countries have negligible return ratios, so there is no sign of decline in product qualities
3. This analysis would be efficient if done on yearly basis

5. **How does net sales revenue vary by year (2015, 2016, and 2017)?**

```
select year(s.OrderDate) as Year, sum(s.OrderQuantity * pr.ProductPrice) -
sum(s.OrderQuantity * pr.ProductCost) as NetRevenue
from sales as s
left join products as pr
```

```
on s.ProductKey = pr.ProductKey
group by Year;
```

Result:



	Year	NetRevenue
▶	2015	2601602.3323997813
	2016	3967084.127003327
	2017	3889028.974802956

1. The Revenue dropped slightly in 2017
2. The year 2016 has the highest revenue

6. **Which product categories and subcategories contribute the most to total revenue?**

```
select
   sub.SubcategoryName,
   cats.CategoryName,
   round(sum(s.OrderQuantity * pr.ProductPrice))  as Revenue,
   round(sum(s.OrderQuantity * (pr.ProductPrice - pr.ProductCost))) as Gro
ssProfit,
   round(sum(s.OrderQuantity * (pr.ProductPrice - pr.ProductCost)) * 1.0 /
        sum(s.OrderQuantity * pr.ProductPrice), 4) as Margin
from sales as s
left join products as pr
on pr.ProductKey = s.ProductKey
left join product_subcategories as sub
on sub.ProductSubcategoryKey = pr.ProductSubcategoryKey
left join product_categories as cats
on cats.ProductCategoryKey = sub.ProductCategoryKey
group by cats.CategoryName, sub.SubcategoryName
```

```
order by Revenue desc
limit 5;
```

Result:

	SubcategoryName	CategoryName	Revenue	GrossProfit	Margin
▶	Road Bikes	Bikes	11287183	4368347	0.387
	Mountain Bikes	Bikes	8583748	3930662	0.4579
	Touring Bikes	Bikes	3771565	1427160	0.3784
	Tires and Tubes	Accessories	380827	238396	0.626
	Helmets	Accessories	205834	131036	0.6366

1. Road Bikes and Mountain Bikes have the highest Revenue and Net Revenue generated
2. Even though Accessories have less sales, they have high margins

7. Is our customer base expanding, or are we relying on a narrowing group of repeat buyers?”

```
select year(sales.OrderDate) as Year, count(distinct(sales.CustomerKey)) a
s TotalCustomerCount
from sales
group by Year;
```

Result:

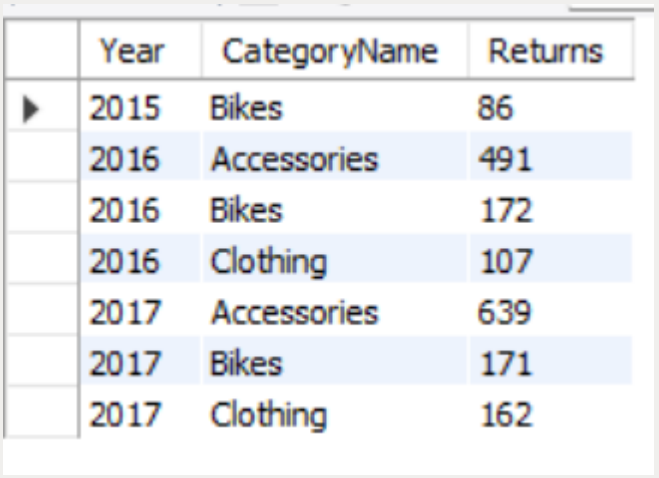
	Year	TotalCustomerCount
▶	2015	2630
	2016	9133
	2017	10502

1. The total customers of the company are increasing throughout the year

8. "What is the total quantity of products returned annually, broken down by product category?"

```
select year(returns.ReturnDate) as Year, product_categories.CategoryNam
e, sum(returns.ReturnQuantity) as Returns
from returns
left join products
on products.ProductKey = returns.ProductKey
left join product_subcategories
on products.ProductSubcategoryKey = product_subcategories.ProductSub
categoryKey
left join product_categories
on product_categories.ProductCategoryKey = product_subcategories.Prod
uctCategoryKey
group by Year, product_categories.CategoryName
order by Year;
```

Result:



	Year	CategoryName	Returns
▶	2015	Bikes	86
	2016	Accessories	491
	2016	Bikes	172
	2016	Clothing	107
	2017	Accessories	639
	2017	Bikes	171
	2017	Clothing	162

1. Only Bike products were returned in the first year and their return counts increased for next two year at a high level
2. The Accessory and Clothing products shows increasing return rate

Follow up:

```
select r.Year, r.CategoryName, (r.TotalReturns / s.TotalSales) * 100 as Retur
nToSalesRatio
from vw_annualcategoryreturns as r
left join vw_annualcategorysales as s
on r.Year = s.Year and r.CategoryName = s.CategoryName;
```

	Year	CategoryName	ReturnToSalesRatio
▶	2015	Bikes	3.2700
	2016	Bikes	3.0660
	2016	Accessories	1.9395
	2016	Clothing	2.0173
	2017	Accessories	1.9666
	2017	Bikes	3.0058
	2017	Clothing	2.2715

1. The avg ratio of return to sales of all products are very small, so there is no real impact of returns for every products
2. Although, Bikes have high return rate followed by clothing and Accessories

---

4. 2 Part B – Understanding Customer Asks

1. “Is there real demand for our products ?”

Statement :
To evaluate whether there is real and sustained customer demand, I will analyze **yearly sales trends from 2015 to 2017**. The analysis will focus on **order quantity and revenue** to determine whether sales are consistent over time and whether products are being purchased across different customer income segments, including lower-income customers.
Time: Yearly
KPIs:

1. Order Quantity
2. Total Revenue

Dimensions:

1. Customers
2. Time

Sub-Questions:

1. Does order quantity show consistent trends month over month?
2. Which customer segments contribute the highest order quantities overall?
3. How does order quantity change year over year across countries?
4. Are monthly order quantities growing across countries?

2. “Which parts of our business are healthy and which ones worry you?”

Statement:
To assess the business's health, I will analyze the Year-Over-Year trends of Net Revenue and Return Rate for each product. This comparison helps me identify "Healthy" products that are truly profitable and growing, while spotting "Worrying" ones where high returns are canceling out our sales performance.
Time: Yearly
KPIs:

1. Net Revenue
2. Order Quantity
3. Return Rate

Dimensions:

1. Product
2. Product category
3. Product subcategory
4. Time

Sub-Questions:

1. What is the Year-Over-Year trend for Net Revenue by product?
2. How are product Return Rates changing Year-Over-Year?
3. Is Order Quantity growing Year-Over-Year?

3. "Do we have stable growth in every country?”

Statement:
I will analyze monthly data to see if our growth is reliable or risky in each country. I want to identify countries where sales are "jumping" unpredictably (volatility) versus those growing steadily. I will also check if high sales volume is actually turning into real profit, or if we are losing that value to returns.
KPIs:

1. Revenue
2. Net Revenue
3. Order Quantity
4. No of customers

Dimensions:

1. Country
2. Time (Month-Year)

Sub-Questions:

1. **Financial Trend:** What is the trend in **Net Revenue vs. Total Revenue** for each country over the years? (Are we growing profitably, or just losing money to returns and marketing effort?)
2. **Volume Stability:** How does the **Sales Quantity** fluctuate in each country month-over-month? (Is demand consistent or random?)
3. **Market Expansion:** Are we **acquiring more customers** in each country over time? (Is our active user base growing, or are we stagnating?)

4. “Are there any regions where we are pushing hard but not seeing profitable growth?”

Statement:
To identify regions where increased sales effort does not result in profitable growth, I will analyze **monthly performance**. The analysis will focus on **sales volume, net revenue, and returns** to understand whether regions with high order quantities are also generating strong net revenue, or whether returns are reducing the overall value of sales.
Time: Yearly
KPIs:

1. Order Quantity
2. Net Revenue
3. Return Rate

Dimensions:

1. Region
2. Time (Month)

Sub-Questions:

1. **Volume Trend:** What is the trend in sales volume for each region over the years? (Are we successfully increasing our market presence?)
2. **Revenue Growth:** What is the trend in net revenue for each region over the years? (Is the increased volume translating into actual profit?)
3. **Value Efficiency:** What is the return-to-sales ratio in each region over the years? (How much monetary value are we losing to returns relative to our sales?)
4. **Quality Check:** What is the trend in return rates for each region over the years? (Is customer dissatisfaction rising in specific regions?)

the products sold in different regions may vary

5. “Is growth coming from more customers or higher purchase quantities per customer?”

Statement:
I will analyze whether our business growth is driven by finding new customers or by selling more to the ones we already have. By comparing the trend in the **Number of Unique Customers** against the trends in **Order Quantity** and **Net Revenue**, I can determine if we are expanding our market reach or simply increasing the size of each order.
KPIs:

1. Distinct Customer count
2. Order Quantity
3. Net Revenue

Dimensions:

1. Time (year or month)

Sub Questions:

1. **Market Reach:** What is the trend in the total number of unique customers year-over-year? (Are we effectively acquiring new users to expand our base?)
2. **Purchase Intensity:** What is the trend in the average quantity purchased per customer year-over-year? (Are customers adding more items to their carts over time?)
3. **Transaction Value:** What is the trend in **Net AOV** over the years? (Is the actual realized value of each individual order increasing, or are we processing more orders for less profit?)

6. "Which customer income segment contributes the most to our growth?”

Statement:
I will group our customers into three income levels: **Low**, **Medium**, and **High**. By tracking how much **Net Revenue** each group generates year-over-year, I will see if our business relies more on everyday buyers or wealthy customers.
Segments:(Example : This assessment dataset choosen by me)

- **Low Income:** Up to $50,000
- **Medium Income:** $60,000 – $100,000
- **High Income:** Over $100,000

KPIs:

1. Net Revenue
2. Order Quantity

Dimensions:

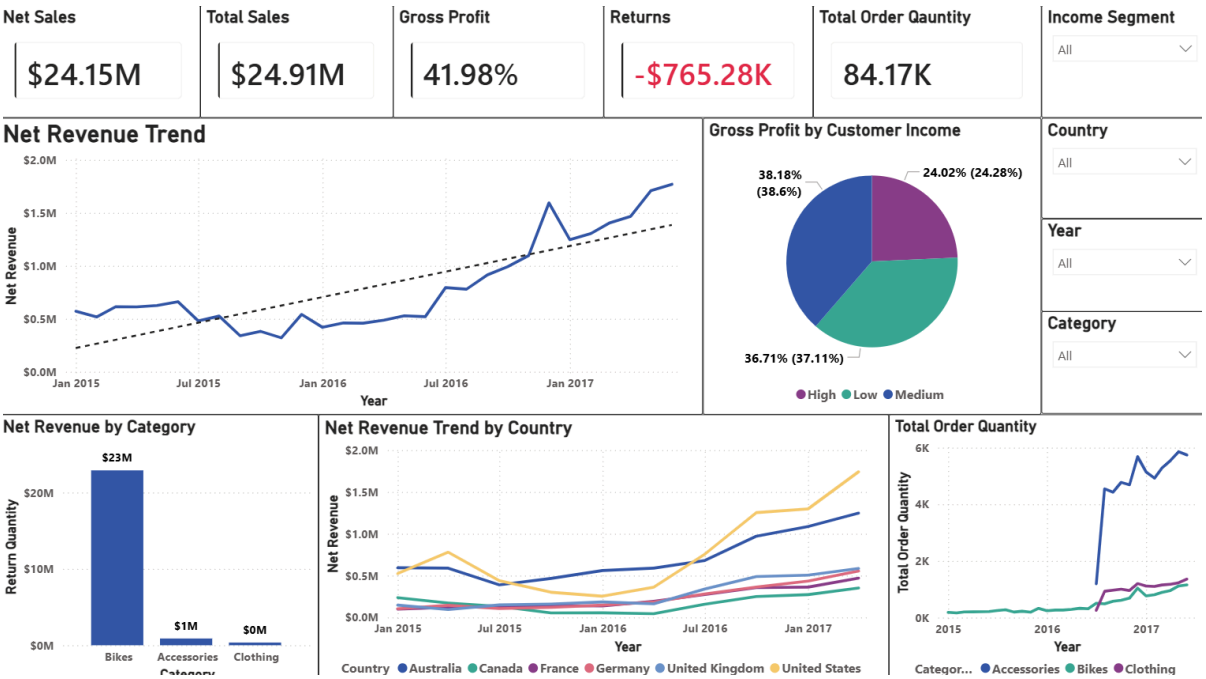
1. **Income Segment**
2. **Time (Year)**

Sub Questions:

1. **How much** Net Revenue did each income group contribute each year?
2. **How many** items did each group buy each year?
3. **Which** segment spends the most money per person on average?

4.3 Part C - Dashboard & Storytelling

1. Dashboard Design



2. Story in Plain English

Executive Summary: High-Margin Growth & Healthy Operations

The business is in a strong financial position with **$24.15M in Net Sales** and a healthy **Gross Profit margin of 41.98%**. Operationally, we are highly efficient: our **overall return-to-sales ratio is approximately 3%** ($765K returns on $24.9M total sales). This is significantly lower than typical e-commerce benchmarks, indicating strong product quality and customer satisfaction. The data reveals a clear "split story": while our operational volume is dominated by Accessories, our actual revenue—and 96% of our return value—is driven by Bikes. Additionally, our growth is being powered by the American market and the Low-to-Medium income consumer segments.

1. The "Volume vs. Value" Split:
We have a massive divergence between what we sell the *most of* versus what *makes us money*.

- **Volume Driver:** The "Total Order Quantity" chart shows a massive explosion in **Accessories** sales starting in mid-2016 (the blue line spike). We are shipping thousands of accessory units.
- **Revenue Driver:** Despite that volume, **Bikes** are the undisputed revenue engine, generating **$23M** (approx. 95% of total sales), compared to just **$1M** from Accessories.

- **Takeaway:** We are effectively a Bike company that uses Accessories as a high-frequency customer touchpoint.

2. The 2016 "US" Inflection Point:
Mid-2016 marked a major turning point for the business.

- **Trend:** Revenue shifted from a flat trend ($0.5M monthly) to an aggressive growth curve, nearly tripling to $1.5M+ monthly by 2017.
- **Driver:** This growth is disproportionately fueled by the **United States** (yellow line), which separated from the pack and skyrocketed. Other regions like Australia and the UK are growing, but at a much more modest pace.

3. Revenue Volume is Dominated by the "Mass Market":
Contrary to the expectation that high-ticket products (Bikes) rely on wealthy buyers, our top-line revenue is overwhelmingly fueled by the "everyman" consumer.

- **The "Everyman" Core:** Low and Medium Income customers combined generate ~87% of our Total Revenue (38.97% Low + 48.27% Medium).
- **The High-End Niche:** The "High" income segment accounts for just 12.76% ($3.18M) of sales, making it our smallest source of turnover.
- **Takeaway:** Our cash flow is primarily sustained by value-conscious and mid-tier buyers, rather than the luxury market.

Strategic Recommendations for the CXO

1. **Investigate & Stabilize Volatility:** While the overall trend is positive, the data shows sharp drops in revenue and volume (specifically in mid-2015 and late 2016). We must perform a root-cause analysis on these periods—were they caused by **stockouts, seasonality, or operational failures**? Identifying the cause is the only way to prevent similar revenue leakage in the future.
2. **Activate "Sleeping" Markets (Non-US/UK):** While the US and UK are driving our surge, other major markets (Australia, Canada, France, Germany) are seeing a much slower pace of growth. We need to execute targeted sales activities and localized marketing in these regions to replicate the high-growth trajectory currently seen in the US.

3. **Protect the "Bike" Supply Chain:** Since $23M of our $24M revenue sits in one category, any supply chain disruption in "Bikes" is a critical risk. We need to ensure our inventory levels match the US demand surge.
4. **Monetize the "Accessory Boom" (Smartly):**
   Accessories have huge volume but generate very little revenue ($1M vs. $23M).

- **The Trap:** Do **not** simply raise prices. Our data shows our customers are "Low & Medium Income," so a price hike could scare them away.
- **The Fix:** Use **Bundling**. Create "Starter Kits" (e.g., Bike + Helmet + Lock) to force an upsell at the point of purchase. This increases the total cash value of every order without looking like a price hike to our budget-conscious buyers.

Questions I Would Ask You (Feedback Orientation)

To refine this dashboard into a sharper decision-making tool, I would ask:

1. **Data Hygiene (The "Components" Category):** I noticed the "Components" category appears in the filters but contains zero sales data. **Should I permanently hide this category** to declutter the dashboard, or is it a placeholder for an upcoming product line that should remain visible?
2. Currently, the dashboard shows "Returns" as a large red dollar amount (-$765K). **Would you prefer to see this as a "Return Rate %" trend line?** This would help you distinguish between a natural rise in returns due to higher sales volume versus an actual drop in product quality.
3. Currently, the 'Total Order Quantity' is a line chart to show the 2016 spike in Accessories. However, this compresses the view for Bikes and Clothing. **Would you prefer a Bar Chart here?** This would clearly rank the total volume of each category, though we would lose the visibility of the month-over-month growth trend.
4. **Granularity (Subcategories):** Currently, the dashboard allows filtering by broad Categories (e.g., Bikes, Accessories), but I have not included a specific Subcategory filter. **Do you require the ability to drill down into specific product types (e.g., Mountain Bikes vs. Road Bikes), or is the high-level Category view sufficient for your strategic needs?**
5. Would you like a Total Revenue vs Net Revenue line chart to understand how effectively sales are converting into net revenue, especially

considering product returns?