Assignment 1 : Descriptive Analytics

Presented to
Instructor Vivek Saahil
TA Vaibhav Vishal

By
Delisle, Audrey - 261142504

INSY 695 Advanced Visualization using Power BI - Section 077

McGill University - Desautels Faculty of Management
Wednesday January 24th 2023

Questions 1: What is the total sales revenue generated by each product category in 2017, and how does it compare to the previous years i.e. 2016 and 2015? FAQ Note: • Returns are NOT to be accounted in the revenue calculation

Numerical Answer Q1: (View Appendix B)

	Accessories	Bikes	Clothing
2017	0.5M	8.5M	0.2M
2016	0.4M	8.8M	0.2M
2015	0M	6.4M	0M

i) Steps to solution:

Datasets Used:

- 'Sales-2015', 'Sales-2016', 'Sales-2017': Contain sales data including OrderDate, ProductKey, and OrderQuantity.
- 'Products': Includes details about products, particularly ProductPrice.
- 'Product Categories': Contains product categories.

Power Query Steps:

- Merge Sales Data: Merged 'Sales-2015', 'Sales-2016', and 'Sales-2017' into a single table for a consolidated view. I did this in get data, I clicked file, and combined all 3 sales csvs into 1 table.
- Merge with Products: Merged the consolidated sales table with the 'Products' table to link each sale with its product price.

Relationship Mapping: (View Appendix A)

1) Established relationships between the merged sales table and the 'Product Categories' table using ProductKey.
2) Product Categories linked to Product Sub-categories through product category key
3) Product Sub-Categories linked to Products through product subcategorykey

Calculation Steps:

- I created a calculated column for 'Total Revenue' in the merged sales table using the formula: Total Revenue = 'Sales'[OrderQuantity] * RELATED(Products[ProductPrice])

Final Visualization:

I used a bar chart to display the total sales revenue for each product category across the years 2015, 2016, and 2017. I placed Category Name in X-axis, Sum of Total Revenue in Y-Axis, and OrderDate (only year) in Legend.

ii) Chart Selection: Stacked Bar Chart

Bar charts are ideal for comparing quantitative data across different categories. In this case, it allows for an effective comparison of total sales revenue across different product categories and years. The bar length visually represents the revenue amount, making it easy to compare and contrast between

categories and over time (stacked). Also, since we were not drilling-down in this example, a bar chart was perfect because it is simple.

iii) Business Applications:

Trend Analysis: The visualization helps stakeholders identify trends in sales revenue across different product categories over multiple years. This insight is important for understanding which categories are performing well and which are not. For example, we can quickly tell that bikes is the best performing category, hence maybe they should invest more advertising budget into the other categories.

Resource Allocation: By knowing which product categories generate the most revenue, the business can allocate resources more effectively, whether in marketing, stock management, or new product development. Once again, this relates to my example above.

Strategic Planning: The data can inform long-term business strategies. For instance, if certain categories show consistent growth, they might be worth investing in more heavily.

Market Analysis: Understanding how different product categories perform over time can also offer insights into market changes and customer preferences, guiding more informed decisions about product offerings. I see in the visuals that in 2015, the only category sold was bikes, hence we can assume that the other 2 categories were a response to the market changing.

Question 2: Which territories, product categories and product subcategories have the highest return rates? FAQ Note: • The business stakeholders would like the solution to be granular and have the ability to drilldown by Territory → Product Category → Product Sub-Category

Numerical Answer Q2: (View Appendix C)

	Highest Return Rate
Territory	France (2.3658%)
Product Category	Bikes (3.0799%)
Product SubCategory	Shorts (4.2372%)

i) Steps to solution:

Datasets Used:

- 'Sales': Contains details of sales transactions, including TerritoryKey, ProductKey, and OrderQuantity.
- 'Returns': Includes data on returned items, particularly ReturnQuantity.
- 'Territories', 'Product Categories', 'Product Subcategories': Provide context and categorization for sales and returns.

Power Query Steps:

- I did not do any significant transformations in Power Query apart from ensuring data cleanliness and correct formats.
-

Relationship Mapping: (View Appendix A)

- Relationships between 'Sales' and 'Territories' through SalesTerritoryKey
- Relationships between Sales to Products through ProductKey
- Relationships between 'Product Categories', and 'Product Subcategories' through ProductCategoryKey
- Relationships between ProductSubcategories to Products through ProductSubcategoryKey
- Relationships between 'Returns' and Products through ProductKey
- Relationships between Returns and Territories through SalesTerritoryKey

Calculation Steps:

- Total Units Sold Measure: I created a measure to calculate the total units sold in each category, subcategory, and territory. Total Units Sold = SUM(Sales[OrderQuantity])
- Total Units Returned Measure: I created a measure to sum up the total units returned, matching the above categories. Total Units Returned = SUM(Returns[ReturnQuantity])
- Return Rate Measure: I created a measure to calculate the return rate as Total Units Returned / Total Units Sold * 100. Return Rate = DIVIDE([Total Units Returned], [Total Units Sold], 0) * 100
- Note: I used measures in this case instead of new columns because they can aggregate data across different tables without needing a row-by-row correspondence, which is a limitation for calculated columns.

Final Visualization:

I chose a pie chart to represent the return rates across different territories, product categories, and subcategories. I placed Region, CategoryName and SubcategoryName in the legend (in that specific order to be able to drill-down), and I placed my return rate in values. There is the drill-down possibility on this visual.

ii) Chart Selection: Pie Chart

Pie charts are excellent for showing the proportion of parts to a whole. All regions are part of 1 whole (the geographical distribution of the company), all product categories are part of 1 whole product line, and all product sub-categories are part of 1 whole product-subcategories line. In this scenario, it helps in understanding how each territory, product category, and subcategory contributes to the overall return rates. This visualization makes it easier to identify which segments have disproportionately high return rates compared to others.

iii) Business Applications:

Identify Problem Areas: High return rates can indicate issues with product quality, customer satisfaction, or other operational challenges. This analysis highlights regions that need attention.

Resource Optimization: Understanding which areas have higher return rates allows the company to allocate resources effectively, whether for quality checks, customer service improvements, or inventory adjustments.

Product and Service Improvement: Insights from return rates can drive changes in product development, procurement, or vendor selection, especially in categories or territories with high returns.

For example, in the visual, I dont see much variation between regions, but I see a big increase in bike returns compared to other products. Hence, perhaps that gives the idea that the bikes need to be improved.

Customer Satisfaction: Targeted strategies can be developed to improve customer satisfaction in regions or product lines with high return rates, which is crucial for long-term customer loyalty and brand reputation. For example, in a region with a low return-rate, there is more opportunity for a reward system, as customers are more loyal.

Question 3: What is the customer demographics breakdown by gender, age and education level in 2017 and how does it compare to the previous years i.e. 2016 and 2015? FAQ Note: • The business stakeholders would like the solution to be granular and have the ability to drilldown by Gender → Age → Education Level

Numerical Answer Q3: (View Appendix D)

		2017	2016	2015
Gender	F	5.1K	4.5K	1.3K
	M	5.3K	4.6K	1.3K
	NA	0.1K	0.1K	0K
Age Group	40-50	1.2K	1.1K	0.3K
	50-65	5.2K	4.6K	1.4K
	65+	4.1K	3.4K	1.0K
Education Level	Partial HS	860	774	169
	HS	1872	1507	458
	Partial College	2900	2400	729
	Bachelors	3105	2815	817
	Graduate	1765	1637	457

i) Steps to solution:

Datasets Used:

- 'Customers': Contains customer demographics including BirthDate, Gender, and Education Level.
- 'Sales': Sales data across various years with CustomerKey to link to the 'Customers' table.

Power Query Steps:

- Date Conversion: In the 'Customers' table, transform the 'BirthDate' column to the Date format using the locale setting for the Date UK.

Relationship Mapping: (View Appendix A)

- Between 'Customers' and 'Sales': Established a relationship based on 'CustomerKey'.

Calculation Steps:

- Age Bucket Column: I created an 'Age Group' column using a DAX formula that calculates the current age and categorizes it into age buckets. I decided to make 6 buckets, making jumps of about 10 years each time. (Age Group = VAR CurrentYear = YEAR(TODAY()) VAR BirthYear = YEAR([BirthDate]) VAR Age = CurrentYear - BirthYear RETURN SWITCH(TRUE(),
  Age < 18, "Under 18",
  Age >= 18 && Age < 24, "18-24",
  Age >= 24 && Age < 30, "24-30",
  Age >= 30 && Age < 40, "30-40",
  Age >= 40 && Age < 50, "40-50",
  Age >= 50 && Age < 65, "50-65",
  Age >= 65, "65+",
  "Unknown")
- Order Year Column: I added a column in the Sales table called OrderYear, for to simplify the next step
- Yearly Purchase Columns: I created columns in the Customers table called 'Purchased in 2017', 'Purchased in 2016', and 'Purchased in 2015' to indicate if customers made purchases in these years. (Purchased in 2017 = IF(CALCULATE(COUNTROWS('Sales'),
  'Sales'[CustomerKey] = EARLIER('Customers'[CustomerKey]),
  'Sales'[Order Year] = 2017
) > 0, 1, 0)
- Yearly Measures: I developed measures for each year (2015, 2016, 2017) to sum up the 'Purchased in [Year]' columns.

Final Visualization:

I used a bar chart to display the breakdown of customer demographics by gender, age group, and education level. I put gender, age group and education level in the x-axis, and my 3 created measures (2015,2016 and 2017) in the y-axis. There is the drill-down possibility on this visual.

ii) Chart Selection: Stacked Bar Chart

I chose a stacked bar charts because they are highly effective for comparing categorical data. In this analysis, the bar chart shows a clear comparison of customer demographics across gender, age groups, and education levels for different years. This format is clear to understand if you want to observe trends and differences in demographics over the years. The stacked option was selected because I wanted the data seen year over year.

iii) Business Applications:

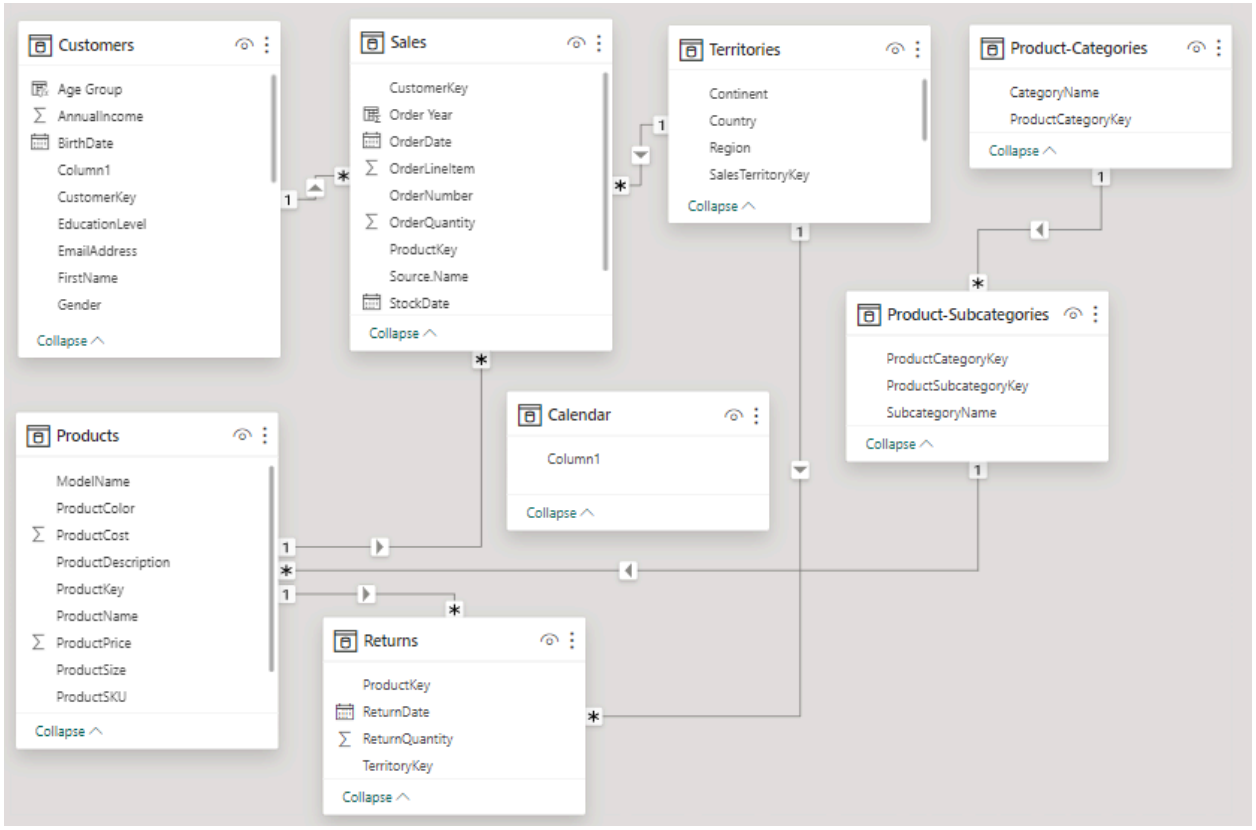
Marketing and Targeting: Understanding the demographic breakdown helps in tailoring marketing strategies and targeting specific customer segments more effectively. For example, we can see that the age group 50-65 is where most of the customer lie. Hence, when marketing, the team might decide to invest more money into the radio and newspapers instead of social media.

Trend Analysis: By comparing demographic data over several years, businesses can identify trends in their customer base, such as shifting age demographics or changes in education levels.

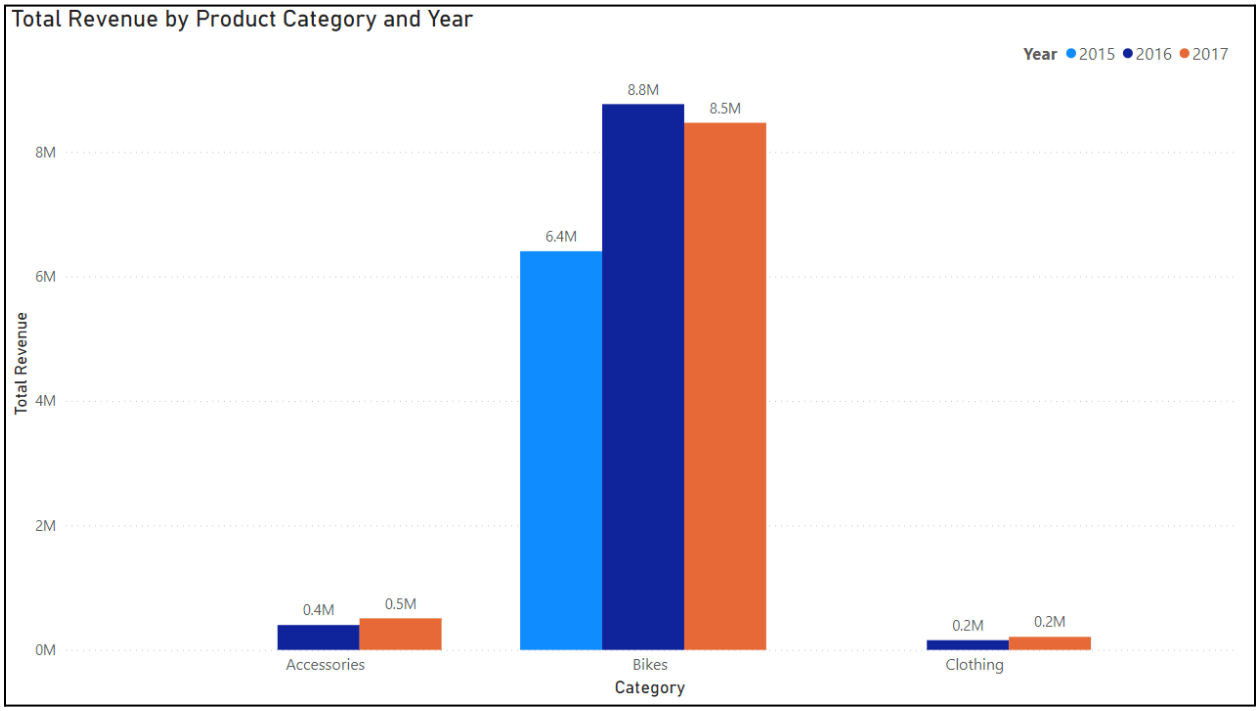
Product Development and Sales Strategy: Insights into the demographic composition can guide product development and sales strategies to cater to the predominant customer groups. For example, they might create a bike with more safety feature knowing their demographic is older and values safety.

Resource Allocation: The demographic breakdown can inform where to allocate resources for customer engagement, support, and outreach initiatives.

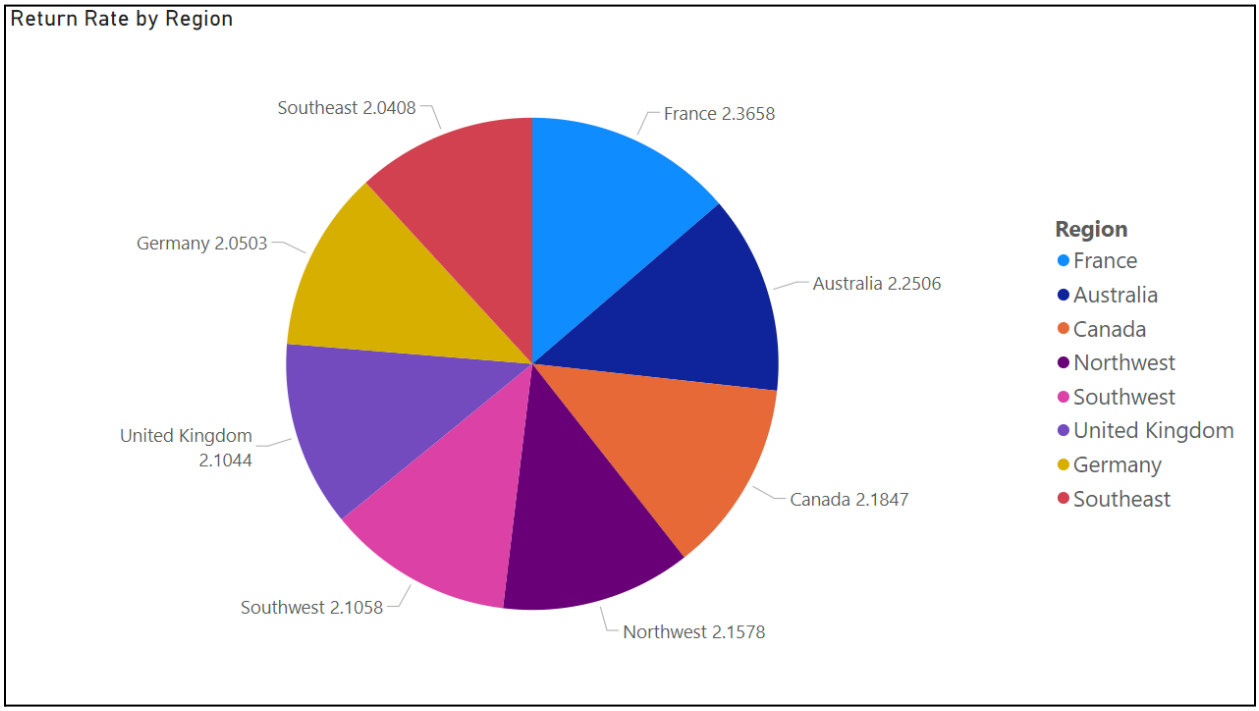
Appendix A : Relationship Map



Appendix B: Question 1 Visual Screenshot



Appendix C: Question 2 Visual Screenshot



Appendix D: Question 3 Visual Screenshot