Question 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)?

(i) The datasets used for this question were Sales-2015, Sales-2016, Sales-2017, Calendar, Products, Product-Categories and Product-Subcategories. First, a new table called Sales was created which appended Sales-2015, Sales-2016 and Sales-2017. For the new Sales table, a 'Sales Year' column was added which extracted the year from the order date column. For the Calendar table, the first row was used as a header and then the actual dates needed to be converted to the proper format (i.e. yyyy-mm-dd). The following relationships were created: Product-Categories and Product-Subcategories (using product category key), Product-Subcategories and Products (using product subcategory key), and Sales and Products (using product key). For calculation steps, in the Sales tables, a calculation was done to find the revenue for each row. This was done by multiplying the Order Quantity from the Sales table with the Product Profit from the Product table. Finally, the visualization (a bar graph) was created. The format for this bar graph was to graph the total sales revenue for each year and separating it by product category. The total sales revenue in 2017 generated was $8,468,854.53 for Bikes, $507,330.99 for Accessories, and $209,263.93 for Clothing. Comparing this to the sales revenue generated in 2016, a higher amount was generated for Bikes at $8,768,706.98, a lower amount was generated for Accessories at $399,342.12, and a lower amount was generated for Clothing at $156,154.69. Comparing to the sales revenue generated in 2015, the only category that generated revenue in 2015 was Bikes which generated less sales revenue than in 2017 at $6,404,933.58. Also, Components had no sales in any of the years.

(ii) The reason for choosing a bar graph was because it is a suitable choice for comparing values across different product categories and showing trends over time (specifically years). We only have 2 metrics that we want to use (product categories and years) to show total sales revenue.

(iii) The bar graph visualizing total sales revenue by product category over three years enables stakeholders to understand trends and performance of product segments. This can guide strategic decisions such as allocating resources to high-performing categories (for example, bikes), adjusting marketing strategies, or identifying potential areas for product development (for example, no sales of category "components"). Understanding year-over-year changes in sales revenue helps in forecasting future trends and making informed pricing and stocking decisions.

Question 2: Which territories, product categories and product subcategories have the highest return rates?

(i) The datasets used for this question were Sales (as described above), Product-Categories, Product-Subcategories, Returns and Territories. In addition to the relationships that were created for question 1, some additional relationships needed to be created here to account for the new tables. These included Returns and Products (product key), Territories and Returns (territory key for Returns and sales territory key for Territories) and Territories and Sales (territory key). Then I created a measure for return rate. This was done by dividing the sum of the total return quantity from the

Returns table by the total order quantity from the Sales table. The final visualization to answer this question is a matrix. In the matrix created, the rows chosen are region, category name, subcategory name (in that order), and the values were chosen to be the return rate. By implementing these columns and numbers into the matrix, a visualization was created which allows the user to view the return rates that came from each of the categories listed above. In this visual there is a "go to next level in the hierarchy" option which allows the user to drill-down for each category. By doing this, it was determined that France was the country with the highest return rates, Bikes was the category with the highest return rates, and Shorts was the subcategory with the highest return rates. Each row in the matrix provides the return rate for each specific region, category and subcategory.

(ii) The reason for choosing a matrix to visualize this question is because it provides a granular view of return rates, allowing stakeholders to drill down from Territories to Product Categories to Product Subcategories. This format helps in presenting detailed information and supports an in-depth analysis of return rates across different dimensions.

(iii) By identifying the territories and product categories/subcategories with the highest return rates, stakeholders can investigate issues in product quality and customer satisfaction. This data enables targeted interventions to reduce returns, such as enhancing product quality controls, providing better customer education on product use, or tailoring the product mix to suit regional preferences.

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?

(i) The datasets used for this question were Calendar, Customers and Sales (as described above). The additional relationship created for this question was Customers to Sales (customer key). The calculation steps included calculating the customers age at the time the order was placed. This was done by creating a new column in the Customers table which took the Birth Date from the Customers table and subtracting today's date (January 23, 2024). After that, another column was created to group the ages (for example, less than 18, 18-24, 25-34, etc.). Then I created binary columns in the Customers table indicating whether a customer made a purchase in each year from 2015 to 2017. I then created a new column in the Sales table to identify the order year for each transaction. Then I created three distinct measures to tally the number of unique customers per year, which I then put into a matrix to provide a yearly breakdown of customer demographics by gender, age, and education level. As a brief overview, 2017 had the greatest number of customers and in terms of gender, most were male, in terms of age group, most were aged 65+, and in terms of educational level, most had a Bachelors' degree. It seems as though in 2015 and 2016, the most frequent age group and education levels were the same but there were more female than male customers in 2015.

(ii) A matrix was chosen to visualize this question because it provides a multi-dimensional analysis of customer demographics across years. The hierarchical nature of the matrix supports the required drill-down capabilities.

(iii) By analyzing the customer demographics breakdown over three years, business stakeholders can pinpoint shifts in their market composition and adjust their customer engagement strategies accordingly. Insights from the data can inform targeted marketing campaigns and product development to better cater to the predominant demographic segments. Additionally, this granular view allows for a refined approach to inventory planning and promotional activities, aligned with their specific customer base.