

INTRO

This document serves as a complement to the *AdventureWorks_sql_project_queries.sql* file which includes all the code for queries written on the data. The goal is to provide screenshots of query output as a way of showing that the code works, and adding brief comments on the output itself where I see fit. If anyone is reading this, thank you for taking the time!

A note on the data:

This Data is from the Microsoft AdventureWorks database and contains data on a bike store with the following tables:

- Calendar
- Customers
- Products
- Product Categories
- Product Subcategories
- Sales (From 2020, 2021 and 2022)
- Returns Data (2020, 2021, 2022)
- Territories

This project focuses on the answering of questions that require more complex queries. Though some may be more general, my analyses are very adaptable to the hypothetical client's needs and to more specific questions. Conclusions are also limited to what can be inferred from the data

Questions, Output and Comments:

1. Top 10 Products by Total Profit

| | model_name character varying (50) 🔒 | total_profit text 🔒 | profit_ranking bigint 🔒 |
|----|--|------------------------|----------------------------|
| 1 | Mountain-200 | 3,301,095.68 | 1 |
| 2 | Road-250 | 1,640,671.19 | 2 |
| 3 | Road-150 | 1,142,464.35 | 3 |
| 4 | Touring-1000 | 1,112,328.88 | 4 |
| 5 | Road-350-W | 563,435.28 | 5 |
| 6 | Road-550-W | 536,122.51 | 6 |
| 7 | Mountain-100 | 324,534.14 | 7 |
| 8 | Road-750 | 279,196.05 | 8 |
| 9 | Road-650 | 206,457.27 | 9 |
| 10 | Mountain-400-W | 186,396.28 | 10 |

2. Return Rate by Product Subcategory

| | subcategory_name character varying (50) 🔒 | return_rate text 🔒 |
|----|--|-----------------------|
| 1 | Shorts | 4.237% |
| 2 | Vests | 3.647% |
| 3 | Hydration Packs | 3.597% |
| 4 | Bike Stands | 3.419% |
| 5 | Touring Bikes | 3.296% |
| 6 | Road Bikes | 3.141% |
| 7 | Helmets | 3.116% |
| 8 | Jerseys | 2.987% |
| 9 | Mountain Bikes | 2.890% |
| 10 | Bike Racks | 2.649% |
| 11 | Socks | 2.070% |
| 12 | Bottles and Cages | 1.907% |
| 13 | Gloves | 1.853% |
| 14 | Tires and Tubes | 1.794% |
| 15 | Cleaners | 1.465% |
| 16 | Fenders | 1.364% |
| 17 | Caps | 1.108% |

Shorts are returned most often as people most likely order the wrong size and replace them. None of these return rates seem to be alarmingly high, though in a real-life scenario I would need to check with someone at the company and see what a threshold level could be. From a google search, bike store return rates vary from 8% to 10% so in comparison to that, AdventureWorks seems to be doing quite well.

3. For each Country and Region, compute Total Sales, Number of (unique) customers and Average Order Quantity (Output for the follow-up query which includes output for the original query):

| | country character varying (50) 🔒 | region character varying (50) 🔒 | total_sales text 🔒 | unique_customers text 🔒 | avg_order_quantity text 🔒 | sales_per_customer text 🔒 |
|----|-------------------------------------|------------------------------------|-----------------------|----------------------------|------------------------------|------------------------------|
| 1 | Australia | Australia | 7,416,456.20 | 3,480 | 1.45 | 2,131.17 |
| 2 | Canada | Canada | 1,769,245.81 | 1,499 | 1.58 | 1,180.28 |
| 3 | France | France | 2,362,643.32 | 1,705 | 1.50 | 1,385.71 |
| 4 | Germany | Germany | 2,524,679.97 | 1,675 | 1.50 | 1,507.27 |
| 5 | United Kingdom | United Kingdom | 2,902,562.09 | 1,822 | 1.51 | 1,593.06 |
| 6 | United States | Central | 3,143.06 | 8 | 1.50 | 392.88 |
| 7 | United States | Northeast | 6,401.57 | 8 | 1.48 | 800.20 |
| 8 | United States | Northwest | 3,095,074.47 | 3,075 | 1.51 | 1,006.53 |
| 9 | United States | Southeast | 11,585.62 | 10 | 1.44 | 1,158.56 |
| 10 | United States | Southwest | 4,822,794.70 | 4,134 | 1.50 | 1,166.62 |

An average order quantity lingering around 1.5 indicates that clients' purchases likely include either just a bike, or a bike accompanied by an attachment or other accessory. **Australia** being at the top of sales results from having the second most customers and the **highest sales per customer at \$2,131.17**.

Sales_per_customer could potentially be explained by factors such as the average income of customers in those countries and regions as well as their spending habits (segmentation is done in a later question).

4. Monthly Sales Trends with Year-Over-Year Growth:

| | year integer | month_name character varying (15) | month_sales text | prior_year integer | prior_year_month_sales text | yoy_pct_change text |
|----|-----------------|--------------------------------------|---------------------|-----------------------|--------------------------------|------------------------|
| 1 | 2022 | June | 1,826,987.14 | 2021 | 533,824.98 | 242.24% |
| 2 | 2022 | May | 1,768,432.51 | 2021 | 545,534.74 | 224.16% |
| 3 | 2022 | April | 1,527,813.72 | 2021 | 494,957.42 | 208.68% |
| 4 | 2022 | March | 1,448,596.12 | 2021 | 471,961.88 | 206.93% |
| 5 | 2022 | February | 1,339,241.29 | 2021 | 474,162.79 | 182.44% |
| 6 | 2022 | January | 1,274,378.67 | 2021 | 432,425.74 | 194.70% |
| 7 | 2021 | December | 1,635,308.80 | 2020 | 563,761.53 | 190.07% |
| 8 | 2021 | November | 1,133,913.05 | 2020 | 326,611.15 | 247.18% |
| 9 | 2021 | October | 1,029,821.05 | 2020 | 404,276.60 | 154.73% |
| 10 | 2021 | September | 952,743.49 | 2020 | 344,062.87 | 176.91% |
| 11 | 2021 | August | 804,193.39 | 2020 | 536,452.82 | 49.91% |
| 12 | 2021 | July | 815,356.47 | 2020 | 486,115.01 | 67.73% |
| 13 | 2021 | June | 533,824.98 | 2020 | 669,988.67 | -20.32% |
| 14 | 2021 | May | 545,534.74 | 2020 | 659,325.90 | -17.26% |
| 15 | 2021 | April | 494,957.42 | 2020 | 653,364.04 | -24.24% |
| 16 | 2021 | March | 471,961.88 | 2020 | 643,436.10 | -26.65% |
| 17 | 2021 | February | 474,162.79 | 2020 | 532,226.25 | -10.91% |
| 18 | 2021 | January | 432,425.74 | 2020 | 585,312.65 | -26.12% |

I don't have enough data or information to be able to explain the drastic fall in sales from 2020 to 2021 and subsequent rise going into 2022, but COVID-19 is most likely the culprit.

5. Customer Lifetime Value Segmentation (Top 10% as top-tier, Next 40% as mid-tier, bottom 50% as lower-tier)

| | customer_name character varying | customer_lifetime_value text | ltv_ranking bigint | customer_tier text | total_ltv_pct_contribution text |
|----|------------------------------------|---------------------------------|-----------------------|-----------------------|------------------------------------|
| 1 | Jordan Turner | 15,570.81 | 1 | Top Tier | 0.06% |
| 2 | Maurice Shan | 12,407.95 | 2 | Top Tier | 0.05% |
| 3 | Janet Munoz | 12,015.40 | 3 | Top Tier | 0.05% |
| 4 | Lisa Cai | 11,330.45 | 4 | Top Tier | 0.05% |
| 5 | Lacey Zheng | 11,085.75 | 5 | Top Tier | 0.04% |
| 6 | Franklin Xu | 10,863.44 | 6 | Top Tier | 0.04% |
| 7 | Larry Munoz | 10,852.03 | 7 | Top Tier | 0.04% |
| 8 | Kate Anand | 10,436.51 | 8 | Top Tier | 0.04% |
| 9 | Larry Vazquez | 10,394.98 | 9 | Top Tier | 0.04% |
| 10 | Ariana Gray | 10,391.43 | 10 | Top Tier | 0.04% |
| 11 | Clarence Gao | 10,331.73 | 11 | Top Tier | 0.04% |
| 12 | Aaron Wright | 10,329.23 | 12 | Top Tier | 0.04% |
| 13 | Ethan Bryant | 10,308.52 | 13 | Top Tier | 0.04% |

Output to follow-up query (per-tier contribution to revenue, number of customers in each tier)

| | customer_tier text | total_ltv text | pct_of_revenue text | customers_in_tier text | ltv_per_customer text |
|---|-----------------------|-------------------|------------------------|---------------------------|--------------------------|
| 1 | Mid-Tier | 14,068,959.55 | 56.47% | 6,938.00 | 2,027.81 |
| 2 | Top Tier | 10,057,524.24 | 40.37% | 1,735.00 | 5,796.84 |
| 3 | Bottom-Tier | 788,101.27 | 3.16% | 8,670.00 | 90.90 |

Mid-Tier Customers are the highest contributors to company revenue. This statistic is subject to change based on how the tiers are determined, as if the top-tier were to consist of the customers in the highest 20% in terms of lifetime value, it would surely become the top revenue-contributing group. The top 10% of customers with regards to lifetime value do have a significantly higher lifetime value per customer compared to the mid and bottom tiers (again, subject to change based on segmentation approach), and the retention of these customers is paramount to the company's success.

6. Top 5 Products with Longest Average Lead Time

| | model_name character varying (50) | avg_lead_time_days numeric | diff_from_avg_in_days numeric | lead_time_rank bigint |
|---|--------------------------------------|-------------------------------|----------------------------------|--------------------------|
| 1 | All-Purpose Bike Stand | 70.95 | 3.60 | 1 |
| 2 | Touring-3000 | 69.99 | 2.64 | 2 |
| 3 | Bike Wash | 68.69 | 1.34 | 3 |
| 4 | ML Mountain Tire | 68.69 | 1.34 | 3 |
| 5 | Mountain Tire Tube | 68.68 | 1.33 | 4 |
| 6 | Touring-2000 | 68.34 | 0.99 | 5 |

It is worth noting that the fastest lead time is about 65 days, so not a huge difference compared to the slowest lead times. If there were to be more variation in lead times (i.e. product turnover), the management of inventory would vary to a higher degree by product. If a product had a significantly faster turnover rate, re-stocks would need to happen more frequently and vice-versa.

7. Contribution of Top Products to Total Revenue (Pareto Analysis) In other words, seeing if top 20% of products contribute to ~80% of revenue:

| | top_20_pct_revenue text | total_revenue text | top_20_pct_contribution text |
|---|----------------------------|-----------------------|---------------------------------|
| 1 | 21,634,561.40 | 24,914,586.82 | 86.83% |

The top 20% of products contributing to 86.83% of revenue confirms the pareto principle. For reference, here are the top 20% themselves along with their subcategories:

| | model_name character varying (50) | subcategory_name character varying (50) | revenue text | pctile_group integer |
|---|--------------------------------------|--|-----------------|-------------------------|
| 1 | Mountain-200 | Mountain Bikes | 7,170,946.50 | 1 |
| 2 | Road-250 | Road Bikes | 4,200,816.75 | 1 |
| 3 | Touring-1000 | Touring Bikes | 2,939,558.31 | 1 |
| 4 | Road-150 | Road Bikes | 2,905,555.24 | 1 |
| 5 | Road-350-W | Road Bikes | 1,549,601.89 | 1 |
| 6 | Road-550-W | Road Bikes | 1,358,594.13 | 1 |
| 7 | Road-750 | Road Bikes | 767,865.78 | 1 |
| 8 | Mountain-100 | Mountain Bikes | 741,622.81 | 1 |

8. Sales Loss from Returns by Territory

| | country character varying (50) 🔒 | region character varying (50) 🔒 | return_losses numeric 🔒 |
|---|-------------------------------------|------------------------------------|----------------------------|
| 1 | Australia | Australia | 240180.46 |
| 2 | United States | Southwest | 133684.93 |
| 3 | United States | Northwest | 95059.68 |
| 4 | Germany | Germany | 86615.51 |
| 5 | United Kingdom | United Kingdom | 84167.38 |
| 6 | France | France | 81035.01 |
| 7 | Canada | Canada | 44502.27 |
| 8 | United States | Southeast | 32.60 |

Again, this is under the assumption that returns are a sunk cost, which isn't the case most of the time (in electronics they refurbish products, and surely a similar concept for bikes). There may be a much higher return amount in Australia and the Southwest, but this comes naturally with the sales and revenue they bring in, which is head and shoulders above the rest of the territories.

9. Impact of Customer Demographics on Spending (By education and occupation)

| | education_level character varying (100) 🔒 | occupation character varying (100) 🔒 | amount_spent text 🔒 | pct_contribution_to_revenue text 🔒 |
|----|--|---|------------------------|---------------------------------------|
| 1 | Bachelors | Professional | 3,422,464.75 | 13.74% |
| 2 | Bachelors | Management | 2,193,612.22 | 8.80% |
| 3 | Partial College | Professional | 2,155,846.78 | 8.65% |
| 4 | Graduate Degree | Management | 1,690,557.48 | 6.79% |
| 5 | Partial College | Skilled Manual | 1,682,416.57 | 6.75% |
| 6 | Partial College | Clerical | 1,596,745.81 | 6.41% |
| 7 | Bachelors | Skilled Manual | 1,551,588.64 | 6.23% |
| 8 | Graduate Degree | Professional | 1,265,574.05 | 5.08% |
| 9 | High School | Professional | 1,243,682.27 | 4.99% |
| 10 | Bachelors | Clerical | 1,204,877.49 | 4.84% |
| 11 | High School | Skilled Manual | 1,066,934.65 | 4.28% |
| 12 | Partial College | Manual | 961,529.42 | 3.86% |

Output to additional query exploring the different income brackets, how much they've spent and % revenue contribution:

| | income_group text | group_spend text | pct_revenue_contribution text |
|---|----------------------|---------------------|----------------------------------|
| 1 | Top 20% | 6,917,437.92 | 27.76% |
| 2 | Upper-Middle 20% | 5,667,567.57 | 22.75% |
| 3 | Lower-Middle 20% | 4,912,785.91 | 19.72% |
| 4 | Middle 20% | 4,033,888.52 | 16.19% |
| 5 | Bottom 20% | 3,382,906.90 | 13.58% |

Output to query showing countries and region with the most high-income customers:

| | country character varying (50) | region character varying (50) | high_income_customers bigint |
|----|-----------------------------------|----------------------------------|---------------------------------|
| 1 | United States | Southwest | 2053 |
| 2 | Australia | Australia | 2050 |
| 3 | United States | Northwest | 1300 |
| 4 | Canada | Canada | 547 |
| 5 | United Kingdom | United Kingdom | 414 |
| 6 | Germany | Germany | 346 |
| 7 | France | France | 313 |
| 8 | United States | Southeast | 5 |
| 9 | United States | Central | 2 |
| 10 | United States | Northeast | 1 |

This could provide insight as to where marketing efforts could be focused given the percentage contribution of high income customers to revenue, keeping in mind the still significant contribution of other income brackets

10. Top Territories by Return-Adjusted Profit

Note: The return sunk cost assumption is still at play here

| | country character varying (50) 🔒 | region character varying (50) 🔒 | returns_adj_profit text 🔒 |
|----|-------------------------------------|------------------------------------|------------------------------|
| 1 | France | France | 941,899.45 |
| 2 | Canada | Canada | 732,050.14 |
| 3 | United States | Southeast | 5,121.10 |
| 4 | Australia | Australia | 2,936,587.87 |
| 5 | United States | Northeast | 2,874.58 |
| 6 | United States | Southwest | 1,962,493.15 |
| 7 | United States | Central | 1,437.44 |
| 8 | United States | Northwest | 1,260,396.83 |
| 9 | United Kingdom | United Kingdom | 1,165,795.65 |
| 10 | Germany | Germany | 1,002,636.43 |

11. Identifying Frequently Returned Products with Low Profit Margins (Candidates for Discontinuation)

| | model_name character varying (50) 🔒 | profit_margin numeric 🔒 | return_rate text 🔒 |
|---|--|----------------------------|-----------------------|
| 1 | HL Mountain Tire | 21.91 | 3.75% |

Before discontinuing, more investigation would need to be done with respect to volume sold and how much revenue the product brings in, as though it may have appeared here, 3.75% is still quite a low return rate

12. Per-Product Subcategory 3-Month Moving Average

| | year integer | month integer | subcategory_name character varying (50) | revenue_for_month text | moving_avg_3month text |
|----|-----------------|------------------|--|---------------------------|---------------------------|
| 1 | 2021 | 7 | Bike Racks | 960.00 | 960.00 |
| 2 | 2021 | 8 | Bike Racks | 3,000.00 | 1,980.00 |
| 3 | 2021 | 9 | Bike Racks | 3,960.00 | 2,640.00 |
| 4 | 2021 | 10 | Bike Racks | 2,400.00 | 3,120.00 |
| 5 | 2021 | 11 | Bike Racks | 2,760.00 | 3,040.00 |
| 6 | 2021 | 12 | Bike Racks | 3,120.00 | 2,760.00 |
| 7 | 2022 | 1 | Bike Racks | 2,160.00 | 2,680.00 |
| 8 | 2022 | 2 | Bike Racks | 2,760.00 | 2,680.00 |
| 9 | 2022 | 3 | Bike Racks | 3,600.00 | 2,840.00 |
| 10 | 2022 | 4 | Bike Racks | 2,520.00 | 2,960.00 |
| 11 | 2022 | 5 | Bike Racks | 5,280.00 | 3,800.00 |
| 12 | 2022 | 6 | Bike Racks | 3,720.00 | 3,840.00 |

Output screenshot just shows results for one of the subcategories

13. Cross-Category Purchase Analysis (Top 5 Most frequently purchased combination of subcategories)

| | subcategory_1 character varying (50) | subcategory_2 character varying (50) | pair_purchase_count bigint | pct_of_multi_subcat_customers text |
|---|---|---|-------------------------------|---------------------------------------|
| 1 | Helmets | Tires and Tubes | 2682 | 20.20% |
| 2 | Helmets | Road Bikes | 1870 | 14.08% |
| 3 | Mountain Bikes | Road Bikes | 1725 | 12.99% |
| 4 | Bottles and Cages | Road Bikes | 1455 | 10.96% |
| 5 | Bottles and Cages | Helmets | 1354 | 10.20% |

Information such as this could serve to help the company in making recommendations to clients based on the initial product they purchase, though full recommendation systems can be developed with python.