Analysis Project Features List Template

Analysis Features	Category	Possible Points	Points	Comments
Documentation (15 total points)				
Executive Summary (up to 1 page)	Core	2		
Feature List	Core	2		
List of Web and AI sources used	Core	3		
Summary Report (up to 15 pages)	Core	8		
Analysis: Customers by Region (10 total points)				
Overview of approach with assumptions	Done	3		
Predicted sales for 2023+2024 for each province and at least two charts or visualizations	Done	4		
Recommendation/summary paragraph for management	Done	3		
Analysis: Best and Worst Products (10 points)				
Overview of approach with assumptions	Done	3		
Predicted sales for 2023+2024 for top 5 and bottom 5 products with at least two charts or visualizations	Done	4		
Recommendation/summary paragraph for management	Done	3		
Analysis: Payment Methods (up to 5 total points)				
Overview of approach with assumptions	Done	1		
Visualization of payment methods used in 2022	Done	2		
Visualization/prediction of payment methods for 2024	Done	2		
Analysis: Inventory Management (up to 10 total points)				
Overview of approach with assumptions	Done	3		
Perform analysis with at least two charts or visualizations	Done	4		
Recommendation paragraph to management	Done	3		
Analysis: Your Choice (up to 10 total points)				
Overview of approach with assumptions	Done	3		
Perform analysis with at least two charts or visualizations	Done	4		
Recommendation paragraph to management	Done	3		
Analysis: Your Choice (up to 10 total points)				
Overview of approach with assumptions	Done	3		
Perform analysis with at least two charts or visualizations	Done	4		
Recommendation paragraph to management	Done	3		
General and Open Category (up to 20 total points)				
Suggest database improvements	Done	3		
Other (your suggestions including using tools besides Excel)	Partially	Up to 20		
Total: (out of 50 with max of 10 bonus. i.e. 60 is max.)				

Group Members: Gavin Kendal-Freedman and Thomas Craig

Executive Summary

This executive summary contains the key insights of from a recent analysis of ACME's customer, product, and warehouse data. The analysis was completed using visualizations with some basic predictive modelling to create a series of data-driven recommendations. A brief overview of the findings is laid out below, followed by the full analysis.

Purpose

While ACME stores a vast amount of data regarding its customers, products, and warehouse inventories, this data has not yet been used to provide insights for potential decisions regarding strategy and policy. This lack of utilization of the available data will have caused gaps in business knowledge, and likely incorrect assumptions about the correct business strategy and policy to expand and optimize ACME's business and revenue. This executive summary and subsequent analysis seek to show the benefits of data-driven decisions and potential impacts on business performance.

Scope of Analysis Conducted

The analysis was conducted on data from 2020 through mid/late-2023, on customers, products, warehouse inventories, and reviews of ACME products. A key focus of the analysis was to enable and create predictions about future years and data for consideration.

Key Findings and Recommendations

While reviewing the data and analysis, the key findings and recommendations set out in the report can be summarized in the following points, in no specific order:

- Review differences in strategic priorities and processes between high performing provinces like Manitoba and Nova Scotia, to the underperforming provinces like Alberta and British Columbia to determine changes to strategic plans and policies,
- Review any ongoing or potential promotions or campaigns with payment processors, especially Visa and Mastercard to see if they are increasing sales or just increasing the portion of sales that go through those payment processors,
- Review warehousing processes and procedures to make sure that expensive and/or slow-moving products are not overstocked, and that products are properly distributed across warehouses based on where demand for specific products is,
- Reorganize or enhance the data process to both normalize the stored data for efficiency and security, as well as reduce liability by linking the payment method used for a specific order to the order.

Conclusion

We recommend the general adoption of data based decision-making processes throughout the ACME Corporation's processes, to proactively identify areas for improvement, and what strategies and decisions are working well. Furthermore, a general yearly review of procedures through a higher-level analysis of the company like this to identify any blind spots in more comprehensive and specific analysis of area specific data are found and rectified. Doing more analysis and strategic decisions grounded in analyses of company data will allow for more optimal growth permitting decisions and priorities, increasing year-on-year revenue and performance.

List of Web and AI sources used

- https://support.microsoft.com/en-us/office/create-a-map-chart-in-excel-f2cfed55-d622-42cd-8ec9-ec8a358b593b
- https://support.microsoft.com/en-us/office/forecast-and-forecast-linear-functions-50ca49c9-7b40-4892-94e4-7ad38bbeda99
- Office 365 Word Text Prediction while Writing the Report/Executive Summary (May count as AI?): https://support.microsoft.com/en-us/office/editor-text-predictions-in-word-7afcb4f3-4aa2-443a-9b08-125a5d692576
- https://learn.microsoft.com/en-us/sql/t-sql/queries/from-using-pivot-and-unpivot?view=sql-server-ver16

Region-Based Analysis of Sales

Overview and assumptions

Data for yearly sales volume by province was retrieved via an SQL query that summed the total dollar value of each province, for 2020-2023. The data for 2023 sales was increased by 20% as a base assumption for what the actual EOY values for 2023 sales would be for each province. A simple linear forecast in excel was then used to predict the total sales volume for each province for 2023 and 2024 using the data from 2020 through 2022. We then performed a geographical based analysis of 2022 sales, and plotting of the data and predictions was then performed to check if the predicted sales values were reasonable.

Predictions: 2023 and 2024

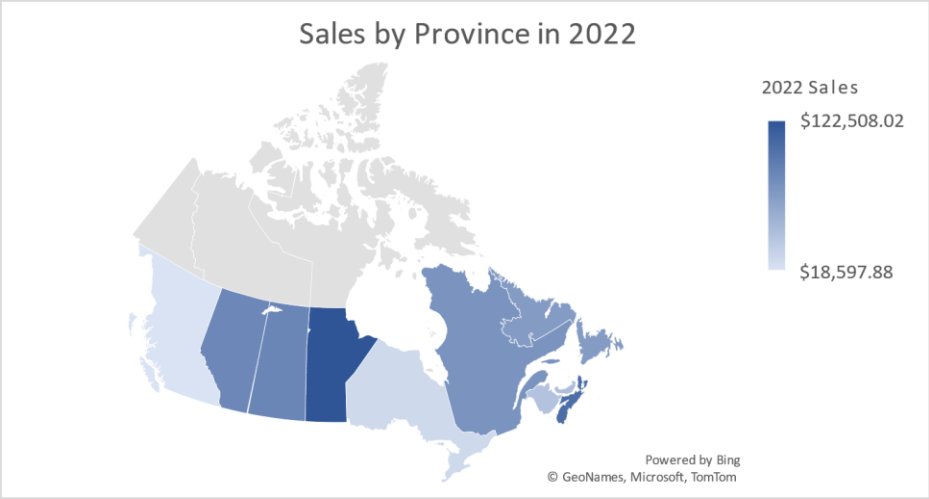
Using Excel's FORECAST.LINEAR function, the following predictions for 2023 and 2024 were generated:

Province	2020	2021	2022	2023 (Adjusted)	2023 (Predicted)	2024 (Predicted)
AB	$107,874.65	$ 129,388.94	$ 84,504.53	$ 111,461.92	$ 83,885.92	$ 72,200.86
BC	$ 38,275.86	$ 33,369.75	$ 18,597.88	$ 36,223.72	$ 10,403.18	$ 564.19
MB	$ 32,430.38	$ 70,422.49	$ 122,508.02	$ 64,843.24	$ 165,197.94	$ 210,236.76
NB	$ 54,677.71	$ 25,795.73	$ 42,169.79	$ 42,256.85	$ 28,373.16	$ 22,119.20
NF	$ 65,332.50	$ 71,624.36	$ 70,661.37	$ 73,371.89	$ 74,534.95	$ 77,199.38
NS	$ 87,858.38	$ 121,721.89	$ 104,665.09	$ 47,712.00	$ 121,555.16	$ 129,958.52
ON	$ 29,598.82	$ 39,533.94	$ 25,873.94	$ 33,378.43	$ 27,944.02	$ 26,081.58
PE	$ 50,369.32	$ 90,113.32	$ 44,909.82	$ 79,613.28	$ 56,337.99	$ 53,608.24
QC	$ 53,720.90	$ 46,438.46	$ 76,367.01	$ 95,684.88	$ 81,488.23	$ 92,811.29
SK	$ 97,501.09	$ 77,365.62	$ 86,051.54	$ 103,450.66	$ 75,523.20	$ 69,798.42

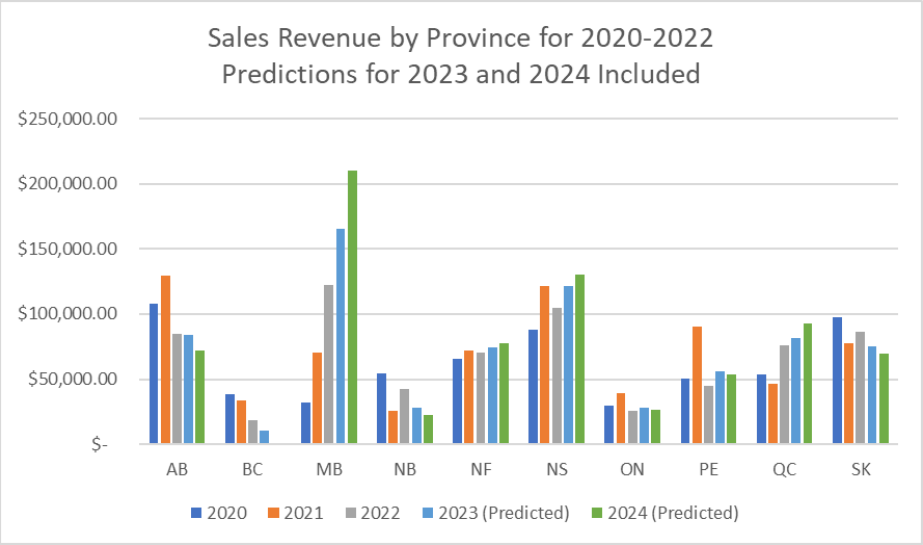
Based off the linear forecast for 2023 using the 2020 through 2022 data, for all provinces but British Columbia, the 2023 adjusted values were not used to calculate the prediction for 2024.

Based on the linear forecast, the province predicted to have the highest sales volume in 2023 and 2024 is Manitoba, predicted to break $200,000 in sales in 2024.

Using the data from 2022, we then geographically mapped the sales volume to get a rough density approximation for where ACME's sales come from, creating the following map:



Then, the data 2020 through 2022 and the 2023 and 2024 predictions were plotted to see if the trends were visibly reasonable.



This plot showed that the linear predictions were roughly reasonable for most provinces.

Summary and Recommendation

It is recommended that the company take measures to investigate appropriate strategic measures for provinces showing declines and stagnation in sales volume, namely Alberta, British Columbia, and Ontario. Furthermore, it is recommended to investigate what marketing and sales efforts have been succeeding in high growth provinces like Manitoba and Nova Scotia and see how strategic methods can be developed to apply similar approaches to other provinces.

Analysis of Best and Worst Performing Products

Overview and assumptions

This data was pulled via an SQL query, aggregating from a pivot over the year for 2020, 2021, and 2022, while 2023 and 2024 was predicted using the predict.lm function in excel. The sales in 2023 were not complete and would create a bias in the model underestimating the number of sales in the products. By introducing a multiplier on all the products, seasonal products that tend to have higher sales during the Christmas season would be underestimated, and non-seasonal products would be overestimated using a general linear model by years.

Overview and assumptions

Since the model was trained a relatively few data points, (2020, 2021 and 2022 sales per product only) The highest and lowest performing products selected are those highlighted in green and red respectively, along with their product sales.

Top Performing Products

Largest Sales	ProdId	ProdName	Non-null
42603.32	121	Gorgeous Plastic Clock	2
33727.30	16	Heavy Duty Bronze Lamp	2
29131.65	81	Aerodynamic Copper Bag	3
23030.67	36	Intelligent Concrete Wallet	3
16565.64	164	Mediocre Plastic Pants	3
15506.72	49	Awesome Cotton Watch	2
15426.62	122	Sleek Plastic Computer	3
14551.02	150	Incredible Wool Plate	3
14337.76	44	Mediocre Iron Chair	2

14091.25	118	Durable Leather Shoes	3
14049.21	112	Awesome Rubber Computer	3
13910.71	120	Incredible Marble Coat	3
13872.56	88	Awesome Aluminum Bag	3
13614.77	11	Intelligent Granite Plate	2
13598.37	65	Synergistic Silk Hat	3

This chart returns the top 15 total performing sales values, while highlighted rows are those with sales in all 3 years and should be more accurate in predicting sales compared to the non-highlighted rows, which was trained on only 2 out of the 3 years.

Worst Performing Products

Smallest Sales	ProdId	ProdName	Non-null
-13613.53	76	Gorgeous Wooden Pants	3
-12434.88	54	Small Plastic Bottle	2
-9163.005	147	Lightweight Wooden Table	3
-6748.838333	39	Heavy Duty Rubber Keyboard	3
-6474.996667	113	Sleek Steel Pants	3
-5056.756667	46	Mediocre Plastic Table	3
-4349.48	157	Fantastic Leather Bench	2
-4254.325	34	Incredible Cotton Shirt	3
-4137.215	73	Small Aluminum Bench	3
-4135.198333	4	Intelligent Iron Shoes	3
-4019.986667	191	Heavy Duty Bronze Watch	3
-3947.535	200	Heavy Duty Silk Plate	3
-3889.323333	12	Incredible Silk Bottle	3
-3777.431667	197	Fantastic Wooden Chair	3
-3661.005	160	Practical Cotton Wallet	3

This chart returns the worst 15 total performing sales, while highlighted values are those with sales in all 3 years, and should be more accurate in predicting sales compared to the non-highlighted rows, which was trained on only 2 out of the 3 years.

Recommendations and summary

The top five products are the Aerodynamic Copper bag, intelligent concrete wallet, mediocre plastic pants, sleek plastic computer, and incredible wooden plate, as there is some growth year to year, (not just between two years, which may be biased). It is also important to consider those non-highlighted rows, are potential high sellers, but without more data year-to-year data or a different modeling methodology, these predictions are likely to over-estimate the 2024 data. Likewise, the 5 worse projects: Gorgeous Wooden Pants, Lightweight Wooden Table, Heavy Duty Rubber Keyboard, Sleek steel pants and mediocre plastic table are the 5 worse performing products. The small plastic bottle product only appears on 2/3 years (2020 to 2022), and likewise may have bias to underestimate the year if it was discontinued in 2022, on which it has no sales for that year.

Therefore, it may be beneficial to discontinue Gorgeous Wooden Pants, Lightweight wooden table, heavy duty rubber keyboard, sleek steel pans, and mediocre plastic table should be discontinued, along with the small plastic bottle product if it hasn't already. Also, it would be beneficial to ensure that there is sufficient stock in the warehouses for the top 5 selling products, and some other products such as gorgeous plastic clock, and heavy duty bronze lamps, which model may overestimate their sales, as potential good sellers.

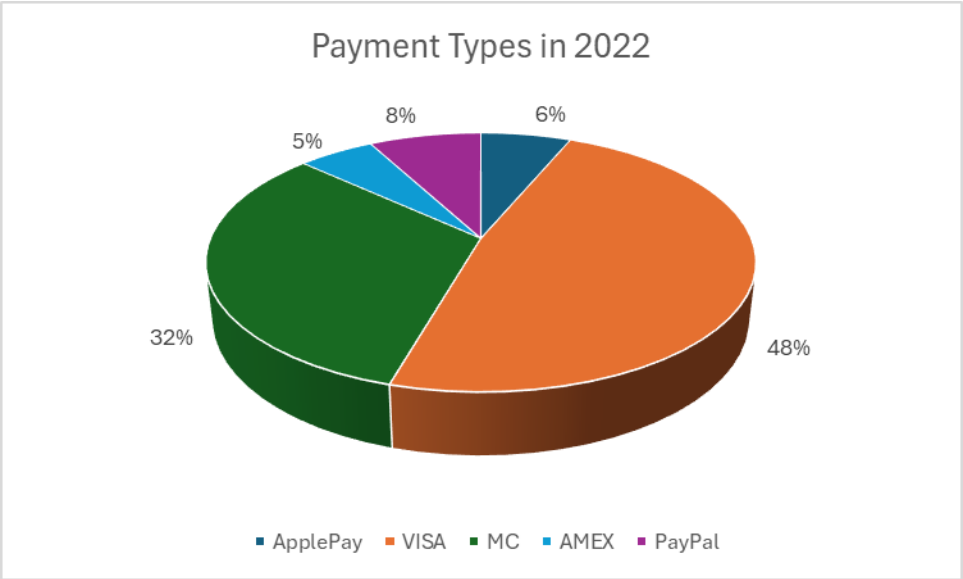
Analysis of Payment Methods

Overview and assumptions

The data was pulled via an SQL query, aggregating via a pivot over the year and payment method for all sales. It was assumed that the first payment method created by a customer was the payment method for any customer with multiple payment methods as payment methods are attached to customers, not orders. Similarly to the regional analysis, data for 2023 was increased by 20% before usage in predictions due to it being incomplete.

Payment Methods in 2022

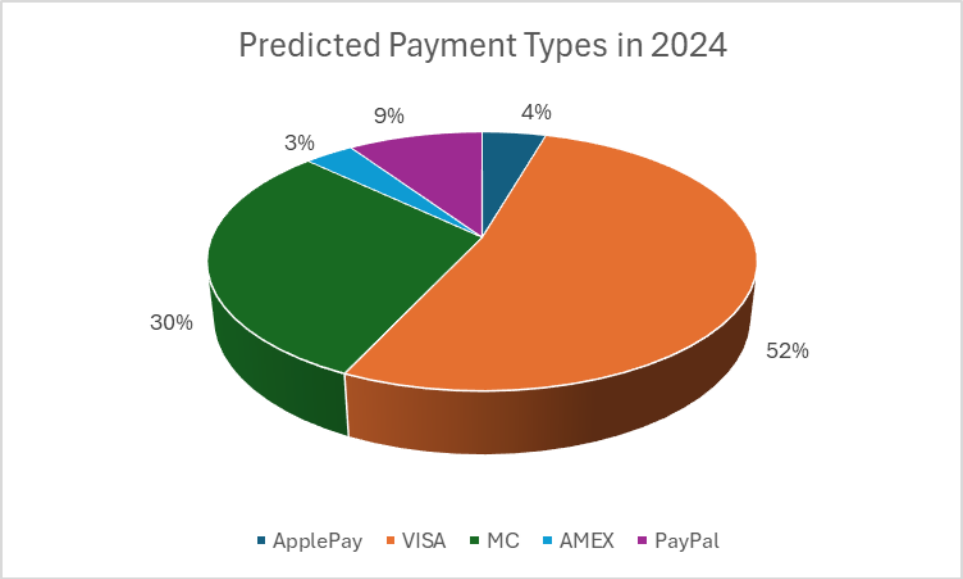
A pie chart visualizing the percentage of all sales revenue for each payment method was created to see how much each available payment method was used:



Almost half of all sales revenue came from Visa cards, with another third of all sales revenue coming from MasterCard. The remaining payment processors all individually accounted for less than 10% of total revenue.

Predicted Payment Methods in 2024

Using a simple linear forecast, the following prediction for the breakdown of payment methods for 2024 was created:



We predict that there will be very little overall change, however it is predicted that VISA will account for over 50% of 2024's revenue, with MasterCard shrinking slightly, and American Express shrinking to a marginal portion of only 3%.

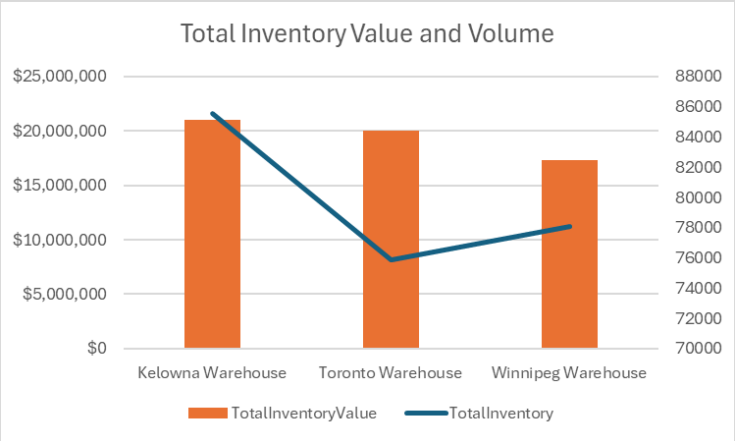
Inventory Management

Overview and assumptions

Two analyses were performed – first, a comparison of the total inventory volume and inventory value at each warehouse, and second, an analysis of the total cost of the top 10 products by total inventory volume. The only assumption made was that the price for all instances of the same item at the same warehouse is identical, as inventory is represented as just price and quantity for each item at each warehouse.

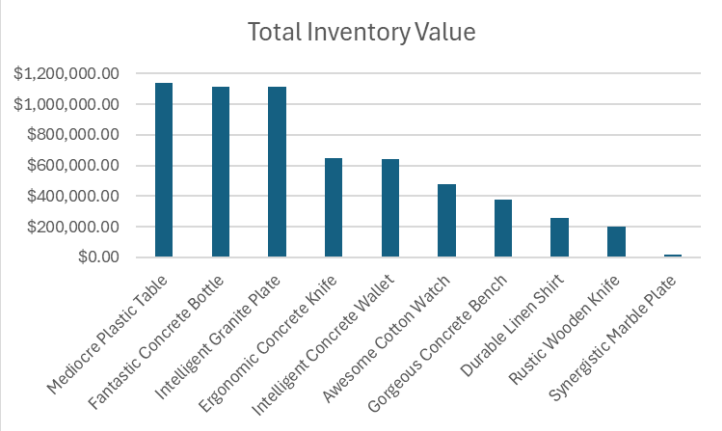
Analysis

First analysis:



Warehouse	Total Inventory	Total Inventory Value
Kelowna	85570	$21,059,976.20
Toronto	75847	$20,019,354.70
Winnipeg	78065	$17,353,783.98

Second Analysis



Product Name	Total Inventory Value	Total Inventory
Mediocre Plastic Table	$1,140,807.71	2767
Fantastic Concrete Bottle	$1,115,084.06	2717
Intelligent Granite Plate	$1,112,265.00	2486
Ergonomic Concrete Knife	$645,345.22	2498
Intelligent Concrete Wallet	$638,834.82	2354
Awesome Cotton Watch	$480,492.09	2549
Gorgeous Concrete Bench	$378,958.53	2534
Durable Linen Shirt	$257,100.80	2536
Rustic Wooden Knife	$198,661.68	2359
Synergistic Marble Plate	$18,955.43	2496

Recommendations and Summary

Reviewing each warehouse's inventory, they are all storing a similar overall value, but the Kelowna Warehouse is storing much lower value items compared to the Toronto and Winnipeg warehouses. Reviewing the total value of the inventory for the top 10 stored products by volume, no account seems to have been made to the cost of each item when determining the quantity to keep stored, as there are some items with over $1 million total value, while other items with a similar quantity have a total dollar value in tens or hundreds of thousands of dollars.

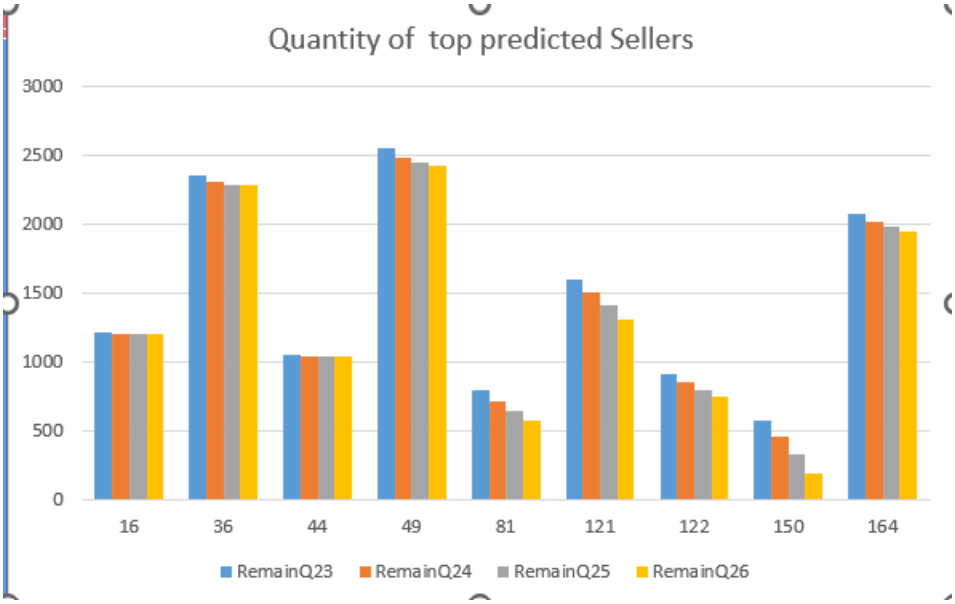
It is recommended that the company review the inventory management policies to determine how much of any item to store at any warehouse, compared to both the cost of an individual item, and how quickly each item sells. Furthermore, a risk assessment of inventory should be conducted to determine if anything catastrophic happened to any of the warehouses, the overall effect it would have on the business and if appropriate mitigation strategies are in place, given the low number of warehouses overall.

Your Choice 1: Predicting product inventory

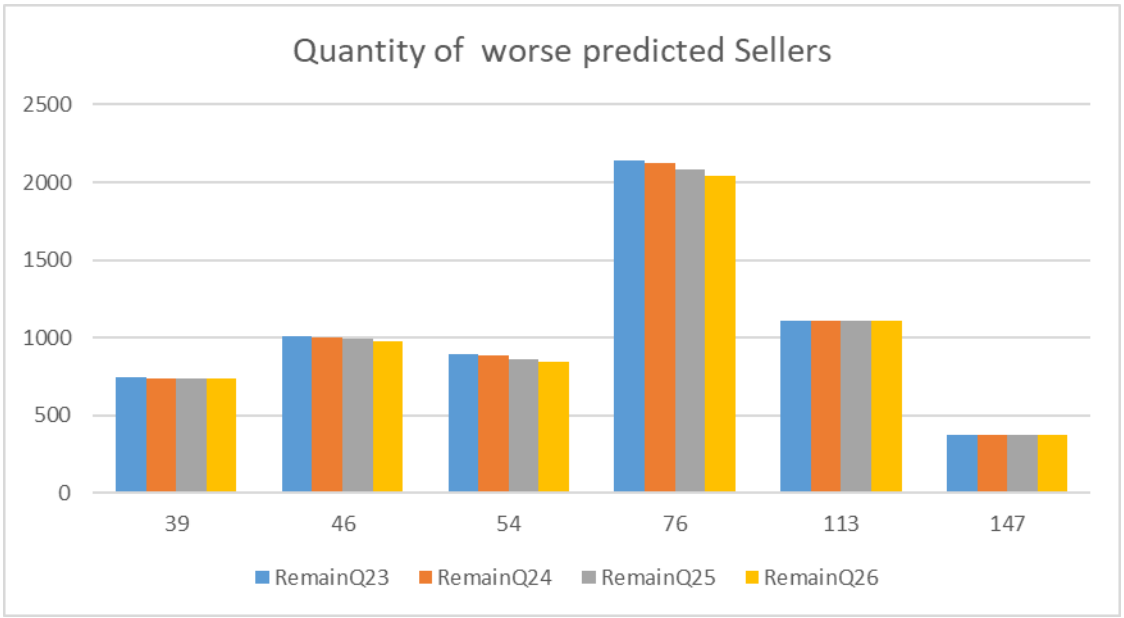
Overview and assumptions

Looking at the quantity currently in stock, and the quantity sold by year starting from 2020 to 2023, and introducing a 1.2 multiplier for 2023 quantity sold to estimate the remaining year, this analysis attempts to see if the potential top-selling products and worse-selling products would need to be restocked within the next 3 years. From the analysis above, we've identified 5 reliable top-selling products for 2024, and 5 products along with 3 more products that may sell exceptionally well, but we don't have data from all 3 years. Likewise, there is one more product that has predicted sales that is within the top five reliably worse selling products that should be included. In total, there are 8 products that are predicted to have good sales, and 6 that are not. The quantity in the warehouse or the initial quantity was taken directly from the quantity column that does not have any sales from Nov. to Dec. 2023, meaning that this prediction may not be the most accurate, as when fitting a model by month leaves far too many null values to do proper data analysis.

Analysis



Attached above are the predicted sales for the top 8 potentially high selling products. Items 81: Aerodynamic Copper Bag, 36: Intelligent Concrete Wallet and 164: Mediocre plastic pants are the more reliable predictions and the less reliable are 121: Gorgeous Plastic Clock, 16: Heavy Duty Bronze Lamp, 49: Awesome Cotton Watch.



Attached above are the predicted sales for the top 8 potentially high selling products. Items 76: Gorgeous Wooden Pants, 147: Lightweight Wooden Table, 39: Heavy Duty Rubber Keyboard, 113: Sleek Steel Pants, 46: Mediocre Plastic Table are the more accurate predictors of being worse sellers, likewise 54: small plastic bottle doesn't have data from all 3 years but is predicted to have sales within the worst 5 reliable sales data.

Recommendations and Summary

A short review of this data, one can see that the worse performing products inventory generally tends to stay the same, with the notable exception being 76, 54 and to a lesser extent 46. All the top selling products will lose products, however 16 and 44 seem that it doesn't significantly compared to the rest. Keep in mind that the products chosen have the best/worst predicted sales, not necessarily the best/worst quantity sold, meaning that items that are predicted to do well if their inventory is generally stable (few units sold), must get their profit from having relatively high prices, compared to units sold. Likewise, we can see that product 76 has many units sold but is losing money, indicating that perhaps the price/unit is too high.

Another general analysis is over the course of 3 years, product quantities don't change that much and would benefit from investing more in factors that would benefit sales, rather than restocking inventory. From the items selected, only item 150: incredible wool plate seems that it would need to be restocked within the next 5 years.

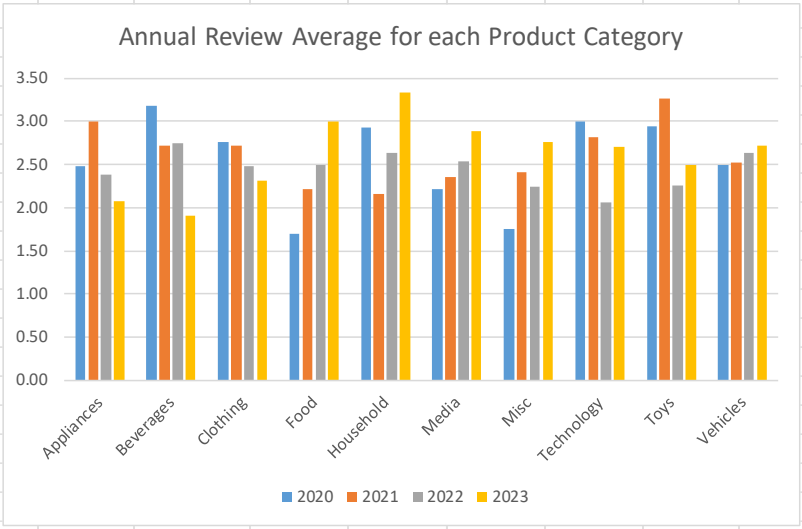
Your Choice 2: Reviews and Sales Performance for each Product Category

Overview and assumptions

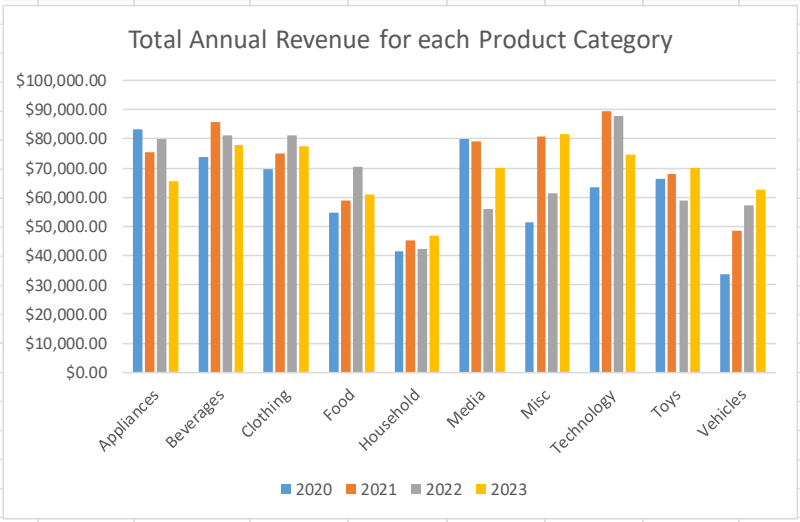
A short review of review and sales performance for each Product Category was conducted to determine if there are any trends or correlations between each Product Category's average annual reviews and annual sales revenue. As with all other analysis performed on 2023 sales, the 2023 numbers were increased by 20% to offset the effects of incomplete data.

Analysis

Category	2020	2021	2022	2023
Appliances	2.48	3.00	2.38	2.08
Beverages	3.18	2.72	2.75	1.91
Clothing	2.76	2.72	2.48	2.32
Food	1.70	2.22	2.50	3.00
Household	2.92	2.15	2.64	3.33
Media	2.22	2.36	2.54	2.89
Misc	1.75	2.41	2.25	2.76
Technology	3.00	2.81	2.06	2.70
Toys	2.94	3.27	2.26	2.50
Vehicles	2.50	2.52	2.64	2.71



Category	2020	2021	2022	2023
Appliances	$83,071.50	$75,531.22	$79,999.57	$65,606.24
Beverages	$73,627.43	$85,645.11	$81,037.59	$78,059.94
Clothing	$69,737.16	$75,081.16	$81,294.10	$77,471.45
Food	$54,869.63	$58,940.26	$70,486.35	$60,927.10
Household	$41,594.40	$45,171.55	$42,254.56	$47,045.40
Media	$79,946.78	$78,966.79	$55,903.22	$70,067.36
Misc	$51,559.06	$80,922.61	$61,521.36	$81,677.64
Technology	$63,236.17	$89,260.46	$87,920.48	$74,757.84
Toys	$66,378.21	$67,823.73	$58,785.59	$69,949.22
Vehicles	$33,619.27	$48,431.61	$57,106.17	$62,434.66



Looking at the reviews, the food, household, vehicles, and misc. categories have the most increasing trends, while technology, beverages, food, and appliances are trending downwards the most. Looking at annual sales revenue, vehicles, clothing, and misc. have the most increasing effect, while appliances, media, and beverages are the most obvious categories with a level decreasing revenue.

Recommendations and Summary

Comparing the trends and plots for reviews and revenue, the data suggests that there may be a low level of correlation between the two for some products or categories. It is recommended that further analysis be done in more detail, looking at similar products within each category, to see if products with different levels of positive reviews are selling in greater or lesser numbers, or contributing greater or lesser amounts to revenue than other similar products.

General Suggestions and Recommendations

Potential Database Improvements

We suggest investigating whether any performance or security benefits would be gained from normalization of the database schema are available, as normalizing can save storage space by reducing redundancy, which can increase efficiency with potential expansions in the business in the future.

Furthermore, we recommend recording the specific payment method used for an order, for both data integrity and legal liability concerns. With the current method, we could be assuming a customer is using a now invalid or expired payment method, and by not

keeping track of what payment method was used for a specific order, we open ourselves up to potential liability or litigation in the event we improperly process refund(s).

General Suggestions for Analysis

Using tools like Tableau, PowerBI, or Python with libraries like Pandas/Polars, Matplotlib and Seaborn, more efficient, reusable, and in-depth analysis and predictions could be done.

Furthermore, there is potential in python with a library like TensorFlow to do basic machine learning and predictions of future sales of products based on sales data and product reviews, to estimate future sales of products based on how likely people would be to purchase it based on reviews of the item, not just raw historical sales.

Further analysis into potential avenues of sales modeling could be done through either Principal Component Analysis (PCA), or Entropy to analyze what the actual drivers of sales and variation between product sales, between the product category, cost, purchaser location, time of year, etc. Doing this could provide a better understanding of the strengths and weaknesses of ACME's product lineup and business strategies, as well as suggest potential areas for improvement.