

Exploring Data

Jake O'Connor

MS544

Assignment 3 – Exploring, Interpreting, and Recommending

Introduction

This report will explore the data contained within the company's sales dataset and try to identify any meaningful patterns found within. The dataset contains sixty distinct sales entries across seven product categories and should serve as a good foundation to draw meaningful conclusions for the company. Within this report, each value of the dataset will be explained and evaluated for usefulness, additional metrics will be developed, and analyses will be conducted on the whole dataset. Finally, based on the conclusions drawn from analysis, recommendations will be provided to the company's stakeholders.

Raw Data

The original dataset came with six tracked values, one of which was a derived value. These values cover the sales of seven disparate products to sixty different customers in markets with five different levels of competition.

Customer

In the dataset, customers are given unique numbers to track their sales. These are ordinal values since they function as both named variables and are ordered, presumably chronologically, but as unique customer identifiers there is not enough information to draw meaningful intervals between values.

Product Code

In the dataset, each of the seven products is given a unique integer code. These are ordinal values since they are unique named variables and are ordered, presumably by release date, but do not carry any inherent interval meaning between product codes.

Competitive Rating

In the dataset, each sale entry is tagged with an integer value denoting the current competitiveness of the market in which the sale took place. A low competitiveness rating indicates that the sale took place in a market with low competition, and a high rating indicates that the market was saturated with competition. These competitive ratings are ordinal variables, as the interval between each value cannot be guaranteed.

Gross Sales

In the dataset, each entry tracks the gross monetary value of the sale. As monetary values, these are top-notch ratio data that is ordered, proportionate, and has a zero value.

Gross Profit

In the dataset, each entry tracks the gross profit value of the sale. As monetary values, these are top-notch ratio data that is ordered, proportionate, and has a zero value. Gross profit values of sales are those which include variable costs (costs which scale with the size of the transaction, e.g. tax, materials, shipping, etc.) but not fixed costs (costs which do not scale with the size of the transaction, e.g. rent, utilities, manufacturing equipment, etc.).

Percent of Gross Profit

In the dataset, each entry has a calculated percent of gross profit (profit margin). This is the ratio between of gross profits to gross sales for the single sale. As a ratio, this metric within the dataset is top-tier data.

Metrics

For the sake of deriving meaningful data from this dataset, I've elected to add multiple metrics within each sales entry. These metrics are detailed below and are used in later calculations to offer recommendations to the company.

Percent of Product Profit

For each entry in the dataset, I've added a calculated value which is the ratio of the sale's profit to the total profits of all sales for the same product in the dataset. This value can help to identify trends in the meaningful profits for each product over time.

Percent of Total Profit

For each entry in the dataset, I've added a calculated value which is the ratio of the sale's profit to the total profits for all sales within the dataset. This value can help to identify trends in the meaningful profits for each sale over time.

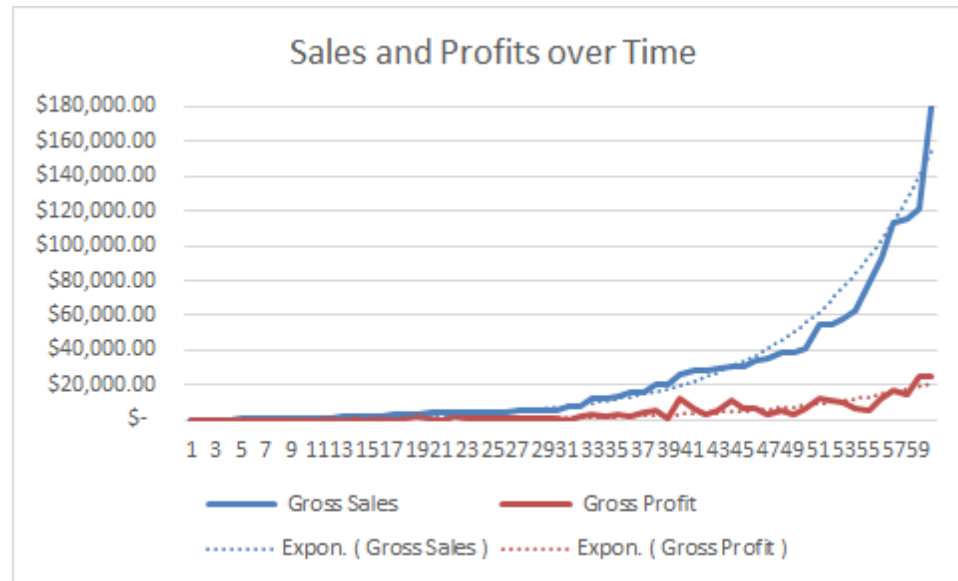
Analysis

Taking both the raw dataset and derived metrics together, there are a few important analyses and corresponding recommendations that can be drawn. Analyzing the dataset reveals that there are some negative trends within the business's sales that need to be corrected for the overall benefit of the company's future.

Sales and Profit over Time

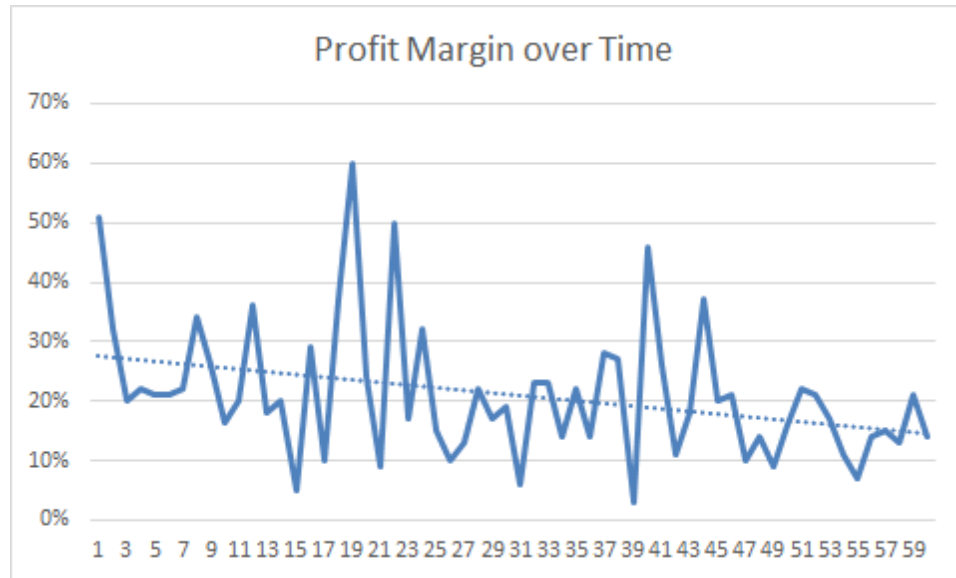
Simple analysis of the dataset reveals that both gross revenue and gross profits are reliably increasing over time with each new sale and customer. This is good! What isn't good is that revenue and profits are not growing at the same rate. Where the gross sales over time are increasing at an

exponential rate, the gross profit from those sales is increasing at a much more linear rate. Since exponential growth in sales is not sustainable in anything but the near term, it is important to plan for an inevitable slowdown in the growth rate of gross sales.



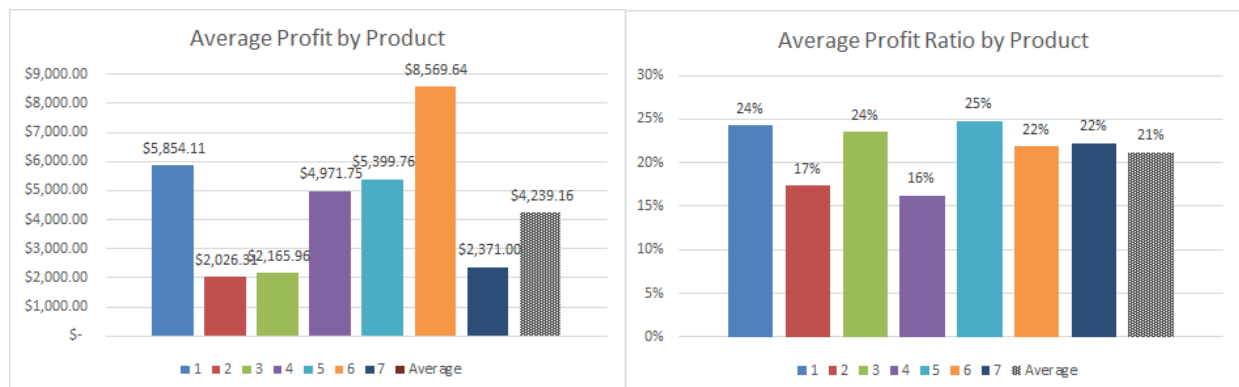
Profit Margins over Time

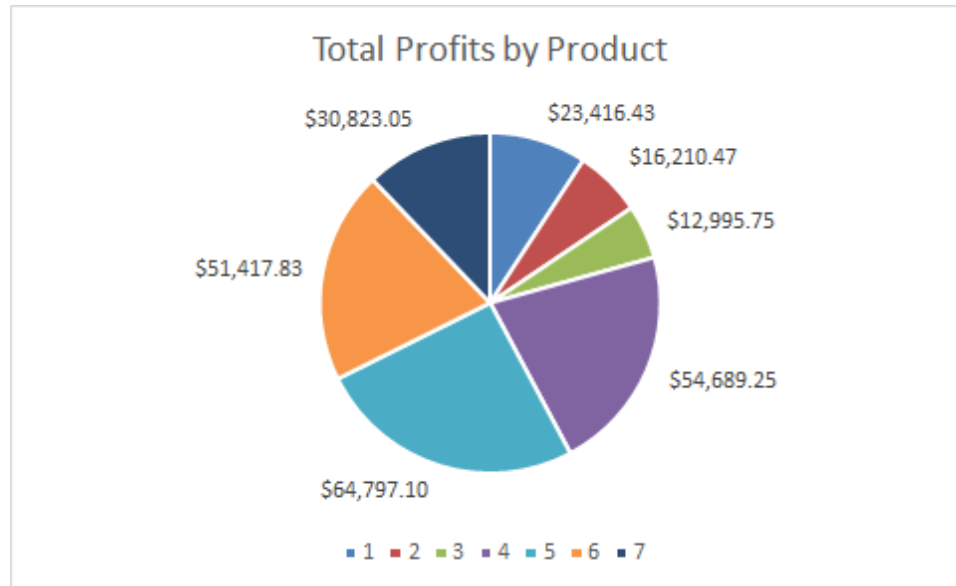
A startling truth extracted from the dataset is that the profit margins of the company's sales are decreasing rapidly. The average profit margin of all sales in the dataset is 21%, but the data is trending quickly downwards and will become unsustainable for the company's continued operations. As the revenue for individual sales grows larger and larger, the corresponding profits are rapidly declining. It would behoove the company to ensure that profit margins at least remain consistent, and ideally increase as unprofitable products are discontinued and operations are optimized.



Per-Product Profits

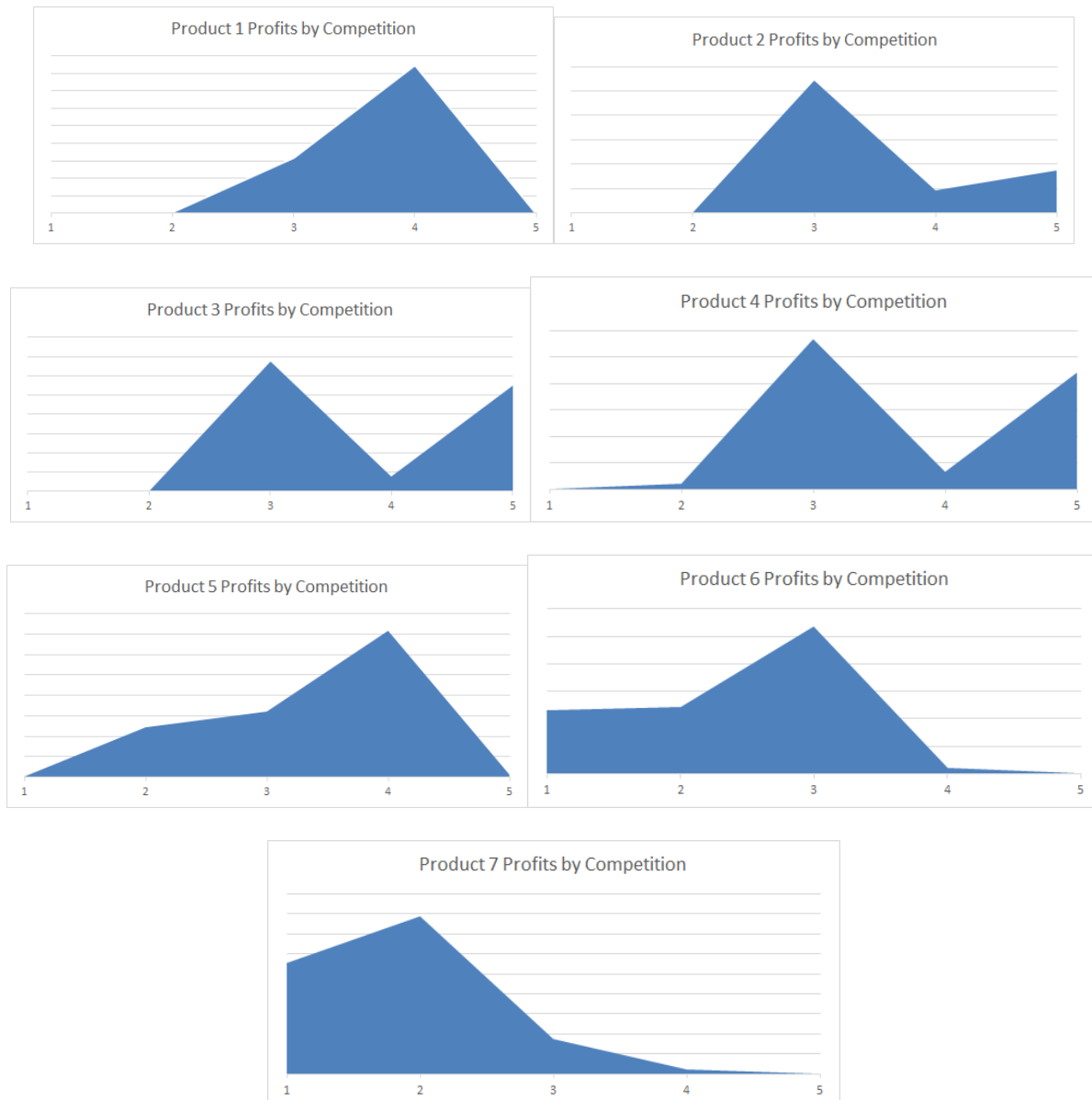
Over the entire dataset, certain products are demonstrably less profitable than others. Products 4, 5, and 6 make up a large majority of the company's profits and have higher-than-average per-sale profits. Products 1, 2, and 3 make up only a small portion of the company's profits. Product 7 makes up a significant portion of the company's total profits but has very low average per-sale profits. Products 2 and 4 have significantly lower average margins than the other products in the dataset, but not startlingly lower to the point where those products would need to be removed from the catalog.





Product Market Competition

Certain products in the dataset have been demonstrated to be more profitable than others in markets with certain levels of competition. Products 1-4 do well in markets with medium-high competition (ratings 3-5), product 5 does well in almost all competition ranges, and products 6 and 7 do much better in lower competition markets (ratings 1-3). These patterns in the profitability of certain markets can point us to doubling down on marketing efforts in markets with certain levels of competition on a product-by-product basis in order to maximize sales.



Conclusion

The most meaningful recommendation I can make from this dataset is to stem the bleeding of the rapidly shrinking profit margins. Profit margins on all sales are trending downwards incredibly quickly, and if they follow projections will be below 10% soon. Since there are no major differences between the average profit margins of each product, that leads me to believe that either per-unit costs are increasing at an uncontrolled rate or the leeway of salespeople to offer discounts for bulk purchases

is too large. This distinction could be easily drawn with additional datasets that include more data about the product manufacturing pipeline, including itemized details on each of the variable costs, and more data about the sales process, including MSRP values and applied discounts. Using these two new datasets, an exact cause of the rapidly declining profit margins could be identified before the company becomes unsustainable.