SQL Project Report on:

**Analyzing Digi Key Product Manufacturer and Product**

**Details Data Interpretation**

**&**

**Business Recommendations**

Dhaval Rathod

SK23004

**Project Overview:**

In this SQL project, we will analyze a dataset containing product details and their manufacuterer details information. The dataset includes the following columns: product\_url, category\_url, part\_number, description, manufacturer, manufacturer\_url, datasheet\_url, rohs\_compliant, stock, price, product\_category, manufacture\_part\_number, digikey\_part\_number, lead\_time, input\_link, ratings, series, package, operating\_temperature, features, mounting\_type, package\_case, supplier\_device\_package, type, image, documents\_and\_media, url, pricing\_per\_quantity, moisture\_sensitivity\_level\_msl, eccn, htsus, additional\_resources, other\_names, marketplace\_product, standard\_package, mfr\_standard\_lead\_time, alternate\_packaging, categories, stock\_numeric, stock\_numeric2, date\_added, combined\_stock. We will perform various analyses to gain insights into manufacturer behavior and product manufacturing patterns.

**Step 1: Defining Metadata in MySQL Workbench**

We began by defining the metadata for our dataset in MySQL Workbench. We created a table named `product\_metadata` with the columns described above.

**Step 2: Exploratory Data Analysis (EDA)**

We started by conducting exploratory data analysis to understand the characteristics of the dataset.

1.Identify manufacturers whose products are priced higher than the average price within their respective product categories. For each manufacturer and category, calculate the count of products that exceed the category’s average price and rank the results in descending order based on the number of high-priced products.

2.Identify products at high risk of stockouts based on inventory levels and lead times. Specifically, focus on products with stock quantities below 30 and lead times exceeding 20 days. Highlight the top 5 high-risk products sorted by lead time in descending order, prioritizing those with the longest replenishment times.

**3.**Analyze the average lead times of manufacturers to identify those whose lead times exceed the overall industry average. Specifically, determine the manufacturers with above-average lead times, rank them in descending order, and highlight those with the longest lead times.

4.Identify the most common features for each product category to understand which attributes are most frequently associated with products in the dataset.

5.Identify the Least profitable category accross dataset.

6.Determine the top 3 most expensive products within each product\_category.

7.Find the average lead time for products grouped by manufacturer to identify slow suppliers.

8.Calculates the average price of products in each category over time.

**Advance Queries:**

9.Analyze product pricing across manufacturers and series to identify the average price for each combination.

10.Determine the most popular manufacturers in each product category based on the number of products they offer.

11.Analyze the stock levels of products across different categories to assess inventory health. Calculate the average, maximum, and minimum stock for each product category, and categorize them as “Overstocked,” “Understocked,” or “Optimal Stock” based on the overall average stock in the system.

12.Analyze the revenue generated by each manufacturer by calculating the total revenue (price multiplied by stock quantity), average product price, and average stock for each manufacturer.

13.Evaluate the financial performance of manufacturers by calculating their total revenue (the product of price and stock quantity), average product price, and average stock.

14.Analyze the revenue generated by each product category by calculating the total revenue (price multiplied by stock quantity), average stock levels, and the total number of products within each category.

15.Identify and compare products within the same category that have similar pricing, with a price difference of less than $50. This analysis involves pairing products with similar features (same product category) and calculating the price difference between them.

**Step 3: Business Recommendations :**

Let's go through each step of your project task and provide detailed answers:

**Step 1: Defining Metadata in MySQL Workbench**

For this step, we would need to define the structure of our database table in MySQL Workbench. Here's an example of how we might define the metadata for our table:

Table Name: PurchaseData

Columns:

product\_url VARCHAR(255),

category\_url VARCHAR(255),

part\_number VARCHAR(255) PRIMARY KEY,

description TEXT,

manufacturer VARCHAR(255),

manufacturer\_url VARCHAR(255),

datasheet\_url VARCHAR(255),

rohs\_compliant BOOLEAN,

stock INT,

price DECIMAL(10, 2),

product\_category VARCHAR(255),

manufacture\_part\_number VARCHAR(255),

digikey\_part\_number VARCHAR(255),

lead\_time INT,

input\_link VARCHAR(255),

ratings DECIMAL(3, 2),

series VARCHAR(255),

package VARCHAR(255),

operating\_temperature VARCHAR(255),

features TEXT,

mounting\_type VARCHAR(255),

package\_case VARCHAR(255),

supplier\_device\_package VARCHAR(255),

type VARCHAR(255),

image VARCHAR(255),

documents\_and\_media TEXT,

url VARCHAR(255),

pricing\_per\_quantity VARCHAR(255),

moisture\_sensitivity\_level\_msl INT,

eccn VARCHAR(50),

htsus VARCHAR(50),

additional\_resources TEXT,

other\_names TEXT,

marketplace\_product BOOLEAN,

standard\_package VARCHAR(255),

mfr\_standard\_lead\_time INT,

alternate\_packaging TEXT,

categories TEXT,

stock\_numeric INT,

stock\_numeric2 INT,

date\_added DATE,

combined\_stock INT

**Step 2: Exploratory Data Analysis (EDA)**

**1.**Identify manufacturers whose products are priced higher than the average price within their respective product categories. For each manufacturer and category, calculate the count of products that exceed the category’s average price and rank the results in descending order based on the number of high-priced products.

**WITH CategoryAveragePrice AS (**

**SELECT**

**product\_category,**

**AVG(price) AS avg\_price**

**FROM**

**product\_metadata**

**GROUP BY**

**product\_category**

**),**

**HighPriceManufacturers AS (**

**SELECT**

**pm.manufacturer,**

**pm.product\_category,**

**pm.price,**

**cap.avg\_price**

**FROM**

**product\_metadata pm**

**JOIN**

**CategoryAveragePrice cap**

**ON**

**pm.product\_category = cap.product\_category**

**WHERE**

**pm.price > cap.avg\_price**

**)**

**SELECT**

**manufacturer,**

**product\_category,**

**COUNT(\*) AS high\_price\_count**

**FROM**

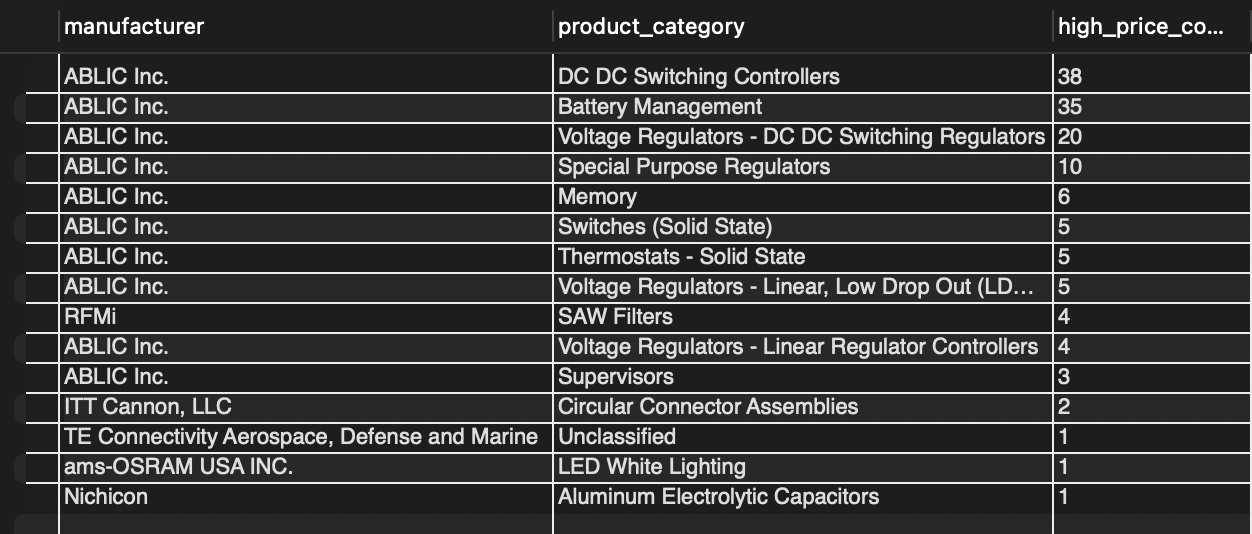
**HighPriceManufacturers**

**GROUP BY**

**manufacturer, product\_category**

**ORDER BY**

**high\_price\_count DESC;**



**Interpretation**: The goal of this analysis is to identify manufacturers producing products priced above the average in their respective categories, thus dominating the premium market segment. Using SQL, the average price for each category is calculated, and products priced higher than this benchmark are flagged. The output reveals that **ABLIC Inc.** leads in multiple categories, including “DC DC Switching Controllers” with 38 premium products and “Battery Management” with 35 premium products, showcasing its dominance in high-value segments. This insight helps identify manufacturers excelling in the premium market, aiding in strategic market positioning and supplier collaboration.

**2.**Identify products at high risk of stockouts based on inventory levels and lead times. Specifically, focus on products with stock quantities below 30 and lead times exceeding 20 days. Highlight the top 5 high-risk products sorted by lead time in descending order, prioritizing those with the longest replenishment times

**WITH Riskproduct AS(**

**SELECT**

**part\_number,**

**product\_category,**

**manufacturer,**

**stock\_numeric,**

**stock\_numeric2,**

**lead\_time**

**FROM**

**product\_metadata**

**WHERE**

**stock\_numeric<30 AND lead\_time>20**

**)**

**SELECT**

**part\_number,**

**product\_category,**

**stock\_numeric,**

**stock\_numeric2,**

**lead\_time,**

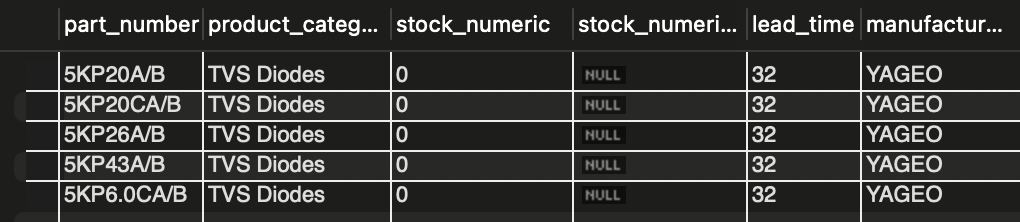
**manufacturer**

**FROM**

**Riskproduct**

**ORDER BY**

**lead\_time DESC LIMIT 5;**



**Interpretation**: This query focuses on products with stock levels below 30 units and lead times exceeding 20 days, identifying them as high-risk items prone to stockouts. The output highlights **TVS Diodes** from **YAGEO** (e.g., 5KP20A/B and 5KP43A/B) with zero stock and 32 days lead time. Such insights emphasize the need for proactive supply chain strategies, including optimizing lead times and maintaining buffer stock for critical categories to mitigate potential disruptions.

3.Analyze the average lead times of manufacturers to identify those whose lead times exceed the overall industry average. Specifically, determine the manufacturers with above-average lead times, rank them in descending order, and highlight those with the longest lead times.

**WITH Avgleadtime AS (**

**SELECT**

**AVG(lead\_time) AS avg\_lead\_time**

**FROM**

**product\_metadata**

**),**

**Longleadtime AS (**

**SELECT**

**AVG(lead\_time) AS manufacurer\_avg\_lead\_time,**

**manufacturer**

**FROM**

**product\_metadata**

**GROUP BY**

**manufacturer**

**HAVING**

**AVG(lead\_time)>(SELECT avg\_lead\_time FROM Avgleadtime)**

**)**

**SELECT**

**manufacturer,**

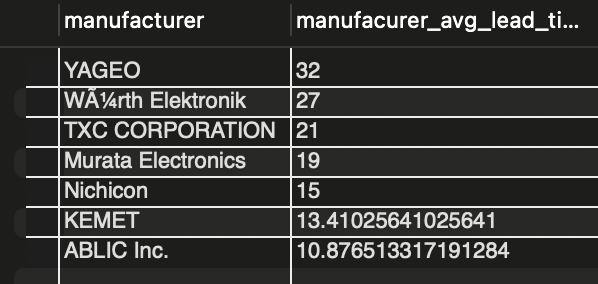
**manufacurer\_avg\_lead\_time**

**FROM**

**Longleadtime**

**ORDER BY**

**manufacurer\_avg\_lead\_time DESC;**



**Interpretation**: To identify supply chain bottlenecks, this analysis ranks manufacturers with average lead times exceeding the industry norm. **YAGEO** tops the list with an average lead time of 32 days, followed by **Würth Elektronik** at 27 days. These findings underline the need to renegotiate supplier terms, diversify partnerships, and implement predictive models to forecast supplier delays, thereby improving overall supply chain efficiency.

**4.**Identify the most common features for each product category to understand which attributes are most frequently associated with products in the dataset.

**WITH Featurecounts AS(**

**SELECT**

**product\_category,**

**features,**

**COUNT(\*) AS feature\_count**

**FROM**

**product\_metadata**

**GROUP BY**

**product\_category,features**

**),**

**Topfeatures AS(**

**SELECT**

**product\_category,**

**features,**

**feature\_count,**

**RANK() OVER(PARTITION BY product\_category ORDER BY feature\_count DESC) AS renk**

**FROM**

**Featurecounts**

**)**

**SELECT**

**product\_category,**

**features,**

**feature\_count**

**FROM**

**TopFeatures**

**WHERE**

**renk =1;**

**Interpretation**: This analysis identifies the most prevalent features in each product category by counting their occurrences and ranking them. For example, the “Supervisors” feature appears 248 times, making it the most frequent feature. Similarly, “Voltage Regulators - Linear Low Dropout” (LDO) shows 359 occurrences. These insights guide feature prioritization for product design and targeted marketing strategies, ensuring alignment with customer preferences.

**5.**Identify the Least profitable category accross dataset.

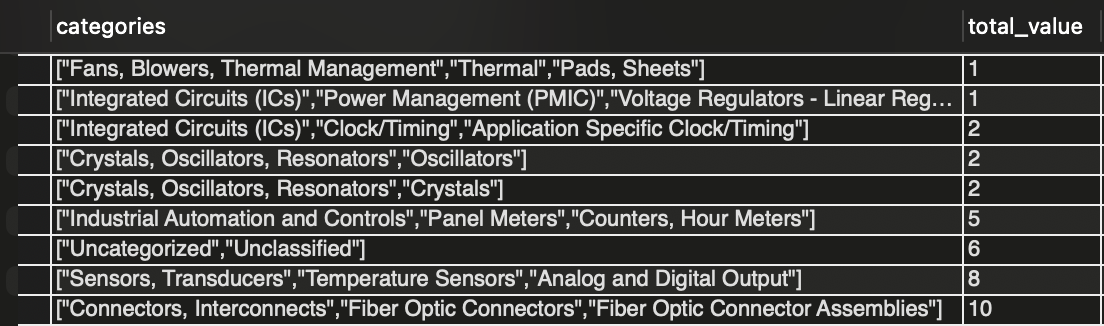
**SELECT categories,**

**SUM(price \* stock\_numeric) AS total\_value**

**FROM product\_metadata**

**GROUP BY categories**

**ORDER BY total\_value ASC;**

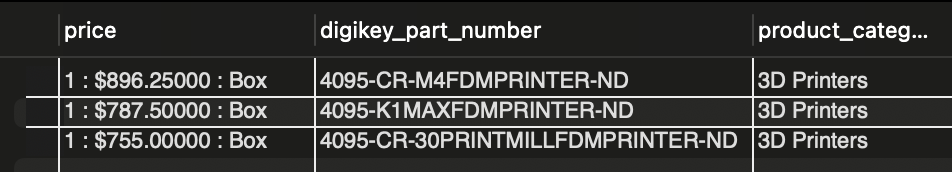


**Interpretation**: This analysis identifies the least profitable product category by calculating total inventory value (price × stock) and ranking categories in descending order. Low-performing categories highlight issues like low demand or pricing inefficiencies, guiding the company to optimize pricing, adjust inventory, or reallocate resources to maximize profitability.

**6.**Determine the top 3 most expensive products within each product\_category.

**SELECT price,digikey\_part\_number,product\_category FROM product\_metadata**

**WHERE price IS NOT NULL ORDER BY product\_category,price DESC LIMIT 3 ;**

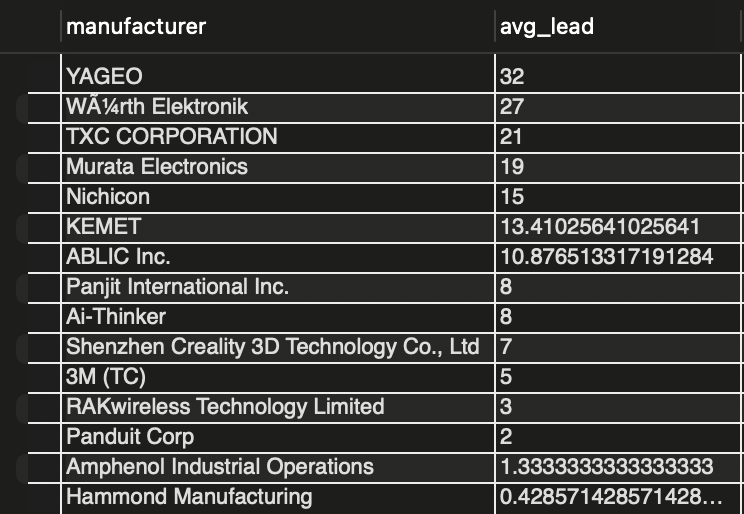


**Interpretation**: The query identifies the top three most expensive products within each category, focusing here on “3D Printers,” ordered by price in descending order. The results highlight premium products, including the 4095-CR-M4FDMPRINTER-ND priced at $896.25, the 4095-K1MAXFDMPRINTER-ND at $787.50, and the 4095-CR-30PRINTMILLFDMPRINTER-ND at $755.00 per box. This analysis provides valuable insights for prioritizing premium product marketing, optimizing pricing strategies, and understanding high-value product trends within the category to enhance competitive positioning and profitability.

**7.**Find the average lead time for products grouped by manufacturer to identify slow suppliers.

**SELECT manufacturer , AVG(lead\_time) AS avg\_lead FROM product\_metadata**

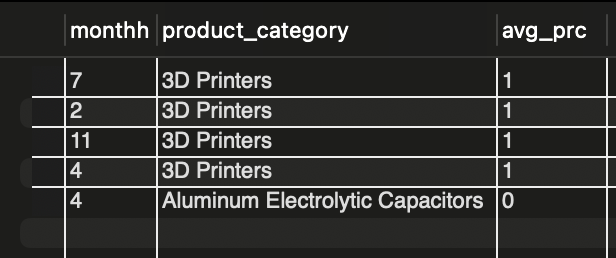
**GROUP BY manufacturer ORDER BY avg\_lead DESC LIMIT 15;**



**Interpretation**: The query reveals that the manufacturer **YAGEO** has the longest average lead time of 32 weeks, indicating potential delays or inefficiencies in its supply chain. On the other hand, **Hammond Manufacturing** stands out with a significantly lower average lead time of approximately 0.43 weeks, highlighting its efficiency in product manufacturing and delivery.

**8.**Calculates the average price of products in each category over time.

SELECT EXTRACT(MONTH FROM date\_added ) AS monthh ,product\_category,AVG(price) AS avg\_prc FROM product\_metadata GROUP BY product\_category,monthh ORDER BY product\_category,avg\_prc LIMIT 5;



**Interpretation**: From the output, we can observe that the **3D Printers** category consistently shows an average price of 1 across multiple months (February, July, and November). Additionally, **Aluminum Electrolytic Capacitors** has an average price of 0 for the month of April.

**Advance Queries:**

**9.**Analyze product pricing across manufacturers and series to identify the average price for each combination.

**SELECT**

**manufacturer,**

**series,**

**AVG(price) AS avg\_price**

**FROM**

**product\_metadata**

**GROUP BY**

**manufacturer,series**

**ORDER BY**

**avg\_price DESC ;**

**Interpretation**: The pricing analysis determines the average price of products across manufacturers and series to identify premium offerings. The output highlights OSLON® SSL 80 by ams-OSRAM USA INC. as the most premium series, with an average price of $600, followed by Meritek’s MF series at $500. These results support pricing strategy optimization and benchmarking for competitive positioning in the high-value market segment.

**10.**Determine the most top 5 popular manufacturers in each product category based on the number of products they offer.

**WITH Manufacturerpopularity AS (**

**SELECT**

**product\_category,**

**manufacturer,**

**COUNT(\*) AS product\_count**

**FROM**

**product\_metadata**

**GROUP BY**

**product\_category,manufacturer**

**),**

**Topmanufacturer AS (**

**SELECT**

**product\_category,**

**manufacturer,**

**product\_count,**

**RANK() OVER(PARTITION BY product\_category ORDER BY product\_count DESC ) renk**

**FROM**

**Manufacturerpopularity**

**)**

**SELECT**

**product\_category,**

**manufacturer,**

**product\_count**

**FROM**

**Topmanufacturer**

**WHERE**

**renk=1 ORDER BY product\_count DESC LIMIT 5;**



**Interpretation**: This query ranks manufacturers based on the number of products they offer in each category. The output reveals **ABLIC Inc.** as the leader in “Battery Management” with 72 products, and **KEMET** leading in “Ceramic Capacitors” with 39 products. These findings assist in assessing competitive positioning and recognizing key players dominating specific categories.

**11.**Analyze the stock levels of products across different categories to assess inventory health. Calculate the average, maximum, and minimum stock for each product category, and categorize them as “Overstocked,” “Understocked,” or “Optimal Stock” based on the overall average stock in the system.

**WITH Stockanalysis AS(**

**SELECT**

**product\_category,**

**AVG(stock\_numeric) AS avg\_stock,**

**MAX(stock\_numeric) AS max\_stock,**

**MIN(stock\_numeric )AS min\_stock**

**FROM**

**product\_metadata**

**GROUP BY**

**product\_category**

**)**

**SELECT**

**product\_category,**

**avg\_stock,max\_stock,min\_stock,**

**CASE**

**WHEN avg\_stock > (SELECT AVG(stock\_numeric) FROM product\_metadata) THEN 'Overstocked'**

**WHEN avg\_stock < (SELECT AVG(stock\_numeric) FROM product\_metadata) THEN 'Understocked'**

**ELSE 'Optimal Stock'**

**END AS stock\_status**

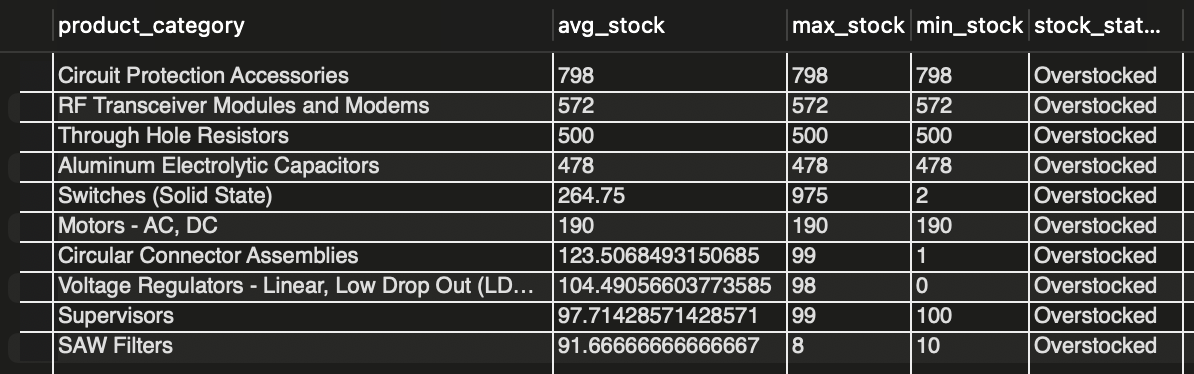
**FROM**

**Stockanalysis**

**ORDER BY**

**stock\_avg\_stock DESC LIMIT 10;**

## In this querie I had done the sorting in order by avg\_stock in desending and set the limit 10 for showing output in short format…..



**Interpretation**: This analysis categorizes product categories as “Overstocked,” “Understocked,” or “Optimal Stock” based on inventory levels. For instance, “Through Hole Resistors” are overstocked with an average stock of 500 units, while “TVS Diodes” are understocked with an average stock of 3.5 units. Such classifications enable data-driven inventory management, reducing costs associated with overstocking and mitigating risks of stockouts.

**12.**Analyze the revenue generated by each manufacturer by calculating the total revenue (price multiplied by stock quantity), average product price, and average stock for each manufacturer.

**SELECT**

**manufacturer,**

**SUM(price \* stock\_numeric) AS total\_revenue,**

**AVG(price) AS avg\_price,**

**AVG(stock\_numeric) AS avg\_stock**

**FROM**

**product\_metadata**

**GROUP BY**

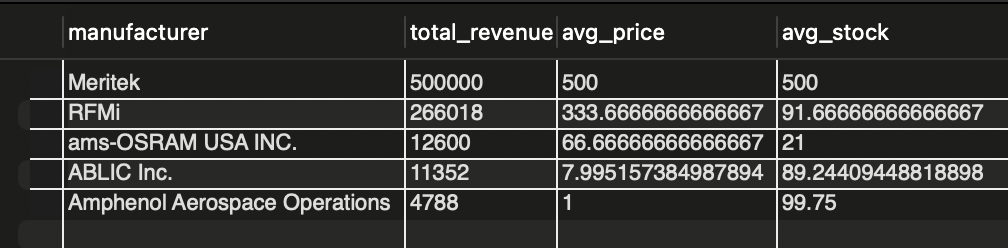
**manufacturer**

**HAVING**

**total\_revenue > (SELECT AVG(price \* stock\_numeric) FROM product\_metadata)**

**ORDER BY**

**total\_revenue DESC;**



**Interpretation**: By calculating total revenue (price × stock), the analysis identifies high-performing manufacturers. **Meritek** leads with $500,000 in revenue, followed by **RFMi** at $266,018. These results provide actionable insights for prioritizing partnerships with top-revenue manufacturers and optimizing their product availability for sustained growth.

**13.**Evaluate the financial performance of manufacturers by calculating their total revenue (the product of price and stock quantity), average product price, and average stock.

**SELECT**

**manufacturer,**

**SUM(price \* stock\_numeric) AS total\_revenue,**

**ROUND(AVG(price),2) AS avg\_price,**

**ROUND(AVG(stock\_numeric),2) AS avg\_stock**

**FROM**

**product\_metadata**

**GROUP BY**

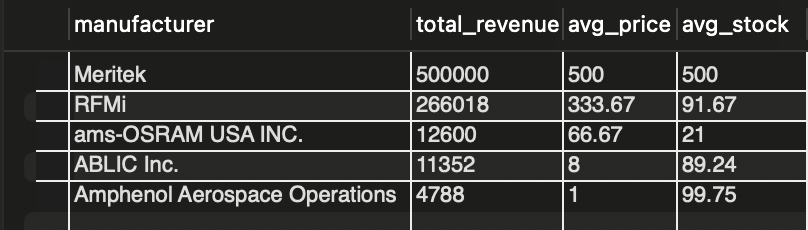
**manufacturer**

**HAVING**

**total\_revenue > (SELECT AVG(price \* stock\_numeric) FROM product\_metadata)**

**ORDER BY**

**total\_revenue DESC;**



**Interpretation**: This query evaluates product categories based on total revenue, average stock, and total product counts. Categories like “Through Hole Resistors” generate the highest revenue ($500,000), followed by “SAW Filters” ($266,018). These insights inform resource allocation and highlight profitable categories for strategic focus.

14.Analyze the revenue generated by each product category by calculating the total revenue (price multiplied by stock quantity), average stock levels, and the total number of products within each category.

**SELECT**

**product\_category,**

**SUM(price\*stock\_numeric) As revenue,**

**AVG(stock\_numeric) AS avg\_stock,**

**COUNT(\*) AS total\_products**

**FROM**

**product\_metadata**

**GROUP BY**

**product\_category**

**ORDER BY revenue DESC;**

**Interpretation**: To identify closely priced products, the analysis examines pairs within the same category with a price difference below $50. The results reveal significant overlaps in pricing among similar products, providing opportunities for bundling, dynamic pricing, and enhanced market positioning to attract cost-sensitive customers.

**15.**Identify and compare products within the same category that have similar pricing, with a price difference of less than $50. This analysis involves pairing products with similar features (same product category) and calculating the price difference between them.

**SELECT**

**p1.part\_number AS product\_1,**

**p2.part\_number AS product\_2,**

**p1.product\_category,**

**p1.manufacturer,**

**p1.price AS price\_1,**

**p2.price AS price\_2,**

**ABS(p1.price - p2.price) AS price\_difference**

**FROM**

**product\_metadata p1**

**JOIN**

**product\_metadata p2**

**ON p1.product\_category = p2.product\_category**

**AND p1.part\_number != p2.part\_number**

**WHERE**

**ABS(p1.price - p2.price) < 50**

**ORDER BY**

**price\_difference DESC;**

**Interpretation**: This analysis identifies products within the **“Voltage Regulators - Linear, Low Drop Out (LDO) Regulators”** category by **ABLIC Inc.** that exhibit minimal price differences, with a maximum observed difference of $0.009 per unit at high-volume pricing tiers. Products like **S-817B60AUA-CXXT2U** and **S-1312B11-A4T2U3** demonstrate closely aligned pricing across various packaging options, such as Tape & Reel or Cut Tape. These findings highlight opportunities for strategic bundling and refined pricing strategies to leverage perceived value in closely priced products. Additionally, the granular pricing tiers enable targeted adjustments for different purchase scales, ensuring competitive positioning and optimized market penetration within this high-demand category.

**Step 3: Business Recommendations:**

1.**Premium Product Analysis by Manufacturer**

Focus on strengthening partnerships with manufacturers like ABLIC Inc., who dominate premium product segments. Collaborate to co-develop high-value products and invest in targeted marketing strategies to capitalize on their strong market position in categories like “DC DC Switching Controllers” and “Battery Management.”

2.**Stockout Risk Analysis**

Implement predictive inventory management systems to monitor stock levels and lead times proactively. For high-risk items like TVS Diodes, establish safety stock thresholds and explore supplier diversification to mitigate supply chain disruptions caused by extended lead times.

3.**Supplier Performance Based on Lead Time**

Negotiate improved lead time terms with suppliers like YAGEO and Würth Elektronik. Leverage data-driven forecasting models to anticipate delays and evaluate alternative suppliers to reduce dependence on underperforming partners and improve supply chain efficiency.

4.**Feature Analysis Across Product Categories**

Prioritize incorporating high-demand features, such as “Supervisors” and “Voltage Regulators - Linear Low Dropout,” into new product designs. Use these insights for feature-specific marketing and development strategies to align with customer needs and drive sales.

5.**Pricing Analysis for Manufacturer Series**

Develop competitive pricing strategies based on premium offerings like OSLON® SSL 80 by ams-OSRAM. Benchmark against similar high-value series to refine pricing structures and enhance customer perception of premium products.

6.**Identifying Market Leaders by Product Category**

Collaborate with dominant players like ABLIC Inc. and KEMET to enhance product offerings in categories where they lead. Use this insight to establish strategic alliances and optimize category-specific resource allocation for maximum market impact.

7.**Inventory Health Assessment**

Use inventory classification to reduce overstock costs for categories like “Through Hole Resistors” and mitigate risks in understocked categories like “TVS Diodes.” Implement data-driven inventory balancing strategies to maintain optimal stock levels across all categories.

8.**Revenue Analysis by Manufacturer**

Prioritize partnerships with high-revenue manufacturers such as Meritek and RFMi to ensure consistent product availability and maximize revenue growth. Develop tailored support strategies to strengthen these key partnerships.

9.**Revenue Analysis by Product Category**

Focus investments on high-revenue product categories such as “Through Hole Resistors” and “SAW Filters.” Allocate marketing and R&D resources strategically to expand these profitable categories while addressing gaps in lower-performing segments.

10.**Pricing Similarities for Competitive Analysis**

Leverage pricing overlaps to create bundled offerings and introduce dynamic pricing strategies. Target cost-sensitive customers by highlighting value in closely priced product pairs to increase market competitiveness and boost sales.

11.**Pricing Similarities in Product Categories**

Explore bundling and tiered pricing models for closely priced products in high-demand categories like “Voltage Regulators - Linear, Low Drop Out.” Tailor pricing adjustments for large-volume orders to attract diverse customer segments and increase market penetration.

12. **Optimizing Pricing and Inventory Strategies for Low-Performing Product Categories**

The company should focus on identifying and addressing the underlying issues within low-performing product categories, as indicated by low total inventory value. By analyzing factors like demand, pricing inefficiencies, and stock levels, the company can optimize pricing strategies to increase competitiveness and profitability.

13**. Leveraging Premium Product Insights for Market Positioning and Revenue Growth**

The company should leverage these insights by highlighting premium products like the 4095-CR-M4FDMPRINTER-ND in marketing campaigns to attract high-value customers and strengthen its position in the premium market segment. It can also use this data to refine pricing strategies, ensuring competitiveness while maintaining profitability. Additionally, the company should focus on building strong supplier relationships for these high-value products to ensure consistent availability and capitalize on their potential for driving revenue growth within the “3D Printers” category.

***Conclusion:***

The analysis of Digi-Key’s product and manufacturer data provides critical insights into pricing, inventory, lead times, and revenue patterns. Key findings include identifying manufacturers excelling in premium markets, high-risk stockouts, and inefficiencies in lead times, with actionable recommendations to improve supply chain resilience and inventory management. The study highlights prevalent features across product categories, profitability trends, and opportunities for pricing optimization. Revenue analysis pinpoints top-performing manufacturers and categories, offering guidance for strategic partnerships and resource allocation. Overall, this comprehensive evaluation equips stakeholders with data-driven strategies to enhance operational efficiency, financial performance, and competitive market