**Business Insights**

**EDA-1 (SQL)**

|  |  |
| --- | --- |
| Date | the date of the transaction |
| FY | the fiscal year |
| Products | the name of the product sold |
| dia | the diameter of the product in millimeters |
| dia group | the diameter group of the product |
| grade | the grade of the product |
| type | the type of product |
| length | the length of the product in meters |
| Voucher Type | the type of voucher used for the transaction |
| Quantity | the quantity of the product sold in tons |
| Rate | the rate of the product in INR per ton |
| Value | the value of the product sold in INR |

The insights each mode can provide in the context of a business, possibly in the construction or manufacturing industry involving steel products:

* **ModeDate ('1/9/2020'):** The most frequent date is '1/9/2020'. This could represent a significant date in the business context, possibly marking an event, transaction, or a period of high activity. Analyzing the activities around this date could provide insights into patterns, trends, or seasonality in the business operations.
* **ModeFY ('FY 20'):** The mode fiscal year is 'FY 20'. This could be the most recent or relevant fiscal year in the business. It's crucial to analyze financial data and performance for this fiscal year to understand revenue, expenses, and overall business health during that period.
* **ModeProducts ('08MM TATA TISCON FE500D (S)'):** The most frequent product is '08MM TATA TISCON FE500D (S)'. This indicates a particular steel product that is in high demand or commonly used. Understanding why this product is popular can help focus marketing efforts and ensure sufficient supply.
* **ModeDia ('08 MM'):** The mode diameter is '08 MM'. This might signify a standard or popular diameter for steel rods. Identifying the reasons for its popularity could help streamline manufacturing processes and cater to customer preferences.
* **ModeDiaGroup ('12 MM - 32 MM'):** The most frequent diameter group is '12 MM - 32 MM'. This could represent a range of diameters commonly used for specific applications. Understanding the demand for this range could guide inventory management and sales strategies.
* **ModeGrade ('500D'):** The most frequent grade is '500D'. This suggests that '500D' grade steel rods are widely used or preferred by customers. Exploring why this grade is popular could inform decisions about sourcing and marketing.
* **ModeType ('FULL LENGTH'):** The most frequent product type is 'FULL LENGTH'. This might indicate that full-length steel rods are commonly requested or produced. Understanding customer preferences for rod lengths can influence production planning.
* **ModeLength ('12 METER'**): The most frequent length is '12 METER'. This could indicate a standard length for steel rods that's in high demand. Ensuring a consistent supply of this length could be important for customer satisfaction.
* **ModeVoucherType ('Sales A/c GST'):** The most frequent voucher type is 'Sales A/c GST'. This could be a common type of transaction, possibly representing standard sales. Analyzing the sales process for this voucher type could highlight any potential areas for improvement.
* **ModeQuantity ('2'):** The most frequent quantity sold is '2'. This suggests that selling units of 2 steel rods might be a common order size. Analyzing customer preferences for order quantities could optimize inventory management.
* **ModeRate ('44000'):** The most frequent rate is '44000'. This could be the common pricing for the product, possibly per ton or per unit. Understanding the pricing strategy for this rate could ensure competitiveness and profitability.
* **ModeValue ('126000'):** The most frequent sales value is '126000'. This could indicate that transactions around this value are common. Analyzing sales at this value could reveal customer segments or product combinations that drive revenue.

The business insights for each column based on the Std dev values:

* **Std dev** **Quantity:** The relatively lowStd dev in quantity of steel rods could indicate that the manufacturing or distribution process is relatively stable. There might be minimal fluctuations in the number of steel rods produced or shipped. This could be a positive sign of consistent production and supply chain management.
* **Std dev** **Rate (Rate of Steel Rods per Ton):** The high Std dev rate in the cost of steel rods per ton suggests that the rate of producing steel rods is fluctuating significantly. This could be due to fluctuations in the cost of raw materials, energy prices, or manufacturing processes. It's important for the business to closely monitor and analyse these cost variations to ensure profitability and competitiveness. Implementing cost-effective procurement strategies and exploring ways to optimize production processes might help manage this variance effectively.
* **Std dev** **Value (Sales Figure of Steel Rods):** The substantial Std dev value in the sales figure of steel rods indicates significant fluctuations in revenue generated from steel rod sales. This could be attributed to changes in market demand, competitive pricing, or shifts in customer preferences. It's essential to analyze the factors contributing to these sales variations and develop strategies to mitigate negative impacts. This could involve diversifying the customer base, offering value-added services, or exploring new market segments.

The business insights for each column based on the skewness values:

* **Skewness Quantity (2.2027269752405503):** The skewness value for the quantity column is positive (greater than 0), indicating a positively skewed distribution. In a business context, this could suggest that there are more data points with lower quantities and fewer data points with higher quantities. This might indicate that the majority of the steel rod quantities are clustered towards the lower end, with occasional instances of larger quantities. It could be useful to analyze the reasons behind this skewness, such as specific customer preferences, order patterns, or market demands.
* **Skewness Rate (0.7046215796373417):** The skewness value for the rate column is positive (greater than 0), indicating a positively skewed distribution. This suggests that there are more data points with lower rates and fewer data points with higher rates. In the context of cost per ton of steel rods, this skewness might imply that the majority of the costs are lower, with a few instances of higher costs. Investigating the reasons for this skewness could involve analyzing the cost components, production efficiencies, and external factors affecting costs.
* **Skewness Value (2.580217755611119):** The skewness value for the value column is positive (greater than 0), indicating a positively skewed distribution. This could mean that there are more data points with lower sales values and fewer data points with higher sales values. In the context of sales figures for steel rods, this skewness might reflect that most of the sales figures are on the lower side, with occasional higher sales values. Understanding the factors driving this skewness could involve analyzing customer segments, marketing strategies, and market dynamics.

The business insights for each column based on the kurtosis values:

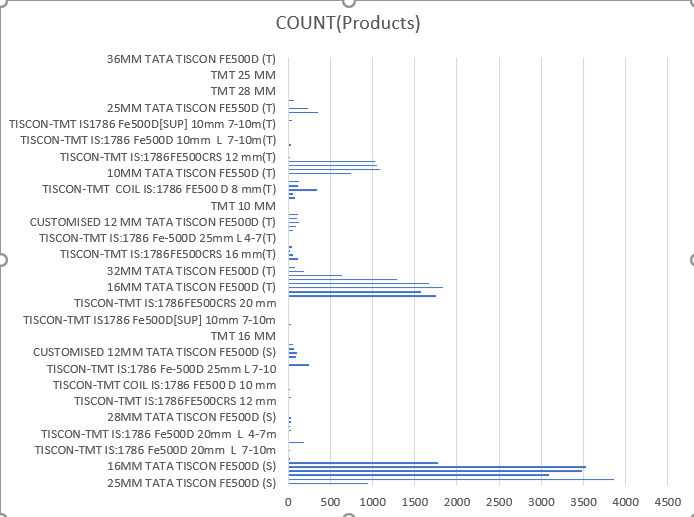
* **Kurtosis Quantity (5.002034936252743):** The positive kurtosis value indicates a leptokurtic distribution. In a business context, this suggests that the quantity distribution of steel rods has heavier tails and more extreme values compared to a normal distribution. This might mean that there are occasional instances of significantly high or low quantities of steel rods. Understanding the reasons behind such extreme values could involve investigating production anomalies, order patterns, or unique customer demands.
* **Kurtosis Rate (-0.08063364352235203):** The negative kurtosis value indicates a platykurtic distribution. This means that the rate distribution (cost per ton) of steel rods has lighter tails and fewer extreme values compared to a normal distribution. In a business context, this could imply that the cost rates are relatively stable, with fewer instances of extremely high or low costs. This stability might be a positive sign, but it's important to ensure that the costs are competitive and aligned with market conditions.
* **Kurtosis Value (8.177512246847437):** The high positive kurtosis value suggests a very leptokurtic distribution for the value column (sales figures of steel rods). This implies that the sales value distribution has very heavy tails, indicating occasional instances of extremely high or low sales values. In a business context, this could indicate some unique sales opportunities or challenges that result in significant deviations from the average. Understanding the drivers of such extreme values is crucial for effective sales planning and resource allocation.

**Univariate Analysis**

1. Count the no. of occurences of each product.

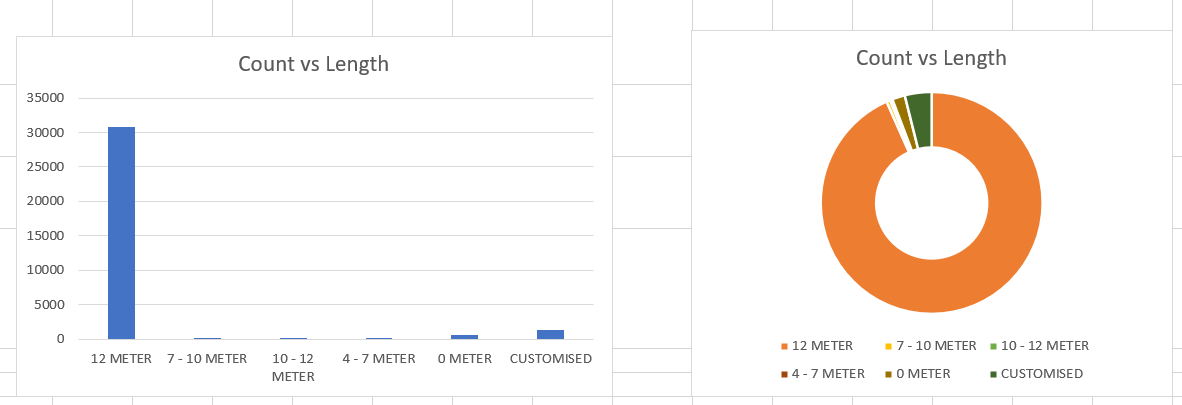
SELECT Products, COUNT(Products) FROM steel\_data

GROUP BY Products;



2. Most frequently ordered length

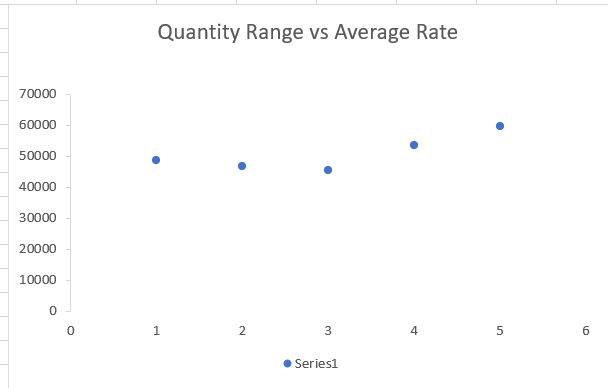
SELECT length, COUNT(length) FROM steel\_data

GROUP BY length;

**Bivariate Analysis** (Numeric - Numeric)

1. How does the rate vary as the quantity of product changes?

|  |  |
| --- | --- |
| Quantity\_Range (in tons) | Average\_Rate |
| 0-10 | 48781.35 |
| 10 20 | 46847.18 |
| 20-30 | 45327.98 |
| 30-40 | 53590.83 |
| 40-50 | 59750 |



**Data Pre-processing-1**

Checking for Duplicates

* A total of 66 duplicate rows have been detected within the dataset. It's important to acknowledge that appropriate actions will be taken to address these duplications.

Treating Outliers

* During the dataset examination, the presence of outliers in certain columns was detected, necessitating prompt action. These outliers were effectively addressed through data standardization methods. Moving forward, a vigilant monitoring strategy will be employed to identify any future outliers, with a commitment to applying requisite treatments as needed. This approach is crucial to uphold the dataset's integrity, supporting well-informed business decisions.

**EDA-2(SQL)**

The business insights for each column based on the std dev values:

* **std dev** **Quantity:** The reduced std dev in quantity of steel rods still suggests a relatively stable manufacturing or distribution process. The smaller variance indicates that the fluctuations in the number of steel rods produced or shipped are even lower. This strengthened stability in production and supply chain management continues to be a positive indicator, potentially leading to streamlined operations and effective resource allocation.
* **Updated std dev Rate with Lower Value:** The increased but still substantial std dev rate in the cost of steel rods per ton continues to suggest significant fluctuations in production costs. This might be indicative of challenges related to raw material sourcing, energy price volatility, or manufacturing inefficiencies. The business should maintain a close watch on cost trends, implementing robust cost-control measures, negotiating better supplier contracts, and investing in process optimization. Such efforts could help maintain profitability and enhance the business's competitive position.
* **Updated std dev** **Value with Lower Value:** The reduced std dev value in the sales figure of steel rods still indicates notable fluctuations in revenue from steel rod sales. This could be due to varying market conditions, dynamic pricing strategies, or changing customer preferences. To address this, the business should conduct comprehensive market research, develop flexible pricing models, and create tailored marketing campaigns for different customer segments. By understanding and adapting to market dynamics, the business can better manage revenue variability and foster sustained growth.

The business insights for each column based on the skewness values:

1. **Skewness Quantity:**

Positive skewness suggests more lower quantities and fewer higher quantities of steel rods. This might indicate customer preferences or market demands, with most quantities being lower but some larger orders occurring occasionally.

2. **Skewness Rate:**

Positive skewness implies more lower cost rates and fewer higher rates per ton. This likely signifies that most cost rates are lower, but some higher costs may be due to specific factors affecting production efficiency or external influences.

3. **Skewness Value:**

Positive skewness indicates more lower sales values and fewer higher values. This could highlight that most sales figures are lower, while occasional higher sales could be influenced by certain customer segments or marketing efforts.

The business insights for each column based on the Kurtosis values:

* **Updated Kurtosis Quantity:** Leptokurtic distribution still indicates occasional extreme quantity values. Investigating these outliers can reveal opportunities or challenges in production and demand.
* **Updated Kurtosis Value:** Leptokurtic distribution continues to suggest occasional extreme sales values. Understanding these occurrences guides marketing and resource decisions. This hints at occasional instances of significantly high or low sales values. These occurrences could be attributed to unique market trends, seasonality, or specific customer needs.
* **Updated Kurtosis Rate:** Platykurtic distribution indicates even more stable cost rates. Maintaining competitiveness remains essential while benefiting from cost stability.

**Data Pre-processing-2 (SQL)**

Checking for Duplicates

* All duplicate rows have been successfully removed from the dataset.

Treating Outliers

* In the process of dataset examination, no instances necessitating outlier treatment were discovered. Moving ahead, consistent monitoring of the data will be upheld, and any identified outliers will be promptly subjected to the required corrective actions.

--------------------------------------------------------------------------------------------------------------------------

**EDA-1 (Python)**

The business insights for each column based on the std dev values:

* **Std dev** **Quantity:** The relatively low std dev in quantity of steel rods could indicate that the manufacturing or distribution process is relatively stable. There might be minimal fluctuations in the number of steel rods produced or shipped. This could be a positive sign of consistent production and supply chain management.
* **Std dev** **Rate (Rate of Steel Rods per Ton):** The high std dev rate in the cost of steel rods per ton suggests that the rate of producing steel rods is fluctuating significantly. This could be due to fluctuations in the cost of raw materials, energy prices, or manufacturing processes. It's important for the business to closely monitor and analyse these cost variations to ensure profitability and competitiveness. Implementing cost-effective procurement strategies and exploring ways to optimize production processes might help manage this variance effectively.
* **Std dev** **Value (Sales Figure of Steel Rods):** The substantial std dev value in the sales figure of steel rods indicates significant fluctuations in revenue generated from steel rod sales. This could be attributed to changes in market demand, competitive pricing, or shifts in customer preferences. It's essential to analyze the factors contributing to these sales variations and develop strategies to mitigate negative impacts. This could involve diversifying the customer base, offering value-added services, or exploring new market segments.

The business insights for each column based on the skewness values:

* **Skewness Quantity (2.2027269752405503):** The skewness value for the quantity column is positive (greater than 0), indicating a positively skewed distribution. In a business context, this could suggest that there are more data points with lower quantities and fewer data points with higher quantities. This might indicate that the majority of the steel rod quantities are clustered towards the lower end, with occasional instances of larger quantities. It could be useful to analyze the reasons behind this skewness, such as specific customer preferences, order patterns, or market demands.
* **Skewness Rate (0.7046215796373417):** The skewness value for the rate column is positive (greater than 0), indicating a positively skewed distribution. This suggests that there are more data points with lower rates and fewer data points with higher rates. In the context of cost per ton of steel rods, this skewness might imply that the majority of the costs are lower, with a few instances of higher costs. Investigating the reasons for this skewness could involve analyzing the cost components, production efficiencies, and external factors affecting costs.
* **Skewness Value (2.580217755611119):** The skewness value for the value column is positive (greater than 0), indicating a positively skewed distribution. This could mean that there are more data points with lower sales values and fewer data points with higher sales values. In the context of sales figures for steel rods, this skewness might reflect that most of the sales figures are on the lower side, with occasional higher sales values. Understanding the factors driving this skewness could involve analyzing customer segments, marketing strategies, and market dynamics.

The business insights for each column based on the kurtosis values:

* **Kurtosis Quantity:** The positive kurtosis value indicates a leptokurtic distribution. In a business context, this suggests that the quantity distribution of steel rods has heavier tails and more extreme values compared to a normal distribution. This might mean that there are occasional instances of significantly high or low quantities of steel rods. Understanding the reasons behind such extreme values could involve investigating production anomalies, order patterns, or unique customer demands.
* **Kurtosis Rate:** The negative kurtosis value indicates a platykurtic distribution. This means that the rate distribution (cost per ton) of steel rods has lighter tails and fewer extreme values compared to a normal distribution. In a business context, this could imply that the cost rates are relatively stable, with fewer instances of extremely high or low costs. This stability might be a positive sign, but it's important to ensure that the costs are competitive and aligned with market conditions.
* **Kurtosis Value:** The high positive kurtosis value suggests a very leptokurtic distribution for the value column (sales figures of steel rods). This implies that the sales value distribution has very heavy tails, indicating occasional instances of extremely high or low sales values. In a business context, this could indicate some unique sales opportunities or challenges that result in significant deviations from the average. Understanding the drivers of such extreme values is crucial for effective sales planning and resource allocation.

**Data Pre-processing-1 (Python)**

Checking for Duplicates

* A total of 66 duplicate rows have been detected within the dataset. It's important to acknowledge that appropriate actions will be taken to address these duplications.

Treating Outliers

* During the dataset examination, the presence of outliers in certain columns was detected, necessitating prompt action. These outliers were effectively addressed through Winsorization method. Moving forward, a vigilant monitoring strategy will be employed to identify any future outliers, with a commitment to applying requisite treatments as needed. This approach is crucial to uphold the dataset's integrity, supporting well-informed business decisions.

**EDA-2 (Python)**

The business insights for each column based on the std dev values:

* **Std dev** **Quantity:** The lower std dev in steel rod quantities suggests a stable manufacturing and distribution process. This stability leads to streamlined operations and efficient resource allocation.
* **Std dev Rate (Cost):** The reduction in std dev for the cost of steel rods per ton signals an improvement in cost consistency. This could reflect better control over production expenses or increased stability in raw material sourcing. The business should continue monitoring cost trends to ensure sustainable profitability and operational efficiency.
* **Std dev Value (Sales):** The reduced std dev in sales revenue indicates improved stability. Adapting to market dynamics, flexible pricing, and targeted marketing efforts can manage revenue fluctuations and drive growth.

The business insights for each column based on the reduced kurtosis values after pre-processing:

* **Kurtosis Quantity:** The decreased kurtosis value for steel rod quantities indicates a shift towards a more moderate distribution. This suggests that extreme deviations from the average quantity are less frequent. This enhanced stability can lead to smoother production planning and optimized inventory management.
* **Kurtosis Rate (Cost):** With a lower kurtosis value for the cost of steel rods per ton, the distribution is becoming less heavy-tailed. This could imply reduced occurrences of extreme cost fluctuations. The business can benefit from improved cost predictability and better budgeting.
* **Kurtosis Value (Sales):** The decreased kurtosis value for sales revenue suggests a move towards a more normal distribution. This indicates a reduction in extreme revenue variations. A smoother revenue pattern can enable more accurate financial planning and enhance the business's overall financial stability.

The business insights for each column based on the reduced skewness values:

* **Skewness Quantity:** Reduced skewness suggests improved balance in steel rod quantities, enhancing inventory management and resource allocation.
* **Skewness Rate (Cost):** Decreased skewness implies smoother cost patterns, aiding accurate cost estimation and financial planning.
* **Skewness Value (Sales):** Lowered skewness signifies more stable revenue patterns, leading to better financial forecasting and management.

**Data Pre-processing-2 (Python)**

Checking for Duplicates

* All duplicate rows have been successfully removed from the dataset.

Treating Outliers

* In the process of dataset examination, no instances necessitating outlier treatment were discovered. Moving ahead, consistent monitoring of the data will be upheld, and any identified outliers will be promptly subjected to the required corrective actions.