Here's a brief overview of some of the columns included in the dataset:

- **WorkOrder**: Identifier for the work order.
- **Item No.**: Unique number assigned to the item.
- **Item**: Description of the item.
- **Shipment**: Shipment method or status.
- **DateClosed**: The date the work order was closed.
- **Order Qty_x** and **Completed Qty**: Quantity ordered and the quantity that was completed.
- **FOH (Factory or Fixed Overhead?)** and **VOH (Variable Overhead)**: Costs associated with production, possibly factory and variable overheads.
- **ActualMaterialCost**: The actual cost of materials used.
- **Labor Hours**, **Setup Hours**: Hours spent on labor and setup.
- **Machines**: The machines used for the work order.
- **Cycle Time**, **Predicted Hours**, **remainHours**: Metrics related to the time spent or predicted for the production.
- **Employee**: The employees involved in the work order.
- **WorkOrderID**: Another identifier for the work order, which seems redundant with the WorkOrder column.

Given the diverse range of data, we could conduct several analyses, including productivity assessments, cost analysis, time efficiency evaluation, and optimization insights for future work orders.

For a labor trend analysis, we generally need data on labor hours, dates, and potentially the number of workers or specific tasks completed over time. Looking at the provided dataset, there are a few relevant columns that could be used for such an analysis:

- **DateClosed**: This could help us understand when tasks were completed.
- **Labor Hours**, **Setup Hours**: These columns give us insight into the amount of labor time invested in tasks.
- **Employee**: Knowing who worked on what could allow us to dive deeper into labor efficiency, distribution of work, and potentially identify trends in productivity based on individual or team performance.
- **WorkOrder** or **WorkOrderID**: These identifiers can help us track labor trends on a per-work-order basis, which could be insightful for understanding productivity trends across different types of tasks.

To perform a basic labor trend analysis, we could look at total labor hours over time to identify any trends in labor usage. For instance, we could analyze whether there's an increase in labor efficiency (decreasing labor hours for the same amount of work) or identify certain periods with high labor demands. We can also analyze labor distribution among employees over time if the data allows.

We will start with a simple analysis to visualize total labor hours (combining Labor Hours and Setup Hours) over time.

The dataset seems to be relatively clean but does require some attention before proceeding with in-depth analyses, especially for labor trend analysis. Here's a summary of the findings:

- **Missing Values**: There are missing values in several columns. Notably, 'Shipment' has 164 missing entries, and 'Cycle Time', 'Predicted Hours', and 'remainHours' each have 27 missing entries. The 'Item' column has 1 missing entry, and a few other columns have a small number of missing entries as well.
- **Duplicate Rows**: There are no duplicate rows in the dataset, which is a good sign for the integrity of the data.
- **Data Types**: Some columns that you would expect to be numeric (e.g., 'Cycle Time', 'Predicted Hours', 'remainHours') are listed as object types, which often indicates they contain non-numeric data or mixed types. This needs correction for any analysis involving these metrics.

Given these observations, the next steps should include:

1. **Handling Missing Values**: Depending on the analysis, you might fill in missing values, drop them, or impute them based on other data points. For labor trend analysis, particularly missing 'Cycle Time', 'Predicted Hours', and 'remainHours' might be critical depending on what aspect of labor you're focusing on.
2. **Correcting Data Types**: Converting columns to the appropriate data types is necessary for any calculations or modeling. For instance, 'Cycle Time', 'Predicted Hours', and 'remainHours' should likely be converted to numeric types if they're intended to be used in the analysis.
3. **Data Cleaning Specific Columns**: The presence of non-numeric characters or inconsistencies within columns that should be numeric needs to be addressed. This involves cleaning up the 'MaterialPerUnit', 'LaborPerUnit', 'OHPerUnit', 'CostPerUnit', 'Order Qty_y', 'Hours in date range', 'Up to date hours', 'Cycle Time', 'Predicted Hours', and 'remainHours' columns, as they are currently object types likely due to mixed data entries.

Given the goal of performing a labor trend analysis with the current dataset, the best approach to proceed involves several key steps tailored to both cleaning the data and preparing it for analysis. This approach ensures the dataset is in the right shape for extracting meaningful insights about labor trends. Here's a structured plan:

1. Handling Missing Values

- **Critical Columns**: For labor trend analysis, focus on columns directly related to labor, such as 'Cycle Time', 'Predicted Hours', 'remainHours', and 'Employee'. Decide whether to fill missing values (e.g., with the median or mean, if applicable) or drop them based on the proportion of missing data and its potential impact on analysis.
- **Other Columns**: For columns like 'Shipment' or 'Program' with missing values that are less critical to labor trends, consider whether these columns will be used in the analysis. If not, they might be dropped or left as is, depending on the analysis needs.

2. Correcting Data Types

- Convert columns that should be numeric but are currently recognized as objects due to mixed types or formatting issues ('Cycle Time', 'Predicted Hours', 'remainHours', etc.) to numeric types. This often involves cleaning the data to remove any non-numeric characters and then converting the column.

3. Data Cleaning for Analysis

- **Uniformity**: Ensure all labor-related columns are consistently formatted and represent the data accurately, enabling straightforward aggregation and analysis.
- **Date Handling**: Convert the 'DateClosed' column to a datetime format to facilitate time series analysis and trend observation.
- **Aggregate Labor Hours**: Consider creating a new column that combines all relevant labor hours ('Labor Hours', 'Setup Hours') for a comprehensive view of total labor invested per work order.

4. Exploratory Data Analysis (EDA)

- Perform an initial EDA to understand the distribution of labor hours, identify outliers, and understand the temporal distribution of work orders. This can help in refining the analysis approach.

5. Labor Trend Analysis

- **Time Series Analysis**: Analyze labor hours over time to identify trends, patterns, or seasonal variations.
- **Employee Productivity**: Assess labor distribution among employees over time, which could highlight efficiency, training needs, or workload distribution issues.
- **Efficiency Trends**: Look for trends in labor efficiency by comparing labor hours against completed quantities over time.

6. Visualization

- Use graphs and charts to visualize labor trends, making it easier to communicate findings and insights. Time series plots, histograms, and box plots can be particularly useful here.

7. Insights and Recommendations

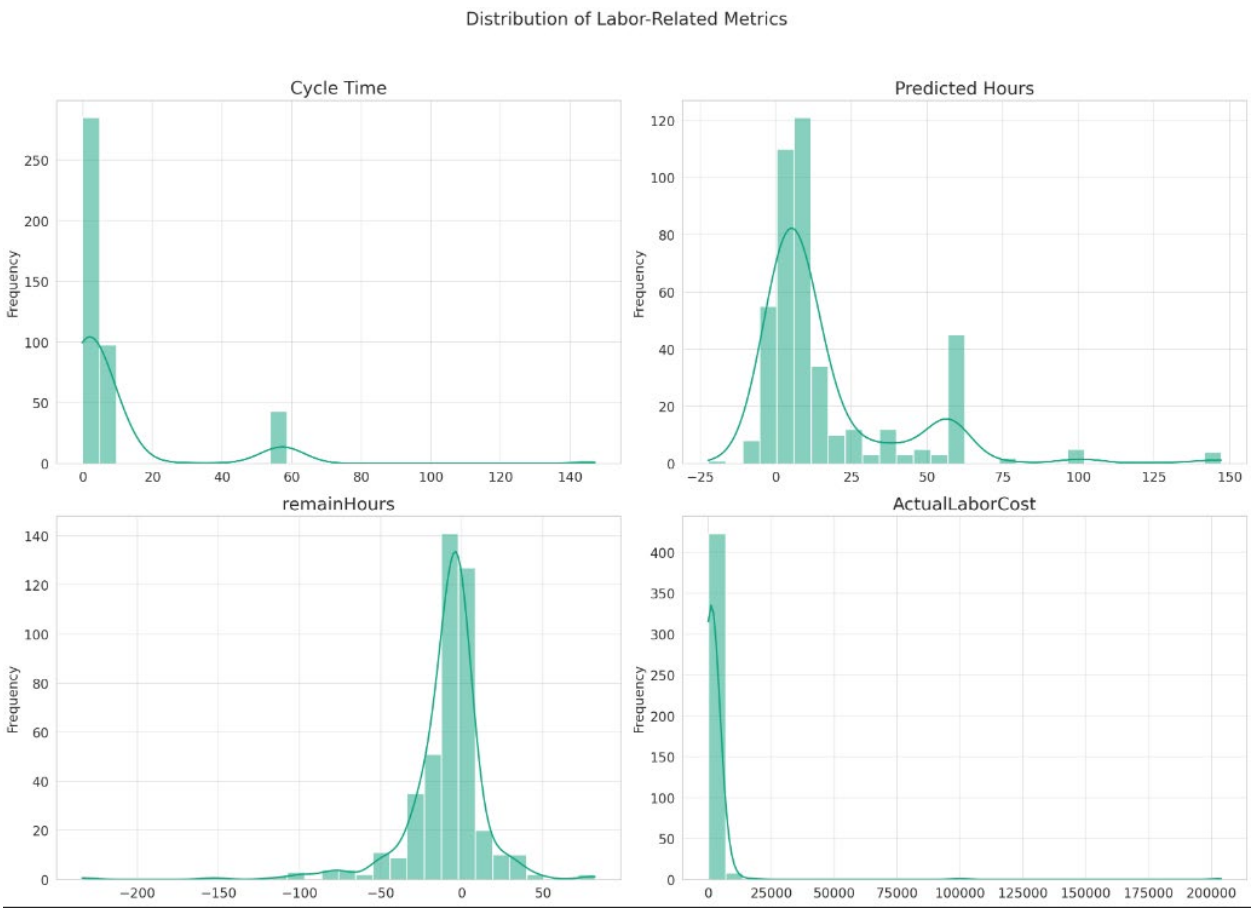
- Based on the analysis, draw insights about labor trends, identify potential areas for improvement, and make actionable recommendations.

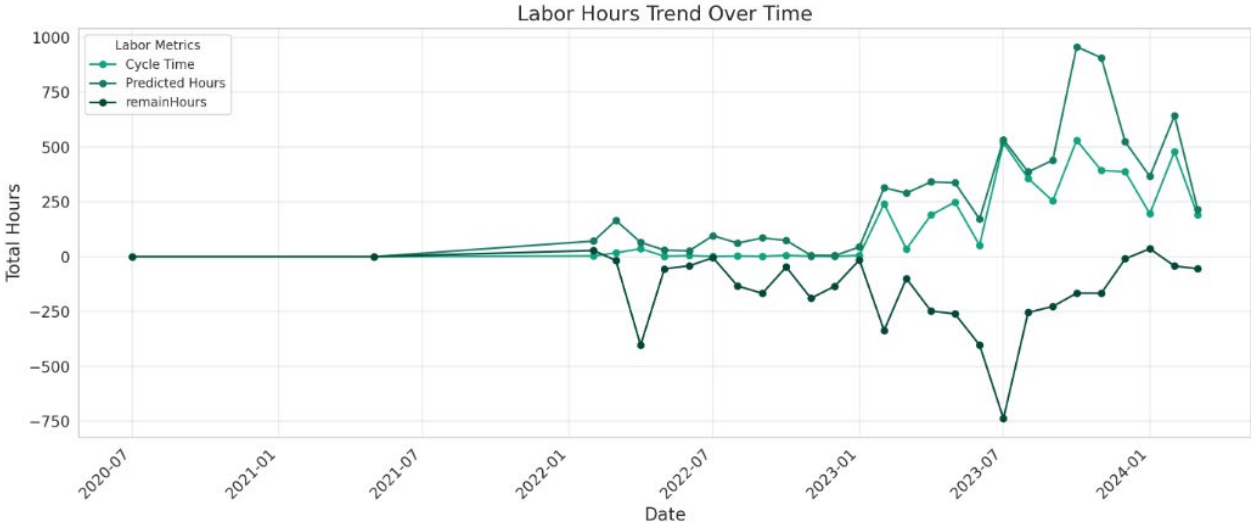
The data cleaning process has successfully addressed the immediate concerns:

1. **Missing Values**:
    - Missing values in 'Cycle Time', 'Predicted Hours', and 'remainHours' have been filled with 0, assuming no time was recorded or spent where data was missing. This is a simplification and may need to be revisited with a deeper understanding of the context.
    - Missing values in 'Shipment' and 'Program' have been filled with 'Unknown' to maintain the integrity of these categorical fields.
    - There remains a single missing value in the 'Item' column and a couple in 'Routing Number', which may require specific domain knowledge to address appropriately.
2. **Data Types**:
    - The 'DateClosed' column has been successfully converted to the datetime format, facilitating any time-series analysis or temporal observations.
    - 'Cycle Time', 'Predicted Hours', and 'remainHours' have been converted to numeric types (float64), allowing for numerical analysis and operations.

The dataset is now in a better shape for conducting labor trend analysis. The next steps would involve exploratory data analysis (EDA) to understand the distribution and trends within the labor-related data, followed by more detailed analysis focused on identifying labor trends over time and across different variables.

The exploratory data analysis (EDA) provides several insights into the labor-related metrics of the dataset:





Distribution of Labor-Related Metrics:

- The histograms for 'Cycle Time', 'Predicted Hours', and 'remainHours' show a wide range of values with most data concentrated at the lower end of the scale, indicating that many tasks have relatively short cycle and predicted times but there are outliers with significantly

higher values. This skewness suggests a few work orders are much more labor-intensive than others.

- The 'ActualLaborCost' distribution also indicates a heavy right skew, with most of the labor costs being relatively low, but a few cases having very high labor costs. This could point towards specific work orders or projects that are significantly more resource-intensive.

Labor Hours Trend Over Time:

- The trend over time for 'Cycle Time', 'Predicted Hours', and 'remainHours' shows variations in labor metrics across different months. This could indicate fluctuations in workload, project complexity, or efficiency changes over time.

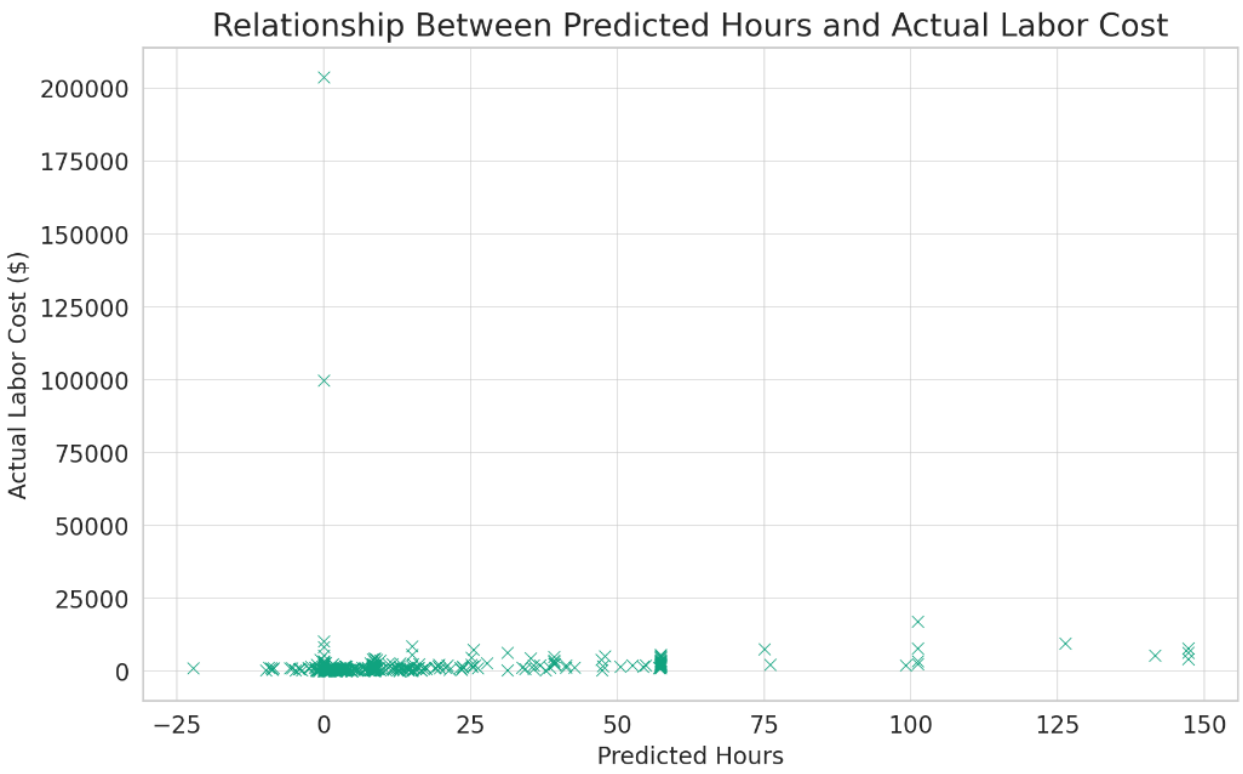
Summary Statistics:

- **Cycle Time**: The average cycle time is approximately 9.58 hours, but with a high standard deviation, indicating significant variability among work orders.
- **Predicted Hours**: On average, the predicted hours for tasks are around 16.49 hours, also with a considerable spread, highlighting the diversity in work order complexities or durations.
- **Remain Hours**: The negative mean value suggests that on average, tasks tend to take less time than predicted, or it could also reflect data recording practices.
- **Actual Labor Cost**: The mean actual labor cost is quite high, but the extremely high standard deviation (max value at 203,835.68) points to a few work orders with exceptionally high labor costs, possibly skewing the average.

These observations suggest there are significant variations in labor requirements and costs across different work orders and over time. The presence of outliers in both labor time and cost metrics might warrant further investigation to understand the causes behind these extremes—whether they're due to specific project requirements, efficiency issues, or data recording practices.

Given these insights, the next steps could involve a deeper dive into:

- The relationship between predicted and actual labor metrics.
- Identifying specific work orders or time periods that contribute most to the observed variability and outliers.
- Examining factors that may influence labor cost and efficiency, such as the type of work, machinery used, or the employees involved.



Analysis Insights

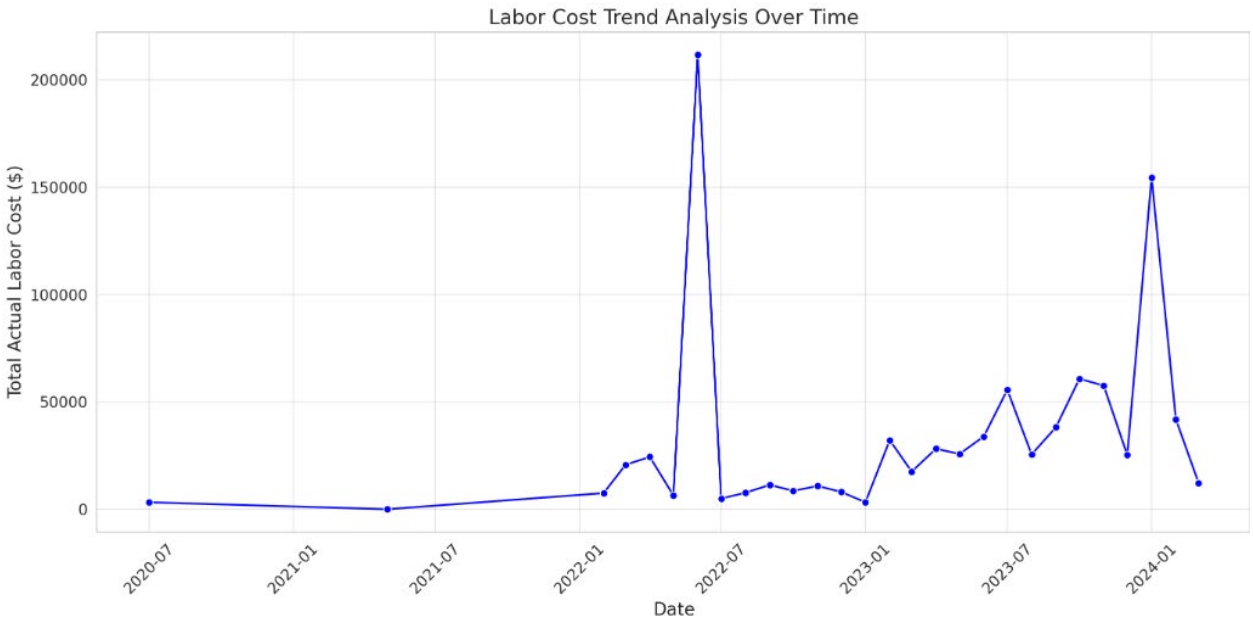
1. **Relationship Between Predicted Hours and Actual Labor Cost**:
    - The scatter plot reveals no clear linear relationship between predicted hours and actual labor cost, suggesting that the cost isn't solely determined by the hours predicted for a task. Factors beyond just the predicted time, such as the complexity of the work, material costs, or unexpected issues, may significantly influence the actual labor cost.
2. **Work Orders with Highest Labor Costs**:
    - The analysis highlights specific work orders that have exceptionally high labor costs, with the top work order (WorkOrder ID: 5966) showing an actual labor cost of $203,835.68, followed by WorkOrder ID: 12785 with $99,903.04. These work orders significantly deviate from the norm and could be outliers due to their unique requirements or challenges.
    - The dates associated with these high-cost work orders range from mid-2022 to early 2024, indicating that such outliers are not confined to a specific time period.
3. **Factors Influencing Labor Cost and Efficiency**:
    - The correlation matrix indicates a modest positive correlation between 'Predicted Hours' and 'Cycle Time' with the 'Actual Labor Cost', albeit the correlation coefficients are relatively low. This suggests only a weak relationship between these

time metrics and the actual labor costs, again implying that other factors play a significant role in determining the cost.

- Interestingly, 'remainHours' shows a slight negative correlation with 'Actual Labor Cost', indicating that tasks with negative remaining hours (i.e., completed faster than predicted) do not necessarily correlate with higher labor costs.

Conclusions and Next Steps

- The absence of a strong linear correlation between predicted hours and actual labor costs, along with the presence of significant outliers in labor costs, underscores the complexity of accurately predicting labor costs based solely on predicted hours. This complexity could stem from varying task difficulties, inefficiencies, or unanticipated complications during execution.
- The highlighted outliers in labor costs warrant further investigation to understand the underlying causes—whether they are due to project-specific challenges, estimation inaccuracies, or other factors.
- A deeper dive into the qualitative aspects of work orders (e.g., the type of task, machinery used, or employee skill levels) could provide more insights into the factors driving labor costs and efficiency. This may involve analyzing textual data in the dataset or integrating additional data sources for a more comprehensive understanding.

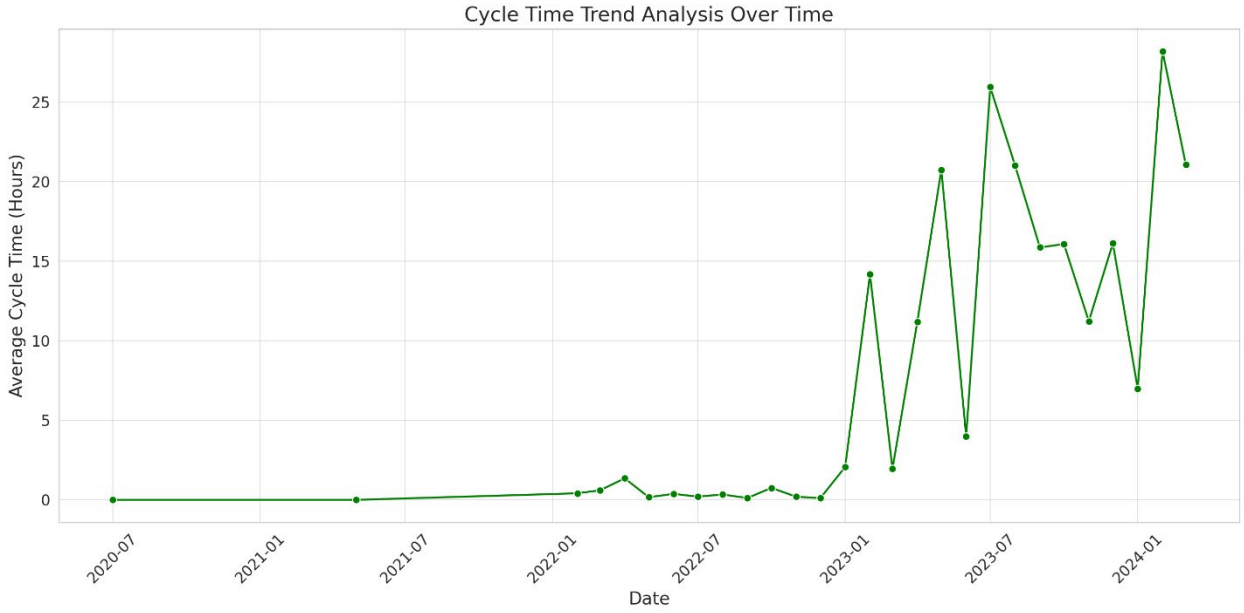


The labor cost trend analysis over time reveals the fluctuation in total actual labor costs on a monthly basis. This visualization highlights several key observations:

- **Variability**: There is noticeable variability in labor costs from month to month. Some months show significant spikes in labor costs, which could correspond to periods of high work volume, more complex projects, or perhaps inefficiencies in labor management.

- **Trend**: While there's variability, it's crucial to look for any underlying trends. Depending on the range of data available and the specific months highlighted, you might identify periods of increasing or decreasing labor costs that could indicate broader trends in efficiency, productivity, or business activity.
- **Outliers**: The spikes in labor costs in certain months may be considered outliers. These could be driven by large, labor-intensive projects or other exceptional circumstances. Identifying and understanding these outliers can provide valuable insights into labor cost management and project planning.

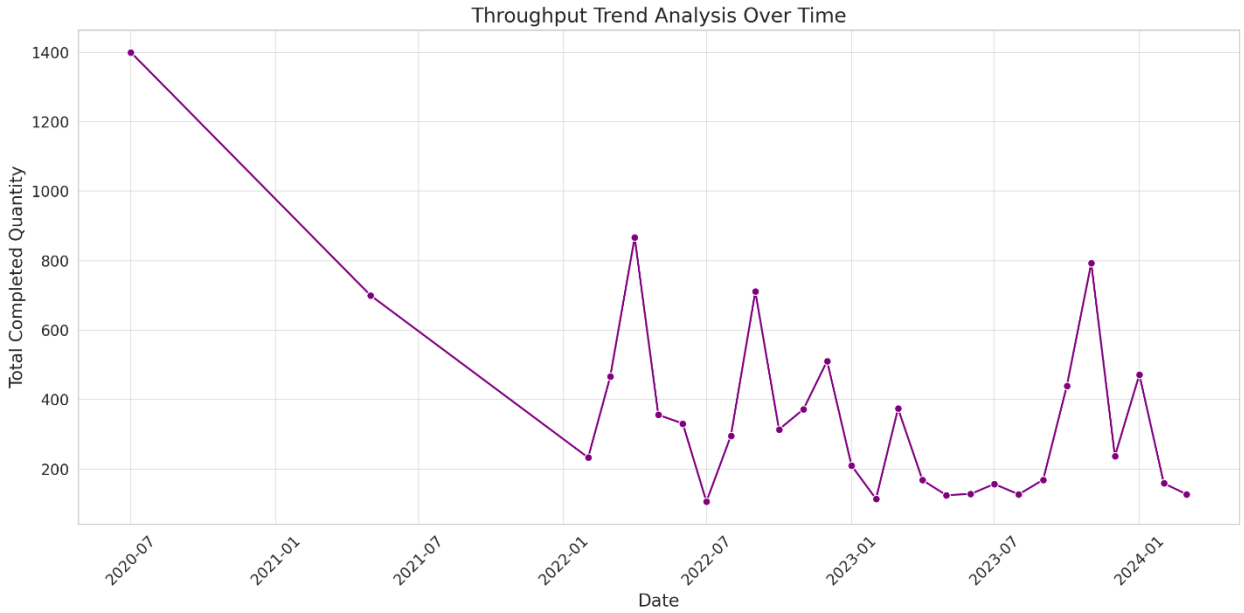
This trend analysis is a foundational step in understanding how labor costs evolve over time within the organization. Further analysis could segment these costs by different departments, types of work, or other relevant categories to gain deeper insights into what drives changes in labor costs.



The cycle time trend analysis over time provides insights into the average time spent on tasks each month. From the plot, several key observations can be made:

- **Fluctuations**: Similar to labor costs, cycle times exhibit fluctuations across different months. These variations could reflect changes in the nature of work being performed, operational efficiencies, or the complexity of tasks undertaken in those periods.
- **Trend Identification**: While the data points fluctuate, identifying a clear long-term trend (e.g., increasing or decreasing cycle times) would require examining the data within a broader context or over a longer time span. A trend in cycle times can indicate changes in operational efficiency, process improvements, or alterations in the type of work being performed.
- **Seasonality and Patterns**: Any recurring patterns that align with specific times of the year could suggest seasonality in work volume or types of tasks. Understanding such patterns can aid in planning and resource allocation to manage cycle times effectively.

This analysis sheds light on how the time required to complete tasks evolves and can highlight areas for further investigation, such as periods with unusually long or short cycle times. Investigating the causes behind these trends can provide actionable insights into improving operational efficiency and task management.



The throughput trend analysis over time visualizes the total completed quantity of work orders each month, offering insights into the productivity and operational capacity of the organization. Key observations from the plot include:

- **Variability in Throughput**: The graph shows fluctuations in the total completed quantity from month to month, indicating changes in production output. These fluctuations might be due to various factors, such as changes in demand, capacity constraints, or operational challenges.
- **Identifying Peaks and Troughs**: Specific months show higher throughput, which could be indicative of peak production periods, possibly driven by high demand, seasonal factors, or special projects. Conversely, months with lower throughput might signal reduced demand, operational issues, or periods of maintenance and downtime.
- **Long-term Trends**: While monthly fluctuations are evident, identifying a long-term trend (e.g., an overall increase or decrease in throughput) would require examining the data within the context of the organization's operational changes, market demand, and capacity enhancements over time.

This analysis is crucial for understanding the operational dynamics and can help in identifying bottlenecks, planning for capacity adjustments, and improving overall operational efficiency. Further investigation into the causes behind significant peaks or troughs in throughput can provide valuable insights for decision-making and strategic planning.