

The Effects of Volume Bid Contracts on Service Level Agreements

Introduction

This project assesses the impact of implementing volume bid contracts on service-level agreements governing the maintenance of data center equipment. Specifically, this study investigates whether the use of volume bid contracts negatively influences the average equipment repair times in data centers as compared to traditional contracting approaches.

Table of Contents

- Introduction
- Section 1 - Set Environment
- Section 2 - Load and Process Data
- Section 3 - Visual Analysis
- Section 4 - Outlier Analysis
- Section 5 - Data Distribution Analysis
- Section 6 - Data Transformation Analysis
- Section 7 - Independent T-Test Analysis
- Section 8 - Mann-Whitney U Test Analysis
- Section 9 - Parametric and Nonparametric Analysis
- Section 10 - Chi-Square Analysis
- Conclusion

Section 1 - Set Environment

Purpose

The purpose of this section is to load the necessary Python packages, establish a path to the application directory, and initialize the Analysis Main class object.

Section Contents

1. Import Packages and Libraries
2. Set Path to Python Files
3. Create the Analysis Main Object

Set Environment 1 - Import Packages and Libraries

```
In [1]: import os  
import sys
```

Set Environment 2 - Set Path to Python Files

```
In [2]: module_path = os.path.abspath(os.path.join('..', 'Python'))  
if module_path not in sys.path: sys.path.append(module_path)
```

Set Environment 3 - Create the Analysis Main Object

```
In [3]: from data_analysis_main import AnalysisMain  
am = AnalysisMain()
```

Section 2 - Load and Process Data

Purpose

This section focuses on uploading individual data files and integrating them into a comprehensive dataset, as well as conducting feature engineering to facilitate thorough data analysis.

This section aligns with the following:

- Objective 1.1 of the Project Plan

Section Contents

1. Load Data
2. Process Data

Load and Process Data 1 - Load Data from CSV Files

```
In [4]: am.load_data()
```

Overview of Loaded Data

Dataset 1 - Historic Data - 28385 Rows, 15 Columns

Column	Column Type	Data Type	Null Values	Non-Null Values	Unique Values
work_order_id	original feature	string	0	28385	28385
work_order_status	original feature	string	0	28385	1
region	original feature	string	0	28385	1
coverage_area	original feature	string	0	28385	4
availability_zone	original feature	string	0	28385	18
data_center	original feature	string	0	28385	52
work_category	original feature	string	0	28385	11
sub_category	original feature	string	0	28385	106
vendor	original feature	string	0	28385	90
priority	original feature	string	0	28385	10
date_submitted	original feature	string	0	28385	26609
date_completed	original feature	string	0	28385	26866
due_date	original feature	string	0	28385	19389
dataset_number	engineered feature	integer	0	28385	1
dataset_name	engineered feature	string	0	28385	1

Dataset 2 - Current Data - 36906 Rows, 15 Columns

Column	Column Type	Data Type	Null Values	Non-Null Values	Unique Values
work_order_id	original feature	string	0	36906	36906
work_order_status	original feature	string	0	36906	1
region	original feature	string	0	36906	1
coverage_area	original feature	string	0	36906	4
availability_zone	original feature	string	0	36906	18
data_center	original feature	string	0	36906	53
work_category	original feature	string	0	36906	10
sub_category	original feature	string	0	36906	85
vendor	original feature	string	0	36906	40
priority	original feature	string	0	36906	10
date_submitted	original feature	string	0	36906	34868
date_completed	original feature	string	0	36906	35404
due_date	original feature	string	0	36906	35910
dataset_number	engineered feature	integer	0	36906	1
dataset_name	engineered feature	string	0	36906	1

Load and Process Data 2 - Process Loaded Data

```
In [5]: am.process_data()
```

Overview of Processed Data

Dataset - Combined Data - 65291 Rows, 23 Columns

Column	Column Type	Data Type	Null Values	Non-Null Values	Unique Values
work_order_id	original feature	string	0	65291	65291
work_order_status	original feature	string	0	65291	1
region	original feature	string	0	65291	1
coverage_area	original feature	string	0	65291	4
availability_zone	original feature	string	0	65291	18
data_center	original feature	string	0	65291	54
work_category	original feature	string	0	65291	11
sub_category	original feature	string	0	65291	123
vendor	original feature	string	0	65291	104
priority	original feature	string	0	65291	10
date_submitted	original feature	string	0	65291	61477
date_completed	original feature	string	0	65291	62270
due_date	original feature	string	0	65291	55299
dataset_number	engineered feature	integer	0	65291	2
dataset_name	engineered feature	string	0	65291	2
dateSubmitted	engineered feature	datetime	0	65291	61477
dateCompleted	engineered feature	datetime	0	65291	62270
dueDate	engineered feature	datetime	0	65291	55299
hours_to_complete	engineered feature	float	0	65291	28294
days_past_due	engineered feature	float	0	65291	13268
completed_on_time	engineered feature	boolean	0	65291	2
sla_group	engineered feature	integer	0	65291	2
sla_label	engineered feature	string	0	65291	2

Observations

Following the successful completion of the Data Collection, Processing, and Feature Engineering phases, the subsequent observations were documented.

- The historical dataset contained a total of 28,385 unique entries.
- The current dataset contained a total of 36,906 unique entries.
- The historical dataset contained 90 unique vendors.
- The current dataset contained 40 unique vendors.
- The processed dataset contains 65,291 unique entries.
- The processed dataset contains 13 original features and 10 engineered features.

Section 3 - Visual Analysis

Purpose

The objective of this section is to conduct a thorough visual examination of the data using histograms, box plots, bar plots, violin plots, and KDE plots. Furthermore, it encompasses performing descriptive statistical analyses and presenting the findings in a tabular format.

This section aligns with the following:

- Objective 1.2 of the Project Plan
- Section 1 of the Data Analysis Process

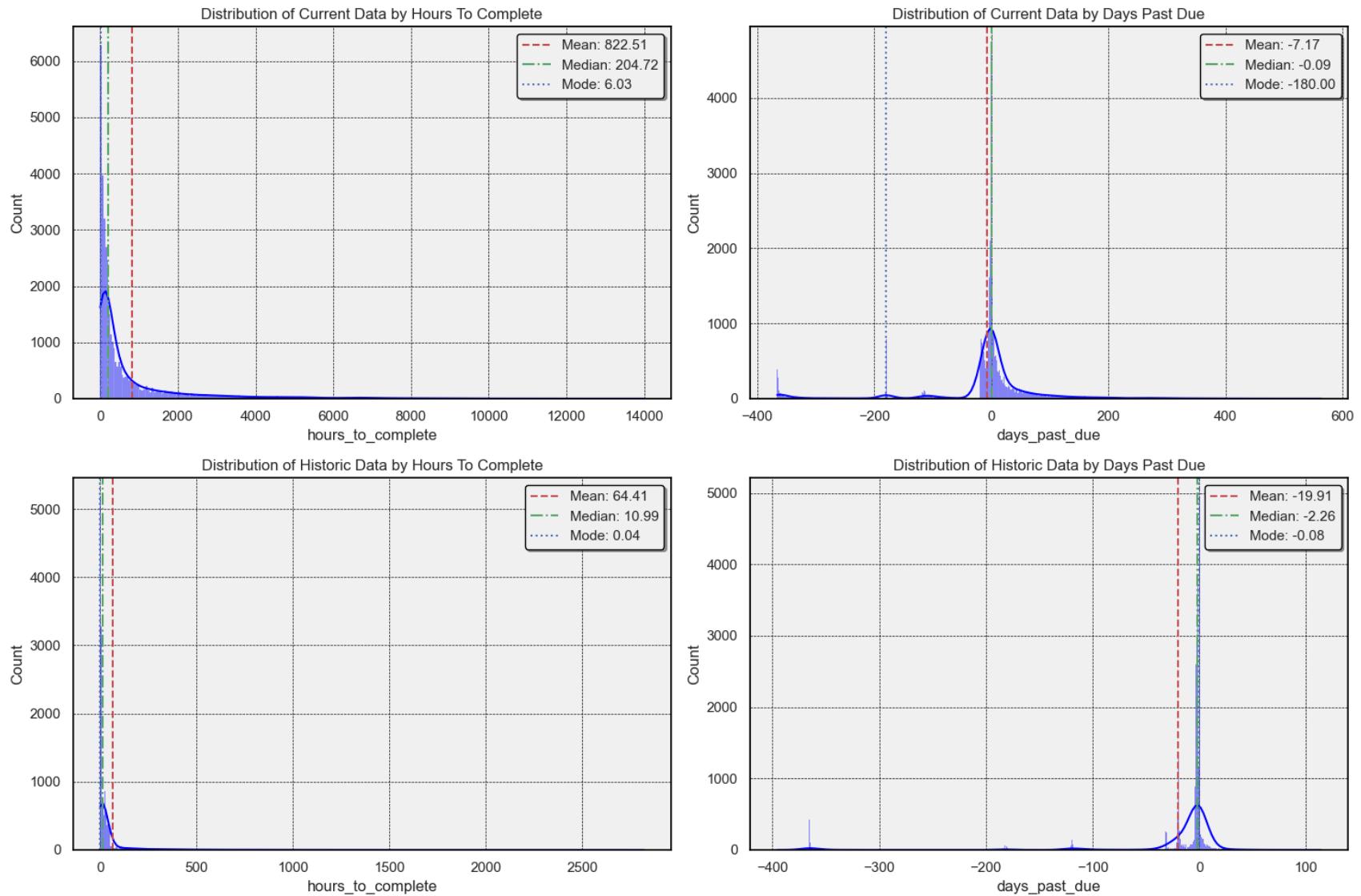
Section Contents

1. Histogram Plot Analysis
2. Box Plot Analysis
3. Count Plot Analysis of Priority Levels
4. Violin Plot Analysis of Hours to Complete by Priority Levels
5. Violin Plot Analysis of Days Past Due by Priority Levels
6. Count Plot Analysis of SLA Label
7. KDE Plot Analysis of SLA Label
8. Statistical Summary of Hours to Complete by SLA Label
9. Statistical Summary of Days Past Due by SLA Label

Visual Analysis 1 - Histogram Plot Analysis of Data

In [6]: `am.histogram_plot_analysis()`

Measure of Central Tendency for Selected Features



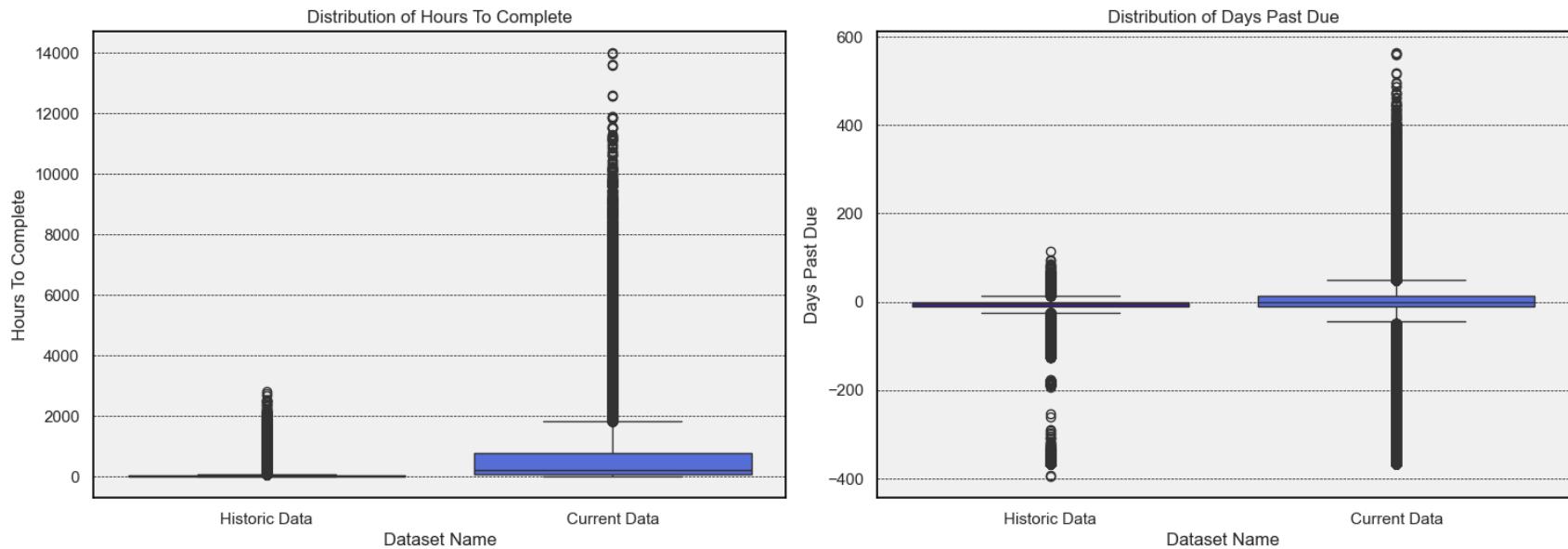
Measure of Central Tendency for Selected Features in Each Dataset Name

Dataset Name	Feature	Mean	Median	Mode
Historic Data	Hours To Complete	64.41	10.99	0.04
Historic Data	Days Past Due	-19.91	-2.26	-0.08
Current Data	Hours To Complete	822.51	204.72	6.03
Current Data	Days Past Due	-7.17	-0.09	-180.0

Visual Analysis 2 - Box Plot Analysis of Data

In [7]: `am.box_plot_analysis()`

Box Plots with Data Points for Selected Features

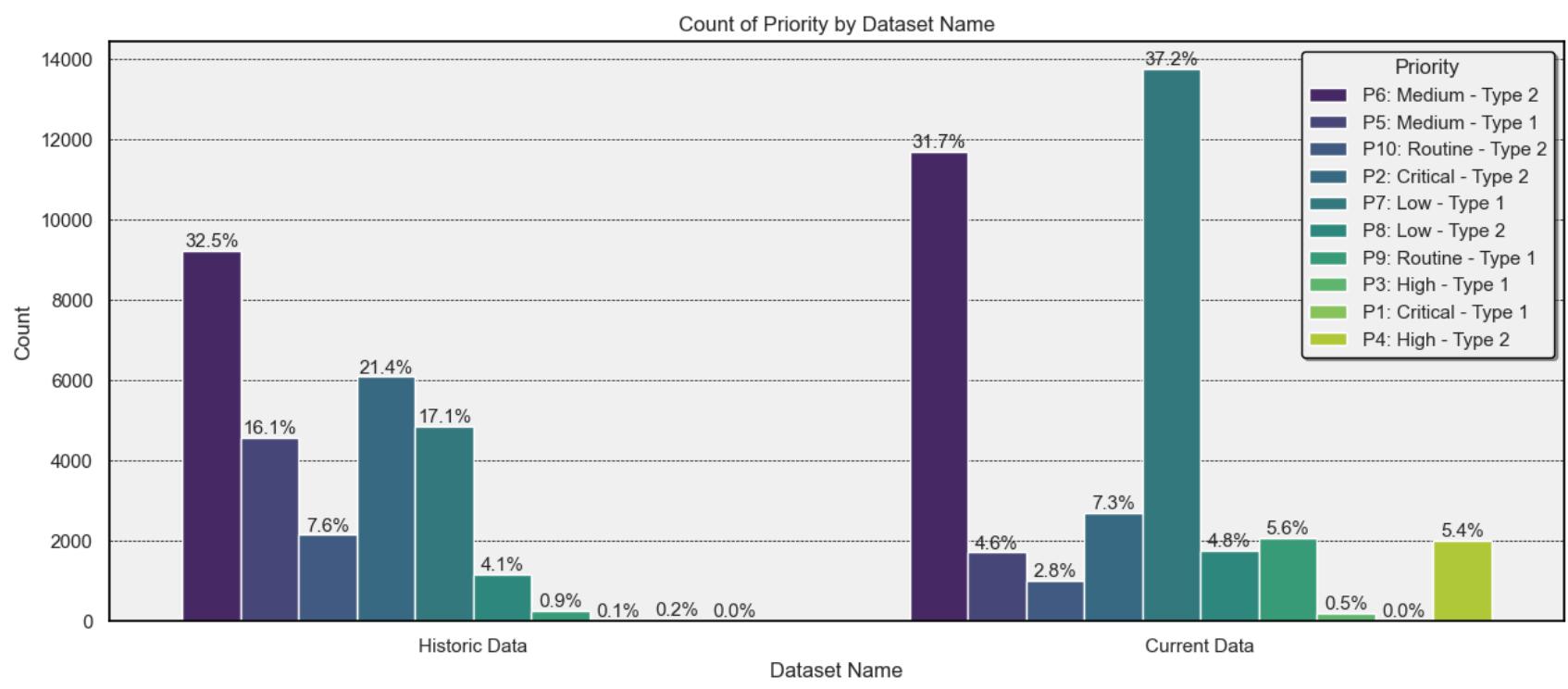


Quartile Values for Selected Features in Each Dataset Name

Dataset Name	Feature	First Quartile	Second Quartile	Third Quartile
Historic Data	Hours To Complete	3.07	10.99	34.12
Current Data	Hours To Complete	75.9	204.72	780.14
Historic Data	Days Past Due	-10.0	-2.26	-0.21
Current Data	Days Past Due	-10.54	-0.09	13.22

Visual Analysis 3 - Count Plot Analysis of Priority Levels

In [8]: `am.count_plot_analysis_1()`

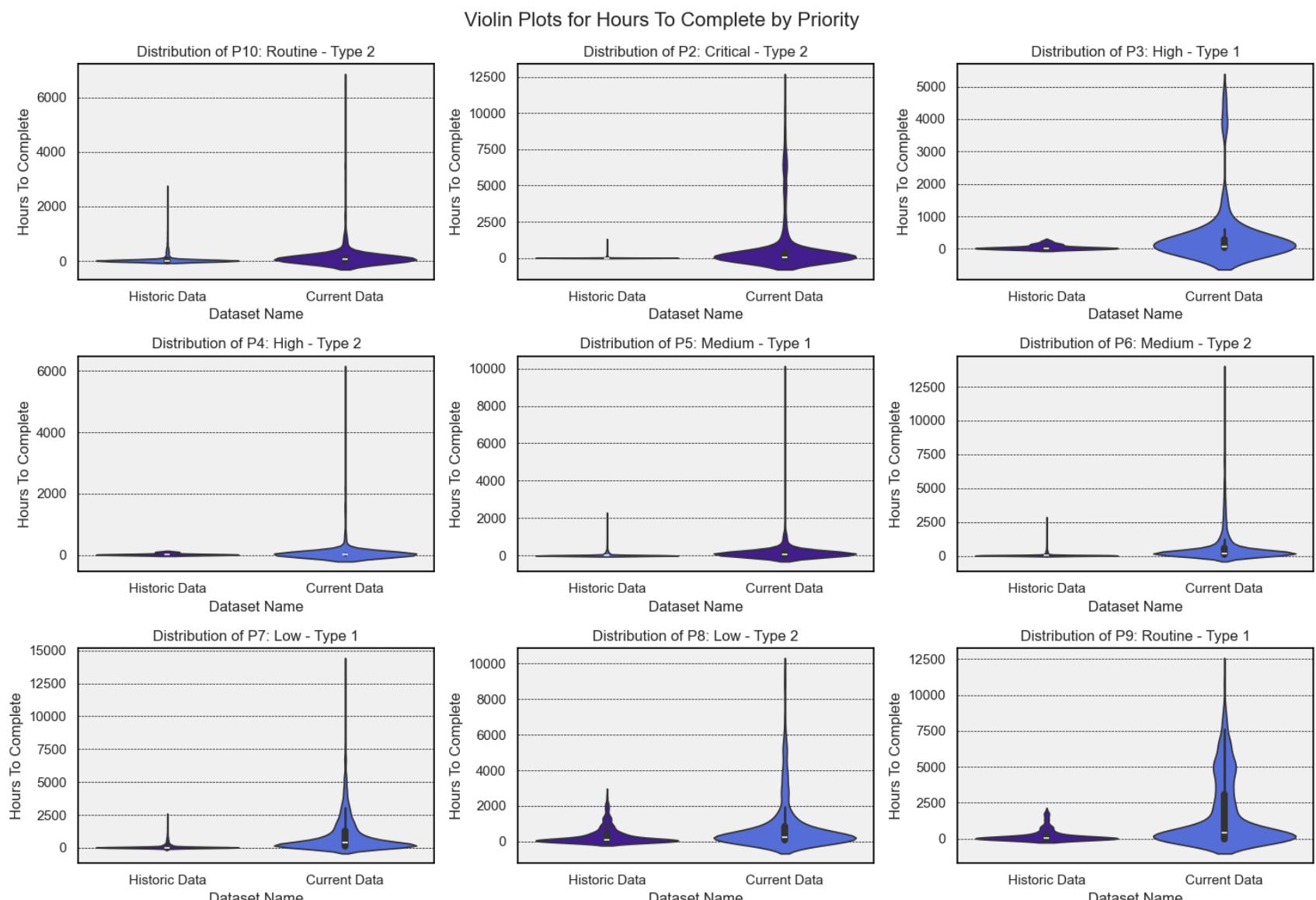


Total Number of Priority Values for Each Dataset Name

Dataset Name	P6	P5	P1	P2	P7	P8	P9	P3	P4
Historic Data	9222	4583	50	6087	4855	1160	249	20	7
Current Data	11686	1714	14	2692	13746	1760	2070	198	2006

Visual Analysis 4 - Violin Plot Analysis of Hours to Complete by Priority Levels

In [9]: `am.violin_plot_analysis_1()`

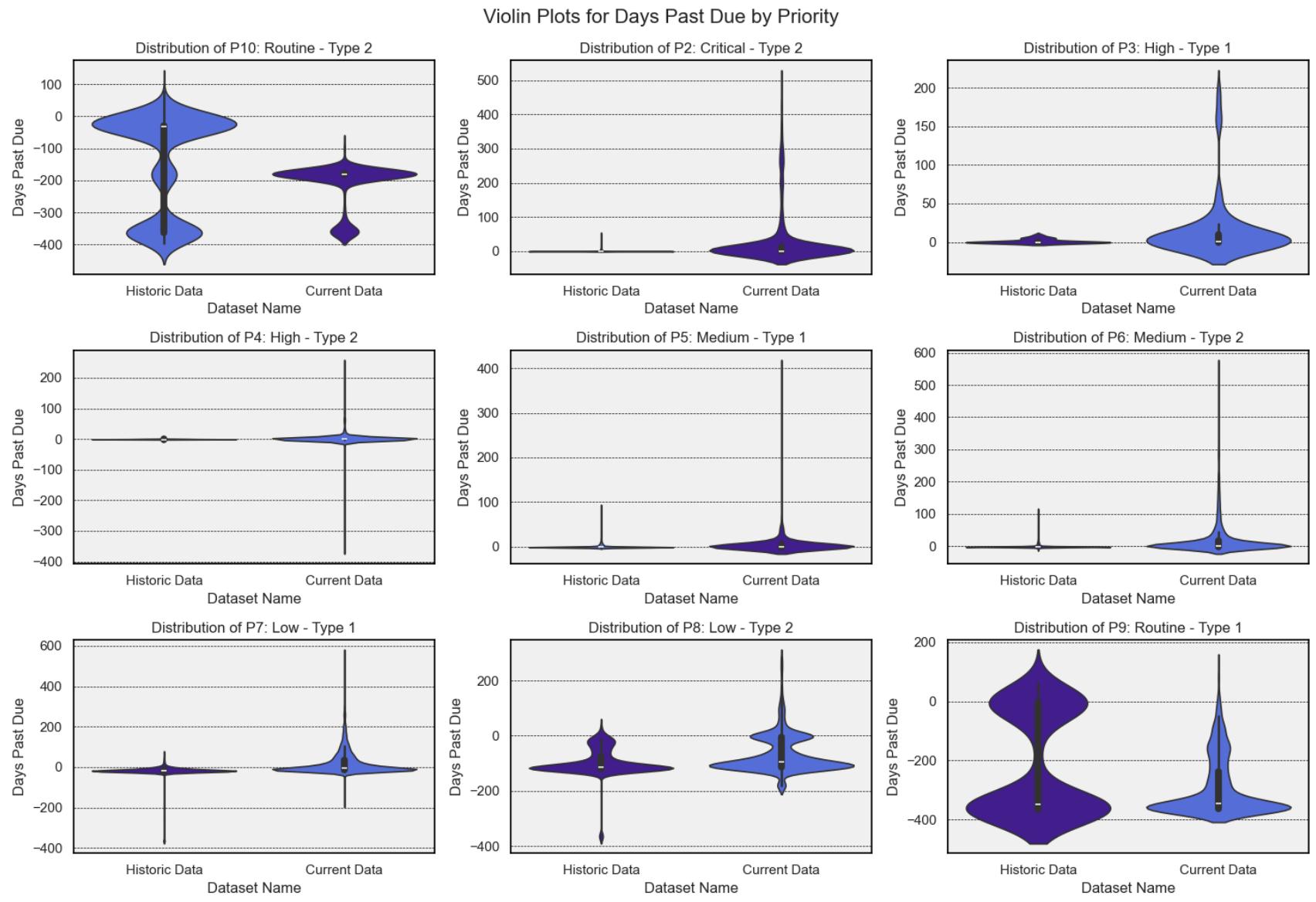


Violin Plot Details for Each Priority by Each Dataset Name

Priority	Dataset Name	Median	Interquartile Range	Range
P10: Routine - Type 2	Historic Data	0.09	23.45	2683.42
P10: Routine - Type 2	Current Data	56.56	55.78	6542.45
P2: Critical - Type 2	Historic Data	3.41	2.72	1296.27
P2: Critical - Type 2	Current Data	28.55	257.55	11908.67
P3: High - Type 1	Historic Data	20.91	40.34	234.14
P3: High - Type 1	Current Data	74.13	246.65	4746.55
P4: High - Type 2	Historic Data	3.86	10.34	85.34
P4: High - Type 2	Current Data	11.28	15.36	5947.45
P5: Medium - Type 1	Historic Data	3.84	3.62	2253.21
P5: Medium - Type 1	Current Data	79.4	140.14	9820.45
P6: Medium - Type 2	Historic Data	25.03	28.06	2820.21
P6: Medium - Type 2	Current Data	201.68	453.14	13615.83
P7: Low - Type 1	Historic Data	25.91	32.67	2520.78
P7: Low - Type 1	Current Data	417.2	1156.74	13996.57
P8: Low - Type 2	Historic Data	89.57	386.38	2740.59
P8: Low - Type 2	Current Data	253.84	749.59	9633.4
P9: Routine - Type 1	Historic Data	19.85	240.68	1875.81
P9: Routine - Type 1	Current Data	439.99	3065.93	11551.46

Visual Analysis 5 - Violin Plot Analysis of Days Past Due by Priority Levels

In [10]: `am.violin_plot_analysis_2()`

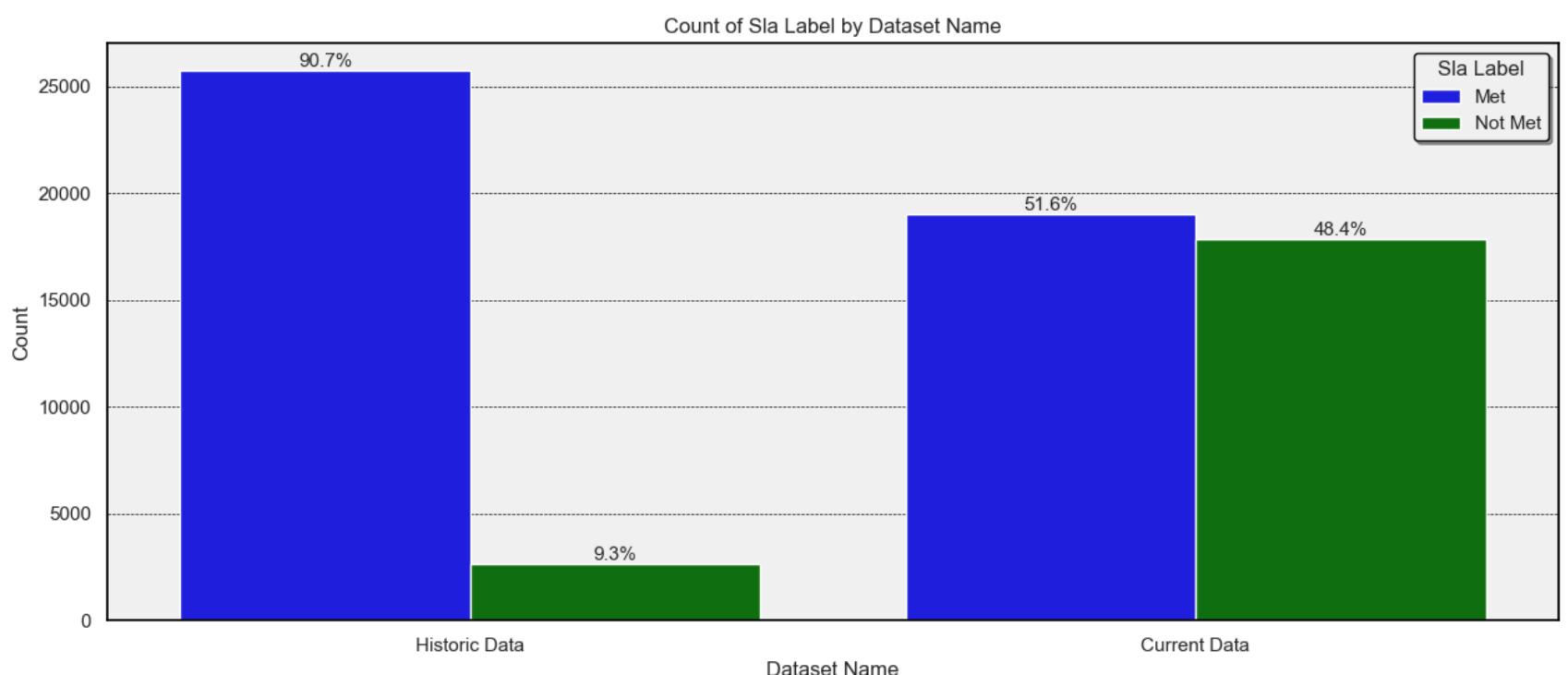


Violin Plot Details for Each Priority by Each Dataset Name

Priority	Dataset Name	Median	Interquartile Range	Range
P10: Routine - Type 2	Historic Data	-31.64	333.34	473.02
P10: Routine - Type 2	Current Data	-180.0	1.0	272.39
P2: Critical - Type 2	Historic Data	-0.11	0.12	54.97
P2: Critical - Type 2	Current Data	0.78	10.73	500.61
P3: High - Type 1	Historic Data	-0.04	1.62	9.69
P3: High - Type 1	Current Data	1.09	10.28	197.77
P4: High - Type 2	Historic Data	-1.09	0.73	1.58
P4: High - Type 2	Current Data	-0.03	0.65	611.26
P5: Medium - Type 1	Historic Data	-1.31	0.23	95.38
P5: Medium - Type 1	Current Data	-0.03	5.85	409.18
P6: Medium - Type 2	Historic Data	-2.57	1.18	127.01
P6: Medium - Type 2	Current Data	1.4	18.89	567.33
P7: Low - Type 1	Historic Data	-19.26	4.8	432.74
P7: Low - Type 1	Current Data	-2.6	48.21	743.21
P8: Low - Type 2	Historic Data	-115.34	43.02	401.28
P8: Low - Type 2	Current Data	-95.16	108.9	462.26
P9: Routine - Type 1	Historic Data	-348.68	360.75	425.99
P9: Routine - Type 1	Current Data	-346.04	126.54	482.06

Visual Analysis 6 - Count Plot Analysis of SLA Label

In [11]: `am.count_plot_analysis_2()`

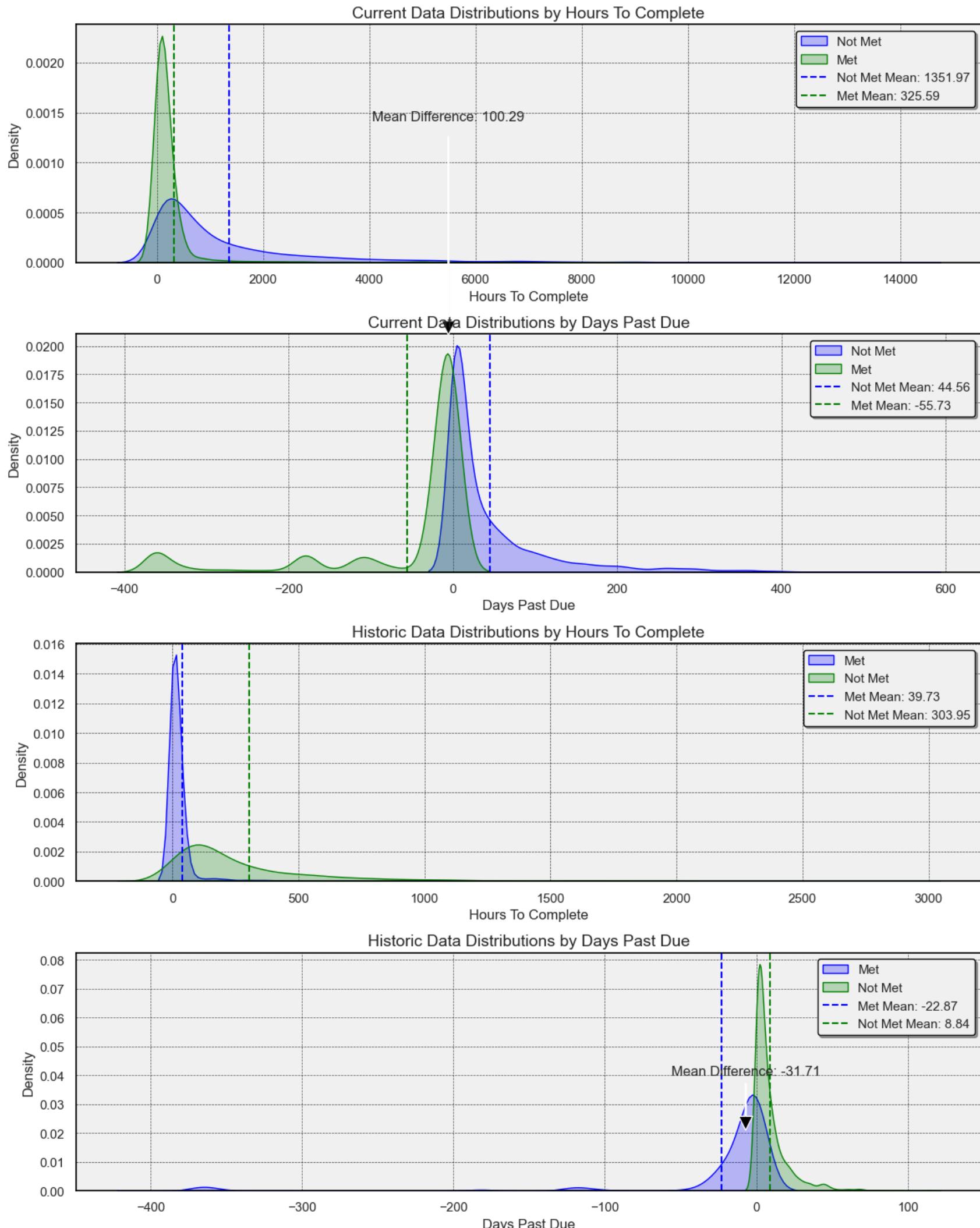


Total Number of Sla Label Values for Each Dataset Name

Dataset Name	Met	Not Met
Historic Data	25733	2652
Current Data	19038	17868

Visual Analysis 7 - KDE Plot Analysis of SLA Label

In [12]: `am.kde_plot_analysis()`



Visual Analysis 8 - Statistical Summary of Hours to Complete by SLA Label

In [13]: `am.statistic_summary_1()`

Statistical Summary of Hours To Complete for Sla Label (Met) by each Dataset Name

Dataset Name	Met mean	Met median	Met std	Met min	Met max
Current Data	325.59	103.48	913.45	0.02	8761.37
Historic Data	39.73	6.43	144.77	0.0	2740.61

Statistical Summary of Hours To Complete for Sla Label (Not Met) by each Dataset Name

Dataset Name	Not Met mean	Not Met median	Not Met std	Not Met min	Not Met max
Current Data	1351.97	656.86	1796.31	6.63	13996.95
Historic Data	303.95	165.13	364.38	6.01	2820.21

Summary of Comparison Analysis for Hours To Complete by Target Group

Target Group	Key Observations from Comparisons
Current Data	Met has lower average Hours To Complete values (by 315.2%) than Not Met
Historic Data	Met has lower average Hours To Complete values (by 665.1%) than Not Met
Met	Current Data has higher average Hours To Complete values (by 87.8%) than Historic Data
Not Met	Current Data has higher average Hours To Complete values (by 77.5%) than Historic Data

Visual Analysis 9 - Statistical Summary of Days Past Due by SLA Label

In [14]: `am.statistic_summary_2()`

Statistical Summary of Days Past Due for Sla Label (Met) by each Dataset Name

Dataset Name	Met mean	Met median	Met std	Met min	Met max
Current Data	-55.73	-9.6	103.59	-365.79	-0.0
Historic Data	-22.87	-2.48	67.73	-395.57	-0.0

Statistical Summary of Days Past Due for Sla Label (Not Met) by each Dataset Name

Dataset Name	Not Met mean	Not Met median	Not Met std	Not Met min	Not Met max
Current Data	44.56	14.3	70.8	0.0	563.21
Historic Data	8.84	4.38	12.16	0.0	114.01

Summary of Comparison Analysis for Days Past Due by Target Group

Target Group	Key Observations from Comparisons
Current Data	Met has lower average Days Past Due values (by 180.0%) than Not Met
Historic Data	Met has lower average Days Past Due values (by 138.6%) than Not Met
Met	Current Data has lower average Days Past Due values (by 59.0%) than Historic Data
Not Met	Current Data has higher average Days Past Due values (by 80.2%) than Historic Data

Observations

Following the successful completion of the visual and statistical analysis, the subsequent observations were documented.

- The means of the historical data are significantly lower than those of the current data.
- The 'Hours to Complete' attribute is substantially right-skewed in both datasets.
- The historic data exhibits less dispersion and fewer outliers compared to the current data. P2 and P6 are the priority levels with the most significant number of entries in both datasets.
- The historical data met the SLA target 90.7% of the time compared to only 51.6% for the current data.
- The means for both SLA Met and SLA Not Met were noticeably lower in the historical data than in the current data.

Section 4 - Outlier Analysis

Purpose

The objective of this section entails a thorough analysis of the dataset to identify outliers and potential anomalies. The findings are presented using box plots, histograms, and comprehensive statistical tables.

This section aligns with the following:

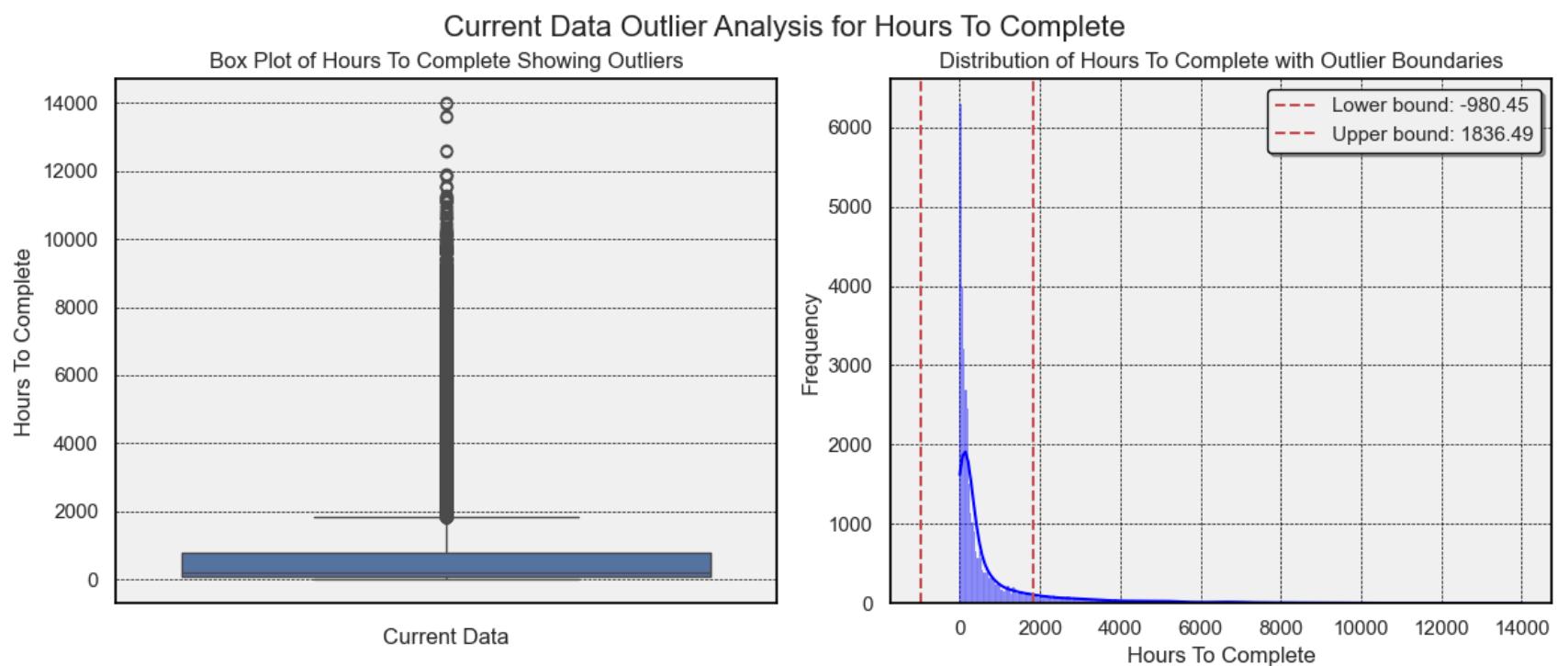
- Objective 1.3 of the Project Plan
- Section 2 of the Data Analysis Process

Section Contents

- Analysis of Hours to Complete for the Current Data
- Analysis of Hours to Complete for the Historic Data
- Analysis of Days Past Due for the Current Data
- Analysis of Days Past Due for the Historic Data

Outlier Analysis 1 - Analysis of Hours to Complete for the Current Data

In [15]: `am.outlier_analysis_1()`

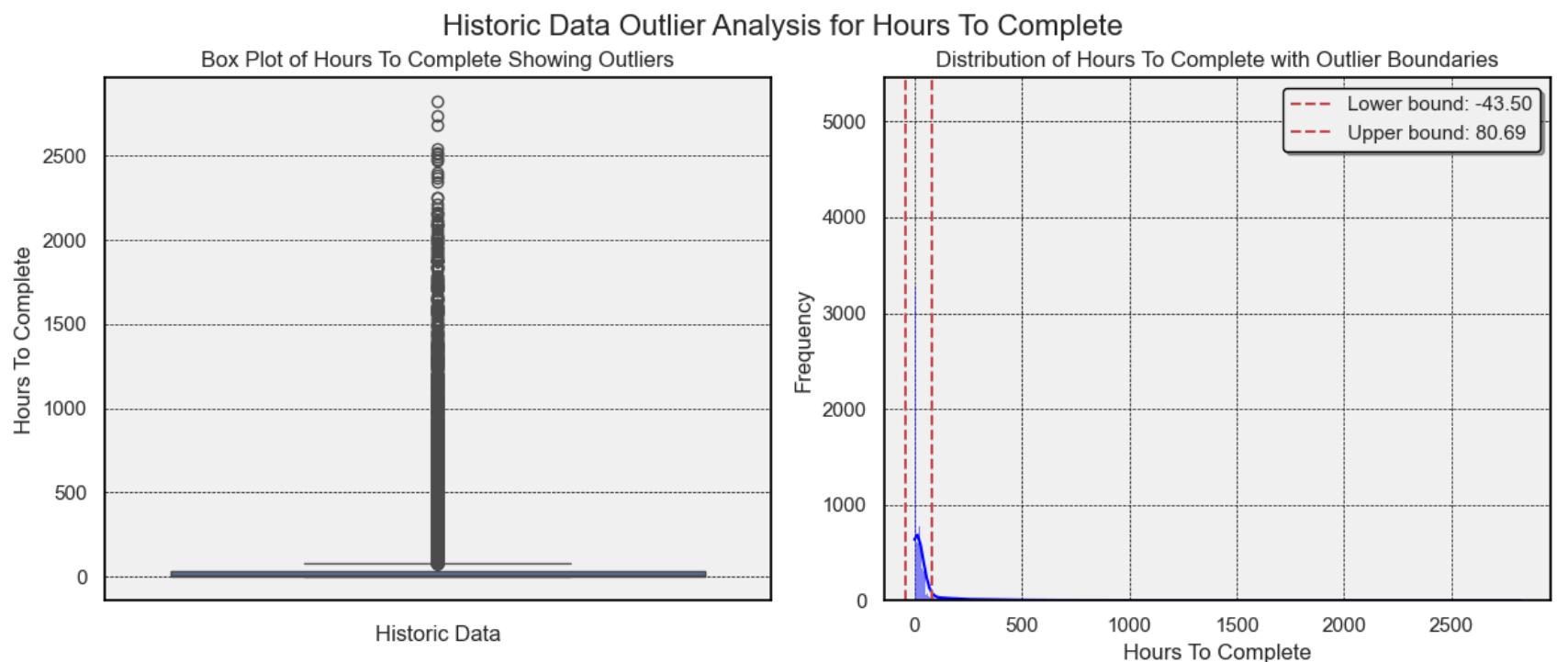


Summary of Outlier Analysis for Hours To Complete in Current Data

Statistic	Value
Number of Outliers	4901
Percentage of Outliers	13.28
Upper Bound	1,836.49
Lower Bound	-980.45
Confidence Interval	2,816.94

Outlier Analysis 2 - Analysis of Hours to Complete for the Historic Data

In [16]: `am.outlier_analysis_2()`

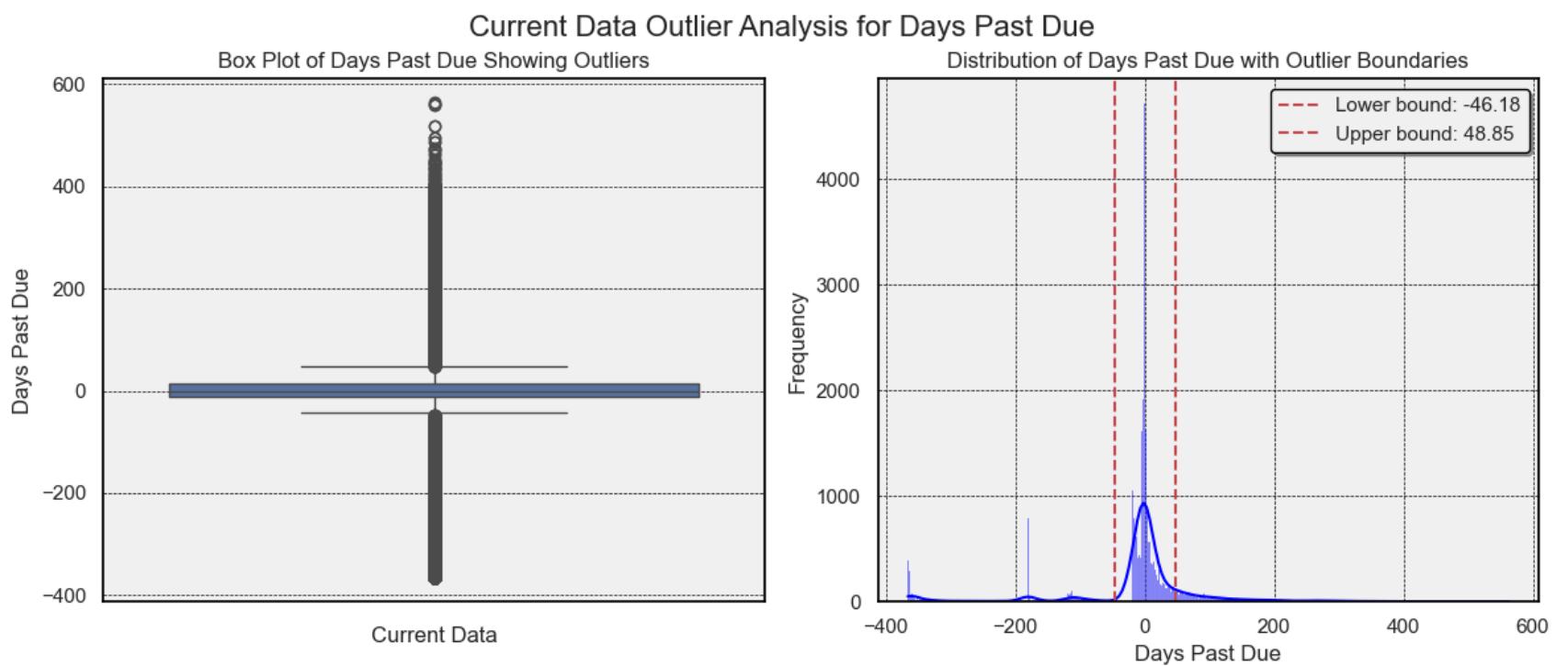


Summary of Outlier Analysis for Hours To Complete in Historic Data

Statistic	Value
Number of Outliers	3,687
Percentage of Outliers	12.99
Upper Bound	80.69
Lower Bound	-43.50
Confidence Interval	124.20

Outlier Analysis 3 - Analysis of Days Past Due for the Current Data

In [17]: `am.outlier_analysis_3()`

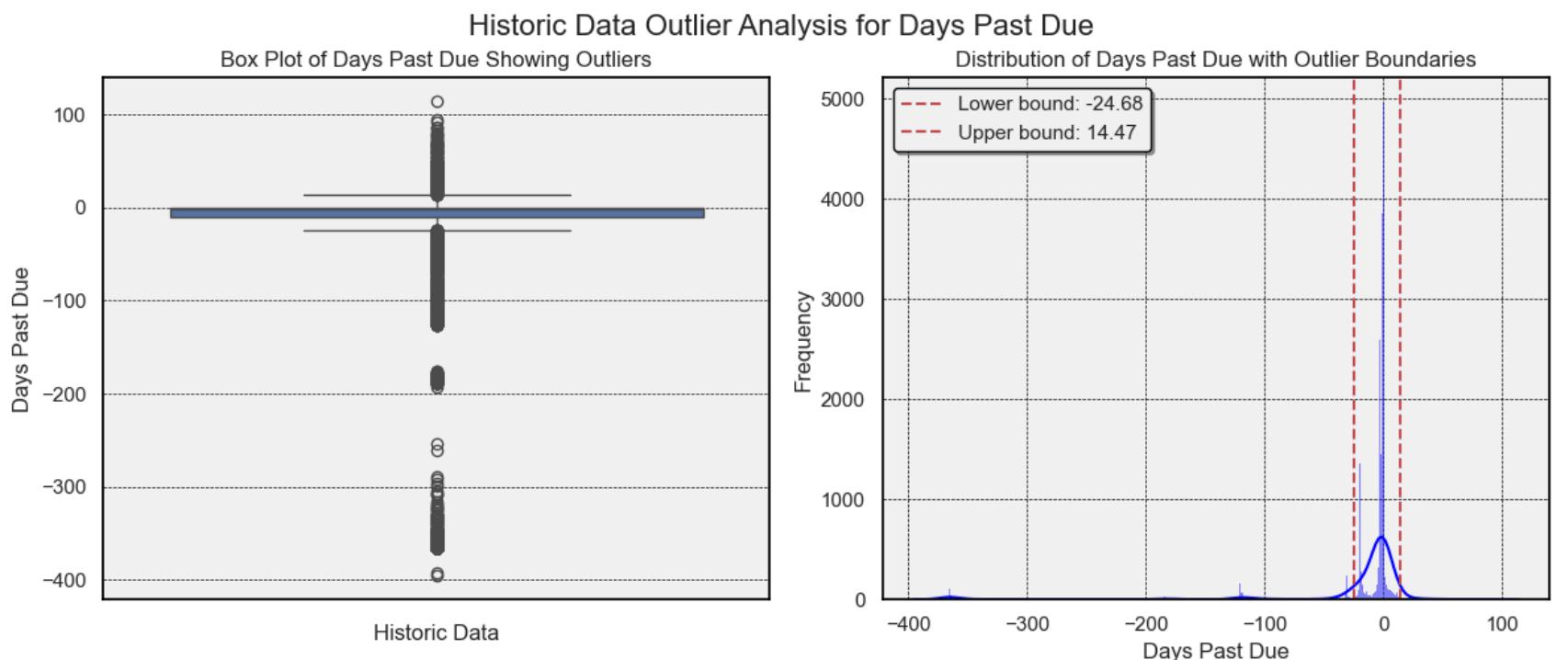


Summary of Outlier Analysis for Days Past Due in Current Data

Statistic	Value
Number of Outliers	9156
Percentage of Outliers	24.81
Upper Bound	48.85
Lower Bound	-46.18
Confidence Interval	95.03

Outlier Analysis 4 - Analysis of Days Past Due for the Historic Data

```
In [18]: am.outlier_analysis_4()
```



Summary of Outlier Analysis for Days Past Due in Historic Data

Statistic	Value
Number of Outliers	3480
Percentage of Outliers	12.26
Upper Bound	14.47
Lower Bound	-24.68
Confidence Interval	39.16

Observations

Following the successful completion of the data outlier analysis, the subsequent observations were documented.

- The total number of outliers was significantly lower in the historical data compared to the current data.
- The current data has 4255 more outliers in the 'Days Past Due' attribute than in the 'Hours to Complete' attribute.
- The historic data has 207 more outliers in the 'Hours to Complete' attribute than in the 'Days Past Due' attribute.
- The confidence interval is considerably smaller in the historical data than in the current data for both 'Hours to Complete' and 'Days Past Due' attributes.

Section 5 - Data Distribution Analysis

Purpose

The objective of this section is to conduct a comprehensive distribution analysis of the dataset to assess its normality, employing both the Shapiro-Wilk and Anderson-Darling tests. The evaluation additionally encompasses an examination of skewness and kurtosis.

This section aligns with the following:

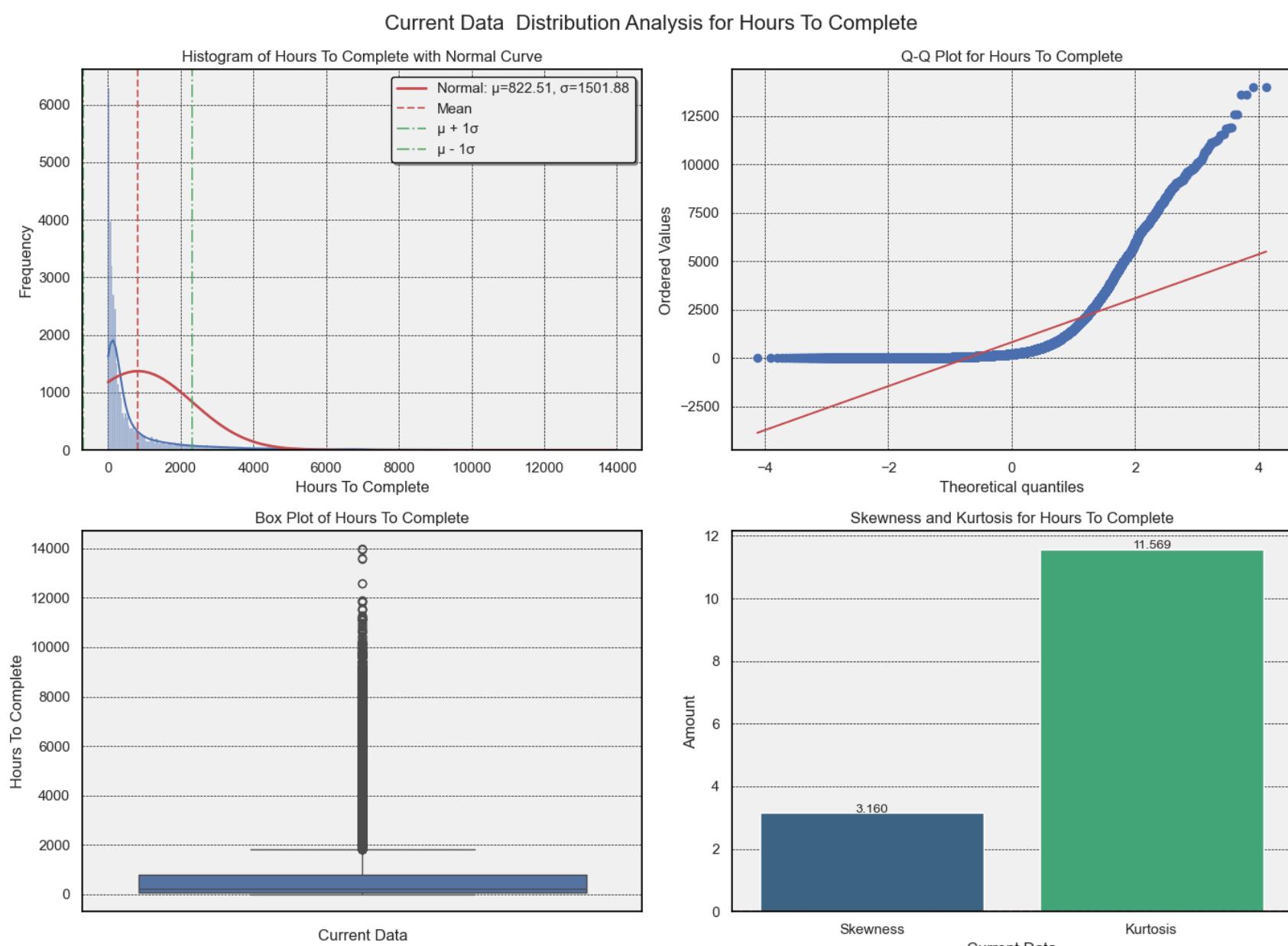
- Objective 1.4 of the Project Plan
- Section 3 of the Data Analysis Process

Section Contents

1. Analysis of Hours to Complete for the Current Data
2. Analysis of Hours to Complete for the Historic Data
3. Analysis of Days Past Due for the Current Data
4. Analysis of Days Past Due for the Historic Data

Distribution Analysis 1 - Analysis of Hours to Complete for the Current Data

In [19]: `am.distribution_analysis_1()`



Shapiro-Wilk Test Results by Priority

Target Group	Statistic	P_Value	Test Interpretation
P1	0.250599	0.000000	Data does not look normally distributed
P2	0.449000	0.000000	Data does not look normally distributed
P3	0.449861	0.000000	Data does not look normally distributed
P4	0.201324	0.000000	Data does not look normally distributed
P5	0.209352	0.000000	Data does not look normally distributed
P6	0.496201	0.000000	Data does not look normally distributed
P7	0.674730	0.000000	Data does not look normally distributed
P8	0.587820	0.000000	Data does not look normally distributed
P9	0.761266	0.000000	Data does not look normally distributed

Shapiro-Wilk Test Average Results

Test Result	Value
Average Statistic	0.453350
Average P-Value	0.000000

Overall Assessment of Shapiro-Wilk Test

Interpretation
Reject H ₀ : Data does not look normally distributed

Anderson-Darling Test Result

Anderson Result	Value
Statistic	5546.1183
P-Value	0.000000

Critical values at significance levels

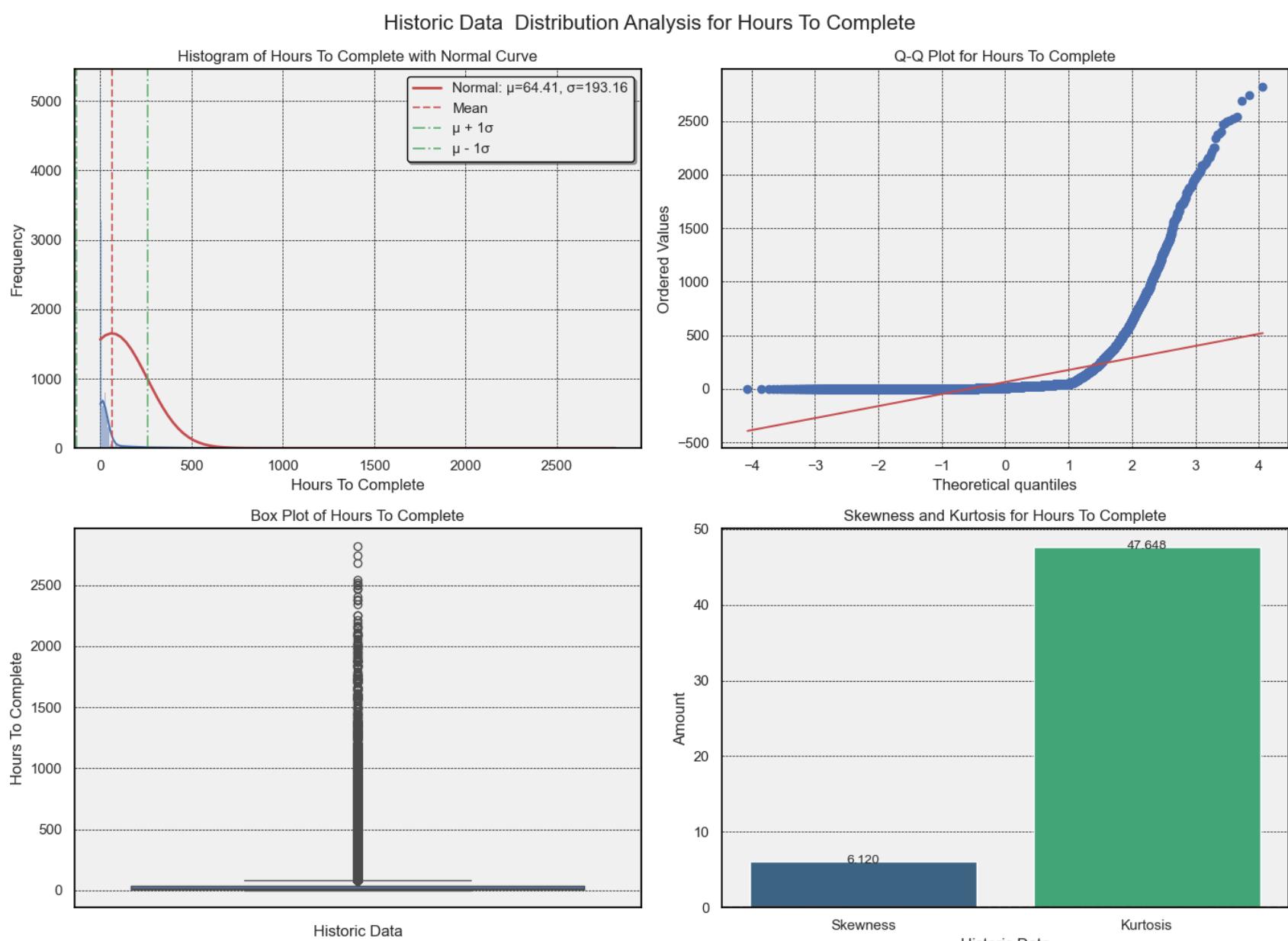
Significance Level	Critical Value	Test Interpretation
15.0%	0.576000	Reject H ₀ : Data does NOT look normally distributed
10.0%	0.656000	Reject H ₀ : Data does NOT look normally distributed
5.0%	0.787000	Reject H ₀ : Data does NOT look normally distributed
2.5%	0.918000	Reject H ₀ : Data does NOT look normally distributed
1.0%	1.092000	Reject H ₀ : Data does NOT look normally distributed

Overall Assessment of Skewness and Kurtosis

Test	Result	Interpretation
Skewness	3.160301	The distribution is positively skewed (right-tailed)
Kurtosis	11.568701	The data distribution has heavier tails than normal (leptokurtic)

Distribution Analysis 2 - Analysis of Hours to Complete for the Historic Data

In [20]: `am.distribution_analysis_2()`



Shapiro-Wilk Test Results by Priority

Target Group	Statistic	P_Value	Test Interpretation
P1	0.320423	0.000000	Data does not look normally distributed
P1	0.357515	0.000000	Data does not look normally distributed
P2	0.253069	0.000000	Data does not look normally distributed
P3	0.710796	0.000053	Data does not look normally distributed
P5	0.281995	0.000000	Data does not look normally distributed
P6	0.326000	0.000000	Data does not look normally distributed
P7	0.416051	0.000000	Data does not look normally distributed
P8	0.692789	0.000000	Data does not look normally distributed
P9	0.597420	0.000000	Data does not look normally distributed

Shapiro-Wilk Test Average Results

Test Result	Value
Average Statistic	0.439562
Average P-Value	0.000006

Overall Assessment of Shapiro-Wilk Test

Interpretation
Reject H ₀ : Data does not look normally distributed

Anderson-Darling Test Result

Anderson Result	Value
Statistic	6581.0305
P-Value	0.000000

Critical values at significance levels

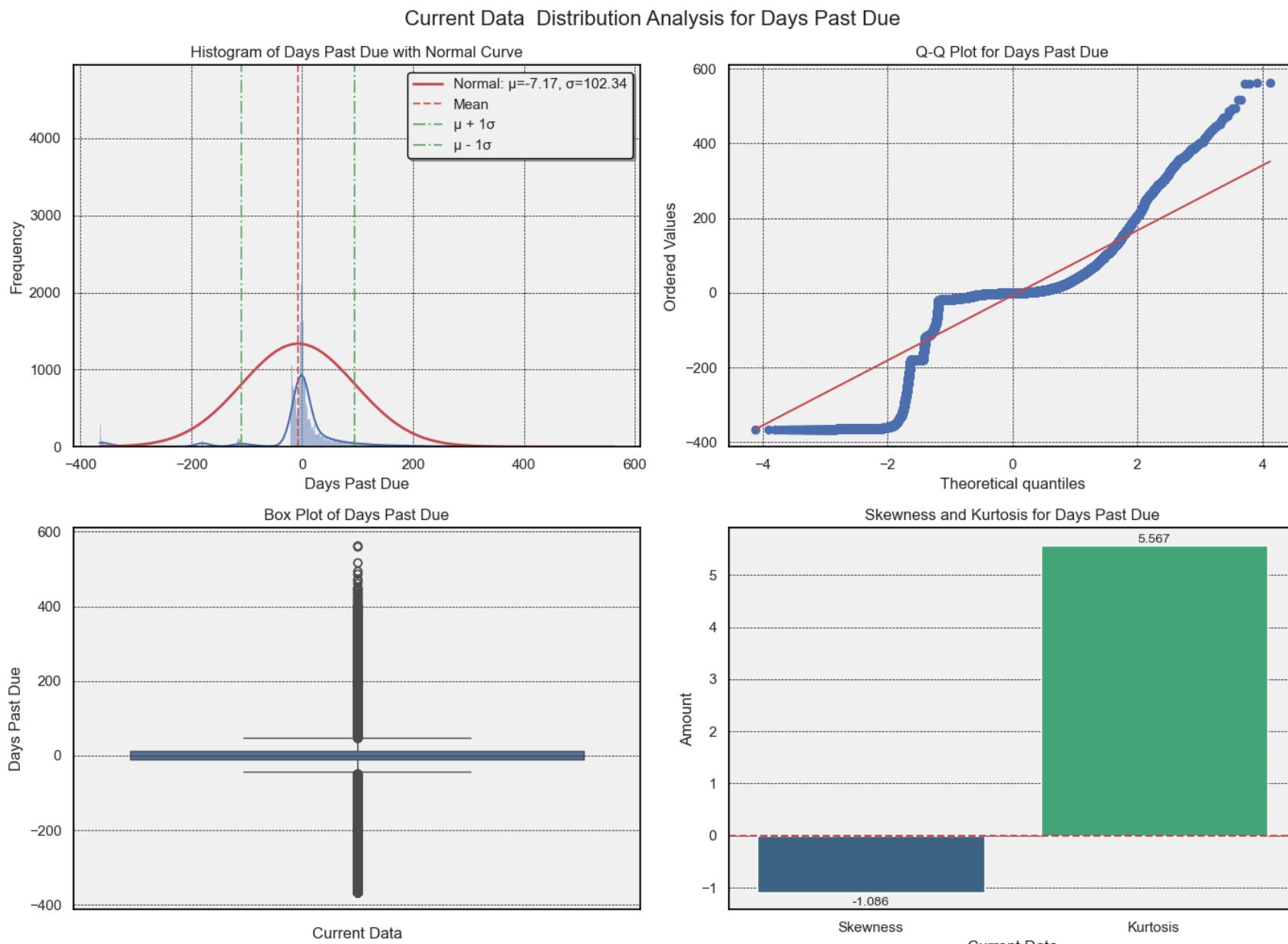
Significance Level	Critical Value	Test Interpretation
15.0%	0.576000	Reject H ₀ : Data does NOT look normally distributed
10.0%	0.656000	Reject H ₀ : Data does NOT look normally distributed
5.0%	0.787000	Reject H ₀ : Data does NOT look normally distributed
2.5%	0.918000	Reject H ₀ : Data does NOT look normally distributed
1.0%	1.092000	Reject H ₀ : Data does NOT look normally distributed

Overall Assessment of Skewness and Kurtosis

Test	Result	Interpretation
Skewness	6.119577	The distribution is positively skewed (right-tailed)
Kurtosis	47.647690	The data distribution has heavier tails than normal (leptokurtic)

Distribution Analysis 3 - Analysis of Days Past Due for the Current Data

In [21]: `am.distribution_analysis_3()`



Shapiro-Wilk Test Results by Priority

Target Group	Statistic	P_Value	Test Interpretation
P1	0.519374	0.000000	Data does not look normally distributed
P2	0.449177	0.000000	Data does not look normally distributed
P3	0.449572	0.000000	Data does not look normally distributed
P4	0.219187	0.000000	Data does not look normally distributed
P5	0.209210	0.000000	Data does not look normally distributed
P6	0.496166	0.000000	Data does not look normally distributed
P7	0.684769	0.000000	Data does not look normally distributed
P8	0.820719	0.000000	Data does not look normally distributed
P9	0.762938	0.000000	Data does not look normally distributed

Shapiro-Wilk Test Average Results

Test Result	Value
Average Statistic	0.512346
Average P-Value	0.000000

Overall Assessment of Shapiro-Wilk Test

Interpretation
Reject H ₀ : Data does not look normally distributed

Anderson-Darling Test Result

Anderson Result	Value
Statistic	4366.0470
P-Value	0.000000

Critical values at significance levels

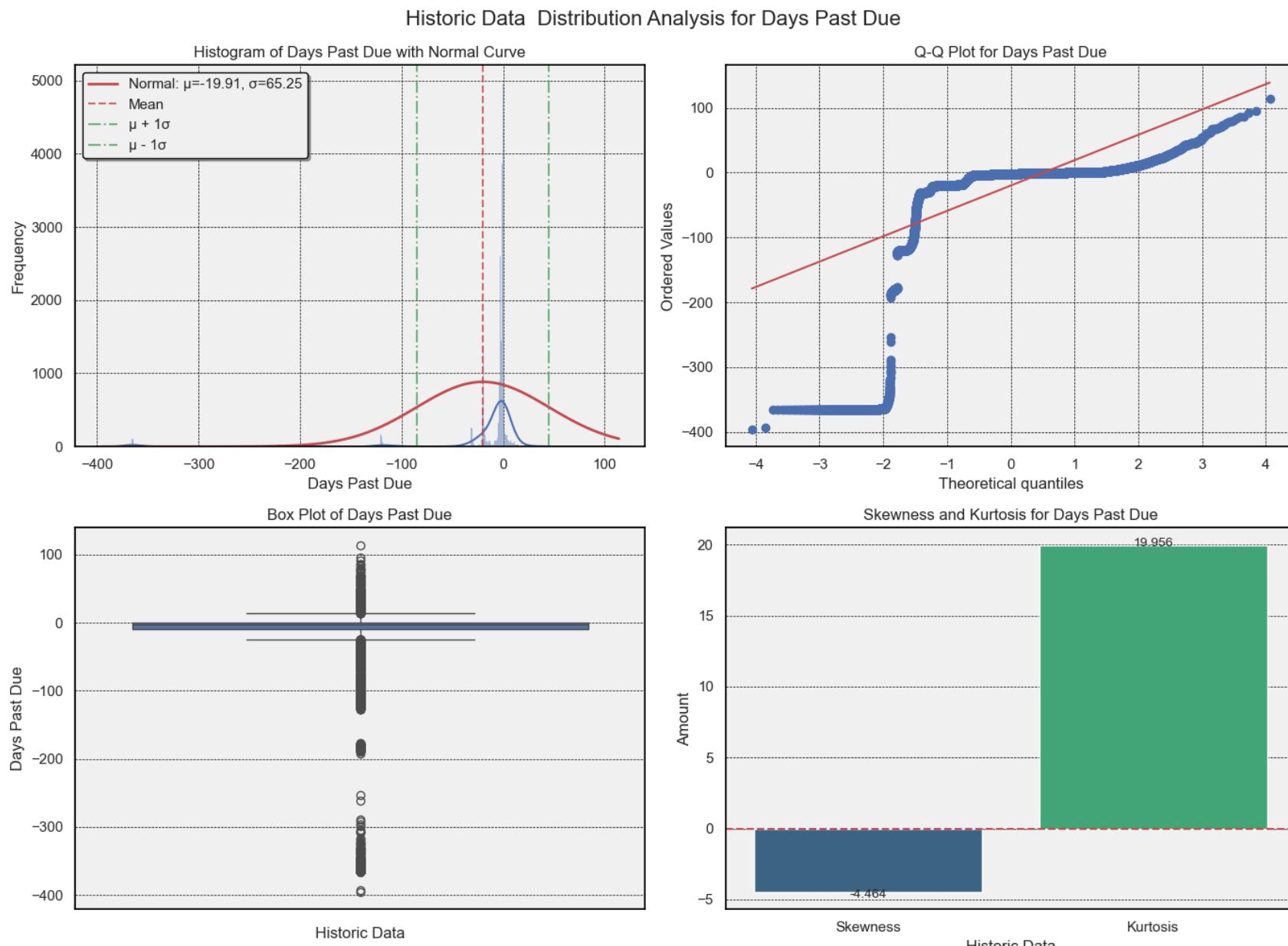
Significance Level	Critical Value	Test Interpretation
15.0%	0.576000	Reject H ₀ : Data does NOT look normally distributed
10.0%	0.656000	Reject H ₀ : Data does NOT look normally distributed
5.0%	0.787000	Reject H ₀ : Data does NOT look normally distributed
2.5%	0.918000	Reject H ₀ : Data does NOT look normally distributed
1.0%	1.092000	Reject H ₀ : Data does NOT look normally distributed

Overall Assessment of Skewness and Kurtosis

Test	Result	Interpretation
Skewness	-1.086275	The distribution is negatively skewed (left-tailed)
Kurtosis	5.567246	The data distribution has heavier tails than normal (leptokurtic)

Distribution Analysis 4 - Analysis of Days Past Due for the Historic Data

In [22]: `am.distribution_analysis_4()`



Shapiro-Wilk Test Results by Priority

Target Group	Statistic	P_Value	Test Interpretation
P1	0.721644	0.000000	Data does not look normally distributed
P1	0.498321	0.000000	Data does not look normally distributed
P2	0.255672	0.000000	Data does not look normally distributed
P3	0.707790	0.000049	Data does not look normally distributed
P5	0.287623	0.000000	Data does not look normally distributed
P6	0.359023	0.000000	Data does not look normally distributed
P7	0.201839	0.000000	Data does not look normally distributed
P8	0.701051	0.000000	Data does not look normally distributed
P9	0.692980	0.000000	Data does not look normally distributed

Shapiro-Wilk Test Average Results

Test Result	Value
Average Statistic	0.491771
Average P-Value	0.000005

Overall Assessment of Shapiro-Wilk Test

Interpretation
Reject H ₀ : Data does not look normally distributed

Anderson-Darling Test Result

Anderson Result	Value
Statistic	6938.0329
P-Value	0.000000

Critical values at significance levels

Significance Level	Critical Value	Test Interpretation
15.0%	0.576000	Reject H ₀ : Data does NOT look normally distributed
10.0%	0.656000	Reject H ₀ : Data does NOT look normally distributed
5.0%	0.787000	Reject H ₀ : Data does NOT look normally distributed
2.5%	0.918000	Reject H ₀ : Data does NOT look normally distributed
1.0%	1.092000	Reject H ₀ : Data does NOT look normally distributed

Overall Assessment of Skewness and Kurtosis

Test	Result	Interpretation
Skewness	-4.463955	The distribution is negatively skewed (left-tailed)
Kurtosis	19.956373	The data distribution has heavier tails than normal (leptokurtic)

Observations

Following the successful completion of the data distribution analysis, the subsequent observations were documented.

- The current and historical data are non-normally distributed for both 'Hours to Complete' and 'Days Past Due' attributes.
- The p-value for the Shapiro-Wilk test was less than 0.000000 for both current and historical datasets in the 'Hours to Complete' and 'Days Past Due' attributes.
- The p-value for the Anderson-Darling test was less than 0.000000 for the current and historical datasets in both 'Hours to Complete' and 'Days Past Due' attributes.
- Both the current and historical datasets are positively skewed for the 'Hours to Complete' attribute.
- Both the current and historical datasets are negatively skewed for the 'Days Past Due' attribute.
- The data distribution has heavier tails than a normal distribution for both current and historical datasets in the 'Hours to Complete' and 'Days Past Due' attributes.

Section 6 - Data Transformation Analysis

Purpose

The objective of this section is to execute a comprehensive transformational analysis on the dataset by applying five established transformation techniques to determine the most appropriate methodology for the data. The evaluation results are presented through histograms, Q-Q plots, and statistical tables.

This section aligns with the following:

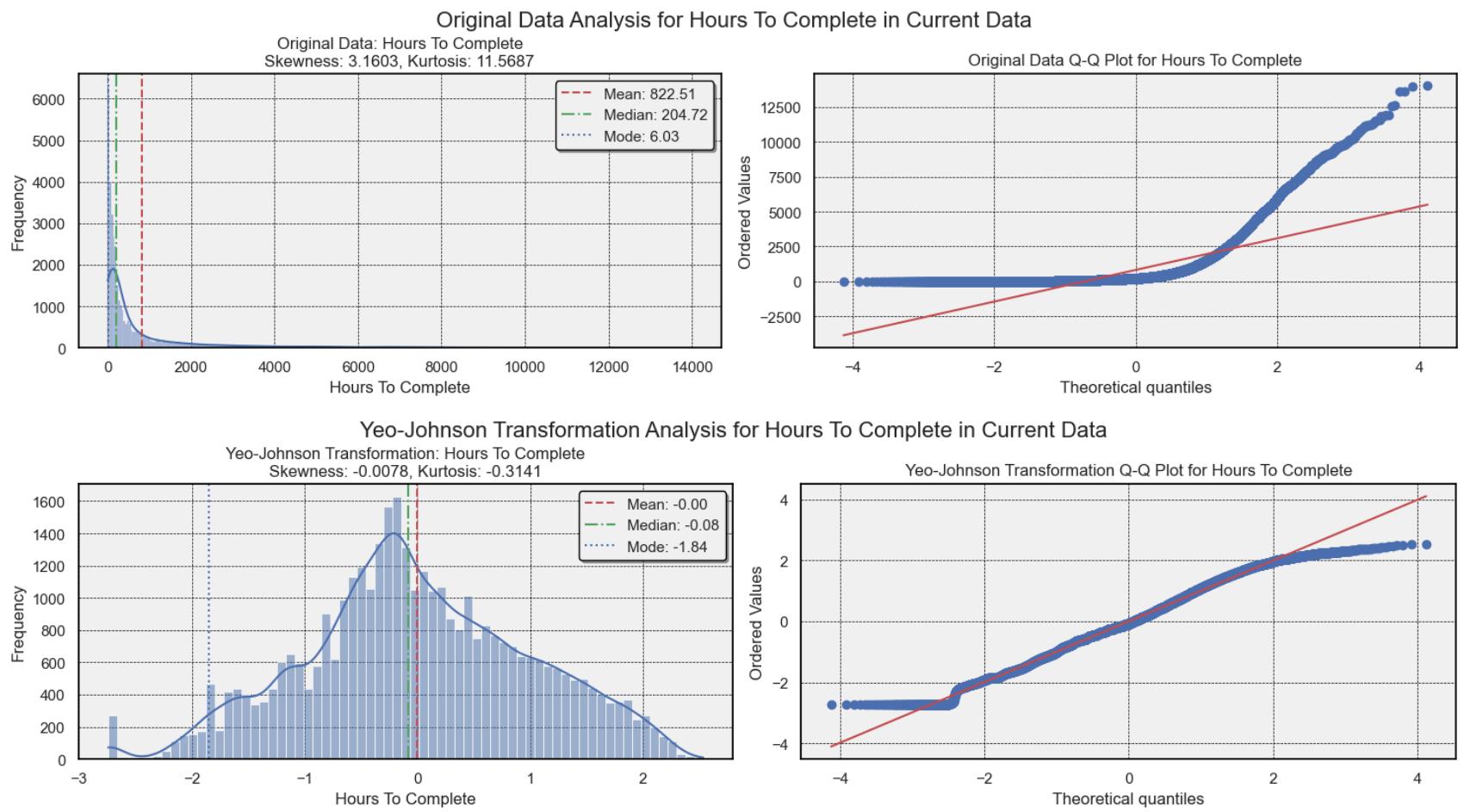
- Objective 1.5 of the Project Plan
- Section 4 of the Data Analysis Process

Section Contents

- Analysis of Hours to Complete for the Current Data
- Analysis of Hours to Complete for the Historic Data
- Analysis of Days Past Due for the Current Data
- Analysis of Days Past Due for the Historic Data

Transformation Analysis 1 - Analysis of Hours to Complete for the Current Data

In [23]: `am.transformation_analysis_1()`



Summary of Transformation Analysis for Hours To Complete in Current Data		
Transformation	Skewness	Kurtosis
Original	3.160301	11.568701
Log	-0.730245	2.122002
Square Root	1.652728	2.547439
Reciprocal	13.671169	197.054778
Box-Cox	0.015728	0.104887
Yeo-Johnson	-0.007844	-0.314056

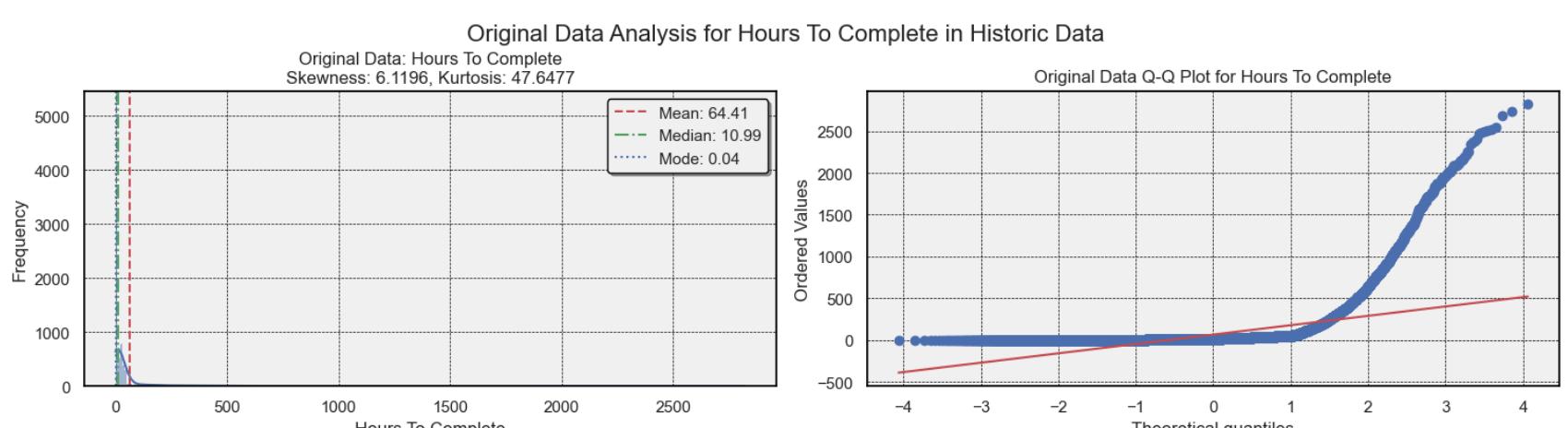
Transformation Results for Normalizing Hours To Complete in Current Data	
Metric	Result
Best Transformation	Yeo-Johnson
Improvement	Skewness reduced from 3.1603 to -0.0078

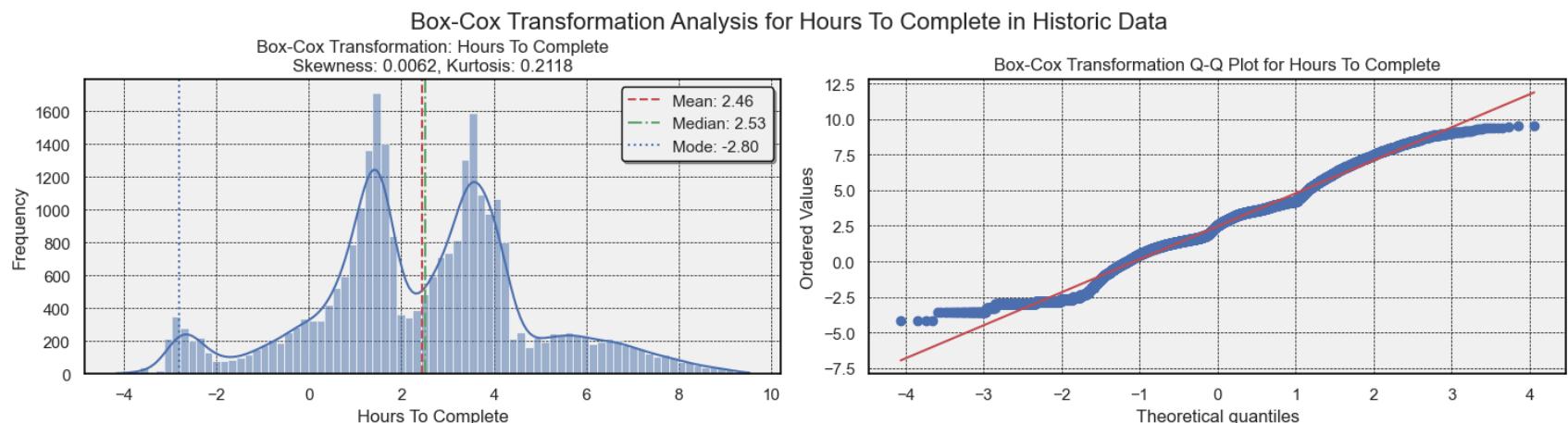
Normality Metrics Comparison for Hours To Complete in Current Data			
Metric	Original Data	Yeo-Johnson	Improvement
Shapiro-Wilk	0.000000	0.000000	Yes
D'Agostino's K^2	0.000000	0.000000	Yes
Kolmogorov-Smirnov	0.000000	0.000000	Yes
Skewness	3.160301	-0.007844	Yes
Kurtosis	11.568701	-0.314056	Yes

Overall Assessment of Transformation Analysis on Hours To Complete in Current Data	
Metric	Result
Analysis Results	5 out of 5 metrics show improvement
Transformation Results	Transformation improved normality
Final Conclusion	The Yeo-Johnson transformation significantly improved the dataset's normality

Transformation Analysis 2 - Analysis of Hours to Complete for the Historic Data

In [24]: `am.transformation_analysis_2()`





Summary of Transformation Analysis for Hours To Complete in Historic Data

Transformation	Skewness	Kurtosis
Original	6.119577	47.647690
Log	0.582592	-0.074652
Square Root	3.080892	11.673482
Reciprocal	1.770340	2.297795
Box-Cox	0.006209	0.211801
Yeo-Johnson	0.056021	-0.689088

Transformation Results for Normalizing Hours To Complete in Historic Data

Metric	Result
Best Transformation	Box-Cox
Improvement	Skewness reduced from 6.1196 to 0.0062

Normality Metrics Comparison for Hours To Complete in Historic Data

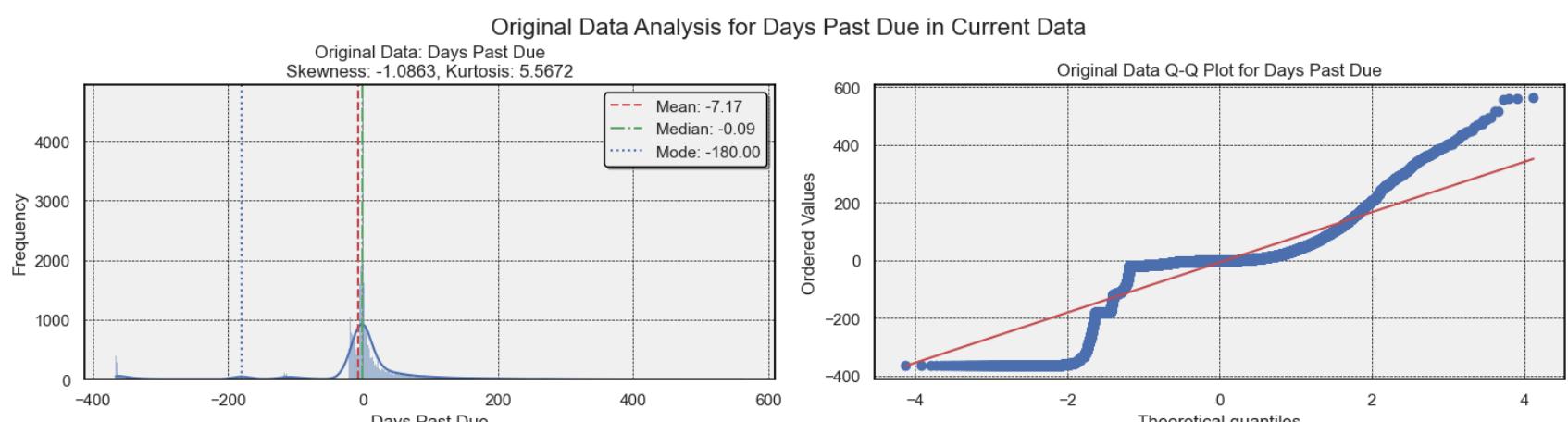
Metric	Original Data	Box-Cox	Improvement
Shapiro-Wilk	0.000000	0.000000	Yes
D'Agostino's K^2	0.000000	0.000000	Yes
Kolmogorov-Smirnov	0.000000	0.000000	Yes
Skewness	6.119577	0.006209	Yes
Kurtosis	47.647690	0.211801	Yes

Overall Assessment of Transformation Analysis on Hours To Complete in Historic Data

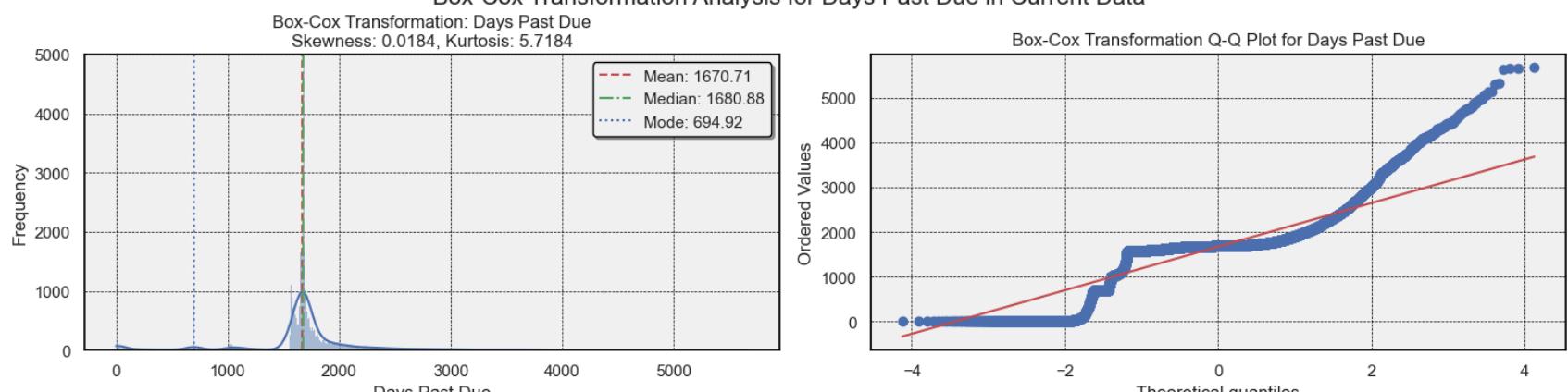
Metric	Result
Analysis Results	5 out of 5 metrics show improvement
Transformation Results	Transformation improved normality
Final Conclusion	The Box-Cox transformation significantly improved the dataset's normality

Transformation Analysis 3 - Analysis of Days Past Due for the Current Data

In [25]: `am.transformation_analysis_3()`



Box-Cox Transformation Analysis for Days Past Due in Current Data



Summary of Transformation Analysis for Days Past Due in Current Data

Transformation	Skewness	Kurtosis
Original	-1.086275	5.567246
Log	nan	nan
Square Root	nan	nan
Reciprocal	nan	nan
Box-Cox	0.018351	5.718447
Yeo-Johnson	0.324630	6.351324

Transformation Results for Normalizing Days Past Due in Current Data

Metric	Result
Best Transformation	Box-Cox
Improvement	Skewness reduced from -1.0863 to 0.0184

Normality Metrics Comparison for Days Past Due in Current Data

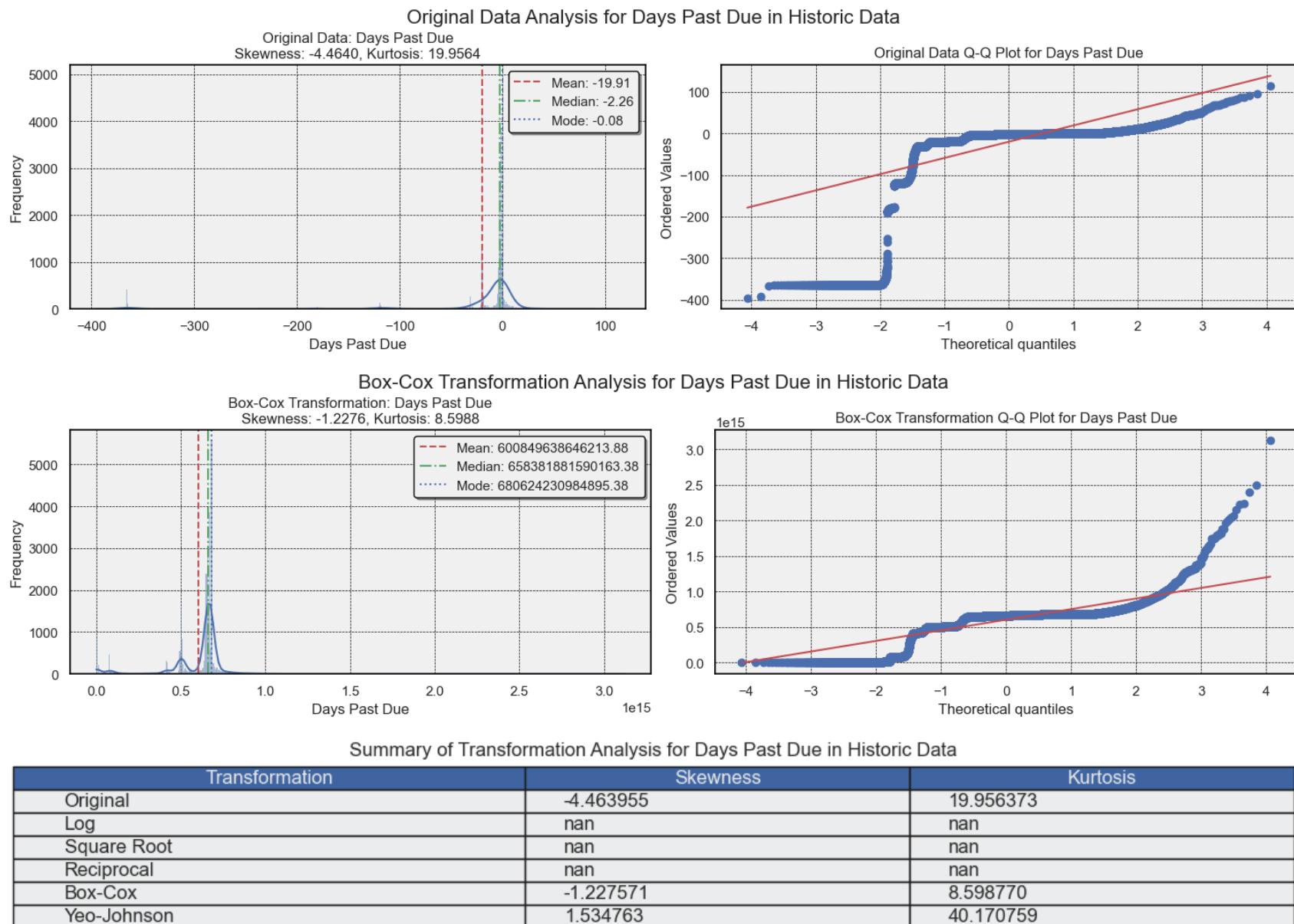
Metric	Original Data	Box-Cox	Improvement
Shapiro-Wilk	0.000000	0.000000	Yes
D'Agostino's K^2	0.000000	0.000000	No
Kolmogorov-Smirnov	0.000000	0.000000	No
Skewness	-1.086275	0.018351	Yes
Kurtosis	5.567246	5.718447	No

Overall Assessment of Transformation Analysis on Days Past Due in Current Data

Metric	Result
Analysis Results	2 out of 5 metrics show improvement
Transformation Results	Transformation did not substantially improve normality
Final Conclusion	Another transformation technique may be required

Transformation Analysis 4 - Analysis of Days Past Due for the Historic Data

In [26]: `am.transformation_analysis_4()`



Transformation Results for Normalizing Days Past Due in Historic Data

Metric	Result
Best Transformation	Box-Cox
Improvement	Skewness reduced from -4.4640 to -1.2276

Normality Metrics Comparison for Days Past Due in Historic Data

Metric	Original Data	Box-Cox	Improvement
Shapiro-Wilk	0.000000	0.000000	Yes
D'Agostino's K^2	0.000000	0.000000	No
Kolmogorov-Smirnov	0.000000	0.000000	No
Skewness	-4.463955	-1.227571	Yes
Kurtosis	19.956373	8.598770	Yes

Overall Assessment of Transformation Analysis on Days Past Due in Historic Data

Metric	Result
Analysis Results	3 out of 5 metrics show improvement
Transformation Results	Transformation improved normality
Final Conclusion	The Box-Cox transformation helped but did not fully normalize the dataset

Observations

Following the successful completion of the transformational analysis, the subsequent observations were documented.

- The Yeo-Johnson method produced the best results in transforming the current data with the 'Hours to Complete' attribute.
- The Box-Cox method yielded the best results in transforming the historic data with the 'Hours to Complete' attribute.
- The Box-Cox method prompted the best results in transforming the current data with the 'Days Past Due' attribute.

- The Box-Cox method generated the best results in transforming the historical data with the 'Days Past Due' attribute.
- The 'Hours to Complete' attribute transformed more effectively than the 'Days Past Due' attribute.

Section 7 - Independent T-Test Analysis

Purpose

The objective of this section involves implementing Independent Samples T-Tests on the dataset, assessing statistical significance by comparing current data to historical data. The findings were presented via box plots, violin plots, histograms, and tabular summaries.

This section aligns with the following:

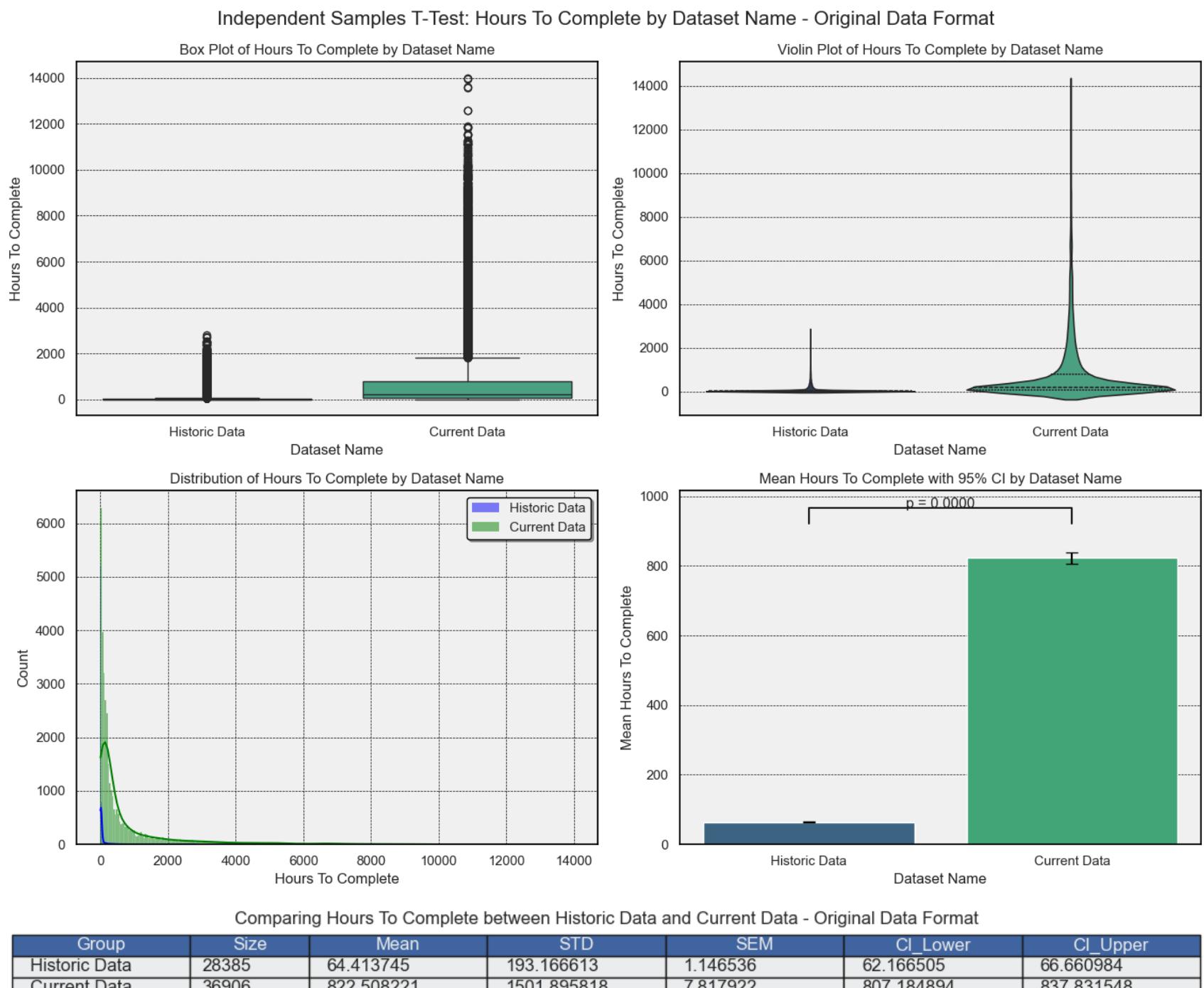
- Objective 1.6 of the Project Plan
- Section 5 of the Data Analysis Process

Section Contents

1. Analysis of Hours to Complete with the Original Data Format
2. Analysis of Days Past Due with the Original Data Format
3. Analysis of Hours to Complete with the Transformed Data Format
4. Analysis of Days Past Due for the Historic Data

Independent T-Test Analysis 1 - Analysis of Hours to Complete with the Original Data Format

In [27]: `am.independent_ttest_analysis_1()`



Independent Samples T-Test Results	
Statistic	Value
Mean Difference	758.09
Lower CI	742.61
Upper CI	773.58
T_Statistic	-95.9425
P_Value	0.0000
Cohen's d	0.7080
Interpretation	medium effect
Power	1.0000

Overall Assessment of Independent T-Test Analysis - Original Data Format

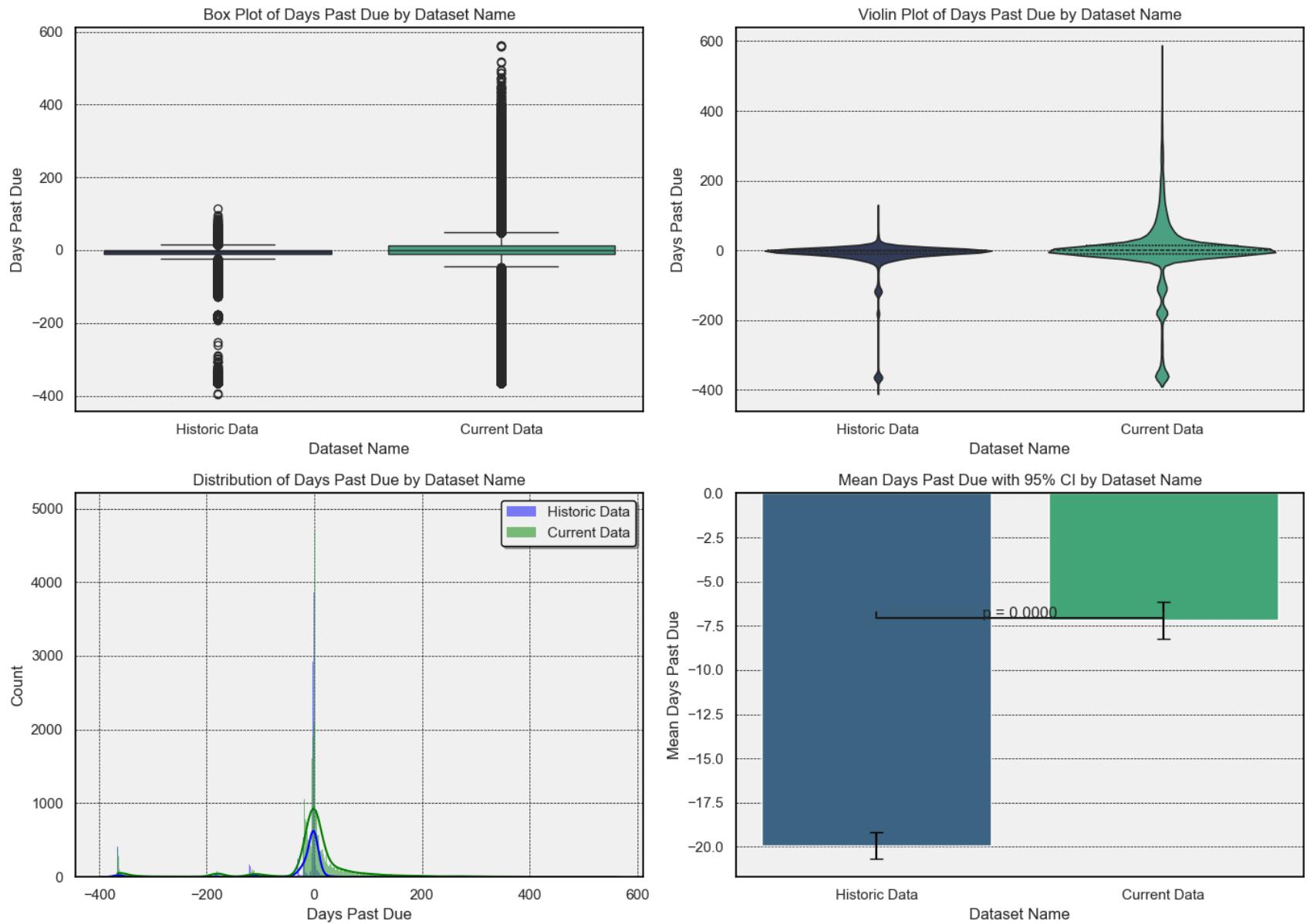
Interpretation

There is a statistically significant difference in Hours To Complete between Historic Data and Current Data

Independent T-Test Analysis 2 - Analysis of Days Past Due with the Original Data Format

In [28]: `am.independent_ttest_analysis_2()`

Independent Samples T-Test: Days Past Due by Dataset Name - Original Data Format



Comparing Days Past Due between Historic Data and Current Data - Original Data Format

Group	Size	Mean	STD	SEM	CI_Lower	CI_Upper
Historic Data	28385	-19.908372	65.248585	0.387281	-20.667445	-19.1493
Current Data	36906	-7.17134	102.342717	0.532732	-8.215494	-6.127185

Independent Samples T-Test Results

Statistic	Value
Mean Difference	12.74
Lower CI	11.45
Upper CI	14.03
T_Statistic	-19.3388
P_Value	0.0000
Cohen's d	0.1484
Interpretation	negligible effect
Power	1.0000

Overall Assessment of Independent T-Test Analysis - Original Data Format

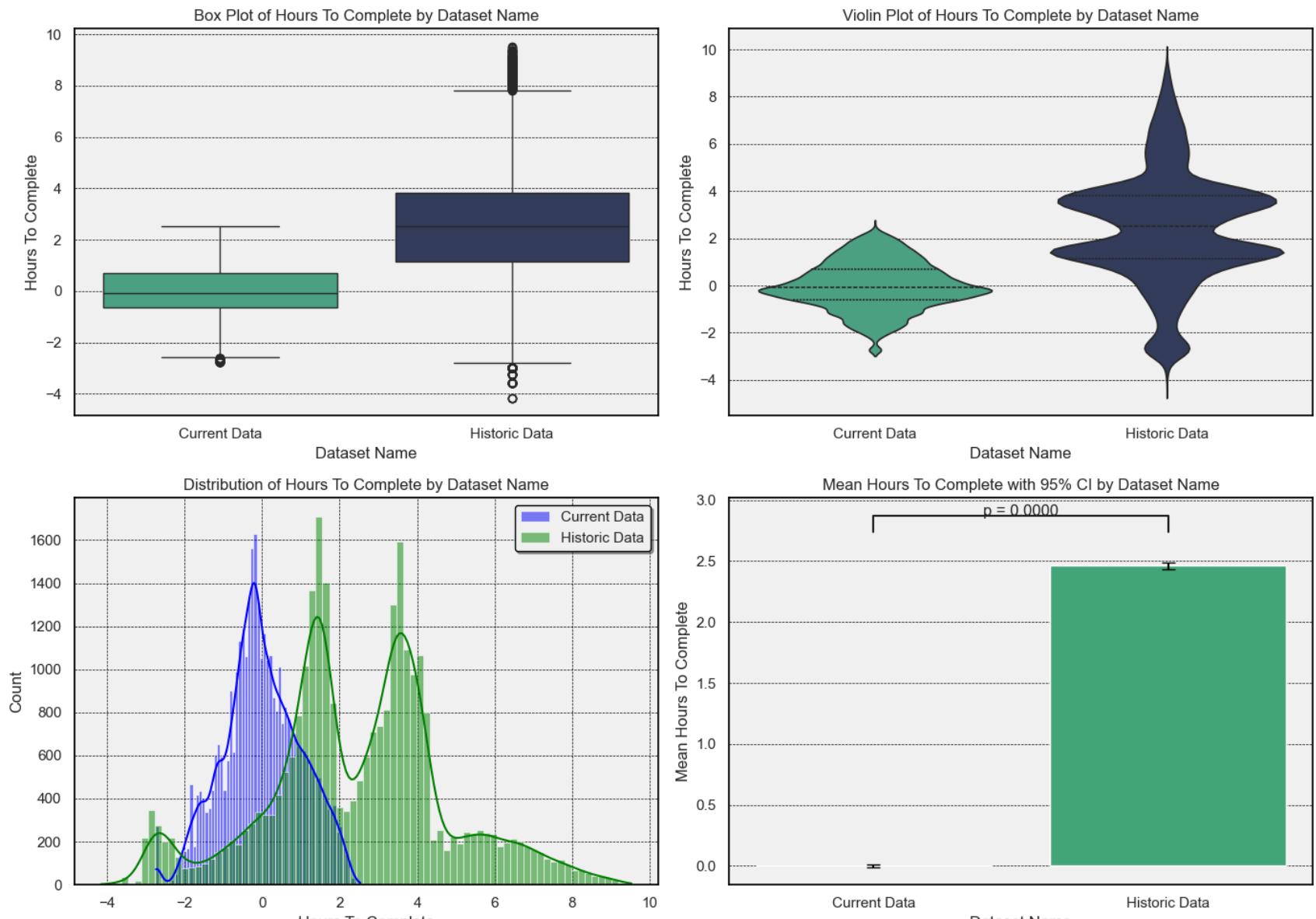
Interpretation

There is a statistically significant difference in Days Past Due between Historic Data and Current Data

Independent T-Test Analysis 3 - Analysis of Hours to Complete with the Transformed Data Format

In [29]: `am.independent_ttest_analysis_3()`

Independent Samples T-Test: Hours To Complete by Dataset Name - Transformed Data Format



Comparing Hours To Complete between Current Data and Historic Data - Transformed Data Format

Group	Size	Mean	STD	SEM	CI_Lower	CI_Upper
Current Data	36906	-0.0	1.000014	0.005205	-0.010203	0.010203
Historic Data	28385	2.460473	2.334476	0.013856	2.433315	2.487632

Independent Samples T-Test Results

Statistic	Value
Mean Difference	2.46
Lower CI	2.43
Upper CI	2.49
T_Statistic	-166.2286
P_Value	0.0000
Cohen's d	1.3701
Interpretation	large effect
Power	1.0000

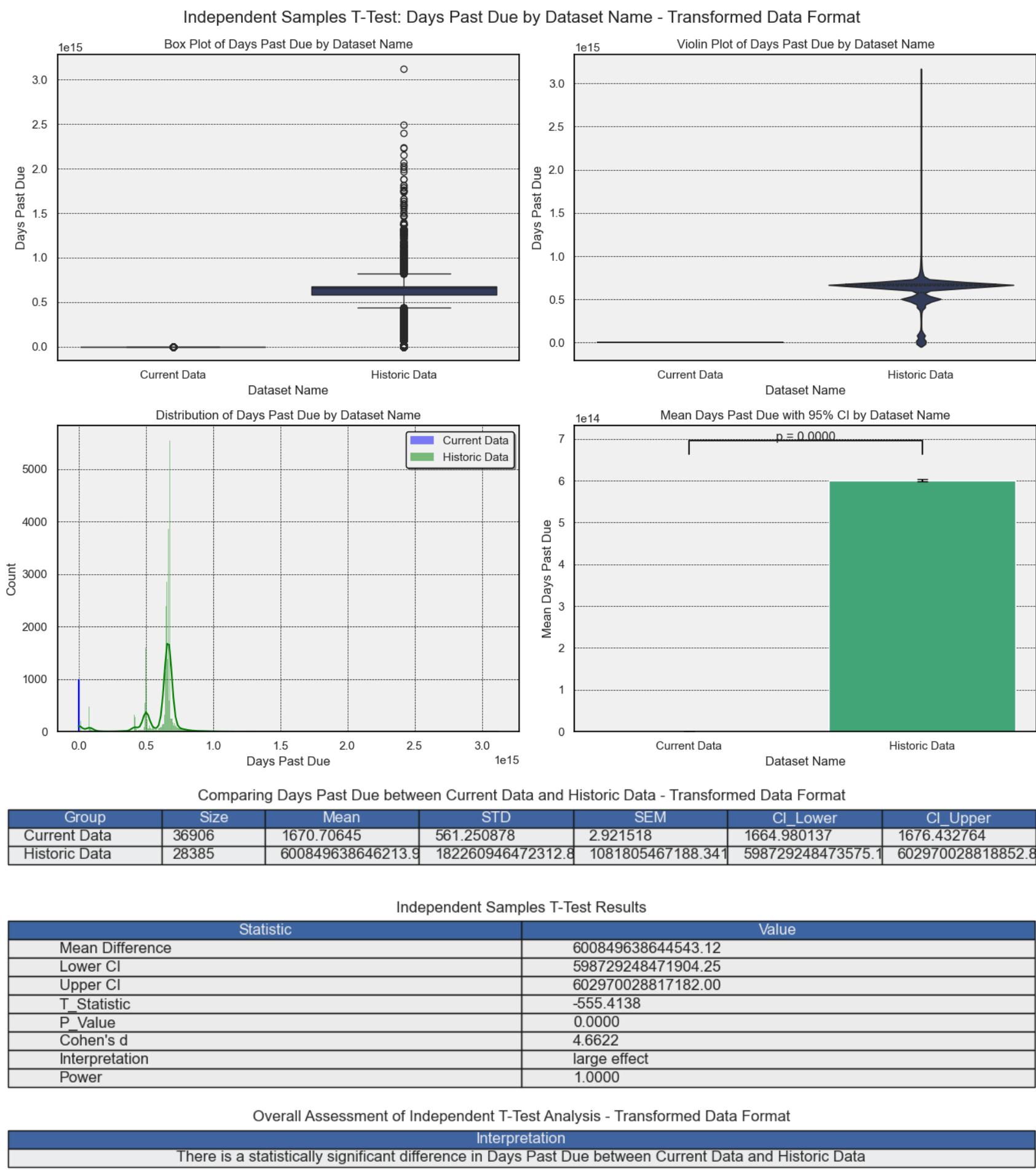
Overall Assessment of Independent T-Test Analysis - Transformed Data Format

Interpretation

There is a statistically significant difference in Hours To Complete between Current Data and Historic Data

Independent T-Test Analysis 4 - Analysis of Days Past Due with the Transformed Data Format

In [30]: `am.independent_ttest_analysis_4()`



Observations

Following the successful completion of the Independent Samples T-Test analysis, the subsequent observations were documented.

- The p-value for the T-Tests was less than 0.000000 for both original and transformed versions of the data in the 'Hours to Complete' and 'Days Past Due' attributes when comparing the current and historical datasets.
- The effect size, Cohen's d, is significant for both the 'Hours to Complete' attribute and the 'Days Past Due' attribute.
- There is a statistically significant difference in both original and transformed versions of the data for the 'Hours to Complete' and 'Days Past Due' attributes when comparing the current and historical datasets.
- These results justify the rejection of the null hypothesis in favor of the alternative hypothesis.

Section 8 - Mann-Whitney U Test Analysis

Purpose

The objective of this section involves conducting a comprehensive non-parametric analysis of the data utilizing the Mann-Whitney U Test to evaluate statistical differences between current and historical datasets. The results are illustrated through box plots, violin plots, histograms, ECDF plots, and tabular summaries.

This section aligns with the following:

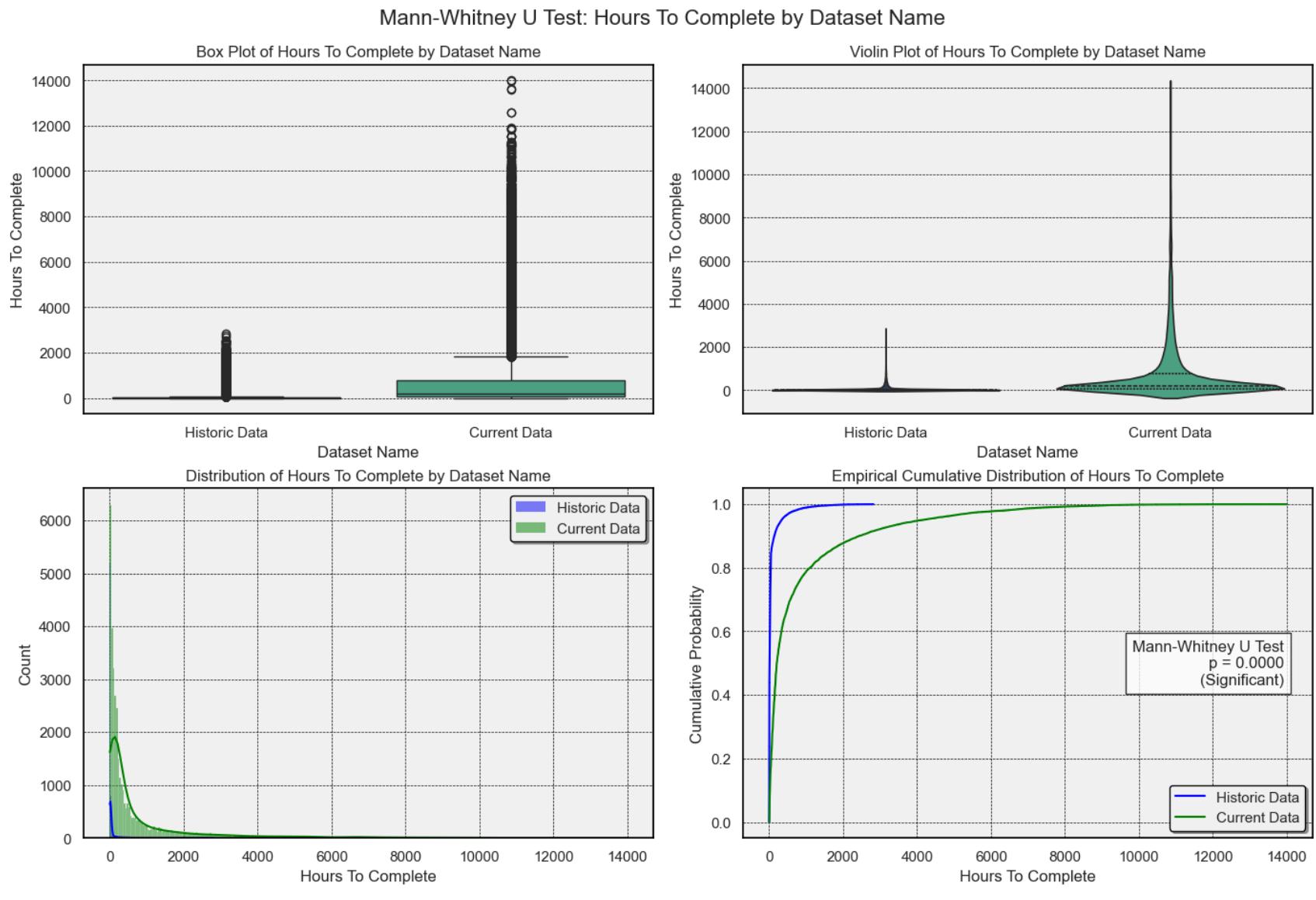
- Objective 1.7 of the Project Plan
- Section 6 of the Data Analysis Process

Section Contents

1. Analysis of Hours to Complete
2. Analysis of Days Past Due

Mann-Whitney U Test Analysis 1 - Analysis of Hours to Complete

In [31]: `am.mannwhitney_utest_analysis_1()`



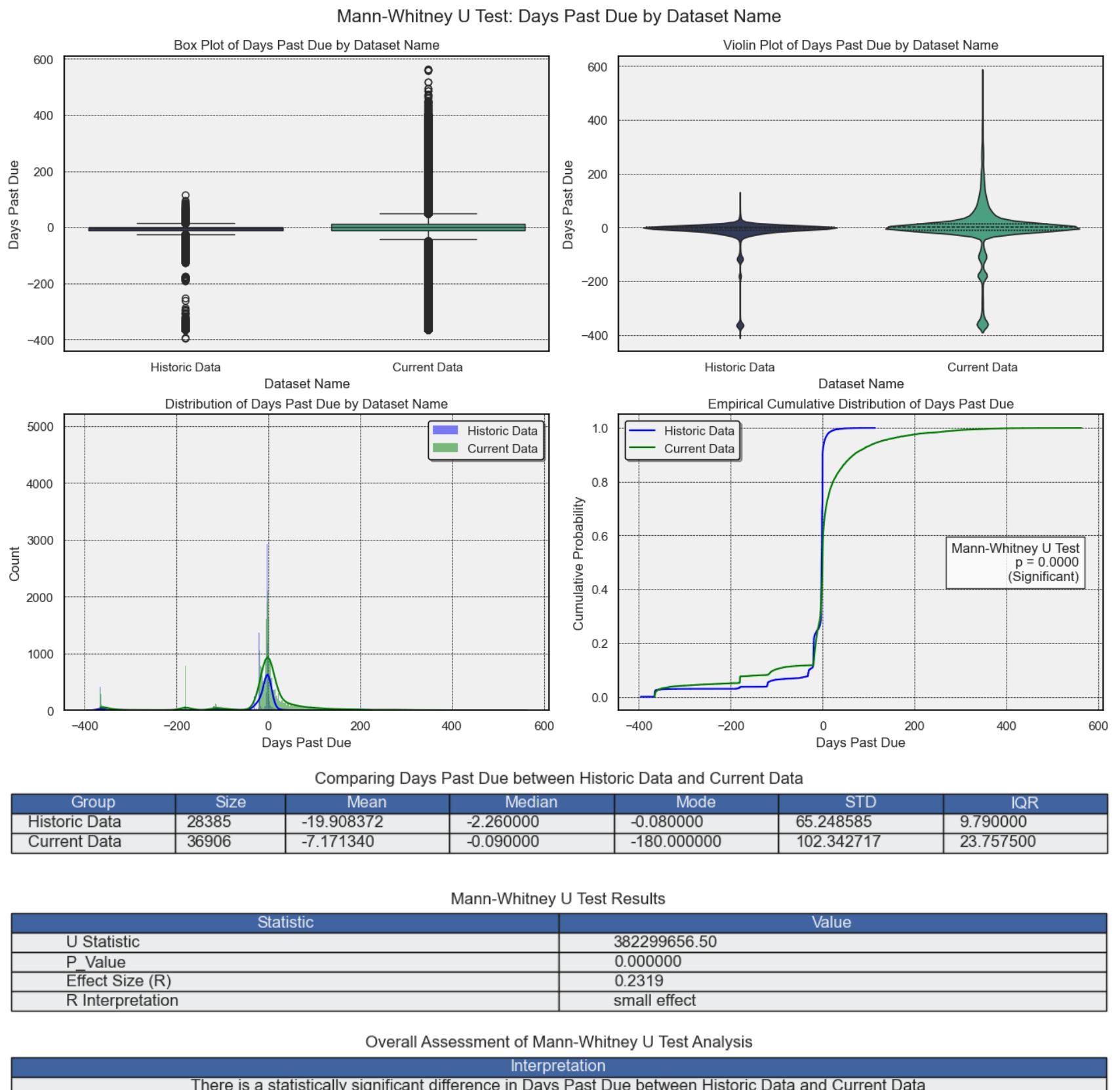
Comparing Hours To Complete between Historic Data and Current Data						
Group	Size	Mean	Median	Mode	STD	IQR
Historic Data	28385	64.413745	10.990000	0.040000	193.166613	31.050000
Current Data	36906	822.508221	204.725000	6.030000	1501.895818	704.235000

Mann-Whitney U Test Results	
Statistic	Value
U Statistic	133624984.50
P_Value	0.000000
Effect Size (R)	0.6396
R Interpretation	large effect

Overall Assessment of Mann-Whitney U Test Analysis	
Interpretation	
There is a statistically significant difference in Hours To Complete between Historic Data and Current Data	

Mann-Whitney U Test Analysis 2 - Analysis of Days Past Due

In [32]: `am.mannwhitney_utest_analysis_2()`



Observations

Following the successful completion of the Mann-Whitney U Test analysis, the subsequent observations were documented.

- The p-value for the U-Tests was less than 0.000000 for both the 'Hours to Complete' and 'Days Past Due' attributes when comparing the current and historical datasets.
- The effect size R is significant for the 'Hours to Complete' attribute; however, it is relatively small for the 'Days Past Due' attribute.
- There is a statistically significant difference in both the 'Hours to Complete' and 'Days Past Due' attributes when comparing the current and historical datasets.
- These results justify the rejection of the null hypothesis in favor of the alternative hypothesis.

Section 9 - Parametric and Nonparametric Analysis

Purpose

The objective of this section is to evaluate the effectiveness of parametric and non-parametric tests applied to the dataset, to determine the most appropriate analytical approach. Results are systematically presented in tabulated form.

This section aligns with the following:

- Objective 1.8 of the Project Plan
- Section 7 of the Data Analysis Process

Section Contents

1. Analysis of Hours to Complete
2. Analysis of Days Past Due

Parametric and Nonparametric Analysis 1 - Analysis of Hours to Complete

In [33]: `am.parametric_nonparametric_analysis_1()`

Comparing Parametric and Non-Parametric Tests for Hours To Complete

Analysis of Normality Test on Historic Data	
Statistic	Value
Test	Shapiro-Wilk
Statistic	0.439562
P_Value	0.000006
Interpretation	Non-Normal Distribution

Analysis of Equal Variance Test - Original	
Statistic	Value
Test	Levene's Test
P_Value	0.000000
Interpretation	Unequal Variances

Analysis of Equal Variance Test - Transformed	
Statistic	Value
Test	Levene's Test
P_Value	0.000000
Interpretation	Unequal Variances

Analysis of Non-Parametric Test	
Statistic	Value
Test	Mann-Whitney U Test
U Statistic	133624984.5000
P_Value	0.000000
Effect Size R	0.6396
Interpretation	Significant

Overall Assessment of Parametric vs Non-Parametric Analysis		
Interpretation		
All tests produce similar results with respect to a statistical significance		
The p-values are comparable with each other in all tests		

Parametric and Nonparametric Analysis 2 - Analysis of Days Past Due

In [34]: `am.parametric_nonparametric_analysis_2()`

Comparing Parametric and Non-Parametric Tests for Days Past Due

Analysis of Normality Test on Historic Data	
Statistic	Value
Test	Shapiro-Wilk
Statistic	0.491771
P_Value	0.000005
Interpretation	Non-Normal Distribution

Analysis of Equal Variance Test - Original	
Statistic	Value
Test	Levene's Test
P_Value	0.000000
Interpretation	Unequal Variances

Analysis of Equal Variance Test - Transformed	
Statistic	Value
Test	Levene's Test
P_Value	0.000000
Interpretation	Unequal Variances

Analysis of Non-Parametric Test	
Statistic	Value
Test	Mann-Whitney U Test
U Statistic	382299656.5000
P_Value	0.000000
Effect Size R	0.2319
Interpretation	Significant

Statistical Analysis for Historic Data and Current Data		
Statistic	Historic Data	Current Data
Size	28385	36906
Mean	64.413745	822.508221
Median	10.990000	204.725000
Mode	0.040000	6.030000
STD	193.166613	1501.895818
IQR	31.050000	704.235000

Overall Assessment of Parametric vs Non-Parametric Analysis		
Interpretation		
All tests produce similar results with respect to a statistical significance		
The p-values are comparable with each other in all tests		

Interpretation
All tests produce similar results with respect to a statistical significance
The p-values are comparable with each other in all tests

Observations

Following the successful completion of the parametric and non-parametric analysis, the subsequent observations were documented.

- The current and historical data are non-normally distributed for both 'Hours to Complete' and 'Days Past Due' attributes.
- The p-value for the T-Tests was less than 0.000000 with the original version of the data for both the 'Hours to Complete' and 'Days Past Due' attributes when comparing the current and historical datasets.
- The p-value for the T-Tests was less than 0.000000 with the transformed versions of the data for both the 'Hours to Complete' and 'Days Past Due' attributes when comparing the current and historical datasets.
- The p-value for the U-Tests was less than 0.000000 for both the 'Hours to Complete' and 'Days Past Due' attributes when comparing the current and historical datasets.
- All tests yielded similar results regarding statistical significance.
- These findings conclusively support the rejection of the null hypothesis in favor of the alternative hypothesis.

Section 10 - Chi-Square Analysis

Purpose

The objective of this section is to conduct a Chi-Squared Analysis of the dataset to determine whether there are statistically significant deviations between observed and expected outcomes. The analysis results are presented using a heatmap visualization and summarized in comprehensive statistical tables.

This section aligns with the following:

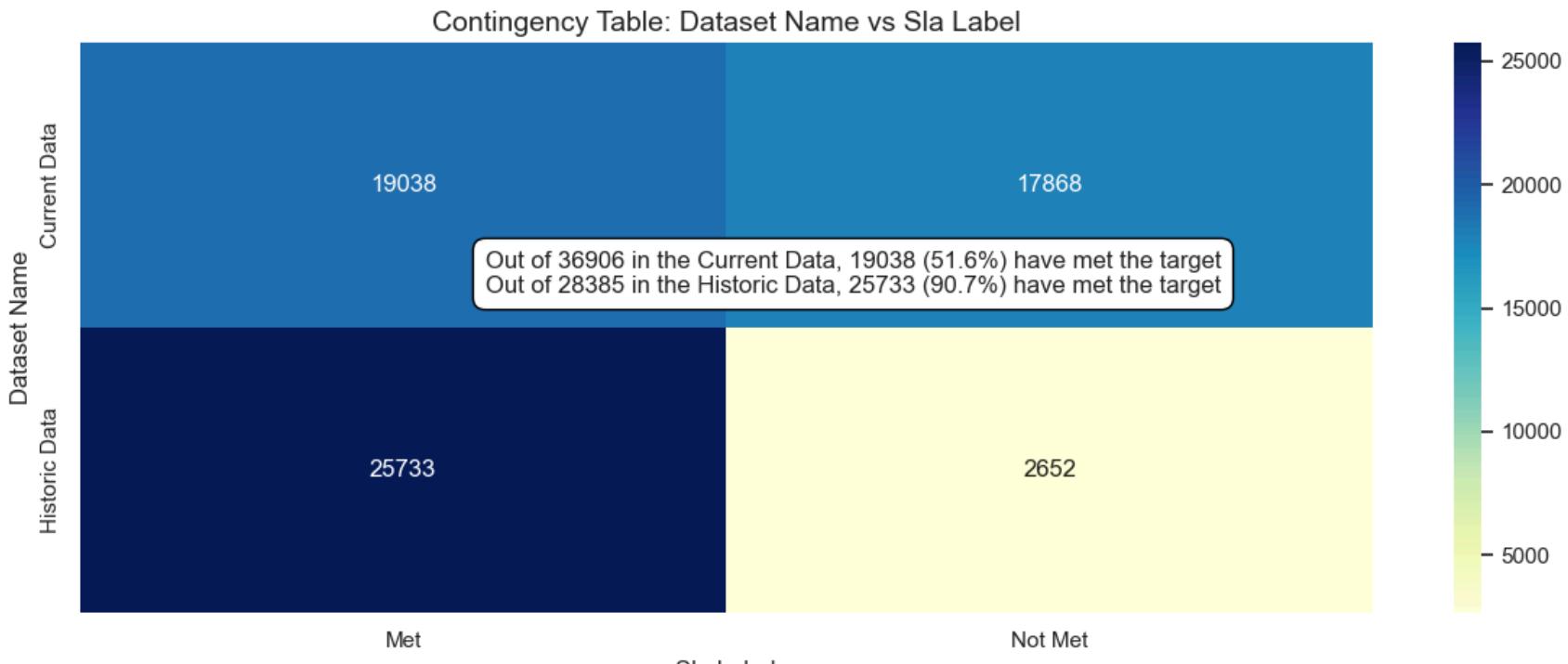
- Objective 1.9 of the Project Plan
- Section 8 of the Data Analysis Process

Section Contents

1. Analysis of Dataset Name and SLA Label

Chi-Square Analysis 1 - Analysis of Dataset Name and SLA Label

In [35]: `am.chi_square_test_analysis_1()`



Analysis of Expected Counts			
Dataset Name	Met	Not Met	Total
Current Data	25306.99	11599.01	36906.0
Historic Data	19464.01	8920.99	28385.0

Analysis of Observed Row Percentages

Dataset Name	Met	Not Met	Total
Current Data	51.59	48.41	100.0
Historic Data	90.66	9.34	100.0

Analysis of Chi-Square Statistic

Dataset Name	Sla Label	Expected Results	Observed Results	Chi-Square Statistic
Current Data	Met	25307.0	19038	1552.94
Current Data	Not Met	11599.0	17868	3388.24
Historic Data	Met	19464.0	25733	2019.12
Historic Data	Not Met	8921.0	2652	4405.36

Sum of All Chi-Square Statistics

Metric	Result
Sum Statistics	11365.66

Analysis of Chi-Square Test Results

Statistic	Value
Chi-Square Statistic	11365.66
Degrees of Freedom	1
P_Value	0.000000
Interpretation	Reject H ₀

Overall Assessment of Chi-Square Test Analysis

Interpretation
There is a statistically significant association between Dataset Name and Sla Label

Observations

Following the successful completion of the Chi-Squared analysis, the subsequent observations were documented.

- The historical data met the SLA target 90.7% of the time compared to only 51.6% for the current data.
- The p-value from the test was less than 0.000000, which is substantially lower than the pre-established alpha of 0.05.
- The findings provide sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis.
- The analysis concluded that a statistically significant association exists between the 'data-name' and the 'SLA-label' attributes.

Conclusion

This project aimed to evaluate the impact of implementing volume bid contracts on service-level agreements that oversee the maintenance of data center equipment. Specifically, this study examined whether the use of volume bid contracts adversely affected the average repair times of equipment in data centers compared with conventional contracting methods.

The study demonstrated that a statistically significant difference exists in the means of both attributes when comparing current and historical contract data. This conclusion was reached by applying the Independent Samples T-Test to both the original and transformed data sets, and the Mann-Whitney U test to the original data from both contract models. All tests yielded probability values approximating zero, indicating that the observed differences are unlikely due to random variation.

An additional examination of the data, through the application of chi-squared analysis, demonstrated that there was a statistically significant association between current and historical data and the 'SLA label' attribute. The analysis results indicated that the current contract model for maintenance tasks achieved their SLA objectives only 51.6% of the time, in contrast to the 90.7% success rate observed in the historical contract model.

From a pragmatic standpoint, recognizing the existence of such a discrepancy indicates an issue affecting the overall effectiveness of the current contractual strategy. The observed decline in efficiency suggests a diminished ability among service providers to fulfill the organization's service level agreements.

In []: