**Box Plot DR and CD:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Department | Median | Highest cycle time | Lowest Cycle time | Range |
| CD | 4.9 | 6.2 | 2.3 | 3.9 |
| DR | 1.6 | 1.8 | 1.2 | 0.6 |

From the box plot below, I can infer the following:

* CD = median of 4.9 (meaning that 50% of the time, reports are cleansed in 4.9 days)
* DR = median of 1.6 (meaning that 50% of the time, reports are cleansed in 1.6 days)

My interpretation would be that out of all the 2 activities, the CD activity has the greatest variation or inconsistencies because of the range. Most of the data is bunched up to the right which means it is negatively skewed with a left tail and we have more data observations between the median and the upper quartile. Although the DR activity exhibits an outlier, which is showing that there is a data point that deviates from other members of the DR sample set. However, instead on focusing on the outlier to understand why it is removed, the analysis will focus on CD so as to understand the variation.

**Histograms CD & DR:**

CD data is not normally distributed which I can tell from the P-value and the shape of the curve which is flat. However, for DR, P-value >= 0.05 which means it is normally distributed and it has a bell shaped curve.

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | **Time (Days) - Tasks: CD  Count = 126 Mean = 4.444 StDev = 2.240 Range = 8.22  Minimum = 0.205645 25th Percentile (Q1) = 2.281 50th Percentile (Median) = 4.880 75th Percentile (Q3) = 6.230 Maximum = 8.43  95% CI Mean = 4.05 to 4.84 95% CI Sigma = 1.99 to 2.56  Anderson-Darling Normality Test: A-Squared = 2.391; P-Value = 0.0000** |
|  | **Time (Days) - Tasks: DR  Count = 376 Mean = 1.599 StDev = 0.482083 Range = 2.55  Minimum = 0.441744 25th Percentile (Q1) = 1.254 50th Percentile (Median) = 1.616 75th Percentile (Q3) = 1.942 Maximum = 2.990593  95% CI Mean = 1.55 to 1.65 95% CI Sigma = 0.45 to 0.52  Anderson-Darling Normality Test: A-Squared = 0.412329; P-Value = 0.3376** |

**Multi-Vari Chart CD & DR:**

Using multi-vari charts to understand time-time variation (temporal variation), it illustrates that the CD exhibits more variation than DR task hence is the dominant source of variation. It shows a higher mean value of 4.5 with a skew to the left i.e. a skew to the lower values. DR task looks much more symmetric.

**Variable Data Analysis and Interpretation (Departments):**

Looking at the 5 departments:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Department | Median | Highest cycle time | Lowest Cycle time | Range |
| Exploration | 3.6 | 5.4 | 0.7 | 4.7 |
| Finance | 2.8 | 5.9 | 0.8 | 5.1 |
| Production Dept | 6 | 8.5 | 0.2 | 8.3 |
| Production Planning | 4.5 | 6.9 | 0.2 | 6.7 |
| Sales | 2.8 | 4.8 | 0.4 | 4.4 |

From the box plot below, I can infer the following:

* Exploration = 3.6 (meaning that 50% of the time, reports are cleansed in 3.6 days)
* Finance = 2.8 (meaning that 50% of the time, reports are cleansed in 2.8 days)
* Product depart = 6 (meaning that 50% of the time, reports are cleansed in 6 days);
* Production Plan = 4.5 (meaning that 50% of the time, reports are cleansed in 4.5 days)
* Sales = 2.8 (meaning that 50% of the time, reports are cleansed in 2.8 days)

My interpretation would be that out of all the 5 departments, the production department has the greatest variation or inconsistencies because of the range. Most of the data is bunched up to the right which means it is negatively skewed with a left tail and we have more data observations between the median and the upper quartile.