NOTES AND CONSIDERATIONS

Initial Observations regarding the provided data:

- Data is for period 12/06 to 11/08. There are blank cells mostly corresponding to Sat and Sun.
- Total Sample Size: 61 Days. Usable Sample Size with data: 41 Days
- No associate has worked on Sundays. Only Raj has worked on one Saturday.
- Arya has not missed any Daily Team Review.
- On Weekdays with no data on Leads and Daily Team Review, it seems that the associates have taken Leave.
- The No. of Incomplete Leads are mostly zero with maximum no. of values for Raj followed with Arya and Ali.
- A general trend of Leads increasing with Time Spent can be seen if the data is arranged in ascending order.

1. **Data Cleaning Method:**

- The original data was kept unchanged in the process so as to create a universal way that will facilitate report generation even if the base data is changed.
- A new column was created "Daily Team Review Clean", where for blank data, days when Leads were generated were marked as "Missed", days when Leads were zero and was not Sat OR Sun were marked as "Leave", else if Sat or Sun then were marked as "Weekend".

  Formula: =IF(ISBLANK(@[Daily Team Review]), IF(@[Leads]>0, "Missed", IF(OR(@[Day]="SAT",@[Day]="SUN"),"Weekend", "Leave")),@[Daily Team Review])

- All blank data for leads and time on Weekends and Leave were kept blank unless required for any calculation where they were considered as zero.
- All other blank data for Attended and Missed days should be considered zero. E.g. Ali on 27/07 and Raj on 01/08. However, these data (Leads, Time Spent) = (0,0) are outliers as they are extreme values and it skews the calculations and predictions. So, they were not considered.
- All the calculations are done considering only days when the associates were working. Blank data on Attended, Missed, Weekdays, and Leave were not considered.

2. **Outliers**:

The outliers were found using the quartile functions.

Leads	**Raj**	**Arya**	**Ali**	**Formula**
Quartile 1 (QTL 1)	10	10	10	=QUARTILE.INC(@[Leads],1)
Quartile 3 (QTL 3)	12	12	15	=QUARTILE.INC(@[Leads],3)
Inter Quartile Range (IQR)	2	2	5	=QTL 3 – QTL 1
Lower Limit (LL)	7	7	2.5	=QTL 1 – IQR * 1.5
Upper Limit (UL)	15	15	22.5	=QTL 3 + IQR * 1.5

The Outliers of Leads were sorted using the formula: = OR(@[Leads]<LL, @[Leads]>UL)

Time Spent	**Raj**	**Arya**	**Ali**	**Formula**
Quartile 1 (QTL 1)	240	123	170	=QUARTILE.INC(@[Time spent on LG (mins)],1)
Quartile 3 (QTL 3)	280	150	260	=QUARTILE.INC(@[Time spent on LG (mins)],3)
Inter Quartile Range (IQR)	40	27	90	=QTL 3 – QTL 1
Lower Limit (LL)	180	82.5	35	=QTL 1 – IQR * 1.5
Upper Limit (UL)	340	191	395	=QTL 3 + IQR * 1.5

The Outliers of Time Spent were sorted using the formula: = OR(@[Time spent on LG (mins)]<LL, @[Time spent on LG (mins)]>UL)

Instead of just removing the outlies from the data, they were normalized by changing them to the nearest whole number lead just below the UL or above the LL. The time spent corresponding to these outlier leads was also modified using the Average Time Spent on LG and vice versa for outlier time spent. This new cleaned data was used for regression in order to get better forecast of associate data.

3. Regression:

The cleaned data after normalizing the outlier Leads and Time Spent was used for calculating regression. As the leads cannot be nonzero if the time spent on LG is zero, the intercept was set to zero. This was also supported by the fact that the R^2 value for zero intercept is near 1 indicating a better fit of the model to the data.

4. Correlation:

"Correlation does not imply Causation."

To find the correlation between the between the average time per lead and the total number of leads generated per day for each associate, Pearson's correlation coefficient was used and p-value was used to find if the result is statistically significant.

Pearson's correlation coefficient (r):

- -1 indicates a perfectly negative linear correlation between two variables.
- 0 indicates no linear correlation between two variables.
- 1 indicates a perfectly positive linear correlation between two variables.

P-value:

- If the p-value is low (generally less than 0.05), then the calculated correlation is statistically significant, and we can use the calculated Pearson coefficient.
- If the p-value is not low (generally higher than 0.05), then the calculated correlation is not statistically significant (it might have happened just by chance) and we should not rely upon your Pearson coefficient.

For calculating Pearson's correlation coefficient and P-value, the following methods can be used. Here sample size is 41 as we have data for 41 days for each associate.

Formulas:

Pearson (r)	=PEARSON(@[Avg Time Per Lead (mins)],@[Leads])
t-stat	=r*SQRT(Sample Size-2)/SQRT(1-r^2)
P-value	= =T.DIST.2T(ABS(t)),Sample Size-2)

Data Analysis Tab:

Pearson (r)	Regression, t-Test: Paired Two Sample for Means
P-value	Regression, t-Test: Paired Two Sample for Means

5. Miscellaneous:

- Same colors are used for each associate throughout the report for easy comparison of each associate with themselves and others.
- In the Time Management Analysis all associate data is plotted on the same graph in order to facilitate visual comparison of the analysis for each associate.