Associate Performance Data Analysis Report

Key Concepts & Libraries Used

The analysis involved cleaning and processing the raw Excel dataset, followed by statistical and visual exploration using Python. Key libraries:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

Methodology

1. Data Collection and Cleaning

- Merged three employee sheets into one dataset.
- Handled missing values and standardized column names.
- Assessed outliers
- Converted date to datetime and extracted relevant features (e.g., weekday/weekend).

2. Exploratory Data Analysis (EDA)

- Box plots, bar charts, and line plots to visualize performance.

3. Statistical Measures

- Standard deviation for performance variability.
- Coefficient of variation for consistency.
- Linear regression to model and predict lead outcomes.

Results & Insights

1. Lead Generation Efficiency

Employee_Name	Leads	Time spent on LG (mins)	Efficiency (Leads per Minute)
Arya	474.0	5569.0	0.085114
Ali	488.0	9235.0	0.052842
Raj	447.0	10585.0	

2. Daily Performance Variability

Employee with highest daily performance variability:

Employee_Name   Ali
Lead Std Dev   3.039777

3. Time Management Analysis

Correlation Coefficient (Avg Time Per Lead vs Leads) for each employee:
Raj: -0.34
Arya: -0.51
Ali: -0.36

4. Compare the average number of leads generated on days when daily team reviews were attended versus missed for each associate.

Average Leads on Review Days (Attended vs Missed) and % Difference:

Daily Team Review Attended Missed % Difference (Missed vs Attended)

Employee_Name

Ali	11.92	11.0	-7.76
Arya	11.56	NaN	NaN
Raj	10.92	10.0	

5. Incomplete Leads Reduction Over Time

{'Raj': np.float64(-0.019788172326712147),
 'Arya': np.float64(-0.012477852030637442),
 'Ali': np.float64(-0.006304323562649105)}

All three associates show improvement, as indicated by the negative slopes (fewer incomplete leads over time). However, the degree of improvement varies:

- Raj has shown the most significant reduction in incomplete leads.
- Arya shows a moderate improvement.
- Ali shows a slight, but still positive, trend toward reducing incomplete leads.

6. Performance Consistency

Coefficient of Variation for Daily Leads (lower = more consistent):

Employee_Name mean   std   CV

Ali	11.902439	3.039777	0.255
Arya	11.560976	2.156951	0.187
Raj	10.900000	2.687101	0.247

Most consistent performer: Arya (CV = 0.187)

7. High-Performance Days

Average Time Spent on Top 10% Lead Days:

Raj: 314.0 mins

Arya: 161.2 mins

Ali: 251.43 mins

8. Impact of Longer Lead Generation Time

Estimated Optimal Time Thresholds (per employee):

Raj: 337.5 mins

Arya: 91.3 mins

Ali: 240.0 mins

Metric	Raj	Arya	Ali
Avg Leads	10.9	11.56	11.9
Std Dev (Leads)	2.69	2.15	**3.04**
CV (Leads)	0.247	**0.187**	0.255
Optimal Time (mins)	337.5	**91.3**	240.0
Avg Top 10% Time	314.0	161.2	251.4
Review Days vs Missed (% Diff)	-8.45%	-	-7.76%
Optimal Time Thresholds (mins)	337.5	91.3	240

9. Comparative Day Analysis

Average Leads on Weekdays vs Weekends

Employee_Name Weekday Weekend

Ali	11.90	NaN
Arya	11.56	NaN
Raj	10.77	16.0

10. Predictive Analysis

Lead Prediction Model Summary (per employee) [with Normalization]:

Raj: {'R²': 0.426, 'MSE': 4.04, 'Intercept': np.float64(10.9), 'Slope': np.float64(1.73)}

Arya: {'R²': 0.1, 'MSE': 4.09, 'Intercept': np.float64(11.56), 'Slope': np.float64(0.67)}

Ali: {'R²': 0.272, 'MSE': 6.56, 'Intercept': np.float64(11.9), 'Slope': np.float64(1.57)}

The model performs best for Raj, with a reasonably good R² (0.426) and low MSE (4.04).

Ali's model is moderately accurate, but the higher MSE (6.56) means predictions are less reliable.

Arya's model has a very low R² (0.10), indicating that factors other than time spent may be driving lead generation in Arya's case.

Performance Observations

- **Ali** is the most aggressive performer, but also has the highest variability.
- **Arya** is the most **consistent** and performs well with less time.
- **Raj** tends to generate better leads with extended hours (optimal: ~338 mins).
- Leads drop marginally when daily reviews are missed.
- Linear regression models show a moderate correlation between time and leads (R² ranging from 0.1 to 0.42).

Visual Observations (Power BI)

- **Performance Trend with Attendance**
  - **Insight**: Leads are consistently higher on days when team reviews are attended.
  - **Action**: Encourage regular attendance to boost productivity.
- **Time Distribution Analysis**

- **Insight**:
  - **Arya**: Most time-efficient and consistent.
  - **Raj**: Spends the most time, but inconsistently.
  - **Ali**: Moderate time, moderate variability.
- **Action**: Raj may need time-management support.
- **Time Spent on Leads (Heatmap)**
  - **Insight**: Most productive efforts fall in the **200–250 mins/day** range.
  - **Action**: Use this range as a benchmark.
- **Monthly Performance Comparison**
  - **Insight**: July was the peak month. Drop in August, especially for Raj.
  - **Action**: Investigate external factors affecting August.
- **Incomplete Leads Trend**
  - **Insight**: Incomplete leads mainly linked to **Ali**.
  - **Action**: Support Ali with quality control processes.

Tentative Projections for Next Month

Assuming average daily time investment and 22 working days:

Employee	Avg Time (mins/day)	Projected Leads
Raj	300	**430**
Arya	180	**299**
Ali	250	**406**

These are based on each associate's regression model using:

Leads = Slope × Time(hours) + Intercept

Recommendations

1. **Encourage Consistent Review Attendance:** Leads drop ~8% on missed review days.
2. **Optimize Time Allocation:**
   - Arya can produce more leads by slightly increasing effort/time.
   - Raj and Ali should be monitored to avoid burnout from long hours.
3. **Focus on Quality over Quantity:** Incomplete leads are relatively low, but tracking this over time can enhance client conversion.
4. **Continue Time-Based Monitoring:** Periodic recalibration of optimal time per employee helps maintain productivity.

Conclusion

This analysis provided a clear, actionable view of associate performance. Arya stands out in consistency, Ali in output, and Raj in time efficiency. The business development team can use these insights to set realistic expectations, fine-tune workload distributions, and promote practices that drive sustainable lead generation.