Data-Driven Analysis

Answer Predictive Questions

For HR Data Analysis

DATA DYNAMOS

Harnessing the Power of Data to Drive Innovation

Data-Driven Analysis

This report explains the entire analytical workflow, including data loading and preparation, KPI analysis, and turnover analysis. The code is organized into sections with clear objectives, supporting visualizations, and key insights.

Stage 1: Data Loading and Preprocessing

Objective

Load and prepare HR data from an Excel file, converting date columns appropriately and verifying the initial data samples.

Methodology:

- The Excel file HrData.xlsx is loaded using pandas and all necessary sheets (Employee, PerformanceRating, SatisfiedLevel) are read into separate DataFrames.
- Date columns such as HireDate and ReviewDate are converted to datetime format to facilitate time based analysis.
- The first few rows of the Employee and Performance Rating data are printed to verify that the data has been loaded correctly.

Results

A preview of the first few rows of both the Employee and Performance Rating data.

Confirmation that date columns are in datetime format.

```python
# Load the Excel file
file_path = 'HrData.xlsx'
xls = pd.ExcelFile(file_path)

# Read necessary sheets
employee_df = pd.read_excel(xls, sheet_name='Employee')
performance_df = pd.read_excel(xls, sheet_name='PerformanceRating')
df_satisfaction = pd.read_excel(xls, sheet_name='SatisfiedLevel')

# Convert date columns to datetime
employee_df['HireDate'] = pd.to_datetime(employee_df['HireDate'])
performance_df['ReviewDate'] = pd.to_datetime(performance_df['ReviewDate'])

# Verify data loading (optional)
print("Employee Data Sample:")
print(employee_df.head())
print("\nPerformance Rating Data Sample:")
print(performance_df.head())
```

```
Employee Data Sample:
  EmployeeID   FirstName LastName      Gender  Age BusinessTravel  \
0  3012-1A41    Leonelle    Simco      Female   30    Some Travel
1  CBCB-9C9D     Leonerd    Aland        Male   38    Some Travel
2  95D7-1CE9       Ahmed    Sykes        Male   43    Some Travel
3  47A0-559B  Ermentrude   Berrie  Non-Binary   39    Some Travel
4  42CC-040A       Stace   Savege      Female   29    Some Travel

        Department  DistanceFromHome State                Ethnicity  ...  \
0            Sales                27    IL                    White  ...
1            Sales                23    CA                    White  ...
2  Human Resources                29    CA  Asian or Asian American  ...
3       Technology                12    IL                    White  ...
4  Human Resources                29    CA                    White  ...

   MaritalStatus  Salary StockOptionLevel OverTime   HireDate  Attrition  \
0       Divorced  102059                1       No 2012-01-03         No
1         Single  157718                0      Yes 2012-01-04         No
2        Married  309964                1       No 2012-01-04         No
3        Married  293132                0       No 2012-01-05         No
4         Single   49606                0       No 2012-01-05        Yes

  YearsAtCompany YearsInMostRecentRole YearsSinceLastPromotion  \
0             10                     4                       9
1             10                     6                      10
...
1             Valid
2             Valid
3             Valid
4             Valid
```

Analysis 1: KPI Analysis

Objective

Measure HR performance using key performance indicators (KPIs), such as:

Employee Turnover Rate: The percentage of employees who have left the company.

Employee Satisfaction: The average satisfaction score calculated from various satisfaction metrics.

(Additional KPIs like Absenteeism Rate and Hiring Efficiency are mentioned as placeholders for future expansion.)

Methodology

Turnover Rate Calculation:

- Calculate the percentage of employees who have left (i.e., those with Attrition == 'Yes').

Employee Satisfaction Calculation:

- Identify the latest performance review for each employee based on the most recent ReviewDate.
- Compute an overall satisfaction score by averaging selected satisfaction columns (e.g., EnvironmentSatisfaction, JobSatisfaction, RelationshipSatisfaction, and WorkLifeBalance).

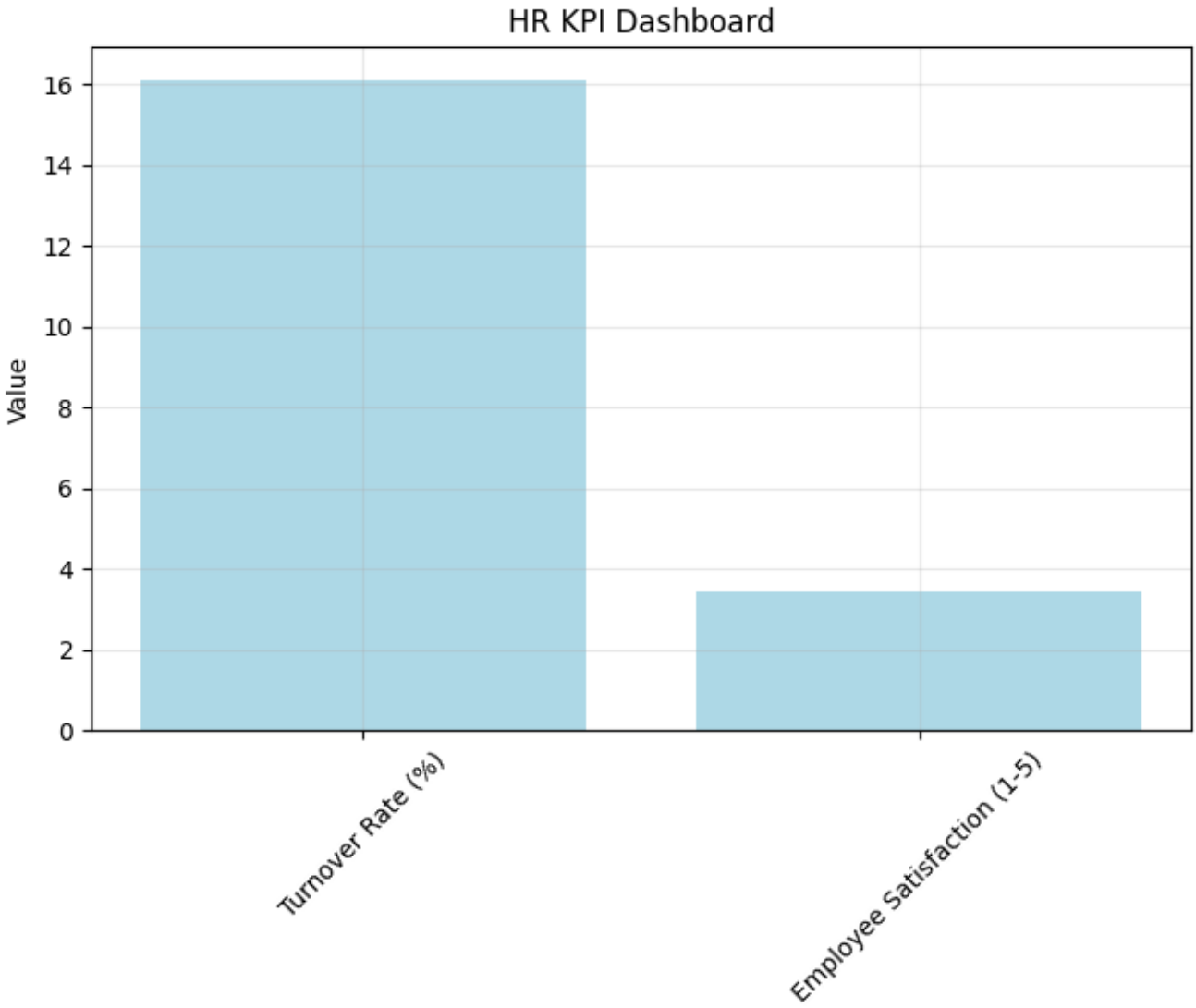
Visualization:

- Create a bar chart that summarizes the calculated KPIs.

Insight:

- The printed turnover rate and average employee satisfaction score.
- A bar chart labeled "HR KPI Dashboard" that visualizes these KPIs.

```python
# --- Turnover Rate ---
# Calculate the percentage of employees who left (Attrition = 'Yes')
turnover_rate = (employee_df['Attrition'] == 'Yes').mean() * 100
print(f"Employee Turnover Rate: {turnover_rate:.2f}%")

# --- Employee Satisfaction ---
# Get the latest performance review for each employee based on ReviewDate
latest_performance = performance_df.loc[performance_df.groupby('EmployeeID')['ReviewDate'].idxmax()]

# Define columns to average for overall satisfaction
satisfaction_cols = ['EnvironmentSatisfaction', 'JobSatisfaction',
                     'RelationshipSatisfaction', 'WorkLifeBalance']

# Calculate overall satisfaction as the mean of satisfaction columns
latest_performance['OverallSatisfaction'] = latest_performance[satisfaction_cols].mean(axis=1)
overall_satisfaction = latest_performance['OverallSatisfaction'].mean()
print(f"Average Employee Satisfaction (1-5 scale): {overall_satisfaction:.2f}")

# --- Visualization: KPI Dashboard ---
# Create a dictionary with the available KPIs
kpi_data = {
    'Turnover Rate (%)': turnover_rate,
    'Employee Satisfaction (1-5)': overall_satisfaction
}
# **Print Results**
print("=== KPIs Analysis ===")
for key, value in kpi_data.items():
    print(f"{key}: {value:.2f}")

# Plot the KPIs as a bar chart
plt.figure(figsize=(8, 5))
plt.bar(kpi_data.keys(), kpi_data.values(), color='lightblue')
plt.title('HR KPI Dashboard')
plt.ylabel('Value')
plt.xticks(rotation=45)
plt.grid(True, alpha=0.3)
plt.show()

```




```
Employee Turnover Rate: 16.12%
Average Employee Satisfaction (1-5 scale): 3.46
=== KPIs Analysis ===
Turnover Rate (%): 16.12
Employee Satisfaction (1-5): 3.46
```

Analysis 2: Turnover Analysis

Objective

Analyze employee turnover to understand attrition patterns, focusing on:

- Identifying why employees leave.
- Determining whether a significant percentage of employees leave within their first year, which could indicate issues with onboarding or training.

Methodology:

1. Filter Attrited Employees: Create a subset of employees who have left (i.e., where Attrition == 'Yes'). Calculate their tenure using the YearsAtCompany column.

2. Tenure Distribution Visualization: Plot a histogram showing the distribution of tenure among attrited employees.

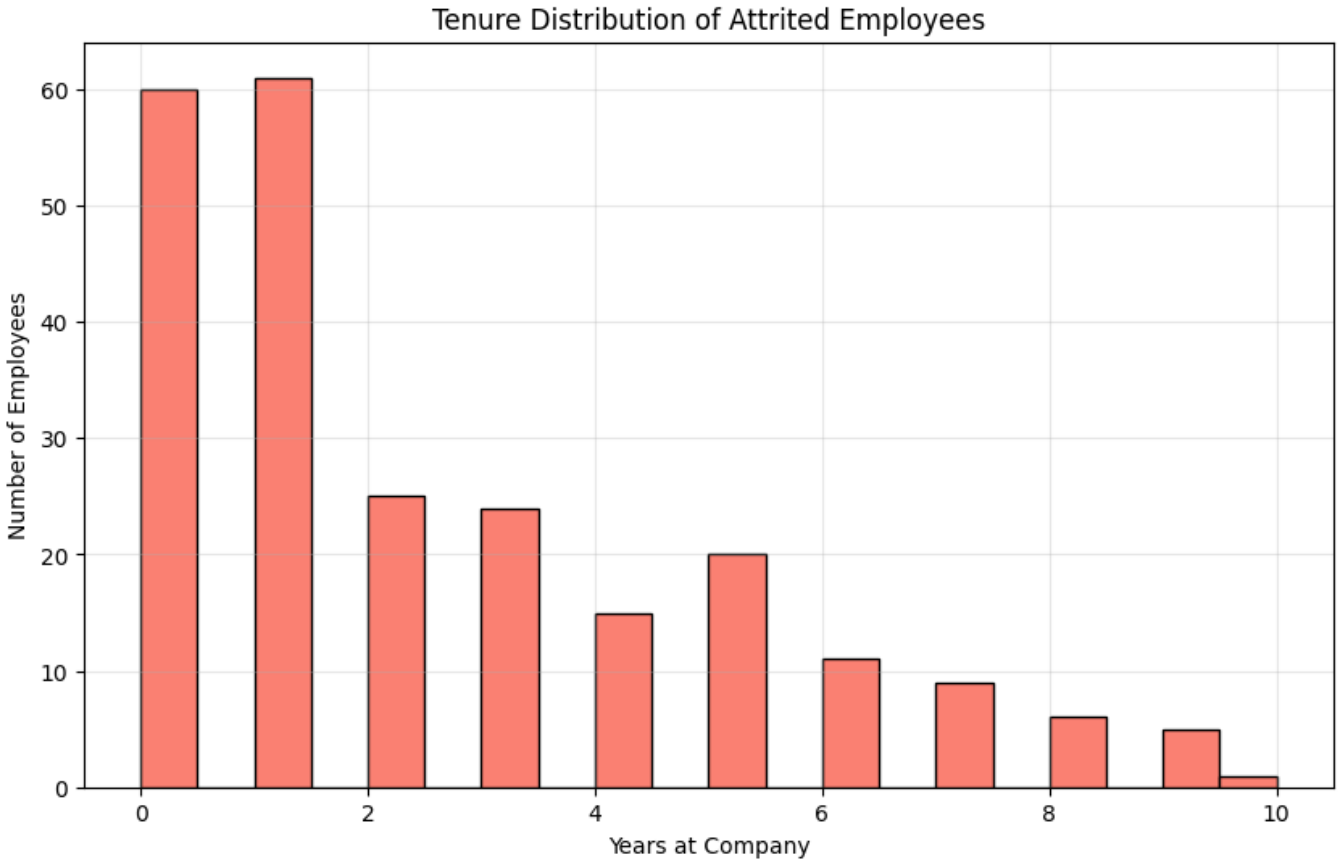
3. First-Year Attrition Calculation: Calculate the percentage of attrited employees with a tenure of one year or less.

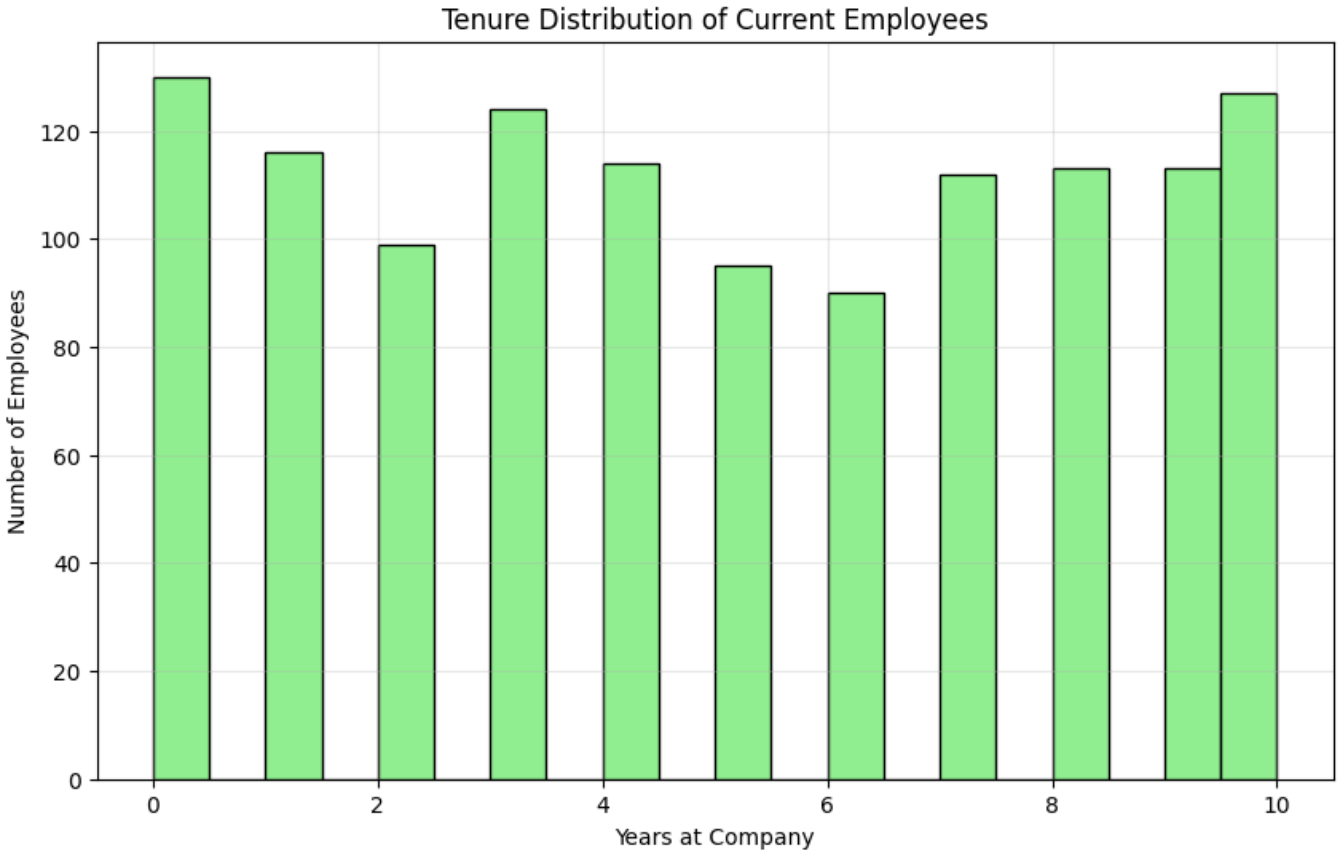
4. Comparison with Current Employees: Similarly, analyze the tenure distribution for current employees. Compare the average tenure between attrited and current employees.

Insight:

- A histogram showing the distribution of tenure among attrited employees.
- The calculated percentage of employees leaving within their first year.
- A histogram showing the tenure distribution for current employees.
- Printed average tenure for both attrited and current employee groups.

```python
# --- Filter Attrited Employees ---
attrited_df = employee_df[employee_df['Attrition'] == 'Yes'].copy()
attrited_df['Tenure'] = attrited_df['YearsAtCompany']

# --- Tenure Distribution of Attrited Employees ---
plt.figure(figsize=(10, 6))
plt.hist(attrited_df['Tenure'], bins=20, color='salmon', edgecolor='black')
plt.title('Tenure Distribution of Attrited Employees')
plt.xlabel('Years at Company')
plt.ylabel('Number of Employees')
plt.grid(True, alpha=0.3)
plt.show()

# --- First-Year Attrition ---
first_year_attrition_pct = (attrited_df['Tenure'] <= 1).mean() * 100
print(f"Percentage of employees leaving within the first year: {first_year_attrition_pct:.2f}%")

# --- Compare with Current Employees ---
current_df = employee_df[employee_df['Attrition'] == 'No'].copy()
current_df['Tenure'] = current_df['YearsAtCompany']

plt.figure(figsize=(10, 6))
plt.hist(current_df['Tenure'], bins=20, color='lightgreen', edgecolor='black')
plt.title('Tenure Distribution of Current Employees')
plt.xlabel('Years at Company')
plt.ylabel('Number of Employees')
plt.grid(True, alpha=0.3)
plt.show()

# --- Average Tenure Comparison ---
avg_tenure_attrited = attrited_df['Tenure'].mean()
avg_tenure_current = current_df['Tenure'].mean()
print(f"Average tenure of attrited employees: {avg_tenure_attrited:.2f} years")
print(f"Average tenure of current employees: {avg_tenure_current:.2f} years")

# **Print Results**

print("\n=== Turnover Analysis ===")
for key, value in turnover_results.items():
    print(f"{key}: {value}")
# Displays the calculated KPIs and turnover insights in a structured format.

```


Tenure Distribution of Attrited Employees
Number of Employees
0
10
20
30
40
50
60
0
2
4
6
8
10
Years at Company


Tenure Distribution of Current Employees
Number of Employees
0
20
40
60
80
100
120
0
2
4
6
8
10
Years at Company

```
Percentage of employees leaving within the first year: 51.05%
```

```
Average tenure of attrited employees: 2.43 years
Average tenure of current employees: 4.97 years

=== Turnover Analysis ===
Turnover by Department: {'Technology': 133, 'Sales': 92, 'Human Resources': 12}
Median Salary of Ex-Employees: 50660.0
Median Experience of Ex-Employees (Years): 1.0
```

Conclusion

KPI Analysis:

- The turnover rate and average employee satisfaction are key metrics that help assess HR performance.
- The KPI dashboard visualization provides a quick, digestible view of these metrics.

Turnover Analysis:

- The tenure distribution of attrited employees indicates the time at which most employees leave the organization.
- A notable percentage of first-year attrition may signal issues in onboarding, training, or early employee engagement.
- Comparing the tenure of attrited versus current employees can guide HR in understanding retention challenges.

Data Dynamos