HR Analytics: Identifying Key Factors Influencing Employee Training Success

Bootcamp Data Analyst with SQL & Python using Google Platform

Dewi Fitradano Adam - PS15180

DOLab

live Class

CONTENTS

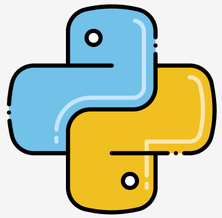
INTRODUCTION

- The HR Analytics dataset provide valuable insight into employee training programs, engagement level, and workplace satisfaction. It contains various attributes such as employee data, performance and survey metrics, training information, dan diversity details.

SCOPE OF ANALYSIS

- To identify important features in each training score that affect training outcomes.
- To identify other factors that affect training success.
- To identify distribution of employee who joined training.
- To identify how much training cost were spent on each program.
- To identify duration of training in each role.
- To identify impact of marital status on performance scores.

TOOLS USED



DATA LOADING

Upload dataset to google colab notebook. Then import corresponding packages such as pandas, numpy, seaborn, matplotilib.pyplot

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
data = pd.read_excel('/content/HR_Data_Analysis.xlsx')
data = pd.DataFrame(data)
data.head()
```

DATA UNDERSTANDING

Dataset is about key factors influencing training success. There are 28 columns in dataset and total 2845 rows of data. Dataset does not contain null value and duplicate entries.

```
# Check duplicate entries based on Employee ID
duplicate_id = data[data.duplicated(subset=["Employee ID"], keep=False)]

# Show total of duplicate ID
print(f"Total duplicate Employee ID: {duplicate_id.shape[0]}")
print(duplicate_id)
```

```
Total duplicate Employee ID: 0
Empty DataFrame
Columns: [Employee ID, StartDate, Title, BusinessUnit, EmployeeStatus, EmployeeType,
Index: []
```

```
Data columns (total 28 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Employee ID                 2845 non-null   int64
 1   StartDate                   2845 non-null   datetime64[ns]
 2   Title                       2845 non-null   object
 3   BusinessUnit                2845 non-null   object
 4   EmployeeStatus              2845 non-null   object
 5   EmployeeType                2845 non-null   object
 6   PayZone                     2845 non-null   object
 7   EmployeeClassificationType  2845 non-null   object
 8   DepartmentType              2845 non-null   object
 9   Division                    2845 non-null   object
 10  DOB                         2845 non-null   datetime64[ns]
 11  State                       2845 non-null   object
 12  GenderCode                  2845 non-null   object
 13  RaceDesc                    2845 non-null   object
 14  MaritalDesc                 2845 non-null   object
 15  Performance Score           2845 non-null   object
 16  Current Employee Rating     2845 non-null   int64
 17  Survey Date                 2845 non-null   datetime64[ns]
 18  Engagement Score            2845 non-null   int64
 19  Satisfaction Score          2845 non-null   int64
 20  Work-Life Balance Score     2845 non-null   int64
 21  Training Date               2845 non-null   datetime64[ns]
 22  Training Program Name       2845 non-null   object
 23  Training Type               2845 non-null   object
 24  Training Outcome            2845 non-null   object
 25  Training Duration(Days)     2845 non-null   int64
 26  Training Cost               2845 non-null   float64
 27  Age                         2845 non-null   int64
```

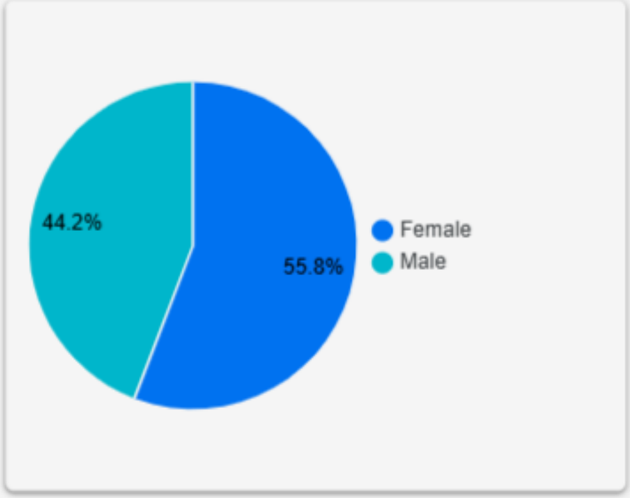
Analysis: Feature Importance in Predicting Training Outcome

Analyse important features that affect training outcome prediction using Random Forest Classifier.

- Work-Life Balance Score having the biggest impact to predict training outcome (35.35%).
- Satisfaction Score is (33.41%).
- Engagement Score has less impact (31.23%).

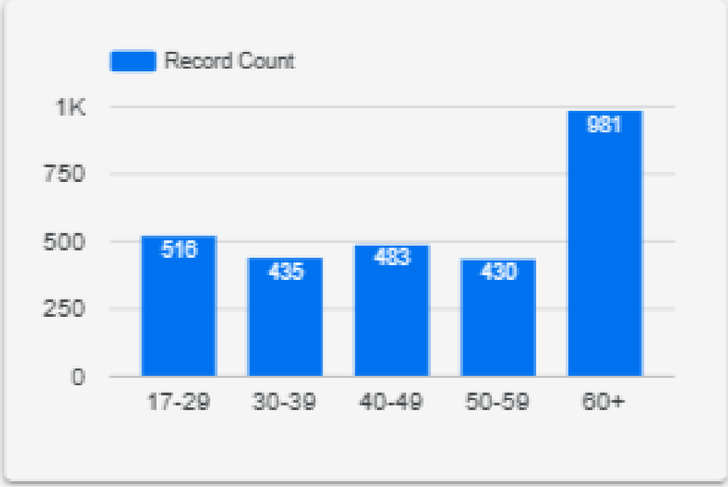
Analysis: Employee Distribution

Gender Distribution



The number of **female** employee is greater than male.

Age Categories



Most of employee age is around **60+ years old** and the second is dominated by 17-29 years old.

Job Title Distribution

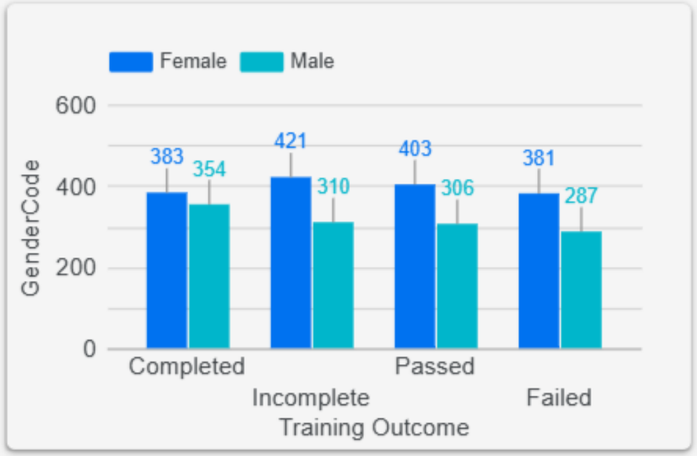
	Title	Record Count
1.	Production Technician I	1,241
2.	Production Technician II	486
3.	Area Sales Manager	277
4.	Production Manager	154
5.	Software Engineer	89
6.	IT Support	88

1 - 32 / 32

Proportion of job title is dominated by role of **Product Technician I**

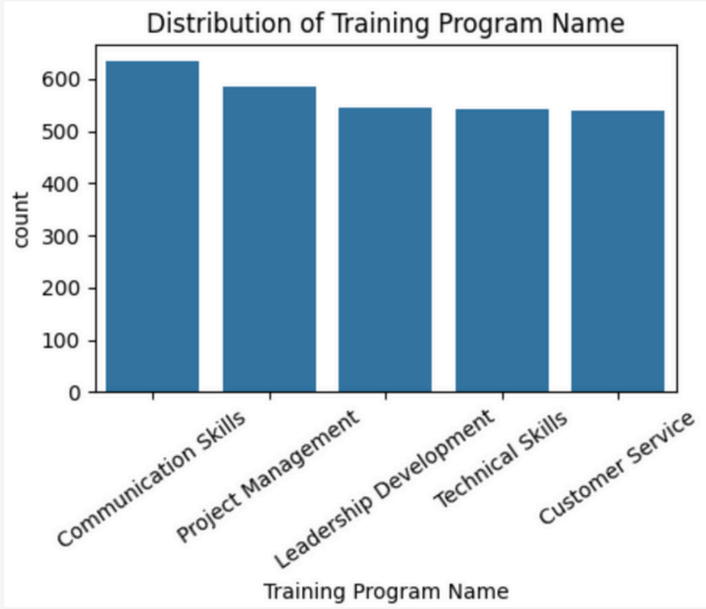
Analysis: Training Program Distribution

Training Outcomes By Gender



Even women have a higher number of participants in all categories, men have a higher success rate than women in numbers of Incomplete and Passed.

Distribution of Training Program




```
Training Program Name
Communication Skills       633
Project Management         585
Leadership Development     544
Technical Skills           543
Customer Service           540
Name: count, dtype: int64
```

Communication Skills Training has the most participants between all training programs.

Analysis: Training Cost & Duration

Average Training Cost per Program

Training Program Name	Training Cost
Communication Skills	544.303318
Customer Service	569.882463
Leadership Development	561.423658
Project Management	562.922735
Technical Skills	560.117551

Average Daliy Training Cost

Training Program Name	
Communication Skills	250.196554
Leadership Development	259.949839
Project Management	262.460960
Customer Service	265.523125
Technical Skills	266.485986

Name: Cost Per Day, dtype: float64

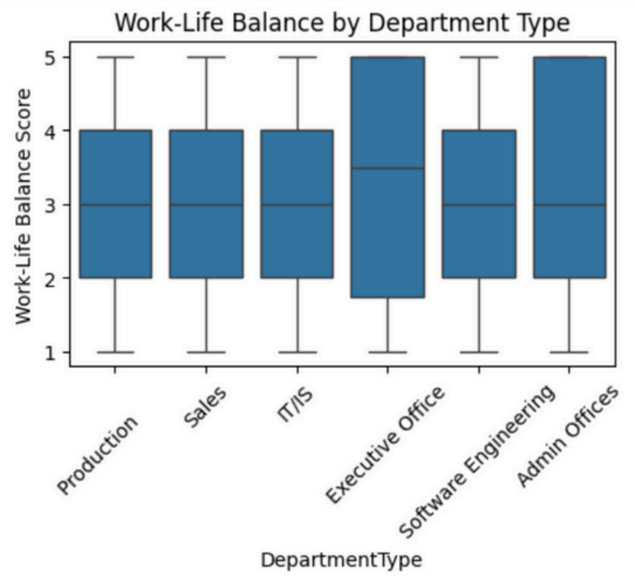
- Customer Service Training has the higest training cost.
- Technical skill has the higest average daily training cost spent on its program.

Average Training Duration by Job Title

- IT Manager role spent the shortest training duration day, averaging in 2.64.
- Senior BI is on the second place on training duration day.

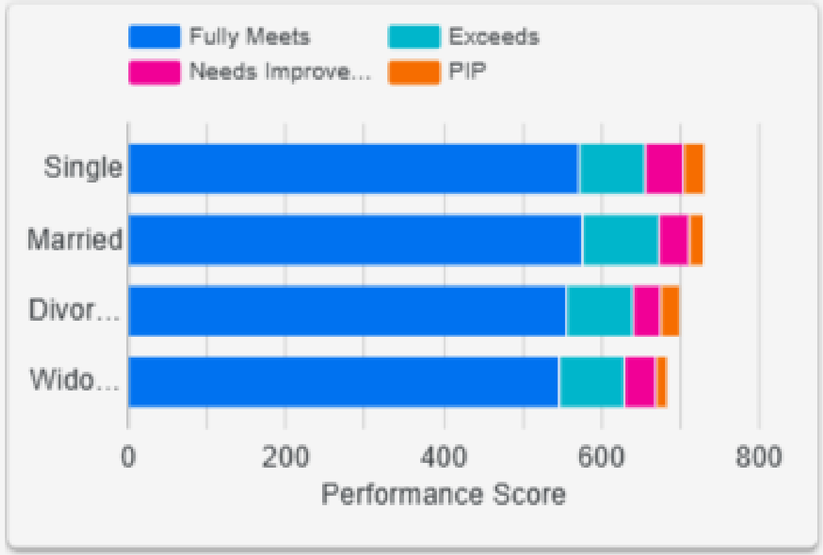
Analysis: Work-Life Balance & Marital Status

Work Life Balance by Department



- Median for most departments is around 3 or 4, meaning employees generally have work-life balance to well.
- Executive Office and Admin Offices have very high of work-life balance rate.
- Software Engineering generally experiences a lower work-life balance.

Performance Score by Marital Status



There are no major differences in performance trends between marital statuses. All groups display a similar pattern, meaning that marital status does not appear to significantly impact performance.

CONCLUSION & RECOMMENDATIONS

Conclusion

1. **Key Features in Training Outcome Prediction:**
   - Work-Life Balance Score has the highest impact (35.35%).
   - Satisfaction Score follows closely (33.41%).
   - Engagement Score has a lower impact (31.23%).
2. **Employee Distribution:**
   - The majority of employees are 60+ years old.
   - Total of female employee is larger than male.
   - The most common job title is Product Technician I.
3. **Training Program Distribution:**
   - Female employee participate more in training program, but male have a higher success rate in training success.
   - Communication Skills Training is the most attended training program.
4. **Training Cost & Duration:**
   - Customer Service Training has the highest training cost.
   - Technical Skill Training has the highest daily training cost.
   - IT Managers have the shortest training duration, averaging 2.64.
5. **Work-Life Balance & Performance:**
   - Most employees rate their work-life balance between 3 and 4 (out of 5).
   - Executive Office and Admin Offices have the highest work-life balance, and Software Engineering experiences the lowest work life balance.
   - Marital status does not significantly impact performance.

Recommendations

1. Address lower work-life balance in Software Engineering through flexible hours or hybrid work models.
2. Conduct periodic employee well-being surveys to monitor and enhance balance.
3. Focus on work-life balance and satisfaction to improve training outcomes.
4. Provide additional support for female employees to increase their training success.
5. Evaluate the cost-effectiveness of customer service training and technical skill training to ensure a high return on investment.
6. Consider shorter, intensive training for roles with extended training duration.
7. Use adaptive learning techniques to improve engagement and satisfaction.
8. Provide mentorship program to support new employees in key training program.

Attached Links

Google Colab

Looker Studio

Dataset Source on Kaggle

DQLab

live Class

THANK YOU!