HR analytics dashboard Report

Infosys Springboard Internship 5.0 Batch 1

Submitted By

NAME- VISHAL SINGH AIREE

EMAIL-vishalbist45@gmail.com

CHAPTER 1
METHODS

2.1 Dataset

The dataset utilized in this project is robust, comprising 26 columns covering a wide range of employee metrics. These metrics are carefully categorized to offer comprehensive insights into key areas such as demographics, job roles, job satisfaction, and engagement levels. Key features of the dataset include:

- **Demographics:**
  - Name
  - Office Location
  - Age
  - Gender
  - Marital Status
- **Job Metrics:**
  - Attrition Status
  - Business Travel Frequency
  - Department
  - Job Role
  - Job Level
  - Monthly Income
  - Total Years at Company
- **Satisfaction Metrics:**
  - Relationship Satisfaction
  - Job Satisfaction
  - Environment Satisfaction
- **Work Metrics:**
  - Training Hours in the Last Year
  - Work-Life Balance Rating
  - Job Involvement Score

EmployeeNumber	RelationshipSatisfaction	TrainingTimesLastYear	WorkLifeBalance	JobSatisfaction	EnvironmentSatisfaction	JobInvolvement
11	2	2	3	3	4	3
79	4	2	3	4	4	3
84	3	2	3	3	4	3
134	4	2	3	1	4	3
193	2	2	3	3	4	3
253	2	2	3	3	4	3
293	4	2	3	3	4	3
349	4	2	3	4	4	3
374	3	2	3	1	4	3
393	2	2	3	4	4	3
410	2	2	3	2	4	3
428	2	2	3	2	4	3
441	4	2	3	4	4	3
447	2	2	3	3	4	3
482	3	2	3	3	4	3
485	2	2	3	3	4	3
497	2	2	3	2	4	3
586	3	2	3	3	4	3
728	4	2	3	1	4	3
738	4	2	3	2	4	3
766	4	2	3	2	4	3
787	4	2	3	1	4	3
825	4	2	3	2	4	3
864	3	2	3	1	4	3
882	2	2	3	4	4	3
895	2	2	3	4	4	3
964	4	2	3	2	4	3

- This **<u>Engagement Survey sheet</u>** measures employee satisfaction and engagement.
- Key Information:
- - Satisfaction Metrics: Relationship Satisfaction, Job Satisfaction, Environment Satisfaction
- - Work Metrics: Training Times Last Year, Work Life Balance, Job Involvement

The dataset is enriched with mappings of attrition trends, satisfaction ratings, and engagement survey results to enable meaningful and actionable visualizations.

2.2 Technologies Used

2.3 Architecture

The dashboard follows a modular architecture, ensuring scalability and clarity. The key components include:

1. Data Layer:

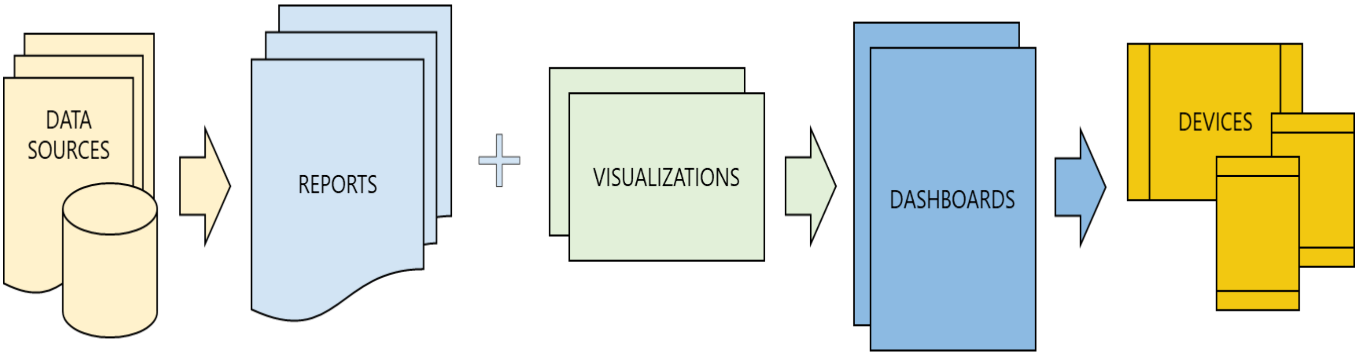
- Ingestion of data from internal HR systems or external files.
- Preprocessing steps such as cleaning, handling missing values, and encoding categorical data to ensure high-quality inputs.

2. Analytics Layer:

- Power BI's advanced DAX expressions for custom measures and calculated fields.
- Python scripts for predictive analytics, including attrition forecasting and feature importance analysis.

3. Visualization Layer:

- Interactive dashboards designed to showcase trends and patterns in employee data.
- Integration of Python-based visuals within Power BI for deeper analytical insights.



Definition of Power BI:

Definition of Power BI

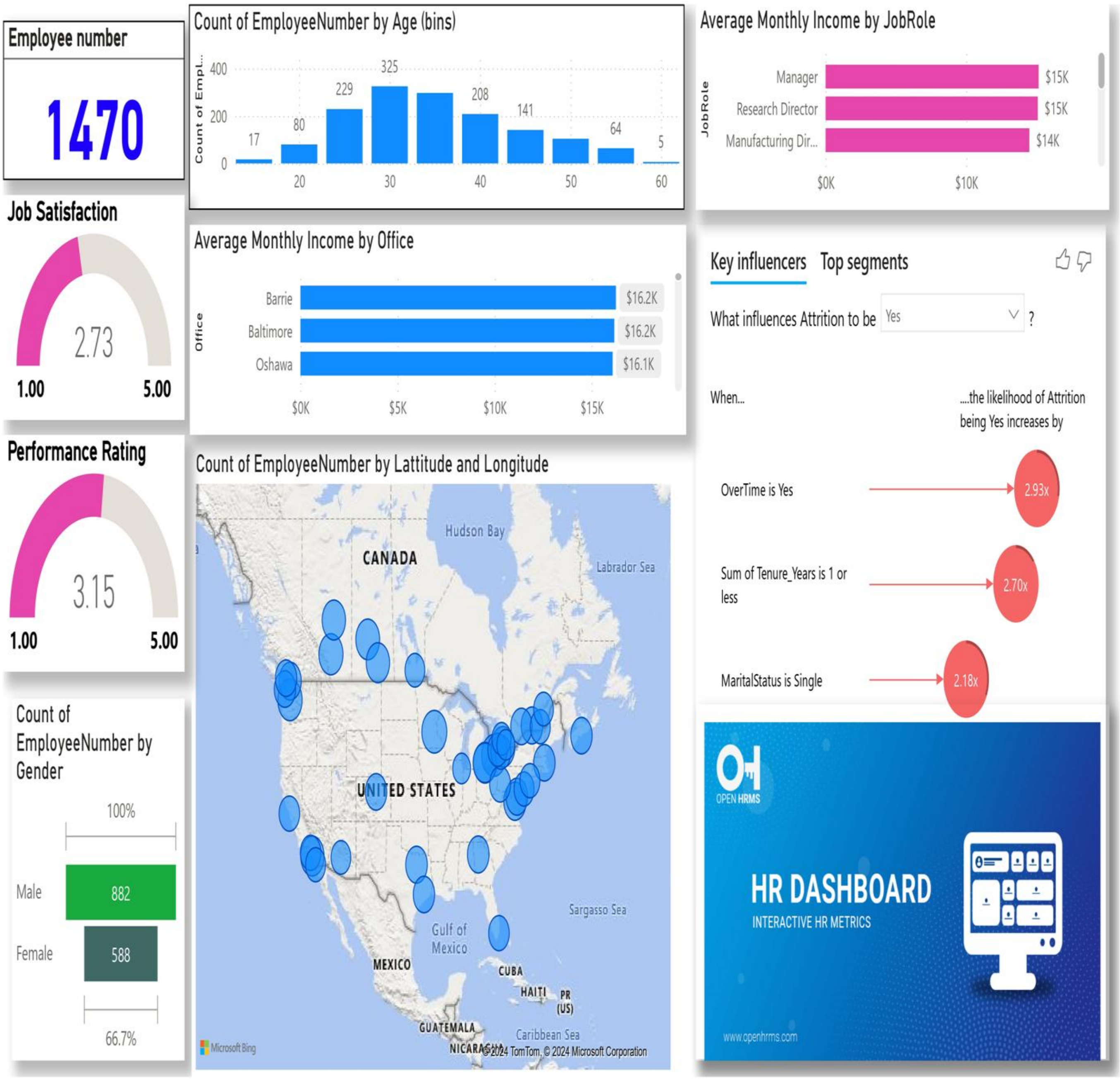
Power BI is a business intelligence and analytics platform by Microsoft that enables users to create interactive dashboards and reports by connecting to diverse data sources, performing data modeling, and visualizing insights.

Project Overview

The HR Analytics Dashboard integrates **business intelligence** with **machine learning**, offering the following benefits:

- Real-time insights into employee metrics.
- Predictive modeling of attrition to enable proactive measures.
- Interactive and user-friendly visuals that cater to diverse stakeholders.

3.1 OUTPUT:



Employee number
1470
Count of EmployeeNumber by Age (bins)
Count of Empl...
400
200
0
17
80
229
325
208
141
64
5
20
30
40
50
60
Average Monthly Income by JobRole
JobRole
Manager
Research Director
Manufacturing Dir...
$15K
$15K
$14K
$0K
$10K
Job Satisfaction
2.73
1.00
5.00
Average Monthly Income by Office
Office
Barrie
Baltimore
Oshawa
$16.2K
$16.2K
$16.1K
$0K
$5K
$10K
$15K
Key influencers
Top segments
What influences Attrition to be
Yes
?
When...
....the likelihood of Attrition being Yes increases by
OverTime is Yes
2.93x
Sum of Tenure_Years is 1 or less
2.70x
MaritalStatus is Single
2.18x
Performance Rating
3.15
1.00
5.00
Count of EmployeeNumber by Lattitude and Longitude
Hudson Bay
CANADA
Labrador Sea
UNITED STATES
Sargasso Sea
Gulf of Mexico
MEXICO
CUBA
HAITI
PR (US)
GUATEMALA
Caribbean Sea
NICARAGUA
Microsoft Bing
© 2024 TomTom, © 2024 Microsoft Corporation
Count of EmployeeNumber by Gender
100%
Male
882
Female
588
66.7%
OPEN HRMS
HR DASHBOARD
INTERACTIVE HR METRICS
www.openhrms.com

This Power BI dashboard is an **HR Analytics Dashboard** that provides an overview of key employee metrics, with a mix of summary statistics, charts, and interactive visualizations to enable workforce analysis. Here’s a detailed breakdown of the elements in the dashboard:

---

1. Employee Number

- **Metric Shown:** Total number of employees in the organization (1,470).
- **Purpose:** A quick, high-level summary of the organization's workforce size.

---

2. Job Satisfaction

- **Visualization Type:** Gauge Chart.
- **Metric Shown:** Average job satisfaction score, which is **2.73 on a scale of 1 to 5**.
- **Purpose:** Indicates the overall satisfaction level of employees with their job, where lower values suggest potential dissatisfaction and the need for improvement initiatives.

---

3. Performance Rating

- **Visualization Type:** Gauge Chart.
- **Metric Shown:** Average performance rating, which is **3.15 on a scale of 1 to 5**.
- **Purpose:** Provides an overall measure of employee performance, useful for identifying areas where support or recognition may be needed.

---

4. Count of Employee Number by Gender

- **Visualization Type:** Horizontal Bar Chart.
- **Metric Shown:** Gender distribution of employees:
  - Male: **882 employees (66.7%)**.
  - Female: **588 employees (33.3%)**.
- **Purpose:** Highlights gender diversity within the organization.

---

5. Count of Employee Number by Age (Bins)

- **Visualization Type:** Histogram/Bar Chart.
- **Metric Shown:** Number of employees across different age ranges:
  - The majority of employees are in their 20s (325) and 30s (229), with fewer employees aged 50+.

- **Purpose:** Helps understand the age demographics of the workforce, which can inform workforce planning and succession strategies.

---

6. Average Monthly Income by Job Role

- **Visualization Type:** Bar Chart.
- **Metric Shown:** Average monthly income for different job roles:
  - Managers and Research Directors earn approximately **$15K per month**, while Manufacturing Directors earn around **$14K**.
- **Purpose:** Allows analysis of income distribution across roles, which may help in evaluating pay equity and designing compensation policies.

---

7. Average Monthly Income by Office

- **Visualization Type:** Bar Chart.
- **Metric Shown:** Average monthly income across offices:
  - Offices in **Barrie, Baltimore, and Oshawa** have an average income of **$16.2K to $16.1K**.
- **Purpose:** Provides insight into salary variations across geographic locations, potentially useful for regional compensation strategies.

---

8. Count of Employee Numbers by Latitude and Longitude

- **Visualization Type:** Geographical Map.
- **Metric Shown:** Distribution of employees across office locations in North America.
- **Purpose:** Visually represents where employees are geographically concentrated, which is useful for resource allocation and regional HR planning.

---

9. Key Influencers Analysis

- **Visualization Type:** Key Influencers Chart.
- **Metric Shown:** Factors influencing employee attrition:
  - **Overtime being Yes** increases the likelihood of attrition by 2.93X.
  - Employees with **tenure of 1 year or less** are 2.70X more likely to leave.
  - **Single marital status** increases attrition likelihood by 2.18X.
- **Purpose:** Identifies the key drivers of employee turnover, helping HR teams focus on these factors to reduce attrition.

---

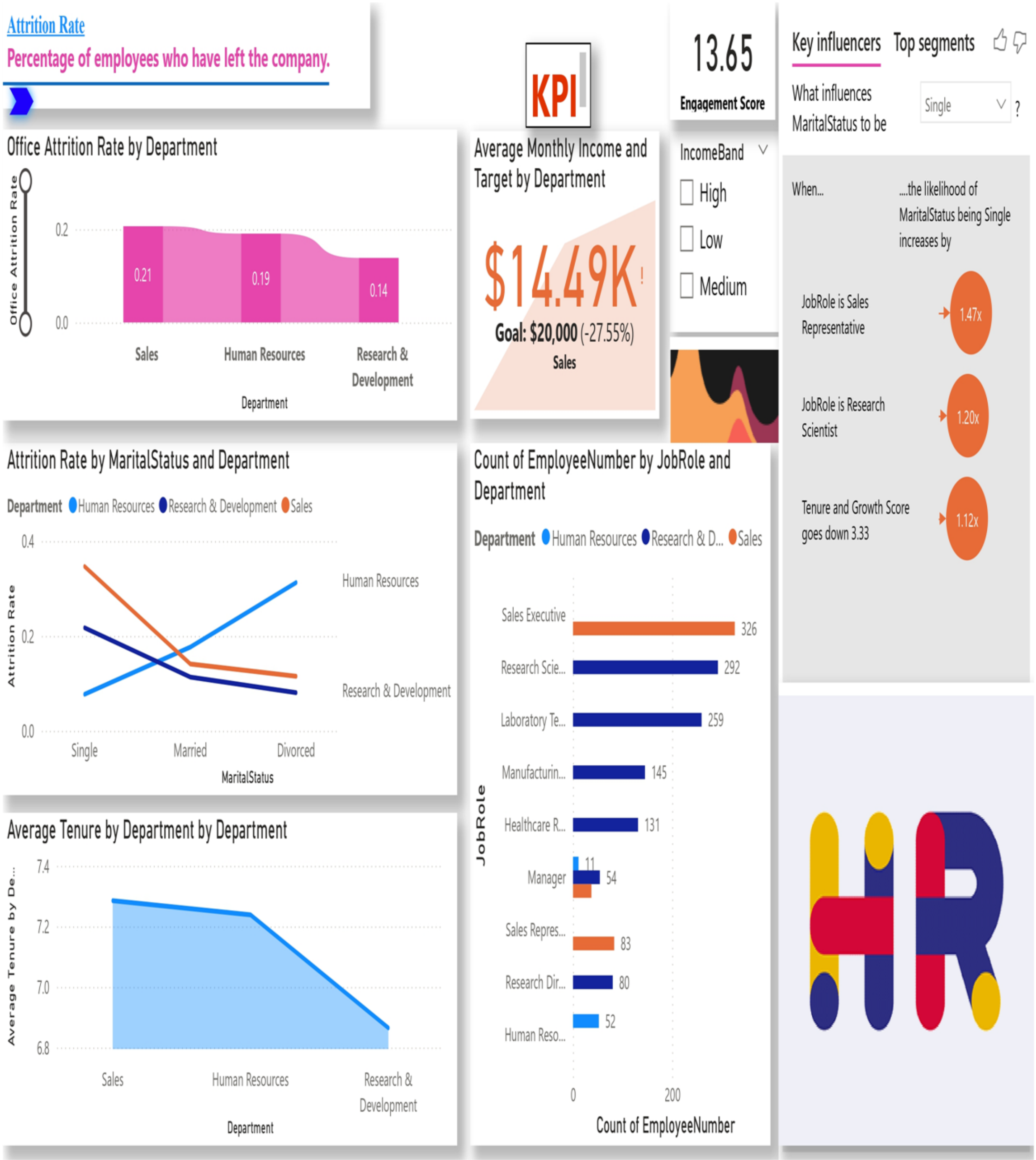
10. HR Dashboard Branding

- **Component:** Title Banner.
- **Purpose:** Serves as the dashboard's branding and provides a professional look, with the title "HR DASHBOARD – Interactive HR Metrics."

Overall, Purpose of the Dashboard

This dashboard provides an interactive and data-driven overview of key HR metrics, enabling stakeholders to:

1. Analyse workforce demographics and satisfaction levels.
2. Understand pay trends and employee distribution across locations.
3. Identify critical factors driving attrition.
4. Use actionable insights for decision-making to improve retention, satisfaction, and workforce efficiency.



This Power BI dashboard focuses on **employee attrition and engagement analytics**, with key metrics and visualizations offering insights into workforce trends and department-specific dynamics. Here's a breakdown of the dashboard:

---

1. Attrition Rate

- **Metric Shown:** A summary of the percentage of employees who have left the company.

- **Purpose:** Provides a clear and concise definition of "Attrition Rate" for viewers, making the dashboard more intuitive.

---

2. Office Attrition Rate by Department

- **Visualization Type:** Bar Chart.
- **Metric Shown:** Attrition rates across departments:
  - Sales has the highest attrition rate (0.21).
  - Research & Development has the lowest attrition rate (0.14).
- **Purpose:** Highlights departmental differences in attrition, enabling HR to focus on areas with higher turnover.

---

3. Average Monthly Income and Target by Department

- **Visualization Type:** KPI Card.
- **Metric Shown:** The average monthly income for the **Sales department** is $14.49K, which falls short of the target of $20K (-27.55%).
- **Purpose:** Displays income performance compared to goals, useful for assessing compensation strategy and goal alignment.

---

4. Engagement Score

- **Metric Shown:** A consolidated engagement score of **13.65**.
- **Purpose:** Offers a high-level indicator of employee engagement, which can correlate with productivity and retention.

---

5. Attrition Rate by Marital Status and Department

- **Visualization Type:** Line Chart.
- **Metric Shown:** Attrition rates based on marital status for each department:
  - **Single employees** have higher attrition rates in Sales and Human Resources.
  - Married and divorced employees tend to have lower attrition rates across most departments.
- **Purpose:** Explores how personal demographics (marital status) influence turnover across departments, allowing for targeted retention strategies.

---

6. Key Influencers Analysis

- **Visualization Type:** Key Influencers Chart.

- **Metric Shown:** Factors influencing marital status to be "Single":
    - Being a **Sales Representative** increases the likelihood by 1.47x.
    - Being a **Research Scientist** increases the likelihood by 1.20x.
    - A decrease in "Tenure and Growth Score" correlates with a 1.12x higher likelihood of being single.
- **Purpose:** Offers insights into workforce demographics and career patterns, helping HR design tailored engagement strategies.

---

7. Average Tenure by Department

- **Visualization Type:** Line Chart.
- **Metric Shown:** Average tenure for employees in each department:
    - Sales has the highest tenure (~7.4 years).
    - Research & Development has the lowest tenure (~6.8 years).
- **Purpose:** Analyses workforce stability and experience levels across departments.

---

8. Count of Employees by Job Role and Department

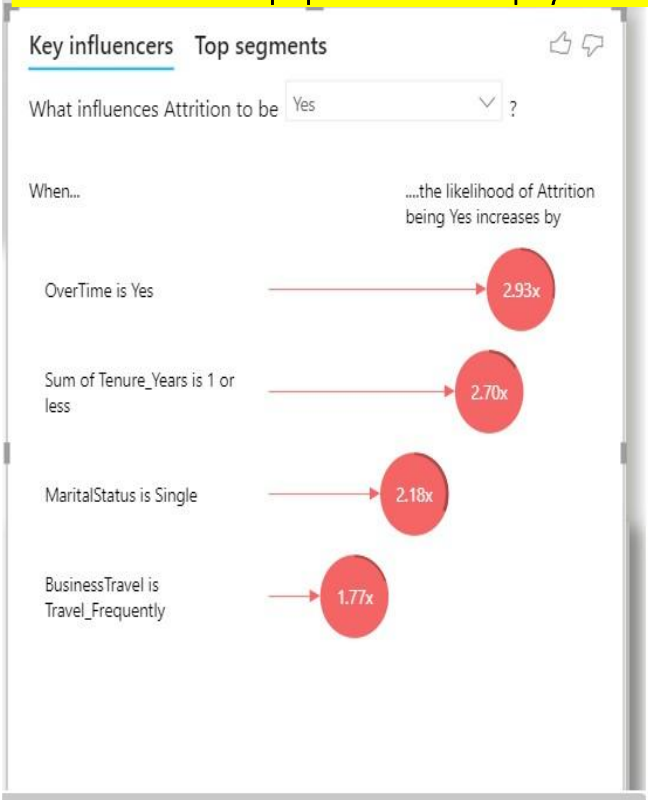
- **Visualization Type:** Stacked Bar Chart.
- **Metric Shown:** Number of employees in various roles across departments:
    - **Sales Executives** dominate the Sales department.
    - **Research Scientists** are the most common in Research & Development.
    - Fewer employees are in managerial roles across all departments.
- **Purpose:** Helps understand workforce distribution, highlighting departmental specialization and potential resource gaps.

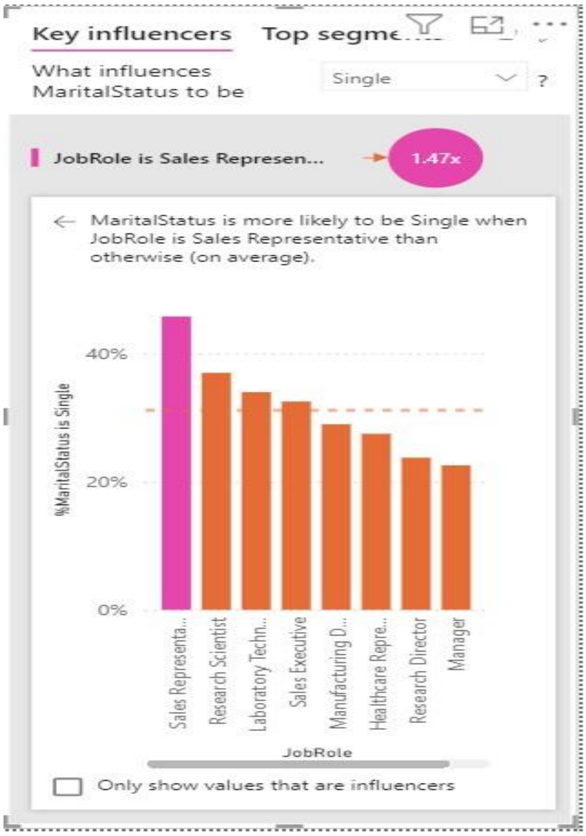
---

9. Branding

- **Component:** HR logo in bold and vibrant colors.
- **Purpose:** Adds a professional, branded element to the dashboard, making it visually appealing.

if overtime is less than the people will leave the company almost 3x





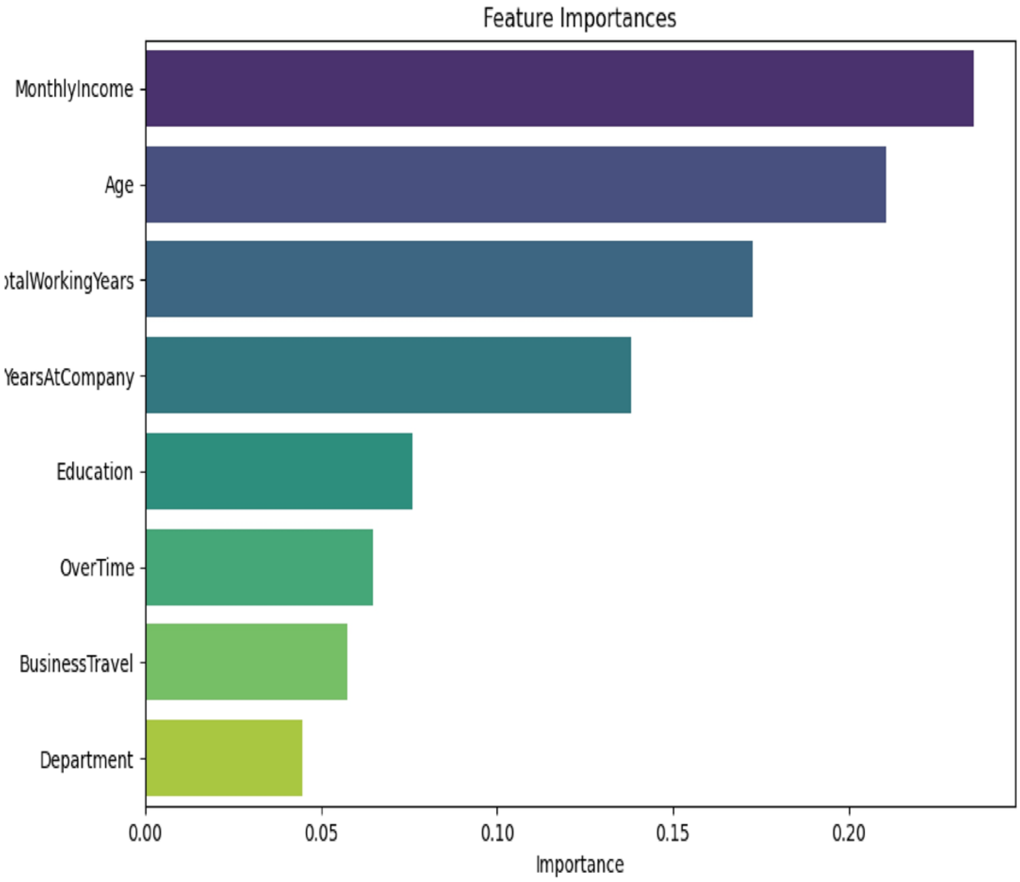
if overtime is yes than the people will leave the company almost 3x

Feature Importance Plot:

This plot shows which features have the most impact on predicting attrition.

- X-Axis: Feature importance scores—higher scores mean greater influence.
- Y-Axis: Features ranked by importance.

Feature and Importance



This visual, embedded in a Power BI dashboard, uses **Python's data visualization capabilities** (likely Matplotlib or Seaborn) to display the **Feature Importance Plot** for predicting employee attrition using a **Random Forest model**. Here's an explanation:

What the Plot Represents

- **Objective:** Identify which features (employee attributes or metrics) have the most impact on predicting attrition (whether employees leave the company).
- **Model Used:** Random Forest, a machine learning algorithm that determines the importance of features based on how much they reduce uncertainty (or split quality) in decision trees.

Axes

1. **X-Axis (Importance):**
   - Displays the relative importance score of each feature.
   - Higher scores indicate greater influence on predicting attrition.
2. **Y-Axis (Features):**
   - Lists the features ranked in descending order of importance, with the most influential at the top.

Key Features and Insights

1. **Monthly Income** (Most Important):
   - High influence suggests that employee compensation is a critical factor in attrition. Low salaries may be associated with higher turnover.
2. **Age:**
   - Employees' age impacts attrition, possibly reflecting differences in career priorities or opportunities between younger and older employees.
3. **Total Working Years:**
   - The cumulative experience of an employee plays a significant role, possibly linked to job satisfaction or retirement.
4. **Years at Company:**
   - Employees with shorter tenure might have a higher likelihood of leaving, indicative of onboarding or cultural mismatch issues.
5. **Education:**
   - Educational qualifications influence attrition, potentially reflecting alignment (or lack thereof) between job requirements and employee capabilities.
6. **Overtime:**
   - Indicates the impact of workload on employee decisions to stay or leave. Employees frequently required to work overtime may experience burnout.
7. **Business Travel:**

- Frequent travel might be a driver for attrition, especially if it affects work-life balance.

8. **Department (Least Important):**
   - The department an employee works in has comparatively lower influence on attrition, but it still matters in the broader context.
   - **Code**

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from scipy.stats import binom

# Load the Excel file
file_path = r'C:\Users\visha\Videos\HRM database PowerBI.xlsx'
excel_data = pd.ExcelFile(file_path)

# Load the HRIS sheet
hris_data = excel_data.parse('HRIS')

# Select relevant columns for attrition prediction
columns_to_use = ['Attrition', 'Department', 'Age', 'BusinessTravel', 'Education',
            'YearsAtCompany', 'TotalWorkingYears', 'OverTime', 'MonthlyIncome']
data = hris_data[columns_to_use].dropna()

# Encode categorical variables
```

```python
label_encoders = {}
for col in ['Attrition', 'Department', 'BusinessTravel', 'OverTime']:
    le = LabelEncoder()
    data[col] = le.fit_transform(data[col])
    label_encoders[col] = le

# Split data into features and target
X = data.drop('Attrition', axis=1)
y = data['Attrition']

# Split into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Train a Random Forest model
rf_model = RandomForestClassifier(random_state=42)
rf_model.fit(X_train, y_train)

# Make predictions
y_pred = rf_model.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
classification_rep = classification_report(y_test, y_pred, output_dict=True)

# Extract key metrics from the classification report
precision = classification_rep['weighted avg']['precision']
recall = classification_rep['weighted avg']['recall']
f1_score = classification_rep['weighted avg']['f1-score']

# Simulate attrition using a binomial distribution
n_trials = len(y_test)  # Number of employees in the test set
p_success = accuracy     # Probability of success (correct prediction)
```

```python
# Binomial distribution: Simulate the number of correctly predicted attritions
binom_simulation = binom.rvs(n=n_trials, p=p_success, size=1000)

# Prepare feature importance data
feature_importances = pd.DataFrame({
    'Feature': X.columns,
    'Importance': rf_model.feature_importances_
}).sort_values(by='Importance', ascending=False)

# Plot feature importance with all outputs
plt.figure(figsize=(12, 8))
sns.barplot(data=feature_importances, x='Importance', y='Feature', palette='viridis')
plt.title('Feature Importances with Model Performance', fontsize=16)
plt.xlabel('Importance', fontsize=12)
plt.ylabel('Features', fontsize=12)

# Add model evaluation metrics as text
plt.text(0.6, len(feature_importances) - 1,
         f"Model Accuracy: {accuracy:.2f}\n"
         f"Precision: {precision:.2f}\n"
         f"Recall: {recall:.2f}\n"
         f"F1-Score: {f1_score:.2f}",
         fontsize=12, bbox=dict(facecolor='white', alpha=0.6))

# Add binomial simulation insights
expected_success = n_trials * p_success
plt.text(0.6, len(feature_importances) - 2.5,
         f"Binomial Expected Successes: {expected_success:.0f}\n"
         f"Simulation Range: {min(binom_simulation)}-{max(binom_simulation)}",
         fontsize=12, bbox=dict(facecolor='white', alpha=0.6))
```

```
plt.tight_layout()
plt.show()
```

Data Analysis Expressions Code

So, what is DAX, The DAX is a collection of functions, operators, and constants that can be used in a formula, or expression, to calculate and return one or more values.

```
Office Attrition Rate =
DIVIDE(
   COUNTROWS(FILTER(HRIS, HRIS[Attrition] = "Yes")),
   COUNTROWS(HRIS),
   0
)
```

Inner Filter:

FILTER(HRIS, HRIS[Attrition] = "Yes"): This part filters the HRIS table to only include rows where the Attrition column equals "Yes".

Counting Attrited Employees:

COUNTROWS(FILTER(HRIS, HRIS[Attrition] = “Yes”)): This counts the number of rows (employees) in the filtered table, giving us the total number of employees who have attrited.

Counting Total Employees:

COUNTROWS(HRIS): This counts the total number of rows (employees) in the entire HRIS table.

Calculating Attrition Rate:

DIVIDE: This function divides the number of attrited employees by the total number of employees. The 0 as the third argument ensures that the formula returns a result even if the divisor is zero, preventing potential errors.

In simpler terms:

The formula calculates the attrition rate by dividing the number of employees who left the company by the total number of employees.

Example:

If you have an HRIS table with 100 employees, and 10 of them have attrited (indicated by "Yes" in the Attrition column), the formula would calculate:

DIVIDE(10, 100, 0) = 0.1

```
Tenure and Growth Score =



AVERAGEX(
   HRIS,
   DIVIDE(
      [TotalWorkingYears] + [YearsWithCurrManager] +
      [YearsSinceLastPromotion] + [YearsInCurrentRole] +
      [Tenure_Years] + [PercentSalaryHike],
      6
   )
)
```

Step-by-Step Explanation:

1. **Calculating the Sum of Relevant Factors:**
    - The formula adds up several factors related to an employee's tenure and growth:
        - TotalWorkingYears
        - YearsWithCurrManager
        - YearsSinceLastPromotion
        - YearsInCurrentRole
        - Tenure_Years
        - PercentSalaryHike
2. **Normalizing the Sum:**
    - The sum is divided by 6 to normalize the values and bring them to a common scale. This helps in comparing employees with different tenures and growth trajectories.
3. **Calculating the Average:**
    - AVERAGEX iterates through each row of the HRIS table, calculates the normalized score for each employee, and then calculates the average of these scores across all employees.

Interpretation of the Score:

- A higher score indicates a longer tenure, more stability, and potential for growth within the organization.
- A lower score might suggest a shorter tenure, recent job changes, or slower career progression.

Potential Use Cases:

- **Identifying High-Potential Employees:** Employees with higher scores might be identified as high-potential candidates for leadership roles or promotions.
- **Retention Strategies:** Analyzing the scores can help identify employees who might be at risk of leaving, allowing for targeted retention strategies.
- **Performance Management:** The score can be used as one of the factors in performance reviews, along with other performance metrics.
- **Succession Planning:** By understanding the tenure and growth of employees, organizations can identify potential successors for key roles.

REFERENCES

Power BI

- **Official Documentation:**
  - Microsoft Power BI: https://powerbi.microsoft.com
  - Learn about DAX (Data Analysis Expressions): https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-quickstart-learn-dax-basics
  - Python Integration with Power BI: https://learn.microsoft.com/en-us/power-bi/connect-data/service-python-visuals

2. Random Forest Algorithm

- **Concept and Theory:**
  - Leo Breiman, "Random Forests": https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf
  - Scikit-learn Documentation (Random Forest): https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

3. Feature Importance in Machine Learning

- **Overview of Feature Importance:**
  - Christoph Molnar, "Interpretable Machine Learning": https://christophm.github.io/interpretable-ml-book/
  - SHAP (SHapley Additive exPlanations) and Feature Importance: https://shap.readthedocs.io/en/latest/

4. Data Visualization with Python

- **Visualization Libraries:**
  - Matplotlib Documentation: https://matplotlib.org/stable/contents.html
  - Seaborn for Statistical Graphics: https://seaborn.pydata.org/

5. HR Analytics and Employee Attrition

- **HR Analytics Overview:**
  - Predictive HR Analytics: By Martin R. Edwards and Kirsten Edwards, Book.
  - Case Studies in HR Analytics: By Mohan Thite, Book.
  - McKinsey Report on People Analytics: https://www.mckinsey.com
- **Attrition Modeling:**
  - Kaggle: Employee Attrition Dataset Example: https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

Online Resources:

[1] Infosys certification program : https://www.infosys.com/