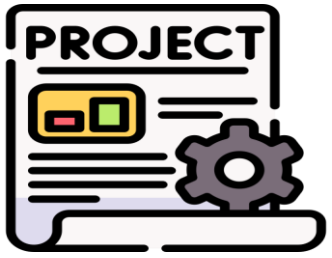
Graduation Project

رواد مصر الرقمية

Human Resources Analysis



رواد مصر الرقمية

Data Analysis Project Team_DEPI

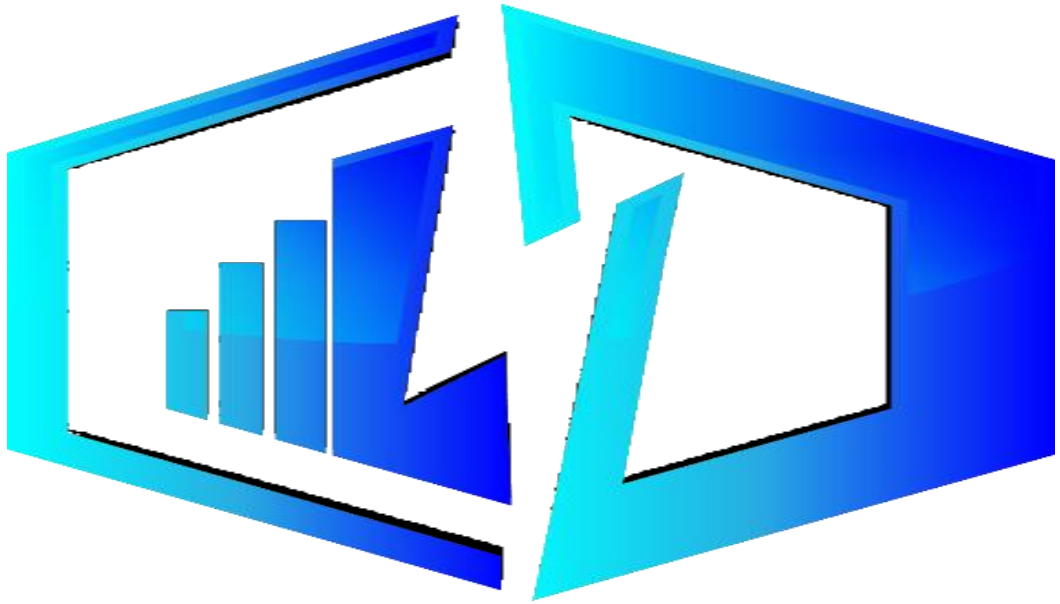
DEPI_GIZ2_DAT1_S5

Under Supervision of:

Dr. Sepa Mostafa

Dr. Rami Magdi

Data Dynamos Team	
1	Ahmed Emad Habib
2	Mohammed Amgd Marzouk
3	Peter Rafat Adel
4	Mohammed Elhossiny Abrabo
5	Zyad Sameh Mohamed



DATA DYNAMOS

Harnessing the Power of Data to Drive Innovation

About Dr. Seba

Dr. Seba Mostafa

Data Analyst

I am Sepa, I have a BSc in Science and have been working as a senior data analyst and data analytics trainer for the past 5 years.

I have been working with both global and local companies (Udacity, Orange, Afri Labs, Carerha, ...).

- Published a paper on machine learning in an international journal.
- Trained over 1000 Students.
- Students who graduated got job offers.
- Udacity Top-Rated Instructor.
- Over 1000 hours of training done.

Meet Our Team

Ahmed Emad Habib

Mohammed Elhossiny Abrabo

Peter Rafat Adel

Meet Our Team

Mohammed Amgd Marzouk

Zyad Sameh Mohamed

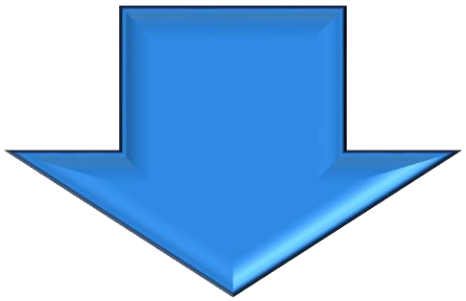
Introduction

This project focuses on analyzing the HR data of a company to gain a comprehensive understanding of employee demographics, salary trends, promotions, and work-life factors. By examining key areas such as employee turnover, salary distributions across gender and ethnicity, and correlations between education, marital status, and overtime, the project aims to provide data-driven insights. It will also explore employee-manager relationships through performance ratings and the impact of training opportunities on promotions and salaries. Various statistical methods and visual representations will be used to identify trends and patterns in the data.

Objectives

The primary objective of this project is to conduct a comprehensive analysis of HR data to uncover actionable insights that can improve employee retention, enhance job satisfaction, and optimize workforce performance. By leveraging advanced data analysis techniques and predictive modeling.

Tools: Python, Power BI, Tableau, SQL, Excel, local studio.

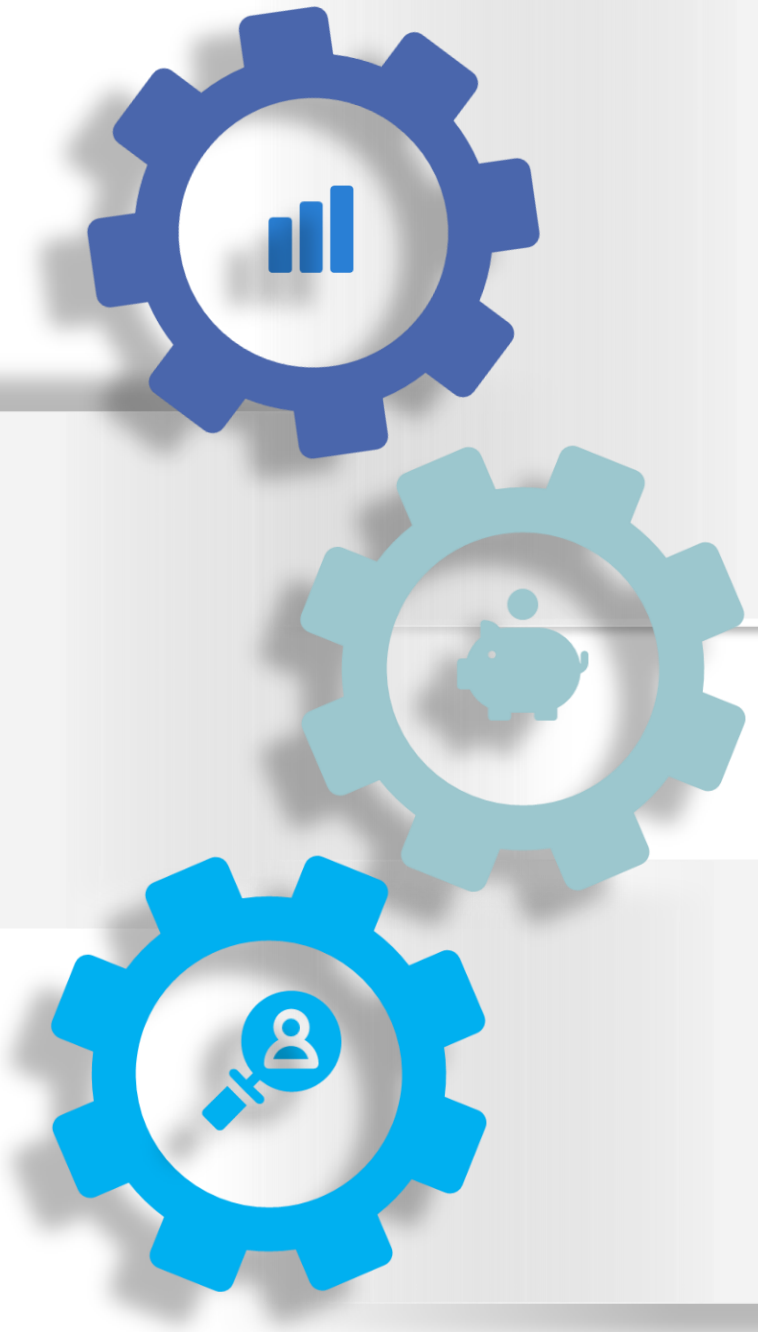


Analyze Age and Tenure:

- Calculate the average, highest, and lowest employee ages.
- Identify the year with the highest employee turnover.
- Calculate the average duration employees have been with the company.
- Explore the correlation between years at the company and marital status, as well as years at the company and salary.

Examine Salary Distribution:

- Analyze salaries based on gender and ethnicity.
- Identify the highest-paid department and job role.
- Investigate the relationship between marital status and salaries, as well as distance from home and overtime.



Department and Role Analysis:

- Determine the department with the fastest promotions.
- Assess the average age of employees by state.
- Analyze the correlation between work-life balance and marital status.

Performance and Promotions:

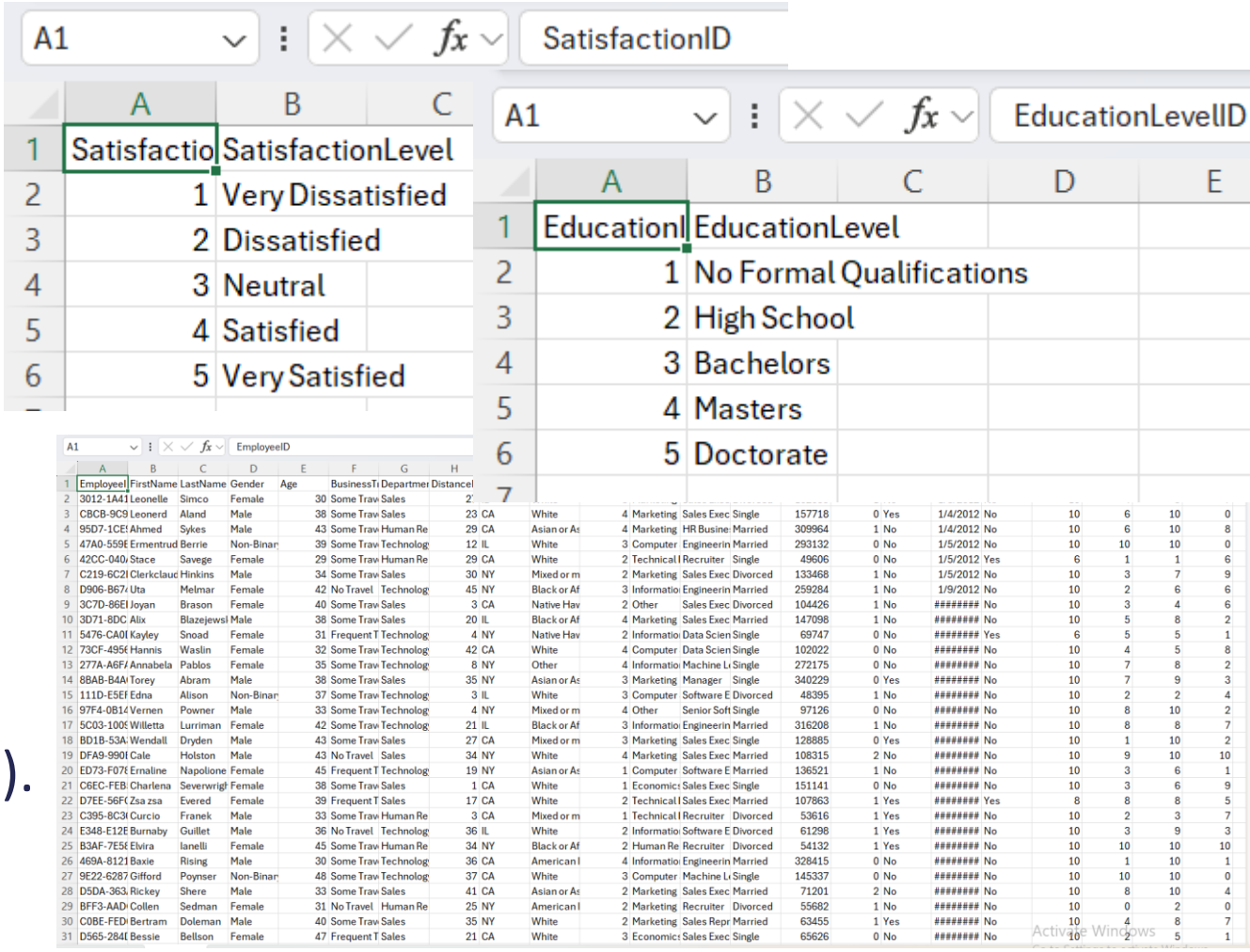
- Examine the relationship between manager ratings and promotions.
- Calculate the difference between self-ratings and manager ratings.
- Evaluate the correlation between training opportunities taken and salaries or promotions.

Visualize Employment Trends:

- Use a line chart to display hire dates by employee count.
- Present marital status distribution with a pie chart, highlighting any correlation with salaries and overtime.

Data Collection

- **Data Sources:**
  - Employee demographics (Gender, Age, Ethnicity, Marital Status).
  - Job details (Department, DistanceFromHome, BusinessTravel, HireDate).
  - Performance and tenure (PerformanceRating, YearsAtCompany, YearsSinceLastPromotion).
  - Compensation (Salary, StockOptionLevel).
  - Other factors (OverTime, Attrition).



A1 | SatisfactionID

	A	B	C
1	SatisfactionID	SatisfactionLevel	
2	1	Very Dissatisfied	
3	2	Dissatisfied	
4	3	Neutral	
5	4	Satisfied	
6	5	Very Satisfied	

A1 | EducationLevelID

	A	B	C	D	E
1	EducationLevelID	EducationLevel			
2	1	No Formal Qualifications			
3	2	High School			
4	3	Bachelors			
5	4	Masters			
6	5	Doctorate			

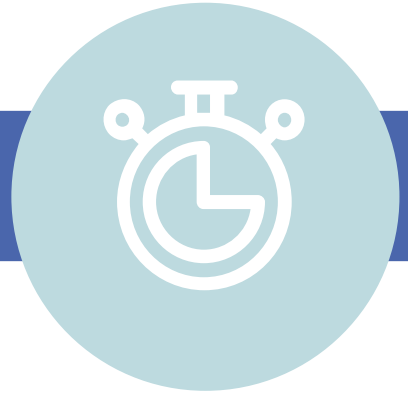
A1 | EmployeeID

	A	B	C	D	E	F	G	H
1	EmployeeID	FirstName	LastName	Gender	Age	BusinessTravel	Department	DistanceFromHome
2	3012-1A41	Leonelle	Simco	Female	30	Some Travel	Sales	2
3	CBCB-9C9	Leonerd	Aland	Male	38	Some Travel	Sales	23
4	95D7-1CE	Ahmed	Sykes	Male	43	Some Travel	Human Resources	29
5	47A0-559	Ermentrud	Berrie	Non-Binary	39	Some Travel	Technology	12
6	42CC-040	Stace	Savege	Female	29	Some Travel	Human Resources	29
7	C219-6C2	Clerkclaud	Hinkins	Male	34	Some Travel	Sales	30
8	D906-B67	Uta	Melmar	Female	42	No Travel	Technology	45
9	3C7D-86E	Joyan	Brason	Female	40	Some Travel	Sales	3
10	3D71-8DC	Alix	Blazejewski	Male	38	Some Travel	Sales	20
11	5476-CA0	Kayley	Snoad	Female	31	Frequent Travel	Technology	4
12	73CF-495	Hannis	Waslin	Female	32	Some Travel	Technology	42
13	277A-A6F	Annabela	Pablos	Female	35	Some Travel	Technology	8
14	8BAB-B4A	Torey	Abram	Male	38	Some Travel	Sales	35
15	111D-E5E	Edna	Alison	Non-Binary	37	Some Travel	Technology	3
16	97F4-0B1	Vernen	Powner	Male	33	Some Travel	Technology	4
17	5C03-100	Willetta	Lurriman	Female	42	Some Travel	Technology	21
18	BD1B-53A	Wendall	Dryden	Male	43	Some Travel	Sales	27
19	DFA9-990	Cale	Holston	Male	43	No Travel	Sales	34
20	ED73-F07	Ernaline	Napolione	Female	45	Frequent Travel	Technology	19
21	C6EC-FEB	Charlena	Severwright	Female	38	Some Travel	Sales	1
22	D7EE-56F	Zsa zsa	Evered	Female	39	Frequent Travel	Sales	17
23	C395-8C3	Curcio	Franek	Male	33	Some Travel	Human Resources	3
24	E348-E12	Burnaby	Guillet	Male	36	No Travel	Technology	36
25	B3AF-7E5	Elvira	Ianelli	Female	45	Some Travel	Human Resources	34
26	469A-8121	Baxie	Rising	Male	30	Some Travel	Technology	36
27	9E22-6287	Gifford	Poynser	Non-Binary	48	Some Travel	Technology	37
28	D5DA-363	Rickey	Shere	Male	33	Some Travel	Sales	41
29	BFF3-AAD	Collen	Sedman	Female	31	No Travel	Human Resources	25
30	C0BE-FED	Bertram	Doleman	Male	40	Some Travel	Sales	35
31	D565-284	Bessie	Bellson	Female	47	Frequent Travel	Sales	21

CA	White	4	Marketing	Sales Exec	Single	157718	0	Yes	1/4/2012	No	10	6	10	0
CA	Asian or As	4	Marketing	HR Busine	Married	309964	1	No	1/4/2012	No	10	6	10	8
IL	White	3	Computer	Engineerin	Married	293132	0	No	1/5/2012	No	10	10	10	0
CA	White	2	Technical I	Recruiter	Single	49606	0	No	1/5/2012	Yes	6	1	1	6
NY	Mixed or m	2	Marketing	Sales Exec	Divorced	133468	1	No	1/5/2012	No	10	3	7	9
NY	Black or Af	3	Informatio	Engineerin	Married	259284	1	No	1/9/2012	No	10	2	6	6
CA	Native Hav	2	Other	Sales Exec	Divorced	104426	1	No	########	No	10	3	4	6
IL	Black or Af	4	Marketing	Sales Exec	Married	147098	1	No	########	No	10	5	8	2
NY	Native Hav	2	Informatio	Data Scien	Single	69747	0	No	########	Yes	6	5	5	1
CA	White	4	Computer	Data Scien	Single	102022	0	No	########	No	10	4	5	8
NY	Other	4	Informatio	Machine Le	Single	272175	0	No	########	No	10	7	8	2
NY	Asian or As	3	Marketing	Manager	Single	340229	0	Yes	########	No	10	7	9	3
IL	White	3	Computer	Software E	Divorced	48395	1	No	########	No	10	2	2	4
NY	Mixed or m	4	Other	Senior Soft	Single	97126	0	No	########	No	10	8	10	2
IL	Black or Af	3	Informatio	Engineerin	Married	316208	1	No	########	No	10	8	8	7
CA	Mixed or m	3	Marketing	Sales Exec	Single	128885	0	Yes	########	No	10	1	10	2
NY	White	4	Marketing	Sales Exec	Married	108315	2	No	########	No	10	9	10	10
NY	Asian or As	1	Computer	Software E	Married	136521	1	No	########	No	10	3	6	1
CA	White	1	Economics	Sales Exec	Single	151141	0	No	########	No	10	3	6	9
CA	White	2	Technical I	Sales Exec	Married	107863	1	Yes	########	Yes	8	8	8	5
CA	Mixed or m	1	Technical I	Recruiter	Divorced	53616	1	Yes	########	No	10	2	3	7
IL	White	2	Informatio	Software E	Divorced	61298	1	Yes	########	No	10	3	9	3
NY	Black or Af	2	Human Re	Recruiter	Divorced	54132	1	Yes	########	No	10	10	10	10
CA	American I	4	Informatio	Engineerin	Married	328415	0	No	########	No	10	1	10	1
CA	White	3	Computer	Machine Le	Single	145337	0	No	########	No	10	10	10	0
CA	Asian or As	2	Marketing	Sales Exec	Married	71201	2	No	########	No	10	8	10	4
NY	American I	2	Marketing	Recruiter	Divorced	55682	1	No	########	No	10	0	2	0
NY	White	2	Marketing	Sales Repr	Married	63455	1	Yes	########	No	10	4	8	7
CA	White	3	Economics	Sales Exec	Single	65626	0	No	########	No	10	2	5	1

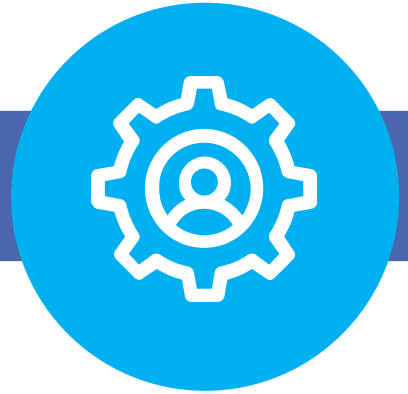
Data Wrangling

Merging Datasets:



Combined Employee, PerformanceRating, and EducationLevel datasets.

Handling Nested Data:



Used data explosion techniques to analyze nested data.

Key Output:

Clean, merged dataset ready for analysis.

Data Cleaning & Processing

- **Handling Missing Values:**
  - Removed or imputed missing data.
- **Correcting Errors:**
  - Fixed inconsistencies in data formats.
- **Removing Duplicates:**
  - Ensured data integrity by removing duplicate records.
- **Standardizing Formats:**
  - Normalized data for consistency.

Data cleaning

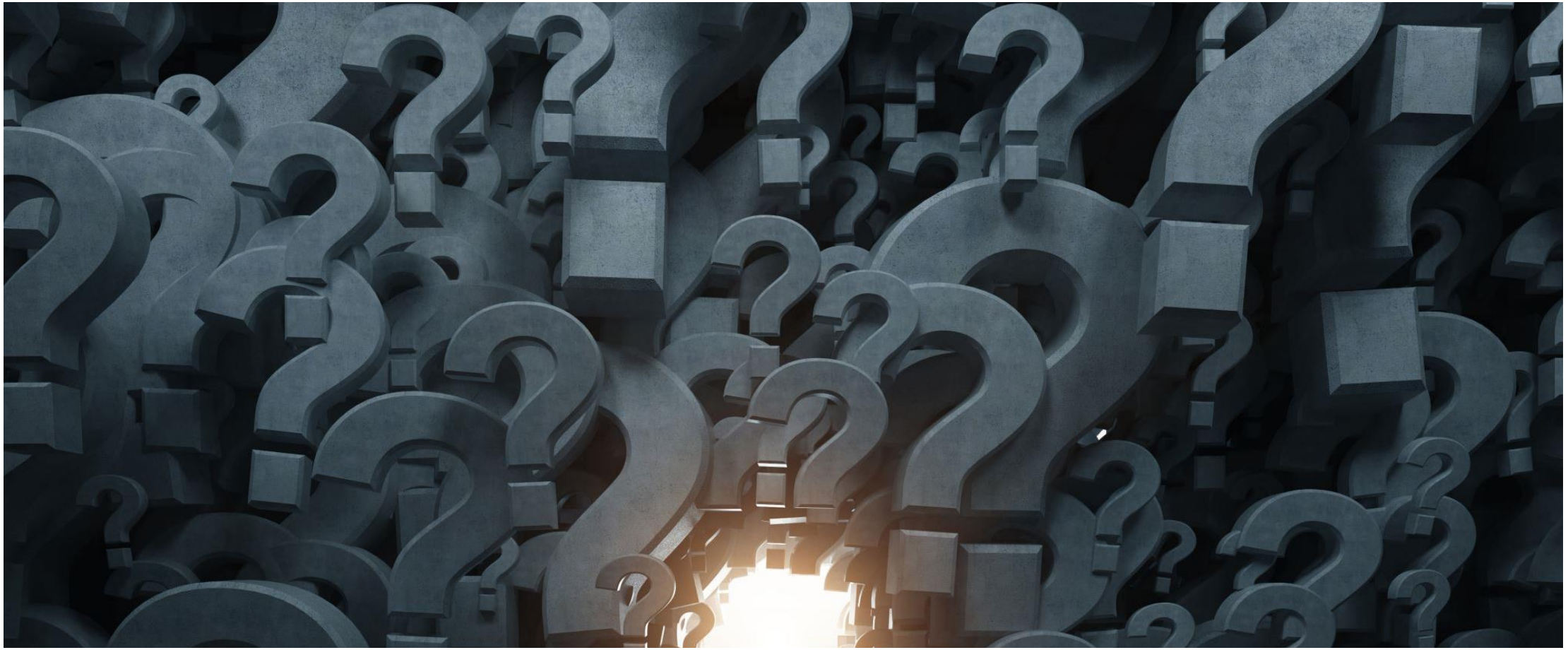
EmployeeID	FirstName	LastName	Gender	Age	BusinessTravel	Department	DistanceFromHome	State	Ethnicity	Education	EducationField	JobRole	MaritalStatus	Salary	StockOptionLevel	OverTime	HireDate	Attrition	YearsAtCompany	YearsInMostRecentRole	YearsSinceLastPromotion	YearsWithCurrManager
3012-1A41	Leonelle	Simco	Female	30	Some Travel	Sales	27	IL	White	5	Marketing	Sales Executive	Divorced	102059	1	No	1/3/2012	No	10	4	9	7
CBCB-9C9D	Leonerd	Aland	Male	38	Some Travel	Sales	23	CA	White	4	Marketing	Sales Executive	Single	157718	0	Yes	1/4/2012	No	10	6	10	0
95D7-1CE9	Ahmed	Sykes	Male	43	Some Travel	Human Resources	29	CA	Asian or Asian American	4	Marketing	HR Business Partner	Married	309964	1	No	1/4/2012	No	10	6	10	8
47A0-559B	Ermentrude	Berrie	Non-Binary	39	Some Travel	Technology	12	IL	White	3	Computer Science	Engineering Manager	Married	293132	0	No	1/5/2012	No	10	10	10	0
42CC-040A	Stace	Savege	Female	29	Some Travel	Human Resources	29	CA	White	2	Technical Degree	Recruiter	Single	49606	0	No	1/5/2012	Yes	6	1	1	6
C219-6C2E	Clerkclaude	Hinkins	Male	34	Some Travel	Sales	30	NY	Mixed or multiple ethnic groups	2	Marketing	Sales Executive	Divorced	133468	1	No	1/5/2012	No	10	3	7	9
D906-B674	Uta	Melmar	Female	42	No Travel	Technology	45	NY	Black or African American	3	Information Systems	Engineering Manager	Married	259284	1	No	1/9/2012	No	10	2	6	6
3C7D-86ED	Joyan	Brason	Female	40	Some Travel	Sales	3	CA	Native Hawaiian	2	Other	Sales Executive	Divorced	104426	1	No	1/11/2012	No	10	3	4	6
3D71-8DC2	Alix	Blazejewski	Male	38	Some Travel	Sales	20	IL	Black or African American	4	Marketing	Sales Executive	Married	147098	1	No	1/11/2012	No	10	5	8	2
5476-CA0D	Kayley	Snoad	Female	31	Frequent Traveller	Technology	4	NY	Native Hawaiian	2	Information Systems	Data Scientist	Single	69747	0	No	1/12/2012	Yes	6	5	5	1
73CF-4956	Hannis	Waslin	Female	32	Some Travel	Technology	42	CA	White	4	Computer Science	Data Scientist	Single	102022	0	No	1/12/2012	No	10	4	5	8
277A-A6FA	Annabela	Pablos	Female	35	Some Travel	Technology	8	NY	Other	4	Information Systems	Machine Learning Engineer	Single	272175	0	No	1/12/2012	No	10	7	8	2
8BAB-B4A6	Torey	Abram	Male	38	Some Travel	Sales	35	NY	Asian or Asian American	3	Marketing	Manager	Single	340229	0	Yes	1/13/2012	No	10	7	9	3
111D-E5EF	Edna	Alison	Non-Binary	37	Some Travel	Technology	3	IL	White	3	Computer Science	Software Engineer	Divorced	48395	1	No	1/15/2012	No	10	2	2	4
97F4-0B14	Vernen	Powner	Male	33	Some Travel	Technology	4	NY	Mixed or multiple ethnic groups	4	Other	Senior Software Engineer	Single	97126	0	No	1/15/2012	No	10	8	10	2
5C03-1009	Willetta	Lurriman	Female	42	Some Travel	Technology	21	IL	Black or African American	3	Information Systems	Engineering Manager	Married	316208	1	No	1/17/2012	No	10	8	8	7
BD1B-53A3	Wendall	Dryden	Male	43	Some Travel	Sales	27	CA	Mixed or multiple ethnic groups	3	Marketing	Sales Executive	Single	128885	0	Yes	1/17/2012	No	10	1	10	2
DFA9-990E	Cale	Holston	Male	43	No Travel	Sales	34	NY	White	4	Marketing	Sales Executive	Married	108315	2	No	1/17/2012	No	10	9	10	10
ED73-F078	Ernaline	Napolione	Female	45	Frequent Traveller	Technology	19	NY	Asian or Asian American	1	Computer Science	Software Engineer	Married	136521	1	No	1/18/2012	No	10	3	6	1
C6EC-FEB5	Charlena	Severwright	Female	38	Some Travel	Sales	1	CA	White	1	Economics	Sales Executive	Single	151141	0	No	1/19/2012	No	10	3	6	9
D7EE-56FC	Zsa zsa	Evered	Female	39	Frequent Traveller	Sales	17	CA	White	2	Technical Degree	Sales Executive	Married	107863	1	Yes	1/19/2012	Yes	8	8	8	5
C395-8C36	Curcio	Franek	Male	33	Some Travel	Human Resources	3	CA	Mixed or multiple ethnic groups	1	Technical Degree	Recruiter	Divorced	53616	1	Yes	1/19/2012	No	10	2	3	7
E348-E12B	Burnaby	Guillet	Male	36	No Travel	Technology	36	IL	White	2	Information Systems	Software Engineer	Divorced	61298	1	Yes	1/19/2012	No	10	3	9	3
B3AF-7E58	Elvira	Ianelli	Female	45	Some Travel	Human Resources	34	NY	Black or African American	2	Human Resources	Recruiter	Divorced	54132	1	Yes	1/20/2012	No	10	10	10	10
469A-8121	Baxie	Rising	Male	30	Some Travel	Technology	36	CA	American Indian or Alaska Native	4	Information Systems	Engineering Manager	Married	328415	0	No	1/20/2012	No	10	1	10	1
9E22-6287	Gifford	Poynser	Non-Binary	48	Some Travel	Technology	37	CA	White	3	Computer Science	Machine Learning Engineer	Single	145337	0	No	1/21/2012	No	10	10	10	0
D5DA-363A	Rickey	Shere	Male	33	Some Travel	Sales	41	CA	Asian or Asian American	2	Marketing	Sales Executive	Married	71201	2	No	1/24/2012	No	10	8	10	4
BFF3-AADC	Collen	Sedman	Female	31	No Travel	Human Resources	25	NY	American Indian or Alaska Native	2	Marketing	Recruiter	Divorced	55682	1	No	1/25/2012	No	10	0	2	0
C0BE-FED0	Bertram	Doleman	Male	40	Some Travel	Sales	35	NY	White	2	Marketing	Sales Representative	Married	63455	1	Yes	1/26/2012	No	10	4	8	7
D565-284D	Bessie	Bellson	Female	47	Frequent Traveller	Sales	21	CA	White	3	Economics	Sales Executive	Single	65626	0	No	1/27/2012	No	10	2	5	1
00D4-DD53	Joyce	Goor	Female	30	Frequent Traveller	Technology	44	CA	Black or African American	1	Computer Science	Software Engineer	Single	68508	0	Yes	1/28/2012	Yes	5	4	4	4
7FFD-C810	Claresta	Impy	Female	32	Some Travel	Technology	22	IL	White	2	Other	Data Scientist	Married	75821	1	No	1/30/2012	No	10	1	5	1
07B2-D67A	Rossie	Everleigh	Male	44	Some Travel	Sales	40	NY	White	2	Economics	Manager	Divorced	285620	1	No	1/30/2012	No	10	6	8	4
1749-81A2	Elora	Bentjens	Female	48	Some Travel	Technology	17	CA	Black or African American	3	Information Systems	Software Engineer	Single	65860	0	No	2/1/2012	No	10	6	7	1
FF14-A43E	Koenraad	Nannizzi	Male	47	Some Travel	Technology	41	NY	American Indian or Alaska Native	1	Computer Science	Data Scientist	Married	40786	0	No	2/1/2012	No	10	3	5	5
3CD6-5587	Dorise	Klishin	Female	31	Some Travel	Technology	20	CA	White	2	Computer Science	Software Engineer	Single	59697	0	Yes	2/2/2012	Yes	7	5	6	0
4FC2-A40D	Gayle	Riseley	Female	32	No Travel	Technology	12	NY	Black or African American	3	Computer Science	Engineering Manager	Married	316725	1	No	2/2/2012	No	10	6	10	2
9AAB-DCD0	Staci	Leith	Female	28	Some Travel	Technology	14	IL	White	2	Business Studies	Software Engineer	Divorced	105984	2	Yes	2/4/2012	No	10	7	9	10
427A-8C69	Pasquale	Abreheart	Male	32	Some Travel	Technology	15	CA	Mixed or multiple ethnic groups	3	Other	Analytics Manager	Married	393294	1	No	2/4/2012	No	10	2	2	2
A923-9EF6	Alaine	Hinrichsen	Female	41	Frequent Traveller	Technology	7	CA	Black or African American	1	Business Studies	Machine Learning Engineer	Single	107008	0	No	2/5/2012	No	10	9	10	9
5C61-8F3A	Dyana	Gallie	Female	34	Some Travel	Technology	5	CA	White	1	Information Systems	Data Scientist	Single	127432	0	Yes	2/6/2012	No	10	8	9	9
CF2F-8CA3	Lindy	Rawstorne	Male	45	Some Travel	Sales	20	CA	White	2	Marketing	Sales Executive	Married	76693	1	No	2/6/2012	No	10	2	9	10
40A9-8EB0	Joannes	McFadden	Female	39	Some Travel	Technology	31	CA	White	3	Information Systems	Software Engineer	Single	90017	0	Yes	2/7/2012	No	10	7	7	3
2E72-4BF1	Grace	Gohier	Male	44	Some Travel	Sales	21	CA	White	3	Marketing	Manager	Married	426142	0	No	2/7/2012	No	10	4	5	4
EE39-C8F2	Jaclin	Chadburn	Female	47	Some Travel	Technology	17	NY	White	2	Computer Science	Software Engineer	Married	46514	2	No	2/12/2012	No	10	9	9	2
C0C7-A3EF	Hagen	Worge	Male	48	No Travel	Sales	9	CA	White	5	Economics	Sales Executive	Single	125587	0	No	2/12/2012	No	10	10	10	4
79F3-8AA5	Lacy	Domerc	Female	47	Some Travel	Human Resources	21	NY	Mixed or multiple ethnic groups	3	Marketing	HR Executive	Single	98608	0	No	2/14/2012	No	10	5	10	4
9528-AA17	Jaquelin	Kite	Female	48	Some Travel	Technology	30	NY	Black or African American	3	Information Systems	Software Engineer	Married	72794	1	No	2/14/2012	No	10	4	5	3
9C71-34AE	Yvor	Walaron	Non-Binary	40	Some Travel	Technology	44	CA	White	2	Computer Science	Software Engineer	Single	69119	0	No	2/15/2012	No	10	8	10	5
C79A-E5E1	Alvera	Kulver	Female	32	Frequent Traveller	Technology	13	CA	White	1	Computer Science	Data Scientist	Divorced	48835	1	Yes	2/17/2012	Yes	8	5	7	0
5468-EEE1	Forbes	Toretta	Male	28	No Travel	Technology	44	NY	White	2	Computer Science	Senior Software Engineer	Married	218406	1	No	2/17/2012	No	10	10	10	10
B50B-5A18	Paulina	Senecaut	Female	40	Frequent Traveller	Technology	27	NY	Black or African American	4	Computer Science	Software Engineer	Divorced	42333	1	No	2/19/2012	No	10	4	5	5
11AE-3E4D	Sheffie	Tonkin	Male	39	Some Travel	Sales	38	CA	Mixed or multiple ethnic groups	2	Technical Degree	Manager	Married	300299	1	Yes	2/19/2012	No	10	7	7	10
151D-1FF2	Kai	Lush	Female	34	Frequent Traveller	Sales	15	CA	White	2	Marketing	Sales Executive	Single	173246	0	Yes	2/20/2012	Yes	9	4	7	7
D584-4E31	Patrica	Lamyman	Female	35	Some Travel	Human Resources	22	CA	Mixed or multiple ethnic groups	1	Economics	Recruiter	Married	67264	0	No	2/22/2012	No	10	8	8	2
0FC8-561B	Caryl	Loving	Male	29	Frequent Traveller	Technology	14	IL	Asian or Asian American	3	Computer Science	Senior Software Engineer	Married	93591	1	No	2/22/2012	No	10	4	10	3
A531-8159	Carlyn	Pavese	Female	37	No Travel	Technology	16	NY	Mixed or multiple ethnic groups	3	Information Systems	Data Scientist	Divorced	82099	2	No	2/23/2012	No	10	1	7	6
E7A6-33A3	Arnye	Townes	Female	38	Some Travel	Technology	39	IL	White	4	Information Systems	Data Scientist	Married	83590	1	No	2/23/2012	No	10	2	3	9
E4F4-11B9	Susanne	Wisam	Female	29	Some Travel	Sales	19	NY	White	2	Marketing	Sales Executive	Single	122811	0	Yes	2/27/2012	Yes	2	1	1	1
586E-5C14	Gunther	Keays	Male	40	Some Travel	Technology	9	NY	Black or African American	1	Computer Science	Software Engineer	Single	58551	0	No	2/27/2012	No	10	6	7	3
8A40-C85A	Reggie	Matthewman	Female	44	Some Travel	Technology	28	CA	Black or African American	3	Information Systems	Software Engineer	Married	53325	1	Yes	2/28/2012	No	10	4	8	4
2B08-7D3D	Roderic	Daddow	Male	45	Some Travel	Sales	19	IL	White	4	Marketing	Sales Representative	Married	47326	1	No	2/29/2012	No	10	10	10	10

Employee | PerformanceRating | EducationLevel | RatingLevel | SatisfiedLevel

Data Exploration & Transformation

- **Exploratory Data Analysis (EDA):**
  - Analyzed data distribution, patterns, and relationships.
- **Data Transformation:**
  - Created new variables (e.g., tenure groups, age groups).
- **Visualizations:**
  - Used histograms, scatter plots, and box plots to understand data.

Key Insights
Business Questions

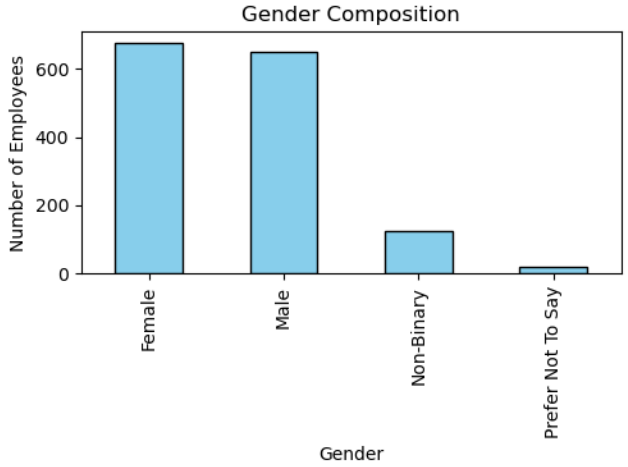


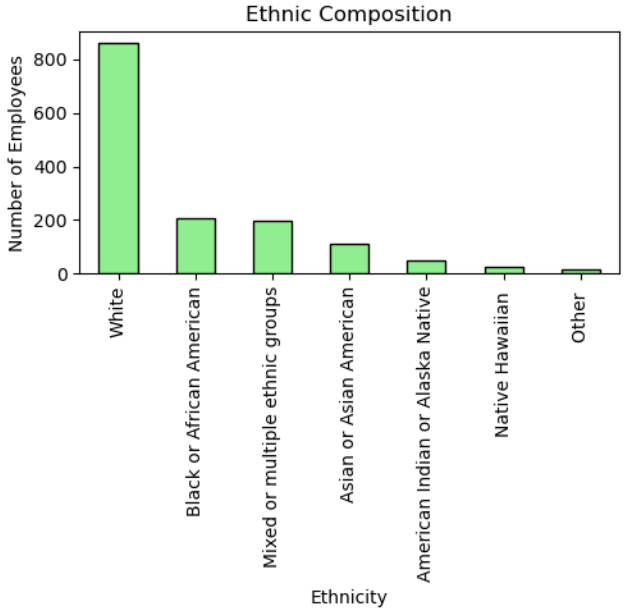
Key Insights
Workforce Demographics

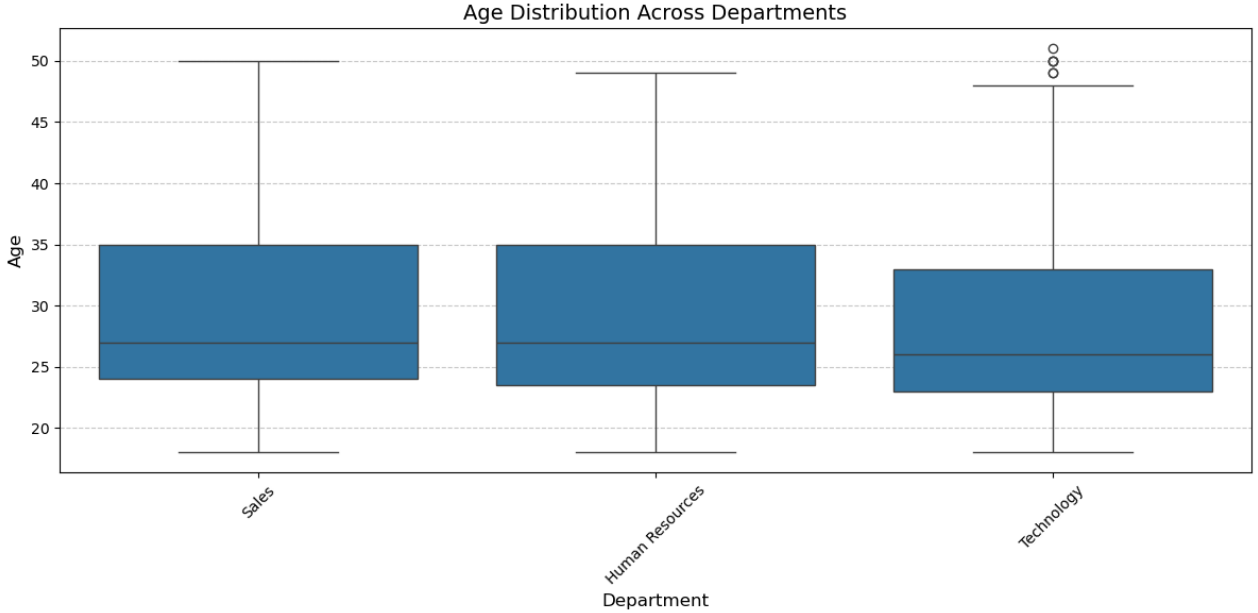
```python
import pandas as pd
import matplotlib.pyplot as plt

# Correct file path
file_path = r'F:\data analysis\BQ\HR Data.xlsx'
sheet_name = 'Employee'

# Load the data
employee_data = pd.read_excel(file_path, sheet_name=sheet_name)

# Analyze gender and ethnic composition
def workforce_demographics(data):

    gender_counts = data['Gender'].value_counts()
    ethnicity_counts = data['Ethnicity'].value_counts()

    # Plot gender composition
    plt.figure(figsize=(10, 5))
    plt.subplot(1, 2, 1)
    gender_counts.plot(kind='bar', color='skyblue', edgecolor='black')
    plt.title('Gender Composition')
    plt.ylabel('Number of Employees')
    plt.xlabel('Gender')

    # Plot ethnicity composition
    plt.subplot(1, 2, 2)
    ethnicity_counts.plot(kind='bar', color='lightgreen', edgecolor='black')
    plt.title('Ethnic Composition')
    plt.ylabel('Number of Employees')
    plt.xlabel('Ethnicity')

    plt.tight_layout()
    plt.savefig('workforce_demographics.png')  # Save the plot
    plt.show()

# Call the function
workforce_demographics(employee_data)

```

```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Set the file path and sheet name
file_path = r'F:\data analysis\BQ\HR Data.xlsx'
sheet_name = 'Employee'

# Load the data
employee_data = pd.read_excel(file_path, sheet_name=sheet_name)

# Ensure the required columns exist
if 'Age' in employee_data.columns and 'Department' in employee_data.columns:
    # Plot the age distribution across departments
    plt.figure(figsize=(12, 6))
    sns.boxplot(data=employee_data, x='Department', y='Age')
    plt.title('Age Distribution Across Departments', fontsize=14)
    plt.xlabel('Department', fontsize=12)
    plt.ylabel('Age', fontsize=12)
    plt.xticks(rotation=45)
    plt.grid(axis='y', linestyle='--', alpha=0.7)
    plt.tight_layout()
    plt.show()
else:
    print("Make sure the columns 'Age' and 'Department' exist in the dataset.")

```

Key Insights
Workforce Demographics

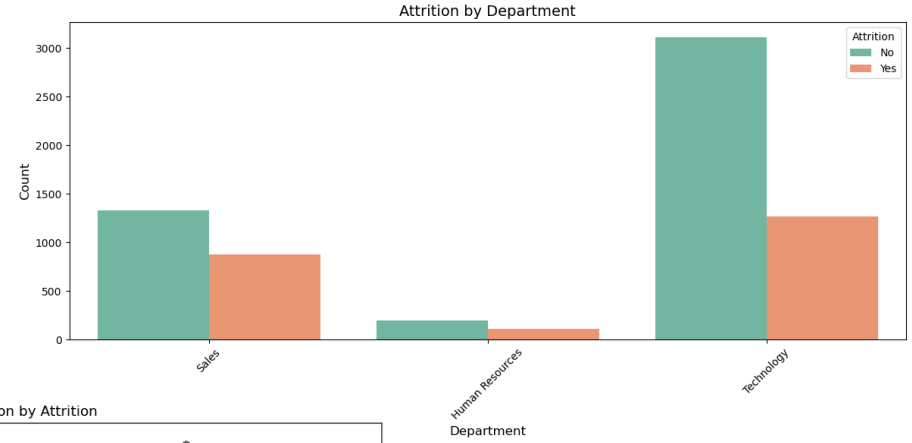


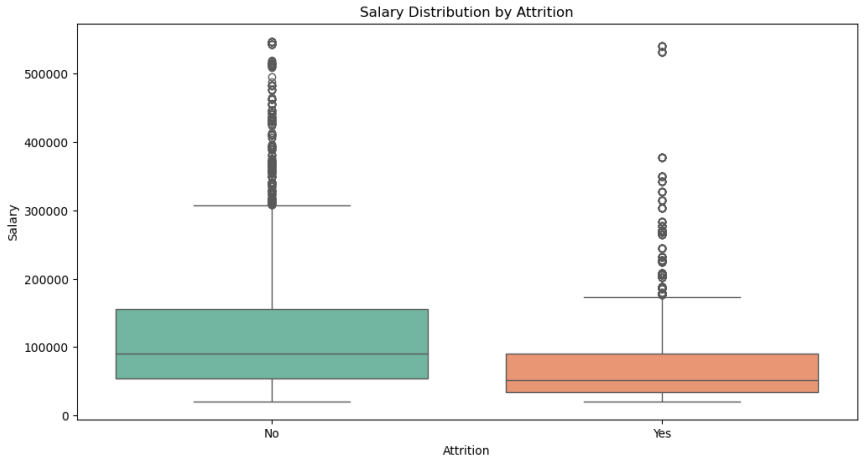


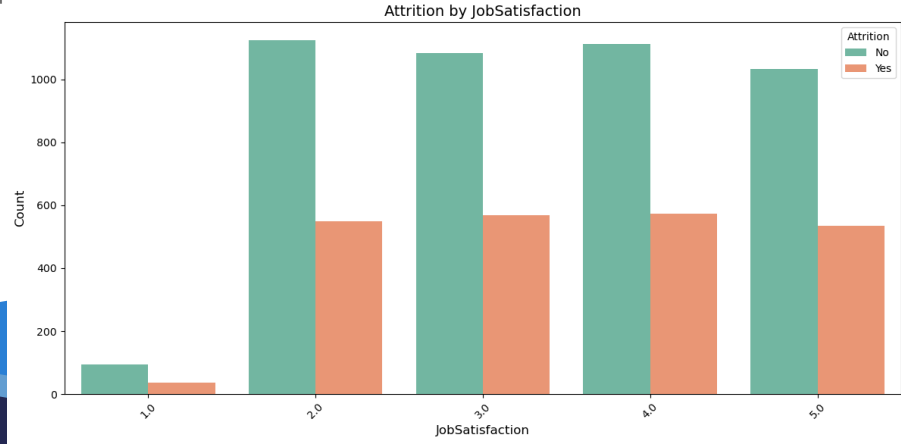


Key Insights - Employee Attrition

```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Set the file path and sheet names
file_path = r'F:\data analysis\BQ\HR Data.xlsx'
employee_sheet = 'Employee'
performance_sheet = 'PerformanceRating'

# Load the data
employee_data = pd.read_excel(file_path, sheet_name=employee_sheet)
performance_data = pd.read_excel(file_path, sheet_name=performance_sheet)

# Merge performance data to get JobSatisfaction
if 'EmployeeID' in employee_data.columns and 'EmployeeID' in performance_data.columns:
    employee_data = employee_data.merge(performance_data[['EmployeeID', 'JobSatisfaction']], on='EmployeeID', how='left')
else:
    print("EmployeeID column not found in one of the datasets.")

# Rename MonthlyIncome to salary
if 'MonthlyIncome' in employee_data.columns:
    employee_data.rename(columns={'MonthlyIncome': 'Salary'}, inplace=True)

    # Drop rows where salary is missing
    employee_data = employee_data.dropna(subset=['Salary'])

    # Categorize salary into bins
    bins = [0, 3000, 7000, 12000, 20000, employee_data['Salary'].max()]
    labels = ['Low', 'Lower-Mid', 'Mid', 'Upper-Mid', 'High']
    employee_data['salary_category'] = pd.cut(employee_data['Salary'], bins=bins, labels=labels)

# Check for the 'Attrition' column
if 'Attrition' in employee_data.columns:
    # Ensure salary_category exists before analysis
    if 'salary_category' in employee_data.columns:
        key_factors = ['Age', 'Department', 'JobRole', 'Salary', 'JobSatisfaction']
    else:
        key_factors = ['Age', 'Department', 'JobRole', 'JobSatisfaction']

    for factor in key_factors:
        if factor in employee_data.columns:
            plt.figure(figsize=(12, 6))
            sns.countplot(data=employee_data, x=factor, hue='Attrition', palette='Set2')
            plt.title(f'Attrition by {factor}', fontsize=14)
            plt.xlabel(factor, fontsize=12)
            plt.ylabel('Count', fontsize=12)
            plt.xticks(rotation=45)
            plt.legend(title='Attrition', loc='upper right')
            plt.tight_layout()
            plt.show()
        else:
            print(f"Column '{factor}' not found in the dataset.")
else:
    print("The dataset does not contain an 'Attrition' column.")
```

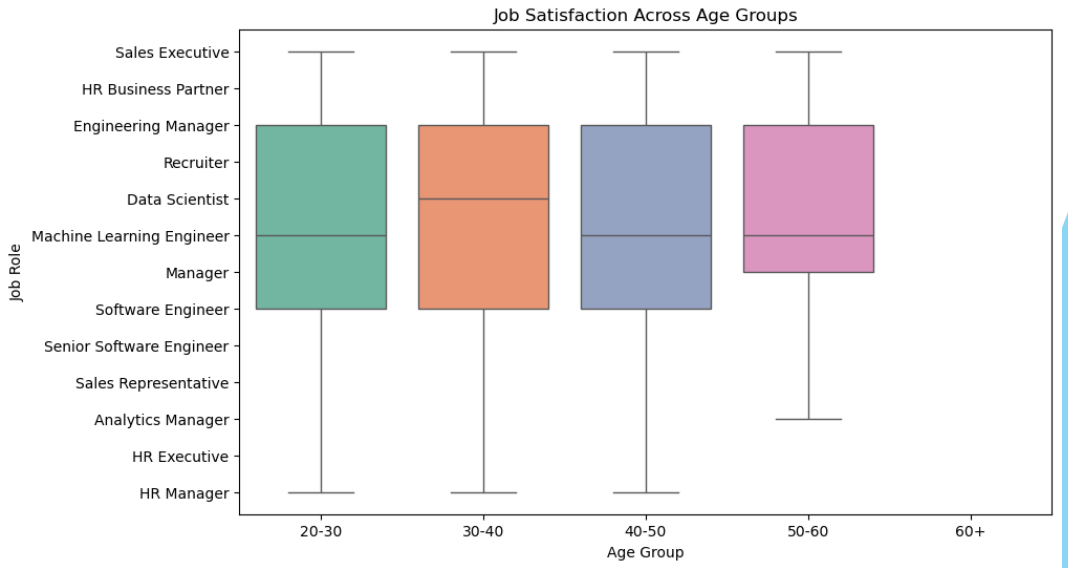


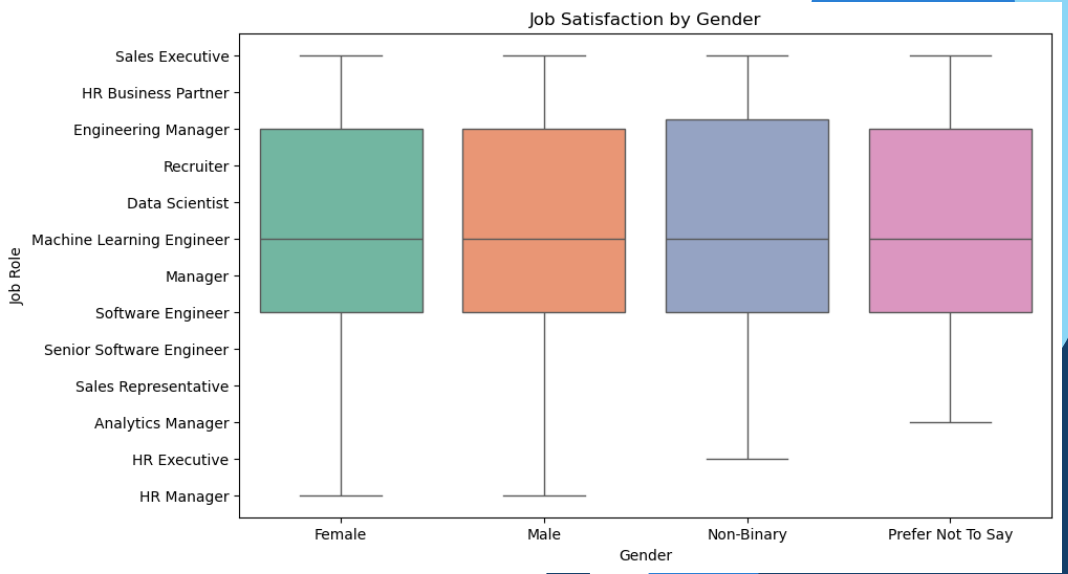




Key Insights - Performance & Satisfaction

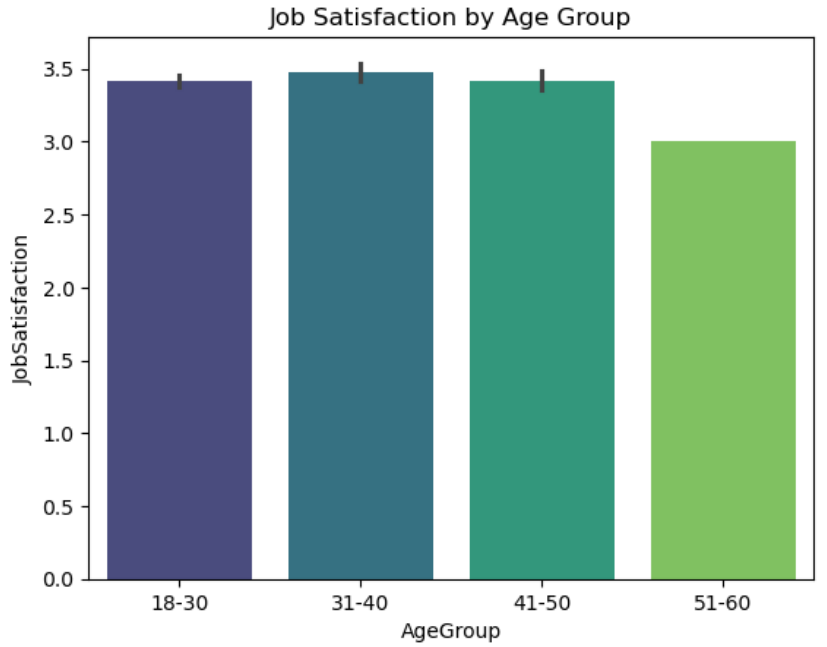
```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats

# Load your dataset
file_path = r'F:\data analysis\BQ\HR Data.xlsx'
sheet_name = 'Employee'
employee_data = pd.read_excel(file_path, sheet_name=sheet_name)

# Ensure necessary columns exist
required_columns = ['Attrition', 'Age', 'JobRole', 'Department', 'Gender']
if all(col in employee_data.columns for col in required_columns):

    # Convert Attrition to numerical values (Yes=1, No=0) if not already done
    if 'AttritionNumeric' not in employee_data.columns:
        employee_data['AttritionNumeric'] = employee_data['Attrition'].map({'Yes': 1, 'No': 0})

    # 8. Age Impact on Job Satisfaction and Attrition:
    print("\nAge Impact on Job Satisfaction and Attrition:")

    # Create age groups (e.g., 20-30, 30-40, 40-50, 50+)
    age_bins = [20, 30, 40, 50, 60, 70]
    age_labels = ['20-30', '30-40', '40-50', '50-60', '60+']
    employee_data['AgeGroup'] = pd.cut(employee_data['Age'], bins=age_bins, labels=age_labels, right=False)

    # 1. How does job satisfaction vary across different age groups?
    # Assuming 'JobRole' or other columns are related to job satisfaction
    # If you have a specific column for job satisfaction, you can use it here
    plt.figure(figsize=(10, 6))
    sns.boxplot(x='AgeGroup', y='JobRole', data=employee_data, palette='Set2')
    plt.title('Job Satisfaction Across Age Groups')
    plt.xlabel('Age Group')
    plt.ylabel('Job Role')
    plt.show()

    # 2. What is the attrition rate for different age groups?
    attrition_rate_by_age = employee_data.groupby('AgeGroup')['AttritionNumeric'].mean()
    print("\nAttrition Rate by Age Group:")
    print(attrition_rate_by_age)

    # Visualize Attrition Rate by Age Group
    plt.figure(figsize=(10, 6))
    sns.barplot(x=attrition_rate_by_age.index, y=attrition_rate_by_age.values, palette='Set2')
    plt.title('Attrition Rate by Age Group')
    plt.xlabel('Age Group')
    plt.ylabel('Attrition Rate')
    plt.show()

else:
    print("Some required columns are missing.")
```

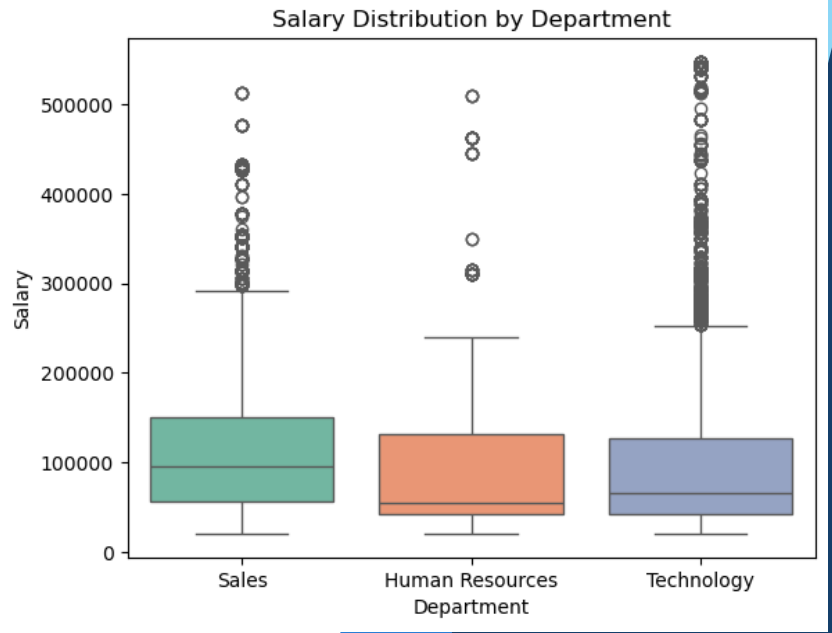




Key Insights - Compensation Analysis

```
...
    Salary Analysis Across Demographics and Departments:
    Average Salary by Department:
    Department
    Human Resources    105804.270627
    Sales              121045.081899
    Technology         106396.216115
    Name: Salary, dtype: float64
    ANOVA Test for Salary by Department: F-statistic = 16.40, P-value = 0.0000
    T-test for Salary by Gender: T-statistic = 1.61, P-value = 0.1079

    Stock Option Level and Retention (Attrition):
    Correlation between StockOptionLevel and Attrition: -0.17

    Age Impact on Job Satisfaction and Attrition:
    Average Job Satisfaction by Age Group:
    AgeGroup
    18-30    3.416400
    31-40    3.477931
    41-50    3.420327
    51-60    3.000000
    Name: JobSatisfaction, dtype: float64
    Attrition Rate by Age Group:
    AgeGroup
    18-30    0.470574
    ...
    Divorced    3.405199
    Married     3.440098
    Single      3.433018
    Name: JobSatisfaction, dtype: float64
```

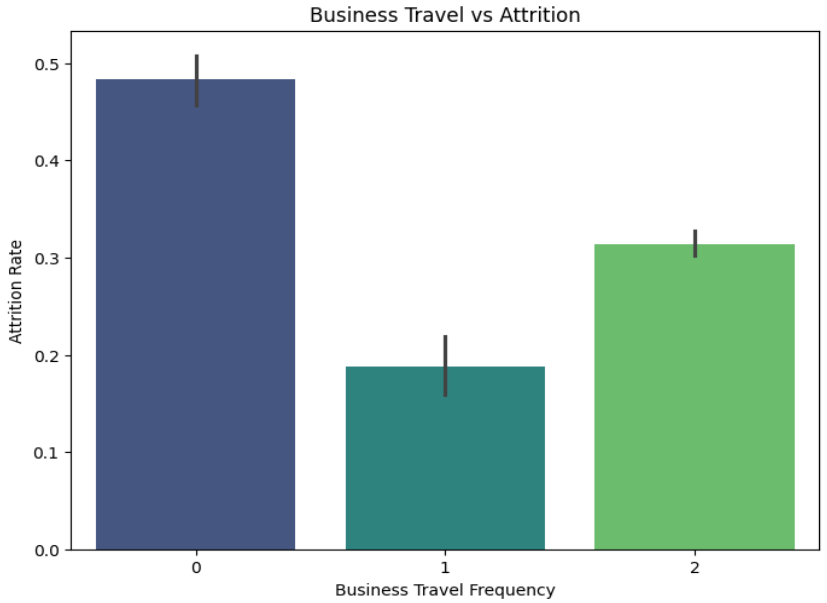


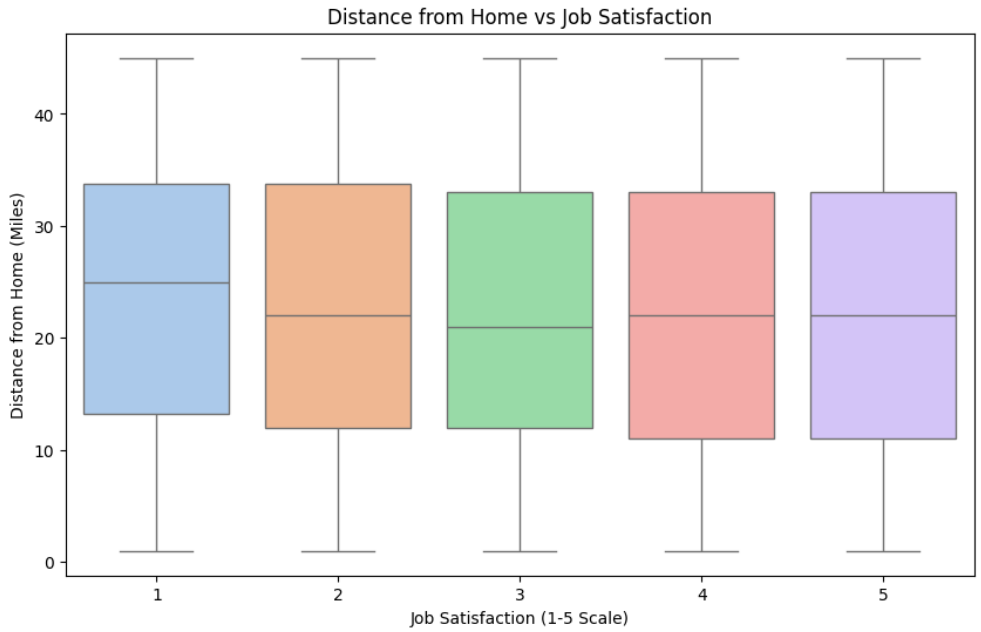


Key Insights - Geographic Insights

```python
# Average Job Satisfaction and Attrition by Business Travel category
travel_satisfaction = df.groupby('BusinessTravel')[['JobSatisfaction', 'Attrition']].mean()
print(f"Business Travel and Job Satisfaction:\n{travel_satisfaction}")

# Visualization: Business Travel vs Job Satisfaction
plt.figure(figsize=(8, 6))
sns.barplot(x='BusinessTravel', y='JobSatisfaction', data=df, palette='viridis')
plt.title('Business Travel vs Job Satisfaction')
plt.xlabel('Business Travel Frequency')
plt.ylabel('Average Job Satisfaction')
plt.show()

# Visualization: Business Travel vs Attrition
plt.figure(figsize=(8, 6))
sns.barplot(x='BusinessTravel', y='Attrition', data=df, palette='viridis')
plt.title('Business Travel vs Attrition')
plt.xlabel('Business Travel Frequency')
plt.ylabel('Attrition Rate')
plt.show()
```

```
Business Travel and Job Satisfaction:
               JobSatisfaction  Attrition
BusinessTravel
0                     3.410162   0.483063
1                     3.489362   0.188216
2                     3.428903   0.314346
```





Data Modeling
Predictive Analysis

Predictive Analysis

Attrition Prediction:

•The model demonstrates strong accuracy in predicting employee attrition.

• Key factors influencing attrition include overtime hours, salary levels, and the number of years spent at the company.

Performance Prediction:

•Salary, age, and tenure at the company are the most significant factors affecting employee performance, with salary being the most influential.

Promotion Readiness:

Employees with longer tenure and higher manager ratings are more likely to be promoted.

Salary Prediction:

The model estimates salaries by analyzing department, years of experience, and job role, maintaining a reasonable margin of error in its predictions.

Data Forecasting

•Future Trends:

•**New hires show high attrition risk** without preventive measures.

•**Overtime employees face concerning resignation trends**.

•**Hybrid work models effectively reduce turnover** and boost satisfaction.

Data Visualization

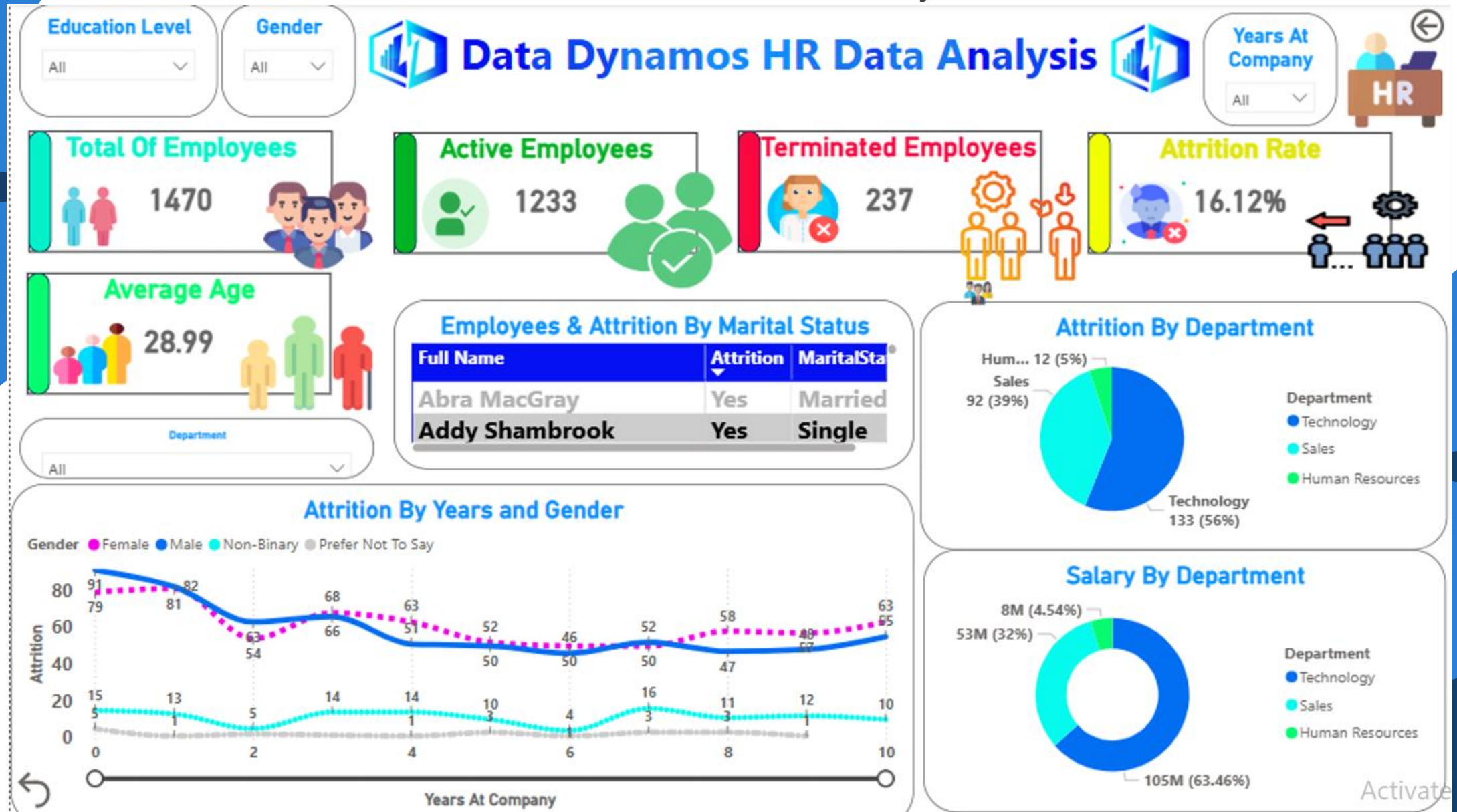
Tools: Tableau, Power BI & Excel

Data Visualization

Tools: Tableau, Power BI & Excel

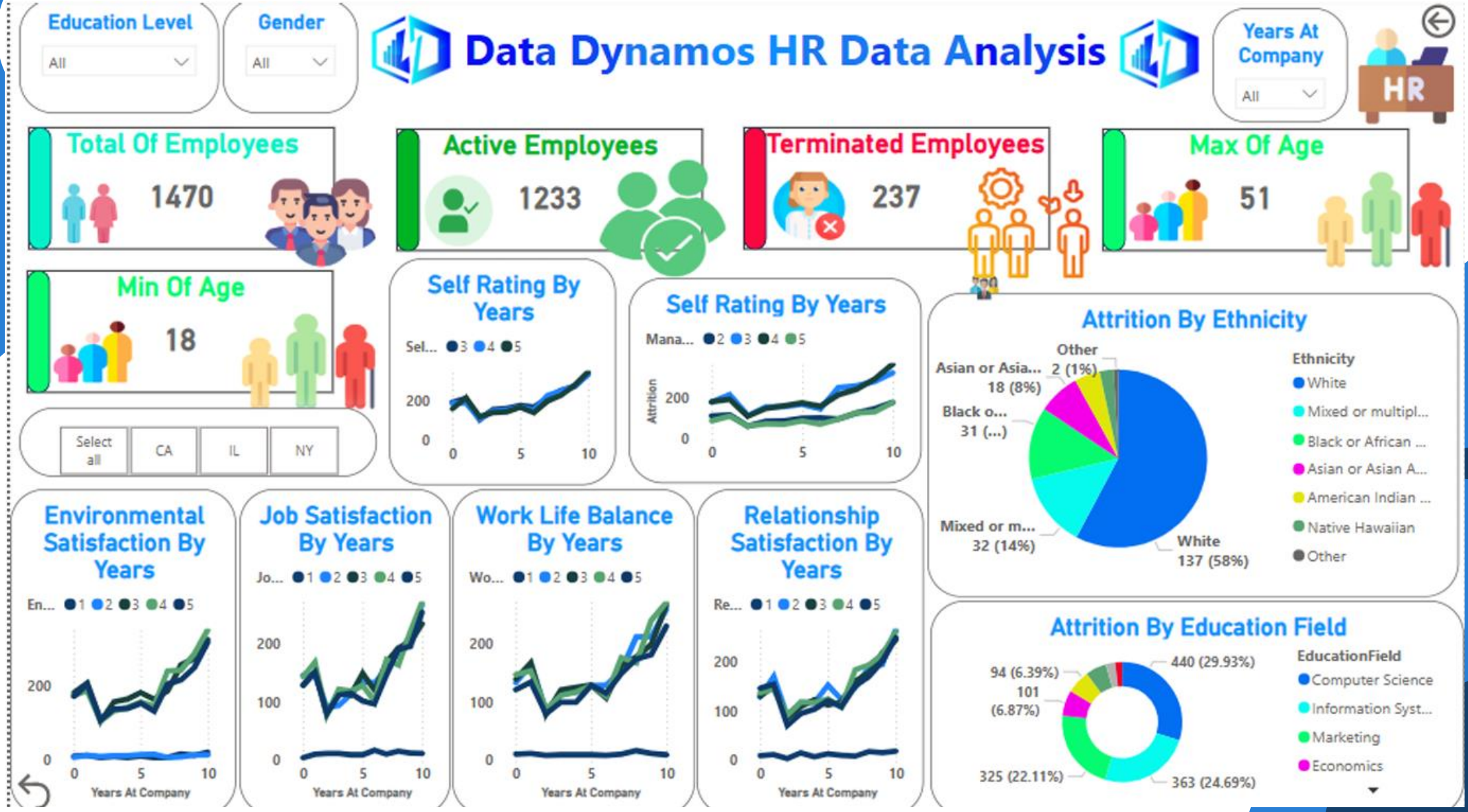
Data Visualization

Power BI for Data Analysis



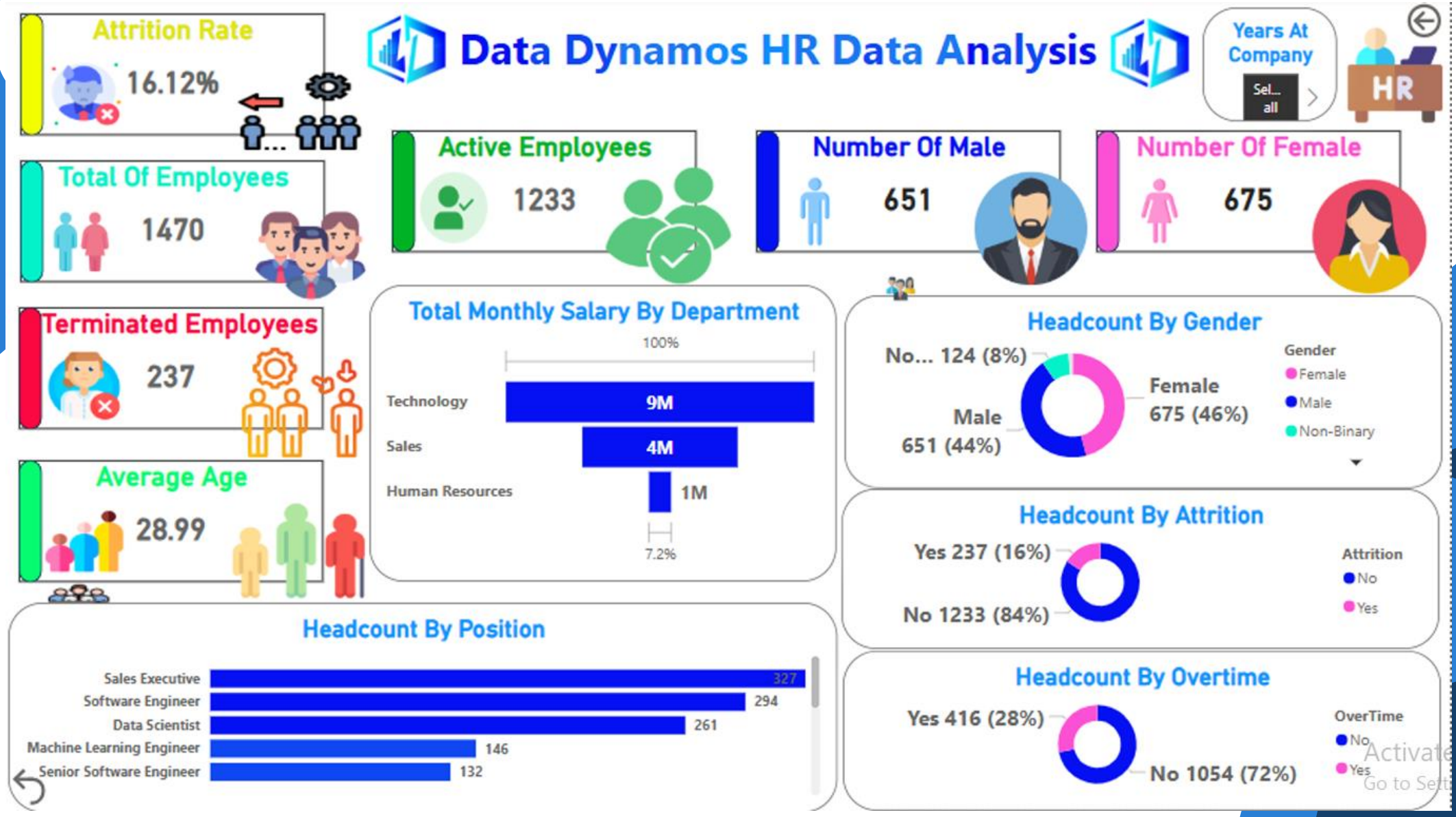
Data Visualization

Power BI for Data Analysis



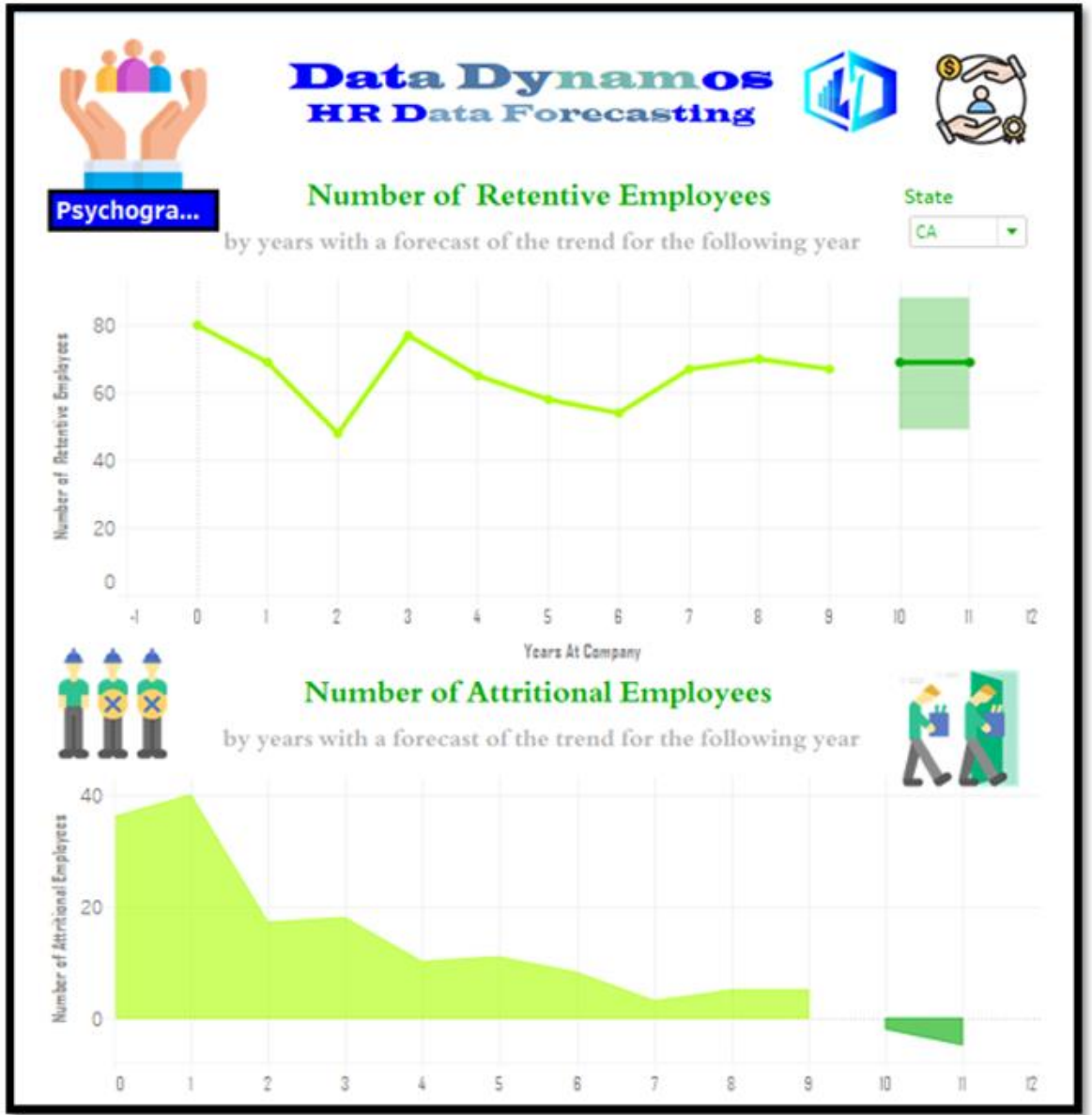
Data Visualization

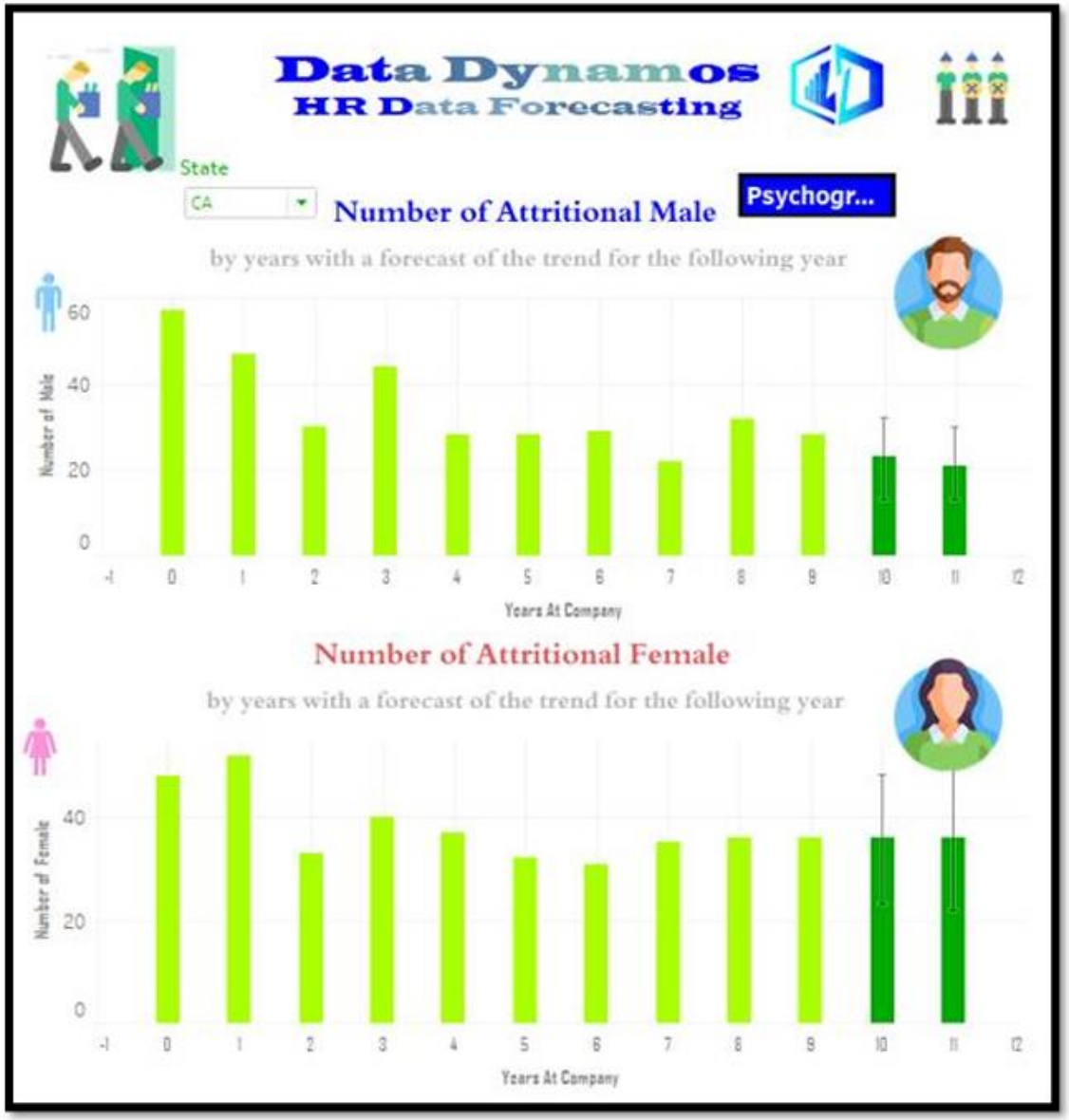
Power BI for Data Analysis



Data Visualization

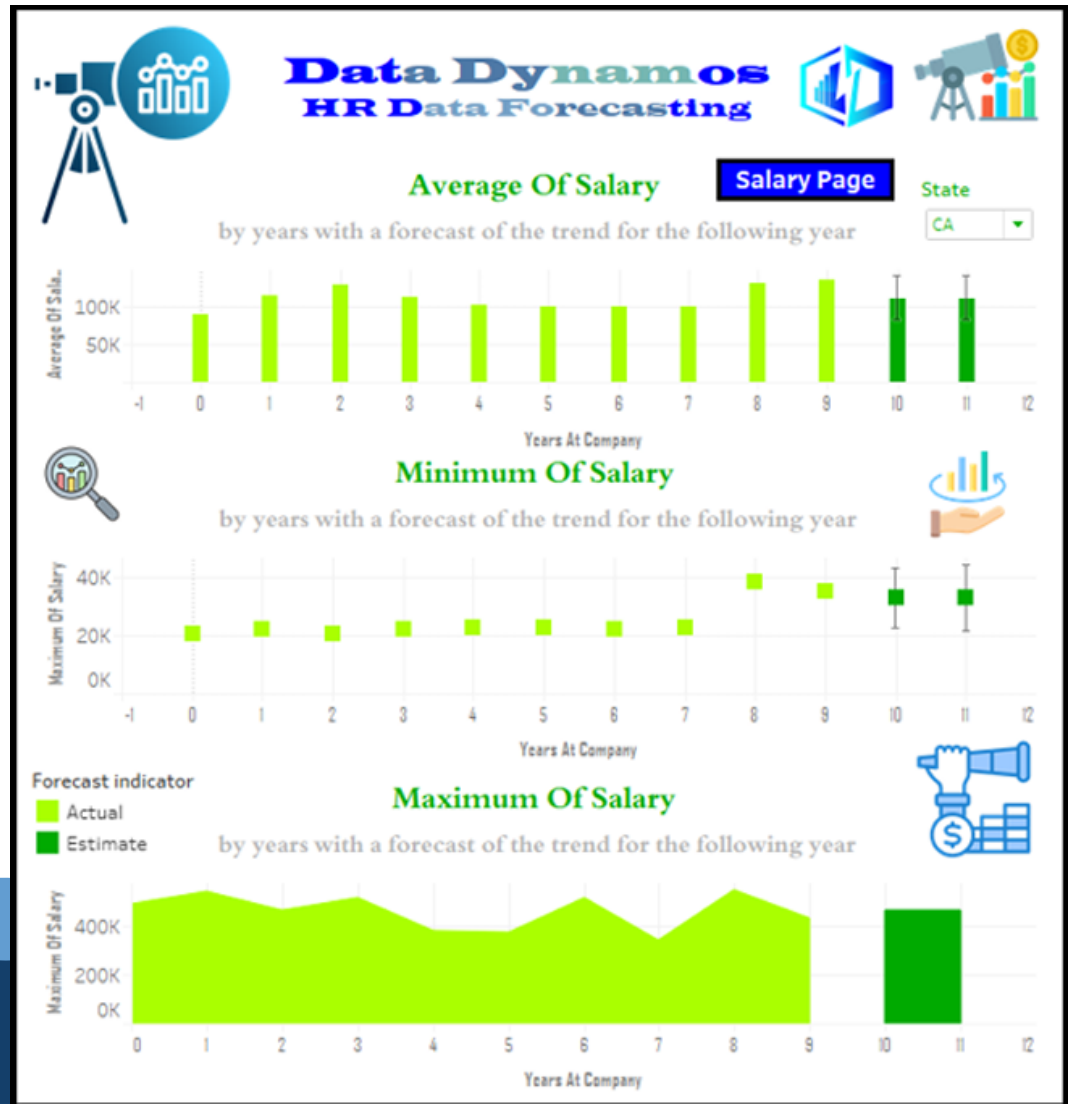
Tableau for Data Forecasting

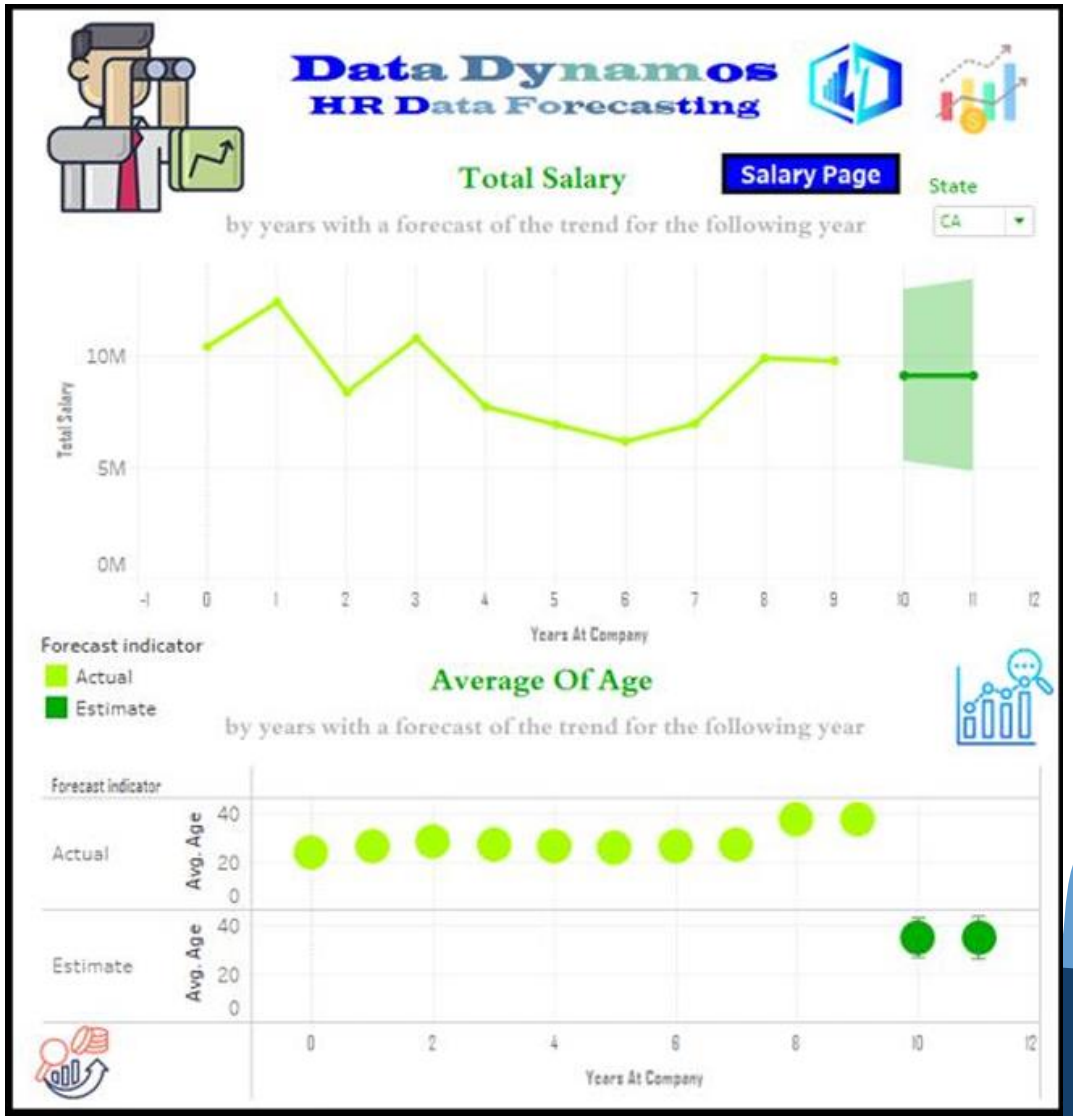




Data Visualization

Tableau for Data Forecasting



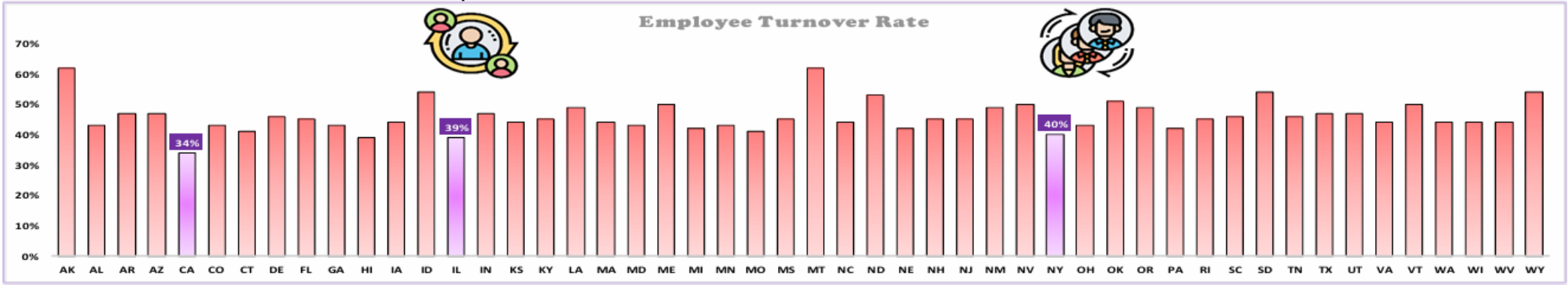


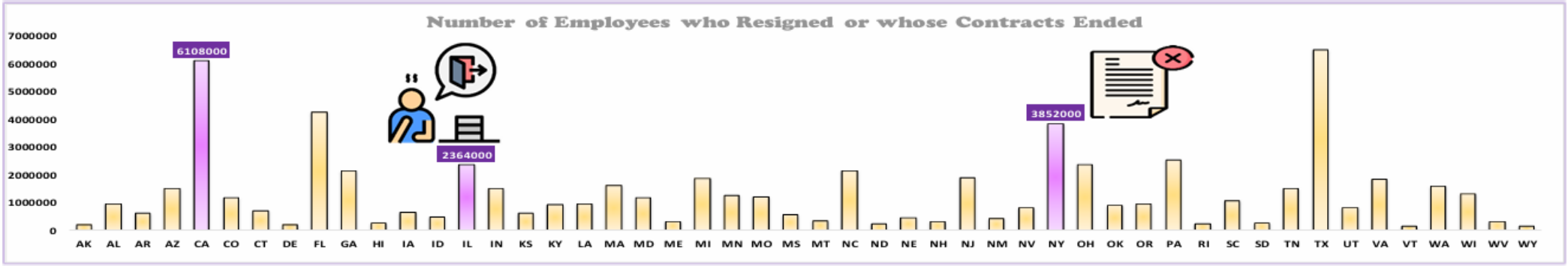
Data Visualization

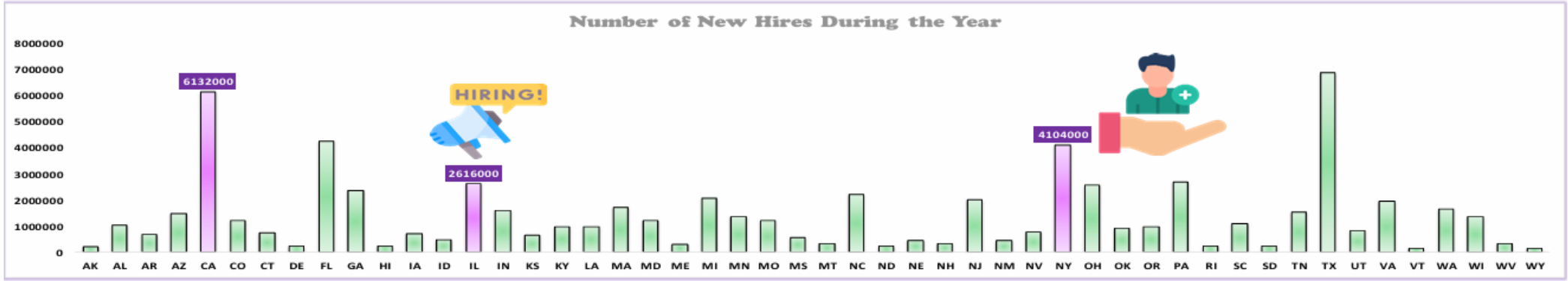
excel for marketing research

9

MARKET RESEARCH



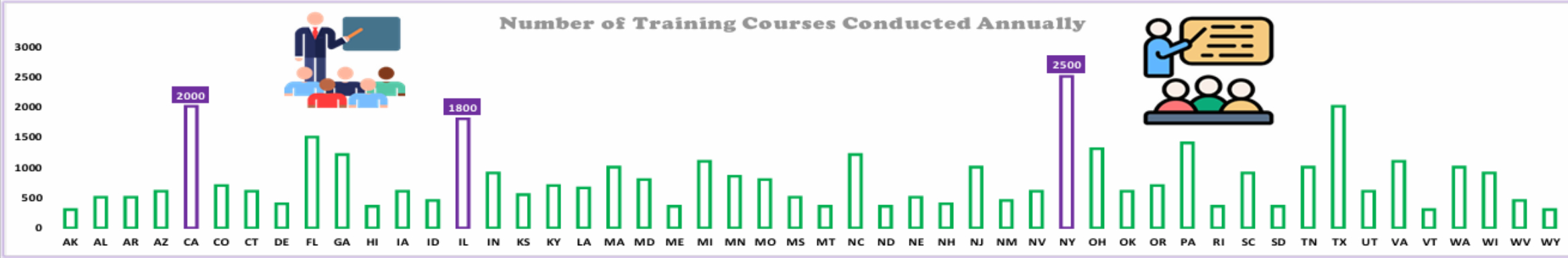




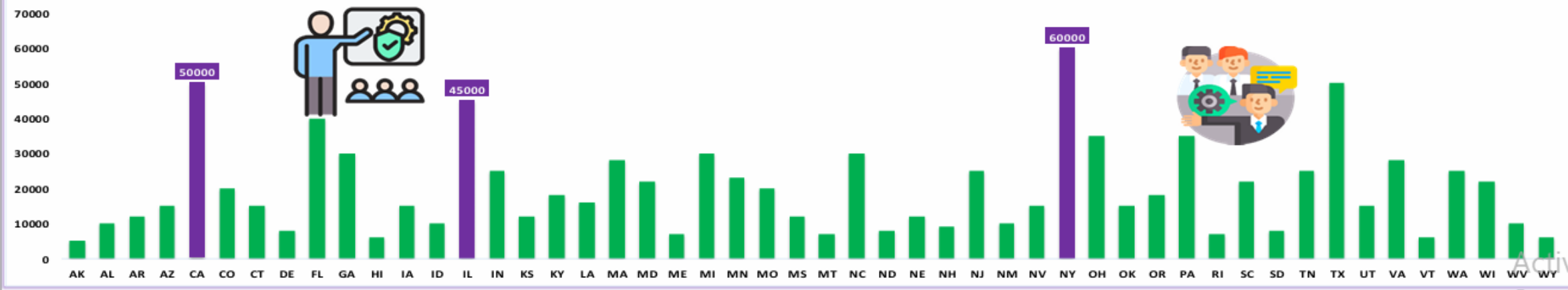
MARKET RESEARCH

Training Budget

Number of Training Courses Conducted Annually

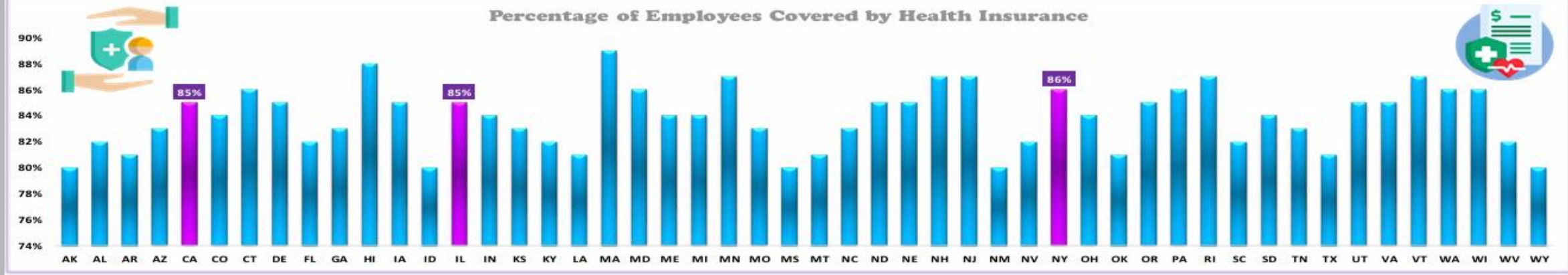


Number of Employees who Received Training

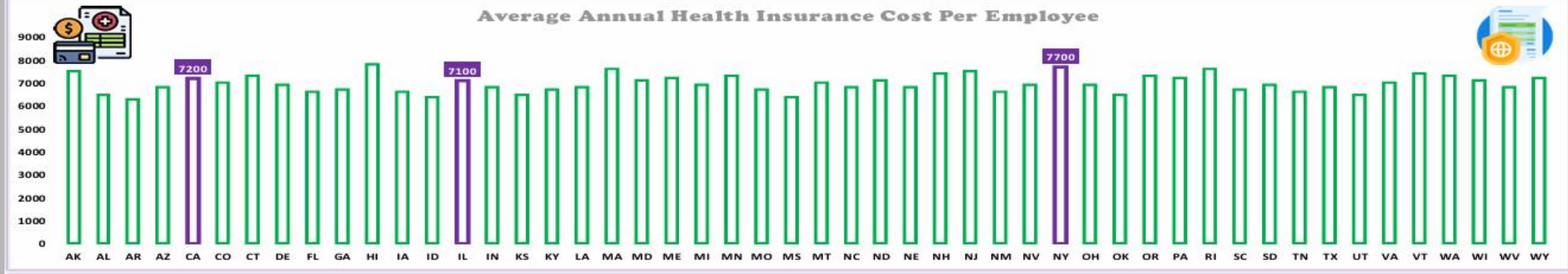


MARKET RESEARCH

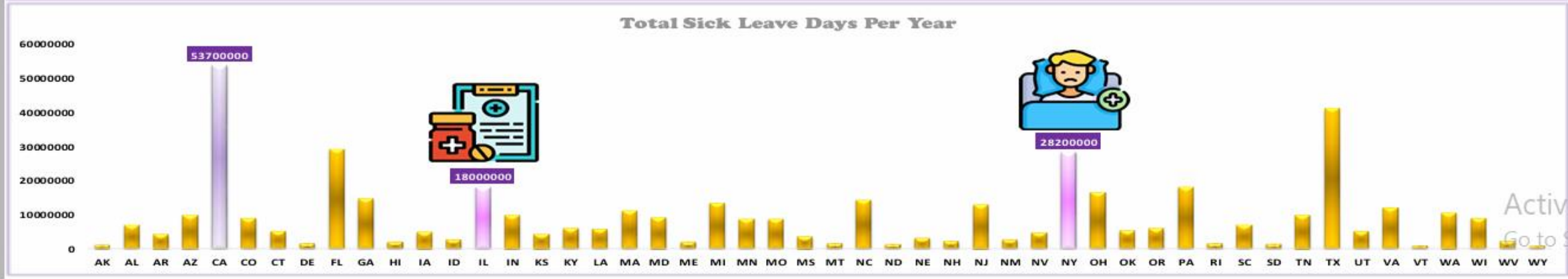
Percentage of Employees Covered by Health Insurance



Average Annual Health Insurance Cost Per Employee



Total Sick Leave Days Per Year



Data Mining

Data Mining Techniques

1 Clustering

- Grouping employees based on similar characteristics (e.g., job satisfaction, performance ratings).
- Example: Identifying high-risk groups for attrition.

2 Classification

- Predicting categorical outcomes (e.g., Will an employee leave? Yes/No).
- Example: Attrition prediction using Random Forest Classifier.

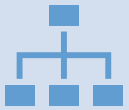
3 Association Rule Mining

- Discovering relationships between variables (e.g., Overtime → Attrition).
- Example: Employees working overtime are 3x more likely to leave.

4 Regression Analysis

- Predicting continuous outcomes (e.g., Salary, Tenure).
- Example: Salary prediction based on job role and experience.

Data Mining Insights



1. High-Risk Groups for Attrition:

Employees with low salaries and short tenure are at higher risk of leaving.

Employees working overtime are 3x more likely to leave.

2. Performance Trends:

High-performing employees tend to have higher salaries and longer tenure.

Employees with lower performance ratings (3/5) have the highest resignation rates.

3. Training Effectiveness:

Employees who participate in training programs show improved performance and higher promotion rates.

Training participation is positively correlated with job satisfaction.

Data-Driven Decision Making

- **Recommendations:**
  - Improve work-life balance and reduce overtime.
  - Enhance compensation and benefits.
  - Provide clear career paths and training opportunities.
  - Strengthen diversity and inclusion initiatives.

Data Storage

Platforms Used:	• GitHub and Google Drive.
Repository Structure:	• Organized by project stages (Data Collection, Data Cleaning, etc.).
Accessibility:	• Secure and easily accessible for team members.

Market Research Analysis

Labor Force

•Key states demonstrate strong workforce participation, indicating robust talent availability for strategic hiring initiatives.(like: CA, IL, NY)

Unemployment Rate

•The current tight labor market creates competitive hiring conditions, requiring careful consideration of timing, compensation, and training investments.

Average Annual Salaries

•Significant regional salary differences emerge, highlighting the importance of market-aligned compensation strategies for talent retention and productivity.

Employment Type

•The mix of employment arrangements provides valuable insights for optimizing benefits and workforce structure to balance organizational flexibility with operational stability.

Salary Structure

•Disparities in minimum and maximum pay reveal equity and skill demand.
•Important for compensation strategy and budget planning.

Performance Evaluation

- Supports promotion planning, training needs, and employee engagement.

Insurance & Benefits

•Comprehensive benefits and insurance coverage serve as a competitive advantage in both attracting top talent and improving employee retention.

Training & Development

•Significant organizational investment in continuous learning programs demonstrates commitment to workforce capability building and future readiness.

Turnover & Recruitment

•Employee retention patterns directly influence recruitment strategies, particularly during periods of business growth or workforce instability.

Absenteeism & Leave

•Indicates job satisfaction and organizational health.
•Helps optimize workforce planning.

Total Employees

•Reflects business expansion or downsizing.
•Aids in resource planning and performance analysis

PESTEL Analysis

P

E

S

Political

- Government policies on labor laws, wages, insurance, and leave directly impact hiring and compensation.
- Political stability and tax policies influence operational costs.
- Union influence and potential strikes may affect workforce stability.

Economic

- Significant variations in compensation levels highlight the need for equitable pay structures.
- External economic factors including inflation and currency fluctuations may pressure salary budgets, while overall economic conditions directly shape talent acquisition strategies.

Social

- Strong employee benefits and work-life balance provisions contribute to workforce satisfaction.
- Rising expectations for diversity, flexible work, and career development.
- Sick leave policies vary by location and are loosely regulated.

Technological

- Substantial investments in digital upskilling programs position the workforce for future readiness.
- Tools like Power BI and Tableau enhance decision-making.
- Automation and cybersecurity are key areas for improvement.

Environmental

- Remote work can reduce environmental impact from commuting.
- Opportunities exist for CSR and eco-friendly initiatives.
- New environmental regulations may require office policy updates.

Legal

- High health insurance coverage shows compliance.
- Adherence to labor laws, data privacy, and anti-discrimination regulations is essential.
- Regular contract reviews and legal audits are recommended.

SWOT Analysis

STRENGTHS

- Competitive salaries and long employee tenure.
- High job satisfaction and productivity.

WEAKNESSES

- High turnover and weak incentive programs.
- Salary disparities and lack of training opportunities..

OPPORTUNITIES

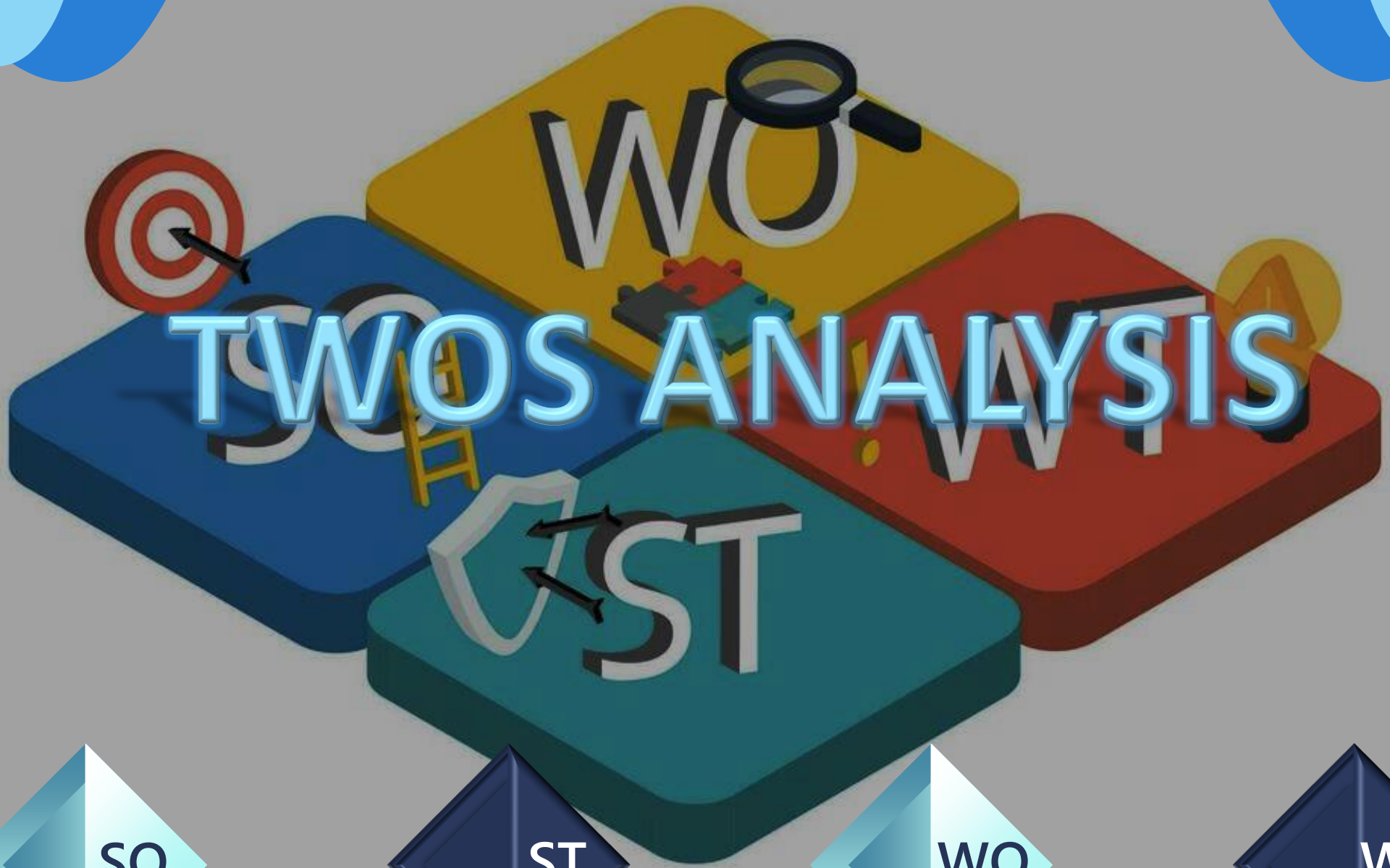
- Improve training programs and offer remote work.
- Leverage technology and expand into new markets.

THREATS

- Intense competition for talent and rising recruitment costs.
- Work pressures and new regulations


WO
SO
!WT
ST

TWOS ANALYSIS



SO

ST

WO

WT

TWOS ANALYSIS

SO

1. Develop training programs to enhance skills based on evaluations.
2. Leverage competitive salaries to attract and retain talent.
3. Increase travel and job exchange opportunities to improve employee efficiency.

ST

WO

WT

TWOS ANALYSIS

SO

1. Develop training programs to enhance skills based on evaluations.
2. Leverage competitive salaries to attract and retain talent.
3. Increase travel and job exchange opportunities to improve employee efficiency.

ST

1. Enhance job stability through long-term incentive policies.
2. Develop an attractive work environment to reduce external competition for skilled employees.
3. Improve promotion policies to ensure high-performing employees stay.

WO

WT

TWOS ANALYSIS

SO

1. Develop training programs to enhance skills based on evaluations.
2. Leverage competitive salaries to attract and retain talent.
3. Increase travel and job exchange opportunities to improve employee efficiency.

ST

1. Enhance job stability through long-term incentive policies.
2. Develop an attractive work environment to reduce external competition for skilled employees.
3. Improve promotion policies to ensure high-performing employees stay.

WO

1. Improve financial incentives and rewards to create a more competitive work environment.
2. Implement sustainable professional development plans to empower employees.
3. Enhance work-life balance through flexible policies.

WT

TWOS ANALYSIS

SO

1. Develop training programs to enhance skills based on evaluations.
2. Leverage competitive salaries to attract and retain talent.
3. Increase travel and job exchange opportunities to improve employee efficiency.

ST

1. Enhance job stability through long-term incentive policies.
2. Develop an attractive work environment to reduce external competition for skilled employees.
3. Improve promotion policies to ensure high-performing employees stay.

WO

1. Improve financial incentives and rewards to create a more competitive work environment.
2. Implement sustainable professional development plans to empower employees.
3. Enhance work-life balance through flexible policies.

WT

1. Reduce attrition rates through work environment improvements and stress reduction initiatives.
2. Review recruitment and retention policies to ensure workforce stability.
3. Introduce psychological support and career counseling programs to reduce job stress and increase productivity.

Strengths, Opportunities, Aspirations, and Resul


Strengths
Results
SOAR ANALYSIS
Opportunities
Aspirations

SOAR ANALYSIS

Strengths

- **High Workforce Satisfaction:** Strong employee approval of workplace relationships and organizational culture.
- **Sustainable Work-Life Integration:** Established policies supporting employee well-being and balance.
- **Talent Retention Edge:** Competitive compensation packages and lower turnover rates than industry peers.

Opportunities

SOAR ANALYSIS

- Enhance training programs and attract young talent.
- Use performance analytics and expand job benefits.

gths

SOAR ANALYSIS

Aspirations

- Become a regional leader in flexible work innovation
- Prioritize internal talent growth and decision-making

Strengths

Results

- Improved employee retention and skill development
- Streamlined HR processes through tech adoption

SOAR ANALYSIS

Strengths

VRIO Analysis

1) Value:

High-performing employees, strong training programs, and job satisfaction contribute to valuable human capital.

2) Rarity:

Specialized expertise, multilingual talent, and technical proficiency create differentiation.

3) Imitability:

- Internal training programs and learning culture are hard to replicate.
- Significant investment in employee development increases uniqueness.

4) Organization:

Stable teams, effective HR strategies, and low attrition ensure alignment with business goals.

Conclusion

•Summary of Insights:

- The company has valuable and rare human resources with strong training programs.
- High turnover and work-life balance issues need addressing.
- Predictive models and strategic recommendations can improve retention and performance.

•Next Steps:

- Implement recommended strategies to improve retention and performance.
- Continuously monitor workforce trends and adjust policies as needed.

Any Question?

THANKS!