Nexasoft

رواد مصر الرقمية

HR Analysis project

DEPI Round 2 (2024, 2025)

Made By

Ahmed Wael
Ali Nader
Hazem Amr
Mahmoud Alaa
Sameh Shahin
Omar abdelhamid

رواد مصر الرقمية

Agenda

Expressing our Work through this presentation

01

Introduction

02

Objective

03

Methodology

04

Insights

رواد مصر الرقمية

01

Introduction

02

Objective

03

Methodology

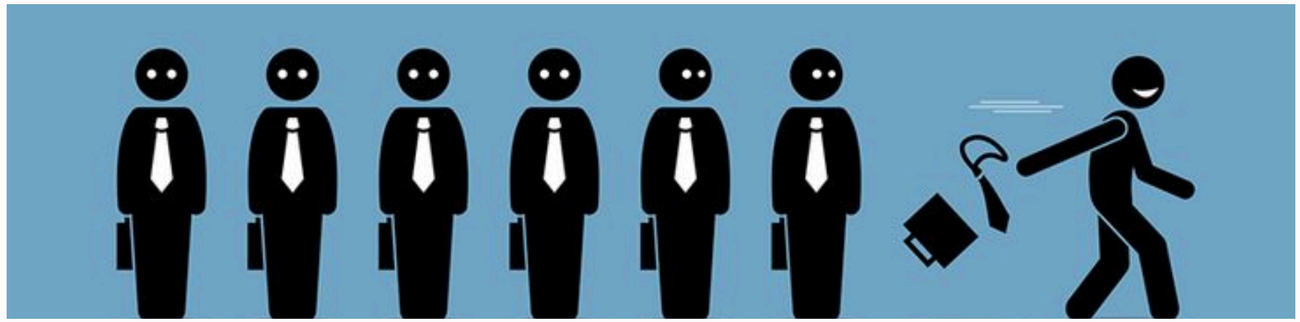
04

Insights

رواد مصر الرقمية

Introduction

Our dataset represents a software company and provides a detailed overview of its workforce. This analysis aims to explore patterns and trends related to employee departures, helping to understand the underlying causes and impacts.



رواد مصر الرقمية

01

Introduction

02

Objective

03

Methodology

04

Insights

Objective

Explore Key Factors

Determine the key factors that contribute to employee attrition within the company

Analyze Variables

Conduct thorough analysis of various variables including employee satisfaction, years at company, education level, job role, department, and salary

Gain Insights

Gain valuable insights into the reasons behind employee departures

رواد مصر الرقمية

01

Introduction

02

Objective

03

Methodology

04

Insights

Methodology

dataset

- About dataset
- Dataset description

Data preprocessing

- Data Preparation
- Data Cleaning & Preprocessing
- Data modeling

Data Analysis

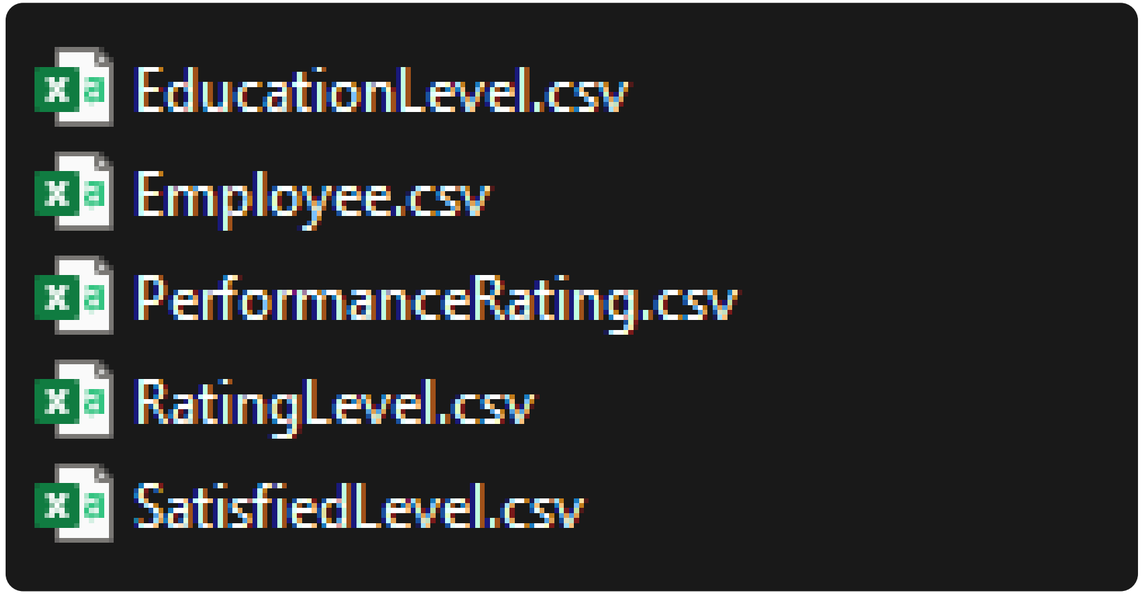
- EDA
- Data visualization
- Key Insights

رواد مصر الرقمية

DATASET

About the dataset

The dataset is provided by DEPI in 5 separated sheets in csv format named by



رواد مصر الرقمية

DATASET DESCRIPTION

DATASET

Employee

Total Records: 1,470 employees

Total Columns: 23

EmployeeID	Name	Gender	Age	BusinessTravel	DistanceFromHome	State	Ethnicity	EducationLevelID	EducationField
001A-8F88	Christy Jumel	Male	22	Some Travel	40	California	White	4	Information Systems
005C-E0FB	Fin O'Halleghane	Non-Binary	24	Frequent Traveller	17	California	White	4	Marketing
00A3-2445	Wyatt Ziehm	Male	30	Some Travel	6	California	Black or African American	2	Computer Science
00B0-F199	Trueman Jirasek	Male	23	Some Travel	35	California	White	1	Marketing
00D4-DD53	Joyce Goor	Female	30	Frequent Traveller	44	California	Black or African American	1	Computer Science
00E4-3D60	Sherilyn Girke	Female	30	Frequent Traveller	37	California	White	1	Business Studies
0145-DBFC	Keelia Studde	Female	34	No Travel	8	California	Asian or Asian American	4	Information Systems
0172-B5D2	Ketti Keighley	Female	27	Some Travel	28	California	White	4	Computer Science
017D-0B97	Maison Werndley	Male	20	Some Travel	39	California	White	2	Technical Degree
018B-DF1E	Otha Sopper	Female	19	Some Travel	1	California	White	1	Economics
0210-E0D8	Yankee Charteris	Male	36	Some Travel	7	New York	White	3	Technical Degree
022A-0219	Francine Fernez	Female	32	Frequent Traveller	14	Illinois	American Indian or Alaska Native	3	Business Studies
0240-4D29	Kaylil Blenkiron	Non-Binary	35	Some Travel	31	California	Asian or Asian American	4	Computer Science
02DA-7A72	Tadeas Lackney	Male	29	Frequent Traveller	42	California	White	3	Technical Degree
0317-B7BC	Ronni Beechcraft	Female	23	Some Travel	7	Illinois	White	4	Marketing
0322-D46B	Nikolas Leslie	Male	32	Some Travel	39	California	Black or African American	4	Marketing
0375-2F04	Lani Raddenbury	Female	19	No Travel	38	Illinois	White	4	Technical Degree
03C5-51AD	Maxie Banker	Female	25	Some Travel	1	Illinois	White	2	Computer Science
03D3-AA88	Mycah Brolechan	Male	23	Some Travel	8	New York	White	4	Other
03D5-622D	Aurora Whate	Female	21	No Travel	43	California	White	1	Marketing
03D9-FF08	Lynnette Portinari	Non-Binary	22	Some Travel	30	California	White	4	Other
03EF-5904	Shepperd Brittain	Male	31	Some Travel	45	Illinois	American Indian or Alaska Native	3	Other
040C-EDED	Lonnie Braxton	Male	22	Frequent Traveller	45	California	White	2	Computer Science
041A-31B0	Gerri Fullerlove	Female	34	Some Travel	41	New York	Black or African American	3	Computer Science
04D3-B8D3	Ann Sivorn	Female	24	Frequent Traveller	23	New York	White	4	Computer Science

Performance Rating Dataset

Total Records: 6,709 performance reviews

Columns (11): Captures employee satisfaction and rating evaluation.

PerformanceID	EmployeeID	ReviewDate	EnvironmentSatisfaction	JobSatisfaction	RelationshipSatisfaction	TrainingOpportunitiesWithinYear	TrainingOpportunitiesTaken	WorkLifeBalance	SelfRating
PR01	79F7-78EC	01/02/2013	5	4	5	1	0	4	4
PR02	B61E-0F26	01/03/2013	5	4	4	1	3	4	4
PR03	F5E3-48BB	01/03/2013	3	4	5	3	2	3	5
PR04	0678-748A	01/04/2013	5	3	2	2	0	2	3
PR05	541F-3E19	01/04/2013	5	2	3	1	0	4	4
PR06	F93E-BDEF	01/04/2013	3	3	2	2	0	4	4
PR07	9E7A-1F70	01/08/2013	3	4	5	2	1	5	4
PR08	05ED-92F1	01/10/2013	4	5	4	1	1	3	3
PR09	F72D-261D	01/10/2013	4	5	2	1	1	4	5
PR10	774E-685D	01/11/2013	5	4	3	2	3	4	5
PR100	B013-7D0C	04/10/2013	4	3	3	2	0	4	3
PR1000	528C-3E0D	3/16/2016	4	4	2	2	2	4	5
PR1001	D077-169C	3/17/2016	3	5	3	2	2	3	5
PR1002	9727-BC84	3/18/2016	4	3	3	2	2	2	4
PR1003	DA8E-9496	3/18/2016	3	5	4	1	0	5	5
PR1004	DEC5-9319	3/18/2016	3	4	3	2	3	2	4
PR1005	88B8-EB84	3/19/2016	3	4	2	3	1	4	5
PR1006	9C57-828C	3/19/2016	5	4	2	1	1	2	3
PR1007	E1B4-9AA1	3/22/2016	5	4	3	3	2	3	4
PR1008	3CD6-5587	3/23/2016	5	4	2	2	0	4	4
PR1009	BAFA-86DF	3/23/2016	3	3	4	2	1	2	3
PR101	152E-8DB1	04/12/2013	5	2	5	1	0	5	5
PR1010	C6D7-A568	3/23/2016	4	2	3	3	0	2	4
PR1011	5160-53BD	3/24/2016	4	2	5	1	2	2	5
PR1012	81FF-8D6C	3/24/2016	3	4	3	2	0	4	5

Education Level Dataset

EducationLevelID	EducationLevel
1	No Formal Qualifications
2	High School
3	Bachelors
4	Masters
5	Doctorate

Rating Level Dataset

RatingID	RatingLevel
1	Unacceptable
2	Needs Improvement
3	Meets Expectation
4	Exceeds Expectation
5	Above and Beyond

Satisfaction Level Dataset

SatisfactionID	SatisfactionLevel
1	Very Dissatisfied
2	Dissatisfied
3	Neutral
4	Satisfied
5	Very Satisfied

Any Questions?

رواد مصر الرقمية

DATA PREPROCESSING

DATASET

DATASET DESCRIPTION

Data Preparation

رواد مصر الرقمية

EmployeeID	Name	Gender	Age	BusinessTravel	DistanceFromHome	State	Ethnicity
Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 10 Error Empty
1000 distinct, 1000 unique	1000 distinct, 1000 unique	4 distinct, 0 unique	34 distinct, 1 unique	3 distinct, 0 unique	45 distinct, 0 unique	3 distinct, 0 unique	7 distinct, 0 unique
001A-8F88	Christy Jumel	Male	22	Some Travel	40	California	White
005C-E0FB	Fin O'Halleghane	Non-Binary	24	Frequent Traveller	17	California	White
00A3-2445	Wyatt Ziehm	Male	30	Some Travel	6	California	Black or African American
00B0-F199	Trueman Jirasek	Male	23	Some Travel	35	California	White
00D4-DD53	Joyce Goor	Female	30	Frequent Traveller	44	California	Black or African American
00E4-3D60	Sherilyn Girke	Female	30	Frequent Traveller	37	California	White
0145-DBFC	Keelia Studde	Female	34	No Travel	8	California	Asian or Asian American
0172-B5D2	Ketti Keighley	Female	27	Some Travel	28	California	White
017D-0B97	Maison Werndley	Male	20	Some Travel	39	California	White
018B-DF1E	Otha Sopper	Female	19	Some Travel	1	California	White
0210-E0D8	Yankee Charteris	Male	36	Some Travel	7	New York	White
022A-0219	Francine Fernez	Female	32	Frequent Traveller	14	Illinois	American Indian or Alaska Native
0240-4D29	Kaylil Blenkiron	Non-Binary	35	Some Travel	31	California	Asian or Asian American
02DA-7A72	Tadeas Lackney	Male	29	Frequent Traveller	42	California	White
0317-B7BC	Ronni Beechcraft	Female	23	Some Travel	7	Illinois	White
0322-D46B	Nikolas Leslie	Male	32	Some Travel	39	California	Black or African American
0375-2F04	Lani Raddenbury	Female	19	No Travel	38	Illinois	White

Data cleaning and Processing

- Handling Missing data
- Detecting duplicated values
- Data Type Correction

Data Processing

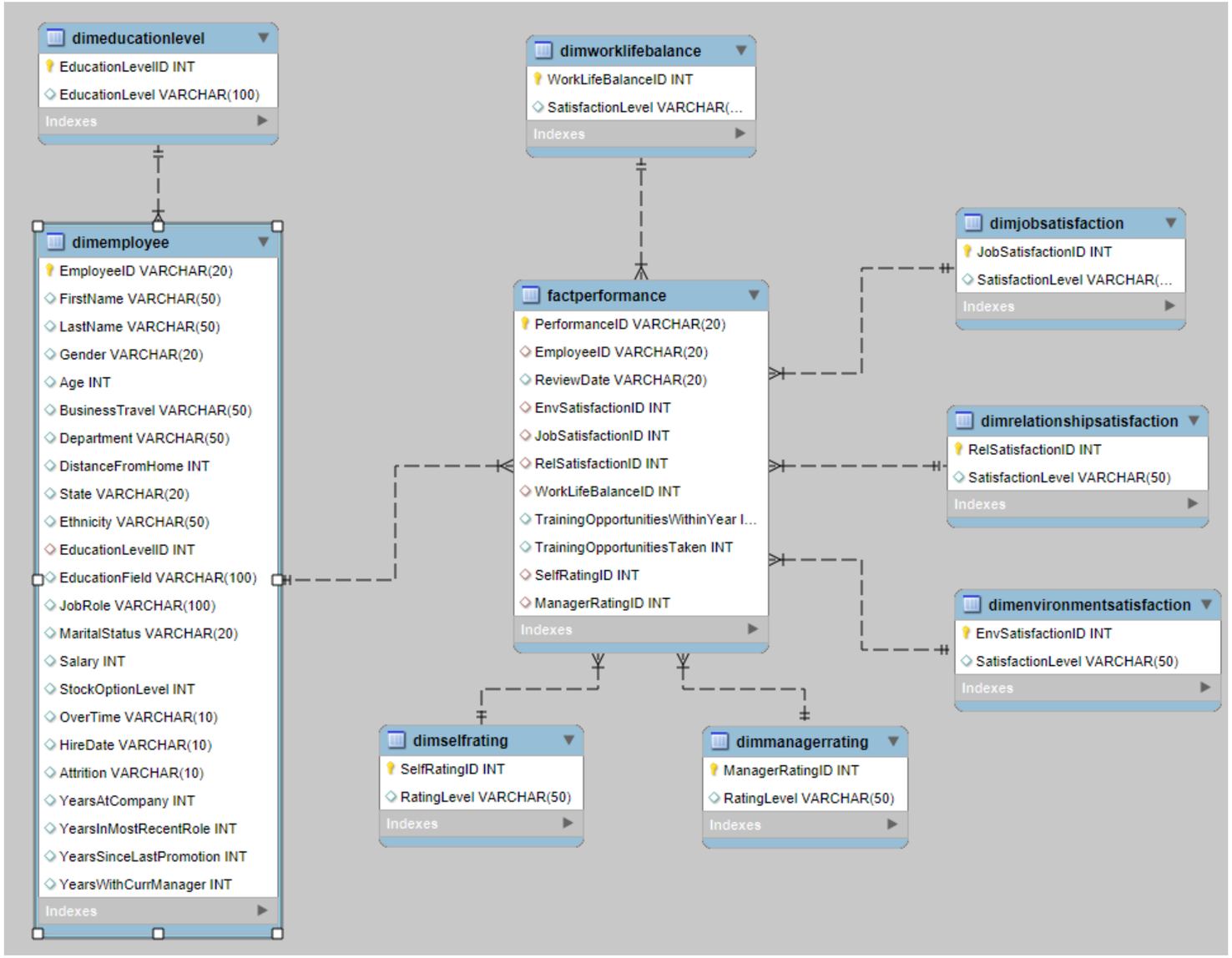
- **Data Mismatch**

EducationField	Department	JobRole	MaritalStatu	Salary
Marketing	Technology	Sales Executive	Single	140158

- **Identified Non-evaluted Employees**

Any Questions?

Data Modelling



Data Modelling

رواد مصر الرقمية

1-Preparing data

- Using MySQL For creating database
- creating tables for each csv file

2-Data Engineering

- Developing 4 new Dimension tables to avoid data Redunducy

3-Modeling Outcome

- Fact table: Performance Table
- 7 Dimensions connected to the fact

رواد مصر الرقمية

DATA ANALYSIS

DATASET

DATASET DESCRIPTION

DATA PREPROCESSING

Exploratory data analysis

Questions to Ask in a Presentation:

1. **What are the key factors influencing employee attrition?**
2. **Which group (age, salary, stock option level) has the highest attrition?**
3. **How does department type impact attrition?**
4. **Do stock options help retain employees?**

You can introduce these in your EDA or Analysis section to show deeper insight:

Which job roles experience the highest and lowest attrition rates?

How does job satisfaction correlate with attrition?

How do performance ratings (from managers vs. self) relate to attrition?

Which departments show the highest satisfaction or dissatisfaction levels?

How does work-life balance differ between those who left and stayed?

How many employees received training, and did it affect retention?

What is the average tenure for each department or job role?

Is there a difference in attrition between genders?

How does relationship satisfaction with managers affect retention?

Do employees with higher education stay longer or leave sooner?

Exploratory data analysis(4 W’s)

1

What happened? (Descriptive Analytics)
EDA Question: What is the overall attrition rate in the company?

2

Why did it happen? (Diagnostic Analytics)
EDA Question: What factors contribute most to employee attrition?

3

What will happen? (Predictive Analytics)
EDA Question: If trends continue, which departments will have the highest attrition next year?

4

What should we do? (Prescriptive Analytics)
EDA Question: How can HR reduce employee attrition?

Exploratory data analysis

رواد مصر الرقمية

1

How the problem happened?

EDA Question: What are the main reasons employees leave the company?

2

How long it may last?

EDA Question: If no changes are made, how long will high attrition persist?

3

How to solve the problem?

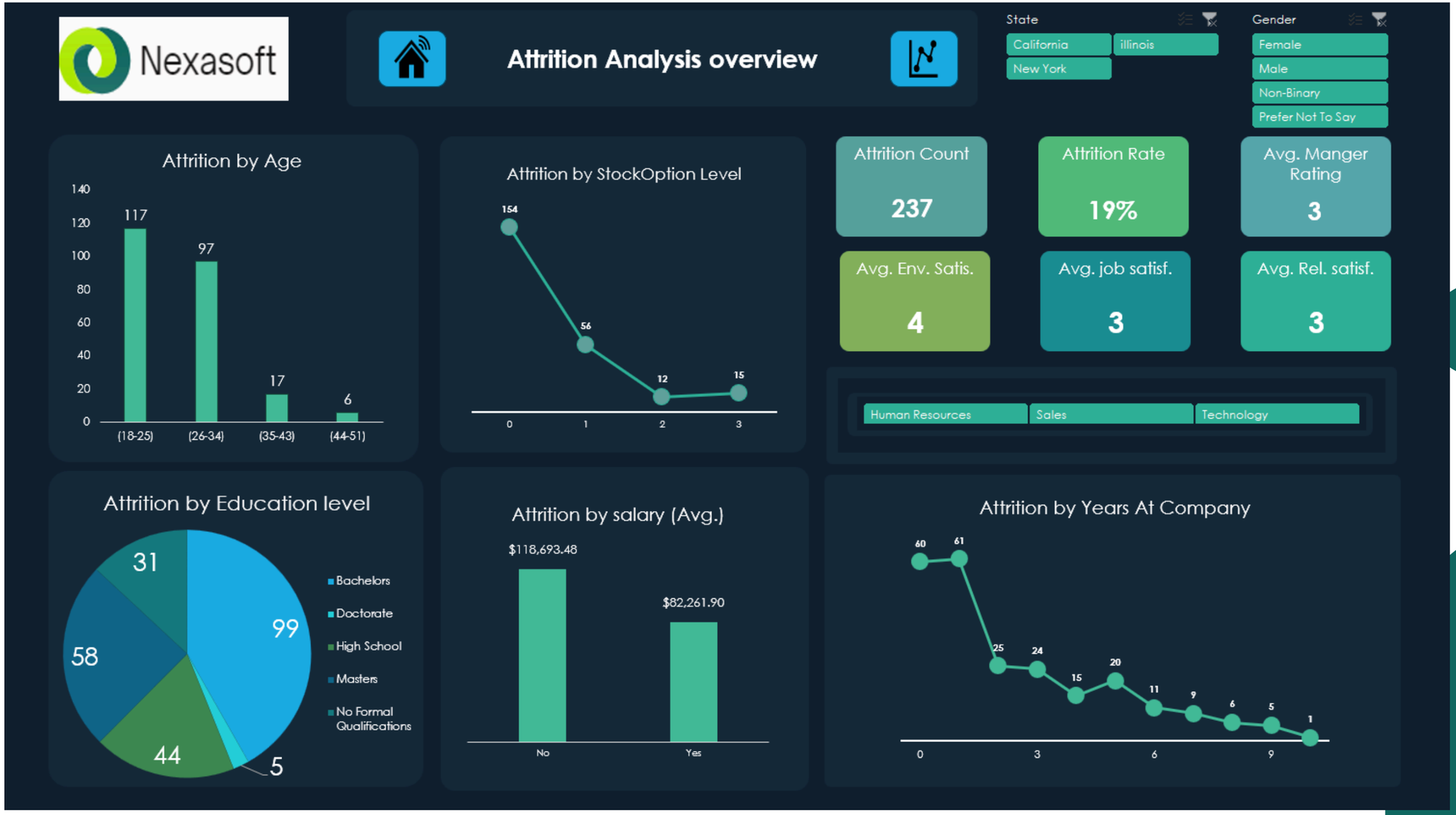
EDA Question: What policies or incentives can reduce employee turnover?

Data visualization

Employee Overview

Data visualization

Attrition Analysis



رواد مصر الرقمية

01

Introduction

02

Objective

03

Methodology

04

Insights

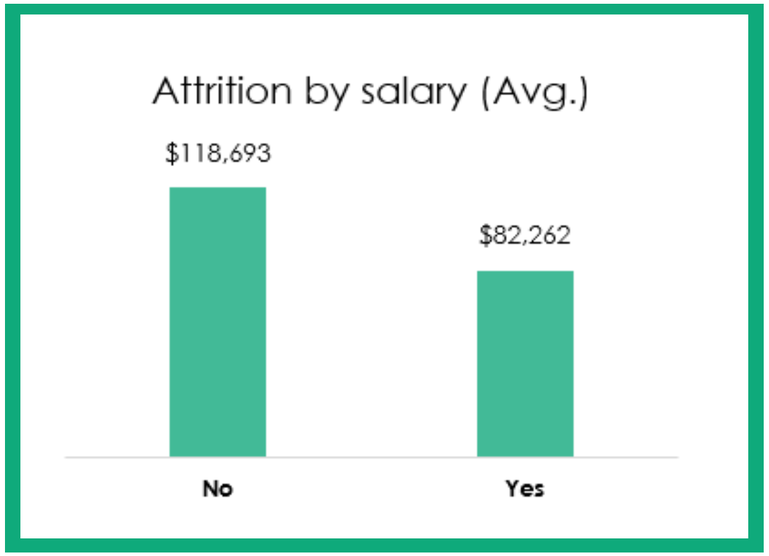
Data Visualization and key insights

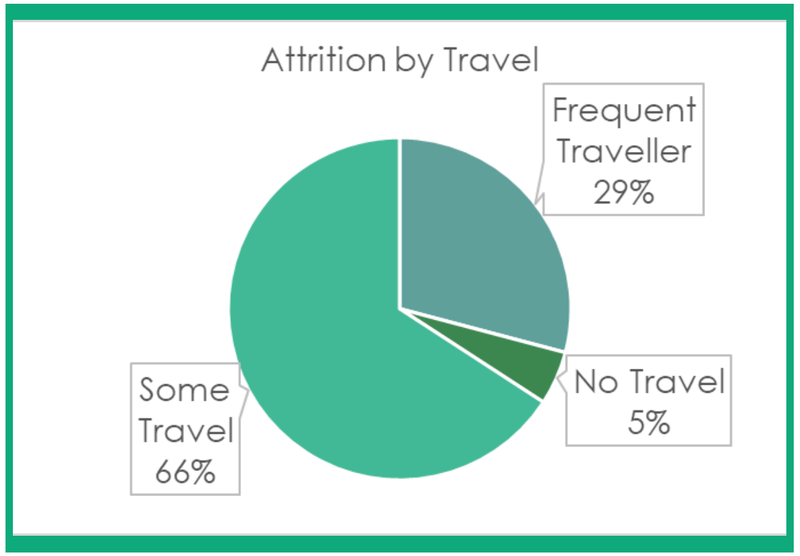
What is the overall attrition rate in the company?

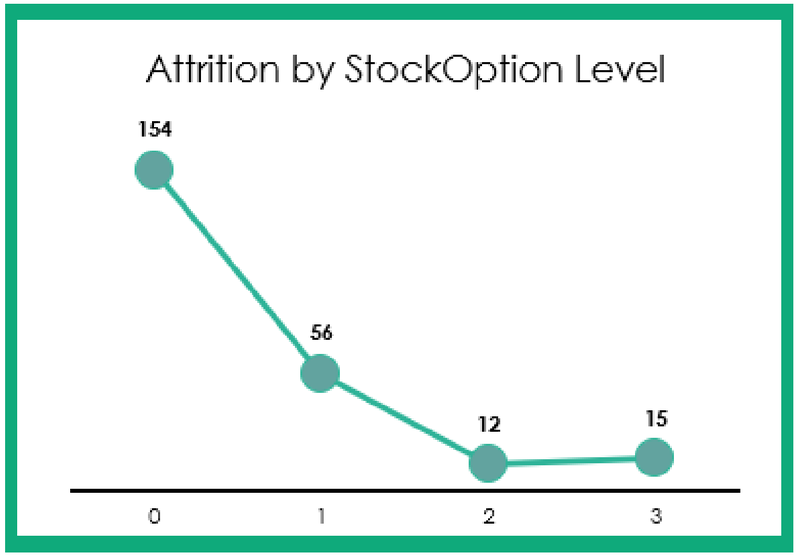
Company Reputation is at high risk

Data Visualization and key insights

What factors contribute most to employee attrition?







30% Salary Gap may be the highest factor affecting the attrition rate

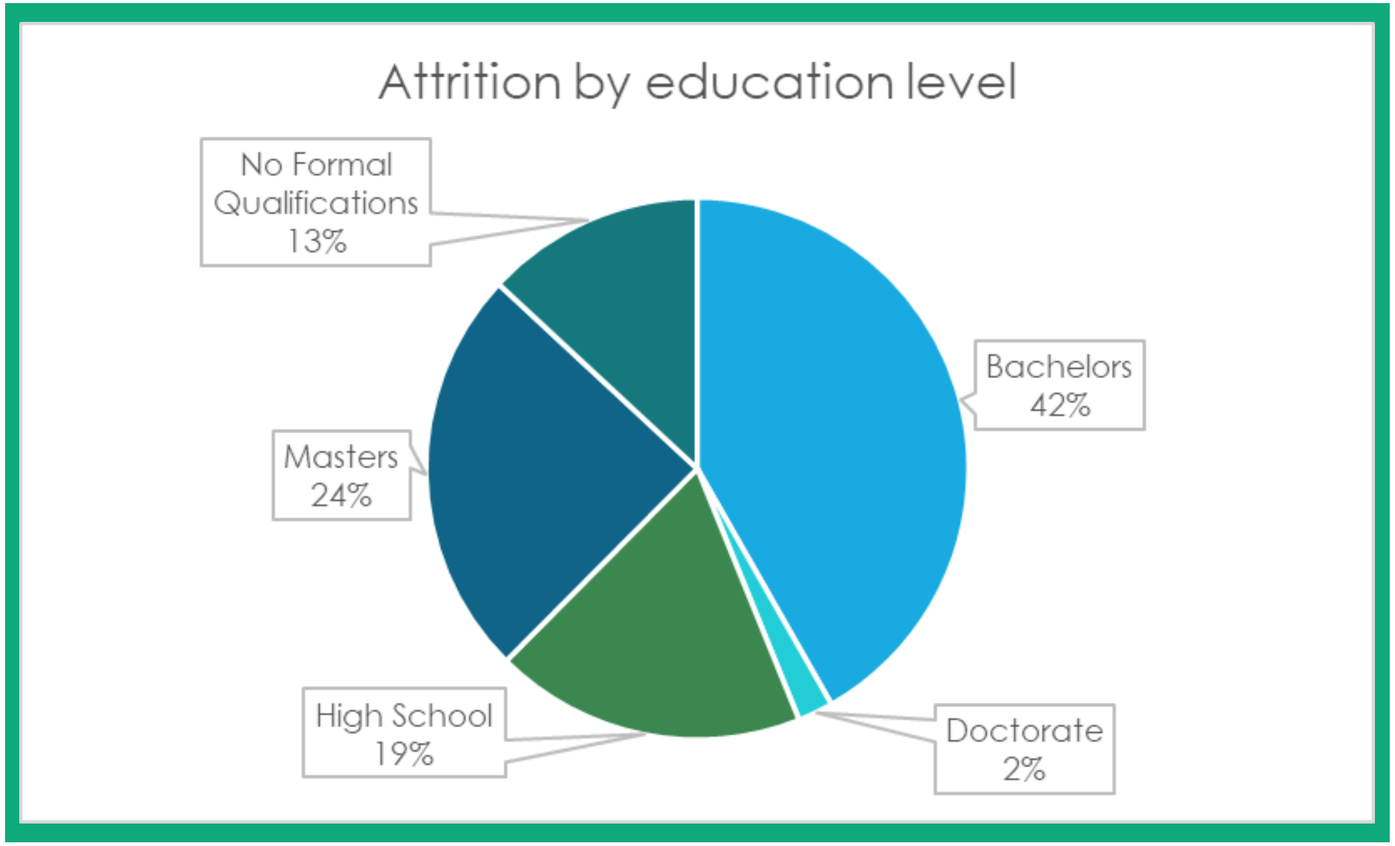
As Some travel is the most, this may not be a Strong factor affecting the attrition

Lack of Stock option level may cause employees to leave the company

Data Visualization and key insights

What factors contribute most to employee attrition?

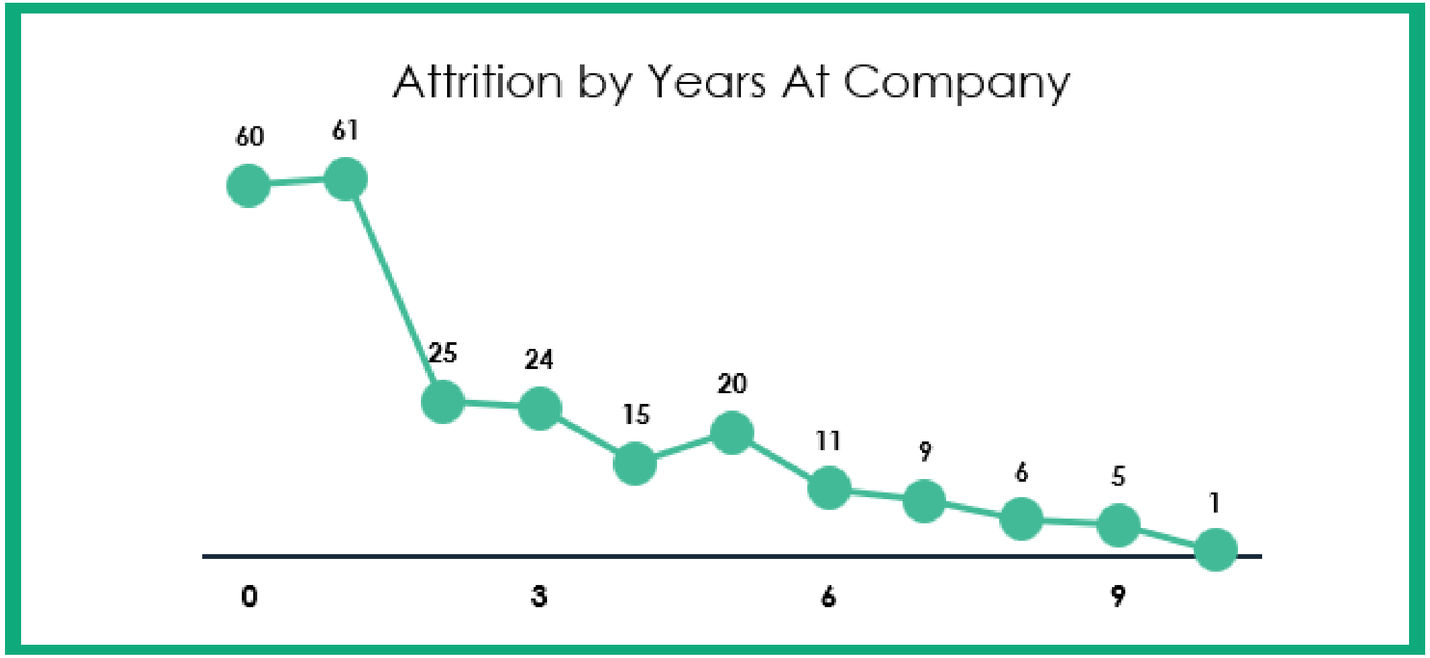
Education level show seems to be in equal with high ratio at bachelor, which may be also a weak factor for attrition



Data Visualization and key insights

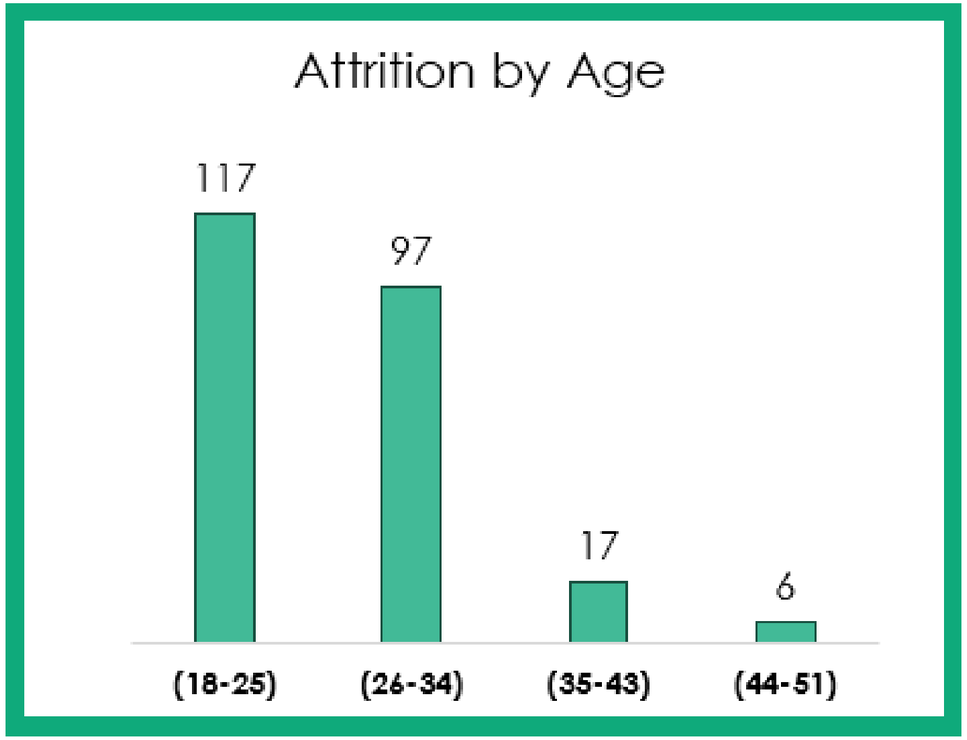
What factors contribute most to employee attrition?

Years at company could be a crucial factor for de-influencing attrition rate



Data Visualization and key insights

High-Risk Groups: Employees aged 18-34, single employees, and those with no stock options.



Conclusion and Key Insights

- Establish New salary policies to maintain balance between employees
- Improve onboarding & career development for new hires
- Offer stock options to more employees to improve long-term retention.
- Targeted retention programs for younger employees and high-turnover roles (sales, recruitment).


رواد مصر الرقمية
Q&A

THANK YOU