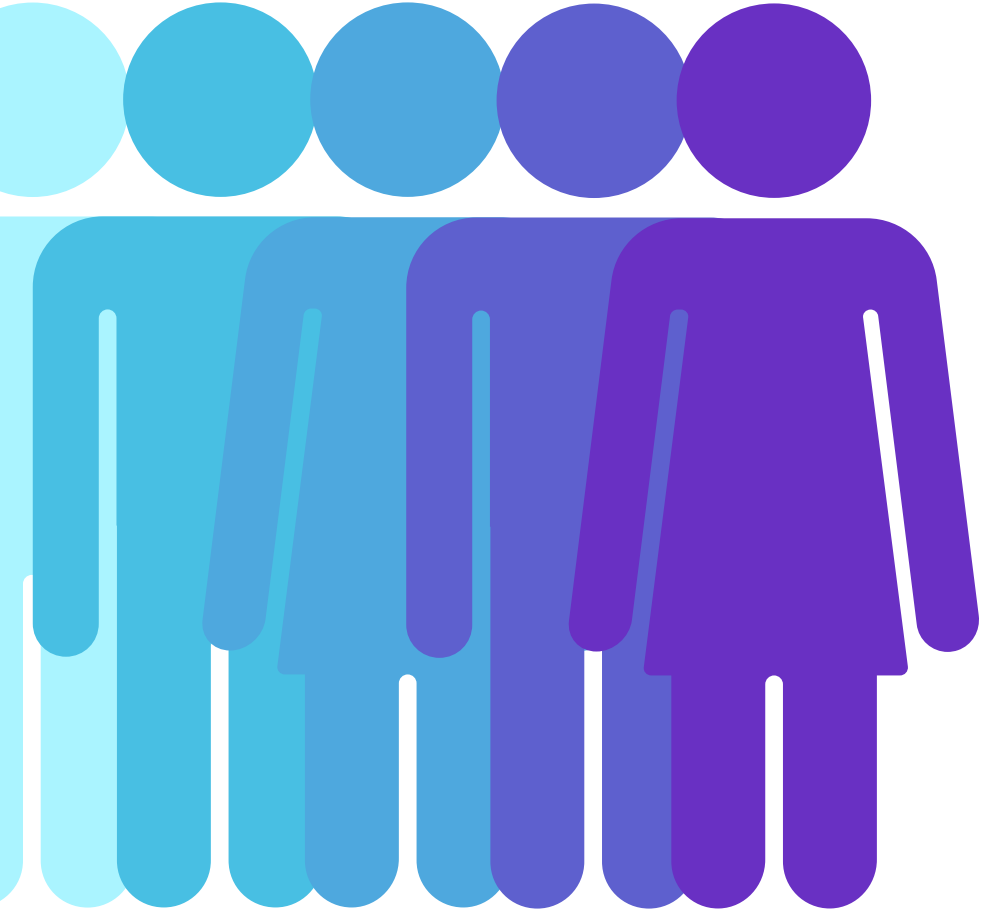


Mini Personal Project

People Analytics

Azriel Akbar Al Fajri

A Look into Azriel's Profile

A resilient and content-pursuer person, currently pursuing an undergraduate degree in Economics at the University of Brawijaya. He has various experiences in data analytics initiatives, such as a Data Analyst Intern at KNEKS, Data Analyst Generasi Gigih Apprenticeship at GoTo Impact Foundation, Incoming Capstonian with Jobhun (Generasi Gigih Industry Partner's Capstone Project), etc.

He agrees that both *data* and *new things* are akin to oil—valuable resources that can infuse life with excitement. But, without the ability to refine them, their value remains untapped.

His ultimate goal is to make the world a much better place to live by improving the accuracy of human and business decision-making that's being reflected in Belajar Manusia

Hard-skill :

MySQL Looker +ableau STATA Office

People Analytics

Definition

People Analytics is a **process of collecting, analyzing, and interpreting human** data in the field of human resources that **can unlock the true potential of your workforce.**

Purpose of Analysis

This project is designed to extract valuable insights about performance and attrition of employees and provide recommendations to address any issues that may arise.

About Dataset

The dataset can be obtained on Kaggle through this Employee Dataset link. The dataset is a fabricated dataset designed for the purpose of experimenting with different data analysis and machine learning approaches in the field of human resources and employee management. While this dataset mimics the format and attributes of genuine employee data, it's important to note that all the information it contains is entirely fictional and has been generated for demonstration purposes only.

Data Preprocessing

Preview of Data

Head of the data

	EmpID	FirstName	LastName	StartDate	ExitDate	Title	Supervisor	ADEmail	BusinessUnit	EmployeeStatus	...	Division	DOB	State	JobFunctionDescription	GenderCode	LocationCode	RaceDesc	MaritalDesc	Performance Score	Current Employee Rating
0	3427	Uriah	Bridges	20-Sep-19	NaN	Production Technician I	Peter Oneill	uriah.bridges@bilearner.com	CCDR	Active	...	Finance & Accounting	07-10-1969	MA	Accounting	Female	34904	White	Widowed	Fully Meets	4
1	3428	Paula	Small	11-Feb-23	NaN	Production Technician I	Renee Mccormick	paula.small@bilearner.com	EW	Active	...	Aerial	30-08-1965	MA	Labor	Male	6593	Hispanic	Widowed	Fully Meets	3
2	3429	Edward	Buck	10-Dec-18	NaN	Area Sales Manager	Crystal Walker	edward.buck@bilearner.com	PL	Active	...	General - Sga	06-10-1991	MA	Assistant	Male	2330	Hispanic	Widowed	Fully Meets	4
3	3430	Michael	Riordan	21-Jun-21	NaN	Area Sales Manager	Rebekah Wright	michael.riordan@bilearner.com	CCDR	Active	...	Finance & Accounting	04-04-1998	ND	Clerk	Male	58782	Other	Single	Fully Meets	2
4	3431	Jasmine	Onque	29-Jun-19	NaN	Area Sales Manager	Jason Kim	jasmine.onque@bilearner.com	TNS	Active	...	General - Con	29-08-1969	FL	Laborer	Female	33174	Other	Married	Fully Meets	3

5 rows × 26 columns

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3000 entries, 0 to 2999
Data columns (total 26 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   EmpID                       3000 non-null   int64
 1   FirstName                   3000 non-null   object
 2   LastName                    3000 non-null   object
 3   StartDate                   3000 non-null   object
 4   ExitDate                    1533 non-null   object
 5   Title                       3000 non-null   object
 6   Supervisor                  3000 non-null   object
 7   ADEmail                     3000 non-null   object
 8   BusinessUnit                3000 non-null   object
 9   EmployeeStatus              3000 non-null   object
 10  EmployeeType                3000 non-null   object
 11  PayZone                     3000 non-null   object
 12  EmployeeClassificationType  3000 non-null   object
 13  TerminationType             3000 non-null   object
 14  TerminationDescription      1533 non-null   object
 15  DepartmentType              3000 non-null   object
 16  Division                    3000 non-null   object
 17  DOB                         3000 non-null   object
 18  State                       3000 non-null   object
 19  JobFunctionDescription      3000 non-null   object
 20  GenderCode                  3000 non-null   object
 21  LocationCode                3000 non-null   int64
 22  RaceDesc                    3000 non-null   object
 23  MaritalDesc                 3000 non-null   object
 24  Performance Score           3000 non-null   object
 25  Current Employee Rating     3000 non-null   int64
dtypes: int64(3), object(23)
memory usage: 609.5+ KB
```

Data information:

Column: 26 columns
Row: 3000 rows

There are 2 columns containing null values, but still make sense

Unique values in each column

```
EmpID                        3000
FirstName                    1597
LastName                     1137
StartDate                    1496
ExitDate                      820
Title                          32
Supervisor                   2952
ADEmail                      2998
BusinessUnit                   10
EmployeeStatus                  5
EmployeeType                    3
PayZone                         3
EmployeeClassificationType      3
TerminationType                 5
TerminationDescription       1533
DepartmentType                  6
Division                       25
DOB                          2800
State                          28
JobFunctionDescription         83
GenderCode                      2
LocationCode                 2821
RaceDesc                        5
MaritalDesc                     4
Performance Score               4
Current Employee Rating         5
dtype: int64
```

Data Manipulation

```
Column:  Title
Unique Values:  ['Accountant I' 'Administrative Assistant' 'Area Sales Manager'
 'BI Developer' 'BI Director' 'CIO' 'Data Analyst' 'Data Analyst '
 'Data Architect' 'Database Administrator' 'Director of Operations'
 'Director of Sales' 'Enterprise Architect' 'IT Director'
 'IT Manager - DB' 'IT Manager - Infra' 'IT Manager - Support'
 'IT Support' 'Network Engineer' 'President & CEO'
 'Principal Data Architect' 'Production Manager' 'Production Technician I'
 'Production Technician II' 'Sales Manager' 'Senior BI Developer'
 'Shared Services Manager' 'Software Engineer'
 'Software Engineering Manager' 'Sr. Accountant' 'Sr. DBA'
 'Sr. Network Engineer']
```

```
Column:  DepartmentType
Unique Values:  ['Admin Offices' 'Executive Office' 'IT/IS' 'Production        ' '
 'Software Engineering']
```

1. Handling White Space Issues

Handling white space issues to avoid double counting on **Title** column and inappropriate records on **DepartmentType** column

2. Change the objected-date type to datetime

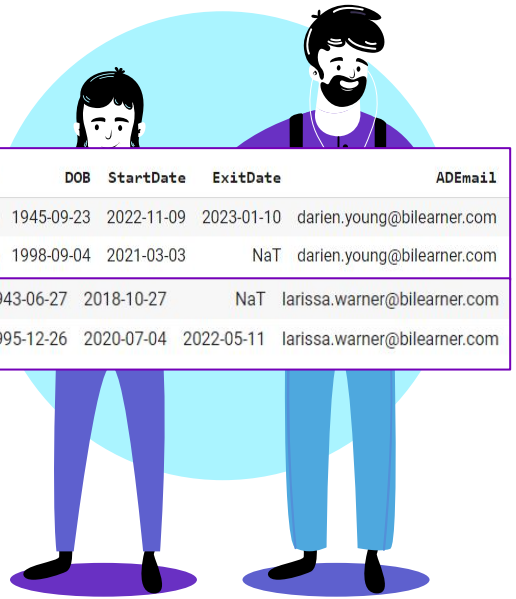
3. Inspect duplicated records

It's become a problem if employees have the same email and still have an active status. The company need to do a real-time validation when new employees register their emails. Employees, need to ask to related department if not-only-once unknown email is being delivered to them

	EmployeeStatus	DOB	StartDate	ExitDate	ADEmail
1974	Terminated for Cause	1945-09-23	2022-11-09	2023-01-10	darien.young@bilearner.com
2204	Active	1998-09-04	2021-03-03	NaT	darien.young@bilearner.com
865	Active	1943-06-27	2018-10-27	NaT	larissa.warner@bilearner.com
1742	Active	1995-12-26	2020-07-04	2022-05-11	larissa.warner@bilearner.com

4. Add 2 new columns

To support analysis, column `Age` and `Tenure` was added to the dataframe



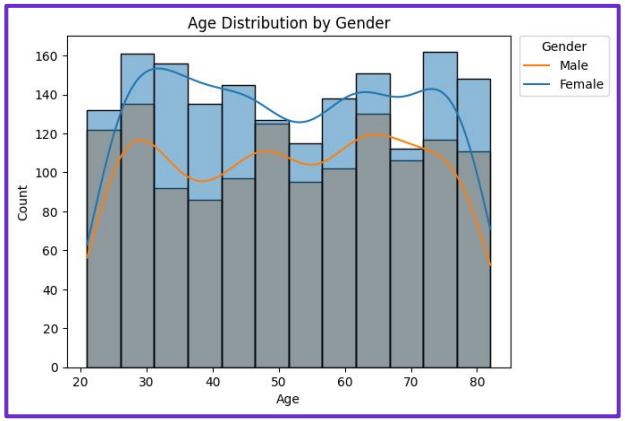
Exploratory Data Analysis

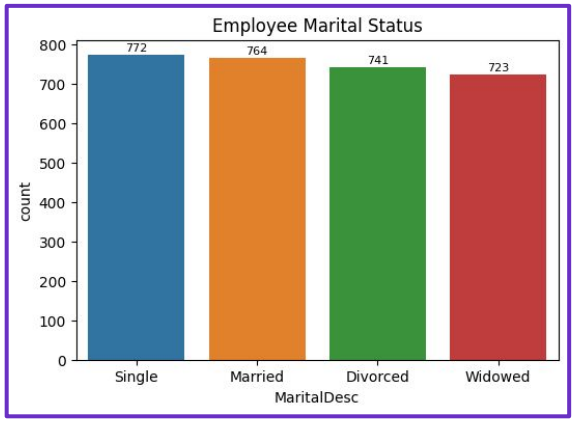
Descriptive and Performance Analysis

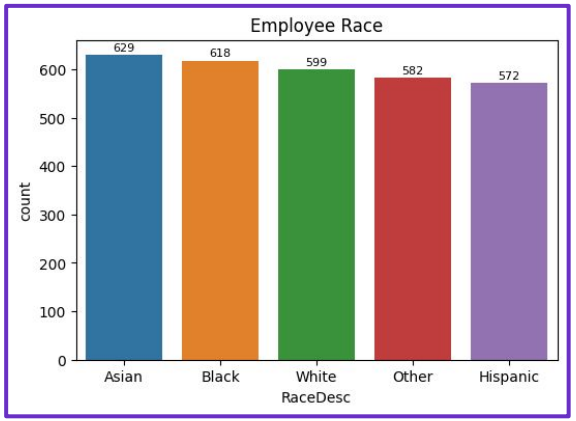
Attrition and Retention Analysis

- **Descriptive and Performance Analysis**

Employees Basic Demographics Information



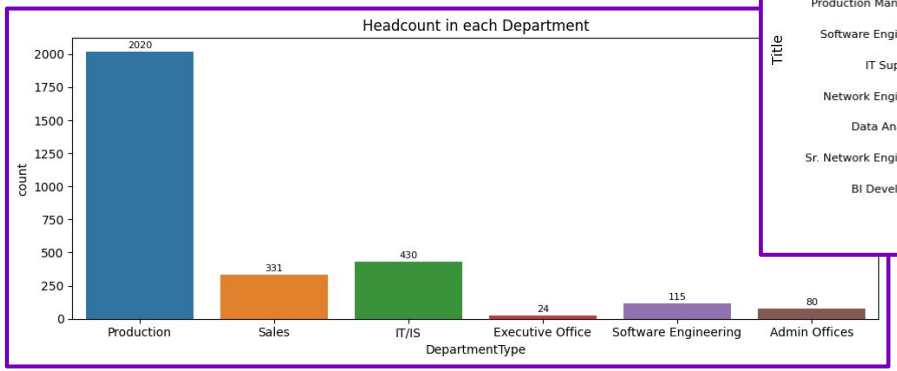


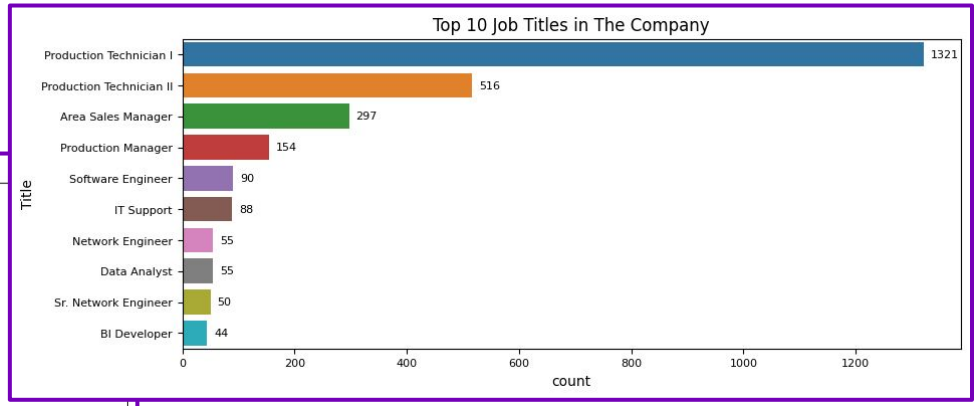


If we look at the first graph, we can see the number of Females is higher than Males. Then, the distribution is slightly similar among ages, marital statuses, and races.

- **Descriptive and Performance Analysis**

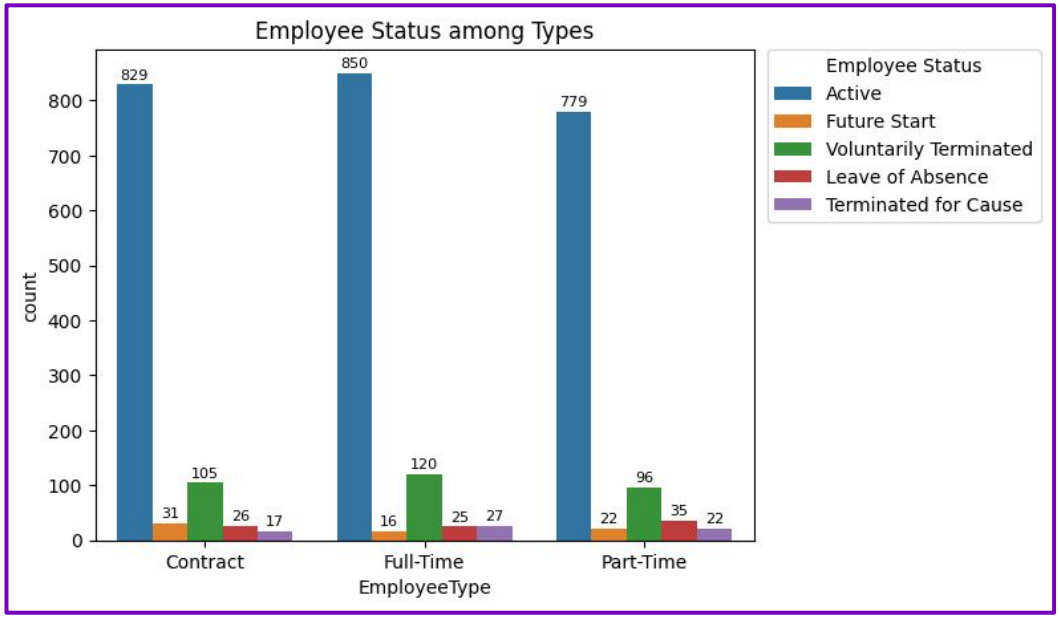
Headcount of all employees that currently working or have worked at this company





Top 10 job titles of employees that currently working or have worked at this company

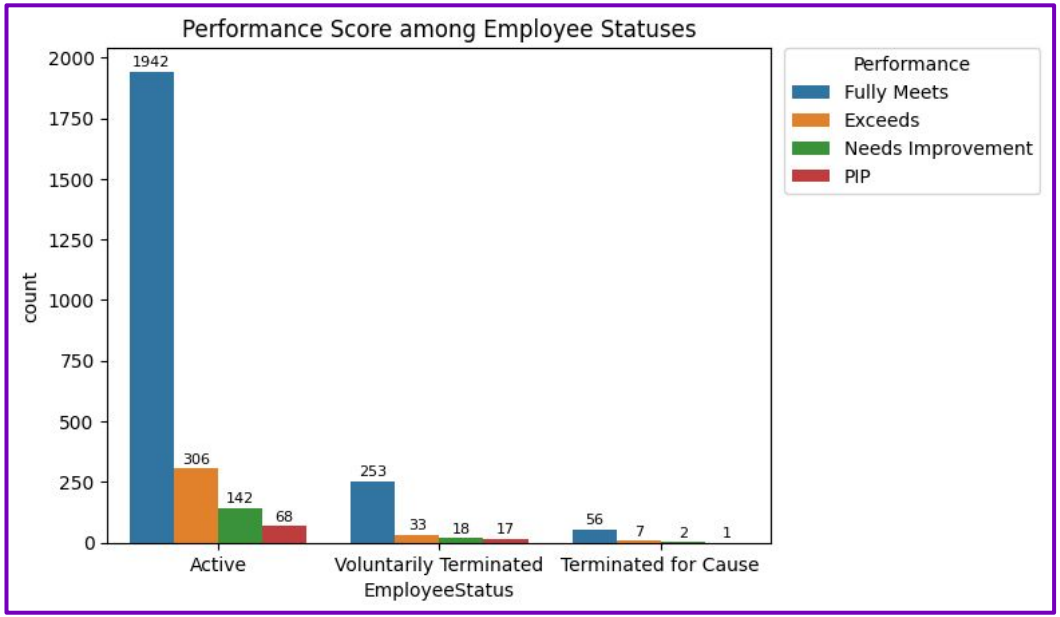
What is the headcount in each employment classification type and employee status?



On all employee classifications, there are 2458 active employees, 387 employees who have been terminated (both voluntarily or involuntarily), and 69 will start soon.

Descriptive and Performance Analysis

How is the performance between active employees and terminated employees?



On all statuses, most employees were scored **Fully Meets**, but this is not clear enough. So, **Current Employee Rating** will be used to see the average rating or evaluation of all employee's performance score

Descriptive and Performance Analysis

What is the average employee rating between active and terminated employees?

EmployeeStatus	Active	Terminated for Cause	Voluntarily Terminated	All
EmployeeClassificationType				
Full-Time	2.967857	2.791667	3.019608	2.968944
Part-Time	3.016688	3.600000	3.095238	3.035595
Temporary	2.897497	3.074074	3.087719	2.924490
All	2.959317	3.090909	3.068536	2.974692

Overall, the average employee performance rating is **2.9 out of 5.0**, which is —we can say, met the company expectations. However, active employees have slightly lower scores compared to other groups.

Descriptive and Performance Analysis

How long have active or terminated employees stayed (tenure)?

EmployeeStatus	Active	Terminated for Cause	Voluntarily Terminated	All
EmployeeClassificationType				
Full-Time	509.193353	534.083333	545.098039	518.514223
Part-Time	478.394040	293.800000	478.647619	471.895735
Temporary	439.377095	549.000000	550.017544	470.585170
All	474.586276	485.575758	525.109034	486.881713

Both involuntarily/voluntarily terminated employee have an **average tenure about 505 days or about a year and a quarter** before they leave. Looking back the employees' ratings, they tend to have higher ratings compared to active employees. Many factors may contribute to the progress of employees' performance before they leave. Such as, the company's training and development programs, company's culture, bonuses, and more. However, it's crucial to the company to assess its employee retention and turnover, **are they good or bad?**

- **Retention and Attrition Analysis**

Make a new dataset and build some metrics

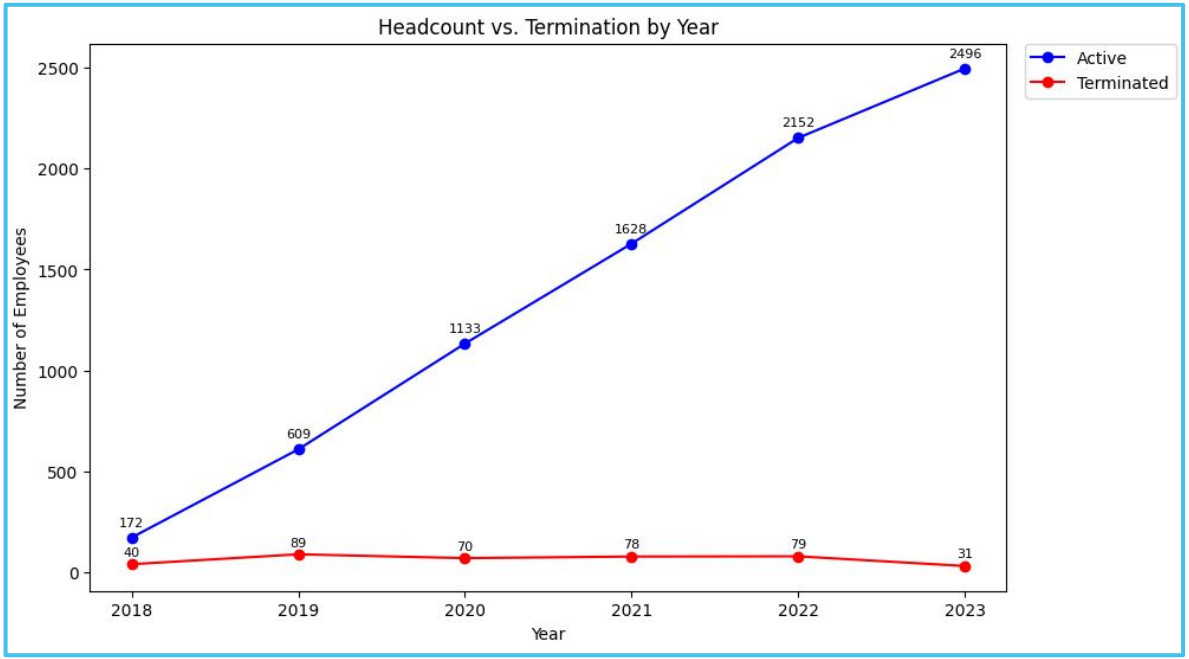
	Year	New Employee	Terminated Employee	Employee Start	Employee End	Attrition Rate	Retention Rate
0	2018	212	40	212	172	20.83	81.13
1	2019	486	89	698	609	13.62	87.25
2	2020	505	70	1203	1133	5.99	94.18
3	2021	503	78	1706	1628	4.68	95.43
4	2022	525	79	2231	2152	3.60	96.46
5	2023	296	31	2527	2496	1.23	98.77

Retention and Attrition Rate Metrics

```
emp_atre['Attrition Rate'] = round((emp_atre['Terminated Employee']) / ((emp_atre['Employee Start'] + emp_atre['Employee End'])/2) * 100,2)
emp_atre['Retention Rate'] = round((emp_atre['Employee End']) / emp_atre['Employee Start'] * 100,2)
```

Note: In 2018, the initial count for Employee Start was setted as 212, reflecting the company's need to have a foundational workforce in place to initiate and support its operations

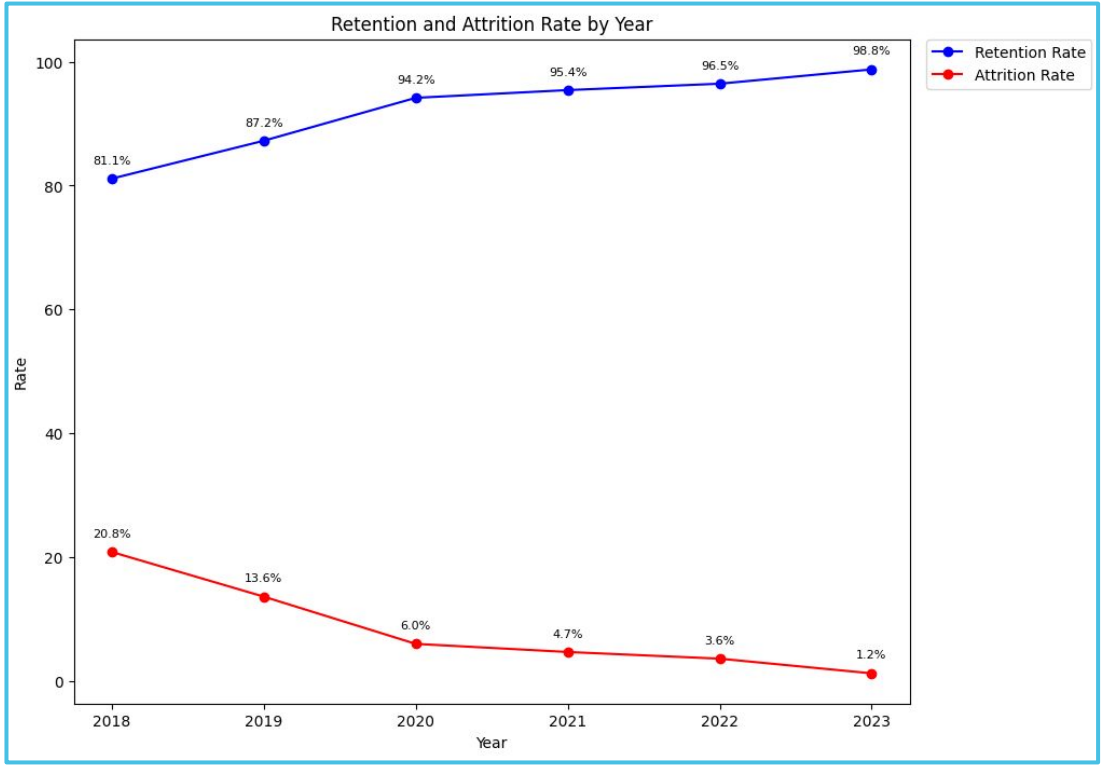
Retention and Attrition Analysis



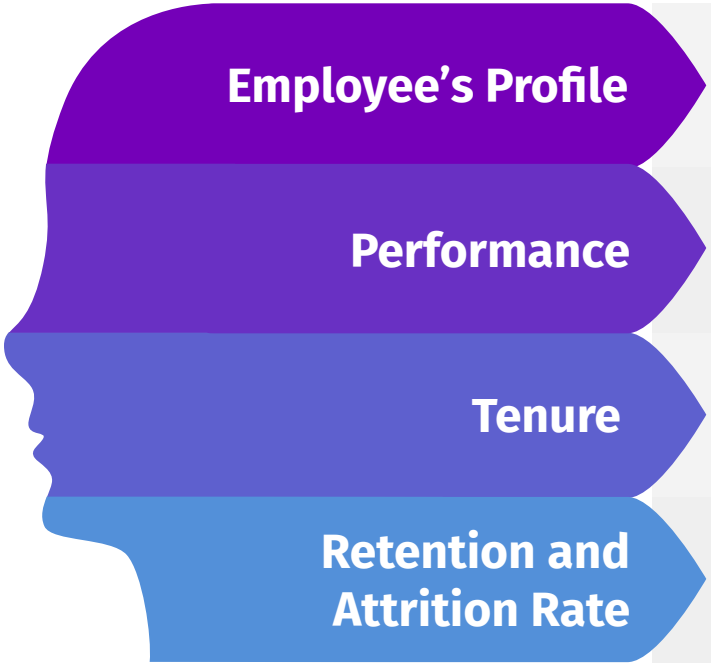
With the significant growth of active employees each year and also a relative stagnant of terminated employees. It may caused by many factors, such as a massive expansion of business, and the company's own strategies. For clearer insight(s) we can see through the retention and attrition rate

Retention and Attrition Analysis

Based on **People Keep**, the business should maintain an attrition rate **lower than 10% and aimed retention rate above 90%**. The company experienced a very high turnover rate at the beginning of business but become better as time flies



Insights



Employee's Profile

With 2458 active employees, 387 employees who have been terminated (both voluntarily or involuntarily), and 69 will start soon. The company have a lot employees who work as Technician, Sales, and Engineer. Also, women are dominating in this company

Performance

About 79% of employees that currently working or have worked at this company, successfully met the company's expectations. Moreover, 12% exceeds the expectations. However, active employees have slightly lower rating scores compared to other groups.

Tenure

Employees typically stayed for an average of a year and three months. When examining the average ratings, it becomes apparent that those who stayed longer tend to have a slightly higher performance levels.

Retention and Attrition Rate

Retention rate tells a good number. On the other hand, The company experienced a very high turnover rate at the beginning of business but become better as time flies

Recommendation

- The company has improved its retention rate over the years, which is a positive sign. However, recognizing the trend of employees staying for around a year and three months it's important to retain employees for longer periods and improve overall employee satisfaction, consider conducting employee engagement surveys, career growth opportunity, build positive environment, and performance feedback sessions to identify areas for improvement in the workplace.
- Data from Glassdoor, a good onboarding program can improve new hire retention by 82%. So, designing a strategic onboarding program for incoming new employees to learn about the company's culture and details about their job description may help the retention of employee.
- To smooth the operational processes, it's also important to regularly check company's data quality

Thank you.

Related info:

linkedin.com/in/azriel-akbar-alfajri

Access to Google Colab*

*you are welcome to giving comment to the content