



CASE STUDY #2

Employee Data Warehouse & Data Mart

Created by Latif

Student of Business Intelligence Bootcamp Batch 9

Dibimbing.id

Table of Content.

- 1** Introduction
- 2** Business Case
- 3** Problems
- 4** Objectives
- 5** About Data Warehouse to use
- 6** Key Metrics and Report
- 7** Building Data Mart
- 8** Conclusion



Latif

Final Year Student of Computer Engineering
Politeknik Elektronika Negeri Surabaya (PENS)

Introduction

Hello everyone, introduce myself Latif student in Bootcamp Dibimbing.id Business Intelligence batch 9. I am happy to share my mini project with you.

In this mini project, I use PostgreSQL tools to create a data warehouse and data mart by taking a case study of a company's employee data.

The main purpose of this mini project is to analyse the performance of a company's employees. The hope is that stakeholders will get various insights related to employees based on the data presented.

Business Case

A company wanted to optimise human resource management and performance tracking. They aim to leverage their data warehouse, which consists of employee, department, and education information as well as employee performance metrics, to make decisions and gain insights based on that data regarding employee performance, department efficiency, and education trends within the organization.

Problems

The company lacks a comprehensive understanding of employee performance across different departments and their educational backgrounds. They struggle to identify patterns and trends that can help in optimising human resource management and improving overall performance.

Objectives

- Gain insight into employee performance across different departments and educational backgrounds.
- Identify the best performing employees, departments and education levels.
- Analyse performance trends over time to understand areas of improvement.
- Optimize employee allocation based on performance and department needs.
- Assess the impact of education level on employee performance.
- Create data-driven strategies to improve overall employee efficiency and productivity.

About Data Warehouse to use

The following is the **employee** data warehouse used. This data warehouse is visualised in the form of an Entity Relational Diagram (ERD). In the ERD used there is one **fact table** and three **dimension tables** with **snowflake schema**.

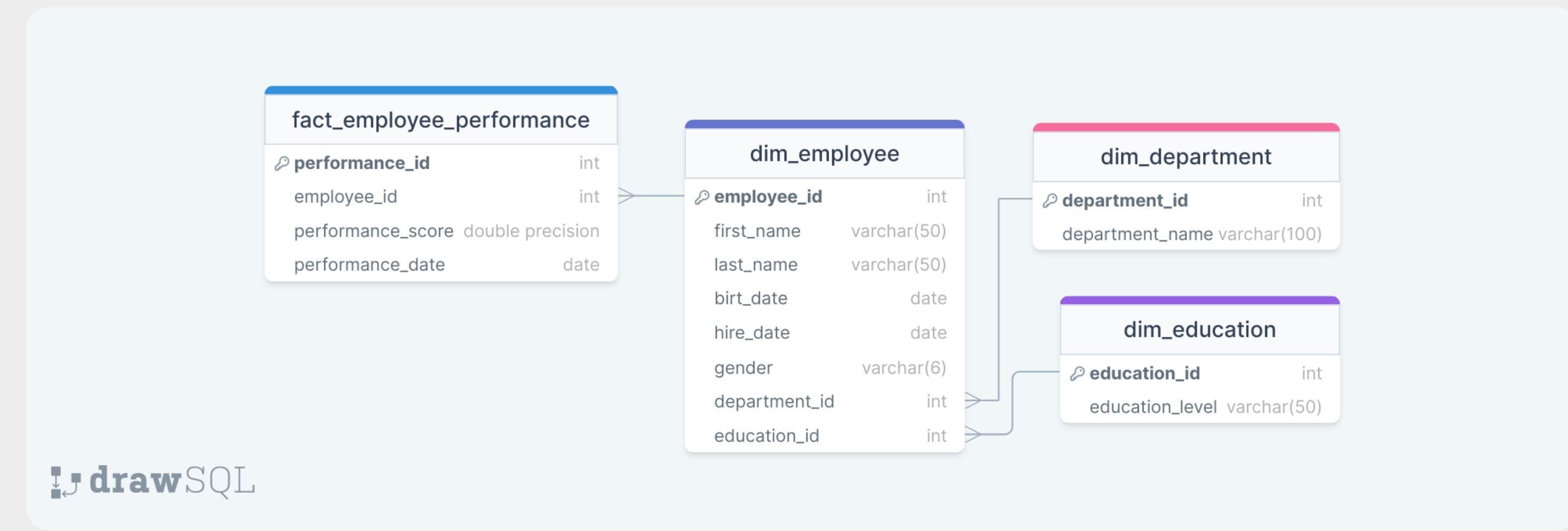
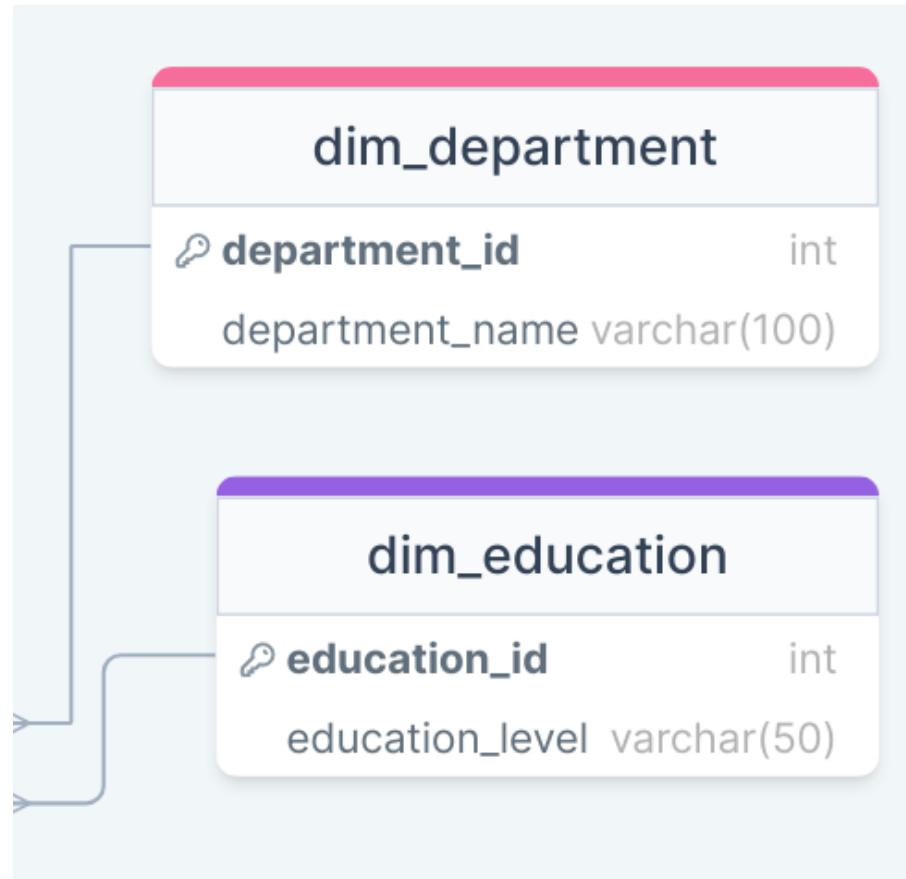


Table Detail on ERD



The **dim_department**, is a dimension table for storing information on the department where employees work. This table has a one-to-many relationship to the **dim_employee** table through the foreign key **department_id** column. this table contains the following columns:

- **department_id**, contains the employee's department id (int) - Primary Key.
- **department_name**, contains the name of the department in the company (varchar).

The **dim_education**, is a dimension table for storing employee education information. This table has a one-to-many relationship to the **dim_employee** table through the foreign key column **education_id**. this table contains the following columns:

- **education_id**, contains the employee's education id (int) - Primary Key.
- **education_level**, contains the employee's education level (varchar).

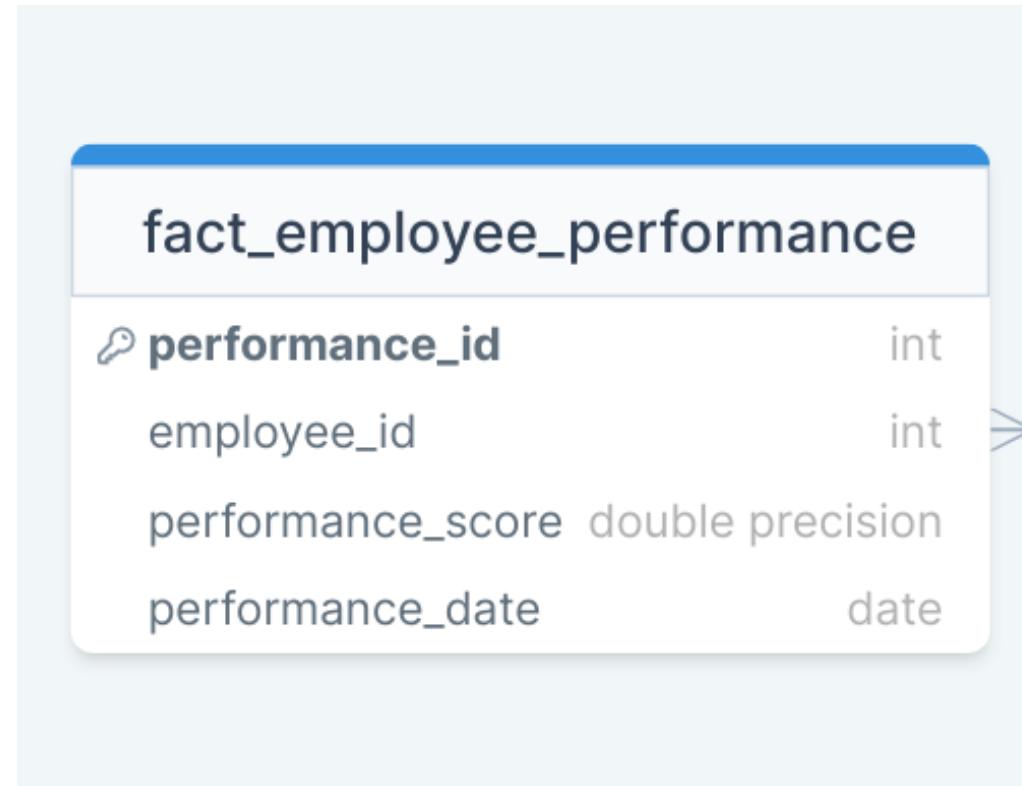
Table Detail on ERD

dim_employee		
employee_id	int	PK
first_name	varchar(50)	
last_name	varchar(50)	
birt_date	date	
hire_date	date	
gender	varchar(6)	
department_id	int	FK
education_id	int	FK

The **dim_employee**, is a dimension table to save unique employee information. This table has a one-to-many relationship to the **fact_employee_performance** table through the foreign key column **employee_id**. This table contains the following columns:

- **employee_id**, Identifier for each employee (integer) - Primary Key.
- **first_name**, contains the employee's first name (varchar).
- **last_name**, contains the employee's last name (varchar).
- **birt_date**, contains the employee's date of birth (date).
- **hire_date**, contains the date the employee was hired (date).
- **gender**, contains the employee's gender (varchar).
- **department_id**, contains the employee's department id (int).
- **education_id**, contains the id that represents the employee's education level (int).

Table Detail on ERD



The **fact_employee_performance**, is a central fact table that saves aggregate information and information related to employee performance. This table has a many-to-one relationship to the **dim_employee** table through the foreign key column **employee_id**. this table contains the following columns:

- **performance_id**, identifier for each employee performance (int) - Primary Key.
- **employee_id**, contains the employee id (int).
- **performance_score**, contains the employee's performance score based on the KPI (double precision).
- **performance_date**, contains the date on which the performance score was assigned (date).

Key Metrics & Report

- **Employee Performance by Department:**

Metric: Average performance score

Report: Performance summary by department

- **Employee Performance Trends:**

Metrics: Performance scores over time

Report: Analyse performance trends over a specific period

- **Best Employee:**

Metric: Highest performance score

Report: List of top employees and their performance scores

- **Performance Distribution by Education Level:**

Metrics: Distribution of performance scores by education level

Report: Performance scores categorised by education level

- **Employee Performance by Gender:**

Metric: Average performance score by gender

Report: Performance analysis by gender

- **Performance Deviation Analysis:**

Metric: Standard deviation of performance scores

Report: Analysis of department performance score deviation

- **New Hire Performance:**

Metrics: New hire performance evaluation

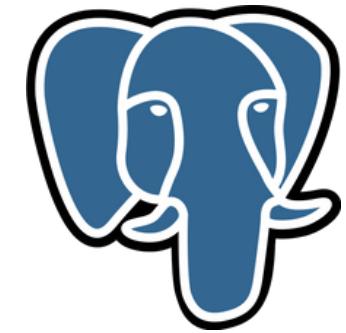
Report: New hire performance scores by time period hired

- **Employee Age Demographics:**

Metric: Employee age distribution

Report: Employee age distribution analysis by date of birth

Building Data Mart



PostgreSQL

Employee Performance by Department:

SQL Query:

```
CREATE VIEW dm.employee_performance_by_department AS
SELECT
    d.department_id,
    d.department_name,
    CAST(AVG(f.performance_score) AS DECIMAL(10, 2)) AS avg_performance_score
FROM
    dwh.dim_department d
INNER JOIN
    dwh.dim_employee e ON d.department_id = e.department_id
INNER JOIN
    dwh.fact_employee_performance f ON e.employee_id = f.employee_id
GROUP BY
    d.department_id, d.department_name
ORDER BY
    avg_performance_score DESC;
```

Employee Performance by Department:

Data Output:

	department_id integer	department_name character varying (100)	avg_performance_score numeric (10,2)
1	10	Administration	53.90
2	5	Operations	51.32
3	2	Finance	50.71
4	3	Marketing	50.18
5	4	IT	50.11
6	6	Sales	49.45
7	9	Customer Service	49.03
8	7	Legal	47.76
9	1	HR	47.00
10	8	Research and Development	46.76

Report:

Based on the average performance score data, the department that has the best performance is the Administration department, then the Operations department in second place and the Finance department in third place. Meanwhile, the department with the lowest performance ranking is the Research and Development department.

Recommendation:

The department with the best performance can be given a reward to trigger other departments to improve their performance. As for the department with the lowest performance, further analysis can be carried out regarding the performance of the performance in the department.

Employee Performance Trends:

SQL Query:

```
CREATE VIEW dm.employee_performance_trend AS
SELECT
    EXTRACT(YEAR FROM performance_date) AS performance_year,
    EXTRACT(MONTH FROM performance_date) AS performance_month,
    CAST( AVG(performance_score) AS DECIMAL(10, 2)) AS avg_performance_score
FROM
    dwh.fact_employee_performance
GROUP BY
    performance_year, performance_month
ORDER BY
    performance_year, performance_month;
```

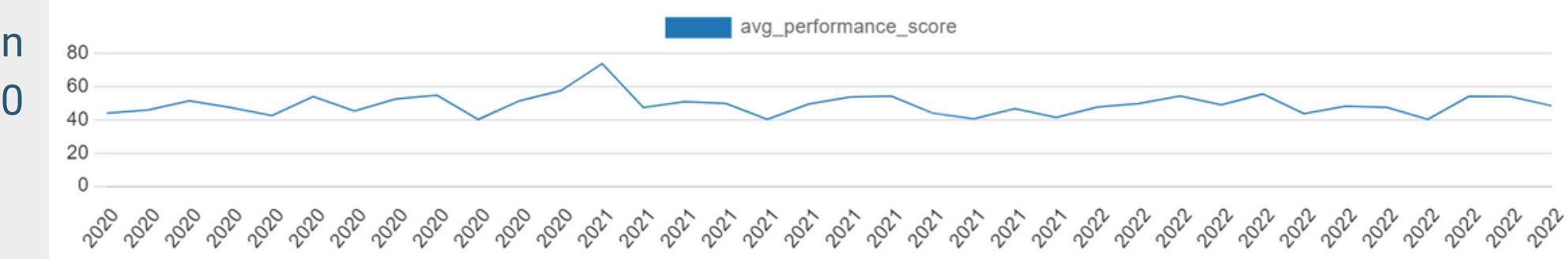
Employee Performance Trends:

Data Output:

	performance_year numeric	performance_month numeric	avg_performance_score numeric (10,2)		performance_year numeric	performance_month numeric	avg_performance_score numeric (10,2)		performance_year numeric	performance_month numeric	avg_performance_score numeric (10,2)	
1	2020	1	44.05		13	2021	1	73.86		25	2022	1
2	2020	2	46.01		14	2021	2	47.38		26	2022	2
3	2020	3	51.41		15	2021	3	51.00		27	2022	3
4	2020	4	47.46		16	2021	4	49.85		28	2022	4
5	2020	5	42.63		17	2021	5	40.33		29	2022	5
6	2020	6	54.02		18	2021	6	49.51		30	2022	6
7	2020	7	45.27		19	2021	7	53.69		31	2022	7
8	2020	8	52.55		20	2021	8	54.42		32	2022	8
9	2020	9	54.87		21	2021	9	44.10		33	2022	9
10	2020	10	40.27		22	2021	10	40.57		34	2022	10
11	2020	11	51.52		23	2021	11	46.77		35	2022	11
12	2020	12	57.61		24	2021	12	41.40		36	2022	12

Report:

based on three years of employee performance trend data, it can be concluded that there is no significant difference in improving employee performance during 2020 to 2022, which has an average range between 40 - 55.



Recommendation:

HRD team can evaluate employee performance every month.

Best Employee:

SQL Query:

```
CREATE VIEW dm.top_performing_employees AS
SELECT
    e.employee_id,
    e.first_name,
    e.last_name,
    e.department_id,
    f.performance_score
FROM
    dwh.dim_employee e
INNER JOIN
    dwh.fact_employee_performance f ON e.employee_id = f.employee_id
ORDER BY
    f.performance_score DESC
LIMIT 20;
```

Best Employee:

Data Output:

	employee_id integer	first_name character varying (50)	last_name character varying (50)	department_id integer	performance_score double precision
1	356	Armando	Whyberd	2	99.99345392794964
2	689	Rori	MacDonough	9	99.95289276267422
3	812	Arliene	Brecknall	9	99.48211848805857
4	406	Brandie	Binford	1	99.47554551437605
5	556	Loutitia	Alessandone	3	99.14297853905703
6	211	Napoleon	Cromarty	6	99.128372523017
7	45	Jerri	Cholton	9	99.12381793011704
8	765	Verla	Brashaw	1	99.018703456024
9	770	Avigdor	Howie	6	98.73801632977144
10	936	Jeffy	Wehnerr	1	98.71055826020742
11	695	Lucia	Farmacy	9	98.7029161475595
12	829	Claus	Orringe	7	98.54524518444978
13	707	Diena	Ryott	10	98.53116992937917
14	759	Birgit	Laurand	5	98.40633043792111
15	698	Barrett	Skin	8	98.3939999617546
16	549	Philbert	Cattonnet	9	98.35832772852058
17	21	Hillery	Parris	2	98.34831252250262
18	240	Feodora	Claiden	4	97.95346518123212
19	830	Gwennie	Djakovic	9	97.77338070592766
20	370	Vinita	Statersfield	10	97.67308977894228

Report:

The following is data on employees with the best performance performance scores based on data from 2020 to 2023. the data presented is 20 employees who get the best performance performance scores in the company.

Recommendation:

Companies can consider giving rewards to employees with the best performance scores, in order to foster enthusiasm at work and can maintain employee performance.

Performance Distribution by Education Level:

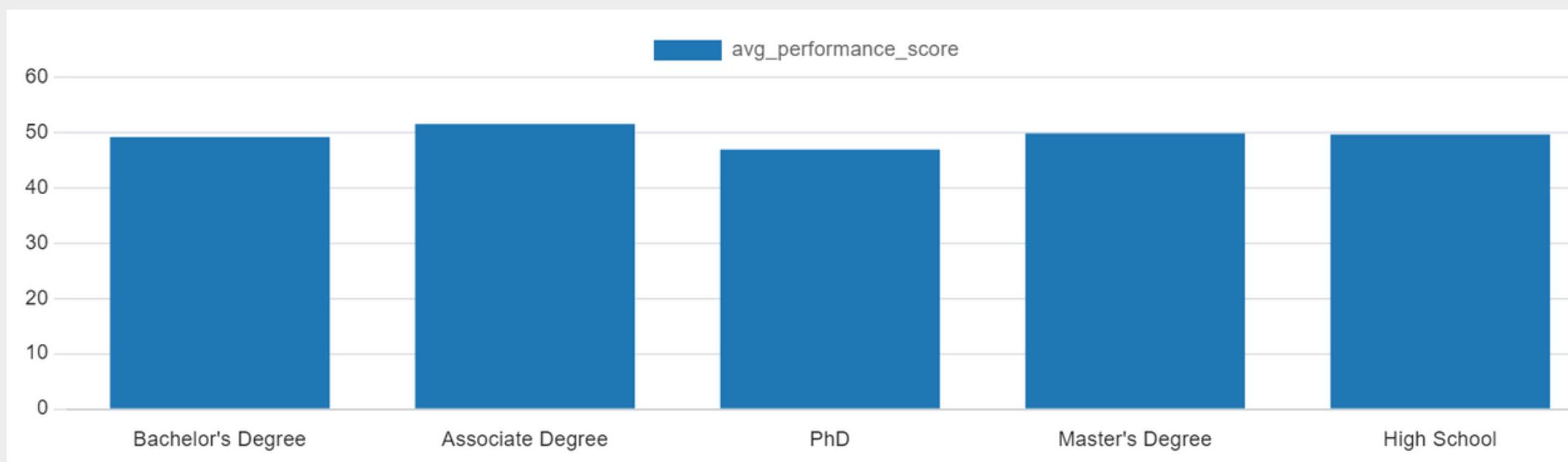
SQL Query:

```
CREATE VIEW dm.performance_distribution_by_education AS
SELECT
    ed.education_level,
    COUNT(*) AS employee_count,
    AVG(f.performance_score) AS avg_performance_score
FROM
    dwh.dim_education ed
INNER JOIN
    dwh.dim_employee e ON ed.education_id = e.education_id
INNER JOIN
    dwh.fact_employee_performance f ON e.employee_id = f.employee_id
GROUP BY
    ed.education_level;
```

Performance Distribution by Education Level:

Data Output:

	education_level character varying (50) 	employee_count bigint 	avg_performance_score double precision 
1	Bachelor's Degree	209	49.228151487257485
2	Associate Degree	214	51.57069875349984
3	PhD	190	46.95257079630433
4	Master's Degree	193	49.86262604346513
5	High School	194	49.68513783950901



Report:

Based on the data presented, the company has the most employees with Associate Degree level with 214 employees and has the highest average performance score compared to other education levels.

Recommendation:

The company can consider hiring employees with Associate Degree graduates if needed because based on data from education graduates at the Associate Degree level have an average employee performance score that is quite good in the company.

Employee Performance by Gender:

SQL Query:

```
CREATE VIEW dm.employee_performance_by_gender AS
SELECT
    e.gender,
    AVG(f.performance_score) AS avg_performance_score
FROM
    dwh.dim_employee e
INNER JOIN
    dwh.fact_employee_performance f ON e.employee_id = f.employee_id
GROUP BY
    e.gender;
```

Employee Performance by Gender:

Data Output:

	gender character varying (6) 	avg_performance_score double precision 
1	Female	49.33558038574339
2	Male	49.57237713045363

Report:

Based on the data displayed, employees with male gender have a slightly higher average performance score than employees with female gender. However, the average difference between the two is so small that it can be said that this average difference does not have a significant impact on employee performance.

Recommendation:

The company does not need to make gender the main consideration related to employee performance, there may be other factors that can be used as the main consideration in assessing employee performance.

Performance Deviation Analysis:

SQL Query:

```
CREATE VIEW dm.performance_deviation AS
SELECT
    e.department_id,
    d.department_name,
    STDDEV(f.performance_score) AS performance_deviation
FROM
    dwh.dim_department d
INNER JOIN
    dwh.dim_employee e ON d.department_id = e.department_id
INNER JOIN
    dwh.fact_employee_performance f ON e.employee_id = f.employee_id
GROUP BY
    e.department_id, d.department_name;
```

Performance Deviation Analysis:

Data Output:

	department_id integer	department_name character varying (100)	performance_deviation double precision
1	7	Legal	27.59674444683699
2	5	Operations	29.722555138389705
3	9	Customer Service	30.034912860465095
4	10	Administration	29.093595974403797
5	2	Finance	32.42385747661925
6	1	HR	29.969338903746515
7	3	Marketing	27.434204214781985
8	8	Research and Development	26.970204214369293
9	6	Sales	29.5644133060435
10	4	IT	29.335081748138208

Report:

Based on the data, the performance deviation analysis shows the variation in employee performance across different departments. The Finance department has the highest level of performance deviation (32.42), signalling significant variation in employee performance in the department. On the other hand, the Research and Development Department has the lowest level of performance deviation (26.97), indicating higher consistency in employee performance.

Recommendation:

Based on the analysis of performance deviations, it is recommended that department managers immediately identify the causes of significant performance variations and take corrective actions. Collaboration between the HR team, managers, organizational leaders, and employees will help design appropriate development strategies to improve overall performance.

New Hire Performance:

SQL Query:

```
CREATE VIEW dm.new_hire_performance AS
SELECT
    e.employee_id,
    e.first_name,
    e.last_name,
    e.hire_date,
    f.performance_score
FROM
    dwh.dim_employee e
INNER JOIN
    dwh.fact_employee_performance f ON e.employee_id = f.employee_id
ORDER BY
    e.hire_date DESC;
```

New Hire Performance:

Data Output:

	employee_id	first_name	last_name	hire_date	performance_score
1	402	Gerty	Whittle	2024-02-09	62.287908572941156
2	239	Lenee	Goney	2024-02-09	54.91229265869961
3	2	Oralla	Minnette	2024-02-05	38.40222708316501
4	540	Alida	Mabbott	2024-02-03	18.447920735270262
5	310	Selena	Shevels	2024-02-02	0.8403086532049242
6	524	Petra	Crolly	2024-02-02	85.43309400837364
7	72	Dionisio	Aggiss	2024-02-02	1.7370520309907267
8	375	Rayna	Layson	2024-01-30	55.06338221001374
9	395	Natalina	Bidgod	2024-01-30	65.3594466018852
10	213	Cullin	Holdron	2024-01-29	76.31564970766442
11	563	Claudie	Woof	2024-01-28	92.62709787546204
12	403	Rawley	Stinson	2024-01-27	3.590171654784502
13	154	Sonny	Stading	2024-01-27	40.21589573447677
14	247	Othello	Gooch	2024-01-26	40.140322010870364
15	974	Welbie	Bauer	2024-01-25	18.844970930727722

Total rows: 1000 of 1000

Query complete 00:00:00.305

Report:

The data on the side shows that the performance analysis of newly hired employees shows variations in their performance since they were hired. There is a wide range of performance, from employees who achieve high performance after being recruited to employees who take longer to adapt and achieve expected performance.

Recommendation:

The company can ensure that the recruitment and onboarding process runs smoothly and that new employees can quickly contribute to the success of the organization.

Employee Age Demographics:

SQL Query:

```
CREATE VIEW dm.employee_age_distribution AS
SELECT
    CASE
        WHEN EXTRACT(YEAR FROM CURRENT_DATE) - EXTRACT(YEAR FROM e.birt_date) -
            (CASE
                WHEN TO_CHAR(CURRENT_DATE, 'MM-DD') < TO_CHAR(e.birt_date, 'MM-DD') THEN 1
                ELSE 0
            END) >= 0 THEN EXTRACT(YEAR FROM CURRENT_DATE) - EXTRACT(YEAR FROM e.birt_date) -
            (CASE
                WHEN TO_CHAR(CURRENT_DATE, 'MM-DD') < TO_CHAR(e.birt_date, 'MM-DD') THEN 1
                ELSE 0
            END)
        ELSE NULL
    END AS age,
    COUNT(*) AS employee_count
FROM
    dwh.dim_employee e
GROUP BY
    age;
```

Employee Age Demographics:

Data Output:

	age numeric	employee_count bigint
1	31	118
2	24	98
3	27	101
4	28	96
5	26	103
6	34	2
7	25	95
8	33	115
9	32	100
10	29	92
11	30	80

Report:

Based on the data presented, employees working in this company have a varied age range between the ages of 24 years - 34 years. This distribution can provide insight that there is a diversity of age and work experience in the organization.

Recommendation:

By paying attention to the age diversity of employees and taking appropriate action, organisations can strengthen an inclusive culture, improve performance and ensure long-term success.

Conclusion

Based on the mart data presented, the company can better **understand** and **manage** employee performance, gain **insights** into performance trends, identify **factors** affecting performance, and understand employee demographics and characteristics to **support** better **strategic** decision-making.



Thank You

-gladly accept comments and discuss together-

www.linkedin.com/in/latif-/