

Laporan Portofolio “SQL for Data Analysis”

Pendahuluan

Proyek ini merupakan bagian dari kursus online SQL yang saya ikuti. Tujuan proyek ini adalah untuk menganalisis data menggunakan query SQL dan memberikan wawasan mengenai dataset yang digunakan.

Deskripsi Proyek

Judul : Analisis gaji seorang data sains

Sumber Data : Data diperoleh dari platform Kaggle. Berikut link dataset

<https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries>

Tujuan : Menganalisis wawasan apa saja yang bisa diperoleh dari dataset tersebut

Tool yang digunakan : MySQL Workbench 8.0 CE

Struktur database

Dataset ini terdiri dari beberapa fitur, yaitu:

1. **Work_year** : tahun gaji dibayarkan
2. **Experience_level** : tingkat pengalaman kerja (EN Entry-Level, MI Mid-level, SE Senior-level, EX Expert-level)
3. **Employement_type** : jenis pekerjaan (paruh waktu, penuh waktu, kontrak, freelance)
4. **Job_title** : jabatan pekerjaan
5. **Salary** : jumlah gaji kotor
6. **Salary_currency** : mata uang gaji yang dibayarkan sebagai kode mata uang ISO 4217
7. **Salary_in_usd** : gaji dalam USD
8. **Employee_residence** : negara tempat tinggal karyawan
9. **Remote_ratio** : jumlah keseluruhan pekerjaan yang dilakukan jarak jauh
10. **Company_location** : negara kantor pusat atau cabang kontrak
11. **Company_size** : rata-rata jumlah orang yang bekerja di perusahaan selama setahun

Implementasi Query

Berikut adalah query yang digunakan dalam menganalisis dataset tersebut:

- Melakukan pengecekan terhadap data null

```
SELECT * FROM ds_salaries
WHERE work_year IS NULL
OR experience_level IS NULL
OR employment_type IS NULL
OR job_title IS NULL
OR salary IS NULL
OR salary_currency IS NULL
OR salary_in_usd IS NULL
OR employee_residence IS NULL
OR remote_ratio IS NULL
OR company_location IS NULL
OR company_size IS NULL;
```

- Daftar job title

```
SELECT DISTINCT job_title FROM ds_salaries ORDER BY job_title;
```

- Menampilkan job title yang berkaitan dengan data analis

```
SELECT DISTINCT job_title FROM ds_salaries WHERE job_title LIKE '%data analyst' ORDER BY job_title;
```

- Menampilkan rata-rata gaji seorang data analis perbulannya dalam rupiah

```
SELECT (AVG(salary_in_usd)*15000)/12 AS salary_in_rp_monthly FROM ds_salaries;
```

- Menampilkan rata-rata gaji data analis berdasarkan experience level

```
SELECT experience_level, (AVG(salary_in_usd)*15000)/12 AS salary_in_rp_monthly
FROM ds_salaries
GROUP BY experience_level;
```

- Menampilkan rata-rata gaji data analis berdasarkan experience level dan jenis employee nya

```
SELECT experience_level, employment_type, (AVG(salary_in_usd)*15000)/12 AS salary_in_rp_monthly
FROM ds_salaries
GROUP BY experience_level, employment_type
ORDER BY experience_level, employment_type;
```

- Negara dengan gaji menarik untuk posisi Data Analyst (full-time) dengan experience level entry dan mid-level

```
SELECT company_location, AVG(salary_in_usd) AS avg_sal_in_usd FROM ds_salaries
WHERE job_title LIKE '%data analyst%' AND employment_type = 'FT' AND experience_level IN ('MI', 'EN')
GROUP BY company_location
HAVING avg_sal_in_usd >= 20000;
```

- Di tahun berapa kenaikan gaji dari mid-level ke senior (EX) memiliki kenaikan yang tertinggi pada posisi yang berkaitan dengan data analis (fulltime)

```
WITH ds_1 AS (
    SELECT work_year, AVG(salary_in_usd) AS sal_in_usd_ex
    FROM ds_salaries
    WHERE employment_type = 'FT' AND experience_level = 'EX' AND job_title LIKE '%data analyst%'
    GROUP BY work_year
), ds_2 AS (
    SELECT work_year, AVG(salary_in_usd) AS sal_in_usd_mid
    FROM ds_salaries
    WHERE employment_type = 'FT' AND experience_level = 'MI' AND job_title LIKE '%data analyst%'
    GROUP BY work_year
), t_year AS (
    SELECT DISTINCT work_year FROM ds_salaries
) SELECT t_year.work_year, ds_1.sal_in_usd_ex,
    ds_2.sal_in_usd_mid, ds_1.sal_in_usd_ex - ds_2.sal_in_usd_mid difference
FROM t_year
LEFT JOIN ds_1 ON ds_1.work_year = t_year.work_year
LEFT JOIN ds_2 ON ds_2.work_year = t_year.work_year;
```

Hasil dan Analisis

Dari hasil query yang dijalankan, ditemukan bahwa:

- Tidak ada data yang null
- Terdapat 50 job title
- Terdapat 9 job title yang berkaitan dengan data analis
- Rata-rata gaji seorang data analis perbulannya adalah sekitar 140 juta rupiah
- Rata-rata gaji data analis pada level junior sekitar 110 juta rupiah, level senior sekitar 173 juta rupiah, level pemula sekitar 77 juta rupiah, dan level profesional sekitar 249 juta rupiah.
- Gaji yang tertinggi adalah pada mid-level untuk fulltime mencapai sekitar 110 juta rupiah
- Negara dengan gaji yang menarik untuk posisi data analis adalah Amerika Serikat dengan rata-rata gaji 101 ribu usd
- Kenaikan gaji tertinggi berada pada tahun 2022 dengan jumlah kenaikan 51 ribu usd

Kesimpulan

Pada proyek ini, menunjukkan bagaimana SQL dapat digunakan untuk menganalisis data secara efektif. Suatu perusahaan dapat mengambil keputusan yang lebih baik untuk strategi bisnisnya.

Referensi

File sql : <https://github.com/annsalsha/SQL-Data-Science-/blob/main/data%20science.sql>