

**LAPORAN PRAKTIKUM**  
**UJIAN AKHIR SEMESTER**  
**DATA WAREHOUSE**



Disusun Oleh:

<b>Dzulfikar Muhammad Al Ghifari</b>	<b>2341760071</b>
<b>Muhammad Afiq Firdaus</b>	<b>2341760189</b>
<b>Nimas Septiandini</b>	<b>2341760087</b>
<b>Zaki Muhammad Athallah Erlangga</b>	<b>2341760018</b>

**JURUSAN TEKNOLOGI INFORMASI**  
**D-IV SISTEM INFORMASI BISNIS**  
**POLITEKNIK NEGERI MALANG**  
**2025**

## 1. Studi Kasus Gaji Karyawan

Dataset Employee Salaries – 2024 pada Data.gov berisi data gaji tahunan karyawan tetap dan aktif Pemerintah Montgomery County, Maryland, Amerika Serikat, untuk tahun kalender 2024. Isi dari dataset ini mencakup atribut seperti: Department, Department Name, Division, Gender, Base Salary, 2024 Overtime Pay, 2024 Longevity Pay, Grade

Dataset ini nantinya dapat digunakan untuk membangun proyek ETL dan data warehouse yang berfokus pada analisis penggajian, evaluasi anggaran departemen, pemantauan distribusi jabatan, serta pelaporan performa kepegawaian pemerintah daerah secara lebih efisien dan terstruktur.

Proyek ETL ini bertujuan mengubah data penggajian karyawan pemerintah menjadi data warehouse terstruktur berbasis star schema. Hasilnya dapat digunakan untuk analisis gaji, pelaporan keuangan, dan pengambilan keputusan strategis terkait pengelolaan SDM di lingkup organisasi pemerintahan.

## 2. Dataset yang dipilih (*penjelasan ttg isi dataset*)

Dataset: [Montgomery County of Maryland - Employee Salaries - 2024](#)

Dataset “Montgomery County of Maryland - Employee Salaries - 2024” adalah dataset terbuka (open data) yang disediakan oleh pemerintah Kabupaten Montgomery, negara bagian Maryland, Amerika Serikat. Dataset ini memuat informasi gaji tahunan para pegawai negeri sipil (PNS) di wilayah tersebut untuk tahun fiskal 2024. Data yang tersedia mencakup nama pegawai, jabatan atau posisi pekerjaan, departemen tempat bekerja, tingkat atau golongan gaji, status kepegawaian (seperti full-time atau part-time), tanggal mulai bekerja, hingga jumlah gaji kotor aktual yang diterima selama periode tersebut.

Tujuan utama dari publikasi data ini adalah untuk meningkatkan transparansi dan akuntabilitas dalam pengelolaan anggaran pemerintah, khususnya dalam hal pengeluaran untuk sumber daya manusia. Selain itu, dataset ini dapat dimanfaatkan untuk berbagai keperluan analisis seperti perbandingan struktur gaji antar departemen, distribusi penghasilan berdasarkan posisi atau status kerja, hingga bahan evaluasi kebijakan penggajian oleh pihak legislatif, akademisi, maupun publik. Dataset ini sangat cocok digunakan sebagai dasar dalam proyek ETL (Extract, Transform, Load)

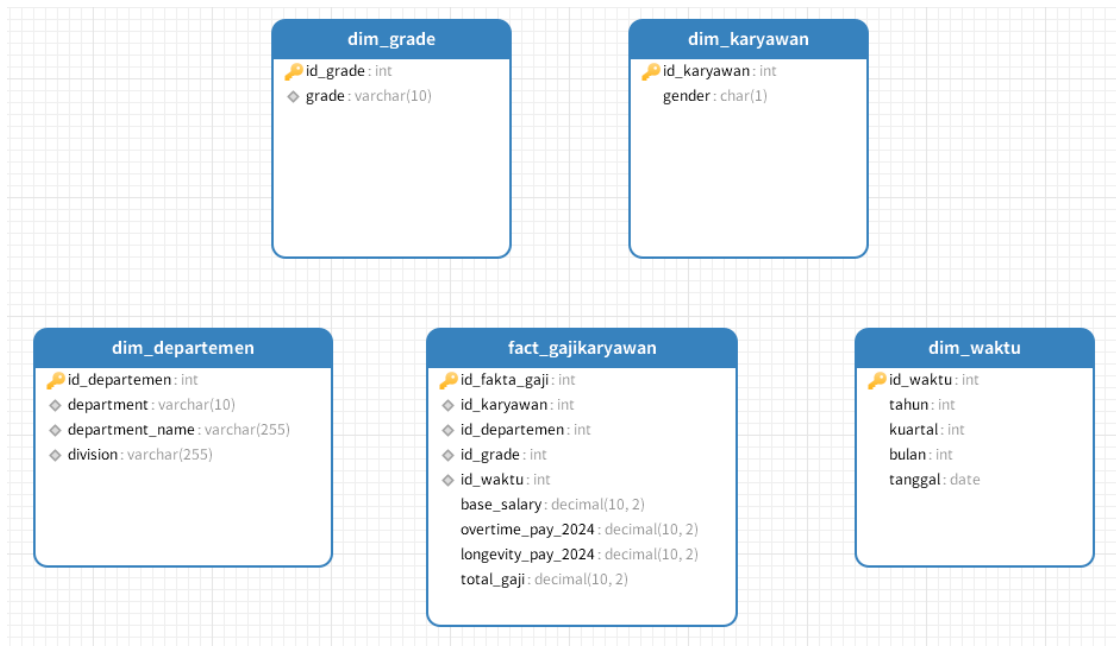
dan implementasi data warehouse karena memiliki struktur yang konsisten dan mencerminkan proses administratif yang nyata.

### 3. Data Gaji Karyawan

id # int	department # varchar(10)	department_name # text	division # text	gender # char(1)	base_salary # decimal(10,2)	_2024_overtime_pay # decimal(10,2)	_2024_longevity_pay # decimal(10,2)	grade # varchar(10)
1	ABS	Alcohol Beverage Services	ABS 85 IT Administration	M	169555,00	0,00	0,00	M3
2	CUS	Community Use of Public	CUS 70 Finance and Adm	F	141876,93	0,00	0,00	M3
3	DGS	Department of General Ser	DGS 36 Fleet Management	F	87523,00	748,37	2701,48	16
4	CEX	Offices of the County Exec	CEX 15 Chief Administrativ	F	207000,00	0,00	0,00	EX2
5	HHS	Department of Health and	HHS 60 Office of Eligibility	F	91392,00	0,00	5733,27	N17
6	OAG	Office of Agriculture	OAG 78 Agricultural Servic	F	108899,00	0,00	3361,18	N21
7	COR	Correction and Rehabilitati	COR 42 MSD Fiscal Service	F	160088,68	0,00	0,00	M3
8	HHS	Department of Health and	HHS 60 Public Health Servi	F	130181,00	0,00	8166,53	N25
9	HHS	Department of Health and	HHS 60 Office of Eligibility	F	104197,00	0,00	3216,00	20
10	DGS	Department of General Ser	DGS 36 Real Estate Office	F	169555,00	0,00	0,00	M3
11	DOT	Department of Transportat	DOT 50 Transit COS Taxica	F	108899,00	0,00	4045,18	21
12	SHF	Sheriff's Office	SHF 48 Administration	F	125647,51	1053,19	0,00	N25
13	TBS	Department of Technology	TBS 34 CIO Finance and Hi	F	119032,00	0,00	1184,88	N23
14	HHS	Department of Health and	HHS 60 Chief Operating Oi	M	194980,00	0,00	0,00	M2
15	OHR	Office of Human Resource	OHR 33 Administrative Ser	F	194980,00	0,00	0,00	M2
16	HHS	Department of Health and	HHS 60 Public Health Servi	F	119032,00	0,00	7467,13	N23
17	HHS	Department of Health and	HHS 60 Chief Operating Oi	F	130181,00	0,00	8166,53	N25
18	HHS	Department of Health and	HHS 60 Office of Eligibility	F	95458,00	0,00	5988,20	18
19	IGR	Office of Intergovernment	IGR 20 Office of Intergover	F	108899,00	21,28	0,00	N21
20	DOT	Department of Transportat	DOT 50 Director Managem	F	119032,00	0,00	7467,13	N23
21	DGS	Department of General Ser	DGS 36 Building Design an	F	149112,00	0,00	9354,05	28
22	HHS	Department of Health and	HHS 60 Chief Operating Oi	M	217359,00	0,00	0,00	M1
23	FIN	Department of Finance	FIN 32 Accounts Payable	M	130181,00	0,00	650,92	N25
24	HHS	Department of Health and	HHS 60 Services to End an	F	136188,00	0,00	0,00	N26
25	POL	Department of Police	POL 47 MSB Training and I	M	115223,00	4110,90	12285,18	P4
26	DOT	Department of Transportat	DOT 50 Highway Services	F	83837,00	17425,48	5259,22	15
27	DGS	Department of General Ser	DGS 36 Building Design an	F	138535,00	0,00	0,00	N28
28	DGS	Department of General Ser	DGS 36 Fleet Management	M	194980,00	0,00	0,00	M2
29	DEP	Department of Environmer	DEP 80 Waste Reduction ar	F	169555,00	0,00	0,00	M3
30	BOA	Board of Appeals Departm	BOA 02 Board of Appeals C	F	95458,00	0,00	5988,20	N18
31	COR	Correction and Rehabilitati	COR 42 MSD Fiscal Service	F	127509,43	0,00	7845,63	N26
32	HHS	Department of Health and	HHS 60 Aging & Disability	F	108899,00	0,00	9070,24	21
33	HCA	Department of Housing an	HCA 76 Finance and Admi	F	119032,00	0,00	7467,13	N23

Dataset Montgomery County of Maryland - Employee Salaries - 2024 berisi informasi gaji pegawai pemerintah di berbagai departemen wilayah Montgomery County. Setiap entri mencakup data seperti nama departemen, divisi kerja, jenis kelamin, gaji pokok tahunan, total lembur dan tunjangan masa kerja selama tahun 2024, serta grade atau level jabatan. Dataset ini berguna untuk menganalisis struktur kompensasi, distribusi gaji antar departemen, dan kesetaraan gender dalam penghasilan di sektor pemerintahan lokal.

#### 4. Rancangan Star Schema



Gambar tersebut menunjukkan rancangan star schema untuk data warehouse gaji karyawan Pemerintah Montgomery County. Di pusatnya terdapat tabel fakta **fact\_gajikaryawan** yang menyimpan data numerik seperti gaji pokok, lembur, tunjangan masa kerja, dan total gaji. Tabel ini terhubung ke empat tabel dimensi: **dim\_karyawan** (berisi informasi jenis kelamin karyawan), **dim\_departemen** (menyimpan data departemen, nama departemen, dan divisi), **dim\_grade** (berisi level jabatan karyawan), dan **dim\_waktu** (yang memuat atribut waktu seperti tahun, kuartal, bulan, dan tanggal). Skema ini memungkinkan analisis data gaji dari berbagai perspektif, seperti tren gaji per bulan, perbedaan gaji berdasarkan gender, departemen, atau grade jabatan, serta evaluasi biaya lembur sepanjang tahun. Struktur ini sangat mendukung kebutuhan pelaporan dan pengambilan keputusan berbasis data secara cepat dan efisien.

#### 5. Rancang Proyek ETL Berdasarkan Star Schema



Dalam proyek ETL ini, terdapat 6 proses utama yang dirancang secara berurutan untuk memindahkan data dari sumber menuju data warehouse. Berikut penjelasan tiap langkah:

1. Start, Langkah awal dalam job ETL untuk menandai permulaan proses otomatisasi.
2. Transformation 1 - dim\_departemen\_load, Memuat data dari sumber (dataset) ke dalam tabel dim\_departemen. Data ini mencakup kode departemen, nama lengkap departemen, dan divisi.
3. Transformation 2 - dim\_karyawan\_load, Memuat informasi karyawan, seperti ID dan gender, ke dalam tabel dim\_karyawan.
4. Transformation 3 - dim\_grade\_load, Memuat level atau grade jabatan dari masing-masing karyawan ke dalam tabel dim\_grade.
5. Transformation 4 - dim\_waktu\_load, Menghasilkan dimensi waktu (tahun, kuartal, bulan, tanggal) secara otomatis dan memuatnya ke tabel dim\_waktu. Ini mendukung analisis temporal.
6. Transformation 5 - fact\_gaji\_karyawan\_load, Merupakan transformasi utama yang memuat data numerik ke tabel fact\_gajikaryawan, termasuk:
  - i. Gaji pokok (base\_salary)
  - ii. Lembur (overtime\_pay\_2024)
  - iii. Tunjangan masa kerja (longevity\_pay\_2024)
  - iv. Total gaji (hasil kalkulasi)

Tabel ini menghubungkan seluruh dimensi melalui foreign key.

## **6. Langkah-Langkah Proses ETL**

### **A. Ekstraksi (E) – Ekstraksi Data**

Tujuan utama dari tahap ekstraksi data adalah untuk mengambil seluruh data mentah dari sumbernya agar siap untuk diolah lebih lanjut. Dalam proses ini, data akan diambil dari file SQL bernama uas-dw-salary.sql. Tabel-tabel yang diekstrak, beserta peran masing-masing, adalah sebagai berikut:

- employees: Berisi data utama terkait gaji, gender, grade karyawan, dan departemen tempat karyawan bekerja.

- `dim_karyawan`: Berisi informasi spesifik mengenai gender per ID karyawan.
- `dim_grade`: Berisi informasi tingkatan jabatan atau grade.
- `dim_departemen`: Berisi nama dan divisi departemen.
- `dim_waktu`: Berisi informasi waktu, khusus untuk tahun 2024.

## **B. Transformasi (T) – Transformasi Data**

Tahap transformasi data bertujuan untuk menyiapkan data agar siap dimuat ke dalam gudang data secara terstruktur. Ini dilakukan agar data sesuai dengan kebutuhan analisis dan struktur skema bintang. Proses ini melibatkan serangkaian langkah krusial, meliputi pembersihan data, normalisasi dan penyesuaian, serta validasi data.

### **1. Pembersihan Data**

Langkah-langkah pembersihan data meliputi:

- **Penghapusan Duplikasi:** Duplikasi data pada tabel `employees` akan dihapus untuk memastikan integritas dan keunikan setiap catatan.
- **Penanganan Nilai Kosong:** Nilai kosong (`null`) pada kolom `salary`, `overtime`, dan `longevity` akan diubah menjadi 0 untuk menghindari kesalahan perhitungan dan analisis.
- **Konversi Tipe Data:** Memastikan semua nilai numerik yang relevan (seperti `salary`, `overtime`, `longevity`, `total_salary`) memiliki tipe data `float` atau `integer` yang sesuai.

### **2. Normalisasi dan Penyesuaian**

Langkah-langkah normalisasi dan penyesuaian data meliputi:

- **Perhitungan `total_salary`:** Kolom `total_salary` akan dihitung berdasarkan formula: `base_salary + overtime_pay + longevity_pay`.
- **Pemetaan grade ke `id_grade`:** Nama grade dari tabel `employees` akan dipetakan ke `id_grade` melalui proses *join* dengan tabel `dim_grade`.
- **Pemetaan Nama Departemen:** Nama departemen dari tabel `employees` akan dipetakan ke `id_departemen` melalui proses *join* dengan tabel `dim_departemen`.

- Penugasan id\_waktu: Semua data akan ditetapkan dengan id\_waktu yang sesuai dari tabel dim\_waktu. Dalam konteks ini, semua data penggajian tahun 2024 akan diberi id\_waktu = 2024.

### **3. Validasi Data**

Langkah-langkah validasi data meliputi:

- Pengecekan Kunci Asing (Foreign Key): Memastikan validitas dan integritas hubungan antar tabel melalui pengecekan kunci asing (id\_grade, id\_departemen, id\_karyawan, id\_waktu).
- Konsistensi Gender: Memastikan konsistensi data gender antara tabel employees dan dim\_karyawan.

## **C. Pemuatan (L) – Pemuatan Data**

Tujuan utama dari tahap pemuatan adalah untuk memasukkan data yang telah dibersihkan dan ditransformasi ke dalam skema bintang di gudang data. Pada tahap ini, tabel fakta bernama Fact\_Penggajian akan dibuat.

### **1. Pembuatan Tabel Fact\_Penggajian**

Tabel Fact\_Penggajian akan memiliki atribut-atribut berikut:

- id\_karyawan
- id\_departemen
- id\_grade
- id\_waktu
- base\_salary
- overtime\_pay
- longevity\_pay
- total\_salary

### **2. Pemuatan Data**

Data yang telah melalui proses transformasi akan dimasukkan ke dalam tabel Fact\_Penggajian.

### 3. Penegasan Relasi

Memastikan bahwa relasi antara tabel fakta (Fact\_Penggajian) dan tabel dimensi (dim\_karyawan, dim\_departemen, dim\_grade, dim\_waktu) melalui kunci asing berjalan dengan benar dan sesuai dengan struktur skema bintang.

### 7. Analisis KPI (Key Performance Indicators)

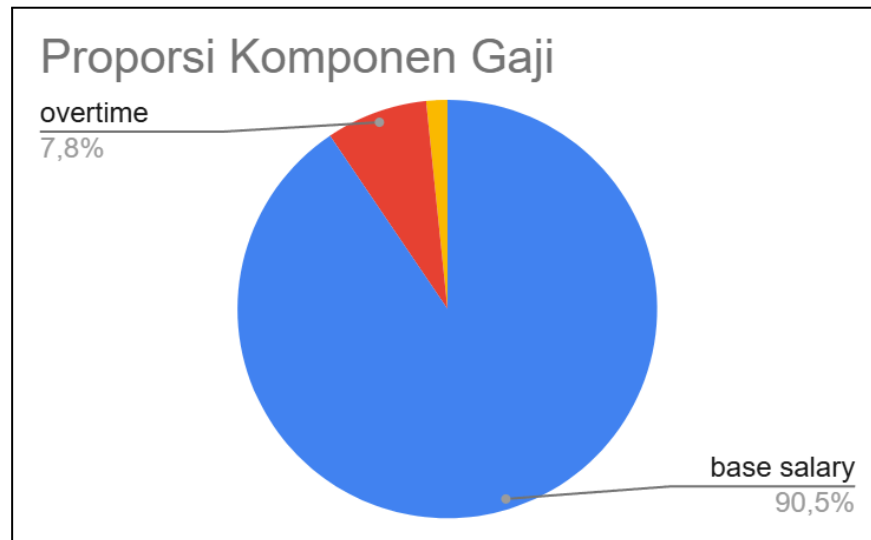
Key Result	KPI	Bobot	Target	Realisasi	Skor	Final Score
Proses ETL	Waktu total proses ETL	25	< 20 detik	± 8 detik	100	25
Konsistensi data	Semua foreign key pada tabel fakta tervalidasi (lookup berhasil)	20	Tidak ada error join	Semua lookup berhasil	100	20
Akurasi relasi data	Key lookup berhasil mengaitkan dengan dimensi yang sesuai	20	Tidak ada missing key	Tidak ditemukan missing key	100	20
Kelengkapan data warehouse	Semua entitas (dimensi dan fakta) berhasil dimuat tanpa error	15	Semua tabel dimuat	Semua data berhasil diinsert	100	15



Efektivitas struktur skema	Star schema diterapkan dengan baik (FK relasi & indexing)	20	Star schema lengkap	Dimensi & fakta terhubung rapi	100	20
Total		100	Total			100

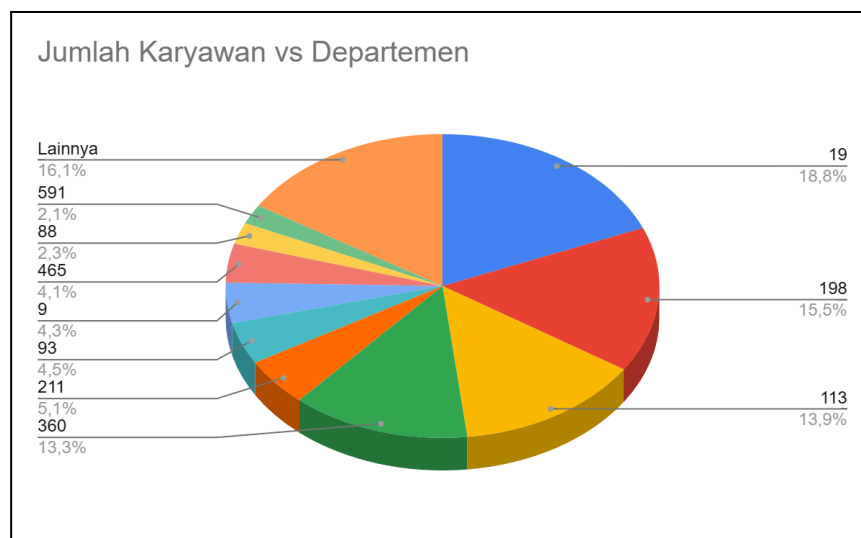
## 8. Visualisasi dan Analisis Fact\_Gaji\_Karyawan

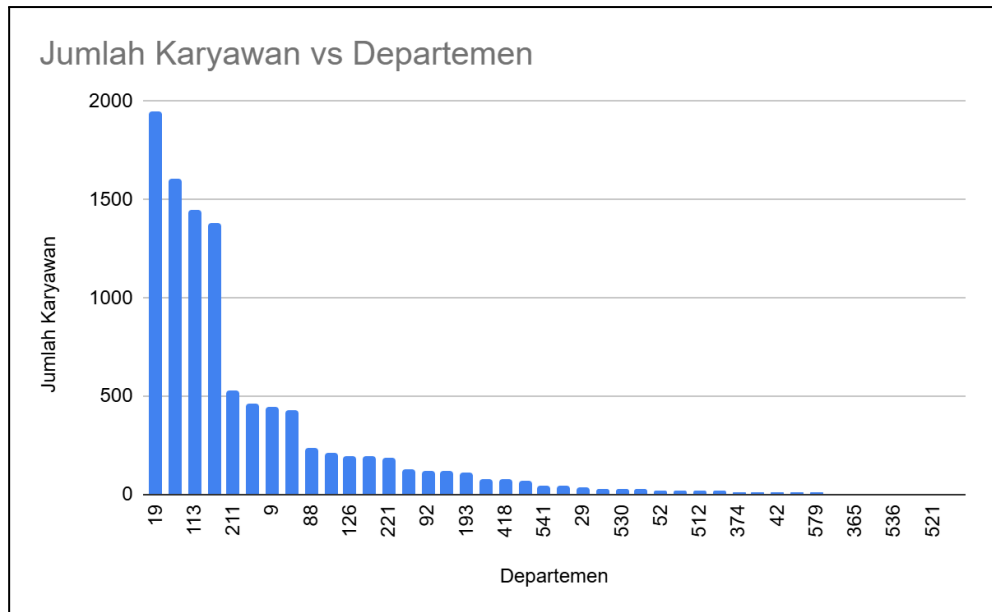
### a. Proporsi Komponen Gaji



Pie chart tersebut menunjukkan bahwa mayoritas komponen gaji berasal dari base salary sebesar 90,5%, sedangkan overtime hanya menyumbang 7,8%. Hal ini menunjukkan bahwa gaji pokok merupakan sumber utama pendapatan karyawan, sementara lembur hanya memberikan kontribusi kecil.

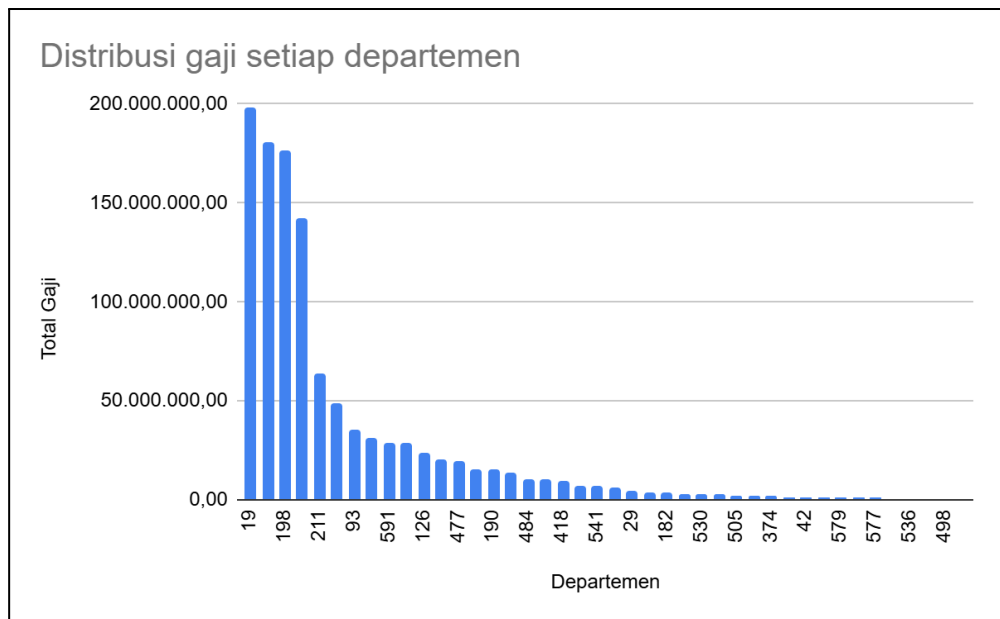
### b. Distribusi Karyawan Tiap Departemen





Jumlah karyawan tidak merata antar departemen, dengan sebagian besar terkonsentrasi pada tiga departemen utama (119, 113, dan 198), sementara sisanya terbagi kecil ke departemen lain.

### c. Distribusi Gaji Tiap Departemen



Grafik menunjukkan bahwa total gaji terbesar terdapat pada Departemen 119, 198, dan 113, selaras dengan jumlah karyawan terbanyak pada departemen tersebut.