

# Web Scraping Data Lowongan Kerja dari JobStreet Menggunakan Selenium & BeautifulSoup

*Fuad Hasyim*

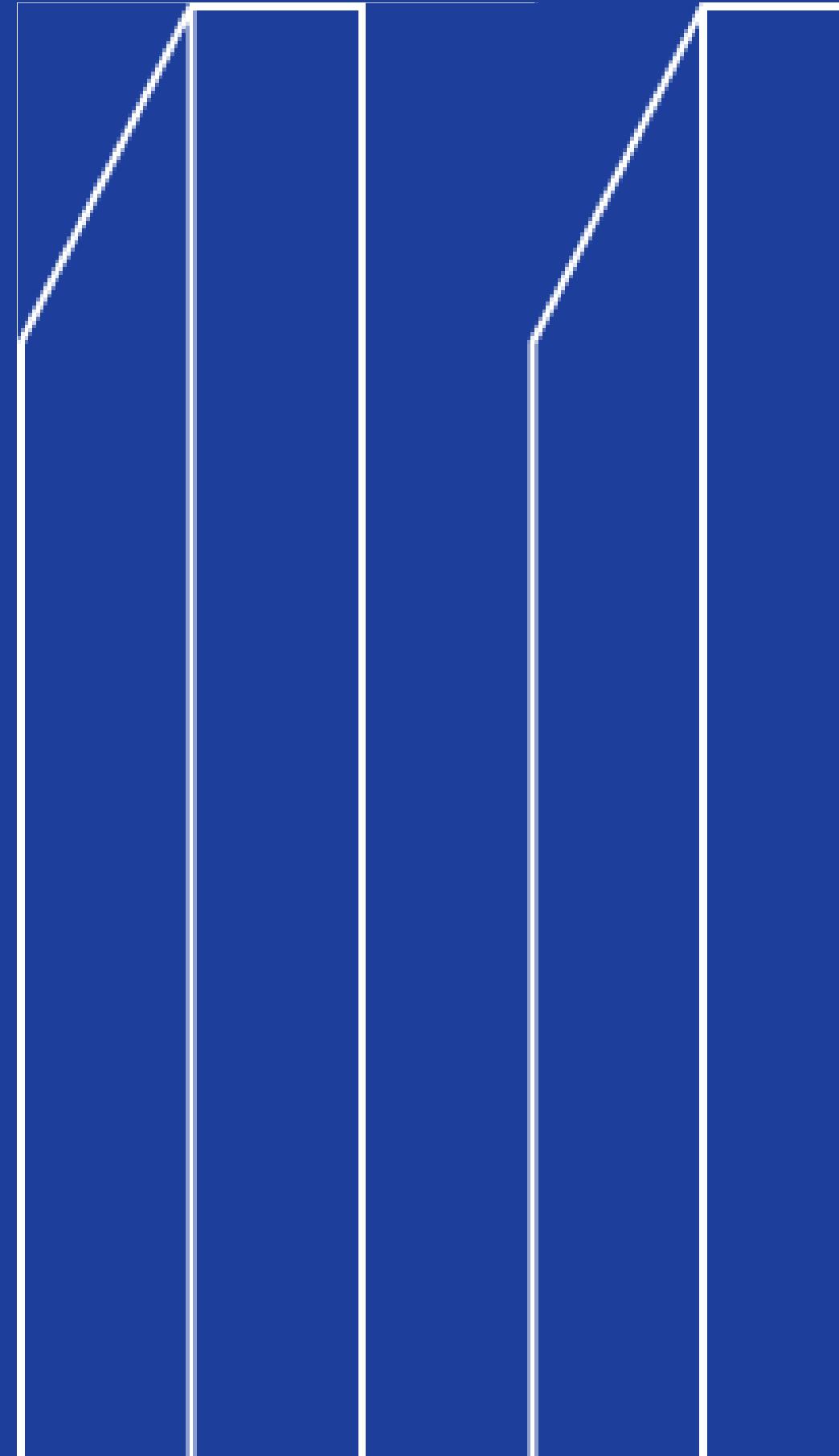
- WRITE TO US *Mengotomatisasi Pengumpulan Data Lowongan Kerja untuk Analisis Data*



# Tujuan Proyek

---

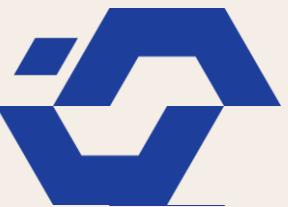
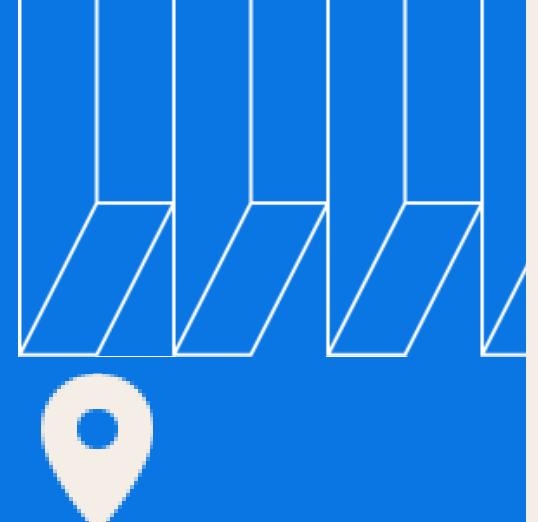
- Mengotomatisasi proses pengumpulan data lowongan kerja dari JobStreet Indonesia.
- Mengekstrak informasi penting seperti:
  - Judul Pekerjaan
  - Nama Perusahaan
  - Lokasi
  - Gaji
  - Tautan Lowongan
- Membangun dataset untuk analisis tren pekerjaan dan kebutuhan pasar kerja.



## Tools & Library

Library	Fungsi
Selenium	Mengontrol browser dan mengambil data dari website dinamis
BeautifulSoup	Memproses dan mengekstrak data dari HTML
Pandas / NumPy	Mengelola dan membersihkan data
Matplotlib / Seaborn	Membuat visualisasi data
WordCloud	Menampilkan frekuensi kata dari deskripsi pekerjaan

# Alur Kerja (Workflow)



1

*Inisialisasi Selenium WebDriver (Chrome).*



2

*Mengakses URL untuk tiap posisi pekerjaan*



3

*Melakukan scroll agar seluruh konten termuat.*



4

*Parsing HTML dengan BeautifulSoup*



5

*Mengambil detail pekerjaan (title, company, location, salary, link).*



6

*Menyimpan hasil ke dalam list.*



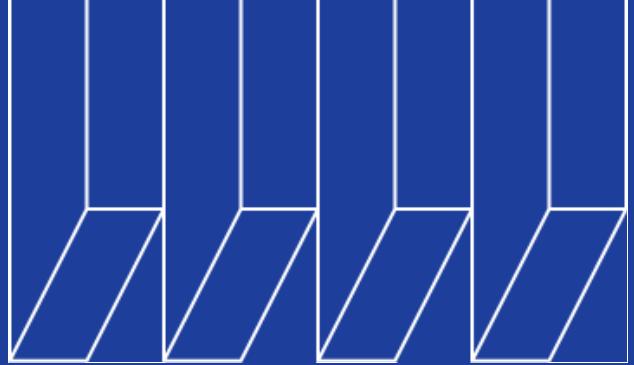
7

*Mengubah hasil ke DataFrame dan menyimpannya ke CSV.*

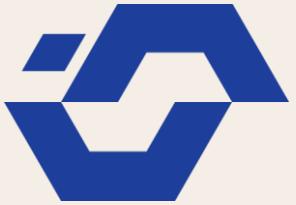


8

*Menyimpan ke database untuk analisis lanjutan dan visualisasikan*



# Logika Web Scraping

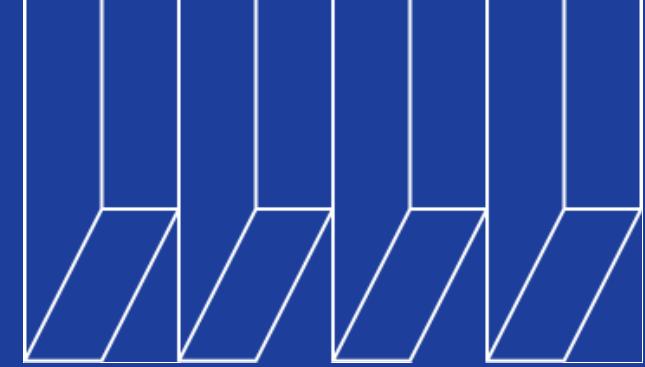


```
option = webdriver.ChromeOptions()  
driver = webdriver.Chrome(options=option)
```

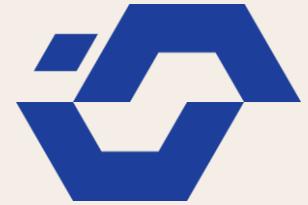
*Setup Selenium  
webdriver.Chrome()  
berperan seperti “browser  
otomatis”.*

```
for page in range(1, MAX_PAGES + 1):  
    url = f"https://id.jobstreet.com/id/data-scientist-jobs?page={page}"  
    driver.get(url)  
    time.sleep(3)  
    driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
```

*Mengambil data dari setiap  
halaman dan memastikan semua  
card lowongan termuat sebelum  
diproses.*



# Ekstraksi Informasi Pekerjaan



```
soup = BeautifulSoup(driver.page_source, 'html.parser')
jobs = soup.find_all('article')

for job in jobs:
    title = job.find('a', {'data-automation': 'jobTitle'})
    company = job.find('a', {'data-automation': 'jobCompany'})
    salary = job.find('span', {'data-automation': 'jobSalary'})
```

*Parsing HTML dengan BeautifulSoup*

*Mengambil data dari setiap halaman dan memastikan semua kartu lowongan termuat sebelum diproses.*

```
job_info = {
    'title': title.text.strip(),
    'company': company.text.strip(),
    'location': location.text.strip(),
    'salary': salary.text.strip() if salary else 'NaN',
    'link': link
}
results.append(job_info)
```

*Semua hasil disimpan dalam list of dictionaries, lalu diubah menjadi DataFrame Pandas.*



# Ekspor ke CSV

```
df_job = pd.DataFrame(results)
df_job.to_csv('jobstreet_data_engineer_all_pages.csv', index=False)
print(f"Total lowongan terkumpul: {len(df_job)}")
```

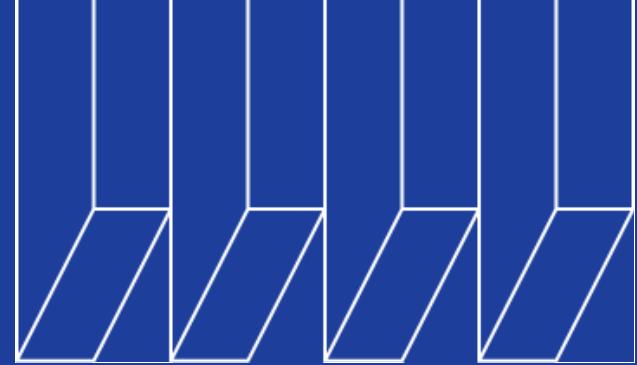
In [39]:	# Convert to DataFrame df_job = pd.DataFrame(results) df_job.head()				
Out[39]:	title	company	location	salary	link
	Database Finance Administrator	KSO Sucofindo-Surveyor Indonesia	Jakarta Selatan	NaN	<a href="https://id.jobstreet.com/id/job/86382708?type=standard&amp;ref=search-standalone&amp;origin=cardTitle#sol=8b6c3ec88ab4bac7d3372d0d8e4f6f19a8577d0f">https://id.jobstreet.com/id/job/86382708?type=standard&amp;ref=search-standalone&amp;origin=cardTitle#sol=8b6c3ec88ab4bac7d3372d0d8e4f6f19a8577d0f</a>
	Senior Database Administrator Engineer - Engineering Platform	PT GOTO GOJEK TOKOPEDIA TBK	Jakarta Raya	NaN	<a href="https://id.jobstreet.com/id/job/86494164?type=standard&amp;ref=search-standalone&amp;origin=cardTitle#sol=9802c8e825f33fadea5bb7aac9e0c9cb8df20038">https://id.jobstreet.com/id/job/86494164?type=standard&amp;ref=search-standalone&amp;origin=cardTitle#sol=9802c8e825f33fadea5bb7aac9e0c9cb8df20038</a>
	SENIOR DATABASE ENGINEER	PT JAYA AGUNG TEKNOLOGI	Jakarta Utara	NaN	<a href="https://id.jobstreet.com/id/job/86535509?type=standard&amp;ref=search-standalone&amp;origin=cardTitle#sol=a8c4d2d4973d3504bd66916030c7bee2d743fe9c">https://id.jobstreet.com/id/job/86535509?type=standard&amp;ref=search-standalone&amp;origin=cardTitle#sol=a8c4d2d4973d3504bd66916030c7bee2d743fe9c</a>

Menyimpan hasil scraping ke file CSV lalu gabungkan semua hasil web scrapping menjadi 1, agar bisa dianalisis lebih lanjut atau diunggah ke database.

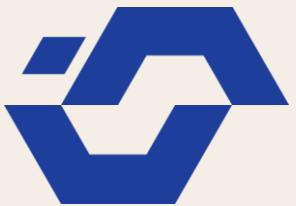
## Contoh Hasil

- Data Posisi yang Diambil
- Business Analyst — 30 halaman
- Data Scientist — 9 halaman
- Data Engineer — 30 halaman
- Database Administrator — 4 halaman

Total: ±70 halaman lowongan dikumpulkan secara otomatis



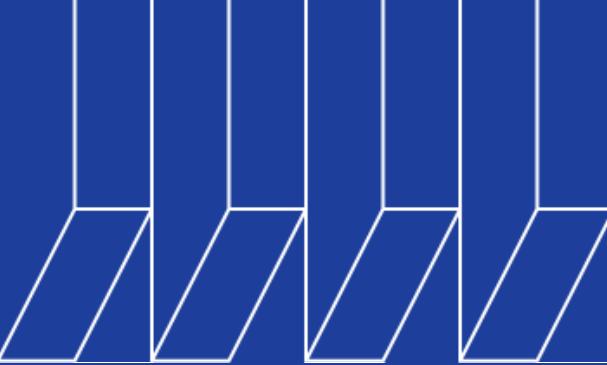
# Data Cleaning dan selection



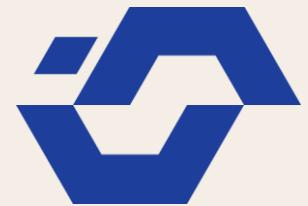
1. *Handling Duplicate*
2. *Kategorisasi dan Menghapus Pekerjaan yang Tidak Relevan*

Meskipun sudah menggunakan kata kunci "Data Scientist, Data Engineer, Data Base Administrator dan Bussines analyst", ternyata tidak semua pekerjaan yang muncul benar-benar terkait dengan keempat hal tersebut.

```
ds_list = ['data scientist', 'data science', 'machine learning','artificial intelligence', 'ai/ml', 'ml engineer','ai engineer']
de_list = ['data engineer', 'big data engineer', 'etl engineer','data pipeline', 'data warehouse', 'hadoop', 'spark engineer']
dba_list = ['database administrator', 'dba', 'sql server','oracle dba', 'mysql administrator', 'postgresql administrator','db administrator']
ba_list = ['business analyst', 'bisnis analis', 'system analyst','business intelligence','bi analyst']
length = len(df['title'])
df_title = df['title']
```



# Ekploratory Data



DBeaver 25.1.1 - <postgres> Script-10

File Edit Navigate Search SQL Editor Database Window Help

Database Navigator X

Filter connections by name

> DBeaver Sample Database (SQLite)

> dibimbing ep-flat-mouse-a5xmvitt.us-east-2.aw

postgres localhost:5432

Databases

Jobstreet

Schemas

public

Tables

all\_jobs\_clean 632K

Foreign Tables

Views

Materialized Views

Indexes

Functions

Sequences

Data types

Aggregate functions

Event Triggers

Extensions

Storage

System Info

Roles

dibimbing

dvdrental

postgres

Administrator

System Info

SQL Commit Rollback Auto postgres public@Jobstreet

example\_client\_b... example\_client\_tr... handson\_rider\_stat... handson\_rider\_team \*<postgres> Script... \*<postgres> Script...

select title from all\_jobs\_clean ajc;

-- Top 5 lokasi dengan lowongan terbanyak

```
SELECT "location", COUNT(*) AS jumlah_lowongan
FROM all_jobs_clean ajc
GROUP BY "location"
ORDER BY jumlah_lowongan DESC
LIMIT 5;
```

-- Jumlah lowongan per perusahaan

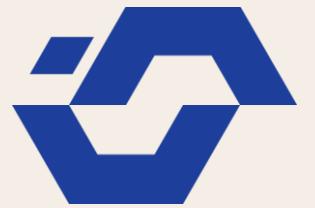
```
SELECT company, COUNT(*) AS jumlah_lowongan
FROM all_jobs_clean
GROUP BY company
ORDER BY jumlah_lowongan DESC
LIMIT 10;
```

all\_jobs\_clean 1

select title from all\_jobs\_clean ajc Enter a SQL expression to filter results (use Ctrl+Space)

Record	A-Z title
1	business analyst
2	business analyst
3	business analyst
4	it business analyst (ba)
5	business analyst
6	business analyst staff
7	corporate business strategic analyst
8	it bpm business analyst

Refresh Save Cancel Export data 200 200+ WIB en Writable Smart Insert 1:38:37



# Ekploratory Data

## TOP 5 job by Location

DBeaver 25.1.1 - <postgres> Script-10

File Edit Navigate Search SQL Editor Database Window Help

Rollback Auto postgres public@Jobstreet

Database Navigator X

Filter connections by name

> DBeaver Sample Database (SQLite)

> dibimbing ep-flat-mouse-a5xmvitt.us-east-2.aw

< postgres localhost:5432

Databases Jobstreet Schemas

public

- Tables (all\_jobs\_clean, Foreign Tables, Views, Materialized Views, Indexes, Functions, Sequences, Data types, Aggregate functions)
- Event Triggers, Extensions, Storage, System Info, Roles
- dibimbing, dvdrental, postgres
- Administer, System Info

SQL Editor

```
select title from all_jobs_clean ajc;
-- Top 5 lokasi dengan lowongan terbanyak
SELECT "location", COUNT(*) AS jumlah_lowongan
FROM all_jobs_clean ajc
GROUP by "location"
ORDER BY jumlah_lowongan DESC
LIMIT 5;
```

--Jumlah lowongan per perusahaan

```
SELECT company, COUNT(*) AS jumlah_lowongan
FROM all_jobs_clean
GROUP BY company
ORDER BY jumlah_lowongan DESC
LIMIT 10;
```

all\_jobs\_clean 1 X

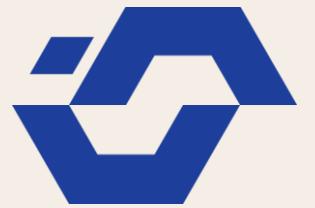
SELECT "location", COUNT(\*) AS jumlah\_lowongan Enter a SQL expression to filter results (use Ctrl+Space)

location	jumlah_lowongan
Jakarta Raya	699
Jakarta Selatan	225
Jakarta Pusat	105
Jakarta Barat	104
Jakarta Utara	90

Grid Text Record

Refresh Save Cancel Export data 200 5

WIB en Writable Smart Insert 8:9:219



# Ekploratory Data

## Job Title by Company

DBeaver 25.1.1 - <postgres> Script-10

File Edit Navigate Search SQL Editor Database Window Help

SQL Commit Rollback Auto postgres public@Jobstreet

Database Navigator

example\_client\_b... example\_client\_tr... handson\_rider\_stat... handson\_rider\_team \*<postgres> Script... \*<postgres> Script...

Filter connections by name

DBeaver Sample Database (SQLite)

dibimb... ep-flat-mouse-a5xmvitt.us-east-2.aw

postgres localhost:5432

Databases

Jobstreet

Schemas

public

- Tables (632K)
- Foreign Tables
- Views
- Materialized Views
- Indexes
- Functions
- Sequences
- Data types
- Aggregate functions
- Event Triggers
- Extensions
- Storage
- System Info
- Roles

dibimb... dvrental postgres Administer System Info

```
SELECT company, COUNT(*) AS jumlah_lowongan
FROM all_jobs_clean
GROUP BY company
ORDER BY jumlah_lowongan DESC
LIMIT 10;
```

--Cek range (min & max) salary per posisi

```
SELECT title,
       MIN(salary::numeric) AS min_salary,
       MAX(salary::numeric) AS max_salary
  from all_jobs_clean ajc
 WHERE salary IS NOT NULL
```

all\_jobs\_clean 1

SELECT company, COUNT(\*) AS jumlah\_lowongan Enter a SQL expression to filter results (use Ctrl+Space)

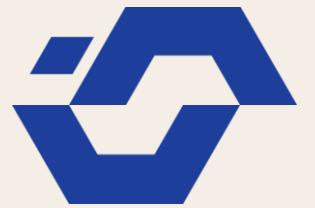
	A-Z company	123 jumlah_lowongan
1	PT Trinusa Travelindo	50
2	RGF HR Agent Indonesia	35
3	PT. SIGMA GLOBAL TEKNOLOGI (SIGMATECH)	30
4	PT Amarta Mikro Fintek (Jakarta)	30
5	PT Solusi Transportasi Indonesia	25
6	PT SMART,Tbk	23
7	PT. Metrodata Electronics, Tbk	23
8	DNC Indonesia	21

Refresh Save Cancel Export data 200 10

WIB en Writable Smart Insert 15 : 10 : 381

The screenshot shows the DBeaver interface with a database connection to 'postgres' on 'public@Jobstreet'. In the SQL Editor, there are two queries. The first query retrieves the count of job listings for each company, ordered by the count in descending order with a limit of 10. The second query checks the minimum and maximum salary for each job position. Below the SQL Editor, the results of the first query are displayed in a grid. The grid has columns for rank, company name, and the count of job listings ('jumlah\_lowongan'). The companies listed are PT Trinusa Travelindo, RGF HR Agent Indonesia, PT. SIGMA GLOBAL TEKNOLOGI (SIGMATECH), PT Amarta Mikro Fintek (Jakarta), PT Solusi Transportasi Indonesia, PT SMART,Tbk, and PT. Metrodata Electronics, Tbk.

Rank	Company	Jumlah Lowongan
1	PT Trinusa Travelindo	50
2	RGF HR Agent Indonesia	35
3	PT. SIGMA GLOBAL TEKNOLOGI (SIGMATECH)	30
4	PT Amarta Mikro Fintek (Jakarta)	30
5	PT Solusi Transportasi Indonesia	25
6	PT SMART,Tbk	23
7	PT. Metrodata Electronics, Tbk	23
8	DNC Indonesia	21



# Ekploratory Data

## Job by Title Category

DBeaver 25.1.1 - <postgres> Script-10

File Edit Navigate Search SQL Editor Database Window Help

SQL Commit Rollback Auto postgres public@Jobstreet

Database Navigator

Filter connections by name

DBeaver Sample Database (SQLite)

dibimbing ep-flat-mouse-a5xmvitt.us-east-2.aw

postgres localhost:5432

Databases

Jobstreet

Schemas

public

- Tables (all\_jobs\_clean, 640K)
- Foreign Tables
- Views
- Materialized Views
- Indexes
- Functions
- Sequences
- Data types
- Aggregate functions
- Event Triggers
- Extensions
- Storage
- System Info
- Roles

dibimbing

dvdrental

postgres

Administer

System Info

```
--Jumlah lowongan per perusahaan
SELECT company, COUNT(*) AS jumlah_lowongan
FROM all_jobs_clean
GROUP BY company
ORDER BY jumlah_lowongan DESC
LIMIT 10;

--SELECT title_category, COUNT(*) AS total_jobs
FROM all_jobs_clean ajc
WHERE title_category IN ('Data Science', 'Data Engineer', 'Data Base Administrator', 'Business Analyst')
GROUP BY title_category
ORDER BY total_jobs DESC;
```

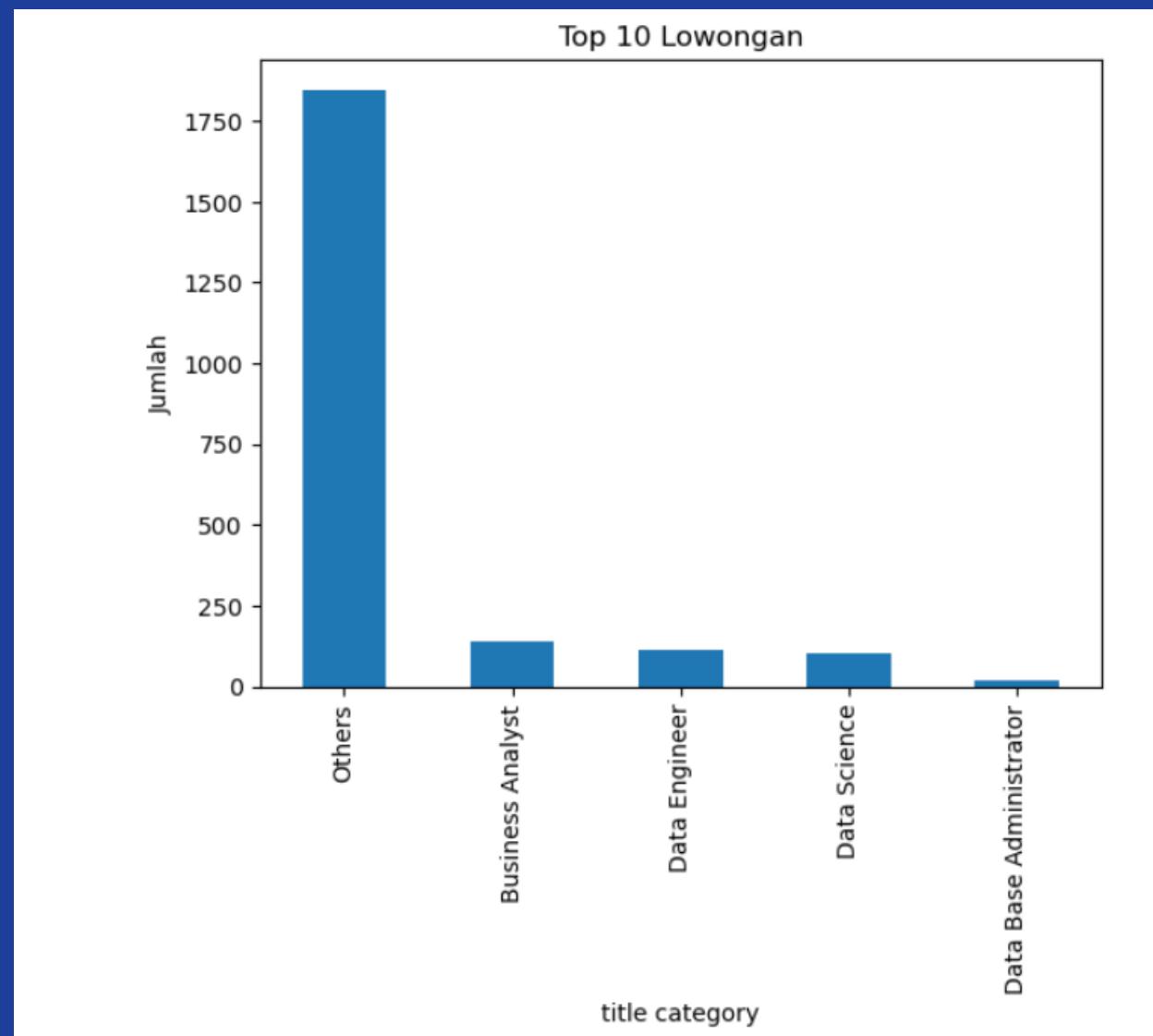
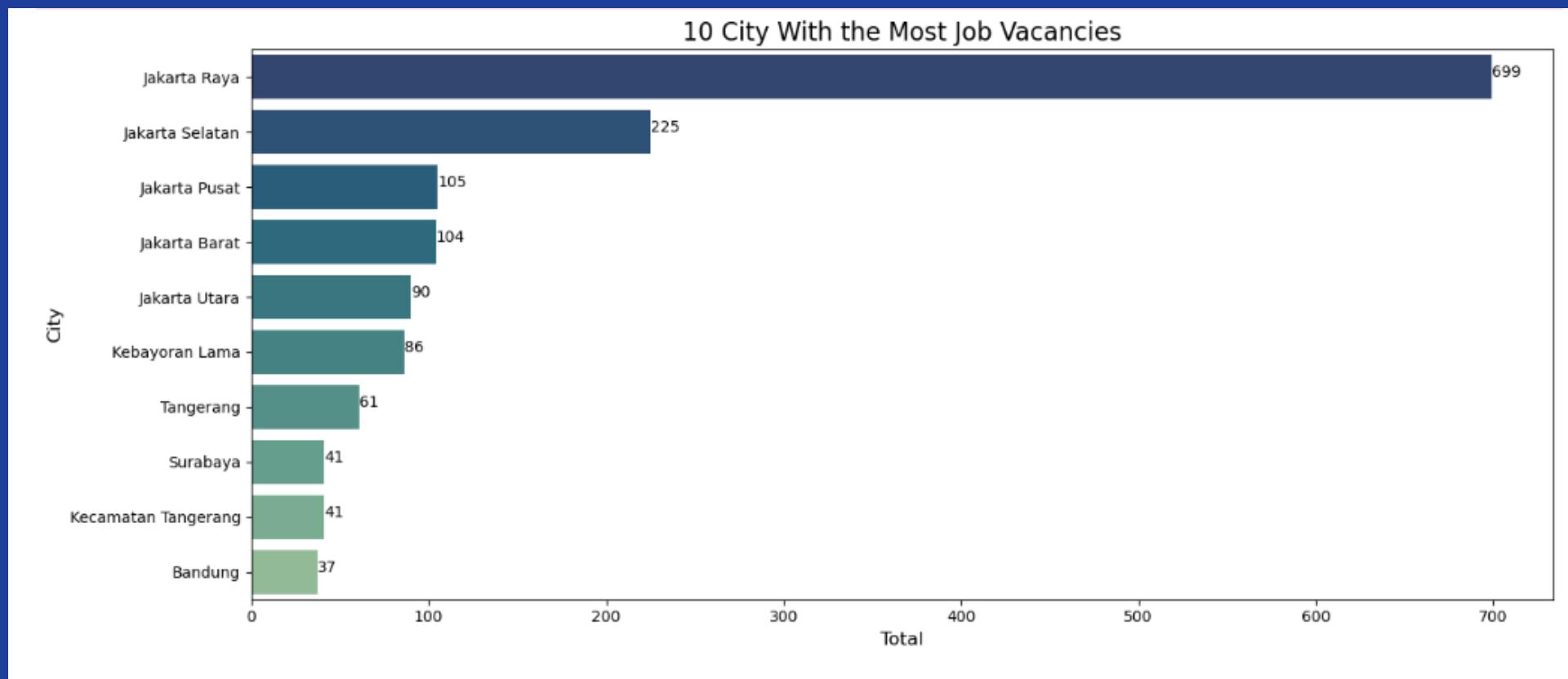
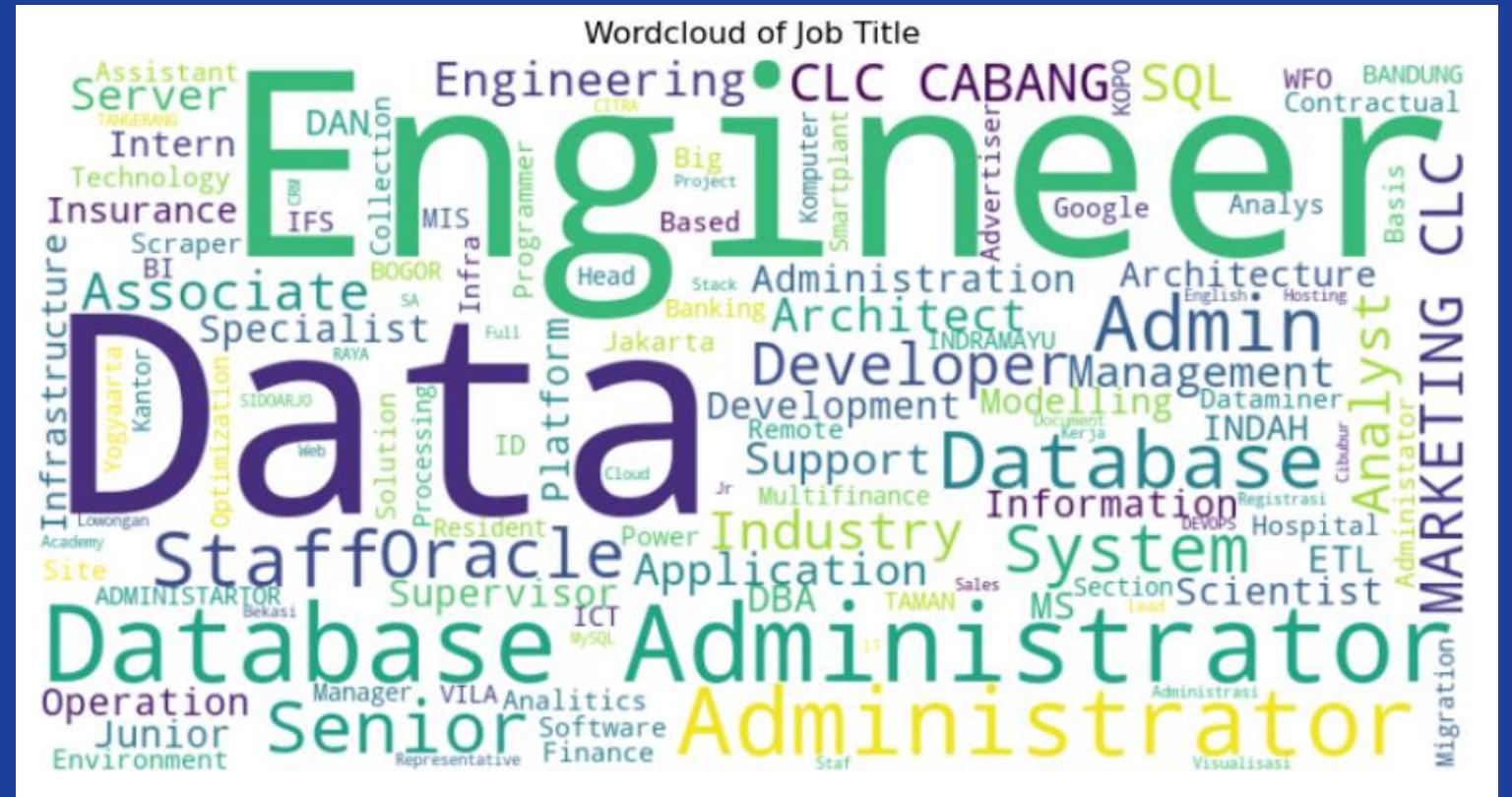
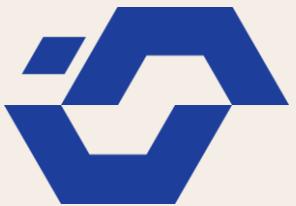
all\_jobs\_clean 1

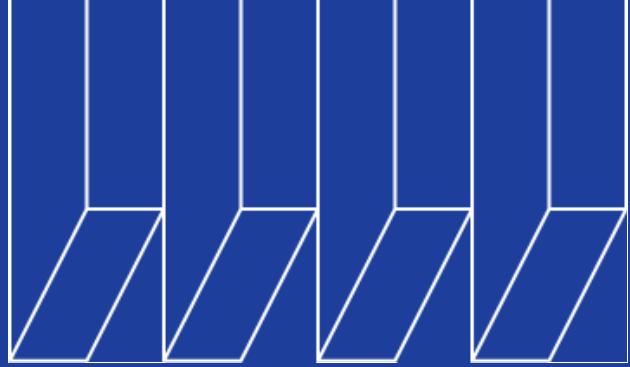
	A-Z title_category	123 total_jobs
1	Business Analyst	142
2	Data Engineer	112
3	Data Science	103
4	Data Base Administrator	22

Refresh Save Cancel Export data 200 4

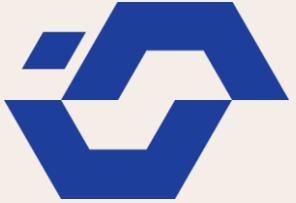
WIB en Writable Smart Insert 21:26:614

# Vizualization





# Conclusion



## ***Top 5 job Data by location***

- *Jakarta Raya* : 699 job
- *Jakarta Selatan* : 225 job
- *Jakarta Pusat* : 105 job
- *Jakarta Barat* : 104 job
- *Jakarta Utara* : 90 job

## ***Top 5 job Data by Company***

- *PT Trinusa Travelindo* : 50 job
- *RGF HR Indonesia* : 35 job
- *PT Sigma Global Teknologi (SIGMATECH)*: 30 job
- *PT Amartha mikro Fintech* : 30 job
- *PT Solusi Transportasi Indonesia*: 25 job

## ***Top 5 job Data by Title Job***

- *Data Analys* : 699 job
- *Data Engineer*: 225 job
- *Data Science* : 105 job
- *Database Administrator*: 104 job



# Terima kasih.



*Silahkan bertanya*

*Tapi, jangan yang susah” ya!!!!*

