

Project Manajemen Data Statistika



VISUALIZING CAREER PATHWAYS

FROM WEB SCRAPPING DATA

Rahmi Anadra



rahmiandr



rahmianadra



rahmianadraa_



TABLE OF CONTENT



INTRODUCTION

COLLECT DATA

DATA PORCESSING

ANALYZING DATA

CONCLUSION

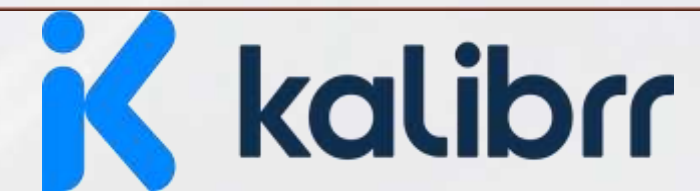
INSIGHT



INTRODUCTION

Kalibrr merupakan sumber informasi yang dapat diandalkan untuk mencari dan menilai peluang karier dalam industri, serta digunakan oleh pencari kerja dan profesional sumber daya manusia untuk menemukan dan menilai potensi perusahaan dan posisi pekerjaan yang sesuai dengan kebutuhan mereka.

Pada project kali ini dilakukan scraping pada website <https://www.kalibrr.id/>. Kalibrr adalah platform yang menyediakan informasi tentang lowongan kerja, perusahaan, dan kesempatan karier lainnya. Doyek ini bertujuan untuk mengumpulkan data dari Kalibrr menggunakan teknik web scraping, yang nantinya akan digunakan untuk analisis dan visualisasi data terkait informasi lowongan kerja dan perusahaan yang tersedia. Proyek ini akan menggunakan teknik web scraping untuk mengumpulkan beberapa informasi yang relevan dari website Kalibrr. Setelah data berhasil dikumpulkan, langkah selanjutnya adalah melakukan analisis data dan membuat visualisasi yang informatif, seperti grafik untuk menggambarkan tren lowongan kerja berdasarkan sektor industri, tingkat pendidikan yang dibutuhkan, atau lokasi pekerjaan yang paling banyak dicari. Visualisasi ini akan membantu dalam memahami dinamika pasar kerja dan memberikan wawasan berharga bagi para pencari kerja dan perusahaan.





COLLECT DATA

● TOOLS



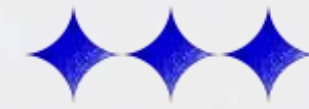
Digunakan untuk menyimpan dan mengelola data hasil scraping secara aman dan terstruktur di cloud.



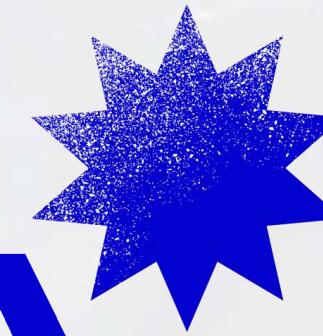
Digunakan untuk menulis, mengembangkan, dan menjalankan skrip R yang melakukan scraping, pembersihan, dan analisis data dari Kalibrr.



Digunakan untuk mengotomatisasi scraping data setiap hari pada jam 01.00 dengan menjalankan skrip R secara terjadwal.



COLLECT DATA

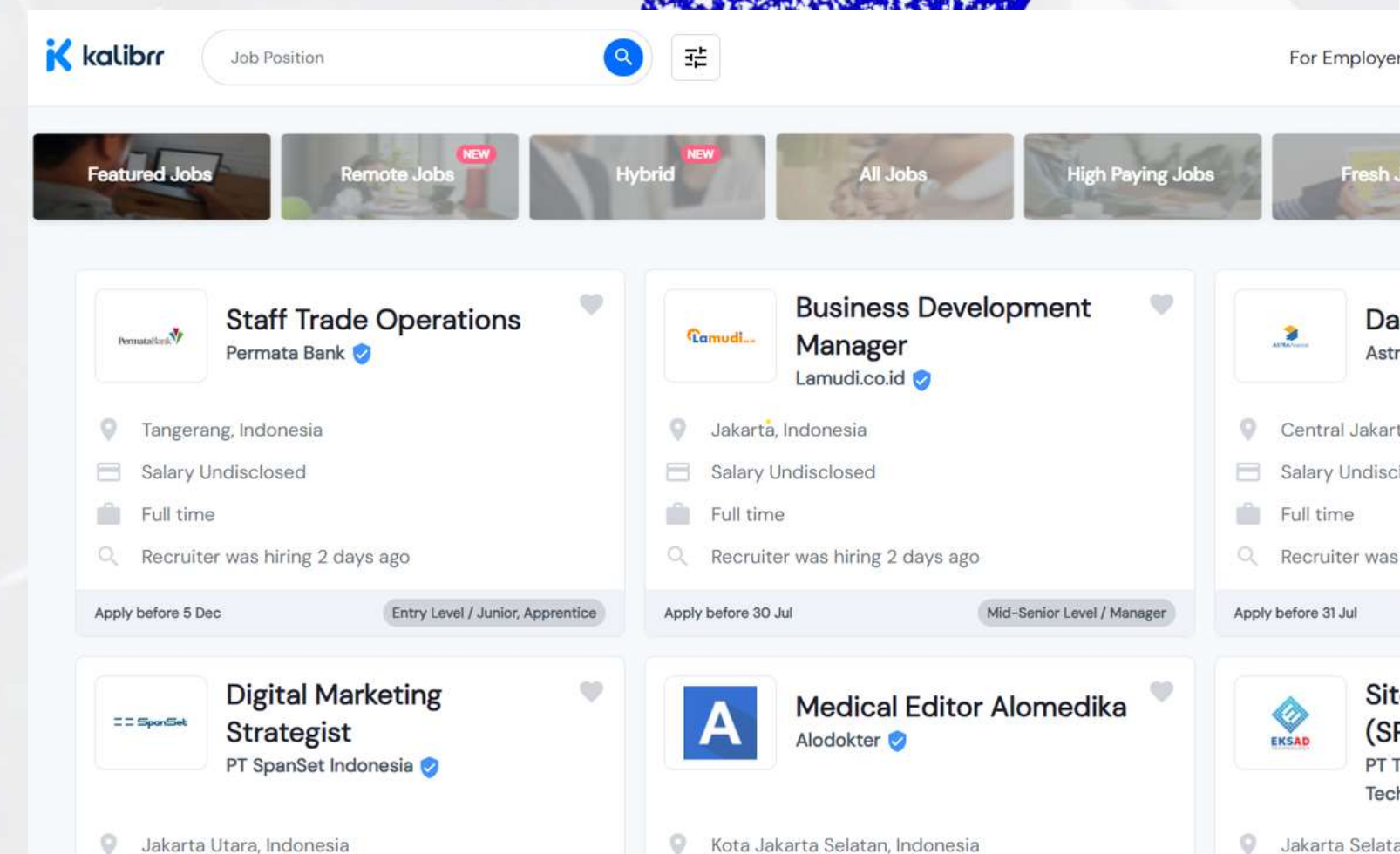


● SCRAPING DATA

Dalam hal ini, data yang akan dilakukan scraping berkaitan dengan :

- Posisi: Posisi atau jabatan pekerjaan yang ditawarkan.
- Perusahaan: meliputi kenyamanan lounge, kebersihan, katering makanan, toilet, layanan staf, dll.
- Lokasi: Lokasi tempat kerja.
- Gaji: Informasi tentang gaji yang ditawarkan untuk posisi tersebut.
- Jenis: Jenis pekerjaan, seperti full time, part time, internship.
- Batas: Tanggal batas pengajuan berkas lamaran
- Level: Tingkat pengalaman yang dibutuhkan untuk posisi tersebut, seperti Entry Level / Junior, Apprentice.

Scraping dilakukan dengan software R menggunakan packages rvest dan tidyverse sebagai cleaning tools pada data yang telah diambil. Data yang telah discraping dari website akan disimpan dalam MongoDB Atlas dan dijadwalkan setiap hari di jam 01.00 akan diambil secara acak lima data lowongan pekerjaan. Scraping terjadwal dilakukan dengan membuat workflow di Github Action.



rahmianadr



rahmianadra



rahmianadraa

Project Manajemen Data Statistika

COLLECT DATA

HASIL SCRAPPING

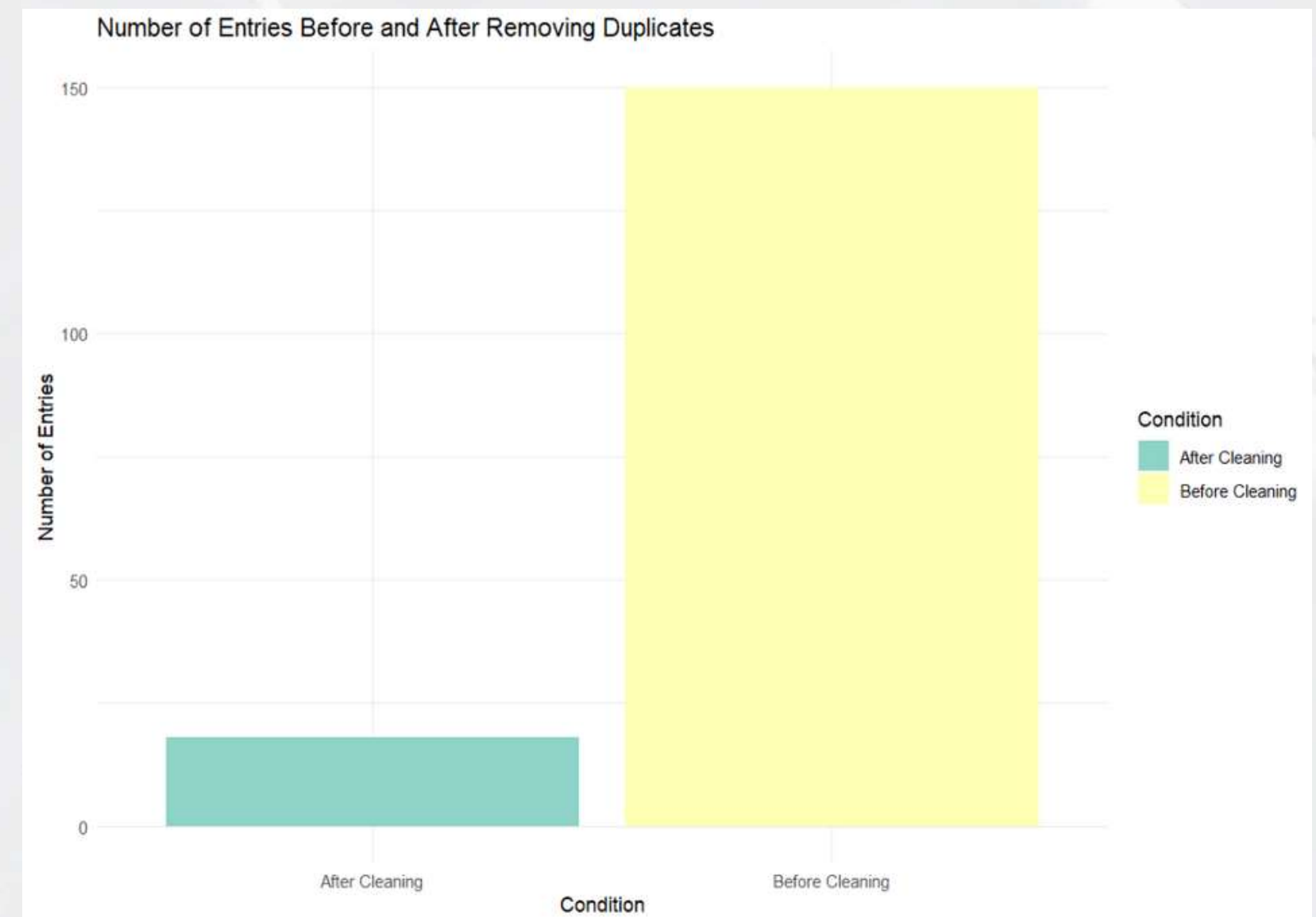
The screenshot displays the MongoDB Atlas web interface. The top navigation bar includes the Atlas logo, organization name 'Rahmi's Org ...', and links for 'Access Manager' and 'Billing'. The main navigation menu on the left lists various services like 'Overview', 'DEPLOYMENT', 'Database', 'Data Lake', 'SERVICES', 'Device & Edge Sync', 'Triggers', 'Data API', 'Data Federation', 'Atlas Search', 'Stream Processing', 'Migration', 'SECURITY', 'Quickstart', 'Backup', 'Database Access', 'Network Access', and 'Advanced'. The 'Database' section is expanded, showing a list of databases under 'MDS_Scrape', including 'kalibrr'. The 'kalibrr' database is selected, and its details are shown on the right. The details include 'MDS_Scrape.kalibrr' (CLUSTERED), storage size (44KB), logical data size (39.2KB), total documents (150), and indexes total size (0B). Below this, there are tabs for 'Find', 'Indexes', 'Schema Anti-Patterns', 'Aggregation', and 'Search Indexes'. The 'Find' tab is active, showing a query filter input field with the text 'Type a query: { field: 'value' }'. Below the filter, the query results are displayed, showing a single document with the following fields: '_id: ObjectId('665984ccc50e47cfec03aec1')', 'time_scraped: 2024-05-31T08:05:29.308+00:00', 'posisi: "Officer Development Program (ODP) Batch 2024"', 'perusahaan: "Bank Negara Indonesia"', 'lokasi: "Jakarta Pusat, Indonesia"', 'gaji: "Salary Undisclosed"', 'jenis: "FULL_TIME"', 'batas: "Apply before 29 Jun"', and 'level: "Entry Level / Junior, Apprentice"'. The results are paginated, showing '1-20 of many results'.

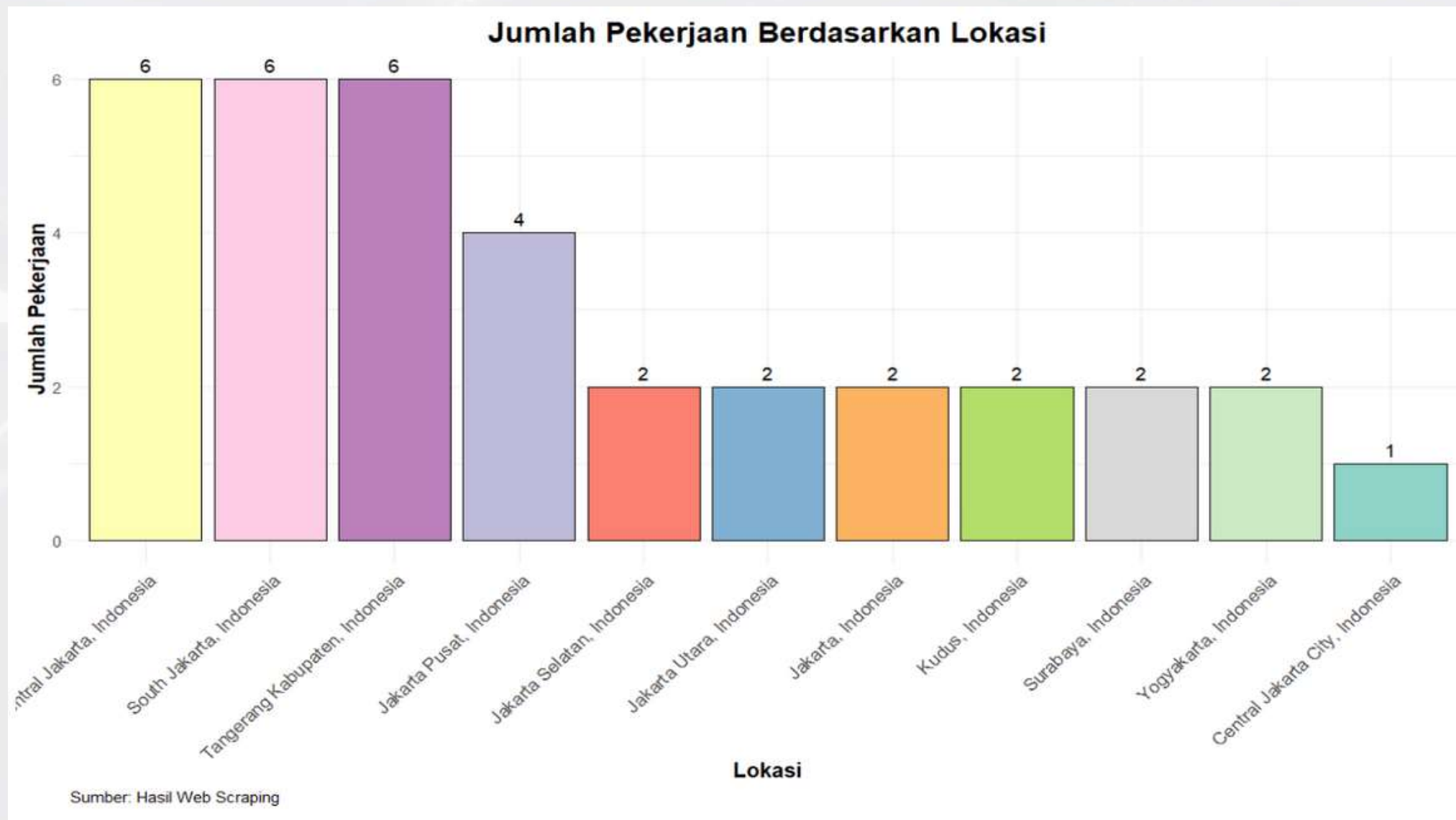
DATA PROCESSING

DELETE DUPLIKACE DATA

Menghapus data duplikat memastikan analisis dan visualisasi yang dihasilkan berdasarkan data yang bersih, akurat, dan dapat diandalkan, sehingga menghasilkan wawasan yang valid dan berguna.

Data yang awalnya berjumlah 150 setelah dilakukan penghapusan data duplikat menjadi 18 data. Hal ini terjadi karena banyak data yang terambil sama setiap harinya atau tidak adanya update lowongan pekerjaan per hari.





JUMLAH PEKERJAAN BERDASARKAN LOKASI

Sebagian besar pekerjaan dalam dataset tersebar di Jakarta dan sekitarnya, terutama di "Central Jakarta", "South Jakarta", dan "Tangerang Kabupaten". Hal ini menunjukkan bahwa peluang pekerjaan lebih banyak di Jakarta, pusat ekonomi Indonesia. "Kudus", "Surabaya", dan "Yogyakarta" memiliki sedikit pekerjaan, menunjukkan peluang kerja lebih terbatas di daerah-daerah tersebut. Informasi ini penting bagi pencari kerja untuk memilih lokasi dengan peluang kerja lebih tinggi.

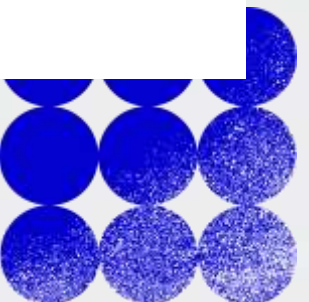
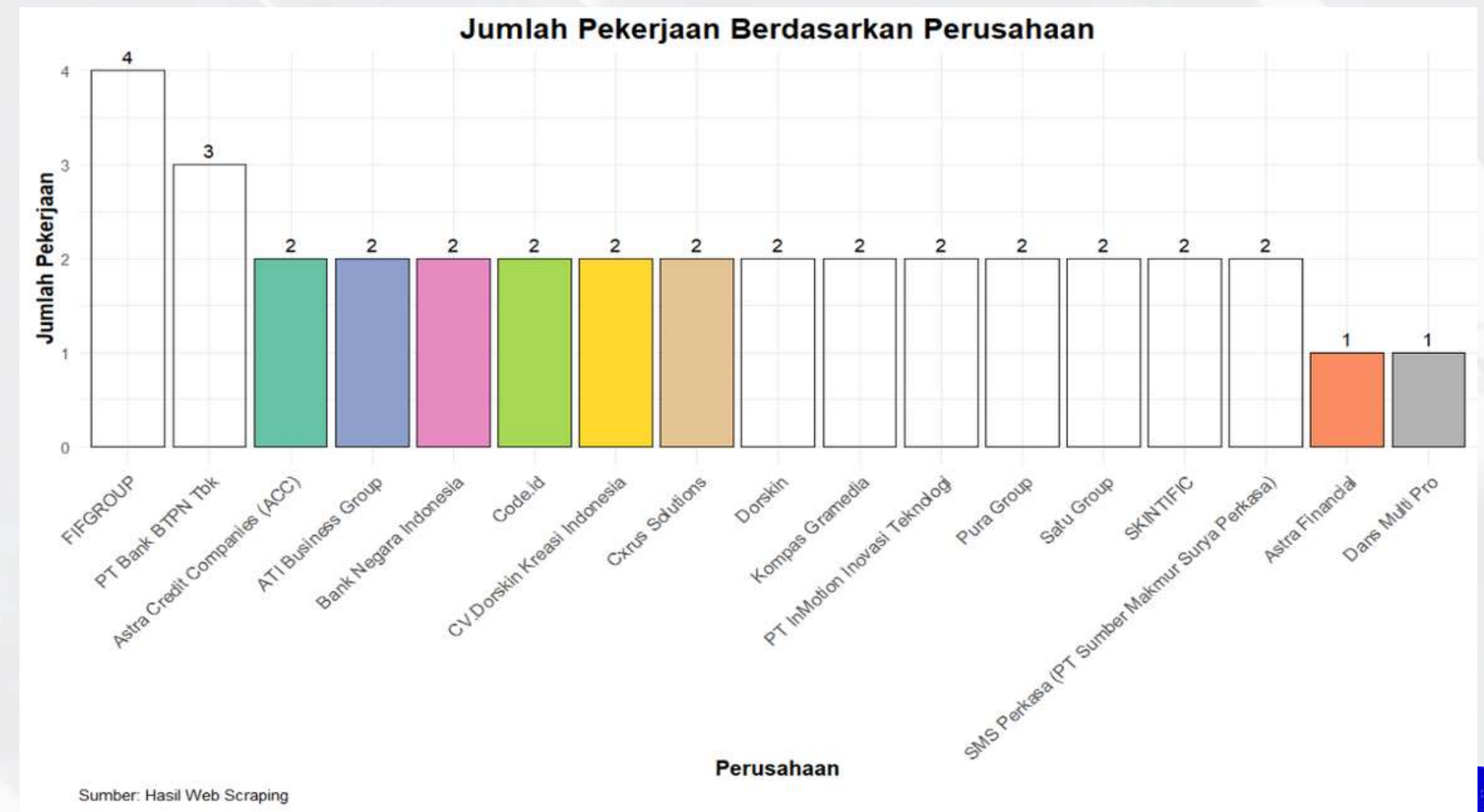
ANALYZING DATA



ANALYZING DATA

JUMLAH PEKERJAAN BERDASARKAN PERUSAHAAN

FIFGROUP mendominasi dengan dua lowongan pekerjaan, sedangkan perusahaan lain seperti Astra Credit Companies (ACC), Astra Financial, dan Bank Negara Indonesia masing-masing menawarkan satu pekerjaan. Hal ini menunjukkan bahwa meskipun ada keragaman perusahaan yang membuka lowongan, mayoritas perusahaan hanya menawarkan satu pekerjaan. Kesimpulannya, peluang kerja tersebar merata di berbagai perusahaan, namun jumlah lowongan per perusahaan cenderung rendah.

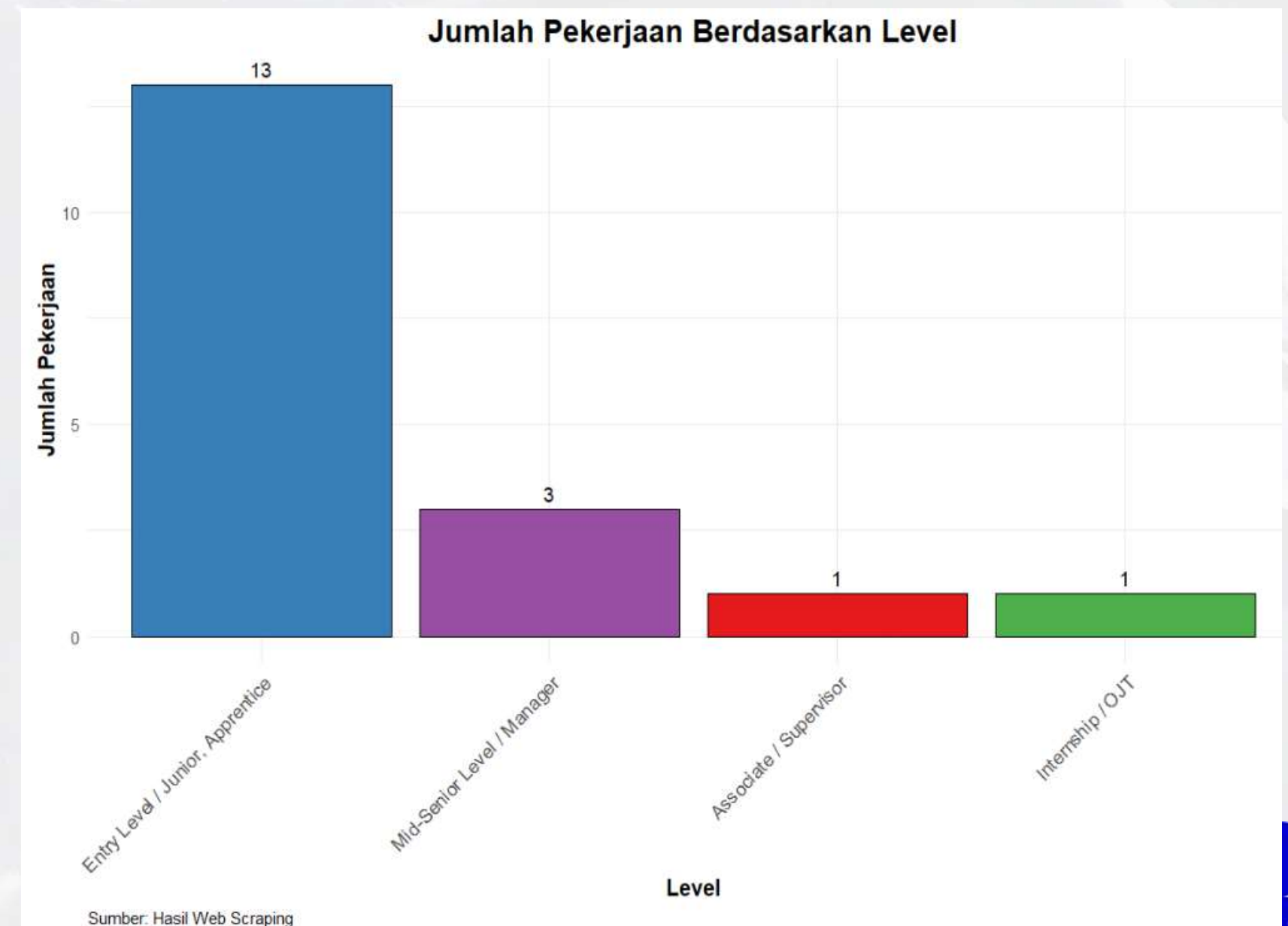




ANALYZING DATA

JUMLAH PEKERJAAN BERDASARKAN LEVEL

Mayoritas pekerjaan yang tersedia adalah untuk level Entry Level/Junior/Apprentice dengan jumlah 13 pekerjaan. Pekerjaan untuk level Mid-Senior/Manager tersedia 3 lowongan, sedangkan Associate/Supervisor dan Internship/OJT masing-masing memiliki 1 lowongan. Kesimpulannya, lowongan pekerjaan di Kalibrr didominasi oleh posisi entry level, menunjukkan bahwa peluang kerja lebih banyak tersedia untuk posisi pemula, sementara peluang untuk level yang lebih tinggi lebih terbatas.

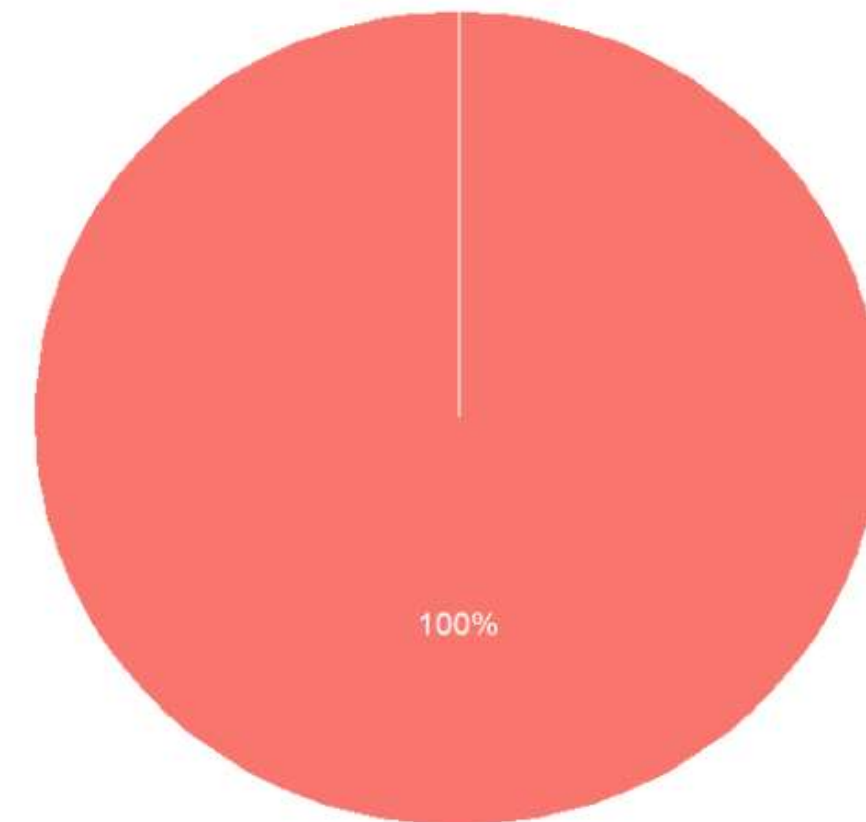




DISTRIBUSI PEKERJAAN BERDASARKAN JENIS PEKERJAAN

Visualisasi ini jelas menunjukkan bahwa semua pekerjaan yang dicatat dalam dataset ini adalah pekerjaan penuh waktu (FULL_TIME). Hal ini memberikan informasi penting bagi para pencari kerja dan perusahaan bahwa data yang dikumpulkan hanya mencakup pekerjaan penuh waktu, yang bisa menjadi batasan atau keunggulan tergantung pada kebutuhan mereka.

Distribusi Pekerjaan Berdasarkan Jenis Pekerjaan



Jenis Pekerjaan ■ FULL_TIME

Sumber: Hasil Web Scraping

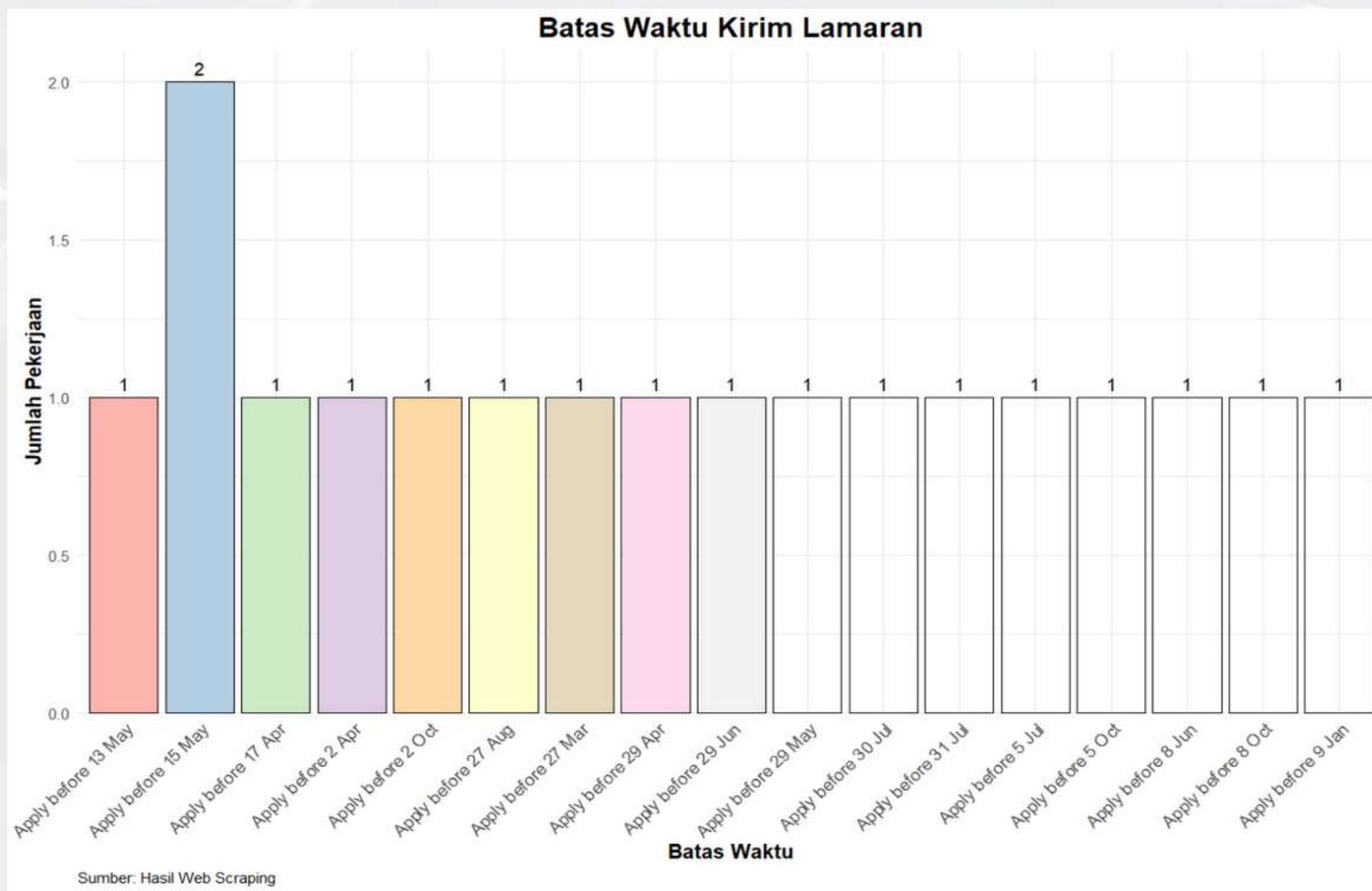
ANALYZING DATA



ANALYZING DATA

BATAS WAKTU KIRIM LAMARAN

Sebagian besar pekerjaan memiliki batas waktu pengiriman lamaran yang bervariasi. Tanggal 15 Mei memiliki jumlah lowongan pekerjaan terbanyak dengan dua lowongan, sementara tanggal lainnya masing-masing hanya memiliki satu lowongan pekerjaan. Kesimpulannya, lowongan pekerjaan di Kalibrr tersebar merata dengan beberapa tanggal batas waktu pengiriman lamaran yang berbeda, namun ada satu tanggal tertentu yang memiliki lebih banyak lowongan dibandingkan yang lain, menunjukkan peluang lebih besar untuk pelamar pada tanggal tersebut.

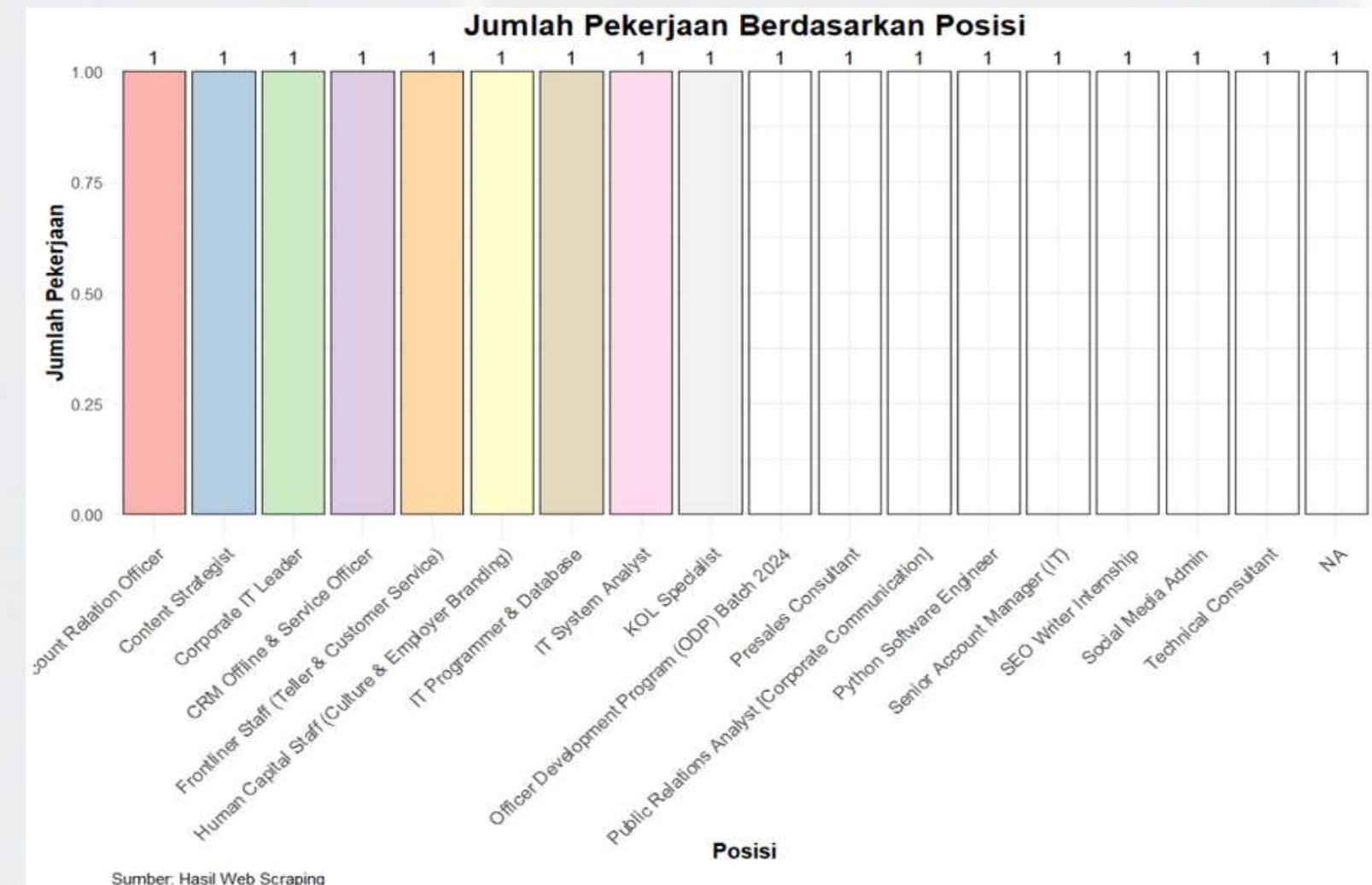




ANALYZING DATA

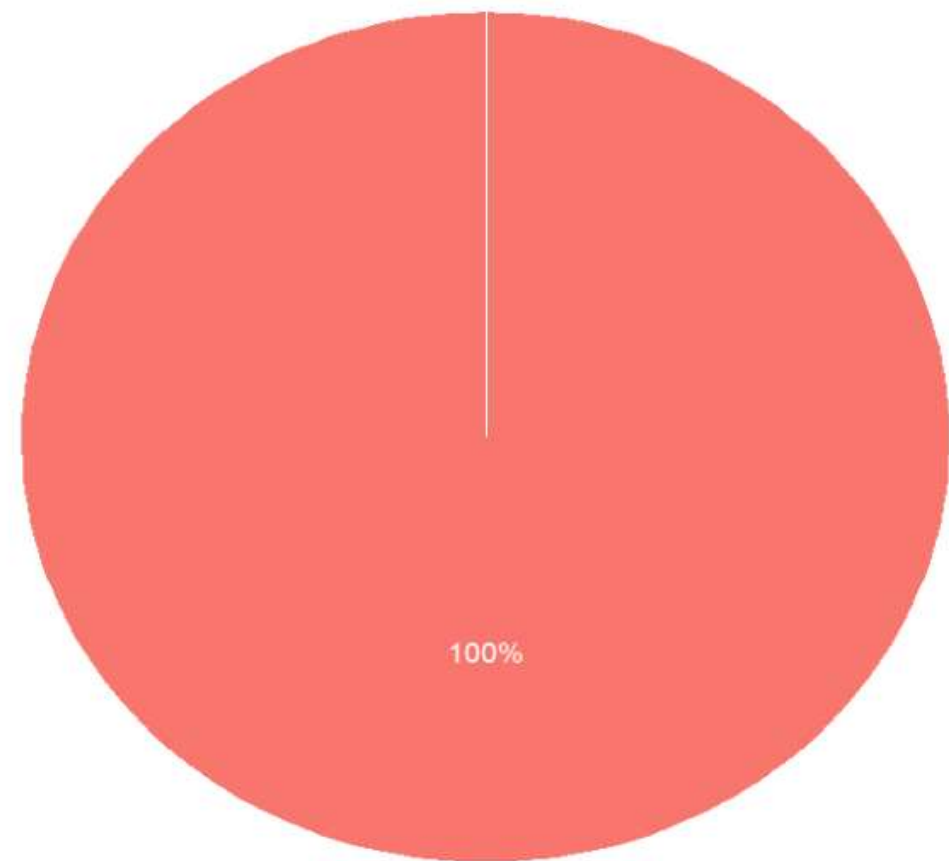
JUMLAH PEKERJAAN BERDASARKAN POSISI

Setiap posisi tersedia dengan jumlah lowongan yang sama, yaitu satu. Posisi mencakup berbagai bidang seperti Account Relation Officer, Content Strategist, Corporate IT Leader, dan lainnya. Tidak ada dominasi posisi tertentu dalam jumlah lowongan pekerjaan, menunjukkan distribusi yang merata dari berbagai posisi pekerjaan yang ditawarkan di Kalibrr.





Distribusi Pekerjaan Berdasarkan Rentang Gaji



Rentang Gaji ■ Salary Undisclosed

Sumber: Hasil Web Scrapping

DISTRIBUSI PEKERJAAN BERDASARKAN RENTANG GAJI

Visualisasi ini jelas menunjukkan bahwa semua pekerjaan yang dicatat dalam dataset ini memiliki kategori gaji "Salary Undisclosed". Hal ini memberikan informasi penting bahwa tidak ada pekerjaan dalam dataset yang menyebutkan gaji, yang bisa menjadi batasan bagi para pencari kerja yang memerlukan informasi gaji untuk pengambilan keputusan.

ANALYZING DATA



[rahmianadr](#)



[rahmianadra](#)

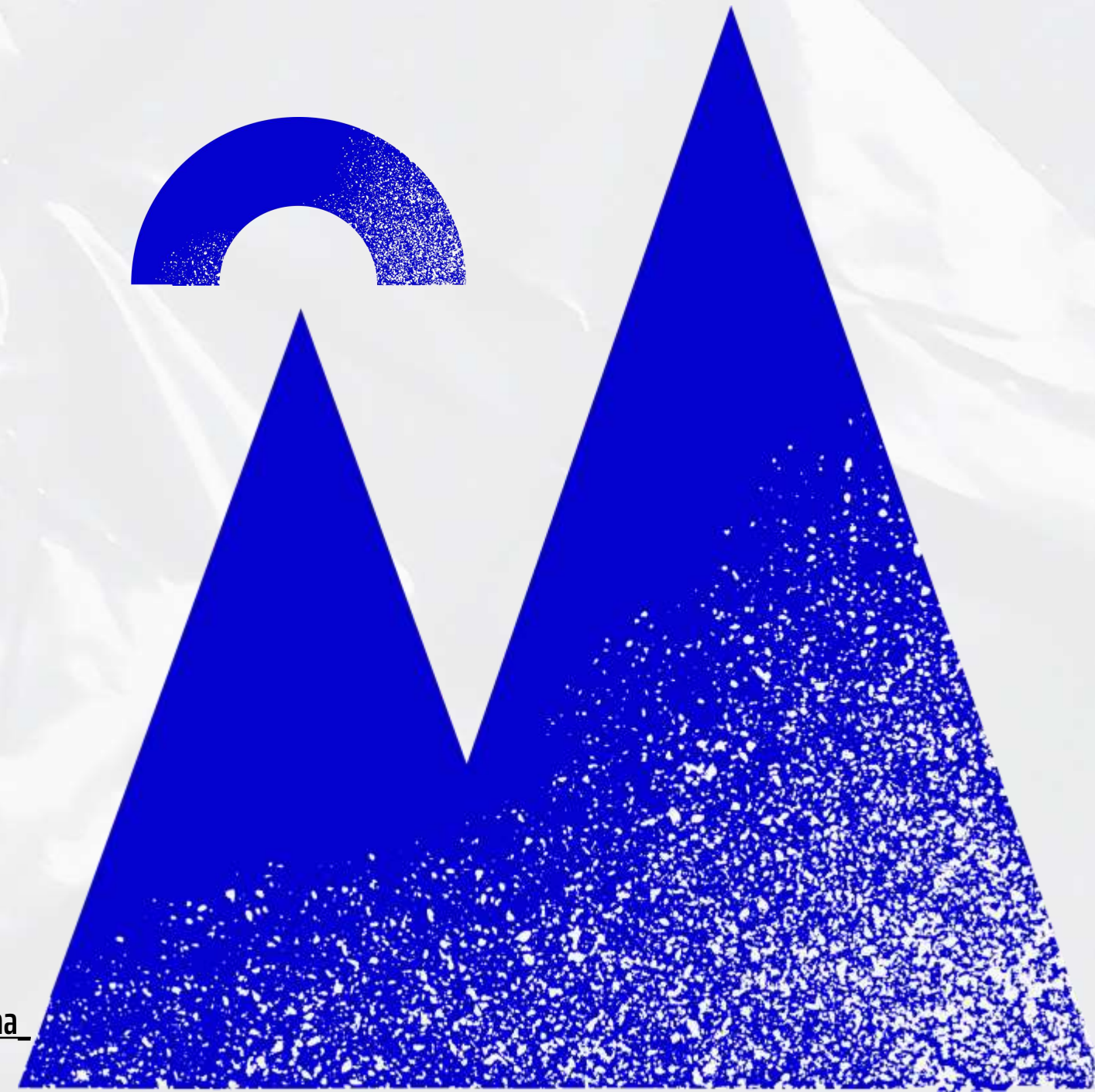
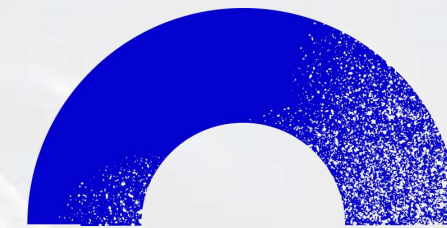


[rahmianadraa](#)



CONCLUSION

Dataset pekerjaan dari Kalibrr menunjukkan bahwa peluang kerja terbanyak berada di Jakarta dan sekitarnya, dengan mayoritas pekerjaan untuk level entry dan semuanya adalah pekerjaan penuh waktu tanpa informasi gaji. Lowongan pekerjaan tersebar merata di berbagai perusahaan dan posisi, dengan beberapa tanggal batas waktu lamaran yang memiliki peluang lebih tinggi. Informasi ini berguna bagi pencari kerja dalam memilih lokasi, jenis pekerjaan, dan waktu pengiriman lamaran yang tepat.



[rahmianadr](#)



[rahmianadra](#)



[rahmianadraa](#)



INSIGHT

Insight untuk Perusahaan

1. Strategi Rekrutmen di Luar Jakarta: Perusahaan bisa mempertimbangkan untuk membuka lebih banyak lowongan di kota-kota lain untuk mendiversifikasi lokasi kerja dan menjangkau lebih banyak talenta.
2. Transparansi Gaji: Memberikan informasi gaji yang jelas dapat meningkatkan daya tarik lowongan pekerjaan.
3. Rekrutmen untuk Level Senior: Meningkatkan jumlah lowongan untuk posisi Mid-Senior/Manager untuk menarik profesional berpengalaman.

Insight untuk Pencari Kerja

1. Fokus pada Jakarta dan Sekitarnya: Menargetkan lowongan pekerjaan di Jakarta dapat meningkatkan peluang mendapatkan pekerjaan.
2. Persiapan untuk Pekerjaan Penuh Waktu: Mempersiapkan diri untuk posisi penuh waktu dan negosiasi gaji.
3. Memanfaatkan Entry Level Opportunities: Fresh graduates harus memanfaatkan banyaknya lowongan entry level yang tersedia.



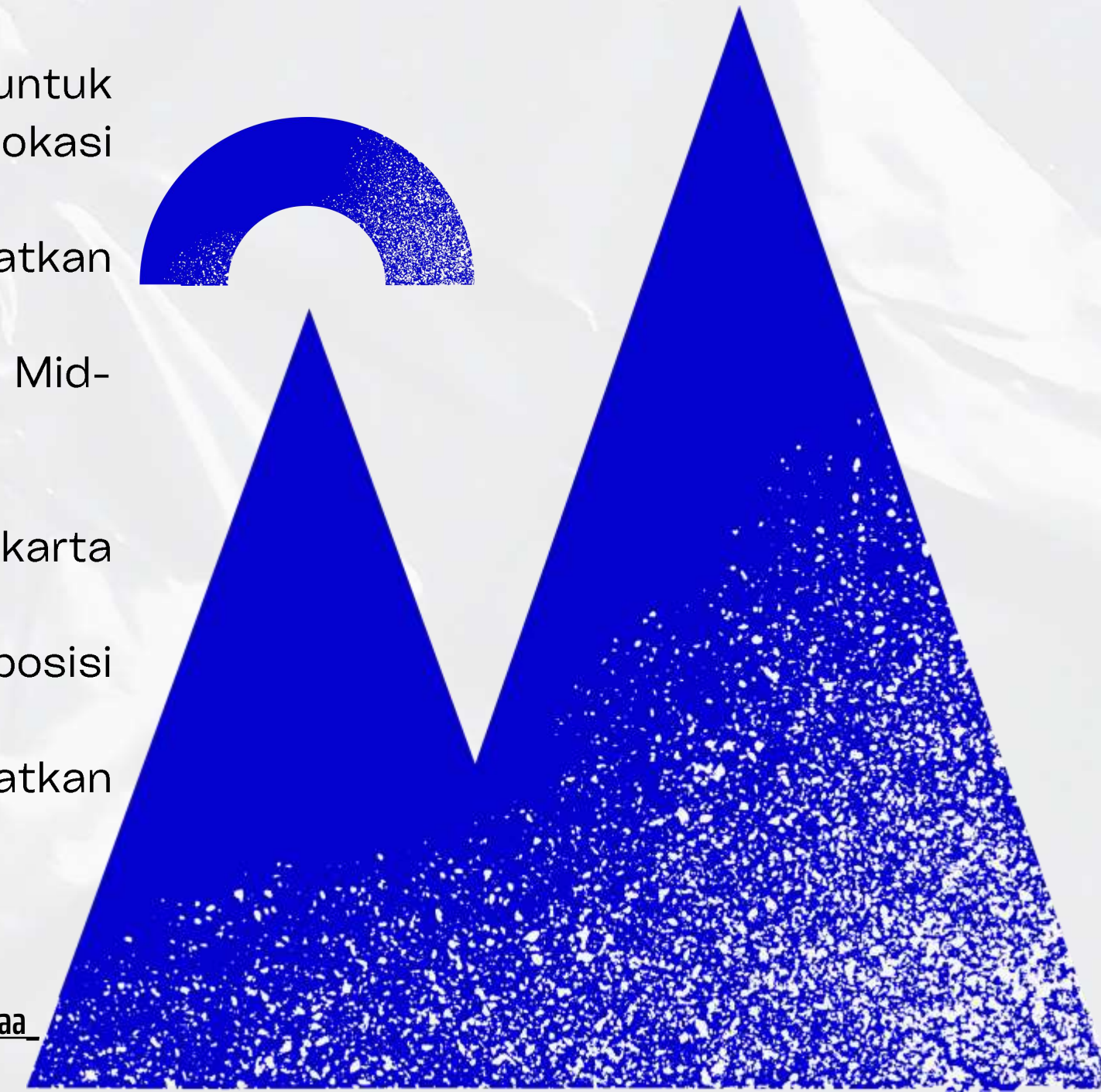
[rahmianadr](#)



[rahmianadra](#)



[rahmianadraa](#)



THANK YOU.

Project Manajemen Data Statistika



RAHMI ANADRA

Seorang data enthusiastic yang tengah menempuh pascasarjana Statistika dan Sains Data di IPB University. Tertarik pada bidang big data dengan berbagai software statistik seperti R Studio, Python, Tableau, dan lainnya.

Konten ini dibuat untuk tugas Praktikum Manajemen Data Statistika dan sebagai sumber pembelajaran bagi siapa pun yang tertarik. Selamat belajar!



rahmiandr



rahmianadra



rahmianadraa_