



RUANG BELAJAR DATA EPS. 5

Exploratory Data Analysis (**EDA**)

Klik untuk memulai proyek!





Perkenalkan.. Saya Muhammad Iqbal Pillie!

Analyst • Presenter • Content Creator

Pekerjaan:

- Data Analytics Consultant, [StarCore Analytics](#)
- Instructor, [RevoU](#)

Pendidikan:

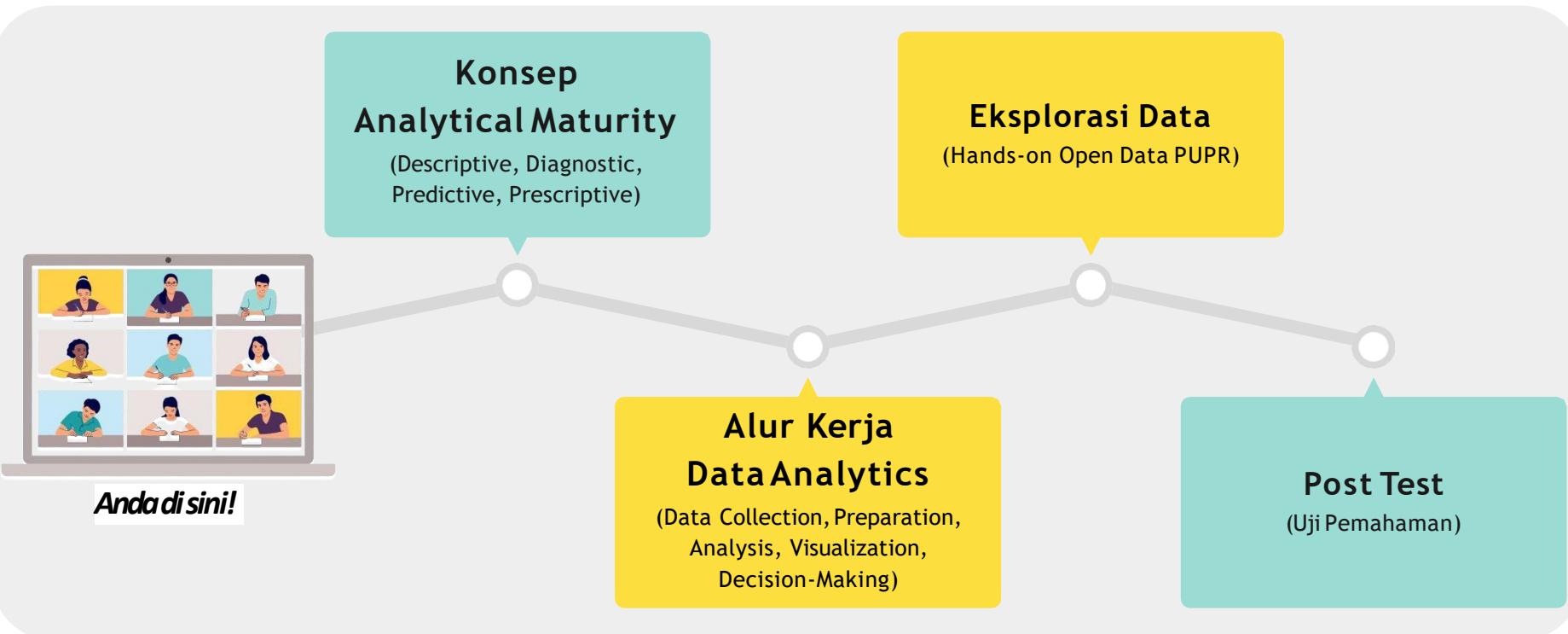
- Statistika, [IPB University](#)

Pengalaman Trainer:

- Statistika UNJ
- Jakarta Smart City
- BAPPEDA DKI Jakarta
- Bank BJB Syariah
- Mitra Mandiri Informatika



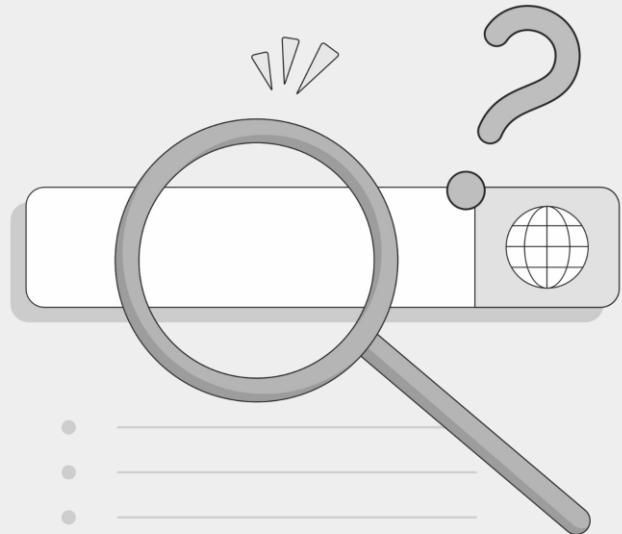
Rute Perjalanan



Ada yang kenal?



Ada yang kenal?



Carikan 1 dari 7 kembaran beliau!



HASIL DITEMUKAN !



Similarity score: 95%*



*Cerita ini hanyalah fiktif belaka, jika ada kesamaan dresscode, pose, dan gaya rambut adalah sebuah kebetulan dan memang ide cerita ini sangat memaksa.



Clive Humby

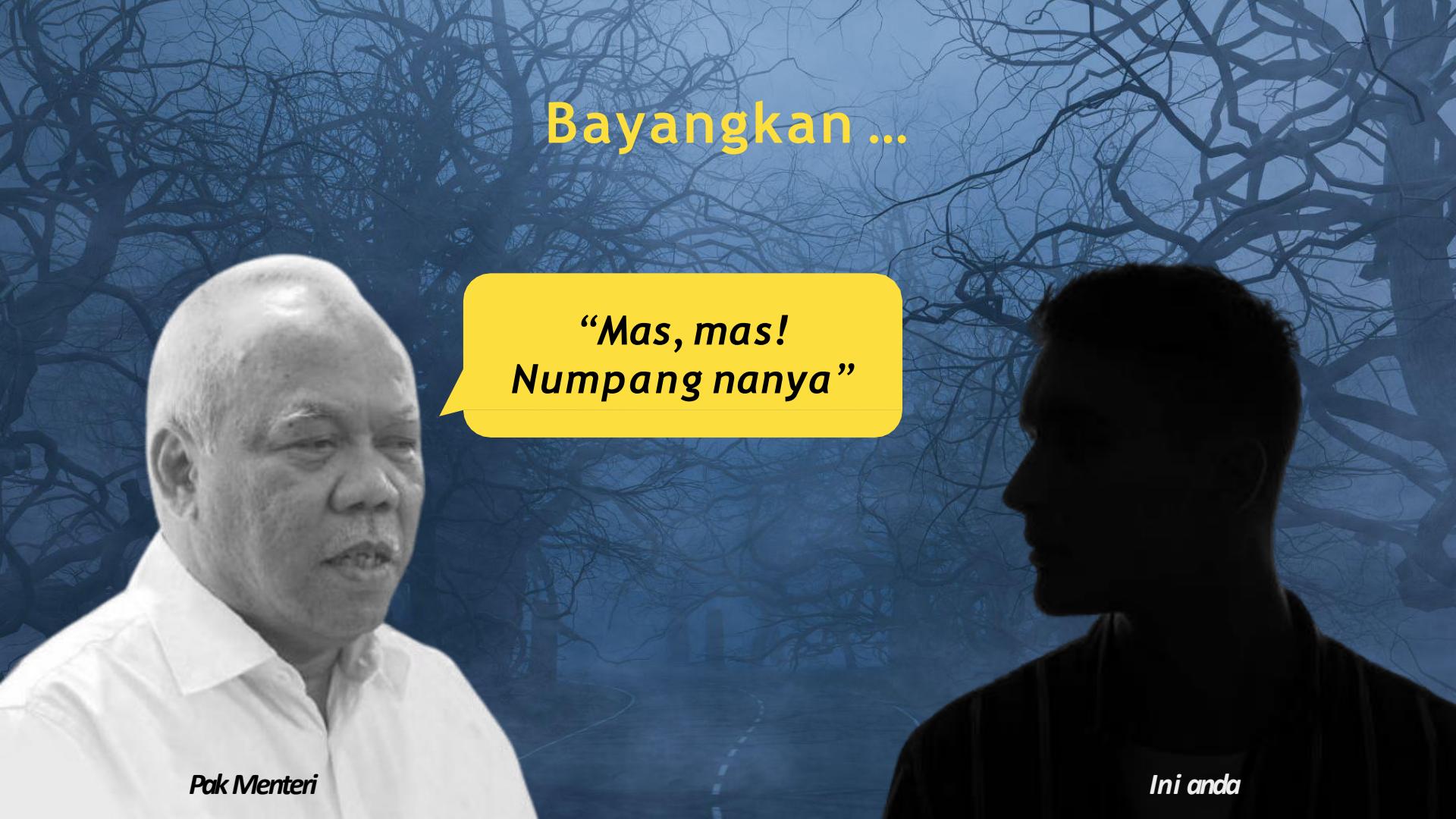
“Data is the new oil.” (2006)



Data is the new oil. Like oil, data is valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity. so, must data be broken down, analysed for it to have value.

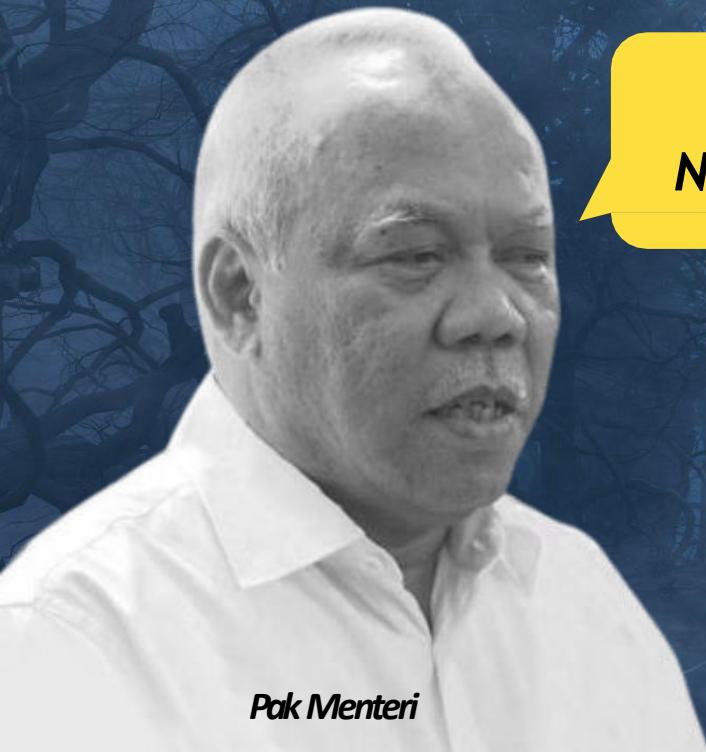


Dan lain-lain



Bayangkan ...

*“Mas, mas!
Numpang nanya”*



Pak Menteri



Ini anda



Role Play



“Apakah banyak proyek yang selesai tepat waktu tahun ini?”

Metode:

Membagi jumlah proyek selesai dengan total proyek tahun ini.

Analisis:

Tabel rekap bulanan, line chart, dll.



Role Play



*“Mengapa banyak proyek
yang **tertunda** tahun ini?”*

Metode:

Menganalisis faktor-faktor yang menyebabkan penundaan, seperti fluktuasi harga bahan baku, cuaca buruk, atau masalah manajemen proyek.

Analisis:

Analisis korelasi, ANOVA, analisis regresi linier, dll.



Role Play



“Apa yang akan terjadi jika kita menambah jumlah pekerja di proyek-proyek besar?”

Metode:

Membuat model prediksi untuk mengestimasi dampak penambahan jumlah pekerja terhadap waktu penyelesaian proyek-proyek besar.

Analisis:

Analisis time-series, forecasting, machine learning (Random Forest, Gradient Boosting, SVM), dll.



Role Play



“Bagaimana caranya kita memastikan proyek selesai tepat waktu?”

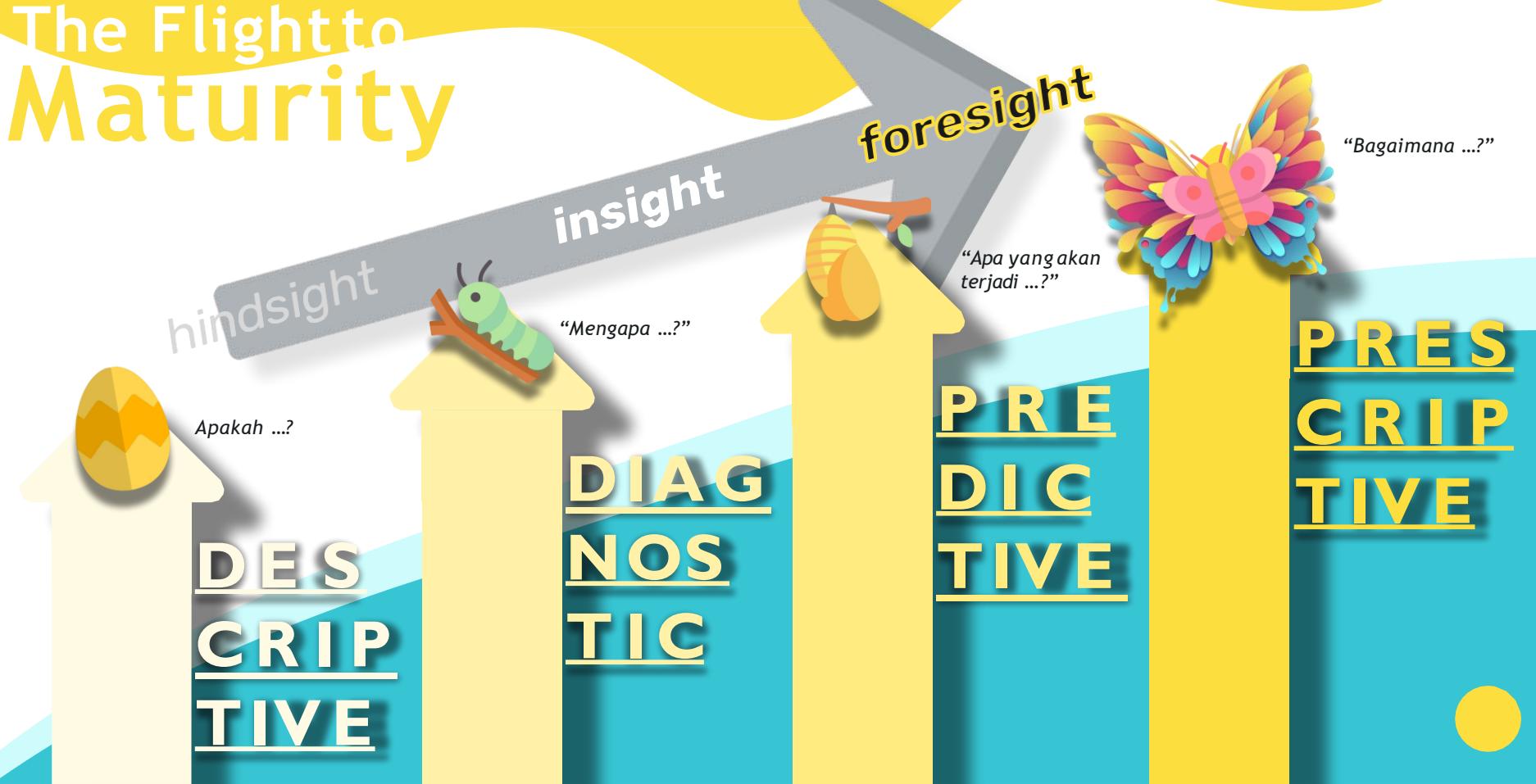
Metode:

Menggunakan model optimasi untuk menentukan alokasi sumber daya terbaik, seperti jumlah pekerja, waktu kerja, dan distribusi bahan.

Analisis:

Analisis optimasi (Linear Programming, Integer Programming), simulasi Monte Carlo (untuk menguji berbagai skenario), dll.

The Flight to Maturity



#SadarkahAnda?



"Apakah banyak proyek yang selesai **tepat waktu** tahun ini?"



Descriptive



"Mengapa banyak proyek yang **tertunda** tahun ini?"



Diagnostic



"Apa yang akan terjadi jika kita **menambah jumlah pekerja** di proyek-proyek besar?"



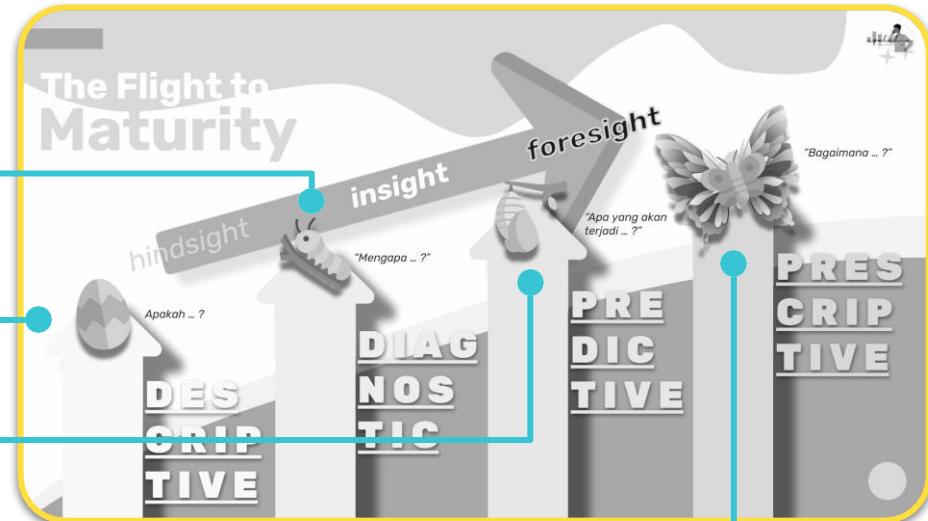
Predictive



"Bagaimana cara kita memastikan proyek **selesai tepat waktu**?"



Prescriptive



Konsep: [Five Stages of Analytics Maturity \(Davenport & Harris, 2007\)](#)

Gambar disunting dari: [Ngolah.in](#)



Jalan Tol Semarang-Demak Seksi 2
(Progress 93%) - Sep, 2022

Kira-kira, dimanakah peran data dalam pembangunan infrastruktur seperti jalan tol ini?

Sumber:

<https://jatengprov.go.id/publik/kostruksi-jalan-tol-semarang-demak-seksi-2-capai-93-persen/>





Data Collection

Atau proses pengumpulan data.

Sumber Data:

Direktorat Jenderal Bina Marga, Sumber Daya Air, Cipta Karya, Perumahan, Bina Konstruksi, dll. → [Open Data PUPR](#)

Contoh data:

Data progres proyek, anggaran, penggunaan sumberdaya, kondisi lapangan.

The screenshot shows the homepage of the Open Data PUPR website. The header features the PUPR logo and navigation links for Datasets, Organisasi, Stories, Topik, Visualisasi, Infografis, Produk, Belajar Data, and Bantuan. The main banner has a dark background with a network of nodes and the text "Selamat datang di OPEN DATA PUPR". Below the banner is a search bar with the placeholder "car" and a magnifying glass icon. At the bottom, there is a row of icons representing different data categories: APBN (Rp), Konektivitas, SDM Konstruksi, Ketahanan Air, Kebencanaan, and Permukiman. A red circular icon with a question mark is located in the bottom right corner.



Data Preparation

Atau proses **penyiapan data**.

Unit Kerja:

**Misalnya, Direktorat Bina Teknik Jalan dan Jembatan
(Direktorat Jenderal Bina Marga)**

Kegiatan:

Validasi dan konsolidasi data dari berbagai Direktorat Jenderal untuk memastikan akurasi.

Permen PUPR Nomor 13 Tahun 2020

BAB VI

DIREKTORAT JENDERAL BINA MARGA

Bagian Kesepuluh

Direktorat Bina Teknik Jalan dan Jembatan

Pasal 274

Subdirektorat Data dan Pengembangan Sistem Informasi Jalan dan Jembatan mempunyai tugas melaksanakan pembinaan dan pengembangan dan penyelenggaraan sistem manajemen jalan dan jembatan, pengolahan, analisis dan **validasi data** jalan daerah, pembinaan dan pengembangan teknologi informasi dan komunikasi bidang jalan dan jembatan, pengelolaan, pelayanan dan keamanan pangkalan data dan informasi spasial tematik bidang teknik jalan dan jembatan, pelaksanaan diseminasi dan penyiapan bahan informasi publik hasil pengembangan dan penerapan teknologi jalan dan jembatan, pengelolaan perpustakaan, jurnal, publikasi, dan dokumentasi kebinateknikan jalan dan jembatan, pengelolaan sistem manajemen pengetahuan (*knowledge management*) di bidang teknik jalan dan jembatan, pengembangan *Building Information Modelling*, pembinaan dan pengelolaan data leger jalan nasional, jalan bebas hambatan dan jalan daerah, koordinasi dan penyusunan program dan anggaran tahunan Direktorat dan unit pelaksana teknis/Balai Bina Teknik, pelaksanaan pemantauan, evaluasi, dan pelaporan direktorat.



Data Analysis

Atau proses **analisis data**.

Unit Kerja:

**Misalnya, Direktorat Bina Teknik Jalan dan Jembatan
(Direktorat Jenderal Bina Marga)**

Kegiatan:

Analisis tren progres proyek, identifikasi hambatan, dan peluang efisiensi.

Permen PUPR Nomor 13 Tahun 2020

BAB VI

DIREKTORAT JENDERAL BINA MARGA

Bagian Kesepuluh

Direktorat Bina Teknik Jalan dan Jembatan

Pasal 268

Dalam melaksanakan tugasnya sebagaimana dimaksud dalam Pasal 267, Direktorat Bina Teknik Jalan dan Jembatan menyelenggarakan fungsi:

- j. pelaksanaan **analisa data** dan pengembangan sistem manajemen jalan dan jembatan;
- k. pengembangan dan penyelenggaraan sistem informasi manajemen teknik jalan dan jembatan, pelaksanaan pangkalan data dasar, data statistik, infrastruktur teknologi informasi serta informasi spasial tematik bidang teknik jalan dan jembatan;

Pasal 269

Direktorat Bina Teknik Jalan dan Jembatan terdiri atas:

- a. Subdirektorat Teknologi dan Peralatan Infrastruktur Bina Marga;
- b. Subdirektorat Keselamatan dan Keamanan Jalan dan Jembatan;
- c. Subdirektorat **Data** dan Pengembangan Sistem Informasi Jalan dan Jembatan;
- d. Subdirektorat Pengelolaan Jabatan Fungsional Bidang Jalan dan Pengembangan Profesi; dan
- e. Subbagian Tata Usaha.



Data Visualization

Atau proses **visualisasi data**.

Unit Kerja:

**Misalnya, Direktorat Bina Teknik Jalan dan Jembatan
(Direktorat Jenderal Bina Marga)**

Kegiatan:

Pembuatan dashboard untuk pemantauan real-time proyek-proyek infrastruktur.

Permen PUPR Nomor 13 Tahun 2020

BAB VI

DIREKTORAT JENDERAL BINA MARGA

Bagian Kesepuluh

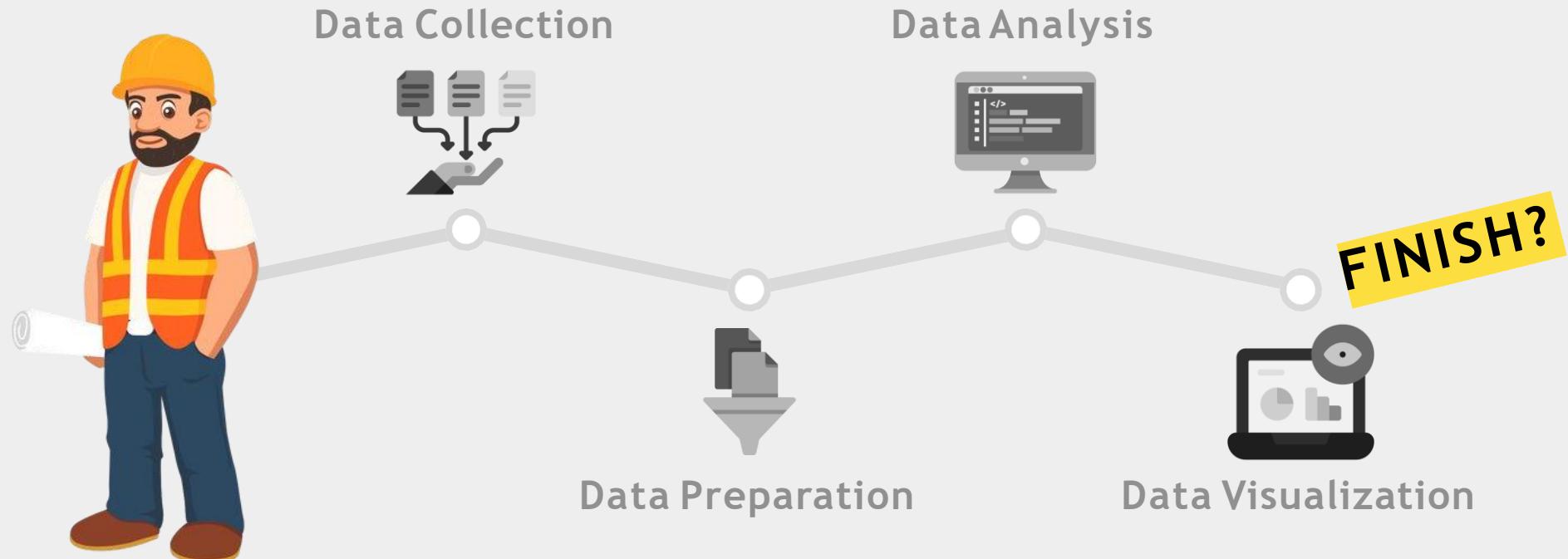
Direktorat Bina Teknik Jalan dan Jembatan

Pasal 274

Subdirektorat Data dan Pengembangan Sistem Informasi Jalan dan Jembatan mempunyai tugas melaksanakan pembinaan dan pengembangan dan penyelenggaraan sistem manajemen jalan dan jembatan, pengolahan, analisis dan validasi data jalan daerah, pembinaan dan pengembangan teknologi informasi dan komunikasi bidang jalan dan jembatan, pengelolaan, pelayanan dan keamanan pangkalan data dan informasi spasial tematik bidang teknik jalan dan jembatan, pelaksanaan diseminasi dan penyiapan bahan **informasi publik** hasil pengembangan dan penerapan teknologi jalan dan jembatan, pengelolaan perpustakaan, jurnal, publikasi, dan dokumentasi kebinatan teknik jalan dan jembatan, pengelolaan sistem manajemen pengetahuan (*knowledge management*) di bidang teknik jalan dan jembatan, pengembangan *Building Information Modelling*, pembinaan dan pengelolaan data leger jalan nasional, jalan bebas hambatan dan jalan daerah, koordinasi dan penyusunan program dan anggaran tahunan Direktorat dan unit pelaksana teknis/Balai Bina Teknik, pelaksanaan pemantauan, evaluasi, dan pelaporan direktorat.



Alur Kerja Data Analytics



Jangan terlewat!

Decision Making

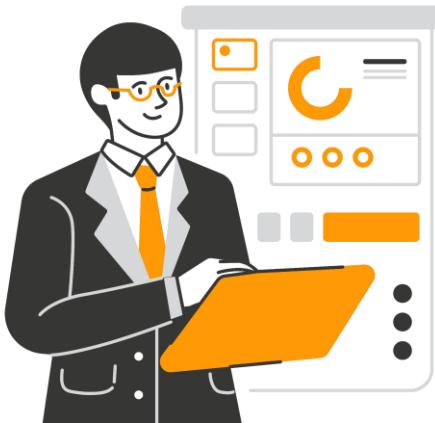
Atau proses pengambilan kesimpulan berbasis data.

Unit Kerja:

Misalnya, Direktorat Sistem & Strategi Penyelenggaraan Jalan & Jembatan (Direktorat Jenderal Bina Marga)

Kegiatan:

Perumusan kebijakan penyelenggaraan jalan berbasis data dan dapat dimonitor melalui real-time dashboard.



Permen PUPR Nomor 13 Tahun 2020

BAB VI

DIREKTORAT JENDERAL BINA MARGA

Bagian Keempat

Direktorat Sistem dan Strategi Penyelenggaraan Jalan dan Jembatan

Pasal 195

Direktorat Sistem dan Strategi Penyelenggaraan Jalan dan Jembatan mempunyai tugas melaksanakan perumusan kebijakan penyelenggaraan jalan di bidang keterpaduan sistem jaringan jalan dan jembatan, strategi program dan anggaran, pembinaan teknis jalan daerah, pengelolaan pinjaman dan hibah luar negeri serta pemantauan dan evaluasi penyelenggaraan jalan dan jembatan termasuk kawasan metropolitan dan kota besar.



Alur Kerja Data Analytics + Alat Kerja

Data Collection

Google Form, SurveyMonkey,
Google Analytics, etc.



Data Analysis

Python, R, Google Sheets,
Excel, SAS, SPSS, etc.



Decision Making

Orang dengan kemampuan
Stat/Math, Computer Science,
& Business Expertise



Data Preparation

Python, R, Google Sheets, Excel,
Google BigQuery, MySQL, etc.



Data Visualization

SAS Viya, PowerBI, Tableau,
Looker Studio, QlikView, etc.



Data Analytics merupakan **tahapan yang panjang** bukan?

→ Jawabannya: **BUKAN!**

→ Kunci jawaban: Lebih panjang **MACET!**

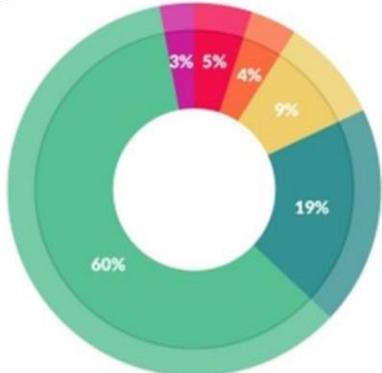
Fun fact: MACET ≈ DATA CLEANING *(sama-sama makan waktu!)*

“Macet dianggap **seni** dalam mudik (*tidak bagi saya!*)
Data Cleaning adalah **seni** dalam analisis data.”

*dianggap = subjektif, adalah = mutlak

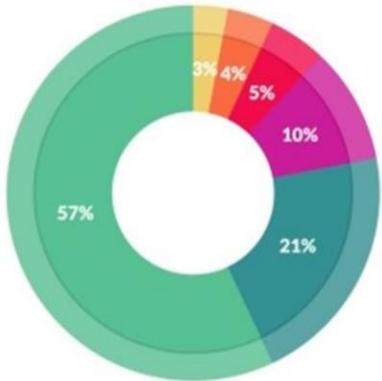


- Iqbal, 2024



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60% (highlighted)
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%



What's the least enjoyable part of data science?

- Building training sets: 10%
- Cleaning and organizing data: 57% (highlighted)
- Collecting data sets: 21%
- Mining data for patterns: 3%
- Refining algorithms: 4%
- Other: 5%

DATA CLEANING:

Dianggap paling makan waktu,
tapi paling makan hati!

Source: [Survey by CrowdFlower \(2016\)](#)

Kenapa Harus Data Cleaning?



GIGO

“Garbage in, garbage out”

Hasil analisis Anda akan sama bagusnya dengan data Anda yang dapat diandalkan, atau malah seburuk data Anda yang cacat.

#MariBeranalogi



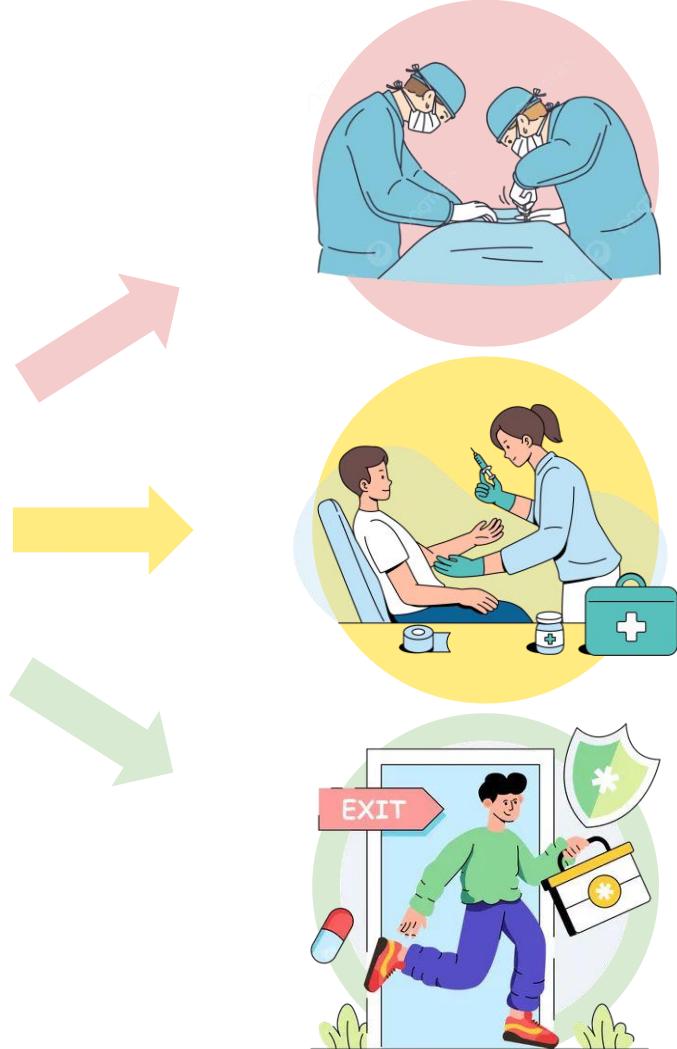
*EDA dapat dikategorikan ke tahapan
Data Cleaning maupun Data Analysis*

Apa yang dokter lakukan pertama kali pada pasien yang datang?



Dokter
Diagnosa awal

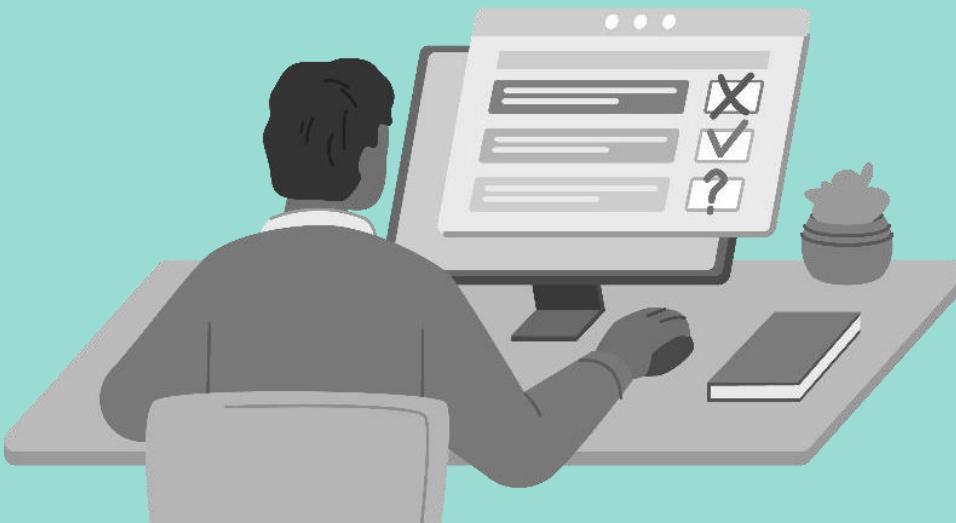
Analyst
≈ Eksplorasi data



Hands-on

Eksplorasi Data (EDA)

- Kondisi Permukaan Jalan Kab/Kota
- Jenis Permukaan Jalan Kab/Kota
- Panjang Kemantapan Jembatan Nasional



Silakan akses dataset di sini:
bit.ly/DataRBD5

Step 1: Data Validation (Cek Konsistensi Keberadaan Nilai ID/Primary Key)

Data:

Kondisi Permukaan Jalan Kab/Kota 2020-2023

2020	2021	2022	2023
Kd_Prov Provinsi	Kd_Prov Provinsi	Kd_Prov Provinsi	Kode Provinsi Provinsi
11 Aceh	11 Aceh	11 Aceh	11 Aceh
12 Sumatera Utara	12 Sumatera Utara	51 Bali	12 Sumatera Utara
13 Sumatera Barat	13 Sumatera Barat	36 Banten	13 Sumatera Barat
14 Riau	14 Riau	17 Bengkulu	14 Riau
15 Jambi	15 Jambi	34 D.I. Yogyakarta	15 Jambi
16 Sumatera Selatan	16 Sumatera Selatan	75 Gorontalo	16 Sumatera Selatan
17 Bengkulu	17 Bengkulu	15 Jambi	17 Bengkulu
18 Lampung	18 Lampung	32 Jawa Barat	18 Lampung
19 Kepulauan Bangka Belitung	19 Kepulauan Bangka Belitung	33 Jawa Tengah	19 Kepulauan Bangka Belitung
21 Kepulauan Riau	21 Kepulauan Riau	35 Jawa Timur	21 Kepulauan Riau
31 DKI Jakarta	32 Jawa Barat	61 Kalimantan Barat	32 Jawa Barat
32 Jawa Barat	33 Jawa Tengah	63 Kalimantan Selatan	33 Jawa Tengah
33 Jawa Tengah	34 D.I. Yogyakarta	62 Kalimantan Tengah	34 D.I. Yogyakarta
34 D.I. Yogyakarta	35 Jawa Timur	64 Kalimantan Timur	35 Jawa Timur
35 Jawa Timur	36 Banten	65 Kalimantan Utara	36 Banten
36 Banten	51 Bali	19 Kepulauan Bangka Belitung	51 Bali
51 Bali	52 Nusa Tenggara Barat	21 Kepulauan Riau	52 Nusa Tenggara Barat
52 Nusa Tenggara Barat	53 Nusa Tenggara Timur	18 Lampung	53 Nusa Tenggara Timur
53 Nusa Tenggara Timur	61 Kalimantan Barat	81 Maluku	61 Kalimantan Barat
61 Kalimantan Barat	62 Kalimantan Tengah	82 Maluku Utara	62 Kalimantan Tengah
62 Kalimantan Tengah	63 Kalimantan Selatan	52 Nusa Tenggara Barat	63 Kalimantan Selatan
63 Kalimantan Selatan	64 Kalimantan Timur	53 Nusa Tenggara Timur	64 Kalimantan Timur
64 Kalimantan Timur	65 Kalimantan Utara	94 Papua	65 Kalimantan Utara
65 Kalimantan Utara	71 Sulawesi Utara	91 Papua Barat	71 Sulawesi Utara
71 Sulawesi Utara	72 Sulawesi Tengah	14 Riau	72 Sulawesi Tengah
72 Sulawesi Tengah	73 Sulawesi Selatan	76 Sulawesi Barat	73 Sulawesi Selatan
73 Sulawesi Selatan	74 Sulawesi Tenggara	73 Sulawesi Selatan	74 Sulawesi Tenggara
74 Sulawesi Tenggara	75 Gorontalo	72 Sulawesi Tengah	75 Gorontalo
75 Gorontalo	76 Sulawesi Barat	74 Sulawesi Tenggara	76 Sulawesi Barat
76 Sulawesi Barat	81 Maluku	71 Sulawesi Utara	81 Maluku
81 Maluku	82 Maluku Utara	13 Sumatera Barat	82 Maluku Utara
82 Maluku Utara	91 Papua Barat	16 Sumatera Selatan	91 Papua Barat
91 Papua Barat	94 Papua	12 Sumatera Utara	94 Papua
94 Papua	31 DKI Jakarta	31 DKI Jakarta	31 DKI Jakarta

Permasalahan:

Misalkan terdapat **inkonsistensi keberadaan nilai ID (Kode Provinsi atau Provinsi)** pada tahun yang berbeda.

Untuk integrasi tabel, dibutuhkan ID yang konsisten.

Dari kasus ini, terdapat 2 hal yang menyebabkan inkonsistensi:

1. **DKI Jakarta** tidak berada pada dataset tahun 2021-2023 (**Anomali**) → **Perlu ditelusuri alasannya.**
2. Terdapat **Papua Selatan, Papua Tengah, Papua Pegunungan, Papua Barat Daya** pada dataset tahun 2023 (**Normal**) → **Pemekaran Provinsi Papua 2022.**

Solusi:

Membuat dataset gabungan (**master data**) dengan **Provinsi lengkap**, yakni sebanyak **38** agar dapat menampung seluruh kemungkinan dataset yang tersedia pada tahun 2020-2023.

Step 2: Data Validation (Cek Konsistensi Nama Kolom)

Data:

Kondisi Permukaan Jalan Kab/Kota 2020-2023

2020	2021	2022	2023
Kolom	Kolom	Kolom	Kolom
Kondisi_Baik	Kondisi_Baik	Kondisi_Baik	Baik km
Baik_%	Baik_%	Baik_%	Baik %
Kondisi_Sedang	Kondisi_Sedang	Kondisi_Sedang	Sedang km
Sedang_%	Sedang_%	Sedang_%	Sedang %
Kondisi_RR	Kondisi_RR	Kondisi_RR	RR km
RR_%	RR_%	RR_%	RR %
Kondisi_RB	Kondisi_RB	Kondisi_RB	RB km
RB_%	RB_%	RB_%	RB %
Total_Panjang	Total_Panjang	Total_Panjang	Total panjang

*Warna abu-abu → Penulisan nama kolom berbeda sedikit dari yang seharusnya → Dapat diperbaiki langsung oleh Data Analyst/Data Engineer.

*Warna merah → Penulisan nama kolom cukup berbeda dari yang seharusnya → Dapat diperbaiki setelah memastikan isi kolom sesuai dengan padanannya, bahkan di kondisi tertentu perlu dikonsultasikan dengan penyedia data atau business expert.

Permasalahan:

Misalkan terdapat **inkonsistensi nama kolom** pada tahun yang berbeda. Untuk integrasi tabel, sebaiknya nama kolom distandarisasi, meskipun secara logika dapat dengan mudah diperbaiki*.

*Namun, saat terdapat banyak sekali kolom, proses standarisasi akan memakan waktu yang panjang.

Solusi:

- **Jangka pendek:** Melakukan perbaikan penulisan mengikuti penulisan paling dominan.
- **Jangka panjang:** Melakukan standarisasi pada keseluruhan **database** mengikuti aturan penulisan yang berlaku → *Proses ini dapat melibatkan berbagai tim dan unit kerja dalam direktorat untuk memastikan proses ini benar-benar diterapkan.*

Step 3: Data Validation (Cek Konsistensi Antar Kolom)

Data:

Kondisi Permukaan Jalan Kab/Kota 2020-2023

2020-2022

Kd_Prov	Provinsi
11	Aceh
12	Sumatera Utara
13	Sumatera Barat
14	Riau
15	Jambi
16	Sumatera Selatan
17	Bengkulu
18	Lampung
19	Kepulauan Bangka Belitung
21	Kepulauan Riau
32	Jawa Barat
33	Jawa Tengah
34	D.I. Yogyakarta
35	Jawa Timur
36	Banten
51	Bali
52	Nusa Tenggara Barat
53	Nusa Tenggara Timur
61	Kalimantan Barat
62	Kalimantan Tengah
63	Kalimantan Selatan
64	Kalimantan Timur
65	Kalimantan Utara
71	Sulawesi Utara
72	Sulawesi Tengah
73	Sulawesi Selatan
74	Sulawesi Tenggara
75	Gorontalo
76	Sulawesi Barat
81	Maluku
82	Maluku Utara
91	Papua Barat
94	Papua

2023

Kode Provinsi	Provinsi
11	Aceh
12	Sumatera Utara
13	Sumatera Barat
14	Riau
15	Jambi
16	Sumatera Selatan
17	Bengkulu
18	Lampung
19	Kepulauan Bangka Belitung
21	Kepulauan Riau
32	Jawa Barat
33	Jawa Tengah
34	D.I. Yogyakarta
35	Jawa Timur
36	Banten
51	Bali
52	Nusa Tenggara Barat
53	Nusa Tenggara Timur
61	Kalimantan Barat
62	Kalimantan Tengah
63	Kalimantan Selatan
64	Kalimantan Timur
65	Kalimantan Utara
71	Sulawesi Utara
72	Sulawesi Tengah
73	Sulawesi Selatan
74	Sulawesi Tenggara
75	Gorontalo
76	Sulawesi Barat
81	Maluku
82	Maluku Utara
91	Papua
92	Papua Barat
93	Papua Selatan
94	Papua Tengah
95	Papua Pegunungan
96	Papua Barat Daya

Permasalahan:

Misalkan terdapat **inkonsistensi Kode Provinsi terhadap Nama Provinsi** pada tahun yang berbeda. Pada tahun 2023, Kode Provinsi “91” dan “94” mengalami perubahan pasangan Nama Provinsinya dibandingkan dengan tahun-tahun sebelumnya.

Solusi:

- Opsi 1: Menyeragamkan Nama Provinsi untuk Kode “91” dan “94”. Ini dapat mengikuti penulisan tahun 2023, ataupun penulisan tahun 2020-2022, tergantung pada referensi atau kesepakatan tim.
- Opsi 2: Melakukan integrasi (join table) dengan merujuk pada variabel Provinsi, ketimbang variabel Kode Provinsi → Untuk kasus ini, opsi 2 lebih aman secara substansi. Namun apabila diaplikasikan pada nama orang/customer, ataupun nama produk yang memiliki kemungkinan penulisan sama, opsi ini perlu dipertimbangkan lagi.

Step 4: Data Validation (Cek Konsistensi Nilai)

Data:

Kondisi Permukaan Jalan Kab/Kota 2020-2023

2020-2022

Kd_Prov	Provinsi
11	Aceh
12	Sumatera Utara
13	Sumatera Barat
14	Riau
15	Jambi
16	Sumatera Selatan
17	Bengkulu
18	Lampung
19	Kepulauan Bangka Belitung
21	Kepulauan Riau
32	Jawa Barat
33	Jawa Tengah
34	D.I. Yogyakarta
35	Jawa Timur
36	Banten
51	Bali
52	Nusa Tenggara Barat
53	Nusa Tenggara Timur
61	Kalimantan Barat
62	Kalimantan Tengah
63	Kalimantan Selatan
64	Kalimantan Timur
65	Kalimantan Utara
71	Sulawesi Utara
72	Sulawesi Tengah
73	Sulawesi Selatan
74	Sulawesi Tenggara
75	Gorontalo
76	Sulawesi Barat
81	Maluku
82	Maluku Utara
91	Papua Barat
94	Papua

2023

Kode Provinsi	Provinsi
11	Aceh
12	Sumatera Utara
13	Sumatera Barat
14	Riau
15	Jambi
16	Sumatera Selatan
17	Bengkulu
18	Lampung
19	Kepulauan Bangka Belitung
21	Kepulauan Riau
32	Jawa Barat
33	Jawa Tengah
34	D.I. Yogyakarta
35	Jawa Timur
36	Banten
51	Bali
52	Nusa Tenggara Barat
53	Nusa Tenggara Timur
61	Kalimantan Barat
62	Kalimantan Tengah
63	Kalimantan Selatan
64	Kalimantan Timur
65	Kalimantan Utara
71	Sulawesi Utara
72	Sulawesi Tengah
73	Sulawesi Selatan
74	Sulawesi Tenggara
75	Gorontalo
76	Sulawesi Barat
81	Maluku
82	Maluku Utara
91	Papua
92	Papua Barat
93	Papua Selatan
94	Papua Tengah
95	Papua Pegunungan
96	Papua Barat Daya

Permasalahan:

Apabila mengikuti Opsi 2 pada langkah sebelumnya (Step 3), masih terdapat permasalahan lainnya, yaitu **inkonsistensi penulisan nama Provinsi** pada tahun yang berbeda. **Pada tahun 2023**, penulisan Provinsi DIY sedikit mengalami perubahan dari yang awalnya tertulis “D.I. Yogyakarta” menjadi “DI Yogyakarta”. Hal ini terlihat sepele (hanya perbedaan pada titik), namun apabila bermain dengan data yang besar, hal seperti ini mungkin bisa terlewatkhan, dan bahkan DIY akan dianggap sebagai 2 provinsi yang berbeda oleh software.

Solusi:

- **Jangka pendek:** Melakukan perbaikan penulisan **mengikuti** penulisan pada tahun terbaru.
- **Jangka panjang:** Melakukan **standarisasi** pada keseluruhan **database** mengikuti aturan penulisan yang berlaku → *Proses ini dapat melibatkan berbagai tim dan unit kerja dalam direktorat untuk memastikan proses ini benar-benar diterapkan.*

Step 5: Data Exploration (Cek Potensi Outliers/Anomali)

Data: Kondisi Permukaan Jalan Kab/Kota 2020-2023

Permasalahan: Misalkan Data Tahun **2021 memiliki nilai maximum yang mencurigakan**, artinya terdapat Provinsi dengan panjang jalan (dalam kondisi baik) yang terlalu besar dibandingkan dengan provinsi lainnya.

2020

Kondisi_Baik	
Mean	5,081.8
Standard Error	804.7
Median	3,256.6
Mode	21,095.5
Standard Deviation	4,692.3
Sample Variance	22,018,109.1
Kurtosis	3.3
Skewness	1.8
Range	21,095.5
Minimum	0.0
Maximum	21,095.5
Sum	172,781.5
Count	34.0
Largest(1)	21,095.5
Smallest(1)	0.0
Confidence Level(95)	1,637.2

2021

Kondisi_Baik	
Mean	679,116.6
Standard Error	321,018.7
Median	3,773.7
Mode	7,157,929.0
Standard Deviation	1,844,112.0
Sample Variance	3,400,749,087.0
Kurtosis	7.4
Skewness	2.9
Range	7,157,152.2
Minimum	776.8
Maximum	7,157,929.0
Sum	22,410,848.8
Count	33.0
Largest(1)	7,157,929.0
Smallest(1)	776.8
Confidence Level(95)	653,893.7

2022

Kondisi_Baik	
Mean	5,256.3
Standard Error	793.1
Median	3,379.1
Mode	20,148.3
Standard Deviation	4,555.8
Sample Variance	20,755,296.5
Kurtosis	2.6
Skewness	1.7
Range	19,263.3
Minimum	885.0
Maximum	20,148.3
Sum	173,456.5
Count	33.0
Largest(1)	20,148.3
Smallest(1)	885.0
Confidence Level(95)	1,615.4

2023

Baik km	
Mean	4,458.8
Standard Error	695.6
Median	3,051.4
Mode	18,635.2
Standard Deviation	4,231.5
Sample Variance	17,905,210.3
Kurtosis	2.7
Skewness	1.6
Range	18,382.1
Minimum	253.0
Maximum	18,635.2
Sum	164,976.8
Count	37.0
Largest(1)	18,635.2
Smallest(1)	253.0
Confidence Level(95)	1,410.8

*Statistik deskriptif di atas diperoleh melalui Add-ons XLMiner Analysis Toolpak ([klik link ini](#) untuk install di Google Sheets-mu!)

Solusi:

- Cek sumber data, atau bahkan hubungi penyedia data.
- Cek referensi dari sumber eksternal untuk memvalidasi nilai, misalnya berita di internet.

Step 6: Data Verification (Cek Sumber Data Internal)

→ Melakukan pengecekan ke sumber data (internal)

Kondisi Permukaan Jalan Kabupaten-Kota Tahun 2021.csv

CSV

Data Preview: Note that by default the preview only displays up to 100 records. Use the pager to flip through more records or adjust the start and end fields to display the number of records you wish to see.

Grid Graph Map 34 records « 1 34 » Q Search data ... Go » Filters Fields

Kd_P...	Provinsi	Kondisi_Baik	Baik...	Kond...	Seda...	Kond...	RR_%	Kond...	RB_%	Total
15	Jambar	2455.90	51.07	2354....	50.1	1077....	24.21	1067....	14.02	771
16	Sumatera Selatan	4157621	33.85	44082...	35.89	24301...	19.79	12858...	10.47	12
17	Bengkulu	2239.12	32.53	1775....	25.8	2002....	29.09	865.88	12.58	688
18	Lampung	4878.12	27.45	5206....	29.29	5539....	31.16	2150....	12.1	17
19	Kepulauan Bangka B...	2024.95	70.03	314.52	10.88	460.52	15.93	91.7	3.17	289
21	Kepulauan Riau	2120.35	46.03	1332....	28.94	927.36	20.13	225.5	4.9	460
32	Jawa Barat	15220.95	62.63	4454....	18.33	2371....	9.76	2255....	9.28	24
33	Jawa Tengah	14826.72	55.63	6739....	25.28	2920....	10.96	2167....	8.13	260
34	D.I. Yogyakarta	1744.67	51.9	798.06	23.74	414.05	12.32	404.83	12.04	330
35	Jawa Timur	20198.57	64.02	6097....	19.33	3257.4	10.33	1994....	6.32	319
36	Banten	1651.22	62.45	337.72	12.77	300.62	11.37	354.65	13.41	264
51	Bali	4387.59	61.23	1440.4	20.1	587.67	8.2	750.41	10.47	710
52	Nusa Tenggara Barat	3366.67	55.33	639.26	10.51	1026....	16.87	1052....	17.3	608
53	Nusa Tenggara Timur	7157929	46.27	14339....	9.27	14129....	9.13	54651....	35.33	154
61	Kalimantan Barat	2934040	18.28	54172....	33.75	41166....	25.64	35848....	22.33	160
62	Kalimantan Tengah	3773.71	25.74	3823....	26.08	2776....	18.94	4289....	29.25	140
63	Kalimantan Selatan	4413.82	38.45	3159....	27.52	2256.1	19.65	1650.7	14.38	114
64	Kalimantan Timur	2838.81	30.44	3135....	33.62	2420....	25.95	932.27	10	93
65	Kalimantan Utara	776.76	31.52	592.21	24.03	641.18	26.02	454.24	18.43	246
71	Sulawesi Utara	3799.31	54.97	960.78	13.9	845.47	12.23	1305....	18.89	691

Temuan:

Misalkan ternyata, nilai maksimum panjang jalan kab/kota (kondisi baik) yang mencurigakan (tahun 2021) merupakan data Provinsi Nusa Tenggara Timur, yakni sepanjang 7.157,929 km.

Pertanyaan:

Apakah panjang jalan kab/kota (kondisi baik) di Provinsi NTT sepanjang 7.1juta km dianggap masuk akal?

Step 7: Data Verification (Cek Referensi Data Eksternal)

→ Melakukan pengecekan referensi dari sumber eksternal, yaitu BPS maupun Google Search.

Provinsi	Panjang Jalan Negara (km) (km)	Panjang Jalan Provinsi (km) (km)	Panjang Jalan Kabupaten/Kota (km) (km)	Jumlah Panjang Jalan(km) (km)
DI Yogyakarta	248	760	3.330	4.338
Jawa Timur	2.361	1.421	38.739	42.521
Banten	565	762	4.385	5.712
Bali	629	743	7.415	8.787
Nusa Tenggara Barat	935	1.484	6.085	8.504
Nusa Tenggara Timur	1.858	2.650	19.148	23.656
Kalimantan Barat	2.118	1.535	14.662	18.315
Kalimantan Tengah	2.002	1.272	15.535	18.809
Kalimantan Selatan	1.204	756	11.480	13.440
Kalimantan Timur	1.711	895	10.664	13.270
Kalimantan Utara	585	852	3.477	4.914
...

Sumber:

<https://www.bps.go.id/id/statistics-table/3/U0VOeFZEZFNIvNByUkdGMINrOTFVVGRRHY1ZkVGR6MDkjMw==/panjang-jalan-menurut-provinsi-dan-tingkat-kewenangan-pemerintahan--km---2021.html?year=2021>

Berapa km jalan di Indonesia?

Badan Pusat Statistik (BPS) mencatat, panjang jalan di Indonesia mencapai **549.161 kilometer** pada 2022.



Indonesia Baik

<https://indonesiabaik.id/infografis/provinsi-dengan-jalan-paling-banyak-di-indonesia/>

Temuan:

Merujuk pada data BPS tahun 2021, panjang jalan kab/kota di Provinsi **Nusa Tenggara Timur** hanya sepanjang **19,148 km**, sangat jauh dari data sebelumnya, yakni 7,157,929 km.

Terlebih, pada tahun 2022, BPS mencatat bahwa **total panjang jalan di Indonesia** hanyalah mencapai **549,161 km**.

Solusi:

- Jangka pendek: Mengganti data mencurigakan dengan sumber eksternal, misalnya BPS.
- Jangka panjang: Melakukan pengecekan ulang terhadap penyedia data terkait, lalu melakukan evaluasi terhadap proses pengumpulan & penyajian data di website yang pemakaiannya bersifat publik sehingga validitas data wajib dipastikan terlebih dahulu.

Step 8: Data Validation (Cek Validitas Angka)

Data: Kondisi Permukaan Jalan Kab/Kota 2021

Permasalahan: Misalkan Terdapat data **total panjang jalan mencurigakan** pada beberapa provinsi.

Kd_Prov	Provinsi	Kondisi_Baik	Baik_%	Kondisi_Sedan	Sedang_%	Kondisi_RR	RR_%	Kondisi_RB	RB_%	Total_Panjang
11	Aceh	7,583.6	39.2	3,092.3	16.0	5,471.0	28.3	3,207.0	16.6	19,353.8
12	Sumatera Utara	11,258.9	37.1	6,504.0	21.4	5,062.2	16.7	7,518.9	24.8	30,343.9
13	Sumatera Barat	6,725,646.0	42.8	2,871,269.0	18.3	2,712,690.0	17.3	3,400,494.0	21.7	15,710,099.0
14	Riau	5,605.5	27.6	6,523.7	32.1	4,764.4	23.5	3,419.0	16.8	20,312.6
15	Jambi	2,456.0	31.7	2,334.8	30.1	1,877.9	24.2	1,087.4	14.0	7,755.9
16	Sumatera Selatan	4,157,621.0	33.9	4,408,226.0	35.9	2,430,156.0	19.8	1,285,869.0	10.5	12,281,872.0
17	Bengkulu	2,239.1	32.5	1,775.9	25.8	2,002.1	29.1	865.9	12.6	6,882.9
18	Lampung	4,878.1	27.5	5,206.9	29.3	5,539.1	31.2	2,150.1	12.1	17,774.2
19	Kepulauan Bangka Belitung	2,025.0	70.0	314.5	10.9	460.5	15.9	91.7	3.2	2,891.7
21	Kepulauan Riau	2,120.4	46.0	1,332.9	28.9	927.4	20.1	225.5	4.9	4,606.1
32	Jawa Barat	15,221.0	62.6	4,454.2	18.3	2,371.6	9.8	2,255.4	9.3	24,302.1
33	Jawa Tengah	14,826.7	55.6	6,739.0	25.3	2,920.9	11.0	2,167.0	8.1	26,653.6
34	D.I. Yogyakarta	1,744.7	51.9	798.1	23.7	414.1	12.3	404.8	12.0	3,361.6
35	Jawa Timur	20,198.6	64.0	6,097.8	19.3	3,257.4	10.3	1,994.5	6.3	31,548.2
36	Banten	1,651.2	62.5	337.7	12.8	300.6	11.4	354.7	13.4	2,644.2
51	Bali	4,387.6	61.2	1,440.4	20.1	587.7	8.2	750.4	10.5	7,166.1
52	Nusa Tenggara Barat	3,366.7	55.3	639.3	10.5	1,026.4	16.9	1,052.7	17.3	6,085.0
53	Nusa Tenggara Timur	7,157,929.0	46.3	1,433,935.0	9.3	1,412,917.0	9.1	5,465,101.0	35.3	15,469,882.0
61	Kalimantan Barat	2,934,040.0	18.3	5,417,221.0	33.8	4,116,683.0	25.6	3,584,853.0	22.3	16,052,797.0
62	Kalimantan Tengah	3,773.7	25.7	3,823.5	26.1	2,776.8	18.9	4,289.1	29.3	14,663.1
63	Kalimantan Selatan	4,413.8	38.5	3,159.7	27.5	2,256.1	19.7	1,650.7	14.4	11,480.3
64	Kalimantan Timur	2,838.8	30.4	3,135.2	33.6	2,420.3	26.0	932.3	10.0	9,326.5
65	Kalimantan Utara	776.8	31.5	592.2	24.0	641.2	26.0	454.2	18.4	2,464.4
71	Sulawesi Utara	3,799.3	55.0	960.8	13.9	845.5	12.2	1,305.6	18.9	6,911.1
72	Sulawesi Tengah	2,900.1	22.5	4,265.3	33.1	2,578.2	20.0	3,147.0	24.4	12,890.6
73	Sulawesi Selatan	11,098.3	43.9	5,318.2	21.0	5,962.4	23.6	2,933.7	11.6	25,312.6
74	Sulawesi Tenggara	2,796.5	29.5	3,250.8	34.3	2,099.4	22.1	1,340.0	14.1	9,486.6
75	Gorontalo	1,295,191.0	40.9	662,470.0	20.9	491,834.0	15.5	721,436.0	22.8	3,170,931.0
76	Sulawesi Barat	1,314.3	27.8	1,323.2	28.0	1,242.9	26.3	842.3	17.8	4,722.7
81	Maluku	2,111.0	27.6	2,146.3	28.1	1,362.7	17.8	2,018.4	26.4	7,638.5
82	Maluku Utara	1,485.0	30.5	1,134.1	23.3	1,042.5	21.4	1,212.7	24.9	4,874.3
91	Papua Barat	1,307.4	13.9	3,531.5	37.7	1,923.8	20.5	2,616.6	27.9	9,379.3
94	Papua	2,244.0	16.3	5,668.5	41.1	2,696.6	19.5	3,192.5	23.1	13,801.6

Temuan:

Ternyata, Nusa Tenggara Timur, terdapat beberapa provinsi dengan total panjang jalan mencurigakan (lebih dari total panjang jalan di Indonesia tahun 2022 versi BPS, yakni 549,161 km):

- **Sumatera Barat (15,710,099 km)**
- **Sumatera Selatan (12,281,872 km)**
- **Kalimantan Barat (16,052,797 km)**
- **Gorontalo (3,170,931km)**

Step 9: Data Validation (Cek Konsistensi Antar Kolom)

Data: Kondisi Permukaan Jalan Kab/Kota 2021

Permasalahan: Misalkan Terdapat potensi ketidaksesuaian total panjang jalan pada beberapa provinsi.

Kd_Prov	Provinsi	Kondisi_Baik	Baik_%	Kondisi_Sedang	Sedang_%	Kondisi_RR	RR_%	Kondisi_RB	RB_%	Total_Panjang	Kesesuaian Panjang
11	Aceh	7,583.6	39.2	3,092.3	16.0	5,471.0	28.3	3,207.0	16.6	19,353.8	Sesuai
12	Sumatera Utara	11,258.9	37.1	6,504.0	21.4	5,062.2	16.7	7,516.9	24.8	30,343.9	Sesuai
13	Sumatera Barat	6,725,646.0	42.8	2,871,269.0	18.3	2,712,690.0	17.3	3,400,494.0	21.7	15,710,099.0	Sesuai
14	Riau	5,605.5	27.6	6,523.7	32.1	4,764.4	23.5	3,419.0	16.8	20,312.6	Sesuai
15	Jambi	2,456.0	31.7	2,334.8	30.1	1,877.9	24.2	1,087.4	14.0	7,755.9	Tidak
16	Sumatera Selatan	4,157,621.0	33.9	4,408,226.0	35.9	2,430,156.0	19.8	1,285,869.0	10.5	12,281,872.0	Sesuai
17	Bengkulu	2,239.1	32.5	1,775.9	25.8	2,002.1	29.1	865.9	12.6	6,882.9	Sesuai
18	Lampung	4,878.1	27.5	5,206.9	29.3	5,539.1	31.2	2,150.1	12.1	17,774.2	Sesuai
19	Kepulauan Bangka Belitung	2,025.0	70.0	314.5	10.9	460.5	15.9	91.7	3.2	2,891.7	Sesuai
21	Kepulauan Riau	2,120.4	46.0	1,332.9	28.9	927.4	20.1	225.5	4.9	4,606.1	Sesuai
32	Jawa Barat	15,221.0	62.6	4,454.2	18.3	2,371.6	9.8	2,255.4	9.3	24,302.1	Sesuai
33	Jawa Tengah	14,826.7	55.6	6,739.0	25.3	2,920.9	11.0	2,167.0	8.1	26,653.6	Tidak
34	D.I. Yogyakarta	1,744.7	51.9	798.1	23.7	414.1	12.3	404.8	12.0	3,361.6	Sesuai
35	Jawa Timur	20,198.6	64.0	6,097.8	19.3	3,257.4	10.3	1,994.5	6.3	31,548.2	Tidak
36	Banten	1,651.2	62.5	337.7	12.8	300.6	11.4	354.7	13.4	2,644.2	Sesuai
51	Bali	4,387.6	61.2	1,440.4	20.1	587.7	8.2	750.4	10.5	7,166.1	Sesuai
52	Nusa Tenggara Barat	3,366.7	55.3	639.3	10.5	1,026.4	16.9	1,052.7	17.3	6,085.0	Tidak
53	Nusa Tenggara Timur	7,157,929.0	46.3	1,433,935.0	9.3	1,412,917.0	9.1	5,465,101.0	35.3	15,469,882.0	Sesuai
61	Kalimantan Barat	2,934,040.0	18.3	5,417,221.0	33.8	4,116,683.0	25.6	3,584,853.0	22.3	16,052,797.0	Sesuai
62	Kalimantan Tengah	3,773.7	25.7	3,823.5	26.1	2,776.8	18.9	4,289.1	29.3	14,663.1	Tidak
63	Kalimantan Selatan	4,413.8	38.5	3,159.7	27.5	2,256.1	19.7	1,650.7	14.4	11,480.3	Sesuai
64	Kalimantan Timur	2,838.8	30.4	3,135.2	33.6	2,420.3	26.0	932.3	10.0	9,326.5	Sesuai
65	Kalimantan Utara	776.8	31.5	592.2	24.0	641.2	26.0	454.2	18.4	2,464.4	Sesuai
71	Sulawesi Utara	3,799.3	55.0	960.8	13.9	845.5	12.2	1,305.6	18.9	6,911.1	Tidak
72	Sulawesi Tengah	2,900.1	22.5	4,265.3	33.1	2,578.2	20.0	3,147.0	24.4	12,890.6	Sesuai
73	Sulawesi Selatan	11,098.3	43.9	5,318.2	21.0	5,962.4	23.6	2,933.7	11.6	25,312.6	Sesuai
74	Sulawesi Tenggara	2,796.5	29.5	3,250.8	34.3	2,099.4	22.1	1,340.0	14.1	9,486.6	Tidak
75	Gorontalo	1,295,191.0	40.9	662,470.0	20.9	491,834.0	15.5	721,436.0	22.6	3,170,931.0	Sesuai
76	Sulawesi Barat	1,314.3	27.8	1,323.2	28.0	1,242.9	26.3	842.3	17.8	4,722.7	Sesuai
81	Maluku	2,111.0	27.6	2,146.3	28.1	1,362.7	17.8	2,018.4	26.4	7,638.5	Tidak
82	Maluku Utara	1,485.0	30.5	1,134.1	23.3	1,042.5	21.4	1,212.7	24.9	4,874.3	Sesuai
91	Papua Barat	1,307.4	13.9	3,531.5	37.7	1,923.8	20.5	2,616.6	27.9	9,379.3	Sesuai
94	Papua	2,244.0	16.3	5,668.5	41.1	2,696.6	19.5	3,192.5	23.1	13,801.6	Sesuai

Temuan:

Kolom “Kesesuaian Panjang” pada dataset seharusnya merupakan penjumlahan antara panjang jalan dengan Kondisi_Baik + Kondisi_Sedang + Kondisi_RR + Kondisi_RB.

Akan tetapi, ditemukan beberapa provinsi yang tidak sesuai, namun semuanya justru merupakan provinsi yang total panjang jalan kab/kota-nya tidak mencurigakan.

Setelah ditelusuri, ternyata ketidaksesuaian ini hanya dikarenakan pembulatan desimal saja. Hal ini dapat terjadi karena input angka dilakukan dengan pembulatan yang berbeda pada setiap kondisi jalannya. Ketika pembulatan dilakukan dengan ceiling atau flooring, maka total panjang semua provinsi menjadi sesuai. Oleh karena itu, ketidaksesuaian ini dapat diabaikan saja.

Step 10: Data Validation (Cek Validitas Angka)

Data: Jenis Permukaan Jalan Kab/Kota 2021 & 2022

Permasalahan: Misalkan Data Tahun 2022 memiliki potensi kesalahan pencatatan.

2021

Kd_Prov	Provinsi	Paved	Paved_%	UnPaved	UnPaved_%	Jumlah
11	Aceh	9.824,6	50,8	9.529,2	49,2	19.353,8
12	Sumatera Utara	13.754,2	45,3	16.589,5	54,7	30.343,7
13	Sumatera Barat	10.201,7	59,7	7.021,8	40,8	17.223,4
14	Riau	8.344,6	41,1	11.968,1	58,9	20.312,6
21	Kepulauan Riau	2.644,2	57,4	1.961,9	42,6	4.606,1
15	Jambi	3.848,8	49,6	3.907,1	50,4	7.755,9
17	Bengkulu	2.965,5	43,1	3.917,4	56,9	6.882,9
16	Sumatera Selatan	8.373,9	64,9	4.527,7	35,1	12.901,6
19	Kepulauan Bangka Belitung	2.455,5	84,9	436,2	15,1	2.891,7
18	Lampung	8.839,5	49,7	8.934,6	50,3	17.774,1
31	DKI Jakarta	0,0	0,0	0,0	0,0	0,0
36	Banten	3.569,2	88,0	487,7	12,0	4.056,9
32	Jawa Barat	20.941,0	86,2	3.361,2	13,8	24.302,1
33	Jawa Tengah	23.106,0	86,7	3.547,7	13,3	26.653,6
34	DI Yogyakarta	2.526,6	75,2	835,0	24,8	3.361,6
35	Jawa Timur	25.757,8	81,7	5.790,5	18,4	31.548,3
61	Kalimantan Barat	3.371,3	20,2	13.362,6	79,9	16.733,9
62	Kalimantan Tengah	4.621,9	31,5	10.041,1	68,5	14.663,1
64	Kalimantan Timur	4.166,9	44,7	5.159,7	55,3	9.326,5
65	Kalimantan Utara	892,5	36,2	1.571,9	63,8	2.464,4
63	Kalimantan Selatan	6.907,9	60,2	4.572,4	39,8	11.480,3
51	Bali	6.099,5	85,1	1.066,6	14,9	7.166,1
52	Nusa Tenggara I	4.181,4	68,7	1.903,6	31,3	6.085,0
53	Nusa Tenggara II	4.066,1	25,5	11.853,3	74,5	15.919,3
71	Sulawesi Utara	4.474,4	64,7	2.436,8	35,3	6.911,1
75	Gorontalo	1.757,3	47,6	1.931,8	52,4	3.689,1
72	Sulawesi Tengah	3.155,1	24,5	9.735,5	75,5	12.890,6
76	Sulawesi Barat	1.900,8	40,3	2.821,9	59,8	4.722,7
73	Sulawesi Selatan	15.939,8	63,0	9.372,7	37,0	25.312,6
74	Sulawesi Tenggara	3.636,6	38,4	5.948,4	61,7	9.486,6
81	Maluku	1.932,0	25,3	5.706,5	74,7	7.638,5
82	Maluku Utara	1.442,0	29,6	3.432,3	70,4	4.874,3
94	Papua	3.232,2	23,4	10.569,4	76,6	13.801,6
91	Papua Barat	1.665,7	17,8	7.713,6	82,2	9.379,3

2022

Kode Prov	Provinsi	Paved	Paved_%	UnPaved	UnPaved_%	Jumlah
11	Aceh	8.848,1	55,2	7.189,2	44,8	16.037,3
12	Sumatera Utara	19.186,0	62,5	11.513,0	37,5	30.699,0
13	Sumatera Barat	11.851,1	64,6	6.503,1	35,4	18.354,2
14	Riau	7.601,3	49,3	7.811,2	50,7	15.412,5
15	Jambi	3.417,4	47,2	3.819,8	52,8	7.237,2
16	Sumatera Selatan	9.946,0	63,6	5.682,9	36,4	15.628,9
17	Bengkulu	3.670,0	62,6	2.193,4	37,4	5.863,3
18	Lampung	12.066,4	67,9	5.707,8	32,1	17.774,2
19	Kep. Bangka Belitung	3.915,8	85,4	667,2	14,6	4.583,1
21	Kep. Riau	2.286,6	54,7	1.896,5	45,3	4.183,1
31	DKI Jakarta	2.496,7	91,8	224,3	8,2	2.721,0
32	Jawa Barat	19.190,6	93,0	1.445,8	7,0	20.637,4
33	Jawa Tengah	24.142,3	91,9	2.121,0	8,1	26.263,4
34	DI Yogyakarta	2.924,0	88,9	364,5	11,1	3.288,5
35	Jawa Timur	29.070,9	92,0	2.541,3	8,0	31.612,2
36	Banten	2.220,7	97,7	52,8	2,3	2.273,5
51	Bali	6.688,1	93,5	465,1	6,5	7.163,2
52	Nusa Tenggara I	4.244,3	80,9	1.003,5	19,1	5.247,7
53	Nusa Tenggara II	11.321,7	65,2	6.032,3	34,8	17.354,0
61	Kalimantan Barat	5.847,9	39,1	9.099,8	60,9	14.947,7
62	Kalimantan Tengah	5.072,3	34,9	9.464,6	65,1	14.536,9
63	Kalimantan Selatan	8.072,7	70,7	3.339,5	29,3	11.412,2
64	Kalimantan Timur	4.649,4	49,9	4.677,1	50,1	9.326,5
65	Kalimantan Utara	1.178,5	47,8	1.285,9	52,2	2.464,4
71	Sulawesi Utara	4.896,0	74,3	1.690,7	25,7	6.586,7
72	Sulawesi Tengah	5.509,6	42,8	7.377,0	57,2	12.886,6
73	Sulawesi Selatan	18.436,7	60,7	11.924,5	39,3	30.361,2
74	Sulawesi Tenggara	3.930,5	46,8	4.470,6	53,2	8.401,1
75	Gorontalo	2.162,3	51,0	2.077,7	49,0	4.240,8
76	Sulawesi Barat	2.041,9	43,2	2.680,8	56,8	4.722,7
81	Maluku	3.144,4	44,2	3.964,5	55,8	7.108,9
82	Maluku Utara	1.873,1	42,8	2.507,2	57,2	4.380,3
91	Papua Barat	2.165,5	23,1	7.227,9	76,9	9.393,4
94	Papua	3.652,7	25,4	10.737,8	74,6	14.390,6

2022

Provinsi	Panjang Jalan Negara (km) (km)	Panjang Jalan Provinsi (km) (km)	Panjang Jalan Kabupaten/Kota (km) (km)	Jumlah Panjang Jalan(km) (km)
Sumatera Selatan	1.581	1.514	16.492	19.587
Bengkulu	782	1.563	6.883	9.228
Lampung	1.298	1.693	17.774	20.765
Kepulauan Bangka Belitung	599	851	4.441	5.891
Kepulauan Riau	589	896	4.203	5.688
DKI Jakarta	53	6.432	-	6.485
Jawa Barat	1.783	2.361	24.249	28.393
Jawa Tengah	1.581	2.501	26.737	30.819
DI Yogyakarta	307	760	3.382	4.449
Jawa Timur	2.262	1.421	38.739	42.422
Banten	568	762	4.385	5.715
Bali	590	743	7.323	8.656

Sumber:

<https://www.bps.go.id/id/statistics-table/3/U0VOeFZEZFNIvNByUkdGMNrOTFVVGRHY1ZkVGR6MDkjMw==/panjang-jalan-menurut-provinsi-dan-tingkat-kewenangan-pemerintahan--km---2021.html?year=2022>

Temuan:

Merujuk pada data BPS, DKI Jakarta tidak memiliki ruas jalan kab/kota sehingga panjangnya adalah nol (0 km). Sedangkan, pada website PUPR, DKI Jakarta pada tahun 2022 memiliki persentase jalan diaspal (Paved) dan tidak diaspal (Unpaved), yang artinya memiliki jalan kab/kota.

Solusi:

Diperlukan kajian lebih lanjut untuk menentukan sumber yang dapat digunakan selanjutnya. Sebagai informasi, BPS menuliskan bahwa sumber data yang mereka gunakan adalah Keputusan Menteri Pekerjaan Umum dan Perumahan Rakyat No. 430/KPTS/M/2022.

Step 11: Data Verification (Cek Kelengkapan)

Data:

Panjang Kemantapan Jembatan Nasional Tahun 2020-2023

Permasalahan:

Misalkan Data Tahun **2022 tidak tersedia.**

Data and Resources

**Panjang Kemantapan Jembatan Nasional Tahun 2023**
Resource ini berisi data Panjang Kemantapan Jembatan Nasional di Indonesia...
[Pratinjau](#) [Unduh](#)

**Panjang Kemantapan Jembatan Nasional Tahun 2021**
Resource ini berisi data Panjang Kemantapan Jembatan Nasional...
[Pratinjau](#) [Unduh](#)

**Panjang Kemantapan Jembatan Nasional Tahun 2020**
Resource ini berisi data Panjang Kemantapan Jembatan Nasional...
[Pratinjau](#) [Unduh](#)

Solusi:

- Untuk keperluan **analisis**, dapat melakukan **data imputation***, yakni mengisi missing values dengan modus (untuk data kategorik), ataupun mean/median/modus (untuk data numerik).
**Data Imputation sebaiknya dilakukan apabila data tersedia dalam jangka waktu yang cukup panjang, dengan tujuan agar hasil imputation tidak mempengaruhi kesimpulan secara signifikan. Tujuan data imputation di sini hanya untuk menghindari error saat melakukan analisis data karena adanya missing values.*
- Untuk keperluan **dashboard/visualisasi**, dapat **dibiarkan saja**. Hal ini lumrah dilakukan, sebab dashboard/visualisasi umumnya memiliki fitur filter untuk memisahkan dataset berdasarkan waktu.

Step 12: Data Validation (Cek Format Data)

Data:

Panjang Kemantapan Jembatan Nasional Tahun 2023

Kd_Prov	PROVINSI	Total_Panjang	Mantap_m	Mantap_%	TMantap_m	TMantap_%
11	ACEH	25427,75	22526,7	88,59	2901,05	11,41
12	SUMATERA UTARA	24169,9	21410,7	88,58	2759,2	11,42
13	SUMATERA BARAT	18.361.846,0	17.344.346,0	94,46	1017,5	5,54
14	RIAU	15785,8	14538,8	92,1		1.247,0 7,9
15	JAMBI	11017,8	9441,8	85,7		1.576,0 14,3
16	SUMATERA SELATAN	18807,16	12863,81	68,4	5943,35	31,6
17	BENGKULU	9754,2		8.531,0 87,46	1223,2	12,54
18	LAMPUNG	11112,3	9265,45	83,38	1846,85	16,62
19	KEPULAUAN BANGKA BELITUNG	2881,28	2437,78	84,61	443,5	15,39
21	KEPULAUAN RIAU	7425,2		7.339,0 98,84	86,2	1,16
32	JAWA BARAT	34919,7	32742,9	93,77	2176,8	6,23
33	JAWA TENGAH	30507,55	28910,95	94,77	1596,6	5,23
34	DI YOGYAKARTA	6334,4	6334,4		100,0	0,0
35	JAWA TIMUR	32222,63	28827,94	89,46	3394,69	10,54
36	BANTEN	6755,4		6.452,0 95,51	303,4	4,49
51	BALI	9577,95	8907,8	93,84	590,15	6,16
52	NUSA TENGGARA BARAT	8330,3	7685,8	92,26	644,5	7,74
53	NUSA TENGGARA TIMUR	10980,6		8.673,0 78,98	2307,6	21,02
61	KALIMANTAN BARAT	25042,86	21120,97	84,34	3921,89	15,66
62	KALIMANTAN TENGAH	30799,38	18103,58	58,78	12695,8	41,22
63	KALIMANTAN SELATAN	19074,7	14976,11	78,51	4098,59	21,49
64	KALIMANTAN TIMUR	8287,4	7728,1	93,25	559,3	6,75
65	KALIMANTAN UTARA		6.019,0	5.652,0 93,9		367,0 6,1
71	SULAWESI UTARA	16137,8	14219,4	88,11	1918,4	11,89
72	SULAWESI TENGAH	22422,28	8905,91	39,72	13516,37	60,28
73	SULAWESI SELATAN	15658,9	12134,45	77,49	3524,45	22,51
74	SULAWESI TENGGARA	13404,67	13111,17	97,81	293,5	2,19
75	GORONTALO	6201,5	4693,5	75,68		1.508,0 24,32
76	SULAWESI BARAT	7858,9	6316,8	80,38	1542,1	19,62
81	MALUKU	25240,9	21289,5	84,35	3951,4	15,65
82	MALUKU UTARA	13566,96	11000,96	81,09		2.566,0 18,91
91	PAPUA	29704,6	23598,15	79,44	6106,45	20,56
92	PAPUA BARAT	12793,1	9249,5	72,3	3543,6	27,7

Permasalahan:

Misalkan terdapat **Format penulisan angka tidak sesuai**, seharusnya desimal dituliskan dengan titik (.), bukan koma (,). Oleh karena itu, sebagian besar nilai yang dituliskan dengan koma (,) terbaca sebagai teks dan terlihat rata kiri (*diberikan warna merah pada screenshot*).

Kemungkinan Penyebab:

- Adanya **pergantian software** pada tahun sebelumnya yang memiliki **default berbeda** untuk angka.
- Pergantian standar baku penulisan** dari internasional (.) menjadi standar baku penulisan bahasa Indonesia, yaitu (,).

Solusi:

- Jangka pendek:** Melakukan **find & replace** koma (,) menjadi titik (.), lalu membuat **dokumentasi & komunikasi** mengenai kasus ini terhadap tim terkait.
- Jangka panjang:** Menerapkan **standarisasi** format data, melakukan proses **otomasi ETL**, serta mengadakan **sosialisasi & pelatihan**.

Step 13: Data Validation (Cek Format Data)

Data:

Panjang Kemantapan Jembatan Nasional Tahun 2023

Sebelum

Kd_Prov	PROVINSI	Total_Panjang	Mantap_m	Mantap_%	TMantap_m	TMantap_%
11	ACEH	25427.75	22526.7	88.59	2901.05	11.41
12	SUMATERA UTARA	24169.9	21410.7	88.58	2759.2	11.42
13	SUMATERA BARAT	18.361.846.0	17.344.346.0	94.46	1017.5	5.54
14	RIAU	15785.8	14538.8	92.1		1.247.0 7.9
15	JAMBI	11017.8	9441.8	85.7		1.576.0 14.3
16	SUMATERA SELATAN	18807.16	12863.81	68.4	5943.35	31.6
17	BENGKULU	9754.2		8.531.0 87.46	1223.2	12.54
18	LAMPUNG	11112.3	9265.45	83.38	1846.85	16.62
19	KEPULAUAN BANGKA BELITUNG	2881.28	2437.78	84.61	443.5	15.39
21	KEPULAUAN RIAU	7425.2		7.339.0 98.84	86.2	1.16
32	JAWA BARAT	34919.7	32742.9	93.77	2176.8	6.23
33	JAWA TENGAH	30507.55	28910.95	94.77	1596.6	5.23
34	DI YOGYAKARTA	6334.4	6334.4		100.0	0.0
35	JAWA TIMUR	32222.63	28827.94	89.46	3394.69	10.54
36	BANTEN	6755.4		6.452.0 95.51	303.4	4.49
51	BALI	9577.95	8987.8	93.84	590.15	6.16
52	NUSA TENGGARA BARAT	8330.3	7685.8	92.26	644.5	7.74
53	NUSA TENGGARA TIMUR	10980.6		8.673.0 78.98	2307.6	21.02
61	KALIMANTAN BARAT	25042.86	21120.97	84.34	3921.89	15.66
62	KALIMANTAN TENGAH	30799.38	18103.58	58.78	12695.8	41.22
63	KALIMANTAN SELATAN	19074.7	14976.11	78.51	4098.59	21.49
64	KALIMANTAN TIMUR	8287.4	7728.1	93.25	559.3	6.75
65	KALIMANTAN UTARA	6.019.0	5.652.0	93.9		367.0 6.1
71	SULAWESI UTARA	16137.8	14219.4	88.11	1918.4	11.89
72	SULAWESI TENGAH	22422.28	8905.91	39.72	13516.37	60.28
73	SULAWESI SELATAN	15658.9	12134.45	77.49	3524.45	22.51
74	SULAWESI TENGGARA	13404.67	13111.17	97.81		293.5 2.19
75	GORONTALO	6201.5	4693.5	75.68		1.508.0 24.32
76	SULAWESI BARAT	7858.9	6316.8	80.38	1542.1	19.62
81	MALUKU	25240.9	21289.5	84.35	3951.4	15.65
82	MALUKU UTARA	13566.96	11000.96	81.09		2.566.0 18.91
91	PAPUA	29704.6	23598.15	79.44	6106.45	20.56
92	PAPUA BARAT	12793.1	9249.5	72.3	3543.6	27.7

Sesudah (find & replace)

Kd_Prov	PROVINSI	Total_Panjang	Mantap_m	Mantap_%	TMantap_m	TMantap_%
11	ACEH	25.427.8	22.526.7	88.6	2.901.1	11.4
12	SUMATERA UTARA	24.169.9	21.410.7	88.6	2.759.2	11.4
13	SUMATERA BARAT	18.361.846.0	17.344.346.0	94.5	1.017.5	5.5
14	RIAU	15.785.8	14.538.8	92.1	1.247.0	7.9
15	JAMBI	11.017.8	9.441.8	85.7	1.576.0	14.3
16	SUMATERA SELATAN	18.807.2	12.863.8	68.4	5.943.4	31.6
17	BENGKULU	9.754.2	8.531.0	87.5	1.223.2	12.5
18	LAMPUNG	11.112.3	9.265.5	83.4	1.846.9	16.6
19	KEPULAUAN BANGKA BELITUNG	2.881.3	2.437.8	84.6	443.5	15.4
21	KEPULAUAN RIAU	7.425.2	7.339.0	98.8	86.2	1.2
32	JAWA BARAT	34.919.7	32.742.9	93.8	2.176.8	6.2
33	JAWA TENGAH	30.507.6	28.911.0	94.8	1.596.6	5.2
34	DI YOGYAKARTA	6.334.4	6.334.4	100.0	0.0	0.0
35	JAWA TIMUR	32.222.6	28.827.9	89.5	3.394.7	10.5
36	BANTEN	6.755.4	6.452.0	95.5	303.4	4.5
51	BALI	9.577.95	8.987.8	93.8	590.2	6.2
52	NUSA TENGGARA BARAT	8.330.3	7.685.8	92.3	644.5	7.7
53	NUSA TENGGARA TIMUR	10.980.6	8.673.0	79.0	2.307.6	21.0
61	KALIMANTAN BARAT	25.042.9	21.121.0	84.3	3.921.9	15.7
62	KALIMANTAN TENGAH	30.799.4	18.103.6	58.8	12.695.8	41.2
63	KALIMANTAN SELATAN	19.074.7	14.976.1	78.5	4.098.6	21.5
64	KALIMANTAN TIMUR	8.287.4	7.728.1	93.3	559.3	6.8
65	KALIMANTAN UTARA	6.019.0	5.652.0	93.9	367.0	6.1
71	SULAWESI UTARA	16.137.8	14.219.4	88.1	1.918.4	11.9
72	SULAWESI TENGAH	22.422.3	8.905.9	39.7	13.516.4	60.3
73	SULAWESI SELATAN	15.658.9	12.134.5	77.5	3.524.5	22.5
74	SULAWESI TENGGARA	13.404.7	13.111.2	97.8	293.5	2.2
75	GORONTALO	6.201.5	4.693.5	75.7	1.508.0	24.3
76	SULAWESI BARAT	7.858.9	6.316.8	80.4	1.542.1	19.6
81	MALUKU	25.240.9	21.289.5	84.4	3.951.4	15.7
82	MALUKU UTARA	13.567.0	11.001.0	81.1	2.566.0	18.9
91	PAPUA	29.704.6	23.598.2	79.4	6.106.5	20.6
92	PAPUA BARAT	12.793.1	9.249.5	72.3	3.543.6	27.7



Step 14: Data Verification (Cek Format Data)

Data: Panjang Kemantapan Jembatan Nasional Tahun 2023

Permasalahan: Misalkan Terdapat salah pembacaan desimal pada data Provinsi Sumatera Barat ketika diunduh ke format csv yang membuat angka terlihat sangat besar dibandingkan provinsi lainnya sehingga cenderung tidak masuk akal.

Website PUPR

Panjang Kemantapan Jembatan Nasional tahun 2023.csv						
Data Preview: Note that by default the preview only displays up to 100 records. Use the pager to flip through more records or adjust the start and end fields to display the number of records you wish to see.						
Grid	Graph	33 records	1 - 33	Search data ...	Go »	Filters Fields
Kd_Prov	PROVINSI	Total_Panjang	Mantap_mm	Mantap_%	TMatap_mm	TMatap_%
11	ACEH	25427,75	22526,7	88,59	2901,...	11,41
12	SUMATERA UTARA	24169,9	21410,7	88,58	2759,2	11,42
13	SUMATERA BARAT	18361,846	17344,346	94,46	1017,5	5,54
14	RIAU	15785,8	14538,8	92,1	1247	7,9
15	JAMBI	11017,8	9441,8	85,7	1576	14,3
16	SUMATERA SELATAN	18807,16	12863,81	68,4	5943,...	31,6
17	BENGKULU	9754,2	8531	87,46	1223,2	12,54
18	LAMPUNG	11112,3	9265,45	83,38	1846,...	16,62
19	KEPULAUAN BANGKA BELITUNG	2881,28	2437,78	84,61	443,5	15,39
21	KEPULAUAN RIAU	7425,2	7339	98,84	86,2	1,16
32	JAWA BARAT	34919,7	32742,9	93,77	2176,8	6,23
33	JAWA TENGAH	30507,55	28910,95	94,77	1596,6	5,23
34	DI YOGYAKARTA	6334,4	6334,4	100	0	0
35	JAWA TIMUR	32222,63	28827,94	89,46	3394,...	10,54
36	BANTEN	6755,4	6452	95,51	303,4	4,49
51	BALI	9577,95	8987,8	93,84	590,15	6,16
52	NUSA TENGGARA BARAT	8330,3	7685,8	92,26	644,5	7,74
53	NUSA TENGGARA TIMUR	10980,6	8673	78,98	2307,6	21,02
61	KALIMANTAN BARAT	25042,86	21120,97	84,34	3921,...	15,66
62	KALIMANTAN TENGAH	30799,38	18103,58	58,78	12695,...	41,22
63	KALIMANTAN SELATAN	10071,7	14074,44	70,44	4000	21,60



Spreadsheet						
Kd_Prov	PROVINSI	Total_Panjang	Mantap_mm	Mantap_%	TMatap_mm	TMatap_%
11	ACEH	25.427,8	22.526,7	88,6	2.901,1	11,4
12	SUMATERA UTARA	24.169,9	21.410,7	88,6	2.759,2	11,4
13	SUMATERA BARAT	18.361,846	17.344,346	94,5	1.017,5	5,5
14	RIAU	15.785,8	14.538,8	92,1	1.247,0	7,9
15	JAMBI	11.017,8	9.441,8	85,7	1.576,0	14,3
16	SUMATERA SELATAN	18.807,2	12.863,8	68,4	5.943,4	31,6
17	BENGKULU	9.754,2	8.531,0	87,5	1.223,2	12,5
18	LAMPUNG	11.112,3	9.265,5	83,4	1.846,9	16,6
19	KEPULAUAN BANGKA BELITUNG	2.881,3	2.437,8	84,6	443,5	15,4
21	KEPULAUAN RIAU	7.425,2	7.339,0	98,8	86,2	1,2
32	JAWA BARAT	34.919,7	32.742,9	93,8	2.176,8	6,2
33	JAWA TENGAH	30.507,6	28.911,0	94,8	1.596,6	5,2
34	DI YOGYAKARTA	6.334,4	6.334,4	100,0	0,0	0,0
35	JAWA TIMUR	32.222,63	28.827,94	89,46	3.394,7	10,5
36	BANTEN	32.226,6	28.827,9	89,5	3.394,7	10,5
51	BALI	6.755,4	6.452,0	95,5	303,4	4,5
52	NUSA TENGGARA BARAT	9.578,0	8.987,8	93,8	590,2	6,2
53	NUSA TENGGARA TIMUR	8.330,3	7.685,8	92,3	644,5	7,7
61	KALIMANTAN BARAT	10.980,6	8.673,0	79,0	2.307,6	21,0
62	KALIMANTAN TENGAH	25.042,9	21.121,0	84,3	3.921,9	15,7
63	KALIMANTAN SELATAN	30.799,4	18.103,6	58,8	12.695,8	41,2
64	KALIMANTAN TIMUR	19.074,7	14.976,1	78,5	4.098,6	21,5
65	KALIMANTAN UTARA	8.287,4	7.728,1	93,3	559,3	6,8
71	SULAWESI UTARA	6.019,0	5.652,0	93,9	367,0	6,1
72	SULAWESI TENGAH	16.137,8	14.219,4	88,1	1.918,4	11,9
73	SULAWESI SELATAN	22.422,3	8.905,9	39,7	13.516,4	60,3
74	SULAWESI TENGGARA	15.658,9	12.134,5	77,5	3.524,5	22,5
75	GORONTALO	13.404,7	13.111,2	97,8	293,5	2,2
76	SULAWESI BARAT	6.201,5	4.693,5	75,7	1.508,0	24,3
81	MALUKU	7.858,9	6.316,8	80,4	1.542,1	19,6
82	MALUKU UTARA	25.240,9	21.289,5	84,4	3.951,4	15,7
91	PAPUA	13.567,0	11.001,0	81,1	2.566,0	18,9
92	PAPUA BARAT	29.704,6	23.598,2	79,4	6.106,5	20,6
93	PAPUA BARAT	12.793,1	9.249,5	72,3	3.543,6	27,7

Temuan: Setelah diunduh ke format csv, format desimal data Provinsi Sumatera Barat berbeda dari yang lainnya, yakni 3 digit di belakang koma, sedangkan yang lainnya maksimal 2 digit di belakang koma.

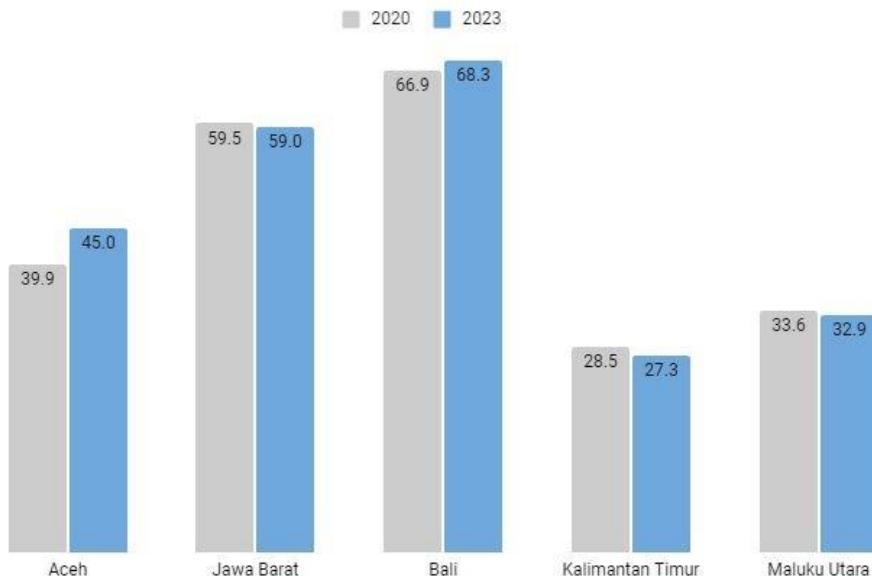
Solusi: Koreksi manual (jangka pendek), Evaluasi format data yang disediakan di website (jangka panjang)

Step 15: Data Exploration (Analisis Tren/Progres)

Data:

Kondisi Permukaan Jalan Kab/Kota 2020 & 2023 (5 Provinsi Tertentu)

Progres Persentase Jalan Kab/Kota dalam Kondisi Baik (2020 vs 2023)



Jenis Grafik:

Double Bar Chart.

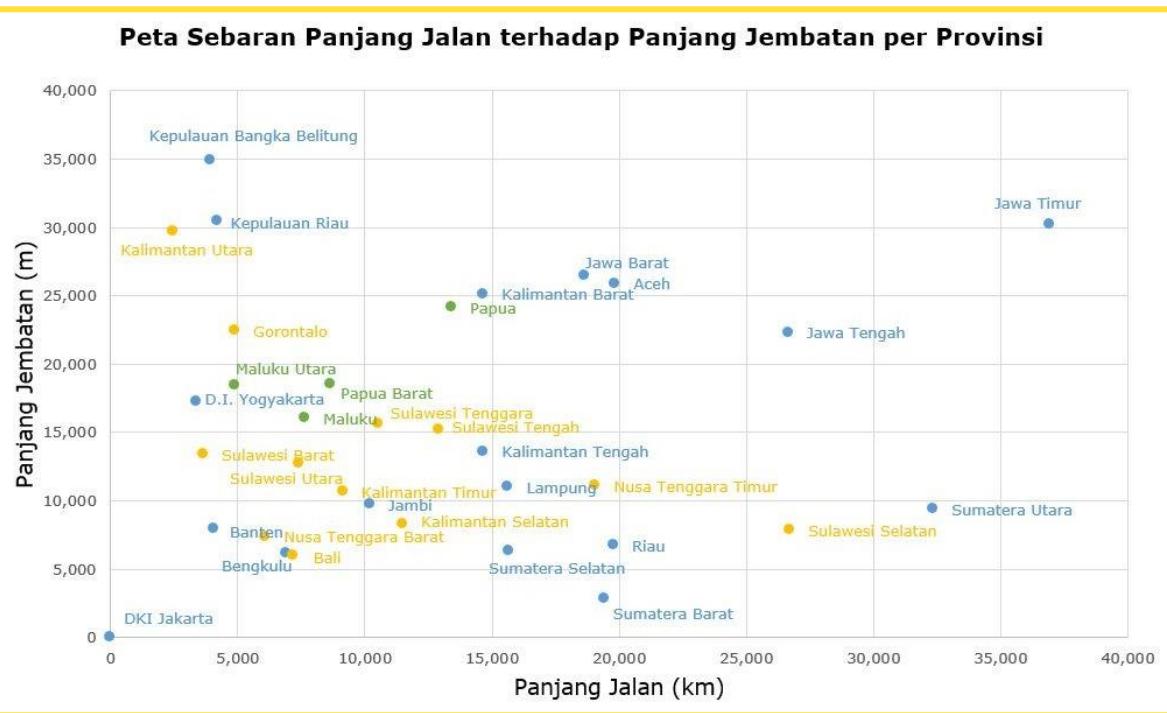
Insight:

- Dari 5 provinsi terpilih, ternyata **sebagian besar mengalami penurunan persentase** jalan kab/kota dalam kondisi baik. Hal ini dapat disebabkan oleh 2 hal:
 1. **Perawatan kurang baik** terhadap jalan yang sudah ada,
 2. **Penambahan jalan kab/kota baru**, namun belum dimaksimalkan kondisinya.
- Provinsi di wilayah **Jawa-Bali** memiliki persentase jalan dalam **kondisi baik tertinggi** dibandingkan provinsi di pulau-pulau lainnya.
- Ibu Kota Nusantara (**IKN**) yang bermarkas di Kalimantan Timur **perlu diperhatikan** agar dapat mendukung proses jalannya pemerintahan di masa mendatang.

Step 16: Data Exploration (Analisis Hubungan 2 Variabel)

Data:

Join Table antara Kondisi Permukaan Jalan Kab/Kota 2020 & Panjang Kemantapan Jembatan Nasional 2020



Jenis Grafik:

Scatterplot, dibedakan dengan warna per wilayah.

Insight:

- Secara persebaran, nampaknya wilayah **Indonesia Barat** memiliki **kondisi infrastruktur yang beragam**, tetapi **relatif lebih baik**, terlihat dari posisi titik warna biru yang **menyebar di ujung kanan** (panjang jalan terbesar) dan **ujung atas** (panjang jembatan terbesar) pada scatterplot.
- Wilayah **Indonesia Timur** cenderung memiliki karakteristik yang **relatif homogen**, ditandai dengan posisi titik warna hijau yang **berkumpul di tengah** scatterplot (panjang jalan 5-15rb km dan panjang jembatan 15-25rb m).

Step 17: Data Exploration (Analisis Hubungan 2 Variabel)

Data:

Join Table antara Kondisi Permukaan Jalan Kab/Kota 2020 & Panjang Kemantapan Jembatan Nasional 2020



Jenis Grafik:

Scatterplot, highlight pada provinsi tertentu.

Insight:

- Jawa Timur memiliki kondisi infrastruktur paling baik, ditandai dengan posisi titik yang berada di paling ujung kanan-atas pada scatterplot. Provinsi ini memiliki panjang jalan sekitar 37rb km dan panjang jembatan sekitar 30rb m.
- DKI Jakarta seakan memiliki kondisi infrastruktur paling buruk, ditandai dengan posisi titik yang berada di paling ujung kiri-bawah pada scatterplot, bahkan berada pada titik koordinat (0,0). Setelah ditelusuri, ternyata DKI Jakarta memang tidak memiliki jalan kab/kota, maupun jembatan nasional → Data Analyst memiliki peran penting untuk meluruskan hal seperti ini, atau bahkan hapus saja datanya!

Kita barusaja membuat

Deck EDA



31



32



33



34



35



36



37



38



39



40



41



42



43



44



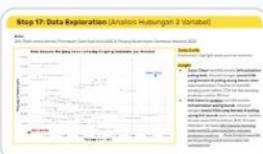
45



46



47



48



Documentation



Client Engagement



Build Trust



When You Skip EDA in Real Life..

Data Collection

Google Form, SurveyMonkey
Google Analytics, etc.



Python, R, Google Sheets,
Google BigQuery, MySQL, etc.



Decision Making

Orang dengan kemampuan
Stat/Math, Computer Science,
& Business Expertise



SAS Viya, Tableau, Power BI, Looker, QlikView, etc.

Data Preparation

Data Visualization

CONTACT ME



Domicile:
Jababeka Residence,
Bekasi Regency.



+62 852-1147-4471

M **miqbalpillie18@gmail.com**

THANK YOU & SEE YOU!