

Matakuliah Visualisasi Data  
Tanggal 05 12 2023

**Visualisasi Distribusi Air Bersih di Indonesia (2015, 2017-2021):  
Tren, Perbandingan, dan Tantangan Berdasarkan Data BPS**

Oleh

Rika Ajeng Finatih (121450036)

Veni Zahara Kartika (121450075)

Elsyah Sapyrah (121450096)

Program Studi Sains Data  
Institut Teknologi Sumatera  
Lampung Selatan  
2023

## SURAT PERNYATAAN BEBAS PLAGIAT

Saya yang bertanda tangan di bawah ini :

Nama : Rika Ajeng Finatih

NIM : 121450036

Jabatan : Ketua

Nama : Veni Zahara Kartika

NIM : 121450075

Jabatan : Anggota

Nama : Elsyah Sapyrah

NIM : 121450096

Jabatan : Anggota

Dengan ini menyatakan bahwa judul tugas Analisis Visual Distribusi Air Bersih di Indonesia (2018-2021): Tren, Perbandingan, dan Tantangan Berdasarkan Data BPS benar bebas dari plagiat, dan apabila pernyataan ini terbukti tidak benar maka saya bersedia menerima sanksi sesuai ketentuan yang berlaku yaitu **nilai tugas ini bernilai 0**.

Demikian surat pernyataan ini saya buat untuk dipergunakan sebagaimana mestinya.

Lampung Selatan, 02 Desember 2023

Yang membuat pernyataan,



Rika Ajeng Finatih

## **1. Definisi permasalahan**

Penelitian ini difokuskan pada permasalahan ketersediaan akses air bersih yang layak dan aman dikonsumsi, terutama di pedesaan Indonesia yang masih menghadapi tantangan signifikan dalam pelayanan air bersih. Data Badan Pusat Statistik (BPS) dari 2015, 2018-2021 menjadi dasar analisis untuk mengidentifikasi pola distribusi akses air bersih. Pemanfaatan data BPS memungkinkan pemahaman dinamika perubahan akses air bersih dari waktu ke waktu, memberikan gambaran yang akurat. Aspek kualitas air menjadi fokus, karena berdampak langsung pada kesehatan masyarakat. Integrasi temuan penelitian dengan data BPS dapat memberikan kontribusi berharga dalam merumuskan kebijakan yang efektif terkait penyediaan air bersih di wilayah Indonesia.

Pengembangan visualisasi dalam bentuk dasbor (dashboard) akan menjadi alat utama untuk mempresentasikan temuan penelitian secara komprehensif. Dasbor akan memvisualisasikan secara interaktif pola distribusi akses air bersih, perubahan seiring waktu, dan kualitas air di berbagai wilayah. Ini tidak hanya meningkatkan pemahaman akan data, tetapi juga memudahkan pemangku kebijakan dalam mengambil keputusan yang lebih tepat dan efektif untuk meningkatkan pelayanan air bersih di pedesaan Indonesia. Visualisasi ini akan memungkinkan para pengambil keputusan untuk melihat dan memahami dinamika permasalahan secara lebih cepat dan efisien.

## **2. Purpose & Parameter**

### **2.1 Latar Belakang**

Distribusi air bersih merupakan salah satu aspek penting dalam menjaga kesehatan dan keberlangsungan hidup masyarakat. Akses terhadap air bersih masih menjadi tantangan utama, terutama di daerah pedesaan Indonesia. Meskipun ada upaya dari pemerintah dan berbagai pihak terkait untuk meningkatkan akses terhadap air bersih, masih terdapat disparitas yang signifikan antara daerah perkotaan dan pedesaan, serta antara pulau-pulau di Indonesia.

Berdasarkan permasalahan ketersediaan akses air bersih yang layak dan aman dikonsumsi yang sudah dijelaskan sebelumnya. Oleh karena itu, penelitian kami kali ini akan menganalisis dan memvisualisasikan beberapa data dalam bentuk dashboard.

Visualisasi Menurut Mesterjon (2012:171) “merupakan rekayasa dalam pembuatan gambar, diagram atau animasi untuk menampilkan suatu informasi. Berdasarkan beberapa pengertian tersebut visualisasi adalah segala sesuatu yang dirancang dalam bentuk rekayasa berupa gambar, tulisan, grafik, ataupun gerakan [1].

Data distribusi air bersih yang diterbitkan oleh Badan Pusat Statistik (BPS) dari tahun 2015, 2017 - 2021 menjadi sumber utama kami pada penelitian kali ini. Namun, banyaknya jumlah data setiap tahunnya akan sulit untuk dilakukan pengolahan data secara manual. Terlebih lagi untuk dilakukan analisis dan prediksi terkait data distribusi air bersih tersebut. Untuk itu, diperlukan suatu teknik untuk mengolah data secara cepat

dan akurat, serta mudah dipahami. Melalui analisis visual, informasi data dari BPS dapat disajikan dengan cara yang lebih mudah dipahami, memberikan gambaran yang komprehensif terkait tren, perbandingan, dan tantangan yang dihadapi dalam distribusi air bersih selama periode tersebut. Pemerintah, lembaga non-profit, dan masyarakat umum juga dapat mengidentifikasi area-area yang memerlukan perhatian lebih dalam upaya meningkatkan akses terhadap air bersih di Indonesia.

## 2.2 Tujuan

Berikut merupakan tujuan utama dari analisis distribusi air bersih di Indonesia:

1. Mengetahui sebaran distribusi air bersih pada tahun 2015, 2017 - 2021 di 35 Provinsi Indonesia
2. Memperlihatkan tren dan perbandingan distribusi air bersih pada tahun 2015, 2017 - 2021 di 35 Provinsi Indonesia

Dari laporan ini, diharapkan dapat diperoleh pandangan yang lebih komprehensif mengenai kondisi distribusi air bersih di Indonesia, membantu dalam mengidentifikasi daerah-daerah prioritas untuk perbaikan infrastruktur air bersih. Selain itu, diharapkan laporan ini juga dapat memberikan inspirasi bagi pembuat kebijakan untuk merumuskan strategi baru atau meningkatkan kebijakan yang sudah ada guna mendukung distribusi air bersih yang lebih baik di Indonesia.

## 3. Prepare & Explore data

Preprocessing adalah tahapan untuk melakukan pembersihan data (data cleaning) dan menggabungkan data yang sudah ada (integrasi data) [2]. Pembersihan data juga akan mempengaruhi performansi dari sistem data mining karena data yang akan ditangani akan berkurang jumlah dan kompleksitasnya [3].

Sebelum data diolah, kami melakukan preprocessing terlebih dahulu dengan 2 tahapan, yaitu :

1. Integrasi Data

```
df1 = pd.read_csv('/content/Nilai Air Bersih yang Disalurkan 2015-2018.csv')
df1.head() # Menampilkan 5 data teratas dari dataset
```

	Provinsi	2015	2017	2018
0	ACEH	106402.0	104834.0	101737.0
1	SUMATERA UTARA	771685.0	883017.0	932260.0
2	SUMATERA BARAT	255501.0	256875.0	281808.0
3	RIAU	72133.0	53557.0	80125.0
4	JAMBI	67329.0	138531.0	180663.0

Gambar 1. Dataset Nilai Air Bersih yang Disalurkan 2015-2018

```
df2 = pd.read_csv('/content/Nilai Air Bersih yang Disalurkan 2019-2021.csv')
df2.head() # Menampilkan 5 data teratas dari dataset
```

	Provinsi	2019	2020	2021
0	ACEH	152839.0	158695.0	155402.0
1	SUMATERA UTARA	936629.0	960930.0	952824.0
2	SUMATERA BARAT	309177.0	344803.0	367149.0
3	RIAU	76836.0	86512.0	88400.0
4	JAMBI	165555.0	185147.0	203887.0

Gambar 2. Dataset Nilai Air Bersih yang Disalurkan 2019-2021

```
df = pd.merge(df1, df2, on = 'Provinsi') # Mengintegrasikan data pada dataframe 1 dan 2 menggunakan k
df
```

	Provinsi	2015	2017	2018	2019	2020	2021
0	ACEH	106402.0	104834.0	101737.0	152839.0	158695.0	155402.0
1	SUMATERA UTARA	771685.0	883017.0	932260.0	936629.0	960930.0	952824.0
2	SUMATERA BARAT	255501.0	256875.0	281808.0	309177.0	344803.0	367149.0
3	RIAU	72133.0	53557.0	80125.0	76836.0	86512.0	88400.0
4	JAMBI	67329.0	138531.0	180663.0	165555.0	185147.0	203887.0
5	SUMATERA SELATAN	507544.0	565299.0	604885.0	628045.0	661481.0	675702.0
6	BENGKULU	50587.0	72973.0	78637.0	80143.0	79828.0	78735.0

Gambar 3. Integrasi Data pada kedua Data Nilai Air Bersih yang Disalurkan

Data yang kami gunakan bersumber dari BPS, namun disajikan dalam beberapa file dengan kelompok tahun tertentu. Oleh karena itu, sebelum dilakukan analisis, kami melakukan integrasi data atau proses menggabungkan 2 data dengan kolom tertentu yang sama. Dalam hal ini, 2 file yang kami gunakan adalah data distribusi air bersih yang disalurkan tahun 2015-2018 (Gambar. 1) dan data distribusi air bersih yang disalurkan tahun 2019-2021 (Gambar. 2). Kemudian, data digabungkan dengan menggunakan kolom 'Provinsi' sebagai kunci. Hal ini, disebabkan kedua data memiliki kolom 'Provinsi' dengan nilai-nilai yang sama yaitu 35 Provinsi di Indonesia. Setelah dilakukan penggabungan, maka diperoleh satu data (Gambar. 3) yang kemudian dapat dilakukan pembersihan data terlebih dahulu agar memudahkan dalam analisis dan visualisasi data.

## 2. Data Cleaning

Setelah dilakukan penggabungan data, kemudian dilakukan pembersihan data.

### a. Mengidentifikasi Missing Value

```
df.isnull().sum()

Provinsi    0
2015        0
2017        0
2018        0
2019        0
2020        0
2021        0
dtype: int64
```

Gambar. 4 Missing Value

Pada Gambar. 4 dilakukan pengecekan terhadap missing value dan diperoleh nilai 0 pada semua atribut. Artinya, pada kolom Provinsi, 2015, 2017, 2018, 2019, 2020, dan 2021 tidak terdapat missing value satu pun.

b. Mengidentifikasi Data Duplikat

```
df.duplicated().sum()

0
```

Gambar. 5 Data Duplikat

Pada Gambar. 5 dilakukan pengecekan terhadap data duplikat. Diperoleh nilai 0, artinya pada data yang kami gunakan tidak terdapat data duplikat.

c. Mengidentifikasi Statistika Deskriptif Data

```
df.describe() # Cek deskripsi data
```

	2015	2017	2018	2019	2020	2021
<b>count</b>	3.500000e+01	3.500000e+01	3.500000e+01	3.500000e+01	3.500000e+01	3.500000e+01
<b>mean</b>	7.492275e+05	8.360221e+05	9.272886e+05	9.955821e+05	1.018462e+06	1.044357e+06
<b>std</b>	2.223990e+06	2.480028e+06	2.742540e+06	2.944159e+06	3.005986e+06	3.083055e+06
<b>min</b>	1.046100e+04	2.198800e+04	2.916900e+04	3.311700e+04	2.950300e+04	2.699100e+04
<b>25%</b>	6.140700e+04	7.215750e+04	7.716700e+04	7.848950e+04	8.345350e+04	8.465050e+04
<b>50%</b>	1.104460e+05	1.385310e+05	1.714000e+05	1.656480e+05	1.851470e+05	2.038870e+05
<b>75%</b>	5.017405e+05	5.416715e+05	6.079710e+05	6.628395e+05	6.708380e+05	6.757615e+05
<b>max</b>	1.311148e+07	1.463039e+07	1.622755e+07	1.742269e+07	1.782309e+07	1.827625e+07

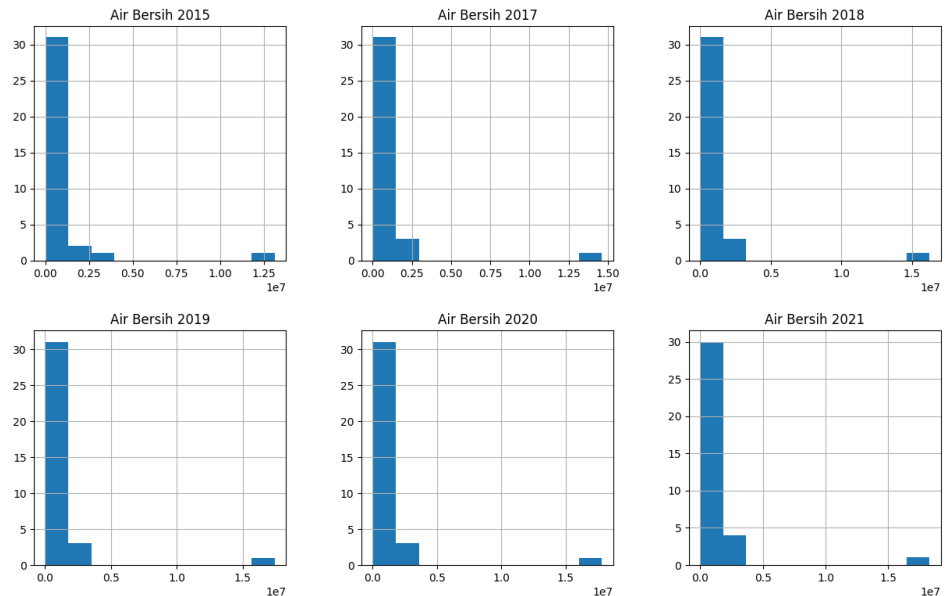
Gambar. 6 Statistika Deskriptif Data

Pada Gambar. 6 dilakukan identifikasi terhadap nilai statistika deskriptif menggunakan fungsi describe(). Fungsi ini akan menghasilkan beberapa statistik umum dalam dataframe, yaitu :

- Count: jumlah entri non-null dalam setiap kolom.
- Mean: rata-rata dari nilai dalam setiap kolom.
- Std: standar deviasi, mengukur seberapa tersebar data dari rata-rata.
- Min: nilai minimum dalam setiap kolom.
- 25th percentile (Q1): nilai di mana 25% dari data berada di bawahnya (kuartil pertama).

- 50th percentile (Q2): median atau nilai di tengah dataset (kuartil kedua).
- 75th percentile (Q3): nilai di mana 75% dari data berada di bawahnya (kuartil ketiga).
- Max: nilai maksimum dalam setiap kolom.

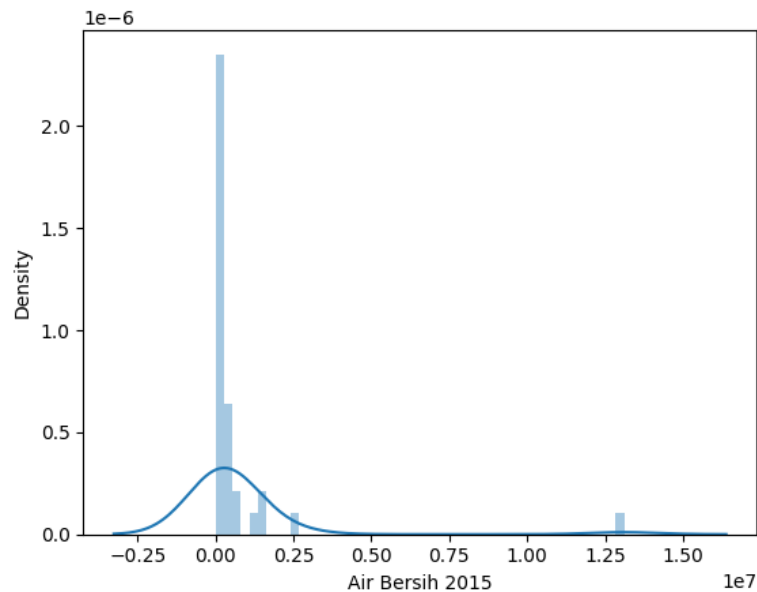
d. Membuat Visualisasi Histogram dari masing-masing atribut



Gambar. 7 Histogram dari masing-masing atribut

Pada Gambar. 7 dilakukan visualisasi dari setiap kolom numerik dari data menggunakan fungsi `hist()`. Plot diatur dengan lebar 15 dan tinggi 14. Diperoleh 7 histogram yang menampilkan distribusi dari masing-masing kolom.

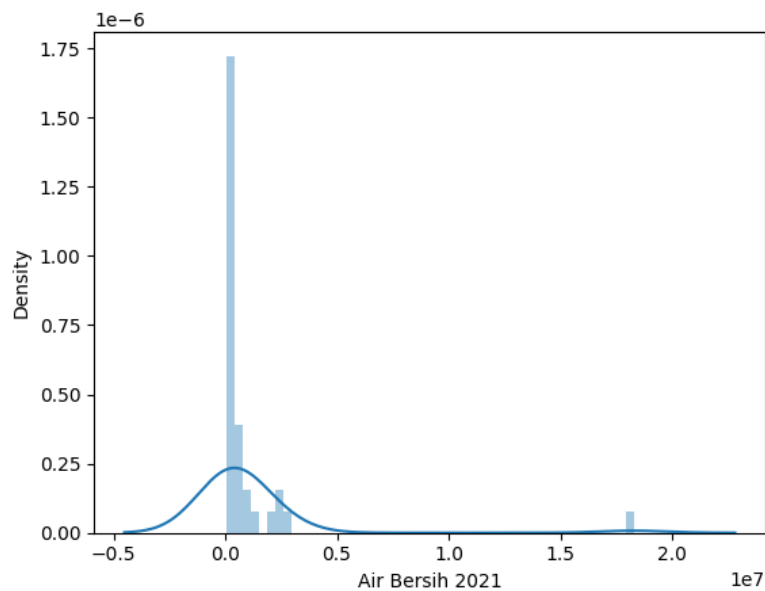
- e. Membuat Visualisasi Distplot dari kolom ‘Air Bersih 2015’



Gambar. 8 Distribusi Plot dari kolom ‘Air Bersih 2015’

Pada Gambar. 8 dilakukan visualisasi distribusi plot (distplot) dari kolom ‘Air Bersih 2015’ menggunakan fungsi `sns.distplot()` dari library Seaborn. Plot yang dihasilkan berupa distribusi dari kolom ‘Air Bersih 2015’ serta kepadatan yang memberikan gambaran tentang distribusi probabilitas dari data. Terlihat pada rentang 0 - 0,1 memiliki kepadatan tertinggi yakni diatas 2.

- f. Membuat Visualisasi Distplot dari kolom ‘Air Bersih 2021’

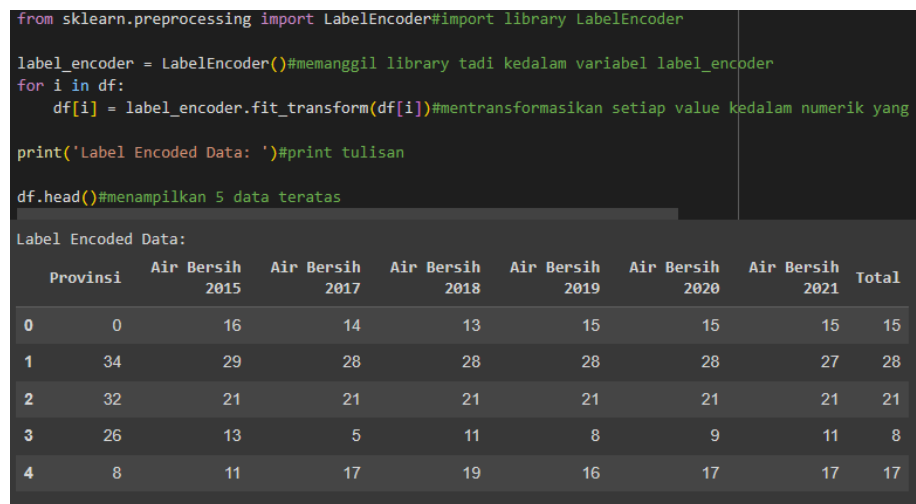


Gambar 9 Distribusi Plot dari kolom ‘Air Bersih 2021’



Pada Gambar. 9 dilakukan visualisasi distribusi plot (distplot) dari kolom 'Air Bersih 2021' menggunakan fungsi sns.distplot() dari library Seaborn. Plot yang dihasilkan berupa distribusi dari kolom 'Air Bersih 2021' serta kepadatan yang memberikan gambaran tentang distribusi probabilitas dari data. Terlihat pada rentang 0 - 0,1 memiliki kepadatan tertinggi yakni sebesar 1,75.

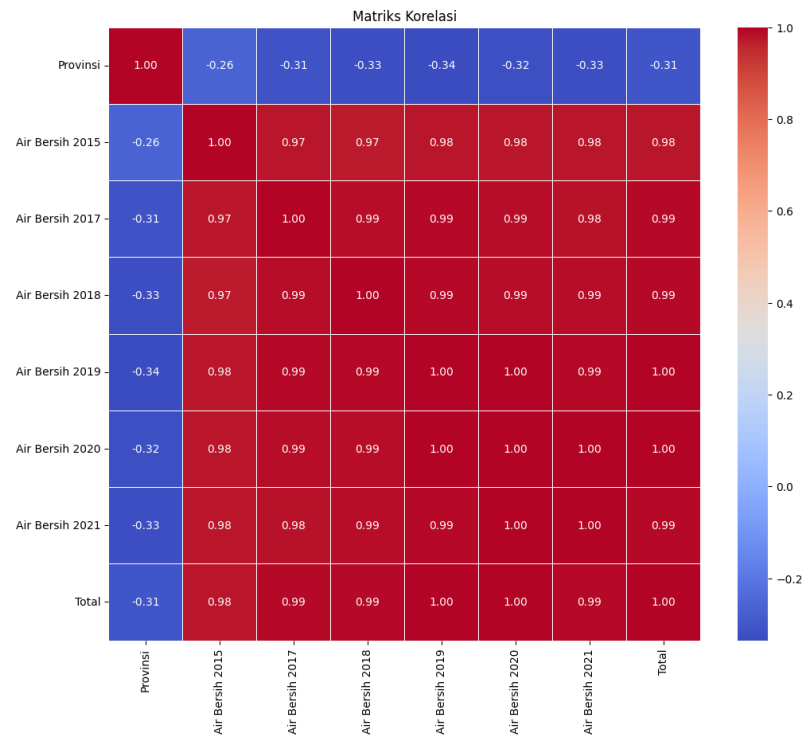
g. Membuat Label Encoded



Gambar 10. Label Encoded

Pada Gambar 10 dilakukan label encoded untuk mengubah nilai-nilai dalam setiap kolom dari tipe data kategori menjadi numerik. Kode menggunakan 'LabelEncoder' dari library scikit-learn. Dari kode tersebut, akan ditampilkan lima baris pertama dari dataframe setelah nilai-nilai kategorik pada setiap kolom diubah menjadi numerik. Nilai-nilai akan menjadi bilangan bulat yang mewakili nilai-nilai asli dari data kategori.

#### h. Membuat Matriks Korelasi



Gambar 11. Matriks Korelasi

Pada Gambar 11 dilakukan perhitungan berupa matriks korelasi dari setiap atribut pada data. Data dilakukan scaling terlebih dahulu agar memiliki rata-rata 0 dan varians 1. Kemudian, membuat matriks korelasi menggunakan fungsi `sns.heatmap()` dari Seaborn. Output yang dihasilkan adalah sebuah matriks korelasi antar kolom dari data. Warna dan angka di dalam setiap sel heatmap menunjukkan seberapa kuat korelasi antar kolom-kolom tersebut. Semakin dekat warnanya ke merah, semakin kuat korelasinya (positif). Sebaliknya, semakin dekat ke biru menunjukkan korelasi negatif yang kuat. Angka-angka dalam sel adalah nilai koefisien korelasi antar kolomnya.

#### 4. Formulate Question Design

Berikut merupakan pertanyaan terkait dengan dashboard distribusi akses air bersih di Indonesia yang mungkin diformulasikan:

- 1) Bagaimana sebaran distribusi air bersih di 35 Provinsi Indonesia pada tahun 2015?
- 2) Bagaimana sebaran distribusi air bersih di 35 Provinsi Indonesia pada tahun 2018?
- 3) Bagaimana sebaran distribusi air bersih di 35 Provinsi Indonesia pada tahun 2019?

- 4) Bagaimana sebaran distribusi air bersih di 35 Provinsi Indonesia pada tahun 2020?
- 5) Bagaimana sebaran distribusi air bersih di 35 Provinsi Indonesia pada tahun 2021?
- 6) Bagaimana tren distribusi air bersih di tiap provinsi di tahun 2015 hingga 2021?
- 7) Apakah terdapat perbedaan signifikan dalam distribusi air bersih antar provinsi?
- 8) Apa saja faktor-faktor utama yang mempengaruhi distribusi air bersih di Indonesia?
- 9) Bagaimana kualitas air bersih berkembang di berbagai wilayah selama periode penelitian?

## 5. Concepting

Sebelum memulai pembuatan visualisasi, perlu dilakukannya konsep dan pemilihan grafik yang sesuai dengan tujuan analisis. Berikut merupakan konsep visualisasi yang akan diterapkan dalam dasbor distribusi air bersih di Indonesia.

- 1) Penyajian judul besar “Distribusi Air Bersih di Indonesia”, bertujuan untuk menyajikan secara komprehensif dan informatif mengenai kondisi dan perkembangan distribusi air bersih di seluruh wilayah Indonesia.
- 2) Tampilan “Total” bertujuan untuk memberikan gambaran menyeluruh tentang total biaya per tahun dalam bantuan air bersih dari setiap provinsi. Tujuan utama dari tampilan ini adalah memberikan pemahaman komprehensif tentang kontribusi finansial yang diberikan oleh masing-masing provinsi dalam mendukung distribusi air bersih di seluruh Indonesia.
- 3) Tampilan “Provinsi” digunakan untuk melihat total dari provinsi yang terdapat di Indonesia dalam mendapatkan air bersih. Tujuannya adalah memberikan gambaran holistik tentang kontribusi masing-masing provinsi terhadap distribusi air bersih di Indonesia.
- 4) Peta sebaran akan digunakan untuk memberikan pandangan visual tentang distribusi air bersih di seluruh Indonesia. Provinsi-provinsi akan diwarnai untuk mencerminkan tingkat distribusi air bersih dan memungkinkan pengguna untuk dengan cepat mengidentifikasi pola dan disparitas antar wilayah.
- 5) Plot bar chart dan grafik batang akan digunakan untuk membandingkan jumlah bantuan air bersih tiap tahun pada setiap 35 provinsi di Indonesia. Hal ini memberikan gambaran komparatif yang jelas tentang perkembangan distribusi air bersih dari tahun ke tahun di berbagai wilayah.
- 6) Plot Bar Chart Tahun 2021, memberikan fokus pada data terkini. Pengguna dapat melihat perbandingan distribusi air bersih di setiap provinsi pada tahun tersebut dan menyoroti trend terbaru serta perubahan yang signifikan.
- 7) Plot Setengah Lingkaran (Gauge), memberikan indikator visual tentang pencapaian target atau standar tertentu dalam distribusi air bersih. Ini dapat

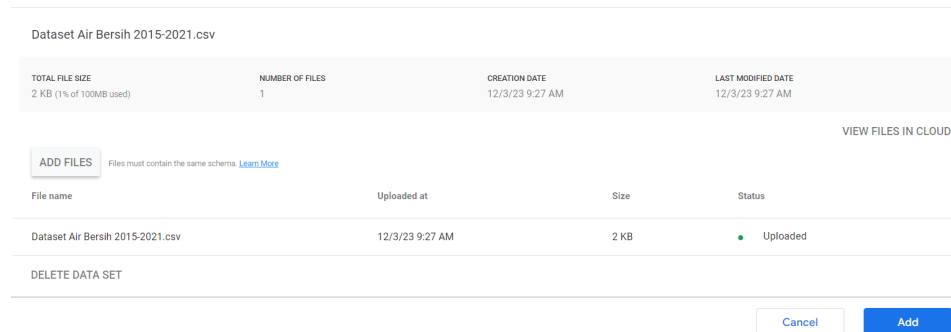
membantu mengevaluasi sejauh mana suatu provinsi telah mencapai tujuan tertentu dalam penyediaan air bersih.

Dengan menggunakan kombinasi visualisasi ini, dasbor akan memberikan gambaran yang komprehensif tentang distribusi air bersih di Indonesia. Setelah konsep disusun, langkah selanjutnya adalah mengimplementasikan dalam pembuatan visualisasi menggunakan software Google Data Studio.

## 6. Construct

Setelah melalui proses preprocessing data untuk memastikan keberhasilan dan kesiapan data untuk tahap visualisasi dalam pembuatan dashboard dan juga menentukan konsep dari konten yang akan ditampilkan di dalam dasbor, langkah berikutnya adalah menyimpan dan mengimpor data tersebut ke Google Data Studio. Dalam konteks pembuatan dasbor distribusi air di Indonesia, berikut adalah langkah-langkahnya.

### 1) Memasukkan dataset dengan memilih menu New -> Report -> File Upload



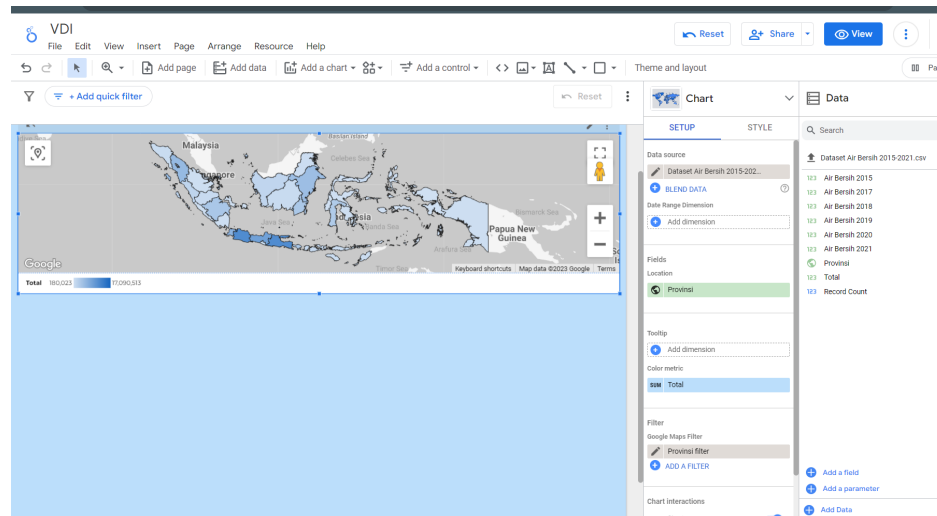
Gambar 12. Upload dataset

### 2) Membuat judul dengan tools text



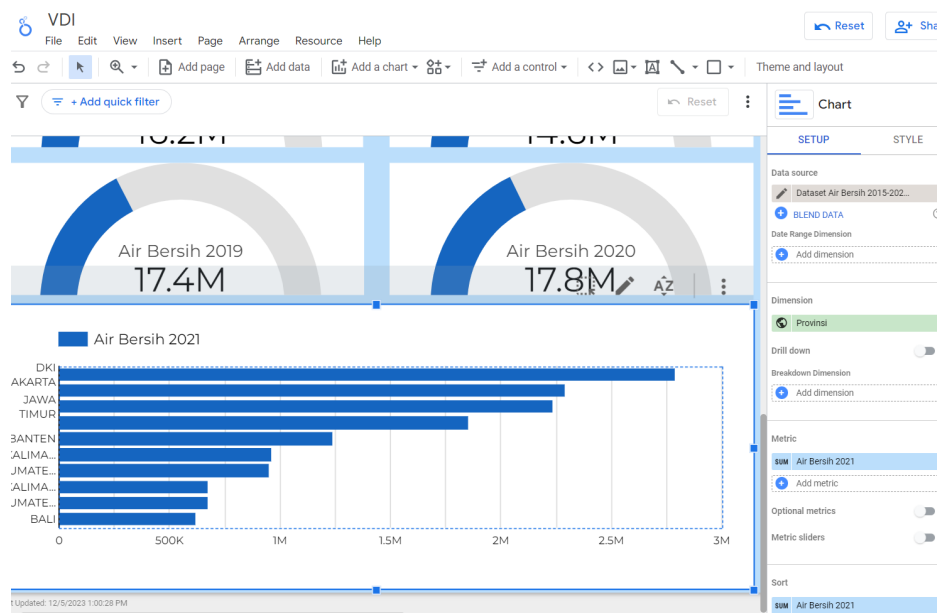
Gambar 13. Judul Dashboard

### 3) Membuat visualisasi peta sebaran air di Provinsi Indonesia menggunakan metrics kolom 'Total' dan 'Provinsi'



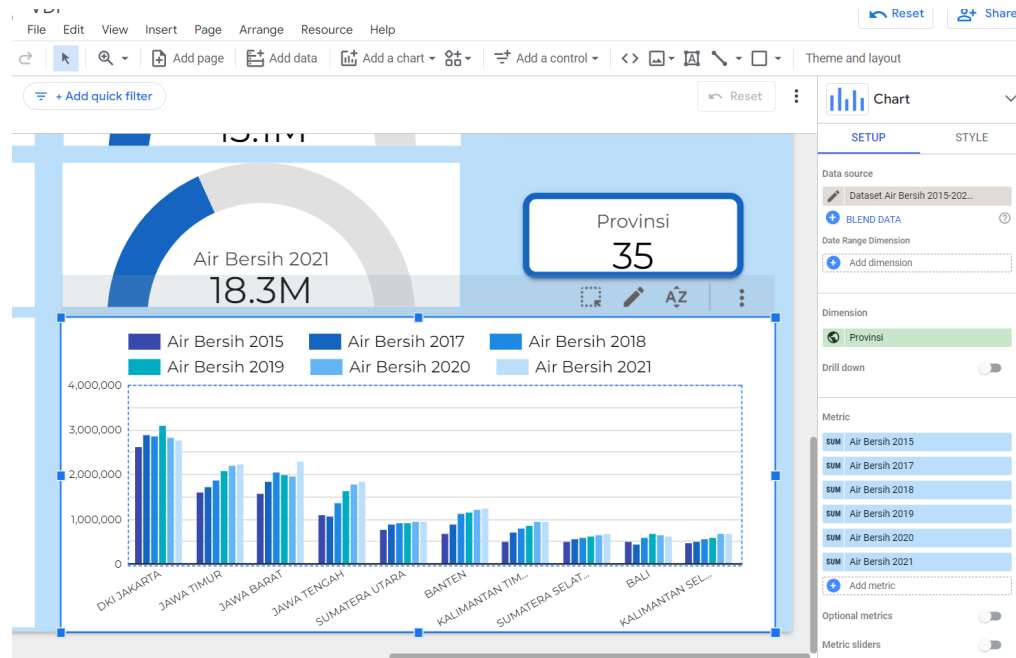
Gambar 14. Peta sebaran

4) Membuat Barchart secara vertikal untuk data air bersih yang tersebar di tahun 2021 menggunakan metric 'Air Bersih 2021' dan dimension 'Provinsi'



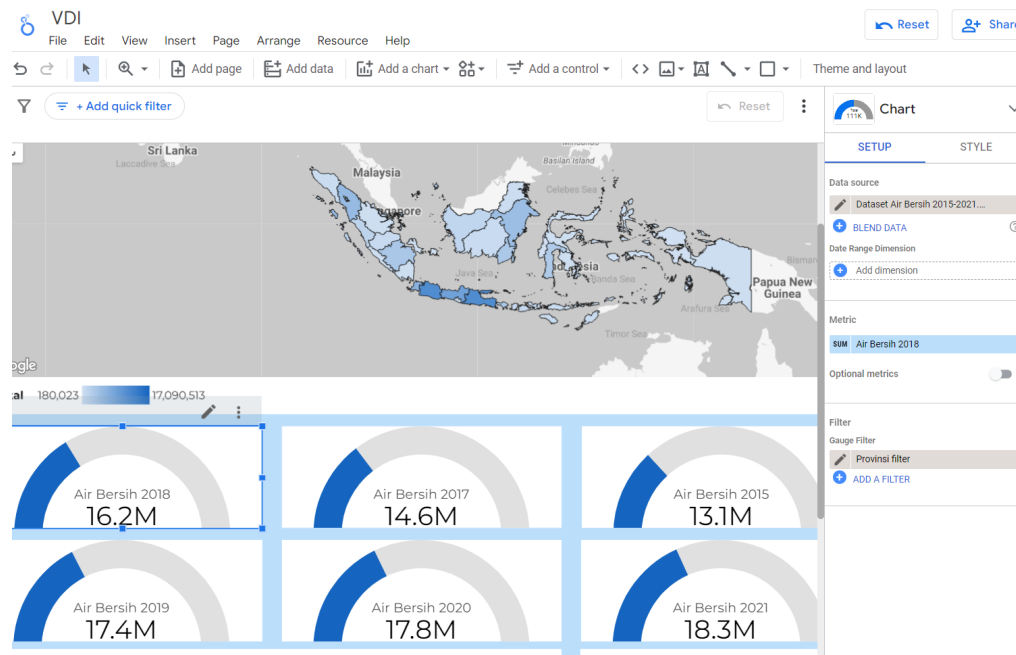
Gambar 15. Barchart sebaran air bersih 2015

5) Membuat Bar Chart untuk membandingkan sebaran air bersih di tahun 2015, 2017-2021 menggunakan dimension kolom 'Provinsi' dan metric kolom 'Air Bersih 2015', 'Air Bersih 2017', 'Air Bersih 2018', 'Air Bersih 2019', 'Air Bersih 2020', dan 'Air Bersih 2021'.



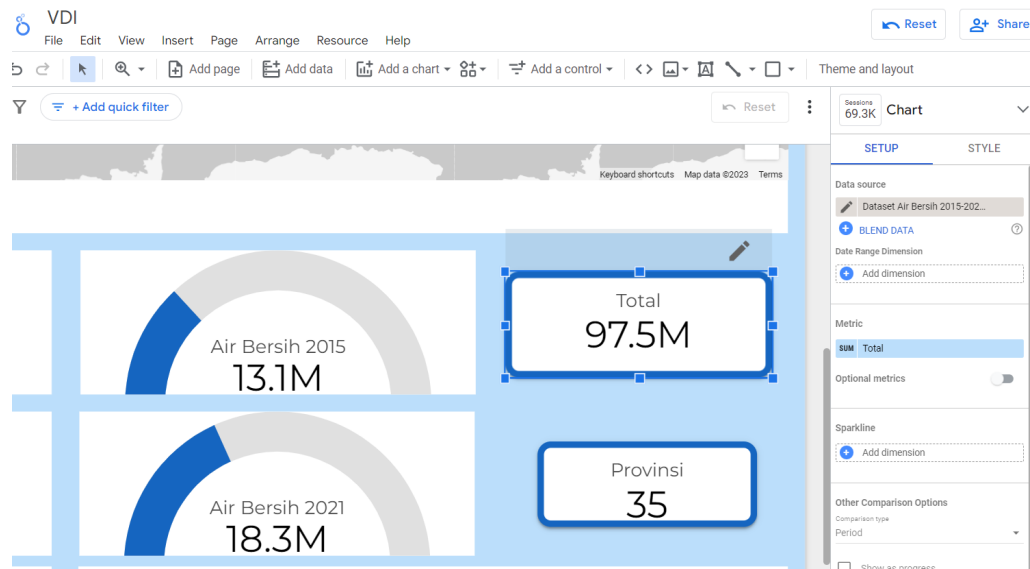
Gambar 16. Barchart Sebaran air bersih setiap tahun

6) Membuat Plot Setengah Lingkaran (Gauge) untuk total sebaran air di setiap tahun menggunakan metrics kolom 'Air Bersih 2015', Air Bersih 2017', Air Bersih 2018', Air Bersih 2019', Air Bersih 2020', dan Air Bersih 2021'.



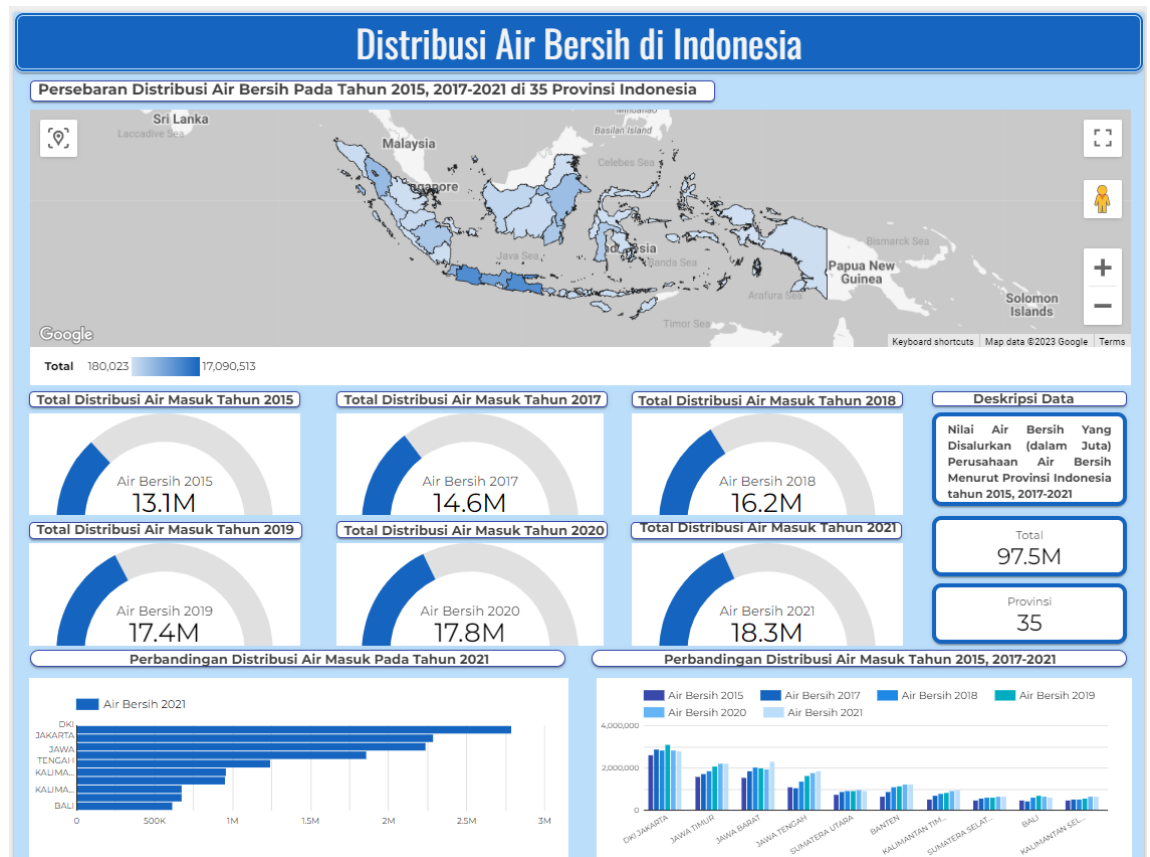
Gambar 17. Plot gauge sebaran air tiap tahun

## 7) Membuat tampilan total dan data provinsi



Gambar 18. Tampilan total air dan provinsi

## 8) Tampilan Visualisasi Dashboard



Link: <https://lookerstudio.google.com/s/u43ILfIvczA>

## 7. Feedback

Kelompok	Kelas	Masukan
5	RA	Memberikan deskripsi data singkat pada visualisasi
8	RA	Tidak ada
8	RA	-
10	RA	Untuk bagian distribusi air pertahun, kurang tepat penggunaan jenis visualisasinya karena kurang memperlihatkan perbedaan proporsi dari masing-masing tahun dan cukup sulit untuk membandingkan antara tahun tersebut.
87	RA	Mungkin bisa diberi label pada data nantinya, seperti di mapnya itu maksudnya biru tua itu apa? n biru muda apa? kasih label. terus total 16.2 juta itu apa? juta air bersih dalam mili kah ato apa? kasih tau juga. Gw rasa layoutnya juga bisa lebih di improve lagi, n visualisasi/design dashboardnya secara keseluruhan.
6	RA	-
7	RB	Diberi judul di setiap visualisasi
3	RC	-
14	RA	Petanya diperbesar supaya keliatan
5	RA	Tidak ada

## Referensi

[1] S. Zaid Romegar Mair, "VISUALISASI DEMOGRAFI WILAYAH," *Jurnal TIPS : Jurnal Teknologi Informasi dan Komputer Politeknik Sekayu*, vol. 9, pp. 1-9, 2019.

[2] A. D. A. & Andri, "Pemanfaatan Data Mining Dalam Memprediksi Produksi Pada PTPupuk Sriwidjaja Palembang Menggunakan Algoritma Regresi Linier Berganda," *Jurnal Nasional Ilmu Komputer*, vol. 2, pp. 51-63, 2021.



[3] I. Nurdiyanto, O. Nurdiawan, N. R, A. I. Purnamasari dan D. A. Kurnia, “Penentuan Keputusan Pemberian Pinjaman Kredit Menggunakan Algoritma C.45,” *JURNAL DATA SCIENCE & INFORMATIKA (JDSI)*, vol. 2, pp. 1-5, 2022.