

# **LAPORAN TUGAS BESAR II**

## **IF4061 VISUALISASI DATA**

### **Visualisasi Data Interaktif**

Diajukan sebagai tugas Mata Kuliah IF4061 Visualisasi Data



Disusun Oleh:

Jason Rivalino	13521008
Muhammad Equilibrie Fajria	13521047
M. Farrel Danendra Rachim	13521048

**SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA**  
**INSTITUT TEKNOLOGI BANDUNG**  
**BANDUNG**  
**2025**

## DAFTAR ISI

<b>DAFTAR ISI</b>	<b>2</b>
<b>BAB I</b>	<b>3</b>
<b>RUMUSAN PERANCANGAN DASHBOARD</b>	<b>3</b>
A. Pertanyaan Kunci	3
B. Jenis Dashboard	3
C. Target Pengguna	4
<b>BAB II</b>	<b>5</b>
<b>KAKAS DAN JENIS GRAFIK DASHBOARD</b>	<b>5</b>
A. Framework dan Library yang digunakan	5
B. Jenis Grafik pada Dashboard	5
<b>BAB III</b>	<b>8</b>
<b>EKSPLORASI DAN PERSIAPAN DATA</b>	<b>8</b>
A. Data Acquisition	8
B. Data Examination	8
C. Penjelasan Tipe Data	9
D. Data Transforming	11
E. Consolidating	14
<b>BAB IV</b>	<b>15</b>
<b>KONSEP DESAIN</b>	<b>15</b>
A. Struktur Informasi	15
B. Desain Dashboard	16
<b>BAB V</b>	<b>21</b>
<b>PENGUJIAN DAN EVALUASI DASHBOARD</b>	<b>21</b>
A. Pengujian Rancangan Awal Dashboard	21
B. Evaluasi Hasil Pengujian	21
C. Perbaikan dan Penyempurnaan Dashboard	22
D. Evaluasi Akhir Dashboard	26
<b>LAMPIRAN</b>	<b>27</b>
A. Link Dashboard Interaktif	27
B. Referensi	27
C. Pembagian Tugas	27
D. Github Link	27

## **BAB I**

### **RUMUSAN PERANCANGAN *DASHBOARD***

#### **A. Pertanyaan Kunci**

Pertanyaan kunci yang ingin dijawab melalui pembuatan *Dashboard* pada tugas besar ini adalah untuk mengetahui analisis terkait dengan tren keterlambatan penerbangan yang terjadi di Amerika Serikat selama periode 2013 hingga 2023 dengan peninjauan berdasarkan berbagai komponen antara lain faktor-faktor utama yang menyebabkan keterlambatan, distribusi pelayanan penerbangan di masing-masing negara bagian, serta maskapai yang beroperasi. Faktor utama keterlambatan menjadi rumusan masalah utama dalam hal ini, sedangkan distribusi pelayanan penerbangan dan maskapai yang beroperasi menjadi rumusan masalah pendukung.

Proses pengembangan *Dashboard* ini memiliki tujuan utama yaitu untuk membantu pengguna dalam memahami kondisi keterlambatan penerbangan yang terjadi secara menyeluruh, sehingga dapat digunakan sebagai dasar dalam mengevaluasi dan merumuskan solusi terhadap permasalahan keterlambatan yang terjadi selama periode tersebut.

#### **B. Jenis Dashboard**

Jenis *Dashboard* yang akan dibangun dalam pengerjaan tugas besar ini adalah *Dashboard* berjenis analitis. Pemilihan jenis *Dashboard* ini didasarkan pada tujuannya untuk memberikan fleksibilitas kepada pengguna dalam melakukan eksplorasi terhadap berbagai variabel yang terdapat dalam data. Dengan demikian, pengguna dapat menganalisis informasi secara lebih mendalam dan menyesuaikan proses analisis sesuai dengan kebutuhan spesifik yang ingin dicapai.

Melalui *Dashboard* yang ada, pengguna dapat menemukan sejumlah hasil analisis yang menarik sehingga *Dashboard* yang ada tidak hanya menjadi alat penyaji data, tetapi juga alat untuk pengambilan keputusan yang membantu pengguna dalam memahami penyebab utama

dan tren yang terjadi selama periode 2013–2023 terkait keterlambatan penerbangan di Amerika Serikat.

### **C. Target Pengguna**

Terdapat tiga target pengguna yang ingin dituju dari proses visualisasi data interaktif yang akan dilakukan yaitu sebagai berikut.

a. Penumpang pesawat secara umum

Dengan adanya visualisasi data terkait topik yang dibahas, informasi yang ada diharapkan dapat membantu penumpang pesawat secara umum untuk melakukan pemilihan terhadap maskapai atau bandara dengan tingkat keterlambatan yang lebih rendah.

b. Agen perencanaan perjalanan

Dengan adanya visualisasi data terkait topik yang dibahas, informasi yang ada diharapkan dapat memberikan wawasan kepada agen perencanaan perjalanan untuk memberikan rekomendasi saran terkait waktu penerbangan secara lebih akurat.

c. Ahli atau penggemar penerbangan

Dengan adanya visualisasi data terkait topik yang dibahas, informasi yang ada diharapkan dapat membantu para ahli ataupun penggemar penerbangan untuk menentukan solusi dari permasalahan penundaan serta membantu pengembangan dalam bidang industri penerbangan di Amerika Serikat yang lebih baik untuk kedepannya.

## BAB II

### KAKAS DAN JENIS GRAFIK DASHBOARD

#### A. Framework dan Library yang digunakan

Pembangunan *Dashboard* interaktif akan dilakukan dengan menggunakan *Framework Streamlit* yang berbasis bahasa pemrograman Python. *Framework* ini menyediakan antarmuka *website* yang sesuai dengan kebutuhan visualisasi data dan analisis eksplorasi yang diperlukan. Selain penggunaan *Streamlit*, terdapat beberapa *library* pendukung yang digunakan untuk membantu visualisasi data antara lain sebagai berikut.

a. *Numpy*

*Library* ini digunakan untuk mengelola operasi perhitungan numerik yang dibutuhkan untuk proses visualisasi data.

b. *Pandas*

*Library* ini digunakan untuk melakukan proses pembacaan *file dataset* CSV untuk kemudian dilakukan berbagai proses operasi analisis untuk visualisasi data.

c. *Plotly*

*Library* ini digunakan untuk menampilkan keseluruhan visualisasi data dalam bentuk grafik yang bersifat interaktif.

d. *Datetime*

*Library* ini digunakan untuk membantu pemrosesan data berdasarkan waktu.

#### B. Jenis Grafik pada *Dashboard*

Terdapat beberapa jenis grafik yang akan digunakan pada proses visualisasi data dengan menggunakan *Dashboard* yang bertujuan untuk membantu pengguna dalam memahami pola, tren, dan hubungan yang ada pada data secara lebih mudah. Pemilihan jenis grafik pada visualisasi akan disesuaikan dengan karakteristik data serta tujuan analisis yang ingin dicapai. Beberapa jenis grafik yang digunakan antara lain sebagai berikut.

a. *Horizontal Stacked Bar Chart*

*Horizontal Stacked Bar Chart* digunakan untuk memvisualisasikan proporsi data berdasarkan beberapa kategori tertentu. Pada *Dashboard* yang dikembangkan, *Horizontal Bar Chart* digunakan untuk menunjukkan proporsi terkait dengan faktor-faktor yang

menyebabkan terjadinya keterlambatan pada penerbangan untuk tahun terkini. Data ini direpresentasikan untuk menyampaikan informasi kepada pembaca terkait dengan penyebab keterlambatan utama untuk tahun terbaru sebagai salah satu informasi yang menjadi prioritas bagi pengetahuan pengguna. Visualisasi ini bertujuan untuk memberikan gambaran ringkas mengenai kontribusi setiap faktor terhadap keseluruhan keterlambatan yang terjadi pada tahun paling baru.

b. *Line Chart*

*Line Chart* digunakan untuk memvisualisasikan data yang berkaitan dengan periode atau jangka waktu tertentu. Pada *Dashboard* yang dikembangkan, *Line Chart* digunakan untuk menampilkan tren jumlah total keterlambatan penerbangan yang terjadi setiap tahun selama periode 2013 hingga 2023. Visualisasi ini bertujuan untuk membantu pengguna dalam mengidentifikasi pola trend dari keterlambatan penerbangan di Amerika Serikat yang terjadi dari tahun ke tahunnya.

c. *Donut Chart*

*Donut Chart* digunakan untuk memvisualisasikan proporsi data berdasarkan beberapa kategori tertentu. Pada *Dashboard* yang dikembangkan, *Donut Chart* digunakan untuk menunjukkan proporsi terkait dengan faktor-faktor yang menyebabkan terjadinya keterlambatan pada penerbangan untuk semua tahun. Visualisasi ini bertujuan untuk memberikan gambaran ringkas mengenai kontribusi setiap faktor terhadap keseluruhan keterlambatan yang terjadi dalam jangka tahun tertentu.

d. *Stacked Column Chart*

*Stacked Column Chart* digunakan untuk memvisualisasikan perbandingan antara kategori tertentu dalam bentuk diagram batang, yang digabungkan dengan perbandingan proporsi kategori lain yang dimiliki setiap kategori tersebut. Pada *Dashboard* yang dikembangkan, *Stacked Column Chart* digunakan untuk membandingkan total keterlambatan penerbangan yang direpresentasikan oleh setiap batang, dan faktor keterlambatan setiap tahun direpresentasikan oleh proporsi warna dari setiap batang. Visualisasi ini membantu pengguna dalam mengenali proporsi apa saja yang menjadi penyebab paling dominan untuk setiap tahun.

e. *Bar Column Chart*

*Bar Column Chart* digunakan untuk memvisualisasikan perbandingan antara kategori tertentu dalam bentuk diagram batang. Pada *Dashboard* yang dikembangkan, *Column Chart* digunakan untuk membandingkan rata-rata keterlambatan yang terjadi pada berbagai maskapai penerbangan. Visualisasi ini membantu pengguna dalam mengidentifikasi maskapai dengan tingkat keterlambatan tertinggi maupun terendah secara lebih mudah.

f. *Choropleth Map*

*Choropleth Map* digunakan untuk memvisualisasikan data spasial berdasarkan wilayah geografis. Pada *Dashboard* yang dikembangkan, *Choropleth Map* digunakan untuk mengetahui persebaran terkait dengan keterlambatan penerbangan yang terjadi pada tiap negara bagian yang ada di Amerika Serikat. Visualisasi ini memungkinkan pengguna untuk melihat secara geografis wilayah mana saja yang memiliki tingkat keterlambatan tinggi atau rendah, sehingga memudahkan dalam mengidentifikasi pola keterlambatan berdasarkan lokasi.

## BAB III

### EKSPLORASI DAN PERSIAPAN DATA

#### A. *Data Acquisition*

Proses *Data Acquisition* dilakukan dengan melakukan pengambilan data awal yang berasal dari *dataset* yang bersifat *Open Source* sehingga telah tersedia secara langsung dan dapat diakses melalui platform Kaggle dengan link berikut.

Kaggle link: <https://www.kaggle.com/datasets/sriharshaedala/airline-delay>

Adapun *dataset* yang terdapat pada *Kaggle Link* ini juga mengambil referensi sumber data yang berasal dari pemerintah Amerika (*U.S. Government Works*), tepatnya pada instansi pemerintah bagian Bureau of Transportation Statistics, Airline Service Quality Performance 234 dengan referensi link pemerintah adalah berikut.

US Government link: <https://www.usa.gov/government-copyright>

#### B. *Data Examination*

Proses *Data Examination* dilakukan untuk menentukan seberapa tingkat kepercayaan pada *dataset* yang akan dianalisis dan divisualisasikan. Proses *Data Examination* akan meninjau dua hal utama yaitu aspek *Completeness* dan *Quality* dengan rincian penjelasan sebagai berikut.

##### a. *Completeness*

Pada aspek *Completeness*, beberapa hasil yang didapat dari pemeriksaan yang dilakukan terkait dengan kelengkapan kebutuhan data adalah sebagai berikut.

- i. Data yang ada sudah memiliki kelengkapan yang sesuai dengan setiap kategori atribut data yang diperlukan telah mencukupi untuk kebutuhan proses visualisasi data.
- ii. Data yang ada sudah mencakup periode waktu secara jelas dan lengkap dengan rentang waktu yang terdapat dalam data berada diantara Agustus 2013 hingga Agustus 2023.
- iii. Data yang ada sudah mencakup keseluruhan aspek variabel dengan total keseluruhan kolom pada data yaitu sebanyak 21 kolom.



- iv. Data yang ada sudah menyediakan total jumlah *records* yang sudah cukup untuk dianalisis secara menyeluruh dengan total sebanyak 171.666 baris data.

b. *Quality*

Pada aspek *Quality*, beberapa hasil yang didapat dari pemeriksaan yang dilakukan terkait dengan kelengkapan kebutuhan data adalah sebagai berikut.

- i. Data yang ada sudah cukup bagus dan tidak mengandung eror.
- ii. Data yang ada sudah memberikan informasi yang jelas terkait dengan klasifikasi atribut yang dimiliki masing-masing.
- iii. Data yang ada secara umum sudah memiliki format penulisan yang terstruktur untuk semua *records* dan tidak ada permasalahan terkait dengan karakter data.
- iv. Masih terdapat *records* data yang mengandung *missing value* dengan beberapa atribut tertentu memiliki nilai yang bersifat NULL. Untuk rincian atribut dengan *missing value* yaitu terdapat 6 atribut dengan seluruh *records* terisi lengkap, 14 atribut dengan total 240 *records* yang mengalami *missing value*, dan 1 atribut dengan total 443 *records* yang mengalami *missing value*.
- v. Tidak ada *records* yang mengalami duplikat pada keseluruhan data yang ada.
- vi. Akurasi data sudah terukur dengan baik, meskipun ada beberapa data yang tergolong dalam *extreme outlier*.
- vii. Terdapat sekitar 19.000-22.000 nilai data yang termasuk *outlier* untuk setiap kolom jika dilakukan pengecekan dengan menggunakan metode IQR dan sekitar 1.000-3.000 nilai data *outlier* jika dilakukan pengecekan dengan menggunakan metode Z-Score. Walaupun begitu, hanya sekitar 100-200 data yang dapat dikatakan tergolong dalam *extreme outlier*, yaitu *outlier* yang memiliki nilai yang sangat tidak realistis dibandingkan *outlier* lainnya.

### C. Penjelasan Tipe Data

Berbagai atribut yang ada dalam data yang akan divisualisasikan memiliki jenis tipe dengan kelompok yang berbeda-beda dengan rincian pembagiannya adalah sebagai berikut.

a. *Categorical Nominal*

Beberapa atribut yang termasuk ke dalam tipe data ini adalah berikut.

- *Carrier*: kode maskapai penerbangan
- *Carrier\_name*: nama maskapai penerbangan

- *Airport*: kode bandara
- *Airport\_name*: nama bandara

b. *Categorical Ordinal*

Tidak ada atribut yang termasuk ke dalam tipe data ini.

c. *Quantitative (interval-scale)*

Beberapa atribut yang termasuk ke dalam tipe data ini adalah berikut.

- *Year*: data tahun penerbangan
- *Month*: data bulan penerbangan

d. *Quantitative (ratio-scale)*

Beberapa atribut yang termasuk ke dalam tipe data ini adalah berikut.

- *Arr\_flights*: banyaknya penerbangan kedatangan
- *Arr\_del15*: banyaknya penerbangan kedatangan yang mengalami keterlambatan (*delayed*) selama 15 menit atau lebih
- *Carrier\_ct*: jumlah keterlambatan selama 15 menit atau lebih akibat faktor maskapai penerbangan
- *Weather\_ct*: jumlah keterlambatan selama 15 menit atau lebih akibat faktor cuaca
- *Nas\_ct*: jumlah keterlambatan selama 15 menit atau lebih akibat faktor *National Airspace System* (NAS)
- *Security\_ct*: jumlah keterlambatan selama 15 menit atau lebih akibat faktor keamanan
- *Late\_aircraft\_ct*: jumlah keterlambatan selama 15 menit atau lebih akibat kedatangan pesawat yang terlambat
- *Arr\_cancelled*: jumlah penerbangan yang dibatalkan
- *Arr\_diverted*: jumlah penerbangan yang dialihkan
- *Arr\_delay*: banyaknya total menit dari penundaan penerbangan
- *Carrier\_delay*: total menit penundaan penerbangan akibat faktor maskapai penerbangan
- *Weather\_delay*: total menit penundaan penerbangan akibat faktor cuaca
- *Nas\_delay*: total menit penundaan penerbangan akibat faktor *National Airspace System* (NAS)
- *Security\_delay*: total menit penundaan penerbangan akibat faktor keamanan

- *Late\_aircraft\_delay*: total menit penundaan penerbangan akibat kedatangan pesawat yang terlambat

#### **D. Data Transforming**

##### **a. Transforming for Quality**

Proses *Transforming for Quality* dilakukan dengan tujuan untuk melakukan penyesuaian dan pembersihan terhadap data yang mengalami kesalahan. Pada data yang digunakan ini, terdapat sejumlah proses penyesuaian dan pembersihan yang dilakukan antara lain sebagai berikut.

- i. Melakukan penghapusan pada data yang mengalami duplikat. Dikarenakan tidak ada duplikat pada data yang digunakan, maka proses ini tidak perlu dilakukan.
- ii. Memproses beberapa data yang mengalami *missing value*. Pada data yang ada, dapat dilakukan penghapusan terhadap baris data yang mengalami *missing value* dikarenakan proporsi data yang memiliki *missing value* hanya sedikit yaitu sebanyak 443 *records* dibanding total keseluruhan data yaitu 171.666 *records* (~0.26%).
- iii. Data-data yang memiliki nilai *outlier* dibiarkan karena asumsi data yang diperoleh bernilai realistis dan proporsi data yang termasuk "*outlier*" sangat banyak sehingga tidak mungkin disingkirkan atau diganti nilainya.
- iv. Melakukan *handling* berupa perubahan terhadap kesalahan format penulisan karakter. Dikarenakan tidak ada kesalahan format penulisan karakter, maka proses ini tidak perlu dilakukan.

##### **b. Transforming for Analysis**

Proses *Transforming for Analysis* dilakukan dengan tujuan untuk mempersiapkan dan menyempurnakan data yang didapat untuk analisis dan visualisasi. Pada data yang digunakan ini, terdapat sejumlah proses penyempurnaan yang dilakukan antara lain sebagai berikut.

- i. Melakukan *parsing* variabel yang terdapat pada atribut kolom "*airport\_name*" dengan mengekstraksi nama kota dan nama *state* menjadi dua kolom baru yang terpisah dari nama bandara karena data pada kolom tersebut memiliki tiga variabel yang berbeda dengan format yang *parseable* yaitu {kota tempat lokasi bandara}, {*state* tempat lokasi bandara}: {nama bandara}. Setelah dilakukan

proses *parsing*, kolom "*airport\_name*" yang awalnya hanya satu saja akan dibagi menjadi tiga kolom yaitu "*airport\_city*", "*airport\_state*", dan "*airport\_name*". Berikut merupakan contoh implementasi terhadap datanya.

Contoh data: "Albany, GA: Southwest Georgia Regional".

Hasil parsing:

<i>airport_city</i>	<i>airport_state</i>	<i>airport_name</i>
Albany	GA	Southwest Georgia Regional

- ii. Melakukan penggabungan terhadap beberapa variabel untuk membentuk variabel yang baru. Berdasarkan data yang ada, proses ini tidak diperlukan dikarenakan variabel-variabel yang tersedia sudah cukup sesuai dalam memberikan informasi yang dibutuhkan.
- iii. Mengubah data yang bersifat kualitatif menjadi kode atau kata kunci. Berdasarkan data yang ada, beberapa atribut sudah memiliki isi data dalam bentuk kode atau kata kunci seperti "*carrier\_code*" ataupun "*airport\_code*" sehingga proses ini tidak diperlukan.
- iv. Menentukan nilai baru lainnya. Berdasarkan data yang ada, proses ini tidak diperlukan dikarenakan variabel-variabel yang tersedia sudah cukup sesuai dalam memberikan informasi yang dibutuhkan.
- v. Melakukan perubahan terhadap beberapa kolom data menjadi proporsi persentase. Kalkulasi dilakukan karena jika dilakukan agregasi terhadap nilai keseluruhan dari kolom-kolom tersebut, maka hasilnya akan sama dengan sebuah kolom data yang menampung nilai-nilai data tersebut. Dengan kata lain, ada beberapa atribut data yang menjadi sub kategori dari atribut data tertentu (kategori induk). Nilai kolom sebaiknya diwakilkan menjadi persentase (desimal) untuk kemudahan analisis dan visualisasi proporsi data yang nantinya akan membandingkan kuantitas setiap sub kategori tersebut (misal perbandingan penyebab penundaan penerbangan). Proses ini menjadi penting karena untuk nilai data yang termasuk ke dalam "kategori induk" memiliki nilai yang berbeda-beda untuk setiap baris data. Contoh kondisi perbaikan yang dilakukan adalah berikut.

- Atribut "*arr\_flights*" yang menandakan banyak kedatangan penerbangan suatu pesawat, memiliki subkategori "*arr\_del15*" (banyak kedatangan penerbangan yang tertunda selama 15 menit atau lebih), "*arr\_diverted*" (banyak penerbangan yang dialihkan) dan "*arr\_cancelled*" (banyak penerbangan yang dibatalkan). Oleh karena itu, perlu dibuat kolom baru yang berisi persentase data penerbangan dari total penerbangan "*arr\_flights*", yakni "*arr\_del15\_percentage*", "*arr\_diverted\_percentage*", dan "*arr\_cancelled\_percentage*".
- Atribut "*arr\_del15*" memiliki subkategori banyaknya penerbangan yang ditunda berdasarkan penyebab penundaan penerbangan, yakni "*carrier\_ct*", "*weather\_ct*", "*nas\_ct*", "*security\_ct*", dan "*late\_aircraft\_ct*". Oleh karena itu, perlu dibuat kolom baru yang berisi persentase dari total penerbangan yang ditunda "*arr\_del15*", yakni "*carrier\_ct\_percentage*", "*weather\_ct\_percentage*", "*nas\_ct\_percentage*", "*security\_ct\_percentage*", dan "*late\_aircraft\_ct\_percentage*".
- Atribut "*arr\_delay*" yang menandakan total menit penundaan dari sebuah pesawat memiliki subkategori banyaknya menit penerbangan yang ditunda berdasarkan penyebab penundaan penerbangan, yakni "*carrier\_delay*", "*weather\_delay*", "*nas\_delay*", "*security\_delay*", dan "*late\_aircraft\_delay*". Oleh karena itu, perlu dibuat kolom baru yang berisi persentase dari total menit penundaan "*arr\_delay*", yakni "*carrier\_delay\_percentage*", "*weather\_delay\_percentage*", "*nas\_delay\_percentage*", "*security\_delay\_percentage*", dan "*late\_aircraft\_delay\_percentage*".

Contoh implementasi transformasi:

arr_delay	carrier_delay	weather_delay	nas_delay	security_delay	late_aircraft_delay
1604.00	207.00	0.00	1049.00	0.00	348.00

Agregasi nilai subkategori sama dengan nilai kategori induk "*arr\_delay*":

$$207 + 0 + 1049 + 0 + 348 = 1604.$$

Berdasarkan hasil ini, maka nilai kolom-kolom baru yang telah dikalkulasi proporsinya adalah:

- carrier\_delay\_percentage:  $\frac{207}{1604} * 100\% = 12.90\%$
- weather\_delay\_percentage:  $\frac{0}{1604} * 100\% = 0.00\%$
- nas\_delay\_percentage:  $\frac{1049}{1604} * 100\% = 65.40\%$
- security\_delay\_percentage:  $\frac{0}{1604} * 100\% = 0.00\%$
- late\_aircraft\_delay\_percentage:  $\frac{348}{1604} * 100\% = 21.70\%$

Dari hasil yang didapatkan, untuk kondisi dari keseluruhan tabel baru yang terbentuk adalah sebagai berikut.

arr_delay	carrier_delay	carrier_delay_percentage	weather_delay	weather_delay_percentage
1604.00	207.00	12.90	0.00	0.00

nas_delay	nas_delay_percentage	security_delay	security_delay_percentage	late_aircraft_delay	late_aircraft_delay_percentage
1049.00	65.40	0.00	0.00	348.00	21.70

- vi. Melakukan penghapusan pada data *redundant* yang tidak diperlukan. Berdasarkan data yang ada, proses ini tidak diperlukan dikarenakan variabel-variabel yang tersedia sudah cukup sesuai dalam memberikan informasi yang dibutuhkan.
- vii. Resolusi yang diterapkan pada data ini yaitu *Full Resolution* yang menampilkan data individual secara lengkap.

### E. Consolidating

Berdasarkan data yang digunakan, data hanya diambil dari satu sumber saja dan tidak mengambil referensi dari sumber lainnya sehingga proses *Consolidating* tidak diperlukan dalam visualisasi data yang akan dilakukan.

## BAB IV

### KONSEP DESAIN

#### A. Struktur Informasi

Berikut adalah struktur *Dashboard* per halaman yang dibuat dalam urutan logika intuitif, yaitu secara *top-down*, *left-right*.

##### a. Halaman utama

Level	Komponen <i>Dashboard</i>	Contoh Isi
Overview	Tahun terakhir, total keterlambatan penerbangan pada tahun terakhir, total keterlambatan penerbangan sejak tahun 2013, dan proporsi faktor keterlambatan tahun terakhir.	Menunjukkan faktor apa yang paling mempengaruhi keterlambatan tahun terakhir.
Segmentasi	Tren persentase keterlambatan penerbangan per tahun dan proporsi faktor keterlambatan 2013-2023.	Mendapatkan informasi terkait dengan peningkatan signifikan persentase keterlambatan penerbangan pada tahun 2021/2022.
Graf Pendukung	Perbandingan jumlah keterlambatan penerbangan per tahun beserta penyebabnya.	Mendapatkan informasi bahwa pesawat yang terlambat datang menjadi salah satu faktor utama keterlambatan penerbangan pada setiap tahun.

##### b. Halaman analisis *carrier delay*

Level	Komponen <i>Dashboard</i>	Contoh Isi
Overview	Rata-rata persentase keterlambatan, maskapai dengan persentase keterlambatan tertinggi dan terendah.	JetBlue Airways memiliki persentase keterlambatan tertinggi.
Segmentasi	Grafik perbandingan keterlambatan penerbangan berdasarkan maskapai.	Allegiant Air memiliki persentase keterlambatan lebih rendah dibandingkan JetBlue Airways.

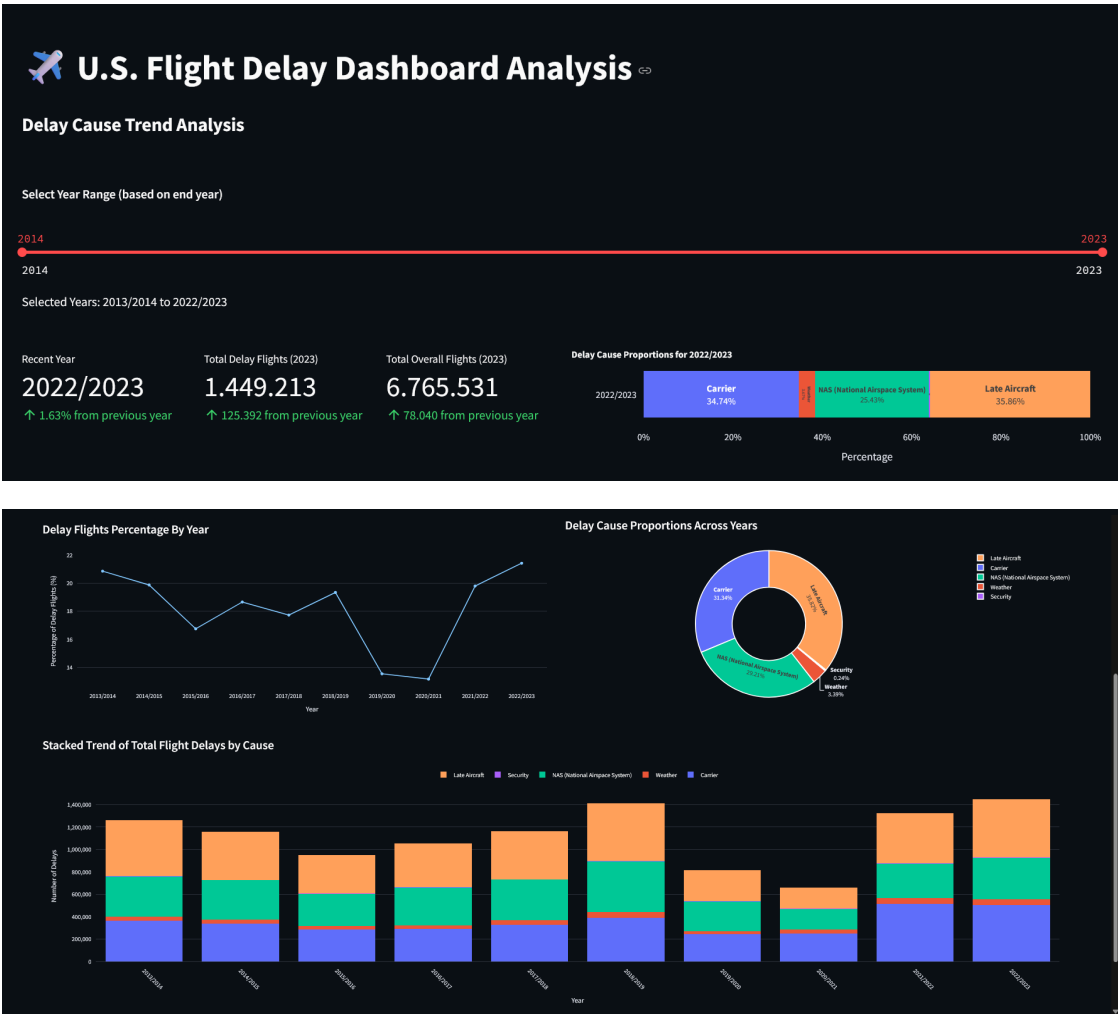
c. Halaman analisis *state delay*

Level	Komponen <i>Dashboard</i>	Contoh Isi
Overview	Rata-rata persentase keterlambatan setiap <i>state</i> , <i>state</i> dengan persentase keterlambatan tertinggi dan terendah.	Delaware memiliki persentase keterlambatan tertinggi.
Segmentasi	Grafik perbandingan keterlambatan penerbangan berdasarkan <i>state</i> .	Banyak negara bagian di Amerika yang memiliki persentase keterlambatan sekitar 20%.

B. Desain Dashboard

Berikut adalah desain awal dashboard kami, yang dipisahkan berdasarkan halaman dashboard.

a. Halaman utama





Halaman utama berisi segala informasi mengenai penyebab dan jumlah keterlambatan yang terjadi setiap tahunnya. Halaman ini merupakan rumusan masalah utama yang ingin disampaikan kepada pembaca pada pembuatan *dashboard* ini. Dalam segi *layout*, halaman ini hanya memiliki maksimal empat grafik beserta KPI yang berguna sebagai *overview*, agar pengguna hanya memperoleh informasi yang paling penting. Dengan membatasi jumlah grafik dalam satu halaman, gambaran umum tidak akan hilang dalam segala detail yang redundan.

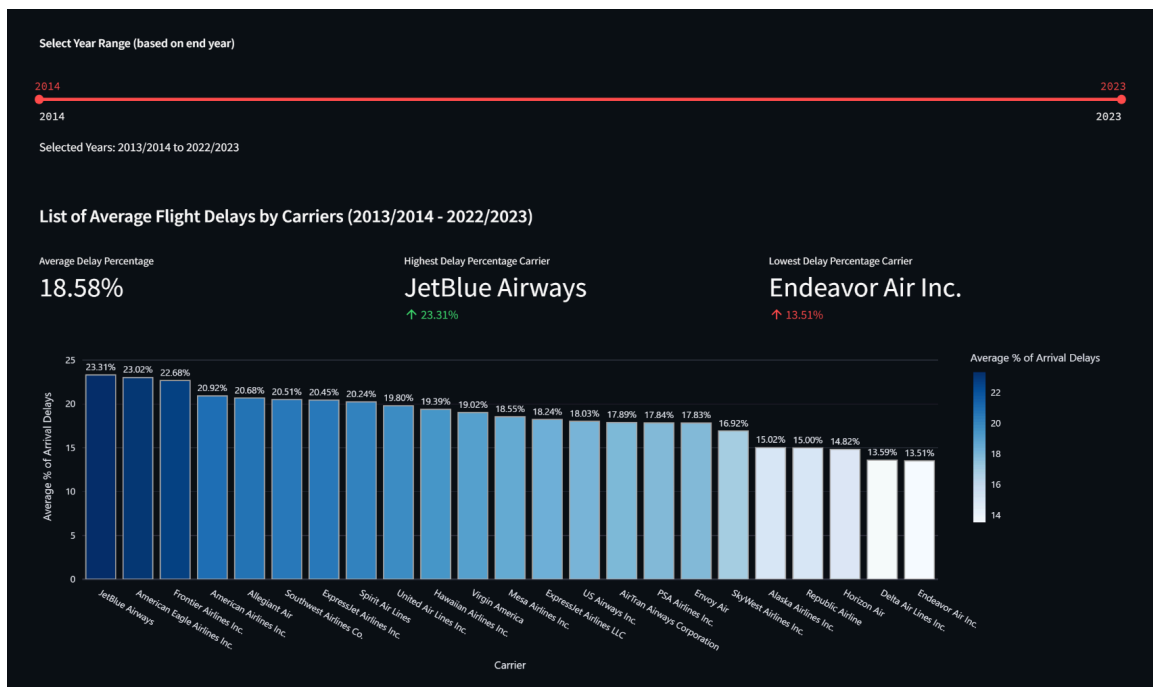
Urutan *dashboard* yang intuitif juga menjadi faktor penting agar pembaca memahami informasi, dari tingkat umum sampai tingkat yang lebih kompleks. Implementasi dari pengurutan pada *Dashboard* yang digunakan yaitu dengan urutan dari atas ke bawah dan dari kiri ke kanan. Tujuan penerapan urutan ini agar pembaca dapat memahami tahun terbaru dari informasi data, serta berbagai penyebab dan jumlah keterlambatan penerbangan dengan *overview* data diletakkan pada bagian atas *dashboard*. Setelah itu, data yang lebih kompleks namun masih penting seperti tren banyak keterlambatan per tahun (*line chart*) dan proporsi keterlambatan untuk seluruh tahun (*donut chart*) diletakkan pada bagian tengah *dashboard*. *Line chart* menampilkan data jumlah penerbangan yang relatif lebih penting dibandingkan *donut chart*, sehingga *line chart* diletakkan pada bagian kiri. Pada bagian paling bawah, diletakkan *chart* pendukung yang menampilkan secara lebih jelas proporsi penyebab keterlambatan penerbangan untuk setiap tahun. Data ini memiliki waktu lebih lama untuk dipahami dibandingkan *chart* lain, sehingga *chart* ini ditaruh pada bagian paling bawah.

Untuk halaman ini, *filter* untuk rentang tahun dibuat dalam bentuk *slider* yang sangat cocok untuk disesuaikan dan digerakkan jika pengguna hanya ingin menggerakkan *slider* hanya untuk setahun atau dua tahun. Filter diletakkan di bagian paling atas karena filter ini akan di-*apply* untuk semua *view* dalam halaman *dashboard* ini. Jika filter tahun digerakkan, data yang ditampilkan juga akan berubah sesuai rentang tahun. Selain itu, judul teks seperti “Selected Years: start\_year to end\_year” juga dibuat dinamis sesuai seleksi *filter* agar pembaca mengetahui bahwa data sudah difilter serta rentang tahun yang dipilih.

Dengan berbagai warna yang digunakan untuk merepresentasikan faktor keterlambatan, digunakan legenda untuk memberi petunjuk apa faktor yang direpresentasikan warna tersebut (misal hijau untuk NAS dan biru untuk *carrier*). Legenda untuk *stacked column chart* dan *donut chart* diletakkan sedekat mungkin dengan masing-masing *chart* karena letak dua *chart* tersebut yang agak berjauhan dan agar dapat diamati pengguna dengan mudah.

Adapun terdapat juga elemen pendukung graf lain seperti *tooltip* dan sumbu graf. *Tooltip* merupakan kotak teks yang menunjukkan informasi tambahan pada suatu objek data saat di-*hover*. Hal ini membantu menghilangkan kebingungan akibat *information overload* dan memastikan hanya menampilkan informasi yang penting saja pada pandangan pertama. Misalkan, titik data tahun 2016/2017 pada *line chart* memiliki *tooltip* yang berguna untuk menampilkan jumlah eksak keterlambatan penerbangan pada tahun 2016/2017. Selain itu, sumbu-X dan Y graf sudah memiliki label dan format yang sesuai, serta unit yang dicantumkan pada judul sumbu.

b. Halaman analisis *carrier delay*

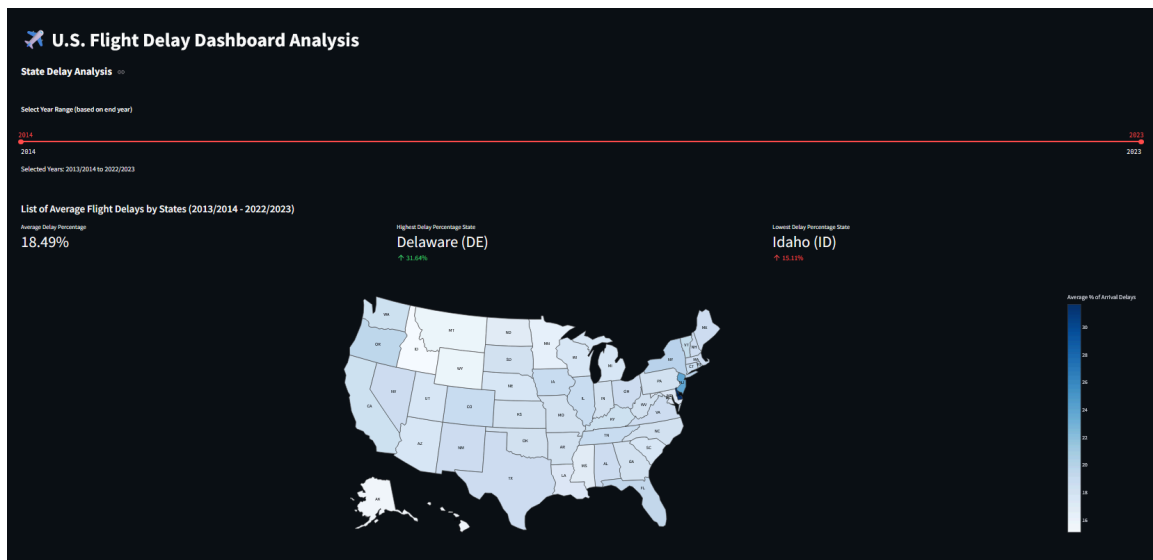


Halaman analisis *carrier delay* berisi informasi mengenai perbandingan persentase keterlambatan berdasarkan maskapai. Informasi ini ditampilkan dalam halaman terpisah

karena memiliki fokus tujuan yang berbeda terhadap fokus utama halaman utama yaitu untuk mengetahui analisis berdasarkan maskapai penerbangan. Halaman ini hanya berisi *overview* informasi dasar berupa persentase keterlambatan rata-rata, maskapai dengan persentase keterlambatan tertinggi, maskapai dengan persentase keterlambatan terendah, dan *column chart* yang menunjukkan banyak persentase keterlambatan seluruh maskapai.

Halaman ini juga memiliki filter rentang tahun berupa *slider* yang diletakkan pada bagian paling atas untuk mempengaruhi perubahan informasi *overview* dan graf. Teks subjudul seperti “List of Average Flight Delays by Carriers (start\_year - end\_year)” juga dibuat dinamis agar pembaca mengetahui bahwa data sudah dilakukan *filtering* serta rentang tahun yang dipilih. Selain *tooltip*, graf juga memiliki gradasi warna terurut untuk *chart* yang dipilih dalam warna biru untuk menunjukkan lebih lanjut bahwa data diurutkan dari keterlambatan penerbangan tertinggi sampai terendah.

c. Halaman analisis *state delay*



Halaman analisis *state delay* berisi informasi mengenai perbandingan persentase keterlambatan berdasarkan maskapai. Informasi ini ditampilkan dalam halaman terpisah karena memiliki fokus tujuan yang berbeda terhadap fokus utama halaman utama yaitu untuk mengetahui analisis berdasarkan negara bagian. Halaman ini hanya berisi *overview* informasi dasar berupa persentase keterlambatan rata-rata, *state* dengan persentase keterlambatan tertinggi, *state* dengan persentase keterlambatan terendah, dan *choropleth*

*map* yang menunjukkan gradasi warna untuk menunjukkan banyaknya persentase keterlambatan per *state*.

Halaman ini juga memiliki filter rentang tahun berupa *slider* yang diletakkan pada bagian paling atas untuk mempengaruhi perubahan informasi *overview* dan peta *choropleth*. Teks subjudul seperti “List of Average Flight Delays by States (start\_year - end\_year)” juga dibuat dinamis agar pembaca mengetahui bahwa data sudah difilter serta rentang tahun yang dipilih. Peta *choropleth* juga dilengkapi dengan *tooltip* yang menunjukkan nama lengkap *state* dan persentase keterlambatan saat pengguna fokus terhadap suatu *state*.

## BAB V

### PENGUJIAN DAN EVALUASI *DASHBOARD*

#### A. Pengujian Rancangan Awal *Dashboard*

Pengujian terhadap rancangan awal *Dashboard* dilakukan dengan menerima *feedback* dari responden yang menjadi target dalam *dashboard* ini. Kami mengumpulkan sebanyak 15 orang penggemar penerbangan dari ITB untuk mencoba *dashboard* versi awal kami. Responden kemudian akan memberi *feedback* terhadap beberapa pertanyaan tersebut.

- Apa informasi yang ditampilkan sudah memiliki *insight* yang menarik dan menjawab pertanyaan rumusan masalah?
- Bagaimana waktu responsivitas *dashboard*?
- Bagaimana *layout dashboard* kami? Apa informasi dapat dicerna dengan mudah dan terorganisir?
- Apa penggunaan tipe dan warna *chart* sudah tepat?
- Apa grafik dapat dibaca urutannya secara logis dan intuitif?
- Apa Anda dapat mengambil kesimpulan analisis dari dashboard ini dengan baik?

#### B. Evaluasi Hasil Pengujian

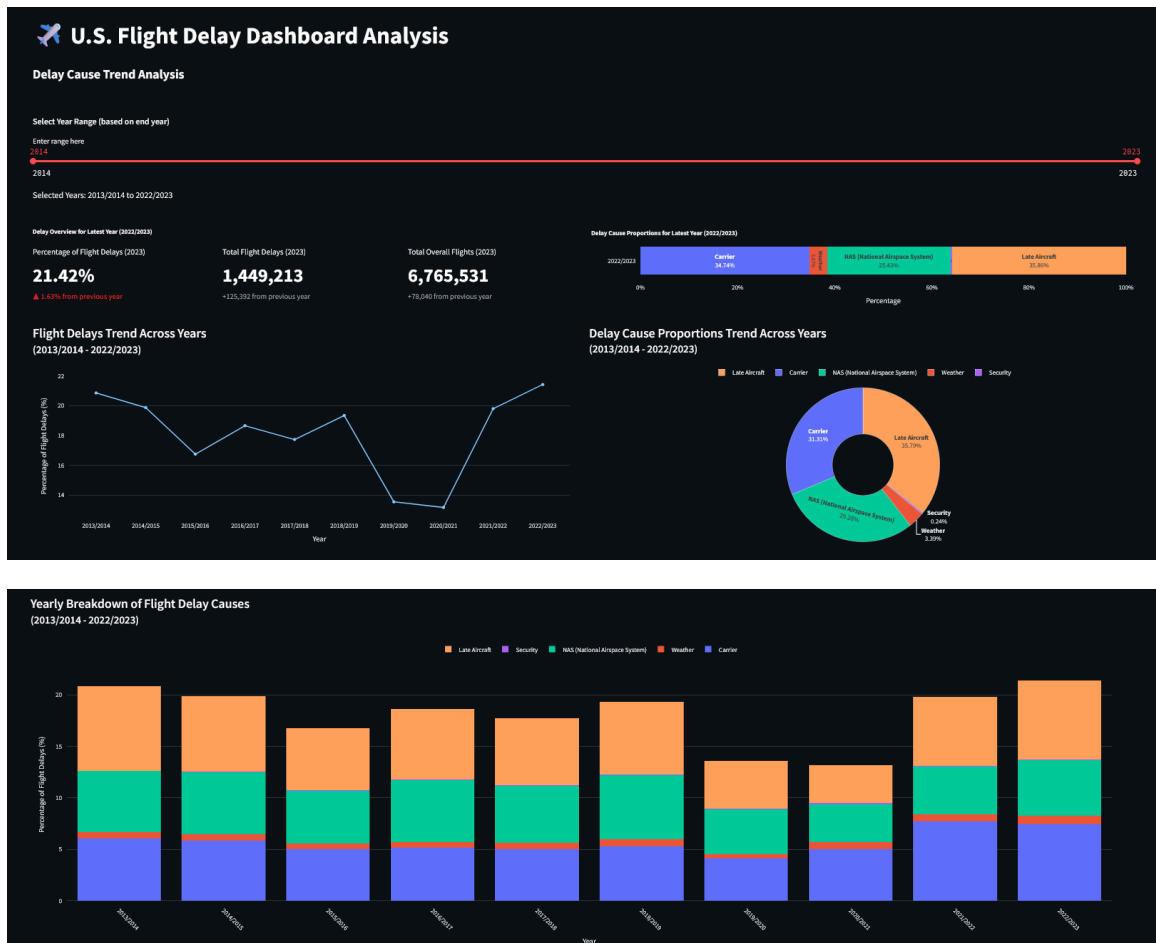
Berdasarkan pengujian, berikut adalah beberapa hal penting yang dapat diperbaiki pada *dashboard*:

- Jika keterlambatan menurun, seharusnya merupakan hal yang bagus. Warna di *dashboard* merah saat keterlambatan menurun seakan-akan itu adalah hal yang buruk.
- Sebaiknya judul dinamis ditambahkan ke setiap *chart* agar interpretasi *chart* lebih mudah dilakukan sesuai konteks judulnya.
- Untuk informasi pada baris pertama (tren penyebab keterlambatan) sedikit kurang intuitif karena yang di-input pengguna berupa rentang tahun, namun yang dimunculkan pada *dashboard* adalah tahun paling *recent*.
- Dapat ditambahkan informasi tambahan lagi untuk analisis halaman *state* dan *carrier* seperti perbandingan keterlambatan per maskapai untuk setiap tahunnya.

### C. Perbaikan dan Penyempurnaan *Dashboard*

Berikut adalah beberapa perbaikan yang kami lakukan terhadap halaman *dashboard* berdasarkan masukan dari responden.

#### a. Halaman utama



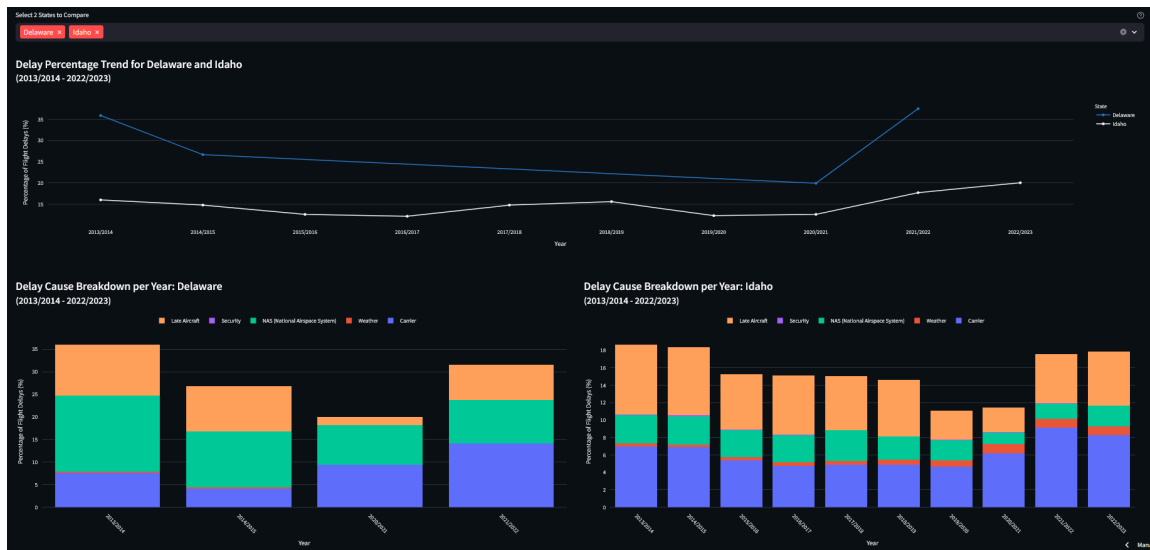
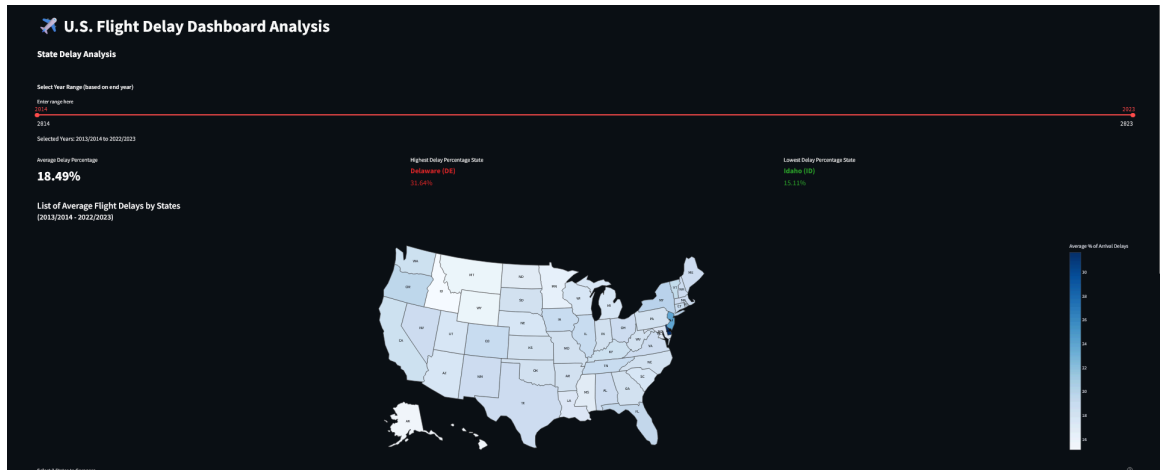
- Informasi *overview* di-**bold** karena lebih penting dibandingkan teks-teks lainnya.
- Baris pertama *dashboard* (*overview* dan *bar chart*) dikasih judul kecil yang bersifat dinamis agar pembaca memahami bahwa data yang ditampilkan di baris atas mewakili data tahun paling terakhir sesuai filter *slider*. Misalkan, pada contoh gambar di atas, tahun terakhir di *slider* adalah 2022/2023, sehingga data yang ditampilkan di *overview* dan *bar chart* adalah data dari 2023.
- Pada setiap *chart* lain juga ditambahkan rentang tahun yang ditampilkan oleh grafik tersebut agar pengguna langsung dapat mengetahui rentang tahun apa saja yang ditampilkan berdasarkan *slider* tanpa perlu meng-*scroll* ke atas untuk mengecek.



- Informasi *overview* di-*bold* karena lebih penting dibandingkan teks-teks lainnya. Seperti halaman utama, warna penanda *overview* disesuaikan dengan konotasi sifat keterlambatan maskapai; warna merah menandakan maskapai yang memiliki persentase keterlambatan terbesar (negatif), sedangkan warna hijau menandakan maskapai yang memiliki persentase keterlambatan terkecil (positif).
- Legenda gradasi warna *column chart* ditiadakan karena redundant.
- Ditambahkan dua tipe grafik baru yang berfungsi membandingkan maksimal dua maskapai berdasarkan persentase keterlambatannya. Seperti halaman utama, grafik yang ditampilkan adalah *line chart* untuk membandingkan persentase keterlambatan maskapai-maskapai per tahun (dengan tiap garis menggambarkan maskapai yang dibandingkan), serta dua grafik *stacked column chart* yang menggambarkan *breakdown* penyebab keterlambatan masing-masing maskapai tersebut. Pengguna dapat memilih maskapai menggunakan menu *multiselect*. *Slider* rentang tahun juga berlaku untuk graf-graf ini. Hal ini agar pengguna memiliki kebebasan lebih lanjut untuk mengeksplorasi lebih dalam mengenai bagaimana sebuah maskapai memiliki persentase keterlambatan yang lebih besar/kecil dibandingkan maskapai lain.



c. Halaman analisis *state delay*



- Informasi *overview* di-**bold** karena lebih penting dibandingkan teks-teks lainnya. Seperti halaman utama, warna penanda *overview* disesuaikan dengan konotasi sifat keterlambatan state; warna merah menandakan state yang memiliki persentase keterlambatan terbesar (negatif), sedangkan warna hijau menandakan state yang memiliki persentase keterlambatan terkecil (positif).
- Ditambahkan dua tipe grafik baru yang berfungsi membandingkan maksimal dua state berdasarkan persentase keterlambatannya. Seperti halaman utama, grafik yang ditampilkan adalah *line chart* untuk membandingkan persentase keterlambatan *state-state* per tahun (dengan tiap garis menggambarkan state yang dibandingkan), serta dua grafik *stacked column chart* yang menggambarkan *breakdown* penyebab keterlambatan masing-masing *state* tersebut. Pengguna dapat memilih *state*

menggunakan menu *multiselect*. *Slider* rentang tahun juga berlaku untuk graf-graf ini. Hal ini agar pengguna memiliki kebebasan lebih lanjut untuk mengeksplorasi lebih dalam mengenai bagaimana sebuah *state* memiliki persentase keterlambatan yang lebih besar/kecil dibandingkan *state* lain.

#### **D. Evaluasi Akhir *Dashboard***

Berdasarkan perbaikan yang sudah dilakukan, berikut adalah evaluasi akhir *dashboard* dari segi fungsi, kemudahan penggunaan, dan estetika.

##### **a. Fungsi**

- Visualisasi sudah menjawab semua pertanyaan pada rumusan masalah, yang telah dibedakan berdasarkan halaman *dashboard*.
- Tujuan visualisasi sudah dijelaskan secara singkat di judul dan teks sekitarnya.
- Kami sudah memilih *chart* yang paling tepat untuk dianalisis.
- Data yang paling penting seperti persentase keterlambatan sudah ditampilkan di sumbu-X dan sumbu-Y, sedangkan data yang kurang penting seperti penyebab keterlambatan ditampilkan sebagai warna.
- *Tooltip* sudah bersifat informatif dan mudah dibaca karena formatnya yang rapi.

##### **b. Kemudahan penggunaan**

- Pengguna dapat menggunakan filter seperti *slider* dengan mudah dan intuitif.
- Pengguna dapat memahami judul, legenda, dan informasi yang ditampilkan oleh *chart*.
- Untuk setiap *view* tidak ada *scrollbar*.

##### **c. Estetika**

- Graf sudah diletakkan dengan rapi dengan urutan logis, dengan graf paling penting di bagian atas kiri.
- Warna di *dashboard* kelihatan rapi tanpa adanya warna yang terlalu bertabrakan.
- Font yang digunakan sudah konsisten, hanya digunakan maksimal 2 *font*.

## LAMPIRAN

### A. Link *Dashboard* Interaktif

Berikut merupakan lampiran dari link *dashboard* interaktif dari pengerjaan tugas besar ini.

Link: <https://us-flight-delay-1323-dashboard-analysis.streamlit.app/>

### B. Referensi

Sumber Dataset:

<https://www.kaggle.com/datasets/sriharshaedala/airline-delay>

### C. Pembagian Tugas

Berikut pembagian tugas setiap anggota kelompok:

NIM	Nama	Kontribusi
13521008	Jason Rivalino	- Mengerjakan tab State Delay Analysis - Mengerjakan laporan
13521047	Muhammad Equilibrie Fajria	- Mengerjakan tab Carrier Delay Analysis - Mengerjakan laporan
13521048	M. Farrel Danendra Rachim	- Mengerjakan tab Delay Cause Trend Analysis - Mengerjakan laporan

### D. Github Link

Berikut merupakan lampiran dari link Github yang digunakan untuk proses penyusunan program untuk perancangan *Dashboard* interaktif dari pengerjaan tugas besar ini.

Link: [https://github.com/jasonrivalino/IF4061\\_InteractiveData\\_USAirlineDelay](https://github.com/jasonrivalino/IF4061_InteractiveData_USAirlineDelay)