

Analisis Data Eksploratorif Pada Studi Kasus US Flight Routes and Fares

<https://doi.org/10.28932/jste.v1i1.13174>

Received: 28 Agustus 2025 | Revised: 27 September 2025 | Accepted: 29 September 2025

Creative Commons License 4.0 (CC BY – NC)



¹Nathan Joshua^{✉#1}, Setia Budi^{*2}

[#]Program Studi Teknik Informatika, Fakultas Teknologi dan Rekayasa Cerdas, Universitas Kristen Maranatha
Jl. Prof. drg. Surya Sumantri No.65, Bandung, Jawa Barat 40164, Indonesia

¹2172026@maranatha.ac.id

^{*}Program Studi Magister Ilmu Komputer, Fakultas Teknologi dan Rekayasa Cerdas, Universitas Kristen Maranatha
Jl. Prof. drg. Surya Sumantri No.65, Bandung, Jawa Barat 40164, Indonesia

²setia.budi@it.maranatha.edu

[✉]Corresponding author: 2172026@maranatha.ac.id

How to cite this article:

N. Joshua, S. Budi, "Analisis Data Eksploratorif Pada Studi Kasus US Flight Routes and Fares," *Journal of Smart Technology and Engineering*, vol. 1, no. 1, pp. 38-57, 2025, <https://doi.org/10.28932/jste.v1i1.13174>

Abstrak — Industri penerbangan merupakan salah satu sektor ekonomi global yang memiliki pengaruh sangat besar, untuk Amerika Serikat sendiri penerbangan merupakan salah satu sektor ekonomi terbesar, mulai dari penerbangan nasional hingga internasional dengan distribusi harga tiket penerbangan yang tidak menentu. Analisis dilakukan untuk mencari faktor penentu hingga faktor utama yang menentukan harga tiket penerbangan domestik di Amerika Serikat. Analisis dilakukan dengan menggunakan metode EDA (Exploratory Data Analysis), dan dikombinasikan dengan rekayasa fitur, data eksternal berupa berita, dan artikel-artikel yang relevan. Terdapat beberapa temuan yang didapatkan setelah melakukan analisis terhadap dataset US Flight Routes and Fares (1993-2023), yang merupakan dataset yang menampung catatan penerbangan domestik Amerika Serikat. Temuan yang didapatkan adalah harga tiket penerbangan akan semakin mahal jika menggunakan bandara eksklusif sebagai asal atau tujuan, ada juga beberapa bandara yang memiliki harga tiket penerbangan yang tinggi dikarenakan hal-hal seperti tempat yang strategis dan kebijakan pemerintah setempat yang membuat harga tiket penerbangan menjadi mahal. Pemilihan kota mempengaruhi harga tiket pesawat, namun semakin populer kota tersebut tidak membuat harga tiket pesawat menjadi lebih mahal, seperti Los Angeles, dan New York merupakan kota yang populer, namun tidak masuk ke dalam daftar kota yang memiliki harga tiket pesawat yang mahal. Jumlah jarak penerbangan tidak menentukan harga tiket pesawat, namun harga tiket pesawat untuk jarak penerbangan di atas 1118 mil, yang merupakan rata-rata dalam dataset, memiliki harga tiket pesawat yang tinggi. Peristiwa besar seperti Covid-19, 9/11, dan Inflasi sangat mempengaruhi naik turunnya harga tiket pesawat, terbukti pada tahun 2020 saat Covid-19 merajalela, beberapa bandara mengeluarkan kebijakan pembatasan penerbangan yang membuat aktivitas penerbangan menjadi sulit, kurangnya peminat, dan kekurangan staf yang membuat harga tiket pesawat menjadi sangat rendah. Pemilihan waktu (kuartal) mempengaruhi harga tiket penerbangan, harga tiket penerbangan akan meningkat pada saat hari libur nasional seperti pada saat kuartal ke-3 atau pada saat liburan musim panas, sehingga membuat permintaan menjadi tinggi dan tentu saja membuat harga tiket penerbangan menjadi naik. Secara eksplisit tidak ditemukan faktor yang menonjol, namun dengan menggabungkan beberapa faktor seperti pemilihan kota, bandara, dan waktu penerbangan maka akan didapatkan tren atau pola harga tiket pesawat.

Kata Kunci— airport; analisis; biaya, data; faktor; penerbangan.

Exploratory Data Analysis in a Case Study of US Flight Routes and Fares

Abstract— The aviation industry is one of the sectors of the global economy that has a huge influence, for the United States itself aviation is one of the largest economic sectors, ranging from national to international flights with uncertain distribution of flight ticket prices. The analysis was conducted to find the determining factors to the main factors that determine the price of domestic flight tickets in the United States. The analysis was carried out using the EDA (Exploratory Data Analysis) method, and combined with feature engineering, external

data in the form of news, and relevant articles. There are several findings obtained after analysing the US Flight Routes and Fares (1993-2023) dataset, which is a dataset that accommodates domestic flight records of the United States. The discovery obtained is that flight ticket prices will be more expensive if you use exclusive airports as origins or destinations, there are also some airports that have high flight ticket prices due to things such as strategic places and local government policies that make flight ticket prices expensive. The choice of city affects the price of airline tickets, but the more popular the city is, it doesn't make it more expensive, like Los Angeles, and New York is a popular city, but it doesn't make it to the list of cities that have expensive flight ticket prices. The number of flight distances does not determine the price of the flight ticket, but the flight ticket price for a flight distance above 1118 miles, which is the average in the dataset, has a high average flight ticket price. Major events such as Covid-19, 9/11, and Inflation greatly affect the ups and downs of flight ticket prices, as evidenced by the fact that in 2020 when Covid-19 was rampant, several airports issued policies to restrict flights which made flight activities difficult, lack of interest, and staff shortages that made flight ticket prices very low. The choice of time (quarter) affects the price of flight tickets, flight ticket prices will increase during national holidays such as during the 3rd quarter or at the time of summer holidays, making demand high and of course increasing flight ticket prices. Explicitly, no prominent factors were found, but by combining several factors such as the selection of cities, airports, and flight times, a trend or pattern for flight ticket prices will be obtained.

Keywords— *airport; analysis; data; factor; flight; fare.*

I. PENDAHULUAN

Industri penerbangan merupakan salah satu sektor perekonomian global yang sangat berpengaruh dan penting. Tidak hanya digunakan dalam konteks penggunaan komersial untuk kepentingan wisatawan seperti traveling, namun seiring berjalannya waktu demand dari sektor penerbangan juga semakin meningkat dalam sektor bisnis, pekerjaan, atau kepentingan lainnya. Tingginya demand tersebut membuat harga penerbangan tiket pesawat yang kian melonjak dari tahun ke tahunnya[1]. Tidak hanya demand yang menjadi variable atau faktor penentu untuk harga penerbangan yang sangat bervariasi. Penentuan harga penerbangan sering disebut sangat complex, hal ini dikarenakan pasar atau market yang sangat luas, kapasitas penerbangan, kualitas penerbangan, dan juga dikarenakan permintaan agregat yang tidak pasti [2]

Banyak faktor-faktor yang menjadi penentu harga penerbangan, namun tidak menutup kemungkinan bisa menemukan faktor-faktor utama yang menjadi penentu harga penerbangan. Dalam konteks penelitian ini, data diambil dari data penerbangan Amerika Serikat, pada tahun 1993-2024. Dataset ini berisikan informasi terkait rute penerbangan domestik di Amerika Serikat dan harga tiketnya yang sudah dikumpulkan selama lebih dari 30 tahun. Dataset ini diambil dari website Kaggle dengan judul "US Airline Flight Routes and Fares (1993-2024).[3]

Penggunaan data dari tahun ke tahun semakin marak digunakan. Tidak bisa dipungkiri bahwa salah satu alasan mengapa penggunaan data semakin meningkat dari tahun ke tahunnya adalah salah satu sumber yang bisa dipercaya. Namun apalah arti data jika hanya dikumpulkan dan tidak dipergunakan semaksimal mungkin. Exploratory Data Analysis merupakan salah satu metode dimana sebuah data yang sudah dikumpulkan, akan di-eksplor dan akan dilakukan sebuah analisis yang terkakulasi sehingga menghasilkan sebuah jawaban dari sebuah tren, anomaly, ataupun ketidakpastian sebuah kumpulan data

Terdapat banyak alasan mengapa suatu penerbangan lebih mahal atau lebih murah dari penerbangan yang lainnya jika hanya melihat secara sekilas. Penelitian ini menggunakan kumpulan data selama kurang lebih 30 tahun. Tentunya sebuah anomaly, tren yang tidak tentu, dan ketidakpastian data akan terlihat, maka dari itu penelitian ini menggunakan metode Exploratory Data Analysis (EDA). Tujuan dilakukannya penelitian ini ialah untuk meng-analisis pola, tren, dan faktor-faktor yang mempengaruhi harga tiket penerbangan di Amerika Serikat. Metode EDA memungkinkan untuk menyelidiki karakteristik utama data secara menyeluruh, mendeteksi anomaly, dan menemukan atau mengungkapkan pola yang tersembunyi. Hasil dari analisis ini bisa digunakan untuk membantu maskapai penerbangan untuk menyusun strategi untuk menentukan harga agar bisa bersaing di pasarnya, serta membantu konsumen untuk memahami faktor-faktor yang menjadi penentu utama dalam penentuan harga penerbangan

II. KAJIAN TEORI

A. Teori Analisis Data

Hidup manusia tidak lepas dari data, secara harfiah hasil dari apa yang telah dilakukan manusia dapat dijadikan menjadikumpulan data. Data sendiri diartikan sebagai kumpulan fakta, pengukuran, gambaran, atau pengamatan yang dikumpulkan melalui sebuah proses yang telah terstruktur. Data sendiri dapat berbentuk angka, teks, gambar, suara, atau segala bentuk yang dapat dianalisis. Jika analisis data digambarkan sebagai proses kegiatan memasak, maka data merupakan bahan utama atau bahan mentah untuk analisis dan pengambilan Keputusan. Data juga bisa diartikan sebagai elemen dasar dalam sebuah penelitian. Data ini merupakan faktor yang mempengaruhi desain penelitian serta teknik pengumpulan analisis apa yang akan digunakan. Pemahaman tentang data bagi seorang melakukan analisis sangat penting untuk proses pemilihan metode penelitian yang sesuai, baik melewati pendekatan kuantitatif, kualitatif, atau campuran. [17]

a. Pembagian Jenis Data

Data dapat diklasifikasikan atau dibagi sesuai atau berdasarkan sifat, skala pengukuran data tersebut, dan bagaimana data tersebut akan diolah. Menurut teori statistik dan ilmu data yang sudah ada, data diklasifikasikan sebagai berikut:

1. Data Kuantitatif Secara singkat data kuantitatif merupakan data yang bersifat numerik atau berbentuk angka yang dapat dihitung dan diukur. Data kuantitatif sendiri sering digunakan dalam analisis statistika. Data kuantitatif sendiri bisa diklasifikasikan lebih lanjut sebagai berikut:

- a. Data Diskrit Data ini terdiri dari nilai-nilai tertentu yang dapat dihitung dan tidak berbentuk pecahan atau desimal. Contoh dari data diskrit sendiri Universitas Kristen Maranatha 25 seperti jumlah penumpang, jumlah pesawat yang melewati jalur tertentu [15]
 - b. Data Kontinu Data ini berbeda dari data diskrit, rentang nilai data ini lebih luas dari data diskrit atau dalam kata lain data ini bisa berbentuk desimal atau pecahan, contohnya jarak antara bandara awal ke bandara tujuan, durasi waktu, dan berat luggage pesawat.
 2. Data Kualitatif Sering juga disebut sebagai data kategoris, merupakan data yang bersifat deskriptif, data kualitatif secara umum tidak dinyatakan dalam angka. Data kualitatif sendiri bisa diklasifikasikan sebagai berikut:
 - a. Data Nominal Digunakan untuk pengelompokan data yang tanpa urutan atau berbentuk hierarki tertentu. Contohnya seperti jenis kelamin, warna pesawat, atau nama kota asal [15]
 - b. Data Ordinal Menunjukkan data urutan atau peringkat dari kumpulan data, namun jarak antar data sendiri tidak dapat diukur secara numerik. Contohnya seperti tingkat kepuasan pelanggan.
 3. Data Berdasarkan Sumber Data juga bisa dibagi sesuai dari sumber data tersebut dikumpulkan atau diperoleh, dapat diklasifikasikan sebagai berikut:
 - a. Data Primer Data primer sendiri dikumpulkan langsung oleh seorang yang akan melakukan penelitian, data primer biasanya bisa didapatkan melalui wawancara, survei ataupun eksperimen.
 - b. Data Sekunder Data sekunder sendiri dikumpulkan dari data atau sumber data yang sudah ada, seperti laporan, jurnal, ataupun database publik [15]. Universitas Kristen Maranatha 26 Dalam konteks penelitian ini, data didapatkan dari dataset yang sudah ada, yang diambil dari website kaggle. Maka data yang digunakan dalam penelitian ini diklasifikasikan sebagai data sekunder [1].
- b. Pengumpulan Data
- Pengumpulan data merupakan langkah penting yang dilakukan setelah mengetahui jenis data apa yang akan digunakan. Dalam kasus penelitian ini data yang digunakan merupakan sumber yang sudah tersedia, maka sumber data yang relevan bisa diklasifikasikan dalam bentuk data sekunder. Terdapat beberapa teknik pengumpulan data yang dapat dilakukan, pemilihan pengumpulan data yang tepat disesuaikan dengan tipe data apa yang akan dikumpulkan dan diperoleh. Terdapat beberapa teknik pengumpulan data yang dapat dilakukan yaitu teknik pengumpulan data kuantitatif, kualitatif, dan mixed methods.
- i. Penelitian Kuantitatif
- Dalam penelitian kuantitatif, data dapat dikumpulkan dengan menggunakan instrumen yang dapat diukur secara numerik, beberapa teknik yang umum dilakukan dalam penelitian kuantitatif merupakan survei, kuesioner, dan tes. Survei sendiri merupakan metode yang paling umum digunakan untuk sampel data yang lebih besar, hal tersebut dikarenakan efisiensinya untuk mengumpulkan data dalam waktu yang singkat [15]. Instrumen survei sendiri biasanya berbentuk pertanyaan tertutup yang dapat mudah diukur dan dianalisis secara statistik. Selain survei, terdapat juga metode observasi yang sering digunakan dalam penelitian kuantitatif, hal ini dilakukan disaat peneliti mengamati perilaku atau kejadian yang telah ditentukan sebelumnya dalam lingkungan yang telah terkontrol [18].
- ii. Penelitian Kualitatif
- Penelitian kualitatif sendiri cenderung menggunakan metode pengumpulan data yang lebih fleksibel dan mendalam, contohnya seperti wawancara, observasi partisipan, dan analisis dokumen. Wawancara merupakan metode yang umum digunakan dimana wawancara memungkinkan peneliti untuk lebih bisa eksplor pandangan, pengalaman dan pemahaman subjek lebih dalam lagi. Selain wawancara, observasi partisipatif merupakan metode yang umum juga, metode ini memungkinkan peneliti terlibat langsung dalam lingkungan atau konteks yang sedang diteliti. Metode ini memungkinkan pengamatan perilaku dan interaksi secara langsung.
- iii. Mixed Method
- Metode campuran atau mixed methods menggabungkan kedua pendekatan sebelumnya, hal ini dilakukan untuk mendapatkan data yang lebih komprehensif. Pengumpulan data ini melibatkan kombinasi survei dan wawancara atau kombinasi lainnya yang bisa dilakukan dalam method method sebelumnya. Pendekatan ini memungkinkan untuk mendapatkan data kuantitatif yang kuat dan juga data kualitatif yang lebih mendalam.
- c. Pengolahan Data
- Pengolahan data merupakan tahap selanjutnya setelah mengumpulkan data. Pengolahan data sendiri terdiri dari beberapa rangkaian proses, secara singkat pengolahan data merupakan proses pengubahan data mentah menjadi informasi yang digunakan dalam analisis data dan pengambilan keputusan. Terdapat beberapa tahapan dalam pengolahan data, yaitu pembersihan data, transformasi, agregasi, dan validasi data [19].
- i. Pembersihan Data
- Pembersihan data merupakan tahapan atau proses awal dalam pengolahan data, proses ini bertujuan untuk menghilangkan Kumpulan anomaly yang terdapat dalam data seperti data yang tidak lengkap, invalid, tidak konsisten, outliers, dan data duplikat. Pembersihan data sangat penting agar hasil dari analisis data memiliki tingkat akurasi yang tinggi [20].
- ii. Transformasi Data
- Transformasi data merupakan proses untuk mengubah data ke dalam format yang disesuaikan untuk kegiatan analisis lebih lanjut. Proses ini melibatkan normalisasi data, standarisasi, encoding, pengkategorian, dan pembuatan

fitur baru dalam data. Tujuannya adalah untuk meningkatkan kualitas dan kegunaan data dalam model yang akan digunakan [17].

- iii. Agregasi Data
Proses ini dilakukan untuk meringkas data dalam bentuk yang lebih sederhana, seperti menghitung rata-rata, total, ataupun jumlah data tertentu dalam sebuah kategori. Langkah ini membantu dalam menganalisis pola atau tren yang terdapat dalam kumpulan data [17].
- iv. Validasi Data
Proses ini memastikan bahwa data yang telah diolah melalui proses sebelumnya memenuhi standar yang ditentukan. Langkah ini mencakup pemeriksaan tipe data, struktur data, dan rentang nilai. Validasi data ditujukan untuk mencegah kesalahan dalam analisis agar menghasilkan informasi yang akurat. Proses pengolahan data yang dilakukan dengan baik dan terstruktur sangat penting dalam penelitian data science, hal ini dikarenakan untuk meminimalisir kesalahan dalam kegiatan analisis data dan menghasilkan informasi yang kredibel. Proses pengolahan data juga membantu dalam menentukan pola yang tidak terlihat dengan penglihatan secara seksama, agar memungkinkan pengambilan keputusan yang berbasis fakta yaitu data [20].

d. *Exploratory Data Analysis*

Exploratory Data Analysis merupakan langkah yang essential, EDA sendiri memiliki goals atau tujuan sebagai menemukan dan menentukan distribusi data, data outliers, anomali data, dan memperkuat hasil dari hipotesis yang dibuat. Hasil dari kegiatan EDA juga dapat membantu untuk lebih mengenali pattern data yang dianalisis dengan menggunakan visualisasi data. [23]

Hasil dari EDA juga dapat diperkuat dengan informasi tambahan external yang bisa didapatkan dari berita-berita ataupun artikel yang relevan. Hal ini dilakukan untuk memperkuat dan memvalidasi hasil analisis, hipotesa, ataupun kesimpulan yang dihasilkan dari kegiatan analisis data, agar hasilnya lebih akurat, valid, dan relevant dengan apa apa yang dicari dari kegiatan analisis data.

B. Perangkat Pendukung Analisis: Python Pandas

Python Pandas merupakan salah satu library dari Python yang sangat populer digunakan dalam kegiatan analisis dan manipulasi data. Library Pandas menyediakan sistem struktur data yang relatif kuat dan mudah digunakan. DataFrame dan Series merupakan sistem yang mudah digunakan dan memungkinkan pengguna untuk menyusun, menganalisis, dan melakukan visualisasi data secara mudah atau efisien dan efektif. Pandas memungkinkan proses analisis lebih efisien dan terstruktur, hal ini dikarenakan memiliki fungsi bawaan untuk membaca, menulis, manipulasi, dan membersihkan data dalam berbagai format seperti CSV, Excel, SQL, ataupun JSON

a. Struktur Data pada Pandas

Library Pandas menyediakan dua struktur data utama yang dapat digunakan untuk kegiatan analisis data secara mudah dan efisien:

1. Series Merupakan struktur data satu dimensi, dapat digunakan untuk menyimpan berbagai jenis data seperti integer, float, dan objek. Series sendiri sering digunakan untuk menyimpan kolom dari DataFrame ataupun data numerik sederhana.
2. DataFrame Merupakan Struktur data dua dimensi, bentuk dari struktur data ini mirip dengan table pada database. DataFrame sendiri terdiri dari baris dan kolom dan memungkinkan pengguna melakukan kegiatan analisis, manipulasi, dan transformasi data dengan mudah [19].

b. Fungsi Utama Pandas

Terdapat tiga fungsi utama Pandas dalam melakukan analisis data, ketiga fungsi ini sangat krusial dalam kegiatan analisis data. Tiga fungsi utama tersebut adalah pembersihan data, manipulasi data, dan visualisasi data, ketiga fungsi ini memiliki pengertian dan tugas yang berbeda.

1. Pembersihan Data Pandas menyediakan fungsi untuk menangani beberapa anomaly data yang harus dibersihkan seperti, missing values, dan duplicate. Dengan fungsi ini bisa dilakukan transformasi data seperti normalisasi, labeling, dan encoding variabel kategori. Seperti contohnya terdapat fungsi `dropna()` yang digunakan untuk menghapus baris yang memiliki missing values, selanjutnya terdapat `fillna()` untuk mengisi nilai kosong dengan nilai tertentu.
2. Manipulasi Data Terdapat beberapa fungsi yang bisa digunakan untuk melakukan manipulasi data seperti `groupby()`, `merge()`, dan `pivot()`. Dengan ini Pandas memudahkan melakukan manipulasi data untuk kegiatan analisis data terutama saat menggabungkan berbagai table ataupun melakukan agregasi data [19].
3. Visualisasi Data Pandas mendukung beberapa integrasi dalam pustaka visualisasi, beberapa pustaka yang sering digunakan merupakan Matplotlib dan Seaborn untuk pembuatan grafik dan visualisasi data.

c. Keunggulan dan Keterbatasan Pandas

Python Pandas memberikan efisiensi tinggi untuk melakukan analisis data, hal ini dilakukan dengan fungsi-fungsi yang disediakan yang membuat kegiatan analisis data lebih efisien dan efektif, namun terdapat keterbatasan utama Pandas ialah kinerja yang kurang mendukung untuk big data. Dalam kasus dataset yang sangat besar dapat menyebabkan Pandas bekerja secara lambat karena Pandas sendiri sepenuhnya menggunakan memori untuk menyimpan data

C. Visualisasi Data

Visualisasi data merupakan proses penyajian data atau informasi dengan menggunakan atau dalam bentuk grafis. Tujuan dilakukannya visualisasi data sendiri untuk memudahkan, memperjelas, dan menyederhanakan penyajian data atau informasi sehingga lebih memudahkan pemahaman data yang kompleks menggunakan visual seperti grafik, peta, dan diagram. Visualisasi data juga dapat membantu men-identifikasi pola, hubungan antara data dengan data yang lainnya, relevansi data, dan pencarian anomaly data [23]. Tidak hanya untuk memudahkan, menyederhanakan penyampaian data atau informasi, dan pendukung pengambilan keputusan, Visualisasi data juga memperkuat komunikasi antar pengguna dan peneliti walaupun memiliki perbedaan latar belakang. Hal ini dikarenakan visual data memungkinkan setiap orang memahami data secara visual dasar tanpa memerlukan pemahaman statistik yang mendalam[23].

a. Varian Visualisasi Data

Visualisasi data memiliki beberapa jenis yang memiliki kelebihan dan kekurangan masing-masing. Pemilihan varian visualisasi data disesuaikan dengan data apa yang ingin direpresentasikan dalam visualisasi data. Berikut merupakan beberapa varian visualisasi data yang umum digunakan:

1. *Line Chart* digunakan untuk menunjukkan tren atau perubahan nilai dari waktu ke waktu seperti contoh: tren perubahan harga terbang pesawat dari tahun ke tahun.
2. *Histogram/Bar Chart* digunakan untuk menunjukkan distribusi data ataupun komparasi antar kategori data yang sudah dikumpulkan.
3. *Scatter Plot* digunakan untuk menunjukkan korelasi antara variabel data yang terdapat didalam dataset.
4. *Box Plot* digunakan untuk menunjukkan distribusi data dan juga digunakan untuk menunjukkan *outliers* data yang terdapat.

b. Alat Untuk Visualisasi Data

Terdapat beberapa alat, library atau framework yang biasanya digunakan untuk mendukung kegiatan visualisasi data. Beberapa alat ini digunakan sesuai dengan platform atau perangkat lunak yang digunakan, pemilihan alat untuk visualisasi data juga disesuaikan dengan konteks penelitian atau eksplorasi data. Beberapa alat tersebut diantara lain:

1. *Matplotlib & Seaborn* digunakan jika proses analisis data dilakukan menggunakan platform Bahasa pemrograman *Python*. Bisa digunakan untuk visualisasi dasar ataupun visualisasi lanjut.
2. *Tableau & PowerBI* alat visualisasi data berbasis perangkat lunak yang memudahkan visualisasi data tanpa pengetahuan pemrograman.
3. *D3.js* Digunakan untuk melakukan visualisasi data melalui web interaktif[22].

D. Harga Tiket Penerbangan

Industri penerbangan merupakan salah satu industri yang bergerak dalam bidang layanan/jasa. Industri penerbangan menjadi salah satu sektor yang paling berpengaruh dalam perkembangan ekonomi, hal ini didukung dengan beberapa alasan, namun alasan utamanya merupakan dengan transportasi udara banyak orang dapat menjangkau tempat-tempat yang relatif sulit dijangkau dengan mode transportasi darat dan laut, hal ini juga didukung dengan waktu tempuh transportasi udara lebih sedikit atau cepat dibandingkan dengan mode transportasi lainnya. Alasan yang sudah dipaparkan menjadi salah satu alasan mengapa banyak orang memilih penerbangan sebagai mode transportasi untuk perjalanan yang jauh atau sulit dijangkau dengan mode transportasi lainnya.

Dengan berjalanannya waktu, demand untuk penggunaan mode transportasi udara semakin meningkat, hal ini membuat harga tiket penerbangan berkembang, namun tidak hanya demand yang tinggi, terdapat beberapa komponen penting yang menjadi penentu harga tiket penerbangan. Komponen-komponen tersebut sangat penting untuk berjalanannya penelitian terkait analisis data harga penerbangan.[24]

a. Maskapai Penerbangan

Maskapai Penerbangan menjadi salah satu komponen penting dalam harga tiket penerbangan, setiap maskapai memiliki kebijakan tersendiri untuk penentuan harga tiket. Namun terdapat satu hal yang menjadi alasan mengapa harga tiket setiap penerbangan berbeda-beda, hal itu merupakan kategori maskapai penerbangan. Secara umum terdapat dua kategori maskapai penerbangan, full service dan low-cost carrier. Maskapai dengan kategori full-service menawarkan layanan yang lebih lengkap dibandingkan dengan kategori low-cost carrier, layanan tersebut termasuk makanan, hiburan, dan bagasi. Sebaliknya maskapai low cost carrier memiliki biaya operasional yang rendah dan memiliki layanan yang dasar, namun memiliki biaya tambahan untuk layanan yang opsional.

Model bisnis yang berbeda ini mempengaruhi strategi harga tiket penerbangan setiap maskapai penerbangan. Penetapan model bisnis untuk setiap maskapai disesuaikan dengan tujuan pasarnya, dimana kategori full-service menargetkan pasar bisnis yang lebih high-end, sebaliknya untuk low-cost carrier menargetkan pasar bisnis yang lebih terbatas

b. Kategori Kelas Penerbangan

Pemilihan kelas penerbangan menjadi salah satu faktor yang mempengaruhi harga tiket penerbangan. Setiap kategori memiliki fitur atau service yang berbeda dengan alasan yang berbeda. Terdapat tiga kelas penerbangan yang biasanya ditawarkan oleh maskapai penerbangan, diantaranya:

1. Ekonomi

Memiliki kelas dan harga yang paling terjangkau, namun menyediakan layanan yang dasar yang memenuhi kebutuhan dasar untuk penerbangan. Ditujukan untuk pelanggan yang tidak memiliki budget yang tinggi dan tidak membutuhkan layanan ekstra.

2. Bisnis
Memiliki ruang yang lebih luas untuk tempat duduk, memiliki layanan makanan yang lebih luas, dan memiliki ruang tunggu premium.
 3. First Class
Kelas yang paling eksklusif, hal ini dikarenakan first class memiliki layanan dan kenyamanan yang lebih baik dibandingkan kelas penerbangan bisnis dan ekonomi[26].
- c. Komponen Penentu Harga Tiket Penerbangan
- Banyak sekali komponen penentu harga tiket penerbangan, tujuan utama dalam penelitian analisis data ini adalah mencari variabel penentu lebih spesifiknya dari dataset yang digunakan, namun secara umum terdapat empat faktor utama yang menjadi komponen penentu harga tiket, diantaranya;
1. Jarak dan Durasi Penerbangan
 2. Permintaan dan Penawaran
 3. Kebijakan Maskapai
 4. Biaya Operasional

III. ANALISIS DAN RANCANGAN SISTEM

A. Analisis Kolom Fare dengan Year

Analisis pertama yang dilakukan adalah mengidentifikasi korelasi kolom fare dan year/quarter. Harga tiket penerbangan mungkin sangat dipengaruhi dengan kejadian yang terjadi pada tahun tersebut, bahkan pada kuartal atau musim tertentu. Hal ini bisa menjadi faktor penting dalam penentuan alasan perubahan harga tiket penerbangan.

Terdapat juga hasil yang diharapkan yaitu untuk menemukan pola waktu tertentu seperti harga cenderung naik setiap tahun, atau harga cenderung naik pada kuartal tertentu seperti pada saat musim liburan, ataupun harga cenderung turun pada waktu tertentu. Dan tujuan selanjutnya adalah mencari tahun dengan distribusi harga terbesar.

B. Analisis Kolom Fare dengan City1/City2

Analisis selanjutnya adalah mengidentifikasi apakah kota tujuan ataupun kota asal memiliki peran penting dalam distribusi harga terlebih lagi di dalam ruang lingkup negara Amerika Serikat. Pada tahun 2023 dilakukan sebuah studi terhadap peningkatan airline fare pada setiap kota di Amerika yang berubah setiap saat, menurut studi tersebut peningkatan airline fare terhadap beberapa kota di Amerika dikarenakan faktor popularitas kota tersebut, ataupun jumlah events yang terjadi di kota tersebut.

Hasil yang diharapkan dari analisis ini adalah menemukan kota tujuan/kota asal di Amerika yang memiliki distribusi harga penerbangan yang paling tinggi, dan paling rendah.

C. Analisis Kolom Fare dengan nsmiles

Analisis yang terakhir adalah menilai pengaruh jarak penerbangan yang bisa didapatkan dari kolom nsmiles terhadap harga tiket rata-rata. Tujuannya adalah untuk memahami apakah harga tiket meningkat secara signifikan atau harga tiket tidak terlalu dipengaruhi jarak penerbangan.

Hasil yang diharapkan dari analisis tersebut adalah apakah harga tiket cenderung meningkat dengan jarak penerbangan?, dan apakah rute jarak yang relatif pendek dapat memiliki variasi harga yang lebih besar atau bervariasi.

D. Analisis Kolom Fare dengan Passengers

Jumlah penumpang bisa diartikan sebagai jumlah demand terhadap penerbangan tersebut. Analisis ini bertujuan untuk memahami bagaimana jumlah penumpang mempengaruhi jumlah tiket penerbangan, terutama dalam konteks permintaan pasar atau demand.

Hasil yang diharapkan dari analisis ini adalah untuk melihat bagaimana jumlah penumpang dalam suatu rute penerbangan mempengaruhi harga penerbangan tersebut

IV. IMPLEMENTASI

A. Persiapan Dataset

Pada sub-bab ini, diperlihatkan untuk bagian persiapan dataset yang akan digunakan, termasuk juga import pustaka(library) yang akan digunakan untuk kepentingan analisis data. Beberapa library yang digunakan diantara lain pandas, NumPy, Matplotlib, dan Seaborn. Dataset yang digunakan sendiri diimport melalui pustaka pandas, yang akan disimpan kedalam sebuah variabel dataframe dengan nama df.

Hal yang selanjutnya dilakukan ialah memeriksa info dari data yang terdapat dari dataset yang digunakan, hal ini termasuk untuk melihat tipe data, dan jumlah data kosong yang terdapat pada kolom Non-Null Count. Untuk melakukan hal tersebut bisa dilakukan dengan cara menjalankan function df.info(). Dengan melihat info dari dataset yang akan digunakan, maka akan terlihat kolom mana saja yang memiliki missing value.

```
Data columns (total 23 columns):
 #  Column          Non-Null Count  Dtype  
 --- 
  0  tbl             245955 non-null  object  
  1  year            245955 non-null  int64  
  2  quarter         245955 non-null  int64  
  3  citymarketid_1  245955 non-null  int64  
  4  citymarketid_2  245955 non-null  int64  
  5  city1           245955 non-null  object  
  6  city2           245955 non-null  object  
  7  airportid_1     245955 non-null  int64  
  8  airportid_2     245955 non-null  int64  
  9  airport_1       245955 non-null  object  
  10  airport_2      245955 non-null  object  
  11  nsmiles         245955 non-null  int64  
  12  passengers       245955 non-null  int64  
  13  fare            245955 non-null  float64 
  14  carrier_lg      244415 non-null  object  
  15  lange_ms        244415 non-null  float64 
  16  fare_lg          244415 non-null  float64 
  17  carrier_low     244343 non-null  object  
  18  lf_ms            244343 non-null  float64 
  19  fare_low         244343 non-null  float64 
  20  Geocoded_City1  206749 non-null  object  
  21  Geocoded_City2  206749 non-null  object  
  22  tbl1apk          245955 non-null  object  
 dtypes: float64(5), int64(8), object(10)
 memory usage: 43.24+ MB
```

Gambar 1. df.info()

Setelah melihat info dari dataset yang akan di-analisa, maka hal yang selanjutnya dilakukan adalah untuk melihat kolom yang mempunyai data null atau missing value. Hal tersebut bisa dilakukan dengan menggunakan fungsi berikut:

```
tbl          0
Year         0
quarter      0
citymarketid_1  0
citymarketid_2  0
city1         0
city2         0
airportid_1    0
airportid_2    0
airport_1      0
airport_2      0
nsmiles        0
passengers     0
fare           0
carrier_lg     1540
lange_ms       1540
fare_lg         1540
carrier_low     1612
lf_ms           1612
fare_low         1612
Geocoded_City1 39206
Geocoded_City2 39206
tbl1apk         0
dtype: int64
```

Gambar 2. Missing Value

Setelah melihat beberapa kolom yang memiliki missing value, hal yang selanjutnya dilakukan ialah melihat terdapatnya data-data yang ter-duplikasi didalam dataset. Hal tersebut bisa dilakukan dengan menggunakan function



Gambar 3. Duplicated Data Sum

B. Analisa Deskriptif Dataset

Setelah melihat data duplikasi yang terdapat pada dataset yang digunakan, selanjutnya yang dilakukan ialah melihat deskripsi dataset, hal tersebut dilakukan dengan built-in function yaitu *df.describe()*. Yang dihasilkan merupakan hasil nilai deskriptif dari dataset, termasuk *count*, *Mean*, *nilai kuartal*, hingga nilai tertinggi untuk setiap kolom. Berikut hasil dari penggunaan *function describe*

[13]:	Year	quarter	citymarketid_1	citymarketid_2	airportid_1	airportid_2	nsmiles	passengers	fare	large_ms
count	245955.000000	245955.000000	245955.000000	245955.000000	245955.000000	245955.000000	245955.000000	245955.000000	245955.000000	244415.000000
mean	2008.524124	2.479153	31556.430201	32180.117086	12437.099986	13249.889525	1189.812319	299.476795	218.979587	0.665252
std	8.703364	1.122149	1089.872880	1232.464184	1431.665257	1425.810159	703.143472	511.389486	82.372486	0.224635
min	1993.000000	1.000000	30135.000000	30189.000000	10135.000000	10466.000000	109.000000	0.000000	50.000000	0.003800
25%	2001.000000	1.000000	30721.000000	30994.000000	11193.000000	12197.000000	626.000000	21.000000	164.620000	0.480000
50%	2008.000000	2.000000	31423.000000	32211.000000	12266.000000	13303.000000	1023.000000	113.000000	209.320000	0.652400
75%	2016.000000	3.000000	32467.000000	33192.000000	13487.000000	14679.000000	1736.000000	339.000000	262.890000	0.871900
max	2024.000000	4.000000	35412.000000	35628.000000	16440.000000	15919.000000	2724.000000	8301.000000	3377.000000	1.000000

Gambar 4. Dataset Describe

C. Analisis Fare dengan Year/Quarter

Analisis ini bertujuan untuk melihat distribusi harga dari tahun ke tahun dengan juga menambahkan informasi waktu kuartil untuk lebih spesifik lagi. Dari hasil analisis sini yang didapatkan ialah informasi tentang tahun-tahun dengan rata-rata harga tiket tertinggi, waktu spesifik terhadap tahun dan kuartil pada tahun tersebut, dan tren distribusi harga tiket penerbangan terhadap tahun dan kuartil penerbangan tersebut terjadi. Dari hasil analisis ini bisa didapatkan beberapa insight yang bisa didapatkan, dan dapat diperkuat dengan pencarian berita yang relevan dan tentunya terkait terhadap insight tersebut.

Untuk analisis yang pertama dilakukan adalah, untuk melihat 10 tahun dalam dataset dengan rata-rata harga tiket penerbangan yang relatif tinggi dan rendah dengan hasil berikut:

	Year	Average Fare
0	2024	257.243160
1	2014	251.657087
2	2023	249.931843
3	2022	249.667511
4	2015	247.992015
5	2013	244.911938
6	2019	241.179140
7	2012	241.111118
8	2016	240.550512
9	2017	240.096521

Gambar 5. Top 10 Expensive Year

Temuan yang didapatkan dari hasil analisis yang dilakukan ialah: Tahun dengan rata – rata harga tiket penerbangan tertinggi dipegang oleh tahun 2024, diambil dari cnbc.com, pada tahun 2024 harga tiket penerbangan amerika meningkat 25% dikarenakan inflasi yang terjadi pada Amerika Serikat, namun jika dibandingkan dengan tahun 2022 yang disebut-sebut memiliki tingkat inflasi tertinggi sejak tahun 1998, tahun 2024 masih tetap memiliki rata-rata harga tiket yang lebih tinggi, jika melihat dari “Consumer Price Index of All Urban Customers: Airline in U.S City Average” tahun 2024 memang memiliki index rata-rata harga tiket yang lebih stabil di konsisten di angka \$257, dibandingkan dengan tahun 2022 yang memiliki kenaikan index yang drastis dari bulan februari dengan index \$226 hingga bulan Juni dengan index \$321 Dolar [4]. Hal ini bisa diperkuat dengan melihat dari dataset yang digunakan.

	Year	quarter	fare
110	2022	2	268.063067
112	2022	4	259.263496
117	2024	1	257.243160
114	2023	2	256.990807
78	2014	2	256.893597
79	2014	3	256.395050
82	2015	2	254.436633
111	2022	3	253.461567
116	2023	4	253.224050
81	2015	1	252.892771

Gambar 6. Top 10 Expensive Year and Quarter

Dilihat dari list yang dihasilkan, pada tahun 2022 pada kuartil 2, merupakan entry data dengan rata-rata harga tiket tertinggi, data tersebut terbukti jika kita melihat index yang digunakan sebelumnya.

Analisa selanjutnya adalah untuk melihat data dengan tahun yang memiliki rata-rata harga tiket terendah.

	Year	quarter	fare
110	2022	2	268.063067
112	2022	4	259.263496
117	2024	1	257.243160
114	2023	2	256.990807
78	2014	2	256.893597
79	2014	3	256.395050
82	2015	2	254.436633
111	2022	3	253.461567
116	2023	4	253.224050
81	2015	1	252.892771

Gambar 7. Top 10 Cheap Year and Quarter

Dilihat dari list yang dihasilkan, pada tahun 2022 pada kuartil 2, merupakan entry data dengan rata-rata harga tiket tertinggi, data tersebut terbukti jika kita melihat index yang digunakan sebelumnya.

Analisa selanjutnya adalah untuk melihat data dengan tahun yang memiliki rata-rata harga tiket terendah.

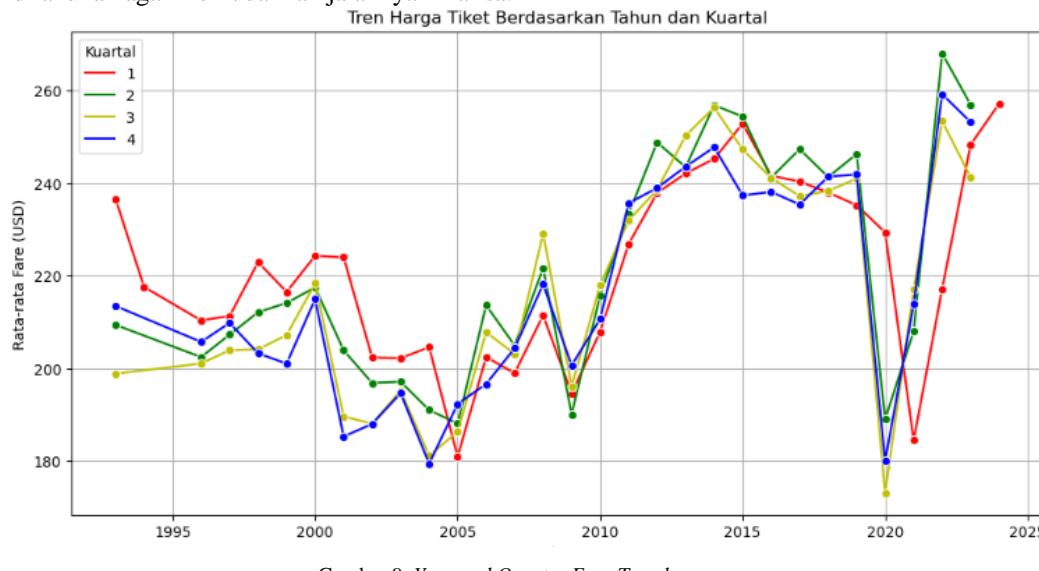
	Year	Average Fare
0	2005	186.991831
1	2004	189.021593
2	2020	193.615156
3	2002	193.810930
4	2009	195.406621
5	2003	197.277303
6	2001	200.860409
7	2007	202.886365
8	1996	204.900283
9	2006	205.197657

Gambar 8. Top 10 Cheap Year

Dari hasil yang didapatkan, bisa disimpulkan rentang tahun 2001-2007 merupakan rentang tahun dengan rata-rata harga tiket terendah, hal ini diperkuat dengan terjadinya krisis ekonomi dot-com bubble [6] pada tahun 2001, hal ini menyebabkan harga tiket penerbangan mengalami penurunan harga atau penetapan harga yang rendah, serta terjadinya tragedy 9/11 yang terjadi pada tahun 2001 bulan September [7]. Hal yang bisa mempengaruhi selanjutnya ialah munculnya beberapa *low-cost carrier* yang menyebabkan persaingan harga pasar pada harga tiket penerbangan.

Tahun 2020 merupakan satu-satunya tahun diatas 2010 yang muncul dalam list tahun yang memiliki rata-rata harga yang termurah, hal tersebut terjadi dikarenakan munculnya pandemic Covid-19. Pandemic ini menyebabkan penurunan demand untuk kegiatan travel, perubahan regulasi penerbangan, dan pada saatnya belum ditemukan *recovery patterns* untuk kegiatan penerbangan, hal ini tentunya sangat mempengaruhi harga tiket penerbangan. [8].

Analisa selanjutnya menggunakan visualisasi lineplot untuk melihat tren harga tiket penerbangan berdasarkan tahun dan kuartalnya, hal ini dilakukan agar memudahkan jalannya Analisa.



Gambar 9. Year and Quarter Fare Trend

Temuan yang didapatkan dari visualisasi yang dilakukan ialah, terdapat penurunan harga secara signifikan dari tahun 2019-2020, hal ini dikaitkan dengan menyebarnya pandemi covid-19 dimana membuat banyaknya regulasi yang dibuat pemerintah sekitar untuk kepentingan kesehatan dan keamanan masyarakat negara Amerika Serikat untuk kegiatan penerbangan pada masanya. [8] sehingga membuat kegiatan penerbangan sangat minim dan menyebabkan harga tiket penerbangan sangat rendah.

Penemuan selanjutnya merupakan pada tahun 2001 pada kuartal 3 (September), penurunan harga tiket penerbangan, hal tersebut bisa dipastikan dikarenakan terjadinya peristiwa yang sekarang disebut 9/11 yang merupakan peristiwa penyerangan teroris dengan menggunakan pesawat yang ditabrakan ke Gedung twin towers. Kejadian ini memberikan dampak nasional hingga internasional. Dampak dari peristiwa ini juga membuat pemerintah AS setempat gentar dan membuat peraturan darurat untuk menghentikan hingga memperketat pemeriksaan jalur penerbangan. [9]

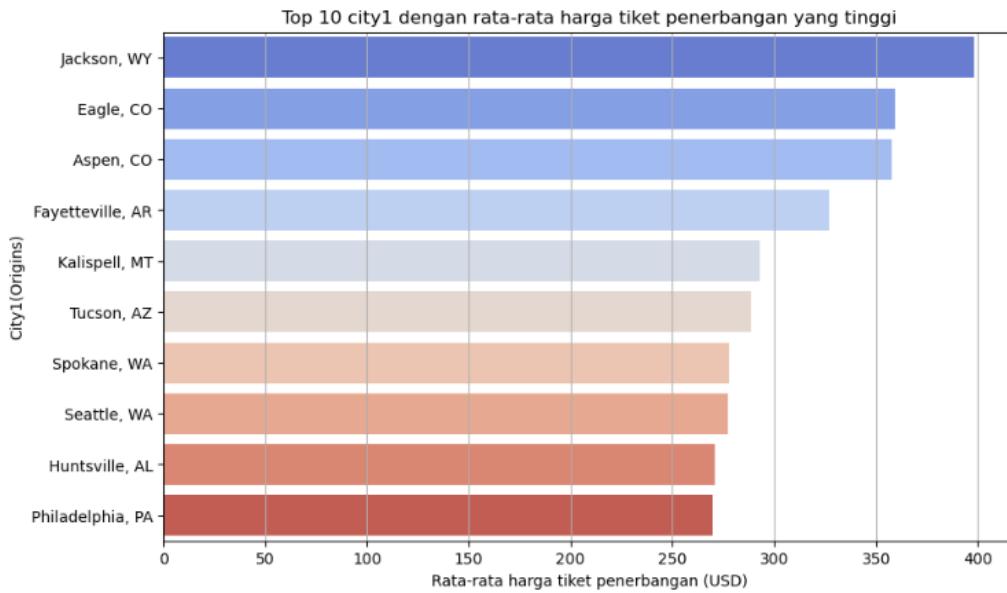
Terdapat 2 kuartal yang memiliki rata-rata harga yang tinggi, hal ini bisa dikaitkan dengan waktu libur nasional yang dilakukan pada negara Amerika Serikat, pada Kuartal 1 (Januari-Maret) terdapat beberapa waktu liburan, yaitu new-years (Januari), Saint-Patrick Day(Maret), yang dijadikan hari libur nasional. Kuartal 3 memiliki tren yang relatif tinggi, hal ini dikarenakan pada kuartal 3 terdapat hari libur nasional summer vacation yang biasanya bisa terjadi beberapa minggu hingga beberapa bulan, hal ini tentunya menaikkan demand, yang nantinya akan dianalisa lebih dalam lagi dengan kolom passenger.

D. Analisis Kolom Fare dengan City1/City2

Analisis berikutnya adalah untuk melihat beberapa kota(keberangkatan dan destinasi) dengan rata-rata harga yang tinggi, beberapa kota dalam negara Amerika Serikat memiliki popularitas yang lebih tinggi dibandingkan dengan kota yang lainnya di dalam negara Amerika Serikat, hal tersebut dikarenakan beberapa kota tersebut merupakan tujuan destinasi, tujuan business, dan metropolitan, dengan tujuan penerbangan tersebut, dapat meningkatkan harga tiket penerbangan terhadap kota tersebut.

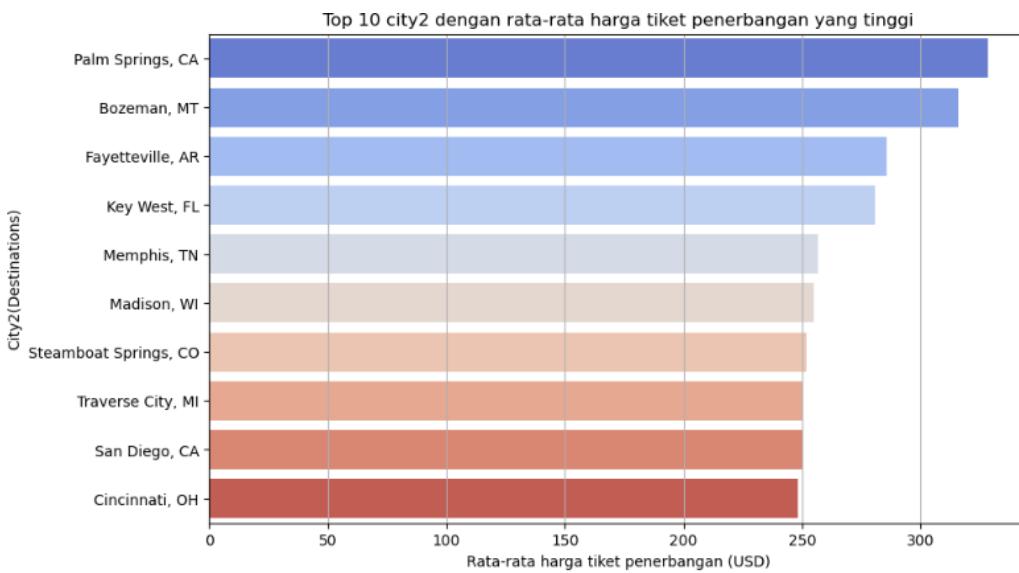
Maka dari itu analisa ini menghasilkan visualisasi untuk 10 kota sering muncul dalam dataset yang digunakan dengan harga tiket penerbangan yang tinggi, dan akan dilakukan visualisasi untuk melihat distribusi harga penerbangan tersebut.

Untuk visualisasi akan dipisah, terdapat visualisasi khusus untuk city1 dan khusus untuk city2. Yang pertama akan dilihat adalah visualisasi city1 sebagai berikut:



Gambar 10. Top 10 Expensive City Origins

Penemuan yang didapatkan dari visualisasi diatas merupakan, beberapa kota yang ditampilkan di dalam visualisasi yang dilakukan merupakan kota-kota wisata, hal tersebut membuat rata-rata harga tiket penerbangan lebih tinggi dibandingkan dengan kota-kota lainnya, hal ini didukung dengan sangat sedikitnya jumlah rute penerbangan yang mengarah ke kota-kota yang terdapat di list hasil visualisasi. Kota-kota seperti *Jackson Wyoming*, *Eagle Colorado*, dan *Aspen Colorado* merupakan beberapa kota yang disebut sebagai tujuan wisata, seperti yang sudah dibahas sebelumnya diatas, kota-kota tersebut memiliki rute yang sedikit atau terbatas, serta ukuran airport yang digunakan relative kecil, sehingga tidak bisa menampung banyaknya penerbangan yang akan masuk ataupun keluar dari airport tersebut. Dikutip dari salah satu komentar yang didapatkan dari website *TripAdvisor* “Airport kota *Jackson* terletak di dalam taman nasional, airport tersebut tidak memiliki banyak penerbangan” [10]. Sama seperti halnya seperti kota *Eagle Colorado* dan *Aspen Colorado*, kota-kota tersebut memiliki rata-rata harga tiket penerbangan yang tinggi yang disebabkan sedikitnya jalur rute penerbangan yang menuju kota tersebut, serta tingginya demand penerbangan tersebut disaat *High Peak Season* dibandingkan sedikitnya jumlah penerbangan yang tersedia, hal ini juga ditambahkan kota-kota tersebut digolongkan sebagai kota yang “Luxurious” [11].



Gambar 11. Top 10 Expensive City Destination

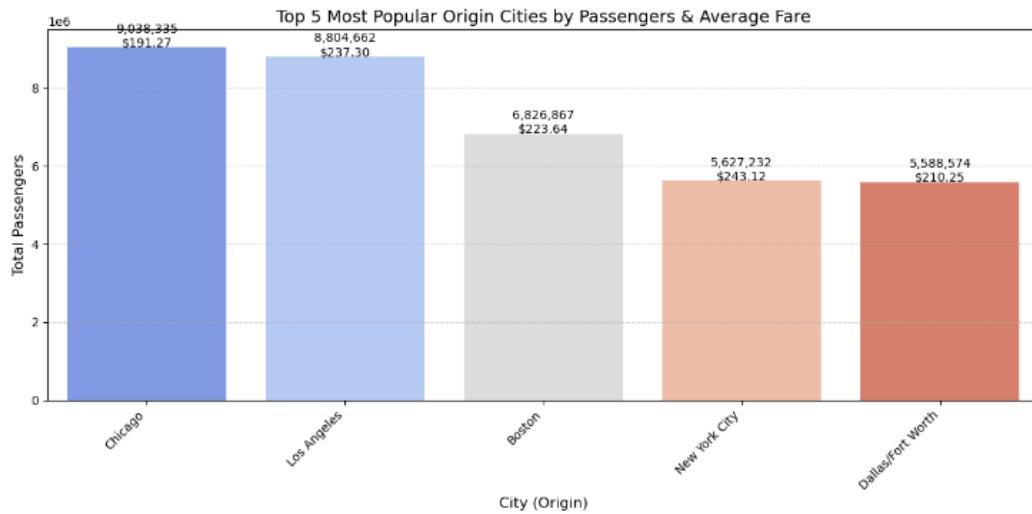
Penemuan yang didapatkan dari visualisasi diatas merupakan, beberapa kota yang ditampilkan didalam visualisasi yang dilakukan merupakan kota-kota wisata, hal tersebut membuat rata-rata harga tiket penerbangan lebih tinggi dibandingkan dengan kota-kota lainnya, hal ini didukung dengan sangat sedikitnya jumlah rute penerbangan yang mengarah ke kota-kota yang terdapat di list hasil visualisasi. Kota-kota seperti Palmsprings California, Bozeman, dan Fayetteville memiliki alasan yang sama untuk mengapa kota-kota tersebut memiliki rata-rata harga tiket penerbangan yang sangat tinggi, yaitu sedikitnya rute penerbangan terhadap kota tersebut, ditambah dengan high demand disaat *High Peak Season* atau saat libur nasional.

Penemuan yang ditemukan dari kedua analisis yang sudah dilakukan adalah kota-kota terkenal seperti Los Angeles, New York, dan Chicago tidak termasuk kedalam list sepuluh kota Amerika dengan rata-rata harga tiket penerbangan yang tinggi

1. Analisis Distribusi Penumpang Penerbangan (City)

Analisis mendalam yang dilakukan terhadap kolom city1 dan city2 adalah bagaimana distribusi harga terhadap lima kota (keberangkatan & tujuan) yang paling popular, tujuan dilakukannya analisis ini merupakan untuk melihat tren distribusi penumpang terhadap harga tiket penerbangan dengan asal ataupun tujuan kota tersebut. Analisa yang dilakukan juga akan menjadi bahan pendukung untuk point demand menjadi salah satu variable yang menentukan harga tiket penerbangan

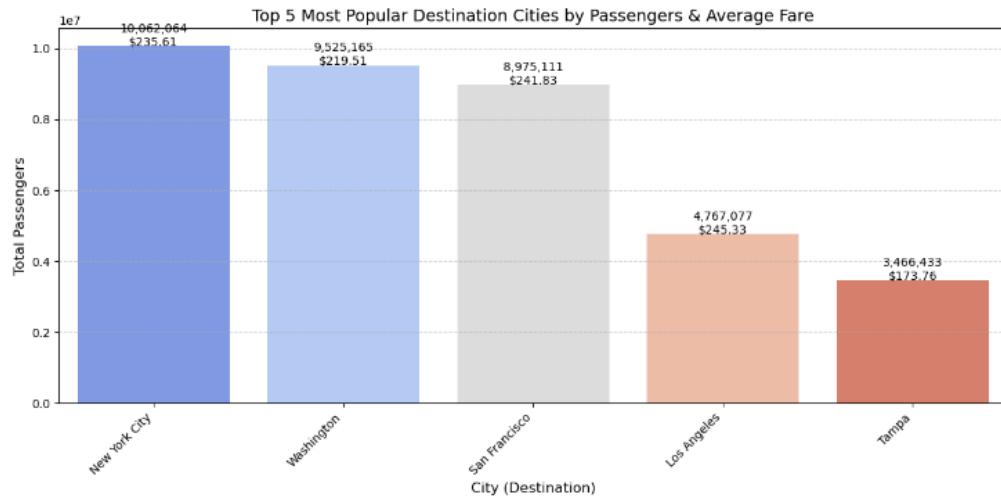
Analisa ini menggunakan *data grouping* dan *visualisasi*. Analisa yang pertama dilakukan adalah untuk melihat distribusi harga terhadap lima kota keberangkatan yang paling popular.



Gambar 12. Passanger and Fare distribution through City

Penemuan yang didapatkan dari analisa yang dilakukan adalah, untuk kota keberangkatan paling popular dengan rata-rata penumpang Sembilan juta tiga puluh delapan ribu tiga ratus tiga puluh lima, memiliki harga penerbangan yang paling rendah diantara empat kota lainnya. Sedangkan kota New York dengan kota ter-populer urutan keempat memiliki rata rata harga penerbangan paling tinggi.

Analisa selanjutnya yang dilakukan adalah distribusi harga terhadap kota tujuan popular. Analisa tersebut dilakukan dengan line of codes yang menggunakan data grouping dan visualisasi yang sama dengan analisa yang sebelumnya dilakukan.



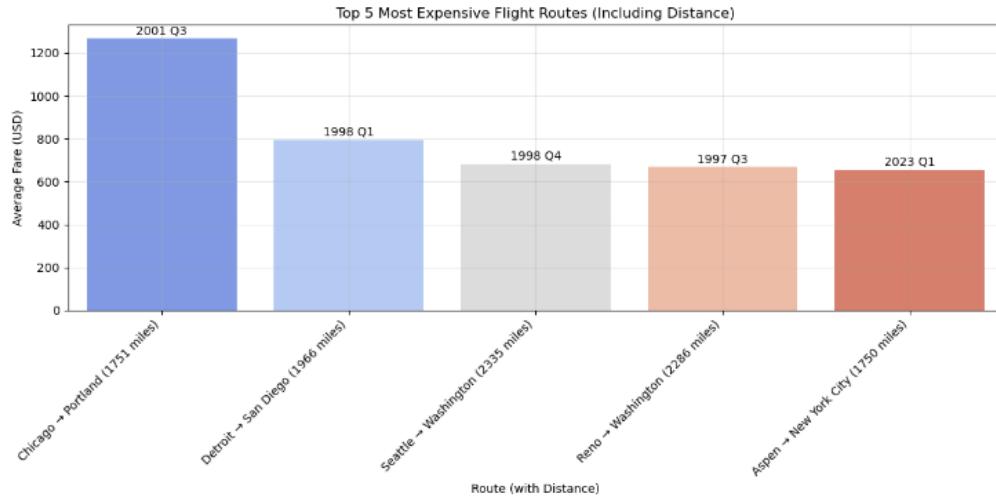
Gambar 13. Passanger and Fare distribution through City Destination

Penemuan yang ditemukan dari analisa yang dilakukan adalah, New York merupakan kota yang paling popular dan memiliki rata-rata harga penerbangan yang relatif tinggi, namun bukan merupakan harga yang paling tinggi dibanding kota Los Angeles dengan rata-rata harga \$245. Tampa merupakan kota yang paling tidak popular didalam konteks lima kota popular, dan memiliki rata-rata harga yang yang paling rendah dengan harga 173\$.

2. Analisis Jalur Penerbangan (By City)

Analisa ini bertujuan untuk melihat lima rute penerbangan termahal yang pernah dicatat didalam dataset yang digunakan, didalam analisa yang dilakukan juga dicantumkan jumlah jarak penerbangan rute tersebut, tahun dan kuartal penerbangan tersebut terjadi. Terdapat juga beberapa penemuan menarik yang didukung dengan berita-berita yang relevan disaat penerbangan tersebut terjadi.

Analisa ini dilakukan dengan data grouping dan visualisasi dengan menggunakan bar-plot. Penerbangan yang tercatat didalam rute penerbangan merupakan rata-rata hasil kumpulan penerbangan pada rute dan waktu penerbangan tersebut



Gambar 14. Top 5 Expensive Flight Routes

Penemuan yang ditemukan adalah, rute penerbangan dengan rata-rata harga tiket penerbangan termahal yang tercatat didalam dataset merupakan, rute penerbangan dengan keberangkatan kota Chicago dan tujuan kota Portland yang terjadi pada tahun 2001 pada kuartal tiga dengan jarak penerbangan 1751 miles. Hal yang menarik dari rute penerbangan ini adalah, rute penerbangan ini tidak memiliki jumlah jarak penerbangan yang paling tinggi, kota tujuan yang terbukti tidak masuk ke dalam beberapa kota yang menjadi kota tujuan yang mahal ataupun popular. Setelah melakukan analisa mendalam dengan melihat berita-berita yang relevan, terdapat sebuah kejadian yang mempengaruhi kota keberangkatan yaitu kota Chicago, kejadian tersebut merupakan peristiwa penyerangan terrorist pada 11 september 2001, waktu kejadian selaras dengan data yang diberikan, terjadi pada tahun 2001 kuartal 3, walaupun kejadian tersebut terjadi pada kota New York, namun kejadian tersebut membuat pemerintah dari kota Chicago memperketat bahkan menutup aktivitas penerbangan keluar maupun masuk kedalam kota Chicago. [15]

E. Analisis Kolom Fare dengan nsmiles

Analisa selanjutnya yang dilakukan adalah melihat beberapa jarak penerbangan dengan harga tertinggi, dan juga akan dilakukan analisis korelasi terhadap kolom nsmiles dengan fare. Yang pertama dilakukan adalah melakukan data grouping untuk kolom nsmiles dan fare, dan melihat 10 data terbesar.

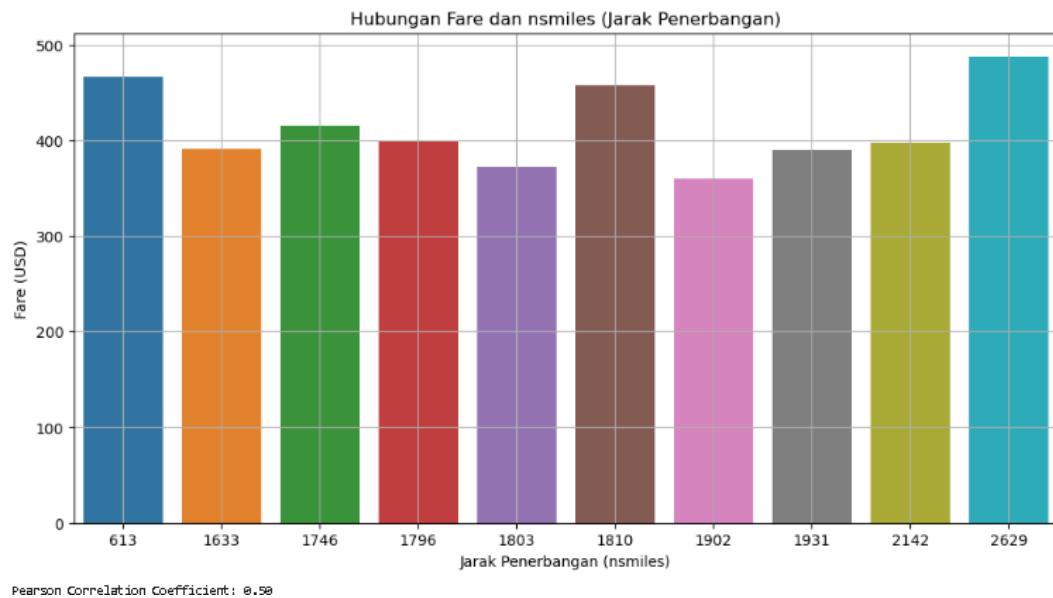
	nsmiles	fare
1150	2629	487.567500
287	613	466.480000
944	1810	457.124583
917	1746	415.280676
939	1796	398.735000
1049	2142	397.281250
883	1633	391.130000
982	1931	390.189000
941	1803	372.231333
970	1902	360.347963

Gambar 15. Top 10 Expensive Distances

Maka penemuan yang didapatkan adalah, dalam urutan pertama dipegang dengan data dengan jarak penerbangan 2629 mil dengan rata-rata harga penerbangan \$487.567, namun terdapat penemuan yang unik dimana penerbangan dengan jumlah mil 613

memiliki rata-rata harga tertinggi kedua dengan harga \$466,48. Dan penerbangan dengan mil > 1500 relatif memiliki rata-rata harga yang bisa dibilang diatas rata-rata.

Untuk mempermudah atau melanjutkan analisa yang sudah dilakukan maka dilakukan visualisasi untuk data grouping tersebut dan juga dilakukan analisis korelasi dengan *pearson correlation*.

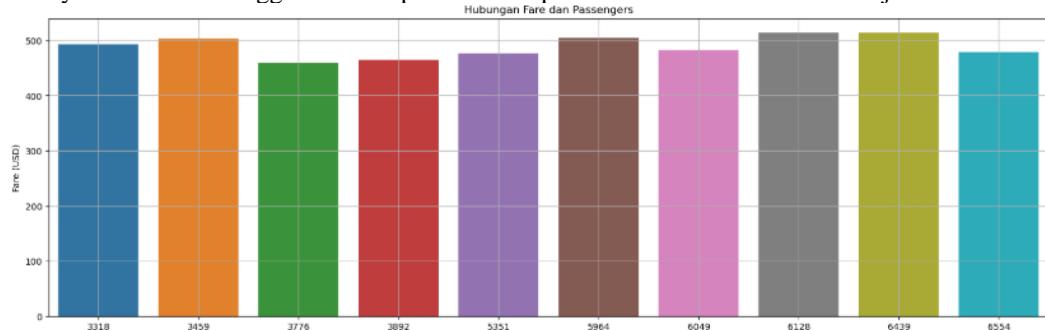


Gambar 16. *Distance and Fare Distribution*

Dilihat dari hasil visualisasi jarak yang lebih tinggi tidak selalu menentukan harga tiket yang tinggi, namun sudah dipastikan jika jarak penerbangan diatas 1500 akan memiliki harga penerbangan yang tinggi. Didapatkan juga hasil dari *Pearson Correlation Coefficient* yaitu : 0.50, dengan hasil ini bisa disimpulkan bahwa harga tiket penerbangan tidak terlalu dipengaruhi jarak penerbangan, namun menjadi salah satu dari banyak alasan untuk harga tiket penerbangan.

F. Analisis Kolom Fare dengan Passengers

Setelah membahas hasil dari analisis fare dengan city1 & city2, yang selanjutnya dilakukan adalah analisis fare dengan Passengers, analisis yang dilakukan ini bertujuan untuk melihat distribusi harga tiket penerbangan terhadap jumlah penumpang penerbangan tersebut, dan untuk melihat tren atau pola harga yang muncul jika dibandingkan dengan jumlah penumpang penerbangan.. untuk melihat hal tersebut, yang selanjutnya dilakukan adalah data grouping kolom passengers dengan kolom fare dan dilakukannya visualisasi menggunakan barplot untuk dapat melakukan analisis lebih lanjut



Gambar 17. *Fare and Passanger Distribution*

Dari hasil yang didapatkan, bisa disimpulkan jumlah penumpang yang tinggi tidak mempengaruhi, hal itu dikarenakan tren yang tidak stabil, jumlah passenger tertinggi tidak memiliki harga yang paling tinggi, hal ini juga dibuktikan dengan hasil dari *Pearson Correlation Coefficient* yaitu -0,17, yang menandakan bahwa tingkat korelasi antara jumlah penumpang dan harga tiket penerbangan tidak tinggi.

G. Analisis Kolom Fare dengan Airport1/Airport2

Analisis ini berfokus kepada kolom Airport1(Airport keberangkatan) dan Airport2(Airport tujuan), isi data dari kolom tersebut merupakan kode IATA yang sudah digunakan untuk menjadi kode pengenal untuk kepentingan penerbangan. Didalam dataset tidak hanya terdapat airport commercial/public, namun terdapat juga beberapa airport yang digunakan untuk kepentingan pribadi, seperti Heliport, dan Amphibious Airport. Kedua airport tersebut memiliki rata-rata harga penerbangan yang tinggi, dikarenakan hasil penemuan tersebut, untuk analisa airport selanjutnya akan dilakukan dengan dataframe yang sudah di-filter dan sebelum

di-filter, hal ini dilakukan karena data-data dengan airport tersebut menyebabkan ketidaksesuaian data atau menyebabkan data outlier.

Beberapa Airport yang menyebabkan data outlier merupakan : TSS (East 34th Street Airport) yang merupakan Heliport yang berada di kota New York, JRB(Downtown Manhattan Heliport), yang merupakan Heliport yang berada di kota Manhattan, dan MPB(Miami Seaplane Base) yang digunakan untuk pesawat amphibious.

1. Filterisasi *Dataframe*

Seperti yang sudah dijelaskan pada bagian sebelumnya, terdapat beberapa airport yang tidak setara dengan airport-airport commercial/public, airport-airport tersebut bisa berupa heliport yaitu landasan yang khusus untuk kegiatan penerbangan helikopter, dan terdapat juga seaplane base yang merupakan landasan amphibious dimana pesawat bisa lepas-landas di air dan bukan di daratan. Airport tersebut lebih sering dipakai untuk kepentingan private yang membuatnya mempunyai rata-rata harga yang tinggi, dengan hal tersebut maka menimbulkan kesenjangan data yang tidak konsisten.

Untuk melakukan filterisasi data tersebut maka harus dilakukannya pencarian airport-airport tersebut, karena isi data kolom airport1-airport2 merupakan kode IATA, tidak terdapatnya kode unik yang menandakan airport tersebut berbeda dari airport pada umumnya. Setelah dilakukan pencarian intensif di dalam dataset terdapat 3 airport yang menimbulkan kesenjangan data. Berikut merupakan line of codes yang digunakan untuk melakukan filterisasi data.

```
excluded_airports = ["TSS", "JRB", "MPB"]  
  
df_filtered=df[~df['airport_1'].isin(excluded_airports)&~df['airport_2'].isin(excluded_airports)]
```

Kode 1Filterisasi Dataset

TSS sendiri merupakan kode airport untuk East 34th Street Heliport in New York City yang merupakan landasan transportasi udara yang diutamakan untuk helikopter, dan private jets [12] yang terletak di kota New York, jenis penerbangan yang tersedia disini hanyalah untuk kegiatan private atau private flight. Hal tersebut membuat harga tiket yang ditawarkan mahal atau diatas rata-rata harga yang ditawarkan untuk penerbangan komersial.

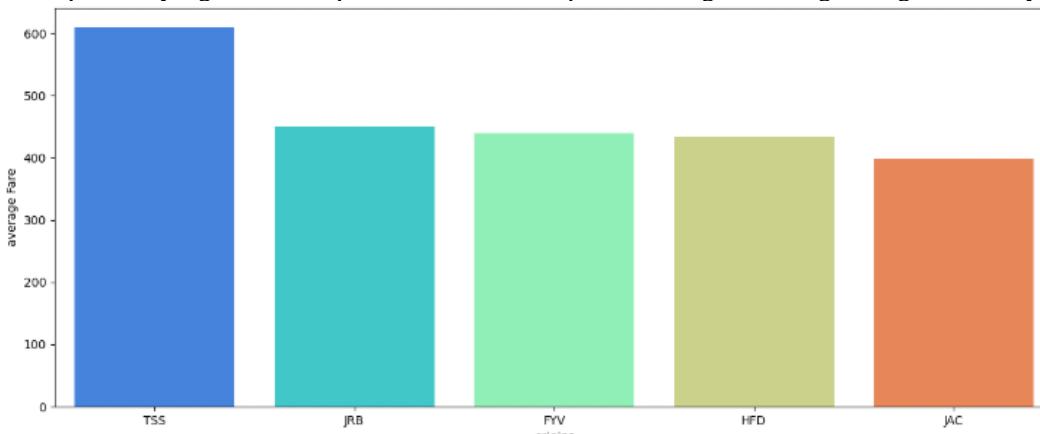
JRB atau Downtown Manhattan Heliport merupakan heliport yang terletak di kota New York. Sama halnya dengan TSS Airport, JRB merupakan heliport yang membuat rata-rata harga tiket penerbangannya diatas rata-rata.

Airport selanjutnya yang masuk kedalam list airport yang menimbulkan kesenjangan data ialah MPB Airport atau Miami Seaplane Base yang merupakan Amphibious airport atau airport yang digunakan dengan pesawat yang bisa mendarat dan lepas landas diatas permukaan air. Dilansir dari website miamiseaplane.com, airport ini juga menjadi salah satu tujuan kunjungan wisata di kota miami [13]. Walaupun disebut sebagai airport umum, namun pesawat yang digunakan ini berbeda dengan pesawat komersial lainnya, dengan kapasitas yang lebih sedikit, hal-hal tersebut membuat penerbangan dengan airport ini memiliki harga yang tinggi.

2. Analisis Distribusi Harga *Airport1/Airport2*

Analisis yang pertama dilakukan adalah untuk melihat 5 airport dengan rata-rata harga penerbangan paling mahal, analisis dilakukan dengan menggunakan dataframe yang sudah di filter dan yang tidak dilakukan filter, dengan dilakukannya hal tersebut, terdapat 2 hasil analisis yang berbeda dan tentunya mendapatkan penemuan yang akan dikuatkan dengan artikel-artikel yang relevan.

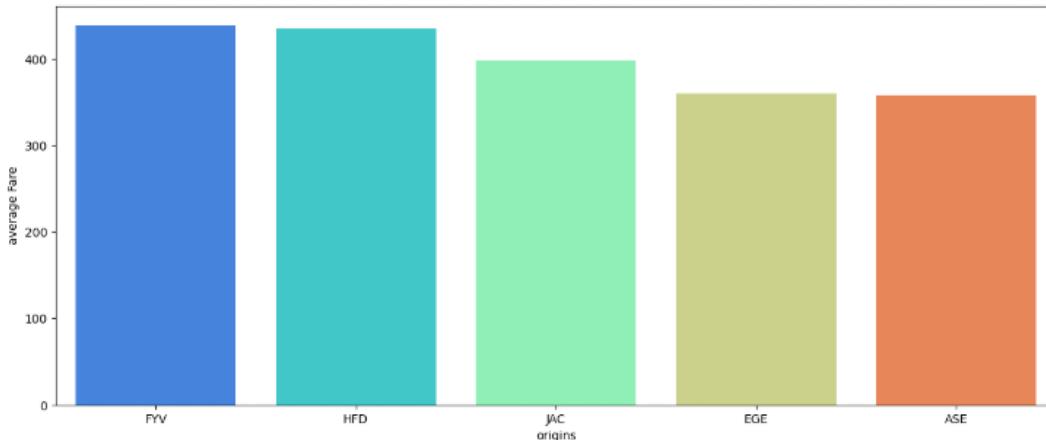
Hasil analisis pertama yang akan ditampilkan adalah lima airport keberangkatan dengan harga rata-rata yang tinggi



Gambar 18. Top 5 Expensive Origins Airport (Unfiltered)

Penemuan yang ditemukan adalah, 2 airport dengan rata-rata harga tertinggi yaitu TSS dan JRB merupakan heliport yang menyebabkan harga rata-rata dari kedua airport tersebut relatif lebih tinggi dibandingkan dengan airport public/commercial lainnya.

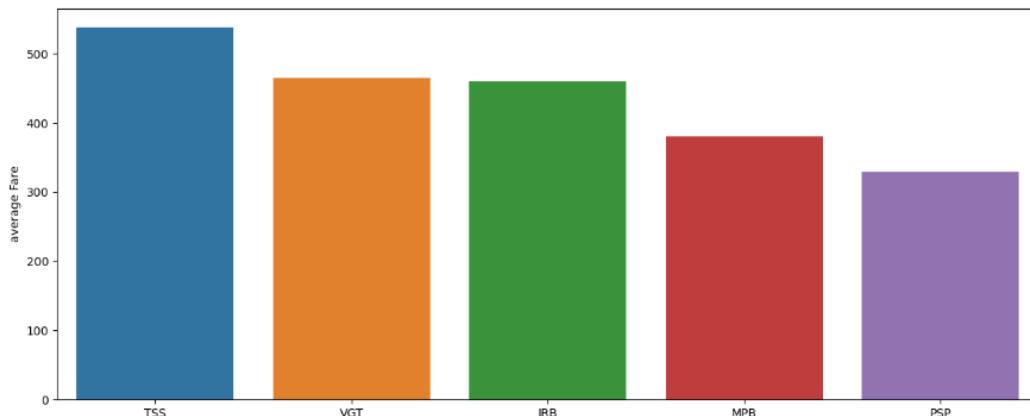
Selanjutnya merupakan visualisasi untuk menunjukkan lima airport dengan rata-rata harga yang tinggi, analisis ini dilakukan dengan dataframe yang sudah difilter sebelumnya pada sub bab Filterisasi Dataframe.



Gambar 19. Top 5 Expensive Origins Airport (Filtered)

Penemuan yang didapatkan adalah, airport dengan rata-rata harga penerbangan yang paling tinggi adalah airport FYV(Drake Field Airport) yang terletak di Arkansas City, dengan rata rata harga \$400-\$500, hal ini mungkin dikarenakan karena Drake Field Airport sendiri merupakan public executive Airport yang menyediakan penerbangan public dan juga private executive sehingga memungkinkan untuk beberapa penerbangannya memiliki harga yang relatif tinggi..

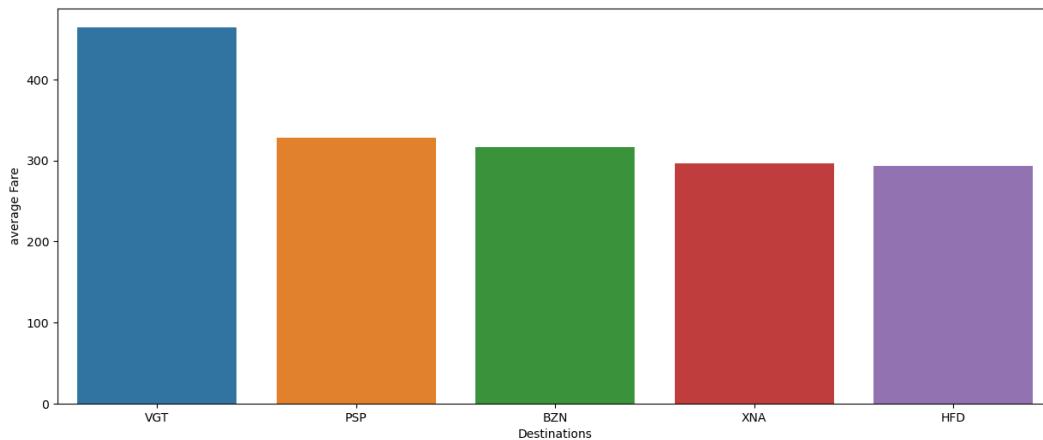
Setelah dilakukannya analisis untuk airport keberangkatan, maka yang selanjutnya dilakukan adalah analisis untuk airport tujuan. Seperti yang sudah dilakukan pada analisis sebelumnya, akan terdapat dua hasil analisis, yaitu setelah dan sebelum dilakukannya filterisasi data. Berikut merupakan line of codes yang digunakan untuk melakukan analisis distribusi harga untuk lima airport tujuan termahal:



Gambar 20. Top 5 Destination Airport (Unfiltered)

Penemuan yang didapatkan adalah, Kembali terdapatnya tiga airport yang bukan merupakan airport commercial/public, kedua airport tersebut merupakan TSS dan JRB yang merupakan sebuah Heliport, dan MPB yang digunakan untuk kegiatan lepas landas seaplane atau pesawat amfibi. Dikarenakan ketiga airport tersebut memiliki mayoritas penerbangan privat atau untuk kepentingan privat, yang menyebabkan harga tiket penerbangan tersebut relatif lebih mahal dibandingkan dengan airport public pada umumnya.

Setelah penemuan yang didapatkan, yang selanjutnya dilakukan adalah untuk melakukan visualisasi yang sama namun dengan menggunakan dataframe yang sudah difilter, agar hasil yang didapatkan hanya untuk airport commercial/public. Untuk melakukan visualisasi tersebut.



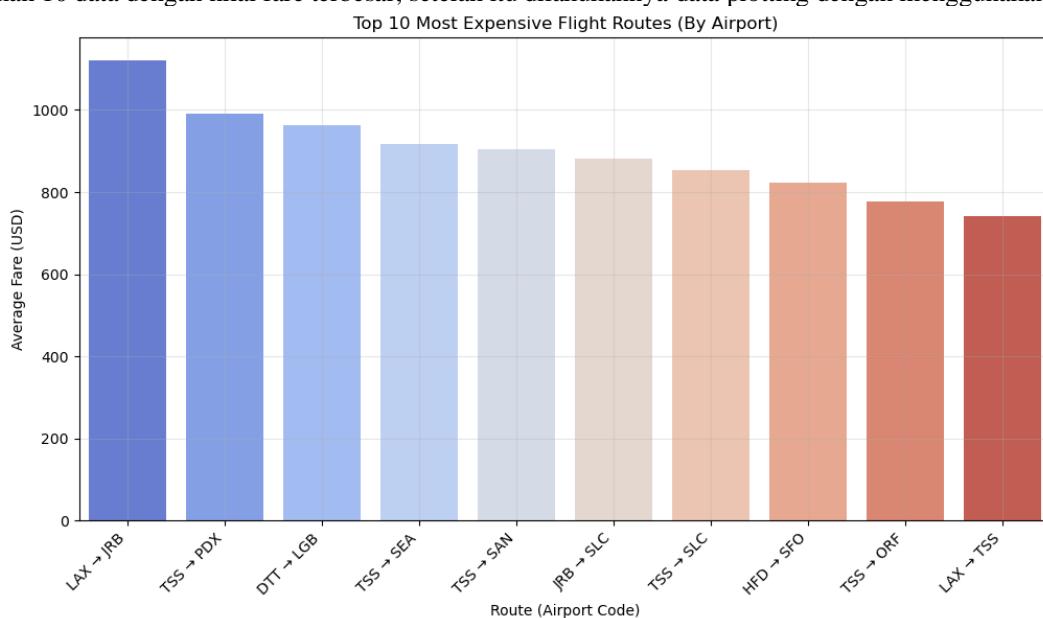
Gambar 21. Top 5 Destination Airport (Filtered)

Penemuan yang didapatkan dari analisis yang dilakukan adalah, terdapat satu airport tujuan yang menonjol yaitu VGT yang memiliki rata-rata harga penerbangan yang relatif tinggi dengan kisaran harga \$400-\$500. VGT sendiri merupakan airport public yang terletak di kota North Las Vegas, yang seperti kita ketahui merupakan salah satu kota yang terkenal untuk pusat bisnis dan rekreasi. Kota Las Vegas juga sering digunakan sebagai tempat untuk event-event special seperti konser music, Super Bowl, F1 Racing Event, etc. Penerbangan dengan menggunakan airport VGT memiliki harga tiket penerbangan yang mahal dikarenakan beberapa faktor, yaitu : lokasi airport yang dekat dengan pusat kota sehingga membuatnya strategis, Charter flight atau penerbangan charter dimana airport VGT bisa dipakai untuk kepentingan penerbangan dengan private jet yang membuatnya memiliki harga tiket yang tinggi, dan kebijakan airport untuk pajak yang dikenakan dalam harga tiket penerbangan disaat terdapatnya special event, pajak tersebut ialah special event fees, yang tentunya meningkatkan harga tiket penerbangan. [14]

3. Analisis Rute Penerbangan (Airport)

Analisis rute penerbangan ini bertujuan untuk melihat beberapa rute-rute penerbangan dengan menggunakan kolom airport beserta harga tiket penerbangan yang relatif tinggi dan akan dilakukan dengan dataframe yang sudah dan belum dilakukan filterisasi data kolom airport. Tujuan yang diharapkan dari analisis ini adalah untuk mendapatkan 10 rute penerbangan dengan rata-rata harga penerbangan yang tinggi setelah dan sebelum dilakukannya filterisasi data.

Yang pertama dilakukan adalah untuk meng-analisa rute penerbangan sebelum dilakukannya filterisasi data, untuk melakukannya perlu dilakukan data grouping untuk kolom airport1 & airport2 dengan kolom fare dan hanya menampilkan 10 data dengan nilai fare terbesar, setelah itu dilakukannya data plotting dengan menggunakan barplot.



Gambar 22. Top 10 Routes Unfiltered

Penemuan yang didapatkan adalah, mayoritas penerbangan dengan rata-rata harga penerbangan tertinggi didominasi dengan *airport* keberangkatan/penerbangan yang merupakan *heliport*, hal ini sudah dibuktikan sebelumnya untuk rata-rata harga penerbangan.

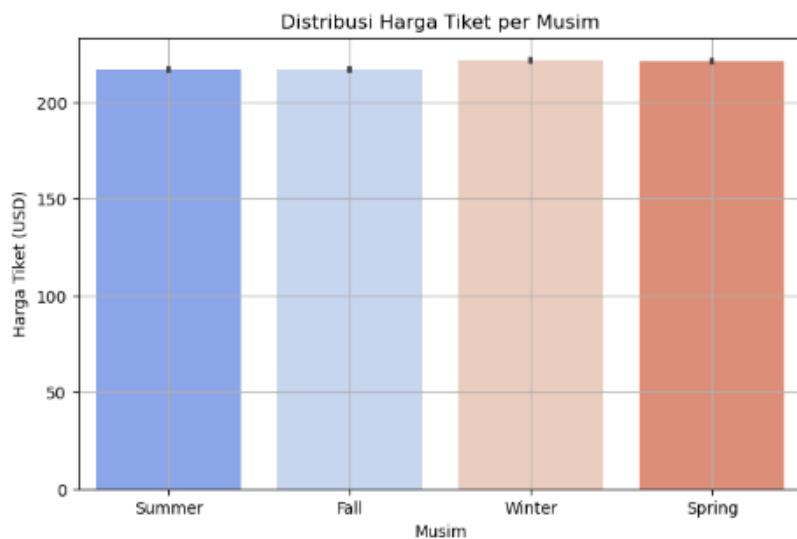
H. Feature Engineering

Feature Engineering merupakan proses mengubah data mentah menjadi fitur yang relevan dan bermakna, sehingga hasilnya lebih mudah untuk melakukan analisis yang lebih dalam untuk analisis yang sudah dilakukan sebelumnya dan memperkuat hasil analisis yang sudah dibuat sebelumnya. Terdapat beberapa *Feature Engineering* yang dilakukan, yang menjadi proses yang memperkuat hasil analisis data yang sudah didapatkan sebelumnya

1. *Feature Engineering*: Distribusi Harga Tiket Per Musim

Feature Engineering ini bertujuan untuk melihat distribusi harga tiket penerbangan dengan mapping data untuk setiap kuartilnya. Kuartil pertama memiliki musim winter, musim dingin sendiri berjalan dimulai dari tanggal 20 Desember hingga 20 Maret. Kuartil kedua memiliki musim Spring atau musim gugur yang berjalan dimulai dari tanggal 21 Maret hingga 20 Juni. Kuartil ketiga akan di-mapping menjadi musim summer atau musim panas yang berjalan pada rentang waktu bulan Juni hingga September, dan untuk kuartil terakhir akan dijadikan musim gugur atau fall. *Feature Engineering* ini juga bertujuan untuk memperkuat hasil analisis yang sudah dilakukan pada sub-bab *Analisis Fare dan Year & Quarter*.

Untuk melakukan *Feature Engineering* ini, yang pertama dilakukan adalah membuat feature baru dengan judul *Seasons*, setelah itu dilakukan data mapping terhadap kolom *Quarter*, hasil dari data mapping tersebut akan dimasukan kedalam feature *seasons*. Setelah itu dilakukannya plotting dengan menggunakan barplot agar bisa memudahkan proses analisa *feature engineering* yang dilakukan.



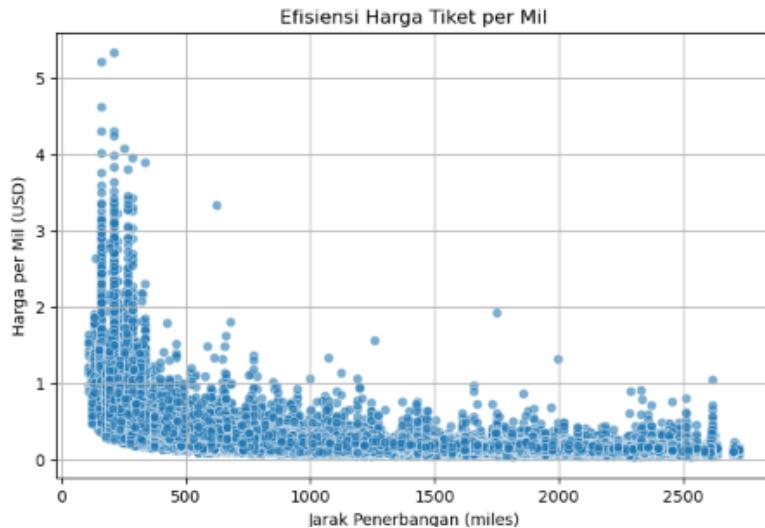
Gambar 23. *Feature Engineering: Quartil Mapping*

Penemuan yang didapatkan adalah musim *winter* (kuartal 1) dan musim *spring* (kuartal 2) merupakan musim yang memiliki distribusi rata-rata harga tiket yang lebih tinggi dibandingkan dengan musim-musim lainnya, namun setiap tahunnya atau seiring berjalannya waktu, season atau musim di Amerika akan mengalami *cycle* yang disebabkan perubahan iklim, perubahan orbital bumi. Namun perubahan atau *cycle* tersebut tidak menyebabkan waktu yang signifikan

2. *Feature Engineering*: Efisiensi Harga Tiket Per Mil

Feature Engineering selanjutnya adalah untuk melihat bagaimana efisiensi harga tiket penerbangan per-mil, hasil atau penemuan yang diharapkan adalah bagaimana pembagian harga tiket penerbangan/mil atau jarak penerbangan tersebut, proses ini juga merupakan analisa tambahan atau lanjutan dari sub-bab *Fare and nsmiles*, dimana hasil yang sebelumnya didapatkan adalah jarak penerbangan tidak terlalu mempengaruhi harga tiket penerbangan yang terlihat dari ketidakstabilan hasil dari visualisasi yang dilakukan.

Untuk melakukan *Feature Engineering* ini yang pertama dilakukan adalah membuat sebuah feature baru yang akan menampung data dari hasil pembagian *fare/nsmiles*, setelah itu hasil data yang terdapat di feature akan dilakukannya visualisasi dengan menggunakan scatterplot.



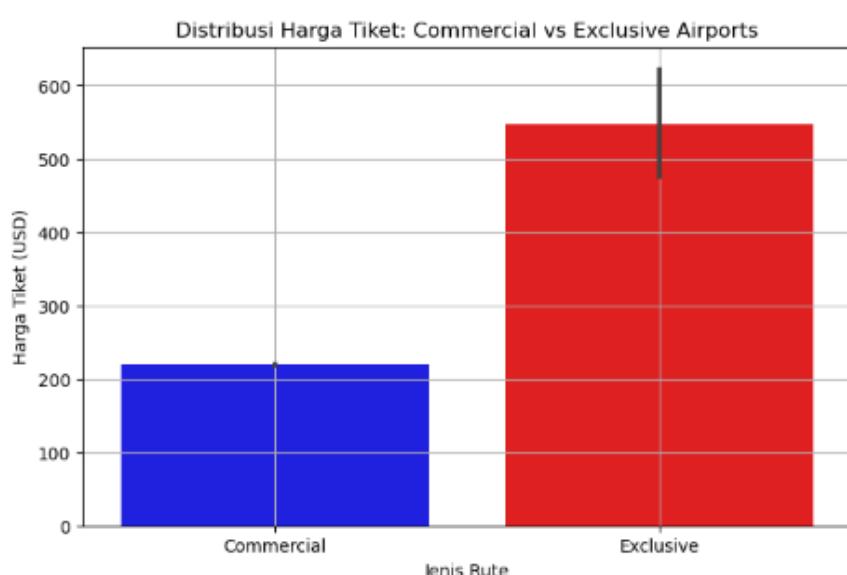
Gambar 24. Feature Engineering: Fare/Miles

Penemuan yang didapatkan adalah, efisiensi harga tiket penerbangan akan semakin meningkat dengan jarak yang lebih jauh, penerbangan dengan jarak penerbangan yang lebih jauh memiliki harga tiket penerbangan/mil lebih murah dibandingkan dengan jarak yang pendek, hal ini tentunya dikarenakan atau disebabkan oleh kebijakan maskapai penerbangan. Hasil dari visualisasi menunjukkan bahwa jarak penerbangan yang lebih jauh akan memiliki harga yang lebih efisien dibandingkan penerbangan yang memiliki jarak yang lebih pendek

3. Feature Engineering: Distribusi Harga Tiket *Airport Commercial* dan *Exclusive*

Feature Engineering selanjutnya merupakan faktor penguatan untuk tujuan dilakukannya filterisasi data yang dilakukan pada sub-bab Filterisasi Dataframe untuk beberapa analisa yang dilakukan dengan kolom airport seperti pada sub-bab Analisis Distribusi Harga *Airport1/Airport2* dan pada sub-bab Analisis Rute Penerbangan (*By Airport*). *Feature Engineering* tersebut juga dilakukan untuk melihat perbandingan distribusi harga tiket penerbangan yang didalam rute penerbangan tersebut menggunakan *airport exclusive* untuk keberangkatan ataupun tujuan dengan harga tiket penerbangan dengan rute penerbangan yang hanya menggunakan *airport commercial/public*.

Untuk melakukan feature engineering ini yang pertama dilakukan membuat variabel yang menampung sebuah list untuk kode *airport exclusive*, setelah itu dilakukan pembuatan feature baru *exclusive_airport1* dan *exclusive_airport2* yang didalamnya menggunakan lambda untuk melakukan pengecekan data *airport*, jika terdapat data *airport* yang terdapat didalam list *exclusive_airports*, maka data akan dimasukan ke dalam feature tersebut, dan selanjutnya kedua feature tersebut akan digabungkan ke dalam *feature exclusive route*. Visualisasi juga akan dilakukan menggunakan barplot agar mempermudah jalannya analisis.



Gambar 25. Feature Engineering: Commercial vs Exclusive Airport

Penemuan yang didapatkan adalah, rute penerbangan yang didalamnya terdapat menggunakan airport exclusive memiliki harga tiket penerbangan yang jauh lebih mahal dibandingkan dengan yang hanya menggunakan airport commercial/public, dengan perbedaan yang signifikan, dimana rata-rata harga tiket penerbangan dengan rute exclusive lebih tinggi 2 kali lipat dibandingkan rute penerbangan *commercial/public*.

V. KESIMPULAN

Penelitian ini telah selesai dan berhasil untuk melakukan eksplorasi dataset untuk menemukan beberapa faktor yang mempengaruhi harga tiket penerbangan domestik Amerika Serikat dengan menggunakan metode EDA (Exploratory Data Analysis). Berdasarkan beberapa hasil yang didapatkan dari analisis data yang dilakukan, terdapat beberapa kesimpulan yang telah dirangkum sebagai berikut. Tidak ditemukan variable yang paling menonjol, namun didapatkannya beberapa variable seperti pemilihan waktu penerbangan, *events/kejadian* yang mendampaki suatu kota, negara, satu dunia. Pemilihan airport, jenis airport dan kota juga menjadi kunci penting dalam menentukan harga tiket penerbangan. Jumlah jarak penerbangan dan jumlah penumpang tidak menjadi kunci penting. Didapatkan juga semakin jauh jarak suatu penerbangan, maka harga tiket penerbangan tersebut semakin efisien.

DAFTAR PUSTAKA

- [1] B. Jikadara, "US Airline Flight Routes and Fares 1993-2024," Kaggle, 10 July 2024. [Online]. Available: <https://www.kaggle.com/datasets/bhavikjikadara/us-airline-flight-routes-and-fares-1993-2024/code>. [Accessed 1 10 2024].
- [2] H. Nassaji, "Qualitative and descriptive research: Data type versus data analysis," *Language teaching research*, vol. 19, pp. 129-132, 2015.
- [3] B. Volodymyr, G. A. A. P. and C. A, "Airline pricing under different market conditions: Evidence from European Low-Cost Carriers," *Tourism Management*, pp. 152-163, 2015.
- [4] B. Holzhauer, "Airline ticket prices are up 25%, outpacing inflation — here are the ways you can still save," CNBC, 22 11 2024. [Online]. Available: <https://www.cnbc.com/select/airline-ticket-prices-are-up-25-percent-why-and-how-to-save/>. [Accessed 14 March 2025].
- [5] FRED, "Customer Price Index of All Urban Customers : Airline Fares in U.S City," Federal Reserve Bank of St. Louis, [Online]. Available: <https://fred.stlouisfed.org/series/CUSR0000SETG01>. [Accessed 14 March 2025].
- [6] A. Hayes, "Dotcom Bubble Definition," Investopedia, 31 May 2024. [Online]. Available: <https://www.investopedia.com/terms/d/dotcom-bubble.asp>. [Accessed 14 March 2025].
- [7] N. H. a. H. Command, "The 9/11 Terrorist Attacks," Naval History and Heritage Command, 07 September 2023. [Online]. Available: <https://www.history.navy.mil/browse-by-topic/wars-conflicts-and-operations/sept-11-attack.html>. [Accessed 14 March 2025].
- [8] Z. Peregrine, "Understanding How COVID-19 Has Influenced Flight Ticket Prices and What Travelers Need to Know," Wowfare.com, 12 December 2024. [Online]. Available: <https://wowfare.com/blog/understanding-how-covid-19-has-influenced-flight-ticket-prices-and-what-travelers-need-to-know/>. [Accessed 14 March 2025].
- [9] D. E. Clark, J. M. McGibany and A. Myers, "The Effects of 9/11 on the Airline Travel Industry," Springer, [Online]. Available: https://link.springer.com/chapter/10.1057/9780230100060_7. [Accessed 10 June 2025].
- [10] Tripadvisor, "Flying into Jackson Hole...why is it so expensive!," Tripadvisor, 11 September 2021. [Online]. Available: https://www.tripadvisor.com>ShowTopic-g60999-i481-k13542093-Flying_into_Jackson_Hole_why_is_it_so_expensive-Yellowstone_National_Park_Wyoming.html. [Accessed 18 May 2025].
- [11] M. Humphries, "15 over-the-top examples of wealth and luxury I saw on my first trip to Aspen, Colorado," Business Insider, 29 January 2025. [Online]. Available: <https://www.businessinsider.com/signs-of-aspen-colorado-luxury-homes-resorts-2022-2?type=social>. [Accessed 19 May 2025].
- [12] AirNav, "East 34th Street Heliport," AirNav, [Online]. Available: <https://airnav.com/airport/6N5>. [Accessed 10 June 2025].
- [13] MiamiSeaplane, "Elevate your Journey," MiamiSeaplane, [Online]. Available: <https://miamiseaplane.com/>. [Accessed 10 June 2025].
- [14] Gel and Lillian, "AOPA fights price gouging at Las Vegas valley airports," AOPA, Las Vegas, 2023.
- [15] A. Martin, "This week in history: Terror hits home," Chicago SunTimes, Chicago, 2020.
- [16] J. W. Creswell and J. D. Creswell, Research design: Qualitative, quantitative, and mixed methods approaches, Sage publications, 2017.
- [17] P. Bruce, B. Andrew and P. Gedeck, Practical statistics for data scientists: 50+ essential concepts using R and Python, O'Reilly Media, 2020.
- [18] R. M. Groves and Peytcheva, Survey Methodology: Key Concepts and Current Challenges. In C. Wolf, D. Joye, T. W. Smith, & Y. Fu (Eds.), SAGE Publications, 2016.
- [19] W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, O'Reilly Media, 2017.
- [20] J. Cabera and A. McDougall, Statistical Consulting, Springer, 2017.
- [21] M. a. Z. Liu, H. a. Xu and K. Zeshui and Ding, "The Fusion of Fuzzy Theories and Natural Language Processing: A State-of-the-Art Survey," *Applied Soft Computing*, p. 111818, 2024.
- [22] J. VanderPlas, Python Data Science Handbook: Essential Tools for Working with Data, O'Reilly Media., 2016.
- [23] Data, MIT Critical and Komorowski, Matthieu and Marshall, Dominic C and Saliccioli, Justin D and Crutain and Yves, "Exploratory data analysis," *Secondary analysis of electronic health records*, pp. 185--203, 2016.
- [24] R. F. Suryawan and M. Fatchoelqorib, "Penerbangan perintis dalam mengembangkan perekonomian di pulau karimunjawa," *Jurnal Manajemen Transportasi & Logistik (JMTRANSLOG)*, vol. 5, pp. 1-2, 2018.
- [25] IATA, "Airline and Airport Code Search," IATA, [Online]. Available: <https://www.iata.org/en/publications/directories/code-search/>. [Accessed 3 March 2025].