

Tugas UAS Presentasi dan Visualisasi Data

Flight Ticket Prediction



Oleh

Chandra Viriya H. C14210075

Alfred Wisana C14210177

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS KRISTEN PETRA
SURABAYA

2024

Latar Belakang

Masa-masa liburan merupakan masa yang paling diminati oleh banyak orang. Namun, liburan sendiri memiliki beban finansial yang cukup tinggi. Salah satu faktor terbesar yang berkontribusi dalam beban tersebut adalah biaya penerbangan yang terkenal dengan sifatnya yang fluktuatif. Berdasarkan artikel dari AlternativesAirlines.com, industri penerbangan sangat bergantung pada supply and demand dan untuk memaksimalkan keuntungan maskapai penerbangan akan menyesuaikan harga tiket mereka sesuai dengan demand yang ada. Selain itu berdasarkan data prosedur penjualan tiket Garuda Indonesia pada tahun 2018, faktor yang mempengaruhi harga adalah jarak, waktu keberangkatan, destinasi, dan beberapa faktor eksternal lainnya.

Ketidakpastian harga ini seringkali membuat banyak konsumen untuk memikirkan kembali rencana berlibur mereka dan beralih kepada mode transportasi lainnya. Dengan demikian diperlukan pedoman dalam memprediksi biaya penerbangan agar konsumen dapat memilih keputusan yang paling tepat.

Tujuan

Tujuan utama dari analisis ini adalah memberikan kontribusi kepada masyarakat dalam mencari dan memahami tren yang menjadi kontributor utama dalam fluktuasi harga penerbangan. Masyarakat jadi mengetahui faktor apa saja berpengaruh dalam perubahan harga tiket pesawat. Dengan Mengetahui faktor yang mempengaruhi perubahan harga, masyarakat dapat melakukan antisipasi lonjakan harga yang tak terduga.

Melalui analisis ini masyarakat diharapkan dapat mengambil keputusan yang lebih terarah dalam mencari dan memprediksi harga penerbangan. Dengan demikian masyarakat dapat lebih matang dalam persiapannya sebelum melakukan penerbangan dan bisa mendapatkan tiket di harga yang optimal sehingga bisa meningkatkan kualitas dan frekuensi penerbangan mereka.

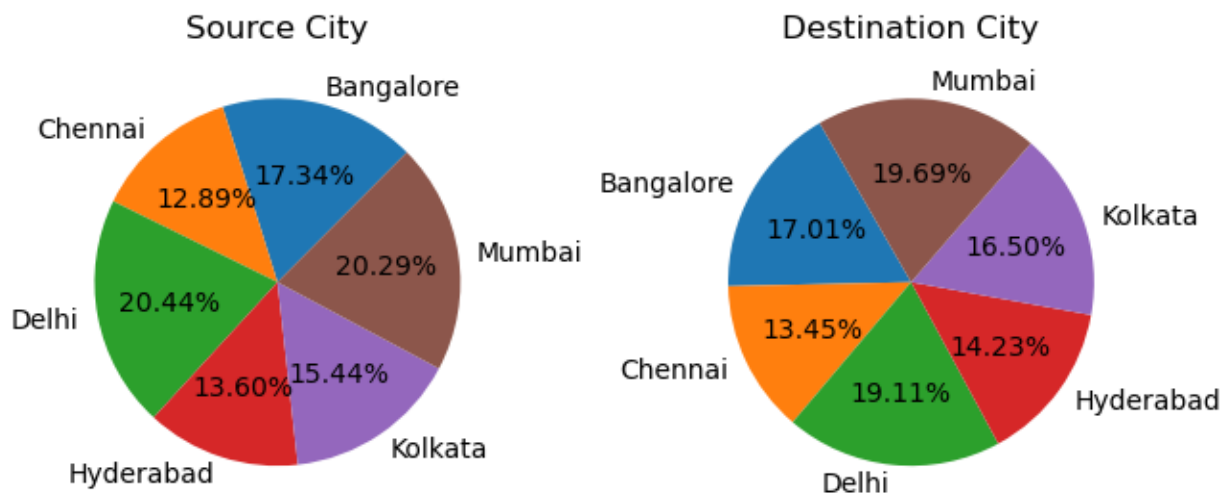
Analisa Data

Dalam analisis ini dataset yang digunakan adalah data penerbangan yang diambil dari situs kaggle. Dataset yang digunakan terdiri dari 30.000 data yang diambil dari data penerbangan domestik di India dan memiliki 11 fitur yaitu *airline*, *flight*, *source city*, *departure time*, *stops*, *arrival time*, *destination city*, *class*, *duration*, *days left*, dan *price*.

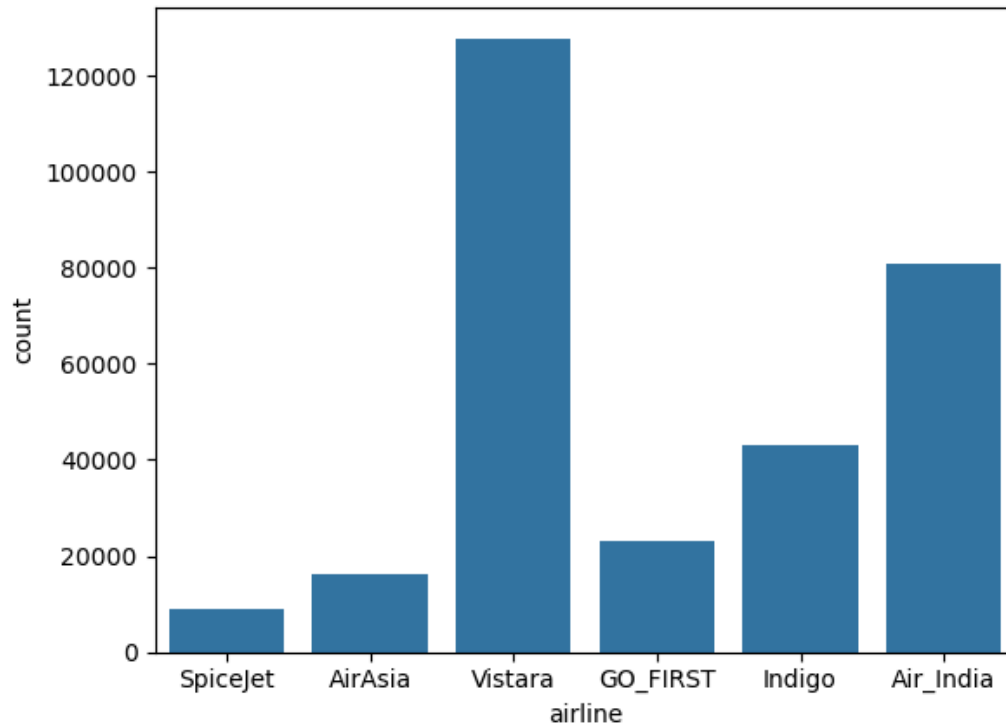
```
RangeIndex: 300153 entries, 0 to 300152
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   Unnamed: 0            300153 non-null  int64  
 1   airline               300153 non-null  object  
 2   flight               300153 non-null  object  
 3   source_city          300153 non-null  object  
 4   departure_time       300153 non-null  object  
 5   stops               300153 non-null  object  
 6   arrival_time        300153 non-null  object  
 7   destination_city     300153 non-null  object  
 8   class               300153 non-null  object  
 9   duration             300153 non-null  float64 
10   days_left           300153 non-null  int64  
11   price               300153 non-null  int64  
```

Field	Description
<i>Airline</i>	Airline
<i>Flight</i>	Unique flight id
<i>Source_city</i>	Departure city
<i>Departure_time</i>	Departure time
<i>Stops</i>	Number of stops along the way
<i>Arrival_time</i>	Arrival time
<i>Destination_city</i>	Flight destination
<i>Class</i>	Seating class
<i>Duration</i>	Overall amount of time it takes to travel between cities in hours
<i>Days_left</i>	The number of days between the booking date and the trip date
<i>Price</i>	Price

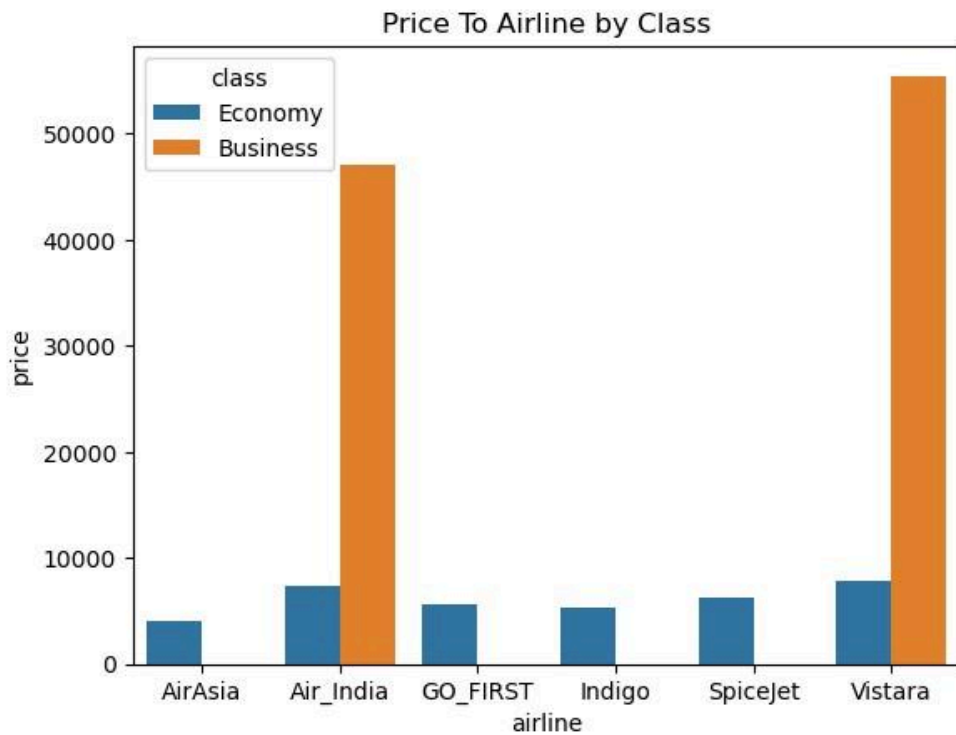
Dari fitur yang ada 1 fitur yang tidak mempengaruhi terhadap penentuan harga yaitu *flight* jadi kita akan drop fitur tersebut.



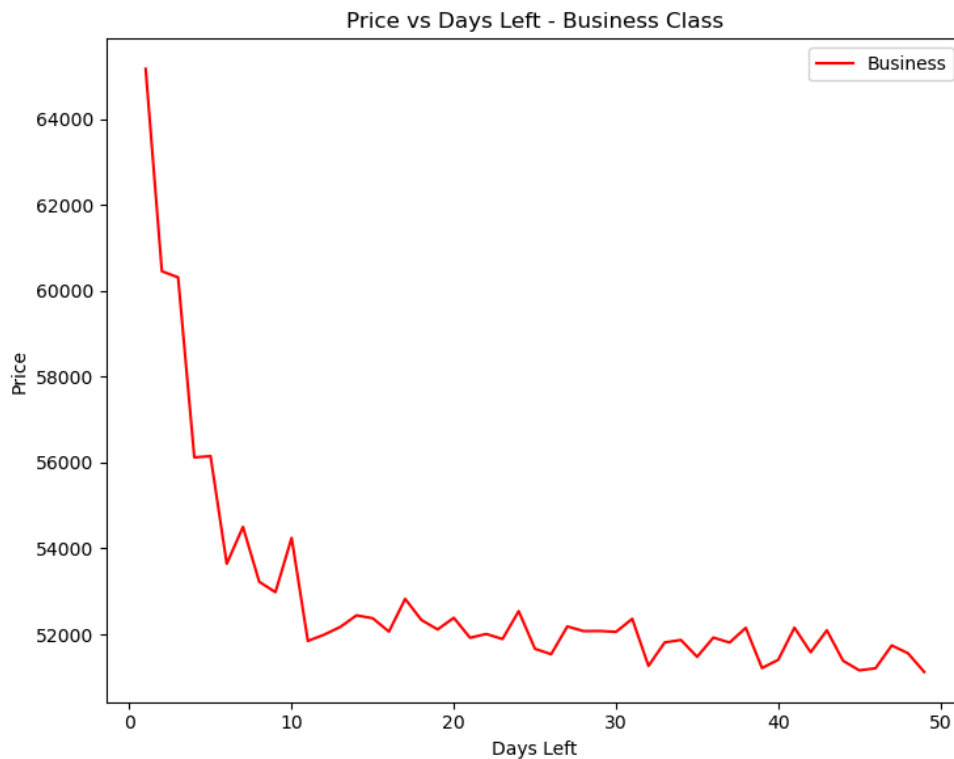
Dari visualisasi diatas dapat dilihat distribusi kota pada dataset cukup merata menandakan bahwa adanya jumlah minat yang cukup merata untuk setiap destinasi dan dari setiap kota asal.



Dari visualisasi diatas dapat dilihat bahwa maskapai dengan penerbangan terbanyak adalah Vistara lalu diikuti oleh Air India, Indigo, Go First, AirAsia, dan SpiceJet. Hal ini menunjukan bahwa Vistara merupakan maskapai yang paling populer.

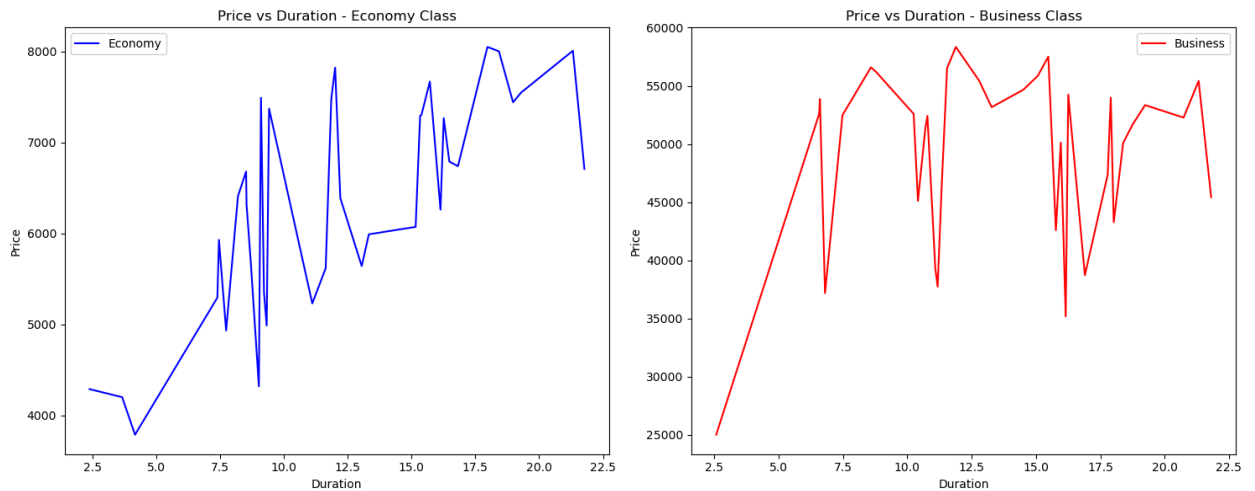


Dengan adanya kepopuleran tersebut terdapat indikasi bahwa maskapai dapat menaikkan harga dibandingkan dengan kompetitor dimana tiket kelas bisnis memiliki kenaikan yang lebih signifikan daripada kenaikan harga tiket kelas ekonomi.

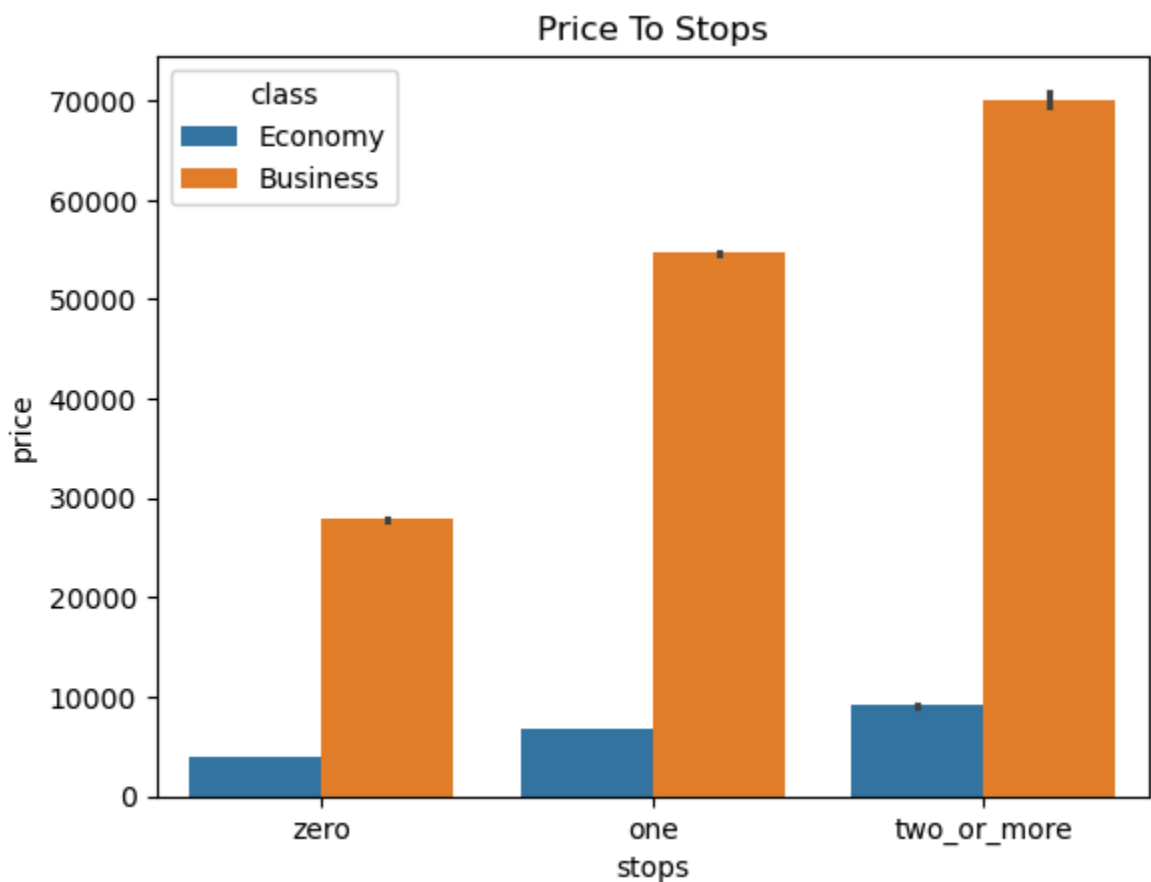


Berbicara mengenai kenaikan harga, berdasarkan kedua visualisasi diatas bisa dilihat bahwa semakin kecil jeda waktu antara tanggal pembelian dan penerbangan maka semakin tinggi biaya yang harus dikeluarkan. Untuk kelas ekonomi ada peningkatan yang signifikan mulai dari hari ke-15 sebelum penerbangan dan untuk

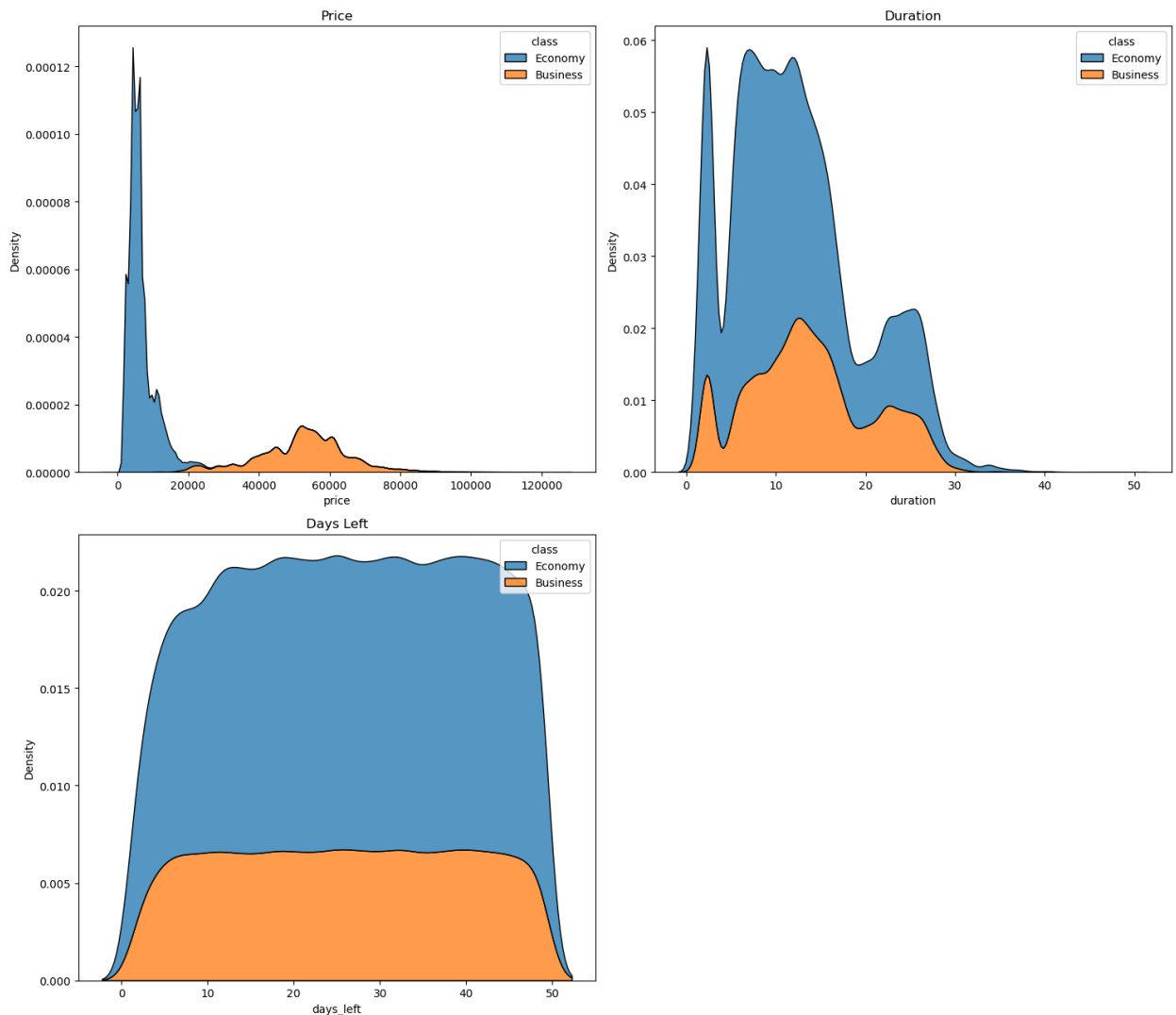
kelas bisnis ada peningkatan yang signifikan mulai dari hari ke-10 sebelum keberangkatan..



Berdasarkan Visualisasi diatas dapat dilihat bahwa, semakin lama waktu penerbangan semakin tinggi harga dari tiket tersebut. Kenaikan harga ticket pada lama penerbangan akan berpengaruh pada kebutuhan pada maintenance pesawat.



Berdasarkan visualisasi diatas terdapat indikasi jika penerbangan dengan transit akan memiliki harga yang lebih tinggi. Dimana semakin banyak transit maka harga menjadi lebih tinggi. Namun kita juga dapat menyimpulkan semakin lama waktu penerbangan maka semakin mahal harga tiket.



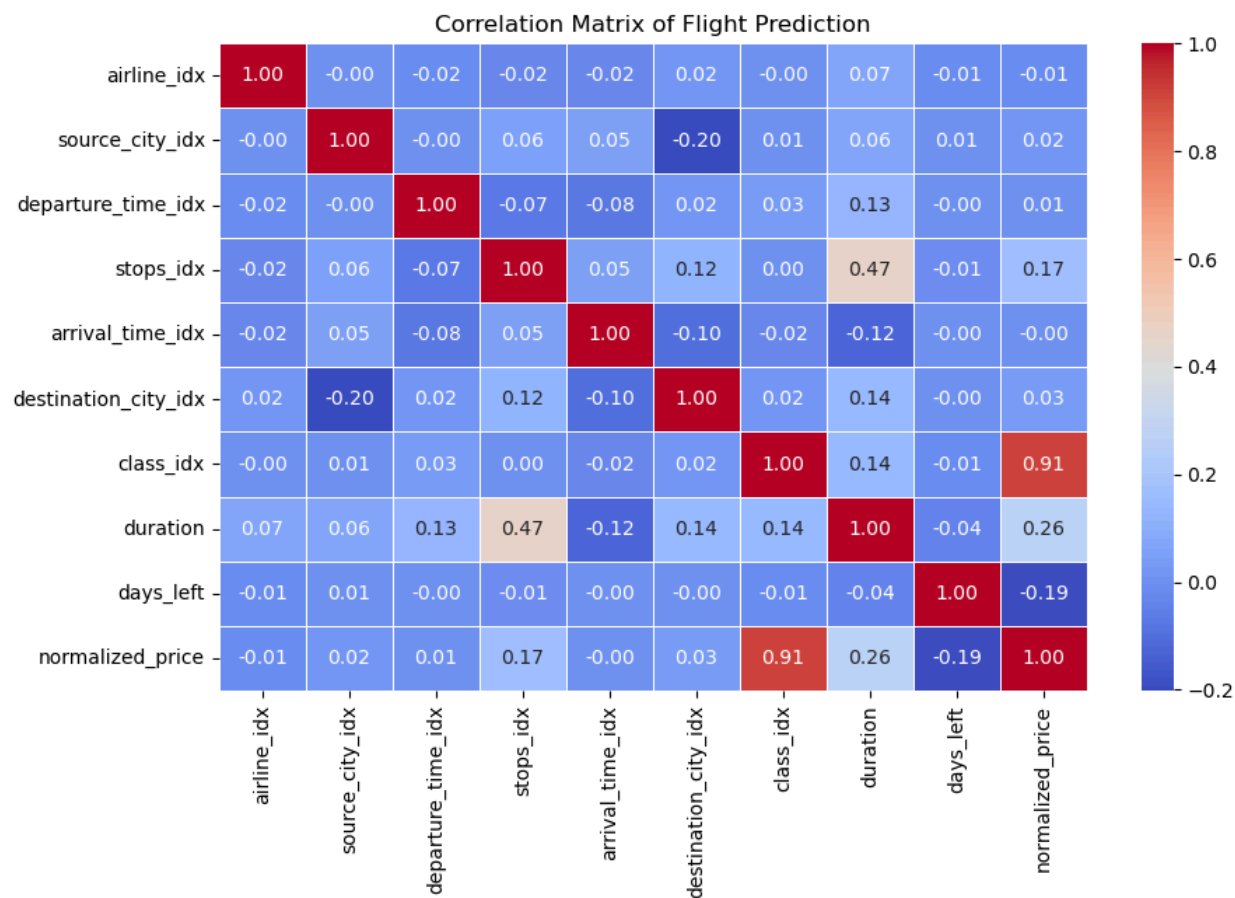
Berdasarkan Density Plot Diatas dapat diketahui bahwa:

1. Kelas Ekonomi memiliki distribusi data yang pada harga yang highly skewed, sehingga memiliki beberapa bukit dan lembah.
2. Data Kelas Bisnis memiliki data dengan distribusi yang lebih lebar sehingga distribusi memiliki puncack yang jelas
3. Data Kelas Ekonomi dan Bisnis memiliki beberapa lama jam penerbangan yang umum sehingga memiliki beberapa bukit dan lembah. Kelas

Ekonomi memiliki lebih banyak penerbangan dengan durasi yang lebih singkat.

- 4. Kelas Bisnis memiliki durasi penerbangan yang lebih lama dibandingkan dengan kelas ekonomi
- 5. Distribusi Kelas Ekonomi dan kelas Bisnis dalam pembelian ticket hampir seragam.

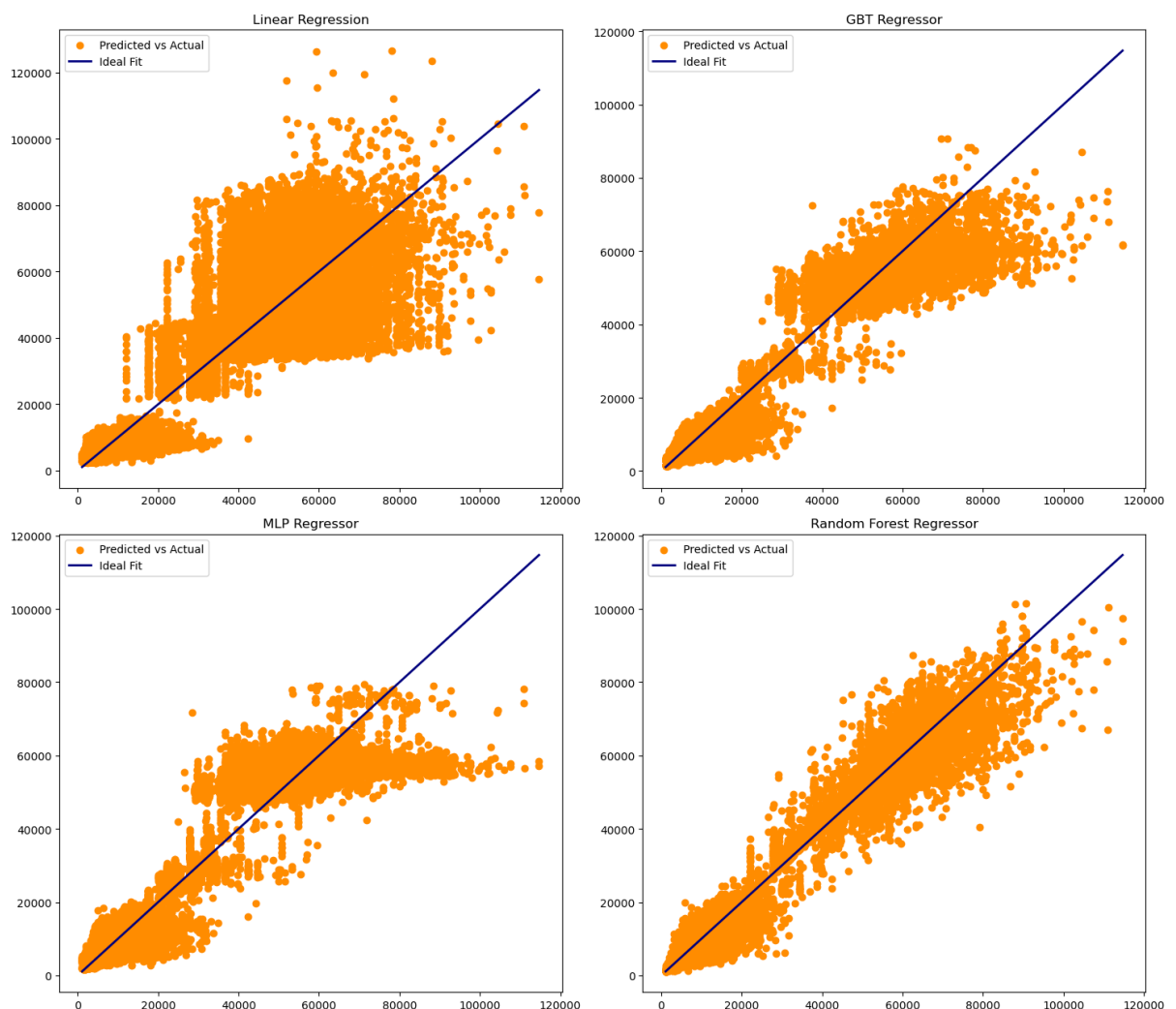
Untuk mencari korelasi antara setiap atribut kita akan menggunakan heatmap.



Dari heatmap diatas dapat dilihat bahwa kelas, durasi, jumlah transit, tujuan, asal, dan waktu penerbangan berpengaruh positif terhadap harga, sedangkan tanggal pembelian tiket maskapai penerbangan berpengaruh negatif terhadap harga. Pengecualian untuk waktu kedatangan yang tidak mempengaruhi harga.

Waktu penerbaangan dengan korelasi negatif menandakan semakin dekat tanggal pembelian tiket dengan jadwal penerbangan, maka harga akan semakin mahal. Untuk maskapai semakin terkenal sebuah maskapai akan mengakibatkan harga yang lebih mahal.

Selanjutnya kita akan menggunakan machine learning untuk menemukan atribut apa yang paling penting untuk menentukan biaya penerbangan. Kita mencoba untuk menggunakan 4 model machine learning yaitu Linear Regression, Gradient Boosting Tree Regressor, Multi Layer Perceptron Regressor, dan Random Forest Regressor. Berikut adalah hasil prediksi model.



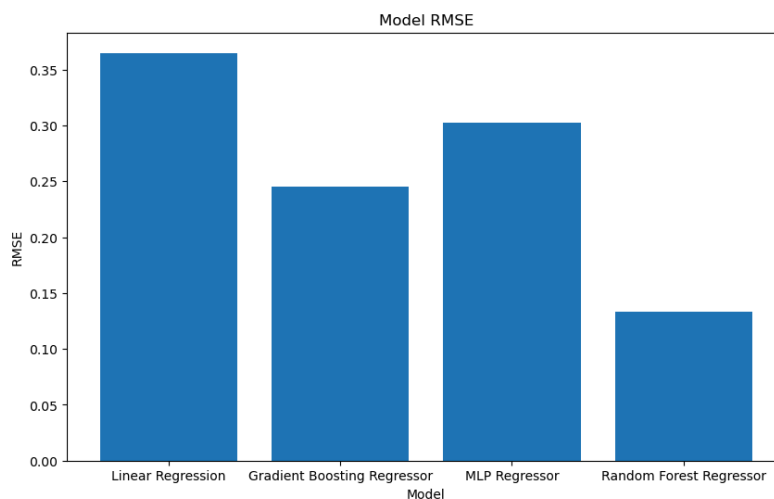
Berdasarkan 4 Model Regresi yang dilakukan dalam pengolahan data, dapat diketahui bahwa:

1. Model Linear Regression: Banyak Data yang menjauhi dari garis regresi sehingga, model menunjukkan hasil yang kurang bagus. Kepadatan

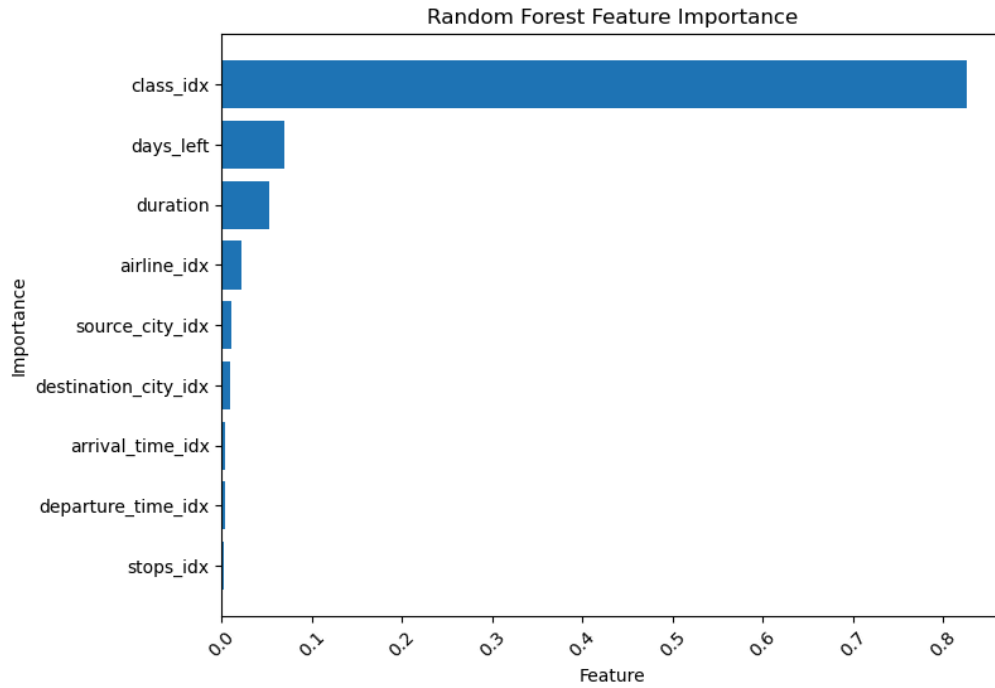
Penyebaran data juga menunjukan linear regression lebih padat pada posisi yang menjauhi garis regresi.

2. Model GBT: Model GBT menunjukkan hasil yang lebih baik, karena data cukup padat mendekati garis regresi.
3. Model MLP: Model MLP menunjukan hasil yang kurang baik, karena kepadatan data lebih tersebar menjauhi garis regresi.
4. Model Random Forest: Model Random Forest menunjukan hasil yang baik dengan kepadatan data lebih banyak yang mendekati garis regresi.

Model akan semakin bagus, jika lebih banyak data yang mendekati garis regresi. Model Regresi selain random forest menunjukan hasil prediksi yang menghasilkan prediksi yang sangat jauh terhadap garis regresi



Dari keempat model yang dicoba, Random Forest Regressor memiliki peforma yang paling bagus dengan nilai RMSE yang terkecil yaitu 0.132 dan berdasarkan model Random Forest Regressor berikut adalah fitur-fitur yang paling mempengaruhi harga.



Dari visualisasi tersebut diperkirakan bahwa 5 fitur terpenting adalah Class, Days_left, Duration, Airline, dan Source_city sedangkan 4 fitur terburuk adalah Destination_city, Arrival_time, Departure_time, dan Stops.

Kesimpulan dan Rekomendasi

Dari analisis yang telah dilakukan disimpulkan bahwa kelas penerbangan, tanggal pembelian tiket, durasi penerbangan dan maskapai penerbangan adalah hal terbesar yang mempengaruhi harga penerbangan. Maskapai yang paling banyak diminati masyarakat memiliki harga yang lebih tinggi dibanding kompetitornya dan secara keseluruhan penerbangan dengan kelas bisnis akan selalu memiliki harga diatas kelas ekonomi dan akan memiliki peningkatan yang lebih tinggi daripada ekonomi. Waktu pembelian tiket juga sangat berkontribusi dalam menentukan harga, semakin jauh tanggal pembelian dari tanggal keberangkatan maka semakin kecil harga tiket yang didapat.

Fitur asal penerbangan, tujuan penerbangan, dan jumlah transit memiliki importance yang rendah dalam penentuan harga. Namun perlu diingat ketiga hal tersebut memiliki korelasi terhadap durasi penerbangan yang sangat mempengaruhi harga, dimana semakin lama durasi penerbangan maka semakin tinggi harga tiket.

Dalam proses pemilihan penerbangan yang ideal bagi konsumen, ada beberapa rekomendasi yang perlu dipertimbangkan:

1. Jangan pernah terbang di kelas bisnis dengan budget minim
2. Pertimbangkan maskapai yang kurang populer karena harga yang ditawarkan bisa lebih murah
3. Belilah tiket dari jauh-jauh hari, setidaknya 2 minggu sebelum tanggal penerbangan
4. Usahakan untuk mencari tiket dengan durasi penerbangan terendah, usahakan untuk mencari penerbangan langsung tanpa transit.