



# AIRLINE PASSENGER SATISFACTION PROJECT

By Adristi Shafiya



# Table of Contents

Introduction

Data Preparation

Exploratory Data Analysis

Data Preprocessing

Machine Learning Model

Recommendation



# Introduction

Perlunya dilakukan evaluasi terhadap pelayanan yang diberikan oleh maskapai untuk mengetahui apakah penumpang merasa puas atau tidak. Dikarenakan hal itu bisa menyebabkan berkurangnya penumpang yang menggunakan maskapai.

Project ini bertujuan untuk

1. Menentukan faktor apa saja yang bisa mempengaruhi tingkat satisfaction penumpang
2. Memprediksi kepuasan pelanggan
3. Memberikan rekomendasi untuk meningkatkan performa maskapai





# DATA PREPARATION

## About Dataset



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103904 entries, 0 to 103903
Data columns (total 25 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Unnamed: 0                               103904 non-null  int64
1   id                                         103904 non-null  int64
2   Gender                                   103904 non-null  object
3   Customer Type                            103904 non-null  object
4   Age                                       103904 non-null  int64
5   Type of Travel                           103904 non-null  object
6   Class                                    103904 non-null  object
7   Flight Distance                          103904 non-null  int64
8   Inflight wifi service                   103904 non-null  int64
9   Departure/Arrival time convenient       103904 non-null  int64
10  Ease of Online booking                  103904 non-null  int64
11  Gate location                           103904 non-null  int64
12  Food and drink                          103904 non-null  int64
13  Online boarding                         103904 non-null  int64
14  Seat comfort                            103904 non-null  int64
15  Inflight entertainment                  103904 non-null  int64
16  On-board service                        103904 non-null  int64
17  Leg room service                        103904 non-null  int64
18  Baggage handling                        103904 non-null  int64
19  Checkin service                         103904 non-null  int64
20  Inflight service                        103904 non-null  int64
21  Cleanliness                             103904 non-null  int64
22  Departure Delay in Minutes               103904 non-null  int64
23  Arrival Delay in Minutes                 103594 non-null  float64
24  satisfaction                             103904 non-null  object
dtypes: float64(1), int64(19), object(5)
memory usage: 19.8+ MB
```

- Data memiliki 25 kolom dan 103904 baris
- Tipe data pada semua kolom sudah sesuai
- Terdapat beberapa kolom yang merupakan kolom dengan values berupa rank (1 – 5)
- Kolom customer ID akan di drop karena mengandung banyak unique value





# Missing Value & Duplicated



## MISSING VALUE

Arrival Delay in Minutes	310	0.298352
--------------------------	-----	----------

Handling Missing Value  
menggunakan rerata (mean)



## DUPLICATED VALUE

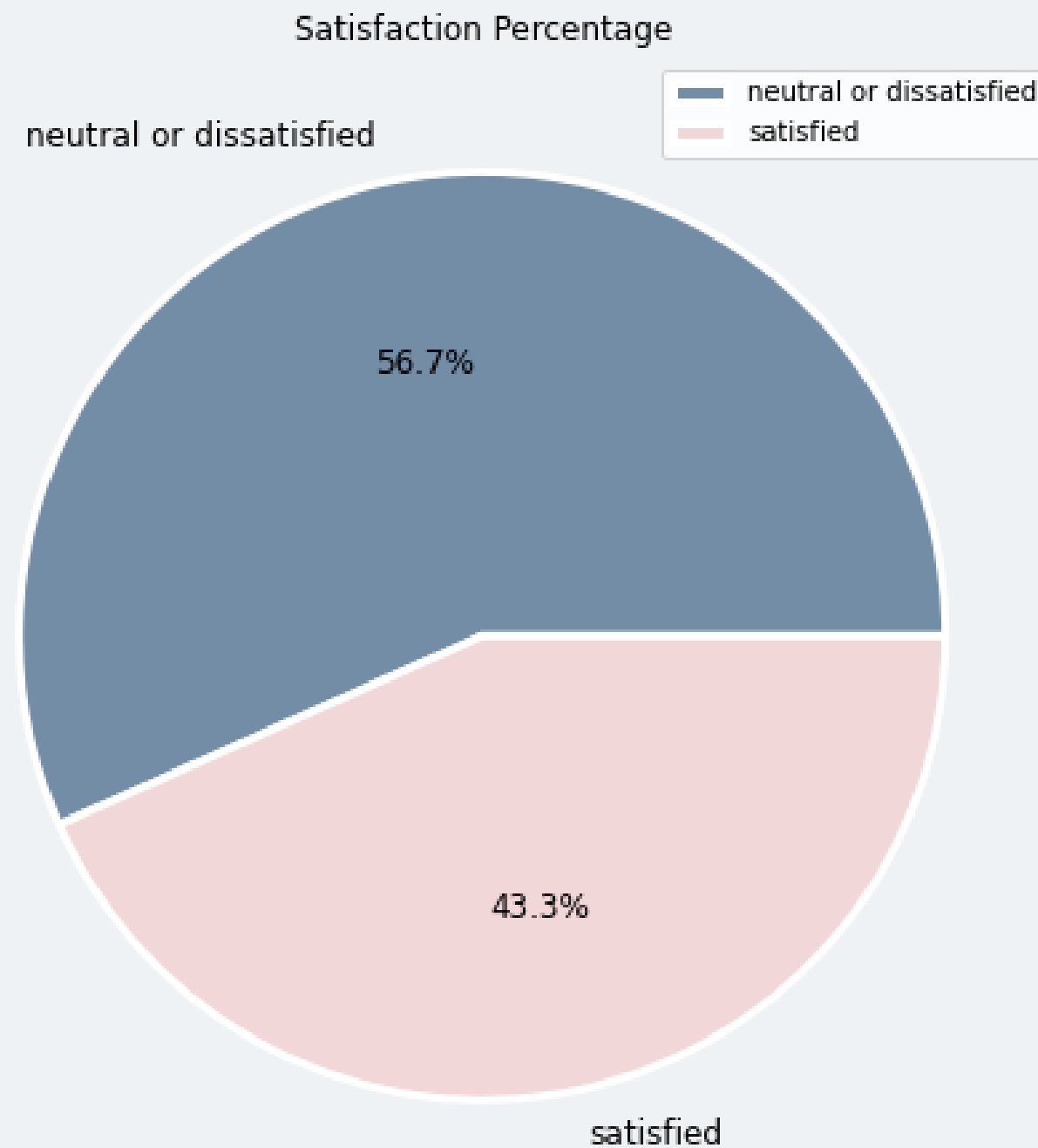
```
satisf_baru.duplicated().sum()  
  
0
```



Tidak ada Missing Value!



# EXPLORATORY DATA ANALYSIS



- Dari total 103904, terdapat 56,7% atau 58879 penumpang dalam kategori netral atau tidak puas
- Dan sebanyak 43,3% atau 45025 penumpang merasa puas terhadap maskapai

# Statistical Summary

## Categorical Column

1. Penumpang terbanyak adalah dengan jenis kelamin **perempuan**
2. Tipe customer yang seringkali menggunakan maskapai ini adalah tipe **loyal customer**
3. Kebanyakan tujuan penumpang menggunakan transportasi pesawat dengan maskapai ini adalah untuk **tujuan bisnis**
4. **Business class** merupakan tipe kelas yang seringkali dipilih oleh kebanyakan penumpang
5. Tingkat kepuasan penumpang terhadap maskapai kebanyakan **netral atau tidak puas**

# Univariate Analysis

Value counts of -- outlier\_flight\_distance-- column

not-outlier 101613

2.25%

outlier 2291

Name: outlier\_flight\_distance, dtype: int64

Value counts of -- outlier\_checkin\_service-- column

not-outlier 91013

14.16%

outlier 12891

Name: outlier\_checkin\_service, dtype: int64

Value counts of -- outlier\_departure\_delay-- column

not-outlier 89375

16.26%

outlier 14529

Name: outlier\_departure\_delay, dtype: int64

Value counts of -- outlier\_arrival\_delay-- column

not-outlier 89950

15.51%

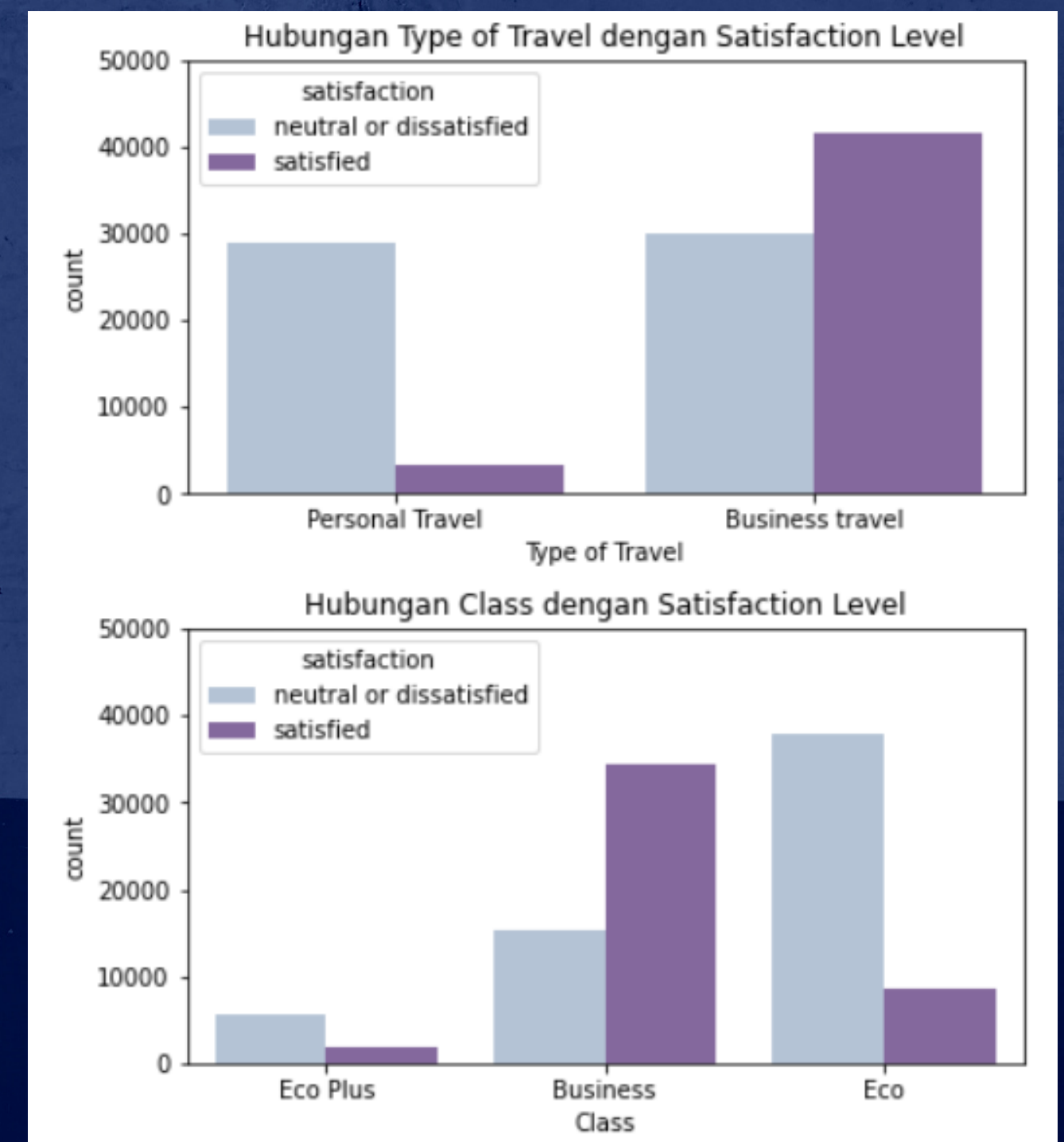
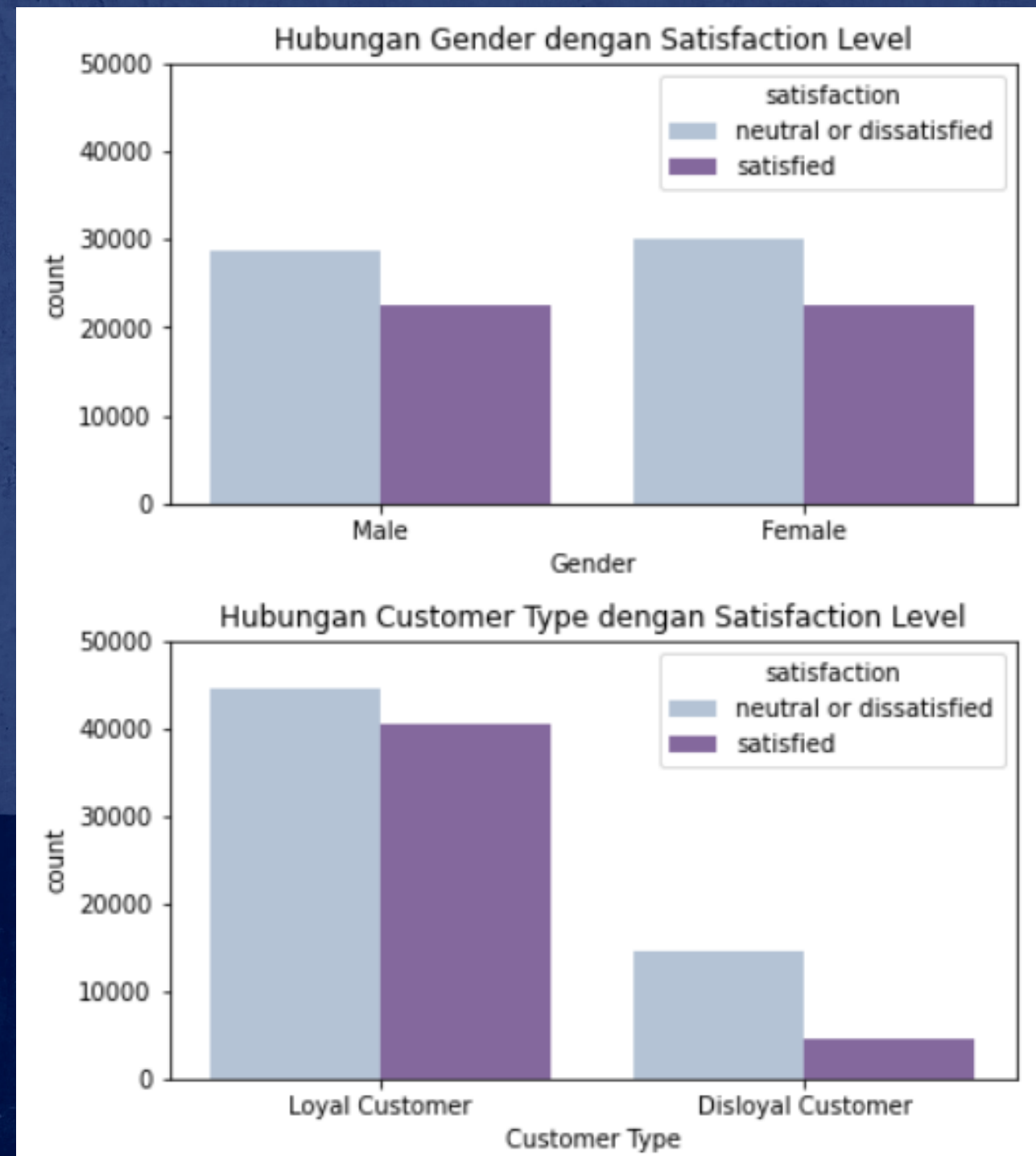
outlier 13954

Name: outlier\_arrival\_delay, dtype: int64

Handling outlier dengan log transformation



# Categorical Analysis





# DEEP DIVE QUESTION

01

Bagaimana pengaruh tingkat kepuasan penumpang berdasarkan umur?

02

Bagaimana pengaruh tipe perjalanan, tipe penumpang, dan tipe kelas terhadap tingkat kepuasan penumpang?

03

Bagaimana pengaruh jarak perjalanan terhadap tipe perjalanan dan tipe kelas yang dipilih oleh penumpang loyal dan tidak loyal?

04

Bagaimana pengaruh keterlambatan keberangkatan dan kedatangan terhadap tingkat kepuasan penumpang?

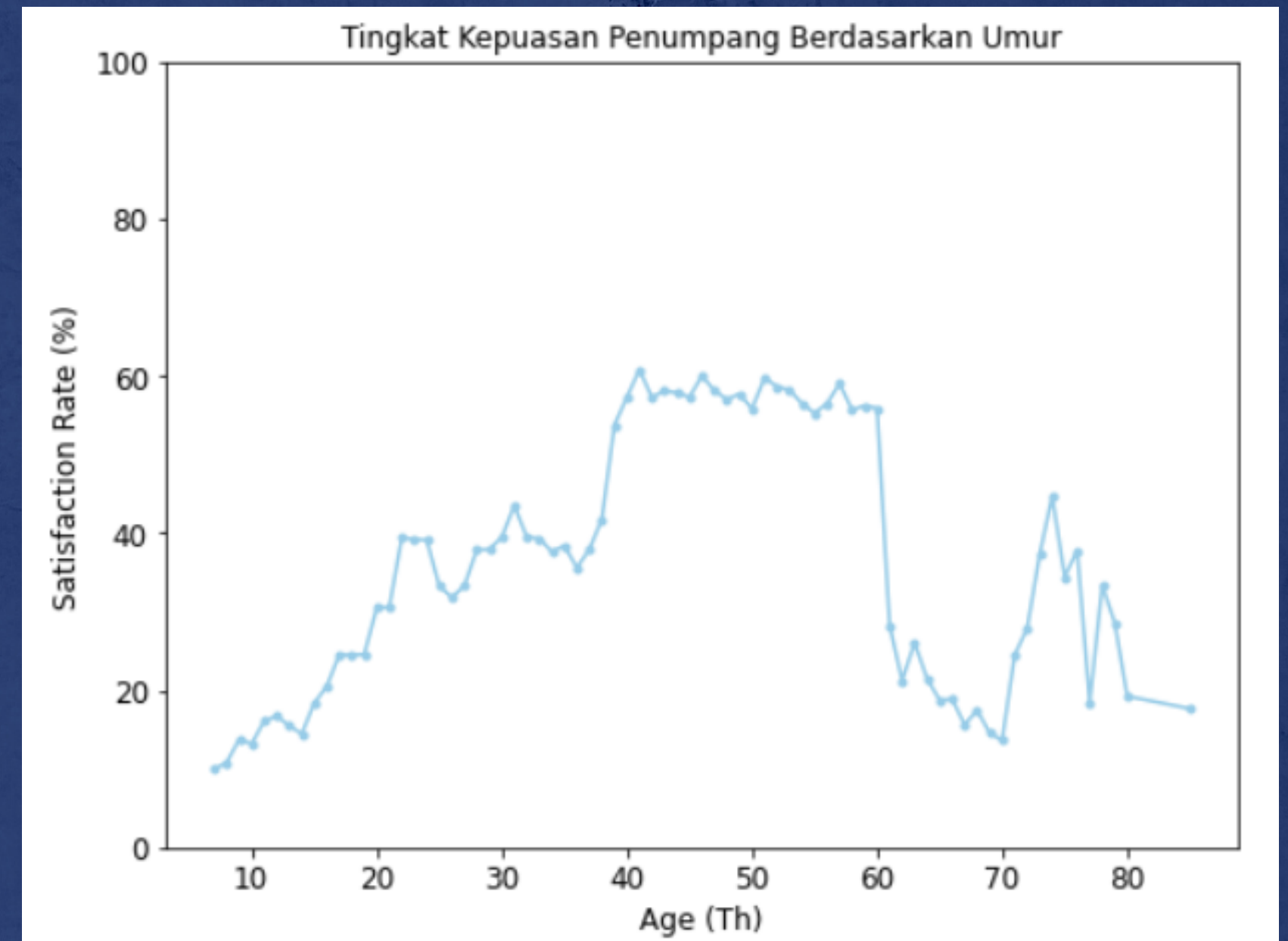
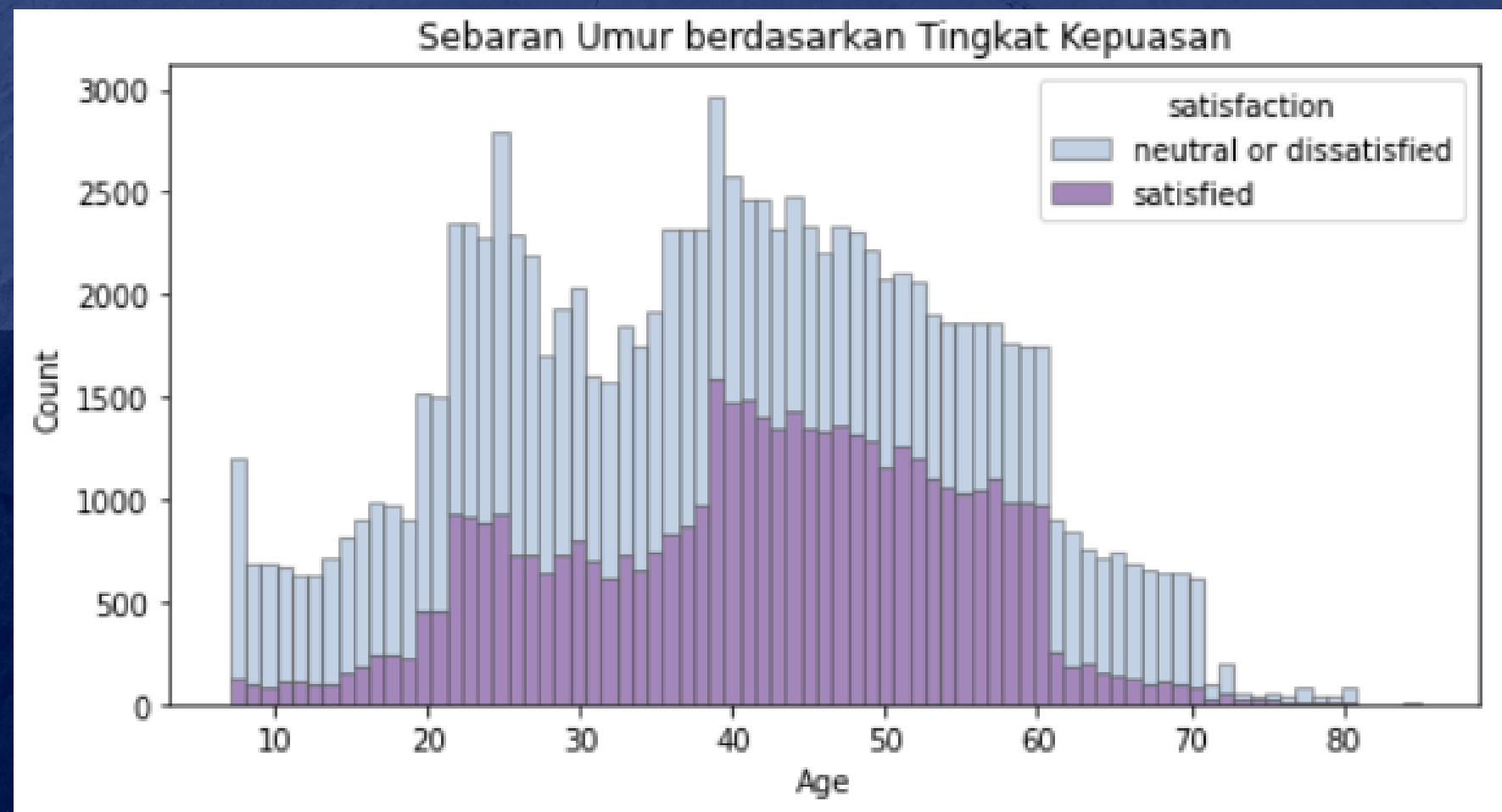
05

Apa saja service yang memiliki rate paling buruk dan paling baik?

06

Bagaimana rerata jarak tujuan penumpang terhadap tipe kelas pesawat yang dipilih?

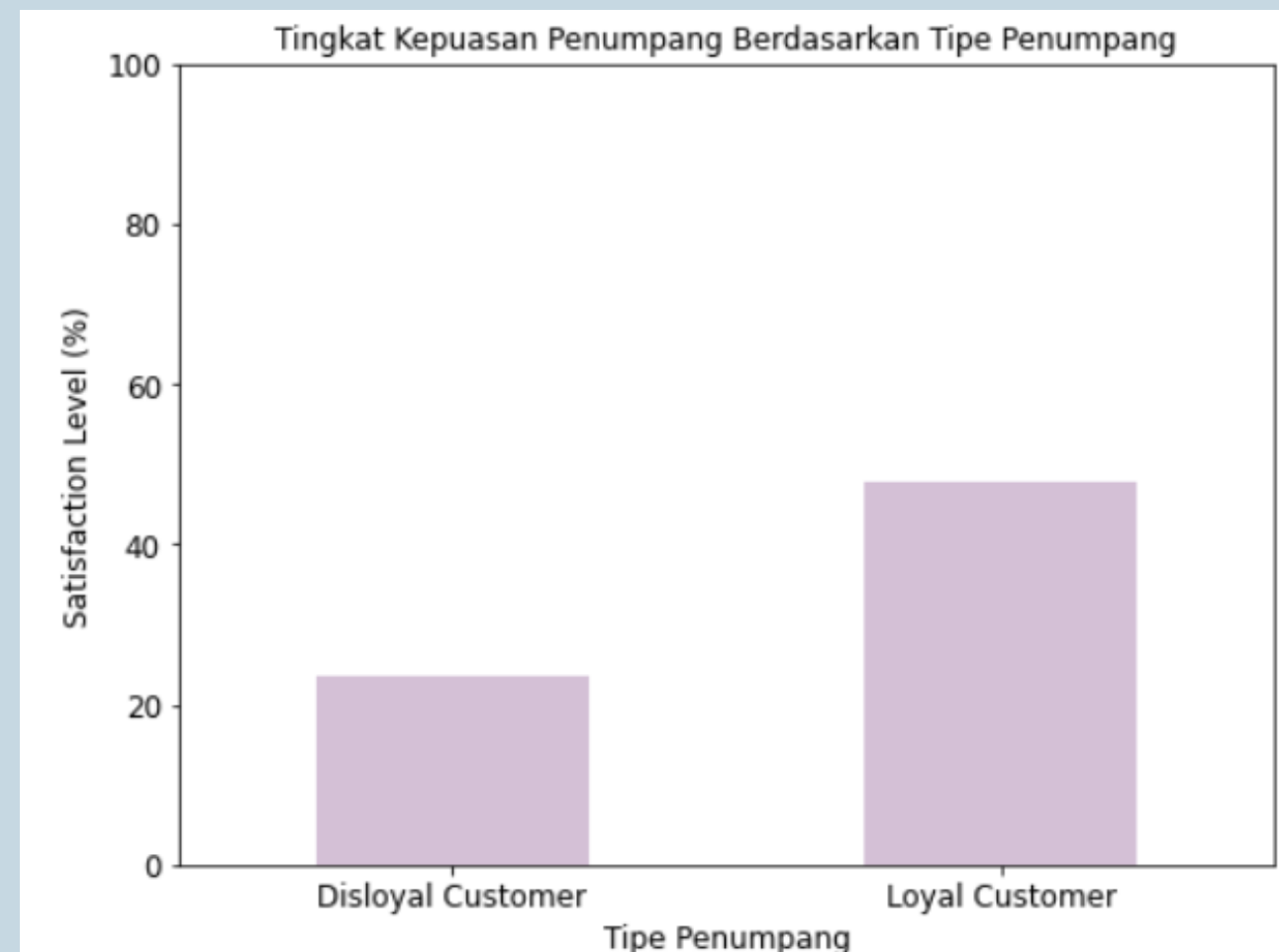
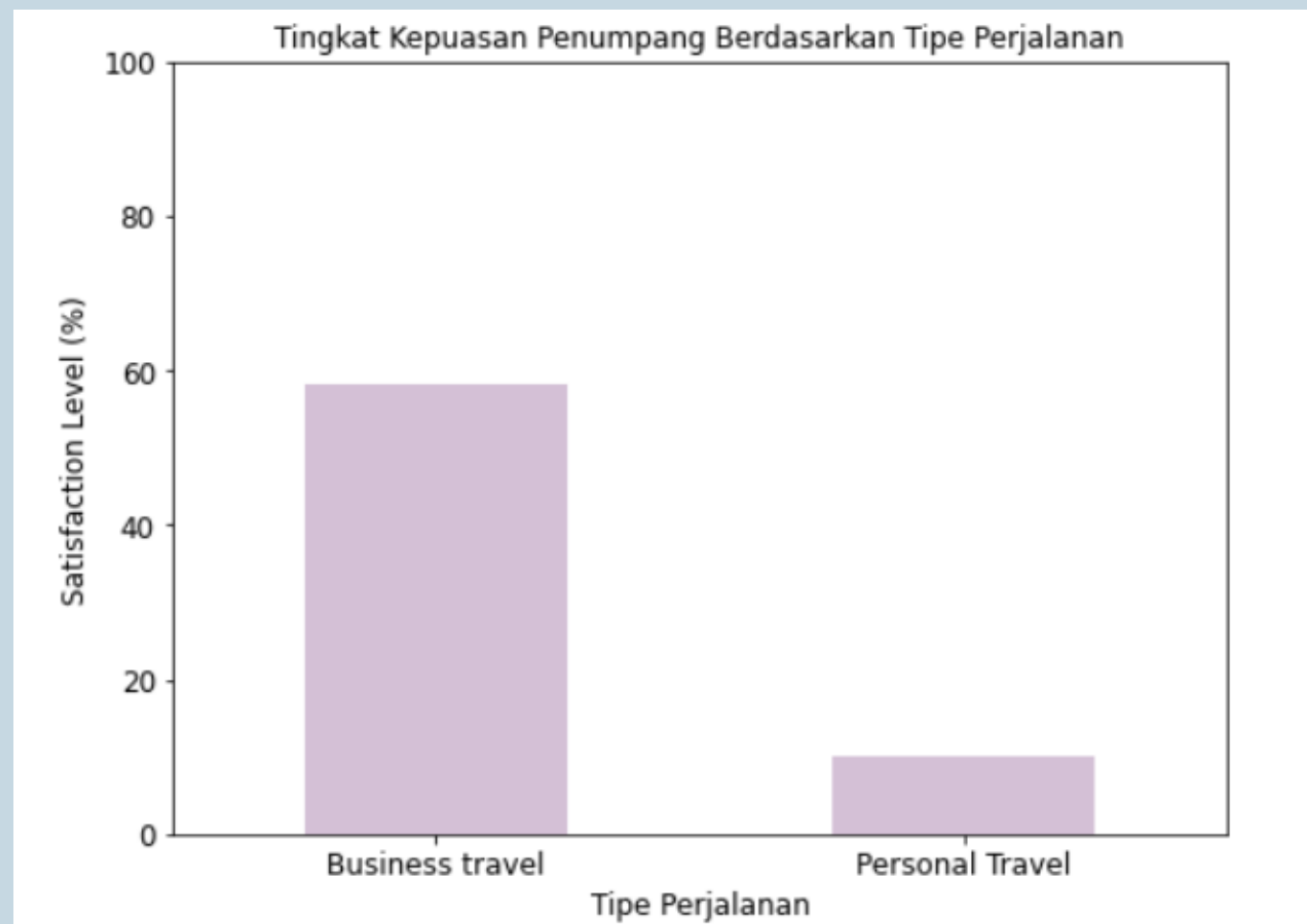
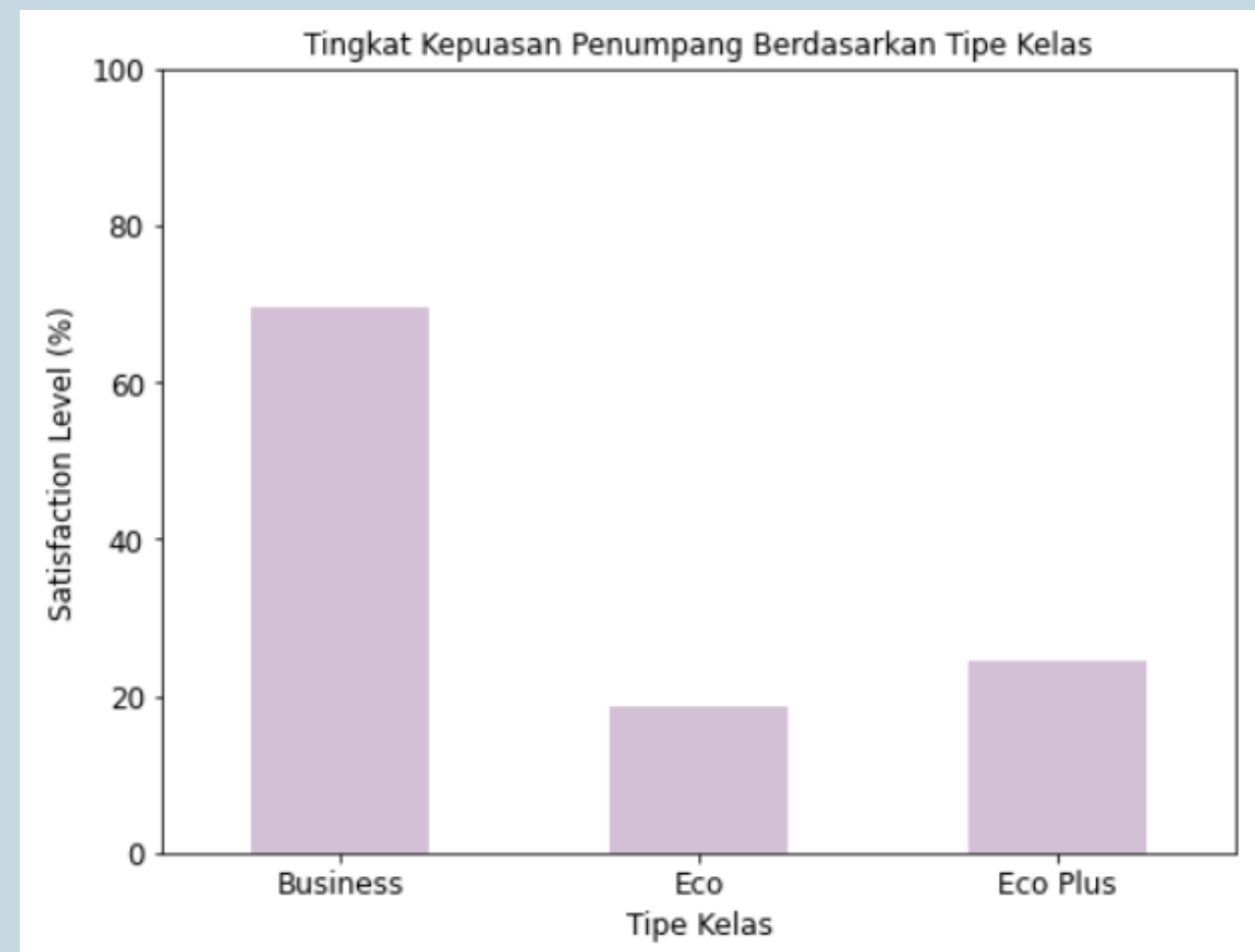
# Bagaimana pengaruh tingkat kepuasan penumpang berdasarkan umur?



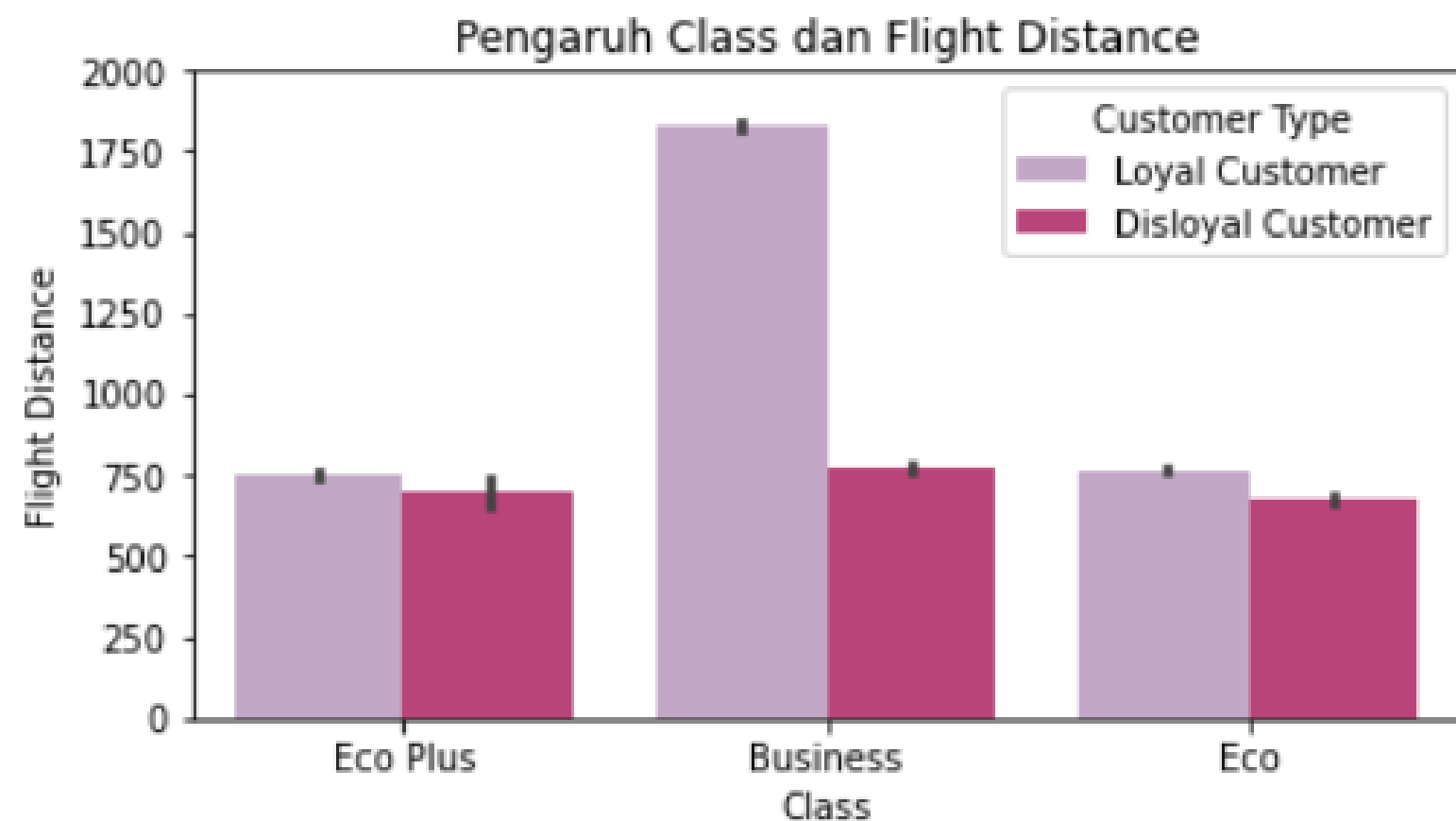
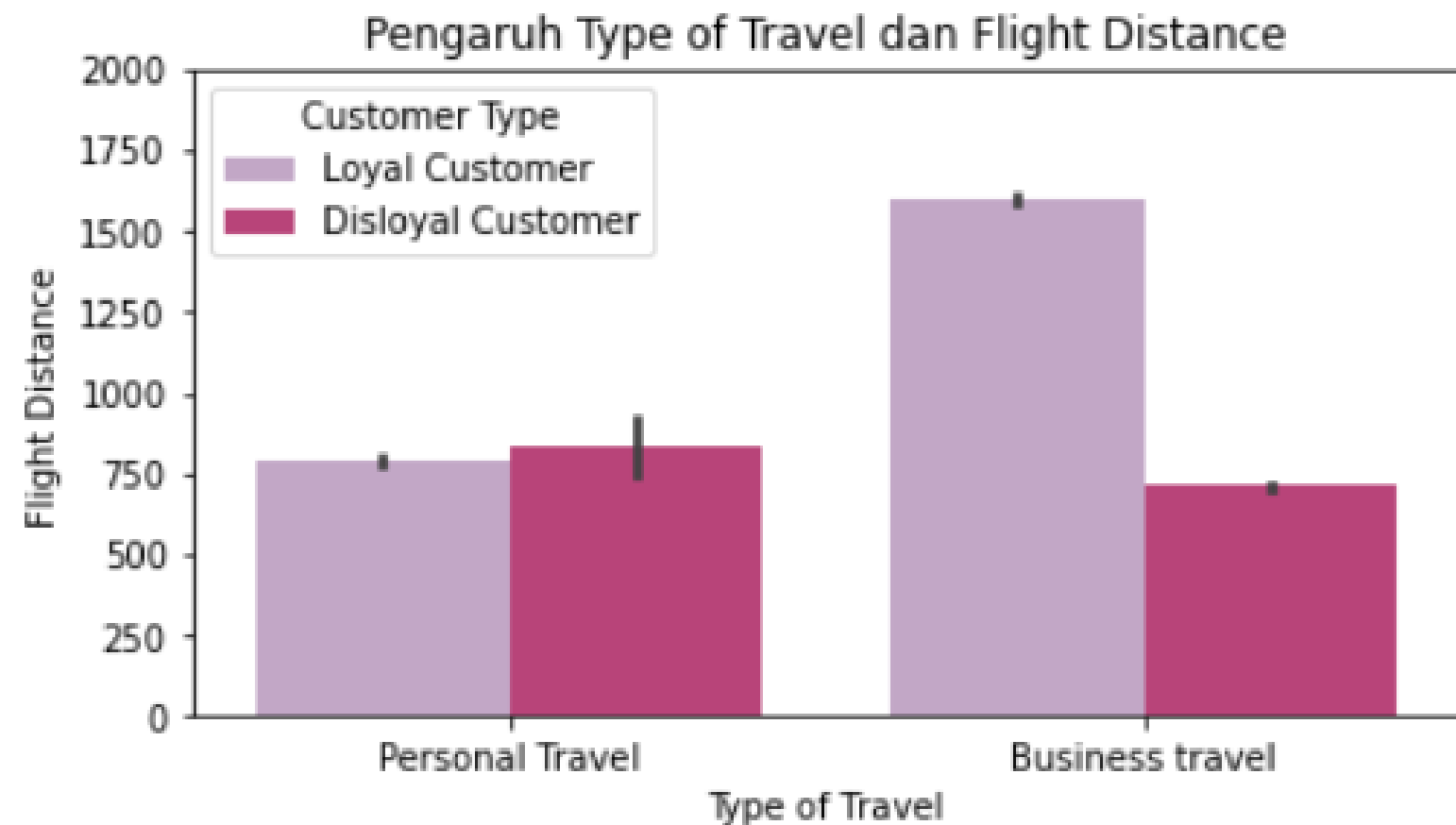
- Kebanyakan penumpang yang berusia antara 40-60 cenderung merasa puas terhadap maskapai
- Penumpang yang berusia dibawah 40 tahun dan diatas 60 tahun cenderung merasa tidak puas terhadap maskapai



Bagaimana pengaruh tipe perjalanan, tipe penumpang, dan tipe kelas terhadap tingkat kepuasan penumpang?



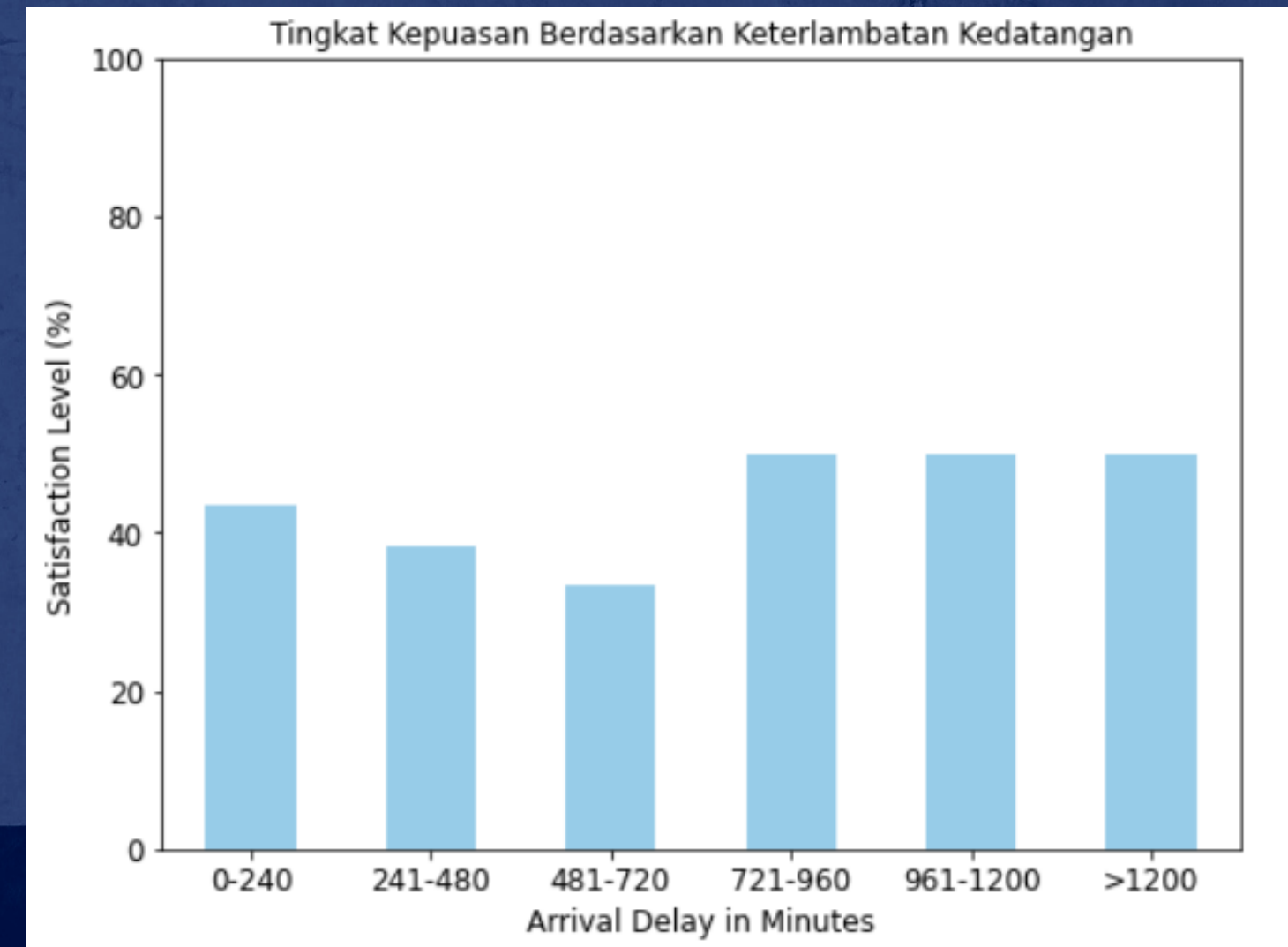
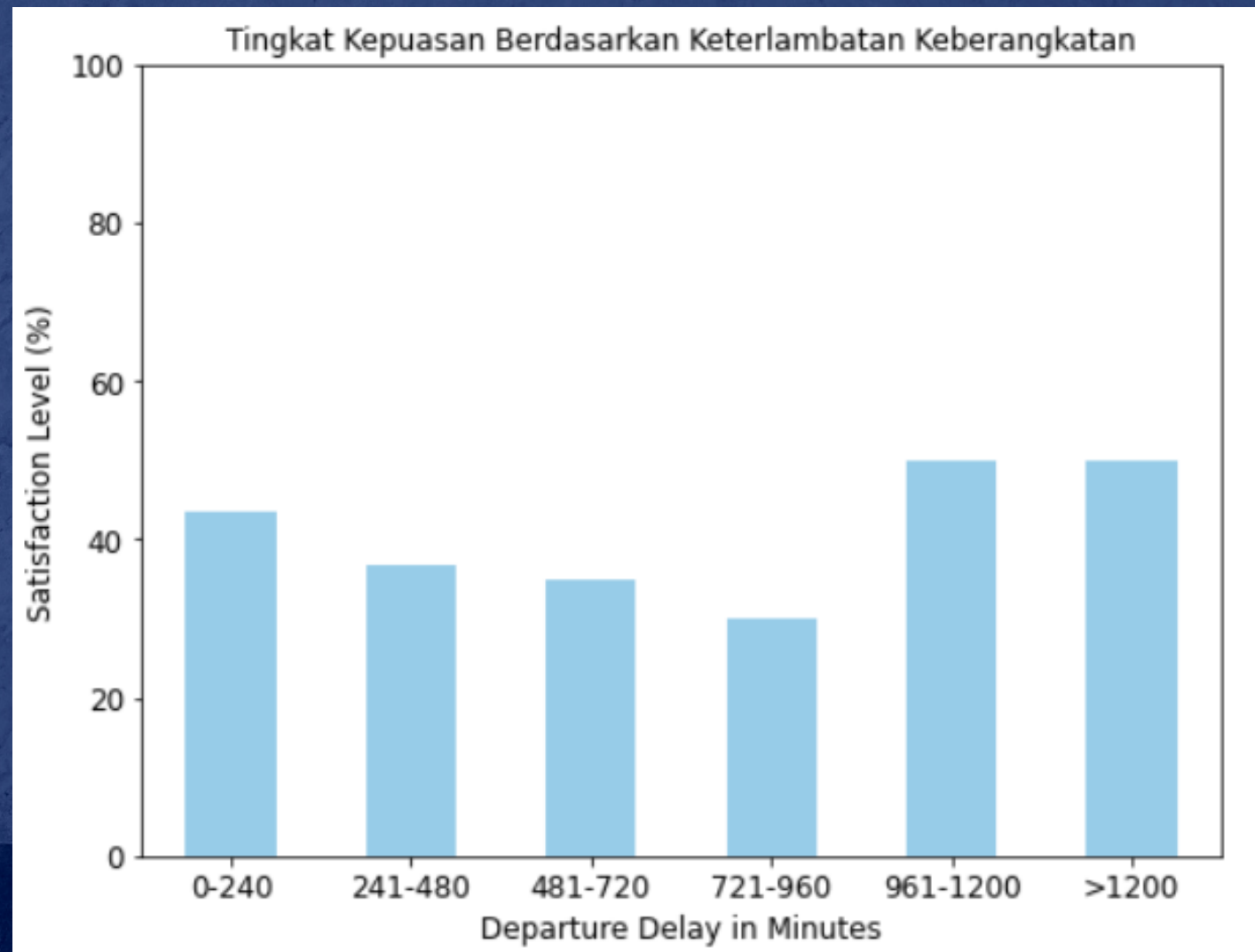
Bagaimana pengaruh jarak perjalanan terhadap tipe perjalanan dan tipe kelas yang dipilih oleh penumpang loyal dan tidak loyal?



- Penumpang loyal kebanyakan menggunakan maskapai ini untuk tujuan bisnis yang membutuhkan perjalanan sekitar 1600-1750 km
- Penumpang loyal biasanya menggunakan tipe kelas bisnis karena jarak perjalanan yang jauh sekitar 1750 - 2000 km

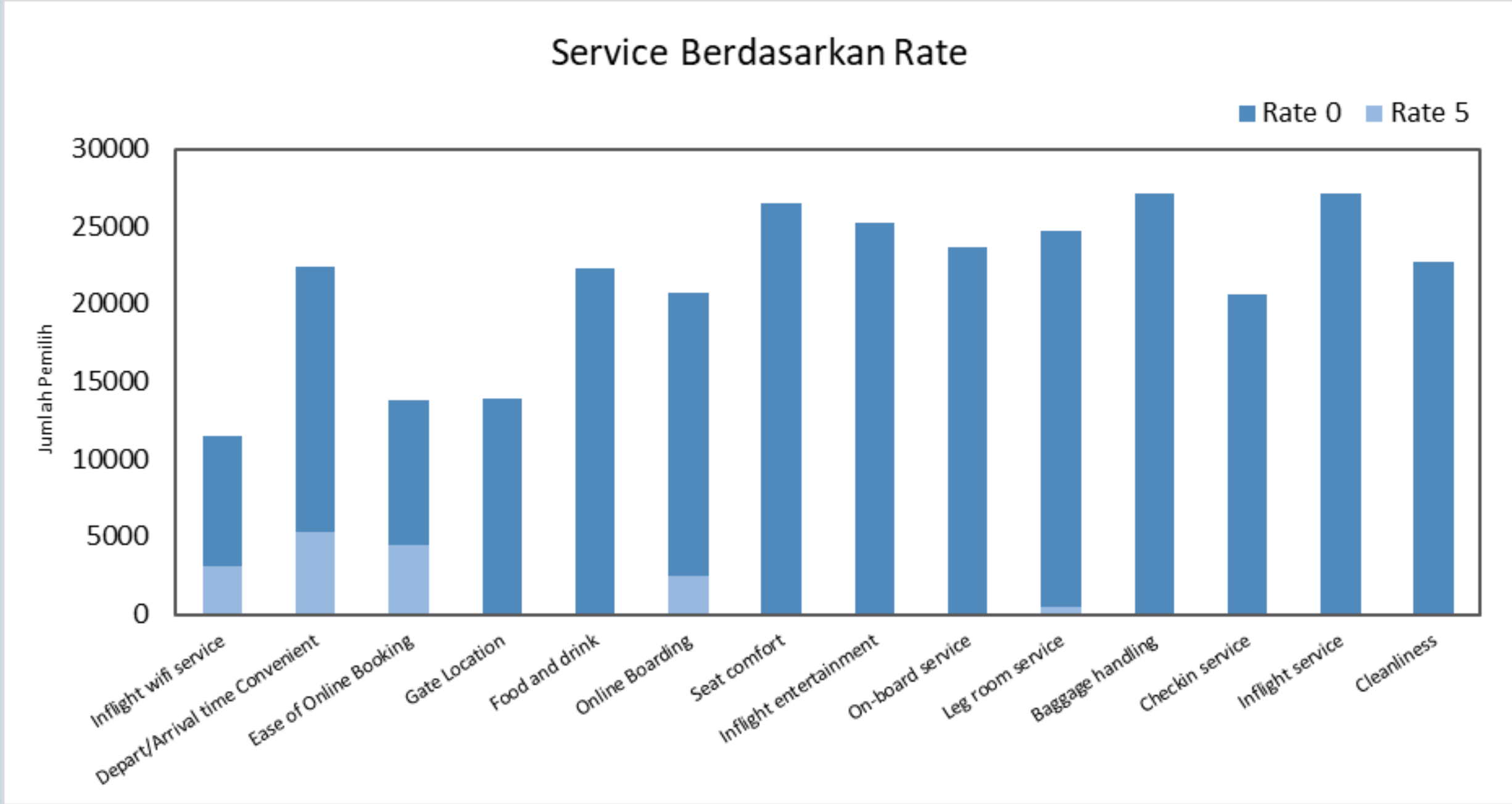


# Bagaimana pengaruh keterlambatan keberangkatan dan kedatangan terhadap tingkat kepuasan penumpang?



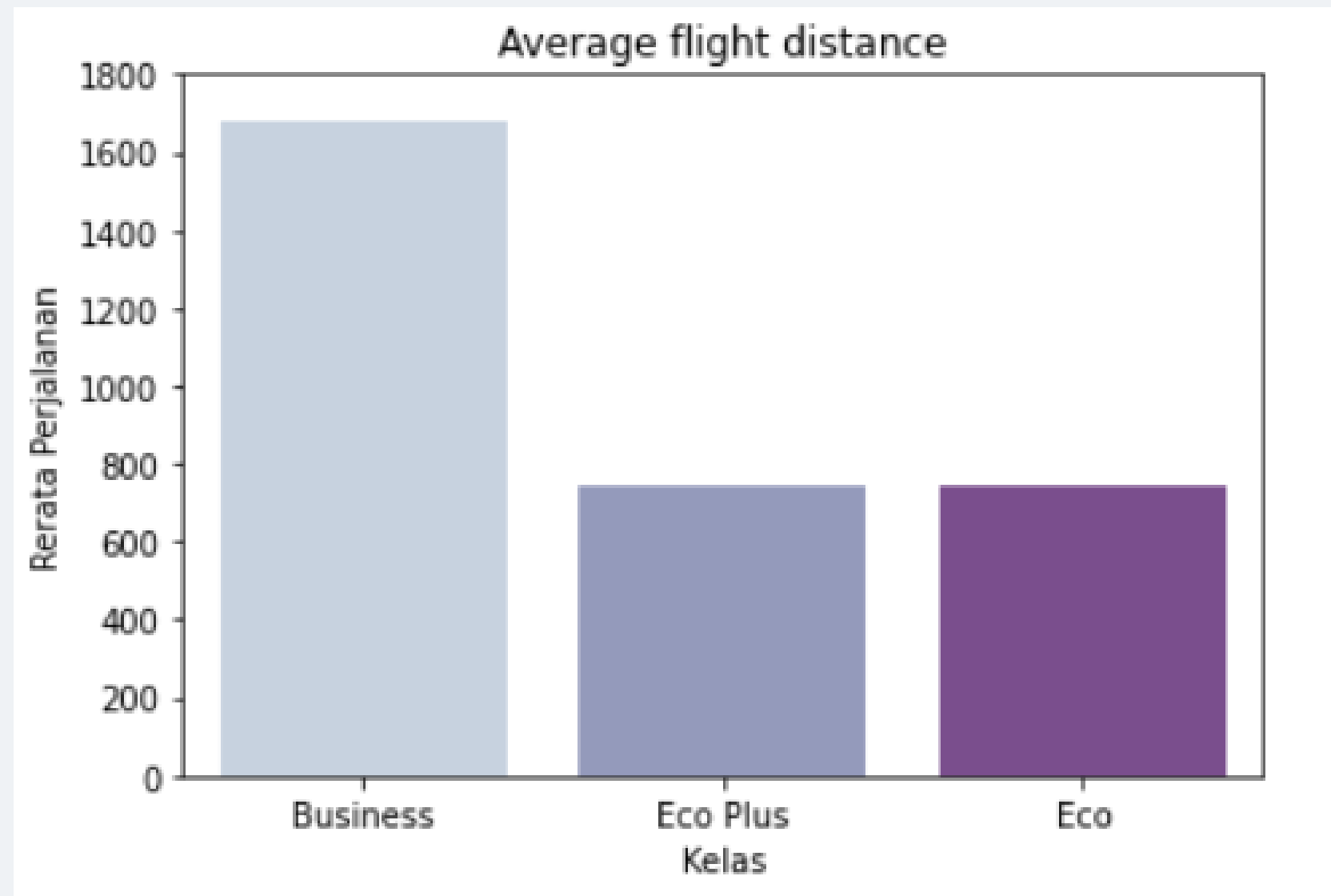
- Rerata tingkat kepuasan tertinggi adalah pada keterlambatan keberangkatan pada menit ke 0-240, menit ke 721 – 960, 961-1200 dan menit ke >1200
- Dalam hal ini bisa juga ditarik suatu asumsi bahwa tingkat kepuasan yang tinggi pada keterlambatan keberangkatan dengan durasi diatas 721 menit biasanya pihak maskapai memberikan informasi lebih awal.
- Berdasarkan gambar diatas, tingkat kepuasan berdasarkan keterlambatan kedatangan dan keberangkatan tidak jauh berbeda karena memang berdasarkan heatmap correlation antara departure dan arrival saling berkorelasi cukup tinggi

# Apa saja service yang memiliki rate paling buruk dan paling baik?





# Bagaimana rerata jarak tujuan penumpang terhadap tipe kelas pesawat yang dipilih?



- Tipe kelas bisnis merupakan tipe kelas yang banyak dipilih penumpang dengan rerata jarak perjalanannya adalah sebesar 1675,98 km



# Machine Learning Model

---

- Logistic Regression
- Decision Tree
- K-Nearest Neighbor
- Random Forest



## **What are we going to do?**

### **TARGET**

Meminimalisir False Positive

Fokus kasus yaitu meminimalisir false positif karena orang yang aslinya tidak puas tapi diprediksi model puas sehingga yang aslinya tidak puas tadi tidak diberikan kompensasi atas ketidak puasanya dan menyebabkan penumpang tidak lagi mau menggunakan maskapai tersebut dan menginformasikan kepada orang lain

# ANALYSIS SCORE

**K-Nearest  
Neighbor**

**Accuracy: 92.7%**  
**Precision: 94.6%**

**Logistic  
Regression**

**Accuracy: 87.7%**  
**Precision: 87.0%**

**Random Forest  
Undersampling**

**Accuracy: 95.9%**  
**Precision: 95.6%**

**Decision Tree**

**Accuracy: 94.48%**  
**Precision: 92.9%**

**Random Forest  
Oversampling**

**Accuracy: 96.0%**  
**Precision: 96.13%**

**Random Forest  
Biasa**

**Accuracy: 96.2%**  
**Precision: 97%**

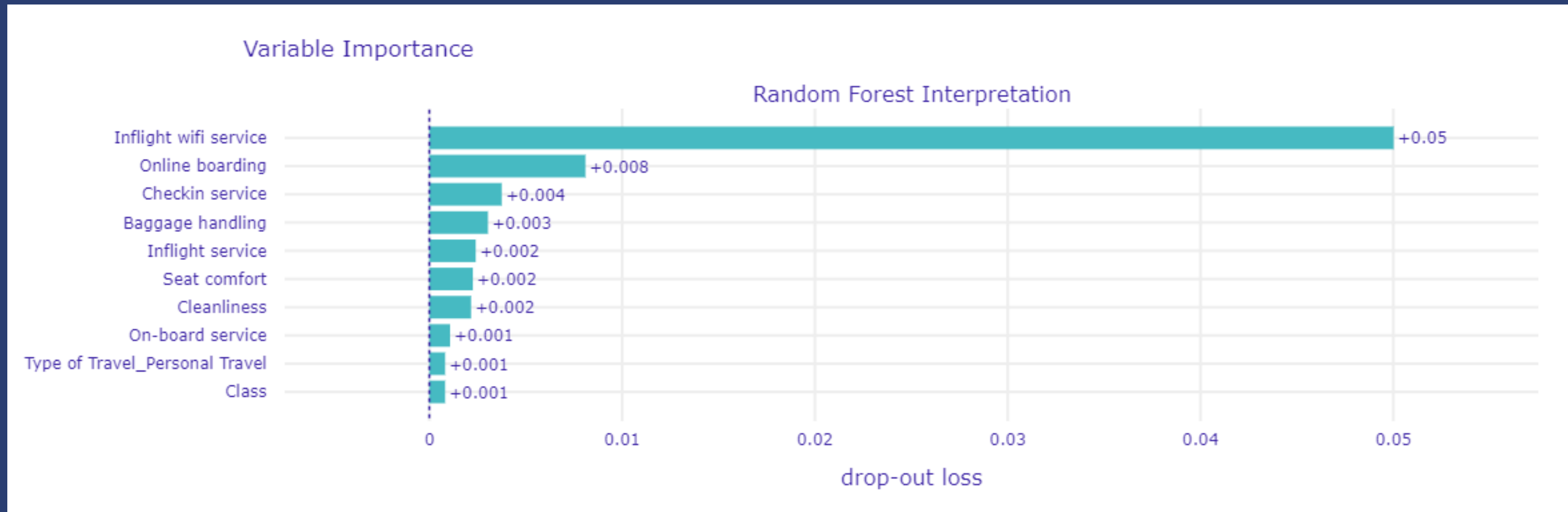
**Random Forest  
Combining**

**Accuracy: 96.1%**  
**Precision: 96.3%**



**Best Model**

# FEATURE IMPORTANCE



- Berdasarkan feature importance didapatkan bahwa terdapat 10 fitur yang mempengaruhi tingkat kepuasan penumpang.
- Fitur yang paling berpengaruh kebanyakan memang terkait dengan pelayanan yang diberikan maskapai. Dan berdasarkan gambar diatas inflight wifi service merupakan fitur yang paling berpengaruh terhadap tingkat kepuasan penumpang



# Rekomendasi

1. Meningkatkan pelayanan untuk pesawat tipe kelas Eco.
2. Perlu ditingkatkannya pelayanan terkait dengan Depart/Arrival time convenient, Ease of Online booking, dan Online boarding dikarenakan banyak sekali penumpang yang memberikan rate kecil.
3. Pelayanan Food and drink dan Leg room service masih termasuk dalam kategori baik karena hanya ada beberapa penumpang yang memberikan rate kecil namun pada kedua pelayanan ini masih harus ditinjau kembali
4. Meminta feedback kepada penumpang terkait service apa yang harus ditingkatkan, diperbaiki, dan dipertahankan untuk bisa meminimalisir ketidakpuasan penumpang





Thank you.

