

Airline Passenger Satisfaction

EXPLORATORY DATA ANALYSIS



06.04.2023

DATA ANALYSIS

ABOUT US

TEAM FOUR

MEMBERSHIP: 4 STUDENTS
GROUPS: KM-03 / KM-02



DOCUMENTATION & TEAMWORK

NAZARII KUTSALABA
-CODEBOOK
-FILE MANAGEMENT
-VERSION CONTROL

DATA
MAKSYM KHOK
-COLLECTION
-MERGE
-PREPROCESSING



OLEKSII HARMASH

EDA
-ANALYSIS
-VISUALISING
-BUSINESS BENEFITS



REVIEW & RESULTS

NAZAR MASLOV
-REVIEW
-CORRECTION
-APPEARANCE



RESEARCH **MOTIVATION**

4 STEP SCHEMA:



RESEARCH MOTIVATION

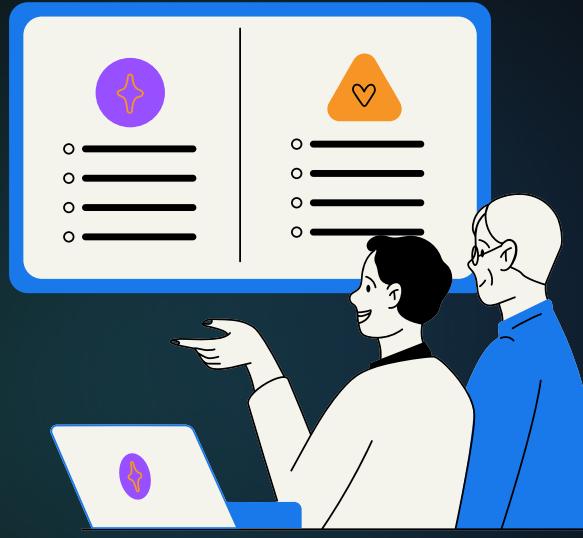
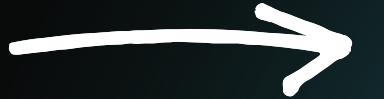
4 STEP SCHEMA:



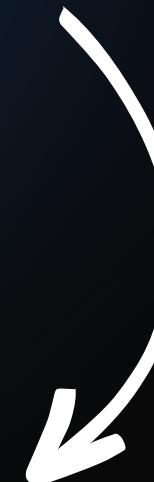
INCOME
INCREASING



Most required and valuable trip type



Airline competition



RESEARCH MOTIVATION

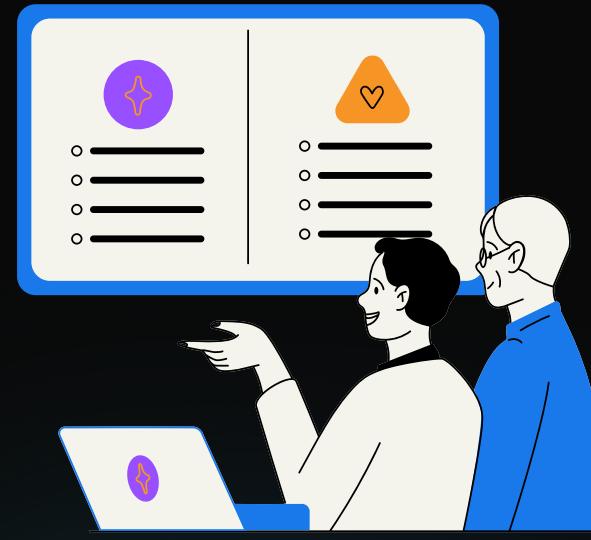
4 STEP SCHEMA:



INCOME
INCREASING



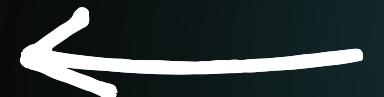
Most required and valuable trip type



Airline competition



Passenger satisfaction



RESEARCH MOTIVATION

4 STEP SCHEMA:



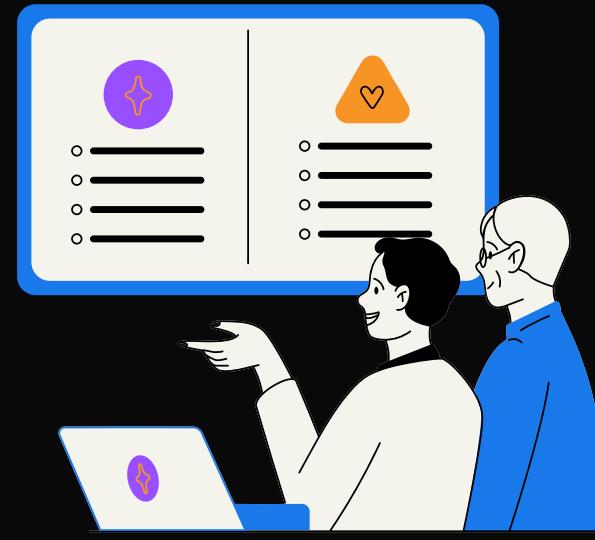
INCOME
INCREASING



Most required and valuable trip type



Service quality



Airline competition



Passenger satisfaction



RESEARCH QUESTIONS

& BUSINESS **BENEFITS**



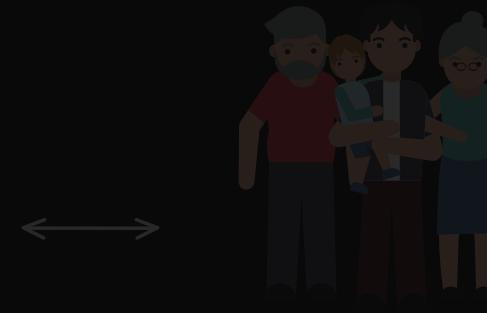
Satisfaction level



Age group



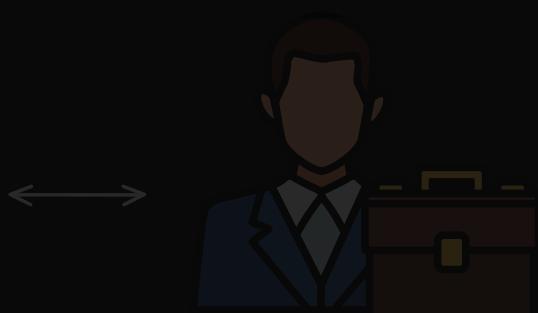
Factor



Age group



Satisfaction level



Type of travel



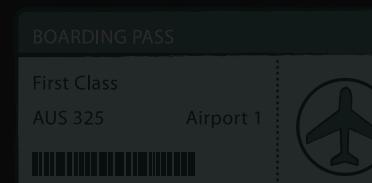
Seat comfort



Flight distance



Satisfaction level



Class

RESEARCH QUESTIONS

& BUSINESS **BENEFITS**



Satisfaction level



Age group



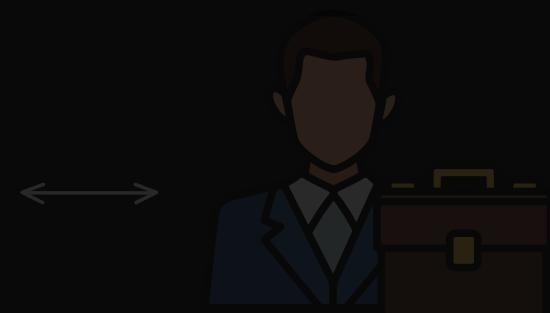
Factor



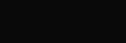
Age group



Satisfaction level



Type of travel



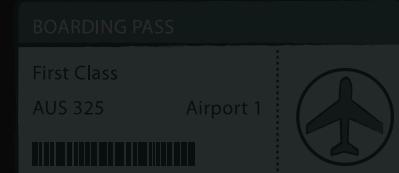
Seat comfort



Flight distance



Satisfaction level



Class



RESEARCH QUESTIONS

& BUSINESS **BENEFITS**



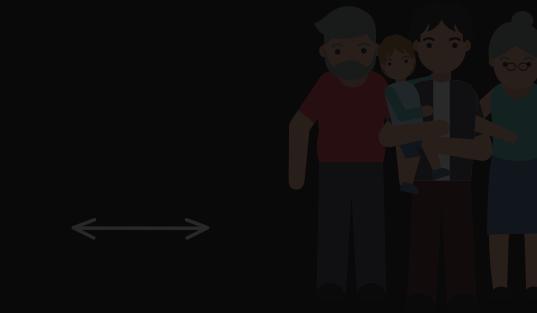
Satisfaction level



Age group



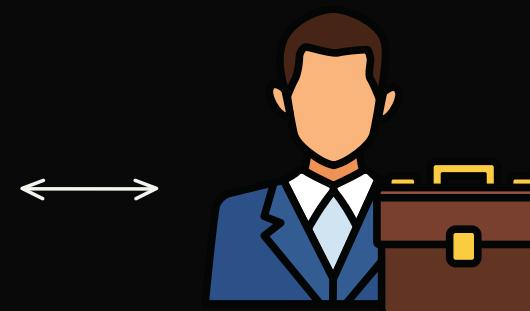
Factor



Age group



Satisfaction level



Type of travel



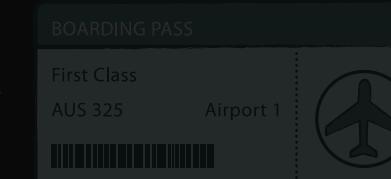
Seat comfort



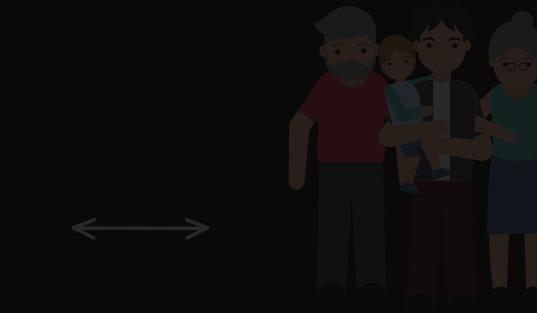
Flight distance



Satisfaction level



Class



Age group

RESEARCH QUESTIONS

& BUSINESS **BENEFITS**



Satisfaction level



Age group



Factor



Age group



Satisfaction level



Type of travel



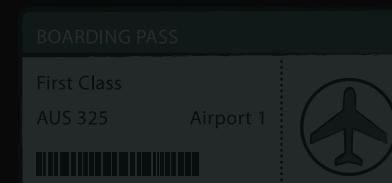
Seat comfort



Flight distance



Satisfaction level



Class

RESEARCH QUESTIONS

& BUSINESS **BENEFITS**



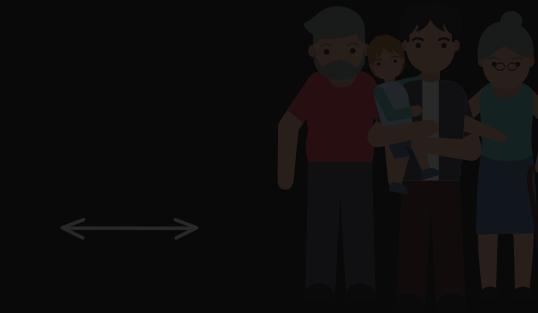
Satisfaction level



Age group



Factor



Age group



Satisfaction level



Type of travel



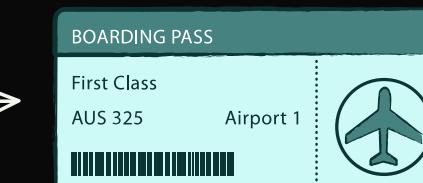
Seat comfort



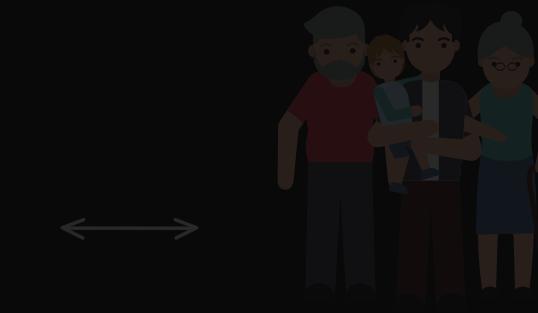
Flight distance



Satisfaction level



Class



Age group

DATA PREPARING

1. PREPROCESSING & CLEANING

1.1 Concatenation

1.2 Label encoding

2. HANDLING MISSING DATA

2.1 Arrival Delay

3. OUTLIERS DETECTION

3.1 Departure/Arrival delay

3.2 Age

3.3 Flight Distance



DATA SOURCE & PROPERTIES



SIZE	129.880 rows
FEATURES	23
NOMINAL	4
NUMERICAL	4
ORDINAL	14
N/A	393 rows
FEATURES N/A	1
OUTLIERS	33
SOURCE	KAGGLE.COM

DATA PREPARING

1. PREPROCESSING & CLEANING

1.1 Concatenation

1.2 Label encoding

2. HANDLING MISSING DATA

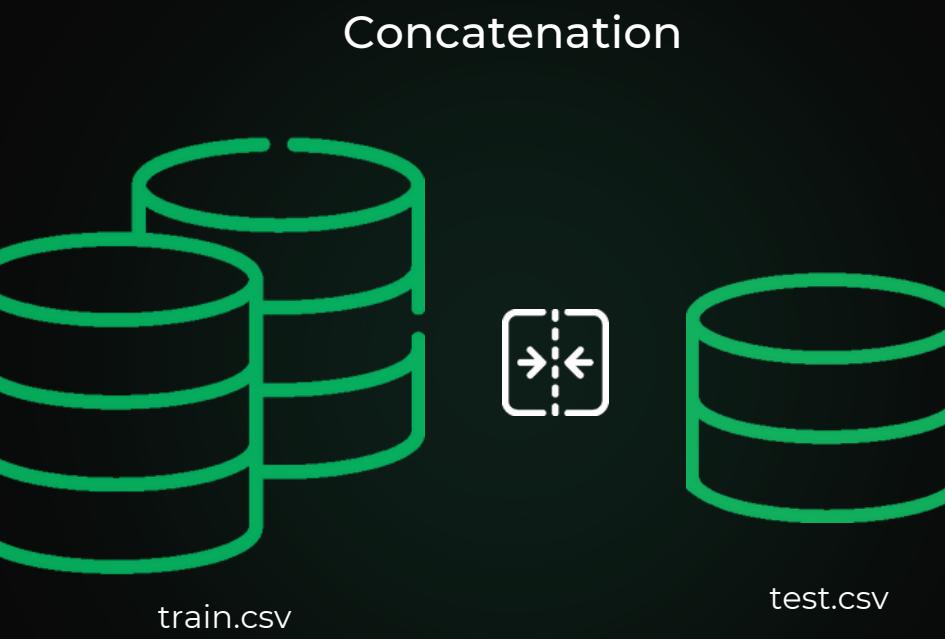
2.1 Arrival Delay

3. OUTLIERS DETECTION

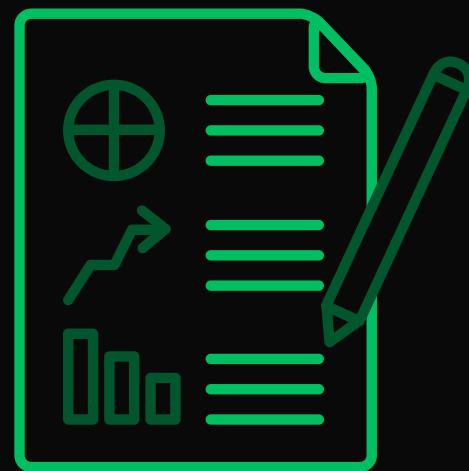
3.1 Flight Distance

3.2 Age

3.3 Departure/Arrival delay



DATA SOURCE & PROPERTIES



SIZE	129.880 rows
FEATURES	23
NOMINAL	4
NUMERICAL	4
ORDINAL	14
N/A	393 rows
FEATURES N/A	1
OUTLIERS	33
SOURCE	KAGGLE.COM

DATA PREPARING

1. PREPROCESSING & CLEANING

1.1 Concatenation

1.2 Label encoding

2. HANDLING MISSING DATA

2.1 Arrival Delay

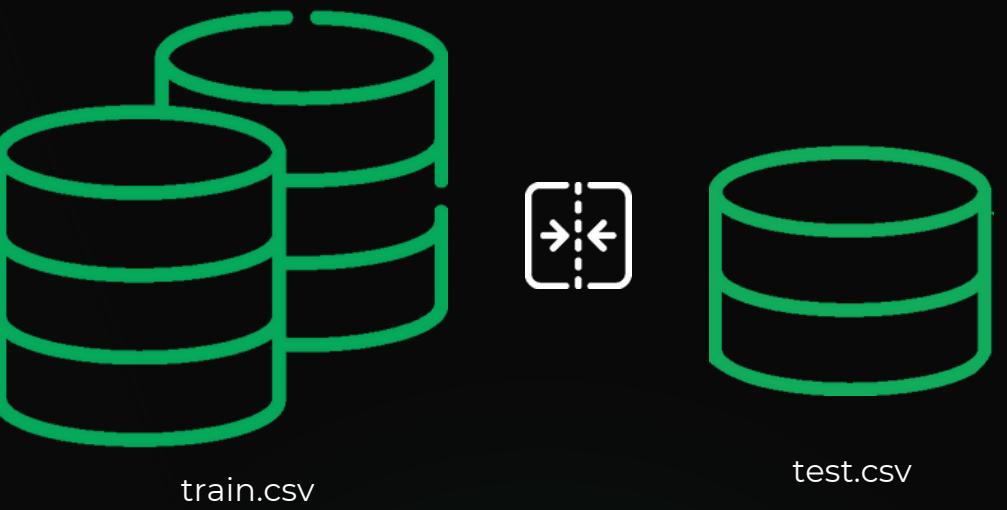
3. OUTLIERS DETECTION

3.1 Flight Distance

3.2 Age

3.3 Departure/Arrival delay

Concatenation



Label encoding

SATISFACTION

neutral or dissatisfied > 0

satisfied > 1

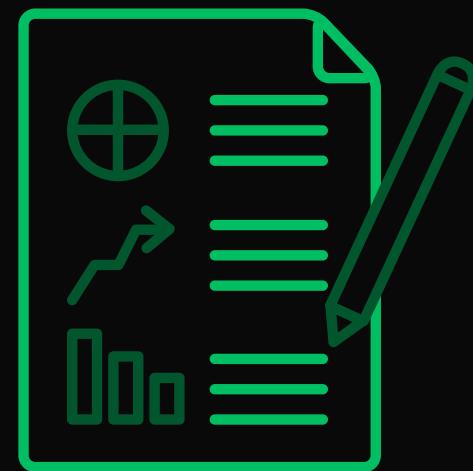
TYPE OF TRAVEL

Business travel > *Business Travel*

CUSTOMER TYPE

disloyal Customer > *Disloyal Customer*

DATA SOURCE & PROPERTIES



SIZE	129.880 rows
FEATURES	23
NOMINAL	4
NUMERICAL	4
ORDINAL	14
N/A	393 rows
FEATURES N/A	1
OUTLIERS	33
SOURCE	KAGGLE.COM

DATA PREPARING

1. PREPROCESSING & CLEANING

1.1 Concatenation

1.2 Label encoding

2. HANDLING MISSING DATA

2.1 [Arrival Delay](#)

3. OUTLIERS DETECTION

3.1 Flight Distance

3.2 Age

3.3 Departure/Arrival delay

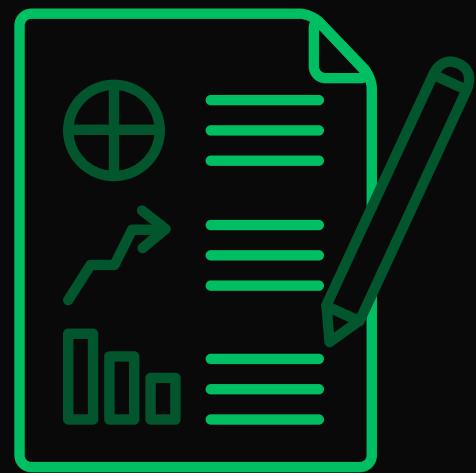
Arrival Delay



393 missing rows
0.3 %

Median: 7min

DATA SOURCE & PROPERTIES



SIZE	129.880 rows
FEATURES	23
NOMINAL	4
NUMERICAL	4
ORDINAL	14
N/A	393 rows
FEATURES N/A	1
OUTLIERS	33
SOURCE	KAGGLE.COM

DATA PREPARING

1. PREPROCESSING & CLEANING

1.1 Concatenation

1.2 Label encoding

2. HANDLING MISSING DATA

2.1 [Arrival Delay](#)

3. OUTLIERS DETECTION

3.1 Flight Distance

3.2 Age

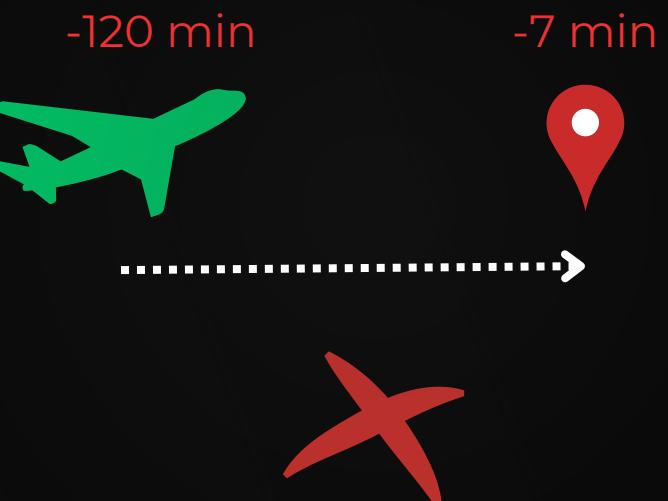
3.3 Departure/Arrival delay

Arrival Delay

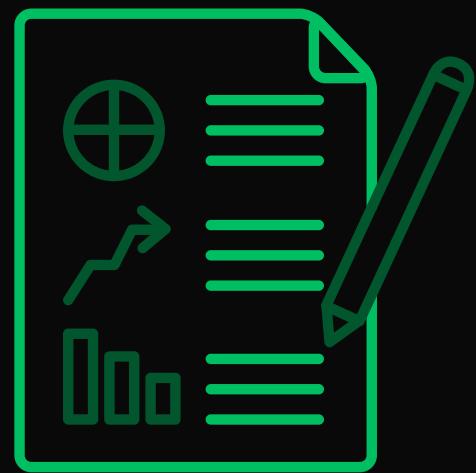


393 missing rows
0.3 %

Median: 7min



DATA SOURCE & PROPERTIES



SIZE	129.880 rows
FEATURES	23
NOMINAL	4
NUMERICAL	4
ORDINAL	14
N/A	393 rows
FEATURES N/A	1
OUTLIERS	33
SOURCE	KAGGLE.COM

DATA PREPARING

1. PREPROCESSING & CLEANING

1.1 Concatenation

1.2 Label encoding

2. HANDLING MISSING DATA

2.1 [Arrival Delay](#)

3. OUTLIERS DETECTION

3.1 Flight Distance

3.2 Age

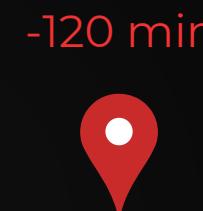
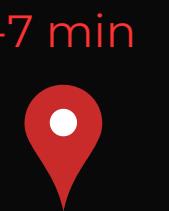
3.3 Departure/Arrival delay

Arrival Delay

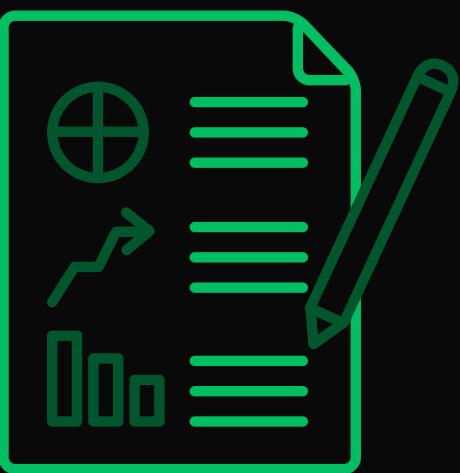


393 missing rows
0.3 %

Median: 7min



DATA SOURCE & PROPERTIES



SIZE	129.880 rows
FEATURES	23
NOMINAL	4
NUMERICAL	4
ORDINAL	14
N/A	393 rows
FEATURES N/A	1
OUTLIERS	33
SOURCE	KAGGLE.COM

DATA PREPARING

1. PREPROCESSING & CLEANING

1.1 Concatenation

1.2 Label encoding

2. HANDLING MISSING DATA

2.1 Arrival Delay

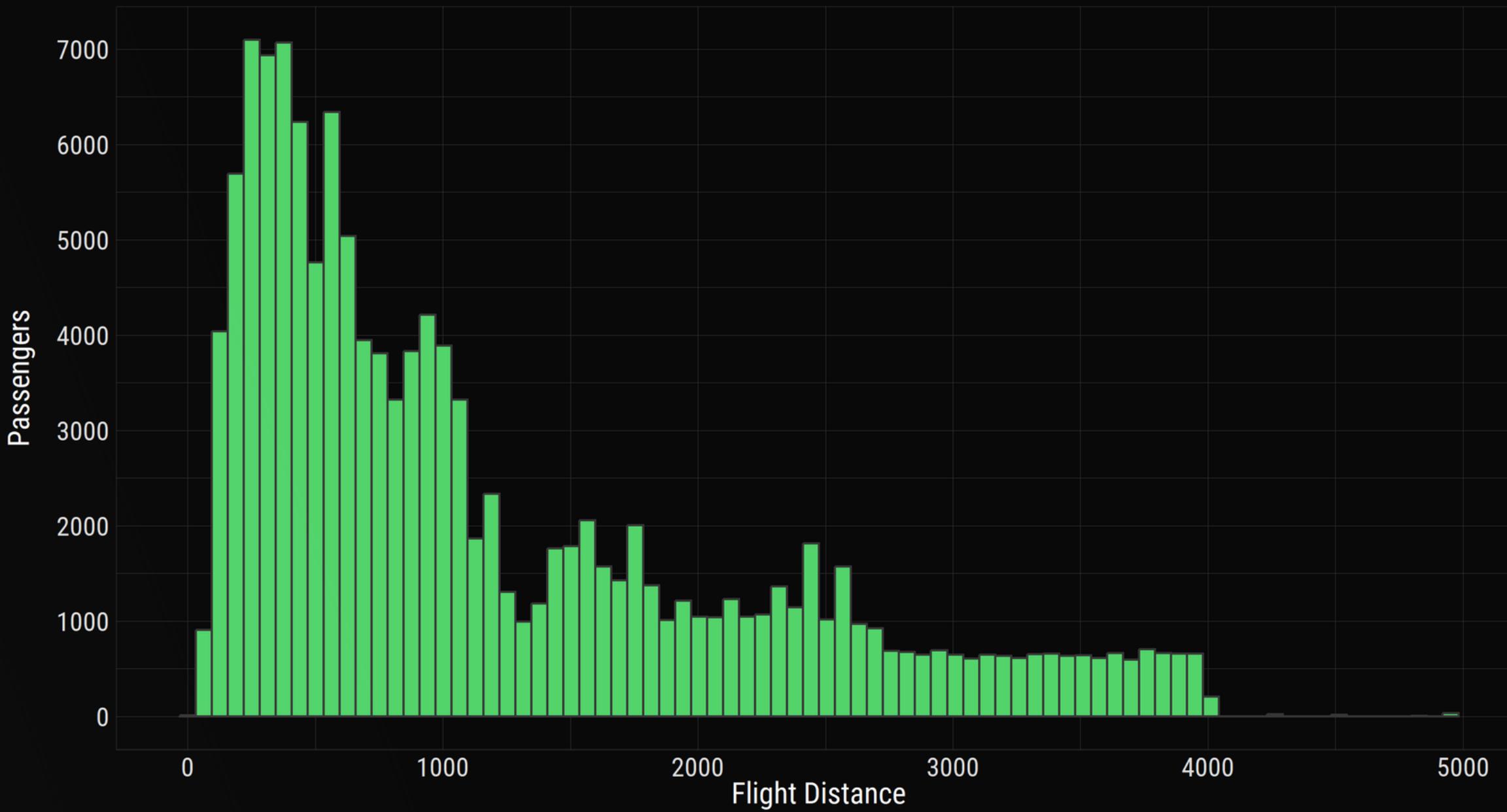
3. OUTLIERS DETECTION

3.1 Flight Distance

3.2 Age

3.3 Departure/Arrival delay

Approximately exponential distribution



DATA PREPARING

1. PREPROCESSING & CLEANING

1.1 Concatenation

1.2 Label encoding

2. HANDLING MISSING DATA

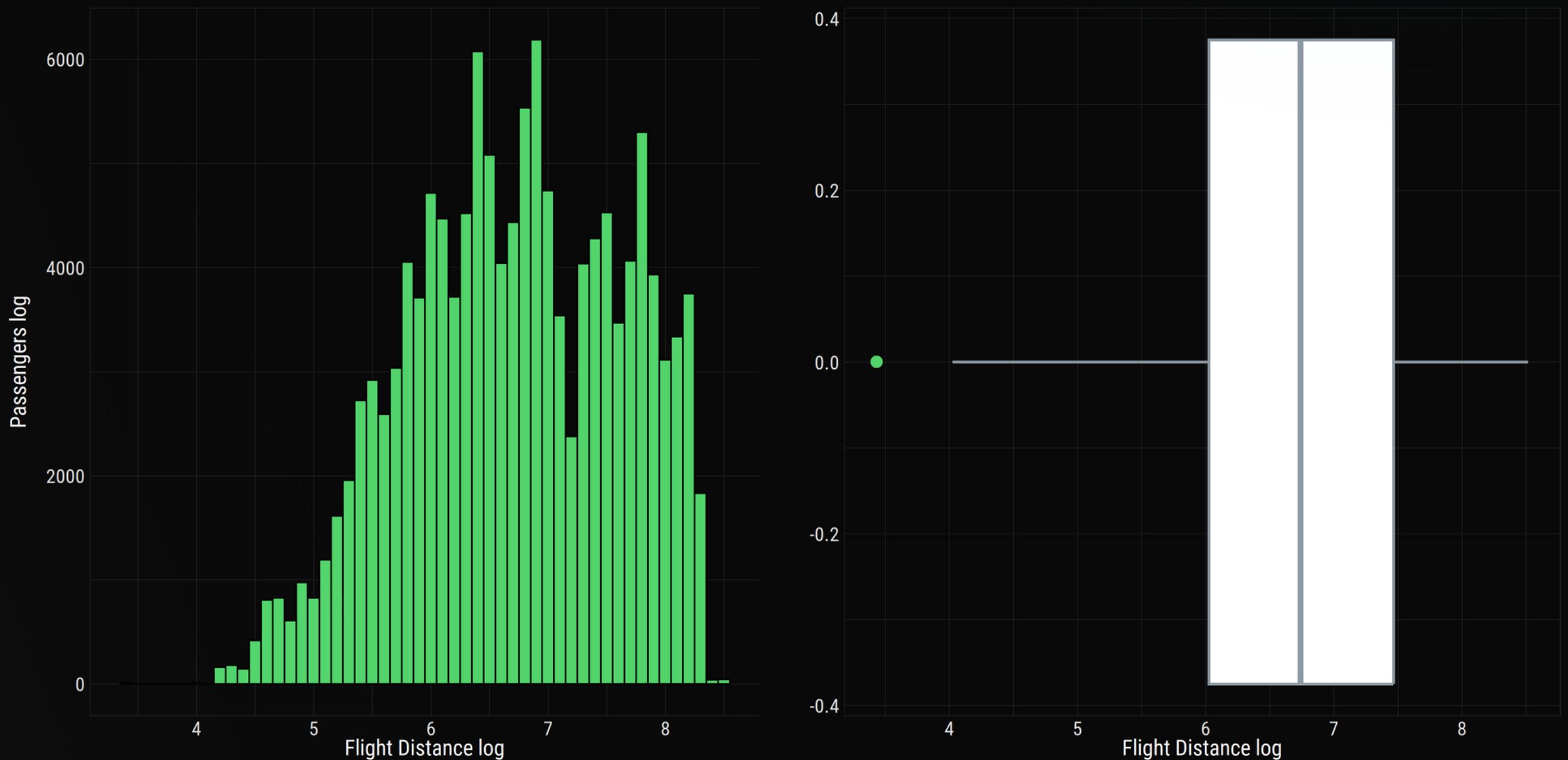
2.1 Arrival Delay

3. OUTLIERS DETECTION

3.1 Flight Distance

3.2 Age

3.3 Departure/Arrival delay



Q1: 6.025
Q3: 7.464

Lower range: 3.869
Upper range: 9.621

DATA PREPARING

* 11 rows

- ## 1. PREPROCESSING & CLEANING

- ## 1.1 Concatenation

2. HANDLING MISSING DATA

- ## 2.1 Arrival Delay

3. OUTLIERS DETECTION

- ### 3.1 Flight Distance

- ### 3.2 Age

- ### 3.3 Departure/Arrival delay

id	Gender	Customer_Type	Age	Type_of_Travel	Class	Flight Distance
29816	Female	Loyal Customer	38	Business Travel	Eco	3.433987
29824	Female	Disloyal Customer	23	Business Travel	Eco	3.433987
29863	Female	Loyal Customer	53	Business Travel	Eco	3.433987
29992	Female	Disloyal Customer	26	Business Travel	Eco	3.433987
...
30184	Male	Loyal Customer	43	Business Travel	Eco Plus	3.433987

DATA PREPARING

1. PREPROCESSING & CLEANING

1.1 Concatenation

1.2 Label encoding

2. HANDLING MISSING DATA

2.1 Arrival Delay

3. OUTLIERS DETECTION

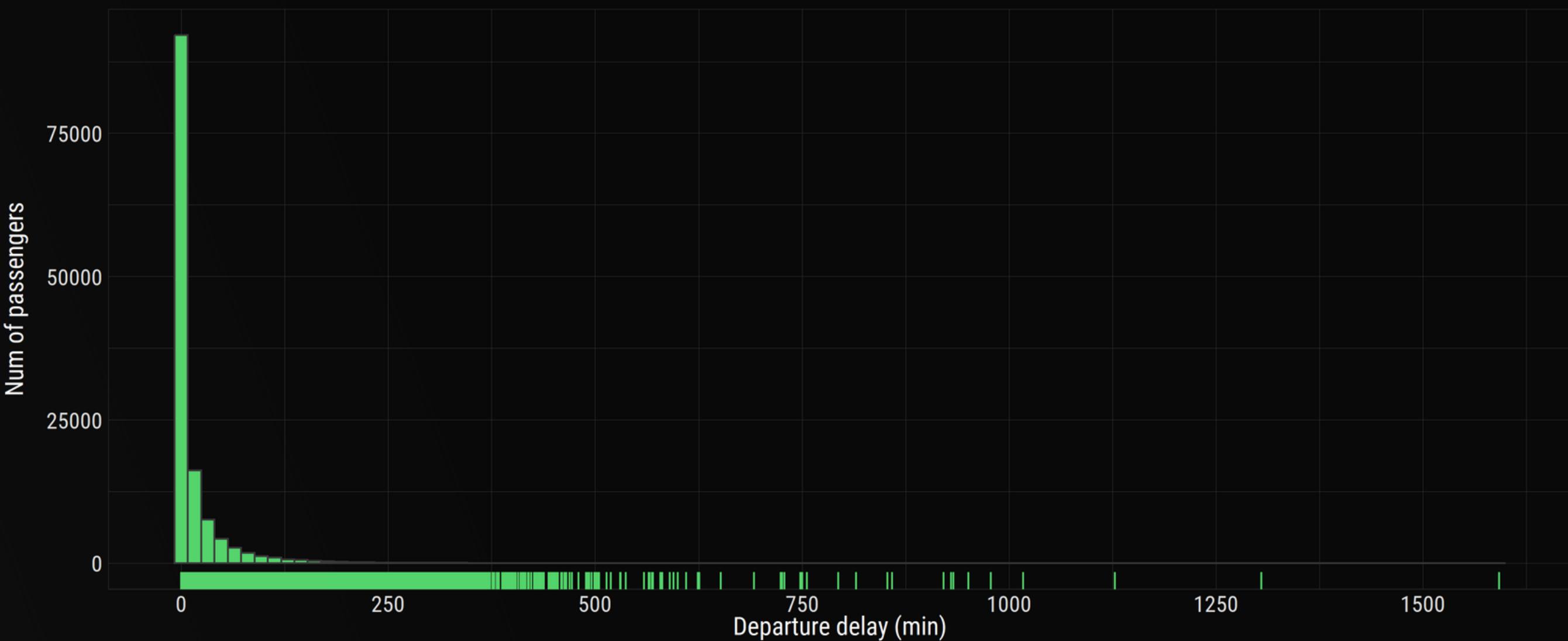
3.1 Flight Distance

3.2 Age

3.3 Departure/Arrival delay

Exponential distribution

Most flights were without delays



SEE MORE IN

CODEBOOK

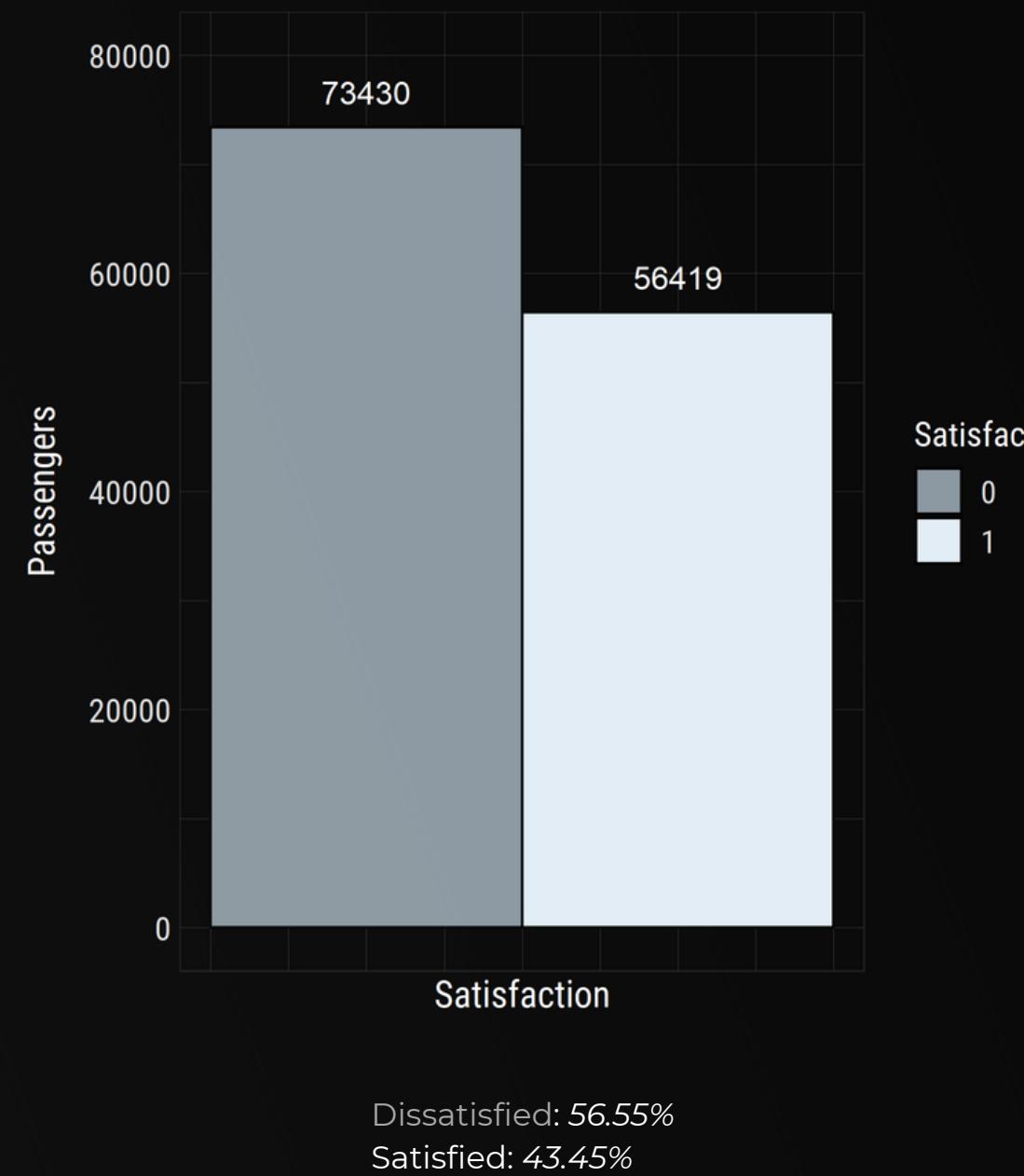
VARIABLE INFORMATION



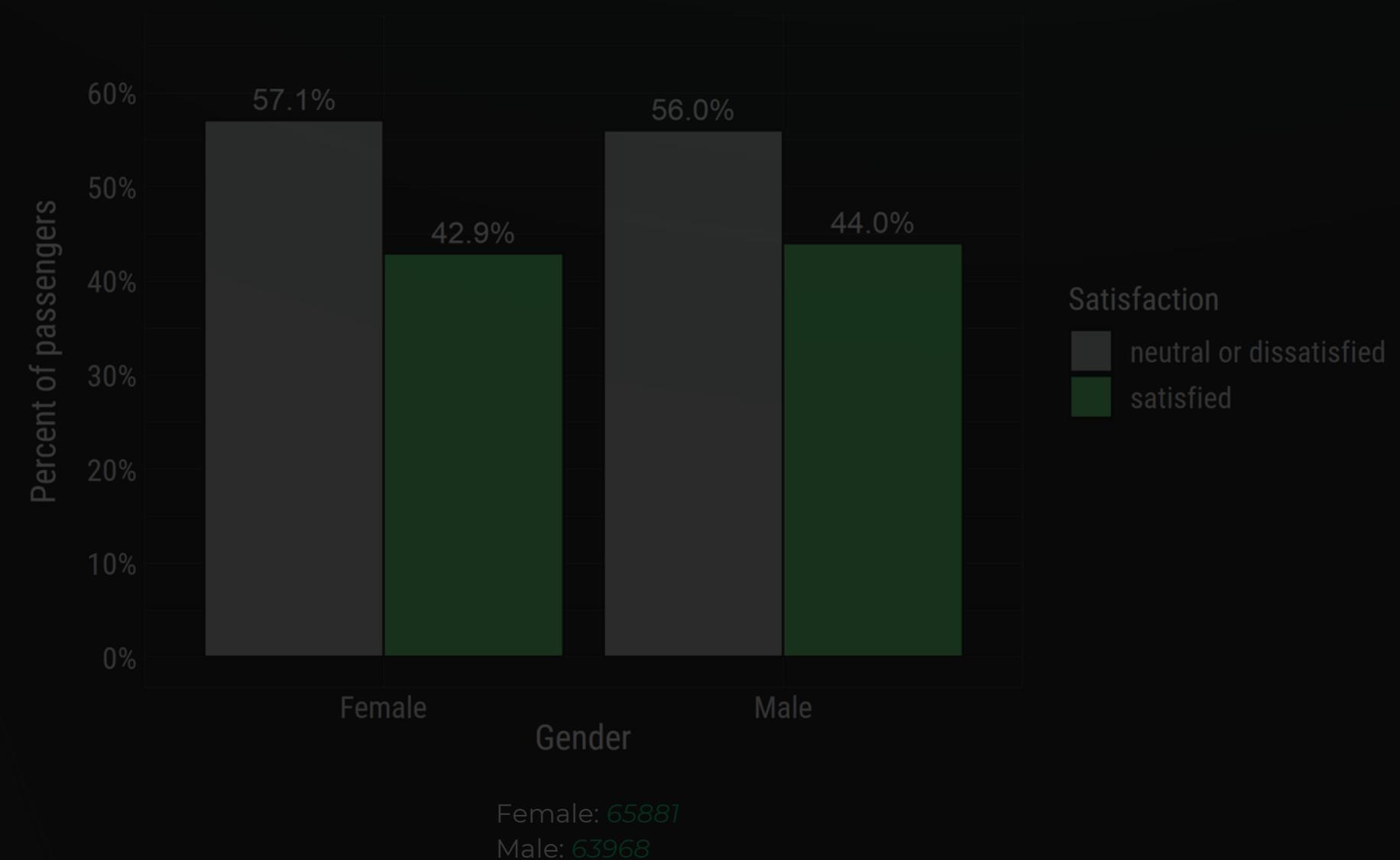
CODEBOOK						
Airline Passenger Satisfaction						
Team Four						
1 Variable Information						
Nº	Variable	Attribute Type	Label	Units of measurement	Details	
1.	id	numerical	id number of the passengers	-	not used in our research	
2.	Satisfaction	binary class label	airline satisfaction level	-	2 class labels: "dissatisfied" and "satisfied"	
3.	Gender	nominal	gender of the passengers	-	2 terms: "female" and "male"	
4.	Customer type	nominal	the customer type	-	2 terms: "loyal customer" and "disloyal customer"	
5.	Age	numerical	the actual age of the passengers	year	integer numbers from 7 to 85	
6.	Type of travel	nominal	purpose of the flight of the passengers	-	2 terms: "personal travel" and "business travel"	
7.	Class	nominal	travel class in the plane of the passengers	-		
8.	Flight distance	ordinal	the flight distance of this journey	-		
9.	Inflight Wi-Fi service convenient	ordinal	-	-		
10.	Departure/Arrival time convenient	ordinal	-	-		

EDA: Overview

The difference is 17 thousand people

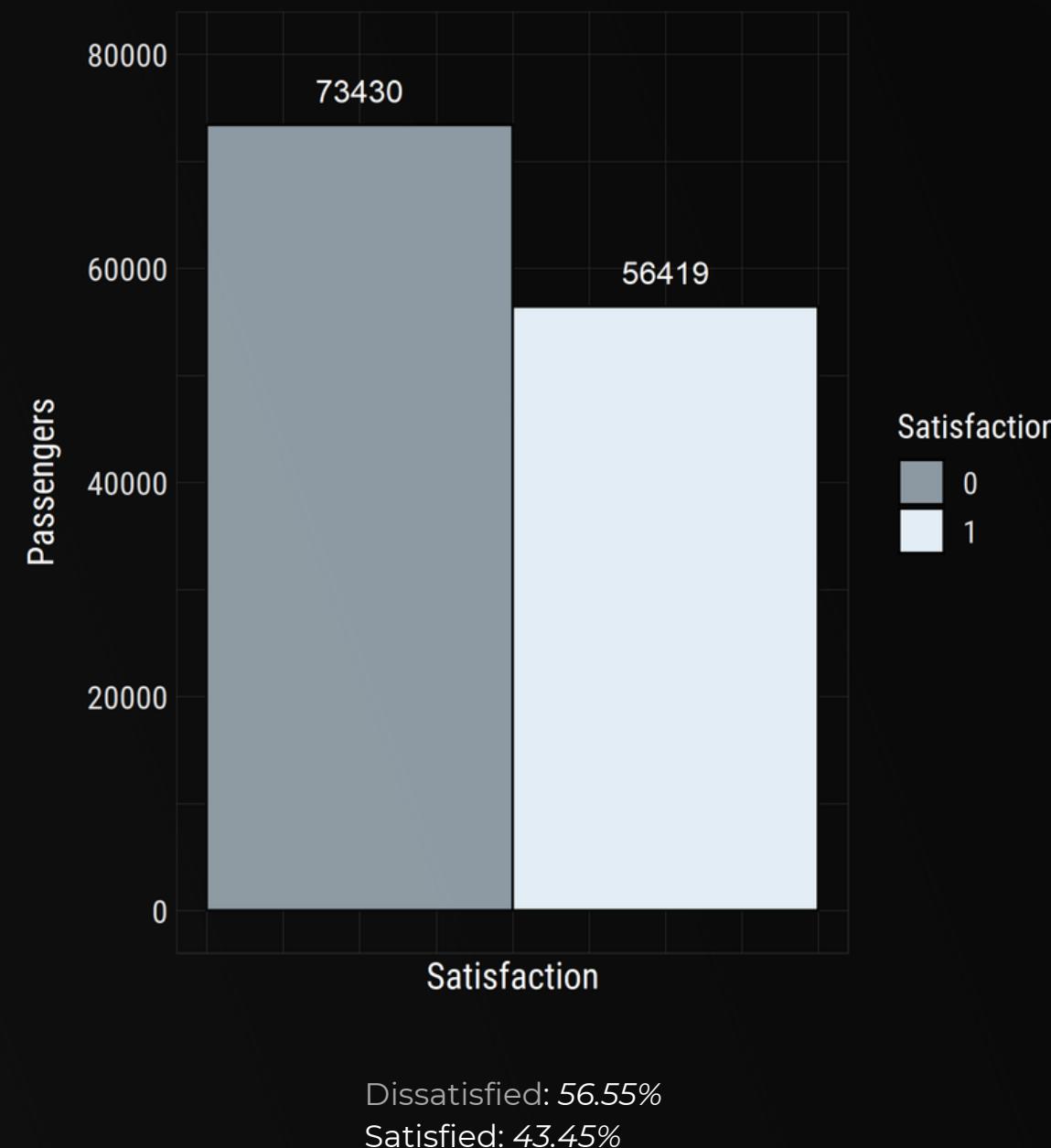


There is no gender problem

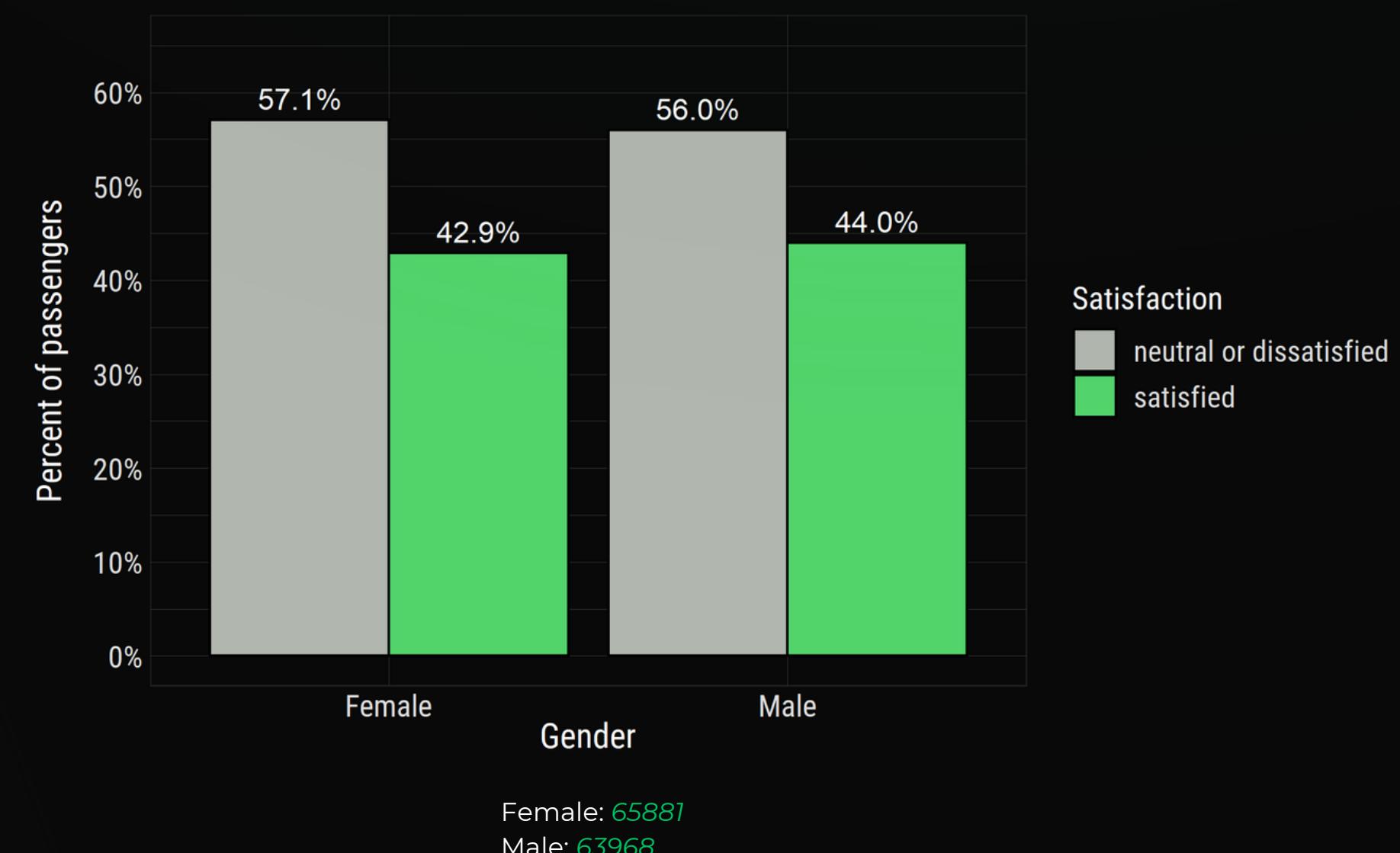


EDA: Overview

The difference is 17 thousand people



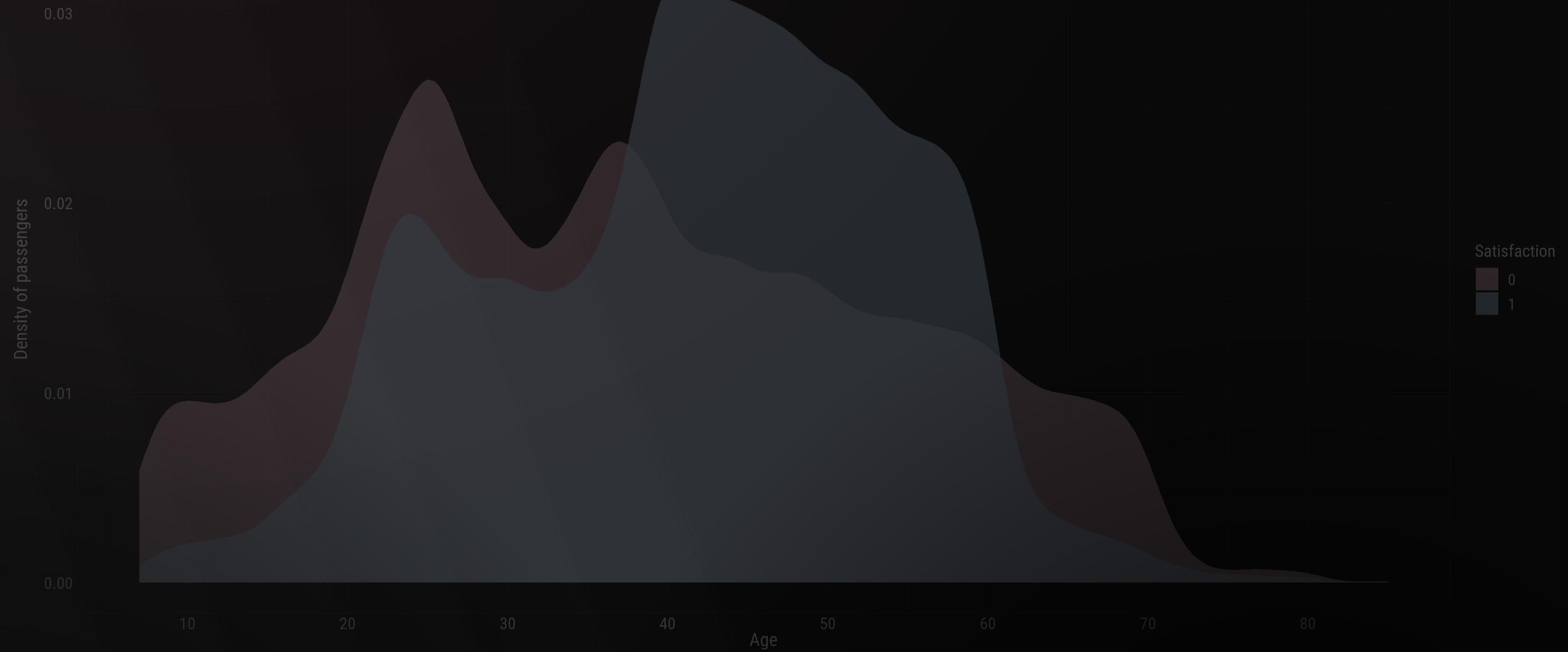
There is no gender problem



Q: Is there a dependence between flight satisfaction and age categories of passengers?

H: There are age categories that have a **lower** level of satisfaction of flights than others

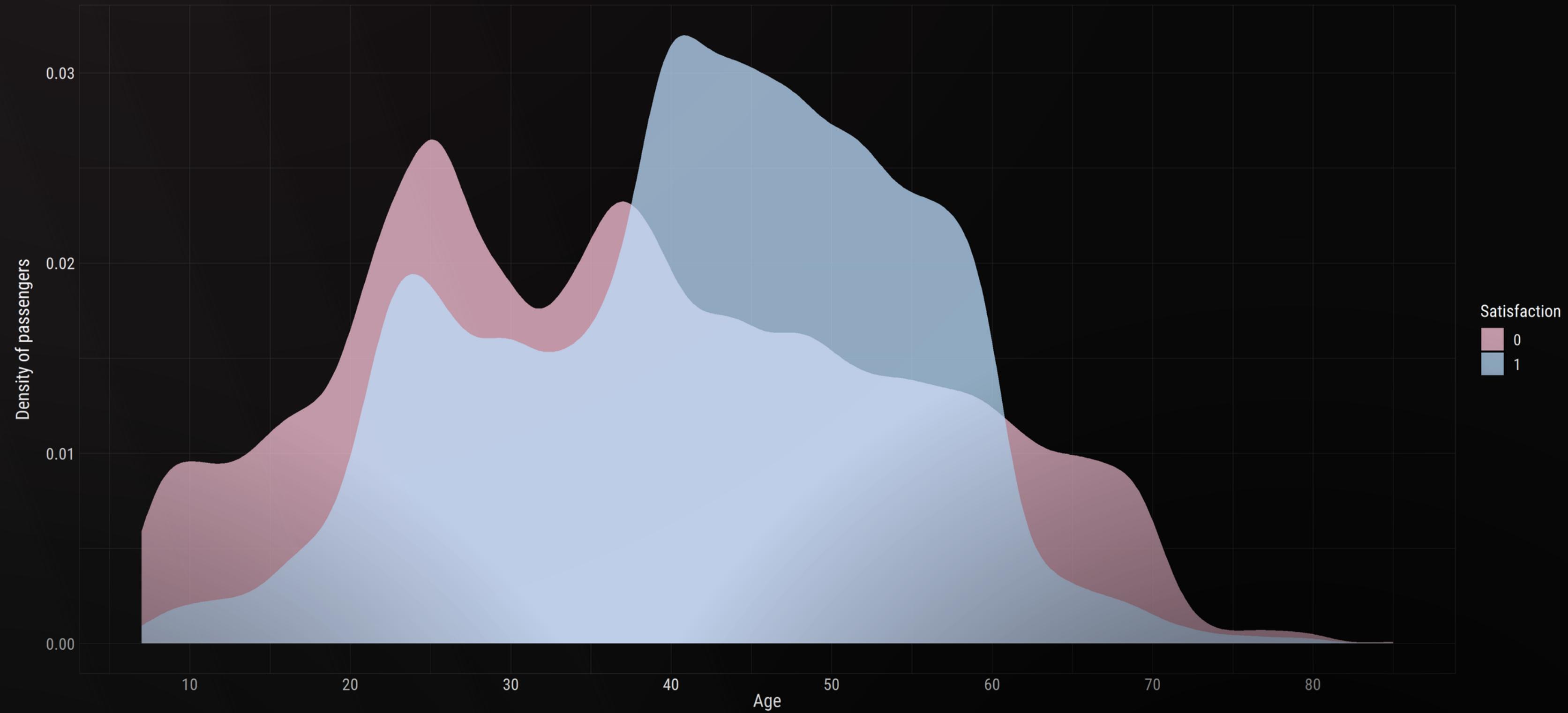
The growth of satisfaction begins after the age of 37



Q: Is there a dependence between flight satisfaction and age categories of passengers?

H: There are age categories that have a lower level of satisfaction of flights than others

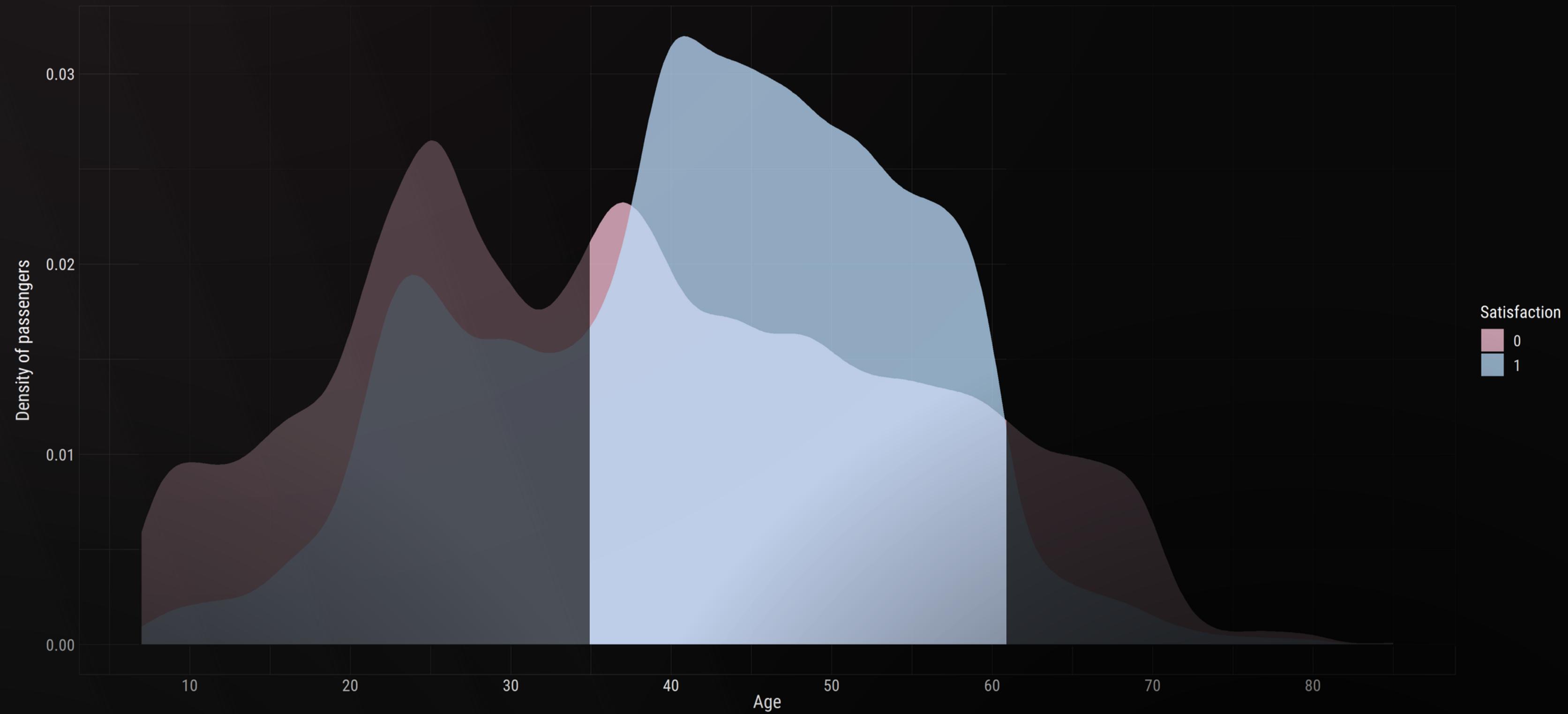
The growth of satisfaction begins after the age of 37



Q: Is there a dependence between flight satisfaction and age categories of passengers?

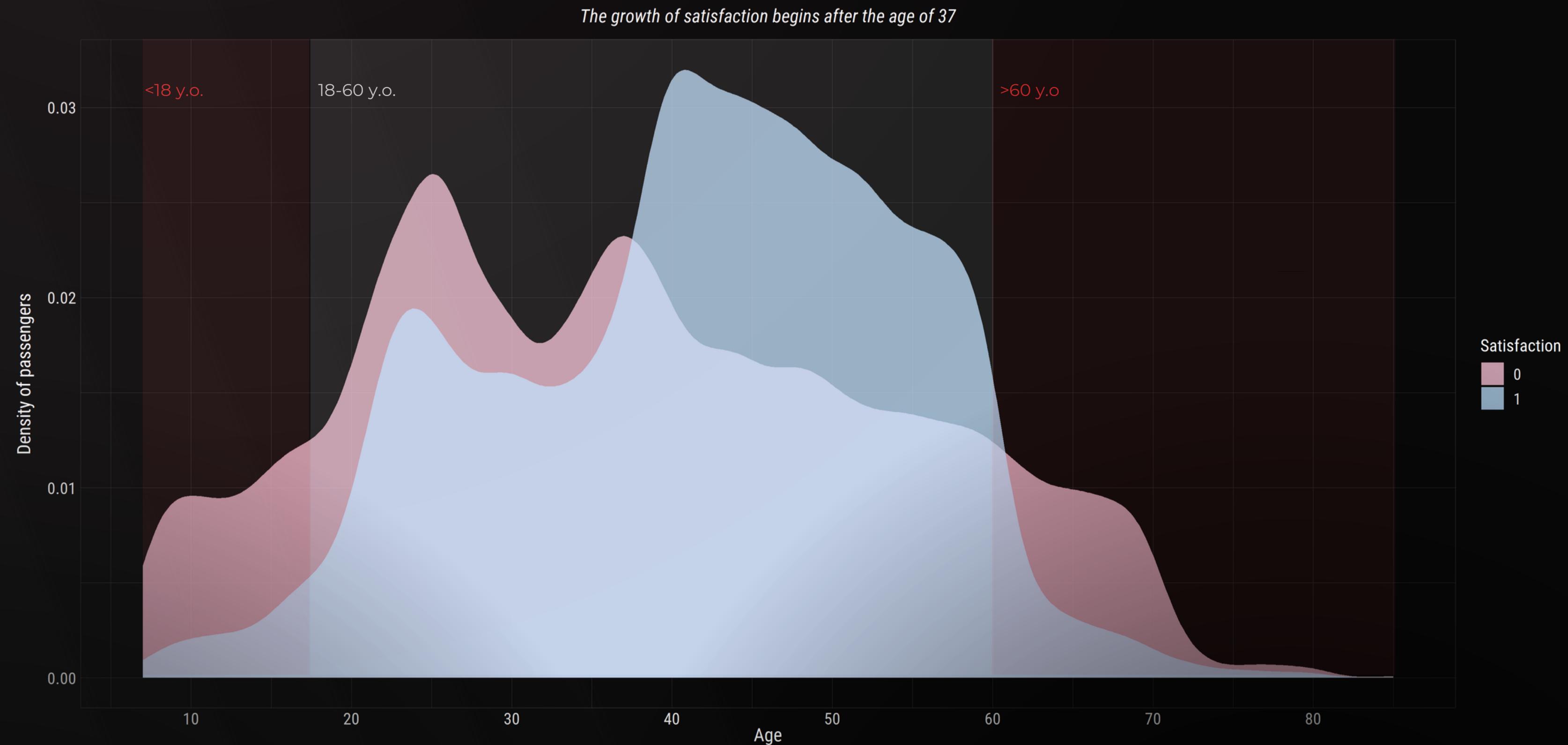
H: There are age categories that have a **lower** level of satisfaction of flights than others

The growth of satisfaction begins after the age of 37



Q: Is there a dependence between flight satisfaction and age categories of passengers?

H: There are age categories that have a lower level of satisfaction of flights than others



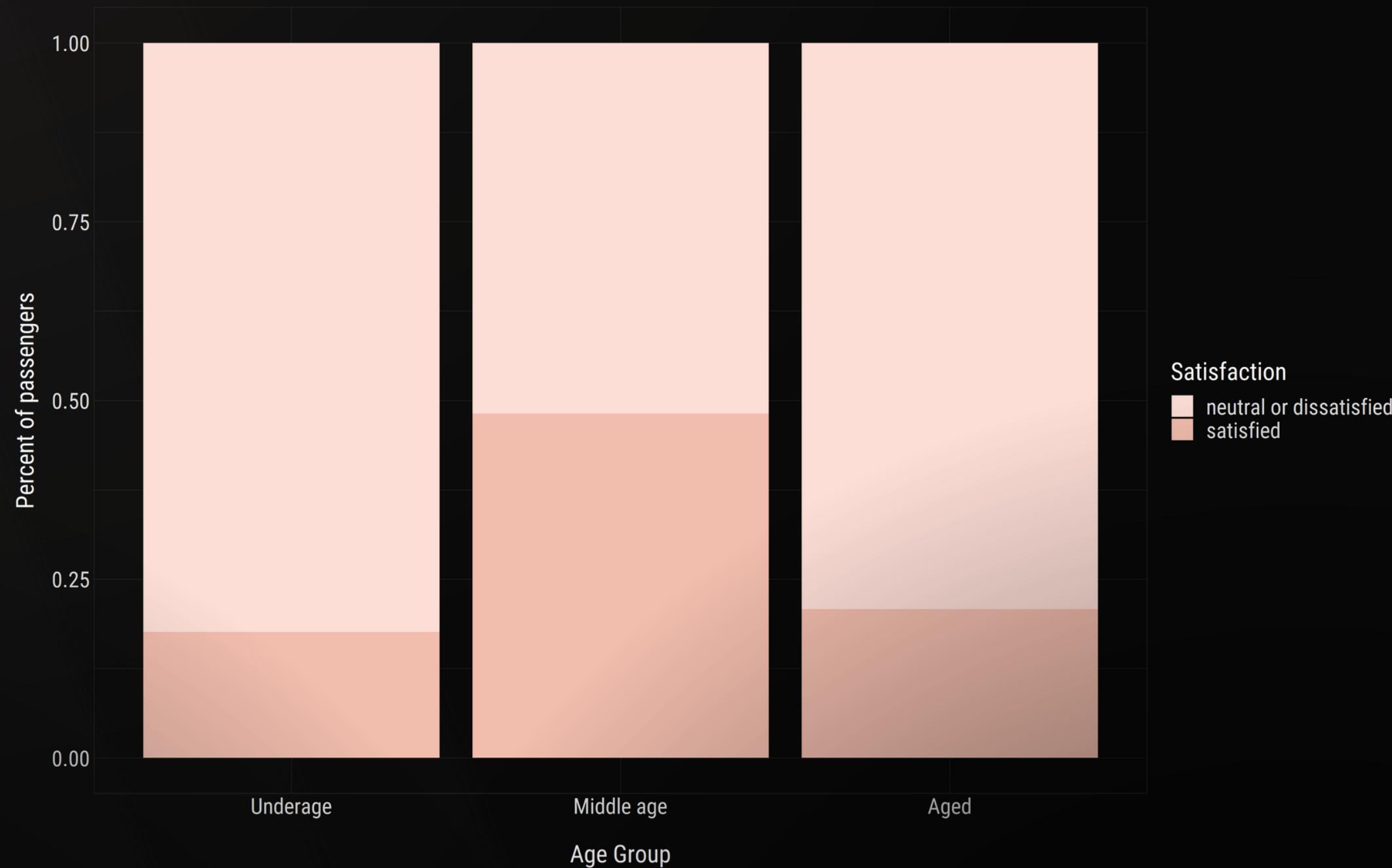
Q: Is there a dependence between flight satisfaction and age categories of passengers?

H: There are age categories that have a lower level of satisfaction of flights than others

There is a difference in the level of satisfaction

Middle age:
Satisfied: 51163
Dissatisfied: 55407
Satisfied / amount of pas.: 48.01%

Underage & Aged:
Satisfied: 5256
Dissatisfied: 18023
Satisfied / amount of pas.: 22.58%



Q: Is there a dependence between flight satisfaction and age categories of passengers?

H: There are age categories that have a lower level of satisfaction of flights than others

There is a difference in the level of satisfaction

Middle age:
Satisfied: 57163
Dissatisfied: 55407
Satisfied / amount of pas.: 48.01%

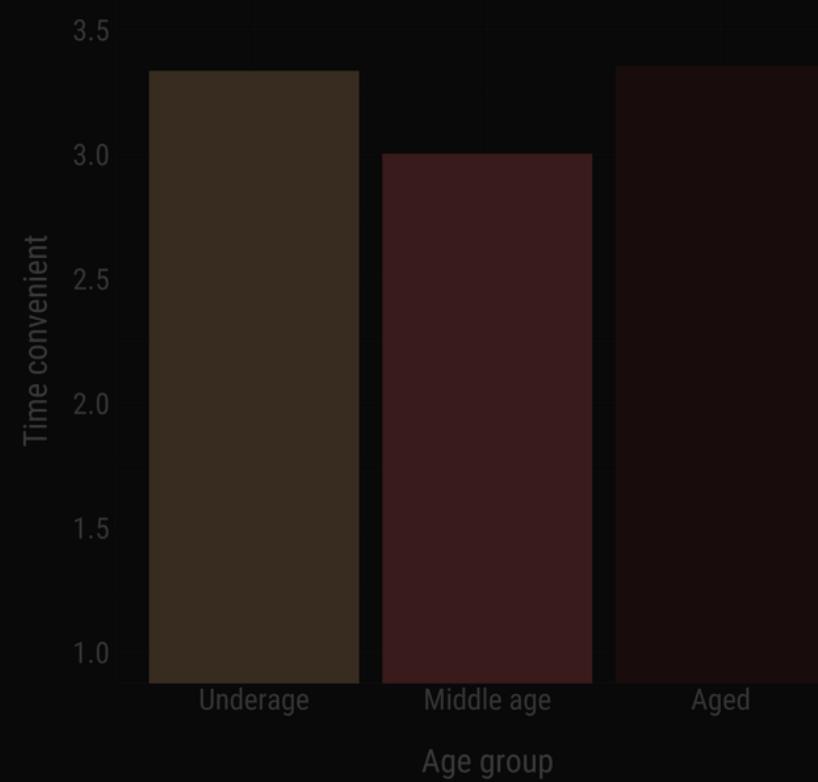
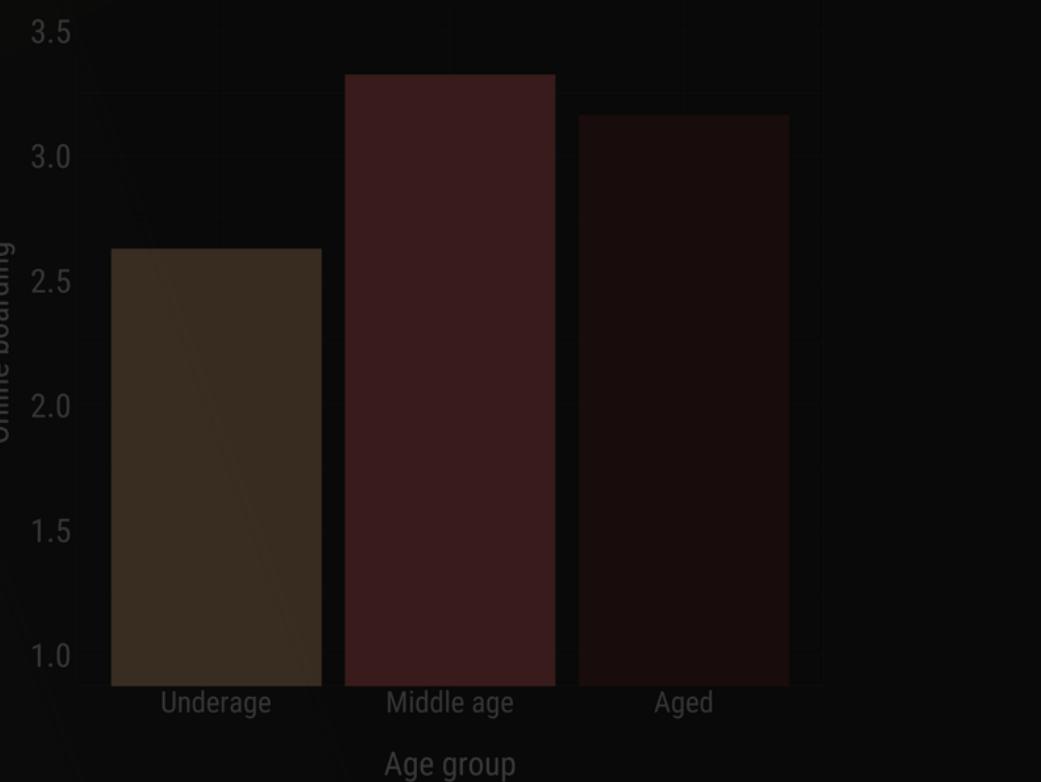
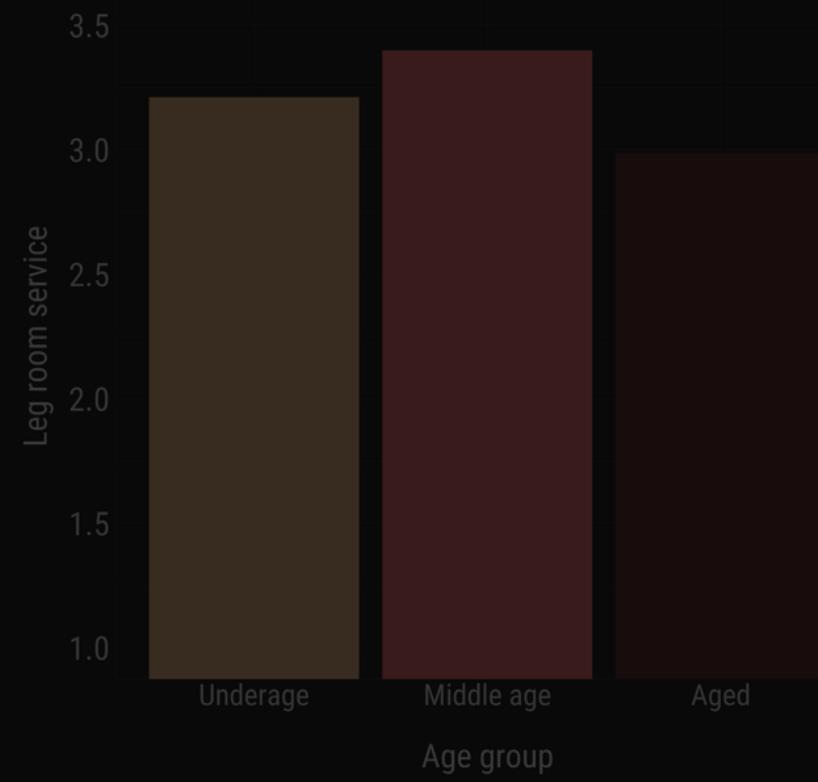
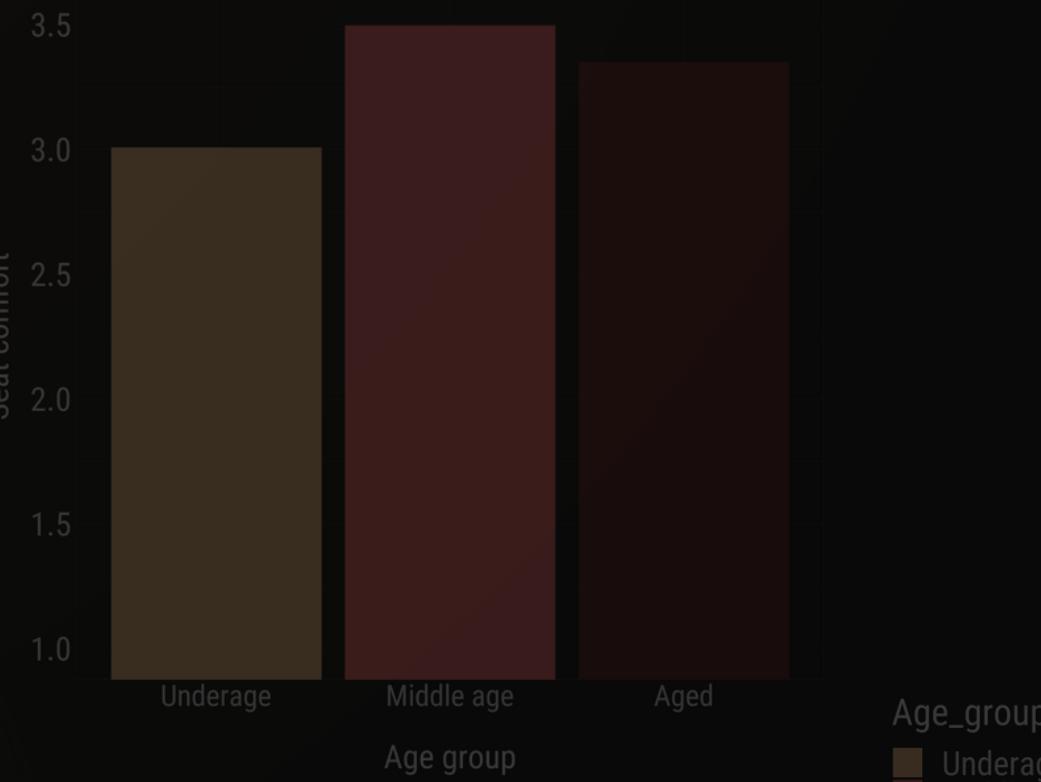
Underage & Aged:
Satisfied: 5256
Dissatisfied: 78023
Satisfied / amount of pas.: 22.58%



Q: Is there a dependence between specific flight satisfaction factors and age categories of passengers?

H: There are age categories with less effect of some satisfaction factors than others.

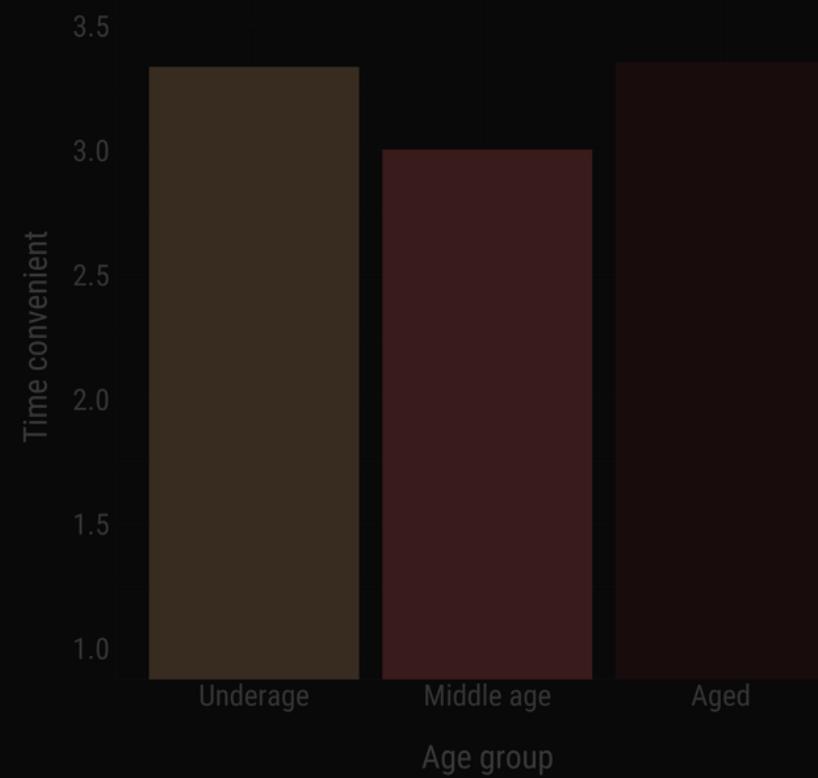
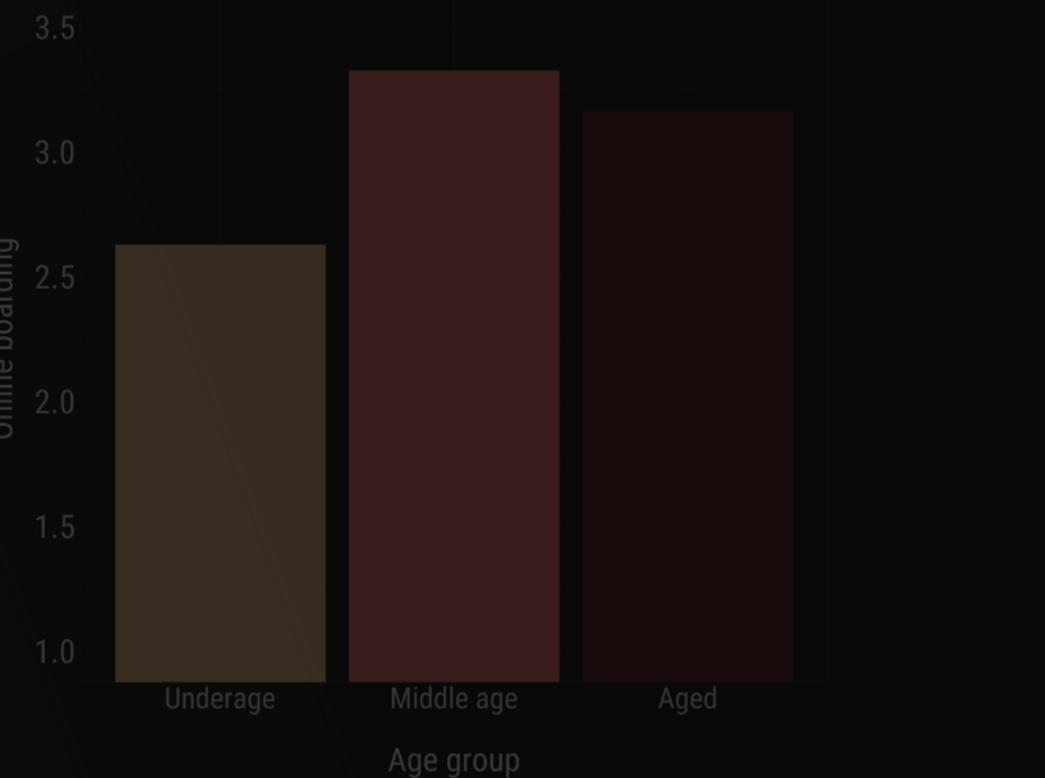
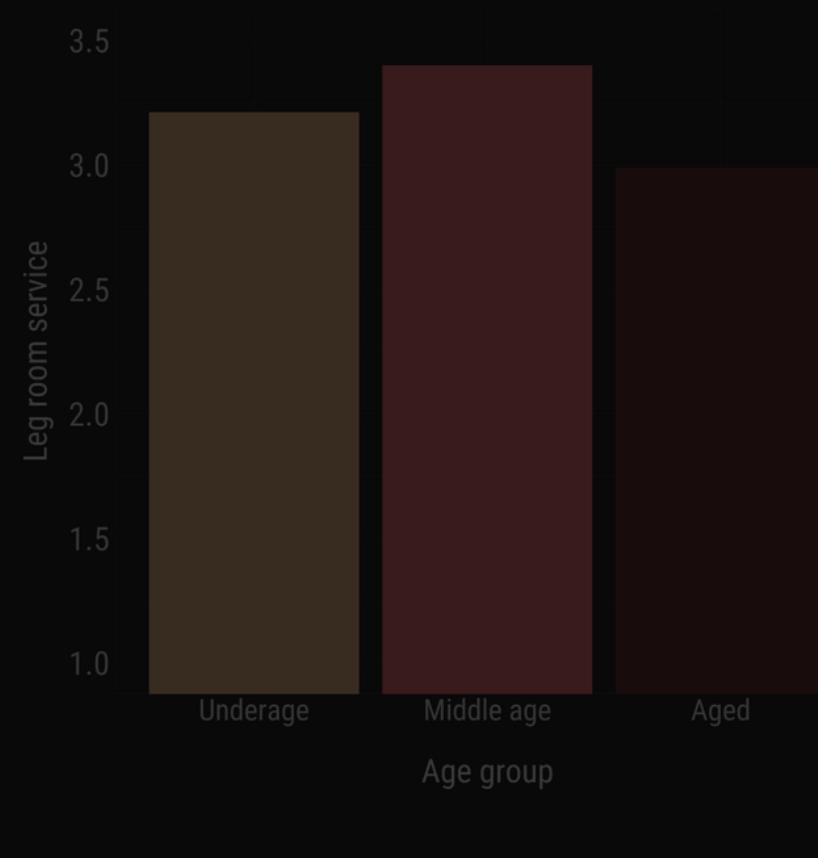
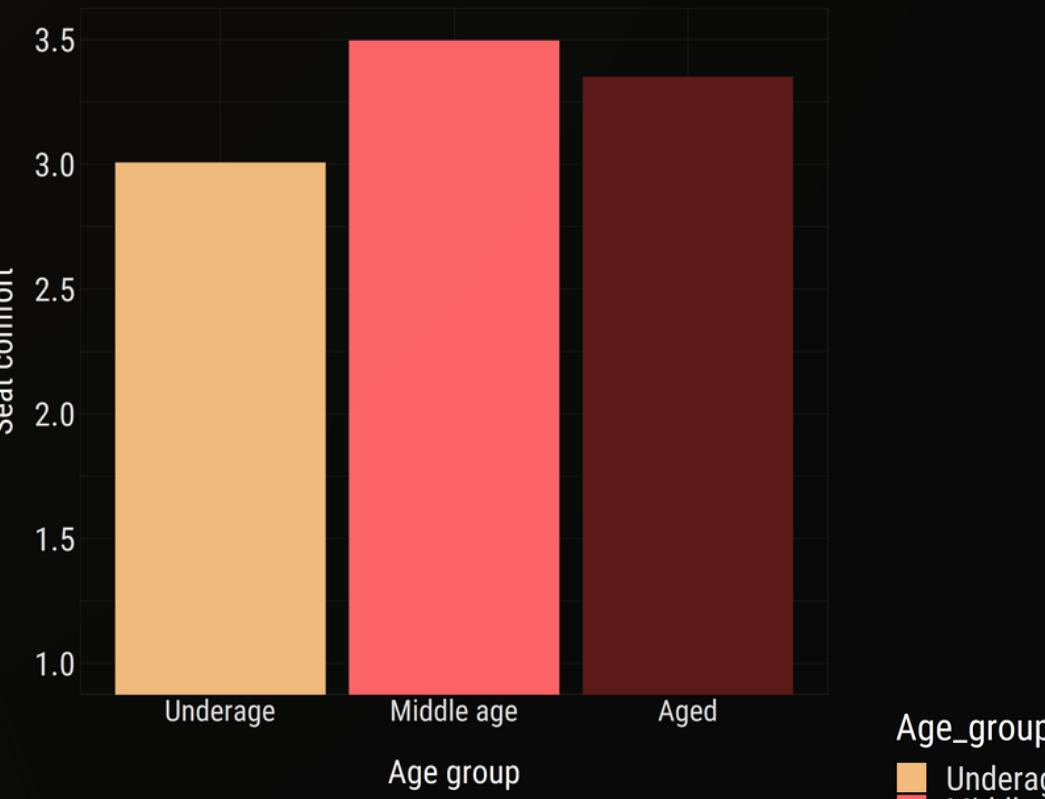
SEAT COMFORT



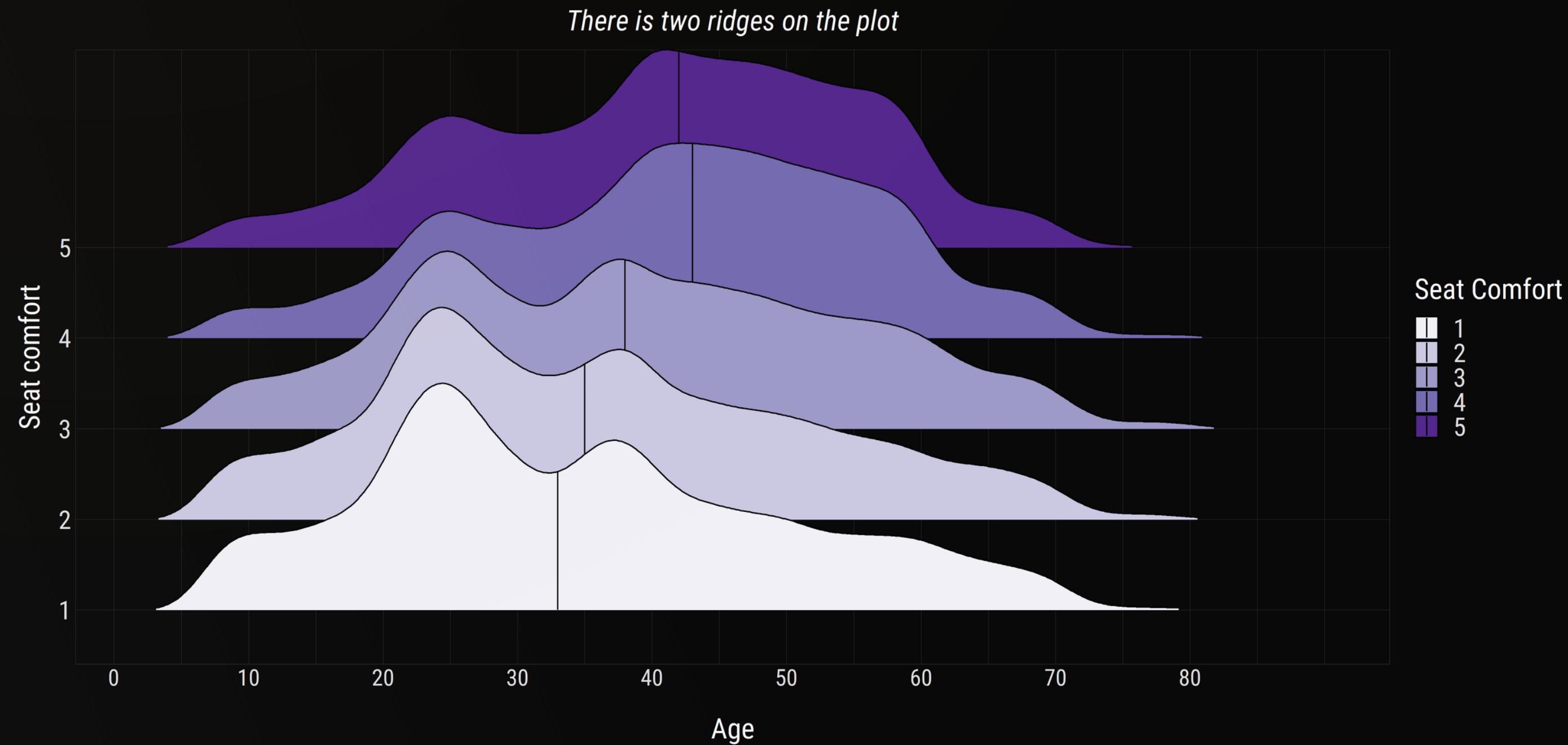
Q: Is there a dependence between specific flight satisfaction factors and age categories of passengers?

H: There are age categories with less effect of some satisfaction factors than others.

SEAT COMFORT



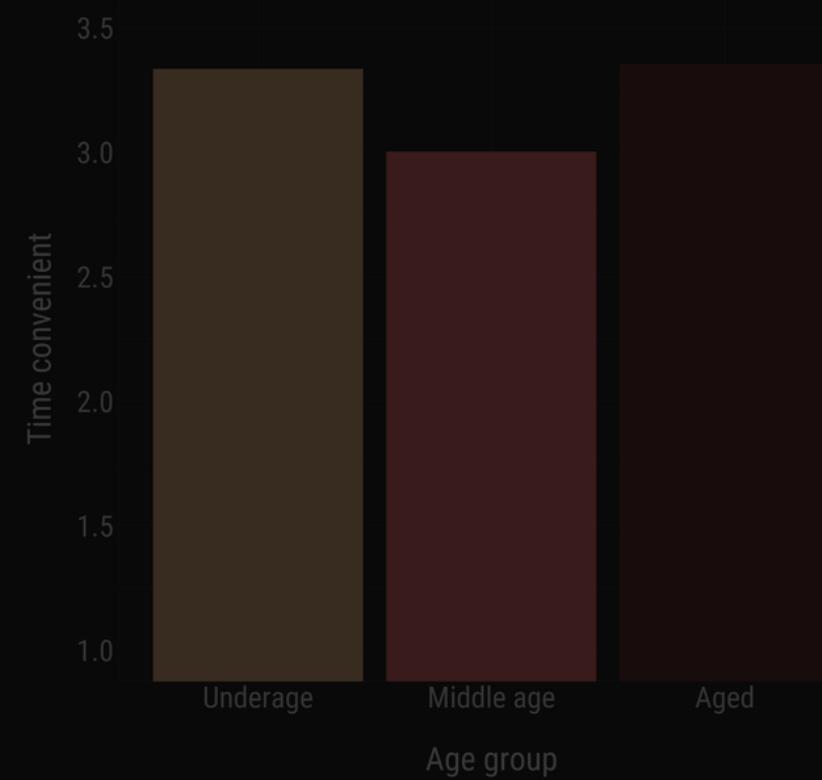
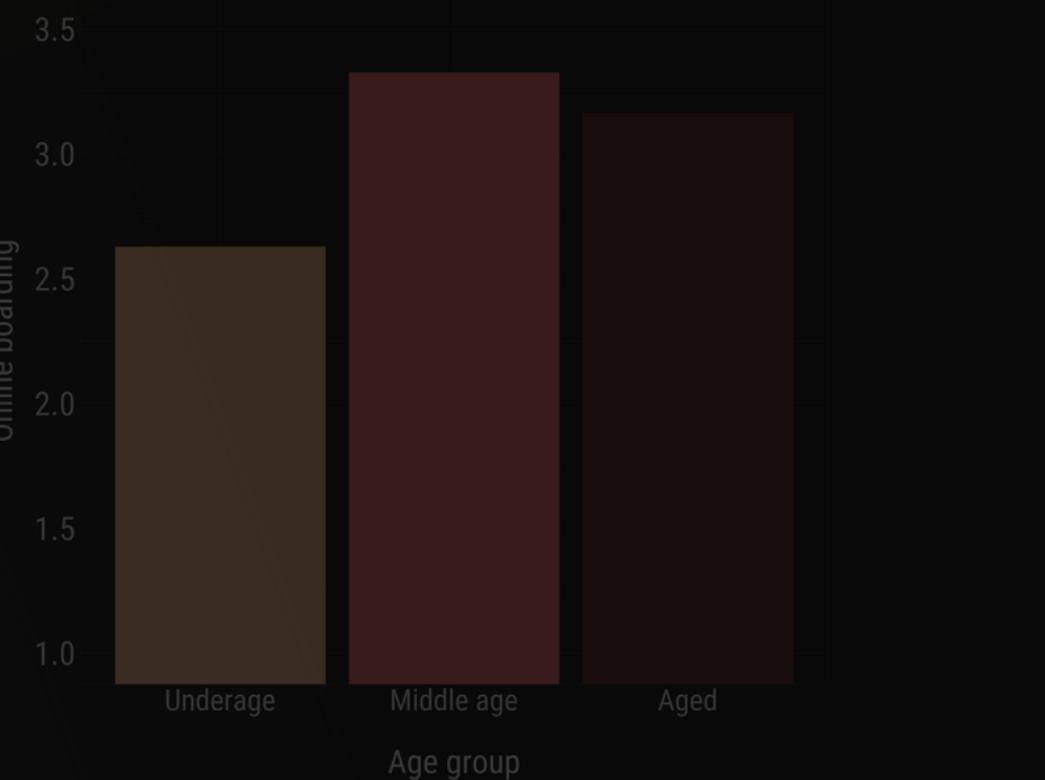
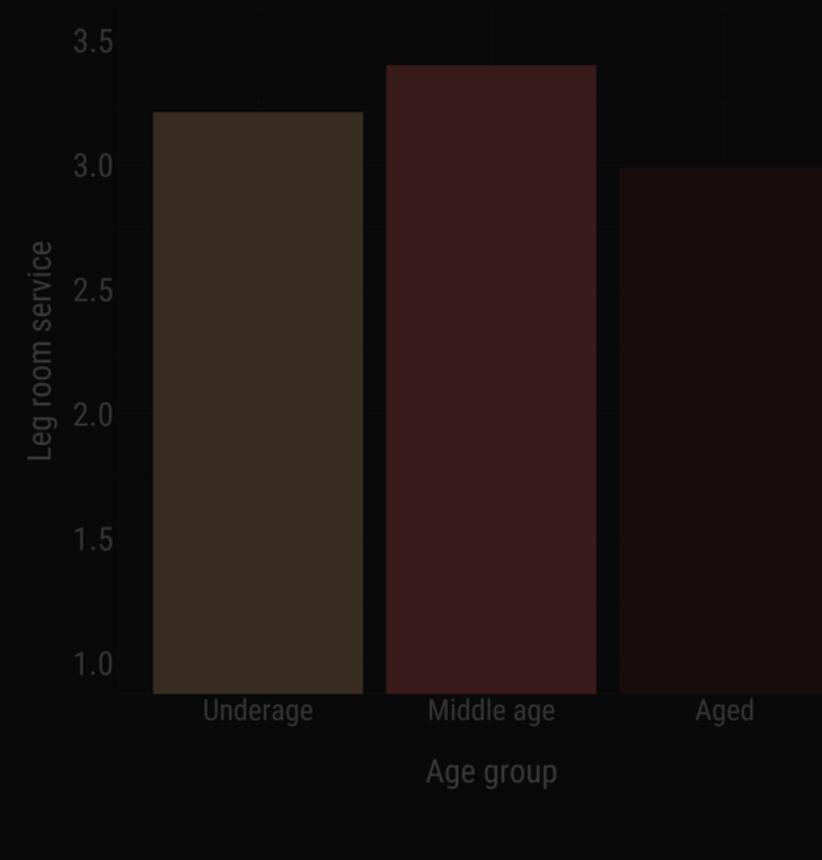
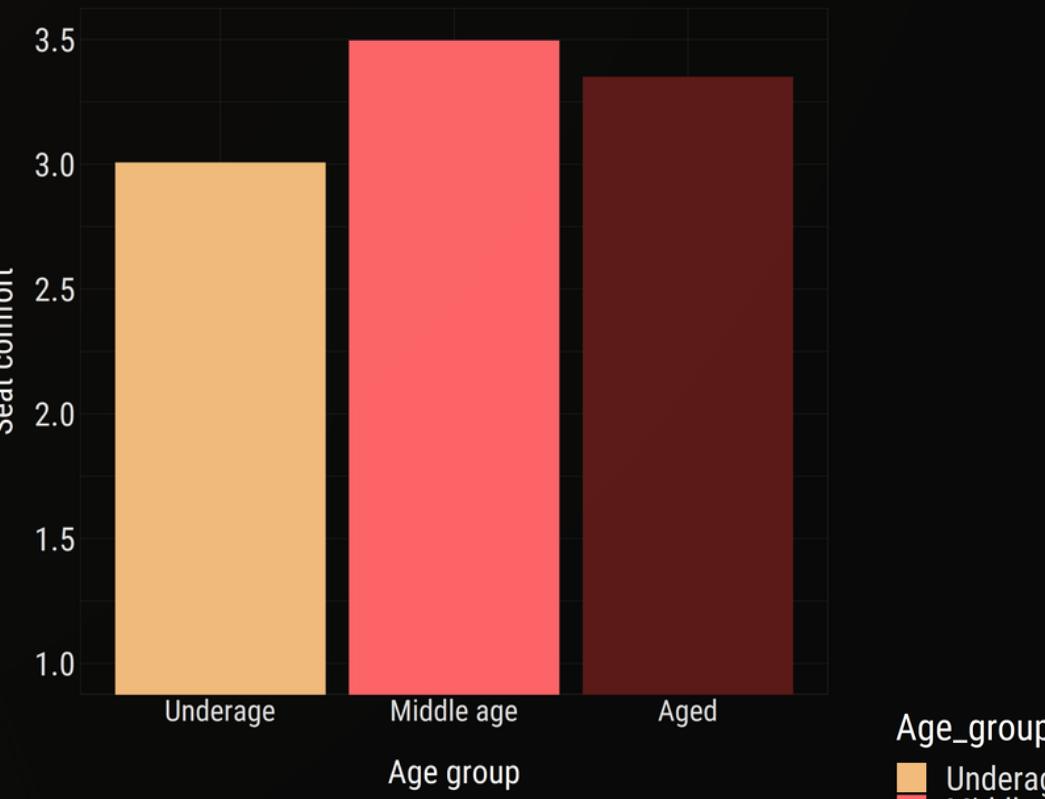
Q: Is there a dependence between specific flight satisfaction factors and age categories of passengers?
H: There are age categories with less effect of some satisfaction factors than others.



Q: Is there a dependence between specific flight satisfaction factors and age categories of passengers?

H: There are age categories with less effect of some satisfaction factors than others.

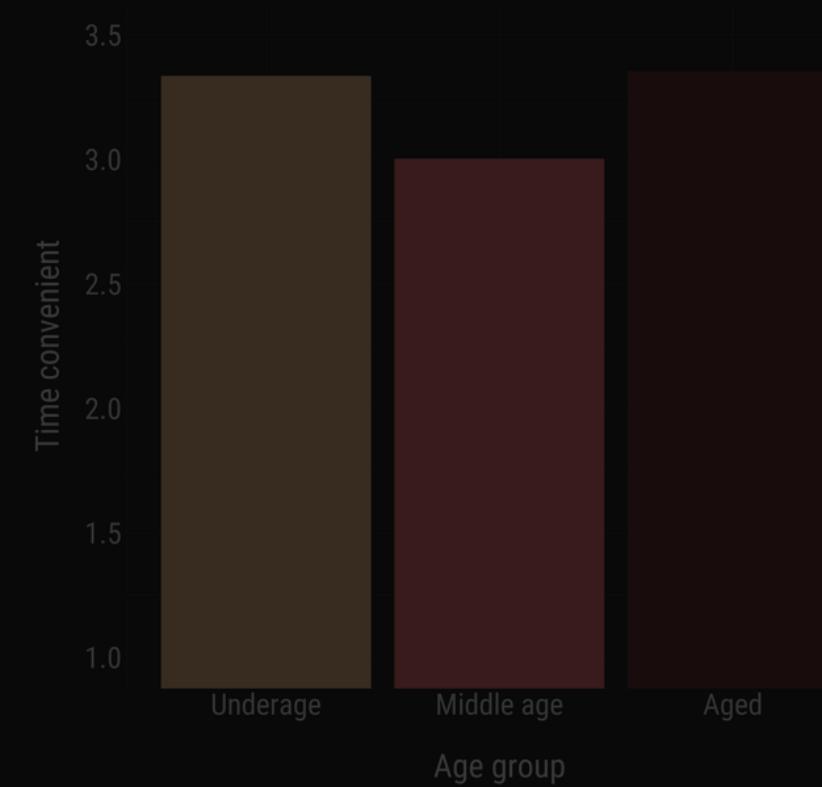
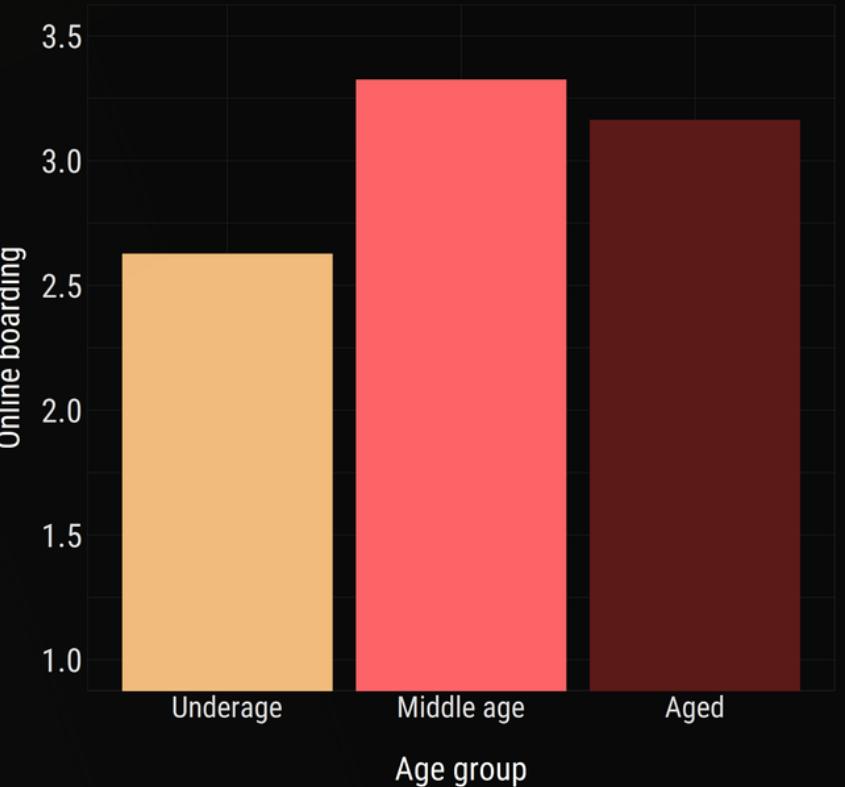
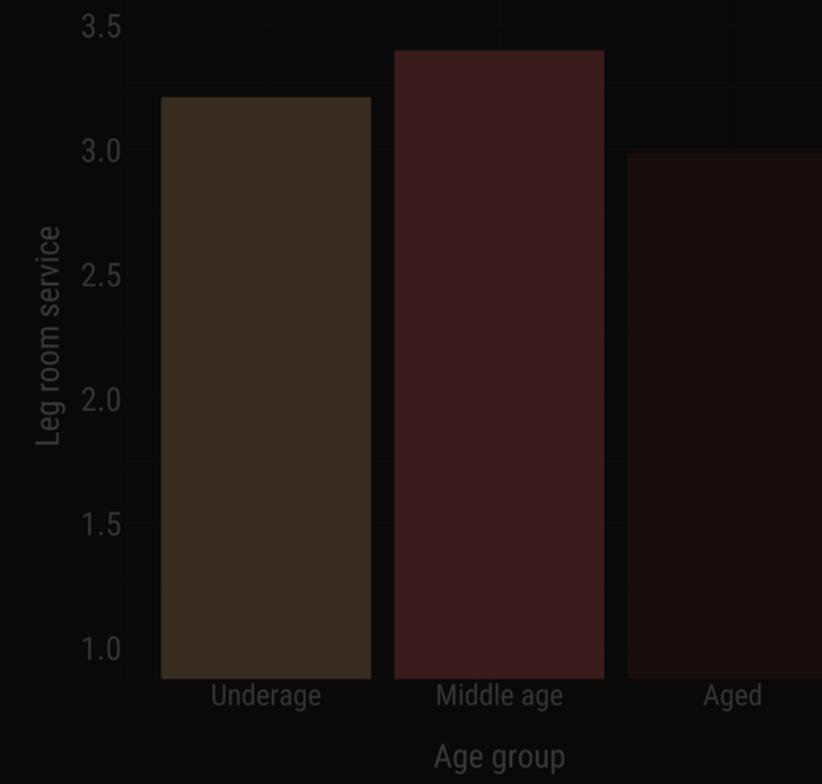
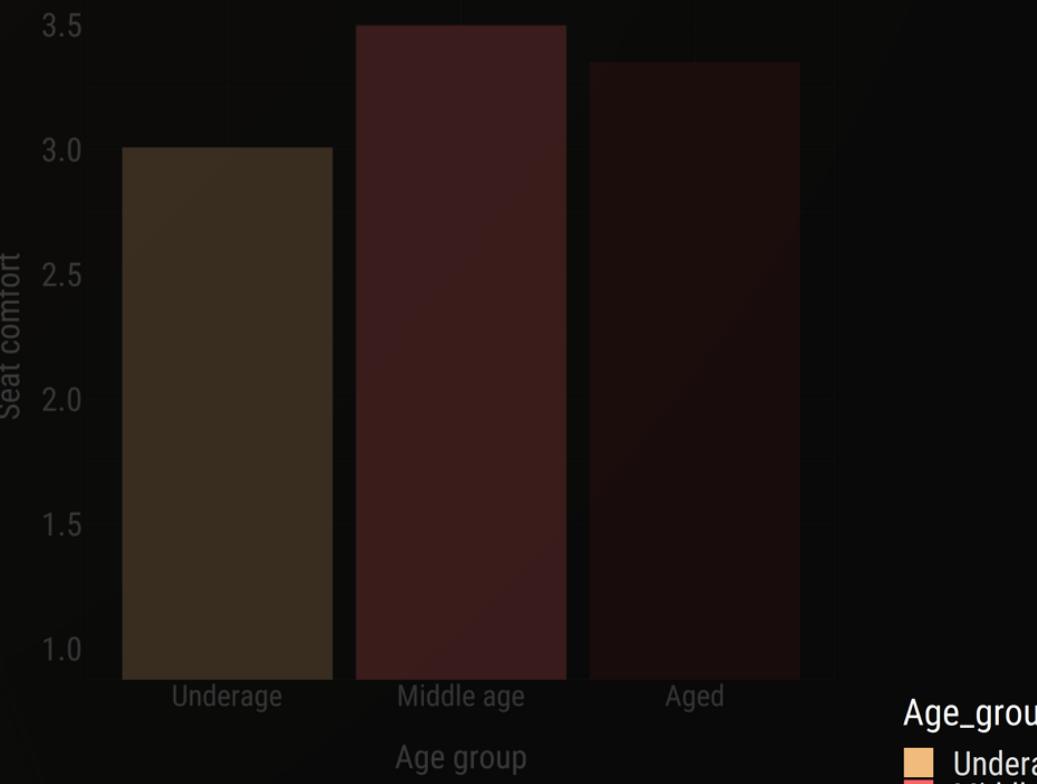
SEAT COMFORT



Q: Is there a dependence between specific flight satisfaction factors and age categories of passengers?

H: There are age categories with less effect of some satisfaction factors than others.

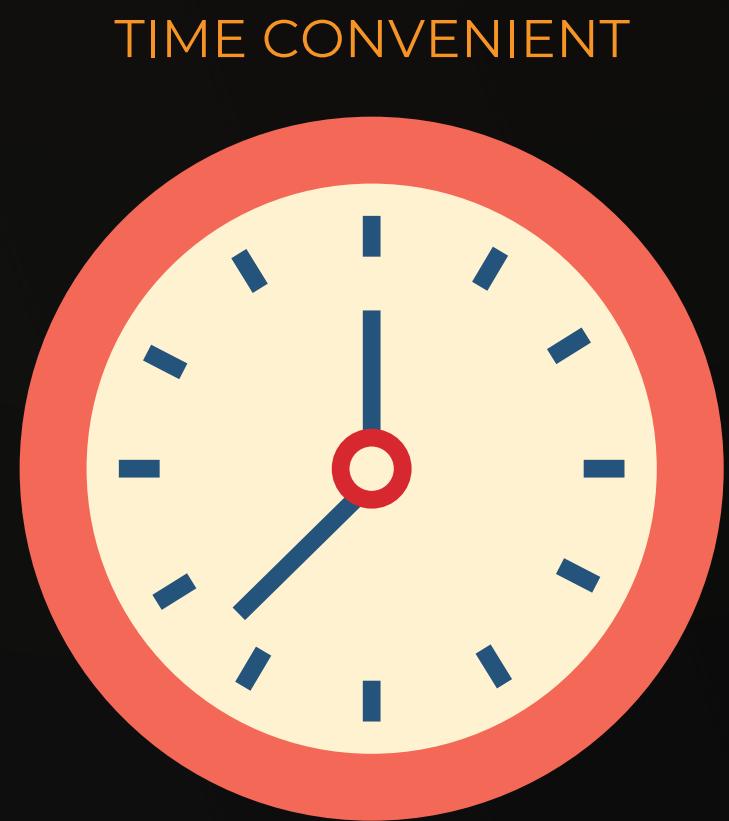
ONLINE BOARDING



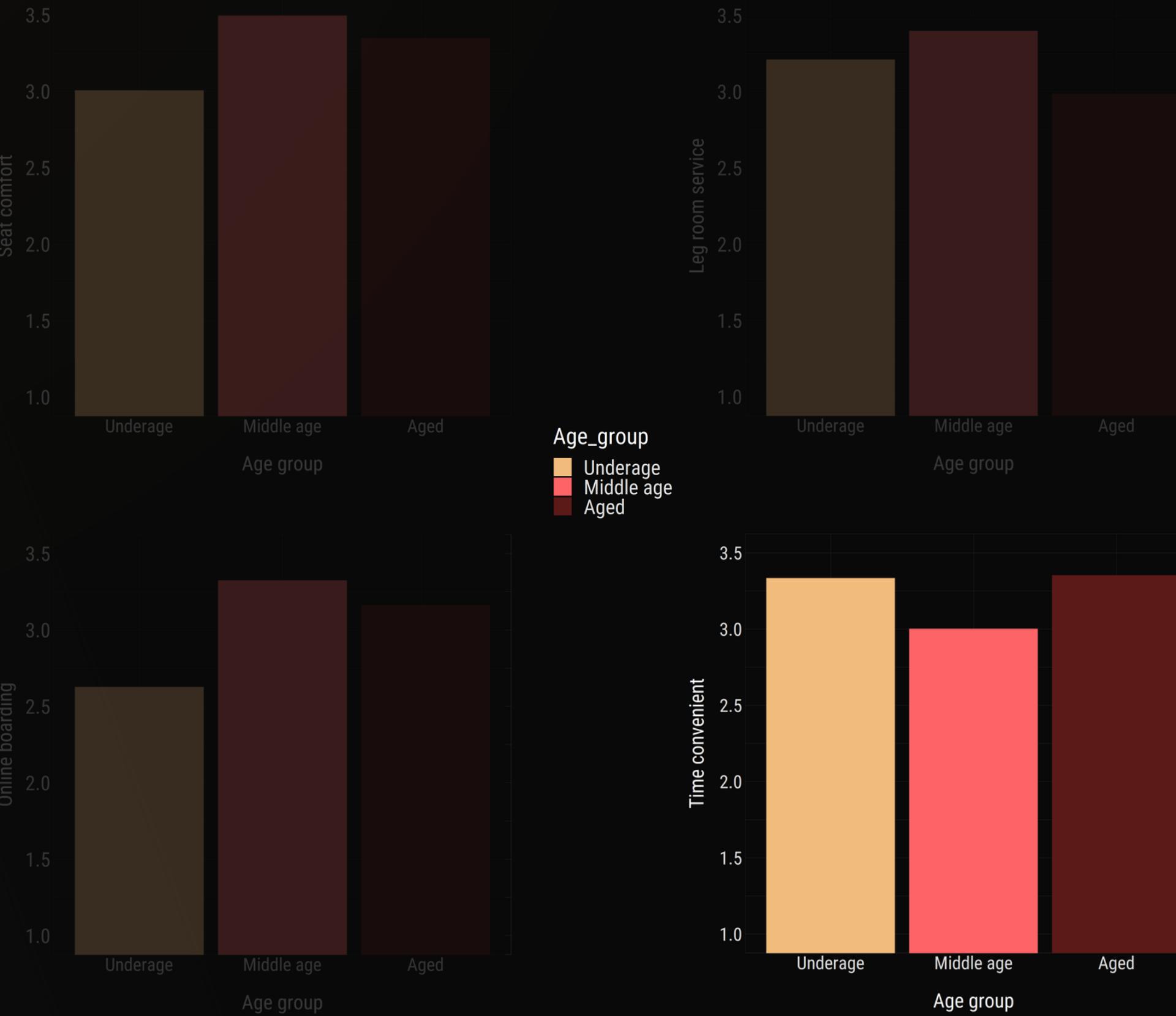
Age_group

- Underage
- Middle age
- Aged

Q: Is there a dependence between specific flight satisfaction factors and age categories of passengers?
H: There are age categories with less effect of some satisfaction factors than others.

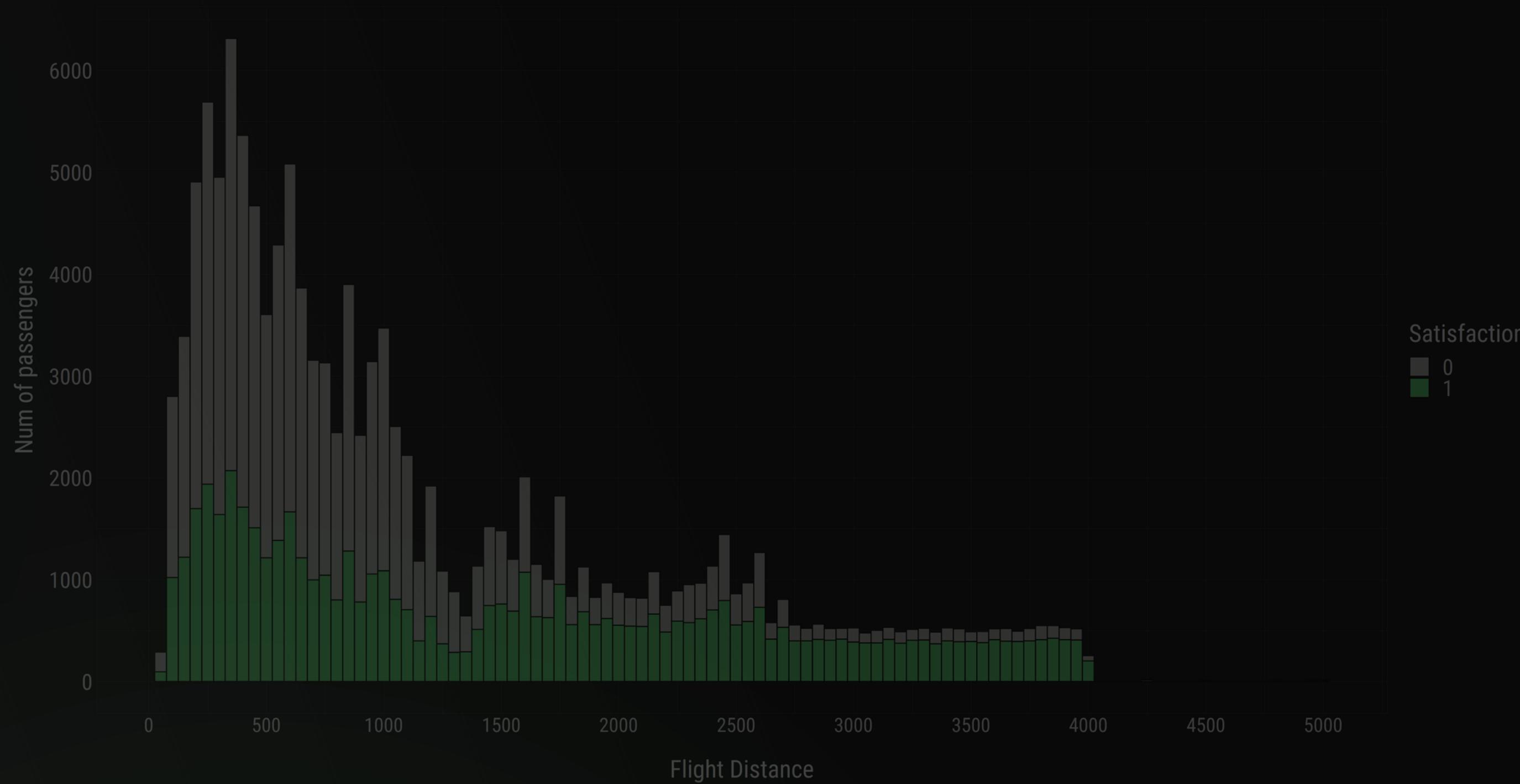


TIME CONVENIENT



Q: Is there a dependence between flight passenger **satisfaction** and flight **goals**?
H: Passengers traveling for personal purposes are **less** satisfied with the flight.

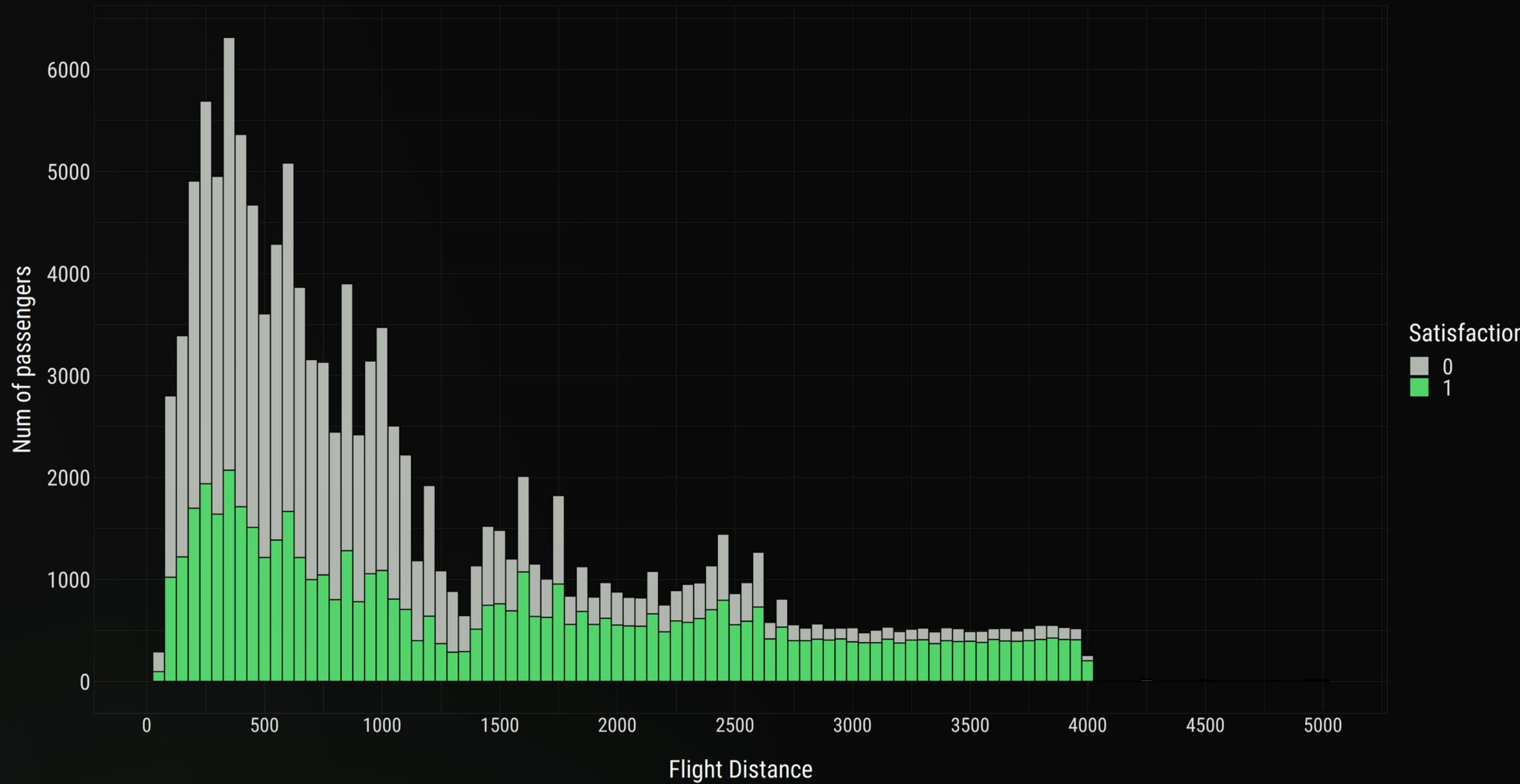
Long-haul flights are more satisfying



Q: Is there a dependence between flight passenger **satisfaction** and flight **goals**?

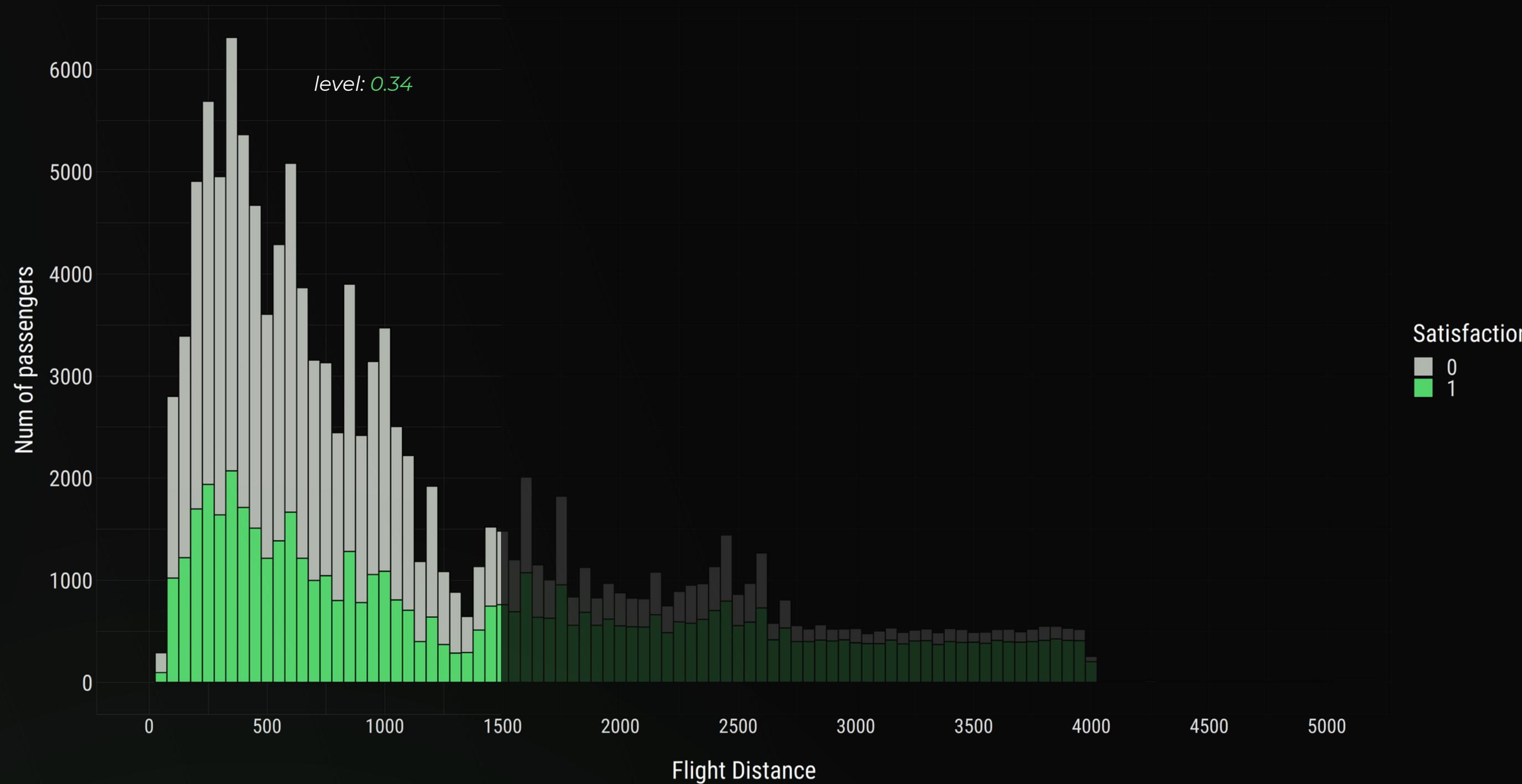
H: Passengers traveling for personal purposes are **less** satisfied with the flight.

Long-haul flights are more satisfying

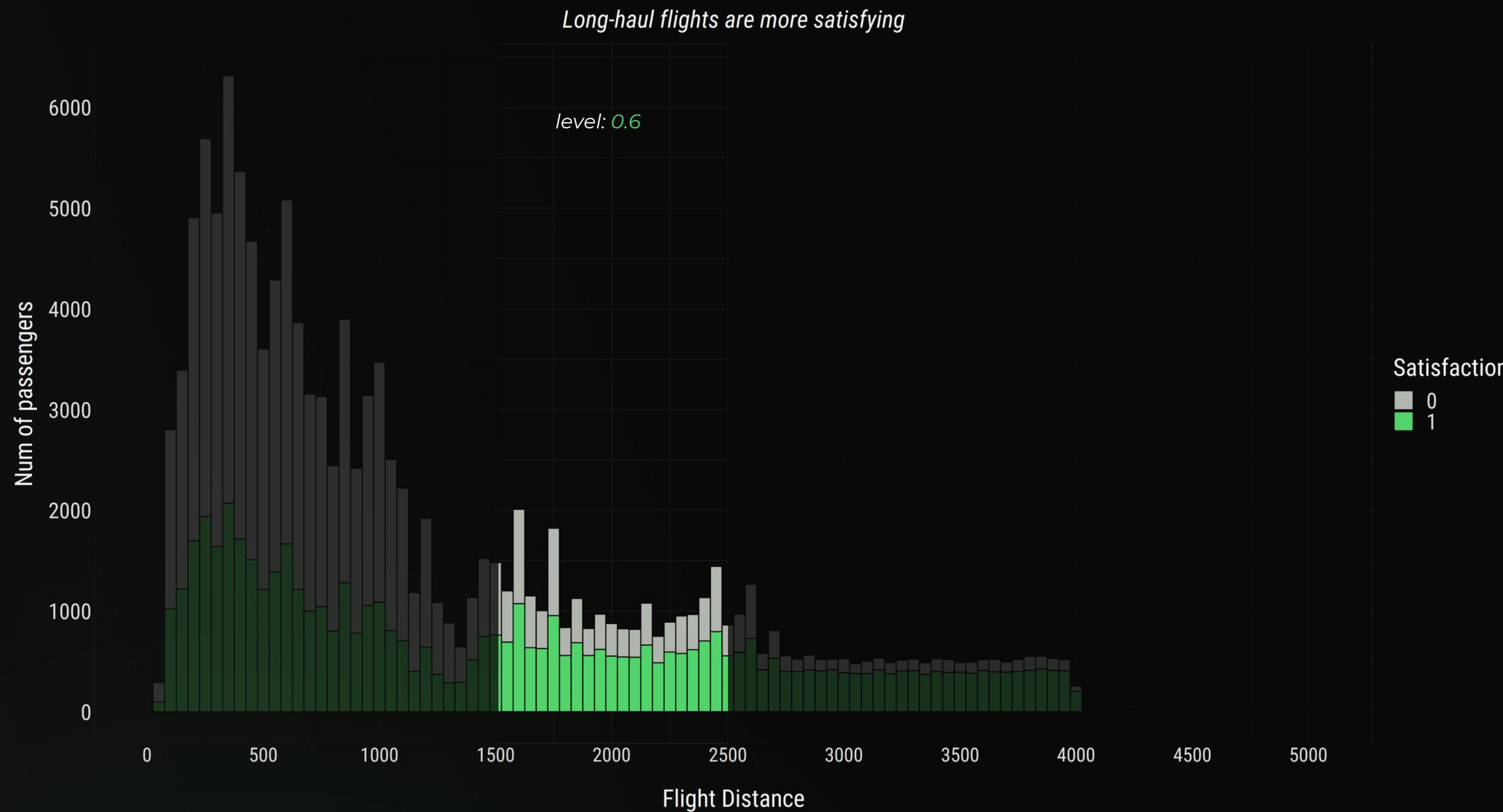


Q: Is there a dependence between flight passenger **satisfaction** and flight **goals**?
H: Passengers traveling for personal purposes are **less** satisfied with the flight.

Long-haul flights are more satisfying

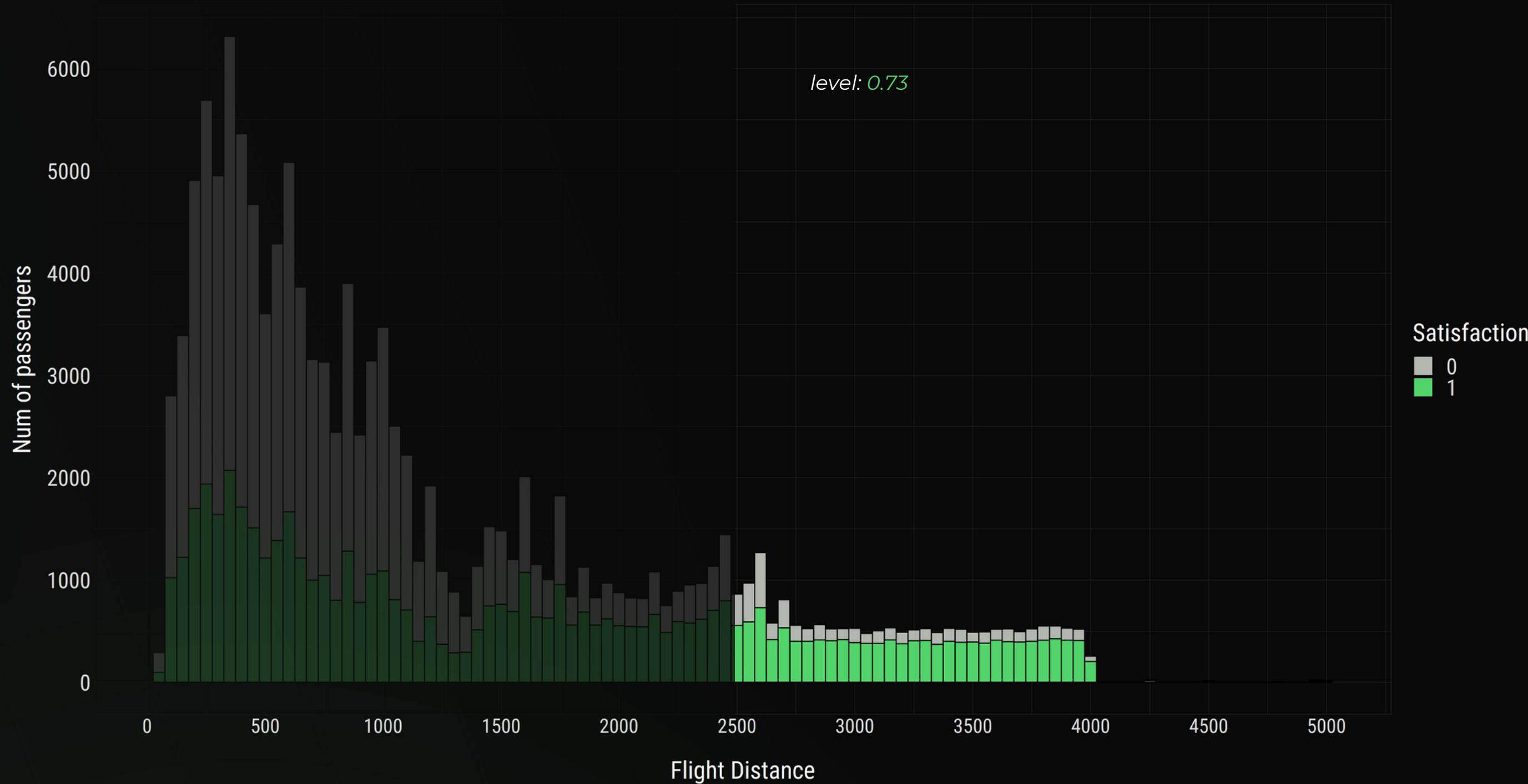


Q: Is there a dependence between flight passenger **satisfaction** and flight **goals**?
H: Passengers traveling for personal purposes are **less** satisfied with the flight.

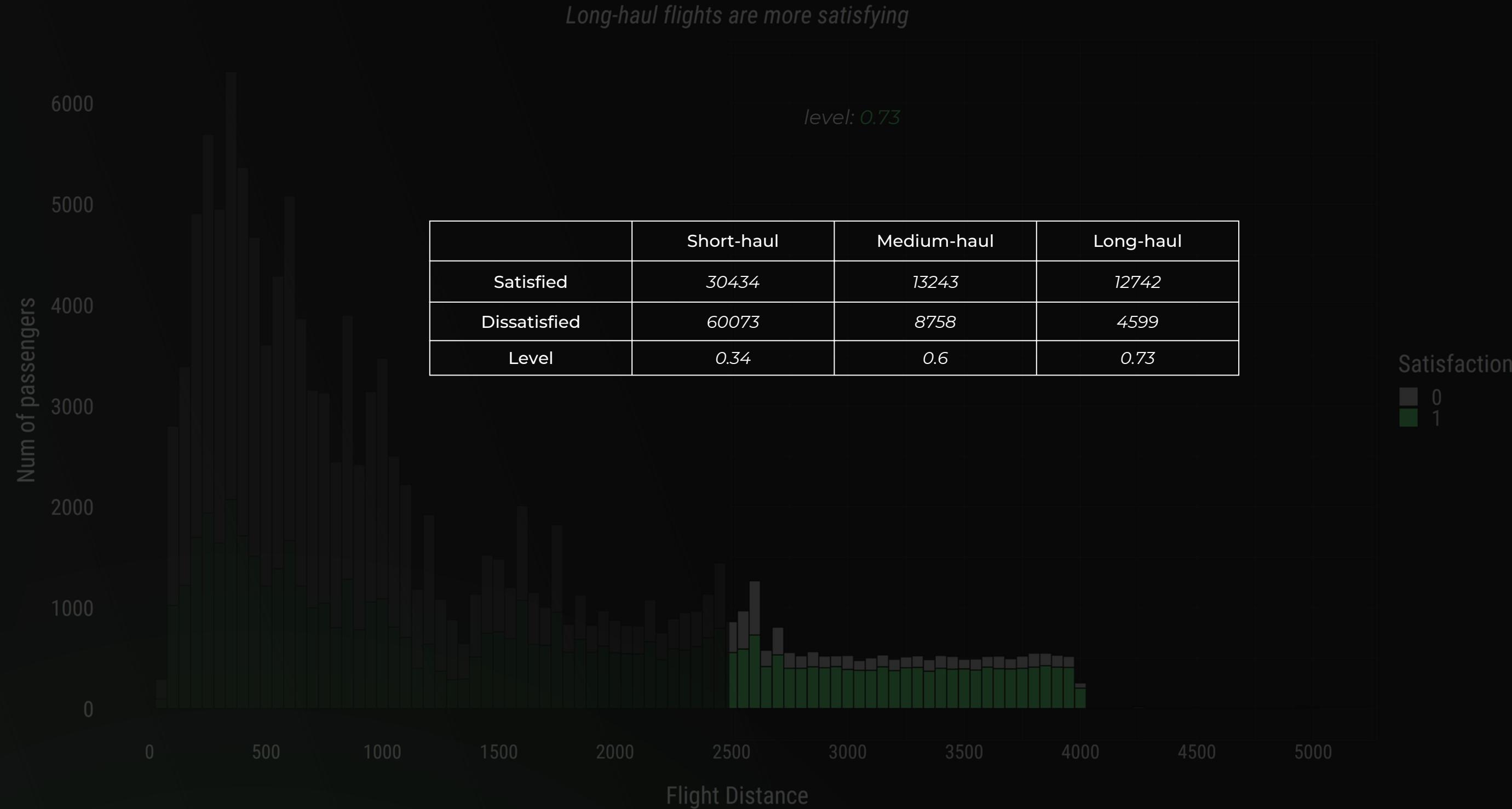


Q: Is there a dependence between flight passenger **satisfaction** and flight **goals**?
H: Passengers traveling for personal purposes are **less** satisfied with the flight.

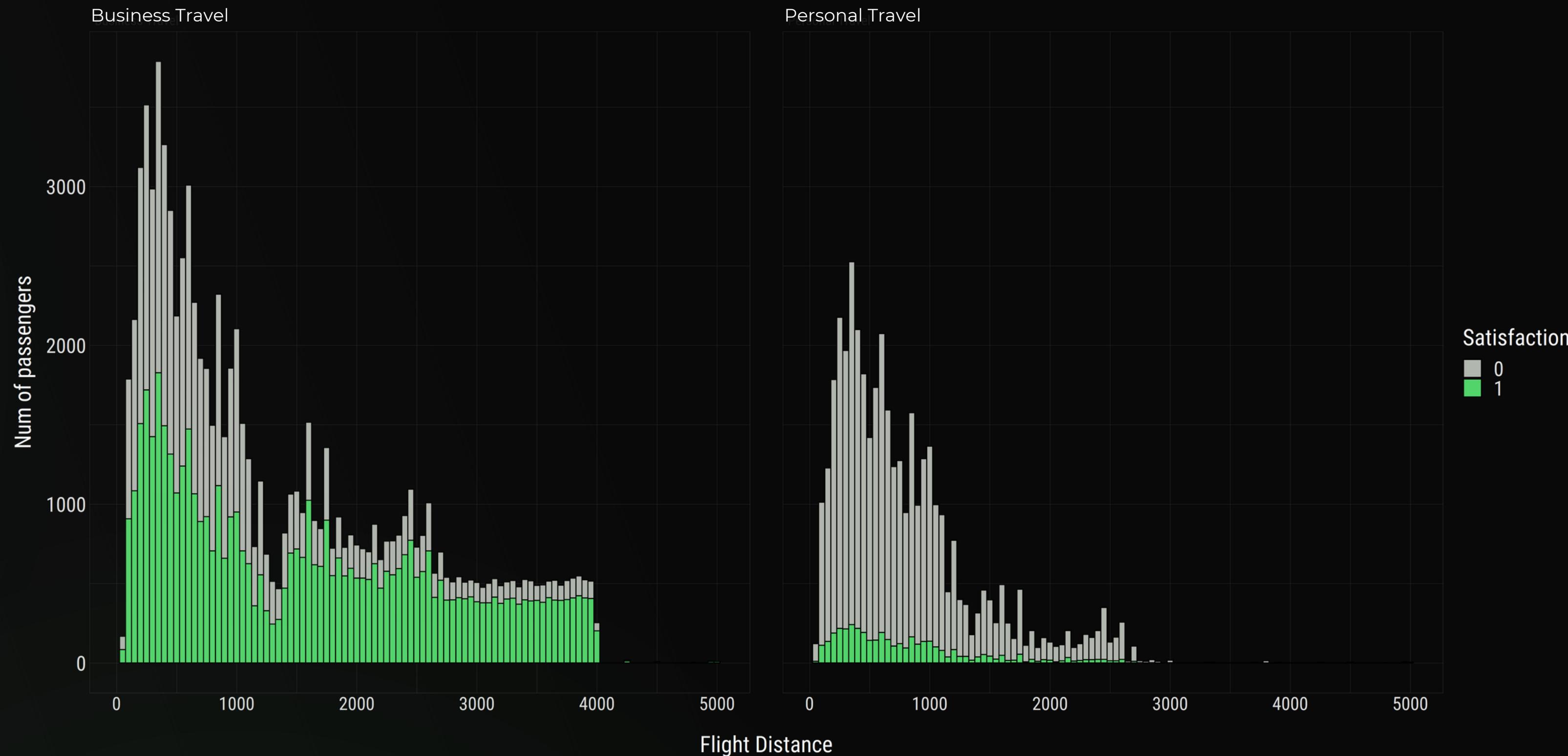
Long-haul flights are more satisfying



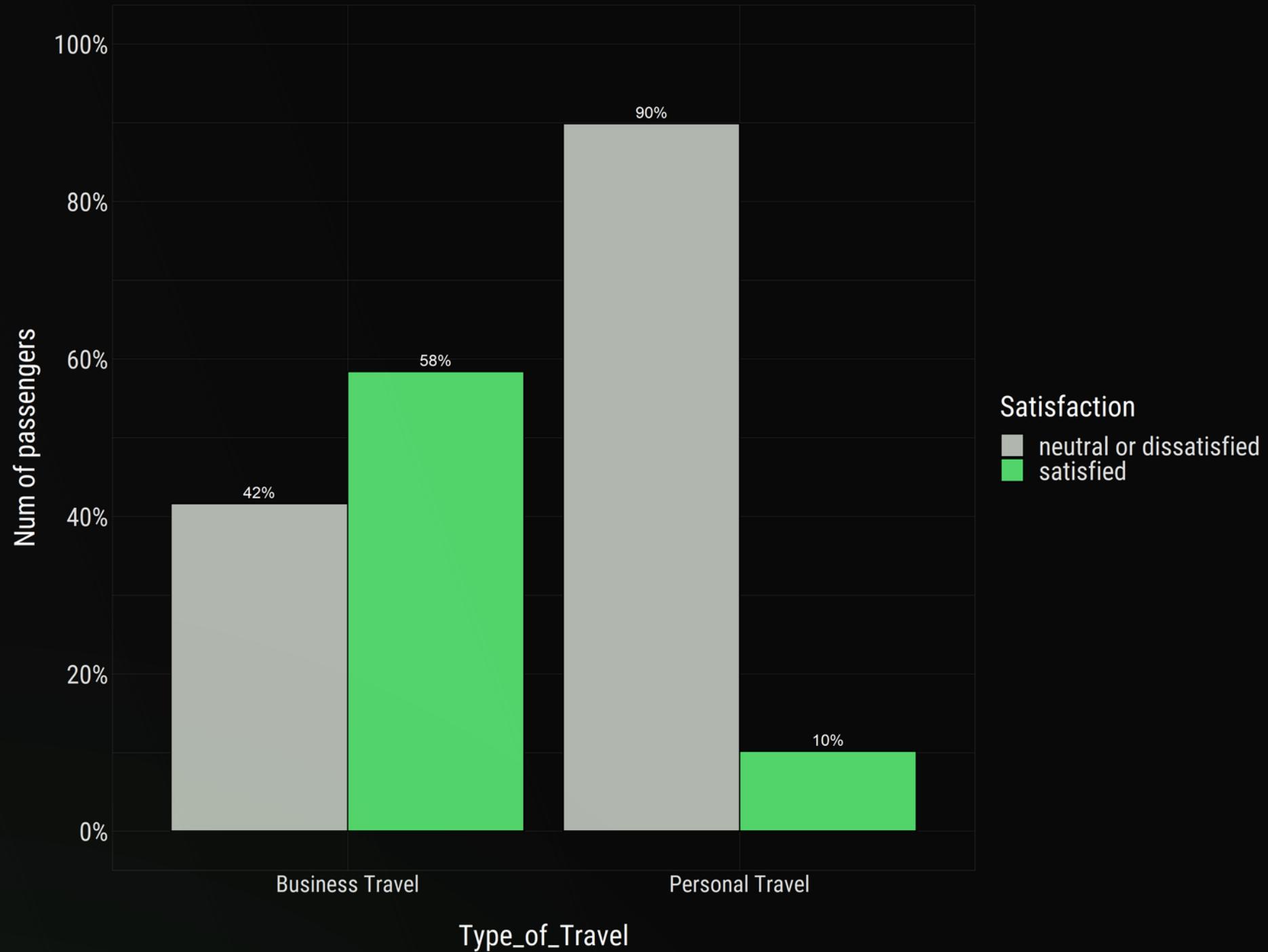
Q: Is there a dependence between flight passenger **satisfaction** and flight **goals**?
H: Passengers traveling for personal purposes are **less** satisfied with the flight.



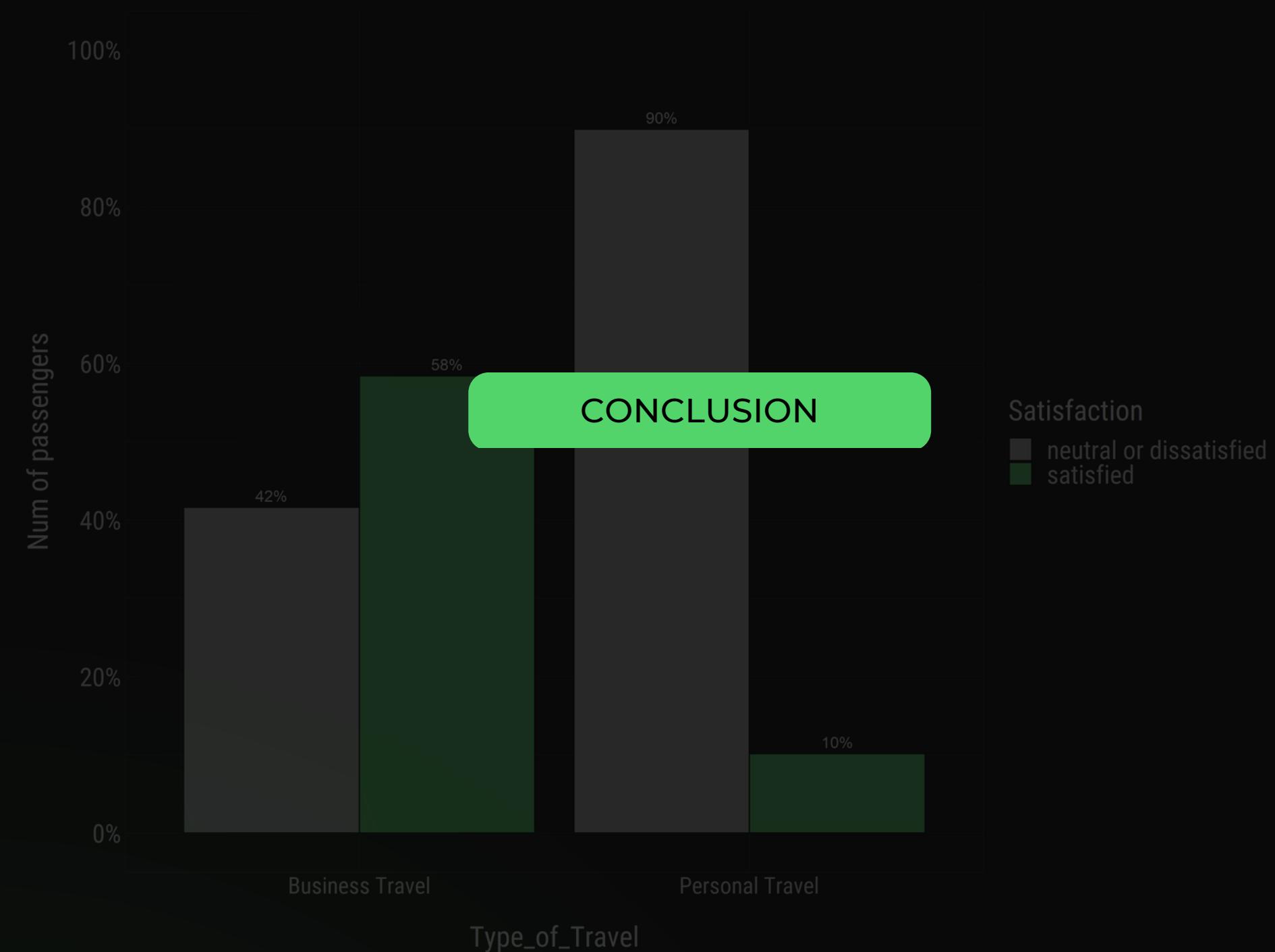
Q: Is there a dependence between flight passenger **satisfaction** and flight **goals**?
H: Passengers traveling for personal purposes are **less** satisfied with the flight.



Q: Is there a dependence between flight passenger **satisfaction** and flight **goals**?
H: Passengers traveling for personal purposes are **less** satisfied with the flight.

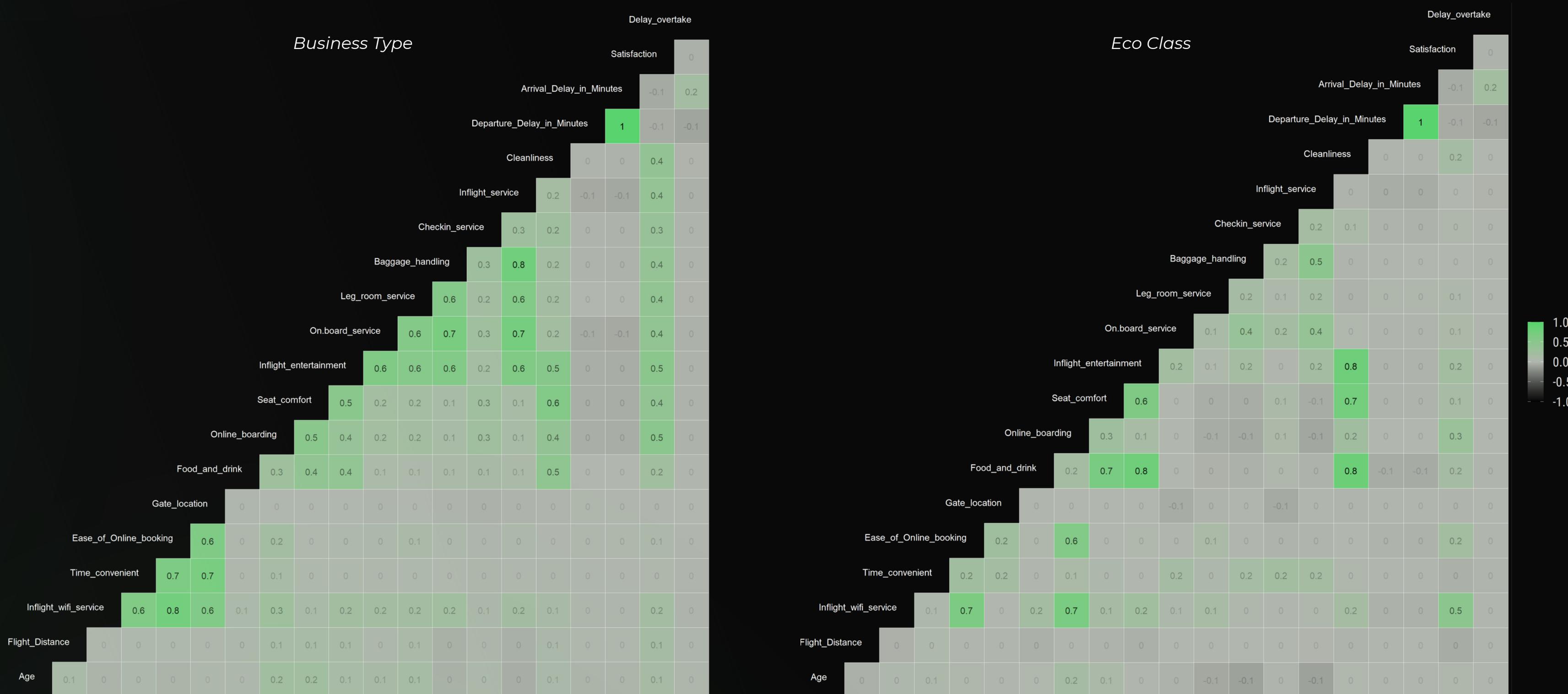


Q: Is there a dependence between flight passenger **satisfaction** and flight **goals**?
H: Passengers traveling for personal purposes are **less satisfied** with the flight.



Q: What attributes have the greatest/least impact on business/economy class passenger satisfaction?

H: There are factors that affect people traveling in different classes differently



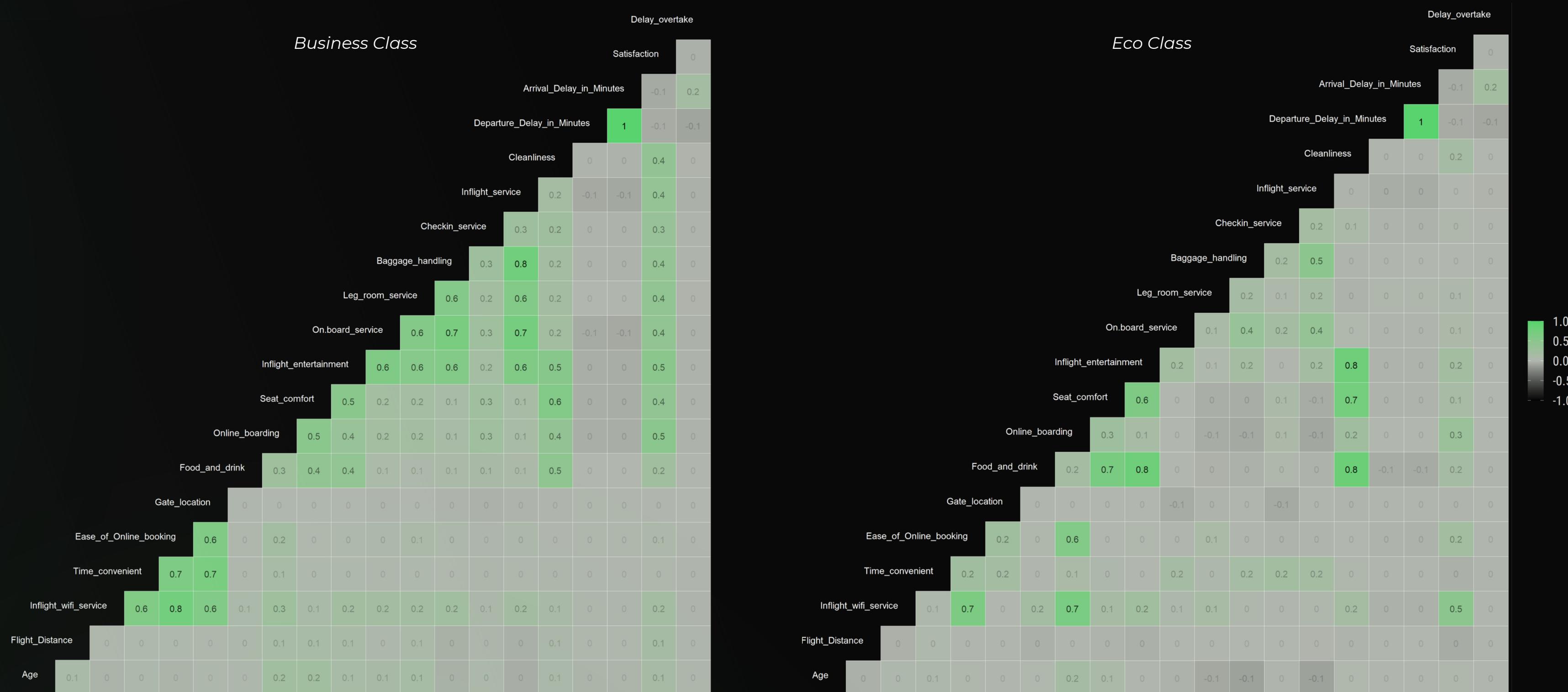
Q: What attributes have the greatest/least impact on business/economy class passenger satisfaction?

H: There are factors that affect people traveling in different classes differently



Q: What attributes have the greatest/least impact on business/economy class passenger satisfaction?

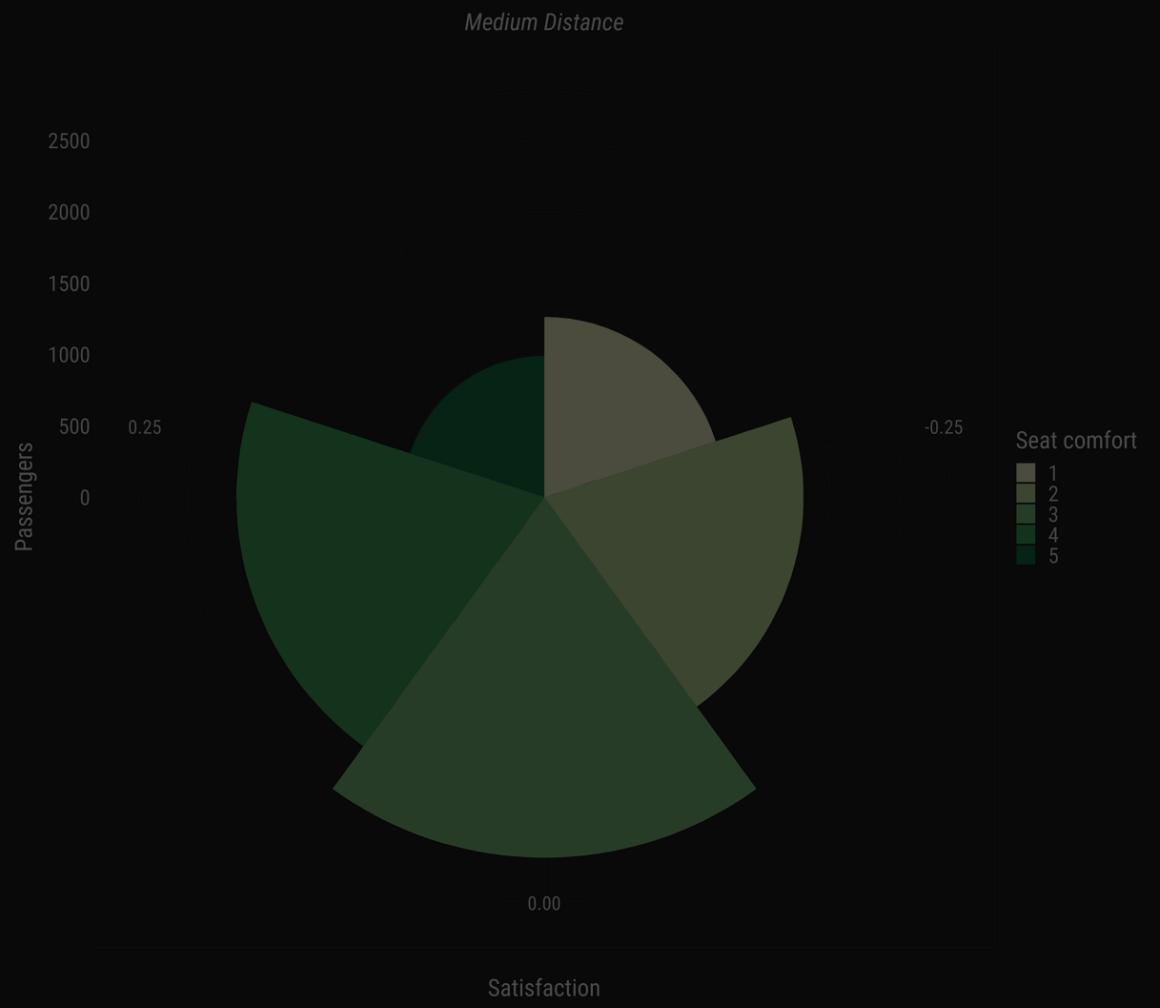
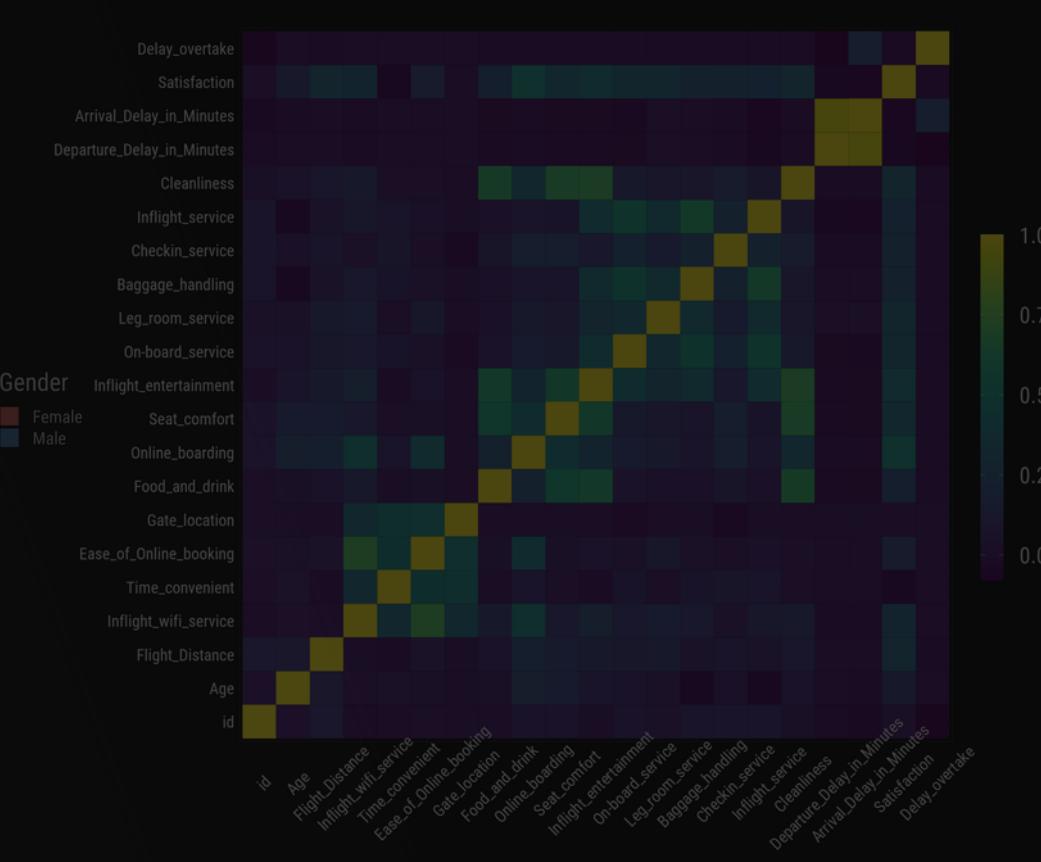
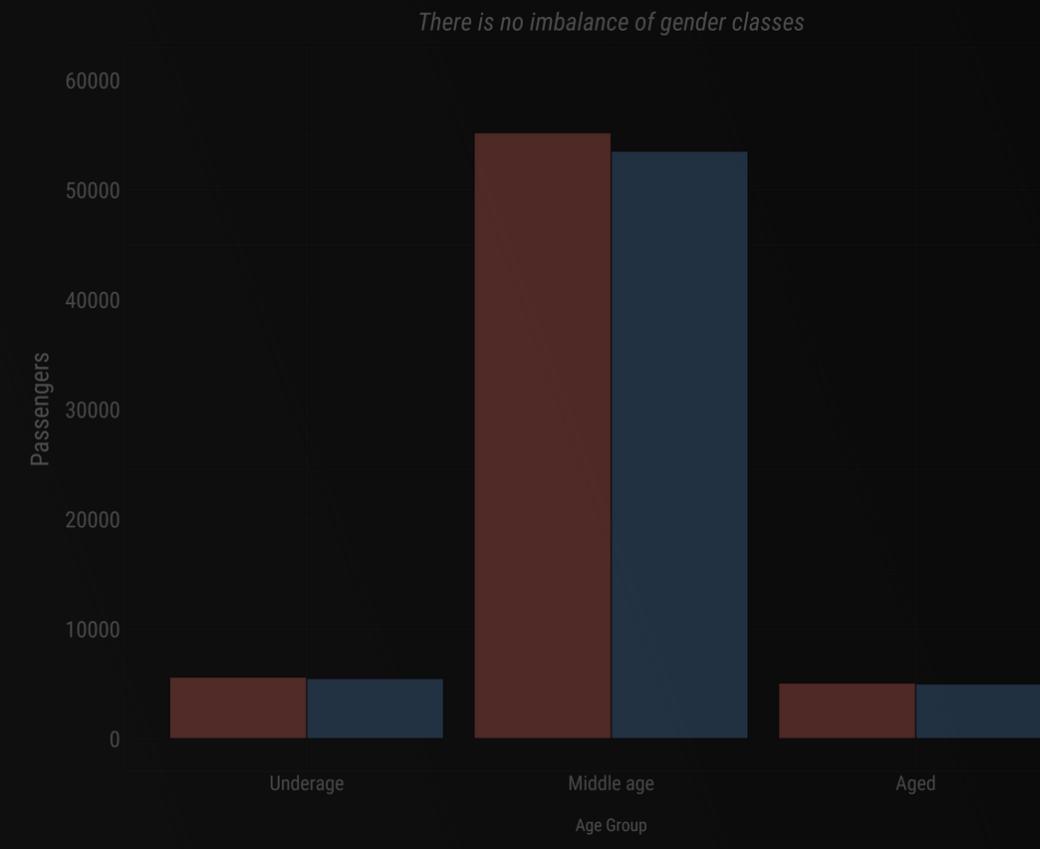
H: There are factors that affect people traveling in different classes differently



Q: What attributes have the greatest/least impact on business/economy class passenger satisfaction?

H: There are factors that affect people traveling in different classes differently





RESEARCH **RESULTS**

& BUSINESS **BENEFITS**



Children & aged mostly dissatisfied of airline

There is problem in seat comfort in airplanes for children



Low satisfaction of online-boarding among aged group

Airline targeting on business trips



Most impact factors - online boarding, inflight entertainment, wifi service

RESEARCH **RESULTS**

& BUSINESS **BENEFITS**



Children & aged mostly dissatisfied of airline

There is problem in seat comfort in airplanes for children



Low satisfaction of online-boarding among aged group

Airline targeting on business trips



Most impact factors - online boarding, inflight entertainment, wifi service

RESEARCH **RESULTS**

& BUSINESS **BENEFITS**



Children & aged mostly dissatisfied of airline

There is problem in seat comfort in airplanes for children



Low satisfaction of online-boarding among aged group

Airline targeting on business trips



Most impact factors - online boarding, inflight entertainment, wifi service

RESEARCH **RESULTS**

& BUSINESS **BENEFITS**



Children & aged mostly dissatisfied of airline

There is problem in seat comfort in airplanes for children



Low satisfaction of online-boarding among aged group

Airline targeting on business trips



Most impact factors - online boarding, inflight entertainment, wifi service

RESEARCH **RESULTS**

& BUSINESS **BENEFITS**



Children & aged mostly dissatisfied of airline

There is problem in seat comfort in airplanes for children



Low satisfaction of online-boarding among aged group

Airline targeting on business trips



Most impact factors - online boarding, inflight entertainment, wifi service

REALIZED **FEATURES**



& SOURCE

«Business Intelligence in Airline Passenger Satisfaction Study — A Fuzzy-Genetic Approach with Optimized Interpretability-Accuracy Trade-Off» - Marian B. Gorzałczany, Filip Rudzinski, and Jakub Piekoszewski, Department of Electrical and Computer Engineering, Kielce University of Technology, Poland, 2021

«Investigating airline passenger satisfaction: Data mining method» - Tri Noviantoro, Jen-Peng Huang, College of Business, Southern Taiwan University of Science and Technology, Taiwan, 2022

«Feature Analysis on Airline Passenger Satisfaction using Orange Tool» - Hannah Susan Mathew, Department of Computer Science, Rajagiri College of Social Sciences, Kochi, India, 2022.

REALIZED **FEATURES**



& SOURCE

«*Business Intelligence in Airline Passenger Satisfaction Study — A Fuzzy-Genetic Approach with Optimized Interpretability-Accuracy Trade-Off*» - Marian B. Gorzałczany, Filip Rudzinski, and Jakub Piekoszewski, Department of Electrical and Computer Engineering, Kielce University of Technology, Poland, 2021

«*Investigating airline passenger satisfaction: Data mining method*» - Tri Noviantoro, Jen-Peng Huang, College of Business, Southern Taiwan University of Science and Technology, Taiwan, 2022

«*Feature Analysis on Airline Passenger Satisfaction using Orange Tool*» - Hannah Susan Mathew, Department of Computer Science, Rajagiri College of Social Sciences, Kochi, India, 2022.