AIRLINE PASSENGER SATISFACTION

Project 4: MSU Bootcamp

INTRODUCTION

Michigan-based Airlines – **JEKS Air** – ran a survey to determine which factors impact passenger satisfaction. The survey includes this data:

- Gender
- Customer Type
- Age
- Type of Travel
- Class
- Flight distance
- Inflight Wi-Fi service
- Departure/Arrival time convenient
- Ease of Online booking
- Gate location
- Food and drink
- Online boarding
- Seat comfort
- Inflight entertainment
- On-board service
- Leg room service
- Baggage handling
- Check-in service
- Inflight service
- Cleanliness
- Departure Delay in Minutes
- Arrival Delay in Minutes

Aside from the overall satisfaction, all other satisfaction metrics are on a scale from 1 to 5 with 0 representing “not applicable”.

JEKS Air is now asking a team of Data Scientist to help them find **an algorithm that could predict customer satisfaction** and **help the airlines deliver better service**.

JEKS Air is also interested in building **a new set of tools that will allow them to visualize their survey data**. They collect a massive amount of data from all over the world each day, but they lack a meaningful way of displaying it.

MEET OUR CREW

STEPHANIE WORTMAN

ELIZABETH HANSEN

JULIE EREMEEVA

KATHRYN KESSLER

WELCOME ABOARD

PRIMARY GOALS

RESEARCH GOALS

№ 1

Define top 3 factors that affect satisfaction levels the most.

Where do the airlines need to invest more money?

№ 2

Define top 3 factors that affect satisfaction levels the least.

Where is it safe to cut down some costs?

№ 3

Develop a Machine Learning Model that can predict airline satisfaction based on the survey results with 90% + accuracy.

PROJECT STEPS

EXPLORING

Getting familiar with the data and asking questions

CLEANING

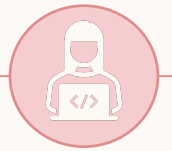
Cleaning the data to prepare it for further analysis

VISUALIZING

Getting the most meaning out of the raw data

SUMMARIZING TRENDS

Making conclusions on what we have explored



BUILD A ML MODEL

Deploying a machine learning algorithm to predict customer satisfaction

TECHNOLOGIES

OUR TOOLKIT


X


+ a b | e a u


pandas


NumPy


S3


APACHE
Spark


scikit
learn


TensorFlow

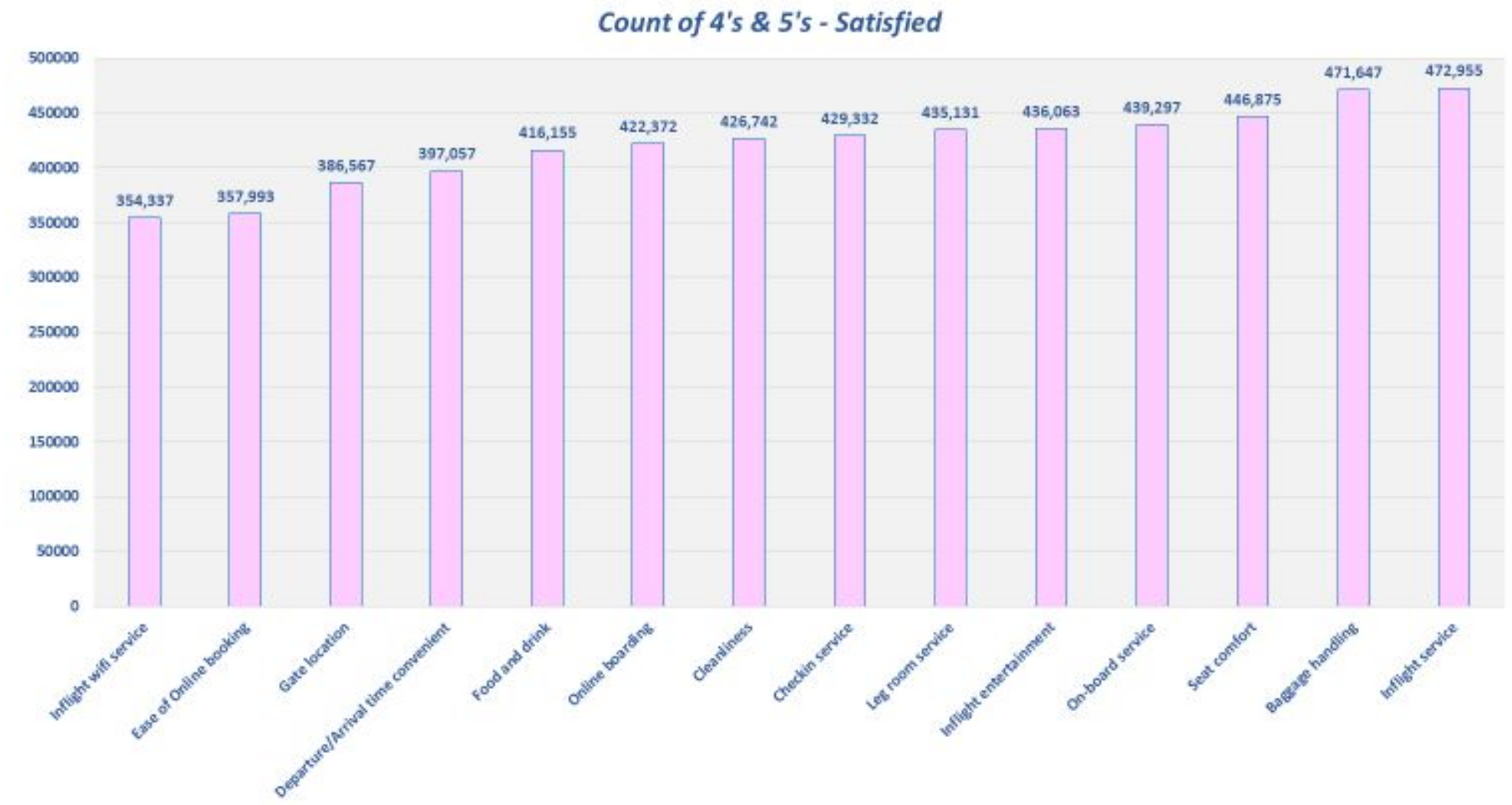

colab

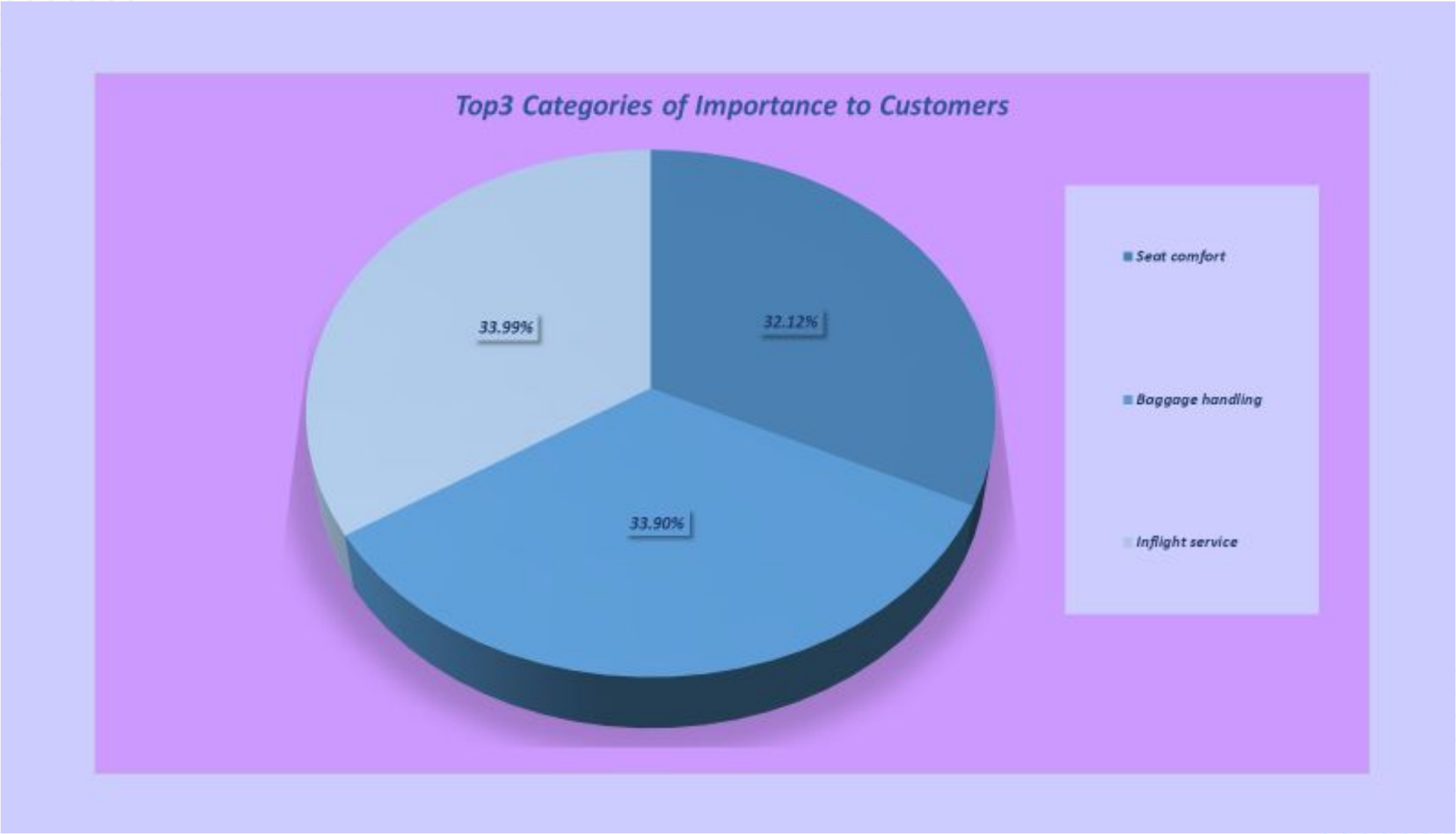

matplotlib

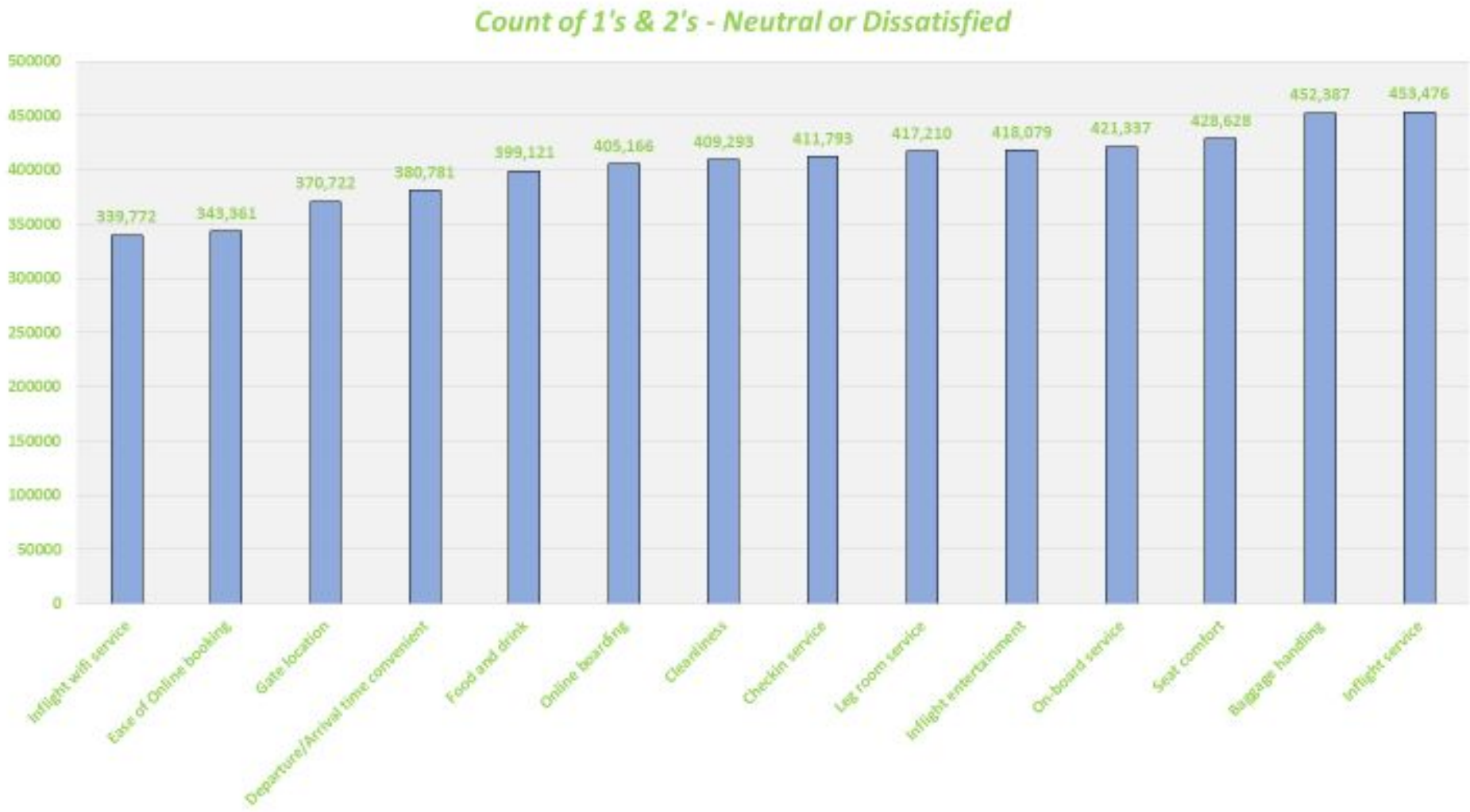

seaborn

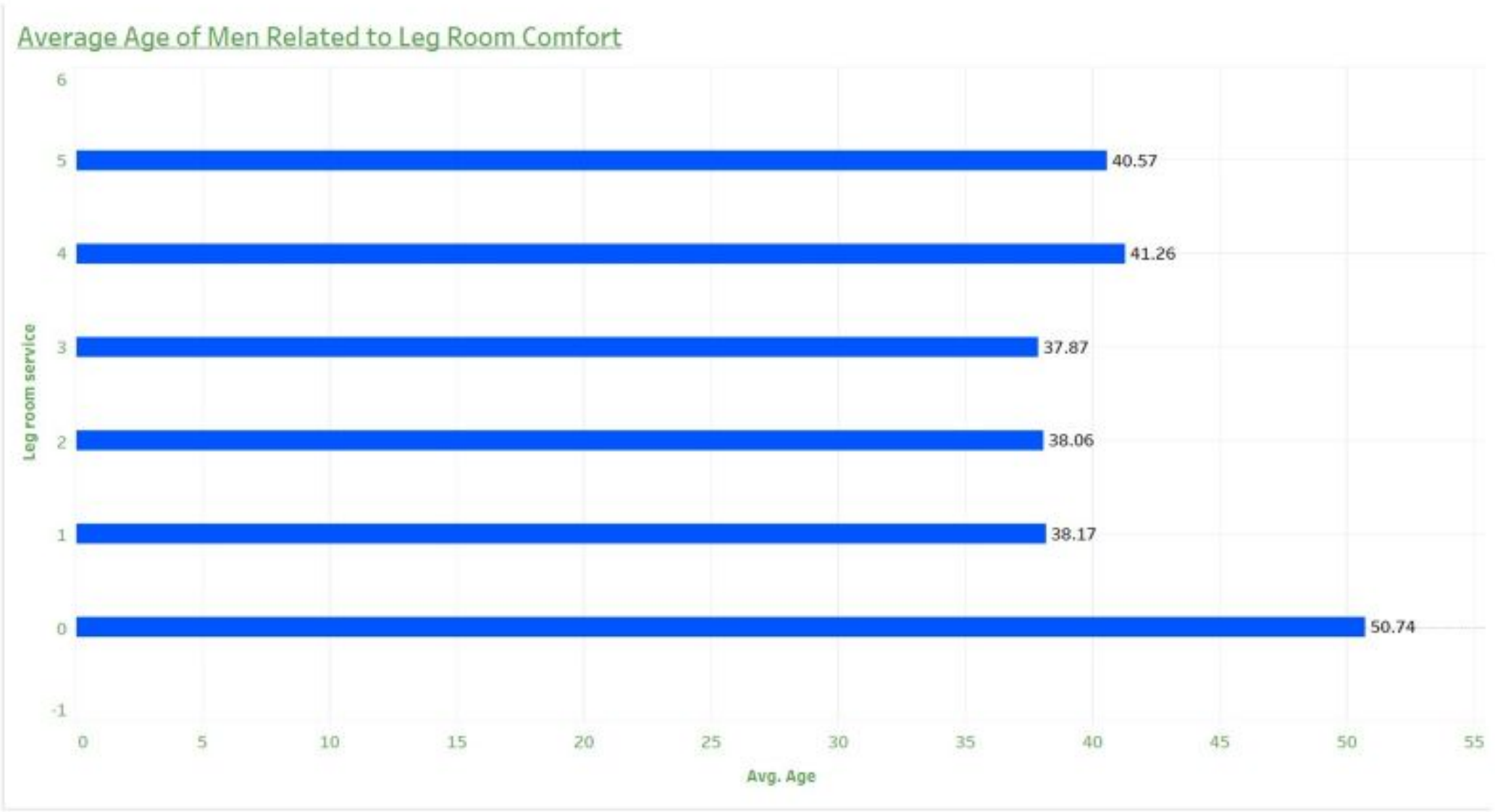
DATA VISUALIZATION

Tableau Dashboard Demo


Count of 4's & 5's - Satisfied
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
354,337
357,993
386,567
397,057
416,155
422,372
426,742
429,332
435,131
436,063
439,297
446,875
471,647
472,955
Inflight wifi service
Ease of Online booking
Gate location
Departure/Arrival time convenient
Food and drink
Online boarding
Cleanliness
Checkin service
Leg room service
Inflight entertainment
On-board service
Seat comfort
Baggage handling
Inflight service


Top3 Categories of Importance to Customers
33.99%
32.12%
33.90%
Seat comfort
Baggage handling
Inflight service


Count of 1's & 2's - Neutral or Dissatisfied
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
339,772
343,361
370,722
380,781
399,121
405,166
409,293
411,793
417,210
418,079
421,337
428,628
452,387
453,476
Inflight wifi service
Ease of Online booking
Gate location
Departure/Arrival time convenient
Food and drink
Online boarding
Cleanliness
Checkin service
Leg room service
Inflight entertainment
On-board service
Seat comfort
Baggage handling
Inflight service


Average Age of Men Related to Leg Room Comfort
Leg room service
6
5
4
3
2
1
0
-1
40.57
41.26
37.87
38.06
38.17
50.74
0
5
10
15
20
25
30
35
40
45
50
55
Avg. Age

MACHINE LEARNING

Report № 1: Logistic Regression

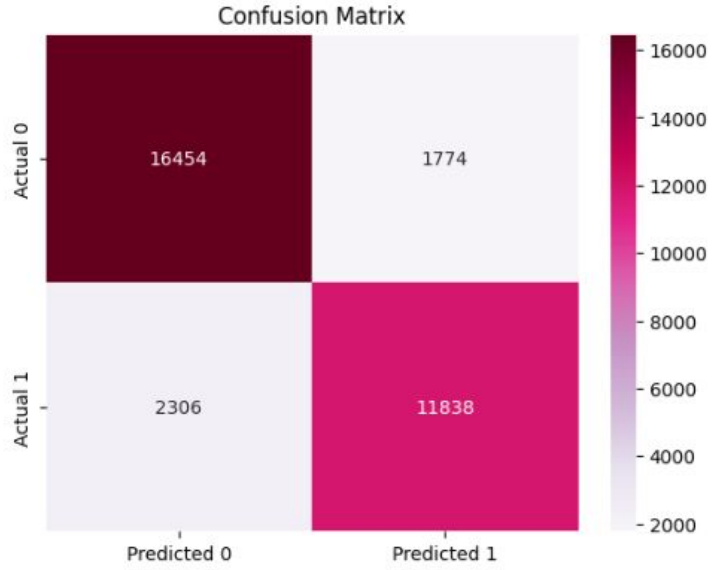
Confusion Matrix

	Predicted 0	Predicted 1
Actual 0	16571	1816
Actual 1	2334	11749

Accuracy Score : 0.8721897135817678

Classification Report

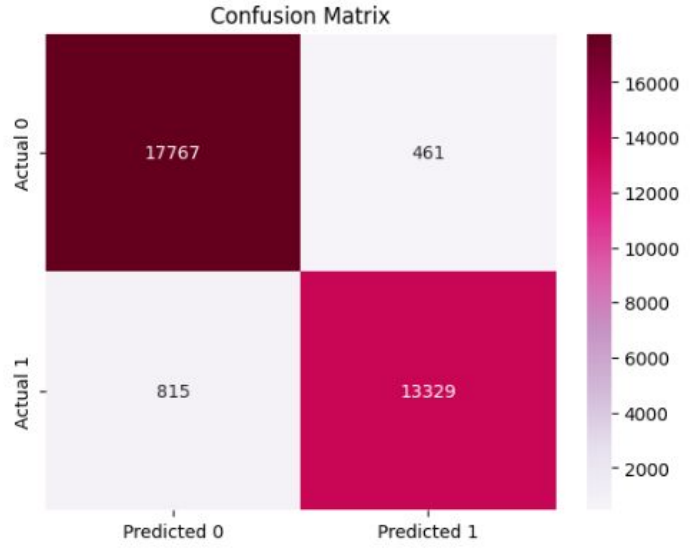
	precision	recall	f1-score	support
0	0.88	0.90	0.89	18387
1	0.87	0.83	0.85	14083
accuracy			0.87	32470
macro avg	0.87	0.87	0.87	32470
weighted avg	0.87	0.87	0.87	32470

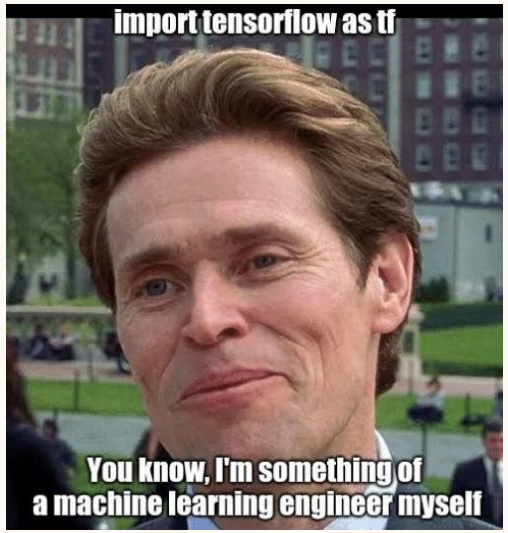


Report № 2: Neural Network

```
# Evaluate the model using the test data
model_loss, model_accuracy = nn.evaluate(X_test_scaled,y_test,verbose=2)
print(f"Loss: {model_loss}, Accuracy: {model_accuracy}")
```

```
1015/1015 - 2s - loss: 0.0992 - accuracy: 0.9610 - 2s/epoch - 2ms/step
Loss: 0.0992468073964119, Accuracy: 0.9610409736633301
```



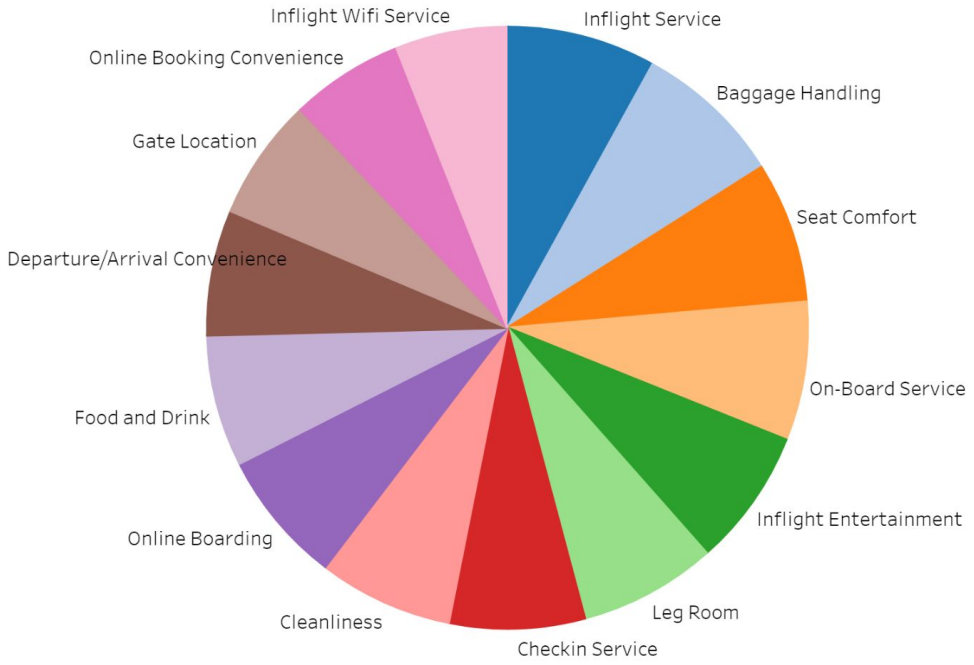


LIMITATIONS

- No data on location or airlines (JEKS AIR was made up by us).
- The answer “neutral” and “dissatisfied” were combined even though they have very different meaning.
- A lot of “N/A” inputs. For some columns, keeping N/As was important (e.g., delay in departure/arrival). For other columns, it was unnecessary and might have affected the ML model.
- No information on how the survey was conducted or how representative the population is.

Recommendations

- Machine Learning Accuracy
- Top and Bottom Parameters



Top 3 Parameters:
Inflight Service
Baggage Handling
Seat Comfort

Bottom 3 Parameters:
Inflight Wifi Service
Online Booking Convenience
Gate Location

THANK YOU