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 Scientists 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.

WELCOME ABOARD
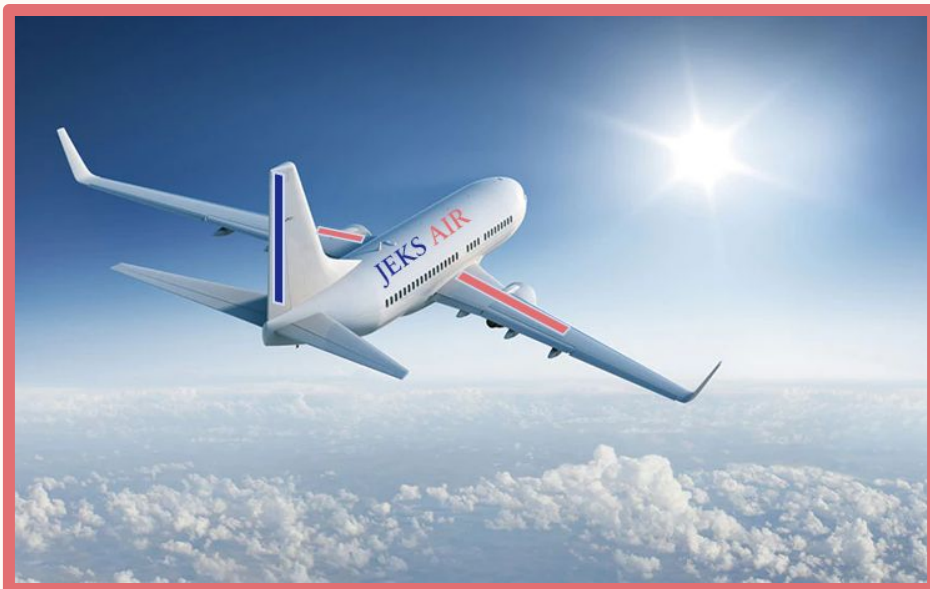
MEET OUR CREW

STEPHANIE WORTMAN

ELIZABETH HANSEN

JULIE EREMEEVA

KATHRYN KESSLER

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

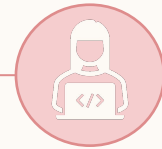Cleaning the data to prepare it for further analysis

VISUALIZING

Getting the most meaning out of the raw data

SUMMARIZING TRENDS

Making conclusions on what we have explored

BUILD A ML MODEL

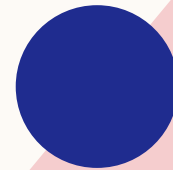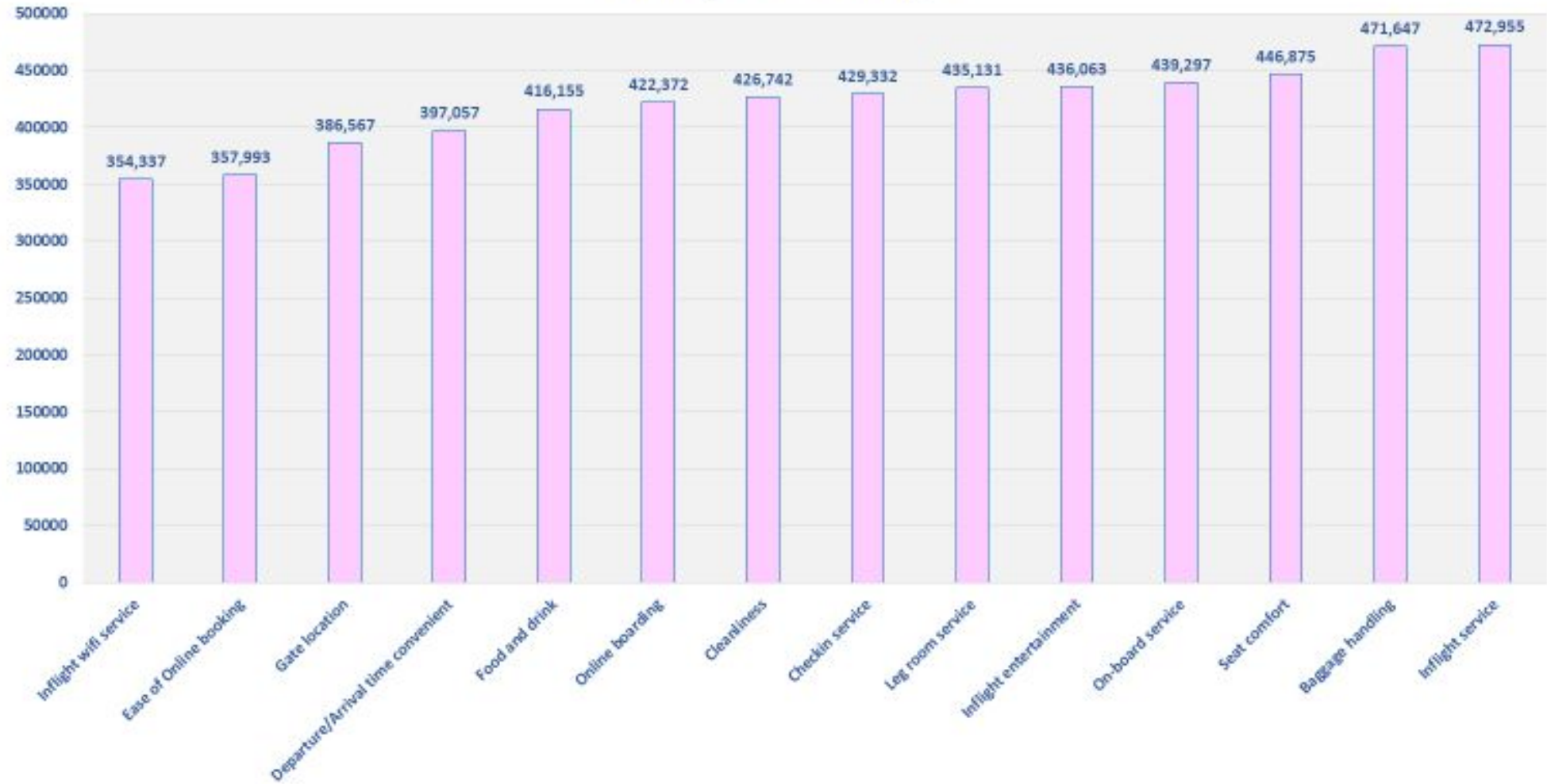Deploying a machine learning algorithm to predict customer satisfaction
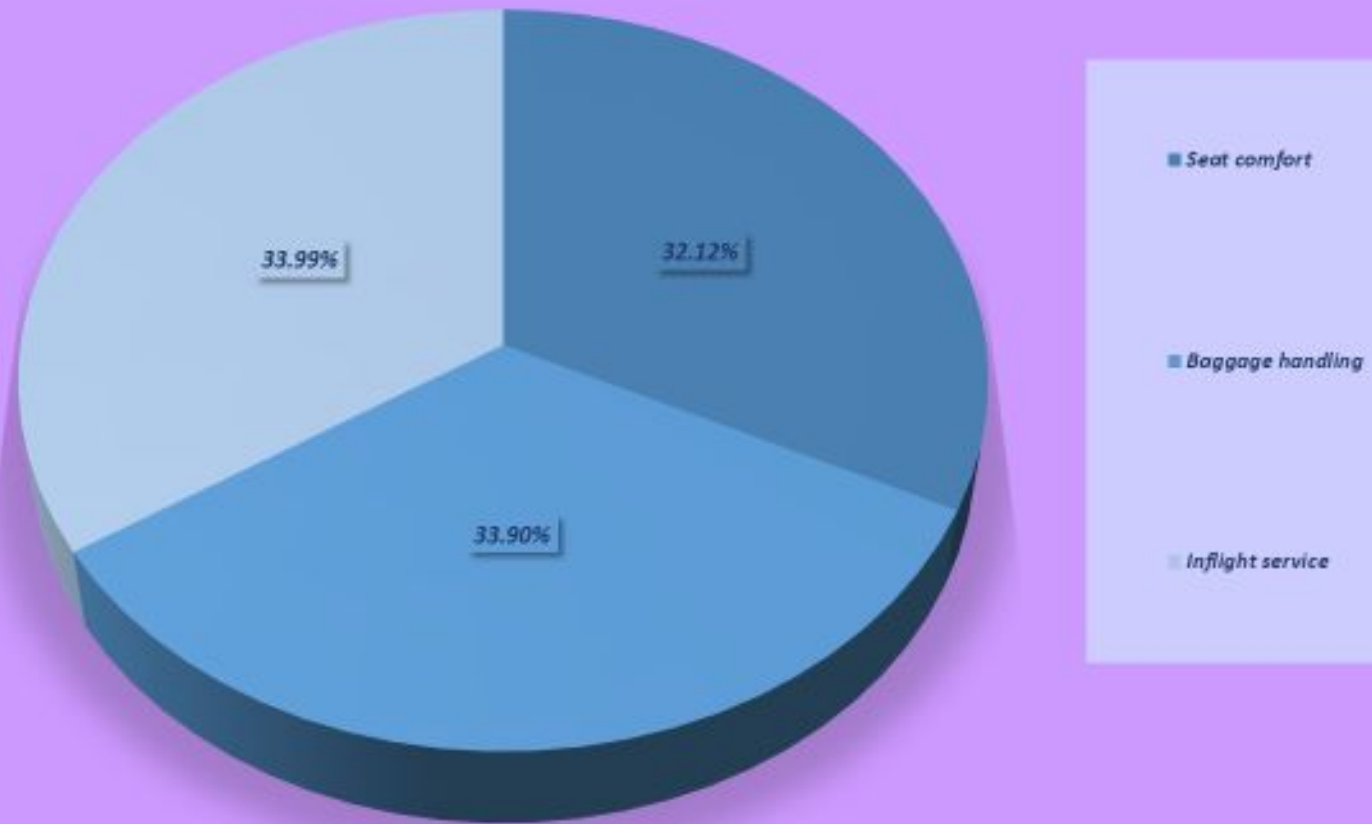
TECHNOLOGIES

OUR TOOLKIT

DATA VISUALIZATION

Count of 4's & 5's - Satisfied

Top3 Categories of Importance to Customers

- Seat comfort — 32.12%
- Baggage handling — 33.90%
- Inflight service — 33.99%

Count of 1's & 2's - Neutral or Dissatisfied
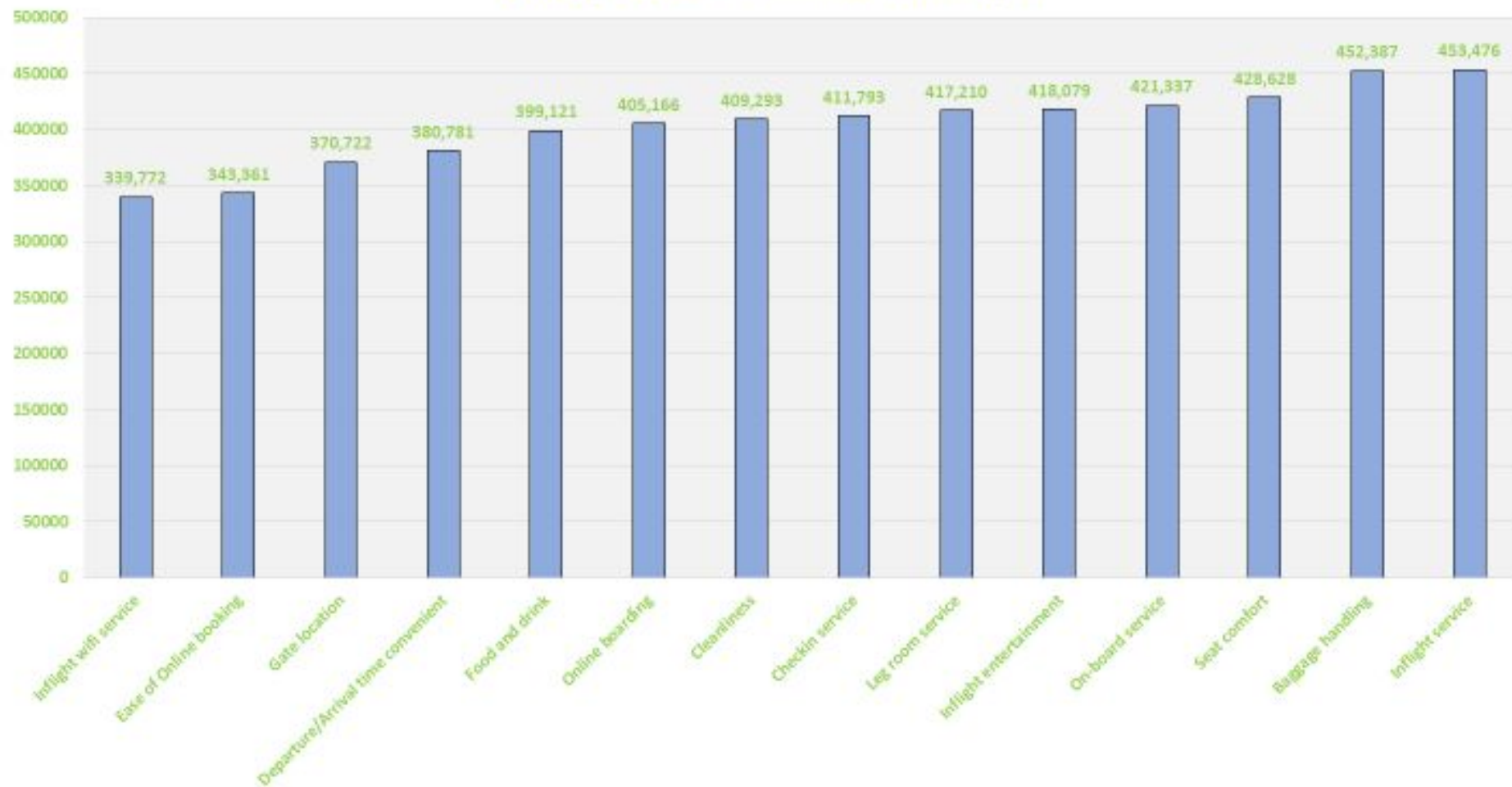
Average Age of Men Related to Leg Room Comfort
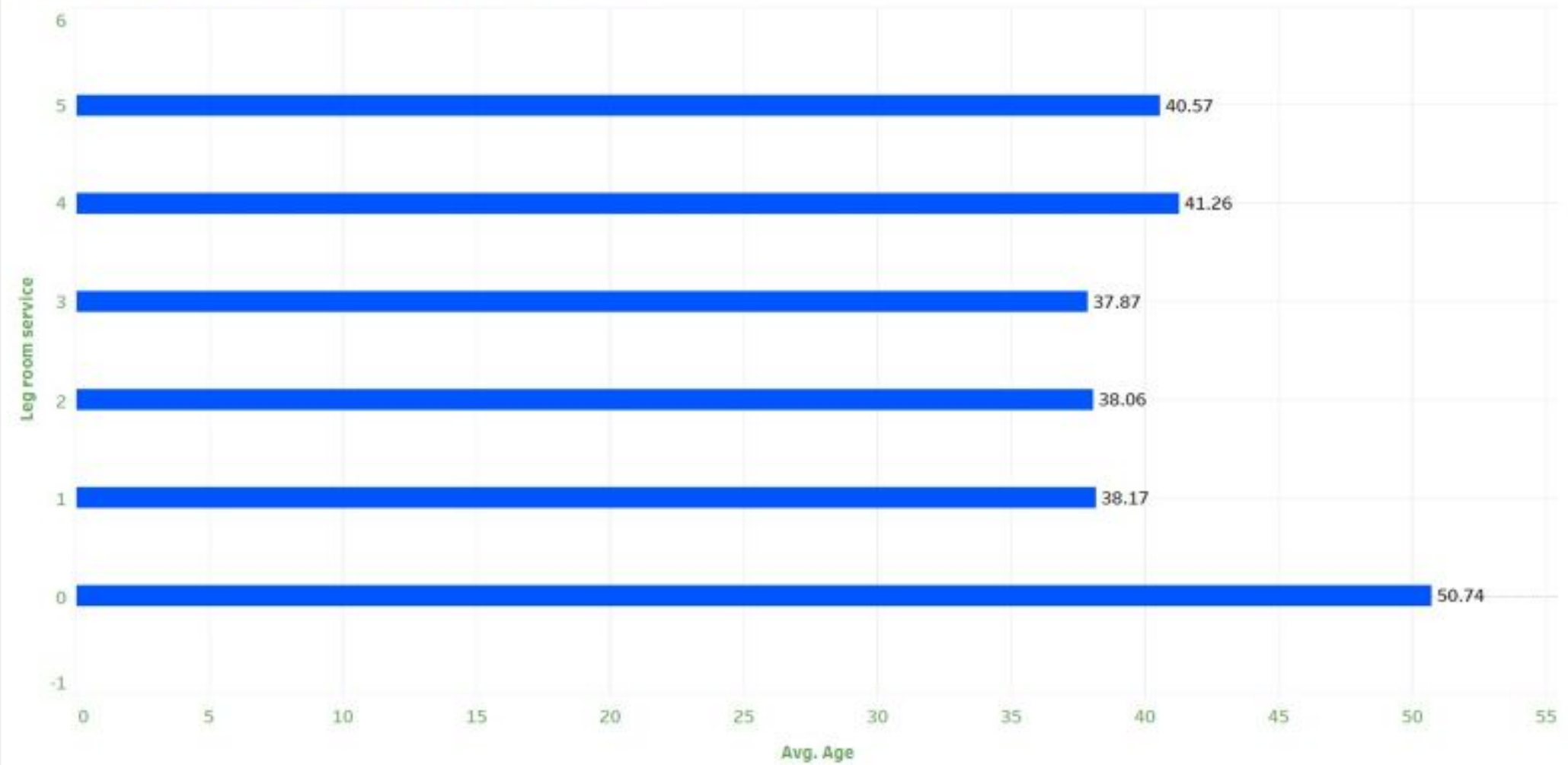
MACHINE LEARNING

Report № 1: Logistic Regression
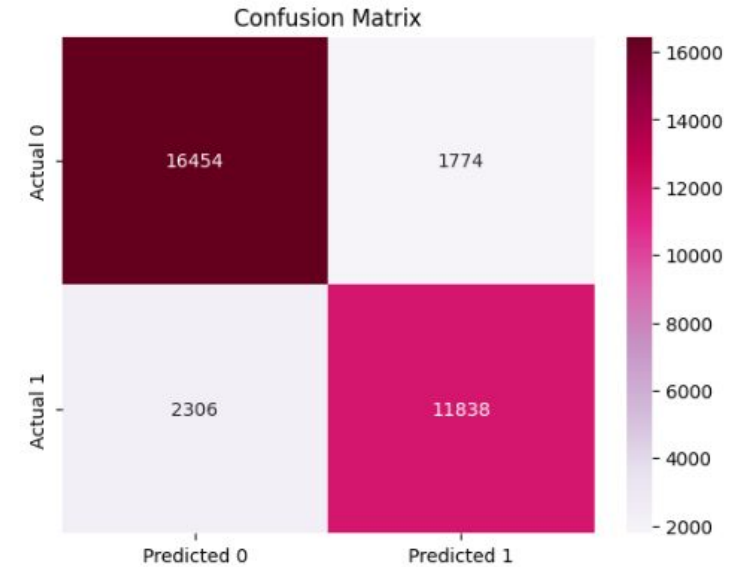
```
Confusion Matrix
          Predicted 0  Predicted 1
Actual 0      16571        1816
Actual 1       2334        11749
Accuracy Score : 0.8721897135817678
Classification Report
           precision    recall  f1-score   support

        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
```
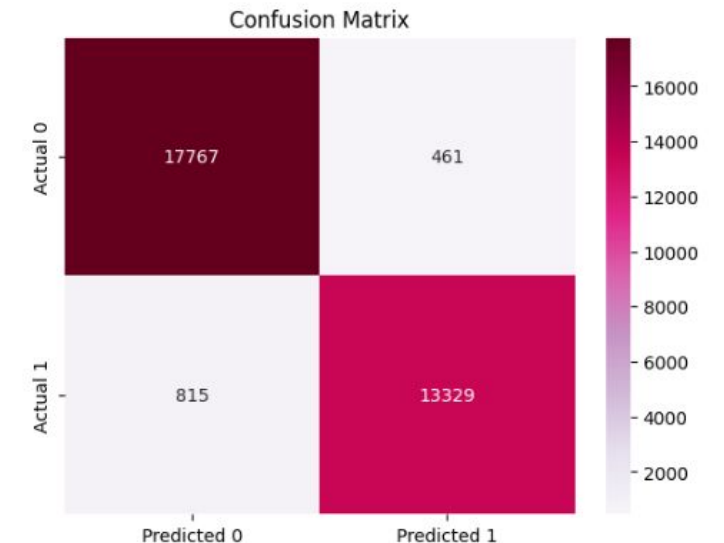
Report № 2: Neural Network

```
# 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
```
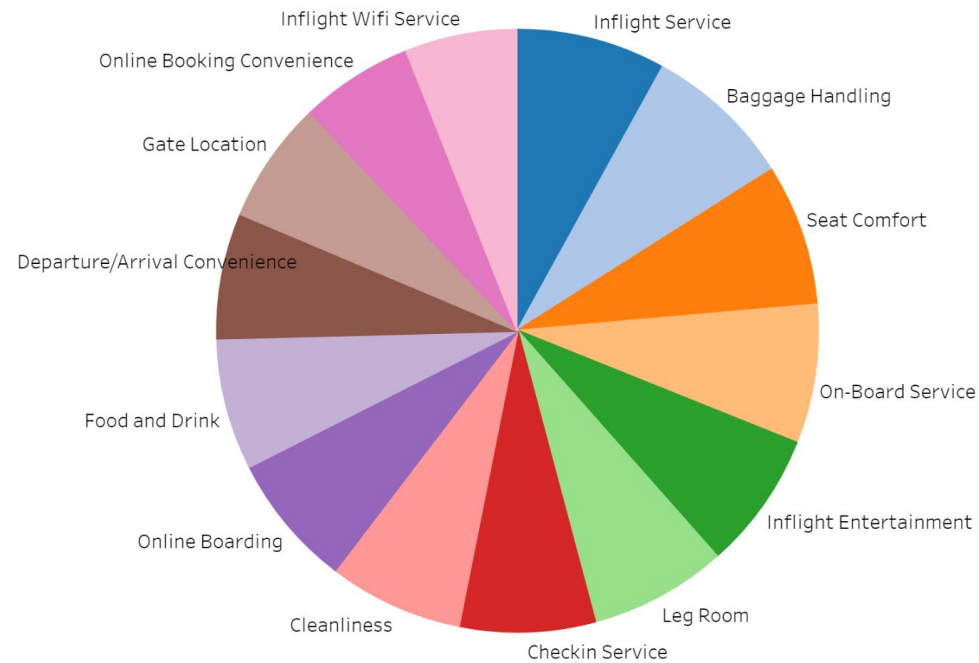

Confusion Matrix

Confusion Matrix

Recommendations

➢ Machine Learning Accuracy
➢ Top and Bottom Parameters

Pie chart segments (clockwise from top): Inflight Wifi Service, Inflight Service, Baggage Handling, Seat Comfort, On-Board Service, Inflight Entertainment, Leg Room, Checkin Service, Cleanliness, Online Boarding, Food and Drink, Departure/Arrival Convenience, Gate Location, Online Booking Convenience
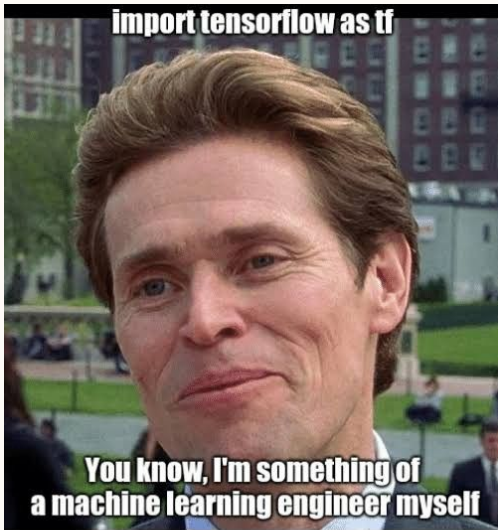
Top 3 Parameters:
Inflight Service
Baggage Handling
Seat Comfort

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

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.

THANK YOU

Me watching my deep learning model train