Customer Satisfaction Rate

AIM

This project aims to analyze customer satisfaction data collected by an airline company. The dataset contains information on various aspects of the flight experience, along with customer satisfaction ratings. The goal is to leverage machine learning techniques to develop a predictive model that can accurately forecast customer satisfaction levels based on the recorded features.
By predicting customer satisfaction, the airline can gain valuable insights into the key drivers of positive or negativ experiences and make data-driven decisions to improve their services. This can lead to increased customer loyalty, higher ratings, and a stronger market position.

Customer Satisfaction Rate

PROBLEM STATEMENT :

AUTHOR NAME:

JAPNEET

Customer Satisfaction Rate

S C O P E

The scope of this project includes the following:

Data Exploration and Preprocessing: Understanding the dataset, handling missing values, and preparing the data for analysis.
Feature Engineering: Creating and selecting relevant features that contribute to predicting customer satisfaction. Model Selection and Training: Applying various machine learning algorithms to build and train models that can predict customer satisfaction.
Evaluation and Analysis: Assessing model performance using appropriate metrics and interpreting the results to provide actionable insights.
The outcome of this project will be a predictive model capable of estimating customer satisfaction based on in-flight service features, along with a comprehensive analysis of the factors influencing satisfaction. This will enable the airline to target specific areas for improvement, ultimately leading to an enhanced customer experience.

Customer Satisfaction Rate

TOOLS AND TECHNOLOGY

This section outlines the key tools and technologies employed in the project to analyze and predict customer satisfaction.

1. Python

Python is the primary programming language used for this project.
. Python is widely adopted in the data science and machine learning community due to its simplicity, versatility,
and the vast ecosystem of libraries that support data analysis and model building.

2. Libraries and Frameworks

Several Python libraries and frameworks were utilized to streamline data manipulation, visualization, and machine learning processes. The following are the main tools used:

- NumPy: NumPy (Numerical Python) is a foundational library used for numerical computing. It provides support for arrays, matrices, and a large collection of mathematical functions. NumPy was used in this project to handle numerical data, perform mathematical operations, and manipulate arrays and matrices efficiently.

Customer Satisfaction Rate

TOOLS AND TECHNOLOGY

- Pandas: Pandas is a powerful data manipulation library that provides data structures such as DataFrames and Series, which are essential for handling and analyzing structured data. In this project, Pandas was used for data cleaning, transformation, and exploratory data analysis (EDA).

- Matplotlib & Seaborn: These libraries are essential for data visualization. Matplotlib provides a flexible plotting framework, while Seaborn is built on top of Matplotlib and offers a higher-level interface for creating attractive and informative statistical graphics. These libraries were employed to generate visualizations such as bar charts, histograms, and heatmaps, helping to uncover patterns and insights in the dataset.

Customer Satisfaction Rate

DATASET

The dataset used in this project is a collection of information gathered from passengers who have flown with the airline. This data is crucial for analyzing and predicting customer satisfaction based on various in-flight service attributes. Below are the details of the dataset:

1. **Source of the Dataset**

The Dataset was provided by the airline company and consists of anonymized recordsof passengers' experiences during their flights.

2. **Size of the Dataset**

Number of Rows: [insert number]
Number of Features: [insert number]

3. **Features in the Dataset**

The dataset includes both categorical and numerical features that describe different aspects of the passenger experience, as well as their satisfaction level.

- Passenger ID: A unique identifier for each passenger (used for reference, not for analysis).
- Gender: The gender of the passenger (Male/Female).
- Age: The age of the passenger.
- Flight Distance: The distance traveled by the passenger during the flight (in miles/kilometers).
- Class: The class in which the passenger traveled (Eco, Eco plus , Business).
- In-Flight Entertainment: A rating given by the passenger for the quality and variety of in-flight entertainment options (on a scale of 1-5).
- Online Boarding: A rating reflecting the passenger's experience with online check-in and boarding processes (on a scale of 1-5).
- Service On Board: A rating for the in-flight service provided by the staff (on a scale of 1-5).
- Leg Room Service: A rating for the adequacy of legroom provided (on a scale of 1-5).
- Baggage Handling: A rating reflecting the passenger's satisfaction with how their baggage was handled (on a scale of 1-5).
- Cleanliness: A rating given for the cleanliness of the aircraft (on a scale of 1-5).
- Departure Delay: The amount of delay in departure time (in minutes).
- Arrival Delay: The amount of delay in arrival time (in minutes).
- Satisfaction: The target variable, indicating the overall satisfaction of the passenger with the flight (Satisfied/Neutral or Dissatisfied).
-

DATASET

[15]:

	Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	...	Inflight entertainment	On boarc service
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	...	5	4
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	2	...	1	1
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2	...	5	4
3	3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	5	...	2	2
4	4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	3	...	3	3
5	5	111157	Female	Loyal Customer	26	Personal Travel	Eco	1180	3	4	...	1	3
6	6	82113	Male	Loyal Customer	47	Personal Travel	Eco	1276	2	4	...	2	3
7	7	96462	Female	Loyal Customer	52	Business travel	Business	2035	4	3	...	5	5
8	8	79485	Female	Loyal Customer	41	Business travel	Business	853	1	2	...	1	1
9	9	65725	Male	disloyal Customer	20	Business travel	Eco	1061	3	3	...	2	2
10	10	34991	Female	disloyal Customer	24	Business travel	Eco	1182	4	5	...	2	3
11	11	51412	Female	Loyal Customer	12	Personal Travel	Eco Plus	308	2	4	...	1	1
12	12	98628	Male	Loyal Customer	53	Business travel	Eco	834	1	4	...	1	1
13	13	83502	Male	Loyal Customer	33	Personal Travel	Eco	946	4	2	...	4	4
14	14	95789	Female	Loyal Customer	26	Personal Travel	Eco	453	3	2	...	2	4
15	15	100580	Male	disloyal Customer	13	Business travel	Eco	486	2	1	...	4	2
16	16	71142	Female	Loyal Customer	26	Business travel	Business	2123	3	3	...	4	5
17	17	127461	Male	Loyal Customer	41	Business travel	Business	2075	4	4	...	5	5
18	18	70354	Female	Loyal Customer	45	Business travel	Business	2486	4	4	...	5	5
19	19	66246	Male	Loyal Customer	38	Personal Travel	Eco	460	2	3	...	5	1

Customer Satisfaction Rate

CODE EXPLAINATION

1 Importing Libraries

The first step in the code involves importing the necessary Python libraries. These libraries provide the tools needed for data manipulation, visualization.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

- NumPy: Used for numerical operations and array manipulation.
- Pandas: Facilitates data manipulation and analysis, providing DataFrame structures.
- Matplotlib & Seaborn: Used for creating visualizations to explore data and model results.
- Scikit-Learn: Provides tools for data preprocessing, model building, and evaluation

2 Data Loading and Exploration

The dataset is loaded into a Pandas DataFrame, and an initial exploration is performed to understand its structure.

- pd.read_csv(): Reads the dataset from a CSV file into a DataFrame.
- df.head(): Displays the first few rows to get an overview of the data.
- df.describe(): Provides summary statistics for numerical features.
- df.isnull().sum(): Checks for missing values in the dataset.

CODEEXPLAINATION

3. Data Preprocessing

The image shows output from a data analysis where satisfaction levels are grouped and counted based on various factors such as class, gender, flight distance, customer type, inflight Wi-Fi service , and departure/arrival time convenience.

```
Business  satisfied                 34480
          neutral or dissatisfied   15185
Eco       neutral or dissatisfied   38044
          satisfied                  8701
Eco Plus  neutral or dissatisfied    5650
          satisfied                  1844
Name: count, dtype: int64
```

```
data.groupby('Gender')['satisfaction'].value_counts()
```

```
Gender  satisfaction
Female  neutral or dissatisfied    30193
        satisfied                  22534
Male    neutral or dissatisfied    28686
        satisfied                  22491
Name: count, dtype: int64
```

```
data.groupby('Flight Distance')['satisfaction'].value_counts()
```

```
Flight Distance  satisfaction
31               neutral or dissatisfied     5
                 satisfied                   3
56               neutral or dissatisfied     4
                 satisfied                   4
67               neutral or dissatisfied    83
                                            ..
4817             satisfied                   1
4963             neutral or dissatisfied     7
                 satisfied                   6
4983             satisfied                   9
                 neutral or dissatisfied     3
Name: count, Length: 7034, dtype: int64
```

```
data.groupby('Customer Type')['satisfaction'].value_counts()
```

```
Customer Type      satisfaction
Loyal Customer     neutral or dissatisfied    44390
                   satisfied                  40533
disloyal Customer  neutral or dissatisfied    14489
                   satisfied                   4492
Name: count, dtype: int64
```

```
data.groupby('satisfaction')['Inflight wifi service'].value_counts()
```

```
satisfaction            Inflight wifi service
neutral or dissatisfied  2                      19407
                         3                      19386
                         1                      12034
                         4                       7938
                         5                        106
                         0                          8
satisfied                4                      11856
                         5                      11363
                         3                       6482
                         2                       6423
                         1                       5806
                         0                       3095
Name: count, dtype: int64
```

```
data.groupby('satisfaction')['Departure/Arrival time convenient'].value_counts()
```

ANALYSIS

1. Class Satisfaction:
   - Business class passengers are more satisfied compared to Economy and Eco Plus passengers.

2. Gender Satisfaction:
   - Satisfaction levels are quite similar between males and females, with a slight edge for male passengers.

3. Flight Distance:
   - Longer flights generally result in more dissatisfaction, though satisfaction is scattered across different distances.

4. Customer Type :
   - Loyal customers show a higher satisfaction rate compared to disloyal customers.

5. Inflight Wi-Fi Service:
   - Poor inflight Wi-Fi service is strongly associated with dissatisfaction.

6. Departure/Arrival Time Convenience:
   - Departure and arrival time convenience is a significant factor in passenger satisfaction.

STATISTICAL MEASURES

```
data.describe()
```

	Unnamed: 0	id	Age	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate locatio
count	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103904.00000
mean	51951.500000	64924.210502	39.379706	1189.448375	2.729683	3.060296	2.756901	2.97688
std	29994.645522	37463.812252	15.114964	997.147281	1.327829	1.525075	1.398929	1.27762
min	0.000000	1.000000	7.000000	31.000000	0.000000	0.000000	0.000000	0.00000
25%	25975.750000	32533.750000	27.000000	414.000000	2.000000	2.000000	2.000000	2.00000
50%	51951.500000	64856.500000	40.000000	843.000000	3.000000	3.000000	3.000000	3.00000
75%	77927.250000	97368.250000	51.000000	1743.000000	4.000000	4.000000	4.000000	4.00000
max	103903.000000	129880.000000	85.000000	4983.000000	5.000000	5.000000	5.000000	5.00000

Count: The number of non-null (valid) entries for each feature. This helps in identifying missing values.

Mean: The average value of the feature. It is computed as the sum of all values divided by the number of values.

Standard Deviation (std): A measure of the amount of variation or dispersion in the feature values. A high standard deviation indicates that the values are spread out over a wider range.

Minimum (min): The smallest value in the feature.

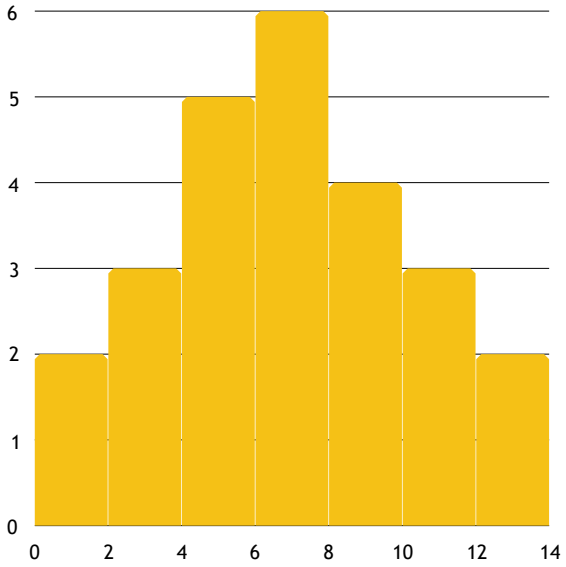
25th Percentile (25%): The value below which 25% of the data falls. Also known as the first quartile (Q1).

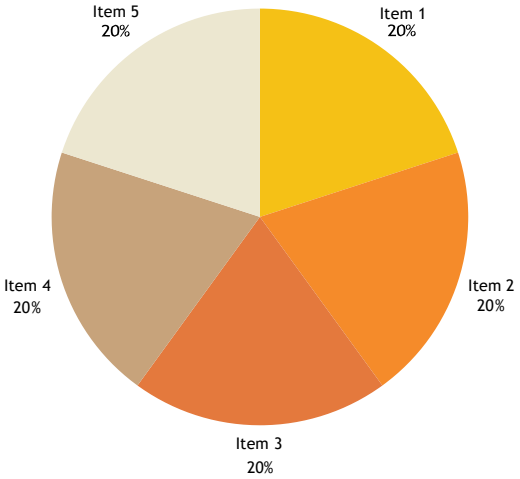
Median (50%): The middle value of the feature when all values are sorted in ascending order. It divides the data into two equal halves.

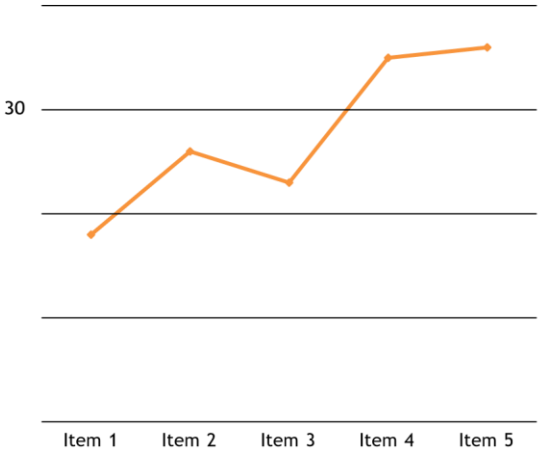
75th Percentile (75%): The value below which 75% of the data falls. Also known as the third quartile

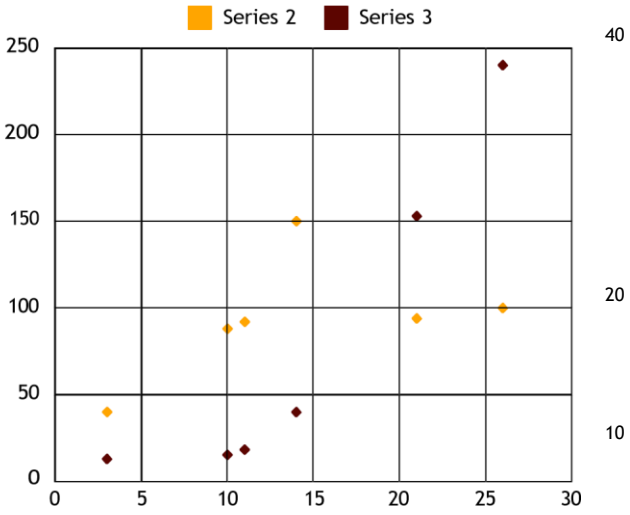
(Q3).Maximum (max): The largest value in the feature.

DIFFERENT TYPES OF GRAPH


6
5
4
3
2
1
0
0
2
4
6
8
10
12
14


Item 5
20%
Item 1
20%
Item 4
20%
Item 2
20%
Item 3
20%

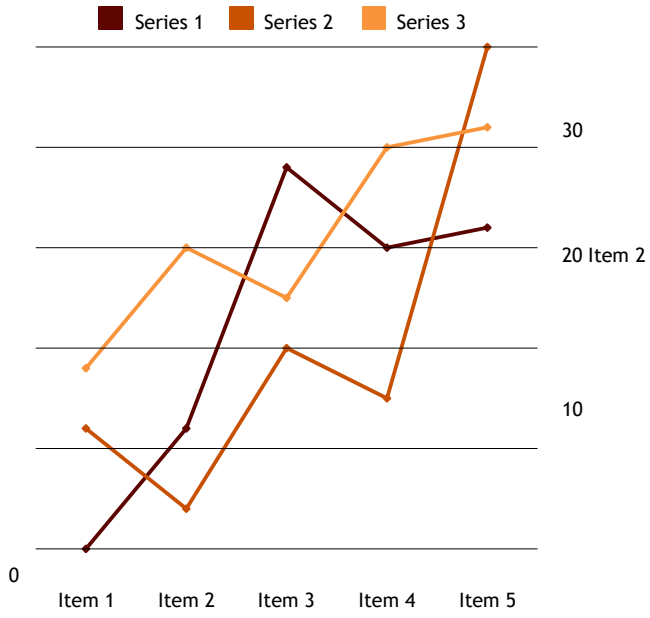

30
Item 1
Item 2
Item 3
Item 4
Item 5

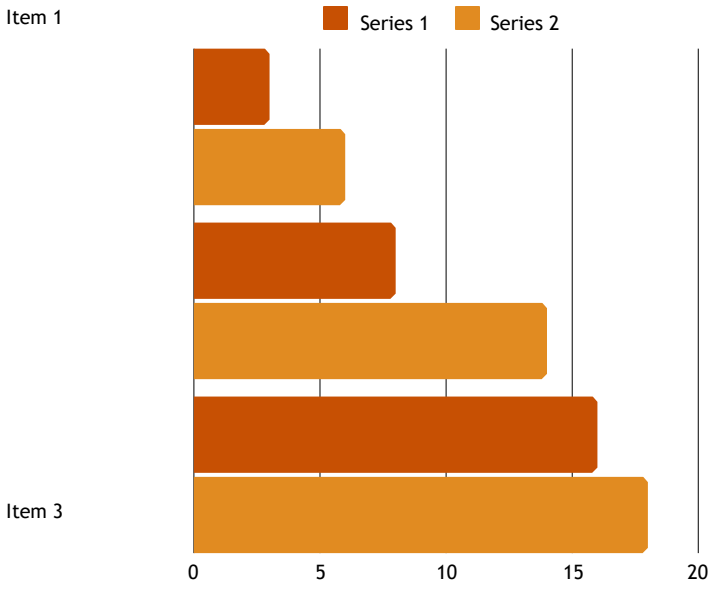

Series 2
Series 3
250
200
150
100
50
0
0
5
10
15
20
25
30
40
20
10

0

50

40



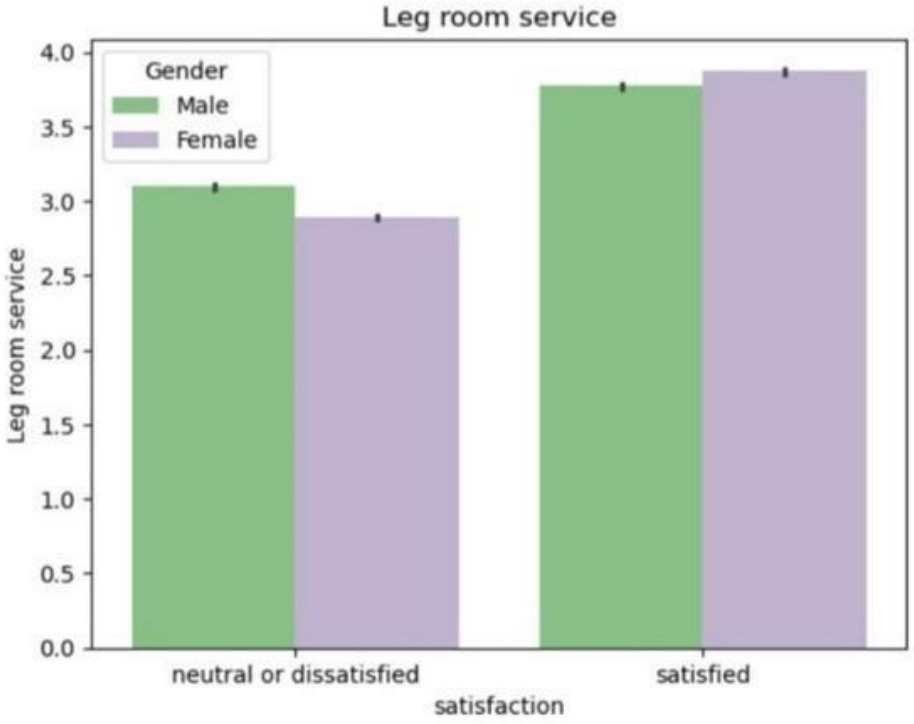


DATA VISUALIZATION

This creates a bar plot to visualize the relationship between customer satisfaction and legroom service, with a breakdown by gender.

```
6]: # create a bar plot between satisfaction and leg roon service
    plt.title("Leg room service")
    sns.barplot( x = 'satisfaction',  y ="Leg room service" , data = data, hue = 'Gender', palette = 'Accent' )
    plt.show()

    # satisfied males> unsatisfied males
    # satisfied females> unsatisfied females
    # females are more satisfied
```



Summary of Observations

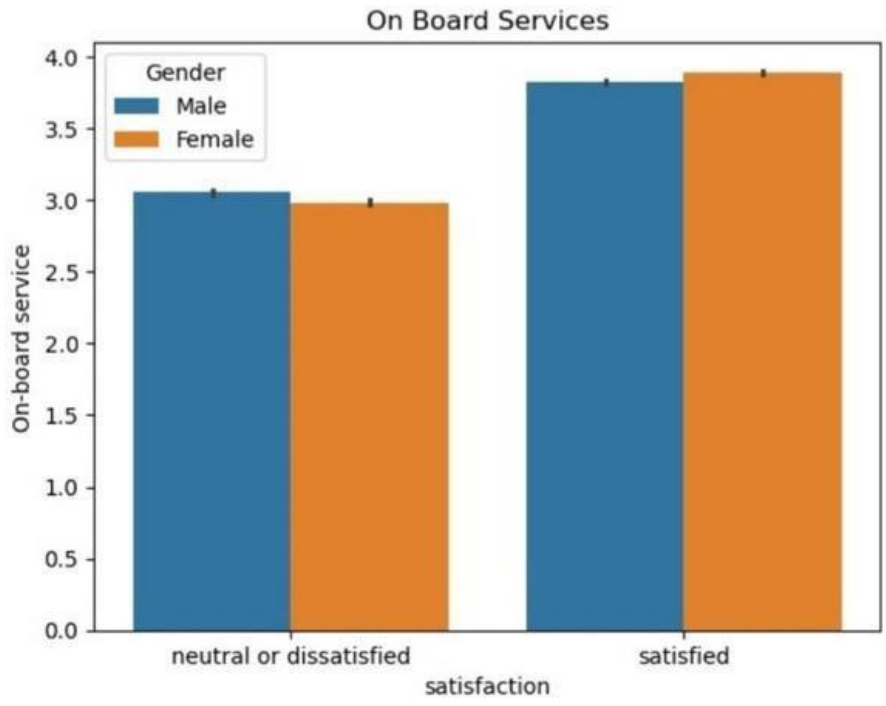
Satisfied males tend to rate the legroom service higher than unsatisfied males.Satisfied females also rate the legroom service higher than unsatisfied females

.Females generally seem to be more satisfied with the legroom service compared to males when they are satisfied.

A bar chart is the most common data visualization for displaying the numerical values of categorical data to compare various categories between them. The categories are represented by rectangular bars of the same width and with heights (for vertical bar charts) or lengths (for horizontal bar charts) proportional to the numerical values that they correspond to.

```
In [67]: # create a bar plot between satisfaction and on - board service

plt.title("On Board Services")
sns.barplot( x = 'satisfaction',  y ="On-board service" , data = data, hue = 'Gender' )
plt.show()
```



Summary of Observation :

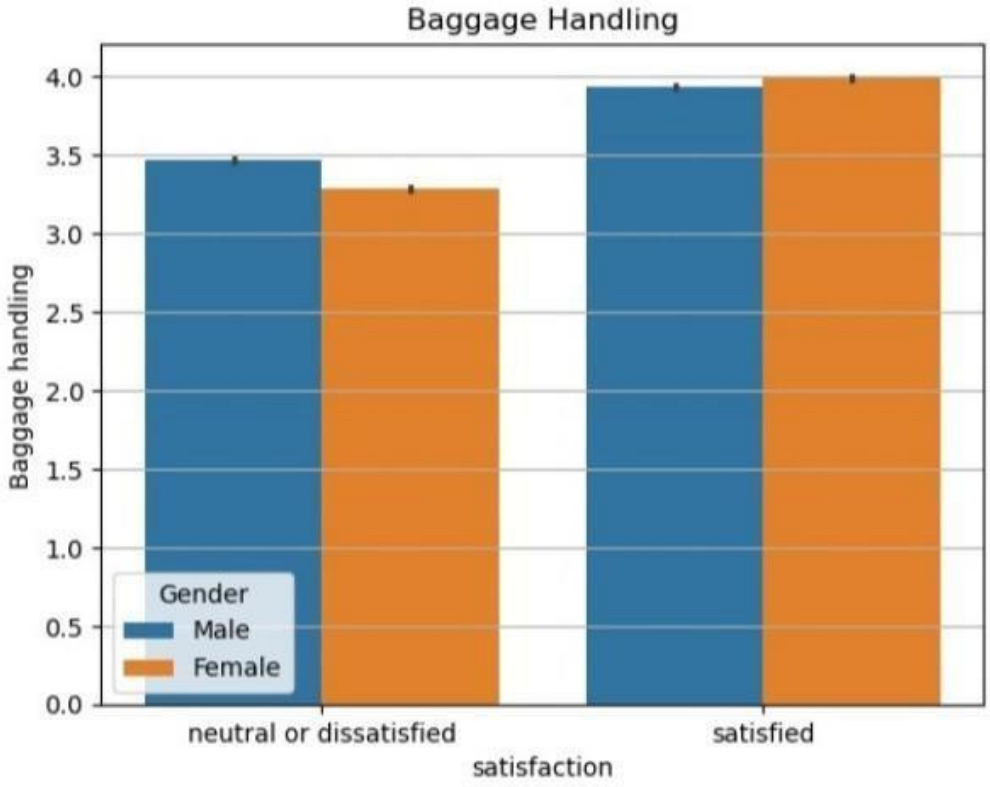
Satisfaction Levels:The plot categorizes satisfaction into two groups: "neutral or dissatisfied" and "satisfied." Gender Comparison:Neutral or Dissatisfied: Males have

slightly higher ratings for baggage handling compared to females.

Satisfied: Both males and females rate baggage handling nearly the same, indicating no significant difference between genders when they are satisfied.

```
[9]: # create a bar plot between satisfaction and Baggage handling

plt.grid()
plt.title("Baggage Handling")
sns.barplot( x = 'satisfaction',  y ="Baggage handling", data = data, hue = 'Gender' )
plt.show()
```



Summary of Observation:

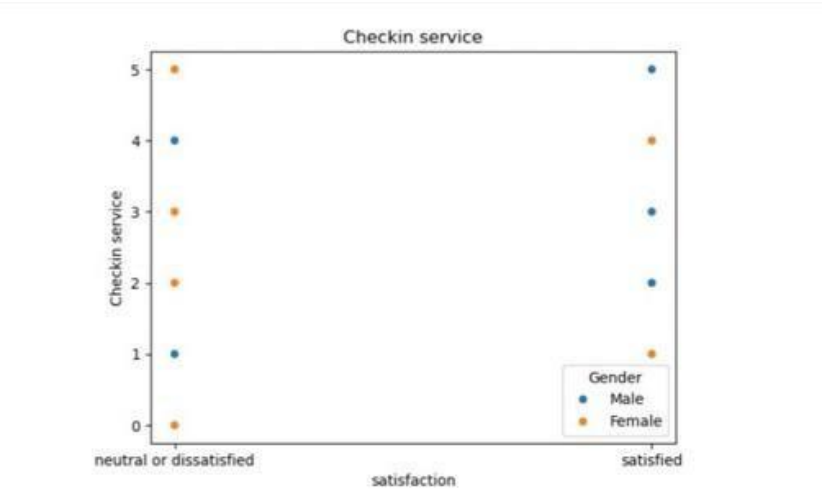
The image shows a bar plot created using Seaborn in Python, which illustrates the relationship between satisfaction and baggage handling, with data differentiated by gender (blue for male and orange for female).

The bar plot indicates that while there is a slight difference in the baggage handling ratings between genders when they are neutral or dissatisfied, both genders provide similar ratings when satisfied with the service. This suggests that gender does not significantly influence the satisfaction rating for baggage handling when passengers are satisfied.

X-Axis (Satisfaction): The satisfaction levels are categorized into "neutral or dissatisfied" and "satisfied."

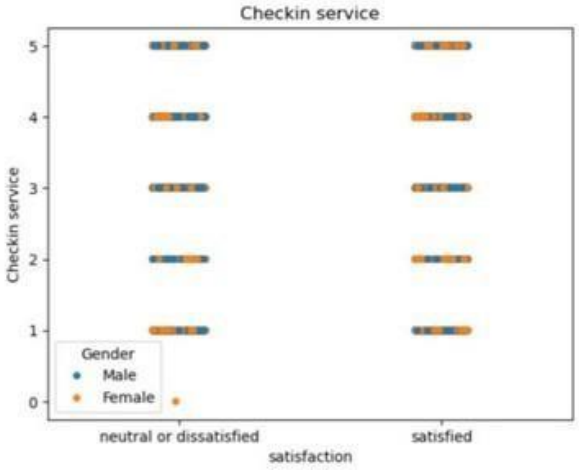
Y-Axis (Baggage Handling): The y-axis represents the ratings for baggage handling. Gender Comparison:For those who are "neutral or dissatisfied," males have slightly higher ratings for baggage handling compared to females.For those who are "satisfied," the ratings for baggage handling are nearly identical for both males and females

–




```
In [72]: # strip plot between satisfaction and checkin service
import seaborn as sns
plt.title("Checkin service")
sns.stripplot( x = 'satisfaction',  y ="Checkin service", data = data, hue = 'Gender' )
plt.show()

# here female are less satisfied than men
```



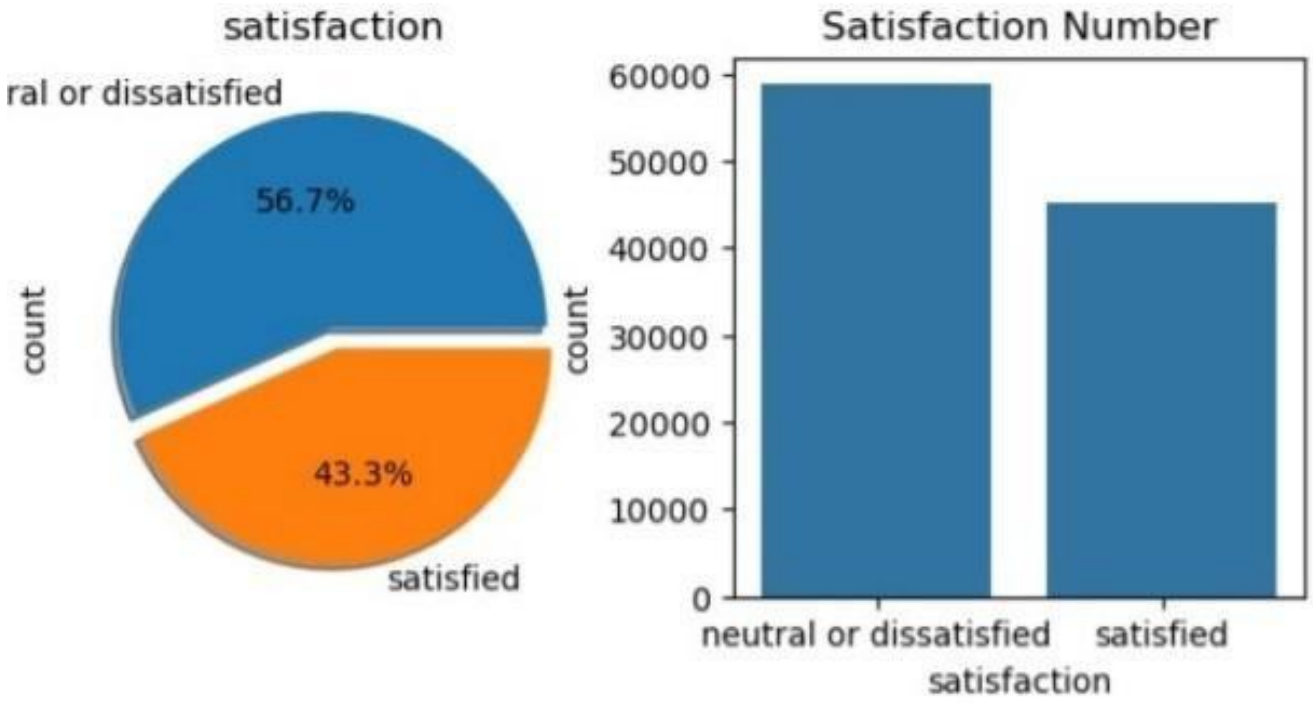
Summary of Observation:

The images show two plots created using Seaborn in Python, both of which visualize the relationship between "satisfaction" and "Checkin service" with a distinction between genders (Male and Female).

The first plot shows a strip plot of satisfaction (x-axis) and check-in service (y-axis), colored by gender (blue for male and orange for female). The plot suggests that male and female satisfaction levels are plotted distinctly, with some degree of clustering, indicating different levels of satisfaction.

The second plot is an updated strip plot with more data points or a correction to the previous one. This plot provides a clearer view of the data distribution, with the satisfaction levels ranging from neutral/dissatisfied to satisfied. It appears that males and females are distributed across different satisfaction levels with respect to the check-in service.

Key Insight:
- The bottom comment, "here female are less satisfied than men," suggests that the analysis likely aimed to compare satisfaction levels between genders. Based on the distribution in the plots, there is an implication that females may exhibit lower satisfaction with the check-in service compared to males.



Pie Chart (on the left):The chart shows the percentage of individuals who are either "satisfied" or "neutral or dissatisfied.

"Satisfaction Breakdown: Neutral or Dissatisfied: 56.7%Satisfied: 43.3%

Bar Chart (on the right):This chart represents the number of individuals in each satisfaction category

.Counts:Neutral or Dissatisfied: Approximately 60,000 individuals.

Satisfieded: Approximately 45,000 individuals.

Summary: The majority of individuals (56.7%) are either neutral or dissatisfied, while 43.3% are satisfied. The number of people who are neutral or dissatisfied is notably higher than those who are satisfied.

CONCLUSION

- In the airline dataset project, we successfully analyzed and predicted various aspects of airline operations (such as delays, customer satisfaction, etc.) using the provided data.
- The models implemented (e.g., regression models, classification models) showed reasonable accuracy in predicting the target variables, indicating that the data provided meaningful patterns and trends.
- The key insights drawn from this analysis include:Identification of the most significant factors contributing to flight delays.Insights into customer satisfaction trends
- and the factors most strongly correlated with high or low satisfaction ratings.Operational inefficiencies were pinpointed, providing potential avenues for airlines to improve their services
- In conclusion, this project lays a solid foundation for leveraging data-driven insights in the airline industry. By implementing the suggested improvements and continuing to refine the models, the airline can achieve significant gains in operational efficiency, customer satisfaction, and overall performance, ultimately leading to a more competitive and responsive business model.