**Air line Passenger statisfaction**

**Data Description :-**

**Gender**: Gender of the passengers (Female, Male)

**Customer Type**: The customer type (Loyal customer, disloyal customer)

**Age**: The actual age of the passengers

**Type of Travel**: Purpose of the flight of the passengers (Personal Travel, Business Travel)

**Class**: Travel class in the plane of the passengers (Business, Eco, Eco Plus)

**Flight distance**: The flight distance of this journey

**Inflight wifi service**: Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)

**Departure/Arrival time convenient**: Satisfaction level of Departure/Arrival time convenient

**Online booking**: Satisfaction level of online booking

**Gate location**: Satisfaction level of Gate location

**Food and drink**: Satisfaction level of Food and drink

**Online boarding**: Satisfaction level of online boarding

**Seat comfort**: Satisfaction level of Seat comfort

**Inflight entertainment**: Satisfaction level of inflight entertainment

**On-board service**: Satisfaction level of On-board service

**Leg room service**: Satisfaction level of Leg room service

**Baggage handling**: Satisfaction level of baggage handling

**Check-in service**: Satisfaction level of Check-in service

**Inflight service**: Satisfaction level of inflight service

**Cleanliness:** Satisfaction level of Cleanliness

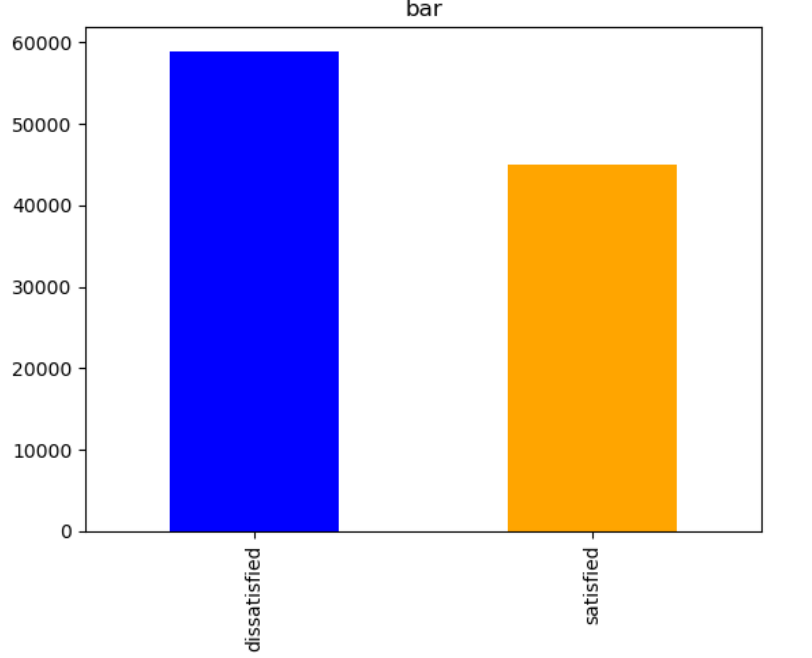
**Departure Delay in Minutes**: Minutes delayed when departure

**Arrival Delay in Minutes**: Minutes delayed when Arrival

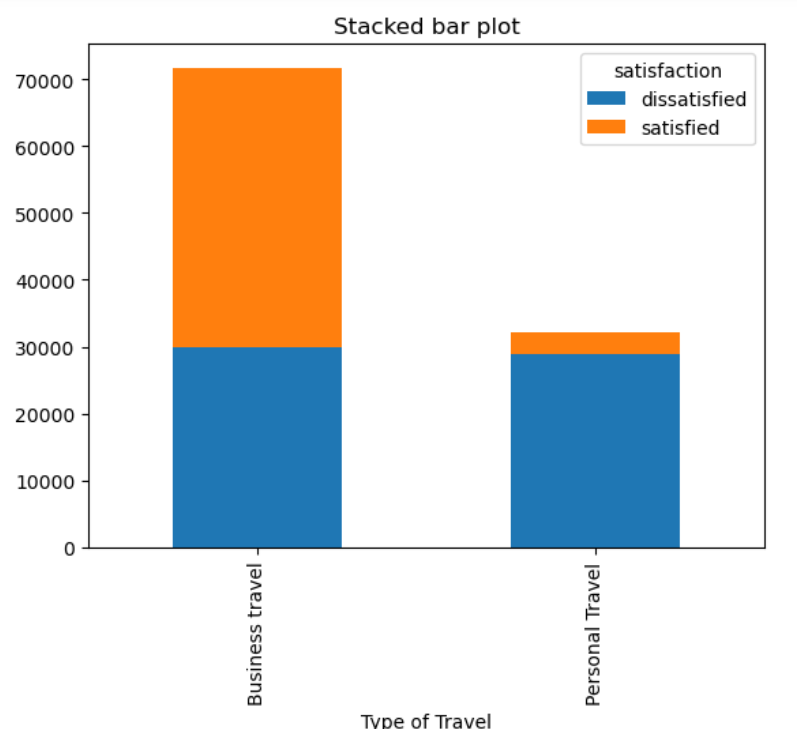
**Satisfaction**: Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

**Exploratory data analysis:-**

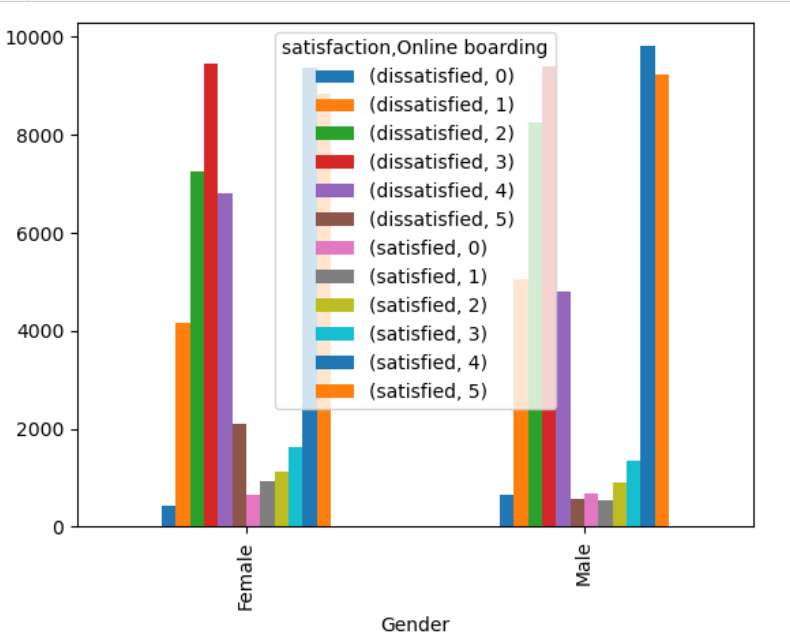
**Satisfaction**:- This is target feature we see most of Passenger are dissatisfied as we below in bar



**Type of Travel and statisfaction :-** Here this columns tell us passenger are travel for bussinees puropose or personal. Most of the Business travel passengers are statisfied and very less passengers who are traveling for personal purpose are less statisfied.



**Gender and Online boarding and statisfaction: -** Most of the passengers are gives satisfied with “0” rating service of air line in online boarding .



**Data Cleaning :-**

* In the data set contains 103903 rows and 24 columns.
* **Arrival Delay in Minutes** columns contains 310 “Null values” so here filling of null values using median because of median not effected by outliers . so perfered filling null values with median .
* In this data set contains 4 columns “**Flight Distance”,”Checkin service”**,” **Departure Delay in Minutes**”, “**Arrival Delay in Minutes”** outlilers and treatments using IQR Method

**Model Building :-**

**Objective of Project :-**  Building a ann model to predict whether a passenger is satisfied or dis satisfied with services of air lines.

Deep learning model building is utilized to leverage the power of artificial neural networks, which can capture complex patterns and relationships in data. This is particularly valuable for tasks where traditional machine learning models may struggle, such as image and speech recognition, natural language processing, and other intricate, high-dimensional data problems. Deep learning models can automatically learn and extract features from data, making them highly effective for a wide range of tasks.

**Data Splitting:** Data was split into training and testing sets using a 80-20% ratio (train-test split) to evaluate the model’s performance. This separation ensures that the model is trained on one subset of the data and tested on an independent subset, allowing us to assess its generalization capabilities and avoid overfitting

**Data Preprocessing:**

**Separating Numerical and Categorical Columns:** In the initial data, we had a dataset with 86959 rows and 23 columns. To prepare the data for deep learning, we first separated the dataset into two types of columns: numerical and categorical. This separation is crucial because different preprocessing techniques are applied to each type.

**Preprocessing the data:** Before training and evaluating any model, it’s essential to preprocess both the training and test data. This ensures that the model performs optimally and consistently across all datasets. The preprocessing steps for both the training and test data include the following:

**Scaling the Numerical Columns:** We standardized the numerical columns in both the training and test data using the StandardScaler from the scikit-learn library. Standardization transforms the data to have a mean of 0 and a standard deviation of 1. This step is crucial because it ensures that all numerical features have the same scale, preventing some features from dominating the learning process.

**One-Hot Encoding for Categorical Features:** For the categorical columns in both the training and test data, we used one-hot encoding. One-hot encoding coverts categorical variables into binary (0/1) format, where each category becomes a separate binary column. This technique ensures that the deep learning algorithm can work with categorical data effectively.

**Label Encoding for Discrete Numerical Features**:- For the numerical discrete columns in both the training and test data, we used label encoding. Label encoding coverts categorical variables .

**Concatenating Numerical and Categorical Features:** After scaling and one-hot encoding, we concatenated the processed numerical and categorical features back together for both the training and test datasets.

**Balancing the data:** In deep learning, an imbalanced dataset can lead to model bias, where the model may perform poorly on the minority class. To address this issue, we used the Synthetic Majority Over-Sampling Technique (SMOTE) to balance the dataset.  **Building logical structure using keras Tuners:-**

* Keras Tuner is a library that automates the process of hyper parameter tuning, helping to find the optimal configuration for your deep learning models.And it will randomly initializers weights and activation function .
* The final model was trained on the preprocessed dataset with 10 as the batch size and epochs are 40 and a validation split of 10% was used for to check performance of model .
* After doing predictions on test data used accuracy ,confusion matrix evaluate of model .
* The model is giving accuracy of 88%.
* The model's performance was further evaluated using a confusion matrix and it was to maximize the diagonal elements of the matrix, indicating correct predictions.
* In confusion matrix diagonal elements count high means the model's capability to make accurate classifications.

