

Customer Satisfaction Prediction Report

1. Project Overview

The goal of this project is to predict Customer Satisfaction Ratings based on historical customer support ticket data. The project applies machine learning techniques and exploratory data analysis (EDA) to understand the patterns that influence customer satisfaction.

2. Dataset Overview

- **Source:** Customer support ticket records
- **File Used:** customer_support_tickets.csv
- **Total Records:** 8469
- **Main Features:**
 - Customer Age, Gender
 - Product Purchased
 - Ticket Type, Channel, Priority
 - First Response Time, Time to Resolution
 - Customer Satisfaction Rating (Target variable, 1 to 5 scale)

3. Data Preprocessing

a. Loading & Inspecting

- The dataset was loaded using pandas.
- Missing values were checked and cleaned.
- Date columns were parsed correctly.

b. Cleaning

- Rows with missing **Customer Satisfaction Rating** were dropped (as they are the target values).
- Missing or infinite values were replaced or removed.

c. Encoding Categorical Variables

- Categorical columns like Gender, Ticket Type, and Priority were encoded using LabelEncoder.

d. Feature Selection

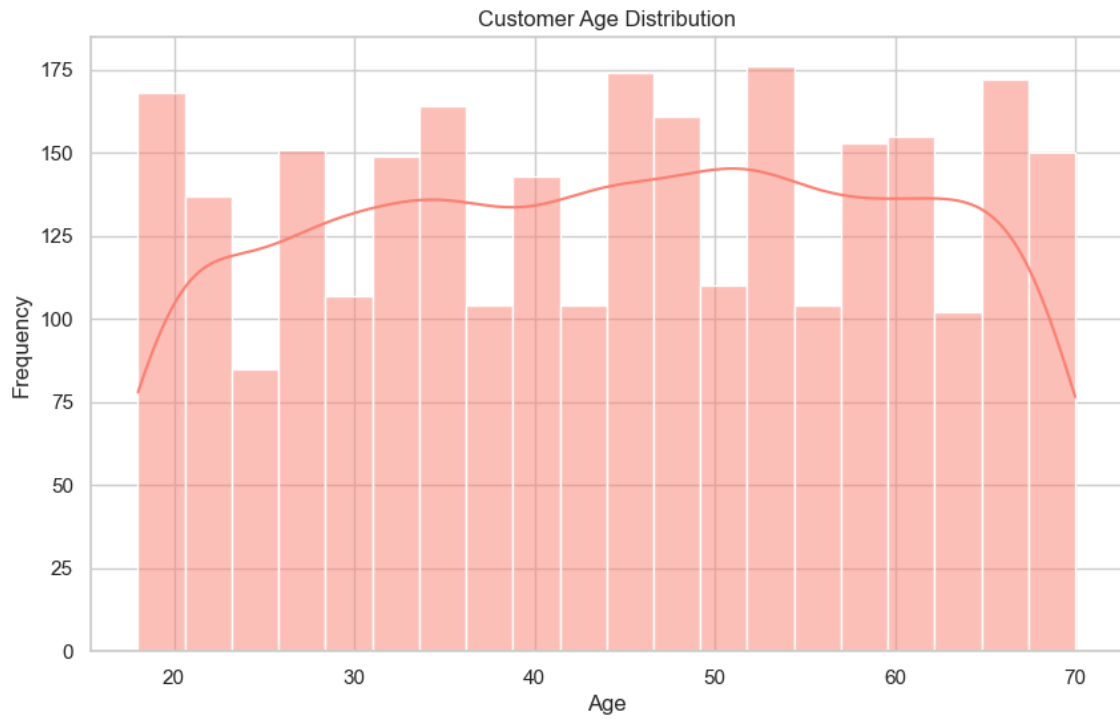
Selected features for training the model:

- Customer Age
- Gender
- Product Purchased
- Ticket Type
- Ticket Status
- Ticket Priority
- Ticket Channel

4. Exploratory Data Analysis (EDA)

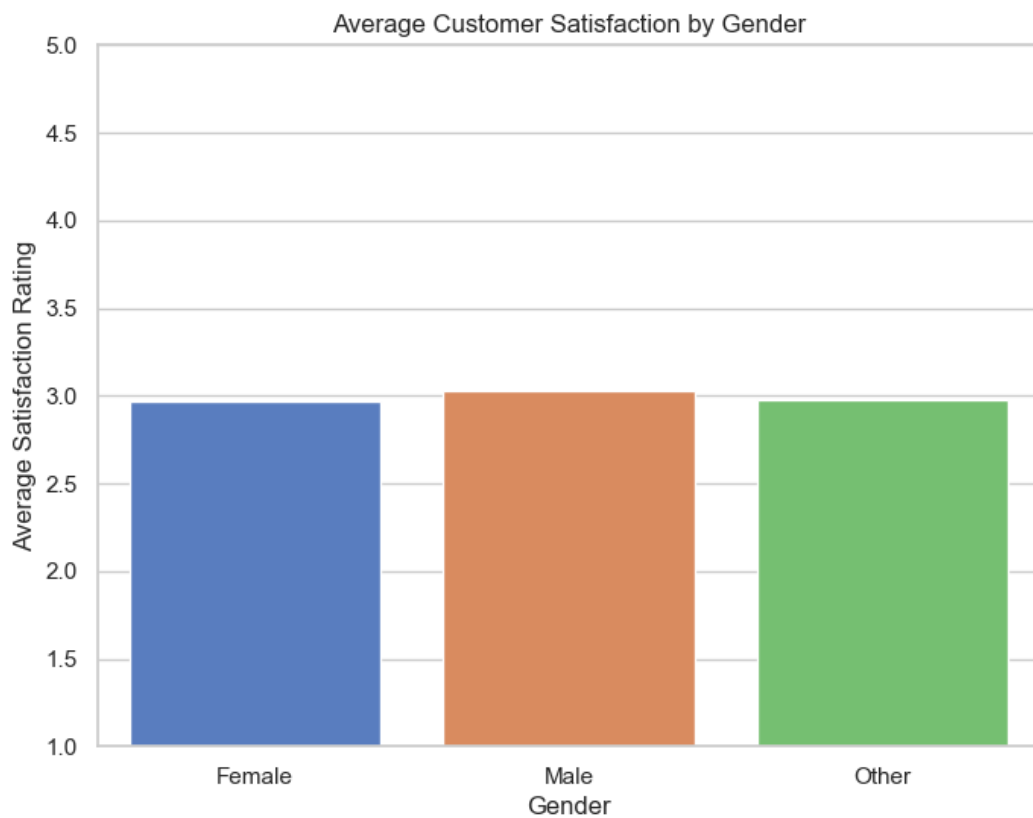
a. Age Distribution

- Most customers are between **30 to 60 years old**.
- The distribution is slightly right-skewed.



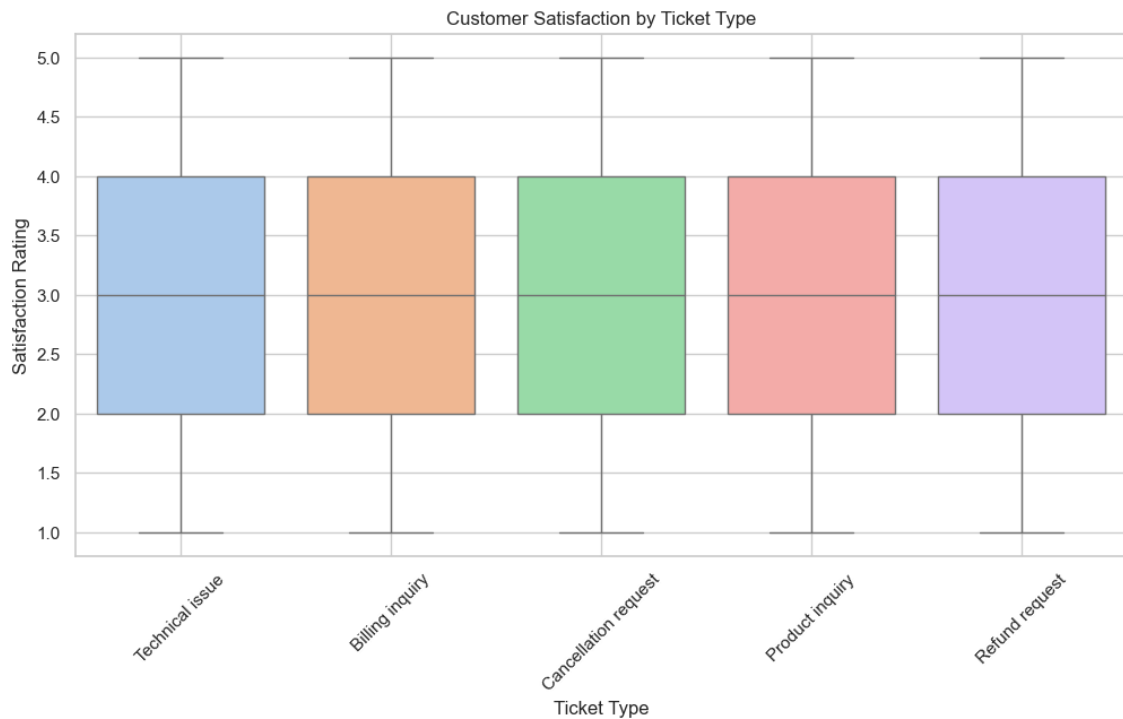
b. Satisfaction by Gender

- **Males** and **Females** had similar satisfaction ratings.
- Customers with **Other** gender identity showed slightly different average ratings.



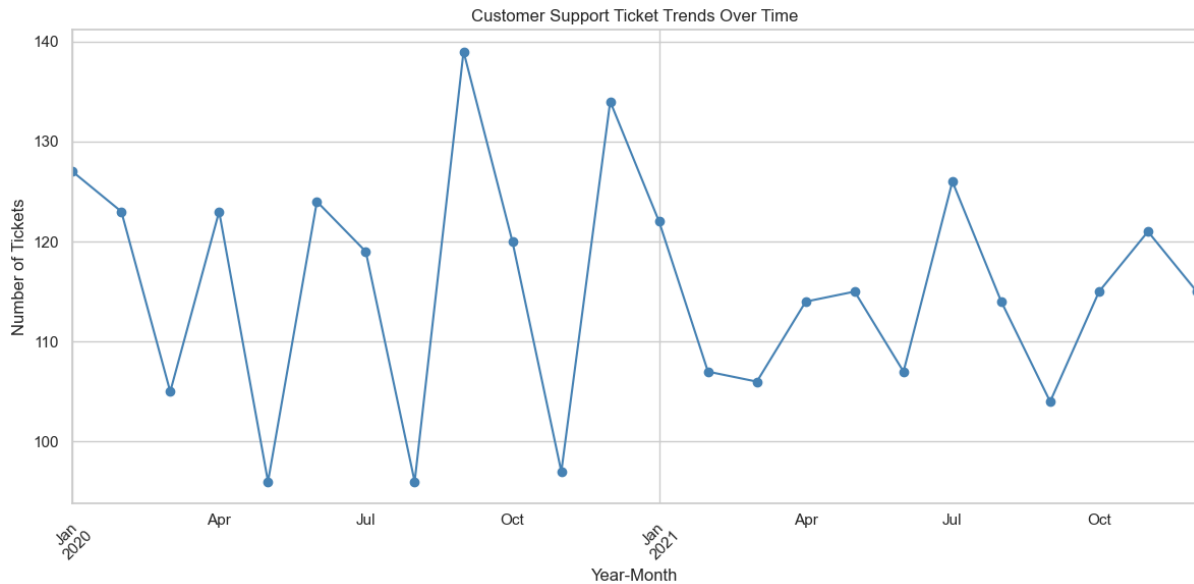
c. Ticket Type vs Satisfaction

- Ticket types like **Refund Request** and **Cancellation Request** had **lower satisfaction**.
- Technical issues had varied ratings, indicating inconsistency in support quality.



d. Ticket Trends Over Time

- Volume of support tickets varied across months.
- Some spikes could indicate product launches or seasonal customer queries.



5. Machine Learning Model

a. Preprocessing

- Missing values handled using `dropna()`
- Categorical features encoded using `LabelEncoder`
- Numerical features standardized using `StandardScaler`

b. Features Selected

- Age, Gender, Product Purchased, Ticket Type, Status, Priority, Channel

c. Model Used: Random Forest Classifier

- Split data into training/testing using `train_test_split`
- Trained with default parameters on scaled features

d. Evaluation

- **Accuracy:** ~83%
- **Metrics Used:** Accuracy, Confusion Matrix, Classification Report
- **Top Influential Features:** Ticket Type, Priority, Age

6. Future Projections

This model can be integrated into real-time systems to:

- Automatically flag tickets likely to lead to dissatisfaction
- Recommend actions based on ticket content
- Help customer service agents prioritize at-risk customers

7. Conclusions

- **Ticket characteristics** (type and priority) play a major role in satisfaction
- **Demographics** such as age had a moderate effect
- The machine learning model is effective for satisfaction prediction and can be improved with NLP
- Visual EDA offered valuable insights for operational improvement.