Customer Support Ticket Analysis and Resolution Time Prediction

1. Introduction

Customer support plays an important role in keeping customers satisfied. When tickets are resolved quickly, customers trust the service more. This project focuses on analysing customer support tickets and understanding how different factors affect ticket resolution time. The goal is to study ticket patterns and build simple machine learning models to predict ticket resolution behaviour.

---

2. Objective of the Project

The main objectives of this project are:

- To analyse customer support ticket data
- To understand how ticket priority, type, agent, and category affect resolution time
- To predict ticket resolution time using Linear Regression
- To classify tickets as fast or slow resolution using Logistic Regression
- To draw meaningful insights that can help improve customer support performance

---

3. Dataset Description

The dataset contains **1500 customer support tickets**.
Each row represents one support ticket.

Key columns in the dataset:

- ticket_id: Unique ID for each ticket
- category: Type of issue (Login Issue, Payment Failure, etc.)
- priority: Ticket urgency level (Low, Medium, High, Urgent)
- agent: Support agent handling the ticket

- type: Ticket type (Incident, Question, Request)
- queue: Department handling the ticket
- language: Language used by the customer
- sentiment: Customer sentiment
- resolution_hours: Time taken to resolve the ticket

---

4. Data Preprocessing

The dataset required moderate preprocessing before analysis:

- Checked for missing values
- Converted categorical columns into numerical values using Label Encoding
- Verified data types for analysis
- Created a new column resolved_fast:
    - Value = 1 if resolution time ≤ 48 hours
    - Value = 0 if resolution time > 48 hours

This helped in building classification models later.

---

5. Exploratory Data Analysis (EDA)

EDA was performed to understand ticket behaviour and patterns.

5.1 Ticket Priority Distribution

- Tickets are almost evenly distributed across priority levels
- Low and High priority tickets appear slightly more frequent

5.2 Average Resolution Time by Priority

- Urgent tickets have higher average resolution time
- Medium priority tickets tend to be resolved faster

- Priority clearly impacts resolution duration

5.3 Ticket Share by Priority

- Pie chart visualization shows the percentage contribution of each priority
- This helps understand workload distribution across priorities

EDA helped identify important features for modeling.

---

6. Linear Regression Model

6.1 Purpose

Linear Regression was used to predict **resolution time (in hours)** based on ticket features.

6.2 Model Inputs

Independent variables:

- Priority code
- Agent code
- Queue code
- Type code
- Language code
- Sentiment code

Target variable:

- Resolution hours

6.3 Model Evaluation

- Mean Absolute Error (MAE) was calculated
- R^2 score was evaluated

6.4 Observations

- The R^2 score was low, showing poor prediction accuracy

- Predictions were mostly clustered around average values
- Resolution time depends on many human and operational factors

6.5 Conclusion for Linear Regression

Linear Regression was **not suitable** for accurate prediction in this case. The relationship between features and resolution time is complex and non-linear.

---

7. Logistic Regression Model

7.1 Purpose

Logistic Regression was used to classify tickets as:

- **Fast resolution (≤ 48 hours)**
- **Slow resolution (> 48 hours)**

7.2 Model Inputs

Independent variables:

- Priority
- Agent
- Queue
- Ticket type
- Language
- Sentiment

Target variable:

- Resolved fast (0 or 1)

7.3 Model Evaluation

- Accuracy score
- Confusion matrix

- Classification report (precision, recall, F1-score)

7.4 Observations

- Logistic Regression performed better than Linear Regression
- Accuracy was moderate and close to baseline
- Model captured some meaningful patterns
- Classification approach suits business decision-making

---

8. Model Comparison

Model	Purpose	Performance
Linear Regression	Predict exact resolution time	Poor
Logistic Regression	Classify fast vs slow tickets	Better

Logistic Regression is the better model for this project because it provides clear, interpretable results.

---

9. Business Insights

- Ticket priority strongly affects resolution time
- Urgent tickets require better resource planning
- Classification models are more useful than exact time prediction
- This analysis can help:
  - Improve agent allocation
  - Reduce customer waiting time
  - Enhance service quality

---

10. Conclusion

This project analyzed customer support ticket data to understand resolution behavior. Exploratory analysis provided clear insights into ticket priorities and resolution patterns. Linear Regression was tested but showed limited performance due to complex data relationships. Logistic Regression proved more effective by classifying tickets into fast and slow resolution categories. Overall, the project demonstrates how data analysis and machine learning can support better decision-making in customer support systems.

---

11. Future Improvements

- Use more advanced models like Random Forest or XGBoost
- Include agent workload and time-of-day features
- Handle class imbalance using sampling techniques
- Deploy the model as a real-time dashboard