Customer Support Data Analysis & CSAT Prediction

A Data Analytics & Machine Learning Project

Introduction & Business Problem

Objective: To analyse customer support data to understand customer satisfaction (CSAT) and build a machine learning model to predict CSAT scores.

Business Problem: Customer satisfaction is critical for business success. Analyzing customer interactions helps identify key areas for improvement, reduce customer churn, and ultimately, improve the overall customer experience.

Goal: To provide data-driven insights and a predictive model to help the company make better business decisions.

Data & Methodology

Data Source: `Customer_support_data.csv`

Data Overview:

- **Key Columns:** `CSAT Score`, `Product_category`, `channel_name`, and `Resolution_Time_Minutes`.

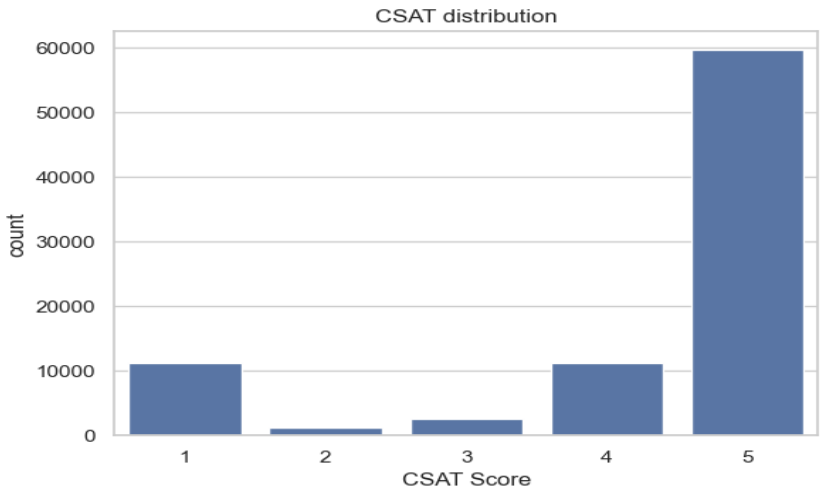
Methodology:

1. **Data Cleaning:** Cleaning the data, handling missing values, and engineering new features.
2. **Exploratory Data Analysis (EDA):** Using visualizations to find patterns and insights.
3. **Machine Learning:** Training a classification model to predict `CSAT Score`.

Key Insights from EDA

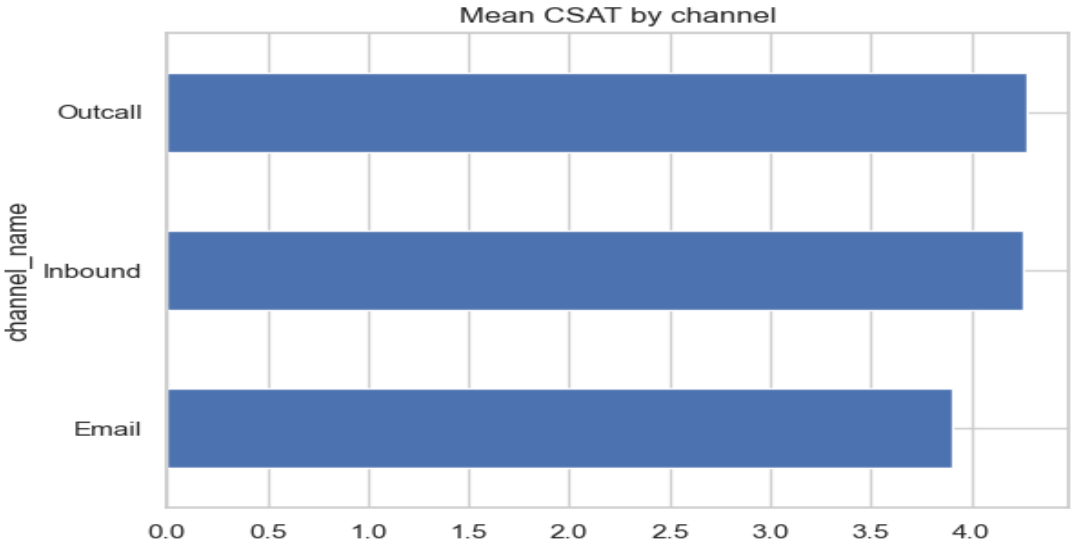
1. Distribution of CSAT Scores

- **Insight:** A majority of customers gave a perfect score of 5, indicating high overall satisfaction.
- **Visualization:**



2. Customer Satisfaction by Channel

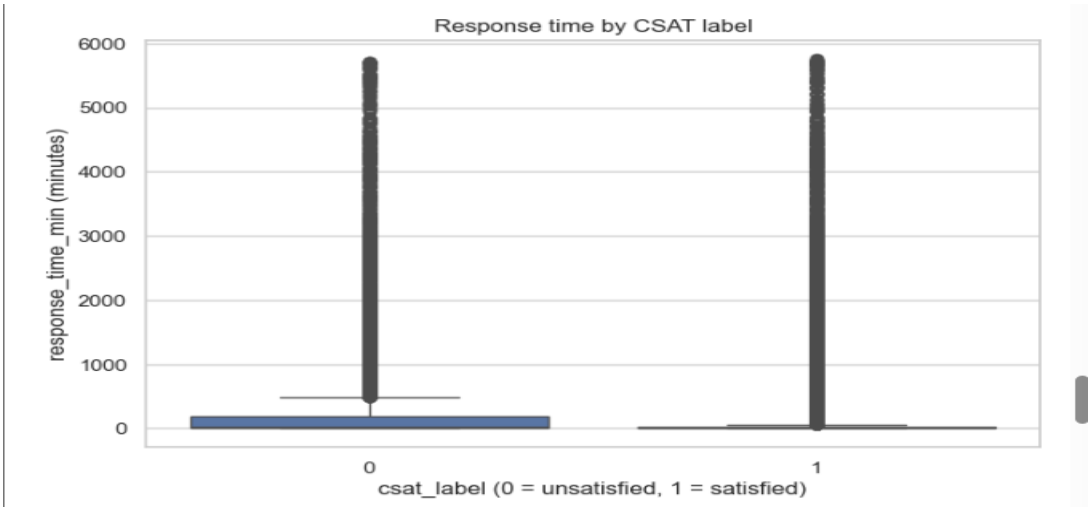
- **Insight:** Customers contacting via the `Inbound` channel tend to have slightly lower CSAT scores compared to `Outcall`.
- **Visualization:**



Key Insights from EDA (Relationship Analysis)

1. Resolution Time vs. CSAT Score

- **Insight:** There is a slight negative correlation between resolution time and CSAT score, suggesting that faster issue resolution can lead to higher customer satisfaction.
- **Visualization:**



2. Agent Performance Metrics

- **Insight:** Agents with more experience (`>90` days) tend to achieve higher average CSAT scores.

3. Item Price vs. CSAT Score

- **Insight:** There is no significant relationship between the item's price and the customer's satisfaction level.

Machine Learning Model

Objective: To predict the `CSAT Score` based on other features in the dataset.

Model Used: Random Forest Classifier.

Why this model?

- An ensemble model that is robust.
- Handles various data types well.
- Less prone to overfitting.

Model Process:

- **Feature Selection:** Used features like `channel_name`, `Product_category`, and `Resolution_Time_Minutes`.
- **Data Split:** Data was split into a training set (80%) and a testing set (20%).
- **Evaluation:** Used metrics like `Accuracy`, `Precision`, and `Recall`.

Model Performance & Results

Accuracy Score: The model achieved an accuracy of [0.6942730764753812].

Classification Report:

Classification Report:

	precision	recall	f1-score	support
1	0.51	0.08	0.14	2241
2	0.50	0.00	0.01	227
3	0.00	0.00	0.00	503
4	0.12	0.00	0.01	2277
5	0.70	0.98	0.82	11934
accuracy			0.69	17182
macro avg	0.37	0.21	0.19	17182
weighted avg	0.58	0.69	0.59	17182

Interpretation:

- Model predicts CSAT=5 (satisfied) very well (Recall = 98%, F1 = 0.82).
- This model can be used to identify potential low-CSAT cases proactively.
- Overall accuracy = 69%

Conclusion & Recommendations

Key Findings:

- Customers generally give high CSAT scores.
- Channels like inbound may require more attention.
- Faster issue resolution is linked to higher satisfaction.

Business Recommendations:

- **Actionable Insight:** Focus on optimizing the inbound support process to improve satisfaction.
- **Model Deployment:** Deploy the predictive model to automatically flag cases that are likely to result in a low CSAT score, allowing supervisors to intervene proactively.

Contact: *[sawantvruttika26gmail.com]*