Student Name: Dishma R.R
Register Number: 212923243010
Institution: St. Joseph College Of Engineering
Department: Artificial Intelligence And Data Science
Date of Submission: 07/05/2025
GitHub Repository Link :

---

1. Problem Statement

The objective of this project is to analyze customer survey data to identify the key drivers of customer satisfaction. By leveraging statistical and visual analytics, the project aims to discover which factors most significantly influence a customer's overall satisfaction rating. This analysis is crucial for businesses aiming to improve customer retention, service quality, and brand loyalty. The project involves **diagnostic and exploratory analytics**, as we aim to not only understand current satisfaction levels but also diagnose contributing factors.

---

2. Project Objectives

- Identify which survey attributes (e.g., service speed, staff behavior, product quality) correlate most strongly with customer satisfaction.
- Understand patterns across demographic or geographic segments.
- Recommend actions to improve satisfaction scores based on findings.
- Deliverables include statistical insights, visual dashboards, and factor analysis.
- Objectives remained consistent but evolved to include demographic-specific analysis after EDA.

---

3. Flowchart of the Project Workflow

java

CopyEdit

Survey Data Collection

↓

Data Preprocessing

↓

Exploratory Data Analysis (EDA)

↓

Correlation & Feature Importance Analysis

↓

Visualization & Insight Generation

↓

Recommendations & Reporting

---

4. Data Description

- **Dataset Name & Source:** Customer Satisfaction Survey Data – Internal Company Data
- **Data Type:** Structured
- **Rows & Columns:** Approximately 5,000 responses × 15 attributes
- **Dataset Nature:** Static
- **Key Fields:** Satisfaction Score, Service Time, Product Quality, Staff Interaction, Demographics (Age, Gender, Region), Net Promoter Score, Complaint Resolution

---

5. Data Preprocessing

- **Missing Values:** Handled via mean/mode imputation and logical removal for critical blanks.
- **Duplicates:** Removed 80 duplicate entries.
- **Formatting:** Parsed dates, standardized column names.
- **Encoding:** Applied Label Encoding for ordinal responses and One-Hot Encoding for nominal fields.
- **Outliers:** Identified using IQR and z-score; extreme anomalies were excluded.

---

6. Exploratory Data Analysis (EDA)

- **Univariate Analysis:** Histograms for satisfaction scores, bar charts for categorical responses.
- **Bivariate Analysis:** Heatmaps and correlation matrix to identify feature relationships.
- **Multivariate Analysis:** Pairplots and regression plots for deeper relationship exploration.
- **Key Insights:**
    - Product Quality and Staff Behavior had the strongest correlation with satisfaction.
    - Resolution time negatively impacted satisfaction in older demographics.
    - NPS correlated positively with quick issue resolution and respectful staff behavior.

---

7. Tools and Technologies Used

- **Programming Language:** Python
- **Notebook/IDE:** Jupyter Notebook

- **Libraries:** pandas, numpy, seaborn, matplotlib, scikit-learn, plotly
- **Automation Tools:** pandas-profiling (for EDA)

---

8. Team Members and Contributions

Name	Contribution
Santina T	Data preprocessing & EDA
Dishma R R	Visualization & correlation analysis
Deyon Michaela K F	Reporting & Recommendations

Would you like this in a downloadable Word document format with your name and details added in?

4o