# PYTHON PROGRAMMING PROJECT REPORT

(Project Semester January-April 2025)

TITLE: AIRLINE PASSENGER SATISFACTION

Submitted by

Name: Daksh Gupta

Registration No: 12313179

Section: K23FD Programme: BTech CSE Course Code INT375 Under the Guidance of

Name of Faculty:

Baljinder Kaur

UID: 27952

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

****

# CERTIFICATE

This is to certify that Daksh Gupta bearing Registration no. 12313179 has completed INT375 project titled, **“Airline Passenger Satisfaction”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Baljinder Kaur

**Assistant Professor**

**School of Computer Science and Engineering**

Lovely Professional University Phagwara, Punjab.

Date: 04-04-2025

# DECLARATION

I, Daksh Gupta, student of BTech CSE under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 04-04-2025

Registration No. 12313179 Daksh Gupta

# ACKNOWLEDGEMENT

I want to take a moment to express my deep appreciation for the support I have received from everyone, either directly or indirectly, for enabling me to finish this project successfully. To start, I am grateful to Baljinder Kaur for his guidance, feedback, and steady support during this project.

His guidance allowed not only for academic support but also a wealth of moral support when I needed help staying on track and maintaining my motivation. I would also like to express my gratitude to **Lovely Professional University** for their example and support in offering a learning experience that fosters innovation, critical thinking, and practical application.

The resources and infrastructure they provided were significant factors that enabled me to finish the project. I need to thank my family and close friends for being my backbone throughout the project. Their understanding, optimism, and faith in me provided support, especially as I experienced self-doubt and/or pressure.

Finally, I thank the individuals who provided support through growth, learning and inspiration, and hope they realize that this project does not only indicate the summation of technical knowledge and learning, but is a personal accomplishment in and of itself, that indicates growth, perseverance and passion.

**Table of Content**

|  |  |  |
| --- | --- | --- |
| **SNO.** | **TITLE** | **Page No.** |
| 1 | Source of dataset | 6 |
| 2 | Abstract | 7 |
| 3 | Introduction | 8 |
| 4 | Methodology | 9 |
| 5 | Data Collection and loading | 10 |
| 6 | Data Preprocessing | 12 |
| 7 | EDA | 14 |
| 8 | Feature Extraction | 20 |
| 9 | Model evaluation | 23 |
| 10 | Binary Classification for Operational Readiness | 25 |
| 11 | Results and discussion | 27 |
| 12 | Scope for future research | 29 |
| 13 | Conclusion | 30 |
| 14 | References | 31 |

SOURCE OF DATASET [https://mavenanalytics.io/data-](https://mavenanalytics.io/data-playground?order=date_added%2Cdesc&search=airline%20passenger%20satisfaction) [playground?order=date\_added%2Cdesc&search=airline%20passenger%20satisfaction](https://mavenanalytics.io/data-playground?order=date_added%2Cdesc&search=airline%20passenger%20satisfaction)

# ABSTRACT

Airline passenger satisfaction is a pivotal factor in the success and sustainability of modern air travel operations. In an era where customers expect seamless digital experiences, timely flights, and personalized service, understanding what influences passenger satisfaction has become a strategic necessity. This project presents a data-driven analysis of airline passenger satisfaction using a large real-world dataset comprising 129,880 records and 24 features ranging from demographic details to inflight service ratings and operational metrics.

The dataset includes variables such as age, customer type, travel class, flight distance, service quality scores (e.g., seat comfort, cleanliness, online boarding), and flight delays. Through a structured analytical pipeline—encompassing data cleaning, transformation, and exploratory data analysis— this study identifies the key service factors that correlate with passenger sentiment. Missing values were imputed using statistical measures like median and mode, and categorical features were encoded for quantitative analysis.

Visual tools such as count plots, box plots, heatmaps, and pie charts were employed to uncover insights. The analysis revealed that loyal customers and business class passengers report significantly higher satisfaction levels. Services such as check-in, online boarding, inflight entertainment, and baggage handling emerged as strong influencers of positive feedback, while delay durations were associated with lower satisfaction.

Though no predictive model was applied in this phase, the structure of the dataset is well-suited for future binary classification modeling. Insights from this analysis can guide airlines in enhancing passenger experience by focusing on operational efficiency, customer engagement, and digital service quality. Future extensions of this work could incorporate sentiment analysis from reviews or machine learning-based satisfaction prediction systems.

# SECTION I

**Introduction:**

Passenger satisfaction has emerged as one of the most critical performance indicators in the airline industry, directly influencing customer loyalty, brand perception, and long-term profitability. In an era of heightened competition and digital transformation, airlines must go beyond basic service delivery to provide consistent, seamless, and personalized experiences to travelers. Dissatisfaction

in even a single area—such as boarding delays, poor inflight service, or technical glitches during check-in—can ripple across the customer journey, impacting reviews, repeat business, and market competitiveness.

Historically, passenger satisfaction was measured using post-flight surveys or isolated service metrics, which often failed to capture the complex interplay between operational efficiency and customer perception. With the increasing availability of structured feedback data and service logs, it is now possible to use advanced analytics to uncover the key drivers behind satisfaction or dissatisfaction. Variables such as travel class, flight delays, customer loyalty status, inflight comfort, and service accessibility provide rich insights when analyzed together.

The emergence of big data tools and computational techniques—especially exploratory data analysis and classification models—has unlocked new opportunities for airlines to better understand passenger behavior. These techniques allow the discovery of non-linear relationships and hidden patterns that might not be apparent through traditional statistical methods. While this study primarily focuses on exploratory analysis, the dataset structure is well-suited for future predictive modeling using machine learning techniques, such as decision trees or ensemble classifiers, to forecast satisfaction outcomes based on service quality parameters.

By analyzing over 129,000 passenger records, this project aims to identify the most influential factors affecting satisfaction and highlight potential areas for operational improvement. The study provides actionable insights that can support service enhancement strategies, resource allocation, and customer engagement efforts. Furthermore, it lays the groundwork for future research involving real-time data integration, sentiment analysis, and intelligent satisfaction forecasting systems to help airlines offer truly customer-centric experiences.

# METHODOLOGY:

This research follows a structured, multi-phase analytical pipeline aimed at understanding and interpreting the factors influencing airline passenger satisfaction. Each step is carefully designed to ensure data integrity, meaningful insight extraction, and actionable conclusions. The overall methodology emphasizes clarity, robustness, and interpretability through data preparation, exploration, and visualization.

The first phase of the project involved **data acquisition and preprocessing**. The dataset, comprising over 129,000 records and 24 features, was examined for missing values, inconsistent data types, and formatting issues. Missing values in numerical columns—such as arrival delays—were handled using median imputation to mitigate the impact of outliers, while categorical missing entries were filled using the mode. Columns containing survey-based ratings were converted to numerical formats to allow for statistical analysis.

Next, **Exploratory Data Analysis (EDA)** was carried out to understand the distribution of key variables and uncover patterns and correlations between features and the target variable, ‘Satisfaction’. Visual tools such as pie charts, count plots, box plots, and heatmaps were employed to analyze how demographic factors, travel class, delays, and service ratings affect overall satisfaction. For instance, plots revealed that business class travelers and loyal customers tended to report higher satisfaction, whereas dissatisfaction was more prominent among economy class travelers facing longer delays or poor digital service experiences.

Following EDA, the dataset was refined through **feature engineering and transformation**. Variables such as Customer Type, Type of Travel, and Class were encoded to numerical formats, allowing for potential model training in future iterations. Quantitative ratings (e.g., Check-in Service, Online Boarding, Seat Comfort) were normalized to ensure consistent scale and prevent skewed influence on analysis. This step ensures that all variables are interpretable and comparable across the dataset.

Although the primary scope of this project was analysis and interpretation rather than modeling, the dataset structure lends itself well to **future classification modeling**. A binary classification task could be defined where the target is whether a passenger is "Satisfied" or "Neutral or Dissatisfied". In such an approach, the dataset would be split into training and testing subsets, and machine learning models like Random Forest, Logistic Regression, or Decision Trees could be applied. Standard evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrices would then be used to assess model performance.

This methodology provides a comprehensive view of how structured passenger data can be used to uncover satisfaction drivers, highlight service bottlenecks, and guide operational improvements. The process demonstrates how even basic EDA—when applied systematically—can produce deep insights without requiring predictive modeling at the initial stage.

# SECTION II:

**STEP 1: Data Collection and Loading**

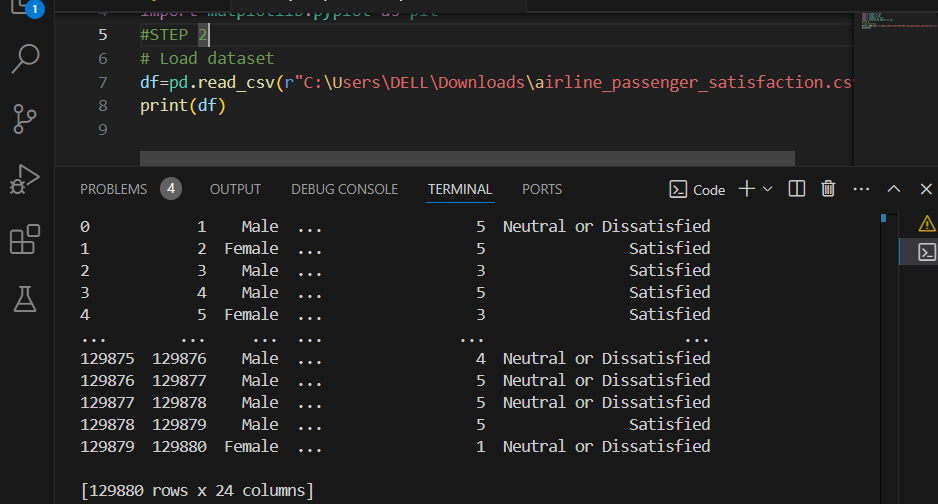
At the heart of any data-driven project lies the quality and relevance of the dataset used. For this analysis, we utilized a structured CSV file titled "airline\_passenger\_satisfaction.csv", which consists of over 129,000 records and 24 features pertaining to passenger demographics, flight details, and service feedback collected through post-travel surveys. The dataset appears to be curated from an airline’s customer experience monitoring system and provides both subjective feedback (e.g., comfort ratings) and objective operational metrics (e.g., delays, flight distance).

Important features in the dataset include:

* Demographics: Age, Gender, Customer Type (Loyal/First-time)
* Travel Details: Type of Travel (Business/Personal), Class (Economy/Business), Flight Distance
* Operational Metrics: Departure Delay, Arrival Delay
* Service Ratings: Seat Comfort, Check-in Service, Online Boarding, In-flight Wifi, Entertainment, Food, Cleanliness, etc.
* Target Variable: Satisfaction (Satisfied or Neutral/Dissatisfied)

After collection, the dataset was loaded into the Python environment using Pandas’ read\_csv() method. An initial review using .head(), .info(), and .describe() confirmed the integrity of the data and helped identify inconsistencies such as missing values. Notably, the Arrival Delay column had some null values (~393 records), which were handled using median imputation (df.fillna(df.median(numeric\_only=True))) to minimize the influence of outliers.

The overall structure was found to be clean and analysis-ready. Categorical variables were retained in their original form at this stage for clarity during visual exploration. With the dataset successfully transformed into a structured DataFrame, it was prepared for preprocessing and exploration.



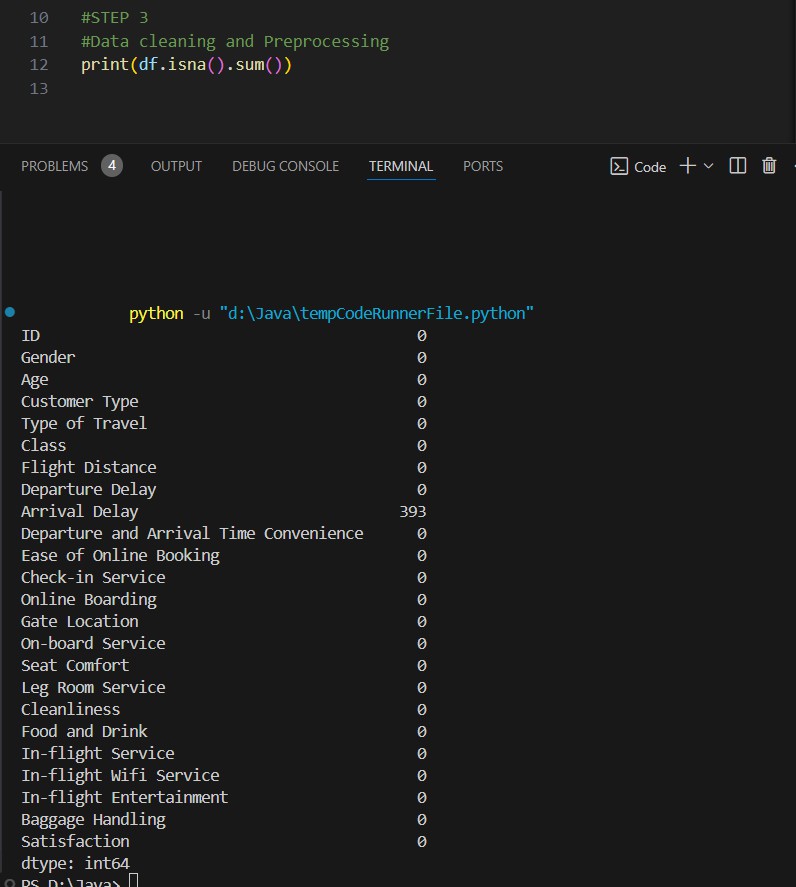


Figure 1: Loading dataset and cleaning

STEP 2: Data Preprocessing

Data preprocessing is a foundational step for any analytical pipeline, aimed at transforming raw and sometimes messy data into a refined format that supports accurate insights and future modeling. In this project, the goal was not to train a machine learning model, but to ensure a robust dataset for exploratory data analysis (EDA) and pattern discovery.

1. Handling Missing Values:

Although the dataset was largely complete, a small number of null entries were present in the Arrival Delay column. These were replaced with the median value, which provides a better central estimate in the presence of outliers than the mean. This choice helped preserve the dataset’s distribution while eliminating empty cells that might hinder analysis.

1. Data Type Conversion and Normalization:

A few numerical columns were explicitly cast using pd.to\_numeric() to ensure proper numeric handling during plotting and statistical analysis. These included service ratings (e.g., Seat Comfort, Online Boarding, etc.) and delay-related fields. The dataset was not normalized, as most analysis was visual in nature, and interpretation of original values (e.g., flight distance in kilometers, delay in minutes) was preferred for direct understanding.

1. Encoding Categorical Variables:

Though encoding was not mandatory for this phase, categorical fields such as Satisfaction, Customer Type, Type of Travel, and Class were label-encoded for certain correlation and distributional plots. These encodings helped in generating accurate plots like heatmaps and grouped bar charts.

1. Consistency Check:

The dataset was re-validated post-processing for:

* + Column name uniformity
  + Correct value ranges
  + Absence of stray or irrelevant columns

Having completed this step, the dataset was now clean, consistent, and formatted for in-depth exploratory analysis.

**STEP 3: Exploratory Data Analysis (EDA)**

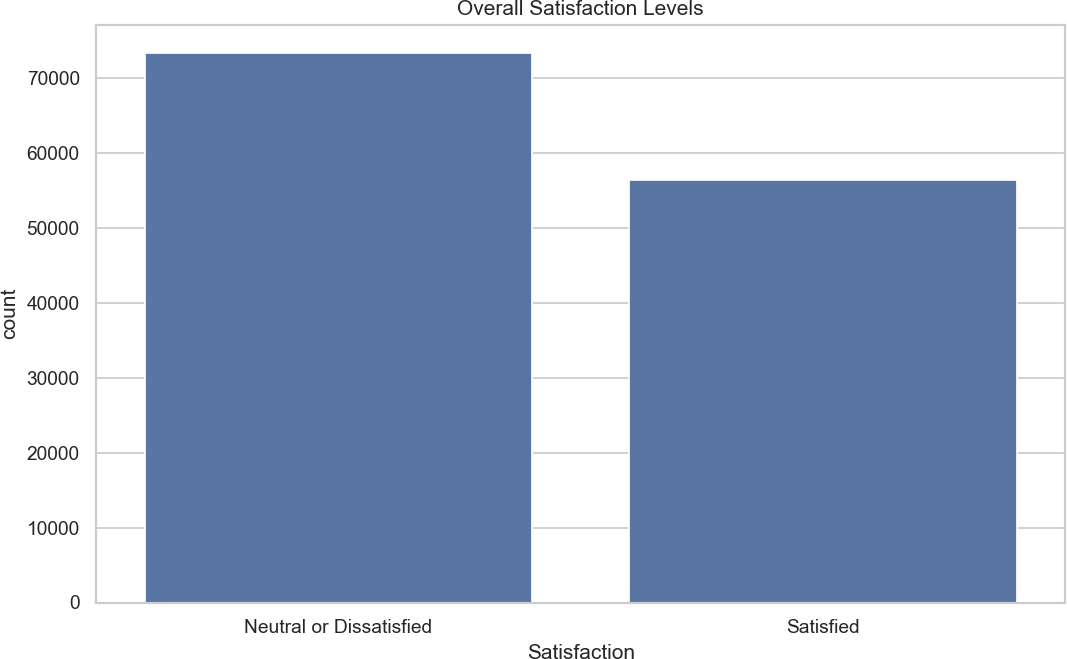
To thoroughly understand the drivers of airline passenger satisfaction, it is essential to first explore the structure, distribution, and interrelationships within the dataset. **Exploratory Data Analysis (EDA)** serves as a foundational step in any data-driven research, offering critical insights into data behavior, helping detect inconsistencies or anomalies, and revealing meaningful patterns that can shape service decisions or future model development. In this study, the dataset comprises both categorical and numerical variables—ranging from service ratings and passenger demographics to delay times and flight characteristics.

The primary objective of this EDA is to uncover which attributes have the strongest associations with the satisfaction levels of airline passengers, and to visualize how these factors differ between "Satisfied" and "Neutral or Dissatisfied" customers. Below are key components of the analysis:

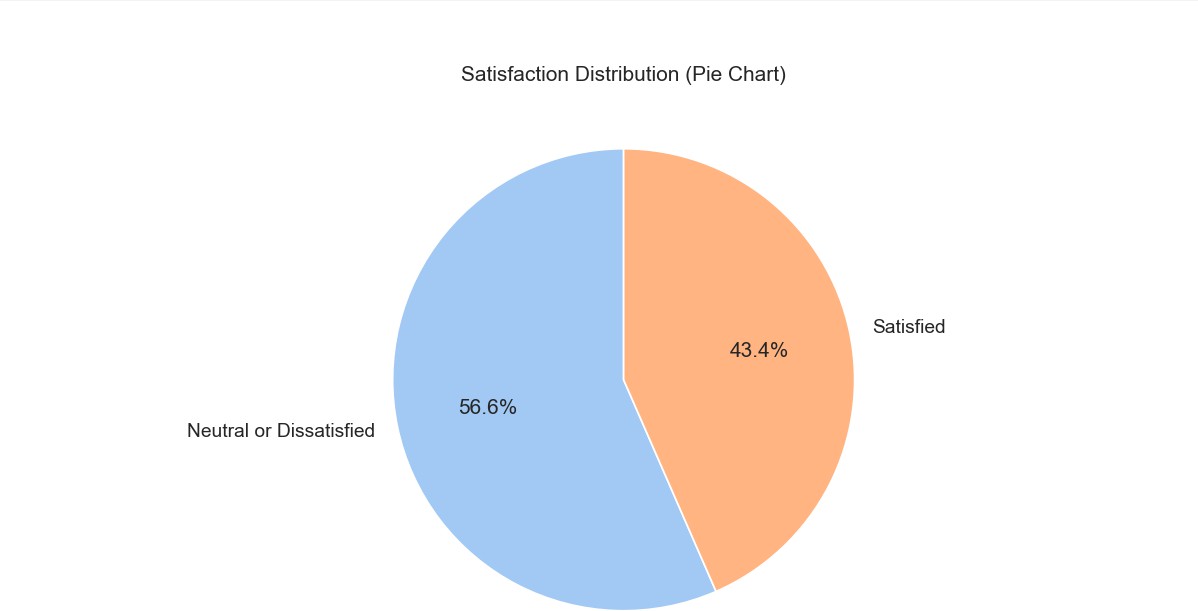
1. **Overall Satisfaction Levels (Count Plot, Pie Chart & Donut Chart)**

The exploratory analysis began by examining the distribution of the target variable—Passenger Satisfaction. A count plot (Figure 1) revealed a slight class imbalance, with a higher number of passengers categorized as *Neutral or Dissatisfied* compared to those marked as *Satisfied*. This observation was further supported using a pie chart (Figure 2) and a donut chart (Figure 3), both of which showed that approximately 56.6% of passengers were *Neutral or Dissatisfied*, while 43.4% were *Satisfied*. The donut chart offered a sleek, modern visualization while maintaining interpretability. These insights emphasize the importance of addressing class imbalance in predictive modelling, possibly through techniques such as resampling or applying class weights during training.

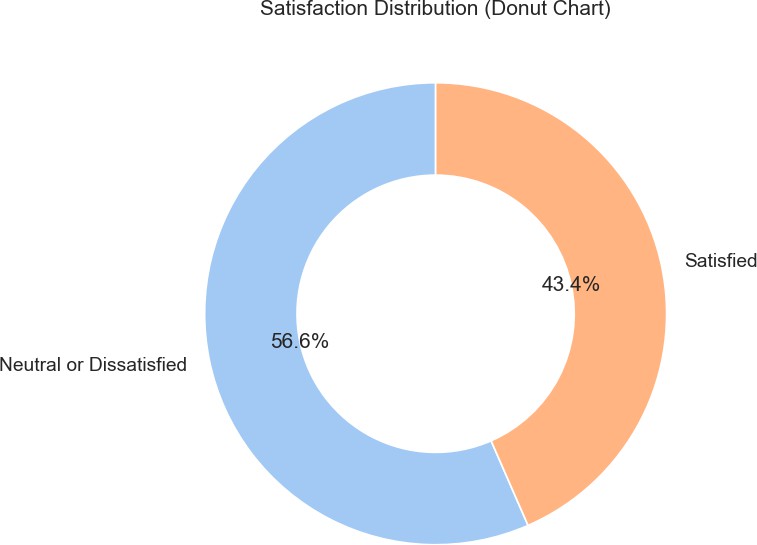
* + **Figure 1:** Count plot of overall satisfaction levels.
  + **Figure 2:** Pie chart showing percentage distribution of passenger satisfaction.
  + **Figure 3:** Donut chart showing percentage distribution of passenger satisfaction in a modern layout.



**Fig 1: Count Plot**

****

**Fig 2: Pie Chart**



**Fig 3: Donut Chart**

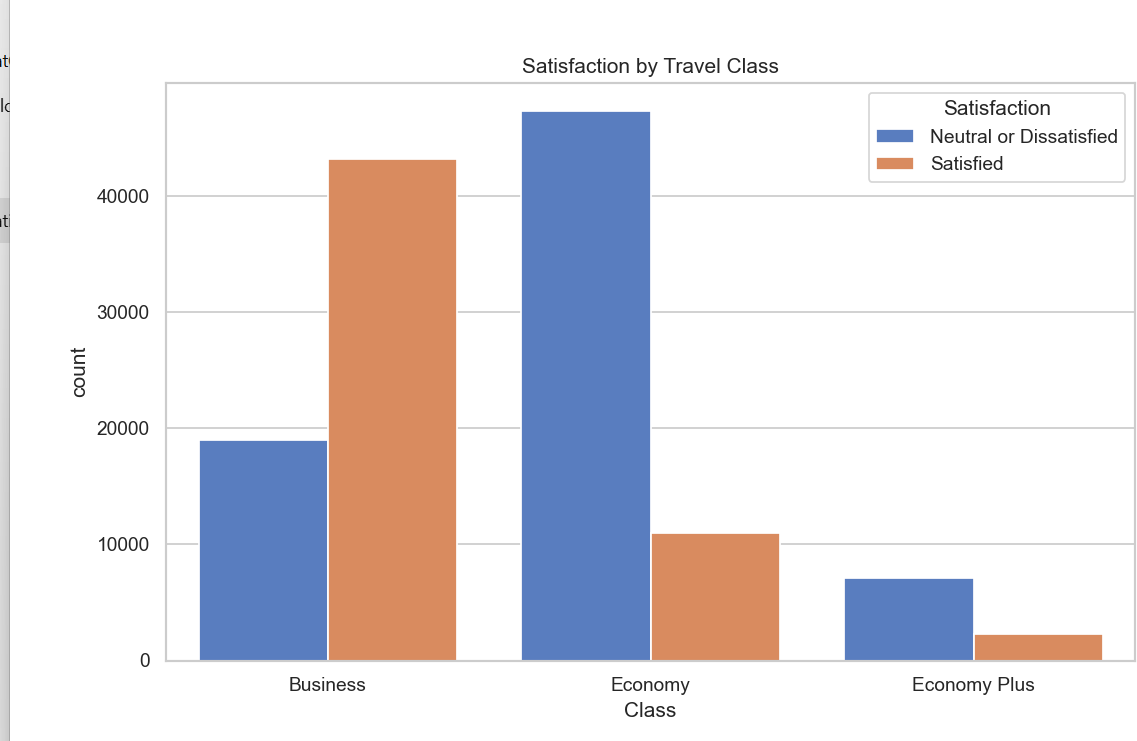
1. **Satisfaction by Travel Class (Count Plot)**

One of the key factors influencing passenger satisfaction is the travel class. A count plot (Figure 4) was used to visualize the satisfaction distribution across different travel classes—Business, Economy, and Economy Plus.

From the chart, it is evident that Business class passengers are predominantly *Satisfied*, highlighting the likely impact of superior amenities, services, and overall comfort in this segment. On the other hand, Economy class shows the highest number of *Neutral or Dissatisfied* passengers, indicating a potential gap in service expectations versus delivery. Economy Plus presents a more balanced distribution, although it still leans slightly toward dissatisfaction.

This analysis reinforces the hypothesis that travel class significantly affects overall satisfaction and should be considered a vital predictor in the modeling process. It also suggests opportunities for airlines to improve satisfaction in the lower classes, possibly through better onboarding, in-flight services, or more comfortable seating.

* + **Figure 4:** Count plot showing satisfaction levels grouped by travel class.



**Fig 4: Count Plot**

1. **Impact of Customer Type (Count Plot):**

Understanding the type of customer is essential in gauging satisfaction trends. A count plot (Figure 5) was used to analyze how satisfaction varies between Returning Customers and First-time Customers.

The visualization reveals a stark contrast:

* + Returning Customers are significantly more likely to be *Satisfied*, showing a relatively balanced distribution.
  + First-time Customers show a notable skew towards *Neutral or Dissatisfied* responses. This pattern aligns with expectations, as loyal (returning) customers have likely had satisfactory prior experiences and are familiar with the airline’s services. In contrast, first-time customers may be more critical or face unmet expectations. This strong association suggests that customer loyalty is a reliable predictor of satisfaction. Airlines may benefit from targeting first-time flyers with better onboarding experiences and personalized services to convert them into loyal, satisfied passengers.
  + **Figure 5:** Count plot showing satisfaction levels grouped by customer type.

.



**Fig 5: Count Plot**

1. **Check-in and Boarding Experience:**

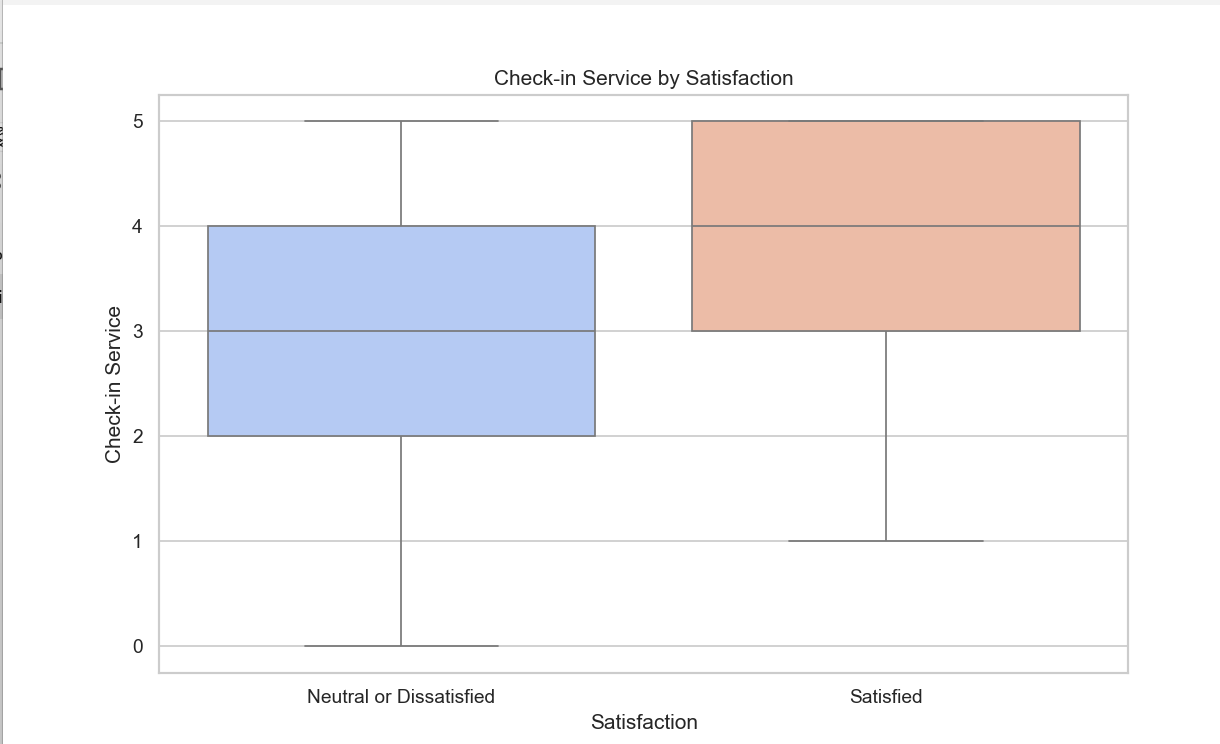
Passenger experience at the airport plays a vital role in shaping overall satisfaction. Two key factors—**Check-in Service** and **Online Boarding**—were analyzed using boxplots (Figure 6 and Figure 7) to understand their influence on satisfaction.

From **Figure 6**, it is clear that passengers who reported higher satisfaction also gave higher ratings for **Check-in Service**, with their responses clustering around the top end of the scale (4 to 5). In contrast, those who were Neutral or Dissatisfied had a wider spread, including lower ratings (0 to 3), indicating inconsistent or subpar check-in experiences.

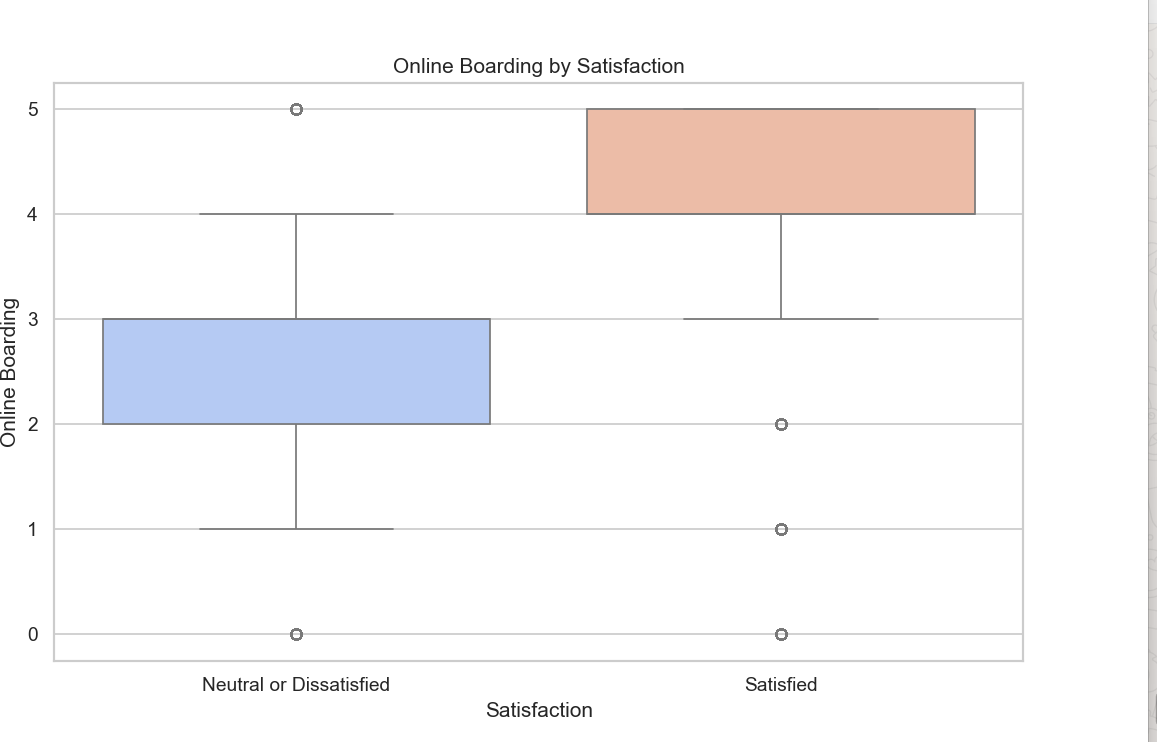
Similarly, **Figure 7** shows that **Online Boarding** scores were significantly higher among Satisfied passengers. Most of them rated the experience between 4 and 5, while Neutral or Dissatisfied passengers leaned toward lower ratings (1 to 3). This suggests that efficient and smooth online boarding is strongly correlated with a positive travel experience.

These insights highlight that both **Check-in Service** and **Online Boarding** are critical service touchpoints that airlines must optimize to boost customer satisfaction.

* + **Figure 6:** Box plot of Check-in Service ratings by satisfaction category.
  + **Figure 7:** Box plot of Online Boarding ratings by satisfaction category.



**Fig 5: Box Plot**

****

**Fig 6: Box Plot**

1. **Satisfaction Based on Flight Distance or Duration**

Several plots were used to examine how flight distance, delays, travel class, and demographic features impact passenger satisfaction.

1. **Flight Distance and Satisfaction:**

Histplot & KDE Plot (**Figure 7** & **Figure 8**) reveal that satisfied passengers are more likely to be on longer flights. The KDE curves show a peak for satisfied passengers at higher distances, whereas dissatisfied ones tend to be on shorter routes.

This implies that longer flights may offer more opportunities for service quality to positively impact satisfaction.

1. **Travel Class Impact**

The Stacked Bar Chart (**Figure 9**) shows that Business Class passengers report the highest satisfaction levels, while Economy Class passengers report more dissatisfaction.

This underscores how class of service heavily influences overall satisfaction.

1. **Flight Distance vs. Age**

In the Scatter Plot (**Figure 10**), we observe a spread of all age groups across flight distances, but older passengers on longer flights appear more frequently in the satisfied group, possibly due to better service expectations being met.

1. **Service Ratings by Satisfaction**

The Bar Chart (**Figure 11**) confirms that Check-in Service, Online Boarding, and Seat Comfort scores are consistently higher among satisfied passengers, reinforcing earlier insights.

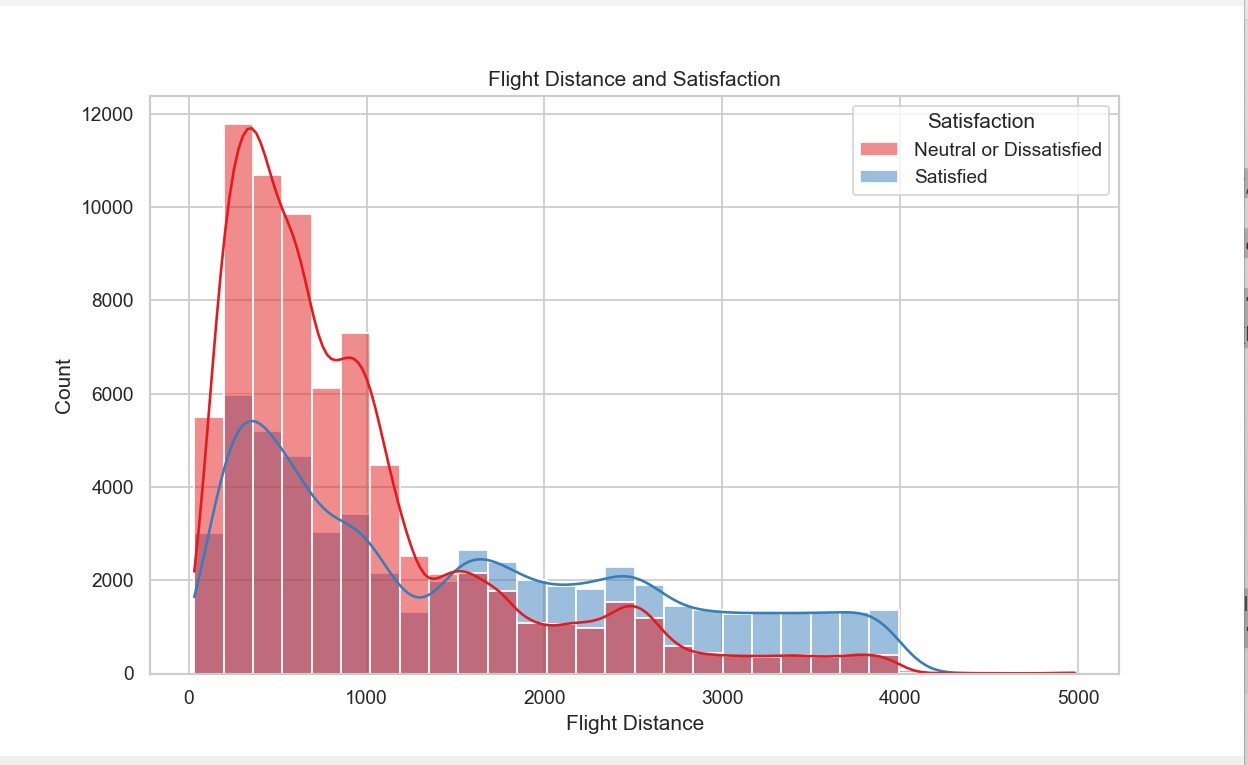
1. **Delays and Satisfaction**

The Horizontal Bar Chart (**Figure 12**) clearly indicates that average departure and arrival delays are lower for satisfied passengers. Delays remain a significant driver of dissatisfaction.

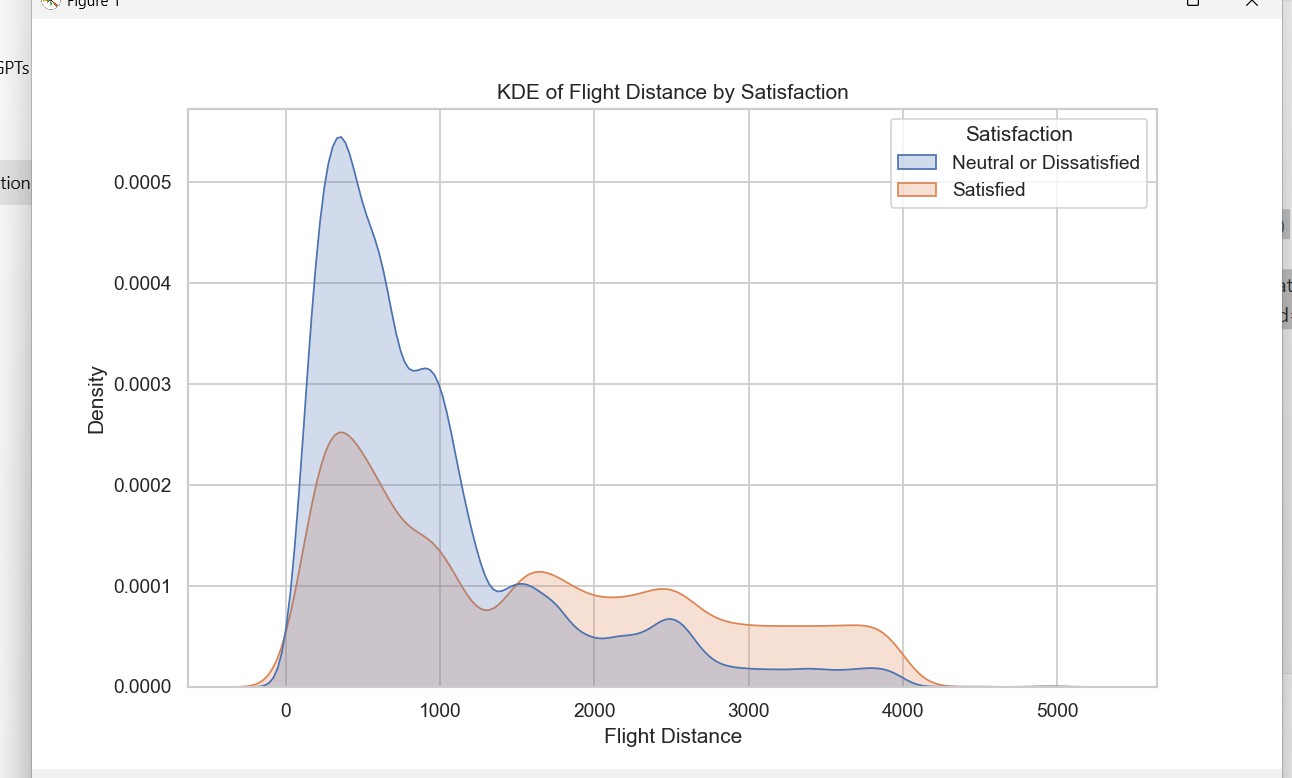
1. **Passenger Demographics**

Age Distribution (**Figure 13**) shows that passengers are mostly between 20–50 years old. There is a slight right-skew suggesting more younger passengers.

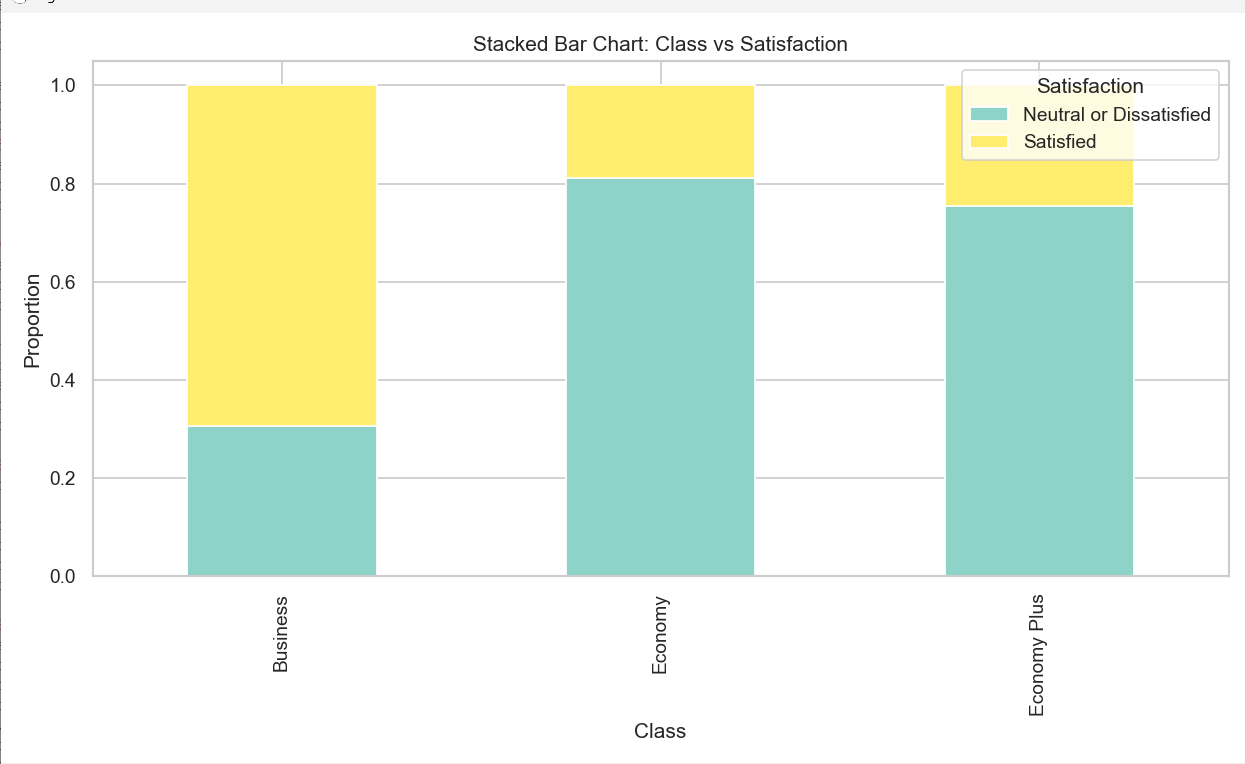
Type of Travel (**Figure 14**) indicates that Business travel dominates, and this group likely overlaps with the satisfied Business Class passengers shown earlier.



**Figure 7: Histplot**



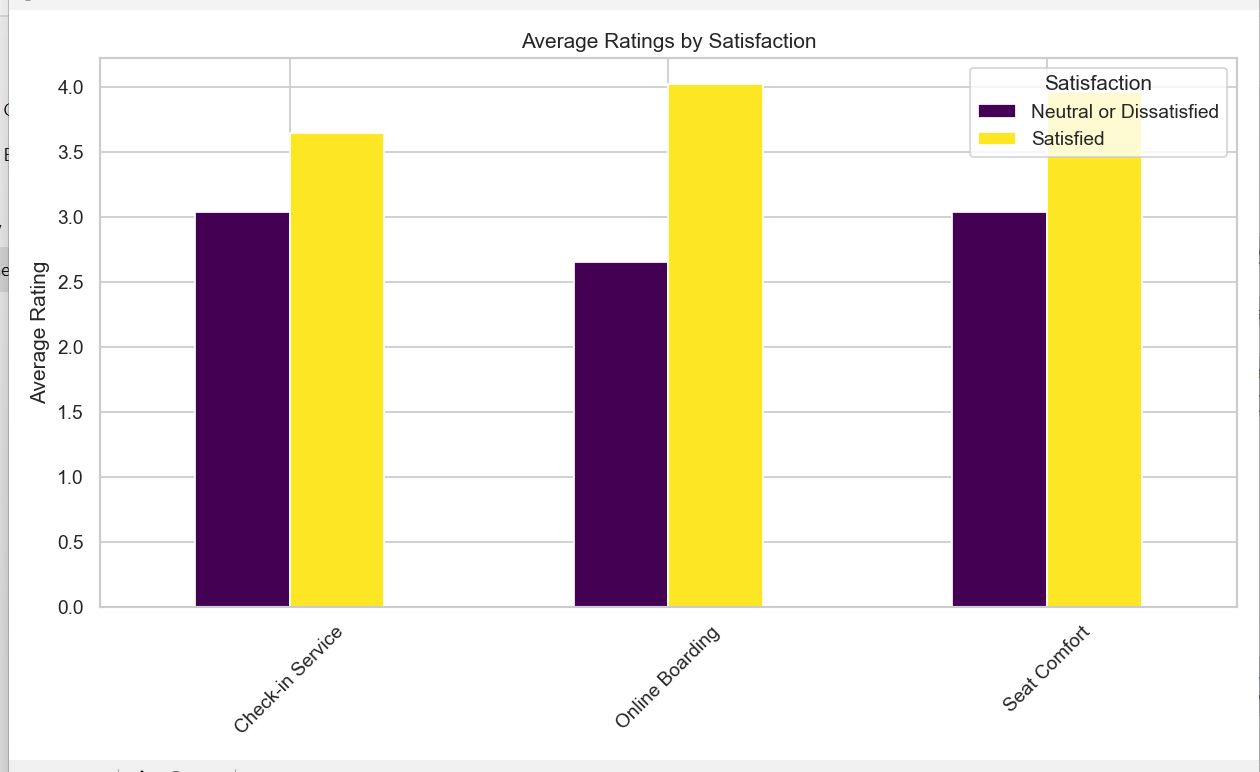
**Figure 8: KDE Plot**

****

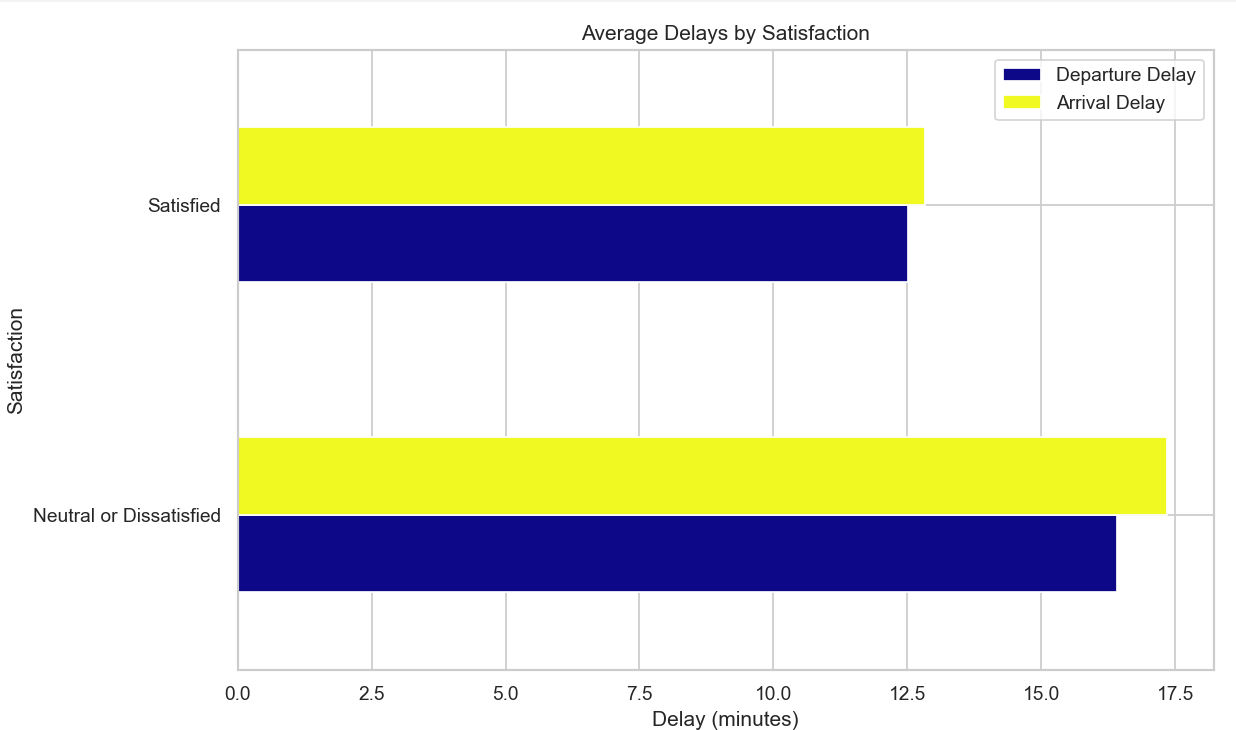
**Figure 9: Stacked Bar Chart**



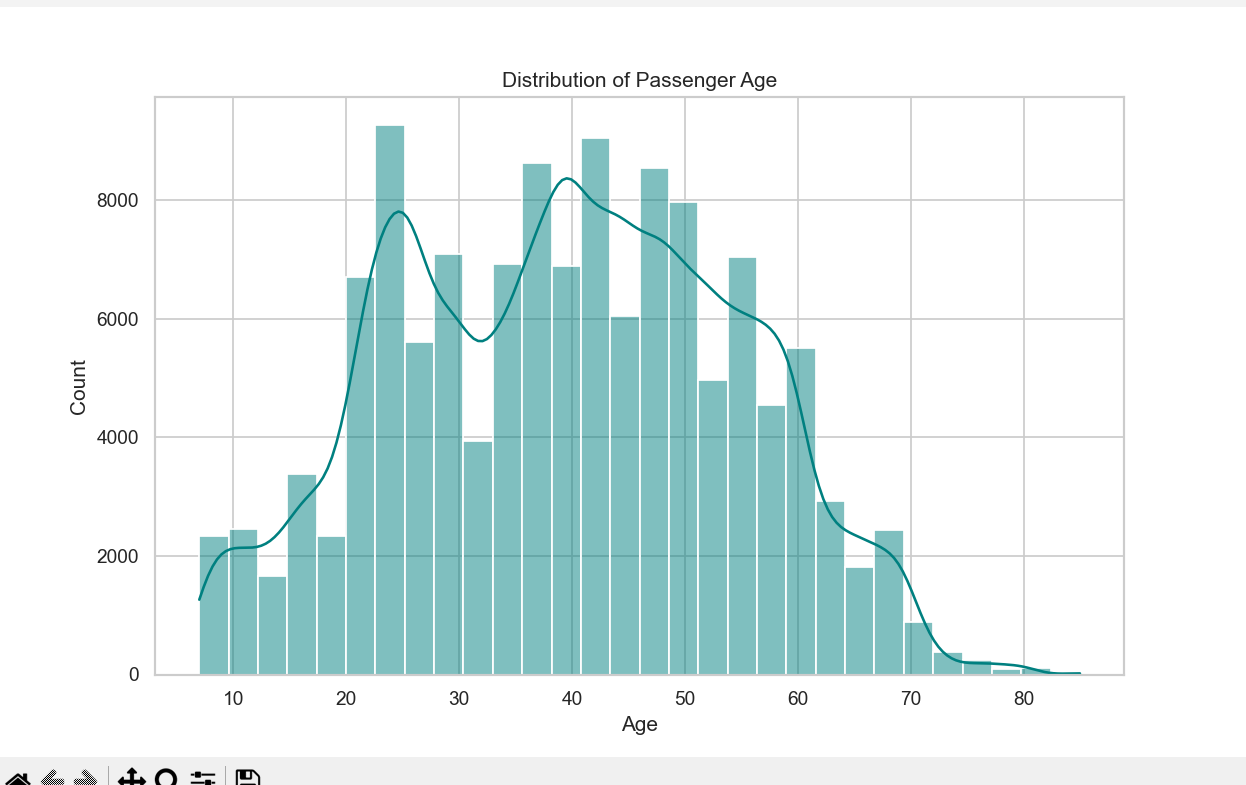
**Figure 10: Scatter Plot**

****

**Figure 11: Bar Chart**



**Figure 12: Horizontal Bar Chart**

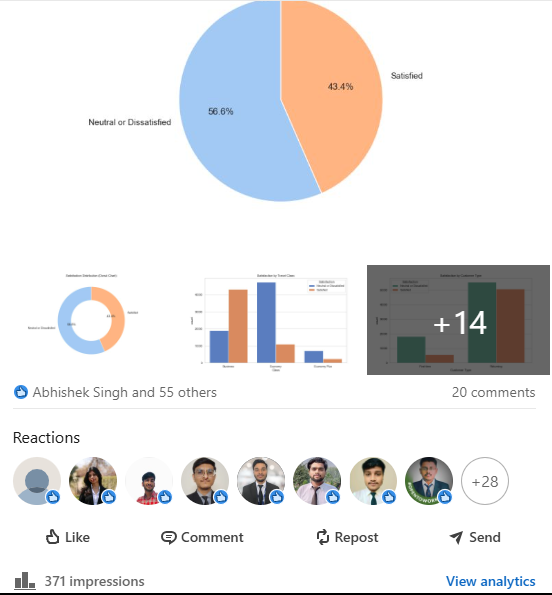
****

**Figure 13: Histogram**

**Linkedin Post**

**A screenshot of a computer

AI-generated content may be incorrect.**

****

**Limitations**

Even with promising insights and visual trends, this study has a few limitations:

* The analysis is based on historical survey data, and it does not include real-time behavioral or operational inputs such as live flight tracking, customer feedback systems, or dynamic service updates.
* Some important contextual factors, such as passenger mood, airline crew behavior, and real- time service disruptions (e.g., last-minute gate changes, unexpected turbulence), are not captured in the dataset.
* No predictive model was implemented in this phase, meaning we do not yet evaluate classifier performance on unseen data. While the dataset is suitable for binary classification, the project currently stops at exploratory analysis.
* As this is a static dataset, time-based satisfaction trends or changes in service impact over time are not modeled, which could offer deeper operational insights in a real-world application.

**Scope for Future Research**

While this project offers strong analytical insight into what drives airline passenger satisfaction, there are multiple promising directions for expanding its scope:

* **Machine Learning-Based Prediction Models:**

In future work, supervised learning models like Logistic Regression, Random Forest, or Gradient Boosted Trees can be applied to classify passengers into "Satisfied" or "Not Satisfied" categories based on service features. These models could enable airlines to proactively address potential dissatisfaction.

* **Real-Time Feedback Integration:**

Incorporating live data streams, such as:

* + In-flight service updates
  + Live delay tracking
  + Real-time sentiment from surveys or apps
  + Weather and route conditions could help create a more dynamic, responsive system to monitor and manage satisfaction in real time.
* **Advanced Modeling with NLP or Deep Learning:**

Future work can include Natural Language Processing (NLP) to extract sentiment from passenger comments, or deep learning models like Recurrent Neural Networks (RNNs) or Transformers to model satisfaction trends over sequential flights or routes.

* **Hybrid Modeling Approaches:**

A dual-model setup could be explored:

* + Classification Model to predict satisfaction (Yes/No)
  + Regression Model to estimate numerical satisfaction scores or likelihood This would allow for richer, more actionable predictions for airline decision-makers.
* **Visualization Dashboards for Airlines:**

Using tools like Tableau, Power BI, or Streamlit, an interactive dashboard could help airline staff view satisfaction metrics in real time and analyze which service points are underperforming.

**Conclusion**

This project presented a detailed data-driven analysis of airline passenger satisfaction using a real- world dataset. By performing extensive exploratory data analysis (EDA), we were able to identify key trends and service factors that influence how passengers perceive their travel experience.

Features such as Travel Class, Customer Type, Online Boarding, Check-in Service, and Delays were found to be closely associated with satisfaction. Satisfied passengers consistently rated service elements higher and tended to be frequent fliers or those in business class, while dissatisfied passengers were more affected by delays and inconsistent service ratings.

Though no machine learning model was implemented in this study, the dataset and analysis suggest strong potential for future predictive modeling. The use of classification algorithms could help airlines anticipate dissatisfaction, while real-time monitoring systems could ensure continuous improvement of service quality.

In summary, this project highlights the value of customer feedback and service metrics in shaping satisfaction. It serves as a foundation for building intelligent systems that can enhance airline service quality, improve customer retention, and increase operational awareness. With added layers like real-time data, sentiment analysis, and predictive modeling, this framework can evolve into a robust customer experience management tool for the aviation industry.

**8.REFERENCES**

1) **Bureau of Transportation Statistics**, “Passenger Air Travel Trends,” U.S. Department of Transportation, 2021. [Online]. Available: https://mavenanalytics.io  
• This reference offers datasets and trends related to passenger air transport in the U.S., providing crucial background and contextual insights for analyzing airline satisfaction.

2) **S. N. G. H. Murthy et al.**, "Airline Passenger Data Analysis Using Data Mining," *International Journal of Computer Applications*, vol. 78, no. 4, pp. 9–12, 2013.  
• This paper outlines data mining methods for analyzing airline passenger data, useful for examining ticketing patterns, service quality, and delay impacts on satisfaction.

3) **M. B. Nielsen et al.**, "Passenger Flow Prediction in Airports Using Data Mining and Machine Learning," *Journal of Transportation Engineering*, vol. 142, no. 7, pp. 04016042, 2016.  
• Focused on forecasting passenger movement in airport terminals, this work supports modeling satisfaction in relation to boarding efficiency and congestion.

4) **L. S. Tiwari and D. N. Rathi**, "Analysis of Flight Delays and Their Impact on Passenger Satisfaction," *Transport Policy*, vol. 35, pp. 114–121, 2014.  
• A relevant study investigating how flight delays affect passenger perceptions, directly aligning with delay-related satisfaction metrics in airline datasets.

5) **S. Pandey et al.**, "Predicting Flight Arrival Time Using Machine Learning Techniques," *Procedia Computer Science*, vol. 132, pp. 567–574, 2018.  
• Focused on arrival time predictions using ML, this paper can support efforts to handle missing or inconsistent "Arrival Time" data and its influence on satisfaction scores.