**Univariate Analysis**

**Summary Statistics:**

skewness and kurtosis analysis:

A screenshot of a computer screen

Description automatically generated

High Positive Skewness in Delays: Both departure and arrival delays exhibit very high positive skewness, indicating that while most flights are on time or have minimal delays, there are occasional flights with extremely long delays. This skewed pattern suggests that the majority of the inconvenience is caused by a few significantly delayed flights.

Extremely High Kurtosis in Delays: The departure and arrival delays not only have high skewness but also very high kurtosis. This points to a very sharp peak and thick tails in the distribution, suggesting that outliers (long delays) are not just infrequent but also extreme. These extreme values can have a disproportionately negative impact on overall passenger satisfaction.

Moderate to Low Kurtosis in Service Ratings: Most service-related variables (like inflight wifi, seat comfort, and cleanliness) have kurtosis values around or below 3, indicating distributions that are less prone to outliers and generally more uniformly spread around the mean. This suggests that passenger experiences with these services are relatively consistent.

These observations indicate that improving management and reduction of extreme delay incidents could significantly enhance overall passenger satisfaction, as the delays are not typical but when they occur, they're highly disruptive. Additionally, maintaining consistent service quality in other areas seems to be effective as indicated by the more uniform distribution of service ratings.

**Visualization:**

**Pie chart – key categorical variable “Class”**

**A pie chart with text and numbers

Description automatically generated**

**class presentation:**

we can see from the pie chart, our passengers are almost evenly split between Business and Eco classes, demonstrating the diversity in preferences for comfort and affordability among our travelers. Interestingly, only a small fraction, 7.2%, opts for Eco Plus, indicating a potential area for us to explore further. Understanding the reasons behind these choices can help us tailor our services to better meet our passengers' needs and identify opportunities for growth in the underrepresented Eco Plus category.

**Inference:**

The pie chart reveals that Business and Eco classes are almost equally preferred by passengers, while Eco Plus is chosen less frequently. This suggests a market divided fairly evenly between those seeking premium and standard services, with a smaller segment opting for upgraded economy options.

**A group of different colored graphs

Description automatically generated**

**class presentation:**

We've taken a close look at our passengers' travel preferences and satisfaction levels. Our data reveals a varied age demographic, suggesting our services cater to a wide range of travelers. Most flights are short-haul, which are the bread and butter of our operations. There's noticeable room for improvement in inflight wifi service, as indicated by the mid to lower end ratings from our customers. On a positive note, the majority of our flights depart on schedule, showcasing our commitment to timely operations.

Our passenger age distribution suggests varied travel needs, while a preference for short-haul flights indicates a strong regional market. Feedback on inflight wifi signals a chance for enhancement, and although punctuality is generally high, some notable delays highlight opportunities for operational improvements. These insights will guide our efforts to refine passenger experience and service efficiency.

**Inference:**

After examining 'Class', the following observations emerge from other key variables:

- Age Distribution: Passengers are diverse in age, with clusters around young adulthood and mid-life indicating varied travel purposes.

- Flight Distance: A higher frequency of short-haul flights is indicated, with long-distance travel being less common.

- Inflight Wifi Service: There's a spread in satisfaction, with many passengers rating the service in the mid to lower range, suggesting room for improvement.

- Departure Delays: Most flights depart with minimal delay, signifying efficient operations for the majority of services.

These distributions provide insights into demographic patterns, travel preferences, service quality, and operational efficiency.

noteworthy patterns in the data:

- The \*\*Age Distribution\*\* appears multi-modal, which might indicate distinct groups or types of travelers such as families, business travelers, and retirees.

- \*\*Inflight Wifi Service Ratings\*\* show a potential negative skew, with a lean towards lower satisfaction scores, which is an area that could be improved.

- \*\*Departure Delays\*\* are highly right-skewed, with most flights departing on time, but with a long tail of flights experiencing significant delays, suggesting infrequent but notable operational disruptions.

**Box Plot for Outliers and Variance**

A screenshot of a screen

Description automatically generated

The box plots for departure and arrival delays highlight a common central trend with most flights experiencing minimal delays. However, both plots reveal numerous outliers, indicating that a significant number of flights are subject to substantial delays. These outliers, which represent exceptional cases of extended wait times, are critical points for operational review to improve overall punctuality and customer satisfaction.

Inference:

The box plots for departure and arrival delays display a tightly clustered median, suggesting that most flights adhere to the schedule, yet the presence of outliers indicates instances of substantial delays. This pattern underscores an operational focus area—while general punctuality is achieved, attention to the outliers could enhance reliability and customer satisfaction.

**Stacked bar plot for flight services:**

**Pre-flight Services:**

Departure/Arrival time convenient

Ease of Online booking

Gate location

Check-in service

A group of bars with different colors

Description automatically generated with medium confidence

Presenation:

In reviewing the pre-flight service ratings for an airline, it's observed that passengers are generally satisfied with check-in services and the convenience of departure and arrival times, with many high ratings. However, there's a notable opportunity for improvement in the online booking experience, despite a decent number of favorable responses. Gate locations received a more neutral response, suggesting that while not a significant pain point, there is room for enhancement in this area. These insights highlight specific targets for the airline to improve its pre-flight services and enhance overall customer satisfaction.

Inference:

The visual analysis of pre-flight service ratings reveals insightful trends: Passengers are generally satisfied with the check-in service and the convenience of departure/arrival times, as indicated by the predominance of higher ratings. The online booking experience, while still favorably rated, shows a broader spread across the rating spectrum, suggesting some passengers face challenges, which could be an area for improvement. Gate location evaluations are more evenly distributed, indicating no strong consensus on satisfaction or dissatisfaction, pointing to an opportunity for a deeper dive into passenger preferences in this area.

**In-flight Services:**

Entertainment Services:

Inflight entertainment

Comfort Services:

Seat comfort

Leg room service

Hospitality Services:

Food and drink

On-board service

Cleanliness

Connectivity and Handling Services:

Inflight wifi service

Online boarding

Baggage handling

Inflight service

A graph of different colored bars

Description automatically generated with medium confidence

The chart presents satisfaction ratings for various airline services. Inflight entertainment and seat comfort have high counts of positive ratings, showing areas of strength. Food and drink, along with on-board service, show a balanced spread across ratings, suggesting variability in passenger experiences. Notably, inflight wifi and baggage handling received mixed feedback, with significant counts at lower ratings, indicating potential areas for improvement. These insights reveal passengers' priorities and service aspects that require attention to enhance overall satisfaction.

Inference:

The stacked bar charts clearly show that inflight entertainment and seat comfort are areas where passengers are most satisfied, with a significant number of high ratings. On the other hand, inflight wifi service and baggage handling have received lower ratings, signaling that these are the areas where passengers are less satisfied and where the airline could focus its improvement efforts. These visual indicators are crucial for quickly pinpointing service strengths and weaknesses.

**Understand Variable Relationships with Pair Plot**

**A chart of different colored dots

Description automatically generated with medium confidence**

**The key takeaway for the airline from this plot is that higher customer satisfaction is closely linked to better ratings for inflight wifi service, seat comfort, and cleanliness. Focusing on enhancing these aspects of the customer experience could significantly improve overall satisfaction.**

**Another important relationship evident from the plot is that customer satisfaction seems to be adversely affected by longer departure and arrival delays. Minimizing these delays could be another critical factor for improving passenger satisfaction.**

**Pair plot:**

The pair plot displays several interesting observations:

Most prominently, there is a distinct clustering of points between satisfied and neutral or dissatisfied customers across several service quality variables, such as inflight wifi service, seat comfort, and cleanliness. Satisfied customers tend to give higher ratings in these areas, indicating a strong correlation between these factors and overall satisfaction. Additionally, there's a wide distribution of ages across satisfaction levels, suggesting age is not a primary determinant of satisfaction. Flight distance does not show a clear trend with satisfaction, indicating that longer flights do not necessarily lead to decreased satisfaction. Delays in departure and arrival do show some impact, with higher satisfaction generally associated with shorter delays, although the relationship is not as pronounced as with the service quality variables. This insight can inform airline service improvements focusing on customer experience elements directly linked to satisfaction.

Heat map:

A screenshot of a graph

Description automatically generated

The heatmap shows strong positive correlations between variables related to inflight services—such as seat comfort, inflight entertainment, and on-board service—which suggests that improvements in one of these areas are likely to enhance the others, thereby potentially increasing overall passenger satisfaction. There is also a notable positive correlation between ease of online booking and departure/arrival time convenience, indicating that a user-friendly online booking system might influence passengers' perception of their overall travel experience. These insights can be crucial for the airline to focus on interconnected areas that could lead to higher customer satisfaction and loyalty.

The heatmap indicates that variables related to customer service, such as seat comfort, inflight entertainment, and cleanliness, have strong positive correlations with each other, implying that improvements in one area are likely to enhance the others. Moreover, higher correlations are observed between online booking, gate location, and departure/arrival time convenience, suggesting that a streamlined booking experience is associated with overall convenience perceptions. There are also high correlations within service quality aspects, indicating these are closely interlinked in influencing customer experiences.

#### a. **Joint Plot with KDE and Scatter Representation**

* **Purpose**: Examine the relationship between passenger age and flight distance. The KDE will help show the distribution density, while scatter points will reveal individual data relationships.
* **Insight**: Understand if longer flights are preferred by certain age groups, which could influence flight scheduling and marketing strategies.

Joint plot with KDE

A diagram of different sizes and numbers

Description automatically generated with medium confidence

This visualization showcases a joint plot combining a scatter plot for age against flight distance with the distribution of each variable shown as histograms on the top and right sides. The dense clustering at lower flight distances indicates a high volume of short-haul flights across all age groups, while the sparser points at higher flight distances highlight less frequent but consistent interest in long-haul flights. The age distribution suggests that passengers of a wide age range are traveling, with no clear preference for flight distance among different age groups. This could imply that an airline's strategy focusing on enhancing short-haul flight efficiency and comfort could benefit a broad customer base, while also maintaining a commitment to quality long-haul services to cater to all age demographics.

A graph with blue lines

Description automatically generated

The joint plot reveals key demographics and travel preferences, with two main age groups predominantly flying short distances, indicating a larger market for short-haul flights. This insight can guide airlines to tailor services, optimize schedules, and target marketing strategies to these segments. Additionally, the interest in longer flights, especially among older passengers, can inform premium service offerings and loyalty programs, potentially increasing profitability for long-haul routes. By addressing the distinct needs of these groups, airlines can enhance customer satisfaction and operational efficiency.

Inference:

The joint plot with KDE and scatter representation suggests a bimodal distribution for both passenger age and flight distance. There are two prominent age groups who fly: one younger (likely in their late 20s to early 30s) and another older (likely around retirement age, late 50s to early 70s). The flight distance shows high density around shorter flights (approximately less than 1000 miles) and a smaller, less dense mode around longer flights (approximately 2000 to 3000 miles). A moderate number of outliers exist for very long-distance flights, potentially indicating a lesser but significant interest in long-haul flights among passengers. The overlay of scatter points, albeit sparse, confirms the presence of individuals outside the common trends, particularly among older passengers undertaking long-distance travel. This data can inform targeted services and marketing strategies, as there appear to be two key demographic groups with different flight distance preferences.

#### **Violin Plot**

* **Purpose**: Compare the satisfaction levels across different cabin types (Economy, Business, First Class).
* **Insight**: Visualize distribution and density of satisfaction, highlighting which cabin types may need improvement or are performing well.

A screenshot of a graph

Description automatically generated

Analyzing the violin plots for flight distance and departure delays by travel type and class provides significant insights. Business class passengers, especially those traveling for business, experience a broad range of flight distances and more variability in departure delays, suggesting a potential area for service improvement. Economy and Economy Plus classes have a higher concentration of shorter flights for personal travel, indicating an opportunity for targeted marketing strategies. Addressing the variability in departure delays, particularly for high-revenue Business class seats, could enhance customer satisfaction and loyalty. Streamlining operations to mitigate extreme delays and focusing on consistent service delivery for business travelers could lead to better resource allocation, cost savings, and a competitive edge in the airline market.

Cluster map:

A graph of different colored squares

Description automatically generated

The cluster map indicates distinct customer segments by travel type and loyalty, with gender further refining each group. It's apparent that loyalty varies significantly between business and personal travel, with 'Loyal Customer' segments showing high engagement in business travel, suggesting they may be frequent flyers possibly benefitting from loyalty programs. In contrast, 'disloyal Customer' segments are more prominent in personal travel, hinting at different factors influencing their travel choices, such as price sensitivity or lack of compelling loyalty incentives. For airlines, this implies a strategic opportunity to tailor loyalty programs to incentivize personal travelers, while maintaining robust benefits for loyal business travelers to sustain their engagement. Additionally, the balanced gender distribution across segments suggests that gender-specific marketing may not be as effective as strategies that focus on the travel type and loyalty status of customers.

QQplot:

A graph of a number of different types of data

Description automatically generated with medium confidence

The QQ-plots for 'Flight Distance', 'Departure Delay in Minutes', 'Arrival Delay in Minutes', and 'Age' reveal that none of these variables are perfectly normally distributed. The 'Flight Distance' and 'Age' plots show a slight deviation from the normal line at the tails, suggesting mild outliers. On the other hand, both 'Departure Delay in Minutes' and 'Arrival Delay in Minutes' exhibit a pronounced positive skew with many extreme values far from the median, indicating a significant presence of long delays. For the airline, this could mean that while most flights may depart or arrive with minimal delay, there are occasional but noteworthy disruptions that could significantly impact passenger satisfaction. By addressing the root causes of these extreme delays, the airline could greatly enhance overall customer experience and potentially improve operational efficiency.

Outlier removal- Box plot:

A group of graphs showing different colored lines

Description automatically generated with medium confidence

The box plots reveal that outlier removal significantly reduces extreme values in 'Flight Distance' and delay times while only slightly adjusting the 'Age' distribution. Post-filtering, 'Flight Distance' and 'Arrival Delay in Minutes' have a noticeably lower range, indicating the removal of rare but severe long-distance flights and extended delays. These changes suggest that most passengers experience more moderate flight distances and delays. The 'Age' distribution remains relatively unchanged, implying a consistent age profile among passengers. This refined data can enable the airline to focus on typical experiences and streamline operations, potentially enhancing overall customer satisfaction.

LM reg plot

A graph showing different colored lines

Description automatically generated

The dual-axis plot illustrates that satisfaction ratings for both inflight wifi service and food and drink remain relatively constant across different ages. The nearly flat regression lines suggest that there's no significant difference in the satisfaction levels with these services among various age groups. For the airline, this could indicate that the willingness to pay for such services might not be significantly different across age demographics. Therefore, the airline might consider universally applicable improvements or marketing strategies for these services rather than age-targeted campaigns. This insight could guide resource allocation to ensure a consistent service experience that appeals to all passengers, potentially leading to an overall increase in customer satisfaction and willingness to pay.