Final Project Report
Airline Passenger Satisfaction
Group - DatAlchemy

Uday Kiran G. S. | Chandana Bhat T.

Naveen Kumar Reddy V | Mai N. | Guntaka Satish H. R.

DePaul University - College of Computing and Digital Media (CDM)

DSC 465 Data Visualization

Prof. Jodi Hill

June 10, 2024

Tables of Contents

Introduction

Due to the unique characteristics and fixed nature of passengers' flight journeys, airline companies are continuously seeking ways to improve customer satisfaction, from booking flight tickets to baggage claim. In this project, our objective is to identify areas for improvement in airline services to enhance passenger satisfaction. Specifically, through to two exploration approaches of passenger Age Groups and Satisfaction Levels. Therefore, the majority of the visualizations will emphasize how Satisfaction Levels are recorded among different passenger demographics. The framework will focus on identifying passenger segments with low service ratings and analyzing their survey responses to pinpoint the services that need improvement.
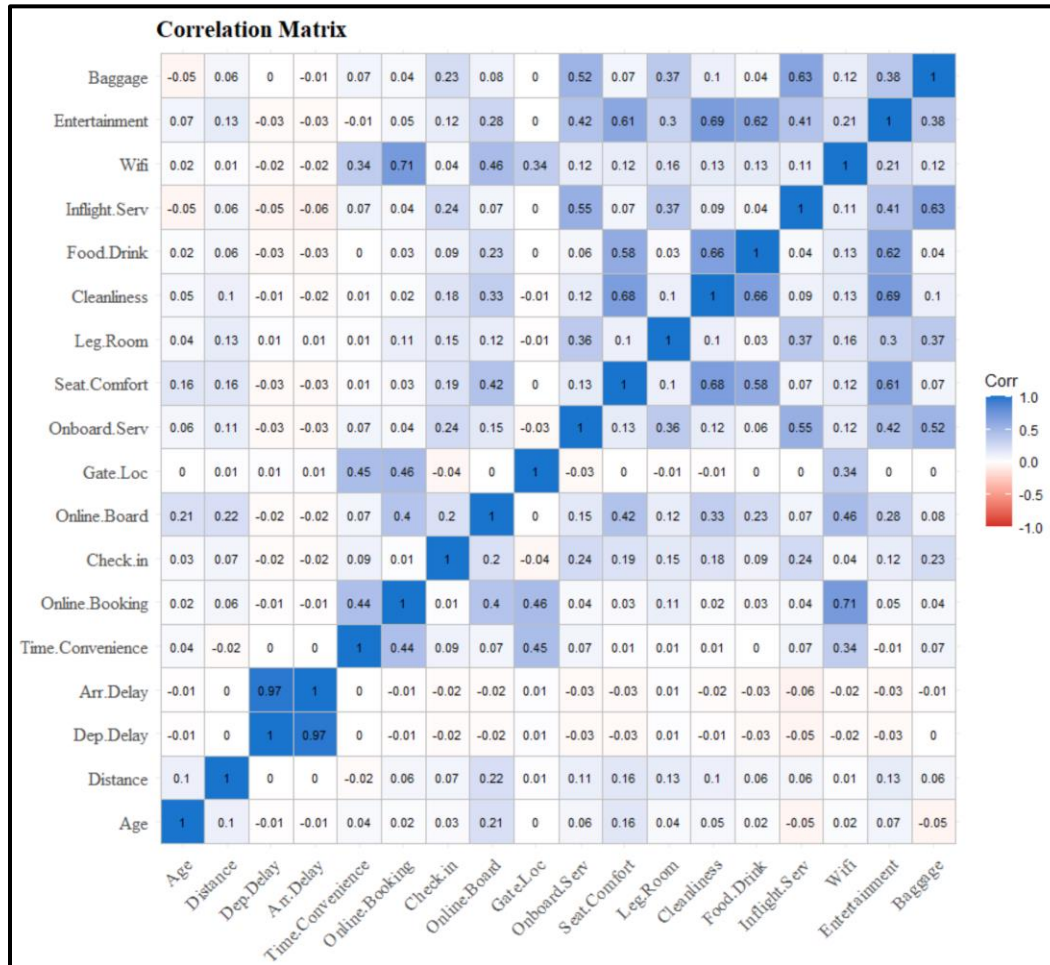
The data was originally retrieved from Kaggle's Airline Passenger Satisfaction dataset. There are 129,880 observations with 24 attributes, encompassing various data types that allow robust visualizations. Furthermore, a fixed color palette was used to maintain aesthetic consistency, and data scaling and plotting techniques were meticulously revised multiple times to ensure readability for non-technical audiences and to effectively convey the story. Lastly, visualization interactivity was not implemented in this project due to simplicity for non-technical viewers to understand and analyze. However, since the majority of the visuals were created from multiple dimensions, interactivity could be added in the future to provide a more comprehensive view of the data.

Data Description

Field	Description	Data type
ID	Unique passenger identifier	**Numerical**
Age	Age of the passenger	
Flight Distance	Flight distance in miles	
Departure Delay	Flight departure delay in minutes	
Arrival Delay	Flight arrival delay in minutes	
Gender	Gender of the passenger (Female/Male)	**Categorical**
Customer Type	Type of airline customer (First-time/Returning)	
Type of Travel	Purpose of the flight (Business/Personal)	
Class	Travel class in the airplane for the passenger seat	
Satisfaction	Overall satisfaction (Satisfied-Neutral/Unsatisfied)	
Departure and Arrival Time | Convenience | Ease of Online Booking | Check-in Service | Online Boarding | Gate Location | Seat Comfort | Leg Room Service | Cleanliness | Food and Drink | In-flight Service | In-flight Wi-Fi Service | In-flight Entertainment | Baggage Handling | On-board Service	14 Passenger Survey Reponses Satisfaction levels: 1 (lowest) to 5 (highest) - 0 ("not applicable")	**Ordinal**

Exploratory Analysis

Variable Correlation

Correlation Matrix

|	Age	Distance	Dep.Delay	Arr.Delay	Time.Convenience	Online.Booking	Check.in	Online.Board	Gate.Loc	Onboard.Serv	Seat.Comfort	Leg.Room	Cleanliness	Food.Drink	Inflight.Serv	Wifi	Entertainment	Baggage
Baggage	-0.05	0.06	0	-0.01	0.07	0.04	0.23	0.08	0	0.52	0.07	0.37	0.1	0.04	0.63	0.12	0.38	1
Entertainment	0.07	0.13	-0.03	-0.03	-0.01	0.05	0.12	0.28	0	0.42	0.61	0.3	0.69	0.62	0.41	0.21	1	0.38
Wifi	0.02	0.01	-0.02	-0.02	0.34	0.71	0.04	0.46	0.34	0.12	0.12	0.16	0.13	0.13	0.11	1	0.21	0.12
Inflight.Serv	-0.05	0.06	-0.05	-0.06	0.07	0.04	0.24	0.07	0	0.55	0.07	0.37	0.09	0.04	1	0.11	0.41	0.63
Food.Drink	0.02	0.06	-0.03	-0.03	0	0.03	0.09	0.23	0	0.06	0.58	0.03	0.66	1	0.04	0.13	0.62	0.04
Cleanliness	0.05	0.1	-0.01	-0.02	0.01	0.02	0.18	0.33	-0.01	0.12	0.68	0.1	1	0.66	0.09	0.13	0.69	0.1
Leg.Room	0.04	0.13	0.01	0.01	0.01	0.11	0.15	0.12	-0.01	0.36	0.1	1	0.1	0.03	0.37	0.16	0.3	0.37
Seat.Comfort	0.16	0.16	-0.03	-0.03	0.01	0.03	0.19	0.42	0	0.13	1	0.1	0.68	0.58	0.07	0.12	0.61	0.07
Onboard.Serv	0.06	0.11	-0.03	-0.03	0.07	0.04	0.24	0.15	-0.03	1	0.13	0.36	0.12	0.06	0.55	0.12	0.42	0.52
Gate.Loc	0	0.01	0.01	0.01	0.45	0.46	-0.04	0	1	-0.03	0	-0.01	-0.01	0	0	0.34	0	0
Online.Board	0.21	0.22	-0.02	-0.02	0.07	0.4	0.2	1	0	0.15	0.42	0.12	0.33	0.23	0.07	0.46	0.28	0.08
Check.in	0.03	0.07	-0.02	-0.02	0.09	0.01	1	0.2	-0.04	0.24	0.19	0.15	0.18	0.09	0.24	0.04	0.12	0.23
Online.Booking	0.02	0.06	-0.01	-0.01	0.44	1	0.01	0.4	0.46	0.04	0.03	0.11	0.02	0.03	0.04	0.71	0.05	0.04
Time.Convenience	0.04	-0.02	0	0	1	0.44	0.09	0.07	0.45	0.07	0.01	0.01	0.01	0	0.07	0.34	-0.01	0.07
Arr.Delay	-0.01	0	0.97	1	0	-0.01	-0.02	-0.02	0.01	-0.03	-0.03	0.01	-0.02	-0.03	-0.06	-0.02	-0.03	-0.01
Dep.Delay	-0.01	0	1	0.97	0	-0.01	-0.02	-0.02	0.01	-0.03	-0.03	0.01	-0.01	-0.03	-0.05	-0.02	-0.03	0
Distance	0.1	1	0	0	-0.02	0.06	0.07	0.22	0.01	0.11	0.16	0.13	0.1	0.06	0.06	0.01	0.13	0.06
Age	1	0.1	-0.01	-0.01	0.04	0.02	0.03	0.21	0	0.06	0.16	0.04	0.05	0.02	-0.05	0.02	0.07	-0.05

Corr
1.0
0.5
0.0
-0.5
-1.0

We have plotted a correlation matrix to understand the relationships between various factors in the dataset. Key variables include 18 numerical columns such as Age, Flight Distance, Delays, Service ratings (e.g., in-flight service, seat comfort, cleanliness), and overall satisfaction indicators. The visualization was created using the R programming language within the RStudio environment using 'dplyr' and 'ggcorrplot' libraries. This correlation matrix reveals several interesting patterns and relationships among variables related to passenger satisfaction. Notably, there are strong positive correlations between "Seat Comfort" and "Cleanliness" (0.70), indicating that improving both can significantly enhance passenger satisfaction. Similarly, "Seat Comfort" and "In-flight Service" (0.43), as well as "In-flight Service" and "Baggage Handling" (0.65), show that high-quality service is interconnected across different areas. There's also a noticeable positive correlation between "Online Booking" and "WiFi Service" (0.64), suggesting that technological conveniences play a vital role in passenger satisfaction. On the other hand, there is a very strong

negative correlation between "Departure Delay" and "Arrival Delay" (-0.97), indicating that addressing departure delays can substantially reduce arrival delays.

This matrix helps us understand the relationships between different variables in the dataset, highlighting which factors are most strongly associated with passenger satisfaction. By identifying both positive and negative correlations, we can pinpoint key areas for improvement, such as enhancing service quality, seat comfort, and in-flight amenities, or addressing flight delays. This comprehensive view of data correlations provides valuable insights that are essential for making data-driven decisions to improve overall passenger satisfaction.

Variable Distribution
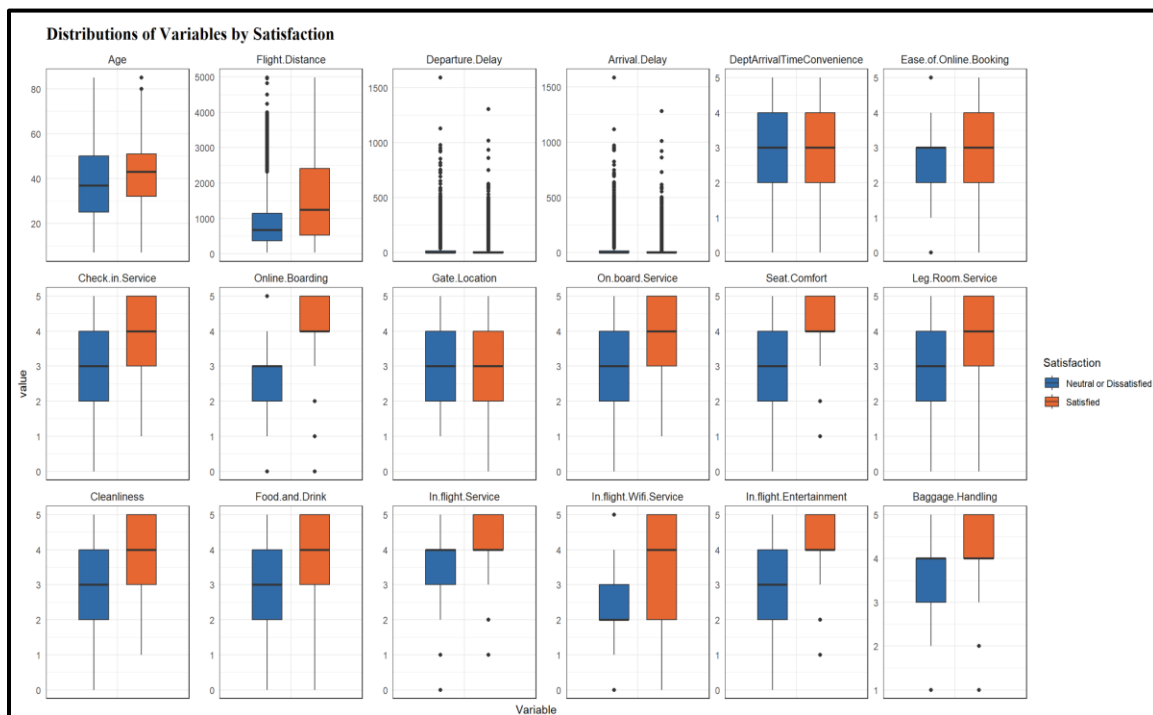
The panel box plot explores customer satisfaction data segmented by age, flight distance, delays, and various service ratings. Each plot contrasts satisfaction levels between "Neutral or Dissatisfied" and "Satisfied" customers. Notable differences are seen in Online Boarding, Seat Comfort, In-flight Service, In-flight Entertainment, and Baggage Handling, where satisfied customers rate higher. Conversely, no noticeable difference is observed in Departure and Arrival Time Convenience and Gate Location. This suggests that service quality impacts customer satisfaction more than operational factors like delays, leading to a focus on service quality.

Box plots were chosen for their effectiveness in displaying distribution and variability, making them suitable for exploring continuous variables and comparing distributions. Benefits include showing the median, quartiles, and potential outliers, providing a rich summary compared to simpler graphs. The methods used create rich displays by highlighting variability and outliers, offering a nuanced understanding beyond simple averages. Design criteria included a clear color

palette, with discrete color scheme for categorical variables: "Neutral or Dissatisfied" and "Satisfied," ensuring distinguishability and maintaining clarity. The visualizations were organized to facilitate easy comparison.

Using R with libraries like ggplot2 enabled precise control over visuals, facilitating the creation of cohesive and informative panel box plot. Data manipulation with tidyverse streamlined the preprocessing steps. Reshape2 was instrumental in transforming the data for effective visualization.

Survey Responses Pattern

Distribution of Survey Responses for Airline Services
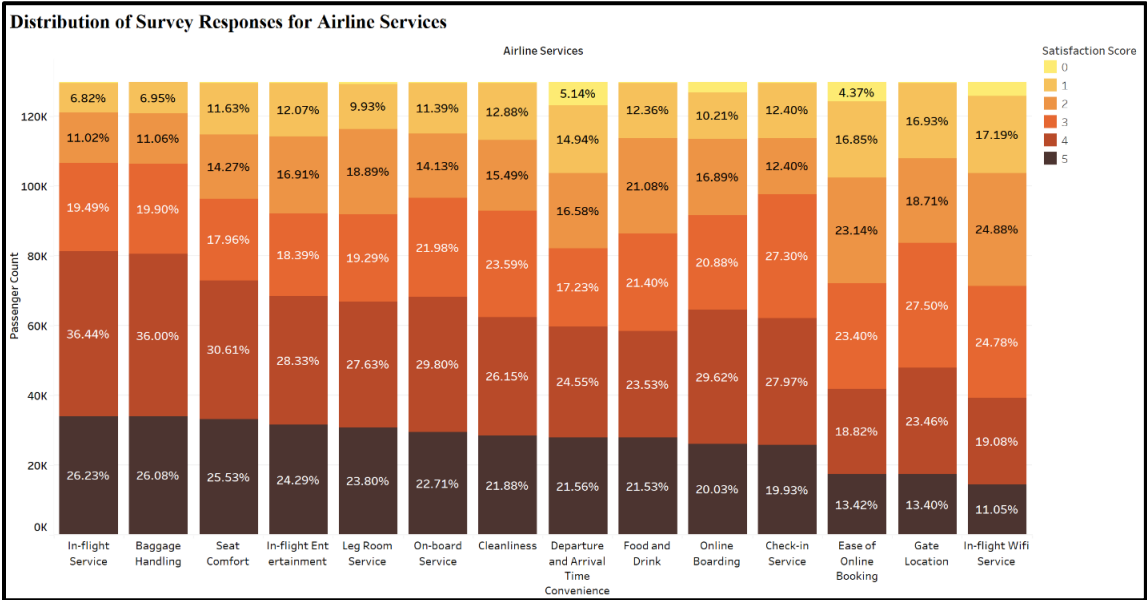
Given the survey scale of 1 for lowest to 5 for highest, and 0 means "not applicable", the graph shows the percentage of survey responses falling into each airline service. The survey results indicate that passengers are generally neutral or slightly satisfied among all airline services, with scores of 3 and 4, and around 50% of responses rating their satisfaction at levels 4 or 5. Specifically, In-flight Service, Baggage Handling and Seat Comfort have relatively higher satisfaction scores. However, some services fall short of these high expectations. Specifically, Ease of Online Booking, Gate Location, and In-flight Wi-Fi Service tend to receive ratings between 3 and 4. Additionally, the distribution of satisfaction scores varies significantly across different services, highlighting specific area of services where passengers are more or less satisfied. For instance, In-flight Entertainment and Food and Drink services have a more balanced distribution of satisfaction scores, showing both high and low satisfaction levels.

Noticeably, the highest satisfaction level observed across all services is 4, which may reflect the complexity and variability of the airline industry's service offerings due to partnerships with various third parties like airports, cleaning services, and online booking platforms. This pattern emphasizes an ongoing challenge for airline business to maintain consistent service standards across different operation platforms. Overall, this visual provides a comprehensive overview of

how passengers rate various aspects of their flight experience, enabling targeted strategies to enhance specific areas that impact overall satisfaction. Given the information, a stacked bar graph was created using Tableau with an ordinal color palette. The sequential color scheme was applied to clearly depict the ordered survey score levels and effectively deliver the information.

Passenger Distribution
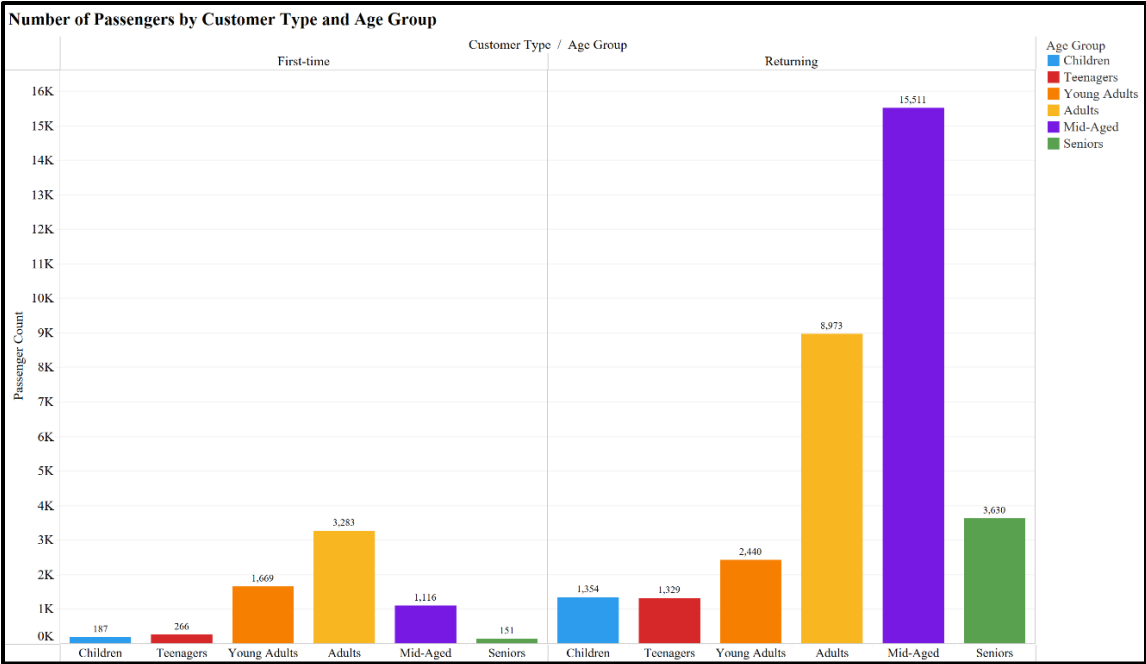
The above bar chart was created using Tableau. This chart illustrates the number of airline passengers segmented by Customer Type (first-time and returning) and Age Group. Among First-time passengers, adults are the largest group with 3,283 passengers, followed by young adults with 1,669 passengers, while children are the least represented group with only 187 passengers. In contrast, among Returning passengers, the mid-aged group has the highest count with 15,511 passengers, followed by adults with 8,973 passengers, and teenagers are the least represented group with 1,329 passengers. This segmentation reveals significant differences in passenger demographics, highlighting that returning passengers are predominantly mid-aged and adults, whereas first-time passengers are mostly adults and young adults. Bar charts were chosen for this analysis because they effectively display categorical data, making it easy to compare different groups at a glance. This technique is appropriate for exploring the distribution of passengers across various age groups and customer types, providing a clear visual representation of the differences and commonalities within the data. For this analysis, we have grouped the Age column, which ranges from 7 to 85, into six different categories for clearer insights, and you can find these categories in the table we have provided.

Age Groups					
Children: 0 - 12	**Teenagers:** 13 - 17	**Young Adults:** 18 - 24	**Adults:** 25 - 39	**Mid-Aged:** 40 - 59	**Seniors:** > 60

Explanatory Visualizations

Passenger Demographics


Age Distribution of Travelers by Flight Distance
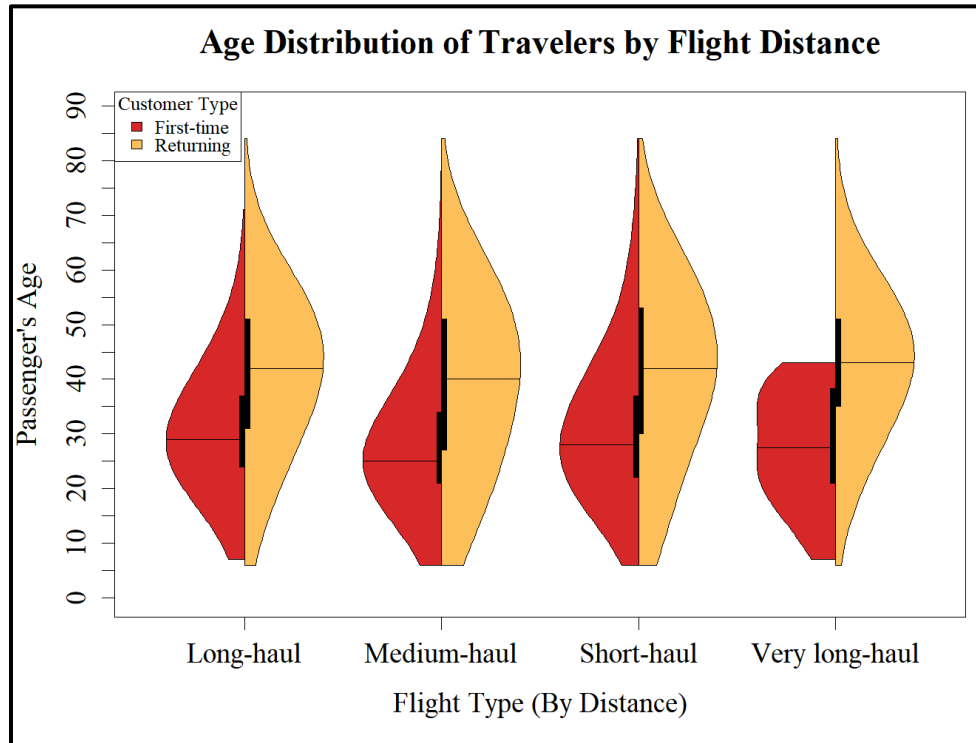
To profile first-time and returning travelers by their age, we use a split violin plot. The figure above shows the age distribution of travelers across different flight distances. The violin plot is split into two halves, first-time and returning travelers. Using this plot, we can identify the age demographics of the travelers to various destinations. The horizontal line in the plot denotes the median age and the vertical bar is the interquartile range.

Since, we want to analyze and infer the age distribution across first-time and returning passengers, plots like violin plot, box plot and kernel density curves are useful to visualize summary statistics (mean, median, mode and quartile) and distribution of samples. We decided to use violin plot for this visualization because it offers the best features of both box plot and density curves i.e. variation in distribution, summary statistics and outliers more intuitively. The violin plot is split into customer type, a variable with only two possible values, to allow finer comparison between the two groups. To create this visualization, the flight distance column, a numerical variable was binned into 4 groups of distance type or "hauls". The dataset is then grouped by the variable customer type (first-time or returning). This processed dataset is used to create the split violin plot.

Flight Distance Groups			
Short haul:	**Medium haul:**	**Long haul:**	**Very long haul:**
< 500	<= 1,500	<= 3,000	> 3,000

## Passenger Satisfaction Level
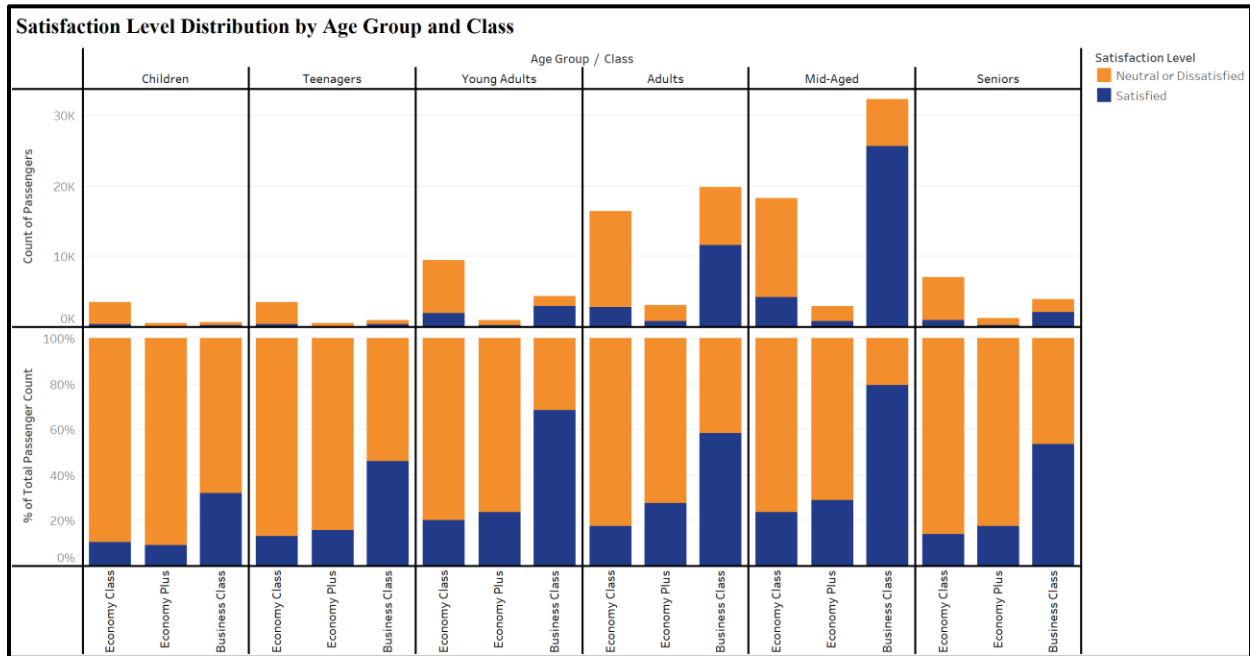
The objective of this visual is to show passenger satisfaction distribution with respect to Age Group and Class attributes. To fully capture this information, preliminary calculation steps were taken to create rich and deep displays of the data. Note that the Satisfaction column has two values: 'Neutral or Dissatisfied' and 'Satisfied'.

- **Step 1:** Calculate total passengers count per sub-group using Class, Age Group and Satisfaction attributes.
- **Step 2:** For each sub-group, calculate the percentage of total passenger count with respect to Satisfaction attribute. Attribute name: '% of Passenger along Satisfaction'.
- **Step 3:** For each sub-group, calculate the number of passengers by Age Group and Class using Create Calculated Field - formula:**{FIXED[Age Group],[Class]:COUNT([ID])}**, set field to attribute type. Attribute name: 'Number of Passengers by Age Group/Class'.

Stacked and proportional stacked bar plots were chosen to display newly calculated columns. Specifically, stacked bar plot emphasizes count of passengers of each sub-group with respect to the entire dataset. While proportional stacked bar plot focuses in displaying passenger satisfaction level as percentages within each sub-group. For other visual alternatives, small multiples were considered. However, the overall visual will then be very scatter and does not give a comprehensive view as determined visual choice. As for aesthetic enhancements, additional techniques such as font size, color, naming convention were implemented to ensure readability and understanding for non-technical audiences.

### *Service Satisfaction Across Airline Classes*
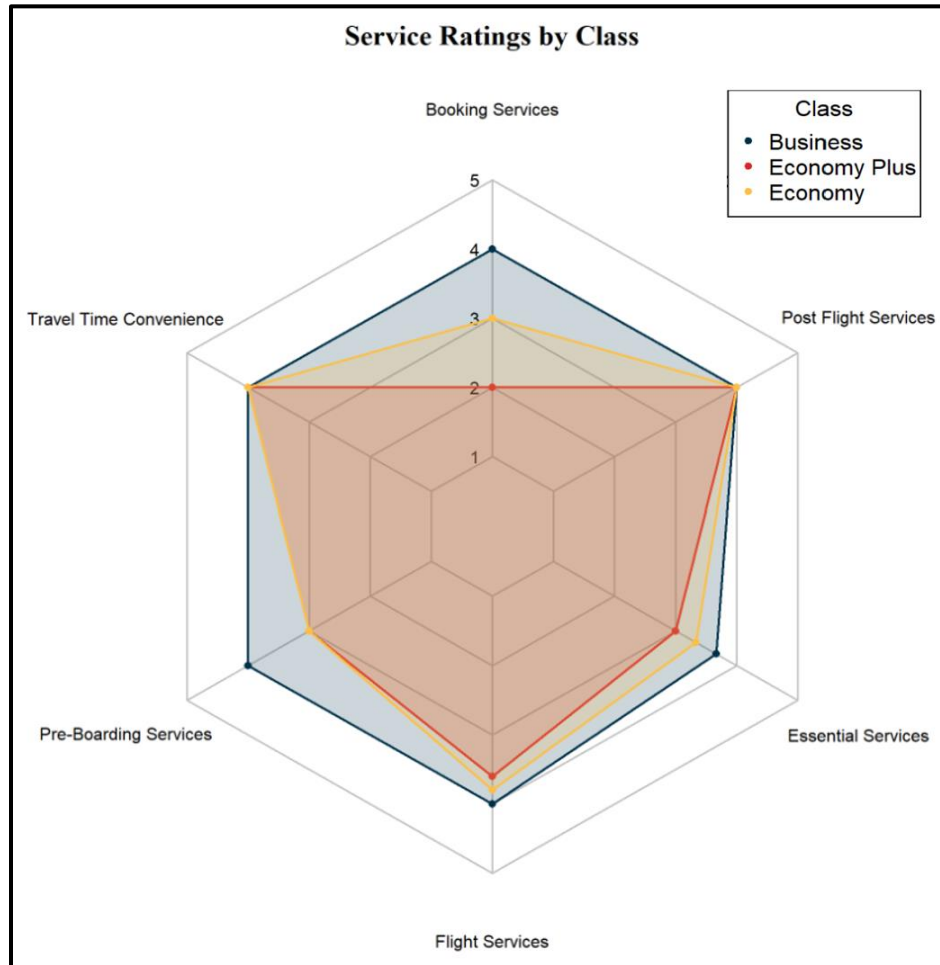
The radar chart shown above maps service ratings across different classes of airline service: Business, Economy Plus, and Economy. Each axis represents a different service category, including Booking Services, Travel Time Convenience, Post Flight Services, Essential Services, Flight Services, and Pre-Boarding Services. By addressing these aspects, the radar chart serves as an integral part of the explanatory data analysis, providing valuable insights that guide strategic decisions to improve overall passenger satisfaction.

We chose this type of chart because it effectively displays multivariate data and compares multiple variables simultaneously. It highlights the strengths and weaknesses in each service category, making it ideal for analyzing passenger satisfaction across different classes.

We grouped all 14 service ratings into comprehensive categories for a holistic view of passenger experience across classes. To enhance clarity and highlight key patterns, several refinements were made: labels were added, titles were adjusted, and distinct colors were used for each class (Business: #003049, Economy Plus: #D62828, Economy: #FCBF49). This radar chart was created using R in RStudio with libraries such as 'dplyr', 'scales', and 'fmsb', involving coding to set up the chart, define axes, and plot data points for each class.

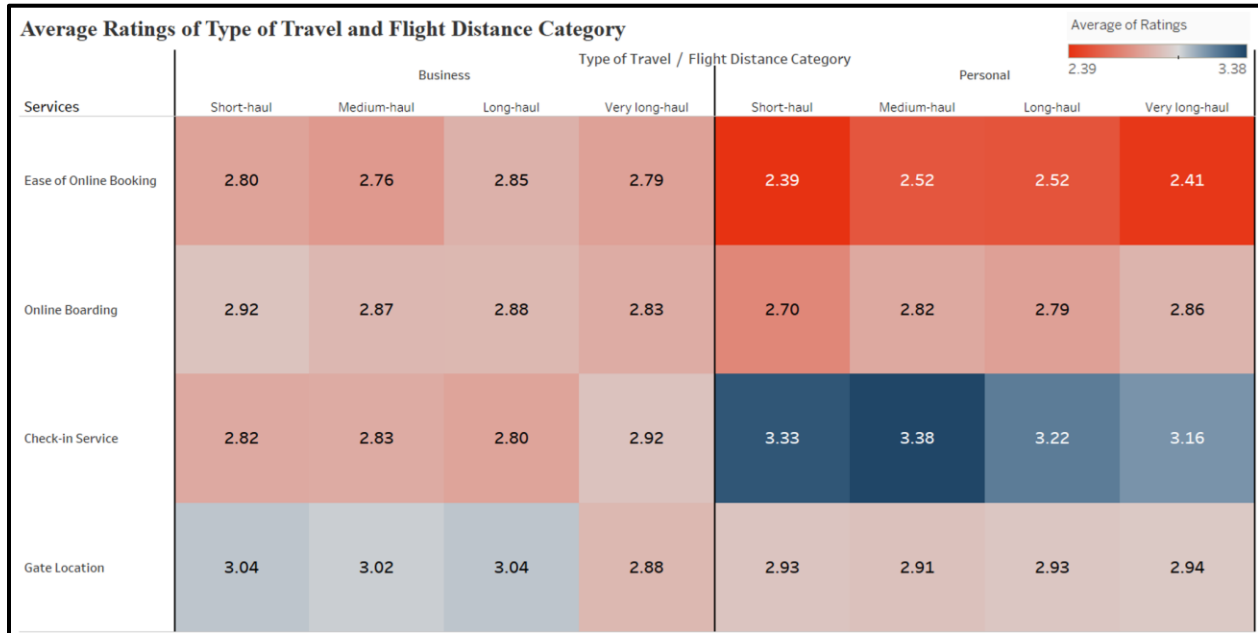## *Service Satisfaction by Age Group*
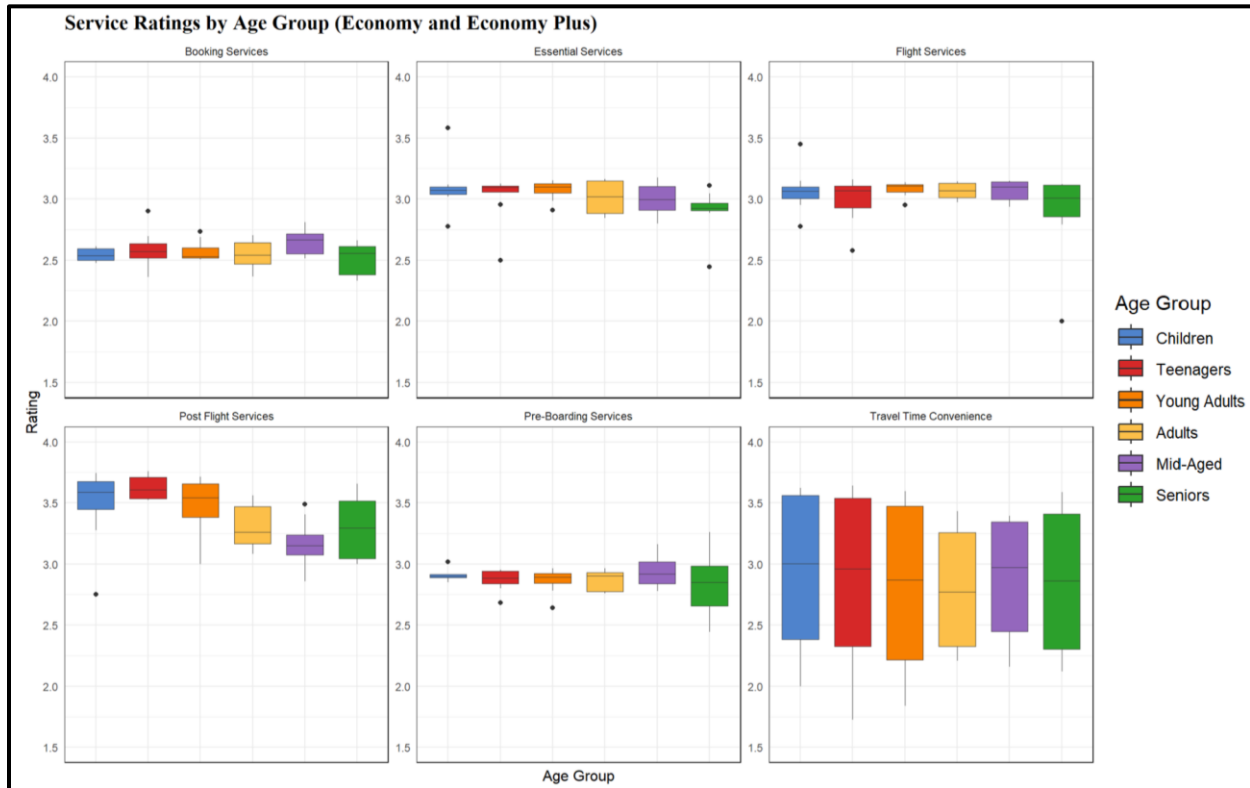
The above image is a series of box plots that display service ratings across different age groups within the Economy and Economy Plus classes. Each box plot represents a different service category: Booking Services, Essential Services, Flight Services, Post Flight Services, Pre-Boarding Services, and Travel Time Convenience.

This plot was created by first calculating the mode for each individual service rating and then computing the mean of these grouped service categories. This approach allowed us to visualize the overall distribution and compare service satisfaction across different age groups within Economy and Economy Plus classes. To enhance clarity and highlight key patterns, several refinements were made: distinct colors were used for each age group to clearly differentiate the service ratings, and outliers were shown to indicate variations and extremes in the data for each age group. Additionally, a common baseline was used for all service categories to facilitate better comparison.

We chose the box plot because it effectively displays the distribution and variability of service ratings across age groups, allowing for clear comparison of medians, quartiles, and outliers. This detailed view identifies specific service areas with significant satisfaction differences among age groups, making it ideal for highlighting passenger satisfaction in Economy and Economy Plus classes.

Service Needs Improvement

Services	Business				Personal			
	Short-haul	Medium-haul	Long-haul	Very long-haul	Short-haul	Medium-haul	Long-haul	Very long-haul
Ease of Online Booking	2.80	2.76	2.85	2.79	2.39	2.52	2.52	2.41
Online Boarding	2.92	2.87	2.88	2.83	2.70	2.82	2.79	2.86
Check-in Service	2.82	2.83	2.80	2.92	3.33	3.38	3.22	3.16
Gate Location	3.04	3.02	3.04	2.88	2.93	2.91	2.93	2.94

The heatmap visualizes average ratings for various services across travel types (Business, Personal) and flight distance categories (Short-haul, Medium-haul, Long-haul, Very long-haul). It covers services like Ease of Online Booking, Online Boarding, Check-in Service, and Gate Location. Created using Tableau, data preprocessing involved pivoting, aggregating ratings by service, and rounding values to two decimals for accuracy and readability. Key variables include the Type of Travel, Flight Distance Category, and Services.

This heatmap effectively displays average rating magnitudes across two dimensions through color gradients, allowing for quick comparisons and pattern identification. Initially, it included ratings for all passengers and services, but was refined by filtering out groups with lower ratings and focusing on Economy and Economy Plus classes based on the insights generated from the previous plots. Enhancements included a diverging color scheme, separating lines between travel types, and a selected color palette for readability and visual appeal.

Organized with services on the vertical axis and travel-distance combinations on the horizontal, the heatmap provides a detailed, easily interpretable overview of passenger satisfaction. It complements other visualizations, offering a richer, nuanced display of how travel type and distance influence service ratings, making it essential for identifying key patterns and insights. Compared to bar plots and tree maps, the heatmap excels in comparing multiple variables with fine detail, facilitating quick identification of patterns and enhancing the understanding of factors influencing passenger satisfaction.

Analysis and Discussion

Passengers Demographic

From the plot we can observe that the median age of first-time travelers across all flight distance types is below 30. Whereas, for returning customers, it is above 40. Furthermore, based on the interquartile range of the violin plots, it can be inferred that 50% (IQR) of the very long-haul flyers are in the age range 35-50, and more than half of first-time travelers are younger than 40. Based on these observations, the majority of first-time flyers are in the adult (25-39) age group and returning flyers are mostly middle-aged (40-59). So, airline marketing and branding to acquire new customers should target adult flyers. The airlines should focus upon retaining adult flyers.

Based on these insights, we can narrow down our analysis to the service ratings and satisfaction levels for the first-time adult flyers and middle-aged returning flyers. This will help us observing the service where the flyers have rated highly and services which have a below par rating, to identify the areas of improvement.

Passenger Satisfaction Level

Given six Age Groups, the above stacked bar charts display the number of passengers distributed across sub-groups. Noticeably, Adults and Mid Aged passengers (25-59 years old) represent the majority of the data, particularly Mid-Aged passengers flying Business Class with 32,000 passengers, indicating a key demographic for airlines. The proportional stacked bar charts below emphasize the satisfaction level distribution as percentages within each sub-group. The overall pattern shows that satisfaction levels tend to increase with age, showing that older age groups report higher satisfaction with their flight experiences.

A notable pattern is that the majority of passengers are more Neutral or Dissatisfied with their flight experience, suggesting that airlines need to review their services and make improvement plans. Indeed, across three classes, Business Class has more Satisfied feedback. This suggests that Business Class services meet the expectations of these passengers better. While Economy and Economy Plus show the same pattern, with slightly higher satisfaction for Economy Plus, which is reasonable considering passengers pay a slight premium for the upgrade. Given the information provided in this visual, further exploration of satisfaction level scores will be emphasized to identify services need improvement.

Service Satisfaction Across Airline Classes

The radar chart visualizes service ratings across Business, Economy Plus, and Economy classes, covering aspects like Booking Services, Travel Time Convenience, Post Flight Services, Essential Services, Flight Services, and Pre-Boarding Services. Business class consistently shows higher ratings across all categories, indicating greater passenger satisfaction, especially in Pre-Boarding and Post Flight Services. Economy Plus has moderate ratings, higher than Economy but lower than Business, with balanced satisfaction levels. Economy class shows the lowest ratings,

particularly in Essential Services and Flight Services, highlighting areas for improvement. Enhancing these specific areas in Economy could significantly elevate overall passenger satisfaction.

Based on this analysis, we will focus on improving Economy and Economy Plus classes, as Business class already has high satisfaction levels, to strategically enhance overall passenger satisfaction. In further analysis, we will investigate service rating satisfaction by different age groups within Economy and Economy Plus.

Service Satisfaction by Age Group

The box plot visualizes service ratings across different age groups within the Economy and Economy Plus classes for various service categories. Ratings for Booking Services are relatively low across all age groups, with adults and mid-aged passengers rating them slightly higher. Essential Services vary, with seniors rating them lower and others average. Flight Services are consistent, with adults rating higher and some teenagers and seniors lower. Post-Flight Services receive average to higher ratings from all age groups. Pre-Boarding Services have generally low ratings from all age groups except some of the mid-aged passengers who rated them slightly above average. Travel Time Convenience has relatively high and consistent ratings across all age groups.

Key insights include mid-aged passengers consistently rating most service categories higher, suggesting that services might be better tailored for mid-aged passengers. However, Booking Services and Pre-Boarding Services receive lower ratings across most age groups, indicating areas for improvement. Since ratings of 3 and above are considered average and good, we decided to focus on improving Booking Services and Pre-Boarding Services, as these were below 3 compared to other categories. By focusing on improving Booking Services and Pre-Boarding Services, we aim to enhance satisfaction across all age groups. In further analysis, we will focus on these two services in depth by examining the individual services within these categories from different angles.

Service Needs Improvement

The heatmap reveals several key insights regarding passenger ratings for various services across different travel types and flight distance categories. Notably, Ease of Online Booking is the service with the lowest average ratings across all services, particularly among personal travel passengers. The lowest average rating is observed in the Personal Short-haul category for Ease of Online Booking, with a rating of 2.39. This indicates a significant area for improvement in the online booking process for personal travelers.

Interestingly, while business passengers generally provide higher average ratings across most services, the Check-in Service stands out as an exception. Personal travelers, especially in the Short-haul category, rate the Check-in Service higher, with a notable rating of 3.33 compared to business travelers. This suggests that personal travelers find the check-in process more satisfactory, highlighting a potential strength in this service area.

Conclusion

Based on our analysis of the dataset, we found that:

- Adults and middle-aged passengers are a key demographic for the airlines. Adults make a significant fraction of first-time passengers. There is a positive trend between the increasing age of passengers and corresponding satisfaction levels.
- Business flying passengers are the most satisfied seating class. Whereas feedback from economy and economy plus passengers shows that customer expectations are not being met.
- Services such as booking and pre-boarding services need significant improvement.

The objective of this analysis is to identify areas needing improvement and provide recommendations to enhance passenger satisfaction. Keeping that mind we provide the following recommendations for improving customer satisfaction.

- Loyalty programs geared towards the adult (20-39) age group to incentivize them to fly with the airline more.
- For Economy and Economy Plus classes, develop an improvement plan for Booking and Pre-Boarding services. Specifically, convenience in booking online. Working with third-party ticket vendor or other airlines.
- Improve Check-in service for Business class.

Furthermore, it is recommended to enhance the online booking experience for personal travelers by simplifying the process, providing a more user-friendly interface, and offering better support and information. Additionally, the high ratings for Check-in Service among personal travelers should be analyzed to understand what aspects are particularly effective, with the aim of extending these best practices to other services and travel segments to improve overall satisfaction. These targeted improvements can help address the specific needs of different traveler segments and elevate the overall travel experience.

Future Scope

Given that the project timeline is half an academic quarter, the visualization and analysis of the dataset developed for the project could be more comprehensive. If there was more time, it would have been used for more analysis into trends and patterns in the data and further introspection over the visualizations created.

Furthermore, we would be interested into finding datasets of airline customer satisfaction but more with variables such as flight number, flight origin and destination and seating configuration of flights i.e. Percentage of business, economy and economy plus seats in a flight.

Appendix – Individual Report

Uday Kiran Galla Srinivasalu

Role and Contributions

I was very fortunate to work with such a talented and collaborative team, and I truly enjoyed every moment of our partnership. Meanwhile I played a crucial role in our group, driving the project from inception to completion. As the coordinator, I created a group on WhatsApp for daily communication and set up a Zoom meeting every Wednesday at 3 PM for updates and planning, effectively acting as a scrum master. I was responsible for setting deadlines and ensuring tasks were completed on time. I collaborated with my team members to finalize the dataset, where we each selected 1 or 2 datasets, pitched our ideas, and conducted a voting process to make the final selection.

Before starting the Exploratory Data Analysis, I took the initiative to clean the data by managing nulls and NAs and performed a summary and statistical analysis to understand the entire dataset. In Milestone 2, I conducted extensive EDA on almost all variables, analyzing distributions such as Passenger Age by a bar graph, Passengers by Class and Gender, and a Treemap for passengers by customer type and age group. I explored all possible directions to extract meaningful insights from the data. In Milestone 3, I collaborated with Chandana to create complex radar plots, examining multiple groups like Ratings by Age and Ratings by Class. I plotted radar charts for various variables to understand different patterns from multiple perspectives. Finally, I decided to focus on Service Ratings by class and worked on grouping the age categories and services.

Throughout this process, I played a pivotal role in connecting all the dots, deciding the flow and structure of our story. I guided my team to form a coherent narrative, starting from overall distribution and then focusing on Airline class, narrowing it down to finer details. I provided constructive feedback on each of my teammates' plots and ideas, respected their work, and learned different techniques from them.

Reflection Summary

This project taught me valuable lessons in team management, technical skills, and storytelling. Coordinating the group and setting up effective communication channels was crucial for our success. By setting up a WhatsApp group and Zoom meetings, I ensured everyone stayed on track and met deadlines. Acting as a scrum master, setting deadlines, and ensuring task completion improved my project management skills. I learned the importance of collaboration and constructive feedback in refining our work and achieving our goals. My feedback on the radar plot significantly enhanced its clarity and effectiveness. Working closely with my teammates, I gained insights into different analytical techniques and data visualization tools, such as using R for data cleaning and creating complex visuals with new packages, which enhanced my technical expertise.

The process of cleaning the data and conducting comprehensive EDA allowed me to develop a deeper understanding of the dataset and the importance of thorough data preparation. Creating complex visuals like radar plots and analyzing service ratings by class helped me appreciate the power of detailed, multidimensional analysis in uncovering valuable insights. Additionally, I learned the significance of organizing information and structuring a compelling narrative to present our findings effectively.

The project has not only improved my technical skills but also enhanced my ability to lead a team and manage projects efficiently. The iterative process of creating and refining visualizations based on feedback emphasized the importance of continuous improvement and adaptability. Overall, this experience has prepared me to become a better data analyst by equipping me with the skills needed for effective data analysis, visualization, and storytelling.

Chandana Bhat Thambrahalli

Role and Contributions

Throughout our "Airlines Passenger Satisfaction" project, I've played a key role in shaping our understanding of the dataset and driving our analysis towards actionable insights. My proactive engagement and strong analytical skills have been fundamental to the quality of our work.

In the early stages post-midterm, I was dedicated to refining the dataset alongside my team members, laying a solid foundation for our analysis. I didn't settle for surface-level analysis; instead, I delved deep into the dataset uncovering patterns and trends creating exploratory visuals paved the way for deeper investigation.

In Milestone 2, I took the lead in creating a correlation matrix using 18 numerical columns, demonstrating my commitment to understanding the complex relationships within the data. This analytical approach provided clarity on dependencies between columns and facilitated more nuanced insights in subsequent stages.

Collaborating with Uday in Milestone 3, I brought creativity and strategic thinking to the table. I proposed the creation of radar plots, grouping numerical columns into service categories based on satisfaction levels (1-5). This innovative approach not only added depth to our analysis but also enhanced the visual appeal of our deliverables.

But I didn't stop there. I looked beyond the data, envisioning tangible strategies that airlines could implement to boost passenger satisfaction. My recommendations weren't just theoretical — they were pragmatic solutions grounded in real-world insights. By bridging the gap between analysis and action.

In summary, my contributions have been instrumental in shaping the trajectory of our project. From data refinement to insightful analysis and actionable recommendations, my proactive approach, analytical prowess, and strategic thinking have greatly enriched our collaboration and contributed to the success of our endeavor.

During this course, my main goal was to become better at telling stories with data and to learn how to use Tableau effectively. Working on this project with my team taught me a lot about the importance of sticking to deadlines, working together, and thinking creatively to solve problems.

We made sure to meet regularly and share the workload evenly. Giving each other feedback helped us improve our work and make sure it was the best it could be. We even practiced our presentation to make sure we got our message across clearly to our audience.

One challenge we faced was creating a radar plot, but working with my teammate Uday taught me new ways to approach problems and analyze data. Even though we only worked together for a short time, I learned a lot about how different people work and think. We respected each other's ideas and made decisions together.

In the future, I want to carry forward the lessons I learned from this project, especially the importance of teamwork and considering different perspectives. I believe that good communication, respect, and working together toward a common goal are key to success in any project.

Overall, this project has not only helped me improve my technical skills but has also shown me the power of collaboration and continuous learning. I'm excited to apply these lessons in my future work and keep growing as a data analyst and team player.

Naveen Kumar Reddy Veeramreddy

Role and Contributions

I played an important role in all phases of the project as a member of our data visualization project team, contributing significantly to both the exploratory and explanatory analyses. I was involved in developing and refining various visualizations to uncover and present key insights from the data. Throughout the project, I created several draft visuals, including box plots, scatter plots, mosaic plots, and parallel coordinates plots, to explore the data comprehensively.

In the exploratory phase, I plotted a panel box plot to identify the distribution and outliers within the dataset, which helped us understand the spread of the data across different variables. I also explored various combinations such as the number of passengers satisfied in different classes, correlations between departure and arrival delays, and rating variations across different members using parallel coordinates plots. These visualizations were instrumental in uncovering important trends and relationships within the data, allowing us to form a coherent story and extract meaningful insights.

For the explanatory phase, I designed a heatmap to visualize average ratings for various services across different travel types (Business, Personal) and flight distance categories (Short-haul, Medium-haul, Long-haul, Very long-haul). This heatmap highlighted key patterns and

differences in passenger satisfaction, providing insights that enabled us to conclude with actionable recommendations.

The panel box plot and the heatmap were particularly impactful in our analysis. The panel box plot allowed us to see the distribution and identify outliers, while the heatmap provided an intuitive visualization of average ratings, making it easier to identify areas of improvement and strengths. These two visualizations, combined with the insights from other exploratory visualizations, ensured that our final narrative was supported by solid data and compelling visual evidence.

Reflection Summary

Through this project, I have gained a deeper understanding of the power and intricacies of data visualization. One of the most significant lessons learned is the importance of choosing the right type of visualization for the data and the message we want to convey. For instance, while scatter plots and box plots are excellent for exploratory analysis, heatmaps provide a more intuitive and comprehensive view for explanatory purposes.

I particularly learned about the complexities of plotting multidimensional data and the significance of using appropriate color schemes to enhance interpretability. Employing small multiples allowed us to show all variations in the data, making it easier to compare different subsets. Using appropriate encoding techniques ensured that the visualizations were not only informative but also accessible and easy to understand.

The process of using advanced visualization techniques to tell the story of the data was eye-opening. Techniques such as parallel coordinates plots enabled us to display multivariate data effectively, revealing hidden patterns and correlations. Additionally, the iterative process of creating multiple draft visuals and refining them based on feedback emphasized the importance of continuous improvement in data visualization.

Moreover, collaborating with the team taught me the importance of combining different visual perspectives to form a complete and nuanced understanding of the data. Each visualization type has its strengths and can reveal different facets of the data. By integrating various visualizations, we were able to tell a more compelling and comprehensive story.

In summary, this project has enriched my skills in data visualization, reinforcing the need for careful selection of visualization techniques, iterative refinement, and collaborative storytelling. These experiences, especially in plotting multidimensional data and using advanced visualization techniques, will undoubtedly benefit my future work in data analysis and visualization.

Mai Phuong Tran Ngo

Role and Contributions

I believe every person in our group plays an important role in this project and made great contributions. After the dataset was finalized, I created a document with the distribution of each

variable at a basic elementary level so that our team could understand the data and use it as supplementary documentation for milestone 2. Diving into milestone 3, I looked further into the data attributes and wanted to create calculations from the original dataset, rather than using the original data itself. So, I mapped and applied the calculation logic for homework 4 since we used the same dataset. Once I successfully achieved the desired plot, I discussed with my teammates my choice of variables and moved forward from there. Additionally, I tend to pay very high attention to detail in my work, so every step I took, both data and visuals, were meticulously decided and high technique so that I could produce good content without needing to redo my work. Note that every week, we all met up twice and discussed each other's progress to get feedback and approval from others.

Additionally, I created the PowerPoint template from scratch and focused on the aesthetic aspects of the deliveries. Ensuring everything looks good is essential since the course focuses on visualization for non-technical audiences. I want to ensure everything we deliver, from a simple text paragraph to a visual, is on point and harmonized as a whole group, including detailed elements such as fonts, colors, and the resolution of the visuals. Overall, I was in charge of editing content for all of our submissions, making sure everything was consistent and high quality. As for the project content and story, I gathered the initial data story with the objective of narrowing down the dataset for finer details. Nonetheless, every person in our group had equal contributions. We had meetings that took five hours to create a concrete story message, listening to each other's ideas and suggestions.

Reflection Summary

Throughout this project, I truly loved working with everyone. We all have the same mindset and are very good at communication. There was no hiccup in our discussions at all. We started the project right after the midterms, so everything was paced out and planned properly due to ample time. Given the fact that we were meeting twice a week, it was super easy for everyone throughout the entire journey. Furthermore, the WhatsApp group played an important role. We could discuss any new ideas or approaches for the project, and everybody was super responsive and highly responsible.

Overall, I learned a lot of new things not just from this project but also from homework 4. I created new data aside from the original data, generated high-tech visuals like mosaic plots, and learned more about how Tableau works. Prior to this class, I already had good experience using R, so for this project, I did everything using Tableau to learn more about the software, including how to edit and manipulate data through Tableau. I am very satisfied with the visual outputs and the experience I gained working on this project. Furthermore, I have always paid high attention to detail, but after taking this class and completing this project, I am even more mindful when creating data visualizations, considering the targeted audience and the story message.

As I mentioned in the first week's discussion, my objective for this course was to learn how to create visuals for a non-technical audience - someone who knows nothing about the data yet can still understand the story. I certainly have achieved that goal.

Guntaka Satish Harshavardhan Reddy

Role and Contributions

As the group liaison, I was tasked with making milestone submissions on behalf of the group. Working on this project along with the group has been an immense pleasure. Constant constructive feedback was shared with each member. In order to maintain uniformity across all the visualizations, we decided to use a common color palette.

The visualization that I created, and which ended up the final report is split violin plot. I had made this kind of visualization for a homework. Given the dataset and the context, I decided to use the split violin plot. The visualization went under a lot of rounds of drafts with constant feedback from the group members on every iteration.

Other contributions include helping with R code, providing critique on visualizations and data analysis. I also contributed to collating milestone and final reports. Specifically, analysis and discussion and the conclusion and recommendations parts in the presentation and the final report.

Reflection Summary

Although the focus was on data visualization, I learnt a great deal about data analysis as well in this project. This was my first foray in the R programming language. All my visualizations have been created in R. Though most of the visualizations didn't make out of the drafts, I am very satisfied with the final plots included. This project and class helped with honing my skills to interpret data through visualizations and critically analyze a visualization on the aspects of aesthetics, message and context.

As a group, we were focused on cohesion and uniformity since the beginning of the project. To that end, we have used a common color palette for visualizations and presentations, tried interlinking the visualizations. This helped in creating a clear narrative in our visualizations with relative ease.