The influence of the length of airplane flights on the perceived satisfaction among airplane passengers

Part 1: Introduction

In the current period marked by swift evolution in the airline industry, propelled by technological advancements and shifts in consumer preferences, grasping the determinants of passenger satisfaction is essential. Air travel, which was once considered a luxury, has transformed into a widely used mode of transport, presenting the challenge of upholding high standards of customer satisfaction. The airline sector is faced with a distinct array of challenges, encompassing aspects like operational efficiency, safety, and the enhancement of the passenger experience. Moreover, similar to other industries, it is susceptible to the fluctuations of the economy and has been impacted by downturns, a notable recent example being the Covid pandemic. In such a dynamic setting, the Airline Passenger Satisfaction (APS) Survey, utilizing the Maven Analytics Dataset, serves as a pivotal metric, indicating the extent to which airlines meet or surpass passenger expectations in various service facets. It is noteworthy, however, that while this dataset is comprehensive, it omits specific details about airlines or flights, which complicates conducting more detailed analyses of the study's results in subsequent sections.

The research question I am exploring is whether the duration of airplane flights has an impact on passengers' perceived satisfaction. Certain control variables will also be included in the analysis, these will be detailed later in the text. The outcomes of this study are expected to provide insights into the airline industry's efforts to enhance passenger satisfaction, contributing to strategies that cater to diverse passenger needs and preferences in the ever evolving, and sometimes hectic landscape of air travel.

Part 2: Data

For the analysis conducted in this study, the dataset was sourced from the Maven Analytics Data Playground, specifically focusing on airline passenger satisfaction. The dataset, initially comprising 129,880 observations, represents individual airline passengers' satisfaction records. Each observation includes various attributes related to the passenger's flight experience. After preprocessing, including handling missing values and creating necessary variables, the dataset was streamlined for the regression analysis. The pre-processing also included drastically trimming down the number of values to comply with the exercise, a method of true random sampling (or as close as could be achieved in python at my skill level) was used to get 5000 values to be used for the analysis.

The primary dependent variable for this study is airline passenger satisfaction, represented as a binary outcome in the 'satisfied_customer' variable, where '1' indicates satisfaction and '0' represents neutral or dissatisfied responses. This binary format simplifies the interpretation of the logistic regression analysis. The Explanatory Variable is the Flight Distance, the primary factor of interest, measuring the length of the flight in miles. Control variables in the study are systematically categorized into distinct groups to facilitate a structured analysis. Under the category of Demographic Indicators, we consider 'Age', which refers to the age of the passenger, and 'Female', a variable indicating whether the passenger is a woman. The analysis also includes Travel Type/Class Indicators. These are 'Class_Economy', a dummy variable to denote if the passenger traveled in Economy class, 'Class_Economy Plus', signifying travel in Economy Plus class, and 'Type of Travel_Personal', distinguishing whether the travel was for personal reasons. In the realm of In-Flight Service Indicators, various aspects are evaluated. These include 'In-flight Service', assessing the overall in-flight service quality, 'Seat Comfort', evaluating the comfort of the seating, 'Leg Room Service', rating the adequacy of legroom provided, and 'Food and Drink', measuring the quality of food and drink services offered during the flight. Lastly, Delay Indicators are considered, encompassing 'Departure Delay' and 'Arrival Delay', which respectively track delays in departure and arrival, providing a comprehensive view of punctuality and operational efficiency in the context of the overall travel experience .

Each of these variables was carefully chosen and prepared to capture different aspects of the airline travel experience that could influence passenger satisfaction. The transformation of categorical variables into dummy variables and the treatment of continuous variables (like flight distance and age) were key steps in the feature engineering process. This structured approach to categorizing the variables facilitates a more nuanced understanding of how various factors, from flight length to in-flight amenities, impact overall passenger satisfaction in the airline industry.

Part 3: Regression

In this segment of our research, our objective is to authenticate and measure the influence of each category of variables outlined in "Part 1" on airline passenger contentment. The Exploratory Data Analysis (EDA) phase unveiled several crucial findings, notably the pronounced right-skewed distribution of flight length. To counteract this skewness and normalize the data, thereby facilitating more efficacious modelling, we shall implement a logarithmic transformation to the flight length variable. Our methodology entails the application of simple linear regression models to scrutinize the correlations between the transformed flight length and other explanatory variables, and their impact on passenger satisfaction. We will employ heteroscedasticity-consistent standard error estimators, specifically opting for the HC1 covariance type. This selection is strategically made to yield more reliable standard errors in the presence of heteroscedasticity, thereby augmenting the validity and interpretability of our regression outcomes.

3.2 Regression models

3.2.1 Model 1: X and Y only, no control variables

The regression analysis for model 1, examining the impact of log-transformed flight distance on customer satisfaction in the airline industry reveals a significant positive relationship, with a coefficient of 0.1405. This indicates that an increase in the log of flight distance is associated with a rise in customer satisfaction. However, the model's modest explanatory power, indicated by an R-squared value of 0.067, suggests that flight distance alone does not fully capture the nuances of customer satisfaction. This finding highlights the necessity of including additional variables in the analysis to better understand the complex factors influencing customer satisfaction in the airline sector.

3.2.2 Model 2: Demographic Indicators

The regression analysis for Model 2, exploring the influence of log-transformed flight distance, age, and gender on customer satisfaction in the airline industry, demonstrates a nuanced relationship among these factors. The model shows a significant positive correlation between log-transformed flight distance and customer satisfaction, with a coefficient of 0.1356, implying that longer flights are likely to yield higher passenger satisfaction. Additionally, age has a positive effect (coefficient: 0.0042), suggesting that satisfaction slightly increases with age. The gender variable ('Female'), with a coefficient of -0.0125, does not show a significant impact on satisfaction. Despite the model's R-squared value of 0.083, indicating that about 8.3% of the variance in customer satisfaction is explained, the modest explanatory power highlights the complexity of the factors affecting satisfaction. This suggests the need to consider more variables to gain a comprehensive understanding of customer satisfaction dynamics in the airline sector.

3.2.3 Model 3: Travel Type/Class Indicators

The regression analysis for Model 3, examining various factors affecting customer satisfaction in the airline industry, reveals a more comprehensive understanding of the determinants of passenger experience. The model, with an R-squared value of 0.320, indicates that approximately 32% of the variance in customer satisfaction is explained by the variables included. A notable finding is the positive influence of the log-transformed flight distance on satisfaction, evidenced by a coefficient of 0.0397. This suggests that an increase in flight distance is associated with a modest increase in the probability of a passenger being satisfied. The model also reveals significant impacts of travel class on satisfaction. Passengers in Economy and Economy Plus classes are less likely to be satisfied compared to those in higher classes, as indicated by negative coefficients of -0.3466 and -0.3151, respectively. Additionally, the type of travel plays a crucial role, with personal travel negatively impacting satisfaction (coefficient: -0.2785). These findings underscore the multifaceted nature of customer satisfaction in air travel, influenced not only by the journey's length but also by the class of travel and the purpose of the trip.

3.2.4 Model 4: In-Flight Service Indicators

The regression analysis for Model 4, examining the impact of log-transformed flight distance and various in-flight service factors on customer satisfaction in the airline industry, reveals significant relationships between these variables and passenger satisfaction. The model indicates that log-transformed flight distance is positively correlated with customer satisfaction, with a coefficient of 0.0958. This suggests that longer flights tend to be associated with higher levels of satisfaction. In-flight service factors also demonstrate substantial influence. Specifically, the coefficients for in-flight service (0.0604), seat comfort (0.0993), and leg room service (0.0826) all indicate positive effects on satisfaction, highlighting the importance of these aspects in the overall passenger experience. The food and drink variable, while also positive, shows a smaller effect (coefficient: 0.0100) and is less statistically significant. With an R-squared value of 0.249, the model explains about 24.9% of the variance in customer satisfaction. While this points to a moderate explanatory power, it also underscores the multifaceted nature of customer satisfaction and the potential for other unexplored variables to contribute to the model.

3.2.5 Model 5: Delay Indicators

The regression analysis for Model 1, reveals a significant positive correlation between the log-transformed flight distance and customer satisfaction, with a coefficient of 0.1400, indicating that increased flight distance is likely associated with higher passenger satisfaction. Additionally, the model examines the effects of both departure and arrival delays. The coefficient for departure delay (0.0014) suggests a slight positive impact on satisfaction, although this effect is marginally significant. Conversely, arrival delay demonstrates a negative effect on customer satisfaction, with a coefficient of -0.0020, indicating that longer arrival delays tend to decrease satisfaction. With an R-squared value of 0.070, the model explains approximately 7% of the variance in customer satisfaction. This modest explanatory power implies that while flight distance and delays are relevant, they capture only a small portion of the factors affecting satisfaction.

Part 4: Generalization and External Validity

In our statistical exploration of airline passenger satisfaction, we applied predictive models to the variables in our Ordinary Least Squares (OLS) models, aiming to validate their associations with customer satisfaction levels. These models were instrumental in establishing the relationship between log-transformed flight distance and customer satisfaction, confirming the foundational hypothesis of Model 1. Further, in Models 2 to 5, we examined the interaction between satisfaction and additional variables such as age, gender, travel class, in-flight service quality, and travel delays. This comprehensive approach illuminated the complex dynamics at play: for instance, the predictive models suggested that certain demographics, service quality, and travel class significantly impact satisfaction. In examining the R-squared and adjusted R-squared values across the models, we observed variations in explanatory power, reflecting the multifaceted nature of factors affecting passenger satisfaction. Models incorporating a broader range of variables, particularly those related to in-flight experience and operational efficiency, showed enhanced predictive accuracy. These findings suggest that passenger satisfaction is not merely a function of travel duration but is also intricately linked to service quality, demographic factors, and travel conditions. The external validity of these models, reinforced through robustness checks, affirms their applicability across various airline industry contexts. This analysis underscores the importance of a holistic approach in evaluating passenger satisfaction, offering valuable insights for strategic decision-making aimed at optimizing the passenger experience.

Part 5: Causal interpretation / main summary

In our comprehensive exploration of airline passenger contentment, we meticulously examined the intricate interplay of a myriad of factors, encompassing the log-transformed distance of flights, age, gender, the various classes of travel, and the caliber of in-flight services. Our investigation, particularly focused on Model 3, unearthed a significant correlation between the level of satisfaction expressed by passengers and their class of travel. This finding suggests that passengers situated in superior travel classes, such as Economy Plus, tend to report heightened levels of satisfaction. Additionally, Model 4 shed light on the pivotal role played by in-flight service elements, with variables such as seat comfort and the quality of food demonstrating a pronounced influence on the satisfaction levels of passengers.

Nonetheless, it is imperative to acknowledge the limitations in the causal interpretation of these relationships. While our regression models do suggest robust associations—for instance, an extended flight distance (as elucidated in Model 1) exhibiting a mild correlation with heightened satisfaction, and operational elements like travel delays (as depicted in Model 5) adversely affecting passenger experiences. These conclusions are predominantly correlative rather than causative in nature. Given that our models are predicated on observational data, they are adept at revealing patterns and associations, but they fall short in providing a robust degree of generalizability, as is evidenced by the relatively low R values. This limitation may be attributable to the potential exclusion of other pertinent variables not encompassed within the scope of this study. Thus, while our research offers valuable insights into the plethora of factors that influence passenger satisfaction, it is crucial to interpret these findings as indicative trends rather than definitive causes of satisfaction. To unravel these relationships with greater precision and establish causal connections, further research is warranted. This subsequent inquiry could potentially adopt experimental designs, which would facilitate a more in-depth exploration and understanding of these intricate dynamics.

Part 6: Findings and Conclusion

This exhaustive investigation into airline passenger satisfaction yields essential insights, pivotal for guiding both strategic business initiatives and policy formulations within the airline industry. The results of this study cast light on critical aspects that airlines might target to augment passenger satisfaction. The analysis intriguingly revealed a subtle yet impactful influence of flight distance on satisfaction. This highlights an opportunity for airlines operating longer flights to boost passenger experience, potentially through augmented services or amenities specifically designed for prolonged travel durations. Foremost, the significance of travel class in shaping satisfaction levels is profound. Passengers in premium classes, such as Economy Plus, consistently reported elevated satisfaction levels. This underscores a pronounced preference for enhanced comfort and additional amenities. Airlines might contemplate enhancements in service quality within economy classes or introducing more enticing features in premium classes to attract a wider spectrum of passengers. The quality of in-flight services, encompassing aspects like seat comfort and food quality, is also instrumental in crafting passenger experiences. Investment in elevating these facets is likely to lead to a notable increase in customer satisfaction. Such improvements could include advanced training for cabin crew, seating upgrades, or refining the quality of meals provided. Operational efficiency, especially in the context of travel delays, is a crucial determinant. Delays were found to have a negative correlation with satisfaction, underscoring the imperative for airlines to focus on timely departures and arrivals. The adoption of efficient scheduling and the formulation of effective contingency strategies to curtail delays could become a central focus.

Nonetheless, it is essential to acknowledge the constraints of our analysis in deducing causality. While the associations identified are robust, they ought to be perceived as precursors for further research rather than conclusive evidence. In summary, our study offers valuable, actionable insights for airlines to refine their service offerings, boost operational efficiency, and ultimately enhance customer satisfaction. Such enhancements, albeit necessitating investment, can lead to increased customer loyalty, favourable referrals, and consequently, heightened profitability in the competitive realm of the airline industry. Moreover, these insights can be instrumental for policymakers in developing regulations and standards aimed at elevating the passenger experience industry wide.

Part 7: Reference(s)

Maven Analytics. (05/18/2022). Airline Passenger Satisfaction. https://mavenanalytics.io/data-playground?search=Airline%20Passenger&tags=3cgwtPRmxwnqScuhpPMCoF

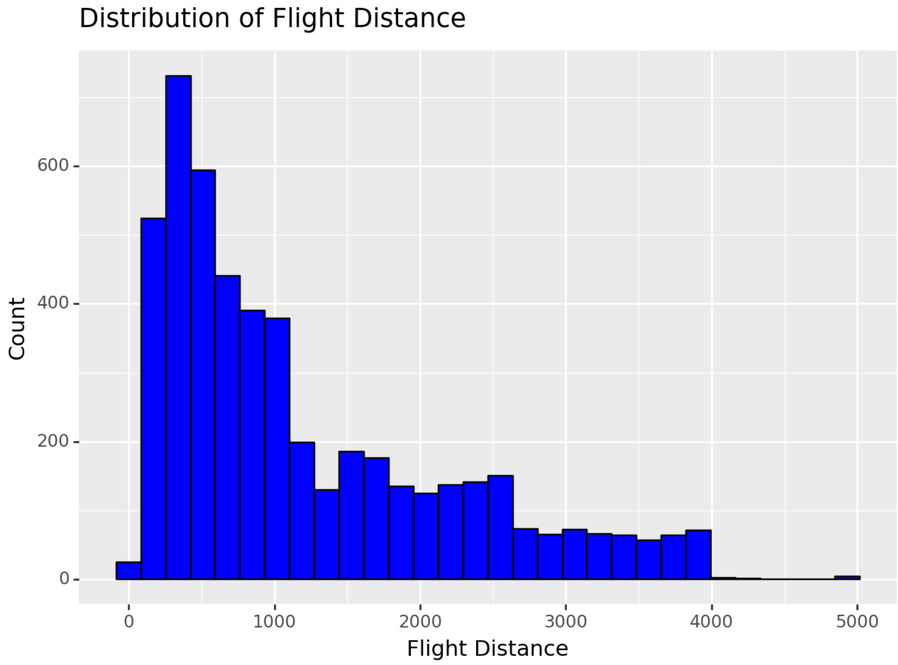
Appendix

1. Summary of Statistics

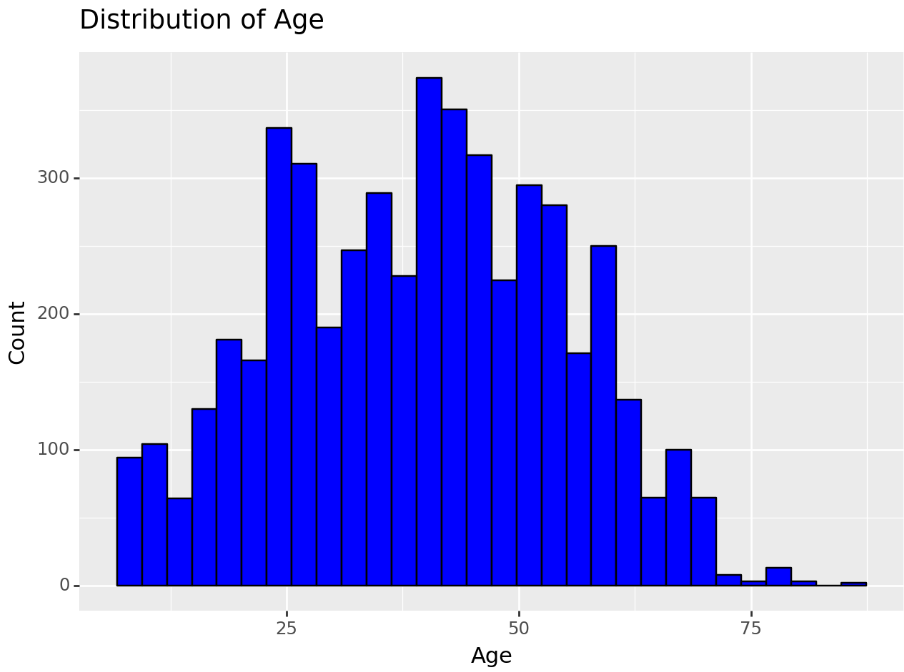
	count	mean	std	min	25%	50%	75%	max
satisfied_customer	5000.0	0.4404	0.496485	0.0	0.0	0.0	1.0	1.0
Flight_Distance	5000.0	1197.0888	992.846421	31.0	419.0	853.0	1750.0	4963.0
Age	5000.0	39.3752	15.204942	7.0	27.0	40.0	51.0	85.0
Female	5000.0	0.5094	0.499962	0.0	0.0	1.0	1.0	1.0
Class_Economy	5000.0	0.4464	0.497168	0.0	0.0	0.0	1.0	1.0
Class_Economy_Plus	5000.0	0.0720	0.258514	0.0	0.0	0.0	0.0	1.0
Type_of_Travel_Personal	5000.0	0.3118	0.463275	0.0	0.0	0.0	1.0	1.0
In_flight_Service	5000.0	3.6486	1.179575	1.0	3.0	4.0	5.0	5.0
Seat_Comfort	5000.0	3.4458	1.328689	1.0	2.0	4.0	5.0	5.0
Leg_Room_Service	5000.0	3.3708	1.317594	0.0	2.0	4.0	4.0	5.0
Food_and_Drink	5000.0	3.1782	1.328607	0.0	2.0	3.0	4.0	5.0
Departure_Delay	5000.0	14.6550	40.551987	0.0	0.0	0.0	12.0	1305.0
Arrival_Delay	5000.0	14.8924	40.607918	0.0	0.0	0.0	12.0	1280.0

2. Distribution Plots

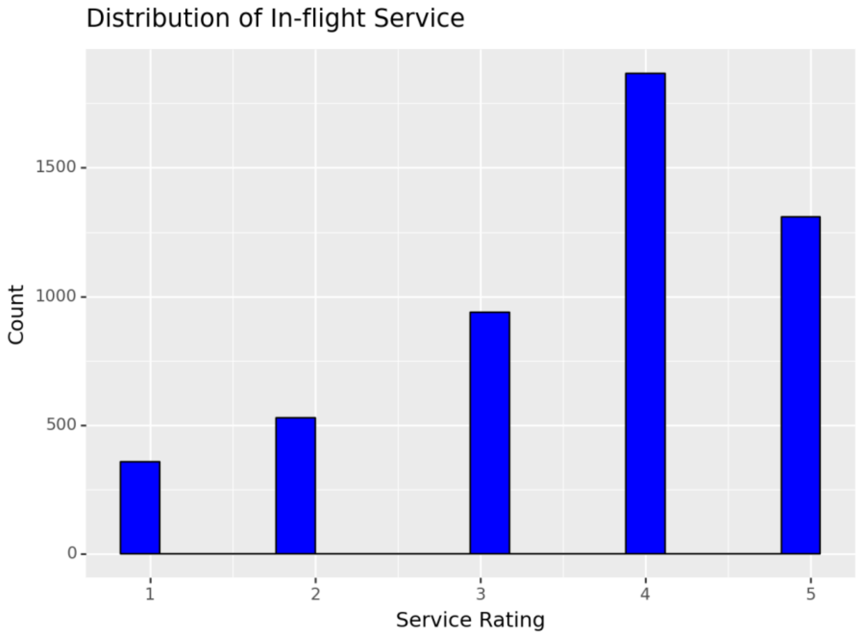
2.1 Distribution of explanatory variable (X)



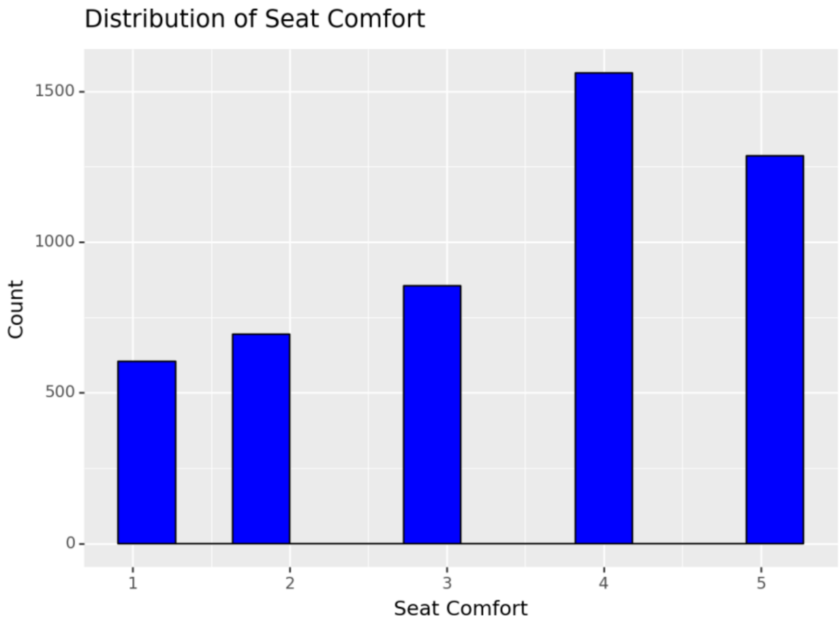
2.2 Distribution of dependent variable "Age"



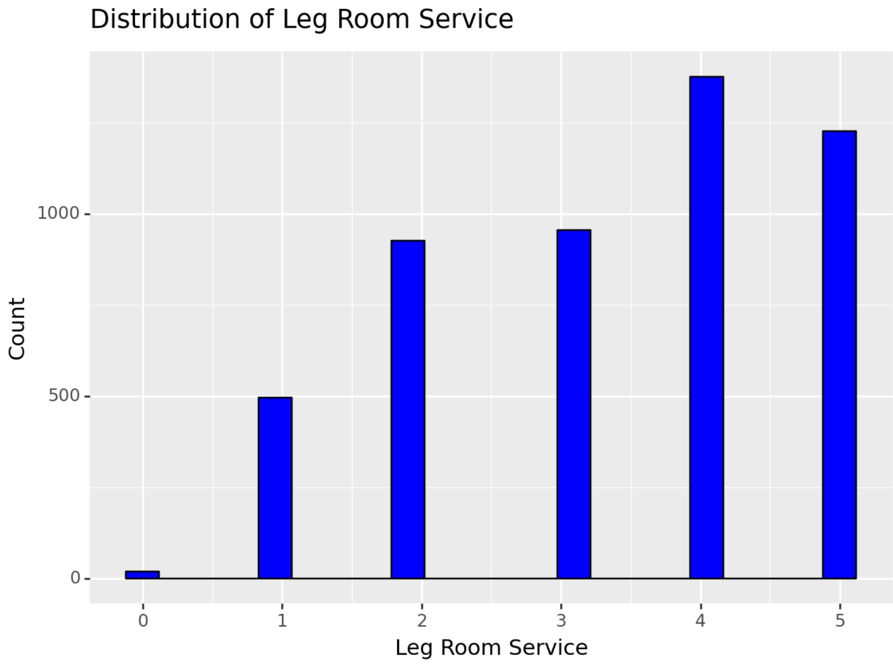
2.3 Distribution of dependent variable "In-flight Service"



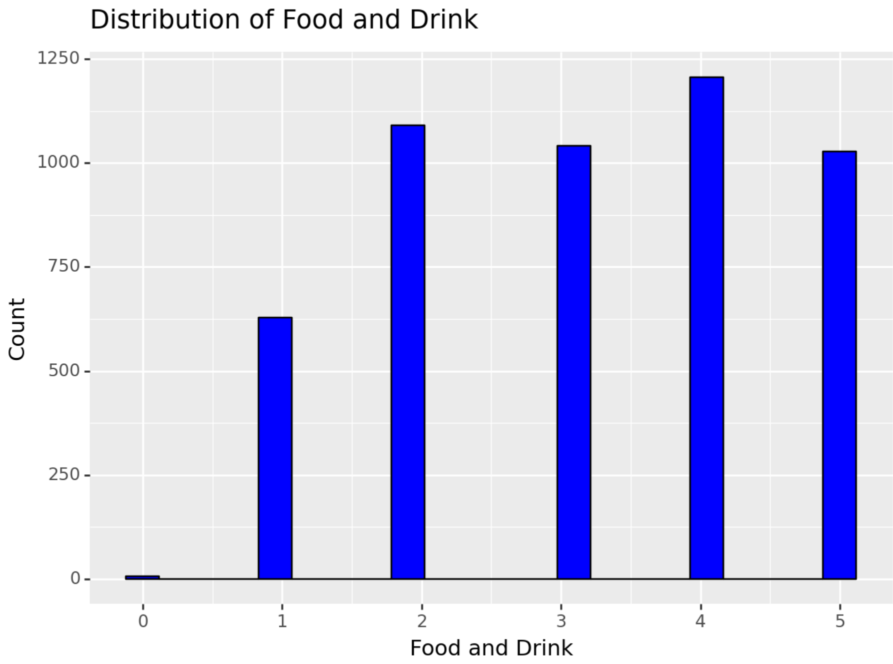
2.4 Distribution of dependent variable "Seat_Comfort"



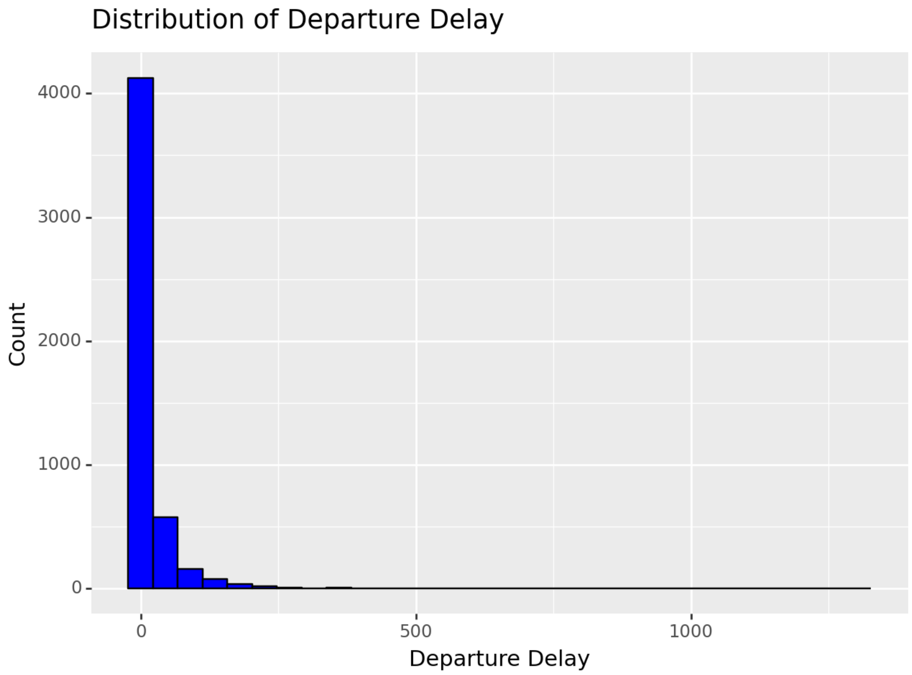
2.5 Distribution of dependent variable "Leg_Room_Service"



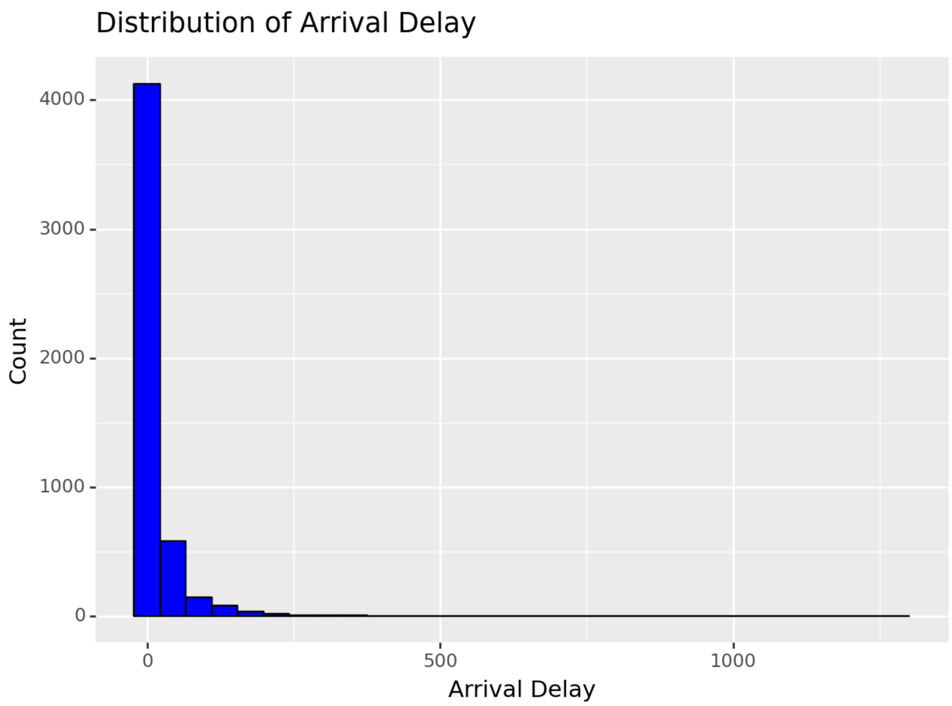
2.6 Distribution of dependent variable "Food_and_Drink"



2.7 Distribution of dependent variable "Departure_Delay"



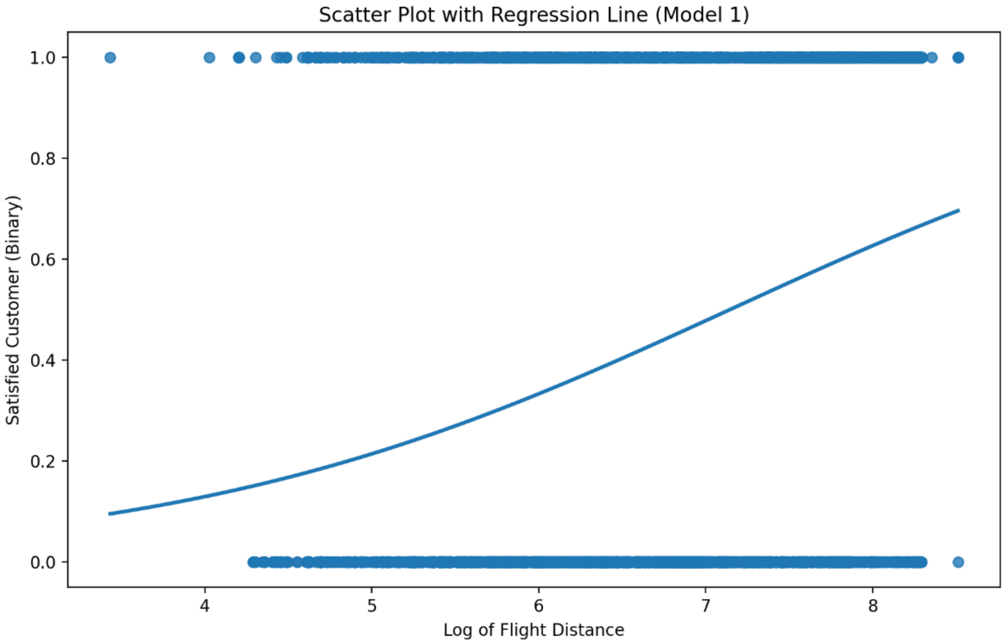
2.8 Distribution of dependent variable "Arrival_Delay"



3. Regression Table (Stargazer Summary)

	Dependent variable: satisfied_customer				
	(1)	(2)	(3)	(4)	(5)
Intercept	-0.503***	-0.629***	0.438***	-1.076***	-0.491***
	(0.050)	(0.051)	(0.053)	(0.047)	(0.051)
Log_Flight_Distance	0.140***	0.136***	0.040***	0.096***	0.140***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Age		0.004***			
		(0.000)			
Female		-0.013			
		(0.013)			
Class_Economy			-0.347***		
			(0.017)		
Class_Economy_Plus			-0.315***		
			(0.026)		
Type_of_Travel_Personal			-0.279***		
			(0.015)		
In_flight_Service				0.060***	
				(0.006)	
Seat_Comfort				0.099***	
				(0.006)	
Leg_Room_Service				0.083***	
				(0.005)	
Food_and_Drink				0.010*	
				(0.006)	
Departure_Delay					0.001*
					(0.001)
Arrival_Delay					-0.002***
					(0.001)
Observations	5000	5000	5000	5000	5000
R^2	0.067	0.083	0.320	0.249	0.070
Adjusted R^2	0.066	0.082	0.319	0.248	0.069
Residual Std. Error	0.480 (df=4998)	0.476 (df=4996)	0.410 (df=4995)	0.430 (df=4994)	0.479 (df=4996)
F Statistic	357.871*** (df=1; 4998)	161.230*** (df=3; 4996)	738.076*** (df=4; 4995)	432.017*** (df=5; 4994)	125.066*** (df=3; 4996)
Note:					*p<0.1; **p<0.05; ***p<0.01

4. Regression Visualization (Only X and Y, No Controls)



5. Comparison of Fit

Model	Model 1	Model 2	Model 3	Model 4	Model 5
R-squared	0.1	0.1	0.3	0.2	0.1
Adjusted R-squared	0.1	0.1	0.3	0.2	0.1
F-statistic	357.9	161.2	738.1	432.0	125.1