Question: Is the flight delayed based on the distance the airplane is from its destination and departure time the plane leaves from its current destination?

Data Overview:



Utilizing this code to see who the predictors should be for predicting our y variable: arrival time.

A screenshot of a computer screen

AI-generated content may be incorrect.

Viewing the means and standard deviations, it seems as the departure delay and arrival delay are close to each other. This would indicate that the departure delay would be a good predictor for arrival time. Distance seems very far off however considering sometimes distance can cause an arrival delay. These will be my two predictors for the arrival time.

A screenshot of a computer

AI-generated content may be incorrect.

My code consists of reading the csv which includes various data regarding airplane flights. The y variable is the arrival delay and I chose to use distance as the predictor. Creating an OLS in order to find the relationship between the two variables will pop up below:

A screenshot of a computer

AI-generated content may be incorrect.

Looking at the data we have a constant of -0.0018 and a y intercept of 64.8901. This means our equation is y=64.8901 – 0.0018(distance) which means for every one mile the arrival delay will decrease .0018. The pval is less than .05 meaning that our data is statistically significant, and we reject the null hypothesis. Please view the scatter plot below:

A graph of blue dots

AI-generated content may be incorrect.

This scatter plot shows the correlation between the distance of the plane to its destination and the effect it may have on the arrival delay. It has a downward correlation and is skewed to the left. It seems that the closer the distance the more likely there will be an arrival delay. Therefore, predictor on the outside may look like it wont be a good predictor as the distance isn’t much of a cause to arrival delay. So, what other factors could there be?

A screenshot of a computer screen

AI-generated content may be incorrect.

This is departure delay. It contains an equation with y=10.0360 + 0.9242(x). Meaning that every minute the time to departure is delayed the arrival delay increases 0.9242. The pval is once again less than 0.05 meaning that it is statistically significant, and we can reject the null. It is likely that the departure time is a good predictor for arrival time.

A graph with blue dots

AI-generated content may be incorrect.

There is a strong positive linear correlation between arrival time and departure delay. If the departure time increases the arrival time delay is more likely to occur. Therefore, departure delay is a good predictor for arrival time.