

Flight Price Analysis

As Centre College students, a strong desire to travel has been instilled in us as a way to learn about other cultures and encounter new perspectives. However, because we are college students, many of us also have limited financial resources for traveling. This led us to wonder when it is cheapest to travel, where it is cheapest to travel to, and which major airline it is cheapest to travel on with the ultimate goal of planning an inexpensive trip abroad. To further investigate our questions, we decided to examine flight data from February, March, and April of 2022. We specifically selected flights departing from NYC, one of the major hub cities in the United States for international travel. We then decided to look at flights traveling to Paris, France (a common destination for U.S. citizens) and Moscow, Russia (a less common destination for U.S. citizens). Ultimately, we endeavored to answer the following questions through our statistical analysis:

- Is it cheapest to fly in February, March, or April?
- Is it cheaper to fly from NYC to Paris or to Moscow?
- Is American Airlines or United Airlines cheaper to fly?

To begin, we used data collected from Kayak using a web scraper developed by Meshal Alamr. The scraped data included the airline, the duration, the total number of stops, the price, and the date for all flights from New York City to Paris and New York City to Moscow during the months of February, March, and April in 2022. We then uploaded the data to Python in order to clean the data into a more usable form. We converted the price of the flights to USD and separated the flight date into individual day, month, and year columns. To visualize our data, we created a variety of boxplots and found the sample mean of each data set in R-Studio.

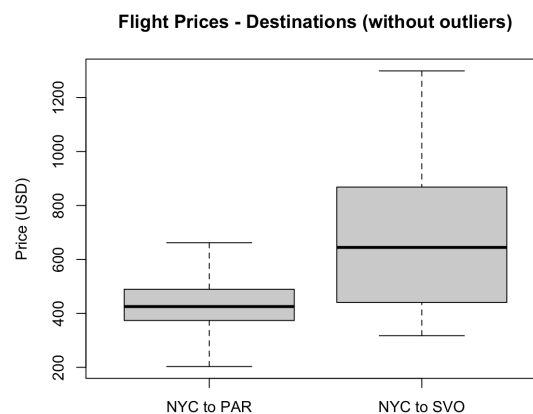


Figure 1. Box plots comparing the distribution of flight prices from NYC to Paris (left) with flight prices from NYC to Moscow (right), excluding outliers.

The sample mean of flight prices from NYC to Paris is 537.58 USD and the sample mean of flight prices from NYC to Moscow is 1157.96 USD, including outliers.

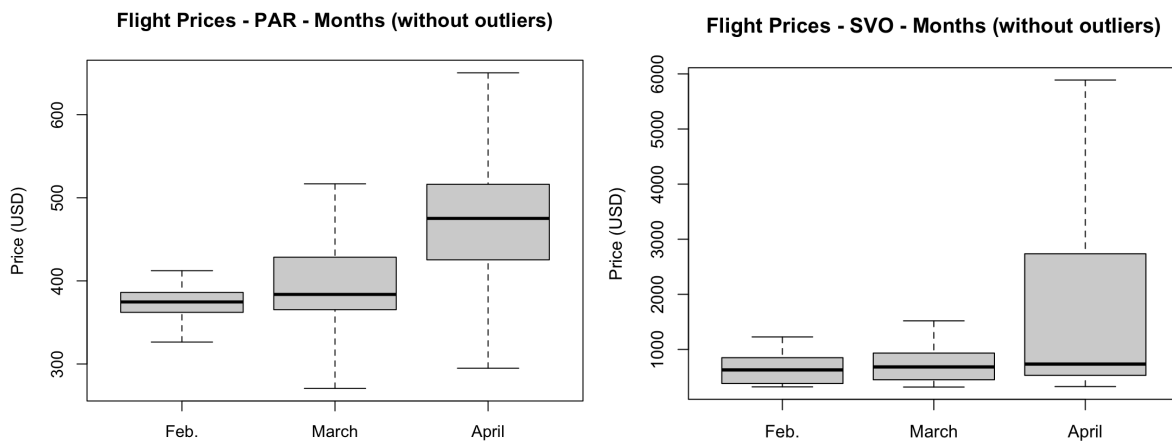


Figure 2 (Left). Box plots comparing the distribution of flight prices from NYC to Paris during February (left), March (middle), and April (right), excluding outliers.

Figure 3 (Right). Box plots comparing the distribution of flight prices from NYC to Moscow during February (left), March (middle), and April (right), excluding outliers.

The sample mean of flight prices from NYC to Paris is 416.37 USD in February, 490.74 USD in March, and 612.57 USD in April, including outliers. The sample mean of flight prices from NYC to Moscow is 978.21 USD in February, 1193.05 USD in March, and 1626.71 USD in April, including outliers.

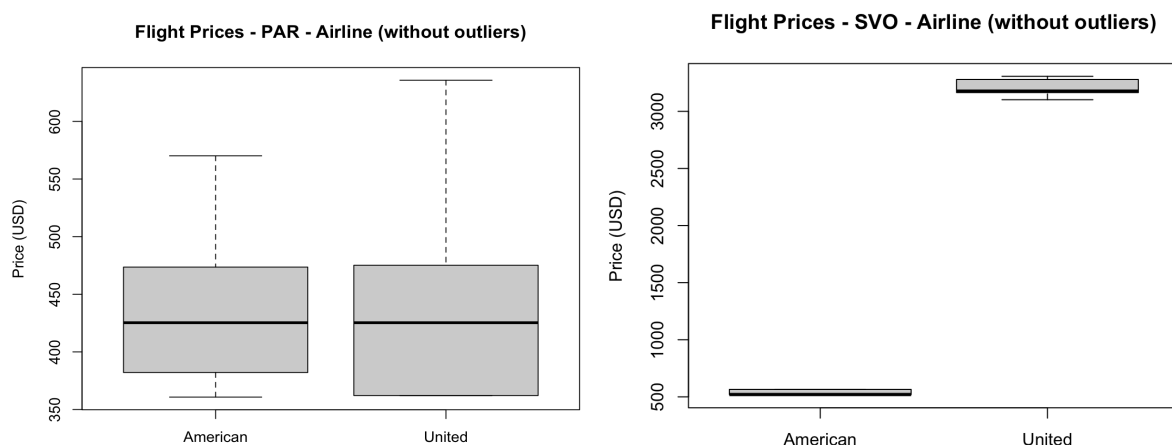


Figure 4 (Left). Box plots comparing the distribution of flight prices from NYC to Paris on American Airlines (left) and United Airlines (right), excluding outliers.

Figure 5 (Right). Box plots comparing the distribution of flight prices from NYC to Moscow on American Airlines (left) and United Airlines (right), excluding outliers.

The sample mean of flight prices from NYC to Paris is 427.72 USD on American Airlines and 452.17 USD on United Airlines, including outliers. The sample mean of flight prices from NYC to Moscow is 619.67 USD on American Airlines and 3315.03 USD on United Airlines, including outliers.

Is it cheapest to fly in February, March, or April?

In order to answer our first question, we ran a series of three hypothesis tests. Prior to the tests, our intuition was that it would be cheapest to fly during the month of February because of the month's winter weather. For the NYC to Paris data, we defined the mean price of February flights as μ_F , the mean price of March flights as μ_M , and the mean price of April flights as μ_A . We then ran a left-tailed T-test where $H_0: \mu_F = \mu_M$ and $H_a: \mu_F < \mu_M$. That is, we tested to see if the mean flight price was cheaper in February than it was in March. The result of this test was a p-value of 0.01. Thus, at the five percent significance level, we concluded that the mean flight price in February was cheaper than in March. Next, we ran a T-test where $H_0: \mu_M = \mu_A$ and $H_a: \mu_M < \mu_A$. That is, we tested to see if the mean flight price was cheaper in March than it was in April. The test resulted in a p-value of $1.01 \cdot 10^{-8}$. Thus, at the five percent significance level, we concluded that the mean flight price in March was cheaper than in April. Finally, we ran a T-test where $H_0: \mu_F = \mu_A$ and $H_a: \mu_F < \mu_A$. That is, we tested to see if the mean flight price was cheaper in February than it was in April. The test resulted in a p-value of $2.20 \cdot 10^{-16}$. Thus, at the five percent significance level, we concluded that the mean flight price in February was cheaper than in April. After running these three tests for the NYC to Paris data, we repeated the exact same series of tests for the NYC to Moscow data. These tests resulted in p-values of 0.0004, $1.80 \cdot 10^{-5}$, and $2.10 \cdot 10^{-11}$ respectively. Thus, we reached the same conclusions for the Moscow data that we reached for the Paris data. Therefore, of our three months, we determined that the cheapest month to fly was February, the middle-priced month to fly was March, and the most expensive month to fly was April.

Is it cheaper to fly from NYC to Paris or to Moscow?

In order to answer our second question, we ran one hypothesis test. Prior to the test, our intuition was that it would be cheaper to fly from NYC to Paris rather than from NYC to Moscow because Paris is closer to NYC than Moscow is. We defined the mean price of flights from NYC to Paris as μ_P and the mean price of flights from NYC to Moscow as μ_M . We ran a left-tailed T-test where $H_0: \mu_P = \mu_M$ and $H_a: \mu_P < \mu_M$. That is, we tested to see if the mean flight price from NYC to Paris was cheaper than the mean flight price from NYC to Moscow. The result of this test was a p-value of $2.20 \cdot 10^{-16}$. Thus, at the five percent significance level, we concluded that the mean flight price from NYC to Paris was less than the mean flight price from NYC to Moscow. Therefore, of our two destinations, we determined that the cheaper place to fly to was Paris.

Is American Airlines or United Airlines cheaper to fly?

In order to answer our third question, we ran one hypothesis test. Prior to our test, our intuition was that there would be no difference between American Airlines and United Airlines flight prices since the airlines compete directly with one another. But, our box plots and sample means indicated that American Airlines may be cheaper than United Airlines which ultimately influenced our choice of test. We defined the mean price of flights from NYC to Paris on American Airlines as μ_{AA} and the mean price of flights from NYC to Paris on United Airlines as μ_{UA} . We ran a left-tailed T-test where $H_0: \mu_{AA} = \mu_{UA}$ and $H_a: \mu_{AA} < \mu_{UA}$. That is, we tested to see if the mean flight price from NYC to Paris on American Airlines was cheaper than the mean flight price from NYC to Paris on United Airlines. The result of this test was a p-value of 7.70×10^{-6} . Thus, at the five percent significance level, we concluded that the mean flight price from NYC to Paris on American Airlines was less than the mean flight price from NYC to Paris on United Airlines. After running this test for the NYC to Paris data, we repeated the exact same test for the NYC to Moscow data. There were no direct flights on American Airlines or United Airlines from NYC to Moscow, so we used data on any flights with at least one leg operated by either airline. This test resulted in a p-value of 2.20×10^{-16} . Thus, we reached the same conclusions for the Moscow data that we reached for the Paris data. Therefore, of our two airlines, we determined that it was cheaper to fly on American Airlines.

Combining the conclusions reached while answering our three primary questions, we decided that we wanted to determine a confidence interval for the price of flights from NYC to Paris on American Airlines during the month of February. Based on our confidence interval, we are 95% confident that the true mean price of flights from NYC to Paris on American Airlines during the month of February is between 378.72 USD and 381.21 USD.

Based on our findings, we would recommend traveling during the month of February from NYC to Paris on American Airlines in order to minimize flight cost. When making this recommendation, we are assuming that 2024 flight prices will follow similar trends to 2022 flight prices, which, despite being a reasonable assumption, we know is not guaranteed to be the case. One drawback of our investigation is that we were limited by our data to only three months and two destinations. We also only tested two airlines. Therefore, we are unable to make conclusions about flight prices at different times, to different locations, using different airlines. Another potential drawback to our conclusion is that we tested flight cost in terms of month, destination, and airline independently. Thus, while February is the cheapest month to fly, Paris is the cheapest destination to fly to, and American Airlines is the cheapest airline to fly on, it is not necessarily true that it is cheapest to fly in February to Paris on American Airlines. Additionally, while our hypothesis tests all yielded statistically significant results, our results are not guaranteed to have practical significance. For example, while we concluded that it is cheaper to fly in February than it is to fly in March, if the mean flight price in these two months only differed by 10 USD it might not impact which flights we ultimately decide to take. Lastly, we want to mention the role of outliers. Some of the flights in our data set are extremely expensive

meaning our sample means are at risk of being higher than is truly representative of our data set. If we were to continue our project in the future, we could expand our analysis to encompass more destinations, airlines, and months. We could also analyze the impact of new variables like weekday and flight duration. Finally, we could remove outliers from our dataset and observe how the removal of outliers affects the conclusions of our hypothesis tests.

Bibliography

Alamr, Meshal. "Flight Price Prediction." *GitHub*, 1 Nov. 2021, github.com/MeshalAlamr/flight-price-prediction.