United Airlines Usability Test

Prioritizing Airfare Affordability

Background

Current Workflow

Flights are sorted by "**best match**."

This includes factors like travel **time**, number of **stops**, and **cabin type**.

The Problem

Many users value **only affordability**.

It's difficult to sort flights by just **price**.
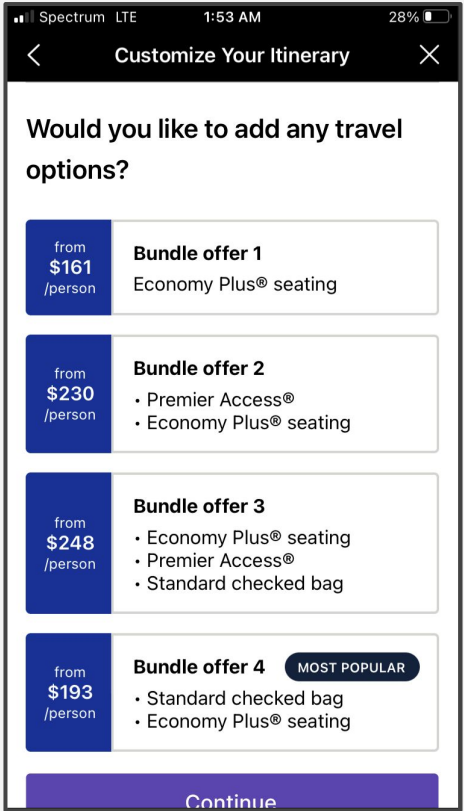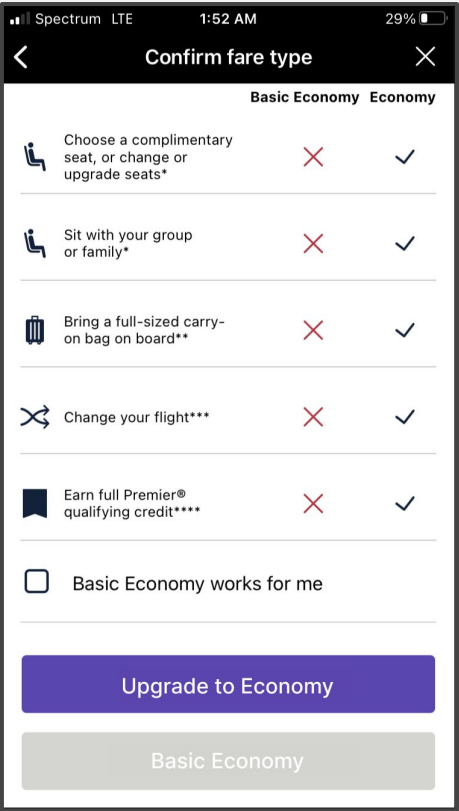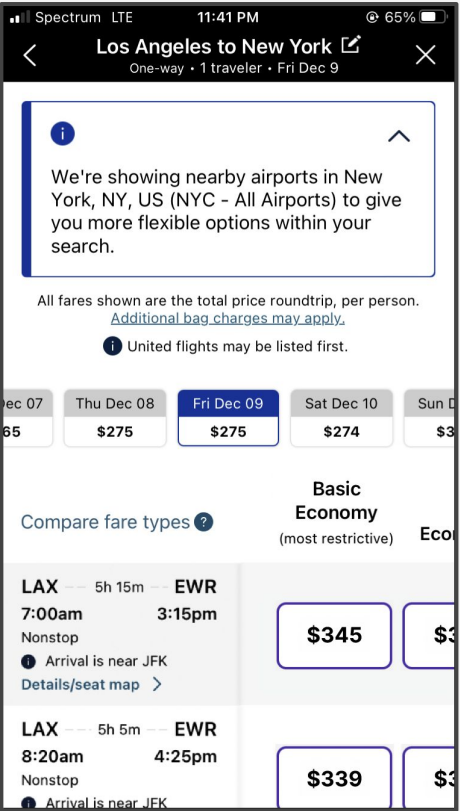
Upsells **complicate** the process.

The Solution

Modifications that prioritize affordability.
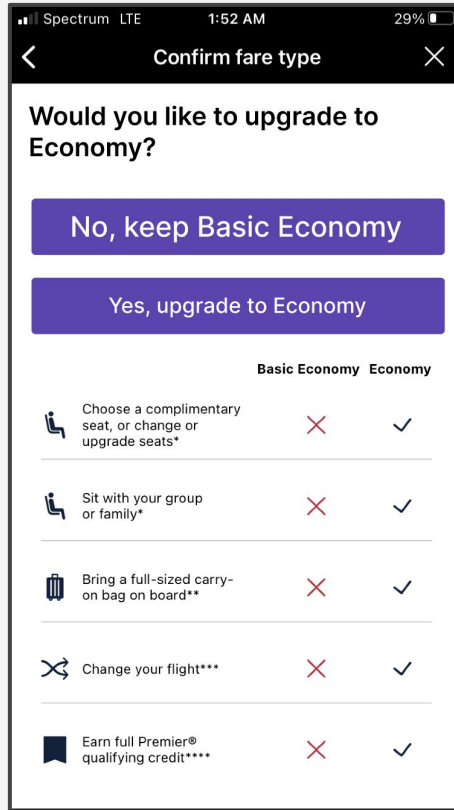
visual cues that flights are sorted by price.

removal of upsells.

Original Workflow

Original Workflow
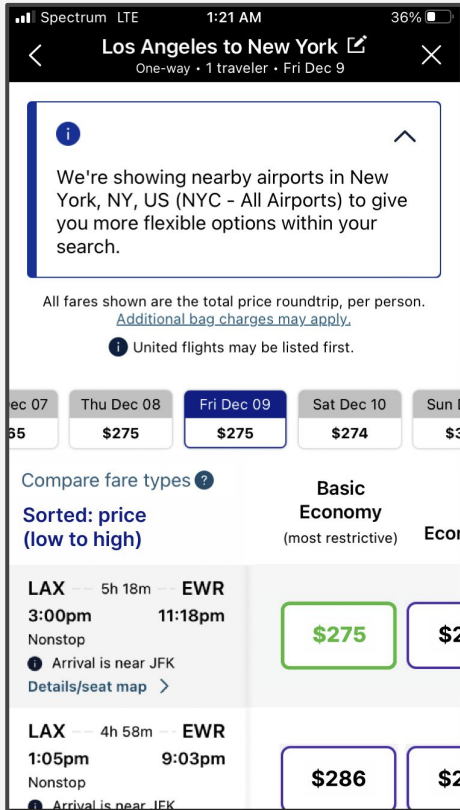
Hypotheses

These changes will:

1. **increase** app usability
2. **improve** user attitudes toward United Airlines
3. **not affect** user spending

Confirming these trends would allow for changes to the app that benefit the user without costing the airline.

Methods

What was done

Participants

Task Description

Participants

Convenience sample of 15 participants; **two** were excluded from the analysis. This study used a **between-subjects design** to compare app versions.

Participants were **screened** before random assignment. Those included were:
- frequent fliers
- not familiar with the United Airlines app
- interested in using flight booking apps

The task was **unmoderated**; some completed it remotely, others in person.

Task Description

Participants were asked to **complete the flight selection process** for a trip from Los Angeles to New York. This involved:
- **selecting** a flight
- **confirming** the fare selection
- **choosing** between or opting out of upsells

Post-experiment questionnaire consisted of:
- five items measuring positive attitudes toward United Airlines
- the Single Ease Questionnaire
- one question assessing user considerations when choosing a flight

Results

What was found

App Usability

User Attitudes

User Spending

————

App Usability

Measures

- number of interactions
- time spent
- SEQ rating

Trends

The modified app version showed **modest to large improvements** in all measures of usability.

Number of Interactions

Participants who used the modified app performed **40% fewer** interactions than those who used the original version.
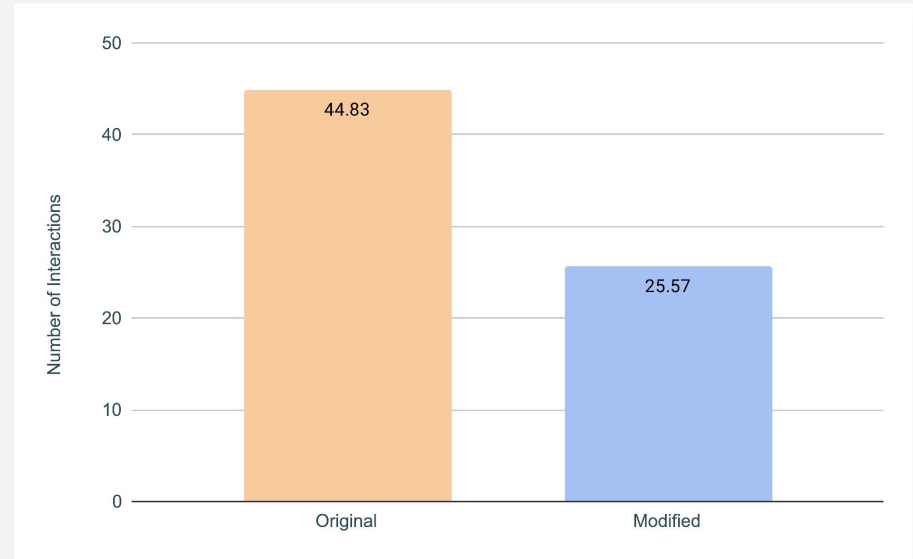
Time Spent

Participants who used the modified app completed the process **30% faster** than those who used the original version.
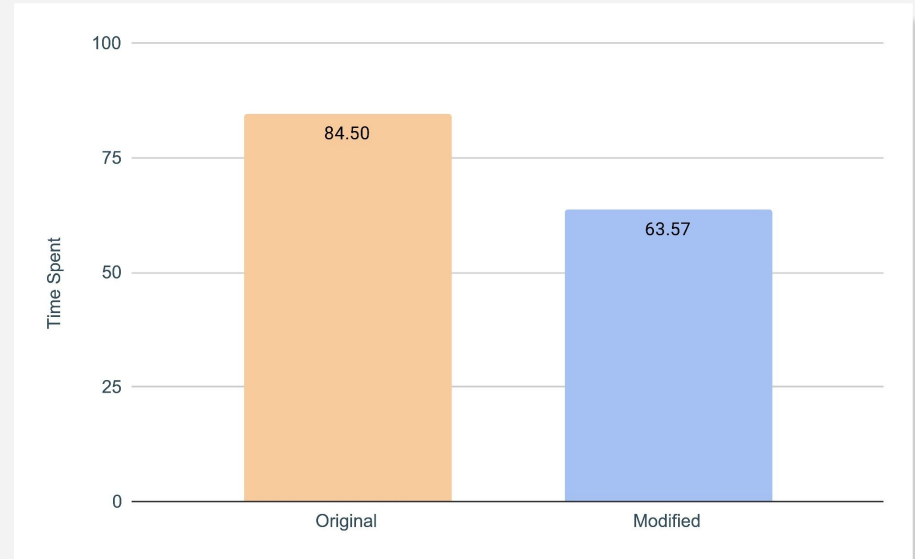
SEQ Rating

Participants who used the modified app reported an SEQ rating **20% higher** than those who used the original version.
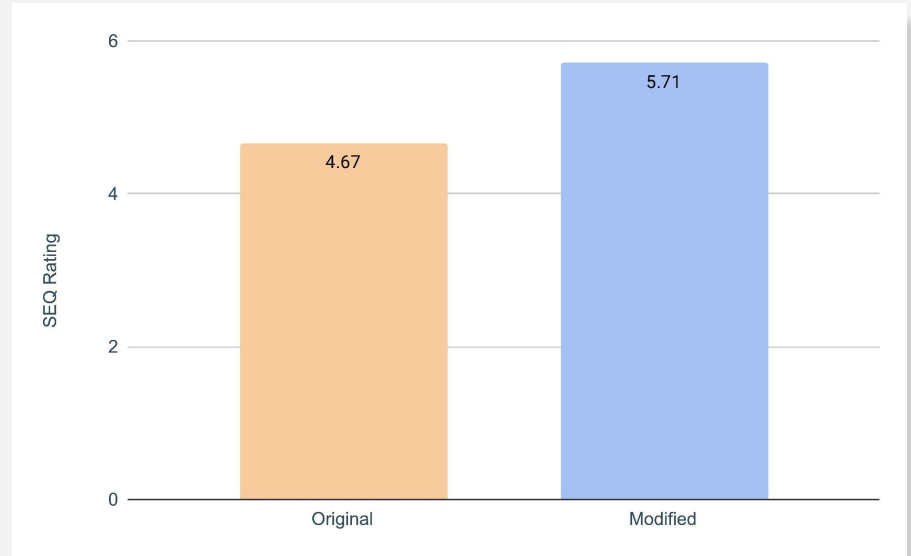
User Attitudes

Measures

- five items assessing positive attitudes toward United Airlines

Trends

The modified app **improved attitudes** of certain user.

Attitudes

Users who considered only price and used the modified app indicated more positive attitudes on **all** questionnaire items.
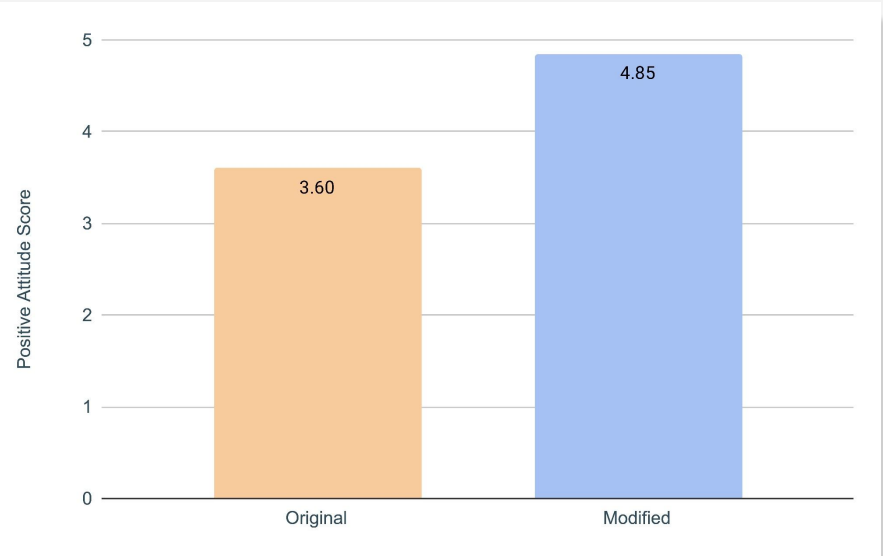
User Spending

Measures

- amount spent on airfare
- amount spent on upsells

Trends

The modified app **did not affect** user spending.

Money Spent

Those who used the modified app **did not** spend less money on flights and upsells than those who used the original version.
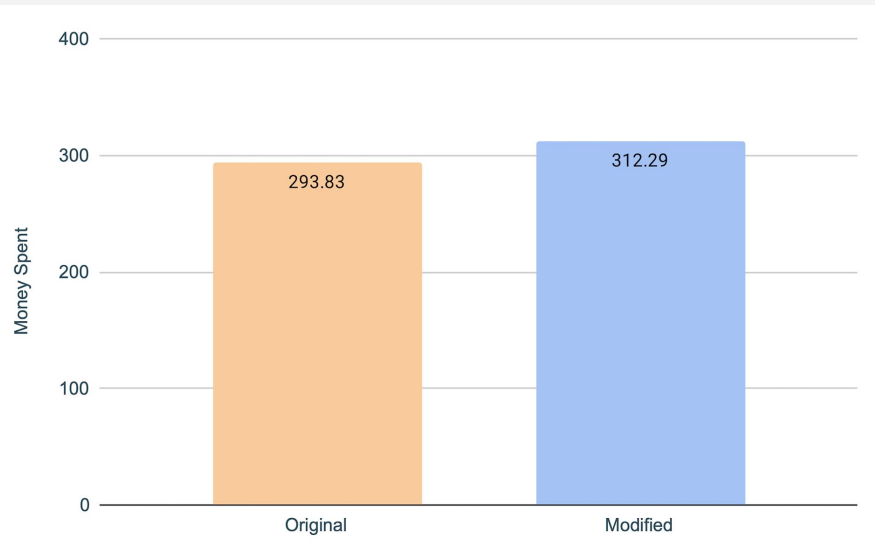
Recommendations

These simple design changes could be implemented now to **streamline** the flight selection process at **no cost** to the airline.

Beta testing on a **larger sample** of United Airlines app users could be done before a full rollout to **verify** these findings.
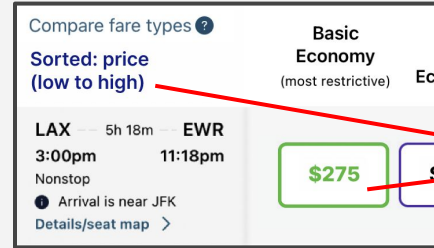
visual
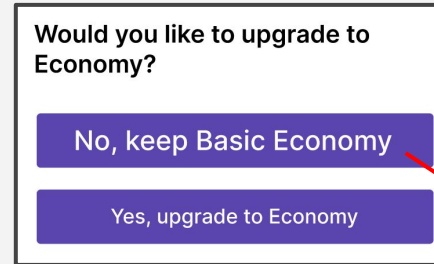
structural
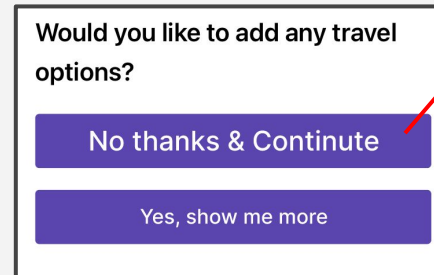
Key Findings

Modifications that prioritize affordability:

- **increased** app usability
- **improved** user attitudes toward United Airlines
- **did not affect** user spending

Not all of these differences were statistically significant, but moderate and large effect sizes support these trends.

Test Plan

Project Summary

At present, United Airlines returns flight options to users sorted by "Best matches" by default. How "Best matches" is determined is unclear, but appears to be an optimization of several factors which gets offers users itineraries with few stop, short total travel time, convenient departure and arrival times, low price, and nearby airports which may have better options. For many users, especially low income users, cost is by far the most important consideration when selecting fights, regardless of almost all other factors, making the concept of "Best matches" much less useful.

At present, users have to select the "Economy" option, near the bottom of the sort filter, *after they've searched for fights*, if they want to sort flights by cost (low to high). There's no "Sorted: price (low to high)" option. Then, after the user has selected a flight, they're prompted to select from a list of "bundle offers" to add amenities and services like "Premier Access" and checked luggage. There's not a "No thank you" or "None" option, you just have to select "Continue" at the bottom of the page.

The aim of this comparative usability test was to design an affordability-prioritizing workflow that helps users quickly find cheap flights and easily bypass additional purchases. Hopefully, these modifications will create a more usable app and foster more positive attitudes toward United Airlines *without* decreasing sales.

Research Goals

Specifically, my interest is in determining if modifications to the flight selection process (table 1) which *prioritize affordability* can be used to:
1. change the number of actions carried out by users
2. change time spent by users selecting flights
3. influence users' attitudes toward United Airlines
4. impact ease of use
5. affect how much money users spend

To examine this, users were assigned to either the "Original workflow," a prototype identical to the current version of the Untied Airlines app, or the "Modified workflow," a prototype with modifications intended to streamline the flight selection process for those who value affordability.

Participants

Sampling

A sample size estimation for an independent sample t-test was calculated using the primary outcome variable of interest — number of interactions required to complete the task. This variable is particularly important, as it likely impacts not only time taken on the task, but also users' responses to questionnaire items, as it's fundamental to the user experience.

To estimate the critical difference needed to calculate minimum sample size, I asked two volunteers to complete the task using both the original and modified workflow (with order counterbalanced). They completed the original workflow with 17 and 18 actions, and the modified with only 11 and 9 actions. I opted for a repeated measures design because I lacked the resources to carry out a larger pilot study. While this method is bound to under estimate the true variance of each group, I felt it was better than forgoing an estimation entirely.

From this "pilot study," I estimate the mean number of interactions required to complete the original workflow to be 17.5, and the mean of the modified workflow to be 10, with variances of 0.5 and 2 respectively. This corresponds with a pooled variance of 1.25 and a pooled standard deviation of roughly 1.12. Therefore, the size of this observed effect is large (d=6.69), but very likely not representative of the true effect, given the small sample and dependent observations. With that in mind, I would like to be able to detect an effect even when the difference between group means is only one standard deviation (d=1), which corresponds with a raw difference of 1.12 interactions. After iterating through the sample size iteration procedure five times, I found the estimated minimum sample size to be 6 participants per group.

Initially, this minimum sample size was met, as data from 15 participants (seven in the original condition and eight in the modified) were collected and data from 13 participants (six in the original condition and seven in the modified) used. However, the analysis became more complex, as discovered a meaningful blocking variable and opted to conduct a series of factorial ANOVAs instead. Given the modified analysis, this study was underpowered, with only three to four participants per cell, greatly limiting the my ability to find statistically signifiant effects.

Screener

Participants were screened before random assignment to app version. To qualify for inclusion in this study, participants had to be frequent fliers who were not familiar with the United Airlines app, but did have experience with and interest in using flight booking apps. The task was unmoderated, though some participants completed it remotely and others in person.

The inclusion criteria were assessed using the following questions:
1. Do you fly often (at least once a year)?
2. Have you flown with United Airlines?
3. Did you purchase your tickets using their app?
4. Would you purchase your tickets using their app?

Design and Procedures

Differences Between Workflows

Original workflow	Modified workflow
Flights are sorted by "Best match"	Flights are sorted by price, which is made clear to the user with salient visual elements including a written indication that flights are "Sorted: price (low to high)" and green highlighting of the cheapest flight.
If a user selects a Basic Economy fare, they are prompted to upgrade to Economy by completing an action to opt out (selecting "Basic Economy works for me") and move forward with the Basic Economy fare. Additionally, "Upgrade to Economy" is above "Basic Economy," even after the user indicates that they want to continue with Basic Economy.	Users are now asked explicitly if they'd like to upgrade to Economy. The options are "No, keep Basic Economy" and "Yes, upgrade to Economy," which appear in that order, near the top of the page above the Basic Economy and Economy comparison chart.
Users are presented with four "Bundle offers" at various price points. To continue to checkout, users must click the "Continue" button below the Bundle offers. There's no option that clearly indicates users are not interested in purchasing a bundle.	There are now only two options on the Bundle offers page: "No thanks & continue," and "Yes, show me more." The former takes users to the end of the task, while the latter takes them to a second Bundle offers page where they can choose which bundle they'd like.

Table 1. Workflows are identical apart from these modifications; the design is experimental, between subjects, and employs mixed methods.

Hypotheses

Compared to those who use the original app version, those who use the modified app version, designed to prioritize affordability, will:

1) perform fewer interactions; (the combined number of swipes, cursor movements, and clicks); H_0: $\mu_{original} = \mu_{modified}$; H_1: $\mu_{original} > \mu_{modified}$
2) spend less time selecting flights; H_0: $\mu_{original} = \mu_{modified}$; H_1: $\mu_{original} > \mu_{modified}$
3) report higher average scores on the five questionnaire items assessing positive attitude toward United Airlines; H_0: $\mu_{original} = \mu_{modified}$; H_1: $\mu_{original} < \mu_{modified}$
4) report a higher SEQ rating of the app; H_0: $\mu_{original} = \mu_{modified}$; H_1: $\mu_{original} < \mu_{modified}$
5) spend no less money on flights and upsells; H_0: $\mu_{original} = \mu_{modified}$; H_1: $\mu_{original} \neq \mu_{modified}$

General Procedures

Introduction, NDA, Consent to Record

Participants were read and responded to the following script after completing the screening assessment and before starting the task:

"Thank you for volunteering to participate in this study! In a moment, you will be asked to complete the flight-selection process for a hypothetical trip you might take. Sharing information about this study may compromise our results and negatively affect the associated design recommendations. Please do not discuss details of this study with anyone. Is this alright with you?

The risks associated with participation in this study are minimal. Do you consent to participation in this study?

Can we use data collected from you during this study in our analysis? You can opt out later if you change your minds."

Task Description
Once participants were assigned to a treatment, informed of the study, and given consent, they were asked to complete the flight selection process for a nonstop, one-way trip from Los Angeles to New York. This involved choosing from one of eight flights, confirming that fare selection, then choosing between or opting out of various "bundle offers." Participants were instructed to "complete this process as they would if they were actually booking a flight." The task instructions script that was read to participants is included in the next section.

The task was carried out on participants' personal computers and screen recorded for data collection. After participants competed the task, they completed a post-experiment questionnaire.

Task Script
"You're about to complete the flight selection process for a one-way, nonstop flight from Los Angeles to New York. There are several options to choose from, all depart on the same date, from the same departure airport, and arrive at the same destination airport. For the sake of this task, just imagine that you chose these constraints and actually do want to travel on this date from and to these locations. Please select a flight that you would actually select if you were to take this trip."

Post-task Questionnaire
"Please indicate your level of agreement with the following questions
(1 = not at all, 7 = very strongly)
1. United Airlines values my wants and needs.
2. United Airlines values my time.
3. I feel positively toward United Airlines.
4. I would like to fly with United Airlines in the future.
5. I would use this app to book flights in the future.

Please answer the following questions
1. What qualities did you prioritize when selecting flights?
    ☐ Price
    ☐ Departure time
    ☐ Arrival time
    ☐ Fare type
    ☐ Other

Overall, how difficult (1) or easy (7) was this task to complete?

Debrief

"Thank you for your participation in this usability test! I am assessing the overall ease of use and money spent during the flight selection process of the United Airlines app, as well as users' priorities when selecting flights and general attitudes toward United Airlines. To study this, modifications intended to help users quickly find cheap flights and easily bypass additional purchases were made to the Untied Airlines app. Your contribution is greatly appreciated. If you have any questions or concerns regarding this study, please contact samuel.light@cgu.edu.

Analysis

Video files were coded and the data entered into a spreadsheet (Google Sheets). After cleaning, data analysis will done using R Studio. Initially, the analysis was going to consist of a series of independent sample t-tests. However, while cleaning the data, I discovered that a large proportion of participants (46%) considered factors other than price while selecting a flight. This variable was correlated with most outcome variables, so I treated it as a blocking variable and ran a factorial ANOVAs in place of t-tests, when appropriate, to control for whether or not participants considered *only* price when selecting a flight. This allowed for isolation of the effect of app version alone. Post-hoc analyses were conducted on all omnibus tests with at least one significant effect. Additionally, the internal consistency of the questionnaire was assessed.

Research Report

Summary

- Modifications designed to prioritize affordability **decrease** the number of interactions required of users while completing the flight selection process, and allowed users to do so **more quickly**. These modification may also **improve** certain users' attitudes toward United Airlines.
- These modifications **did not** decrease user spending on flights or upsells.
- Modifications that implicitly and explicitly prioritize affordability have the potential to streamline the flight selection process and improve users' attitudes toward United Airlines *without* compromising revenue.

Background

By default, the United Airlines mobile app displays flights that match the filter option "best match." This option appears to optimize for several factors including departure and arrival time and city, travel duration, number of stops, and fare price. However, it's difficult to find and use the filter feature to sort flights by any of these factors independently, particularly by price (there's no option that allows users to explicitly sort by price, only by fare type).

Because cost is a limiting factor for many travers, I designed a modified flight selection workflow that prioritizes affordability, rather than "best match." These changes include salient visual indications that flights are sorted by price, and the structural modifications which remove barriers to affordable decision making. This study compared the modified and original versions of the app to assess the potential utility of these changes. Ideally, these changes will benefit the user without costing the airline.

Hypotheses

Modifications to the United Airlines app that prioritize affordability will:
1) Increase app usability
2) Improve user attitudes toward United Airlines
3) Not affect user spending

Methods

Participants
Data were collected from a convenience sample of 15 participants. Two participants were excluded from the analysis; one due to technical issues and the other due to a misunderstanding of the task. Participants were screened before random assignment to app version. To qualify for inclusion in this study, participants had to be frequent fliers who were not familiar with the United Airlines app, but did have experience with and interest in using flight booking apps. The task was unmoderated, though some participants completed it remotely and others in person.

Task description

Participants were asked to complete the flight selection process for a one-way trip from Los Angeles to New York. This involved choosing a flight, confirming the fare selection, then choosing between or opting out of additional purchases called "bundle offers." Participants were instructed to "complete this process as they would if they were actually booking a flight."

The task was carried out on participants' personal computers and screen recorded for data collection. After participants completed the task, they were given a post-experiment questionnaire consisting of: five statements regarding attitudes toward United Airlines, one question assessing factors they considered while selection flights, and the Single Ease Question. Items assessing attitudes and ease of use were scored on a seven-point Likert scale, while the assessment of factors considered by participants during the task was a check all that apply item.

Analysis

After data collection, coding, and cleaning, a series of Factorial ANOVAs were independently performed on all relevant outcome variables by app version, controlling for whether or not participants considered *only* price when selecting a flight. This covariate was treated as a blocking variable, allowing for isolation of the effect of app version alone. A single MANCOVA could have been performed given a sufficiently large sample. Post-hoc analyses were conducted on all omnibus tests with at least one significant effect. Additionally, the internal consistency of the questionnaire was assessed.

Results

Hypothesis 1

A Factorial ANOVA revealed a large, significant main effect of app version on number of interactions performed by users, when controlling for whether or not users considered only price while selecting flights, $f(1,9)=15.20$, p=.004, $\eta^2p=.628$. Specifically, those who used the modified app version carried out significantly fewer interactions ($M=25.57$) than those who used the original app version ($M=44.83$), $t(9)=3,90$, p=.004, $d=2.18$. This provides *strong* evidence in support of hypothesis one.

A Factorial ANOVA revealed a moderate, though insignificant, main effect of app version on time taken to complete the task, when controlling for whether or not users considered only price, $f(1,9)=2.20$, p=.172, $\eta^2p=.196$. Specifically, those who used the modified app version completed the process more quickly ($M=65.42$ seconds) than those who used the original app version ($M=84.50$ seconds), $t(9)=1.48$, p=.178, $d=.83$. Taken with the large effect size, this provides tentative evidence in support of hypothesis one.

Finally, a two-sample t-test comparing SEQ scores of both app versions suggested that the modified version was rated as easier to use, $t(11)=.93$, p=.371, $d=.52$. This difference was not significant, however the moderate effect size is worth noting and likely indicative of a modest effect obscured here by lack of power.

Hypothesis 2
A factorial ANOVA revealed no main effect of app version on positive attitudes toward United Airlines. However, a marginally signifiant interaction of moderate effect size was found, $f(1,9)=3.05$, p=.115, $η^2p$=.253. Specifically, those who considered only price while selecting flights and used the modified app version expressed warmer attitudes toward United Airlines, $t(6)=1.19$, p=.284, d=.92. While not statistically significant, the magnitude of this difference was large. This effect was especially strong for Item 4, $t(5)=2.29$, p=.155, d=1.75.

Hypothesis 3
A two-sample t-test found no significant or sufficiently large effect of app version on amount of money spent by users, $t(11)=.55$, p=.591, d=.31.

Questionnaire
The internal consistency of the questionnaire was found to be very high (5 items; $α = .91$), suggesting that the items could be combined and a composite score meaningfully interpreted.

Summary
The modifications designed to prioritize affordability, assessed in the present study, effectively decrease the number of interactions required of users, and *may* decrease the amount of time taken by users, when completing the flight selection process. There is tentative evidence that these modifications produce a substantial increase in self-reported intent to "fly with United Airlines in the future" among users who considered only price while selecting flights. Finally, these quantitative and qualitative benefits to the user come at no cost to United Airlines, as there was no effect of app version on amount of money spent by users. Together, these findings provide justification for the implementation of features which prioritize airfare affordability during the flight selection process, as these features likely improve user experience at no cost to the airline.

Limitations
The prototype built for this study was fairly restrictive, and the sample size was small. A follow-up study with a larger sample size may expand on the present study by allowing greater autonomy of users during the flight selection process.