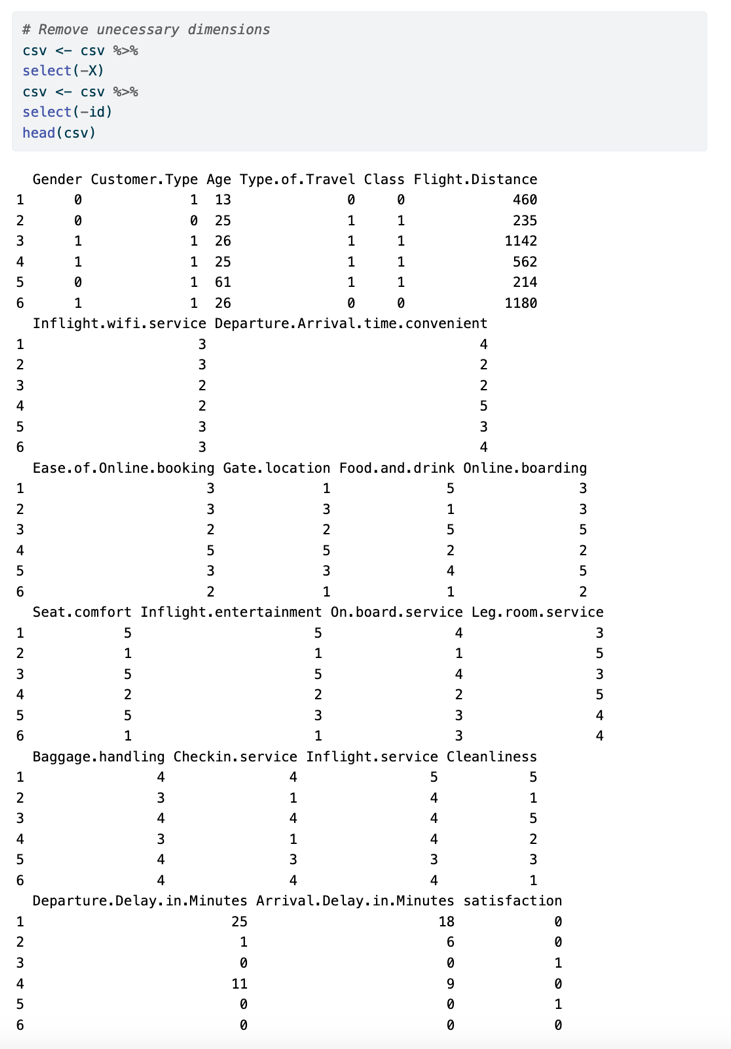
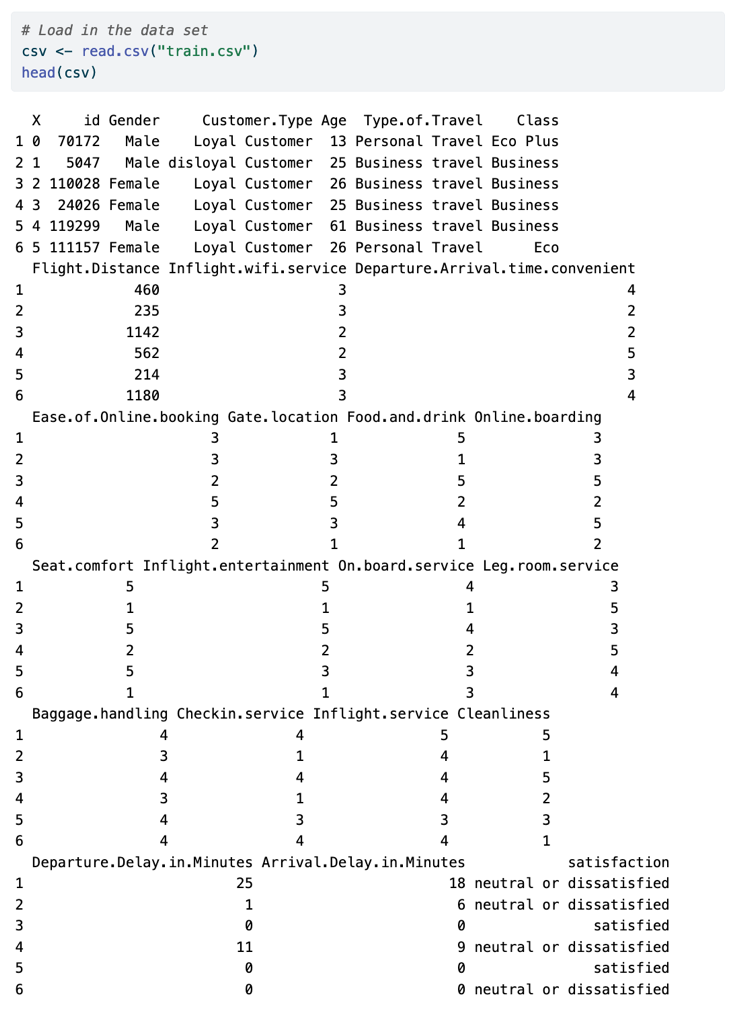
**Exploratory Factor Analysis**

A screenshot of a computer program

Description automatically generatedA computer code with text

Description automatically generatedI completed this analysis in R and will lay out my steps in this section. First, I imported the necessary libraries needed for exploratory factor analysis.

Then, I loaded the data set as a CSV and did the necessary data wrangling to complete the analysis- which included mutating the categorical variables into binary (0,1) and removing the identifier variables that weren't usable in the analysis.

Then, I had to reorder the columns in our dataset to align with the structural equation model and be in the same order.

A screenshot of a computer program

Description automatically generated

Next, I computed the correlation and covariance matrices for the factor rotation and structural equation model.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Based on Kaiser and Jolliffe's criterion, I then summed the eigenvalues in the correlation matrix to see how many factors would be needed. I ended up with ten factors by Jolliffe's criterion to include values greater than 0.7.

A screenshot of a computer

Description automatically generated

I then computed the loading matrix using orthogonal rotation and did a correlation plot of the results.

A screenshot of a computer

Description automatically generatedA close-up of a document

Description automatically generatedA screen shot of a graph

Description automatically generatedTo check if oblique factor rotation was necessary, I looked at the correlations between factors, and since none of them were above the common threshold of 0.3, it was unnecessary, and I stuck with orthogonal rotation.

Next, I checked which of the ten factors each dimension loaded onto to use to create a path diagram and then named the factors by their loadings.

A screenshot of a computer program

Description automatically generatedA list of flight information

Description automatically generated

A diagram of a business

Description automatically generated with medium confidenceI then created a path diagram to illustrate the relationship between dimensions and factors to predict airplane passenger satisfaction. I found that the sixth factor, which I named Ease, could be removed to limit co-variability.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a document

Description automatically generatedThen, I created the Structural Equation Model according to the path diagram to calculate the optimized parameter estimates.

**Parameter Estimates**

Food and Drink had a 0.892 loading onto Hospitality, Seat Comfort had a 0.9 loading onto it, and Inflight Entertainment had a 0.683. Legroom Service has a positive 0.6 loading onto Accommodations and Inflight Entertainment with 0.382, while On Board Service had a loading of 0.94 and Inflight Service has 1.051 onto it. Ease of Online Booking loads onto Facilities with 1.056, Online Boarding with 0.472, and Departure/Arrival Time Convenience loads onto Convenience with 0.978 and Online Boarding with 0.049. Arrival Delay in Minutes loads onto Delays with 1.118. Privilege has loadings from Class with 1.738, Type of travel with 1.205, and Customer type of 0.119.

One variable for each factor was fixed: Cleanliness for Hospitality, Baggage Handling for Accommodations, Inflight Wi-Fi service for Facilities, Gate Location for Convenience, Departure Delay for Delays, and Flight Distance for Privilege.

The strongest loading on Hospitality is Seat Comfort, Accommodations is Inflight Service, and Facilities is Online Boarding, Convenience with Departure/Arrival time convenience, and Privilege with Class.

The loadings onto satisfaction are as follows: Hospitality with 0.220, Accommodations with 0.238, Facilities with 0.625, Convenience with -0.638, Delays with -0.030, Privilege with 0.758, Age with 0.026, Gender with -0.011, and Check in service with 0.089. The strongest loading on satisfaction is Privilege, and the weakest is Gender.

A screenshot of a computer

Description automatically generated**Fit Indices**

From Comparative Fit Indices, the Normed Fit Index (NFI) evaluates the fit of a model by comparing its chi-square statistic to the chi-square statistic of the corresponding independence model and has a standard threshold of values above 0.9 or 0.95, indicating the model is a good fit of the data. Our NFI was 0.843, meaning this model does not adequately fit the data compared to the baseline model. The following indices I looked at were the Root Mean Square Error of Approximation (RMSEA), which measures the disadvantage of using the actual model compared to the saturated model. The standard threshold for providing a good fit to the covariance structure of the data matrix is any value below 0.08, and our model had a value of 0.09, which is close but not enough to indicate a well-fitting model. The last indices I looked at was the Goodness-of-Fit Index (GFI), which is typically interpreted as analogous to the coefficient of determination R^2. The GFI was equal to 0.847 for our model, meaning our structural equation model can explain approximately 84.7% of the variance and covariance in the original data. Overall, it does not seem like our model adequately represented the original data, which can happen. In the future, I would use different variables and try a different structural equation model.