|  |  |
| --- | --- |
| Table W3.2: Evaluation of Research Fit and Resource Use | |
| 1) Sampling and Generalizability | |
| 1.1. How can instances of entities be sampled from the site? | The instances of entities will be sampled from the site through random sampling from the reviews(instances) 2019 and onwards covering the COVID-19 period as well as the most popular airports and airlines. Selecting this time span does not necessarily have a high bias as we want to take inferences about how COVID-19 changed the traveler behavior. |
| 1.2. What sample size is required to test the predictions? |  |
| 1.3. what is the technically feasible sample size? |  |
| 2) Construct Measurement | |
| 2.1. Can the constructs of interest be measured with the available data? | Data is clear and transparent that makes it available to calculate constructs of interest. |
| 3) Data Structure and Preparation | |
| 3.1. What data is required to answer the desired research question? | The type of data provided by the website is panel data it is multi-dimensional over time. This type of data is sufficient to answer the RQ, and provides a ground for future inferences. |
| 3.2 How can the raw data be converted to dataset that can be analyzed? | Do not have information yet on what is needed to transform the data. But we aim at storing the data in the type of CSV file or Google Drive, and preferably in the structure of rows and columns, and graphics to visualize the data better. |
| 4)Resource use | |
| 4.1. What are the development costs? | There are no foreseen development costs to scrape the data from the website |
| 4.2 How much will it cost to run the scraper? | No cost foreseen to run the scraper |
| 4.3 How costly is it to maintain extraction software? | No foreseen cost to maintain the extraction software |
| 4.4 Opportunity costs | To answer the desired RQ novel data is seen to be a better fit. |