# What do your three files look like online?  Which one looks best?

# Do you think there are any problems with your dataset? Explain.

* **Missing Values:**
  + Your notebook checks for missing values using table1.isna().sum(), which suggests there might be missing data.
  + If important columns have missing values, it could affect analysis.
* **Data Distribution Issues:**
  + There is a histogram (sns.histplot(df["total"]...) that visualizes the distribution of "Total Travel Cost." If the data is heavily skewed, this could impact predictive modeling.
* **Data Types and Format:**
  + The command table1.describe(include=('all')) suggests your dataset contains categorical and numerical columns.
  + If categorical values have inconsistent formats (e.g., "USA" vs. "U.S."), this might need cleaning.
* **Potential Outliers:**
  + The histogram may reveal extreme values. Outliers in travel cost, for example, could skew the analysis.

# What Questions do you have about the data? Is there a problem you would want to solve? Something to predict?

**Questions About the Data**

* **What does each column represent?**
  + Are all variables meaningful, or are there redundant ones?
* **What is the goal of this dataset?**
  + Is it for analysis, forecasting, or classification?
* **Are there any unexpected values?**
  + Example: Negative travel costs or incorrect country names.

**Possible Problems to Solve**

* **Predicting Travel Costs:**
  + Can we predict total travel cost based on factors like distance, number of travelers, and destination?
* **Detecting Anomalies:**
  + Are there unusually high or low travel costs that need investigation?
* **Trend Analysis:**
  + Are there seasonal patterns in travel expenses?