

TSA Case Study: Data and Report Requirements

To prepare and analyze the data, you follow these requirements given to you by your supervisor. Be aware that these requirements are only provided for this case study. They are not an accurate representation of TSA requirements.

Data Requirements

- The raw data file **TSAClaims2002_2017.csv** must be imported into SAS and stored in a file named **claims_cleaned** in the **tsa** library.
- Entirely duplicated records must be removed from the table.
- All missing and '-' values in the columns **Claim_Type**, **Claim_Site**, and **Disposition** must be changed to *Unknown*.
- If the claim is separated into two types by a slash, **Claim_Type** is the first type. Values in the column **Claim_Type** must be one of 14 valid values:
 - Bus Terminal
 - Complaint
 - Compliment
 - Employee Loss (MPCECA)
 - Missed Flight
 - Motor Vehicle
 - Not Provided
 - Passenger Property Loss
 - Passenger Theft
 - Personal Injury
 - Property Damage
 - Property Loss
 - Unknown
 - Wrongful Death
- Values in the column **Claim_Site** must be one of 8 valid values:
 - Bus Station
 - Checked Baggage
 - Checkpoint
 - Motor Vehicle
 - Not Provided
 - Other
 - Pre-Check
 - Unknown
- Values in the column **Disposition** must be one of 10 valid values:
 - *Insufficient
 - Approve in Full
 - Closed:Canceled
 - Closed:Contractor Claim

- Deny
- In Review
- Pending Payment
- Received
- Settle
- Unknown
- All **StateName** values should be in proper case.
- All **State** values should be in uppercase.
- The table must include a new column named **Date_Issues** with a value of *Needs Review* to indicate that a row has a date issue. Date issues consist of the following:
 - a missing value for **Incident_Date** or **Date_Received**
 - an **Incident_Date** or **Date_Received** value out of the predefined year range of 2002 through 2017
 - an **Incident_Date** value that occurs after the **Date_Received** value
- The **County** and **City** columns should not be included in the output table.
- Currency should be permanently formatted with a dollar sign and include two decimal points.
- All dates should be permanently formatted in the style 01JAN2000.
- Permanent labels should be assigned to columns by replacing any underscore with a space.
- Final data should be sorted in ascending order by **Incident_Date**.

Report Requirements

The final single PDF report must answer the following questions:

1. How many date issues are in the overall data?

For the remaining analyses, **exclude all rows with date issues**.

2. How many claims per year of **Incident_Date** are in the overall data? Be sure to include a plot.
3. Lastly, a user should be able to dynamically input a specific state value and answer the following:
 - a. What are the frequency values for **Claim_Type** for the selected state?
 - b. What are the frequency values for **Claim_Site** for the selected state?
 - c. What are the frequency values for **Disposition** for the selected state?
 - d. What is the mean, minimum, maximum, and sum of **Close_Amount** for the selected state? The statistics should be rounded to the nearest integer.