# Overview

Your task is to prepare and analyze Transportation Security Administration (TSA) Airport Claims data from 2002 through 2017. The TSA is an agency of the United States Department of Homeland Security that has authority over the security of the traveling public. A claim is filed if you are injured, or your property is lost or damaged during the screening process at an airport. The data file is attached in the dropbox folder.

# Data Description

|  |  |
| --- | --- |
| Column | Description |
| Claim\_Number | The number for each claim. Some claims can have duplicate claim numbers but different information for each claim. Those claims are considered valid for this case study.  Any duplicate rows should be removed from the data. |
| Date\_Received | Date the claim was received. Date\_Received must occur after Incident\_Date.  Range: From 2002 through 2017 |
| Incident\_Date | Date the incident occurred. Incident\_Date must occur before Date\_Received.  Range: From 2002 through 2017 |
| Airport\_Code | Airport code three-letter abbreviation. |
| Airport\_Name | Full name of the airport. |
| Claim\_Type | Category of the claim. If the claim is separated into two types by a slash, Claim\_Type is the first type. For example, Personal Property Loss/Injury is considered Personal Property Loss.  Possible values (14):  • 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 |
| Claim\_Site | Airport location of the claim.  Possible values (8):  • Bus Station • Checked Baggage • Checkpoint • Motor Vehicle  • Not Provided • Other • Pre-Check • Unknown |
| Item\_Category | Type of items that have been filed in the claim. Depending on the year of the data, the Item\_Category values are input differently. Because of varying consistency, you do not clean this column for the case study. |
| Close\_Amount | The dollar amount a claim was closed for. |
| Disposition | The final settlement of the claim.  Possible values (10):  • \*Insufficient • Approve in Full • Closed: Canceled  • Closed: Contractor Claim • Deny • In Review  • Pending Payment • Received • Settle  • Unknown  *\*Insufficient is the value from the raw data.* |
| StateName | Associated airport state name (for example, NEW YORK).  Requirements: Values should be in all proper case. (Original data is in all uppercase.) |
| State | Associated airport state code. This is the standard two-letter abbreviation the post office uses for US states and territories (for example, IL, PR, CQ).  Requirements: Values should be in all uppercase. |
| County | Airport-associated county (or parish) name (for example, Cook). |
| City | Associated airport city name (for example, CHICAGO). |

## Data Requirements

* Import the raw data file TSAClaims2002\_2017.csv.
* The final data should be in the WORK library, and the data set should be named claims\_cleaned.
* Entirely duplicated records need to be removed from the data set.
* All missing and “-“ values in the columns Claim\_Type, Claim\_Site, and Disposition must be changed to Unknown.
* Values in the columns Claim\_Type, Claim\_Site, and Disposition must follow the requirements in the Data Description section.
* All StateName values should be in the proper case.
* All State values should be in uppercase.
* You create 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
* Remove the County and City columns.
* Currency should be permanently formatted with a dollar sign and include two decimal points.
* All dates should be permanently formatted in style 01JAN2000.
* Permanent labels should be assigned columns by replacing the underscores with a space.
* Final data should be sorted in ascending order by Incident\_Date.

## Report Requirements

The final single PDF report needs to exclude all rows with date issues in the analysis and answer the following questions:

1. How many date issues are in the overall data?
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 (this implies the use of macro variables):
   1. What are the frequency values for Claim\_Type for the selected state?
   2. What are the frequency values for Claim\_Site for the selected state?
   3. What are the frequency values for Disposition for the selected state?
   4. What is the mean, minimum, maximum, and sum of Close\_Amount for the selected state? Round to the nearest integer.

The questions should be answered with results sets of proper SAS procedures.

## Access Data

1. Import the TSAClaims2002\_2017.csv file.

## Explore Data

1. Preview the data.
2. Explore the following columns and make note of any adjustments needed using the information from the Data Description and Requirements sections above.
   1. Claim\_Site
   2. Disposition
   3. Claim\_Type
   4. Date\_Received
   5. Incident\_Date

## Prepare Data

1. Remove duplicate rows.
2. Sort the data by ascending Incident\_Date.
3. Clean the Claim\_Site column.
4. Clean the Disposition column.
5. Clean the Claim\_Type column.
6. Convert all State values to uppercase and all StateName values to proper case.
7. Create a new column to indicate date issues.
8. Add permanent labels and formats.
9. Drop County and City.

## Analyze Data

1. Analyze the overall data to answer the business questions specified in the Report Requirements section. Be sure to add appropriate titles.
2. Analyze the state-level data to answer the business questions specified in the Report Requirements section. Be sure to add appropriate titles.

## Export Data

1. Export the end results into a single PDF named ClaimReports with a style of your choice.
2. Customize the procedure labels in your report.