**Data Requirements**

**Importing Data:**

* *Import the raw data file TSAClaims2002\_2017.csv.*

**Code (Screenshot):**

Graphical user interface, text

Description automatically generated

**Code (Script):**

%let path=/home/u62734652/EPG1V2/data/;   
libname tsa "/home/u62734652/EPG1V2/data/";   
  
options validvarname=v7;   
  
/\* Import the raw data file TSAClaims2002\_2017.csv. \*/  
proc import datafile="&path/TSAClaims2002\_2017.csv"   
 dbms=csv out=tsa.ClaimsImport   
 replace;   
 guessingrows=max;   
run;

**Data Cleaning:**

* *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 layout (see attached).*
* *All StateName values should be in the proper case.*
* *All State values should be in uppercase.*

**Code (Screenshot):**

**Graphical user interface, text, application

Description automatically generated**

**Code (Script):**

/\* Entirely duplicated records need to be removed from the data set. \*/  
proc sort data=tsa.ClaimsImport   
 out=tsa.claims\_duplRem  
 nodupkey;  
 by \_all\_;  
run;  
  
/\* All missing and "-" values in the columns Claim\_Type, Claim\_Site, and Disposition must be changed to Unknown. \*/  
data tsa.claimsunknown;  
 set tsa.claims\_duplRem;  
 if Claim\_Type in ("","-") then Claim\_Type="Unknown";  
 if Claim\_Site in ("","-") then Claim\_Site="Unknown";  
 if Disposition in ("","-") then Disposition='Unknown';  
run;  
  
  
/\* All StateName values should be in the proper case. \*/  
/\* All State values should be in uppercase. \*/  
data tsa.claimsproper;  
 set tsa.claimsunknown;  
 StateName = PROPCASE (Statename);  
 State = UPCASE (State);  
run;

**Data wrangling and further cleaning:**

* *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 the 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.*

**Code(screenshot):**

Text

Description automatically generated

**Text

Description automatically generated**

**Code(script):**

/\* You create a new column named Date\_Issues with a value of Needs Review to indicate that a row has a date issue. \*/  
data tsa.claimsdissues;  
 set tsa.claimsproper;  
 if ( Incident\_Date = "" or  
 Date\_Received = "" or  
 year(Incident\_Date) < 2002 or   
 year(Incident\_Date) > 2017 or  
 year(Date\_Received) < 2002 or  
 year(Date\_Received) > 2017 or  
 Incident\_Date>Date\_Received) then Date\_Issues = "Needs Review";  
run;  
  
/\* Remove the County and City columns. \*/  
data tsa.claimsremove;  
 set tsa.claimsdissues;  
 drop County City;  
run;  
  
/\* Currency should be permanently formatted with a dollar sign and include two decimal points. \*/  
data tsa.claimscformat;  
 set tsa.claimsremove;  
 format Close\_Amount Dollar10.2;  
run;  
  
/\* All dates should be permanently formatted in the style 01JAN2000. \*/  
data tsa.claimsdformat;  
 set tsa.claimscformat;  
 format Date\_Received Incident\_Date Date9.;  
run;  
  
/\* Permanent labels should be assigned columns by replacing the underscores with a space. \*/  
data tsa.claimslabels;  
 set tsa.claimsdformat;  
 label   
 Claim\_Number = "Claim Number"  
 Date\_Received = "Date Received"  
 Incident\_Date = "Incident Date"  
 Airport\_Code = "Airport Code"  
 Airport\_Name = "Airport Name"  
 Claim\_Type = "Claim Type"  
 Claim\_Site = "Claim Site"  
 Item\_Category = "Item Category"  
 Close\_Amount = "Close Amount"  
 Date\_Issues = "Date Issues";  
run;  
  
/\* Final data should be sorted in ascending order by Incident\_Date. \*/  
/\* The final report needs to exclude all rows with date issues in the analysis \*/  
/\* The final data should be in the permanent library TSA, and the data set should be named claims\_cleaned. \*/  
proc sort data=tsa.claimslabels out=tsa.claims\_cleaned;  
 by Incident\_Date;  
run;

**Report Requirements**

*The final report needs to exclude all rows with date issues in the analysis.*

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

**Code (Screenshot):**

**Text

Description automatically generated**

**Code (Script):**

/\* The final report needs to exclude all rows with date issues in the analysis \*/  
/\* 1.How many date issues are in the overall data? \*/  
proc freq data=tsa.claims\_cleaned;  
 tables Date\_Issues;  
run;

**Output:**

Table

Description automatically generated

1. *How many claims per year of Incident\_Date are in the overall data? Be sure to include a plot.*

**Code (Screenshot):**

***Text

Description automatically generated***

**Code (Script):**

/\* 2.How many claims per year of Incident\_Date are in the overall data? Be sure to include a plot. \*/  
ods graphics on;   
title "Overall Claims by Year";   
proc freq data=tsa.claims\_cleaned;  
 tables Incident\_Date / plots=freqplot;  
 format Incident\_Date year4.;  
 where Date\_Issues is null;  
run;

**Output:**

Chart, histogram

Description automatically generated

1. *Lastly, dynamically input a specific state value and answer the following:*
   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?*

**Code (Screenshot):**

**A picture containing table

Description automatically generated**

**Code (Script):**

/\* 3.Lastly, dynamically input a specific state value and answer the following: \*/  
/\* Selected state: New York \*/  
/\* a\b\c.What are the frequency values for Claim\_Type,Claim\_Site, and desposition for the selected state? \*/  
ods graphics on;   
title"Claims by City:New York";  
%let StateName=New York;  
proc freq data=tsa.claims\_cleaned;  
 tables Claim\_Type Claim\_Site Disposition;  
 where StateName="&StateName" and   
 Date\_Issues is null;  
run;

**Output:**

**Table

Description automatically generated**

* 1. *What is the mean, minimum, maximum, and sum of Close\_Amount for the selected state? Round to the nearest integer.*

**Code (Screenshot):**

**Text

Description automatically generated with low confidence**

**Code (Script):**

/\* d.What is the mean, minimum, maximum, and sum of Close\_Amount for the selected state? Round to the nearest integer. \*/

ods graphics on;   
title"Means by City:New York";  
%let StateName=New York;  
proc means data=tsa.claims\_cleaned mean min max sum  
 maxdec=0;  
 var Close\_Amount;  
 where StateName="&StateName";  
run;

**Output:**

**Table

Description automatically generated**