```
*access data;
libname TSA "/home/u62098731/EPG194/output";
options validvarname=v7;
*import data;
proc import datafile="/home/u62098731/EPG194/output/TSAClaims2002_2017.csv" dbms=csv
out=TSA.claims_cleaned replace;guessingrows=max;
*Explore Data;
proc print data=tsa.claims_cleaned(obs=5000);
var Claim_Site Disposition Claim_Type Date_Received Incident_Date;
*Prepare Data;
*1&2 Remove duplicate rows and Sort the data by ascending Incident_Date;
proc sort data=tsa.claims_cleaned nodupkey out=tsa.nodup;by Incident_Date;
format Incident_Date Date_Received Date9.;
*3 Clean the Claim_Site column;
data tsa.claims_cleaned;
set tsa.nodup;
if Claim_Site=" " then Claim_Site="Unknown";
*4 Clean the Disposition column;
if Disposition=" " or Disposition="-" then Disposition="Unknown";
else if Disposition="Closed: Canceled" then Disposition="Closed:Canceled";
*proc print data=tsa.cleaned_data;
*var Disposition;
*5 Clean the Claim_Type column;
if Claim_Type=" " or Claim_Type="-" then Claim_Type="Unknown";
else if Claim_Type="Property Damage/Personal Injury" then Claim_Type="Property Damage";
else if Claim_Type="Passenger Property Loss/Personal Injury" then Claim_Type="Passenger Property Loss";
*6 Convert all State values to uppercase and all StateName values to proper case;
State=upcase(State);
StateName=lowcase(StateName);
*7 Create a new column to indicate date issues;
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";
*8 Add permanent labels and formats;
```

```
format Incident_Date Date_Received Date9.;
label Claim_Site="CLAIM SITE"
Disposition="DISPOSITION"
Claim_Type="CLAIM TYPE"
Date_Received="DATE RECEIVED"
Incident_Date="INCIDENT DATE";
*9 Drop County and City;
drop County City;
*Analyze Data;
*1 Analyze the overall data to answer the business questions Be sure to add appropriate titles;
*1.How many date issues are in the overall data?;
ods pdf file="/home/u62098731/EPG194/output/ClaimsReport.pdf" style=meadow;
title "Date Issues";
proc freq data=tsa.claims_cleaned;
tables Date_Issues/nocum nopercent;
run;
title;
*2 How many claims per year of Incident_Date are in the overall data? Be sure to include a plot;
title "claims per year of Incident_Date";
proc freq data=tsa.claims_cleaned;
tables Incident_Date/nocum nopercent plots=freqplot(orient=horizontal scale=percent);
format Incident_Date year.;
where Date_Issues is missing;
run;
title;
*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?;
*Round to the nearest integer;
*export data;
*1 Export the end results into a single PDF named ClaimReports that has a style of your choice;
*2. Customize the procedure labels in your report;
%let sstate=GA;
title "freq of Claim_Type,Claim_Site and Disposition";
proc freq data=tsa.claims_cleaned;
tables Claim_Type Claim_Site Disposition/nocum nopercent;
```

```
where state="&sstate" and Date_Issues is missing;
run;
title;
title "statistics for Close_Amount for the selected state";
proc means data=tsa.claims_cleaned maxdec=0 mean min max sum;
var Close_Amount;
where state="&sstate" and Date_Issues is missing;
run;
title;
ods pdf close;
```