SAS Base Programming 1:
Case Study Submission

Submitted by:
Mariam Ayman Moustafa Mahmoud

```
LIBNAME tsa '~';


proc import datafile = '/home/u63981419/EPG1V2/data/TSAClaims2002_2017.csv'
DBMS = csv
out = tsa.ClaimsImport
replace;
guessingrows = max;

run;


proc print data = tsa.claimsimport (obs = 30);
run;


proc contents data = tsa.claimsimport;
run;

/* PREPARE DATA */

/* Remove duplicate rows */
proc sort data = tsa.claimsimport NODUPKEY;
by _ALL_;
run;

/* Sort the data by ascending Incident Date */
proc sort data = tsa.claimsimport;
by Incident_Date;
run;

/* Clean the Claim_Site column */
data tsa.claims_cleaned;
set tsa.claimsimport;
IF Claim_Site IN ('-', '')
THEN Claim_Site = 'Unknown';

/* Clean the Airport_Name and Airport_Code columns */
IF Airport_Name IN ('-', '')
THEN Airport_Name = 'Unknown';

*length Airport_Code $ 5;
IF Airport_Code IN ('-', '')
THEN Airport_Code  = 'Unknown';


/* Clean the Disposition column */
IF Disposition IN ('-', '')
```

```
THEN Disposition = 'Unknown';
ELSE IF Disposition = 'losed: Contractor Claim'
THEN Disposition = 'Closed:Contractor Claim';
ELSE IF Disposition = 'Closed: Canceled'
THEN Disposition = 'Closed:Canceled';

/* Clean the Claim_Type column */
IF Claim_Type IN ('-', '')
THEN Claim_Type = 'Unknown';
ELSE IF Claim_Type IN ('Passenger Property Loss/Personal Injur', 'Passenger Property
Loss/Personal Injury')
THEN Claim_Type = 'Passenger Property Loss';
ELSE IF Claim_Type = 'Property Damage/Personal Injury'
THEN Claim_Type = 'Property Damage';

/* Convert all State values to uppercase and all StateName values to proper case */
State = UPCASE(State);
StateName = PROPCASE(StateName, ' ');
/* 7. Create a new column to indicate date issues */
IF year(Incident_Date) < 2002 OR year(Incident_Date) > 2017
   OR year(Date_Received) < 2002 OR year(Date_Received) > 2017
   OR Date_Received < Incident_Date
   OR Date_Received = '.' OR Incident_Date = '.'
THEN Date_Issues = 'Needs Review';

/* Add permanent labels and formats */
format Incident_Date Date_Received date9. Close_Amount dollar20.2;
label Airport_Code     =        'Airport Code'
            Airport_Name        =        'Airport Name'
            Claim_Number        =        'Claim Number'
            Claim_Site    =        'Claim Site'
            Claim_Type  =        'Claim Type'
            Close_Amount        =        'Close Amount'
            Date_Issues =        'Date Issues'
            Date_Recevied =      'Date Received'
            Incident_Date =      'Incident Date'
            Item_Category =      'Item Category';

/* Drop County and City columns */
drop County City;

/* Check columns to see if they are transformed properly */
proc freq data = tsa.claims_cleaned;
table Claim_Site Disposition Claim_Type Date_Issues/nocum nopercent;
run;


/* ANALYZE DATA & EXPORT INTO PDF REPORT */
```

```
ODS pdf file = '~/ClaimsReport.pdf' style = meadow pdftoc = 1;  *PDF table of content is in 1 level;
ODS noproctitle;

/* How many date issues are in the overall data? */
ODS proclabel 'Overall Date Issues';
title 'Overall Date Issues in the Data';
proc freq data = tsa.claims_cleaned;
table Date_Issues/missing nocum nopercent;
run;

/* How many claims per year of Incident_Date are in the overall data? Be sure to include a plot. */
ODS proclabel 'Overall Claims by Year';
title 'Overall Claims by Year';
proc freq data = tsa.claims_cleaned;
table Incident_Date/nocum nopercent plots = freqplot;
format Incident_Date year4.;
where Date_Issues = '';
run;


/*
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.)
*/
%let SelectedState = Texas;
%let SelectedState2 = California;

ODS proclabel "&SelectedState Claims Overview";
title "&SelectedState: Claim Types, Claim Sites and Disposition";
proc freq data = tsa.claims_cleaned;
table Claim_Type Claim_Site Disposition/nocum nopercent;
where Date_Issues = '' AND StateName = "&SelectedState";
run;

ODS proclabel "Close Amount Statistics for &SelectedState";
title "&SelectedState: Close_Amount Statistics";
proc means data = tsa.claims_cleaned MAXDEC = 0 mean min max sum;
var Close_Amount;
where Date_Issues = '' AND StateName = "&SelectedState";
run;
```

```
ODS proclabel "&SelectedState2 Claims Overview";
title "&SelectedState2: Claim Types, Claim Sites and Disposition";
proc freq data = tsa.claims_cleaned;
table Claim_Type Claim_Site Disposition/nocum nopercent;
where Date_Issues = '' AND StateName = "&SelectedState2";
run;

ODS proclabel "Close Amount Statistics for &SelectedState2";
title "&SelectedState2: Close_Amount Statistics";
proc means data = tsa.claims_cleaned MAXDEC = 0 mean min max sum;
var Close_Amount;
where Date_Issues = '' AND StateName = "&SelectedState2";
run;

ODS pdf close;
```