```
%let path="/home/u63408142/TSA Project 2/TSAClaims2002_2017.csv";
libname tsa "/home/u63408142/TSAClaims2002_2017";
options validvarname=v7;

proc import datafile="/home/u63408142/TSA Project 2/TSAClaims2002_2017.csv"
dbms=csv
out=tsa.ClaimsImport
replace;
guessingrows=max;
run;

/*preview to print the first 20 observations of the dataset*/
proc print data=tsa.ClaimsImport (obs=20);
run;

/*print the content of the data*/
proc contents data=tsa.ClaimsImport varnum;
run;

/*Exploring columns & making note of adjustments and Requirements.*/
proc freq data=tsa.ClaimsImport;
    tables Claim_Site Disposition Claim_Type / nocum nopercent;
    tables Date_Received Incident_Date / nocum nopercent;
    format Date_Received Incident_Date year4.;
run;

/*remove dups andstore them in Claims_NoDups*/
proc sort data=tsa.ClaimsImport out=tsa.Claims_NoDups nodupkey;
    by _all_;
    run;

/*EXPLORING FINISHED*/
/*sort the claimsnodups*/
proc sort data=tsa.Claims_NoDups;
    by Incident_Date;
    run;


/*create claimscleans and copy the edition of the data in it*/
/*a - Change missing values and "-" to Unknown on Claim_Site */
data tsa.Claims_Cleans;
    set tsa.Claims_NoDups;

    keep Claim_Number Date_Received Incident_Date Airport_Code Airport_Name
        Claim_Type Claim_Site Item_Category Close_Amount Disposition StateName State
        County City;

    if Claim_Site = "" OR Claim_Site = "-" then Claim_Site = "Unknown";
run;

/*c1 - Change missing values and "-" to Unknown on Disposition */
/*c2 - Remove the leading space in Closed: Canceled.*/
/*c3 - Fix the missing character C and the leading space in losed: Contractor Claim.*/
data tsa.Claims_Cleans;
    set tsa.Claims_Cleans;

    keep Claim_Number Date_Received Incident_Date Airport_Code Airport_Name
        Claim_Type Claim_Site Item_Category Close_Amount Disposition StateName State
        County City;

    if Disposition = " " OR Disposition = "-" then Disposition = "Unknown";
    else if Disposition = "Closed: Canceled" then Dispositon = "Closed:Canceled";
    else if Disposition = "losed: Contractor Claim" then Dispositon = "Closed:Contractor Claim";
run;

/*b1 - Change missing values and "-" to Unknown on Claim_Type & */
/*b2 - Change Passenger Property Loss/Injury values to Passenger Property Loss.*/
/*b3 - Change Property Damage/Personal Injury values to Property Damage.*/
data tsa.Claims_Cleans;
    set tsa.Claims_Cleans;

    keep Claim_Number Date_Received Incident_Date Airport_Code Airport_Name
        Claim_Type Claim_Site Item_Category Close_Amount Disposition StateName State
        County City;

    if Claim_Type = " " OR Claim_Type = "-" then Claim_Type = "Unknown";
    else if Claim_Type = "Passenger Property Loss/Injury"  then Claim_Type= "Passenger Property Loss";
    else if Claim_Type = "Property Damage/Personal Injury" then Claim_Type= "Property Damage";
run;

/*6. Convert all State values to uppercase and all StateName values to proper case.*/
data tsa.Claims_Cleans;
    set tsa.Claims_Cleans;

    keep Claim_Number Date_Received Incident_Date Airport_Code Airport_Name
        Claim_Type Claim_Site Item_Category Close_Amount Disposition StateName State
        County City;
```

```
    State=upcase(state);
    StateName=propcase(StateName);
run;


/*7. Create a new column that indicates date issues.
d1 - Notice that many dates are after the year 2017 or missing on Date_Received*/
data tsa.Claims_Cleans;
    set tsa.Claims_Cleans;

    if (Incident_Date > Date_Received or 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) then Date_Issues="Needs Review";
run;


/*8. Add permanent labels and formats.*/
data tsa.Claims_Cleans;
    set tsa.Claims_Cleans;

    keep Claim_Number Date_Received Incident_Date Airport_Code Airport_Name
        Claim_Type Claim_Site Item_Category Close_Amount Disposition StateName State
        County City Date_Issues;

    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_Received="Date Received"
    Incident_Date="Incident Date"
    Item_Category="Item Category";
    run;

/*9. Drop County and City.*/
data tsa.Claims_Cleans;
    set tsa.Claims_Cleans;
    drop County City;
    run;

/*PREPARING FINISHED*/

*EXPORTING;
ods pdf file="/home/u63408142/TSAClaims2002_2017/Report.pdf" style=journal startpage=no pdftoc=1;
ods noproctitle;

*ANALYZING;
title "Overall Date Issues in the Data";

proc freq data=tsa.Claims_Cleans;
    table Date_Issues / nocum nopercent;
run;

title;
ods graphics on;
title "Overall Claims by Year";

proc freq data=tsa.Claims_Cleans;
    table Incident_Date / nocum nopercent plots=freqplot;
    format Incident_Date year4.;
    where Date_Issues is null;
run;



title;
%let StateName=California;
title "&StateName Claim Types, Claim Sites and Disposition
Frequencies";

proc freq data=tsa.Claims_Cleans order=freq;
    table Claim_Type Claim_Site Disposition / nocum nopercent;
    where StateName="&StateName" and Date_Issues is null;
run;

title "Close_Amount Statistics for &StateName";

proc means data=tsa.Claims_Cleans mean min max sum maxdec=0;
    var Close_Amount;
    where StateName="&StateName" and Date_Issues is null;
run;

title;
```

```
*exporting;
title;
ods pdf close;
```