```
/* 1. Accessing Data */

%let path=/home/u41140628/EPG194/ECRB94/data;
%let statename=North Carolina;
libname tsa "&path";
options validvarname=7;
proc import datafile="&path/TSAClaims2002_2017.csv" dbms=csv out=tsa.claimsreport replace;
        guessingrows=max;
run;

/* 2. Exploring data */

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

/* Better understanding of tables and columns */

proc contents data=tsa.claimsreport;
run;

/* Observation, some of the dates are formatted as Best 12 which needs t0 be changed (Prepare data
stage) */

/* Explore categorical variables using frequency procedure */

proc freq data=tsa.claimsreport ;
        tables claim_site disposition claim_type Date_Received incident_date /nocum nopercent;
        format Date_Received incident_date year4.;
run;
```

```
/* Observation, checking disposition column, missing values and hiphen (-), also spelling issues in
category */


proc print data=tsa.claimsreport;
        where Date_Received<incident_date;
        format Date_Received incident_date date9.;
run;


/* 3. Preparing Data */


/* Remove duplicate rows  (From log 5 duplicate observations where deleted)*/
proc sort data=tsa.claimsreport out=tsa.claims_nodups noduprecs;
        by _all_;
run;
/* Sort the data by ascending incident date */
proc sort data=tsa.claims_nodups;
        by incident_date;
run;


data tsa.claims_cleaned;
        set tsa.claims_nodups;
/* Clean the claim site column */
        if claim_site in ('-','') then claim_site="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='Passenger Property Loss/Personal Injur' then claim_type='Passenger
Property Loss';
                else if claim_type='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 state name values to proper case */
        state=upcase(state);
        statename=propcase(statename);
/* create a new column to indicate date issue */
        if (incident_date > date_received or
        date_received=. or incident_date =. 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";
/* Add permenant labels and formats */
        format incident_date date_received date9. close_amout 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";
/* Drop county and city */
        drop county city;
```

```
run;

/* Check if the changes are done properly (frequence procedure) */

proc freq data=tsa.claims_cleaned order=freq ;
        tables claim_site disposition claim_type Date_issues /nocum nopercent;
run;

/* 4. Analyzing data*/

%let outpath=/home/u41140628/EPG194/ECRB94/data;
ods graphics on;
ods pdf file="&outpath/ClaimsReport.pdf" style=meadow pdftoc=1;
ods noproctitle;

/* How may Date issues are there in overall data */
ods proclabel "Overall Date issues";
title "Overall Date issues in the data";
proc freq data=tsa.claims_cleaned;
        tables date_issues /missing nocum nopercent;
run;
title;
/* How may claims per year of incident date are there in overall data with a plot */

ods proclabel "Overall claims by year";
title "Overall claims by year";
proc freq data=tsa.claims_cleaned;
        tables incident_date / nocum nopercent plots=freqplot;
        format incident_date year4.;
```

```
        where date_issues is null;
run;
title;

/* Specific state analysis */
/* A user should be able to dynamically input a specific state value and below questions */
/* 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 */

ods proclabel "&statename claim Overview";
title "&statename claim types,claim sites and disposition";
proc freq data=tsa.claims_cleaned order =freq;
        tables claim_type claim_site disposition;
        where statename="&statename" and date_issues is null;
run;
title;

/* Observation: */

/* d. what is the mean, minimum, maximum and sum of closed amount for the selected state */
ods proclabel "&statename close amount statistics";
title "&statename claim types,claim sites and disposition";
proc means data=tsa.claims_cleaned mean min max sum maxdec=0;
        var close_amount;
        where statename="&statename" and date_issues is null;
run;
title;
```

```
/* 5. Export to pdf*/


ods pdf close;
```