

Index of pages

Catalog

Portfolio Case Study 1.pdf	1
Code_ Case Study 1.sas.pdf	3
ClaimsReport.pdf	5
Total Number of Date Issues	5
Table Date_Issues	5
Number of Claims per year	6
Table Incident_Date	6
Frequency of Claim Type per state	7
Table Claim_Type	7
Frequency of Claim Site per state	8
Table Claim_Site	8
Frequency of Disposition per state	9
Table Disposition	9
Means	10
Summary statistics	10

Portfolio Case Study 1

SAS Programming

Completed by Jonathan Stiglingh

Objective of portfolio

The objective of the following code and output is to showcase my understanding of the basics of SAS programming for the purposes of hiring managers for verification purposes only. Please note that I do know more than what was required for this Case Study, however this was the first case study I decided to complete for my portfolio.

```

1  /***** FIRST STEP *****/
2
3  /* Create library and filename */
4
5  libname tsa '/folders/myfolders/case study/';
6  filename study '/folders/myfolders/case study/TSAClaims2002_2017.csv';
7
8  /* Import the data */
9  proc import datafile=study dbms=csv out=tsa.clean replace;
10     guessingrows=max;
11     getnames=Yes;
12 run;
13
14 /* Preview the data*/
15 proc print data=tsa.Clean (obs=20);
16 run;
17
18 proc contents data=tsa.Clean varnum;
19 run;
20
21
22 /* Sort and remove duplicates */
23 proc sort data=tsa.clean noduprecs;
24     by _all_;
25 /* format the dates I did this here to help with the date calculations in the Data step*/
26     format Date_Received Incident_Date Date9.;
27 run;
28
29 /* Change missing and "-" values to Unknown */
30 data cleaning;
31     set tsa.clean;
32     if Claim_Type in ("-", " ", "") then Claim_Type = 'Unknown';
33     if Claim_Site in ("-", " ", "") then Claim_Site = 'Unknown';
34     if Disposition in ("-", " ", "") then Disposition = 'Unknown';
35 /* Use only the first word */
36     Claim_Type=scan(Claim_Type,1,"/");
37 /* Correct Type/Spelling errors */
38     if Disposition = "Closed: Canceled" then Disposition = "Closed:Canceled";
39     if Disposition = "losed: Contractor Claim" then Disposition = "Closed:Canceled";
40 /* Change the Case of the text */
41     StateName=Propcase(StateName);
42     State=Propcase(State);
43
44
45 /* Create new coloumn to identify date issues */
46     IYear = Year(Incident_Date);
47     DYear = Year(Date_Received);
48     if Incident_Date =. then Date_Issues="Needs Review";
49     if Date_Received =. then Date_Issues="Needs Review";
50     if IYear <2002 or IYear >2017 then Date_Issues="Needs Review2";
51     if DYear <2002 or DYear >2017 then Date_Issues="Needs Review2";
52     IF Incident_Date > Date_Received then Date_Issues="Needs Review2";
53
54 /* Drop Unnecessary columns */
55     Drop IYear DYear County City;
56
57 /* Format currency to be dollar with 2 decimal */
58     format Close_Amount Dollar12.2;
59
60 /* Adding Labels */
61     Label Claim_Number = "Claim Number"
62           Date_Received = "Date Received"
63           Incident_Date = "Incident Date"
64           Airport_Code = "Airport Code"
65           Airport_Name = "Airport Name"
66           Claim_Type = "Claim Type"
67           Claim_Site = "Claim Site"
68           Item_Category = "Item Category"
69           Close_Amount = "Close Amount"
70           Disposition = "Disposition"
71           StateName = "State Name"
72           State = "State"
73           Date_Issues = "Date Issues";
74
75 run;
76
77
78 proc sort data=cleaning;
79     by Incident_Date;
80 run;
81
82 /* Good Practise to clear libname and filename */
83

```

```

84 /* libname tsa clear; */
85 /* filename study clear; */
86
87
88
89 /* Error checking */
90
91 /* Displays data set descriptor information variables and permanent labels */
92 proc contents data=cleaning;
93 run;
94
95 Data Test1;
96     Set cleaning;
97     where IYear > DYear;
98 run;
99
100
101 proc sort data=cleaning out=test;
102     by Date_Issues;
103 run;
104
105 proc freq data=cleaning ;
106     table Date_Issues IYear;
107 run;
108
109
110 /***** SECOND STEP *****/
111
112
113
114 proc sort data=cleaning out=test;
115     by Date_Issues;
116     where Date_Issues is missing;
117 run;
118
119
120 %let outpath=/folders/myfolders/case study/;
121 ods pdf file="&outpath\ClaimsReport.pdf" style=Meadow bookmarkgen=Yes;
122 /* ods proclabel "" */
123
124 Title "Total Number of Date Issues";
125 ods proclabel "Total Number of Date Issues";
126 options nodate; /*removes the date*/
127 ODS NOPROCTITLE;
128
129
130 proc freq data=cleaning;
131     table Date_Issues /nocum nopercnt ;
132 run;
133
134 Title "Number of Claims per year";
135 ods proclabel "Number of Claims per year";
136
137 proc freq data=test;
138     table Incident_Date /nocum nopercnt plots=freqplot; /*this can also be done using the Proc SGPLOTS step*/
139     format Incident_Date year.;
140     where '01Jan2002'd <= Incident_Date <= '31Dec2017'd;
141 run;
142
143 %let State="California"; /*this is the macro to use to substitute the State value in the code below*/
144
145 Title "Frequency of Claim Type per state";
146 ods proclabel "Frequency of Claim Type per state";
147
148 proc freq data=test order=freq;
149     table Claim_Type /nocum nopercnt;
150     where StateName=&State and '01Jan2002'd <= Incident_Date <= '31Dec2017'd;
151 run;
152
153 Title "Frequency of Claim Site per state";
154 ods proclabel "Frequency of Claim Site per state";
155
156 proc freq data=test order=freq;
157     table Claim_Site /nocum nopercnt;
158     where StateName=&State and '01Jan2002'd <= Incident_Date <= '31Dec2017'd;
159 run;
160
161 Title "Frequency of Disposition per state";
162 ods proclabel "Frequency of Disposition per state";
163
164 proc freq data=test order=freq;
165     table Disposition /nocum nopercnt;
166     where StateName=&State and '01Jan2002'd <= Incident_Date <= '31Dec2017'd;
167 run;

```

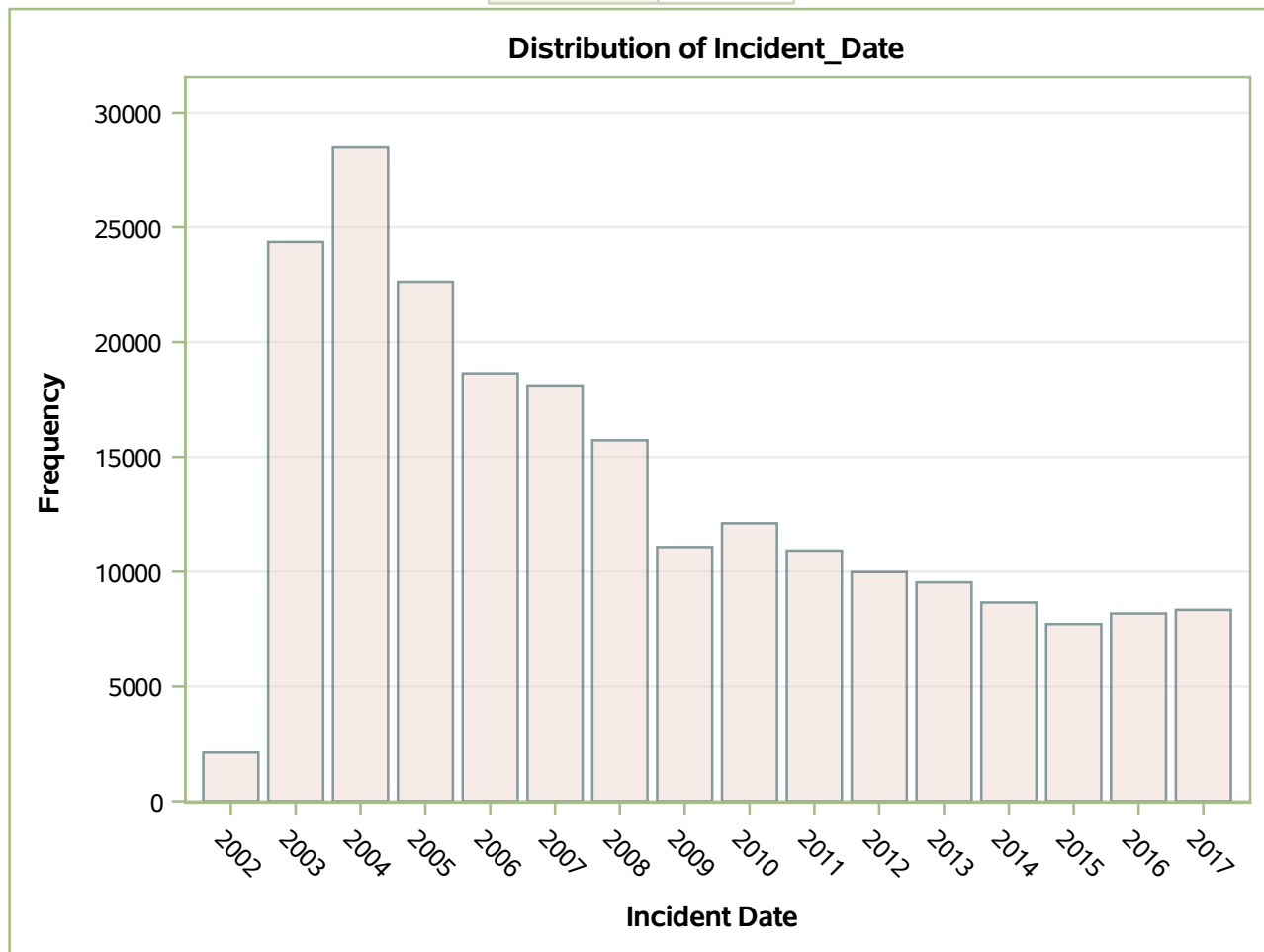
Total Number of Date Issues

Date Issues	
Date_Issues	Frequency
Needs Review	4241
Frequency	
Missing = 216609	

Number of Claims per year

2

Incident Date	
Incident_Date	Frequency
2002	2123
2003	24359
2004	28484
2005	22631
2006	18643
2007	18116
2008	15727
2009	11075
2010	12108
2011	10921
2012	9984
2013	9536
2014	8659
2015	7721
2016	8182
2017	8340



Frequency of Claim Type per state

Claim Type	
Claim_Type	Frequency
Passenger Property Loss	14892
Property Damage	8996
Unknown	756
Personal Injury	194
Passenger Theft	55
Employee Loss (MPCECA)	51
Motor Vehicle	17
Complaint	7
Compliment	2
Missed Flight	1
Property Loss	1

Frequency of Claim Site per state

Claim Site	
Claim_Site	Frequency
Checked Baggage	19553
Checkpoint	5142
Other	187
Unknown	68
Motor Vehicle	19
Pre-Check	2
Bus Station	1

Frequency of Disposition per state

5

Disposition	
Disposition	Frequency
Deny	10918
Approve in Full	5266
Settle	4192
Unknown	3188
In Review	1086
*Insufficient	187
Closed:Canceled	91
Closed:Contractor Claim	41
Received	3

Frequency of Disposition per state

6

Analysis Variable : Close_Amount Close Amount			
Mean	Minimum	Maximum	Sum
98	0	14519	2096386