SAS Project 1

Creating an Analysis Report on USA TSA Claims Dataset using SAS.

***Pooja More**, Data Analyst Intern, HSBC*
***PS ID :** 45115139*
***Email ID:** pooja.more@praxis.ac.in*

A. Business Objective :

Prepare and analyze the claims data of USA Transportation Security Administration (TSA).

B. Business Scenario:

The TSA is an agency of the USA Department of Homeland Security that has the authority over the security of the travelling public. Claims are filed if the travelers are injured or their property is lost or damaged during the screening process at an airport. We have data for claims filed between 2002 and 2017. We need to analyze and report on the overall data and also by any specified state.

C. Data Overview:

Data Source: https://www.kaggle.com/
Created from publicly available data from the TSA and Federal Aviation Administration(FAA). The TSA data has information about claims and the FAA data has information about USA airport facilities.
No. of Columns : 14
No. of Rows : 220855
Details of Columns:

- Claims_Number – Number for each claim. Some claims have duplicate numbers but different information for each claim. Those claims are still considered valid.
- Incident_Date – Date the incident occurred.
- Date_Received – Date the claim was filed.
- Claim_Type – 14 valid claim types.
- Claim_Site – 8 valid claim sites.
- Item_Category – Type of item in the claim. Keeps on updating every year, hence not used in the analysis.

- Disposition – Final settlement of the claim.
- Close_Amount – Dollar amount of the settlement.
- Airport_Code, Airport_Name – where the incident occurred.
- County, City – Location of the airport
- State – 2 letter state code.
- Statename – Full statename.

Obs	Claim_Number	Date_Received	Incident_Date	Airport_Code	Airport_Name	Claim_Type	Claim_Site	Item_Category	Close_Amount	Disposition	StateName	State	County	City
1	2006081611123	17027	17006			Passenger Property Loss	Checked Baggage	Candles - Decorative and other; Clothing - Shoes, belts, accessories, etc.; Dishes, Pottery, Glassware, Plasticware	.					
2	2006062108380	16972	16957			Passenger Property Loss	Checkpoint	Jewelry - Fine	.					
3	2006062008258	16972	16938			Passenger Property Loss	Checked Baggage	Cosmetics - Perfume, toilet articles, medicines, soaps, etc.; Medicines	.					
4	2006010699056	16831	16793			Passenger Property Loss	Checked Baggage	Other	.					
5	2006032303625	16880	16861			Property Damage	Checked Baggage	Luggage (all types including footlockers)	.					
6	2006081511005	17021	.			Property Damage	Checked Baggage	Locks; Luggage (all types including footlockers)	.					
7	2006082111318	17029	17015			Property Damage	Checked Baggage	Locks	.					
8	2006070709101	16988	16979			Passenger Property Loss	Checkpoint	Jewelry - Fine	.					
9	2006072109924	17003	16996			Passenger Property Loss	Checkpoint	DVD/CD Players; Other	.					
10	2006041704637	16898	.			Passenger Property Loss	Checked Baggage	Eyeglasses - (including contact lenses)	.					
11	2006081010689	17016	16866			Property Damage	Checked Baggage	Luggage (all types including footlockers)	.					
12	2006042505214	16905	16896			Property Damage	Checkpoint	Cell Phones	.					
13	2006061408002	16966	.					Luggage (all types including footlockers)	.					
14	2006081110819	17022	17014			Passenger Property Loss	Checked Baggage	Other	.					
15	2006051806567	16937	16924			Property Damage	Checked Baggage	Electrical and Gas Appliances Minor - $200 or less (humidifiers, tv's, etc)	.					
16	2006040504066	16877	16855			Passenger Property Loss	Checked Baggage	Other	.					
17	2006073110251	17008	17001			Passenger Property Loss	Checked Baggage	Clothing - Shoes, belts, accessories, etc.	.					
18	2006042505160	16909	16863			Passenger Property Loss	Checkpoint	Eyeglasses - (including contact lenses)	.					
19	2006020801205	16834	16828			Passenger Property Loss	Checked Baggage	Cameras - Digital	0	Deny				

D. Case Study Process:

1) Access Data

To access the dataset in the required tool (SAS Studio in this case)

- Use of import procedure to access the .csv file of the dataset in the SAS studio.

2) Explore Data

To understand what I am working with and what are the inconsistencies, issues in the data I need to fix or resolve.

- Print procedure to have a quick snap shot of the dataset and the columns.
- Contents procedure to get a closer look at the properties of the columns.
- Frequency procedure to properly understand the categorical variables and also making use of format statement to understand the year wise distribution of the claims.

- Use of where conditions to understand the faults and anomalies in the data (e.g. Date_Received < Incident_Date).

3) Prepare Data

To actually resolve the issues using various concepts so that the data is validated as per the requirements. This is the most time consuming and important process to create accurate and precise reports.

- Use of sort procedure to remove the duplicate rows and sort the data in ascending order of the Incident_Date.
- Data step to prepare the data as required.
- Use of 'if-then-else', 'if-then do' to clean the columns, to create required new columns, to convert misspelled values to valid values.
- Use of 'format' and 'label' statements to uniformly format the values and the label of the columns.
- Use of 'keep' and 'drop' statements to keep the required columns and drop the unrequired or nonessential columns.

4) Analyze Data and Report Preparation

To answer the business questions, create summary statistics and visualizations.

- Frequency procedure to finally analyze the well-prepared data.
- Use of 'title' and 'footnote' to make the analysis results easy to read and understand.
- Use of 'ods graphics' and 'freqplot' to graphically plot the findings in the report.
- Finally use of macros to make the report more dynamic and user friendly.

5) Explore Results

To create a presentable, readable business report.

- Use of 'ods pdf' to export the end results into a single PDF file that has a style of our choice.
- 'ods noproctitle' and 'ods proclabel' to customize the procedure labels in the report.

Kindly refer to 'SAS Project 1 Code' file for SAS code.

Note: *The following pages is the final analysis report.*

Overall date issues in the data

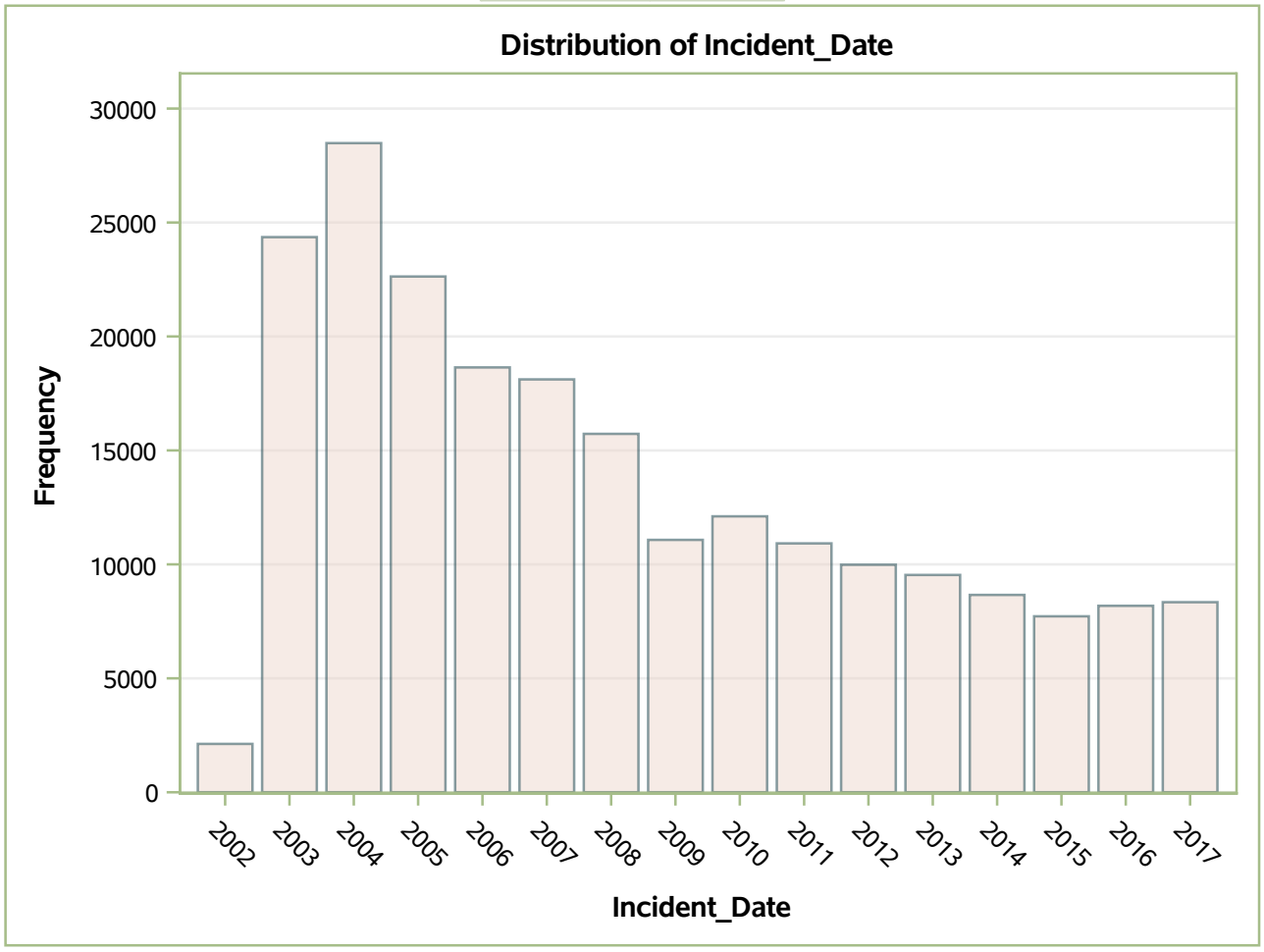
date_issues	Frequency
	216609
Needs Review	4241

Overall states in the data

StateName	Frequency
Alabama	755
Alaska	1637
American Samoa	1
Arizona	6494
Arkansas	537
California	25306
Colorado	5457
Connecticut	1146
Delaware	9
Dist. Of Columbia	6033
Florida	27966
Georgia	7234
Guam	126
Hawaii	4425
Idaho	543
Illinois	9468
Indiana	1416
Iowa	513
Kansas	390
Kentucky	1890
Louisiana	2119
Maine	408
Maryland	3141
Massachusetts	4601
Michigan	4088
Minnesota	3601
Mississippi	509
Missouri	2078
Montana	687
N Mariana Islands	14
Nebraska	931
Nevada	6469
New Hampshire	707
New Jersey	7654
New Mexico	1062
New York	17800
North Carolina	4082
North Dakota	187
Ohio	2816
Oklahoma	1005
Oregon	2649
Pennsylvania	6928
Puerto Rico	1741
Rhode Island	727
South Carolina	1420
South Dakota	299
Tennessee	2606
Texas	14377
Utah	1601
Vermont	338
Virgin Islands	481
Virginia	1928
Washington	6953
West Virginia	137
Wisconsin	1527
Wyoming	164
Frequency Missing = 11669	

Overall claims by year

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



Hawaii Claim_type claim_site & disposition

Claim_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Passenger Property Loss	2762	63.16	2762	63.16
Property Damage	1481	33.87	4243	97.03
Unknown	70	1.60	4313	98.63
Personal Injury	35	0.80	4348	99.43
Employee Loss (MPCECA)	13	0.30	4361	99.73
Passenger Theft	7	0.16	4368	99.89
Complaint	2	0.05	4370	99.93
Motor Vehicle	2	0.05	4372	99.98
Missed Flight	1	0.02	4373	100.00

Claim_Site	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Checked Baggage	3241	74.11	3241	74.11
Checkpoint	1099	25.13	4340	99.25
Other	20	0.46	4360	99.70
Unknown	11	0.25	4371	99.95
Motor Vehicle	2	0.05	4373	100.00

Disposition	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Deny	2030	46.42	2030	46.42
Approve in Full	1017	23.26	3047	69.68
Settle	674	15.41	3721	85.09
Unknown	401	9.17	4122	94.26
In Review	207	4.73	4329	98.99
*Insufficient	34	0.78	4363	99.77
Closed:Canceled	10	0.23	4373	100.00