TSA CLAIMS PROJECT

DATA PREPARATION AND ANALYSIS

AGENDA

Introduction

Project Objectives

Data Preparation Process

Data Cleaning Process

Analysis and Statistical Summary

Project Outcomes

PROJECT INTRODUCTION

THE TRANSPORTATION SECURITY ADMINISTRATION (TSA) PLAYS A CRITICAL ROLE IN ENSURING THE SAFETY OF TRAVELERS ACROSS U.S. AIRPORTS. WHEN PROPERTY IS LOST, DAMAGED, OR INDIVIDUALS ARE INJURED DURING SECURITY SCREENINGS, CLAIMS ARE FILED. THIS PROJECT FOCUSES ON PREPARING AND ANALYZING TSA AIRPORT CLAIMS DATA FROM 2002 TO 2017 TO SUPPORT INFORMED DECISION-MAKING AND IMPROVE TRANSPARENCY IN CLAIM MANAGEMENT PROCESSES.

PROJECT OBJECTIVES

- **Data Preparation**: Clean and transform raw data from TSAClaims2002_2017.csv by addressing missing values, duplicates, and formatting inconsistencies.
- **Data Quality Assurance**: Ensure compliance with technical requirements, such as correcting column values (e.g., Claim_Type, StateName), handling date discrepancies, and removing redundant columns (County, City).
- **Standardization**: Apply permanent formatting for dates (e.g., *01JAN2000*), currency (e.g., $100.00), and column labels to enhance readability.
- **Dynamic Reporting**: Generate a PDF report that provides an overview of claim trends and a detailed analysis for a user-specified state.

DATA PREPARATION PROCESS

Import the Data

Removing Duplicates

Handling Missing Values

IMPORT THE DATA

```
* Initiate and setup a Libirary ;
libname TSA "/home/u63766878/TSA Project";

* Import the Data File ;
options validvarname=v7;

proc import datafile="//home/u63766878/TSA Project/TSAClaims2002_2017.csv"
    dbms=csv
    out= tsa.ClaimsImport
    replace;
    guessingrows=max;
run;
```

```
* Reviewing the Data and Info about it ;
proc print data=tsa.ClaimsImport (obs=20);
run;
proc contents data=tsa.ClaimsImport;
run;
```

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	2.00608E+12	17027	17006			Passenger Property Loss	Checked Baggage	Candles - Decorative and other; Clothing - Shoes, belts, accessories, etc.; Dishes, Pottery, Glassware, Plasticware	.					
2	2.00606E+12	16972	16957			Passenger Property Loss	Checkpoint	Jewelry - Fine	.					
3	2.00606E+12	16972	16938			Passenger Property Loss	Checked Baggage	Cosmetics - Perfume, toilet articles, medicines, soaps, etc.; Medicines	.					
4	2.00601E+12	16831	16793			Passenger Property Loss	Checked Baggage	Other	.					
5	2.00603E+12	16880	16861			Property Damage	Checked Baggage	Luggage (all types including footlockers)	.					
6	2.00608E+12	17021	.			Property Damage	Checked Baggage	Locks; Luggage (all types including footlockers)	.					
7	2.00608E+12	17029	17015			Property Damage	Checked Baggage	Locks	.					
8	2.00607E+12	16988	16979			Passenger Property Loss	Checkpoint	Jewelry - Fine	.					
9	2.00607E+12	17003	16996			Passenger Property Loss	Checkpoint	DVD/CD Players; Other	.					
10	2.00604E+12	16898	.			Passenger Property Loss	Checked Baggage	Eyeglasses - (including contact lenses)	.					

The CONTENTS Procedure

Data Set Name	TSA.CLAIMSIMPORT	Observations	220855
Member Type	DATA	Variables	14
Engine	V9	Indexes	0
Created	04/24/2025 00:19:18	Observation Length	1072
Last Modified	04/24/2025 00:19:18	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1811
First Data Page	1
Max Obs per Page	122
Obs in First Data Page	118
Number of Data Set Repairs	0
Filename	/home/u63766878/TSA Project/claimsimport.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	11177273502
Access Permission	rw-r--r--
Owner Name	u63766878
File Size	227MB
File Size (bytes)	237502464

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
4	Airport_Code	Char	3	$3.	$3.
5	Airport_Name	Char	48	$48.	$48.
14	City	Char	33	$33.	$33.
1	Claim_Number	Char	11	$11.	$11.
7	Claim_Site	Char	15	$15.	$15.
6	Claim_Type	Char	39	$39.	$39.
9	Close_Amount	Num	8	BEST12.	BEST32.
13	County	Char	20	$20.	$20.
2	Date_Received	Num	8	BEST12.	BEST32.
10	Disposition	Char	23	$23.	$23.
3	Incident_Date	Num	8	BEST12.	BEST32.
8	Item_Category	Char	834	$834.	$834.
12	State	Char	2	$2.	$2.
11	StateName	Char	17	$17.	$17.

REMOVE DUPLICATES

```
20 * Removing the Duplicates and save into new file ;
21 proc sort data=tsa.ClaimsImport
22     OUT = tsa.claims_nodups
23     nodupkey;
24     by _ALL_;
25 run;
```

```
NOTE: There were 220855 observations read from the data set TSA.CLAIMSIMPORT.
NOTE: 209 observations with duplicate key values were deleted.
NOTE: The data set TSA.CLAIMS_NODUPS has 220646 observations and 14 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time               1.09 seconds
      user cpu time           0.25 seconds
      system cpu time         0.45 seconds
      memory                  486249.37k
      OS Memory               509996.00k
      Timestamp               04/23/2025 10:33:44 PM
      Step Count                        42  Switch Count  5
      Page Faults                       0
      Page Reclaims                     119248
      Page Swaps                        0
      Voluntary Context Switches        6239
      Involuntary Context Switches      9
      Block Input Operations            0
      Block Output Operations           463120


74
75         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
85
```

HANDLING MISSING VALUES

```
32 * Handling Missing Values ;
33 data tsa.claims_cleaned;
34     set tsa.claims_nodups;
35     if missing(Claim_Type) or Claim_Type='-' then Claim_type="Unknown";
36     if missing(Claim_Site) or  Claim_Site='-' then Claim_Site="Unknown";
37     if missing(Disposition) or Disposition='-' then Disposition="Unknown";
38 run;
39
40
41
42
```

Date_Received	Incident_Date	Airport_Code	Airport_Name	Claim_Type	Claim_Site	Item_Category	Close_Amount	Disposition	StateName	State	County	City
16072	16074	BWI	Baltimore/Washington Intl Thurgood Marshall	Passenger Property Loss	Checked Baggage	Clothing - Shoes, belts, accessories, etc.	0	Unknown	MARYLAND	MD	ANNE ARUNDEL	BALTIMORE
15707	15694	SJC	Norman Y Mineta San Jose International	Unknown	Other	Clothing - Shoes, belts, accessories, etc.	0	Unknown	CALIFORNIA	CA	SANTA CLARA	SAN JOSE
15707	15681			Passenger Property Loss	Checked Baggage	Clothing - Shoes, belts, accessories, etc.	0	Unknown				
15707	15645	PHL	Philadelphia International Airport	Passenger Property Loss	Other	Jewelry - Fine	0	Unknown	PENNSYLVANIA	PA	PHILADELPHIA	PHILADELPHIA
15707	15691	SAN	San Diego International	Unknown	Other	Clothing - Shoes, belts, accessories, etc.	0	Unknown	CALIFORNIA	CA	SAN DIEGO	SAN DIEGO
15707	15684	ORF	Norfolk International	Property Damage	Other	Luggage (all types including footlockers)	.	Unknown	VIRGINIA	VA	NORFOLK	NORFOLK
16075	16024	JAX	Jacksonville International	Unknown	Checked Baggage	Other	70	Settle	FLORIDA	FL	DUVAL	JACKSONVILLE
16075	16044	HNL	Honolulu International Airport	Unknown	Checked Baggage	Other	64.98	Settle	HAWAII	HI	HONOLULU	HONOLULU
16075	16033	ABQ	Albuquerque International Sunport Airport	Unknown	Checked Baggage	Jewelry - Fine	0	Deny	NEW MEXICO	NM	BERNALILLO	ALBUQUERQUE
16075	16021	LGA	LaGuardia	Unknown	Checkpoint	Other	94	Approve in Full	NEW YORK	NY	QUEENS	NEW YORK
16075	16049	ORD	Chicago O'Hare International Airport	Unknown	Checked Baggage	Cell Phones	100	Settle	ILLINOIS	IL	COOK	CHICAGO
16075	16044	RIC	Richmond International	Unknown	Checked Baggage	Clothing - Shoes, belts,	0	Unknown	VIRGINIA	VA	HENRICO	RICHMOND

FREQUENCY BEFORE CLEANING PROCESS

```
44
45 * Reviewing the Freq from some columns before the Cleaning Process ;
46 proc freq data=tsa.claims_cleaned;
47     tables Claim_Type Claim_Site Disposition;
48 run;
49
```

The FREQ Procedure

Claim_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Bus Terminal	1	0.00	1	0.00
Complaint	73	0.03	74	0.03
Compliment	3	0.00	77	0.03
Employee Loss (MPCECA)	492	0.22	569	0.26
Missed Flight	32	0.01	601	0.27
Motor Vehicle	418	0.19	1019	0.46
Not Provided	2	0.00	1021	0.46
Passenger Property Loss	126720	57.43	127741	57.89
Passenger Property Loss/Personal Injur	8	0.00	127749	57.90
Passenger Property Loss/Personal Injury	13	0.01	127762	57.90
Passenger Theft	479	0.22	128241	58.12
Personal Injury	1605	0.73	129846	58.85
Property Damage	82549	37.41	212395	96.26
Property Damage/Personal Injury	12	0.01	212407	96.27
Property Loss	17	0.01	212424	96.27
Unknown	8218	3.72	220642	100.00
Wrongful Death	4	0.00	220646	100.00

Disposition	Frequency	Percent	Cumulative Frequency	Cumu Pe
*Insufficient	1730	0.78	1730	
Approve in Full	48948	22.18	50678	
Closed: Canceled	167	0.08	50845	
Closed:Canceled	283	0.13	51128	
Closed:Contractor Claim	115	0.05	51243	
Deny	99522	45.10	150765	
In Review	8922	4.04	159687	
Pending Payment	1	0.00	159688	
Received	14	0.01	159702	
Settle	34417	15.60	194119	
Unknown	26454	11.99	220573	
losed: Contractor Claim	73	0.03	220646	1

Claim_Site	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Bus Station	20	0.01	20	0.01
Checked Baggage	171387	77.68	171407	77.68
Checkpoint	44730	20.27	216137	97.96
Motor Vehicle	528	0.24	216665	98.
Not Provided	1	0.00	216666	98.
Other	2911	1.32	219577	99.
Pre-Check	8	0.00	219585	99.
Unknown	1061	0.48	220646	100.

DATA CLEANING PROCESS

Valid Values

Data Formatting

Date Issue Column

FIXING SPACES, CASES, AND FORMATTING

```
50 * Data Cleaning Process ;
51     data tsa.claims_cleaned;
52     set tsa.claims_cleaned;
53     if index(Claim_Type, '/') then Claim_Type= substr(Claim_Type, 1, index(Claim_Type, '/')-1);
54 
55     */ Fixing the Spaces /*;
56     if Disposition="Closed: Canceled" then Disposition="Closed:Canceled";
57     if Disposition="losed: Contractor Claim" then Disposition="Closed:Contractor Claim";
58 
59     */ Fixing proper and upper case names */;
60     StateName = propcase(StateName);
61     State = upcase(State);
62 
63     */ Fixing Format of the dates */ ;
64     format Date_Received Incident_Date date9.;
65 run;
```

DATE ISSUES AND FORMATTING

```
* Creating a new column that indicates date issues ;
data tsa.claims_cleaned;
    set tsa.claims_cleaned;
    if Date_Received<Incident_Date
    or Date_Received< "01JAN2002"d
    or Incident_Date< "01JAN2002"d
    or Date_Received> "31DEC2017"d
    or Incident_Date> "31DEC2017"d
    or missing(Date_Received)
    or missing(Incident_Date)
    then Date_Issues="Needs Review";
    */ Fomating the Curruncy /*;
    format Close_Amount dollar10.2;
    */ Drop unwanted column /*;
    drop County City;
run;
```

LABELING COLUMNS

```
93
94  * Adding Labels for every column ;
95  data tsa.claims_cleaned;
96      set tsa.claims_cleaned;
97      label Claim_Number= "Claim Number"
98      Date_Received = "Date Received"
99      Incident_Date = "Incident Date"
100     Airport_Code = "Airport Code"
101     Airport_Name = "Airport Name"
102     Claim_Type = "Claim Type"
103     Claim_Site = "Claim Site"
104     Item_Category = "Item Category"
105     Close_Amount = "Close Amount"
106     Date_Issues = "Date Issues";
107 run;
108
```

ANALYSIS AND STATISTICAL SUMMARY

Analysis Questions

Statistical Summary

OVERALL DATE ISSUES

```
108
109 */ Analyze the data and answer the questions /* ;
110
111 /* How many Data Issues */
112 title 'Overall Date Issues';
113 proc freq data = tsa.claims_cleaned;
114     table Date_Issues ;
115 run;
```

Overall Date Issues

The FREQ Procedure

Date Issues				
Date_Issues	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Needs Review	4175	100.00	4175	100.00
Frequency Missing = 216471				

OVERALL CLAIMS PER YEAR

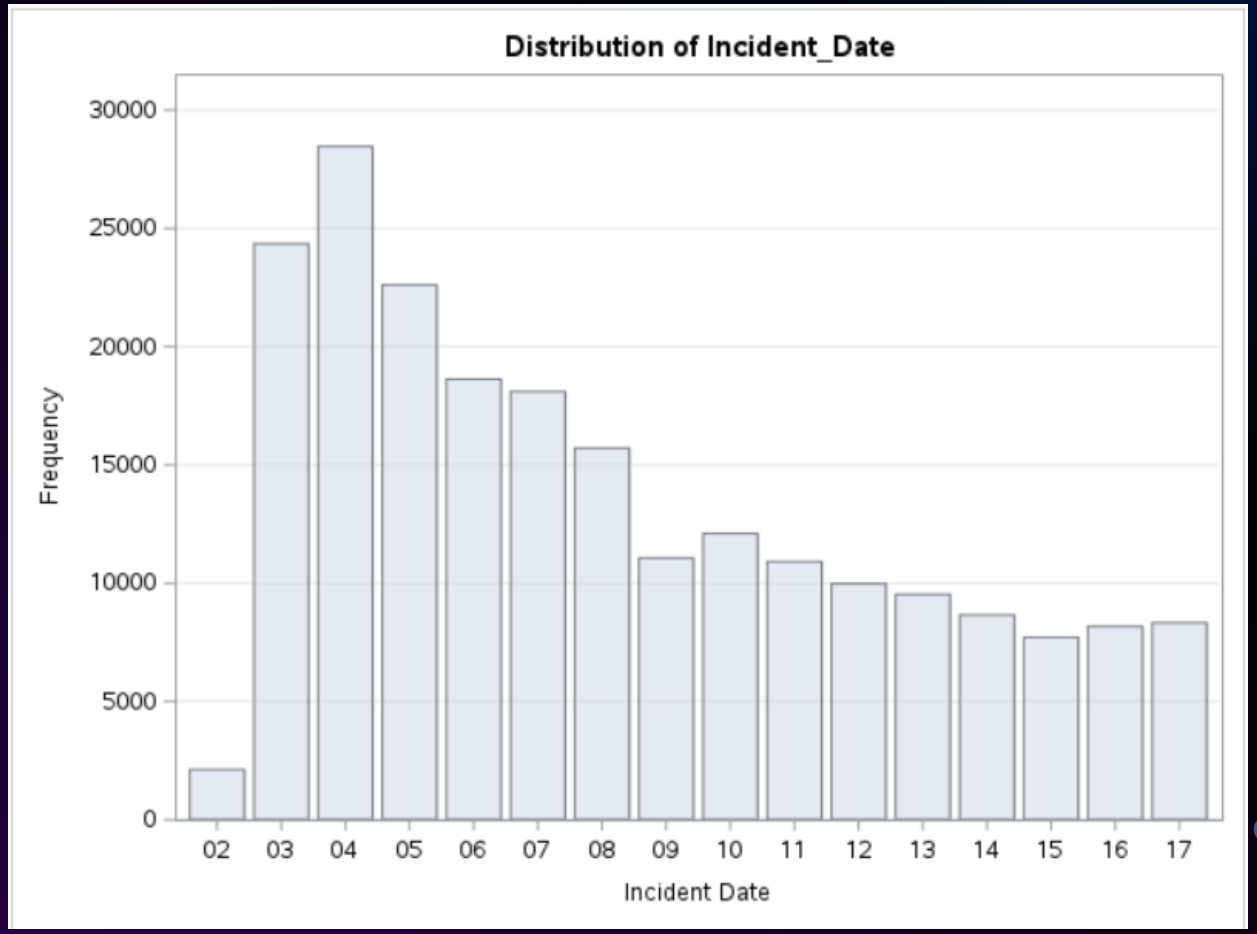
```
17 /* How many claims per year */
18 title 'Overall Claims per Year';
19 proc freq data = tsa.claims_cleaned;
20     table Incident_Date/ plots = freqplot;
21     format Incident_Date year2.;
22     where Date_Issues = '';
23 run;
```

OVERALL CLAIMS PER YEAR

Overall Claims per Year

The FREQ Procedure

Incident Date				
Incident_Date	Frequency	Percent	Cumulative Frequency	Cumulative Percent
02	2123	0.98	2123	0.98
03	24359	11.25	26482	12.23
04	28466	13.15	54948	25.38
05	22611	10.45	77559	35.83
06	18630	8.61	96189	44.44
07	18106	8.36	114295	52.80
08	15715	7.26	130010	60.06
09	11071	5.11	141081	65.17
10	12100	5.59	153181	70.76
11	10913	5.04	164094	75.80
12	9980	4.61	174074	80.41
13	9527	4.40	183601	84.82
14	8655	4.00	192256	88.81
15	7715	3.56	199971	92.38
16	8173	3.78	208144	96.15
17	8327	3.85	216471	100.00



FREQUENCY ANALYSIS

```
125 /* Showing the Freq Values for Claim_Type Claim_Site Disposition by spicific state */
126
127 * Choosing the state ;
128 %let SelectedState = Washington;
129
130 title "Frequancy for Claim Types, Claim Sites and Disposition by Washington";
131 proc freq data = tsa.claims_cleaned;
132     table Claim_Type Claim_Site Disposition ;
133     where Date_Issues = '' AND StateName = "&SelectedState";
134 run;
```

FREQUENCY ANALYSIS

Frequancy for Claim Types, Claim Sites and Disposition by Washington

The FREQ Procedure

Claim Type				
Claim_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Complaint	2	0.03	2	0.03
Employee Loss (MPCECA)	20	0.29	22	0.32
Missed Flight	1	0.01	23	0.33
Passenger Property Loss	3943	57.23	3966	57.56
Passenger Theft	20	0.29	3986	57.85
Personal Injury	37	0.54	4023	58.39
Property Damage	2654	38.52	6677	96.91
Unknown	213	3.09	6890	100.00

125 /* Sho

Claim Site				
Claim_Site	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Checked Baggage	5534	80.32	5534	80.32
Checkpoint	1306	18.96	6840	99.27
Motor Vehicle	2	0.03	6842	99.30
Other	38	0.55	6880	99.85
Unknown	10	0.15	6890	100.00

Disposition	Frequency	Percent	Cumulative Frequency	Cumulative Percent
*Insufficient	30	0.44	30	0.44
Approve in Full	1900	27.58	1930	28.01
Closed:Canceled	12	0.17	1942	28.19
Deny	2750	39.91	4692	68.10
In Review	281	4.08	4973	72.18
Settle	1265	18.36	6238	90.54
Unknown	652	9.46	6890	100.00

STATISTICAL SUMMARY

```
136 /* Statistical Analysis for Close_Amount*/
137 title "Close_Amount Statistics for Washington";
138 proc means data = tsa.claims_cleaned MAXDEC = 0 mean min max sum;
139 var Close_Amount;
140 where Date_Issues = '' AND StateName = "&SelectedState";
141 run;
```

Close_Amount Statistics for Washington

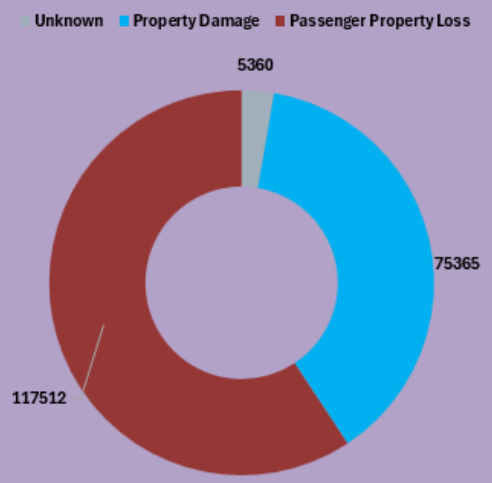
The MEANS Procedure

Analysis Variable : Close_Amount Close Amount			
Mean	Minimum	Maximum	Sum
98	0	9325	586416

ANALYSIS DASHBOARD

TSA CLAIMS REPORT

What are the most common types of claims?

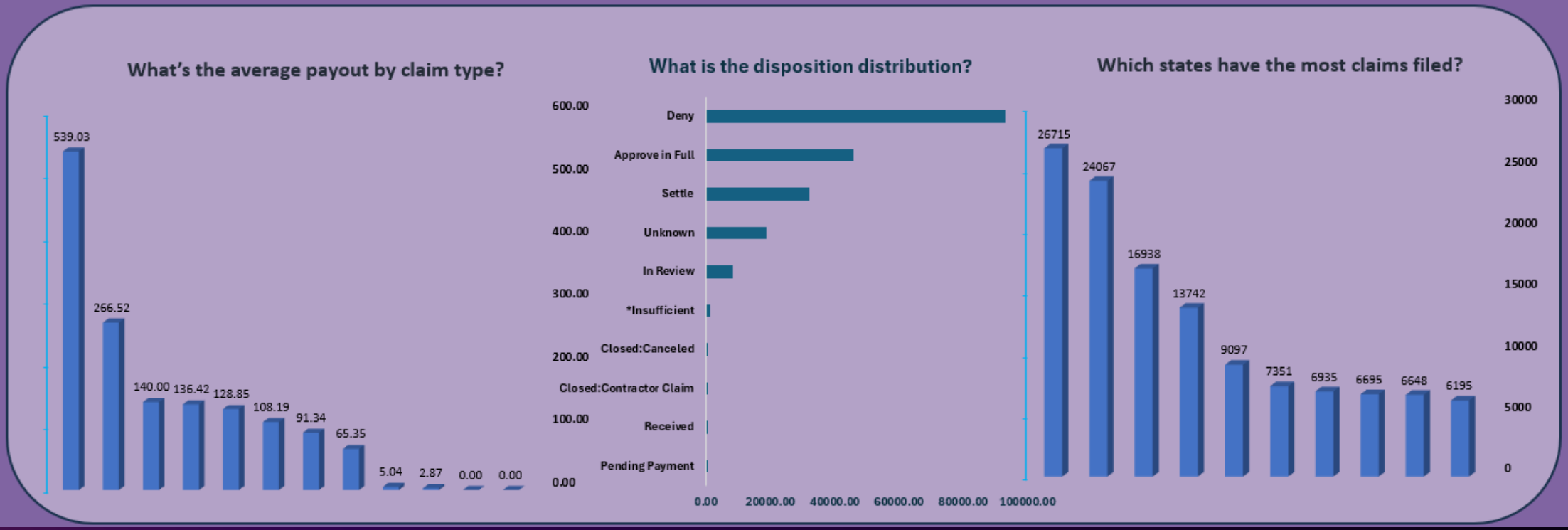


Top 10 airports have the highest number of claims?

What is the most claimed item category?

ANALYSIS DASHBOARD

TSA CLAIMS REPORT



PROJECT OUTCOMES

Key Deliverables

- Cleaned, standardized dataset ready for analysis.
- Insights into claims patterns, date inconsistencies, and financial impacts.
- Final report for stakeholders to guide TSA claims management decisions.

Impact

- Ensured data integrity for accurate decision-making.
- Highlighted trends and anomalies to prioritize operational improvements.

Data Analysis Dashboard

THANK YOU