TSA Complaints - Strategic Analysis of TSA Complaint Data: A Data-Driven Narrative for Change

Audience

The audience of this analysis is a leadership body of senior TSA executives and, potentially, Department of Homeland Security (DHS) stakeholders. The audience has some distinctive characteristics, including being mission-oriented, data-literate but time-constrained, and familiar with the TSA operational context, but not specifically with the detailed nuances of complaint trends.

Purpose

This data story has one purpose: to get an executive to approve the funding for a new training program across the system to improve the passenger experience with prompts for procedural consistency. The data story is persuasive, not informational. The goal is for the audience to leave with an understanding of three things:

1. Passenger discontent is not random acts or individual instances, but a real, quantifiable, and increasing systemic issue.

2. Specific, identifiable areas (both procedural and behavioral) generate complaints.

3. A responsible investment in training is a direct, logical means of addressing the root cause identified in the data.

Medium

The foundational analysis and narrative were developed in a Jupyter Notebook. Using a notebook for the analytic stage was ideal since it offered transparency, reproducibility, making it easy to integrate the code, narrative, and visualizations in one location. However, for the executive audience, the notebook is a complete, detailed, evidence-based source document that would be simplified into a more direct platform: a short visual presentation. The notebook shows the visualizations and summary findings, and it would serve as an appendix for those who want to dive deep into the analysis methodology.

Design

The design of the ten visualizations in the notebook was guided by Gestalt principles to ensure clarity and immediate comprehension. The strategy was to create a visual funnel, starting with a high-level national overview and progressively drilling down to specific, actionable details.

- **Color**: Color was used intentionally to convey meaning. In the choropleth map and heatmap, a sequential color scale (Viridis) was used to intuitively represent density and intensity, allowing leadership to identify hotspots instantly. For bar charts, contrasting but complementary colors were used to ensure readability and clear separation between categories.

- **Text:** Every visualization comes with a descriptive title and labelled axes. Additionally, markdown cells provide context to each visual explaining why they were chosen and how they build upon the central call to action. This information provides a textual link between the data points.

- **Alignment and Sizing:** All elements are aligned and have a polished, professional appearance. I chose horizontal bar charts rather than vertical ones for the category comparisons, as horizontal charts are easier to read and to add labels. Sizing provides a visual hierarchy. For example, I used font size in the word cloud to represent the frequency of complaint words. This allows any common pain points to jump off the page right away.

- **Spacing**: Ample white space was used between visualizations and text blocks to reduce cognitive load and allow each point to be absorbed individually before moving to the next.

- **Visuals**: The visualizations were selected to build a compelling argument. As required, the analysis includes a Choropleth Map (Visualization 1), a Heatmap (Visualization 3), and a Box Plot (Visualization 4). These were supplemented with line charts to show trends, bar charts to compare rankings, and a word cloud to add a qualitative, human element to the data.

Ethical Considerations

Strict ethical considerations were made throughout the data management and reporting process.

- **Data Source and Acquisition:** It was assumed that the data had been previously made available to the public by the TSA through a FOIA request or data portal; therefore, it is ethical to acquire it. Its credibility is high because it comes from the source organization, with probably only the biases inherent in how the data were collected.

- **Data Changes and Assumptions:** All transformations were documented. Column names were renamed for clarity (e.g., airport_x to iata_code). A critical assumption was made when handling missing location data in the airport merge; these were filled with the value 'Unknown'. This was done to retain all complaint counts while acknowledging the data's incompleteness. No data was filtered or removed from the analysis; "Top 10" charts are a clearly labeled form of presentational filtering, not a covert removal of inconvenient data. The most significant assumption is that complaint volume is a valid proxy for overall passenger dissatisfaction.

- **Risks and Potential for Misuse:** Using these data presentation formats presents risks. The choropleth map could, for example, unfairly penalize the states with large hub airports, as it presents absolute counts, not counts normalizing for the volume of passenger boarding. The "Top 10 Airports" chart could similarly invoke some blame directed to some base locations, while ignoring relevant and legitimate issues like staffing or equipment age. The word cloud illustrates a big picture, but is emotionally impactful and will exaggerate certain emotionally charged words out of context.

- **Mitigation Strategies:** To mitigate these risks, the following steps were taken and would be emphasized in a live presentation:

1. **Promote Systemic Focus:** The narrative is intentionally framed around improving processes and training, not blaming individuals or specific airports. The data is presented as evidence of systemic friction, not localized failure.

2. **Recommend Further Analysis:** The presentation would explicitly state that a more nuanced view would require normalizing complaint data against passenger throughput, recommending this as a follow-up analysis.

3. **Maintain Full Transparency:** All cleaning steps and assumptions are documented in the source notebook, ensuring the audience can understand the analysis's limitations.

4. **Uphold Privacy:** The dataset was confirmed to contain no Personally Identifiable Information (PII), adhering to fundamental data privacy principles.

5. **Contextualize All Visuals:** Each visual is explained, preventing it from being interpreted in isolation. The story guides the audience toward a constructive conclusion, rather than leaving them to draw potentially harmful ones.

Visualizations

Visualization 1: Monthly Trend of Total Complaints (Line Chart)

Why it was chosen: A line chart is the best tool for showing trends over time. It answers the question: 'Is the problem getting better or worse?'

What it does: This chart plots the monthly complaints received. This allows us to see the overall trajectory and any seasonal patterns (e.g., spikes during holiday travel seasons).

How it helps the call to action: A direct upward trend on this chart produces a strong sense of urgency. It visually illustrates to leadership that passenger dissatisfaction is not a static issue, but one that is growing and therefore requires 'immediate action', which certainly increases our case for the training program we proposed.
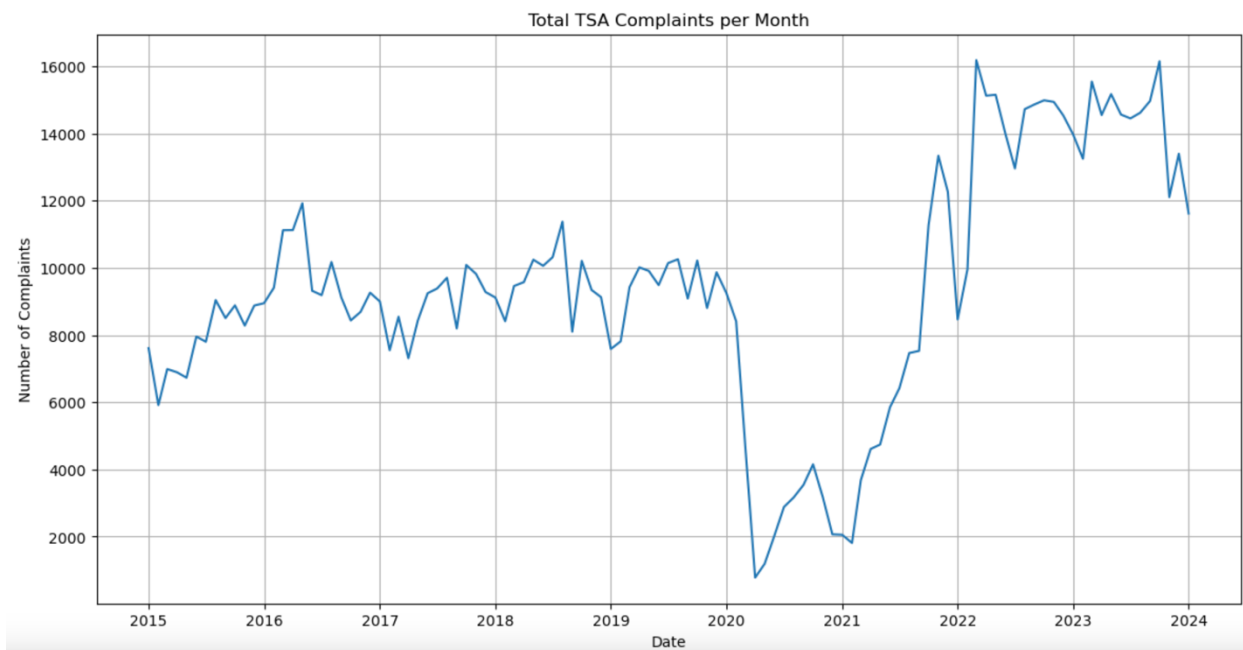
Visualization 2: Year-Over-Year Complaint Growth Rate

Why it was chosen: A raw trend line shows growth, but a percentage growth chart highlights the rate of change, which can be a more powerful metric for conveying urgency to leadership.

What it does: This bar chart calculates the total number of complaints yearly and then displays the yearly percentage increase or decrease.

How it helps the call to action: A chart showing, for example, a 25% jump in complaints last year is a stark, undeniable signal that the problem is growing and accelerating. It frames the issue in business terms (growth rates) that leadership understands well and provides a powerful final argument for immediate investment in the proposed training program.
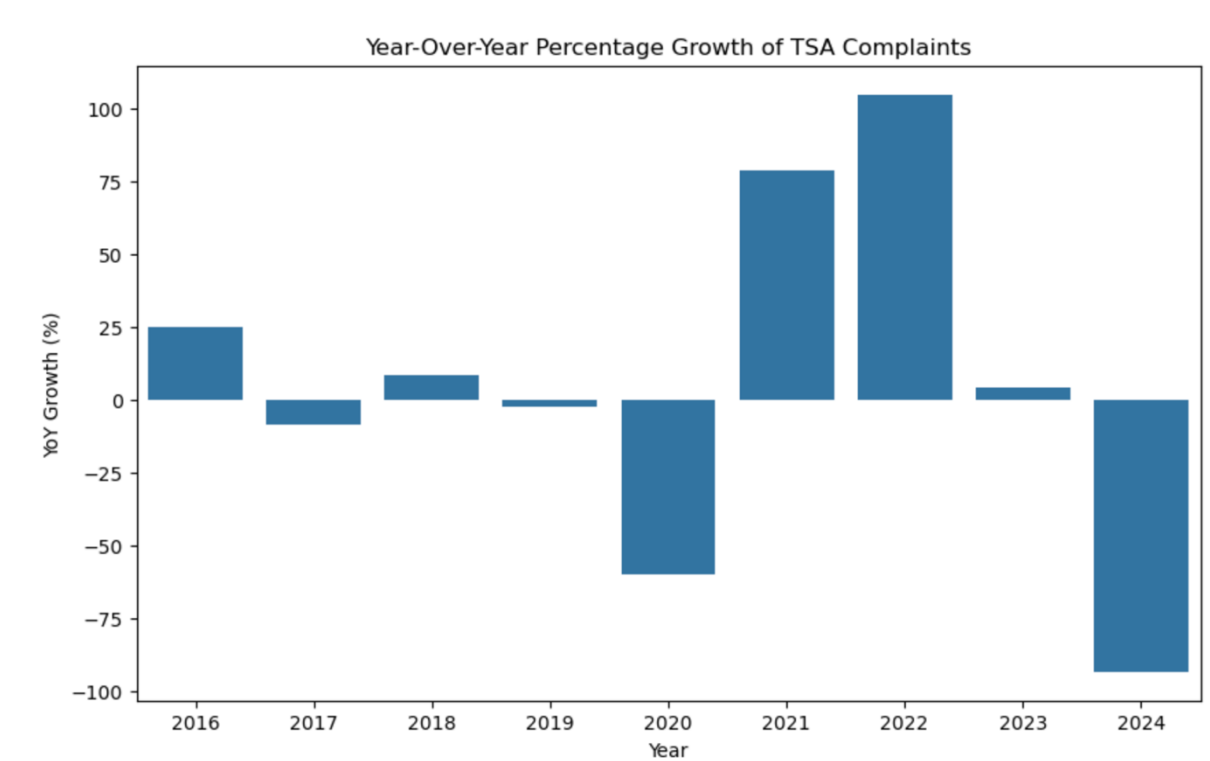
Visualization 3: Top 10 Overall Complaint Subcategories

Why it was chosen: Understanding the higher-level categories is essential, but fundamental changes

occur at the lower levels. This chart shows the form of the most frequent, specific complaints in each

area.

What it does: This horizontal bar chart displays the top 10 most common complaint subcategories from

the entire dataset, providing a clear rank-order of the most pressing specific issues.

How it helps the call to action: This graphic offers a specific, actionable list for leaders. Breaking down

broad categories, such as 'Screening Process,' into issue topics like 'Divestiture,' 'Staffing,' or 'Screening

of Persons,' allows leaders to create specific training modules to mitigate friction at the most common

friction points.
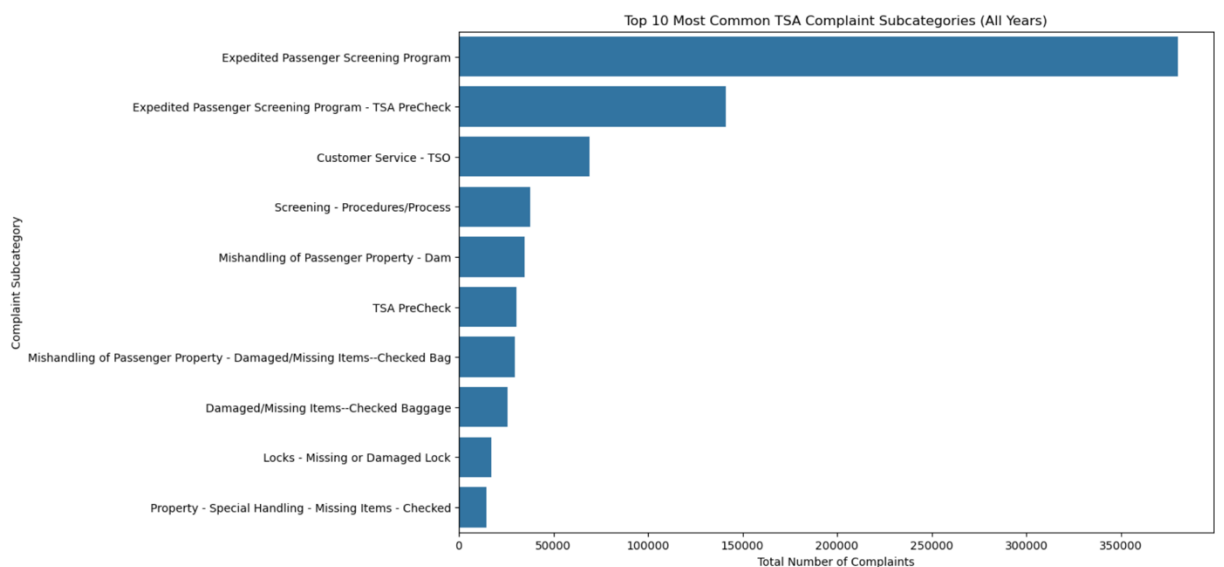
Visualization 4: Word Cloud of Complaint Subcategories

Why it was chosen: A word cloud provides a qualitative, high-impact look at the specific words and phrases passengers use in their complaints. It adds a human element to the quantitative data.

What it does: These visuals display the most frequent terms in the 'subcategory' field. The size of each word is proportional to its frequency.

How it helps the call to action: Seeing words like 'Rude,' 'Unprofessional,' or 'Confusing' displayed prominently gives leadership a direct, unfiltered view into the passenger's perspective. It moves beyond abstract categories to the language of frustration, making the need for improved interpersonal skills training undeniable and emotionally resonant.

Most Common Words in Complaint Subcategories

Visualization 5: Complaint Categories Breakdown (Horizontal Bar Chart)

Why it was chosen: A horizontal bar chart is far superior to a pie chart for comparing proportions, especially with several categories. It avoids label overlap and makes it easy to rank categories.

What it does: This chart shows the total count of complaints for each major category, ordered from most to least common. It answers: 'What are the biggest drivers of complaints?'

How it helps the call to action: These visuals rank problem areas. If 'Passenger Courteousness' and 'Screening Process' are at the top, it provides a powerful, simple visual that tells leadership exactly where to focus their improvement efforts. It validates that our proposed training addresses the most significant part of the problem.
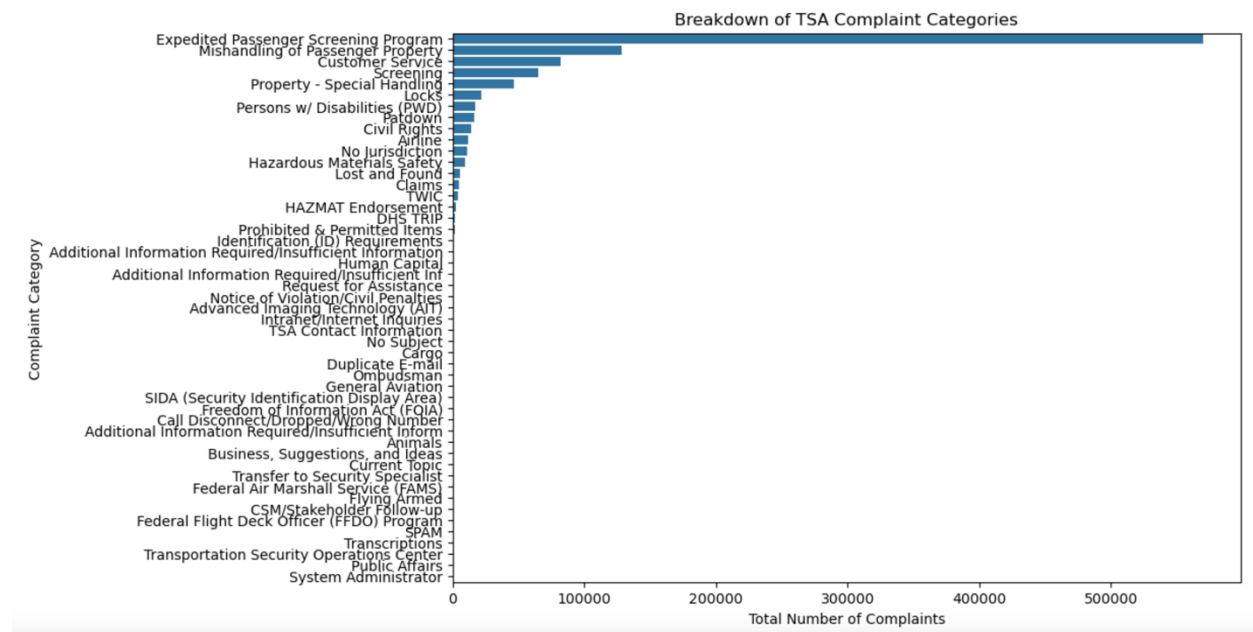
Visualization 6: Bar Chart of Top 10 Airports by Complaint Volume

Why it was chosen: A bar chart is highly effective for directly comparing. It clearly and immediately identifies the airports with the highest complaints.

What it does: This chart displays the top 10 airports ranked by their total number of complaints over all years.

How it helps the call to action: This visualization helps leadership identify which major hubs are the most significant sources of passenger friction. This allows us to suggest that the pilot program for our new training could be rolled out at these specific high-impact locations for the most immediate and measurable results.
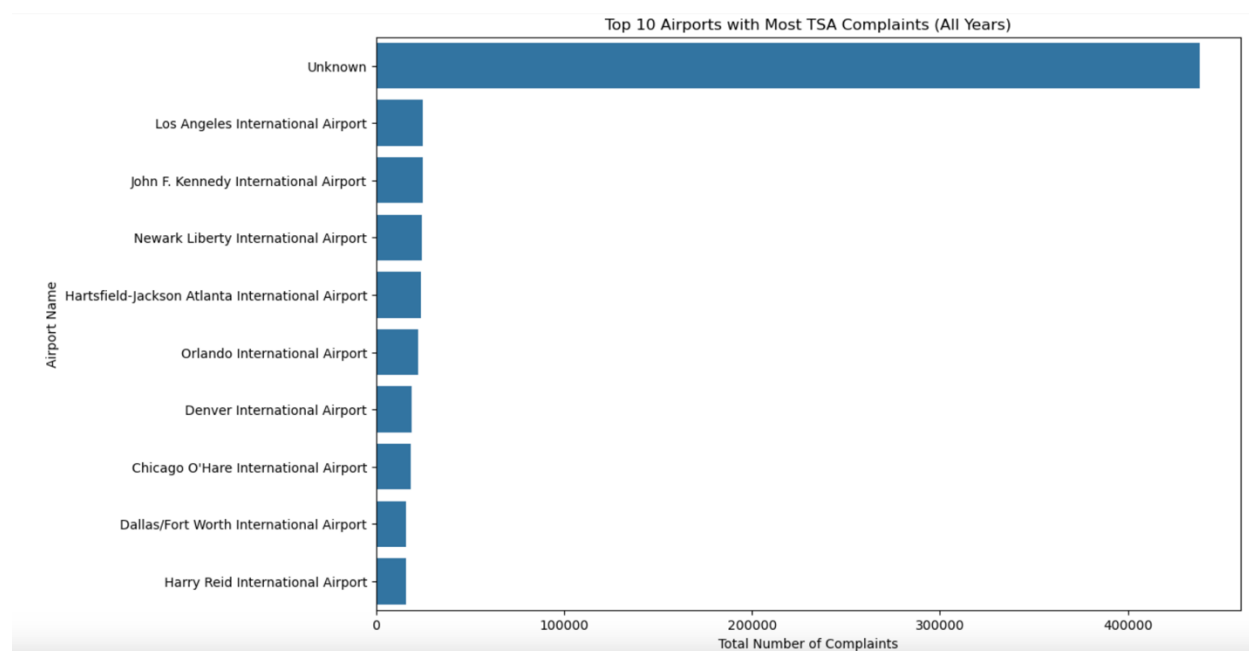
Visualization 7: Box Plot of Complaint Distribution by Year

Why it was chosen: A box plot shows the distribution and spread of data. It goes beyond simple averages to show each year's range, median, and potential outliers in complaint volumes.

What it does: This visualization shows a box plot for each year, representing the distribution of monthly complaint counts across all airports. We can see if the median number of complaints is rising.

How it helps the call to action: An upward trend in the median demonstrates that high complaint volumes are not just isolated incidents but are becoming the 'new normal' across the board. This worsening systemic issue creates urgency in our call for a new training initiative.

Distribution of Monthly TSA Complaints per Airport by Year
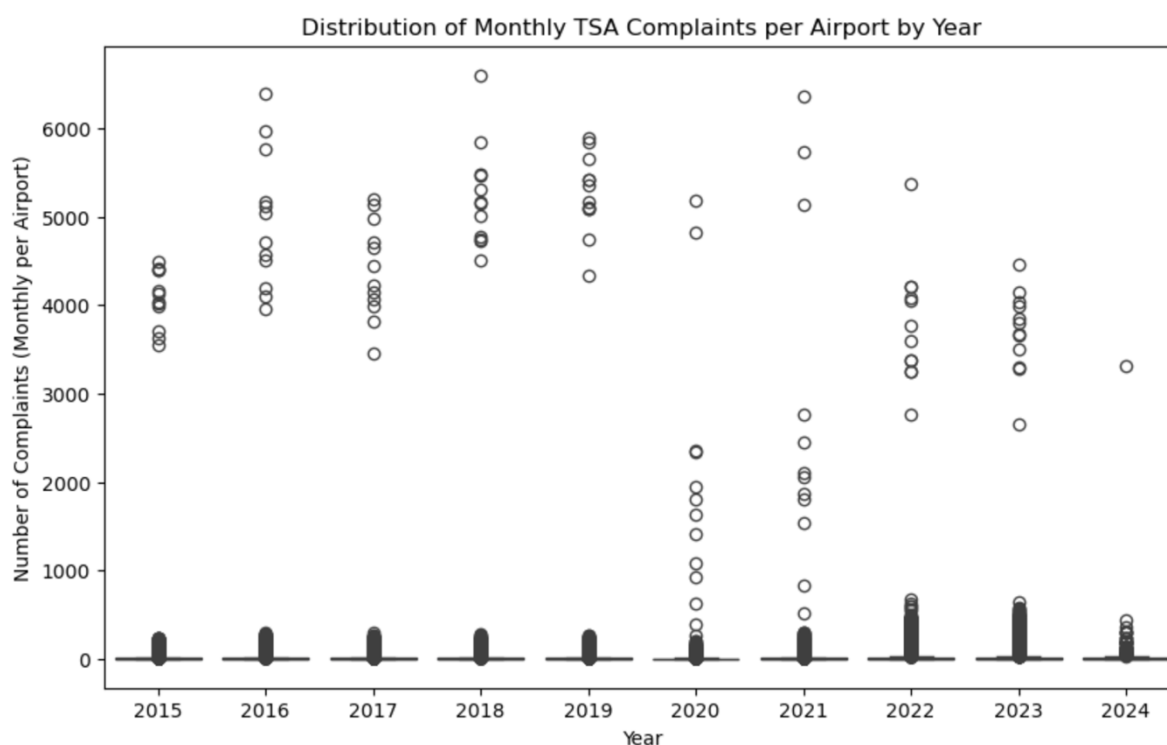
Visualization 8: Heatmap of Complaints by Category and Year

Why it was chosen: A heatmap shows the relationship between two categorical variables. It allows us to see how the composition of complaints has changed over time.

What it does: This heatmap displays the total number of complaints for each major category (y-axis) for each year (x-axis). The color intensity reveals which categories are growing sources of friction.

How it helps the call to action: If the heatmap shows a consistent darkening for categories like 'Passenger Courteousness,' it provides direct evidence that the *quality* of interactions is a growing problem. This directly supports our call to action for improved customer service training.

TSA Complaints by Category and Year

Complaint Category	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Additional Information Required/Insufficient Inf	50	46	185		26	26	61	94	150	8
Additional Information Required/Insufficient Information	98	129	54	82	125	105	72	41	33	62
	72	54	82	86	72	19	2	10	9	1
Airline	1983	2003	1837	1773	1772	931	288	493	408	49
	10	4	2	1	2	4	8	7	7	1
Business, Suggestions, and Ideas	1	7	9	0	2	3	6	3	4	0
	6									
Call Disconnect/Dropped/Wrong Number	22							27	0	
	49	29	29	32	13	2	13		14	0
Civil Rights	1063	1620	1527	2033	2084	910	1446	1510	1720	164
	904	601	858	813	741	431	226	176	118	9
Current Topic	2	16	2	1	3	1	7	4	3	
DHS TRIP	6603	10980	10706	10724	9657	4165	7615	9355	11196	1167
	623	387	260	210	189	63	55	102	110	12
	1					11	122	2		0
Expedited Passenger Screening Program	33597	46810	40002	49694	49173	17832	54672	134190	135469	8504
	4	6	1		1	2	3	2	3	0
Federal Flight Deck Officer (FFDO) Program	1	0	1	0	1	2	3	0		1
Freedom of Information Act (FOIA)	5	3	3	6	11	5	4	7	9	0
	17	31	19	8	13	10	15	0	2	7
HAZMAT Endorsement	1282	597	178	117	85	59	43	191	115	7
	884	1201	1561	2259	2266	1218	126	136	110	9
Human Capital	88	86	44	45	68	31	96	115	91	4
	191	224	146	78	75	69	331	159	2	0
Intranet/Internet Inquiries	71	131	45	44	37	21	5	10	15	35
	3852	4085	3889	3730	3798	1284	369	378	469	
Lost and Found	657	1151	945	783	832	388	317	410	324	53
	30740	31607	22406	15934	13862	5651	2178	2408	3166	319
No Jurisdiction	830	1237	1250	1239	2185	854	712	1015	1510	119
	22	42	9	12	5	15	52	10	28	7
Notice of Violation/Civil Penalties	98	60	57	40	32	34	46	104	90	7
	27	30	19	10	7	6	6	13	19	2
Patdown	897	2221	2783	2554	2374	646	1062	1831	2066	120
	954	1707	2134	2761	2909	1041	1502	1824	2217	167
Prohibited & Permitted Items	145	149	231	91	86	104	583	8	19	1
		0	7413	10766	11658	5537	4277	3500	3149	310
Public Affairs	2	65	93	168	124	23	21	30	50	3
	0						0	1		1
SIDA (Security Identification Display Area)	40	24	14	11	8	1	1	5	7	1
		0	1	0	1	1				
Screening	5133	8837	7588	9077	8238	3820	4599	7311	9818	518
	1					1	0	1		
TSA Contact Information	84	60	48	23	10	6	5	10	0	1
	2369	471	171	90	84	38	83	378	170	14
Transcriptions	1			3					3	
		0	0			0	17	0		0
Transportation Security Operations Center	1					1	1	3	0	

Colorbar scale: 0 – 20000 – 40000 – 60000 – 80000 – 100000 – 120000