

Estimate Refund amount for Claim made to TSA

* (Alternate idea: estimate probability of losing an item at TSA and/or with airline baggage handlers. Basically estimate likelihood of having a baggage issue at some point on a flight)

Introduction

The Transportation Security Administration was created in 20XX, largely motivated as a change in flight security policy post the September 11, 2001, attacks. Passing through TSA checkpoints has been routine and ubiquitous in all United States airports since then, as every one of the millions of US passengers that fly daily must be checked at a TSA security screening since then.

Sometimes passengers encounter unwanted incidents when going through the TSA process. They may leave an item behind, suffer damage to property, have an item stolen, or more. As mandated by congressional law, TSA has setup a claims process for those who feel they have suffered a wrong doing by file a claim for monetary compensation.

TSA provides data of each claim, including the category of incident, date it occurred, date it happened, category of item damaged/stolen/lost, the amount of compensation the claimer is seeking, the decision of the claim, and how much money compensated (if applicable).

Additionally, I could **potentially** find data on various airlines and how often they lose or damage checked bags. If the data is good, this could be combined with the TSA.

Question/Need

What is the expected compensation value one can get from filing a claim to TSA?

OR

How likely is it that you will have a TSA incident based on other details of your upcoming flight?

If I can obtain good data on checked baggage issues and their claims from AIRLINES, then:

Alt 1: Predict expected compensation value (regression) from either/both TSA and/or airline for an incident

Alt 2: Predict probability of ANY kind of baggage incident on your next flight.

Data

TSA releases claims data on a yearly basis. In aggregate, there is a table of over 200,000 claims from 2002-2015 with the following values I plan to eventually create a table like the following:

claim_number	date_received	incident_date	airport_code	airport_name	airline_name	claim_type	claim_site	item	claim_amount	status
string	datetime	datetime	string	string	string	string	string	string	integer	string

Another dataset to pair will be obtaining passenger throughput through TSA or through airport. This data will allow to calculate # of claims per passenger for each airport, thus putting smaller and bigger airports on the same scale. Dataset still needs to be obtained.

Another dataset would be baggage_claim incidents dealt with between *airlines* and *passengers*. If good data is obtained, it can be used in combination with the TSA data to predict the probability any given passenger will **have a baggage related event** from the point they check their bag, to the point they arrive at their destination and pick up their bag from the baggage claim conveyor belt.

See what I can get from the Lost & Found info that TSA has.

Potential data issues

- TSA only reported `claim_amount` for 2002-2006. Afterward, they only released how much compensation money, if any, was given for a given claim.
- The TSA claim data has data released for 2016 and 2017, however it is in PDF format, which will require additional effort to manipulate.

Characteristics of each row of data

Each row of table is for 1 claim. The most interesting target areas will be like the `claim_amount` (what the claimer was requesting), the status and disposition of the claim (approved, denied, in process) and if so to what extent (approved in full? settled, etc.), and finally the `close_amount` which is what the claimer actually was rewarded.

The features will be used to model for any patterns that exist in who gets claims and why.

Known Unknowns

- What went INTO the actual claim (receipts, documentation, haggling, follow-up attempts, crying, etc)

- Not everyone has an incident files a claim. Thus actual incident rate will be higher by an unmeasurable amount.
- Relying on the "accused" - in this case TSA - to provide accurate data. They have a bias. If and how it affects the true data is unknown.
- TSA also does Lost & Found. These would certainly be counted as incidents. Need t
- Not sure if I can get airline-specific baggage complaint data, compensation values to customers for lost/damaged baggage, etc.