



Asher Wright & Austin Zaccor

## Project Purpose

**Purpose: Detect fraudulently created airline bookings**  
Travel agents have access to legacy booking systems when booking flights. These legacy booking systems are not equipped to enforce all booking rules, and some agents abuse this fact to book flights at cheaper rates than are available.

## Actions

- After detection, Air Canada can:
1. Cancel the fraudulent booking, if not yet ticketed
  2. Send debit memos to the booking agency (if part of IATA)
  3. Inhibit the booking agency from booking in the future

## Data

**Data Axiom:** To know if a booking is fraudulent, we must know the complete history of how that booking was created by the agent

### Bookings

**Bookings, or Passenger Name Records,** contain both trip and segment-level data.

**Trip-level data:** number of passengers, frequent flyer number, whether the booking has been ticketed, and more.

**Segment-level data:** origin and destination, date, flight number, seat class, and more.



### Sell Requests (+ Graylogs)

When travel agents assemble bookings, they often experiment, e.g. by changing flights or seat classes. Their tinkering gets stored as requests (what they asked) and responses (what Air Canada responded) in **Sell Requests**.

**Graylogs** contain even more information about these requests & responses, but are in a raw, unparsed format.

## Data Engineering Methodology

**Problem:** There is no straightforward way to find which **sell requests** are related to which **bookings** (or which **sell requests** are related to one another). *For example, the closest thing to a Travel Agent Identifier is the Pseudo City Code (PCC), which is far from unique given that an agency as large as Expedia has only one PCC.*

### Overview

Joining **sell requests** and bookings was a complex task that took months to solve in a scalable way.

In the end, to link **bookings** & **sell requests** we:

1. Clean both datasets
2. Find the final **sell request** that led to the **booking** (Phase I)
3. Find all related **sell requests** to that **final request** (Phase II)

### Phase I

Finding the final **sell request** that led to a **booking** means finding the **last request** an agent made before booking.

We do this by finding **bookings** and **sell requests** close in time with the same attributes, like flight numbers, departure dates, and seat classes.

If we have multiple eligible matches, we choose the best one and repeat the process with unmatched **bookings** and **sell requests**.

### Phase II

Finding all other **sell requests** related to the **requests** found in Phase I is done by following the predefined rules of Air Canada's **sell request** system.

For example, if the **last request** had three segments, and one of which they marked as "already have", then they must have asked for that segment previously. Thus, we simply follow the system's logic back in time until we have identified all related **requests**.

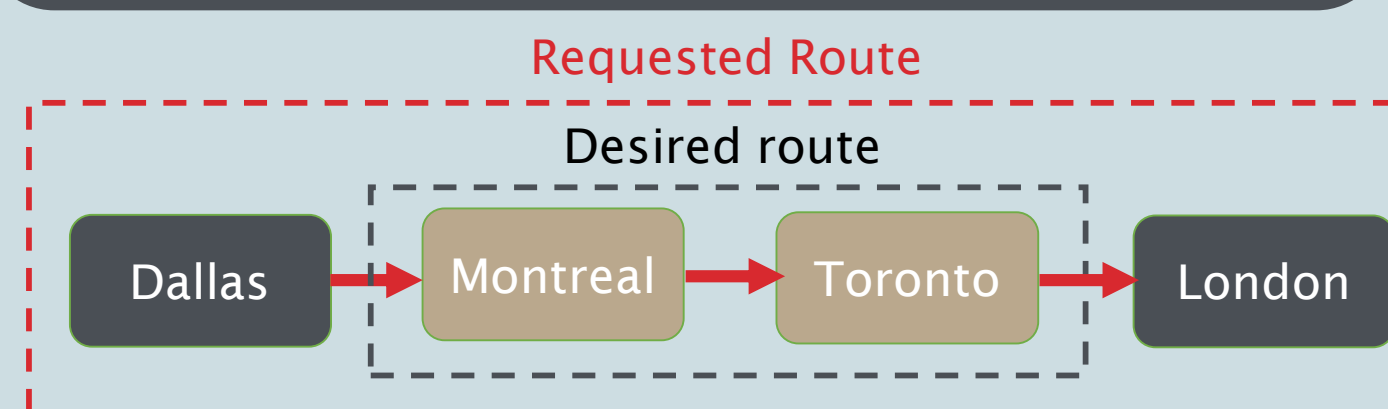
## Fraud Detection Methodology

Once we knew what the travel agent did to create the booking, we could determine if it was fraudulent by checking for known types of abuse:

1. Hidden City Abuse
2. Out of Order Abuse

Using our new dataset, we tried to find other types of abuse as well, but found that these two were the only types.

**Hidden City Abuse** occurs when a travel agent tries to make it seem as though they are booking another, longer segment in order to get better fares for the shorter segments they actually desire.



**Out of Order Abuse** occurs when a travel agent tries to make it seem as though they are booking a roundtrip flight in the reverse direction of what they actually want in order to get cheaper fares.



## Product

We gave **two deliverables** to Air Canada:

1. An auto-updating data pipeline, such that the data processes are run daily and new bookings can be marked as fraudulent
2. An interactive dashboard showing the fraudulent bookings and allowing for exploration and examination of them

## Impact

For the amount of fraud found in July, we project:

- \$255k protected revenue
- Cancelled fraudulent tickets protect revenue from lower fares
- **\$1.2M increased revenue**
- Debiting agencies can yield \$1.1M in fines
- Rebooked tickets yields \$100k

## Next Steps

Currently, the Revenue Integrity Team takes actions on bookings manually.

To maximize the impact of this work, we suggest automating the following:

1. If the booking is not yet ticketed, cancel it
2. If agency can be debited, do so
3. Else, inhibit the agency from making future bookings

## Acknowledgements

A very special thanks to the many people who helped us at Air Canada, including but not limited to: Soad El-Taji, Katrina Hizon, Farnoush Farnia, Fatemeh Farnia, Rezaei Mohammad, and Arpit Dhindsa as well as our MIT advisors Prof. Arnie Barnett and Dr. Carine Simon