skyscanner

FIB · UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

SKYSCANNER

IN COLLABORATION WITH

UNIVERSITAT POLITÈCNICA DE CATALUNYA (UPC)

FINAL DEGREE PROJECT

Skyscanner Heatmap

Skyscanner's data domain representation

Author:
Fèlix Arribas

Director:
Francisco López
University supervisor:
Maria José Casañ

A Project for the Computer Engineering Degree *in the*

Software Engineering and Information Systems department

Facultat d'Informàtica de Barcelona (FIB)

working with

DeLorean squad *from* Marketplace Engine tribe

Wednesday 10th January, 2018

Universitat Politècnica de Catalunya (UPC)

Abstract

Facultat d'Informàtica de Barcelona (FIB)
Software Engineering and Information Systems department

Computer Engineering Degree

Skyscanner Heatmap

by Fèlix Arribas

In the last century, the world has became smaller. Communications are easier and faster than fifty years ago. Back then, you could talk through a fix phone, but you were not able to send any kind of media, like photos, videos, etc. Only the latest technology of that moment was able to do that. Since the smart phone revolution in 2007 almost everyone can text messages, sending images, share live videos or almost whatever you can imagine in less than a second.

But the internet, phones and communications are not the only thing that made the world smaller. Ways of traveling helped to this earth flattering too. In 1918 visiting another place was very difficult. If you wanted to go through the sea, you had to do it by boat. The fastest way to travel very far in a continent was by train, but not all places were connected with rails. Nowadays, all along with the internet revolution, anyone can travel to the other side of the world in less than a day by plane. Even for traveling inside the same country people use planes.

Is the air industry as efficient as it could? Are all airplane users satisfied with their purchases and possibilities? Skyscanner provides an easy to use tool to search cheap flights from any airport to another. Sadly, sometimes is difficult for users to find what they really want.

This project wants to help solving this problem, providing a heat map to explore differences and similarities between what users search, finds and finally buys and what airlines provides. Being able to compare between specific dates to guess user behavior.

Contents

Chapter 1

Introduction

This is a project developed in *Skyscanner* and evaluated by the *Universitat Politècnica de Catalunya (UPC)* as a Final Degree Project. The main goal is creating a tool for *Skyscanner* providers to ease the comparison of routes by different parameters (for instance, user demand and flights provided), so then the flights advertisement is improved according to user demand. Also being programmer oriented, so the company can develop complex software using the Application Program Interface it will provide.

1.1 Context

1.1.1 Economy

1.1.2 Evolution

1.1.3 Complexity

1.2 Motivation

1.3 Vision

Chapter 2

Market

Chapter 3

Project scope

3.1 Objectives

3.2 Scope

3.2.1 Data warehouse

3.2.2 Visual representation

3.2.3 Other applications

3.3 Risks

3.4 Methodology

3.4.1 Tribes and Squads

3.4.2 Agile

3.4.3 GitLab

3.4.4 Jira

3.4.5 Other tools

Chapter 4

Requirements analysis

4.1 Actors

Initially it seemed difficult to find stakeholders and actors in these project apart from the providers. It is not a tool for the user of Skyscanner so, as explained before, one risk of these project was not finding enough support.

After walking with the Squad Lead and then the Product Own of DeLorean squad a lot of stakeholders appeared: DeLorean Squad, Marketing Automation Squad, State Machine Squad, etc. Each of these stakeholders has different use cases and the project became very interesting for a considerable part of Skyscanner .

4.1.1 DeLorean squad

The mission of DeLorean squad is to provide the best data and services around the routes, timetables and modes of transportation to go from one point on Earth to another.

Right now, DeLorean squad provides a very fast service that serves flights logistic information between a given origin and destination. Some information you can find in a route is the fight number, carriers, stops, date ranges, etc. These flights are just **single ticket flights**. The squad is currently working on more complex routes, constructed routes[1]. The constructed routes timetable is under construction, so is not available for these project.

The Heatmap relies on the timetable of Single Flight Number service, also known as Timetable SFN Service. This service provides all the **current** flights. This is a little bit of a problem when trying to get past routes: Timetable SFN Service does not provide past flights information, it is always **up-to-date**. In order to get this data it is needed to go one step back in the whole DeLorean data processing: Timetable Pipeline.

The heatmap must reference old versions of the file created by the Timetable Pipeline to get older routes.

Timetable SFN Service

The *timetable SFN* endpoint returns details for time tabled Single Flight Number itineraries series. Note that SFNs are not ticket-able, so they do not include itineraries

---

[1]Contructed Timetable contains constructed routes, routes composed by two or more single ticket flights.

which cannot be bought on their own, neither the price nor restrictions. Timetable SFN Service provides

Timetable Pipeline

This phase, basically collects all the OAG[2] from a provider and maps it into routes in JSON[2] format. For each different version of the OAG file, the pipeline creates a new file with all the routes.

Product Owner

Jen Agerton is the Product Owner of DeLorean Squad.

DeLorean's Squad Lead

Francisco López, who is also the supervisor of this project, and I has the initial idea for this project. He oriented it for a Machine Learning purpose.

4.1.2   Fuel RaTS squad

Routes and Timetable Servies Squad provides the best data and services around the routes, timetables and modes of transportation to go from one point on Earth to another. Fuel RaTS has the same mission as DeLorean Squad, but develop different services. Since Fuel RaTS provides basic routes data, pricing, live update information and multi-destination combinationcs, DeLorean squad provides a very fast service for only routes.

4.1.3   Marketing Automation squad

Marketing Automation squad enables scalable growth by automating workflows, and the collection of insightful data. They have three main goals:

- Provide data to support decision making

- Automated, data driven campaign management

- Budget process automation

4.1.4   User

4.2   Functional requirements

4.3   Non functional requirements

4.4   Use cases

---

[2]OAG file (also know as WTF file), is a CSV[1] file which each row represents a timetable for a Single Flight Number.

Name	Routes offer and demand comparison **heatmap**
ID	UC0
Description	Heatmap of the comparison between providers offer and user demand. The heat is represented by the *over requests* of a route.
Actors	User
Organzational Benefits	General map that allows identify in a general view non profitable and most requested routes according to the offer. Guides the user to select choose the correct quety for UC1 and UC2.
Frequency of Use	TBD
Triggers	Home page.
Precondition	
Postcondition	Wolrd heatmap with most relevant routes and their heat.
Main Course	
Alternate Courses	
Exceptions	

TABLE 4.1: *Routes offer and demand comparison **heatmap** use case*

Name	Offer and demand plot of route
ID	UC1
Description	Compare the user demand and the providers offer of a specific route from city A to city B in a given date in a plot with two data sets, offer and demand.
Actors	User
Organzational Benefits	Helps users understanding which routes are *over requested* and learn about tendencies.
Frequency of Use	TBD
Triggers	Request to get comparison of route from city A to city B in a specific date.
Precondition	City A and city B exists and there is some connection (SFN or Constructed) in the date.
Postcondition	Plot with the evolution through time of the user demand and air carrier offer. Time limit goes from fist offer apperance to arrival date or current date, depending which comes first.
Main Course	1. System provides a list of cities under *origin* tag. 2. User selects an origin city. 3. System provides another list of cities. Now with *destination* tag. 4. User selects destination (See exception 1). 5. System provides an interactive calendar. 6. User selects a date of the calendar (See exception 2). 7. System provides the plot of the demand and offer evolution of the route.
Alternate Courses	Alternate course 1 1. User **changes** destination city (See exception 1). 2. Return to Main Course step 6. Alternate course 2 1. User **changes** date (See exception 2). 2. Return to Main Course step 7.
Exceptions	1. There are no connections between to given cities. 2. There are connections between to given cities, but not in the given date.

TABLE 4.2: *Offer and demand plot of route* use case

Name	Offer and demand data set of route
ID	UC2
Description	Data set of the evolution of the user demand and providers offer in order to create metrics, alerts, etc.
Actors	Marketing Automation squad, DeLorean squad
Organzational Benefits	With the data provided Marketing Automation squad could be able to create alerts to warn them of notable changes in user demand DeLorean squad may be able to use this data as training in future Machine Learnings algorithms applied to Constructed Routes Pipeline
Frequency of Use	TBD
Triggers	Request to get data set of route from city A to city B in a specific date.
Precondition	City A and city B exists and there is some connection (SFN or Constructed) in the date
Postcondition	Plot with the evolution through time of the user demand and air carrier offer. Time limit goes from fist offer apperance to arrival date or current date, depending which comes first.
Main Course	1. System provides an HTTP endpoint to request data. 2. The developer does a GET request to the endpoint with an origin, destination and a date (See exception 1). 3. System provides a data set in JSON format with all the demand and offers of the entity.
Alternate Courses	
Exceptions	1. There no connections between city A and city B in the given date.

TABLE 4.3: *Offer and demand data set of route* use case

Name	name
ID	id
Description	description
Actors	actors
Organzational benefits	benefits
Frequency of use	frequency
Triggers	trigger
Precondition	pre
Postcondition	post
Main course	main
Alternate courses	alt
Exceptions	exc

TABLE 4.4: *title* use case

Bibliography

[1] Y. Shafranovich. *Common Format and MIME Type for Comma-Separated Values (CSV) Files*. 2005. URL: https://tools.ietf.org/html/rfc4180 (visited on 01/30/2018).

[2] JSON.org. *The JavaScript Object Notation (JSON) Data Interchange Format*. 2017. URL: https://tools.ietf.org/html/rfc8259 (visited on 01/30/2018).