Data Analytics 2018-2019

Case study: Airline Company Analysis

The marketing department of an airline wants to analyze the activity of the flights of each member of its frequent flyer program. This department is interested in analyzing which flights the frequent passengers of the company choose, what airplanes they use to travel, what rates they pay, how they earn and redeem their points, what is the duration of your stays at destination, etc.

Below is the process of developing an analytic repository to cover the needs of this department.

Step 1: Select the business process to model

The management of flights made by frequent passengers of a company is selected. We **do not focus** on booking flights that do not result in the boarding of a frequent passenger on the plane.

Step 2: Select the granularity of the fact that the process

In this case study, the airline captures data at the level of journeys. A path represents a plane taking off at an airport and landing on another airport without intermediate stops. Information is known about the date of the flight, departure time, arrival time, origin and destination airport, the base price of the trip, the points that are obtained (or are spent) when performing the route and the accumulated delay for the journey. The plane that performs the route is also known as well as the base fare and the seats available for each type on each flight.

The airline also collects data from the itinerary, which equals the reservation or flight ticket with its corresponding reservation confirmation number and a date backup. An itinerary consists of a series of routes or paths. For example, if a client book an itinerary to travel from Alicante to Malaga and then again to Alicante, this itinerary can consist of the following routes: Alicante-Madrid, Madrid-Malaga, Malaga-Alicante. Actually the passenger only wants to go from Alicante to Malaga and return without worrying about connections (i.e. in Madrid). However, the company needs data about intermediate stops more complete analyses.

Each fact entry will correspond to the information collected on the boarding pass from every frequent passenger.

Step 3: Choose the dimensions that will be applied to each fact

If the information of the fact corresponds with the data of a boarding pass, according to the above, the dimensions should be the following:

- The type of payment that identifies how the itinerary was purchased (telephone, Internet, etc.).
- The airports of origin and destination for each journey (including information of the name, type and type of radar).
- For frequent passengers, all the information that is important for the airline, including the type of frequent passenger.
- The type of plane use for each trip.
- The base fare for each trip.
- The different dates must be added as temporary dimensions.

Step 4: Identify the measures in the fact table

According to the requirements, the measurements for each would be: the price of each trip, the points obtained (or used, depending on the type of base rate) per trip and the accumulated delay each way. The derived measures can be: total delay of each itinerary, price total of the flight, points of each itinerary, duration in minutes of each journey, etc.

Warning: Do not expect data sources to match 1:1 the requirements posed by a company or by your superiors in most cases. You will be required to compare and adapt your model to the data at hand following a best-effort approach.