Skyscanner Heatmap

Goals description

The main goal of this project is creating a tool for Skyscanner to ease the routes and airports comparison by different parameters, taking into account values like **user demand** and **flights provided** by airlines.

In any team of Skyscanner, user queries and providers data is compared in order to guess valuable trends.

Found that gap, a bunch of new ideas appeared. After some talks with product owners of different squads and some senior engineers, a promising idea showed up:

Comparison of **user demand** and **flights offered** by airlines, enabling finding **over-requested** routes or airports. Those routes or airports with more user demand than offers by the providers.

My team manages a huge amount of data: All flights planned for the next two years, this are more than 75 million records. The database of all user queries in the website or mobile application is even bigger (For instance, if there were only one query per visitor the database would have 4 million new records per day). Not much more information needed to say that this is **Big Data** problem.

With team's product owner help, we found some use cases for the processing of those 75 million routes and all user session's queries to get some significant results:

Provide a **visual tool** to find routes and airports with much **more demand than offer** and be able to observe the evolution of it through time:

- A route or airport with a lot of demand, but not enough offer to cover it, will be **over-requested**.

- A route or airport with much more offer, but not that amount of demand, will be **non-profitable**.

Project scope

Merging both data sources (providers and users) generates a lot of new valuable data with a lot of different application: From simply selling it to providers, to complex deep learning systems.

The final goal of this project is displaying the comparison in a simple Web UI for Marketing squads or tribes. This can be split in three smaller goals or components: The pipeline that processes the data, service giving proper comparison to the website and web user interface with the graphical representation.

This project will **not** have prices or quotes; carriers, cities and countries and the user will not be able to create, update or delete data.

Relevant concepts

First of all, this project will be developed following well defined software methodologies, Extreme Programming (AS, GPS, PES), so some knowledge about these field may be needed.

The application itself is a big data problem (CAIM). To get a good performance, it is needed to find bests algorithms (EDA) for data processing, language (LP), machines (SOAD) and systems where ran the software (AS).

Last but not least, this project is using a lot of data that can be interpreted from many different ways and can be used in much more ways, so it is important to define very well the risks, the requirements, the scope of the project (ER, GPS, PES).

From the beginning to the end it has to be explained and well documented for other interested employees of the company to understand what is this about (PES).

Software engineering project

In my whole student life, I have work in a lot of different project, from a two weeks project to a six months application.

I have also worked in professional companies, but it has never been as professional as Skyscanner. This company combines accuracy and professionalism in their software and agile, open minded and modern methodologies.

This allows me to say that from the beginning to the end, the development of this project is being different from almost all other software I have ever wrote.

Starting with an idea good enough to work on it for six months, write a great elevator pitch to convince the product owners and evaluating all the requirements are some examples of the professionalism the university taught me.

Apart from all this inception, the continuous deployment to demo every two weeks helps me to work properly in an agile methodology. The code needs to be well documented too, so other developers could take this project in the future and improve or adding new functionalities.

Talking about the code, Skyscanner changes a lot every day and despite these changes stills being the best travel aggregator. This is thanks to the code it is behind, it has a very good design, architecture and then it is very efficient. The code is being designed thinking about the future and not about the final product planed for delivering in three months.

To sum up, I can finally see almost all the knowledge I learnt in the software engineering major, summarized in one project.

Technical competences

As explained before, the Heatmap has relevant concepts about Software Engineering to take into account.

It will develop and evaluate **complex software systems**, it is a distributed application, luckily, Amazon Web Service save the development a lot of work.

Some **risks** may appear, like changes in the data source model, but since the company has strong internal contracts with all APIs it is a minor issue.

This application will be fully **distributed in the cloud**, the pipelines what process the data and the service that compares the sources will be distributed in the network.

All the processed data must be **stored in the most efficient** way, so the database is a critical part of the project. A lot of different databases should be evaluated and choose the appropriated one for this kind of data.

The comparison of the heatmap must be **reliable**, and the quality of the software must be good. It necessary to take this into account and have a proof (tests or metrics, for instance) that shows that the software is working fine.

Finally, this project may have an economic, social and ethic impact. The data it is using is very valuable and critical, you can take a lot of information out of it. Combining both sources, the economic and social impact becomes very important.