DataScientest

Holiday Itinerary

Program : Data Engineer

Difficulty : 8/10

Description:

The objective of the project is to create an application that proposes an itinerary according to certain criteria
The user of the application chooses areas / points of interest to visit during his next trip, as well as the duration of the stay and the app proposes a detailed itinerary optimizing his travel and stay time.
The project can be improved by also using the categories of the chosen places, or the ratings on TripAdvisor (for example).

Step	Description	Goal	Modules/ Masterclass/Templates	Conditions of validation
1	Collecting the data	With this site https://opentripmap.io/ https://holidayapi.com/ (we can choose another website in order to collect this kind of data) A dataset listing all the activities that can be carried out in a community, called a POI (Place/Point of Interest). Basically, we don't have the theme of the POI, this information is present in the URL, so we have to do some text-mining on it. We also have geographical data. We will have to do some webscraping on it to get this data.	131 - Text Mining Web Scraping (Selenium, Beautiful Soup)	CSV file JSON file Treatment explanation file (doc/pdf)
2	Data modeling	Organize the data into a relational database. It will therefore be necessary to think about how to segment the data into several tables.	142 - SQL	A relational database UML Diagram A file who creates and queries the SQL database .
		Implement a graph-oriented database with geographical coordinates. The aim would be to implement an interface to be able to apply graph theory to our problem	Neo4j	Same files for the Neo4j database.
3	Data consumption Deployment	Applying clustering to be able to make routes in the clustering, the aim would be to manage to have a route, then it is necessary to visualise a route via Dash. They will then be able to couple their dashboard with an API to use the Neo4j database. They will be able to recover the work of a previous group for the clustering part. It's the production part that should take precedence here.	DE121 Flask, FastAPI Dash, Plotly Docker	Jupyter Notebook API Dash app

4	Automation of flows (optional)	The activities are updated, so you have to collect regularly to have an up-to-date and useful application.	Airflow	Python file for Airflow
5	Defense	Demonstrate their application and explain the reasoning behind their project.	X	Defense Documentation

Bibliography :

Some links who will help you for the Machine Learning part :

- https://www.researchgate.net/publication/340909418_Improving_itinerary_recommendations_for_tourists_through_metaheuristic_algorithms_an_optimization_proposal
- https://towardsdatascience.com/using-unsupervised-learning-to-plan-a-paris-vacation-geo-location-clustering-d0337b4210de