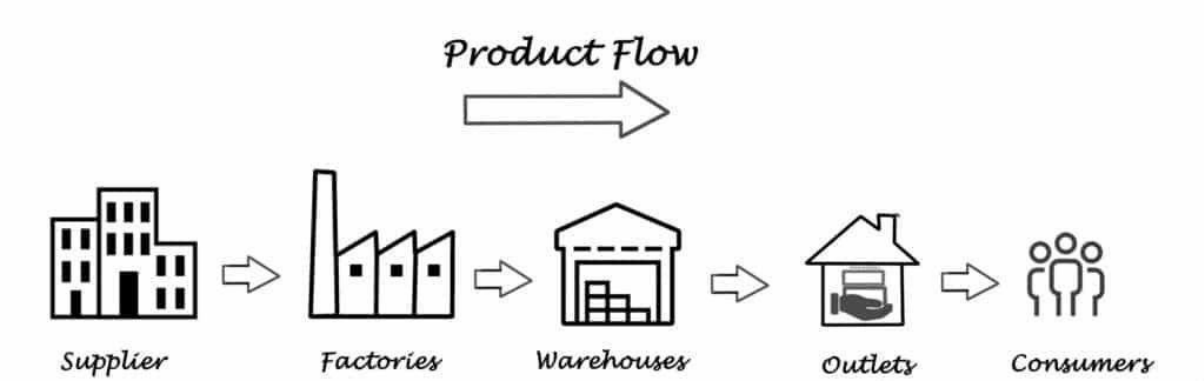
DataScientest

Supply Chain - Customer satisfaction

Program : Data Engineer

Difficulty : 8/10

Description détaillée:
The supply chain represents the stages of procurement, production process and distribution of goods.



Downstream of this process, it can be interesting to evaluate customer satisfaction for :

- *Quality study on the supply chain: design problems, delivery, inappropriate price, durability, etc.*
- *Study whether the product/service corresponds to the market's expectations.*
- *Synthesise feedback, improvements from customers.*
- *Assist in the response or redirection of dissatisfied customers...*

For many products/services, customer satisfaction is measured on the basis of comments and opinions on dedicated sites (Trustpilot, distributor site, twitter, etc.).

It can be long, tedious but important to read and analyse the verbatim comments which are essential to understanding customer satisfaction, but in the absence of tools that allow these opinions to be synthesised, the processes are generally based on sampling.

Step	Description	Goal	Modules/ Masterclass/Templates	Conditions of validation
1	Collecting the data	Collect two types of data through web scraping. One gathering general information about companies (the domain, the number of reviews, the Trustscore, the percentages on each class of reviews (the percentage of Excellent reviews))	131 - Text Mining Webscraping (Selenium, Beautiful Soup)	CSV file Treatment explanation file (doc/pdf)

		The companies are organised in themes https://www.trustpilot.com/categories/atm which list different companies in the field of atm(bank) The other grouping all the comments of a company with more than 10000 reviews (ShowRoom for example), with the information related to the review (number of stars, if the company has responded to the negative review)		
2	Data Modeling	Organize the first type of data into a relational database. It will therefore be necessary to think about how to segment the data into several tables. Implémenter une base de données orientée document avec les commentaires. Le but serait d'implémenter un Dashboard Kibana qui permettait de requêter sur ces commentaires et d'avoir un "suivi" de l'entreprise qui a été commentée.	142 - SQL ElasticSearch	A relational database UML Diagram A file who creates and queries the SQL database . Same files for a Elastic/Mongo/ DataBase
3	Data consumption	Do some fairly basic Machine Learning to do sentiment analysis. Then, recover the work of Data Scientist to make an API of these two models. We will couple the API with Dash to also have a "report" of the company	DE121 Flask, FastAPI Dash, Plotly	ML notebook API Dash app
4	Automation of flow	Automate data scraping and do it on a daily basis with Airflow. This data scraping will update the Dash/Kibana	Airflow	Dash/Kibana actualisé + Rapport Final
5	Defense	Demonstration of their application and explanation of the reasoning behind their project	X	Defense Documentation