

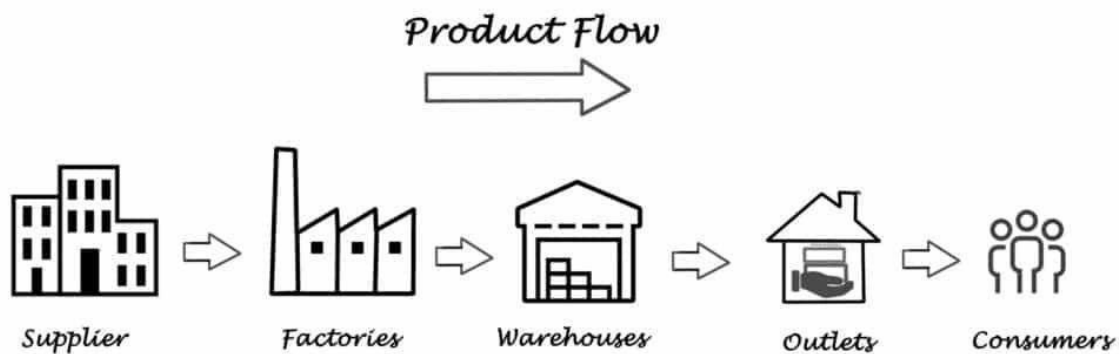
Supply Chain - Satisfaction des clients

Cursus concerné : Data Engineer

Difficulté : 8/10

Description détaillée:

La supply chain représente les étapes d'approvisionnement, du processus productif et de distribution de la marchandise.



En aval de ce processus, il peut être intéressant d'évaluer la satisfaction client pour :

- **Étude de qualité sur la supply chain** : problème de conception, livraison, prix non adapté, durabilité...
- **Étudier** si le produit/service correspond bien à l'**attente du marché**.
- **Synthétiser** les feedback, améliorations des clients.
- **Aider à la réponse** ou à la redirection des clients insatisfaits...

Pour de nombreux produits/services, la satisfaction des clients se mesure sur les commentaires, avis de sites dédiés (Trustpilot, site distributeur, twitter...).

Il peut être long, fastidieux mais important de lire et analyser les verbatim¹ qui sont essentiels à la compréhension de la satisfaction client, mais en l'absence d'outils qui permettent de synthétiser ces avis, les procédés sont généralement par échantillonnage.

¹ Il s'agit de la reproduction intégrale des propos prononcés par l'interviewé ; c'est-à-dire le compte rendu fidèle de l'avis du client.

Etape	Description	Objectif	Modules/ Masterclass/ Templates	Conditions de validation du projet
1	Récolte des données	<p>Récolter deux types de données en passant par le web scraping.</p> <p>Une regroupant des informations générales sur des entreprises (le domaine, le nombre d'avis, la note Trustscore, les pourcentages sur chaque classe de commentaires (le pourcentage d'avis Excellent))</p> <p>Les entreprises sont organisées en thèmes https://www.trustpilot.com/categories/atm qui listent différentes entreprises du domaine atm(banque)</p> <p>L'autre regroupant l'ensemble des commentaires d'une entreprise avec plus de 10000 avis (ShowRoom par exemple), avec les informations liées à l'avis(nombre d'étoile, si l'entreprise a répondu à l'avis négatif)</p>	(131 - Text Mining) 133 - Web Scraping	<p>Fichiers CSV</p> <p>Fichier JSON ?</p> <p>Fichier explicatif du traitement (doc/pdf)</p>
2	Organisation de la donnée	Organiser les données du premier type en une base de données relationnelle. Il faudra donc réfléchir comment segmenter la donnée en plusieurs tables.	142 - SQL (Architecture des données)	<p>Une base de données relationnelle</p> <p>Un fichier de requête SQL pour montrer que c'est bien fonctionnel</p> <p>Fichier pour créer la BDD ES</p> <p>Dashboard Kibana</p> <p>Requête pertinente sur ES</p>
		Implémenter une base de données orientée document avec les commentaires. Le but serait d'implémenter un Dashboard Kibana qui permettrait de requêter sur ces commentaires et d'avoir un "suivi" de l'entreprise qui a été commentée.	ElasticSearch MongoDB	
3	Consommation de la donnée	<p>Faire du Machine Learning assez basique pour faire de l'analyse de sentiment.</p> <p>Vous pouvez aussi faire un Dash pour réaliser un "compte rendu" de l'entreprise.</p>	DE 120 Dash, Plotly	<p>Notebook sommaire de Machine Learning commenté</p> <p>Dashboard avec Dash</p>
4	Mise en production	<p>Récupérer le travail de Data Scientist pour faire une API de ces deux modélisations.</p> <p>Il faudra aussi dockeriser votre projet pour qu'il soit prêt pour être mis en production</p>	Flask, FastAPI Docker	<p>API</p> <p>Docker-compose</p> <p>Dockerfile</p>
5	Automatisation du flux de la donnée	<p>ÉTAPE FACULTATIVE</p> <p>Automatiser le scraping des données et le faire à un rythme journalier avec Airflow. Ce scraping des données actualisera le Dash/Kibana</p>	Airflow Oozie	Dash/Kibana actualisé + Rapport Final
6	Soutenance	Démonstration de leur appli et explication du raisonnement effectué lors de leur projet	X	Soutenance Rapport



Bibliographie:

Il n'y a pas de bibliographie, mais on peut s'inspirer du projet Twitter Sentiment Analysis with Spark pour certaines parties.