

Contexte d'étude

Analyse de données Yelp



Yelp (<https://www.yelp.com>) est un site participatif d'avis sur les commerces locaux et de réseautage social. Le site comprend des pages consacrées aux établissements locaux, tels que des restaurants ou des écoles, où les utilisateurs Yelp peuvent soumettre un avis portant sur leurs produits ou services en ayant recours à un système de notation entre une à cinq étoiles. Les entreprises peuvent également mettre à jour leurs coordonnées, leurs heures d'ouverture et autres renseignements de base ou ajouter des informations particulières. En plus d'écrire des avis, les utilisateurs peuvent réagir à des avis, planifier des événements ou parler de leur vie personnelle¹.

Nous nous intéresserons dans ce contexte d'étude à un sous-ensemble des fonctionnalités offertes par le site Yelp au travers du jeu de données Yelp Open Dataset (<https://www.yelp.com/dataset>) mis à disposition par la société. Nous nous limiterons dans ce contexte d'étude aux avis donnés par les utilisateurs sur les restaurants, via un sous-ensemble extrait de ce jeu de données.

Le schéma 1 illustre le fonctionnement de la plateforme sur lequel nous nous penchons.



Figure 1: Fonctionnement de la plateforme participative d'avis Yelp

Les restaurants sont recensés avec différentes informations, certaines présentes pour tous les restaurants comme par exemple l'identifiant yelp, le nom, l'adresse, la ville, l'état, d'autres variables suivant les restaurants comme par exemple les styles d'ambiance proposés).

Les utilisateurs sont décrits par diverses informations, certaines présentes pour tous les utilisateurs comme par exemple l'identifiant yelp et le nom, d'autres variables comme la liste des années d'obtention du statut élite.

¹Source: <https://fr.wikipedia.org/wiki/Yelp>, accédé le 22/07/2022.

Les utilisateurs peuvent poster des avis (« reviews ») sur les restaurants constitués notamment d'une note globale (sur une échelle de 1 à 5) et d'un commentaire. Les différents avis sont horodatés.

Les utilisateurs peuvent posséder ou non une liste d'amis qui sont également des utilisateurs de la plateforme Yelp.

Trois fichiers de données au format JSON relatifs aux restaurants (`yelp_restaurants.json`), aux utilisateurs (`yelp_user.json`) et aux avis (`yelp_review.json`) sont disponibles en téléchargement (suivant les instructions communiquées par les enseignants). Ces fichiers de données sont utilisables comme sources de données pour alimenter la base de données définie suivant une modélisation adaptée au problème donné.