HANSER Florian  
MANCIAUX Romain

Projet hadoop

# Jeux de données

Le jeu de données est le suivant : https://www.kaggle.com/jtrofe/beer-recipes#recipeData.csv. Ce jeu de données est une liste des bières « homemade » avec leur origine, leur type de brassage et des informations comme quantité de sucre d'amorçage utilisée, la couleur ou l’amertume des bières. Il contient 75.000 entrées provenant du site Brewer's Friend.

# Description des données

Deux fichiers :

* StyleData.csv : fichiers de correspondance entre identifiant de type de bière et nom du type de bière

|  |  |
| --- | --- |
| Nom de la colonne | Description |
| Style | Nom du type de bière |
| StyleId | Identifiant du type de bière |

Les types de bière peuvent être répertorié en 7 grande Famille : Ale, IPA, Stout, Lager, Porter, Bittler, Cider.

* RecipeData.csv : fichier principal comprenant 75.000 recettes fait maison posté par les utilisateurs du site Brewer’s Friend.

|  |  |
| --- | --- |
| Nom de la colonne | Description |
| BeerID | Identifiant de la recette |
| Name | Nom de la bière |
| URL | Adresse de la recette sur le site https://www.brewersfriend.com |
| Style | Nom du type de bière |
| StyleID | Identifiant du type de bière |
| Size(L) | Quantité de bière produite avec la recette |
| OG | Poids spécifique du moût avant la fermentation |
| FG | Poids spécifique du moût après la fermentation |
| ABV | Pourcentage d’alcool |
| IBU | International Bittering Units pour mesurer l'amertume |
| Color | Couleur de la bière, claire vers foncé, ex 40 : noir |
| BoilSize | Quantité de liquide à ébullition |
| BoilTime | Temps durant lequel le moût est bouilli |
| BoilGravity | Poids spécifique du moût avant ébullition |
| Efficiency | Efficacité de l'extraction du moût de bière - extraire les sucres du grain pendant le moût |
| MashThickness | Quantité d’eau par livre de graine |
| SugarScale | Échelle pour déterminer la concentration de solides dissous dans le moût |
| BrewMethod | Techniques de brassage |
| PitchRate | Levure ajoutée au fermenteur - M cellules / ml / deg P |
| PrimaryTemp | Température au stade de fermentation |
| PrimingMethod | Méthode d’amorçage |
| PrimingAmount | Quantité de sucre d'amorçage utilisée |
| UserId | Identifiant utilisateur du site |

# Question

* Pour chaque famille de bières (API, ALE), par quelles méthodes de brassage obtient-on la plus forte ?
* Quelles sont les bières les plus sombres ayant consommé le moins d’eau par pound de grain ?