

Pràctica Kaggle APC

All products available on the Sephora
website

<https://github.com/josepmdc/ApC-Kaggle/>

Josep Maria Domingo Catafal
NIU: 1599946

Introducció al dataset

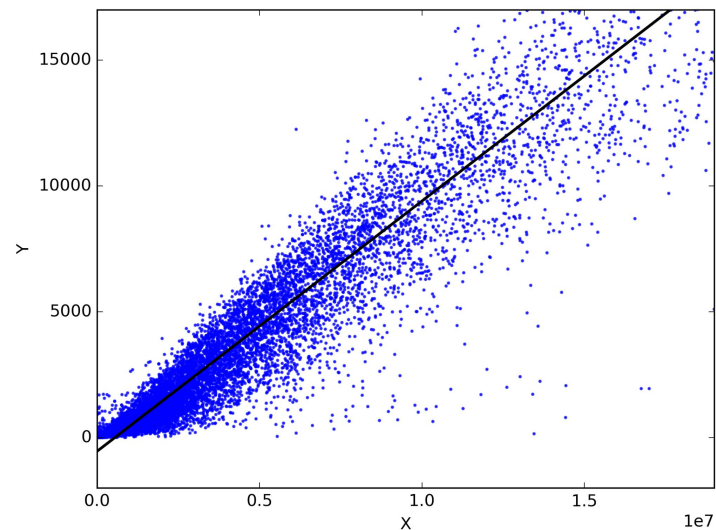
- 9168 productes
- 20 attributes
 - Marca
 - Categoria
 - Ingredients
 - Preu
 - **Núm. persones que els hi agrada el producte**

Preprocessing

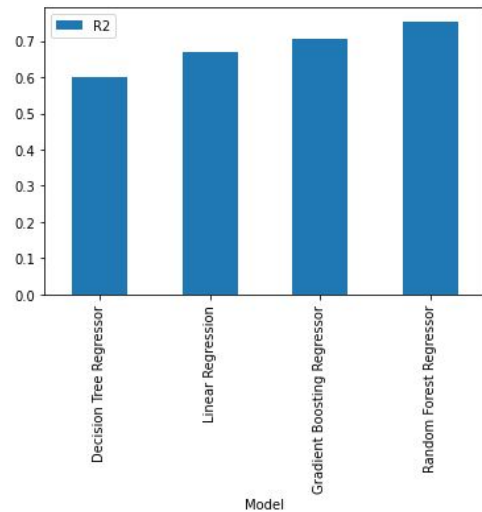
- Estandarització
- Codificació d'atributs categorics
- No hi havia valors nulls

Models aplicats

- Decision Tree Regressor
- Random Forest Regressor
- Linear Regression
- Gradient Boosting regression



Resultats



Conclusions i futures millores

- Productes pel llavis i cremes de la cara són el que agraden més als clients.
- Les edicions limitades són més populars que els productes normals.
- Random Forest és el model que funciona millor per el nostre dataset.
- Gradient Boosting és una altre opció viable com a alternativa al Random Forest.
- De cara al futur es podriem provar altres models o processar les dades de manera diferent a veure si milloren els resultat