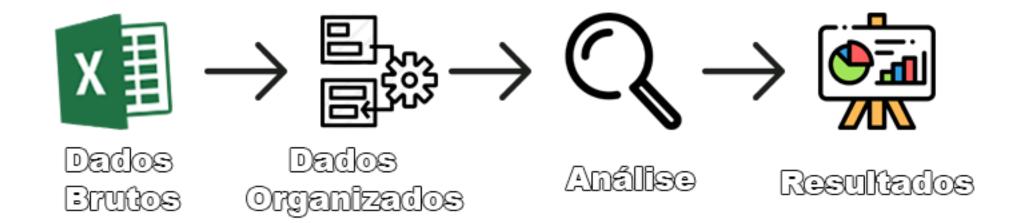


Etapas para resolução do problema



Dados Brutos XII



Os dados brutos desse projeto foi a base de dados disponibilizadas pela Ambev/ABInBev contendo:

- 73.671 linhas, que trazem informação a respeito de pedidos realizados contendo a data, número de pedidos, ID do comprador, receita liquida por produto vendido, marca, submarca, geolocalização, recipiente entre outros vários dados que serão utilizados para fazer a filtragem colaborativa.
- Também será utilizado uma base de dados a ser montada com as características dos produtos da Ambev/ABInBev para filtragem baseada em conteúdo.

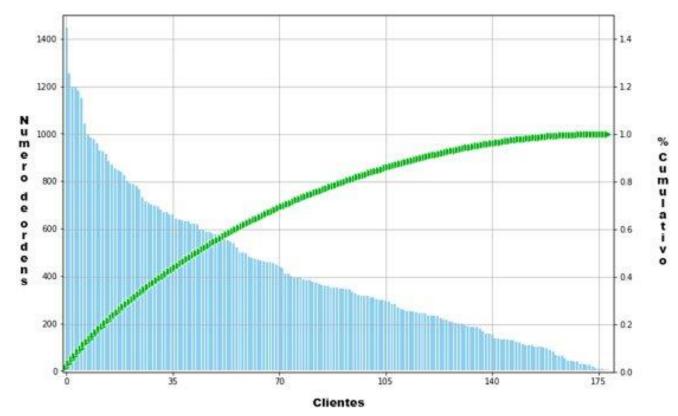
Dados organizados 島線



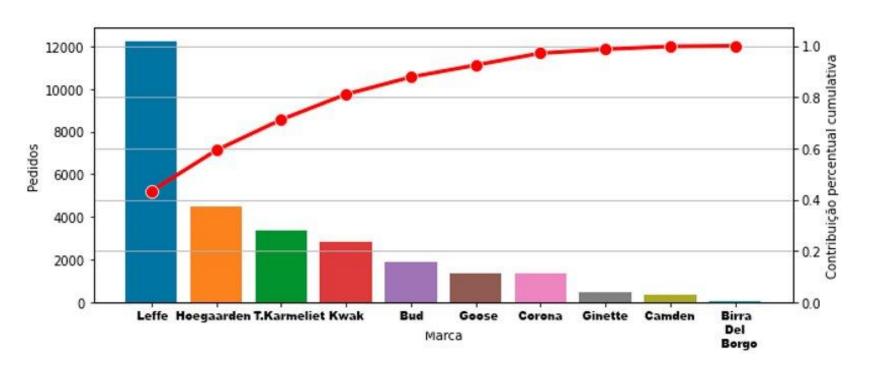
Das informações extraídas da base de dados disponibilizadas inicialmente mostraram especialmente relevante para o processo a geolocalização, quantidade de ordens, itens enviados por ordem, quantidade de clientes, marcas, submarcas, receita liquida por produto vendido e volume do recipiente.

Análise Q

Foi identificado 179 clientes com pontos de vendas espalhados pela França e divididos em diversos núcleos. Foram realizados ao todo 4.220.713 ordens, das quais 80% são representadas por metade dos clientes, comparada com a análise de Pareto.



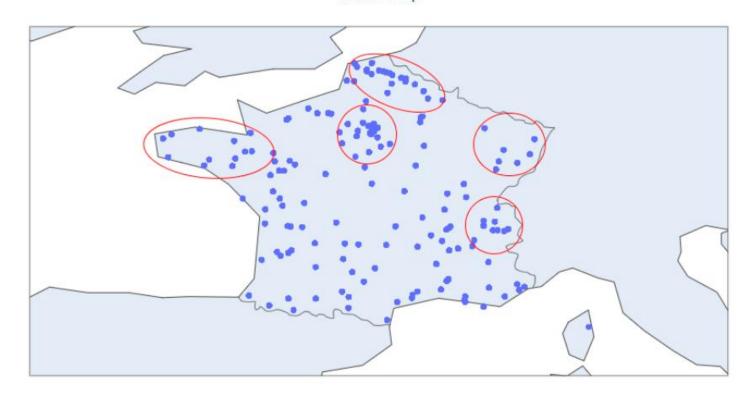
Do total de pedidos, 45.354 contêm marcas e submarcas. Nota-se a preponderância da Leffe, representando c. 43% do volume total dos pedidos, seguido pela Hoegaarden que representa apenas cerca de 15%. A Leffe também apresenta a maior variedade de submarcas, sendo a Blonde a mais comercializada.



Os recipientes de 6 litros representam c. 43% dos dados constantes, seguido dos de 330 ml, que representam cerca de 14,67%. Futuramente se aprofundará na relação das marcas com o volume dos recipientes.

Todos os 163 pontos geográficos se encontram em território francês, apresentando algumas concentrações nas regiões de Flandres, île-de-France, Borgonha, Auvérnia, Bretanha, dentre outros. Esses núcleos comporão a base principal da filtragem colaborativa, que dispensarão as informações distantes, poupando tempo e esforço computacional. Para regiões difusas como de Toulouse, aplicar-se-á um algoritmo de clusterização.

World map



MACO/HL é constante para um mesmo tipo de cerveja para um mesmo recipiente, submarca e variedade, e.g. as cervejas Leffe Royale de 6 litros da variedade IPA apresentam 211,63 MACO/HL.

Resultados 🚎



As análises preliminares indicam que há uma concentração da maioria dos pedidos em uma pequena parcela de clientes, o que aponta para pelo menos dois padrões de consumo distintos.

Aliado a isso, a dispersão geográfica virá como outra ferramenta para aumentar a acurácia das recomendações que serão sugeridas, visto haver núcleos de pontos de venda.

A distribuição temporal da quantidade de ordens será levada em conta para corroborar para a tipificação dos clientes ABInBev e garantir que as sugestões sejam as mais adequadas.

A variável Maco/hl será considerada pelo algoritmo por visar o maior lucro nas vendas. Junto disso, o fato de que cerca de 43% das bebidas vendidas são da marca Leffe e de que 174 dos 179 (97%) clientes ordenaram ao menos 1 vez a mesma, tal produto consistirá no ponto de partida da maioria das recomendações.

Para o processo de agrupamento por geolocalização, pensa-se em utilizar-se o algoritmo K-nn para identificar os grupos pela proximidade.

Será criado um sistema de recomendação próprio, que terá como inspiração o Recommender Systems I – ARL disponível no Kaggle, que comporte as variáveis e atenda o que foi pedido.