

Desafio Ecommerce Events

O notebook 'ecommerce_events.ipynb' foi implementado em Python 3 e utiliza as bibliotecas pandas e matplotlib para resolver 5 desafios e apresentar suas respectivas soluções.

Para executar, é necessário extrair os dados do arquivo 'dados.zip', obtendo uma pasta 'dados' que contém o arquivo 'ecommerce_events.ndjson'.

1. What was the total revenue?

Para responder essa pergunta, foi realizado um filtro para buscar todos os eventos que eram do tipo "transaction", depois pega os itens de cada transação e multiplica o preço pela quantidade, resultado no total de cada item. Somando tudo dá a receita total, respondendo assim a pergunta.

2. What percentage of visitors used a mobile device?

Para apresentar a solução desse desafio, os dados foram filtrados pelo tipo de dispositivo igual a 'mobile', dividido pelo número total de eventos e multiplicado por 100 para obter o percentual.

3. What search query had the highest click-through rate?

Primeiramente, busca os eventos do tipo 'product' e 'search', retirando todos os que possuem o atributo 'searchItems' igual a null, ou seja, buscas que não retornaram nenhum resultado. Depois, é filtrado os eventos de busca e de produto que apenas possuem visitantes em comum e é retirado todos os eventos de 'search' que possuem menos de 15 instâncias. Após esse filtro, é feito outro para restar apenas os produtos que foram retornados pelo 'searchItems' e vice-versa.

Todos esses filtros eliminaram vários dados para, logo depois, poder realizar a contagem de clicks sem demorar tanto. Depois da contagem, os eventos de busca são agrupados pela 'query' contando quantos clicks cada uma teve.

Para poder ficar mais fácil de visualizar, apenas as queries que possuem mais de 250 clicks são mostradas no gráfico.

4. When is the site most busy?

Para saber em que parte do dia o site é mais acessado, foi transformado o atributo 'date' para o horário e depois conta quantos acessos teve nesses horários para formar um gráfico. Já para saber o dia da semana, foi necessário converter o atributo 'date' para o dia da semana e depois contar os acessos para formar o gráfico.

5. What is the share of revenue among categories brought by Campaign_2?

Primeiramente, filtra os eventos do tipo 'product' que são da campanha 2 e eventos do tipo 'transaction'. Depois, é filtrado os eventos de transação e de produto que apenas possuem visitantes em comum. Após esse filtro, é feito outro para restar apenas os produtos que estão tanto no evento de transação quanto no de produto.

A receita é calculada multiplicando a quantidade do produto pelo preço dele. Após o cálculo, uma busca é realizada para pegar a categoria de cada produto para agrupar o valor dos itens pelas categorias.

Depois, é calculado o percentual do preço total de cada categoria, dividindo pela soma de todas as categorias e multiplicando por 100.

6. Estimate the impact of unavailable products

Foram filtrados todos os produtos indisponíveis do evento do tipo 'product' e contando quantos produtos indisponíveis os eventos do tipo 'search' que possuem no 'searchItems'. O percentual de produtos indisponíveis em cada evento do tipo 'search' foi calculado dividindo a quantidade de produtos indisponíveis pela quantidade total de produtos no 'searchItems' e para estimar o impacto dos produtos indisponíveis, foi dividido a quantidade de buscas que tiveram mais de 30% e mais de 60% de produtos indisponíveis pela quantidade total de buscas, resultando no percentual de buscas que tiveram mais de 30% de produtos indisponíveis no resultado da busca e no percentual de buscas que tiveram mais de 60% de produtos indisponíveis no resultado da busca.