

Relatório

Análise e Enriquecimento

Palhoça, 29 de outubro de 2024

SUMÁRIO

| | | |
|---|---------------|---|
| 1 | Sumário | 2 |
| 2 | Implementação | 2 |
| 3 | Resultados | 3 |

1 Sumário

A análise teve início com uma abordagem descritiva simples, direcionada para séries temporais, dada a natureza temporal das requisições. Essa exploração inicial revelou picos de requisições com características não orgânicas, esse entendimento pode ter surgido de uma compreensão limitada sobre o negócio. Diante desse cenário, algumas ideias foram descartadas, como a previsão da série, enquanto outras foram formuladas, incluindo a detecção de anomalias e a análise de "uplift" em picos com causas conhecidas. Essas abordagens permitem prever demandas futuras e facilitar o preparo da infraestrutura de servidores. Contudo, para proporcionar resultados mais robustos e aplicáveis, o foco foi em enriquecer os dados com informações de renda dos setores censitários. Essa escolha foi devido a natureza do negócio, pois o produto "anúncio" demanda segmentação por renda para otimizar sua venda.

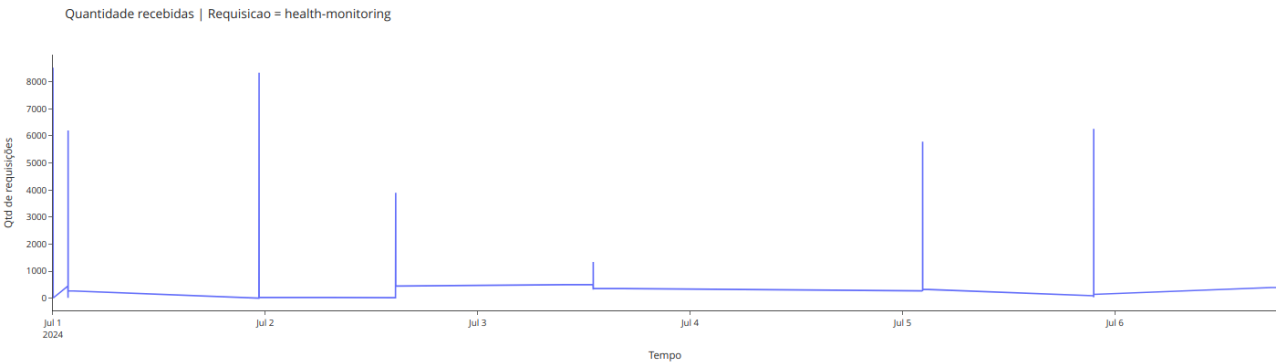
Além do enriquecimento por renda, essa abordagem permite explorar outras formas de segmentação, como o uso de dados demográficos de gênero, contribuindo para aprimorar a segmentação de anúncios. Com essas informações, é possível realizar clusterização procurando uma maior eficácia das estratégias de entrega de marketing e publicidade.

2 Implementação

Para a implementação deste projeto, foram utilizadas diversas bibliotecas que facilitaram a manipulação e a visualização dos dados. A biblioteca Pandas para a manipulação e análise de dados tabulares. Além disso, a biblioteca GeoPandas foi empregada para lidar com dados geoespaciais, possibilitando a leitura e o processamento de arquivos shapefile, que contêm informações geográficas importantes, como coordenadas de setores censitários. A integração com Shapely permitiu a criação de objetos de geometria, facilitando a manipulação das coordenadas geográficas em análises subsequentes. Para visualização, foi utilizado Plotly Express.

A análise começou com o entendimento dos dados e visualizações básicas, considerando que minha ideia inicial era trabalhar com séries temporais. As séries que se mostraram mais interessantes para análise, a princípio, foram as de quantidade de requisições, devido às possibilidades de insights que os dados, como região e tipo de requisição, poderiam oferecer. No entanto, essa ideia foi rapidamente frustrada pelo formato da série, que apresenta certa sazonalidade com picos, mas que parecem não ser orgânicos, ou seja, não ocorrem de maneira natural. Esse tipo de sinal é difícil de prever.

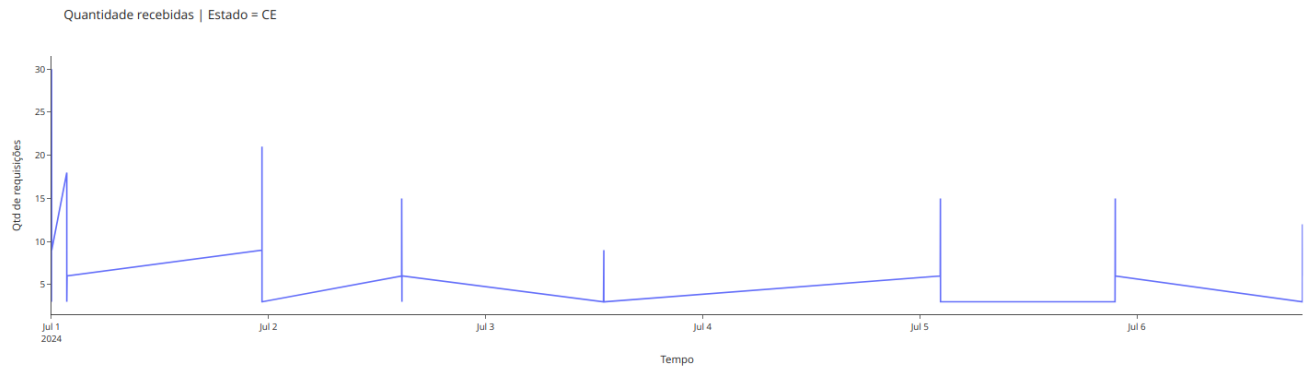
Figura 1: Série temporal da Requisição health-monitoring



Todavia, ao desagregar as séries temporais, poderia ser possível identificar as séries responsáveis por "sujar" a série de maior agregação, possibilitando a previsão dessas séries mais

segmentadas. No entanto, ao analisar as séries por tipo de requisição, local de navegação e estado, não foi possível encontrar a causa dos picos. Diante disso, a abordagem foi alterada, também considerando o tempo limitado disponível para a análise..

Figura 2: Série temporal aberta por estado



Observando o negócio e as possibilidades, optou-se por enriquecer a série com dados públicos. A demonstração foi realizada com dados do Paraná, já que este estado apresenta o maior número de requisições. O enriquecimento incluiu a renda média do setor censitário, fornecida pelo IBGE. Esse tipo de dado pode ser utilizado para clusterização e compreensão do mercado, com o objetivo de segmentar melhor as estratégias de marketing. Além disso, essa abordagem abre possibilidades para agregar outras informações futuras, como gênero, escolaridade, entre outras.

3 Resultados

Os resultados obtidos foram satisfatórios dentro do período estabelecido e em relação ao que foi proposto. Existem diversas possibilidades de implementações futuras para o enriquecimento dos dados, conforme mencionado anteriormente. Em relação às séries temporais, ainda é viável investigar mais a fundo suas componentes, criar variáveis dummy para momentos de alta, entre outras técnicas que possam auxiliar na previsão. No entanto, considerando o escopo de um teste técnico, não seria possível explorar todas essas abordagens. Além disso, o enriquecimento com dados censitários de renda foi efetivamente realizado.