

Relatório

B4 Bank

Análise de dados e
predição via
Machine Learning

Matheus Terra

Relatório B4 Bank

Aplicou-se análise de dados sobre o desempenho do aplicativo financeiro B4 Bank e após a análise, utilizou-se um método de Machine Learning baseado em Support Vector Machines (SVR) para dimensionamento da quantidade de visitas ao app e instalações para as próximas duas semanas.

01

Contextualização do Relatório

01- Introdução

O B4 Bank representa um aplicativo do setor financeiro que tem como objetivo conseguir uma maior visibilidade e retenção de usuários de forma orgânica. Como esse mercado tem crescido muito e é muito competitivo, resolveu-se investir em ASO com o RankMyApp, focando seus esforços não só em visualizações, mas também em instalações, ou seja, não apenas os usuários visualizarão o app na loja como também irão instalá-lo e utilizá-lo por um longo período de tempo. (Retenção de 15-30 dias).

02

Objetivos da Análise

Através do nosso sistema, recebemos informações relacionadas às instalações, concorrentes, performance das palavras-chave (keywords), notas e comentários deixados na loja.

02- Objetivos

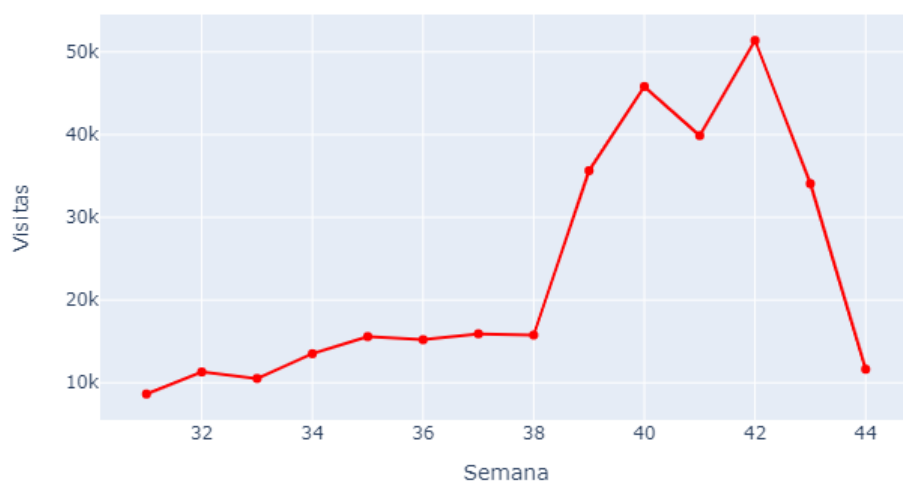
O cliente deseja ter insights sobre o que os usuários estão comentando na loja e qual a sua situação diante dos seus concorrentes. Realize um estudo utilizando Python para fazer a limpeza e manipulação de dados e crie um modelo de machine learning com uma das bases. Depois crie um relatório ou dashboard com a ferramenta que desejar para apresentar os dados.

Para alcançar os desafios almejados, planeja-se por meio da análise de dados, **identificar aspectos atrelados à maior visibilidade e cases de maior retenção. Filtrar ideias dos comentários e utilizar SVR para predição do fluxo orgânico para as próximas duas semanas.**

Análise de Dados

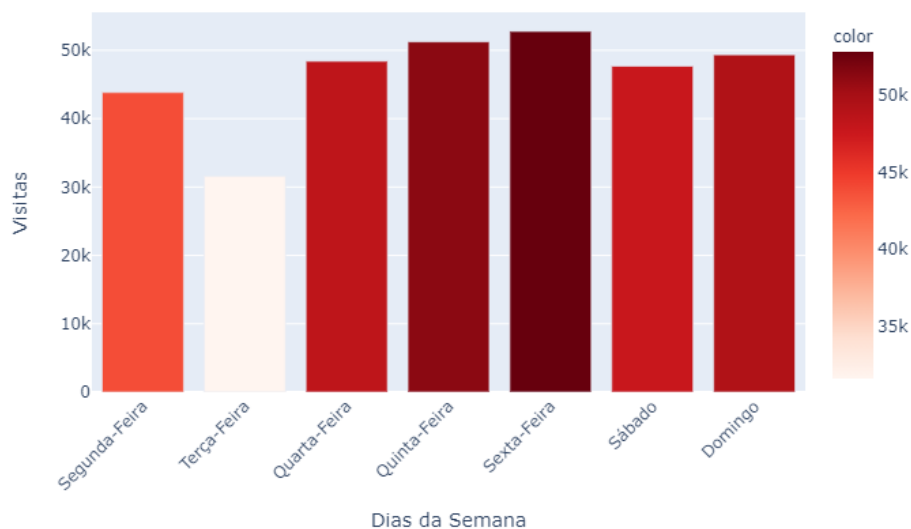
Baseado no dataset disponível do cliente, faz-se necessário como primeira análise, uma interpretação da quantidade de visitas que o app recebeu em sua loja, no decorrer do tempo. Para tal, realizou-se uma manipulação dos dados para agrupamento por semana, havendo uma estimativa mais compacta e que não leva em consideração fatores individuais de cada dia, mas sim um levantamento semanal. Ressalta-se que a numeração da semana, corresponde ao número que a semana recebe no ano.

Visitas ao APP por semana



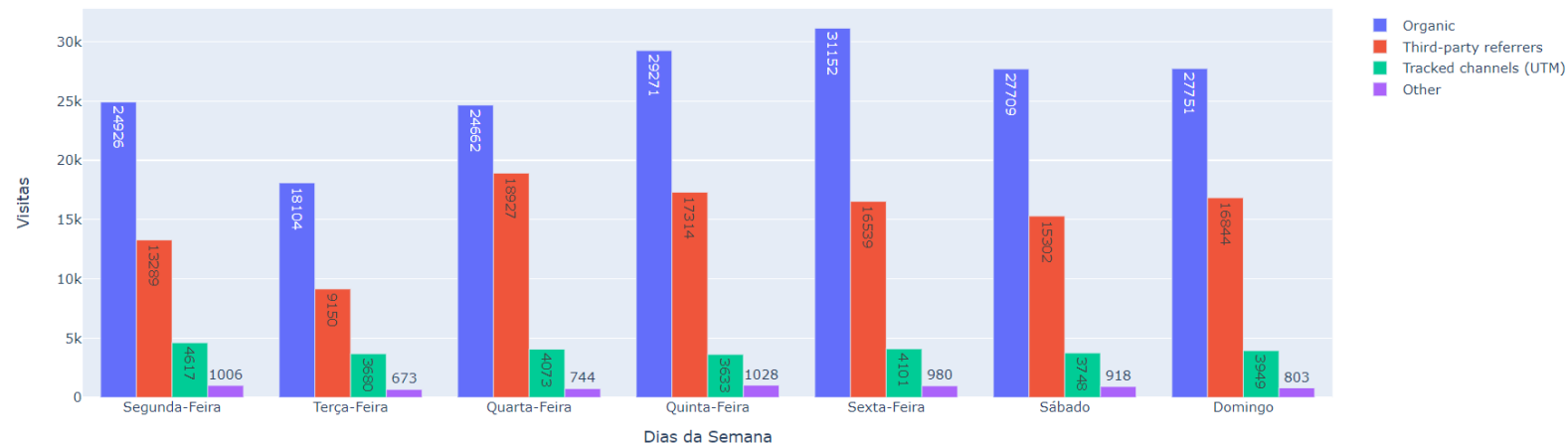
Tais informações são relevantes para indicar a tendência de volume de acessos ao decorrer do tempo. Afim de estimar as ocorrências futuras. Após a análise das visualizações ao decorrer do tempo, é realizado a quantificação baseada nos dias da semana:

Visitas por dia da Semana



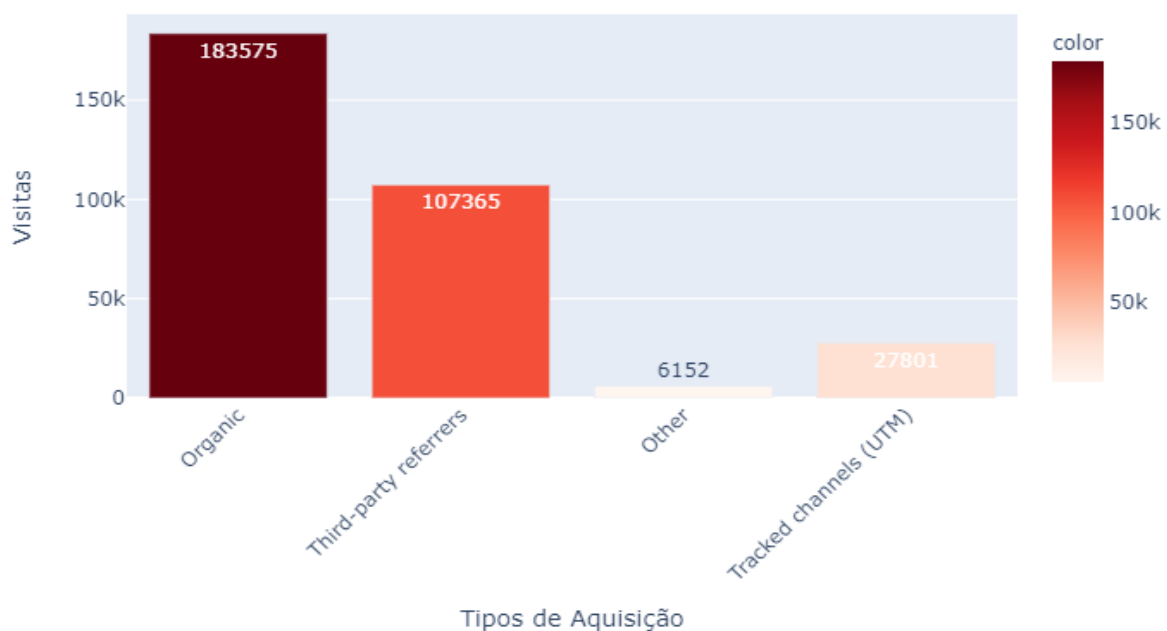
Uma consideração importante é que para criação dos gráficos, foi utilizado a biblioteca Plotly, do python, o que torna o gráfico interativo, ao passo que ao mover o cursor sobre as barras ou pontos, é apresentado informações a respeito.

Visibilidade por dia da Semana



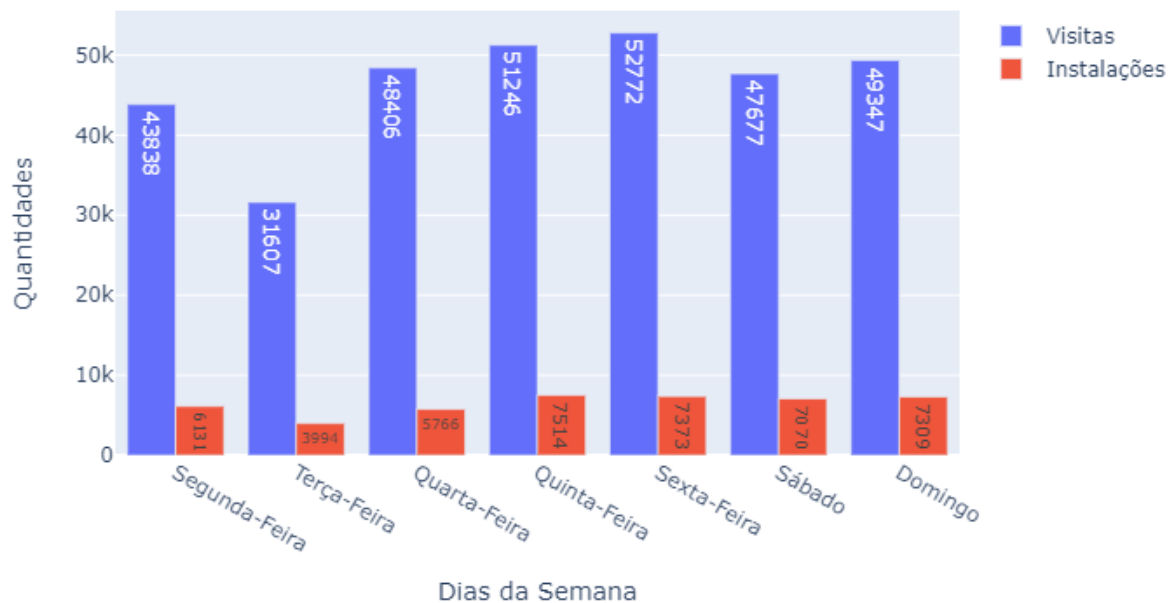
O gráfico acima indica a quantidade de visitas por dias da semana separados por categorias de canais de aquisição. De posse destas informações é possível identificar a origem do tráfego, assim como a intensidade por dia da semana. Em seguida, verificou-se a origem do tráfego no decorrer de todo período analisado, visando identificar pontos fortes e possíveis áreas de investimento para otimização das estratégias de divulgação:

Visitas por método de Aquisição



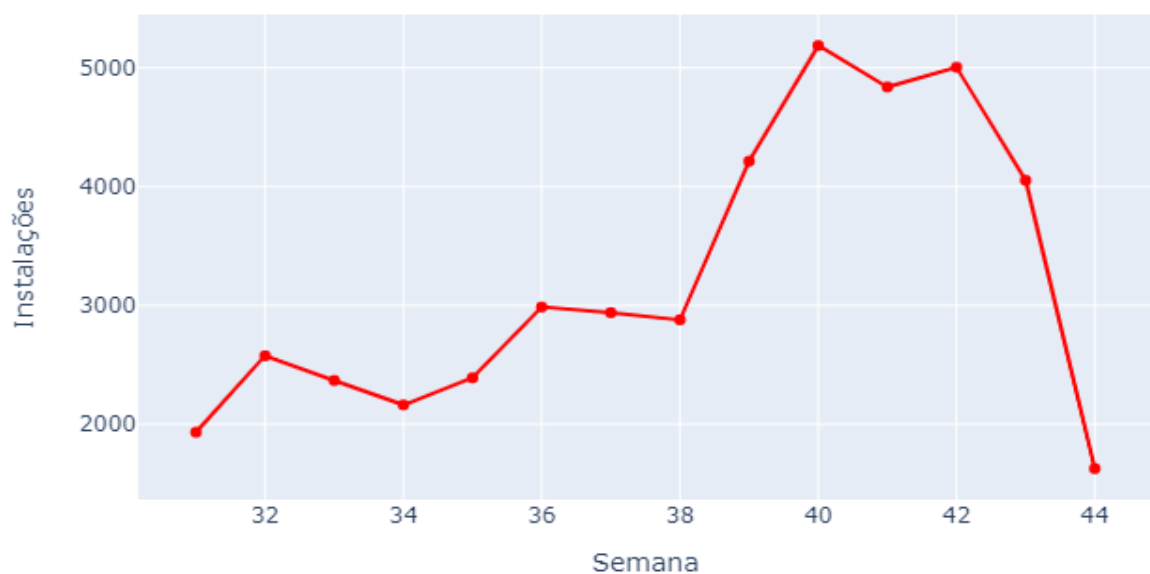
Ainda considerando uma visão total diária, é apresentado um comparativo entre visualizações do App X instalações:

Visibilidade X Instalações



O intuito do gráfico acima é indicar visualmente a taxa de aceitação do app em cada um dos dias da semana.

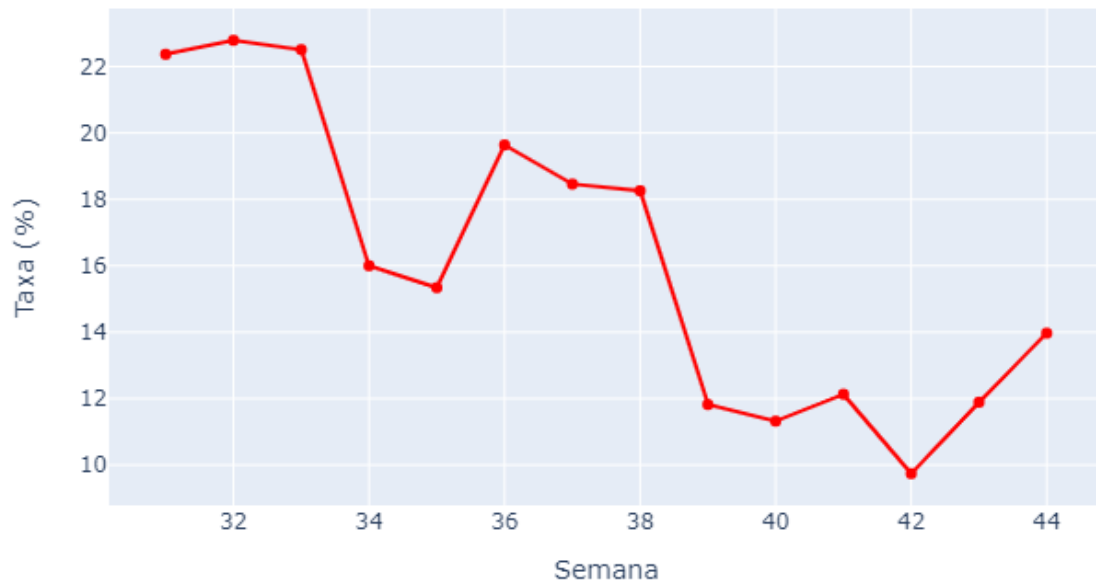
Instalações do APP por semana



Posteriormente, apresentou-se uma linha do tempo em escala semanal para indicação do histórico e tendência de instalações. Tais linhas-do-tempo são cruciais para identificação de motivos para crescimento do engajamento, atrelado à fatores sazonais.

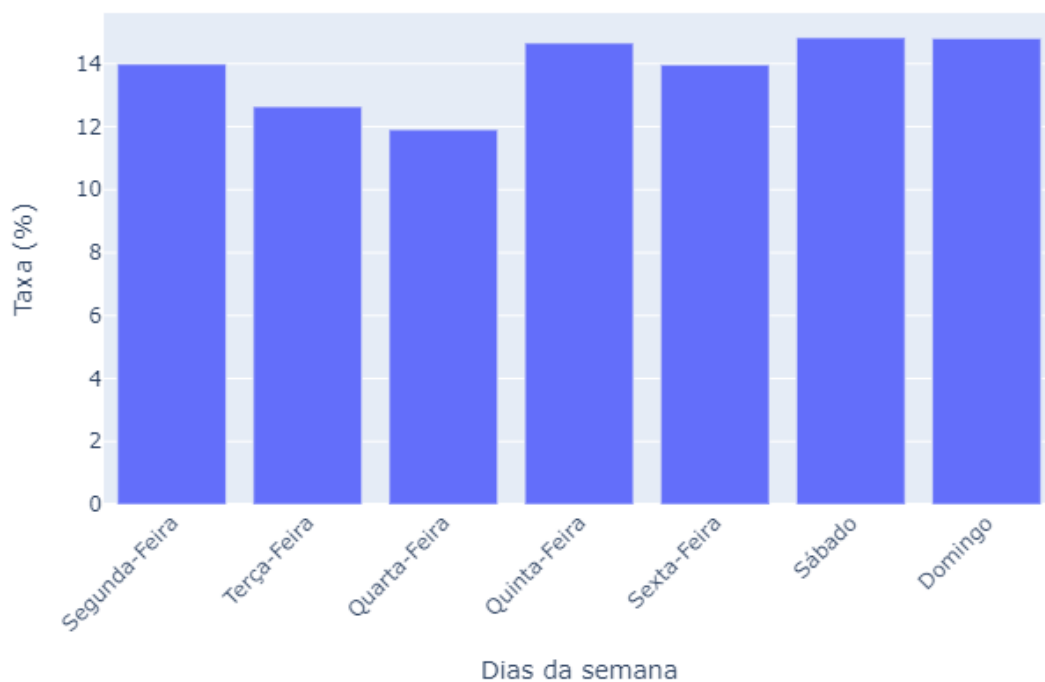
Por conseguinte, identificou-se uma linha do tempo em relação à taxa de conversão entre as visualizações X instalações no decorrer das semanas.

Taxa de Conversão do APP por semana

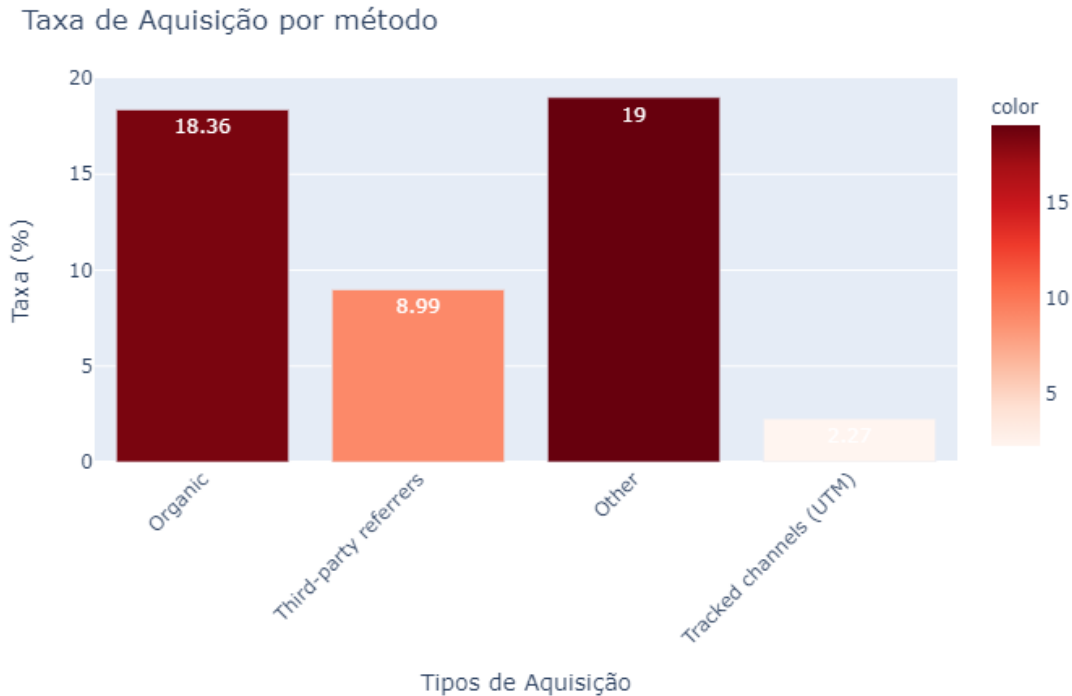


É indicado também a taxa por cada dia da semana afim de indicar um padrão de maior engajamento.

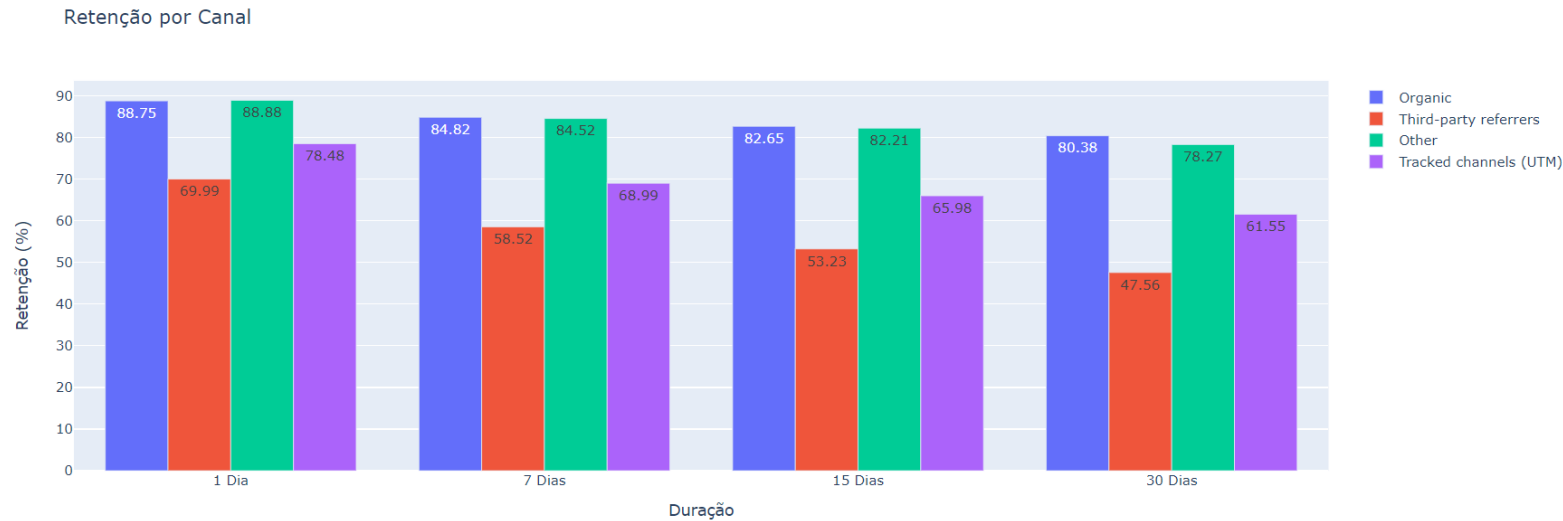
Média de Taxa de Aquisição por dia



O último estudo que trata da conversão, aponta a taxa baseado em canais de aquisição. Informação importante para identificar qual canal é mais efetivo afim de trabalhar em novas estratégias de marketing.



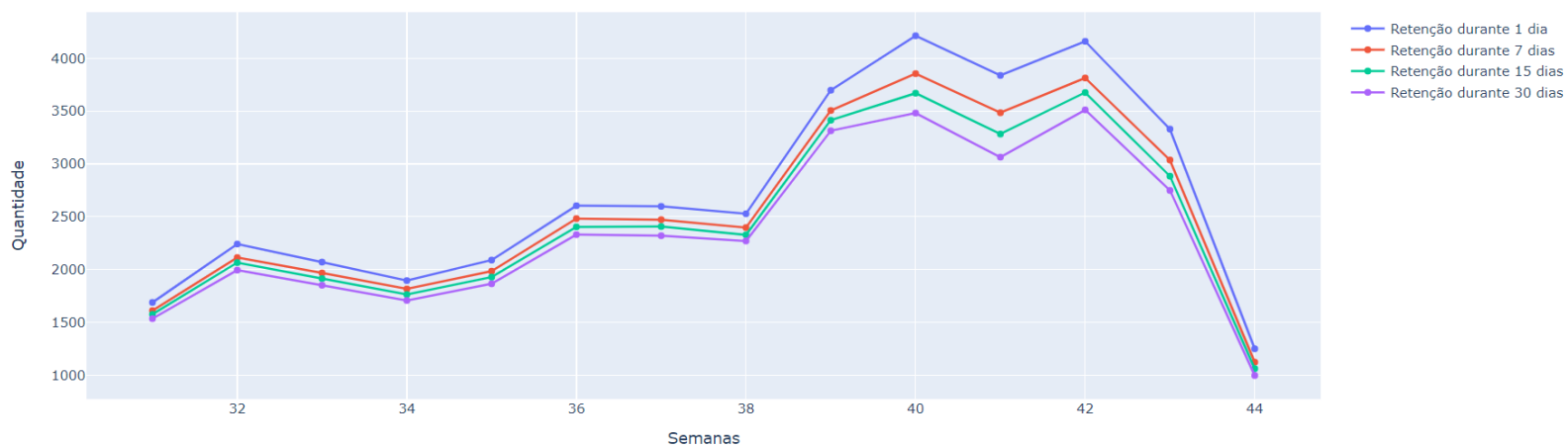
Outra informação imprescindível, além da taxa de conversão, diz respeito à retenção. Tal parâmetro aponta o período de tempo que as instalações foram retidas, variando entre períodos de 1 dia, 7 dias, 15 dias e 30 dias. Sendo o objetivo de reter o usuário por maior período de tempo.



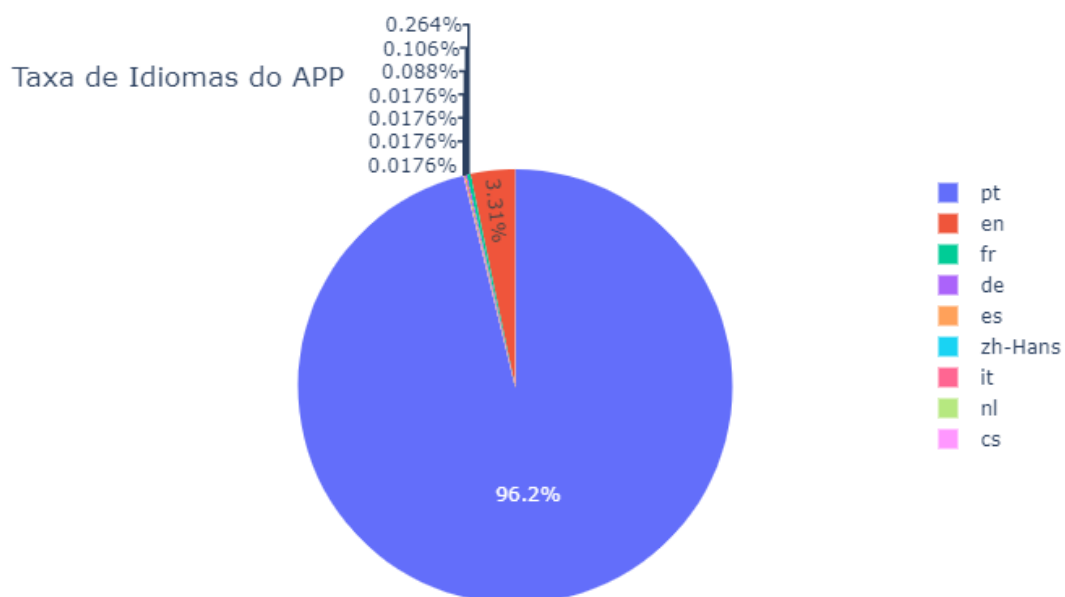
O gráfico acima aponta a porcentagem de retenção baseado na quantidade de instalações por cada um dos canais de aquisição.

É apresentado também linhas de tendências indicando a quantidade de retenções, baseadas no histórico semanal:

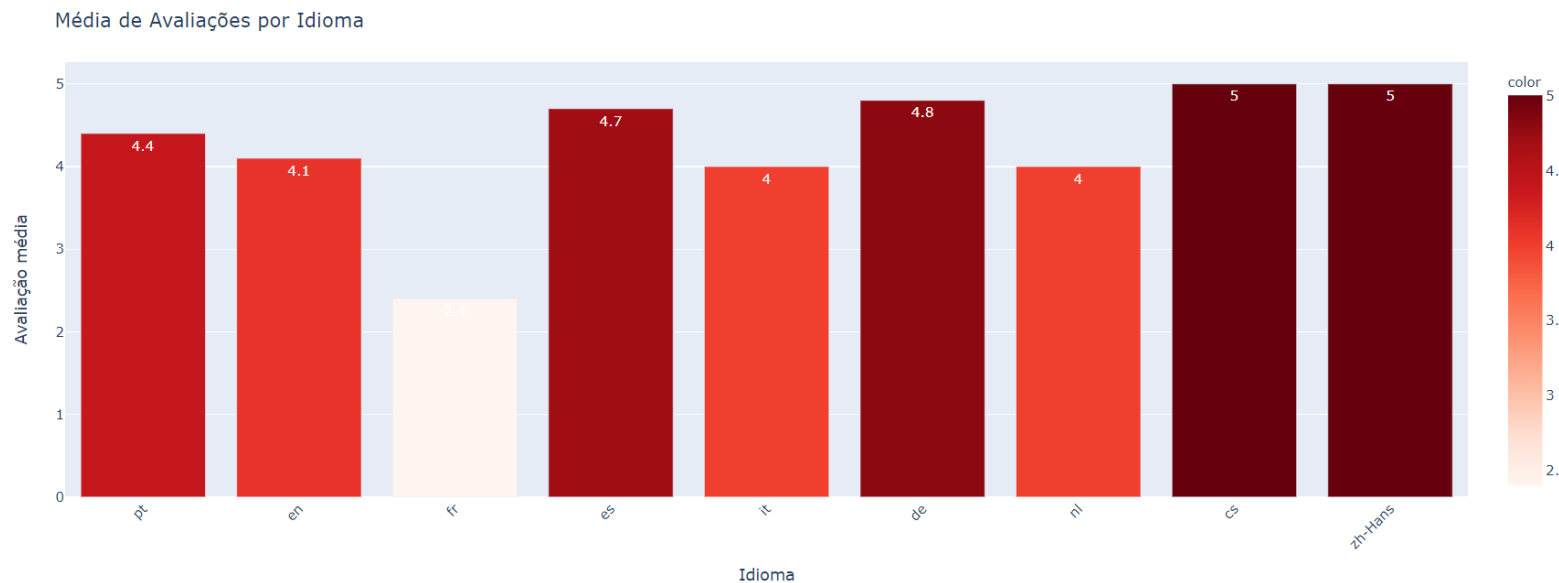
Retenção por Semanas



Partindo de uma análise dos usuários e das instalações, identifica-se o percentual de idiomas utilizados pelo app, sendo esta a única informação que permite estimar (de forma precária) uma possível localização por país, dos usuários.

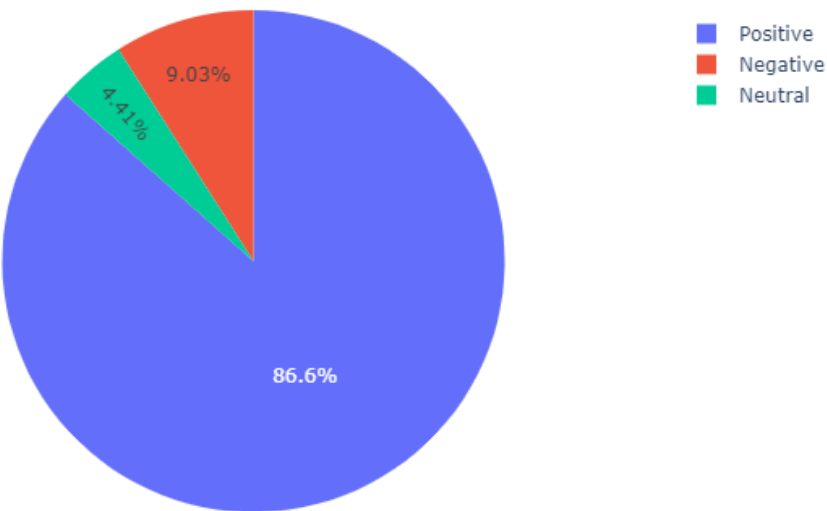


São apresentadas pela análise abaixo, médias de avaliações do app, por cada categoria de idiomas afim de traçar geograficamente índices de rejeição do app.



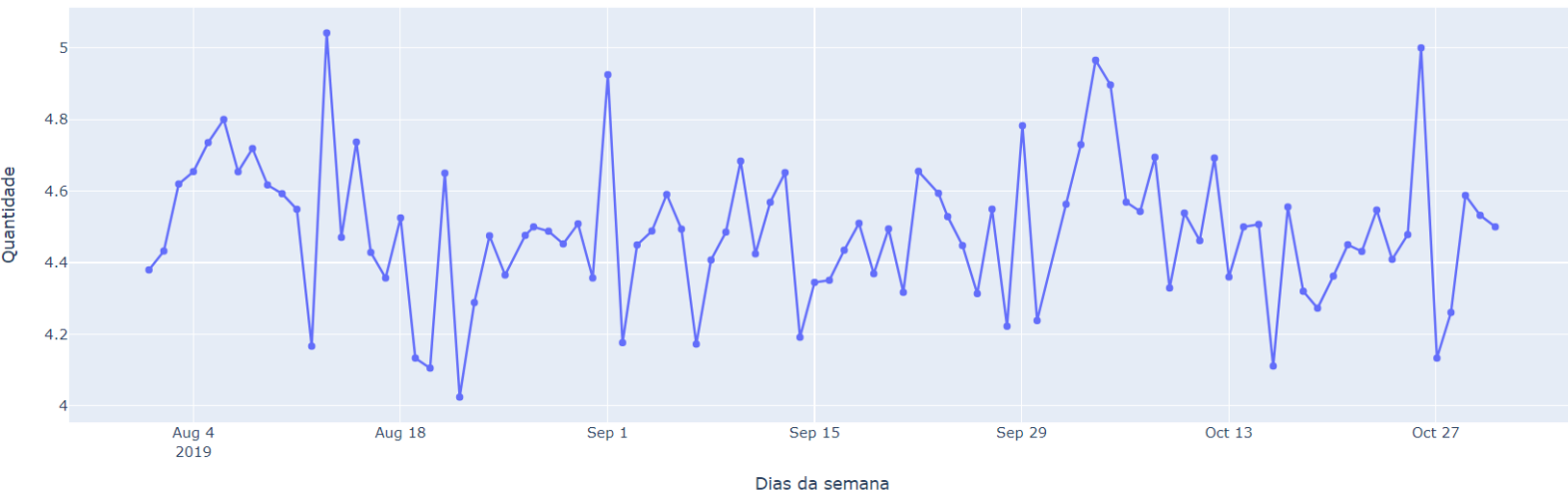
É também estimado valores percentuais de sentimentos dispostos no dataset. Sendo possível levantar o grau de satisfação dos usuários.

Taxa de aceitação dos Sentimentos disponíveis



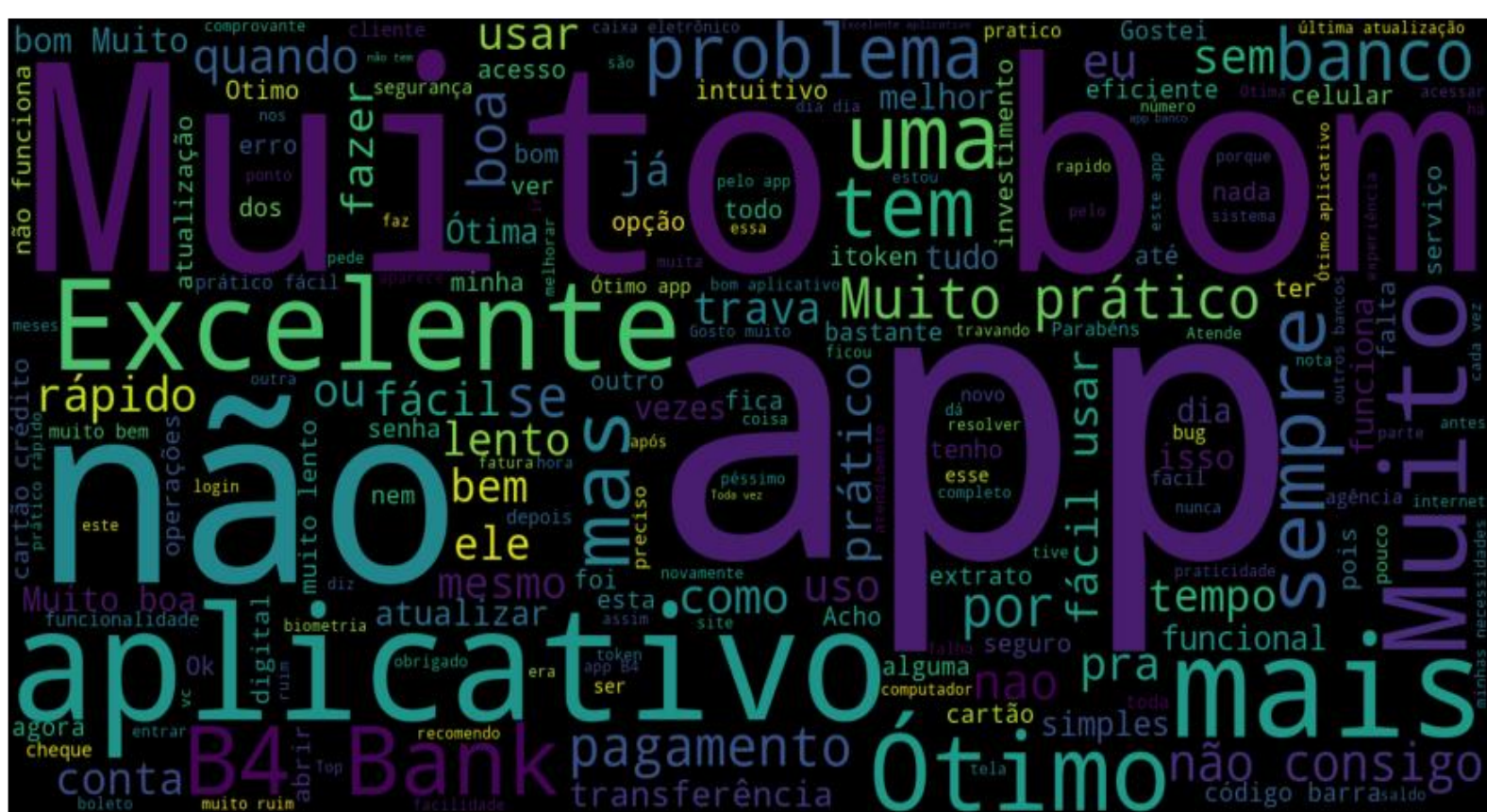
Foi considerado também uma linha do tempo diária contendo a média das avaliações em cada um dos dias, para cruzamento com outros parâmetros apresentados anteriormente:

Variação da Taxa ao longo dos dias



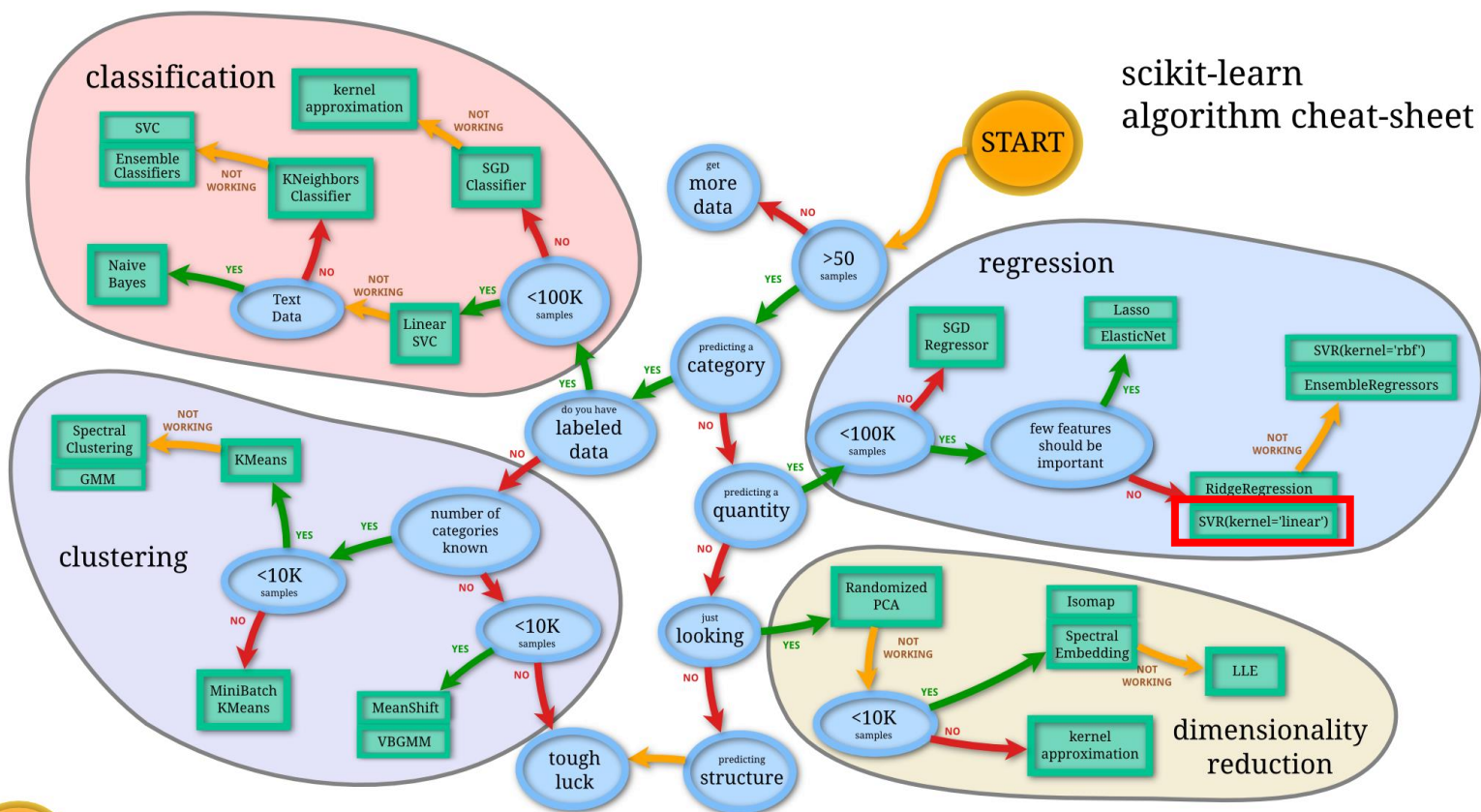
Para identificar insights contido nos comentários, a maneira mais simples, e que surte uma eficiência comprovada (vide canais noticiários) é por meio do Word Cloud, uma técnica que identifica a intensidade de termos e palavras utilizadas no decorrer dos comentários, e apresenta-os em forma de nuvem com diferentes tamanhos representando a incidência do termo/palavra.

Utilizou-se então tal conceito para retratar as opiniões comentadas dos usuários:



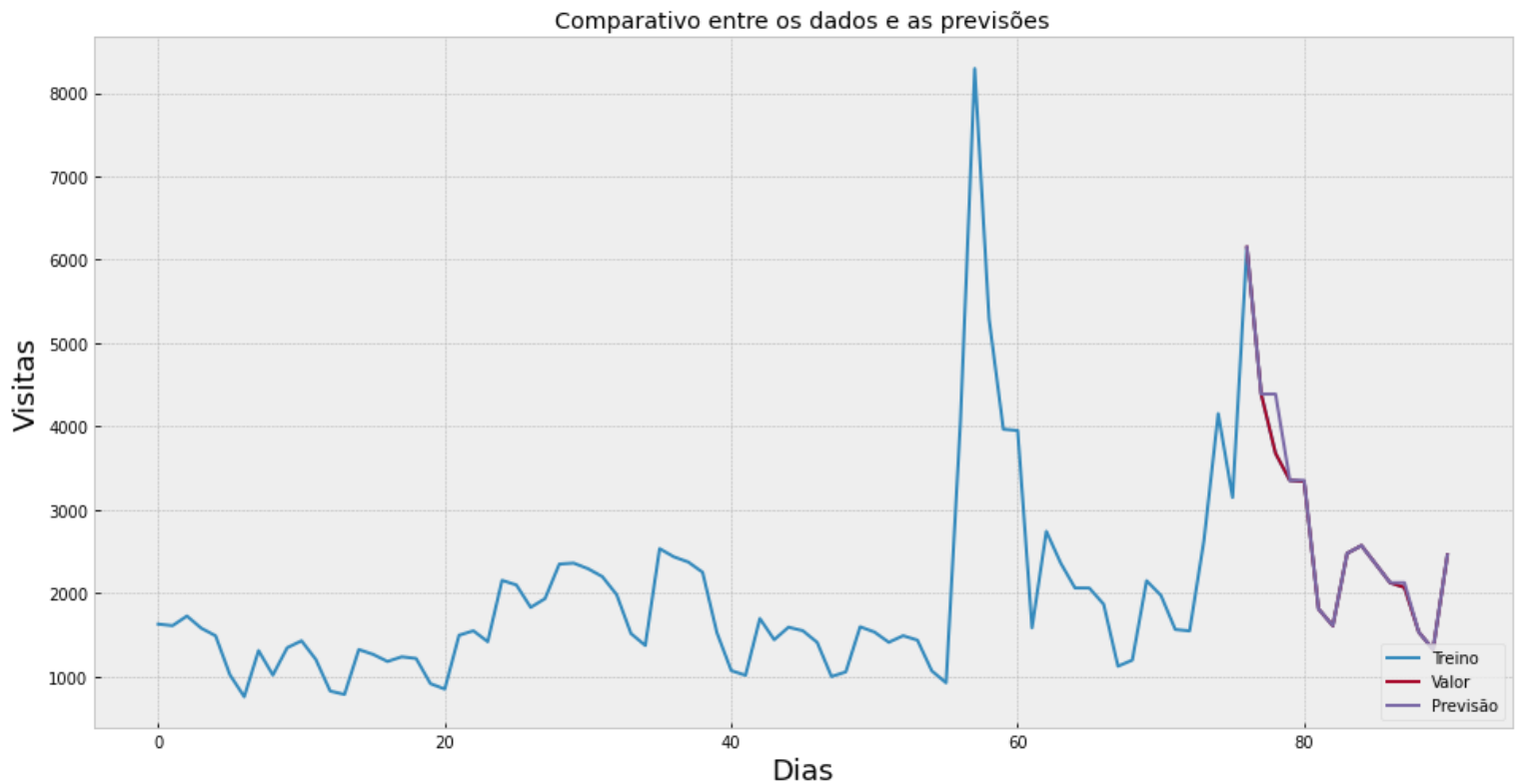
Machine Learning

Baseado nas interpretações retiradas da análise de dados e também devido o interesse do cliente em obter mais visualizações assim como instalações, de forma orgânica, achou-se adequado a previsão aproximada de acessos e instalações de forma orgânica para as próximas duas semanas. Aplicou-se o método SVR de predição Machine Learning, da biblioteca Scikit-learn. Escolheu-se este método devido à adequação dos requisitos, pelo mapeamento de machine learning, oferecido pela biblioteca.



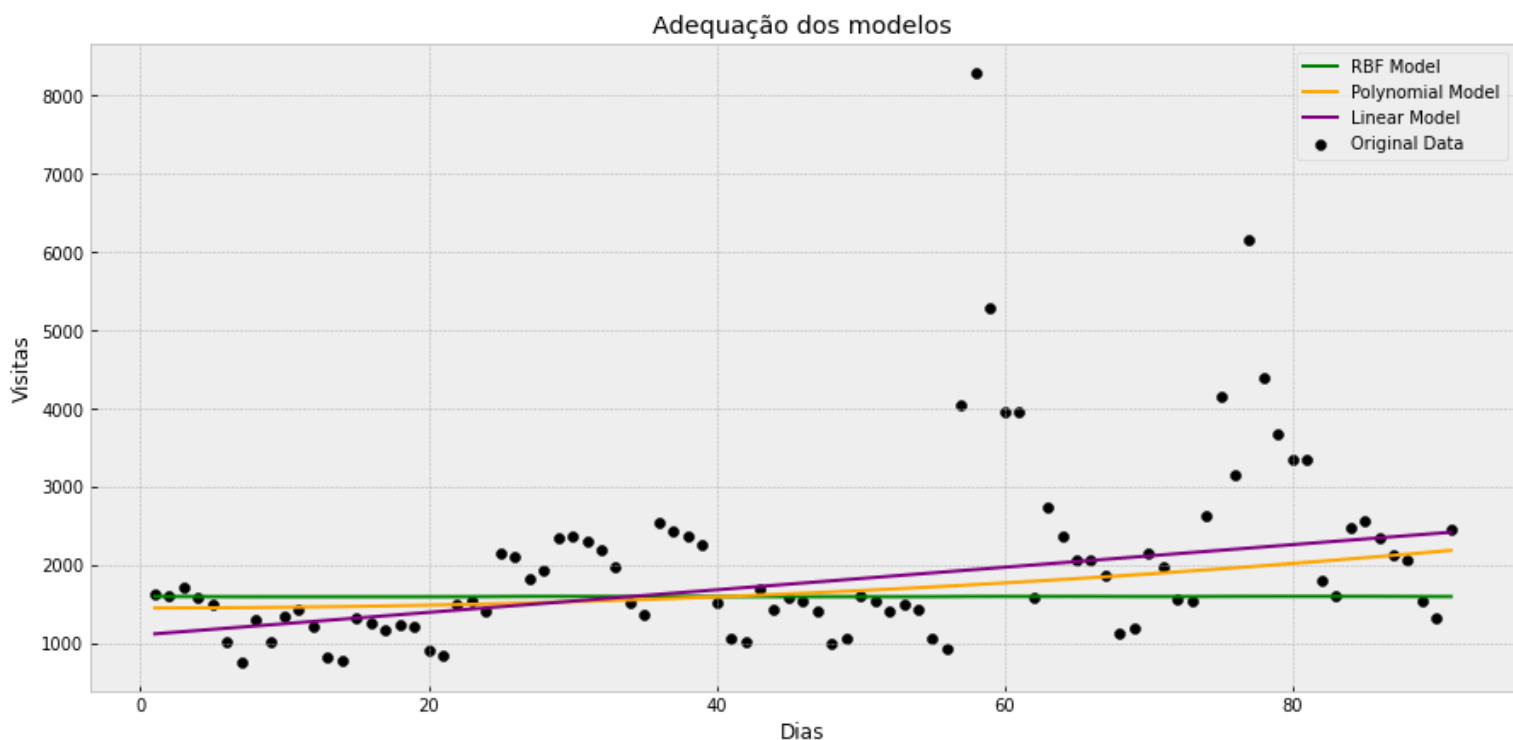
Como a previsão a ser realizada possui mais de 50 instâncias de dados, refere-se à um parâmetro quantitativo, possui menos de 100 mil ocorrências e tanto a quantidade de visualizações quanto instalações não estão atrelados à outros parâmetros do nosso dataset, chega-se até o algoritmo de machine learning SVR(kernel="linear"), disponibilizado pela biblioteca.

Antes de aplicar o método indicado pela biblioteca, utilizou-se um método conhecido como Árvore de Decisões. Ao realizar a estimativa da precisão, obteve-se até 80% de adequação da previsão, com os dados presentes. O gráfico abaixo, representa os valores previstos, assim como os dados originais:



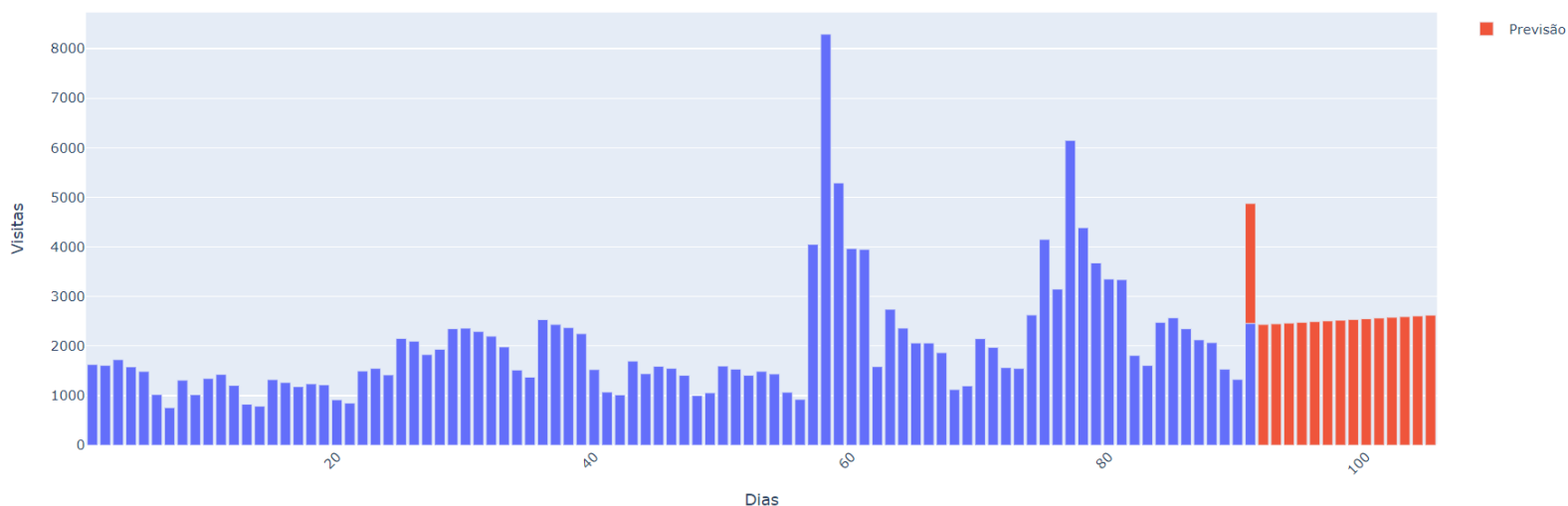
Com uma precisão tão adequada, observa-se que o método de Árvore de decisões representa uma boa opção para representação das previsões.

Para verificação do outro método indicado pela biblioteca, representou-se os valores das visitas ao app por pontos no decorrer do tempo, ao passo que representou-se também a linha de regressão linear, assim como a polinomial e também a regressão pelo método RBF:



Utilizou-se então a regressão linear para predição dos valores de visitas para os próximos 14 dias após o último dia presente no dataset:

Previsões de acessos orgânicos

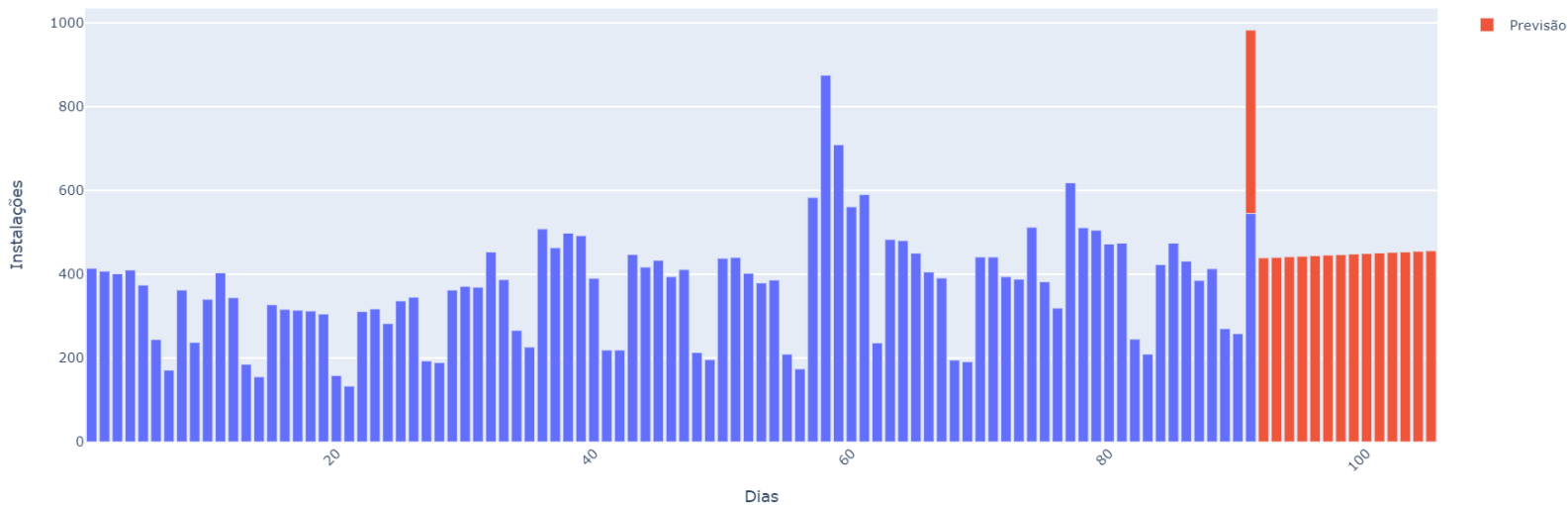


O dia 91 (último dia do dataset), também foi requisitado na previsão, para uma verificação de discrepância entre o valor real, e o valor estimado, obtendo uma porcentagem aproximada de 98%.

Observa-se com essa estimativa, uma tendência de alta, indicando possível acréscimo da quantidade de visitas, em relação à recente baixa que havia tendo.

A análise posterior corresponde à quantidade de instalações de forma orgânica, onde obteve-se a seguinte predição:

Previsões de instalações orgânicas



De mesma forma, utilizou-se a sobreposição da previsão do dia 91, sobre o dado real de instalações neste dia, apenas para fins de estimativa da previsão do algoritmo. O valor comparativo retorna uma precisão de 80,31%.

A previsão então, nos indica um crescimento da quantidade de instalação. Ressalta-se que o gráfico apresentado originalmente encontra-se de forma interativa, sendo possível encontrar o valor exato de cada barra, ao mover o cursor sobre a barra desejada.

Os métodos apresentados acima correspondem à aproximações e previsões baseados no histórico de instalações e visitas ao app no decorrer dos dias. Ressalta-se que quanto, relativamente, maior o nosso dataset, mais precisa se torna nossa previsão. Sendo útil para estabelecer insight's e ideias acerca da tendência do mercado.

Por último, calculou-se o valor total de visitas e instalações do app, previstas para os próximos 14 dias:

O total de visitas previstas para as próximas 2 semanas, de forma orgânica é: 37817.1
Com um total de instalações previstas de: 6700.5

Todos os gráficos, cálculos e manipulações podem ser encontradas pelo notebook online desenvolvido no Google colab:

[NOTEBOOK ONLINE](#)

Link para o Repositório do GitHub:

[REPOSITÓRIO ONLINE](#)

Dashboard Power BI Online:

[DASHBOARD ONLINE](#)