Contexto :

1.0 O contexto (Ficticio) ?

Em uma reunião de relatório trimestral na Rossmann, a diretoria identificou um aumento nas aberturas de lojas concorrentes.

Para evitar que os concorrentes participem da fatia de mercado, o conselho C-Level decidiu realizar uma pesquisa no local com todos os clientes que estavam comprando nas lojas para entender os seguintes pontos:

• Por que o cliente preferiu comprar da Rossmann e não da concorrência?

• De 0 a 10, qual é o nível de satisfação do cliente com os produtos e serviços da loja Rossmann? Por quê? (CSAT)

• Por que havia certos produtos que o cliente preferia comprar da concorrência e não da Rossmann?

• Quais foram, na opinião dos clientes, os três pontos fortes e os três pontos fracos das lojas Rossmann? Por quê?

• De 0 a 10, quanto o cliente recomendaria Rossmann para um amigo ou parente? Por quê? (NPS)

Após a pesquisa com o cliente, os seguintes insights foram identificados:

• Os clientes preferiram comprar da concorrência porque o serviço era melhor.

• Os clientes compraram o mesmo produto da concorrência porque ele ofereceu preços melhores.

• Os clientes preferiram ir à loja da concorrência porque o ambiente físico era mais agradável e alguns até tinham uma rede de lanchonetes dentro deles.

**2.0 O problema**

Com base nas percepções da pesquisa, as iniciativas foram delineadas como:

Redesenhar todo o programa de treinamento para gerentes e atendentes de loja.

Reveja toda a estratégia de preços.

Faça parcerias com redes de cafeterias e padarias.

Implante totens de autoatendimento para clientes que preferiram fazer compras sem o auxílio de atendentes.

Porém, para realizar todas essas iniciativas, um planejamento financeiro bem estruturado e com margem mínima de erro deve ser feito para que a empresa não desperdice dinheiro na implantação das iniciativas.

O CFO responsável por este plano teve uma grande dificuldade em executá-lo porque era preciso saber quanto cada loja estava vendendo e quanto poderia vender no curto prazo. Além disso, não havia uma maneira fácil, automatizada ou conveniente de obter essas informações.

**3.0 A solução**

A solução entregue foi um bot do Telegram.

O usuário só precisava enviar o número da loja e o bot responderia ao total de vendas previsto que a loja faria ao final das seis semanas seguintes.

.**1 O que levou à solução**

**.1 O que levou à solução**

**Análise exploratória de dados**

**Pontos chave:**

**• O valor médio de vendas é de cerca de US $ 5.773,82.**

**• Houve um máximo de 7.388 clientes em um único dia.**

**• Além disso, há lojas concorrentes bem próximas (20 m) das lojas da Rossmann.**

**Mapa de Hipóteses**

**Delineamos esse mapa para nos ajudar a decidir quais variáveis ​​precisamos para validar as hipóteses.**

**Análise Univariada**

**como podemos observar, a maioria das vendas gira em torno de US $ 5.700.**

**Além disso, a distribuição é moderadamente assimétrica (skewness = 0,641460) e apresenta uma curtose positiva (1,778375), o que significa que temos alguns possíveis outliers em nosso conjunto de dados. Portanto, a distribuição não segue uma distribuição normal.**

**Validação de hipótese - Análise bivariada**

**Hipóteses Principais**

**H2. Lojas com concorrentes mais próximos devem ter vendas menores.**

**Como podemos observar no gráfico de barras, as lojas com concorrentes mais próximos têm vendas mais altas. Além disso, podemos observar no gráfico de dispersão que temos uma maior concentração de vendas à medida que diminuímos a distância da concorrência.**

**Portanto, nossa hipótese é FALSA.**

**Correlações**

**Conforme observado nos resultados, o coeficiente de correlação de Pearson entre a distância\_competição e as vendas é -0,23, o que nos diz que é uma correlação negativa fraca. Apesar da fraqueza, podemos incluir a concorrência\_distância porque ela tem alguma influência na variável-alvo (vendas).**

H4. Lojas com maior período de tempo em promoção devem ter vendas maiores.

Como há muitos dados, dividimos o conjunto de dados em dois períodos: promoção regular e promoção estendida.

s podemos observar no Total de vendas x Semanas na promoção estendida, há um período em que a promoção estendida resulta em mais vendas, então após um período de tempo as vendas totais começam a diminuir.

Do Total de vendas x Semanas em promoção regular, podemos observar que à medida que o deslocamento fica cada vez mais próximo de zero, as vendas começam a aumentar.

Assim, lojas com maior tempo de promoção não têm vendas maiores, pois as vendas vão diminuindo conforme a promoção aumenta.

Portanto, a hipótese é FALSA.

Correlações

Além disso, a partir do Mapa de calor de correlação obtivemos um coeficiente de -0,029 que é bem próximo de zero. Portanto, temos uma correlação super fraca, o que faz sentido porque, olhando para nossos dados, temos longos períodos de vendas totais quase constantes (consulte Vendas totais x Semanas na promoção estendida).

H6. Lojas com promoções consecutivas mais altas devem ter vendas mais altas.

Observando os resultados parece que as lojas com maiores promoções consecutivas não apresentam maiores vendas.

Portanto, nossa hipótese é FALSA.  
H9. As lojas devem ter vendas maiores no segundo semestre do ano.

Conforme observado nos resultados anteriores, as lojas não apresentam vendas maiores no segundo semestre do ano. Além disso, observando o coeficiente de correlação de Pearson de -0,75, podemos verificar que existe uma forte correlação negativa entre mês e vendas.

Portanto, nossa hipótese é FALSA.

Resumo de hipóteses

Também delineamos outras hipóteses, mas as apresentadas anteriormente foram as mais perspicazes.

Conforme observado na matriz acima:

Essas correlações nos orientaram na seleção de quais variáveis ​​incluir no modelo. Isso poderia se traduzir em mais assertividade nas previsões, o que significa melhor planejamento orçamentário e menos desperdício de dinheiro.

3.1.3 Aprendizado de Máquina

Conforme observado nos resultados, o Random Forest Regressor teve o menor MAE (837,5 + - 218,9). No entanto, fizemos um ajuste fino no XGBoost Regressor para experimentar se também poderia ser um candidato potencial.

Depois de ajustar o modelo:

Como podemos ver, houve uma grande melhora em relação aos resultados anteriores.

Desempenho dos negócios

Como podemos observar na amostra acima, temos o melhor e o pior cenário para o total de vendas, portanto, podemos ter um intervalo no qual podemos basear as estimativas de orçamento.

MAPE x erro

Existem lojas que são mais difíceis de fazer as previsões (circuladas em vermelho). Assim, algumas estratégias que podem resolver este desafio na próxima iteração do projeto podem ser:

• Olhando mais de perto as variáveis ​​(adicionar ou remover).

• Experimente outros métodos e outras técnicas para melhorar as previsões.

Observando os resultados, podemos ver que:

• Observando os gráficos de primeira e segunda linha, podemos ver que as previsões ou nosso modelo está muito próximo do valor real de vendas. Por outro lado, a taxa de erro tem alguma variação.

• Ao observar o histograma, a distribuição do erro quase segue uma distribuição normal.

• Ao observar o gráfico de dispersão dos erros, os pontos parecem bem encaixados em um tubo horizontal, o que significa que há algumas variações no erro. Se os pontos formaram qualquer outra forma (por exemplo, cone de abertura / fechamento ou arco), isso significaria que os erros seguem uma tendência e precisaríamos revisar nosso modelo.

4.0 Próximas etapas

• Experimente outros algoritmos de aprendizado de máquina para melhorar o desempenho dos negócios em 10%.

• Experimente selecionar outros recursos para ver o quanto o RMSE é afetado.

• Experimente outras estratégias de ajuste fino de hiperparâmetros para ver o quanto o RMSE é impactado.

• Melhorar as mensagens do bot e sua interação com os usuários.