

Mini Projeto - Walmart
Imóveis DEX04 - Grupo 21
Projeto 2, Slot 4

1.0 A Questão de Negócio

1.1 Meta do Projeto

Fornecer consultoria estratégica para a empresa Walmart na tomada de decisões com o objetivo de realizar um levantamento do faturamento das lojas nos EUA e apontar qual loja é melhor para expandir seu tamanho.

1.2 Metodologia

Por meio da análise dos dados e arquivos fornecidos pela empresa, utilizando bibliotecas Python e metodologias específicas da Ciência de Dados, as informações foram analisadas para identificar quais são os dados mais relevantes para atingir o objetivo de forma mais precisa.

Inicialmente, analisou-se os outliers dos dados e verificou-se que, devido ao tamanho da base, não foi necessário retirá-los da análise. Em seguida, calculou-se a correlação entre os dados, para observar se havia algum indicativo de relacionamento entre eles. Contudo, nenhuma das relações apresentou um valor que poderia ser considerado relevante para a análise, sendo a primeira hipótese, descartada.

A segunda hipótese refere-se ao peso dos feriados para as vendas. Apesar de haver alguma relação, não foi identificado um peso maior dos feriados quando comparados aos dias comuns. Dessa forma, levantou-se a hipótese de existir picos de venda ao longo do período mas também não foi identificada uma variação discrepante entre os períodos ao longo dos anos.

Em seguida, foi analisado se havia qualquer semelhança entre as lojas que mais venderam e que poderiam beneficiar-se de uma expansão, mas este padrão não foi encontrado. A hipótese cinco analisou semelhanças das vendas nas lojas considerando feriados e dias comuns, porém não se verificou qualquer relação, sendo o volume de vendas bem dividido.

Após essas análises, determinou-se que um ponto de análise relevante é analisar o CPI que pudesse indicar as maiores vendas, mas as diferenças encontradas não foram suficientes para basear o projeto nessa informação. A partir disso, levantou-se a hipótese de que poderiam existir algumas semanas em que todas as lojas tiveram uma quantidade de vendas baixa e outras uma quantidade de vendas altas, contudo, este padrão não foi observado nos anos subsequentes.

Por último, analisou-se a loja com menor amplitude que poderia ter, entre outros fatores, a falta de espaço como fator limitante. Com base nisso, chegou-se na conclusão que esta loja é a 33, que apresentou-se como uma possibilidade de expansão.

2.0 O Entendimento do Negócio

2.1 Dados disponíveis

O arquivo disponibilizado pela empresa em forma de CSV possui dados que apresentam as vendas semanais de 05/02/10 até 01/11/12 em 45 lojas varejistas da rede Walmart e também algumas métricas econômicas e meteorológicas por semana. A descrição destas características estão abaixo:

Holiday Events	Semanas
Super Bowl	12-Feb-10, 11-Feb-11, 10-Feb-12
Labour Day	10-Sep-10, 09-Sep-11, 07-Sep-12
Thanksgiving	26-Nov-10, 25-Nov-11, 23-Nov-12
Christmas	31-Dec-10, 30-Dec-11, 28-Dec-12

Variável	Descrição
Store	Número da Loja
Date	Semana de Venda
Weekly_Sales	Venda Naquela Semana
Holiday_Flag	Flag se é ou não semana com feriado (1 - Holiday Week 0 - Non-Holiday Week)
Temperature	Temperatura do dia em °F
Fuel_Price	Preço do combustível na região da loja
CPI	Índice de preços ao consumidor
Unemployment	Taxa de desemprego

2.2 Relevância dos dados para o nosso objetivo

Os dados fornecidos pela empresa mostraram-se relevantes, visto que nos permite analisar o escopo do problema a ser resolvido. Por meio dessas informações foi possível aplicar as metodologias existentes na Ciência de Dados de forma a nos gerar *insights* importantes para o negócio analisado.

2.3 Sobre a influência dos dados na solução

Os dados fornecidos não nos trazem informações diretas, então foi necessário que realizássemos alguns tratamentos para identificarmos quais são os mais relevantes para o nosso objetivo.

2.4 Setor do negócio analisado

Setor comercial.

3.0 A Coleta de Dados

3.1 Sobre a acepção dos dados

Os dados fornecidos fazem sentido com relação ao negócio analisado, porém, foi necessário realizar um tratamento bem elaborado para identificar quais dados serão mais importantes para a análise. Por exemplo, foi a partir das informações e da amplitude da coluna `Weekly_Sales` que foi possível determinar a loja candidata a receber o investimento de expansão.

3.2 Sobre o formato dos dados fornecidos

Sobre os dados fornecidos, algumas colunas possuem dados em formatos e padrões diferentes uns dos outros, sendo necessário tratar tais ocorrências. Um exemplo é o campo `data` que está no formato `object`, e deverá ser alterado para o formato `date`. Com essas alterações poderemos usar as funções existentes para a análise dos dados

3.3 Informações importantes que são possíveis de se obter por meio dos dados

Por meio dos dados fornecidos foi possível verificar, por exemplo, picos de venda ao longo do período, se poderiam existir algumas semanas em que todas as lojas tiveram uma quantidade de vendas baixa e outras uma quantidade de vendas altas, entre outros.

4.0 Limpeza de Dados

4.1 Remoção de dados outliers e valores ausentes

Os outliers dos dados foram analisados e verificou-se que, devido ao tamanho da base, não foi necessário retirá-los do projeto.

5.0 Exploração de Dados

5.1 Visualização dos dados

Por meio da biblioteca Seaborn plotou-se gráficos para encontrar as correlações nos dados. Por serem próximos a zero existe pouca correlação entre as colunas, sendo necessário analisar diversas hipóteses para identificar a loja escolhida.

5.2 Nossas métricas essenciais.

Nossas métricas essenciais para chegarmos aos nossos objetivos foram vendas semanais e amplitude destas, de forma a identificar a loja de interesse, também utilizamos o CPI e as datas comuns e de feriados.

5.3 Decisão que tomamos com base nos dados

Verificamos que a loja 33 é a loja que teve a menor amplitude (Aproximadamente 121187), conforme analisamos na hipótese 8. Isso indica que algum fator está limitando as vendas dessa loja mais do que as outras, podendo este ser o seu tamanho. Por isso, ela é a nossa indicação para receber o investimento da expansão.

Precisariamos de mais informações como: tamanho da loja, período de funcionamento, população, qualidade do serviço, satisfação do cliente, estacionamento, concorrência local entre outros para afirmar com mais acurácia se essa realmente é a melhor loja para expansão.