

Relatório e Insights do Projeto

Os dados iniciais não apresentavam dados nulos ou faltantes, não havendo a necessidade de alguma medida para isso. A remoção de duplicatas se torna desnecessária para um cenário de vendas, pois uma coincidência pode ocorrer e um tratamento desse tipo afetaria a precisão da análise de dados em um cenário real.

A relação entre a soma diária de vendas ao longo do tempo tem um comportamento oscilante. Isso é evidenciado pela reta da regressão linear ter inclinação pouco válida para considerar uma relação relevante. Ainda é mais evidenciado pelo histograma da relação de vendas com o mês, em que há um curioso comportamento de crescimento em meses de numeração ímpar e queda em meses de numeração par. A análise das categorias de produtos demonstra que os produtos mais vendidos são eletrodomésticos. Isso é mostrado tanto pelo histograma quanto pelo gráfico de linhas (em que a amplitude de vendas de eletrodomésticos é geralmente maior que a das outras categorias).

A relação entre a média diária de vendas ao longo do tempo aparenta ter mais relação com a série temporal, visto que a inclinação da reta de regressão linear ter um coeficiente angular mais evidente. Os meses de janeiro, setembro, novembro e dezembro são os que mais possuem variabilidade de dados, pois têm alto desvio-padrão, além de serem os mesmos com maior média de vendas. O mesmo ocorre com os eletrodomésticos: são os produtos com mais soma de vendas, os com a maior média de vendas e os com o maior desvio-padrão.

O produto mais vendido, considerando todas as modelos para o mesmo tipo de produto ao longo do período, é a Geladeira Duplex, enquanto o menos vendido no mesmo período é o celular Sony Experia XA. Estatisticamente, o maior investimento em eletrodomésticos é mais viável.