Para o desafio deste projeto utilizei o <u>modelo 3</u> disponível pela plataforma da DIO (Dataset – 1000 com preço variável e renovação de estoque). Os produtos usados nesta análise, são computadores, com variações em <u>marca</u>, <u>modelo</u>, <u>tamanho da tela</u>, <u>Peso</u> e <u>número identificador</u>. Entretanto em nosso Dataset não há precificação de nenhum produto, ou seja, a análise deste estudo preditivo não contemplará uma previsão profunda de variações de preços, mas, sim, como a relação entre os componentes do pacote geram impacto na precificação, modelagem do pacote e forma que o produto é lançado no mercado.

As marcas disponíveis nesta análise eram: Bell, Howell, MicroChip e Orange.

Os **modelos** de computadores analisados eram: <u>Base</u>, <u>Perfomance</u> e <u>Standard</u>.

Os <u>tamanhos de tela</u> eram: 10.1 | 11.6 | 12.3 | 12.5 | 13.3 | 13.5 | 14 | 15.6 | 17.3 e 18.4

- O <u>Peso do Pacote</u> de cada produto variava entre 4.0 8.3 Kg, sendo essas as variações de peso: 4.0 | 4.2 | 4.3 | 4.4 | 4.5 | 4.6 | 4.7 | 4.8 | 4.9 | 5.0 | 5.1 | 5.2 | 5.3 | 5.4 | 5.5 | 5.6 | 5.7 | 5.8 | 5.9 | 6.0 | 6.1 | 6.2 | 6.3 | 6.4 | 6.5 | 6.6 | 6.7 | 6.8 | 6.9 | 7.0 | 7.1 | 7.5 | 7.6 | 7.7 | 7.8 | 7.9 | 8.0 | 8.1 | 8.2 e 8.3.
- O <u>Número Identificador</u> é gerado a partir de valores aleatórios ou pseudoaleatórios, para garantir que sejam únicos e imprevisíveis, o sistema gerador pode ser o mesmo responsável por tecnologias de segurança como os <u>UUIDS</u> (Identificadores Único Universais), <u>sistema de token</u> ou <u>sistema de hashing</u>.

Exemplo:

Marca do Computador: BellModelo do Computador: Base

Tamanho da Tela: 10.1Peso do Produto: 4.2

• <u>Números Identificador</u>: <u>c1e69a69-3373-4e7a-9e54-96b5f9babb22</u>

Visão Geral:

Por meio da análise do SageMaker conseguimos chegar à conclusão que alguns fatores causavam forte impacto na mudança de peso e posteriormente no aumento ou diminuição de preços no nosso Dataset.

Essa variação mudava conforme o <u>tamanho da tela</u>, esse era o principal fator de alteração, pois, quanto maior a tela do aparelho, maior a demanda de custo e de material para a produção do produto. O peso do produto aumenta conforme o tamanho da tela, sendo 4.0 Kg o menor valor estipulado e 8.3 Kg o maior.

Análise:

Essas margens são importantes para entendermos um pouco como o SageMaker Canvas analisa e prevê as possibilidades baseadas em tendências de mercado, mas, também se utilizando de fórmulas matemáticas, como probabilidade e estatística e métricas de machine learning e inteligência Artificial.

Análise dos Gráficos:

Quatro gráficos foram gerados para fazer a previsão da renovação do estoque e identificar os impactos capazes de causar mudanças significativamente nos preços e na forma que os pacotes seriam distribuídos no mercado.

Os gráficos da Overview (visão geral) trazem diferentes formas de análise utilizando-se de diferentes métricas para calcular os impactos e gerar uma previsão capaz de tomar as melhores decisões em cenários adversos mediante as várias tendências de mercado.

Gráfico 01: Impacto do <u>Peso do Pacote</u> na previsão do tamanho da tela

Neste gráfico, o impacto é calculado por meio de uma fórmula matemática que usa o valor inicial de referências e um cálculo de previsão ajustada.

1. Valor inicial de referência:

Este é o <u>valor do tamanho da tela antes de aplicar o impacto do</u> <u>peso do pacote</u>. Vamos chamar este valor de <u>Tamanho Inicial</u>.

<u>Tamanho Inicial</u> = Valor do Tamanho da Tela (Sem o impacto do Peso do Pacote)

2. Cálculo de Previsão Ajustada:

Este <u>aplica o impacto calculado ao valor inicial para obter a</u> <u>previsão</u> ajustada do tamanho da tela.

Aplicação: Impacto calculado + Valor Inicial

Fórmula:

Tamanho previsto – Tamanho inicial + (Impacto X Diferença no Valor) =

Exemplo:

Peso do Pacote: <u>4.500 Kg</u> = Tamanho Inicial: <u>15.6 polegadas</u>

Impacto: -3.804

Cálculo

Tamanho Previsto -15.6 + (-3.804)

Tamanho Previsto – 15.6 – 11.796

*Com o valor do impacto sobre a tela obtido (11.796), deve-se calcular se o tamanho da tela está adequado. Para tal deve-se fazer um cálculo de comparação junto a tabela de tamanhos da tela.

Tamanhos da Tela: 10.1 | <u>11.6</u> | 12.3 | 12.5 | 13.3 | 13.5 | 14 | 15.6 | 17.3 e 18.4

Cálculo de Comparação:

11.6:
$$11.796 - 11.6 = 0.196$$

12.3:
$$11.796 - 12.3 = 0.504$$

12.5:
$$11.796 - 12.5 = 0.704$$

13.3:
$$11.796 - 13.3 = 1.504$$

13.5:
$$11.796 - 13.5 = 1.704$$

14:
$$11.796 - 14 = 2.204$$

15.6:
$$11.796 - 15.6 = 3.804$$

17.3:
$$11.796 - 17.3 = 5.504$$

18.4:
$$11.796 - 18.4 = 6.604$$

O valor de tamanho que mais se aproximar do valor obtido, é o adequado para se usar, para obter-se a previsão. Neste caso o tamanho de tela que mais se aproximou do valor foi o de 11.6 polegadas, que obteve a diferença de 0.196, a mais baixa em comparação ao valor obtido.

Tamanho Previsto = *Tela de* **11.6** *Polegadas*

O <u>Peso do pacote</u> (Package Weigth) conforme o aumento do <u>tamanho da tela</u> (Screensize) causa um impacto de <u>82.03%</u> na variação de preço e porventura no encarecimento do produto.

Gráfico 02: Impacto da <u>Marca do Computador</u> na previsão do tamanho da tela

No Segundo caso de previsão foi analisado o impacto que a marca do produto teria na previsão de variação de preço e alteração do pacote, em comparação com os outros possíveis impactadores.

MicroChip:

Mean -0.212

Max -0.136

75% -0.172

Median -0.192

25% -0.253 Min -0.312

Orange:

Mean -0.741 Max 1.036 75% -0.92 Median -0.637 25% -0.582 Min -0.556

Howell:

Mean -0.155 Max -0.105 75% -0.124 Median -0.169 25% -0.178 Min -0.192

Bell:

Mean -0.039 Max 0.107 75% -0.01 Median -0.027 25% -0.076 Min -0.021

Análise de Impacto

Os valores mostram como cada marca de computador impacta a previsão do tamanho da tela. Os <u>valores negativos indicam uma redução no tamanho da</u> **tela**, enquanto **valores positivos indicam um aumento**.

MicroChip:

A média de -0.212 indica que, em geral, os computadores da marca MicroChip tendem a reduzir o tamanho da tela em 0.212 unidades.

A variação (Min à Max) mostra que o impacto pode variar de -0.312 a -0.136.

Orange:

A média de -0.741 indica uma <u>redução mais significativa no tamanho da tela</u>, em média, para os computadores da marca Orange.

A ampla variação (Min à Max) de -0.741 a 1.036 indica que, <u>embora</u> geralmente <u>reduza o tamanho da tela, em alguns casos pode aumentá-lo significativamente</u>.

Howell:

A média de -0.155 sugere que os computadores da marca Howell <u>tendem a</u> <u>reduzir o tamanho da tela</u> em 0.155 unidades, em média.

A variação é mais estreita, de -0.192 a -0.105, indicando um **impacto mais consistente**.

Bell:

A média de -0.039 indica um <u>impacto muito pequeno na redução do tamanho</u> da tela.

Com variação de -0.039 a 0.107, os computadores da marca Bell <u>podem</u> <u>ocasionalmente aumentar o tamanho da tela</u>.

MicroChip e Howell: Tendem a ter um impacto negativo moderado e mais consistente na previsão do tamanho da tela.

Orange: Tem uma média de impacto negativa mais significativa, mas com grande variabilidade, indicando que pode tanto aumentar quanto diminuir o tamanho da tela.

Bell: Tem o menor impacto médio, com possibilidade de aumento ou diminuição no tamanho da tela.

Esses insights podem ajudar a prever como a marca do computador afeta o tamanho da tela e fazer ajustes mais informados em estratégias de estoque e marketing.

A <u>Marca do Produto</u> (Compunter Brand) é o segundo principal fator de impacto entre os analisados, correspondendo a <u>13.033%</u>.

Gráfico 03: Impacto do <u>ID do produto</u> na previsão do tamanho da tela

O número de identificação como foi mencionado acima é gerado por um sistema de hash ou token e se trata de uma sequência de números aleatórios ou pseudoaleatórios (ex: c1e69a69-3373-4e7a-9e54-96b5f9babb22), o uso escolhido de uma sequência numérica natural para representar o ID dos produtos na explicação deste gráfico foi de minha escolha, apenas para fins didáticos de melhor entendimento e visualização.

ID 01:

Mean -0.029 Max -0.029 75% -0.029 Median -0.029 25% -0.029

Min -0.029

ID 02:

Mean -0.102

Max -0.102

75% -0.102

Median -0.102

25% -0.102

Min -0.102

ID 03:

Mean -0.014

Max -0.014

75% -0.014

Median -0.014

25% -0.014

Min -0.014

ID 04:

Mean -0.111

Max -0.111

75% -0.111

Median -0.111

25% -0.111

Min -0.111

ID 05:

Mean 0

Max 0

75% 0

Median 0

25% 0

Min 0

ID 06:

Mean -0.017

Max -0.017

75% -0.017

Median -0.017

25% -0.017

Min -0.017

ID 07:

Mean -0.105

Max -0.105

75% -0.105

Median -0.105

25% -0.105

Min -0.105

ID 08:

Mean -0.01

Max -0.01

75% -0.01

Median -0.01

25% -0.01

Min -0.01

ID 09:

Mean -0.119

Max -0.119

75% -0.119

Median -0.119

25% -0.119

Min -0.119

ID 10:

Mean -0.033

Max -0.033

75% -0.033

Median -0.033

25% -0.033

Min -0.033

ID 11:

Mean -0.033

Max -0.033

75% -0.033

Median -0.033

25% -0.033

Min -0.033

ID 12:

Mean -0.105

Max -0.105

75% -0.105

Median -0.105

25% -0.105

Min -0.105

ID 13:

Mean -0.223

Max -0.223

75% -0.223

Median -0.223

25% -0.223

Min -0.223

ID 14:

Mean -0.012

Max -0.012

75% -0.012

Median -0.012

25% -0.012

Min -0.012

ID 15:

Mean 0

Max 0

75% 0

Median 0

25% 0

Min 0

ID 16:

Mean -0.015

Max -0.015

75% -0.015

Median -0.015

25% -0.015

Min -0.015

ID 17:

Mean -0.021

Max -0.021

75% -0.021

Median -0.021

25% -0.021 Min -0.021

ID 18:

Mean -0.027 Max -0.027 75% -0.027 Median -0.027 25% -0.027 Min -0.027

ID 19:

Mean -0.101 Max -0.101 75% -0.101 Median -0.101 25% -0.101 Min -0.101

ID 20:

Mean -0.063 Max -0.063 75% -0.063 Median -0.063 25% -0.063 Min -0.063

ID 21:

Mean -0.015 Max -0.015 75% -0.015 Median -0.015 25% -0.015 Min -0.015

ID 22:

Mean -0.111 Max -0.111 75% -0.111 Median -0.111 25% -0.111 Min -0.111

ID 23:

Mean -0.087 Max -0.087 75% -0.087 Median -0.087 25% -0.087 Min -0.087

Os números fornecidos para cada ID indicam a <u>média</u> (Mean), o valor <u>máximo</u> (Max), o valor <u>mínimo</u> (Min), e os <u>percentis</u> (25%, 50% - <u>Mediana</u>, 75%) <u>do impacto no tamanho da tela</u>. Vamos analisar alguns aspectos importantes:

- 1. **Impacto Constante**: Para muitos IDs, os valores de impacto são constantes (todos os valores iguais), indicando que para esses IDs específicos, o impacto é sempre o mesmo.
- 2. **Variação Zero**: Alguns IDs têm impacto zero, significando que esses IDs não influenciam a previsão do tamanho da tela.

Análise por ID

IDs com Impacto Constante

- ID 01, 02, 03, 04, 06, 07, 08, 09, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23:
 - Esses IDs têm impactos negativos constantes, o que significa que cada vez que esses IDs são usados, eles reduzem o tamanho da tela por um valor fixo.
 - Exemplos:

• **ID 01**: -0.029

• **ID 02**: -0.102

• **ID 13**: -0.223

■ **ID 23**: -0.087

IDs com Impacto Zero

- ID 05 e 15:
 - Esses IDs têm impacto zero em todos os casos, indicando que não afetam o tamanho da tela.

Relação dos Números Entre Si

- Consistência: A consistência dos valores para cada ID sugere que o sistema de identificação (hash/token) tem uma relação direta e previsível com o impacto no tamanho da tela.
- Magnitude do Impacto: A magnitude dos impactos varia entre os IDs, indicando que alguns IDs têm um impacto mais significativo do que outros.

Impacto nas Previsões de Estoque e Variação de Preço

- Previsão de Estoque: Os impactos fornecem insights sobre como diferentes IDs de produtos influenciam o tamanho da tela, que pode ser um fator importante na gestão de estoque. Por exemplo, produtos com IDs que resultam em tamanhos de tela menores podem ter um padrão de demanda diferente dos produtos com IDs que resultam em tamanhos maiores.
- Variação de Preço: A variação no tamanho da tela pode afetar os preços dos produtos. IDs que reduzem significativamente o tamanho da tela pode estar associado a produtos de menor preço, enquanto IDs com impacto menor ou nulo podem estar associados a produtos de preço mais alto.

Impacto na Composição Final do Produto

Configuração do Produto: O número de identificação (ID) influencia a configuração final do produto ao afetar o tamanho da tela. Isso pode ser usado para categorizar produtos em diferentes segmentos de mercado.
Estratégia de Marketing e Vendas: Compreender o impacto do ID no tamanho da tela pode ajudar a ajustar estratégias de marketing e vendas, posicionando produtos de acordo com suas características e os impactos dos seus IDs.

Exemplo de Aplicação Prática ID 13:

- **Impacto**: -0.223 (constante)
- Interpretação: Para produtos com ID 13, o tamanho da tela será sempre reduzido em 0.223 unidades. Isso pode significar que produtos com ID 13 são tipicamente menores, talvez mais portáteis e possivelmente mais baratos.
- Previsão de Estoque e Preço: Produtos com ID 13 podem ter uma demanda diferente devido ao seu tamanho reduzido. Ajustar o estoque de acordo com a demanda prevista e considerar o impacto no preço pode ajudar a otimizar a estratégia de vendas.

A previsão de impacto gerado pelo <u>número de identificação</u> (Product Id) no tamanho da tela de modo geral foi responsável por uma margem de <u>3.058%</u> segundo a análise do SageMaker. Apesar de não ser uma margem muito grande em comparação com outros dois gráficos acima, ela representou uma mudança pequena, mas, suficiente para que houvesse uma mudança nos preços dos produtos.

Gráfico 04: Impacto do modelo de computador na previsão do tamanho da tela

Nos gráficos de visão geral, vamos para o último fator de impacto desta previsão, o **Modelo do Computador** (Computer Model) que teve uma margem de **1.879%**.

Base:

Mean -0.106

Max -0.345

75% -0.104

Median -0.09

25% -0.052

Min -0.018

Standard:

Mean -0.016

Max -0.007

75% -0.011

Median -0.015

25% -0.019

Min -0.036

Perfomance:

Mean -0.006

Max -0.002

75% -0.003

Median -0.006

25% -0.007

Min -0.01

Análise dos Impactos

Modelo Base

Média de -0.106: Em média, o modelo Base reduz o tamanho da tela em 0.106 unidades.

Máximo de -0.345 e **Mínimo de -0.018**: O impacto pode variar significativamente, com uma redução máxima de 0.345 unidades e mínima de 0.018 unidades.

Distribuição dos Impactos: A maioria dos impactos está entre -0.052 (25%) e -0.104 (75%), com uma mediana de -0.09. Isso indica que a maior parte dos impactos é relativamente uniforme em torno da média.

Modelo Standard

Média de -0.016: Em média, o modelo Standard reduz o tamanho da tela em 0.016 unidades.

Máximo de -0.007 e **Mínimo de -0.036**: O impacto varia menos que no modelo Base, com uma redução máxima de 0.007 unidades e mínima de 0.036 unidades.

Distribuição dos Impactos: A maioria dos impactos está entre -0.019 (25%) e -0.011 (75%), com uma mediana de -0.015. Isso indica uma distribuição mais concentrada e menor variação em comparação ao modelo Base.

Modelo Performance

Média de -0.006: Em média, o modelo Performance reduz o tamanho da tela em 0.006 unidades.

Máximo de -0.002 e **Mínimo de -0.01**: O impacto é mínimo, com uma redução máxima de 0.002 unidades e mínima de 0.01 unidades.

Distribuição dos Impactos: A maioria dos impactos está entre -0.007 (25%) e -0.003 (75%), com uma mediana de -0.006. Isso indica que o impacto é quase negligenciável e muito consistente.

Relação dos Números Entre Si

- Modelo Base: Apresenta a maior variação e impacto no tamanho da tela.
- **Modelo Standard**: Apresenta impacto menor e mais consistente em comparação ao modelo Base.
- **Modelo Performance**: Tem o menor impacto e a menor variação, indicando que quase não afeta o tamanho da tela.

Impacto nas Previsões de Estoque e Variação de Preço

- Previsão de Estoque: Compreender como cada modelo de computador impacta o tamanho da tela pode ajudar a prever quais modelos terão maior ou menor demanda, permitindo ajustes de estoque mais precisos. Por exemplo, se os modelos Base são menos populares devido ao impacto negativo no tamanho da tela, o estoque pode ser ajustado para refletir essa preferência.
- Variação de Preço: Modelos com menores impactos no tamanho da tela (como o Performance) podem ser valorizados mais no mercado, justificando preços mais altos. Por outro lado, modelos com maior impacto negativo (como o Base) podem precisar de preços mais competitivos para atrair compradores.

Análise Geral:

Os 4 gráficos acima detalham como cada fator impactou no tamanho da tela do produto, e por consequência como estes fatores afetariam positiva ou negativamente o balanço de vendas de um estoque de produtos de computação.

Por meio da tecnologia de previsão e análise inteligente dos recursos AWS como o SageMaker Canvas, foi possível fazer uma projeção de mercado hipotética para a gestão de negócios, se utilizando de recursos de Inteligência Artificial, métricas e modelos de Machine Learning, além, de fórmulas matemáticas. Dessa forma foi possível gerar soluções gráficas de forma rápida e eficiente que pudessem ser compreendidas pelos solicitantes, sem a necessidade de compreensão prévia em programação, ou ciência de dados.

Pensando neste público que a AWS desenvolveu recursos como os encontrados na plataforma SageMaker, o resultado foi esse, alto poder desenvolvimento e análise sem a necessidade de código ou profissional treinado na área da tecnologia. Os recursos distribuídos na plataforma são suficientes para fornecer vários tipos de análise e previsões para o mercado.

Métrica RMSE:

o <u>RMSE</u> (Root Mean-Square Error [inglês] Raiz Quadrada do Erro-Médio [RQEM - português), que é uma <u>métrica de avaliação</u> amplamente utilizada e reconhecida na comunidade de <u>Machine Learning</u>, <u>é usada para medir o</u> desempenho de modelos de regressão.

Ela é calculada tomando-se a raiz quadrada da média dos quadrados dos erros, onde o <u>erro bruto</u> é a <u>diferença entre o valor previsto pelo modelo</u> e o <u>valor real</u>.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - p_i)^2}$$

Onde:

- <u>n</u> é o número de amostras
- y i é o valor observado para a amostra i
- p i é o valor previsto pelo modelo para a amostra i

O RSME pode ser interpretado como o desvio médio que as previsões têm do alvo.

Quanto menor o RMSE, melhor o modelo está se saindo. Porém, um RMSE baixo significa que pode haver uma variação dependendo do contexto.

O valor métrico usado nesta análise é um RSME de <u>+/- 0,668</u>, que é um RMSE baixo, isto é, à medida que a espessura da banda RSME em um modelo

aumenta, menor é a precisão da predição, ou seja, quanto mais próximo de **0** mais eficiente se torna a previsão.

A previsão criada neste insight pelo SageMaker Canvas utiliza um cálculo de previsão baseada em 3 pilares: **Previsão**, **valor atual** e **residual**.

O cálculo consiste em uma subtração usando os valores desses pilares.

- Previsão: Armazena o valor previsto (cria uma estimativa) para o estoque futuro.
- Valor atual: Armazena o valor real do estoque.
- Residual: Armazena o resultado da subtração entre o Valor Real e o Valor Previsto, ou seja, é a diferença entre esses dois valores estimados.

Exemplo:

 $Valor\ Atual - Previsão = Residual$

Usando alguns exemplos do gráfico de **Previsão VS Real** (gráfico de pontuação) do SageMaker temos os seguintes cálculos:

Caso 1:

Predição: 10.352 Atual: 11.6 Residual: 1.248

Caso 2:

Predição: 10.937 Atual: 11.6 Residual: 0.663

Caso 3:

Predição: 11.366 Atual: 10.1 Residual: -1.266

<u>Caso 4</u>:

Predição: 12.098 Atual: 12.3 Residual: 0.202

Caso 5:

Predição: 12.213 Atual: 12.3 Residual: 0.087

Os valores residuais nos casos 1, 2, 4 e $\underline{\mathbf{5}}$ se mantiveram dentro da margem do RMSE ($\underline{\mathbf{0.668}}$) com uma certa diferença entre si, mas, com uma probabilidade

alta de terem êxito em sua previsão. No caso 3 o valor residual ficou fora da margem o que poderia indicar que a previsão para o caso 3 teria uma alta taxa de probabilidade de estar errada.

- 1) Caso 5: Residual de 0.087
- 2) Caso 4: Residual de 0.202
- 3) Caso 2: Residual de 0.663
- 4) Caso 1: Residual de 1.248
- 5) Caso 3: Residual de -1.266

O caso 5 foi o que mais se aproximou de $\underline{\mathbf{0}}$, portanto, foi o que $\underline{\mathbf{teve}}$ $\underline{\mathbf{uma}}$ $\underline{\mathbf{previsão}}$ $\underline{\mathbf{mais}}$ $\underline{\mathbf{assertiva}}$.

Na prática, o residual indica o <u>erro na previsão feita pelo modelo</u>. Um residual de <u>1,248</u> (<u>caso 1</u>) significa que a previsão do modelo estava <u>1,248</u> <u>unidades abaixo do valor real do estoque</u>. Isso ajuda a entender a precisão do modelo: quanto menor o residual, melhor a previsão.

Esses valores são úteis para avaliar a performance do modelo e fazer ajustes se necessário. Se você está obtendo resíduos grandes de forma consistente, pode ser um sinal de que o modelo precisa ser ajustado ou que outros fatores precisam ser considerados na previsão.

Interpretação dos Valores:

- 1. Comparação entre Predição e Atual: O valor real (11,6) é maior do que a previsão (10,35). Isso sugere que o estoque real é maior do que o previsto pelo modelo. Em outras palavras, a previsão estava subestimando o nível de estoque.
- 2. **Análise do Residual**: O residual de 1,248 unidades indica que a previsão estava abaixo do valor real. No entanto, <u>um residual sozinho não informa se há uma tendência de queda ou aumento no estoque; apenas mostra o erro na previsão para esse ponto específico.</u>

Autor:

Está foi a análise feita por: <u>James da Fonseca e Silva</u>, entre os dias 1, 2 e 3 de agosto de 2024, para o desafio de Projeto: Previsão de Estoque Inteligente na AWS com SageMaker Canvas. Proposto pela Digital Innovation One (DIO), dentro do Bootcamp Nexa – Machine Learning para Iniciantes na AWS.