

QUESTÃO 01 (10,0 pontos)

A figura 01 ilustra o processo de marketing estratégico.

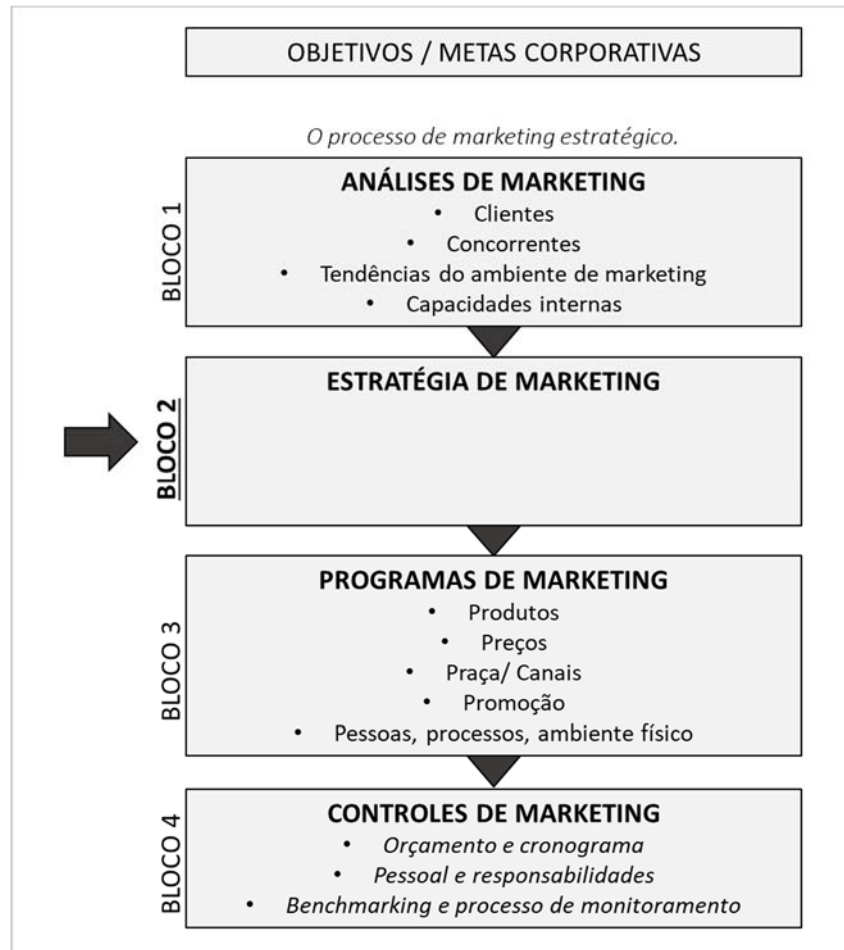


Figura 01: O processo de marketing estratégico. Adaptado de Simking (2008) apud Baker e Saren (2010)

O bloco 2, indicado com uma seta na figura 01, encontra-se incompleto. Diante disso:

- a) Cite e descreva quais informações estão faltando no bloco 02 da figura 01 para que a figura passe a representar de forma completa o processo de marketing estratégico.

RESPOSTA:

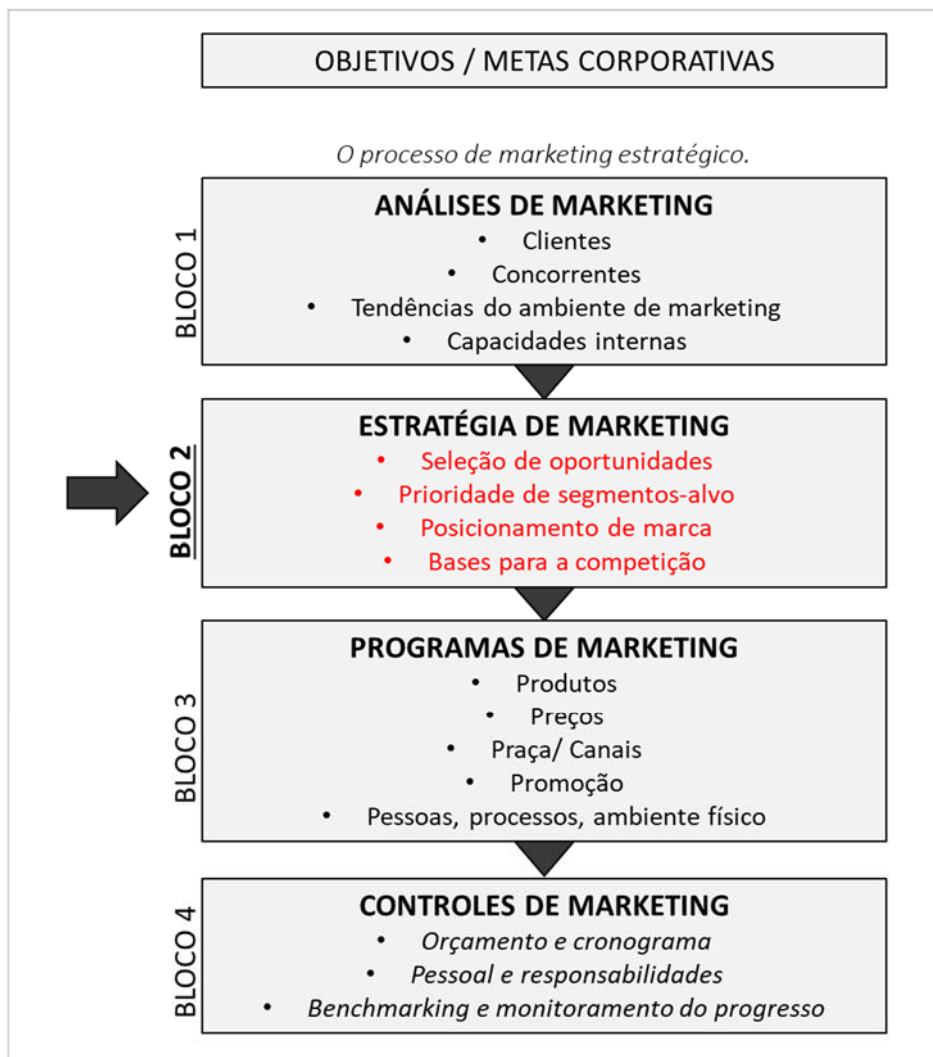


Figura 01: O processo de marketing estratégico. Adaptado de Simking (2008) apud Baker e Saren (2010)
PAGINA 244

A estratégia de marketing refere-se à tomada de decisões. A partir das análises mercadológicas é possível **selecionar quais oportunidades** de mercado explorar com vista a alcançar uma vantagem econômica duradoura. Ao mesmo tempo, deve-se **priorizar segmentos-alvo**, que consiste em decidir quais grupos de consumidores apresentam maior oportunidade. Por sua vez, o **posicionamento de marca**, implica em desenvolver, mensurar e gerir a marca com o objetivo de maximizar seu valor. A estratégia de marketing contempla, ainda, a definição de **bases para competição**, selecionando a proposta de valor adequada ao segmento-alvo.

NOTA: A estratégia de marketing visa fornecer uma resposta a uma das questões dos mercados competitivos: como reconhecer e alcançar uma vantagem econômica duradoura. A estratégia de marketing consiste no equilíbrio entre a formulação da estratégia e a implementação estratégica, esta última representada pelos programas de marketing. A segmentação, seleção de alvo (*targeting*) e posicionamento é a essência do marketing estratégico.

b) Explique de que forma a segmentação de mercado influencia o processo de marketing estratégico.

RESPOSTA:

A segmentação de mercado é fundamental para o processo de marketing estratégico. Como as necessidades dos clientes tornam-se cada vez mais diversificadas, a segmentação de mercado oferece às organizações meios para lidar com essa complexidade. É a partir da segmentação que as organizações tomam decisões estratégicas de posicionamento mercadológico, alocação eficiente de recursos e programas de marketing mais adequados às necessidades dos clientes.

QUESTÃO 02 (10,0 pontos)

Uma das temáticas de comportamento do consumidor, que instiga investigações, refere-se aos processos de comportamento pós-compra.

Explique a diferença entre comportamento de compra repetida e fidelidade à marca.

RESPOSTA:

Comportamento de compra repetida significa que o consumidor está comprando um produto repetidamente, sem nenhum sentimento especial pelo mesmo. O conceito de fidelidade à marca, em contrapartida, implica que o consumidor tem alguma preferência real pela marca.

QUESTÃO 03 (10,0 pontos)

“As técnicas de amostragem são genericamente classificadas como não probabilísticas e probabilísticas” (Malhotra, 2010, p.274).

Definições sobre o modelo de amostragem podem fazer parte do processo de elaboração de uma pesquisa de marketing. Uma dessas definições consiste na escolha da(s) técnicas(s) de amostragem.

a) Diferencie amostragem não probabilística de amostragem probabilística

RESPOSTA:

A amostragem não probabilística confia no julgamento pessoal do pesquisador, e não no acaso, para selecionar os elementos da amostra. Dessa forma, não utiliza seleção aleatória. Por sua vez, a amostragem probabilística é o processo de amostragem em que cada elemento da população tem uma chance fixa de ser incluída na amostra.

b) Cite e explique duas técnicas de amostragem não probabilística.

RESPOSTA:

- Amostragem por conveniência: procura obter uma amostra de elementos convenientes. A seleção das unidades amostrais é deixada a cargo do entrevistador. Com frequência, os entrevistados são escolhidos porque se encontram no lugar exato no momento certo.

- Amostragem por julgamento: é uma forma de amostragem por conveniência em que os elementos da população são selecionados com base no julgamento do pesquisador. Este, exercendo seu julgamento ou aplicando sua experiência, escolhe os elementos a serem incluídos na amostra, pois os considera representativos da população de interesse ou apropriados por algum motivo.

- Amostragem por quotas: consiste numa amostragem em dois estágios. O primeiro estágio consiste em definir as categorias, ou quotas, com base em características relevantes de controle, que podem incluir, gênero, idade, dentre outros. As quotas buscam assegurar que a composição da amostra seja a mesma da população de interesse, com respeito às características definidas. No segundo estágio, os elementos da amostra são selecionados com base em conveniência ou julgamento.

- Amostragem bola de neve ou autogerada: escolhe-se um grupo inicial de entrevistados e se solicita que identifiquem outros que pertençam à população-alvo de interesse. Os entrevistados subsequentes são selecionados com base nessas referências.

c) Cite e explique duas técnicas de amostragem probabilística.

RESPOSTA:

- Amostragem aleatória simples: cada elemento da população tem uma probabilidade conhecida e igual de ser escolhido. Cada elemento é escolhido independente de qualquer outro elemento. A amostra é extraída por um processo aleatório.

- Amostragem sistemática: escolhe-se uma amostra selecionando um ponto de partida aleatório e, em seguida, extraindo cada *i*-ésimo elemento sucessivamente do arcabouço amostral.

- Amostragem estratificada: processo de amostragem em duas etapas em que a população é dividida em subpopulações ou estratos. Todo elemento da população deve ser distribuído para somente um estrato e nenhum elemento deve ser omitido. Na segunda etapa, os elementos são selecionados de cada estrato por um processo aleatório.

- Amostragem por grupo ou por *cluster*: divide-se primeiro a população-alvo em subpopulações mutuamente excludentes e coletivamente exaustivas, ou *clusters* (conglomerados). Seleciona-se então uma amostra aleatória de *clusters* com base em uma técnica de amostragem probabilística. Para cada *cluster* selecionado, incluem-se todos os elementos na amostra, ou se extrai uma amostra de elementos de forma probabilística.

QUESTÃO 04 (10,0 pontos)

“Experimentação é a técnica de pesquisa usada na pesquisa causal” (Malhotra, 2010, p.161).
Considerando a citação acima:

a) Explique o conceito de causalidade numa pesquisa experimental.

RESPOSTA:

Causalidade ou inferência causal é relativa a uma mudança em uma variável que produz uma mudança em outra variável. Numa pesquisa experimental o pesquisador manipula e controla uma ou mais variáveis independentes e, em seguida, observa os efeitos que essas variáveis manipuladas têm sobre as variáveis dependentes.

b) Cite três condições para a causalidade.

RESPOSTA:

Pelo menos três condições devem ser satisfeitas para justificar a inferência de uma relação causal entre duas variáveis: (1) variação concomitante; (2) ordem de tempo da ocorrência das variáveis; e (3) ausência de outros possíveis fatores causais.

c) Uma relação causal pode ser demonstrada de forma conclusiva? Justifique sua resposta.

RESPOSTA:

Não. A combinação de evidência da causalidade na forma de variação concomitante, ordem de tempo da ocorrência das variáveis e eliminação de outros possíveis fatores causais não garante a existência de uma relação causal. Entretanto, um acúmulo de evidências consistentes aumenta nossa confiança de que exista uma relação causal.

QUESTÃO 05 (10,0 pontos)

No contexto em transformação influenciado pelo digital, pode-se dizer que a sociedade e o consumidor mudaram. Segundo Gabriel (2010), o marketing também mudou, levando em consideração novas ferramentas e plataformas que se tornaram disponíveis.

Considere as afirmações abaixo relacionadas ao marketing digital e responda às questões apresentadas a seguir.

**a) “Com a proliferação e difusão das plataformas e mídias digitais no cotidiano das pessoas, misturando-se a plataformas e mídias tradicionais, dois fenômenos importantes passaram a ocorrer intensamente: a crossmedia (ou transmídia) e a convergência” (Gabriel, 2010. p. 110).
Explique os dois fenômenos citados: a transmídia e a convergência.**

RESPOSTA:

Transmídia é o uso integrado das mídias, de forma que uma história ou mensagens ultrapasse os limites de um único meio. Pode ser definida com o suporte colaborativo de múltiplas mídias para entregar uma única história ou tema. Enquanto a transmídia envolve a distribuição de um conteúdo por diversas mídias, a convergência envolve o processo contrário. A convergência ocorre quando tecnologias que eram usadas separadas passam a compartilhar o mesmo meio e interagem umas com as outras de forma sinérgica, criando novas funcionalidades.

b) “Conhecer e compreender as possibilidades que as realidades mistas oferecem podem trazer vantagem competitiva estratégica para ações com diversos públicos-alvo em marketing” (Gabriel, 2010. p. 129).

Defina o que é realidade mista e explique como as realidades mistas podem trazer novas opções de ações mercadológicas para as empresas.

RESPOSTA:

Define-se realidade mista como a fusão dos mundos físico e virtual para produzir novos ambientes e visualizações onde objetos físicos e digitais coexistem e interagem em tempo real. São várias as aplicabilidades da realidade mista às ações mercadológicas, compreendendo realidade virtual e realidade aumentada, por exemplo, que tem como efeito a mobilidade, ao permitir trazer ao ambiente de realidade física as camadas digitais virtuais que a ampliam. Tecnologias de geolocalização, entretenimento, educação e treinamento, virtual demo (demonstrações virtuais de produtos e projetos), eventos aumentados, e jogos imersivos constituem aplicações da realidade mista e são opções de ações mercadológicas para as empresas.

c) Segundo Gabriel (2010), após o desenvolvimento das ações de marketing digital, é essencial estabelecer quais indicadores serão mensurados para analisar os resultados e verificar se os objetivos iniciais do plano de marketing foram atendidos. A autora tipifica os KPI (*Key Performance Indicators*) em: volume, engajamento e conversão.

Considerando cada tipo de KPI, relacione dois indicadores para: volume, engajamento e conversão.

RESPOSTA:

Volume: estão relacionados à mensuração de quantidades puras. Exemplos: Número de fãs, *followers* (seguidores), tweets por minuto, *trending topics*.

Engajamento: estão relacionados às mensurações de ações que envolvam engajamento, influência e relevância. Exemplos: quantidade de retweets, quantidade de assinantes, sentimento, tempo gasto, comentários, réplicas e menções.

Conversão: são os que medem as ações de conversão do público-alvo em função dos objetivos estipulados. Exemplos: vendas, CTR (*click through*), geração de *leads* (cadastro de prospectos).

QUESTÃO 06 (10,0 pontos)

Sites, como o Google, Yahoo, Bing e MSN, são exemplos de plataformas de busca que podem ser utilizadas de forma estratégica, para objetivos específicos de aumento de tráfego para as páginas das empresas.

Considere as afirmações abaixo relacionadas à temática de SEM (*Search Engine Marketing*) e responda às questões apresentadas a seguir.

a) Os links apresentados como resultados naturais em sites de busca são obtidos por uma análise e ranqueamento de páginas em função da incidência da palavra-chave na busca.

Explique como melhorar o tráfego de um site por meio de buscas orgânicas, utilizando palavras-chave.

RESPOSTA:

A determinação e seleção das palavras-chave consistem numa etapa decisiva de SEM. Por isso, para melhorar o tráfego de um site por meio de buscas orgânicas, primeiramente deve-se determinar quais as palavras-chaves importantes para o negócio e fazer uma minuciosa pesquisa sobre quais delas realmente são usadas pelo público-alvo do website. Além disso, para aperfeiçoar e ser mais assertiva, esta estratégia deve: incluir testes na seleção das melhores palavras; atentar-se para possíveis erros de digitação nas buscas por parte do usuário; evitar palavras chaves muito concorridas (neste caso sugere-se comprar a palavra-chave); balancear o uso de palavras-chave genéricas e específicas; e buscar não otimizar organicamente um site para campanhas temporárias (o que pode ser aperfeiçoado por meio de links patrocinados).

b) Além das palavras-chave, há também outro fator externo à página que afeta o posicionamento dela nas buscas: sua relevância (*page rank*).

Explique como esse fator influencia no posicionamento do site em uma busca.

RESPOSTA:

A relevância de uma página diz respeito aos fatores relativos ao relacionamento que a página atrai e não está associada a palavras-chave e sim a importância geral que uma página tem na web. A relevância é um índice de 0 a 10 e este índice é um dos critérios utilizados para determinação do posicionamento do site em uma busca. Sendo assim, pode-se dizer que o posicionamento do site em uma busca também é determinado pelo quanto uma página é relevante em comparação a todas as outras páginas existentes.

QUESTÃO 07 (10,0 pontos)

Um pesquisador deseja explicar a atitude em relação aos carros esportivos em termos do número de anos em que o entrevistado tenha tido um carro esportivo. A atitude é medida em uma escala de 11 pontos (1 = não gosto de carros esportivos e 11 = gosto muito de carros esportivos). A duração da propriedade do carro é medida pelo número de anos em que o entrevistado teve um ou mais carros esportivos. Usou-se um programa de computador para gerar a regressão da atitude sobre a duração da propriedade de carros esportivos. Os resultados obtidos na realização de um pré-teste são demonstrados na tabela 01.

| <i>Estatística de regressão</i> | | | | | |
|---------------------------------|----------------------|---------------------------|-----------------------|----------|----------------|
| R múltiplo | | 0,9361 | | | |
| Quadrado de R | | 0,8762 | | | |
| Quadrado de R ajustado | | 0,8639 | | | |
| Erro-padrão | | 1,2233 | | | |
| ANOVA | | | | | |
| | <i>gl</i> | <i>Soma dos quadrados</i> | <i>Quadrado médio</i> | | |
| Regressão | 1 | 105,9522 | 105,9522 | | |
| Residual | 10 | 14,9644 | 1,4964 | | |
| $F = 70,8027$ | | valor $p = 0,0000$ | | | |
| | <i>Coefficientes</i> | <i>Erro-padrão</i> | <i>Beta (B)</i> | <i>t</i> | <i>Valor p</i> |
| (Constante) | 1,0793 | 0,7434 | | 1,452 | 0,1772 |
| Duração | 0,5897 | 0,7000 | 0,9361 | 8,414 | 0,0000 |

Tabela 01: Regressão bivariada

a) Qual o percentual (%) da variação de atitude em relação aos carros esportivos é explicada pela variação de duração da propriedade do carro?

RESPOSTA:

87,62% ou 86,39%. O coeficiente de determinação (R^2 ou quadrado de R) significa a proporção do total de variação em Y que é considerada pela variação de X. Também é considerado o quadrado de R ajustado.

b) Existe uma relação estatisticamente significativa, com 95% de confiança, entre a variável dependente e independente? Qual o indicador utilizado para embasar sua resposta?

RESPOSTA:

Sim, a relação é significativa em $\alpha = 0,05$. A estatística F calculada ou os resultados do teste t podem ser utilizados para embasar a resposta, tendo em vista que o valor p é menor que α (0,05).

c) Escreva a equação estimada do modelo de regressão bivariada.

RESPOSTA:

Atitude (Y) = $1,0793 + 0,5897x$, onde x = tempo de duração da propriedade do carro

QUESTÃO 08 (10,0 pontos)

A Análise Fatorial Exploratória (AFE) foi utilizada para desenvolver uma escala para mensurar a “Orientação para Marketing Analytics (OMA)”, coletando dados em uma amostra de 127 profissionais de marketing e especialistas da área. Uma AFE inicial foi realizada com 29 itens da escala, utilizando a extração por componentes principais e retraindo fatores com autovalores maiores que 1, resultando na extração de 6 fatores. Esta solução apresentou uma medida Kaiser-Meyer-Olkin (KMO) de adequação de amostra igual a 0,913 e Teste de esfericidade de Bartlett significativo (com 5% de significância).

Análises adicionais indicaram que o mais adequado seria a extração de 2 fatores, além de ser necessário excluir 8 itens da escala original. Assim, fez-se uma solução final da AFE com os itens remanescentes, extraíndo 2 fatores e aplicando a rotação direct oblím. Nesta solução final, os autovalores, durante a extração dos fatores, foram de 12,000 e 2,000, aproximadamente. Considerando o exposto, as condições e os critérios para aplicação da AFE definidos por Hair et al. (2014), responda:

a) Quais fontes de variância (total, comum e do erro) o autor utilizou em sua solução da AFE inicial?

RESPOSTA:

Segundo Hair et al. (2014) a extração por “componentes principais”, considera a variância total e deriva fatores que contêm pequenas proporções de variância única e, em alguns casos, variância de erros. Na análise fatorial apresentada foi utilizado o método de extração por componentes principais, onde todas as fontes de variância foram consideradas, ou seja, foi utilizada a variância total. Sabendo que a variância total é a soma da variância comum, única e da variância do erro.

b) Quais conclusões sobre as condições para a realização da AFE podem ser extraídas de acordo com os índices KMO e com o teste de esfericidade de Bartlett?

RESPOSTA:

A medida KMO se encontra acima de 0,900 indicando ótimos grau de intercorrelação entre os indicadores e a rejeição da hipótese nula do teste de esfericidade de Bartlett revela condições apropriadas para aplicação da AFE.

c) O tamanho da amostra para aplicar a AFE ao conjunto de variáveis da escala proposta é adequado? Justifique sua resposta.

RESPOSTA:

Com 29 itens e uma amostra de 127 respondentes, existem 4,38 observações na amostra para cada indicador na AFE, valores ligeiramente inferiores ao patamar mínimo de 5 sugerido por Hair et al (2014). No entanto, como destacam os autores:

“Em relação à questão do tamanho da amostra, o pesquisador geralmente não analisaria uma amostra de menos de 50 observações, e preferencialmente o tamanho da amostra deve ser 100 ou maior.” (p. 100). Assim, pode-se julgar como aceitável o tamanho da amostra para aplicação da AFE neste caso, especialmente após a exclusão de 8 itens da escala original (neste caso haveria 6,05 observações na amostra para cada variável).

d) O método de rotação aplicado na solução da AFE final é ortogonal ou oblíquo? Quais as implicações do método utilizado para fins da interpretação final da escala?

RESPOSTA:

Oblíquo. Segundo Hair et al. (2014) as rotações oblíquas são semelhantes às rotações ortogonais, exceto que rotações oblíquas permitem fatores correlacionados em vez de manter a independência entre os fatores rotativos. Empregou-se a rotação "direct oblimin", o que implica dizer que é um método que permite e assume que os fatores estejam e possam estar correlacionados.

e) Qual o percentual de variância extraída na solução final da AFE? Explique, com base neste indicador, o grau de adequação da solução final da AFE.

RESPOSTA:

Pode-se estimar o percentual de variância como a soma dos autovalores dos fatores extraídos (12+2 = 14), divididos pela soma total dos autovalores, que é igual ao número de itens da escala (21). Ou seja, o percentual de variância extraída é de 66,67%, valor que está acima do patamar sugerido para as ciências sociais, conforme detalhado abaixo:

"Em contraste, nas ciências sociais, onde a informação é muitas vezes menos precisa, não é incomum considerar uma solução que representa 60% da variação total (e, em alguns casos, ainda menos) como satisfatória." (HAIR et al., 2014, p. 107)

QUESTÃO 09 (10,0 pontos)

A seguir são apresentados três cenários fictícios relacionados aos conceitos de marketing analytics.

a) Market Share

No ano de 2021, uma multinacional do setor de eletrônicos vendeu 20 milhões de celulares no Brasil, faturando R\$ 20 bilhões. Nesse ano, a empresa obteve 40% de participação de mercado em termos de volume (unidades vendidas). Parte de sua estratégia é vender produtos com preço abaixo da média de mercado: os preços dos seus celulares são em média 37,5% inferiores ao preço médio dos celulares do mercado.

Com base nessas informações, calcule o *market share* da empresa em função do faturamento.

Etapas de cálculo:

- Encontra-se a demanda de mercado em volume, dividindo o volume de vendas da empresa, 20 milhões, pelo seu *market share* de volume, 40%.

$$\text{Demanda de mercado: } \frac{20.000.000}{0,40} = 50.000.000 \text{ ou } 50 \text{ milhões}$$

- Encontra-se o preço médio dos celulares praticado pela empresa, dividindo o faturamento total, pelo número de unidades vendidas.

$$\text{Preço Médio da Empresa: } \frac{20 \text{ bilhões}}{20 \text{ milhões}} = 1000$$

- Baseado na informação de que o preço dos celulares é em média 37,5% inferior que a média de mercado, chega-se ao preço médio do mercado.

$$\text{Preço Médio de Mercado (PMM)} = \frac{\text{Preço Médio da Empresa (1.000)}}{(1 - 0,375)}$$

$$\text{Preço Médio de Mercado (PMM)} = 1600$$

- Sabendo que o Preço Médio de Mercado (PMM) é 1.600, estima-se a demanda total do mercado, em termos de faturamento, multiplicando o preço médio de mercado pelo volume total de mercado.

$$\begin{aligned} \text{Demanda total do mercado (faturamento)} &= 50.000.000 \times 1600 \\ &= 80.0000.000.000 \text{ ou } 80 \text{ bilhões} \end{aligned}$$

- Finalmente, chega-se ao *marketing share* por faturamento da empresa, dividindo o faturamento da empresa pela demanda total de mercado, em termos de faturamento.

$$\text{Market share (faturamento)} = \frac{20.000.000.000}{80.0000.000.000}$$

$$\text{Market share (faturamento)} = 0,25, \text{ ou seja, } 25\%$$

b) Customer Lifetime Value

A taxa de retenção de uma operadora de celular é de 90% por ano. A empresa aplica uma taxa de desconto de 10% anual sobre seus investimentos e a margem média dos clientes é de R\$ 200 mensais.

Usando a fórmula do multiplicador de margem, determine o Valor da Vida Útil do Cliente (Customer Lifetime Value – CLTV).

RESPOSTA:

$$CLTV = \frac{\text{Margem} \times \text{Retenção}}{(1 + \text{Taxa de desconto} - \text{Retenção})}$$

$$CLTV = \frac{R\$200 \times 0,90}{(1+0,10-0,90)}$$

$$CLTV = R\$ 900$$

c) Return on Marketing Investment

Um e-commerce de moda feminina recebe em média 50 mil visitas mensais em seu site. O ticket médio do site é de R\$ 100 e a taxa de conversão é de 2% das visitas em vendas com faturamento efetivo (*check-out* com pagamento). A empresa pratica um *markup* de 100% em relação ao custo de aquisição das mercadorias. Outros custos fixos do site somam R\$ 10 mil mensais. A empresa está avaliando a contratação de uma consultoria de otimização da taxa de conversão (*Conversion Rate Optimization* - CRO) que custará R\$ 20 mil. A consultoria tem o objetivo de duplicar a taxa de conversão do site, que chegará a 4%.

Considerando o modelo proposto por Farris et al, (2010), calcule a expectativa de retorno sobre o investimento de marketing (Marketing Return On Investment) da consultoria de CRO.

É dado:

$$\text{Retorno sobre o investimento de marketing (ROMI)} = \frac{\left[\text{Receita incremental atribuída ao Marketing (\$)} * \text{Margem de contribuição (\%)} - \text{Gastos em marketing (\$)} \right]}{\text{Gastos em marketing (\$)}}$$

RESPOSTA:

| | Cenário atual | Campanha | Incremental |
|--|---|--|---|
| Visitas no site | 50.000 | 50.000 | |
| Ticket médio do site | \$100 | \$100 | |
| Taxa de conversão | 2% | 4% | |
| Check-outs <i>Visitas x conversão</i> | 1.000 (50.000 x 0,02) | 2.000 (50.000 x 0,04) | |
| Faturamento <i>check-outs x ticket médio</i> | \$100.000 (1.000 x 100) | \$200.000 (2.000 x 100) | \$100.000 <i>Receita incremental atribuída ao marketing</i> |
| Mark up | 100% | 100% | |
| Custo variável <i>ticket médio</i> <i>(1 + mark up)</i> | \$50 $\frac{100}{(1 + 1)}$ | \$50 $\frac{100}{(1 + 1)}$ | |
| Custo fixo | \$10.000 | \$10.000 | |
| Custo total <i>(Custo variável x check out)+ Custo fixo</i> | \$60.000 (50 x 1.000) +10.000 | \$110.000 (50 x 2.000) +10.000 | |
| Investimento em marketing incremental | - | \$20.000 | |
| Margem Bruta <i>Faturamento – Custo total</i> | \$40.000 (100.000- 60.000) | \$90.000 (200.000 - 110.000) | \$70.000 (90.000- 20.000*) *investimento em marketing |
| Margem percentual <i>margem bruta</i> <i>faturamento</i> | 40% $\frac{40.000}{100.000}$ | 45% $\frac{90.000}{200.000}$ | |

Portanto,

(i) Receita incremental atribuída ao marketing = \$100.000

(ii) Margem de contribuição (%) = 45%

(iii) Gastos em marketing = \$ 20.000

Aplicando a fórmula:

$$\text{Retorno sobre o investimento de marketing (ROMI)} = \frac{\left[\text{Receita incremental atribuída ao Marketing (\$)} * \text{Margem de contribuição (\%)} - \text{Gastos em marketing (\$)} \right]}{\text{Gastos em marketing (\$)}}$$

$$ROMI = \frac{(100.000 \times 0,45) - 20.000}{20.000}$$

$$ROMI = \frac{25.000}{20.000} \quad ROMI = 1,25$$

Expectativa de retorno = 125%

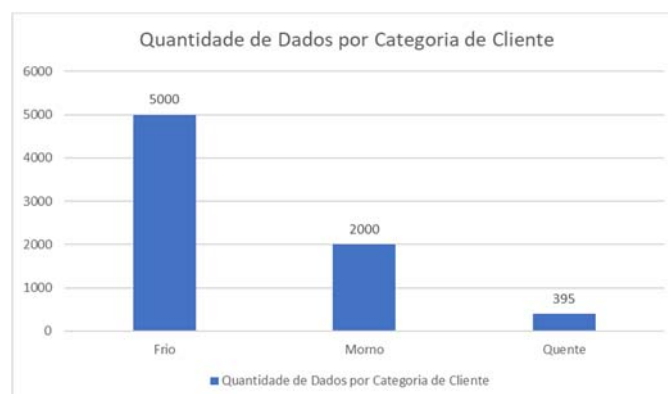
QUESTÃO 10 (10 pontos)

Visando direcionar as campanhas promocionais de marketing para os clientes com maior capacidade de retenção, a empresa PJDECOR extraiu 03 (três) meses de dados do seu sistema de faturamento.

Manualmente, a PJDECOR classificou os seus clientes nas seguintes categorias: QUENTE, MORNO e FRIO. Deste modo, a PJDECOR irá oferecer maiores descontos e promoções para os clientes classificados como QUENTES.

Com essa base de dados, a PJDECOR solicita que seja elaborado um modelo de aprendizagem de máquina que seja capaz de classificar se um novo cliente, recém captado pela empresa, será QUENTE, MORNO ou FRIO para fidelizá-lo e garantir a sua retenção.

Na etapa de pré-processamento do modelo, a quantidade de dados apresentados pela PJDECOR para cada categoria de clientes foi assim representada (GRÁFICO 01):



Para gerar o modelo de classificação, utilizou-se da técnica de *holdout* para dividir a base de dados entre Treino e Teste (70% Treino e 30% Teste), sem técnicas de *over-sampling* ou *under-sampling*, mantendo a base de dados original.

Após o processamento do modelo de aprendizado supervisionado, foi elaborada a matriz de confusão (FIGURA 02).

| | | | | |
|-------------------|--------|----------------|-------|--------|
| CLASSE VERDADEIRA | Frio | 1421 | 150 | 50 |
| | Morno | 80 | 399 | 30 |
| | Quente | 8 | 50 | 31 |
| | | Frio | Morno | Quente |
| | | CLASSE PREDITA | | |

Considere:

| | |
|--|--|
| <p>Taxa de Erro</p> $err(\hat{f}) = \frac{FP + FN}{n}$ | <p>Acurácia</p> $Ac(\hat{f}) = \frac{VP + VN}{n}$ |
| <p>Precisão</p> $prec(\hat{f}) = \frac{VP}{VP + FP}$ | <p>Sensibilidade</p> $sens(\hat{f}) = \frac{VP}{VP + FN}$ |
| <p>Especificidade</p> $esp(\hat{f}) = \frac{VN}{VN + FP}$ | |

VN = Verdadeiro Negativo
 VP = Verdadeiro Positivo
 FP = Falso Positivo
 FN = Falso Negativo
 n = total de elementos

A partir do resultado da Matriz de Confusão, responda as seguintes questões:

a) Explique o funcionamento da Matriz de Confusão e descreva como devem ser interpretados os seus valores.

RESPOSTA:

A Matriz de Confusão ajuda a compreender o desempenho de um modelo de classificação.

Na matriz podem ser observados:

1. Verdadeiros Positivos (VP) que corresponde ao número de exemplos da classe positiva que foram classificados corretamente.

2. Verdadeiros Negativo (VN) que corresponde ao número de exemplos da classe negativa que foram classificados corretamente.
3. Falso Positivos (FP) que corresponde ao número de exemplos que foram classificados erroneamente como pertencentes a uma classe Positiva quando são pertencentes à classe negativa. Também conhecido como Erro do Tipo I no Teste de Hipóteses, ou seja, rejeitar a hipótese nula (Ho) quando ela é verdadeira e absoluta.
4. Falso Negativo (FN) que corresponde ao número de exemplos pertencentes originalmente à classe positiva mas foram incorretamente preditos como classe negativa. Também conhecido como Erro do Tipo II no Teste de Hipóteses, ou seja, a hipótese nula (Ho) é falsa e não foi rejeitada.

A soma dos VP, VN, FP e FN é o total de elementos utilizados pela base de teste para fazer a previsão do modelo. Na diagonal da Matriz encontram-se os VPs. No caso da existência de mais de uma classe, considera-se como Classe Positiva, a classe que está sendo analisada (Frio, Morno e Quente) e as demais como Classes Negativas. Exemplo: se estou analisando a Classe Frio, as classes Morno e Quente são consideradas como classes negativas em relação à Classe Quente. Assim encontramos os Falsos Positivos e Falsos Negativos

| | | | | | |
|-------------------|---------------|-----------------|----------------|---------------|----------------|
| Classe Verdadeira | Frio | 1421 (VP/VN) | 80 | 8 | 88 |
| | Morno | 150 | 399 (VP/VN) | 50 | 200 |
| | Quente | 50 | 30 | 31 (VP/VN) | 80 |
| Falso Positivo | | Frio | Morno | Quente | Falso Negativo |
| | | Classe Predita | | | |
| | | 200 | 110 | 58 | |

O modelo proposto teve o seguinte resultado:

Para a Classe Fria, previu 1421 elementos como Verdadeiros Positivos dos 1509 possíveis. Classificou erroneamente como pertencentes à classe Fria, ou seja, Falso Positivo, 200 elementos. Por sua vez, classificou como Falso Negativo 88 elementos que deveriam ter sido classificados como Frio, mas foram considerados como pertencentes a outras classes. Teve uma taxa de Erro considerada muito baixa, cerca de 5%.

$$Taxa\ de\ Erro = \frac{Falsos\ Negativos}{Verdadeiros\ Positivos + Falsos\ Negativos}$$

Para a Classe Morno, previu 399 elementos como Verdadeiros Positivos dos 599 possíveis. Classificou erroneamente como pertencentes à classe Morna, ou seja, Falso Positivo, 110 elementos. Por sua vez, classificou como Falso Negativo 200 elementos que deveriam ter sido classificados como Morno, mas foram considerados como pertencentes a outras classes. Teve uma taxa de Erro de cerca de 33%.

Para a Classe Quente, previu 31 elementos como Verdadeiros Positivos dos 111 possíveis. Classificou erroneamente como pertencentes à classe Quente, ou seja, Falso Positivo, 58 elementos. Por sua vez, classificou como Falso Negativo 80 elementos que deveriam ter sido

classificados como Quente, mas foram considerados como pertencentes a outras classes. Teve uma taxa de Erro considerada muito altíssima, cerca de 72%.

b) O desempenho do modelo atende à necessidade da PJDECOR? Justifique com a(s) métrica(s) de desempenho que fundamentem a sua decisão.

RESPOSTA:

A acurácia do Modelo foi de 83%. Isso se deu pelo grande número de acertos realizados pelo modelo na Classe Fria e pelo fato da base de dados estar desbalanceada. Quando analisamos a taxa de erro por cada classe, nota-se que o modelo não irá atingir o resultado esperado pelo Gestor, que é acertar a Classe Quente.

$$Acurácia = \frac{VP + VN}{n}$$

No caso em questão, a Matriz de Confusão demonstrou um grande número de acertos da Classe Frio, onde a Taxa de Erro foi de 5%. Na Classe Morno, a Taxa de Erro sobe para 33% e, chama a atenção a Taxa de Erro da Classe Quente (que é a classe interesse do Gestor) que tem a Taxa de Erro de 72%.

$$Taxa de Erro = \frac{Falsos Negativos}{Verdadeiros Positivos + Falsos Negativos}$$

Por isso, a Acurácia não é a métrica ideal medir o desempenho desse modelo. Deve-se substituir a Acurácia por outra métrica mais estratégica que irá verificar os acertos nas classes minoritárias.

Para tal, podemos usar métricas alternativas como precisão e recall (revocação ou sensibilidade), que permitem que o desempenho do modelo seja considerado com foco na classe minoritária (neste caso, a Classe Quente). A precisão pode ser vista como um medida de exatidão do modelo e a sensibilidade como uma medida de sua completude.

A precisão do modelo para a Classe Quente é de 35% portanto, muito baixa para um modelo de previsão.

$$Precisão = \frac{VP}{VP + FP}$$

E o Recall do modelo para a Classe Quente é de 28%. Também baixa para um modelo de previsão.

$$Recall = \frac{VP}{VP + FN}$$

Deste modo, apesar da acurácia do modelo ser de 83%, as métricas de precisão e recall demonstram que o desbalanceamento dos dados influenciou na modelo proposto. Globalmente, o modelo acerta muito a Classe Fria enquanto a Classe Quente, que é do interesse da PJDECOR tem uma taxa de erro da ordem de 72%, Precisão de 33%, e Recall de 27%.

O desempenho do modelo proposto não atende às necessidades da PJDECOR.