

SERVIÇO BRASILEIRO DE APOIO ÀS MICRO E PEQUENAS EMPRESAS (SEBRAE) PROCESSO SELETIVO EXTERNO PARA O PROVIMENTO DE VAGAS E A FORMAÇÃO DE CADASTRO DE RESERVA EM PERFIS DE ANALISTA TÉCNICO

PERFIL 14: ANALISTA TÉCNICO II – CIENTISTA DE DADOS

Prova Discursiva – Estudo de Caso

Aplicação: 08/09/2024

PADRÃO DE RESPOSTA DEFINITIVO

Entre as técnicas estatísticas, há um teste de hipóteses (teste estatístico ou teste de significância), como o teste para a média com variância desconhecida para avaliar a significância estatística sobre o aumento médio do faturamento. Outro teste que poderia ser considerado é o teste para amostras pareadas que compara o antes e depois da aplicação da política de incentivo para a empresa (nesse caso cada unidade amostral é uma empresa diferente). Ambos os testes consideram a distribuição t . Esses testes podem ser utilizados para testar suposições sobre parâmetros populacionais. O teste t é apropriado para comparar médias e determinar se o impacto observado é estatisticamente significativo, crucial para validar os efeitos dos incentivos. **Também poderia ser utilizado a análise de séries temporais interrompidas (ITSA) - *Interrupted Time Series* - ITS, pois é utilizada quando se tem dados sobre um resultado ao longo do tempo e se deseja compreender se e como o resultado mudou após uma intervenção, política ou programa implementado para a população inteira num determinado momento.**

Um possível modelo a ser utilizado é o modelo de regressão linear para analisar o comportamento do faturamento com base em variáveis como setor e localização, e um modelo ARIMA ou ARIMAX para análise de séries temporais em que as variáveis independentes podem ser também setor e localização, além de avaliar as autocorrelações presentes no comportamento histórico. Também poderia ser considerado um modelo de suavização exponencial dentro de séries temporais. A regressão linear é adequada para descobrir relações entre variáveis, enquanto o ARIMA ou ARIMAX é eficaz para prever padrões temporais com base em dados históricos e relações com outras variáveis, caso necessário.

O SQL seria utilizado para consultas e agregações no banco de dados e uma linguagem de programação, ou as duas, seria utilizada para modelagem e visualização dos dados. O SQL é eficiente para manipulação de grandes volumes de dados e as linguagens de programação oferecem flexibilidade e poder na modelagem estatística e visualização.

Para o caso apresentado, o Power BI seria o escolhido para construção de *dashboards*, pois é robusto e integrado, ideal para *dashboards* interativos e oferece serviços de modelagem de dados para análise preditiva. No que se refere às bibliotecas apresentadas, qualquer uma delas pode ser utilizada para a criação de visualização detalhadas, pois ambas são poderosas para criar visualizações customizadas e informativas.

QUESITOS AVALIADOS

Quesito 2 – Desenvolvimento do tema

Quesito 2.1 – Aprofundamento e domínio do tema

Conceito 0 – Não apresenta aprofundamento e domínio do tema.

Conceito 1 – Apresenta aprofundamento e domínio medianos do tema.

Conceito 2 – Apresenta aprofundamento e domínio satisfatórios do tema.

Quesito 2.2 – Argumentação e criticidade

Conceito 0 – Não argumenta nem demonstra criticidade no desenvolvimento do tema.

Conceito 1 – Apresenta argumentação e criticidade medianas.

Conceito 2 – Apresenta argumentação e criticidade satisfatórias.

Quesito 2.3 – Uso de exemplos e(ou) dados e(ou) citações

Conceito 0 – Não apresenta nenhum exemplo, dado ou citação.

Conceito 1 – Apresenta exemplo, dado ou citação, mas de forma precária.

Conceito 2 – Apresenta exemplo, dado ou citação de forma satisfatória.

Quesito 2.4 – Riqueza e consistência das ideias

Conceito 0 – Não apresenta riqueza nem consistência das ideias.

Conceito 1 – Apresenta riqueza e consistência medianas.
Conceito 2 – Apresenta riqueza e consistência satisfatórias.

Quesito 3 – Conhecimento técnico do tema

Quesito 3.1 – Uma técnica de inferência estatística que pode ser utilizada para testar a significância do impacto da política de incentivo à empresa

Conceito 0 – Não atendeu ao solicitado ou o fez de forma totalmente incorreta.
Conceito 1 – Apenas apresentou corretamente a técnica estatística, sem justificar o seu uso.
Conceito 2 – Apresentou corretamente a técnica estatística e justificou o seu uso.

Quesito 3.2 – Um tipo de modelo linear ou de série temporal que pode ser utilizado para prever o faturamento futuro da empresa: (i) modelo linear; (ii) série temporal

Conceito 0 – Não apresentou nenhum tipo de modelo ou série ou o fez de forma totalmente incorreta.
Conceito 1 – Apresentou qualquer um dos itens (i ou ii), sem justificar o seu uso.
Conceito 2 – Apresentou qualquer um dos itens (i ou ii) e justificou o seu uso.

Quesito 3.3 – Forma como o SQL e uma linguagem de programação (R ou Python) podem ser utilizados para realizar a análise de dados e desenvolver os modelos preditivos: (i) SQL e (ii) R/Python

Conceito 0 – Não atendeu ao solicitado ou o fez de forma totalmente incorreta.
Conceito 1 – Explicou corretamente apenas como um dos itens (i ou ii) pode ser utilizado na análise de dados e no desenvolvimento de modelos preditivos.
Conceito 2 – Explicou de forma parcialmente correta como os dois itens podem ser utilizados na análise de dados e no desenvolvimento de modelos preditivos.
Conceito 3 – Explicou corretamente como os dois itens podem ser utilizados na análise de dados e no desenvolvimento de modelos preditivos.

Quesito 3.4 – Ferramenta (Power BI ou Qlik Sense) e biblioteca (ggplot em R ou matplotlib em Python) a serem escolhidas para visualizar e comunicar os resultados: (i) Power BI ou Qlik Sense; (ii) ggplot ou matplotlib

Conceito 0 – Não indicou nem a ferramenta nem a biblioteca ou o fez de forma totalmente incorreta ou não justificada.
Conceito 1 – Indicou e justificou corretamente apenas um dos itens (i ou ii).
Conceito 2 – Indicou e justificou corretamente os dois itens (i ou ii).