



# Instituto Infnet

GRADUAÇÃO EM BANCO DE DADOS

ENGENHARIA DE PROMPTS PARA CIÊNCIA DE  
DADOS [24E4\_4]

WANDERSON RAFAEL MENDONÇA BATISTA

TESTE DE PERFORMANCE – TP2

PROF. THIAGO CIODARO XAVIER

RIO DE JANEIRO, 2024

Para acessar os fontes utilizado, consulte o git:

<https://github.com/wanderaf/promptTP02.git>

## Exercício 1 - Princípios de Criação de Prompts Eficazes

Alguns princípios são essenciais para a construção de prompts claros e obtenção de respostas precisas. Identifique e explique quais princípios foram utilizados no prompt abaixo:

Prompt

Como um especialista em turismo, liste 3 atrações turísticas imperdíveis de Paris e forneça uma breve descrição para cada uma delas. Aqui está um exemplo do formato esperado:

###

Atração 1: Torre Eiffel

1. Especificação do papel: define claramente o papel que o modelo deve assumir, ao solicitar “Como um especialista em turismo” é orientado que a resposta do modelo seja de um conhecedor do assunto;
2. Definição clara: A solicitação de “3 atrações turísticas imperdíveis de Paris e forneça uma breve descrição”, especifica o que precisa ser apresentado, sem deixar margem para equívoco;
3. Exemplo de resposta: Ao incluir um exemplo “Atração 1: Torre Eiffel”, o modelo demonstra como deseja receber as respostas, orientando a estrutura do resultado;
4. Contextualização: Ao limitar a abrangência territorial “Paris” é oferecido um contexto ao modelo, para que ele foque apenas neste local;

## Exercício 2 - Simulação de Atendimento ao Cliente com Diferentes Personas

Vamos simular uma situação de telemarketing onde persona e tom são críticos para o sucesso da interação. Para criarmos uma base de treinamento dos operadores, simule duas respostas para o mesmo atendimento, ora com um cliente agressivo, ora com um cliente tranquilo. Crie e teste um prompt (**Poe.com**) para cada situação a partir do atendimento:

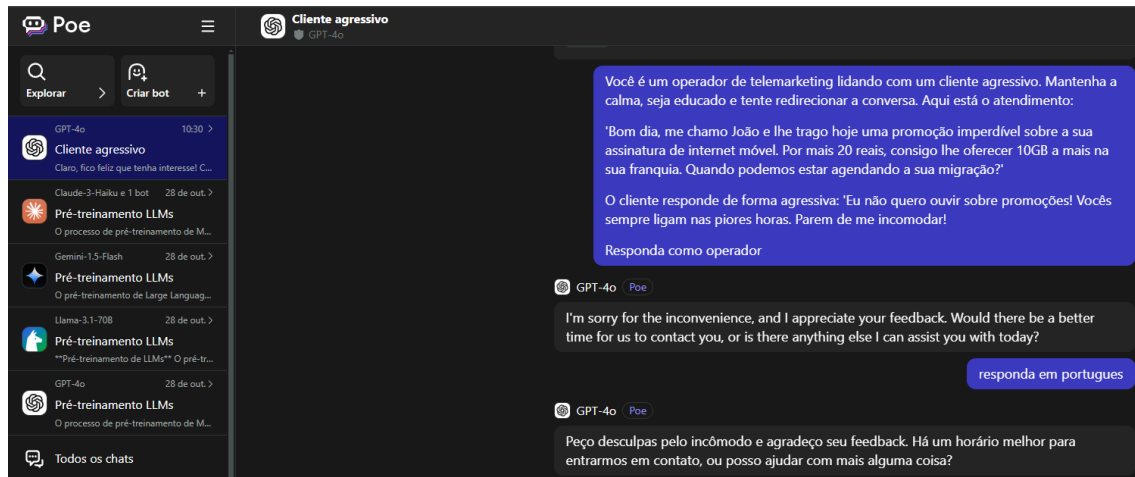
- “Bom dia, me chamo João e lhe trago hoje uma promoção imperdível sobre a sua assinatura de internet móvel. Por mais 20 reais, consigo lhe oferecer 10GB a mais na sua franquia. Quando podemos estar agendando a sua migração?”

Descreva o comportamento do cliente nas duas situações? As respostas saíram como o esperado?

Para realização desta atividade fiquei na dúvida em relação a “simule duas respostas para o mesmo atendimento”, entendo que aqui é para que eu elabore uma resposta do cliente

agressivo e tranquilo e com base na pergunta e resposta seja testado no Poe.com qual seria a resposta do operador (uma vez que a ideia é treinar os operadores).

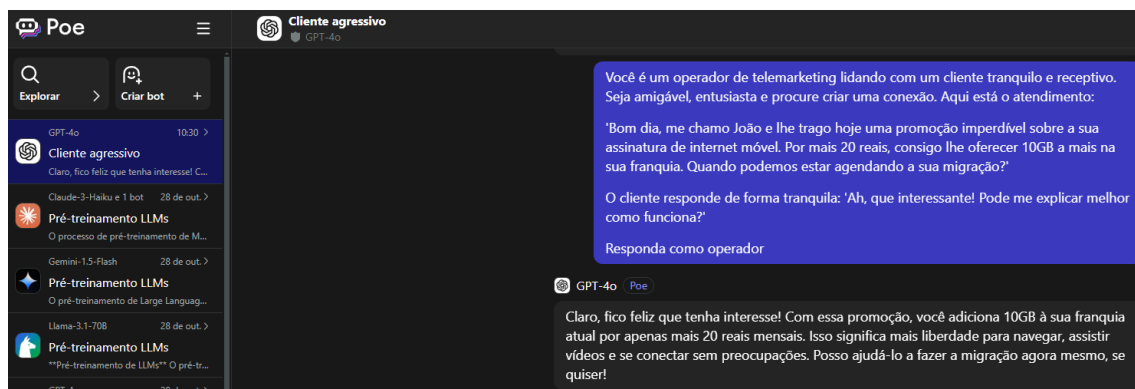
### Simulação com resposta agressiva



Neste caso a simulação aconteceu com um cliente que reclama do horário da ligação, demonstrando desinteresse e irritabilidade.

A resposta esperada do operador é um tom conciliador, paciente de forma a evitar a entrar em conflito, a resposta dada foi nesse sentido.

### Simulação com resposta tranquila:



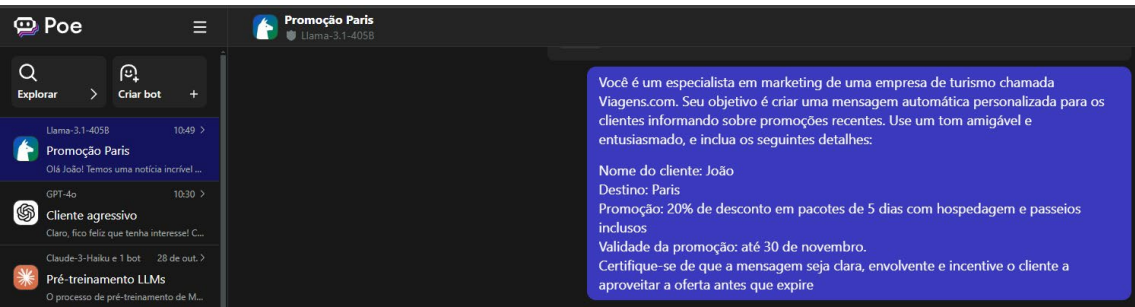
Já na simulação com o cliente tranquilo o memo foi amigável e pediu mais informações, neste caso a resposta esperada do operador deve ser em um tom animador, amigável, de forma que estimule o cliente a finalizar a mudança do plano, observa-se que a resposta também seguiu essas diretrizes.

### Exercício 3 - Automação de Mensagens Promocionais com Princípios de Prompting

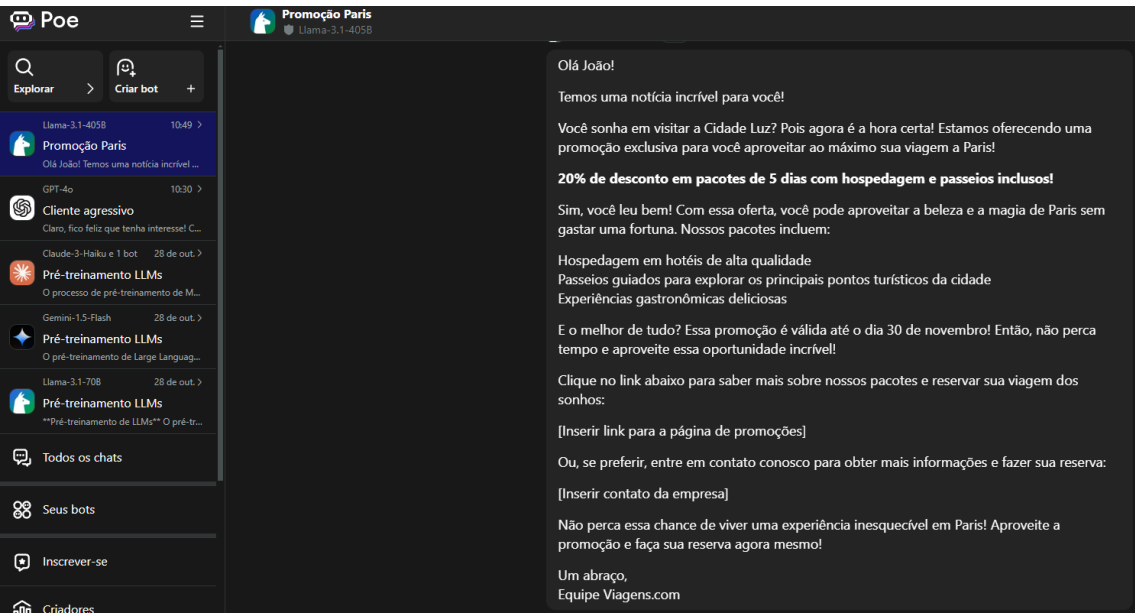
A empresa de turismo viagens.com deseja automatizar o processo de criação de envio de mensagens para clientes, informando-lhes sobre promoções recentes. Escreva um prompt

que contenha os três princípios de 1) persona, 2) dados e 3) tom para gerar a mensagem automática. Execute o prompt usando a sua conta Poe.com e justifique se o LLM respondeu como esperado.

Prompt utilizado:



Resposta:



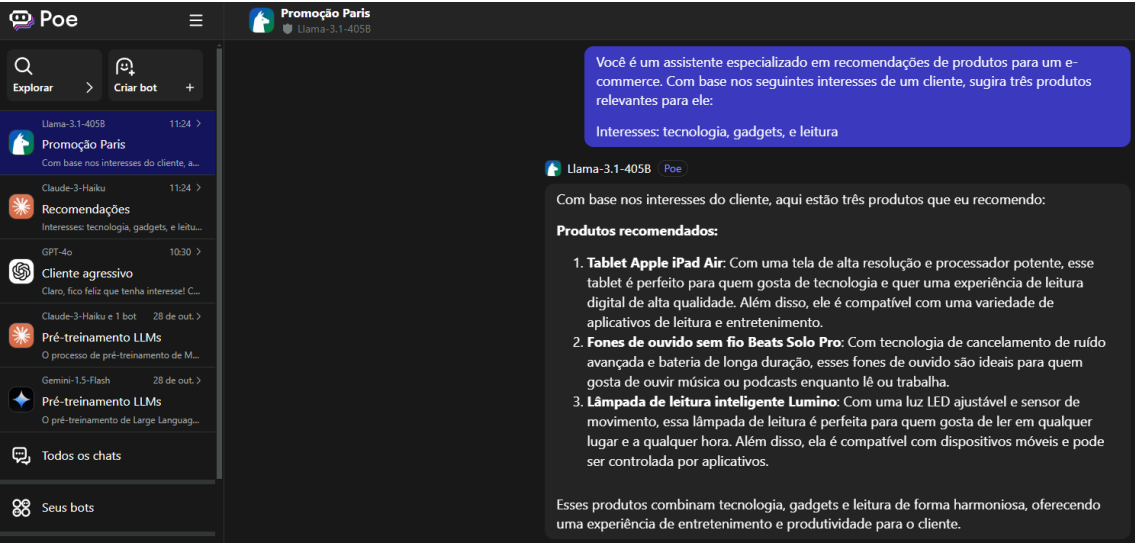
O modelo respondeu como esperado uma vez que adotou a persona (especialista em marketing), apresentou os dados (cliente, destino, desconto, validade da promoção) e utilizou um tom de entusiasmo para utilização da promoção.

Exercício 4 - Desenvolvimento de Prompts para Recomendação de Produtos

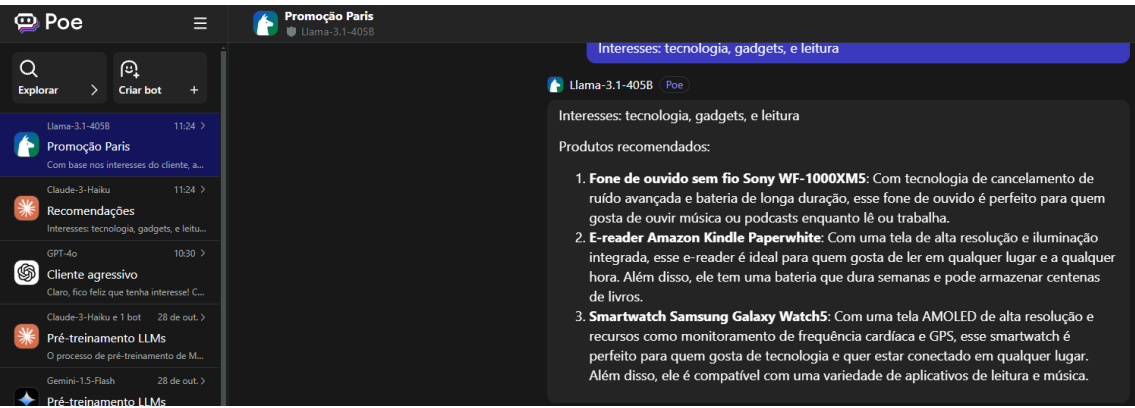
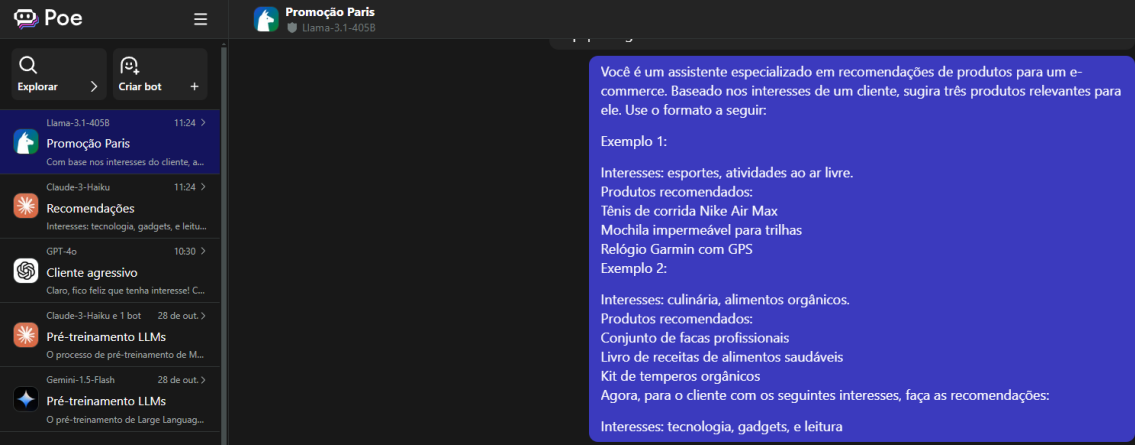
Você está desenvolvendo um sistema de recomendações de produtos para um e-commerce. Crie dois prompts com base nos princípios de prompting, sendo o primeiro um prompt simples e o segundo um prompt com exemplos (few-shot prompting). Utilize o Poe.com para testar os prompts com um LLM, comparando os resultados gerados. Qual dos prompts foi mais eficaz e por quê?



## Prompt Simples



## Prompt com exemplos



O prompt com exemplos foi mais eficiente, uma vez que os exemplos deixam a resposta mais clara, o prompt simples teve uma variação de resultados, o com exemplos apresentou uma sequencia melhor de sugestões.

## Exercício 5 - Listagem de Componentes de Computadores de Alto Desempenho

Utilize o Poe.com para criar um prompt simples que faça o LLM listar os principais componentes de um computador de alto desempenho, suas capacidades computacionais (HD, RAM, CPU, GPU...), marcas, modelos e preços. Teste o prompt e explique se o resultado atendeu às expectativas: as marcas e modelos existem? As capacidades do computador são de alto desempenho? Compare com fontes encontradas na internet.

Promoção Paris  
Llama-3.1-405B

Liste os principais componentes de um computador de alto desempenho, incluindo suas capacidades (HD, RAM, CPU, GPU), marcas, modelos e preços

Llama-3.1-405B Poe

Aqui estão os principais componentes de um computador de alto desempenho, incluindo suas capacidades, marcas, modelos e preços:

- 1. CPU (Processador)**
  - Marca: AMD Ryzen 9 7950X
  - Modelo: 16 núcleos, 32 threads, 4.5 GHz
  - Capacidade: Processamento de alta velocidade e multitarefa
  - Preço: cerca de R\$ 4.500,00
- 2. GPU (Placa de Vídeo)**
  - Marca: NVIDIA GeForce RTX 4090
  - Modelo: 24 GB de memória GDDR6X, 1.8 GHz
  - Capacidade: Renderização de gráficos em alta resolução e jogos em 4K
  - Preço: cerca de R\$ 10.000,00
- 3. RAM (Memória)**
  - Marca: Corsair Vengeance LPX
  - Modelo: 64 GB (4x16 GB), DDR5, 5200 MHz
  - Capacidade: Armazenamento de dados temporários e execução de aplicativos
  - Preço: cerca de R\$ 1.500,00

#### 4. HD (Disco Rígido)

- Marca: Samsung 970 EVO Plus
- Modelo: 1 TB, M.2 NVMe, 3000 MB/s
- Capacidade: Armazenamento de dados e programas
- Preço: cerca de R\$ 800,00

#### 5. Placa-Mãe

- Marca: ASUS PRIME X670E-UD
- Modelo: Suporte a CPU Ryzen 9, 4 slots de RAM DDR5, 2 slots de PCIe x16
- Capacidade: Conexão de todos os componentes do sistema
- Preço: cerca de R\$ 2.000,00

#### 6. Fonte de Alimentação

- Marca: EVGA SuperNOVA 850 GA, 80+ Gold
- Modelo: 850 W, 80 PLUS Gold, modular
- Capacidade: Alimentação de todos os componentes do sistema
- Preço: cerca de R\$ 500,00

#### 7. Gabinete

- Marca: Corsair Obsidian Series 7700
- Modelo: Suporte a placas-mãe ATX, 2 slots de PCIe x16, 4 slots de RAM DDR5
- Capacidade: Acomodação de todos os componentes do sistema
- Preço: cerca de R\$ 800,00

**Preço Total:** cerca de R\$ 20.100,00

Observação: Os preços podem variar dependendo do fornecedor e da região. Além disso, é importante verificar a compatibilidade entre os componentes antes de realizar a compra.

Em uma busca pelas marcas apresentadas, verificou-se que todas existem, agora ao buscar por “computador de alto desempenho” na internet, apresenta-se o conceito de “computação de alto desempenho” (HPC), que é bem diferente do que foi apresentado pela resposta do prompt, entretanto ao buscar por “configurações para um computador de alta performance” é apresentado sugestões de configurações de computadores gamers que são semelhantes às apresentadas pela resposta do prompt, como pode ser consultado em: <https://br.crucial.com/articles/for-gamers/the-best-specs-for-a-gaming-pc>.


### Exercício 6 - Análise dos Benefícios da Inteligência Artificial para Pequenas Empresas

Desenvolva um prompt simples para que o LLM forneça uma análise rápida sobre os benefícios de usar inteligência artificial em pequenas empresas. Qual foi o resultado gerado pelo LLM? Aplique os princípios de prompt para obter resultados mais precisos.



Teste os prompts na sua conta do Poe.com e copie os prompts e respectivas saídas, junto com a explicação do que foi feito entre um prompt e outro.

Prompt simples



IA pequenas empresas

Gemini-1.5-Pro

Quais são os benefícios de usar inteligência artificial em pequenas empresas?

Gemini-1.5-Pro

Poe


A inteligência artificial (IA) oferece uma série de benefícios para pequenas empresas, permitindo que elas otimizem operações, melhorem o atendimento ao cliente e impulsionem o crescimento. Aqui estão alguns dos principais benefícios:

1. Automação de Tarefas Repetitivas:

- **Redução de custos:** A IA pode automatizar tarefas rotineiras e manuais, como entrada de dados, agendamento de compromissos e atendimento inicial ao cliente, liberando funcionários para se concentrarem em atividades mais estratégicas e criativas. Isso reduz custos com mão de obra e minimiza erros humanos.
- **Aumento da eficiência:** A automação agiliza processos, permitindo que as pequenas empresas façam mais em menos tempo. Isso aumenta a produtividade e a eficiência operacional.

2. Melhora no Atendimento ao Cliente:

- **Disponibilidade 24/7:** Chatbots com IA podem fornecer suporte ao cliente 24 horas por dia, 7 dias por semana, respondendo a perguntas frequentes, resolvendo problemas simples e direcionando consultas mais complexas para a equipe humana. Isso melhora a satisfação do cliente e a fidelização.
- **Personalização da experiência:** A IA pode analisar dados do cliente para personalizar a interação e oferecer recomendações de produtos ou serviços relevantes, aumentando as chances de conversão.



IA pequenas empresas

Gemini-1.5-Pro

3. Análise de Dados e Tomada de Decisões:

- **Insights de negócios:** Ferramentas de IA podem analisar grandes volumes de dados para identificar tendências de mercado, padrões de comportamento do consumidor e oportunidades de crescimento. Isso permite que as pequenas empresas tomem decisões mais informadas e estratégicas.
- **Previsão de demanda:** A IA pode prever a demanda futura de produtos ou serviços, ajudando as empresas a otimizar o estoque, evitar desperdícios e atender às necessidades dos clientes de forma mais eficiente.

4. Marketing e Vendas Mais Eficientes:


- **Segmentação de público:** A IA pode segmentar o público-alvo com base em dados demográficos, interesses e comportamento online, permitindo que as pequenas empresas direcionem suas campanhas de marketing de forma mais precisa e eficaz.
- **Automação de marketing:** Ferramentas de IA podem automatizar tarefas de marketing, como envio de e-mails personalizados, gerenciamento de redes sociais e criação de anúncios online, otimizando o alcance e o retorno do investimento.

5. Redução de Riscos:

- **Deteção de fraudes:** A IA pode identificar padrões suspeitos e anomalias em transações financeiras, ajudando a prevenir fraudes e proteger as pequenas empresas contra perdas.
- **Análise de crédito:** Algoritmos de IA podem avaliar o risco de crédito de clientes em potencial, auxiliando na tomada de decisões sobre concessão de crédito e minimizando o risco de inadimplência.

Apesar dos inúmeros benefícios, é importante lembrar que a implementação de IA requer planejamento e investimento. As pequenas empresas devem avaliar suas necessidades específicas, escolher as ferramentas adequadas e treinar seus funcionários para utilizar a tecnologia de forma eficaz. Com a estratégia certa, a IA pode ser uma poderosa ferramenta para impulsionar o crescimento e o sucesso das pequenas empresas.



## Prompt com aplicação dos princípios de prompt

**IA pequenas empresas**  
Gemini-1.5-Pro

Você é um consultor especializado em tecnologia e pequenos negócios. Explique os benefícios de usar inteligência artificial em pequenas empresas, fornecendo exemplos práticos de como ela pode ajudar a:

- Reduzir custos operacionais.
- Melhorar o atendimento ao cliente.
- Aumentar a eficiência nos processos internos.

Use um tom simples e acessível, ideal para empreendedores iniciantes


 Gemini-1.5-Pro 

Olá, empreendedor! Sou um consultor especializado em tecnologia para pequenas empresas e estou aqui para te mostrar como a Inteligência Artificial (IA) pode ser uma grande aliada do seu negócio, mesmo que você esteja começando agora. A IA não é só coisa de filme de ficção científica, ela já é uma realidade acessível e pode trazer muitos benefícios, como redução de custos, melhor atendimento ao cliente e processos mais eficientes. Vamos ver alguns exemplos práticos:

**1. Reduzindo custos operacionais:**

Imagine que você tem uma loja online e gasta muito tempo respondendo as mesmas perguntas dos clientes, como "qual o prazo de entrega?" ou "como faço para trocar um produto?". Com um chatbot com IA, você pode automatizar essas respostas. O chatbot funciona como um atendente virtual, disponível 24 horas por dia, 7 dias por semana, respondendo às perguntas mais frequentes e liberando você para se concentrar em outras tarefas mais estratégicas, como desenvolver novos produtos ou planejar campanhas de marketing. Isso reduz o tempo gasto com atendimento e, consequentemente, os custos operacionais.

Outro exemplo: a IA pode te ajudar a prever a demanda de produtos. Imagine que você vende sorvetes. A IA pode analisar dados históricos de vendas, clima e até mesmo eventos locais para prever quantos sorvetes você provavelmente venderá na próxima semana. Com essa informação, você pode comprar a quantidade certa de ingredientes, evitando desperdícios e reduzindo custos com estoque.

**IA pequenas empresas**  
Gemini-1.5-Pro

desperdícios e reduzindo custos com estoque.

**2. Melhorando o atendimento ao cliente:**

Além dos chatbots, a IA pode personalizar a experiência do cliente. Por exemplo, se você tem uma loja de roupas online, a IA pode analisar o histórico de compras e as preferências de um cliente para recomendar produtos que ele provavelmente gostará. Isso torna a experiência de compra mais agradável e aumenta as chances de fidelizar o cliente.

A IA também pode ajudar a identificar clientes insatisfeitos. Analisando comentários em redes sociais e avaliações de produtos, a IA pode detectar reclamações e te alertar para que você possa tomar medidas para solucionar o problema e evitar que o cliente abandone seu negócio.

**3. Aumentando a eficiência nos processos internos:**

A IA pode automatizar tarefas repetitivas em diversos setores do seu negócio. Por exemplo, no setor financeiro, a IA pode automatizar a conciliação bancária, reduzindo o tempo gasto com essa tarefa e minimizando erros humanos. No setor de recursos humanos, a IA pode ajudar na triagem de currículos, selecionando os candidatos mais adequados para uma vaga e agilizando o processo de contratação.

A IA também pode otimizar a logística. Imagine que você tem uma empresa de entregas. A IA pode analisar rotas, condições de tráfego e outros fatores para definir o melhor trajeto para cada entrega, reduzindo o tempo de entrega e os custos com combustível.

Em resumo, a IA pode ser uma ferramenta poderosa para impulsionar o seu pequeno negócio. Existem diversas soluções de IA disponíveis no mercado, muitas delas acessíveis para pequenas empresas. Comece pesquisando as opções que melhor se adequam às suas necessidades e não tenha medo de experimentar! A IA pode te ajudar a economizar tempo e dinheiro, melhorar o atendimento ao cliente e tornar seus processos mais eficientes.

Ao utilizar os princípios de prompting, foi definido a persona de um “consultor especializado”, especificou-se o que desejava, como os benefícios e solicitou exemplos, além de utilizar um tom acessível de forma que a resposta se adequa-se ao público de “empreendedores iniciantes”.

## Exercício 7 - Resumo de Notícia Utilizando Exemplos em Prompts

Usando a API Gemini, crie um notebook que utilize prompts para resumir uma notícia (escolha da página principal de um portal de notícias e copie no notebook). O prompt deve solicitar um resumo dessa notícia usando o princípio de Exemplos para guiar a resposta do LLM. Teste o código e avalie a qualidade do resumo gerado.

### Código utilizado

```
[17]: import google.generativeai as genai

# Configurar a API Key do Gemini
API_KEY = "AIzaSyD46yC7LbBrLu_ytXyOUwqTMSht2gXn5Ac"
genai.configure(api_key=API_KEY)

# Função para gerar o resumo da notícia
def generate_summary(news_article):
    """
    Gera um resumo de uma notícia utilizando o modelo Gemini.

    Args:
        news_article (str): Texto completo da notícia.

    Returns:
        str: Resumo gerado pelo modelo.
    """
    # Prompt com exemplos (few-shot prompting)
    prompt = f"""
    Resuma a <NOTICIA> em até 2 sentenças com os principais pontos da notícia.
    Aqui está um exemplo de resumo:

    <NOTICIA>
    "A empresa X lançou um novo smartphone no mercado."
    ###
    <RESUMO>
    "A empresa X anunciou o lançamento de um novo smartphone, trazendo inovações tecnológicas e maior duração de bateria."

    Agora, faça o resumo da notícia abaixo:
    <NOTICIA>
    "{news_article}"
    ###
    <RESUMO>
    """

    # Executando o modelo com o prompt
    model = genai.GenerativeModel("gemini-1.5-flash")
    response = model.generate_content(prompt)

# Texto da notícia
news_article = """
Mulher é baleada durante confronto com polícia em operação contra tráfico de drogas no MA
Uma mulher foi baleada durante uma troca de tiros com a Polícia Militar do Maranhão (PM-MA), na cidade de Governador Edison Lobão, a 659 km de São Luís.
A mulher foi atingida no braço e encaminhada ao hospital para receber atendimento médico. De acordo com a polícia, o confronto teve início durante uma op
Segundo a PM, ao perceber a presença dos policiais, o homem começou a atirar contra a guarnição, o que acabou atingido a vítima. O homem conseguiu fugir
A polícia informou que, depois que recebeu atendimento médico, ela foi encaminhada ao Plantão Central e deverá responder por crimes de tráfico de drogas.
"""

# Gerar o resumo
try:
    summary = generate_summary(news_article)
    print("Resumo gerado pelo LLM:")
    print(summary)
except Exception as e:
    print("Erro ao gerar o resumo:", str(e))
```

## Resultado

```
Resumo gerado pelo LLM:
response:
GenerateContentResponse(
  done=True,
  iterator=None,
  result=protos.GenerateContentResponse({
    "candidates": [
      {
        "content": {
          "parts": [
            {
              "text": "Uma mulher foi baleada no braço durante confronto com a PM-PA em Governador Edison Lobato, PA, durante operação antidrogas. A mulher, que transportava drogas com um homem que fugiu, foi atendida e responder por tráfico de drogas."
            }
          ]
        },
        "role": "model"
      },
      {
        "finish_reason": "STOP",
        "avg_logprobs": -0.19846004909939235
      }
    ],
    "usage_metadata": {
      "prompt_token_count": 336,
      "candidates_token_count": 54,
      "total_token_count": 390
    }
  })
)
```

## Exercício - 8 Identificação de Entidades em Notícias com LLM

Uma consultoria lhe contratou para automatizar a descoberta de pessoas mencionadas em notícias de jornal. Com base no Exercício 7, escolha 3 notícias e monte uma aplicação com um prompt para o LLM identificar menções a diferentes entidades em cada notícia (como pessoas, órgãos públicos, empresas...). Implemente um notebook para testar o código usando a API do Gemini. Descreva o prompt, seus resultados e avalie se a resposta do modelo atende às expectativas da consultoria.

## Código utilizado

```
[22]: import google.generativeai as genai

# Configurar a API Key do Gemini
API_KEY = "AIzaSyD46yC7LbBrLu_ytXyOUwqTNSht2gXn5Ac"
genai.configure(api_key=API_KEY)

def identify_entities(news_article):
    """
    Identifica entidades mencionadas em um artigo de notícia utilizando a API Gemini.

    Args:
        news_article (str): Texto completo da notícia.

    Returns:
        str: Resposta do modelo com as entidades identificadas.
    """
    # Prompt para identificação de entidades
    prompt = f"""
    Leia o texto da notícia fornecida e identifique todas as entidades mencionadas, incluindo:
    - Pessoas
    - Órgãos públicos
    - Empresas
    - Outras entidades relevantes

    Aqui está um exemplo:
    <NOTICIA>
    "O presidente da Apple, Tim Cook, anunciou hoje uma parceria com o governo dos EUA para construir novos centros de tecnologia."
    ###
    <ENTIDADES IDENTIFICADAS>
    - Pessoas: Tim Cook
    - Órgãos públicos: Governo dos EUA
    - Empresas: Apple

    Agora, processe a seguinte notícia:
    <NOTICIA>
    "{news_article}"
    ###
    <ENTIDADES IDENTIFICADAS>
    """
```

```

# Executar o modelo com o prompt
model = genai.GenerativeModel("gemini-1.5-flash")
response = model.generate_content(prompt)

return response

# Notícias
news_articles = [
    """
    Mulher é baleada durante confronto com polícia em operação contra tráfico de drogas no MA
    O caso aconteceu na noite dessa sexta-feira (22), em Governador Edison Lobão. O homem, que acompanhava a mulher, conseguiu fugir.
    """,
    """
    Repórter Mirante deste sábado destaca a força do esporte na região dos Lençóis Maranhenses
    Na programação, os hexacampeões do mundo, Datinha , Bobô e Adrielle, que ajudaram a colocar Tutóia no mapa do esporte mundial, também vão abrir um pou
    """,
    """
    Operação do MP apura indícios de contratações irregulares na Prefeitura de São Luís Gonzaga do Maranhão
    A operação 'Illusio' cumpre 15 mandados de busca e apreensão em três municípios. Como parte da operação, a Justiça determinou o bloqueio do valor de
    """,
]

# Função para exibir os resultados
def display_results(news_articles, results):
    for i, (article, result) in enumerate(zip(news_articles, results), start=1):
        print(f"--- Notícia {i} ---")
        print(article.strip())
        print("\nEntidades Identificadas:")

        # Obter as partes da resposta do modelo
        content = result["candidates"][0]["content"]["parts"][0]["text"]

        # Formatar o texto para melhorar a legibilidade
        for line in content.split("\n"):
            if line.strip():
                print(f"    {line.strip()}")
        print("-" * 80)

# Exibir os resultados formatados
display_results(news_articles, results)

```

```

--- Notícia 1 ---
Mulher é baleada durante confronto com polícia em operação contra tráfico de drogas no MA
    O caso aconteceu na noite dessa sexta-feira (22), em Governador Edison Lobão. O homem, que acompanhava a mulher, conseguiu fugir.

Entidades Identificadas:
- Órgãos públicos: Polícia (do MA)
- Localidades: Governador Edison Lobão, MA (Maranhão)
- Outras entidades relevantes: Tráfico de drogas
-----

--- Notícia 2 ---
Repórter Mirante deste sábado destaca a força do esporte na região dos Lençóis Maranhenses
    Na programação, os hexacampeões do mundo, Datinha , Bobô e Adrielle, que ajudaram a colocar Tutóia no mapa do esporte mundial, também vão abrir um pou
    co de suas intimidades e contar sobre a vida e os trabalhos.

Entidades Identificadas:
- Pessoas: Datinha, Bobô, Adrielle
- Empresas/Organizações: Repórter Mirante
- Lugares: Lençóis Maranhenses, Tutóia
-----

--- Notícia 3 ---
Operação do MP apura indícios de contratações irregulares na Prefeitura de São Luís Gonzaga do Maranhão
    A operação 'Illusio' cumpre 15 mandados de busca e apreensão em três municípios. Como parte da operação, a Justiça determinou o bloqueio do valor de
    R$ 1.491.252,45 nas contas dos investigados.

Entidades Identificadas:
- Órgãos públicos: Ministério Público (MP), Prefeitura de São Luís Gonzaga do Maranhão, Justiça
- Municípios: São Luís Gonzaga do Maranhão (e mais dois não nomeados)
-----

```

## Exercício 9 - Cálculo de Tokens em Texto Longo com API Gemini

Implemente um notebook que use a API Gemini para calcular a quantidade de tokens necessários para processar um texto de 5.000 palavras. Baseie-se no modelo de tokenização utilizado por Gemini e explique como a quantidade de tokens influencia o custo e o desempenho da interação com LLMs em textos longos.

Não consegui realizar o código direto da API do Gemini, desta forma foi utilizado o seguinte código:

```
[22]: import tiktoken

# Função para estimar o número de tokens
def estimar_tokens(texto):
    """
    Estima o número de tokens de um texto usando o encoding 'cl100k_base'.

    Args:
        texto (str): Texto de entrada.

    Returns:
        list: Lista de tokens gerados.
    """
    # Inicializar o encoder com o esquema de codificação
    encoder = tiktoken.get_encoding("cl100k_base")
    # Tokenizar o texto
    tokens = encoder.encode(texto)
    return tokens

# Gerar um texto de exemplo com 5.000 palavras
texto_5000_palavras = " ".join(["palavra"] * 5000)

# Estimar tokens
tokens = estimar_tokens(texto_5000_palavras)

# Exibir resultados
print(f"Número total de tokens: {len(tokens)}")
print(f"Primeiros 20 tokens: {tokens[:20]}")



Número total de tokens: 5001
Primeiros 20 tokens: [19866, 63967, 95747, 95747, 95747, 95747, 95747, 95747, 95747, 95747, 95747, 95747, 95747, 95747, 95747, 95747, 95747, 95747, 95747, 95747]
```

A quantidade de tokens impacta diretamente o custo e o desempenho em APIs de modelos de linguagem. Quanto maior o número de tokens processados, maior o custo, especialmente em textos extensos. Além disso, modelos possuem limites de contexto, textos que excedem esse limite podem ser truncados ou rejeitados. Para otimizar o uso, dividir textos longos em blocos menores ajuda a evitar problemas e melhora a eficiência no processamento.



## Exercício 10 - Otimização de Respostas com Role Prompting

No Poe.com, aplique a técnica de Role Prompting para otimizar as respostas do LLM Claude3.5. O cenário é o seguinte: você está desenvolvendo um assistente virtual para uma empresa de consultoria jurídica. Crie um prompt onde o modelo deve assumir o papel de um advogado especializado em direito contábil ao responder perguntas sobre Imposto de Renda de Pessoa Física. Avalie a resposta do modelo para uma mesma pergunta sobre IRPF num prompt com e sem Role Prompting.

Resposta sem Role Prompting

**Deduções IRPF**  
 Claude-3.5-Haiku



O que posso declarar como deduções no Imposto de Renda de Pessoa Física?

 Claude-3.5-Haiku 

No Brasil, a declaração de Imposto de Renda de Pessoa Física (IRPF) permite várias deduções. Aqui estão as principais:

Deduções Legais:

- Dependentes
  - Cônjuge
  - Filhos
  - Pais
  - Outros dependentes legais
  - Valor de dedução por dependente: R\$ 2.275,08 por ano
- Gastos com Educação
  - Mensalidades escolares
  - Cursos regulares
  - Limite de dedução: R\$ 3.561,50 por dependente
- Despesas Médicas
  - Consultas
  - Exames
  - Tratamentos
  - Cirurgias
  - Medicamentos

**Deduções IRPF**  
 Claude-3.5-Haiku

- Exames
- Tratamentos
- Cirurgias
- Medicamentos
- Sem limite definido, mas devem ser comprovadas

4. Previdência Privada

- Contribuições para PGBL
- Limite de 12% da renda anual tributável

5. Pensão Alimentícia

- Comprovada judicialmente
- Integralmente dedutível


Outras Deduções:

- Contribuições ao INSS
- Despesas com instrução
- Doações a fundos controlados



Importante:

- Mantenha sempre os comprovantes
- Verifique os limites atualizados anualmente
- Consulte um contador para orientações específicas

Resposta com Role Prompting

**Deduções IRPF**  
Claude-3.5-Haiku

Você é um advogado especializado em direito contábil e consultoria tributária. Sua função é oferecer respostas claras e completas sobre o Imposto de Renda de Pessoa Física (IRPF). Por favor, responda a esta pergunta como se estivesse orientando um cliente da sua consultoria jurídica:  
  
O que posso declarar como deduções no Imposto de Renda de Pessoa Física?

 Claude-3.5-Haiku 


Como especialista em consultoria tributária, posso orientá-lo sobre as principais deduções permitidas na declaração do Imposto de Renda Pessoa Física (IRPF) para o ano-calendário de 2024:

1. Despesas com Educação

- Gastos com instrução do contribuinte e dependentes
- Limites: até R\$ 3.561,50 por dependente
- Incluem: mensalidades escolares, material didático, cursos

2. Despesas Médicas

- Despesas comprovadas com:
  - Consultas
  - Exames
  - Tratamentos médicos
  - Internações
  - Planos de saúde
- Sem limite específico de dedução

**Deduções IRPF**  
Claude-3.5-Haiku

- Planos de saúde
- Sem limite específico de dedução

3. Previdência Privada

- Contribuições para PGDL ou VGBL
- Limite de dedução: 12% da renda tributável anual

4. Dependentes

- Podem ser deduzidos:
  - Filho(a) até 21 anos
  - Cônjuge
  - Pais sem renda
- Dedução de R\$ 2.475,50 por dependente

5. Pensão Alimentícia

- Valor integral comprovado judicialmente
- Dedutível integralmente

Importante: Todos os valores são referentes à declaração de 2024 (ano-base 2023). Mantenha sempre os comprovantes organizados.

Deseja que eu detalhe algum desses pontos?

A resposta com Role Prompting é mais eficaz para um assistente virtual jurídico, pois transmite autoridade, personalização e clareza, o que melhora a experiência do usuário. Já a resposta sem role, embora técnica, parece menos confiável em um cenário profissional.



## Exercício 11 - Estruturação de Prompts com Arquivo YAML

Crie um arquivo YAML que represente a estrutura:

```
{‘roteiro’: {‘atenas’: ‘prompt’, ‘roma’: ‘prompt’}},
```

onde você deve definir dois prompts: um para listar 3 pontos turísticos de Roma, o número anual de visitantes e uma breve descrição de cada um, e outro para fazer o mesmo com Atenas. Em seguida, crie um notebook que leia esse arquivo YAML e execute os prompts usando o Gemini (uma execução por cidade no roteiro). Qual foi o resultado? Os pontos turísticos são relevantes e suas descrições são precisas?

### Código do yaml

```
roteiro.yaml X
C: > Users > wande > OneDrive > INFNET > 7. 6º Semestre > Engenharia de Prompts para Ciência de Dados > ! roteiro.yaml
1 roteiro:
2   llm_name: "gemini-1.5-flash"
3   prompts:
4     atenas: >
5       Liste 3 pontos turísticos imperdíveis de Atenas, o número anual de visitantes para cada um e uma breve descrição de suas características históricas ou culturais.
6     roma: >
7       Liste 3 pontos turísticos imperdíveis de Roma, o número anual de visitantes para cada um e uma breve descrição de suas características históricas ou culturais.
8
```

### Código utilizado

```
[43]: import google.generativeai as genai
import yaml
import json
import pandas as pd

# Configurar a chave de API do Gemini
API_KEY = "AIzaSyD46yC7Lb8rLu_ytXyOUwqTNSht2gXn5Ac"
genai.configure(api_key=API_KEY)

# Carregar a configuração do YAML
with open("roteiro.yaml", "r") as file:
    config = yaml.safe_load(file)

# Inicializar o modelo do Gemini
model = genai.GenerativeModel(config['roteiro']['llm_name'])

# Processar os prompts de Atenas e Roma
for cidade, prompt in config['roteiro']['prompts'].items():
    print(f"--- Gerando resumo para {cidade.capitalize()} ---")
    response = model.generate_content(prompt)

# Exibir o resumo gerado
print(f"Resumo gerado pelo LLM para {cidade.capitalize()}:")
print(response.text)

# Converter a resposta em um DataFrame (assumindo estrutura JSON na resposta)
try:
    df = pd.DataFrame().from_dict(json.loads(response.text), orient='index')
    df.columns = ['qtd_visitas'] # Renomear a coluna

    # Plotar os dados
    df.sort_values('qtd_visitas', ascending=True).plot.barh(title=f"Pontos Turísticos - {cidade.capitalize()}")
    plt.xlabel('Quantidade de Visitas')
    plt.ylabel('Pontos Turísticos')
    plt.show()
except Exception as e:
    print(f"Erro ao processar a resposta para {cidade.capitalize()}: {e}")
```

```
--- Gerando resumo para Atenas ---
Resumo gerado pelo LLM para Atenas:
É difícil fornecer números precisos de visitantes anuais para sítios arqueológicos específicos em Atenas, pois os dados são frequentemente agregados e nem sempre são publicados consistentemente. As cifras variam de ano para ano e dependem de vários fatores, como a época do ano e eventos globais. No entanto, posso listar 3 pontos turísticos imperdíveis em Atenas com uma estimativa da "ordem de grandeza" de visitantes anuais (que deve ser encarada como uma aproximação) e suas características históricas e culturais:

1. **Acrópole:** (Estimados > 8 milhões de visitantes por ano) O complexo da Acrópole é, sem dúvida, o ponto turístico mais popular de Atenas. Dominando o horizonte da cidade, ele contém monumentos icônicos como o Partenon (templo dedicado à deusa Atena), o Erechtheion (um templo com a famosa "Portico das Cariátides"), o Propileu (entrada monumental) e o Templo de Atena Nike. Sua história se estende por séculos, desde a Idade do Bronze até a era romana, refletindo a importância religiosa, política e cultural de Atenas na Antiguidade. A arquitetura, a escultura e a própria localização na colina proporcionam uma experiência inesquecível.

2. **Antiga Ágora:** (Estimados > 2 milhões de visitantes por ano) O centro cívico e comercial da antiga Atenas, a Antiga Ágora era o coração da vida social e política da cidade. Ruínas de templos, basílicas, stoas (pórticos) e outros edifícios públicos ilustram a vida cotidiana e a estrutura política da democracia ateniense. A Ágora testemunhou debates filosóficos, assembleias políticas e transações comerciais, desempenhando um papel crucial na formação da cultura e da identidade atenienses. A visita oferece uma compreensão valiosa da vida na Grécia Antiga.

3. **Museu da Acrópole:** (Estimados > 1 milhão de visitantes por ano) Embora não seja um sítio arqueológico "in situ", o Museu da Acrópole é essencial para compreender a história e a arte da Acrópole. Ele abriga uma vasta coleção de esculturas, arquitetura e artefatos encontrados no complexo da Acrópole, desde a época micênica até o período romano. A apresentação das peças, contextualizando-as na história da Acrópole, enriquece significativamente a experiência do visitante e oferece uma perspectiva mais completa da história do sítio.

**Importante:** Como mencionado, as estimativas de visitantes são aproximações. Números mais precisos requereriam uma pesquisa de dados oficiais de cada instituição, o que varia em disponibilidade e precisão.

Erro ao processar a resposta para Atenas: Expecting value: line 1 column 1 (char 0)

--- Gerando resumo para Roma ---
Resumo gerado pelo LLM para Roma:
É difícil fornecer números precisos de visitantes anuais para sítios arqueológicos e monumentos em Roma, pois os dados são frequentemente compilados por diferentes órgãos e nem sempre são totalmente públicos ou consistentemente atualizados. No entanto, posso oferecer uma estimativa aproximada e uma ordem de magnitude, além das características históricas e culturais:

1. **Coliseu e Fórum Romano:** (Estimados > 7 milhões de visitantes por ano combinados)

* **Características:** O Coliseu, um anfiteatro oval maciço, é um símbolo icônico de Roma e do Império Romano. Foi palco de combates de gladiadores, execuções públicas e espetáculos grandiosos. O Fórum Romano, situado próximo, era o centro cívico e comercial da Roma Antiga, com ruínas de templos, basílicas, casas e senado. Juntos, representam a grandiosidade e a complexidade da sociedade romana, mostrando a arquitetura, a engenharia e a vida social da época.

2. **Museus Vaticanos e Capela Sistina:** (Estimados > 6 milhões de visitantes por ano combinados)

* **Características:** Os Museus Vaticanos abrigam uma vasta coleção de arte e artefatos reunidos ao longo dos séculos pelos papas. A Capela Sistina, dentro dos museus, é mundialmente famosa pelos afrescos de Michelangelo, incluindo o icônico "A Criação de Adão". O complexo representa a rica história e o poder da Igreja Católica, mostrando um panorama da arte renascentista e o patrimônio cultural do Vaticano.

3. **Fontana di Trevi:** (Estimados > 20 milhões de visitantes por ano)

* **Características:** Embora não seja tão monumental em termos históricos quanto os outros dois, a Fontana di Trevi é um símbolo inegável de Roma, uma obra-prima do barroco. Sua beleza e o mito de lançar uma moeda na fonte para garantir o retorno a Roma a tornam um destino obrigatório para turistas. Sua construção espelha o poder e a estética da época, representando a dominação da água e o triunfo do Papa.

**Observação:** As estimativas de visitantes são aproximadas e podem variar de ano para ano dependendo de fatores como eventos globais e estações turísticas. A popularidade desses locais, no entanto, é indiscutível.

Erro ao processar a resposta para Roma: Expecting value: line 1 column 1 (char 0)
```

## Exercício 12 - Sistema de Envio de Emails com Dados Turísticos em JSON

Queremos montar um sistema que envie emails com os pontos turísticos de cidades num roteiro de viagem. Com base no Exercício 11, altere o prompt para estruturar a saída em formato JSON, com descrições e número de visitantes. Crie um notebook que use a biblioteca pandas para converter esse JSON para dataframe e plot um único gráfico de barras horizontais com todos os pontos turísticos, destacando as cores das barras de acordo com a cidade do ponto turístico (Coliseu, em Roma deve ter uma cor diferente que Parthenon, em Atenas). Coloque as legendas no gráfico.

### Ajuste no arquivo roteiro.yaml

```
roteiro.yaml X
> Users > wande > OneDrive > INFNET > 7. 6º Semestre > Engenharia de Prompts para Ciência de Dados > ! roteiro.yaml
1 roteiro:
2   llm_name: "gemini-1.5-flash"
3   prompts:
4     atenas: >
5       Liste 3 pontos turísticos imperdíveis de Atenas. Para cada ponto turístico, forneça o número estimado de visitantes anuais e uma breve descrição histórica ou cultural.
6       Responda no seguinte formato JSON:
7       {
8         "Nome do Ponto Turístico": {"cidade": "Atenas", "qtd_visitas": número_de_visitantes, "descricao": "descrição histórica/cultural"},
9         ...
10      }
11     roma: >
12       Liste 3 pontos turísticos imperdíveis de Roma. Para cada ponto turístico, forneça o número estimado de visitantes anuais e uma breve descrição histórica ou cultural.
13       Responda no seguinte formato JSON:
14       {
15         "Nome do Ponto Turístico": {"cidade": "Roma", "qtd_visitas": número_de_visitantes, "descricao": "descrição histórica/cultural"},
16         ...
17      }
```

## Código utilizado

```
[89]: import google.generativeai as genai
import os
import yaml
import json
import pandas as pd
import matplotlib.pyplot as plt
import re

# Configurar a chave de API do Gemini
API_KEY = "AIzaSyD46yC7LbBrLu_ytXyOUuqTNSht2gXn5Ac"
genai.configure(api_key=API_KEY)

# Carregar a configuração do YAML
with open("roteiro.yaml", "r") as file:
    config = yaml.safe_load(file)

# Inicializar o modelo do Gemini
model = genai.GenerativeModel(config['roteiro']['llm_name'])

# Função para pré-processar e corrigir a resposta do modelo
def corrigir_json(resposta):
    """
    Limpa e tenta corrigir inconsistências no JSON retornado pelo modelo.
    """
    try:
        # Captura apenas o bloco JSON entre as chaves mais externas
        resposta_corrigida = re.search(r"\{.*\}", resposta, re.DOTALL).group(0)
        # Remover aspas e texto não numérico de valores numéricos
        resposta_corrigida = re.sub(r'"(\d+|\.\d+?)"\s*(.*?)', r'\1', resposta_corrigida)
        return json.loads(resposta_corrigida)
    except Exception as e:
        print(f"Erro ao processar JSON corrigido: {e}")
        return {}

# Função para executar e processar o prompt
def obter_dados_turisticos(prompt):
    try:
        response = model.generate_content(prompt)
        print(f"Resposta bruta:\n{response.text}\n") # Verificar o conteúdo retornado
        return corrigir_json(response.text)
    except Exception as e:
        print(f"Erro ao executar o prompt: {e}")
        return {}

# Lista para acumular dados de ambas as cidades
dados_gerais = []
for cidade, prompt in config['roteiro']['prompts'].items():
    print(f"--- Gerando dados para {cidade.capitalize()} ---")
    dados = obter_dados_turisticos(prompt)
    if dados:
        for ponto, info in dados.items():
            info['nome'] = ponto
            dados_gerais.append(info)
```

```
# Verificar se os dados foram carregados corretamente
if not dados_gerais:
    print("Nenhum dado foi retornado. Verifique os prompts e as respostas do modelo.")
else:
    # Converter os dados em DataFrame
    df = pd.DataFrame(dados_gerais)

    # Corrigir a coluna de visitantes para números inteiros
    df['qtd_visitantes'] = pd.to_numeric(df['qtd_visitantes'], errors='coerce')
    df = df.dropna(subset=['qtd_visitantes'])

    # Exibir o DataFrame
    print("Dados Gerados:")
    print(df)

    # Configurar o gráfico
    cores = {"Atenas": "blue", "Roma": "red"}
    df['cor'] = df['cidade'].map(cores) # Mapear cores de acordo com a cidade

    # Plotar o gráfico de barras horizontais
    plt.figure(figsize=(10, 6))
    plt.barh(df['nome'], df['qtd_visitantes'], color=df['cor'])
    plt.xlabel("Quantidade de Visitantes (Milhões)")
    plt.ylabel("Pontos Turísticos")
    plt.title("Pontos Turísticos por Cidade")
    plt.legend(handles=[plt.Line2D([0], [0], color=cores['Atenas'], lw=4, label='Atenas'),
                        plt.Line2D([0], [0], color=cores['Roma'], lw=4, label='Roma')],
               title="Cidades")
    plt.show()
```

## Resultado do código

```
--- Gerando dados para Atenas ---
Resposta bruta:
```json
{
  "Acrópole de Atenas": {
    "cidade": "Atenas",
    "qtd_visitantes": 8000000,
    "descricao": "Símbolo da cidade e da civilização grega clássica, a Acrópole é um complexo de edifícios antigos, incluindo o Partenon, o Erechtheion, os Propileus e o Templo de Atena Niké. Construída principalmente no século V a.C., durante o auge do Império ateniense sob Péricles, a Acrópole abrigava templos dedicados às deusas Atena e Niké, bem como importantes estruturas administrativas e militares. Representa o ápice da arquitetura grega clássica, com a sua beleza e imponência refletindo a sofisticação política, artística e filosófica da Atenas antiga."
  },
  "Antigo Ágora": {
    "cidade": "Atenas",
    "qtd_visitantes": 2000000,
    "descricao": "O coração da vida política, social e comercial da antiga Atenas. Situado em uma área baixa ao sopé da Acrópole, o Ágora era o centro da cidade, onde os cidadãos se reuniam para discutir política, realizar negócios e participar de eventos públicos. Ruínas de templos, basílicas, estóas e edifícios administrativos podem ser vistos, testemunhando a vida pública da democracia ateniense clássica, que foi o modelo de muitas instituições políticas modernas."
  },
  "Museu da Acrópole": {
    "cidade": "Atenas",
    "qtd_visitantes": 1500000,
    "descricao": "Localizado no sopé da Acrópole, o museu abriga uma vasta coleção de artefatos encontrados no complexo da Acrópole. Apresenta esculturas, cerâmicas, arquitetura e outros objetos que datam da Idade do Bronze até o período romano. O museu fornece um contexto histórico e artístico para os monumentos da Acrópole, ajudando os visitantes a entender melhor a história e a cultura da antiga Atenas. A sua arquitetura moderna, integrando-se com o cenário histórico, também é uma atração."
  }
}
...

**Nota:** As estimativas de número de visitantes são aproximadas e podem variar de ano para ano.
```

```
--- Gerando dados para Roma ---
Resposta bruta:
```json
{
  "Coliseu": {
    "cidade": "Roma",
    "qtd_visitantes": 7000000,
    "descricao": "Anfiteatro oval construído no século I d.C. durante o reinado do imperador Vespasiano. Era usado para gladiatorial combats, execuções públicas e espetáculos grandiosos. Sua arquitetura imponente e a história de violência e entretenimento que abrigou o tornam um símbolo icônico de Roma e do Império Romano."
  },
  "Museus Vaticanos e Capela Sistina": {
    "cidade": "Roma",
    "qtd_visitantes": 6000000,
    "descricao": "Um complexo de museus dentro da Cidade do Vaticano, abrigando uma vasta coleção de arte e artefatos de séculos de história papal. A Capela Sistina, famosa pelas pinturas de Michelangelo na abóbada e no altar, é a atração mais célebre dos museus. Representam um importante centro cultural e religioso do mundo, refletindo a riqueza e o poder da Igreja Católica."
  },
  "Fórum Romano e Palatino": {
    "cidade": "Roma",
    "qtd_visitantes": 4000000,
    "descricao": "Restos arqueológicos do coração da antiga Roma. O Fórum Romano era o centro cívico e social da cidade, onde se encontravam templos, basílicas, mercados e edifícios governamentais. O Palatino, uma das sete colinas de Roma, foi uma das primeiras áreas habitadas e abriga os restos de palácios imperiais. Juntos, oferecem uma visão fascinante sobre a vida política, social e religiosa da Roma antiga."
  }
}
...

**Observação:** As quantidades de visitantes são estimativas e podem variar de ano para ano, dependendo de diversos fatores.
```

Dados Gerados:

	cidade	qtd_visitas	descricao \
0	Atenas	8000000	Símbolo da cidade e da civilização grega cláss...
1	Atenas	2000000	O coração da vida política, social e comercial...
2	Atenas	1500000	Localizado no sopé da Acrópole, o museu abriga...
3	Roma	7000000	Anfiteatro oval construído no século I d.C. du...
4	Roma	6000000	Um complexo de museus dentro da Cidade do Vati...
5	Roma	4000000	Restos arqueológicos do coração da antiga Roma...

	nome
0	Acrópole de Atenas
1	Antigo Ágora
2	Museu da Acrópole
3	Coliseu
4	Museus Vaticanos e Capela Sistina
5	Fórum Romano e Palatino

