#### **ANEXO 3**

# TRILHA ENGENHARIA – REGRAS PARA DESENVOLVIMENTO E ENTREGA DA SOLUÇÃO

O presente documento destina-se ao detalhamento das regras para desenvolvimento e entrega da solução para o desafio da Trilha Engenharia do Programa Embaixadores Analytics BB.

### 1. SOLUÇÃO DE DADOS

1.1. Para fins de avaliação na Trilha Engenharia do Programa Embaixadores Analytics BB, cada participante deverá desenvolver e entregar uma solução de dados para o problema de negócios apresentado.



- 1.2.1. O gestor gostaria de planejar as campanhas de alguns produtos específicos considerando as particularidades de cada município. Ele vai precisar fazer um estudo considerando o público de cada linha de produtos e as características de cada município, como por exemplo o censo municipal, as principais atividades econômicas do município, o PIB total e per capita e o PIB por atividade. Ele espera com isso poder direcionar melhor os recursos para as praças de maior probabilidade de contratação, melhorando assim a eficácia da campanha.
- 1.2.2. Os produtos ou linhas alvo da campanha, bem como seus respectivos públicos são os que relacionamos abaixo:
  - 1.2.2.1. BB Cash:
    - 1.2.2.1.1. **Público:** população com 10 a 17 anos completos
  - 1.2.2.2. Mulheres no Topo Consórcio:
    - 1.2.2.2.1. **Público:** mulheres de 18 a 70 anos não completos
  - 1.2.2.3. BrasilPrev Jr:
    - 1.2.2.3.1. **Público:** população de 0 a 20 anos incompletos nos municípios com maior IDH dentre aqueles que tenham mais de 100 mil habitantes.
  - 1.2.2.4. Custeio agrícola:
    - 1.2.2.4.1. **Público:** municípios com maior participação do Agronegócio no PIB dentre aqueles que tenham mais de 5 mil habitantes















1.3. A solução deverá ser apresentada na forma de um script em Python, para construção de uma ABT (*Analytical Base Table* ou Tabela Base Analítica).

### 2. BASES DE DADOS

2.1. Abaixo relacionamos as bases de dados a serem utilizadas para solução do desafio proposto no item 1:

Arquivo	Nome	Descrição	Ref.	Fonte
idh_novo.csv	IHDM Municípios 2010	Índice de Desenvolvimento Humano Municipal (IDHM) da PNUD - Programa das Nações Unidas para Desenvolvimento. A tabela contém um ranking, ordenação ascendente, dos municípios de melhor pontuação no IDHM do país, incluindo os dados de IDHM Renda, Longevidade e Educação utilizados para o cálculo do índice. Possui dados referentes ao Censo 2010.	2010	https://www.und p.org/pt/brazil/id hm-municípios- 2010
item_artt_uor.c sv	Tabela de Itens de Arquitetura UOR	Representa os Itens de Arquitetura que existem em uma Instituição Organizacional para orientação da estrutura das Unidades Organizacionais (oriundo do sistema IOR).	2023	DB2UOR
municipios_bb. csv	Tabela de Municípios do BB	Tabela com todos os códigos internos das cidades em que o Banco do Brasil está presente e seus respectivos códigos no IBGE	2023	Adaptado da IN-377 (Municípios Limítrofes)













municipios_ibg e.csv	Tabela de Órgãos e Municípios	A Tabela de Órgãos e Municípios (TOM) contém a relação dos municípios brasileiros com a codificação utilizada pelos diversos sistemas da Receita Federal do Brasil (RFB), apresentando a estrutura organizacional e jurisdicional no âmbito da RFB e domínios de endereçamento.	2023	https://dados.gov .br/dados/conjun tos-dados/tabela- de-rgos-e- municpios
tabela1378csv_ parte1.gz e tabela1378csv_ parte2.gz	Tabela 1378	Populacao residente, por situacao do domicilio, sexo e idade, segundo a condicao no domicilio e compartilhamento da responsabilidade pelo domicilio (Censo Demográfico).	2010	https://sidra.ibge. gov.br/tabela/13 78
tabela5938csv. gzip	Tabela 5938	Produto interno bruto (PIB) a preços correntes, impostos, líquidos de subsídios, sobre produtos a preços correntes e valor adicionado bruto a preços correntes total e por atividade econômica, e respectivas participações.	2020	https://sidra.ibge. gov.br/tabela/59 38
uor.zip	UOR - Unidade Organizacional	Representa as Unidades Organizacionais (UOR) do conglomerado Banco do Brasil (no Brasil e Exterior) ou UOR externas que o BB tenha algum tipo de relacionamento ou interesse em identificá-las. Os dados foram anonimizados para utilização para fins didáticos.	2023	DB2UOR











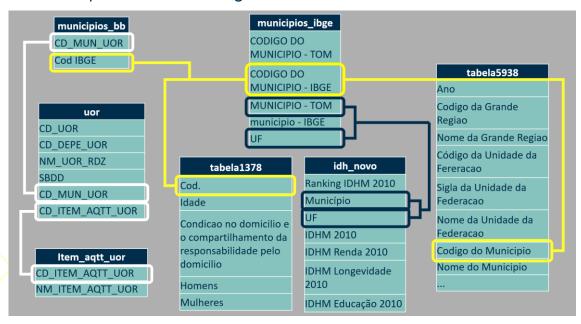


- 2.2. Os arquivos que compõem esse desafio são:
  - 2.2.1. Notebook Modelo (.ipynb)
  - 2.2.2. Bases de dados
  - 2.2.3. Resumo das bases
  - 2.2.4. Documentação das tabelas (metadados)
  - 2.2.5. Glossário de Termos
- 2.3.Clique <u>aqui</u> para efetuar o download dos arquivos necessários para o desenvolvimento do desafio.



#### 3. RELACIONAMENTO ENTRE AS TABELAS

3.1. O relacionamento entre as tabelas disponibilizadas no item 2 está esquematizado na imagem abaixo:



#### 4. REGRAS GERAIS

- 4.1. A participação nos eventos *online* não é obrigatória, porém recomendamos assistir aos vídeos das aulas para melhor se informar sobre tudo o que foi orientado ao longo do programa:
  - 4.1.1. <u>Live 1 SQL Parte 1</u>
  - 4.1.2. Live 2 SQL Parte 2
  - 4.1.3. Live 3 Python para ETL Parte 1
  - 4.1.4. Live 4 Python para ETL Parte 2











- 4.1.5. <u>Live 5 Python para ETL Parte 3</u>
- 4.2. Os materiais de apoio para cada uma das lives também foram disponibilizados nos links acima.
- 4.3. O prazo para envio da solução para verificação automatizada é **até** às 18h do dia 24/07/2023.
- 4.4. Os participantes que não entregarem a solução dentro do prazo estabelecido receberão automaticamente a situação "Não habilitado" para emissão dos certificados.
- 4.5. Somente os funcionários que se inscreveram para a Trilha Engenharia dentro dos prazos propostos no Regulamento terão suas soluções avaliadas para fins de emissão de certificado, conforme as regras detalhadas no presente documento.



- 4.7. Cada participante poderá consultar a situação da sua inscrição e dos seus envios conforme abaixo:
  - 4.7.1. Consultar a situação da inscrição na Trilha e participação nos eventos online: <u>painel de acompanhamento</u>
  - 4.7.2. Consultar a situação dos envios e apontamentos provenientes da verificação automatizada: <u>link de envio das soluções</u>.
- 4.8. A construção dos scripts deve ser individual, bem como as entregas.
  - 4.8.1. **Atenção:** <u>não compartilhe os arquivos</u>! Conforme item 9.14 do Regulamento do Programa, o desenvolvimento de soluções por mais de uma pessoa, ainda que entregue separadamente, implica em desclassificação para todos os envolvidos.

### 5. REGRAS PARA DESENVOLVIMENTO DAS SOLUÇÕES

- 5.1. Baixar e renomear arquivo modelo:
  - 5.1.1. Cada participante deverá baixar o arquivo modelo do Jupyter Notebook (.ipynb) disponibilizado no item 2.3 do presente documento e renomeá-lo de modo a conter:
    - 5.1.1.1. A matrícula com F maiúsculo, sete algarismos e sem dígito (FXXXXXXX)
    - 5.1.1.2. A trilha a qual a entrega se refere em letras minúsculas: engenharia













- 5.1.1.3. O nome não deve conter espaços em branco. Substitua-os por "\_".
- 5.1.1.4. O versionamento será gerado automaticamente na subida do arquivo, por meio do *timestamp* do envio.
- 5.1.1.5. Exemplo: F1234567\_engenharia.ipynb

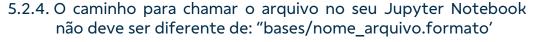
### 5.2. Carregamento das bases de dados:

- 5.2.1. Não carregue no seu Jupyter Notebook nenhum arquivo além das bases de dados disponibilizadas no item 2.2.
- 5.2.2. Não faça nenhuma alteração nos arquivos das bases de dados, seja no conteúdo ou no formato. Não descompacte nem renomeie. Todas as verificações serão feitas a partir dos arquivos originais, mantidos na Plataforma Analítica, portanto qualquer alteração local nos arquivos poderá prejudicar o resultado da avaliação.
- 5.2.3. Antes de carregar as bases de dados no seu Jupyter Notebook, certifique-se de mantê-las numa pasta chamada "bases" (letras minúsculas), obedecendo à seguinte hierarquia:









#### 5.3. Sobre o preenchimento do arquivo modelo

- 5.3.1. Atente-se para todas as orientações constantes no arquivo modelo do Jupyter Notebook. Sua estrutura original e a ordem dos exercícios deve ser mantida sob o risco de invalidar toda a verificação.
- 5.3.2. Não é permitido excluir ou alterar o conteúdo de qualquer das células já preenchidas no notebook, com exceção da seção de **Identificação do Participante**, onde deve ser atribuída a sua











matrícula à variável "chaveF" conforme orientações constantes da própria célula:

- 5.3.2.1. Exemplo: chaveF = "F1234567"
- 5.3.3. É permitido acrescentar células apenas entre as seções do **Exercício 1** ao **Exercício Bônus 3**.
- 5.3.4. Sob nenhuma hipótese deverá ser alterado o conteúdo da seção "**Boletim de Soluções**".
- 5.3.5. Atenha-se aos enunciados dos exercícios. Qualquer novo atributo, transformação ou cálculo além ou aquém do que foi solicitado poderá prejudicar o resultado da avaliação.
- 5.3.6. Recomendamos o desenvolvimento da solução nas plataformas corporativas oficiais para desenvolvimento e execução em Python:



- Acessível em: https://labblite.bb.com.br/
- Permite acesso em qualquer ambiente, inclusive externo ao Banco, por meio de login com senha de portais externos.
- 5.3.6.2. Opção 2: Plataforma Analítica ou AnalyticsLabb
  - Acessível em Plataforma -> Analytics IA -> AnalyticsLabb
  - Acesso somente no ambiente interno, com senha SISBB.
  - Público: funcionários lotados em Unidades Estratégicas ou algumas Unidades de Apoio.
  - Papeis de acesso necessários: GAWMG52, BIGCD001
  - No primeiro acesso, será necessário seguir os passos abaixo:
    - **5.3.6.2..1.** Clique em "Adicionar Projeto"





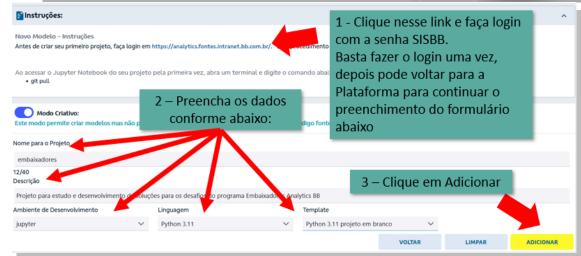


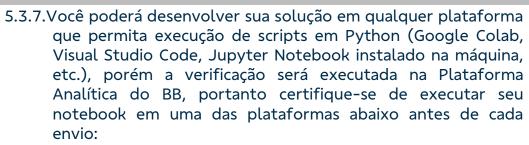














5.3.7.2. Plataforma Analítica (AnalyticsLabb)

5.3.8. Para eventuais instalações de bibliotecas, utilize preferencialmente a instrução %pip (Labblite ou Plataforma Analítica) ou !pip (exclusivo para a Plataforma Analítica).

#### 5.4. Sobre os metadados da ABT

- 5.4.1. Observe atentamente a documentação constante no enunciado de cada exercício, para evitar apontamentos nas verificações.
- 5.4.2. A ordem dos atributos deverá ser obedecida na construção dos DataFrames.
- 5.4.3. A nomenclatura de todos os campos da Tabela Base Analítica (ABT) deverá seguir o padrão de nomenclatura do Banco, conforme consulta disponível no Glossário disponibilizado junto ao Notebook Modelo.
  - 5.4.3.1. O nome físico não deve conter espaços, devendo as abreviações ser separadas por " ".













- 5.4.3.2. Todas as palavras do Nome Lógico deverão ser abreviadas, com exceção apenas das preposições (ex.: de, com, para, até, etc.).
- 5.4.3.3. Ao consultar o Glossário, é importante atentar para a existência de palavras compostas, que devem ser abreviadas juntas.
- 5.4.4. O fato de o atributo aceitar nulos não significa, necessariamente, que ele deve apresentar valores nulos.
- 5.4.5. Eventuais cálculos, formatações, transformações ou arredondamentos dos dados serão detalhados na descrição de cada atributo.

### 6. ENVIO DAS SOLUÇÕES

- 6.1. As soluções deverão ser enviadas no formato .ipynb, nos moldes do arquivo modelo disponibilizado no item 2.3 do presente documento.
- 6.2. Deverá ser enviado **somente o Jupyter Notebook** (arquivo no formato .ipynb).
  - 6.2.1. Não é necessário enviar as bases utilizadas, elas já constam no servidor onde será executada a verificação.
  - 6.2.2. Não é necessário enviar os eventuais arquivos .pkl ou .csv gerados a partir da execução do notebook. Eles serão gerados novamente no servidor para verificação.
  - 6.2.3. Serão desconsiderados quaisquer outros arquivos enviados, em qualquer outro formato diferente do especificado no item 6.2.
- 6.3. Os envios deverão ser efetuados única e exclusivamente por meio do link: <a href="https://envio-embaixadores.bb.com.br/">https://envio-embaixadores.bb.com.br/</a>
- 6.4. O prazo final para envio da solução será <u>até às 18h do dia 24 de</u> <u>julho de 2023</u>, horário de Brasília, conforme Regulamento.
  - 6.4.1. Serão desconsideradas as soluções enviadas fora do prazo.
- 6.5. Os envios poderão ser parciais, ou seja, não é necessário solucionar todos os desafios antes de enviá-lo para verificação, porém é necessário que tenha finalizado pelo menos um exercício para que seja executada a correção.
- 6.6. Não há limite de quantidade de envios por participante. Cada participante poderá reenviar sua solução tantas vezes quantas















forem necessárias, porém para fins de emissão do certificado será considerado o notebook com:

- 6.6.1. Maior quantidade de exercícios concluídos na íntegra.
- 6.6.2. Melhor avaliação qualitativa, se for o caso.

### 7. VERIFICAÇÃO DAS SOLUÇÕES

- 7.1. A verificação do cumprimento dos requisitos será feita de forma automatizada.
- 7.2. A execução dos notebooks entregues para verificação do cumprimento dos requisitos será feita periodicamente, sendo que o período poderá variar de 3 até 24 horas após o envio do arquivo no link especificado no item 6 do presente anexo.



- 7.3. A ocorrência de erros na execução do notebook prejudica a realização da verificação em sua íntegra. Certifique-se de corrigilos antes do envio do notebook.
  - 7.3.1. Havendo erros na execução, será retornado apenas o erro para correção pelo participante.
- 7.4. Todas as verificações serão realizadas a partir do notebook como um todo, portanto certifique-se de que suas variáveis permaneçam íntegras até o final do notebook, não apenas ao término do exercício.

### 8. CRITÉRIOS DE AVALIAÇÃO

- 8.1. A avaliação será dividida em dois quesitos:
  - 8.1.1. Avaliação Técnica
  - 8.1.2. Avaliação Qualitativa
- 8.2. Avaliação Técnica:
  - 8.2.1. Os critérios de avaliação técnica dos scripts serão conforme detalhado abaixo:

Exercício	Descrição	Checagem (S/N)	Principi ante	Desbra vador	Visioná rio	Pontos (0 ou 1)
1	Primeiros atributos da Tabela Base Analítica (ABT)		SIM	OBRIGA TÓRIO	OBRIGA TÓRIO	
2	Atributos do PIB		SIM	OBRIGA TÓRIO	OBRIGA TÓRIO	
3	Atividades com maior valor adicionado		SIM	OBRIGA TÓRIO	OBRIGA TÓRIO	













	bruto				
4	Disponibilidade de prefixos de atendimento em cada município	SIM	OBRIGA TÓRIO	OBRIGA TÓRIO	
5	Quantificar o público das campanhas e população total	SIM	OBRIGA TÓRIO	OBRIGA TÓRIO	
6	Indicadores de população	SIM	OBRIGA TÓRIO	OBRIGA TÓRIO	
Bônus 1	Identificar e tratar municípios emancipados	NÃO	SIM	SIM	
Bônus 2	IDH	NÃO	SIM	SIM	
7	Perguntas	SIM	OBRIGA TÓRIO	OBRIGA TÓRIO	
Bônus 3	Perguntas Extras	NÃO	SIM	SIM	

- 8.2.1.1. Os possíveis resultados da checagem acima são:
  - Sim: exercício validado com sucesso.
  - Não: exercício não concluído, ou seja, apresenta erros na validação.
- 8.2.2. Para cada exercício validado, será atribuído um ponto ao script em questão.
- 8.3. Avaliação Qualitativa:
  - 8.3.1. Os testes qualitativos serão executados exclusivamente para fins de classificação para o Nível Visionário.

### 8.3.2. Pontos extras:

Teste	Descrição	Checagem (S/N)	Percentual de apontamentos	Desbrava dor	Visioná rio	Pontos (0 ou 1)
1	Código reutilizável		n/a	NÃO	OBRIGA TÓRIO	
2	Código comentado		n/a	NÃO	SIM	

- 8.3.2.1. Os possíveis resultados da checagem dos testes 1 a 2 são:
  - Sim: o script foi aprovado no teste.
  - Não: o script não foi aprovado no teste.









- 8.3.2.2. Para cada teste aprovado na checagem, será atribuído um ponto ao participante.
- 8.3.2.3. O Teste 1 consiste na execução do Jupyter Notebook do participante utilizando uma base de dados reduzida e objetiva avaliar se o script está automatizado e flexível o suficiente para se adaptar a eventuais alterações nos dados.
- 8.3.2.4. O Teste 2 consiste na verificação se o script apresenta comentários em todas as células.

### 8.4. Pontos negativos:

Teste	Descrição	Checagem (S/N)	Percentual de Apontam.	Peso	Desbra vador	Visioná rio	Pontos
3	Variáveis não utilizadas		n/a	-4	NÃO	SIM	-4 ou 0
4	Comentários com poucos caracteres	n/a	0 a 1	-2	NÃO	SIM	-2 a 0
5	Células não comentadas	n/a	0 a 1	-1	NÃO	SIM	-1 a 0
6	Blocos de código extensos	n/a	0 a 1	-1	NÃO	SIM	-1 a O
7	Blocos de código duplicados	n/a	0 a 1	-4	NÃO	SIM	-4 a 0

- 8.4.1.1. Os possíveis resultados da checagem do Teste 3 são:
  - Sim: existe uma ou mais variáveis não utilizadas no script
  - Não: não existem variáveis não utilizadas no script
- 8.4.1.2. Havendo uma ou mais variáveis não utilizadas no script, serão deduzidos 4 pontos da pontuação geral do script avaliado.
- 8.4.1.3. Os testes 4 a 7 têm por objetivo levantar o percentual de células com apontamentos nos critérios abaixo relacionados:
  - Teste 4: percentual de células com comentários com menos de 10 caracteres.
  - Teste 5: percentual de células sem comentários.











- Teste 6: percentual de células com blocos de código muito extensos (maior que 30 linhas).
- Teste 7: percentual de células com blocos de código duplicados, indicando pouca reutilização de código.
- 8.4.1.4. O percentual será representado por meio de um índice de 0 a 1, onde 1 corresponde ao número total de células do notebook.
- 8.4.1.5. A pontuação dos testes de 4 a 7 será o resultado da multiplicação do percentual de células com apontamento pelo peso de cada teste.
- 8.4.1.6. O resultado dos testes somente será disponibilizado ao participante após um mínimo de 8 acertos nos exercícios constantes do item 7.1.



- 9.1. Ao final da avaliação automatizada, cada script poderá ter três pontuações:
  - 9.1.1. **Pontuação Técnica:** corresponde ao total de pontos referente a conclusão dos exercícios 1 a 7 e exercícios bônus 1 a 3.
  - 9.1.2. **Pontuação Qualitativa:** corresponde ao total de pontos resultante dos testes 1 a 7.
  - 9.1.3. **Pontuação Total:** corresponde a soma da pontuação Técnica com a Pontuação de Qualitativa.
- 9.2. Para fins de habilitação para o certificado de Nível Principiante ou Desbravador, será considerado o *script* de maior Pontuação Técnica de cada participante.
- 9.3. Para fins de classificação para o certificado de Nível Visionário, será considerado script de maior Pontuação Total de cada participante.

### 10. REGRAS PARA OBTENÇÃO DO CERTIFICADO

- 10.1. Ao final da verificação, cada participante poderá obter um dos resultados abaixo:
  - 10.1.1. **Não Habilitado:** não enviou solução ou não atingiu o aproveitamento mínimo na solução apresentada.
  - 10.1.2. **Habilitado Nível Principiante:** Pontuação Técnica mínima de 5 pontos dos 7 pontos possíveis para o nível em questão, conforme item 7.1.















- 10.1.4. **Classificado Nível Visionário:** 10% dos melhores classificados dentre os participantes Habilitados para o Nível Desbravador.
- 10.2. Para classificação dos Habilitados no Nível Desbravador, será utilizada a Pontuação Total de cada participante.
  - 10.2.1. Caso o participante tenha efetuado mais de um envio, será considerado o script com maior Pontuação Total, dentre os enviados dentro do prazo estabelecido no Regulamento.
- 10.3. Havendo empate na classificação, o desempate será feito pelo menor *timestamp* de envio do notebook avaliado.

### 10.4. Exemplo 1:

Exercícios/Testes	Pontuação
Exercícios 1 a 7	6
Exercícios Bônus 1 a 3	Não habilitado
Testes 1 a 2	Não habilitado
Testes 3 a 7	Não habilitado

Pontuação Técnica: 6

Pontuação Qualitativa: Não se aplica

Pontuação Total: Não se aplica Habilitado Nível Principiante

### 10.5. Exemplo 2:

Exercícios/Testes	Pontuação
Exercícios 1 a 7	7
Exercícios Bônus 1 a 3	2
Testes 1 a 2	1















Testes 3 a 7	-4,5
	,

Pontuação Técnica: 9

Pontuação Qualitativa: 3,5

Pontuação Total: 5,5

Habilitado Nível Desbravador

#### 11. RESSALVA

11.1. A Comissão Organizadora se reserva o direito de efetuar toda e qualquer ação corretiva no presente Anexo para maiores esclarecimentos e ajustes, de modo a evitar ações que venham a interferir no bom funcionamento do Programa.



#### **12. DICAS**

- 12.1. **Não se esqueça de salvar seu projeto**! Sugerimos fazer vários versionamentos dos arquivos, para não correr riscos de perder alterações e para ter um backup, caso algo não saia como o planejado.
- 12.2. Faça envios parciais, tantos quantos forem necessários! Não há limite de quantidade de envios. O ideal é submeter seu notebook a cada exercício concluído, para evitar acúmulo de apontamentos.
- 12.3. Não se assuste com os apontamentos de erros! Observe as orientações e dicas para solucioná-los com maior assertividade.
- 12.4. Muitos exercícios são interdependentes, portanto, na ocorrência de erros, priorize sempre a correção dos exercícios de menos numeração, pois isso pode corrigir alguns dos apontamentos subsequentes.
- 12.5. Priorize a execução dos exercícios de 1 a 7, pois eles fazem parte dos requisitos mínimos para certificação. Após terminá-los, você terá maior tranquilidade para resolver os exercícios Bônus.
- 12.6. Execute o Notebook na íntegra antes de efetuar seus envios para testar a ocorrência de erros, mesmo que ainda não tenha finalizado todos os exercícios ainda!
- 12.7. Utilize o Boletim de Soluções ao final do Notebook Modelo para checar a existência das variáveis de cada exercício. Não se esqueça de informar a sua chave F (chaveF) na seção de Identificação.
- 12.8. Ao fazer junção de tabelas, não se esqueça de conferir as quantidades de linhas e colunas das tabelas originais e tenha













sempre em mente o seu objetivo, antes de escolher o método de junção (interna, esquerda, direita ou total).

- 12.9.Encorajamos a colaboração entre os participantes, porém sempre no sentido de ajudar os demais a chegarem à resposta, nunca respondendo diretamente às perguntas ou compartilhando os arquivos.
- 12.10. Ao copiar seus arquivos, utilize preferencialmente o método copy()
- 12.11.Certifique-se de testar a execução do notebook na íntegra na Plataforma Analítica ou no Labblite antes de submeter seu envio, de modo a testar a execução e verificar a necessidade de instalação de bibliotecas.



12.11.1. **Atenção:** micropip não executa na verificação automatizada! Antes de enviar seu notebook, certifique-se de alterar as instalações de bibliotecas para !pip ou %pip.









