

Processo Seletivo

**Analista de Dados - Pleno**

FIESC - Federação das Indústrias do Estado de Santa Catarina

Florianópolis, Junho de 2024. - **Revisão 0**

|  |  |  |  |
| --- | --- | --- | --- |
| **Processo Seletivo** | **00330/2025 - Analista de Dados - Pleno** | **Etapa** | **Prova Prática** |
| **Entidade** | **IEL** | **Data** | **25/02/2025** |

Dados a serem preenchidos pelo Candidato(a): **(Obrigatório preenchimento)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nome Completo** | **Maurício Gonçalves** | | |
| **E-mail** | [**mauricio@opus10.com.br**](mailto:mauricio@opus10.com.br) | **CPF** | 014.406.009-45 |

Olá, candidato (a)!

Seja bem-vindo à etapa prática do Processo Seletivo - 00330/2025 - Analista de Dados – Pleno.

**Informações Importantes**:

* **Prazo para anexo:** 27/02/2025 às 09h00 – Não serão aceitos envios posteriores ao horário ou enviados por outros meios, como por exemplo e-mail. O não envio conforme orientação acarretará na desclassificação do processo seletivo;
* O arquivo deve estar nomeado com seu nome completo;
* A participação da entrevista que ocorrera em 06/03/2025 está condicionada a aprovação desta etapa;
* A avaliação não poderá ser realizado por celulares ou tablets;
* Não nos responsabilizamos por problemas de instabilidade de internet;
* O envio da resolução reforça o consenso do candidato sobre os critérios de avaliação e participação;
* Link da pasta com o material de apoio:

https://drive.google.com/drive/folders/1mMvxdo9n7TvMQb2hWQh1pJrl0owqemWB?usp=drive\_link

**Sumário**

1. [OBJETIVO 3](#_bookmark0)
2. [ESCOPO DA AVALIAÇÃO 3](#_bookmark1)
   1. [Principais itens avaliados 3](#_bookmark2)
   2. [Pré-requisitos para realização das questões e informações gerais 3](#_bookmark3)
3. [CONTEXTO 3](#_bookmark4)
4. [QUESTÕES 3](#_bookmark5)
   1. [Atendimento de demandas 4](#_bookmark6)
      1. [Demanda 1 – Remuneração média 4](#_bookmark7)
      2. [Demanda 2 – Exportações de Santa Catarina 5](#_bookmark8)
      3. [Demanda 3 – Frota de veículos 6](#_bookmark9)
      4. [Demanda 4 – Estoque de colaboradores 7](#_bookmark10)
   2. [Análise de dados para um projeto 8](#_bookmark11)
   3. [Governança de dados – Validação 8](#_bookmark12)
5. [ENTREGAS 9](#_bookmark13)

# OBJETIVO

O objetivo deste documento é avaliar candidatos(as) à vaga de:

* Analista de Dados - Pleno

# ESCOPO DA AVALIAÇÃO

A avaliação deste processo seletivo está focada em conhecimentos necessários para desempenhar a função de analista de dados pleno, considerando conhecimentos em estatística, programação e pensamento crítico para a resolução dos problemas.

## Principais itens avaliados

* Capacidade do candidato(a) de realizar entregas ad hocs, atendendo o que foi solicitado nas demandas
* Conhecimentos da lógica de programação para resolver os problemas apresentados
* Conhecimento de estatística para realizar as análises
* Capacidade do candidato(a) realizar tratamentos nas bases de dados
* Manipulação de colunas
* Elaboração de visuais
* Construção de storytelling com dados
* Capacidade de trabalhar com strings
* Capacidade de comunicação dos resultados encontrados
* Pensamento crítico e habilidades para detectar e resolver problemas
* Organização e gerenciamento de tempo

## Pré-requisitos para realização das questões e informações gerais

Para realização desta prova, serão fornecidos alguns arquivos de dados para a realização das atividades. Todos os tratamentos de dados devem ser realizados através de linguagem de programação, R ou Python. Tratamentos realizados em planilhas eletrônicas (Microsoft Excel, Google Planilhas) não serão considerados.

Os códigos gerados também serão avaliados. É importante construir os códigos seguindo as boas práticas de programação. Utilize comentários, caso necessário, para deixar claro a lógica utilizada em algumas partes dos códigos.

# CONTEXTO

As questões desta prova foram construídas com a finalidade de simular o dia a dia de um analista de dados do Observatório FIESC. O Observatório FIESC é um centro de inteligência em dados, onde ele possui o objetivo de auxiliar na tomada de decisão dos seus clientes ao fornecer dados e insights. As principais tarefas que são realizadas no Observatório FIESC são a manutenção de produtos de dados padronizados (Faros e Cidade Única), o desenvolvimento de produtos personalizados para resolver um problema de negócio específico dos nossos clientes, o desenvolvimento de projetos de dados e o atendimento de demandas de clientes internos (outras áreas da FIESC) e externos.

Dessa forma, você trabalha na equipe de dados do Observatório FIESC, onde auxilia a equipe no atendimento de demandas, validação de bases de dados e no desenvolvimento de novos projetos.

# QUESTÕES

A seguir, serão apresentados três agrupamentos de questões. Primeiramente, serão apresentadas as questões que simulam o atendimento de demandas do Observatório FIESC. No geral, os nossos clientes encaminham essas demandas para o e-mail do Observatório FIESC, onde é realizada uma triagem dessa demanda e encaminhamento para os responsáveis pelo atendimento. Nesta prova, você será o analista de dados responsável pelo atendimento das demandas.

Após as questões de atendimento de demandas, é apresentado um estudo de caso, onde é simulada a solicitação de uma análise mais aprofundada sobre um problema específico. Nela, será necessário que você desenvolva hipóteses que orientem a resolução do problema apresentado. Além disso, será testado o seu conhecimento em estatística, mais precisamente a análise descritiva e diagnóstica, e a capacidade de comunicação. Os resultados da sua exploração precisam ser inseridos em uma apresentação, respeitando as práticas de storytelling com dados. O público-alvo dessa apresentação são os gerentes da FIESC, que não possuem um entendimento avançado em conceitos relacionados à área de dados.

Por fim, o último agrupamento de questões é referente a uma atividade de validação de base de dados. As validações dos dados que serão utilizados nos produtos do Observatório FIESC são uma das atividades realizadas pela equipe de analista de dados.

## Atendimento de demandas

## Demanda 1 – Remuneração média

Uma empresa gostaria de saber a remuneração média dos trabalhadores que estão atualmente ativos em Joinville. Essa informação é importante, pois eles gostariam de expandir as suas atividades na cidade e precisam ter uma noção do custo com a mão de obra. Eles gostariam de entender a remuneração média praticada por todos os setores. Além disso, eles gostariam de entender a remuneração média praticada por setores industriais. Segundo a equipe de governança, é considerado um setor da indústria os casos em que a coluna cnae inicia com os seguintes valores:

"05","06","07","08","09","10","11","12","13","14","15","16","17","18","19","20","21","22""23","24","25","26","27","28","29",

"30","31","32","33","35","36","37","38","39","41","42""43".

A base bruta que você recebeu para atender a demanda está disponível na pasta “1\_demanda1” com o nome “demanda1.csv”.

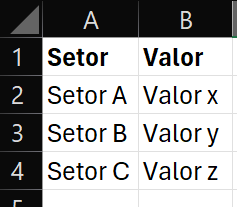
Dicionário das bases:

* + - * nm\_mun: Nome do município
      * cnae: código da classificação nacional de atividades econômicas. Essa coluna precisa ter 7 dígitos;
      * nu\_remuneração: Remuneração média
      * Setor: Identificação setorial de acordo com a classificação do Observatório FIESC

Entrega:

Para essa atividade é necessário que você encaminhe todos os códigos desenvolvidos para realizar o atendimento da demanda. Além disso, é esperado que os resultados sejam salvos em um arquivo no formato “.xlsx”, seguindo os seguintes templates:

Resultado esperado para o arquivo dos setores:



Resultado esperado para os indústria:



## Demanda 2 – Exportações de Santa Catarina

Você recebeu uma demanda, que chegou ao Observatório FIESC, onde o solicitante gostaria de saber o valor das exportações, para o ano de 2024, de carne suína em Santa Catarina. Os resultados desse dado serão utilizados para a construção de uma matéria no jornal.

Você recebeu a base “demanda2.csv” das exportações de Santa Catarina no diretório “2\_demanda2”. Os produtos nessa base estão codificados pelo código do NCM. Diante disso, é necessário que você realize um cruzamento com a base de dimensão dos produtos (base “NCM.csv” do diretório “2\_demanda2”) para coletar a descrição dos produtos.

Dicionário das bases:

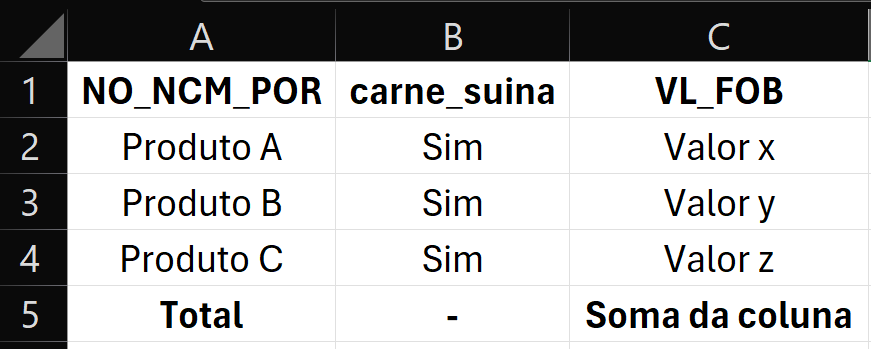
demanda2.csv

* + - * CO\_ANO: Ano das exportações
      * CO\_MES: Mês das exportações
      * CO\_NCM: Código da nomenclatura comum do Mercosu (NCM)l. Esse código precisa ter 8 dígitos e é utilizado para identificar a natureza dos produtos exportados
      * SG\_UF\_NCM: Sigla da Unidade Federativa
      * VL\_FOB: Valor FOB das exportações

NCM.csv

* + - * CO\_NCM: Código da nomenclatura comum do Mercosu (NCM)l. Esse código possui 8 dígitos e é utilizado para identificar a natureza dos produtos exportados
      * NO\_NCM\_POR: Descrição da natureza dos produtos em português Entrega:

Para essa atividade é necessário que você encaminhe todos os códigos desenvolvidos para realizar o atendimento da demanda. Como essa questão exige um julgamento para a classificação dos produtos, detalhe as suas escolhas para encontrar o valor das exportações de carne suína em um documento de texto (e.g. Word). Além disso, é esperado que os resultados sejam salvos em um arquivo no formato “.xlsx”, seguindo o seguinte template:



## Demanda 3 – Frota de veículos

Um telejornal gostaria de ter dados sobre a frota de veículos de Santa Catarina para criar uma matéria. Diante disso, eles encaminharam uma demanda para o Observatório FIESC solicitando um ranking da frota de veículos do Estado. Eles gostariam que fossem encaminhados dois rankings. O primeiro seria um ranking por marca dos veículos e o segundo pela marca e o modelo.

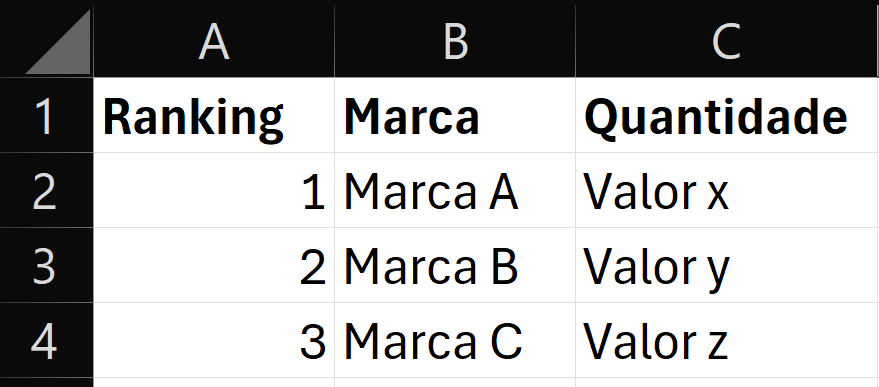
A base bruta da frota de veículos pode ser encontrada no arquivo “demanda3.csv” da pasta “3\_demanda3”

Dicionário da base:

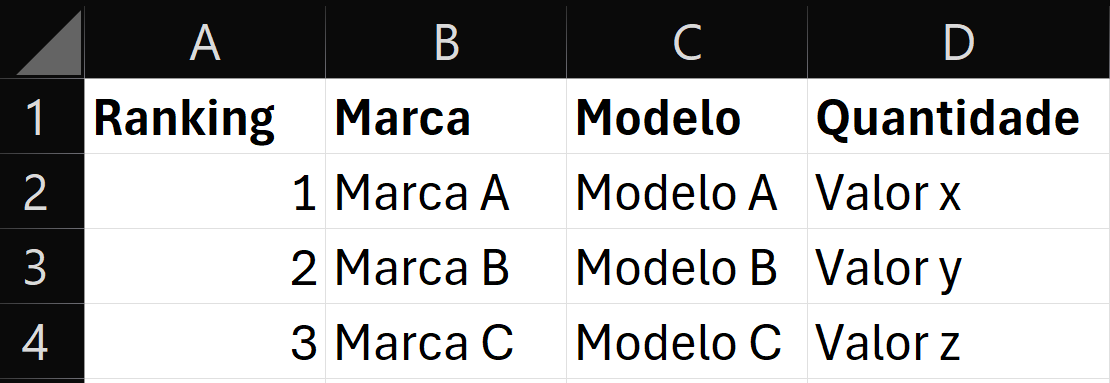
* + - * UF: Estado
      * Marca.Modelo: Informação sobre qual é a marca e o modelo do veículo.
      * Qtd..Veículos: Quantidade de veículos Entrega:

Para essa atividade é necessário que você encaminhe todos os códigos desenvolvidos para realizar o atendimento da demanda.Além disso, é esperado que os resultados sejam salvos em arquivos no formato “.xlsx”, seguindo os seguintes templates:

Ranking por marca:



Ranking por marca e modelo



## Demanda 4 – Estoque de colaboradores

Um gestor público do Estado de Santa Catarina gostaria de saber qual é o estoque de colaboradores de 2024. Ao conversar com a equipe de governança de dados do Observatório FIESC, eles mencionaram que é possível encontrar o estoque de trabalhadores utilizando as informações da RAIS. Entretanto, as últimas informações dessa base de dados são para o ano de 2022. Diante desse cenário, você precisa realizar uma estimativa desse saldo de empregos.

Existem duas opções para realizar a estimativa. A primeira alternativa é construir um modelo preditivo para estimar o estoque dos trabalhadores, de acordo com as informações históricas da Rais. A outra alternativa foi dada pela equipe de Economia do Observatório FIESC. Eles informaram que existe uma outra base de dados, o Novo CAGED, que informa as movimentações de trabalhadores, trazendo a quantidade de trabalhadores que foram admitidos e desligados ao longo dos meses. Dessa forma, você pode realizar o cruzamento das informações da RAIS e do Novo CAGED para encontrar o estoque estimado.

Diante disso, você precisa selecionar uma das abordagens acima para realizar a estimativa do estoque de trabalhadores para o ano de 2024. Justifique a escolha da alternativa. Essa justificativa precisa levar em conta questões relacionadas à confiabilidade dos valores que serão encaminhados ao gestor público.

A base bruta da RAIS pode ser encontrada no arquivo “rais.xlsx” e a do Novo CAGED no arquivo “novo\_caged.xlsx”. Ambas são encontradas na pasta “4\_demanda4”.

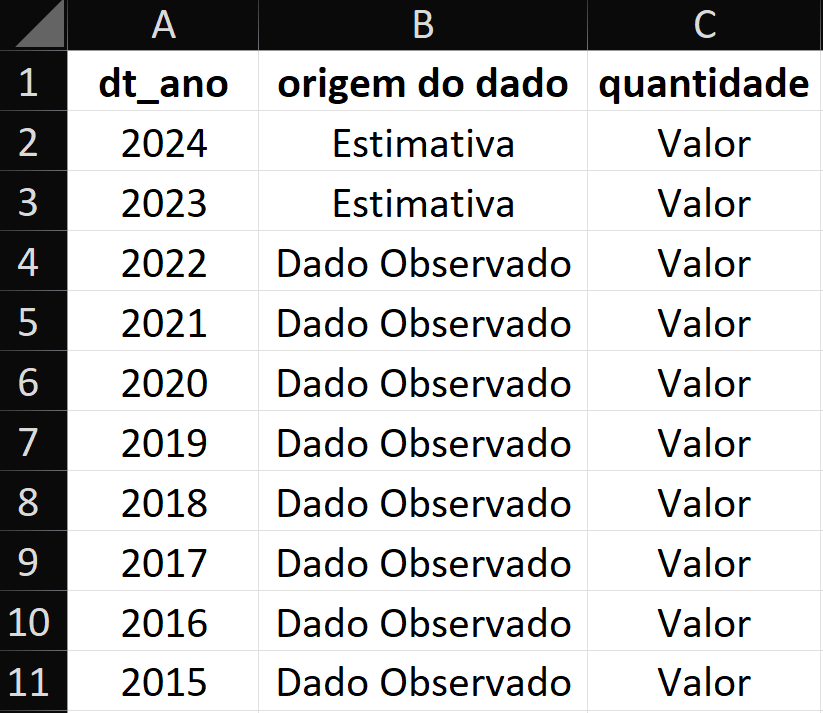
Dicionário das bases: rais

* + - * dt\_ano: Coluna referente ao ano
      * nu\_quantidade: Quantidade de trabalhadores ativos

novo\_caged

* + - * dt\_ano: Coluna referente ao ano
      * nu\_admitidos: Quantidade de trabalhadores admitidos
      * nu\_desligados: Quantidade de trabalhadores desligados Entrega:

Para essa atividade é necessário que você encaminhe todos os códigos desenvolvidos para realizar o atendimento da demanda.Além disso, é esperado que os resultados sejam salvos em um arquivo no formato “.xlsx”, seguindo o seguinte template:



## Análise de dados para um projeto

A área de marketing da FIESC solicitou ao Observatório um estudo para entender melhor o comportamento de vendas de seus produtos para empresas-clientes. Seu líder lhe passou essa demanda já que você possui as habilidades e competências necessárias para concluí-la. Para isso, você receberá uma base de dados contendo o histórico de vendas exportado diretamente do banco de dados. Seu trabalho será aplicar técnicas necessárias para tratar os dados e gerar resultados a partir de métodos estatísticos robustos que possam gerar insights para a equipe de marketing, auxiliando-os na tomada de decisão para o direcionamento de estratégias que possam melhor atender o cliente e impulsionar vendas futuras.

Aqui você será avaliado com base em sua capacidade de manipular, analisar e interpretar uma base de dados de vendas. O teste prático consistirá em uma série de tarefas (descritas a seguir) que envolvem a análise exploratória dos dados e a geração de estatísticas descritivas. Você pode (e deve) sugerir outras tarefas que podem agregar valor ao problema. Os métodos utilizados poderão ser questionados durante a entrevista para avaliar capacidade analítica e pensamento crítico.

Análise exploratória

* Inspecionar a base de dados para entender a estrutura, identificar e corrigir anomalias, formatações, etc;
* Verificar os tipos de dados e transformar caso necessário;
* Analisar a distribuição dos dados e identificar possíveis outliers;

Estatísticas descritivas

* Calcular estatísticas descritivas gerais para as variáveis de interesse;
* Calcular estatísticas descritivas por tempo;

Análises de padrões

* Verificar a presença de padrões em agrupamentos, como por exemplo, por regiões;

Visualização de dados

* Gerar gráficos que representem as análises realizadas e que, principalmente, agreguem valor aos resultados.

Dicionário das bases:

*fat\_vendas*

ID\_cliente: hash de identificação da empresa-cliente data: data da realização da compra (dd-mm-yyyy) produto: nome do produto comprado

valor: valor unitário do produto porte: porte da empresa-cliente

codigo\_municipio: código IBGE do município

*dim\_municipio*

CD7\_IBGE\_MUN: código IBGE do município NM\_MUN: nome do município

SG\_UF: sigla do estado NM\_MICRO: nome da micro-região NM\_MESO: nome da meso-região

Entrega:

A entrega dos resultados gerados deverá ser através de notebook (.ipynb; código e saída) ou arquivo .py/.R com documento de texto (.docx; e.g., Microsoft Word) com prints de código e saída. O candidato(a) deverá fazer uma apresentação em .ppt dos resultados. O arquivo de apresentação deverá ser enviado junto com as demais entregas.

## Governança de dados – Validação

O Observatório FIESC possui uma plataforma chamada Cidade Única. Essa plataforma reúne informações de 5.570 cidades do Brasil, permitindo que os gestores públicos monitorem os principais indicadores econômicos, sociais e demográficos de seus municípios (e de outros de seu interesse).

A validação é uma das etapas do pipeline de dados da plataforma. Essa etapa é importante para garantir a confiabilidade dos dados, garantido que as informações são livres de erros. Essa etapa de validação é realizada pela equipe de analistas de dados do Observatório FIESC.

Você, como analista de dados, recebeu a tarefa de validar o indicador de quantidade de estabelecimentos de saúde. Essa tarefa consiste em detectar e corrigir os problemas encontrados na base de dados.

No diretório “6\_validacao”, você encontrará as pastas com os dados que serão utilizados para alimentar a plataforma. Na pasta de dados, você encontrará três subpastas (bronze, silver e gold). A base de dados bruta (raw), da forma que ela foi coletada na fonte oficial (imutável), é encontrada na pasta bronze. Na pasta silver é encontrada a base de dados com as relações necessárias para inserir na plataforma. Por fim, a pasta gold apresenta as bases de dados tratadas que irão alimentar a plataforma.

Sendo assim, é preciso que você realize a validação dos arquivos que estão na pasta gold. Para essa atividade, você precisa descrever detalhadamente o processo de validação que foi utilizado para a detecção das inconsistências e pontuar as inconsistências na base de dados, caso elas existam. Dentro da pasta “6\_validacao”, existe a pasta “dados\_tabnet”. Dentro dessa pasta existem as informações retiradas da fonte oficial. Utilize esses dados para realizar a validação dos arquivos da pasta “gold”.

Entrega:

Para essa atividade será necessário entregar um documento pontuando o processo utilizado para realizar a validação das bases que estão na gold. Apresente as etapas percorridas para realizar a validação, quais inconsistências foram encontradas e o que foi feito para corrigi-las.

# ENTREGAS

A entrega final deverá ser um arquivo compactado (.zip ou .rar) com o nome do candidato(a) contendo todas as entregas das questões, incluindo a apresentação .ppt do item 4.2.