

Teste Power BI

Descrição:

Você recebeu a demanda de criar um painel em Power BI a partir de um conjunto de dados de **corridas de Fórmula 1**. Uma etapa importante do projeto, antes mesmo da construção do painel, é a de Extração, Transformação e Carga (ETL).

Trata-se de um conjunto de dados público, chamado [Ergast](#). Para garantir que a mesma amostra dos dados será utilizada pelo desenvolvedor e avaliador, foi salva uma amostra de apenas alguns arquivos do conjunto de dados, no formato CSV, em um repositório específico no Github. Portanto, você **não** deverá baixar os arquivos CSV do site do Ergast, mas sim a partir dos links para os arquivos no repositório no Github, listados mais abaixo.

Objetivo:

Criar um projeto para realizar o ETL dos dados, deixando-os prontos para serem lidos pelo Power BI. Em seguida, criar painel para visualização dos dados.

Detalhes:

O avaliador precisará executar seu código-fonte do ETL para testar, bem como abrir o arquivo .pbix para visualizar os dados no painel. Por isso você deverá incluir documentação com:

- Instruções para execução do ETL;
- Caso você tenha habilidades de programação em Python para ETL, recomendamos a utilização desta linguagem. Se for este o caso, você deverá incluir na documentação:
 - Versão do Python utilizada, bem como das bibliotecas utilizadas (de preferência incluir arquivo “requirements.txt” ou equivalente;
 - Serão avaliados alguns aspectos sobre organização e qualidade do código.
- Caso não saiba utilizar Python, poderá utilizar como segunda alternativa preferencial uma outra linguagem de programação, ou então uma ferramenta própria para realização de integração/ETL de dados (gratuita);
- Como terceira opção, baixe os arquivos manualmente, carregue e trate os dados diretamente no Power Query;
- Ao se executar o ETL mais de uma vez, os dados resultantes devem ser os mesmos, sem erros ou duplicidade de dados.

Diferencial (não obrigatório), mas que agrega na avaliação:

- Se souber utilizar Git/Github Você poderá enviar seu código-fonte em um repositório do seu Github, que poderá ser público ou privado (em caso de repositório privado, é necessário fornecer acesso ao usuário do Github “CaioSobreira”). Incluir código fonte do ETL e arquivo .pbix.

*Caso não utilize Github, pode enviar os arquivos por e-mail ou Google Drive/One Drive.

LISTA DE URLS DOS CSVs:

https://github.com/CaioSobreira/dti_arquivos/raw/main/constructors.csv
https://github.com/CaioSobreira/dti_arquivos/raw/main/drivers.csv
https://github.com/CaioSobreira/dti_arquivos/raw/main/races.csv
https://github.com/CaioSobreira/dti_arquivos/raw/main/results.csv

ETAPA 1 - EXTRAÇÃO:

Objetivo: Baixar os arquivos csv da origem em uma pasta denominada “extracao”.

- 1- Seu script deve criar a pasta automaticamente, caso não exista;
- 2- Seu script deve baixar automaticamente os arquivos csv na pasta;

ETAPA 2- TRANSFORMAÇÃO E CARGA:

Objetivo: Preparar os dados para leitura pelo Power BI.

- 1- Seu ETL deve criar uma pasta chamada “saida” automaticamente, caso ela ainda não exista;
- 2- Os arquivos de saída correspondentes a cada CSV de entrada devem ser salvos na pasta saída, mas no formato “**XLSX**” (**Excel**), em vez de CSV;
- 3- Os nomes das colunas da base original estão em inglês, abaixo segue a tradução das mesmas para português. Os seus dados de saída devem estar com os nomes de colunas em português.
OBS: Os dados em si não precisam ser traduzidos, apenas os nomes das colunas.
- 4- Os arquivos CSV originais possuem uma série de colunas desnecessárias para o objetivo deste projeto, as quais devem ser eliminadas nesta etapa. Outras transformações também podem ser necessárias.

Abaixo segue a relação dos arquivos a serem tratados no ETL, com as respectivas listas **apenas das colunas que devem ser mantidas (as demais devem ser descartadas)**, bem como os tipos de dados de cada coluna e demais transformações necessárias:

Arquivo CONSTRUCTORS (montadoras):

Coluna	Tipo de dado	Nome traduzido
constructorId	TEXTO	montadora_id
name	TEXTO	nome
nationality	TEXTO	nacionalidade

***eliminar demais colunas**

Arquivo DRIVERS (pilotos):

***OBS: Apenas nos dados do piloto deve ser realizado um tratamento simples, de juntar os textos das colunas forename (primeiro nome) e surname (sobrenome), separando por um espaço, e criar coluna nomeCompleto**

Coluna	Tipo de dado	Nome traduzido
driverId	TEXTO	piloto_id
nome_completo *coluna criada. Ver observação	TEXTO	nome_completo
nationality	TEXTO	nacionalidade

***Eliminar demais colunas (incluindo forename e surname, após a geração da nova coluna)**

Arquivo RACES (corridas):

Coluna	Tipo de dado	Nome traduzido
raceId	TEXTO	corrida_id
year	NÚMERO INTEIRO	ano
name	TEXTO	nome
date	DATA	corrida_data

***Eliminar demais colunas**

Arquivo RESULTS (resultados):

Coluna	Tipo de dado	Nome traduzido
resultId	NÚMERO INTEIRO	resultado_id
raceId	NÚMERO INTEIRO	corrida_id
driverId	NÚMERO INTEIRO	piloto_id
constructorId	NÚMERO INTEIRO	montadora_id
positionOrder	NÚMERO INTEIRO	posicao_ordem
points	NÚMERO INTEIRO	pontos
fastestLapTime	TEMPO	volta_mais_rapida_tempo

***Eliminar demais colunas**

ETAPA 3 – POWER BI:

Diferencial:

Uma habilidade muito importante para quem trabalha com business intelligence é saber utilizar **modelagem dimensional**. Caso você saiba trabalhar com modelagem dimensional, você deverá criar um modelo do tipo **estrela**, com as seguintes dimensões:

- dim_montadora;
- dim_piloto;
- dim_tempo.

Deverá criar ao menos uma tabela fato, para atender aos requisitos do painel.

*A modelagem dimensional poderá ser feita diretamente no Power Query.

*Caso não saiba utilizar modelagem dimensional, poderá utilizar diretamente os arquivos XLSX da etapa do ETL, mas pontuará menos.

PAINEL:

Você deverá criar um arquivo PBIX com uma ou duas abas (dashboards), que permitam responder, pelo menos, às seguintes perguntas:

- Quantidade de pontos por piloto, por ano;
- Quantidade de pontos por montadora, por ano;
- Quantidade de pontos por nacionalidade dos pilotos, por ano;
- Quantidade de pontos por nacionalidade das montadoras, por ano;

Você pode agregar outras análises que achar interessante. Serão avaliados a usabilidade/utilidade e visual dos painéis.