

Guia de Construção do Desafio

Ps: este documento foi criado com a intenção de transmitir todos os passos utilizados para a construção do dashboard até a apresentação de análise, de forma a enaltecer os procedimentos utilizados.

1. Entendimento do Caso:

De forma a manter uma construção coerente, mesmo o Exame Nacional do Ensino Médio ser o exame mais famoso do país, foi necessário realizar a leitura dos PDFs e Dicionário de Dados contidos na pasta “Leitura Utilizada” para se obter um maior entendimento sobre quais perguntas e aspectos de negócio seriam mais fortemente levados em consideração.

Após a leitura e chegada ao entendimento de pontos principais que poderiam ser explorados como: possíveis índices de desempenho acadêmico (obtido com a média das provas) e perfil socioeconômico do brasileiro, foi observado junto ao Dicionário de Dados quais perguntas de negócios poderiam ser mais viáveis para o caso e adicionado ao mapa de métrica (encontra-se na pasta Mapa de Métrica)

2. Criação de KPI's

Finalizado o entendimento, foi observado que poderiam ser traçados 3 Mapas de Métricas, cada qual com um objetivo específico, sendo eles **(os mapas criados encontram-se na pasta “Mapas de Métricas”)**:

1. Visão Geral do Enem: resumo dos principais dados como perfil dos inscritos, taxa de abandono e medidas gerais.
2. Dados da Prova: relatório mais detalhado sobre as notas, estados com maiores índices e qual a relação disso com estudantes de escola pública e privada, da mesma forma a média dos cursos e cotas.

Obs: os dados de notas dos cursos das faculdades e médias de cortes foram obtidos fora a parte através de um site web e toda a parte de coleta e tratamento foram realizados manualmente com uma lista de 48 universidades federais resultando em mais de 16.000 linhas de dados (pode ser observado na pasta “Lista de Tratamento → lista de universidade.txt”).

3. Perfil Socioeconômico: criado com o objetivo de saber quais os estados com inscritos que possuem menos acesso a internet, saber a renda desse público e entender se ser da zona rural implica para menos acesso a internet, computador e celular.

3. Importação dos CVS's Para Big Query

Os microdados gerados através do arquivo obtido no site do Enem foram importados para o bigquery como seguem as imagens:

Criando Tabelas MICRODADOS_ENEM_2019, ITENS_PROVA_2019 e FACULDADES_MEDIA_CORTE:

Criar tabela

Origem

Criar tabela de
 Google Cloud Storage

Selecione o arquivo do bucket do GCS ou use um [padrão de URL](#) *

☒ faculdades_media_corte/relacao_total_inscritos.csv PROCURAR

Formato do arquivo
 CSV

☐ Particionamento de dados de origem

Destino

Projeto *
 pelagio-bastich-392202 PROCURAR

Conjunto de dados *
 dados_enem2019

Tabela *
 faculdades_media_corte

Letras, números, caracteres especiais, caracteres de escape ou espaços são permitidos.

Create table

Source

Create table from
 Upload

Select file *
 ITENS_PROVA_2019.csv X BROWSE 0

File format
 CSV

Definindo Schemas:

Field name *	Type *	Mode	Description	Max length
SG_AREA	STRING	NULLAB...	Área de Conhecimento do Item	2
CO_ITEM	INTEGER	NULLAB...	Código do Item	
TX_GABARITO	STRING	NULLAB...	Gabarito do Item	1
CO_HABILIDADE	INTEGER	NULLAB...	Habilidade do Item	
IN_ITEM_ABAN	INTEGER	NULLAB...	Indicador de Item abandonado	
TX_MOTIVO_ABAN	STRING	NULLAB...	Motivo para o abandono do	40
NU_PARAM_A	INTEGER	NULLAB...	discriminação: é o poder de	
Field name *	Type *	Mode	Description	

Armazenamento no Cloud Storage

The screenshot shows the Google Cloud Platform interface for a bucket named 'faculdades_media_corte'. The bucket is located in 'us (várias regiões nos Estados Unidos)', has a 'Standard' storage class, 'Não público' public access, and 'Nenhum' protection. The 'OBJETOS' tab is selected, showing a list of objects. One object is visible: 'relacao_total_inscritos.csv', which is 257.4 MB, of type 'text/csv', created on 5 de jun. de 2020, with a 'Standard' storage class, last modified on 5 de jun. de 2020, and 'Não público' access.

Nome	Tamanho	Tipo	Criado	Classe de armazenamento	Última modificação	Acesso público
relacao_total_inscritos.csv	257,4 MB	text/csv	5 de jun. de ...	Standard	5 de jun. de 202...	Não público

TABELAS CRIADAS

The screenshot shows the BigQuery interface for a table named 'faculdades_media_cota'. The 'ESQUEMA' tab is selected, displaying the table's schema. The schema consists of five columns: 'Cidade' (STRING, NULLABLE), 'Cursos' (STRING, NULLABLE), 'Disputa' (STRING, NULLABLE), 'Nota' (FLOAT, NULLABLE), and 'Sigla' (STRING, NULLABLE).

Nome do campo	Tipo	Modo
Cidade	STRING	NULLABLE
Cursos	STRING	NULLABLE
Disputa	STRING	NULLABLE
Nota	FLOAT	NULLABLE
Sigla	STRING	NULLABLE

4. Criação de Queries

Todas as queries e subqueries utilizadas para geração dos dados podem ser verificadas na pasta “Scripts”, a query principal em “script_principal.txt”.

5. Criação dos Dashboards

- O dashboard possui 4 páginas de relatório baseadas nos KPI's definidos
- As cores foram baseadas nas cores da prova



6. Análise do Dashboard

A análise dos dados do enem 2019 foram exposto no PDF/Slide “analise_enem_2019” na pasta principal com o objetivo de analisar o desempenho na prova dos inscritos com diferentes realidades