**Documentação para o Teste de Mesha**

Para a análise dos dados do ENEM 2020 primeiro revisei quais dados são os maias relevantes, olhando as perguntas e quais seriam interessantes de analisar.

No início eu carrego a tabua principal de Microdados do ENEM e seu Dicionário de dados, qual tem nossas tabuas de dimensões.

Logo eu procedi a relacionar as tabuas com um modelo estrela, aonde a tabua principal fica no centro com os campos chave que usarei para relacionar com as tábuas de dimensões. É importante dizer que modifiquei os formatos dos campos que tem só números ou em números decimais em número inteiro e decimal respectivamente.

É importante olhar a completitude dos dados para tomar uma decisão com relação ao analise que eu quero fazer, por exemplo o estado de escola tem uma percentagem menor a 1%, então não é adequado o uso dum gráfico de mapa para ver as médias por estado.

Decidido os campos e feitas as relações posso iniciar a realizar as medidas. Primeiro eu calculo os totais de pessoas inscritas com um COUNT sobre o campo NU\_INSCRICAO que es nosso campo chave na tabua principal.

Num\_Inscritos = count(FAC\_ENEM\_2020[NU\_INSCRICAO])

Depois eu calculo a quantidade de ausências, filtrando os inscritos com os campos de TP\_PRESENCA\_CH, TP\_PRESENCA\_CN, TP\_PRESENCA\_LC, TP\_PRESENCA\_MT.

Ausencia = CALCULATE([Num\_Inscritos], FAC\_ENEM\_2020[TP\_PRESENCA\_CH] = 0 || FAC\_ENEM\_2020[TP\_PRESENCA\_CN] = 0 || FAC\_ENEM\_2020[TP\_PRESENCA\_LC] = 0 || FAC\_ENEM\_2020[TP\_PRESENCA\_MT] = 0)

Com essas duas medidas eu já posso calcular a percentagem de ausência, com um DIVIDE entre as duas medidas. Essas medidas são apresentadas como cartões para facilitar a visualização.

% Ausencia = DIVIDE([Ausencia],[Num\_Inscritos],0)

Para o calculo de médias, iniciei calculado as médias das notas de cada categoria com a função AVERAGE. Prontas a medidas, fiz um gráfico de rosca para visualizá-las.

Ciências da Natureza = AVERAGE(FAC\_ENEM\_2020[NU\_NOTA\_CN])

Ciências Humanas = AVERAGE(FAC\_ENEM\_2020[NU\_NOTA\_CH])

Linguagens = AVERAGE(FAC\_ENEM\_2020[NU\_NOTA\_LC])

Matemática = AVERAGE(FAC\_ENEM\_2020[NU\_NOTA\_MT])

Redacao = AVERAGE(FAC\_ENEM\_2020[NU\_NOTA\_REDACAO])

Com essas 5 medidas se pode calcular a média das provas, dividindo a suma das:

Media\_Geral = DIVIDE([Ciências da Natureza]+[Ciências Humanas]+[Linguagens]+[Matemática]+[Redacao],5,0)

Com a média geral de notas já é possível fazer um gráfico dinâmico da dimensão com um parâmetro de campos que mude a dimensão que você queira visualizar, eu usei o nome AuxMedida e os campos de Sexo, Etnia, Escola, Faixa Etária e Estado Civil. Se usa na barra de tarefas o parâmetro e como fórmula fica:

AuxMedida = {

("Sexo", NAMEOF('DIM\_SEXO'[Descrição]), 0),

("Etnia", NAMEOF('DIM\_COR\_RACA'[Descrição]), 1),

("Escola", NAMEOF('DIM\_ESCOLA'[Descrição]), 2),

("Faixa Etaria", NAMEOF('DIM\_FAIXA\_ETARIA'[Descrição]), 3),

("Estado Civil", NAMEOF(DIM\_ESTADO\_CIVIL[Descrição]), 4)

}

No gráfico eu adicionei a quantidade de pessoas e a média. Para mudar o título do gráfico usei uma nova medida chamada Titulo\_Grafico.

Titulo\_Grafico = "Media Geral e Quantidade de Pessoas por " & VALUES(AuxMedida[AuxMedida])

Para determinar qual é a pessoa com maior média é preciso fazer um ranking, eu usei uma tabua com o número da inscrição e o valor da media, para logo filtrar o ranking por menor que 2. A fórmula do ranking é:

Ranking Pessoa = RANKX(ALL(FAC\_ENEM\_2020[NU\_INSCRICAO]), [Media\_Geral])

Em relação aos novos indicadores que tenha que fazer, decidi adicionar 3 gráficos. Primeiro usei o gráfico de linhas para ver a renda familiar mensal e a relação com a média geral. egundo eu queria saber quantas pessoas tenham internet o 2020, devido a que foi o primeiro ano da pandemia e esse ano a maioria teve aulas online, para o qual usei um gráfico de palavras. Finalmente usei o mesmo gráfico de palavras para saber quantas pessoas que não são brasileiras fizeram o ENEM, já que eu sou estrangeiro.