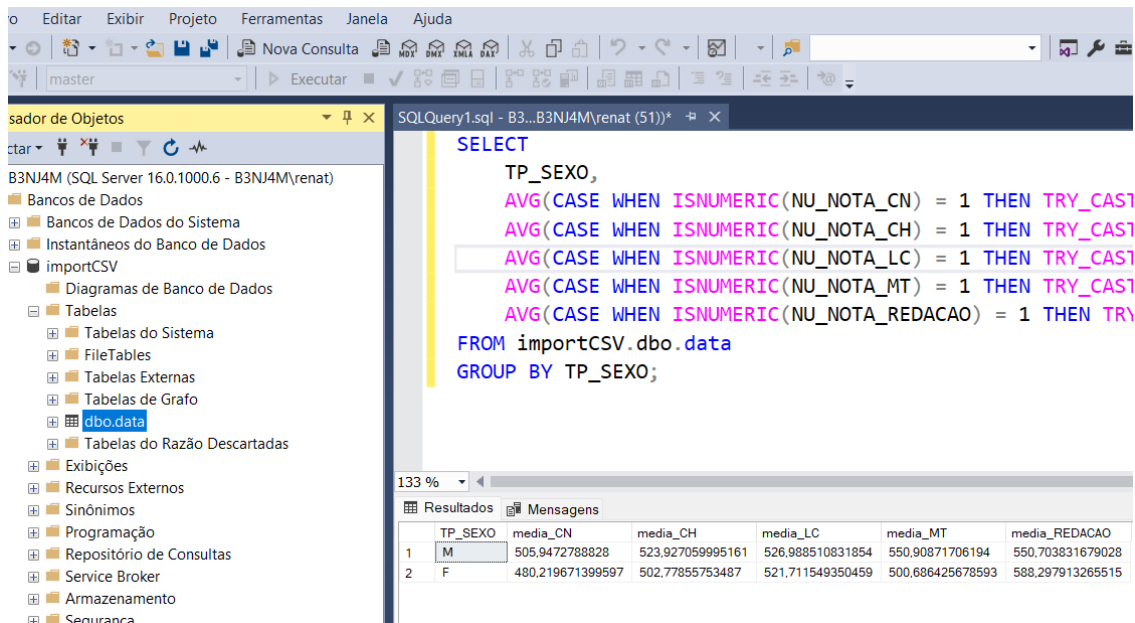


ETL – Áthila Vinicius

Por se tratar de um arquivo muito grande, importei a base CSV para o banco de dados do SQL SERVER



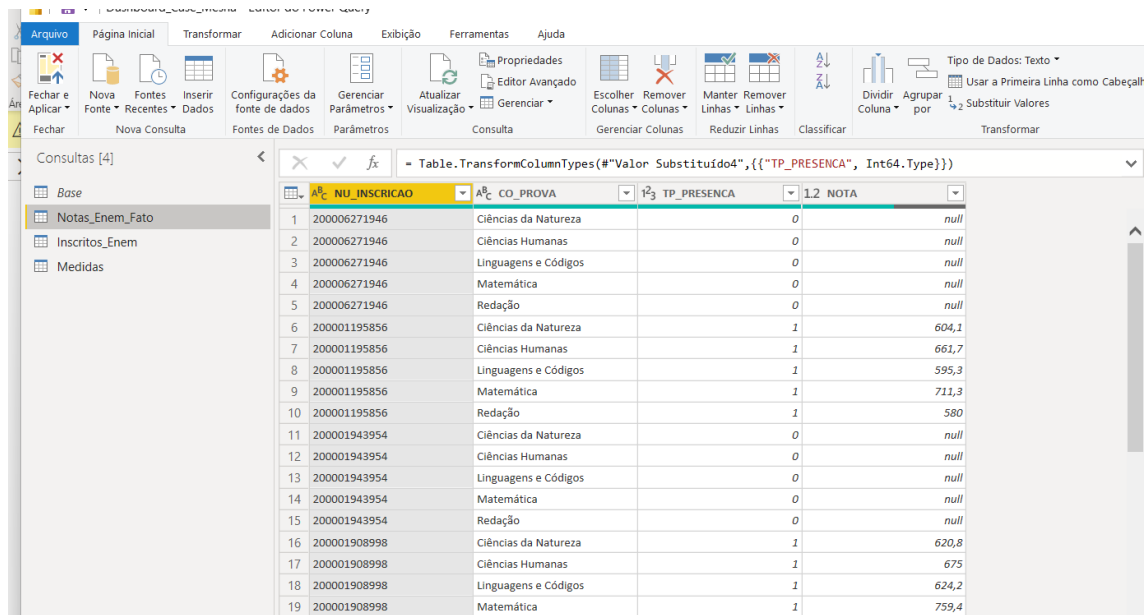
The screenshot shows the SQL Server Enterprise Manager interface. On the left, the 'Servidor de Objetos' (Object Explorer) pane displays the database structure for 'B3NJ4M (SQL Server 16.0.1000.6 - B3NJ4M\renat)'. The 'importCSV' folder is expanded, showing a table named 'dbo.data'. The main window displays a SQL query in the 'SQLQuery1.sql' file:

```
SELECT
    TP_SEXO,
    AVG(CASE WHEN ISNUMERIC(NU_NOTA_CN) = 1 THEN TRY_CAST
    AVG(CASE WHEN ISNUMERIC(NU_NOTA_CH) = 1 THEN TRY_CAST
    AVG(CASE WHEN ISNUMERIC(NU_NOTA_LC) = 1 THEN TRY_CAST
    AVG(CASE WHEN ISNUMERIC(NU_NOTA_MT) = 1 THEN TRY_CAST
    AVG(CASE WHEN ISNUMERIC(NU_NOTA_REDACAO) = 1 THEN TRY
FROM importCSV.dbo.data
GROUP BY TP_SEXO;
```

Below the query, the 'Resultados' (Results) pane shows the output of the query, displaying two rows of data:

| TP_SEXO | media_CN | media_CH | media_LC | media_MT | media_REDACAO |
|---------|------------------|------------------|------------------|------------------|------------------|
| 1 M | 505,9472788828 | 523,927059995161 | 526,988510831854 | 550,90871706194 | 550,703831679028 |
| 2 F | 480,219671399597 | 502,77855753487 | 521,711549350459 | 500,686425678593 | 588,297913265515 |

Ao importar a base para o PowerBI criei duas referencias da tabela Base, uma para registrar os Fato das notas e outra para a Dimensão dos Inscritos



The screenshot shows the Microsoft Power BI Desktop interface. The 'Consultas' (Queries) pane on the left shows a table named 'Base'. The main window displays the data table with the following columns: NU_INSCRICAO, CO_PROVA, TP_PRESENCIA, and NOTA. The data is organized into rows, showing the results of the ETL process.

| NU_INSCRICAO | CO_PROVA | TP_PRESENCIA | NOTA |
|-----------------|----------------------|--------------|-------|
| 1 200006271946 | Ciências da Natureza | 0 | null |
| 2 200006271946 | Ciências Humanas | 0 | null |
| 3 200006271946 | Linguagens e Códigos | 0 | null |
| 4 200006271946 | Matemática | 0 | null |
| 5 200006271946 | Redação | 0 | null |
| 6 200001195856 | Ciências da Natureza | 1 | 604,1 |
| 7 200001195856 | Ciências Humanas | 1 | 661,7 |
| 8 200001195856 | Linguagens e Códigos | 1 | 595,3 |
| 9 200001195856 | Matemática | 1 | 711,3 |
| 10 200001195856 | Redação | 1 | 580 |
| 11 200001943954 | Ciências da Natureza | 0 | null |
| 12 200001943954 | Ciências Humanas | 0 | null |
| 13 200001943954 | Linguagens e Códigos | 0 | null |
| 14 200001943954 | Matemática | 0 | null |
| 15 200001943954 | Redação | 0 | null |
| 16 200001908998 | Ciências da Natureza | 1 | 620,8 |
| 17 200001908998 | Ciências Humanas | 1 | 675 |
| 18 200001908998 | Linguagens e Códigos | 1 | 624,2 |
| 19 200001908998 | Matemática | 1 | 759,4 |

A principal transformação que fiz na tabela nota foi juntar todas provas em uma coluna, mesclar nota com presença e realizar o unpivot.

Registros:

let

```
Fonte = Base,

#"Colunas Mescladas" = Table.CombineColumns(Fonte,{"TP_PRESENCA_CN",
"NU_NOTA_CN"},Combiner.CombineTextByDelimiter(":", QuoteStyle.None),"CN"),

#"Colunas Mescladas1" = Table.CombineColumns(#"Colunas Mescladas",{"TP_PRESENCA_LC",
"NU_NOTA_LC"},Combiner.CombineTextByDelimiter(":", QuoteStyle.None),"LC"),

#"Colunas Mescladas2" = Table.CombineColumns(#"Colunas Mescladas1",{"TP_PRESENCA_MT",
"NU_NOTA_MT"},Combiner.CombineTextByDelimiter(":", QuoteStyle.None),"MT"),

#"Outras Colunas Removidas" = Table.SelectColumns(#"Colunas Mescladas2",{"NU_INSCRICAO", "TP_PRESENCA_CH", "CN",
"NU_NOTA_CH", "LC", "MT", "TP_STATUS_REDACAO", "NU_NOTA_REDACAO"}),

#"Colunas Mescladas3" = Table.CombineColumns(#"Outras Colunas Removidas",{"TP_PRESENCA_CH",
"NU_NOTA_CH"},Combiner.CombineTextByDelimiter(":", QuoteStyle.None),"CH"),

#"Colunas Mescladas4" = Table.CombineColumns(#"Colunas Mescladas3",{"TP_STATUS_REDACAO",
"NU_NOTA_REDACAO"},Combiner.CombineTextByDelimiter(":", QuoteStyle.None),"RD"),

#"Colunas Não Dinâmicas" = Table.UnpivotOtherColumns(#"Colunas Mescladas4", {"NU_INSCRICAO"}, "Atributo", "Valor"),

#"Colunas Renomeadas" = Table.RenameColumns(#"Colunas Não Dinâmicas",{{"Atributo", "CO_PROVA"}}),

#"Dividir Coluna por Delimitador" = Table.SplitColumn(#"Colunas Renomeadas", "Valor", Splitter.SplitTextByDelimiter(":",
QuoteStyle.Csv), {"Valor.1", "Valor.2"}),

#"Colunas Renomeadas1" = Table.RenameColumns(#"Dividir Coluna por Delimitador",{{"Valor.1", "TP_PRESENCA"}, {"Valor.2",
"NOTA"}}),

#"Tipo Alterado com Localidade" = Table.TransformColumnTypes(#"Colunas Renomeadas1", {{"NOTA", type number}}, "en-US"),

#"Valor Substituído" = Table.ReplaceValue(#"Tipo Alterado com Localidade", "CN", "Ciências da
Natureza", Replacer.ReplaceText, {"CO_PROVA"}),

#"Valor Substituído1" = Table.ReplaceValue(#"Valor Substituído", "CH", "Ciências Humanas", Replacer.ReplaceText, {"CO_PROVA"}),

#"Valor Substituído2" = Table.ReplaceValue(#"Valor Substituído1", "MT", "Matemática", Replacer.ReplaceText, {"CO_PROVA"}),

#"Valor Substituído3" = Table.ReplaceValue(#"Valor Substituído2", "LC", "Linguagens e
Códigos", Replacer.ReplaceText, {"CO_PROVA"}),

#"Valor Substituído4" = Table.ReplaceValue(#"Valor Substituído3", "RD", "Redação", Replacer.ReplaceText, {"CO_PROVA"}),

#"Tipo Alterado" = Table.TransformColumnTypes(#"Valor Substituído4", {{"TP_PRESENCA", Int64.Type}}

in

#"Tipo Alterado"
```

Na tabela Inscritos mantive as colunas que iria usar, transformei alguns tipos de dados e substitui alguns valores

Ex.: F para Femino, M para Masculino...

Comecei a importar algumas tabelas dimensão para exibir as UF por extenso

