



SEGUNDA ETAPA3– ANALÍTICO

A área de análise de dados pediu uma visão mais agregada dos dados de churn.

O que deverá ser feito?

Extração Ler dados de *churns* mensais do arquivo 2-CHURN_PERFIL_ENRIQUECIDO.csv que se encontra em C:\ProjetoChurn\Destino

Transformação e Saídas requeridas (Load)

Saída1: Esta saída atenderá a área de análise de dados.

Pede-se gravar os dados no arquivo **3-CHURN_PERFIL_ANALITICO.csv** em C:\ProjetoChurn\Destino

O arquivo poderá ser substituído a cada nova execução no mês.

O Layout deverá ser:

- **Header:** com o nome das colunas abaixo listadas:

	Name	Type	Format
1	genero	String	
!	uf	String	
!	anomes	Integer	#
!	estadocivil	String	
!	escolaridade	String	
!	faixaidade	String	
!	total_churn	Integer	#
!	valor_mensalidade_mes	Integer	#
!	valor_participacao_mes	Integer	#

- **Cabeçalho:** primeira linha nome das colunas
- **Separador das colunas:** ponto e vírgula (;
- **Ecoding:** UTF-8

PROJETO: TRATA CHURN



Transformações/limpeza sobre regras do negócio

- Criar uma faixa de idade seguindo os intervalos:
 - Até 29 anos – Jovem
 - De 30 a 50 anos – Adulto Jovem
 - De 51 a 70 anos – Adulto Maduro
 - Acima de 70 anos – Idoso
- Criar a categoria escolaridade:
Substituir: Pós-Graduação, Mestrado, Doutorado - Stricto/Lato Sensu
- Agregar os dados, gerando um perfil por: gênero, UF, anomes, estado civil, escolaridade e faixa de idade
- Criar as métricas:
 - Total churn – contagem
 - valor_mensalidade_mes – soma das mensalidades
 - valor_participacao_mes - soma das copartipação



Anexo

Regra step formula: Cria faixa etária e escolaridade

```
if(OR([escolaridade]= "Pós-Graduação";[escolaridade]= "Mestrado";[escolaridade]="Doutorado");"Stricto/Lato Sensu";[escolaridade])  
  
if([idade]<=29;"Jovem";if(AND([idade]>= 30;[idade]<= 50);"Adulto Jovem";if(AND([idade]>= 51;[idade]<=70); "Adulto Maduro ";"Idoso")))
```

Help PDI

<https://docs.hitachivantara.com/r/en-us/pentaho-data-integration-and-analytics/10.0.x/mk-95pdia003/pdi-transformation-steps>

UTF-8: é um tipo de codificação Unicode (padrão internacional para a representação e manipulação de texto em diferentes sistemas de escrita ao redor do mundo). UTF-8 é o formato de codificação mais amplamente utilizado e compatível com a maioria dos sistemas e linguagens. Ele é uma codificação variável que usa de 1 a 4 bytes para representar cada caractere.