**ROTEIRO TRATAMENTO DE DADOS**

Passo a passo do que o algoritmo de tratamento dos dados deverá executar.

**Base Consumo (DOUGLAS)**

*Fonte: EPE Anuario Estatistico de Energia*

1. Descartar colunas: B - Tipo Consumidor, F a K – SetorN2, Tensao N1, Tensao N2, Tensao N3, Faixa de consumo.
2. Reformatar coluna A - Data e coluna D - SetorN1
3. Agregar linhas repetidas (Data x Sistema x UF x SetorN1) somando para Consumo e Consumidores
4. Renomear (Data – Mês), (Sistema – Subsistema), (SetorN1 – Setor)

**Base Tarifa (BRUNO)**

*Fonte ANEEL Tarifas Homologadas Distribuidoras*

1. Converter todos os preços para R$/MWh
2. Descartar colunas: A, B, L, N, O (Conferir)
3. Agregar linhas repetidas
4. Verificar outliers da tabela, script com arvore de regras para descarte
5. Criar relação mês a mês de acordo com as datas de início e fim das tarifas.
   1. Mês tal, percorrer linhas da base, caso mês tal entre DataInicio e DataFim, inserir dados na linha, próximo mês.

**Base Capacidade de Geração (ANSELMO)**

*Fonte: ANEEL SIGA Empreendimentos Geração*

1. Descartar colunas: A, C, F, J, K, M, P, Q, R, U, V, W
2. Selecionar apenas linhas onde DscFaseUsina : “Operação”
3. Descartar coluna DscFaseUsina
4. Criar colunas “Início considerado” e “Fim considerado”. Utilizar de árvore de regras para definir quais datas utilizar e preencher colunas:
   1. Se DatEntradaOperacao < 2018, considerar DatInicioVigencia
      1. Se ...
5. Tratar Datas restantes (Colunas início e fim considerado podem ter trazido erros):
   1. Criar coluna “Outlier”
   2. Se InicioConsiderado tiver ano ~ 1900 🡪 Null, e classificar coluna Outlier = True
      1. Se ...
      2. Se ...
6. Criar relação mês a mês de acordo com as datas de início e fim consideradas, com script.
   1. Mês tal, percorrer linhas da base, caso mês tal entre DataInicio e DataFim, inserir dados na linha, próximo **mês..**

**Base Disponibilidade de Geração (SANDERSON)**

*Fonte: ONS Disponibilidade*

1. Anos em arquivos separados, percorrer os arquivos e juntá-los em um só adicionando as linhas de um sob o outro (UNION de tabelas).
2. Base utiliza “.” como separador decimal, caso necessário, converter.

**Base Carga de Energia (SANDERSON)**

*Fonte: ONS Carga de Energia*

1. Anos em arquivos separados, percorrer os arquivos e juntá-los em um só adicionando as linhas de um sob o outro (UNION de tabelas).