

Documentação do Código

Resumo

O código em questão é responsável por realizar uma série de operações de processamento de dados utilizando o Apache Spark. Ele trabalha com conjuntos de dados relativos ao absenteísmo e corrigindo valores. Abaixo, segue uma descrição detalhada das funções e operações realizadas.

Funções

Rais_2008_2018()

Descrição:

Importa arquivos necessários, como `cbo_classificacoes.csv` e `cnae_industrial.parquet`, e realiza operações de limpeza e transformação nos dados.

Parâmetros:

- `var_adls_uri`: URI do Azure Data Lake Storage.
- `cbo_classificacoes_path`: Caminho para o arquivo `cbo_classificacoes.csv`.
- `raw_cnae_industrial_path`: Caminho para o arquivo `cnae_industrial.parquet`.

Retorno:

DataFrame com os dados processados.

CD_CAUSA_AFASTAMENTO_08_18_NAO_INCLUINDO_10_11(x)

Descrição:

Filtra os dados da função `Rais_2008_2018()` para os anos 2008, 2009 e 2012 a 2018.

Parâmetros:

- `x`: Tipo de dado a ser processado.

Retorno:

DataFrame com os dados filtrados.

CD_CAUSA_AFASTAMENTO_10_11(y)

Descrição:

Filtra os dados da função `Rais_2008_2018()` para os anos 2010 e 2011.

Parâmetros:

- `y`: Tipo de dado a ser processado.

Retorno:

DataFrame com os dados filtrados.

Concatenate_DataFrames_2008_2018(x)

Descrição:

Concatena os resultados das funções `CD_CAUSA_AFASTAMENTO_08_18_NAO_INCLUINDO_10_11(x)` e `CD_CAUSA_AFASTAMENTO_10_11(y)`.

Parâmetros:

- `x`: Tipo de dado a ser processado.

Retorno:

DataFrame com os dados concatenados.

absent_custo(x)

Descrição:

Realiza operações de transformação nos dados da função `Concatenate_DataFrames_2008_2018(x)` para calcular o custo direto do absenteísmo por ano, UF, CNAE e CBO.

Parâmetros:

- `x`: Tipo de dado a ser processado.

Retorno:

DataFrame com os custos calculados.

****Indice_Absenteismo_ANO_2008_2018(x, kwargs)**

Descrição:

Calcula o índice de absenteísmo por ano, UF, CNAE e CBO a partir dos dados da função `Concatenate_DataFrames_2008_2018(x)`.

Parâmetros:

- `x`: Tipo de dado a ser processado.
- `kwargs`: Parâmetros adicionais que podem ser passados para customizar o cálculo do índice.

Retorno:

DataFrame com os índices calculados.

****Custo_Direto_ANO_2008_2018(x, kwargs)**

Descrição:

Calcula o custo direto do absenteísmo por ano, UF, CNAE e CBO a partir dos dados da função `Concatenate_DataFrames_2008_2018(x)`.

Parâmetros:

- `x`: Tipo de dado a ser processado.
- `kwargs`: Parâmetros adicionais que podem ser passados para customizar o cálculo do custo direto.

Retorno:

DataFrame com os custos calculados.

****teste(x, kwargs)**

Descrição:

Função de teste que realiza operações de transformação nos dados da função `Concatenate_DataFrames_2008_2018(x)`.

Parâmetros:

- `x`: Tipo de dado a ser processado.
- `kwargs`: Parâmetros adicionais que podem ser passados para customizar a função de teste.

Retorno:

DataFrame com os resultados da função de teste.

Operações

- Importação de arquivos e seleção de registros.
 - Limpeza e transformação de dados.
 - Agrupamento e ordenação de dados.
 - Cálculo de índices e custos.
-

Parâmetros

- `x`: Parâmetro obrigatório que representa o tipo de dado a ser processado.
 - `kwargs`: Parâmetros adicionais que podem ser passados para customizar os cálculos.
-

Retorno

Os resultados dos cálculos são retornados em forma de DataFrames do Apache Spark.

Exemplos de Uso

Rais_2008_2018():

```
df = Rais_2008_2018()
CD_CAUSA_AFASTAMENTO_08_18_NAO_INCLUINDO_10_11(x):
python
Copiar
df = CD_CAUSA_AFASTAMENTO_08_18_NAO_INCLUINDO_10_11('Indústria')
CD_CAUSA_AFASTAMENTO_10_11(y):
python
Copiar
df = CD_CAUSA_AFASTAMENTO_10_11('Não Indústria')
Concatenate_DataFrames_2008_2018(x):
python
Copiar
```

```

df = Concatenate_DataFrames_2008_2018('Indústria')
absent_custo(x):
python
Copiar
df = absent_custo('Indústria')
**Indice_Absenteismo_ANO_2008_2018(x, kwargs):
python
Copiar
df = Indice_Absenteismo_ANO_2008_2018('Indústria', ano=2018)
**Custo_Direto_ANO_2008_2018(x, kwargs):
python
Copiar
df = Custo_Direto_ANO_2008_2018('Indústria', ano=2018)
**teste(x, kwargs):
python
Copiar
df = teste('Indústria', ano=2018)
Notas
Os parâmetros kwargs podem ser utilizados para customizar os cálculos e operações realizadas pelas funções.
Os resultados dos cálculos são retornados em forma de DataFrames do Apache Spark.
As funções podem ser utilizadas para realizar operações de processamento de dados em larga escala.

```

```

<style type="text/css">@media print {
  *, :after, :before {background: 0 0 !important;color: #000 !important;box-shadow: none !important;text-shadow: none !im
  a, a:visited {text-decoration: underline}
  a[href]:after {content: " (" attr(href) ")"}
  abbr[title]:after {content: " (" attr(title) ")"}
  a[href^="#"]:after, a[href^="javascript:"]:after {content: ""}
  blockquote, pre {border: 1px solid #999;page-break-inside: avoid}
  thead {display: table-header-group}
  img, tr {page-break-inside: avoid}
  img {max-width: 100% !important}
  h2, h3, p {orphans: 3;widows: 3}
  h2, h3 {page-break-after: avoid}
}
html {font-size: 12px}
@media screen and (min-width: 32rem) and (max-width: 48rem) {
  html {font-size: 15px}
}
@media screen and (min-width: 48rem) {
  html {font-size: 16px}
}
body {line-height: 1.85}
.air-p, p {font-size: 1rem;margin-bottom: 1.3rem}
.air-h1, .air-h2, .air-h3, .air-h4, h1, h2, h3, h4 {margin: 1.414rem 0 .5rem;font-weight: inherit;line-height: 1.42}
.air-h1, h1 {margin-top: 0;font-size: 3.998rem}
.air-h2, h2 {font-size: 2.827rem}
.air-h3, h3 {font-size: 1.999rem}
.air-h4, h4 {font-size: 1.414rem}
.air-h5, h5 {font-size: 1.121rem}
.air-h6, h6 {font-size: .88rem}
.air-small, small {font-size: .707em}
canvas, iframe, img, select, svg, textarea, video {max-width: 100%}
body {color: #444;font-family: 'Open Sans', Helvetica, sans-serif;font-weight: 300;margin: 0;text-align: center}
img {border-radius: 50%;height: 200px;margin: 0 auto;width: 200px}
a, a:visited {color: #3498db}
a:active, a:focus, a:hover {color: #2980b9}
pre {background-color: #fafafa;padding: 1rem;text-align: left}
blockquote {margin: 0;border-left: 5px solid #7a7a7a;font-style: italic;padding: 1.33em;text-align: left}
li, ol, ul {text-align: left}
p {color: #777}</style>

```