

Fornecedores

August 17, 2023



```
[ ]: # Versão da linguagem Python e arquitetura do Jupyter Notebook
import platform
print('Versão da linguagem Python utilizada neste notebook:', platform.
      ↪python_version())
print('Arquitetura do Jupyter utilizada neste notebook:', platform.
      ↪architecture()[0])

# Outra forma de apresentar a versão da linguagem Python
# import sys
# print('Versão da linguagem Python:', sys.version)
```

1 DA1 - 2023: Limpeza e transformação de dados de fornecedores SC

1.1 Instalando os pacotes do projeto

```
[ ]: # Imports
# Manipulação de dados
import pandas as pd
import numpy as np

# Ignore warning
import sys
import warnings
if not sys.warnoptions:
    warnings.simplefilter("ignore")

[ ]: # Para verificar as versões dos pacotes utilizados neste notebook.
%reload_ext watermark
%watermark -a "vll.adm.br" --iversions
```

1.2 Carregando o dataset de trabalho

Iremos trabalhar com o dataset de Cadastro de Fornecedores, disponibilizado pela Diretoria de Gestão de Licitações e Contratos da Secretaria de Estado da Administração de Santa Catarina - SEA/SC. A base de dados, atualizada em 15/06/23, encontra-se disponível em <https://github.com/SmartSC-SEA/Fornecedores-cadastrados-E-LIC>.

```
[ ]: # Criando o dataframe de trabalho
df = pd.read_csv('datasets/ELIC_fornecedores_cadastro.csv')
```

```
[ ]: # Verificando o tipo do dataframe
type(df)
```

```
[ ]: # Verificando o formato do dataframe
df.shape
```

```
[ ]: # Listando os nomes dos campos
df.columns
```

```
[ ]: # Informações gerais do dataframe
df.info()
```

1.3 Análise exploratória dos dados

```
[ ]: # Listando uma amostra do dataset
df.head()
```

1.3.1 Dividindo a coluna 'produtos habilitados' e criando as colunas 'Grupo' e 'Classe'

```
[ ]: # Dividindo a coluna 'produtos_habilitados' em partes usando o caracter '-'
df1 = df
df1[['Grupo', 'Classe', 'Descricao']] = df['produtos_habilitados'].str.split('-',
↵      ↵, 2, expand=True)
```

```
[ ]: df1.head()
```

```
[ ]: # Dividindo o código do grupo-classe em duas colunas
df2 = df1
df2['Grupo1'] = df2['Grupo'].str[0:2]
df2['Grupo2'] = df2['Grupo'].str[2:]
```

```
[ ]: df2.head()
```

```
[ ]: df2.rename(columns={'Grupo': 'Código GC'}, inplace=True)
```

```
[ ]: df2.head()
```

```
[ ]: df2['Grupo_desc'] = df2['Grupo1']+' - '+df2['Classe']
df2['Classe_desc'] = df2['Grupo2']+' - '+df2['Descricao']

[ ]: df2.head()

[ ]: df2.drop(columns=['Classe','Descricao','Grupo1','Grupo2'], inplace=True)

[ ]: df2.head()

[ ]: df2.rename(columns={'Grupo_desc':'Grupo','Classe_desc':'Classe'},inplace=True)

[ ]: df2.head()

[ ]: # Gerando um arquivo csv do novo dataset df2
df2.to_csv('./Datasets/df2_fornecedores.csv', index=False)
```