

# Relatório de Dados de Impressão (RDI)

Objetivo: integrar informações dos serviços de outsourcing de impressão da Petrobras e oferecer: visualização consolidada e análise de predição de consumo através da aplicação de técnicas de **Machine Learning (ML)** e **Natural Language Processing (NLP)**.

## Etapa1: Leitura dos dados

```
In [71]: #codigo
import pandas as pd

#data = pd.read_csv('entrada.txt', header = None)
data = pd.read_csv('entrada.txt')
data
```

Out[71]:

	CHAVE	LOTACAOEMPREGADO	PRÉDIO EMPREGADO	NOME DOC IMPRESSO	NUMPAG	DATA
0	ur5d	TD/EA	EDISE	How Digital Transformation is changing our world	13	02/02/2019
1	upmz	TIC/TIC-US/GDGAS-US	EDICIN	Enterprise Data Lakes	20	08/07/2018
2	ur5d	TD/EA	EDISE	Python for dummines	89	02/02/2019
3	ur5d	TD/EA	EDISE	relat[orio horas	89	02/02/2019
4	ur5d	TD/EA	EDISE	email joao	89	02/02/2019
5	ur5d	TD/EA	EDISE	foto minha filha fofinha	1	02/09/2018
6	upmz	TIC/TIC-US/GDGAS-US	EDICIN	Upstream Data Management	45	06/07/2018

## Etapa 2: Como estamos hoje? (Visualização Consolidada)

Qual a média de impressão por mês por usuário? (número de páginas)

```
In [102]: #codigo
#data.groupby('MES_IMP')['NUMPAG'].mean()
temp=data.groupby('CHAVE').mean()
temp
```

Out[102]:

	NUMPAG	MES_IMP
CHAVE		
upmz	32.5	7.0
ur5d	56.2	3.4

Qual a média de impressão por mês? (número de páginas)

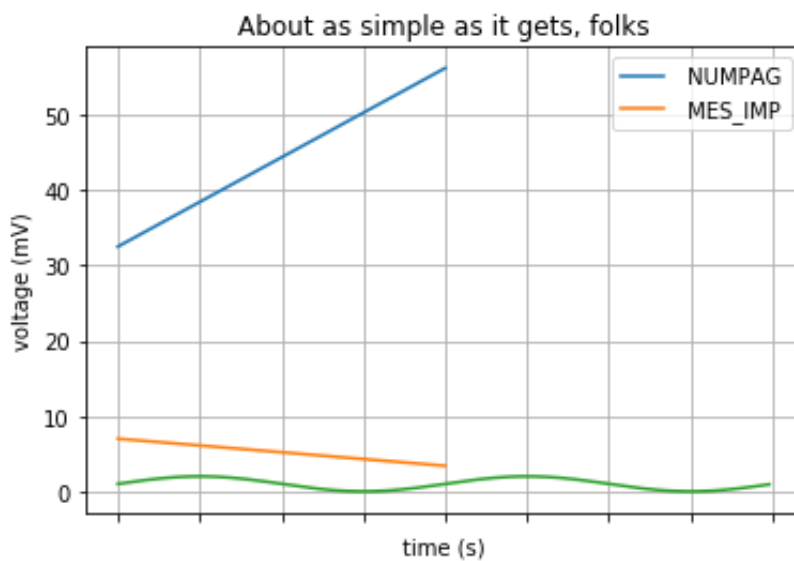
```
In [73]: #codigo
```

Quais são os 10 maiores usuários de impressoras?

```
In [104]: #codigo de um exemplo genérico
import matplotlib.pyplot as plt
import numpy as np

t = np.arange(0.0, 2.0, 0.01)
s = 1 + np.sin(2*np.pi*t)
plt.plot(t, s)

plt.xlabel('time (s)')
plt.ylabel('voltage (mV)')
plt.title('About as simple as it gets, folks')
plt.grid(True)
plt.savefig("test.png")
plt.show()
```



Quais são as 10 maiores gerências usuárias de impressoras?

```
In [88]: #codigo
temp=data.groupby('CHAVE').max('NUMPAG')
temp
```

```
-----
-----
TypeError                                Traceback (most recent c
all last)
<ipython-input-88-47204e4f9626> in <module>()
      1 #codigo
----> 2 temp=data.groupby('CHAVE').max('NUMPAG')
      3 temp

TypeError: f() takes 1 positional argument but 2 were given
```

Qual o percentual de impressões feitas em impressoras equivocadas? (usuário do prédio A seleciona impressora do prédio B)

In [78]: `#codigo`

## Etapa 3: Para onde iremos amanhã? (Predição de consumo)

Qual a previsão de ranking das 10 maiores gerências consumidoras do serviço de impressão?

In [13]: `#codigo`

Aplicando Natural Language Processing é possível identificar impressões de documentos com nomes "estranhos"?

In [14]: `#codigo`

by TD@2019