

## ▼ Prova MAC0459 - 2021

**Aluno:** Daniel Angelo Esteves Lawand - **NUSP:** 10297693

**Professor:** Roberto Hirata

### ▼ 1.0 - Questão 1

Tomaremos o [conjunto de dados de acidentes aéreos no Brasil](#), para responder as seguintes perguntas:

1. Qual o ano com maior quantidade de ocorrências?
2. Qual o mês do ano com maior quantidade de ocorrências?
3. Qual Unidade Federativa teve o maior número de ocorrências?
4. Qual município teve o maior número de ocorrências?
5. Qual tipo de ocorrência é a mais frequente?

### ▼ 1.1 - Tarefa 1 EDA

O dataset observado engloba as planilhas "ocorrencia.csv" e "ocorrencia\_tipo.csv". Sendo que estas são suficientes para responder as perguntas elaboradas. A planilha "ocorrencia.csv" aponta diversos dados sobre as ocorrências aéreas, porém usaremos apenas a data e a localização da ocorrência. A planilha "ocorrencia\_tipo.csv" é uma planilha mais enxuta, porém possui dados descritivos sobre o tipo de ocorrência, e estes dados são os que usaremos na nossa análise.

```
# Importação de bibliotecas
import matplotlib.pyplot as plt
import pandas as pd
import datetime
import seaborn as sns
```

```
# Caso esteja usando o Colaboratory do Google
from google.colab import files
uploaded = files.upload()
```

No file chosen

```
# Criação dos dataframes
df_ocorrencias = pd.read_csv("ocorrencia.csv", sep=";")
df_ocorrencias_tipo = pd.read_csv("ocorrencia_tipo.csv", sep=";")

# Alteração do tipo do dado, de string para datetime
df_ocorrencias["ocorrencia_dia"] = pd.to_datetime(df_ocorrencias["ocorrencia_dia"])
```

```
# Mostra de como são os dados da planilha "ocorrencia.csv"
df_ocorrencias.head()
```

	<b>codigo_ocorrencia</b>	<b>codigo_ocorrencia1</b>	<b>codigo_ocorrencia2</b>	<b>codigo_ocorre</b>
<b>0</b>	40211	40211	40211	
<b>1</b>	40349	40349	40349	
<b>2</b>	40351	40351	40351	
<b>3</b>	39527	39527	39527	
<b>4</b>	40324	40324	40324	



```
# Mostra de como são os dados da planilha "ocorrencia_tipo.csv"
df_ocorrencias_tipo.head()
```

	<b>codigo_ocorrencia1</b>	<b>ocorrencia_tipo</b>	<b>ocorrencia_tipo_categoria</b>	<b>taxon</b>
<b>0</b>	39115	PANE SECA	COMBUSTÍVEL   PANE SECA	
<b>1</b>	39155	VAZAMENTO DE COMBUSTÍVEL	FALHA OU MAU FUNCIONAMENTO DE SISTEMA / COMPON...	
<b>2</b>	39156	FOGO EM VOO	FOGO/FUMAÇA (SEM IMPACTO)   FOGO EM VOO	FALHA OU MAU

```
# LISTA 1.1.0 - Contagem da quantidade de ocorrências por ano
df_ocorrencias["year"] = pd.DatetimeIndex(df_ocorrencias["ocorrencia_dia"]).year
df_ocorrencias["year"].value_counts()
```

```
2013    654
2012    647
2011    576
2014    569
2010    552
2020    510
2019    495
2015    471
2018    444
2017    432
2016    403
2021    361
Name: year, dtype: int64
```

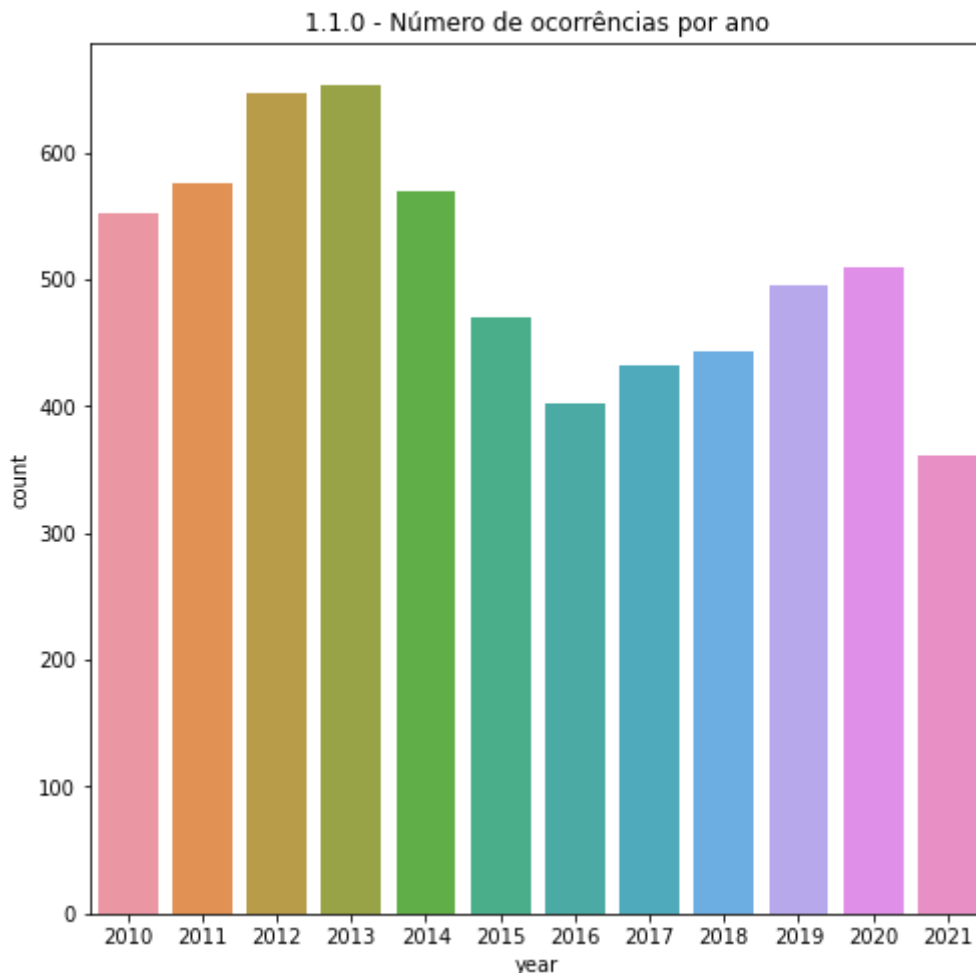
```
# LISTA 1.1.1 - Contagem da quantidade de ocorrências por mês
df_ocorrencias["month"] = pd.DatetimeIndex(df_ocorrencias["ocorrencia_dia"]).month
```

```
df_ocorrencias["month"].value_counts()
```

```
3    579
1    562
9    546
10   509
4    507
7    504
8    501
5    493
11   492
6    484
2    475
12   462
```

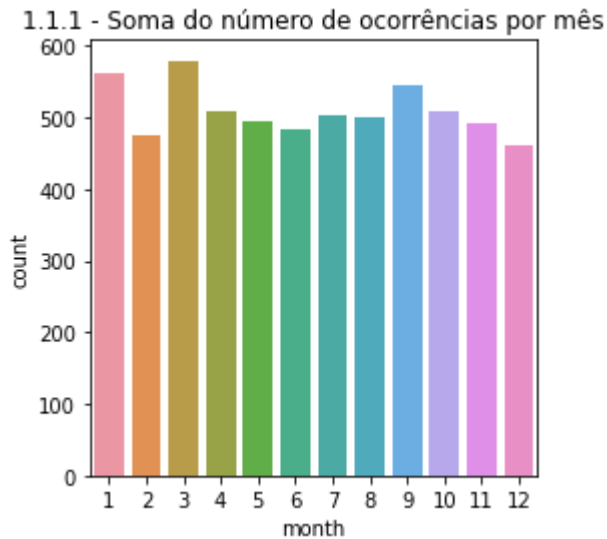
```
Name: month, dtype: int64
```

```
# Plotagem do gráfico da quantidade de ocorrências por ano
plt.rcParams["figure.figsize"] = [8.00, 8.00]
sns.countplot(x='year',data=df_ocorrencias).set_title("1.1.0 - Número de ocorrências por ano")
plt.show()
```



```
# Plotagem do gráfico da quantidade de ocorrências por mês
plt.rcParams["figure.figsize"] = [4.00, 4.00]
sns.countplot(x='month',data=df_ocorrencias).set_title("1.1.1 - Soma do número de ocorrências por mês")
plt.show()
```





```
# LISTA 1.1.2 - Contagem da quantidade de ocorrências por Unidade Federativa
df_ocorrencias["ocorrencia_uf"].value_counts(dropna=False)
```

```
SP      1464
MG       570
RJ       536
PR       502
RS       371
GO       333
MT       320
PA       305
AM       245
BA       230
SC       196
MS       168
DF       142
PE       107
CE        90
ES        83
MA        76
AC        62
RR        62
TO        55
PI        41
RO        41
AL        34
PB        25
SE        22
RN        18
AP        14
***         2
```

```
Name: ocorrencia_uf, dtype: int64
```

```
# LISTA 1.1.3 - Contagem da quantidade de ocorrências por cidade
df_ocorrencias["ocorrencia_cidade"].value_counts()
```

```
RIO DE JANEIRO    331
SÃO PAULO        291
BELO HORIZONTE    196
CAMPINAS          176
```

```

GUARULHOS          163
...
PARAÚNA            1
PRADÓPOLIS        1
MAIRIPORÃ         1
AGUDOS            1
ARENÁPOLIS        1
Name: ocorrencia_cidade, Length: 1133, dtype: int64

```

# LISTA 1.1.4 - Contagem da quantidade de ocorrências por cada tipo de ocorrência  
df\_ocorrencias\_tipo["ocorrencia\_tipo"].value\_counts()

```

FALHA DO MOTOR EM VOO          781
FALHA OU MAU FUNCIONAMENTO DE SISTEMA / COMPONENTE  677
ESTOURO DE PNEU                652
PERDA DE CONTROLE NO SOLO      476
PERDA DE CONTROLE EM VOO       381
...
PERDA DE SEPARAÇÃO / COLISÃO EM VOO          1
RELACIONADO COM SECURITY                1
PERDA DE CONDIÇÕES DE SUSTENTAÇÃO EM ROTA    1
EXPLOSÃO                                1
HIPÓXIA                                  1
Name: ocorrencia_tipo, Length: 81, dtype: int64

```

Pergunta 1: Qual o ano com maior quantidade de ocorrências?

Observando a LISTA 1.1.0, podemos perceber que o ano de mais ocorrências aéreas no dataset é o ano de 2013, com 654 ocorrências.

Pergunta 2: Qual o mês do ano com maior quantidade de ocorrências?

Observando a LISTA 1.1.1, podemos perceber que o mês de mais ocorrências aéreas no dataset é o mês de Março, com 579 ocorrências.

Pergunta 3: Qual Unidade Federativa teve o maior número de ocorrências?

Observando a LISTA 1.1.2, podemos perceber que a Unidade Federativa com mais ocorrências aéreas no dataset é o estado de São Paulo, com 1454 ocorrências.

Pergunta 4: Qual município teve o maior número de ocorrências?

Observando a LISTA 1.1.3, podemos perceber que a cidade com mais ocorrências aéreas no dataset é o município do Rio de Janeiro, com 331 ocorrências.

Pergunta 5: Qual tipo de ocorrência é a mais frequente?

Observando a LISTA 1.1.4, podemos perceber que o tipo de ocorrência mais frequente é a de "FALHA DO MOTOR EM VOO", com frequência de 781.

## Estratégia

A ideia é construir dois dataframes com pandas, cada qual correspondendo a uma das tabelas que usaremos. Introduziremos duas novas colunas no dataframe `df_ocorrencias`, uma indicando o mês e a outra o ano da ocorrência, para manusear melhor os dados. Após isso, iremos utilizar o método `value_counts()`, para adquirir o valor da frequência de incidentes aéreos por atributo selecionado. Após fazer isso, podemos fazer o plot de gráficos. Alguns gráficos não foram gerados, pois não fazia sentido colocá-los.

---

✓ 0s conclusão: 22:55

