

In [2]:

```
!pip install plotly
```

Collecting plotly

Downloading plotly-5.10.0-py2.py3-none-any.whl (15.2 MB)

----- 15.2/15.2 MB 10.6 MB/s eta 0:0

0:00

Collecting tenacity>=6.2.0

Downloading tenacity-8.0.1-py3-none-any.whl (24 kB)

Installing collected packages: tenacity, plotly

Successfully installed plotly-5.10.0 tenacity-8.0.1

In [3]:

```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import numpy as np
import plotly.express as px

dataFrame = pd.read_csv('C:\\Users\\dpits\\Desktop\\eu\\Repositorios\\Python\\BI\\Dia6Fabri
dataFrame.head()
```

Out[3]:

	id	data_inversa	dia_semana	horario	uf	br	km	municipio	causa_acidente
0	331730.0	2021-01-01	sexta-feira	05:30:00	SP	116.0	453	REGISTRO	Reação tardia ou ineficiente do condutor
1	331804.0	2021-01-01	sexta-feira	08:05:00	AM	174.0	937	MANAUS	Reação tardia ou ineficiente do condutor
2	331815.0	2021-01-01	sexta-feira	10:10:00	CE	222.0	128,5	ITAPAJE	Velocidade Incompatível
3	331823.0	2021-01-01	sexta-feira	12:30:00	RJ	493.0	18	MAGE	Ingestão de álcool e/ou substâncias psicoativa...
4	331843.0	2021-01-01	sexta-feira	14:40:00	RJ	393.0	252	BARRA DO PIRAI	Condutor Dormindo

5 rows × 30 columns

In [4]:

```
dataFrame.shape
```

Out[4]:

(64539, 30)

In [6]:

```
# pedido para remover as duplicatas, no caso não há  
dataFrame = dataFrame.drop_duplicates()  
dataFrame.shape
```

Out[6]:

(64539, 30)

In [8]:

```
#verificar qtde valores nulos existentes e retorna em sumário  
dataFrame.isnull().sum()
```

Out[8]:

id	0
data_inversa	0
dia_semana	0
horario	0
uf	0
br	167
km	167
municipio	0
causa_acidente	0
tipo_acidente	0
classificacao_acidente	0
fase_dia	0
sentido_via	0
condicao_metereologica	0
tipo_pista	0
tracado_via	0
uso_solo	0
peessoas	0
mortos	0
feridos_leves	0
feridos_graves	0
ilesos	0
ignorados	0
feridos	0
veiculos	0
latitude	0
longitude	0
regional	0
delegacia	18
uop	132
dtype:	int64

In [10]:

```
#trabalhando os dados = criando colunas para marcar colunas que tinham valores nulos
#dados[''] = np.where(dados[''].isnull(), 'null', 'n/a')

dataFrame['uop'] = np.where(dataFrame['uop'].isnull(), 'null', 'n/a')

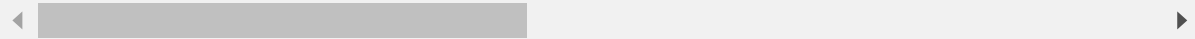
# verificando a base
dataFrame.head()

#Aqui temos o código para criar as colunas de validação se há ou não valores nulos.
#Isso é importante para manter um histórico das alterações realizadas na base.
```

Out[10]:

	id	data_inversa	dia_semana	horario	uf	br	km	municipio	causa_acidente
0	331730.0	2021-01-01	sexta-feira	05:30:00	SP	116.0	453	REGISTRO	Reação tardia ou ineficiente do condutor
1	331804.0	2021-01-01	sexta-feira	08:05:00	AM	174.0	937	MANAUS	Reação tardia ou ineficiente do condutor
2	331815.0	2021-01-01	sexta-feira	10:10:00	CE	222.0	128,5	ITAPAJE	Velocidade Incompatível
3	331823.0	2021-01-01	sexta-feira	12:30:00	RJ	493.0	18	MAGE	Ingestão de álcool e/ou substâncias psicoativa...
4	331843.0	2021-01-01	sexta-feira	14:40:00	RJ	393.0	252	BARRA DO PIRAI	Condutor Dormindo

5 rows × 30 columns



In [12]:

```
#verificar retorno da base após alteração:
```

```
dataFrame.isnull().sum()
```

Out[12]:

```
id                0
data_inversa      0
dia_semana        0
horario           0
uf                0
br                167
km                167
municipio         0
causa_acidente    0
tipo_acidente     0
classificacao_acidente 0
fase_dia          0
sentido_via       0
condicao_metereologica 0
tipo_pista        0
tracado_via       0
uso_solo          0
pessoas           0
mortos            0
feridos_leves     0
feridos_graves    0
ilesos            0
ignorados         0
feridos           0
veiculos          0
latitude          0
longitude         0
regional          0
delegacia         18
uop               0
dtype: int64
```

In [14]:

```
dataFrame['delegacia'] = np.where(dataFrame['delegacia'].isnull(), 'null', 'n/a')
dataFrame['br'] = np.where(dataFrame['br'].isnull(), 'null', 'n/a')
dataFrame['km'] = np.where(dataFrame['km'].isnull(), 'null', 'n/a')
dataFrame.isnull().sum()
```

Out[14]:

id	0
data_inversa	0
dia_semana	0
horario	0
uf	0
br	0
km	0
municipio	0
causa_acidente	0
tipo_acidente	0
classificacao_acidente	0
fase_dia	0
sentido_via	0
condicao_metereologica	0
tipo_pista	0
tracado_via	0
uso_solo	0
pessoas	0
mortos	0
feridos_leves	0
feridos_graves	0
ilesos	0
ignorados	0
feridos	0
veiculos	0
latitude	0
longitude	0
regional	0
delegacia	0
uop	0
dtype: int64	

In [15]:

```
dataFrame.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 64539 entries, 0 to 64538
Data columns (total 30 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                    64539 non-null  float64
1   data_inversa                         64539 non-null  object
2   dia_semana                           64539 non-null  object
3   horario                              64539 non-null  object
4   uf                                    64539 non-null  object
5   br                                    64539 non-null  object
6   km                                    64539 non-null  object
7   municipio                            64539 non-null  object
8   causa_acidente                       64539 non-null  object
9   tipo_acidente                        64539 non-null  object
10  classificacao_acidente                64539 non-null  object
11  fase_dia                             64539 non-null  object
12  sentido_via                           64539 non-null  object
13  condicao_metereologica                 64539 non-null  object
14  tipo_pista                           64539 non-null  object
15  tracado_via                           64539 non-null  object
16  uso_solo                              64539 non-null  object
17  pessoas                              64539 non-null  int64
18  mortos                               64539 non-null  int64
19  feridos_leves                         64539 non-null  int64
20  feridos_graves                       64539 non-null  int64
21  ileso                                 64539 non-null  int64
22  ignorados                            64539 non-null  int64
23  feridos                              64539 non-null  int64
24  veiculos                             64539 non-null  int64
25  latitude                             64539 non-null  object
26  longitude                             64539 non-null  object
27  regional                             64539 non-null  object
28  delegacia                            64539 non-null  object
29  uop                                   64539 non-null  object
dtypes: float64(1), int64(8), object(21)
memory usage: 15.3+ MB
```

In [16]:

```
#Estatísticas Descritivas >> Retorna medidas de tendência central  
#exibe dados estatísticos da base  
#Quantidade, média, desvio padrão, valor mínimo, 25% da coluna, 50% da coluna (média), 75%  
dataFrame.describe()
```

Out[16]:

	id	pessoas	mortos	feridos_leves	feridos_graves	ilesos
count	64539.000000	64539.000000	64539.000000	64539.000000	64539.000000	64539.000000
mean	369745.487194	2.338508	0.083608	0.839508	0.273710	0.981639
std	21954.249547	1.866810	0.340734	1.029122	0.612703	1.440071
min	331693.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	351095.500000	1.000000	0.000000	0.000000	0.000000	0.000000
50%	369790.000000	2.000000	0.000000	1.000000	0.000000	1.000000
75%	388100.500000	3.000000	0.000000	1.000000	0.000000	1.000000
max	465993.000000	70.000000	19.000000	44.000000	26.000000	53.000000

In [19]:

```
#Alternar colunas e linhas - outra visualização da base
#Exibe 6 registros:
```

```
dataFrame.sample(6).T
```

Out[19]:

	53369	22886	54268	38211	53413
id	389096.0	340357.0	390487.0	365413.0	389157.0
data_inversa	2021-10-11	2021-02-13	2021-10-18	2021-06-20	2021-10-11
dia_semana	segunda-feira	sábado	segunda-feira	domingo	segunda-feira
horario	13:55:00	02:50:00	15:15:00	19:50:00	17:10:00
uf	CE	PR	MG	SE	PR
br	n/a	n/a	n/a	n/a	n/a
km	n/a	n/a	n/a	n/a	n/a
municipio	FORTALEZA	PONTA GROSSA	ARAGUARI	ITAPORANGA DAJUDA	CURITIBA
causa_acidente	Manobra de mudança de faixa	Velocidade Incompatível	Ingestão de álcool pelo condutor	Manobra de mudança de faixa	Condutor deixou de manter distância do veículo...
tipo_acidente	Capotamento	Saída de leito carroçável	Saída de leito carroçável	Colisão com objeto	Colisão traseira
classificacao_acidente	Com Vítimas Feridas	Com Vítimas Feridas	Sem Vítimas	Com Vítimas Feridas	Com Vítimas Feridas
fase_dia	Pleno dia	Plena Noite	Pleno dia	Plena Noite	Plena Noite
sentido_via	Decrescente	Decrescente	Crescente	Crescente	Crescente
condicao_meteorologica	Céu Claro	Céu Claro	Sol	Garoa/Chuvisco	Chuva
tipo_pista	Dupla	Dupla	Dupla	Dupla	Dupla
tracado_via	Não Informado	Curva	Reta	Reta	Reta
uso_solo	Sim	Sim	Sim	Não	Não
pessoas	2	2	1	4	2
mortos	0	0	0	0	0
feridos_leves	1	1	0	2	1
feridos_graves	0	1	0	0	0
ilesos	0	0	1	2	1
ignorados	1	0	0	0	0
feridos	1	2	0	2	1
veiculos	1	1	1	1	2
latitude	-3,87839882	-25,15559918	-18,65453092	-11,01647157	-25,5599323
longitude	-38,50066893	-50,13322909	-48,14727203	-37,32010446	-49,30743945

	53369	22886	54268	38211	53413
regional	SPRF-CE	SPRF-PR	SPRF-MG	SPRF-SE	SPRF-PR
delegacia	n/a	n/a	n/a	n/a	n/a
uop	n/a	null	n/a	n/a	n/a

In [20]:

```
#Alternar colunas e linhas - outra visualização da base
#Exibe 5 primeiros registros:
```

```
dataFrame[0:5].T
```

Out[20]:

	0	1	2	3	4
id	331730.0	331804.0	331815.0	331823.0	331843.0
data_inversa	2021-01-01	2021-01-01	2021-01-01	2021-01-01	2021-01-01
dia_semana	sexta-feira	sexta-feira	sexta-feira	sexta-feira	sexta-feira
horario	05:30:00	08:05:00	10:10:00	12:30:00	14:40:00
uf	SP	AM	CE	RJ	RJ
br	n/a	n/a	n/a	n/a	n/a
km	n/a	n/a	n/a	n/a	n/a
municipio	REGISTRO	MANAUS	ITAPAJE	MAGE	BARRA DO PIRAI
causa_acidente	Reação tardia ou insuficiente	Reação tardia ou insuficiente	Velocidade	Ingestão de álcool e/ou	Condutor

In [21]:

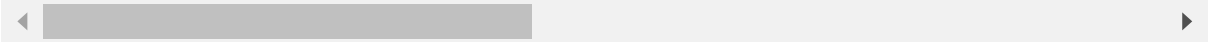
```
#Exibe 5 últimos registros: (linhas)
```

```
dataFrame.tail()
```

Out[21]:

	id	data_inversa	dia_semana	horario	uf	br	km	municipio	causa_acidente
64534	465833.0	2021-11-17	quarta-feira	19:00:00	MG	n/a	n/a	NOVA SERRANA	Transitar no acostamento
64535	465953.0	2021-10-27	quarta-feira	18:20:00	MG	n/a	n/a	MURIAE	Acessar a via sem observar a presença dos outr...
64536	465955.0	2021-10-28	quinta-feira	08:50:00	RJ	n/a	n/a	SAO JOAO DE MERITI	Condutor deixou de manter distância do veículo...
64537	465965.0	2021-11-07	domingo	20:15:00	SP	n/a	n/a	JUQUITIBA	Transitar no acostamento
64538	465993.0	2021-08-25	quarta-feira	15:10:00	RO	n/a	n/a	PORTO VELHO	Acessar a via sem observar a presença dos outr...

5 rows × 30 columns



In [23]:

```
#Alternar colunas e linhas - outra visualização da base
#Exibe as linhas solicitadas:
```

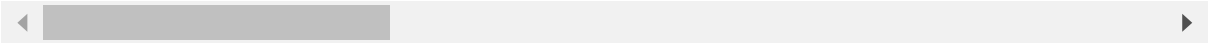
```
dataFrame[100:200].T
```

Out[23]:

	100	101	102	103	104	
id	332575.0	332582.0	332588.0	332590.0	332597.0	
data_inversa	2021-01-05	2021-01-05	2021-01-05	2021-01-02	2021-01-05	
dia_semana	terça-feira	terça-feira	terça-feira	sábado	terça-feira	
horario	11:00:00	12:37:00	11:20:00	19:20:00	13:00:00	
uf	PR	PR	RJ	MG	AL	
br	n/a	n/a	n/a	n/a	n/a	
km	n/a	n/a	n/a	n/a	n/a	
municipio	GUARATUBA	FAZENDA RIO GRANDE	ITATIAIA	LARANJAL	MATA GRANDE	
causa_acidente	Mal súbito do condutor	Condutor desrespeitou a iluminação vermelha do...	Reação tardia ou ineficiente do condutor	Ausência de reação do condutor	Transitar na contramão	
tipo_acidente	Colisão com objeto	Colisão traseira	Saída de leito carroçável	Colisão traseira	Colisão frontal	Coli:
classificacao_acidente	Com Vítimas Feridas	Com Vítimas Feridas	Com Vítimas Feridas	Com Vítimas Feridas	Com Vítimas Feridas	C
fase_dia	Pleno dia	Pleno dia	Pleno dia	Plena Noite	Pleno dia	
sentido_via	Decrescente	Decrescente	Crescente	Crescente	Crescente	C
condicao_meteorologica	Céu Claro	Sol	Céu Claro	Chuva	Céu Claro	
tipo_pista	Dupla	Dupla	Dupla	Simples	Simples	
tracado_via	Reta	Interseção de vias	Reta	Reta	Não Informado	Regi
uso_solo	Não	Sim	Não	Sim	Não	
pessoas	1	3	1	4	2	
mortos	0	0	0	0	0	
feridos_leves	1	1	1	1	0	
feridos_graves	0	0	0	0	1	
ilesos	0	1	0	3	1	
ignorados	0	1	0	0	0	
feridos	1	1	1	1	1	
veiculos	1	2	1	2	2	

	100	101	102	103	104	
latitude	-25,96596616	-25,6367641	-22,46068623	-21,36499168	-8,953151	-4
longitude	-48,89663319	-49,3130591	-44,45262504	-42,47388147	-37,822001	-3
regional	SPRF-PR	SPRF-PR	SPRF-RJ	SPRF-MG	SPRF-AL	
delegacia	n/a	n/a	n/a	n/a	n/a	
uop	n/a	n/a	n/a	n/a	n/a	

30 rows × 100 columns



In [35]:

```
dataFrame = dataFrame.rename(columns = {"causa_acidente": "causa", "classificacao_acidente":  
                                         "condicao_metereologica": "clima"})
```

In [39]:

```
#Selecionando Múltiplas Colunas  
#exibir tabela com colunas escolhidas/selecionadas  
  
selecionados = dataFrame[['uf', 'br', 'causa', 'classificacao', 'clima', 'mortos', 'latitud  
selecionados['clima'].value_counts()
```

Out[39]:

Céu Claro	38983
Nublado	10960
Chuva	6710
Sol	4295
Garoa/Chuveisco	2148
Ignorado	806
Nevoeiro/Neblina	488
Vento	144
Neve	3
Granizo	2
Name: clima, dtype: int64	

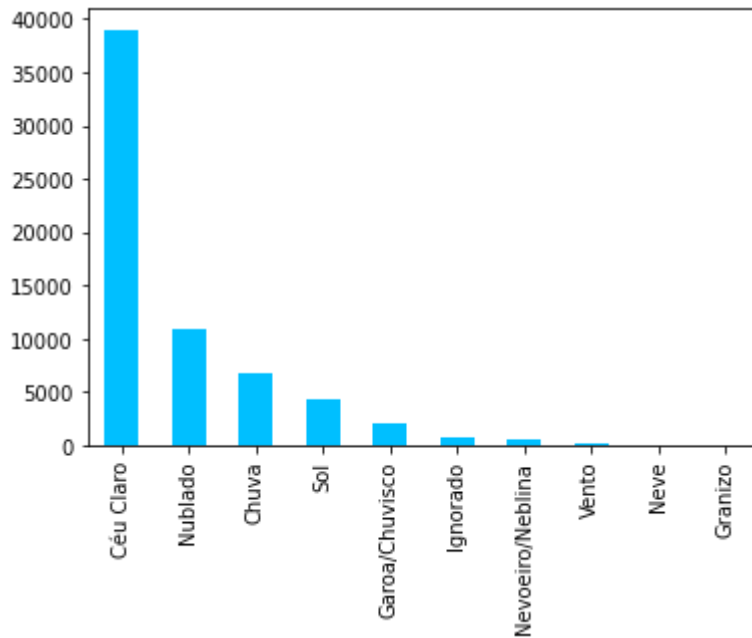
In [79]:

```
#gráfico de barras vertical
```

```
selecionados['clima'].value_counts().plot.bar(color='#00BFFF')
```

Out[79]:

<AxesSubplot:>



In [62]:

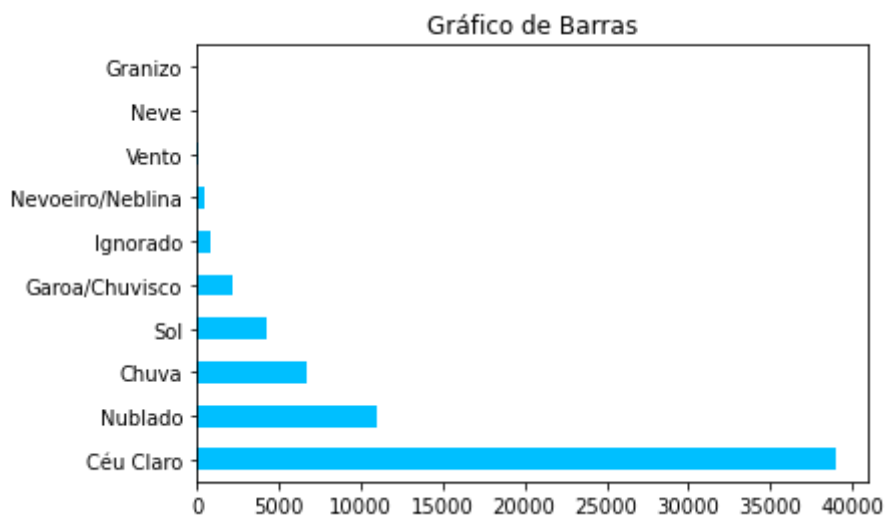
```
#graficos > barras horizontal

grafico1 = selecionados['clima'].value_counts()

grafico1.plot(
    x = 'causa acidente',
    kind = 'barh',
    stacked = True,
    title = 'Gráfico de Barras',
    color='#00BFFF',
    mark_right = True)
```

Out[62]:

<AxesSubplot:title={'center':'Gráfico de Barras'}>



In [61]:

```
selecionados['classificacao'].value_counts()
```

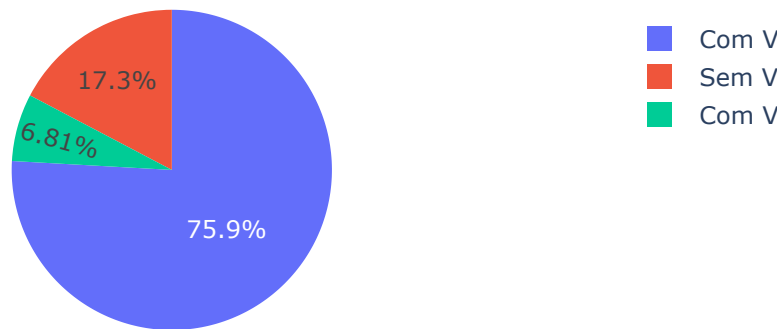
Out[61]:

```
Com Vítimas Feridas    48186
Sem Vítimas            11689
Com Vítimas Fatais     4664
Name: classificacao, dtype: int64
```

In [60]:

```
#lib plot > pizza
dados_x= ['Com Vítimas Feridas', 'Com Vítimas Fatais', 'Sem Vítimas' ]
dados_y= ['51183', '4592', '11671']

fig = px.pie(names=dados_x, values= dados_y, width=700, height=300)
fig.show()
```

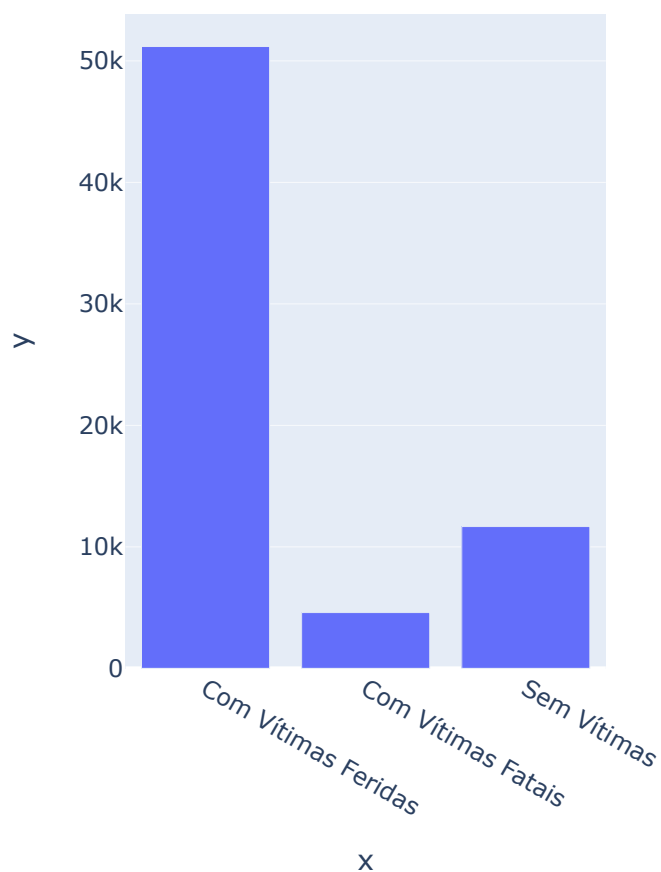


In [63]:

```
#lib plot > barras vertical  
#gráfico interativo
```

```
dados_x= ['Com Vítimas Feridas', 'Com Vítimas Fatais', 'Sem Vítimas' ]  
dados_y= [51183, 4592, 11671]
```

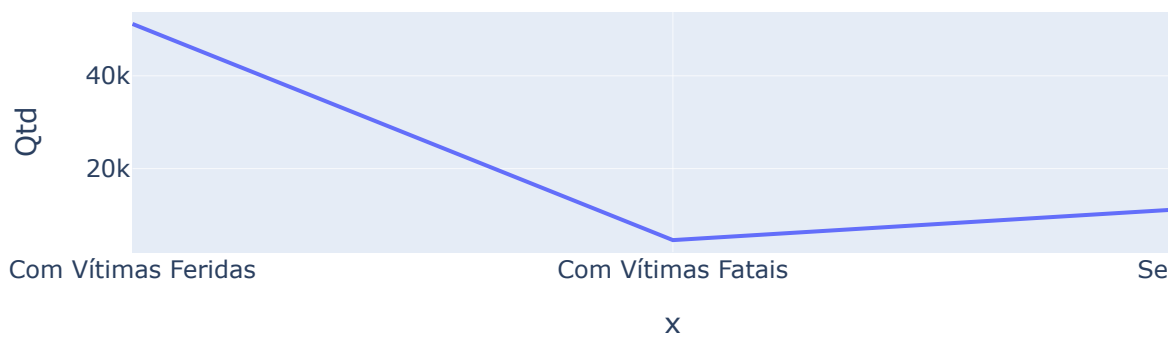
```
fig = px.bar(x=dados_x, y=dados_y, width=400, height=500)  
fig.show()
```



In [64]:

```
#lib plot > linhas  
#grafico interativo  
#gráfico com apenas 1 ano fica incorreto.. utilizar barras  
  
dados_x= ['Com Vítimas Feridas', 'Com Vítimas Fatais', 'Sem Vítimas' ]  
dados_y= [51183, 4592, 11671]  
  
fig = px.line(x=dados_x, y=dados_y, title= 'Acidentes por tipo em 2021', width=700, height=  
fig.update_yaxes(title='Qtd')  
fig.show()
```

Acidentes por tipo em 2021



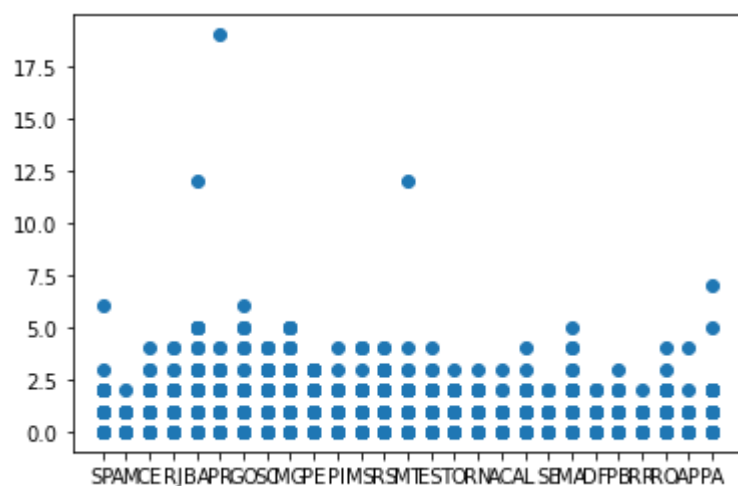
In [65]:

```
dataFrame.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 64539 entries, 0 to 64538
Data columns (total 30 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    64539 non-null  float64
1   data_inversa          64539 non-null  object
2   dia_semana            64539 non-null  object
3   horario               64539 non-null  object
4   uf                    64539 non-null  object
5   br                    64539 non-null  object
6   km                    64539 non-null  object
7   municipio             64539 non-null  object
8   causa                 64539 non-null  object
9   tipo_acidente         64539 non-null  object
10  classificacao          64539 non-null  object
11  fase_dia              64539 non-null  object
12  sentido_via           64539 non-null  object
13  clima                 64539 non-null  object
14  tipo_pista            64539 non-null  object
15  tracado_via           64539 non-null  object
16  uso_solo              64539 non-null  object
17  pessoas               64539 non-null  int64
18  mortos                64539 non-null  int64
19  feridos_leves         64539 non-null  int64
20  feridos_graves        64539 non-null  int64
21  ilesos                64539 non-null  int64
22  ignorados             64539 non-null  int64
23  feridos               64539 non-null  int64
24  veiculos              64539 non-null  int64
25  latitude              64539 non-null  object
26  longitude             64539 non-null  object
27  regional              64539 non-null  object
28  delegacia             64539 non-null  object
29  uop                   64539 non-null  object
dtypes: float64(1), int64(8), object(21)
memory usage: 15.3+ MB
```

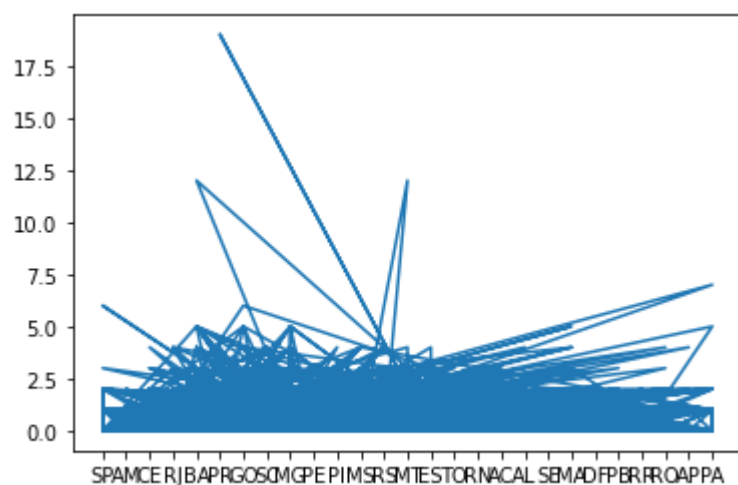
In [66]:

```
uf = dataframe['uf']  
mortos = dataframe['mortos']  
plt.scatter(uf, mortos)  
plt.show()
```



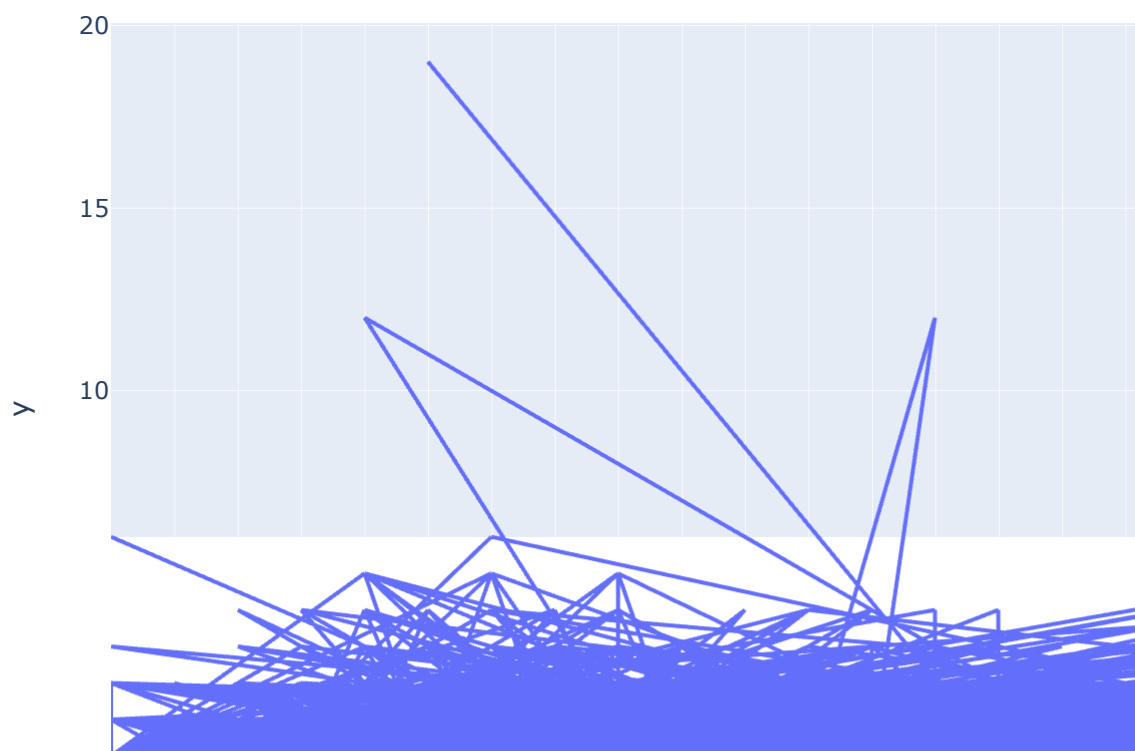
In [67]:

```
plt.plot(uf, mortos)  
plt.show()
```



In [68]:

```
px.line(x=uf, y=mortos)
```



In [75]:

```
selecionados['clima'].value_counts()
```

Out[75]:

Céu Claro	38983
Nublado	10960
Chuva	6710
Sol	4295
Garoa/Chuveisco	2148
Ignorado	806
Nevoeiro/Neblina	488
Vento	144
Neve	3
Granizo	2

Name: clima, dtype: int64

In [76]:

```
nublado = dataframe.loc[dataframe['clima'] == 'Nublado']
nublado.head()
```

Out[76]:

	id	data_inversa	dia_semana	horario	uf	br	km	municipio	causa	tipo_a
0	331730.0	2021-01-01	sexta-feira	05:30:00	SP	n/a	n/a	REGISTRO	Reação tardia ou ineficiente do condutor	Tomb
3	331823.0	2021-01-01	sexta-feira	12:30:00	RJ	n/a	n/a	MAGE	Ingestão de álcool e/ou substâncias psicoativa...	Colisã
4	331843.0	2021-01-01	sexta-feira	14:40:00	RJ	n/a	n/a	BARRA DO PIRAI	Condutor Dormindo	Colisã
8	331864.0	2021-01-01	sexta-feira	17:10:00	SC	n/a	n/a	INDAIAL	Transitar na contramão	Colisã
10	331910.0	2021-01-01	sexta-feira	19:50:00	SC	n/a	n/a	LONTRAS	Acessar a via sem observar a presença dos outr...	trar

5 rows × 30 columns

In [77]:

```
nublado.describe()
```

Out[77]:

	id	pessoas	mortos	feridos_leves	feridos_graves	ilesos
count	10960.000000	10960.000000	10960.000000	10960.000000	10960.000000	10960.000000
mean	369982.090602	2.338412	0.085128	0.846350	0.259672	0.974453
std	23320.304663	2.083481	0.372616	1.099538	0.571471	1.546872
min	331693.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	348867.500000	1.000000	0.000000	0.000000	0.000000	0.000000
50%	369077.500000	2.000000	0.000000	1.000000	0.000000	1.000000
75%	390875.250000	3.000000	0.000000	1.000000	0.000000	1.000000
max	465953.000000	57.000000	19.000000	44.000000	15.000000	52.000000

In [80]:

```
grafico = px.scatter_matrix(dataFrame, dimensions=['causa', 'clima'], color='causa')
grafico.show()
```

