

▼ Análise Exploratória de um Dataset de Casas para Alugar

Dataset name: *houses_to_rent_v2.csv*

Número de linhas: *10962 (número total de imóveis disponíveis para aluguel)*

Número de cidades: *5*

Data dos dados: *20/03/2020*

```
#importando as bibliotecas
import pandas as pd
import matplotlib.pyplot as plt
plt.style.use("seaborn-colorblind")
```

```
#upload do arquivo com os dados as serem analisados
from google.colab import files
arq = files.upload()
```

houses_to_rent_v2.csv

- **houses_to_rent_v2.csv**(application/vnd.ms-excel) - 697786 bytes, last modified: 7/19/2020 - 100% done
Saving houses_to_rent_v2.csv to houses_to_rent_v2 (1).csv

```
#criando o dataframe
df = pd.read_csv("houses_to_rent_v2.csv")
```

```
#visualizando amostra dos dados
df.head(3)
```

	city	area	rooms	bathroom	parking spaces	floor	animal	furniture	hoa (R\$)	rent amount (R\$)	property tax (R\$)	fire insurance (R\$)	total (R\$)
0	São Paulo	70	2	1	1	7	accept	furnished	2065	3300	211	42	5618
1	São	320	4	4	0	20	accept	not	1200	1060	1750	62	7072

```
#explorando a estrutura
df.shape
```

```
(10692, 13)
```

```
df.dtypes
```

```
city                object
area                int64
rooms               int64
bathroom            int64
parking spaces      int64
floor               object
animal              object
furniture            object
hoa (R$)             int64
rent amount (R$)     int64
property tax (R$)    int64
fire insurance (R$)  int64
total (R$)           int64
dtype: object
```

```
#realizando alguns tratamentos
```

```
df = df.rename(columns={"city":"Cidade", "rooms":"Quartos", "bathroom":"Banheiros", "parking spaces":"VagasEst", "floor":"Andar", "an
```

```
df.columns
```

```
Index(['Cidade', 'area', 'Quartos', 'Banheiros', 'VagasEst', 'Andar', 'Pets',
       'Mobiliado', 'Condominio', 'Aluguel', 'IPTU', 'Seg Incendio',
```

```
'total (R$)'],  
dtype='object')
```

```
df.head(2)
```

	Cidade	area	Quartos	Banheiros	VagasEst	Andar	Pets	Mobiliado	Condominio	Aluguel	IPTU	Seg Incendio	total (R\$)
0	São Paulo	70	2	1	1	7	accept	furnished	2065	3300	211	42	5618

```
df.tail(2)
```

	Cidade	area	Quartos	Banheiros	VagasEst	Andar	Pets	Mobiliado	Condominio	Aluguel	IPTU	Seg Incendio	total (R\$)
10690	Rio de Janeiro	120	2	2	2	8	accept	furnished	1585	12000	279	155	14020

Realizando algumas análises

```
#Informações estatísticas do meu dataset  
df.describe()
```

	area	Quartos	Banheiros	VagasEst	Condominio	Aluguel	IPTU	Seg Incendio	total
count	10692.000000	10692.000000	10692.000000	10692.000000	1.069200e+04	10692.000000	10692.000000	10692.000000	1.069200e+04
mean	149.217920	2.506079	2.236813	1.609147	1.174022e+03	3896.247194	366.704358	53.300879	5.490479

#Conhecendo a minha base de dados

```
df["Cidade"].unique()
```

```
array(['São Paulo', 'Porto Alegre', 'Rio de Janeiro', 'Campinas',  
      'Belo Horizonte'], dtype=object)
```

50%	90.000000	2.000000	2.000000	1.000000	5.600000e+02	2661.000000	125.000000	36.000000	3.581500
------------	-----------	----------	----------	----------	--------------	-------------	------------	-----------	----------

#Fatiando os dados

```
SaoPaulo = df.loc[df["Cidade"] == "São Paulo"]
```

```
PortoAlegre = df.loc[df["Cidade"] == "Porto Alegre"]
```

```
RioJaneiro = df.loc[df["Cidade"] == "Rio de Janeiro"]
```

```
Campinas = df.loc[df["Cidade"] == "Campinas"]
```

```
BeloHorizonte = df.loc[df["Cidade"] == "Belo Horizonte"]
```

```
BeloHorizonte.head(1)
```

```
BeloHorizonte["Cidade"].unique()
```

```
array(['Belo Horizonte'], dtype=object)
```

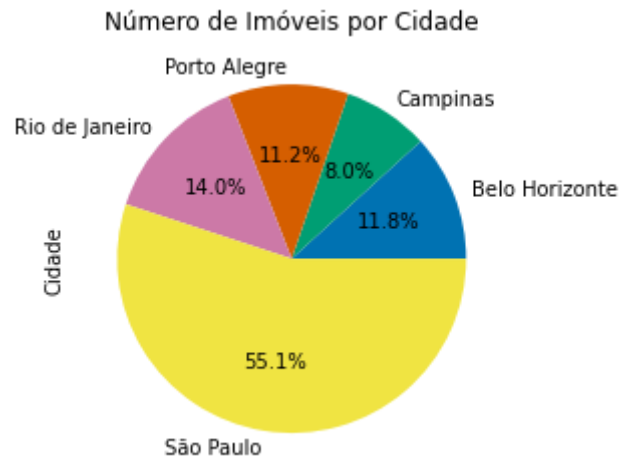
#verificando qual cidade tem mais ofertas de **aluguel** - análise simples sem considerar tamanho da cidade, população e etc

```
df.groupby("Cidade")["Cidade"].count()
```

```
Cidade
Belo Horizonte    1258
Campinas           853
Porto Alegre      1193
Rio de Janeiro    1501
São Paulo         5887
Name: Cidade, dtype: int64
```

#São Paulo e Rio de Janeiro tem o maior número de imóveis para alugar. Mostrando a mesma informação em um gráfico

```
df.groupby("Cidade")["Cidade"].count().plot.pie(title="Número de Imóveis por Cidade", autopct='%1.1f%%');
```



#verificando a média do valor total de **aluguel** por cidade

```
df.groupby("Cidade")["total (R$)"].mean()
```

```
Cidade
Belo Horizonte    6315.242448
Campinas          3173.276671
Porto Alegre      2989.782900
Rio de Janeiro    4611.684877
São Paulo         6380.831833
Name: total (R$), dtype: float64
```

#verificando graficamente a média do valor total de **aluguel** por cidade, sem considerar as características dos imóveis

```
plt.style.use("seaborn")
```

```
df.groupby("Cidade")["total (R$)"].mean().plot.bar(title="Valor total médio de aluguel por cidade", color="green")
```

```
plt.ylabel("Valor médio");
```



#comparando a aceitação de pets entre as cidades
`df.groupby("Cidade")["Pets"].value_counts()`

Cidade	Pets	
Belo Horizonte	accept	931
	not accept	327
Campinas	accept	693
	not accept	160
Porto Alegre	accept	1005
	not accept	188
Rio de Janeiro	accept	1205
	not accept	296
São Paulo	accept	4482
	not accept	1405

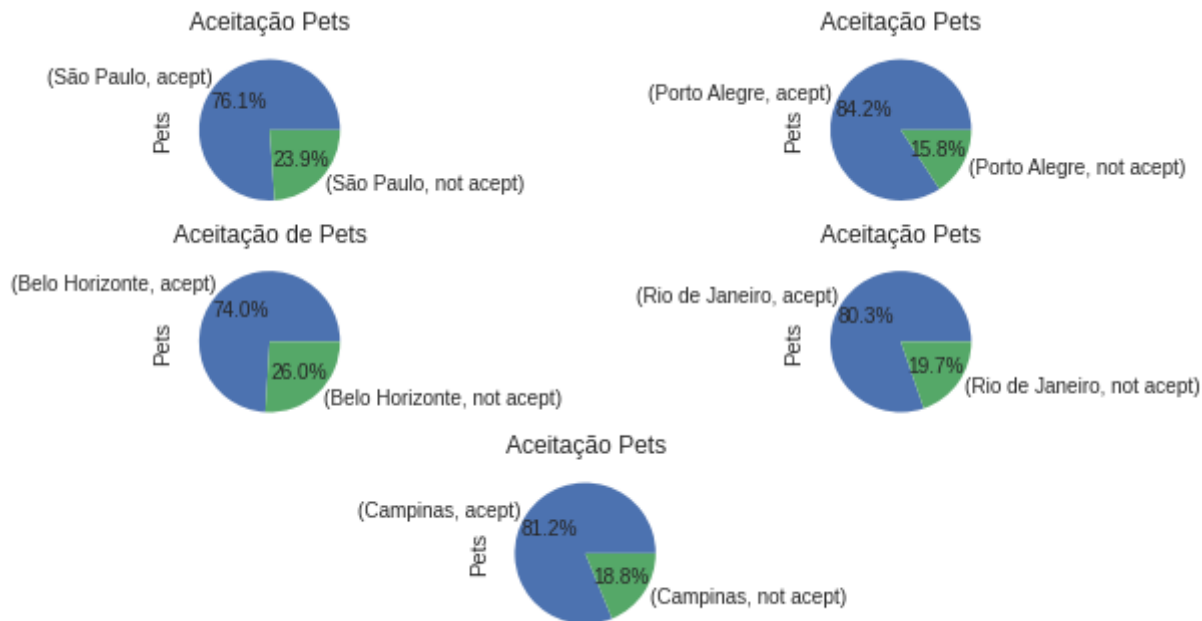
Name: Pets, dtype: int64

#comparando graficamente
`plt.style.use("seaborn")`
`plt.subplot(331)`
`SaoPaulo.groupby("Cidade")["Pets"].value_counts().plot.pie(title="Aceitação·Pets", autopct='%1.1f%%');`

```

plt.subplot(333)
PortoAlegre.groupby("Cidade")["Pets"].value_counts().plot.pie(title="Aceitação·Pets", autopct='%1.1f%%');
plt.subplot(334)
BeloHorizonte.groupby("Cidade")["Pets"].value_counts().plot.pie(title="Aceitação de Pets", autopct='%1.1f%%');
plt.subplot(336)
RioJaneiro.groupby("Cidade")["Pets"].value_counts().plot.pie(title="Aceitação Pets", autopct='%1.1f%%');
plt.subplot(338)
Campinas.groupby("Cidade")["Pets"].value_counts().plot.pie(title="Aceitação Pets", autopct='%1.1f%%');

```



```

#apuração do valor médio de IPTU
df.groupby("Cidade")["IPTU"].mean()

```

```

Cidade
Belo Horizonte    272.782194
Campinas          147.657679
Porto Alegre      124.021794
Rio de Janeiro    256.853431
São Paulo         495.701716
Name: IPTU, dtype: float64

```

```
#apuração do valor médio de Condomínio  
df.groupby("Cidade")["Condominio"].mean()
```

```
Cidade  
Belo Horizonte    2324.197138  
Campinas          628.922626  
Porto Alegre      491.618609  
Rio de Janeiro    1079.432378  
São Paulo         1169.627994  
Name: Condominio, dtype: float64
```

```
#verificando a quantidade de imóveis por número de quartos  
QTdImoveisNumQuartos = df.groupby(["Cidade","Quartos"])[ "Quartos"].count().reset_index(name='QTD Imóveis')  
QTdImoveisNumQuartos
```


	Cidade	Quartos	QTD Imóveis
0	Belo Horizonte	1	150
1	Belo Horizonte	2	286
2	Belo Horizonte	3	389
3	Belo Horizonte	4	317
4	Belo Horizonte	5	78
5	Belo Horizonte	6	20
6	Belo Horizonte	7	13
7	Belo Horizonte	8	4
8	Belo Horizonte	10	1
9	Campinas	1	218
10	Campinas	2	253
11	Campinas	3	271
12	Campinas	4	91
13	Campinas	5	14
14	Campinas	6	4
15	Campinas	7	2
16	Porto Alegre	1	374
17	Porto Alegre	2	424
18	Porto Alegre	3	299
19	Porto Alegre	4	62
20	Porto Alegre	5	22
21	Porto Alegre	6	9

22	Porto Alegre	7	1
23	Porto Alegre	8	2
24	Rio de Janeiro	1	404
25	Rio de Janeiro	2	522
26	Rio de Janeiro	3	422
27	Rio de Janeiro	4	128
28	Rio de Janeiro	5	18
29	Rio de Janeiro	6	3

```
#verificando a quantidade de imoveis com 4 quartos ou menos  
QTdImoveisNumQuartos.loc[QTdImoveisNumQuartos["Quartos"] <= 4]
```

	Cidade	Quartos	QTD Imóveis
0	Belo Horizonte	1	150
1	Belo Horizonte	2	286
2	Belo Horizonte	3	389
3	Belo Horizonte	4	317
9	Campinas	1	218
10	Campinas	2	253
11	Campinas	3	271
12	Campinas	4	91

#verificando a quantidade de imóveis com 4 dormitórios ou menos

```
QTdImoveisNumQuartos.loc[QTdImoveisNumQuartos["Quartos"] <= 4].sum()
```

```
Cidade      Belo HorizonteBelo HorizonteBelo HorizonteBelo...
Quartos                                50
QTD Imóveis                                10284
dtype: object
```

24	Rio de Janeiro	1	404
----	----------------	---	-----

#criando um subdataset apenas com os imóveis com 4 dormitórios ou menos - obs.: 10284 imóveis de um total de 10692

```
subdf = df.loc[df["Quartos"] <= 4]
```

24	Rio de Janeiro	1	404
----	----------------	---	-----

```
subdf.sample(10)
```

	Cidade	area	Quartos	Banheiros	VagasEst	Andar	Pets	Mobiliado	Condominio	Aluguel	IPTU	Seg Incendio	total (R\$)
3503	Campinas	47	1	1	1	18	accept	furnished	582	2700	135	35	3452
354	São Paulo	540	3	3	4	9	accept	not furnished	6200	15000	3500	191	24890
1383	São Paulo	170	3	2	4	6	accept	not furnished	3678	1500	1130	20	6328
								not					

```
subdf['total·(R$)'].describe().apply("{0:.2f}".format)
```

```
count      10284.00
mean        5281.67
std         16597.88
min          499.00
25%         2021.00
50%         3447.00
75%         6381.00
max        112000.00
Name: total (R$), dtype: object
```

#De acordo co o valor max do total de **aluguel** nota-se que valores muito discrepantes

```
subdf.loc[subdf['total·(R$)']>.30000]
```

	Cidade	area	Quartos	Banheiros	VagasEst	Andar	Pets	Mobiliado	Condominio	Aluguel	IPTU	Seg Incendio	total (R\$)
255	Belo Horizonte	155	1	4	0	4	not accept	not furnished	1117000	2790	64	38	1120000
1444	Porto Alegre	42	1	1	0	10	accept	not furnished	32000	700	40	11	32750
1000	São Paulo	100	1	1	0	1	accept	not furnished	10000	10000	10000	100	100000

```
subdf.loc[subdf['total (R$)'] > 30000].count()
```

```
Cidade      11
area        11
Quartos     11
Banheiros   11
VagasEst    11
Andar       11
Pets        11
Mobiliado   11
Condominio  11
Aluguel     11
IPTU        11
Seg Incendio 11
total (R$)  11
dtype: int64
```

#descartando as linhas com valor de **aluguel** discrepante - serão descartados imóveis com valores de **aluguel** acima de 30.000

```
subdf = subdf.loc[subdf['total (R$)'] < 30000]
```

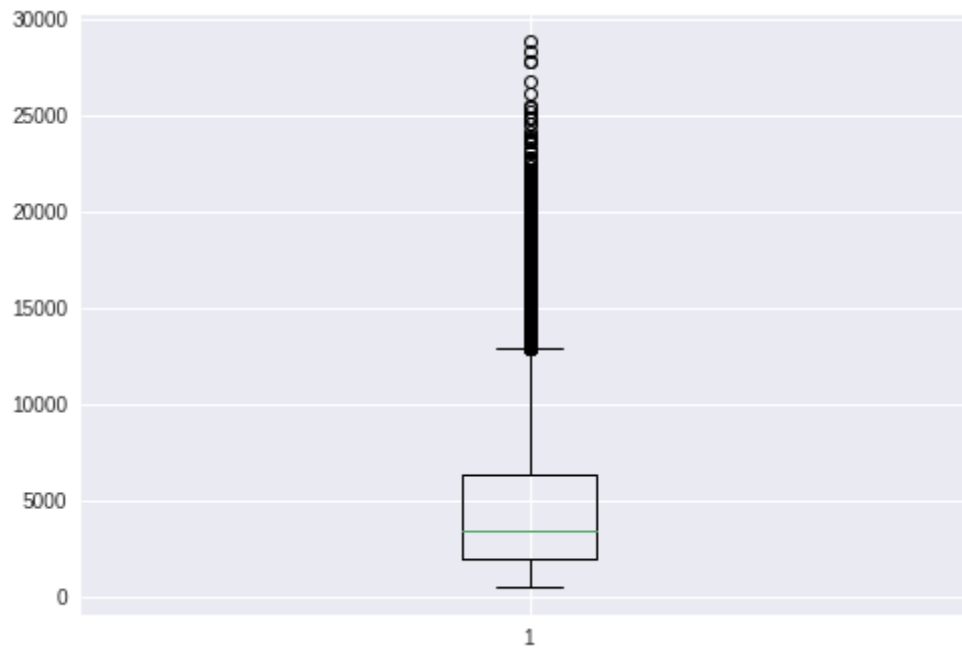
#verificando resultado após a limpeza

```
subdf['total (R$)'].describe().apply("{0:.2f}".format)
```

```
count      10273.00
mean        4986.67
std         4284.18
min         499.00
25%         2021.00
```

```
50%      3444.00
75%      6368.00
max      28790.00
Name: total (R$), dtype: object
```

```
#verificando graficamente
plt.boxplot(subdf['total (R$)']);
```



```
#pelo gráfico anterior é possível que ainda existem muitos valores discrepantes
subdf.loc[subdf['total (R$)'] > 15000]
```

	Cidade	area	Quartos	Banheiros	VagasEst	Andar	Pets	Mobiliado	Condominio	Aluguel	IPTU	Seg Incendio	total (R\$)
8	São Paulo	152	2	2	1	3	accept	furnished	1000	15000	250	191	16440
19	São Paulo	600	4	5	6	-	accept	not furnished	0	12000	9500	181	21680
40	São Paulo	300	4	6	4	20	accept	not furnished	3700	12000	1584	153	17440
86	São Paulo	600	4	4	4	9	accept	not furnished	7400	8500	3368	108	19380
106	São Paulo	300	4	5	4	-	accept	furnished	0	14000	1167	211	15380
...

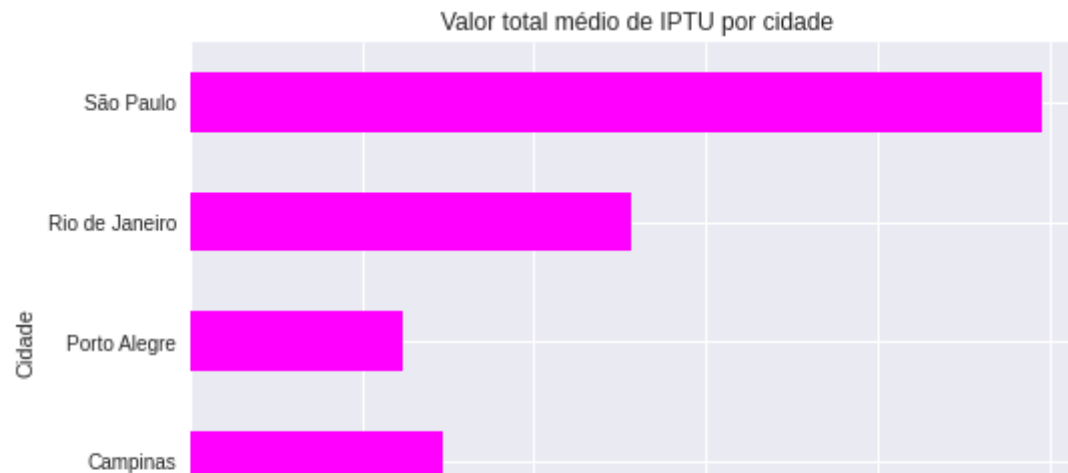
#continuando as análises com a base reduzida - imóveis.com.4.dormitórios.ou.menos

#apuração do valor médio de IPTU

```
df.groupby("Cidade")["IPTU"].mean()
```

```
Cidade
Belo Horizonte    272.782194
Campinas          147.657679
Porto Alegre      124.021794
Rio de Janeiro    256.853431
São Paulo         495.701716
Name: IPTU, dtype: float64
```

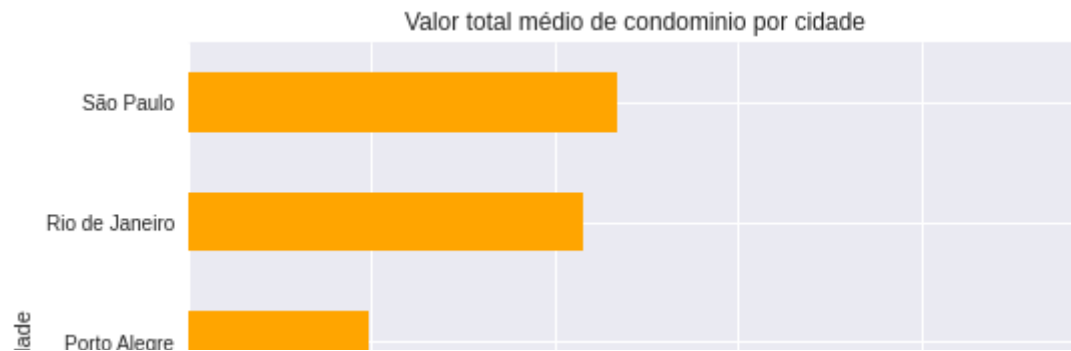
```
df.groupby("Cidade")["IPTU"].mean().plot.barh(title="Valor total médio de IPTU por cidade", color="magenta")
plt.xlabel("Valor médio IPTU");
```



#apuração do valor médio de Condomínio - imóveis com 4 dormitórios ou menos
`df.groupby("Cidade")["Condominio"].mean()`

```
Cidade
Belo Horizonte    2324.197138
Campinas          628.922626
Porto Alegre      491.618609
Rio de Janeiro    1079.432378
São Paulo         1169.627994
Name: Condominio, dtype: float64
```

```
df.groupby("Cidade")["Condominio"].mean().plot.barh(title="Valor total médio de condominio por cidade", color="orange")
plt.xlabel("Valor médio Condomínio");
```

#comparação entre os gastos com IPTU e condomínio

```
plt.subplot(121)
```

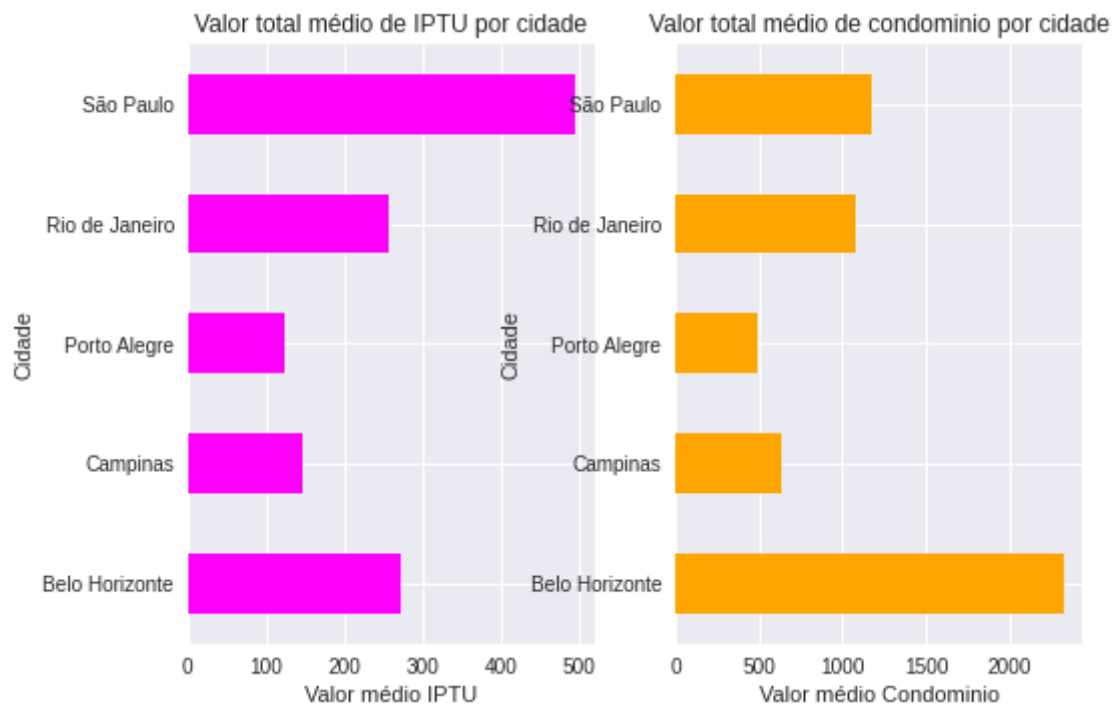
```
df.groupby("Cidade")["IPTU"].mean().plot.barh(title="Valor total médio de IPTU por cidade", color="magenta")
```

```
plt.xlabel("Valor médio IPTU");
```

```
plt.subplot(122)
```

```
df.groupby("Cidade")["Condominio"].mean().plot.barh(title="Valor total médio de condominio por cidade", color="orange")
```

```
plt.xlabel("Valor médio Condomínio");
```



✓ 0s completed at 01:04

