



Respondendo as perguntas de negócio

☰ Live	Ciclo 02: Análise Descritiva
🕒 Created	@April 4, 2023 8:31 PM
☑ Reviewed	<input type="checkbox"/>
# Aula	15
☑ Finished	<input checked="" type="checkbox"/>
☑ Ready	<input checked="" type="checkbox"/>

Objetivo da Aula

- ☐ Recapitulando o problema de negócio
- ☐ Resolução
- ☐ Produto Final
- ☐ Próxima aula

Conteúdo:

▼ 1. Recapitulando o problema de negócio

▼ 1.1 As perguntas do CEO

1. Quais são as **categorias de imóveis que estão cadastradas** dentro da base de dados da cidade de Nova York?
2. **Quantos usuários (Hosts) únicos** cadastrados existem dentro da base de dados da cidade de Nova York?

3. Como é a **variação do preços** dos imóveis em NY?
4. Existem mais **imóveis baratos ou caros**?
5. Qual a **distribuição do número de Reviews**? Existem imóveis com muitos e outro com poucos reviews?

▼ 2. Resolução

▼ 2.1. Planejamento do Processo

| Quais são os passos para encontrar as respostas?

1. Contar todos os nomes distintos que aparecem na coluna “room_type”
2. Contar os cadastros únicos da coluna “host_id”.
3. Calcular o desvio padrão em torno da média de preços dos imóveis.
4. Desenhar um histograma para mostrar o número de apartamentos dentro da faixa do valor do aluguel.
5. Desenhar um histograma para mostrar o número de apartamentos dentro de uma faixa de avaliação.

▼ 2.2. Coleta do conjunto de dados

New York City Airbnb Open Data
Airbnb listings and metrics in NYC, NY, USA (2019)

<https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>



▼ 2.2.1 Comando para ler uma planilha de dados

```
import pandas as pd

data = pd.read_csv( '/content/AB_NYC_2019.csv' )
```

▼ 2.3. Pergunta #1

Quais são as **categorias de imóveis que estão cadastradas** dentro da base de dados da cidade de Nova York?

▼ 2.3.1 Selecionar colunas

```
import pandas as pd
import numpy as np

data = pd.read_csv( '/content/AB_NYC_2019.csv' )

# selecionar o tipo de sala
room_type = data.loc[:, 'room_type']
```

▼ 2.3.2 Contar valores distintos

```
# mostrar os valores únicos
room_type_unique = np.unique( room_type )

# mostrar os valores
print( 'As categorias são:' )
print( room_type_unique )

Resposta: 'Entire home/apt', 'Private room', 'Shared room'
```

▼ 2.4. Pergunta #2

Quantos **usuários (Hosts) únicos** cadastrados existem dentro da base de dados da cidade de Nova York?

▼ 2.4.1 Selecionar colunas

```
import pandas as pd
import numpy as np

# selecionar a coluna host id
host_id = data.loc[:, 'host_id']
```

▼ 2.4.2 Contar valores distintos

```
import pandas as pd
import numpy as np

# selecionar a coluna neighbourhood_group
host_id_unique = np.unique( host_id )regioes_unicas
len( host_id_unique )

R: 37.457 hosts únicos
```

▼ 2.5. Pergunta #3

| Como é a **variação do preços** dos imóveis em NY?

▼ 2.5.1 Selecionar colunas

```
import pandas as pd
import numpy as np

# selecionar a coluna price
price = data.loc[:, 'price']
```

▼ 2.5.2 Encontrar o valor do desvio padrão

```
import pandas as pd
import numpy as np

# calcular o desvio padrão
desvio_padrao = np.std( price )
desvio_padrao

R: Os preços estão disperson em U$ 240.15 em torno da média
```

▼ 2.6. Pergunta #4

| Existem mais **imóveis baratos ou caros**?

▼ 2.6.1 Selecionar colunas

```
import pandas as pd
import numpy as np

# selecionar a coluna price e filtrar linhas
```

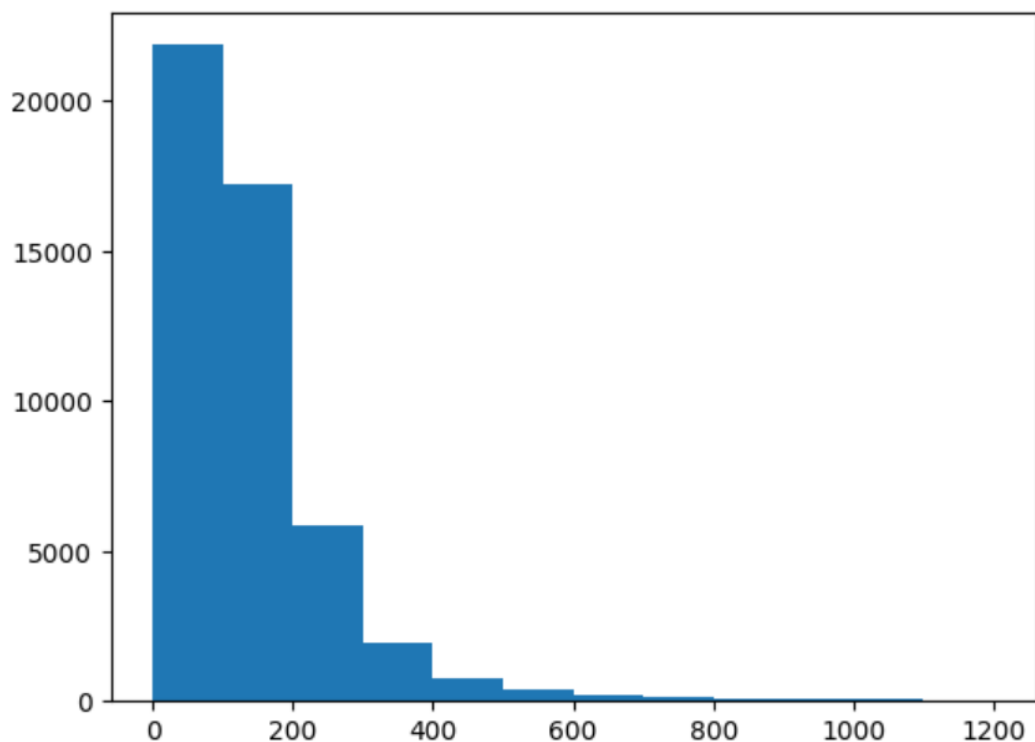
```
linhas = data.loc[:, 'price'] < 1250
price = data.loc[linhas, 'price']
```

▼ 2.6.2 Desenhar o histograma

```
from matplotlib import pyplot as plt

# desenhar o histograma
plt.hist( price, bins=12);

R: Existem mais de 20.000 imóveis com valor de aluguel
de até US$ 100,00
```



▼ 2.7. Pergunta #5

Qual a **distribuição do número de Reviews**? Existem imóveis com muitos e outro com poucos reviews?

▼ 2.7.1 Selecionar colunas

```
import pandas as pd
import numpy as np
```

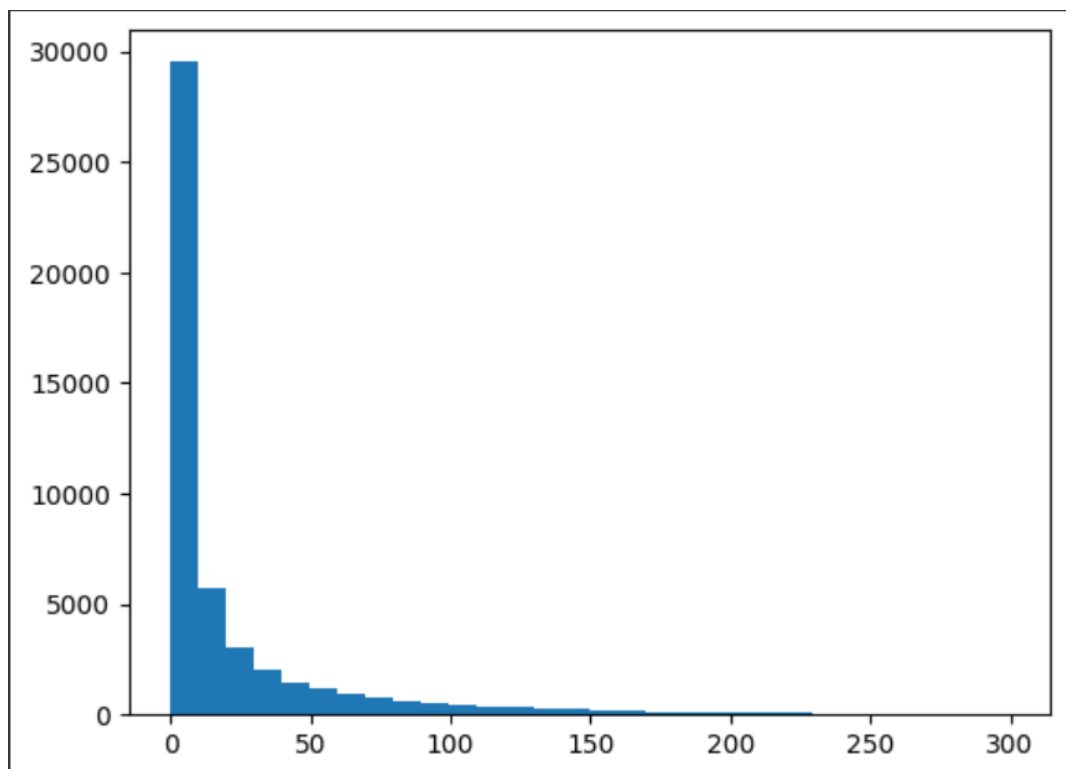
```
# selecionar a coluna price e filtrar linhas
linhas = data.loc[:, 'number_of_reviews'] < 300
number_reviews = data.loc[linhas, 'number_of_reviews']
```

▼ 2.7.2 Desenhar a distribuição de avaliações

```
from matplotlib import pyplot as plt

# desenhar o histograma
plt.hist( price, bins=12);

R: Existem quase 30.000 imóveis com até 10 avaliações.
```



▼ 3. Produto Final

| E-mail com as respostas

Olá CEO, espero que esse e-mail o encontre bem.

Meu nome é Meigarom, sou Cientista de Dados do time de expansão.

Segue abaixo as respostas das novas perguntas de negócio solicitadas, referente a expansão de Nova York.

▼ 1. Quais são as **categorias de imóveis que estão cadastradas** dentro da base de dados da cidade de Nova York?

R: Entire home/apt, Private room, Shared room

▼ 2. **Quantos usuários (Hosts) únicos** cadastrados existem dentro da base de dados da cidade de Nova York?

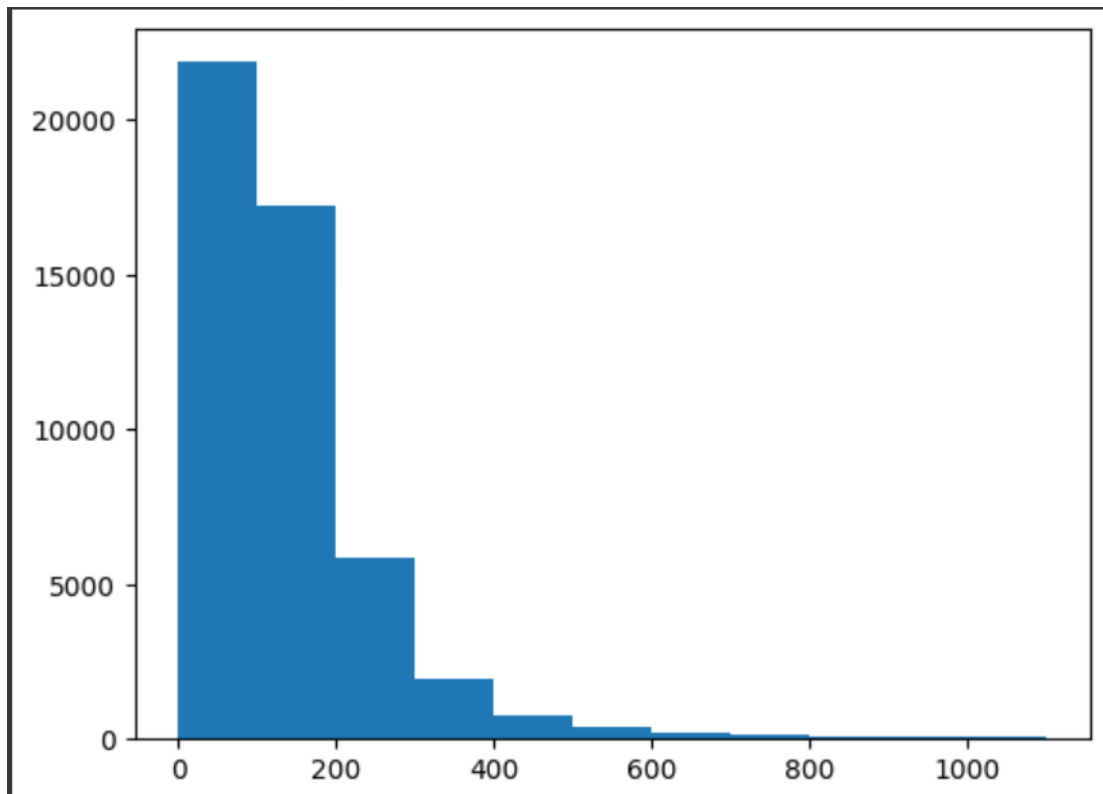
R: 37.457 hosts únicos cadastrados dentro da base de dados.

▼ 3. Como é a **variação do preços** dos imóveis em NY?

R: A variação média dos preços é média de U\$ 152.72 com desvio padrão de U\$ 240.15

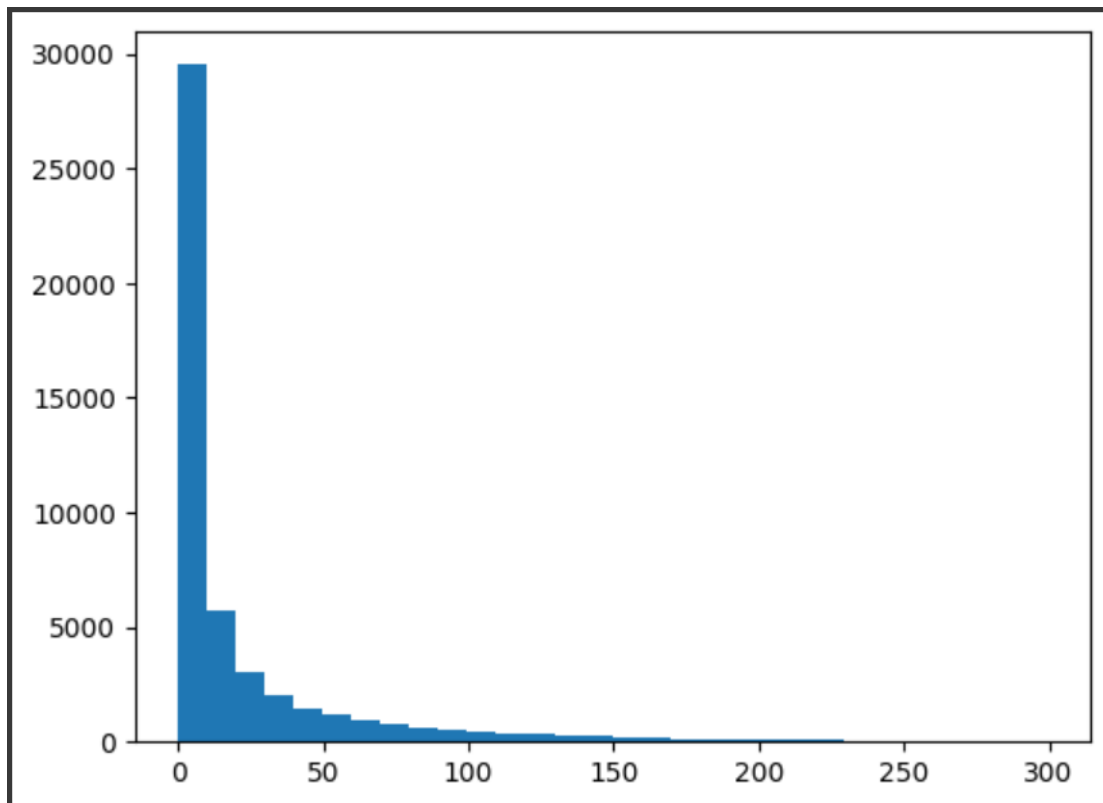
▼ 4. Existem mais **imóveis baratos ou caros**?

R: Existem mais de 2.000 imóveis com o valor de aluguel até U\$ 100 dólares. Ou seja, existe uma concentração de imóveis nos valores baixos de aluguel que caracteriza a maioria como sendo imóveis mais baratos. Veja o histograma abaixo



▼ 5. Qual a **distribuição do número de Reviews**? Existem imóveis com muitos e outro com poucos reviews?

R: Existem quase 30.000 imóveis com até 10 avaliações. Ou seja, existem muitos imóveis com poucas avaliações. Veja o histograma abaixo



▼ 4. Próxima aula

Exercícios