

GROUP 6

```
In [16]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
```

```
In [15]: df=pd.read_csv("Housing.csv")

In [4]: df
```

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
0	13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	furnished
1	12250000	8960	4	4	4	yes	no	no	no	yes	3	no	furnished
2	12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	semi-furnished
3	12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished
4	11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	furnished
...
540	1820000	3000	2	1	1	yes	no	yes	no	no	2	no	unfurnished
541	1767150	2400	3	1	1	no	no	no	no	no	0	no	semi-furnished
542	1750000	3620	2	1	1	yes	no	no	no	no	0	no	unfurnished
543	1750000	2910	3	1	1	no	no	no	no	no	0	no	furnished
544	1750000	3850	3	1	2	yes	no	no	no	no	0	no	unfurnished

545 rows x 13 columns

```
In [5]: df.head()
```

```
Out[5]:
```

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
0	13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	furnished
1	12250000	8960	4	4	4	yes	no	no	no	yes	3	no	furnished
2	12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	semi-furnished
3	12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished
4	11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	furnished

```
In [26]: df.drop_duplicates()
```

```
Out[26]:
```

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
0	13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	furnished
1	12250000	8960	4	4	4	yes	no	no	no	yes	3	no	furnished
2	12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	semi-furnished
3	12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished
4	11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	furnished
...
540	1820000	3000	2	1	1	yes	no	yes	no	no	2	no	unfurnished
541	1767150	2400	3	1	1	no	no	no	no	no	0	no	semi-furnished
542	1750000	3620	2	1	1	yes	no	no	no	no	0	no	unfurnished
543	1750000	2910	3	1	1	no	no	no	no	no	0	no	furnished
544	1750000	3850	3	1	2	yes	no	no	no	no	0	no	unfurnished

545 rows x 13 columns

```
In [6]: df.tail()
```

```
Out[6]:
```

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
540	1820000	3000	2	1	1	yes	no	yes	no	no	2	no	unfurnished
541	1767150	2400	3	1	1	no	no	no	no	no	0	no	semi-furnished
542	1750000	3620	2	1	1	yes	no	no	no	no	0	no	unfurnished
543	1750000	2910	3	1	1	no	no	no	no	no	0	no	furnished
544	1750000	3850	3	1	2	yes	no	no	no	no	0	no	unfurnished

```
In [28]: df.shape

Out[28]: (545, 13)
```

```
In [29]: df.isnull
```

```
Out[29]:
```

```
<bound method DataFrame.isnull of
0    13300000    7420    4    2    3    yes    no    no    no    yes    2    yes    furnished
1    12250000    8960    4    4    4    yes    no    no    no    yes    3    no    furnished
2    12250000    9960    3    2    2    yes    no    yes    no    no    2    yes    semi-furnished
3    12215000    7500    4    2    2    yes    no    yes    no    yes    3    yes    furnished
4    11410000    7420    4    1    2    yes    yes    yes    no    yes    2    no    furnished
...
540    1820000    3000    2    1    1    yes    no    yes    no    no    2    no    unfurnished
541    1767150    2400    3    1    1    no    no    no    no    no    0    no    semi-furnished
542    1750000    3620    2    1    1    yes    no    no    no    no    0    no    unfurnished
543    1750000    2910    3    1    1    no    no    no    no    no    0    no    furnished
544    1750000    3850    3    1    2    yes    no    no    no    no    0    no    unfurnished

hotwaterheating    airconditioning    parking    prefarea    furnishingstatus
0                no                yes    2        yes    furnished
1                no                yes    3        no    furnished
2                no                no    2        yes    semi-furnished
3                no                yes    3        yes    furnished
4                no                yes    2        no    furnished
...
540                no                no    2        no    unfurnished
541                no                no    0        no    semi-furnished
542                no                no    0        no    unfurnished
543                no                no    0        no    unfurnished
544                no                no    0        no    unfurnished

[545 rows x 13 columns]>
```

```
In [31]: df.describe()
```

```
Out[31]:
```

	price	area	bedrooms	bathrooms	stories	parking
count	5.450000e+02	545.000000	545.000000	545.000000	545.000000	545.000000
mean	4.766729e+06	5150.541284	2.965138	1.286239	1.805505	0.693578
std	1.870440e+06	2170.141023	0.738064	0.502470	0.867492	0.861586
min	1.750000e+06	1650.000000	1.000000	1.000000	1.000000	0.000000
25%	3.430000e+06	3600.000000	2.000000	1.000000	1.000000	0.000000
50%	4.340000e+06	4600.000000	3.000000	1.000000	2.000000	0.000000
75%	5.740000e+06	6360.000000	3.000000	2.000000	2.000000	1.000000
max	1.330000e+07	16200.000000	6.000000	4.000000	4.000000	3.000000

```
In [32]: df.head
```

```
Out[32]:
```

```
<bound method NDFrame.head of
0    13300000    7420    4    2    3    yes    no    no    no    yes    2    yes    furnished
1    12250000    8960    4    4    4    yes    no    no    no    yes    3    no    furnished
2    12250000    9960    3    2    2    yes    no    yes    no    no    2    yes    semi-furnished
3    12215000    7500    4    2    2    yes    no    yes    no    yes    3    yes    furnished
4    11410000    7420    4    1    2    yes    yes    yes    no    yes    2    no    furnished
...
540    1820000    3000    2    1    1    yes    no    yes    no    no    2    no    unfurnished
541    1767150    2400    3    1    1    no    no    no    no    no    0    no    semi-furnished
542    1750000    3620    2    1    1    yes    no    no    no    no    0    no    unfurnished
543    1750000    2910    3    1    1    no    no    no    no    no    0    no    furnished
544    1750000    3850    3    1    2    yes    no    no    no    no    0    no    unfurnished

hotwaterheating    airconditioning    parking    prefarea    furnishingstatus
0                no                yes    2        yes    furnished
1                no                yes    3        no    furnished
2                no                no    2        yes    semi-furnished
3                no                yes    3        yes    furnished
4                no                yes    2        no    furnished
...
540                no                no    2        no    unfurnished
541                no                no    0        no    semi-furnished
542                no                no    0        no    unfurnished
543                no                no    0        no    unfurnished
544                no                no    0        no    unfurnished

[545 rows x 13 columns]>
```

```
In [27]: df.all()
```

```
Out[27]:
```

```
price                True
area                 True
bedrooms             True
bathrooms            True
stories              True
mainroad             True
guestroom            True
basement             True
hotwaterheating      True
airconditioning      True
parking              False
prefarea             True
furnishingstatus     True
dtype: bool
```

```
In [11]: df.drop_duplicates()
```

```
Out[11]:
```

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
0	13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	furnished
1	12250000	8960	4	4	4	yes	no	no	no	yes	3	no	furnished
2	12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	semi-furnished
3	12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished
4	11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	furnished
...
540	1820000	3000	2	1	1	yes	no	yes	no	no	2	no	unfurnished
541	1767150	2400	3	1	1	no	no	no	no	no	0	no	semi-furnished
542	1750000	3620	2	1	1	yes	no	no	no	no	0	no	unfurnished
543	1750000	2910	3	1	1	no	no	no	no	no	0	no	furnished
544	1750000	3850	3	1	2	yes	no	no	no	no	0	no	unfurnished

545 rows x 13 columns

```
In [44]: df.nunique()
```

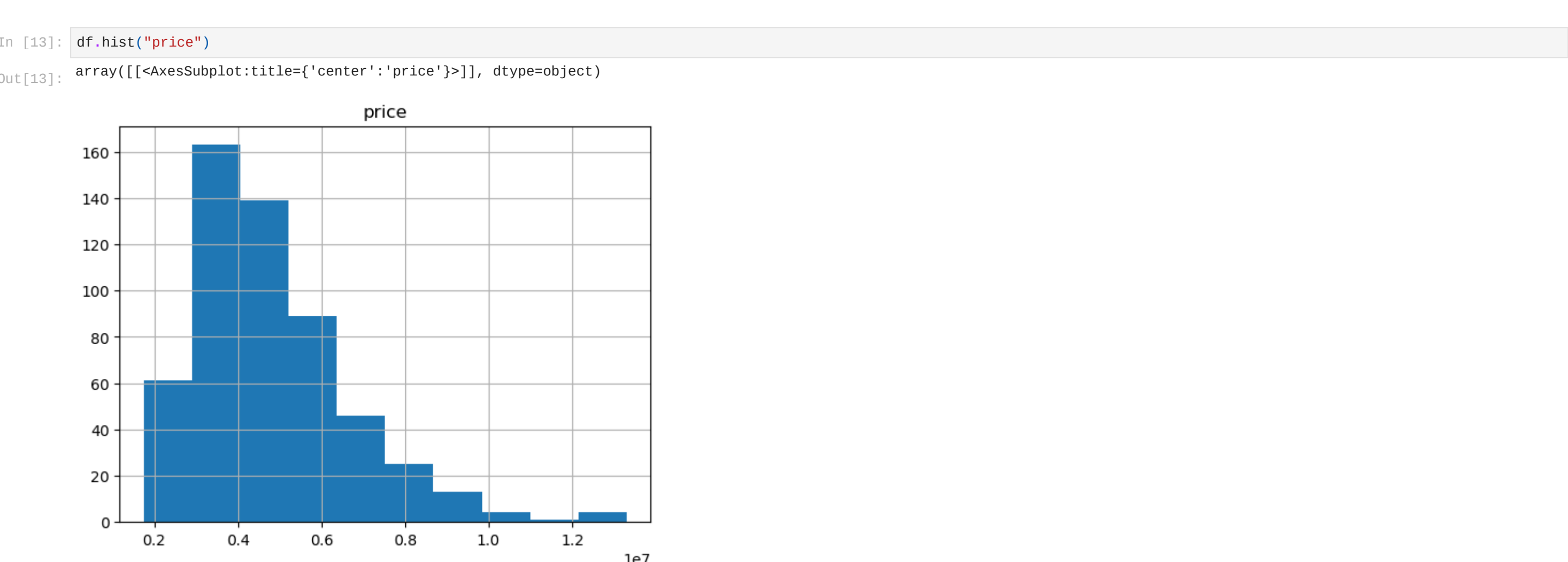
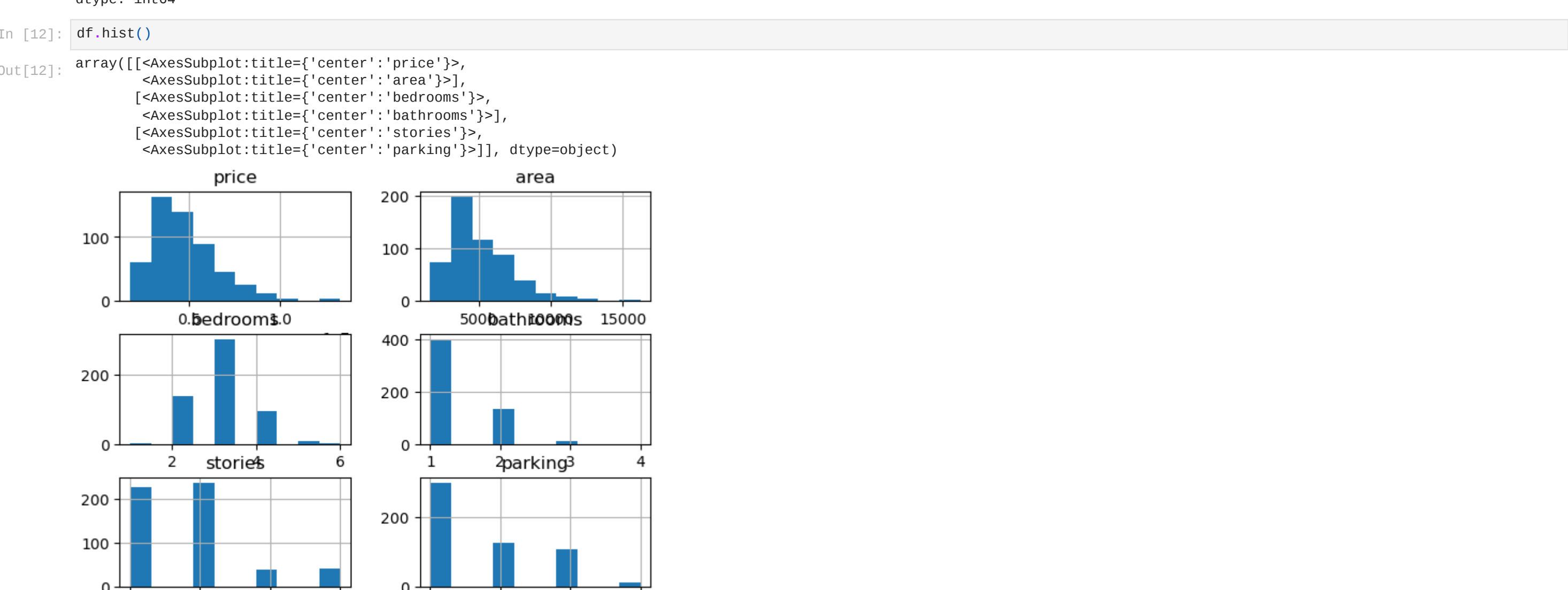
```
Out[44]:
```

```
price                219
area                 284
bedrooms              6
bathrooms             4
stories               4
mainroad              2
guestroom             2
basement              2
hotwaterheating       2
airconditioning       2
parking               4
prefarea              2
furnishingstatus      3
dtype: int64
```

```
In [43]: df.count()
```

```
Out[43]:
```

```
price                545
area                 545
bedrooms             545
bathrooms            545
stories              545
mainroad             545
guestroom            545
basement             545
hotwaterheating      545
airconditioning      545
parking              545
prefarea             545
furnishingstatus     545
dtype: int64
```



```
In [ ]:
```