1-How to Import Libraries?

In [1]:
```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
```

2-How we Read the Data?

In [2]:
```
game = pd.read_csv("android-games.csv")
```

3-How we Call the Data?

In [3]:
```
game.head(10)
```

Out[3]:

	rank	title	total ratings	installs	average rating	growth (30 days)	growth (60 days)	price	category	5 star ratings	4 star ratings
0	1	Garena Free Fire- World Series	86273129	500.0 M	4	2.1	6.9	0.0	GAME ACTION	63546766	4949507
1	2	PUBG MOBILE - Traverse	37276732	500.0 M	4	1.8	3.6	0.0	GAME ACTION	28339753	2164478
2	3	Mobile Legends: Bang Bang	26663595	100.0 M	4	1.5	3.2	0.0	GAME ACTION	18777988	1812094
3	4	Brawl Stars	17971552	100.0 M	4	1.4	4.4	0.0	GAME ACTION	13018610	1552950
4	5	Sniper 3D: Fun Free Online FPS Shooting Game	14464235	500.0 M	4	0.8	1.5	0.0	GAME ACTION	9827328	2124154
5	6	Call of Duty®: Mobile - Season 4: Spurned & Bu...	13572148	100.0 M	4	2.0	4.0	0.0	GAME ACTION	10501443	1274162
6	7	Among Us	11936964	100.0 M	3	1.8	5.6	0.0	GAME ACTION	5954262	1041297
7	8	Temple Run 2	9633929	500.0 M	4	0.3	0.8	0.0	GAME ACTION	6579369	991341
8	9	PUBG MOBILE LITE	7578630	100.0 M	4	1.0	2.5	0.0	GAME ACTION	5382545	500696
9	10	Gangstar Vegas: World of Crime	6268377	100.0 M	4	0.4	1.0	0.0	GAME ACTION	4509647	605510

4-Define the shape of Data?

In [4]:
```
game.shape
```

Out[4]: (1730, 15)

5-How to get whole data Information?

In [5]:
```
game.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1730 entries, 0 to 1729
Data columns (total 15 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   rank              1730 non-null   int64
 1   title             1730 non-null   object
 2   total ratings     1730 non-null   int64
 3   installs          1730 non-null   object
 4   average rating    1730 non-null   int64
 5   growth (30 days)  1730 non-null   float64
 6   growth (60 days)  1730 non-null   float64
 7   price             1730 non-null   float64
 8   category          1730 non-null   object
 9   5 star ratings    1730 non-null   int64
 10  4 star ratings    1730 non-null   int64
 11  3 star ratings    1730 non-null   int64
 12  2 star ratings    1730 non-null   int64
 13  1 star ratings    1730 non-null   int64
 14  paid              1730 non-null   bool
dtypes: bool(1), float64(3), int64(8), object(3)
memory usage: 191.0+ KB
```

6-How to describe Data?

In [6]:
```
game.describe()
```

Out[6]:

	rank	total ratings	average rating	growth (30 days)	growth (60 days)	price	5 star ratin
count	1730.000000	1.730000e+03	1730.000000	1730.000000	1730.000000	1730.000000	1.730000e+0
mean	50.386705	1.064332e+06	3.908092	321.735896	122.554971	0.010942	7.622315e+0
std	28.936742	3.429250e+06	0.290973	6018.914507	2253.891703	0.214987	2.538658e+0
min	1.000000	3.299300e+04	2.000000	0.000000	0.000000	0.000000	1.397500e+0
25%	25.000000	1.759992e+05	4.000000	0.100000	0.200000	0.000000	1.277300e+0
50%	50.000000	4.286065e+05	4.000000	0.500000	1.000000	0.000000	2.964340e+0
75%	75.000000	8.837970e+05	4.000000	1.700000	3.300000	0.000000	6.198358e+0
max	100.000000	8.627313e+07	4.000000	227105.700000	69441.400000	7.490000	6.354677e+0

7-How many number of unique values in a Data?

In [7]:
```
game.nunique()
```

Out[7]:
```
rank                 100
title               1675
total ratings       1699
installs               9
average rating         3
growth (30 days)     186
growth (60 days)     206
price                  6
category              17
5 star ratings      1697
4 star ratings      1692
3 star ratings      1677
2 star ratings      1653
1 star ratings      1686
paid                   2
dtype: int64
```

8-How we do Data cleaning by using pandas and numpy?

Data is already cleaned so there is no need for data cleaning.

9-Count Plot

NOTE: In count plot there is only one numerical X variable we use.

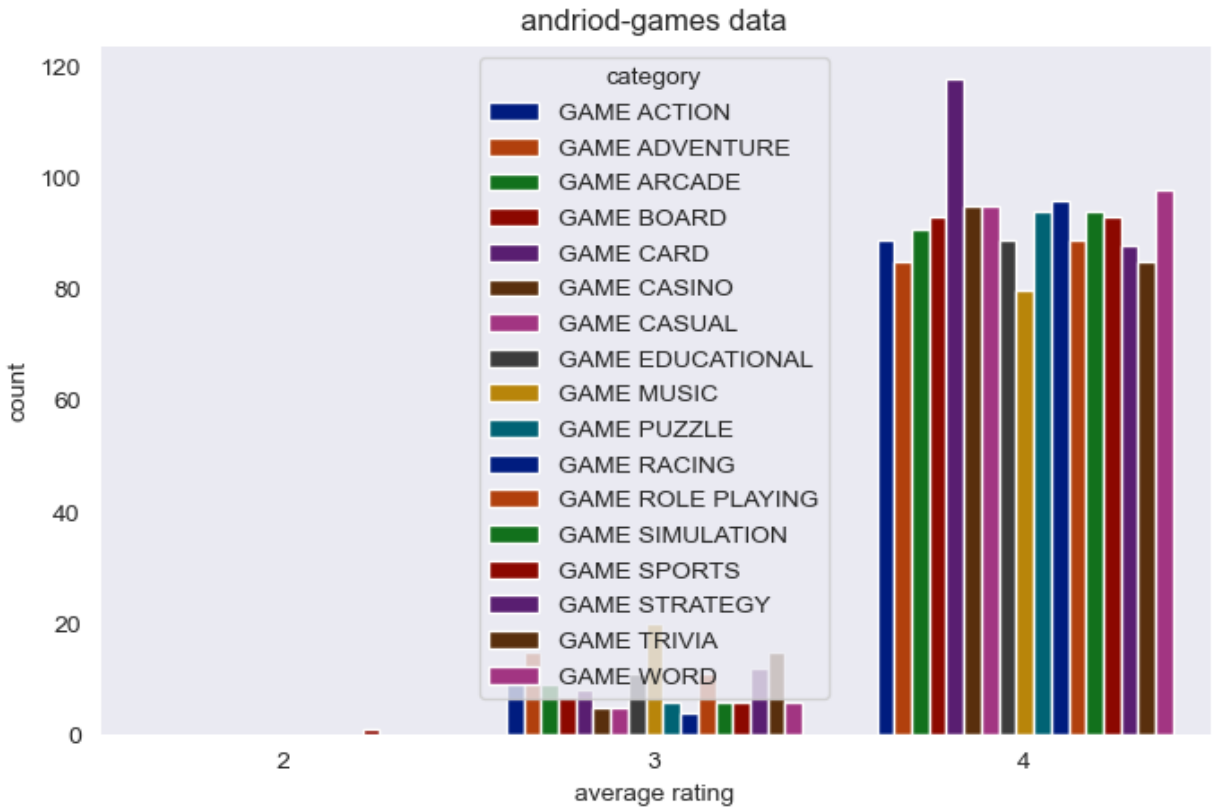
In [3]:
```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

sns.set_style("dark")

game = pd.read_csv("android-games.csv")

plt.figure(figsize=(8,5))

p=sns.countplot(x="average rating", data=game, hue="category", saturation=4, palette='
plt.title("andriod-games data")
plt.show()
```



10-Box/Boxen Plot

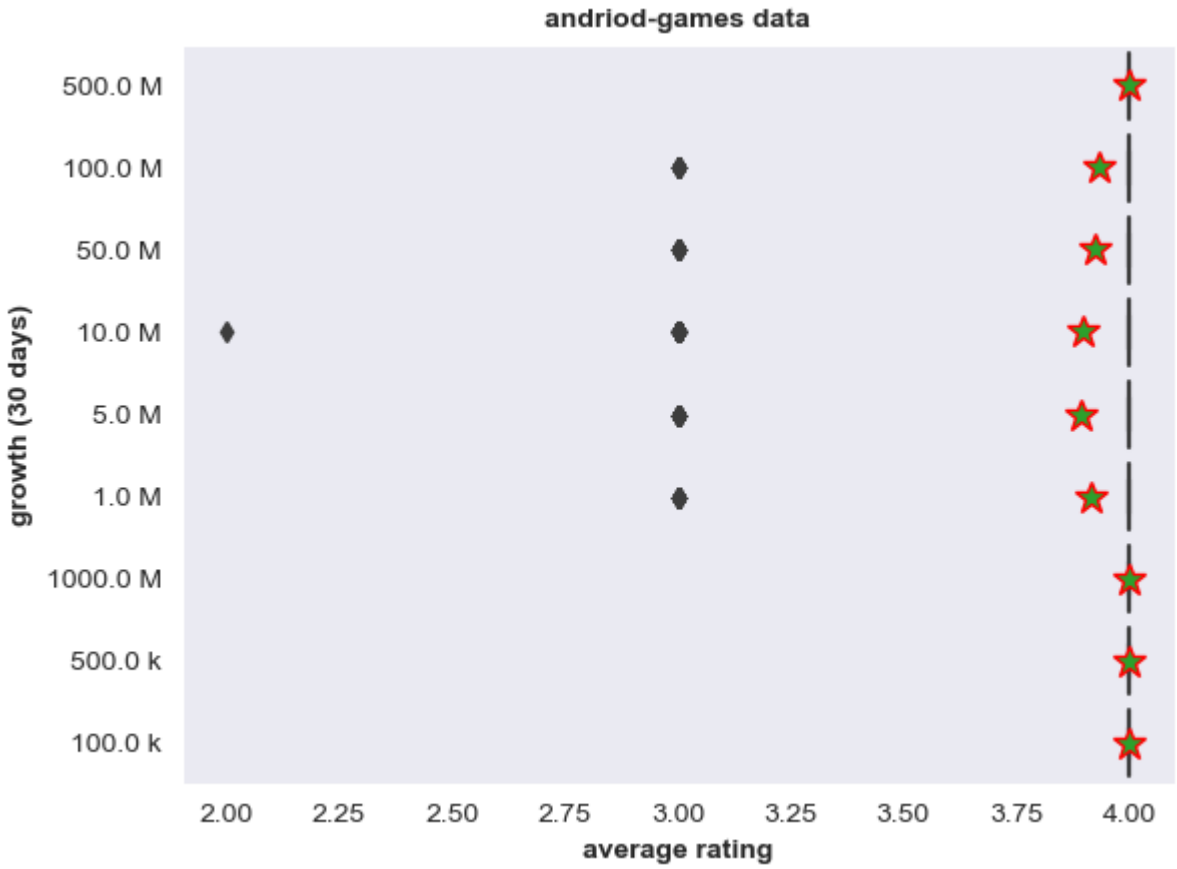
NOTE: In box & Boxen plot we use two X & Y variable may contain one numeric and one cetagorical data. (Quartile, outliyers, median, positive skewed, negative skewed)

In [32]:
```
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

sns.set_style("dark")

game = pd.read_csv("android-games.csv")

sns.boxplot(x="average rating", y="installs", data=game, showmeans=True, meanprops= {'
plt.xlabel("average rating", size=10, weight='bold')
plt.ylabel("growth (30 days)", size=10, weight='bold')
plt.title("andriod-games data", size=10, weight='bold')
plt.show()
```



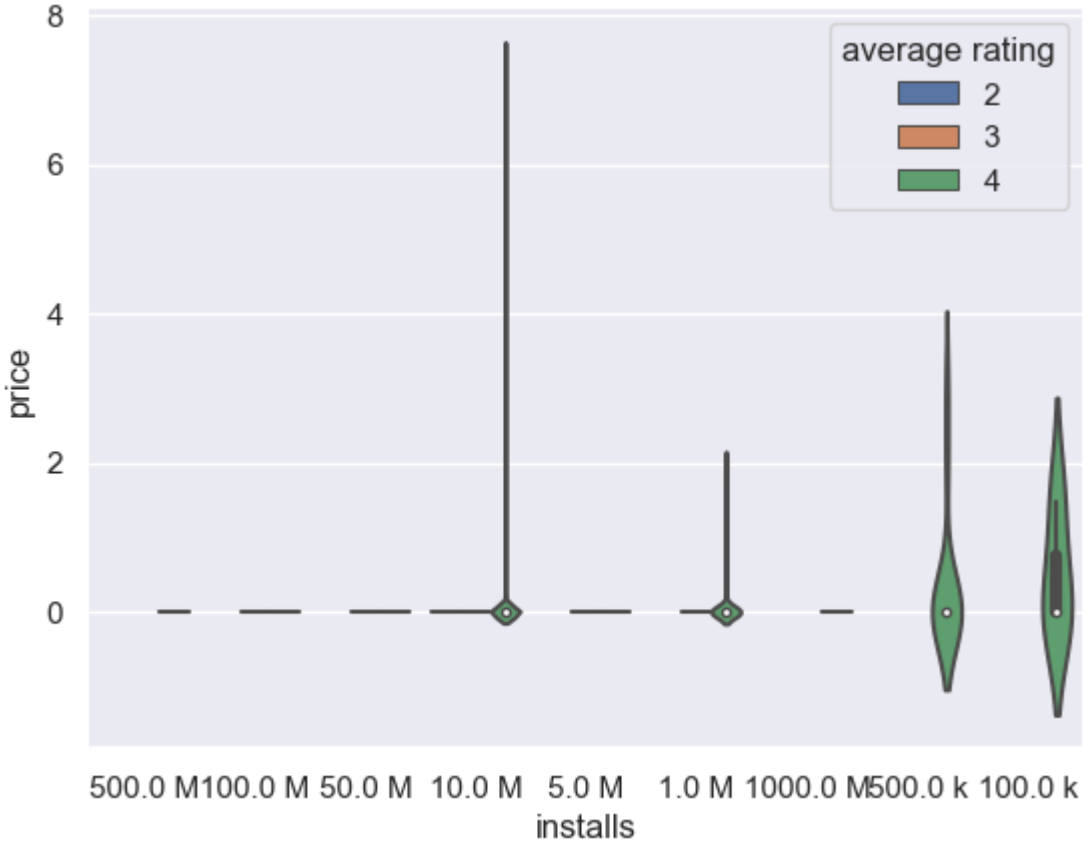
11-Vilion Plot

NOTE: In vilion plot we use two X & Y variable may contain one numeric and one cetagorical data. (data spreading)

In [43]:
```
import seaborn as sns
sns.set_theme(style="darkgrid")

game = pd.read_csv("android-games.csv")


sns.violinplot(data=game, x="installs", y="price", hue="average rating")
#                split=True, linewidth=1, palette={"Yes": "b", "No": ".85"})
sns.despine(left=True)
```


average rating
2
3
4
price
8
6
4
2
0
500.0 M
100.0 M
50.0 M
10.0 M
5.0 M
1.0 M
1000.0 M
500.0 k
100.0 k
installs