

▼ GROUP 4

- Nguyen Duc Dung : 20160674
- Tran Duc Phung : 20166589
- Vu Thanh Tung : 20164571

Question 4. Check the distribution of top platforms for games

- load .csv file to obtain data
- delete rows which have NULL value at 'title' or 'platform' column
- group by platform then select top 10
- use pie chart to show the platform distribution

```
from google.colab import files
```

```
uploaded = files.upload()
```

```
for fn in uploaded.keys():
```

```
    print('User uploaded file "{name}" with length {length} bytes'.format(  
        name=fn, length=len(uploaded[fn])))
```



Choose Files ign.csv

- **ign.csv**(text/csv) - 2032739 bytes, last modified: 11/5/2019 - 100% done
Saving ign.csv to ign (1).csv
User uploaded file "ign.csv" with length 2032739 bytes

```
import pandas as pd
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
dataset = pd.read_csv('ign.csv', sep=';')
```

```
dataset.head()
```

```
# dataset.shape
```



	Unnamed: 0	score_phrase	title	url	platform	score	
0	0	Amazing	LittleBigPlanet PS Vita	/games/littlebigplanet-vita/vita-98907	PlayStation Vita	9.0	Plai
1	1	Amazing	LittleBigPlanet PS Vita -- Marvel Super Hero E...	/games/littlebigplanet-ps-vita-marvel-super-he...	PlayStation Vita	9.0	Plai
2	2	Great	Splice: Tree of Life	/games/splice/ipad-141070	iPad	8.5	
3	3	Great	NHL 13	/games/nhl-13/xbox-360-128182	Xbox 360	8.5	
4	4	Great	NHL 13	/games/nhl-13/ps3-128181	PlayStation 3	8.5	

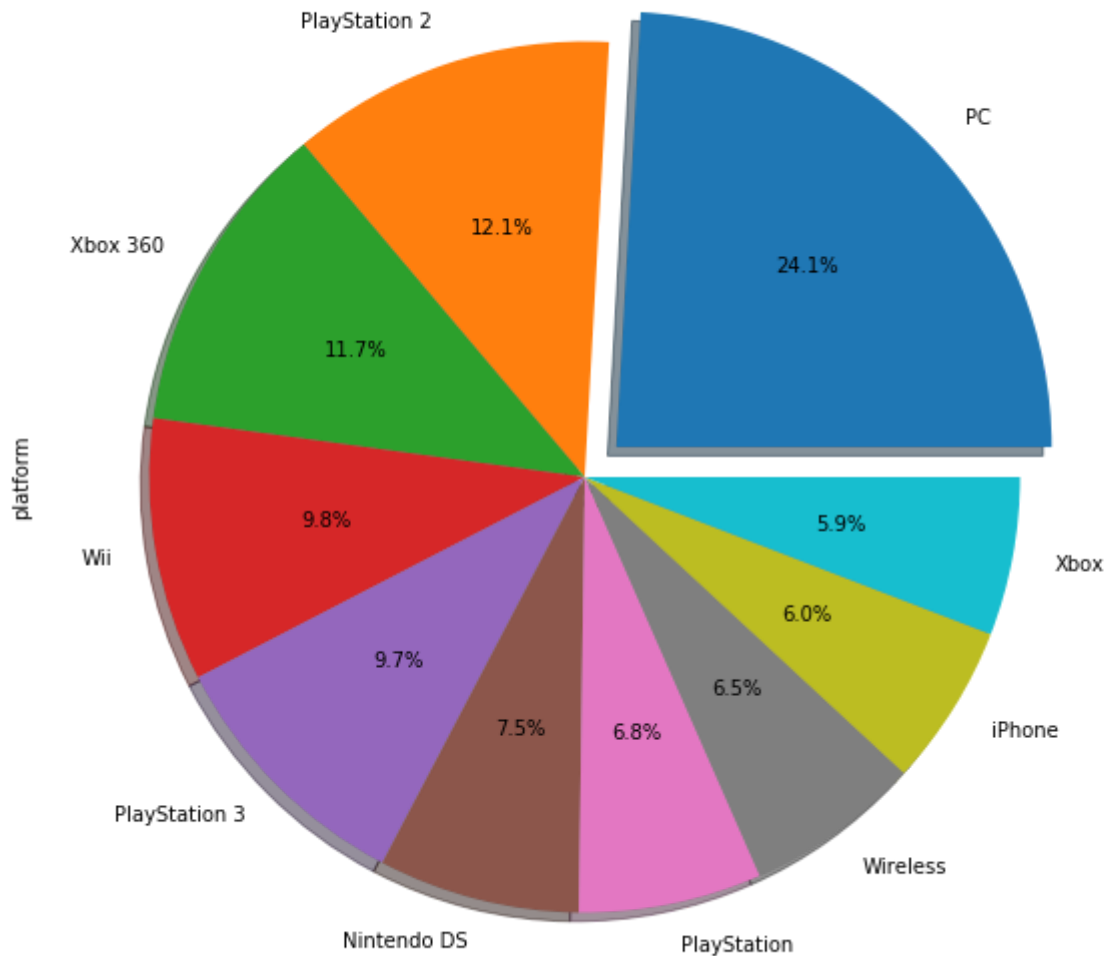
```
# clean data
```

```
platform_dataset = dataset.dropna(how='any', subset=['title', 'platform'])
platform_dataset.shape
```

```
(18624, 11)
```

```
platforms = platform_dataset['platform'].value_counts()[:10]
plt.figure(figsize=(10,10))
platforms.plot(kind='pie', autopct='%1.1f%%', shadow=True, explode=[0.1,0,0,0,0,0,0,0,0,0])
plt.title('Figure 1: Distribution of Top 10 Platform', fontsize = 20, )
plt.legend(loc = 'right', bbox_to_anchor=[1,0,0.5,1])
plt.show()
```

Figure 1: Distribution of Top 10 Platform



As we can see from the chart, PC is the most popular platform with the proportion is 24.1%. The for Xbox 360 (12.1% and 11.7% respectively). Meanwhile, Xbox and iPhone are the least well-known p

Question 16. Check the score distribution of game considering the score phrase is Amazing

- Delete rows which have NULL value at 'score' or 'score_phrase' column
- Select games whose 'score_phrase' attribute = 'Amazing'
- Group by score
- Use histogram to show the frequency

```
score_dataset = dataset.dropna(how='any', subset=['score_phrase', 'score'])
score_dataset.shape
```

```
(18624, 11)
```

```
amazingGames = score_dataset[score_dataset['score_phrase'] == 'Amazing']
amazingGames.shape
```

```
(18624, 11)
```

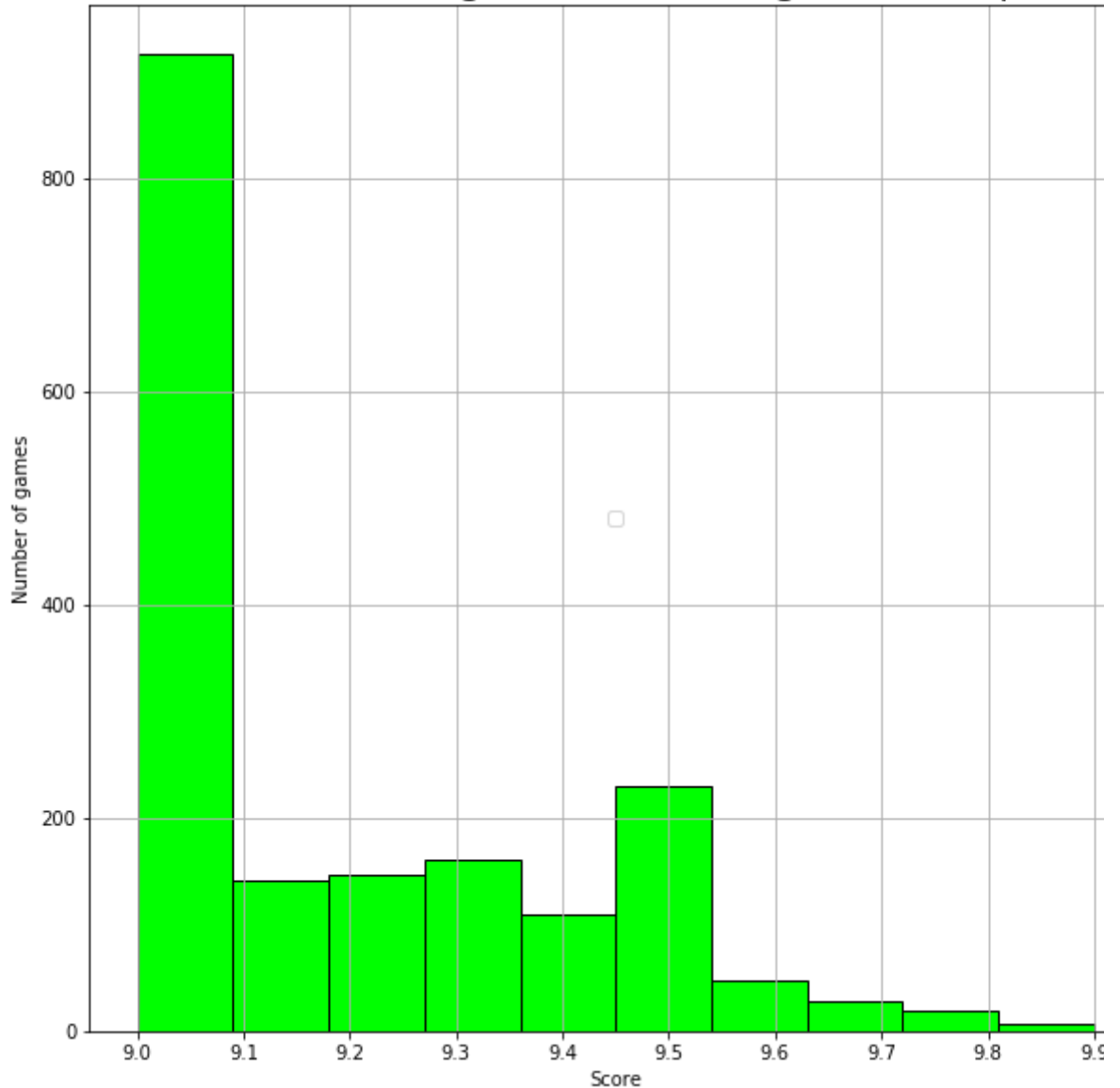
```

(1004, 11)
plt.figure(figsize=(10,10))
amazingGames['score'].hist(color='#00FF00',edgecolor='black')
x = np.arange(9, 10, step=0.1)
plt.xticks(x)
plt.xlabel('Score')
plt.ylabel('Number of games')
plt.title('Figure 2: Score distribution of game considering the score phrase is Am')
plt.legend(loc = 'center')
plt.show()

```

☞ No handles with labels found to put in legend.

Figure 2: Score distribution of game considering the score phras



According to the chart, the game whose score is 9.0 has the greatest number with over 900 games; quite same quantity, about 150 games. With higher score, the bigger the score is, the fewer the number of games which has over 200 games.