Report

One of the main assumptions made on this report is that the total channel views is the most important factor reflecting channel success, more important than for example amount of subscribers.

Conclusions:

Clear conclusions:

* There are specific video categories that seems to be a lot more popular than others. This conclusion is based on two graphs video category id to amount of subscribers and video category id to amount of channel views.
* As expected the most a channel is popular the less dislikes/views he gets.

Section of this type doesn't generate any new information they just confirm already known knowledge.

Unclear conclusions:

* There is no clear correlation between channel elapsed time and total channel views.
* There is no clear correlation between channel elapsed time and dislike/subscriber.
* There is no clear correlation between comments/views and amount of subscribers.
* There is no clear correlation between comments/views and total channel views.
* There is no clear correlation between channel elapsed time and total channel views.
* There is no clear correlation between likes/views and amount of subscribers or between likes/dislikes and total channel views.

This is contrary to what was expected a successful channel is expected to get a large amount of likes compared to his number of views or compared to his likes/dislikes ratio.

Final conclusion

Although it was expected that the data will bring positive results regarding improving YouTube channels, the only definitive advice that can be given to a YouTube channel owner is that specific few video categories are a lot more popular by a substantial margin compared to the remaining video categories.

The Code used to create the graph using jupyter notebook, panda, numpy and matloplib libraries is pasted below.

import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.bar(sample['videoCategoryId'],sample['subscriberCount'])  
plt.xlabel('videoCategoryId')  
plt.ylabel('subscriberCount')  
plt.grid(axis='y')  
plt.show()  
  
import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.scatter(sample['dislikes/views'],sample['subscriberCount'])  
plt.xlabel('dislikes/views')  
plt.ylabel('subscriberCount')  
plt.grid(axis='y')  
plt.show()  
  
import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.scatter(sample['comments/views'],sample['subscriberCount'])  
plt.xlabel('comments/views')  
plt.ylabel('subscriberCount')  
plt.grid(axis='y')  
plt.show()  
  
import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.scatter(sample['likes/views'],sample['subscriberCount'])  
plt.xlabel('likes/views')  
plt.ylabel('subscriberCount')  
plt.grid(axis='y')  
plt.show()  
  
import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.scatter(sample['VideoCommentCount'],sample['subscriberCount'])  
plt.xlabel('Videocommentcount')  
plt.ylabel('subscriberCount')  
plt.grid(axis='y')  
plt.show()  
  
import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.bar(sample['videoCategoryId'],sample['channelViewCount'])  
plt.xlabel('videoCategoryId')  
plt.ylabel('channelViewCount')  
plt.grid(axis='y')  
plt.show()  
  
import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.scatter(sample['likes/dislikes'],sample['channelViewCount'])  
plt.xlabel('likes/dislikes')  
plt.ylabel('channelViewCount')  
plt.grid(axis='y')  
plt.show()  
  
import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.scatter(sample['comments/views'],sample['channelViewCount'])  
plt.xlabel('comments/views')  
plt.ylabel('channelViewCount')  
plt.grid(axis='y')  
plt.show()  
  
import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.scatter(sample['channelelapsedtime'],sample['channelViewCount'])  
plt.xlabel('channelelapsedtime')  
plt.ylabel('channelViewCount')  
plt.grid(axis='y')  
plt.show()  
  
import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.scatter(sample['dislikes/subscriber'],sample['channelViewCount'])  
plt.xlabel('dislikes/subscriber')  
plt.ylabel('channelViewCount')  
plt.grid(axis='y')  
plt.show()  
  
import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.scatter(sample['dislikes/subscriber'],sample['channelelapsedtime'])  
plt.xlabel('dislikes/subscriber')  
plt.ylabel('channelelapsedtime')  
plt.grid(axis='y')  
plt.show()  
  
  
  
  
  
  
  
  
  
  
  
  
import pandas as pd  
import numpy as np  
sample=pd.read\_csv('you.csv')  
from matplotlib import pyplot as plt  
  
  
  
plt.scatter(sample['comments/views'],sample['subscriberCount'])  
plt.xlabel('comments/views')  
plt.ylabel('subscriberCount')  
plt.grid(axis='y')  
plt.show()