12.Downloading Scores:

Next we will be splitting our data into english and non english (universal) accounts. We do this as the API has different scores based on whether an account is english or non english. From this we will create 6 data frames being English: Bot, Human, Organization & Universal: Bot, Human, Organization. We will create 7 columns in this data frame which are the ones described earlier in the tutorial (You can name these dataframes whatever you want). We will accomplish this using the pandas library as pd.DataFrame to create the new dataframes we will be using to do visualizations on the various scores assigned by the RapidAPI. It should look something like this (*ADD NEW PICTURE WITH CORRECT VALUES):

```
training_set = pd.read_csv(r'C:\Users\Esau\PycharmProjects\pythonProject1\split_training_set (version 2).csv')
ids = training_set['id'].tolist()

human_eng = pd.DataFrame(columns=["astroturf", "fake follower",
                         "financial", "other", "overall", "self-declared", "spammer"])
bot_eng = pd.DataFrame(columns=["astroturf", "fake follower",
                         "financial", "other", "overall", "self-declared", "spammer"])
org_eng = pd.DataFrame(columns=["astroturf", "fake follower",
                         "financial", "other", "overall", "self-declared", "spammer"])
human_univ = pd.DataFrame(columns=["astroturf", "fake follower",
                         "financial", "other", "overall", "self-declared", "spammer"])
bot_univ = pd.DataFrame(columns=["astroturf", "fake follower",
                         "financial", "other", "overall", "self-declared", "spammer"])
org_univ = pd.DataFrame(columns=["astroturf", "fake follower",
                         "financial", "other", "overall", "self-declared", "spammer"])
```

Next, we need to change the type of value of the dataframes to create histograms of the data. The dataframes currently have the data saved as string value and we want to change it to float values which are used to create histograms. We do this by type casting the dataframes as floats into another variable to house the dataframes. We then merge all the dataframes into one csv file for english and universal using pd.concat to create a hierarchal indexed dataframe. It should look like this:

```
human_eng = human_eng.astype(float)
bot_eng = bot_eng.astype(float)
org_eng = org_eng.astype(float)
human_univ = human_univ.astype(float)
bot_univ = human_univ.astype(float)
org_univ = human_univ.astype(float)

eng_merged = pd.concat([human_eng, bot_eng, org_eng], axis=1, keys=['human', 'bot', 'organization'])
eng_merged.to_csv(r'C:\Users\Esau\PycharmProjects\pythonProject1\english.png', index=False)
univ_merged = pd.concat([human_univ, bot_univ, org_univ], axis=1, keys=['human', 'bot', 'organization'])
univ_merged.to_csv(r'C:\ s\Esau\PycharmProjects\pythonProject1\universal.png', index=False)
```

Finally, we want to format all the data into visible and readable histograms which we will do utilizing the matplotlib histogram method which converts inputted data into a histogram. We want to create a figure for each of the 6 types of data frames we originally created (human_eng, bot_eng, org_eng, human_univ, bot_univ, org_univ) and additionally a histogram for each of the columns we described beforehand being 7. So, we should have 42 images altogether. The code should look like this:

```python
def create_histograms(df, lang, color1=None, color2=None, color3=None):
    human_df = df[df["label"] == 0]
    human_df = human_df.loc[:, "astroturf":"spammer"].astype(float)
    bot_df = df[df["label"] == 1]
    bot_df = bot_df.loc[:, "astroturf":"spammer"].astype(float)
    org_df = df[df["label"] == 2]
    org_df = org_df.loc[:, "astroturf":"spammer"].astype(float)

    human_df.hist(figsize=(15, 12), color=color1)
    plt.tight_layout(rect=[0, 0.03, 1, 0.95])
    plt.suptitle(f'Human ({lang})')
    plt.savefig(r'C:\Users\Rachel\Documents\twitter_tutorial_2022\{}\human_({})'.format(lang, lang))

    bot_df.hist(figsize=(15, 12), color=color2)
    plt.tight_layout(rect=[0, 0.03, 1, 0.95])
    plt.suptitle(f'Bot ({lang})')
    plt.savefig(r'C:\Users\Rachel\Documents\twitter_tutorial_2022\{}\bot_({})'.format(lang, lang))

    org_df.hist(figsize=(15, 12), color=color3)
    plt.tight_layout(rect=[0, 0.03, 1, 0.95])
    plt.suptitle(f'Organization ({lang})')
    plt.savefig(r'C:\Users\Rachel\Documents\twitter_tutorial_2022\{}\org_({})'.format(lang, lang))

    plt.show()
```

Now we will actually call the methods/functions we created and utilize them to create our histograms. First we will take the input csv and split it up into chunks. Like this:

```python
            plt.show()



    input_csv = r'C:\Users\Rachel\Documents\twitter_tutorial_2022\marked_twitter_id2.csv'
    chunks_eng = [r'C:\Users\Rachel\Documents\twitter_tutorial_2022\english\english{}.csv'.format(i) for i in range(4)]
    chunks_univ = [r'C:\Users\Rachel\Documents\twitter_tutorial_2022\universal\universal{}.csv'.format(i) for i in range(4)]
    output_eng = r'C:\Users\Rachel\Documents\twitter_tutorial_2022\english'
    output_univ = r'C:\Users\Rachel\Documents\twitter_tutorial_2022\universal'
```

Then we will use the create_csv function to create the 4 chunks which if why we had for loop included for the chunks_eng and chunks_univ dfs. It should look like this:

```python
create_csv(r'C:\Users\Rachel\Documents\twitter_tutorial_2022\chunk0.csv', output_eng, output_univ)
create_csv(r'C:\Users\Rachel\Documents\twitter_tutorial_2022\chunk1.csv', output_eng, output_univ)
create_csv(r'C:\Users\Rachel\Documents\twitter_tutorial_2022\chunk2.csv', output_eng, output_univ)
create_csv(r'C:\Users\Rachel\Documents\twitter_tutorial_2022\chunk3.csv', output_eng, output_univ)
```

After we can utilize the merge_csv function to combine them all then create our histograms using our combined csvs. We will have two combined files that being english and universal and the create_histrogram function will be called twice for eng_merged and univ_merged It should look like this:

```python
eng_merged = merge_files(chunks_eng)
univ_merged = merge_files(chunks_univ)

create_histograms(eng_merged, "english", "pink", "darkmagenta", "palevioletred")
create_histograms(univ_merged, "universal", "orchid", "plum", "mediumvioletred")
```

The first parameter takes in the csv file the second applies either the English or universal to the header historgram title, and the rest are color options for easier visualizations. You don't have to use the same colors as you can pick valid colors supported by matplotlib The final histograms should looks something like this:

Bot (universal)