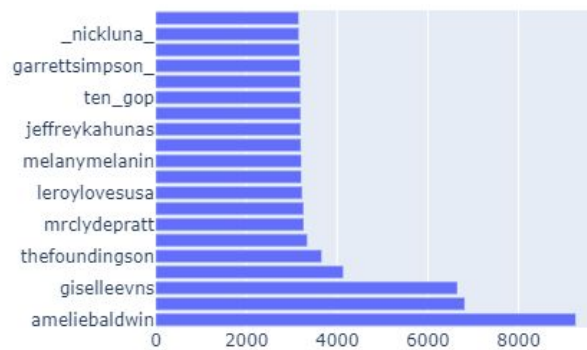


Task 5 Short Report

Tim Mathis, Tanner Abernathy, Evan Branton, Akash Patel

The code that was used in the Twitter Visualization Activity 1 was then modified by changing and declaring different variables. This produced the code below which was used for the plotting of top ten user id or user key with a max retweet count, taken from the tweets.csv. There was not much difficulty in changing the variables because the code from Activity 1 did the same thing that this part was asking for. There was not much teamwork needed for the activity.

Accounts with most tweets



Task 5 Code

```
twts = pd.DataFrame(twitter_tweets.user_key.str.split(' ', 1).tolist(),
                    columns = ['users'])

tweets = twts.groupby('users', as_index=False).size().reset_index(name='counts')

tweets = tweets.sort_values('counts', ascending=False).head(20)

# bar plot

#tweets
top_tweets = twts.groupby('users', as_index=False).size().reset_index(name='counts')
top_tweets = top_tweets.sort_values('counts', ascending=False).head(20)
df4 = go.Bar(
    x=top_tweets['counts'],
    y=top_tweets['users'],
    orientation = 'h',
    name = 'Top Tweets by user'
)

fig = tools.make_subplots(rows=1, cols=1, subplot_titles=(""))
fig.append_trace(df4, 1, 1)
fig['layout'].update(height=400, width=500, title='Accounts with most tweets')
plot(fig, filename = 'top-line')
```