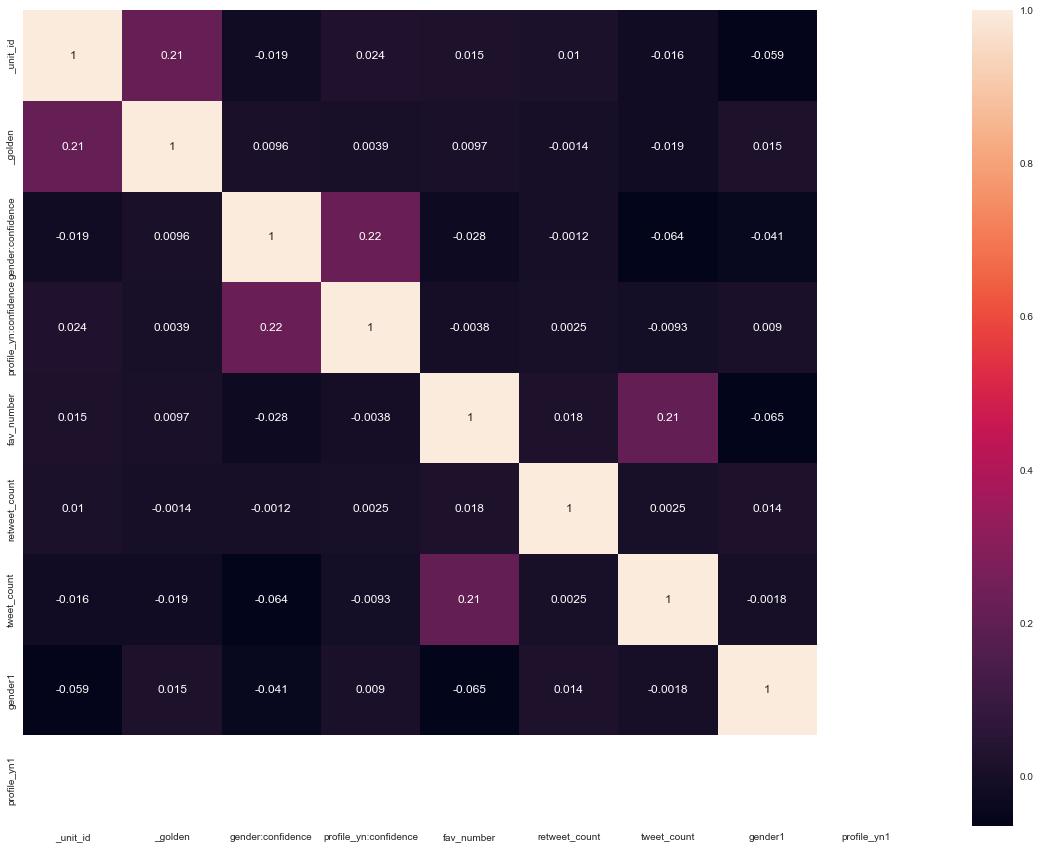
***MPV of the Project***

***study on Twitter***

* We note that there is a relationship between fame and documented accounts.
* We note that there is a relationship between the number of tweets and the number of followers.
* We note the accounts can be brand, male or female.
* I noticed the abundance of brands on Twitter.
* At the end of the project, we can show a histogram of some of the most common words.

 **Code start for the project:**

**Step 0: Imports**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import nltk

from nltk.tokenize import word\_tokenize

from nltk.corpus import stopwords

import string

# step 1 : Read a data

df = pd.read\_csv('data2.csv')

df.head()

df.info()

df.shape

df =df.drop(['gender\_gold','profile\_yn\_gold','tweet\_coord','tweet\_id','tweet\_location','user\_timezone'],axis = 1)

df.head()

df.shape

df['gender'].unique()

df = df[df['gender'] != 'nan']

df.info()

df = df.dropna(subset = ['gender'])

df.info()

df = df[df['gender'] != 'unknown']

df['gender'].unique()

df.shape

df = df.drop(['\_unit\_state', 'created', 'tweet\_created', 'sidebar\_color', 'link\_color', 'profileimage', '\_last\_judgment\_at', '\_trusted\_judgments'],axis = 1)

df.head()

df.shape