

[Open in app ↗](#)

Search



Write



◆ Get unlimited access to the best of Medium for less than \$1/week. [Become a member](#)



# Analyzing LinkedIn Content: A practical application of Data Science

Emmanuel Acheampong · [Follow](#)

5 min read · May 16, 2020



9



...



“Without data  
you’re just  
another person  
with an opinion.”

- W. Edwards Deming,  
Data Scientist

Unlike most people nowadays, I had no idea about it until I was introduced to the field as an intern at Entrepreneurs RoundTable Accelerator and PCB:NG, the summer of 2017. My amazing bosses, Jonathan and Peter, were very interested in my personal and professional growth and wanted to make sure I was very much a part of the team. Unfortunately, I was unfamiliar with almost all of the core technology stack of the company and since *Python* was the only language out of the stack I knew, I was given a data analysis project and since then I developed an interest in data science and practically wrote all my *Python* code in Jupyter Notebooks.

My exploration of *data science* continued after my internship. I began spending the majority of my time learning MOOCs on *data science* and *machine learning*. I became so encapsulated in *data science* and *machine learning* that my social media feed, Google search history, YouTube video recommendations all became about *data science* and *machine learning* at a point.

At that time, I had begun creating content on LinkedIn and seeing a lot of advice on how to increase engagement on LinkedIn. For example, some people stated the best times to post to be on Tuesdays at 11 am. Some people said things like ‘posts with pictures got the most engagement.’ Others stated that tagging famous people in posts got a lot more views and creating video content was better than articles. As much as these pieces of advice were coming from a good place, none of it seemed to be based on data but more based on opinions and experience. And as a person who is data-driven, I decided to work on a project that analyzed the data of my LinkedIn content in order to provide insights.

## Project: Data Analysis of LinkedIn content

### 1. Data Acquisition

## 2. Exploratory Analysis

### 3. Data Cleaning

### 4. Data Analysis

### 5. Data Visualization

### 6. Action

### 7. Results

## Data Acquisition

This was the most important phase of the project. In order for this project to be successful, I needed data currency. Thankfully, I had been consistently producing content and so there was about more than a year's worth of data to analyze.

After clarifying my data source, it was necessary to mine that source. I could have written a *Python* script to scrape my LinkedIn content data but after careful research, I realized I could use Phantom Buster which is an online social media scraping tool. I really like Phantom Buster because it has a lot of interesting APIs for social media particularly LinkedIn which I find very interesting. I used the LinkedIn Activities Extractor API to get the data of my posts. Since I had a lot of content, the process took a while to complete.

## Exploratory Analysis

After acquiring my data set in a '.csv' file, I took a first glimpse of what my data looked like in order to figure out what analysis could be performed.

The dataset had 8 columns and 253 rows. The columns were:

- *profileUrl* — which was my LinkedIn web address where the posts were scrapped from
- *action* — which stated whether a post was made on my profile or in a group
- *postContent* — which was literally the content of each of my post down to even the spacing and punctuation.
- *likeCount* — which indicated the number of likes each post garnered.
- *commentCount* — which stated the number of comments on each post.
- *postDate* — which stated how long the post has lasted and not necessarily the date of the post.
- *timestamp* — stated the time each post was scrapped.
- *postUrl* — which was the web address of the LinkedIn post

## Data Cleaning

In a bid to clean up the data to make the analysis more efficient, I did the following:

- The exploration of the data made me aware that there were columns that weren't necessary to provide insights. As a result, I deleted the *postUrl*, *timestamp*, and *profileUrl* columns in order to make the data cleaner.
- In addition, I converted the data in the *likeCount* and *commentCount* columns from strings to integers so that they could be summed, averaged and aggregated.

## Data Analysis

After exploring and cleaning up the data, I used the standard *Python Pandas* library to perform my initial analysis. Below are the stats obtained:

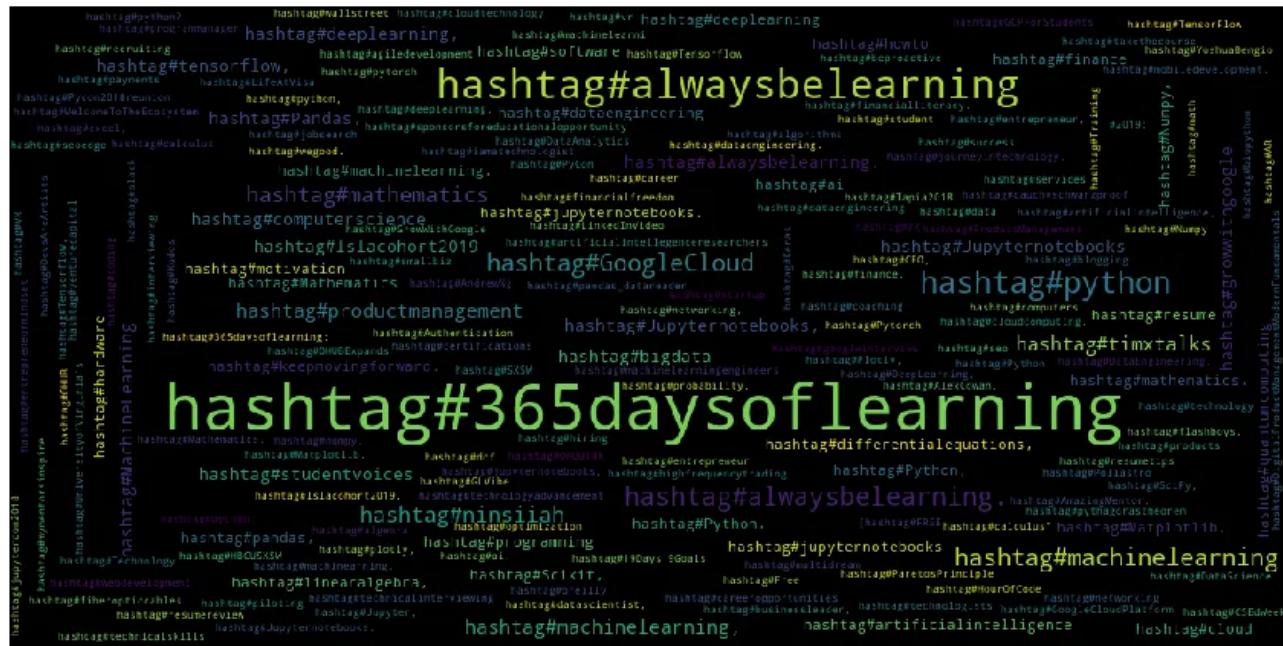
- The total number of posts = 252
- The total number of likes on posts = 8340
- The total number of comments on posts = 727
- The average number of likes on posts = 33
- The average number of comments on posts = 2
- The highest number of likes for a particular post = 513
- The highest number of comments for a particular post = 51

A closer look at these stats made me want to analyze the texts used in my posts and so I began to delve into the content and utilized a bit of NLP to do the analysis. Below are the stats of the content of the texts:

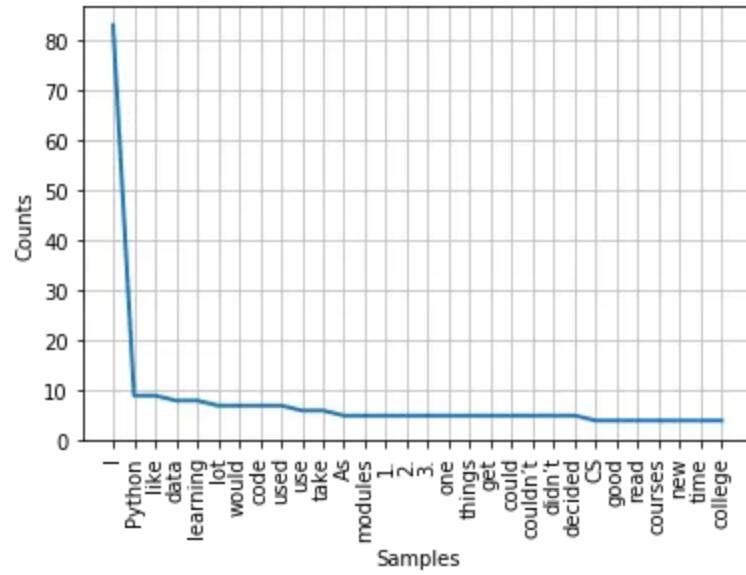
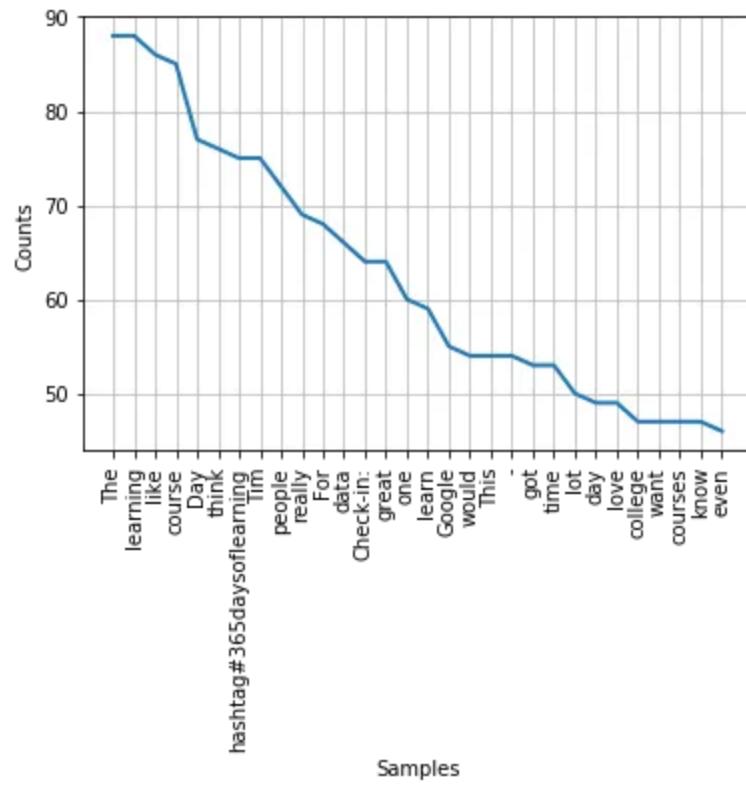
- The number of likes on posts that contain *#365daysoflearning* = 1820, which is about 1/8th of all the likes on my posts
- The highest number of likes on a post that contains *#365daysoflearning* = 153
- The highest number of comments on a post that contains *#365daysoflearning* = 14
- The number of posts that contain the word *python* = 37
- The number of posts that contain the word *data* = 41
- Most used word = 'I'

# Data Visualization

Word Cloud of most used hashtags in my posts



Word Cloud of most used words.



## Actions

These insights helped me make the decision to keep creating consistent content particularly posts that are related to technology which were the ones that received the most engagement.

The insights from this analysis also allowed me to refine the content I created allowing me to focus on telling the story of my journey in tech.

I began to also experiment more with video content which was the form that received the most engagement

## Results

The actions put in place after the insights increased my LinkedIn connections from around 3000 connections to 4185 currently and growing. This insights also made me aware of a need of people who love online learning which led me to create MOOC Anonymous group on LinkedIn.

## Video Presentation of the Jupyter Notebook of the project.



In conclusion, if you're new to data science, gathering your own data and getting insights from it could be an excellent way to explore the skills you're

learning and demonstrate your knowledge to potential employers

You can check out the source code below.

<https://github.com/acheamponge/Data-Analysis-of-LinkedIn-Posts>

Machine Learning

Data Science

Data Analysis

Python

Jupyter Notebook



Written by **Emmanuel Acheampong**

30 Followers

Follow



ML Engineer | Python Developer | AI Hobbyist | 3x Google Cloud Certified | Azure Certified  
AI Engineer

---

More from Emmanuel Acheampong



 Emmanuel Acheampong

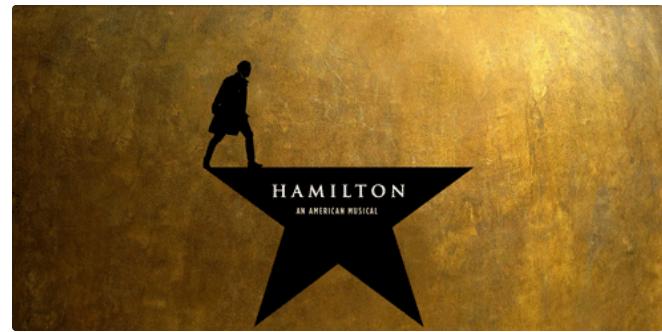
## 100 Awesome LinkedIn Learning Courses You Can Take.

I have curated a list of my top 100 courses I think everyone can take for a competitive...

8 min read · Mar 4, 2019



28



 Emmanuel Acheampong in Chatbots Life

## Using Google Cloud NLP API to analyze writings featured in the...

A Google Cloud Natural Language Processing project focused on the writings featured in t...

7 min read · Aug 10, 2020



62



 Emmanuel Acheampong

## Move Slow To Learn Fast—What I learned about coding at a skating...

Sometimes the greatest epiphanies show up in the most unlikely of places. This is what...

3 min read · May 16, 2020

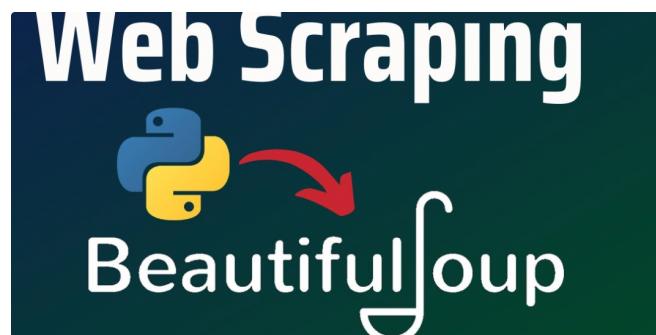


13



[See all from Emmanuel Acheampong](#)

## Recommended from Medium

 Gaby A

### Getting Started with Web Scraping Amazon Reviews with...

Python web scraping introduction: How to extract Amazon Reviews using Python?

7 min read · Jul 24

 103

...

 Rohit

### Top 5 Project Ideas for Beginners in Data Science

Learn data science by doing these five fun and instructive projects that cover different...

8 min read · Jul 23

 95

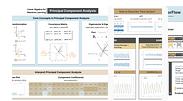
...

## Lists



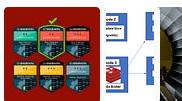
### Predictive Modeling w/ Python

20 stories · 650 saves



### Practical Guides to Machine Learning

10 stories · 729 saves



## New\_Reading\_List

174 stories · 209 saves



## Coding & Development

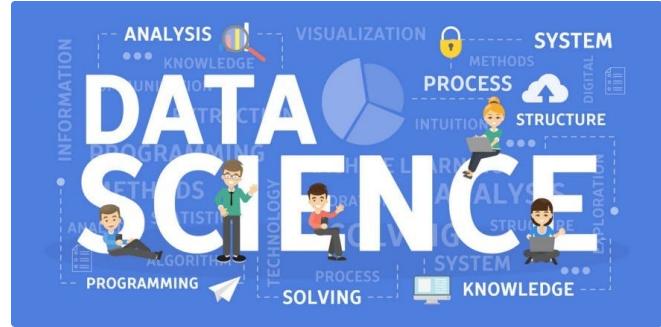
11 stories · 290 saves


 Memmy

## Exciting Web Scraping Projects

Web scraping is an essential technique in the world of data extraction and analysis. It...

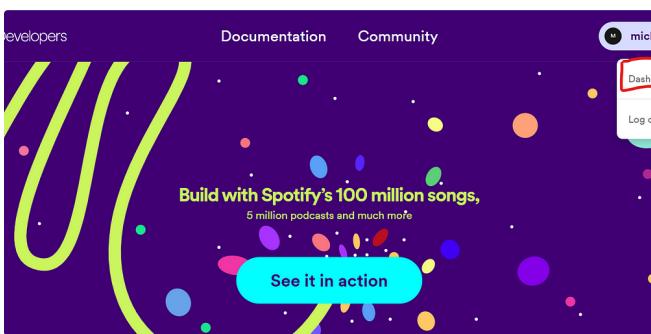
3 min read · Aug 4

 53

 Tony Sebastian

## Data Science Road Map 2023-2024

Data science involves utilizing domain knowledge, programming skills,...

6 min read · Jul 8


 Michaelmiller

## Extracting Song Data from Spotify Using Spotipy

Ever wanted to gain access to song data on Spotify? This will walk you through how to us...

4 min read · Jun 28

```
> twscrape tweet_details 1674894268912087040 | jq
{
  "id": 1674894268912087000,
  "id_str": "1674894268912087040",
  "url": "https://twitter.com/elonmusk/status/1674894268912087040",
  "date": "2023-06-30 21:34:46+00:00",
  "user": {
    "id": 44196397,
    "id_str": "44196397",
    "url": "https://twitter.com/elonmusk",
    "username": "elonmusk",
    "displayname": "Elon Musk",
    "created": "2009-06-02 20:12:29+00:00",
    "followersCount": 145969587,
    ...
    "_type": "snscraper.modules.twitter.User"
  },
  ...
}
```

 vladkens

## How to still scrape millions of tweets in 2023 using twscrape

Twitter is a great place to gather data and assess various trends. Many analytics teams...

4 min read · Jul 1



•••



•••

---

See more recommendations