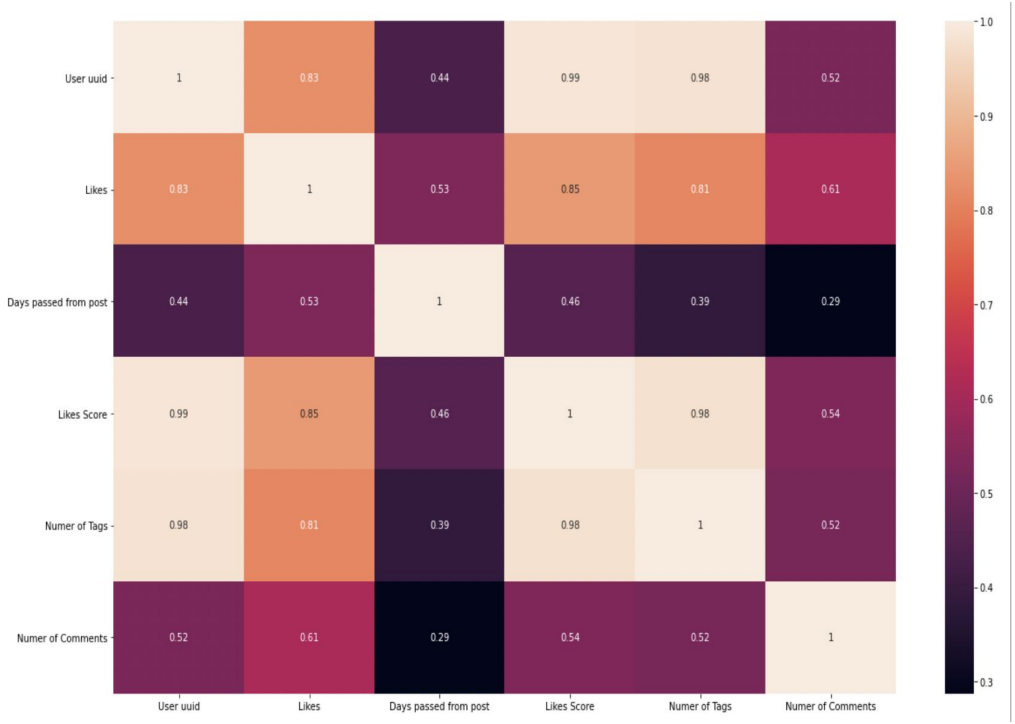
Instagram User Likes Data

Mattie Gittings

The problem

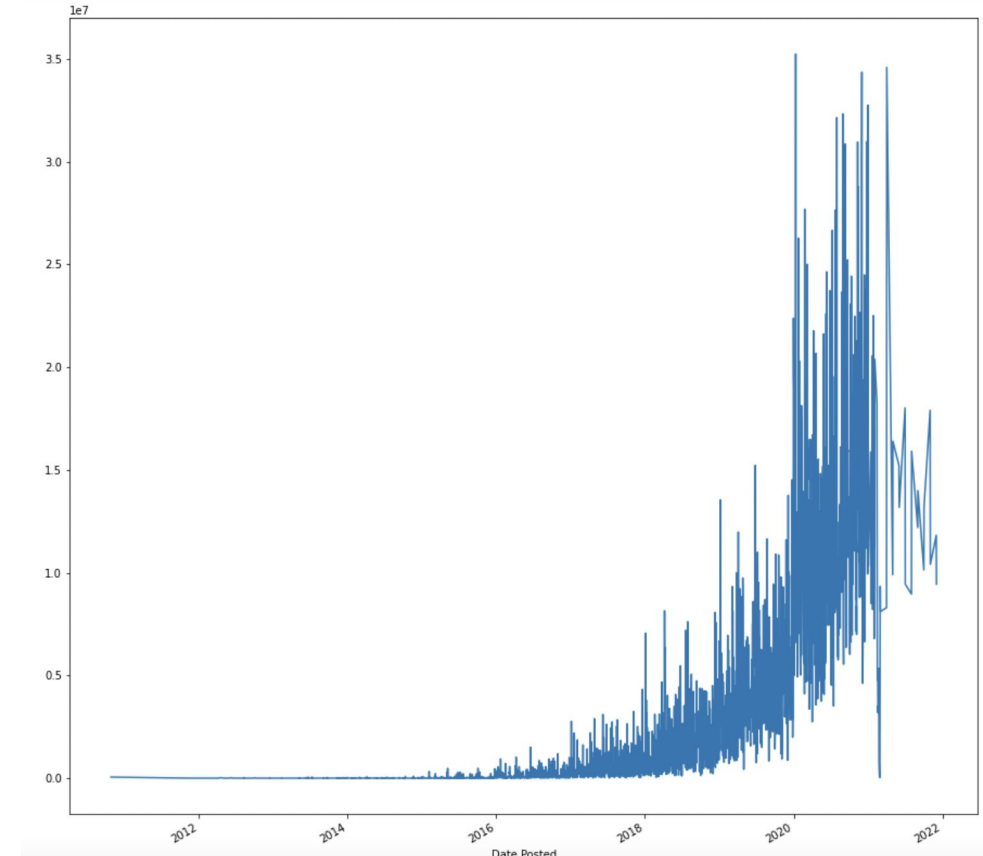
- Data collected from 1,100 active instagram users
  - Non-influencers
- Tracking engagement through likes, comments, etc on their posts
- Model interaction with instagram

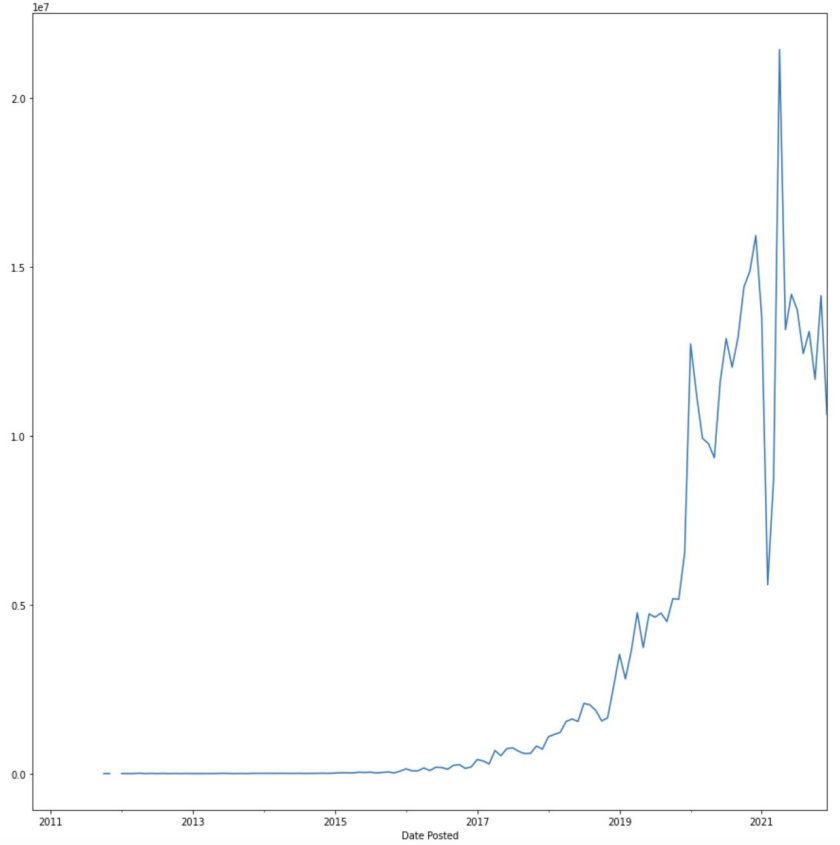
Wrangling the Data



- Combine posts made on the same day by multiple users
- Set Dates as the index

Taking a look at the Data

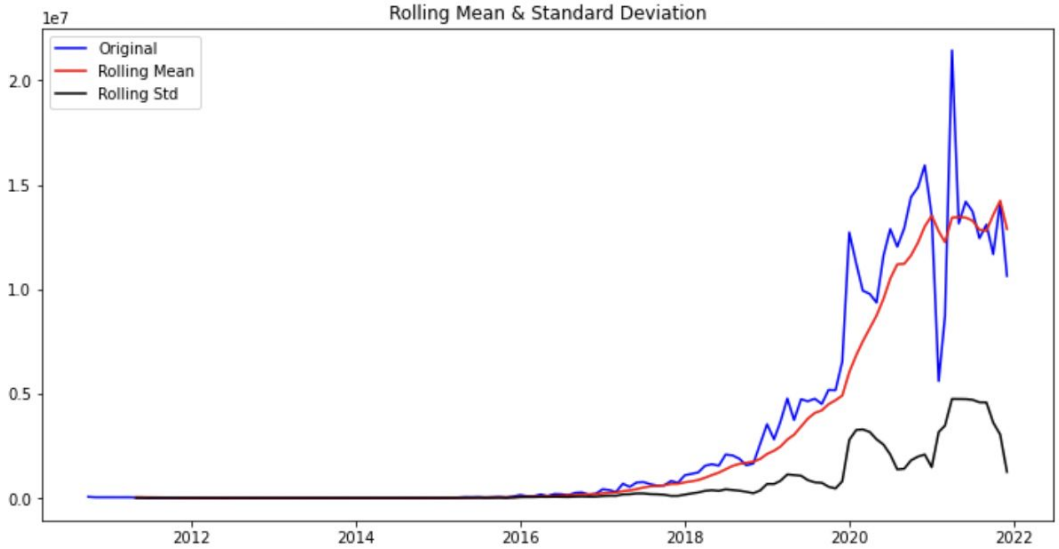




Manipulating for Time Series

- Preliminary Dickey-Fuller Testing

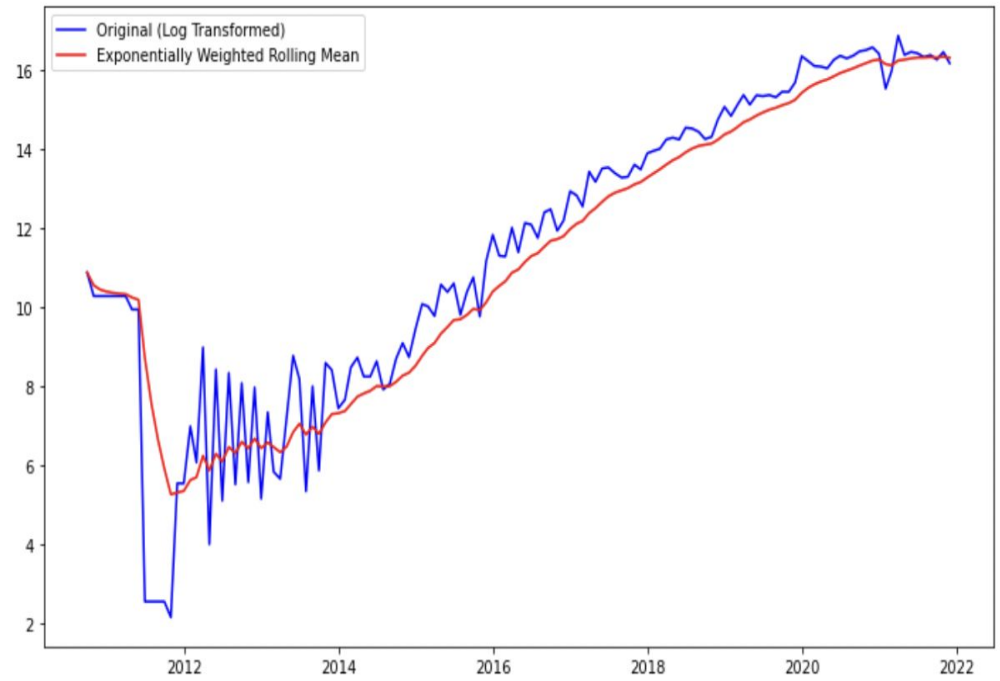
- Data is not stationary, needs to be transformed before we can work with this for time series modelling

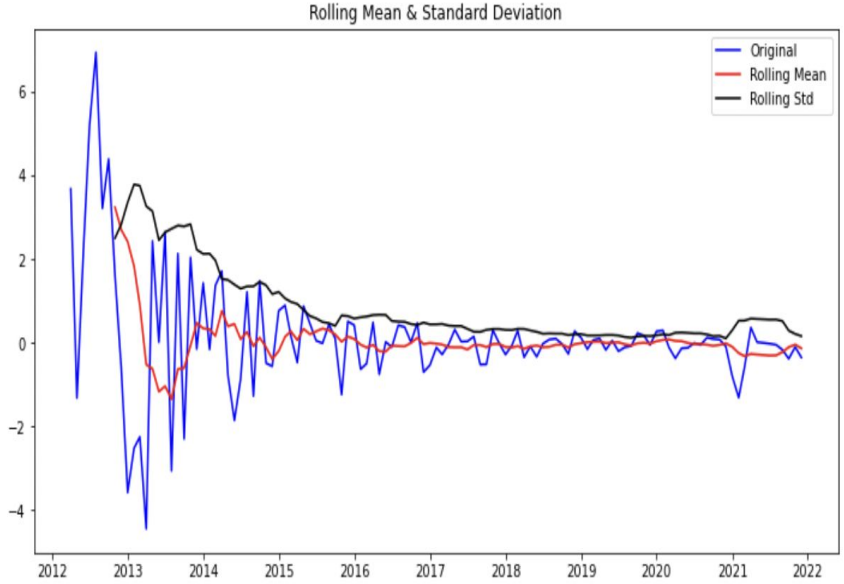



```
Results of Dickey-Fuller Test:

Test Statistic                  -0.496194
p-value                          0.892757
#Lags Used                      13.000000
Number of Observations Used    121.000000
Critical Value (1%)             -3.485585
Critical Value (5%)             -2.885739
Critical Value (10%)            -2.579676
dtype: float64
```

Transforming the Data





Results of Dickey-Fuller Test:

Test Statistic	-3.697710
p-value	0.004145
#Lags Used	13.000000
Number of Observations Used	103.000000
Critical Value (1%)	-3.495493
Critical Value (5%)	-2.890037
Critical Value (10%)	-2.581971

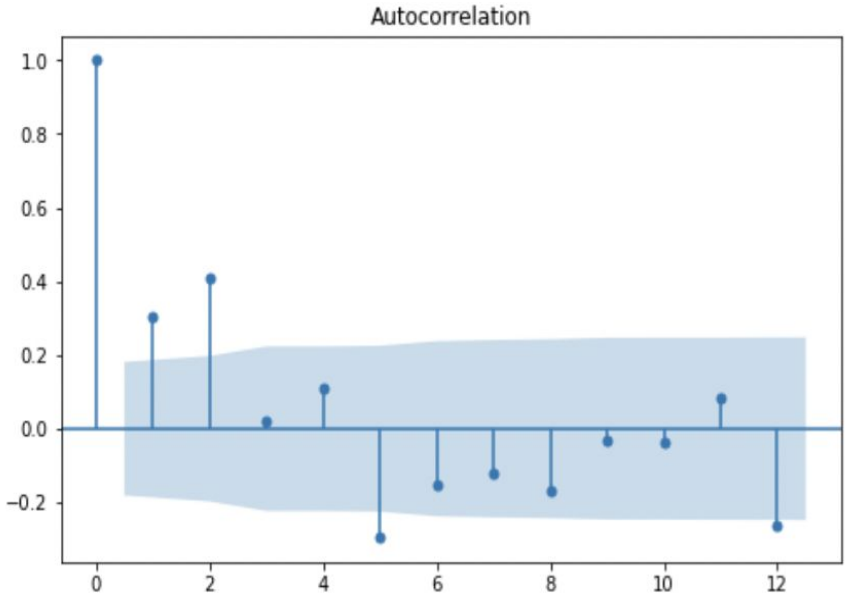
dtype: float64

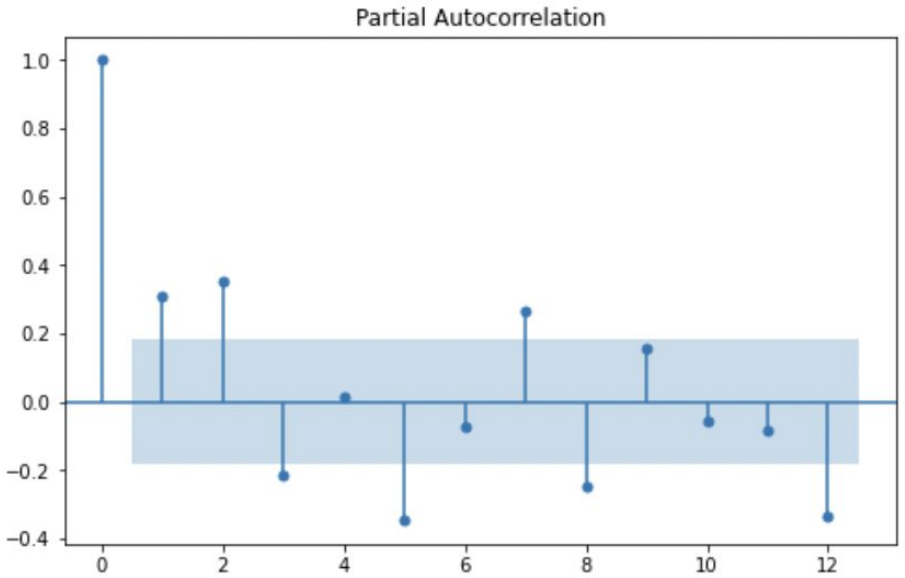
Getting Started on Modelling

- The data is now stationary and we can reject the null hypothesis

- As we are looking at data set that relies on consistent user activity (posting) as well as engagement (liking) I think it would be best to look at an ARMA Model

Autocorrelation





ARMA Model Results

Dep. Variable:	Likes	No. Observations:	117
Model:	ARMA(2, 1)	Log Likelihood	-189.682
Method:	css-mle	S.D. of innovations	1.221
Date:	Fri, 13 May 2022	AIC	389.364
Time:	09:57:18	BIC	403.175
Sample:	04-01-2012 - 12-01-2021	HQIC	394.971

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.1376	0.236	0.583	0.560	-0.325	0.600
ar.L1.Likes	-0.2367	0.116	-2.045	0.041	-0.463	-0.010
ar.L2.Likes	0.5180	0.081	6.359	0.000	0.358	0.678
ma.L1.Likes	0.5204	0.112	4.630	0.000	0.300	0.741

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	-1.1796	+0.0000j	1.1796	0.5000
AR.2	1.6364	+0.0000j	1.6364	0.0000
MA.1	-1.9216	+0.0000j	1.9216	0.5000

Conclusion & Next Steps

- **Using an Autoregression + Moving Average model to look at users can give us a good sense of trends going forward for social media**

- **Using Advanced Clustering to be able to see more user data**

- **More user data in general**