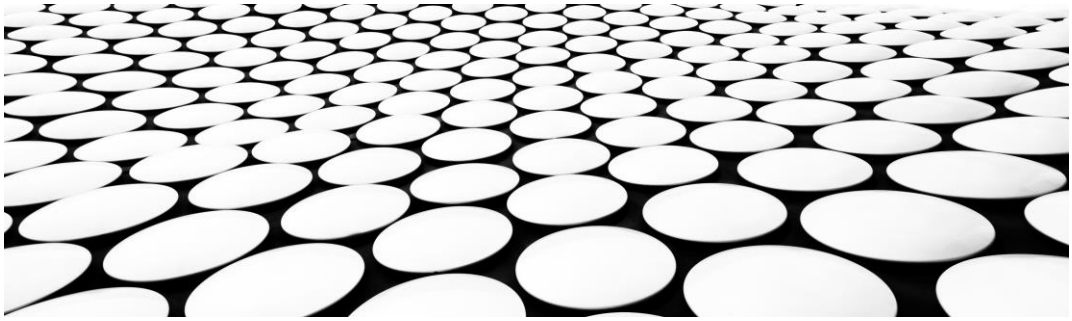
INCREMENT INSTAGRAM INFLUENCE

BY ERIK CASTRO // FLATIRON SCHOOL



OBJECTIVE

Learn the practices that make an Instagram account popular.

This can benefit anyone that wants to build a presence in this social network platform.

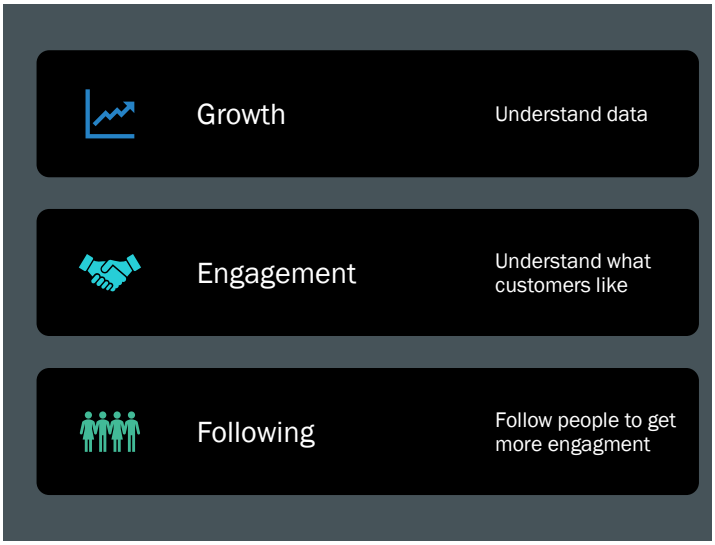
The goal of this study is to use the study to create an Instagram page and follow the recommendations to grow an audience

This project is probably one of the most ambitious projects I have worked so far.
There are different goals, firstly we need to understand what makes a person engage with an Instagram account, this will be specific for the theme I have chosen.
Secondly, this can be used to help someone already stablished with an Instagram account to further improve their engage with their audience and increment their followers
However, the main goal is to use the recommendations given at the end of the study and get an Instagram account as far as possible.
There is going to be involved Natural Language Processing, Time Series analysis (as a bench mark) , Machine learning for Classification and finally some AWS using EC2 and Lambda to run the scripts at certain times.

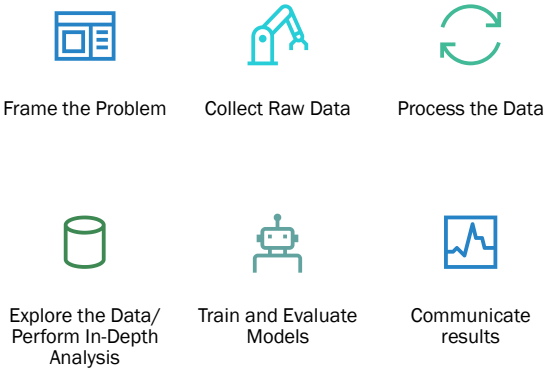
DATA:

- There are two sets of data:
  - Training data: this data consists of around 5.5k rows, consisting of the most important data from posts and the user. Which are:
    - Insta_name
    - Source_image
    - Caption
    - Caption_length
    - Followers
    - Num_of_likes
    - Media_Type
    - Vide_views
    - Number_of_hashtags
    - Num_of_mentions
    - Date
    - Time
  - Test data: this data consists of +6.5k rows, consisting of the same metadata as the training dataset

BUSINESS VALUE



METHODOLOGY



Frame the problem: Identify business priorities and make strategic decisions that will lead my work
Collect raw data: Extract data from the database provided.
Process the Data: Understand the data and proceed to clean it.
Explore the data: Split the data in different ways and use statistics to test and create visualizations to interpret data
Train and Evaluate Models: Create Time Series Models to predict Instagram account growth and create machine learning models to classify images.

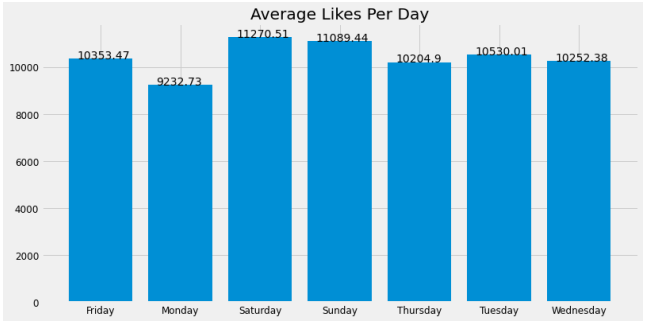
COLLECT RAW DATA AND PROCESS THE DATA

There are +11.000 Instagram metadata

There were 12 features, these were used differently for each of the objectives

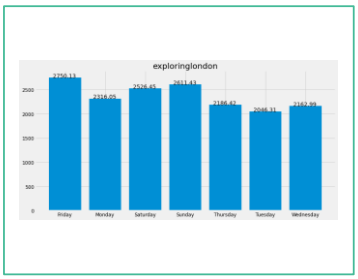
Clean data: Missing values, deal with data for Time Series Analysis, NLP and Image Classification

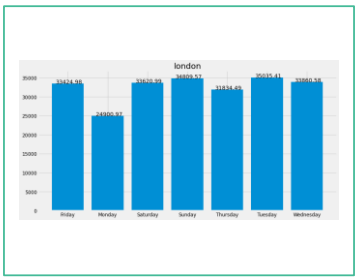
EXPLORE DATA/ PERFORM IN-DEPTH ANALYSIS

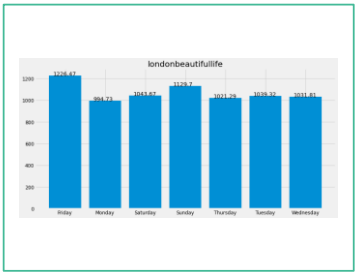


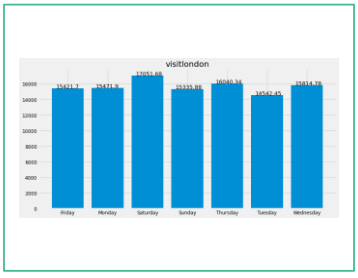
Within my data there were four different Instagram accounts, these accounts however, had the same theme. A theme: is a type of content that the Instagram account owner posts about. For example: cars, motorbikes, country.
The theme I have chosen was a city : London.
Getting the average, we can appreciate that the Instagram users that follow this kind of Instagram pages are more active Friday, Saturday and Sunday.

EXPLORE DATA/ PERFORM IN-DEPTH ANALYSIS

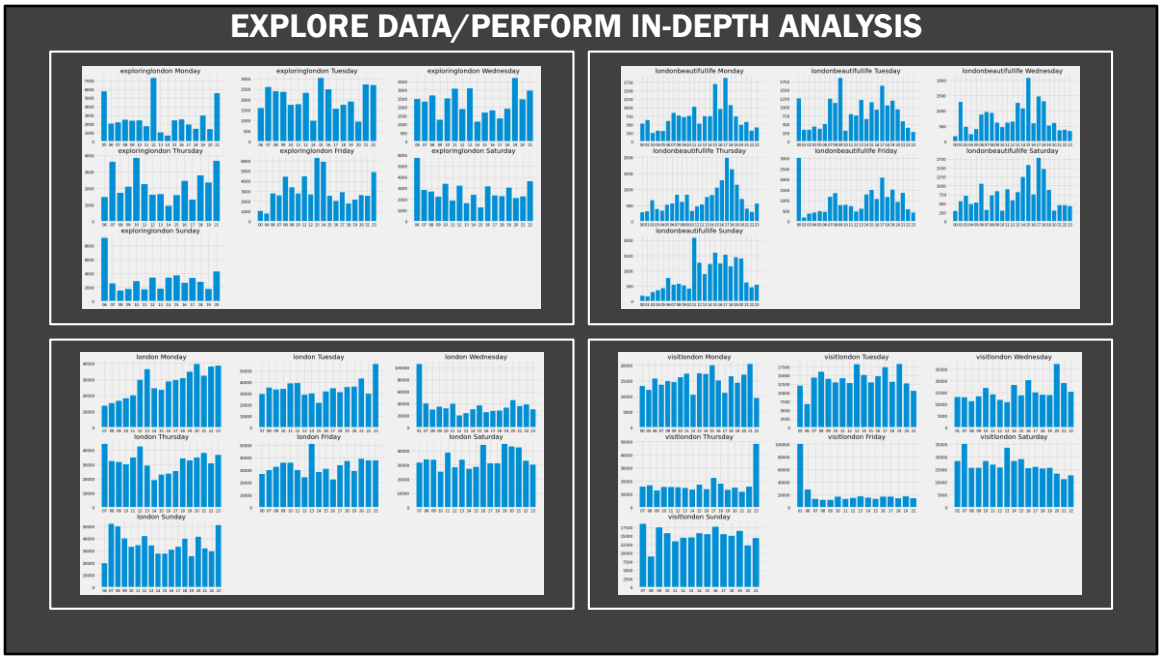








Looking at each Instagram account individually we can differentiate how their followers react to their posts.
Exploring London: gets more likes on Fridays, Saturday and Sundays.
Moving to London: gets more likes on Sundays, Tuesdays and Wednesdays
Londonbeautifullife: on Fridays, Saturdays and Sundays
Lastly, Visit London: Saturday, Thursday and Wednesdays are where the average likes are higher.

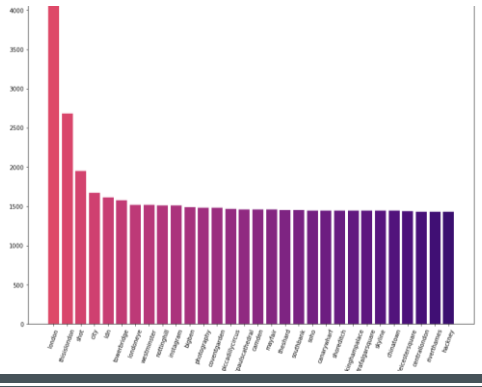


We can further Explore within the day, the time of the day with a higher average of likes,

Due to the time constrain I will not be able to go through all the Instagram pages. However, Exploring London will serve as an example.

We can see that posting on Monday at either 05am, 12pm or 21 gets the account more likes on average.

Friday in the other hand, 13:00 and 14:.00 gets the most likes



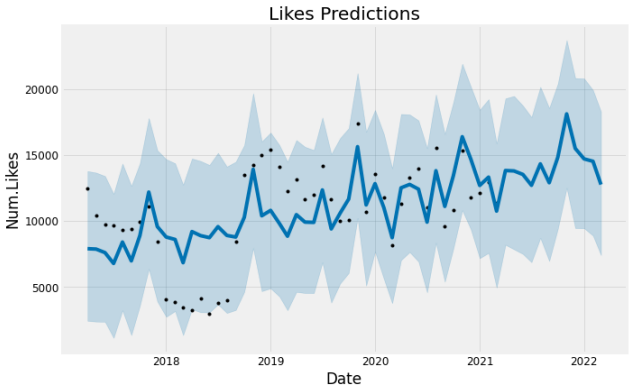
EXPLORE DATA/ PERFORM IN-DEPTH ANALYSIS

TEXT CLASSIFICATION

I conducted some text classification in the caption of the posts. This allowed me to generate the best captions for the pictures.
We can see the top 30 words used in the captions of the pictures. For my surprise Hackney is number 30.

TRAIN AND EVALUATE MODELS

TIME SERIES ANALYSIS



I created this time series analysis to see where the Instagram pages will be in the future if they kept posting in a similar manner as they have had all this time.
We can see that towards the end of the year the account tends to get more likes. We can appreciate that this happens in the predictions. However, the further the predictions are the less reliable it is

TRAIN AND EVALUATE MODELS

IMAGE CLASSIFICATION MACHINE LEARNING

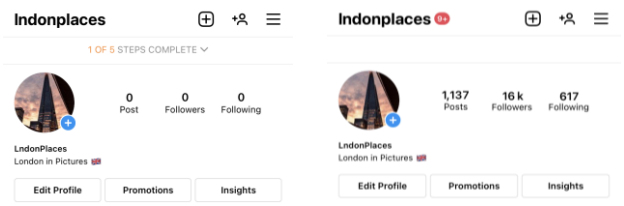
Model	Test F1	Test Recall	Test Precision	Test Accuracy
LR	0.80	0.98	0.67	0.67
GB	0.80	0.91	0.72	0.70
AB	0.79	0.85	0.73	0.69

Image classification was done with a totally different dataset, this dataset constituted of images of New York. The machine learning model was trained on this data but it was later used on my main London dataset.
I had to manually classify 6.5k following a criteria. The model that performed better was Logistic Regression. We can see that the F1 score and the test Recall are higher than in the other models. Therefore, I decided to create my final model using Logistic Regression.

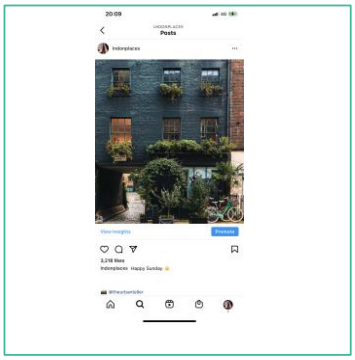
RECOMMENDATIONS

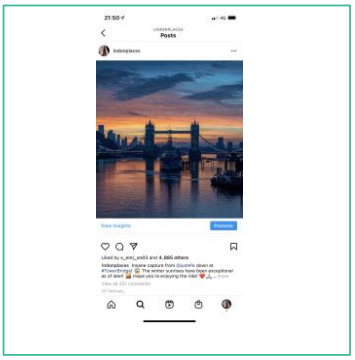
- Post heavily on Fridays, Saturdays and Sundays.
- On the weekends, post between 12:00 and 15:00, also between 19:00 and 21:00
- Use top 30 Hashtags found on the post's captions. These are:
  - London, thisislondon, shot, city, ldn, towerbridge, londoneye, Westminster, nottinghill, Instagram, bigben, photography, coventgarden, picadillycircus, stpaulscathedral, Camden, mayfair, theshard, southbank, soho, canarywharf, shreditch, buckinghampalace, trafalgarsquare, skyline, chinatown, leicestersquare, centrallondon,riverthames and hackney

RESULTS



I have included the results as an extra slide, I created a Instagram page and followed my recommendations. I used the 30 hashtags, posted +4 pictures each day on Friday, Saturday and Sunday. I posted other days as well, but not with the same intensity as on the weekend.
I also followed some people, to make it look more genuine and evade Instagram from blocking my account





RESULTS

We can see that these pictures show the amount of likes on each picture, we have a progressive increase on them. First picture was posted at the beginning of February with 3,218 likes, second picture was posted at the end of February and has 4,885 likes. Last picture was on March and it had 5,755 likes. This means that following the recommendations will help this particular Instagram themed to increase their engage and their followers.

To Finalize this project I decided to sell this account, this was done smoothly and at a really good price.

FUTURE WORK

- This is a project that must be tuned depending on how users react to the posts. Therefore, analysis is required every 3 months.
- Scrapping more images is required every 2/3 weeks.
- Spend more time with time series analysis.

THANK YOU

Erik Castro – Flatiron School

Any Questions?