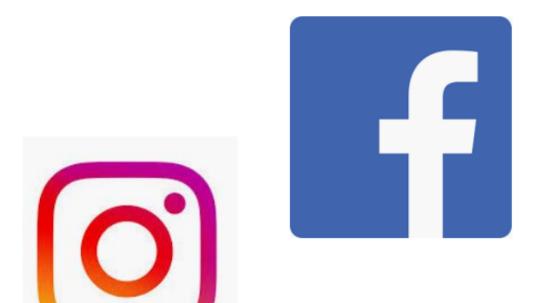
NEWS POPULARITY ON DIFFERENT PLATFORMS

Zhujun Shen
Data Science Initiative, Brown University
October 21, 2019



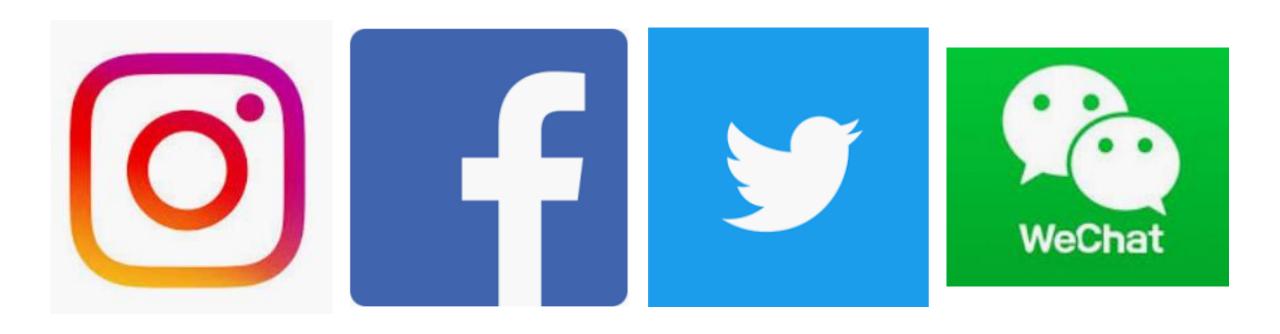






Why important?



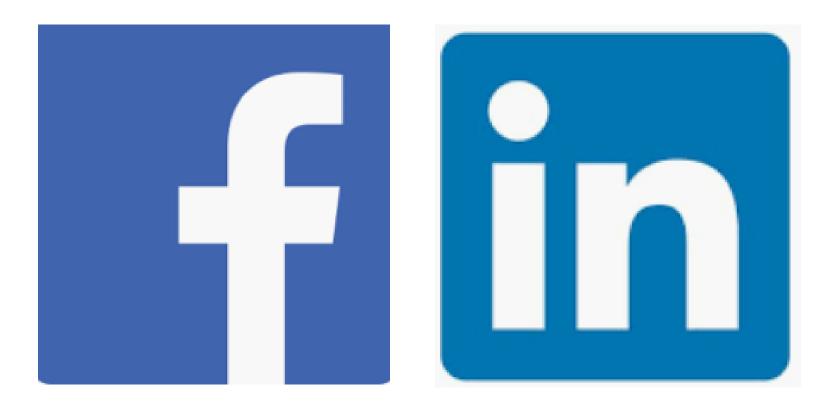


Where to post? What to post? When to post?



News Popularity







What to post?

When to post?

Where to post? (classification)



Data Description

- Features:
- Publish Date
- 2. Topic (economics, Obama, Microsoft, Palestine)
- 3. Title / Sentiment Title (No need to preprocess)
- 4. Headline / Sentiment Headline (No need to preprocess)
- 5. News popularity on different platforms (Facebook, GooglePlus, LinkedIn)



^{*} No missing value in the dataset.

Categorical features

- Features:
- 1. Publish Date
- 2. Topic (economics, Obama, Microsoft, Palestine)

OneHotEncoder

```
'x0_economy', 'x0_microsoft', 'x0_obama', 'x0_palestine'
```

- 3. Title / Sentiment Title
- 4. Headline / Sentiment Headline
- 5. News popularity on different platforms (Facebook, GooglePlus, LinkedIn)



Continuous features

- Features:
- 1. Publish Date Unix time
- 2. Topic (economics, Obama, Microsoft, Palestine)
- 3. Title / Sentiment Title
- 4. Headline / Sentiment Headline
- 5. News popularity on different platforms (Facebook, GooglePlus, LinkedIn)

They are all reasonably bounded. Thus, MinMaxScaler!



Label

- Features:
- Publish Date
- 2. Topic
- 3. Title / Sentiment Title
- Headline / Sentiment Headline
- News popularity on different platforms (Facebook, GooglePlus, LinkedIn)

- The news is only posted on less than two platforms.
 - Facebook has the highest popularity.
 - 2 GooglePlus has the highest popularity.
 - 3 LinkedIn has the highest popularity.
 - 4. The news is post on all three platforms and gain zero popularity.
- **5** Others.



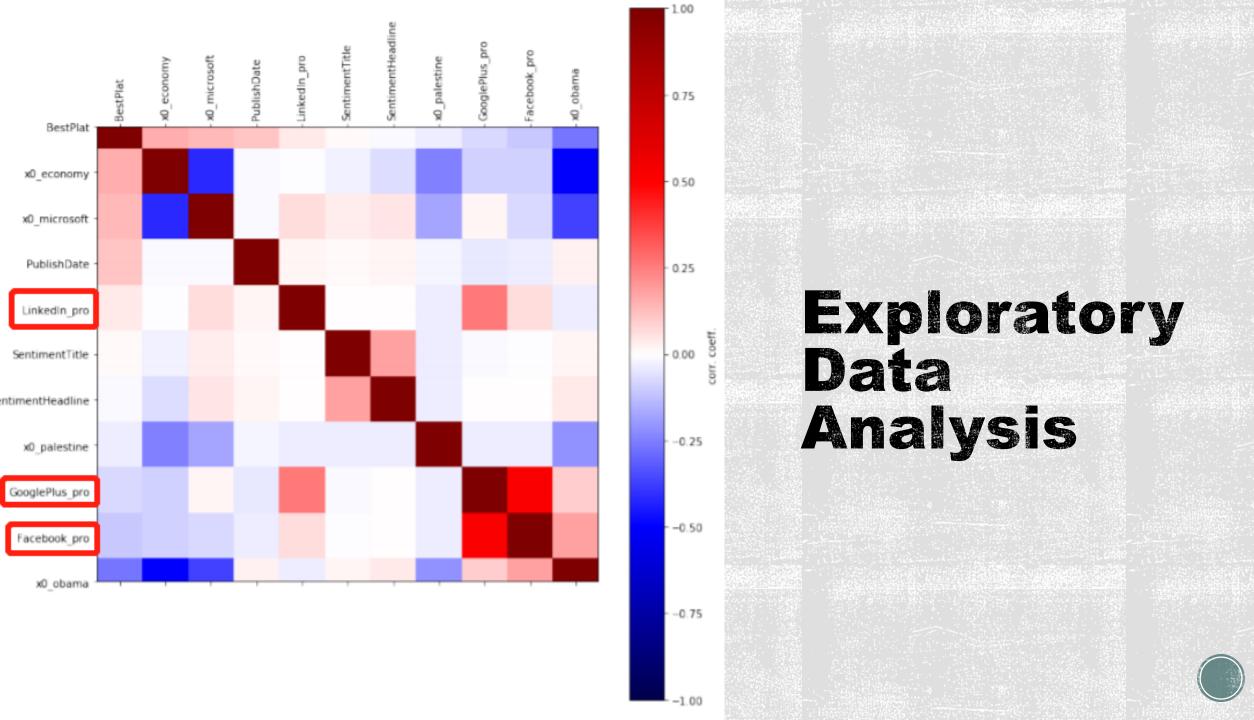
Data Preprocessing:

- No missing value in the dataset.
- Resulting dataset:

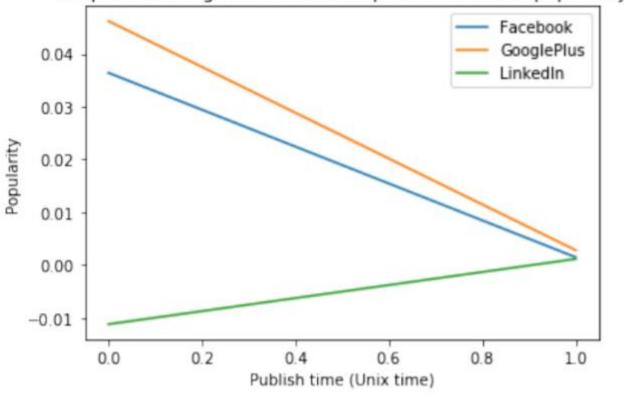
df.shape

(93239, 16)





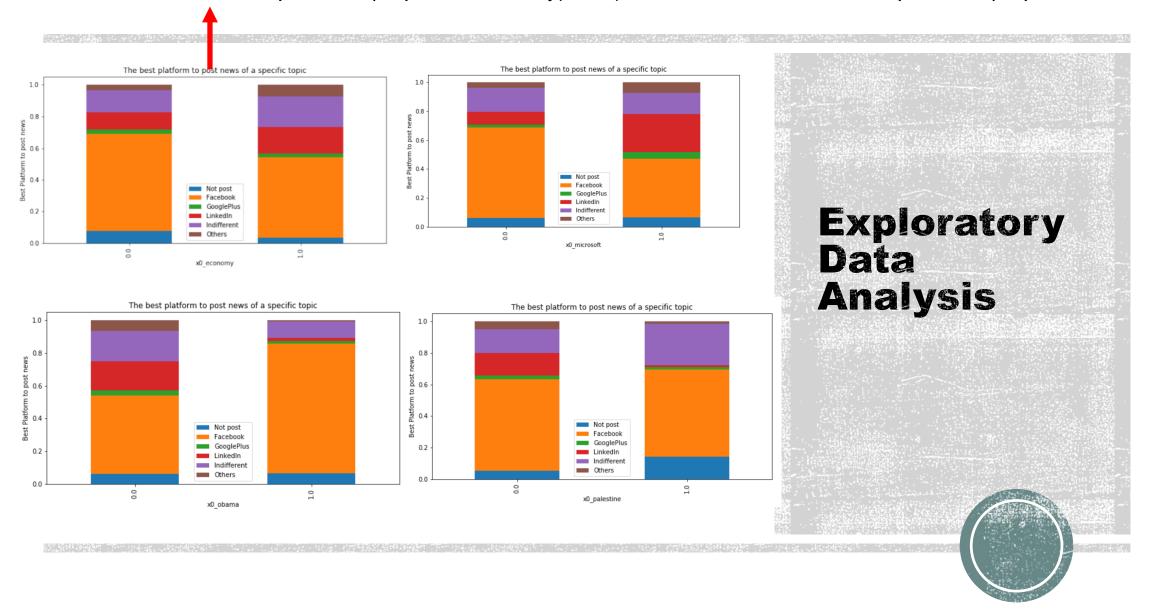
Simple linear regression between publish time and popularity

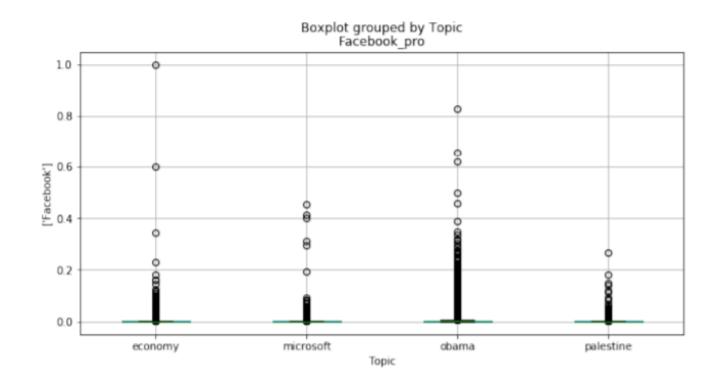


Exploratory Data Analysis



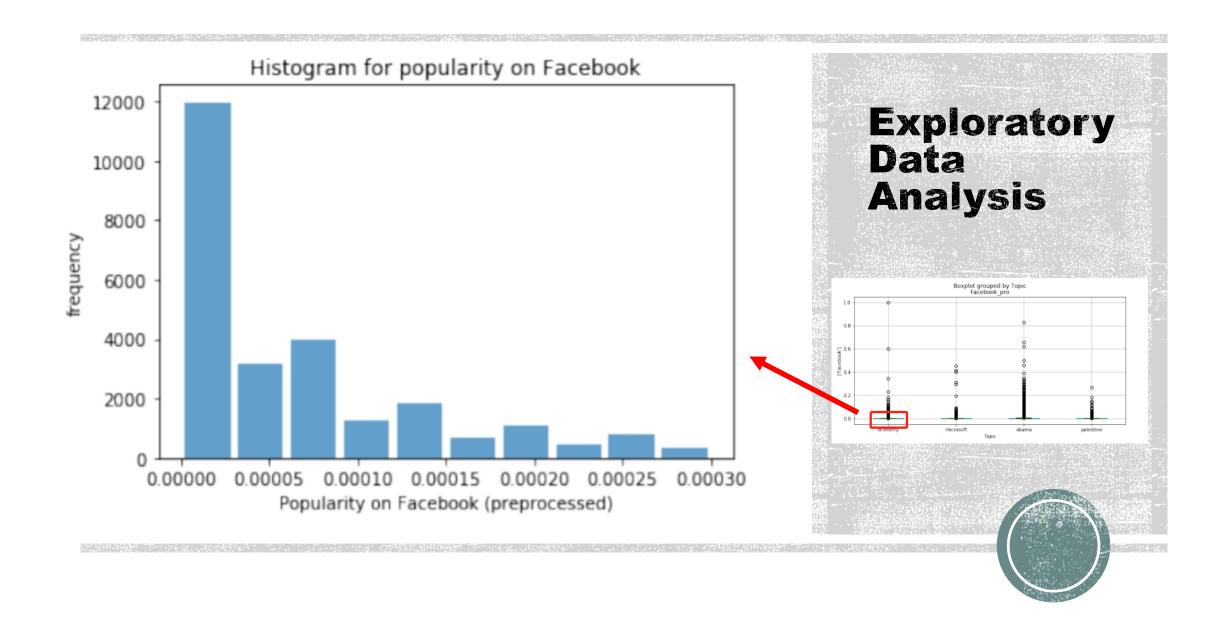
P(choose Facebook as final platform | topic = economy) < P(choose Facebook as final platform | topic != economy)





Exploratory Data Analysis





THE END Zhujun Shen

Data Science Initiative, Brown University October 21, 2019

