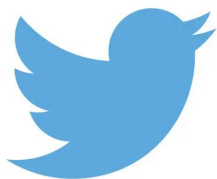


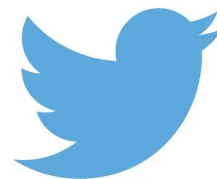
---

---

# Maximizing Tweet Outreach and Predicting Tweet Performance



Angad Pfuetzner



# The Problem

How can musical artists reach the largest audience on Twitter?

Can we predict how a tweet will perform before it is tweeted?

# Who Cares?

Musical Artists



Social Media  
Influencers



Marketers



# Data Sources



Music-Map

The Tourist Map of Music



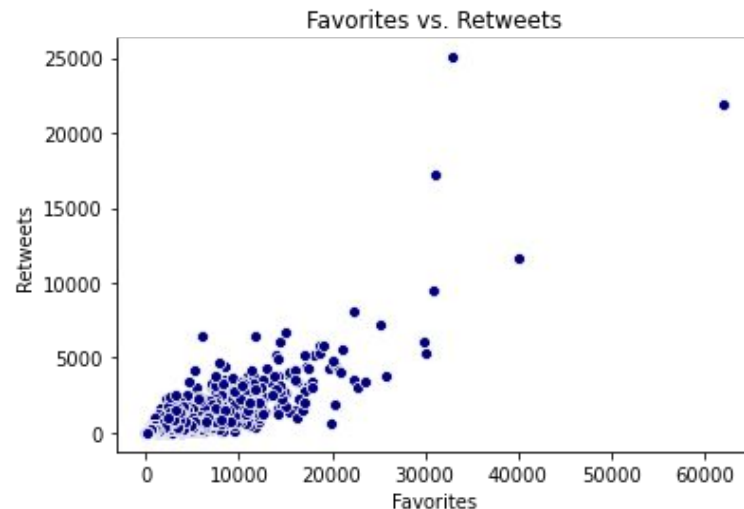
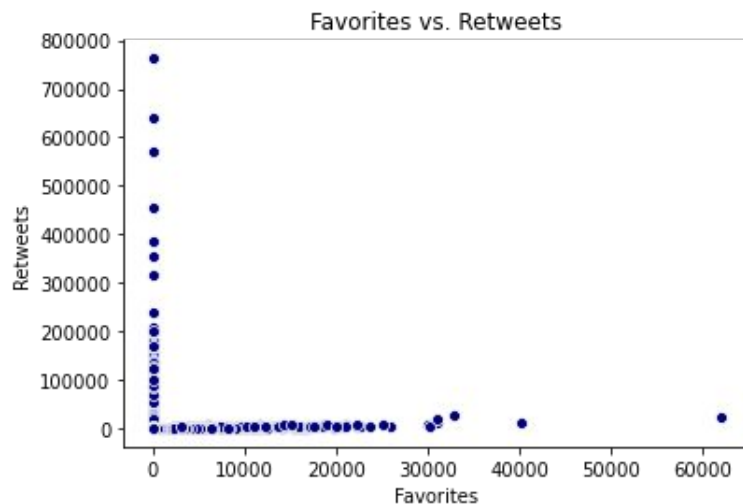
# Artists Selected

- Alison Wonderland
- Galantis
- San Holo
- The Knocks
- Peking Duk
- Porter Robinson
- What So Not
- Louis The Child
- NGHTMRE
- Jai Wolf

31 680 tweets pulled

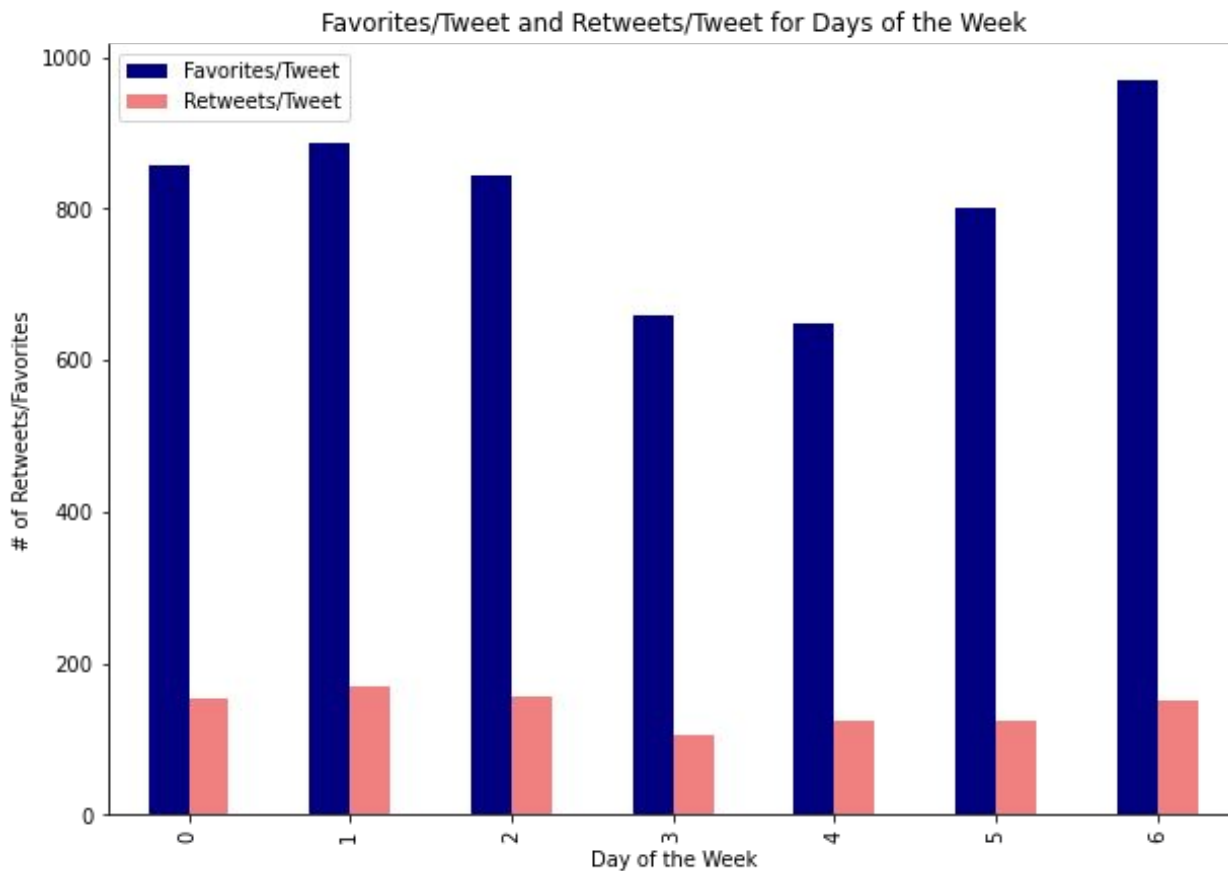
# Data Cleaning

- Retweeted Tweets

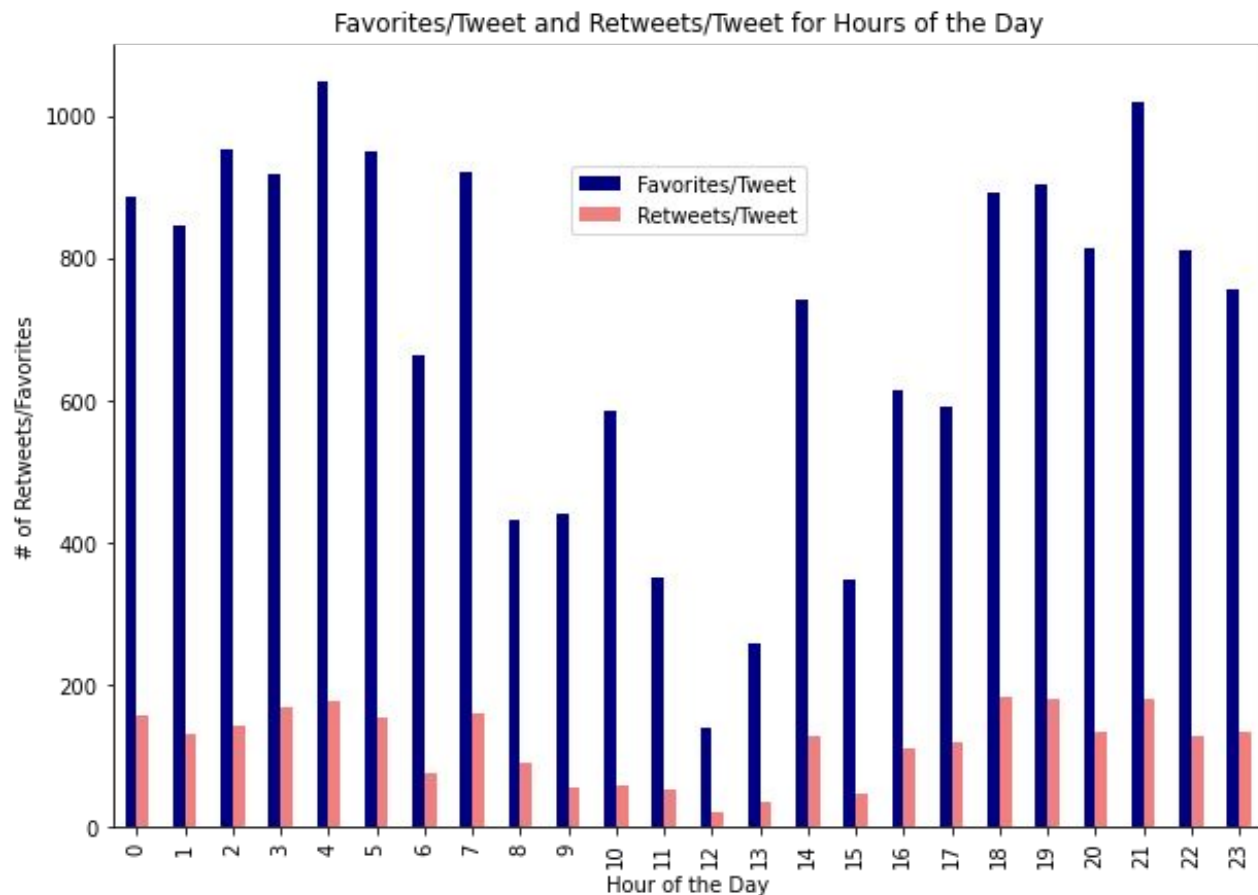


- Reply Tweets
- Adding Day of Week and Time of Day

# Optimal Day of Week



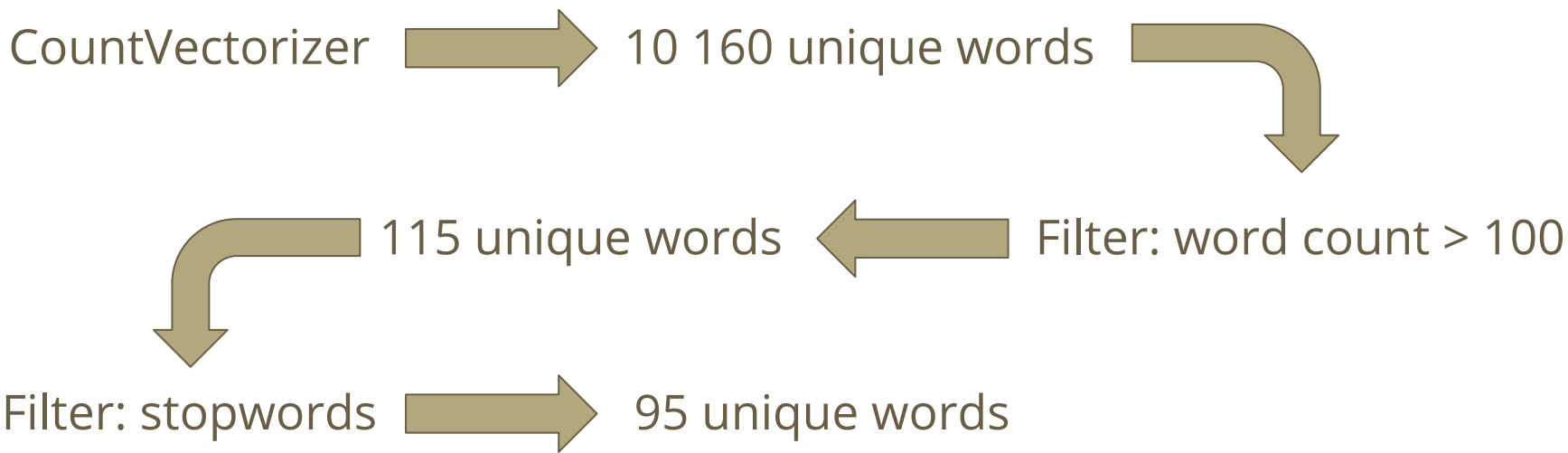
# Optimal Time of Day





# NLP Data Processing

- Convert media links ("http.") to the string "containsmedia"



# General Pre-Processing

- Pandas get\_dummies (screen\_name, day\_of\_week, time\_of\_day)
- Manual selection of relevant words
- Final Features - 70
  - 6 day\_of\_week
  - 23 time\_of\_day
  - 9 screen\_name
  - 32 words

# Determining the Dependent Variable

Should we use favorites or retweets as the variable to predict with the model?

Build Linear Regression Models

$R^2$  using favorites = 0.28

$R^2$  using retweets = 0.16

# Model Selection

	Train MSE	Test MSE	Train Time	Prediction Time
Model				
Linear Regression	3078481.255	3491299.105	1.360	0.005
KNN	4304365.629	4993206.518	0.489	4.400
Random Forest	1551021.994	7007460.686	76.010	27.292

# Hyperparameter Tuning

## ElasticNet Algorithm

- alpha: (0.01, 0.1, 1)
- l1\_ratio: (0, 0.5, 1)
- normalize: (True, False)

## Optimal Parameters

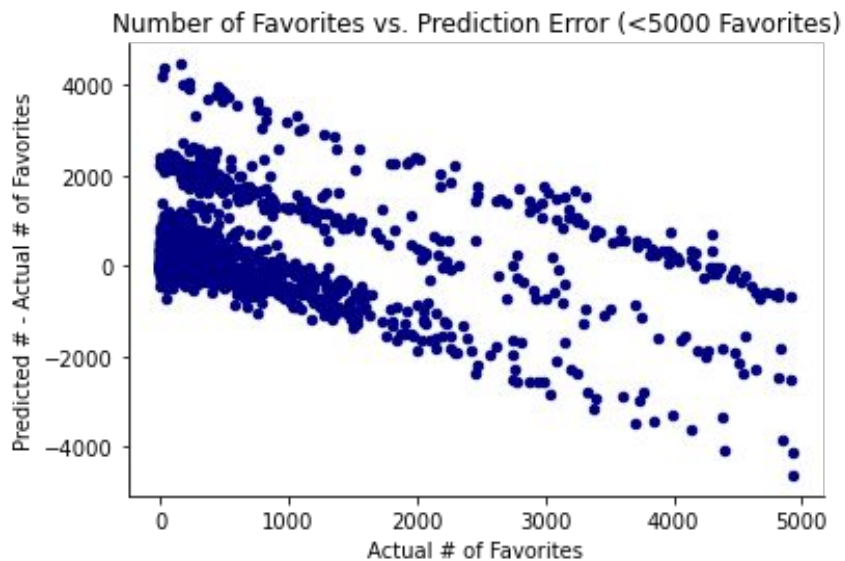
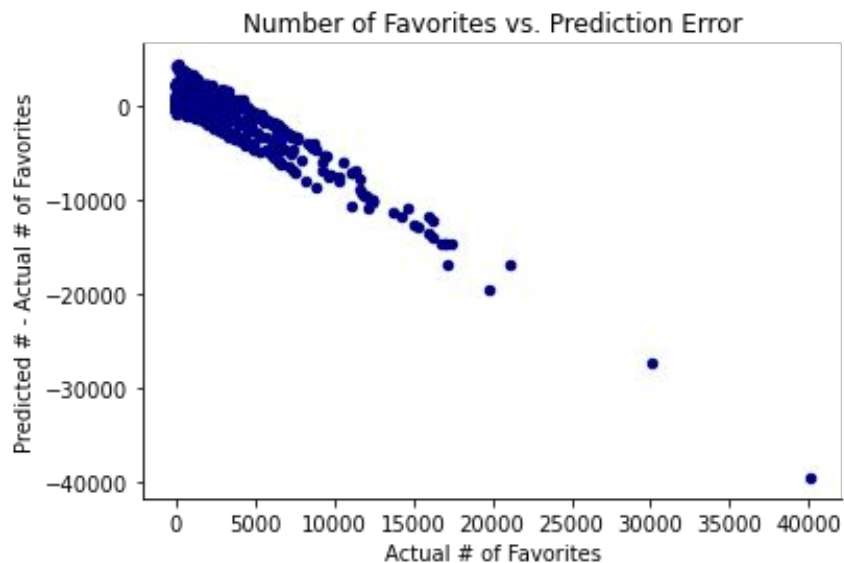
- alpha: 0.1
- l1\_ratio: 1
- normalize: True

# Final Model Coefficients

Feature	Coefficient
screen_name_awonderland	3847.85
screen_name_porterrobinson	1988.78
music	635.84
screen_name_sanholobeats	550.35
going	519.62
album	504.00
virtual	483.02
miss	420.03
song	360.22
first	275.70

# Results

Final MSE: 3 490 805 (MAE = 1868)



## Further Study

- Effect of time on tweet performance
- More advanced NLP analysis



# Key Takeaways

## Optimal Times to Tweet

- Sundays
- 18h - 21h and 0h - 4h

## Predictive model off by ~1900 favorites/tweet

- Needs further development

# Thank You!

Github: [https://github.com/Angad99/Springboard/tree/main/Capstone\\_Two](https://github.com/Angad99/Springboard/tree/main/Capstone_Two)

LinkedIn: <https://www.linkedin.com/in/angadpfuetzner/>

Email: [angad.pfuetzner@mail.mcgill.ca](mailto:angad.pfuetzner@mail.mcgill.ca)