

# A Musician's Alter Ego

Benjamin Lin

June 1st, 2018



# Robert Bryson Hall II

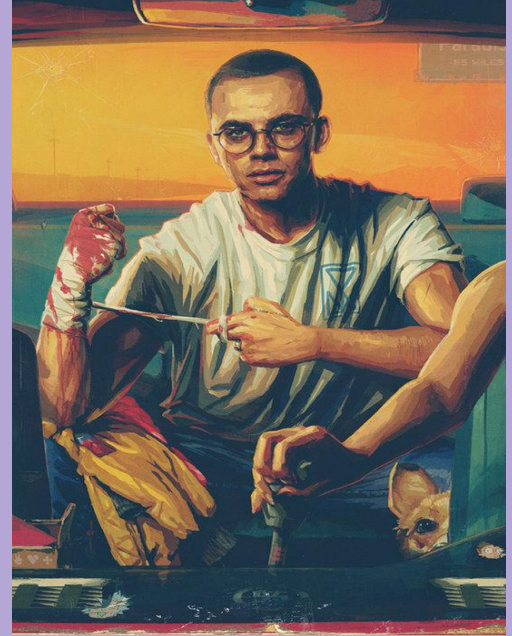
Logic



Young Sinatra



Bobby Tarantino



# Robert Bryson Hall II

Logic



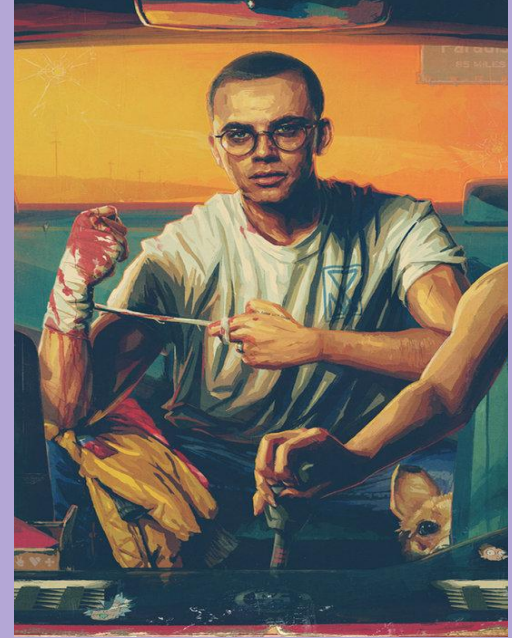
Peace, love, society

Young Sinatra



Success, motivated,  
vengeful

Bobby Tarantino



Outspoken,  
turn-up, carefree



# Objectives

1. Cluster his lyrics into different groups
2. Predict if a line came from Logic, Young Sinatra, or Bobby Tarantino

# Data

## Logic



2,100 rows

## Young Sinatra



2,000 rows

## Bobby Tarantino

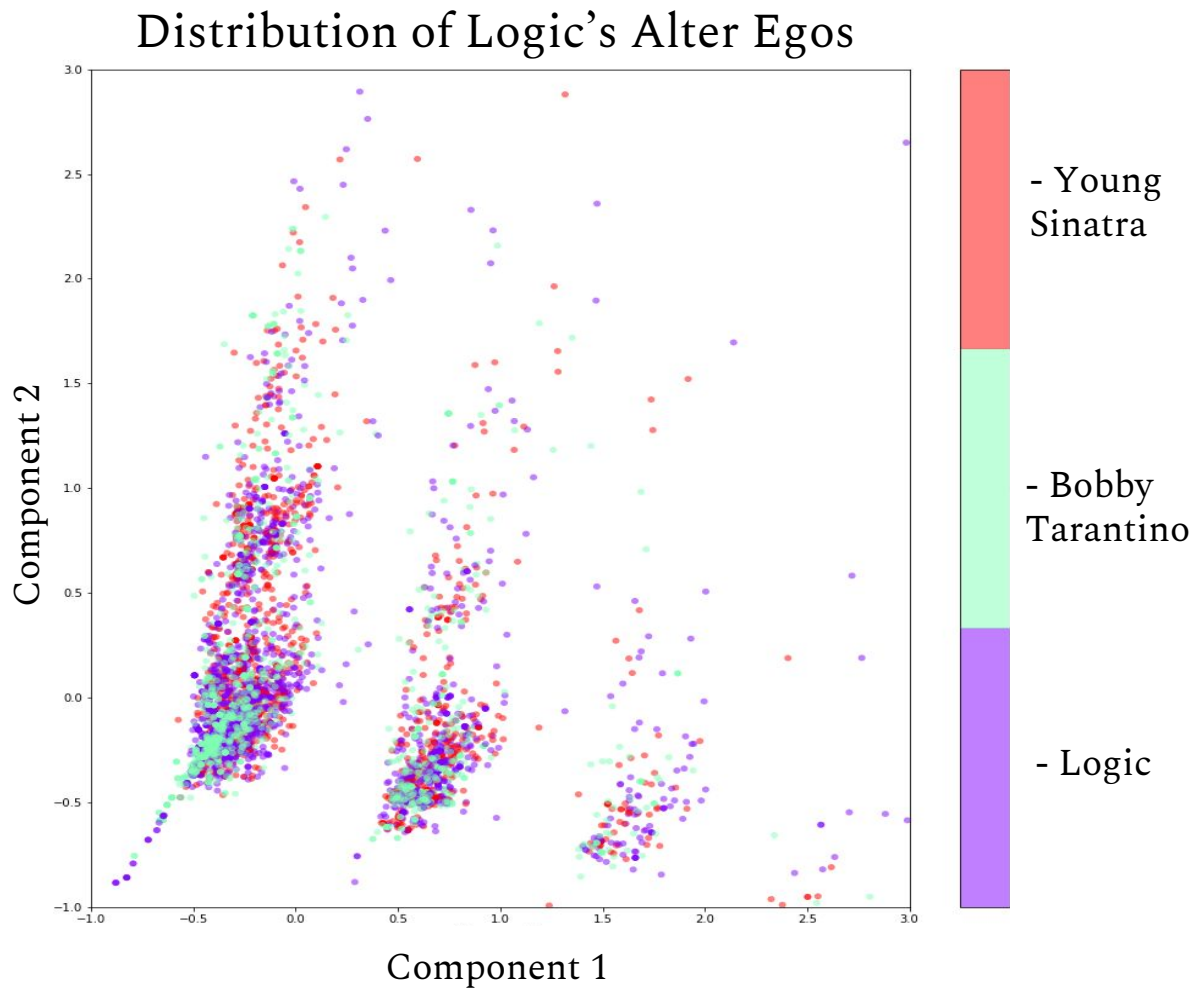


1,800 rows

# Unsupervised Clustering



# PCA



LDA

# Alter Ego Topics

Let's revisit his 3 alter-egos and their corresponding topics:

Logic	Young Sinatra	Bobby T.
<ul style="list-style-type: none"><li>- Peace</li><li>- Love</li><li>- Society</li></ul>	<ul style="list-style-type: none"><li>- Success</li><li>- Motivated</li><li>- Vengeance</li></ul>	<ul style="list-style-type: none"><li>- Outspoken</li><li>- Turn-up</li><li>- Carefree</li></ul>

# LDA Topics

LDA best split the text into 2 topics. Here are the top terms for each topic:

Topic 1	Topic 2
<ul style="list-style-type: none"><li>- Everybody</li><li>- Love</li><li>- Mind</li><li>- People</li><li>- World</li></ul>	<ul style="list-style-type: none"><li>- Money</li><li>- Nobody</li><li>- Livin</li><li>- B*tch</li><li>- Give</li></ul>

# LDA Topics

LDA best split the text into 2 topics. Here are the top terms for each topic:

Topic 1	Topic 2
<ul style="list-style-type: none"><li>- Everybody</li><li>- Love</li><li>- Mind</li><li>- People</li><li>- World</li></ul>	<ul style="list-style-type: none"><li>- Money</li><li>- Nobody</li><li>- Livin</li><li>- B*tch</li><li>- Give</li></ul>
Logic!	Young Sinatra/Bobby T!

# Supervised Classification



# Models

## Features

### Algorithms

	Bag of Words	tf-idf	Spacy Features	Word2Vec
Multi N.B.	0.672	<u>0.674</u>	0.427	0.417
KN Classifier	0.604	0.646	0.532	0.592
SVC	0.601	0.429	0.429	0.574

\*metric = test set accuracy score



The screenshot shows a Jupyter Notebook window with three code cells. The first cell is empty. The second cell contains a function call to `who_said` with a list containing the string `'i want the world to be a better place'` and a list containing the string `'Logic'`. The third cell contains a function call to `who_said` with a list containing the string `'lets drink and party all night'` and a list containing the string `'Bobby Tarantino'`. The fourth cell contains a function call to `who_said` with an empty list `[]`. The notebook interface includes a top bar with tabs for 'Project 4 - 10000', 'Jupyter Notebook', and 'Jupyter Notebook'. The bottom bar shows various icons for file operations and a status bar.

```
In [48]: who_said(['i want the world to be a better place'])  
         ['Logic']  
  
In [49]: who_said(['lets drink and party all night'])  
         ['Bobby Tarantino']  
  
In [ ]: who_said([])
```

# The End!

Email: [benlin100@gmail.com](mailto:benlin100@gmail.com)

LinkedIn: [linkedin.com/in/benjaminlin24/](https://www.linkedin.com/in/benjaminlin24/)

Github: [github.com/benlin100](https://github.com/benlin100)

# Appendix

# Agenda

1. Background & Objectives
2. Data
3. Unsupervised Clustering
4. Supervised Classification
5. Conclusion

# Background & Objectives



# Summary

1. Cluster his lyrics into different groups
  - PCA
  - LDA (Topic Modeling)
2. Predict if a line came from Logic, Young Sinatra, or Bobby Tarantino
  - Features: Bag of Words, SPACY, W2V, tf-idf
  - Modeling Algorithms: Multi N.B, KN Classifier, SVC

# Models

## Features

### Algorithms

	Bag of Words	tf-idf	Spacy Features	Spacy + BoW	Word2Vec
Multi N.B.	0.672	<u>0.674</u>	0.427	0.665	0.417
KN Classifier	0.604	0.646	0.532	0.577	0.592
SVC	0.601	0.429	0.429	0.429	0.574

\*Metric = Test set accuracy score