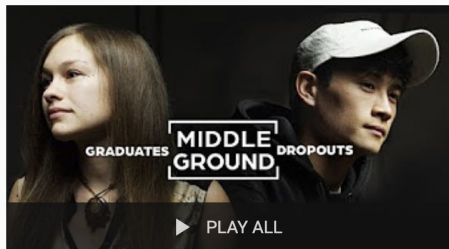


# **Analyzing Youtube Comments**

**Vegans vs Meateaters**

## Goals:

1. Train a model that classifies 25,000+ comments as “vegan” or “meat-eater”
2. See if there’s a difference in sentiment by class



## Middle Ground : Season 2

8 videos • 615,899 views • Last updated on May 24, 2018



Can opposing thought groups find middle ground?  
People holding differing views on topics such as  
education and immigration seek to find out.



Jubilee

SUBSCRIBE 2.9M

### SEASON 2 ▾

- MIDDLE GROUND S2 • E1  
**Dropouts And Graduates: Is College Worth It?**  
Jubilee
- MIDDLE GROUND S2 • E2  
**Pro-Gun Vs. Anti-Gun: Is There Middle Ground?**  
Jubilee
- MIDDLE GROUND S2 • E3  
**Can Trump Supporters And Immigrants See Eye To Eye?**  
Jubilee
- MIDDLE GROUND S2 • E4  
**Vegans Vs. Meat Eaters: What Is The Right Diet?**  
Jubilee
- MIDDLE GROUND S2 • E5  
**Can You Be In Love With Multiple People?**  
Jubilee
- MIDDLE GROUND S2 • E6  
**Can Feminists and Non-Feminists Agree On Gender Equality?**  
Jubilee
- MIDDLE GROUND S2 • E7  
**Can Black Lives Matter & Law Enforcement See Eye To Eye?**  
Jubilee



**Vegan**



**Vegan**

# Imbalanced Classes

```
1 # Baseline Accuracy
2 y.value_counts()/df.shape[0]
```

```
0    0.059349
1    0.005521
```

## Random Forest with TFIDF: 91.5% Accuracy

```
# Random Forest and TFIDF

tf = TfidfVectorizer(stop_words='english', ngram_range=(1,2))
Trees = RandomForestClassifier()

pipe = Pipeline([
    ('tfidf', tf),
    ('trees', Trees)
])

gs = GridSearchCV(pipe, param_grid = {})

gs.fit(X_train, y_train)

print('Grid Search Best Score:', gs.best_score_)
print('Test Score:', gs.score(X_test, y_test))
```

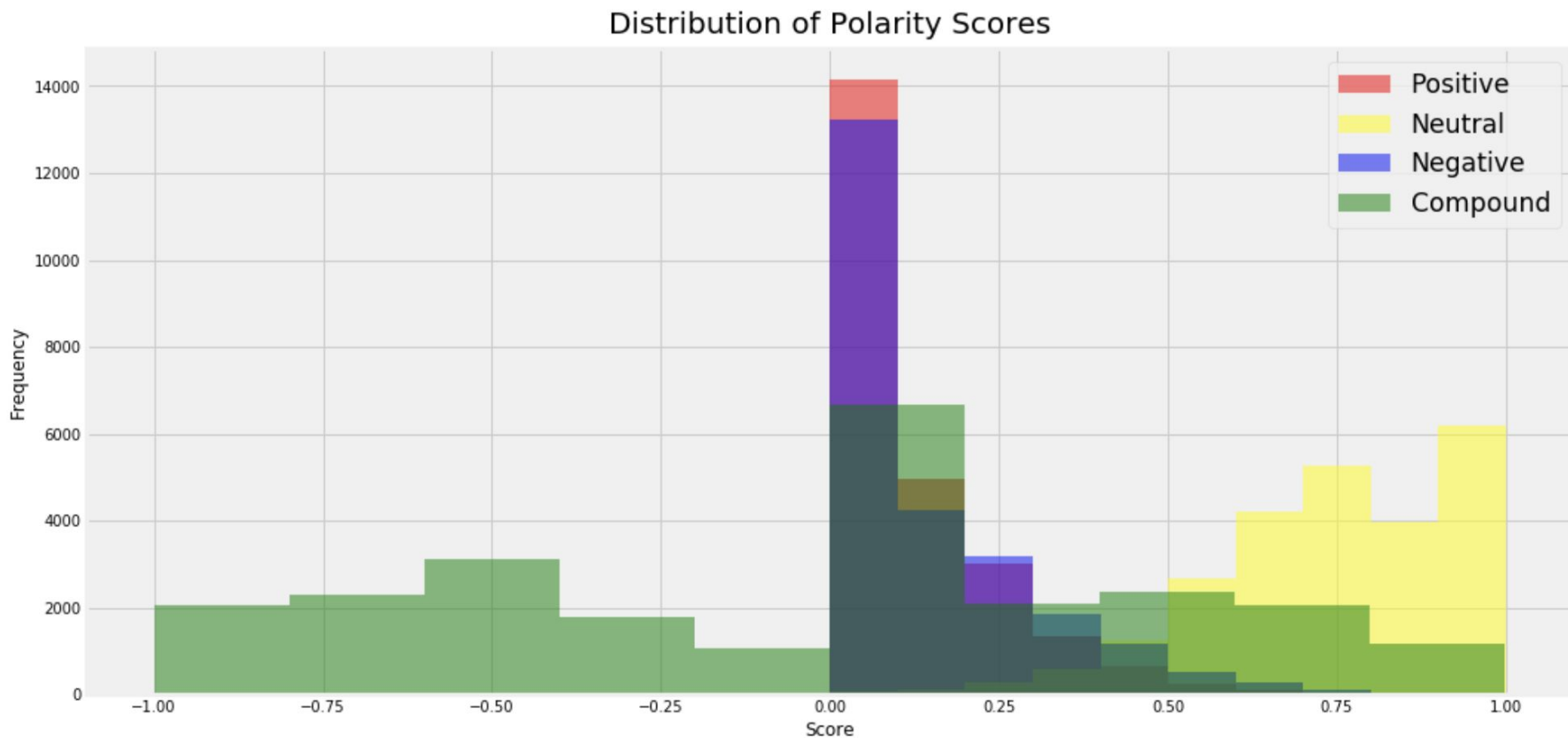
2:06 ima take the shit out that life i  
love meat

THE GUY WITH ORANGE SCARF IN  
HIS HEAD IN THE 1...

Vegans...need meat dont @ me

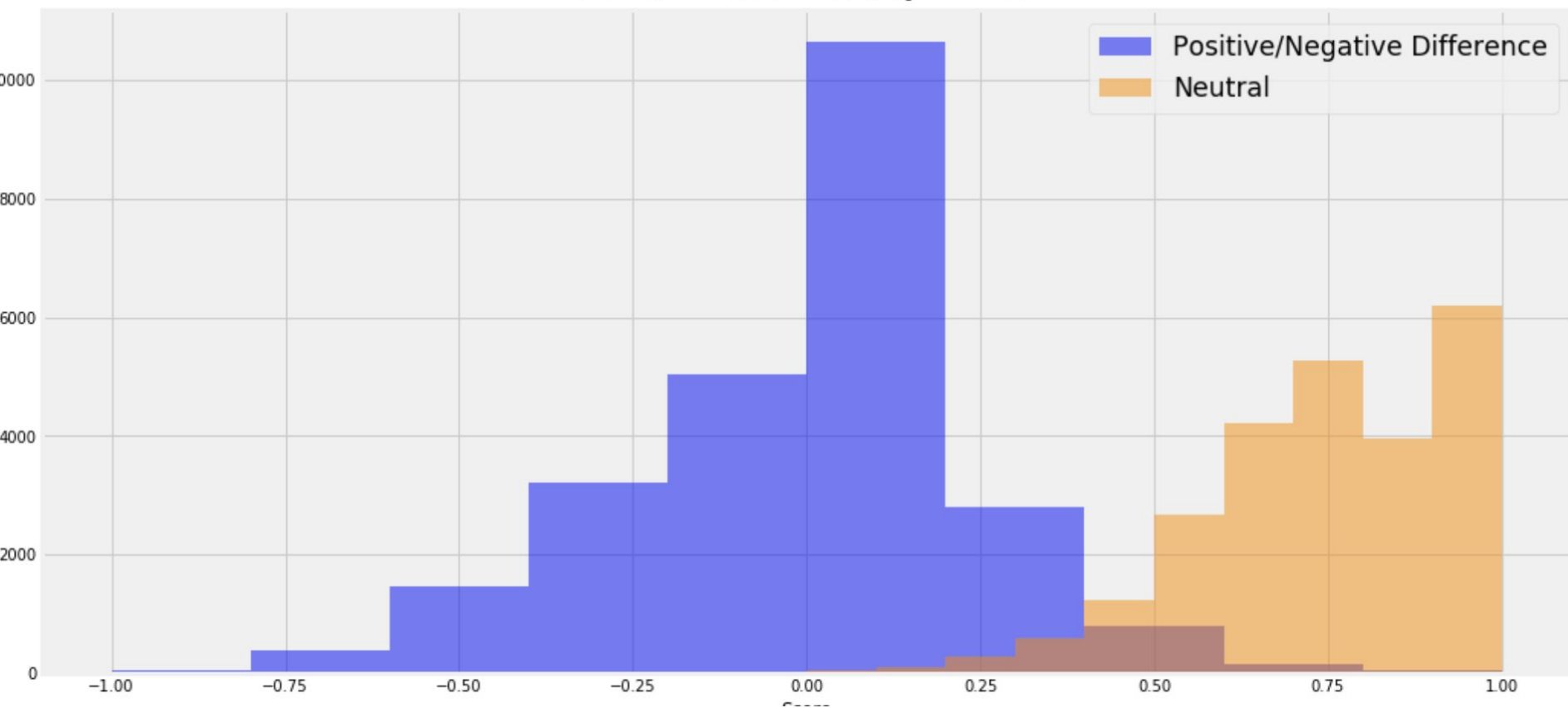
Vegans are delusional

# Polarity distributions of overall comments

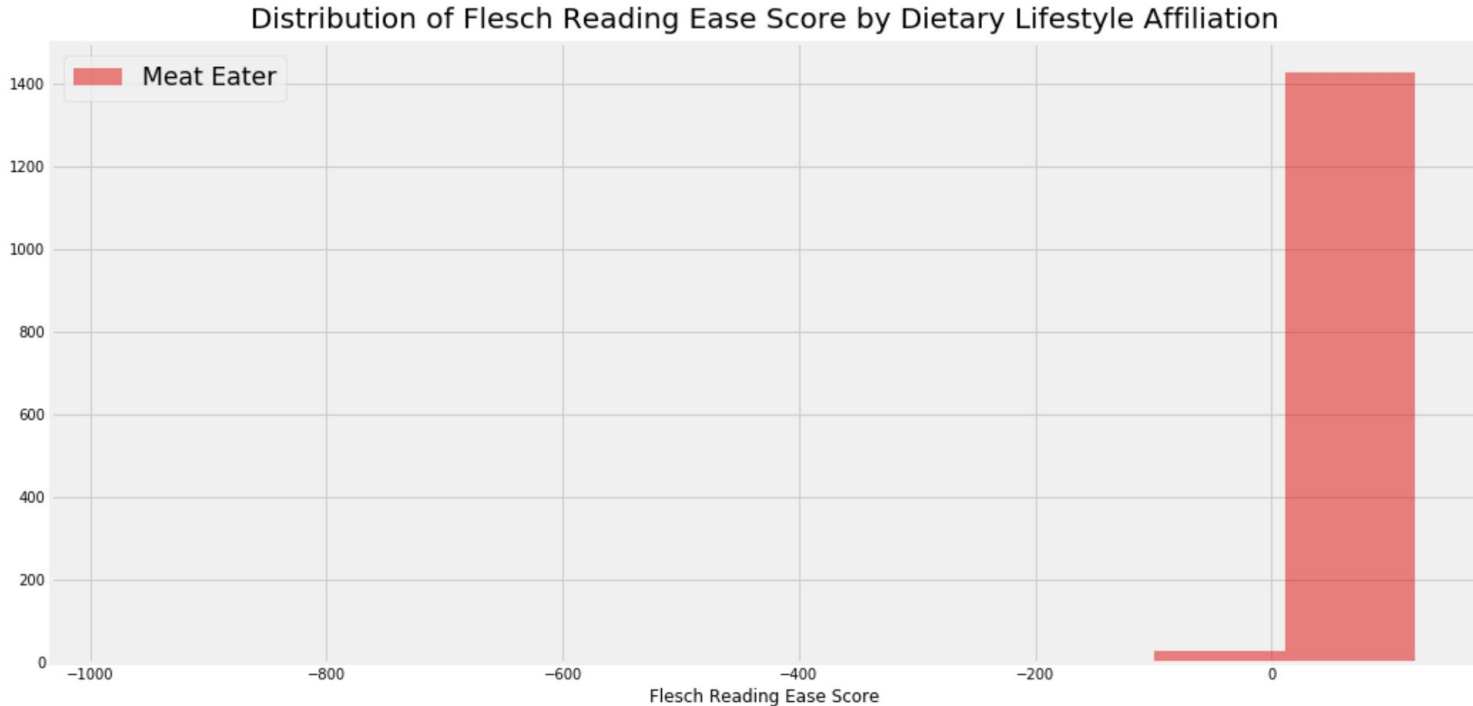




Distribution of Polarity Scores

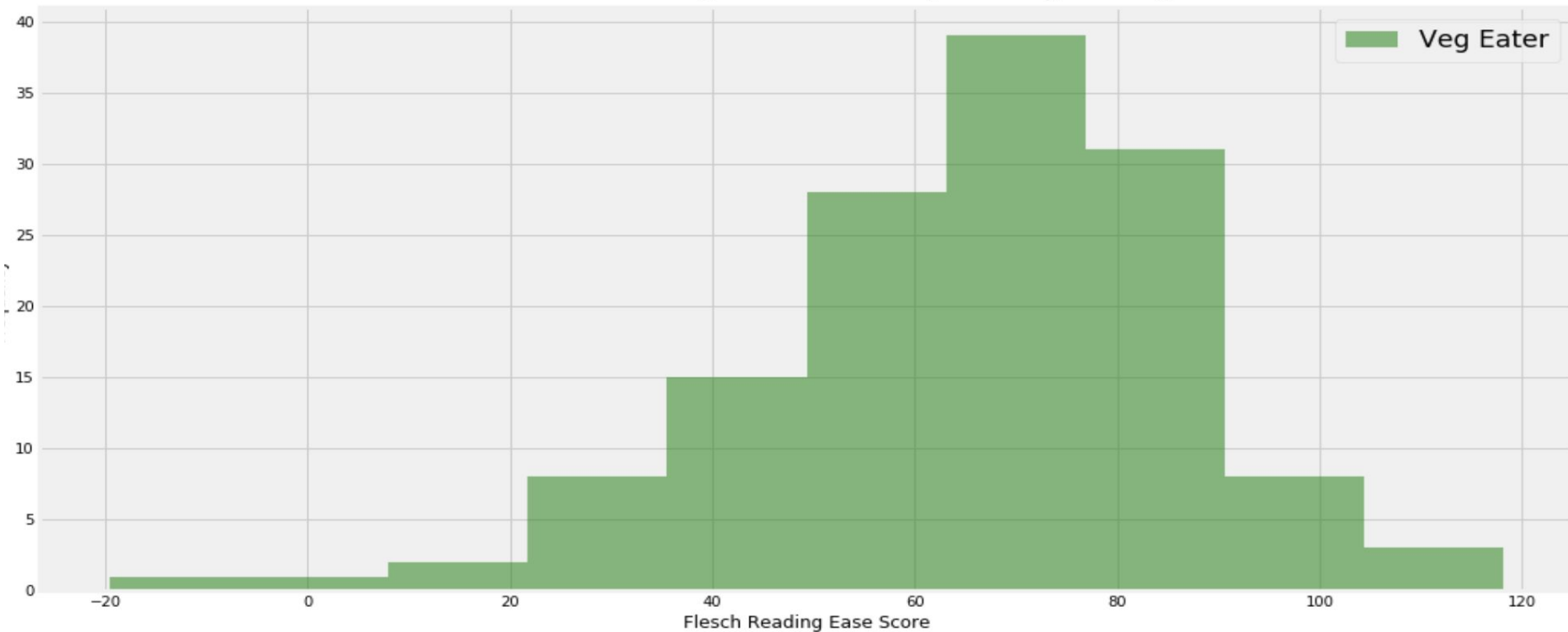


# Flesch Reading score: Lower Score indicates the text is harder to read



# Vegans had a higher score

Distribution of Flesch Reading Ease Score by Dietary Lifestyle Affiliation



# Example :

```
In [18]: 1 sub['commentText'][542]
```

```
Out[18]: '4:56 i kind of agree with him because YOU are stepping into the MOUNTAIN LION'S territory. f
or me it'd be like it's your fault that maybe you got injured. i think you should try and pro
tect yourself, even if it means injuring a huge animal in comparison to yourself. but when ar
e you going to beat a mountain lion or other huge animal without a serious weapon like a gun
or knife. if you tried to protect yourself and maybe it didn't work out as well as you wante
d, and you ended up getting injured by the animal, yes the animal did injure you, but it woul
d be your fault for not coming prepared when you knew(if you did, might want to do research b
efore you go do something like that) that it was a mountain lion's territory. the animal is j
ust trying to protect itself just like you are. it's just like you in your neighborhood and y
ou see a very unusual and maybe dangerous or violent animal(or it might be how they are portr
ayed) then you might harm it if they get to close or something along those lines. people and
animals just want to protect themselves,(&family) and their territory. the animal probably j
ust feels threatened just like you would. \nwasn't expecting on writing this much and it's pr
obably all over the place but oh well'
```

## **More to do!**

- **Unsupervised learning**
  - **PCA**
- **Compare polarity distributions to videos on the same subject & debates on other subjects**