

# Reddit Post Analysis of r/fountainpens and r/pens

...

A Potential Market Research and Sentiment Analysis Tool For the  
Goulet Pen Company

# The Goulet Pen Company - Online Only Fountain Pen Retail Business



- ❑ Started in 2009
- ❑ Specializes in fountain pens, ink, paper, and other accessories centered around fountain pens
- ❑ Small business that has no plans to expand to a 'brick and mortar' business model

# Natural Language Processing - Tool For Market Research & Sentiment Analysis

How can Goulet Pens keep track of customer opinions and what potential customers want?

The r/fountainpens and r/pens community! The r/fountainpens community has 293k member and r/pens has 106k.

Utilizing natural language processing can provide insight into what these communities are interested in as well as opinions of Goulet Pens.



# Testing Models - Pen or Fountain Pen? - No False Positives!

Goal: Create a model that can predict if a post belongs in the r/fountainpens or r/pen community. Emphasis on never classifying a pen post incorrectly as a fountain pen post. Accuracy is also important. r/fountainpens is considered the positive outcome.

	Posts
r/fountainpens	1557
r/pens	1243
Total Posts	2800

\*\* Four posts dropped due to null data.

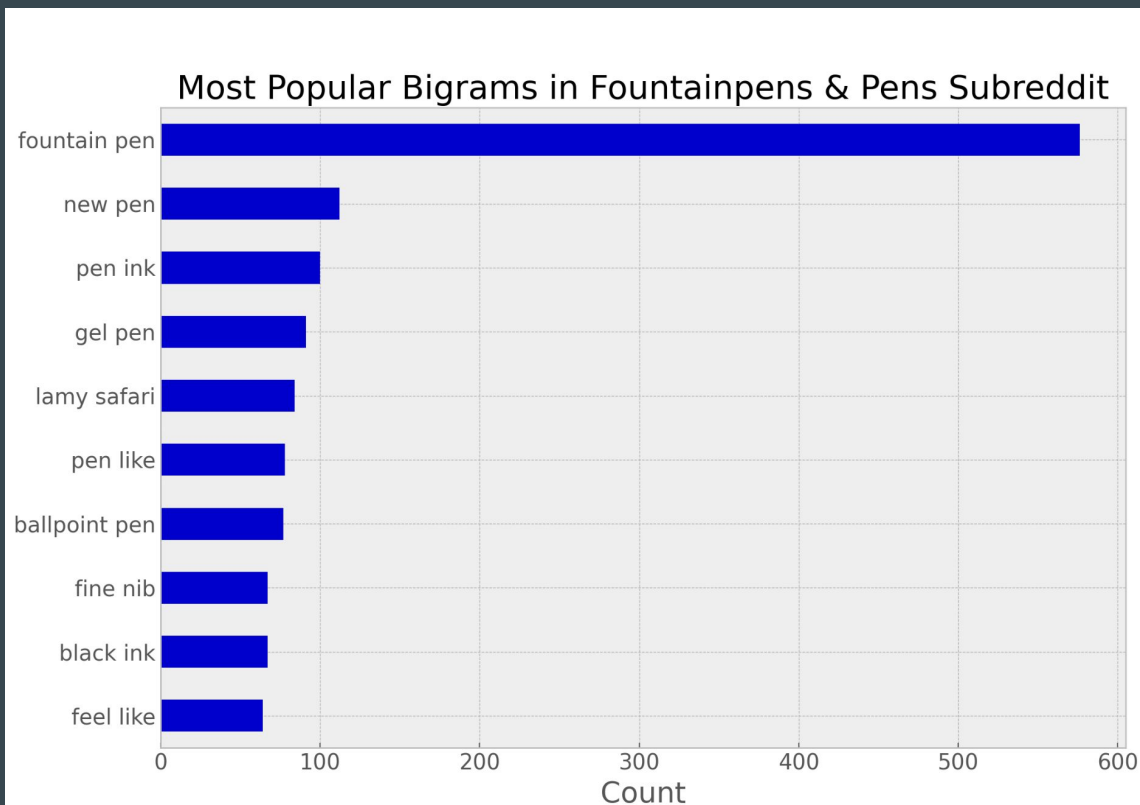
# Model Results - Bernoulli Naive Bayes vs. Logistic Regression

	Training Data Accuracy	Testing Data Accuracy	Prediction Accuracy	Sensitivity/Recall	<i>Specificity (Goal Metric)</i>	Precision
Logistic Regression & CountVectorizer	98%	89%	89%	92%	87%	90%
Logistic Regression & TF IDF	89%	87%	87%	91%	82%	86%
<i>Bernoulli Naive Bayes &amp; CountVectorizer</i>	90%	90%	90%	88%	91%	93%

- ❑ All models implemented SpaCy for natural language processing.
- ❑ Logistic Regression with CountVectorizer is overfit.
- ❑ Logistic Regression with TF IDF generalizes data better, but specificity is lower.
- ❑ Bernoulli Naive Bayes with CountVectorizer performs best with specificity and is generally a nice model.

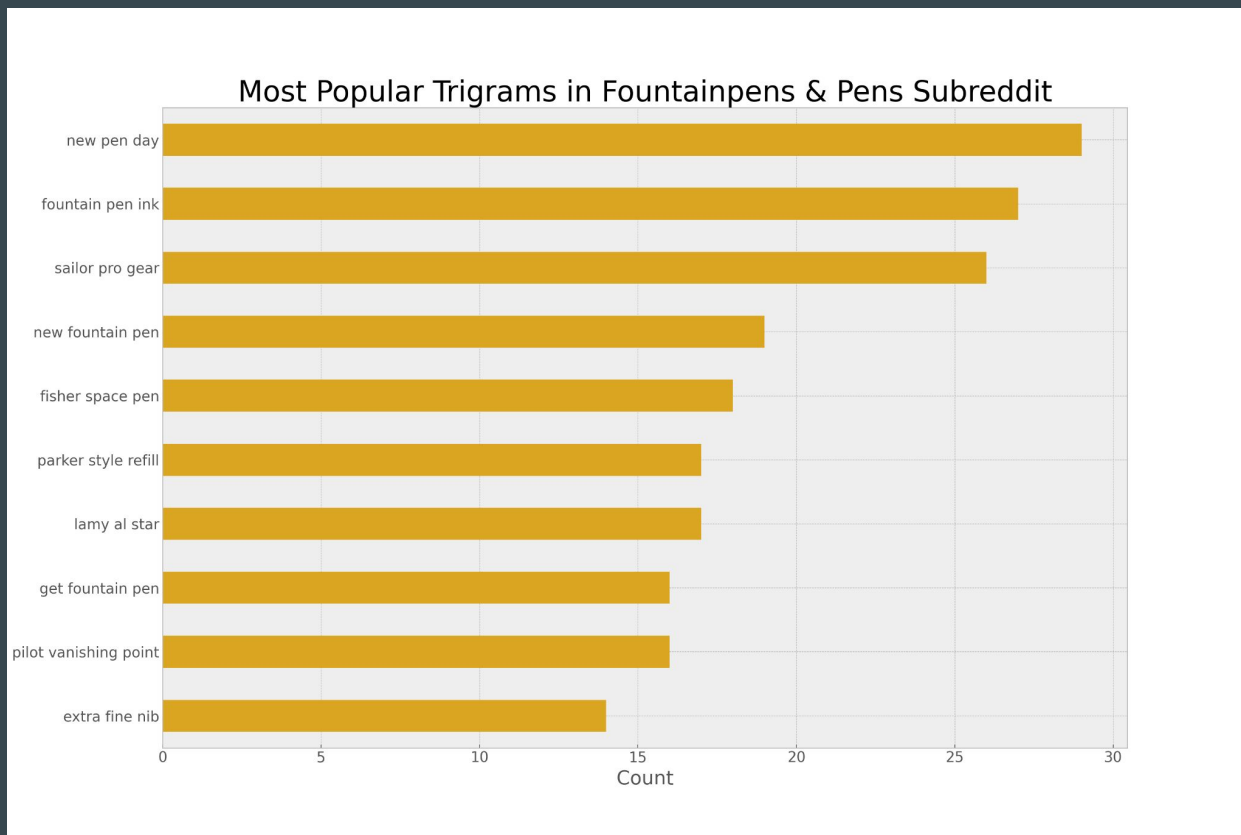
Base model is 56%

# Now For the Fun Stuff - What Are Pen Enthusiasts Talking About?



- ❑ ...it's fountain pens
- ❑ Lamy Safari
- ❑ Gel/Ballpoint pens
- ❑ Black Ink

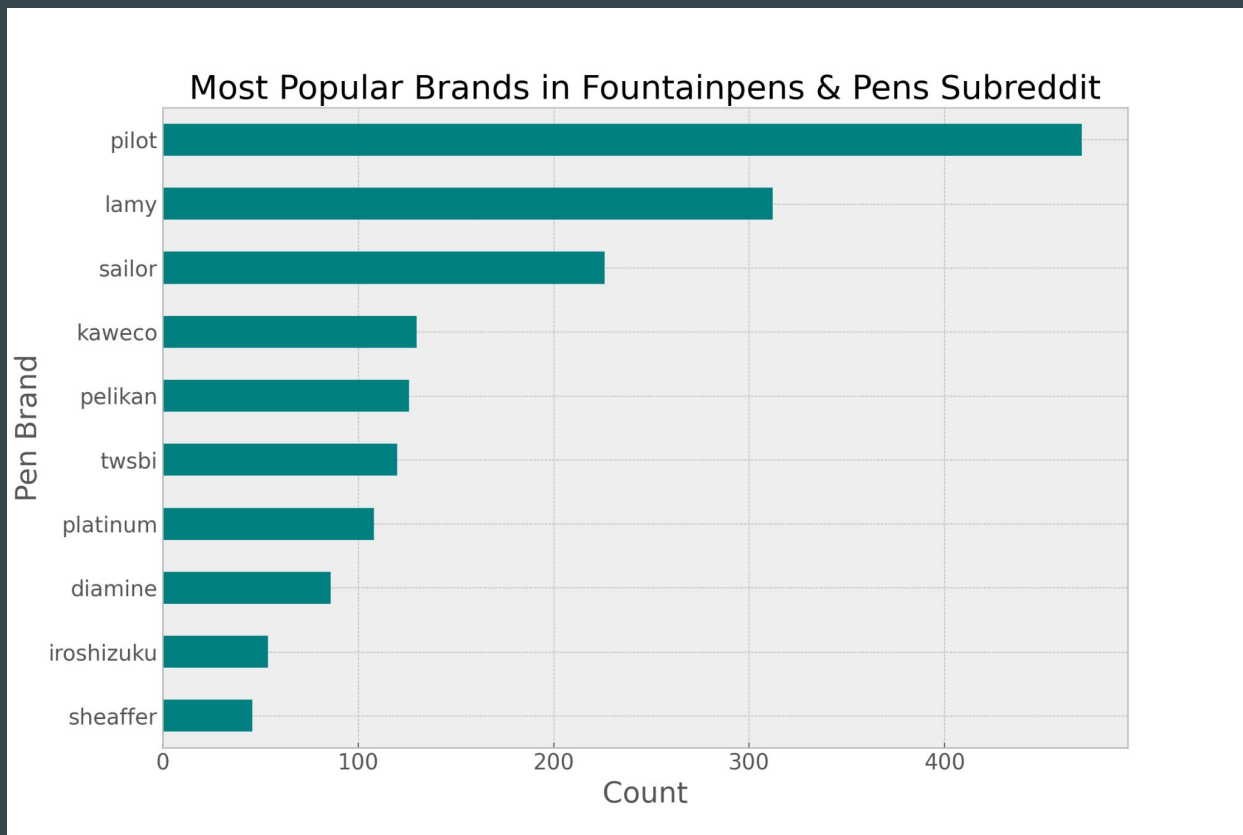
# Triple Threat Words - The Trigram!



- ❑ New Pen Day!
- ❑ Sailor Pro Gear
- ❑ Lamy AL Star
- ❑ Pilot Vanishing Point

Best performing model  
uses trigrams.

# Most Talked About Brands Sold at Goulet Pens



- Pilot pens and ink
- Lamy strikes again
- Sailor also near top

‘Goulet’ was #19 on list of brands generated for analysis.



# Suggestions and Improvements

- ❑ Goulet Pens offer a 'Beginner Set' which includes a Lamy Safari fountain pen and sample black ink.
- ❑ Emphasize marketing on Sailor Pro Gear and Pro Gear Slim pens as new editions are released.
- ❑ Further research implementing SpaCy natural language processing. Comments!
- ❑ Modify model to focus on sentiment analysis for Goulet Pens.

Thanks to Prab Jaswal &  
Radha Mohanty for  
helping me better  
understand SpaCy!



Photo by [Thomas Griggs](#)