

Talk Business

Analysis of conversations of
Small Business owners
and Startup founders.

Marta Ghiglioni

Data Science Immersive - General Assembly

*Slide Credit to SlideCarnival
Images from Freepicks*



A new blog for business owner

**What should we talk about to
attract business owners?**

What are main differences and
similarities of Small Business
owners' and Startup founders'
conversations?



Project Workflow





Reddit is an important platform of peer knowledge sharing, vastly used by professionals and people discussing many different topics.

Choice of Subreddits

Two subreddits were selected, among the most active and popular in Finance & Business:

- ▶ r/Startups
- ▶ r/SmallBusiness

Using PushShift API were performed 100 queries, pulling 500 post each on a span of 20 days.

Collecting a total of 19.993 posts.

Exploratory Data Analysis

Engagement

1. Comments

Analysis were run to investigate differences in the engagement of users in the two subreddits.

No significant differences were found, so no new features were added.

| | count | mean | std | min | 25% | 50% | 75% | max |
|---------------|--------|-----------|-----------|-----|-----|-----|------|-------|
| subreddit | | | | | | | | |
| smallbusiness | 7406.0 | 10.675263 | 20.056605 | 0.0 | 2.0 | 5.0 | 11.0 | 425.0 |
| startups | 4523.0 | 12.445722 | 29.673758 | 0.0 | 2.0 | 5.0 | 13.0 | 692.0 |

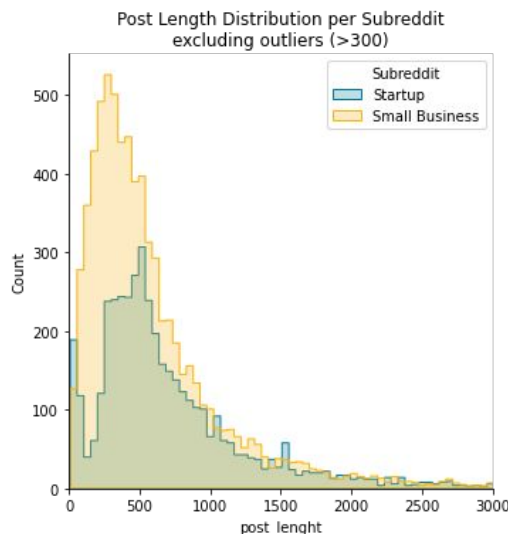
Exploratory Data Analysis

Engagement

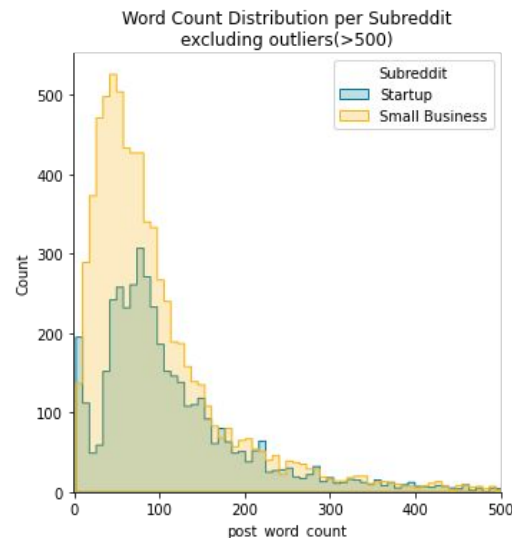
Analysis were run to investigate differences in the engagement of users in the two subreddits.

No significant differences were found, so no new features were added.

2. Words Count



3. Post Length



You can also split your content

Small Business

advice business client company
customer day **employee** good ha
help **idea** just know like looking
make **money** month need new
owner people **product** question
really sale service small start
thanks thing time use wa want
way **website** work year

Startup

app business company customer
experience feedback **founder free**
good help idea know like looking
make marketing **money** month
need new **people** post product
question really **service** start
startup **team** thing think time use
user want way website work
working year

Models

I run the following models

- ▶ Baseline
- ▶ Manual Random Forest
- ▶ Grid Search Random Forest
- ▶ Count Vectorizer MNBayes
- ▶ TF-IDF Vectorier MNBayes
- ▶ Logistic regression

Using Pipelines
to streamline
the workflow

Scoring the models
using classification
report and checking
for Precision and
Recall

Models

Baseline Model - DummyClassifier

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.62 | 1.00 | 0.77 | 1851 |
| 1 | 0.00 | 0.00 | 0.00 | 1131 |
| accuracy | | | 0.62 | 2982 |
| macro avg | 0.31 | 0.50 | 0.38 | 2982 |
| weighted avg | 0.39 | 0.62 | 0.48 | 2982 |

Manual Random Forest

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.84 | 0.93 | 0.88 | 1851 |
| 1 | 0.86 | 0.72 | 0.78 | 1131 |
| accuracy | | | 0.85 | 2982 |
| macro avg | 0.85 | 0.82 | 0.83 | 2982 |
| weighted avg | 0.85 | 0.85 | 0.85 | 2982 |

Grid Searched Random Forest

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.85 | 0.91 | 0.88 | 1851 |
| 1 | 0.83 | 0.74 | 0.78 | 1131 |
| accuracy | | | 0.85 | 2982 |
| macro avg | 0.84 | 0.83 | 0.83 | 2982 |
| weighted avg | 0.84 | 0.85 | 0.84 | 2982 |

Models

Count Vectorizer MNBayes

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.86 | 0.85 | 0.85 | 1851 |
| 1 | 0.76 | 0.78 | 0.77 | 1131 |
| accuracy | | | 0.82 | 2982 |
| macro avg | 0.81 | 0.81 | 0.81 | 2982 |
| weighted avg | 0.82 | 0.82 | 0.82 | 2982 |

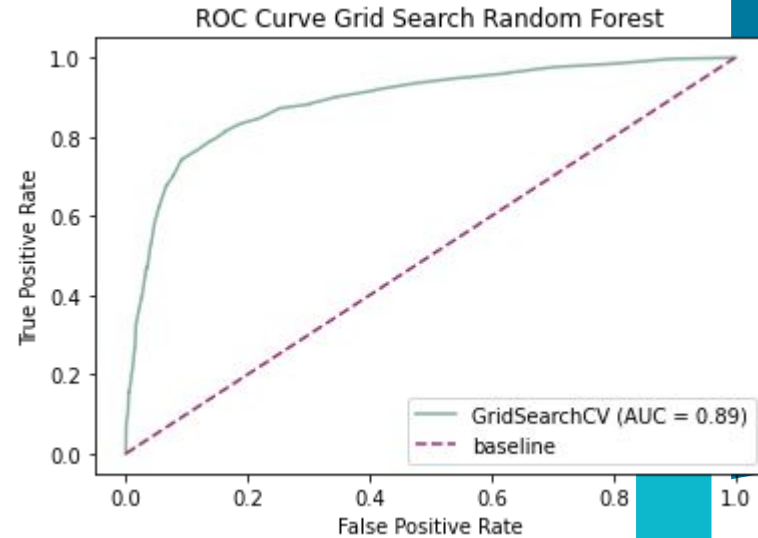
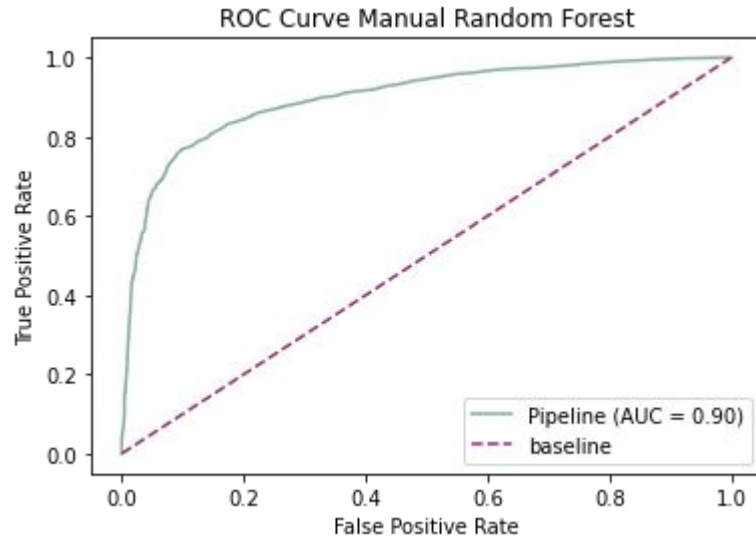
TF-IDF Vectorizer MNBayes

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.84 | 0.91 | 0.87 | 1851 |
| 1 | 0.83 | 0.71 | 0.76 | 1131 |
| accuracy | | | 0.84 | 2982 |
| macro avg | 0.84 | 0.81 | 0.82 | 2982 |
| weighted avg | 0.84 | 0.84 | 0.83 | 2982 |

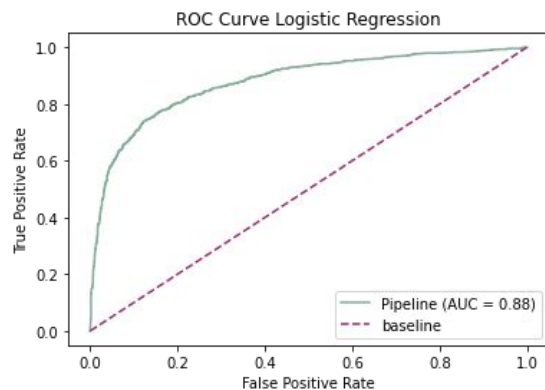
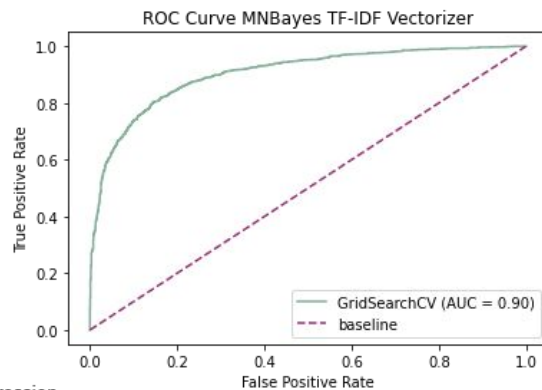
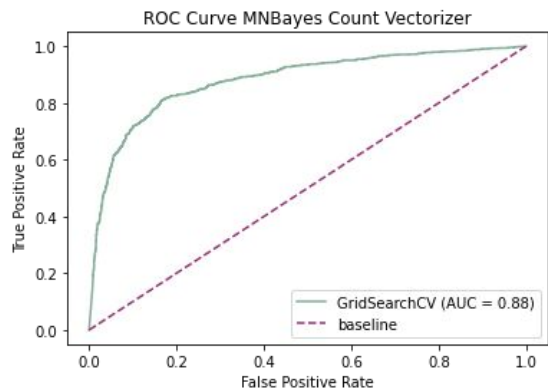
Logistic Regression

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.84 | 0.88 | 0.86 | 1851 |
| 1 | 0.79 | 0.73 | 0.76 | 1131 |
| accuracy | | | 0.82 | 2982 |
| macro avg | 0.82 | 0.81 | 0.81 | 2982 |
| weighted avg | 0.82 | 0.82 | 0.82 | 2982 |

Models



Models



Random Forest

Was the best model

85% Recall

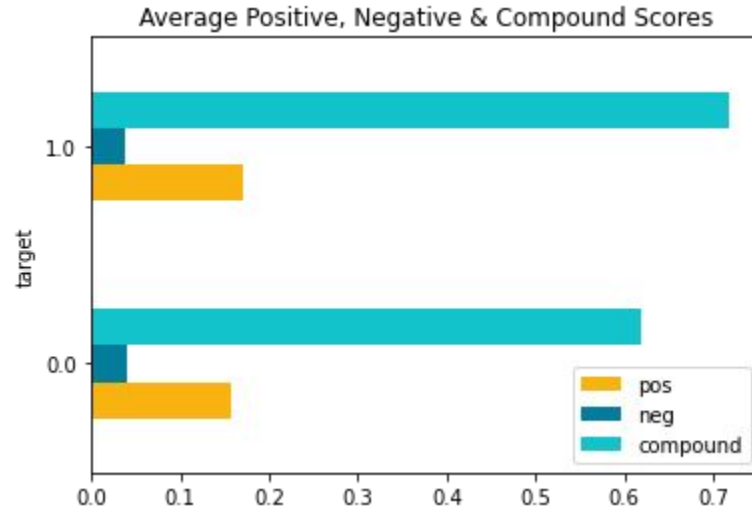
On weighted average

Reliability

The models run all performed in the same range.

Sentiment Analysis

Sentiment In both subreddits are extremely similar. For this reason the sentiment does not appear to be a good predictor of the subreddit. This also confirms how the two categories are similar, in the use that users do of the subreddit and the tone with which they discuss topics.



Take Away

- ▶ Founders and SB Owners have an overlap area of about 15% in their conversations
- ▶ Although they are two very distinct subsects, and they deserve to be separately targeted.
- ▶ Startups conversations are more focused on Investment, Apps and free resources.
- ▶ SB are interested in taxes, advices and industry specific insights (e.g. Coffe, restaurant, etc)



What is next?

We could implement additional features and extend the analysis to other subreddits.

