

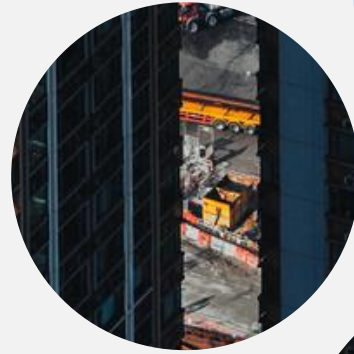


Predicting News Genres

Benjamin Jaron

Table of Contents

- Overview
- Business Problem
- Data Insights
- Model and Score
- Final Recommendation



Overview

Our goal is to use predictive modeling in order to help political readers save time finding online articles.

We deliver a model that is extremely useful at discarding unrelated content.



Business Problem



Readers avoid internet

- A Nielsen Norman Group report shows most readers prefer reading news offline
- They find it hard to find preferred content

Limited time online

- When it comes to news, people try to maximize time by personally selecting articles offline
- Spend 15s looking at an article before moving on

Our solution

- Build a model that discards unrelated content
- This gives an online political reader an easier time selecting preferred articles

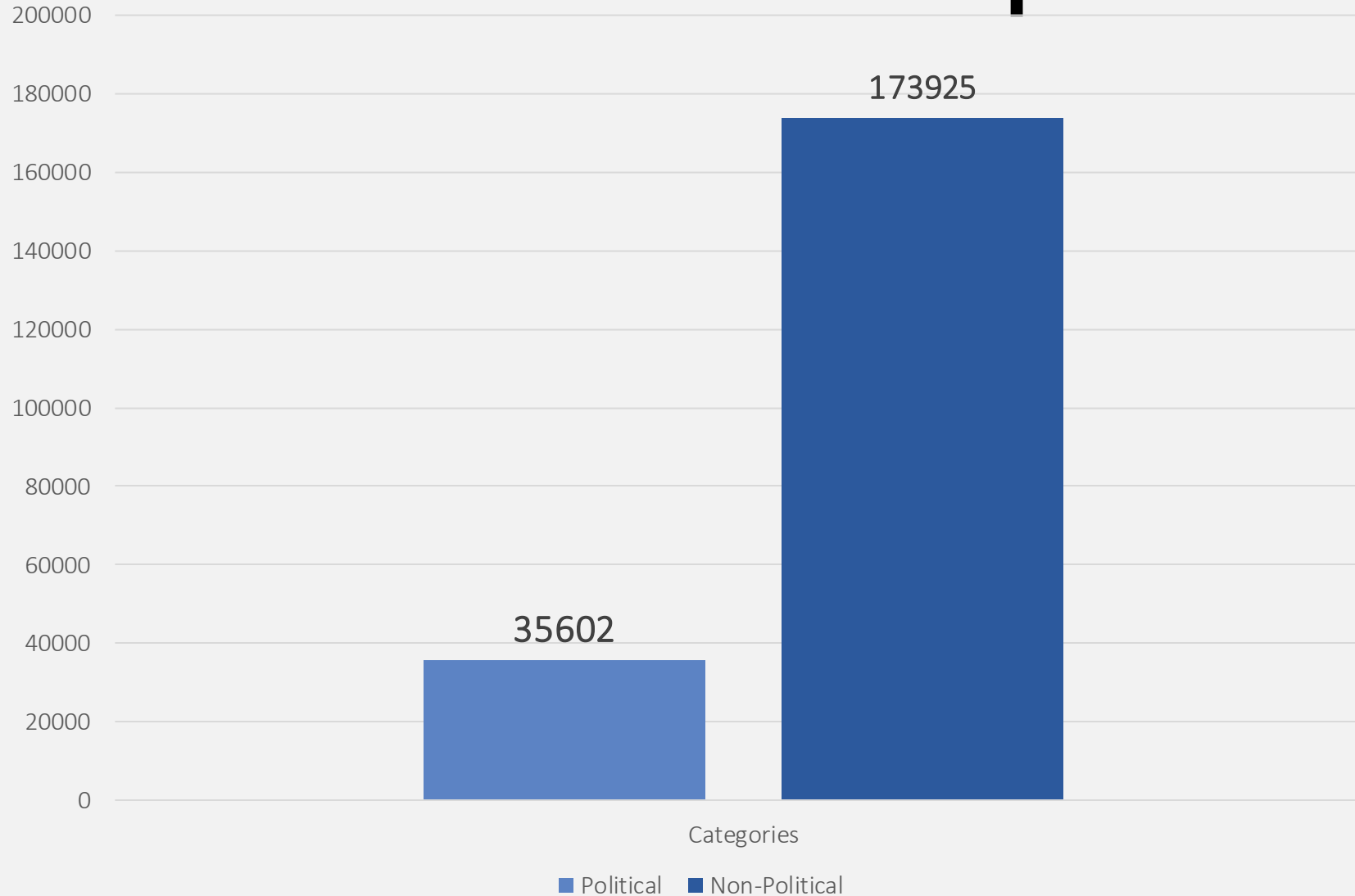


Insights from our data

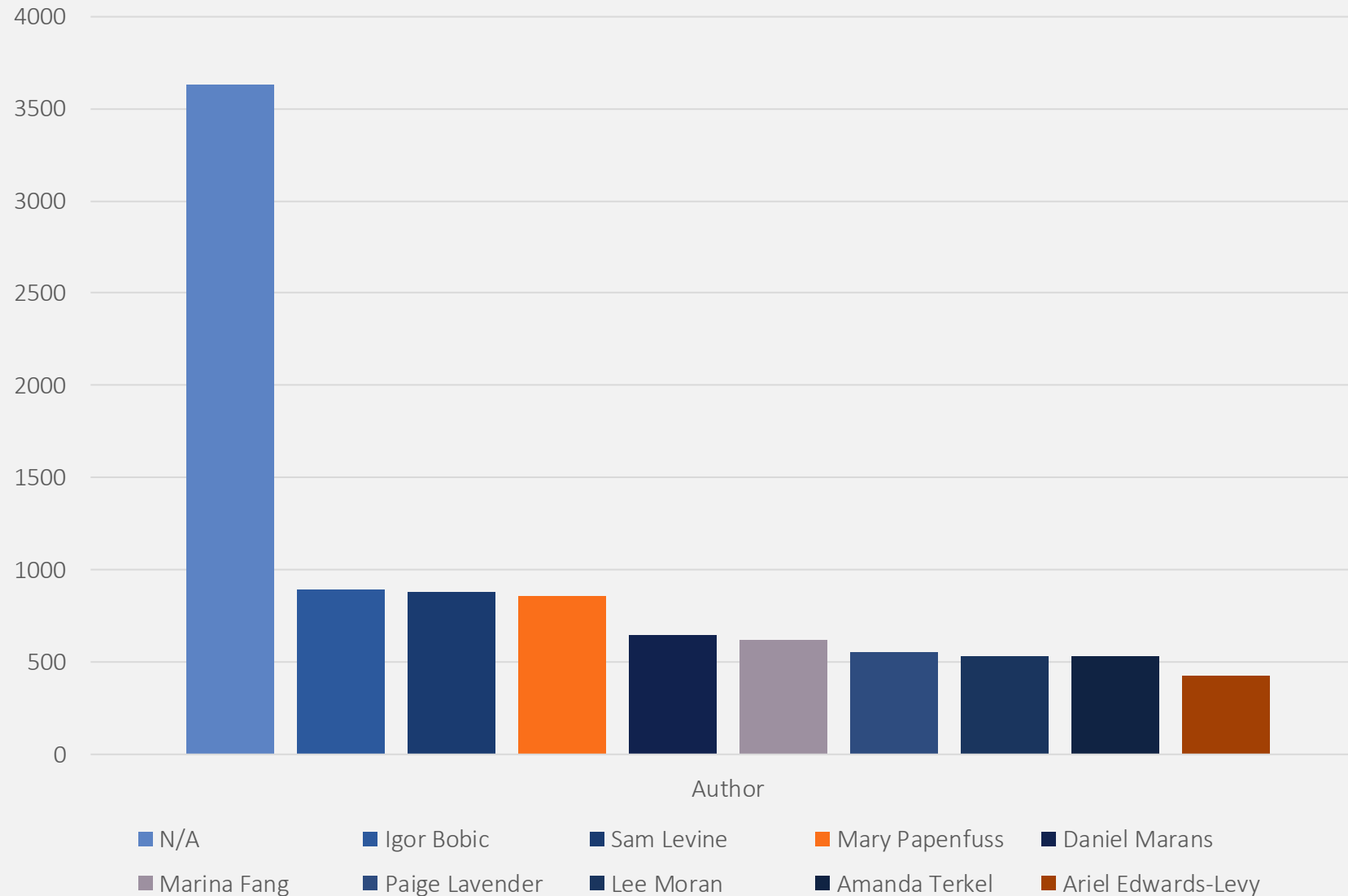
Let's look!



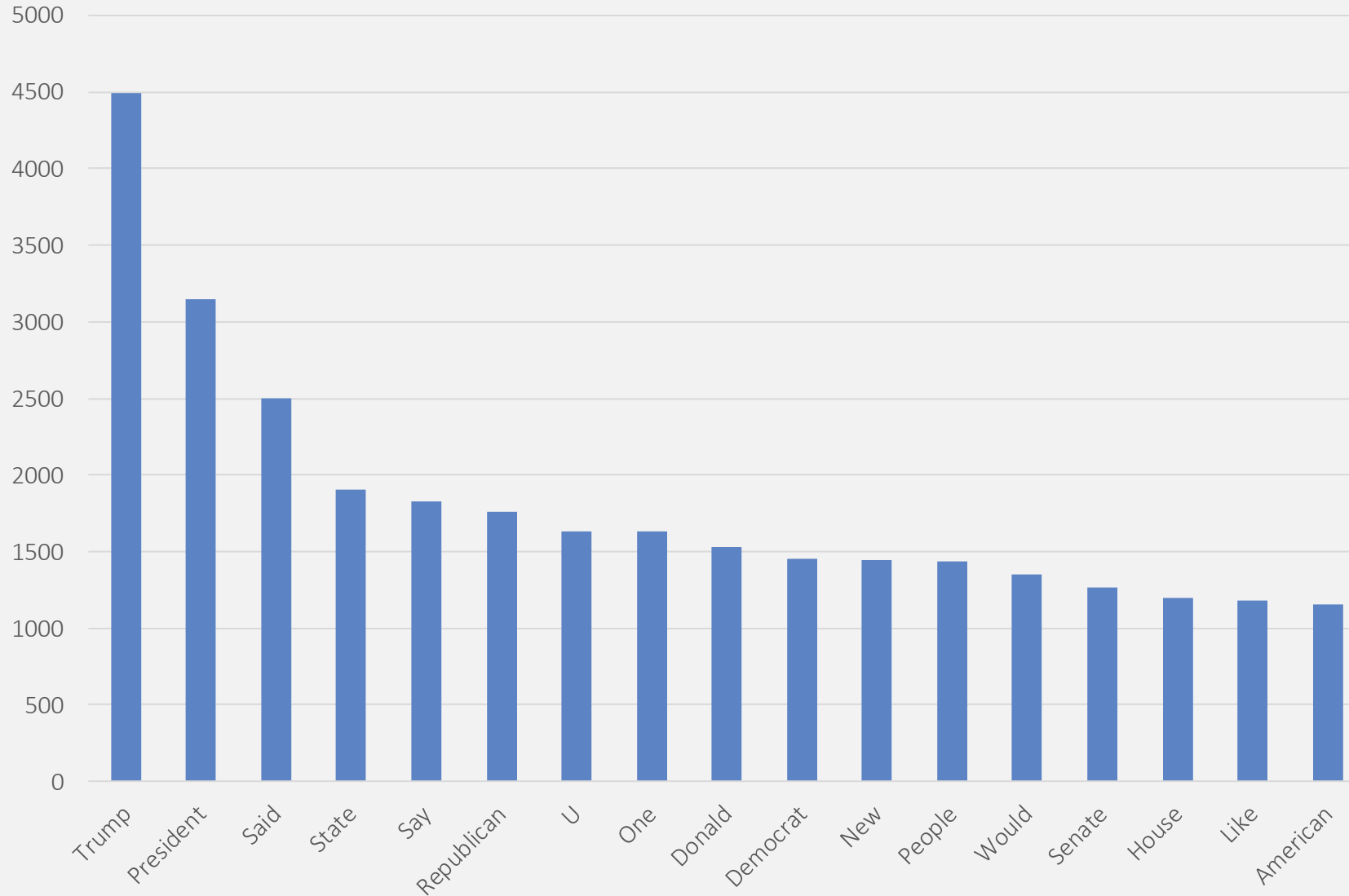
Most articles are non-political



Most articles not authored by individual



Most words in political articles are political

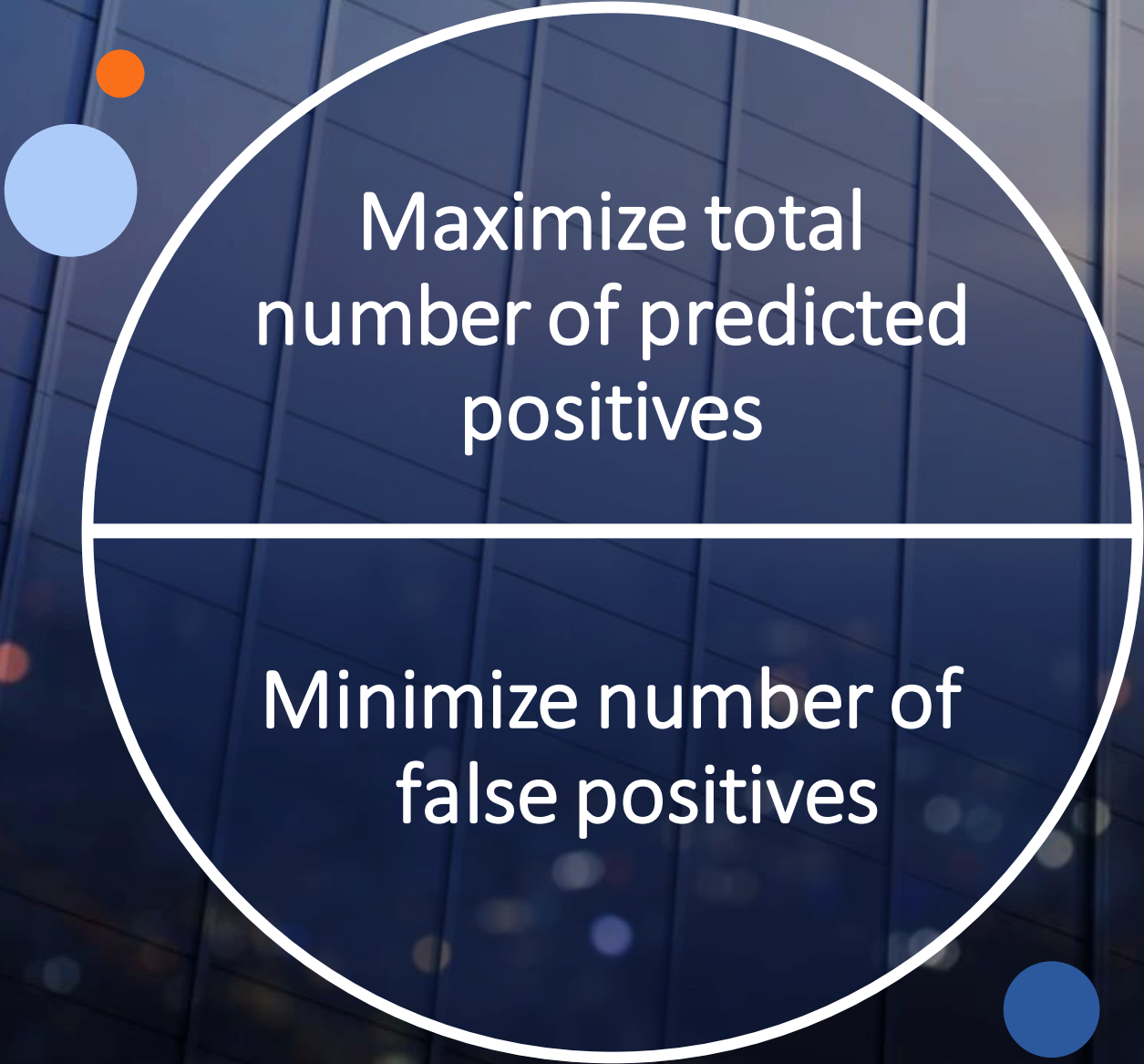




Modeling and score


Possible Outcomes

True Positive Our model correctly predicts a political article	False Positive Our model predicts political, but really article is non-political
False Negative Our model predicts non-political, but really article is political	True Negative Our model correctly predicts a non-political article



Maximize total
number of predicted
positives

Minimize number of
false positives



0.60 score when
trying to balance

94% precision at
eliminating non-
political articles

Recommendations



Use our model in order to eliminate majority of non-political articles



Rely on the reader's natural eye to pick political articles from remainder



Develop more advanced modeling techniques in the future, and work with stakeholder to see what needs improvement

The end

We're open to questions

Benjamin Jaron
bmjaron@gmail.com

