

```
1 !pip install textthero
```

```
↳ collecting textthero
```

```
  Downloading textthero-1.1.0-py3-none-any.whl (24 kB)
```

```
collecting unicode>=1.1.1
```

```
  Downloading Unidecode-1.3.2-py3-none-any.whl (235 kB)
```

```
    |████████████████████| 235 kB 6.2 MB/s
```

```
requirement already satisfied: scikit-learn>=0.22 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: tqdm>=4.3 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: pandas>=1.0.2 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-packages
```

```
collecting nltk>=3.3
```

```
  Downloading nltk-3.6.5-py3-none-any.whl (1.5 MB)
```

```
    |████████████████████| 1.5 MB 9.0 MB/s
```

```
requirement already satisfied: matplotlib>=3.1.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: spacy<3.0.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: gensim<4.0,>=3.6.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: plotly>=4.2.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: wordcloud>=1.5.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: smart-open>=1.2.1 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: scipy>=0.18.1 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: six>=1.5.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: cyclr>=0.10 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages
```

```
collecting regex>=2021.8.3
```

```
  Downloading regex-2021.11.10-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (749 kB)
```

```
    |████████████████████| 749 kB 22.3 MB/s
```

```
requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: retrying>=1.3.3 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: blis<0.5.0,>=0.4.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: srsly<1.1.0,>=1.0.2 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: thinc==7.4.0 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: importlib-metadata>=0.20 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: typing-extensions>=3.6.4 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages
```

```
requirement already satisfied: pillow in /usr/local/lib/python3.7/dist-packages
```

```
installing collected packages: regex, unidecode, nltk, textthero
```

```

Attempting uninstall: regex
Found existing installation: regex 2019.12.20
Uninstalling regex-2019.12.20:
Successfully uninstalled regex-2019.12.20
Attempting uninstall: nltk

```

## Visualize the data

```

1 import textthero as hero
2 import pandas as pd
3 df = pd.read_csv('/content/MIA.csv')
4 df.head(2)

```

```

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.

```

1 to 2 of 2 entries

Filter




index	text	sentiment
0	A proposed change to the flight paths above South Florida could turn the local airspace into a sky-high Interstate 95, according to critics of the Federal Aviation Administration's plan to consolidate current flight paths into narrower and more concentrated bands of traffic	0
1	Members of the Miami International Airport Noise Abatement Advisory Committee, who heard the FAA's pitch for its Next Generation Air Transportation System during a recent meeting, expressed concerns about the added noise pollution that more	1

```
1 df['text'] = hero.clean(df['text'])
```

```
1 df
```

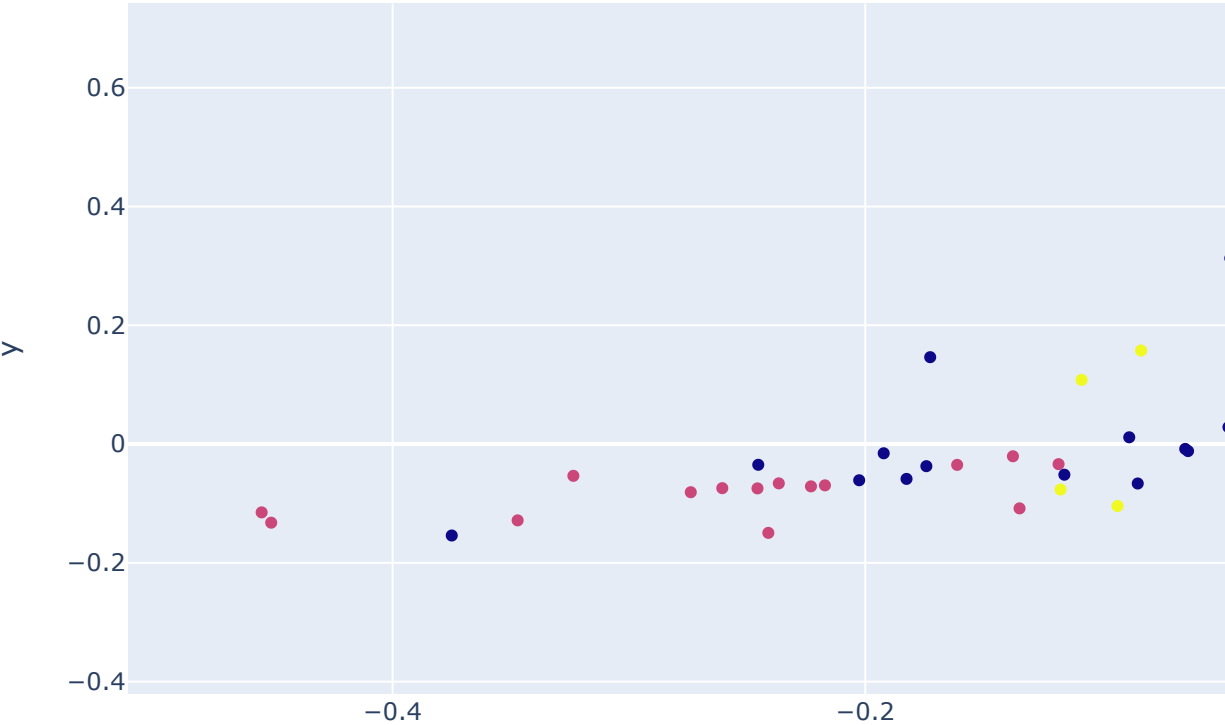
text sentiment 

```
1 df['tfidf'] = (  
2     hero.tfidf(df['text'], max_features=100)  
3 )  
4 df[['tfidf', 'sentiment']].head(2)
```

	tfidf	sentiment	
0	[0.0, 0.0, 0.23592474613902928, 0.0, 0.0, 0.0,...	0	
1	[0.0, 0.0, 0.0, 0.0, 0.46761693057722087, 0.0,...	1	
...	... regardless into day thing really taken ...	...	

```
1 df['pca'] = hero.pca(df['tfidf'])  
2 hero.scatterplot(  
3     df,  
4     col='pca',  
5     color='sentiment',  
6     title="PCA MIA Engagement"  
7 )
```

PCA MIA Engagement



```

1 df['named_entities'] = (
2     hero.named_entities(df['text'])
3 )
4 df[['named_entities', 'sentiment']].head(2)

```

	named_entities	sentiment
0	[(florida, GPE, 35, 42)]	0
1	[(miami international airport, FAC, 8, 35), (f...	1



```

1 NUM_TOP_WORDS = 5
2 hero.top_words(df['text'])[:NUM_TOP_WORDS]

```

```

noise      15
faa         15
flight      14
paths       13
nextgen     10
Name: text, dtype: int64

```

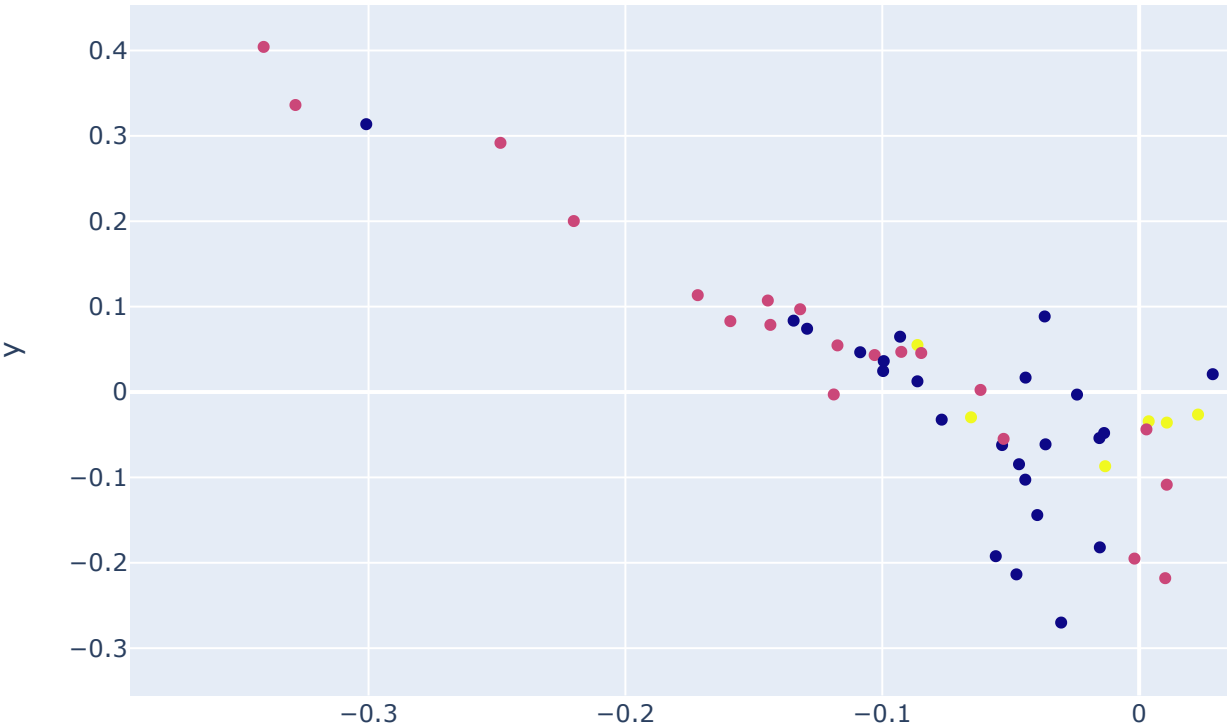
## Summary

```

1 import textthero as hero
2 import pandas as pd
3
4 df = pd.read_csv('/content/MIA.csv')
5 df['pca'] = (
6     df['text']
7     .pipe(hero.clean)
8     .pipe(hero.tfidf)
9     .pipe(hero.pca)
10 )
11
12 hero.scatterplot(df, col='pca', color='sentiment', title="PCA MIA Engagement")

```

PCA MIA Engagement



✓ 0s completed at 10:20 AM

