

Import required packages

24-nov

- t-SNE followed by spectral clustering

TOC

- Modeling and Evaluation 1 :

Train and adjust parameters

- t-SNE
- end of file

```
In [40]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
#%matplotlib inline
import warnings
warnings.simplefilter('ignore',DeprecationWarning)
import seaborn as sns
import time
import copy

from pylab import rcParams
#import hdbscan

from sklearn.model_selection import ShuffleSplit
from sklearn.preprocessing import StandardScaler

#from sklearn.datasets import make_blobs

from sklearn.ensemble import RandomForestClassifier
from sklearn.calibration import CalibratedClassifierCV
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import StratifiedKFold, cross_val_score

from sklearn import metrics
from sklearn import metrics as mt
from sklearn.metrics import log_loss
from sklearn.metrics import accuracy_score as acc
from sklearn.metrics import confusion_matrix as conf
from sklearn.metrics import f1_score, precision_score, recall_score, classification_report
from sklearn.metrics import accuracy_score
from sklearn.metrics import precision_recall_fscore_support as score

from sklearn.cluster import KMeans

from tabulate import tabulate

from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"

from __future__ import print_function
```

Read in cleaned dataset from .csv file

```
In [41]: data_dir = '../data/'  
data_file = 'mashable_clean_dataset_for_lab_03.csv'  
  
file_2_read = data_dir + data_file  
df = pd.read_csv(file_2_read)  
  
df_cluster = copy.deepcopy(df)  
  
#del df_cluster['data_channel']  
  
# ... read in original data set to retrieve 'shares' values  
  
data_file = 'OnlineNewsPopularity.csv'  
  
file_2_read = data_dir + data_file  
df_ONP = pd.read_csv(file_2_read)  
  
df_ONP.columns = df_ONP.columns.str.strip()  
  
df_ONP = df_ONP[['shares']]  
  
df_ONP['ln_shares'] = np.log(df_ONP['shares']+1)  
  
df_ONP['popular'] = np.where(df_ONP['shares'] > 1400, True, False)
```

```
In [42]: col_names = df_cluster.columns.values.tolist()  
col_names
```

```
Out[42]: ['n_tokens_title',
          'num_keywords',
          'data_channel_is_lifestyle',
          'data_channel_is_entertainment',
          'data_channel_is_socmed',
          'kw_avg_max',
          'weekday_is_monday',
          'weekday_is_tuesday',
          'weekday_is_wednesday',
          'weekday_is_thursday',
          'weekday_is_friday',
          'is_weekend',
          'global_subjectivity',
          'global_rate_positive_words',
          'rate_positive_words',
          'max_positive_polarity',
          'min_negative_polarity',
          'max_negative_polarity',
          'title_sentiment_polarity',
          'abs_title_subjectivity',
          'ln_n_tokens_content',
          'ln_num_hrefs',
          'ln_num_imgs',
          'ln_num_videos',
          'ln_kw_min_min',
          'ln_kw_avg_min',
          'ln_kw_min_max',
          'ln_kw_avg_avg',
          'ln_self_reference_avg_shares',
          'ln_LDA_00',
          'ln_LDA_01',
          'ln_LDA_02',
          'ln_LDA_03',
          'ln_LDA_04',
          'ln_global_rate_negative_words',
          'ln_min_positive_polarity',
          'ln_abs_title_sentiment_polarity',
          'ln_shares']
```

Out[42]:

	count	mean	std	min	25%	50%	75%
n_tokens_title	39644.0	10.398749	2.114037	2.0	9.000000	10.000000	12.000000
num_keywords	39644.0	7.223767	1.909130	1.0	6.000000	7.000000	9.000000
data_channel_is_lifestyle	39644.0	0.052946	0.223929	0.0	0.000000	0.000000	0.000000
data_channel_is_entertainment	39644.0	0.178009	0.382525	0.0	0.000000	0.000000	0.000000
data_channel_is_socmed	39644.0	0.058597	0.234871	0.0	0.000000	0.000000	0.000000
kw_avg_max	39644.0	1.913205	1.000000	0.0	1.271003	1.800325	2.442234
weekday_is_monday	39644.0	0.168020	0.373889	0.0	0.000000	0.000000	0.000000
weekday_is_tuesday	39644.0	0.186409	0.389441	0.0	0.000000	0.000000	0.000000
weekday_is_wednesday	39644.0	0.187544	0.390353	0.0	0.000000	0.000000	0.000000
weekday_is_thursday	39644.0	0.183306	0.386922	0.0	0.000000	0.000000	0.000000
weekday_is_friday	39644.0	0.143805	0.350896	0.0	0.000000	0.000000	0.000000
is_weekend	39644.0	0.130915	0.337312	0.0	0.000000	0.000000	0.000000
global_subjectivity	39644.0	0.443370	0.116685	0.0	0.396167	0.453457	0.508333
global_rate_positive_words	39644.0	0.039625	0.017429	0.0	0.028384	0.039023	0.050279
rate_positive_words	39644.0	0.682150	0.190206	0.0	0.600000	0.710526	0.800000
max_positive_polarity	39644.0	0.756728	0.247786	0.0	0.600000	0.800000	1.000000
min_negative_polarity	39644.0	0.478056	0.290290	0.0	0.300000	0.500000	0.700000
max_negative_polarity	39644.0	0.892500	0.095373	0.0	0.875000	0.900000	0.950000
title_sentiment_polarity	39644.0	1.071425	0.265450	0.0	1.000000	1.000000	1.150000
abs_title_subjectivity	39644.0	0.341843	0.188791	0.0	0.166667	0.500000	0.500000
ln_n_tokens_content	39644.0	5.889971	1.255442	0.0	5.509388	6.016157	6.575076
ln_num_refs	39644.0	2.156564	0.809445	0.0	1.609438	2.197225	2.708050
ln_num_imgs	39644.0	1.116427	0.973755	0.0	0.693147	0.693147	1.609438
ln_num_videos	39644.0	0.400420	0.680486	0.0	0.000000	0.000000	0.693147
ln_kw_min_min	39644.0	1.174410	1.733030	0.0	0.000000	0.000000	1.791759
ln_kw_avg_min	39644.0	5.302209	1.132463	0.0	4.968076	5.470168	5.883322
ln_kw_min_max	39644.0	5.045209	4.521016	0.0	0.000000	7.244942	8.974745
ln_kw_avg_avg	39644.0	7.976327	0.489467	0.0	7.776304	7.962442	8.189031
ln_self_reference_avg_shares	39644.0	6.667697	3.280186	0.0	6.889782	7.696667	8.556606
ln_LDA_00	39644.0	0.148724	0.194635	0.0	0.024742	0.032842	0.215884
ln_LDA_01	39644.0	0.117056	0.164989	0.0	0.024705	0.032801	0.140485
ln_LDA_02	39644.0	0.172661	0.207322	0.0	0.028171	0.039224	0.288345
ln_LDA_03	39644.0	0.176795	0.216061	0.0	0.028171	0.039221	0.319008
ln_LDA_04	39644.0	0.186227	0.212166	0.0	0.028173	0.039920	0.336462
ln_global_rate_negative_words	39644.0	0.016419	0.010571	0.0	0.009569	0.015221	0.021506

```
Out[42]: ['n_tokens_title',
 'num_keywords',
 'data_channel_is_lifestyle',
 'data_channel_is_entertainment',
 'data_channel_is_socmed',
 'kw_avg_max',
 'weekday_is_monday',
 'weekday_is_tuesday',
 'weekday_is_wednesday',
 'weekday_is_thursday',
 'weekday_is_friday',
 'is_weekend',
 'global_subjectivity',
 'global_rate_positive_words',
 'rate_positive_words',
 'max_positive_polarity',
 'min_negative_polarity',
 'max_negative_polarity',
 'title_sentiment_polarity',
 'abs_title_subjectivity',
 'ln_n_tokens_content',
 'ln_num_hrefs',
 'ln_num_imgs',
 'ln_num_videos',
 'ln_kw_min_min',
 'ln_kw_avg_min',
 'ln_kw_min_max',
 'ln_kw_avg_avg',
 'ln_self_reference_avg_shares',
 'ln_LDA_00',
 'ln_LDA_01',
 'ln_LDA_02',
 'ln_LDA_03',
 'ln_LDA_04',
 'ln_global_rate_negative_words',
 'ln_min_positive_polarity',
 'ln_abs_title_sentiment_polarity',
 'ln_shares']
```

```
In [43]: # set required variables for model comparison

comparison_tbl = pd.DataFrame(columns = [
    'model_name',
    'n_clusters',
    'inertia',
    'silhouette',
    'process_time'])

i_index = []
i_index = 0

# preparation for cross validation and model comparison, each classifier is
# appended once model is fit

models = []
```

Table of Contents

t-SNE

t-SNE

```
In [44]: from sklearn.manifold import TSNE

X1 = df_cluster
X1['ln_shares'] = df_ONP['ln_shares']
X1['popular'] = df_ONP['popular']

X1 = X1.sample(frac = 0.30)

X1_ln_shares = X1['ln_shares']
X1_popular = X1['popular']

columns_to_drop = ['ln_shares', 'popular']
X1.drop(columns_to_drop, axis = 1, inplace = True)

tic = time.clock()

tsne = TSNE(n_components = 2, verbose = 1, perplexity = 5, n_iter = 300)

tsne_results = tsne.fit_transform(X1)

toc = time.clock()
print (toc - tic)

[t-SNE] Computing 16 nearest neighbors...
[t-SNE] Indexed 11893 samples in 0.028s...
[t-SNE] Computed neighbors for 11893 samples in 6.274s...
[t-SNE] Computed conditional probabilities for sample 1000 / 11893
[t-SNE] Computed conditional probabilities for sample 2000 / 11893
[t-SNE] Computed conditional probabilities for sample 3000 / 11893
[t-SNE] Computed conditional probabilities for sample 4000 / 11893
[t-SNE] Computed conditional probabilities for sample 5000 / 11893
[t-SNE] Computed conditional probabilities for sample 6000 / 11893
[t-SNE] Computed conditional probabilities for sample 7000 / 11893
[t-SNE] Computed conditional probabilities for sample 8000 / 11893
[t-SNE] Computed conditional probabilities for sample 9000 / 11893
[t-SNE] Computed conditional probabilities for sample 10000 / 11893
[t-SNE] Computed conditional probabilities for sample 11000 / 11893
[t-SNE] Computed conditional probabilities for sample 11893 / 11893
[t-SNE] Mean sigma: 0.739581
[t-SNE] KL divergence after 250 iterations with early exaggeration: 96.793800
[t-SNE] Error after 300 iterations: 4.367745
108.82638899999995
```

```
In [67]: from ggplot import *
from ggplot import scale_fill_brewer
from matplotlib import cm

df_tsne = copy.deepcopy(X1)
df_tsne['x-tsne'] = tsne_results[:,0]
df_tsne['y-tsne'] = tsne_results[:,1]

col_names = df_tsne.columns.values.tolist()

for col in col_names :
    plt.figure(figsize=(12, 8));
    plt.subplot(111, axisbg='darkgrey');
    plt.scatter(df_tsne['x-tsne'], df_tsne['y-tsne'],
                c = df_tsne[col],
                cmap = plt.cm.Spectral,
                s = 50,
                linewidths = 0,
                alpha = 0.30)
    plt.colorbar()
    plt.xlabel('t-SNE axis 1')
    plt.ylabel('t-SNE axis 2')
    plt.title(col)
    plt.grid(True)
    plt.show();
```

```
Out[67]: <matplotlib.figure.Figure at 0x7fe7996887f0>

/home/mcdevitt/anaconda3/lib/python3.6/site-packages/matplotlib/cbook.py:136: MatplotlibDeprecationWarning: The axisbg attribute was deprecated in version 2.0. Use facecolor instead.
    warnings.warn(message, mplDeprecation, stacklevel=1)

Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe799e6c860>

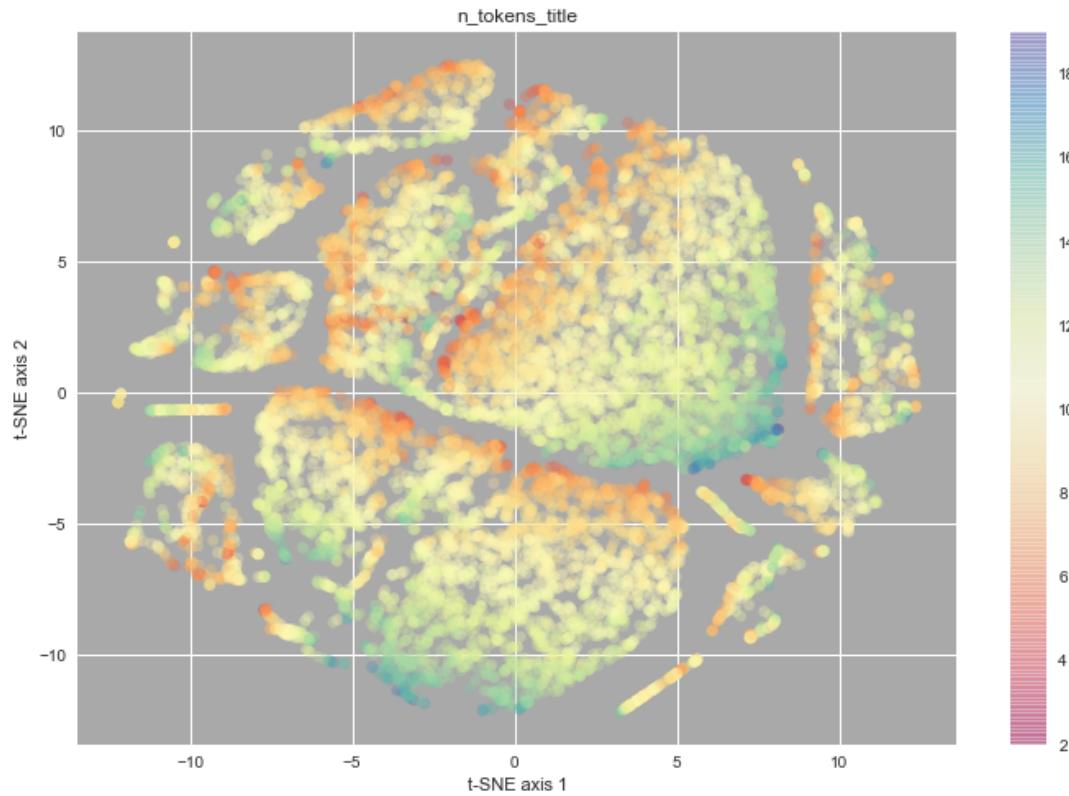
Out[67]: <matplotlib.collections.PathCollection at 0x7fe79a0766d8>

Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7997749e8>

Out[67]: <matplotlib.text.Text at 0x7fe7b6d8cf98>

Out[67]: <matplotlib.text.Text at 0x7fe7b6a09a20>

Out[67]: <matplotlib.text.Text at 0x7fe799407c50>
```



```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6b52390>

Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6b36e80>

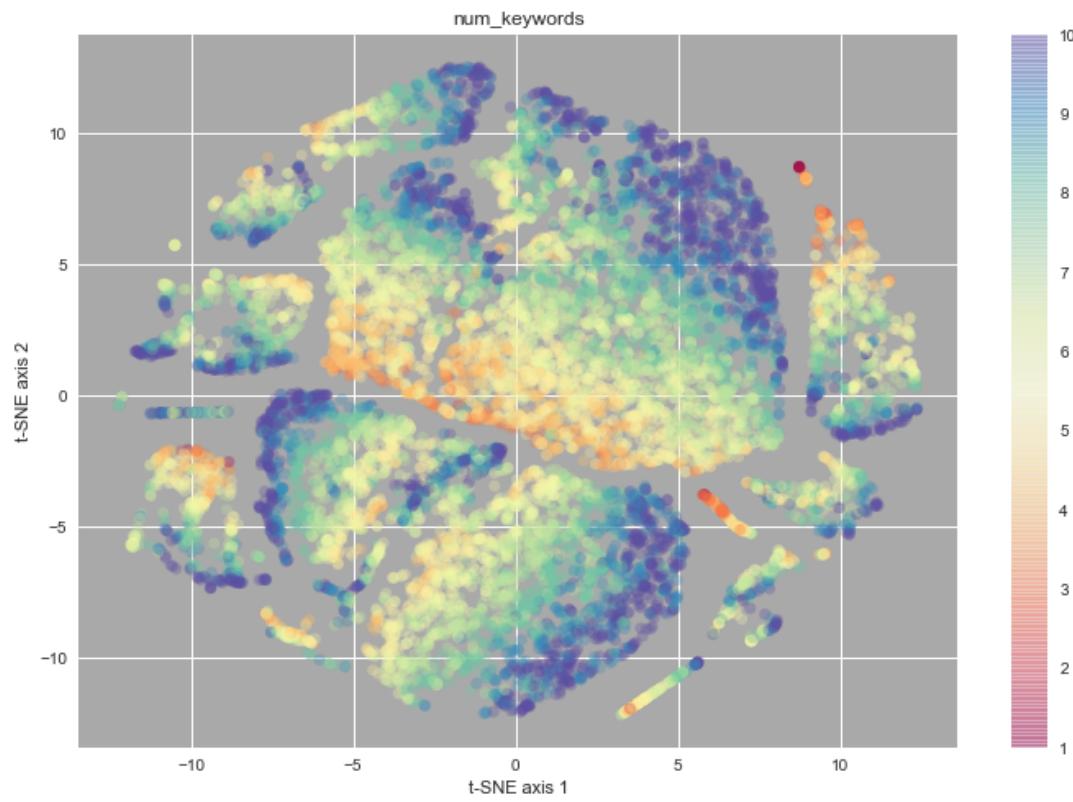
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b67b2c18>

Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b69576d8>

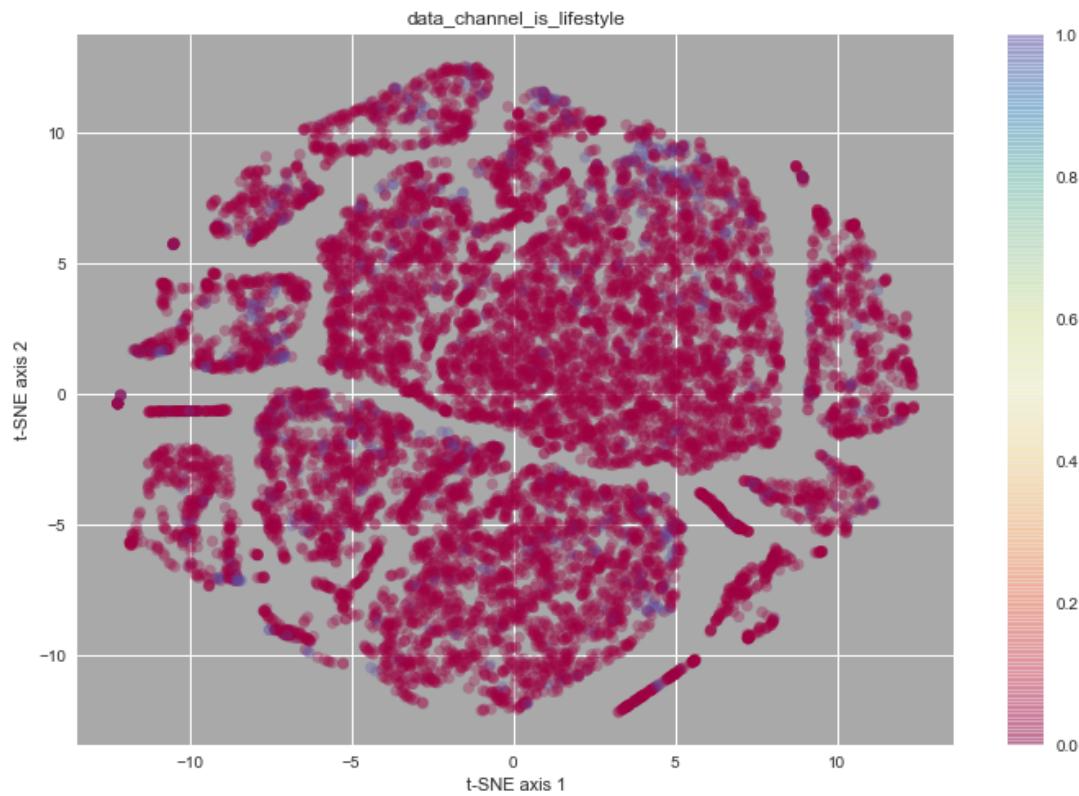
Out[67]: <matplotlib.text.Text at 0x7fe7b6bb8390>

Out[67]: <matplotlib.text.Text at 0x7fe7b6bdd358>

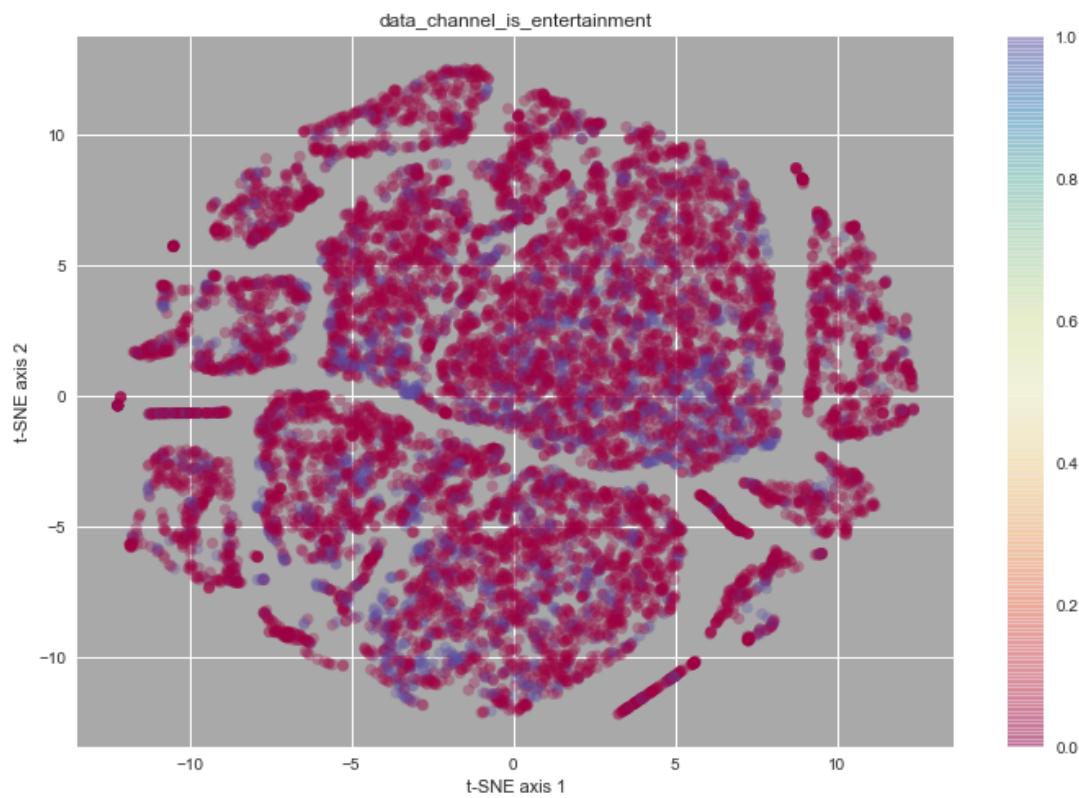
Out[67]: <matplotlib.text.Text at 0x7fe7b6ada128>
```



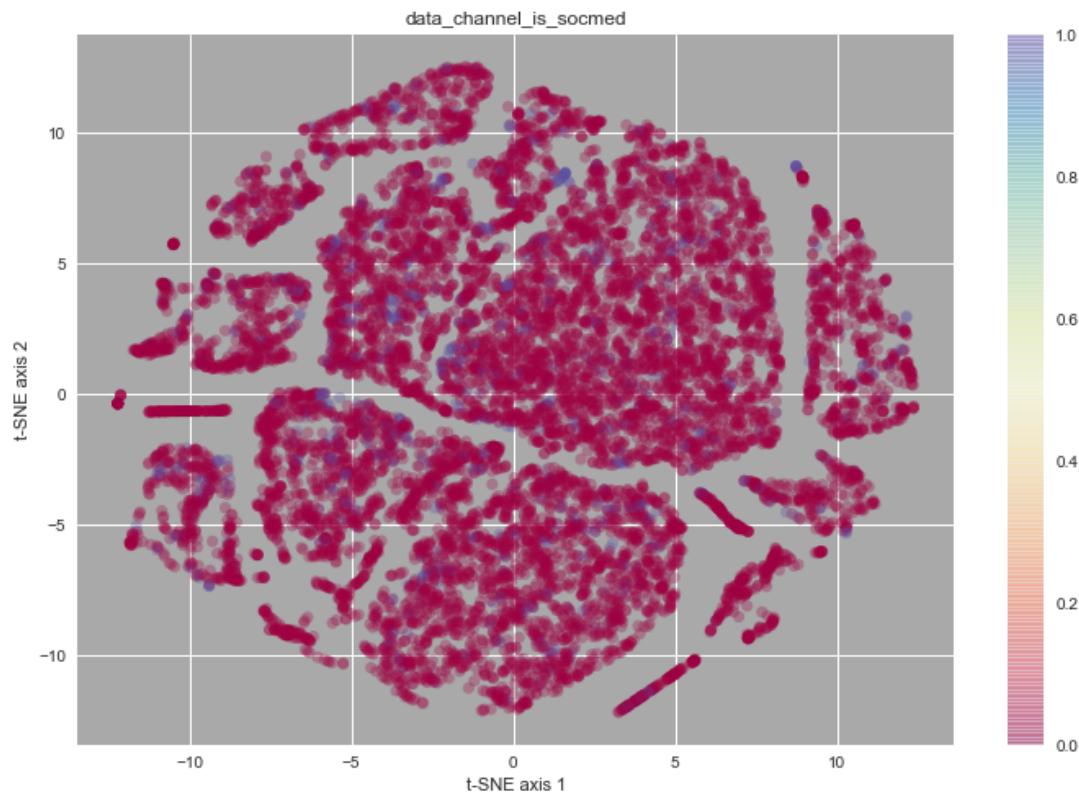
```
Out[67]: <matplotlib.figure.Figure at 0x7fe799bc1748>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b66058d0>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b672feb8>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b693fda0>
Out[67]: <matplotlib.text.Text at 0x7fe7b652e3c8>
Out[67]: <matplotlib.text.Text at 0x7fe7b654b2e8>
Out[67]: <matplotlib.text.Text at 0x7fe7b655ffd0>
```



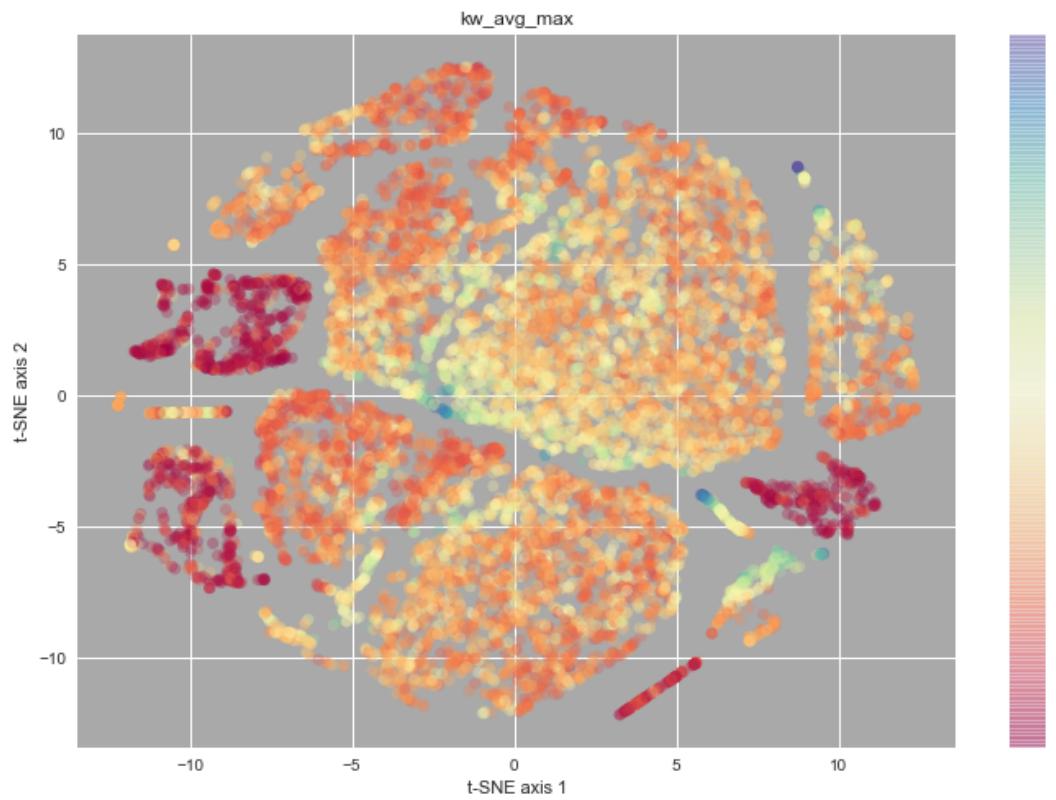
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b68dde48>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6b4e080>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b651db70>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b642db70>
Out[67]: <matplotlib.text.Text at 0x7fe7b68bd7b8>
Out[67]: <matplotlib.text.Text at 0x7fe7b6884160>
Out[67]: <matplotlib.text.Text at 0x7fe7b64ecf60>
```



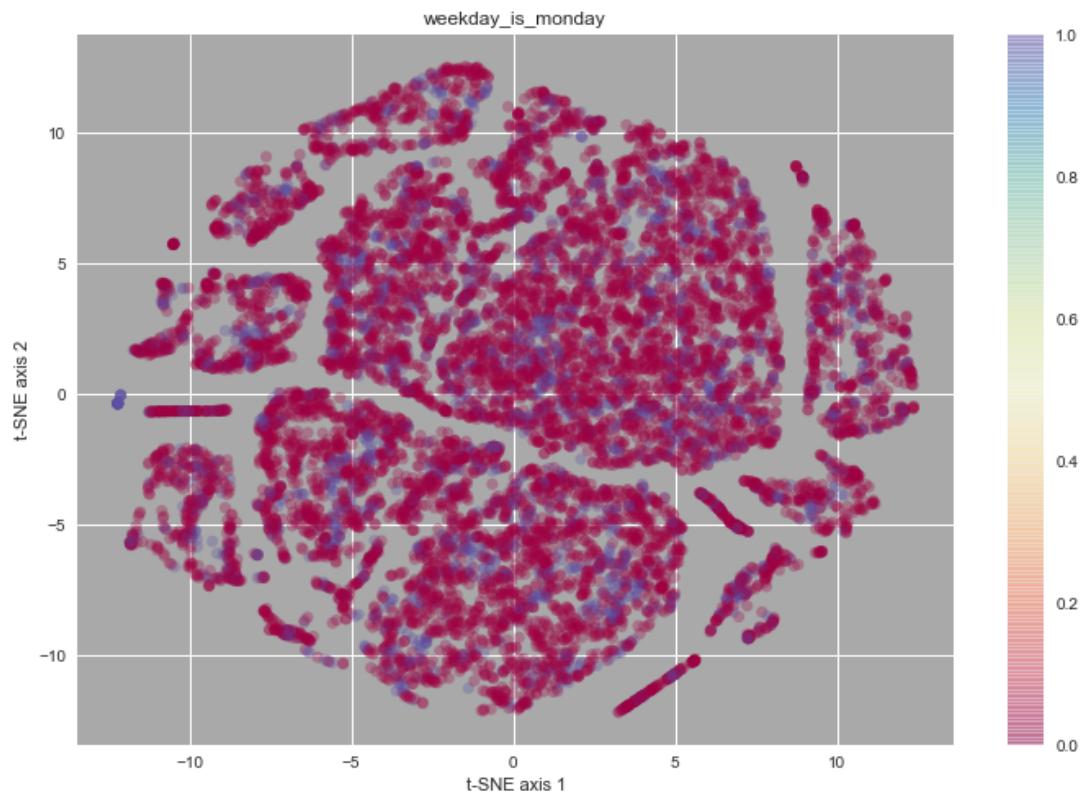
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b654c588>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b63ee978>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b68bdb70>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b65bfe0>
Out[67]: <matplotlib.text.Text at 0x7fe7b644f9b0>
Out[67]: <matplotlib.text.Text at 0x7fe7b65cd518>
Out[67]: <matplotlib.text.Text at 0x7fe7b68cc940>
```



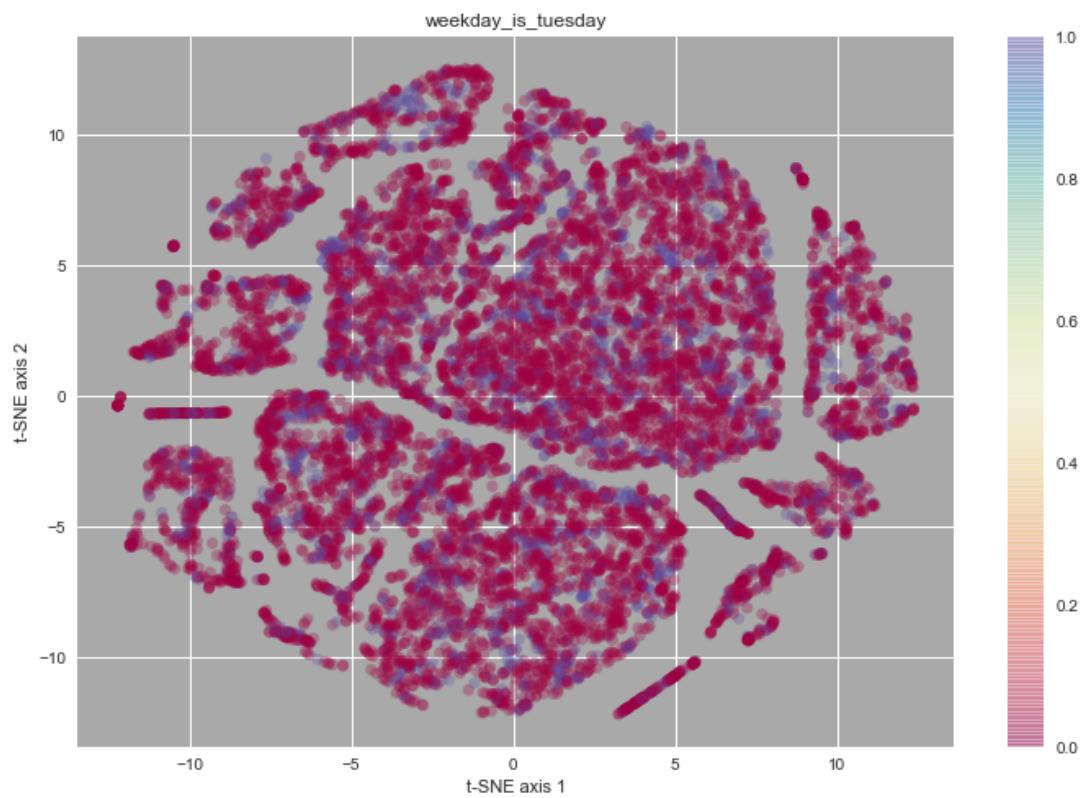
```
Out[67]: <matplotlib.figure.Figure at 0x7fe799f55240>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b65bfcf8>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6bb84e0>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7998e2588>
Out[67]: <matplotlib.text.Text at 0x7fe7b67ce828>
Out[67]: <matplotlib.text.Text at 0x7fe7b6813550>
Out[67]: <matplotlib.text.Text at 0x7fe7b6b4e5f8>
```



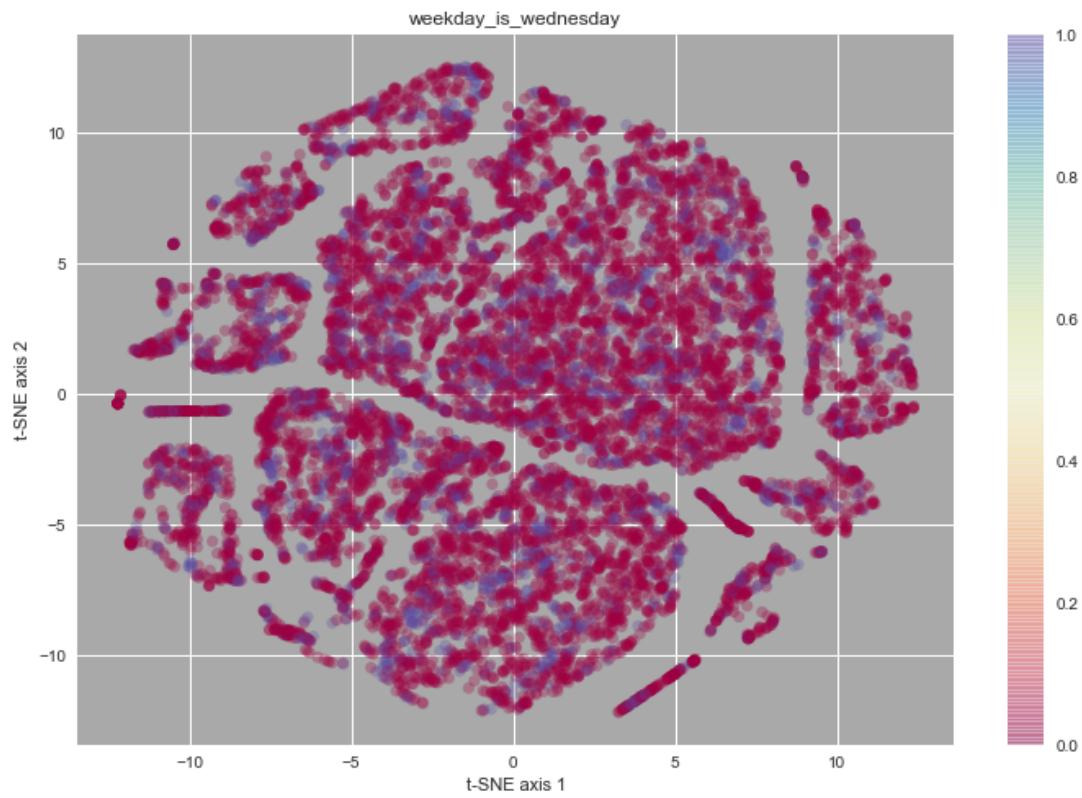
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6cad860>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe799a81198>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7993d3c50>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe79950c2b0>
Out[67]: <matplotlib.text.Text at 0x7fe7b6664908>
Out[67]: <matplotlib.text.Text at 0x7fe7996a7128>
Out[67]: <matplotlib.text.Text at 0x7fe79970d860>
```



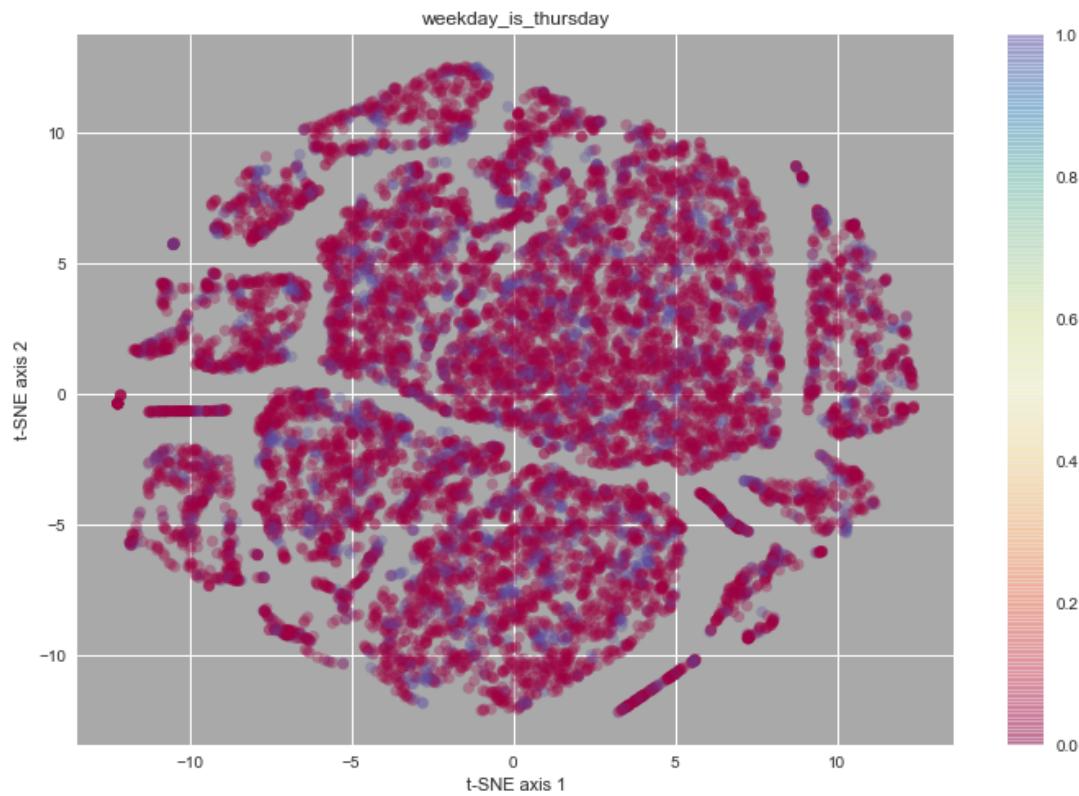
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b698c588>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe79948f208>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe799480860>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6bd7dd8>
Out[67]: <matplotlib.text.Text at 0x7fe7999d87b8>
Out[67]: <matplotlib.text.Text at 0x7fe7b69070b8>
Out[67]: <matplotlib.text.Text at 0x7fe7b68ec550>
```



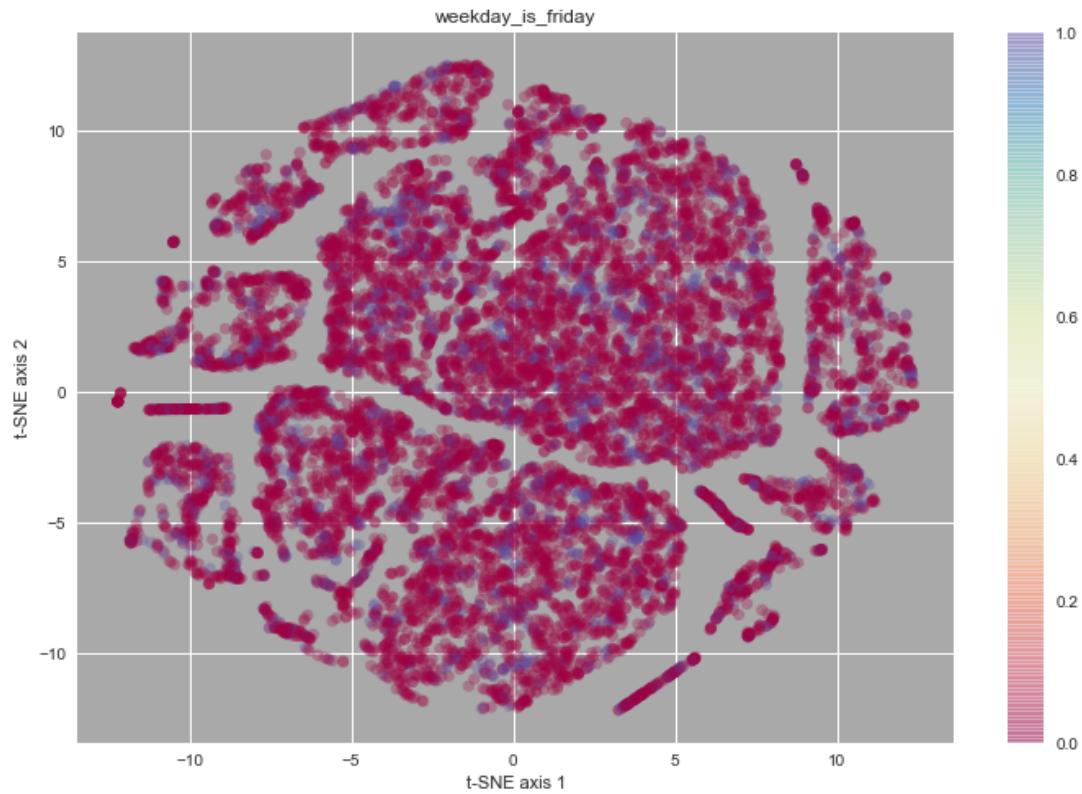
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6404d30>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6419780>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b67b9cf8>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b672b5f8>
Out[67]: <matplotlib.text.Text at 0x7fe7b640ed30>
Out[67]: <matplotlib.text.Text at 0x7fe7b6bcb2e8>
Out[67]: <matplotlib.text.Text at 0x7fe7b655a128>
```



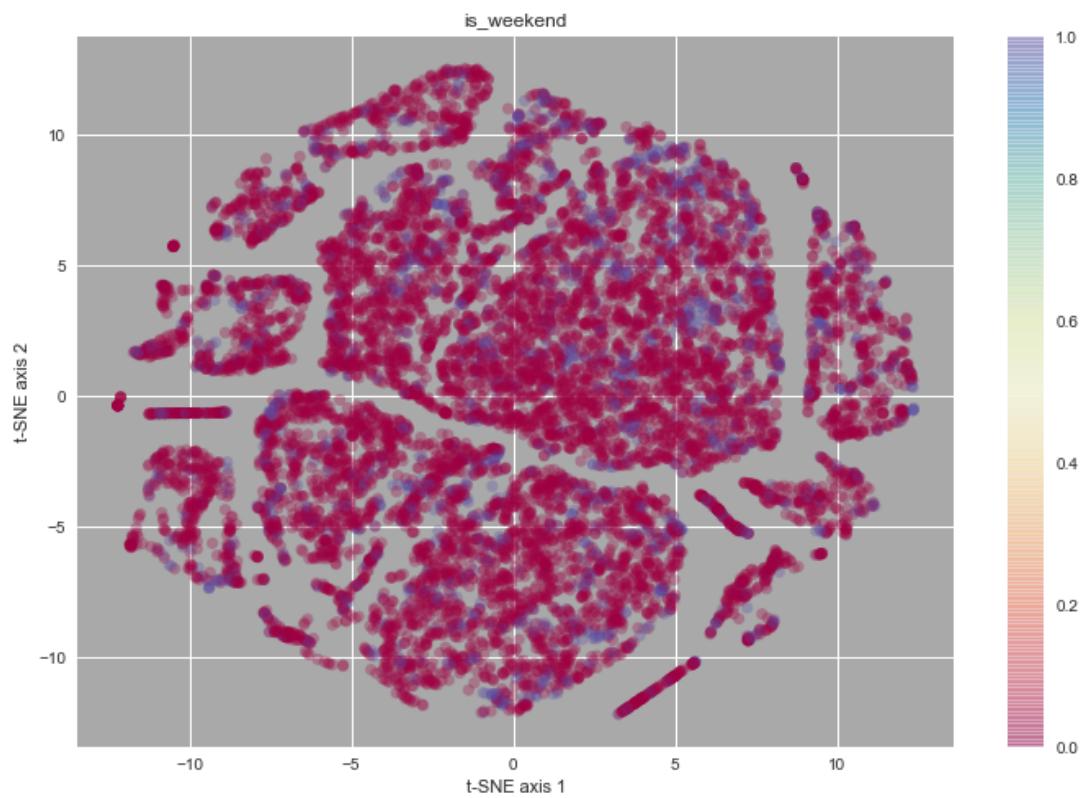
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b652e898>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe799495f60>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b65a9fd0>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b63defd0>
Out[67]: <matplotlib.text.Text at 0x7fe7b6948710>
Out[67]: <matplotlib.text.Text at 0x7fe7b6ab65c0>
Out[67]: <matplotlib.text.Text at 0x7fe7b6629400>
```



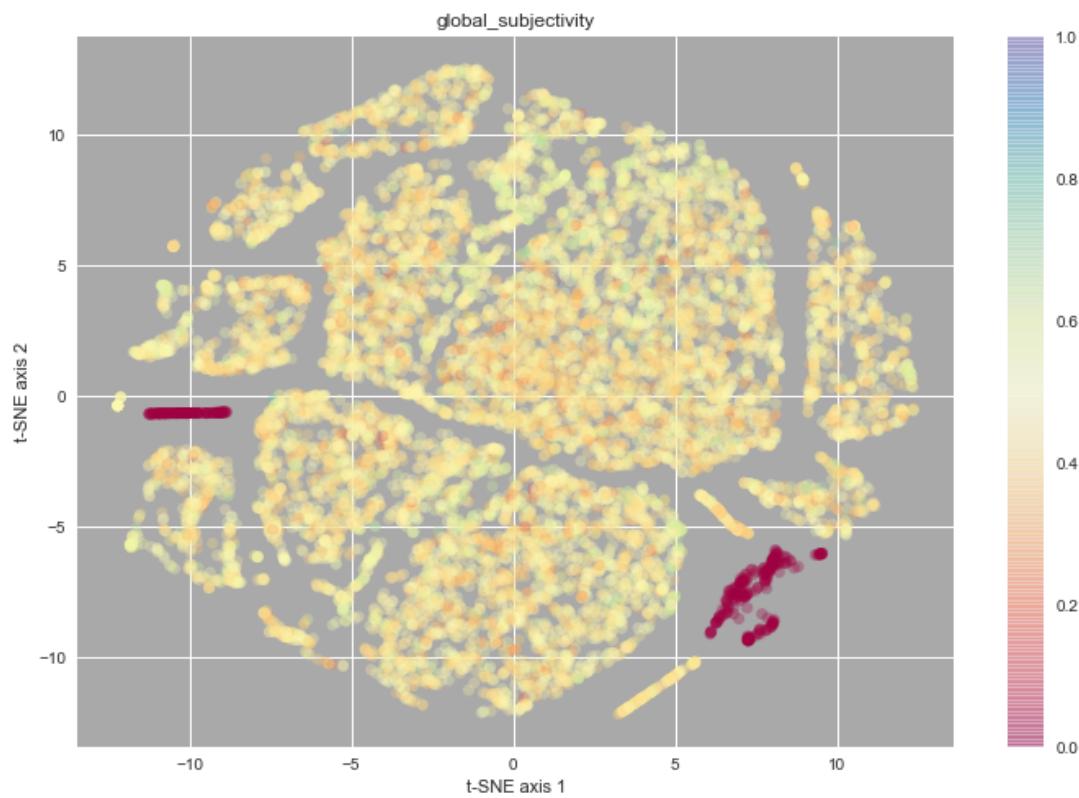
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6929f28>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6399dd8>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b609a2e8>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b60682e8>
Out[67]: <matplotlib.text.Text at 0x7fe7b662b2e8>
Out[67]: <matplotlib.text.Text at 0x7fe7b6345898>
Out[67]: <matplotlib.text.Text at 0x7fe7b62f36d8>
```



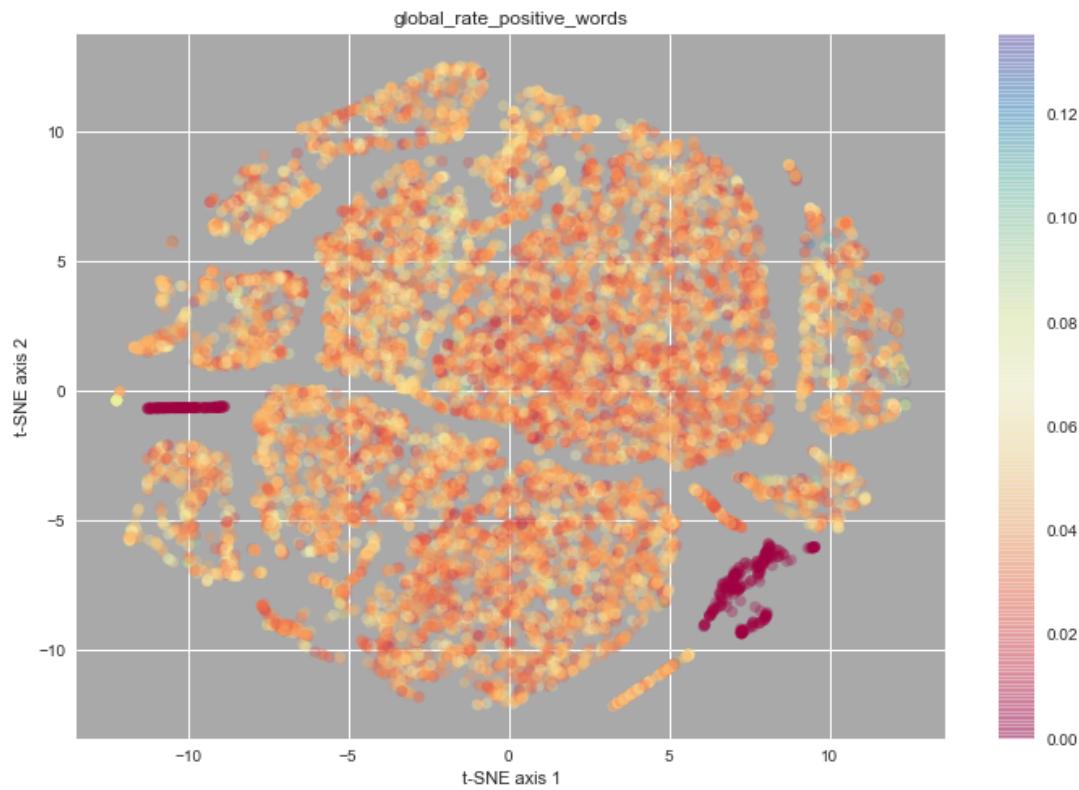
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b632a390>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b60259b0>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b5fa7438>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6025128>
Out[67]: <matplotlib.text.Text at 0x7fe7b607d048>
Out[67]: <matplotlib.text.Text at 0x7fe7b6048b70>
Out[67]: <matplotlib.text.Text at 0x7fe7b5ff59b0>
```



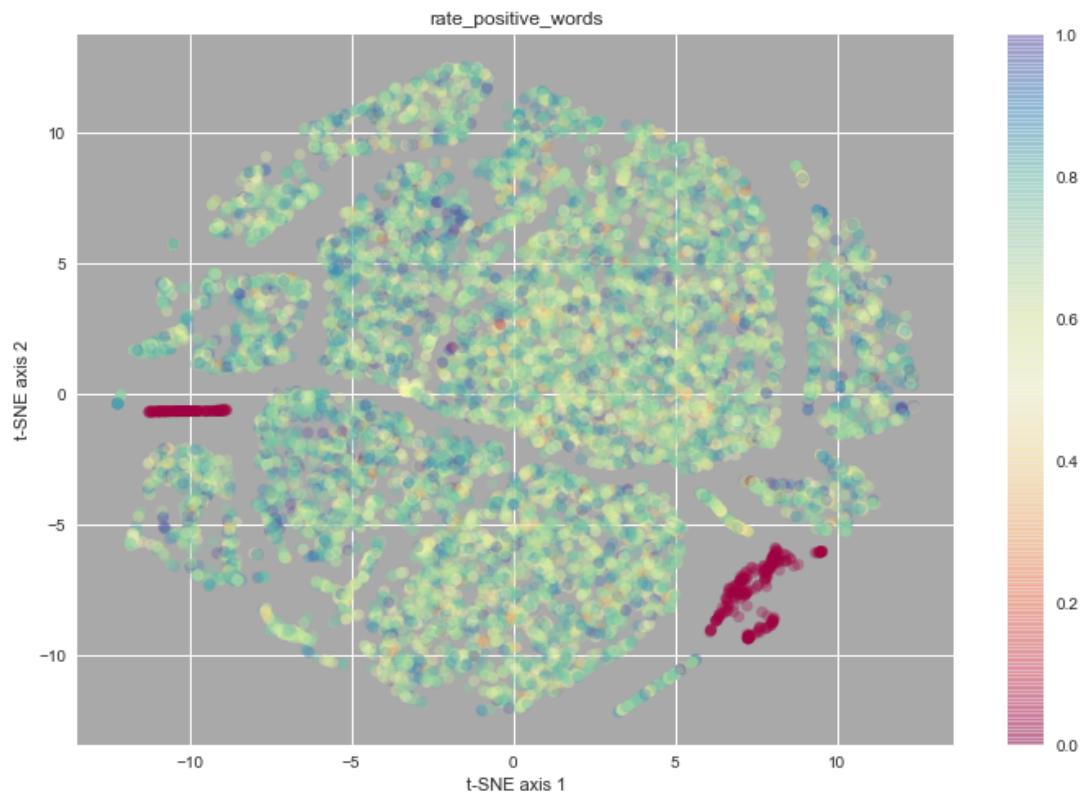
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6082518>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6316da0>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b637dda0>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe79b906048>
Out[67]: <matplotlib.text.Text at 0x7fe7b6345898>
Out[67]: <matplotlib.text.Text at 0x7fe7b609a4a8>
Out[67]: <matplotlib.text.Text at 0x7fe7b6378e10>
```



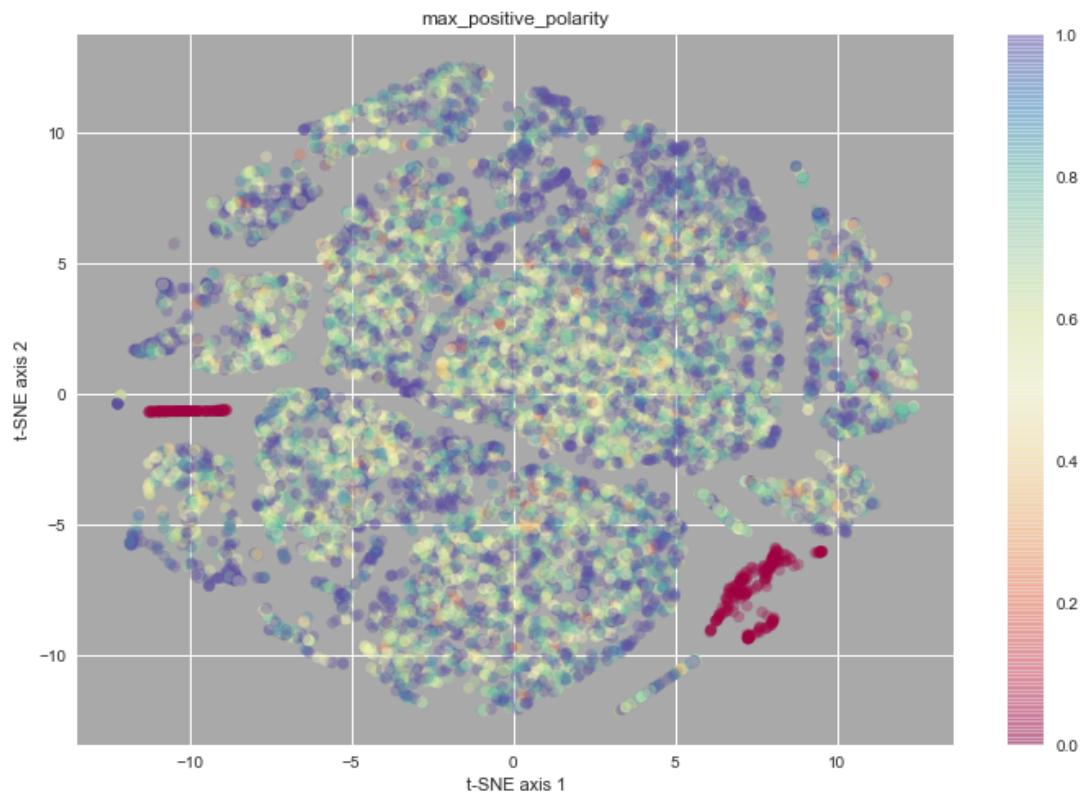
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b633f588>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b66be668>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6a3ad68>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe79950c8d0>
Out[67]: <matplotlib.text.Text at 0x7fe7b6b17eb8>
Out[67]: <matplotlib.text.Text at 0x7fe7b690e780>
Out[67]: <matplotlib.text.Text at 0x7fe7b68ec7f0>
```



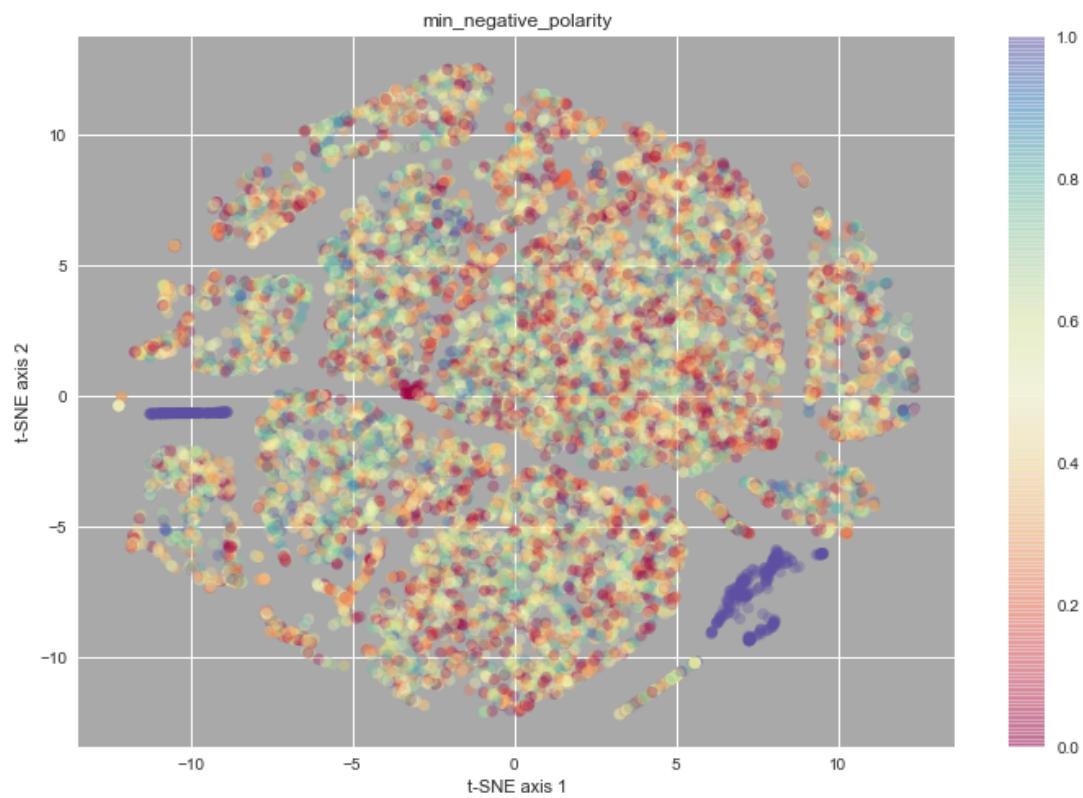
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b607a080>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe79ba4be10>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6c55a58>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b67f8f28>
Out[67]: <matplotlib.text.Text at 0x7fe7b6e13320>
Out[67]: <matplotlib.text.Text at 0x7fe7994faef0>
Out[67]: <matplotlib.text.Text at 0x7fe7b6e030b8>
```



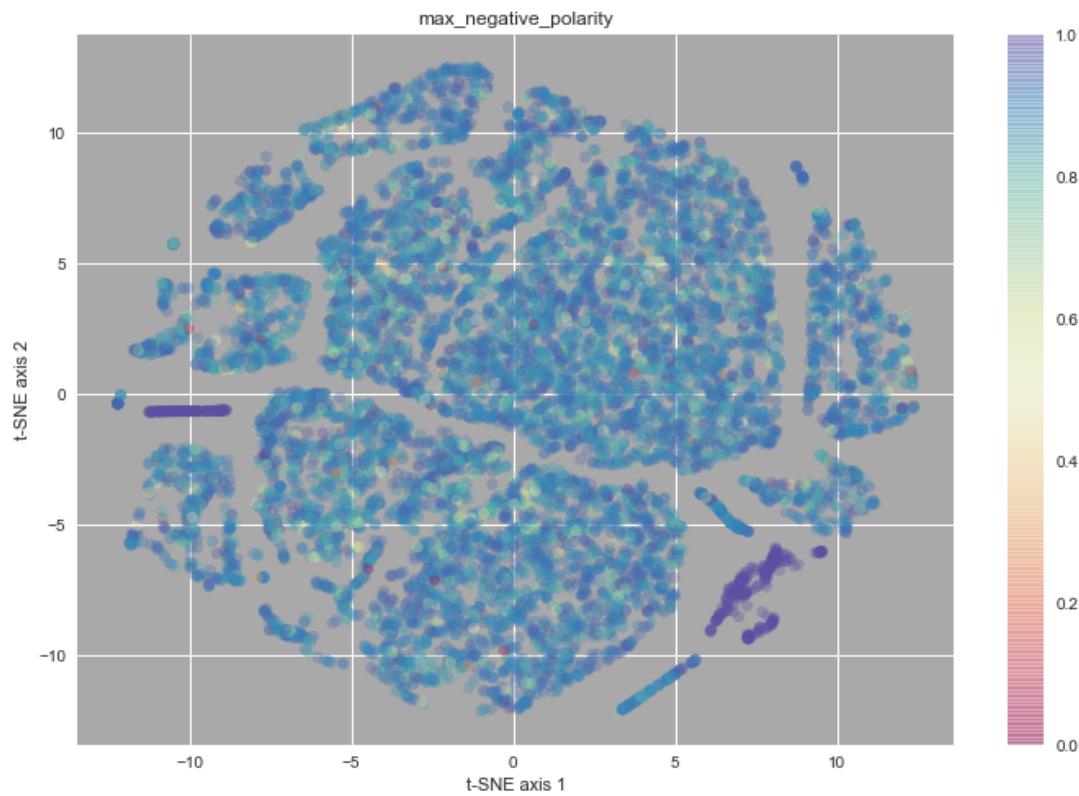
```
Out[67]: <matplotlib.figure.Figure at 0x7fe79a015ba8>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe799641b00>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7994f2860>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b64215f8>
Out[67]: <matplotlib.text.Text at 0x7fe7b6c58828>
Out[67]: <matplotlib.text.Text at 0x7fe7993c1550>
Out[67]: <matplotlib.text.Text at 0x7fe7b666fa58>
```



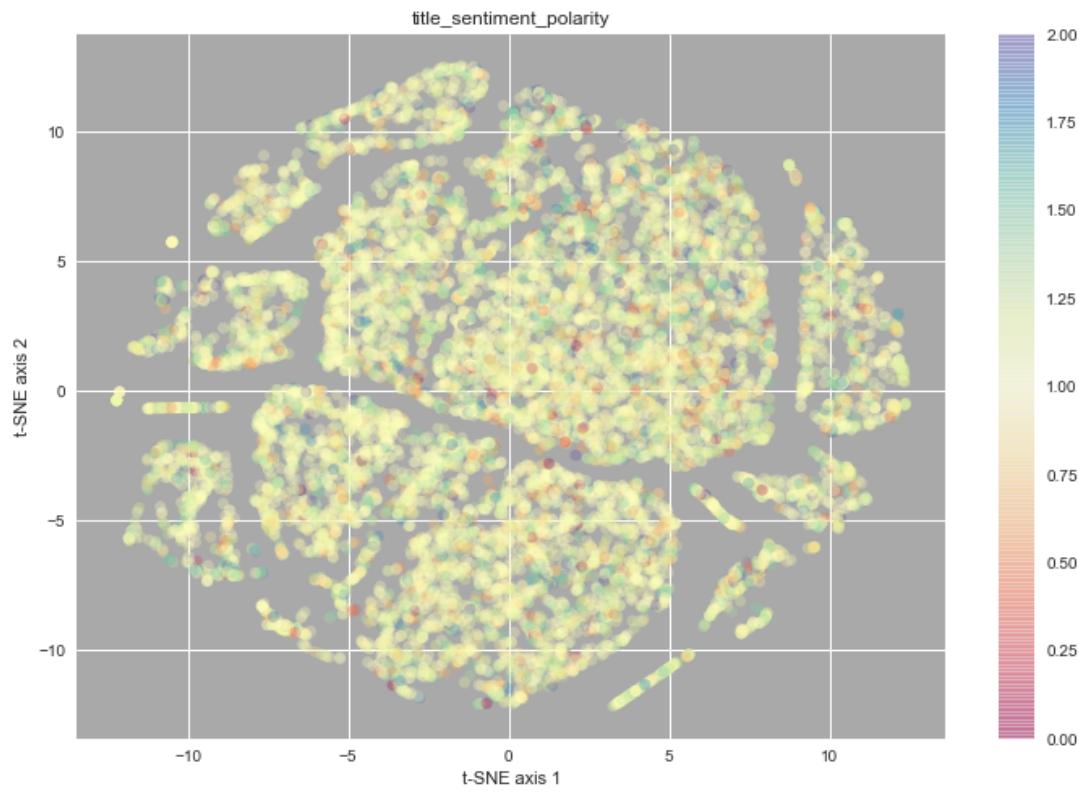
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b62ee860>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6ba2240>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6bbf828>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6d649b0>
Out[67]: <matplotlib.text.Text at 0x7fe7b696c320>
Out[67]: <matplotlib.text.Text at 0x7fe7b6040550>
Out[67]: <matplotlib.text.Text at 0x7fe7b6ab6da0>
```



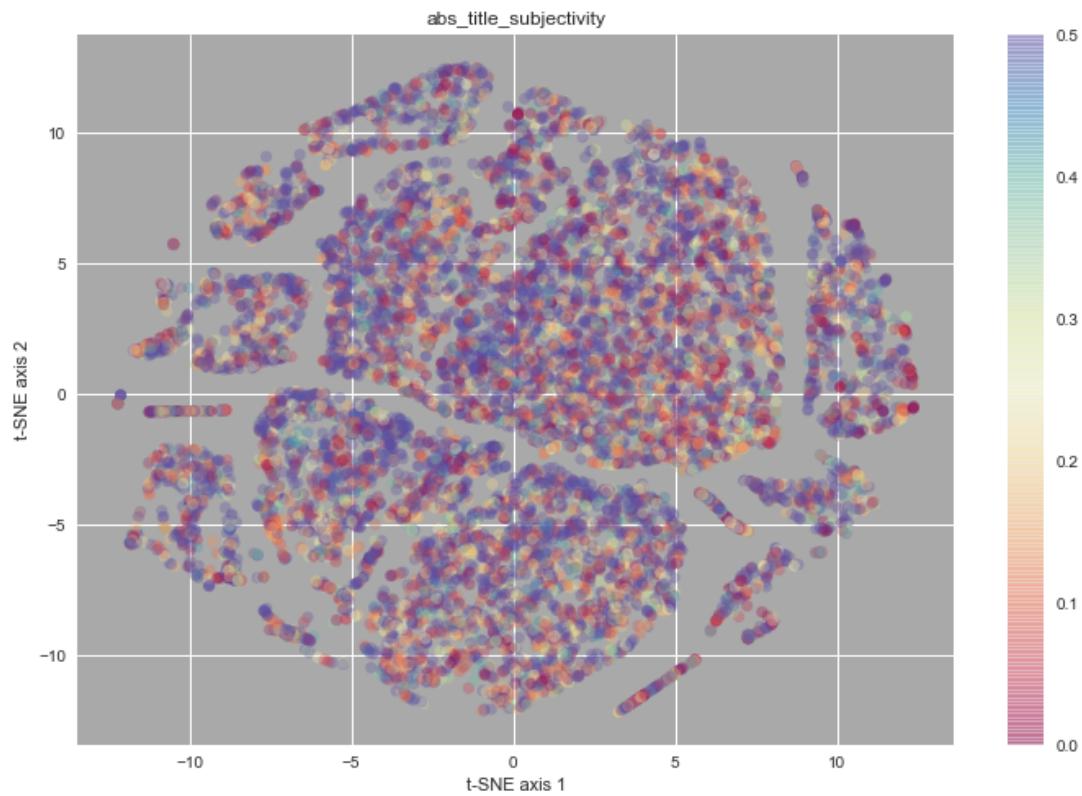
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6bd47b8>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6cdd898>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6653b00>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6754c88>
Out[67]: <matplotlib.text.Text at 0x7fe7b6d90da0>
Out[67]: <matplotlib.text.Text at 0x7fe7b6942278>
Out[67]: <matplotlib.text.Text at 0x7fe7b60a90b8>
```



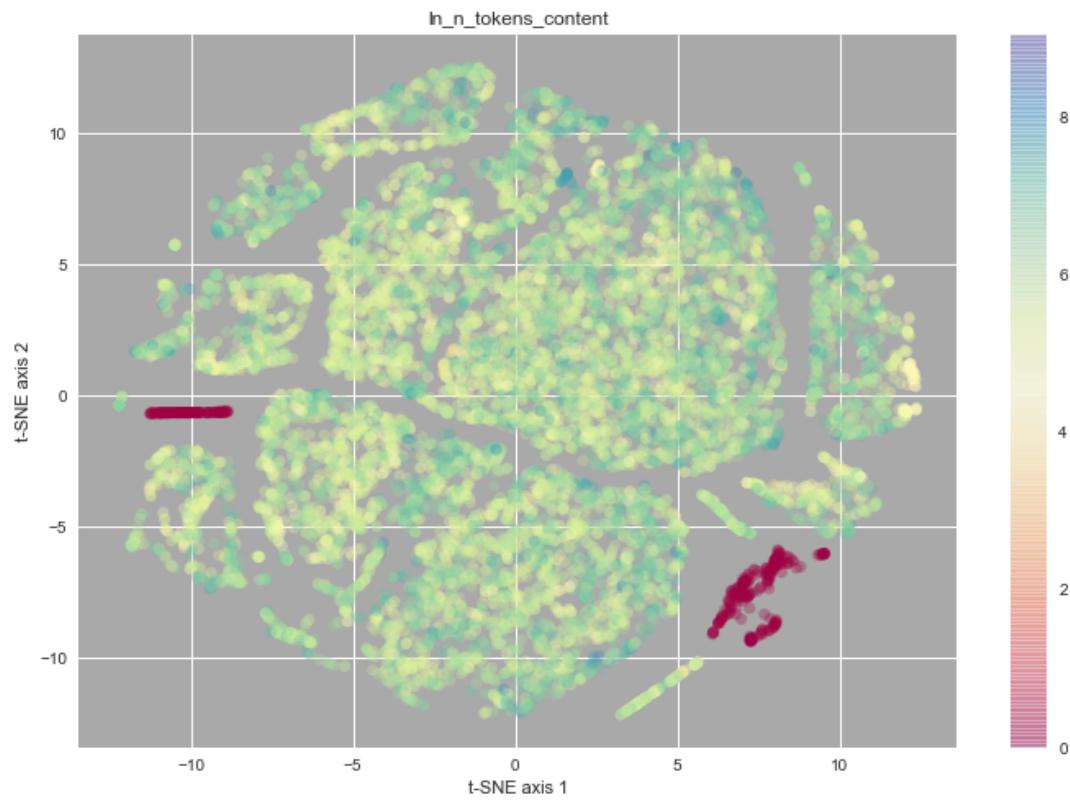
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b67ae128>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b67aeeff0>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b68cbdd8>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b65a4b00>
Out[67]: <matplotlib.text.Text at 0x7fe7b63e43c8>
Out[67]: <matplotlib.text.Text at 0x7fe7b65f7550>
Out[67]: <matplotlib.text.Text at 0x7fe7b6607390>
```



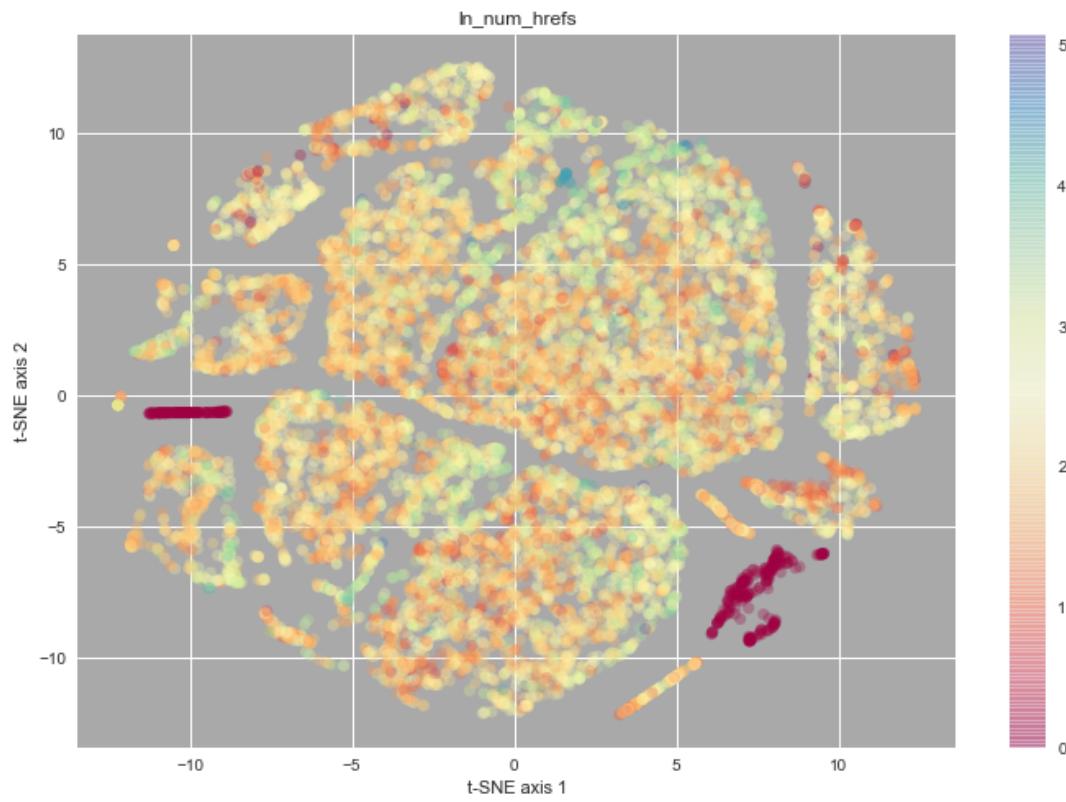
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b693d4e0>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6931dd8>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe79b8f75f8>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7999f5550>
Out[67]: <matplotlib.text.Text at 0x7fe7b63f7fd0>
Out[67]: <matplotlib.text.Text at 0x7fe7b6371e10>
Out[67]: <matplotlib.text.Text at 0x7fe7b6ac9c50>
```



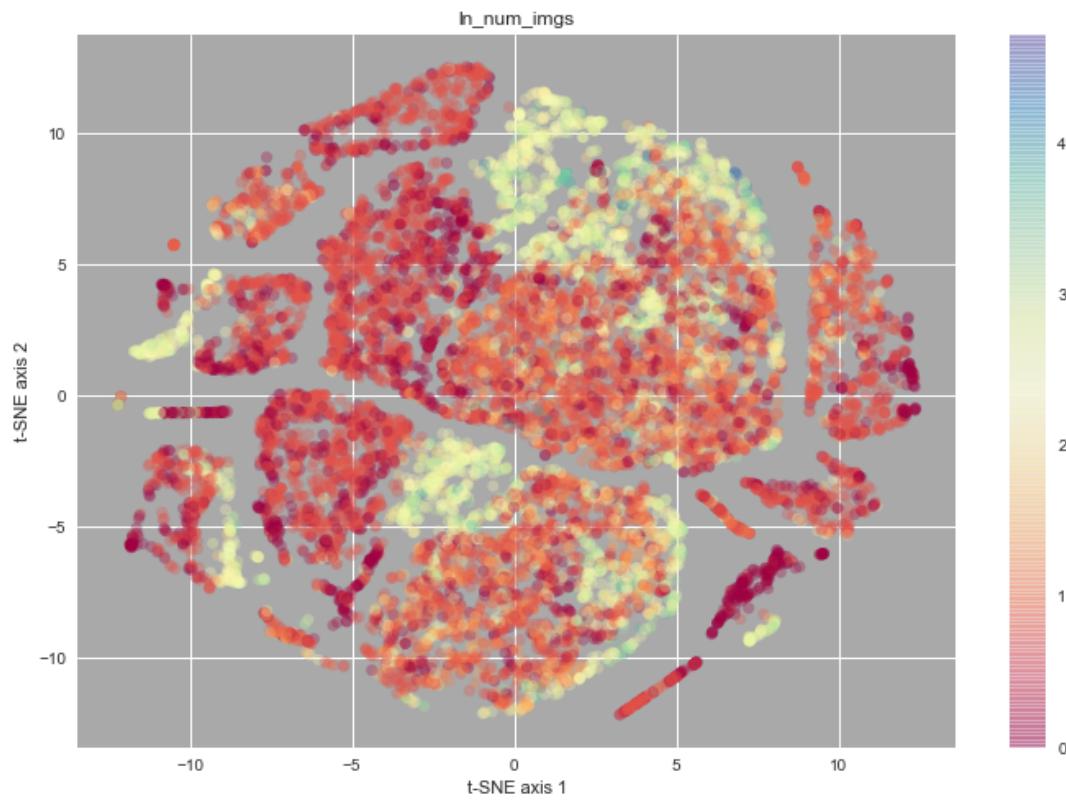
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6a09780>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b63e4358>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe799863048>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6b78ef0>
Out[67]: <matplotlib.text.Text at 0x7fe7b6a0f320>
Out[67]: <matplotlib.text.Text at 0x7fe7b68bde80>
Out[67]: <matplotlib.text.Text at 0x7fe799801470>
```



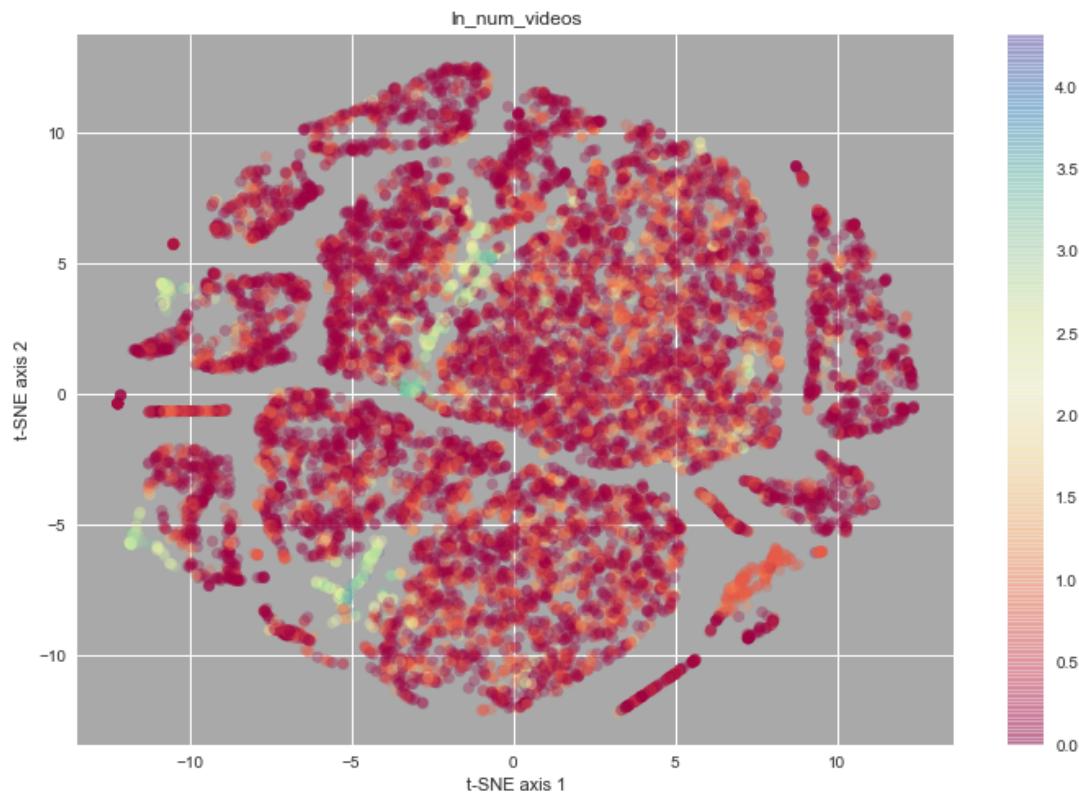
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6ba2400>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe799b3c898>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7996d75c0>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b658e400>
Out[67]: <matplotlib.text.Text at 0x7fe7b6c78dd8>
Out[67]: <matplotlib.text.Text at 0x7fe7995cba90>
Out[67]: <matplotlib.text.Text at 0x7fe799d5f2e8>
```



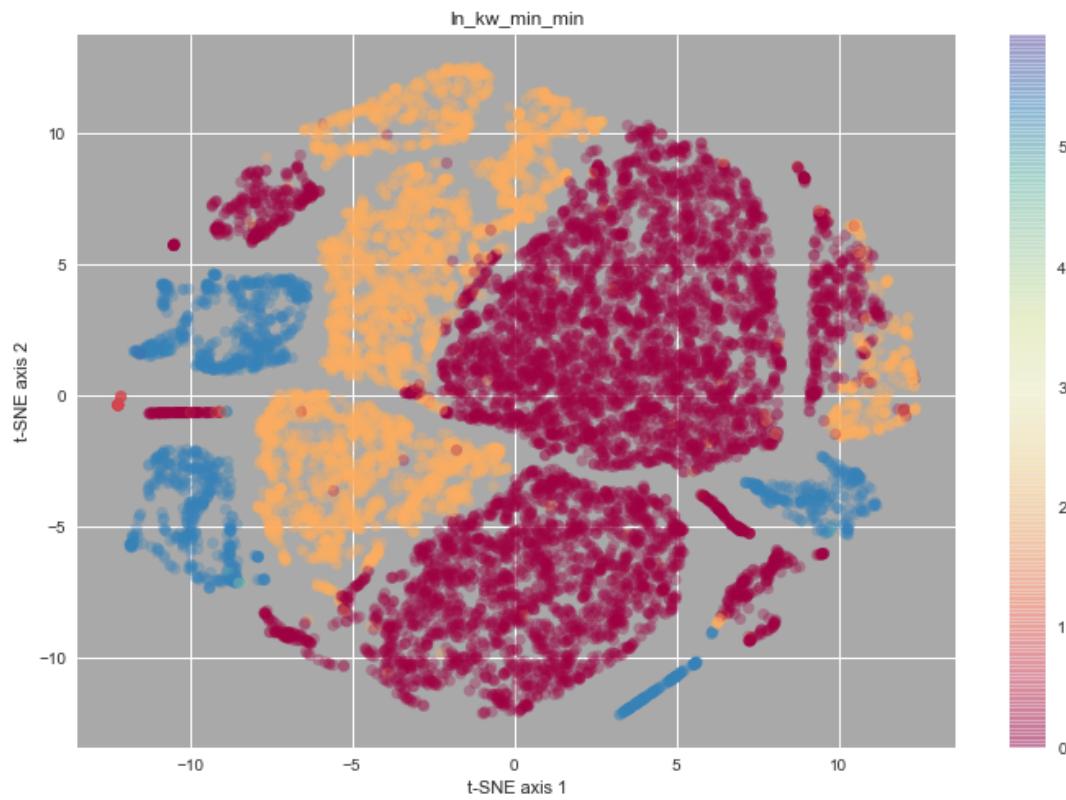
```
Out[67]: <matplotlib.figure.Figure at 0x7fe799b3cac8>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe79a015c50>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b637d048>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6d52f28>
Out[67]: <matplotlib.text.Text at 0x7fe799e7fe10>
Out[67]: <matplotlib.text.Text at 0x7fe7994f3940>
Out[67]: <matplotlib.text.Text at 0x7fe7b6ae1e10>
```



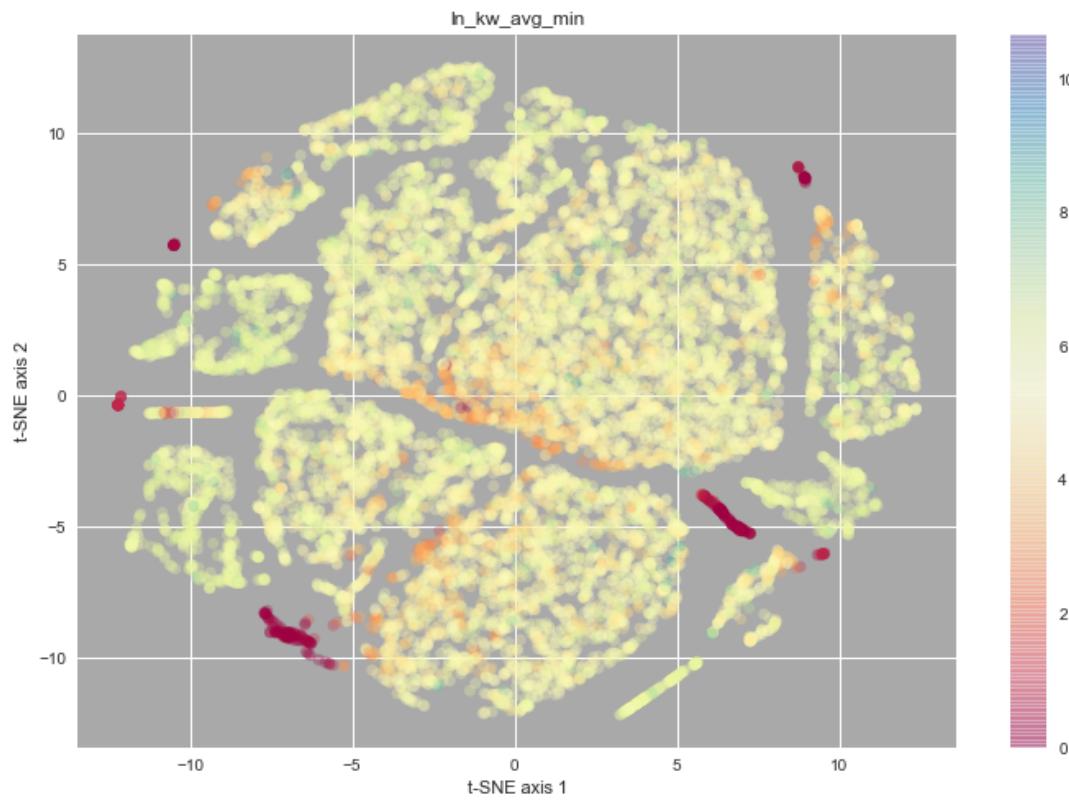
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6557e48>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6410320>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6633630>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b65df7b8>
Out[67]: <matplotlib.text.Text at 0x7fe7b6040048>
Out[67]: <matplotlib.text.Text at 0x7fe7b6be0eb8>
Out[67]: <matplotlib.text.Text at 0x7fe7b6b3eba8>
```



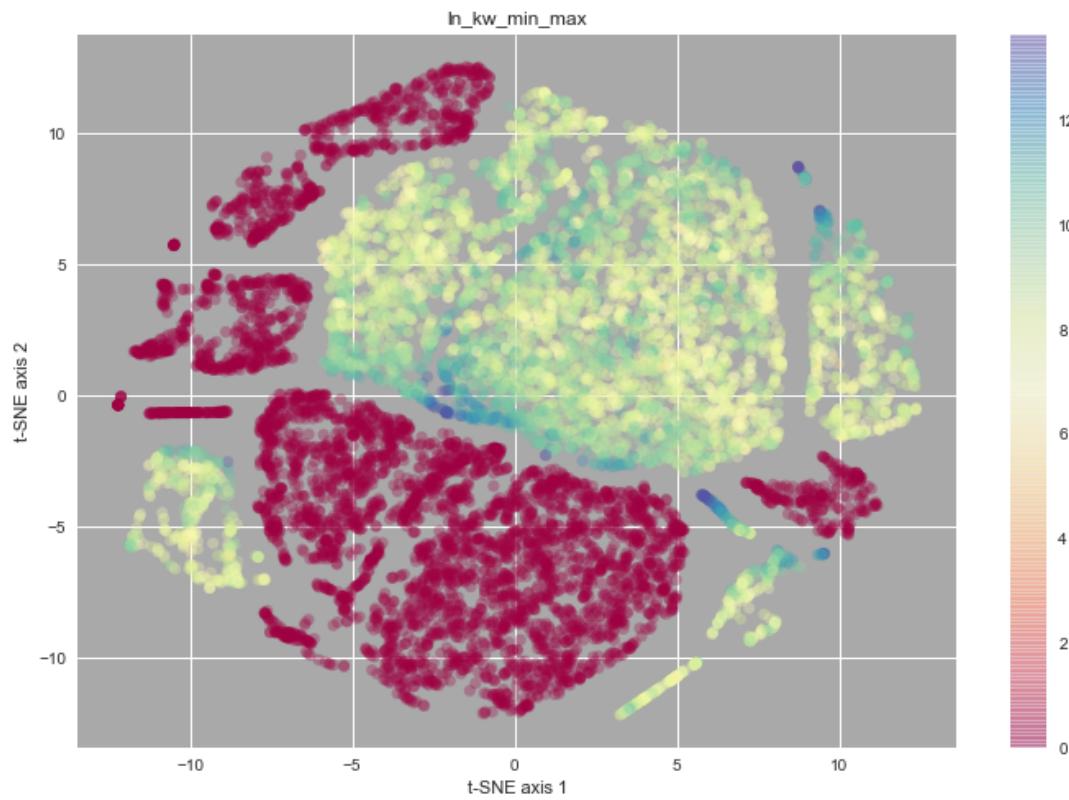
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b60a7f28>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b60a7a90>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b63324e0>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b60990b8>
Out[67]: <matplotlib.text.Text at 0x7fe7b65b3390>
Out[67]: <matplotlib.text.Text at 0x7fe7b6552898>
Out[67]: <matplotlib.text.Text at 0x7fe7b638c5f8>
```



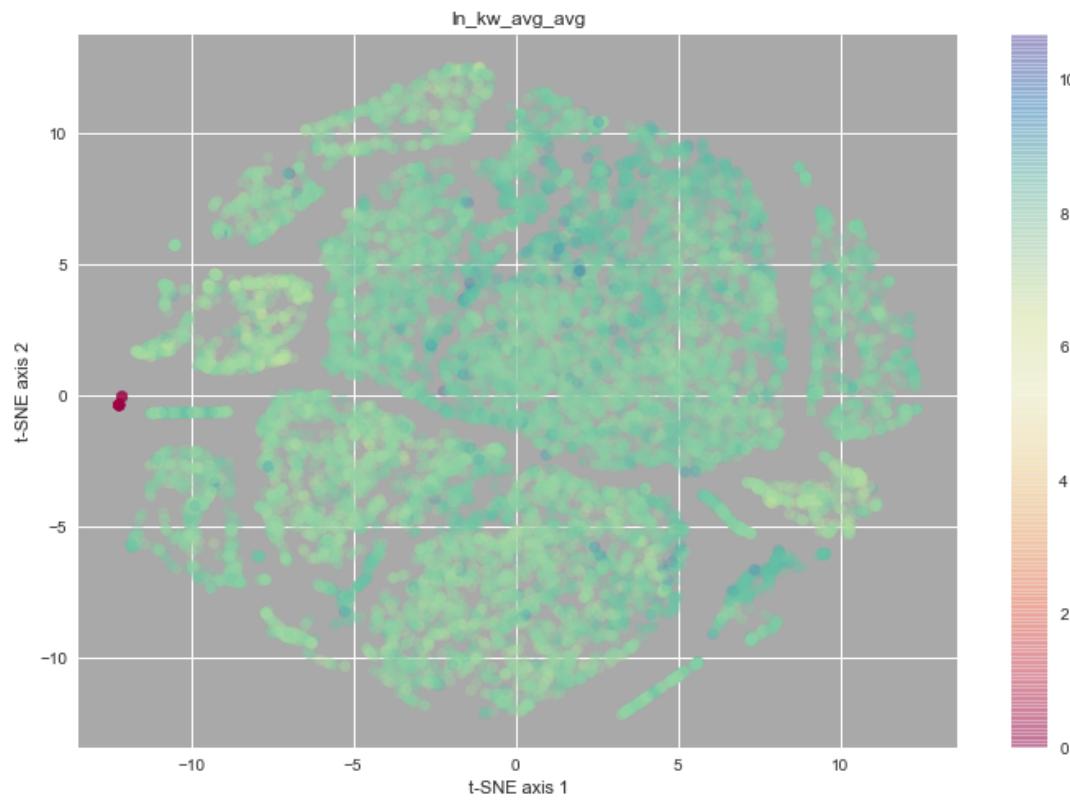
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b604ea20>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6d8cbe0>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b5f6be80>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b608c898>
Out[67]: <matplotlib.text.Text at 0x7fe7b6027780>
Out[67]: <matplotlib.text.Text at 0x7fe7b5fbe5f8>
Out[67]: <matplotlib.text.Text at 0x7fe7b6489438>
```



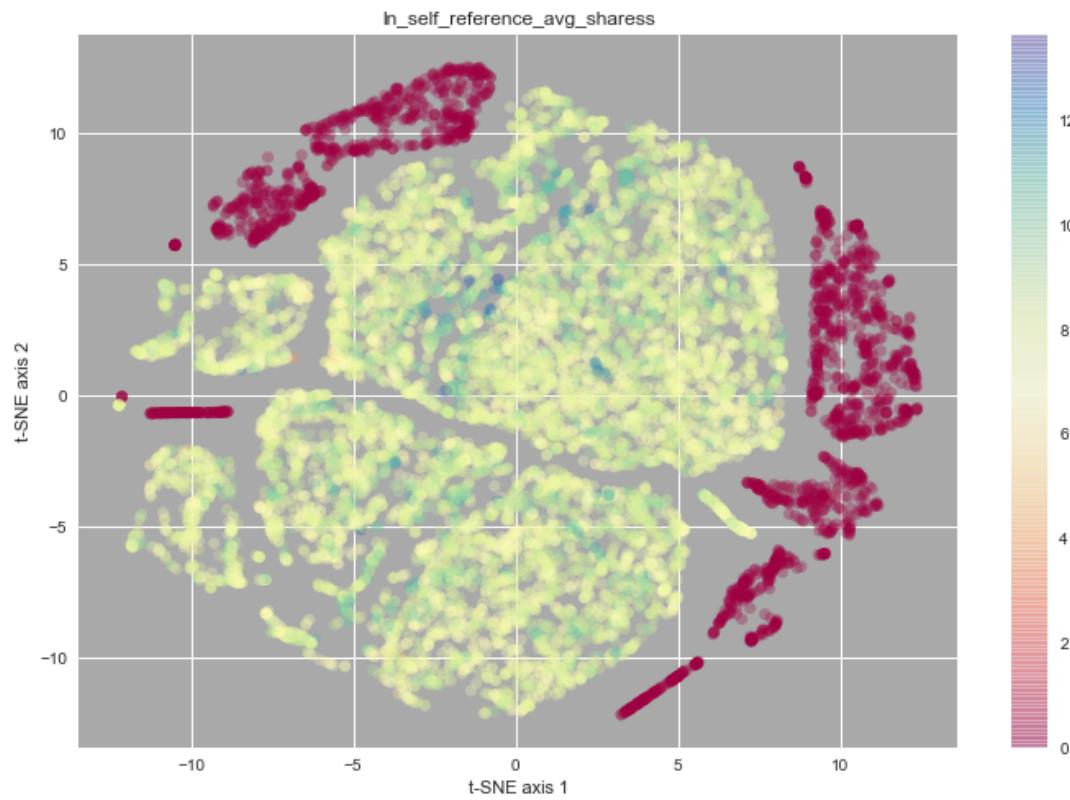
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6349ef0>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6349d68>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe799480cc0>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6ae1d68>
Out[67]: <matplotlib.text.Text at 0x7fe7b6ab8748>
Out[67]: <matplotlib.text.Text at 0x7fe7b6940208>
Out[67]: <matplotlib.text.Text at 0x7fe7b6534588>
```



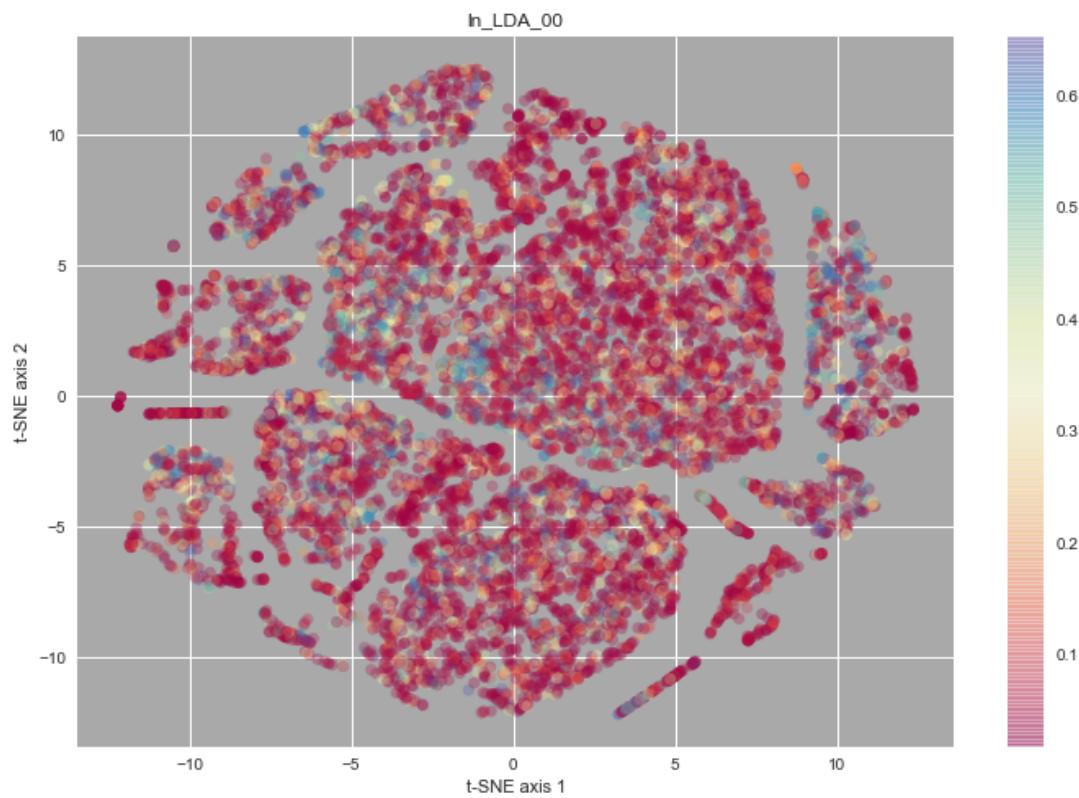
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6a02438>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe799e6deb8>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6cac2b0>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b604e908>
Out[67]: <matplotlib.text.Text at 0x7fe7b60cb828>
Out[67]: <matplotlib.text.Text at 0x7fe7b658e828>
Out[67]: <matplotlib.text.Text at 0x7fe7b6a34e80>
```



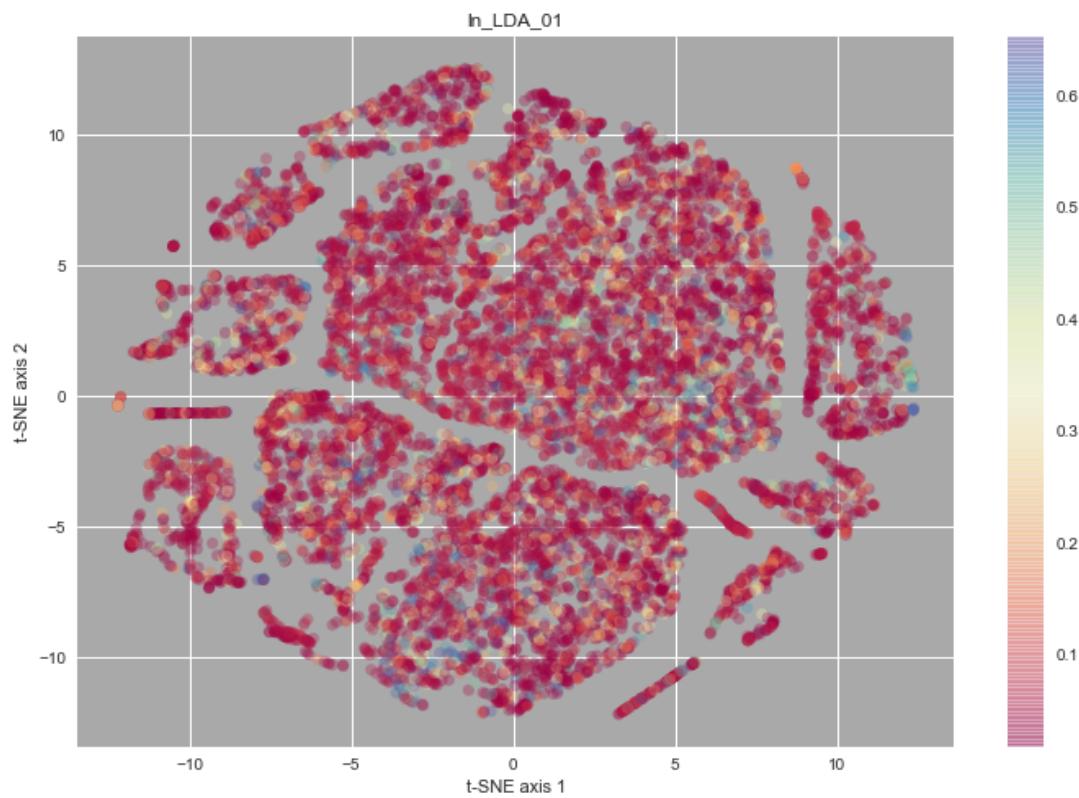
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7994ec0f0>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7996b19b0>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6e11ba8>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6a81e48>
Out[67]: <matplotlib.text.Text at 0x7fe7b6965b38>
Out[67]: <matplotlib.text.Text at 0x7fe799791748>
Out[67]: <matplotlib.text.Text at 0x7fe7b68bd1d0>
```



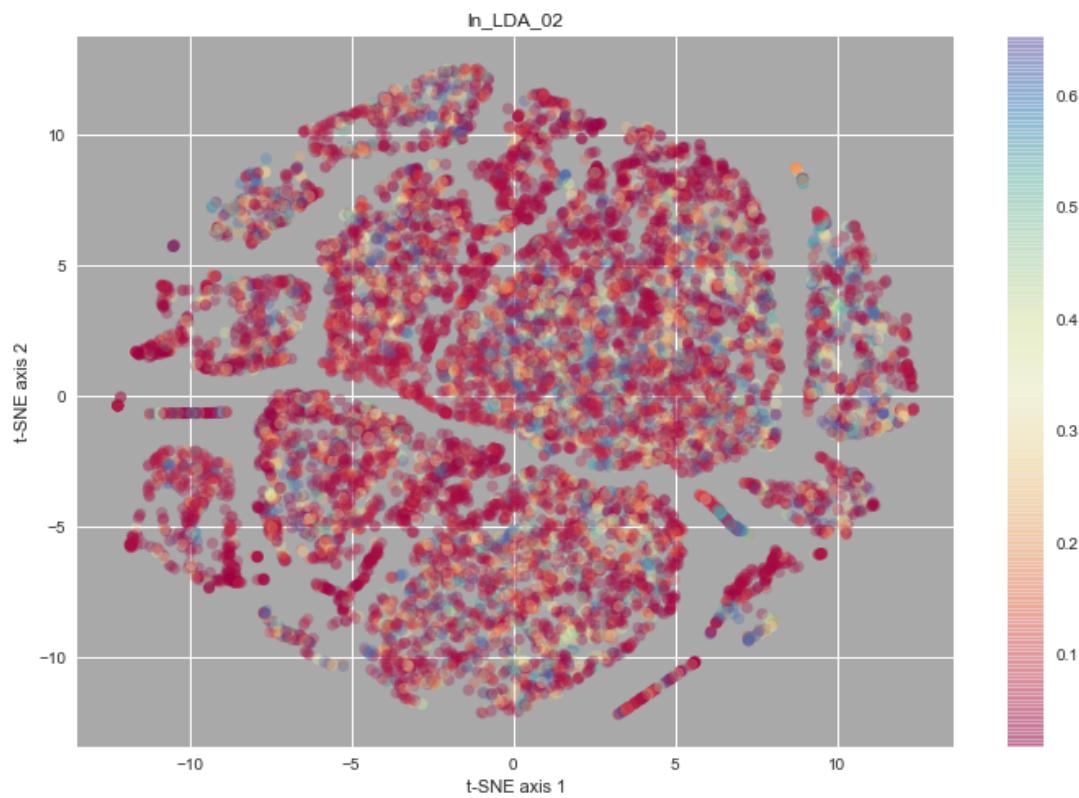
```
Out[67]: <matplotlib.figure.Figure at 0x7fe79950c828>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe799daad30>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6557ba8>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7994d9a20>
Out[67]: <matplotlib.text.Text at 0x7fe79951bac8>
Out[67]: <matplotlib.text.Text at 0x7fe7b6711d30>
Out[67]: <matplotlib.text.Text at 0x7fe7b6a9d828>
```



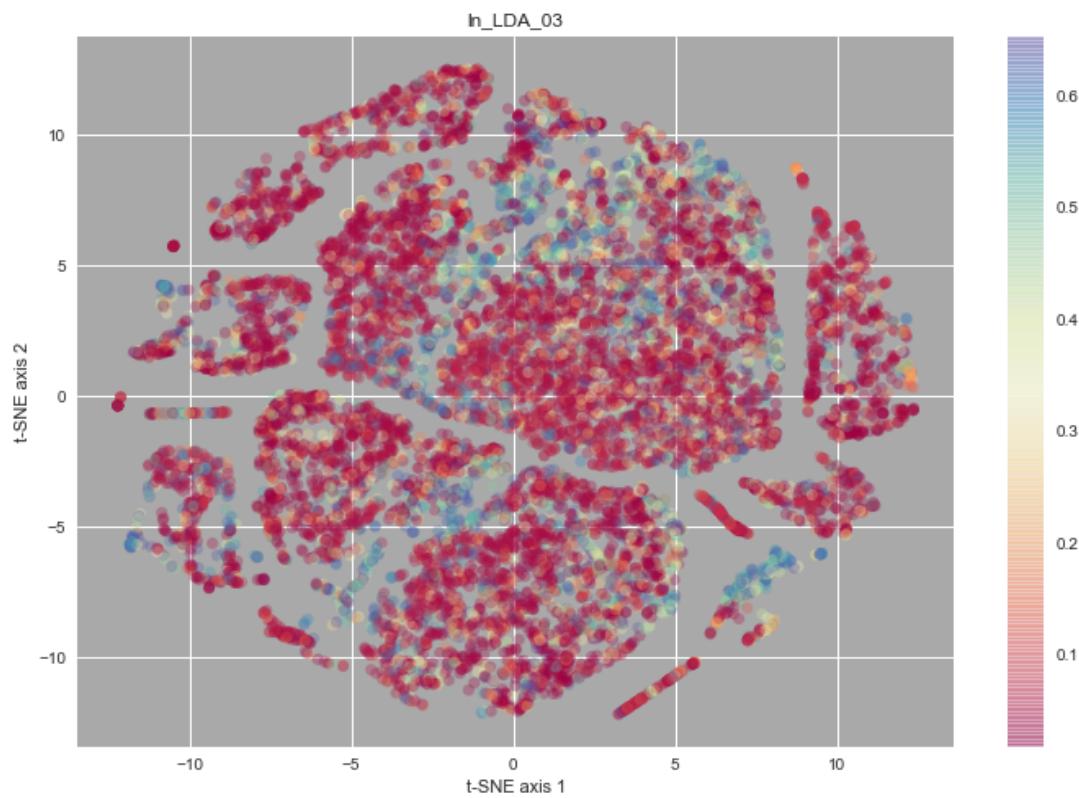
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6ab82b0>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5fce10>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6b33940>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b665bac8>
Out[67]: <matplotlib.text.Text at 0x7fe7b60bd1d0>
Out[67]: <matplotlib.text.Text at 0x7fe7b6be10b8>
Out[67]: <matplotlib.text.Text at 0x7fe7b6538eb8>
```



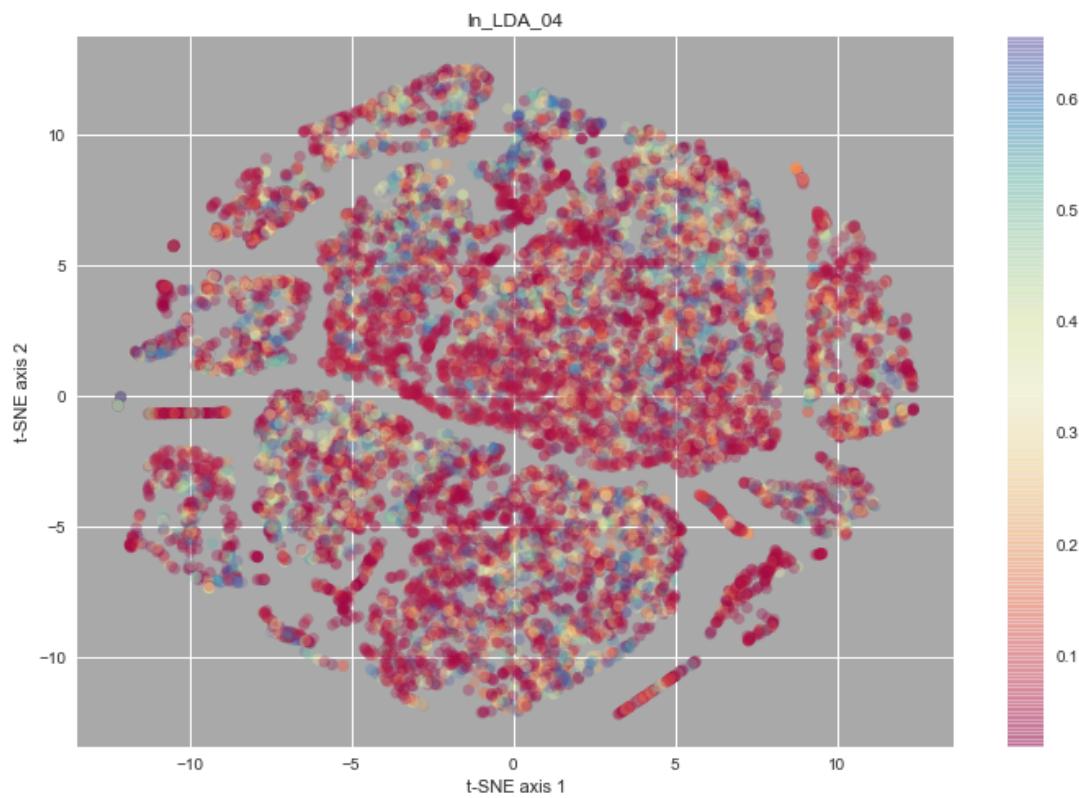
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b65aec88>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b67b78d0>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6338c18>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b65cdda0>
Out[67]: <matplotlib.text.Text at 0x7fe7b67af588>
Out[67]: <matplotlib.text.Text at 0x7fe7b6ad54e0>
Out[67]: <matplotlib.text.Text at 0x7fe7b69301d0>
```



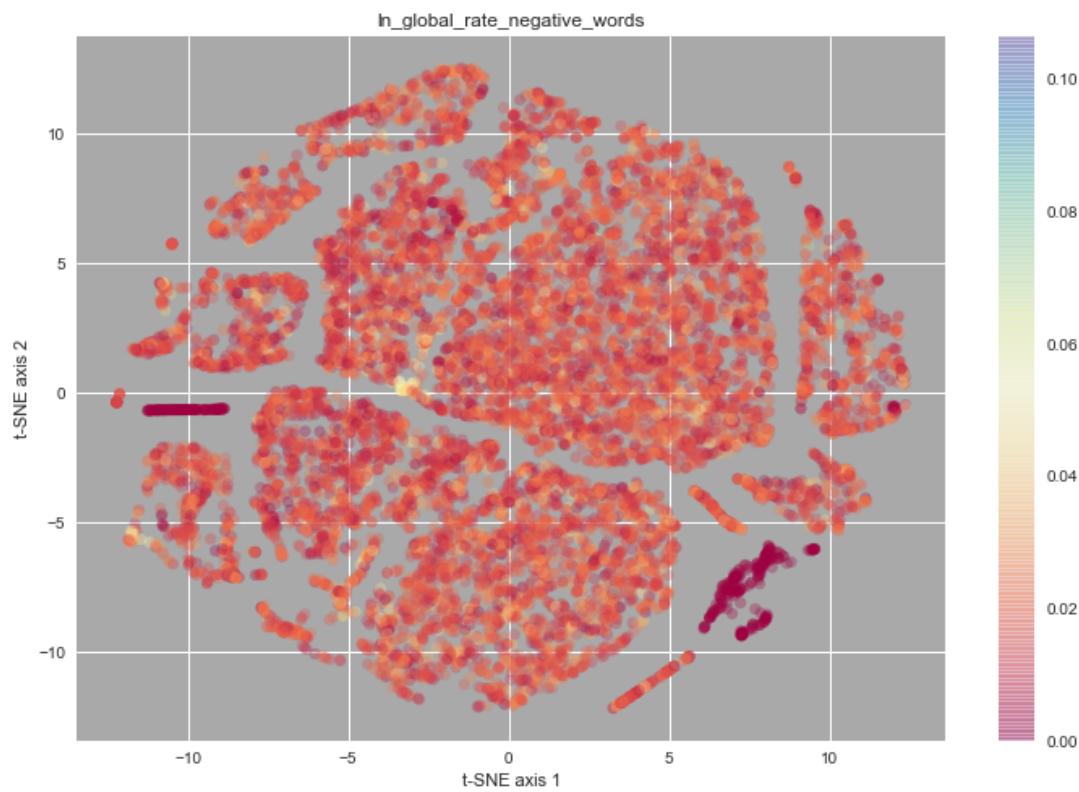
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7994c0390>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b67bb3c8>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6055ef0>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6338f98>
Out[67]: <matplotlib.text.Text at 0x7fe7b641e1d0>
Out[67]: <matplotlib.text.Text at 0x7fe7b68a3668>
Out[67]: <matplotlib.text.Text at 0x7fe7b605a4a8>
```



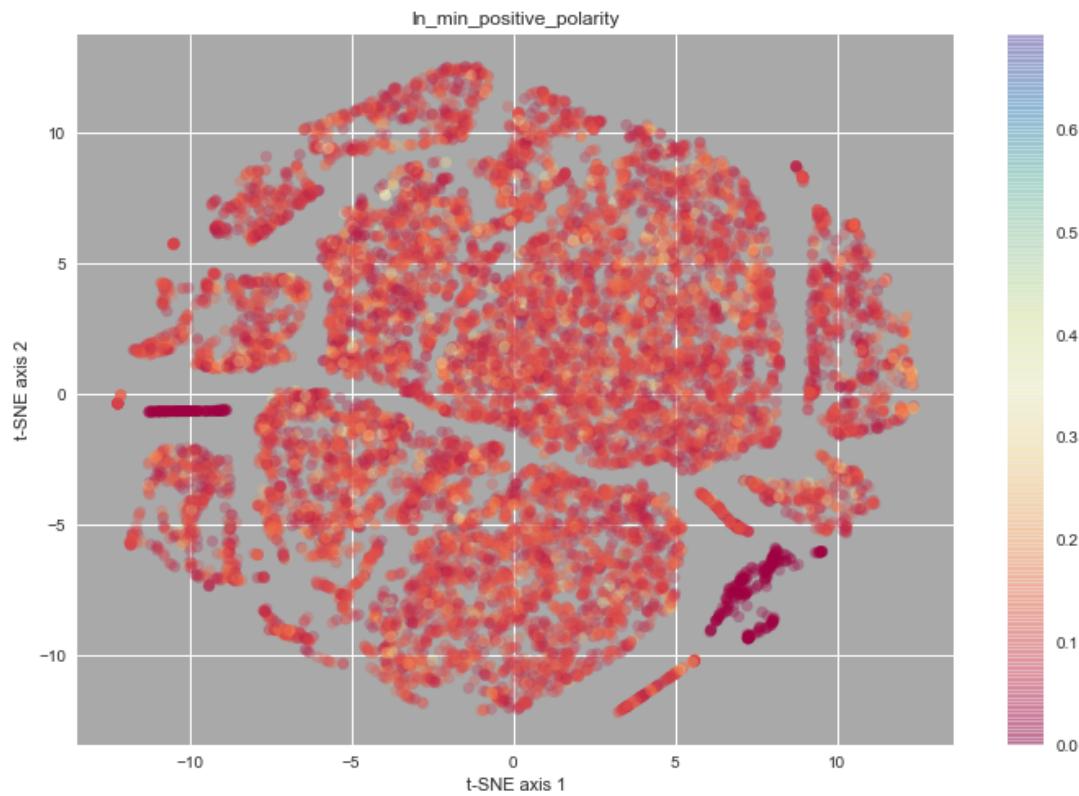
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b63f7588>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b692f780>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b69f5390>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6a09ba8>
Out[67]: <matplotlib.text.Text at 0x7fe7b63edc50>
Out[67]: <matplotlib.text.Text at 0x7fe7b6746e10>
Out[67]: <matplotlib.text.Text at 0x7fe7b631ab70>
```



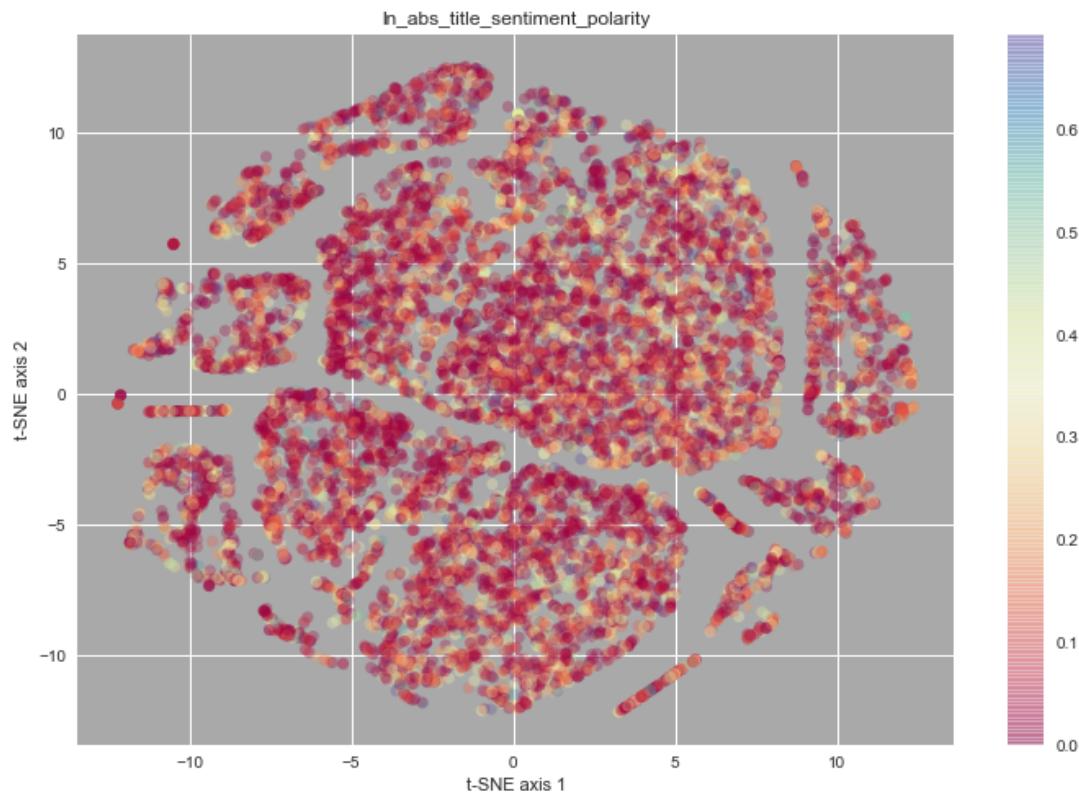
```
Out[67]: <matplotlib.figure.Figure at 0x7fe799a336d8>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7993c1a90>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7999f55c0>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6c37a90>
Out[67]: <matplotlib.text.Text at 0x7fe7b6a9d5f8>
Out[67]: <matplotlib.text.Text at 0x7fe7997e2748>
Out[67]: <matplotlib.text.Text at 0x7fe7b691e4e0>
```



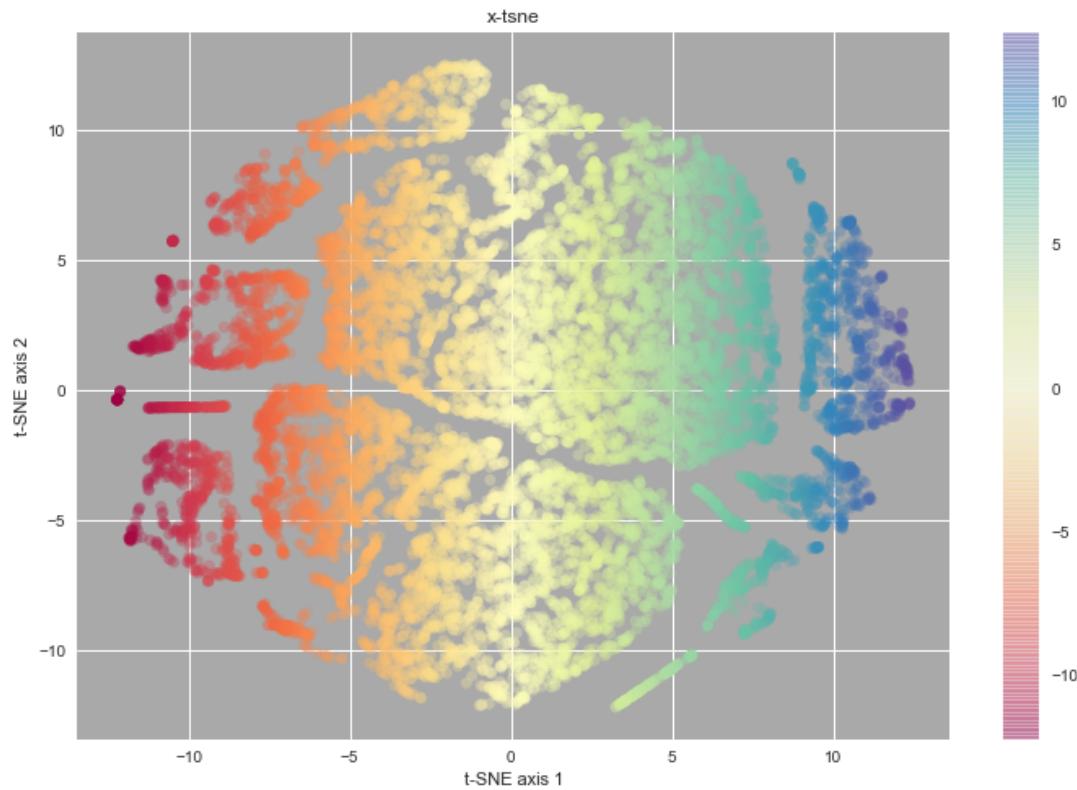
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6bbfba8>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6bb87f0>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe799641630>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7999366a0>
Out[67]: <matplotlib.text.Text at 0x7fe799817390>
Out[67]: <matplotlib.text.Text at 0x7fe7b690e3c8>
Out[67]: <matplotlib.text.Text at 0x7fe7996b1240>
```



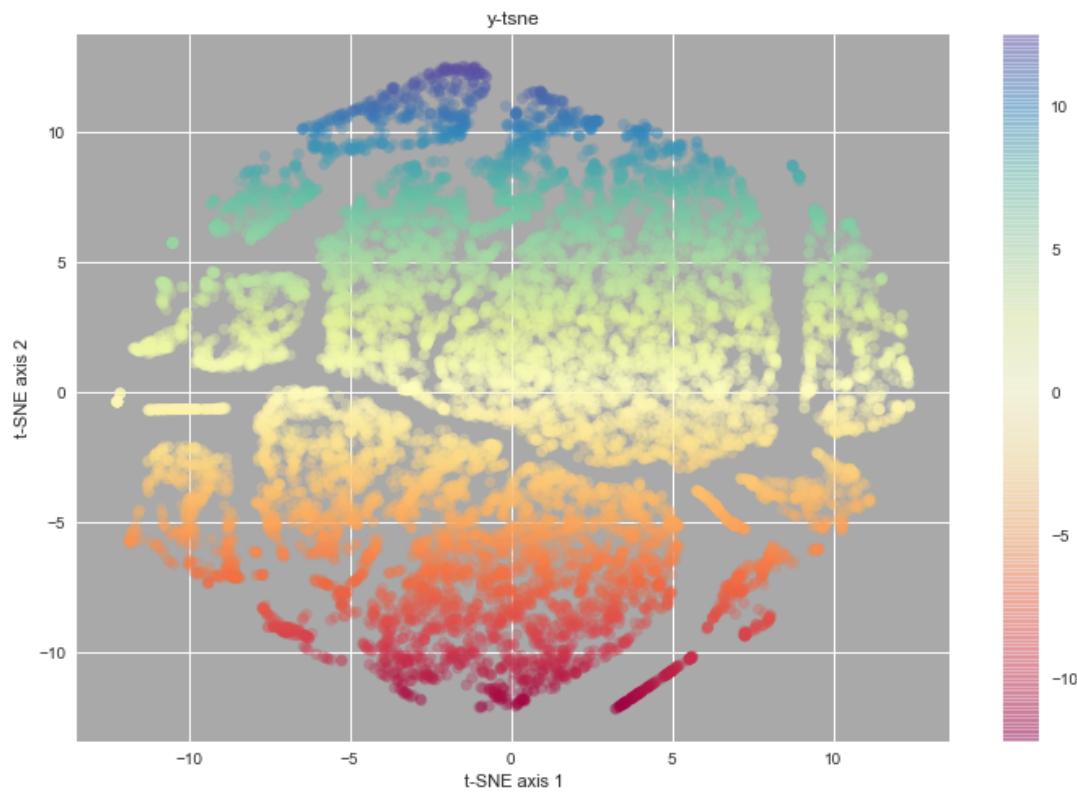
```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6728748>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b69e7a90>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe799534dd8>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6af00f0>
Out[67]: <matplotlib.text.Text at 0x7fe79948acf8>
Out[67]: <matplotlib.text.Text at 0x7fe7b6d525f8>
Out[67]: <matplotlib.text.Text at 0x7fe799e7f0b8>
```



```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b6332240>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6a7a710>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b6b44438>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe799daa748>
Out[67]: <matplotlib.text.Text at 0x7fe7b6d73c88>
Out[67]: <matplotlib.text.Text at 0x7fe7b6416f60>
Out[67]: <matplotlib.text.Text at 0x7fe7b6bd9908>
```



```
Out[67]: <matplotlib.figure.Figure at 0x7fe7b64e9470>
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b60a56d8>
Out[67]: <matplotlib.collections.PathCollection at 0x7fe7b65c5fd0>
Out[67]: <matplotlib.colorbar.Colorbar at 0x7fe7b6336240>
Out[67]: <matplotlib.text.Text at 0x7fe7b634be80>
Out[67]: <matplotlib.text.Text at 0x7fe7b60ca6d8>
Out[67]: <matplotlib.text.Text at 0x7fe7b6934438>
```



```
In [48]: tsne_results  
len(tsne_results)
```

```
Out[48]: array([[ 3.44641972, -6.01625061],  
                 [-7.65256214, -4.43160725],  
                 [ 0.46993127,  0.86966801],  
                 ...,  
                 [-2.87863922,  7.63277388],  
                 [ 1.98842239,  6.56376886],  
                 [-1.59004664, -10.48487186]], dtype=float32)
```

```
Out[48]: 11893
```

```
In [59]: # set required variables for model comparison

tsne_tbl = pd.DataFrame(columns = [
    'model_name',
    'n_clusters',
    'inertia',
    'silhouette',
    'process_time'])

i_index = []
i_index = 0

# preparation for cross validation and model comparison, each classifier is appended once mode
l is fit

models = []
```

Spectral Clustering on the t-SNE 2-D mapping

```
In [64]: # ... spectraclustering on the t-sne vectors

X_tsne = pd.DataFrame(columns=['t1', 't2'])
X_tsne['t1'] = tsne_results[:,0]
X_tsne['t2'] = tsne_results[:,1]

from sklearn.cluster import SpectralClustering

# If a string, this may be one of
# 'nearest_neighbors', 'precomputed', 'rbf'
# or one of the kernels supported by sklearn.metrics.pairwise_kernels

for n_clstr in range(2, 10):

    tic = time.clock()

    print ("n_clusters = ", n_clstr)

    spc = SpectralClustering(n_clusters = n_clstr,
                             affinity = 'nearest_neighbors')
    spc_labels = spc.fit_predict(X_tsne)
    spc_labels
    spc_silhouette = metrics.silhouette_score(X_tsne,
                                                spc_labels,
                                                metric = 'euclidean',
                                                sample_size = 10000)
    print ("silhouette = ", spc_silhouette)

    toc = time.clock()
# ... -----
# ... - save statistics for model comparison
# ... -----
    exe_time = '{0:.4f}'.format(toc-tic)

    raw_data = {
        'model_name' : 'spc - features',
        'n_clusters' : n_clstr,
        'inertia': 0,
        'silhouette': spc_silhouette,
        'process_time' : exe_time
    }

    df_tbl = pd.DataFrame(raw_data,
                          columns = ['model_name', 'n_clusters', 'inertia', 'silhouette', 'process_time'],
                          index = [i_index + 1])

    tsne_tbl = tsne_tbl.append(df_tbl)

# ... -----
# ... - make some plots of clusters
# ... -----
    plt.figure(figsize=(12, 8));
    plt.subplot(111, axisbg='darkgrey');
```

```
n_clusters = 2

/home/mcdevitt/anaconda3/lib/python3.6/site-packages/sklearn/manifold/spectral_embedding_.py:234: UserWarning: Graph is not fully connected, spectral embedding may not work as expected.
warnings.warn("Graph is not fully connected, spectral embedding"

Out[64]: array([0, 0, 0, ..., 0, 0, 0], dtype=int32)

silhouette = 0.20519

Out[64]: <matplotlib.figure.Figure at 0x7fe799d19e10>

/home/mcdevitt/anaconda3/lib/python3.6/site-packages/matplotlib/cbook.py:136: MatplotlibDeprecationWarning: The axisbg attribute was deprecated in version 2.0. Use facecolor instead.
warnings.warn(message, mplDeprecation, stacklevel=1)

Out[64]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7996bf5c0>

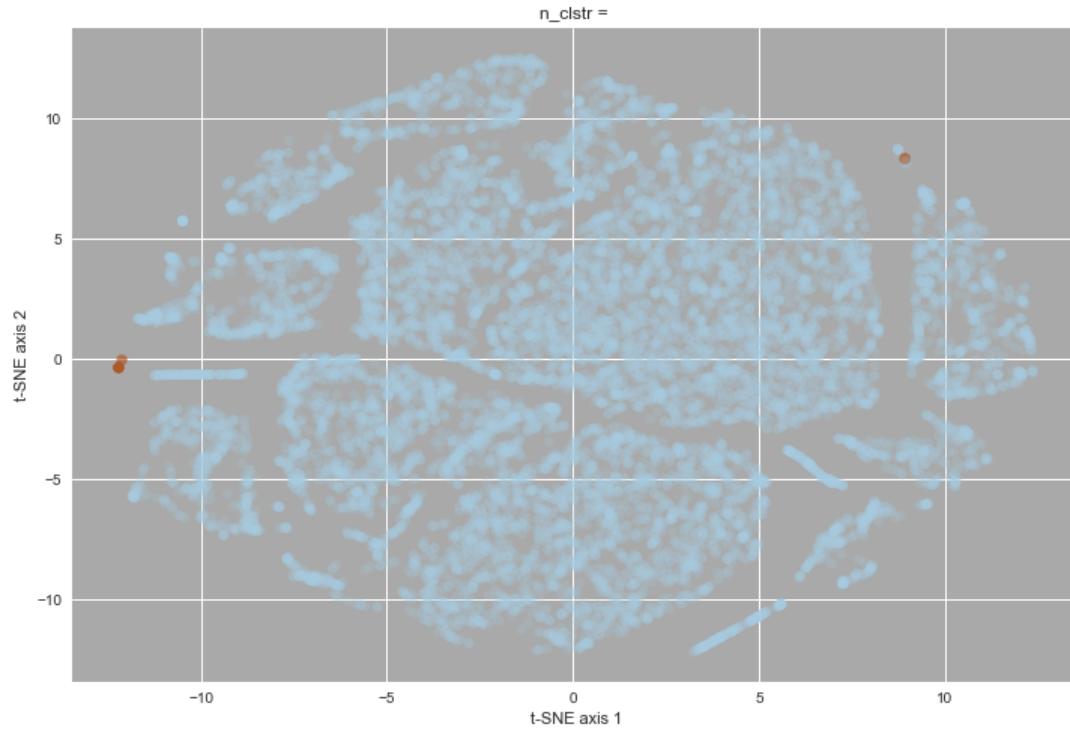
Out[64]: <matplotlib.collections.PathCollection at 0x7fe79951b978>

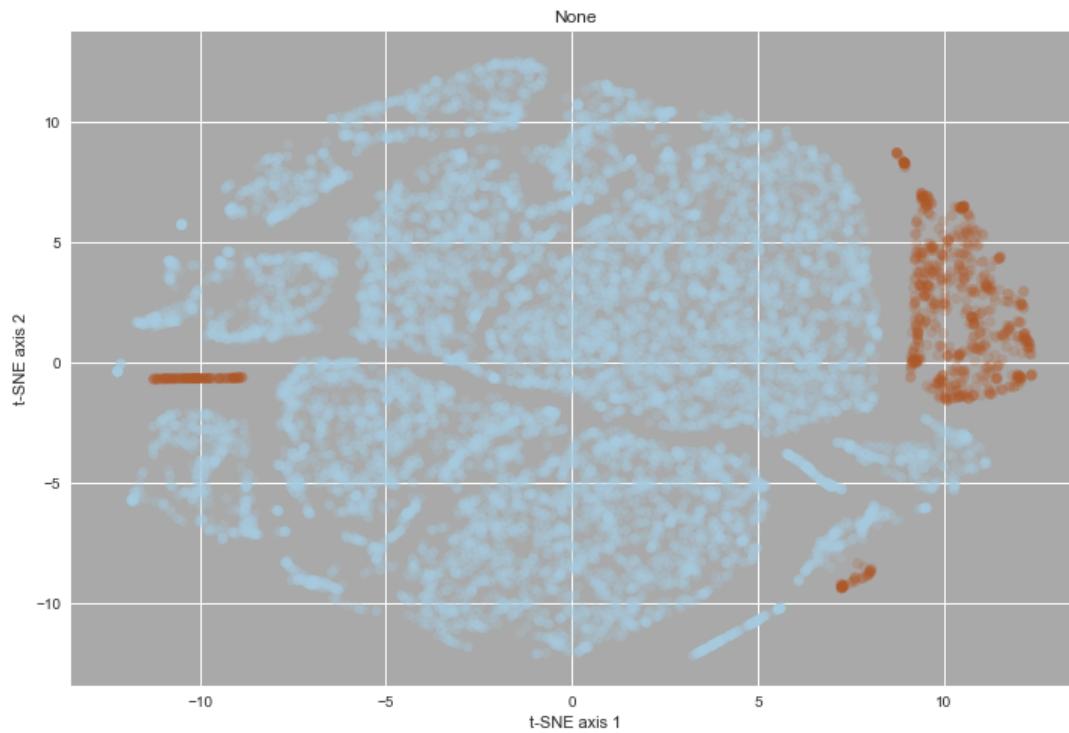
Out[64]: <matplotlib.text.Text at 0x7fe7b6c86f28>

Out[64]: <matplotlib.text.Text at 0x7fe7b6b96fd0>

n_clusters = 2

Out[64]: <matplotlib.text.Text at 0x7fe7996baf28>
```





```
n_clusters = 3

Out[64]: array([1, 1, 0, ..., 0, 0, 1], dtype=int32)

silhouette = 0.145513

Out[64]: <matplotlib.figure.Figure at 0x7fe7996764a8>

Out[64]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b69f5860>

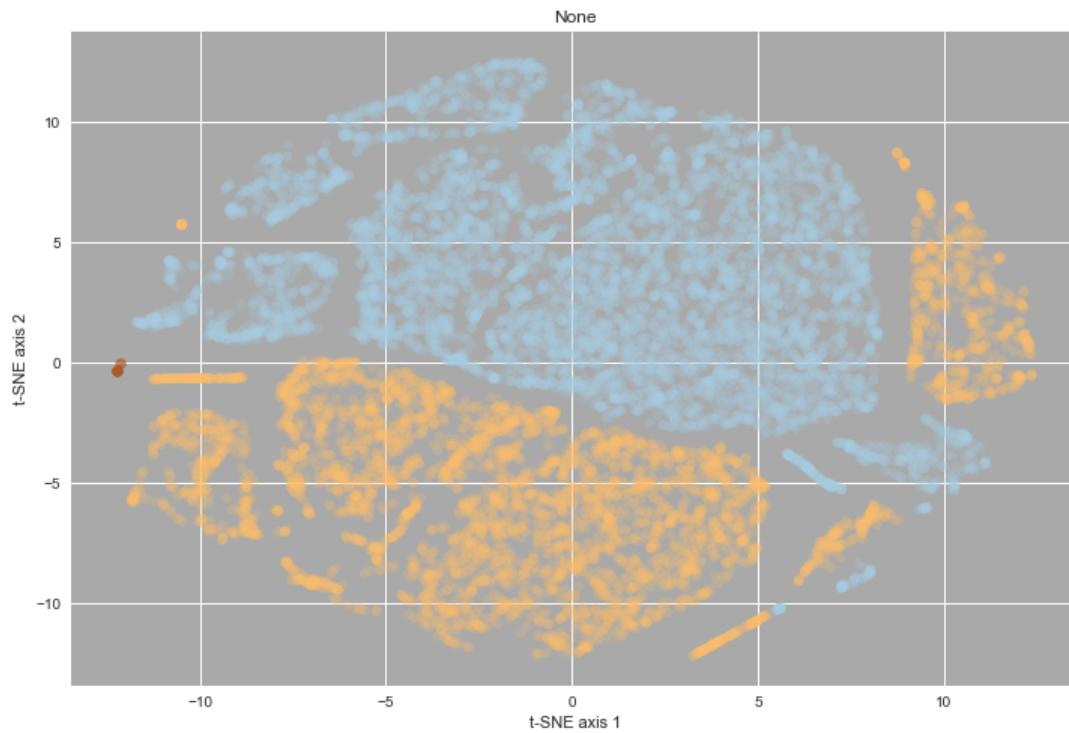
Out[64]: <matplotlib.collections.PathCollection at 0x7fe7b69ddc50>

Out[64]: <matplotlib.text.Text at 0x7fe799608518>

Out[64]: <matplotlib.text.Text at 0x7fe799f550b8>

n_clusters = 3

Out[64]: <matplotlib.text.Text at 0x7fe799801630>
```



```
n_clusters = 4

Out[64]: array([0, 0, 0, ..., 0, 0, 0], dtype=int32)

silhouette = 0.00557496

Out[64]: <matplotlib.figure.Figure at 0x7fe79a0200b8>

Out[64]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6c37c18>

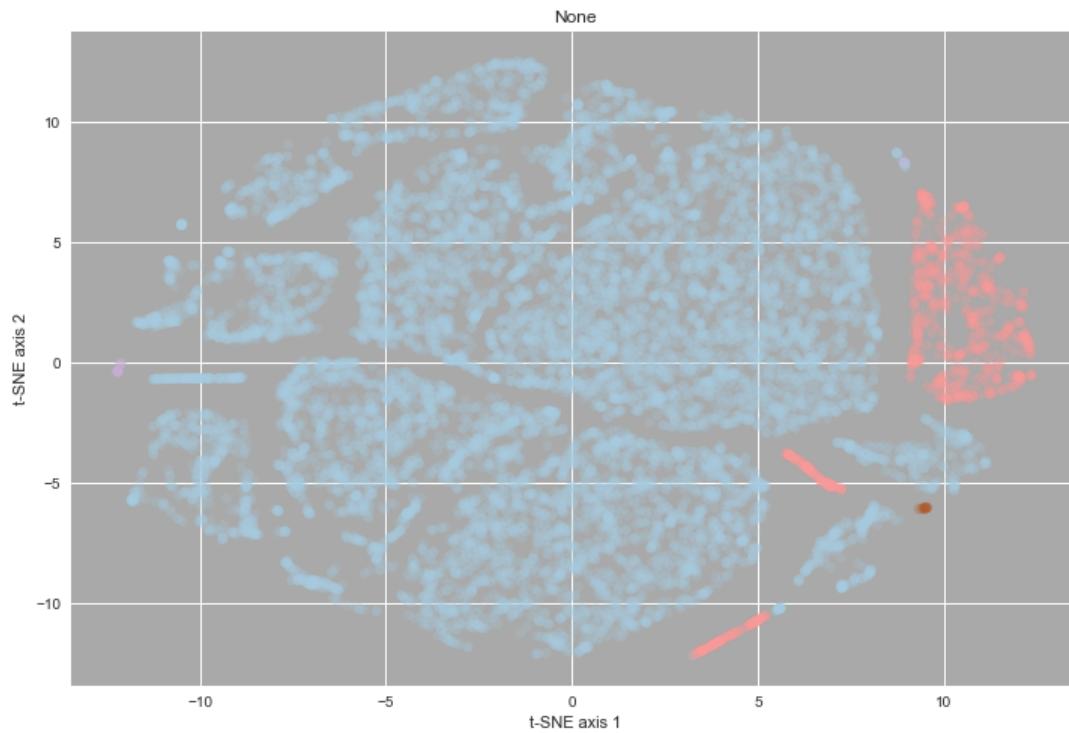
Out[64]: <matplotlib.collections.PathCollection at 0x7fe7b6e044a8>

Out[64]: <matplotlib.text.Text at 0x7fe7993c1b38>

Out[64]: <matplotlib.text.Text at 0x7fe7b6e03828>

n_clusters = 4

Out[64]: <matplotlib.text.Text at 0x7fe7b6c58898>
```



```
n_clusters = 5

Out[64]: array([0, 0, 0, ..., 0, 0, 0], dtype=int32)

silhouette = -0.19882

Out[64]: <matplotlib.figure.Figure at 0x7fe799f8f4a8>

Out[64]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7996c0a90>

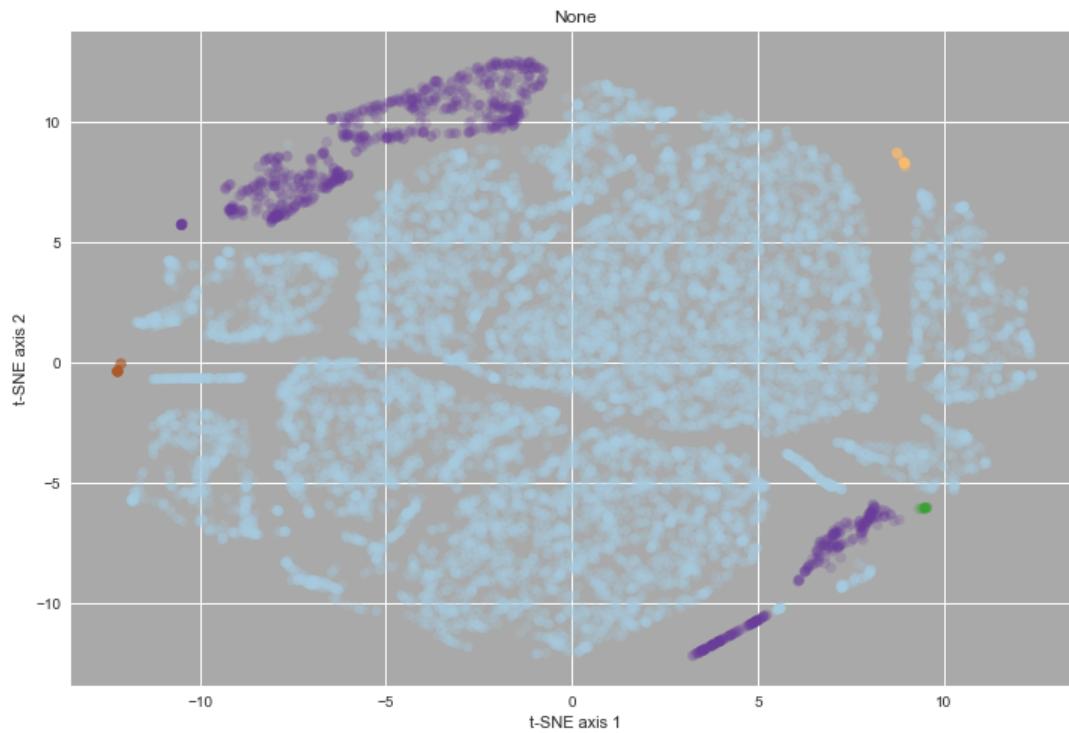
Out[64]: <matplotlib.collections.PathCollection at 0x7fe799d5f390>

Out[64]: <matplotlib.text.Text at 0x7fe7996ace10>

Out[64]: <matplotlib.text.Text at 0x7fe7b6ab2cc0>

n_clusters = 5

Out[64]: <matplotlib.text.Text at 0x7fe7b6a1b940>
```



```
n_clusters = 6

Out[64]: array([1, 0, 1, ..., 5, 1, 1], dtype=int32)

silhouette = -0.220689

Out[64]: <matplotlib.figure.Figure at 0x7fe7b6a0f2e8>

Out[64]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6cdd668>

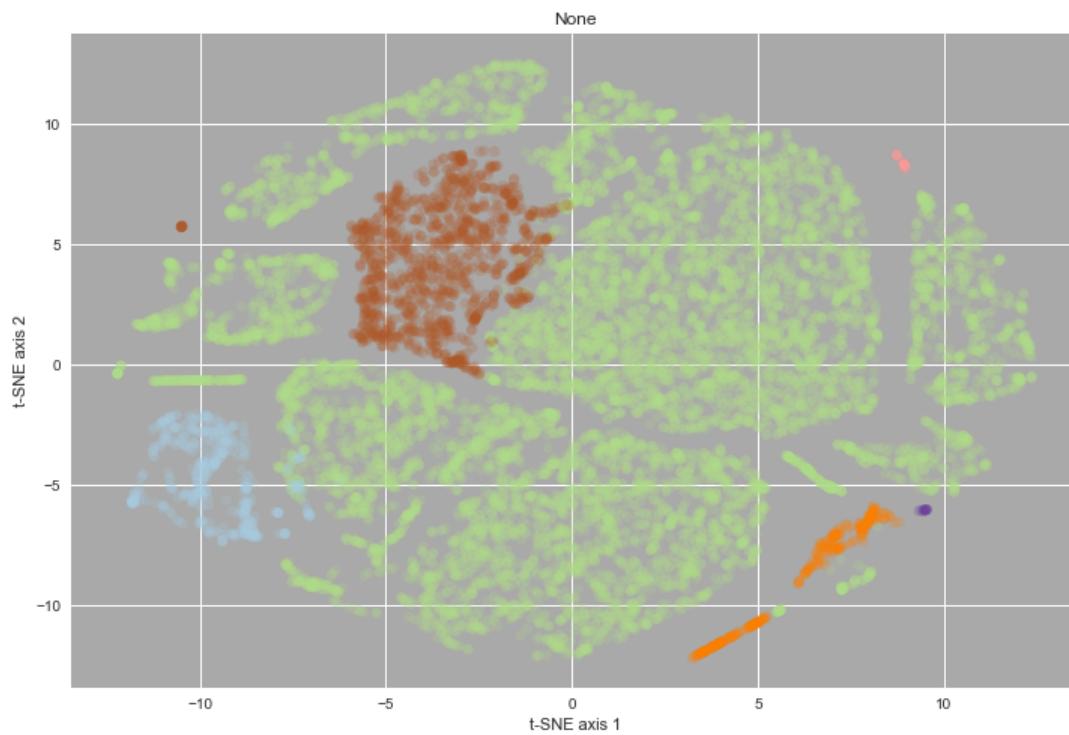
Out[64]: <matplotlib.collections.PathCollection at 0x7fe7b6c5def0>

Out[64]: <matplotlib.text.Text at 0x7fe799817a58>

Out[64]: <matplotlib.text.Text at 0x7fe799936320>

n_clusters = 6

Out[64]: <matplotlib.text.Text at 0x7fe799bb9d68>
```



```
n_clusters = 7

Out[64]: array([0, 2, 0, ..., 0, 0, 0], dtype=int32)

silhouette = -0.2278

Out[64]: <matplotlib.figure.Figure at 0x7fe7994d9dd8>

Out[64]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe79948a860>

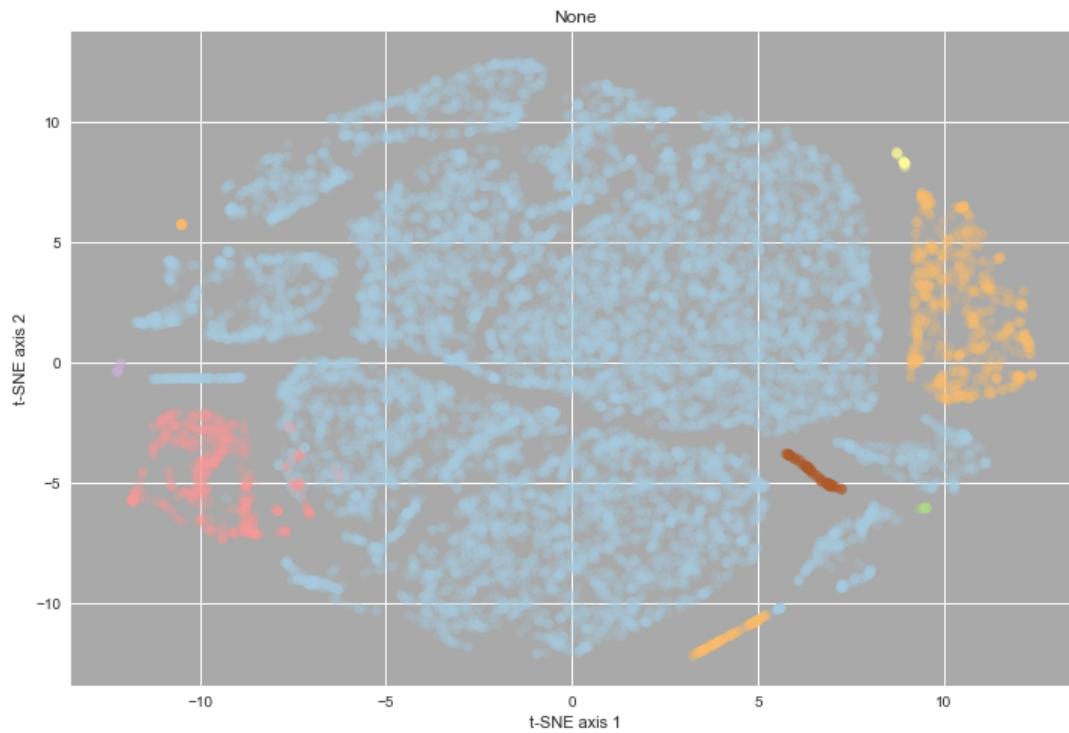
Out[64]: <matplotlib.collections.PathCollection at 0x7fe7b69e5240>

Out[64]: <matplotlib.text.Text at 0x7fe79962b940>

Out[64]: <matplotlib.text.Text at 0x7fe7997ee940>

n_clusters = 7

Out[64]: <matplotlib.text.Text at 0x7fe7994af828>
```



```
n_clusters = 8

Out[64]: array([7, 0, 1, ..., 6, 1, 7], dtype=int32)

silhouette = -0.018651

Out[64]: <matplotlib.figure.Figure at 0x7fe7b6d520b8>

Out[64]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe79948a2b0>

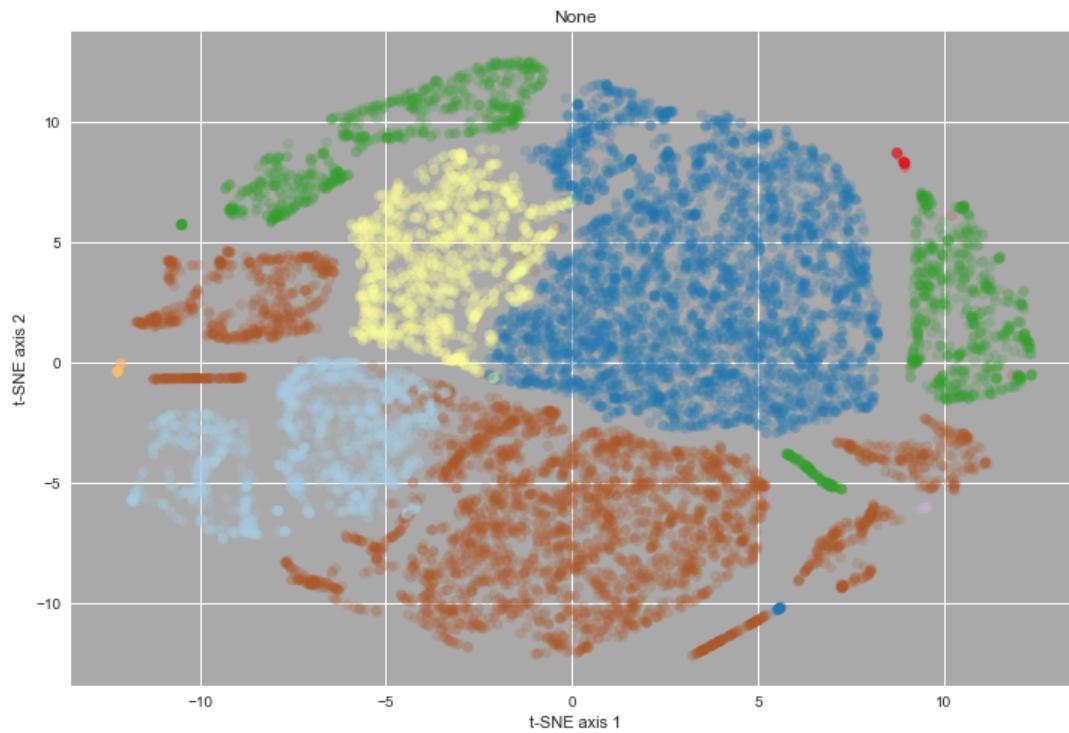
Out[64]: <matplotlib.collections.PathCollection at 0x7fe7b68ecba8>

Out[64]: <matplotlib.text.Text at 0x7fe7b6dd1f28>

Out[64]: <matplotlib.text.Text at 0x7fe7b64a3da0>

n_clusters = 8

Out[64]: <matplotlib.text.Text at 0x7fe7b6a4bcc0>
```



```
n_clusters = 9

Out[64]: array([8, 6, 1, ..., 0, 1, 8], dtype=int32)

silhouette = -0.028307

Out[64]: <matplotlib.figure.Figure at 0x7fe7b6b17550>

Out[64]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6a8bf60>

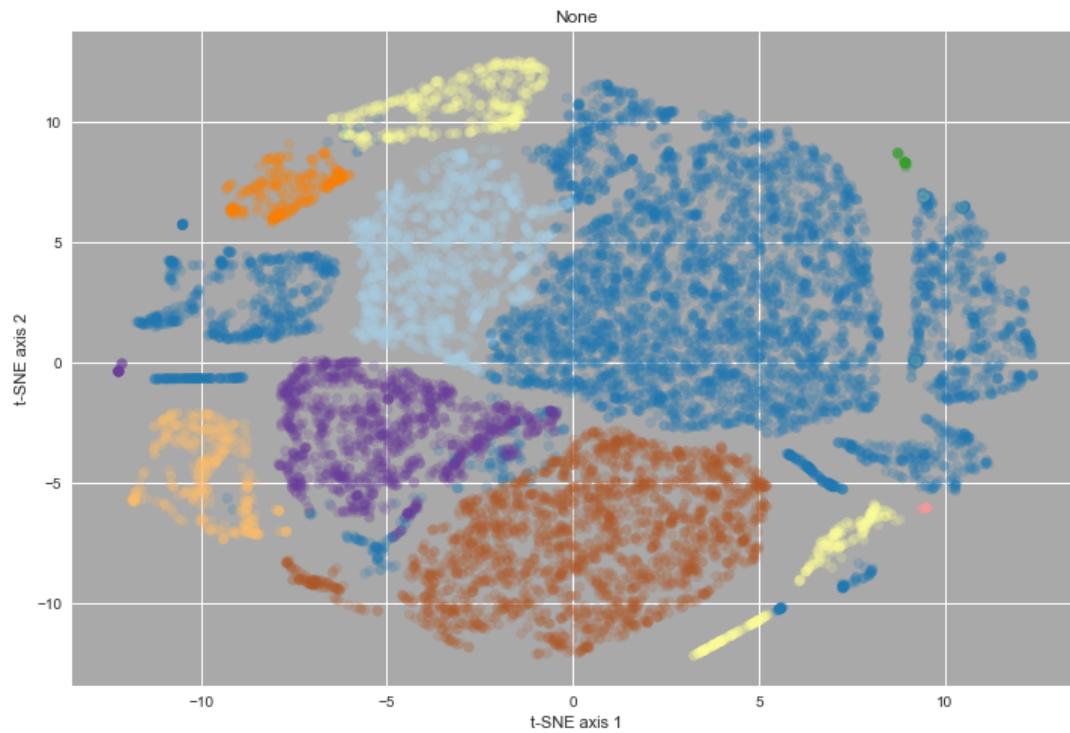
Out[64]: <matplotlib.collections.PathCollection at 0x7fe7b6cec3c8>

Out[64]: <matplotlib.text.Text at 0x7fe7b6d56e80>

Out[64]: <matplotlib.text.Text at 0x7fe7b6570908>

n_clusters = 9

Out[64]: <matplotlib.text.Text at 0x7fe7b69e5ac8>
```



In [65]:

```
# ... -----
# ... - plot metrics across models for comparison
# ... -----
# ... silhouette values

plt.figure(figsize=(16, 6));

# ... silhouette values

plt.subplot(131);
plt.scatter(tsne_tbl['n_clusters'],
            tsne_tbl['silhouette'],
            s = 40,
            linewidths = 1.0,
            marker = '^',
            edgecolors = 'black',
            alpha = 0.90);

plt.plot(tsne_tbl['n_clusters'],
         tsne_tbl['silhouette'])

plt.xlabel('n_clusters'), plt.ylabel('silhouette');
plt.grid();

# ... inertia values

plt.subplot(132);
plt.scatter(tsne_tbl['n_clusters'],
            tsne_tbl['inertia'],
            s = 40,
            linewidths = 1.0,
            marker = '^',
            edgecolors = 'black',
            alpha = 0.90);

plt.plot(tsne_tbl['n_clusters'],
         tsne_tbl['inertia'])

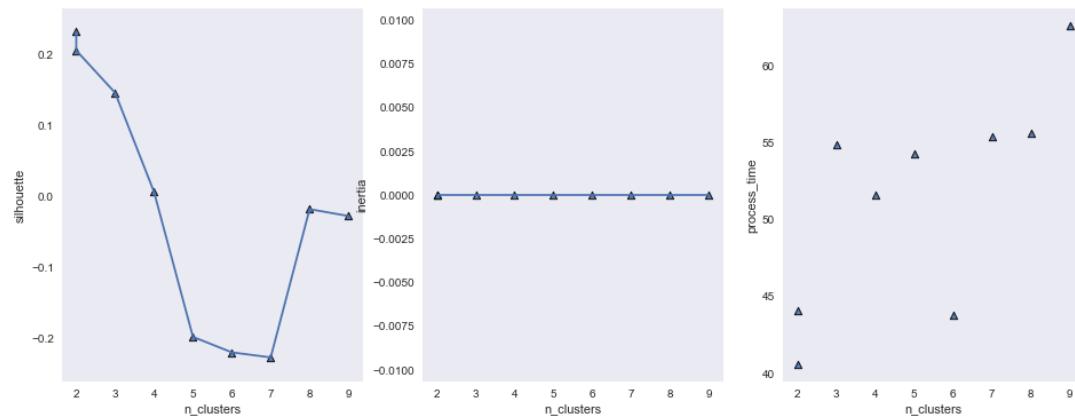
plt.xlabel('n_clusters'), plt.ylabel('inertia');
plt.grid();

# ... process time

plt.subplot(133);
plt.scatter(tsne_tbl['n_clusters'],
            tsne_tbl['process_time'],
            s = 40,
            linewidths = 1.0,
            marker = '^',
            edgecolors = 'black',
            alpha = 0.90);

# plt.plot(tsne_tbl['n_clusters'],
#          tsne_tbl['process_time'])

plt.xlabel('n clusters'), plt.ylabel('process time');
```



- choose spectral clustering with preferred number of clusters

```
In [66]: n_clusters_chosen = 8

for n_clstr in range(n_clusters_chosen, n_clusters_chosen+1):

    tic = time.clock()

    print ("n_clusters = ", n_clstr)

    spc = SpectralClustering(n_clusters = n_clstr,
                            affinity = 'nearest_neighbors')
    spc_labels = spc.fit_predict(X_tsne)
    spc_labels
    spc_silhouette = metrics.silhouette_score(X_tsne,
                                                spc_labels,
                                                metric = 'euclidean',
                                                sample_size = 10000)
    print ("silhouette = ", spc_silhouette)

    toc = time.clock()
```

```
n_clusters = 8

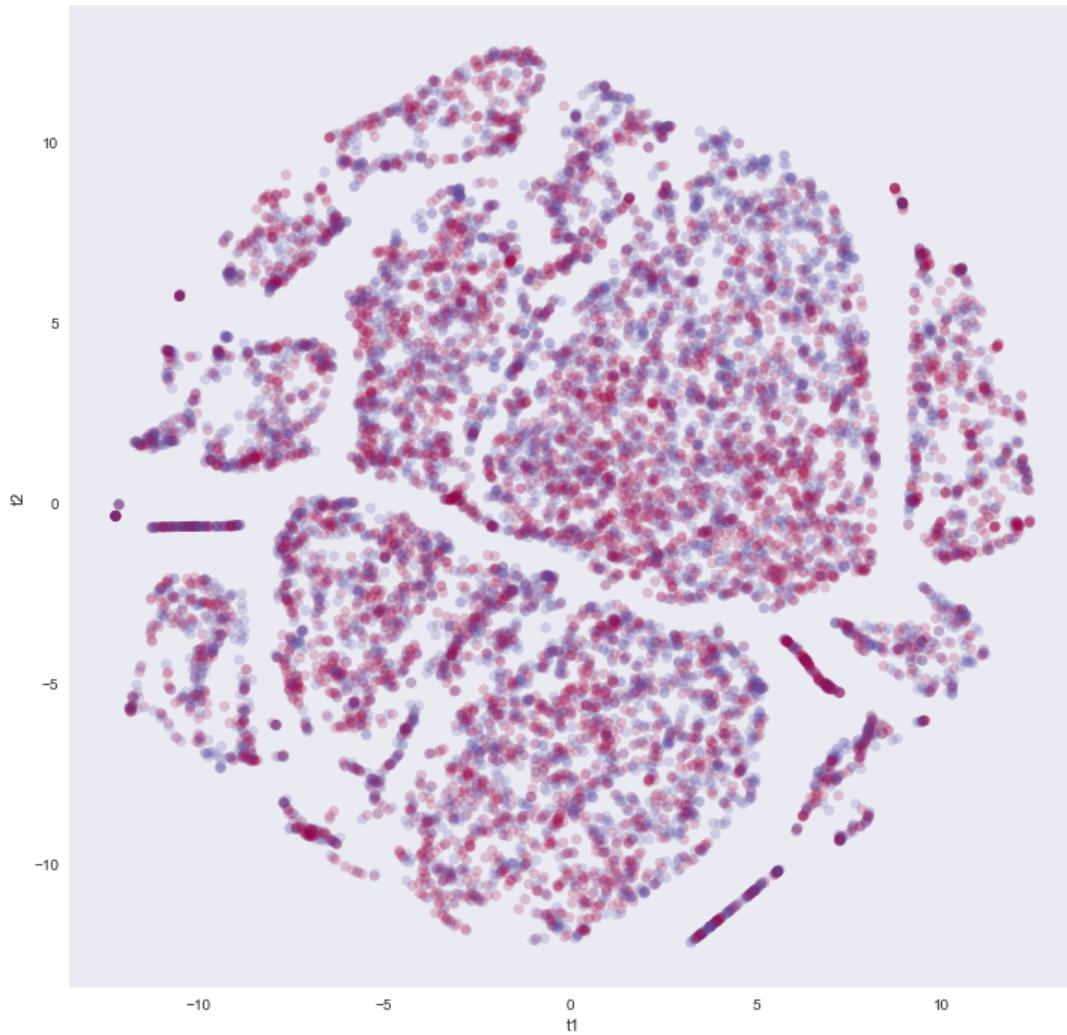
/home/mcdevitt/anaconda3/lib/python3.6/site-packages/sklearn/manifold/spectral_embedding_.py:234: UserWarning: Graph is not fully connected, spectral embedding may not work as expected.
  warnings.warn("Graph is not fully connected, spectral embedding"
Out[66]: array([1, 7, 1, ..., 0, 1, 1], dtype=int32)

silhouette = -0.238869

Out[66]: <matplotlib.figure.Figure at 0x7fe7994956d8>

Out[66]: <matplotlib.collections.PathCollection at 0x7fe7b6c86320>

Out[66]: (<matplotlib.text.Text at 0x7fe799513ac8>,
<matplotlib.text.Text at 0x7fe799bc1898>)
```



```
In [68]: X_all_together = copy.deepcopy(X1)

len(X_all_together)

X_all_together['ln_shares'] = X1_ln_shares
X_all_together['popular'] = X1_popular
X_all_together['spc_labels'] = spc_labels

X_all_together['t1'] = tsne_results[:,0]
X_all_together['t2'] = tsne_results[:,1]
```

```
out[68]: 11893
```

out[68]:

	count	mean	std	min	25%	
n_tokens_title	11893.0	10.376776	2.100059	2.000000	9.000000	10.
num_keywords	11893.0	7.235349	1.909407	1.000000	6.000000	7.
data_channel_is_lifestyle	11893.0	0.052888	0.223820	0.000000	0.000000	0.
data_channel_is_entertainment	11893.0	0.175481	0.380394	0.000000	0.000000	0.
data_channel_is_socmed	11893.0	0.057765	0.233308	0.000000	0.000000	0.
kw_avg_max	11893.0	1.925615	0.995841	0.004170	1.282797	1.
weekday_is_monday	11893.0	0.169091	0.374848	0.000000	0.000000	0.
weekday_is_tuesday	11893.0	0.183217	0.386861	0.000000	0.000000	0.
weekday_is_wednesday	11893.0	0.185319	0.388573	0.000000	0.000000	0.
weekday_is_thursday	11893.0	0.183385	0.386998	0.000000	0.000000	0.
weekday_is_friday	11893.0	0.143193	0.350285	0.000000	0.000000	0.
is_weekend	11893.0	0.135794	0.342584	0.000000	0.000000	0.
global_subjectivity	11893.0	0.444601	0.116513	0.000000	0.396786	0.
global_rate_positive_words	11893.0	0.039731	0.017416	0.000000	0.028549	0.
rate_positive_words	11893.0	0.684280	0.188235	0.000000	0.600000	0.
max_positive_polarity	11893.0	0.756673	0.245948	0.000000	0.600000	0.
min_negative_polarity	11893.0	0.482372	0.289514	0.000000	0.300000	0.
max_negative_polarity	11893.0	0.892134	0.093882	0.000000	0.875000	0.
title_sentiment_polarity	11893.0	1.073321	0.264616	0.000000	1.000000	1.
abs_title_subjectivity	11893.0	0.342324	0.188949	0.000000	0.166667	0.
ln_n_tokens_content	11893.0	5.881019	1.235073	0.000000	5.501258	5.
ln_num_hrefs	11893.0	2.151420	0.800923	0.000000	1.609438	2.
ln_num_imgs	11893.0	1.114733	0.968991	0.000000	0.693147	0.
ln_num_videos	11893.0	0.396894	0.677641	0.000000	0.000000	0.
ln_kw_min_min	11893.0	1.164188	1.712570	0.000000	0.000000	0.
ln_kw_avg_min	11893.0	5.290346	1.141318	0.000000	4.952300	5.
ln_kw_min_max	11893.0	5.032837	4.526671	0.000000	0.000000	7.
ln_kw_avg_avg	11893.0	7.978300	0.483315	0.000000	7.781397	7.
ln_self_reference_avg_shares	11893.0	6.693919	3.242012	0.000000	6.906254	7.
ln_LDA_00	11893.0	0.150010	0.194340	0.018019	0.024750	0.
ln_LDA_01	11893.0	0.116420	0.165225	0.018022	0.024702	0.
ln_LDA_02	11893.0	0.169444	0.206484	0.018019	0.025052	0.
ln_LDA_03	11893.0	0.177447	0.216042	0.018019	0.028171	0.
ln_LDA_04	11893.0	0.188006	0.213576	0.018019	0.028173	0.
ln_global_rate_negative_words	11893.0	0.016316	0.010366	0.000000	0.009569	0.

```
In [74]: # boxplot across clusters for each feature ...
import seaborn as sns

col_names = X_all_together.columns.values.tolist()

for col in col_names :

    _ = plt.figure(figsize=(24, 8));

    # ... feature distribution color map

    _ = plt.scatter(X_all_together['t1'], X_all_together['t2'],
                    c = X_all_together[col],
                    cmap = plt.cm.Spectral,
                    s = 50,
                    linewidths = 0,
                    alpha = 0.30)
    _ = plt.title(col)

    # ... feature boxplots

    _ = plt.subplot(132, axisbg='darkgrey');
    sns.boxplot(x = "spc_labels", y = col, data = X_all_together);

    # ... cluster color map

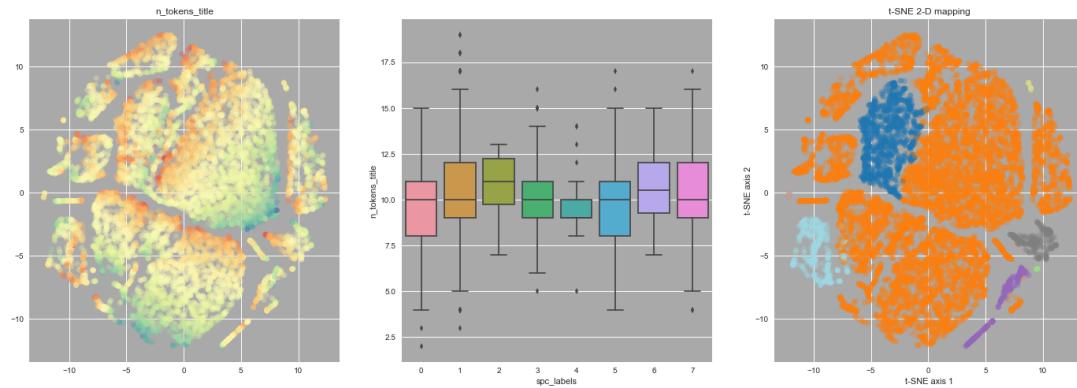
    _ = plt.subplot(133, axisbg='darkgrey');

    _ = plt.scatter(X_all_together['t1'], X_all_together['t2'],
                    c = spc_labels,
                    cmap = plt.cm.tab20,
                    s = 50,
                    linewidths = 0,
                    alpha = 0.30)
    _ = plt.xlabel('t-SNE axis 1')
    _ = plt.ylabel('t-SNE axis 2')
    _ = plt.title('t-SNE 2-D mapping')

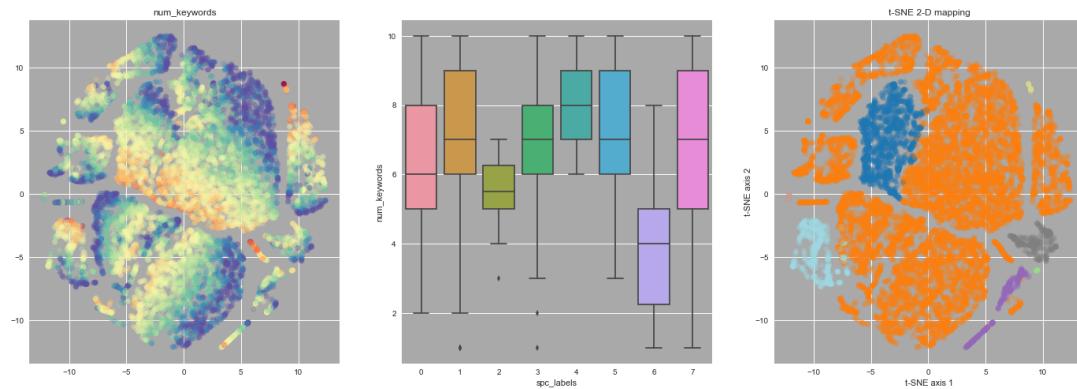
plt.show();
```

```
/home/mcdevitt/anaconda3/lib/python3.6/site-packages/matplotlib/cbook.py:136: MatplotlibDeprecationWarning: The axisbg attribute was deprecated in version 2.0. Use facecolor instead.
  warnings.warn(message, mplDeprecation, stacklevel=1)
```

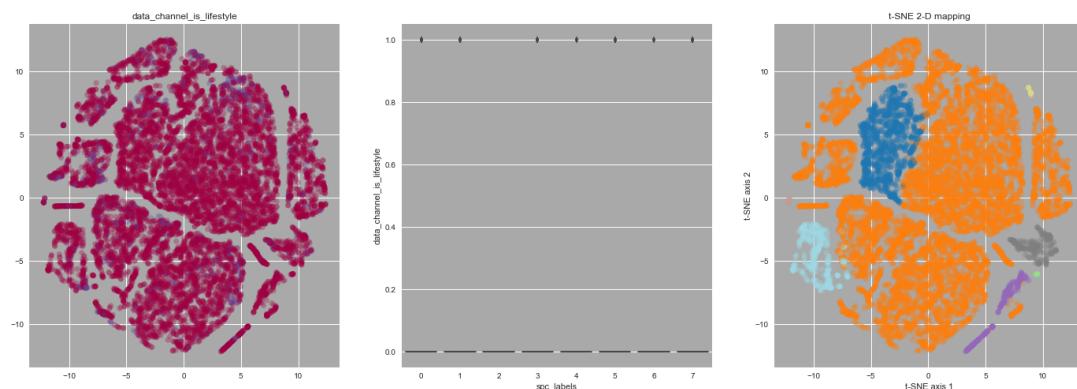
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe79903ddd8>



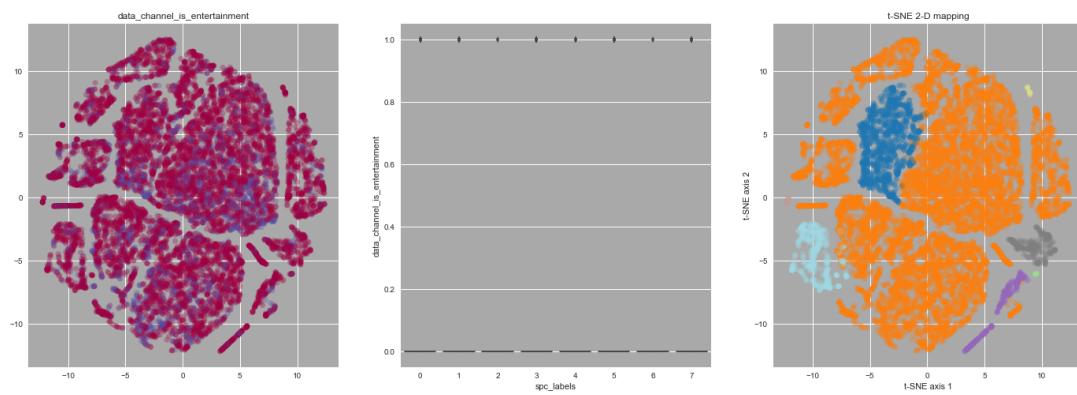
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5acf080>



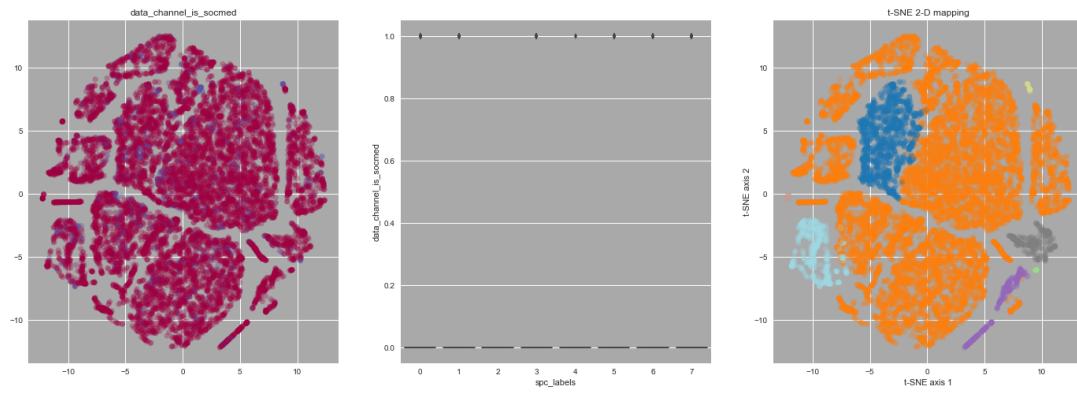
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5ed4438>



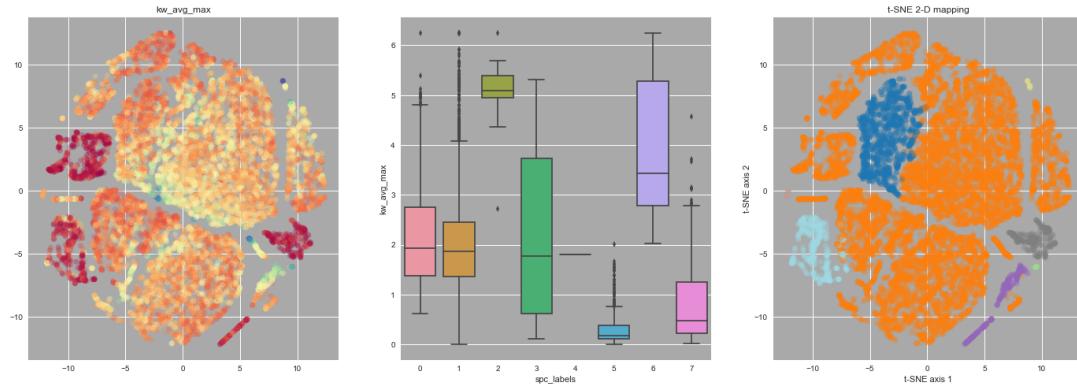
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5ae2a58>



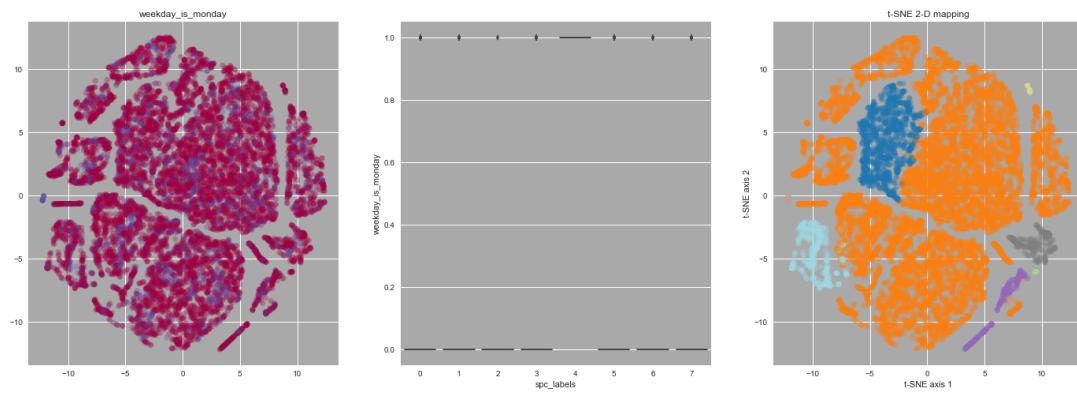
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5aa1518>
```



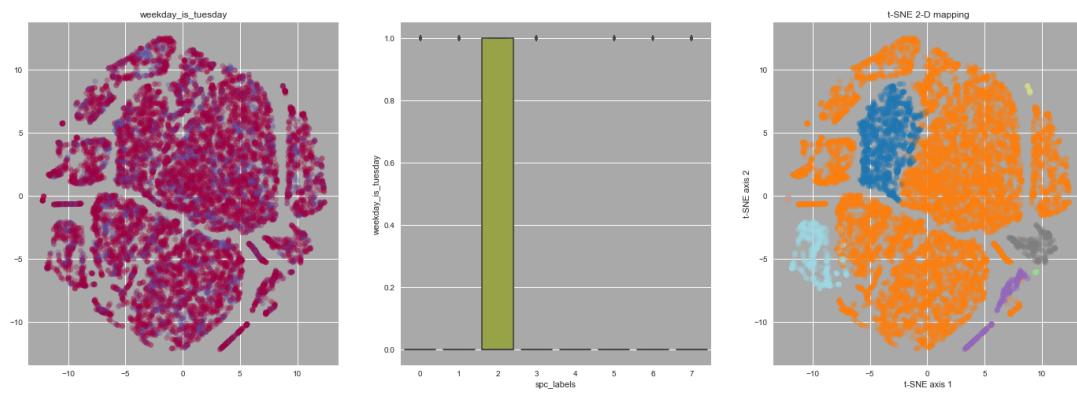
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b64907f0>
```



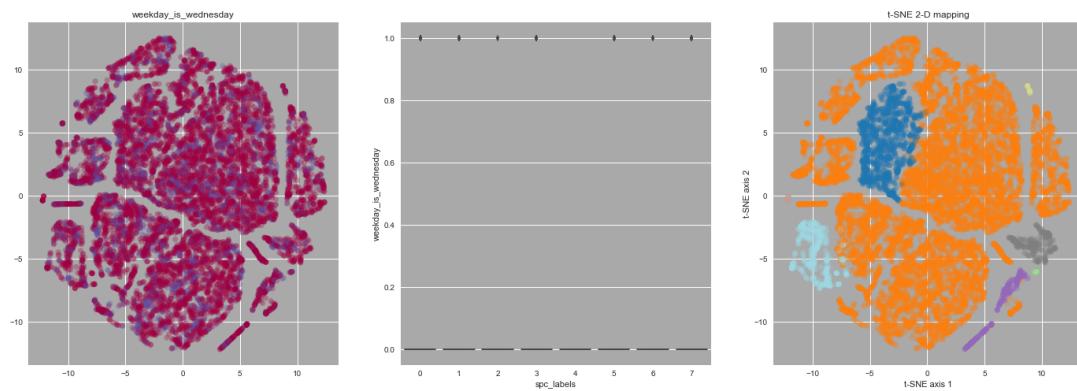
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b55ad208>
```



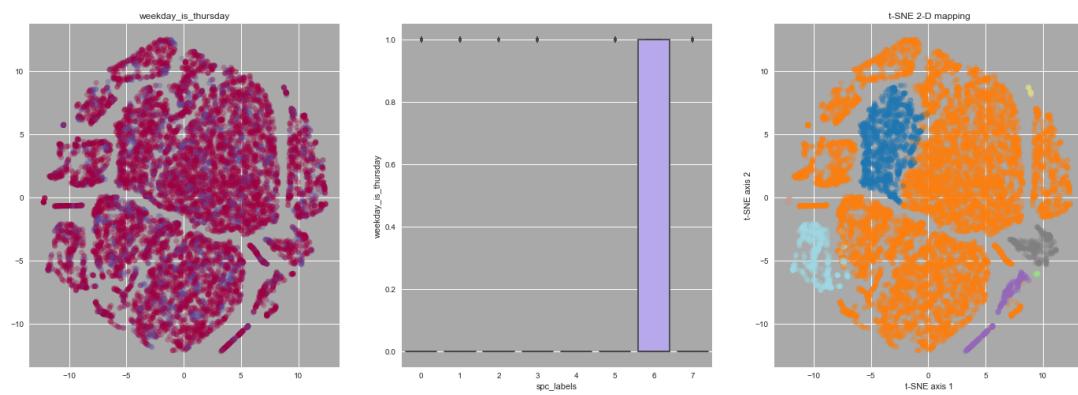
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5fd89b0>



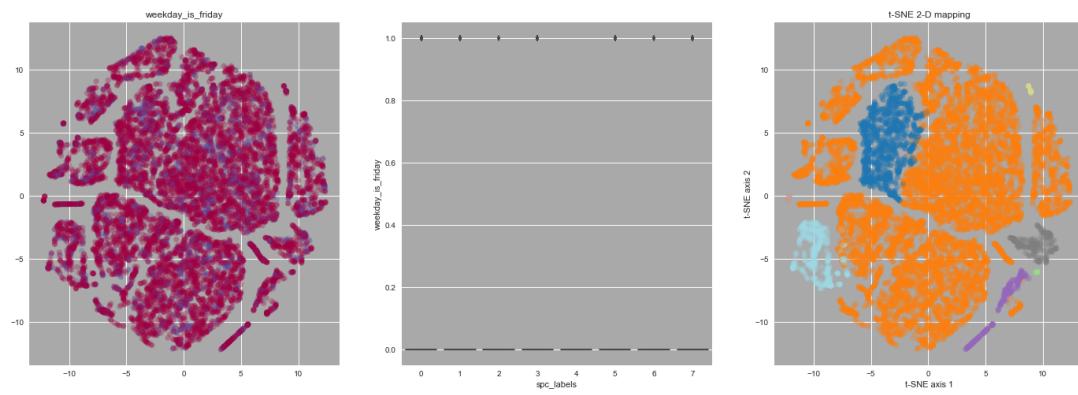
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe798a009b0>



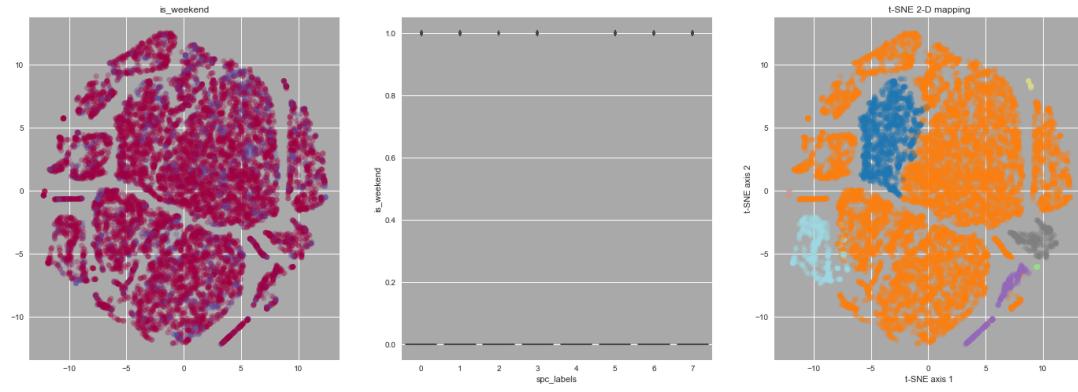
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7989269b0>



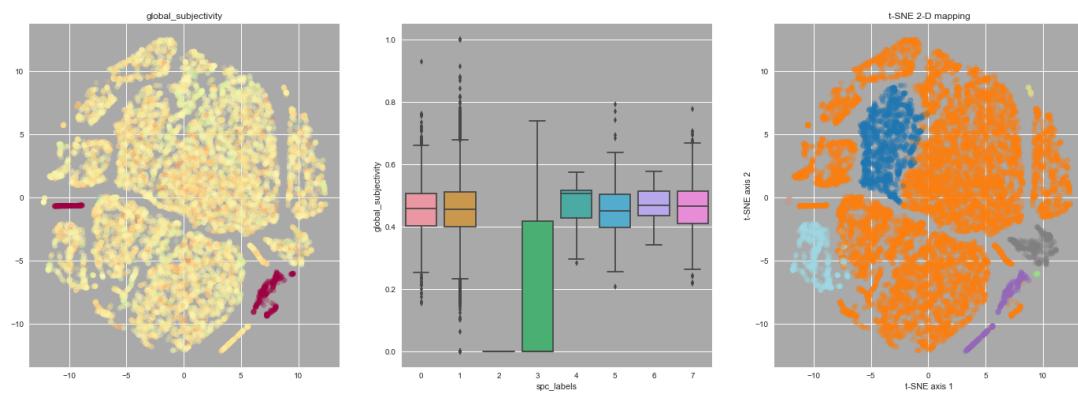
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe798f5b240>



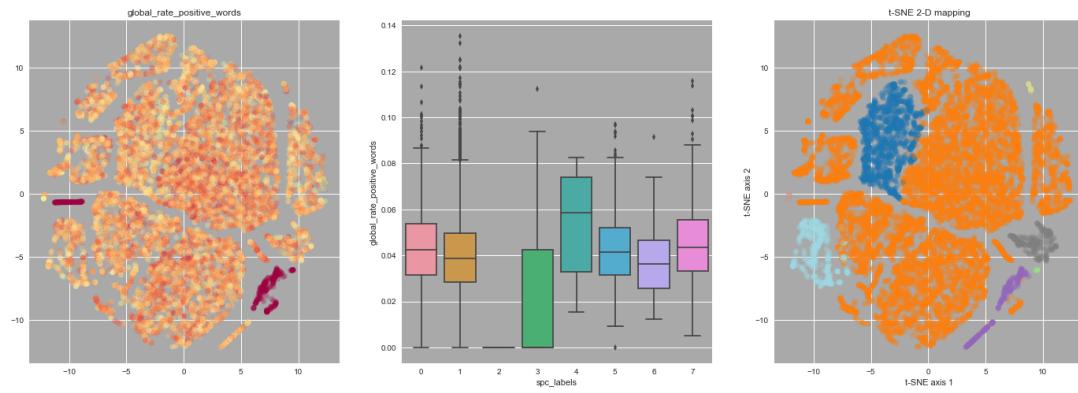
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b58426d8>



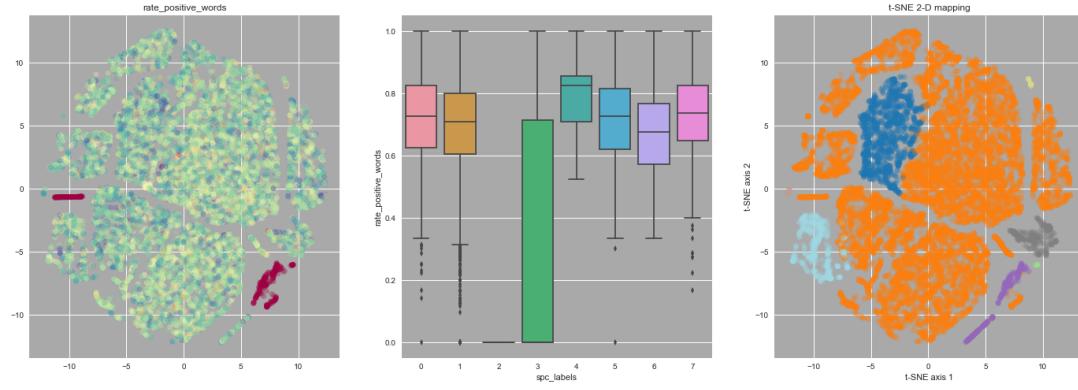
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b592d668>



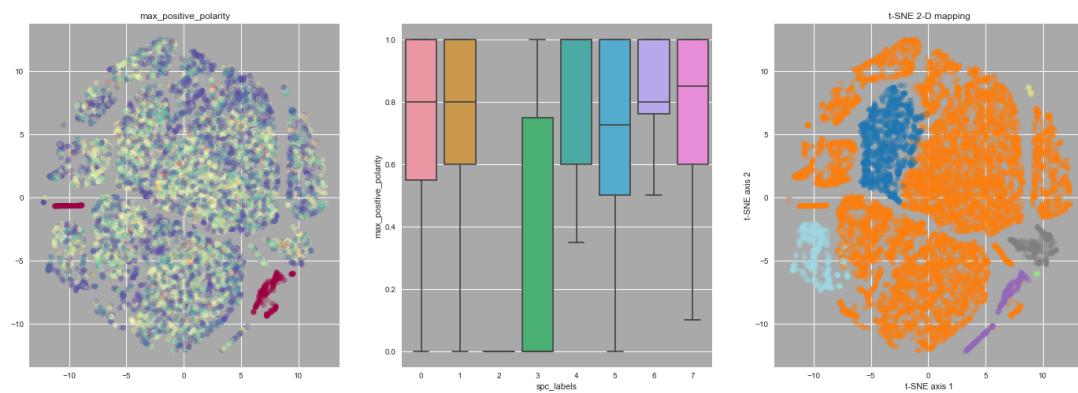
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe798ed0630>
```



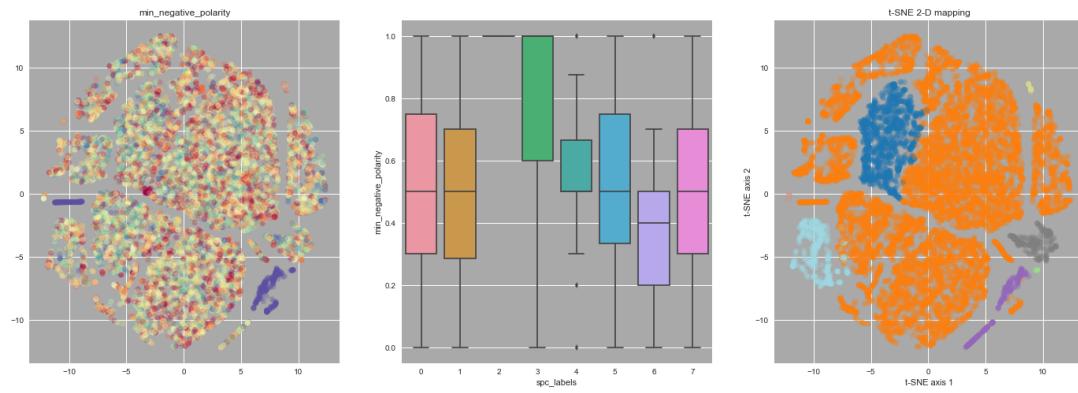
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b57c2a90>
```



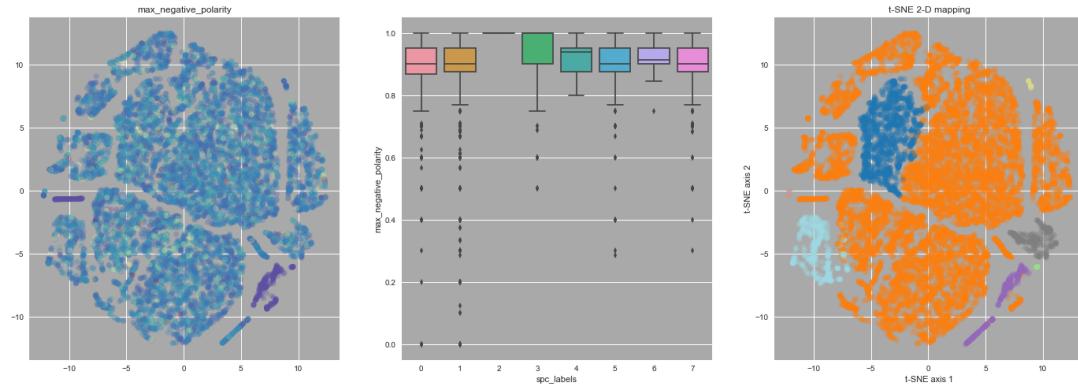
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b58dd940>
```



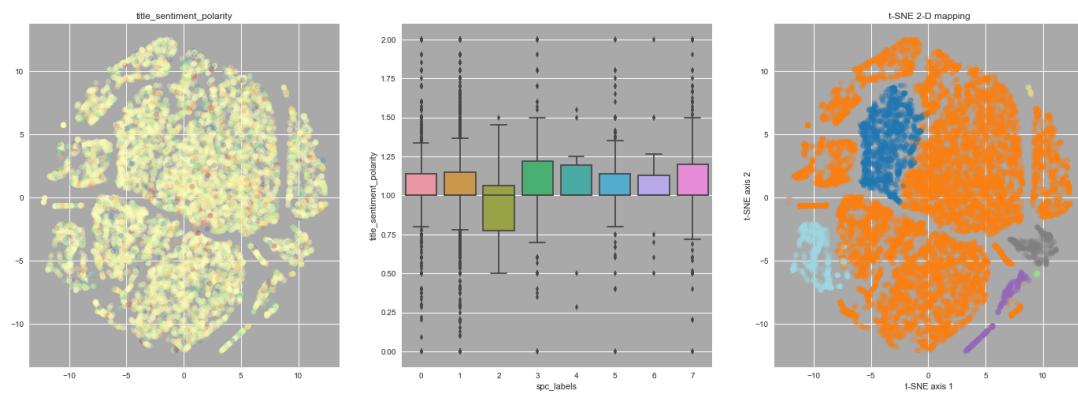
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7990f0710>
```



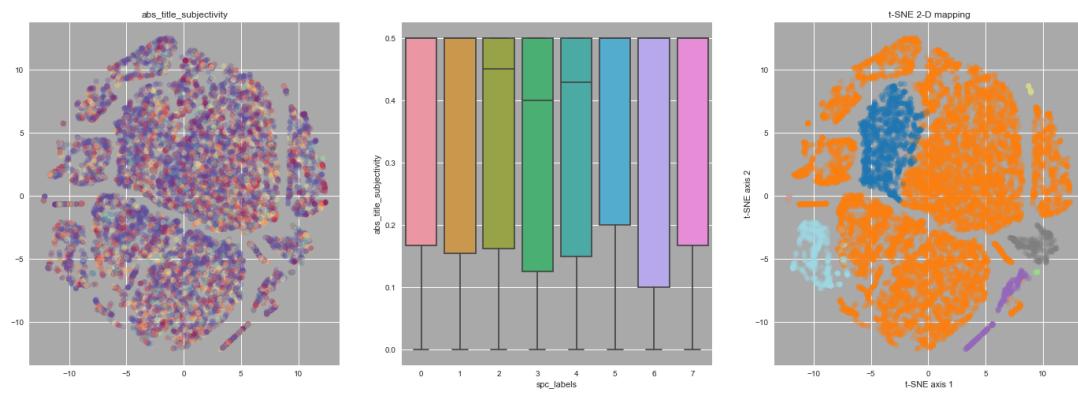
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe798b126d8>
```



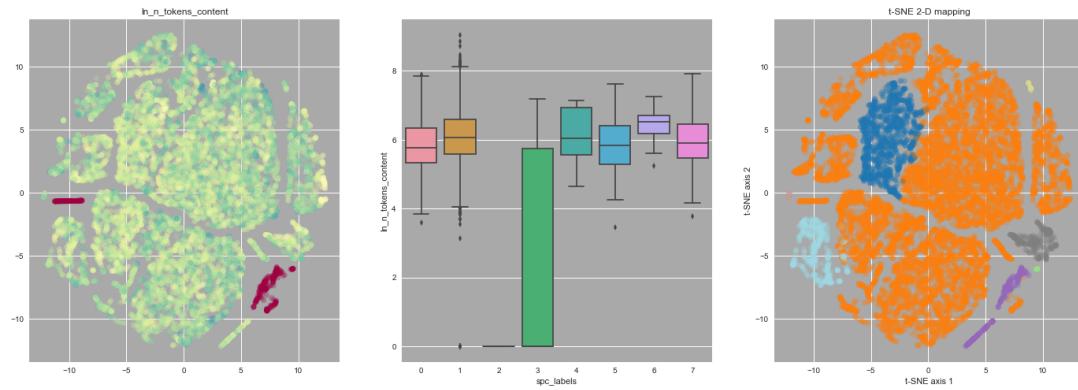
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b65eddd8>
```



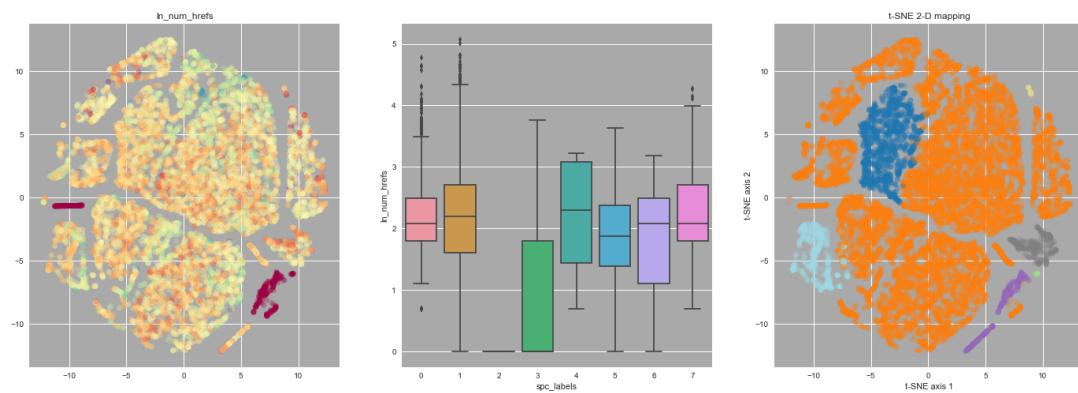
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5944e48>
```



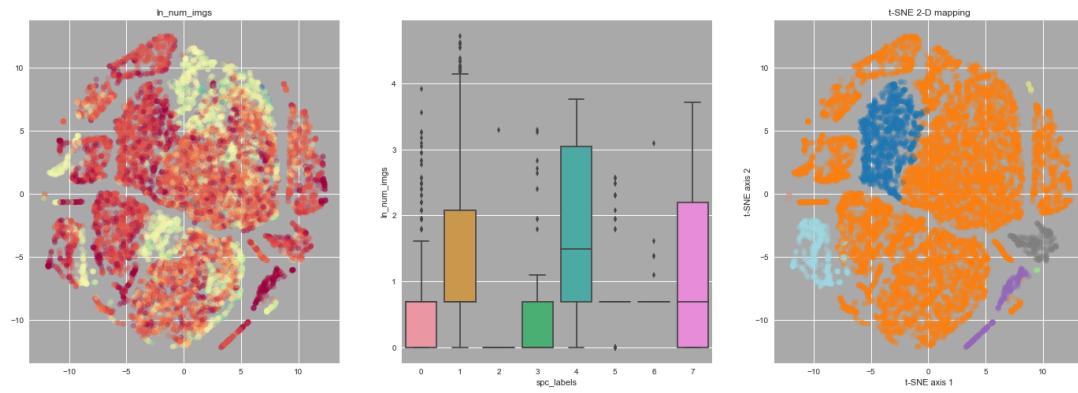
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b6880470>
```



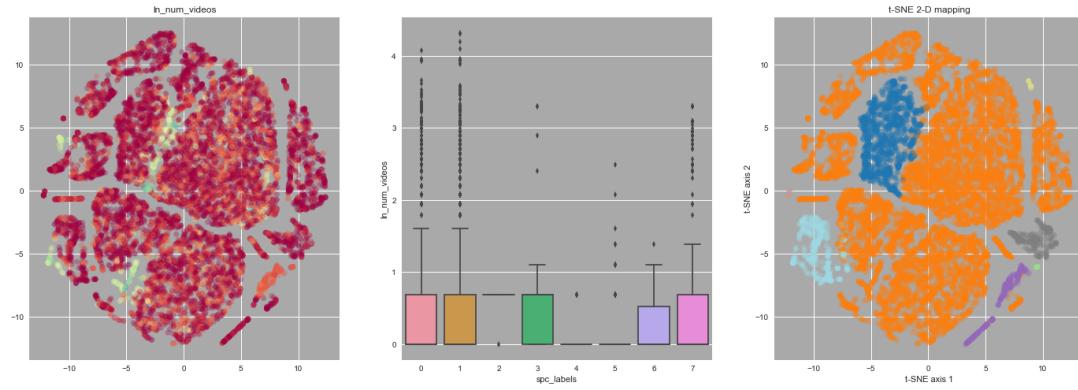
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5967d30>
```



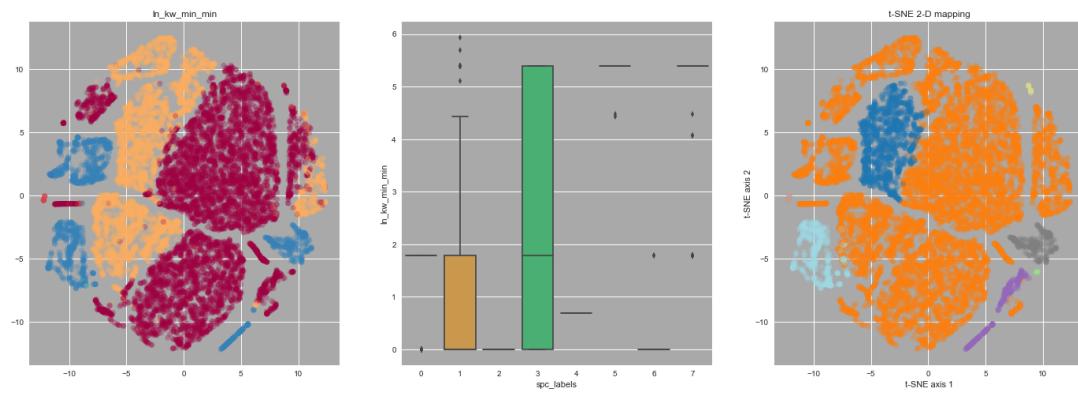
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b60a72e8>
```



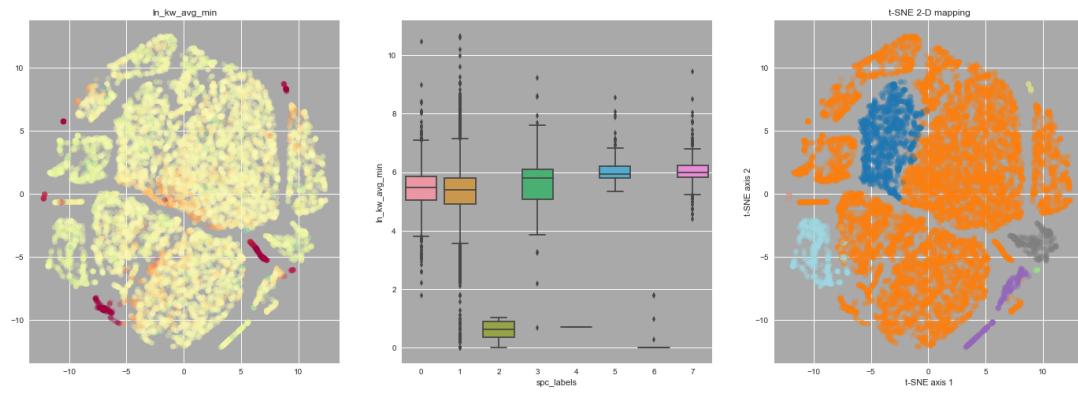
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b559d6d8>
```



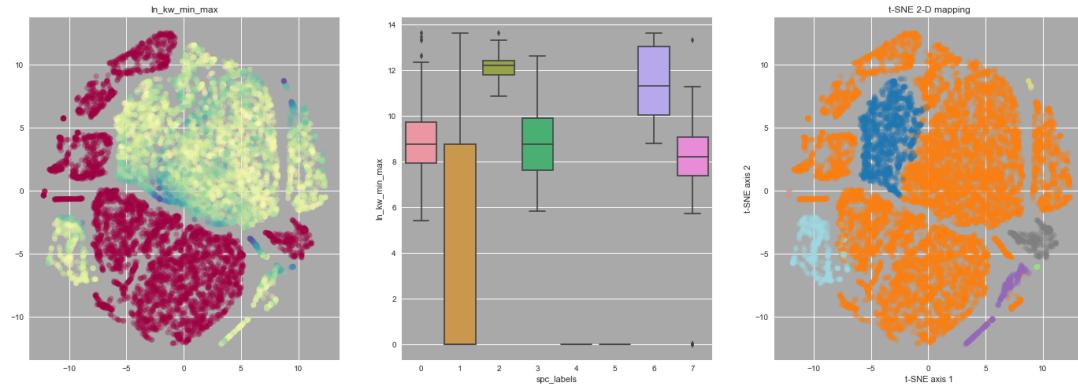
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b600c358>
```



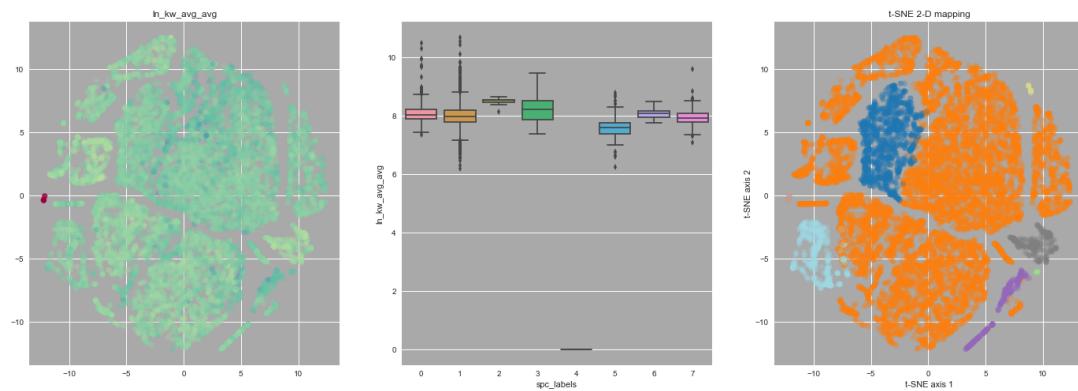
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7991ed940>



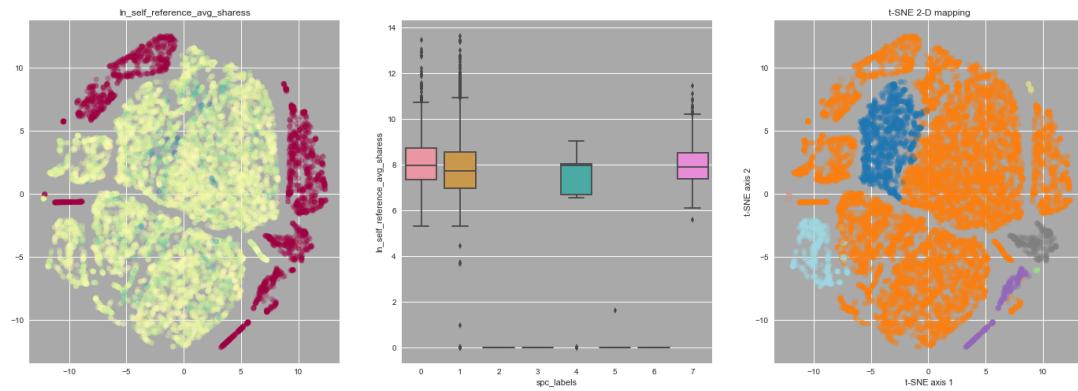
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5dc57f0>



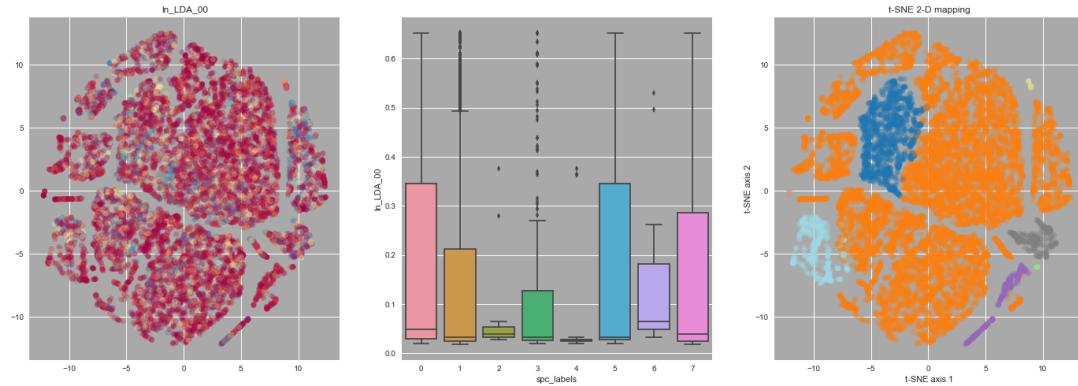
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe798ec8c18>



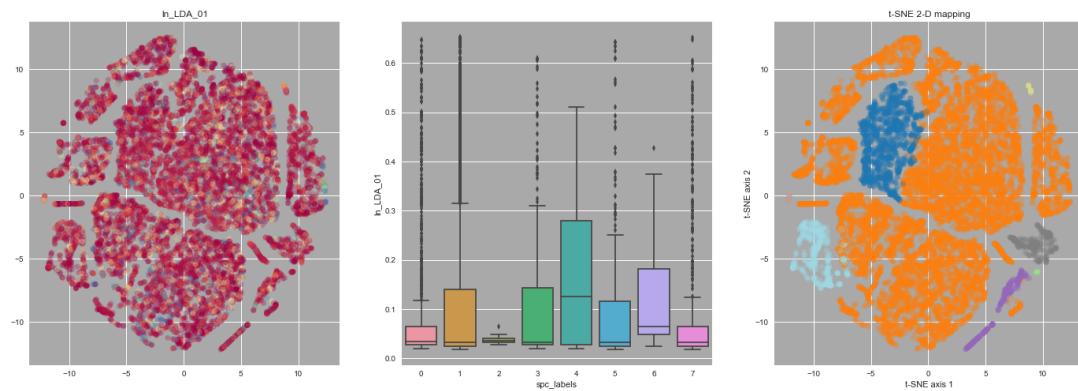
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe799353dd8>
```



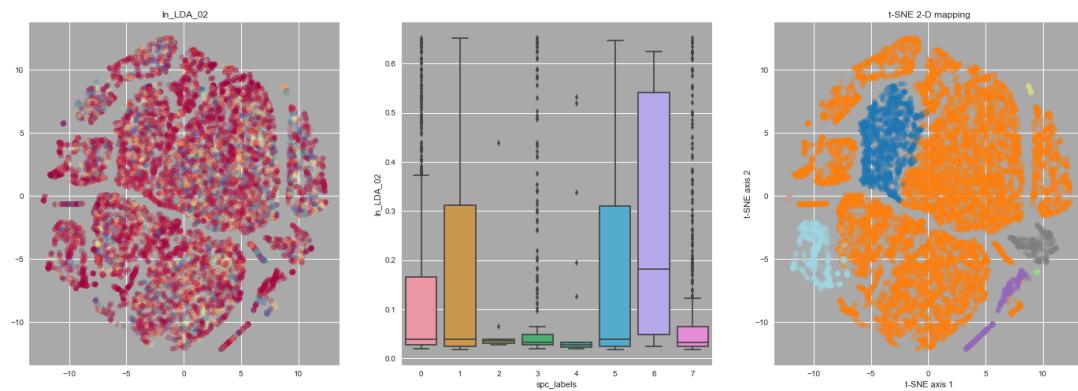
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe798e87c50>
```



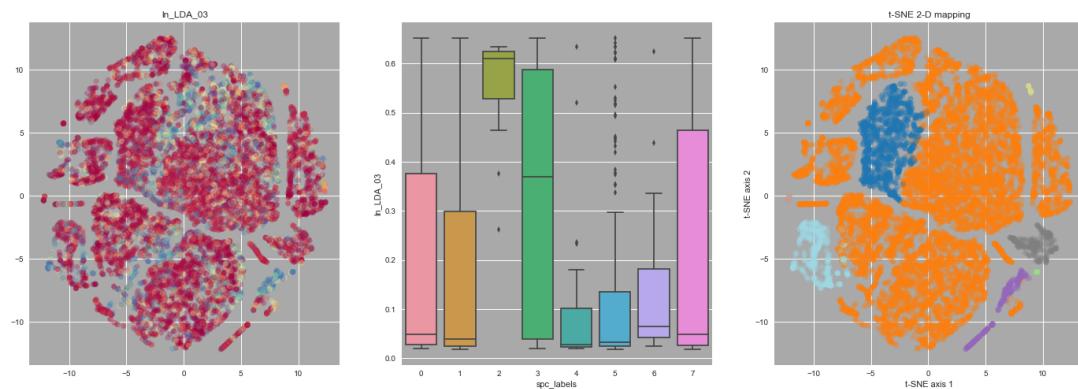
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5ddf550>
```



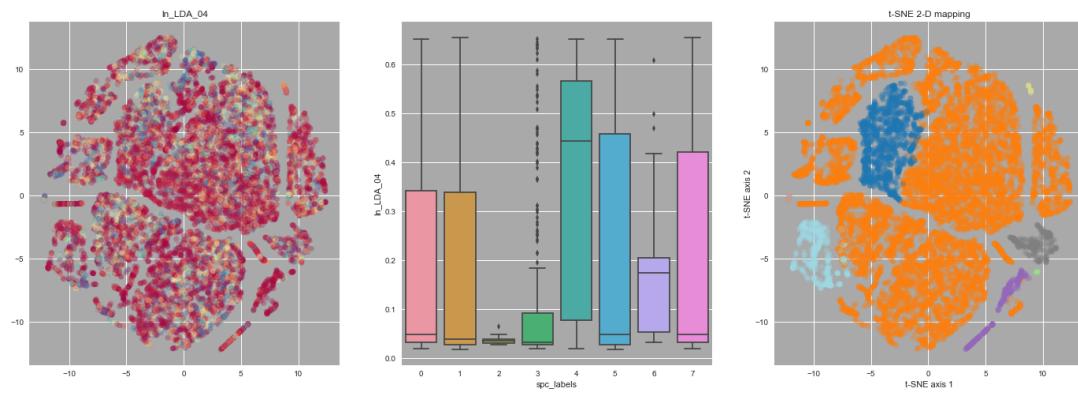
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe798cbe940>
```



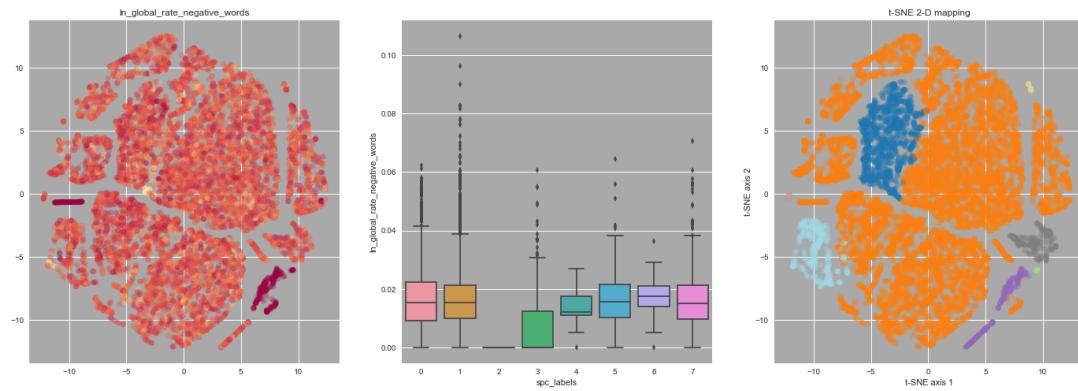
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe79937c160>
```



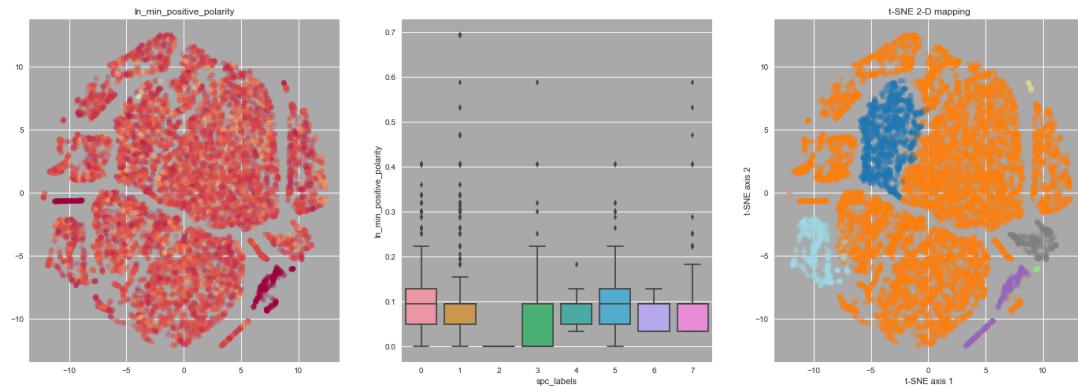
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5b17668>
```



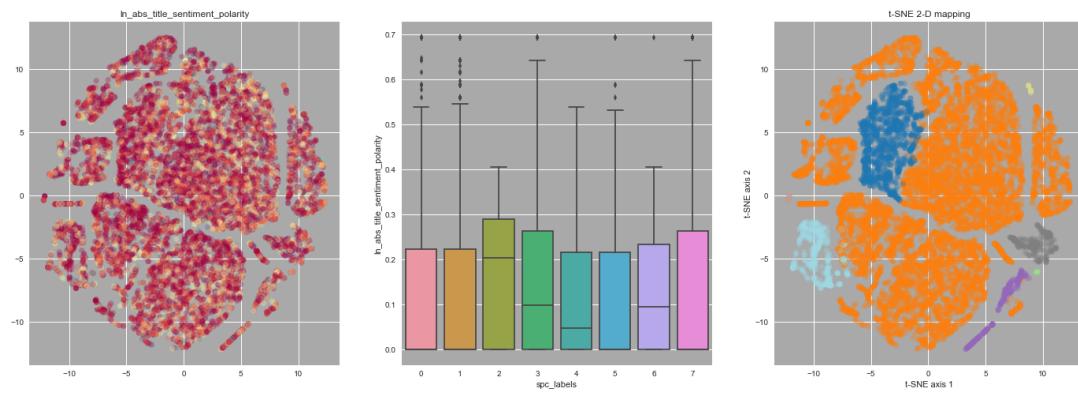
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7991f1e48>
```



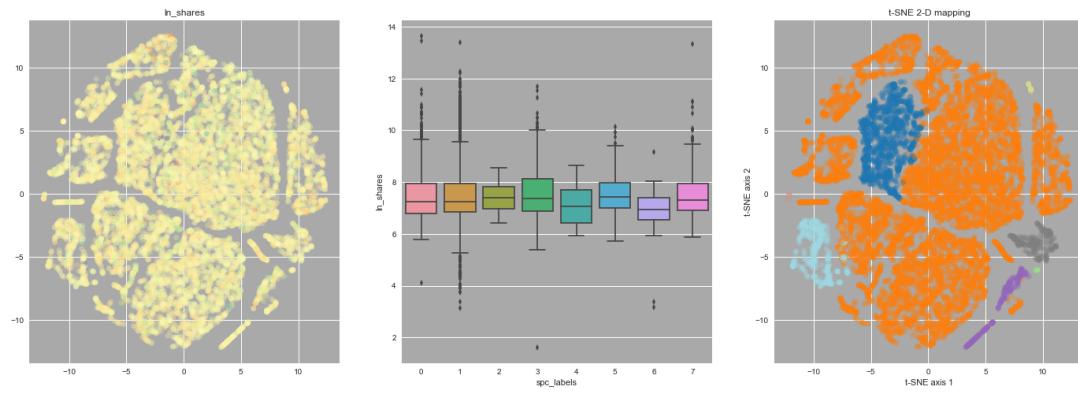
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b55b7898>
```



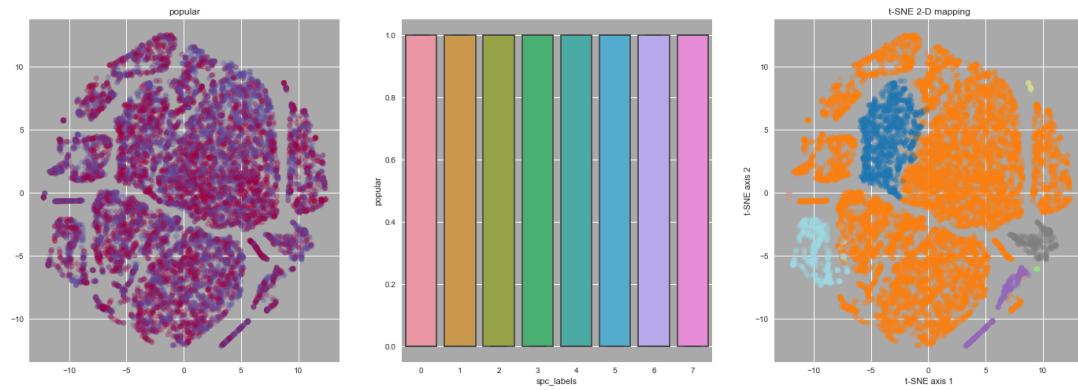
```
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b67b9f60>
```



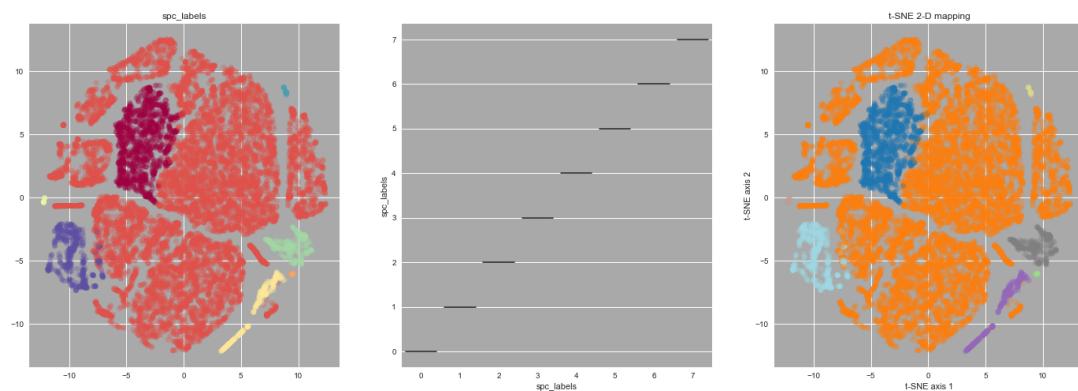
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5a84550>



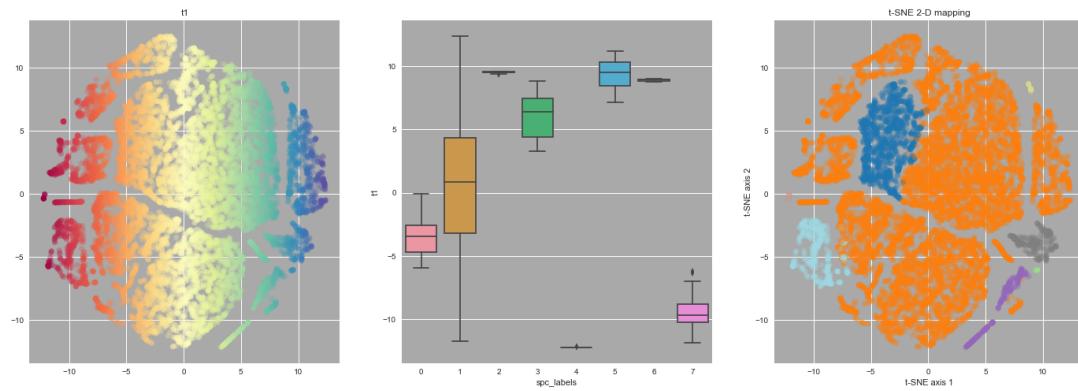
Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe798914588>



Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b60a7278>



Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe7b5930b38>



Out[74]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe798bdcc18>

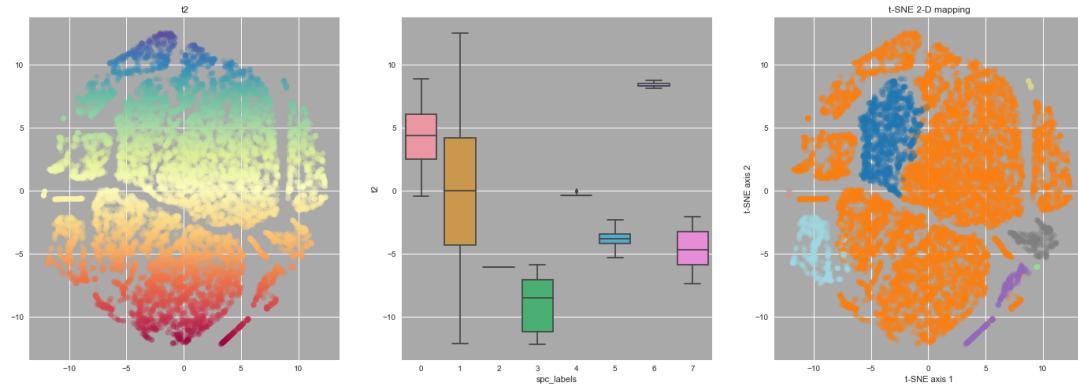


Table of Contents

end of file

In []: