Evaluating Community Building Effectiveness of Transportation Investments

Webinar II: Social Media Data Download and Analysis for Transportation Projects Part 1: Presentation

Presented by
Samiul Hasan, Assistant Professor
Naveen Eluru, Professor
Jiechao Zhang, PhD Student

Civil, Environmental, and Construction Engineering
University of Central Florida

Outline

- Introduction
- Data Collection
- Sentiment Analysis
- Demonstration
- Topic Analysis
- Final Recommendation

Objectives

- Understand community building effectiveness from an alternative data source.
- Collect social media data from Twitter by searching relevant keywords and user accounts
- Perform sentiment analysis to measure public opinion about several ongoing projects in Central Florida
  - Perform subjectivity analysis to determine the level of community opinions of transportation projects
  - Perform polarity analysis to determine the degree of community likeability of transportation projects
- Perform topic analysis to understand specific public perspectives of different transportation projects

Social Media Data Collection

- Source: Twitter
- Data Collection Area: Central Florida Region
- Collection Methods:
  - Using Twitter Search API to search for
    - Keywords
    - User accounts
- **Keywords** are selected to reflect the public opinions and find out their activities around the transportation infrastructures (bus service, train service etc.) of Central Florida, i.e. Sunrail, Suntrail, Lynxbus, juicebike etc.
- **User names** are carefully selected which include some of the most active accounts of public transport and traffic information provider of Central Florida region, fl511_state, I4Ultimate etc.

Social Media Data Collection

- Write Twitter data collection query scripts
- Continue running the scripts throughout the data collection phase (until December of 2018)

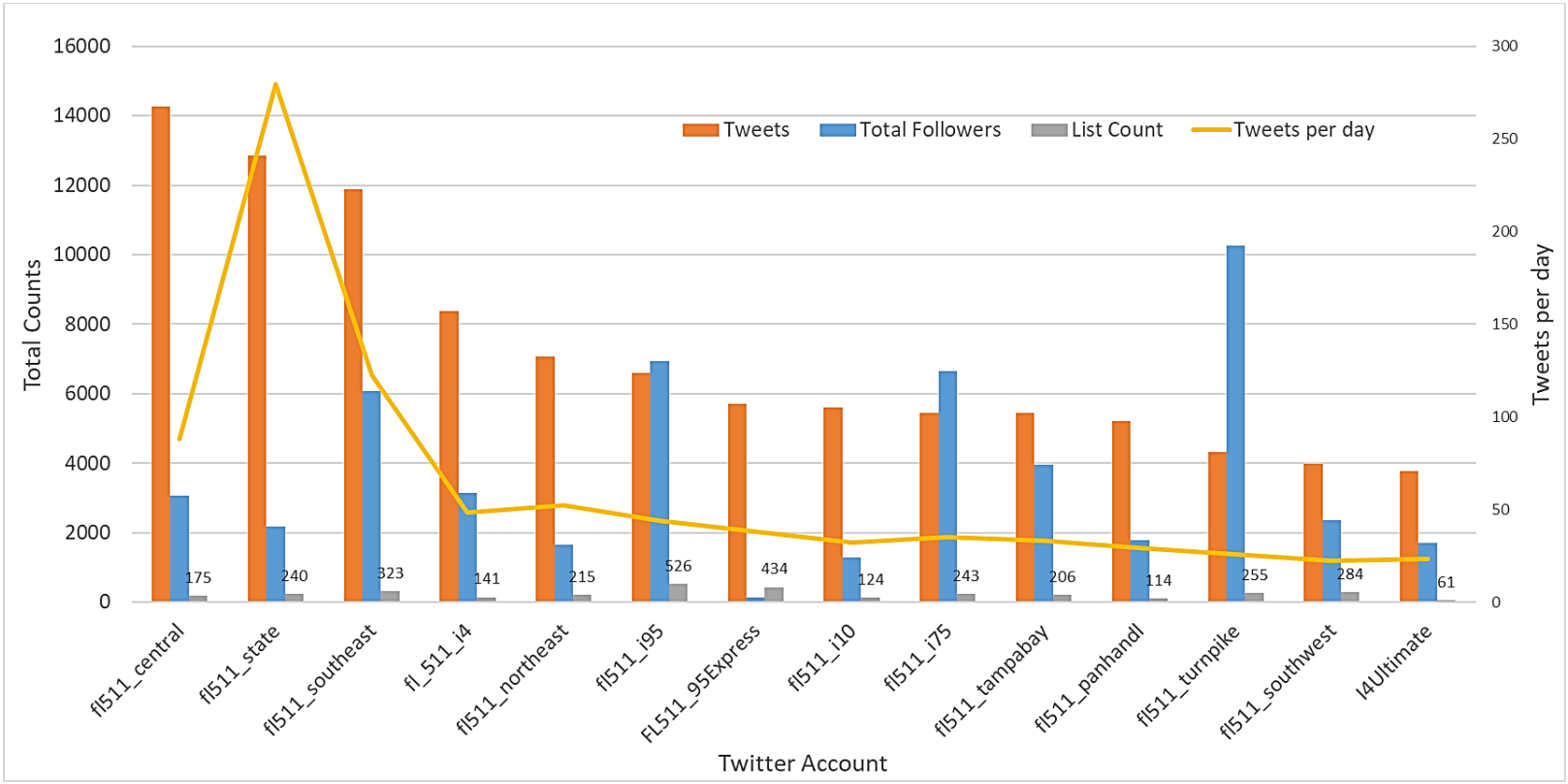
Data Description - User Accounts Search

User Name	Follower Count	Friends Count	Number of Posts Re-tweeted	Created at	Earliest Tweet	Latest Tweet	Total Tweets
321Transit	713	288	460	8/25/2010 15:58	8/25/2010 16:04	12/20/2018 15:49	6218
965traffic	1869	329	719	4/7/2011 13:54	9/29/2016 13:38	12/20/2018 23:38	19315
BikeWalkCFL	1621	1160	2137	8/29/2013 19:02	8/30/2013 17:58	12/20/2018 23:48	17769
fl_511_i4	4314	616	25	1/12/2012 14:48	2/1/2017 21:00	12/20/2018 23:56	61040
FL511_95Express	239	52	31	1/24/2017 19:51	2/24/2017 21:02	12/20/2018 23:53	58032
fl511_central	4316	646	63	10/6/2010 16:54	2/12/2017 0:25	12/20/2018 23:50	61296
fl511_i10	1996	245	23	10/6/2010 17:30	1/30/2017 12:52	12/20/2018 22:44	61249
fl511_i75	9604	523	81	10/6/2010 17:33	2/17/2017 17:31	12/20/2018 23:57	61323
fl511_i95	8656	1020	50	10/6/2010 17:37	2/24/2017 13:14	12/20/2018 23:53	61370
fl511_northeast	2128	196	25	10/7/2010 12:38	3/11/2017 7:24	12/20/2018 22:25	61315
fl511_panhandl	2628	207	82	1/12/2012 14:20	1/28/2017 11:56	12/20/2018 22:44	61289
FL511_SOUTHEAST	7200	367	104	5/10/2017 1:42	4/13/2017 12:58	12/20/2018 23:57	61276
fl511_southwest	3292	71	249	10/6/2010 17:15	1/20/2017 10:41	12/20/2018 21:12	61244
fl511_state	4191	177	14	10/7/2010 12:57	4/29/2017 19:18	12/20/2018 23:57	61192
fl511_tampabay	5086	107	71	10/6/2010 17:01	2/11/2017 17:56	12/20/2018 23:56	61334
fl511_turnpike	13060	360	268	10/6/2010 17:23	2/4/2017 19:34	12/20/2018 23:48	61325
I4Ultimate	2319	149	728	11/25/2014 17:19	1/17/2017 14:20	12/20/2018 1:59	57893
juicebikes	1286	253	151	3/23/2009 22:59	3/19/2011 21:53	12/20/2018 12:30	1360
lakexpress	117	109	8	8/13/2009 20:37	9/29/2010 18:45	10/26/2017 17:45	378
lynxbusorlando	5600	299	2243	6/4/2009 19:39	4/10/2013 13:46	12/20/2018 15:00	19321
RideSunRail	15995	595	1885	5/7/2012 20:50	5/10/2014 15:59	12/20/2018 21:19	16171
SunRailRider	825	27	249	3/24/2011 13:54	4/4/2011 22:41	8/29/2014 11:12	2580
SunTranTDP2017	15	97	5	11/9/2016 15:20	11/9/2016 17:44	6/13/2017 15:08	145
WazeTrafficOrl	129	0	16	11/3/2014 19:32	9/8/2016 6:07	2/24/2017 1:59	58392

Data Description – Keyword Search

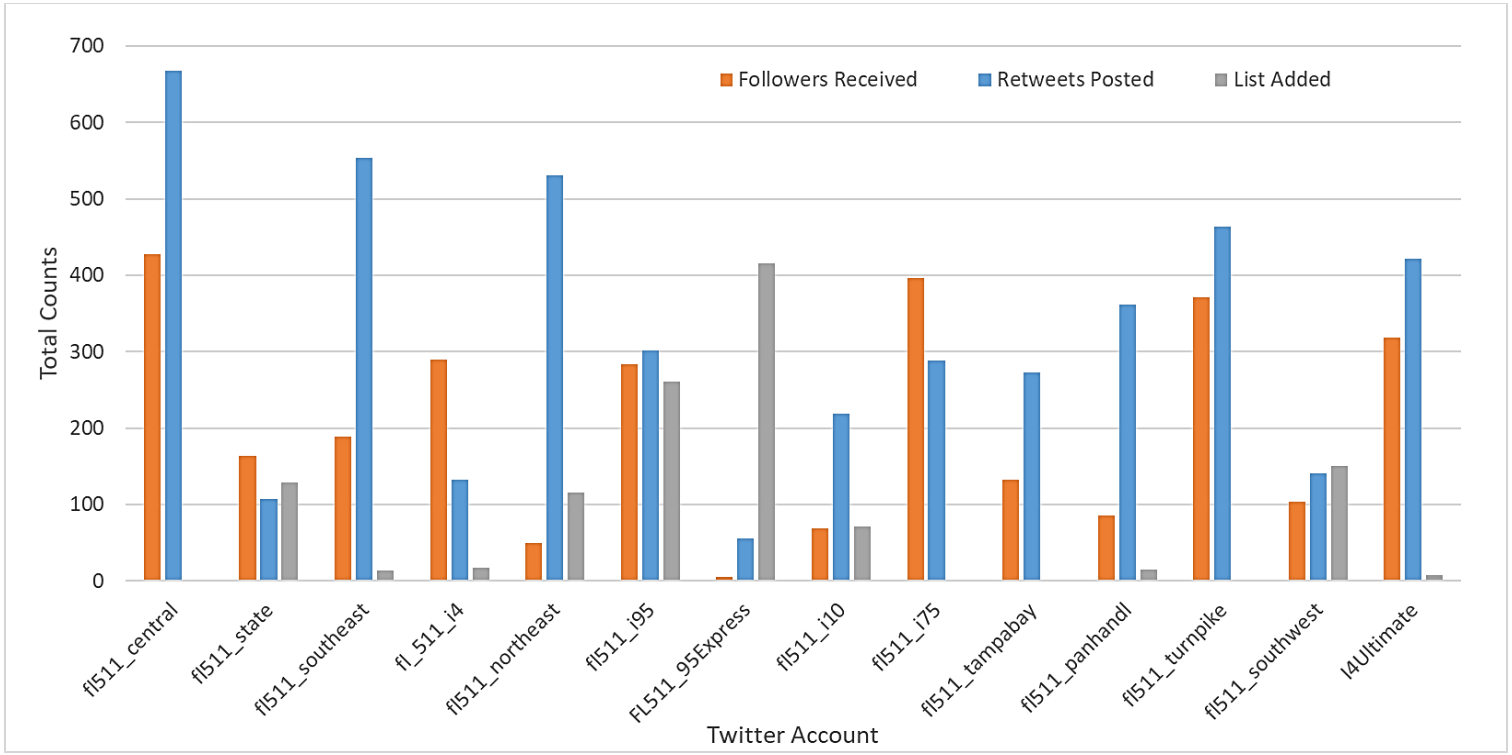
Keyword	Number of tweets	Number of geotagged tweets	Earliest Tweet	Latest Tweet
florida bus	10059	82	4/6/2017 16:40	12/20/2018 23:56
florida crime	40338	19	4/6/2017 16:40	12/20/2018 23:52
florida sidewalk	601	19	4/6/2017 16:40	12/20/2018 15:33
florida walking	29712	160	4/6/2017 16:40	12/20/2018 23:51
I4 Construction	1190	0	2/15/2017 11:56	12/20/2018 17:54
I4 Crash	7659	1	2/15/2017 11:18	12/20/2018 21:51
I4 Ultimate	215	0	2/15/2017 11:18	12/20/2018 19:29
Juice Bike	1963	21	2/15/2017 11:18	12/20/2018 23:24
juicebike	4	0	2/15/2017 11:18	12/20/2018 23:55
lakexpress	33	0	2/15/2017 11:18	11/1/2018 23:55
lynx bus	1399	23	2/15/2017 11:18	12/20/2018 23:56
Lynx Vanpool	3	0	2/15/2017 11:18	10/11/2018 23:55
Space Coast Area Transit	84	4	2/15/2017 11:18	12/18/2018 23:57
sunrail	4378	82	2/15/2017 11:18	12/20/2018 23:42
Sunshine Skyway	5472	260	2/15/2017 11:18	12/20/2018 21:47
suntrail	133	1	2/15/2017 11:18	12/16/2018 23:57
suntran ocala	32	2	2/15/2017 11:18	12/19/2018 23:57

Insights: Activity and influence metrics



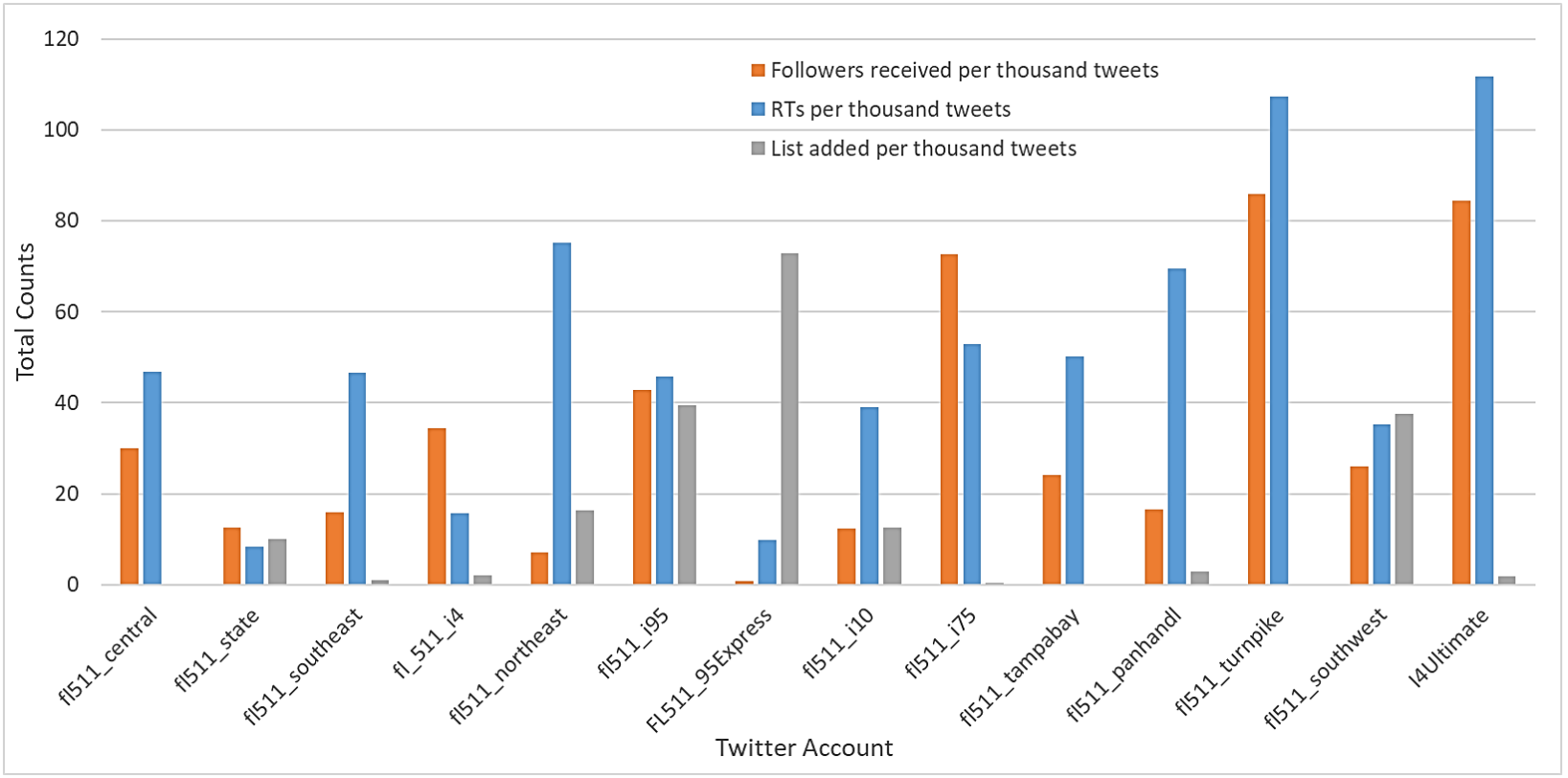
Data collection period: April 2017 to July 2017

Insights: Attention gained



Data collection period: April 2017 to July 2017

Insights: Attention gained per thousand tweets



Data collection period: April 2017 to July 2017

Sentiment Analysis - Introduction

- We applied 'TextBlob', which is a Python library for processing textual data, to analyze the sentiment information of tweets. Based on the tweets text, the projection will return a tuple of form (polarity, subjectivity).

  - Subjectivity: A float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective.
  - Polarity: A float within the range [-1.0, 1.0] where -1.0 is very negative and 1.0 is very positive.

Sentiment Analysis – Positive Examples

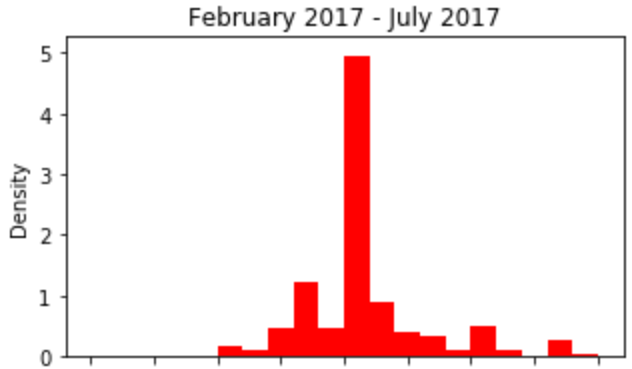
- patboyle60: “RT @PierreTristam: Loving our very first ride on #sunrail. We need more of these French-like car-busting commuter luxuries” (polarity: 0.475; subjectivity: 0.628)

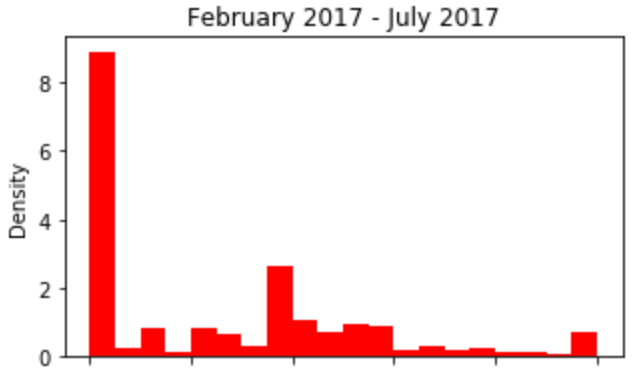
- CrankySlytherin: “Not surprised by Sunrail's numbers. It's still new to the area. Give it time it will grow. Also needs more hours and more destinations.”(polarity: 0.272; subjectivity: 0.589)

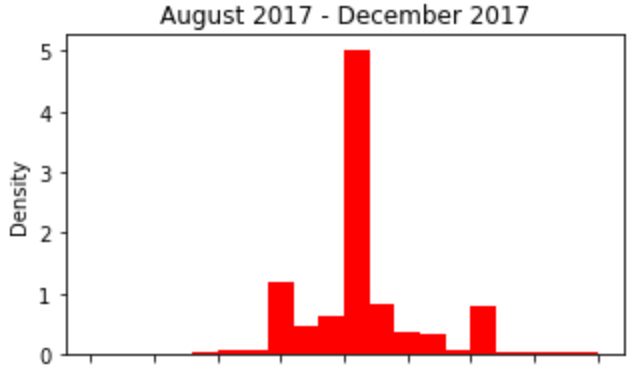
Sentiment Analysis – Negative Examples

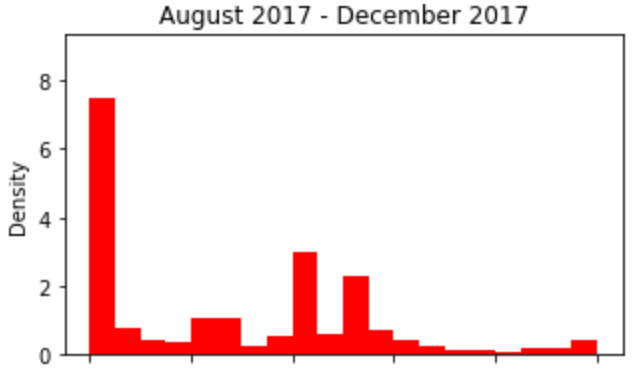
- grumpbump : "@RideSunRail @WESH @WFTV @Fox35News what's going on with Sunrail. No updates for frustrated passengers, on Friday afternoon, no less." (polarity: -0.308; subjectivity: 0.133)
- AllForTransit : “ALERT: #SunRail service is suspended between Altamonte Springs and Longwood Stations due to police activity on the tracks. Expect delays.”(polarity: -0.125; subjectivity: 0.589)

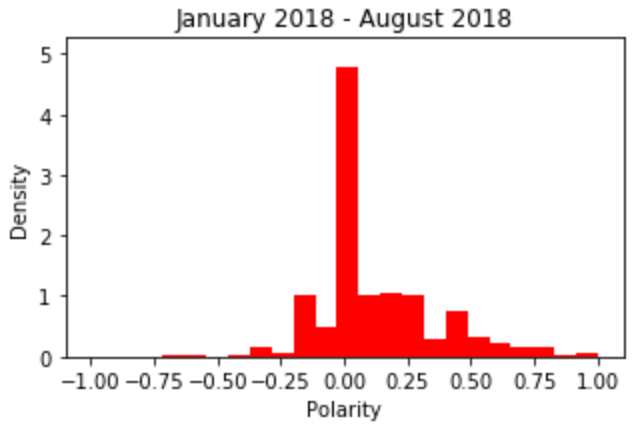
Sentiment Analysis Results - Sunrail

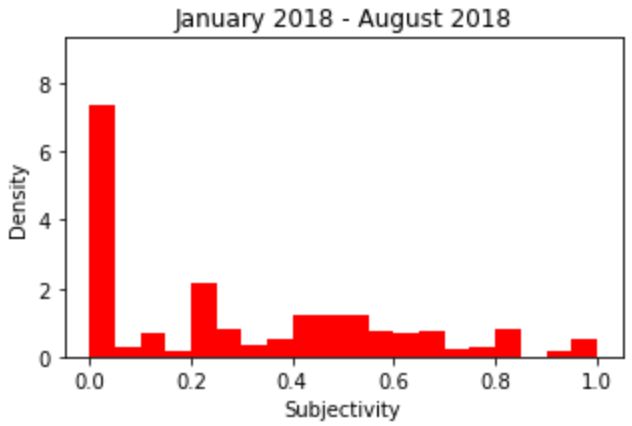




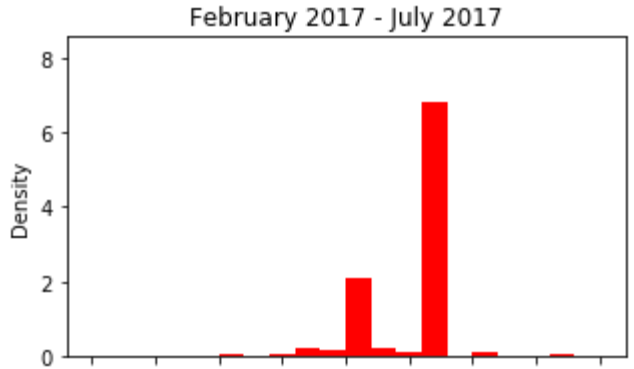


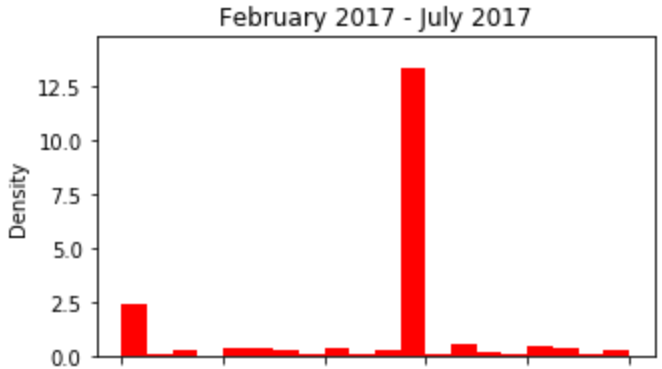


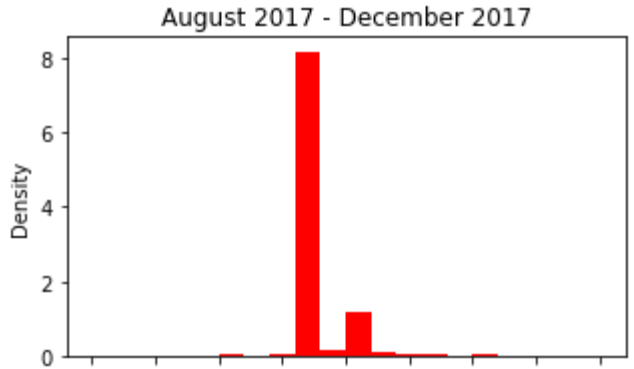


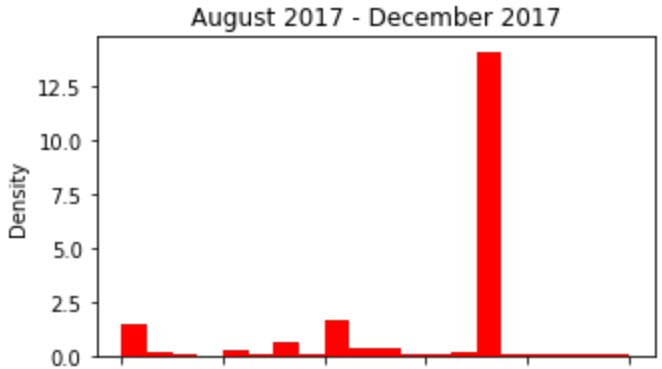


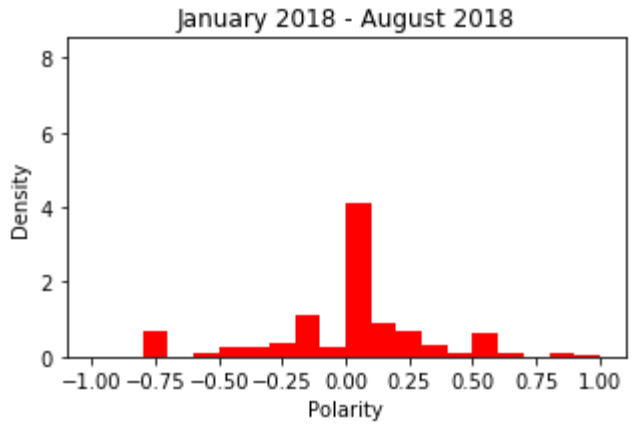
Sentiment Analysis Results – Florida Walking

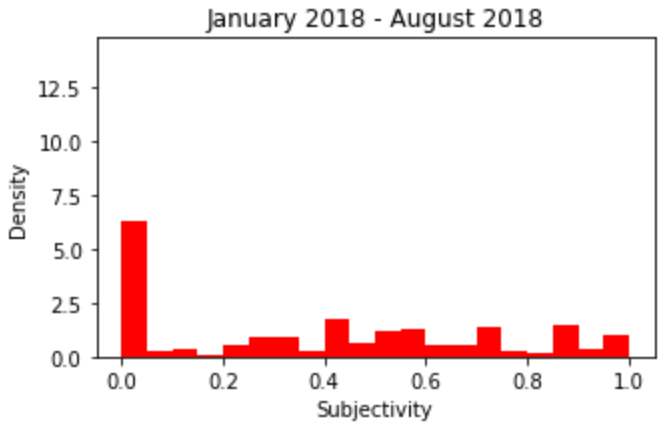




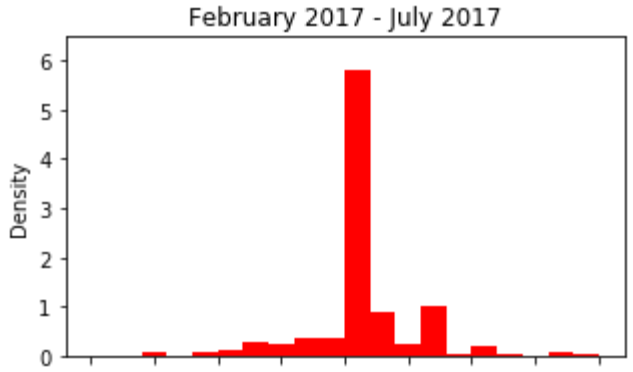


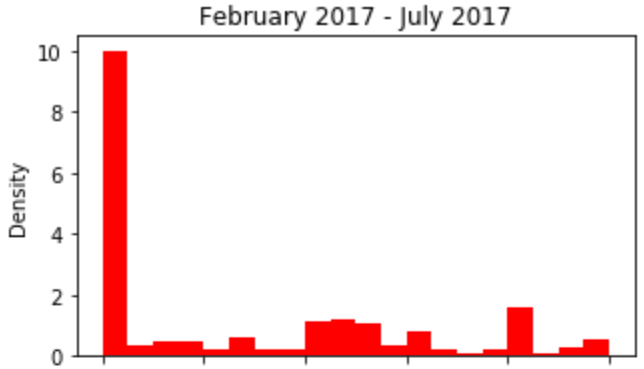


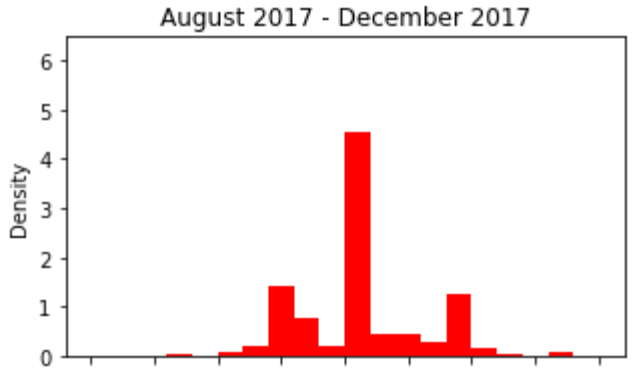


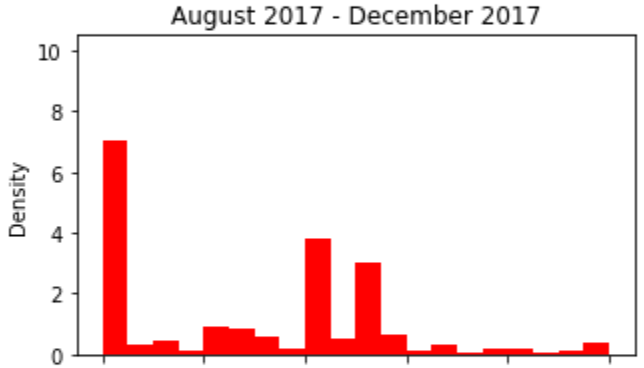


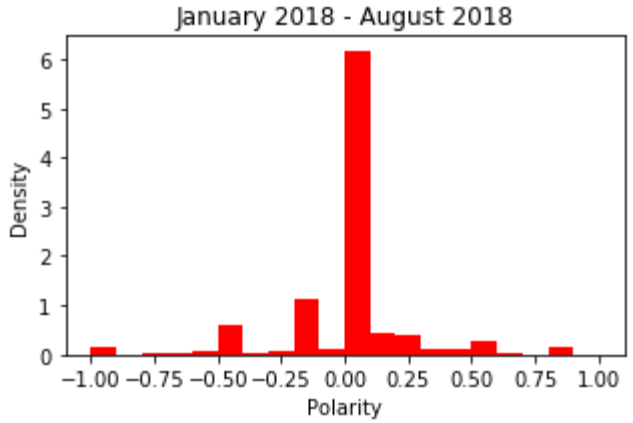
Sentiment Analysis Results – Florida Bus

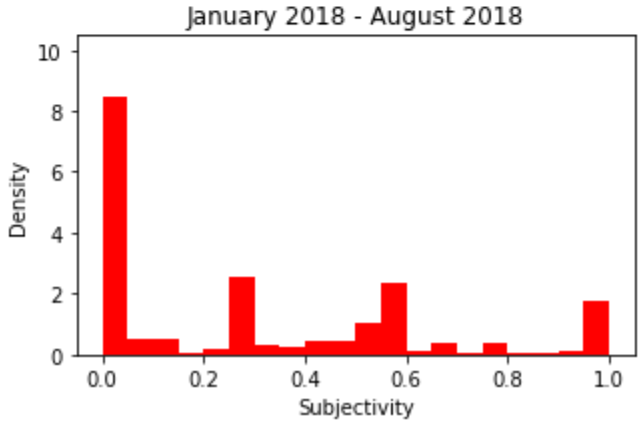




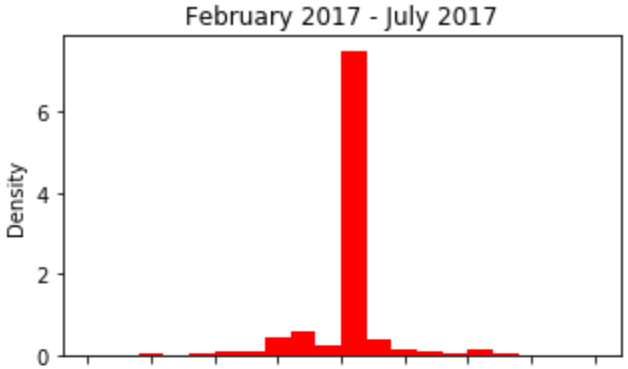


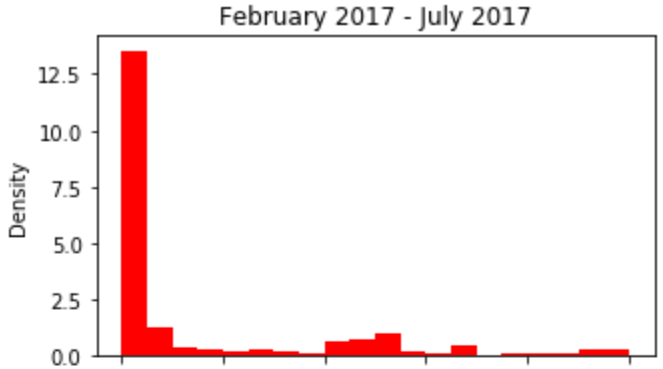


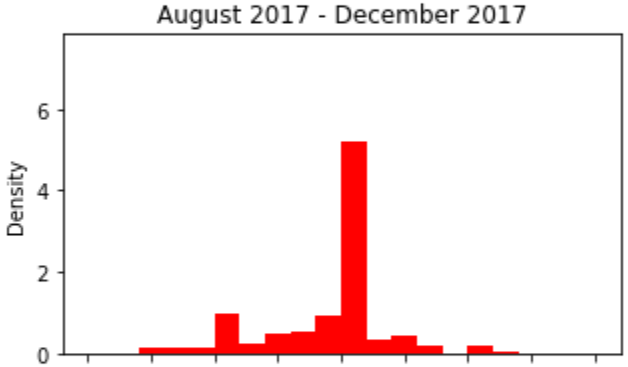


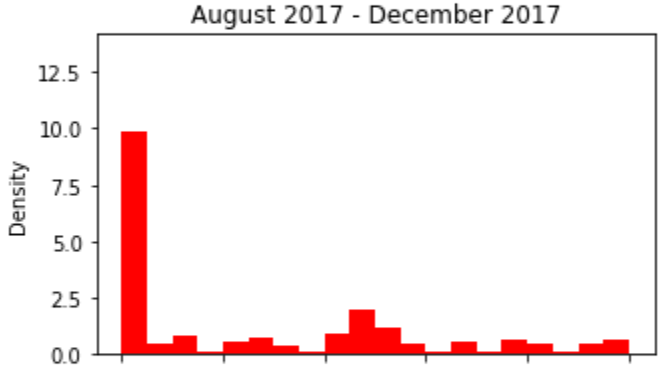


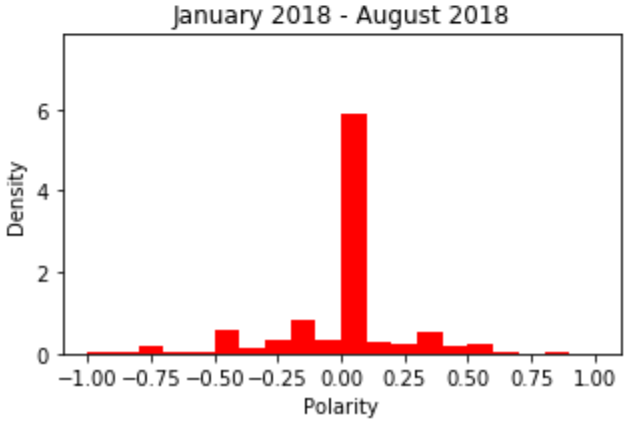
Sentiment Analysis Results – Florida Crime

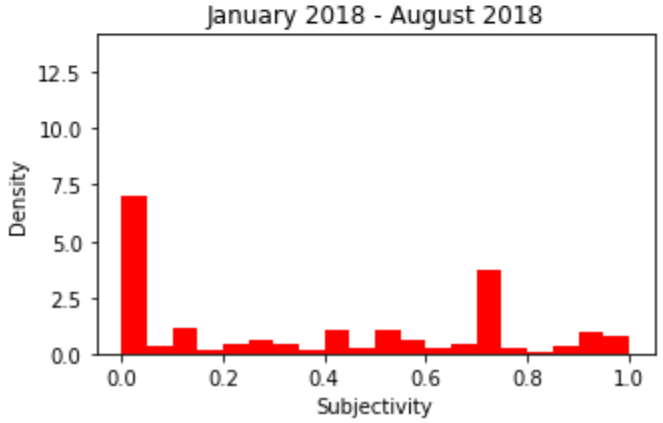




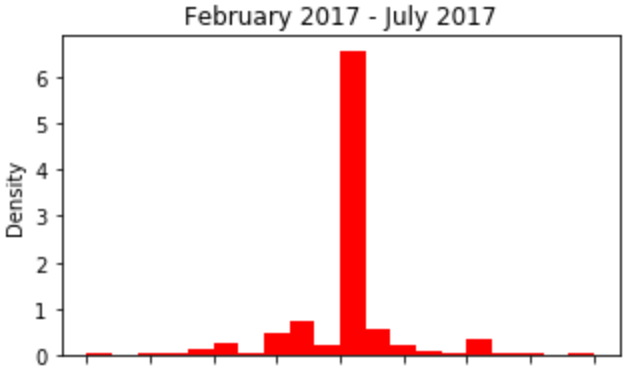


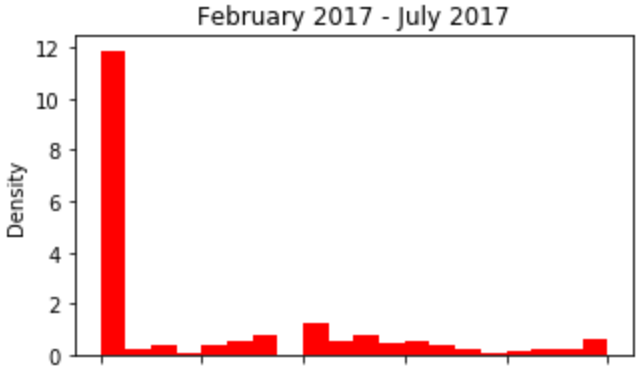


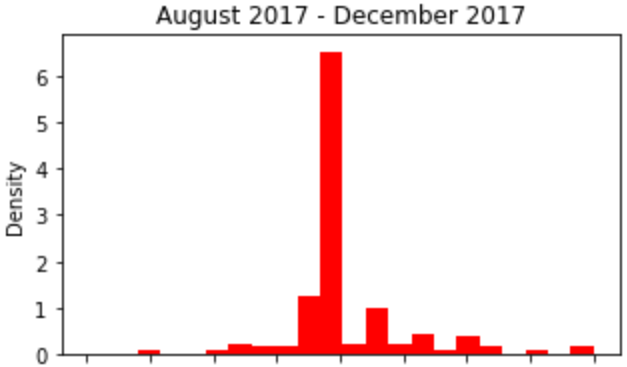


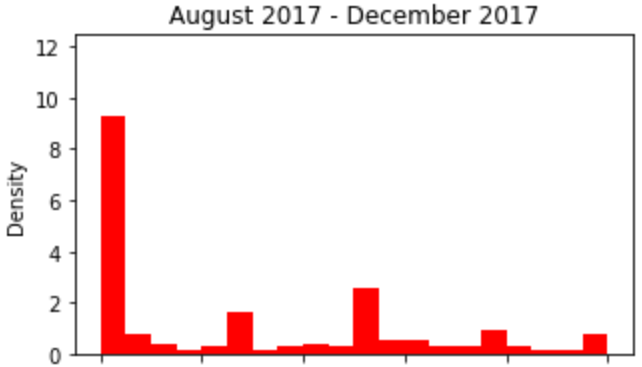


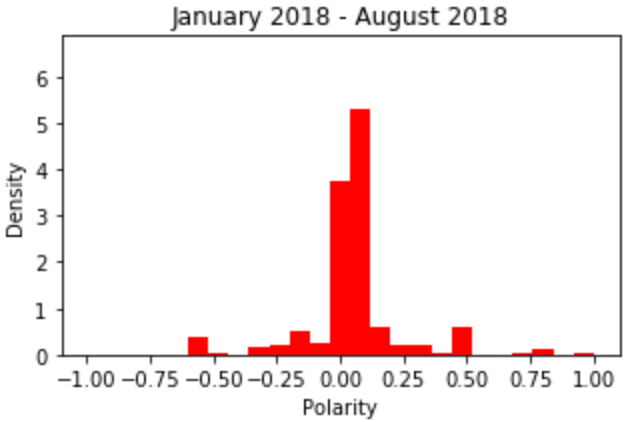
Sentiment Analysis Results – Florida Sidewalk

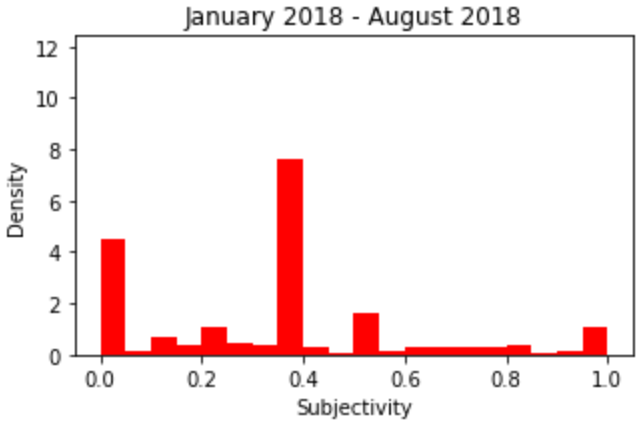




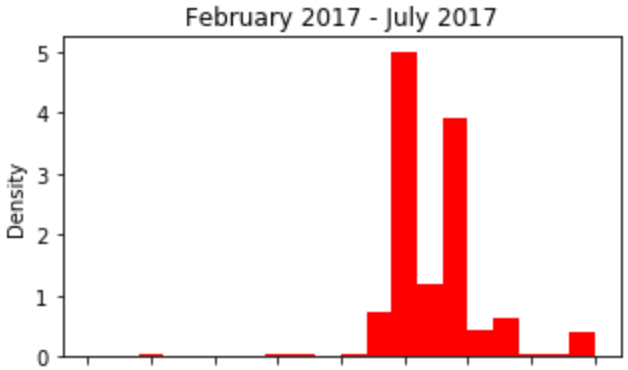


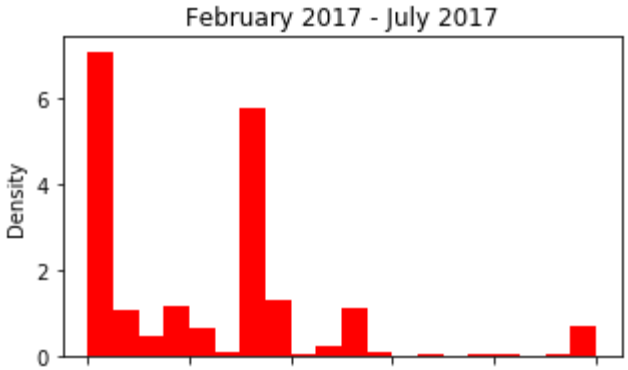


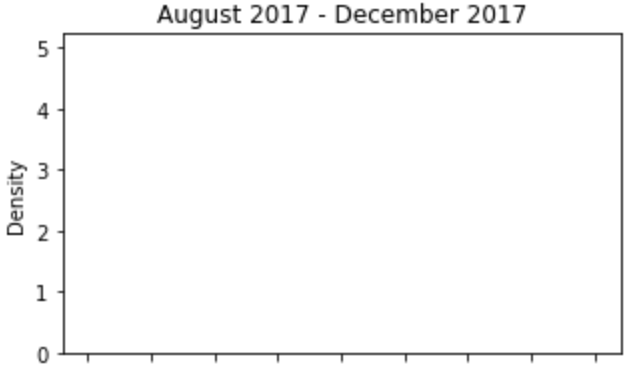


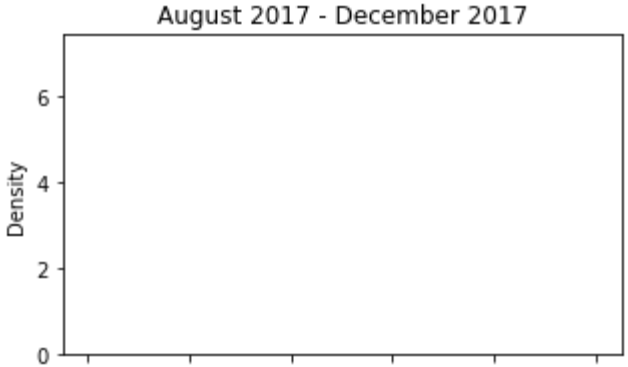


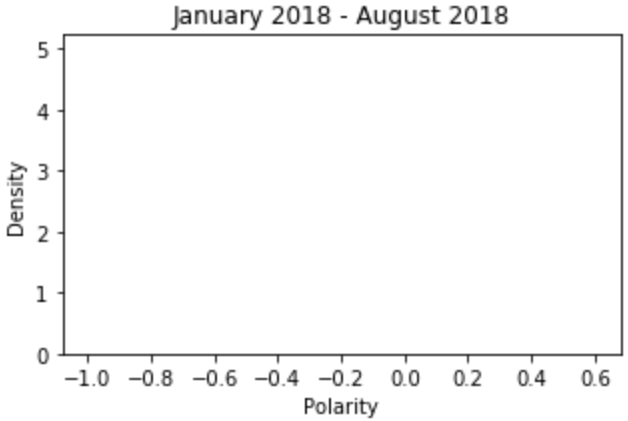
Sentiment Analysis Results – I4 Construction

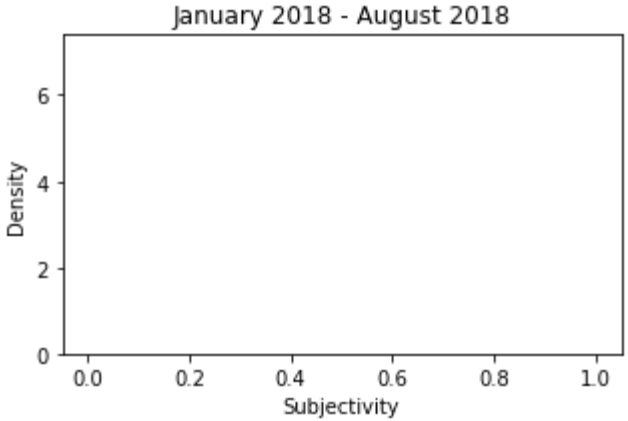




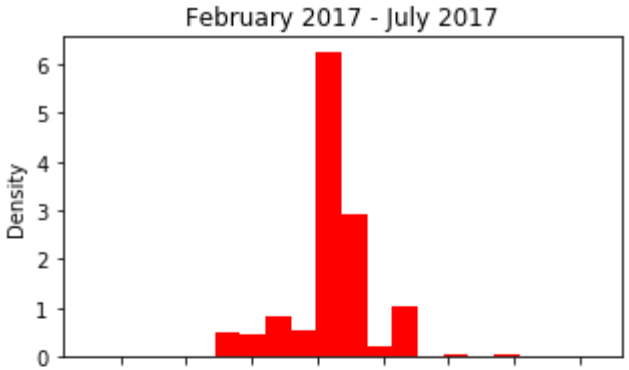


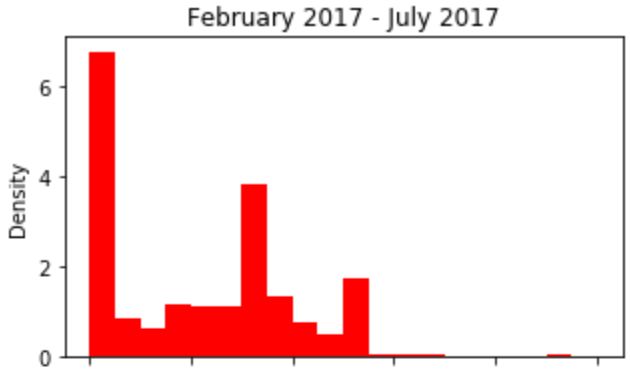


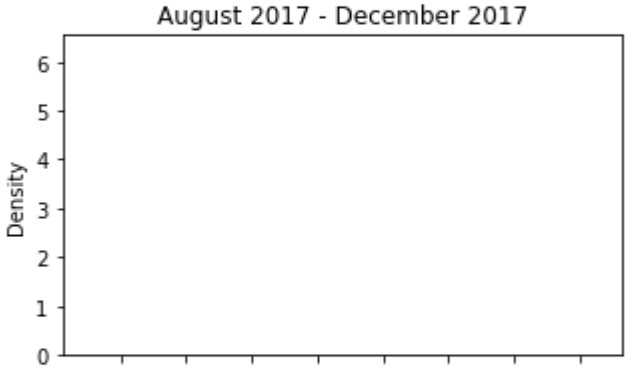


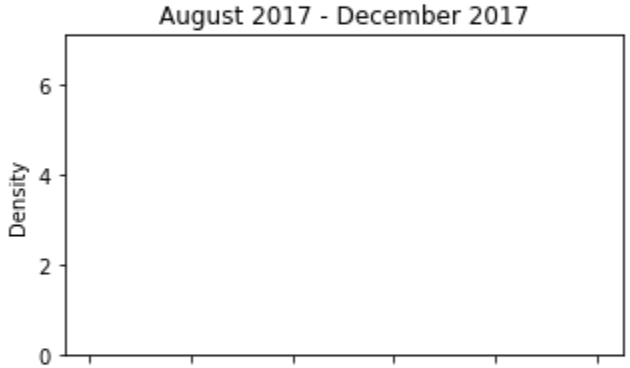


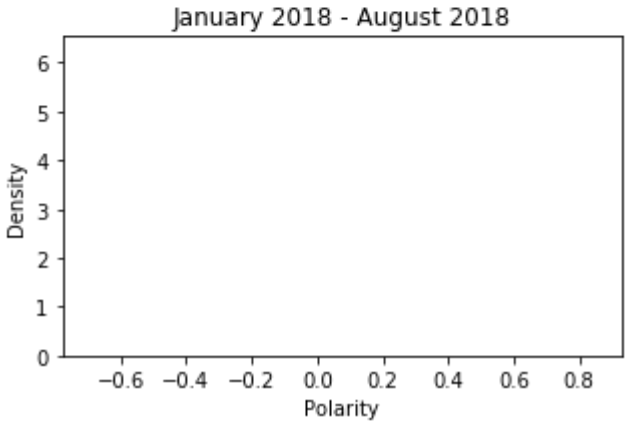
Sentiment Analysis Results – I4 Crash

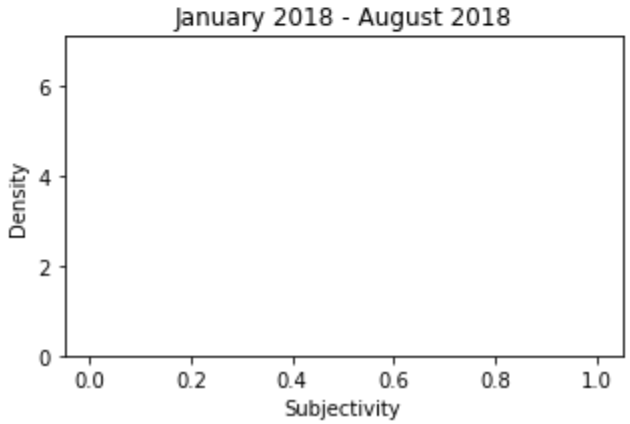




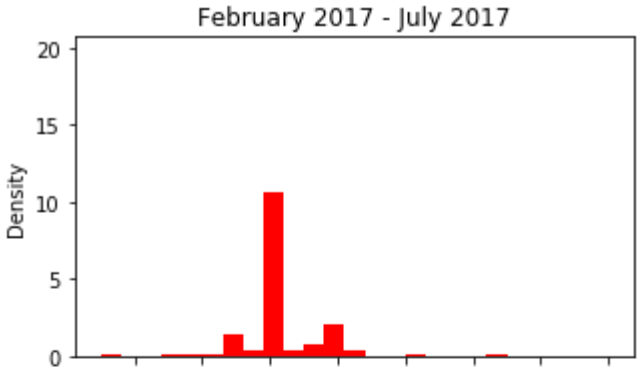


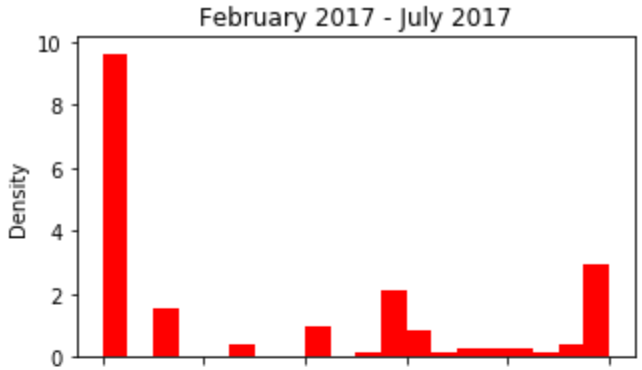


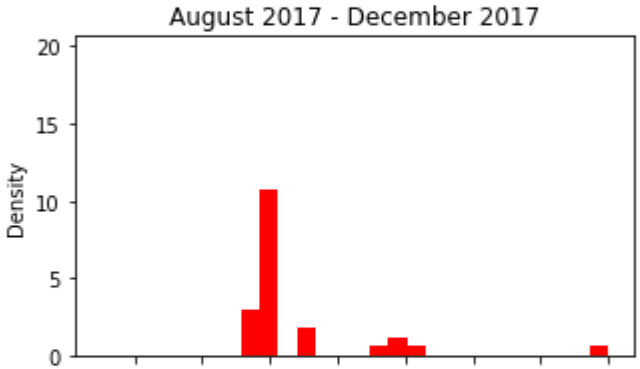


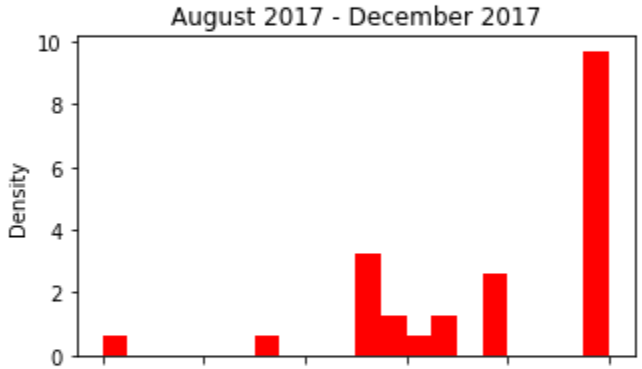


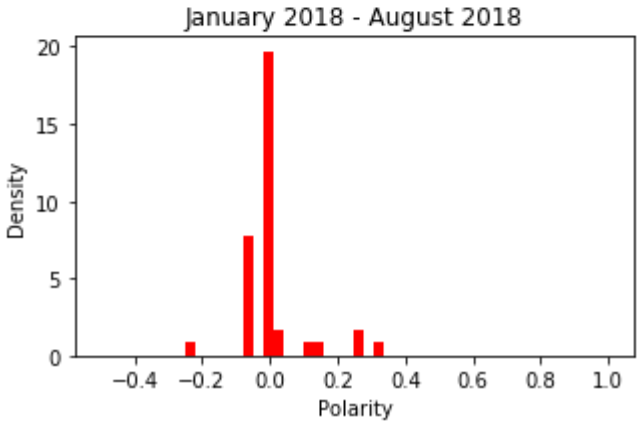
Sentiment Analysis Results – I4 Ultimate

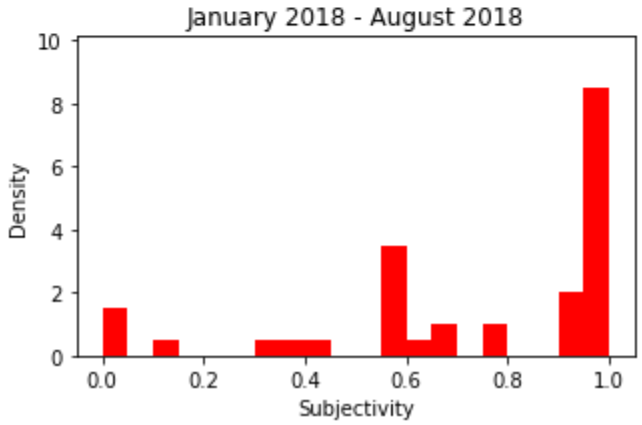




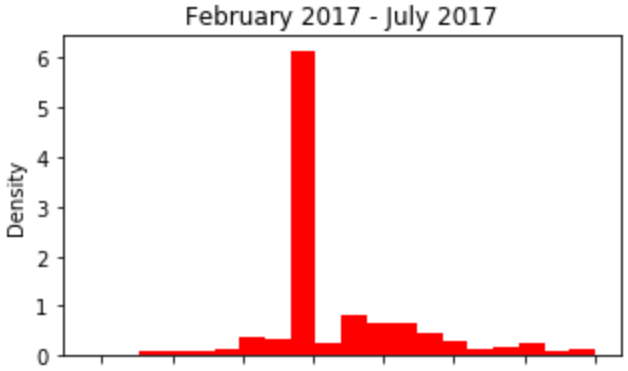


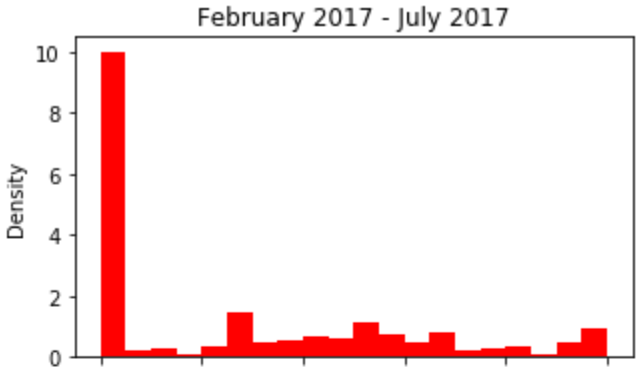


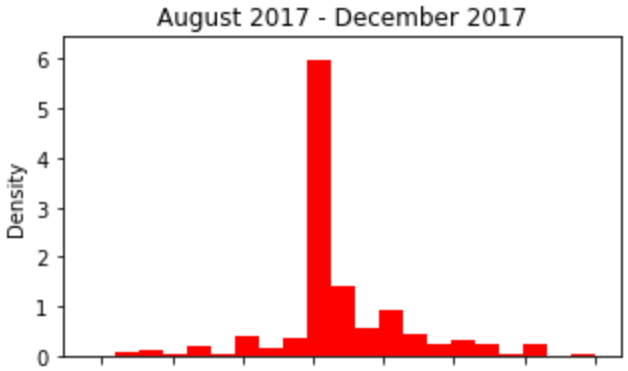


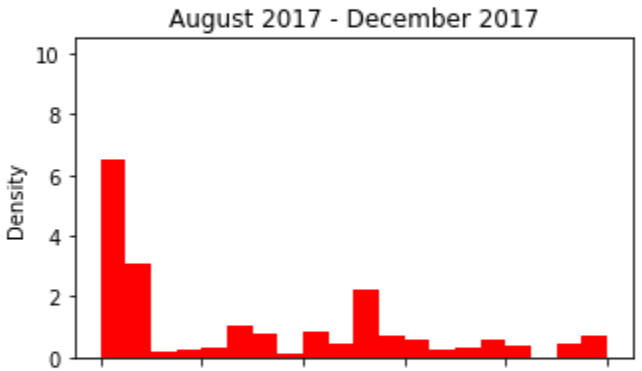


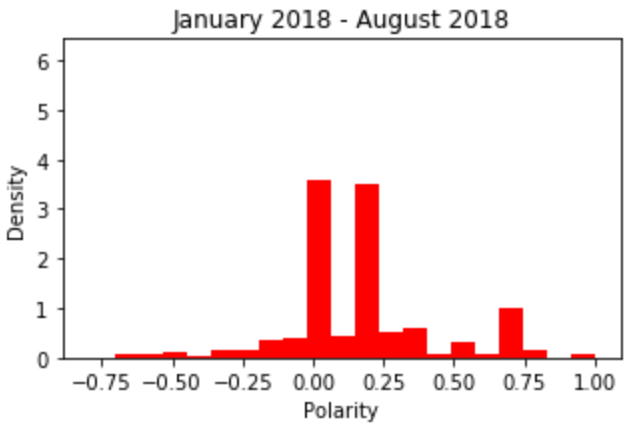
Sentiment Analysis Results – Juice Bike

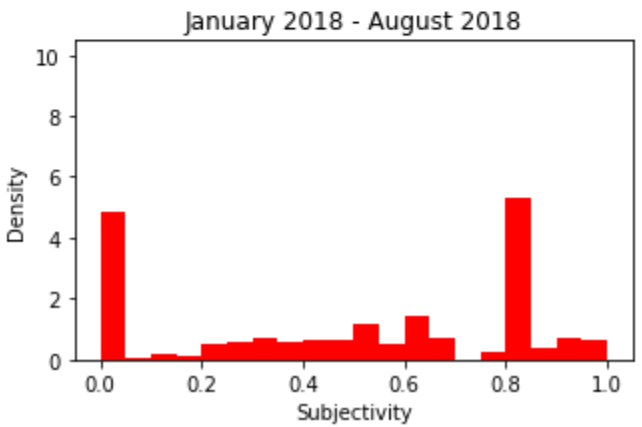




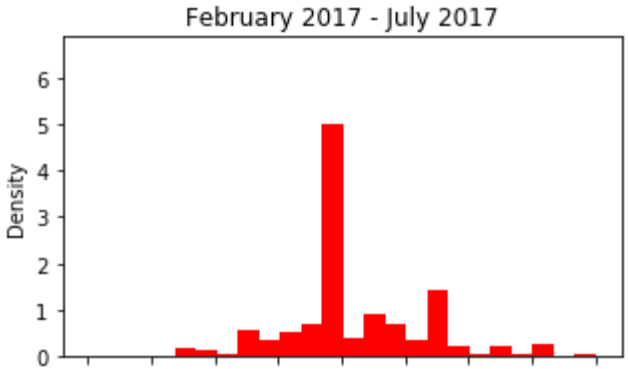


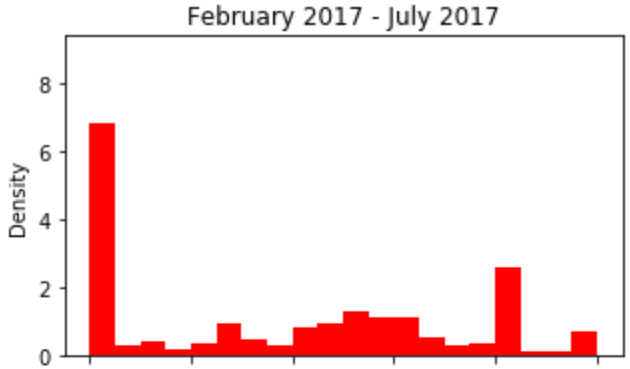


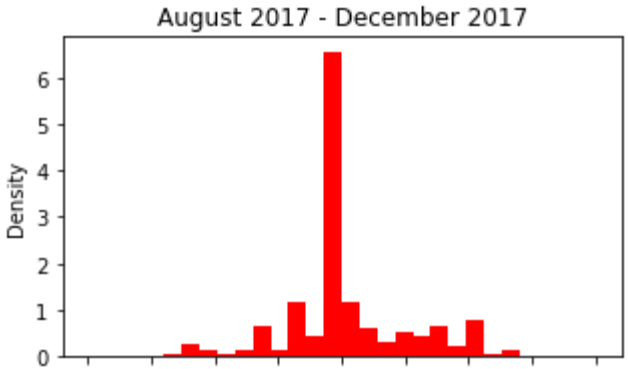


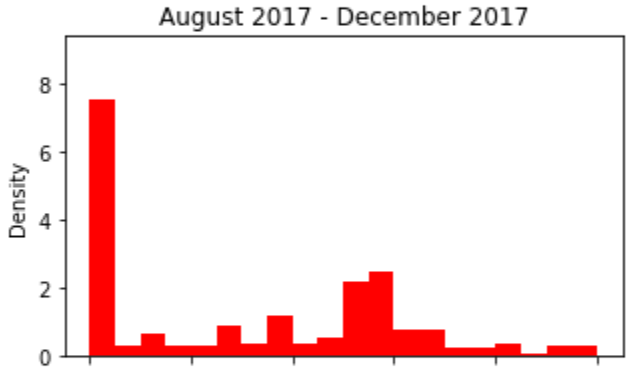


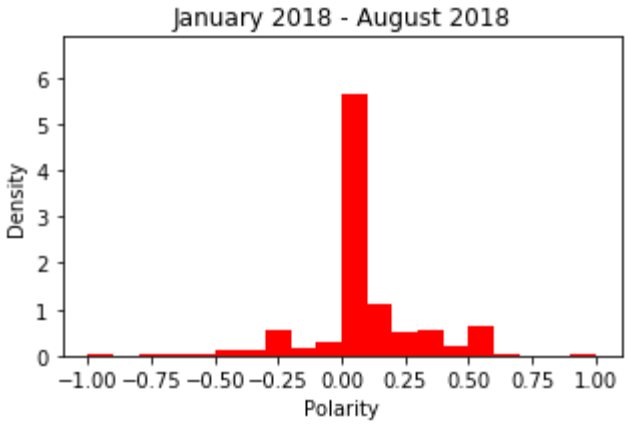
Sentiment Analysis Results – Lynx Bus

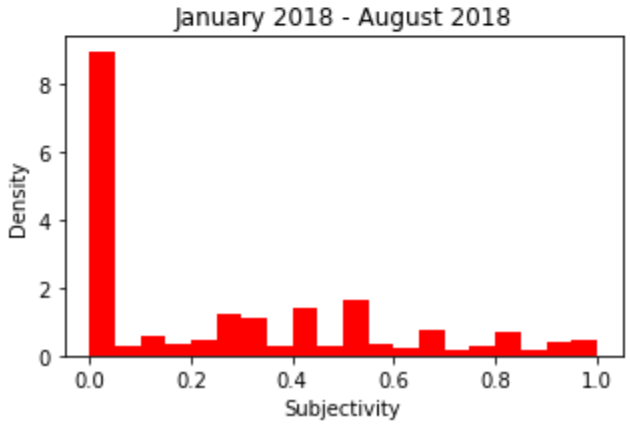




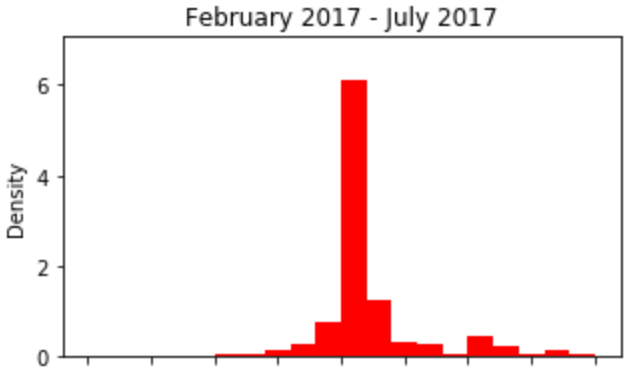


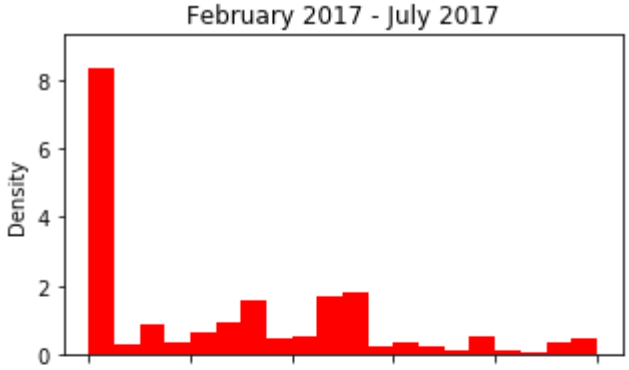


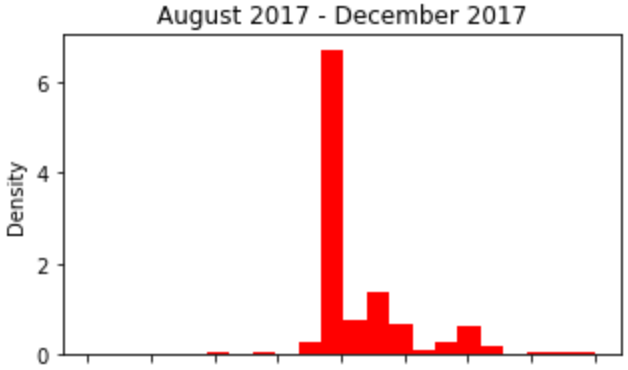


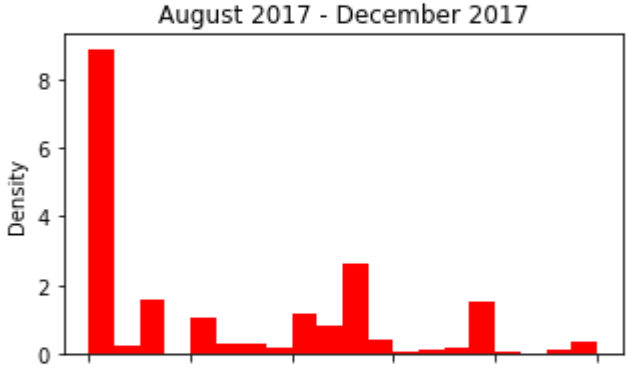


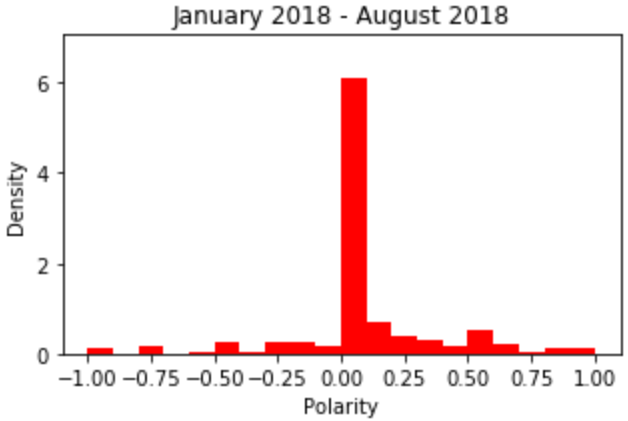
Sentiment Analysis Results – Sunshine Skyway

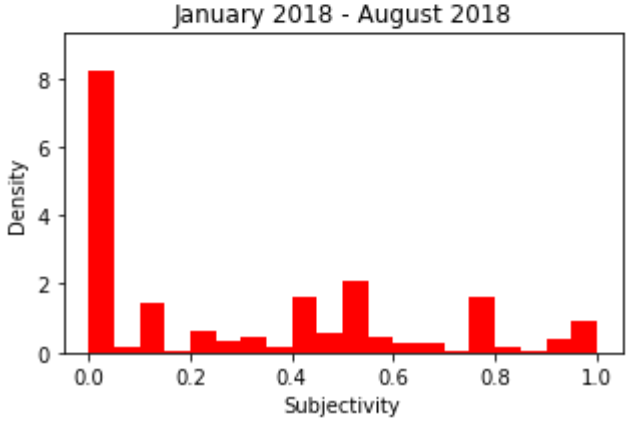




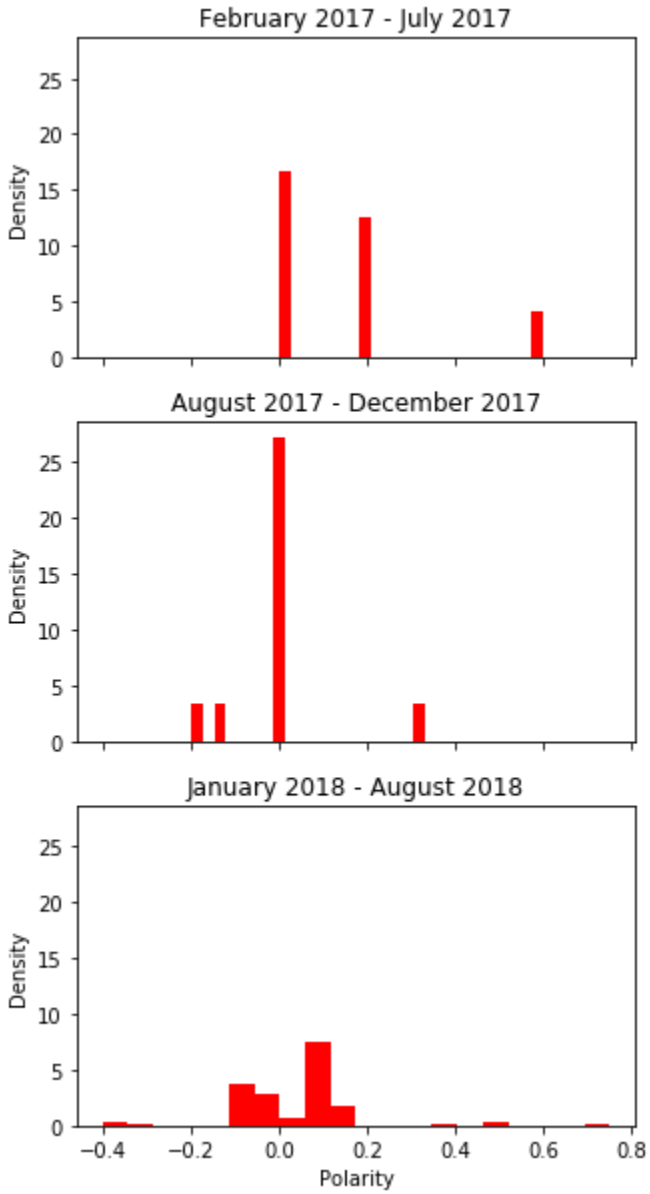


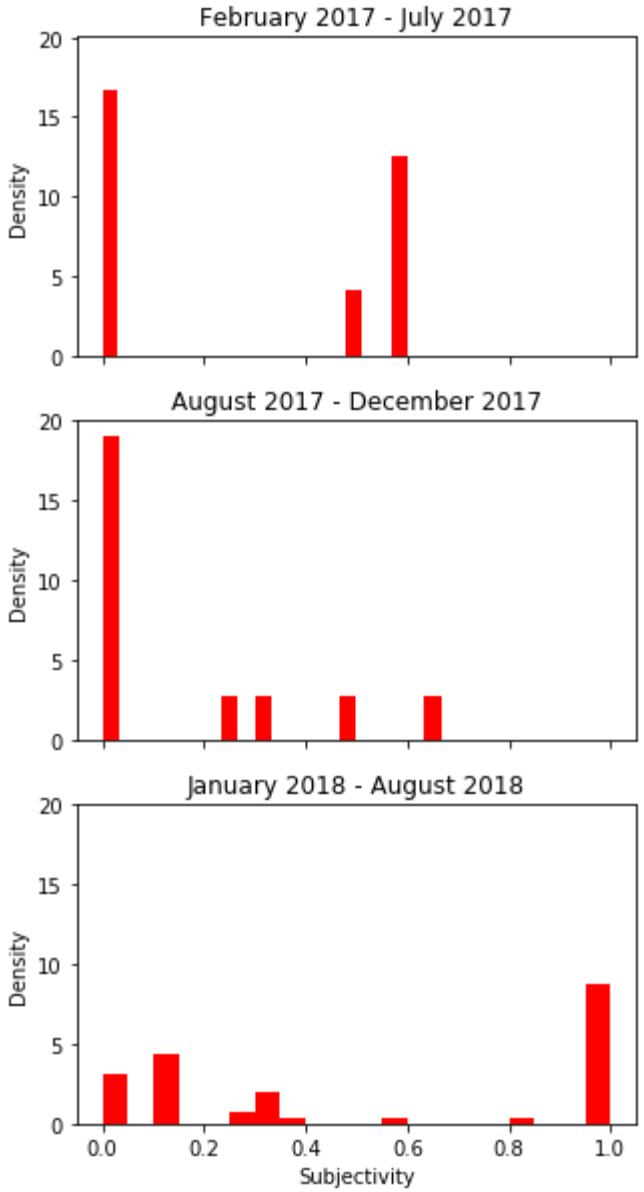




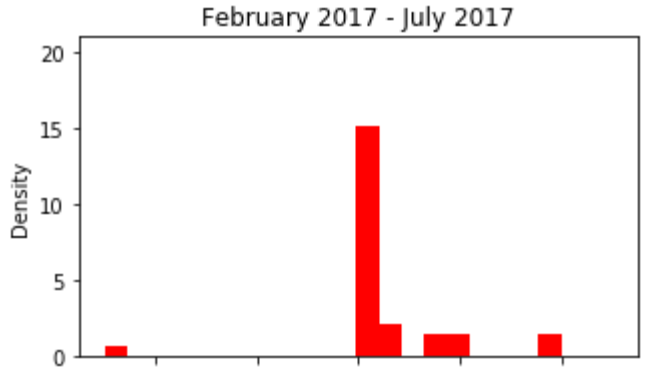


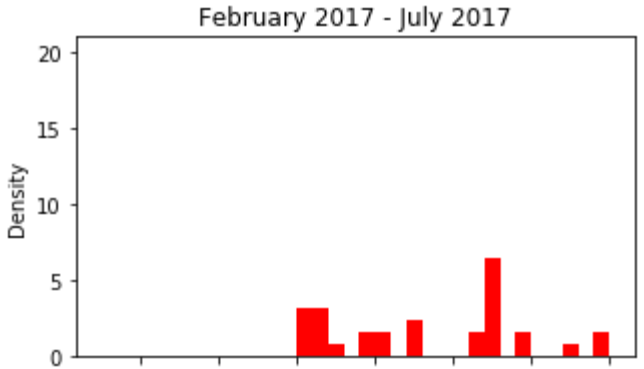
Sentiment Analysis Results – Suntrail

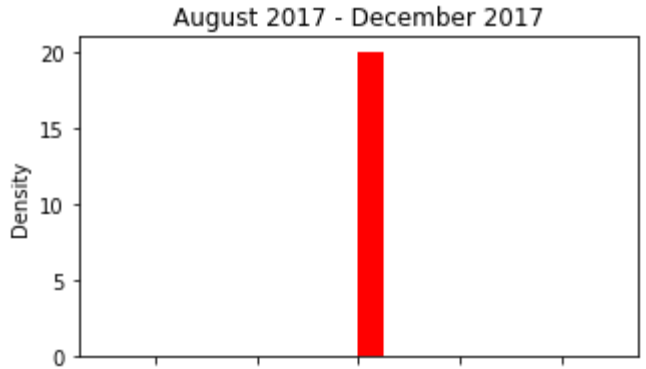


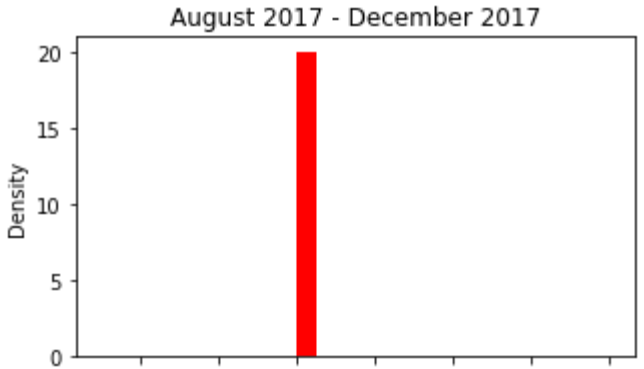


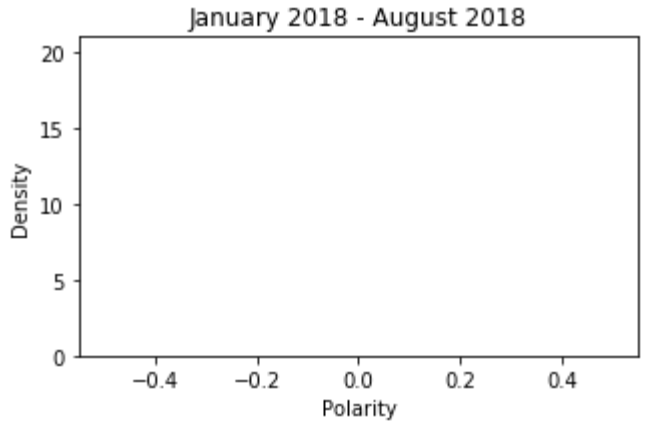
Sentiment Analysis Results – Suntran Ocala

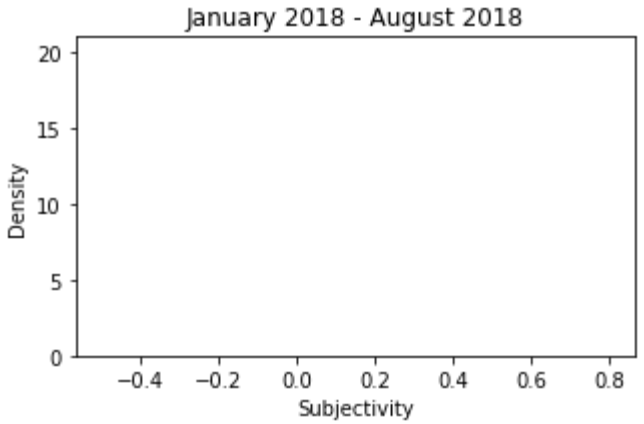










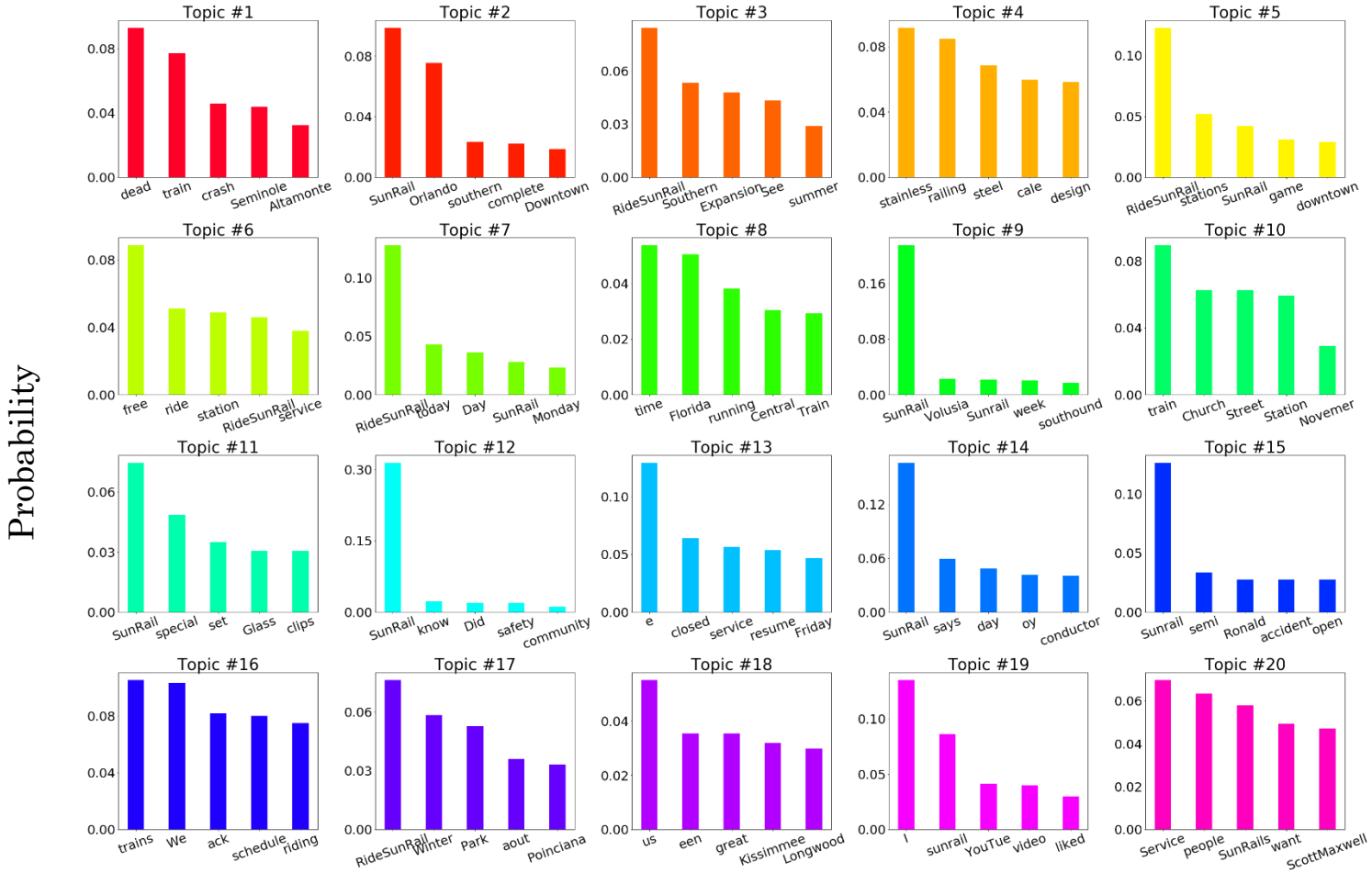


Demonstration

Topic Model - Introduction

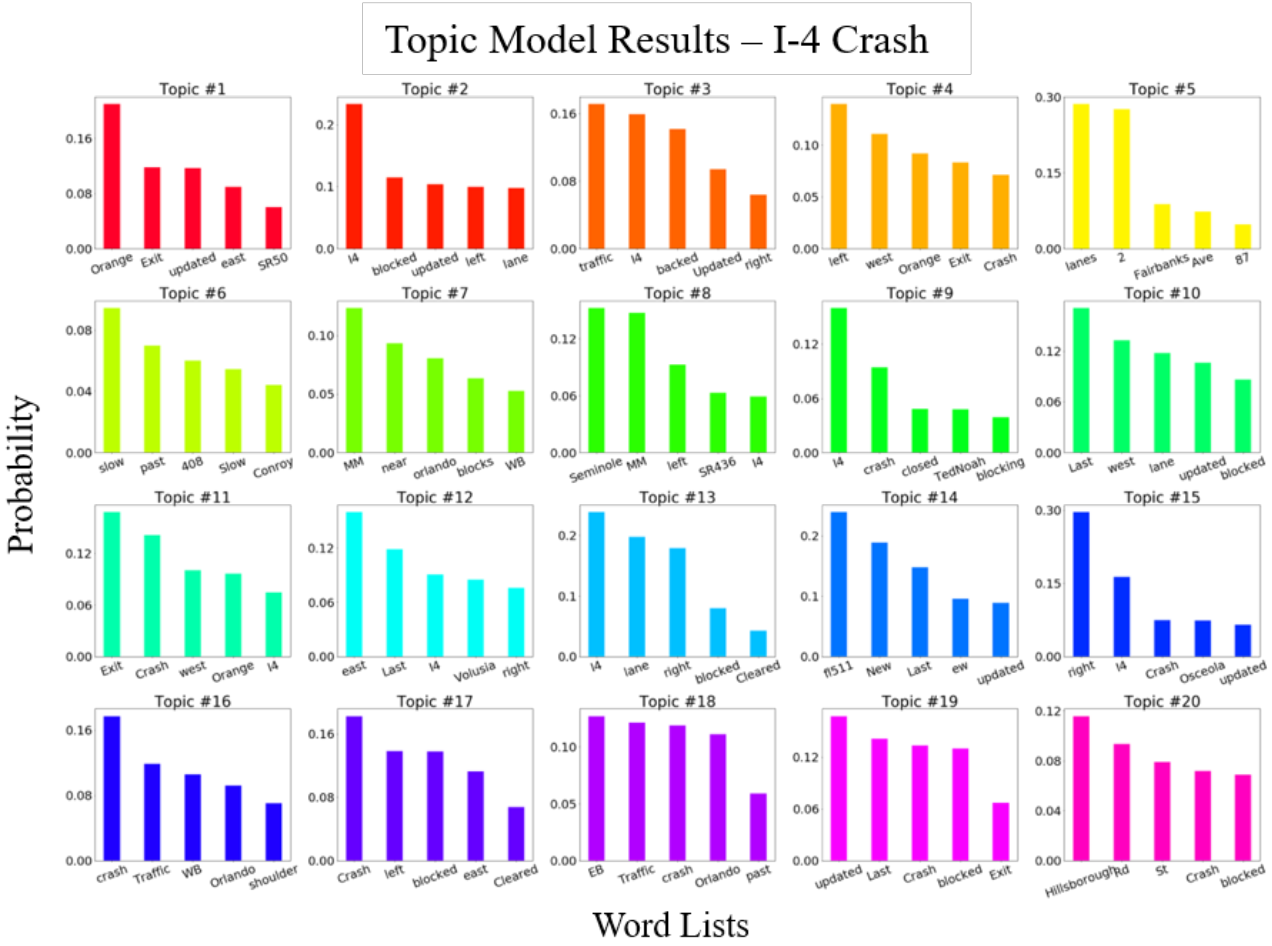
- Topic models are generally used for finding the topics in a large corpus of documents
- Each document is modeled as a mixture of topics
- Each topic is modeled as a distribution of words
  - where the entity “word” represents the basic unit of discrete data.

Topic Model Results – SunRail (subjective tweets)



Topics #1, #12, and #15 are about safety related issues. Topic #4 indicates conversations about Sunrail design. Topics #6, #13, and #20 indicate various aspects of SunRail services (e.g., free ride and closed service).

Topic Model Results – I-4 Crash (subjective tweets)



These topics indicate the local areas affected by crashes in I-4 and related traffic advisory services. Such information services are critical for travelers.

KEYWORDS WITHOUT SUFFICIENT DATA

Keywords	Number of Tweets
juicebike	4
lakexpress	33
lynx bus	1399
Lynx Vanpool	0
Space Coast Area Transit	84
suntrail	133
suntran ocala	32

Final Recommendation

- Useful indicators based on social media data can be constructed to understand community perception on transportation investments.
  - show the community building impacts of transportation projects.
- Data collected over some of the keywords lack sufficient relevant information for running a topic analysis.
- Additional data can be obtained by purchasing data from Twitter and modifying the keywords.
- Twitter data is free, but not free.
  - Customized analytics tool need to be created to harvest, analyze, and interpret data over a longer period to monitor community impacts of transportation investments.