**First Round framework for sentiment analysis**

**1. Create a plan on how to collect social media data concerning customer experiences (AT&T will not provide any data for this competition)**

a. Collecting all social media reviews about AT&T retail stores in Dallas.Taking AT&T store locator as reference and after looking up the social media website likes google reviews ,twitter  
And yelp. We decided to drill down to 11 AT&T store locations which are prominent in the Dallas areas and have considerable number of reviews on social media website. We didn’t take AT&T dealer since the problem focuses on AT&T retail stores. The store locations list is given below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Store**  **Address** | **Store Alias** | **Store Web Business Id (For Twitter and Yelp)** | **Geolocation(For twitter)** | **Zip**  **Code** |
| 208 S Akard Street, Ste 110, Dallas, | Dallas 1 | at-and-t-dallas-6 | 32.779555, -96.8009703 | TX 75202 |
| 3329 Oak Lawn Avenue Dallas | Dallas 2 | at-and-t-dallas-13 | 32.8111091,-96.8092962 | TX 75219 |
| 5616 Lemmon Ave Dallas | Dallas 3 | at-and-t-dallas-17 | 32.8293128,-96.8272358 | TX 75209 |
| 8687 N Central Expressway Suite 2340 | Dallas 4 | at-and-t-dallas-7 | 32.8685017,-96.7757012 | TX 75225 |
| 1152 North Buckner Blvd | Dallas 5 | at-and-t-dallas-9 | 32.8342578,-96.7045404 | TX 75218 |
| 9100 N Central Expressway Suite 105 | Dallas 6 | at-and-t-dallas | 32.8740567,-96.771404 | TX 75231 |
| 5959 Royal Lane Dallas | Dallas 7 | at-and-t-dallas-16 | 32.8957338,-96.8079243 | TX 75230 |
| 7800 N. Macarthur Boulevard Suite 150 | Dallas 8 | at-and-t-irving-2 | 32.913273, -96.958064 | TX 75075 |
| 701 N Central Expy Plano, TX 75075 | Dallas 9 | at-and-t-plano-6 | 33.009892, -96.709061 | TX 75063 |
| 5100 Beltline Road Ste. 1032 | Dallas 10 | at-and-t-addison-2 | 32.953929, -96.821254 | TX 75254 |
| 13710 Dallas Parkway Suite I | Dallas 11 | at-and-t-dallas-5 | 32.934372, -96.820672 | TX 75240 |

**Geolocation**

Since one of the important factors during the case study was store location. Different social media store location in different ways, For twitter the geolocation is in latitude , longitudes coordinates. It becomes necessary to extract geolocation information from a tweet. Moreover , it is not easy to track the store location from twitter. Approximation based on nearest store has to be done to determine the score. In case of google and yelp , the reviews are categorized by store location . But , during the analysis , there should be unified standard . Hence , the geolocation coordinates is the criteria for the location of stores to collect the reviews from different social media platforms.

**Store Alias**

Store alias would serve as a unique identifier for the stores. It would serve as an identifier for the stores during the data collection as well as during the warehousing and analysis process.  
  
**Store Web Business Id (For Twitter and Yelp)**

Since , google and yelp reviews are based on a store location. The data collection for yelp and google can work in combination during the data collection process.

**Data Collection from twitter**

Data collection will happen through the twitter API (<https://api.twitter.com/1.1/search/tweets.json>")

Data will be fetched in json format and it will be converted to .csv format for keeping the data in common format for facilitating the use of Hadoop distributed File System (HDFS). From the perspective of making the application future proof and scalable.   
  
  
The application for data collection has been abstracted is such a way that the api key and other configurations are handled well from a config.ini file.

Base.py - Contains the base script which loads the config for twitter. It also contains the fetch feeds method , which is called by app.py for filtering on the basis of search query

App.py - Search query strings and called to fetch\_twitter methods is done in the script. Query strings are critical for Twitter due to the fact that only customer satisfaction and product experience data should be filtered.

**Data Collection from Google and Yelp**Out of all of the places online that you can review a business, only two places spring to mind for many consumers: Yelp and Google.Both of them are powerhouses with a long history in the business. Both of them are influential.Yelp has a lot of power online. Aside from its community, it’s still the #1 place users go to search for reviews

The yelp-api is used for fetching the reviews , the architecture is similar to that of twitter. The main difference is reviews are being fetched using the yelp business-id. Yelp API supports search by keyword via "term" parameter and filter by category via category\_filter.

Google Places API was used to obtain the reviews by leveraging the business-id from yelp.

A code snippet is shown below which shows how google is leveraging the retail stores fetched during review collection from yelp API.

yelp\_results = yelp\_instance.Search(location='US', term="google", category\_filter="AT&T")  
  
data = []  
  
**for** business **in** yelp\_results.businesses:  
 places = google\_instance.text\_search(  
 business.id,business,name  
 lat\_lng={  
 'lat': business.location.coordinate['latitude'],  
 'lng': business.location.coordinate['longitude']  
 }  
 )  
  
 att\_stores = {}  
 **for** place **in** places.places:  
 place.get\_details()  
 company.update(place.details)  
  
 data.append(att\_stores)

**2) Identify key factors that go into positive or negative customer experiences from the data you collected- list the different social media platforms involved in analysis**

The key factors for analyzing the sentiment was mainly the text of the review ,along with geolocation of the user.

Individual words and short sequences of words (n-grams) and comparing them with a probability model. It can also detect negations in phrases, i.e, the phrase "not bad" will be classified as positive despite having two individual words with a negative sentiment.  
  
AlchemyAPI's sentiment analysis algorithm looks for words that carry a positive or negative connotation then figures out which person, place or thing they are referring to. It also understands negations (i.e. "this car is good" vs. "this car is not good") and modifiers (i.e. "this car is good" vs. "this car is really good").  
  
  
Sentiment Analysis Example for some tokens found in review :

|  |  |  |
| --- | --- | --- |
| Analyzed Item | Target | Sentiment |
| AT &T | Entity | Mixed |
| Quick Service | Keyword | Positive |
| Slow Response | Keyword | Negative |
| ATT Cares | Entity | Positive |
| Happy Customer | Keyword | Positive |
| Customer love | Keyword | Positive |
| Bad Service | Keyword | Negative |