FORDHAM UNIVERSITY

Sentiment Analysis of Fashion Reviews from Vogue 2016

Task 1: Open the file "fashion.csv" using pandas - Sample dataset

```
df = pd.read_csv('fashion.csv')
df.head()
```

	year	season	brand	author of review	location	time	review text
0	2016	Spring	A Dtacher	Kristin Anderson	NEW YORK	September 13, 2015	Detachment was the word of the day at A Dtache...
1	2016	Spring	A.F. Vandevorst	Luke Leitch	PARIS	October 1, 2015	You heard this collection coming long before y...
2	2016	Spring	A.L.C.	Kristin Anderson	NEW YORK	September 21, 2015	August saw the announcement of big news for A....
3	2016	Spring	A.P.C.	Nicole Phelps	PARIS	October 3, 2015	They call me the king of basics, Jean Touitou ...
4	2016	Spring	A.W.A.K.E.	Maya Singer	NEW YORK	October 21, 2015	Natalia Alaverdian is a designer with a lot of...

Task 2: Generating positive, negative, neutral

- Method 1: By counting how many positive words in the review_text as well as how many negative words.

	year	season	brand	author of review	location	time	review text	positive_1	negative_1	neutral_1
0	2016	Spring	A Dtacher	Kristin Anderson	NEW YORK	September 13, 2015	Detachment was the word of the day at A Dtache...	16	12	4
1	2016	Spring	A.F. Vandevorst	Luke Leitch	PARIS	October 1, 2015	You heard this collection coming long before y...	14	18	-4
2	2016	Spring	A.L.C.	Kristin Anderson	NEW YORK	September 21, 2015	August saw the announcement of big news for A....	22	25	-3
3	2016	Spring	A.P.C.	Nicole Phelps	PARIS	October 3, 2015	They call me the king of basics, Jean Touitou ...	11	26	-15
4	2016	Spring	A.W.A.K.E.	Maya Singer	NEW YORK	October 21, 2015	Natalia Alaverdian is a designer with a lot of...	14	20	-6

- Method 2: By using SentimentIntensityAnalyzer from NLTK
  - NLTK.Vader is a lexicon and rule-based sentiment analysis package that is specifically attuned to sentiments expressed in social media.
  - NLTK.Vader returns four columns, which are **'neg', 'neu', 'pos', scores** that reflect how negative, neutral or positive a document is, and 'compound', the sum of all the lexicon ratings, which have been standardized to range between -1 and 1. The following figure is the sample result from NLTK.Vader.

	year	season	brand	author of review	location	time	review text	positive_1	negative_1	neutral_1	positive_2	negative_2	neutral_2	compound
0	2016	Spring	A Dtacher	Kristin Anderson	NEW YORK	September 13, 2015	Detachment was the word of the day at A Dtache...	16	12	4	0.105	0.012	0.883	0.9625
1	2016	Spring	A.F. Vandevorst	Luke Leitch	PARIS	October 1, 2015	You heard this collection coming long before y...	14	18	-4	0.059	0.052	0.889	0.4378
2	2016	Spring	A.L.C.	Kristin Anderson	NEW YORK	September 21, 2015	August saw the announcement of big news for A....	22	25	-3	0.122	0.006	0.872	0.9842
3	2016	Spring	A.P.C.	Nicole Phelps	PARIS	October 3, 2015	They call me the king of basics, Jean Touitou ...	11	26	-15	0.100	0.054	0.846	0.9373
4	2016	Spring	A.W.A.K.E.	Maya Singer	NEW YORK	October 21, 2015	Natalia Alaverdian is a designer with a lot of...	14	20	-6	0.116	0.017	0.866	0.9738

- Method 3: By using Textblob Sentiment analysis
    - TextBlob is a Python library for processing textual data. In this case, I'm using PatternAnalyzer, which based on the pattern library in TextBlob.
    - TextBlob results return a list of **polarity score** within range [-1,1], where 1 for strong positive, -1 for strong negative, 0 for Neutral; and a list of subjectivity within range [0,1] where 0 is very objective, and 1 is very **subjective, emotional**;
    - The following attached figure is the sample results from TextBlob.

brand	author of review	location	time	review text	positive_1	negative_1	neutral_1	positive_2	negative_2	neutral_2	compound	polarity	subjectivity
A Dtacher	Kristin Anderson	NEW YORK	September 13, 2015	Detachment was the word of the day at A Dtache...	16	12	4	0.105	0.012	0.883	0.9625	0.176518	0.506270
A.F. Vandevorst	Luke Leitch	PARIS	October 1, 2015	You heard this collection coming long before y...	14	18	-4	0.059	0.052	0.889	0.4378	0.134861	0.473283
A.L.C.	Kristin Anderson	NEW YORK	September 21, 2015	August saw the announcement of big news for A....	22	25	-3	0.122	0.006	0.872	0.9842	0.178247	0.466122
A.P.C.	Nicole Phelps	PARIS	October 3, 2015	They call me the king of basics, Jean Touitou ...	11	26	-15	0.100	0.054	0.846	0.9373	0.068447	0.383085
A.W.A.K.E.	Maya Singer	NEW YORK	October 21, 2015	Natalia Alaverdian is a designer with a lot of...	14	20	-6	0.116	0.017	0.866	0.9738	0.249673	0.593464

Task 3: Save document with four columns 1. ID, 2. Positive, 3. Negative, and 4. Neutral scores

```
id_ = list(range(1,df.shape[0]+1))
df_task3 = df[['positive_1','negative_1','neutral_1']]
df_task3['ID']=id_
df_task3 = df_task3[['ID','positive_1','negative_1','neutral_1']]
df_task3.to_csv('sentiment_output.csv')
```

Task 4: Merge the output file with the original input file.

ason	brand	author of review	location	time	review text	positive_1	negative_1	neutral_1	positive_2	negative_2	neutral_2	compound	polarity	subjectivity
oring	A Dtacher	Kristin Anderson	NEW YORK	2015-09-13	Detachment was the word of the day at A Dtache...	16	12	4	0.105	0.012	0.883	0.9625	0.176518	0.506270
oring	A.F. Vandevorst	Luke Leitch	PARIS	2015-10-01	You heard this collection coming long before y...	14	18	-4	0.059	0.052	0.889	0.4378	0.134861	0.473283
oring	A.L.C.	Kristin Anderson	NEW YORK	2015-09-21	August saw the announcement of big news for A....	22	25	-3	0.122	0.006	0.872	0.9842	0.178247	0.466122
oring	A.P.C.	Nicole Phelps	PARIS	2015-10-03	They call me the king of basics, Jean Touitou ...	11	26	-15	0.100	0.054	0.846	0.9373	0.068447	0.383085
oring	A.W.A.K.E.	Maya Singer	NEW YORK	2015-10-21	Natalia Alaverdian is a designer with a lot of...	14	20	-6	0.116	0.017	0.866	0.9738	0.249673	0.593464

Task5: Summarize the sentiment scores for each of the four cities – New York, Paris, Milan, and London.

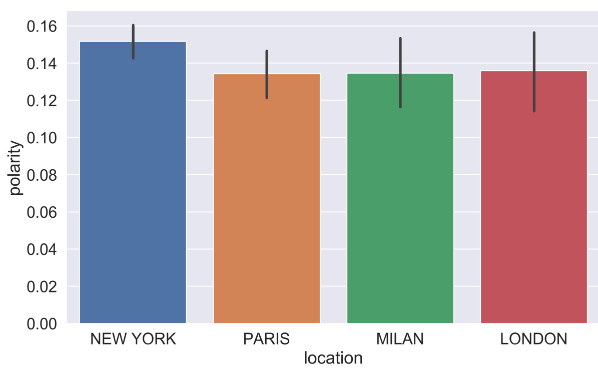
```
sent_by_brand.sort_values('number_reviews').head()
```

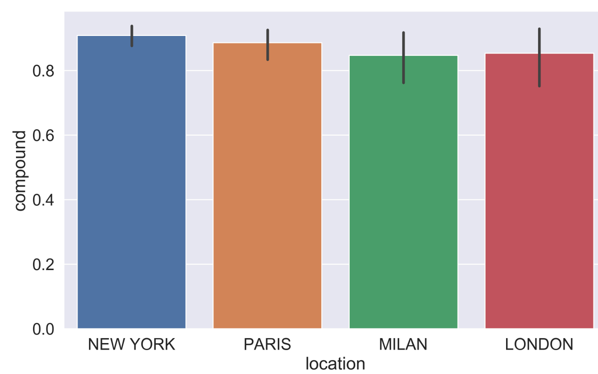
brand	positive_1	negative_1	neutral_1	positive_2	negative_2	neutral_2	compound	polarity	subjectivity	number_reviews
1205	10	17	-7	0.086	0.024	0.890	0.8831	0.064933	0.541516	1
Osklen	22	21	1	0.173	0.038	0.789	0.9950	0.097396	0.417329	1
Oscar de la Renta	34	29	5	0.097	0.020	0.883	0.9876	0.127586	0.426100	1
Orley	15	19	-4	0.099	0.025	0.876	0.9767	0.099147	0.456887	1
Orla Kiely	14	19	-5	0.123	0.057	0.820	0.9565	0.072115	0.459402	1

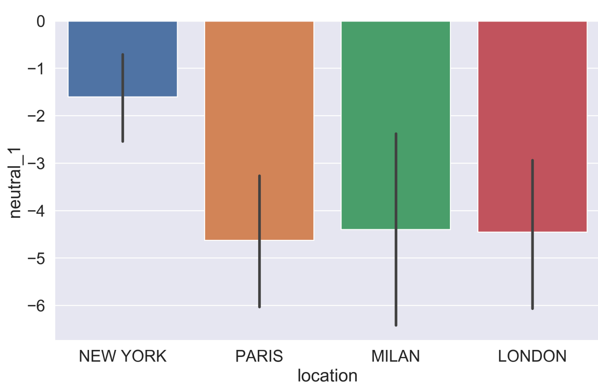
It seems that every brand got only one review, thus we tend to find insightful knowledges based on location.

		Compound_Vader		Polarity_TextBlob		Subjectivity_TextBlob	
	Count	Mean	STD	Mean	STD	Mean	STD
LONDON	62	0.853624	0.34805	0.135992	0.082444	0.496022	0.091253
MILAN	53	0.846585	0.312263	0.134584	0.070444	0.45984	0.074726
NEW YORK	209	0.908633	0.229015	0.151715	0.069223	0.466952	0.070011
PARIS	110	0.886355	0.239649	0.134372	0.067735	0.466143	0.07294

This three methods shows that New York received the highest sentiment score among 4 cities,

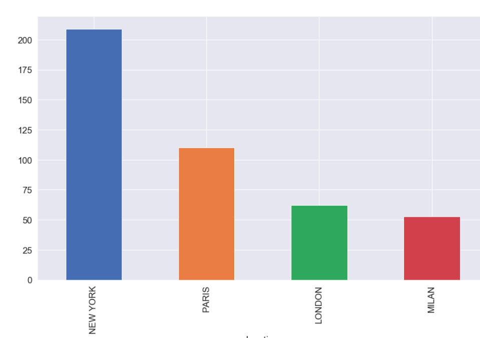






Those graphs show that, for fashion brand or a buyer, New York should be your first choice!

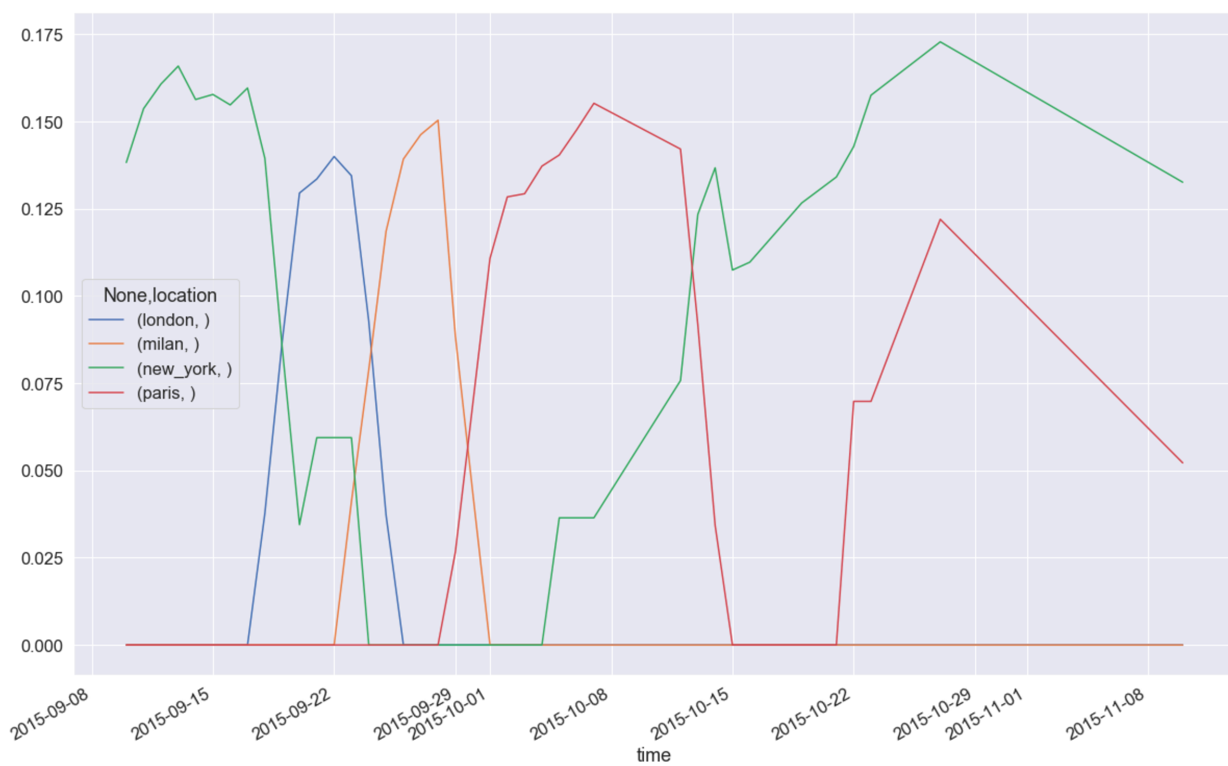
It received more positive sentiment score reviews from the critics.



From count of reviews in each city, we can see that most of reviews come from New York.For business purpose. New York is the best for promotion

6. Brief description of other improvements you can make *(bonus points)*
Time Series Analysis

Time series analysis !



1. Transform time columns into datetime format.
2. Computing the sentiment score by location and date.
3. Plotting result.

Advices:

1.In terms of brand advertising, New York and London are better choices, which enjoys continual and more positive reviews.

2.Milan and Paris is not a good choice.

Other way like generating new rule to delivery a new sentiment score.