

## Steps to run the notebook file:

Step 1) Install tweepy API using the command “pip install tweepy”

Step 2) Write the below in a separate notebook in a separate cell

```
import nltk
```

```
nltk.download()
```

a) By doing this, a pop-up box will be opened. In that pop-up box, go to models menu and search for the identifier “vader\_lexicon”, “punkt”, “averaged\_perceptron” and click on download.

b) In the corpora section, search for the identifier “stopwords” and click on download.

**Vader\_lexicon:**

Collections	Corpora	Models	All Packages
Identifier	Name	Size	Status
maxent_ne_chunker	ACE Named Entity Chunker (Maximum entropy)	12.8 MB	installed
maxent_treebank_pos	Treebank Part of Speech Tagger (Maximum entropy)	9.7 MB	installed
moses_sample	Moses Sample Models	10.5 MB	not installed
mwa_ppdb	The monolingual word aligner (Sultan et al. 2015)	1.5 MB	not installed
perluniprops	perluniprops: Index of Unicode Version 7.0.0 characters	97.9 KB	not installed
porter_test	Porter Stemmer Test Files	195.8 KB	not installed
punkt	Punkt Tokenizer Models	13.0 MB	installed
rsdp	RSLP Stemmer (Removedor de Sufixos da Lingua Portuguesa)	3.7 KB	not installed
sample_grammars	Sample Grammars	19.8 KB	not installed
snowball_data	Snowball Data	6.5 MB	not installed
spanish_grammars	Grammars for Spanish	4.0 KB	not installed
tagsets	Help on Tagsets	33.7 KB	installed
universal_tagset	Mappings to the Universal Part-of-Speech Tagset	18.6 KB	installed
vader_lexicon	VADER Sentiment Lexicon	88.4 KB	installed

Stopwords:

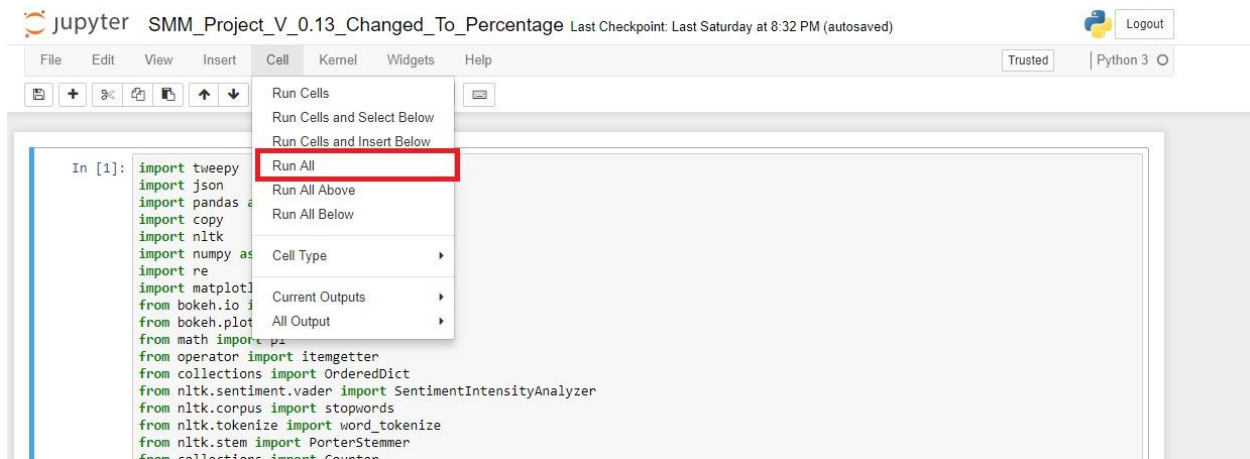
Collections	Corpora	Models	All Packages
Identifier	Name	Size	Status
ptb	Penn Treebank	6.1 KB	not installed
qc	Experimental Data for Question Classification	122.5 KB	not installed
reuters	The Reuters-21578 benchmark corpus, ApteMod v	6.1 MB	installed
rte	PASCAL RTE Challenges 1, 2, and 3	377.2 KB	not installed
semcor	SemCor 3.0	4.2 MB	not installed
senseval	SENSEVAL 2 Corpus: Sense Tagged Text	2.1 MB	installed
sentence_polarity	Sentence Polarity Dataset v1.0	478.8 KB	not installed
sentiwordnet	SentiWordNet	4.5 MB	not installed
shakespeare	Shakespeare XML Corpus Sample	464.3 KB	not installed
sinica_treebank	Sinica Treebank Corpus Sample	878.2 KB	not installed
smultron	SMULTRON Corpus Sample	162.3 KB	not installed
state_union	C-Span State of the Union Address Corpus	789.8 KB	installed
stopwords	Stopwords Corpus	17.1 KB	installed
subjecttree	SubjectTree Dataset v1.0	500.4 KB	not installed

Averaged\_perceptron and Punkt:

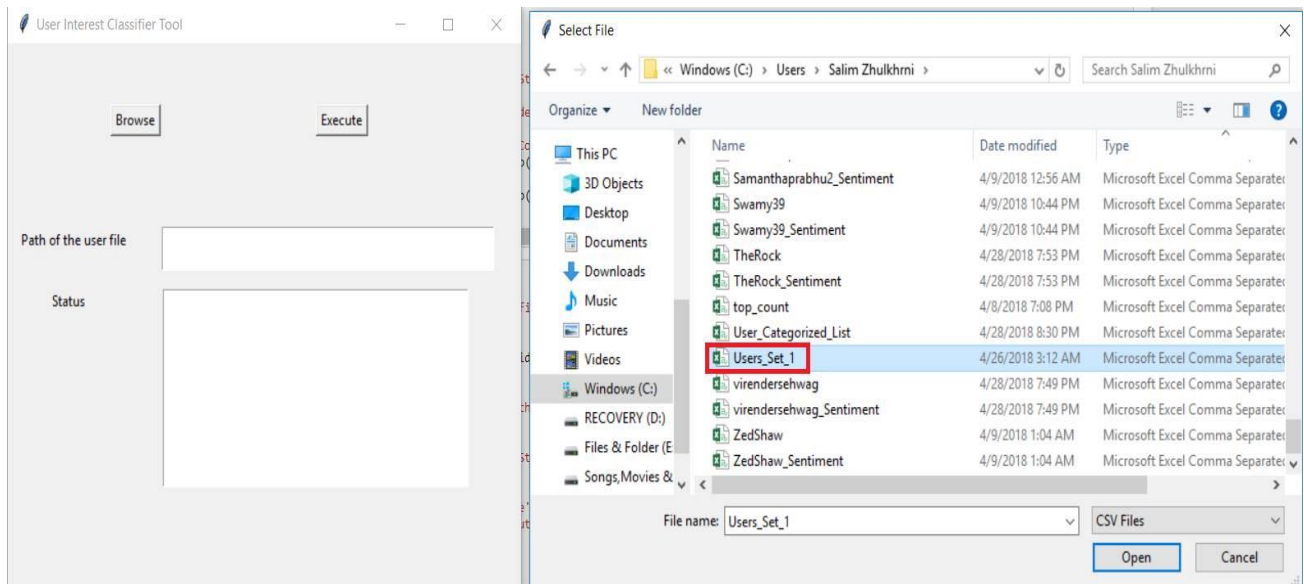
Collections	Corpora	Models	All Packages
Identifier	Name	Size	Status
averaged_perceptron	Averaged Perceptron Tagger	2.4 MB	installed
averaged_perceptron_russian	Averaged Perceptron Tagger (Russian)	8.2 MB	not installed
basque_grammars	Grammars for Basque	4.6 KB	not installed
bllip_wsj_no_aux	BLLIP Parser: WSJ Model	23.4 MB	not installed
book_grammars	Grammars from NLTK Book	8.9 KB	installed
large_grammars	Large context-free and feature-based grammars for	277.1 KB	not installed
maxent_ne_chunker	ACE Named Entity Chunker (Maximum entropy)	12.8 MB	installed
maxent_treebank_pos	Treebank Part of Speech Tagger (Maximum entropy)	9.7 MB	installed
moses_sample	Moses Sample Models	10.5 MB	not installed
mwa_ppdb	The monolingual word aligner (Sultan et al. 2015)	1.5 MB	not installed
perluniprops	perluniprops: Index of Unicode Version 7.0.0 chara	97.9 KB	not installed
porter_test	Porter Stemmer Test Files	195.8 KB	not installed
punkt	Punkt Tokenizer Models	13.0 MB	installed

Step 3) Place the excel sheet named ‘**Category\_List.csv**’ in your present working directory where the provided notebook is referred.

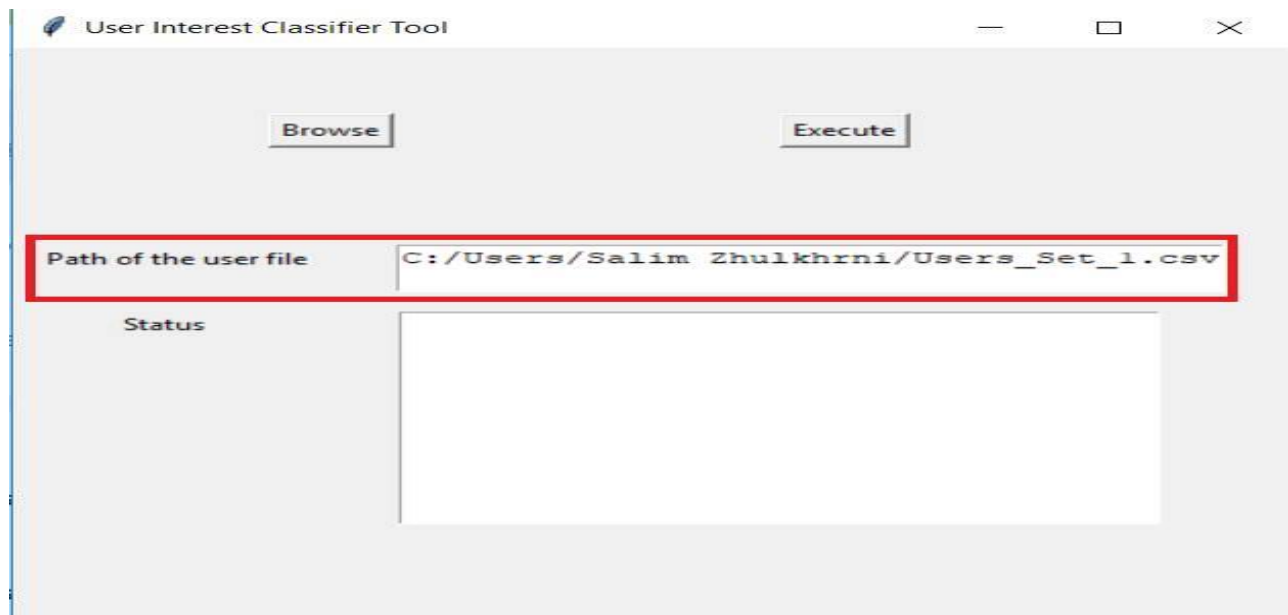
Step 4) Open the provided notebook file and click on ‘**Run All**’ option from the ‘Cell’ menu.



Step 5) A UI dialog box would be opened. In the UI dialog box, click on browse button and select the file name 'User\_Set\_1' and click on Open.



Step 6) Ensure that the path is populated in 'path of the user file' field.



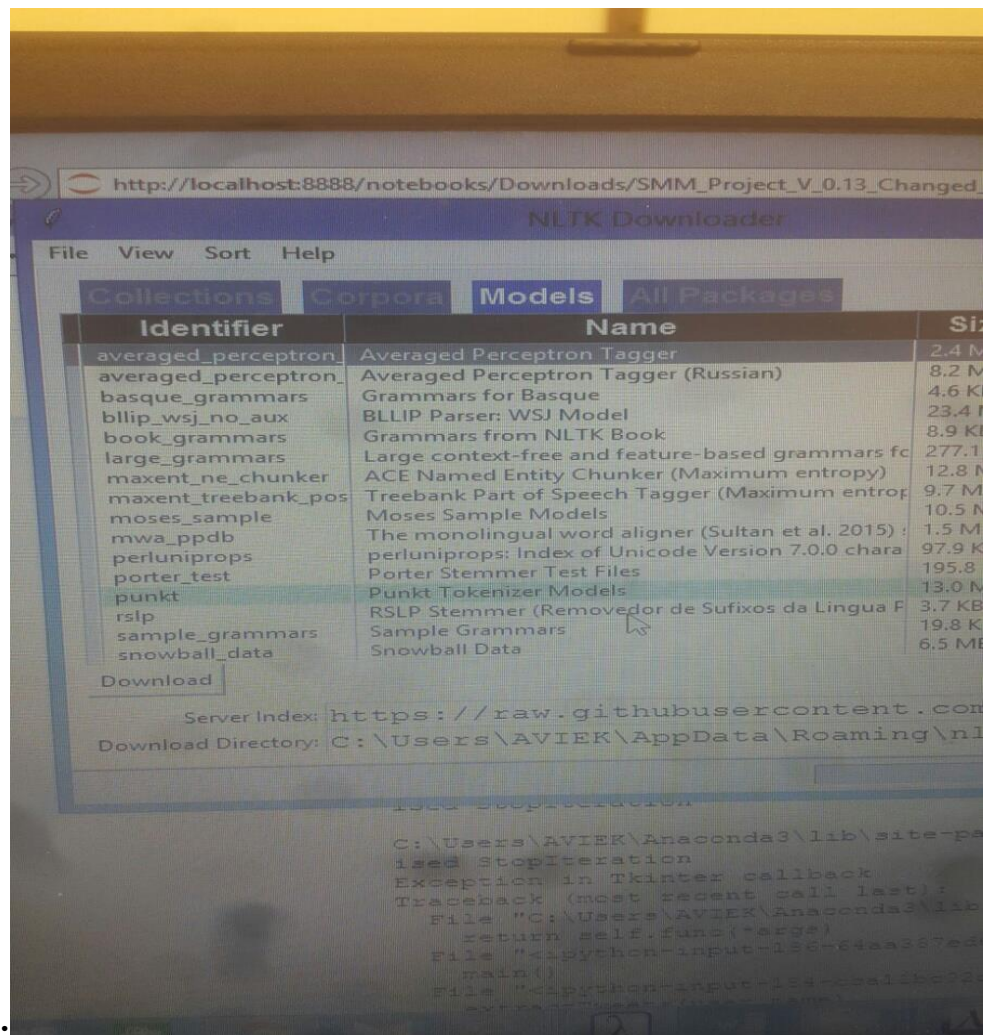
Step 7) Click on execute button and observe the pie charts generated in the notebook file and bar chart being generated in a separate tab.

Step 8) Verify that the file named 'Users\_Catergory\_List.csv' gets generated that contains the numbers of users against each category.

Untitled5.ipynb	4/5/2018 3:59 AM	IPYNB File	1 KB
Untitled6.ipynb	4/19/2018 11:04 PM	IPYNB File	14 KB
Untitled7.ipynb	4/28/2018 4:16 AM	IPYNB File	2 KB
Untitled8.ipynb	4/29/2018 8:42 PM	IPYNB File	1 KB
Untitled9.ipynb	5/1/2018 9:35 PM	IPYNB File	1 KB
<input checked="" type="checkbox"/> User_Categorized_List	4/28/2018 8:30 PM	Microsoft Excel Comma Separated Values File	1 KB
Users_Set_1	4/26/2018 3:12 AM	Microsoft Excel Comma Separated Values File	1 KB
virendersehwag	5/1/2018 9:56 PM	Microsoft Excel Comma Separated Values File	466 KB
virendersehwag_Sentiment	5/1/2018 9:56 PM	Microsoft Excel Comma Separated Values File	501 KB
WelcomeBack-NLP.ipynb	4/8/2018 11:34 PM	IPYNB File	122 KB
ZedShaw	4/9/2018 1:04 AM	Microsoft Excel Comma Separated Values File	493 KB
ZedShaw_Sentiment	4/9/2018 1:04 AM	Microsoft Excel Comma Separated Values File	519 KB







Punkt:

