**CSEE 5542 Big Data Analytics and Appn**

**Lab assignment 3**

Pavankumar Manchala

Class Id: 16

Team - 5

**Objectives:**

1. Image Caption Generation.
2. Data analysis using Unsupervised learning.

Platforms used:

1. Pycharm
2. IntelIJ

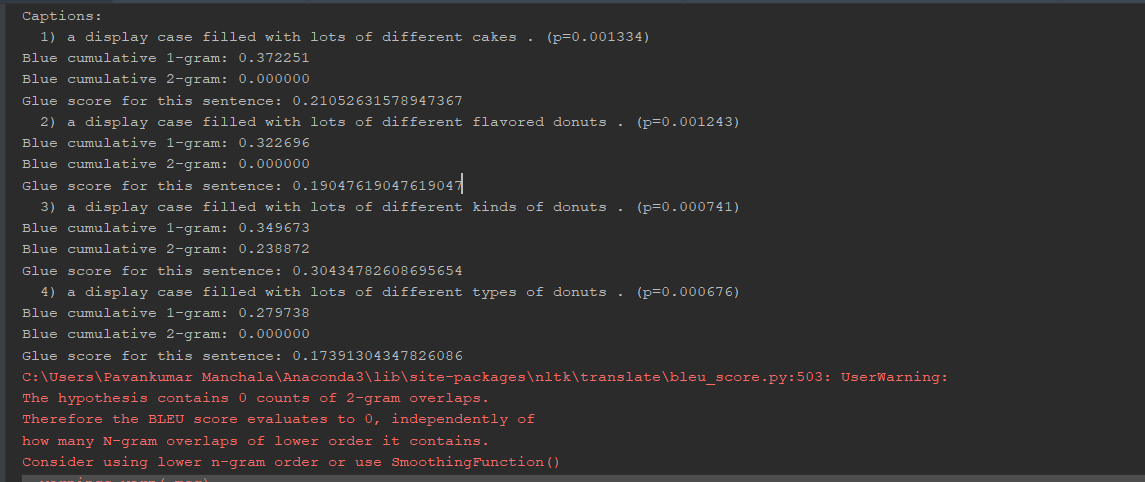
Packages Installed:

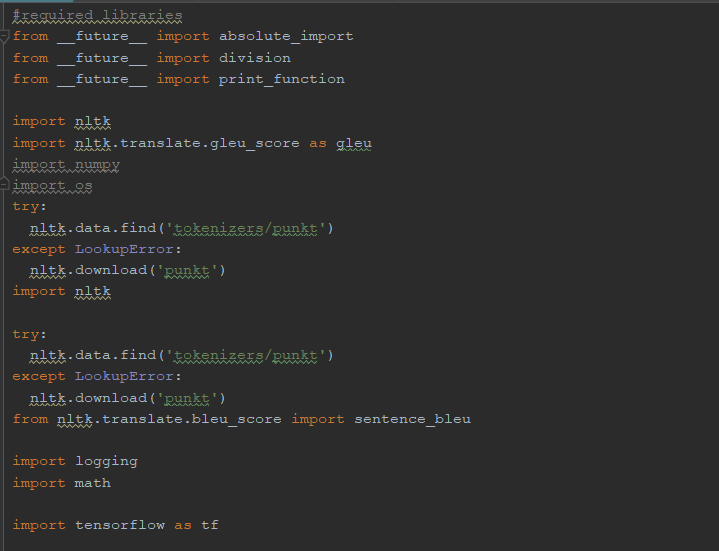
1. opencv-python
2. numpy
3. nltk
4. matplotlib
5. Tensorflow
6. Show and tell model
7. PIL
8. BLEU score
9. logging
10. Heapq

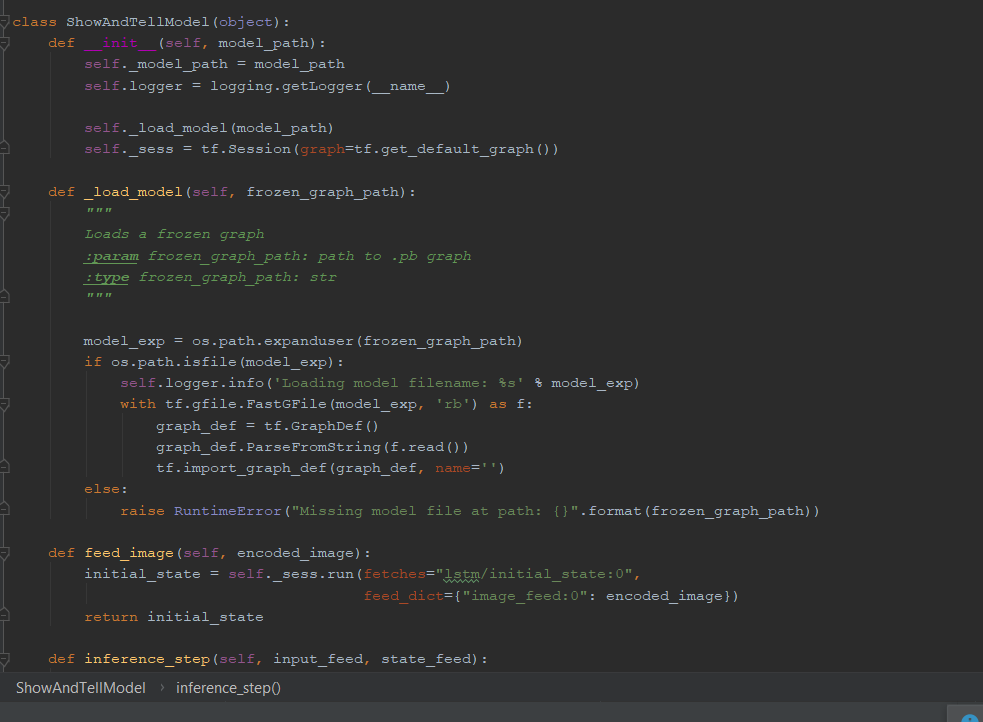
**Task 1:**

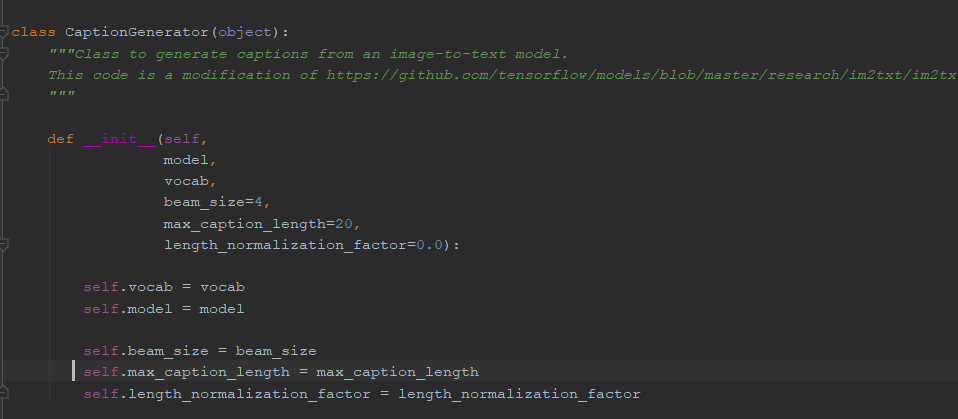
Generating captions for our own dataset using show and tell model. Reporting accuracy in BLEU, CIDER, METEOR and ROGUE measures.

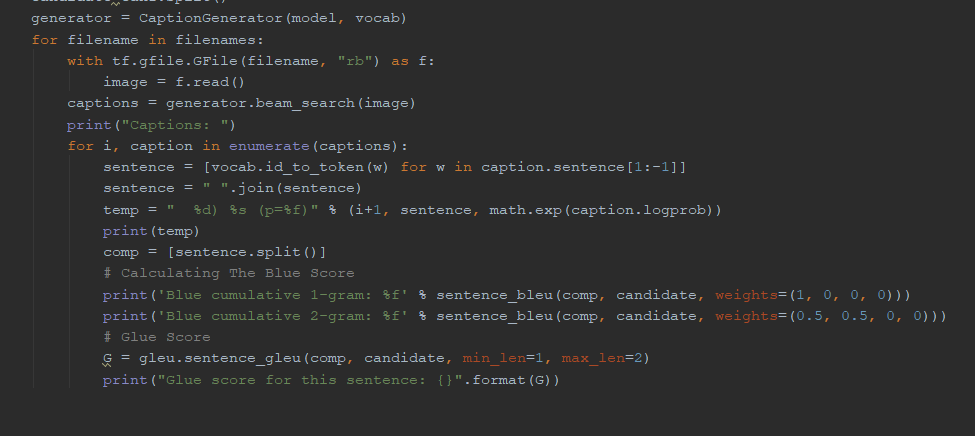
The SBU dataset has flexibility for data and contains caption and image URL files.

Show and tell model creates captions for the dataset. The output is as follows: 

The screenshots of show and tell model as follows: Requirements: 

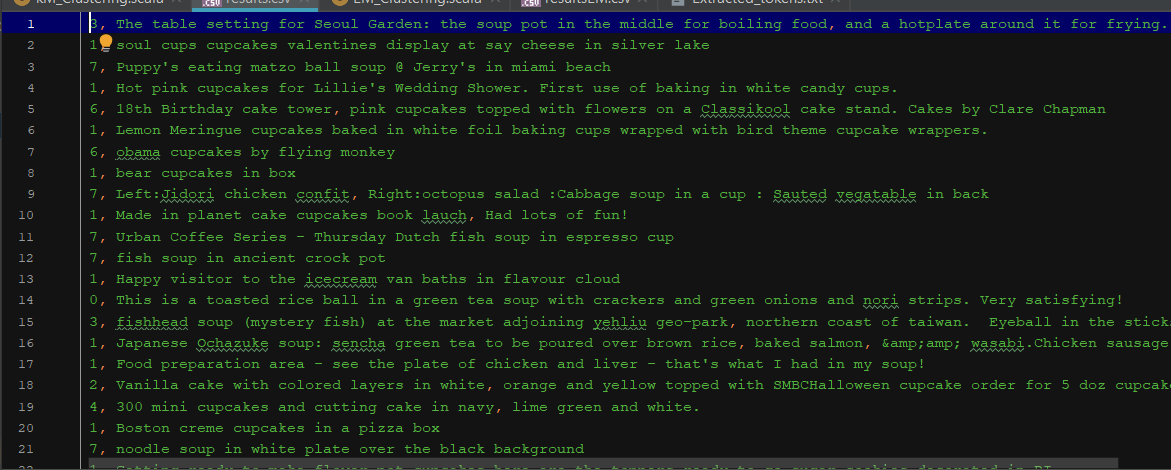
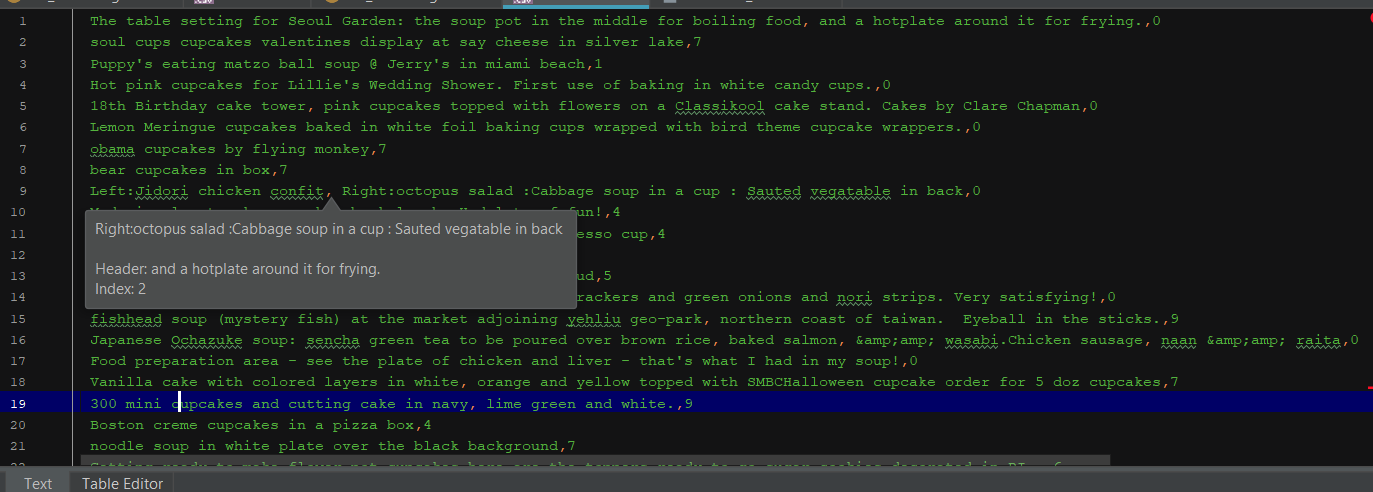
Model funct:

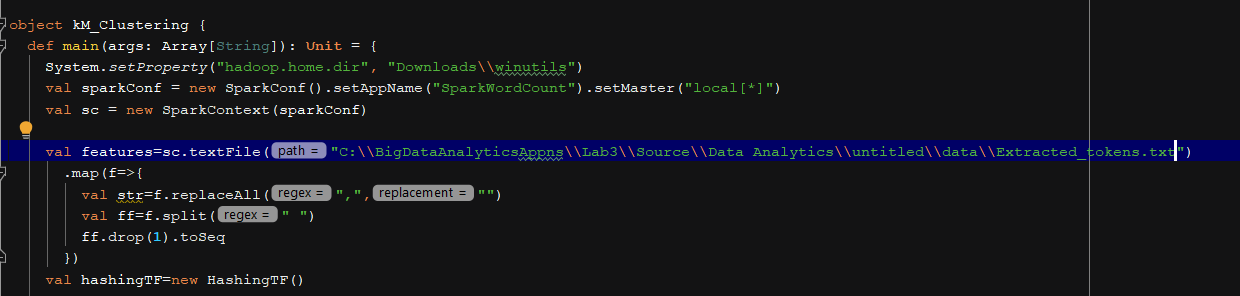
Beam size:The Beam size is 4, so 4 captions for each image generated.

Caption Generator:

**Task 2:**

Data analytics using unsupervised learning. Clustering mechanisms used like KMeans and EM clustering data into clusters respectively.

Result of KM Clustering: Result of EM Clustering:

Code for KM\_Clustering: The input data and hadoop property are initialized. 

The code represents pushing captions into hash map and then clustering them accordingly, values are saved in .csv file. 