# Module 3: Analyzing Sentence Structure

## Case Study

## Case Study

ABC Company wants to perform some text analysis and make visualization for one of their dataset.

The dataset has been taken from Kaggle.
(https://www.kaggle.com/crowdflower/twitter-airline-sentiment/home)

This is a dataset having tweets about 6 US Airlines along with their sentiments: positive, negative and neutral.

You are provided with this dataset named "Tweets.csv".
This is a snapshot of the dataset:

| | text | airline_sentiment |
|---|---|---|
| 0 | @VirginAmerica What @dhepburn said. | neutral |
| 1 | @VirginAmerica plus you've added commercials t... | positive |
| 2 | @VirginAmerica I didn't today... Must mean I n... | neutral |
| 3 | @VirginAmerica it's really aggressive to blast... | negative |
| 4 | @VirginAmerica and it's a really big bad thing... | negative |
| 5 | @VirginAmerica seriously would pay $30 a fligh... | negative |
| 6 | @VirginAmerica yes, nearly every time I fly VX... | positive |

It has tweets in 'text' column and sentiments in 'airline_sentiment' column.

1. Retrieve all tags starting from '@' in the entire dataset and save in a file called "References.txt"

2. Extract all noun phrases from their dataset and save them in different lines in a file named "Noun Phrases for <airline_sentiment> Review .txt" (You can choose your own grammar for noun phrase).
   Here <airline_sentiment> will have 3 different values: positive, negative and neutral so 3 files will be created.

3. Extract all verb phrases from their dataset and save them in different lines in a file named "Verb Phrases for <airline_sentiment> Review .txt"" (You can choose your own grammar for noun phrase).
   Here <airline_sentiment> will have 3 different values: positive , negative and neutral so 3 files will be created.

4. For each sentiment, make a well labelled pie chart showing distribution of Noun Phrases and Verb Phrases of that sentiment from the data set. Use the above created files to get the frequencies.