**Overview:**

Download amazon Camera review data from the link [https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon\_reviews\_us\_Camera\_v1\_00.tsv.gzLinks to an external site.](https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_us_Camera_v1_00.tsv.gz) . After extracting you will get a “.tsv” file where the columns are ‘tab’ separated. The description of the columns can be found in the link[https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt Links to an external site.](https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt) .  
  
For this assignment,

* you must run different queries on Apache HIVEusing this dataset.
* you must submit all your queries along with the screenshot of the results.
* every part is depended on the data insertion in 'part 1'.

**Part 1:**

* Download the dataset and insert into a table named 'CameraTable'.
* For each ‘marketplace’ & ‘product category’, find the total number of ‘review\_id’ & average ‘star\_rating’ where  ‘marketplace’ does not include ‘US’.

**Part 2:**

* Create a table 'CameraTablePart' that supports partitioning on 'star\_rating' field. Fill up data for partitions 'star\_rating=4' and 'star\_rating=2' from the 'CameraTable'.
* For each ‘star\_rating’ from 'CameraTablePart' table, find sum of ‘helpful\_votes’ and sum of ‘total\_votes’ ordered by sum of 'total\_votes' descending.

**Part 3:**

* Create a table 'CameraTableBuck' that supports bucketing on 'review\_date' field. Consider 4 buckets. For each bucket, find the minimum and maximum 'review\_date'.
* For each ‘product\_id’, find the average ‘helpful\_votes’ and average ‘total\_votes’ where average 'helpful\_votes' is greater than 2.