**SCALABLE ANALYTICS**

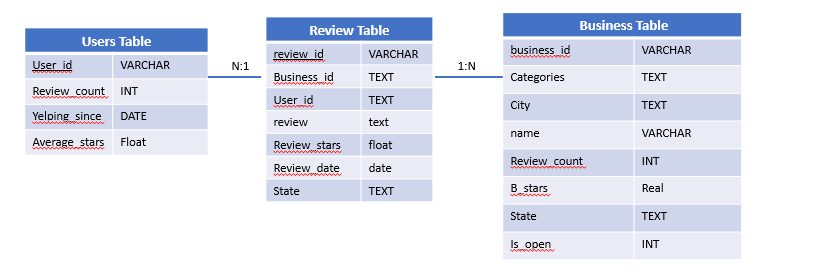
**FINAL PROJECT (TEAM-9)**

**Aim:**

Yelp is an app/website which publish crowd-sourced reviews as well as online reservation service. The objective of this project was to study the available datasets and come up with interesting insights. In this project we also used Google Cloud Platform for data processing and cluster computing

**Data Set:**

Data from the 3 tables to analyzing Yelp users, its affiliated businesses, and their reviews



**Cloud Usage**

Uploading Data Files on Google Cloud

**Text

Description automatically generated**

Python scripts files on Cluster

Text

Description automatically generated

Dataset deployed on Hadoop File System

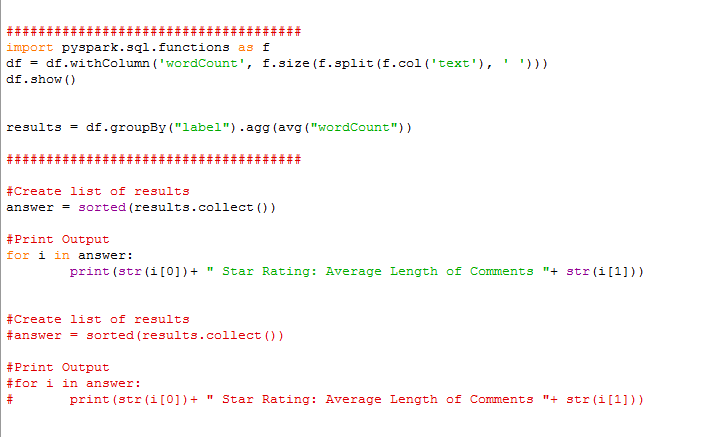
Text

Description automatically generated

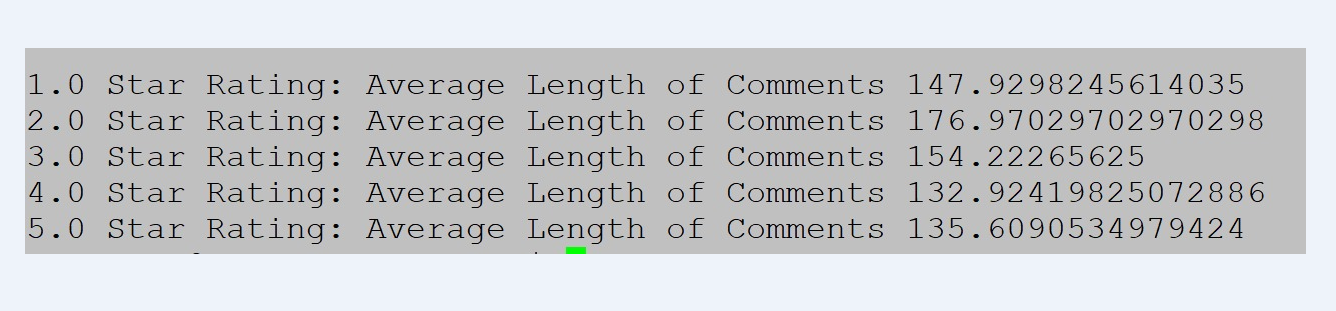
**Insight-1**

**Average Length of Comments: This part of code gives us the average length of comments for each rating**





**Result:**



We can see that 4-star rating and 5-star rating has the lowest average length of review comments=> that user tend to write the more about negative reviews compare to positive reviews.

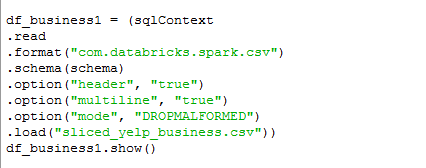
**Insight-2**

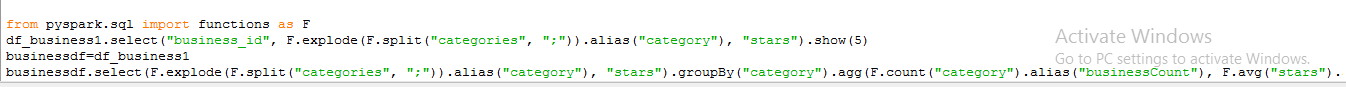
**This part of code is for getting average star rating, min star rating and max star rating for each unique category**



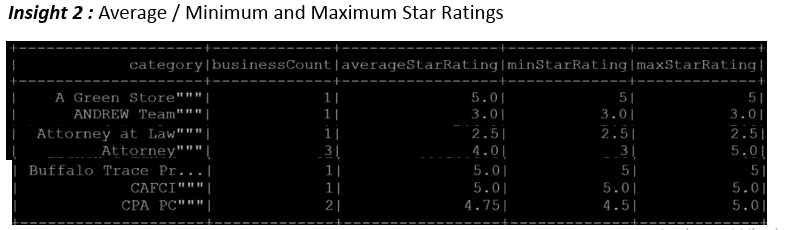






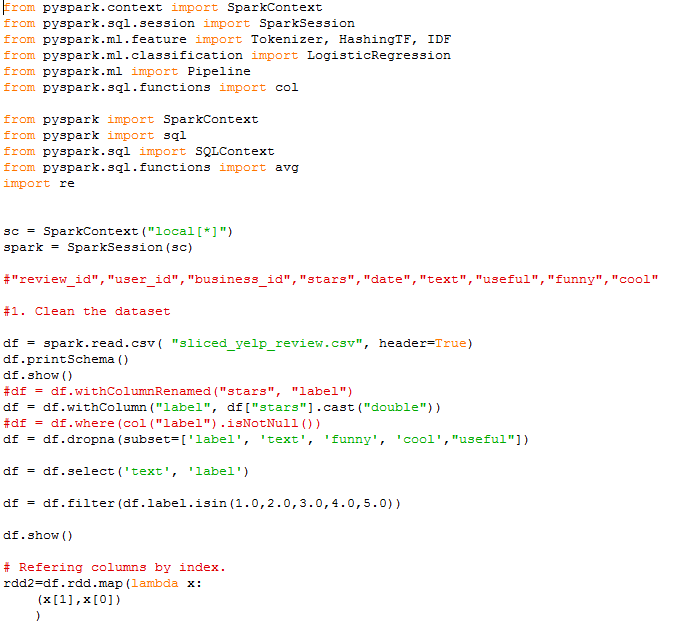


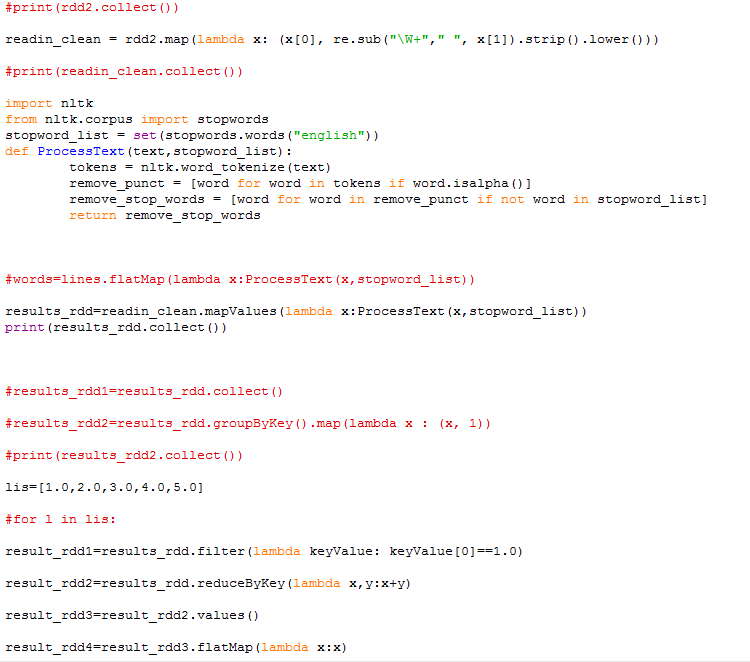
**Result:**

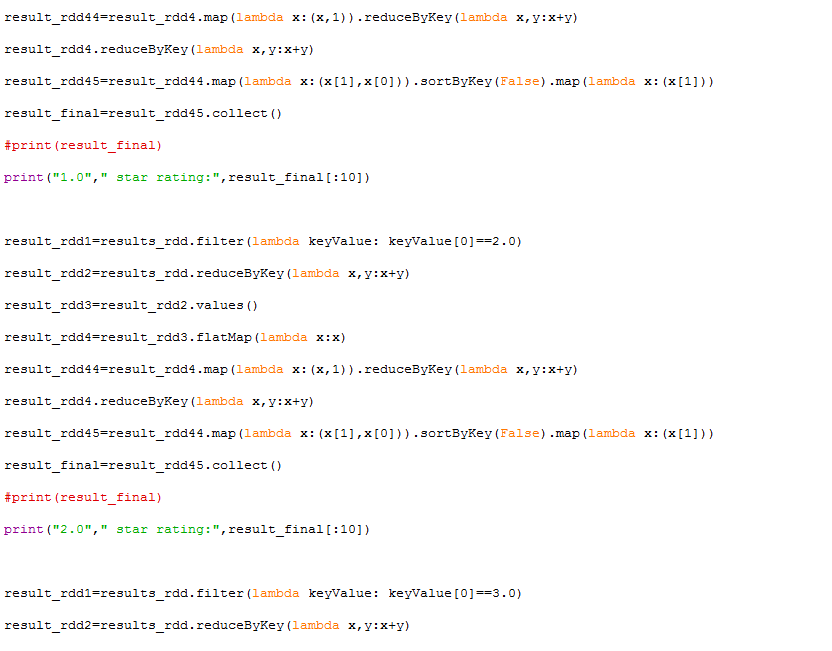


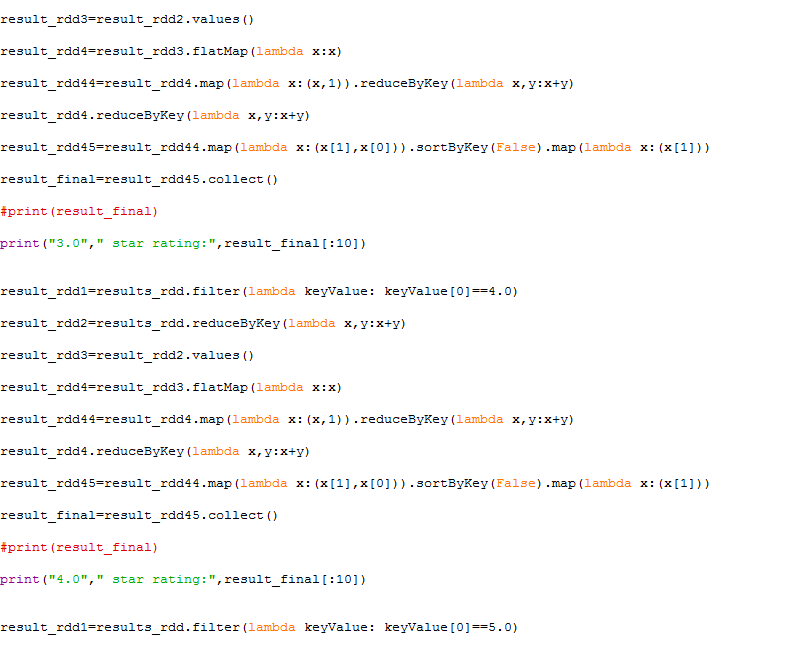
**Insight 3:**

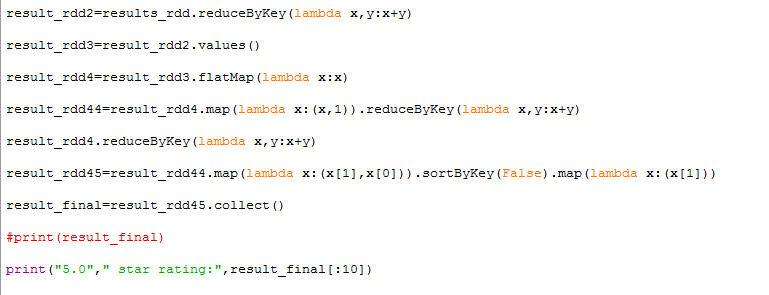
**Review: - This part of code is giving review text for each rating**



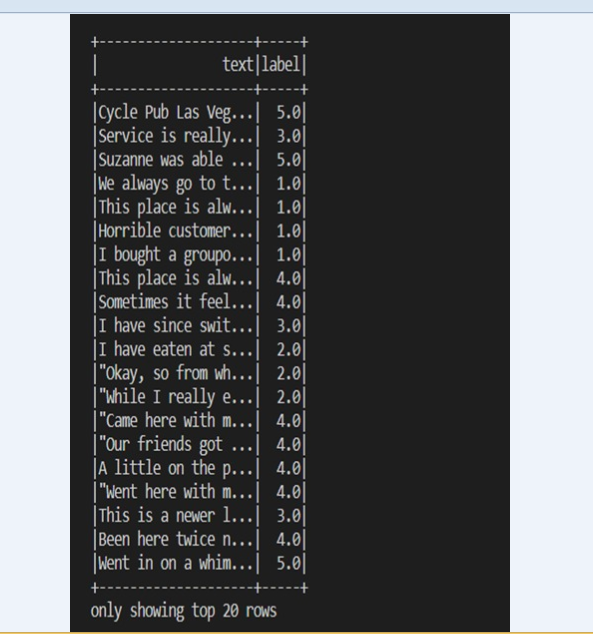








**Result:**



**Insight - 4:**

**This PY-spark code is done for each type: funny, cool, and useful and gives us the review text for each type separately**

**(a) – funny**





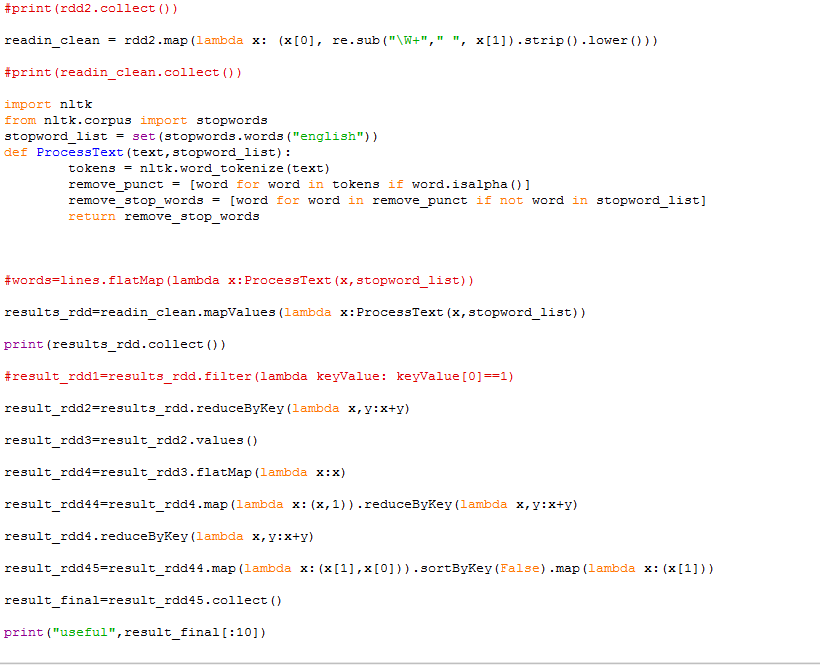
**(b) - Cool**

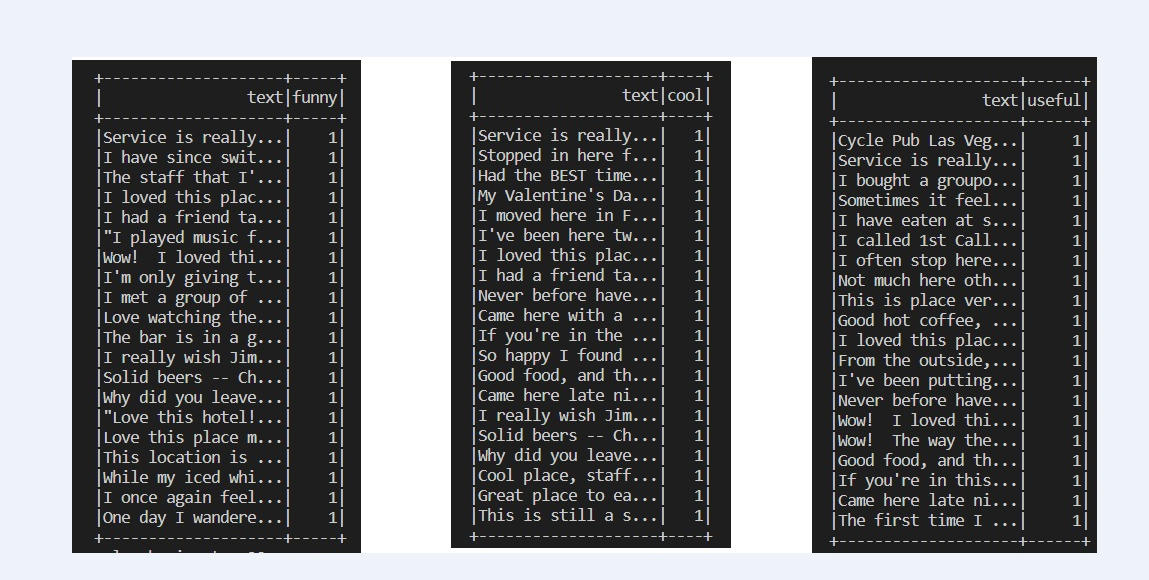




**(C)- Useful**

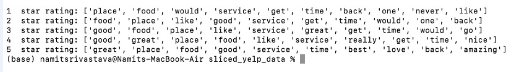




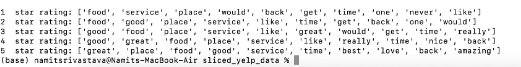


**Top 10 Words for each star rating for funny, cool, and useful respectively**

**Funny=1**



**Cool=1**



**Useful=1**



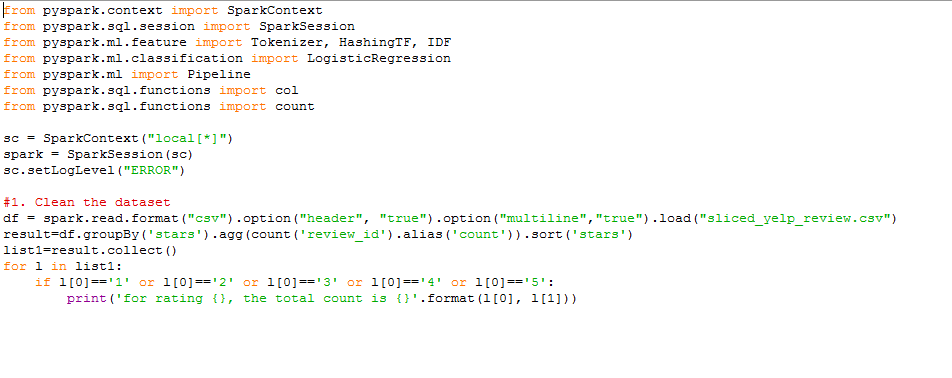
**People use “love” as one of the most frequent word in useful**

**“Never” is used very frequently in 1 star rating**

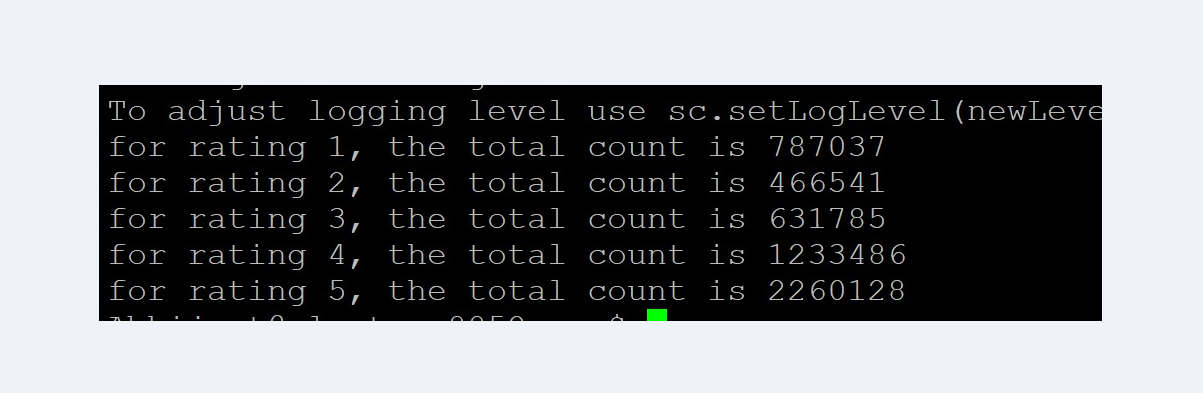
**“Time” – occurs very frequently in low ratings (in 1 & 2)**

**Insight 5:**

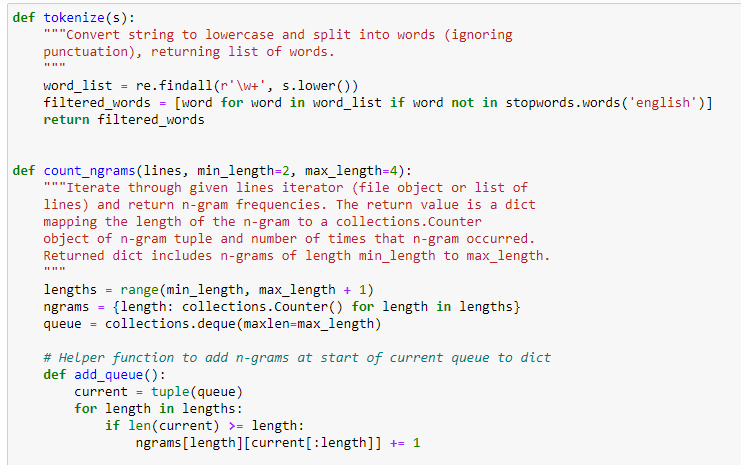
**Rating Count: -Below code gives us total count for each star rating => 5-star rating has the maximum number of counts**

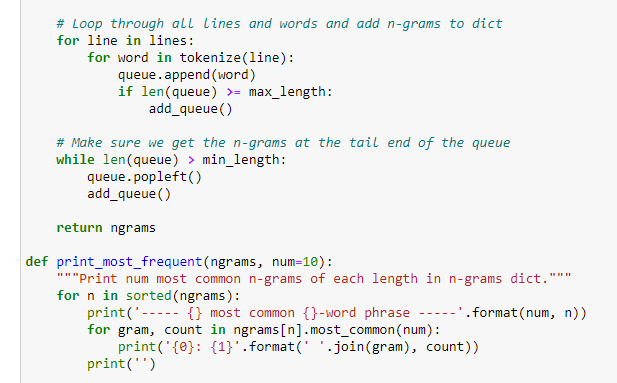


**Result**



**Insight 6:** this is word cloud for positive and negative words in review text, for this code has been written in python jupyter notebook







First screenshot (left hand side) word cloud is giving information for positive words in review text for Cool, funny, and useful and second word cloud is giving information for negative words in review text

