Project Name

REVIEW REPORTING OF A CAR FORUM

GROUP MEMBERS:

Ankit Prasad, Suraj Shaw, Srabani Santra, Anirban Manna, Sumita Maity

Acknowledgement

The achievement that is associated with the successful completion of any task would be incomplete without mentioning the names of those people whose endless cooperation made it possible. Their constant guidance and encouragement made all our efforts successful.

We take this opportunity to express our deep gratitude towards our project mentor, Swarup Kr Ghosh for giving such valuable suggestions, guidance and encouragement during the development of this project work.

We would like to convey our regards to our college '**Netaji Subhash Engineering College'** for giving us such a nice opportunity to work on this project. Last but not the least we are grateful to the HOD, CSE Dept. Dr. Anupam Ghosh who provided us such an opportunity to learn Python and encouraged us to do this project.

CERTIFICATE FROM SUPERVISOR

This is to certify that Mr. Suraj Shaw, Mr. Ankit Prasad, Miss. Srabani Santra ,Miss. Sumita Maity & Mr. Anirban Manna successfully completed the project titled **"Car Review Forum"** under my supervision during the period from November to January which is submitted to Netaji Subhash Engineering College, Kolkata.

Signature and Seal of the HOD
CSE Department

Date:

Overview

By going through a **Car forum** (Carwale, Car Dekho etc) and to perform the following goals at which **word limit is 100K and above**

Goals

1. Extracting the posts of online user discussion forum or review (user and customer review only) for a particular segment like **Hatchback or, Sedan or SUVs** and remove HTML tags and other unnecessary things to mine only user discussion text.
2. Making separate files (text/xlsx file) from the extracted data by **fuel type** (e. g. petrol or diesel) at which file will contain **only the text data** (user name and date, time are not required) and data should be in line and paragraph format as it is.
3. Making a separate file to divide the sentence according to word. (Each line will contain one word only)
4. Extracting all the words which gives some **emotions** such as unpredictable, good, excellent, bad, damage etc (The words must be in separate new line but don't give more than one line space and the punctuations marks are also must be in separate new line).

Tools & Platform

Tools

1. Python 3.5
2. Eclipse

Platform

1. Microsoft Windows 10

Software & Hardware Requirements

- Platform Independent
- Python 2.7 and above
- Text Viewer.
- Machine having
  - HDD 200MB and above
  - Processor Pentium 4 or newer processor

Procedure

1. Installed beutifulsoup at first

2. Using urllib, opened the carwale.com next by using beutifulsoup, extracted the user reviews html file without scripts & html tags

3. Separated the forms of car (Diesel, Petrol) from the extracted html file & saved the different forms into files.

4. Extracted the emotions such as unpredictable, good, excellent, bad, damage etc using regular expression.

5. Created separate files having one word per line of different reviews.

Code

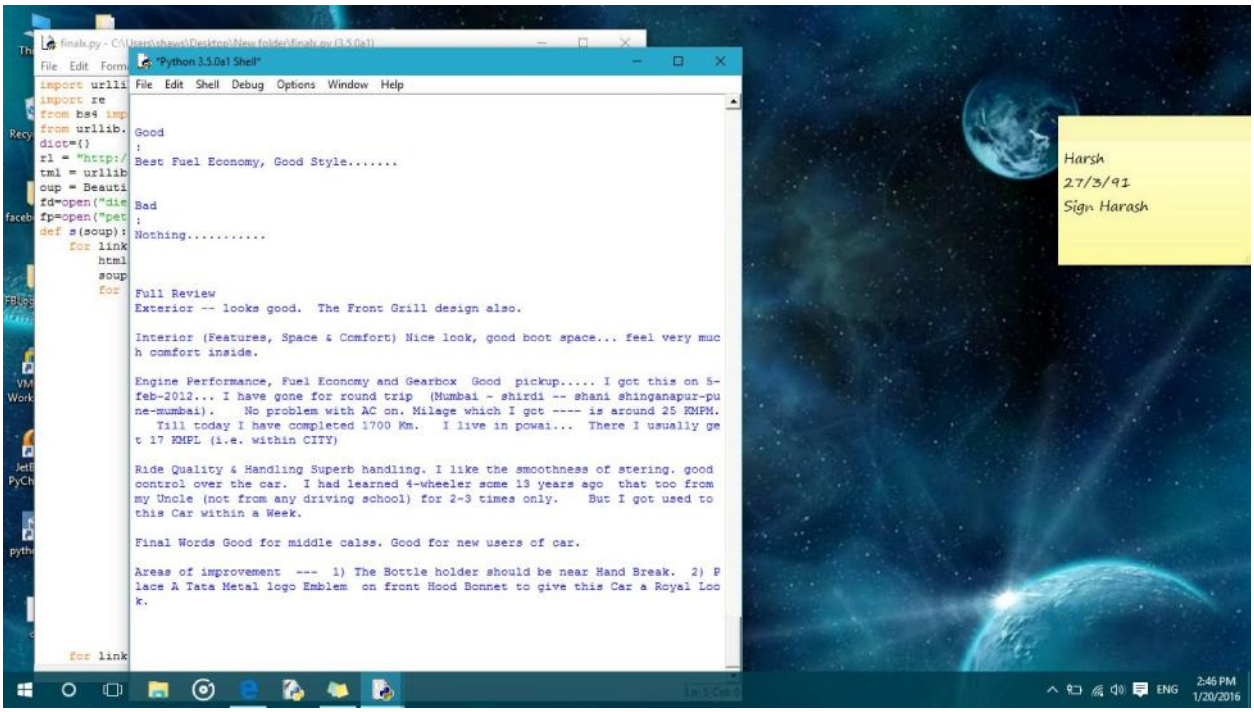
```
import urllib
import re
from bs4 import BeautifulSoup
from urllib.request import urlopen
dict={}
rl = "http://www.carwale.com/newcars/search_result.aspx?bs=1&pn=1"
tml = urllib.request.urlopen(rl).read()
oup = BeautifulSoup(tml)
fd=open("diesel.txt",'w',encoding='utf-8');
fp=open("petrol.txt",'w',encoding='utf-8');
def s(soup):
    for link in soup.findAll("a",href=True,text="read complete review"):
        html2=urllib.request.urlopen("http://www.carwale.com"+link["href"]).read()
        soup2 = BeautifulSoup(html2)
        for soup3 in soup2.findAll("div",{"class":"mid-box margin-top20"}):
            for soupx in soup3.findAll("table",{"width":"100%"}):
                if (soupx(text=re.compile('Version Reviewed')))==[]:
                    pass
                else:
                    for soupy in soupx.findAll("td",{"width":None}):
                        if soupy.get_text() in dict.keys():
                            if dict[soupy.get_text()]=="Diesel":
                                for sou3 in soup2.findAll("div",{"class":"mid-box margin-top20"}):
                                    for soux in sou3.findAll("table",{"width":"100%"}):
```

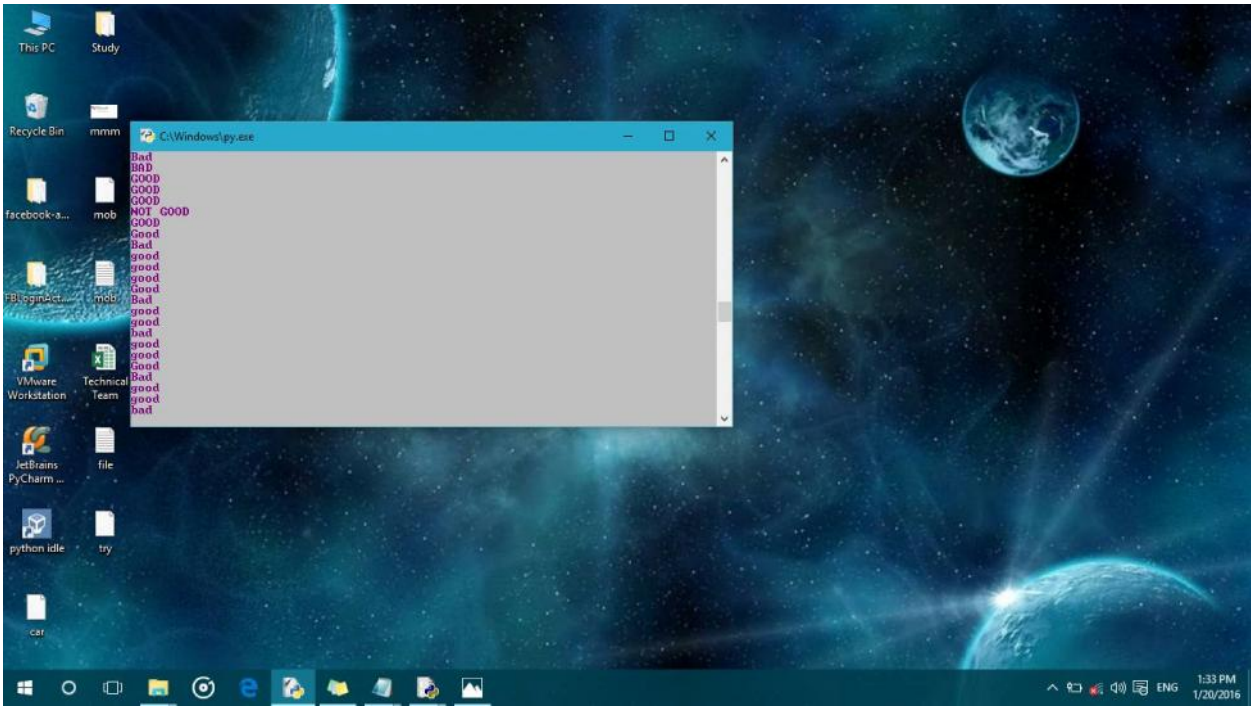
```
                print(soux.get_text())
                fd.write(soux.get_text())
            for sou4 in sou3.findAll("h3",{"class":"font14"}):
                sou5=sou4.parent
                print(sou5.get_text())
                fd.write(sou5.get_text())
        else:
            for sou3 in soup2.findAll("div",{"class":"mid-box margin-top20"}):
                for soux in sou3.findAll("table",{"width":"100%"}):
                    print(soux.get_text())
                    fp.write(soux.get_text())
                for sou4 in sou3.findAll("h3",{"class":"font14"}):
                    sou5=sou4.parent
                    print(sou5.get_text())
                    fp.write(sou5.get_text())
    for links in soup.findAll("a",href=True,text="Next"):
        html=urllib.request.urlopen("http://www.carwale.com"+links["href"]).read()
        soup = BeautifulSoup(html)
        s(soup)
def st(oup):
    for oup1 in oup.findAll("td",{"class":"ver-pdd"}):
        for oup3 in oup1.findAll("a"):
            pass
        for oup2 in oup1.findAll("span",{"class":"text-grey"}):
            if (oup2(text=re.compile('Diesel')))==[]:
                dict[oup3.get_text()]='Petrol'
            else:
                dict[oup3.get_text()]='Diesel'
    for oup1 in oup.findAll("tr",{"class":"model-row version-row"}):
```

```
        for oup3 in oup1.findAll("a",{"class":"href-grey"}):
            tml=urllib.request.urlopen("http://www.carwale.com"+oup3["href"]).read()
            oup1 = BeautifulSoup(tml)
            s(oup1)
    for ink in oup.findAll("a",href=True,text="Next"):
        tml=urllib.request.urlopen("http://www.carwale.com"+ink["href"]).read()
        oup = BeautifulSoup(tml)
        print(dict)
        st(oup)
        break
st(oup)
import re
fx = open("suraj.txt", 'r');
fy = open("sep1.txt",'w');
for line in fx:
    line = re.sub(r'[ , ( )]','\n',line);
    fy.write(line);
fx.close();
fy.close();
li=['good','bad','great'];
fv = open("file.txt", 'r');
fw = open("sep2.txt",'w');
count=0;
for line in fv:
    count += 1;
    for ch in li:
        print(li);
        matchobj = re.search(ch,line,re.M|re.I);
        if matchobj:
```

```
        print(matchobj.group());
        val=ch + " in Line no. "+str(count)+'\n';
        fw.write(val);
fv.close();
fw.close();
fp.close()
fd.close()
```

Snapshots





Milestones

I. Not too long ago, when potential customers wanted to know about the quality of your product or service, they relied on the opinions of friends and family. Word-of-mouth is still alive in today's connected world—it's just become exponentially amplified with the use of online reviews. One good (or bad) review of any business can echo through cyberspace indefinitely.so it will be going to help the consumers definitely.

II. Consumers increasingly look to reviewers when trying a new product, and reviews play a big role in the sales and overall awareness of a product.

Conclusion

The application has been designed, implemented and tested successfully. The project helped us in understanding the challenges involved in developing an online application, the ways to overcome them. It has also helped in understanding the value of designing the components of overall application before implementing them. In addition to that, it has also taught us programming skills and refining the design and implementation logic of the software at every phase of the development life cycle in order to improve the overall performance of the application. This has also taught us to work in a team. Lastly it has boosted our confidence that if we get a similar type of project in future, we will be able to contribute our knowledge and experience

Reference:

1.http://www.tutorialspoint.com/python

2.Learning Python_Mark Lutz, 5th Edition

3. PythonProgrammingfortheAbsoluteBeginner

4. Python_Essential_David Beazley

5. python_book_Dave Kuhlman