Lovely Professional University
Jalandhar, Punjab, India.

INT 404

PROJECT

TV SHOW POPULARITY ANALYSIS

REGISTRATION NO	NAME	ROLL NUMBER
11803304	Vadlamani Pavan Kumar	13
11808127	Karri Chandra Sekhar	08
11808096	Akshat Kumar Jain	07
11803455	Mohmad Aymen Naqash	19

Code of the project is available in the following link

GITHUB LINK:https://github.com/PavanVadlamani/int-404

.

PROJECT DESCRIPTION:

Through this project we built a system which calculates the polarity and subjectivity of the corresponding comments from the public on particular tv show to give a corresponding rating to the show.

For this to be implemented we need an online source from which we can access through the database of the source to collect the people comments.

We created an developers app from the developers site to get the API for that source, followed by the consumer key, consumer secret key, access token, and access secrete token while choosing the **TWITTER** as the source.

Next we installed the tweepy module to use the respective functions like API(), OAuthHandler(), etc....

After collecting the comments from a screen name, we cleaned the data using **re module**

The we stored the refined comments on the data frame and we applied sentiment analysis on each and every comment

Then found the respective polarities and subjectivities from sentiment analyzer using the **TEXTBLOB** library.

Then we draw a graph of positive, negative and neutral comments.

ABSTRACT

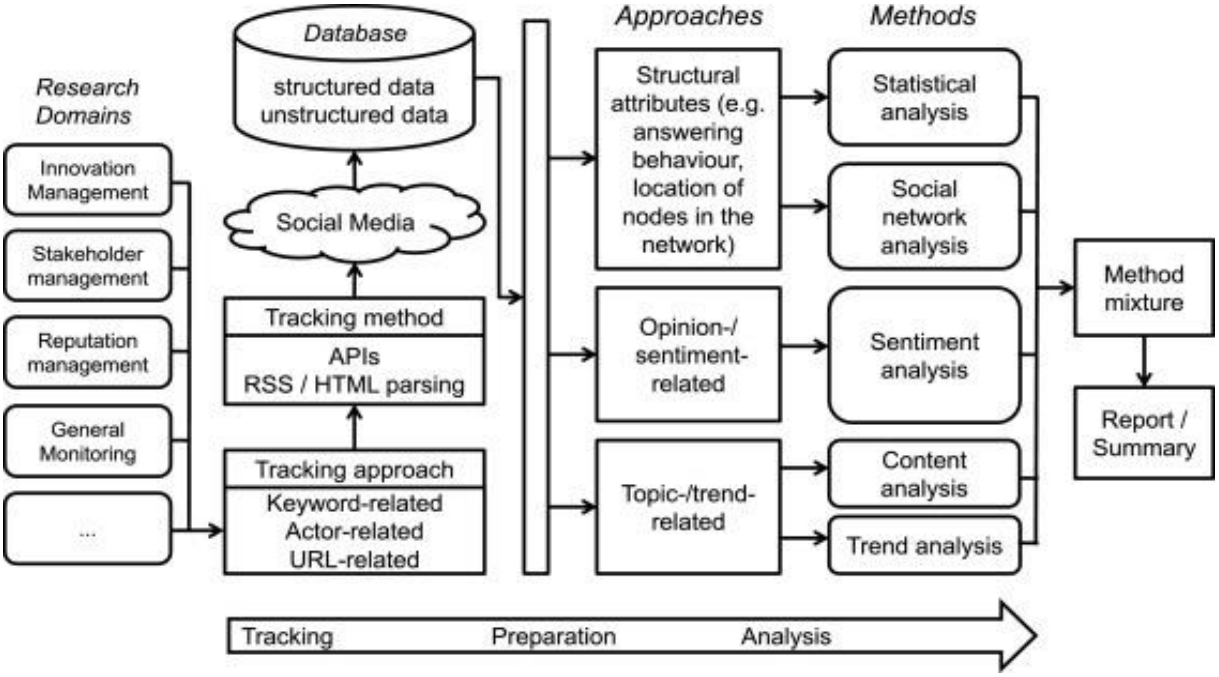
TV Show Popularity prediction using sentiment analysis is one of the most interesting and challenging tasks. Acritical demand along this line is to predict the popularity of online serials, which can enable a wide range of applications, such as online advertising, and serial recommendation. The problem motivation stated above suggest is that it is only the viewer of a program who is responsible for its popularity or failure and if we anyhow can identify the most common features of a program which, the viewers want most, and through some effective scientific methodology could insert these requirements in the proposed TV program well at the time of production.

Pre-processing is one of the important steps in text mining, Natural Language Processing (NLP) and information retrieval (IR). which gives tokenization, normalization. i.e.

remove @, remove #and URL. Data pre-processing is used to extract interesting and non-trivial knowledge from unstructured text data. Information Retrieval is important for deciding which documents in a collection should be retrieved so that we can satisfy a user's need for information.

SENTIMENT ANALYSIS

Sentiment analysis is another primary use case for NLP. Using sentiment analysis, data scientists can assess comments on social media to see how their business's brand is performing, for example, or review notes from customer service teams to identify areas where people want the business to perform better.



MODULES USED IN THE PROJECT:

TWEEPY MODULE:

Tweepy is a Python library for accessing the Twitter API. It is great for simple automation and creating twitter bots. Tweepy has many features.

In this tutorial we will be covering:

- Get tweets from our timeline.
- Creating and deleting Tweets.
- Follow and unfollow users.

TEXT BLOB MODULE:

Text Blob is a python library and offers a simple API to access its methods and perform basic NLP tasks.

A good thing about Text Blob is that they are just like python strings. So, you can transform and play with it same like we did in python. Below, I have shown you below some basic tasks. Don't worry about the syntax, it is just to give you an intuition about how much-related Text Blob is to Python strings.

Sentiment Analysis

Sentiment analysis is basically the process of determining the attitude or the emotion of the writer, i.e., whether it is positive or negative or neutral.

The sentiment function of text blob returns two properties, **polarity**, and **subjectivity**.

Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement. Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective refers to factual information. Subjectivity is also a float which lies in the range of [0,1].

WORD CLOUD MODULE:

Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points can be highlighted using a word cloud. Word clouds are widely used for analyzing data from social network websites.

Pandas Module

Pandas is an open source library in Python. It provides ready to use high-performance data structures and data analysis tools. Pandas module runs on top of NumPy and it is popularly used for data science and data analytics. NumPy is a low-level data structure that supports multi-dimensional arrays and a wide range of mathematical array operations. Pandas has a higher-level interface. It also provides streamlined alignment of tabular data and powerful time series functionality. Data Frame is the key data structure in Pandas. It allows us to store and manipulate tabular data as a 2-D data structure.

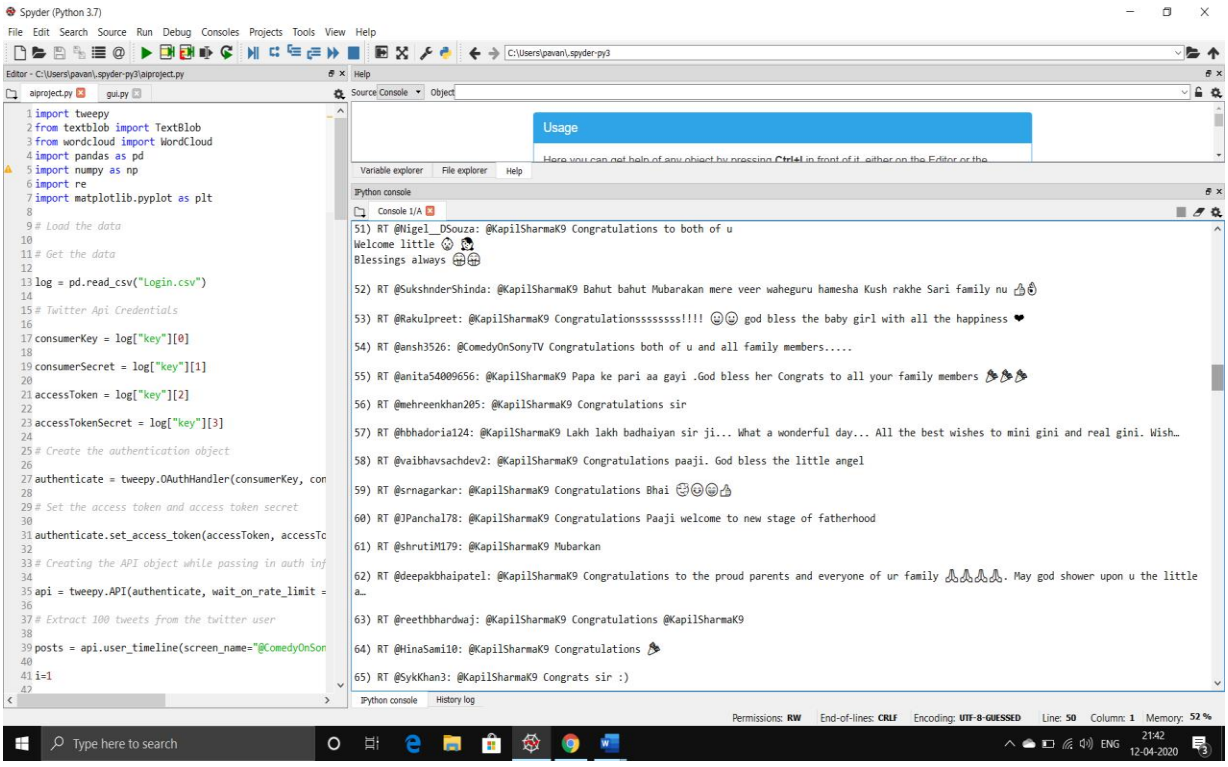
RE MODULE:

The Python module **re** provides full support for Perl-like regular expressions in Python. The re module raises the exception **re**. error if an error occurs while compiling or using a regular expression. We would cover two important functions, which would be used to handle regular expressions.

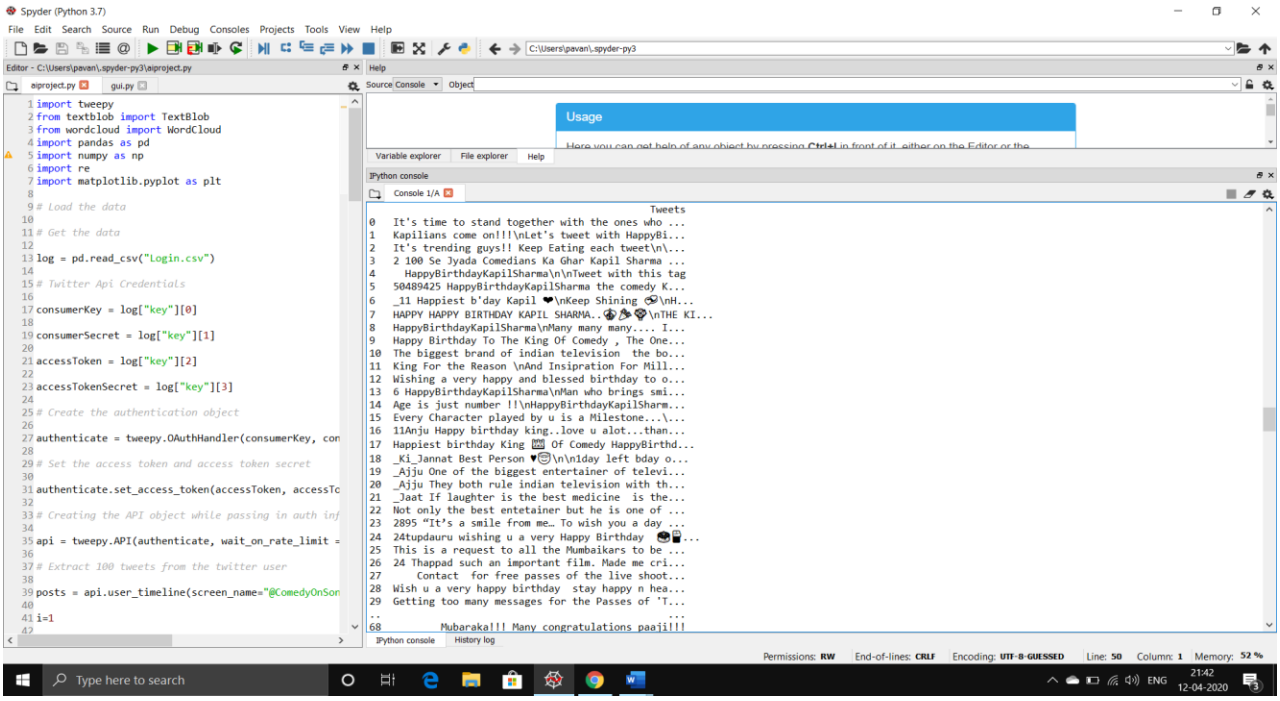
MODULE MATPLOTLIB:

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack.

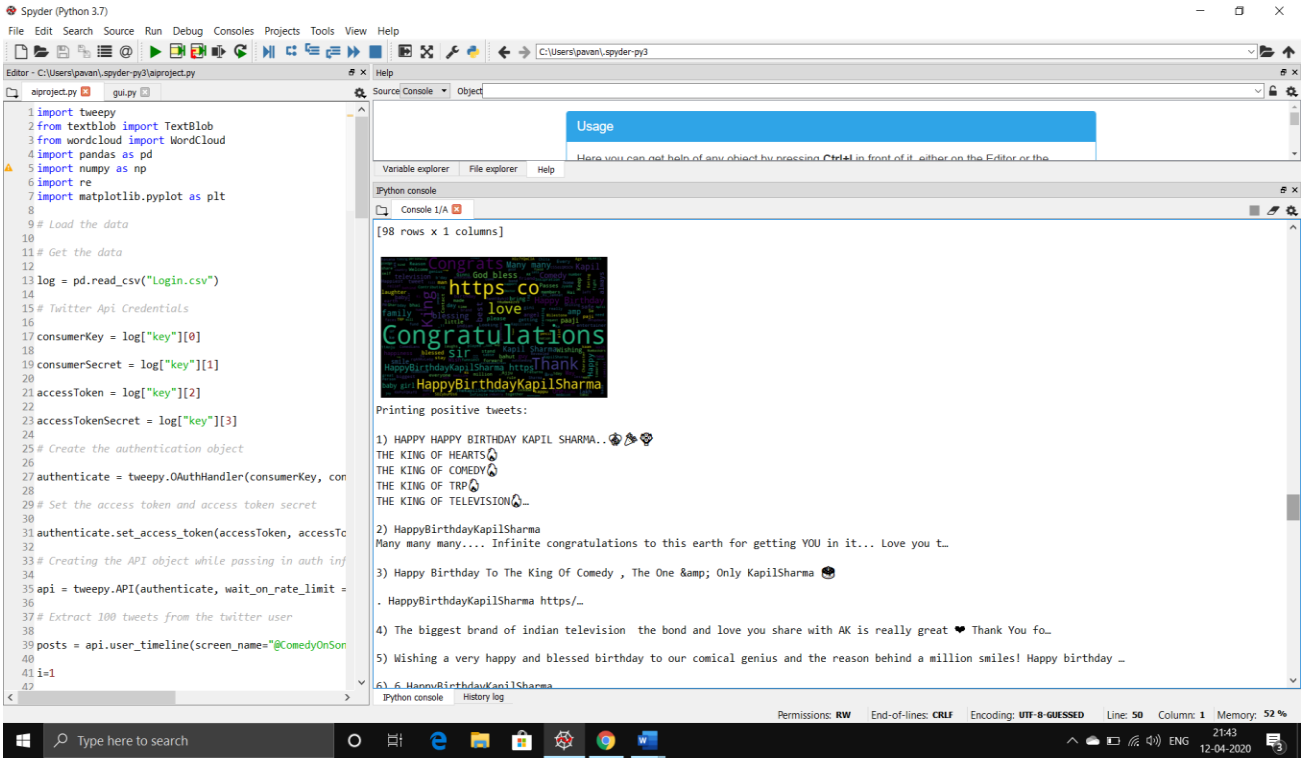
SOME SNAP SHOTS OF THE OUTPUT:



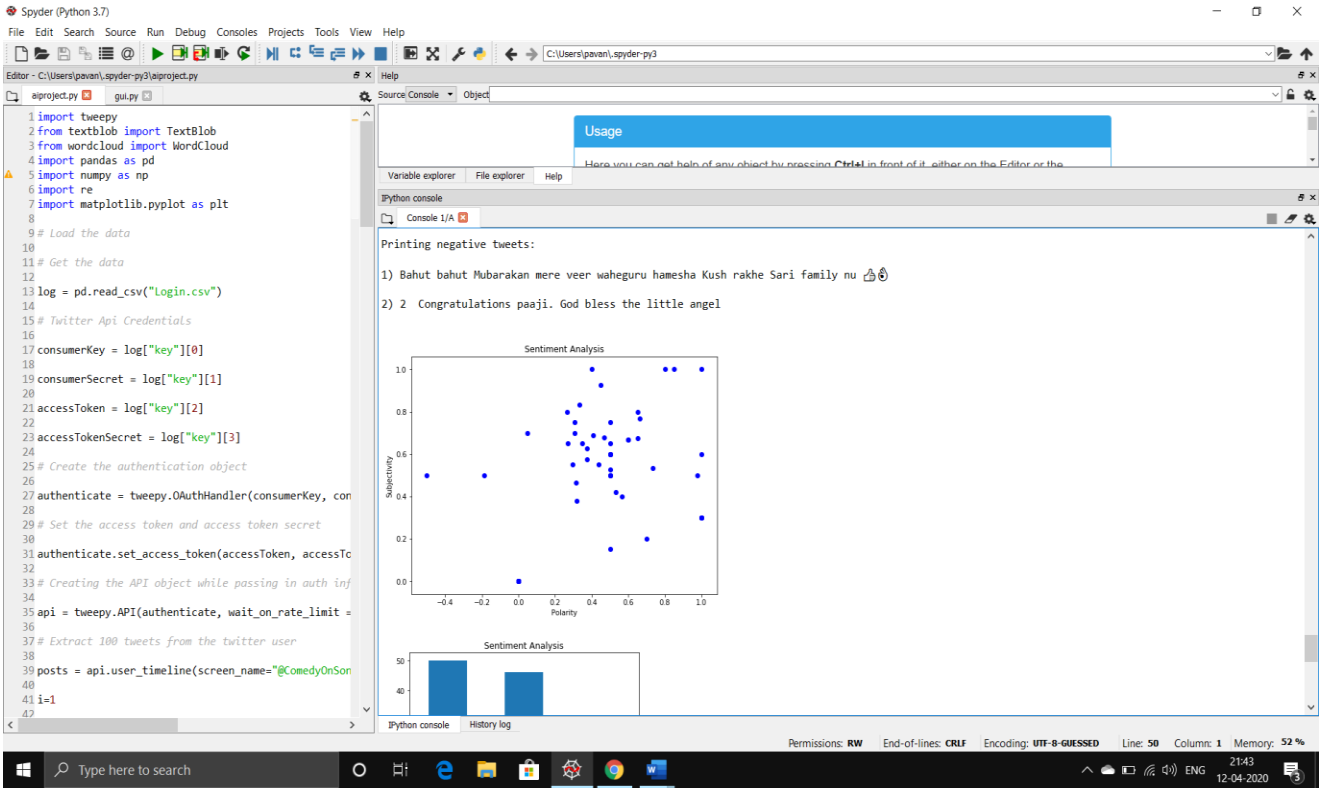
Collecting comments using api.



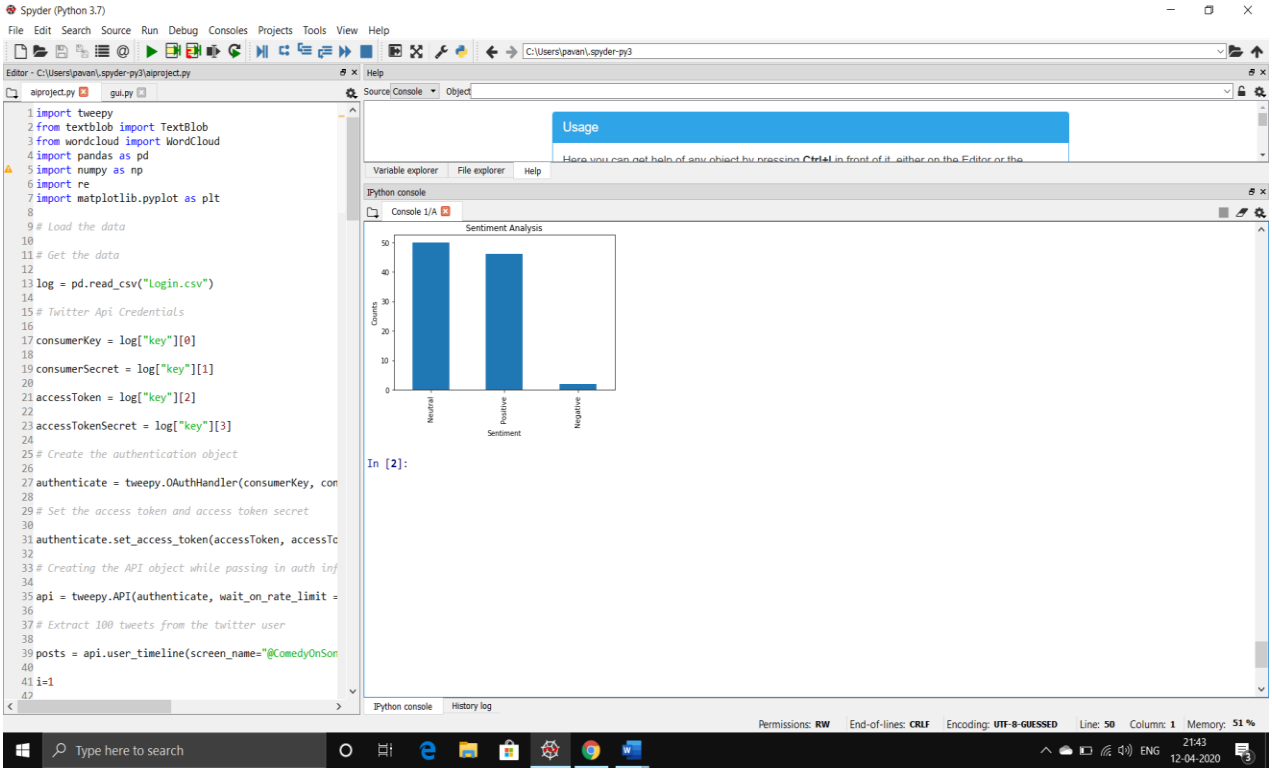
After cleaning the comments we can observe that no URLs , @ characters special symbols are there



This is a word cloud which shows the highly repeated words among all comments is the biggest one also we can observe all the positive tweets are collected



We can also observe the scatter graph of all the positive, negitive and neutral comments


Spyder (Python 3.7)
File Edit Search Source Run Debug Consoles Projects Tools View Help
Editor - C:\Users\pavan\.spyder-py3\aiproject.py
aiproject.py
gui.py
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
# Load the data
# Get the data
log = pd.read_csv("Login.csv")
# Twitter Api Credentials
consumerKey = log["key"][0]
consumerSecret = log["key"][1]
accessToken = log["key"][2]
accessTokenSecret = log["key"][3]
# Create the authentication object
authenticate = tweepy.OAuthHandler(consumerKey, con
# Set the access token and access token secret
authenticate.set_access_token(accessToken, accessTc
# Creating the API object while passing in auth inf
api = tweepy.API(authenticate, wait_on_rate_limit =
# Extract 100 tweets from the twitter user
posts = api.user_timeline(screen_name="@ComedyOnSon
i=1
Help
Usage
Variable explorer
File explorer
Help
IPython console
Console 1/A
Sentiment Analysis
Counts
Neutral
Positive
Negative
Sentiment
In [2]:
IPython console
History log
Permissions: RW
End-of-lines: CRLF
Encoding: UTF-8-GUESSED
Line: 50
Column: 1
Memory: 51 %
Type here to search
ENG
21:43
12-04-2020