Principles of Big Data Management
Computer Science 5540 0001

GitHub URL: https://github.com/rdbgm/PBD_Warm_Up.git

Aim: To retrieve the longest quotation mentioned in any news article, on which day (also by whom)? A quotation is a string with "…"

Dataset: New York Times Articles (https://www.kaggle.com/nzalake52/new-york-times-articles)

This project has undergone two methods in-order give a detailed explanation regarding properand improper data.

Stey-by-step procedure has been explained in the following documentation.

Method 1: When the given data is in appropriate condition

Article="URL: http://www.nytimes.com/2016/06/30/sports/baseball/washington-nationals-max- scherzer-baffles-mets-completing-a-sweep.html WASHINGTON — Stellar pitching kept the Mets afloat in the first half of last season despite their offensive woes. But they cannot produce an encore of their pennant-winning season if their lineup keeps floundering while their pitching is nicked, bruised and stretched thin. "We were going to ride our pitching," Manager Terry Collins said before Wednesday's game. "But we're not riding it right now. We've got as many problems with our pitching as we do anything." Wednesday's 4-2 loss to the Washington Nationals was cruel for the already-limping Mets. Pitching in Steven Matz's place, the spot starter Logan Verrett allowed two runs over five innings. But even that was too large a deficit for the Mets' lineup to overcome against Max Scherzer, the Nationals' starter. "We're not even giving ourselves chances," Collins said, adding later, "We just can't give our pitchers any room to work." The Metsdid not score until the ninth inning,.   "

'Article' is an object which stored the given data.

import   re   result=re.findall(r""""(.*?)"""",

Article)**Explanation:**

Importing the regular expressions(import re) package in-order to identify strings with double quotations.

dct = {i:len(i) for i in result}print(dct)

Explanation:

Dictioinary (dct) is created to store all the strings with double quotations with their respectivelengths.

Output:

`{'We were going to ride our pitching,': 35,

'But we're not riding it right now. We've got as many problems with our pitching as we do anything.': 98,

'We're not even giving ourselves chances,': 40, 'We just can't give our pitchers any room to work.': 49,

'I don't think we've played half our games yet this year,': 56,

'There's still a lot of things left that can and hopefully will happen.': 70,

'If they keep adding pressure on themselves, they're going to continue to struggle,': 82,

.............}

Keymax   =   max(dct,   key=dct.get)

Keymax

Explanation:

From the dictionary-dct, we are retrieving the largest string by using the max function and printing the largest string. Keymax stores the largest string.

Output:

'This issue is a significant part of what the entire debate regarding Puerto Rico is about: Billionaire hedge fund managers who purchase Puerto Rican bonds for pennies on the dollar now want a 100 percent return on their investment, while schools are being shut down in Puerto Rico, while pensions are being threatened with cuts, while children on the island go hungry. That is morally unacceptable'.

 dct2=Article.split("URL: ")

 dct2 = [x for x in dct2 if len(x) >0]   dct2

Explanation:

In order to find in which article the largest string is stored, we are splitting the articles by using the split function and using keyword as URL:

Here, dct2 is the list which stores the separated articles.

Output:

[' http://www.nytimes.com/2016/06/30/sports/baseball/washington-nationals-max-scherzer- baffles-mets-completing-a-sweep.html \nWASHINGTON — Stellar pitching kept the Mets afloat in the first half of last season despite their offensive woes. But they cannot produce an encore of their pennant-winning season if their lineup keeps floundering while their pitching is nicked, bruised and stretched thin.\n"We were going to ride our pitching," Manager Terry Collins said before Wednesday's game. "But we're not riding it right now. We've got as many problems with our pitching as we do anything."\nWednesday's 4-2 loss to the Washington Nationals was cruel for the already-limping Mets. Pitching in Steven Matz's place, the spot starter Logan Verrett allowed two runs over five innings. But even that was too large a deficit for the Mets' lineup to overcome against Max Scherzer, the Nationals' starter.\n"We're not even giving ourselves chances," Collins said, adding later, "We just can't give our pitchers any room to work."\nThe Mets did not score until the ninth inning, when a last-gasp two-run homer by James Loney off Nationals reliever Shawn Kelley snapped a streak of 23 scoreless innings for the team.\nThe Mets were swept in the

three-game series and fell six games behind the Nationals in the National League East. Of late, the Mets have looked worse than their 40-37 record.\n"I don't think we've played half our games yet this year," right fielder Curtis Granderson said. "There's still a lot of things left that can and hopefully will happen."\nScherzer toyed with the Mets, who were initially without Granderson after he was scratched from the lineup with lingering calf tightness. Even though Granderson has been inconsistent this season, he had hit well against Scherzer in the past. Alejandro De Aza, who entered the game with a .165 average, started in right field instead because Collins said the team had few options.\nAfter Scherzer gave up a single to Asdrubal Cabrera and walked Loney in the second inning, he retired the next 18 batters, until an eighth- inning single by Brandon Nimmo.\nThe Mets struggled again with runners on base. After Nimmo and the pinch-hitting Granderson singled in the eighth, pinch-hitter Travis d'Arnaud grounded out, and De Aza struck out.\n"If they keep adding pressure on themselves, they're going to continue to struggle," Collins said. "That's one of the things we try to make sure they have to understand: They have to be themselves."\nGeneral Manager Sandy Alderson, Collins and the coaching staff have met about the offense and discussed the odd dynamics: Some players are performing at or better than their career averages, but the lineup as a whole has struggled immensely, especially with runners in scoring position.\n"We're just not driving in any runs," Collins said. "That's been the frustrating part. It's not that we're striking out. We're popping up, or a double-play ball."\nThe Mets have a power-hitting team, so asking players to bunt or hit and run would go against their strengths.\n"When you start to change a team that's built one way and start to make them do something different, you're going to get your butt beat," Collins said.\nEarlier in the season, the Mets appeared like an all-or-nothing, home-run-driven team. Although they hit only .211 as a team in May, they smashed 40 home runs. They have a higher

average in June, but they have hit only 24 homers, and the inconsistent offense has put a strain on the pitching staff.\nIn the second inning, Verrett gave up a solo home run to the ex-Met Daniel Murphy. Collins wanted to limit the workloads of Addison Reed and Jeurys Familia, so he turned to reliever Sean Gilmartin in the eighth. Gilmartin gave up a two-run homer to Murphy, who has hit .429 (15 for 35) with four home runs against the Mets this season, his first since leaving the team.\n"I felt like I kept us in the game and gave us a chance to come back and win it," Verrett said. "I wish that I wouldn't have given up the two runs."\nVerrett was put in this position because of the effects of bone spurs on the Mets' rotation. The team asked Verrett to start Wednesday and gave Matz an extra day of rest after he received anti-inflammatory medication for the large bone spur in his left elbow. He will try to pitch through it.\nNoah Syndergaard has a smaller and less intrusive bone spur in the back of his right elbow.\n"As long as I'm staying on my anti-inflammatories and my mechanics are on point, I'm able to go out there every five days and compete," Syndergaard said.\nFor the Mets, the immediate road ahead will be even tougher. Matz was expected to pitch Thursday against the Chicago Cubs, one of the best teams in baseball this season.\n\n',
'http://www.nytimes.com/2016/06/30/nyregion/mayor-de-blasios-counsel-to-leave-next-month-to-lead-police-review-board.html\n\nMayor Bill de Blasio's counsel and chief legal adviser, Maya Wiley, is resigning next month from her City Hall position to become the chairwoman of the Civilian Complaint Review Board, New York City's independent oversight agency for the Police Department.\nThe move represents the latest shake-up for the de Blasio administration amid continuing state and federal investigations into the mayor's fund-raising, and fills a two-month vacancy at the police review board created by the resignation of its chairman, Richard D. Emery, in April.\nA civil rights lawyer and advocate for racial and social justice, Ms. Wiley joined the de Blasio administration in early 2014 to focus on legal issues as well as on the mayor's efforts to address issues of inequality. But over time, Ms. Wiley became discouraged over not being part of Mr. de Blasio's inner circle and felt cut out of both legal questions and advocacy, according to a person familiar with her thinking. On the former, Mr. de Blasio often relied instead on the city's corporation counsel and Henry Berger, the mayor's special counsel; on the latter, he favored his top political aides. The person requested anonymity to discuss private conversations.\nMore recently, Ms. Wiley was assigned to help craft the administration's legal response to the state and federal inquiries as well as to requests for the public disclosure of documents, notably emails between Mr. de Blasio and trusted advisers outside the administration.\nIt was in response to a question from reporters

about the withholding of those emails with advisers that Ms. Wiley, defending the practice, described the advisers as "agents of the city" — a designation that appeared novel and resulted in days of unfavorable press coverage.\nIn a statement on Wednesday, Mr. de Blasio thanked Ms. Wiley for her service and congratulated her on her new role.\nThe review board investigates allegations of misconduct by officers and makes recommendations for discipline to the Police Department. Its data on the number of complaints, and its investigations of officers, provide an important barometer of police behavior and a politically important one for Mr. de Blasio, a Democrat who campaigned on improving police- community relations.\nMr. Wiley will also take a position at the New School in Manhattan. Her moves were reported by The Wall Street Journal.\nThe announcement of Ms. Wiley's departure from City Hall followed that of a recently hired director of social media, Scott Kleinberg, who resigned on Saturday, just eight weeks after being hired to bring greater personality to the Twitter, Facebook and other online accounts associated with the mayor's office. His resignation was reported by DNA Info.\nIn a Facebook post that was later removed, Mr. Kleinberg complained of long hours and micromanagement and described his experience with the administration as working with "political hacks plus a boss who just couldn't get it," adding, "It was a bad combination for sure." Mr. Kleinberg declined to comment.\nThe departures came less than two months after Karen Hinton, the mayor's top spokeswoman, announced her resignation from the administration. (She stayed in the position until mid-June.)\n\n',

.........................]

 result2 = [i for i in dct2 if Keymax in i]print(result2)

Explanation

Using dct2 we will find the largest string and store it in the keymax and the article which consists of the big string will be assigned to the result2

Output:

[' http://www.nytimes.com/2016/06/30/us/politics/bernie-sanders.html \nWASHINGTON — The lusty applause that greeted his return to the Capitol is behind him now, as are the pecks on the cheek he received as he sat at his desk on the Senate floor, looking vaguely glum, receiving good wishes like a warrior returned to civilization, injured but intact.\nSenator Bernie Sanders of Vermont is back to giving floor speeches deriding the rich and defending those in misery, writing Op-Ed pieces against trade and giving television interviews during which he declines to fully support Hillary Clinton for president. He spent much of Wednesday vigorously denouncing a rescue bill for Puerto Rico that had the support of two-thirds of his fellow senators.\n"Let us be clear," Mr. Sanders said on the Senate floor Wednesday, reviving his familiar Brooklyn-inflected pedagogy. "This issue is a significant part of what the entire debate regarding Puerto Rico is about: Billionaire hedge fund managers who purchase Puerto Rican bonds for pennies on the dollar now want a 100 percent return on their investment, while schools are being shut down in Puerto Rico, while pensions are being threatened with cuts, while children on the island go hungry. That is morally unacceptable."\nMr. Sanders, who has not withdrawn from the presidential race, has not really left the Democratic primary battlefield. Apparently defeated but decidedly unbowed, he has brought his campaign to Capitol Hill, most visibly in the large security detail that surrounds him as he moves about ("He is very secure," noted Senator Tim Scott, Republican of South Carolina) but also in his comportment with his colleagues.\nEager to expand the left-leaning coalition he has built during his campaign, Mr. Sanders has been pushing his colleagues to take on policy fights that helped propel his base's passion and gave him new gravitas among Democrats.\nIn addition to opposing the measure to aid Puerto Rico, he is working hard to kill trade agreements. He is threatening a bill that would govern the labeling of genetically modified food.\n"Everyone knows the fervency of his opinions on different things," said Senator Harry Reid, the minority leader, who, like most senators, was eager to get on with business and out of town for the Fourth of July.\nThe Democrats with whom the independent Mr. Sanders caucuses have been tolerant of his not-quite-campaign, in no small part because they do not wish

to emulate Republicans, whose wounds have been oozing openly. But many Democratic colleagues, especially the women, are growing weary of his progressive lectures that seem more fit for a dais than a lunchroom encounter, and his unwillingness to energetically back Mrs. Clinton.\n"He feels he has a duty to his followers to raise the flag on the issues they care about," said Senator Claire McCaskill, Democrat of Missouri. "We are all being patient, and we're all hopeful he'll be on the campaign" for Mrs. Clinton.\nFor now, Mr. Sanders seems to be adjusting — if slightly mournfully — at the fork in the road between kind-of former candidate and definitely current senator. He walks through the halls at times emulating Senator Elizabeth Warren of Massachusetts, whose shoulder he gave a squeeze as he bounded through the basement of the Capitol on Wednesday, brushing off reporters with a wave. At other times he becomes chatty again, talking about legislation he despises. "I'm sure it's really hard on him," said Senator Barbara Boxer, Democrat of California. "Losing is awful."\nHe has his eyes on the senior Democratic slot on the Senate's high-profile Committee on Health, Education, Labor and Pensions — as of late very bipartisan — to the dismay of Republicans, who prefer Senator Patty Murray of Washington, the senior Democrat on the committee, and even some Democrats, who fret that Mr. Sanders might try to push the committee to the left on several issues.\n"If the opportunity were to arise," said Michael Briggs, a spokesman for Mr. Sanders, "he would be proud to chair that committee, which deals with so many issues of vital importance to the American people." The decision is Ms. Murray's to make.\nOthers said they can already see Mr. Sanders's influence on Democrats. "He has made a big splash," Mr. Scott said. "So his influence can be seen on Hillary Clinton's campaign, and over the last few months up here, especially on the banking committee, he has moved people to the left."\nMr. Sanders clearly wants to work with Democrats to put together legislation that reflects his priorities — like affordable college — while still working to get much of his agenda in the platform of the party to which he does not belong for its convention next month.\n"As he has said many times, however, he does not believe real change is going to take place until a political revolution occurs and millions of people stand up and fight for their rights," Mr. Briggs said.\nFor now, Mr. Sanders's campaign has given him newfound clout among Democrats. But he risks losing some if he does not endorse Mrs. Clinton soon, and he is still working to figure out how to spend his capital in a caucus where he lacks many close friends.\n"Bernie is going to do what Bernie's going to do," said Senator Dianne Feinstein, Democrat of California, "and the only thing I think is that it should be something constructive and helpful to the ticket."\nSenator Chuck Schumer, Democrat of New York, recognizes that Mr. Sanders's appeal is valuable to the party and will try to harness it while attempting to convince him that uniting around Mrs. Clinton is the best way to defeat Donald J. Trump in November. "In our caucus," Mr. Schumer said, "the general view is that Bernie is a constructive force."\n']

import re

match = re.search(r'(\d+/\d+/\d+)',result2[0])   print("Date Posted",match.group(1))

Explanation:

In order to find on which date the article got published, we used regular expressions (\d+/\d+/\d+) and group() is a generator function which is used to retrieve the required date.

Output:

Date Posted 2016/06/30

m   =   re.match(r'(\w.+-\w+)',result2[0])

print(m.group(1))

Explanation

In order to obtain URL we have used regular expression (\w.+-\w+) and found the author'sname.

Output:

http://www.nytimes.com/2016/06/30/us/politics/bernie-sanders

 from urllib.parse import urlparse

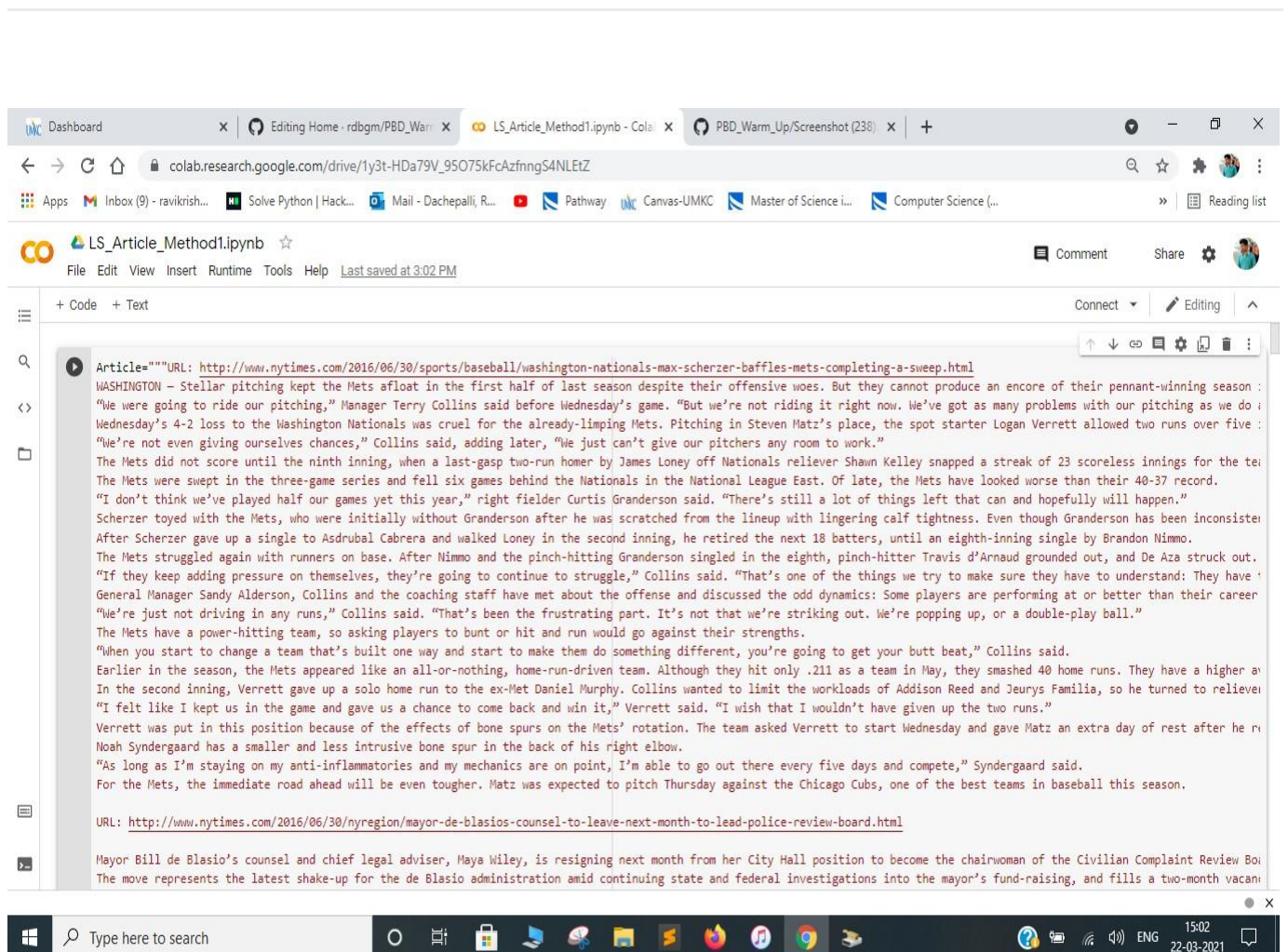 url = m.group(1)   p = urlparse(url)   print(p.path.split('/')[6]

Explanation

To find author's name we have imported urlparse package from urllib.parse. In order to split thetokens in the URL we have used urlparse. The index 6 refers to the bernie sanders from the link http://www.nytimes.com/2016/06/30/us/politics/bernie-sanders

Output:

bernie-sanders

Screenshots of Method-1:

co LS_Article_Method1.ipynb ☆

File Edit View Insert Runtime Tools Help Last saved at 3:02 PM

Comment Share

+ Code + Text

Connect ▼ Editing

```python
        import re
result=re.findall(r""""(.*?)"""", Article)
```

```python
dct = {i:len(i) for i in result}
print(dct)
```

{'We were going to ride our pitching,': 35, 'But we're not riding it right now. We've got as many problems with our pitching as we do anything.': 98, 'We're not even giving ourselves

```python
Keymax = max(dct, key=dct.get)
```

▶ Keymax

'This issue is a significant part of what the entire debate regarding Puerto Rico is about: Billionaire hedge fund managers who purchase Puerto Rican bonds for pennies on the dollar now want a 100 percent return on their investment, while schools are being shut down in Puerto Rico, while pensions are being threatened with cuts, while children on the island go h ungry. That is morally unacceptable.'

```python
dct2=Article.split("URL: ")
dct2 = [x for x in dct2 if len(x) >0]
```

Second screenshot:

```python
dct2=Article.split("URL: ")
dct2 = [x for x in dct2 if len(x) >0]
dct2
```

```
['http://www.nytimes.com/2016/06/30/sports/baseball/washington-nationals-max-scherzer-baffles-mets-completing-a-sweep.html \nWASHINGTON — Stellar pitching kept the Mets afloat in the
'http://www.nytimes.com/2016/06/30/nyregion/mayor-de-blasios-counsel-to-leave-next-month-to-lead-police-review-board.html\n\nMayor Bill de Blasio's counsel and chief legal adviser,
'http://www.nytimes.com/2016/06/30/nyregion/three-men-charged-in-killing-of-cuomo-administration-lawyer.html\n\nIn the early morning hours of Labor Day last year, a group of gunmen
'http://www.nytimes.com/2016/06/30/nyregion/tekserve-precursor-to-the-apple-store-to-close-after-29-years.html\n\nIt was the Apple Store in New York City before there was such a thi
'http://www.nytimes.com/2016/06/30/sports/olympics/once-at-michael-phelpss-feet-and-still-chasing-them.html\n\nOMAHA — The United States Olympic swimming trials are the spectacle th
'http://www.nytimes.com/2016/06/30/sports/olympics/missy-franklin-breaks-through-in-trials-and-earns-a-return-to-olympics.html\n\nOMAHA — In the first three races of her third Olymp
'http://www.nytimes.com/2016/06/30/business/dealbook/lionsgate-is-said-to-be-near-deal-to-buy-starz.html\n\nLionsgate is near a deal to buy Starz, uniting the film studio behind "Th
'http://www.nytimes.com/2016/06/30/nyregion/pool-rules-no-running-no-eating-or-drinking-no-men.html\n\nUnder slate-colored light slanting from the skylights, the women entered the c
'http://www.nytimes.com/2016/06/30/sports/basketball/knicks-look-to-young-blood-and-free-agency-to-patch-porous-roster.html\n\nWINTER GARDEN, Fla. — Jeff Hornacek, the new Knicks co
'http://www.nytimes.com/2016/06/30/nyregion/latest-sign-of-change-in-harlem-its-congressional-candidate.html\n\nAs Washington Heights rejoiced on Wednesday over the apparent victory
'http://www.nytimes.com/2016/06/30/world/africa/a-slow-steady-siege-on-isis-stronghold-in-libya.html\n\nSURT, Libya — Perched on a doorstep, the teenage Juma brothers whiled away th
'http://www.nytimes.com/2016/06/30/us/politics/house-democrats-try-to-sustain-push-for-gun-control.html\n\nFARMINGTON, Conn. — As a pastor, Sam Saylor knows how draining the fight a
'http://www.nytimes.com/2016/06/30/nyregion/operative-tied-to-cuomo-is-accused-of-bribing-judge-to-get-favorable-rulings.html\n\nA political operative with ties to Gov. Andrew M. Cu
'http://www.nytimes.com/2016/06/30/world/europe/father-killed-in-turkey-attacks-was-trying-to-save-his-son-from-isis.html\n\nsearching-for-isis-linked-son.html\n\nPARIS — Fathi Bayoudh was a
'http://www.nytimes.com/2016/06/30/technology/kleiner-perkins-raises-1-4-billion-with-two-funds.html\n\nSAN FRANCISCO — With more start-ups and venture firms working harder to raise
'http://www.nytimes.com/2016/06/30/us/politics/huma-abedin-hillary-clinton-emails.html\n\nHuma Abedin, Hillary Clinton's longtime aide and confidante, acknowledged that Mrs. Clinton
'http://www.nytimes.com/2016/06/30/world/middleeast/turkey-a-conduit-for-fighters-joining-isis-begins-to-feel-its-wrath.html\n\nPARIS — When the bodies of Islamic State fighters are
'http://www.nytimes.com/2016/06/30/nyregion/brooklyn-schools-supporters-say-the-city-bet-against-its-progress.html\n\nWhen the New York City Education Department put a new Success A
'http://www.nytimes.com/2016/06/30/us/politics/bernie-sanders.html \nWASHINGTON — The lusty applause that greeted his return to the Capitol is behind him now, as are the pecks on th
```

```python
result2 = [i for i in dct2 if Keymax in i]
print(result2)
```

```
['http://www.nytimes.com/2016/06/30/us/politics/bernie-sanders.html \nWASHINGTON — The lusty applause that greeted his return to the Capitol is behind him now, as are the pecks on th
```

```
[ ]  import re
     match = re.search(r'(\d+/\d+/\d+)',result2[0])

[ ]  print("Date Posted",match.group(1))

     Date Posted 2016/06/30

     m = re.match(r'(\w.+-\w+)',result2[0])
     print(m.group(1))

     http://www.nytimes.com/2016/06/30/us/politics/bernie-sanders

[ ]  from urllib.parse import urlparse
     url = m.group(1)
     p = urlparse(url)
     print(p.path.split('/')[6])

     bernie-sanders
```
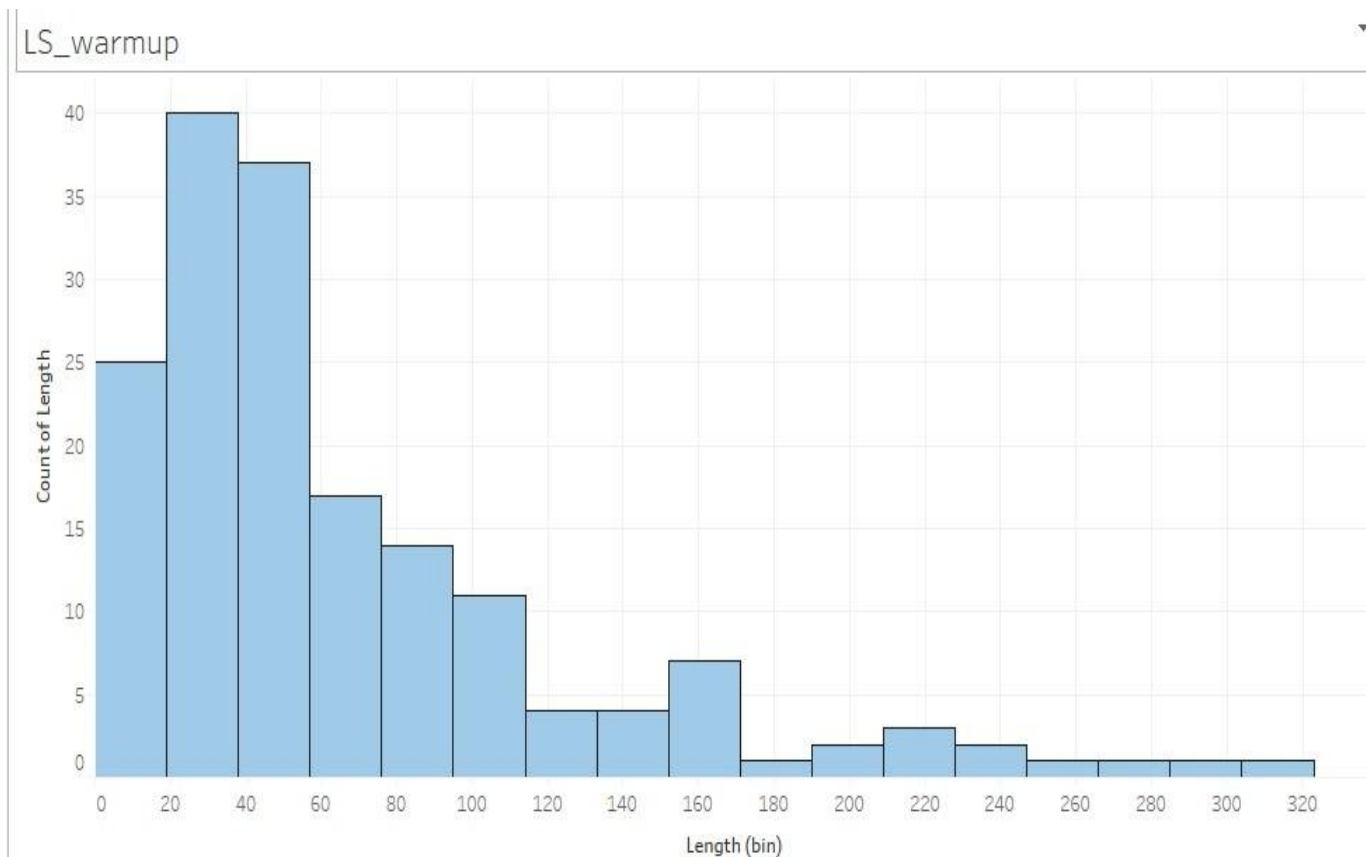

This is a bar chart that shows the length of the string on the x-axis and the count of lengths on the y-axis. Used tableau to build this chart.

Method 2: When the given data is not inappropriate condition.

import re

Explanation:

Importing the required Regular Expressions

 import findsparkfindspark.init()

Explanation:

Importing the findspark and initializing it

 from pyspark.sql import SparkSession

Explanation:

from pyspark.sql importing the Sparksession

 spark = SparkSession.builder \.master("local[*]") \.appName('warmupproject')
\.getOrCreate()

Explanation:

The application name is given as warmupproject and created the sparksession.

 source = spark.sparkContext

Explanation:

Assigning the spark context to the source

 data = source.textFile("E:/UMKC Materials/PBD/WarmUp/yt.txt")

Explanation:

Data from the text file is assigned to the object 'data'.

 data.collect()

Explanation:

To print the data loaded.

 a = ''

 for x in data.collect(): a += x   b=a.replace(''','"')   result=re.findall(r"""""(.*?)"""""",b)

Explanation:

As the given data is irregular, we need some pre-processing. So, we replaced the single quotation marks with the double quotation and findall() is used for identifying the strings with the double quotations and the output is stored in the variable 'result'

```
dct = {i:len(i) for i in result}dct
```

Explanation:

Dictionary (dct) is used to store all the strings with double quotations with their respectivelengths.

Output:

{'We were going to ride our pitching,': 35,

'But we're not riding it right now. We've got as many problems with our pitching as we doanything.': 98,

'We're not even giving ourselves chances,': 40,

'We just can't give our pitchers any room to work.': 49,

'I don't think we've played half our games yet this year,': 56,

'There's still a lot of things left that can and hopefully will happen.': 70,

'If they keep adding pressure on themselves, they're going to continue to struggle,': 82,

'That's one of the things we try to make sure they have to understand: They have to be themselves.': 97,

'We're just not driving in any runs,': 35,

'That's been the frustrating part. It's not that we're striking out. We're popping up, or a double-play ball.': 108,

'When you start to change a team that's built one way and start to make them do somethingdifferent, you're going to get your butt beat,': 135,

'I felt like I kept us in the game and gave us a chance to come back and win it,': 79,'I wish

that I wouldn't have given up the two runs.': 50,

'As long as I'm staying on my anti-inflammatories and my mechanics are on point, I'm able to goout there every five days and compete,': 133,

'agents of the city': 18,
'political hacks plus a boss who just couldn't get it,': 53, ...}

```
Keymax = max(dct, key=dct.get)

Keymax
```

Explanation:

From dictionary(dct) we are retrieving the largest string by using max() function and we are printing largest string (Keymax).

Output:

'and I am proud to stand up for this club and give us another chance together.◆ URL:http://www.nytimes.com/2016/04/21/business/dealbook/yahoos-first-quarter-adds-urgency-to- sale-option.htmlYahoo's first quarter is the last straw. An 11 percent drop in revenue for the first three months of the year puts an exclamation point on the lack of progress in almost four years with Marissa Mayer as chief executive. Google and Facebook are cleaning up in digital advertising, especially mobile. It's time to offload Yahoo's core business before bidders like Verizon lose interest.Ms. Mayer's enthusiasm is unflagging. Even after the poor showing reported on Tuesday, which included an operating loss of $167 million, she keeps talking about her ability to kick-start growth while considering other alternatives. It's hard to see what options are left.Yahoo's so-called mavens are moving at an analog pace. This Mayer-named collection of mobile, social, native and video ads represent a little more than a third of Yahoo's top line and grew only 7 percent in the first quarter. By comparison, Facebook's mobile-ad revenue increased 81 percent in the fourth quarter from a year earlier and accounted for nearly the entire $5.8 billion generated.Most companies put a gloss on themselves when they are up for sale to woo suitors. Yahoo, a $35 billion company that has lurched from strategy to strategy, is struggling to manage even that.

Ms. Mayer is making potential bidders listen to prerecorded messages while providing little financial detail, according to media reports. It suggests that there may be something messy inside or that the chief executive isn't especially keen to sell.Despite having to press for the pitch book, suitors are showing interest nonetheless. The telecom titan Verizon Communications signaled it would bid in an attempt to marry Yahoo with AOL, which it acquired last year for $4 billion. The private-equity firm TPG Capital and the digital-advertising remains of the phone- book publisher Yellow Pages also may be in the mix.Some potential bidders have already hung up on Yahoo. Comcast, AT&T, Time and Barry Diller's Internet hodgepodge IAC are among those that media reports say considered making offers but did not.Yahoo expects the second quarter to be even worse, with at least a 12 percent decline in sales from a year ago. Given that outlook, Ms. Mayer has only one last job to do at Yahoo. She may be running short on time, however, to get this one right. URL:http://www.nytimes.com/2016/04/21/technology/personaltech/apps-to-build-your- understanding-of-the-environment.html EARTH DAY, which takes place on Friday, has become an increasingly well-known event as politicians, billionaires and others take part in building awareness about the environment. You, too, can participate in Earth Day with apps that remind you how to add a touch of green to your life.GoodGuide is particularly helpful in that regard. The idea behind the app is that instead of being a slave to advertising while shopping, people can make informed choices. GoodGuide has a database of more than 200,000 products sold in the United States, including food and skin care products. It details how much particular goods affect the environment and your health, and whether the products are energy-efficient.GoodGuide's clear graphics and simple interface make it easy to search for a particular product. There is even a built-in bar code scanner to find data on various items, which can be handy in a supermarket. The app is free on iOS and Android.It's sometimes easy to forget that we share our planet with other animals, many of whom are threatened by humankind's changes to their habitats and lives. This is something the World Wildlife Fund's WWF Together can tell you about.Using photos, animations and interactive graphics, the app lays out the stories of endangered animals around the globe. The beautiful images speak for themselves in many cases, but the program also has plenty of data — for example, it points out that efforts to protect the endangered giant panda in China have helped increase the number of wild pandas over the last decade. WWF Together also has 360-degree photos of habitats, as well as educational games.A lot of the content is free, but unlocking all the available material costs $2. WWF Together is available on iOS and Android.#Climate is a different sort of environmental awareness app that lets you take part in activism from the comfort of your sofa or office desk. You can customize the app's settings to be told about the kind of environmental issues that matter to you, be they topic-

specific, regional or global. It then pulls together relevant "actions'

```
dct2=a.split("URL: ")
```

```
dct2 = [x for x in dct2 if len(x) >0]   dct2
```

Explanation:

In order to find in which article the largest string stored, splitting the data with split function URL as token.

Output:

['http://www.nytimes.com/2016/06/30/sports/baseball/washington-nationals-max-scherzer-   baffles-mets-completing-a-sweep.htmlWASHINGTON — Stellar pitching kept the Mets afloat in the first half of last season despite their offensive woes. But they cannot produce an encore of their pennant-winning season if their lineup keeps floundering while their pitching is nicked, bruised and stretched thin."We were going to ride our pitching," Manager Terry Collins said before Wednesday's game. "But we're not riding it right now. We've got as many problems with our pitching as we do anything."Wednesday's 4-2 loss to the Washington Nationals was cruel for the already-limping Mets. Pitching in Steven Matz's place, the spot starter Logan Verrett allowed two runs over five innings. But even that was too large a deficit for the Mets' lineup to overcome against Max Scherzer, the Nationals' starter."We're not even giving ourselves chances," Collins said, adding later, "We just can't give our pitchers any room to work."The Mets did not score untilthe ninth inning, when a last-gasp two-run homer by James Loney off Nationals reliever Shawn Kelley snapped a streak of 23 scoreless innings for the team.The Mets were swept in the three- game series and fell six games behind the Nationals in the National League East. Of late, the Mets have looked worse than their 40-37 record."I don't think we've played half our games yet this year," right fielder Curtis Granderson said. "There's still a lot of things left that can and hopefully will happen."Scherzer toyed with the Mets, who were initially without Granderson after he was scratched from the lineup with lingering calf tightness. Even though Granderson has beeninconsistent this season, he had hit well against Scherzer in the past. Alejandro De Aza, who entered the game with a .165 average, started in right field instead because Collins said the teamhad few options.After Scherzer gave up a single to Asdrubal Cabrera and walked Loney in the second inning, he retired the next 18 batters, until an eighth-inning single by Brandon Nimmo.The Mets struggled again with runners on base. After Nimmo and the pinch-hitting Granderson singled in the eighth, pinch-hitter Travis d'Arnaud grounded out, and De Aza struck out."If they keep adding pressure on themselves, they're going to continue to struggle," Collins said. "That's one of the things we try to make sure they have to understand: They have to be themselves."General Manager Sandy Alderson, Collins and the coaching staff have met about the offense and discussed the odd dynamics: Some players are performing at or better than their career averages...

```
result2 = [i for i in dct2 if Keymax in I]result2
```
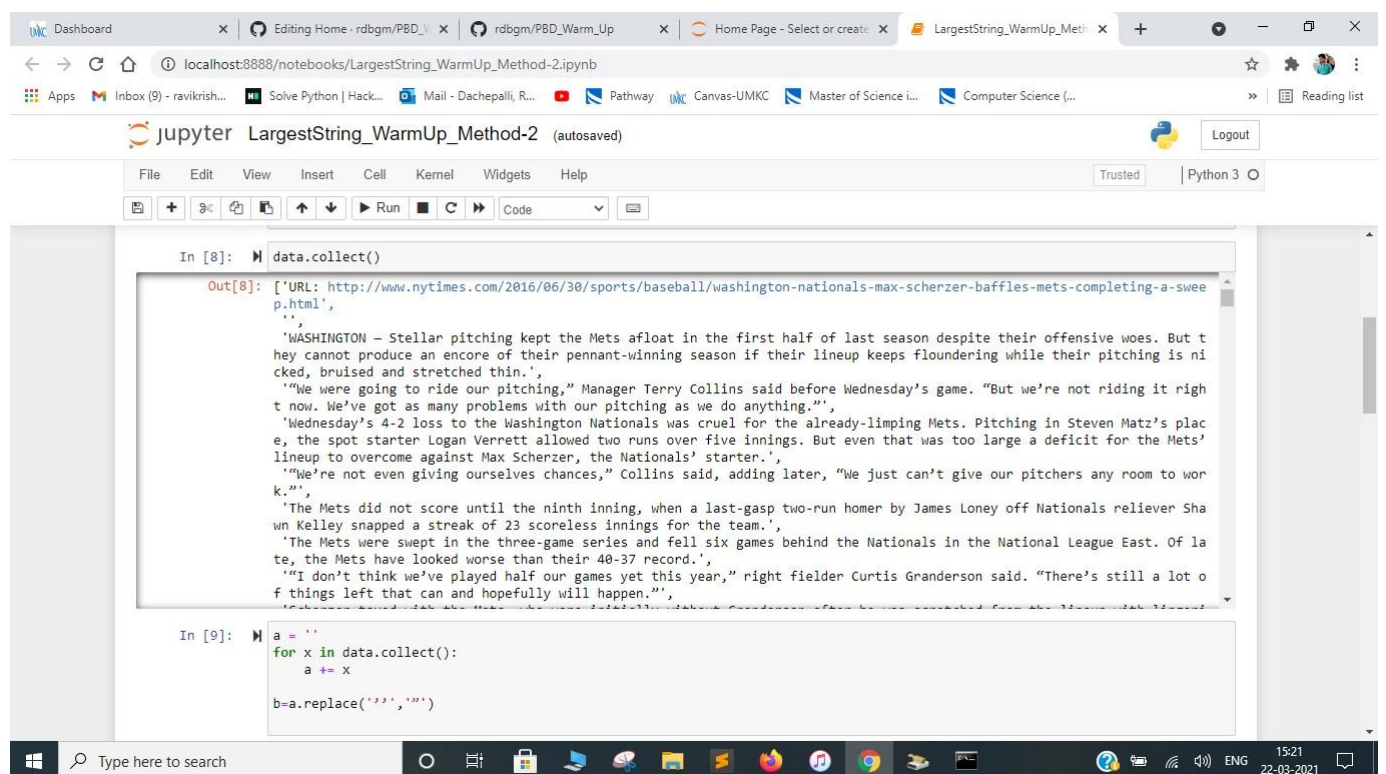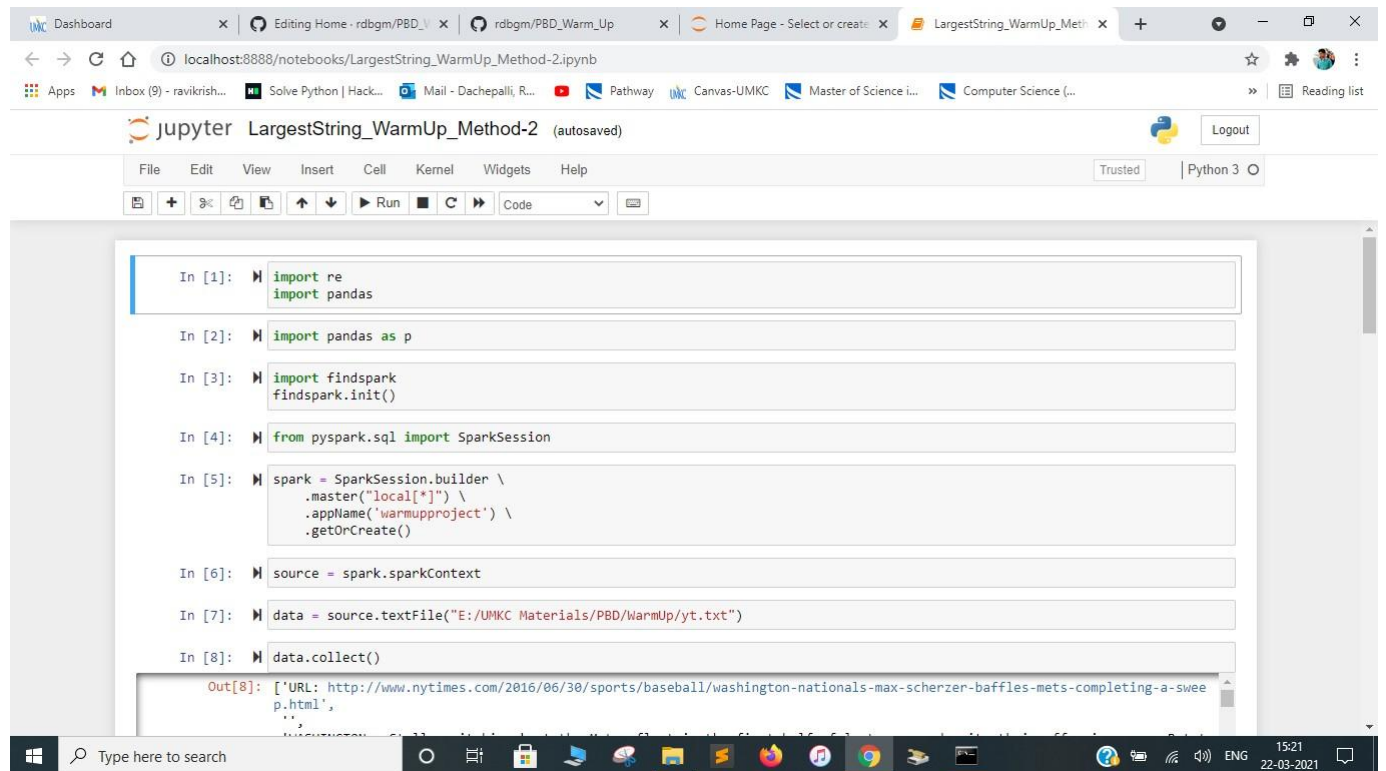
Explanation:

Here, Largest string is a combination of 2 or 3 articles. Therefore, it is unable to pick this largest string from the list where it is ending up with ambiguity and giving output as empty list.This is because of improper close and open of double quotations.

Output:

`[]`

Screenshots for Method-2

Jupyter LargestString_WarmUp_Method-2 (autosaved)   Logout

File   Edit   View   Insert   Cell   Kernel   Widgets   Help   Trusted | Python 3 ○

```
In [10]:  ▶ result=re.findall(r"""(.*?)""", b)
```

```
In [11]:  ▶ dct = {i:len(i) for i in result}
```

```
In [12]:  ▶ dct
```

```
Out[12]:  {'We were going to ride our pitching,': 35,
           'But we're not riding it right now. We've got as many problems with our pitching as we do anything.': 98,
           'We're not even giving ourselves chances,': 40,
           'We just can't give our pitchers any room to work.': 49,
           'I don't think we've played half our games yet this year,': 56,
           'There's still a lot of things left that can and hopefully will happen.': 70,
           'If they keep adding pressure on themselves, they're going to continue to struggle,': 82,
           'That's one of the things we try to make sure they have to understand: They have to be themselves.': 97,
           'We're just not driving in any runs,': 35,
           'That's been the frustrating part. It's not that we're striking out. We're popping up, or a double-play ball.': 108,
           'When you start to change a team that's built one way and start to make them do something different, you're going to get
           your butt beat,': 135,
           'I felt like I kept us in the game and gave us a chance to come back and win it,': 79,
           'I wish that I wouldn't have given up the two runs.': 50,
           'As long as I'm staying on my anti-inflammatories and my mechanics are on point, I'm able to go out there every five day
           s and compete,': 133,
           'agents of the city': 18,
           'political hacks plus a boss who just couldn't get it,': 53,
           'It was a bad combination for sure.': 34,
```

```
In [13]:  ▶ Keymax = max(dct, key=dct.get)
```

Jupyter LargestString_WarmUp_Method-2 (autosaved)   Logout

File   Edit   View   Insert   Cell   Kernel   Widgets   Help   Trusted | Python 3 ○

```
In [14]:  ▶ Keymax
```

```
Out[14]:  'We're all so inundated all the time. We need a place to turn everything off. "The ultimate luxury, he suggested, is not a
          $49.5 million townhouse like this one, it's a good night's sleep.The Kips Bay Decorator Show House is open through June 9 at
          19 East 61st Street; $35, 212-755-5733, kipsbaydecoratorshowhouse.org.URL: http://www.nytimes.com/2016/05/12/technology/pers
          onaltech/running-and-jumping-through-various-virtual-gantlets.htmlGRAPHICS-HEAVY, three-dimensional video games can be overw
          helming as you jump, hit and shoot through various adventures. But a calmer experience can be had in beautiful, stylish game
          s that could be called one dimensional — some just need a single finger for control.Dragon Hills is a fabulous new game in t
          his category. It's an auto-runner, which means your character — a purple-haired, sword-swinging warrior atop a dragon — auto
          matically runs forward through the levels, leaving you in charge only when you jump.The control is simple: Your dragon slith
          ers along the ground, and buries itself in the soil when you hold the screen with one finger. Then it leaps out into the air
          when you let go. Its teeth munch when you tap the screen, helping you take bites out of enemies and obstacles and gather awa
          rds and tokens as you zip along.Dragon Hills' graphics are cartoonlike and colorful, and the way the dragon, terrain and oth
          er characters are animated is eye-pleasing and fun. The play itself feels fluid and easy, but as you advance you have to lea
          rn to burrow under lava puddles and leap chasms to avoid arrows being fired at you by knights.Dragon Hills is equal parts fu
          nny, enjoyable and easy. It is also challenging as you increase the strength of your dragon by adding armor and facing trick
          ier obstacles. Dragon Hills is $2 on iOS and free on Android.Chameleon Run, also new, has a similar setup, with the characte
          r running in a straight line through a series of levels. You must also jump over gaps, leap over obstacles and collect item
          s. But the graphics in Chameleon Run — colorful geometric blocks and shapes that soar through the air — are even more minima
          l than those in Dragon Hills.This game features two controls to manage. As well as using one finger to jump (the longer you
          hold, the higher and farther you jump) you need another finger to change the color of your character to match the color of t
          he block you're going to land on. If the colors do not match, or if you fall into the void because you miss a landing, you e
          xplode and the game is over.It's fast, gorgeously animated and looks deceptively easy. But don't let this apparent simplicit
          y fool you, since it's challenging to get through higher levels. The game is $2 to download on iOS and Android.Whereas Drago
          n Hills is amusing and Chameleon Run is frenetic, Fotonica is psychedelic. Part "Tron"'
```

```
In [15]:  ▶ dct2=a.split("URL: ")
          dct2 = [x for x in dct2 if len(x) >0]
          dct2
```

jupyter LargestString_WarmUp_Method-2 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 ○

Code

```
In [15]:  dct2=a.split("URL: ")
          dct2 = [x for x in dct2 if len(x) >0]
          dct2
```

```
Out[15]:  ['http://www.nytimes.com/2016/06/30/sports/baseball/washington-nationals-max-scherzer-baffles-mets-completing-a-sweep.htm
          lWASHINGTON — Stellar pitching kept the Mets afloat in the first half of last season despite their offensive woes. But th
          ey cannot produce an encore of their pennant-winning season if their lineup keeps floundering while their pitching is nic
          ked, bruised and stretched thin."We were going to ride our pitching," Manager Terry Collins said before Wednesday's game.
          "But we're not riding it right now. We've got as many problems with our pitching as we do anything."Wednesday's 4-2 loss
          to the Washington Nationals was cruel for the already-limping Mets. Pitching in Steven Matz's place, the spot starter Log
          an Verrett allowed two runs over five innings. But even that was too large a deficit for the Mets' lineup to overcome aga
          inst Max Scherzer, the Nationals' starter."We're not even giving ourselves chances," Collins said, adding later, "We just
          can't give our pitchers any room to work."The Mets did not score until the ninth inning, when a last-gasp two-run homer b
          y James Loney off Nationals reliever Shawn Kelley snapped a streak of 23 scoreless innings for the team.The Mets were swe
          pt in the three-game series and fell six games behind the Nationals in the National League East. Of late, the Mets have l
          ooked worse than their 40-37 record."I don't think we've played half our games yet this year," right fielder Curtis Grand
          erson said. "There's still a lot of things left that can and hopefully will happen."Scherzer toyed with the Mets, who wer
          e initially without Granderson after he was scratched from the lineup with lingering calf tightness. Even though Granders
          on has been inconsistent this season, he had hit well against Scherzer in the past. Alejandro De Aza, who entered the gam
          e with a .165 average, started in right field instead because Collins said the team had few options.After Scherzer gave u
          p a single to Asdrubal Cabrera and walked Loney in the second inning, he retired the next 18 batters, until an eighth-inn
          ing single by Brandon Nimmo.The Mets struggled again with runners on base. After Nimmo and the pinch-hitting Granderson s
          ingled in the eighth, pinch-hitter Travis d'Arnaud grounded out, and De Aza struck out."If they keep adding pressure on t
          hemselves. they're going to continue to struggle." Collins said. "That's one of the things we try to make sure they have
```

```
In [16]:  result2 = [i for i in dct2 if Keymax in i]
```

```
In [17]:  result2
```

```
Out[17]:  []
```

```
In [ ]:
```

THANK YOU…!!