

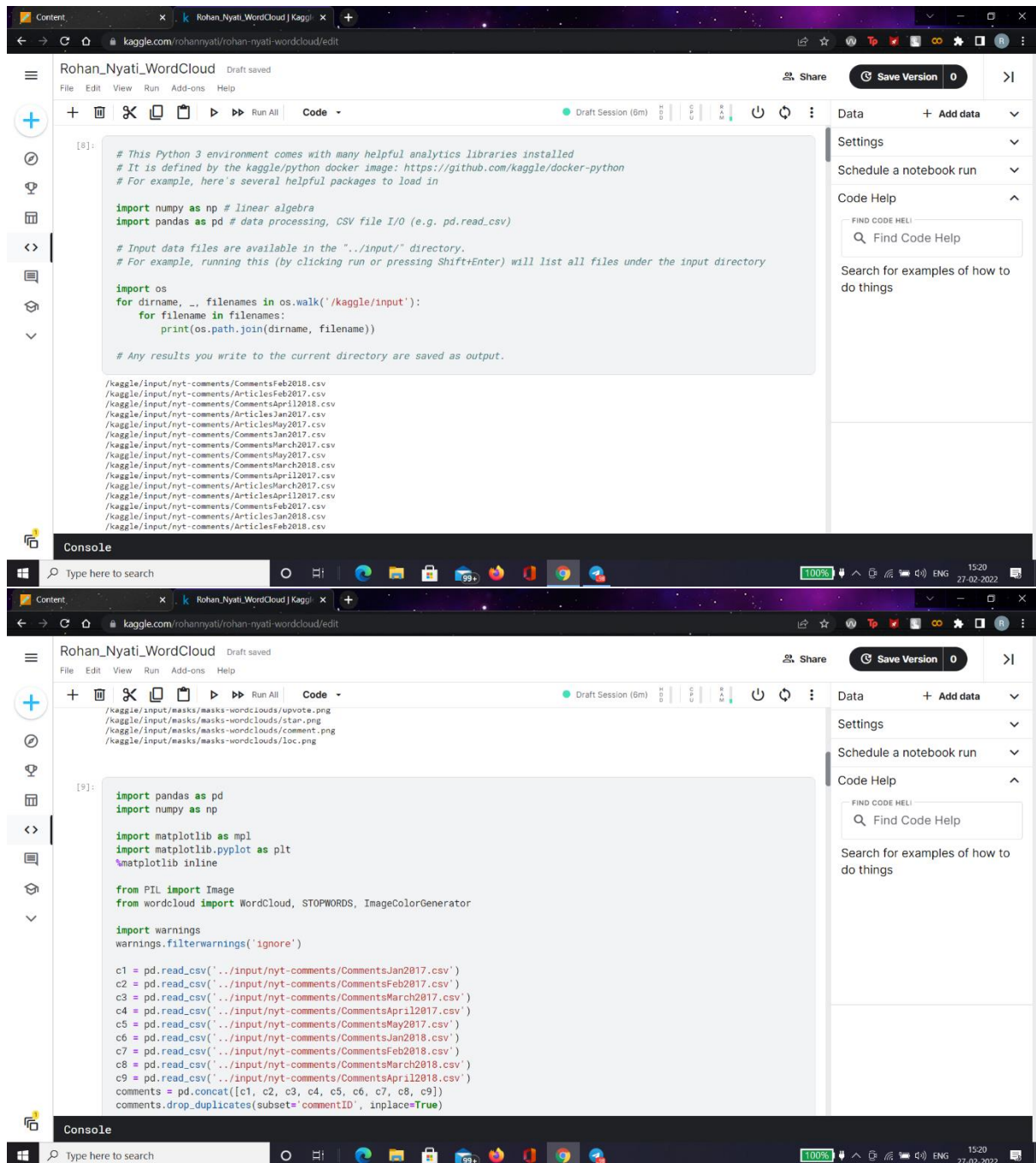
Rohan Nyati

500075940

R177219148

B5-Ai & ML

Experiment -7



Content x Rohan_Nyati_WordCloud | Kaggle x

kaggle.com/rohannyati/rohan-nyati-wordcloud/edit

Rohan_Nyati_WordCloud

Draft saved

File Edit View Run Add-ons Help

Share Save Version 0

Code

Draft Session (8m)

Run All

Data + Add data

Settings

Schedule a notebook run

Code Help

FIND CODE HELI

Find Code Help

Search for examples of how to do things

```
a1 = pd.read_csv('../input/nyt-comments/ArticlesJan2017.csv')
a2 = pd.read_csv('../input/nyt-comments/ArticlesFeb2017.csv')
a3 = pd.read_csv('../input/nyt-comments/ArticlesMarch2017.csv')
a4 = pd.read_csv('../input/nyt-comments/ArticlesApril2017.csv')
a5 = pd.read_csv('../input/nyt-comments/ArticlesMay2017.csv')
a6 = pd.read_csv('../input/nyt-comments/ArticlesJan2018.csv')
a7 = pd.read_csv('../input/nyt-comments/ArticlesFeb2018.csv')
a8 = pd.read_csv('../input/nyt-comments/ArticlesMarch2018.csv')
a9 = pd.read_csv('../input/nyt-comments/ArticlesApril2018.csv')
articles = pd.concat([a1, a2, a3, a4, a5, a6, a7, a8, a9])
articles.drop_duplicates(subset='articleID', inplace=True)
```

[10]: comments.shape

[10]: (2118617, 34)

[11]: articles.shape

[11]: (9298, 16)

[12]:

```
def plot_wordcloud(text, mask=None, max_words=400, max_font_size=120, figure_size=(20.0,10.0),
                  title=None, title_size=40, image_color=False):
    stopwords = set(STOPWORDS)
    more_stopwords = {'one', 'br', 'Po', 'th', 'sayi', 'fo', 'Unknown'}
    stopwords = stopwords.union(more_stopwords)

    wordcloud = WordCloud(background_color='white',
                          stopwords = stopwords,
                          max_words = max_words,
                          max_font_size = max_font_size,
                          random_state = 42,
                          mask = mask)
    wordcloud.generate(text)

    plt.figure(figsize=figure_size)
    if image_color:
        image_colors = ImageColorGenerator(mask);
        plt.imshow(wordcloud.recolor(color_func=image_colors), interpolation="bilinear");
        plt.title(title, fontdict={'size': title_size,
                                   'verticalalignment': 'bottom'})
    else:
        plt.imshow(wordcloud);
        plt.title(title, fontdict={'size': title_size, 'color': 'green',
                                   'verticalalignment': 'bottom'})
    plt.axis('off');
    plt.tight_layout()

d = '../input/masks/masks-wordclouds/'
```

Console

Type here to search

100%

15:20 27-02-2022

Content x Rohan_Nyati_WordCloud | Kaggle x

kaggle.com/rohannyati/rohan-nyati-wordcloud/edit

Rohan_Nyati_WordCloud

Draft saved

File Edit View Run Add-ons Help

Share Save Version 0

Code

Draft Session (8m)

Run All

Data + Add data

Settings

Schedule a notebook run

Code Help

FIND CODE HELI

Find Code Help

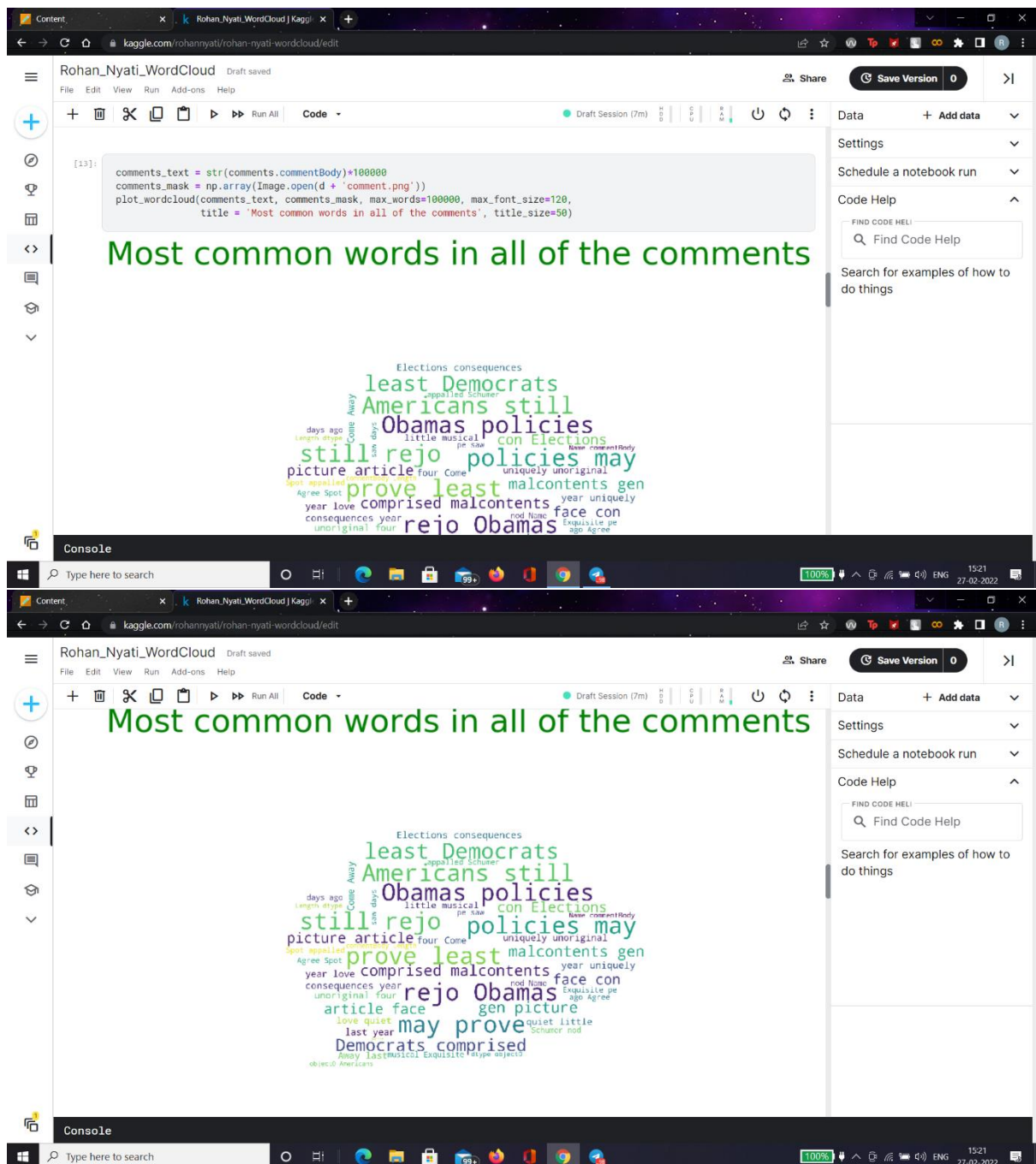
Search for examples of how to do things

Console

Type here to search

100%

15:21 27-02-2022



Content x Rohan_Nyati_WordCloud | Kaggle x

kaggle.com/rohannyati/rohan-nyati-wordcloud/edit

Rohan_Nyati_WordCloud Draft saved

File Edit View Run Add-ons Help

+ Save Version 0

+ Add data

Settings

Schedule a notebook run

Code Help

FIND CODE HELI

Find Code Help

Search for examples of how to do things

```
[14]: n = round(comments.shape[0])
top_recommended_comments_text = str(comments.nlargest(n, columns='recommendations').commentBody)
upvote_mask = np.array(Image.open(d + 'upvote.png'))
plot_wordcloud(top_recommended_comments_text, upvote_mask, max_words=300000, max_font_size=300,
               title = 'Most common words in the top 1% most upvoted comments')
```

Most common words in the top 1% most upvoted comments

Brenda
thi
Monika
Kamal
create
years
Stantcheva
dtype
Dear
People
outrageous
object
thank
Mckenzie
Length
patriot
walked
question
quick
author
dencies
share
lawyer
Jett
provided
President
tendencies
thought
new
mo
sitting
commentBody
Young
interpretation
bit
Written
sickened
follow yo
veteran
Spicer

Console

Type here to search

Content x Rohan_Nyati_WordCloud | Kaggle x

kaggle.com/rohannyati/rohan-nyati-wordcloud/edit

Rohan_Nyati_WordCloud Draft saved

File Edit View Run Add-ons Help

+ Save Version 0

+ Add data

Settings

Schedule a notebook run

Code Help

FIND CODE HELI

Find Code Help

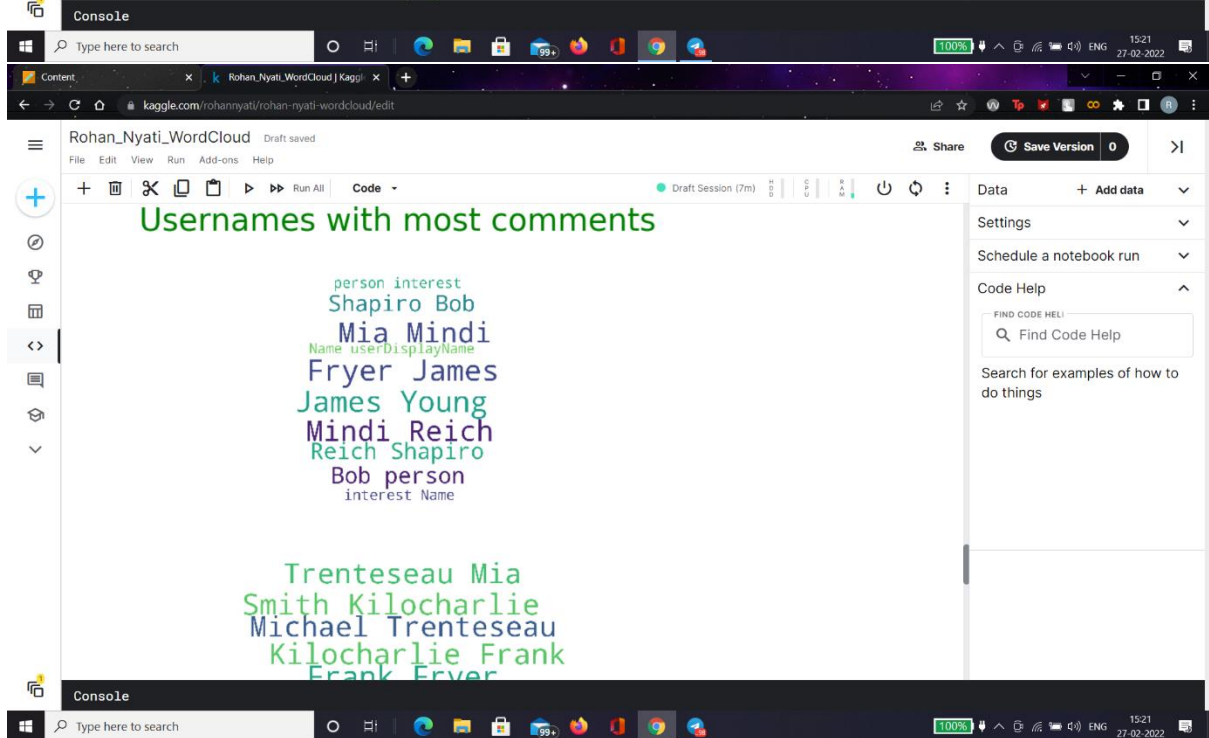
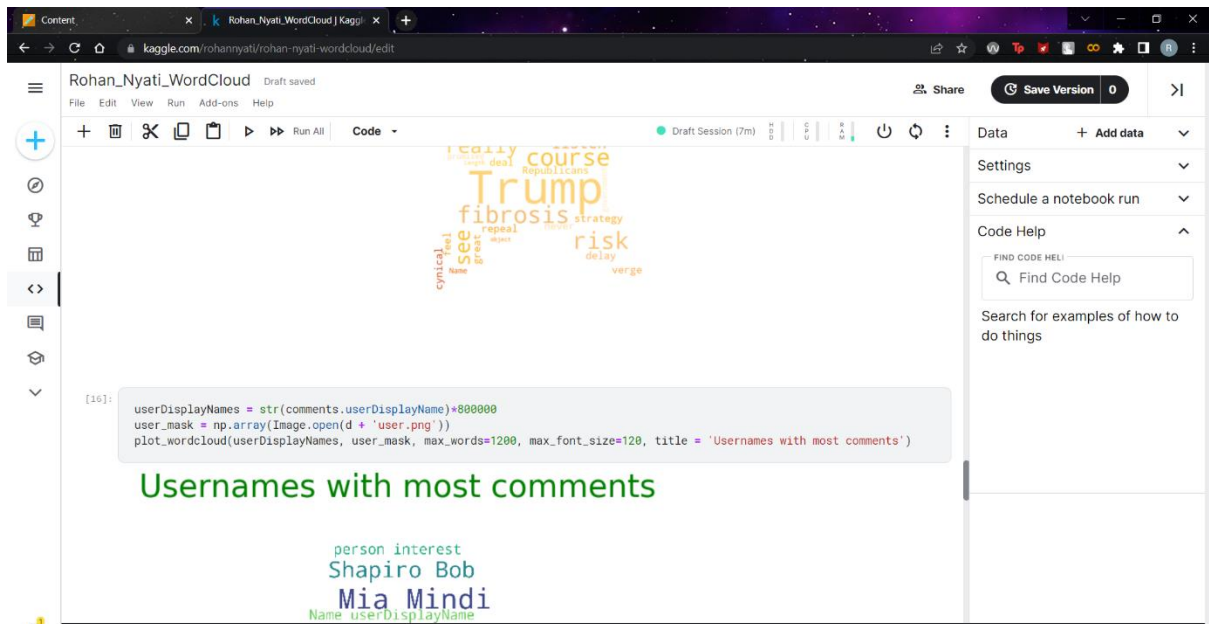
Search for examples of how to do things

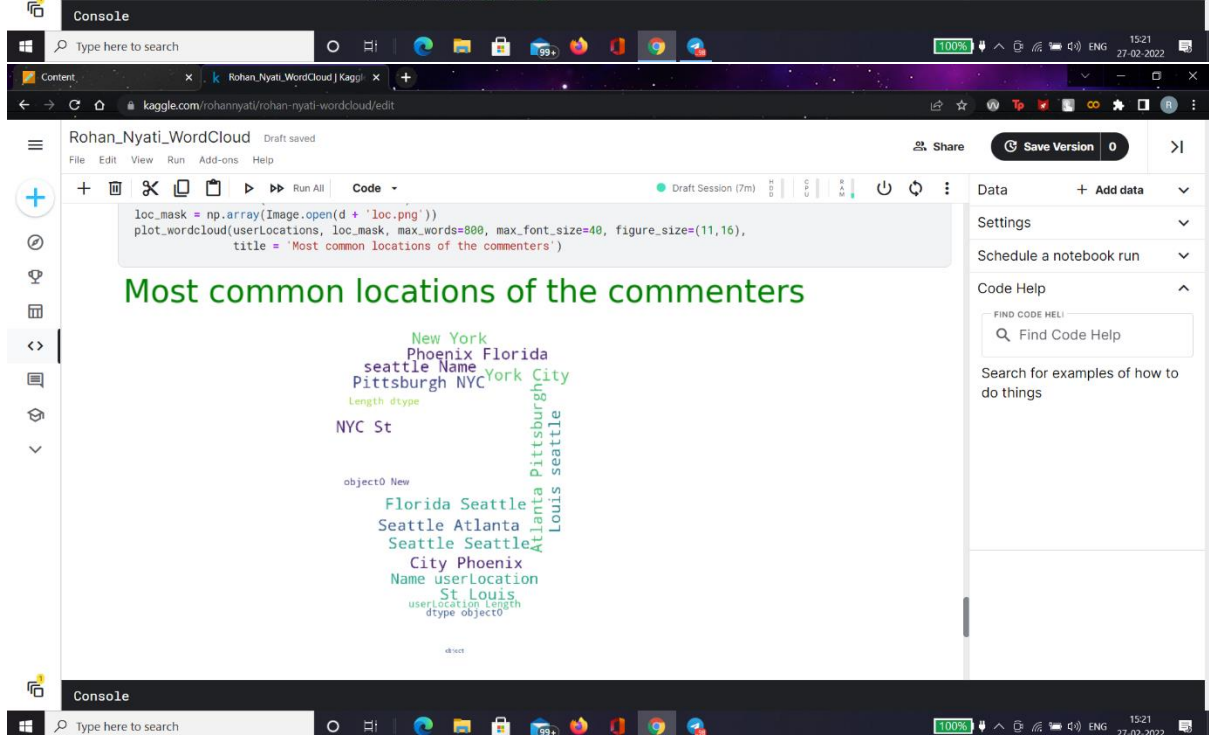
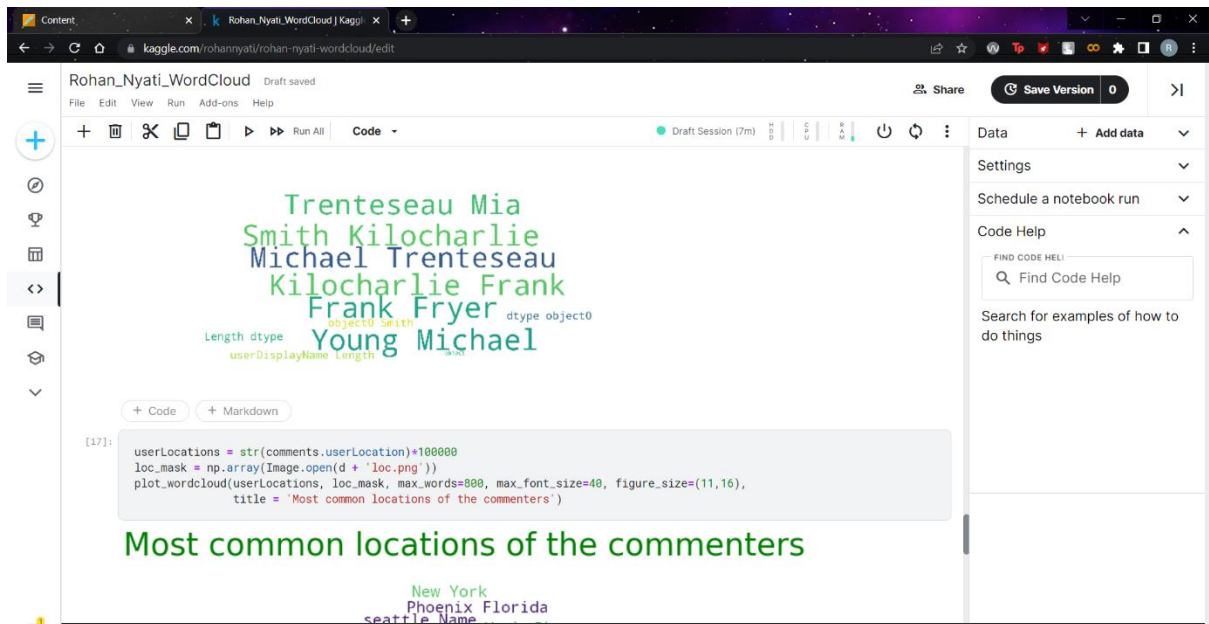
Most common words in the top 1% most upvoted comments

Brenda
thi
Monika
Kamal
create
years
Stantcheva
dtype
Dear
People
outrageous
object
thank
Mckenzie
Length
patriot
walked
question
quick
author
dencies
share
lawyer
Jett
provided
President
tendencies
thought
new
mo
sitting
commentBody
Young
interpretation
bit
Written
sickened
follow yo
veteran
Spicer

Console

Type here to search





Content

Rohan_Nyati_WordCloud | Kaggle

kaggle.com/rohannyati/rohan-nyati-wordcloud/edit

Rohan_Nyati_WordCloud Draft saved

File Edit View Run Add-ons Help

+ Draft Session (7m)

Share Save Version 0

Data Add data

Settings

Schedule a notebook run

Code Help

FIND CODE HELI

Find Code Help

Search for examples of how to do things

```
[18]: headlines = str(articles.headline)*100
      plot_wordcloud(headlines, max_words=800, max_font_size=30, figure_size=(12, 10),
                    title = 'Most common words in the headlines on the articles')
```

Most common words in the headlines on the articles

Leadership Poised Hope Tested Hope

Obama Pi Little Troublemakers

Hope Young Poised Topple

Pi Fractured Troublemakers Angela

Young Pre World Tested

Target Boots Topple Obama

Fractured World

Boots Stranger

+ Code + Markdown

Console