In [2]:
```
import pandas as pd

dataset = pd.read_csv(r"C:\Users\Vaish\Downloads\tweets.csv", encoding = 'ISO-8859-

dataset.head(3)
```

Out[2]:

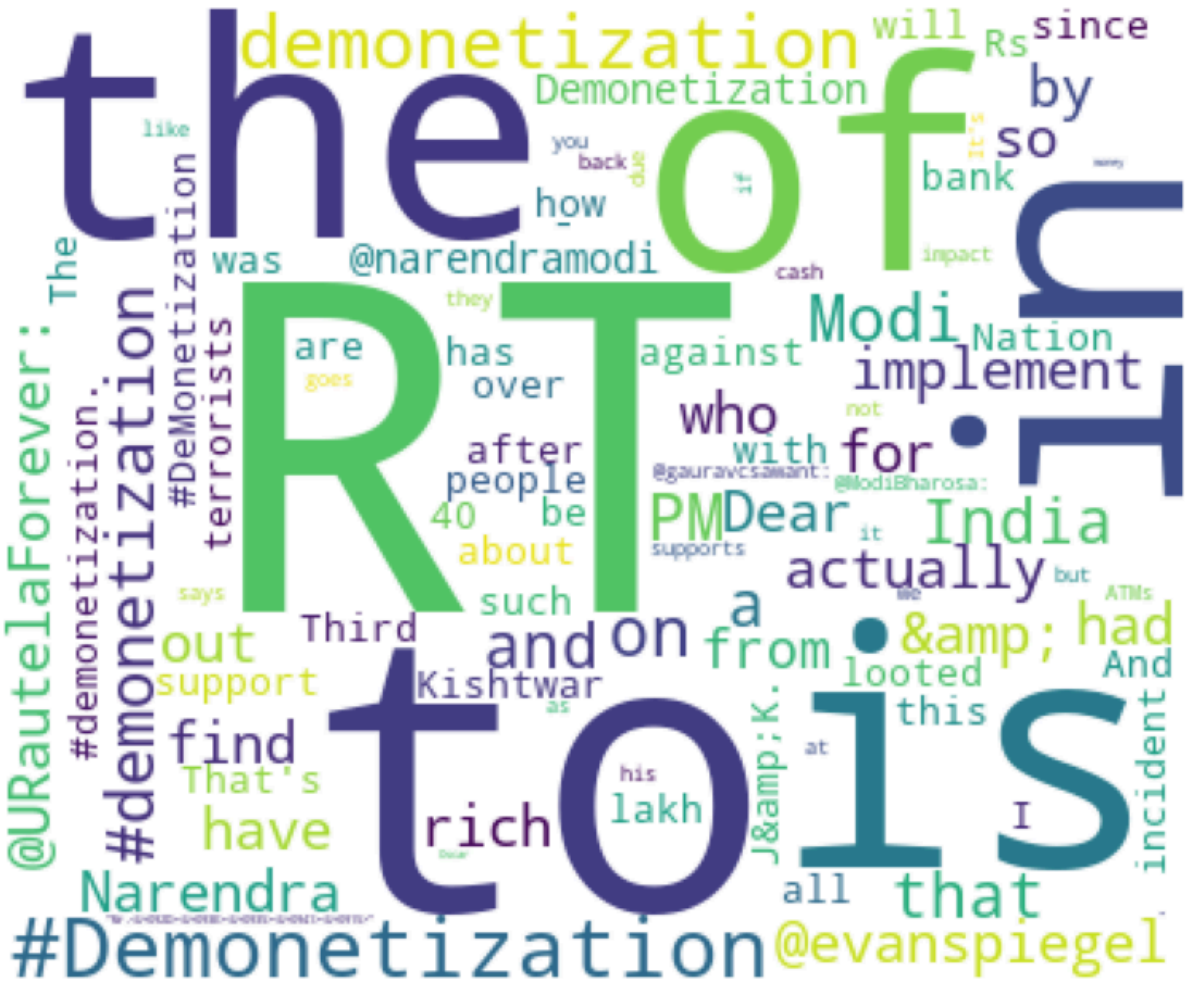
	Unnamed: 0	X	text	favorited	favoriteCount	replyToSN	created	truncate
0	1	1	RT @rssurjewala: Critical question: Was PayTM ...	False	0	NaN	2016-11-23 18:40:30	Fals
1	2	2	RT @Hemant_80: Did you vote on #Demonetization...	False	0	NaN	2016-11-23 18:40:29	Fals
2	3	3	RT @roshankar: Former FinSec, RBI Dy Governor,...	False	0	NaN	2016-11-23 18:40:03	Fals

In [4]:
```
def gen_freq(text):


    word_list = []

    for tw_words in text.split():

        word_list.extend(tw_words)



    #Create word frequencies using word_list

    word_freq = pd.Series(word_list).value_counts()



    #Print top 10 words

    word_freq[:10]

    return word_freq
```

In [6]:
```
word_freq = gen_freq(dataset.text.str)

word_freq
```

Out[6]:
```
RT                          11053
to                           7650
is                           5152
in                           4491
the                          4331
                            ...
#News                           1
notes|                          1
https://t.co/ECl4oIzdHA         1
https://t.co/9MjFtLtCtR         1
https://t.co/hwgqjbqgvG         1
Name: count, Length: 19601, dtype: int64
```

In [8]:
```
#Import libraries

import matplotlib.pyplot as plt

from wordcloud import WordCloud

#Generate word cloud

wc = WordCloud(width=400, height=330, max_words=200,

        background_color='white').generate_from_frequencies(word_freq)



plt.figure(figsize=(12, 8))

plt.imshow(wc)

plt.axis('off')

plt.show()
```



In [14]:

```
import re

def clean_text(text):

    #Remove RT

    text = re.sub(r'RT', '', text)

    #Fix &

    text = re.sub(r'&amp;', '', text)

    #Remove punctuations

    text = re.sub(r'[?!.;:,#@-]', '', text)

    #Convert to lowercase to maintain consistency

    text = text.lower()

    # remove digits

    text = re.sub(r'\d+', '', text)

    #Fix %
```

```
    text = re.sub(r'%', '', text)



    return text
```

In [16]:
```
from wordcloud import STOPWORDS
print(STOPWORDS)
```

```
{'was', 'is', 'their', 'here', 'then', "there's", 'of', 'did', "where's", 'during',
"you're", 'such', 'below', 'doing', 'between', 'under', 'too', 'otherwise', 'in', 'o
nce', 'the', 'no', 'itself', 'nor', 'just', 'she', "wasn't", 'its', "we'll", 'do',
'our', 'so', 'against', 'would', 'had', 'few', 'how', 'them', "doesn't", 'your', 'in
to', 'being', 'other', 'very', 'up', 'those', "you've", 'what', 'same', "i'd", 'mor
e', 'why', "haven't", 'not', 'above', 'than', "you'll", "don't", 'at', 'her', 'abou
t', "they'll", 'with', 'http', "hasn't", 'can', "i'm", 'own', 'myself', "what's", 'm
e', 'but', 'him', 'ourselves', 'all', 'an', "how's", 'www', 'both', 'ought', "we'r
e", 'down', 'hence', 'his', 'my', "why's", 'each', 'on', 'himself', 'themselves', 'a
gain', 'has', 'ever', "she'd", 'as', "he's", 'until', 'because', 'have', 'over', "wh
o's", "i've", 'from', "wouldn't", 'any', "she'll", 'you', 'by', 'shall', 'theirs',
'herself', "he'd", 'he', 'could', "couldn't", 'through', 'i', 'for', "aren't", 'an
d', 'therefore', 'when', "mustn't", "won't", 'been', 'a', 'before', 'only', 'havin
g', 'cannot', 'are', 'yourself', 'if', 'while', "they're", 'however', 'we', 'hers',
"let's", 'ours', 'that', 'further', "i'll", 'does', "it's", 'were', "hadn't", "sha
n't", 'some', "can't", "when's", "didn't", 'be', 'most', 'it', 'k', "that's", 'com',
"isn't", "we'd", 'out', "she's", "shouldn't", 'which', 'this', 'yourselves', "we'v
e", 'who', 'am', 'there', "they'd", 'to', 'or', 'where', 'after', "he'll", 'these',
"you'd", "they've", "weren't", 'also', 'they', 'yours', 'else', 'get', 'like', 'shou
ld', 'whom', "here's", 'r', 'since', 'off'}
```

In [18]:
```
text = dataset.text.apply(lambda x: clean_text(x))

word_freq = gen_freq(text.str)

word_freq = word_freq.drop(labels=STOPWORDS, errors='ignore')



#Generate word cloud

wc = WordCloud(width=350, height=220, max_words=500,

               background_color='white').generate_from_frequencies(word_freq)



plt.figure(figsize=(12, 8))

plt.imshow(wc, interpolation='bilinear')

plt.axis('off')

plt.show()
```

In [20]:
```
import pandas as pd


text = ['Sarah lives in a hut in the village.',
      'She has an apple tree in her backyard.',
      'The apples are red in colour.']


df = pd.DataFrame(text, columns=['Sentence'])


df
```

Out[20]:

	Sentence
0	Sarah lives in a hut in the village.
1	She has an apple tree in her backyard.
2	The apples are red in colour.

In [ ]: