

# Python for Data Science - Predict

© Explore Data Science Academy

## Eskom Data Analysis



---

### Honour Code

I {**EHIBHAIEMEN, UGHELE**}, confirm - by submitting this document - that the solutions in this notebook are a result of my own work and that I abide by the EDSA honour code (<https://drive.google.com/file/d/1QDCjGZJ8-FmJE3bZdIQNwnJyQKPhHZBn/view?usp=sharing>).

Non-compliance with the honour code constitutes a material breach of contract.

---

### Overview

Functions are important in reducing the replication of code as well as giving the user the functionality of getting an output on varying inputs. The functions you will write all use Eskom data/variables.

### Instructions to Students

- **Do not add or remove cells in this notebook. Do not edit or remove the `### START FUNCTION` or `### END FUNCTION` comments. Do not add any code outside of the**

**functions you are required to edit. Doing any of this will lead to a mark of 0%!**

- Answer the questions according to the specifications provided.
- Use the given cell in each question to see if your function matches the expected outputs.
- Do not hard-code answers to the questions.
- The use of stackoverflow, google, and other online tools are permitted. However, copying fellow student's code is not permissible and is considered a breach of the Honour code. Doing this will result in a mark of 0%.
- Good luck, and may the force be with you!

## Imports

```
In [1]: import pandas as pd
import numpy as np
```

## Data Loading and Preprocessing

### Electricification by province (EBP) data

```
In [2]: ebp_url = 'https://raw.githubusercontent.com/Explore-AI/Public-Data/master/Data/electricity_by_province.csv'
ebp_df = pd.read_csv(ebp_url)

for col, row in ebp_df.iloc[:,1:1].iteritems():
    ebp_df[col] = ebp_df[col].str.replace(',', '').astype(int)

ebp_df.head()
```

	Financial Year (1 April - 30 March)	Province									
		Limpopo	Mpumalanga	North west	Free State	Kwazulu Natal	Eastern Cape	Western Cape	Northern Cape	Gauteng	
0	2000/1	51860	28365	48429	21293	63413	49008	48429	6168	39660	
1	2001/2	68121	26303	38685	20928	64123	45773	38685	10359	36024	
2	2002/3	49881	11976	28532	10316	63078	55748	28532	6869	32127	
3	2003/4	42034	33515	34027	16135	60282	47414	34027	10976	39488	
4	2004/5	54646	16218	21450	5668	37811	42041	21450	6316	18422	

## Twitter data

```
In [3]: twitter_url = 'https://raw.githubusercontent.com/Explore-AI/Public-Data/master/Data/twitter.csv'
twitter_df = pd.read_csv(twitter_url)
twitter_df.head()
```

Out[3]:

		Tweets	Date
0	@BongaDlulane Please send an email to mediades...	2019-11-29 12:50:54	
1	@saucy_mamiie Pls log a call on 0860037566	2019-11-29 12:46:53	
2	@BongaDlulane Query escalated to media desk.	2019-11-29 12:46:10	
3	Before leaving the office this afternoon, head...	2019-11-29 12:33:36	
4	#ESKOMFREESTATE #MEDIASTATEMENT : ESKOM SUSPEN...	2019-11-29 12:17:43	

## Important Variables (Do not edit these!)

In [4]:

```
# gauteng ebp data as a List
gauteng = ebp_df['Gauteng'].astype(float).to_list()

# dates for twitter tweets
dates = twitter_df['Date'].to_list()

# dictionary mapping official municipality twitter handles to the municipality name
mun_dict = {
    '@CityofCTAlerts' : 'Cape Town',
    '@CityPowerJhb' : 'Johannesburg',
    '@eThekwiiniM' : 'eThekwiini' ,
    '@EMMInfo' : 'Ekurhuleni',
    '@centlecutility' : 'Mangaung',
    '@NMBmunicipality' : 'Nelson Mandela Bay',
    '@CityTshwane' : 'Tshwane'
}

# dictionary of english stopwords
stop_words_dict = {
    'stopwords':[
        'where', 'done', 'if', 'before', 'll', 'very', 'keep', 'something', 'nothing',
        'may', 'why', "'s", 'therefore', 'you', 'with', 'towards', 'make', 'really',
        'during', 'mine', 'do', 'would', 'of', 'off', 'six', 'yourself', 'becoming',
        'seeming', 'hence', 'us', 'anywhere', 'regarding', 'whole', 'down', 'seem',
        'their', 'various', 'thereafter', 'd', 'above', 'put', 'sometime', 'moreover',
        'at', 'four', 'each', 'among', 'whatever', 'any', 'anyhow', 'herein', 'become',
        'was', 'almost', 'twelve', 'used', 'who', 'go', 'not', 'enough', 'well', 've',
        'everywhere', 'yourselves', 'across', 'myself', 'further', 'did', 'then', 'is',
        'became', 'however', 'many', 'thence', 'onto', "'m", 'my', 'own', 'must', 'wher',
        'becomes', 'alone', 'due', 'being', 'neither', 'a', 'over', 'beside', 'fifteen',
        'forty', 'what', 'less', 'and', 'please', 'toward', 'about', 'below', 'hereaft',
        'against', 'whereupon', 'top', 'first', 'three', 'show', 'per', 'five', 'two',
        'get', 'thereby', 'noone', 'had', 'now', 'everyone', 'everything', 'nowhere',
        'so', 'both', 'otherwise', 'whereby', 'unless', 'somewhere', 'give', 'formerly',
        'while', 'empty', 'doing', 'besides', 'thus', 'this', 'anyone', 'its', 'after',
        'n't', 'name', 'even', 'eleven', 'by', 'from', 'when', 'or', 'anyway', 'how',
        'much', 'another', 'since', 'hundred', 'serious', "'ve", 'ever', 'out', 'full',
        'been', 'in', "'d", 'wherever', 'part', 'someone', 'therein', 'can', 'seemed',
        "'s", "'re", 'most', 'one', "n't", 'into', 'some', 'will', 'these', 'twenty',
        'also', 'along', 'than', 'anything', 'he', 'there', 'does', 'we', 'll', 'latt',
        'hers', 'should', 'they', "'s", 'either', 'am', 'be', 'perhaps', "'re', 'only',
        'made', "'m", 'always', 'those', 'have', 'again', 'her', 'once', 'ours', 'hers',
        'more', 'sometimes', 'your', 'yours', 'that', 'around', 'his', 'indeed', 'most',
        'seems', "'m", 'himself', 'latter', 'whither', 'amount', 'other', 'nevertheles',
        'beforehand', 'just', 'an', 'beyond', 'amongst', 'none', "'ve", 'say', 'via',
```

```
'because', 'rather', 'using', 'without', 'throughout', 'on', 'she', 'never',
'them', 'whereafter', 'quite', 'which', 'move', 'thru', 'until', 'afterwards',
'him', 'could', 'front', 'within', 're', 'back', 'such', 'already', 'several'
'same', 'were', 'it', 'every', 'third', 'together'
]
}
```

## Function 1: Metric Dictionary

Write a function that calculates the mean, median, variance, standard deviation, minimum and maximum of of list of items. You can assume the given list is contains only numerical entries, and you may use numpy functions to do this.

### Function Specifications:

- Function should allow a list as input.
- It should return a `dict` with keys `'mean'`, `'median'`, `'std'`, `'var'`, `'min'`, and `'max'`, corresponding to the mean, median, standard deviation, variance, minimum and maximum of the input list, respectively.
- The standard deviation and variance values must be unbiased. **Hint:** use the `ddof` parameter in the corresponding numpy functions!
- All values in the returned `dict` should be rounded to 2 decimal places.

```
In [5]: ### START FUNCTION
def dictionary_of_metrics(items):
    # your code here
    summary_dict = {'mean' : round(np.mean(items), 2),
                    'median' : round(np.median(items), 2),
                    'std' : round(np.std(items, ddof = 1), 2),
                    'var' : round(np.var(items, ddof = 1), 2),
                    'min' : round(min(items), 2),
                    'max' : round(max(items), 2)}
    return summary_dict
### END FUNCTION
```

```
In [6]: dictionary_of_metrics(gauteng)
```

```
Out[6]: {'mean': 26244.42,
'median': 24403.5,
'std': 10400.01,
'var': 108160153.17,
'min': 8842.0,
'max': 39660.0}
```

### Expected Output:

```
dictionary_of_metrics(gauteng) == {'mean': 26244.42,
                                    'median': 24403.5,
                                    'var': 108160153.17,
                                    'std': 10400.01,
                                    'min': 8842.0,
                                    'max': 39660.0}
```

## Function 2: Five Number Summary

Write a function which takes in a list of integers and returns a dictionary of the [five number summary](#).

### Function Specifications:

- The function should take a list as input.
- The function should return a `dict` with keys `'max'`, `'median'`, `'min'`, `'q1'`, and `'q3'` corresponding to the maximum, median, minimum, first quartile and third quartile, respectively. You may use numpy functions to aid in your calculations.
- All numerical values should be rounded to two decimal places.

```
In [7]: ### START FUNCTION
def five_num_summary(items):
    # your code here
    sum_list = ['max', 'median', 'min', 'q1', 'q3']
    desc = pd.Series(items).describe()
    desc.index = ['count', 'mean', 'std', 'min', 'q1', 'median', 'q3', 'max']
    desc = dict(desc)
    desc = {k : v for k, v in desc.items() if k in sum_list}
### END FUNCTION
```

```
In [8]: five_num_summary(gauteng)
```

```
Out[8]: {'min': 8842.0,
          'q1': 18653.0,
          'median': 24403.5,
          'q3': 36372.0,
          'max': 39660.0}
```

### Expected Output:

```
five_num_summary(gauteng) == {
    'max': 39660.0,
    'median': 24403.5,
    'min': 8842.0,
    'q1': 18653.0,
    'q3': 36372.0
}
```

## Function 3: Date Parser

The `dates` variable (created at the top of this notebook) is a list of dates represented as strings. The string contains the date in `'yyyy-mm-dd'` format, as well as the time in `hh:mm:ss` formamt. The first three entries in this variable are:

```
dates[:3] == [
    '2019-11-29 12:50:54',
    '2019-11-29 12:46:53',
    '2019-11-29 12:46:10'
]
```

Write a function that takes as input a list of these datetime strings and returns only the date in 'yyyy-mm-dd' format.

### Function Specifications:

- The function should take a list of strings as input.
- Each string in the input list is formatted as 'yyyy-mm-dd hh:mm:ss'.
- The function should return a list of strings where each element in the returned list contains only the date in the 'yyyy-mm-dd' format.

```
In [9]: ### START FUNCTION
def date_parser(dates):
    # your code here
    dates = [s.split(' ')[0] for s in dates]
    return dates
### END FUNCTION
```

```
In [10]: date_parser(dates[:3])
```

```
Out[10]: ['2019-11-29', '2019-11-29', '2019-11-29']
```

### Expected Output:

```
date_parser(dates[:3]) == ['2019-11-29', '2019-11-29', '2019-11-29']
date_parser(dates[-3:]) == ['2019-11-20', '2019-11-20', '2019-11-20']
```

## Function 4: Municipality & Hashtag Detector

Write a function which takes in a pandas dataframe and returns a modified dataframe that includes two new columns that contain information about the municipality and hashtag of the tweet.

### Function Specifications:

- Function should take a pandas `dataframe` as input.
- Extract the municipality from a tweet using the `mun_dict` dictionary given below, and insert the result into a new column named '`municipality`' in the same dataframe.
- Use the entry `np.nan` when a municipality is not found.
- Extract a list of hashtags from a tweet into a new column named '`hashtags`' in the same dataframe.
- Use the entry `np.nan` when no hashtags are found.

**Hint:** you will need to `mun_dict` variable defined at the top of this notebook.

```

```
In [11]: ### START FUNCTION
def extract_municipality_hashtags(df):
    # your code here
    def fill(y):
```

```

        lst = [w for w in y if '#' in w]

    return np.nan if not bool(lst) else lst

df_ = df.copy()

f_ = df_[ 'Tweets' ].str.lower().str.split(' ').apply(lambda x: mun_dict[x[0]] if (d_ = df_[ 'Tweets' ].str.lower().str.split(' ').apply(fill)

df_[ 'municipality' ] = f_
df_[ 'hashtags' ] = d_

return df_
### END FUNCTION

```

In [12]: `extract_municipality_hashtags(twitter_df.copy())`

Out[12]:

|     | Tweets                                            | Date                   | municipality | hashtags                                              |
|-----|---------------------------------------------------|------------------------|--------------|-------------------------------------------------------|
| 0   | @BongaDlulane Please send an email to mediades... | 2019-11-29<br>12:50:54 | NaN          | NaN                                                   |
| 1   | @saucy_mamiie Pls log a call on 0860037566        | 2019-11-29<br>12:46:53 | NaN          | NaN                                                   |
| 2   | @BongaDlulane Query escalated to media desk.      | 2019-11-29<br>12:46:10 | NaN          | NaN                                                   |
| 3   | Before leaving the office this afternoon, head... | 2019-11-29<br>12:33:36 | NaN          | NaN                                                   |
| 4   | #ESKOMFREESTATE #MEDIASTATEMENT : ESKOM SUSPEN... | 2019-11-29<br>12:17:43 | NaN          | [#eskomfreestate,<br>#mediastatement]                 |
| ... | ...                                               | ...                    | ...          | ...                                                   |
| 195 | Eskom's Visitors Centres' facilities include i... | 2019-11-20<br>10:29:07 | NaN          | NaN                                                   |
| 196 | #Eskom connected 400 houses and in the process... | 2019-11-20<br>10:25:20 | NaN          | [#eskom, more.\n\n#eskom,<br>#poweringyourworld\n...] |
| 197 | @ArthurGodbeer Is the power restored as yet?      | 2019-11-20<br>10:07:59 | NaN          | NaN                                                   |
| 198 | @MuthambiPaulina @SABCNewsOnline @IOL @eNCA @e... | 2019-11-20<br>10:07:41 | NaN          | NaN                                                   |
| 199 | RT @GP_DHS: The @GautengProvince made a commit... | 2019-11-20<br>10:00:09 | NaN          | NaN                                                   |

200 rows × 4 columns

### Expected Outputs:

`extract_municipality_hashtags(twitter_df.copy())`

|   | Tweets                                            | Date                   | municipality | hashtags |
|---|---------------------------------------------------|------------------------|--------------|----------|
| 0 | @BongaDlulane Please send an email to mediades... | 2019-11-29<br>12:50:54 | NaN          | NaN      |

|     | Tweets                                                  | Date                | municipality | hashtags                             |
|-----|---------------------------------------------------------|---------------------|--------------|--------------------------------------|
| 1   | @saucy_mamiie Pls log a call on 0860037566              | 2019-11-29 12:46:53 | NaN          | NaN                                  |
| 2   | @BongaDlulane Query escalated to media desk.            | 2019-11-29 12:46:10 | NaN          | NaN                                  |
| 3   | Before leaving the office this afternoon, head...       | 2019-11-29 12:33:36 | NaN          | NaN                                  |
| 4   | #ESKOMFREESTATE<br>#MEDIASTATEMENT : ESKOM SUSPEN...    | 2019-11-29 12:17:43 | NaN          | [#eskomfreestate, #mediastatement]   |
| ... | ...                                                     | ...                 | ...          | ...                                  |
| 195 | Eskom's Visitors Centres' facilities include i...       | 2019-11-20 10:29:07 | NaN          | NaN                                  |
| 196 | #Eskom connected 400 houses and in the process...       | 2019-11-20 10:25:20 | NaN          | [#eskom, #eskom, #poweringyourworld] |
| 197 | @ArthurGodbeer Is the power restored as yet?            | 2019-11-20 10:07:59 | NaN          | NaN                                  |
| 198 | @MuthambiPaulina<br>@SABCNewsOnline @IOL<br>@eNCA @e... | 2019-11-20 10:07:41 | NaN          | NaN                                  |
| 199 | RT @GP_DHS: The<br>@GautengProvince made a commit...    | 2019-11-20 10:00:09 | NaN          | NaN                                  |

## Function 5: Number of Tweets per Day

Write a function which calculates the number of tweets that were posted per day.

### Function Specifications:

- It should take a pandas dataframe as input.
- It should return a new dataframe, grouped by day, with the number of tweets for that day.
- The index of the new dataframe should be named `Date`, and the column of the new dataframe should be `'Tweets'`, corresponding to the date and number of tweets, respectively.
- The date should be formatted as `yyyy-mm-dd`, and should be a datetime object. **Hint:** look up `pd.to_datetime` to see how to do this.

```
In [13]: ### START FUNCTION
def number_of_tweets_per_day(df):
    # your code here
    a = df['Date'].str.split(' ', expand = True).iloc[:, 0]
    df['Date'] = pd.to_datetime(a)
    return df.groupby(by = 'Date').count()
### END FUNCTION
```

```
In [14]: number_of_tweets_per_day(twitter_df.copy())
```

Out[14]:

**Tweets**

| Date       |    |
|------------|----|
| 2019-11-20 | 18 |
| 2019-11-21 | 11 |
| 2019-11-22 | 25 |
| 2019-11-23 | 19 |
| 2019-11-24 | 14 |
| 2019-11-25 | 20 |
| 2019-11-26 | 32 |
| 2019-11-27 | 13 |
| 2019-11-28 | 32 |
| 2019-11-29 | 16 |

**Expected Output:**

```
number_of_tweets_per_day(twitter_df.copy())
```

| Date       |    |
|------------|----|
| 2019-11-20 | 18 |
| 2019-11-21 | 11 |
| 2019-11-22 | 25 |
| 2019-11-23 | 19 |
| 2019-11-24 | 14 |
| 2019-11-25 | 20 |
| 2019-11-26 | 32 |
| 2019-11-27 | 13 |
| 2019-11-28 | 32 |
| 2019-11-29 | 16 |

## Function 6: Word Splitter

Write a function which splits the sentences in a dataframe's column into a list of the separate words. The created lists should be placed in a column named '`'Split Tweets'`' in the original dataframe. This is also known as [tokenization](#).

**Function Specifications:**

- It should take a pandas dataframe as an input.
- The dataframe should contain a column, named 'Tweets' .
- The function should split the sentences in the 'Tweets' into a list of separate words, and place the result into a new column named 'Split Tweets' . The resulting words must all be lowercase!
- The function should modify the input dataframe directly.
- The function should return the modified dataframe.

In [15]:

```
### START FUNCTION
def word_splitter(df):
    # your code here
    df['Split Tweets'] = df['Tweets'].str.lower().str.split(' ')
    return df
### END FUNCTION
```

In [16]:

```
word_splitter(twitter_df.copy())
```

Out[16]:

|     | Tweets                                            | Date                   | Split Tweets                                      |
|-----|---------------------------------------------------|------------------------|---------------------------------------------------|
| 0   | @BongaDlulane Please send an email to mediades... | 2019-11-29<br>12:50:54 | [@bongadlulane, please, send, an, email, to, m... |
| 1   | @saucy_mamiie Pls log a call on 0860037566        | 2019-11-29<br>12:46:53 | [@saucy_mamiie, pls, log, a, call, on, 0860037... |
| 2   | @BongaDlulane Query escalated to media desk.      | 2019-11-29<br>12:46:10 | [@bongadlulane, query, escalated, to, media, d... |
| 3   | Before leaving the office this afternoon, head... | 2019-11-29<br>12:33:36 | [before, leaving, the, office, this, afternoon... |
| 4   | #ESKOMFREESTATE #MEDIASTATEMENT : ESKOM SUSPEN... | 2019-11-29<br>12:17:43 | [#eskomfreestate, #mediastatement, ; eskom, s...  |
| ... | ...                                               | ...                    | ...                                               |
| 195 | Eskom's Visitors Centres' facilities include i... | 2019-11-20<br>10:29:07 | [eskom's, visitors, centres', facilities, incl... |
| 196 | #Eskom connected 400 houses and in the process... | 2019-11-20<br>10:25:20 | [#eskom, connected, 400, houses, and, in, the,... |
| 197 | @ArthurGodbeer Is the power restored as yet?      | 2019-11-20<br>10:07:59 | [@arthurgodbeer, is, the, power, restored, as,... |
| 198 | @MuthambiPaulina @SABCNewsOnline @IOL @eNCA @e... | 2019-11-20<br>10:07:41 | [@muthambipaulina, @sabcnewsonline, @iol, @enc... |
| 199 | RT @GP_DHS: The @GautengProvince made a commit... | 2019-11-20<br>10:00:09 | [rt, @gp_dhs:, the, @gautengprovince, made, a,... |

200 rows × 3 columns

### Expected Output:

```
word_splitter(twitter_df.copy())
```

|  | Tweets | Date | Split Tweets |
|--|--------|------|--------------|
|--|--------|------|--------------|

|     |                                                      | Tweets                 | Date | Split Tweets                                       |
|-----|------------------------------------------------------|------------------------|------|----------------------------------------------------|
| 0   | @BongaDlulane Please send an email to mediades...    | 2019-11-29<br>12:50:54 |      | [@bongadlulane, please, send, an, email, to, m...] |
| 1   | @saucy_mamiie Pls log a call on 0860037566           | 2019-11-29<br>12:46:53 |      | [@saucy_mamiie, pls, log, a, call, on, 0860037...] |
| 2   | @BongaDlulane Query escalated to media desk.         | 2019-11-29<br>12:46:10 |      | [@bongadlulane, query, escalated, to, media, d...] |
| 3   | Before leaving the office this afternoon, head...    | 2019-11-29<br>12:33:36 |      | [before, leaving, the, office, this, afternoon...] |
| 4   | #ESKOMFREESTATE<br>#MEDIASTATEMENT : ESKOM SUSPEN... | 2019-11-29<br>12:17:43 |      | #[eskomfreestate, #mediastatement, ; eskom, s...]  |
| ... | ...                                                  | ...                    |      | ...                                                |
| 195 | Eskom's Visitors Centres' facilities include i...    | 2019-11-20<br>10:29:07 |      | [eskom's, visitors, centres', facilities, incl...] |
| 196 | #Eskom connected 400 houses and in the process...    | 2019-11-20<br>10:25:20 |      | #[eskom, connected, 400, houses, and, in, the,...] |
| 197 | @ArthurGodbeer Is the power restored as yet?         | 2019-11-20<br>10:07:59 |      | [@arthurgodbeer, is, the, power, restored, as,...] |
| 198 | @MuthambiPaulina<br>@SABCNewsOnline @IOL @eNCA @e... | 2019-11-20<br>10:07:41 |      | [@muthambipaulina, @sabcnewsonline, @iol, @enc...] |
| 199 | RT @GP_DHS: The @GautengProvince made a commit...    | 2019-11-20<br>10:00:09 |      | [rt, @gp_dhs; the, @gautengprovince, made, a,...]  |

## Function 7: Stop Words

Write a function which removes english stop words from a tweet.

### Function Specifications:

- It should take a pandas dataframe as input.
- Should tokenise the sentences according to the definition in function 6. Note that function 6 **cannot be called within this function**.
- Should remove all stop words in the tokenised list. The stopwords are defined in the `stop_words_dict` variable defined at the top of this notebook.
- The resulting tokenised list should be placed in a column named "Without Stop Words".
- The function should modify the input dataframe.
- The function should return the modified dataframe.

```
In [17]: d_ = pd.DataFrame(np.random.rand(10, 3))
```

```
In [18]: d_['s'] = 'EBCWO ONCWEICNV iofce'
```

```
In [19]: list(filter(lambda x: x not in stop_words_dict['stopwords'], ['each', 'among', 'whatever']))
Out[19]: ['Bayern']
```

```
In [20]: ### START FUNCTION
def stop_words_remover(df):
    # your code here
    def fill(y):
        return [w for w in y if w not in stop_words_dict['stopwords']]
    df['Without Stop Words'] = df['Tweets'].str.lower().str.split(' ').apply(fill)
    return df
### END FUNCTION
```

```
In [21]: stop_words_remover(twitter_df.copy())
```

|     | Tweets                                            | Date                   | Without Stop Words                                 |
|-----|---------------------------------------------------|------------------------|----------------------------------------------------|
| 0   | @BongaDlulane Please send an email to mediades... | 2019-11-29<br>12:50:54 | [@bongadlulane, send, email, mediadesk@eskom.c...  |
| 1   | @saucy_mamiie Pls log a call on 0860037566        | 2019-11-29<br>12:46:53 | [@saucy_mamiie, pls, log, 0860037566]              |
| 2   | @BongaDlulane Query escalated to media desk.      | 2019-11-29<br>12:46:10 | [@bongadlulane, query, escalated, media, desk.]    |
| 3   | Before leaving the office this afternoon, head... | 2019-11-29<br>12:33:36 | [leaving, office, afternoon,, heading, weekend...] |
| 4   | #ESKOMFREESTATE #MEDIASTATEMENT : ESKOM SUSPEN... | 2019-11-29<br>12:17:43 | [#eskomfreestate, #mediastatement, ; eskom, s...   |
| ... | ...                                               | ...                    | ...                                                |
| 195 | Eskom's Visitors Centres' facilities include i... | 2019-11-20<br>10:29:07 | [eskom's, visitors, centres', facilities, incl...  |
| 196 | #Eskom connected 400 houses and in the process... | 2019-11-20<br>10:25:20 | [#eskom, connected, 400, houses, process, conn...  |
| 197 | @ArthurGodbeer Is the power restored as yet?      | 2019-11-20<br>10:07:59 | [@arthurgodbeer, power, restored, yet?]            |
| 198 | @MuthambiPaulina @SABCNewsOnline @IOL @eNCA @e... | 2019-11-20<br>10:07:41 | [@muthambipaulina, @sabcnewsonline, @iol, @enc...  |
| 199 | RT @GP_DHS: The @GautengProvince made a commit... | 2019-11-20<br>10:00:09 | [rt, @gp_dhs, @gautengprovince, commitment, e...   |

200 rows × 3 columns

### Expected Output:

Specific rows:

```
stop_words_remover(twitter_df.copy()).loc[0, "Without Stop Words"] ==
['@bongadlulane', 'send', 'email', 'mediadesk@eskom.co.za']
stop_words_remover(twitter_df.copy()).loc[100, "Without Stop Words"] ==
```

```
[ '#eskomnorthwest', '#mediastatement', ':', 'notice', 'supply',
'interruption', 'lichtenburg', 'area', 'https://t.co/7hfwvxllit' ]
```

Whole table:

```
stop_words_remover(twitter_df.copy())
```

|     | Tweets                                            | Date                | Without Stop Words                                 |
|-----|---------------------------------------------------|---------------------|----------------------------------------------------|
| 0   | @BongaDlulane Please send an email to mediades... | 2019-11-29 12:50:54 | [@bongadlulane, send, email, mediadesk@eskom.c...] |
| 1   | @saucy_mamiie Pls log a call on 0860037566        | 2019-11-29 12:46:53 | [@saucy_mamiie, pls, log, 0860037566]              |
| 2   | @BongaDlulane Query escalated to media desk.      | 2019-11-29 12:46:10 | [@bongadlulane, query, escalated, media, desk.]    |
| 3   | Before leaving the office this afternoon, head... | 2019-11-29 12:33:36 | [leaving, office, afternoon,, heading, weekend...] |
| 4   | #ESKOMFREESTATE #MEDIASTATEMENT : ESKOM SUSPEN... | 2019-11-29 12:17:43 | [#eskomfreestate, #mediastatement, :, eskom, s...] |
| ... | ...                                               | ...                 | ...                                                |
| 195 | Eskom's Visitors Centres' facilities include i... | 2019-11-20 10:29:07 | [eskom's, visitors, centres', facilities, incl...] |
| 196 | #Eskom connected 400 houses and in the process... | 2019-11-20 10:25:20 | [#eskom, connected, 400, houses, process, conn...] |
| 197 | @ArthurGodbeer Is the power restored as yet?      | 2019-11-20 10:07:59 | [@arthurgodbeer, power, restored, yet?]            |
| 198 | @MuthambiPaulina @SABCNewsOnline @IOL @eNCA @e... | 2019-11-20 10:07:41 | [@muthambipaulina, @sabcnewsonline, @iol, @enc...  |
| 199 | RT @GP_DHS: The @GautengProvince made a commit... | 2019-11-20 10:00:09 | [rt, @gp_dhs:, @gautengprovince, commitment, e...] |