

```
In [2]: import xlrd
import pandas as pd
import xlswriter
import glob
import os
import re
import numpy as np
import sys
```

Read xlsx files lists

```
In [ ]: xlsx_files = glob.glob('C:/Users/Rocku/Downloads/SeekingAlpha_EarningCalls/1-80000/*.xlsx')
len(xlsx_files)
# xlsx_files[:10]
```

```
In [ ]: def save_error_urls(error_data, error_log):
    path = 'error_logs.txt'

    with open(path, 'a') as f:
        f.write(error_data + ',' + error_log + '\n')
```

```

In [ ]: success = 0
        fail = 0

        for i, path in enumerate(xlsx_files):
            if (i % 1000 == 0):
                print(i, 'success:', success, 'fail:', fail)

            try:
                data = pd.read_excel(path, header=None, index_col=0, encoding='utf-8')
                regex1 = re.compile('\s\W\s')

                if 'Executives' in data.index and (data.loc[['Executives'],1].count() != 0):
                    data.loc['Executives',1] = data.loc['Executives',1].replace('&', 'and').replace('@', 'at')
                    if (len(data.loc[['Executives'],1]) == 1):
                        if (len(regex1.split(data.loc['Executives',1])) > 1):
                            data.loc['Executives',2] =
regex1.split(data.loc['Executives',1])[1]
                            data.loc['Executives',1] = regex1.split(data.loc['Executives',1])[0]
                        else:
                            data.loc['Executives',2] = data.loc['Executives',1].str.split(r'\s\W\s').str[1]
                            data.loc['Executives',1] = data.loc['Executives',1].str.split(r'\s\W\s').str[0]

                    if ('Analysts' in data.index) and (data.loc[['Analysts'],1].count() != 0):
                        data.loc['Analysts',1] = data.loc['Analysts',1].replace('&', 'and').replace('@', 'at')
                        if (len(data.loc[['Analysts'],1]) == 1):
                            if (len(regex1.split(data.loc['Analysts',1])) > 1):
                                data.loc['Analysts',2] = regex1.split(data.loc['Analysts',1])[1]
                                data.loc['Analysts',1] = regex1.split(data.loc['Analysts',1])[0]
                            else:
                                data.loc['Analysts',2] =
data.loc['Analysts',1].str.split(r'\s\W\s').str[1]
                                data.loc['Analysts',1] = data.loc['Analysts',1].str.split(r'\s\W\s').str[0]

```

```

        data.loc['Analysts',1] =
data.loc['Analysts',1].str.split(r'\s\\s').str[0]

        columns = ['FileName', 'CompanyName', 'Exchange', 'Ticker', 'CallDate', 'Uplo
adDate', 'Role', 'Position', 'Name']
        df = pd.DataFrame(columns=columns)

        FileName = path.split('\\')[-1]
        UploadDate = data.loc['UploadDate',1]
        CompanyName = data.loc['CompanyName',1]
        Exchange = data.loc['Exchange',1]
        Ticker = data.loc['Ticker',1]
        Subtitle = data.loc['Subtitle',1]
        CallDate = data.loc['CallDate',1]

        ex = data.loc[['Executives', 'Analysts'], :]

        for i in ex.iterrows():
            data_dict = {}
            data_dict['FileName'] = FileName
            data_dict['UploadDate'] = UploadDate
            data_dict['CompanyName'] = CompanyName
            data_dict['Exchange'] = Exchange
            data_dict['Ticker'] = Ticker
            data_dict['CallDate'] = CallDate
            data_dict['Role'] = i[0]
            data_dict['Name'] = i[1][1]
            data_dict['Position'] = i[1][2]

            regex = re.compile('\s+')
            if (len(data_dict['Exchange'].split(' ')) != 1):
                data_dict['Exchange'] = data_dict['Exchange'] + ':' + data_dict['Tick
er']
                data_dict['Ticker'] = regex.split(data_dict['Exchange'])
[0].replace(')', '')
                data_dict['CallDate'] = ' '.join(regex.split(data_dict['Exchange'])
[-6:])
                data_dict['Exchange'] = None

            if (pd.isnull(data_dict['Name']) == True):
                pass
            else:
                df1 = pd.DataFrame([data_dict], columns=columns)
                df = df.append(df1)

            if not os.path.isfile('index_file.csv'):
                df.to_csv('index_file.csv', header='column_names', index=False,
encoding='utf-8')
            else:
                df.to_csv('index_file.csv', mode='a', header=False, index=False,
encoding='utf-8')

            success += 1

        except:
            e = sys.exc_info()[1]
            save_error_urls(path.split('\\')[-1], str(e))
            fail += 1

print('-----end-----')

```

Error correction #1 - Remove the error data

```
In [ ]: path = 'C:/Users/Rocku/Dropbox/Data_analysis/seekingalpha/error_logs_upto80000.txt'

        bad_files = data = pd.read_table(path, header=-1, sep=',', encoding='utf-8')

In [ ]: bad_files.iloc[:,0] = 'C:/Users/Rocku/Downloads/SeekingAlpha_EarningCalls/1-80000\\'
        + bad_files.iloc[:,0]
        bad_file = bad_files.iloc[:,0].tolist()
        len(bad_file)

In [ ]: for raw_file in xlsx_files:
        if raw_file in bad_file:
            os.remove(raw_file)

In [ ]: path1 = 'C:/Users/Rocku/Dropbox/Data_analysis/seekingalpha/index_file_1.csv'
        path2 = 'C:/Users/Rocku/Dropbox/Data_analysis/seekingalpha/index_file_2.csv'

        df1 = pd.read_csv(path1)
        df2 = pd.read_csv(path2)

In [ ]: df1.count()

In [ ]: df2.count()

In [ ]: df1.loc[pd.isnull(df1['CompanyName']), ['Ticker']] = "TCKRF"
        df1.loc[pd.isnull(df1['CompanyName']), ['Exchange']] = "OTC"
        df1.loc[pd.isnull(df1['CompanyName']), ['CompanyName']] = "Teck Resources Ltd."

In [ ]: df2[pd.isnull(df2['CompanyName'])].count()

In [ ]: df2[~df2['FileName'].str.startswith('Seeking')]

In [ ]: error_data = df2.loc[pd.isnull(df2['CompanyName']), 'FileName']

In [ ]: error_data_list = list(set(error_data[error_data.str.startswith('Seeking')]))

In [ ]: len(error_data_list)
        error_data_list

In [ ]: df2 = df2[df2['FileName'].str.startswith('Seeking')]

In [ ]: df2 = df2[pd.notnull(df2['CompanyName'])]

In [ ]: df2 = df2[pd.notnull(df2['Name'])]

In [ ]: df2.count()
```

```
In [ ]: for raw_file in error_data_list:
        try:
            os.remove('C:/Users/Rocku/Downloads/SeekingAlpha_EarningCalls/1-80000\\'+raw_
file)
        except:
            print(raw_file)
```

Error correction #2 - Ticker contains lots of information --> split

```
In [2]: path1 = 'C:/Users/Rocku/Dropbox/Data_analysis/seekingalpha/index_file_1_revised.csv'
path2 = 'C:/Users/Rocku/Dropbox/Data_analysis/seekingalpha/index_file_2_revised.csv'

df1 = pd.read_csv(path1)
df2 = pd.read_csv(path2)
```

```
In [6]: df1.count()
```

```
Out[6]: FileName      613940
CompanyName    613940
Exchange       591234
Ticker         613940
CallDate       611077
UploadDate     613940
Role           613940
Position       605875
Name           613940
dtype: int64
```

```
In [7]: df2.count()
```

```
Out[7]: FileName      485163
CompanyName    485163
Exchange       424010
Ticker         485146
CallDate       483393
UploadDate     485163
Role           485163
Position       468545
Name           485163
dtype: int64
```

```
In [8]: df1[pd.notnull(df1['Ticker'].str.split(' ').str[1])].count()
```

```
Out[8]: FileName      52343
CompanyName    52343
Exchange       52343
Ticker         52343
CallDate       49482
UploadDate     52343
Role           52343
Position       52062
Name           52343
dtype: int64
```

```
In [9]: df1[pd.notnull(df1['Ticker'].str.split('\s').str[1])]
```

Out[9]:

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
714	SeekingAlpha_1397181.xlsx	Canfor Pulp Products, Inc.	OTC	CFPUF Q1 2013 Earnings Call May 2, 2013 11	Donald B. Kayne – Chief Executive Officer	2013-05-02 21:15:04 -0400
715	SeekingAlpha_1397181.xlsx	Canfor Pulp Products, Inc.	OTC	CFPUF Q1 2013 Earnings Call May 2, 2013 11	Donald B. Kayne – Chief Executive Officer	2013-05-02 21:15:04 -0400
716	SeekingAlpha_1397181.xlsx	Canfor Pulp Products, Inc.	OTC	CFPUF Q1 2013 Earnings Call May 2, 2013 11	Donald B. Kayne – Chief Executive Officer	2013-05-02 21:15:04 -0400
717	SeekingAlpha_1397181.xlsx	Canfor Pulp Products, Inc.	OTC	CFPUF Q1 2013 Earnings Call May 2, 2013 11	Donald B. Kayne – Chief Executive Officer	2013-05-02 21:15:04 -0400
718	SeekingAlpha_1397181.xlsx	Canfor Pulp Products, Inc.	OTC	CFPUF Q1 2013 Earnings Call May 2, 2013 11	Donald B. Kayne – Chief Executive Officer	2013-05-02 21:15:04 -0400
719	SeekingAlpha_1397181.xlsx	Canfor Pulp Products, Inc.	OTC	CFPUF Q1 2013 Earnings Call May 2, 2013 11	Donald B. Kayne – Chief Executive Officer	2013-05-02 21:15:04 -0400
720	SeekingAlpha_1397181.xlsx	Canfor Pulp Products, Inc.	OTC	CFPUF Q1 2013 Earnings Call May 2, 2013 11	Donald B. Kayne – Chief Executive Officer	2013-05-02 21:15:04 -0400
721	SeekingAlpha_1397181.xlsx	Canfor Pulp Products, Inc.	OTC	CFPUF Q1 2013 Earnings Call May 2, 2013 11	Donald B. Kayne – Chief Executive Officer	2013-05-02 21:15:04 -0400
762	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
763	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
764	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
765	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
766	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
767	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
768	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
769	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
770	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
771	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
772	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
773	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
774	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
775	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
776	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
777	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
778	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
779	SeekingAlpha_1397241.xlsx	BG Group	OTCQX	BRGY Q1 2013 Earnings Call May 2, 2013 7	Chris Finlayson - Chief Executive Officer, Man...	2013-05-02 21:20:07 -0400
1252	SeekingAlpha_1397901.xlsx	Superior Plus Corp	OTC	SUUIF Q1 2013 Earnings Call May 2, 2013 10	Luc Desjardins - President and Chief Executive...	2013-05-03 01:26:05 -0400
1253	SeekingAlpha_1397901.xlsx	Superior Plus Corp	OTC	SUUIF Q1 2013 Earnings Call May 2, 2013 10	Luc Desjardins - President and Chief Executive...	2013-05-03 01:26:05 -0400
1254	SeekingAlpha_1397901.xlsx	Superior Plus Corp	OTC	SUUIF Q1 2013 Earnings Call May 2, 2013 10	Luc Desjardins - President and Chief Executive...	2013-05-03 01:26:05 -0400
1255	SeekingAlpha_1397901.xlsx	Superior Plus Corp	OTC	SUUIF Q1 2013 Earnings Call May 2, 2013 10	Luc Desjardins - President and Chief Executive...	2013-05-03 01:26:05 -0400
...
613844	SeekingAlpha_4085639.xlsx	J Sainsbury plc	OTCQX	JSNSF Q1 2017/2018 Trading Update Earnings Con...	Kevin O'Byrne - Chief Financial Officer	2017-07-04 10:10:06 -0400
613845	SeekingAlpha_4085639.xlsx	J Sainsbury plc	OTCQX	JSNSF Q1 2017/2018 Trading Update Earnings Con...	Kevin O'Byrne - Chief Financial Officer	2017-07-04 10:10:06 -0400
613846	SeekingAlpha_4085639.xlsx	J Sainsbury plc	OTCQX	JSNSF Q1 2017/2018 Trading Update Earnings Con...	Kevin O'Byrne - Chief Financial Officer	2017-07-04 10:10:06 -0400

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
613847	SeekingAlpha_4085639.xlsx	J Sainsbury plc	OTCQX	JSNSF Q1 2017/2018 Trading Update Earnings Con...	Kevin O'Byrne - Chief Financial Officer	2017-07-04 10:10:06 -0400
613848	SeekingAlpha_4085639.xlsx	J Sainsbury plc	OTCQX	JSNSF Q1 2017/2018 Trading Update Earnings Con...	Kevin O'Byrne - Chief Financial Officer	2017-07-04 10:10:06 -0400
613849	SeekingAlpha_4085818.xlsx	Ocado Group plc	OTCPK	OCDGF Q2 2017 Earnings Conference Call July 5...	Stuart Rose - Non- Executive Chairman	2017-07-05 11:30:43 -0400
613850	SeekingAlpha_4085818.xlsx	Ocado Group plc	OTCPK	OCDGF Q2 2017 Earnings Conference Call July 5...	Stuart Rose - Non- Executive Chairman	2017-07-05 11:30:43 -0400
613851	SeekingAlpha_4085818.xlsx	Ocado Group plc	OTCPK	OCDGF Q2 2017 Earnings Conference Call July 5...	Stuart Rose - Non- Executive Chairman	2017-07-05 11:30:43 -0400
613852	SeekingAlpha_4085818.xlsx	Ocado Group plc	OTCPK	OCDGF Q2 2017 Earnings Conference Call July 5...	Stuart Rose - Non- Executive Chairman	2017-07-05 11:30:43 -0400
613853	SeekingAlpha_4085818.xlsx	Ocado Group plc	OTCPK	OCDGF Q2 2017 Earnings Conference Call July 5...	Stuart Rose - Non- Executive Chairman	2017-07-05 11:30:43 -0400
613854	SeekingAlpha_4085818.xlsx	Ocado Group plc	OTCPK	OCDGF Q2 2017 Earnings Conference Call July 5...	Stuart Rose - Non- Executive Chairman	2017-07-05 11:30:43 -0400

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
613855	SeekingAlpha_4085818.xlsx	Ocado Group plc	OTCPK	OCDGF Q2 2017 Earnings Conference Call July 5...	Stuart Rose - Non-Executive Chairman	2017-07-05 11:30:43 -0400
613856	SeekingAlpha_4085818.xlsx	Ocado Group plc	OTCPK	OCDGF Q2 2017 Earnings Conference Call July 5...	Stuart Rose - Non-Executive Chairman	2017-07-05 11:30:43 -0400
613857	SeekingAlpha_4085818.xlsx	Ocado Group plc	OTCPK	OCDGF Q2 2017 Earnings Conference Call July 5...	Stuart Rose - Non-Executive Chairman	2017-07-05 11:30:43 -0400
613858	SeekingAlpha_4085818.xlsx	Ocado Group plc	OTCPK	OCDGF Q2 2017 Earnings Conference Call July 5...	Stuart Rose - Non-Executive Chairman	2017-07-05 11:30:43 -0400
613859	SeekingAlpha_4085818.xlsx	Ocado Group plc	OTCPK	OCDGF Q2 2017 Earnings Conference Call July 5...	Stuart Rose - Non-Executive Chairman	2017-07-05 11:30:43 -0400
613892	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400
613893	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400
613894	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
613895	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400
613896	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400
613897	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400
613898	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400
613899	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400
613900	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400
613901	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400
613902	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
613903	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400
613904	SeekingAlpha_4086142.xlsx	SODEXO	OTC	SDXOF Q3 2017 Earnings Conference Call July 6...	Virginia Jeanson - IR	2017-07-06 15:20:45 -0400
613928	SeekingAlpha_4086177.xlsx	TAG Oil Ltd.	OTCQX	TAOIF Q4 2017 Results Earnings Conference Call...	Toby Pierce - CEO	2017-07-06 18:49:05 -0400

52343 rows × 9 columns



```
In [10]: df1.loc[pd.notnull(df1['Ticker'].str.split(' ').str[1]), 'CallDate'] = \
df1.loc[pd.notnull(df1['Ticker'].str.split(' ').str[1]),
'Ticker'].str.split(r'\s+').str[-4:].str.join(' ')

df1.loc[pd.notnull(df1['Ticker'].str.split(' ').str[1]), 'Ticker'] = \
df1.loc[pd.notnull(df1['Ticker'].str.split(' ').str[1]),
'Ticker'].str.split(r'\s+').str[0]
```

```
In [11]: df1[pd.notnull(df1['Ticker'].str.split(' ').str[1])].count()
```

```
Out[11]: FileName      0
CompanyName    0
Exchange       0
Ticker        0
CallDate      0
UploadDate    0
Role          0
Position      0
Name          0
dtype: int64
```

```
In [12]: df2[pd.notnull(df2['Ticker'].str.split(' ').str[2])].count()
```

```
Out[12]: FileName      8571
CompanyName    8571
Exchange       8571
Ticker        8571
CallDate      6827
UploadDate    8571
Role          8571
Position      8273
Name          8571
dtype: int64
```

```
In [13]: df2[pd.notnull(df2['Ticker'].str.split(' ').str[2])]
```

Out[13]:

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
1328	SeekingAlpha_1003571.xlsx	Petrobank Energy and Resources Ltd.	OTCPK	PBEGF Q3 2012 Earnings Call November 13, 2012 10	John D. Wright ?? President, Chief Executive Of...	2012-11-13 13:35
1329	SeekingAlpha_1003571.xlsx	Petrobank Energy and Resources Ltd.	OTCPK	PBEGF Q3 2012 Earnings Call November 13, 2012 10	John D. Wright ?? President, Chief Executive Of...	2012-11-13 13:35
1330	SeekingAlpha_1003571.xlsx	Petrobank Energy and Resources Ltd.	OTCPK	PBEGF Q3 2012 Earnings Call November 13, 2012 10	John D. Wright ?? President, Chief Executive Of...	2012-11-13 13:35
1331	SeekingAlpha_1003571.xlsx	Petrobank Energy and Resources Ltd.	OTCPK	PBEGF Q3 2012 Earnings Call November 13, 2012 10	John D. Wright ?? President, Chief Executive Of...	2012-11-13 13:35
1332	SeekingAlpha_1003571.xlsx	Petrobank Energy and Resources Ltd.	OTCPK	PBEGF Q3 2012 Earnings Call November 13, 2012 10	John D. Wright ?? President, Chief Executive Of...	2012-11-13 13:35
1573	SeekingAlpha_1004401.xlsx	Aurcana Corporation	OTCPK	AUNFF Q3 2012 Earnings Call November 13, 2012 2	Catalin Chiloflisch - Manager, Investor Commu...	2012-11-13 16:21
1574	SeekingAlpha_1004401.xlsx	Aurcana Corporation	OTCPK	AUNFF Q3 2012 Earnings Call November 13, 2012 2	Catalin Chiloflisch - Manager, Investor Commu...	2012-11-13 16:21
1575	SeekingAlpha_1004401.xlsx	Aurcana Corporation	OTCPK	AUNFF Q3 2012 Earnings Call November 13, 2012 2	Catalin Chiloflisch - Manager, Investor Commu...	2012-11-13 16:21

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
1612	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1613	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1614	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1615	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1616	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1617	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1618	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
1619	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1620	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1621	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1622	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1623	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1624	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1625	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
1626	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1627	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1628	SeekingAlpha_1004631.xlsx	Vivendi	OTC	VIVDY Q3 2012 Earnings Call November 13, 2012 12	Jean-Michel Bonamy - Executive Vice President ...	2012-11-13 17:00
1668	SeekingAlpha_1004871.xlsx	Eon Ag Ads	OTCPK	EONGY Q3 2012 Earnings Conference Call November...	Dr. Johannes Teyssen ?? Chairman and Chief Exec...	2012-11-13 18:31
1669	SeekingAlpha_1004871.xlsx	Eon Ag Ads	OTCPK	EONGY Q3 2012 Earnings Conference Call November...	Dr. Johannes Teyssen ?? Chairman and Chief Exec...	2012-11-13 18:31
1670	SeekingAlpha_1004871.xlsx	Eon Ag Ads	OTCPK	EONGY Q3 2012 Earnings Conference Call November...	Dr. Johannes Teyssen ?? Chairman and Chief Exec...	2012-11-13 18:31
1671	SeekingAlpha_1004871.xlsx	Eon Ag Ads	OTCPK	EONGY Q3 2012 Earnings Conference Call November...	Dr. Johannes Teyssen ?? Chairman and Chief Exec...	2012-11-13 18:31

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
1672	SeekingAlpha_1004871.xlsx	Eon Ag Ads	OTCPK	EONGY Q3 2012 Earnings Conference Call November...	Dr. Johannes Teyssen ?? Chairman and Chief Exec...	2012-11-13 18:31
...
483553	SeekingAlpha_993651.xlsx	Repsol SA	OTCQX	REPLY Q3 2012 Earnings Call November 8, 2012 7	Maria Victoria Zingoni ?? Investor Relations Di...	2012-11-08 21:47
483554	SeekingAlpha_993651.xlsx	Repsol SA	OTCQX	REPLY Q3 2012 Earnings Call November 8, 2012 7	Maria Victoria Zingoni ?? Investor Relations Di...	2012-11-08 21:47
483641	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44
483642	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44
483643	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44
483644	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
483645	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44
483646	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44
483647	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44
483648	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44
483649	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44
483650	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44
483651	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44
483652	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
483653	SeekingAlpha_993861.xlsx	Ecosphere Technologies, Inc.	OTCQB	ESPH Q3 2012 Earnings Call November 8, 2012 4	NaN	2012-11-08 22:44
484008	SeekingAlpha_995221.xlsx	Hickory Tech Corporation	OTCPK	HTCO Q3 2012 Earnings Call November 9, 2012 10	NaN	2012-11-09 13:18
484009	SeekingAlpha_995221.xlsx	Hickory Tech Corporation	OTCPK	HTCO Q3 2012 Earnings Call November 9, 2012 10	NaN	2012-11-09 13:18
484010	SeekingAlpha_995221.xlsx	Hickory Tech Corporation	OTCPK	HTCO Q3 2012 Earnings Call November 9, 2012 10	NaN	2012-11-09 13:18
484011	SeekingAlpha_995221.xlsx	Hickory Tech Corporation	OTCPK	HTCO Q3 2012 Earnings Call November 9, 2012 10	NaN	2012-11-09 13:18
484092	SeekingAlpha_995391.xlsx	International Stem Cell	OTCQB	ISCO Q3 2012 Earnings Call November 9, 2012 11	Mark McPartland	2012-11-09 13:40
484093	SeekingAlpha_995391.xlsx	International Stem Cell	OTCQB	ISCO Q3 2012 Earnings Call November 9, 2012 11	Mark McPartland	2012-11-09 13:40
484094	SeekingAlpha_995391.xlsx	International Stem Cell	OTCQB	ISCO Q3 2012 Earnings Call November 9, 2012 11	Mark McPartland	2012-11-09 13:40

	FileName	CompanyName	Exchange	Ticker	CallDate	UploadDate
484746	SeekingAlpha_997321.xlsx	Gasfrac Energy Services, Inc.	OTC	GSFVF Q3 2012 Earnings Call November 8, 2012 11	NaN	2012-11-09 22:21
484747	SeekingAlpha_997321.xlsx	Gasfrac Energy Services, Inc.	OTC	GSFVF Q3 2012 Earnings Call November 8, 2012 11	NaN	2012-11-09 22:21
484748	SeekingAlpha_997321.xlsx	Gasfrac Energy Services, Inc.	OTC	GSFVF Q3 2012 Earnings Call November 8, 2012 11	NaN	2012-11-09 22:21
484749	SeekingAlpha_997321.xlsx	Gasfrac Energy Services, Inc.	OTC	GSFVF Q3 2012 Earnings Call November 8, 2012 11	NaN	2012-11-09 22:21
484750	SeekingAlpha_997321.xlsx	Gasfrac Energy Services, Inc.	OTC	GSFVF Q3 2012 Earnings Call November 8, 2012 11	NaN	2012-11-09 22:21
484751	SeekingAlpha_997321.xlsx	Gasfrac Energy Services, Inc.	OTC	GSFVF Q3 2012 Earnings Call November 8, 2012 11	NaN	2012-11-09 22:21
484752	SeekingAlpha_997321.xlsx	Gasfrac Energy Services, Inc.	OTC	GSFVF Q3 2012 Earnings Call November 8, 2012 11	NaN	2012-11-09 22:21
484753	SeekingAlpha_997321.xlsx	Gasfrac Energy Services, Inc.	OTC	GSFVF Q3 2012 Earnings Call November 8, 2012 11	NaN	2012-11-09 22:21

8571 rows × 9 columns



```
In [14]: df2.loc[pd.notnull(df2['Ticker'].str.split(' ').str[2]), 'CallDate'] = \
df2.loc[pd.notnull(df2['Ticker'].str.split(' ').str[2]),
'Ticker'].str.split(r'\s+').str[-4:].str.join(' ')

df2.loc[pd.notnull(df2['Ticker'].str.split(' ').str[2]), 'Ticker'] = \
df2.loc[pd.notnull(df2['Ticker'].str.split(' ').str[2]),
'Ticker'].str.split(r'\s+').str[0]
```

```
In [15]: df2[pd.notnull(df2['Ticker'].str.split(' ').str[2])].count()
```

```
Out[15]: FileName      0
CompanyName    0
Exchange       0
Ticker         0
CallDate       0
UploadDate     0
Role           0
Position       0
Name           0
dtype: int64
```

```
In [16]: df1.to_csv('index_file_1_revised2.csv', header='column_names', index=False,
encoding='utf-8')
```

```
In [17]: df2.to_csv('index_file_2_revised2.csv', header='column_names', index=False,
encoding='utf-8')
```

```
In [18]: del df1, df2
```

Concatenate two index files

```
In [19]: path1 = 'C:/Users/Rocku/Dropbox/Data_analysis/seekingalpha/index_file_1_revised2.csv'
path2 = 'C:/Users/Rocku/Dropbox/Data_analysis/seekingalpha/index_file_2_revised2.csv'

df1 = pd.read_csv(path1)
df2 = pd.read_csv(path2)
```

```
In [20]: df2.rename(columns={'\ufeffFileName': 'FileName'}, inplace=True)
```

```
In [21]: df = pd.concat([df1, df2])
```

```
In [ ]: df.to_csv('index_file.csv', header='column_names', index=False, encoding='utf-8')
```

Error correction #3 - remove 'None' Ticker file


```
In [390]: path1 = 'C:/Users/Rocku/Dropbox/Data_analysis/seekingalpha/index_file.csv'
df1 = pd.read_csv(path1)
df1.count()
```

```
Out[390]: FileName      1099103
CompanyName    1099103
Exchange       1015244
Ticker         1099059
CallDate       1099075
UploadDate     1099103
Role           1099103
Position       1074420
Name           1099103
dtype: int64
```

```
In [391]: df1[df1['Ticker'] == 'None'].count()
```

```
Out[391]: FileName      15362
CompanyName    15362
Exchange       15362
Ticker         15362
CallDate       15343
UploadDate     15362
Role           15362
Position       15015
Name           15362
dtype: int64
```

```
In [392]: df = df1[~(df1['Ticker'] == 'None')]
df.count()
```

```
Out[392]: FileName      1083741
CompanyName    1083741
Exchange       999882
Ticker         1083697
CallDate       1083732
UploadDate     1083741
Role           1083741
Position       1059405
Name           1083741
dtype: int64
```

```
In [393]: df.to_csv('index_file2.csv', header='column_names', index=False, encoding='utf-8')
```

```
In [394]: none_data_list = list(set(df1.loc[df1['Ticker'] == 'None', 'FileName']))
len(none_data)
```

```
Out[394]: 1728
```

```
In [398]: cant_remove = []

for raw_file in none_data_list:
    try:
        os.remove('C:/Users/Rocku/Downloads/SeekingAlpha_EarningCalls/raw_xlsx/' + raw_file)
    except:
        cant_remove.append(raw_file)

len(cant_remove)
```

```
Out[398]: 0
```

Divide into two index files

```
In [3]: path1 = 'C:/Users/Rocku/Downloads/SeekingAlpha_EarningCalls/index_file.csv'

df1 = pd.read_csv(path1)
```

```
In [4]: df1.count()
```

```
Out[4]: FileName      1083741
CompanyName    1083741
Exchange        999882
Ticker         1083697
CallDate       1083732
UploadDate     1083741
Role           1083741
Position       1059405
Name           1083741
dtype: int64
```

```
In [18]: df_first = df1[:599984]
df_second = df1[599984:]
```

```
In [22]: df_first.to_csv('index_file_p1.csv', header='column_names', index=False, encoding='utf-8')
df_second.to_csv('index_file_p2.csv', header='column_names', index=False, encoding='utf-8')
```