```
In [1]: import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        import numpy as np
        from scipy.stats import norm
        from sklearn.preprocessing import StandardScaler
        from scipy import stats
        import warnings
        import seaborn as sns
        import sys

        warnings.filterwarnings('ignore')
        %matplotlib inline
```

Loading Data

```
In [2]: df_data = pd.read_csv('connectivity_task.csv')
```

```
In [3]: df_data.head()
```

Out[3]:

	event_time	event_type	organisation_name	place_name	asset_name	asset_type	module_id
0	2021-02-21 05:53:05	Connected	Gonzalez-Hancock	Thompson Group	EB-FVF-50	5	NaN
1	2021-02-19 05:53:02	Disconnected	Larson-Mccall	Cohen PLC	FI-DMS-52	8	NaN
2	2021-02-19 10:53:05	Connected	Larson-Mccall	Cohen PLC	GV-WTD-14	8	NaN
3	2021-02-22 18:30:09	Connected	Taylor, Flores and Douglas	Matthews-Phillips	PD-PFG-92	3	NaN
4	2021-02-16 09:31:11	Connected	Larson-Mccall	Benson Ltd	MO-RQJ-49	8	NaN

```
In [4]: df_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19559 entries, 0 to 19558
Data columns (total 7 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   event_time         19537 non-null  object
 1   event_type         19537 non-null  object
 2   organisation_name  19559 non-null  object
 3   place_name         19559 non-null  object
 4   asset_name         19559 non-null  object
 5   asset_type         19559 non-null  int64
 6   module_id          1531 non-null   object
dtypes: int64(1), object(6)
memory usage: 1.0+ MB
```

- Number of null values

```
In [5]: nan_values = len(df_data) - df_data.count()
        print(nan_values)
```

```
event_time               22
event_type               22
organisation_name         0
place_name                0
asset_name                0
asset_type                0
module_id             18028
dtype: int64
```

Data Cleaning

- Drop rows with nan values at event_time & event_type

```
In [6]: print('number of rows that have nan in event_time column: ', len(df_data[df_data['event_time'].isnull
        ()]))
        rows_with_nan_eventime = df_data[df_data['event_time'].isnull()]
        # drop rows with nan values in event_time
        df_data = df_data[df_data['event_time'].notna()]
        print('number of rows without nan in event_time',len(df_data))
```

```
number of rows that have nan in event_time column:  22
number of rows without nan in event_time 19537
```

- Convert event_time from string to pandas datetime

```
In [7]: df_data["event_time"]= pd.to_datetime(df_data["event_time"],  format= "%Y-%m-%d %H:%M:%S")
```

- set event_time to be the index

```
In [8]: df_data.set_index('event_time', inplace=True)
```

```
In [9]: df_data.head()
```

Out[9]:

event_time	event_type	organisation_name	place_name	asset_name	asset_type	module_id
2021-02-21 05:53:05	Connected	Gonzalez-Hancock	Thompson Group	EB-FVF-50	5	NaN
2021-02-19 05:53:02	Disconnected	Larson-Mccall	Cohen PLC	FI-DMS-52	8	NaN
2021-02-19 10:53:05	Connected	Larson-Mccall	Cohen PLC	GV-WTD-14	8	NaN
2021-02-22 18:30:09	Connected	Taylor, Flores and Douglas	Matthews-Phillips	PD-PFG-92	3	NaN
2021-02-16 09:31:11	Connected	Larson-Mccall	Benson Ltd	MO-RQJ-49	8	NaN

```
In [10]: df_data['asset_type'].unique()
```

```
Out[10]: array([5, 8, 3], dtype=int64)
```

```
In [11]: df_data.describe(include='all')
```

Out[11]:

	event_type	organisation_name	place_name	asset_name	asset_type	module_id
count	19537	19537	19537	19537	19537.000000	1531
unique	2	3	33	60	NaN	9
top	Connected	Larson-Mccall	Cohen PLC	ND-IRJ-15	NaN	598-84-6937
freq	13949	8127	7484	4571	NaN	1173
mean	NaN	NaN	NaN	NaN	6.488714	NaN
std	NaN	NaN	NaN	NaN	2.144644	NaN
min	NaN	NaN	NaN	NaN	3.000000	NaN
25%	NaN	NaN	NaN	NaN	5.000000	NaN
50%	NaN	NaN	NaN	NaN	8.000000	NaN
75%	NaN	NaN	NaN	NaN	8.000000	NaN
max	NaN	NaN	NaN	NaN	8.000000	NaN

```
In [12]: df_data.describe(include = ['O'])
```

Out[12]:

	event_type	organisation_name	place_name	asset_name	module_id
count	19537	19537	19537	19537	1531
unique	2	3	33	60	9
top	Connected	Larson-Mccall	Cohen PLC	ND-IRJ-15	598-84-6937
freq	13949	8127	7484	4571	1173

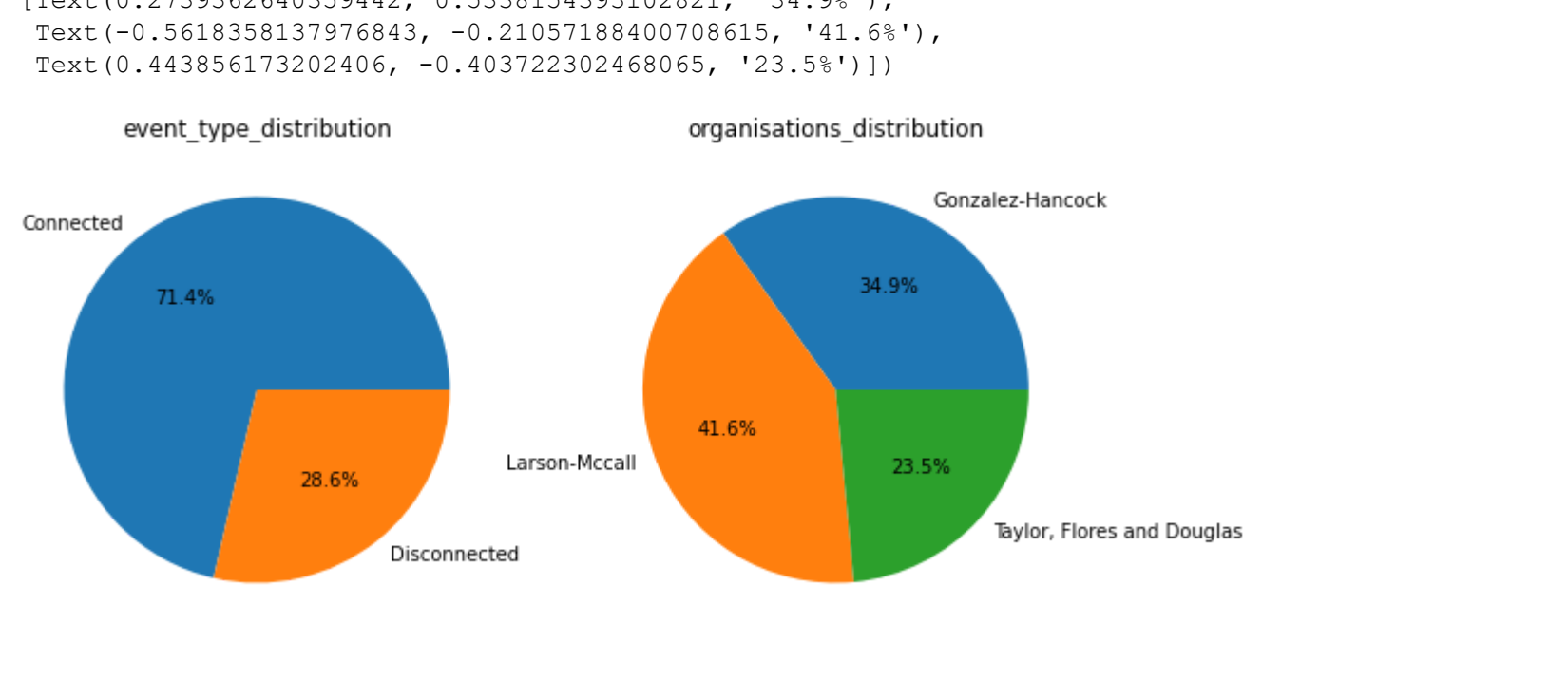
Data Visualization

```
In [13]: event_type_distribution = df_data.groupby('event_type').size()
         print(event_type_distribution)
         labels = ["Connected", "Disconnected"]
         fig1, (ax1,ax2) = plt.subplots(1,2,figsize=(10,10))
         ax1.title.set_text('event_type_distribution')
         ax1.pie(event_type_distribution, labels=labels, autopct='%1.1f%%')

         organisations_distribution = df_data.groupby('organisation_name').size()
         labels2 = df_data['organisation_name'].unique()
         ax2.title.set_text('organisations_distribution')
         ax2.pie(organisations_distribution, labels=labels2, autopct='%1.1f%%')
```

```
event_type
Connected       13949
Disconnected     5588
dtype: int64
```

```
Out[13]: ([<matplotlib.patches.Wedge at 0x1602c4e7760>,
           <matplotlib.patches.Wedge at 0x1602c4e7c10>,
           <matplotlib.patches.Wedge at 0x1602c4f62e0>],
          [Text(0.5022164840658977, 0.9786616387355173, 'Gonzalez-Hancock'),
           Text(-1.0300323252957546, -0.3860484540129913, 'Larson-Mccall'),
           Text(0.8137363175377444, -0.7401575545247859, 'Taylor, Flores and Douglas')],
          [Text(0.2739362640359442, 0.5338154393102821, '34.9%'),
           Text(-0.5618358137976843, -0.21057188400708615, '41.6%'),
           Text(0.443856173202406, -0.403722302468065, '23.5%')])
```



Show time window of the data

```
In [14]: start_date = df_data.index.min()
         end_date = df_data.index.max()
         difference_of_date = end_date - start_date
         print('the time window of the data is: ', difference_of_date)
         print('First record starts at:', start_date)
         print('Last record ends at: ', end_date)
```

```
the time window of the data is:  7 days 19:48:23
First record starts at: 2021-02-15 14:46:00
Last record ends at:  2021-02-23 10:34:23
```

Deep dive into the Data

In [15]:
```python
# Split data by organisation_name
orgs = df_data['organisation_name'].unique()
plt.figure(figsize=(15, 3))
colors = ['b', 'c', 'y', 'm', 'r']

for organisation in orgs:
    org_data = df_data.loc[df_data['organisation_name']==organisation]
    places_within_org = org_data['place_name'].unique()
    print('*****************************************************************************************
**********************')
    print('Organisation name is: ', organisation)
    print(org_data.describe(include = ['O']))

    org_distribution = org_data.groupby('place_name').size()
    labels2 = places_within_org
    if(organisation == 'Taylor, Flores and Douglas'):
        fig1, ax2 = plt.subplots(figsize=(15,20))
    else:
        fig1, ax2 = plt.subplots(figsize=(15,5))
    plc_titles = 'Places within Organisation:' + organisation
    ax2.title.set_text(plc_titles)
    ax2.pie(org_distribution, labels=labels2, autopct='%1.1f%%')

    for place in places_within_org:
        print('--------------------------------------------------------------------------------')
        print('\nplace: ', place)
        plc_data = org_data.loc[org_data['place_name']==place]
        asset_names_within_place = plc_data['asset_name'].unique()

        plc_distribution = plc_data.groupby('asset_name').size()
        labels2 = asset_names_within_place

        fig1, ax2 = plt.subplots(figsize=(15,5))
        asset_titles = 'asset_names within Place:' + place
        ax2.title.set_text(asset_titles)
        ax2.pie(plc_distribution, labels=labels2, autopct='%1.1f%%')

#         print('number of asset_name within_place is :',len(asset_names_within_place))
        print('asset_names_within_place: ', asset_names_within_place)
        plt.figure(figsize=(15, 3))
        asset_types_within_place = plc_data['asset_type'].unique()
        asset_names_within_place = plc_data['asset_name'].unique()
#         module_id__within_place = plc_data.groupby('module_id').size()
        module_id__within_place = plc_data['module_id'].unique()
        print('asset_types_within_place: ', asset_types_within_place)
        print('module_id__within_place: ', module_id__within_place)
        for asset_name in asset_names_within_place:
            asset_data = plc_data.loc[plc_data['asset_name']==asset_name]
            asset_data['event_type'] = asset_data['event_type'].replace(['Connected','Disconnected'],[1
,0])

            graph_title = 'Organisation name is: '+ str(organisation)+', place: '+str(place) + ' with a
sset_name: '\
                    + str(asset_names_within_place) + ' and asset_types: '+ str(asset_types_within_plac
e)
            plt.title(graph_title)
            asset_types_within_asset_name = asset_data['asset_type'].unique()
            for asset_type in asset_types_within_asset_name:
                asset_type_data = asset_data.loc[asset_data['asset_type']==asset_type]
                if(asset_type == 8):
                    clr = 'b'
                elif(asset_type == 5):
                    clr = 'r'
                else:
                    clr = 'y'
                plt.scatter(asset_type_data.index, asset_type_data['event_type'], color=clr)
                plt.legend(asset_types_within_place)
```

```
*****************************************************************************************
**************
Organisation name is:  Gonzalez-Hancock
       event_type organisation_name     place_name asset_name    module_id
count        6820              6820           6820       6820         1531
unique          2                 1              4          8            9
top     Connected  Gonzalez-Hancock  Flynn-Mcgrath  ND-IRJ-15  598-84-6937
freq         5978              6820           4571       4571         1173
--------------------------------------------------------------------------------

place:  Thompson Group
asset_names_within_place:  ['EB-FVF-50' 'SL-UAD-50' 'CH-QKR-84']
asset_types_within_place:  [5 8]
module_id__within_place:  [nan '591-45-2627' '647-80-6730' '598-84-6937' '351-86-2983' '877-10-5115'
 '509-27-8182' '324-50-2046' '478-66-3754']
--------------------------------------------------------------------------------

place:  Flynn-Mcgrath
asset_names_within_place:  ['ND-IRJ-15']
asset_types_within_place:  [8]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Torres, Montoya and Neal
asset_names_within_place:  ['NW-CNY-73']
asset_types_within_place:  [5]
module_id__within_place:  [nan '598-84-6937' '877-10-5115' '324-50-2046' '647-80-6730' '478-66-3754'
 '509-27-8182' '591-45-2627' '351-86-2983']
--------------------------------------------------------------------------------

place:  Bell LLC
asset_names_within_place:  ['OQ-XNB-63' 'OZ-LHY-83' 'IC-INE-34']
asset_types_within_place:  [5 8]
module_id__within_place:  ['598-84-6937' nan '351-86-2983' '324-50-2046' '877-10-5115' '444-13-0230'
 '591-45-2627' '509-27-8182' '647-80-6730' '478-66-3754']
*****************************************************************************************
**************
Organisation name is:  Larson-Mccall
         event_type organisation_name place_name asset_name module_id
count          8127              8127       8127       8127         0
unique            2                 1          3          8         0
top     Disconnected     Larson-Mccall  Cohen PLC  GV-WTD-14       NaN
freq           4175              8127       7484       3822       NaN
--------------------------------------------------------------------------------

place:  Cohen PLC
asset_names_within_place:  ['FI-DMS-52' 'GV-WTD-14']
asset_types_within_place:  [8]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Benson Ltd
asset_names_within_place:  ['MO-RQJ-49' 'GR-TDN-28' 'XX-PXF-22']
asset_types_within_place:  [8]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Aguilar, Price and Clark
asset_names_within_place:  ['CJ-KWS-16' 'BM-LAS-19' 'MC-FRO-53']
asset_types_within_place:  [8]
module_id__within_place:  [nan]
*****************************************************************************************
**************
Organisation name is:  Taylor, Flores and Douglas
       event_type           organisation_name                  place_name  \
count        4590                        4590                        4590
unique          2                           1                          26
top     Connected  Taylor, Flores and Douglas  Bell, Turner and Rodriguez
freq         4019                        4590                         574

       asset_name module_id
count        4590         0
unique         44         0
top     EI-IUO-83       NaN
freq          574       NaN
--------------------------------------------------------------------------------

place:  Matthews-Phillips
asset_names_within_place:  ['PD-PFG-92' 'YW-QRR-91']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Olson Inc
asset_names_within_place:  ['MA-REN-12' 'US-KJX-96']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Torres-Wilson
asset_names_within_place:  ['DA-IXV-17' 'XO-BQH-64']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Hart, Mason and Smith
asset_names_within_place:  ['WN-DQW-72' 'LR-DNI-18']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Martinez, Pruitt and Wheeler
asset_names_within_place:  ['KN-GWX-65']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Cantu-Lane
asset_names_within_place:  ['DT-XSY-00' 'PL-PSP-39']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Brown-Jenkins
asset_names_within_place:  ['CF-MQK-54' 'CK-XNN-42']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Knox, Mills and Keith
asset_names_within_place:  ['WT-JZS-58']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Wade, Ruiz and Harrison
asset_names_within_place:  ['KL-KYK-17' 'FG-YFO-64']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Bell, Turner and Rodriguez
asset_names_within_place:  ['EI-IUO-83']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Knight PLC
asset_names_within_place:  ['XQ-ZSS-50']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Young, Fox and Bryant
asset_names_within_place:  ['UM-XEZ-99']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Ochoa-Chavez
asset_names_within_place:  ['XZ-XXC-35' 'AI-CXD-88']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Booth, Gregory and Saunders
asset_names_within_place:  ['ZU-MOT-42']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Gallagher-Moore
asset_names_within_place:  ['WF-QFJ-09' 'IO-MMI-57']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Alvarado-Alexander
asset_names_within_place:  ['QT-LTC-56' 'US-YHZ-63']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Harris, Gross and Schmidt
asset_names_within_place:  ['EW-ONE-19' 'BN-HAC-74']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Parks, Nash and Espinoza
asset_names_within_place:  ['YN-NZP-03']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Moreno, Butler and Huff
asset_names_within_place:  ['TH-ODX-36' 'KL-RZN-01']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Diaz Inc
asset_names_within_place:  ['ME-SZR-11' 'TY-VOB-39']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Berry, Meyer and Williams
asset_names_within_place:  ['VN-YHO-26' 'DJ-VZR-77']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Lane Inc
asset_names_within_place:  ['WG-AYD-90' 'KH-QQF-62']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Hunter-Shepherd
asset_names_within_place:  ['JL-YFJ-40' 'YS-YEI-18']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Green, Hayes and Taylor
asset_names_within_place:  ['WK-ZGU-09' 'YS-YFI-24']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Chandler PLC
asset_names_within_place:  ['GF-XQD-42' 'PI-PWV-21']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
--------------------------------------------------------------------------------

place:  Smith Ltd
asset_names_within_place:  ['ZE-KHP-90']
asset_types_within_place:  [3]
module_id__within_place:  [nan]
&lt;Figure size 1080x216 with 0 Axes&gt;
```

Places within Organisation:Gonzalez-Hancock
Flynn-Mcgrath
67.0%
Thompson Group
11.1%
10.7%
Bell LLC
11.2%
Torres, Montoya and Neal
asset_names within Place:Thompson Group
SL-UAD-50
95.7%
EB-FVF-50
CH-QKR-84
Organisation name is: Gonzalez-Hancock, place: Thompson Group with asset_name: ['EB-FVF-50' 'SL-UAD-50' 'CH-QKR-84'] and asset_types: [5 8]
asset_names within Place:Flynn-Mcgrath
ND-IRJ-15
100.0%
Organisation name is: Gonzalez-Hancock, place: Flynn-Mcgrath with asset_name: ['ND-IRJ-15'] and asset_types: [8]
asset_names within Place:Torres, Montoya and Neal
NW-CNY-73
100.0%
Organisation name is: Gonzalez-Hancock, place: Torres, Montoya and Neal with asset_name: ['NW-CNY-73'] and asset_types: [5]
asset_names within Place:Bell LLC
OZ-LHY-83
96.8%
OQ-XNB-63
IC-INE-34
Organisation name is: Gonzalez-Hancock, place: Bell LLC with asset_name: ['OQ-XNB-63' 'OZ-LHY-83' 'IC-INE-34'] and asset_types: [5 8]
Places within Organisation:Larson-Mccall
Benson Ltd
Cohen PLC
6.4%
1.5%
92.1%
Aguilar, Price and Clark
asset_names within Place:Cohen PLC
FI-DMS-52
48.9%
51.1%
GV-WTD-14
Organisation name is: Larson-Mccall, place: Cohen PLC with asset_name: ['FI-DMS-52' 'GV-WTD-14'] and asset_types: [8]
asset_names within Place:Benson Ltd

Conclusion

1. The data represents 3 Organisations
2. Each Organisation has unique places
3. Each place within single Organisation contains one or more asset_name and one or more asset_type