

2. Data Cleansing - Write a python function/functions for data cleansing process

```
In [2]: import json
with open("C://Users/monisha/Downloads/airports.json") as f:
    data=json.load(f)
```

```
In [3]: data
```

```
Out[3]: [{ 'code': 'AAA',
  'lat': '-17.3595',
  'lon': '-145.494',
  'name': 'Anaa Airport',
  'city': 'Anaa',
  'state': 'Tuamotu-Gambier',
  'country': 'French Polynesia',
  'woeid': '12512819',
  'tz': 'Pacific/Midway',
  'phone': '',
  'type': 'Airports',
  'email': '',
  'url': '',
  'runway_length': '4921',
  'elev': '7',
  'icao': 'NTGA',
  'direct_flights': '2',
  'carriers': '1'},
  { 'code': 'AAE',
  'lat': '-13.6636',
  'lon': '-17.1333',
  'name': 'Fanning Island Airport',
  'city': 'Fanning Island',
  'state': 'Line Islands',
  'country': 'Micronesia',
  'woeid': '12512819',
  'tz': 'Pacific/Midway',
  'phone': '',
  'type': 'Airports',
  'email': '',
  'url': '',
  'runway_length': '4921',
  'elev': '7',
  'icao': 'NTGA',
  'direct_flights': '2',
  'carriers': '1'}
```

```
In [4]: def int(var):
    i=0
    while i<len(data):
        if data[i][var]==None or data[i][var]=='NULL' or data[i][var]=='':
            data[i][var]=0
        i=i+1
    return data
```

```
In [5]: data = int("woeid")
data = int("runway_length")
data = int("elev")
data = int("direct_flights")
data = int("carriers")
```

In [6]: `print(data)`

```
[{'code': 'AAA', 'lat': '-17.3595', 'lon': '-145.494', 'name': 'Anaa Airpor
t', 'city': 'Anaa', 'state': 'Tuamotu-Gambier', 'country': 'French Polynesi
a', 'woeid': '12512819', 'tz': 'Pacific/Midway', 'phone': '', 'type': 'Airpo
rts', 'email': '', 'url': '', 'runway_length': '4921', 'elev': '7', 'icao':
'NTGA', 'direct_flights': '2', 'carriers': '1'}, {'code': 'AAE', 'lat': '36.
8236', 'lon': '7.8103', 'name': 'El Mellah Airport', 'city': 'El Tarf', 'sta
te': 'Annaba', 'country': 'Algeria', 'woeid': '12510325', 'tz': 'Africa/Algi
ers', 'phone': '', 'type': 'Airports', 'email': '', 'url': '', 'runway_lengt
h': '9843', 'elev': '16', 'icao': 'DABB', 'direct_flights': '6', 'carriers':
'2'}, {'code': 'AAL', 'lat': '57.0952', 'lon': '9.85606', 'name': 'Aalborg A
irport', 'city': 'Norresundby', 'state': 'Nordjylland', 'country': 'Denmar
k', 'woeid': '12512587', 'tz': 'Europe/Copenhagen', 'phone': '', 'type': 'Ai
rports', 'email': '', 'url': 'http://www.aal.dk/', 'runway_length': '8700',
'elev': '10', 'icao': 'EKYT', 'direct_flights': '7', 'carriers': '10'}, {'co
de': 'AAM', 'lat': '-24.8', 'lon': '31.5333', 'name': 'Mala Mala', 'city':
'Mala Mala', 'state': '', 'country': 'South Africa', 'woeid': '55921381', 't
z': 'Africa/Johannesburg', 'phone': '', 'type': 'Airports', 'email': '', 'ur
l': '', 'runway_length': '4420', 'elev': '1200', 'icao': 'FAMD', 'direct_fli
ghts': '1', 'carriers': '1'}, {'code': 'AAN', 'lat': '24.25', 'lon': '55.7
```

In [7]: `def str(var):`
 `i = 0`
 `while i < len(data):`
 `if data[i][var] == None or data[i][var] == 'NULL' or data[i][var]== '':`
 `data[i][var] = 'BLANK'`
 `i = i+1`
 `return data`

In [8]: `data=str("name")`
`data=str("city")`
`data=str("state")`
`data=str("country")`
`data=str("tz")`
`data=int("phone")`
`data=str("type")`
`data=str("email")`
`data=str("url")`
`data=str("icao")`

In [9]: `print(data)`

```
[{'code': 'AAA', 'lat': '-17.3595', 'lon': '-145.494', 'name': 'Anaa Airpor
t', 'city': 'Anaa', 'state': 'Tuamotu-Gambier', 'country': 'French Polynesi
a', 'woeid': '12512819', 'tz': 'Pacific/Midway', 'phone': 0, 'type': 'Airpor
ts', 'email': 'BLANK', 'url': 'BLANK', 'runway_length': '4921', 'elev': '7',
'icao': 'NTGA', 'direct_flights': '2', 'carriers': '1'}, {'code': 'AAE', 'la
t': '36.8236', 'lon': '7.8103', 'name': 'El Mellah Airport', 'city': 'El Tar
f', 'state': 'Annaba', 'country': 'Algeria', 'woeid': '12510325', 'tz': 'Afr
ica/Algiers', 'phone': '', 'type': 'Airports', 'email': 'BLANK', 'url': 'BLA
NK', 'runway_length': '9843', 'elev': '16', 'icao': 'DABB', 'direct_flight
s': '6', 'carriers': '2'}, {'code': 'AAL', 'lat': '57.0952', 'lon': '9.8560
6', 'name': 'Aalborg Airport', 'city': 'Norresundby', 'state': 'Nordjyllan
d', 'country': 'Denmark', 'woeid': '12512587', 'tz': 'Europe/Copenhagen', 'p
hone': '', 'type': 'Airports', 'email': 'BLANK', 'url': 'http://www.aal.d
k/', 'runway_length': '8700', 'elev': '10', 'icao': 'EKYT', 'direct_flight
s': '7', 'carriers': '10'}, {'code': 'AAM', 'lat': '-24.8', 'lon': '31.533
3', 'name': 'Mala Mala', 'city': 'Mala Mala', 'state': 'BLANK', 'country':
'South Africa', 'woeid': '55921381', 'tz': 'Africa/Johannesburg', 'phone':
'', 'type': 'Airports', 'email': 'BLANK', 'url': 'BLANK', 'runway_length':
'4420', 'elev': '1200', 'icao': 'FAMD', 'direct_flights': '1', 'carriers':
'10'}, {'code': 'AAN', 'lat': '13.4351', 'lon': '15.75', 'name': 'AAN', 'ci
ty': 'AAN', 'state': 'AAN', 'country': 'AAN', 'woeid': 'AAN', 'tz': 'AAN', 'p
hone': 'AAN', 'type': 'AAN', 'email': 'AAN', 'url': 'AAN', 'runway_length':
'AAN', 'elev': 'AAN', 'icao': 'AAN', 'direct_flights': 'AAN', 'carriers': 'A
AN'}
```

In [10]: `def float(var):`
 `i=0`
 `while i<len(data):`
 `if data[i][var]==None or data[i][var]=='NULL' or data[i][var]=='':`
 `data[i][var]=0.0`
 `i=i+1`
 `return data`

In [11]: `data=float("lat")`
`data=float("lon")`

