

## Data Generation

### Group Members:

Akash Memon 14884

Muhammad Mujtaba Afzal 14882

### Searching for Data:

After creating the ERD and Database diagram we searched for a relevant dataset online. Since the ERD was created from scratch, we could not find any dataset to fulfill our requirements. In order to use data for our project, we had to create data using Python.

### Generating Data:

The data is generated using Python and saved in CSV format. Libraries used are pandas, random and uuid. The data is randomly generated using functions and loops that we built. The main library used is random.

For example in order to create Contact Numbers we used the following code:

```
def randomContact():  
    return "(021) 921" + str(randrange(10000, 99999))
```

To generate Addresses for places of interest we used:

```
def randomAddress():  
    return (  
        str(randrange(3000))  
        + " Building "  
        + str(randrange(21))  
        + ", Street "  
        + str(randrange(11))  
    )
```

Another method that we have used for tables that have like 3 or 5 rows is that we manually populate the list and then export it in a CSV. Below are example of it:

```
TOI_Type = [  
    [1, "Monument"],  
    [2, "Restaurant"],  
    [3, "Museum"],  
    [4, "Hotel"],  
    [5, "Cafe"],  
]  
  
Event_Type = [  
    [1, "Party"],  
    [2, "Drama"],  
    [3, "Dinner"],  
    [4, "Concert"],  
    [5, "Festival"],  
]
```

To generate a complete table we used loops. Inside these loops the functions created to generate random data for a column are called. An example mentioned below:

```
for x in range(1000000):#organizations
    list_orgs.append(
        [
            (x + 1),
            randomizer(len(list_states)),
            randomizer(len(travelCode)),
            "Organization " + str(x + 1),
            randomAddress(),
            randomContact(),
            randomWebsite(x + 1),
        ]
    )
```

Here we first generate an ID using simple increments.

Then we select a random state from the states list and also a random travel org. Type from its list. We name the Organization by concatenating "organisation" with its ID.

Then the functions are called to generate addresses, contact, and websites.

Lastly, all the lists were converted to their respective CSVs using pandas toCSV.

#### **Data Files:**

The data files are attached with submission in zip format.

Also they can be accessed by [clicking here](#).