

# Data Design for Museum Dataset

Student No. C14489558

## Dataset:

The dataset used for this assignment was the Museum Dataset, containing data about museums in the United States. This can be downloaded here: <https://www.kaggle.com/imls/museum-directory>, however the museums.csv file is included in the 'data' directory of this submission.

## Process:

The first two designs (museum\_flat & museum\_embed\_byid) created for this dataset were created by writing Python programs to generate .js (JavaScript) files which held all of the inserts. These javascript files could then be run in Studio 3T to create and populate the collections. The two .js files are included in the data directory (museum\_embed\_by\_ID.js and museum\_flat.js).

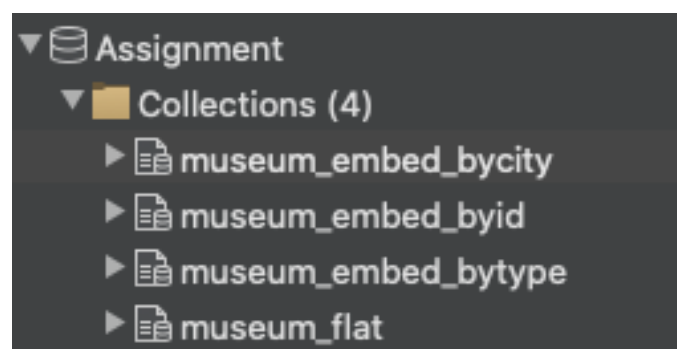
The Python files (GenerateMuseums\_EmbedID.py and GenerateMuseums\_Flat.py) are also included with this submission in the main directory, they are commented to explain the functions and uses of the code within them. The Python programs can be run again if needs be and these will simply overwrite the existing javascript files.

The second two designs (museum\_embed\_bycity & museum\_embed\_bytype) were created using an aggregate function on the museum\_embed\_byid collection. The two aggregates are contained within the GenerateMuseums\_Embed\_City\_Type.js file (this file is included in the 'data' directory).

All four collections are included with this submission in the form of JSON files located within the 'json' directory to demonstrate the collection designs.

## Designs:

As mentioned above, there are four collection designs demonstrated: museum\_flat, museum\_embed\_byid, museum\_embed\_bycity and museum\_embed\_bytype. These are explained below:



The first, museum\_embed\_flat, is a 'flat design', this is simply a collection of documents, where there is no embedded arrays. For each document, all information is listed, one after the other, using the Museum ID as the identifier for each, meaning there is a total of 1,370 documents. Two documents are shown below as examples:

```
{
  "_id" : ObjectId("5f8e61b4168b7773e7578564"),
  "museumID" : [
    8400100163.0
  ],
  "museumName" : "AFRICAN AMERICAN HERITAGE TRAIL",
  "museumType" : "GENERAL MUSEUM",
  "streetAddress" : "204 SOUTH LAWRENCE ST",
  "city" : "MOBILE",
  "state" : "AL",
  "zipCode" : 36633.0,
  "phoneNumber" : "2515332174"
}
{
  "_id" : ObjectId("5f8e61b5168b7773e7578565"),
  "museumID" : [
    8409500238.0
  ],
  "museumName" : "ALABAMA AGRICULTURAL MUSEUM",
  "museumType" : "SCIENCE & TECHNOLOGY MUSEUM OR PLANETARIUM",
  "streetAddress" : "430 LANDMARK DR",
  "city" : "DOTHAN",
  "state" : "AL",
  "zipCode" : 36303.0,
  "phoneNumber" : "3346711078"
}
```

The second, museum\_embed\_byid, is a simple embedded design, this is comprised of the same documents, however for each document in this collection, there is an embedded array for location containing all of the location information for that document. The Museum ID is again used as the identifier, meaning there is a total of 1,370 documents. Two documents are shown below as examples:

```
{
  "_id" : ObjectId("5f8e3b498bac42bf8e0c7142"),
  "museumID" : [
    8400100163.0
  ],
  "museumName" : "AFRICAN AMERICAN HERITAGE TRAIL",
  "museumType" : "GENERAL MUSEUM",
  "location" : [
    {
      "streetAddress" : "204 SOUTH LAWRENCE ST",
      "city" : "MOBILE",
      "state" : "AL",
      "zipCode" : 36633.0,
      "latitude" : 30.68582,
      "longitude" : -88.04614000000001
    }
  ],
  "phoneNumber" : "2515332174"
}
{
  "_id" : ObjectId("5f8e3b4a8bac42bf8e0c7143"),
  "museumID" : [
    8409500238.0
  ],
  "museumName" : "ALABAMA AGRICULTURAL MUSEUM",
  "museumType" : "SCIENCE & TECHNOLOGY MUSEUM OR PLANETARIUM",
  "location" : [
    {
      "streetAddress" : "430 LANDMARK DR",
      "city" : "DOTHAN",
      "state" : "AL",
      "zipCode" : 36303.0,
      "latitude" : 31.27931,
      "longitude" : -85.37361
    }
  ],
  "phoneNumber" : "3346711078"
}
```

The third design, `museum_embed_bycity`, is a collection of documents where they are grouped by the city they are located within. There is a total of 815 documents, one for each city, within each of them is an embedded array of museums containing all of the museums within that city. The start of the document for New York state is shown below as an example:

```
{
  "_id" : {
    "city" : [
      "CARMEL"
    ]
  },
  "museums" : [
    {
      "_id" : ObjectId("5fbe3ba48bac42bf8e0c71cb"),
      "museumID" : [
        8400602018.0
      ],
      "museumName" : "ARTS HABITAT",
      "museumType" : "ART MUSEUM",
      "location" : [
        {
          "streetAddress" : "3785 VILLA NONA MARIE",
          "city" : "CARMEL",
          "state" : "CA",
          "zipCode" : 93922.0,
          "latitude" : 36.53934,
          "longitude" : -121.90593
        }
      ],
      "phoneNumber" : "8316246111"
    },
    {
      "_id" : ObjectId("5fbe3bbb8bac42bf8e0c71f1"),
      "museumID" : [
        8400602263.0
      ],
      "museumName" : "CARMEL HERITAGE",
      "museumType" : "HISTORIC PRESERVATION",
      "location" : [
        {
          "streetAddress" : "LINCOLN ST & 6TH AVE",
          "city" : "CARMEL",
          "state" : "CA",
          "zipCode" : 93921.0,
          "latitude" : 36.562,
          "longitude" : -121.8964
        }
      ],
      "phoneNumber" : "8316244447"
    }
  ],
}
```

The fourth design, `museum_embed_bytype` is a collection of documents where they are grouped by the type of museum. There is a total of nine documents, one for each type of museum, within each of them is an embedded array of museums containing all of the museums classified under that type.

The start of the document for History Museums is shown below as an example:

```
{
  "_id" : {
    "MuseumType" : "ZOO, AQUARIUM, OR WILDLIFE CONSERVATION"
  },
  "museums" : [
    {
      "_id" : ObjectId("5f8e3b7d8bac42bf8e0c718b"),
      "museumID" : [
        8400100213.0
      ],
      "museumName" : "MONTGOMERY AREA ZOOLOGICAL SOCIETY",
      "museumType" : "ZOO, AQUARIUM, OR WILDLIFE CONSERVATION",
      "location" : [
        {
          "streetAddress" : "2301 COLISEUM PKWY",
          "city" : "MONTGOMERY",
          "state" : "AL",
          "zipCode" : 36109.0,
          "latitude" : 32.421079999999996,
          "longitude" : -86.27306
        }
      ],
      "phoneNumber" : "3342404900"
    },
    {
      "_id" : ObjectId("5f8e3ba08bac42bf8e0c71c5"),
      "museumID" : [
        8409500010.0
      ],
      "museumName" : "AQUARIUM OF THE BAY",
      "museumType" : "ZOO, AQUARIUM, OR WILDLIFE CONSERVATION",
      "location" : [
        {
          "streetAddress" : "THE EMBARCADERO AT BEACH STREET PIER 39",
          "city" : "SAN FRANCISCO",
          "state" : "CA",
          "zipCode" : 94133.0,
          "latitude" : 37.80032,
          "longitude" : -122.4104
        }
      ],
      "phoneNumber" : "4156235300"
    }
  ],
}
```