# MongoDB

June 9, 2024

[1]:
```
from pymongo import MongoClient
import pandas as pd
```

## 0.1 Connecting to the database

[2]:
```
# Connect to MongoDB
client = MongoClient('mongodb://localhost:27017/')
```

[6]:
```
db = client['Iris']
db
```

[6]:
```
Database(MongoClient(host=['localhost:27017'], document_class=dict,
tz_aware=False, connect=True), 'Iris')
```

[7]:
```
collection = db['iris']
collection
```

[7]:
```
Collection(Database(MongoClient(host=['localhost:27017'], document_class=dict,
tz_aware=False, connect=True), 'Iris'), 'iris')
```

### 0.1.1 Pulling Results

[8]:
```
# Retrieve the first three records
result = collection.find().limit(3)

# Print the records
for record in result:
    print(record)
```

```
{'_id': ObjectId('6665dfbf0729b193d494ee51'), 'sepal_length': 5.1,
'sepal_width': 3.5, 'petal_length': 1.4, 'petal_width': 0.2, 'species': 'Iris-
setosa'}
{'_id': ObjectId('6665dfbf0729b193d494ee52'), 'sepal_length': 4.9,
'sepal_width': 3.0, 'petal_length': 1.4, 'petal_width': 0.2, 'species': 'Iris-
setosa'}
{'_id': ObjectId('6665dfbf0729b193d494ee53'), 'sepal_length': 4.7,
'sepal_width': 3.2, 'petal_length': 1.3, 'petal_width': 0.2, 'species': 'Iris-
setosa'}
```

[9]:
```
result = collection.find_one({'species':'Iris-setosa'})
print(result)
```

```
{'_id': ObjectId('6665dfbf0729b193d494ee51'), 'sepal_length': 5.1,
'sepal_width': 3.5, 'petal_length': 1.4, 'petal_width': 0.2, 'species': 'Iris-
setosa'}
```

[26]:
```
query = {"species": "Iris-virginica"}
```

[27]:
```
# Retrieve all records
result = collection.find(query)

# Convert the result to a pandas DataFrame
df = pd.DataFrame(list(result))
```

[28]:
```
df['species'].unique()
```

[28]:
```
array(['Iris-virginica'], dtype=object)
```

[29]:
```
df.head()
```

[29]:

| | _id | sepal_length | sepal_width | petal_length | petal_width | species |
|---|---|---|---|---|---|---|
| 0 | 6665dfbf0729b193d494eeb5 | 6.3 | 3.3 | 6.0 | 2.5 | Iris-virginica |
| 1 | 6665dfbf0729b193d494eeb6 | 5.8 | 2.7 | 5.1 | 1.9 | Iris-virginica |
| 2 | 6665dfbf0729b193d494eeb7 | 7.1 | 3.0 | 5.9 | 2.1 | Iris-virginica |
| 3 | 6665dfbf0729b193d494eeb8 | 6.3 | 2.9 | 5.6 | 1.8 | Iris-virginica |
| 4 | 6665dfbf0729b193d494eeb9 | 6.5 | 3.0 | 5.8 | 2.2 | Iris-virginica |

[30]:
```
df.tail()
```

[30]:

| | _id | sepal_length | sepal_width | petal_length | petal_width | species |
|---|---|---|---|---|---|---|
| 45 | 6665dfbf0729b193d494eee2 | 6.7 | 3.0 | 5.2 | 2.3 | Iris-virginica |
| 46 | 6665dfbf0729b193d494eee3 | 6.3 | 2.5 | 5.0 | 1.9 | Iris-virginica |
| 47 | 6665dfbf0729b193d494eee4 | 6.5 | 3.0 | 5.2 | 2.0 | Iris-virginica |
| 48 | 6665dfbf0729b193d494eee5 | 6.2 | 3.4 | 5.4 | | |
| 49 | 6665dfbf0729b193d494eee6 | 5.9 | 3.0 | 5.1 | | |

```
48          2.3  Iris-virginica
49          1.8  Iris-virginica
```

### 0.1.2 Checking the amount of data points for each species

[31]:
```
virginica_amt = collection.count_documents({"species": "Iris-virginica"})
print("Amount of data-points for virginica:", virginica_amt)
```

```
Amount of data-points for virginica: 50
```

[32]:
```
setosa_amt = collection.count_documents({"species": "Iris-setosa"})
print("Amount of data-points for setosa:", setosa_amt)
```

```
Amount of data-points for setosa: 50
```

[33]:
```
versicolor_amt = collection.count_documents({"species": "Iris-versicolor"})
print("Amount of data-points for versicolor:", versicolor_amt)
```

```
Amount of data-points for versicolor: 50
```

[ ]:
```

```