

MongoDB Tutorial

Brian Reicher & Ben Wyant

```
In [1]: from mongo_connection import MongoDriver

In [2]: mongo= MongoDriver('localhost', 27017, 'restaurants')
mongo.connect()

if mongo.collection_size("restaurants_collection") == 0:
    mongo.insert_data('restaurants_collection', 'data/restaurants.json', clear=False)

Database(MongoClient(host='localhost:27017'), document_class=dict, tz_aware=False, connect=True), 'restaurants')
```

Q1) How many restaurants are in each borough?

```
In [3]: print("\n Number of restaurants in each bourough: \n")
mongo.aggregate_query("restaurants_collection", {
    {
        "$group": {
            "_id": '$borough',
            'count': { '$sum': 1 }
        }
    }
})

Number of restaurants in each bourough:
```

```
Out[3]: [{'_id': 'Bronx', 'count': 2338},
{'_id': 'Manhattan', 'count': 10259},
{'_id': 'Missing', 'count': 51},
{'_id': 'Queens', 'count': 5656},
{'_id': 'Staten Island', 'count': 969},
{'_id': 'Brooklyn', 'count': 6086}]
```

Q2) How many McDonald's are in NYC?

```
In [4]: print("\n Number of McDonald's in NYC: \n")
mongo.aggregate_query("restaurants_collection", {
    {
        "$match": {
            "name": "Mcdonald's"
        }
    },
    {
        "$count": "totalMcDonalds"
    }
})

Number of McDonald's in NYC:

Out[4]: [{'totalMcDonalds': 208}]
```

Q3) Which boroughs, sorted by count, have the highest number of Chinese restaurants?

```
In [5]: print("\n Boroughs with the highest number of Chinese restaurants --> give the number of Chinese restaurants in each boroughough: \n")
mongo.aggregate_query("restaurants_collection", {
    {
        "$match": { "cuisine": "Chinese" } },
    {
        "$group": {
            "_id": "$borough",
            "count": { "$sum": 1 }
        }
    },
    {
        "$sort": { "count": -1 }
    }
})

Boroughs with the highest number of Chinese restaurants --> give the number of Chinese restaurants in each boroughough:
```

```
Out[5]: [{'_id': 'Brooklyn', 'count': 763},
{'_id': 'Queens', 'count': 728},
{'_id': 'Manhattan', 'count': 510},
{'_id': 'Bronx', 'count': 323},
{'_id': 'Staten Island', 'count': 88},
{'_id': 'Missing', 'count': 6}]
```

Q4) What are the 5 restaurants with the highest average score, sorted by average score? (min 5 reviews)

```
In [6]: print("\n Top 5 restaurants with the highest average score, sorted by average score (min 5 reviews): \n")
mongo.aggregate_query("restaurants_collection", [
    {
        "$match": {
            "$expr": {
                "$gte": [ { "$size": "$grades" }, 5 ]
            }
        }
    },
    {
        "$unwind": "$grades"
    },
    {
        "$group": {
            "_id": "$restaurant_id",
            "name": { "$first": "$name" },
            "avgScore": { "$avg": "$grades.score" }
        }
    },
    {
        "$sort": { "avgScore": -1 }
    },
    {
        "$limit": 5
    }
])

Top 5 restaurants with the highest average score, sorted by average score (min 5 reviews):

Out[6]: [{'_id': '40393488', 'name': 'Bella Napoli', 'avgScore': 38.6},
{'_id': '41267350', 'name': 'Tenda Asian Fusioin', 'avgScore': 37.4},
{'_id': '41602559', 'name': 'Red Chopstick', 'avgScore': 36.285714285714285},
{'_id': '41550400', 'name': 'El Mxteco', 'avgScore': 34.8},
{'_id': '41683783', 'name': 'Bamboo Restaurant', 'avgScore': 34.0}]
```

Q5) Which restaurants have a zipcode that starts with '10', are Italian or Chinese cuisine, and have been graded 'A' in their latest grade?

```
In [7]: print("\n Restaurants which have a zipcode that starts with '10' and they are of either Italian or Chinese cuisine and have been graded 'A' in their latest grade: ")
mongo.search_query(collection_name="restaurants_collection", qu = {
    "address.zipcode": { "$regex": "^10" },
    "or": {
        { "cuisine": "Italian" },
        { "cuisine": "Chinese" }
    },
    "grades.0.grade": "A"
},
    proj = {
        "_id": 1,
        "name": 1,
    })

Restaurants which have a zipcode that starts with '10' and they are of either Italian or Chinese cuisine and have been graded 'A' in their latest grade:
```

```
Out[7]: [{'name': 'Happy Garden'},
{'name': 'Happy Garden'},
{'name': 'Isle Of Capri Resturant'},
{'name': 'Marchis Restaurant'},
{'name': 'Crystal Room'},
{'name': 'Forlinis Restaurant'},
{'name': 'Angelo Of Mulberry St.'},
{'name': 'Arturo's'},
{'name': 'V & T Restaurant'},
{'name': 'Barbetta Restaurant'}]
```

Q6) Which restaurants are located in the Bronx borough and have an 'American' cuisine?

```
In [8]: print("\n Restaurants that are located in the Bronx borough and have an 'American' cuisine: \n")
mongo.search_query(collection_name="restaurants_collection", qu = {'borough': "Bronx", "cuisine": "American"},
    proj={
        "_id": 0,
        "name": 1,
        "borough": 1,
        "cuisine": 1,
    }, lim=5)

Restaurants that are located in the Bronx borough and have an 'American' cuisine:
```

```
Out[8]: [{'borough': 'Bronx', 'cuisine': 'American', 'name': 'Wild Asia'},
{'borough': 'Bronx', 'cuisine': 'American', 'name': 'Manhem Club'},
{'borough': 'Bronx', 'cuisine': 'American', 'name': 'The New Starling Athletic Club Of The Bronx'},
{'borough': 'Bronx', 'cuisine': 'American', 'name': 'Yankee Tavern'},
{'borough': 'Bronx', 'cuisine': 'American', 'name': 'The Punch Bowl'}]
```

Q7) Which restaurants have a 'Pizza' cuisine and a 'B' grade in their latest inspection?

```
In [9]: print("\n Restaurants that have a 'Pizza' cuisine and a 'B' grade in their latest inspection: \n")
mongo.search_query(collection_name="restaurants_collection", qu={
    "cuisine": "Pizza",
    "grades.0.grade": "B"
},
    proj={
        "_id": 0,
        "name": 1,
        "borough": 1,
        "cuisine": 1,
        "grades": {
            "$elemMatch": { "grade": "B" }
        }
    }, lim=5)

Restaurants that have a 'Pizza' cuisine and a 'B' grade in their latest inspection:
```

```
Out[9]: [{'borough': 'Queens', 'cuisine': 'Pizza', 'name': 'Pizza Town', 'grades': [{'date': {'$date': 1408579200000}, 'grade': 'B', 'score': 15}]},
{'borough': 'Queens', 'cuisine': 'Pizza', 'name': 'Barone Pizza', 'grades': [{'date': {'$date': 1415836800000}, 'grade': 'B', 'score': 16}]},
{'borough': 'Staten Island', 'cuisine': 'Pizza', 'name': 'Pizza D'Oro', 'grades': [{'date': {'$date': 1407369600000}, 'grade': 'B', 'score': 21}]},
{'borough': 'Manhattan', 'cuisine': 'Pizza', 'name': 'Sal And Carmine Pizza', 'grades': [{'date': {'$date': 1409097600000}, 'grade': 'B', 'score': 15}]},
{'borough': 'Manhattan', 'cuisine': 'Pizza', 'name': 'Famous Famiglia Pizza', 'grades': [{'date': {'$date': 1408492800000}, 'grade': 'B', 'score': 25}]}
```

Q8) Filter to show restaurant's names, cuisines, and high scores where the score is greater than or equal to 10

```
In [10]: print("\n Restaurant's names, cuisines, and high scores where the score is greater than or equal to 10: \n")
mongo.aggregate_query("restaurants_collection",[
    {
        "$project": {
            "name": 1,
            "cuisine": 1,
            "high_scores": {
                "$filter": {
                    "input": "$grades",
                    "as": "grade",
                    "cond": { "$gte": [ "$$grade.score", 10 ] }
                }
            }
        }
    },
    { "$limit": 5 }
])

Restaurant's names, cuisines, and high scores where the score is greater than or equal to 10:
```

```
Out[10]: [{'_id': ObjectId('63f6d8ffc7bc914ba5c90071'), 'cuisine': 'Bakery', 'name': 'Morris Park Bake Shop', 'high_scores': [{'date': {'$date': 1358985600000}, 'grade': 'A', 'score': 10}, {'date': {'$date': 1299715200000}, 'grade': 'B', 'score': 14}]}, {'_id': ObjectId('63f6d8ffc7bc914ba5c90072'), 'cuisine': 'Hamburgers', 'name': 'Wendy's', 'high_scores': [{'date': {'$date': 1404172800000}, 'grade': 'B', 'score': 23}, {'date': {'$date': 1367280000000}, 'grade': 'A', 'score': 12}, {'date': {'$date': 1336435200000}, 'grade': 'A', 'score': 12}]}, {'_id': ObjectId('63f6d8ffc7bc914ba5c90073'), 'cuisine': 'Irish', 'name': 'Dj Reynolds Pub And Restaurant', 'high_scores': [{'date': {'$date': 1374451200000}, 'grade': 'A', 'score': 11}, {'date': {'$date': 1343692800000}, 'grade': 'A', 'score': 12}, {'date': {'$date': 1325116800000}, 'grade': 'A', 'score': 12}]}, {'_id': ObjectId('63f6d8ffc7bc914ba5c90074'), 'cuisine': 'American', 'name': 'Riviera Caterer', 'high_scores': [{'date': {'$date': 1334275200000}, 'grade': 'A', 'score': 12}, {'date': {'$date': 1318377600000}, 'grade': 'A', 'score': 12}]}, {'_id': ObjectId('63f6d8ffc7bc914ba5c90075'), 'cuisine': 'Jewish/Kosher', 'name': 'Tov Kosher Kitchen', 'high_scores': [{'date': {'$date': 1416787200000}, 'grade': 'Z', 'score': 20}, {'date': {'$date': 1358380800000}, 'grade': 'A', 'score': 13}, {'date': {'$date': 1343865600000}, 'grade': 'A', 'score': 13}, {'date': {'$date': 1323907200000}, 'grade': 'B', 'score': 25}]}
```

Q9) Which restaurants have a 'Chinese' or 'Japanese' cuisine, are located in the 'Queens' borough, and have a grade 'A' in their latest inspection?

```
In [11]: print("\n Restaurants that have a 'Chinese' or 'Japanese' cuisine, and are located in the 'Queens' borough, and have a grade 'A' in their latest inspection. \n")
mongo.search_query(collection_name="restaurants_collection", qu={
    "borough": "Queens",
    "cuisine": {"$in": ["Chinese", "Japanese"]},
    "grades.0.grade": "A"
},
    proj={
        "name": 1,
        "borough": 1,
        "cuisine": 1,
        "grades": 1,
        "_id": 0
    },
    lim=5)

Restaurants that have a 'Chinese' or 'Japanese' cuisine, and are located in the 'Queens' borough, and have a grade 'A' in their latest inspection.
```

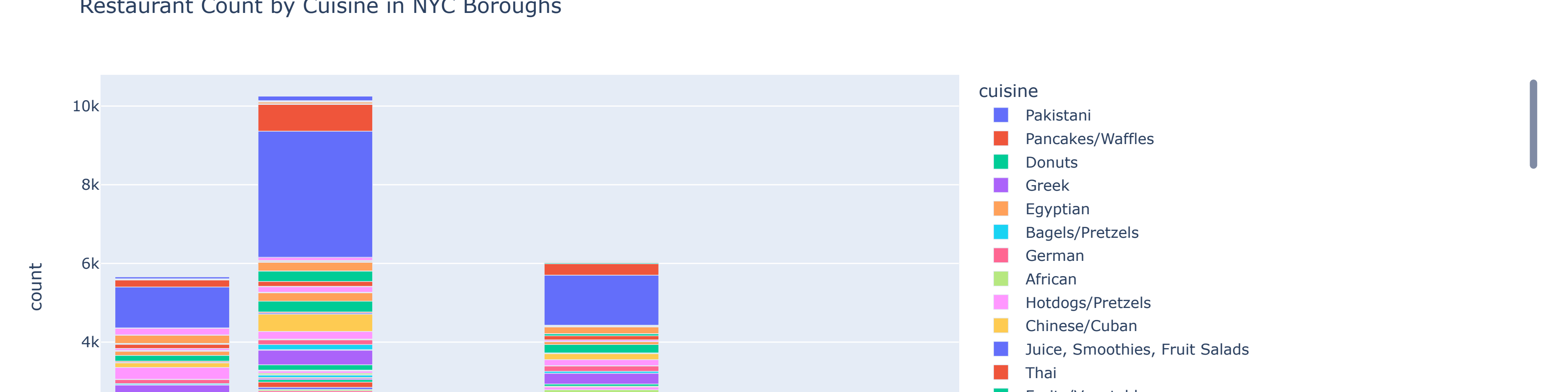
```
Out[11]: [{'borough': 'Queens', 'cuisine': 'Chinese', 'grades': [{'date': {'$date': 1398038400000}, 'grade': 'A', 'score': 10}, {'date': {'$date': 1384214400000}, 'grade': 'A', 'score': 5}, {'date': {'$date': 1370304000000}, 'grade': 'A', 'score': 12}, {'date': {'$date': 1352851200000}, 'grade': 'A', 'score': 10}, {'date': {'$date': 1349913600000}, 'grade': 'P', 'score': 0}, {'date': {'$date': 1337817600000}, 'grade': 'A', 'score': 13}, {'date': {'$date': 1323904000000}, 'grade': 'A', 'score': 12}, {'date': {'$date': 1311120000000}, 'grade': 'A', 'score': 11}, {'name': 'Ho Mei Restaurant'}, {'borough': 'Queens', 'cuisine': 'Japanese', 'grades': [{'date': {'$date': 1403654400000}, 'grade': 'A', 'score': 6}, {'date': {'$date': 1389225600000}, 'grade': 'A', 'score': 10}, {'date': {'$date': 1367452800000}, 'grade': 'A', 'score': 21}, {'date': {'$date': 1347321600000}, 'grade': 'A', 'score': 11}, {'date': {'$date': 1329091200000}, 'grade': 'A', 'score': 4}, {'date': {'$date': 1316044800000}, 'grade': 'B', 'score': 27}, {'name': 'Harita Japanese Restaurant'}, {'borough': 'Queens', 'cuisine': 'Chinese', 'grades': [{'date': {'$date': 1401840000000}, 'grade': 'A', 'score': 13}, {'date': {'$date': 1387843200000}, 'grade': 'A', 'score': 12}, {'date': {'$date': 1370988000000}, 'grade': 'A', 'score': 13}, {'date': {'$date': 1336780800000}, 'grade': 'A', 'score': 13}, {'date': {'$date': 1321488000000}, 'grade': 'A', 'score': 12}, {'date': {'$date': 1310428800000}, 'grade': 'B', 'score': 16}, {'name': 'Baybridge Szechuan Restaurant'}, {'borough': 'Queens', 'cuisine': 'Chinese', 'grades': [{'date': {'$date': 1413417600000}, 'grade': 'A', 'score': 11}, {'date': {'$date': 1379548800000}, 'grade': 'A', 'score': 11}, {'date': {'$date': 1346803200000}, 'grade': 'A', 'score': 9}, {'name': 'Peking Restaurant'}, {'borough': 'Queens', 'cuisine': 'Japanese', 'grades': [{'date': {'$date': 1410307200000}, 'grade': 'A', 'score': 10}, {'date': {'$date': 1394668800000}, 'grade': 'A', 'score': 12}, {'date': {'$date': 1373864000000}, 'grade': 'A', 'score': 9}, {'date': {'$date': 1356566400000}, 'grade': 'A', 'score': 5}, {'date': {'$date': 1337299200000}, 'grade': 'A', 'score': 5}, {'date': {'$date': 1323216000000}, 'grade': 'B', 'score': 14}, {'name': 'Sushi Family'}]}
```

Q10) What is the distribution of restaurants across different cuisines in the NYC boroughs

```
In [12]: print("\n Query to make a visualization of the distribution of restaurants across different cuisines in the NYC boroughs: \n")
res list = mongo.aggregate_query("restaurants_collection", [{"$group": {"_id": {"borough": "$borough", "cuisine": "$cuisine", "count": {"$sum": 1}}}, {"$project": {"borough": "$_id.borough", "cuisine": "$_id.cuisine", "count": "$count", "_id": 0}}], show=False)

for item in res[:10]:
    print(item)

mongo.plot_query(res, x_var="borough", y_var="count", color_on="cuisine", plot_title="Restaurant Count by Cuisine in NYC Boroughs", save_as='../data/mongo_visuali
mongo.disconnect()
```



This visualization shows us the difference between cuisines across the 5 NYC boroughs. We can first note the sheer discrepancy in the amount of restaurants in each borough, with Manhattan dramatically leading the pack, and Staten Island at ~1/10 the amount of Manhattan. There are notably dozens of cuisines in each borough, which is indicative of New York's large population and melting pot of various cultures/immigrants. We are also able to see the largest cluster portion in each bar, that reads the number of American restaurants in each borough. This makes sense given that NYC is still located in America, and would reason to have a significant number of American restaurants. Another popular restaurant cuisine across all boroughs is Chinese food (second largest cluster in most bars), which is extremely popular across the East Coast of the US, and again indicative of New York's large Chinese population.