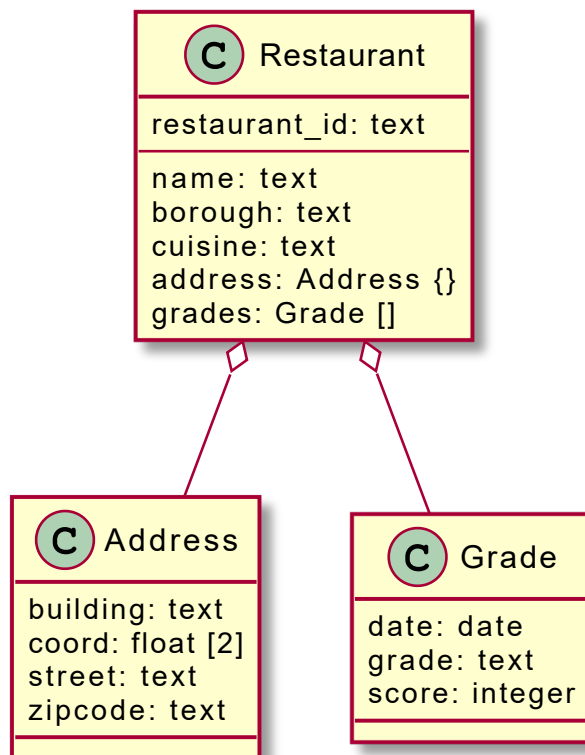


## Sample Project ICNDb

Chuck Norris who?! He was a famous martial arts actor in the '90s: <https://chucknorris.com/>  
(<https://chucknorris.com/>)

### 1 Summary, Overview

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.



Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

## 2 Requirements & Configuration

```
! pip3 list | findstr "pymongo dnspython pandas"
```

dnspython	2.1.0
pandas	1.2.4
pymongo	3.11.3

```
import pymongo
from pprint import pprint
import pandas as pd
import requests
import json
```

```
# pandas configuration
pd.set_option('precision', 2)
pd.set_option('max_rows', 30)
pd.set_option('max_colwidth', 50)
# pd.describe_option('max_rows')
# pd.describe_option('precision')
# pd.describe_option('max_colwidth')
```

```
# API and Database details
API_URL = "https://api.icndb.com/jokes/?exclude=[explicit]"
CNX_STR = "localhost:27017"
DB_NAME = "icndb"
COLL_NAME = "jokes"
```

```
# connection to MongoDB
client = pymongo.MongoClient(CNX_STR)
db = client[DB_NAME]
jokes = db[COLL_NAME]
```

## 3 ETL

### 3.1 Remove all existing documents -> Reset collection

```
jokes.drop()
jokes.count_documents({})
```

0

### 3.2 Fetch data

```
# fetch JSON from API_URL
r = requests.get(API_URL)
data = json.loads(r.text)
```

```
print(r.text[0:500])
```

```
{ "type": "success", "value": [ { "id": 2, "joke": "MacGyver can build an airplane out of gum and paper clips. Chuck Norris can kill him and take it.", "categories": [] }, { "id": 3, "joke": "Chuck Norris doesn't read books. He stares them down until he gets the information he wants.", "categories": [] }, { "id": 4, "joke": "If you ask Chuck Norris what time it is, he always answers 'Two seconds till'". After you ask 'Two seconds to what?', he roundhouse kicks you in the face.
```

### 3.3 Insert into MongoDB

```
# insert the list of jokes in "value" into MongoDB collection "jokes"
jokes.insert_many(data['value']);
```

```
# count number of documents inserted
jokes.count_documents({})
```

559

```
c = jokes.aggregate([
    {"$limit": 1},
])

for doc in c:
    pprint(f"{doc}[:500])
```

```
("{'_id': ObjectId('612d3d433179130d1c6be6b7'), 'id': 2, 'joke': 'MacGyver can
"
'build an airplane out of gum and paper clips. Chuck Norris can kill him and '
'take it.', 'categories': []}"))
```

```
c = jokes.aggregate([
    {"$limit": 5},
])

pd.DataFrame(c)
```

	_id	id	joke	categories
0	612d3d433179130d1c6be6b7	2	MacGyver can build an airplane out of gum and ...	[]
1	612d3d433179130d1c6be6b8	3	Chuck Norris doesn't read books. He stares the...	[]
2	612d3d433179130d1c6be6b9	4	If you ask Chuck Norris what time it is, he al...	[]
3	612d3d433179130d1c6be6ba	6	Since 1940, the year Chuck Norris was born, ro...	[]
4	612d3d433179130d1c6be6bb	12	Chuck Norris sheds his skin twice a year.	[]

### 3.4 Transform

```
# assign id to _id and remove id
c = jokes.aggregate([
    {"$project": {"_id": "$id", "joke": 1, "categories": 1}},
])
pd.DataFrame(c)
```

	joke	categories	_id
0	MacGyver can build an airplane out of gum and ...	[]	2
1	Chuck Norris doesn't read books. He stares the...	[]	3
2	If you ask Chuck Norris what time it is, he al...	[]	4
3	Since 1940, the year Chuck Norris was born, ro...	[]	6
4	Chuck Norris sheds his skin twice a year.	[]	12
...	...	...	...
554	Chuck Norris can build a snowman out of rain.	[]	615
555	Chuck Norris made the sun by rubbing his hands...	[]	616
556	Chuck Norris doesn't turn on his faucet, he st...	[]	617
557	Chuck Norris puts sunglasses on to protect the...	[]	618
558	Chuck Norris can lock a safe and keep the key ...	[]	619

559 rows × 3 columns

## 4 Data analysis

Quisque sit amet turpis lectus. Phasellus tincidunt mi metus, et ornare ipsum consectetur eu. Cras accumsan purus vel leo viverra, at mollis neque interdum. Sed non ultrices odio, vitae sodales neque. Quisque diam odio, gravida quis auctor ut, aliquet ac ex. Integer venenatis elit ex, vitae imperdiet tortor malesuada quis. Vestibulum dignissim est sed libero viverra interdum.

### 4.1 Categories

```
c = jokes.aggregate([
    {"$project": {"joke": 0}},
    {"$unwind": "$categories"},
    {"$group": {"_id": "$categories", "count": {"$sum": 1}}},
])
pd.DataFrame(c)
```

	_id	count
0	nerdy	105

```
c = jokes.aggregate([
    {"$match": {"categories": {"$in" : ["nerdy"]}}},
])

pd.DataFrame(c)
```

	_id	id	joke	categories
0	612cfd7d9d461a7ec9a683d6	20	The Chuck Norris military unit was not used in...	[nerdy]
1	612cfd7d9d461a7ec9a683db	26	Chuck Norris is the only human being to displa...	[nerdy]
2	612cfd7d9d461a7ec9a683e4	36	Chuck Norris originally appeared in the &quot;...	[nerdy]
3	612cfd7d9d461a7ec9a68402	69	Scientists have estimated that the energy give...	[nerdy]
4	612cfd7d9d461a7ec9a68549	412	Chuck Norris knows the last digit of pi.	[nerdy]
...	...	...	...	...
100	612cfd7d9d461a7ec9a685d5	565	Chuck Norris can make a class that is both abs...	[nerdy]
101	612cfd7d9d461a7ec9a685d6	566	Chuck Norris could use anything in java.util.*...	[nerdy]
102	612cfd7d9d461a7ec9a685d7	567	Code runs faster when Chuck Norris watches it.	[nerdy]
103	612cfd7d9d461a7ec9a685d8	584	Only Chuck Norris shuts down websites without ...	[nerdy]
104	612cfd7d9d461a7ec9a685ed	605	Chuck Norris doesn't need a keyboard he tells ...	[nerdy]

105 rows × 4 columns

## 4.2 Jokes

Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec in risus sed augue blandit tincidunt eu nec leo. Phasellus suscipit ex ut luctus auctor. Mauris efficitur finibus nunc, gravida pulvinar metus commodo eget. Quisque quis orci vehicula, maximus tellus sit amet, dignissim ligula. Proin auctor, tellus eget tempus imperdiet, nunc nisi laoreet tellus, nec viverra ipsum quam in quam.

Nam ut pellentesque arcu. Ut faucibus elit enim, nec tincidunt massa mattis id. Cras tortor urna, tempus eu viverra quis, suscipit sed magna. Mauris eget eleifend leo, ut tristique justo. In quis lectus eu neque euismod bibendum non in mi. In lobortis iaculis pulvinar. Morbi et mi neque. Etiam maximus elementum metus, non auctor dui eleifend ac.

```
c = jokes.aggregate([
    {"$match": {"joke": {"$regex" : "chuck", "$options": ""}}},
])

pd.DataFrame(c)
```

	_id	id	joke	categories
0	612cfd7d9d461a7ec9a68404	72	How much wood would a woodchuck chuck if a woo...	[]
1	612cfd7d9d461a7ec9a6856e	456	All browsers support the hex definitions #chuc...	[nerdy]

```
c = jokes.aggregate([
  {"$project": {"joke": 1}},
  {"$match": {"joke": {"$not": {"$regex" : "chuck", "$options": "i"}}}},
])

pd.DataFrame(c)
```

	_id	joke
0	612cfd7d9d461a7ec9a6843b	There is in fact an 'l' in Norris, but there i...
1	612cfd7d9d461a7ec9a6843d	An anagram for Walker Texas Ranger is KARATE W...
2	612cfd7d9d461a7ec9a6844d	Superman once watched an episode of Walker, Te...
3	612cfd7d9d461a7ec9a68450	Movie trivia: The movie "Invasion U.S.A.&...
4	612cfd7d9d461a7ec9a68457	Once you go Norris, you are physically unable ...
5	612cfd7d9d461a7ec9a6848d	Crime does not pay - unless you are an underta...
6	612cfd7d9d461a7ec9a6854a	Those aren't credits that roll after Walker Te...

```
c = jokes.aggregate([
  {"$match": {"joke": {"$regex" : "chuck", "$options": ""}}},
])

pd.DataFrame(c)
```

	_id	id	joke	categories
0	612cfd7d9d461a7ec9a68404	72	How much wood would a woodchuck chuck if a woo...	[]
1	612cfd7d9d461a7ec9a6856e	456	All browsers support the hex definitions #chuc...	[nerdy]

Curabitur vel magna nec ipsum pulvinar imperdiet vitae vitae nisi. Pellentesque mattis ultricies diam eu cursus. Maecenas eleifend ante arcu, at feugiat erat eleifend eu. In volutpat faucibus dui, sed faucibus ligula faucibus et. Maecenas convallis sodales sollicitudin. Ut consectetur, arcu ac imperdiet rutrum, massa nisi sollicitudin odio, vel mattis mi augue et sem.

Fusce semper porta risus, vitae hendrerit mauris congue vitae. Praesent venenatis varius lacus. Cras tempor augue lectus, at iaculis ex pretium sit amet. In hac habitasse platea dictumst. Nunc pharetra est eu pellentesque hendrerit. Ut nec varius sem. Morbi eu elit id lacus laoreet pharetra.

## 5 Conclusions

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata

sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

## **6 Remarks**

### **6.1 Learnings**

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.