My Project Title

1 Summary

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.

2 Requirements & Configuration

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

```
dnspython                     2.2.0
pandas                        1.4.1
pymongo                       4.0.1
```

```
import pymongo
import pprint as pp
import pandas as pd
import requests
import json
import time
import string
```

```
# 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://www.thecocktaildb.com/api/json/v1/1/search.php"
CNX_STR = "localhost:27017"
DB_NAME = "cocktaildb"
COLL_NAME = "drinks"
```

3 ELT Process

3.1 DB Setup

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

```
# Remove all existing documents
drinks.drop()
drinks.count_documents({})
```

```
0
```

3.2 Extract

```
# download all drinks starting with "A"
url = f'{API_URL}?f=a'
r = requests.get(url)
data = r.json()

# inspect first drink
drink = data['drinks'][0]

# hide fields with None values
for key in list(drink.keys()):
    if drink[key] is None:
        del drink[key]

drink
```

```
{'idDrink': '17222',
 'strDrink': 'A1',
 'strCategory': 'Cocktail',
 'strAlcoholic': 'Alcoholic',
 'strGlass': 'Cocktail glass',
 'strInstructions': 'Pour all ingredients into a cocktail shaker, mix and ser
ve over ice into a chilled glass.',
 'strInstructionsES': 'Vierta todos los ingredientes en una coctelera, mezcle
y sirva con hielo en un vaso frío.',
 'strInstructionsDE': 'Alle Zutaten in einen Cocktailshaker geben, mischen un
d über Eis in ein gekühltes Glas servieren.',
 'strInstructionsIT': 'Versare tutti gli ingredienti in uno shaker, mescolare
e servire con ghiaccio in un bicchiere freddo.',
 'strDrinkThumb': 'https://www.thecocktaildb.com/images/media/drink/2x8thr150
4816928.jpg',
 'strIngredient1': 'Gin',
 'strIngredient2': 'Grand Marnier',
 'strIngredient3': 'Lemon Juice',
 'strIngredient4': 'Grenadine',
 'strMeasure1': '1 3/4 shot ',
 'strMeasure2': '1 Shot ',
 'strMeasure3': '1/4 Shot',
 'strMeasure4': '1/8 Shot',
 'strCreativeCommonsConfirmed': 'No',
 'dateModified': '2017-09-07 21:42:09'}
```

```python
# extract values from single drink
def extract_values(drink):
    values = {
        'id': int(drink['idDrink']),
        'name': drink['strDrink'],
        'category': drink['strCategory'],
        'alcoholic': drink['strAlcoholic']=='Alcoholic',
        'modified': drink['dateModified'],
        'ingredients': [],
    }

    for i in range(15):
        key = f'strIngredient{i+1}'
        if key in drink:
            ingredient = drink[key]
            if ingredient is not None:
                values['ingredients'].append(ingredient)

    return values
```

```python
# test extraction for one drink
doc = extract_values(drink)
doc
```

```
{'id': 17222,
 'name': 'A1',
 'category': 'Cocktail',
 'alcoholic': True,
 'modified': '2017-09-07 21:42:09',
 'ingredients': ['Gin', 'Grand Marnier', 'Lemon Juice', 'Grenadine']}
```

```python
# generate list of characters including digits
list_characters = [char for char in string.ascii_lowercase + string.digits]
print(list_characters)
```

```
['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j', 'k', 'l', 'm', 'n', 'o',
'p', 'q', 'r', 's', 't', 'u', 'v', 'w', 'x', 'y', 'z', '0', '1', '2', '3',
'4', '5', '6', '7', '8', '9']
```

```python
# download/search drinks by starting char
list_drinks = []
for i, char in enumerate(list_characters):
    print(f'- {char} ({i+1}/{len(list_characters)})', end='\r')
    url = f'{API_URL}?f={char}'
    r = requests.get(url)
    data = r.json()
    if data['drinks'] is not None:
        for drink in data['drinks']:
            values = extract_values(drink)
            list_drinks.append(values)
    time.sleep(1)  # Be nice and throttle your downloads
```

```
- 9 (36/36)
```

```
# inspect first drink:
list_drinks[0]
```

```
{'id': 17222,
 'name': 'A1',
 'category': 'Cocktail',
 'alcoholic': True,
 'modified': '2017-09-07 21:42:09',
 'ingredients': ['Gin', 'Grand Marnier', 'Lemon Juice', 'Grenadine']}
```

3.3 Load

```
# insert the list of drinks(=documents) into MongoDB collection "drinks"
drinks.insert_many(list_drinks);
```

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

```
441
```

```
# get one drink from MongoDB
drinks.find_one()
```

```
{'_id': ObjectId('6408f09eb0d2b162c64ea00c'),
 'id': 17222,
 'name': 'A1',
 'category': 'Cocktail',
 'alcoholic': True,
 'modified': '2017-09-07 21:42:09',
 'ingredients': ['Gin', 'Grand Marnier', 'Lemon Juice', 'Grenadine']}
```

```
# get 5 drinks from MongoDB and display as datframe
r = drinks.aggregate([
      {"$limit": 5},
])

pd.DataFrame(r)
```

	_id	id	name	category	alcoholic	modified	ingredients
0	6408f09eb0d2b162c64ea00c	17222	A1	Cocktail	True	2017-09-07 21:42:09	[Gin, Grand Marnier, Lemon Juice, Grenadine]
1	6408f09eb0d2b162c64ea00d	13501	ABC	Shot	True	2016-08-31 19:32:08	[Amaretto, Baileys irish cream, Cognac]
2	6408f09eb0d2b162c64ea00e	17225	Ace	Cocktail	True	2017-09-07 22:05:06	[Gin, Grenadine, Heavy cream, Milk, Egg White]
3	6408f09eb0d2b162c64ea00f	14610	ACID	Shot	True	2016-11-15 11:28:37	[151 proof rum, Wild Turkey]
4	6408f09eb0d2b162c64ea010	17837	Adam	Ordinary Drink	True	2016-09-02 11:29:29	[Dark rum, Lemon juice, Grenadine]

3.4 Transform

```
# check document structure
r = drinks.aggregate([
    {"$project": {"_id": 0}},
    {"$limit": 5},
])
pd.DataFrame(r)
```

	id	name	category	alcoholic	modified	ingredients
0	17222	A1	Cocktail	True	2017-09-07 21:42:09	[Gin, Grand Marnier, Lemon Juice, Grenadine]
1	13501	ABC	Shot	True	2016-08-31 19:32:08	[Amaretto, Baileys irish cream, Cognac]
2	17225	Ace	Cocktail	True	2017-09-07 22:05:06	[Gin, Grenadine, Heavy cream, Milk, Egg White]
3	14610	ACID	Shot	True	2016-11-15 11:28:37	[151 proof rum, Wild Turkey]
4	17837	Adam	Ordinary Drink	True	2016-09-02 11:29:29	[Dark rum, Lemon juice, Grenadine]

```
# add field "ingredients_cnt" to collection
r = drinks.aggregate([
    {"$addFields": {"ingredients_cnt": { "$size": "$ingredients"}}},
    {"$out": "drinks"},
])
```

```
# check document structure
r = drinks.aggregate([
    {"$project": {"_id": 0}},
    {"$limit": 5},
])
pd.DataFrame(r)
```

	id	name	category	alcoholic	modified	ingredients	ingredients_cnt
0	17222	A1	Cocktail	True	2017-09-07 21:42:09	[Gin, Grand Marnier, Lemon Juice, Grenadine]	4
1	13501	ABC	Shot	True	2016-08-31 19:32:08	[Amaretto, Baileys irish cream, Cognac]	3
2	17225	Ace	Cocktail	True	2017-09-07 22:05:06	[Gin, Grenadine, Heavy cream, Milk, Egg White]	5
3	14610	ACID	Shot	True	2016-11-15 11:28:37	[151 proof rum, Wild Turkey]	2
4	17837	Adam	Ordinary Drink	True	2016-09-02 11:29:29	[Dark rum, Lemon juice, Grenadine]	3

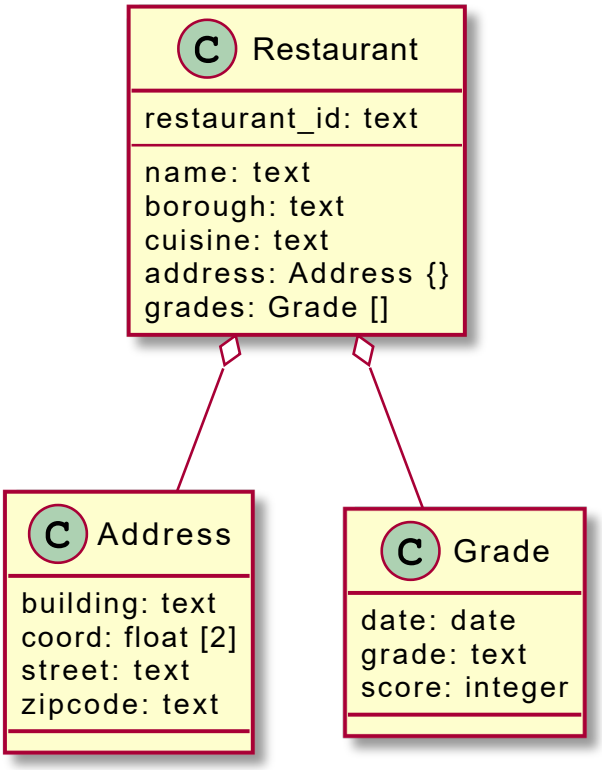
```
# drop field "modified"
r = drinks.aggregate([
    {"$unset": "modified"},
    {"$out": "drinks"},
])
```

```
# check document structure
r = drinks.aggregate([
    {"$project": {"_id": 0}},
    {"$limit": 5},
])
pd.DataFrame(r)
```

	id	name	category	alcoholic	ingredients	ingredients_cnt
0	17222	A1	Cocktail	True	[Gin, Grand Marnier, Lemon Juice, Grenadine]	4
1	13501	ABC	Shot	True	[Amaretto, Baileys irish cream, Cognac]	3
2	17225	Ace	Cocktail	True	[Gin, Grenadine, Heavy cream, Milk, Egg White]	5
3	14610	ACID	Shot	True	[151 proof rum, Wild Turkey]	2
4	17837	Adam	Ordinary Drink	True	[Dark rum, Lemon juice, Grenadine]	3

```
# Don't forget data quality checks!
```

3.5 Datastructure



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, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

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 Top 10 Ingredients

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.

```
r = drinks.aggregate([
    {"$project": {"ingredients": 1}},
    {"$unwind": "$ingredients"},
    {"$group": {"_id": "$ingredients", "count": {"$sum": 1}}},
    {"$project": {"_id": 0, "name": "$_id", "count":1}},
    {"$sort": {"count": -1}},
    {"$limit": 10},
 ])

pd.DataFrame(r)
```

	count	name
0	83	Gin
1	69	Vodka
2	51	Sugar
3	41	Orange juice
4	37	Ice
5	36	Lemon
6	36	Lemon juice
7	31	Grenadine
8	31	Light rum
9	30	Lime

4.2 Top 10 Longest Drink Names

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.

```
r = drinks.aggregate([
    {"$project": {"name": 1}},
    {"$addFields": {"length": {"$strLenCP":"$name"}}},
    {"$project": {"_id": 0, "name": 1, "length": 1}},
    {"$sort": {"length": -1}},
    {"$limit": 10},
 ])

pd.DataFrame(r)
```

	name	length
0	Empellón Cocina's Fat-Washed Mezcal	35
1	57 Chevy with a White License Plate	35
2	Radioactive Long Island Iced Tea	32
3	Owen's Grandmother's Revenge	28
4	Orange Scented Hot Chocolate	28
5	Pineapple Gingerale Smoothie	28
6	Lassi - A South Indian Drink	28
7	3-Mile Long Island Iced Tea	27
8	Lemon Elderflower Spritzer	26
9	Frozen Pineapple Daiquiri	25

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, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

6 Learnings

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.

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, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.