### Examen Elastic search

### Réalisé par Nour TOUNSI

### Le 19/11/2023

## Q1 : Répertoriez tous les quartiers de New York :

Figure 1 : Les quartiers de New York

- ✓ Get ncyrestaurant/-search : Cette ligne indique que nous effectuons une opération GET sur l'index "ncyrestaurant" en utilisant l'API de recherche d'Elasticsearch
- √ "size": 0 : Cela signifie que nous ne voulons pas de documents de retour dans les résultats de la recherche, nous sommes uniquement intéressés par les résultats de l'agrégation
- √ "aggregations" : C'est là que nous définissons les agrégations, qui sont des calculs effectués sur les données de votre index
- √ "avg-of-bord": C'est le nom que nous donnons à notre agrégation. Il peut être n'importe quel nom de votre choix.
- √ "terms" : C'est le type d'agrégation que nous effectuons. Une agrégation de termes rassemble tous les documents qui partagent la même valeur dans un champ spécifique et les regroupe
- √ "field": "BORO" : Cela indique que nous effectuons l'agrégation sur le champ
  "BORO"

### Q2 : Quel quartier compte le plus de restaurants ?

Dans cette requête, l'agrégation "terms" est utilisée pour regrouper les données par le champ "BORO", qui représente les quartiers de New York. Le paramètre "size" est défini sur 1, ce qui limite le nombre de termes renvoyés à 1. Le paramètre "order" est défini pour trier les résultats par nombre d'occurrences dans l'ordre décroissant.



Figure 2 : Le quartier qui compte le plus de restaurants

## Q3 : Que signifie le code de violation « 04N » ?

Cette requête récupère la description de la violation pour le code "04N". Elle utilise une requête "match" pour filtrer les documents avec le code de violation "04N" et renvoie la première description de violation trouvée.

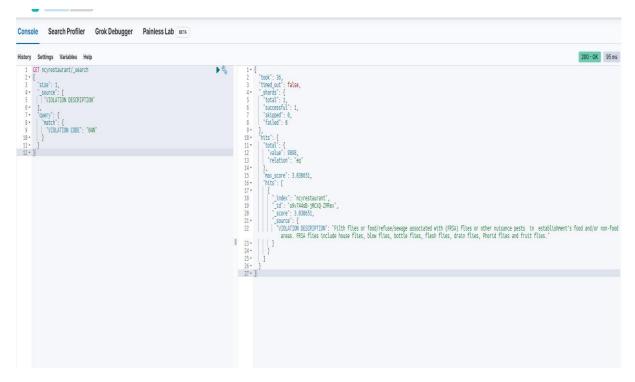


Figure 3 : Signification du code de violation « 04N »

### Q4 : Où se trouvent les restaurants (nom, adresse, quartier) notés A ?

Cette requête récupère les restaurants notés "A" et renvoie les informations de nom, adresse et quartier. Elle utilise une requête "match" pour filtrer les documents avec la note "A".

```
"Index": "ncyrestaurant"
"Id: "#9V7448B-]#CXQ-ZPRex",
"score": 0.3641192,
"source": 8.3641192,
"BUILING": "38",
"BURDING": "38",
"BORO": "Manhattan",
"ZIPCOBE": "18093",
"STREET": "EAST 19 STREET"
                                                                                                                                                                                                                                                                                                                                                                                             "_index": "ncyrestaurant",
"id": "qNv744sB-jMCXQ-ZPRex",
"_score": 0.3641192,
"source": (")B8A": "CASA GALICIA",
"BUTLDING": "3789",
"8000": "Opener"
"_index": "ncyrestaurant",
"id": "NN7448B-]MCQ-ZPRex",
"score": 0.3641192.
"Source": {
"DBA": "PANERA BREAD",
"BUILDING": 2358
"BORD": "Queens"
"ZIPCODE": "BELL BOULEVARD"
                                                                                                                                                                                                                                                                                                                                                                                            "_tndex": "ncyrestaurant",
"_td": "rtv7448-JMCXQ-ZPRex",
"score": 0,364192;
"source": {
"DBA": DUNKIN"
"BUILDING": '7111",
"BUBO: Brooklyn",
"CTRETE": 3 AVENUE"
                                                                                                                                                                                                                                                                                                                                                                                       "tndex": "ncyrestaurant",
"'tdf: "f9V7448B-MCXQ-ZFRex",
"score": 6 3641192.
"Source": ("'DBA': "Yong Kong Street",
"BUILDING": "18085",
"BORO": "Manhattan",
   "index": "ncyrestaurant",
"Ld": "SNV7458-NCXV2-PPREX",
"Score": 8,041192,
"Source": {
"DBA": ZIA MARIA LITTLE ITALY",
"BUILDING": "138",
"BURO": "MANHATEM",
ZIFCODE: "10912"
"STREET": "MULBERRY STREET"
                                                                                                                                                                                                                                                                                                                                                                                                      "tindex": "ncyrestaurant",
"Id": "stv74485-JMCX0-ZPRex",
"score": 8, 364192,
"source": {
"BOLIDING: "HIE GROUNDS OF BROOKLYN",
"BOLIDING: "BOLIDING: "STREET": "NEWKIRK AVENUE"
}
"STREET": "NEWKIRK AVENUE"
                                                                                                                                                                                                                                                                                                                                                                                                   "_index": "ncyrestaurant",
"_td": "tNv744sB-jMCXQ-ZPRex",
"score": 0.3641192,
"source": {
"DBA": "KOMBIT KREYOL",
"BUILDING": "279",
"BORO": "Brooklun".
```

Figure 4-8: Les restaurants (nom, adresse, quartier) notés A

### Q5 : Quelle est la cuisine la plus populaire ?

Cette requête utilise une agrégation "terms" pour déterminer la cuisine la plus populaire en fonction du champ "CUISINE DESCRIPTION". La taille est définie sur 1 pour renvoyer uniquement la cuisine la plus populaire. Pour obtenir la cuisine la plus populaire par quartier, vous devrez ajouter une sous-agrégation "terms" pour le champ "BORO" comme dans l'exemple ci-dessous.

Figure 9: La cuisine la plus populaire

### Et par quartier?

Figure 10 : La cuisine la plus populaire par quartier

## Q6 : Quelle est la date de la dernière inspection ?

Cette requête récupère la date de la dernière inspection en triant les documents par "INSPECTION DATE" dans l'ordre décroissant et en renvoyant le premier document.

```
History Settings Variables Help

| Comparison of Compariso
```

Figure 11 : La date de la dernière inspection

### Q7 : Fournissez la liste des restaurants chinois notés à Brooklyn :

Cette requête utilise une requête "bool" avec deux conditions "must" pour filtrer les restaurants chinois situés à Brooklyn.



Figure 12-14 : La liste des restaurants chinois notés à Brooklyn

### Q8 : Quelle est l'adresse du restaurant LADUREE ?

Cette requête utilise une requête "match" pour filtrer les documents avec le nom "the LADUREE restaurant" et renvoie l'adresse (champ "STREET") du premier document trouvé.

Figure 13 : L'adresse du restaurant LADUREE

# Q9 : Quelle cuisine est la plus touchée par l'infraction « Aliment chaud non maintenu à 140 °F ou plus » ?

Cette requête utilise une requête "match" pour filtrer les documents avec le code de violation "02B" et une agrégation "terms" pour déterminer la cuisine la plus touchée par cette infraction.

```
#Question 9:
                                                                                                                                                            "took": 1,
"timed_out": false,
                                                                                                                           ▶ ৩ৢৢ
3 get ncyrestaurant/_search
                                                                                                                                                              "_shards":
"total":
          {
    "_source": ["CUISINE DESCRIPTION"],
    "size": 0,
    "query": {
        "match": {
        | "VIOLATION DESCRIPTION": "Hot fo
                                                                                                                                                               "successful": 1,
"skipped": 0,
"failed": 0
                                                                                                                                                           hits":
                  "VIOLATION DESCRIPTION": "Hot food item not held at or above 140°
10 ^ | }
11 ^ },
12 * "agg.
13 * | "to
                                                                                                                                                                 its": {
"total": {|
"value": 10000,
"relation": "gte"
          13
                                                                                                                                                               },
"max_score": null,
"hits": []
                  14 -
                                                                                                                                                 15
 15
16
                                                                                                                                                              , aggregations": {
  "top_cuisine": {
   "doc_count_error_upper_bound": 0,
   "sum_other_doc_count": 166008,
   "buckets": [
 17 -
18
19 •
                                                                                                                                                 20
20 ^
21 ^
22 ^
23 ^ }
                                                                                                                                                                     {
    "key": "American",
    "doc_count": 33321
                                                                                                                                                 25
                                                                                                                                                 26 ^
27 ^
28 ^
29 ^
```

Figure 14 : La cuisine est la plus touchée par l'infraction « Aliment chaud non maintenu à 140 °F ou plus »

### Q10 : Quelles sont les violations les plus courantes (Top 5) ?

Ces requêtes sont conçues pour créer une cartographie pour un champ nommé "DESCRIPTION DE VIOLATION", puis effectuer une agrégation pour trouver les cinq premières descriptions de violation dans l'index "ncyrestaurant".

La première requête est une requête PUT qui crée une cartographie pour le champ "DESCRIPTION DE VIOLATION". Ce champ est de type "texte" et a un sous-champ "mot-clé" de type "mot-clé". Cela permet au champ d'être utilisé dans des requêtes au niveau des termes, telles que le tri ou l'agrégation, qui ne sont généralement pas prises en charge par les champs de texte.

La deuxième requête est une requête GET qui effectue une agrégation de termes sur le champ "DESCRIPTION DE VIOLATION.mot-clé". Cela renverra les cinq descriptions de violation les plus courantes dans l'index "ncyrestaurant". Le paramètre "size" est défini sur 0, ce qui signifie qu'aucun document ne sera renvoyé dans la section "hits" de la réponse. Au lieu de cela, la réponse inclura uniquement les résultats de l'agrégation.

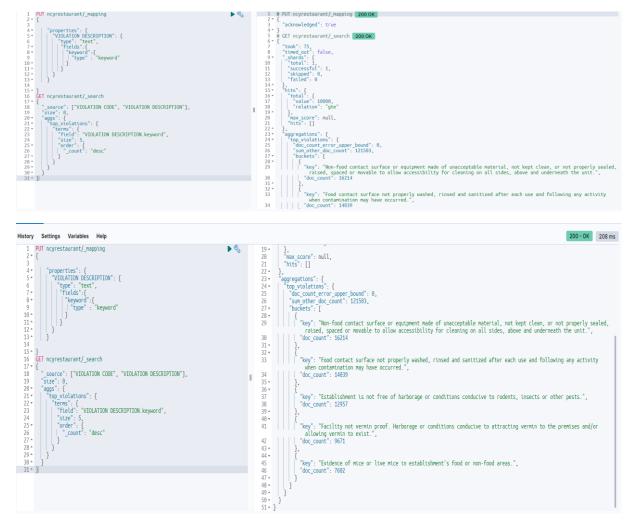


Figure 14: Les violations les plus courantes (Top 5)

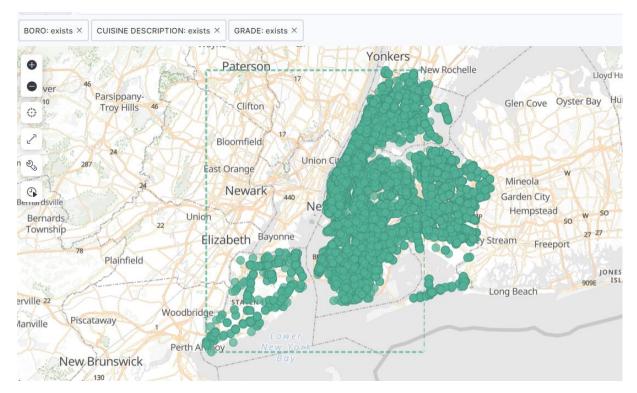
### Q11 : Quelle chaîne de restaurants est la plus populaire ?

Cette série de requêtes crée un nouvel index avec le champ "DBA" en tant que type "keyword", réindexe les données de l'ancien index vers le nouvel index, supprime l'ancien index et ajoute un alias pour le nouvel index. Enfin, elle utilise une agrégation "terms" pour déterminer la chaîne de restaurants la plus populaire en fonction du champ "DBA".

```
istory Settings Variables
                                                                                                                                      # PUT new-ncyrestaurant 200 OK * {
        PUT new-ncyrestaurant
                                                                                                                                             "acknowledged": true,
"shards_acknowledged": true,
"index": "new-ncyrestaurant"
             "mappings": {
                 "properties": {
   "DBA": {
    "type": "keyword"
                                                                                                                                         # POST _reindex 200 OK
                                                                                                                                             "took": 29370,
"timed_out": false,
"total": 208480,
"updated": 0,
"created": 208480,
"deleted": 0,
"batches": 209,
  10 ^ }
  10 ^ }
11 PC
12 * {
13 *
14
         POST _reindex
            "source": {
| "index": "ncyrestaurant"
            ; "dest": {
| "index": "new-ncyrestaurant"
                                                                                                                                              'version_conflicts": 0,
                                                                                                                                              "noops": 0,
"retries": {
"bulk": 0,
"search": 0
  18 -
  19 ^ }
20 DELETE ncyrestaurant
                                                                                                                                             },|
"throttled_millis":
                                                                                                           ► 2/3
            "actions": [
                                                                                                                                             "requests_per_second": -1,
"throttled_until_millis": 0,
23 -
                                                                                                                                  23
  24 <del>*</del> 25 <del>*</del>
                  "remove": {
    "index": "new-ncyrestaurant",
    "alias": "ncyrestaurant"
                                                                                                                                             "failures": []
                                                                                                                                 27 # DELETE ncyrestaurant 200 OK
  26
27
                                                                                                                         \parallel
  28 4
                                                                                                                                             "acknowledged": true
               {
    "add": {
        "index": "new-ncyrestaurant",
        "alias": "ncyrestaurant"
                                                                                                                                  30 - 3
  30 -
  31 ▼
32
                                                                                                                                            POST _aliases 200 OK
                                                                                                                                             "acknowledged": true
  33
                                                                                                                                  33
                                                                                                                                  35 # GET new-ncyrestaurant/_search 200 OK
  35 ^
                                                                                                                                 40 ° l
49     "acknowledged": true
        DELETE ncyrestaurant
          POST _aliases
  22 + {
23 - "actions": [
                                                                                                                                        # GET new-ncyrestaurant/_search 200 OK
                                                                                                                                 51
  24 <del>*</del>
25 <del>*</del>
                {
    "remove": {
        "index": "new-ncyrestaurant",
        "alias": "ncyrestaurant"
                                                                                                                                            "timed_out": false,
                                                                                                                                             "shards": {
  "total": 1,
  "successful": 1,
  "skipped": 0,
  "failed": 0
                                                                                                                                 55 <del>*</del> 56 57 58
                                                                                                                         \parallel
  28 -
  29 ^
30 ₹
                   "add": {
    "index": "new-ncyrestaurant",
    "alias": "ncyrestaurant"
                                                                                                                                          |
| "hits": {
| "total": {
| "value": 10000,
| "relation": "gte"
| " pull,
  31 ▼
32
                                                                                                                                 60 -
  33
  34 ^
35 ^
                                                                                                                                 63
  36 ^
37 ^ }
                                                                                                                                 65 ^
                                                                                                                                              "max_score": null,
"hits": []
                                                                                                                                 66
67
  38
         GET new-ncyrestaurant/_search
                                                                                                                                 68 -
             "size": 0,
  40
                                                                                                                                                ggregations": {
"popular_chains": {
    "doc_count_error_upper_bound": 0,
    "sum_other_doc_count": 205372,
                                                                                                                                              aggregations":
                                                                                                                                 70 🕶
                 43 -
                                                                                                                                 73 -
                                                                                                                                                   "buckets": [
  45
                                                                                                                                                     {
| "key": "DUNKIN"
                                                                                                                                 75
                                                                                                                                                        "doc_count": 2565
  48 -
                                                                                                                                 78 -
  50 ^
                                                                                                                                 80 -
  52 ^ }
```

Figure 14 : La chaîne de restaurants la plus populaire

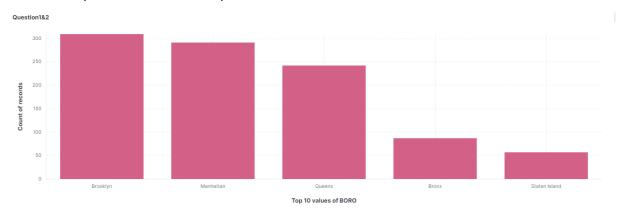
### Réponse à 3éme partie de l'examen :



Kibana offre une variété de filtres qui peuvent être utilisés pour affiner les résultats de recherche. Par exemple, le filtre 'exists' vérifie si un champ spécifique est présent dans un document. Cependant, si vous souhaitez filtrer les résultats en fonction d'une valeur spécifique d'un champ, vous pouvez utiliser des filtres tels que is, is not

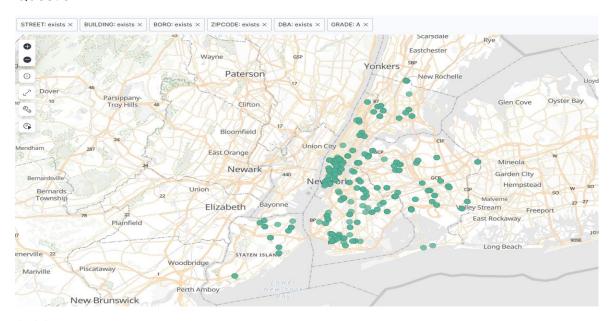
## Visualisation pour les questions de la première partie :

Dans le cadre de ce rapport, j'ai choisi de ne pas visualiser certaines questions car il ne m'a pas semblé pertinent de créer des graphiques à partir d'une seule donnée. Cependant, vous trouverez certaines présentations avec une donnée unique pour illustrer la procédure à suivre pour arriver à la visualisation.





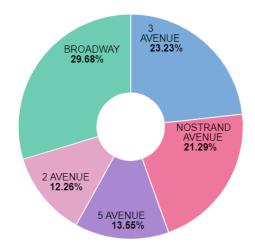
## Question4:



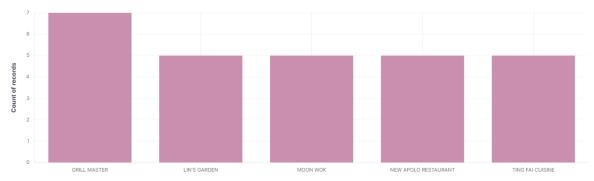


Top value of CUISINE DESCRIPTION

### Question5/Parquartier

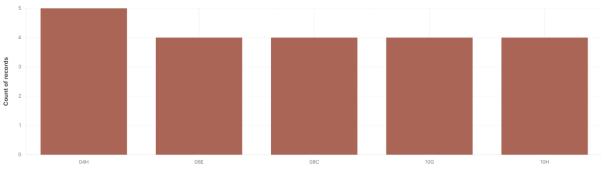


#### Question7



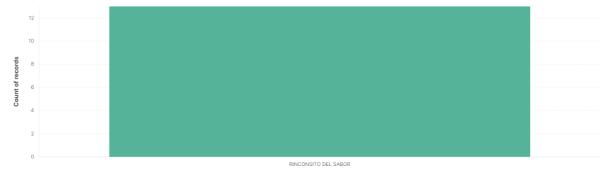
### Top 5 values of DBA

### Question10



Top 5 values of VIOLATION CODE

### Question11



Top value of DBA