



SORBONNE UNIVERSITÉ

DAAR Projet_ElasticSearch

Auteurs:

Hongyu YAN, Liuyi CHEN

Professeur :

Alfred Deivassagayame

Exercice 2

Nous avons choisi de visualiser les données des dix dernières années.

Puisque le champ "DBA" et "CUISINE DESCRIPTION" étaient initialement de type "Text" et que l'agrégation ne pouvait pas être faite pour des champs de ce type. Nous avons donc créé un nouvel index dont ces champs sont en type "keyword". Notre index initiale est nommé `new_york_city_restaurant` et l'index où nous avons changé le type de champs en "keyword" est nommé `nyc_restaurant_type_changed`

```
PUT /nyc_restaurant_type_changed
{
  "mappings" : {
    "properties": {
      "CUISINE DESCRIPTION": {
        "type": "keyword"
      },
      "VIOLATION DESCRIPTION": {
        "type": "keyword"
      },
      "DBA": {
        "type": "keyword"
      }
    }
  }
}
```

```

    }
  }
}

POST /_reindex?wait_for_completion=false
{
  "source": {
    "index": "new_york_city_restaurant"
  },
  "dest": {
    "index": "nyc_restaurant_type_changed"
  }
}

```

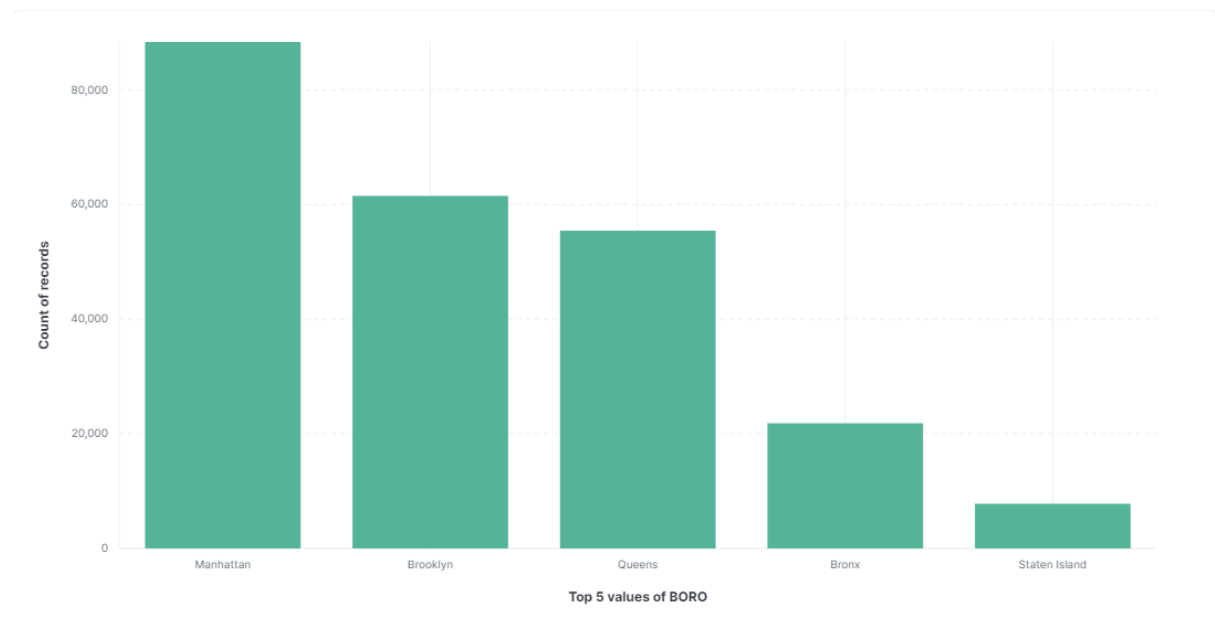
Q1 : Citez tous les quartiers de New-York.

Les quartiers sont: Manhattan, Brooklyn, Queens, Bronx et Staten Island.

```

GET /new_york_city_restaurant/_search
{
  "aggs" : {
    "quartiers" : {
      "terms" : {
        "field": "BORO"
      }
    }
  },
  "size" : 0
}

```



Q2 : Quelle est le quartier qui recense le plus restaurant ?

D'après le résultat de la question précédente, c'est le quartier Manhattan.

```
GET /new_york_city_restaurant/_search
{
  "size": "0",
  "aggs" : {
    "uniq_gender" : {
      "terms" : {
        "field" : "BORO",
        "size": "1"
      }
    }
  }
}
```

Q2_Quartier qui recense le plus restaurant

Top value of BORO	Count of records
Manhattan	88,406

Q3 : A quoi correspond le code de violation “04N”

Ce code correspond à deux descriptions différentes

1. “Filth flies or food/refuse/sewage-associated (FRSA) flies present in facility’s food and/or non-food areas. Filth flies include house flies, little house flies, blow flies, bottle flies and flesh flies. Food/refuse/sewage-associated flies include fruit flies, drain flies and Phorid flies.”
2. "Filth flies or food/refuse/sewage associated with (FRSA) flies or other nuisance pests in establishment’s food and/or non-food areas. FRSA flies include house flies, blow flies, bottle flies, flesh flies, drain flies, Phorid flies and fruit flies.”

```
GET /nyc_restaurant_type_changed/_search
{
  "query":{
    "term" :{
      "VIOLATION CODE.keyword": "04N"
    }
  }
```

```

    },
    "_source": ["VIOLATION CODE", "VIOLATION DESCRIPTION"],
    "collapse": {
      "field": "VIOLATION DESCRIPTION"
    }
  }
}

```

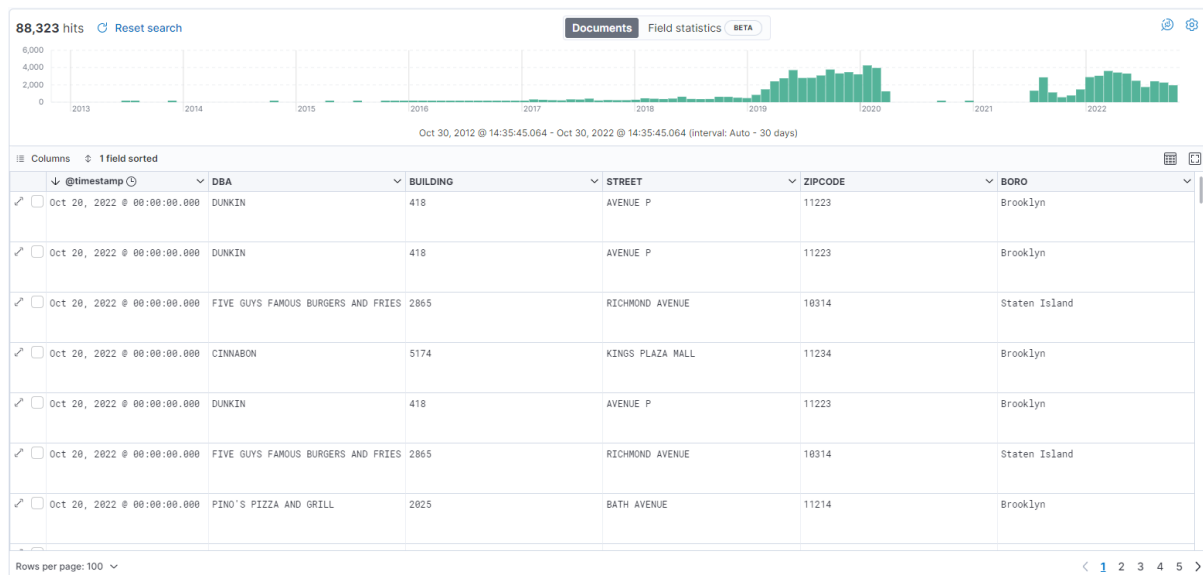
VIOLATION DESCRIPTION	VIOLATION CODE
Filth flies or food/refuse/sewage-associated (FRSA) flies present in facility's food and/or no...	04N
Filth flies or food/refuse/sewage associated with (FRSA) flies or other nuisance pests in...	04N

Q4 : Où se trouve les restaurants (nom, adresse, quartier) dont la grade est A ?

```

GET /new_york_city_restaurant/_search
{
  "query": {
    "bool": {
      "must": [
        {"match": {
          "GRADE": "A"
        }}
      ]
    }
  },
  "_source": [
    "DBA",
    "BUILDING",
    "STREET",
    "ZIPCODE",
    "BORO"
  ]
}

```



Q5 : Quelle est la cuisine la plus populaire ? Et par quartier ?

C'est la cuisine américaine qui est la plus populaire dans la ville et pareil pour les quartiers.

- Q5_1

```
GET /new_york_city_restaurant/_search
{
  "size" : 0,
  "aggs" : {
    "cuisine_pop" : {
      "terms" : {
        "field" : "CUISINE DESCRIPTION",
        "size": 1
      }
    }
  }
}
```

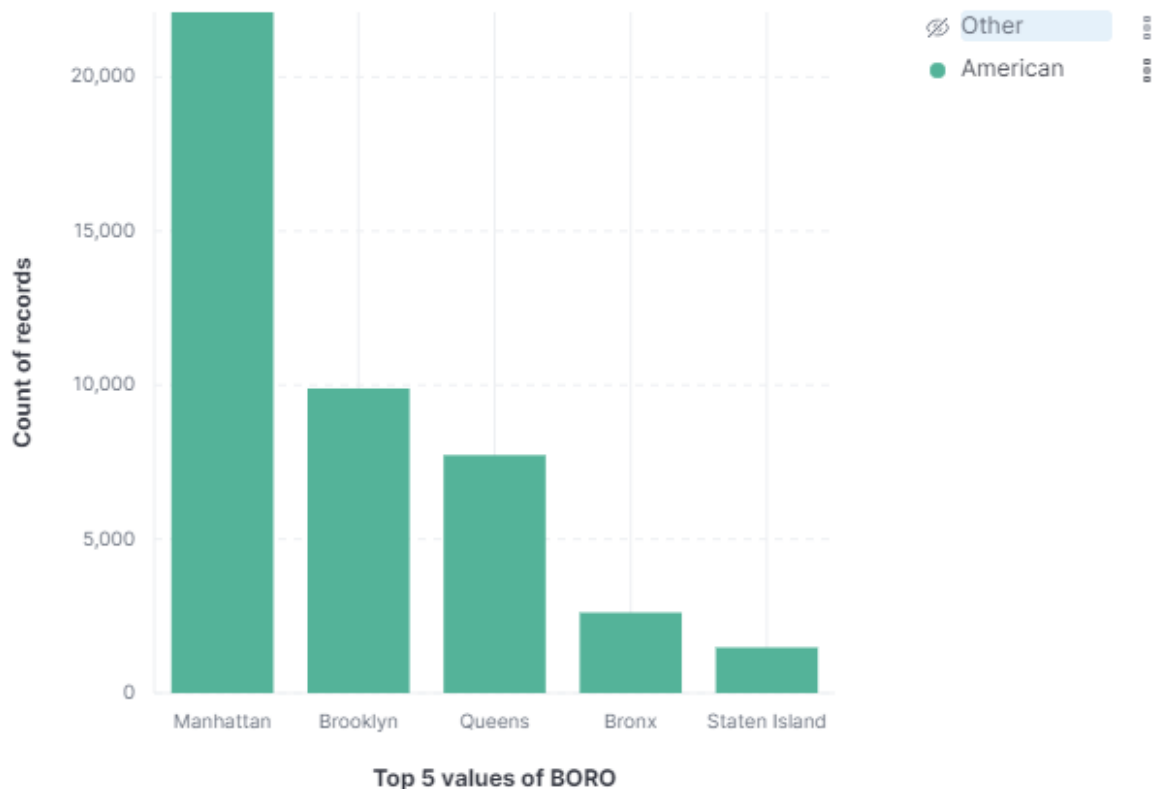
Q5.1_La cuisine la plus populaire à NYC

Top value of CUISINE DESCRIPTION	Count of records
American	43,819

- Q5_2

```
GET /nyc_restaurant_type_changed/_search
{
  "size" : 0,
  "aggs" : {
    "quartier" : {
      "terms": {
        "field" : "BORO.keyword"
      },
      "aggs" : {
        "cuisine_pop" : {
          "terms" : {
            "field" : "CUISINE DESCRIPTION",
            "size": 1
          }
        }
      }
    }
  }
}
```

Q5.2 La top cuisine de chaque quartier



Q6 : Quelle est la date de la dernière inspection ?

La dernière inspection c'est le 21 octobre 2022. Nous trions la date en ordre décroissant pour obtenir la date la plus récente.

```
GET /new_york_city_restaurant/_search
{
  "sort" : [
    {
      "INSPECTION DATE" : {
        "order" : "desc"
      }
    }
  ],
  "size": 1,
  "_source": [
    "INSPECTION DATE"
  ]
}
```

Q6_La date de la dernière inspection



Maximum of INSPECTION DATE

Oct 21, 2022 @ 02:00:00.000

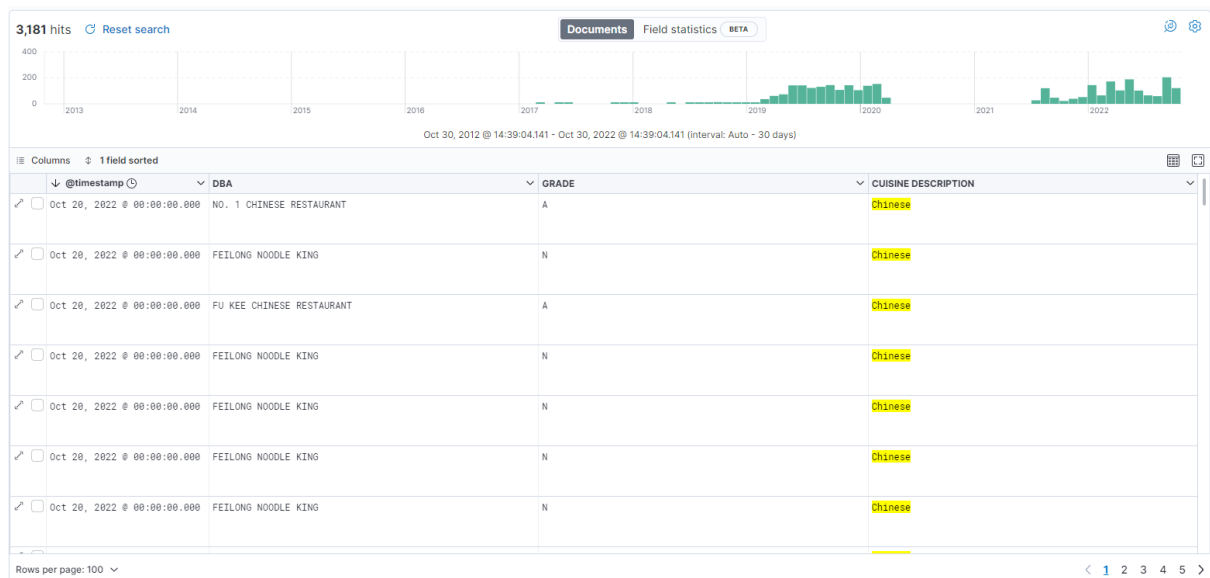
Q7 : Donnez la liste des restaurants chinois gradés à Brooklyn

```
GET /new_york_city_restaurant/_search
{
  "query": {
    "bool": {
      "must": [
        {
          "exists": {
            "field": "GRADE"
          }
        },
        {
          "match": {
            "CUISINE DESCRIPTION": "Chinese"
          }
        },
        {
          "match": {
            "BORO": "Brooklyn"
          }
        }
      ]
    }
  }
}
```

```

    }
  }
]
}
},
"collapse": {
  "field": "CAMIS"
},
"_source": [
  "DBA",
  "CUISINE DESCRIPTION",
  "GRADE"
]
}

```



Q8 : Quelle est l'adresse du restaurant LADUREE ?

Donc l'adresse du restaurant LADUREE est "864, MADISON AVENUE, 10021".

```

GET /nyc_restaurant_type_changed/_search
{
  "size": 0,
  "query": {
    "match": {
      "DBA": "LADUREE"
    }
  },
  "aggs" : {
    "query_adress" : {
      "multi_terms": {
        "terms": [
          {

```



```

        "field": "DBA"
      }, {
        "field": "BUILDING.keyword"
      }, {
        "field": "STREET.keyword"
      }, {
        "field": "ZIPCODE.keyword"
      }
    ]
  }
}

```

Q8_L'adresse du restaurant LADUREE

DBA	BUILDING	STREET	ZIPCODE	Count of records
LADUREE	864	MADISON AVENUE	10021	5

Q9 : Quelle est la cuisine la plus touchée par la violation "Hot food item not held at or above 140° F"

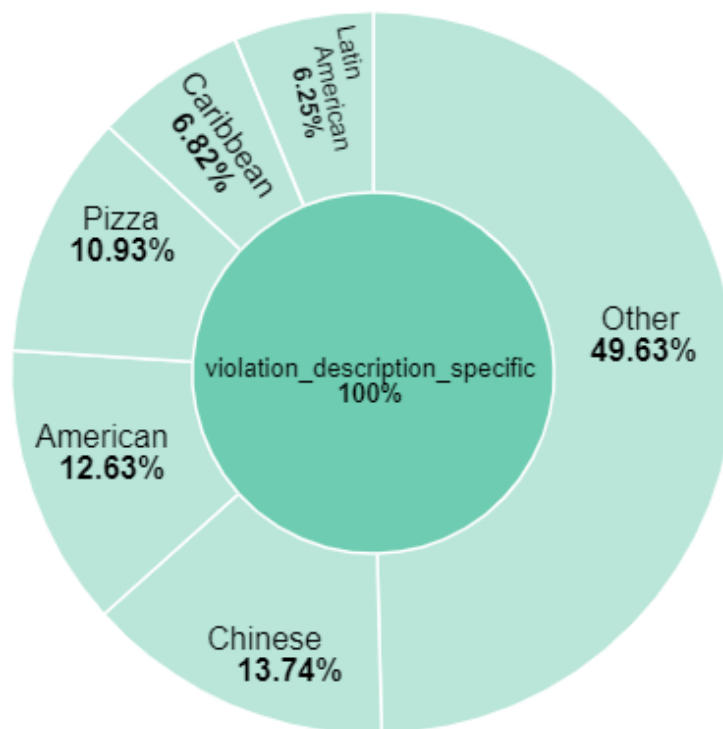
Dans la visualisation nous pouvons voir que c'est la cuisine chinoise qui est la plus touchée par la violation demandée ("Hot food item not held at or above 140° F")

```

GET /new_york_city_restaurant/_search
{
  "size" : 0,
  "query" : {
    "match_phrase": {
      "VIOLATION DESCRIPTION": "Hot food item not held at or above 140° F"
    }
  },
  "aggs" : {
    "cuisine_pop" : {
      "terms" : {
        "field" : "CUISINE DESCRIPTION",
        "size" : 1
      }
    }
  }
}

```

Q9_La cuisine chinoise est la plus touchée par la violation demandée ⓘ



Q10 : Quelle sont les violations les plus courants (Top 5) ?

Les top 5 sont: 10F, 08A, 04L, 06D et 02G

```
GET /new_york_city_restaurant/_search
{
  "size" : 0,
  "aggs" : {
    "top_five_viol" : {
      "terms": {
        "field" : "VIOLATION CODE",
        "size": 5
      }
    }
  }
}
```

Top 5 values of VIOLATION CODE	Count of records
10F	40,785
08A	25,216
04L	15,709
06D	15,592
02G	14,298

Q11 : Quelle est la chaine de restaurants la plus populaire ?

C'est "DUNKIN".

```
GET /nyc_restaurant_type_changed/_search
{
  "size" : 0,
  "aggs" : {
    "type_count" : {
      "terms": {
        "field" : "DBA",
        "size": 1
      }
    }
  }
}
```

Q11_ La chaine de restaurants la plus populaire

Top value of DBA	Count of records
DUNKIN	2,828

Data source : <https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j>

