

PROGETTINO BASI DI DATI

1) SCELTA DEI DATASET

Ho scelto un dataset trovato su Kaggle relativo a tutti i crimini avvenuti a Londra tra il 2009 e il 2017 dato che è di buone dimensioni (circa 800 MB) e potrebbe fare vedere "ampie differenze" per quanto riguarda i tempi di esecuzione. Successivamente vista la presenza di soli 7 campi (Codice ID, quartiere, maggior accusa, minor accusa, numero di reati commessi nel dato quartiere in quel mese dell'anno, mese e anno) ho deciso di integrare la mia base di dati con un altro dataset riguardante le varie religioni presenti nei vari quartieri di Londra (trovato sugli Open Data del comune di Londra).

2) MODELLAZIONE

Avendo solo due documenti (o tabelle) ho deciso di creare con 2 script in pymongo, 2 basi di dati, una basata sugli embeddings del campo quartiere (relazione 1-1) e l'altra basata sui reference sempre del campo quartiere. Gli script in pymongo sono presenti nella cartella della consegna.

```
bruno@ubuntito-ThinkPad-T460s:~$ cd Desktop/Progetto\ Basi/
bruno@ubuntito-ThinkPad-T460s:~/Desktop/Progetto Basi$ ./import_embedded.py
44.57532358169556
bruno@ubuntito-ThinkPad-T460s:~/Desktop/Progetto Basi$ ./import_reference.py
29.823797941207886
bruno@ubuntito-ThinkPad-T460s:~/Desktop/Progetto Basi$
```

Il caricamento della base di dati con embedding ha richiesto il 50% di tempo in più rispetto alla semplice creazione dei due documenti (modello reference)

```
bruno@ubuntito-ThinkPad-T460s:~$ sudo mongo --host 127.0.0.1:27017
MongoDB shell version v3.6.5
connecting to: mongodb://127.0.0.1:27017/
MongoDB server version: 3.6.5
Server has startup warnings:
2018-06-05T13:10:26.892+0200 I STORAGE [initandlisten]
2018-06-05T13:10:26.892+0200 I STORAGE [initandlisten] ** WARNING: Using the XFS
filesystem is strongly recommended with the WiredTiger storage engine
2018-06-05T13:10:26.892+0200 I STORAGE [initandlisten] ** See http://d
ochub.mongodb.org/core/prodnotes-filesystem
2018-06-05T13:10:27.729+0200 I CONTROL [initandlisten]
2018-06-05T13:10:27.729+0200 I CONTROL [initandlisten] ** WARNING: Access contr
ol is not enabled for the database.
2018-06-05T13:10:27.729+0200 I CONTROL [initandlisten] ** Read and wri
te access to data and configuration is unrestricted.
2018-06-05T13:10:27.729+0200 I CONTROL [initandlisten]
> show dbs
LondonCrime          0.816GB
LondonCrimesEmbedded 0.094GB
LondonCrimesReference 0.067GB
admin                 0.000GB
config                0.000GB
local                 0.000GB
>
```

Per quanto riguarda le dimensioni la base memorizzata per embedding pesa circa il 50% in più della base per reference, mentre LondonCrime, il database creato usando la mongoinport pesa 9 volte tanto!!

Per quanto riguarda la struttura dell'unico documento che compone il database per embedding la sua struttura è la seguente:

```
> db.collection.findOne()
{
  "_id" : ObjectId("5b1672d97d6cd135ebbd3a4c"),
  "Crime_Information" : {
    "lsoa_code" : "E01001116",
    "borough" : {
      "All people" : 363378,
      "Christian" : 205022,
      "Buddhist" : 2381,
      "Hindu" : 21739,
      "Jewish" : 709,
      "Muslim" : 29513,
      "Sikh" : 1450,
      "Other religion" : 2153,
      "No religion" : 72654,
      "Religion not stated" : 27757
    },
    "major_category" : "Burglary",
    "minor_category" : "Burglary in Other Buildings",
    "year" : 2016,
    "month" : 11
  }
}
```

Mentre per quanto riguarda la struttura dei due documenti del db per reference:

```
> db.crimes.findOne()
{
  "_id" : ObjectId("5b16730b7d6cd136063bfb0a"),
  "Crime_Information" : {
    "lsoa_code" : "E01001116",
    "borough" : "Croydon",
    "major_category" : "Burglary",
    "minor_category" : "Burglary in Other Buildings",
    "year" : 2016,
    "month" : 11
  }
}
> db.borough.findOne()
{
  "_id" : ObjectId("5b1673017d6cd136063bf8ae"),
  "All people" : 185911,
  "Christian" : 104101,
  "Buddhist" : 842,
  "Hindu" : 4464,
  "Jewish" : 425,
  "Muslim" : 25520,
  "Sikh" : 2952,
  "Other religion" : 533,
  "No religion" : 35106,
  "Religion not stated" : 11968
}
```

3)QUERIES E TEMPI DI ESECUZIONE CON E SENZA INDICI

MODELLO EMBEDDED:

Q1 Senza indice

```
> var prima = new Date()
> db.collection.aggregate([
...   {
...     $group: {
...       _id: {
...         "name" : "$Crime_Information.borough.Name",
...         "number_of_christians" : "$Crime_Information.borough.Christian"
...       }
...     },
...     {
...       $sort: {
...         "_id.number_of_christians" : 1
...       }
...     },
...     {
...       $project: {
...         "_id" : 0,
...         "name" : "$_id.name",
...         "number_of_christians" : "$_id.number_of_christians"
...       }
...     }
...   ]
{ "name" : "City of London", "number_of_christians" : 3344 }
{ "name" : "Tower Hamlets", "number_of_christians" : 75087 }
{ "name" : "Kingston upon Thames", "number_of_christians" : 84684 }
{ "name" : "Camden", "number_of_christians" : 84737 }
{ "name" : "Kensington and Chelsea", "number_of_christians" : 86005 }
{ "name" : "Harrow", "number_of_christians" : 89181 }
{ "name" : "Islington", "number_of_christians" : 90107 }
{ "name" : "Hackney", "number_of_christians" : 95131 }
{ "name" : "Hammersmith and Fulham", "number_of_christians" : 98808 }
{ "name" : "Redbridge", "number_of_christians" : 102755 }
{ "name" : "Richmond upon Thames", "number_of_christians" : 103319 }
{ "name" : "Barking and Dagenham", "number_of_christians" : 104101 }
{ "name" : "Hounslow", "number_of_christians" : 106660 }
{ "name" : "Sutton", "number_of_christians" : 111086 }
{ "name" : "Merton", "number_of_christians" : 111993 }
{ "name" : "Haringey", "number_of_christians" : 114659 }
{ "name" : "Newham", "number_of_christians" : 123119 }
{ "name" : "Waltham Forest", "number_of_christians" : 124939 }
{ "name" : "Brent", "number_of_christians" : 129080 }
{ "name" : "Greenwich", "number_of_christians" : 134632 }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 1.422seconds
> □
```

Dopo aver creato un indice sul campo Crime_information.borough

```
> db.collection.ensureIndex({"Crime_information.borough":1})
{
  "createdCollectionAutomatically" : false,
  "numIndexesBefore" : 1,
  "numIndexesAfter" : 2,
  "ok" : 1
}
```


Q1 Con indice

Rieseguo la query e con l'indice stranamente ci mette 20 centesimi di secondo in più.

```
> var prima = new Date()
> db.collection.aggregate([
...   {
...     $group: {
...       _id: {
...         "name" : "$Crime_Information.borough.Name",
...         "number_of_christians" : "$Crime_Information.borough.Christian"
...       }
...     },
...     {
...       $sort: {
...         "_id.number_of_christians" : 1
...       }
...     },
...     {
...       $project: {
...         "_id" : 0,
...         "name" : "$_id.name",
...         "number_of_christians" : "$_id.number_of_christians"
...       }
...     }
...   ]
... ])

{ "name" : "City of London", "number_of_christians" : 3344 }
{ "name" : "Tower Hamlets", "number_of_christians" : 75087 }
{ "name" : "Kingston upon Thames", "number_of_christians" : 84684 }
{ "name" : "Camden", "number_of_christians" : 84737 }
{ "name" : "Kensington and Chelsea", "number_of_christians" : 86005 }
{ "name" : "Harrow", "number_of_christians" : 89181 }
{ "name" : "Islington", "number_of_christians" : 90107 }
{ "name" : "Hackney", "number_of_christians" : 95131 }
{ "name" : "Hammersmith and Fulham", "number_of_christians" : 98808 }
{ "name" : "Redbridge", "number_of_christians" : 102755 }
{ "name" : "Richmond upon Thames", "number_of_christians" : 103319 }
{ "name" : "Barking and Dagenham", "number_of_christians" : 104101 }
{ "name" : "Hounslow", "number_of_christians" : 106660 }
{ "name" : "Sutton", "number_of_christians" : 111086 }
{ "name" : "Merton", "number_of_christians" : 111993 }
{ "name" : "Haringey", "number_of_christians" : 114659 }
{ "name" : "Newham", "number_of_christians" : 123119 }
{ "name" : "Waltham Forest", "number_of_christians" : 124939 }
{ "name" : "Brent", "number_of_christians" : 129080 }
{ "name" : "Greenwich", "number_of_christians" : 134632 }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 1.442seconds
> □
```

Q2 Senza indice

```
type it for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 1.422seconds
> var prima = new Date()
> db.collection.aggregate([
...   {
...     $group : {
...       "_id": "$Crime_Information.year",
...       "crime" : {
...         $sum : 1
...       }
...     }
...   },
...   {
...     $sort : {
...       "crime" : 1
...     }
...   },
...   {
...     $project : {
...       "_id" : 0,
...       "crime" : 1,
...       "year" : "$_id"
...     }
...   }
... ])
{ "crime" : 112858, "year" : 2016 }
{ "crime" : 112980, "year" : 2015 }
{ "crime" : 113017, "year" : 2014 }
{ "crime" : 113255, "year" : 2009 }
{ "crime" : 113494, "year" : 2011 }
{ "crime" : 113521, "year" : 2012 }
{ "crime" : 113591, "year" : 2010 }
{ "crime" : 113596, "year" : 2008 }
{ "crime" : 113674, "year" : 2013 }
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.853seconds
> 
```

Q2 Con indice

```
21
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 1.442seconds
> var prima = new Date()
> db.collection.aggregate([
...   {
...     $group : {
...       "_id": "$Crime_Information.year",
...       "crime" : {
...         $sum : 1
...       }
...     }
...   },
...   {
...     $sort : {
...       "crime" : 1
...     }
...   },
...   {
...     $project : {
...       "_id" : 0,
...       "crime" : 1,
...       "year" : "$_id"
...     }
...   }
... ])
{ "crime" : 112858, "year" : 2016 }
{ "crime" : 112980, "year" : 2015 }
{ "crime" : 113017, "year" : 2014 }
{ "crime" : 113255, "year" : 2009 }
{ "crime" : 113494, "year" : 2011 }
{ "crime" : 113521, "year" : 2012 }
{ "crime" : 113591, "year" : 2010 }
{ "crime" : 113596, "year" : 2008 }
{ "crime" : 113674, "year" : 2013 }
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.87seconds
> 
```

Anche in questo caso con indice la query ci mette leggermente di più.

Q3 Senza indice

```
> var prima = new Date()
> db.collection.aggregate([
...   {
...     $match : {
...       "Crime_Information.month" : 12,
...       "Crime_Information.year" : 2015
...     }
...   },
...   {
...     $group : {
...       "_id" : "$Crime_Information.borough.Name",
...       "furti" : {$sum:1}
...     }
...   },
...   {
...     $sort : {
...       "furti": -1
...     }
...   },
...   {
...     $project :
...     {
...       _id : 0,
...       borough : "$_id",
...       furti : "$furti"
...     }
...   }
... ])
{ "borough" : "Barnet", "furti" : 421 }
{ "borough" : "Croydon", "furti" : 415 }
{ "borough" : "Ealing", "furti" : 398 }
{ "borough" : "Wandsworth", "furti" : 398 }
{ "borough" : "Lambeth", "furti" : 376 }
{ "borough" : "Enfield", "furti" : 369 }
{ "borough" : "Brent", "furti" : 363 }
{ "borough" : "Lewisham", "furti" : 361 }
{ "borough" : "Bromley", "furti" : 361 }
{ "borough" : "Southwark", "furti" : 355 }
{ "borough" : "Newham", "furti" : 344 }
{ "borough" : "Hillingdon", "furti" : 332 }
{ "borough" : "Redbridge", "furti" : 328 }
{ "borough" : "Greenwich", "furti" : 327 }
{ "borough" : "Hackney", "furti" : 325 }
{ "borough" : "Waltham Forest", "furti" : 298 }
{ "borough" : "Tower Hamlets", "furti" : 279 }
{ "borough" : "Havering", "furti" : 279 }
{ "borough" : "Hounslow", "furti" : 279 }
{ "borough" : "Camden", "furti" : 276 }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.507seconds
> █
```


Q3 Con indice

```
> var prima = new Date()
> db.collection.aggregate([
...   {
...     $match : {
...       "Crime_Information.month" : 12,
...       "Crime_Information.year" : 2015
...     }
...   },
...   {
...     $group : {
...       "_id" : "$Crime_Information.borough.Name",
...       "furti" : {$sum:1}
...     }
...   },
...   {
...     $sort : {
...       "furti": -1
...     }
...   },
...   {
...     $project :
...     {
...       _id : 0,
...       borough : "$_id",
...       furti : "$furti"
...     }
...   }
... ])
{ "borough" : "Barnet", "furti" : 421 }
{ "borough" : "Croydon", "furti" : 415 }
{ "borough" : "Ealing", "furti" : 398 }
{ "borough" : "Wandsworth", "furti" : 398 }
{ "borough" : "Lambeth", "furti" : 376 }
{ "borough" : "Enfield", "furti" : 369 }
{ "borough" : "Brent", "furti" : 363 }
{ "borough" : "Lewisham", "furti" : 361 }
{ "borough" : "Bromley", "furti" : 361 }
{ "borough" : "Southwark", "furti" : 355 }
{ "borough" : "Newham", "furti" : 344 }
{ "borough" : "Hillingdon", "furti" : 332 }
{ "borough" : "Redbridge", "furti" : 328 }
{ "borough" : "Greenwich", "furti" : 327 }
{ "borough" : "Hackney", "furti" : 325 }
{ "borough" : "Waltham Forest", "furti" : 298 }
{ "borough" : "Tower Hamlets", "furti" : 279 }
{ "borough" : "Havering", "furti" : 279 }
{ "borough" : "Hounslow", "furti" : 279 }
{ "borough" : "Camden", "furti" : 276 }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.498seconds
```

In questo caso i tempi migliorano leggermente con l'aggiunta dell'indice

Q4 Senza indice

```
{ "crime" : 8, "borough" : "City of London" }
{ "crime" : 279, "borough" : "Tower Hamlets" }
{ "crime" : 174, "borough" : "Kingston upon Thames" }
{ "crime" : 276, "borough" : "Camden" }
{ "crime" : 193, "borough" : "Kensington and Chelsea" }
{ "crime" : 236, "borough" : "Harrow" }
{ "crime" : 257, "borough" : "Islington" }
{ "crime" : 325, "borough" : "Hackney" }
{ "crime" : 217, "borough" : "Hammersmith and Fulham" }
{ "crime" : 328, "borough" : "Redbridge" }
{ "crime" : 219, "borough" : "Richmond upon Thames" }
{ "crime" : 218, "borough" : "Barking and Dagenham" }
{ "crime" : 279, "borough" : "Hounslow" }
{ "crime" : 219, "borough" : "Sutton" }
{ "crime" : 273, "borough" : "Merton" }
{ "crime" : 258, "borough" : "Haringey" }
{ "crime" : 344, "borough" : "Newham" }
{ "crime" : 298, "borough" : "Waltham Forest" }
{ "crime" : 363, "borough" : "Brent" }
{ "crime" : 327, "borough" : "Greenwich" }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.521seconds
```

Q4 Con indice

```
{ "crime" : 8, "borough" : "City of London" }
{ "crime" : 279, "borough" : "Tower Hamlets" }
{ "crime" : 174, "borough" : "Kingston upon Thames" }
{ "crime" : 276, "borough" : "Camden" }
{ "crime" : 193, "borough" : "Kensington and Chelsea" }
{ "crime" : 236, "borough" : "Harrow" }
{ "crime" : 257, "borough" : "Islington" }
{ "crime" : 325, "borough" : "Hackney" }
{ "crime" : 217, "borough" : "Hammersmith and Fulham" }
{ "crime" : 328, "borough" : "Redbridge" }
{ "crime" : 219, "borough" : "Richmond upon Thames" }
{ "crime" : 218, "borough" : "Barking and Dagenham" }
{ "crime" : 279, "borough" : "Hounslow" }
{ "crime" : 219, "borough" : "Sutton" }
{ "crime" : 273, "borough" : "Merton" }
{ "crime" : 258, "borough" : "Haringey" }
{ "crime" : 344, "borough" : "Newham" }
{ "crime" : 298, "borough" : "Waltham Forest" }
{ "crime" : 363, "borough" : "Brent" }
{ "crime" : 327, "borough" : "Greenwich" }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.502seconds
>
```

Anche in questo caso con l'indice ci mette meno.

Q5 Senza indice

```
{ "crime" : 8, "borough" : "City of London" }
{ "crime" : 270, "borough" : "Bexley" }
{ "crime" : 344, "borough" : "Newham" }
{ "crime" : 218, "borough" : "Barking and Dagenham" }
{ "crime" : 219, "borough" : "Sutton" }
{ "crime" : 327, "borough" : "Greenwich" }
{ "crime" : 361, "borough" : "Lewisham" }
{ "crime" : 279, "borough" : "Hounslow" }
{ "crime" : 415, "borough" : "Croydon" }
{ "crime" : 174, "borough" : "Kingston upon Thames" }
{ "crime" : 273, "borough" : "Merton" }
{ "crime" : 361, "borough" : "Bromley" }
{ "crime" : 355, "borough" : "Southwark" }
{ "crime" : 398, "borough" : "Ealing" }
{ "crime" : 376, "borough" : "Lambeth" }
{ "crime" : 279, "borough" : "Havering" }
{ "crime" : 217, "borough" : "Hammersmith and Fulham" }
{ "crime" : 298, "borough" : "Waltham Forest" }
{ "crime" : 279, "borough" : "Tower Hamlets" }
{ "crime" : 219, "borough" : "Richmond upon Thames" }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.575seconds
> █
```

Q5 Con indice

```
{ "crime" : 8, "borough" : "City of London" }
{ "crime" : 270, "borough" : "Bexley" }
{ "crime" : 344, "borough" : "Newham" }
{ "crime" : 218, "borough" : "Barking and Dagenham" }
{ "crime" : 219, "borough" : "Sutton" }
{ "crime" : 327, "borough" : "Greenwich" }
{ "crime" : 361, "borough" : "Lewisham" }
{ "crime" : 279, "borough" : "Hounslow" }
{ "crime" : 415, "borough" : "Croydon" }
{ "crime" : 174, "borough" : "Kingston upon Thames" }
{ "crime" : 273, "borough" : "Merton" }
{ "crime" : 361, "borough" : "Bromley" }
{ "crime" : 355, "borough" : "Southwark" }
{ "crime" : 398, "borough" : "Ealing" }
{ "crime" : 376, "borough" : "Lambeth" }
{ "crime" : 279, "borough" : "Havering" }
{ "crime" : 217, "borough" : "Hammersmith and Fulham" }
{ "crime" : 298, "borough" : "Waltham Forest" }
{ "crime" : 279, "borough" : "Tower Hamlets" }
{ "crime" : 219, "borough" : "Richmond upon Thames" }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.53seconds
```

L'indice velocizza le prestazioni anche in questo caso.

Q6 Senza indice

```
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.53seconds
>
> var prima = new Date()
> db.collection.aggregate([
...   {
...     $match : {
...       "Crime_Information.year" : 2015
...     }
...   },
...   {
...     $group : {
...       "_id" : {
...         "minor_category" : "$Crime_Information.minor_category",
...         "major_category" : "$Crime_Information.major_category"
...       },
...       "crime" : {
...         $sum : 1
...       }
...     }
...   },
...   {
...     $sort : {
...       "crime" : -1
...     }
...   },
...   {
...     $limit : 1
...   },
...   {
...     $project : {
...       "_id" : 0,
...       "minor_category" : "$_id.minor_category",
...       "major_category" : "$_id.major_category",
...       "crime" : 1
...     }
...   }
... ])
{ "crime" : 4540, "minor_category" : "Common Assault", "major_category" : "Violence Against the Person" }
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.63seconds
> □
```

Q6 Con indice

```
> var prima = new Date()
> db.collection.aggregate([
...   {
...     $match : {
...       "Crime_Information.year" : 2015
...     }
...   },
...   {
...     $group : {
...       "_id" : {
...         "minor_category" : "$Crime_Information.minor_category",
...         "major_category" : "$Crime_Information.major_category"
...       },
...       "crime" : {
...         $sum : 1
...       }
...     }
...   },
...   {
...     $sort : {
...       "crime" : -1
...     }
...   },
...   {
...     $limit : 1
...   },
...   {
...     $project : {
...       "_id" : 0,
...       "minor_category" : "$_id.minor_category",
...       "major_category" : "$_id.major_category",
...       "crime" : 1
...     }
...   }
... ])
{ "crime" : 4540, "minor_category" : "Common Assault", "major_category" : "Violence Against the Person" }
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.581seconds
> 
```

Anche in questo caso con l'indice la query ci mette meno

MODELLO REFERENCE:

Q1 Senza indice

```
> var prima = new Date()
> db.borough.aggregate([
...   {
...     $group : {
...       _id : {
...         "name" : "$Name",
...         "number_of_christians" : "$Christian"
...       }
...     },
...     {
...       $sort : {
...         "_id.number_of_christians" : 1
...       }
...     },
...     {
...       $project : {
...         "_id" : 0,
...         "name" : "$_id.name",
...         "number_of_christians" : "$_id.number_of_christians"
...       }
...     }
...   ])
{ "name" : "City of London", "number_of_christians" : 3344 }
{ "name" : "Tower Hamlets", "number_of_christians" : 75087 }
{ "name" : "Kingston upon Thames", "number_of_christians" : 84684 }
{ "name" : "Camden", "number_of_christians" : 84737 }
{ "name" : "Kensington and Chelsea", "number_of_christians" : 86005 }
{ "name" : "Harrow", "number_of_christians" : 89181 }
{ "name" : "Islington", "number_of_christians" : 90107 }
{ "name" : "Hackney", "number_of_christians" : 95131 }
{ "name" : "Hammersmith and Fulham", "number_of_christians" : 98808 }
{ "name" : "Redbridge", "number_of_christians" : 102755 }
{ "name" : "Richmond upon Thames", "number_of_christians" : 103319 }
{ "name" : "Barking and Dagenham", "number_of_christians" : 104101 }
{ "name" : "Hounslow", "number_of_christians" : 106660 }
{ "name" : "Sutton", "number_of_christians" : 111086 }
{ "name" : "Merton", "number_of_christians" : 111993 }
{ "name" : "Haringey", "number_of_christians" : 114659 }
{ "name" : "Newham", "number_of_christians" : 123119 }
{ "name" : "Waltham Forest", "number_of_christians" : 124939 }
{ "name" : "Brent", "number_of_christians" : 129080 }
{ "name" : "Greenwich", "number_of_christians" : 134632 }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.012seconds
> █
```

Creazione indice sul campo Crime_information.borough

```
> db.crimes.ensureIndex({"Crime_information.borough":1})
{
  "createdCollectionAutomatically" : false,
  "numIndexesBefore" : 1,
  "numIndexesAfter" : 2,
  "ok" : 1
}
>
```

Q1 Con indice

```
> var prima = new Date()
> db.borough.aggregate([
...   {
...     $group : {
...       _id : {
...         "name" : "$Name",
...         "number_of_christians" : "$Christian"
...       }
...     },
...     {
...       $sort : {
...         "_id.number_of_christians" : 1
...       }
...     },
...     {
...       $project : {
...         "_id" : 0,
...         "name" : "$_id.name",
...         "number_of_christians" : "$_id.number_of_christians"
...       }
...     }
...   ])
{ "name" : "City of London", "number_of_christians" : 3344 }
{ "name" : "Tower Hamlets", "number_of_christians" : 75087 }
{ "name" : "Kingston upon Thames", "number_of_christians" : 84684 }
{ "name" : "Camden", "number_of_christians" : 84737 }
{ "name" : "Kensington and Chelsea", "number_of_christians" : 86005 }
{ "name" : "Harrow", "number_of_christians" : 89181 }
{ "name" : "Islington", "number_of_christians" : 90107 }
{ "name" : "Hackney", "number_of_christians" : 95131 }
{ "name" : "Hammersmith and Fulham", "number_of_christians" : 98808 }
{ "name" : "Redbridge", "number_of_christians" : 102755 }
{ "name" : "Richmond upon Thames", "number_of_christians" : 103319 }
{ "name" : "Barking and Dagenham", "number_of_christians" : 104101 }
{ "name" : "Hounslow", "number_of_christians" : 106660 }
{ "name" : "Sutton", "number_of_christians" : 111086 }
{ "name" : "Merton", "number_of_christians" : 111993 }
{ "name" : "Haringey", "number_of_christians" : 114659 }
{ "name" : "Newham", "number_of_christians" : 123119 }
{ "name" : "Waltham Forest", "number_of_christians" : 124939 }
{ "name" : "Brent", "number_of_christians" : 129080 }
{ "name" : "Greenwich", "number_of_christians" : 134632 }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.008seconds
```

Con l'indice migliorano le prestazioni.

Q2 Senza indice

```
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.012seconds
> var prima = new Date()
> db.crimes.aggregate([
...   {
...     $group : {
...       "_id": "$Crime_Information.year",
...       "crime" : {
...         $sum : 1
...       }
...     }
...   },
...   {
...     $sort : {
...       "crime" : 1
...     }
...   },
...   {
...     $project : {
...       "_id" : 0,
...       "crime" : 1,
...       "year" : "$_id"
...     }
...   }
... ])

{ "crime" : 116047, "year" : 2016 }
{ "crime" : 116152, "year" : 2014 }
{ "crime" : 116153, "year" : 2015 }
{ "crime" : 116427, "year" : 2009 }
{ "crime" : 116690, "year" : 2011 }
{ "crime" : 116702, "year" : 2008 }
{ "crime" : 116739, "year" : 2012 }
{ "crime" : 116773, "year" : 2010 }
{ "crime" : 116892, "year" : 2013 }
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 1.232seconds
>
```

Q2 Con indice

```
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 372.489seconds
> var prima = new Date()
> db.crimes.aggregate([
...   {
...     $group : {
...       "_id": "$Crime_Information.year",
...       "crime" : {
...         $sum : 1
...       }
...     }
...   },
...   {
...     $sort : {
...       "crime" : 1
...     }
...   },
...   {
...     $project : {
...       "_id" : 0,
...       "crime" : 1,
...       "year" : "$_id"
...     }
...   }
... ])

{ "crime" : 116047, "year" : 2016 }
{ "crime" : 116152, "year" : 2014 }
{ "crime" : 116153, "year" : 2015 }
{ "crime" : 116427, "year" : 2009 }
{ "crime" : 116690, "year" : 2011 }
{ "crime" : 116702, "year" : 2008 }
{ "crime" : 116739, "year" : 2012 }
{ "crime" : 116773, "year" : 2010 }
{ "crime" : 116892, "year" : 2013 }
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.89seconds
>
```

L'indice ha velocizzato le prestazioni in questo caso.

Q3 Senza indice

```
> var prima = new Date()
> db.crimes.aggregate([
...   {
...     $match : {
...       "Crime_Information.month" : 12,
...       "Crime_Information.year" : 2015
...     }
...   },
...   {
...     $group : {
...       "_id" : "$Crime_Information.borough",
...       "furti" : {
...         $sum : 1
...       }
...     }
...   },
...   {
...     $sort : {
...       "furti": -1
...     }
...   },
...   {
...     $project : {
...       _id : 0,
...       borough : "$_id",
...       furti : "$furti"
...     }
...   }
... ])
{ "borough" : "Barnet", "furti" : 421 }
{ "borough" : "Croydon", "furti" : 415 }
{ "borough" : "Ealing", "furti" : 398 }
{ "borough" : "Wandsworth", "furti" : 398 }
{ "borough" : "Lambeth", "furti" : 376 }
{ "borough" : "Enfield", "furti" : 369 }
{ "borough" : "Brent", "furti" : 363 }
{ "borough" : "Lewisham", "furti" : 361 }
{ "borough" : "Bromley", "furti" : 361 }
{ "borough" : "Southwark", "furti" : 355 }
{ "borough" : "Newham", "furti" : 344 }
{ "borough" : "Hillingdon", "furti" : 332 }
{ "borough" : "Redbridge", "furti" : 328 }
{ "borough" : "Greenwich", "furti" : 327 }
{ "borough" : "Hackney", "furti" : 325 }
{ "borough" : "Waltham Forest", "furti" : 298 }
{ "borough" : "Tower Hamlets", "furti" : 279 }
{ "borough" : "Havering", "furti" : 279 }
{ "borough" : "Hounslow", "furti" : 279 }
{ "borough" : "Camden", "furti" : 276 }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.522seconds
> 
```

Q3 Con indice

```
> var prima = new Date()
> db.crimes.aggregate([
...   {
...     $match : {
...       "Crime_Information.month" : 12,
...       "Crime_Information.year" : 2015
...     }
...   },
...   {
...     $group : {
...       "_id" : "$Crime_Information.borough",
...       "furti" : {
...         $sum : 1
...       }
...     }
...   },
...   {
...     $sort : {
...       "furti": -1
...     }
...   },
...   {
...     $project : {
...       _id : 0,
...       borough : "$_id",
...       furti : "$furti"
...     }
...   }
... ])
{ "borough" : "Barnet", "furti" : 421 }
{ "borough" : "Croydon", "furti" : 415 }
{ "borough" : "Ealing", "furti" : 398 }
{ "borough" : "Wandsworth", "furti" : 398 }
{ "borough" : "Lambeth", "furti" : 376 }
{ "borough" : "Enfield", "furti" : 369 }
{ "borough" : "Brent", "furti" : 363 }
{ "borough" : "Lewisham", "furti" : 361 }
{ "borough" : "Bromley", "furti" : 361 }
{ "borough" : "Southwark", "furti" : 355 }
{ "borough" : "Newham", "furti" : 344 }
{ "borough" : "Hillingdon", "furti" : 332 }
{ "borough" : "Redbridge", "furti" : 328 }
{ "borough" : "Greenwich", "furti" : 327 }
{ "borough" : "Hackney", "furti" : 325 }
{ "borough" : "Waltham Forest", "furti" : 298 }
{ "borough" : "Tower Hamlets", "furti" : 279 }
{ "borough" : "Havering", "furti" : 279 }
{ "borough" : "Hounslow", "furti" : 279 }
{ "borough" : "Camden", "furti" : 276 }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.502seconds
>
```

Anche in questo caso l'indice migliora i tempi di esecuzione della query

Q4 Senza indice

```
{ "crime" : 8, "borough" : "City of London" }
{ "crime" : 279, "borough" : "Tower Hamlets" }
{ "crime" : 174, "borough" : "Kingston upon Thames" }
{ "crime" : 276, "borough" : "Camden" }
{ "crime" : 193, "borough" : "Kensington and Chelsea" }
{ "crime" : 236, "borough" : "Harrow" }
{ "crime" : 257, "borough" : "Islington" }
{ "crime" : 325, "borough" : "Hackney" }
{ "crime" : 217, "borough" : "Hammersmith and Fulham" }
{ "crime" : 328, "borough" : "Redbridge" }
{ "crime" : 219, "borough" : "Richmond upon Thames" }
{ "crime" : 218, "borough" : "Barking and Dagenham" }
{ "crime" : 279, "borough" : "Hounslow" }
{ "crime" : 219, "borough" : "Sutton" }
{ "crime" : 273, "borough" : "Merton" }
{ "crime" : 258, "borough" : "Haringey" }
{ "crime" : 344, "borough" : "Newham" }
{ "crime" : 298, "borough" : "Waltham Forest" }
{ "crime" : 363, "borough" : "Brent" }
{ "crime" : 327, "borough" : "Greenwich" }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.837seconds
>
```

Q4 Con indice

```
{ "crime" : 8, "borough" : "City of London" }
{ "crime" : 279, "borough" : "Tower Hamlets" }
{ "crime" : 174, "borough" : "Kingston upon Thames" }
{ "crime" : 276, "borough" : "Camden" }
{ "crime" : 193, "borough" : "Kensington and Chelsea" }
{ "crime" : 236, "borough" : "Harrow" }
{ "crime" : 257, "borough" : "Islington" }
{ "crime" : 325, "borough" : "Hackney" }
{ "crime" : 217, "borough" : "Hammersmith and Fulham" }
{ "crime" : 328, "borough" : "Redbridge" }
{ "crime" : 219, "borough" : "Richmond upon Thames" }
{ "crime" : 218, "borough" : "Barking and Dagenham" }
{ "crime" : 279, "borough" : "Hounslow" }
{ "crime" : 219, "borough" : "Sutton" }
{ "crime" : 273, "borough" : "Merton" }
{ "crime" : 258, "borough" : "Haringey" }
{ "crime" : 344, "borough" : "Newham" }
{ "crime" : 298, "borough" : "Waltham Forest" }
{ "crime" : 363, "borough" : "Brent" }
{ "crime" : 327, "borough" : "Greenwich" }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.813seconds
>
```

L'indice migliora i tempi

Q5 Senza indice

```
{ "crime" : 8, "borough" : "City of London" }
{ "crime" : 270, "borough" : "Bexley" }
{ "crime" : 344, "borough" : "Newham" }
{ "crime" : 218, "borough" : "Barking and Dagenham" }
{ "crime" : 219, "borough" : "Sutton" }
{ "crime" : 327, "borough" : "Greenwich" }
{ "crime" : 361, "borough" : "Lewisham" }
{ "crime" : 279, "borough" : "Hounslow" }
{ "crime" : 415, "borough" : "Croydon" }
{ "crime" : 174, "borough" : "Kingston upon Thames" }
{ "crime" : 273, "borough" : "Merton" }
{ "crime" : 361, "borough" : "Bromley" }
{ "crime" : 355, "borough" : "Southwark" }
{ "crime" : 398, "borough" : "Ealing" }
{ "crime" : 376, "borough" : "Lambeth" }
{ "crime" : 279, "borough" : "Havering" }
{ "crime" : 217, "borough" : "Hammersmith and Fulham" }
{ "crime" : 298, "borough" : "Waltham Forest" }
{ "crime" : 279, "borough" : "Tower Hamlets" }
{ "crime" : 219, "borough" : "Richmond upon Thames" }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.839seconds
> █
```

Q5 Con indice

```
{ "crime" : 8, "borough" : "City of London" }
{ "crime" : 270, "borough" : "Bexley" }
{ "crime" : 344, "borough" : "Newham" }
{ "crime" : 218, "borough" : "Barking and Dagenham" }
{ "crime" : 219, "borough" : "Sutton" }
{ "crime" : 327, "borough" : "Greenwich" }
{ "crime" : 361, "borough" : "Lewisham" }
{ "crime" : 279, "borough" : "Hounslow" }
{ "crime" : 415, "borough" : "Croydon" }
{ "crime" : 174, "borough" : "Kingston upon Thames" }
{ "crime" : 273, "borough" : "Merton" }
{ "crime" : 361, "borough" : "Bromley" }
{ "crime" : 355, "borough" : "Southwark" }
{ "crime" : 398, "borough" : "Ealing" }
{ "crime" : 376, "borough" : "Lambeth" }
{ "crime" : 279, "borough" : "Havering" }
{ "crime" : 217, "borough" : "Hammersmith and Fulham" }
{ "crime" : 298, "borough" : "Waltham Forest" }
{ "crime" : 279, "borough" : "Tower Hamlets" }
{ "crime" : 219, "borough" : "Richmond upon Thames" }
Type "it" for more
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.84seconds
> █
```

L'indice non migliora i tempi di esecuzione che rimangono pressochè uguali.

Q6 Senza indice

```
> var prima = new Date()
> db.crimes.aggregate([
...   {
...     $match : {
...       "Crime_Information.year" : 2015
...     }
...   },
...   {
...     $group : {
...       "_id" : {
...         "minor_category" : "$Crime_Information.minor_category",
...         "major_category" : "$Crime_Information.major_category"
...       },
...       "crime" : {
...         $sum : 1
...       }
...     }
...   },
...   {
...     $sort : {
...       "crime" : -1
...     }
...   },
...   {
...     $limit : 1
...   },
...   {
...     $project : {
...       "_id" : 0,
...       "minor_category" : "$_id.minor_category",
...       "major_category" : "$_id.major_category",
...       "crime" : 1
...     }
...   }
... ])
{ "crime" : 4676, "minor_category" : "Common Assault", "major_category" : "Violence Against the Person" }
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.599seconds
> 
```

Q6 Con indice

```
> var prima = new Date()
> db.crimes.aggregate([
...   {
...     $match : {
...       "Crime_Information.year" : 2015
...     }
...   },
...   {
...     $group : {
...       "_id" : {
...         "minor_category" : "$Crime_Information.minor_category",
...         "major_category" : "$Crime_Information.major_category"
...       },
...       "crime" : {
...         $sum : 1
...       }
...     }
...   },
...   {
...     $sort : {
...       "crime" : -1
...     }
...   },
...   {
...     $limit : 1
...   },
...   {
...     $project : {
...       "_id" : 0,
...       "minor_category" : "$_id.minor_category",
...       "major_category" : "$_id.major_category",
...       "crime" : 1
...     }
...   }
... ])
{ "crime" : 4676, "minor_category" : "Common Assault", "major_category" : "Violence Against the Person" }
> var dopo = new Date()
> var finale = dopo - prima
> print("took " + finale / 1000 + "seconds")
took 0.601seconds
> 
```

Anche in questo caso l'indice non migliora i tempi di esecuzione che rimangono pressoché uguali.