

# Práctica NoSQL

Mihai Dobrin

2023-10-26

## Tarea a realizar

A continuación se proponen 4 ejercicios a completar. Se recomienda completarlos en script markdown de R que se proporciona.

### Tarea 1

Exploremos otro diagrama de barras con una colección diferente - inspections. Esta recopilación contiene datos sobre las inspecciones de edificios de la ciudad de Nueva York y si pasan o no. Recupere dicha colección en R.

```
print("Completar Tarea 1")
```

```
## [1] "Completar Tarea 1"
```

```
## Setting de la conexión
```

```
conexion_mongomihai <- "mongodb+srv://Mihai:Mnl0Chm832jA4VVi@cluster0.fvn4d8x.mongodb.net/?retryWrites=
conexion_weakdb <- ssl_options(weak_cert_validation = T)
```

```
inspecciones <- mongo(collection = "inspections", db = "sample_training", url = conexion_mongomihai, op
```

### Tarea 2

Suponga que desea verificar el número de empresas que no aprobaron las inspecciones en 2015 en comparación con 2016.

Si ve los datos obtenidos de la colección, notará que el campo de fecha es una Cadena. Convertirlo en tipo de fecha y luego extraer el año requerirá algún procesamiento. Pero, con la canalización de agregación de MongoDB, puede hacer todo en una sola consulta. Para manipular el campo de fecha, use el operador \$addField.

Además, agregue las deficiencias encontradas en las inspecciones por año

```
print("Completar Tarea 2")
```

```
## [1] "Completar Tarea 2"
```

```

InspeccionesNoSuperadas2015y2016 <- inspecciones$aggregate('[{ "$addField":
{"AñoConvertido": {
  "$year": {
    "$dateFromString": {
      "dateString": "$date"}}}}},
{
"$match": {
  "$and": [{ "AñoConvertido": { "$in": [2015, 2016] }}, { "result": "Fail" }]
  }
},
{
"$group": {
  "_id": "$AñoConvertido", "Num de Fails": {"$sum":1}
}
}]')

colnames(InspeccionesNoSuperadas2015y2016) <- c("Año", "Cantidad_que_fracasaron")
InspeccionesNoSuperadas2015y2016

```

```

##      Año Cantidad_que_fracasaron
## 1 2015                      1042
## 2 2016                       58

```

### Tarea 3

Teniendo en cuenta que el resultado de la tarea anterior está agrupando los resultados por año, cree un gráfico de barras.

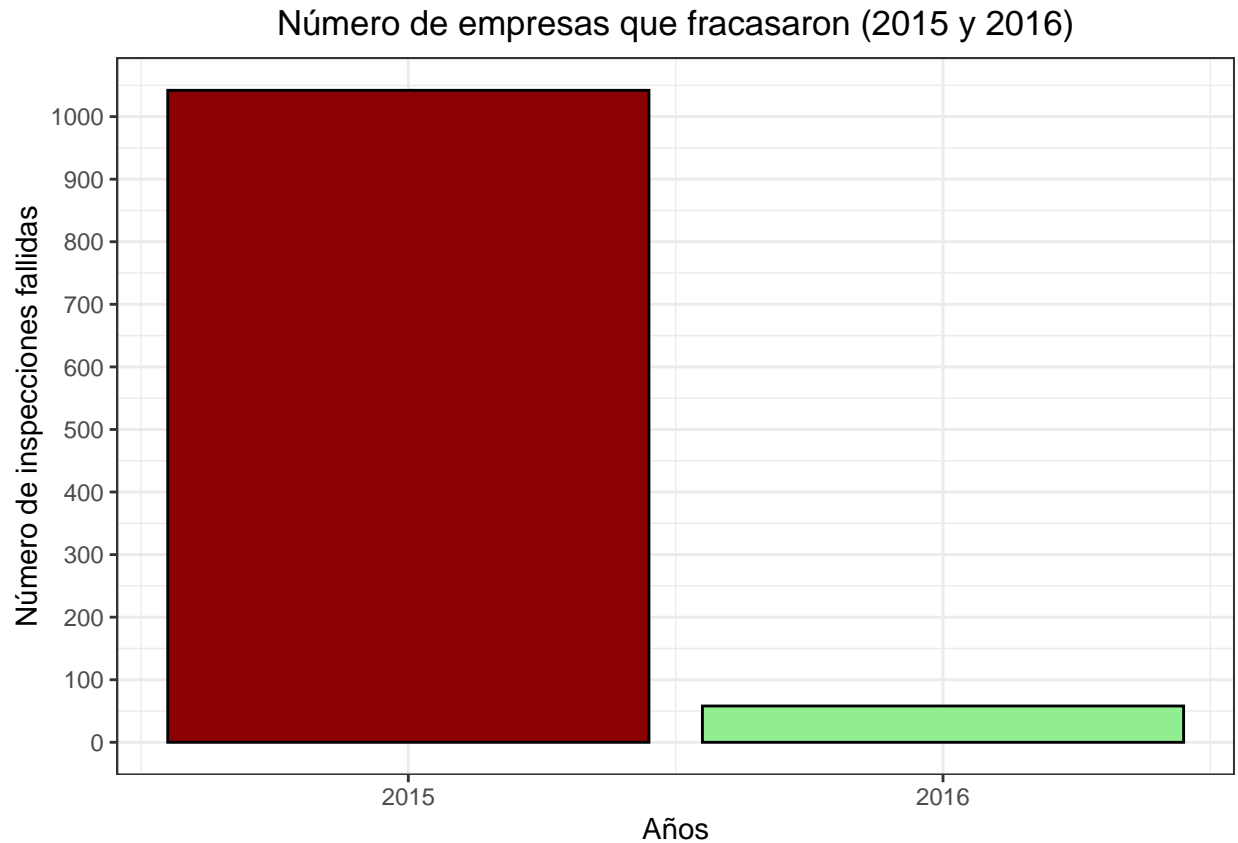
```
print("Completar Tarea 3")
```

```
## [1] "Completar Tarea 3"
```

```

InspeccionesNoSuperadas2015y2016 %>%
  ggplot(aes(Año, Cantidad_que_fracasaron, fill = factor(Año))) +
  geom_bar(stat = "identity", color = "black", fill = c("red4","lightgreen")) +
  theme_bw() +
  scale_y_continuous(breaks = seq(0, 1200, by = 100)) +
  scale_x_continuous(breaks = c(2015,2016)) +
  labs(title = "Número de empresas que fracasaron (2015 y 2016)",
       x = "Años",
       y = "Número de inspecciones fallidas") +
  theme(plot.title = element_text(hjust = 0.5))

```



## Tarea 4

A continuación, se utilizará la colección ‘companies’, que contiene información sobre empresas, como su año de fundación y la dirección de su sede.

Supongamos que desea conocer la tendencia del número de empresas de publicidad (category\_code = ‘advertising’) fundadas a partir de 2000 en adelante. Para ello, utilice el operador relacional \$gt, agrupe los resultados por año de creación (‘founded\_year’) y ordénelos para que se muestren posteriormente en un gráfico de líneas por año.

```
print("Completar Tarea 4")
```

```
## [1] "Completar Tarea 4"
```

```
companies <- mongo(collection = "companies", db = "sample_training", url = conexion_mongomihai, options = list())

EmpresasFundadasPost2000 <- companies$aggregate(['$match': {'category_code': "advertising", 'founded_year': {'$gt' : 1999 }},
{'$group': {'_id':"$founded_year", "Num de Empresas de Publicidad": {'$sum':1}}},
{'$sort': {'_id': 1}}
])
```

```
colnames(EmpresasFundadasPost2000) <- c("Año", "Cantidad_Emp_Fundadas")

tibble(EmpresasFundadasPost2000)
```

```
## # A tibble: 12 x 2
##       Año Cantidad_Emp_Fundadas
##   <int>          <int>
## 1  2000             16
## 2  2001             13
## 3  2002             15
## 4  2003             17
## 5  2004             24
## 6  2005             32
## 7  2006             49
## 8  2007             80
## 9  2008             77
## 10 2009             27
## 11 2012              1
## 12 2013              1
```

```
EmpresasFundadasPost2000 %>%
  ggplot(aes(Año, Cantidad_Emp_Fundadas)) +
  geom_line(color = "darkred", linetype = "longdash") +
  geom_point() +
  # scale_x_continuous(breaks = unique(EmpresasFundadasPost2000$Año)) +
  scale_x_continuous(breaks = seq(2000,2014, by = 1)) +
  labs(title = "Número de empresas que se fundaron apartir de 2000",
       x = "Años",
       y = "Número de empresas fundadas") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_bw()
```

Número de empresas que se fundaron apartir de 2000

