

Taller 4

Consultas básicas

1) Hallar los tweets del usuario con user id '818839458'.

```
collection.find(  
  {  
    "_id": "818839458"  
  }  
)
```

2) Hallar aquellos tweets que tengan más de 500000 retweets

```
collection.find(  
  {  
    "retweet_count": { $gt: 500000 }  
  }  
)
```

3) Mostrar la cantidad de retweets de los tweets que se hayan hecho desde Argentina o Brasil.

```
collection.find(  
  {  
    "place.country": { $in: [ "Argentina", "Brasil" ] }  
  },  
  {  
    "retweet_count": 1  
  }  
)
```

4) Hallar los usuarios que tengan tweets con 200000 o más retweets y sean en idioma español.

```
collection.find(  
  {  
    "lang": "es",  
    "retweet_count": { $gte: 200000 }  
  },  
  {  
    "user": 1  
  }  
)
```

5) Mostrar la cantidad de retweets para los tweets que no se hayan hecho en Argentina ni Brasil, pero si tengan un lugar definido y sean en español.

```
collection.find(  
  {  
    $and: [{"place.country": {$nin: ["Argentina", "Brasil", null]}}, {"lang": "es"}]  
  },  
  {  
    "retweet_count": 1  
  }  
)
```

6) Mostrar los screen name de aquellos usuarios que tengan "Juan" como parte de su nombre.

```
collection.find(  
  {  
    "user.name": /Juan/  
  },  
  {  
    "user.screen_name": 1  
  }  
)
```

7) Mostrar de los 10 tweets con más retweets, su usuario y la cantidad de retweets.

```
collection.find({},  
  {  
    "user": 1,  
    "retweet_count": 1  
  }  
)  
.sort({"retweet_count": -1}).limit(10)
```

Consultas de agregación

1) Mostrar de los 10 tweets con más retweets, su usuario y la cantidad de retweets. Ordenar la salida de forma ascendente.

```
collection.aggregate([
  {
    $project: {retweet_count: 1, user: 1}
  },
  {
    $sort: {retweet_count: -1}
  },
  {
    $limit: 10
  },
  {
    $sort: {retweet_count: 1}
  }
])
```

2) Encontrar los 10 hashtags más usados.

```
collection.aggregate([
  {
    $unwind: "$entities.hashtags"
  },
  {
    $group: {_id: "$entities.hashtags.text", usos: {$count: {}}}
  },
  {
    $sort: {usos: -1}
  },
  {
    $limit: 10
  }
])
```

3) Encontrar a los 5 usuarios más mencionados.

```
collection.aggregate([
  {
    $unwind: "$entities.user_mentions"
  },
  {
    $group: {
      _id: "$entities.user_mentions.id_str",
      count: {$sum: 1}
    }
  },
  {
    $sort: {count: -1}
  },
  {
    $limit: 5
  }
])
```

4) Hallar la cantidad de retweets promedio para los tweets que se hayan hecho desde Argentina y aquellos que no.

```
collection.aggregate([
  {
    $group: {
      _id: {isFromArgentina: {$cond: [{
        $eq: ["$user.location", "Argentina"]}, true, false
      ]}},
      avgRetweets: {$avg: "$retweet_count"}
    }
  }
])
```

5) Por cada usuario obtener una lista de ids de tweets y el largo de la misma.

```
collection.aggregate([
  {
    $group: {
      _id: "$user",
      tweets: {$push: "$_id"},
      cantidad: {$count: {}}
    }
  }
])
```

6) Hallar la máxima cantidad de retweets totales que tuvo algún usuario.

```
collection.aggregate([
  {
    $group: {
      _id: "$user.id_str",
      totalRetweets: {$sum: "$retweet_count"}
    }
  },
  {
    $sort: {totalRetweets: -1}
  },
  {
    $limit: 1
  },
  {
    $project: {
      _id: 0,
      userId: "$_id",
      maxTotalRetweets: "$totalRetweets"
    }
  }
])
```

7) Hallar para cada intervalo de una hora cuantos tweets realizo cada usuario.

```
collection.aggregate([
  {
    $project: {"user": 1, "hora": {"$hour": "$created_at"}}
  },
  {
    $group: {
      "_id": {"usuario": "$user.name", "hora": "$hora"},
      "cantidad": {"$count": {}}
    }
  },
  {
    $group: {
      "_id": "$_id.user_name",
      "horas_tweets": {"$push": {
        "hora": "$_id.hora",
        "cantidad": "$cantidad"
      }}
    }
  }
])
```