

Proyecto.

Análisis de vuelos de 2015 de EU utilizando el sistema de base de datos MONGO DB

Para este proyecto se obtuvo una muestra de tamaño 500,000 de los datos de retrasos y cancelaciones de vuelos de 2015 y datos de análisis de sentimientos en Twitter de las aerolíneas estadounidenses recopilados del 17 al 24 de febrero 2015 provenientes de la plataforma Kaggle a través de la página <https://www.kaggle.com/>

Planteamiento del problema

¿En qué aerolíneas se debe volar para evitar retrasos significativos?

Objetivo:

- Conocer y ubicar la situación de las aerolíneas estadounidenses.
- Comparar dos fuentes de información de vuelos de 2015 de EU, por un lado, datos de retrasos y cancelación de vuelos de 2015; por otro lado, datos de Twitter sobre análisis de sentimientos de viajeros de EU durante febrero de 2015.
- Establecer indicadores que determinen el desempeño en los vuelos de retraso de las aerolíneas de EU.

Estructura de la base en Mongo DB

Collection Name ^	Documents	Avg. Document Size	Total Document Size	Num. Indexes	Total Index Size
Twitter_aerolineas	14,640	538.0 B	7.5 MB	1	156.0 KB
airlines	14	74.4 B	1.0 KB	1	20.0 KB
airports	322	175.8 B	55.3 KB	1	20.0 KB
catalogo_dia_semana	7	53.6 B	375.0 B	1	20.0 KB
sample_flights	500,000	235.5 B	112.3 MB	1	4.6 MB

La base de datos está compuesta por dos archivos:

- Archivo de retrasos y cancelaciones de vuelos de 2015, contiene 4 colecciones de datos:
 1. **sample_flights**. Es una colección de datos de tamaño 500,000 y contiene información de los vuelos de las aerolíneas.
 2. **airlines**. Es un catálogo que contiene información sobre las aerolíneas de EU.
 3. **airports**. Es un catálogo que contiene información sobre los aeropuertos de EU.
 4. **catalogo_dia_semana**. Es un catálogo que contiene la codificación del día de semana.
- Archivo de Twitter sobre análisis de sentimientos de viajeros de EU, solo contiene una colección de datos.
 1. **Twitter_aerolineas**. Colección de datos de análisis de sentimientos de viajeros de EU durante el 17 al 24 de febrero de 2015.

Glosario:

scheduled_departure: salida programada, en hora minuto

departure_delay: retardo de salida, en minutos

arrival_delay: retardo de llegada, en minutos

1. ¿Obtener el número de vuelos con retraso en su llegada por compañía?

```
[{$project: {
  AIRLINE: 1,
  retraso_salida: {
    $cond: {
      'if': {
        $gt: [
          '$DEPARTURE_DELAY',
          0
        ]
      },
      then: 1,
      'else': 0
    }
  },
  retraso_llegada: {
    $cond: {
      'if': {
        $gt: [
          '$ARRIVAL_DELAY',
          0
        ]
      },
      then: 1,
      'else': 0
    }
  }
}}, {$group: {
  _id: '$AIRLINE',
  Total_retraso_Salida: {
    $sum: '$retraso_salida'
  },
  Total_retraso_LLlegada: {
    $sum: '$retraso_llegada'
  }
}}, {$lookup: {
  from: 'airlines',
  localField: '_id',
  foreignField: 'IATA_CODE',
  as: 'airlines'
}}, {$addFields: {
  airlines_obj: {
    $arrayElemAt: [
      '$airlines',
      0
    ]
  }
}}, {$addFields: {
```

```

airline: '$airlines_obj.AIRLINE'
}}, {$project: {
  _id: 0,
  airline: 1,
  Total_retraso_Llegada: 1
}}, {$sort: {
  Total_retraso_Llegada: -1
}}

```

Output after **\$project** stage (Sample of 20 documents)

```

1 * {
2   AIRLINE: 1,
3   retraso_salida: {
4     $cond: {
5       'if': {
6         $gt: [
7           "$DEPARTURE_DELAY",
8           0
9         ]
10      },
11      then: 1,
12      'else': 0
13    },
14    retraso_llegada: {
15      $cond: {
16        'if': {
17          $gt: [
18            "$ARRIVAL_DELAY",
19            0
20          ]
21        },
22        then: 1,
23        'else': 0
24      }
25    }
26  }

```

_id: ObjectId("5f26def4bd7cf51340488e00")

AIRLINE: "US"

retraso_salida: 0

retraso_llegada: 0

_id: ObjectId("5f26def4bd7cf51340488e01")

AIRLINE: "EV"

retraso_salida: 0

retraso_llegada: 0

Output after **\$group** stage (Sample of 14 documents)

```

1 * {
2   _id: '$AIRLINE',
3   Total_retraso_Salida: {
4     $sum: '$retraso_salida'
5   },
6   Total_retraso_Llegada: {
7     $sum: '$retraso_llegada'
8   }
9 }

```

_id: "EV"

Total_retraso_Salida: 2919

Total_retraso_Llegada: 3650

_id: "UA"

Total_retraso_Salida: 4399

Total_retraso_Llegada: 3161

Output after **\$lookup** stage (Sample of 14 documents)

```

1 * {
2   from: 'airlines',
3   localField: '_id',
4   foreignField: 'IATA_CODE',
5   as: 'airlines'
6 }

```

_id: "AS"

Total_retraso_Salida: 757

Total_retraso_Llegada: 952

airlines: Array

_id: "MQ"

Total_retraso_Salida: 1543

Total_retraso_Llegada: 1785

airlines: Array

▼

\$addFields

🔴

🗑️

➕

```
1 = {  
2   airlines_obj: {  
3     $arrayElemAt: [  
4       '$airlines',  
5       0  
6     ]  
7   }  
8 }
```

Output after [\\$addFields](#) stage ⓘ (Sample of 14 documents)

```
_id: "F9"  
Total_retrato_Salida: 597  
Total_retrato_Llegada: 711  
airlines: Array  
airlines_obj: Object
```

```
_id: "NK"  
Total_retrato_Salida: 873  
Total_retrato_Llegada: 931  
airlines: Array  
airlines_obj: Object
```

▼

\$addFields

🔴

🗑️

➕

```
1 = {  
2   airline: '$airlines_obj.AIRLINE'  
3 }
```

Output after [\\$addFields](#) stage ⓘ (Sample of 14 documents)

```
_id: "VX"  
Total_retrato_Salida: 426  
Total_retrato_Llegada: 448  
airlines: Array  
airlines_obj: Object  
airline: "Virgin America"
```

```
_id: "NK"  
Total_retrato_Salida: 873  
Total_retrato_Llegada: 931  
airlines: Array  
airlines_obj: Object  
airline: "Spirit Air Lines"
```

▼

\$project

🔴

🗑️

➕

```
1 = {  
2   _id: 0,  
3   airline: 1,  
4   Total_retrato_Llegada: 1  
5 }
```

Output after [\\$project](#) stage ⓘ (Sample of 14 documents)

```
Total_retrato_Llegada: 1693  
airline: "JetBlue Airways"
```

```
Total_retrato_Llegada: 531  
airline: "Hawaiian Airlines Inc."
```

▼

\$sort

🔴

🗑️

➕

```
1 = {  
2   Total_retrato_Llegada: -1  
3 }
```

Output after [\\$sort](#) stage ⓘ (Sample of 14 documents)

```
Total_retrato_Llegada: 8014  
airline: "Southwest Airlines Co."
```

```
Total_retrato_Llegada: 4419  
airline: "American Airlines Inc."
```

2. Obtener las causas de cancelación: “Aerolinea” y “Condiciones metereologicas” por tipo de aerolínea

```
[{$match: {
  CANCELLATION_REASON: {$in:
    ['Aerolinea',
     'Condiciones metereologicas'
    ]
  }
}}, {$group: {
  _id: {
    aerolinea: '$AIRLINE',
    razon_cancelacion: '$CANCELLATION_REASON'
  },
  total: {
    $sum: 1
  }
}}, {$addFields: {
  aerolinea: '$_id.aerolinea',
  razon_cancelacion: '$_id.razon_cancelacion'
```

```

}}, {$lookup: {
  from: 'airlines',
  localField: 'aerolinea',
  foreignField: 'IATA_CODE',
  as: 'aerolinea'
}}, {$addFields: {
  usr_obj: {
    $arrayElemAt: [
      '$aerolinea',
      0
    ]
  }
}}, {$addFields: {
  aerolinea: '$usr_obj.AIRLINE'
}}, {$project: {
  _id: 0,
  aerolinea: 1,
  razon_cancelacion: 1,
  total: 1
}}, {$sort: {
  aerolinea: 1,
  total: -1
}}

```

\$match

+

Output after [\\$match](#) stage ⓘ (Sample of 20 documents)

```

1 {
2   CANCELLATION_REASON: {$in:
3     ['Aerolinea',
4     'Condiciones metereologicas'
5   ]}
6 }

```

DAY: 9
DAY_OF_WEEK: 1
AIRLINE: "DL"
ARRIVAL_DELAY: NaN
DEPARTURE_DELAY: NaN
SCHEDULED_DEPARTURE: "9:00"
ORIGIN_AIRPORT: "BOS"
DESTINATION_AIRPORT: "LGA"
CANCELLATION_REASON: "Condiciones metereologicas"

DAY: 9
DAY_OF_WEEK: 7
AIRLINE: "EV"
ARRIVAL_DELAY: NaN
DEPARTURE_DELAY: NaN
SCHEDULED_DEPARTURE: "14:33"
ORIGIN_AIRPORT: "BOS"
DESTINATION_AIRPORT: "RIC"
CANCELLATION_REASON: "Aerolinea"

\$group

+

Output after [\\$group](#) stage ⓘ (Sample of 20 documents)

```

1 {
2   _id: {
3     aerolinea: '$AIRLINE',
4     razon_cancelacion: '$CANCELLATION_REASON'
5   },
6   total: {
7     $sum: 1
8   }
9 }

```

_id: Object
total: 240

_id: Object
total: 2

<div> <div> <div>\$addFields</div> <div></div> </div> <div> <div></div> <div></div> </div> </div> <pre> 1 = { 2 aerolinea: '\$_id.aerolinea', 3 razon_cancelacion: '\$_id.razon_cancelacion' 4 } </pre>	<div>Output after \$addFields stage ⓘ (Sample of 20 documents)</div> <div> <div> <div> <div>▶ _id: Object</div> <div>total: 813</div> <div>aerolinea: "MQ"</div> <div>razon_cancelacion: "Condiciones metereologicas"</div> </div> </div> <div> <div> <div>▶ _id: Object</div> <div>total: 55</div> <div>aerolinea: "NK"</div> <div>razon_cancelacion: "Aerolinea"</div> </div> </div> </div>
<div> <div> <div>\$lookup</div> <div></div> </div> <div> <div></div> <div></div> </div> </div> <pre> 1 = { 2 from: 'airlines', 3 localField: 'aerolinea', 4 foreignField: 'IATA_CODE', 5 as: 'aerolinea' 6 } </pre>	<div>Output after \$lookup stage ⓘ (Sample of 20 documents)</div> <div> <div> <div> <div>▶ _id: Object</div> <div>total: 205</div> <div>aerolinea: Array</div> <div>razon_cancelacion: "Aerolinea"</div> </div> </div> <div> <div> <div>▶ _id: Object</div> <div>total: 259</div> <div>aerolinea: Array</div> <div>razon_cancelacion: "Condiciones metereologicas"</div> </div> </div> </div>
<div> <div> <div>\$addFields</div> <div></div> </div> <div> <div></div> <div></div> </div> </div> <pre> 1 = { 2 usr_obj: { 3 \$arrayElemAt: [4 '\$aerolinea', 5 0 6] 7 } 8 } </pre>	<div>Output after \$addFields stage ⓘ (Sample of 20 documents)</div> <div> <div> <div> <div>▶ _id: Object</div> <div>total: 12</div> <div>aerolinea: Array</div> <div>razon_cancelacion: "Aerolinea"</div> <div>usr_obj: Object</div> </div> </div> <div> <div> <div>▶ _id: Object</div> <div>total: 85</div> <div>aerolinea: Array</div> <div>razon_cancelacion: "Condiciones metereologicas"</div> <div>usr_obj: Object</div> </div> </div> </div>
<div> <div> <div>\$addFields</div> <div></div> </div> <div> <div></div> <div></div> </div> </div> <pre> 1 = { 2 aerolinea: '\$usr_obj.AIRLINE' 3 } </pre>	<div>Output after \$addFields stage ⓘ (Sample of 20 documents)</div> <div> <div> <div> <div>▶ _id: Object</div> <div>total: 240</div> <div>aerolinea: "American Airlines Inc."</div> <div>razon_cancelacion: "Aerolinea"</div> <div>usr_obj: Object</div> </div> </div> <div> <div> <div>▶ _id: Object</div> <div>total: 2</div> <div>aerolinea: "Virgin America"</div> <div>razon_cancelacion: "Condiciones metereologicas"</div> <div>usr_obj: Object</div> </div> </div> </div>
<div> <div> <div>\$project</div> <div></div> </div> <div> <div></div> <div></div> </div> </div> <pre> 1 = { 2 _id: 0, 3 aerolinea: 1, 4 razon_cancelacion: 1, 5 total: 1 6 } </pre>	<div>Output after \$project stage ⓘ (Sample of 20 documents)</div> <div> <div> <div>total: 12</div> <div>aerolinea: "Hawaiian Airlines Inc."</div> <div>razon_cancelacion: "Aerolinea"</div> </div> </div> <div> <div>total: 85</div> <div>aerolinea: "Spirit Air Lines"</div> <div>razon_cancelacion: "Condiciones metereologicas"</div> </div>

| \$sort ``` 1 = { 2 aerolinea: 1, 3 total: -1 4 } ``` | Output after [\\$sort](#) stage ⓘ (Sample of 20 documents) total: 29 aerolinea: "Alaska Airlines Inc." razon_cancelacion: "Aerolinea" total: 18 aerolinea: "Alaska Airlines Inc." razon_cancelacion: "Condiciones metereologicas" |

3. Obtener la cantidad de vuelos por los días de la semana

```

[{$group: {
  _id: '$DAY_OF_WEEK',
  total: {

```

```

    $sum: 1
  }
}, {$lookup: {
  from: 'catalogo_dia_semana',
  localField: '_id',
  foreignField: 'id',
  as: 'ds'
}}, {$addFields: {
  ds_obj: {
    $arrayElemAt: [
      '$ds',
      0
    ]
  }
}}, {$addFields: {
  dia_semana: '$ds_obj.dia_semana'
}}, {$project: {
  _id: 0,
  dia_semana: 1,
  total: 1
}}, {$sort: {
  total: -1
}}]

```

\$group

Output after [\\$group](#) stage (Sample of 7 documents)

_id: 3

total: 14618

_id: 2

total: 14499

\$lookup

Output after [\\$lookup](#) stage (Sample of 7 documents)

_id: 6

total: 11876

ds: Array

_id: 2

total: 14499

ds: Array

SaddFields

Output after [SaddFields](#) stage ⓘ (Sample of 7 documents)

```
1 {
2   ds_obj: {
3     $arrayElemAt: [
4       '$ds',
5       0
6     ]
7   }
8 }
```

_id: 6

total: 11876

ds: Array

ds_obj: Object

_id: 7

total: 14241

ds: Array

ds_obj: Object

SaddFields

Output after [SaddFields](#) stage ⓘ (Sample of 7 documents)

```
1 {
2   dia_semana: '$ds_obj.dia_semana'
3 }
```

_id: 7

total: 14241

ds: Array

ds_obj: Object

dia_semana: "domingo"

_id: 1

total: 14823

ds: Array

ds_obj: Object

dia_semana: "lunes"

Sproject

Output after [Sproject](#) stage ⓘ (Sample of 7 documents)

```
1 {
2   id: 0,
3   dia_semana: 1,
4   total: 1
5 }
```

total: 11876

dia_semana: "sabado"

total: 14769

dia_semana: "viernes"

Ssort

Output after [Ssort](#) stage ⓘ (Sample of 7 documents)

```
1 {
2   total: -1
3 }
```

total: 15174

dia_semana: "jueves"

total: 14823

dia_semana: "lunes"

4.Cuál es el aeropuerto con más retrasos en sus vuelos (retraso en salidas y retraso en llegadas)

```
[{$project: {
  ARRIVAL_DELAY:1,
  DEPARTURE_DELAY:1,
  ORIGIN_AIRPORT:1,
  retraso_salida:{
    $cond:{if:{$gt:["$DEPARTURE_DELAY", 0]},
    then: 1,
    else: 0}},
  retraso_llegada:{
    $cond:{if:{$gt:["$ARRIVAL_DELAY", 0]},
    then: 1,
    else: 0},
  }
}}, {$group: {
  _id: "$ORIGIN_AIRPORT",
  total_retardo_salida:{$sum:"$retraso_salida"},
  total_retardo_llegada:{$sum:"$retraso_llegada"}
}}, {$addFields: {
```

```

total: {$sum:["$total_retardo_salida",
"$total_retardo_llegada"]}
}
}}, {$lookup: {
  from: 'airports',
  localField: '_id',
  foreignField: 'IATA_CODE',
  as: 'aeropuertos'
}}, {$addFields: {
  usr_obj: {
    $arrayElemAt: [
      '$aeropuertos',
      0
    ]
  }
}}, {$addFields: {
  aeropuertos:"$usr_obj.AIRPORT"
}}, {$project: {
  _id:0,
  aeropuertos:1,
  total:1,
  total_retardo_salida:1,
  total_retardo_llegada:1
}}, {$sort: {
  total: -1
}}}

```

Output after **\$project** stage (Sample of 20 documents)

```

1 {
2   ARRIVAL_DELAY: 1,
3   DEPARTURE_DELAY: 1,
4   ORIGIN_AIRPORT: 1,
5   retraso_salida: {
6     $cond: {
7       'if': {
8         $gt: [
9           '$DEPARTURE_DELAY',
10          0
11        ]
12      },
13      then: 1,
14      'else': 0
15    }
16  },
17  retraso_llegada: {
18    $cond: {
19      'if': {
20        $gt: [
21          '$ARRIVAL_DELAY',
22          0
23        ]
24      },
25      then: 1,
26      'else': 0
27    }
28  }
29 }

```

Document 1:

```

_id: ObjectId("5f25de4f8aecc42c38471966")
ARRIVAL_DELAY: -11
DEPARTURE_DELAY: -8
ORIGIN_AIRPORT: "IAH"
retraso_salida: 0
retraso_llegada: 0

```

Document 2:

```

_id: ObjectId("5f25de4f8aecc42c38471967")
ARRIVAL_DELAY: 0
DEPARTURE_DELAY: -5
ORIGIN_AIRPORT: "IAH"
retraso_salida: 0
retraso_llegada: 0

```

\$group

1 * {
2 _id: '\$ORIGIN_AIRPORT',
3 total_retardo_salida: {
4 \$sum: '\$retraso_salida'
5 },
6 total_retardo_llegada: {
7 \$sum: '\$retraso_llegada'
8 }
9 }

Output after \$group stage (Sample of 20 documents)

_id: "11695"
total_retardo_salida: 2
total_retardo_llegada: 3

_id: "15H"
total_retardo_salida: 6
total_retardo_llegada: 10

\$addFields

1 * {
2 total: {
3 \$sum: [
4 '\$total_retardo_salida',
5 '\$total_retardo_llegada'
6]
7 }
8 }

Output after \$addFields stage (Sample of 20 documents)

_id: "14683"
total_retardo_salida: 10
total_retardo_llegada: 4
total: 14

_id: "14738"
total_retardo_salida: 5
total_retardo_llegada: 2
total: 7

\$lookup

1 * {
2 from: 'airports',
3 localField: '_id',
4 foreignField: 'IATA_CODE',
5 as: 'aeropuertos'
6 }

Output after \$lookup stage (Sample of 20 documents)

_id: "BDL"
total_retardo_salida: 96
total_retardo_llegada: 92
total: 188
aeropuertos: Array

_id: "ERI"
total_retardo_salida: 4
total_retardo_llegada: 8
total: 12
aeropuertos: Array

\$addFields

1 * {
2 usr_obj: {
3 \$arrayElemAt: [
4 '\$aeropuertos',
5 0
6]
7 }
8 }

Output after \$addFields stage (Sample of 20 documents)

_id: "DAB"
total_retardo_salida: 5
total_retardo_llegada: 7
total: 12
aeropuertos: Array
usr_obj: Object

_id: "SAV"
total_retardo_salida: 44
total_retardo_llegada: 43
total: 87
aeropuertos: Array
usr_obj: Object

\$addFields

1 * {
2 aeropuertos: '\$usr_obj.AIRPORT'
3 }

Output after \$addFields stage (Sample of 20 documents)

_id: "MSN"
total_retardo_salida: 39
total_retardo_llegada: 51
total: 90
aeropuertos: "Dane County Regional Airport"
usr_obj: Object

_id: "12173"
total_retardo_salida: 21
total_retardo_llegada: 21
total: 42

\$project

1 * {
2 _id: 0,
3 aeropuertos: 1,
4 total: 1,
5 total_retardo_salida: 1,
6 total_retardo_llegada: 1
7 }

Output after \$project stage (Sample of 20 documents)

total_retardo_salida: 96
total_retardo_llegada: 92
total: 188
aeropuertos: "Bradley International Airport"

total_retardo_salida: 4
total_retardo_llegada: 8
total: 12
aeropuertos: "Erie International Airport"

The screenshot shows the MongoDB Compass interface. On the left, the 'Sort' stage is configured with the field 'total' and the sort order '-1' (descending). The output of this stage is displayed on the right, showing a sample of 20 documents. The output is divided into two columns. The left column shows documents with 'total_retardo_salida: 2176', 'total_retardo_llegada: 1900', and 'total: 4076'. The right column shows documents with 'total_retardo_salida: 2113', 'total_retardo_llegada: 1933', and 'total: 4046'. Both columns show 'aeropuertos: "Hartsfield-Jackson Atlanta International"' and 'aeropuertos: "Chicago O'Hare International Airport"' respectively.

5. Cuál es el estado con más aeropuertos

```

[{$group: {
  _id: '$ORIGIN_AIRPORT',
  total: {
    $sum: 1
  }
}}, {$lookup: {
  from: 'airports',
  localField: '_id',
  foreignField: 'IATA_CODE',
  as: 'airports'
}}, {$addFields: {
  airports_obj: {
    $arrayElemAt: [
      '$airports',
      0
    ]
  }
}}, {$addFields: {
  STATE: '$airports_obj.STATE',
  AIRPORT: '$airports_obj.AIRPORT'
}}, {$group: {
  _id: '$STATE',
  total: {
    $sum: 1
  }
}}, {$match: {
  _id: {$ne: null}
}}, {$sort: {
  total: -1
}}]

```

\$group

1 * {
2 _id: '\$ORIGIN_AIRPORT',
3 total: {
4 \$sum: 1
5 }
6 }

Output after \$group stage ⓘ (Sample of 20 documents)

_id: "SDF"
total: 178

_id: "BOS"
total: 1887

\$lookup

1 * {
2 from: 'airports',
3 localField: '_id',
4 foreignField: 'IATA_CODE',
5 as: 'airports'
6 }

Output after \$lookup stage ⓘ (Sample of 20 documents)

_id: "14683"
total: 36
▶ airports: Array

_id: "14738"
total: 12
▶ airports: Array

\$addFields

1 * {
2 airports_obj: {
3 \$arrayElemAt: [
4 '\$airports',
5 0
6]
7 }
8 }

Output after \$addFields stage ⓘ (Sample of 20 documents)

_id: "11695"
total: 5
▶ airports: Array

_id: "ISN"
total: 37
▶ airports: Array
▶ airports_obj: Object

\$addFields

1 * {
2 STATE: '\$airports_obj.STATE',
3 AIRPORT: '\$airports_obj.AIRPORT'
4 }

Output after \$addFields stage ⓘ (Sample of 20 documents)

_id: "GEG"
total: 185
▶ airports: Array
▶ airports_obj: Object
STATE: "WA"
AIRPORT: "Spokane International Airport"

_id: "SJU"
total: 437
▶ airports: Array
▶ airports_obj: Object
STATE: "PR"
AIRPORT: "Luis Muñoz Marín International Airport"

\$group

1 * {
2 _id: '\$STATE',
3 total: {
4 \$sum: 1
5 }
6 }

Output after \$group stage ⓘ (Sample of 20 documents)

_id: null
total: 275

_id: "TX"
total: 24

\$match

1 * {
2 _id: {
3 \$ne: null
4 }
5 }

Output after \$match stage ⓘ (Sample of 20 documents)

_id: "MT"
total: 8

_id: "NM"
total: 4



6. Obtener el promedio de retraso de un avión por hora de salida programada

```
[{$project: {
  _id: 0,
  SCHEDULED_DEPARTURE: 1,
  DEPARTURE_DELAY: 1,
  retraso_salida: {
    $cond: {
      'if': {
        $gt: [
          '$DEPARTURE_DELAY',
          0
        ]
      },
      then: 1,
      'else': 0
    }
  }
}}, {$match: {
  retraso_salida: 1
}}, {$group: {
  _id: '$SCHEDULED_DEPARTURE',
  promedio_retraso: {
    $avg: '$DEPARTURE_DELAY'
  },
  n: {
    $sum: 1
  }
}}, {$project: {
  hora_salida_proramada: '$_id',
  promedio_retraso: {$round: [
    '$promedio_retraso', 2]
  },
  n: 1,
  _id: 0
}
}, {$match: {
  n: {
    $gte: 10
  }
}
```

```
}}, {$sort: {
  promedio_retraso: -1
}}]
```

Project

Output after [Project](#) stage (Sample of 20 documents)

```
1 {
2   _id: 0,
3   SCHEDULED_DEPARTURE: 1,
4   DEPARTURE_DELAY: 1,
5   retraso_salida: {
6     $cond: {
7       'if': {
8         $gt: [
9           '$DEPARTURE_DELAY',
10          0
11        ]
12      },
13      then: 1,
14      'else': 0
15    }
16  }
17 }
```

DEPARTURE_DELAY: -8
SCHEDULED_DEPARTURE: "18:05"
retraso_salida: 0

DEPARTURE_DELAY: -5
SCHEDULED_DEPARTURE: "9:10"
retraso_salida: 0

Match

Output after [Match](#) stage (Sample of 20 documents)

```
1 {
2   retraso_salida: 1
3 }
```

DEPARTURE_DELAY: 6
SCHEDULED_DEPARTURE: "11:55"
retraso_salida: 1

DEPARTURE_DELAY: 26
SCHEDULED_DEPARTURE: "18:55"
retraso_salida: 1

Group

Output after [Group](#) stage (Sample of 20 documents)

```
1 {
2   _id: '$SCHEDULED_DEPARTURE',
3   promedio_retraso: {
4     $avg: '$DEPARTURE_DELAY'
5   },
6   n: {
7     $sum: 1
8   }
9 }
```

_id: "12:48"
promedio_retraso: 30.46153846153846
n: 39

_id: "18:18"
promedio_retraso: 40.520833333333336
n: 48

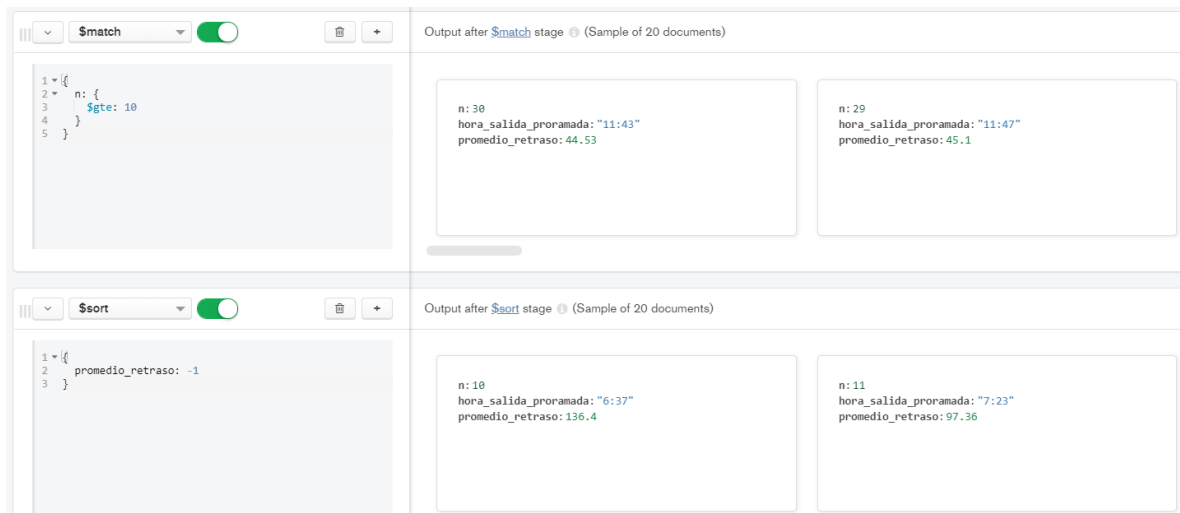
Project

Output after [Project](#) stage (Sample of 20 documents)

```
1 {
2   hora_salida_proramada: '$_id',
3   promedio_retraso: {$round: [
4     '$promedio_retraso', 2]
5   },
6   n: 1,
7   _id: 0
8 }
9
10
```

n: 30
hora_salida_proramada: "11:43"
promedio_retraso: 44.53

n: 29
hora_salida_proramada: "11:47"
promedio_retraso: 45.1



7. Promedio del número de vuelos que salen de los aeropuertos del estado de NY por mes

```
[{$group: {
  _id: {aeropuerto:"$ORIGIN_AIRPORT", mes: "$MONTH"},
  total:{$sum:1}
}}, {$addFields: {
  _id:"$_id.aeropuerto",
  mes:"$_id.mes"
}}, {$lookup: {
  from: 'airports',
  localField: '_id',
  foreignField: 'IATA_CODE',
  as: 'airports'
}}, {$addFields: {
  airports_obj: {
    $arrayElemAt: [
      '$airports', 0
    ]
  }
}}, {$addFields: {
  STATE: '$airports_obj.STATE',
  AIRPORT: '$airports_obj.AIRPORT'
}}, {$match:
{
  STATE:{$eq:"NY"}
}}, {$project:
{
  AIRPORT:1,
  STATE:1,
  mes:1,
  total:1
}}, {$group: {
```



```

_id: "$AIRPORT",
promedio_salidas_mes: {$avg: "$total"}
}}, {$sort: {
promedio_salidas_mes: -1
}}}

```

<div> <div>\$group</div> <div><input checked="" type="checkbox"/></div> <div>🗑️ +</div> </div> <pre> 1 * { 2 _id: { 3 aeropuerto: '\$ORIGIN_AIRPORT', 4 mes: '\$MONTH' 5 }, 6 total: { 7 \$sum: 1 8 } 9 } </pre>	<div>Output after \$group stage (Sample of 20 documents)</div> <div> <div> <div>_id: Object</div> <div>total: 1</div> </div> <div> <div>_id: Object</div> <div>total: 2</div> </div> </div>	
<div> <div>\$addFields</div> <div><input checked="" type="checkbox"/></div> <div>🗑️ +</div> </div> <pre> 1 * { 2 _id: '\$_id.aeropuerto', 3 mes: '\$_id.mes' 4 } </pre>	<div>Output after \$addFields stage (Sample of 20 documents)</div> <div> <div> <div>_id: "BOI"</div> <div>total: 23</div> <div>mes: 3</div> </div> <div> <div>_id: "ROC"</div> <div>total: 11</div> <div>mes: 3</div> </div> </div>	
<div> <div>\$lookup</div> <div><input checked="" type="checkbox"/></div> <div>🗑️ +</div> </div> <pre> 1 * { 2 from: 'airports', 3 localField: '_id', 4 foreignField: 'IATA_CODE', 5 as: 'airports' 6 } </pre>	<div>Output after \$lookup stage (Sample of 20 documents)</div> <div> <div> <div>_id: "11471"</div> <div>total: 1</div> <div>mes: 10</div> <div>airports: Array</div> </div> <div> <div>_id: "FLL"</div> <div>total: 131</div> <div>mes: 12</div> <div>airports: Array</div> </div> </div>	
<div> <div>\$addFields</div> <div><input checked="" type="checkbox"/></div> <div>🗑️ +</div> </div> <pre> 1 * { 2 airports_obj: { 3 \$arrayElemAt: [4 '\$airports', 5 0 6] 7 } 8 } </pre>	<div>Output after \$addFields stage (Sample of 20 documents)</div> <div> <div> <div>_id: "12173"</div> <div>total: 71</div> <div>mes: 10</div> <div>airports: Array</div> </div> <div> <div>_id: "CWA"</div> <div>total: 3</div> <div>mes: 12</div> <div>airports: Array</div> <div>airports_obj: Object</div> </div> </div>	
<div> <div>\$addFields</div> <div><input checked="" type="checkbox"/></div> <div>🗑️ +</div> </div> <pre> 1 * { 2 STATE: '\$airports_obj.STATE', 3 AIRPORT: '\$airports_obj.AIRPORT' 4 } </pre>	<div>Output after \$addFields stage (Sample of 20 documents)</div> <div> <div> <div>_id: "PSP"</div> <div>total: 19</div> <div>mes: 1</div> <div>airports: Array</div> <div>airports_obj: Object</div> <div>STATE: "CA"</div> <div>AIRPORT: "Palm Springs International Airport"</div> </div> <div> <div>_id: "MSY"</div> <div>total: 62</div> <div>mes: 3</div> <div>airports: Array</div> <div>airports_obj: Object</div> <div>STATE: "LA"</div> <div>AIRPORT: "Louis Armstrong New Orleans International Airport"</div> </div> </div>	
<div> <div>\$match</div> <div><input checked="" type="checkbox"/></div> <div>🗑️ +</div> </div> <pre> 1 * { 2 STATE: { 3 \$eq: 'NY' 4 } 5 } </pre>	<div>Output after \$match stage (Sample of 20 documents)</div> <div> <div> <div>_id: "JFK"</div> <div>total: 153</div> <div>mes: 3</div> <div>airports: Array</div> <div>airports_obj: Object</div> <div>STATE: "NY"</div> <div>AIRPORT: "John F. Kennedy International Airport (New York International Airport)"</div> </div> <div> <div>_id: "BUF"</div> <div>total: 25</div> <div>mes: 7</div> <div>airports: Array</div> <div>airports_obj: Object</div> <div>STATE: "NY"</div> <div>AIRPORT: "Buffalo Niagara International Airport"</div> </div> </div>	

\$project

Output after \$project stage (Sample of 20 documents)

```
1 * {
2   AIRPORT: 1,
3   STATE: 1,
4   mes: 1,
5   total: 1
6 }
```

```
_id: "ROC"
total: 11
mes: 3
STATE: "NY"
AIRPORT: "Greater Rochester International Airport"
```

```
_id: "ELM"
total: 2
mes: 1
STATE: "NY"
AIRPORT: "Elmira/Corning Regional Airport"
```

\$group

Output after \$group stage (Sample of 14 documents)

```
1 * {
2   id: '$AIRPORT',
3   promedio_salidas_mes: {
4     $avg: '$total'
5   }
6 }
```

```
_id: "LaGuardia Airport (Marine Air Terminal)"
promedio_salidas_mes: 154.18181818182
```

```
_id: "Long Island MacArthur Airport"
promedio_salidas_mes: 7.454545454545454
```

\$project

Output after \$project stage (Sample of 14 documents)

```
1 * {
2   _id: 0,
3   aeropuerto: '$ id',
4   promedio_salidas_mes: {
5     $round: [
6       '$promedio_salidas_mes',
7       2
8     ]
9   }
}
```

```
aeropuerto: "Greater Rochester International Airport"
promedio_salidas_mes: 10.55
```

```
aeropuerto: "Elmira/Corning Regional Airport"
promedio_salidas_mes: 3.45
```

\$sort

Output after \$sort stage (Sample of 14 documents)

```
1 * {
2   promedio_salidas_mes: -1
3 }
```

```
aeropuerto: "LaGuardia Airport (Marine Air Terminal)"
promedio_salidas_mes: 154.18
```

```
aeropuerto: "John F. Kennedy International Airport (New
International Airport)"
promedio_salidas_mes: 143.09
```

8. Obtén por aerolíneas la cantidad de vuelos cancelados por condiciones meteorológicas del estado de Texas

```
{
  $match: {
    CANCELLATION_REASON: 'Condiciones metereologicas'
  },
  $group: {
    _id: {aeropuerto: "$ORIGIN_AIRPORT",
      aerolinea: "$AIRLINE"},
    total: {
      $sum: 1
    }
  },
  $lookup: {
    from: 'airlines',
    localField: '_id.aerolinea',
    foreignField: 'IATA_CODE',
    as: 'aerolineas'
  },
  $lookup: {
    from: 'airports',
    localField: '_id.aeropuerto',
    foreignField: 'IATA_CODE',

```

```
as: 'aeropuertos'
}}, {$addField: {
aeropuertos_obj: {
  $arrayElemAt: [
    '$aeropuertos',
    0]},
aerolineas_obj: {
  $arrayElemAt: [
    '$aerolineas',
    0]}
}}, {$addField: {
aeropuertos: "$aeropuertos_obj.AIRPORT",
estados: "$aeropuertos_obj.STATE",
aerolineas: "$aerolineas_obj.AIRLINE"
}}, {$match: {
estados:"TX"
}}, {$group: {
  _id: "$aerolineas",
  total:{$sum:"$total"}
}}, {$project: { _id:0,
aerolineas: "$_id",
total:1
}}, {$sort: {
total:-1
}}}
```

\$match ☒

```

1 {
2   CANCELLATION_REASON: 'condiciones metereologicas'
3 }

```

Output after **\$match** stage (Sample of 20 documents)

```

_id: ObjectId("5f25de4f8aecc42c38471979")
YEAR: 2015
MONTH: 8
DAY: 18
DAY_OF_WEEK: 2
AIRLINE: "MQ"
ARRIVAL_DELAY: NaN
DEPARTURE_DELAY: NaN
SCHEDULED_DEPARTURE: "20-20"

```

```

_id: ObjectId("5f25de4f8aecc42c38471987")
YEAR: 2015
MONTH: 1
DAY: 27
DAY_OF_WEEK: 2
AIRLINE: "UA"
ARRIVAL_DELAY: NaN
DEPARTURE_DELAY: NaN
SCHEDULED_DEPARTURE: "27-20"

```

\$group ☒

```

1 {
2   _id: {aeropuerto: "$ORIGIN_AIRPORT",
3     aerolinea: "$AIRLINE"},
4   total: {
5     $sum: 1
6   }
7 }

```

Output after **\$group** stage (Sample of 20 documents)

```

_id: Object
total: 2

```

```

_id: Object
total: 2

```

⌵

\$lookup

🔴

🗑️

+

Output after [\\$lookup](#) stage ⓘ (Sample of 20 documents)

```
1 {
2   from: 'airlines',
3   localField: '_id.aerolinea',
4   foreignField: 'IATA_CODE',
5   as: 'aerolineas'
6 }
```

▶ **_id**: Object
total: 57
▶ aerolineas: Array

▶ **_id**: Object
total: 2
▶ aerolineas: Array

⌵

\$lookup

🔴

🗑️

+

Output after [\\$lookup](#) stage ⓘ (Sample of 20 documents)

```
1 {
2   from: 'airports',
3   localField: '_id.aeropuerto',
4   foreignField: 'IATA_CODE',
5   as: 'aeropuertos'
6 }
```

▶ **_id**: Object
total: 1
▶ aerolineas: Array
▶ aeropuertos: Array

▶ **_id**: Object
total: 2
▶ aerolineas: Array
▶ aeropuertos: Array

⌵

\$addFields

🔴

🗑️

+

Output after [\\$addFields](#) stage ⓘ (Sample of 20 documents)

```
1 {
2   aeropuertos_obj: {
3     $arrayElemAt: [
4       '$aeropuertos',
5       0
6     ]
7   },
8   aerolineas_obj: {
9     $arrayElemAt: [
10      '$aerolineas',
11      0
12    ]
13  }
14 }
```

▶ **_id**: Object
total: 4
▶ aerolineas: Array
▶ aeropuertos: Array
▶ aeropuertos_obj: Object
▶ aerolineas_obj: Object

▶ **_id**: Object
total: 1
▶ aerolineas: Array
▶ aeropuertos: Array
▶ aeropuertos_obj: Object
▶ aerolineas_obj: Object

⌵

\$addFields

🔴

🗑️

+

Output after [\\$addFields](#) stage ⓘ (Sample of 20 documents)

```
1 {
2   aeropuertos: "$aeropuertos_obj.AIRPORT",
3   estados: "$aeropuertos_obj.STATE",
4   aerolineas: "$aerolineas_obj.AIRLINE"
5 }
```

▶ **_id**: Object
total: 3
aerolineas: "American Eagle Airlines Inc."
aeropuertos: "Central Illinois Regional Airport at Bloomington"
▶ aeropuertos_obj: Object
▶ aerolineas_obj: Object
estados: "IL"

▶ **_id**: Object
total: 5
aerolineas: "Spirit Air Lines"
aeropuertos: "Fort Lauderdale-Hollywood International Airport"
▶ aeropuertos_obj: Object
▶ aerolineas_obj: Object
estados: "FL"

⌵

\$match

🔴

🗑️

+

Output after [\\$match](#) stage ⓘ (Sample of 20 documents)

```
1 {
2   estados: "TX"
3 }
```

▶ **_id**: Object
total: 5
aerolineas: "Southwest Airlines Co."
aeropuertos: "Lubbock Preston Smith International Airport"
▶ aeropuertos_obj: Object
▶ aerolineas_obj: Object
estados: "TX"

▶ **_id**: Object
total: 1
aerolineas: "Atlantic Southeast Airlines"
aeropuertos: "Easterwood Airport"
▶ aeropuertos_obj: Object
▶ aerolineas_obj: Object
estados: "TX"

⌵

\$group

🔴

🗑️

+

Output after [\\$group](#) stage ⓘ (Sample of 12 documents)

```
1 {
2   _id: "$aerolineas",
3   total: { $sum: "$total" }
4 }
```

▶ **_id**: "American Eagle Airlines Inc."
total: 246

▶ **_id**: "Southwest Airlines Co."
total: 185

Project

Output after **Project** stage (Sample of 12 documents)

```

1 { _id:0,
2   aerolineas: "$_id",
3   total:1
4 }

```

total:5
aerolineas:"US Airways Inc."

total:1
aerolineas:"Frontier Airlines Inc."

Sort

Output after **Sort** stage (Sample of 12 documents)

```

1 {
2   total:-1
3 }

```

total:246
aerolineas:"American Eagle Airlines Inc."

total:239
aerolineas:"American Airlines Inc."

A partir de esta sección las preguntas corresponden a la base Twitter_aerolineas, y son datos de Tweets de viajeros recopilados del 17 al 24 de febrero 2015

9. ¿Principales emociones negativas de usuarios a través de Twitter sobre las aerolíneas estadounidenses?

```

[{$match: {
  airline_sentiment: 'negative'
}}, {$group: {
  _id: '$negativereason',
  Total: {
    $sum: 1
  }
}}, {$project: {_id:0,
  razon_negativa:"$_id",
  Total:1
}}, {$sort: {
  Total: -1
}}]

```



```
}}, {$sort: {
  Total: -1
}}]
```

\$match

Output after \$match stage (Sample of 20 documents)

```
1 * {
2   airline_sentiment: 'negative'
3 }
```

```
_id: ObjectId("5f25f5a28aecc42c384eba97")
tweet_id: 578301031407624200
airline_sentiment: "negative"
airline_sentiment_confidence: 1
negativereason: "Bad Flight"
negativereason_confidence: 0.7833
airline: "Virgin America"
airline_sentiment_gold: ""
name: "inandino"
```

```
_id: ObjectId("5f25f5a28aecc42c384eba98")
tweet_id: 578300817074462700
airline_sentiment: "negative"
airline_sentiment_confidence: 1
negativereason: "Can't Tell"
negativereason_confidence: 1
airline: "Virgin America"
airline_sentiment_gold: ""
name: "inandino"
```

\$group

Output after \$group stage (Sample of 6 documents)

```
1 * {
2   _id: '$airline',
3   Total: {
4     $sum: 1
5   }
6 }
```

```
_id: "Virgin America"
Total: 181
```

```
_id: "Delta"
Total: 955
```

\$project

Output after \$project stage (Sample of 6 documents)

```
1 * {
2   _id: 0,
3   aerolinea: '$_id',
4   Total: 1
5 }
```

```
Total: 2633
aerolinea: "United"
```

```
Total: 181
aerolinea: "Virgin America"
```

\$sort

Output after \$sort stage (Sample of 6 documents)

```
1 * {
2   Total: -1
3 }
```

```
Total: 2633
aerolinea: "United"
```

```
Total: 2263
aerolinea: "US Airways"
```

11. ¿Hora con mayor incidencia de tweets negativos?

```
[{$addFields: {
  date: {
    $dateFromString: {
      dateString: '$tweet_created'
    }
  }
}}, {$addFields: {
  hora: {
    $hour: '$date'
  }
}}, {$group: {
  _id: '$hora',
  total: {
    $sum: 1
  }
}}
```

```
}
}}, {$project: { _id:0,
  hora:"$_id",
  total:1
}}, {$sort: {
  total: -1
}}]
```

SaddFields

Output after \$addFields stage (Sample of 20 documents)

```
1 {  
2   date: {  
3     $dateFromString: {  
4       dateString: '$tweet_created'  
5     }  
6   }  
7 }
```

_id: ObjectId("5f25f5a28aec42c384eba94")
 tweet_id: 570306133677760500
 airline_sentiment: "neutral"
 airline_sentiment_confidence: 1
 negativereason: ""
 negativereason_confidence: 0
 airline: "Virgin America"
 airline_sentiment_gold: ""
 name: "crandien"

_id: ObjectId("5f25f5a28aec42c384eba95")
 tweet_id: 570301130888122400
 airline_sentiment: "positive"
 airline_sentiment_confidence: 0.3486
 negativereason: ""
 negativereason_confidence: 0
 airline: "Virgin America"
 airline_sentiment_gold: ""
 name: "daandien"

SaddFields

Output after \$addFields stage (Sample of 20 documents)

```
1 {  
2   hora: {  
3     $hour: '$date'  
4   }  
5 }
```

_id: ObjectId("5f25f5a28aec42c384eba94")
 tweet_id: 570306133677760500
 airline_sentiment: "neutral"
 airline_sentiment_confidence: 1
 negativereason: ""
 negativereason_confidence: 0
 airline: "Virgin America"
 airline_sentiment_gold: ""
 name: "crandien"

_id: ObjectId("5f25f5a28aec42c384eba95")
 tweet_id: 570301130888122400
 airline_sentiment: "positive"
 airline_sentiment_confidence: 0.3486
 negativereason: ""
 negativereason_confidence: 0
 airline: "Virgin America"
 airline_sentiment_gold: ""
 name: "daandien"

\$group

Output after \$group stage (Sample of 20 documents)

```
1 {  
2   _id: '$hora',  
3   total: {  
4     $sum: 1  
5   }  
6 }
```

_id: 22
 total: 927

_id: 13
 total: 477

\$project

Output after \$project stage (Sample of 20 documents)

```
1 {  
2   _id: 0,  
3   hora: '$_id',  
4   total: 1  
5 }
```

total: 602
 hora: 4

total: 754
 hora: 1

\$sort

Output after \$sort stage (Sample of 20 documents)

```
1 {  
2   total: -1  
3 }
```

total: 1013
 hora: 17

total: 986
 hora: 19

total: 962 hora: 18	total: 927 hora: 22	total: 926 hora: 16
------------------------	------------------------	------------------------

12. ¿ubicación con mayores publicaciones de tweets negativos?

```
[{$group: {
  _id: '$tweet_location',
  total: {
    $sum: 1
  }
}}, {$match: {
  _id: {
    $ne: "
  }
}}, {$project:
  {_id:0,
   localizacion:"$_id",
   total:1
}}, {$sort: {
  total: -1
}}]
```

1

2

3

4

5

6

\$group

Output after \$group stage (Sample of 20 documents)

_id: "Queenstown, MD"

total: 1

_id: "Somewhere eating swine..."

total: 2

1

2

3

4

5

\$match

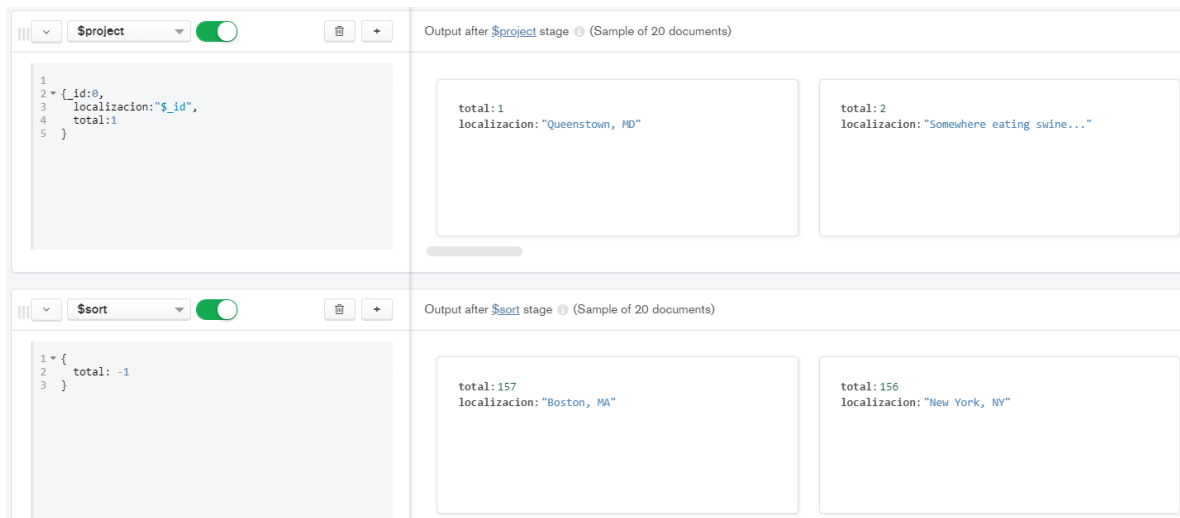
Output after \$match stage (Sample of 20 documents)

_id: "Zamunda "

total: 1

_id: "McKinney TX"

total: 22



13. ¿Obtener por tipo de tweet la cantidad de publicaciones que hay en cada aerolínea estadounidense?

```
[{$group: {
  _id: {
    tipo_tweet: '$airline_sentiment',
    aerolinea: '$airline'
  },
  total: {
    $sum: 1
  }
}}, {$addFields: {
  tipo_tweet: '$_id.tipo_tweet',
  aerolinea: '$_id.aerolinea'
}}, {$project: {
  _id: 0,
  tipo_tweet: 1,
  aerolinea: 1,
  total: 1
}}, {$sort: {
  aerolinea: 1,
  total: -1,
  tipo_tweet: -1
}}]
```



```

      'if': {
        $eq: [
          '$airline_sentiment',
          'positive'
        ]
      },
      then: 1,
      'else': 0
    }
  },
  neutral: {
    $cond: {
      'if': {
        $eq: [
          '$airline_sentiment',
          'neutral'
        ]
      },
      then: 1,
      'else': 0
    }
  }
}, {$addField: {
  total: {
    $sum: 1
  }
}}, {$group: {
  _id: '$airline',
  negativos: {
    $sum: '$negativo'
  },
  positivos: {
    $sum: '$positivo'
  },
  neutral: {
    $sum: '$neutral'
  },
  total: {
    $sum: '$total'
  }
}}, {$addField: {
  indice_negativos: {
    $divide: [
      '$negativos',
      '$total'
    ]
  }
}}, {$project: { _id:0,

```

```

aerolinea:"$_id",
negativos:1,
positivos:1,
neutral:1,
total:1,
indice_negativos: {
  $round: [
    '$indice_negativos', 2
  ]
}
}}, {$sort: {
  indice_negativos:-1
}}]

```

Project

🗑️ ➕

```

1 = {
2   airline: 1,
3   negativo: {
4     $cond: {
5       'if': {
6         $eq: [
7           '$airline_sentiment',
8           'negative'
9         ]
10      },
11      then: 1,
12      'else': 0
13    },
14  },
15  positivo: {
16    $cond: {
17      'if': {
18        $eq: [
19          '$airline_sentiment',
20          'positive'
21        ]
22      },
23      then: 1,
24      'else': 0
25    },
26  },
27  neutral: {
28    $cond: {
29      'if': {
30        $eq: [
31          '$airline_sentiment',
32          'neutral'
33        ]
34      },
35      then: 1,
36      'else': 0
37    },
38  }
39 }

```

Output after \$project stage (Sample of 20 documents)

```

_id: ObjectId("5f25f5a28aecc42c384eba94")
airline: "Virgin America"
negativo: 0
positivo: 0
neutral: 1

```

```

_id: ObjectId("5f25f5a28aecc42c384eba95")
airline: "Virgin America"
negativo: 0
positivo: 1
neutral: 0

```

SaddFields

🗑️ ➕

```

1 = {
2   total: {
3     $sum: 1
4   }
5 }

```

Output after \$addFields stage (Sample of 20 documents)

```

_id: ObjectId("5f25f5a28aecc42c384eba94")
airline: "Virgin America"
negativo: 0
positivo: 0
neutral: 1
total: 1

```

```

_id: ObjectId("5f25f5a28aecc42c384eba95")
airline: "Virgin America"
negativo: 0
positivo: 1
neutral: 0
total: 1

```

\$group

🗑️ ➕

```

1 = {
2   _id: '$airline',
3   negativos: {
4     $sum: '$negativo'
5   },
6   positivos: {
7     $sum: '$positivo'
8   },
9   neutral: {
10    $sum: '$neutral'
11  },
12  total: {
13    $sum: '$total'
14  }
15 }

```

Output after \$group stage (Sample of 6 documents)

```

_id: "Delta"
negativos: 955
positivos: 544
neutral: 723
total: 2222

```

```

_id: "US Airways"
negativos: 2263
positivos: 269
neutral: 381
total: 2913

```

SaddFields

Output after [SaddFields](#) stage (Sample of 6 documents)

```

1 {
2   indice_negativos: {
3     $divide: [
4       '$negativos',
5       '$total'
6     ]
7   }
8 }

```

_id: "Virgin America"

negativos: 181

positivos: 152

neutral: 171

total: 504

indice_negativos: 0.35912698412698413

_id: "Delta"

negativos: 955

positivos: 544

neutral: 723

total: 2222

indice_negativos: 0.4297929792979298

Sproject

Output after [Sproject](#) stage (Sample of 6 documents)

```

1 { _id: 0,
2   aerolinea: "$_id",
3   negativos: 1,
4   positivos: 1,
5   neutral: 1,
6   total: 1,
7   indice_negativos: {
8     $round: [
9       '$indice_negativos', 2
10    ]
11  }
12 }

```

negativos: 955

positivos: 544

neutral: 723

total: 2222

aerolinea: "Delta"

indice_negativos: 0.43

negativos: 2263

positivos: 269

neutral: 381

total: 2913

aerolinea: "US Airways"

indice_negativos: 0.78

Ssort

Output after [Ssort](#) stage (Sample of 6 documents)

```

1 {
2   indice_negativos: -1
3 }

```

negativos: 2263

positivos: 269

neutral: 381

total: 2913

aerolinea: "US Airways"

indice_negativos: 0.78

negativos: 1960

positivos: 336

neutral: 463

total: 2759

aerolinea: "American"

indice_negativos: 0.71

15. Obtener la proporción de emociones negativas vs positivas por aerolínea

```

[{$project: {
  airline: 1,
  negativo: {
    $cond: {
      'if': {
        $eq: [
          '$airline_sentiment',
          'negative'
        ]
      },
      then: 1,
      'else': 0
    }
  },
  positivo: {
    $cond: {
      'if': {
        $eq: [
          '$airline_sentiment',
          'positive'
        ]
      },
      then: 1,
      'else': 0
    }
  }
}]

```

```

    }
  }, { $group: {
    _id: '$airline',
    negativos: {
      $sum: '$negativo'
    },
    positivos: {
      $sum: '$positivo'
    }
  } }, { $addFields: {
    proporcion_negativos: {
      $divide: [
        '$negativos',
        '$positivos'
      ]
    }
  } }, { $project: {
    _id: 0,
    aerolinea: '$_id',
    proporcion_negativos: {
      $round: [
        '$proporcion_negativos',
        2
      ]
    }
  } }, { $sort: {
    proporcion_negativos: -1
  } }
}

```

Project
☒

+

Output after [\\$project](#) stage (Sample of 20 documents)

```

1 {
2   airline: 1,
3   negativo: {
4     $cond: {
5       'if': {
6         $eq: [
7           '$airline_sentiment',
8           'negative'
9         ]
10      },
11      then: 1,
12      'else': 0
13    }
14  },
15  positivo: {
16    $cond: {
17      'if': {
18        $eq: [
19          '$airline_sentiment',
20          'positive'
21        ]
22      },
23      then: 1,
24      'else': 0
25    }
26  }
27 }

```

_id: ObjectId("5f25f5a28aecc42c384eba94")
airline: "Virgin America"
negativo: 0
positivo: 0

_id: ObjectId("5f25f5a28aecc42c384eba95")
airline: "Virgin America"
negativo: 0
positivo: 1

\$group

1 {
2 _id: '\$airline',
3 negativos: {
4 \$sum: '\$negativo'
5 },
6 positivos: {
7 \$sum: '\$positivo'
8 }
9 }

Output after \$group stage ⓘ (Sample of 6 documents)

_id: "Delta"
negativos: 955
positivos: 544

_id: "US Airways"
negativos: 2263
positivos: 269

\$addFields

1 {
2 proporcion_negativos: {
3 \$divide: [
4 '\$negativos',
5 '\$positivos'
6]
7 }
8 }

Output after \$addFields stage ⓘ (Sample of 6 documents)

_id: "Delta"
negativos: 955
positivos: 544
proporcion_negativos: 1.755514705882353

_id: "US Airways"
negativos: 2263
positivos: 269
proporcion_negativos: 8.41263940520446

\$project

1 {
2 _id: 0,
3 aerolinea: '\$_id',
4 proporcion_negativos: {
5 \$round: [
6 '\$proporcion_negativos',
7 2
8]
9 }
10 }

Output after \$project stage ⓘ (Sample of 6 documents)

aerolinea: "United"
proporcion_negativos: 5.35

aerolinea: "Virgin America"
proporcion_negativos: 1.19

\$sort

1 {
2 proporcion_negativos: -1
3 }

Output after \$sort stage ⓘ (Sample of 6 documents)

aerolinea: "US Airways"
proporcion_negativos: 8.41

aerolinea: "American"
proporcion_negativos: 5.83

Vistas en Mongo DB

Vista 1.- `vuelos_retrasos_aerolineas`. La vista corresponde a la pregunta 1.

[Data_Analysis.vuelos_retrasos_aerolineas](#) (view on: `Data_Analysis.sample_flights`)

	Documents	Aggregations	Schema	Explain Plan	Indexes	Validation
FILTER						
VIEW						
vuelos_retrasos_aerolineas						
	Total_retraso_Llegada Int32	airline String				
1	40548	"Southwest Airlines Co."				
2	21800	"American Airlines Inc."				
3	21505	"Delta Air Lines Inc."				
4	19350	"Skywest Airlines Inc."				
5	18291	"Atlantic Southeast Airlines"				
6	15832	"United Air Lines Inc."				
7	8889	"American Eagle Airlines Inc."				
8	8800	"JetBlue Airways"				
9	6536	"US Airways Inc."				
10	4888	"Alaska Airlines Inc."				
11	4723	"Spirit Air Lines"				
12	3604	"Frontier Airlines Inc."				
13	2539	"Hawaiian Airlines Inc."				
14	2099	"Virgin America"				

Vista 2.- `indice_emociones_negativas`. La vista corresponde a la pregunta 14

[Data_Analysis.indice_emociones_negativas](#) (view on: `Data_Analysis.Twitter_aerolineas`) [MODIFY SOURCE](#)

	Documents	Aggregations	Schema	Explain Plan	Indexes	Validation
FILTER						
OPTIONS FIND						
VIEW						
Displaying documents 1 - 6 of 6						
indice_emociones_negativas						
	negativos Int32	positivos Int32	neutral Int32	total Int32	aerolinea String	indice_negativos Dou
1	2263	269	381	2913	"US Airways"	0.78
2	1960	336	463	2759	"American"	0.71
3	2633	492	697	3822	"United"	0.69
4	1186	570	664	2420	"Southwest"	0.49
5	955	544	723	2222	"Delta"	0.43
6	181	152	171	504	"Virgin America"	0.36

Conclusiones:

- La aerolínea con mayor cantidad de retraso en sus llegadas es la aerolínea **Southwest Airlines Co.** con 40,548, seguido con 21,800 la aerolínea **American Airlines Inc.** y **Delta Air Lines Inc.** con 21,505.
- Las principales razones de cancelación de vuelos son por razones de las aerolíneas (errores mecánicos y problemas con los horarios de la tripulación), seguido de las condiciones meteorológicas.
- La mayor cantidad de vuelos se realizan entre los días de semana lunes a viernes.
- El aeropuerto con mayor retraso en sus vuelos tanto en llegadas como salidas es **Hartsfiels-Jackson Atlanta International.**
- El estado con más aeropuertos es Texas.
- Las horas de salida programadas abiertas 6:37, 7:23 que no terminan en 0 o 30, son las horas que presentan el mayor promedio de retraso de salida de un avión.
- Las principales razones negativas de viajeros que expresaron sus emociones en Twitter sobre las aerolíneas son: servicio al cliente, vuelo tardío, vuelo cancelado, equipaje perdido, vuelo malo.
- Las Aerolíneas estadounidenses con más tweets negativos son **United Air Lines Inc.** y **US Airways Inc.**
- Se confirma a través de los indicadores propuestos como índice de emociones y razón de proporción negativa que la aerolínea con más atrasos son **US Airways Inc.** y **American Airlines Inc.**
- Finalmente, las aerolíneas menos recomendables para realizar un vuelo son:
 - **Southwest Airlines Co.**
 - **American Airlines Inc.**
 - **Delta Air Lines Inc.**
 - **US Airways Inc.**