Saulo Gabriel López Antonio

**data-analysis-gdl-20-04**

PROYECTO

“Captación y cuentas de ahorro de la banca mexicana”

# 1.- Definición del proyecto (fuentes y problemas a resolver)

## A) Fuente Base de datos

<https://portafoliodeinformacion.cnbv.gob.mx/bm1/Paginas/infoper.aspx>

Reporte 040-4A-R10 y 040-4A-R11 del 2020

## B) *Problema a resolver*

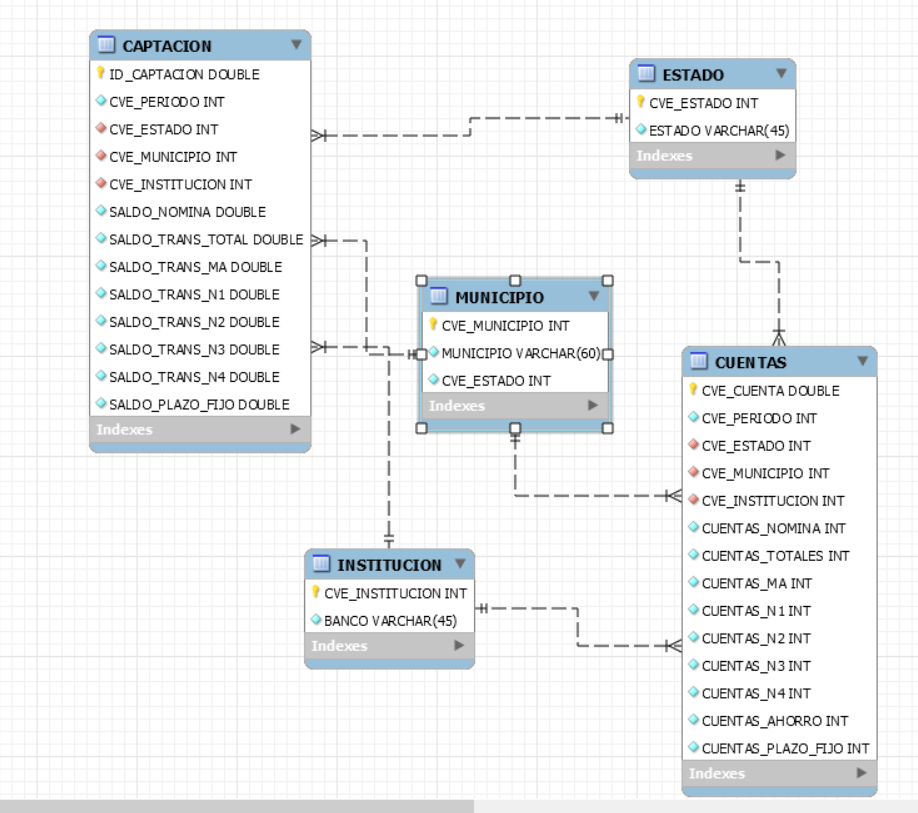
Cierta institución bancaria requiere conocer el mercado de las cuentas de captación y el saldo que hay por estado y/o municipio en México, además de saber cómo es la situación del resto de las instituciones bancarias

# 2.- Definición de la base de datos

## Archivos utilizados:

1. Cuentas.csv (Base con la información histórica de las cuentas de ahorro)
2. Captación.csv (Base con la información histórica del saldo de las cuentas de ahorro)
3. Municipio.csv (Base con nombres de los municipios)
4. Estado.csv (Base con los nombres de los Estados)
5. Institucion.csv (Base con los nombres de los bancos de México)

## Diagrama entidad relación



## Glosario

|  |  |  |
| --- | --- | --- |
| Base Cuentas | | |
| CVE\_CUENTA | | Clave única de la base de cuentas |
| CVE\_PERIODO | | Clave del periodo que tiene el registro |
| CVE\_ESTADO | | Clave del estado al que pertenece el registro |
| CVE\_MUNICIPIO | | Clave del municipio al que pertenece el registro |
| CVE\_INSTITUCION | | Clave de la institución al que pertenece el registro |
| CUENTAS\_NOMINA | | Numero de cuentas de nomina |
| CUENTAS\_TOTALES | | Cuentas de ahorro transaccionales (N1+N2+N3+N4) |
| CUENTAS\_MA | | Cuentas de ahorro transaccionales que no tienen algún crédito con la institución |
| CUNETAS\_NIVEL1 | | Cuentas transaccionales de nivel 1 (N1) |
| CUENTAS\_NIVEL2 | | Cuentas transaccionales de nivel 2 (N2) |
| CUENTAS\_NIVEL3 | | Cuentas transaccionales de nivel 3 (N3) |
| CUENTAS\_NIVEL4 | | Cuentas transaccionales de nivel 4 (N4) |
| CUENTAS\_AHORRO | | Cuentas nivel 4 pertenecientes al grupo de la institución (subuniverso N4) |
| CUENTAS\_PLAZOFIJO | | Cuentas de plazo fijo |
| Base Captacion | | |
| ID\_CAPTACION | Clave única de la base de captación | |
| CVE\_PERIODO | Clave del periodo que tiene el registro | |
| CVE\_ESTADO | Clave del estado al que pertenece el registro | |
| CVE\_MUNICIPIO | Clave del municipio al que pertenece el registro | |
| CVE\_INSTITUCION | Clave de la institución al que pertenece el registro | |
| SALDO\_NOMINA | Saldo de cuentas de nomina | |
| SALDO\_TRANS\_TOTAL | Saldo de ahorro transaccionales (N1+N2+N3+N4) | |
| SALDO\_TRANS\_MA | Saldo cuentas de ahorro transaccionales que no tienen algún crédito con la institución | |
| SALDO\_TRANS\_N1 | Saldo de cuentas transaccionales de nivel 1 (N1) | |
| SALDO\_TRANS\_N2 | Saldo de cuentas transaccionales de nivel 2 (N2) | |
| SALDO\_TRANS\_N3 | Saldo de cuentas transaccionales de nivel 3 (N3) | |
| SALDO\_TRANS\_N4 | Saldo de cuentas transaccionales de nivel 4 (N4) | |
| SALDO\_AHORRO | Saldo nivel 4 pertenecientes al grupo de la institución (subuniverso N4) | |
| SALDO\_PLAZO\_FIJO | Saldo de cuentas de plazo fijo | |

|  |  |
| --- | --- |
| Base Estado | |
| CVE\_ESTADO | Clave única de la base de Estado |
| ESTADO | Nombre del Estado de la República Mexicana |

|  |  |
| --- | --- |
| Base Institución | |
| CVE\_INSTITUCION | Clave única de la base Institución |
| BANCO | Nombre de la Institución bancaria |

|  |  |
| --- | --- |
| Base Municipio | |
| CVE\_MUNICIPIO | Clave única de la base Municipio |
| MUNICIPIO | Nombre del Municipio |
| CVE\_ESTADO | Clave del estado al que pertenece el Municipio |

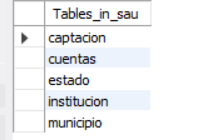
# Sesión 1: Fundamentos de SQL

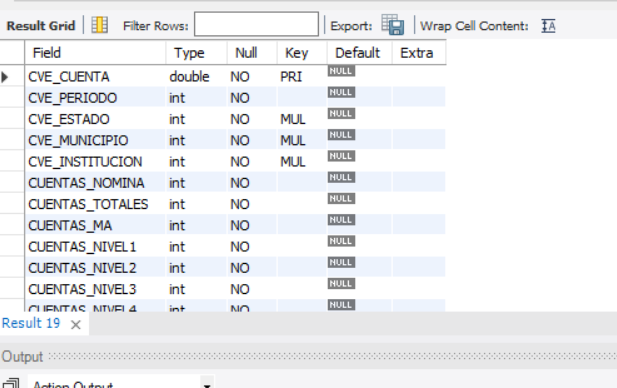
1.1 Usando la base de datos SAU, mostrar las tablas que contienen y la estructura de la tabla "cuenta"

USE SA;

SHOW TABLES;

DESCRIBE CUENTAS;



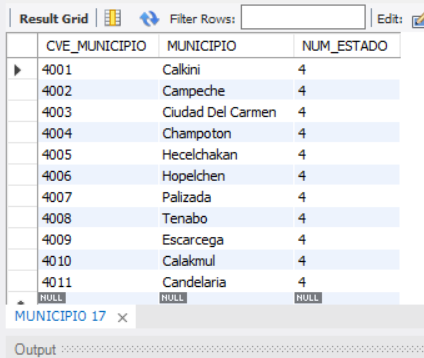


1.2 ¿Cuáles son los municipios del estado 4?

SELECT \*

FROM MUNICIPIO

WHERE num\_estado = 4



1.3 ¿Cuáles son los bancos y que saldo total de cuentas transaccionales se tienen en mayo del 2020 (202005), que no tengan saldo de plazo fijo y que sean del municipio 2002?

SELECT

CVE\_INSTITUCION,

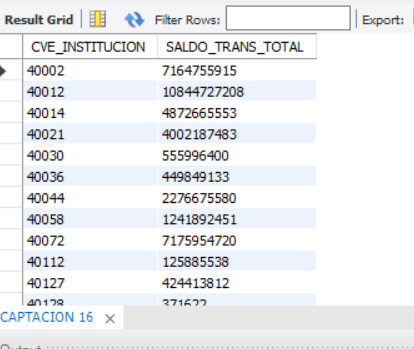
SALDO\_TRANS\_TOTAL

from CAPTACION

WHERE CVE\_PERIODO = 202005

AND CVE\_MUNICIPIO = 2002

AND SALDO\_PLAZO\_FIJO > 0



1.4 ¿Cuál es el top 9 de bancos que tienen más cuentas nivel 2 en abril y que sean del municipio 09014?

SELECT

CVE\_INSTITUCION,

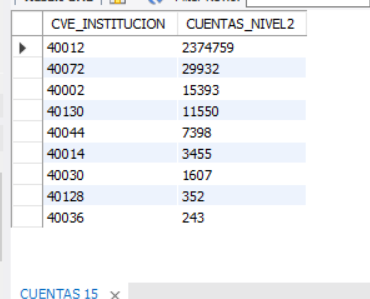
CUENTAS\_NIVEL2

FROM CUENTAS

WHERE CVE\_MUNICIPIO = 9014

AND CVE\_PERIODO = 202004

order by CUENTAS\_NIVEL2 desc limit 9



1.5 ¿Cuáles son los 20 municipios con menos saldo de plazo fijo del banco 40012 en enero del 2020?

select

CVE\_MUNICIPIO,

SALDO\_PLAZO\_FIJO

from CAPTACION

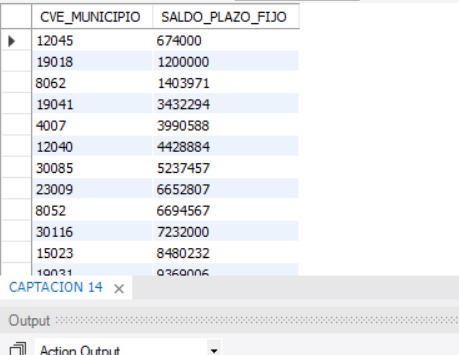
WHERE CVE\_INSTITUCION = 40012

AND CVE\_PERIODO = 202001

AND SALDO\_PLAZO\_FIJO>0

ORDER BY SALDO\_PLAZO\_FIJO

limit 20



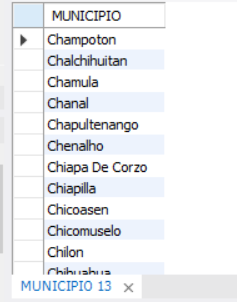
# Sesión 2: Agrupaciones y subconsultas

2.1 ¿Cuáles son todos los municipios que empiezan con 'Ch'?

SELECT MUNICIPIO

FROM MUNICIPIO

WHERE MUNICIPIO LIKE 'Ch%'



2.2 Obtener el total de cuentas (Nomina, plazo fijo y transaccionales) en mayo del 2020 y el total general de estas

select

SUM(CUENTAS\_TOTALES) AS CUENTAS\_TRANSACCIONALES,

SUM(CUENTAS\_NOMINA) AS NOMINA,

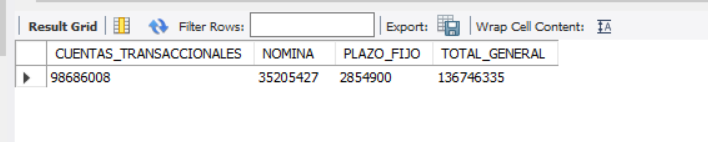
SUM(CUENTAS\_PLAZOFIJO) AS PLAZO\_FIJO,

SUM(CUENTAS\_TOTALES) + SUM(CUENTAS\_NOMINA) + SUM(CUENTAS\_PLAZOFIJO) AS TOTAL\_GENERAL

FROM

CUENTAS

WHERE CVE\_PERIODO = 202005



2.3 Obtener el total de saldo de plazo fijo que hay por institución en cada periodo y el numero de municipios en donde se encuentran

Select

CVE\_PERIODO,

CVE\_INSTITUCION,

SUM(SALDO\_PLAZO\_FIJO) AS SALDO\_PLAZO\_FIJO,

COUNT(CVE\_MUNICIPIO) AS MUNICIPIO

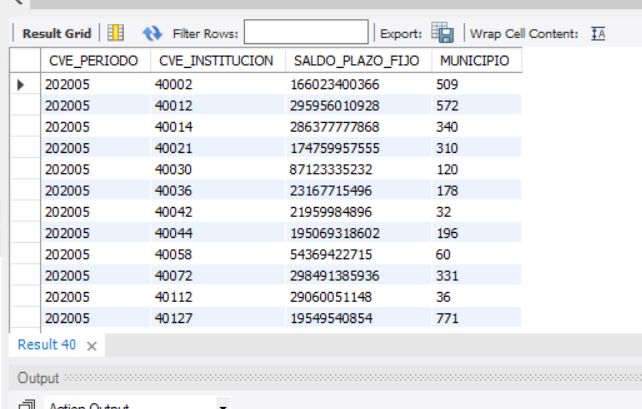
from CAPTACION

WHERE SALDO\_PLAZO\_FIJO > 0

group by

CVE\_PERIODO,

CVE\_INSTITUCION



2.4 ¿Cuál es el nombre las instituciones bancarias que no tienen saldo plazo fijo en el municipio 2002 o tienen cuenta nivel 3 en cualquiera de los periodos?

SELECT BANCO

FROM INSTITUCION

WHERE CVE\_INSTITUCION IN

(select cve\_institucion

FROM

CAPTACION

where SALDO\_PLAZO\_FIJO > 0

AND CVE\_MUNICIPIO = 2002)

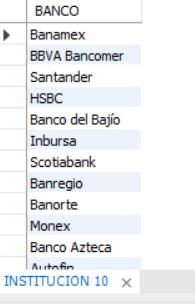
OR CVE\_INSTITUCION IN

(select

cve\_institucion

FROM CUENTAS

WHERE CUENTAS\_NIVEL3 > 0)



2.5 ¿Cuál es el total de cuentas de Compartamos que se tienen por estado y en cuantos municipios tienen cuentas durante mayo?

SELECT

CVE\_ESTADO,

SUM(CUENTAS\_TOTALES) AS CUENTAS,

COUNT(CVE\_MUNICIPIO) AS MUNICIPIOS

FROM CUENTAS

WHERE CVE\_PERIODO = 202005

AND CUENTAS\_TOTALES > 0

AND CVE\_INSTITUCION = (Select CVE\_INSTITUCION

from INSTITUCION

WHERE BANCO LIKE '%compartamos%')

GROUP BY CVE\_ESTADO



# Sesión 3: Joins y Vistas

3.1 ¿Cuáles son los 3 bancos con más saldo de nómina en febrero 2020?

select

B.BANCO,

C.NOMINA AS SALDO

from

(

SELECT

CVE\_INSTITUCION,

SUM(SALDO\_NOMINA) AS NOMINA

FROM

CAPTACION

WHERE CVE\_PERIODO = 202002

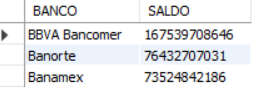
GROUP BY CVE\_INSTITUCION

) AS C

LEFT JOIN INSTITUCION B

ON C.CVE\_INSTITUCION = B.CVE\_INSTITUCION

ORDER BY C.NOMINA DESC LIMIT 3



3.2 ¿Cuántos bancos tienen al menos una cuenta N4 por municipio y estado al cierre de mayo?

SELECT

E.ESTADO,

M.MUNICIPIO,

BANCOS

FROM

(

SELECT

COUNT(CVE\_INSTITUCION) AS BANCOS,

CVE\_MUNICIPIO,

CVE\_ESTADO

from CUENTAS

WHERE CUENTAS\_TOTALES > 0

AND CVE\_PERIODO = 202005

GROUP BY

CVE\_MUNICIPIO,

CVE\_ESTADO

) AS C

LEFT JOIN

MUNICIPIO AS M

ON M.CVE\_MUNICIPIO = C.CVE\_MUNICIPIO

LEFT JOIN

ESTADO AS E

ON E.CVE\_ESTADO = C.CVE\_ESTADO



3.3 Usando la consulta anterior obtener el saldo total que hay por cada municipio

SELECT

E.ESTADO,

M.MUNICIPIO,

BANCOS,

S.SALDO

FROM

(

SELECT

COUNT(CVE\_INSTITUCION) AS BANCOS,

CVE\_MUNICIPIO,

CVE\_ESTADO

from CUENTAS

WHERE CUENTAS\_TOTALES > 0

AND CVE\_PERIODO = 202005

GROUP BY

CVE\_MUNICIPIO,

CVE\_ESTADO

) AS C

LEFT JOIN

MUNICIPIO AS M

ON M.CVE\_MUNICIPIO = C.CVE\_MUNICIPIO

LEFT JOIN

ESTADO AS E

ON E.CVE\_ESTADO = C.CVE\_ESTADO

LEFT JOIN

(SELECT

CVE\_MUNICIPIO,

SUM(SALDO\_TRANS\_TOTAL) AS SALDO

from CAPTACION

WHERE CVE\_PERIODO = 202005

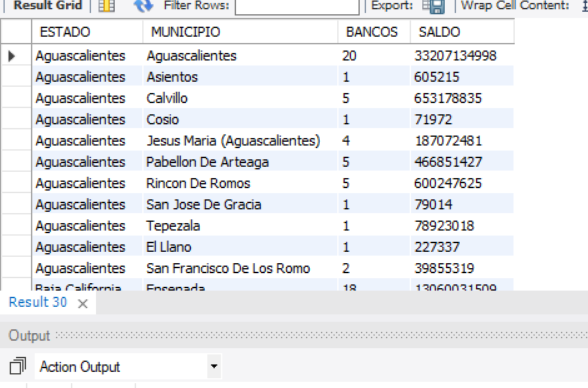
GROUP BY

CVE\_MUNICIPIO,

CVE\_ESTADO

) AS S

ON S.CVE\_MUNICIPIO = C.CVE\_MUNICIPIO



3.4 Hacer la consulta anterior usando vistas

CREATE VIEW TBANCOS\_MAYO AS

(

SELECT

COUNT(CVE\_INSTITUCION) AS BANCOS,

CVE\_MUNICIPIO,

CVE\_ESTADO

from CUENTAS

WHERE CUENTAS\_TOTALES > 0

AND CVE\_PERIODO = 202005

GROUP BY

CVE\_MUNICIPIO,

CVE\_ESTADO

);

CREATE VIEW TSALDO\_MAYO AS

(

SELECT

CVE\_MUNICIPIO,

SUM(SALDO\_TRANS\_TOTAL) AS SALDO

from CAPTACION

WHERE CVE\_PERIODO = 202005

GROUP BY

CVE\_MUNICIPIO,

CVE\_ESTADO

);

SELECT

E.ESTADO,

M.MUNICIPIO,

BANCOS,

S.SALDO

FROM TBANCOS\_MAYO AS C

LEFT JOIN

MUNICIPIO AS M

ON M.CVE\_MUNICIPIO = C.CVE\_MUNICIPIO

LEFT JOIN

ESTADO AS E

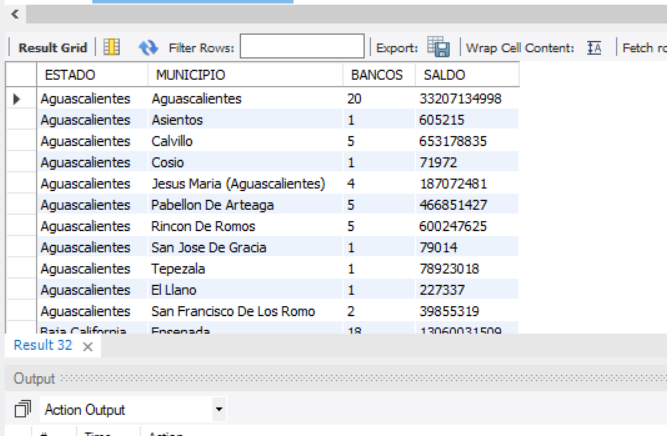
ON E.CVE\_ESTADO = C.CVE\_ESTADO

LEFT JOIN

TSALDO\_MAYO

AS S

ON S.CVE\_MUNICIPIO = C.CVE\_MUNICIPIO;



3.5 Con el ejercicio anterior, agregar el total de cuentas que había en mayo

CREATE VIEW TBANCOS\_MAYO AS

(

SELECT

COUNT(CVE\_INSTITUCION) AS BANCOS,

CVE\_MUNICIPIO,

CVE\_ESTADO

from CUENTAS

WHERE CUENTAS\_TOTALES > 0

AND CVE\_PERIODO = 202005

GROUP BY

CVE\_MUNICIPIO,

CVE\_ESTADO

);

CREATE VIEW TSALDO\_MAYO AS

(

SELECT

CVE\_MUNICIPIO,

SUM(SALDO\_TRANS\_TOTAL) AS SALDO

from CAPTACION

WHERE CVE\_PERIODO = 202005

GROUP BY

CVE\_MUNICIPIO,

CVE\_ESTADO

);

CREATE VIEW TCUENTAS\_MAYO AS

(

SELECT

CVE\_MUNICIPIO,

SUM(CUENTAS\_TOTALES) AS CUENTAS

from CUENTAS

WHERE CVE\_PERIODO = 202005

GROUP BY

CVE\_MUNICIPIO,

CVE\_ESTADO

);

SELECT

E.ESTADO,

M.MUNICIPIO,

B.BANCOS,

S.SALDO,

C.CUENTAS

FROM TBANCOS\_MAYO AS B

LEFT JOIN

MUNICIPIO AS M

ON M.CVE\_MUNICIPIO = B.CVE\_MUNICIPIO

LEFT JOIN

ESTADO AS E

ON E.CVE\_ESTADO = B.CVE\_ESTADO

LEFT JOIN

TSALDO\_MAYO

AS S

ON S.CVE\_MUNICIPIO = B.CVE\_MUNICIPIO

LEFT JOIN

TCUENTAS\_MAYO

AS C

ON C.CVE\_MUNICIPIO = B.CVE\_MUNICIPIO



# Sesión 4: Fundamentos de MongoDB

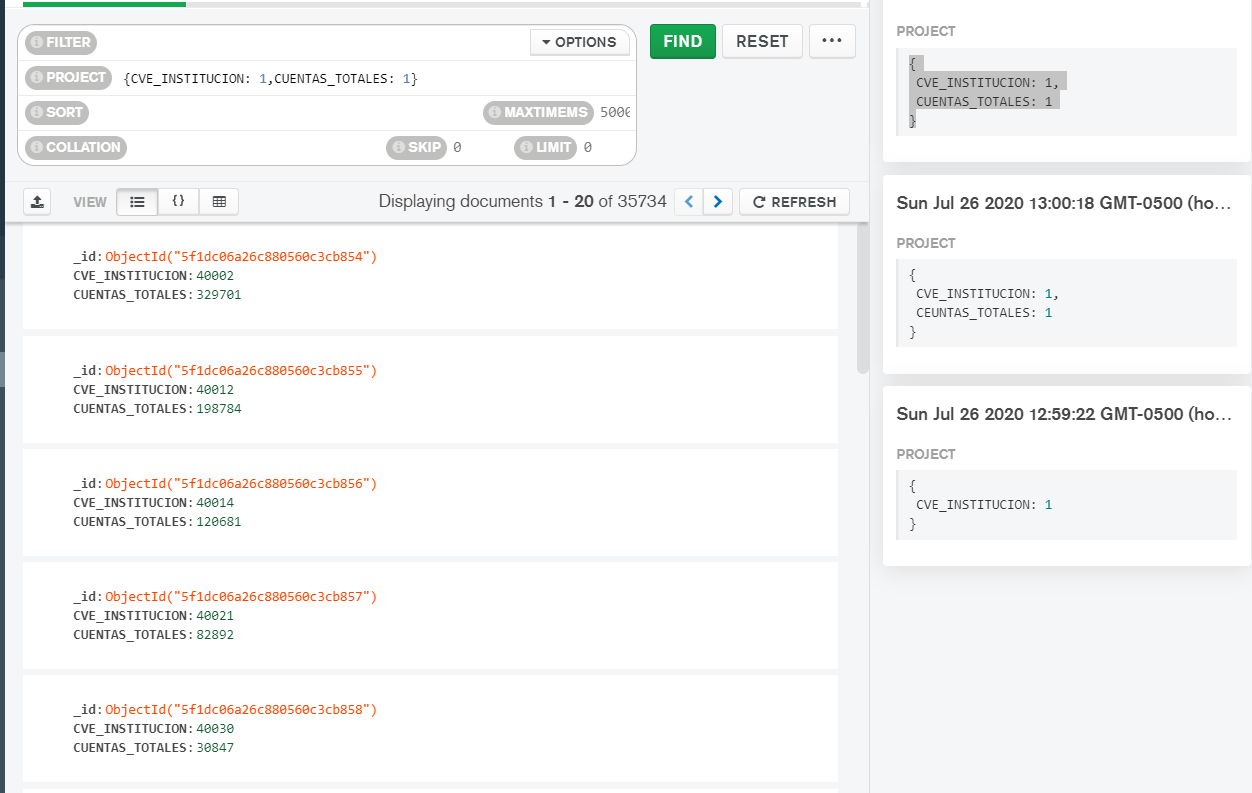
4.1 Mostrar los datos de cuentas totales y de instituciones

{

CVE\_INSTITUCION: 1,

CUENTAS\_TOTALES: 1

}



4.2 ¿Cuáles son los municipios del estado 4?

* **FILTER**

{

CVE\_ESTADO: 4

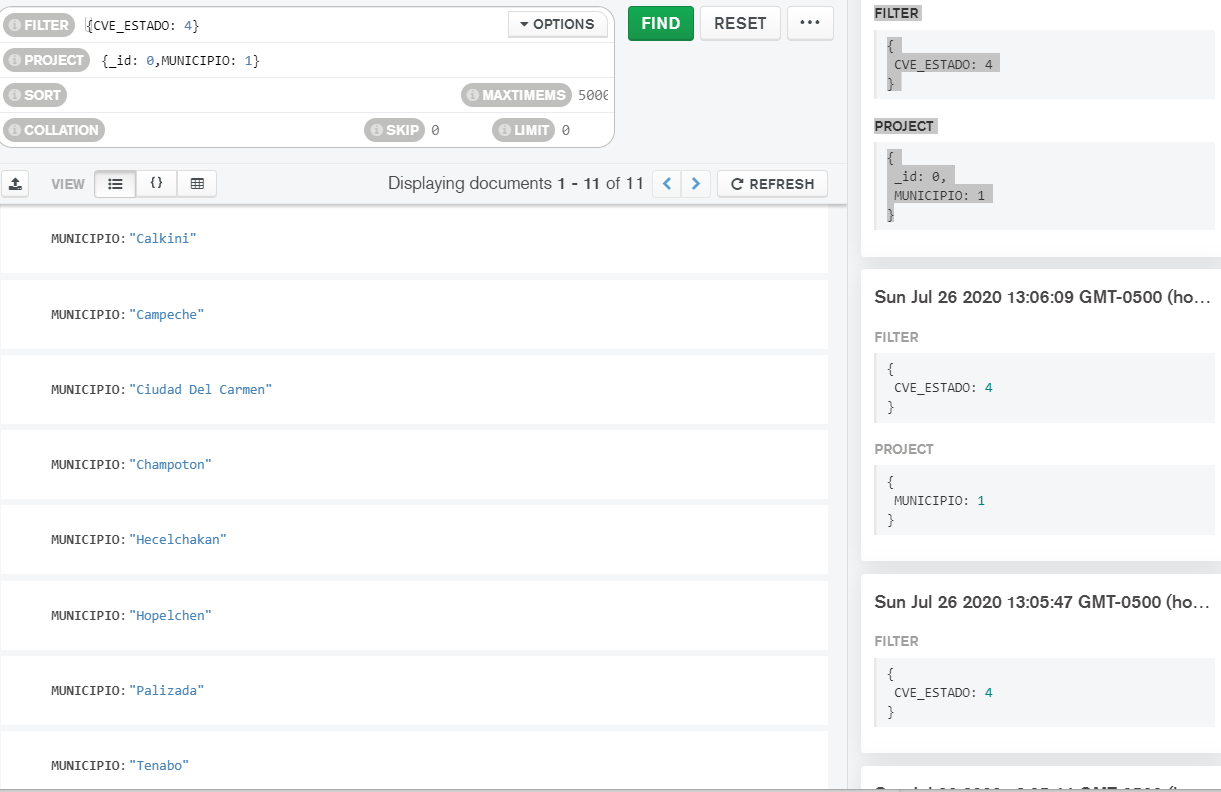
* }
* **PROJECT**

{

\_id: 0,

MUNICIPIO: 1

* }



4.3 ¿Cuáles son los bancos y que saldo total de cuentas transaccionales se tienen en mayo del 2020 (202005), que no tengan saldo de plazo fijo y que sean del municipio 2002?

* **FILTER**

{

CVE\_MUNICIPIO: 2002,

SALDO\_PLAZO\_FIJO: {

$gt: 0

},

CVE\_PERIODO: 202005

* }
* **PROJECT**

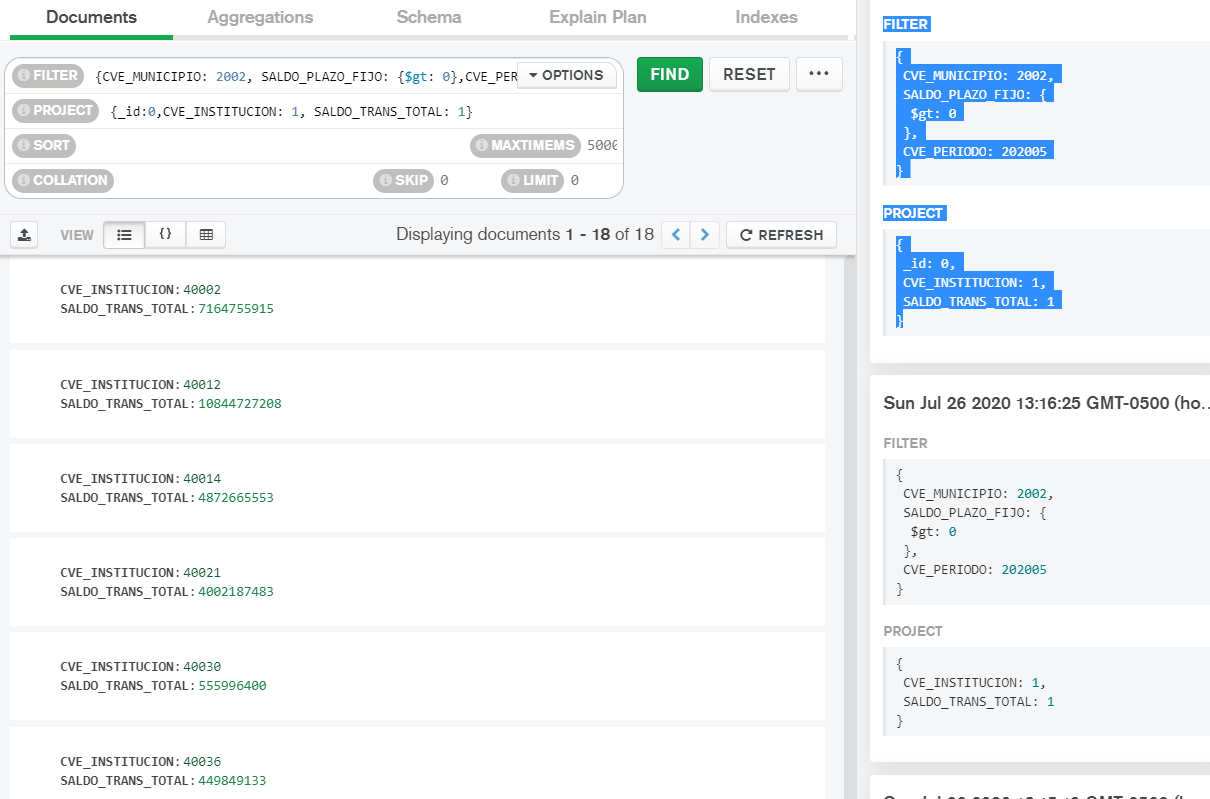
{

\_id: 0,

CVE\_INSTITUCION: 1,

SALDO\_TRANS\_TOTAL: 1

* }



4.4 ¿Cuál es el top 9 de bancos que tienen más cuentas nivel 2 en abril y que sean del municipio 09014?

**FILTER**

{

CVE\_PERIODO: 202004,

CVE\_MUNICIPIO: 9014

}

**PROJECT**

{

\_id: 0,

CVE\_INSTITUCION: 1

}

**SORT**

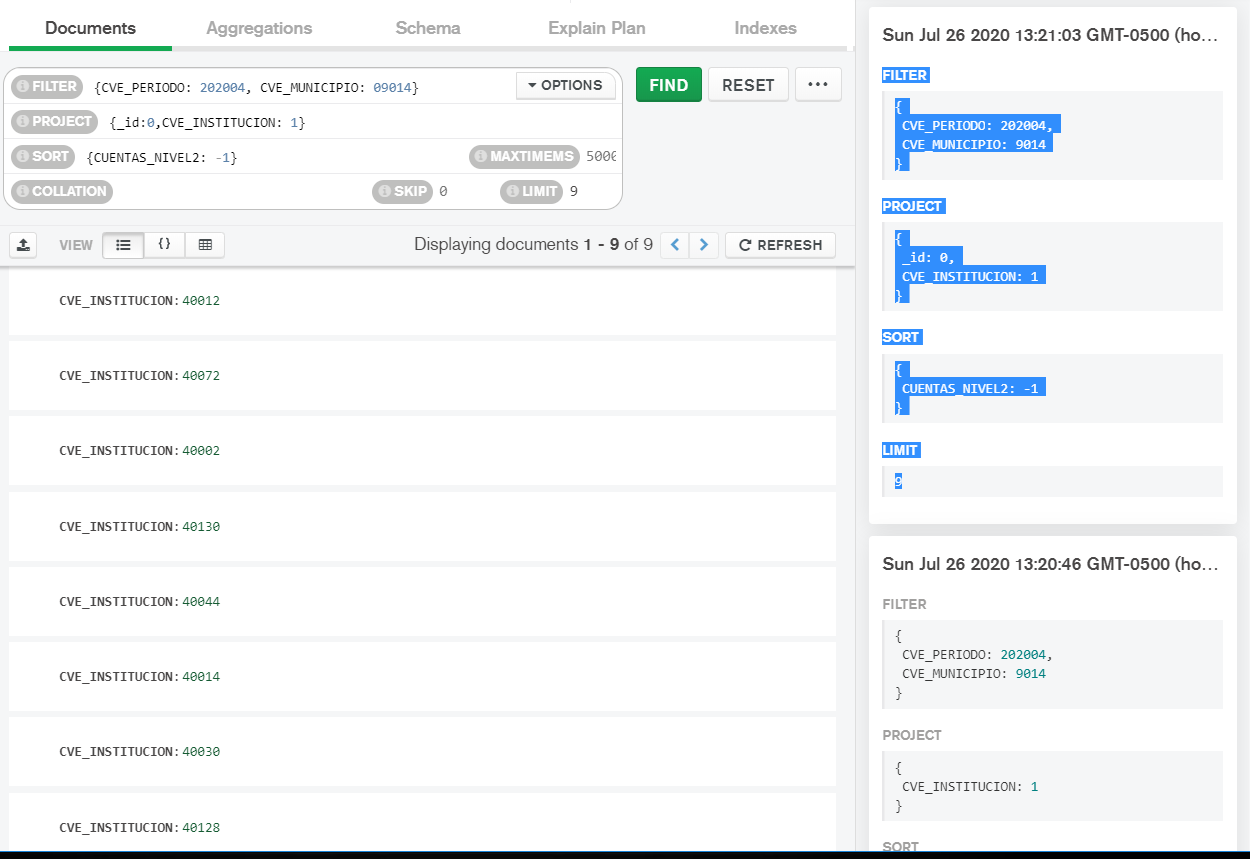
{

CUENTAS\_NIVEL2: -1

}

**LIMIT**

9



4.5 ¿Cuáles son los 20 municipios con menor saldo de plazo fijo del banco 40012 en enero del 2020?

* **FILTER**

{

CVE\_PERIODO: 202004,

CVE\_MUNICIPIO: 9014

* }
* **PROJECT**

{

\_id: 0,

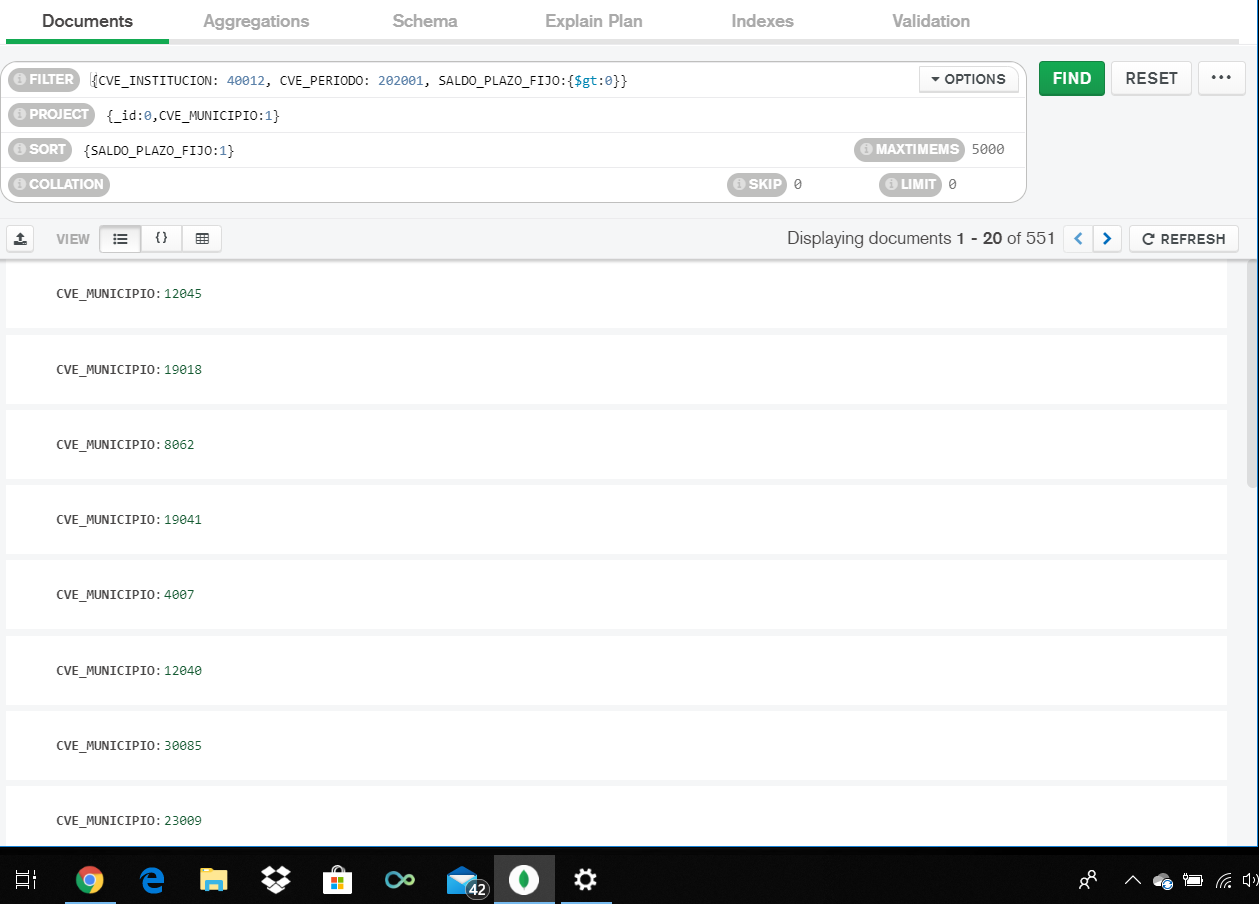
CVE\_INSTITUCION: 1

* }
* **SORT**

{

CUENTAS\_NIVEL2: -1

* }
* **LIMIT**
* 9



# Sesión 5: Consultas en Mongo DB

5.1 Obtener todos los municipios que empiezen con 'Ch'

* **FILTER**

{

MUNICIPIO: RegExp('^Ch', i)

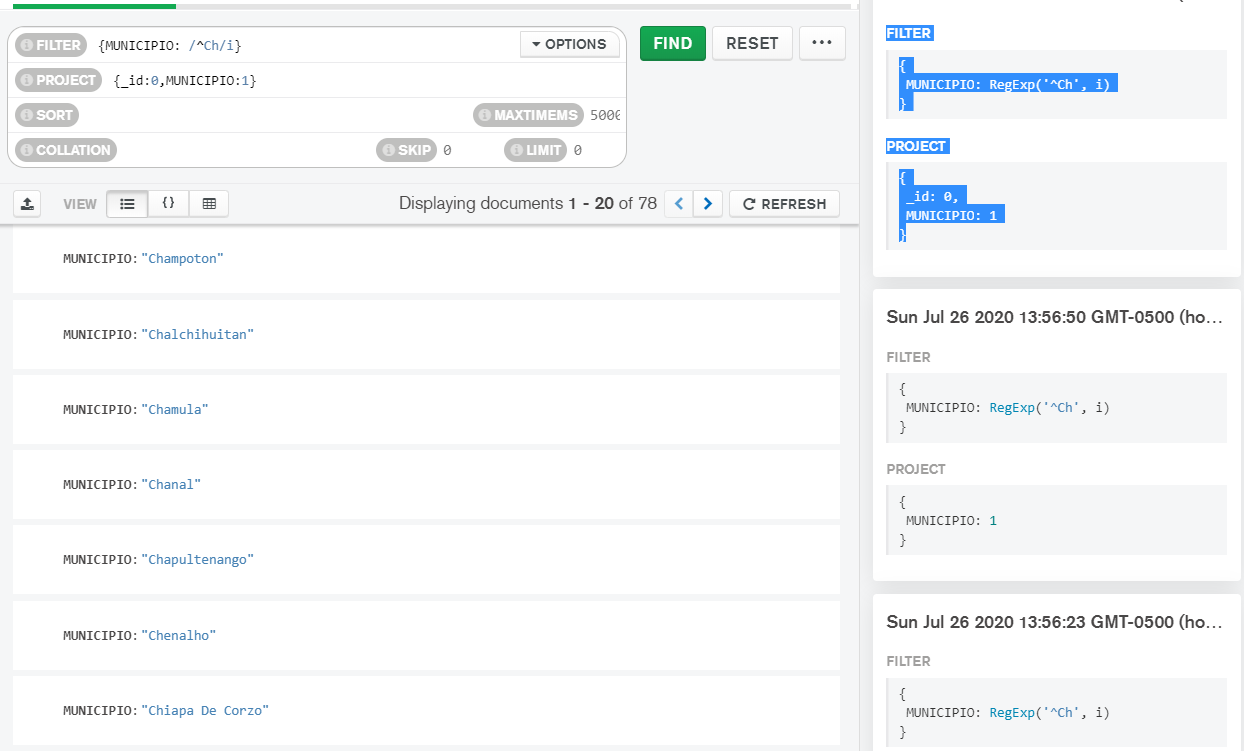
* }
* **PROJECT**

{

\_id: 0,

MUNICIPIO: 1

* }



5.2 Mostrar la institución, estado y municipio de aquellos bancos que tengan cuentas n1 o n2 en enero 2020

[{$match: {

$or: [

{CUNETAS\_NIVEL1: {$gt: 0}}

,

{CUENTAS\_NIVEL2: {$gt: 0}}

]

,

CVE\_PERIODO: 202005

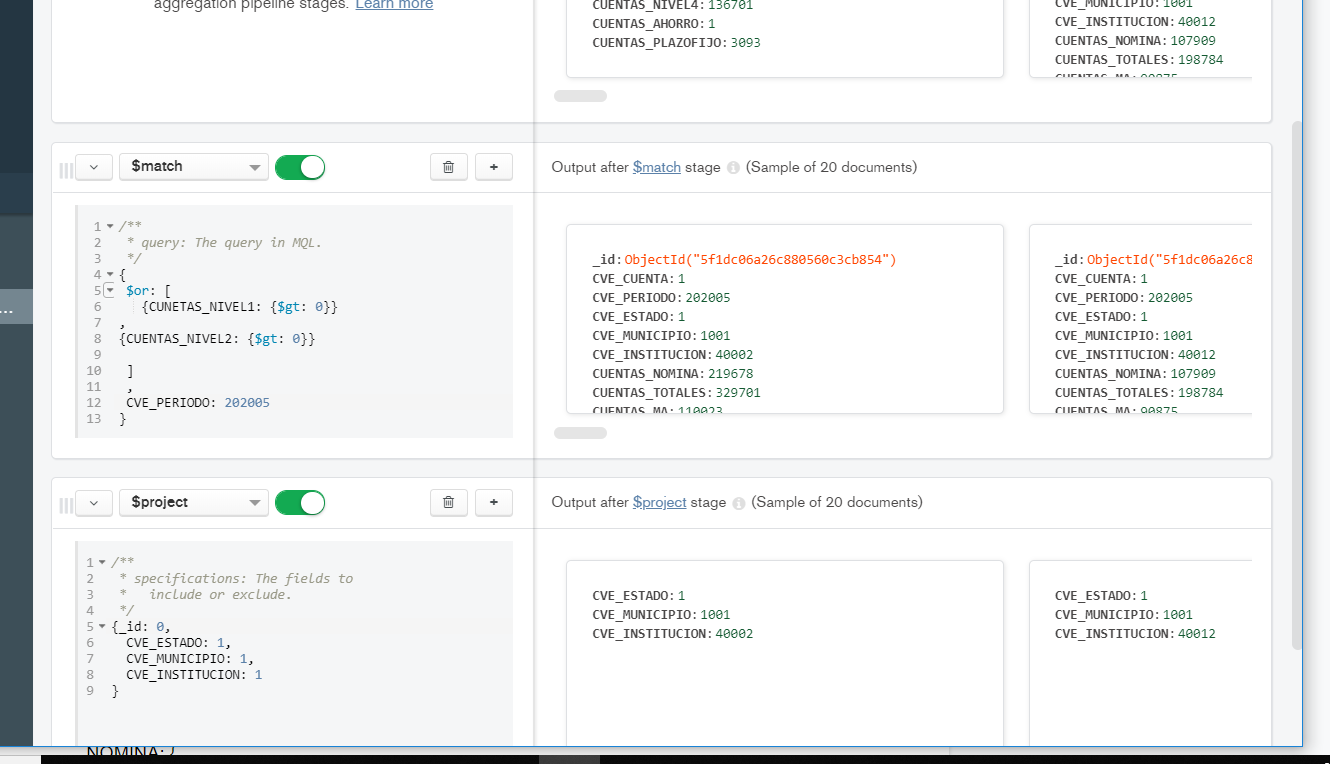
}}, {$project: {\_id: 0,

CVE\_ESTADO: 1,

CVE\_MUNICIPIO: 1,

CVE\_INSTITUCION: 1

}}]



5.3 Obtener el total de cuentas (Nomina, plazo fijo y transaccionales) en mayo del 2020 y el total general de estas

[{$match: {

CVE\_PERIODO: 202005

}}, {$group: {

\_id: null,

NOMINA: {

$sum: '$CUENTAS\_NOMINA'

},

CUENTAS\_TRANS: {

$sum: '$CUENTAS\_TOTALES'

},

CUENTAS\_PLAZO\_FIJO: {

$sum: '$CUENTAS\_PLAZOFIJO'

}

}}, {$project: {\_id:0,

NOMINA: 1,

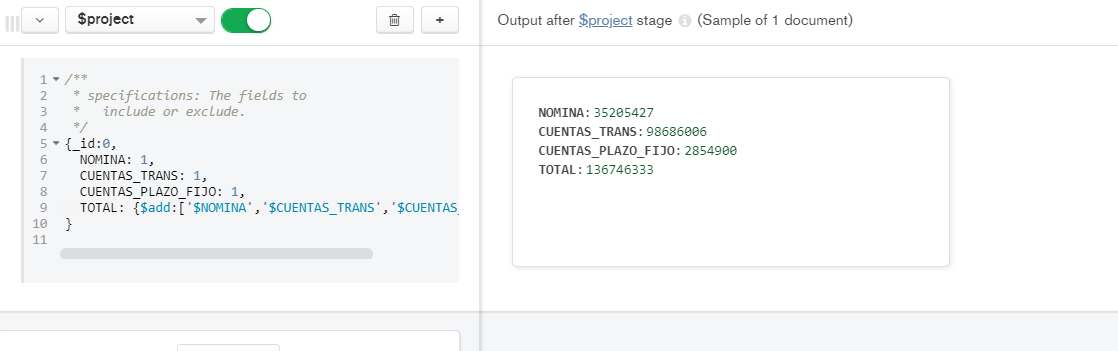
CUENTAS\_TRANS: 1,

CUENTAS\_PLAZO\_FIJO: 1,

TOTAL: {$add:['$NOMINA','$CUENTAS\_TRANS','$CUENTAS\_PLAZO\_FIJO']}

}

}]



5.4 ¿Cuál es el total de saldo de plazo fijo que hay por institución en cada periodo y en cuantos municipios se encuentran?

[{$match: {

SALDO\_PLAZO\_FIJO: {$gt:0}

}}, {$group: {

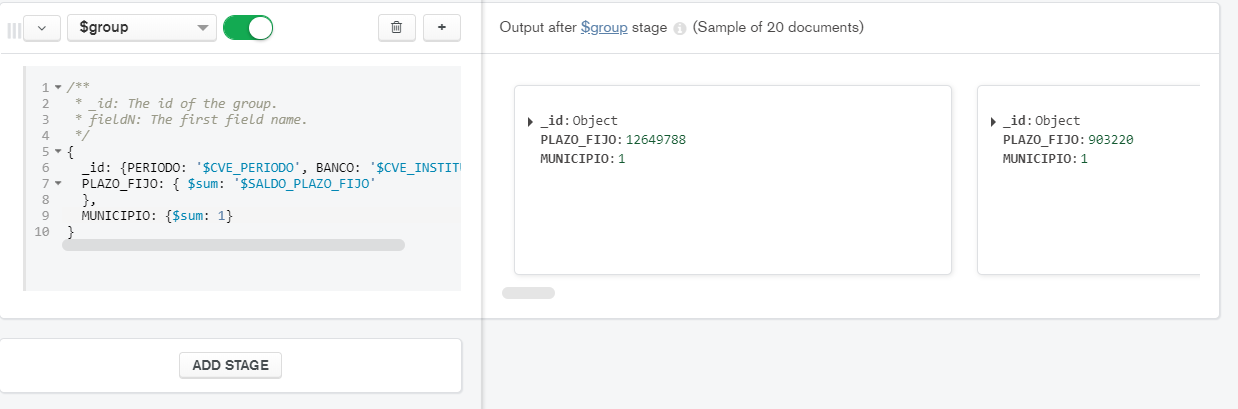
\_id: {PERIODO: '$CVE\_PERIODO', BANCO: '$CVE\_INSTITUCION'},

PLAZO\_FIJO: { $sum: '$SALDO\_PLAZO\_FIJO'

},

MUNICIPIO: {$sum: 1}

}}]



5.5 Obtener el saldo de las cuentas N3 que hay por estado de cada institución en enero del 2020

[{

$match: {

CVE\_PERIODO: 202001

}

}, {

$group: {

\_id: {

ESTADO: '$CVE\_ESTADO',

INSTITUCION: '$CVE\_INSTITUCION'

},

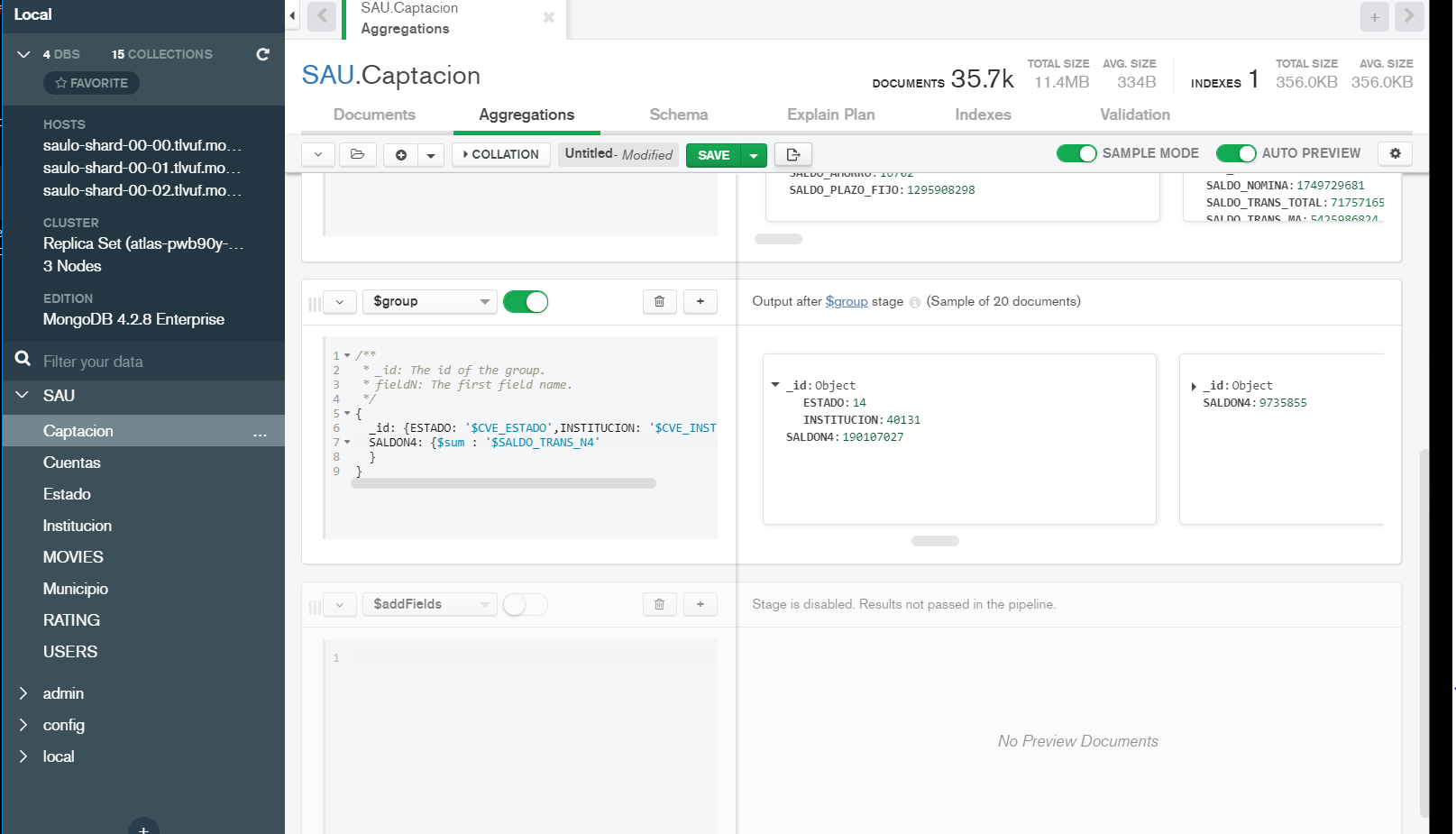
SALDON4: {

$sum: '$SALDO\_TRANS\_N4'

}

}

}]



# Sesión 6: Agregaciones

6.1 Obtener el promedio de cuentas por institución que hay por cada estado, en mayo

[{$match: {

CVE\_PERIODO: 202005,

CUENTAS\_NOMINA: {$gt:0}

}}, {$group: {

\_id: {ESTADO: '$CVE\_ESTADO', INSTITUCION: '$CVE\_INSTITUCION'},

NOMINA: {

$sum: '$CUENTAS\_NOMINA'

}

}}, {$addFields: {

CVE\_ESTADO: '$\_id.ESTADO'

}}, {$group: {

\_id: '$CVE\_ESTADO' ,

NOMINA: {

$sum: '$NOMINA'

},

BANCOS: {$sum: 1}

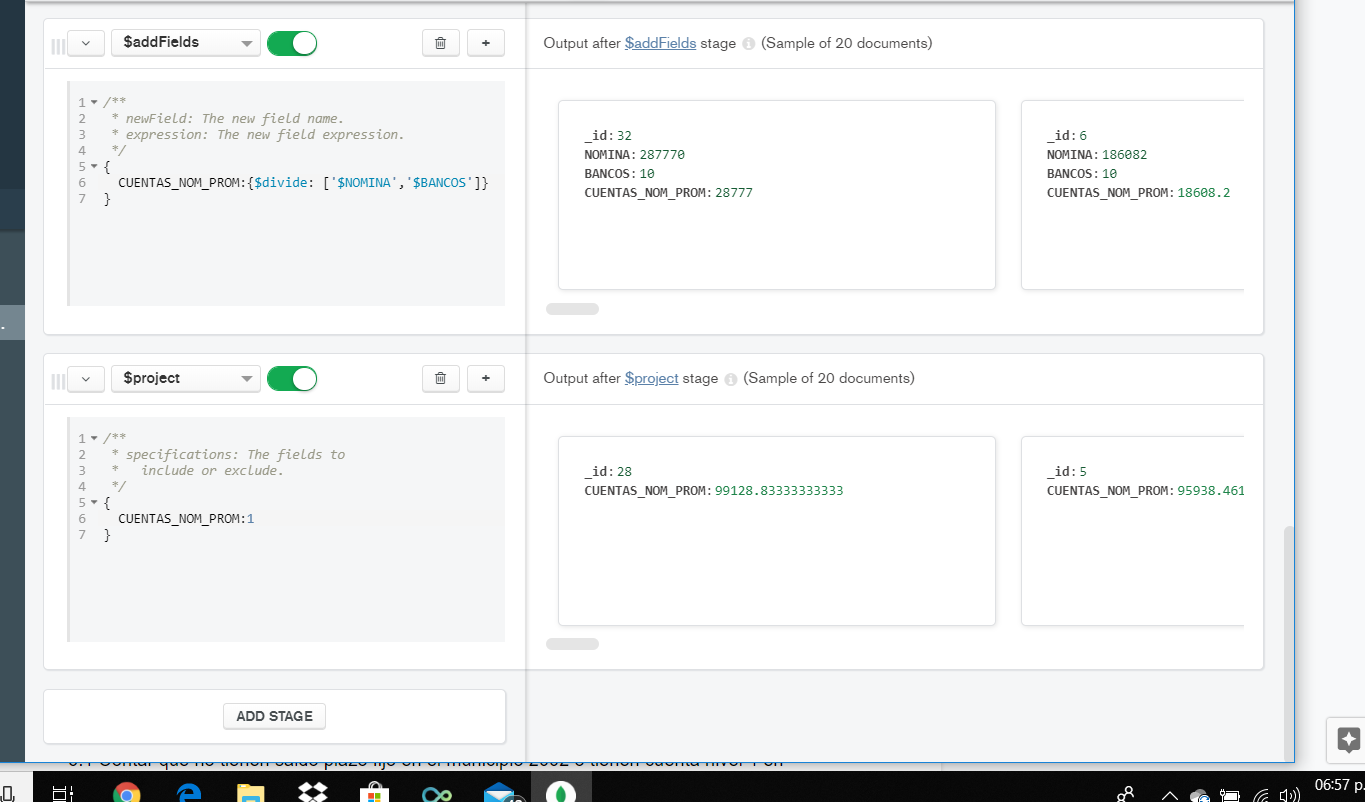
}}, {$addFields: {

CUENTAS\_NOM\_PROM:{$divide: ['$NOMINA','$BANCOS']}

}}, {$project: {

CUENTAS\_NOM\_PROM:1

}}]



6.2 ¿Cuál es el ranking de municipios por el saldo promedio de las instituciones de plazo fijo en febrero 2020?

[{$match: {

SALDO\_PLAZO\_FIJO: {$gt: 0},

CVE\_PERIODO: 202002

}}, {$group: {

\_id: '$CVE\_MUNICIPIO',

SALDO: {

$sum: '$SALDO\_PLAZO\_FIJO'

},

BANCOS: {$sum: 1}

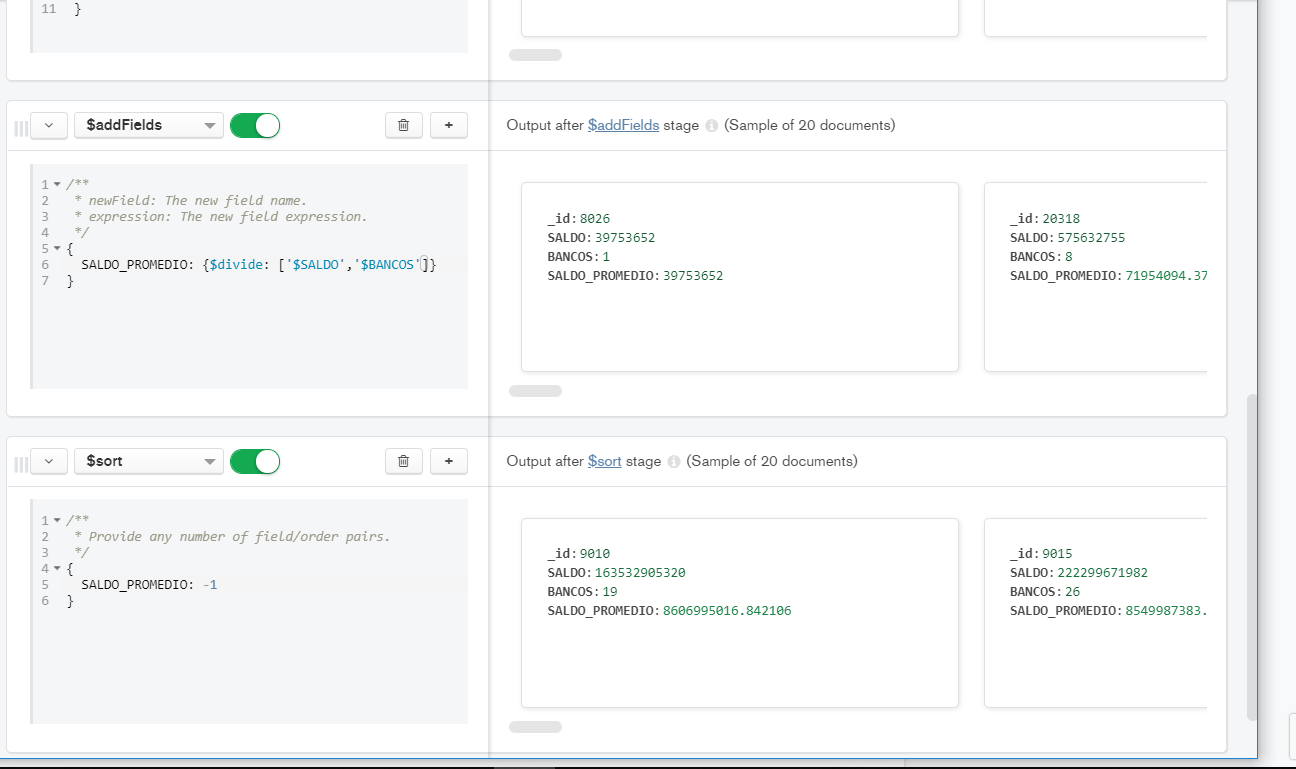
}}, {$addFields: {

SALDO\_PROMEDIO: {$divide: ['$SALDO','$BANCOS']}

}}, {$sort: {

SALDO\_PROMEDIO: -1

}}]



6.3 ¿Cuál es el total de cuentas totales (sin nomina ni plazo fijo) de Compartamos en mayo que se tienen por estado y en cuantos municipios esta?

[{$lookup: {

from: 'Institucion',

localField: 'CVE\_INSTITUCION',

foreignField: 'CVE\_INSTITUCION',

as: 'BANCO\_A'

}}, {$addFields: {

BANCO\_OBJETO: {$arrayElemAt:["$BANCO\_A",0]}

}}, {$addFields: {

BANCO: '$BANCO\_OBJETO.BANCO'

}}, {$match: {

BANCO: /compartamos/i,

CVE\_PERIODO: 202005

}}, {$group: {

\_id: '$CVE\_ESTADO',

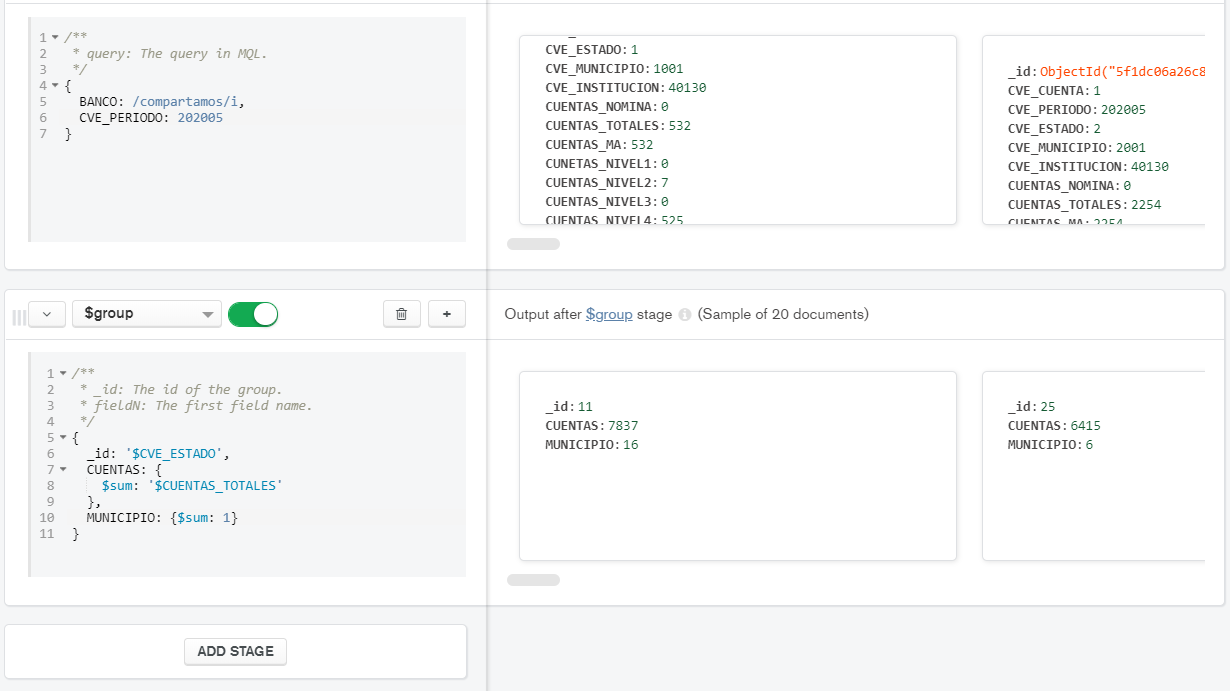
CUENTAS: {

$sum: '$CUENTAS\_TOTALES'

},

MUNICIPIO: {$sum: 1}

}}]



6.4 Usando el ejercicio anterior, crear una vista y poner el nombre del estado

[{$lookup: {

from: 'Estado',

localField: '\_id',

foreignField: 'CVE\_ESTADO',

as: 'ARRE\_ESTADO'

}}, {$addFields: {

ESTADO\_OBJETO: {$arrayElemAt: ['$ARRE\_ESTADO',0]}

}}, {$addFields: {

ESTADO: '$ESTADO\_OBJETO.ESTADO'

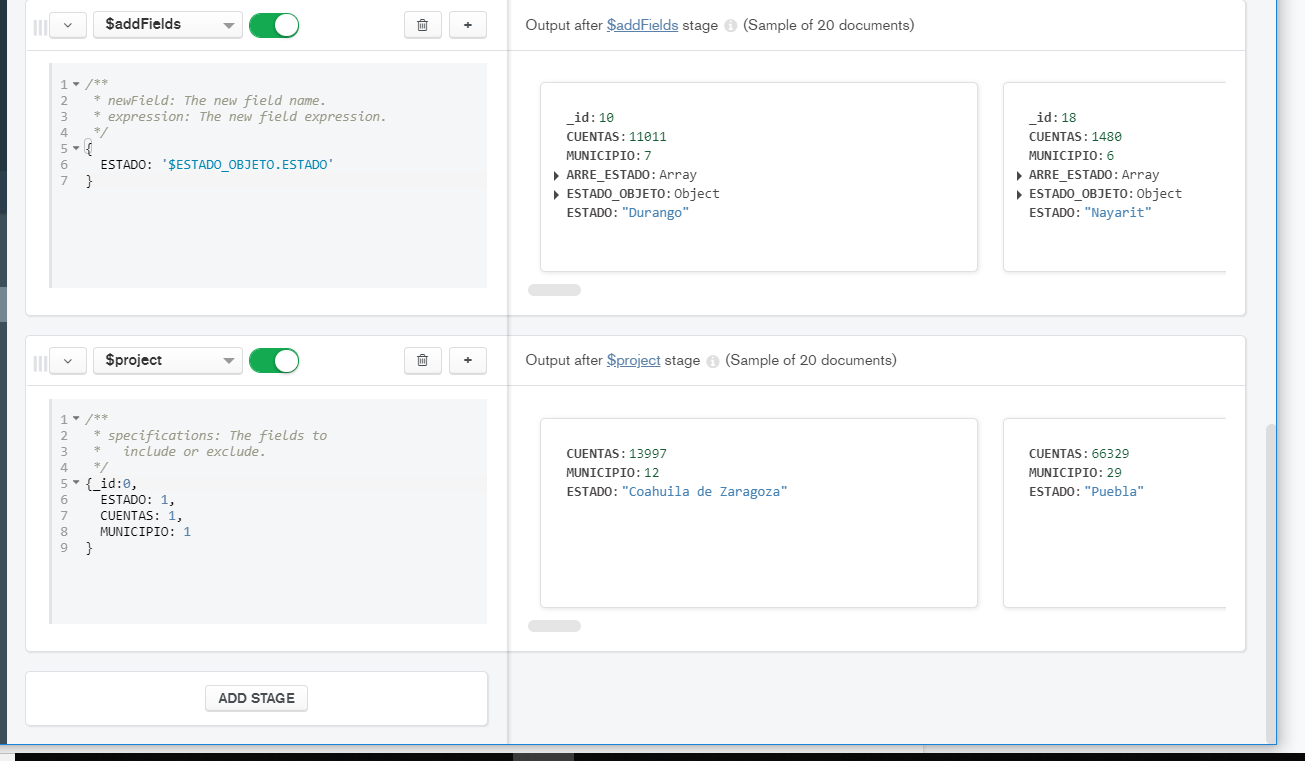
}}, {$project: {\_id:0,

ESTADO: 1,

CUENTAS: 1,

MUNICIPIO: 1

}}]



6.5 Hacer una vista del ejercicio 6.2 y poner el nombre del municipio y filtrar para mostrar los municipios que empiezan con ch

[{$lookup: {

from: 'Municipio',

localField: '\_id',

foreignField: 'CVE\_MUNICIPIO',

as: 'ARRE\_ESTADO'

}}, {$addFields: {

Municipo\_objeto:{$arrayElemAt:['$ARRE\_ESTADO',0]}

}}, {$addFields: {

MUNICIPIO: '$Municipo\_objeto.MUNICIPIO'

}}, {$match: {

MUNICIPIO: /^ch/i

}}, {$project: {

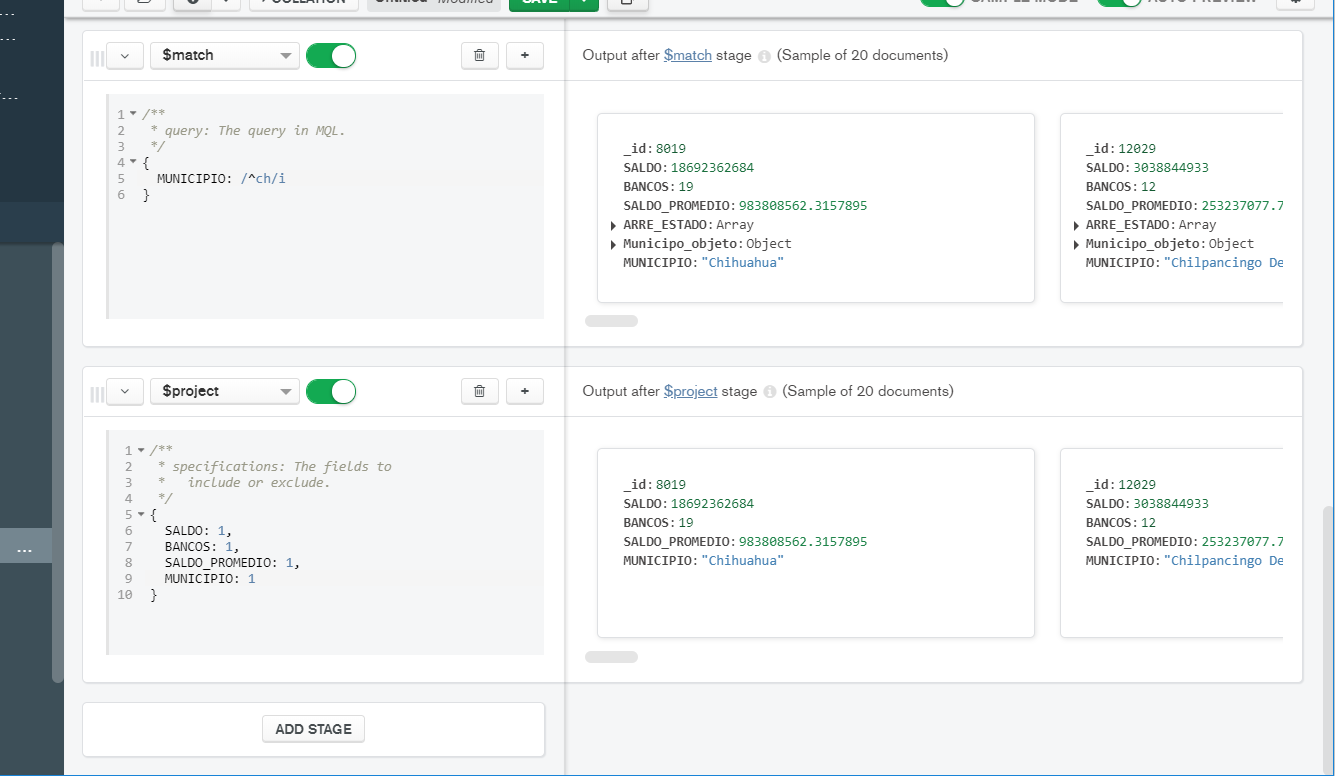
SALDO: 1,

BANCOS: 1,

SALDO\_PROMEDIO: 1,

MUNICIPIO: 1

}}]



# Conclusiones

1. SQL workbench se vuelve lenta en la carga de información, además de que se tiene que hacer las relaciones entre tablas y definir las llaves foráneas, pero una vez terminado este proceso, es muy fácil el hacer consultas que tengan varias relaciones
2. Mongo es una herramienta que puede funcionar muy bien para la carga de información y si se tiene todos los registros en una sola tabla es muy fácil poder consultar, sin embargo, tener diferentes tablas se vuelve complicado el hacer el lookup entre ellas, debido a que se generan objetos con arreglos de datos, por lo que se tienen que hacer más procesos para convertir estos en campos de la tabla
3. Debido a que esta información de la banca, ya está estructurada para ser usada como una base de datos relacional, optaría por hacer todos los procesos en SQL