Методы сбора, хранения, обработки и анализа данных

Лекция 16

СУБД MongoDb – агрегация

Фреймворк агрегации

- Специальный фреймворк для формирования последовательной обработки значений
- Используется для создания последовательного пайплайна
- Используется aggregate()

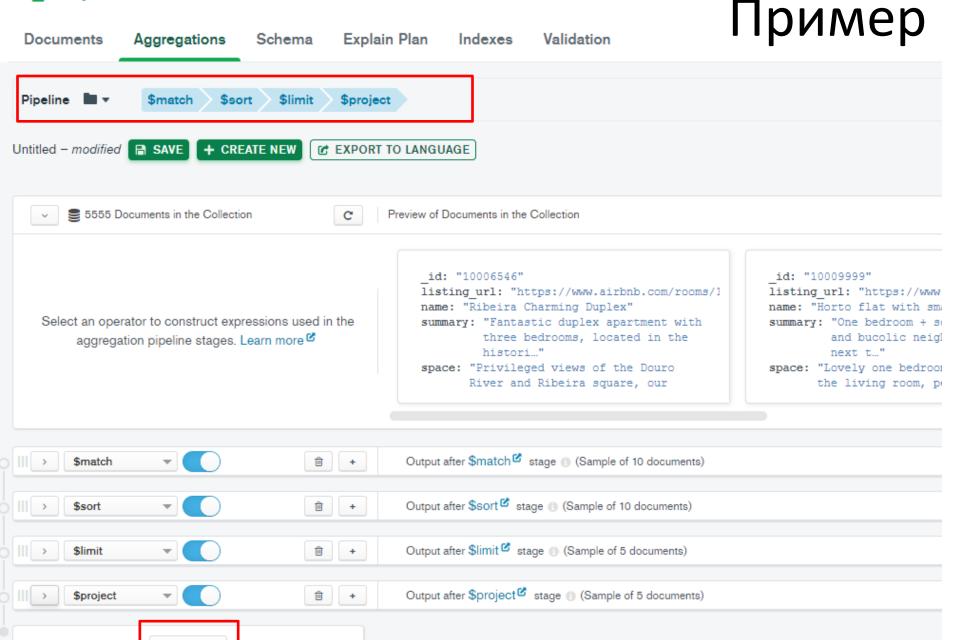
Агрегация

- Пайплайн состоит из стадий
- Классы стадий:
 - Запрос
 - Изменение данных
 - Join
 - Подсчет итогов
 - Ввод/вывод

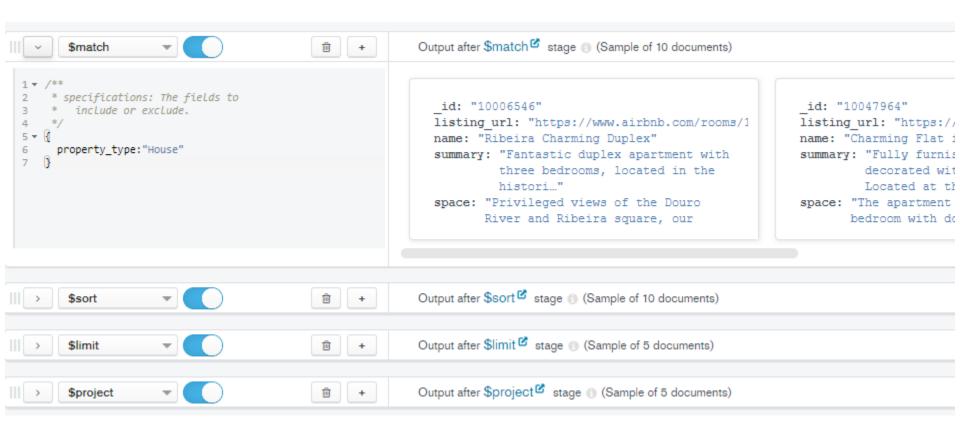
Стадии

Query	Mutate	Summarise/ Itemise	Join	Input/Output
\$geoNear	\$addFields	\$bucket	\$graphLookup	\$documents
\$limit	\$densify	\$bucketAuto	\$lookup	\$merge
\$match	\$fill	\$count	\$unionWith	\$out
\$sample	\$project	\$facet		
\$search	\$redact	\$group		
\$searchMeta	\$replaceRoot	\$sortByCount		
\$skip	\$replaceWith	\$unwind		
\$sort	\$set			
	\$setWindowFields			
	\$unset			

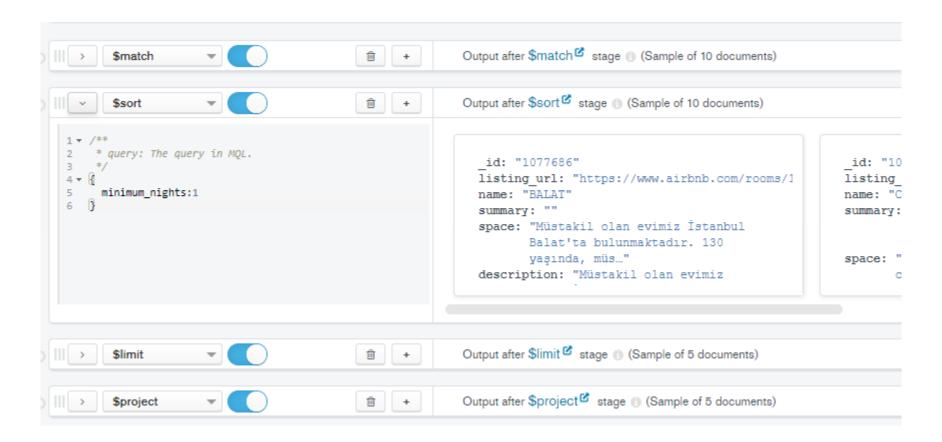
all subjects.reviews



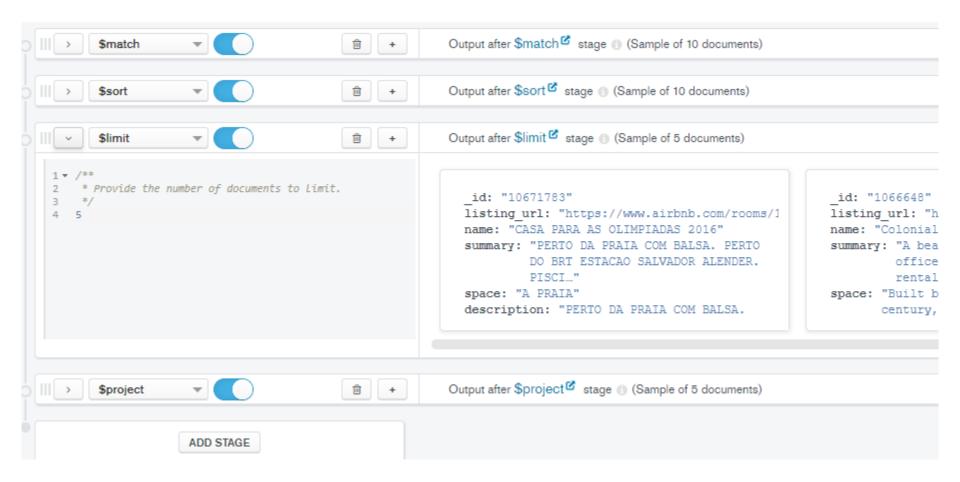
Условие



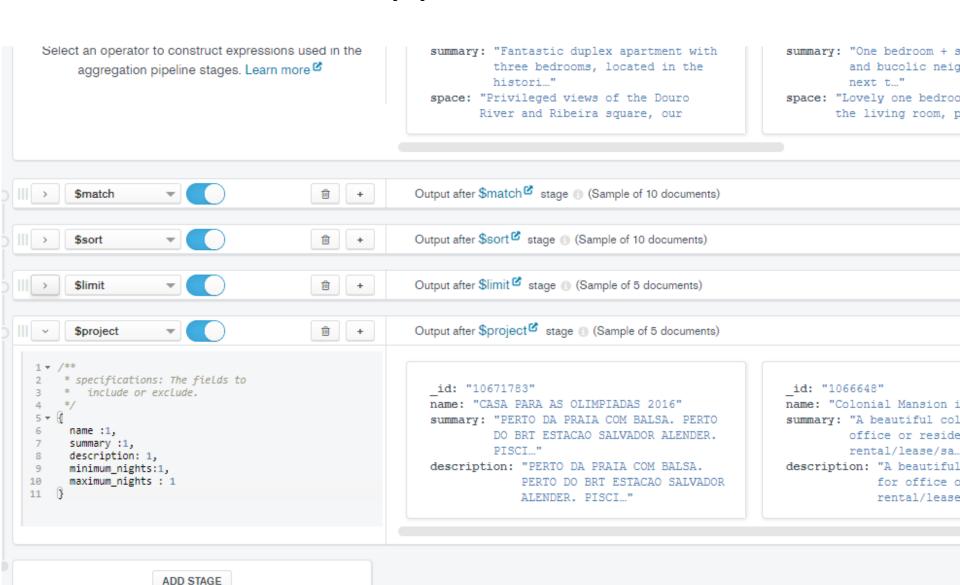
Сортировка



Лимит



Выводимые поля



Результат

```
id: "10671783"
name: "CASA PARA AS OLIMPIADAS 2016"
summary: "PERTO DA PRAIA COM BALSA, PERTO DO BRT ESTACAO SALVADOR ALENDER, PISCI..."
description: "PERTO DA PRAIA COM BALSA, PERTO DO BRT ESTACAO SALVADOR ALENDER, PISCI ..."
minimum nights: "1"
maximum_nights: "1125"
id: "1066648"
name: "Colonial Mansion in Santa Teresa"
summary: "A beautiful colonial mansion for office or residential rental/lease/sa..."
description: "A beautiful colonial mansion for office or residential rental/lease/sa..."
minimum nights: "1"
maximum nights: "1125"
id: "1077686"
name: "BALAT"
summary: ""
description: "Müstakil olan evimiz İstanbul Balat'ta bulunmaktadır. 130 yaşında, müs..."
minimum nights: "1"
maximum nights: "1125"
id: "1073607"
name: "Bright Inner West Terrace Value+"
summary: ""
description: "Come and stay in wonderful Petersham! This amazing and large terrace h ... "
minimum nights: "1"
maximum nights: "1125"
id: "10324377"
name: "Suíte em local tranquilo e seguro"
```

```
Explain Export Run Less Options 

MAX TIME MS 60000

Showing 1 – 5 count results 

VIEW 

(1)
```

Запрос

ص

Export Pipeline To Language

My Pipeline:

```
1 - []{
         $match: {
             property_type: 'House'
5 * }, {
         $sort: {
             minimum_nights: 1
 9 + }, {
         $limit: 5
10
11 7 }, {
        $project: {
12 ×
13
            name: 1,
            summary: 1,
14
15
            description: 1,
            minimum_nights: 1,
16
            maximum_nights: 1
17
```

Export Pipeline To: PYTHON 3

```
1 - [
                                                                ص
             '$match': {
                 'property_type': 'House'
             '$sort': {
                 'minimum_nights': 1
10 ×
         }, {
11
             '$limit': 5
12 -
             '$project': {
13 ▼
                 'name': 1,
                 'summary': 1,
15
                 'description': 1,
                 'minimum nights': 1,
                 'maximum_nights': 1
```

- ☐ Include Import Statements
 - ☐ Include Driver Syntax

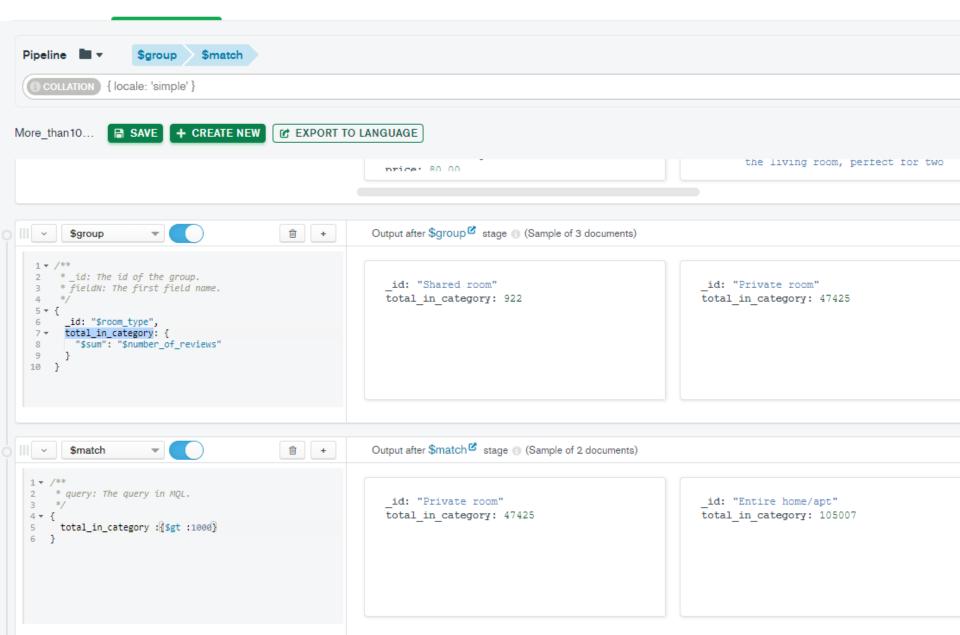
CLOSE

Запрос и результат

```
> db.reviews.aggregate([{
      $match: {
         property type: 'House'
 }, {
      $sort: {
         minimum nights: 1
 }, {
      $limit: 5
 }, {
      $project: {
          name: 1,
          summary: 1,
          description: 1,
         minimum nights: 1,
         maximum nights: 1
 }])
```

```
id: '10671783',
 name: 'CASA PARA AS OLIMPIADAS 2016',
 summary: 'PERTO DA PRAIA COM BALSA. PERTO DO BRT ESTACAO SALVADOR ALENDER. PISCINA SINUCA COND
 description: 'PERTO DA PRAIA COM BALSA. PERTO DO BRT ESTACAO SALVADOR ALENDER. PISCINA SINUCA
 minimum nights: '1',
 maximum nights: '1125' }
{ id: '1066648',
 name: 'Colonial Mansion in Santa Teresa',
 summary: 'A beautiful colonial mansion for office or residential rental/lease/sale - set in a
 description: 'A beautiful colonial mansion for office or residential rental/lease/sale - set i
 minimum nights: '1',
 maximum nights: '1125' }
{ id: '1077686',
 name: 'BALAT',
 description: 'Müstakil olan evimiz İstanbul Balat\'ta bulunmaktadır. 130 yaşında, müstakil bir
 minimum nights: '1',
{ id: '1073607',
 name: 'Bright Inner West Terrace Value+',
 description: 'Come and stay in wonderful Petersham! This amazing and large terrace house is lo
 minimum nights: '1',
 maximum nights: '1125' }
{ id: '10324377',
 name: 'Suite em local tranquilo e seguro',
 summary: 'Casa próxima à Barra em Jacarepaguá, com piscina, churrasqueira e internet. Facilida
 description: 'Casa próxima à Barra em Jacarepaguá, com piscina, churrasqueira e internet. Faci
 minimum nights: '1',
 maximum nights: '1125' }
```

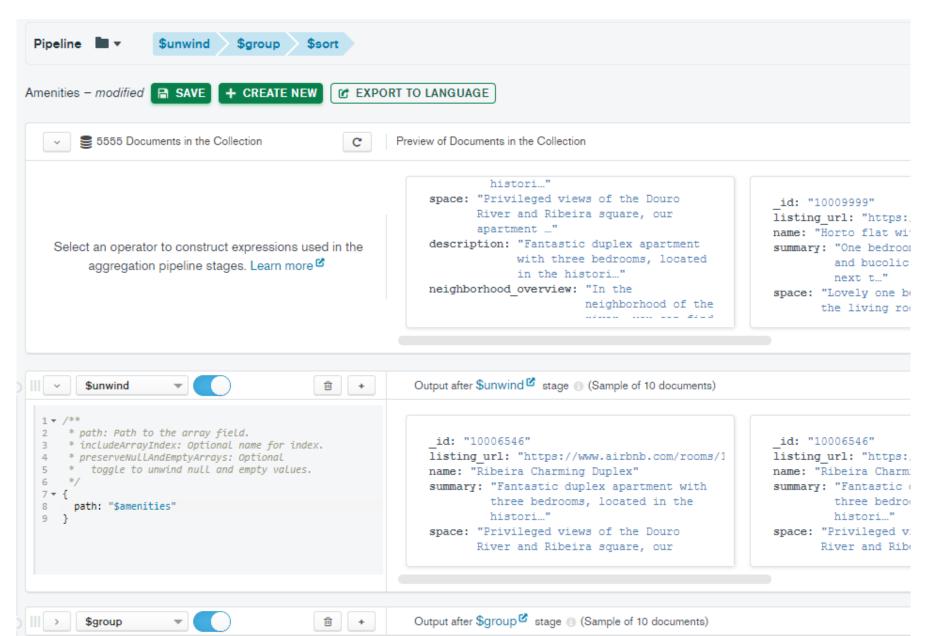
Пример с группировкой



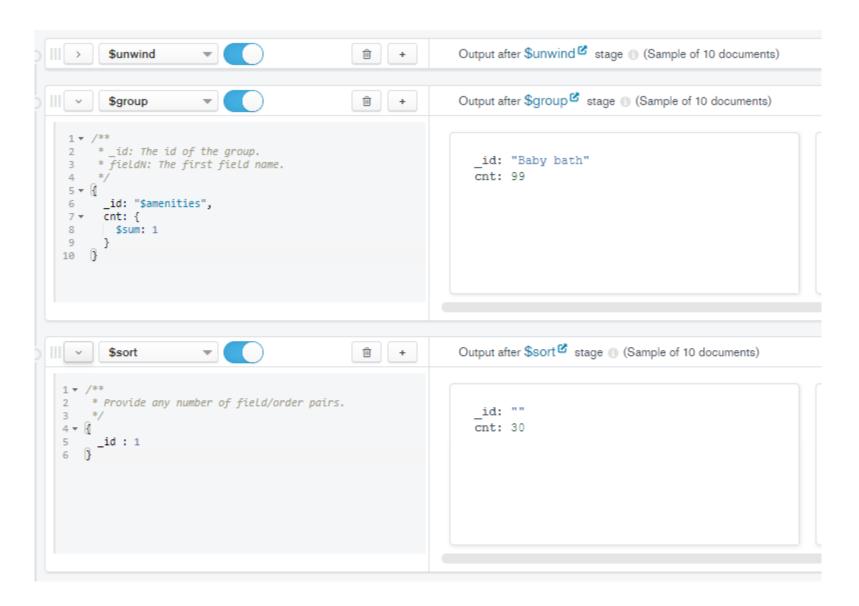
Пример с группировкой

```
> db.reviews.aggregate([{
     $group: {
         id: '$room type',
         total in category: {
             $sum: '$number of reviews'
 }, {
     $match: {
         total in category: {
             $gt: 1000
 }])
{ { id: 'Entire home/apt', total in category: 105007 }
 { id: 'Private room', total in category: 47425 }
```

Стадия \$unwind



Стадия \$unwind

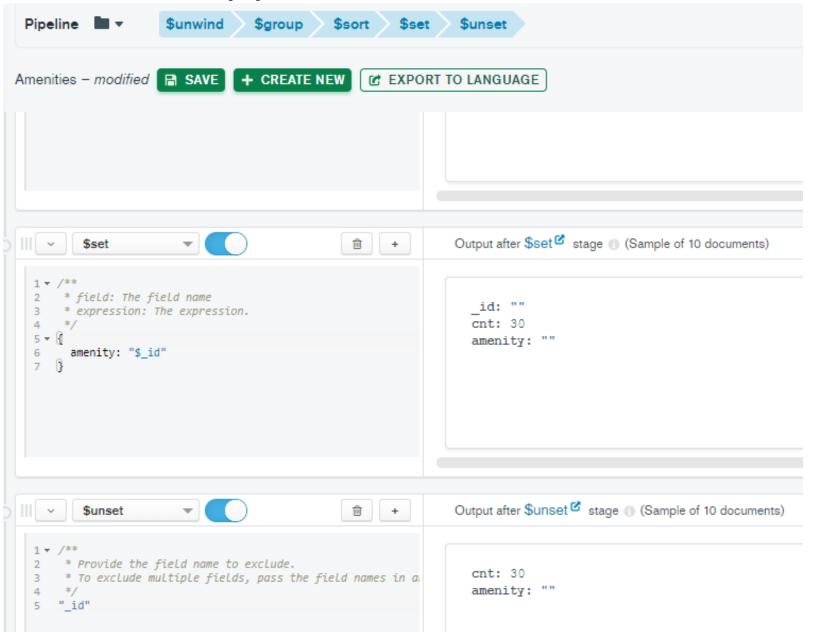


Стадия \$unwind - выполнение

```
> db.reviews.aggregate([{
     $unwind: {
         path: '$amenities'
 }, {
     $group: {
         id: '$amenities',
         cnt: {
              Ssum: 1
 }, {
     $sort: {
         id: 1
 }1)
```

```
< { id: '', cnt: 30 }
 { id: '24-hour check-in', cnt: 748 }
 { id: 'Accessible-height bed', cnt: 248 }
 { id: 'Accessible-height toilet', cnt: 193 }
 { id: 'Air conditioning', cnt: 3431 }
 { id: 'Air purifier', cnt: 3 }
 { id: 'Alfresco shower', cnt: 1 }
 { id: 'BBQ grill', cnt: 422 }
 { id: 'Baby bath', cnt: 99 }
 { id: 'Baby monitor', cnt: 19 }
 { id: 'Babysitter recommendations', cnt: 131 }
 { id: 'Balcony', cnt: 10 }
 { id: 'Bath towel', cnt: 25 }
 { id: 'Bathroom essentials', cnt: 25 }
 { id: 'Bathtub', cnt: 381 }
 { id: 'Bathtub with bath chair', cnt: 19 }
 { id: 'Beach chairs', cnt: 2 }
 { id: 'Beach essentials', cnt: 223 }
 { id: 'Beach view', cnt: 2 }
 { id: 'Beachfront', cnt: 191 }
 Type "it" for more
```

Стадии \$set и \$unset

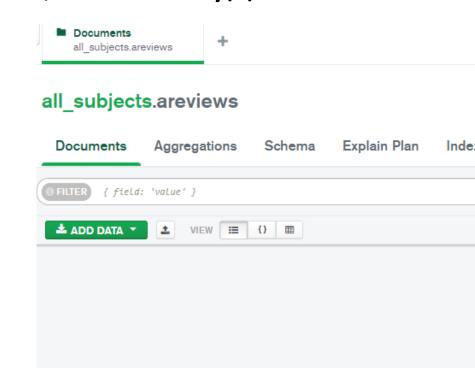


Стадия \$out

• Создаем пустую коллекцию, в которую загрузим данные

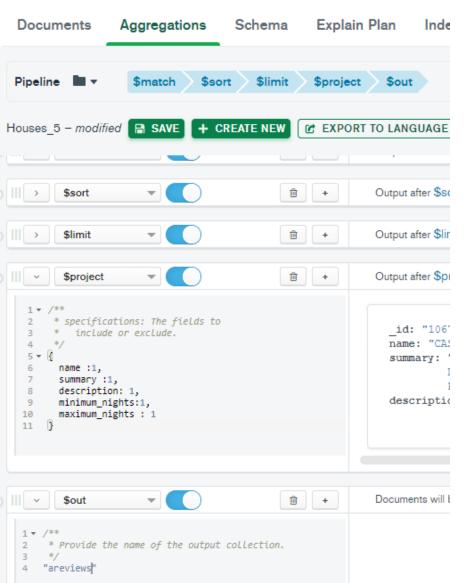
• Если коллекция непуста, то она будет

перезаписана



Стадия \$out

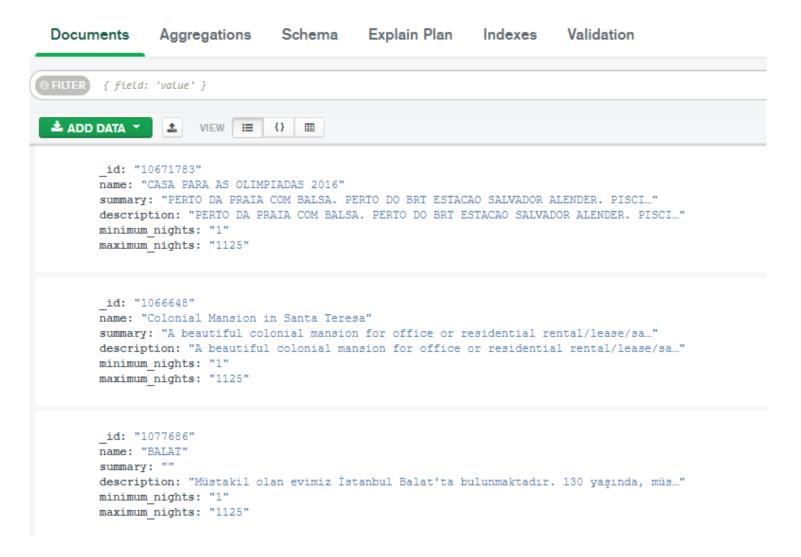
all_subjects.reviews



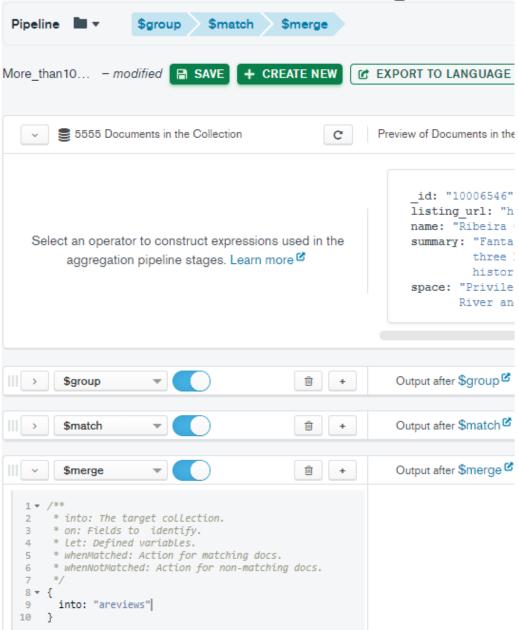
 Добавляем
 финальную стадию в пайплайн

Стадия \$out

all subjects.areviews



Стадия \$merge



Explain Plan Documents Aggregations Schema Indexes Validation FIND RESET Ð { field: 'value' } ▶ OPTIONS ▲ ADD DATA ▼ VIEW := Displaying documents 1 - 7 of 7 < > C REFRESH id: "1066648" name: "Colonial Mansion in Santa Teresa" summary: "A beautiful colonial mansion for office or residential rental/lease/sa..." description: "A beautiful colonial mansion for office or residential rental/lease/sa..." minimum nights: "1" maximum nights: "1125" id: "1077686" name: "BALAT" summary: "" description: "Müstakil olan evimiz İstanbul Balat'ta bulunmaktadır. 130 yaşında, müs..." minimum nights: "1" maximum nights: "1125" id: "1073607" name: "Bright Inner West Terrace Value+" description: "Come and stay in wonderful Petersham! This amazing and large terrace h ... " minimum nights: "1" maximum nights: "1125" id: "10324377" name: "Suíte em local tranquilo e seguro" summary: "Casa próxima à Barra em Jacarepaguá, com piscina, churrasqueira e inte..." description: "Casa próxima à Barra em Jacarepaguá, com piscina, churrasqueira e inte..." minimum nights: "1" maximum nights: "1125" id: "Entire home/apt" total in category: 105007 id: "Private room" total in category: 47425

Вопросы?