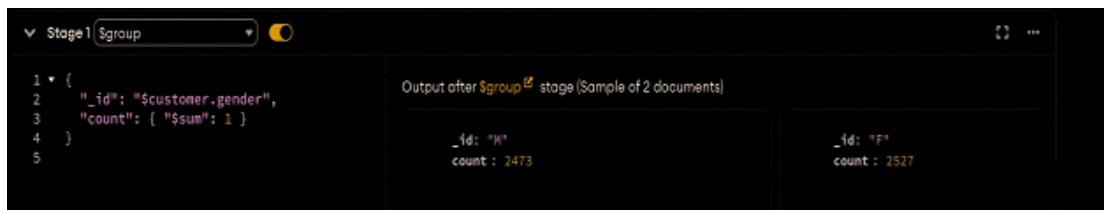**Tazhibayev**

Sultanbay

SE-2326

Assignment 4.2

1. How many customers males and females? Show in percentage of Males and Females among all records.

At this stage, the data is grouped by gender, and the number of records for each category is counted. Then, the percentage of males and females relative to the total number of records is calculated.

The total number of records is determined, and an array is created to store the count of men and women. After that, the percentage of each gender is calculated in relation to the overall dataset.

At this stage, the array is expanded, creating separate entries for men and women. This allows for a more detailed analysis of gender distribution.

The percentage is calculated, and unnecessary fields are removed.

The final result is presented in a structured format:

## 2. Get top 3 customers who most satisfied by store.

Customers are sorted in descending order based on their satisfaction level, with the most satisfied ones appearing first. This helps in identifying top-performing services or areas that meet customer expectations.

Only the top three entries are retained at this stage, allowing for a focused analysis of the highest-ranking results.
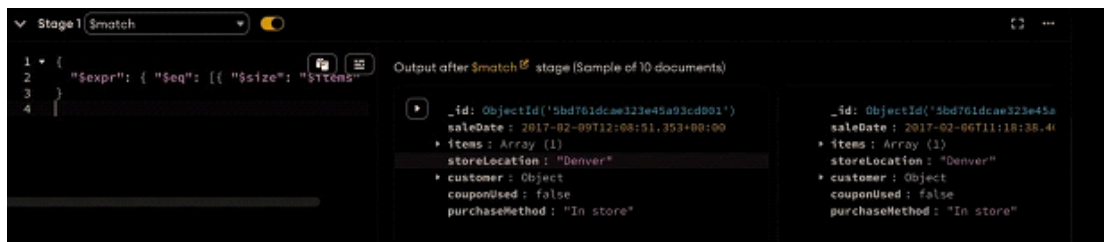

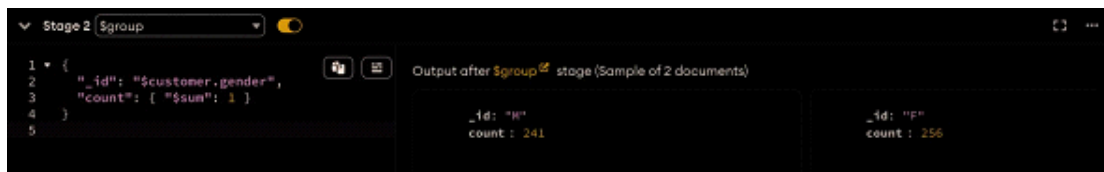*Only the email and satisfaction level are selected at this stage, ensuring a more concise dataset for further analysis.*
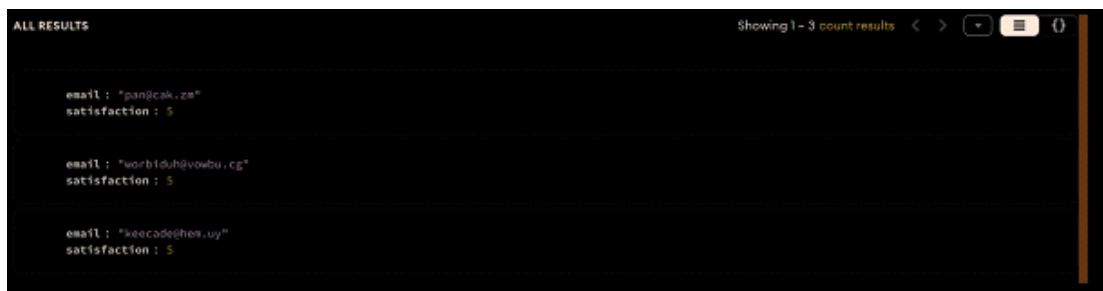
Result:

3. How many costumers bought only one item type (any quantity) from the store.

Records are filtered to include only those where the items array contains a single product, simplifying the dataset for analysis.

Data is grouped by gender, and the number of records in each category is counted, providing insights into gender distribution.
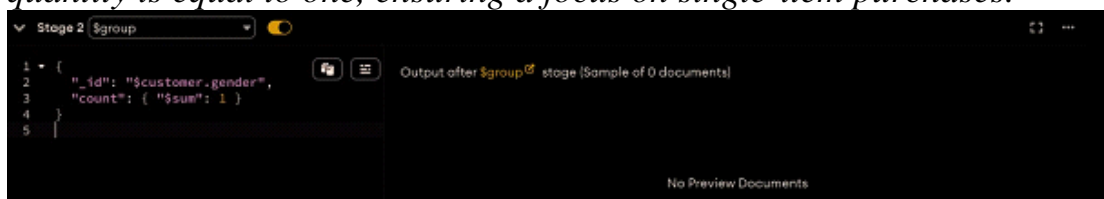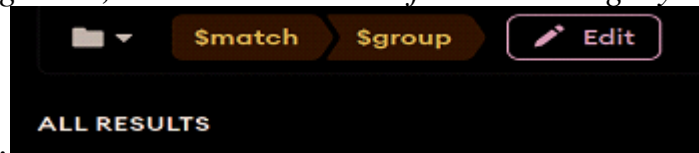
Result:

Result:

4. How many costumers bought only one item type (where quantity is one) from the store. How many of the males and females?

Records are selected where the items array contains only one product, and its quantity is equal to one, ensuring a focus on single-item purchases.

Data is grouped by gender, and the total count for each category is calculated to

analyze distribution.

5. What is the average age of customer where satisfaction is 1? Convert resulting value to an integer.
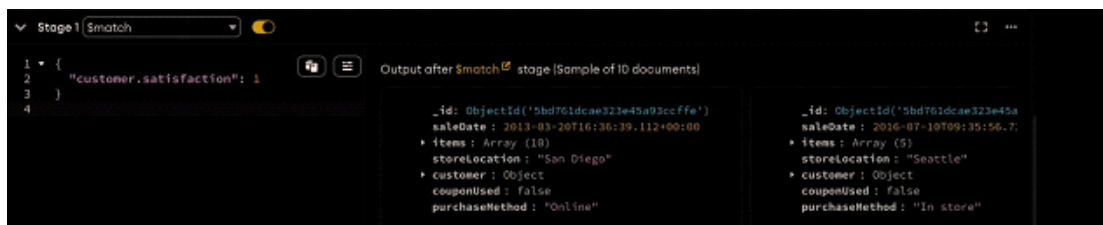
Only records where satisfaction equals 1 are filtered, focusing on the least satisfied customers.

The average age is calculated to analyze the overall age distribution of the dataset.

The result is rounded to ensure a more precise and readable value.

Result:

6. How much money spend 'pan@cak.zm'? Note about quantity. Convert resulting value to an integer.

Stage 1 $match

1 ▾ {
2   "customer.email": "pan@cak.zm"
3   }
4 |

Output after $match stage (Sample of 1 document)

_id: ObjectId('5bd761dcae323e45a93ccfee')
saleDate : 2014-11-11T02:13:51.893+08:00
▸ items : Array (10)
storeLocation : "London"
▸ customer : Object
couponUsed : false
purchaseMethod : "In store"

Purchases are filtered to include only those associated with the specified email, ensuring a targeted analysis.

Stage 2 $unwind

1 ▾ {
2   "path": "$items"
3   }
4

Output after $unwind stage (Sample of 10 documents)

_id: ObjectId('5bd761dcae323e45a93ccfee')   _id: ObjectId('5bd761dcae323e45a
saleDate : 2014-11-11T02:13:51.893+00:00   saleDate : 2014-11-11T02:13:51.8
▸ items : Object   ▸ items : Object
storeLocation : "London"   storeLocation : "London"
▸ customer : Object   ▸ customer : Object
couponUsed : false   couponUsed : false
purchaseMethod : "In store"   purchaseMethod : "In store"

The items array is expanded, creating separate entries for each individual product within the array.

Stage 3 $group

1 ▾ {
2   "_id": null,
3   "totalSpent": { "$sum": { "$multiply":
4   }
5

Output after $group stage (Sample of 1 document)

_id: null
totalSpent : 3492.14

The total amount is calculated by summing up the values of all relevant records.

The amount is rounded to the nearest whole number for better readability and accuracy.

Result:

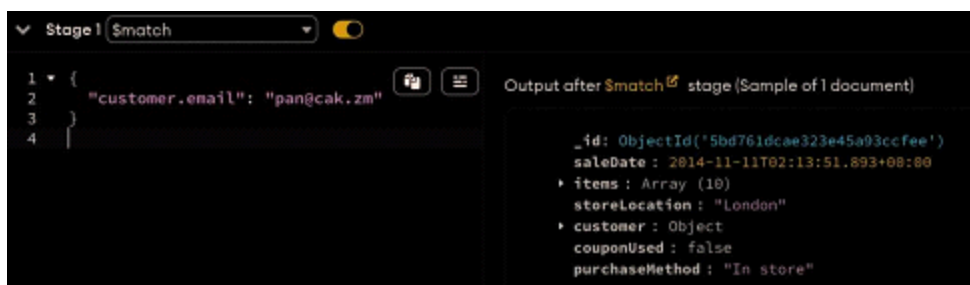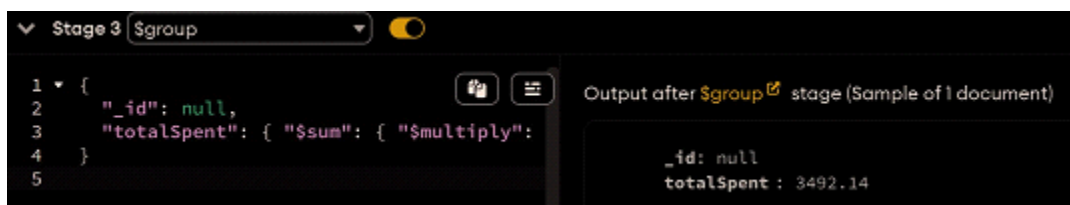# 7. What is the average age of males and females? Convert resulting value to an integer.

Data is grouped by gender, and the average age is calculated for each group to analyze age distribution across genders.

The average age is converted to an integer by rounding it to the nearest whole number.

Result:

8. What is the average money spend by males and females? Convert resulting value to an integer.

The items array is expanded, allowing each purchase to be processed as an individual entry for more detailed analysis.

Data is grouped by gender, and the average purchase amount is calculated for each group to analyze spending patterns.

The amount is converted to an integer by rounding it to the nearest whole number for simplicity and clarity.

ALL RESULTS   Showing 1 - 2 count results < > [▾] [≡] {}

    gender : "M"
    avgSpent : 361

    gender : "F"
    avgSpent : 359

Result:

9. What is the average money spend by using coupon? Convert resulting value to an integer. What average without using coupon? How many males and females used coupon?

Stage 1 `$unwind`

```
1 ▼ {
2     "path": "$items"
3   }
4
```

Output after $unwind stage (Sample of 10 documents)

```
_id: ObjectId('5bd761dcae323e45a93ccfe8')
saleDate : 2015-03-23T21:06:49.506+00:00
▸ items : Object
storeLocation : "Denver"
▸ customer : Object
couponUsed : true
purchaseMethod : "Online"
```

```
_id: ObjectId('5bd761dcae323e45a
saleDate : 2015-03-23T21:06:49.5
▸ items : Object
storeLocation : "Denver"
▸ customer : Object
couponUsed : true
purchaseMethod : "Online"
```

Stage 2 `$group`

```
1 ▼ {
2     "_id": "$couponUsed",
3     "avgSpent": { "$avg": { "$multiply": ["
4   }
5
```

Output after $group stage (Sample of 2 documents)

```
_id: true
avgSpent : 365.2230636055701919450035378246
```

```
_id: false
avgSpent : 360.121649247407287
```

Stage 3 `$project`

```
1 ▼ {
2     "_id": 0,
3     "couponUsed": "$_id",
4     "avgSpent": { "$toInt": "$avgSpent" }
5   }
6
```

Output after $project stage (Sample of 2 documents)

```
couponUsed : true
avgSpent : 365
```

```
couponUsed : false
avgSpent : 360
```

Stage 4 `$match`

```
1 ▼ {
2     "couponUsed": true
3   }
4
```

Output after $match stage (Sample of 1 document)

```
couponUsed : true
avgSpent : 365
```

Stage 5 `$group`

```
1 ▼ {
2     "_id": "$customer.gender",
3     "count": { "$sum": 1 }
4   }
5
```

Output after $group stage (Sample of 1 document)

```
_id: null
count : 1
```

1. Unwinding the array: The `$unwind` stage expands the `items` array, creating separate records for each item.

2. Grouping by coupon usage: The `$group` stage calculates the average spending (`avgSpent`) for customers who used a coupon and those who did not.

3. Rounding values: The `$project` stage converts `avgSpent` to an integer for better readability.

4. Filtering by coupon usage: The `$match` stage retains only records where `couponUsed` is `true`.

5. Grouping by gender: The final `$group` stage counts the number of records based on `customer_gender`.

ALL RESULTS

_id: null
count : 1

Result.

10. What is the average satisfaction where purchaseMethod online and for not online? Get only two numbers after the decimal point in the avgSatisfaction field.

Data is grouped by `purchaseMethod`, and the average satisfaction value is calculated for each method to analyze customer satisfaction across different purchasing channels.

The result is rounded to two decimal places for improved precision and readability.

```
    purchaseMethod : "Online"
    avgSatisfaction : 3.77

▶   purchaseMethod : "Phone"
    avgSatisfaction : 3.84

    purchaseMethod : "In store"
    avgSatisfaction : 3.8
```

Result.