Aggregation Homework Solutions

author: Hubert Łyczek
version: *1.0.0*

This document provides solutions to all the tasks.

Warning: This documents does not include any explanations.

Task1

Task

Count the number of people, divided into countries.

Solution

Query

```
db.people.aggregate([
    {
        $group: {
            _id: '$address.country',
            total: {$sum: 1}
        }
    }
])
```

Result

```
{
    "_id" : "France",
    "total" : 101238.0
}
{
    "_id" : "Poland",
    "total" : 98762.0
}
```

Task2

Task

What is the most popular address, and how many people live there?

Solution

Query

```
db.people.aggregate([
    {
        $group: {
            _id: '$address',
            total: {$sum: 1}
        }
    },
    {
        $sort: {'total': -1}
    },
    {
        $limit: 1
    }
])
```

Result

```
{
    "_id" : {
        "country" : "Poland",
        "city" : "Warsaw",
        "postalCode" : "02-495",
        "street" : "Aleje Jerozolimskie"
    },
    "total" : 5744.0
}
```

Task 3

Task

How many people from each country have visited a restaurant?

Solution

Query

```
db.people.aggregate([
    {
        $match: {
            payments: {
                $elemMatch: {name: 'restaurant'}
            }
        }
    },
    {
        $group: {
            _id: '$address.country',
            visits: {$sum: 1}
        }
    }
])
```

Result

```
{
    "_id" : "France",
    "visits" : 53126.0
}
{
    "_id" : "Poland",
    "visits" : 46971.0
}
```

Task 4

Task

Find the 3 people who have the most in their bank accounts.

Solution

Query

```
db.people.aggregate([
    {
        $unwind: '$wealth.bankAccounts'
    },
    {
        $group: {
            _id: '$_id',
            firstName: {$max: '$firstName'},
            lastName: {$max: '$lastName'},
            totalBalance: {
                $sum: '$wealth.bankAccounts.balance'
            }
        }
    },
    {
        $sort: { totalBalance: -1 }
    },
    {
        $limit: 3
    },
    {
        $project: {
            _id: 0,
            firstName: 1,
            lastName: 1,
            totalBalance: 1
        }
    }
])
```

Result

```
{
    "firstName" : "Mathilde",
    "lastName" : "Bonnet",
    "totalBalance" : 68544.28
}
{
    "firstName" : "Maxime",
    "lastName" : "Michel",
    "totalBalance" : 67416.83
}
{
    "firstName" : "Marie",
    "lastName" : "Petit",
    "totalBalance" : 66765.45
}
```

Task 5

Task

Count the number of all restaurant visits, the total amount spent, and the average amount per visit. All of this should be divided into countries.

Solution

Query

```
db.people.aggregate([
    {
        $match: {payments: {$elemMatch: {name: 'restaurant'}}}
    },
    {
        $project: {
            'address.country': 1,
            paymentsFiltered: {
                $filter: {
                    input: '$payments',
                    as: 'payment',
                    cond: {
                        $eq: ['$$payment.name', 'restaurant']
                    }
                }
            }
        }
    },
    {
        $unwind: '$paymentsFiltered'
    },
    {
        $group: {
            _id: '$address.country',
            totalVisits: {$sum: 1},
            totalAmount: {
                $sum: '$paymentsFiltered.amount'
            }
        }
    },
    {
        $project: {
            _id: 1,
            totalVisits: 1,
            totalAmount: 1,
            avgAmount: {
                $divide: ['$totalAmount', '$totalVisits']
            }
        }
    }
])
```

Result

```
{
    "_id" : "France",
    "totalVisits" : 78627.0,
```

```
    "totalAmount" : 2958352.16000001,
    "avgAmount" : 37.625143525761
}
{
    "_id" : "Poland",
    "totalVisits" : 66489.0,
    "totalAmount" : 2026963.35,
    "avgAmount" : 30.4856946261787
}
```

Task 6

Task

There is one country in which the average payment in a restaurant is the highest, and one in which the average payment in a restaurant is the lowest. How many times do people from the first country spend more than people from the second country?

Solution

You can extend the solution from the previous task, or write the result from the previous task to a new collection, and then perform another query on that collection.
This solution assumes, that the result from the previous task has been written to the collection *task5*, using the *$out* stage.

Query

```
db.task5.aggregate([
    {
        $group: {
            _id: null,
            max: {$max: '$avgAmount'},
            min: {$min: '$avgAmount'}
        }
    },
    {
        $project: {
            diff: {
                $divide: ['$max', '$min']
            }
        }
    }
])
```

Result

```
{
    "_id" : null,
    "diff" : 1.23419013367179
}
```

Task 7

Task

Count the number of people in each country according to age groups:

- **18-29**
- **30-39**
- **40-49**

Solution

Query

```
db.people.aggregate([
    {
        $project: {
            country: '$address.country',
            queryDate: {$literal: ISODate("2016-06-22T00:00:00.000Z")},
            birthDate: {
                $let: {
                    vars: {
                        epochStart: ISODate("1970-01-01T00:00:00.000Z")
                    },
                    in: {
                        $cond: {
                            if: {$lt: ['$birthDate', '$$epochStart']},
                            then: '$$epochStart',
                            else: '$birthDate'
                        }
                    }
                }
            }
        }
    },
    {
        $project: {
            country: 1,
            birthDate: 1,
            age: {
                $let: {
                    vars: {
                        baseAge: {
                            $subtract: [
                                {$year: '$queryDate'},
                                {$year: '$birthDate'}
                            ]
                        },
                        queryMonth: {$month: '$queryDate'},
                        queryDay: {$dayOfMonth: '$queryDate'},
                        birthMonth: {$month: '$birthDate'},
                        birthDay: {$dayOfMonth: '$birthDate'}
                    },
                    in: {
                        $cond: {
                            if: {
                                $gt: ['$$queryMonth', '$$birthMonth']
                            },
```

```
                            then: '$$baseAge',
                            else: {
                                $cond: {
                                    if: {
                                        $eq: ['$$queryMonth', '$$birthMonth']
                                    },
                                    then: {
                                        $cond: {
                                            if: {
                                                $gte: ['$$queryDay', '$$birthDay']
                                            },
                                            then: '$$baseAge',
                                            else: {
                                                $subtract: ['$$baseAge', 1]
                                            }
                                        }
                                    },
                                    else: {
                                        $subtract: ['$$baseAge', 1]
                                    }
                                }
                            }
                        }
                    }
                }
            }
        },
        {
            $project: {
                country: 1,
                range: {
                    $cond: {
                        if: {
                            $lte: ['$age', 29]
                        },
                        then: '18-29',
                        else: {
                            $cond:  {
                                if: {
                                    $gte: ['$age', 40]
                                },
                                then: '40-49',
                                else: '30-39'
                            }
                        }
                    }
                }
            }
        },
        {
            $group: {
                _id: {
                    country: '$country',
                    ageRange: '$range'
                },
                count: {$sum: 1}
            }
        }
```

```
])
```

Result


```
{
    "_id" : {
        "country" : "Poland",
        "ageRange" : "18-29"
    },
    "count" : 30742.0
}
{
    "_id" : {
        "country" : "France",
        "ageRange" : "18-29"
    },
    "count" : 31853.0
}
{
    "_id" : {
        "country" : "Poland",
        "ageRange" : "40-49"
    },
    "count" : 34800.0
}
{
    "_id" : {
        "country" : "France",
        "ageRange" : "40-49"
    },
    "count" : 35758.0
}
{
    "_id" : {
        "country" : "Poland",
        "ageRange" : "30-39"
    },
    "count" : 33220.0
}
{
    "_id" : {
        "country" : "France",
        "ageRange" : "30-39"
    },
    "count" : 33627.0
}
```

Task 8

Task

Calculate what percentage of the population of the country are the age groups:

- **18-29**
- **30-39**
- **40-49**

The result should be rounded to two decimal places.

Solution

Assuming that the result from the previous task has been written to the collection ***task7***.

Query

```
db.task7.aggregate([
    {
        $group: {
            _id: '$_id.country',
            total: {$sum: '$count'},
            ageRanges: {
                $addToSet: {ageRange: '$_id.ageRange', count: '$count'}
            }
        }
    },
    {
        $unwind: '$ageRanges'
    },
    {
        $project: {
            _id: {
                country: '$_id',
                ageRange: '$ageRanges.ageRange'
            },
            percent: {
                $multiply: [
                    {
                        $divide: ['$ageRanges.count', '$total']
                    },
                    100
                ]
            }
        }
    },
    {
        $project: {
            percent: {
                $let: {
                    vars: {
                       p100: {$multiply: ['$percent', 100]}
                    },
                    in: {
                        $divide: [
                            {
```

```
                                $add: [
                                    {$floor: '$$p100'},
                                    {
                                        $cond: {
                                            if: {
                                                $gte: [
                                                    {$mod: ['$$p100', 1]},
                                                    0.5
                                                ]
                                            },
                                            then: 1,
                                            else: 0
                                        }
                                    }
                                ]
                            },
                            100
                        ]
                    }
                }
            }
        }
    }
])
```

Result

```
{
    "_id" : {
        "country" : "Poland",
        "ageRange" : "18-29"
    },
    "percent" : 31.13
}
{
    "_id" : {
        "country" : "Poland",
        "ageRange" : "40-49"
    },
    "percent" : 35.24
}
{
    "_id" : {
        "country" : "Poland",
        "ageRange" : "30-39"
    },
    "percent" : 33.64
}
{
    "_id" : {
        "country" : "France",
        "ageRange" : "18-29"
    },
    "percent" : 31.46
}
{
    "_id" : {
        "country" : "France",
        "ageRange" : "40-49"
    },
```

```
    "percent" : 35.32
}
{
    "_id" : {
        "country" : "France",
        "ageRange" : "30-39"
    },
    "percent" : 33.22
}
```