

Big Data System Group 203 Assignment

Group Members:

Name	BITS ID	Contribution
CHANDUPATLA ANIRUDH REDDY	2022DA04387	100%
MEDAPATI JAYANAGA SURESH REDDY	2022DA04445	100%
YASH RATHORE	2022DA04617	100%

Imported the data from given [Women Clothing Ecommerce Sales Data.csv](#)

#	BDS_Assignment	_id ObjectId	order_id Int32	order_date String	sku Int32	color String	size String	unit_price Int32
1		ObjectId('657efc8e894ce316f32...	3	"2022/6/10 19:59:00"	799	"Dark Blue"	"XL"	264
2		ObjectId('657efc8e894ce316f32...	5	"2022/6/11 14:46:00"	799	"Dark Blue"	"XL"	264
3		ObjectId('657efc8e894ce316f32...	6	"2022/6/11 19:23:00"	799	"Dark Blue"	"XL"	264
4		ObjectId('657efc8e894ce316f32...	9	"2022/6/11 22:53:00"	239	"Black"	No field	258
5		ObjectId('657efc8e894ce316f32...	10	"2022/6/12 17:32:00"	708	"Dark Blue"	"XL"	278
6		ObjectId('657efc8e894ce316f32...	14	"2022/6/14 0:07:00"	799	"Dark Blue"	"M"	264
7		ObjectId('657efc8e894ce316f32...	18	"2022/6/15 13:36:00"	77	"A"	No field	288
8		ObjectId('657efc8e894ce316f32...	23	"2022/6/16 20:12:00"	799	"Dark Blue"	"L"	264
9		ObjectId('657efc8e894ce316f32...	26	"2022/6/18 11:31:00"	708	"Dark Blue"	"3XL"	278
10		ObjectId('657efc8e894ce316f32...	27	"2022/6/18 19:51:00"	708	"Dark Blue"	"XL"	278

1. What is the total revenue generated by the e-commerce store in June 2022?

Ans: 25259

Query:

```
[
{
  $match: {
    order_date: {
      $gte: "2022/6/1",
      $lt: "2022/7/1",
    }
  },
  $group: {
    _id: null,
    total_revenue: {
      $sum: "$revenue"
    }
  }
}
]
```

db.BDS_Assignment

527 DOCUMENTS 1 INDEXES

Documents Aggregations Schema Indexes Validation

Pipeline ▾ \$match \$group Generate aggregation + Explain Export Run More Options ▾

Untitled - modified SAVE + CREATE NEW EXPORT TO LANGUAGE PREVIEW STAGES TEXT

```
1 [
2   {
3     $match: {
4       order_date: {
5         $gte: "2022/6/1",
6         $lt: "2022/7/1",
7       }
8     },
9   },
10  {
11    $group: {
12      _id: null,
13      total_revenue: {
14        $sum: "$revenue"
15      }
16    }
17  }
18 ]
```

PIPELINE OUTPUT

Sample of 1 document

```
_id: null
total_revenue: 25259
```

2. What is the highest revenue of a SKU in a single order?

Ans: 899

Query:

```
[
{
  $group: {
    _id: "$sku",
    highestRevenue: {
      $max: "$revenue"
    }
  }
},
{
  $sort: {
    highestRevenue: -1
  }
},
{
  $limit: 1
}
]
```

db.BDS_Assignment

527 DOCUMENTS 1 INDEXES

Documents Aggregations Schema Indexes Validation

Pipeline ▾ \$group \$sort \$limit Generate aggregation + Explain Export Run More Options ▾

Untitled - modified SAVE + CREATE NEW EXPORT TO LANGUAGE PREVIEW STAGES TEXT

```
1 [
2   {
3     $group: {
4       _id: "$sku",
5       highestRevenue: { $max: "$revenue" }
6     }
7   },
8   {
9     $sort: {
10      highestRevenue: -1
11    }
12  },
13  {
14    $limit: 1
15  }
16 ]
```

PIPELINE OUTPUT

Sample of 1 document

```
_id: 708
highestRevenue: 899
```

3. Find the average unit price per SKU for the “Dark Blue” color.

Ans:

SKU ID	Average Unit Price
9699	288.0
1499	287.0
89	256.55
799	275.8056537102473
708	294.4040404

Query: [

```
{
  $match: {
    color: "Dark Blue"
  },
  $group: {
    _id: "$sku",
    avgUnitPrice: {
      $avg: "$unit_price"
    }
  }
}
```

]

db.BDS_Assignment

527 1
DOCUMENTS INDEXES

Documents Aggregations Schema Indexes Validation

Pipeline ▼ \$match \$group Generate aggregation ⚙ ? Explain Export Run More Options ▶

Untitled - modified SAVE + CREATE NEW EXPORT TO LANGUAGE PREVIEW {} STAGES TEXT ⚙

```
1 [
2   {
3     $match: {
4       color: "Dark Blue"
5     },
6   },
7   {
8     $group: {
9       _id: "$sku",
10      avgUnitPrice: {
11        $avg: "$unit_price"
12      }
13    }
14  }
15 ]
```

PIPELINE OUTPUT

Sample of 5 documents

OUTPUT OPTIONS ▼

_id: 9699
avgUnitPrice: 288

_id: 1499
avgUnitPrice: 287

_id: 89
avgUnitPrice: 256.55555555555554

_id: 799
avgUnitPrice: 275.8056537102473

_id: 708
avgUnitPrice: 294.4040404040404

4. Identify the top 5 best-selling colors based on quantity sold.

Ans:

Color	Quantity
Dark Blue	401
Light Blue	26
Black	15
Grey Wash	9
Mid Wash	8

Query:

```
[
{
  $group: {
    _id: "$color",
    total_quantity: {
      $sum: "$quantity"
    }
  },
},
{ $sort: {
  total_quantity: -1
}
},
{ $limit: 5 },
]
```

db.BDS_Assignment

527 1
DOCUMENTS INDEXES

Documents Aggregations Schema Indexes Validation

Pipeline ▼ \$group +2 Generate aggregation ⚙ ? Explain Export Run More

Untitled - modified SAVE + CREATE NEW EXPORT TO LANGUAGE PREVIEW {} TEXT ⚙

```
1 [
2   {
3     $group: {
4       _id: "$color",
5       total_quantity: {
6         $sum: "$quantity"
7       }
8     },
9   },
10  { $sort: {
11    total_quantity: -1
12  }
13  },
14  { $limit: 5 },
15 ]
```

PIPELINE OUTPUT

Sample of 5 documents

OUTPUT OPTIONS ▼

_id: "Dark Blue"
total_quantity: 401

_id: "Light Blue"
total_quantity: 26

_id: "Black"
total_quantity: 15

_id: "Grey Wash"
total_quantity: 9

_id: "Mid Wash"
total_quantity: 8

5. Calculate the total revenue for orders with "Dark Blue" products in XL size.

Ans: 40759

Query:

```
[
  {
    $match: {
      color: "Dark Blue",
      size: "XL"
    }
  },
  {
    $group: {
      _id: null,
      total_revenue: {
        $sum: "$revenue"
      }
    },
  },
]
```

db.BDS_Assignment

527 1
DOCUMENTS INDEXES

Documents Aggregations Schema Indexes Validation

Pipeline \$match +1 Generate aggregation Explain Export Run More

Untitled - modified SAVE CREATE NEW EXPORT TO LANGUAGE PREVIEW

```
1 [
2   {
3     $match: {
4       color: "Dark Blue",
5       size: "XL"
6     }
7   },
8   {
9     $group: {
10      _id: null,
11      total_revenue: {
12        $sum: "$revenue"
13      }
14    },
15  },
16 ]
```

PIPELINE OUTPUT OUTPUT OPTIONS

Sample of 1 document

```
{
  _id: null
  total_revenue: 40759
}
```

6. What is the average quantity of products per order for "Dark Blue" SKUs?

Ans:

average quantity of products for order ID 47 is 2.0

average quantity of products for order ID 242 is 2.0

average quantity of products for order ID 84 is 1.5

average quantity of products for order ID 83 is 1.333

average quantity of products for all other order ID is 1.0

Query:

```
[
  {
    $match: {
      color: "Dark Blue"
    }
  },
  {
    $group: {
      _id: "$order_id",
      avgQuantityPrice: {
        $avg: "$quantity"
      }
    }
  },
  {
    $sort: {
      avgQuantityPrice: -1
    }
  }
]
```

db.BDS_Assignment

527 1
DOCUMENTS INDEXES

Documents Aggregations Schema Indexes Validation

Pipeline \$match +2 Generate aggregation Explain Export Run More

Untitled - modified SAVE CREATE NEW EXPORT TO LANGUAGE PREVIEW

```
1 [
2   {
3     $match: {
4       color: "Dark Blue"
5     }
6   },
7   {
8     $group: {
9       _id: "$order_id",
10      avgQuantityPrice: {
11        $avg: "$quantity"
12      }
13    }
14  },
15  {
16    $sort: {
17      avgQuantityPrice: -1
18    }
19  }
20 ]
```

PIPELINE OUTPUT OUTPUT OPTIONS

Sample of 50 documents

```
{
  _id: 47
  avgQuantityPrice: 2
}
{
  _id: 242
  avgQuantityPrice: 2
}
{
  _id: 84
  avgQuantityPrice: 1.5
}
{
  _id: 83
  avgQuantityPrice: 1.3333333333333333
}
{
  _id: 223
  avgQuantityPrice: 1
}
```

7. Find the total revenue for each day in June 2022 and identify the highest revenue day.

Ans: Highest revenue of 4338 was on June 6th 2022

Query:

```
[
  { $match: {
    order_date: {
      $gte: "2022/6/1",
      $lt: "2022/7/1",
    },
  },
  { $addFields: {
    day: {
      $dayOfMonth: {
        $toDate: "$order_date"
      },
    },
  },
  { $group: {
    _id: "$day",
    total_revenue_in_a_day: {
      $sum: "$revenue"
    },
  },
  { $sort: { total_revenue_in_a_day: -1 } },
  //{ $limit: 1 },
]
```

```
1 [
2   {
3     $match: {
4       order_date: {
5         $gte: "2022/6/1",
6         $lt: "2022/7/1",
7       },
8     },
9   },
10  {
11    $addFields: {
12      day: {
13        $dayOfMonth: {
14          $toDate: "$order_date"
15        },
16      },
17    },
18  },
19  {
20    $group: {
21      _id: "$day",
22      total_revenue_in_a_day: {
23        $sum: "$revenue"
24      },
25    },
26  },
27  { $sort: { total_revenue_in_a_day: -1 } },
28  //{ $limit: 1 },
29 ]
```

PIPELINE OUTPUT

Sample of 27 documents

_id: 6
total_revenue_in_a_day: 4338

_id: 15
total_revenue_in_a_day: 2654

_id: 11
total_revenue_in_a_day: 1606

_id: 9
total_revenue_in_a_day: 1449

_id: 18
total_revenue_in_a_day: 1380

_id: 24
total_revenue_in_a_day: 1360

_id: 14
total_revenue_in_a_day: 1226

8. Calculate the average unit price for each SKU

Ans:

sku	average unit_price
127	238
1499	287
1719	256.5
218	314.857142857143
229	207
239	262.4375
2499	240.5
29	227
3081	198
339	271
3799	265.56
439	258
539	268
61399	278
628	268
708	294.40
77	253.4375
79	313
799	275.805653710247
8499	233
89	257
897	268
9699	288
bobo	212.75

db.BDS_Assignment

527

1

DOCUMENTS INDEXES

Documents Aggregations Schema Indexes Validation

Pipeline ▼ \$group Generate aggregation ⚙️ ? Explain Export Run More Options

Untitled - modified

SAVE

+ CREATE NEW

EXPORT TO LANGUAGE

PREVIEW

⚙️

```
1 [
2   {
3     $group: {
4       _id: "$sku",
5       average_unit_price: {
6         $avg: "$unit_price"
7       },
8     },
9   },
10 ]
```

PIPELINE OUTPUT

Sample of 24 documents

OUTPUT OPTIONS ▼

_id: 628
average_unit_price: 268

_id: 61399
average_unit_price: 278

_id: 127
average_unit_price: 238

_id: 3081
average_unit_price: 198

_id: 29
average_unit_price: 227

Query: [

```
{
  $group: {
    _id: "$sku",
    average_unit_price: { $avg: "$unit_price" },
  },
}
```

Ans: 287

[

db.BDS_Assignment

527

1

DOCUMENTS INDEXES

Documents Aggregations Schema Indexes Validation

Pipeline

Sgroup

Generate aggregation

Explain

Export

Run

More Optio

Untitled - *modified*

SAVE

+ CREATE NEW

EXPORT TO LANGUAGE

 PREVIEW

```
1 ▼ [
2 ▼ {
3 ▼   $group: {
4 ▼     _id: "$sku",
5 ▼     total_quantity: {
6 ▼       $sum: "$quantity"
7 ▼     },
8 ▼   },
9 ▼ },
10 ▼ {
11 ▼   $sort: {
12 ▼     total_quantity: -1
13 ▼   },
14 ▼ },
15 ▼ { $limit: 1 },
16 ▼ ]
```

PIPELINE OUTPUT

OUTPUT OPTIONS ▾

Sample of 1 document

```
_id: 799
total_quantity: 287
```

Ans:

Navy Blue: 1.155%

[

```
[
  { $group: {
    _id: "$color",
    revenue: { $sum: "$revenue" }
  }
},
{ $group: {
  _id: null,
  total: { $sum: "$revenue" },
  colors: {
    $push: {
      color: "$_id",
      revenue: "$revenue"
    }
  }
}
},
{ $unwind: "$colors" },
{ $match: {
  "colors.color": {
    $in: ["Dark Blue", "Navy Blue"]
  }
}
},
{ $project: {
  _id: 0,
  color: "$colors.color",
  revenue: "$colors.revenue",
  percentage: {
    $multiply: [
      { $divide: ["$colors.revenue", "$total"] },
      100
    ]
  }
}
}
]
```

PIPELINE OUTPUT

Sample of 2 documents

```
color: "Navy Blue"  
revenue: 1693  
percentage: 1.1554815416430635
```

```
color: "Dark Blue"  
revenue: 112364  
percentage: 76.68903009166047
```


11. Calculate the total revenue for each size category.

Ans:

Size	Total Revenue
2XL	12422
3XL	11301
4XL	3205
L	25362
M	25755
One Size	990
One Size	11793
S	227
XL	45868

```
1 [
2   {
3     $group: {
4       _id: "$size",
5       total_revenue: {
6         $sum: "$revenue"
7       }
8     },
9   },
10  {
11    $sort: {
12      total_revenue: -1
13    }
14  }
15 ]
```

PIPELINE OUTPUT

Sample of 10 documents

OUTPUT OPTIONS

_id: "XL"
total_revenue: 45868

_id: "M"
total_revenue: 25755

_id: "L"
total_revenue: 25362

_id: "2XL"
total_revenue: 12422

_id: "One Size"
total_revenue: 11793

_id: "3XL"
total_revenue: 11301

Query: [

```
{
  $group: {
    _id: "$size",
    total_revenue: {
      $sum: "$revenue"
    }
  },
},
{
  $sort: {
    total_revenue: -1
  }
}
]
```

12 . Identify the top 3 SKUs with the highest average revenue per order.

Ans:

SKU	Average revenue per order
218	314.85714285714283
79	313.0
9699	306.0

Query:

```
[
{
  $group: {
    _id: "$sku",
    avg_revenue: {
      $avg: "$revenue",
    },
  },
},
{
  $sort: { avg_revenue: -1 }
},
{ $limit: 3 },
]
```

```
1 [
2   {
3     $group: {
4       _id: "$sku",
5       avg_revenue: {
6         $avg: "$revenue",
7       },
8     },
9   },
10  {
11    $sort: { avg_revenue: -1 }
12  },
13  { $limit: 3 },
14 ]
```

PIPELINE OUTPUT

Sample of 3 documents

OUTPUT OPTIONS

_id: 218
avg_revenue: 314.85714285714283

_id: 79
avg_revenue: 313

_id: 9699
avg_revenue: 306