#### Big Data System Group 203 Assignment

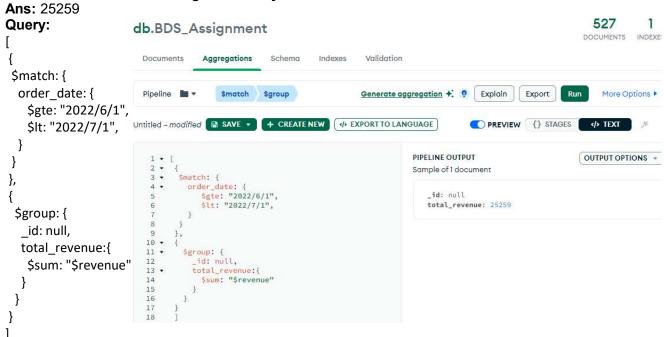
#### Group Members:

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

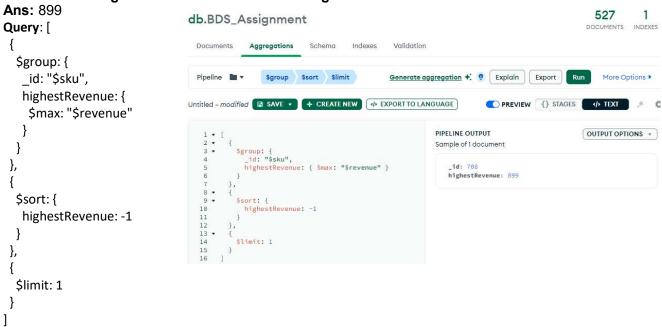
#### Imported the data from given Women Clothing Ecommerce Sales Data.csv

	DOD_11334Billione						
	_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

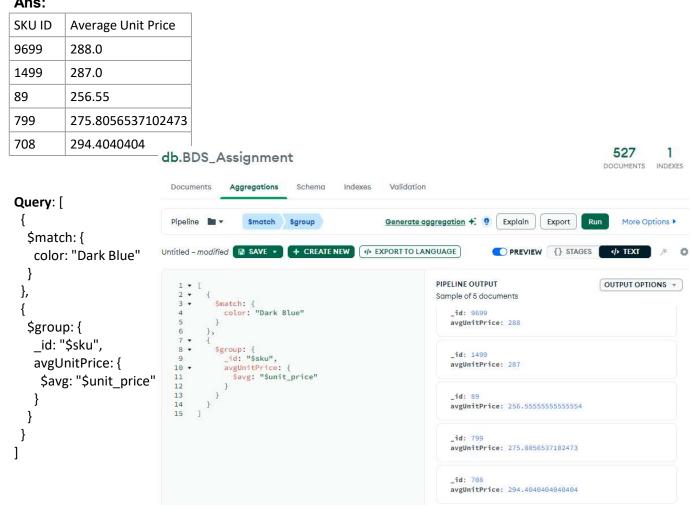
#### 1. What is the total revenue generated by the e-commerce store in June 2022?



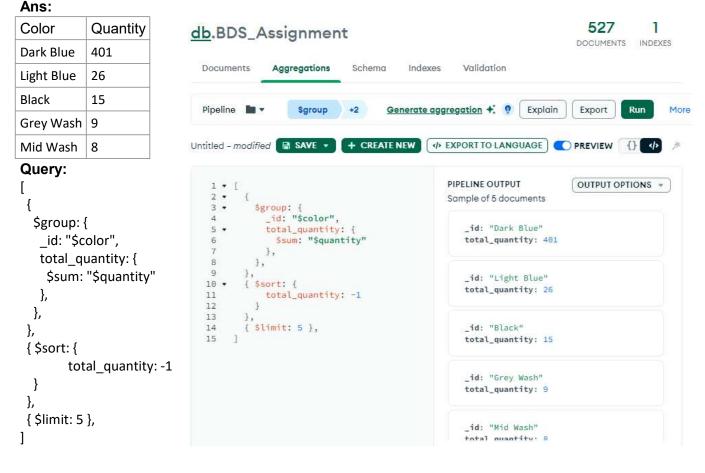
### 2. What is the highest revenue of a SKU in a single order?



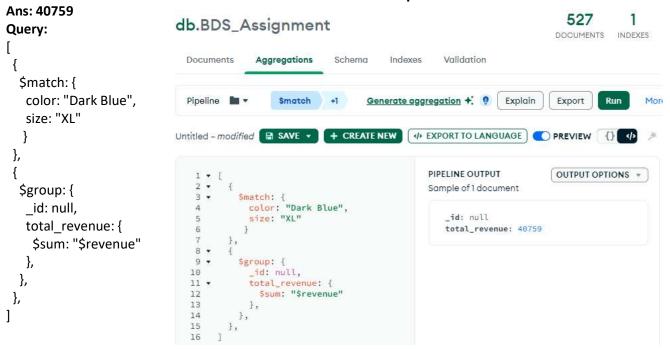
# 3. Find the average unit price per SKU for the "Dark Blue" color.



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



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



#### 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:
                                                                                                            527
                                                                                                                        1
                                db.BDS_Assignment
[
                                                                                                                      INDEXES
                                                                                                          DOCUMENTS
 {
                                                                                    Validation
                                  Documents
                                               Aggregations
                                                               Schema
                                                                         Indexes
  $match: {
   color: "Dark Blue"
                                  Pipeline 🖿 🔻
                                                    Smatch
                                                                     Generate aggregation +: 9
                                                                                                Explain
                                                                                                           Export
 },
                                                                              Untitled - modified
                                                 SAVE .
                                                              + CREATE NEW
  $group: {
    _id: "$order_id",
                                                                                 PIPELINE OUTPUT
                                                                                                          OUTPUT OPTIONS *
                                   1 • [
   avgQuantityPrice: {
                                    2 .
                                                                                 Sample of 50 documents
                                    3
                                            Smatch: {
     $avg: "$quantity"
                                              color: "Dark Blue"
                                    4
                                                                                    id: 47
                                    5
                                                                                    avgQuantityPrice: 2
                                    6
  }
                                    8 .
                                            $group: {
 },
                                              _id: "$order_id",
                                   9
                                                                                     _id: 242
 {
                                              avgQuantityPrice: {
   $avg: "$quantity"
                                   10 .
                                                                                    avgQuantityPrice: 2
                                   11
  $sort: {
                                   12
   avgQuantityPrice: -1
                                   13
                                   14
                                                                                     id: 84
  }
                                   15 *
                                                                                    avgQuantityPrice: 1.5
                                            Ssort: [
                                   16 .
 }
                                   17
                                              avgQuantityPrice: -1
]
                                   1.8
                                   19
                                                                                    avgQuantityPrice: 1.33333333333333333
                                   20
                                                                                    _id: 223
                                                                                    avgQuantityPrice: 1
```

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

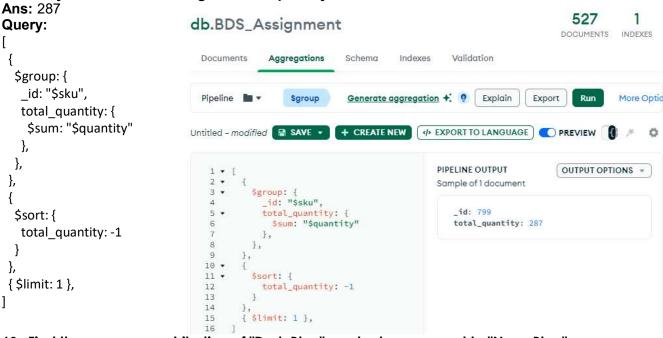
**Ans:** Highest revenue of 4338 was on June 6th 2022 Query:

```
[
                                                                                                           PIPELINE OUTPUT
 { $match: {
                                                                                                           Sample of 27 documents
                                            .
                                                   $match: {
    order date: {
                                                     order_date: {
   $gte: "2022/6/1",
   $lt: "2022/7/1",
                                          4 .
                                          5
     $gte: "2022/6/1",
                                          6
                                                                                                              total_revenue_in_a_day: 4338
     $lt: "2022/7/1",
                                                     },
                                          8
    },
                                          9
                                                                                                               id: 15
                                         10 -
  },
                                                                                                              total_revenue_in_a_day: 2654
                                         11 •
                                                   $addFields: {
                                         12 •
 },
                                                       $dayOfMonth: {
    $toDate: "$order_date"
                                         13 •
 { $addFields: {
                                         14
                                                                                                               id: 11
                                         15
                                                                                                              total_revenue_in_a_day: 1606
    day: {
                                         16
     $dayOfMonth: {
                                         17
                                         18
                                                                                                               _id: 9
       $toDate: "$order date"
                                         19 •
                                                                                                              total_revenue_in_a_day: 1449
                                         20 •
                                                   $group: {
     },
                                         21
                                                     _id: "$day",
                                                     total_revenue_in_a_day: {
   $sum: "$revenue"
                                         22 -
    },
                                         23
                                                                                                               id: 18
  },
                                         24
                                                                                                              total_revenue_in_a_day: 1380
                                                   },
 },
                                         26
                                                  $sort: { total_revenue_in_a_day: -1 } },
 { $group: {
                                         27
                                                                                                               id: 24
                                         28
                                                 //{ $limit: 1 },
    _id: "$day",
                                                                                                              total_revenue_in_a_day: 1360
                                         29
    total_revenue_in_a_day: {
     $sum: "$revenue"
                                                                                                               id: 14
                                                                                                              total_revenue_in_a_day: 1226
    },
  },
 },
 { $sort: { total_revenue_in_a_day: -1 } },
 //{ $limit: 1 },
```

## 8. Calculate the average unit price for each SKU

```
527
                                                                                    1
db.BDS_Assignment
                                                                      DOCUMENTS
                                                                                 INDEXES
  Documents
              Aggregations
                             Schema
                                       Indexes
                                                 Validation
                              Generate aggregation +: 9 Explain
  Pipeline ■ ▼
                   Sgroup
                                                                Export
                                                                                 More Options
Untitled - modified SAVE
                           0
   1 • [
                                              PIPELINE OUTPUT
                                                                     OUTPUT OPTIONS *
         1
                                              Sample of 24 documents
           $group: {
   _id: "$sku",
             average_unit_price: {
    $avg: "$unit_price"
                                                  _id: 628
                                                  average_unit_price: 268
             },
           },
         },
                                                  id: 61399
                                                  average_unit_price: 278
                                                  id: 127
                                                  average_unit_price: 238
                                                  _id: 3081
                                                  average_unit_price: 198
Query: [
                                                  average_unit_price: 227
  $group: {
    id: "$sku",
   average_unit_price: { $avg: "$unit_price" },
  },
 },
]
```

#### 9.Identify the SKU with the highest total quantity sold.



color: "Navy Blue" revenue: 1693

color: "Dark Blue"

revenue: 112364

percentage: 1.1554815416430635

percentage: 76.68903009166047

# 10. Find the revenue contribution of "Dark Blue" products compared to "Navy Blue"

```
products.
Ans:
                                                                                                 PIPELINE OUTPUT
Dark Blue: 76.689%
                                           { $group: {
                                                                                                 Sample of 2 documents
Navy Blue: 1.155%
                                                id: "$color",
                                               revenue: { $sum: "$revenue" }
Query:
[ { $group: {
                                           { $group: {
   _id: "$color",
                                               id: null.
                                               total: { $sum: "$revenue" },
   revenue: { $sum: "$revenue" }
                                                 $push: {
  }
                                                  color: "$_id",
                                                   revenue: "$revenue"
 },
 { $group: {
   _id: null,
                                             $unwind: "$colors" },
   total: { $sum: "$revenue" },
                                           { Smatch: {
                                               "colors.color": {
   colors: {
                                                 $in: ["Dark Blue", "Navy Blue"]
     $push: {
                                             }
      color: "$ id",
      revenue: "$revenue"
                                           { $project: {
                                               _id: 0,
color: "$colors.color",
     }
                                               revenue: "$colors.revenue",
   }
                                               percentage:
                                                 $multiply: [
  }
                                                   { $divide: ["$colors.revenue", "$total"] },
 },
                                                   100
 { $unwind: "$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 ]
  }
 }
]
```

#### 11. Calculate the total revenue for each size category.

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

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

```
PIPELINE OUTPUT
                          OUTPUT OPTIONS +
Sample of 10 documents
   _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:
[
                                1 • [
                                                                               PIPELINE OUTPUT
                                                                                                        OUTPUT OPTIONS *
                                                                  Sample of 3 documents
                                         $group: {
    _id: "$sku",
  $group: {
                                4
                                                                                   id: 218
   _id: "$sku",
                                5 *
                                           avg_revenue: {
                                6
                                             $avg: "$revenue",
                                                                                  avg_revenue: 314.85714285714283
   avg_revenue: {
                                8
                                         },
    $avg: "$revenue",
                                9
                                                                                   _id: 79
   },
                               10 •
                                                                                  avg_revenue: 313
                               11
                                         $sort: { avg_revenue: -1 }
  },
                               12
},
                                       { $limit: 3 },
                               13
                                                                                  _id: 9699
                                                                                  avg_revenue: 306
  $sort: { avg_revenue: -1 }
},
{ $limit: 3 },
```