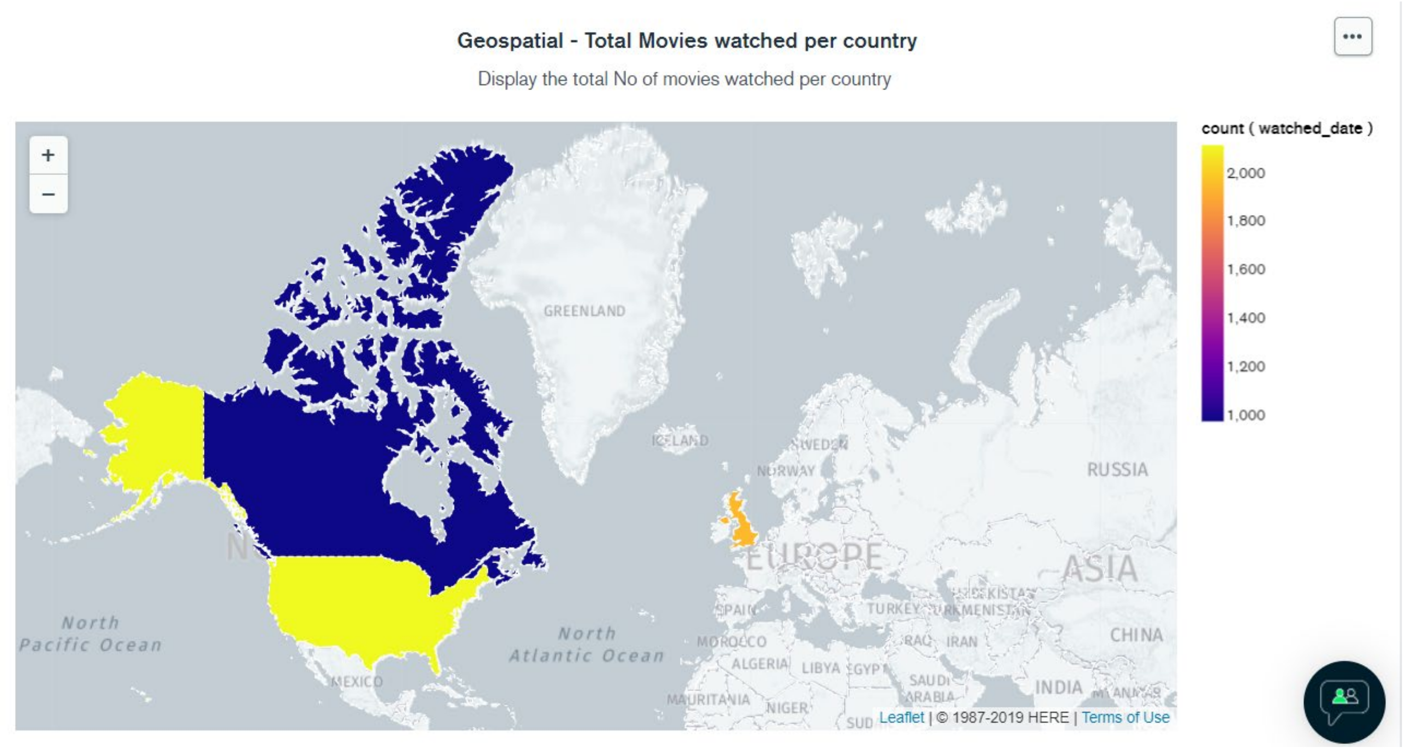
LAB 2 - MONGO DB EXECUTION LOGS

Submitted By : Gayathri Sundareshwar, Keerthana Gopikrishnan and Deepasha Jenamani

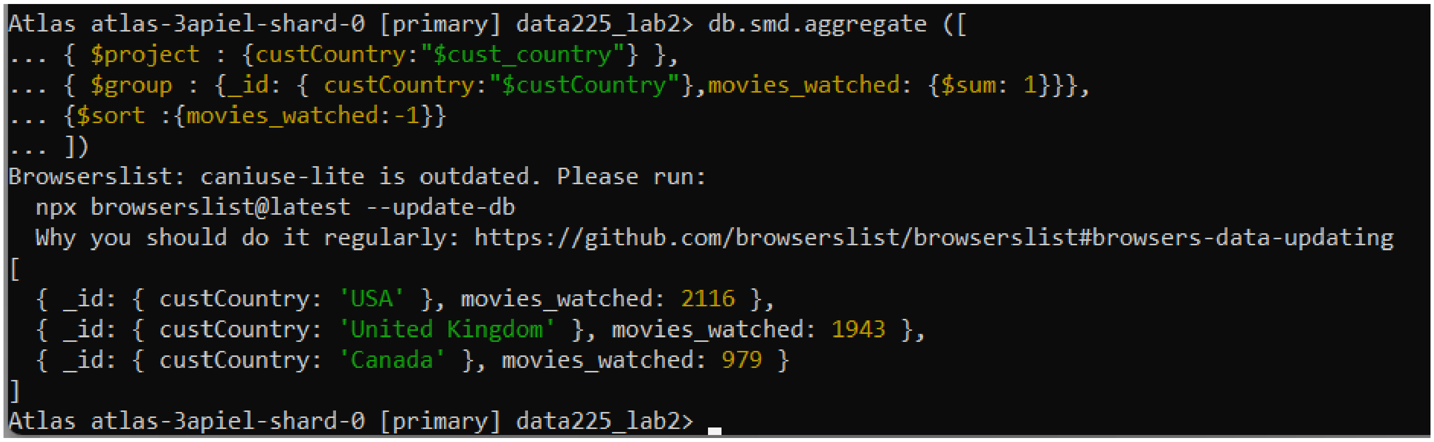
SCENARIO 1 - NO OF MOVIES WATCHED PER COUNTRY

MONGO CHARTS:



QUERY & RESULT (SHELL EXECUTION):

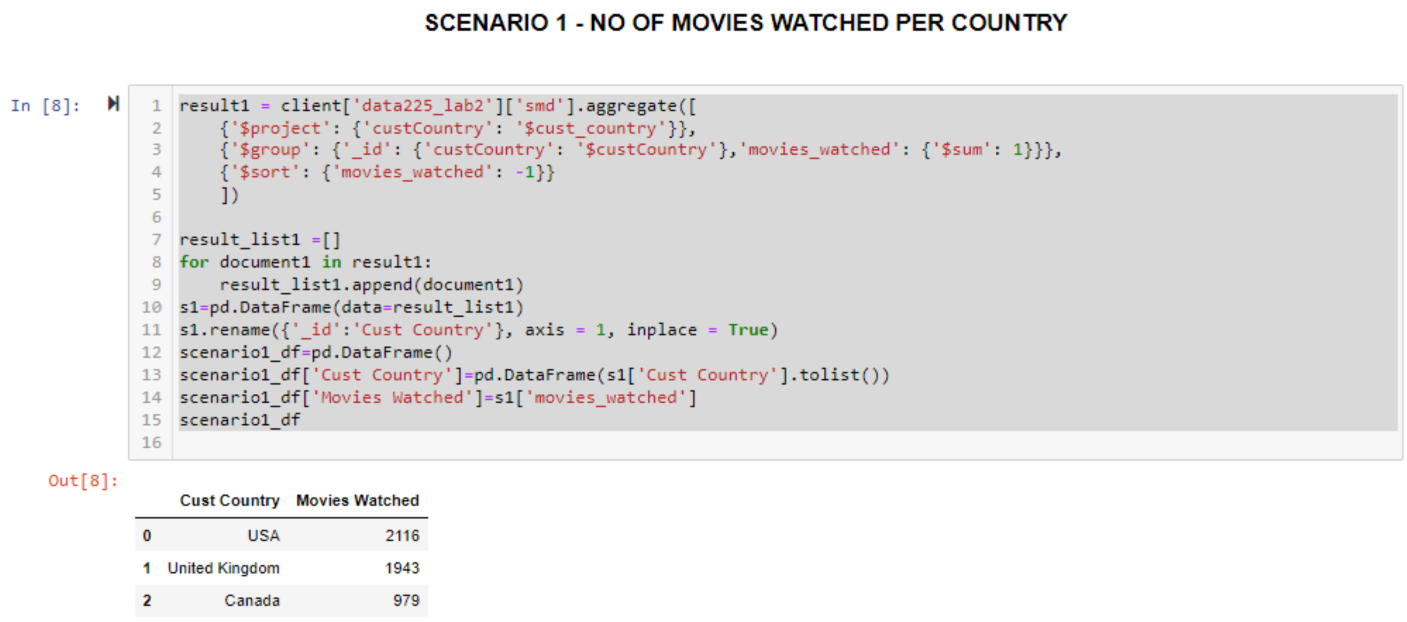
```
db.smd.aggregate ([
{ $project : {custCountry:"$cust_country"} },
{ $group : {_id: { custCountry:"$custCountry"},movies_watched: {$sum: 1}}},
{$sort :{movies_watched:-1}}
])
```



```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate ([
... { $project : {custCountry:"$cust_country"} },
... { $group : {_id: { custCountry:"$custCountry"},movies_watched: {$sum: 1}}},
... {$sort :{movies_watched:-1}}
... ])
Browserslist: caniuse-lite is outdated. Please run:
  npx browserslist@latest --update-db
  Why you should do it regularly: https://github.com/browserslist/browserslist#browsers-data-updating
[
  { _id: { custCountry: 'USA' }, movies_watched: 2116 },
  { _id: { custCountry: 'United Kingdom' }, movies_watched: 1943 },
  { _id: { custCountry: 'Canada' }, movies_watched: 979 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> _
```

QUERY & RESULT (PYTHON EXECUTION):

```
result1 = client['data225_lab2']['smd'].aggregate([
{'$project': {'custCountry': '$cust_country'}},
{'$group': {'_id': {'custCountry': '$custCountry'},'movies_watched': {'$sum': 1}}},
{'$sort': {'movies_watched': -1}}
])

result_list1 =[]
for document1 in result1:
result_list1.append(document1)
s1=pd.DataFrame(data=result_list1)
s1.rename({'_id':'Cust Country'}, axis = 1, inplace = True)
scenario1_df=pd.DataFrame()
scenario1_df['Cust Country']=pd.DataFrame(s1['Cust Country'].tolist())
scenario1_df['Movies Watched']=s1['movies_watched']
scenario1_df
```



SCENARIO 1 - NO OF MOVIES WATCHED PER COUNTRY

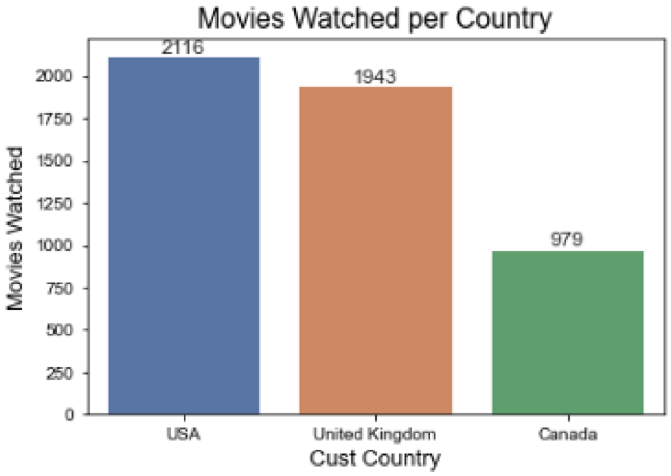
In [8]:

```
result1 = client['data225_lab2']['smd'].aggregate([
    {'$project': {'custCountry': '$cust_country'}},
    {'$group': {'_id': {'custCountry': '$custCountry'},'movies_watched': {'$sum': 1}}},
    {'$sort': {'movies_watched': -1}}
    ])

result_list1 =[]
for document1 in result1:
    result_list1.append(document1)
s1=pd.DataFrame(data=result_list1)
s1.rename({'_id':'Cust Country'}, axis = 1, inplace = True)
scenario1_df=pd.DataFrame()
scenario1_df['Cust Country']=pd.DataFrame(s1['Cust Country'].tolist())
scenario1_df['Movies Watched']=s1['movies_watched']
scenario1_df
```

Out[8]:

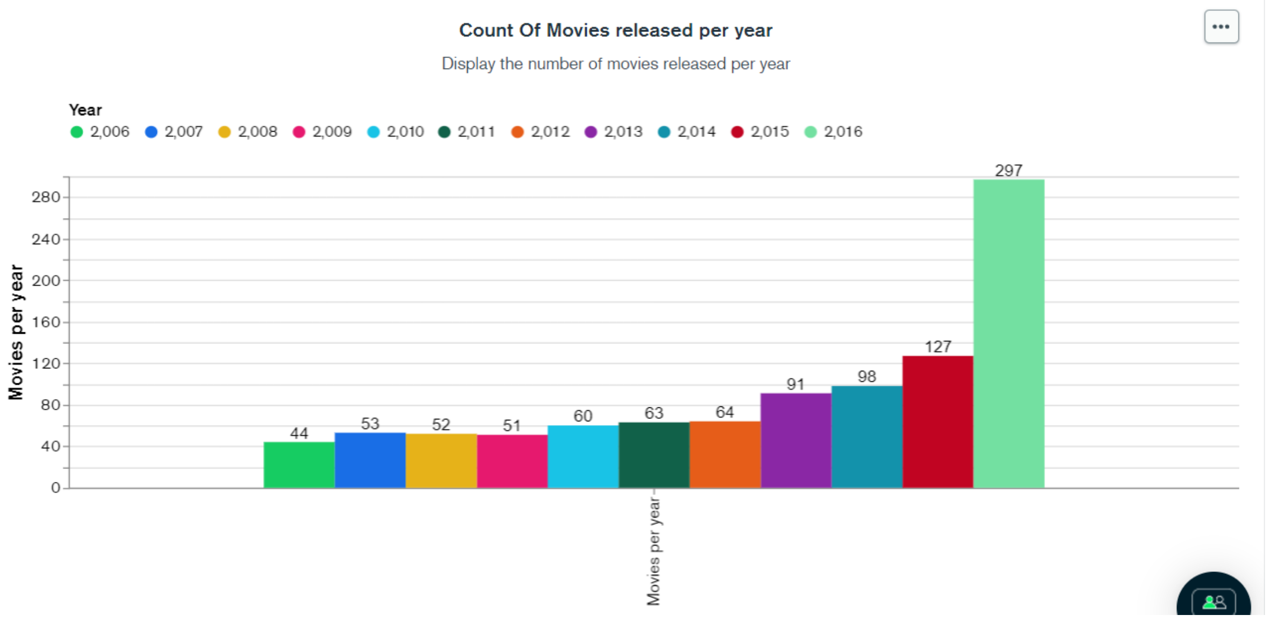
	Cust Country	Movies Watched
0	USA	2116
1	United Kingdom	1943
2	Canada	979

JUPYTER VISUALIZATION:



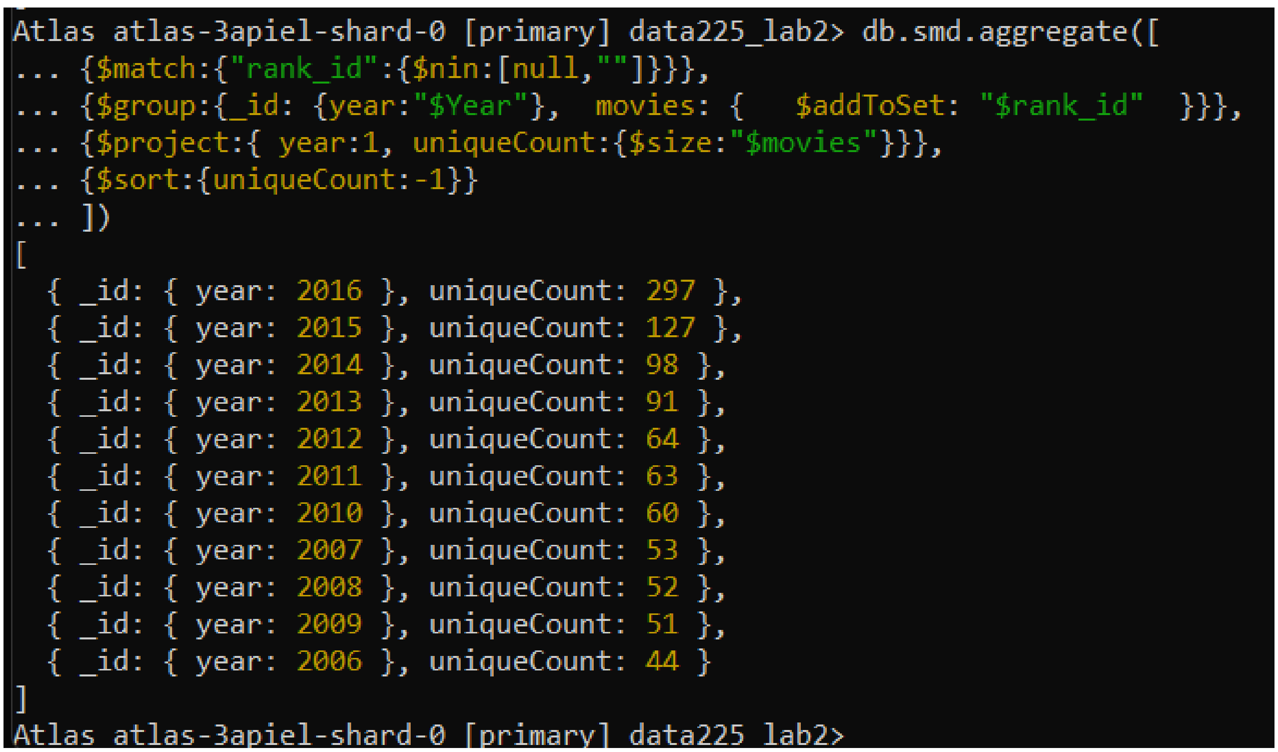
SCENARIO 2 - NO OF MOVIES RELEASED PER YEAR

MONGO CHARTS:



QUERY & RESULT (SHELL EXECUTION):

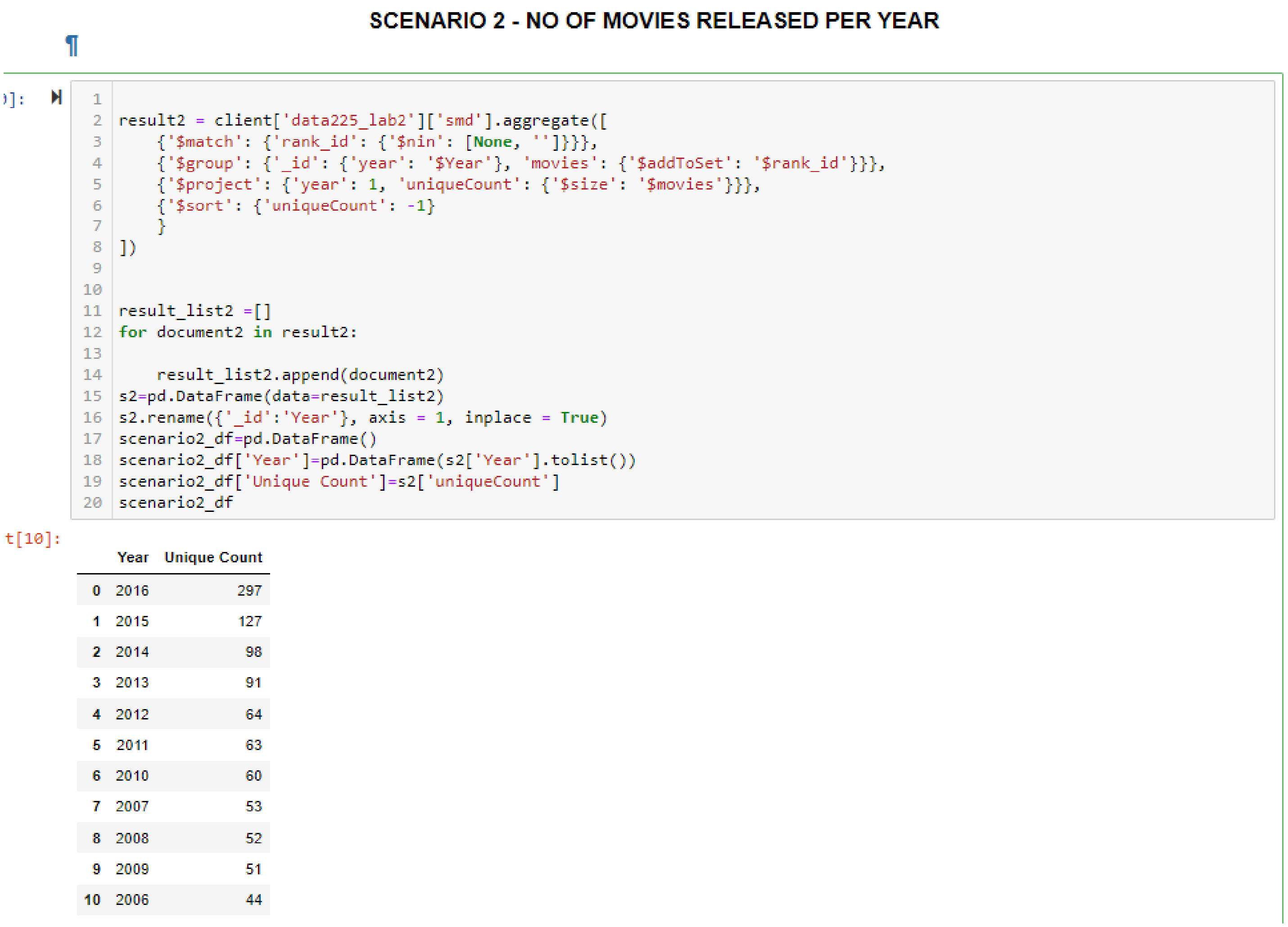
```
db.smd.aggregate([
{$match:{"rank_id":{$nin:[null,""]}}},
{$group:{_id: {year:"$Year"},  movies: {   $addToSet: "$rank_id"  }}},
{$project:{ year:1, uniqueCount:{$size:"$movies"}}},
{$sort:{uniqueCount:-1}}
])
```



```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:{"rank_id":{$nin:[null,""]}}},
... {$group:{_id: {year:"$Year"},  movies: {   $addToSet: "$rank_id"   }}},
... {$project:{ year:1, uniqueCount:{$size:"$movies"}}},
... {$sort:{uniqueCount:-1}}
... ])
[
  { _id: { year: 2016 }, uniqueCount: 297 },
  { _id: { year: 2015 }, uniqueCount: 127 },
  { _id: { year: 2014 }, uniqueCount: 98 },
  { _id: { year: 2013 }, uniqueCount: 91 },
  { _id: { year: 2012 }, uniqueCount: 64 },
  { _id: { year: 2011 }, uniqueCount: 63 },
  { _id: { year: 2010 }, uniqueCount: 60 },
  { _id: { year: 2007 }, uniqueCount: 53 },
  { _id: { year: 2008 }, uniqueCount: 52 },
  { _id: { year: 2009 }, uniqueCount: 51 },
  { _id: { year: 2006 }, uniqueCount: 44 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2>
```

QUERY & RESULT (PYTHON EXECUTION):

```
result2 = client['data225_lab2']['smd'].aggregate([
{'$match': {'rank_id': {'$nin': [None, '']}}},
{'$group': {'_id': {'year': '$Year'}, 'movies': {'$addToSet': '$rank_id'}}},
{'$project': {'year': 1, 'uniqueCount': {'$size': '$movies'}}},
{'$sort': {'uniqueCount': -1}    }
])
result_list2 =[]
for document2 in result2:
result_list2.append(document2)
s2=pd.DataFrame(data=result_list2)
s2.rename({'_id':'Year'}, axis = 1, inplace = True)
scenario2_df=pd.DataFrame()
scenario2_df['Year']=pd.DataFrame(s2['Year'].tolist())
scenario2_df['Unique Count']=s2['uniqueCount']
scenario2_df
```



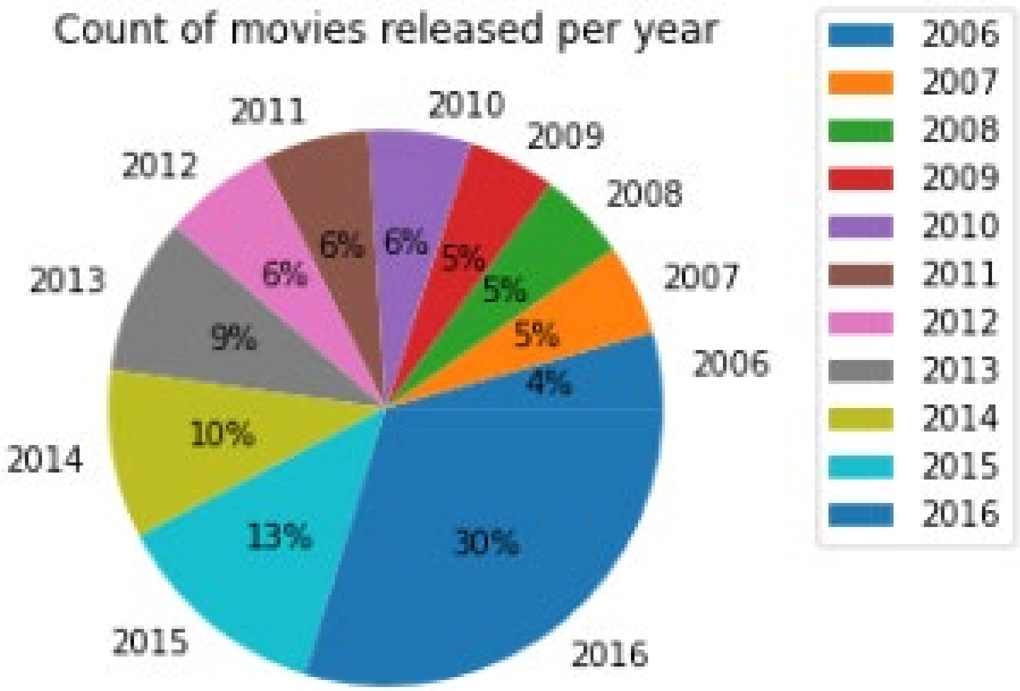
SCENARIO 2 - NO OF MOVIES RELEASED PER YEAR

```

result2 = client['data225_lab2']['smd'].aggregate([
    {'$match': {'rank_id': {'$nin': [None, '']}}},
    {'$group': {'_id': {'year': '$Year'}, 'movies': {'$addToSet': '$rank_id'}}},
    {'$project': {'year': 1, 'uniqueCount': {'$size': '$movies'}}},
    {'$sort': {'uniqueCount': -1}
    }
])


result_list2 =[]
for document2 in result2:

    result_list2.append(document2)
s2=pd.DataFrame(data=result_list2)
s2.rename({'_id':'Year'}, axis = 1, inplace = True)
scenario2_df=pd.DataFrame()
scenario2_df['Year']=pd.DataFrame(s2['Year'].tolist())
scenario2_df['Unique Count']=s2['uniqueCount']
scenario2_df
```

t[10]:

```
    Year  Unique Count
 0  2016           297
 1  2015           127
 2  2014            98
 3  2013            91
 4  2012            64
 5  2011            63
 6  2010            60
 7  2007            53
 8  2008            52
 9  2009            51
10  2006            44
```

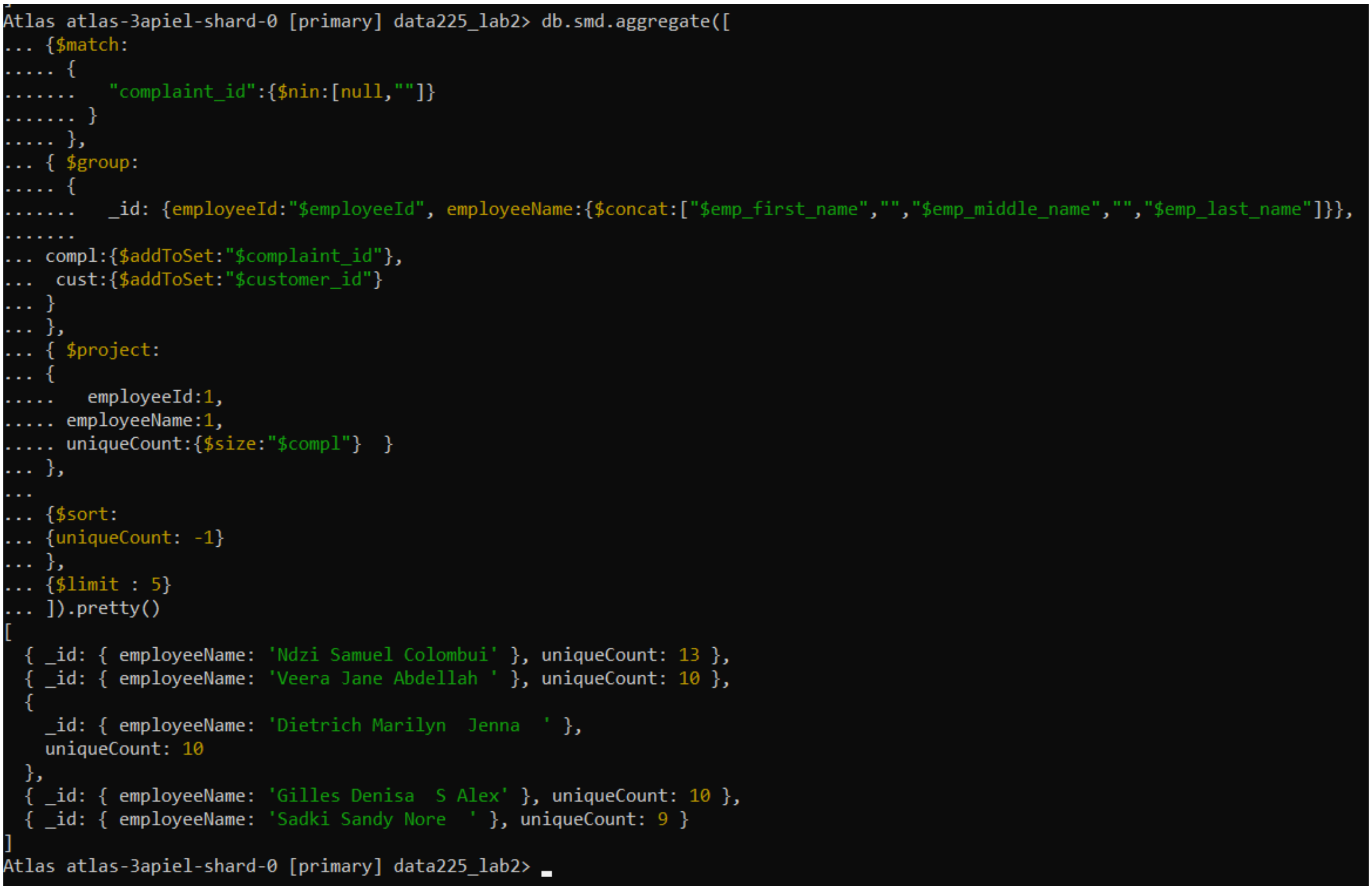
JUPYTER VISUALIZATION:



SCENARIO 3 - TOP 5 EMPLOYEES WHO HAVE RESPONDED TO MOST COMPLAINTS

QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$match:{  "complaint_id":{$nin:[null,""]}}},
{ $group:
{ _id: {employeeId:"$employeeId",
employeeName:{$concat:["$emp_first_name","","$emp_middle_name","","$emp_last_name"]}},
compl:{$addToSet:"$complaint_id"}, cust:{$addToSet:"$customer_id"}
} },
{ $project:{  employeeId:1,employeeName:1,uniqueCount:{$size:"$compl"}  }},
{$sort:{uniqueCount: -1}},
{$limit : 5}
]).pretty()
```

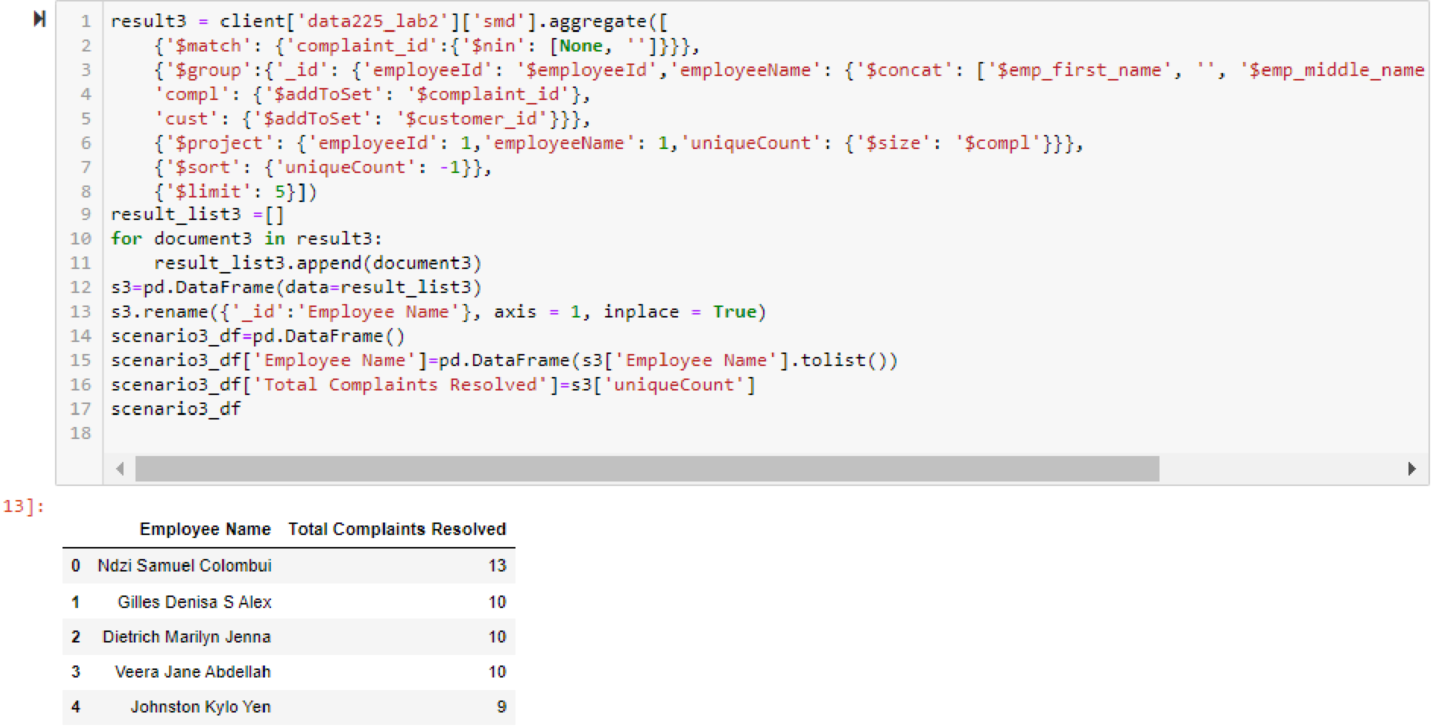


```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:
..... {
.......    "complaint_id":{$nin:[null,""]}
....... }
..... },
... { $group:
..... {
.......    _id: {employeeId:"$employeeId", employeeName:{$concat:["$emp_first_name","","$emp_middle_name","","$emp_last_name"]}},
.......
... compl:{$addToSet:"$complaint_id"},
...  cust:{$addToSet:"$customer_id"}
... }
... },
... { $project:
... {
.....   employeeId:1,
..... employeeName:1,
..... uniqueCount:{$size:"$compl"}  }
... },
...
... {$sort:
... {uniqueCount: -1}
... },
... {$limit : 5}
... ]).pretty()
[
  { _id: { employeeName: 'Ndzi Samuel Colombui' }, uniqueCount: 13 },
  { _id: { employeeName: 'Veera Jane Abdellah ' }, uniqueCount: 10 },
  {
    _id: { employeeName: 'Dietrich Marilyn  Jenna  ' },
    uniqueCount: 10
  },
  { _id: { employeeName: 'Gilles Denisa  S Alex' }, uniqueCount: 10 },
  { _id: { employeeName: 'Sadki Sandy Nore  ' }, uniqueCount: 9 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> _
```

QUERY & RESULT (PYTHON EXECUTION):

```
result3 = client['data225_lab2']['smd'].aggregate([
{'$match': {'complaint_id':{'$nin': [None, '']}}},
{'$group':{'_id': {'employeeId': '$employeeId','employeeName': {'$concat': ['$emp_first_name', '',
'$emp_middle_name', '', '$emp_last_name']}},
'compl': {'$addToSet': '$complaint_id'},
'cust': {'$addToSet': '$customer_id'}}},
{'$project': {'employeeId': 1,'employeeName': 1,'uniqueCount': {'$size': '$compl'}}},
{'$sort': {'uniqueCount': -1}},
{'$limit': 5}])
result_list3 =[]
for document3 in result3:
result_list3.append(document3)
s3=pd.DataFrame(data=result_list3)
s3.rename({'_id':'Employee Name'}, axis = 1, inplace = True)
scenario3_df=pd.DataFrame()
scenario3_df['Employee Name']=pd.DataFrame(s3['Employee Name'].tolist())
scenario3_df['Total Complaints Resolved']=s3['uniqueCount']
scenario3_df
```

SCENARIO 3 - TOP 5 EMPLOYEES WHO HAVE RESPONDED TO MOST COMPLAINTS

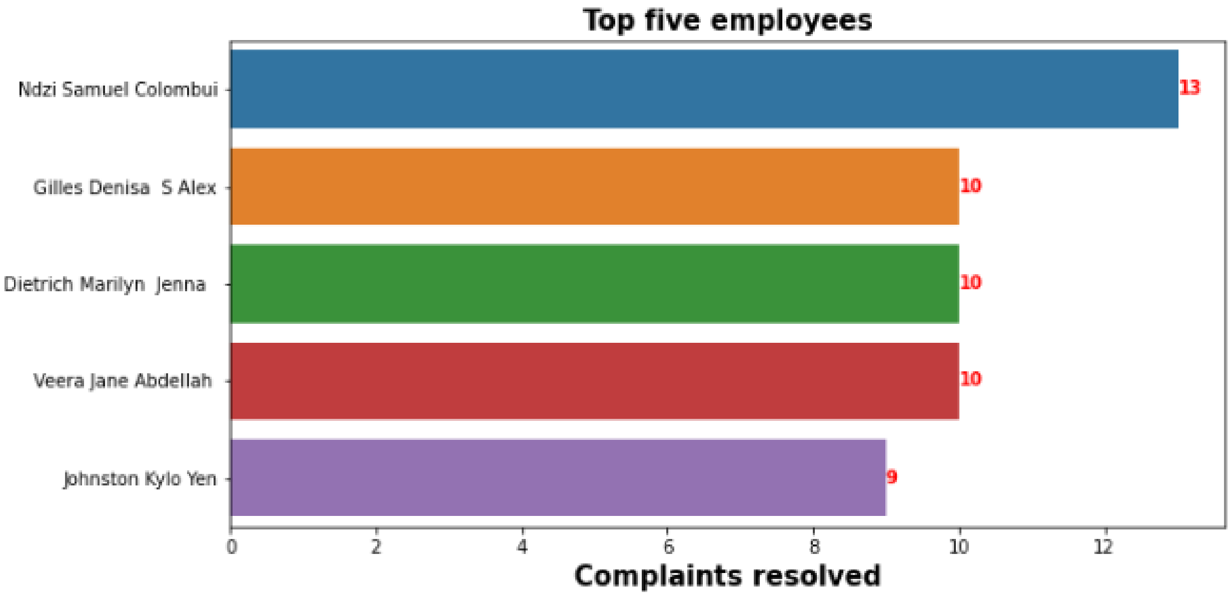


```python
result3 = client['data225_lab2']['smd'].aggregate([
    {'$match': {'complaint_id':{'$nin': [None, '']}}},
    {'$group':{'_id': {'employeeId': '$employeeId','employeeName': {'$concat': ['$emp_first_name', ' ', '$emp_middle_name
    'compl': {'$addToSet': '$complaint_id'},
    'cust': {'$addToSet': '$customer_id'}}},
    {'$project': {'employeeId': 1,'employeeName': 1,'uniqueCount': {'$size': '$compl'}}},
    {'$sort': {'uniqueCount': -1}},
    {'$limit': 5}])
result_list3 =[]
for document3 in result3:
    result_list3.append(document3)
s3=pd.DataFrame(data=result_list3)
s3.rename({'_id':'Employee Name'}, axis = 1, inplace = True)
scenario3_df=pd.DataFrame()
scenario3_df['Employee Name']=pd.DataFrame(s3['Employee Name'].tolist())
scenario3_df['Total Complaints Resolved']=s3['uniqueCount']
scenario3_df

```

13]:

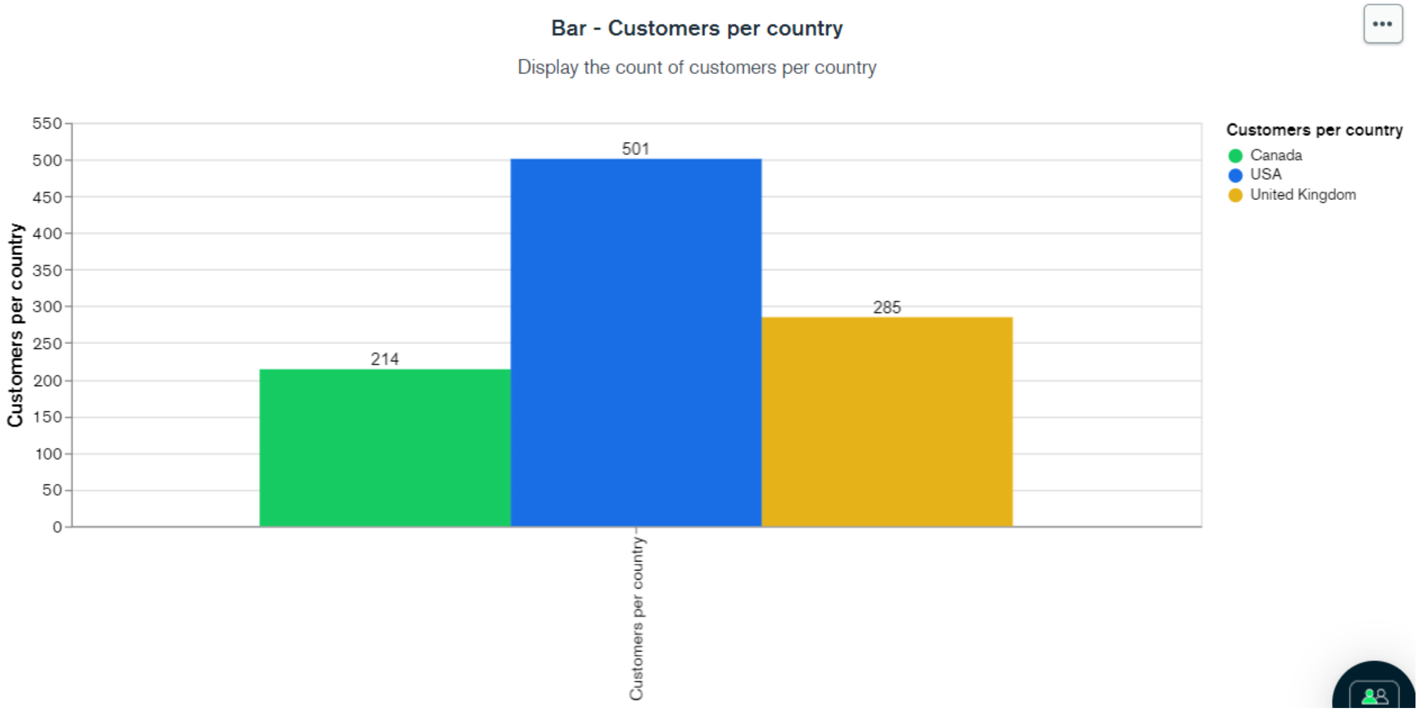
	Employee Name	Total Complaints Resolved
0	Ndzi Samuel Colombui	13
1	Gilles Denisa S Alex	10
2	Dietrich Marilyn Jenna	10
3	Veera Jane Abdellah	10
4	Johnston Kylo Yen	9

JUPYTER VISUALIZATION:



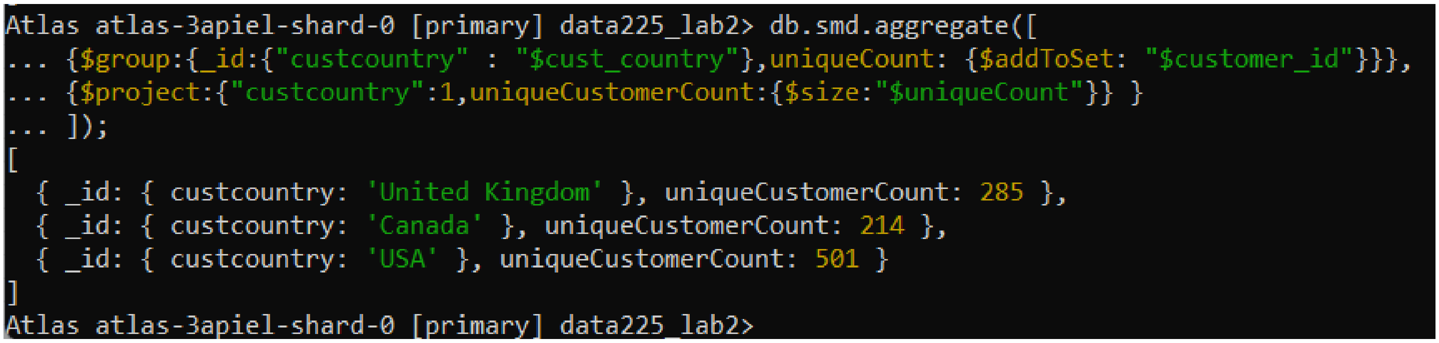
SCENARIO 4 - NO OF CUSTOMERS PER COUNTRY

MONGO CHARTS:



QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$group:{_id:{"custcountry" : "$cust_country"},uniqueCount: {$addToSet: "$customer_id"}}},
{$project:{"custcountry":1,uniqueCustomerCount:{$size:"$uniqueCount"}} }
]);
```

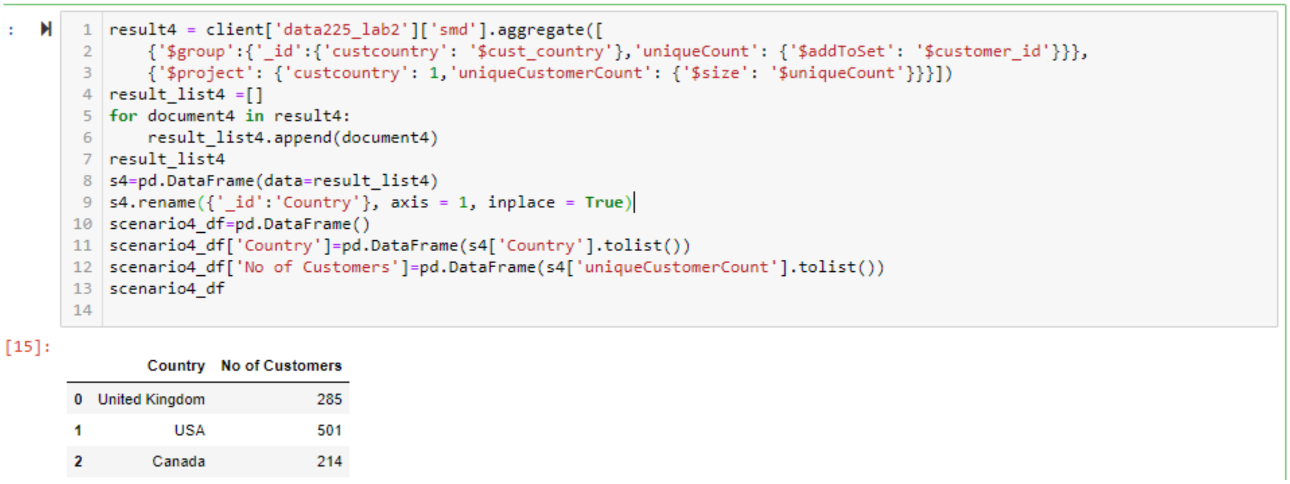


```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$group:{_id:{"custcountry" : "$cust_country"},uniqueCount: {$addToSet: "$customer_id"}}},
... {$project:{"custcountry":1,uniqueCustomerCount:{$size:"$uniqueCount"}} }
... ]);
[
  { _id: { custcountry: 'United Kingdom' }, uniqueCustomerCount: 285 },
  { _id: { custcountry: 'Canada' }, uniqueCustomerCount: 214 },
  { _id: { custcountry: 'USA' }, uniqueCustomerCount: 501 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2>
```

QUERY & RESULT (PYTHON EXECUTION):

```
result4 = client['data225_lab2']['smd'].aggregate([
{'$group':{'_id':{'custcountry': '$cust_country'},'uniqueCount': {'$addToSet': '$customer_id'}}},
{'$project': {'custcountry': 1,'uniqueCustomerCount': {'$size': '$uniqueCount'}}}])
result_list4 =[]
for document4 in result4:
result_list4.append(document4)
result_list4
s4=pd.DataFrame(data=result_list4)
s4.rename({'_id':'Country'}, axis = 1, inplace = True)
scenario4_df=pd.DataFrame()
scenario4_df['Country']=pd.DataFrame(s4['Country'].tolist())
scenario4_df['No of Customers']=pd.DataFrame(s4['uniqueCustomerCount'].tolist())
scenario4_df
```

SCENARIO 4 - NO OF CUSTOMERS PER COUNTRY

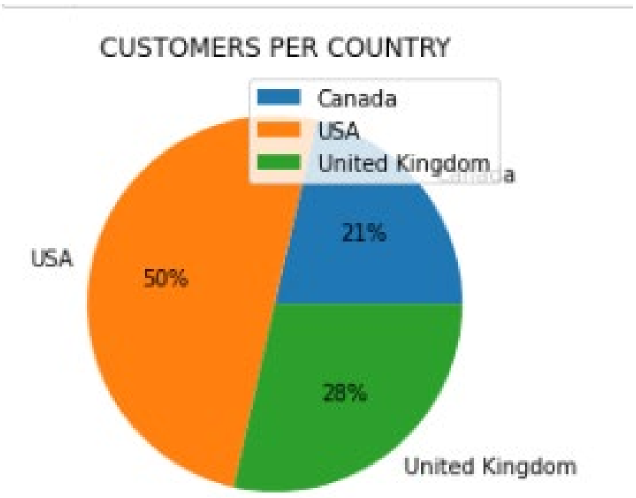


```
result4 = client['data225_lab2']['smd'].aggregate([
    {'$group':{'_id':{'custcountry': '$cust_country'},'uniqueCount': {'$addToSet': '$customer_id'}}},
    {'$project': {'custcountry': 1,'uniqueCustomerCount': {'$size': '$uniqueCount'}}}])
result_list4 =[]
for document4 in result4:
    result_list4.append(document4)
result_list4
s4=pd.DataFrame(data=result_list4)
s4.rename({'_id':'Country'}, axis = 1, inplace = True)
scenario4_df=pd.DataFrame()
scenario4_df['Country']=pd.DataFrame(s4['Country'].tolist())
scenario4_df['No of Customers']=pd.DataFrame(s4['uniqueCustomerCount'].tolist())
scenario4_df
```

[15]:

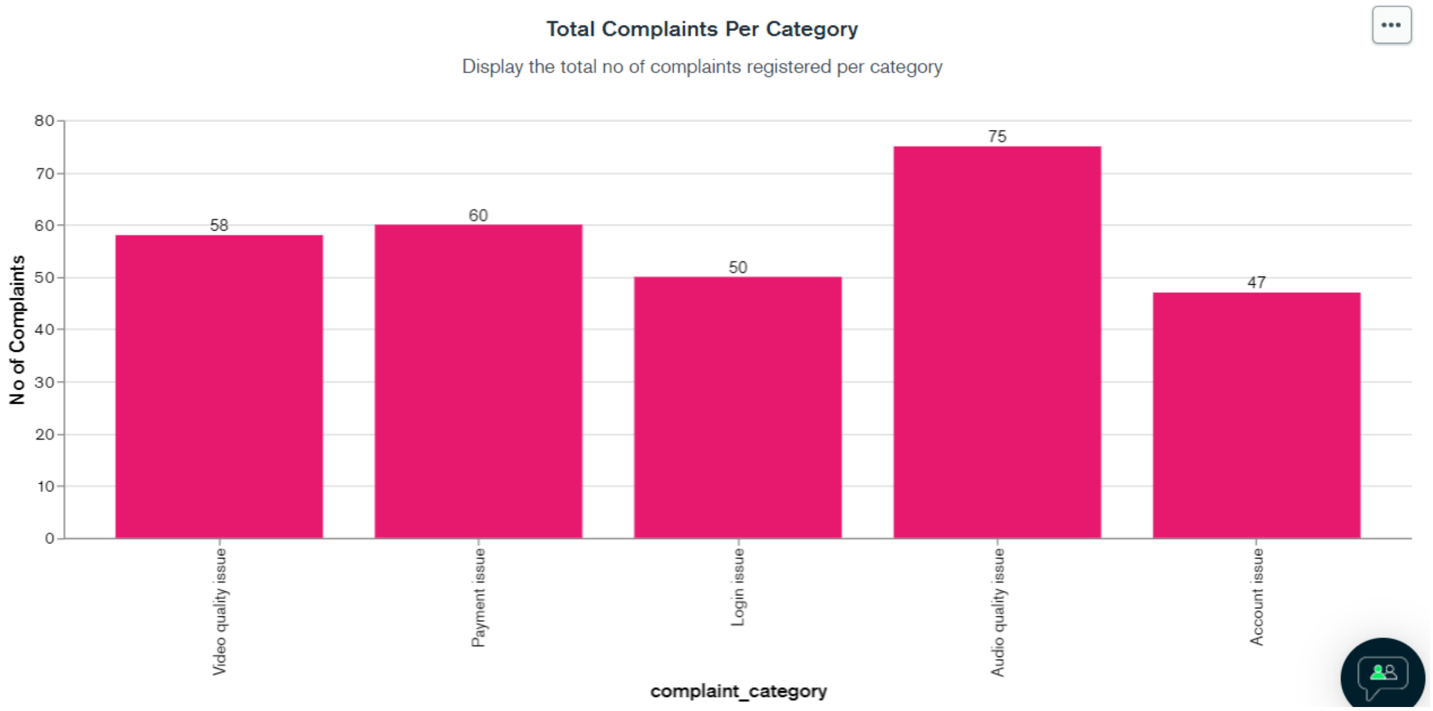
	Country	No of Customers
0	United Kingdom	285
1	USA	501
2	Canada	214

JUPYTER VISUALIZATION:



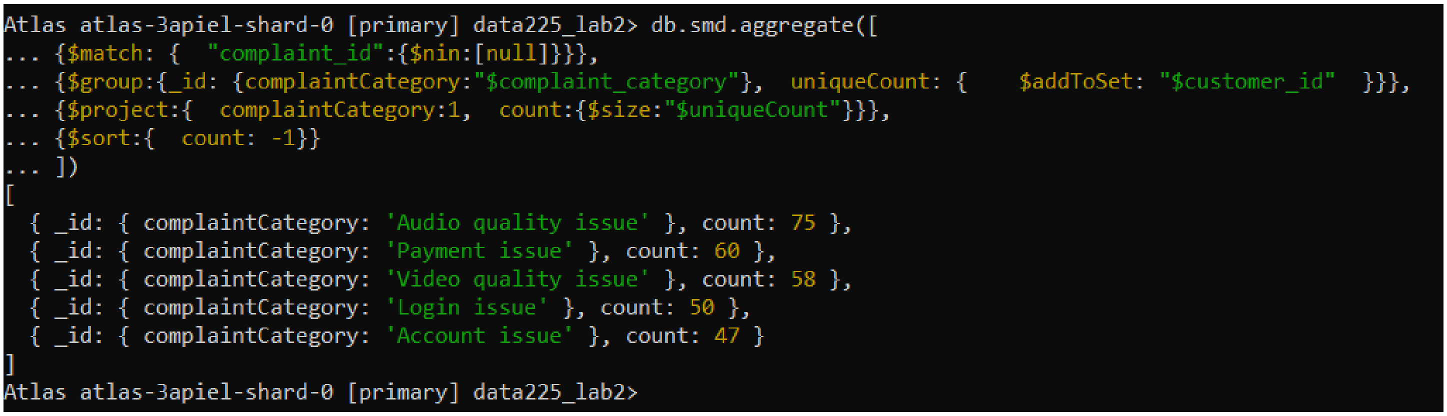
SCENARIO 5 - NO OF COMPLAINTS RECORDED PER COMPLAINT CATEGORY

MONGO CHARTS:



QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$match: {  "complaint_id":{$nin:[null]}}},
{$group:{_id: {complaintCategory:"$complaint_category"},  uniqueCount: {    $addToSet: "$customer_id"
}}},
{$project:{  complaintCategory:1,  count:{$size:"$uniqueCount"}}},
{$sort:{  count: -1}}
])
```

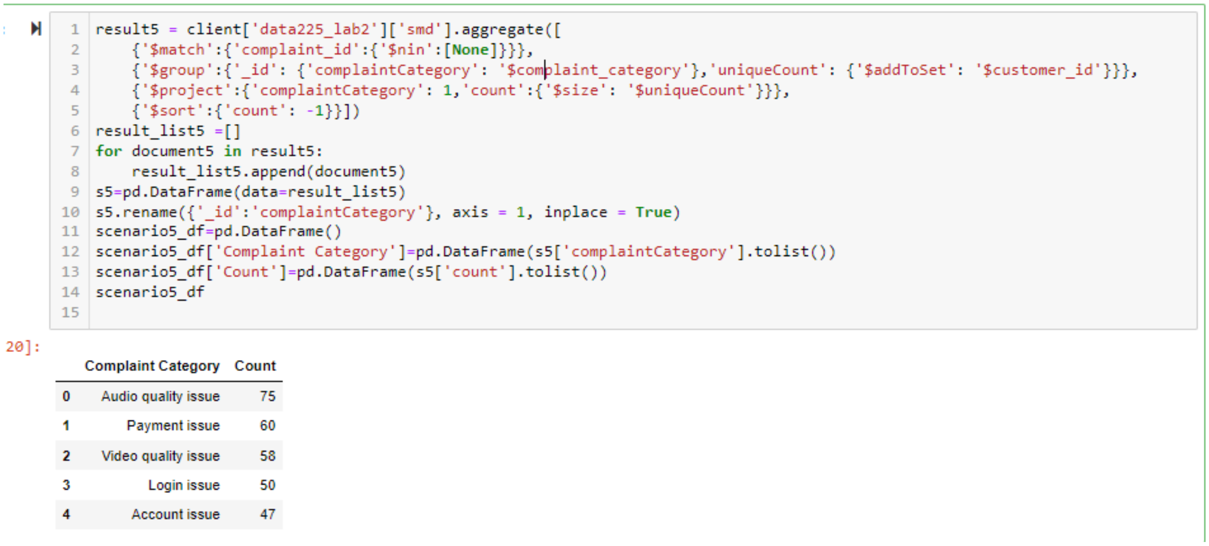


```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match: {  "complaint_id":{$nin:[null]}}},
... {$group:{_id: {complaintCategory:"$complaint_category"},  uniqueCount: {    $addToSet: "$customer_id"  }}},
... {$project:{  complaintCategory:1,  count:{$size:"$uniqueCount"}}},
... {$sort:{  count: -1}}
... ])
[
  { _id: { complaintCategory: 'Audio quality issue' }, count: 75 },
  { _id: { complaintCategory: 'Payment issue' }, count: 60 },
  { _id: { complaintCategory: 'Video quality issue' }, count: 58 },
  { _id: { complaintCategory: 'Login issue' }, count: 50 },
  { _id: { complaintCategory: 'Account issue' }, count: 47 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2>
```

QUERY & RESULT (PYTHON EXECUTION):

```
result5 = client['data225_lab2']['smd'].aggregate([
{'$match':{'complaint_id':{'$nin':[None]}}},
{'$group':{'_id': {'complaintCategory': '$complaint_category'},'uniqueCount': {'$addToSet': '$customer_id'}}},
{'$project':{'complaintCategory': 1,'count':{'$size': '$uniqueCount'}}},
{'$sort':{'count': -1}}])
result_list5 =[]
for document5 in result5:
result_list5.append(document5)
s5=pd.DataFrame(data=result_list5)
s5.rename({'_id':'complaintCategory'}, axis = 1, inplace = True)
scenario5_df=pd.DataFrame()
scenario5_df['Complaint Category']=pd.DataFrame(s5['complaintCategory'].tolist())
scenario5_df['Count']=pd.DataFrame(s5['count'].tolist())
scenario5_df
```

SCENARIO 5 - NO OF COMPLAINTS RECORDED PER COMPLAINT CATEGORY

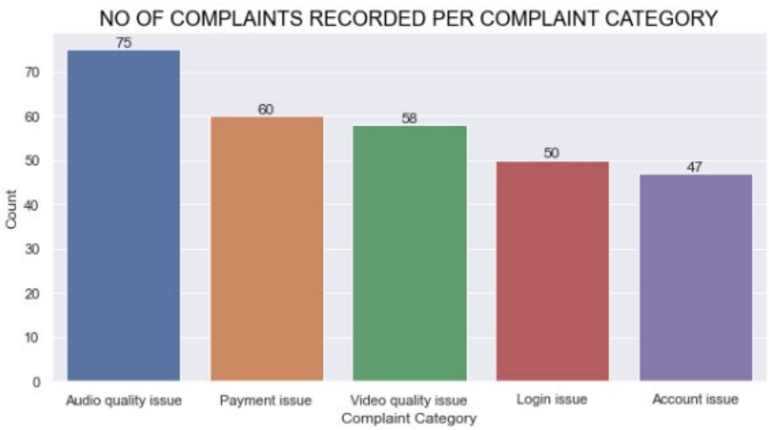


```
result5 = client['data225_lab2']['smd'].aggregate([
    {'$match':{'complaint_id':{'$nin':[None]}}},
    {'$group':{'_id': {'complaintCategory': '$complaint_category'},'uniqueCount': {'$addToSet': '$customer_id'}}},
    {'$project':{'complaintCategory': 1,'count':{'$size': '$uniqueCount'}}},
    {'$sort':{'count': -1}}])
result_list5 =[]
for document5 in result5:
    result_list5.append(document5)
s5=pd.DataFrame(data=result_list5)
s5.rename({'_id':'complaintCategory'}, axis = 1, inplace = True)
scenario5_df=pd.DataFrame()
scenario5_df['Complaint Category']=pd.DataFrame(s5['complaintCategory'].tolist())
scenario5_df['Count']=pd.DataFrame(s5['count'].tolist())
scenario5_df
```

20]:

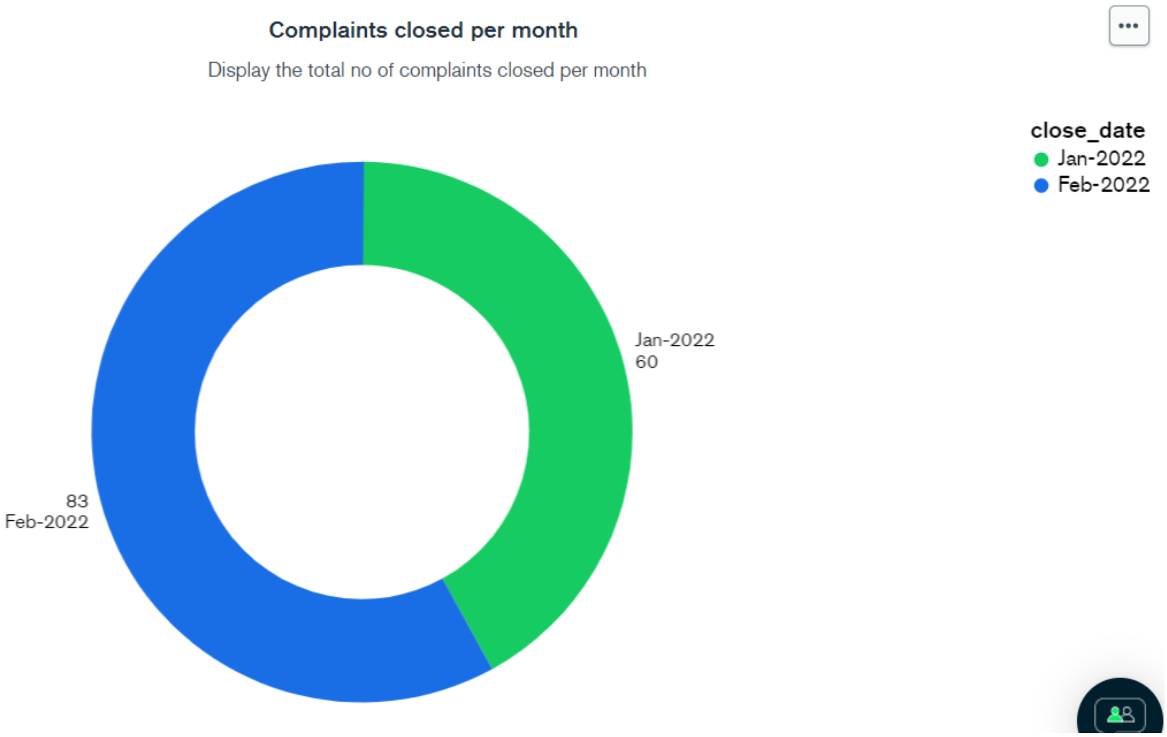
	Complaint Category	Count
0	Audio quality issue	75
1	Payment issue	60
2	Video quality issue	58
3	Login issue	50
4	Account issue	47

JUPYTER VISUALIZATION:



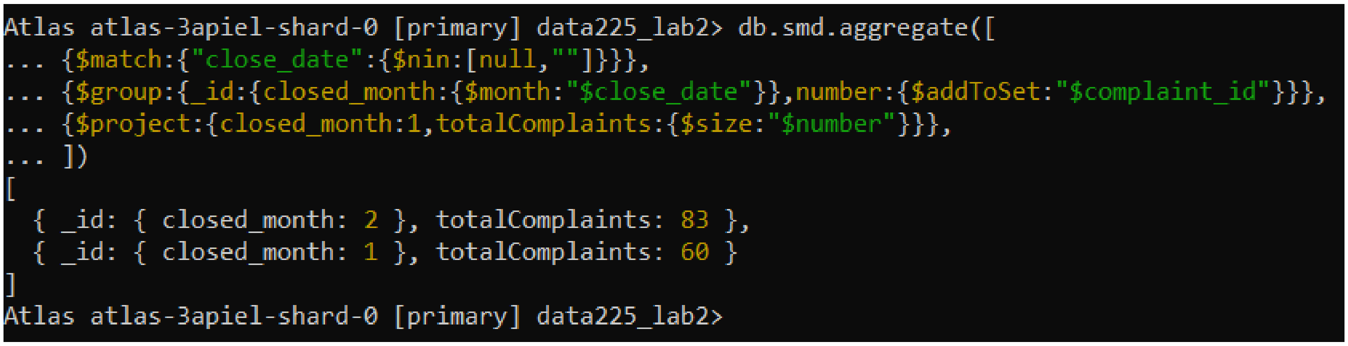
SCENARIO 6 - COMPLAINTS CLOSED PER MONTH

MONGO CHARTS:



QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$match:{"close_date":{$nin:[null,""]}}},
{$group:{_id:{closed_month:{$month:"$close_date"}},number:{$addToSet:"$complaint_id"}}},
{$project:{closed_month:1,totalComplaints:{$size:"$number"}}},
])
```

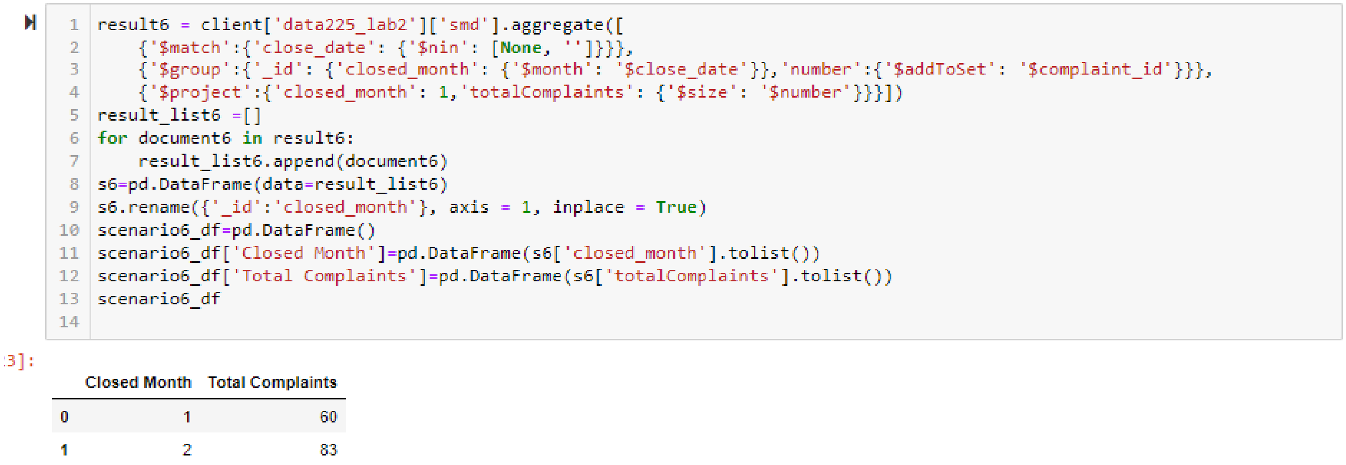


```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:{"close_date":{$nin:[null,""]}}},
... {$group:{_id:{closed_month:{$month:"$close_date"}},number:{$addToSet:"$complaint_id"}}},
... {$project:{closed_month:1,totalComplaints:{$size:"$number"}}},
... ])
[
  { _id: { closed_month: 2 }, totalComplaints: 83 },
  { _id: { closed_month: 1 }, totalComplaints: 60 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2>
```

QUERY & RESULT (PYTHON EXECUTION):

```
result6 = client['data225_lab2']['smd'].aggregate([
{'$match':{'close_date': {'$nin': [None, '']}}},
{'$group':{'_id': {'closed_month': {'$month': '$close_date'}},'number':{'$addToSet': '$complaint_id'}}},
{'$project':{'closed_month': 1,'totalComplaints': {'$size': '$number'}}}])
result_list6 =[]
for document6 in result6:
result_list6.append(document6)
s6=pd.DataFrame(data=result_list6)
s6.rename({'_id':'closed_month'}, axis = 1, inplace = True)
scenario6_df=pd.DataFrame()
scenario6_df['Closed Month']=pd.DataFrame(s6['closed_month'].tolist())
scenario6_df['Total Complaints']=pd.DataFrame(s6['totalComplaints'].tolist())
scenario6_df
```

SCENARIO 6 - COMPLAINTS CLOSED PER MONTH

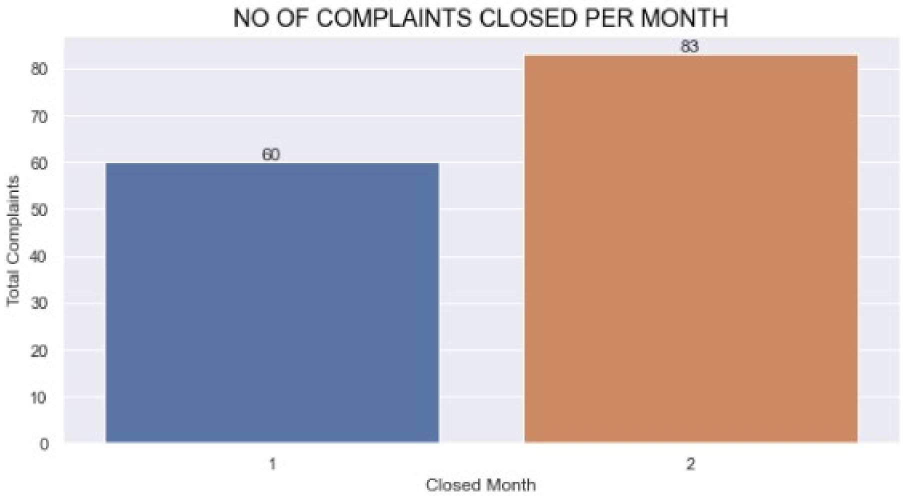


```
result6 = client['data225_lab2']['smd'].aggregate([
    {'$match':{'close_date': {'$nin': [None, '']}}},
    {'$group':{'_id': {'closed_month': {'$month': '$close_date'}},'number':{'$addToSet': '$complaint_id'}}},
    {'$project':{'closed_month': 1,'totalComplaints': {'$size': '$number'}}}])
result_list6 =[]
for document6 in result6:
    result_list6.append(document6)
s6=pd.DataFrame(data=result_list6)
s6.rename({'_id':'closed_month'}, axis = 1, inplace = True)
scenario6_df=pd.DataFrame()
scenario6_df['Closed Month']=pd.DataFrame(s6['closed_month'].tolist())
scenario6_df['Total Complaints']=pd.DataFrame(s6['totalComplaints'].tolist())
scenario6_df
```

3]:

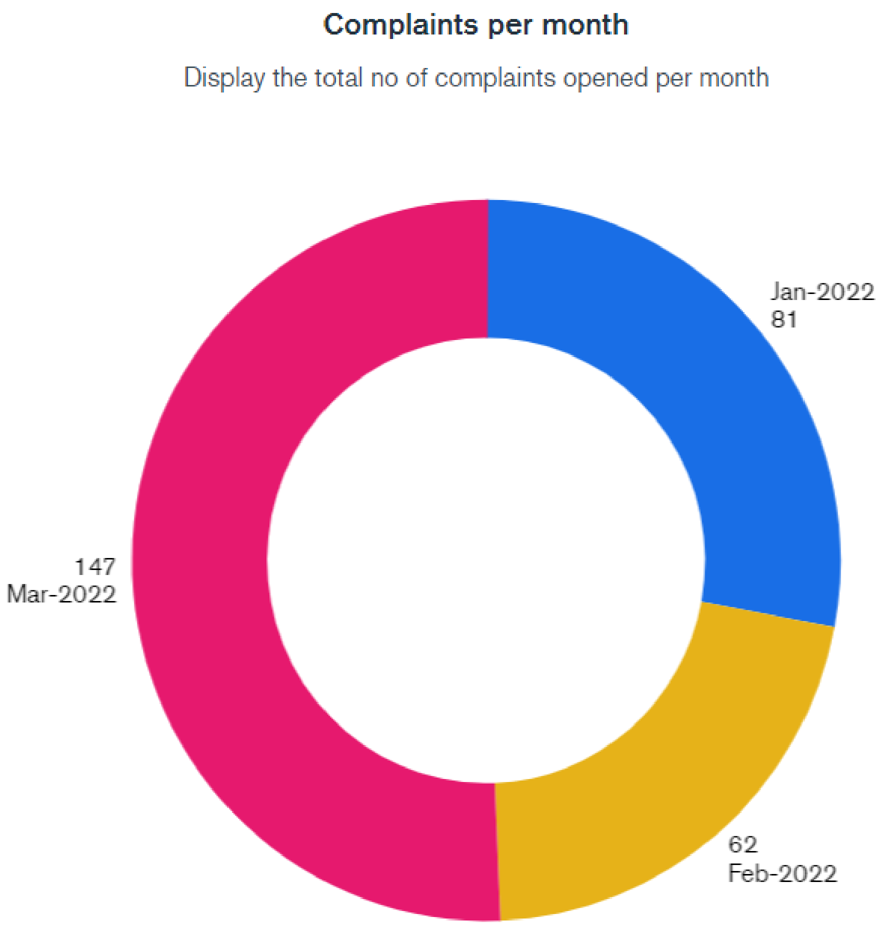
	Closed Month	Total Complaints
0	1	60
1	2	83

JUPYTER VISUALIZATION:



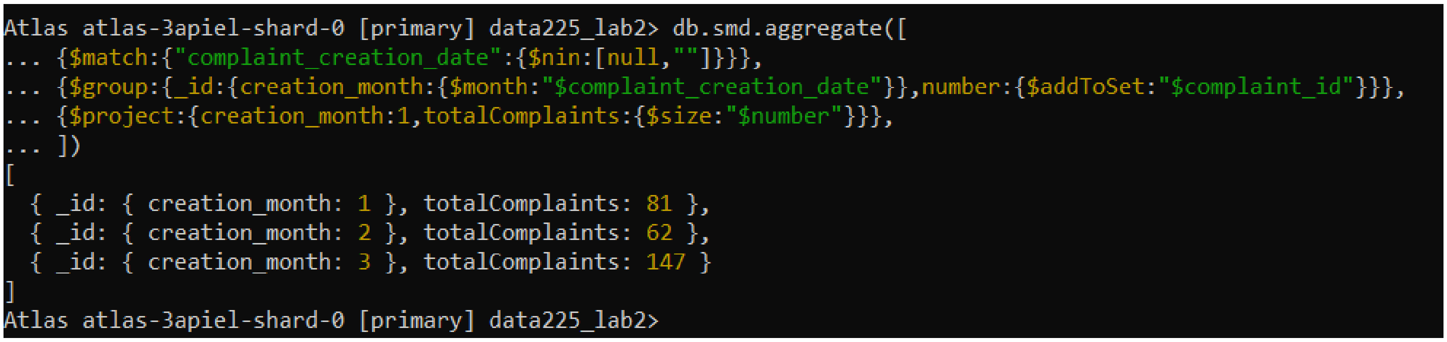
SCENARIO 7 – COMPLAINTS CREATED PER MONTH

MONGO CHARTS:



QUERY & RESULT (SHELL EXECUTION):

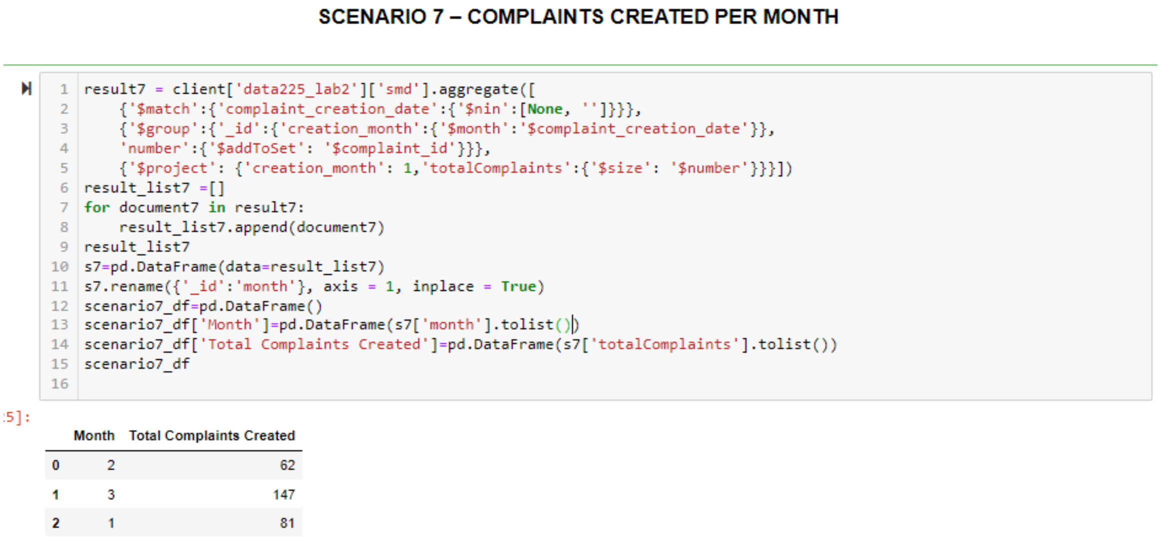
```
db.smd.aggregate([
{$match:{"complaint_creation_date":{$nin:[null,""]}}},
{$group:{_id:{creation_month:{$month:"$complaint_creation_date"}},number:{$addToSet:"$complaint_id"
}}},
{$project:{creation_month:1,totalComplaints:{$size:"$number"}}},
])
```



```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:{"complaint_creation_date":{$nin:[null,""]}}},
... {$group:{_id:{creation_month:{$month:"$complaint_creation_date"}},number:{$addToSet:"$complaint_id"}}},
... {$project:{creation_month:1,totalComplaints:{$size:"$number"}}},
... ])
[
  { _id: { creation_month: 1 }, totalComplaints: 81 },
  { _id: { creation_month: 2 }, totalComplaints: 62 },
  { _id: { creation_month: 3 }, totalComplaints: 147 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2>
```

QUERY & RESULT (PYTHON EXECUTION):

```
result7 = client['data225_lab2']['smd'].aggregate([
{'$match':{'complaint_creation_date':{'$nin':[None, '']}}},
{'$group':{'_id':{'creation_month':{'$month':'$complaint_creation_date'}},
'number':{'$addToSet': '$complaint_id'}}},
{'$project': {'creation_month': 1,'totalComplaints':{'$size': '$number'}}}])
result_list7 =[]
for document7 in result7:
result_list7.append(document7)
result_list7
s7=pd.DataFrame(data=result_list7)
s7.rename({'_id':'month'}, axis = 1, inplace = True)
scenario7_df=pd.DataFrame()
scenario7_df['Month']=pd.DataFrame(s7['month'].tolist())
scenario7_df['Total Complaints Created']=pd.DataFrame(s7['totalComplaints'].tolist())
scenario7_df
```



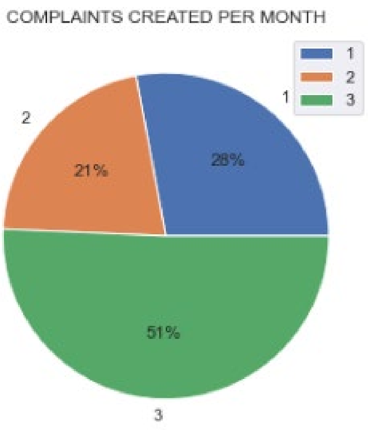
SCENARIO 7 – COMPLAINTS CREATED PER MONTH

```
result7 = client['data225_lab2']['smd'].aggregate([
    {'$match':{'complaint_creation_date':{'$nin':[None, '']}}},
    {'$group':{'_id':{'creation_month':{'$month':'$complaint_creation_date'}},
    'number':{'$addToSet': '$complaint_id'}}},
    {'$project': {'creation_month': 1,'totalComplaints':{'$size': '$number'}}}])
result_list7 =[]
for document7 in result7:
    result_list7.append(document7)
result_list7
s7=pd.DataFrame(data=result_list7)
s7.rename({'_id':'month'}, axis = 1, inplace = True)
scenario7_df=pd.DataFrame()
scenario7_df['Month']=pd.DataFrame(s7['month'].tolist())
scenario7_df['Total Complaints Created']=pd.DataFrame(s7['totalComplaints'].tolist())
scenario7_df
```

5]:

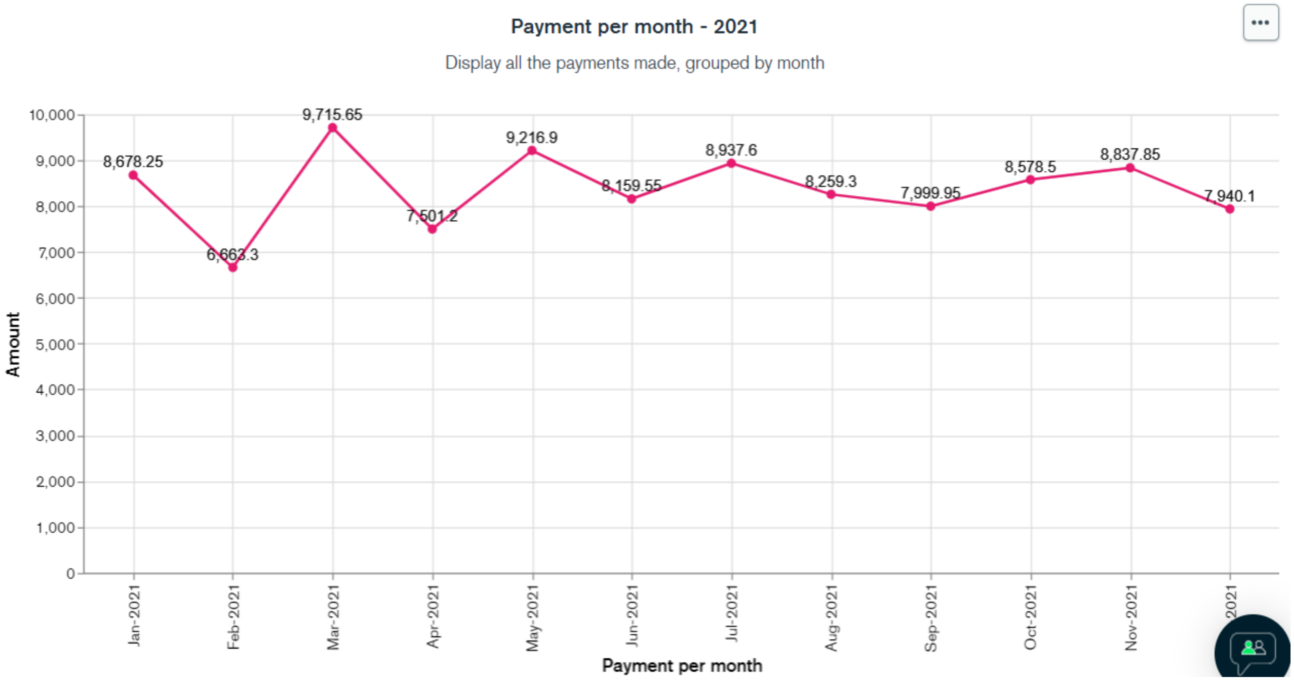
	Month	Total Complaints Created
0	2	62
1	3	147
2	1	81

JUPYTER VISUALIZATION:



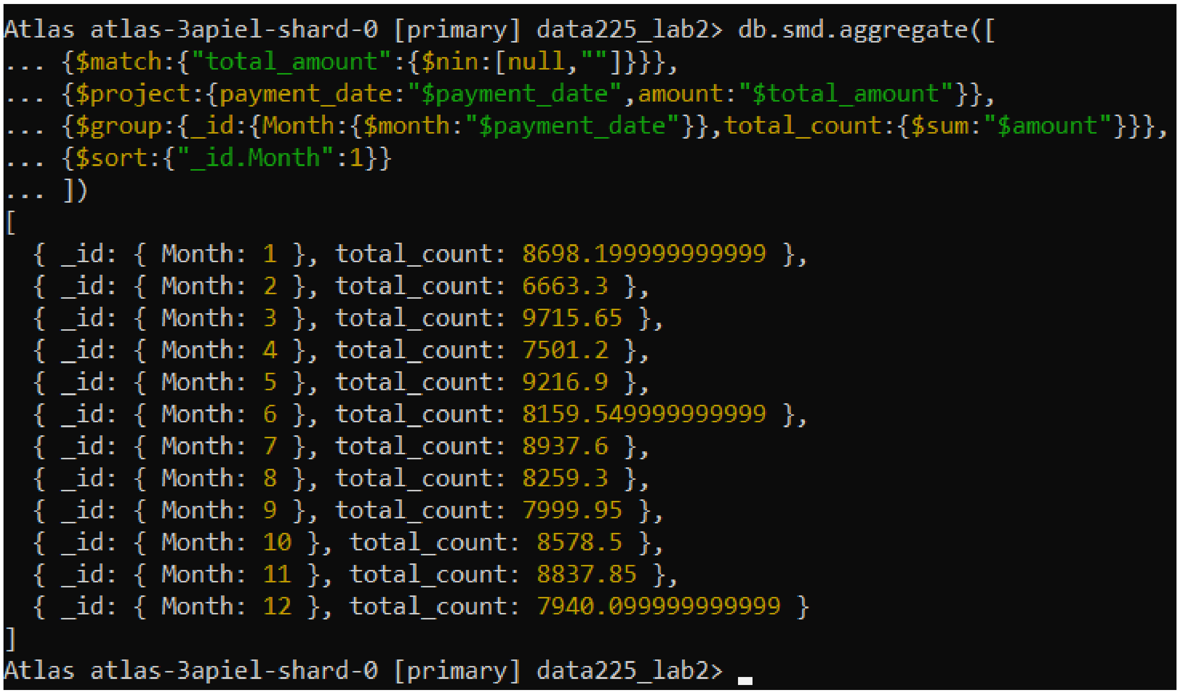
SCENARIO 8 - PAYMENT MADE PER MONTH

MONGO CHARTS:



QUERY & RESULT (SHELL EXECUTION):

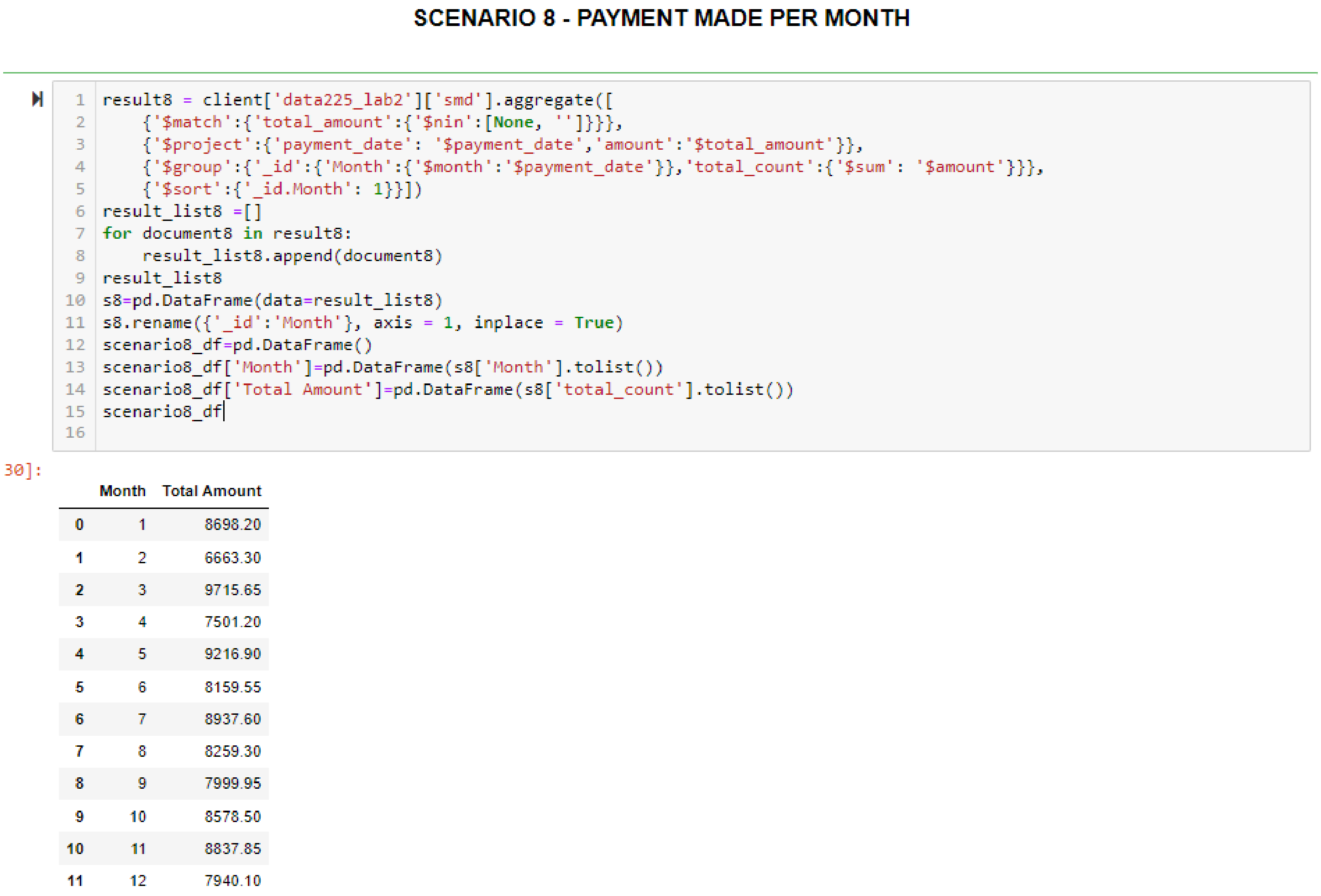
```
db.smd.aggregate([
{$match:{"total_amount":{$nin:[null,""]}}},
{$project:{payment_date:"$payment_date",amount:"$total_amount"}},
{$group:{_id:{Month:{$month:"$payment_date"}},total_count:{$sum:"$amount"}}},
{$sort:{"_id.Month":1}}
])
```



```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:{"total_amount":{$nin:[null,""]}}},
... {$project:{payment_date:"$payment_date",amount:"$total_amount"}},
... {$group:{_id:{Month:{$month:"$payment_date"}},total_count:{$sum:"$amount"}}},
... {$sort:{"_id.Month":1}}
... ])
[
  { _id: { Month: 1 }, total_count: 8698.199999999999 },
  { _id: { Month: 2 }, total_count: 6663.3 },
  { _id: { Month: 3 }, total_count: 9715.65 },
  { _id: { Month: 4 }, total_count: 7501.2 },
  { _id: { Month: 5 }, total_count: 9216.9 },
  { _id: { Month: 6 }, total_count: 8159.549999999999 },
  { _id: { Month: 7 }, total_count: 8937.6 },
  { _id: { Month: 8 }, total_count: 8259.3 },
  { _id: { Month: 9 }, total_count: 7999.95 },
  { _id: { Month: 10 }, total_count: 8578.5 },
  { _id: { Month: 11 }, total_count: 8837.85 },
  { _id: { Month: 12 }, total_count: 7940.099999999999 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2>
```

QUERY & RESULT (PYTHON EXECUTION):

```
result8 = client['data225_lab2']['smd'].aggregate([
{'$match':{'total_amount':{'$nin':[None, '']}}},
{'$project':{'payment_date': '$payment_date','amount':'$total_amount'}},
{'$group':{'_id':{'Month':{'$month':'$payment_date'}},'total_count':{'$sum': '$amount'}}},
{'$sort':{'_id.Month': 1}}])
result_list8 =[]
for document8 in result8:
result_list8.append(document8)
result_list8
s8=pd.DataFrame(data=result_list8)
s8.rename({'_id':'Month'}, axis = 1, inplace = True)
scenario8_df=pd.DataFrame()
scenario8_df['Month']=pd.DataFrame(s8['Month'].tolist())
scenario8_df['Total Amount']=pd.DataFrame(s8['total_count'].tolist())
scenario8_df
```



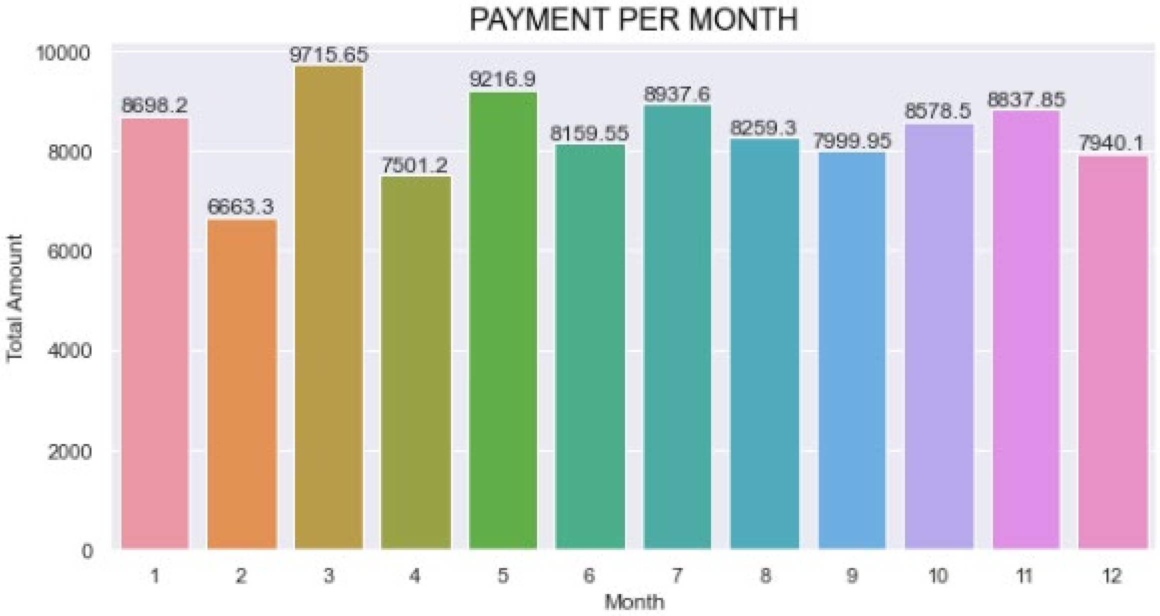
SCENARIO 8 - PAYMENT MADE PER MONTH

```
result8 = client['data225_lab2']['smd'].aggregate([
    {'$match':{'total_amount':{'$nin':[None, '']}}},
    {'$project':{'payment_date': '$payment_date','amount':'$total_amount'}},
    {'$group':{'_id':{'Month':{'$month':'$payment_date'}},'total_count':{'$sum': '$amount'}}},
    {'$sort':{'_id.Month': 1}}])
result_list8 =[]
for document8 in result8:
    result_list8.append(document8)
result_list8
s8=pd.DataFrame(data=result_list8)
s8.rename({'_id':'Month'}, axis = 1, inplace = True)
scenario8_df=pd.DataFrame()
scenario8_df['Month']=pd.DataFrame(s8['Month'].tolist())
scenario8_df['Total Amount']=pd.DataFrame(s8['total_count'].tolist())
scenario8_df
```

30]:

	Month	Total Amount
0	1	8698.20
1	2	6663.30
2	3	9715.65
3	4	7501.20
4	5	9216.90
5	6	8159.55
6	7	8937.60
7	8	8259.30
8	9	7999.95
9	10	8578.50
10	11	8837.85
11	12	7940.10

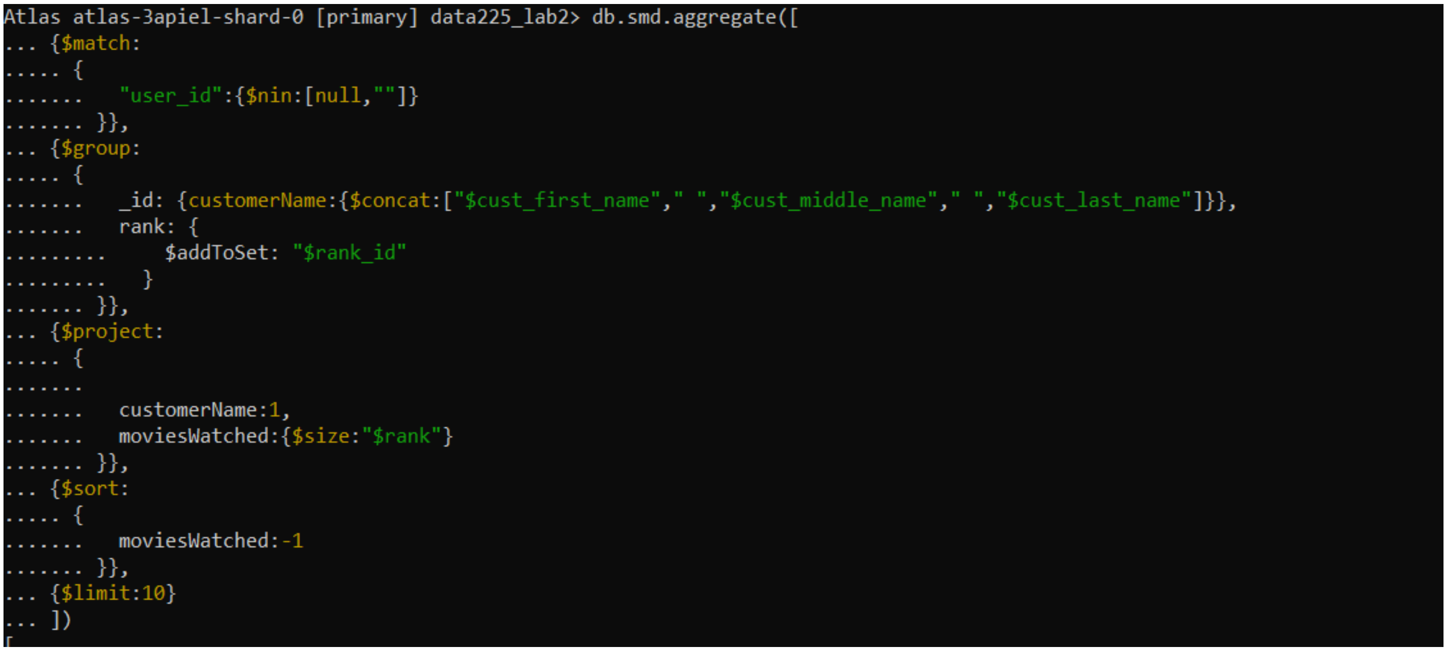
JUPYTER VISUALIZATION:



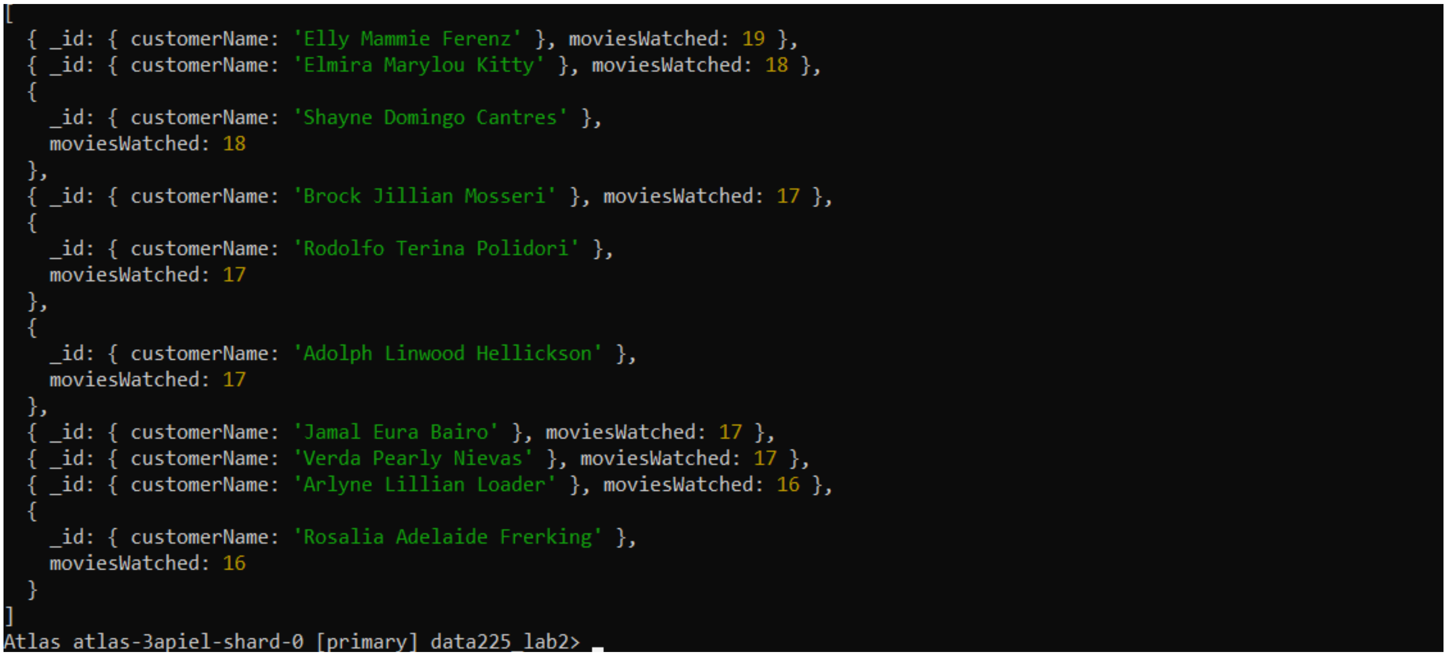
SCENARIO 9 - TOP 10 CUSTOMERS BASED ON MOVIES WATCHED

QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$match:{  "user_id":{$nin:[null,""]}}},
{$group:{  _id: {customerName:{$concat:["$cust_first_name"," ","$cust_middle_name","
","$cust_last_name"]}},  rank: {    $addToSet: "$rank_id" }}},
{$project:{ customerName:1,  moviesWatched:{$size:"$rank"}}},
{$sort:{  moviesWatched:-1}},
{$limit:10}
])
```



```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:
..... {
.......    "user_id":{$nin:[null,""]}
....... }},
... {$group:
..... {
.......    _id: {customerName:{$concat:["$cust_first_name"," ","$cust_middle_name"," ","$cust_last_name"]}},
.......    rank: {
.........       $addToSet: "$rank_id"
.........    }
....... }},
... {$project:
..... {
.......
.......    customerName:1,
.......    moviesWatched:{$size:"$rank"}
....... }},
... {$sort:
..... {
.......    moviesWatched:-1
....... }},
... {$limit:10}
... ])
[
```

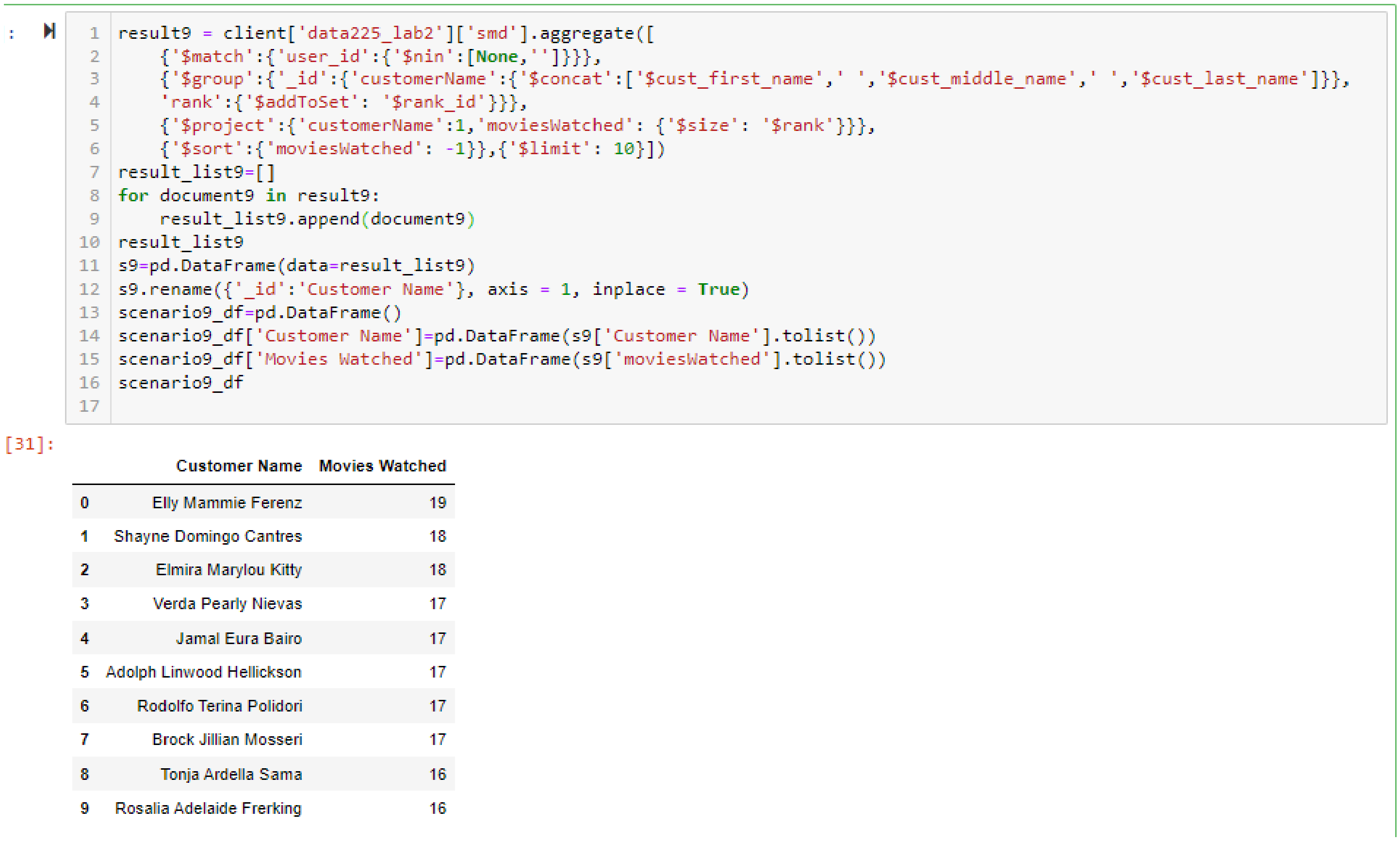


```
[
  { _id: { customerName: 'Elly Mammie Ferenz' }, moviesWatched: 19 },
  { _id: { customerName: 'Elmira Marylou Kitty' }, moviesWatched: 18 },
  {
    _id: { customerName: 'Shayne Domingo Cantres' },
    moviesWatched: 18
  },
  { _id: { customerName: 'Brock Jillian Mosseri' }, moviesWatched: 17 },
  {
    _id: { customerName: 'Rodolfo Terina Polidori' },
    moviesWatched: 17
  },
  {
    _id: { customerName: 'Adolph Linwood Hellickson' },
    moviesWatched: 17
  },
  { _id: { customerName: 'Jamal Eura Bairo' }, moviesWatched: 17 },
  { _id: { customerName: 'Verda Pearly Nievas' }, moviesWatched: 17 },
  { _id: { customerName: 'Arlyne Lillian Loader' }, moviesWatched: 16 },
  {
    _id: { customerName: 'Rosalia Adelaide Frerking' },
    moviesWatched: 16
  }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2>
```

QUERY & RESULT (PYTHON EXECUTION):

```
result9 = client['data225_lab2']['smd'].aggregate([
{'$match':{'user_id':{'$nin':[None,'']}}},
{'$group':{'_id':{'customerName':{'$concat':['$cust_first_name',' ','$cust_middle_name','
','$cust_last_name']}},
'rank':{'$addToSet': '$rank_id'}}},
{'$project':{'customerName':1,'moviesWatched': {'$size': '$rank'}}},
{'$sort':{'moviesWatched': -1}},{'$limit': 10}])
result_list9=[]
for document9 in result9:
result_list9.append(document9)
result_list9
s9=pd.DataFrame(data=result_list9)
s9.rename({'_id':'Customer Name'}, axis = 1, inplace = True)
scenario9_df=pd.DataFrame()
scenario9_df['Customer Name']=pd.DataFrame(s9['Customer Name'].tolist())
scenario9_df['Movies Watched']=pd.DataFrame(s9['moviesWatched'].tolist())
scenario9_df
```

SCENARIO 9 - TOP 10 CUSTOMERS BASED ON MOVIES WATCHED

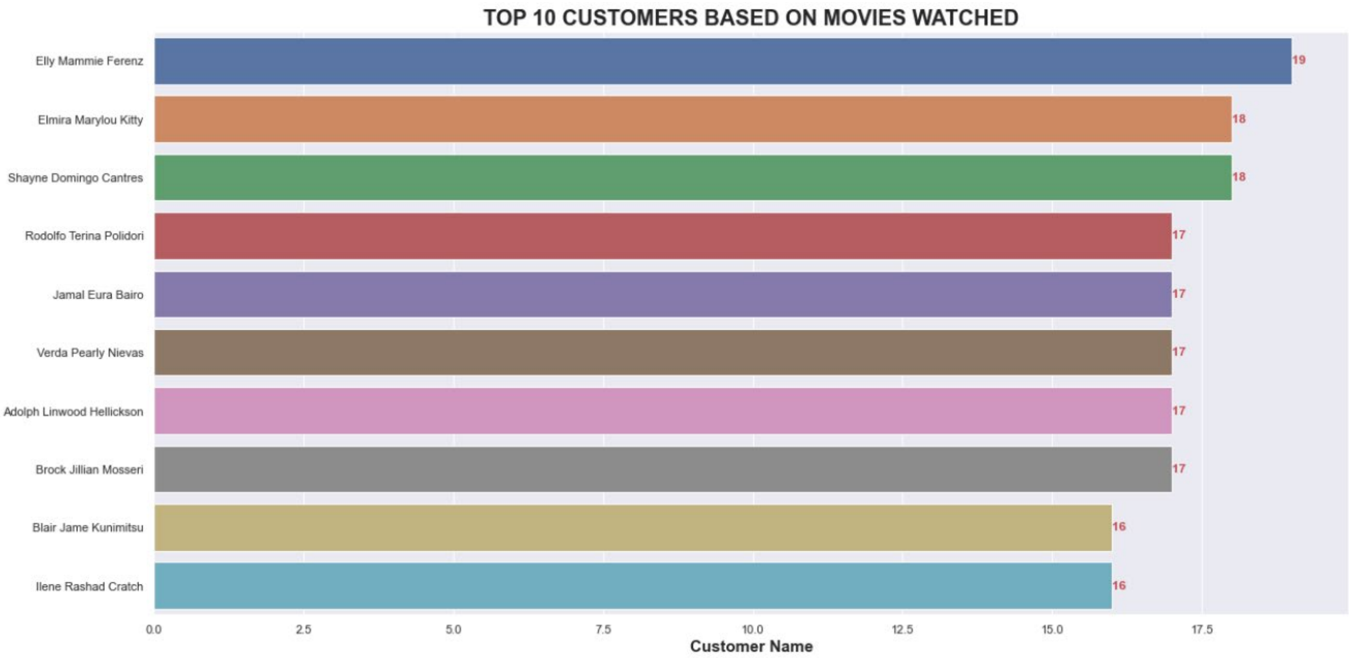


```
result9 = client['data225_lab2']['smd'].aggregate([
    {'$match':{'user_id':{'$nin':[None,'']}}},
    {'$group':{'_id':{'customerName':{'$concat':['$cust_first_name',' ','$cust_middle_name',' ','$cust_last_name']}},
    'rank':{'$addToSet': '$rank_id'}}},
    {'$project':{'customerName':1,'moviesWatched': {'$size': '$rank'}}},
    {'$sort':{'moviesWatched': -1}},{'$limit': 10}])
result_list9=[]
for document9 in result9:
    result_list9.append(document9)
result_list9
s9=pd.DataFrame(data=result_list9)
s9.rename({'_id':'Customer Name'}, axis = 1, inplace = True)
scenario9_df=pd.DataFrame()
scenario9_df['Customer Name']=pd.DataFrame(s9['Customer Name'].tolist())
scenario9_df['Movies Watched']=pd.DataFrame(s9['moviesWatched'].tolist())
scenario9_df
```

[31]:

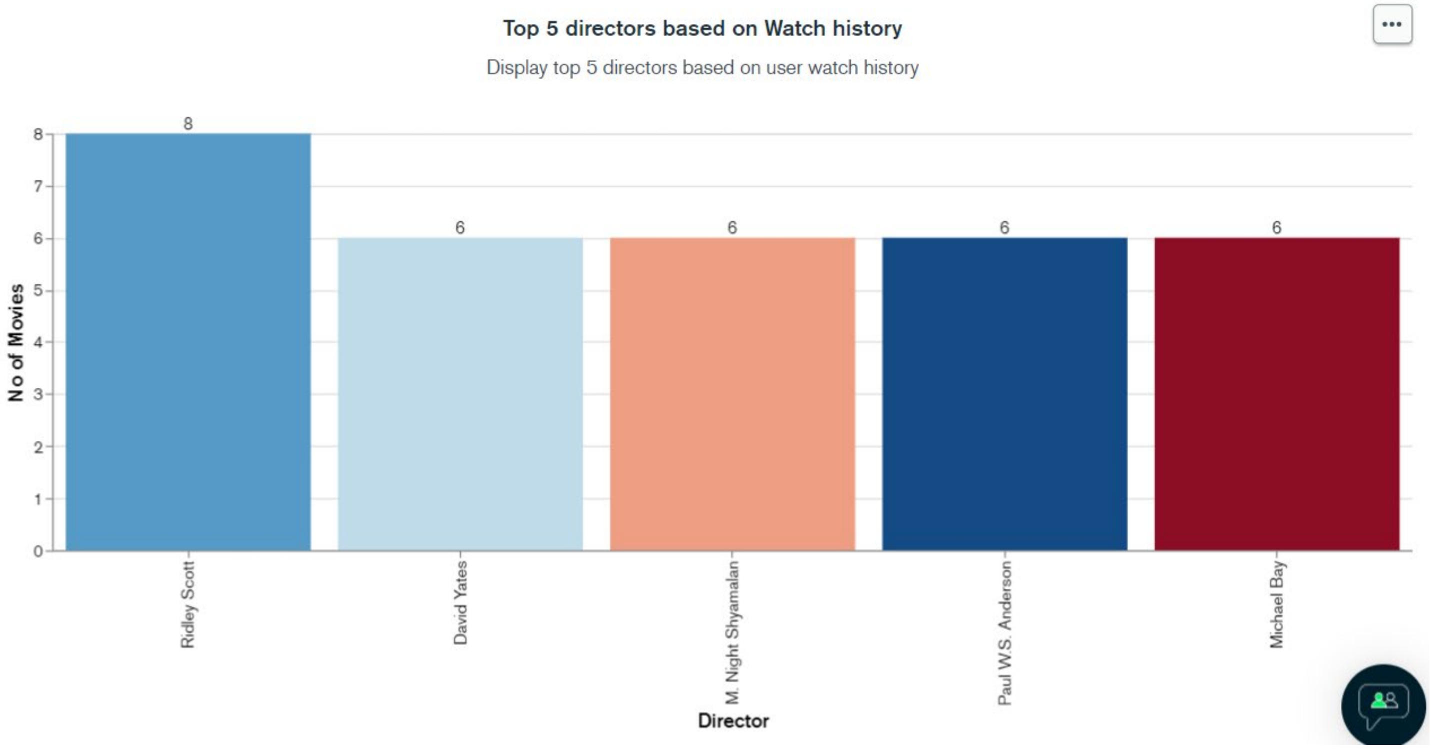
	Customer Name	Movies Watched
0	Elly Mammie Ferenz	19
1	Shayne Domingo Cantres	18
2	Elmira Marylou Kitty	18
3	Verda Pearly Nievas	17
4	Jamal Eura Bairo	17
5	Adolph Linwood Hellickson	17
6	Rodolfo Terina Polidori	17
7	Brock Jillian Mosseri	17
8	Tonja Ardella Sama	16
9	Rosalia Adelaide Frerking	16

JUPYTER VISUALIZATION:



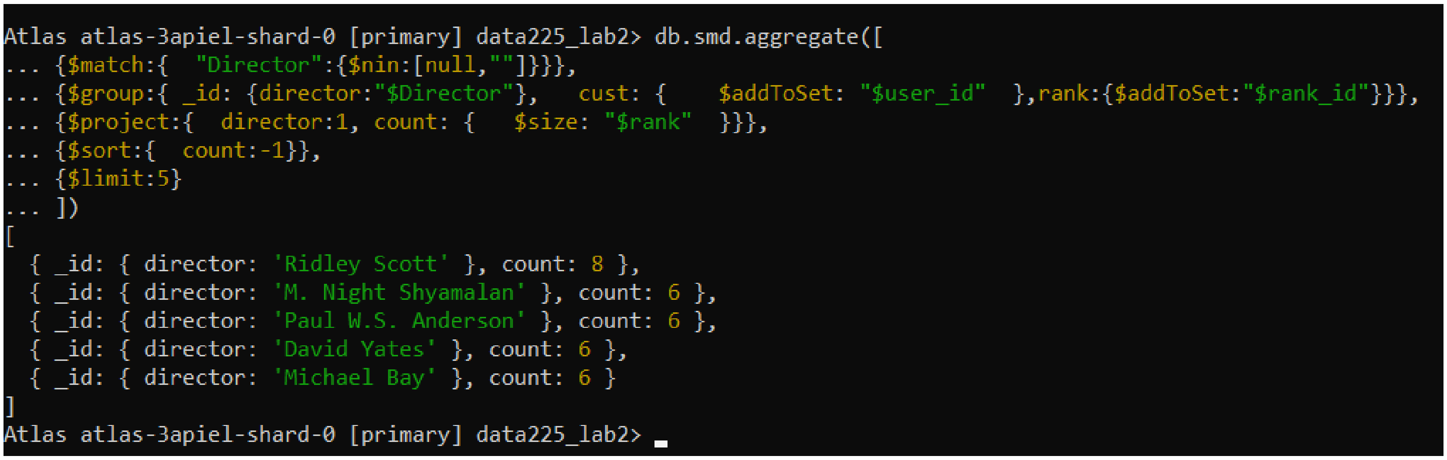
SCENARIO 10 - TOP 10 DIRECTORS BASED ON USER WATCH HISTORY

MONGO CHARTS:



QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$match:{  "Director":{$nin:[null,""]}}},
{$group:{ _id: {director:"$Director"},   cust: {    $addToSet: "$user_id"  },rank:{$addToSet:"$rank_id"}}},
{$project:{  director:1, count: {   $size: "$rank"  }}},
{$sort:{  count:-1}},
{$limit:5}
])
```

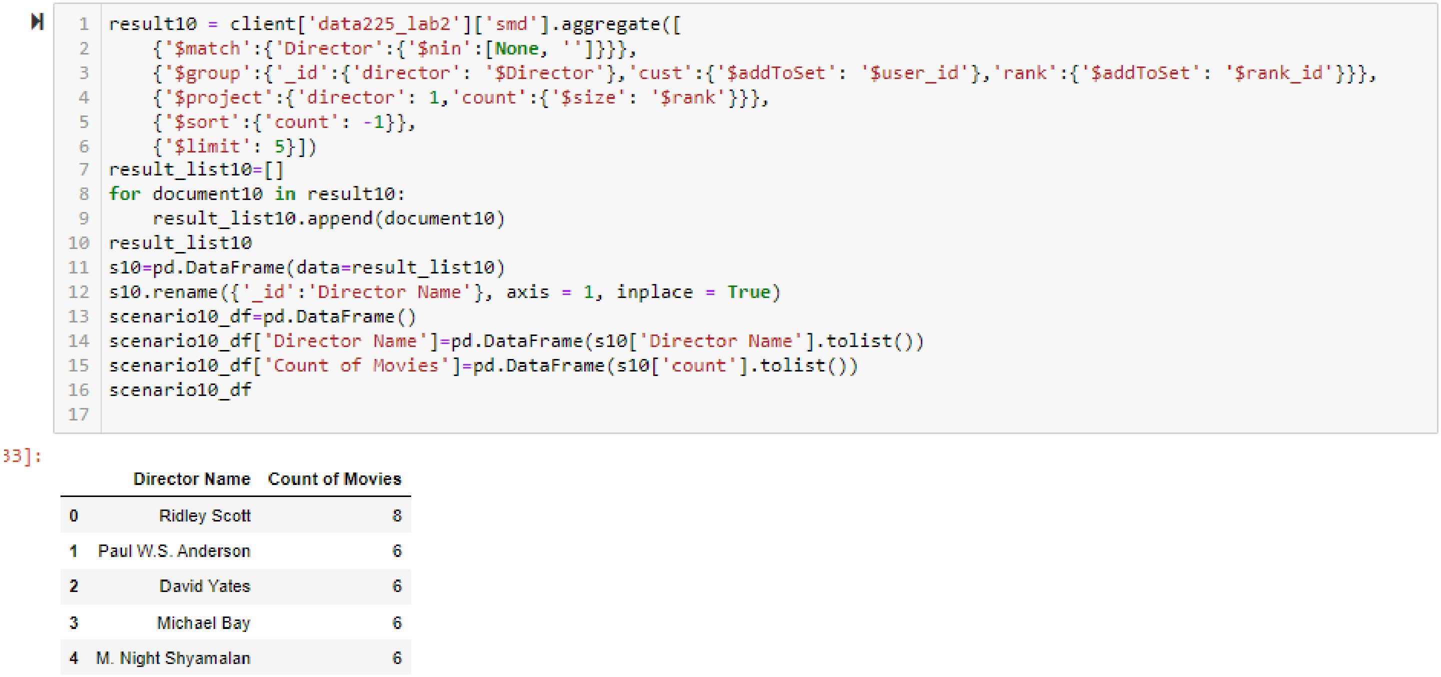


```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:{  "Director":{$nin:[null,""]}}},
... {$group:{ _id: {director:"$Director"},   cust: {    $addToSet: "$user_id"  },rank:{$addToSet:"$rank_id"}}},
... {$project:{  director:1, count: {   $size: "$rank"  }}},
... {$sort:{  count:-1}},
... {$limit:5}
... ])
[
  { _id: { director: 'Ridley Scott' }, count: 8 },
  { _id: { director: 'M. Night Shyamalan' }, count: 6 },
  { _id: { director: 'Paul W.S. Anderson' }, count: 6 },
  { _id: { director: 'David Yates' }, count: 6 },
  { _id: { director: 'Michael Bay' }, count: 6 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> _
```

QUERY & RESULT (PYTHON EXECUTION):

```
result10 = client['data225_lab2']['smd'].aggregate([
{'$match':{'Director':{'$nin':[None, '']}}},
{'$group':{'_id':{'director': '$Director'},'cust':{'$addToSet': '$user_id'},'rank':{'$addToSet': '$rank_id'}}},
{'$project':{'director': 1,'count':{'$size': '$rank'}}},
{'$sort':{'count': -1}},
{'$limit': 5}])
result_list10=[]
for document10 in result10:
result_list10.append(document10)
result_list10
s10=pd.DataFrame(data=result_list10)
s10.rename({'_id':'Director Name'}, axis = 1, inplace = True)
scenario10_df=pd.DataFrame()
scenario10_df['Director Name']=pd.DataFrame(s10['Director Name'].tolist())
scenario10_df['Count of Movies']=pd.DataFrame(s10['count'].tolist())
scenario10_df
```

SCENARIO 10 - TOP 5 DIRECTORS BASED ON USER WATCH HISTORY

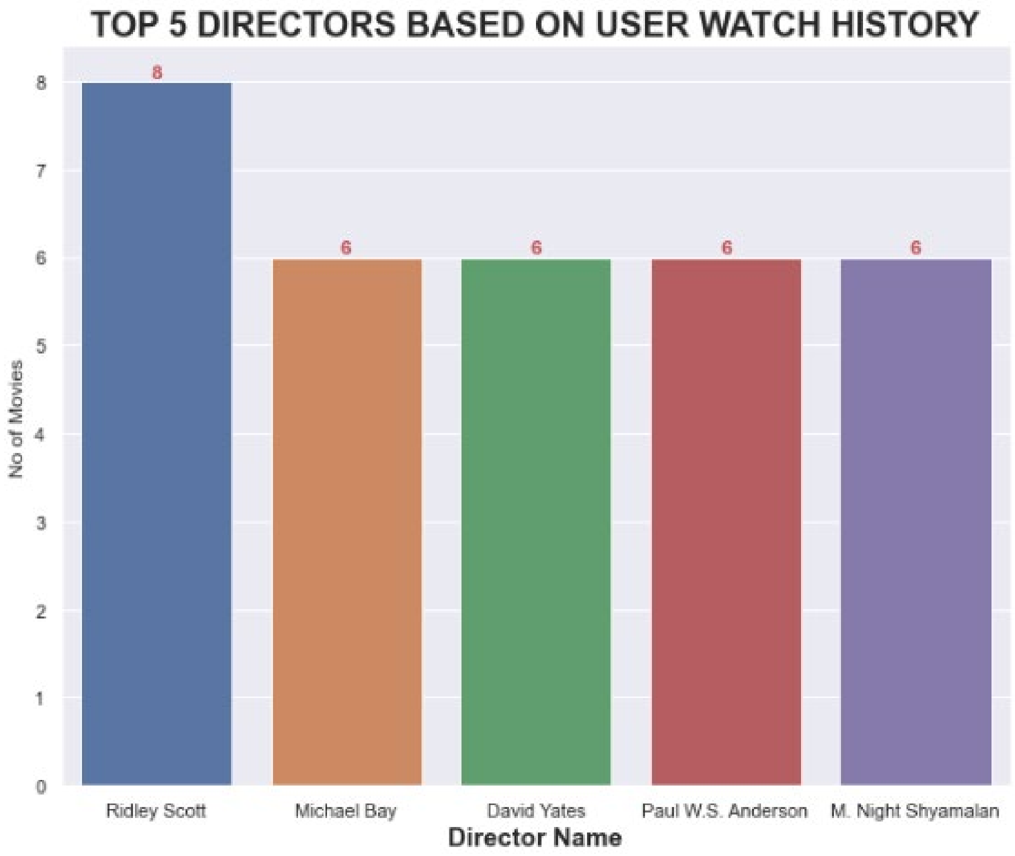


```
result10 = client['data225_lab2']['smd'].aggregate([
    {'$match':{'Director':{'$nin':[None, '']}}},
    {'$group':{'_id':{'director': '$Director'},'cust':{'$addToSet': '$user_id'},'rank':{'$addToSet': '$rank_id'}}},
    {'$project':{'director': 1,'count':{'$size': '$rank'}}},
    {'$sort':{'count': -1}},
    {'$limit': 5}])
result_list10=[]
for document10 in result10:
    result_list10.append(document10)
result_list10
s10=pd.DataFrame(data=result_list10)
s10.rename({'_id':'Director Name'}, axis = 1, inplace = True)
scenario10_df=pd.DataFrame()
scenario10_df['Director Name']=pd.DataFrame(s10['Director Name'].tolist())
scenario10_df['Count of Movies']=pd.DataFrame(s10['count'].tolist())
scenario10_df
```

33]:

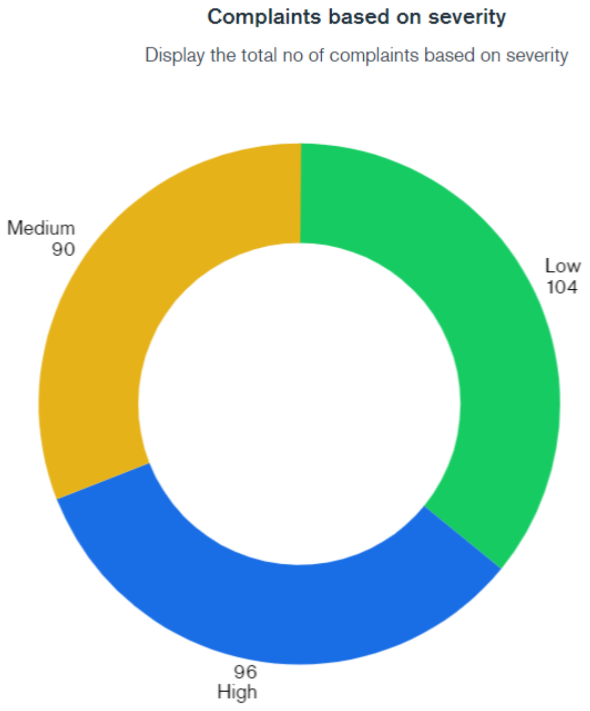
	Director Name	Count of Movies
0	Ridley Scott	8
1	Paul W.S. Anderson	6
2	David Yates	6
3	Michael Bay	6
4	M. Night Shyamalan	6

JUPYTER VISUALIZATION:



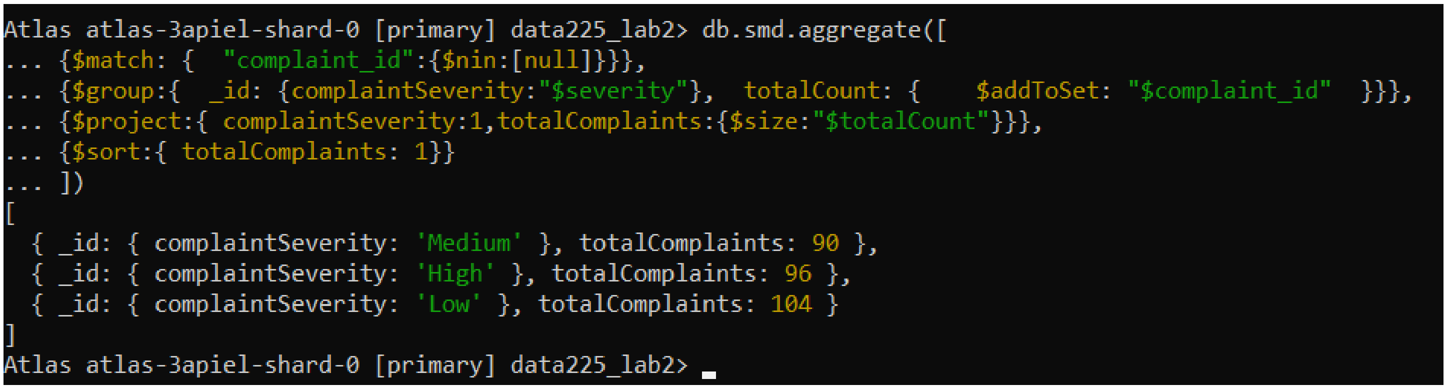
SCENARIO 11 - COMPLAINTS BASED ON SEVERITY

MONGO CHARTS:



QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$match: {  "complaint_id":{$nin:[null]}}},
{$group:{  _id: {complaintSeverity:"$severity"},  totalCount: {    $addToSet: "$complaint_id"  }}},
{$project:{ complaintSeverity:1,totalComplaints:{$size:"$totalCount"}}},
{$sort:{ totalComplaints: 1}}
])
```

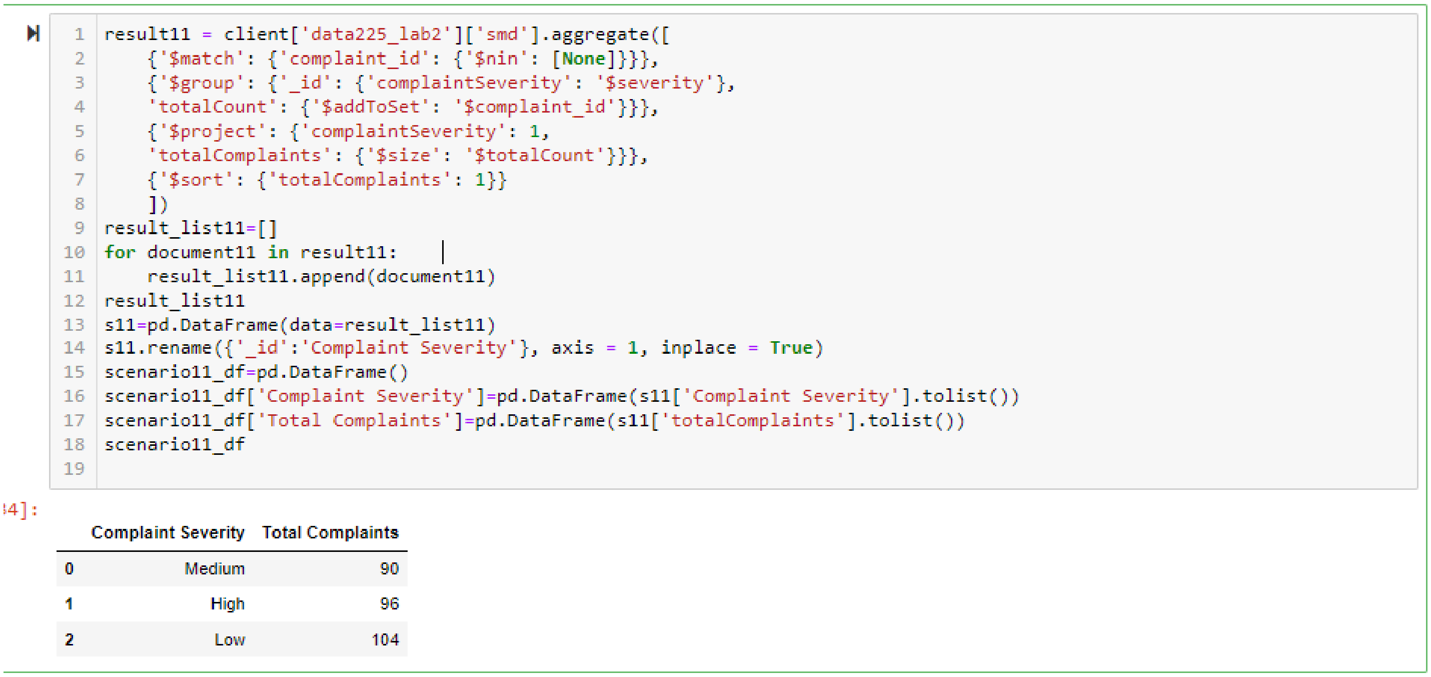


```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match: {  "complaint_id":{$nin:[null]}}},
... {$group:{  _id: {complaintSeverity:"$severity"},  totalCount: {    $addToSet: "$complaint_id"  }}},
... {$project:{ complaintSeverity:1,totalComplaints:{$size:"$totalCount"}}},
... {$sort:{ totalComplaints: 1}}
... ])
[
  { _id: { complaintSeverity: 'Medium' }, totalComplaints: 90 },
  { _id: { complaintSeverity: 'High' }, totalComplaints: 96 },
  { _id: { complaintSeverity: 'Low' }, totalComplaints: 104 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> _
```

QUERY & RESULT (PYTHON EXECUTION):

```
result11 = client['data225_lab2']['smd'].aggregate([
{'$match': {'complaint_id': {'$nin': [None]}}},
{'$group': {'_id': {'complaintSeverity': '$severity'},
'totalCount': {'$addToSet': '$complaint_id'}}},
{'$project': {'complaintSeverity': 1,
'totalComplaints': {'$size': '$totalCount'}}},
{'$sort': {'totalComplaints': 1}}
])
result_list11=[]
for document11 in result11:
result_list11.append(document11)
result_list11
s11=pd.DataFrame(data=result_list11)
s11.rename({'_id':'Complaint Severity'}, axis = 1, inplace = True)
scenario11_df=pd.DataFrame()
scenario11_df['Complaint Severity']=pd.DataFrame(s11['Complaint Severity'].tolist())
scenario11_df['Total Complaints']=pd.DataFrame(s11['totalComplaints'].tolist())
scenario11_df
```

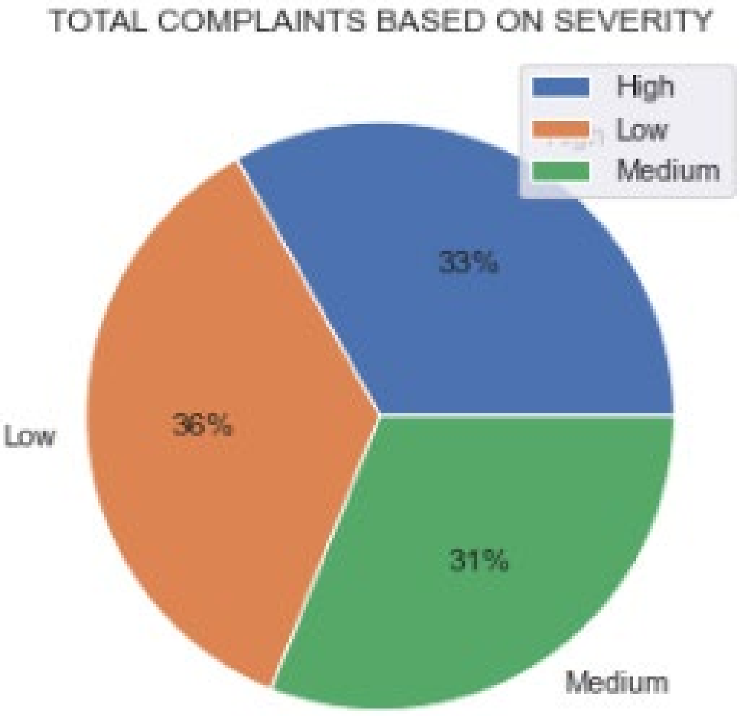
SCENARIO 11 - COMPLAINTS BASED ON SEVERITY



```
result11 = client['data225_lab2']['smd'].aggregate([
    {'$match': {'complaint_id': {'$nin': [None]}}},
    {'$group': {'_id': {'complaintSeverity': '$severity'},
    'totalCount': {'$addToSet': '$complaint_id'}}},
    {'$project': {'complaintSeverity': 1,
    'totalComplaints': {'$size': '$totalCount'}}},
    {'$sort': {'totalComplaints': 1}}
    ])
result_list11=[]
for document11 in result11:
    result_list11.append(document11)
result_list11
s11=pd.DataFrame(data=result_list11)
s11.rename({'_id':'Complaint Severity'}, axis = 1, inplace = True)
scenario11_df=pd.DataFrame()
scenario11_df['Complaint Severity']=pd.DataFrame(s11['Complaint Severity'].tolist())
scenario11_df['Total Complaints']=pd.DataFrame(s11['totalComplaints'].tolist())
scenario11_df

```

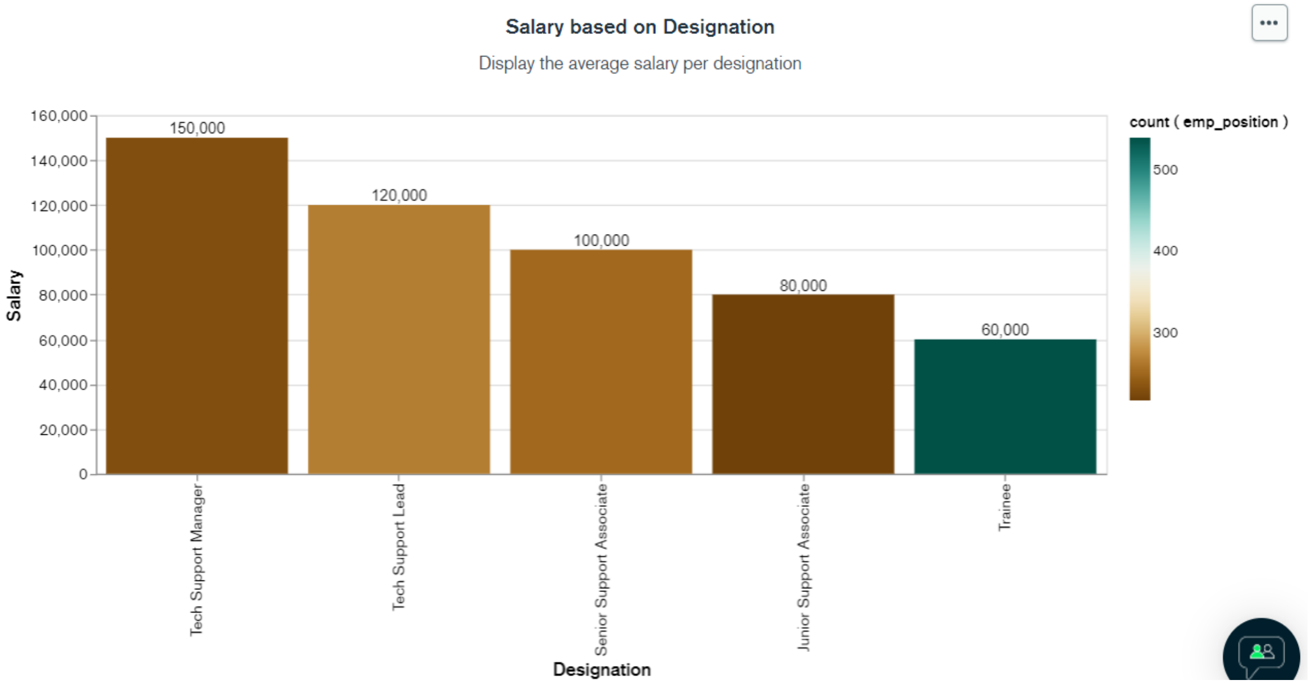
	Complaint Severity	Total Complaints
0	Medium	90
1	High	96
2	Low	104

JUPYTER VISUALIZATION:



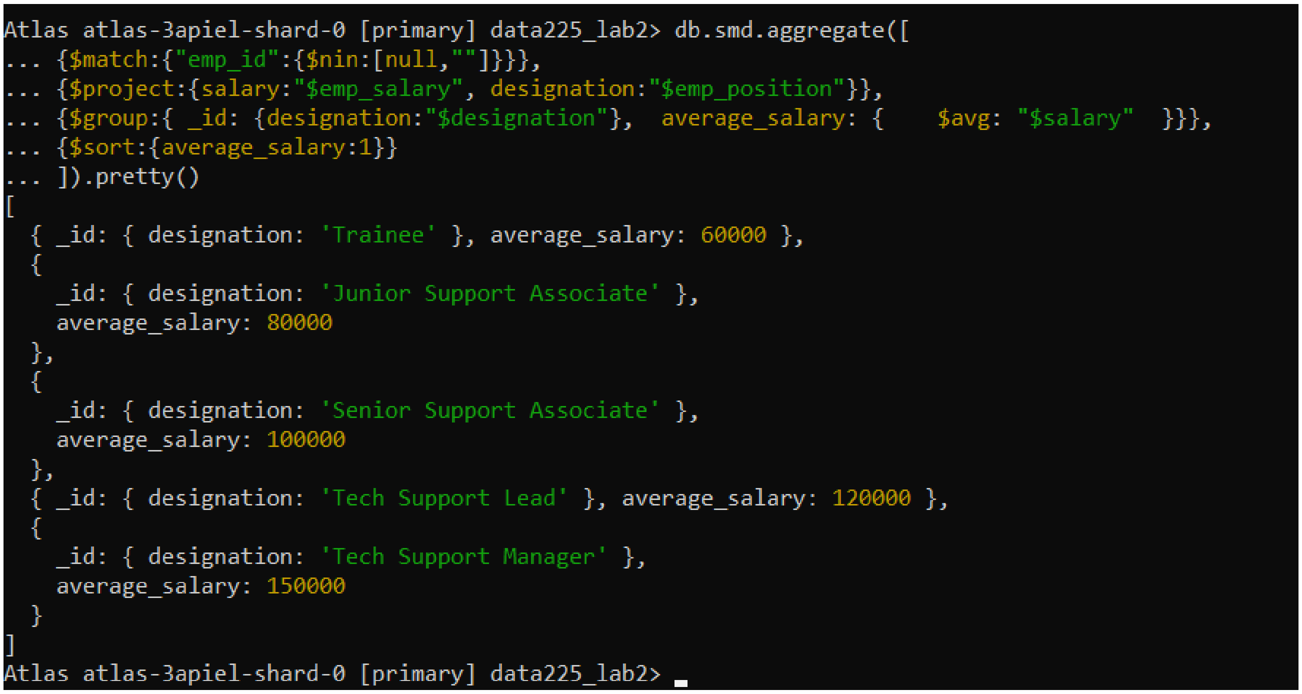
SCENARIO 12 - AVERAGE SALARY BASED ON DESIGNATION

MONGO CHARTS:



QUERY & RESULT (SHELL EXECUTION):

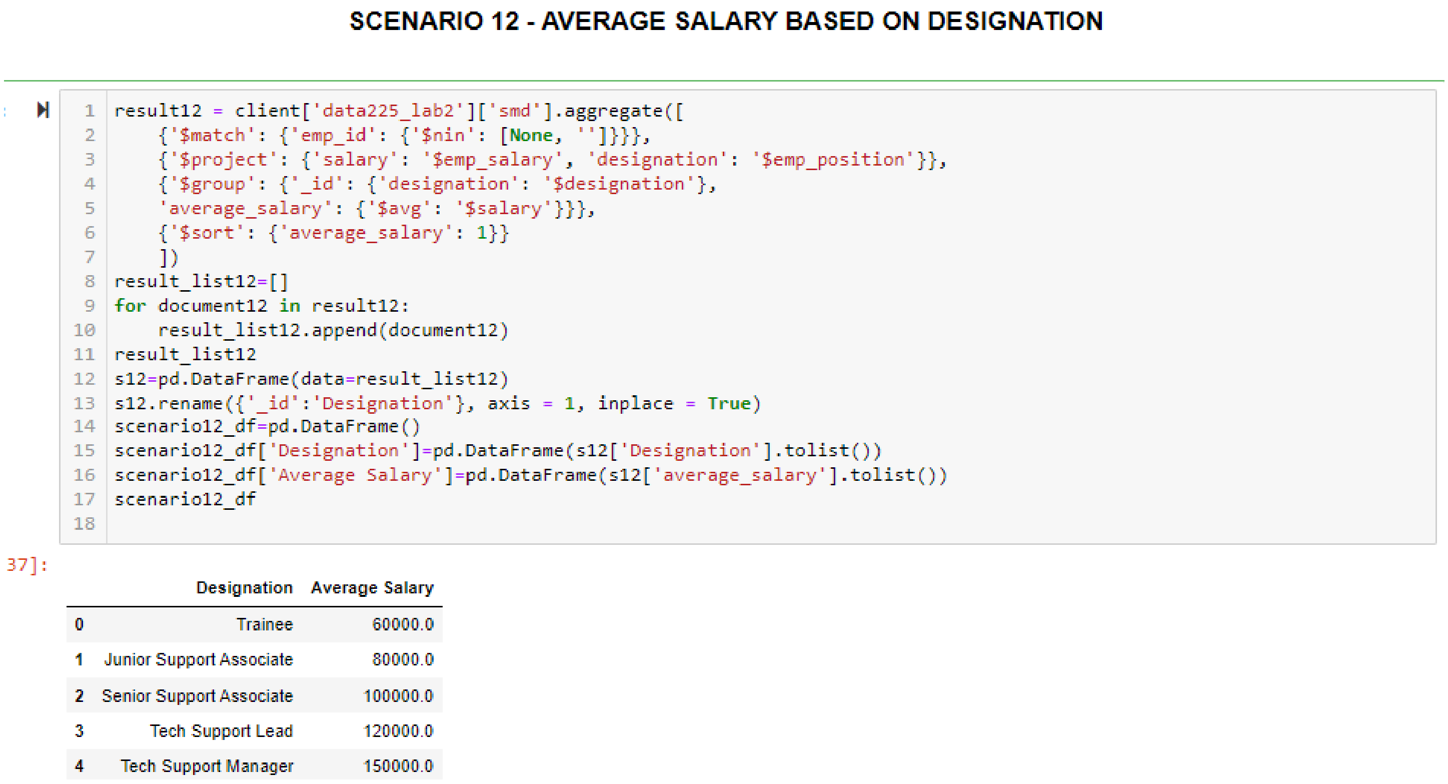
```
db.smd.aggregate([
{$match:{"emp_id":{$nin:[null,""]}}},
{$project:{salary:"$emp_salary", designation:"$emp_position"}},
{$group:{ _id: {designation:"$designation"},  average_salary: {    $avg: "$salary"  }}},
{$sort:{average_salary:1}}
]).pretty()
```



```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:{"emp_id":{$nin:[null,""]}}},
... {$project:{salary:"$emp_salary", designation:"$emp_position"}},
... {$group:{ _id: {designation:"$designation"},  average_salary: {    $avg: "$salary"  }}},
... {$sort:{average_salary:1}}
... ]).pretty()
[
  { _id: { designation: 'Trainee' }, average_salary: 60000 },
  {
    _id: { designation: 'Junior Support Associate' },
    average_salary: 80000
  },
  {
    _id: { designation: 'Senior Support Associate' },
    average_salary: 100000
  },
  { _id: { designation: 'Tech Support Lead' }, average_salary: 120000 },
  {
    _id: { designation: 'Tech Support Manager' },
    average_salary: 150000
  }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> _
```

QUERY & RESULT (PYTHON EXECUTION):

```
result12 = client['data225_lab2']['smd'].aggregate([
   {'$match': {'emp_id': {'$nin': [None, '']}}},
   {'$project': {'salary': '$emp_salary', 'designation': '$emp_position'}},
   {'$group': {'_id': {'designation': '$designation'},
   'average_salary': {'$avg': '$salary'}}},
   {'$sort': {'average_salary': 1}}
   ])
result_list12=[]
for document12 in result12:
   result_list12.append(document12)
result_list12
s12=pd.DataFrame(data=result_list12)
s12.rename({'_id':'Designation'}, axis = 1, inplace = True)
scenario12_df=pd.DataFrame()
scenario12_df['Designation']=pd.DataFrame(s12['Designation'].tolist())
scenario12_df['Average Salary']=pd.DataFrame(s12['average_salary'].tolist())
scenario12_df
```



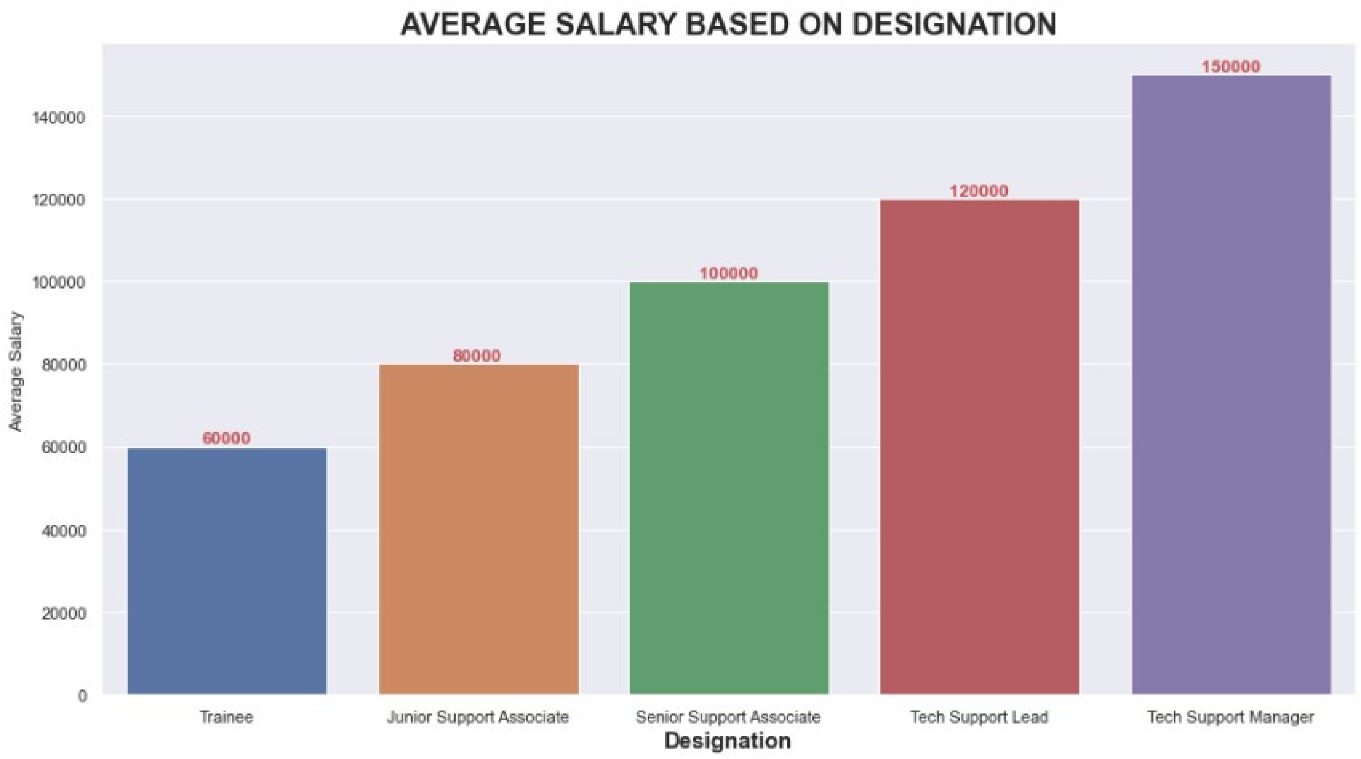
SCENARIO 12 - AVERAGE SALARY BASED ON DESIGNATION

```
result12 = client['data225_lab2']['smd'].aggregate([
    {'$match': {'emp_id': {'$nin': [None, '']}}},
    {'$project': {'salary': '$emp_salary', 'designation': '$emp_position'}},
    {'$group': {'_id': {'designation': '$designation'},
    'average_salary': {'$avg': '$salary'}}},
    {'$sort': {'average_salary': 1}}
    ])
result_list12=[]
for document12 in result12:
    result_list12.append(document12)
result_list12
s12=pd.DataFrame(data=result_list12)
s12.rename({'_id':'Designation'}, axis = 1, inplace = True)
scenario12_df=pd.DataFrame()
scenario12_df['Designation']=pd.DataFrame(s12['Designation'].tolist())
scenario12_df['Average Salary']=pd.DataFrame(s12['average_salary'].tolist())
scenario12_df

```

37]:

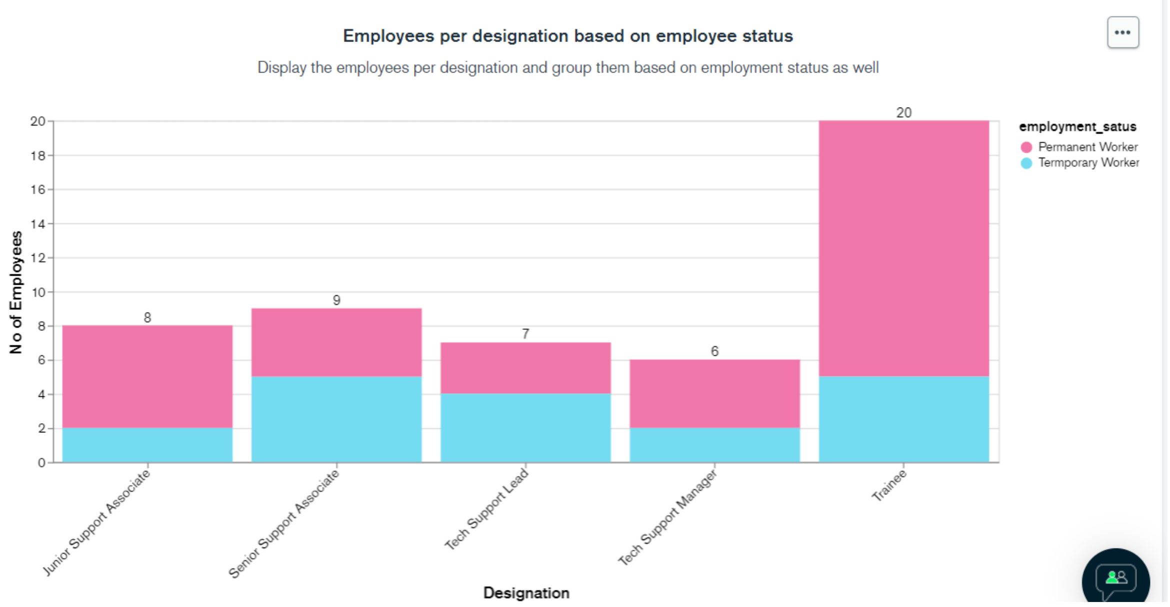
	Designation	Average Salary
0	Trainee	60000.0
1	Junior Support Associate	80000.0
2	Senior Support Associate	100000.0
3	Tech Support Lead	120000.0
4	Tech Support Manager	150000.0

JUPYTER VISUALIZATION:



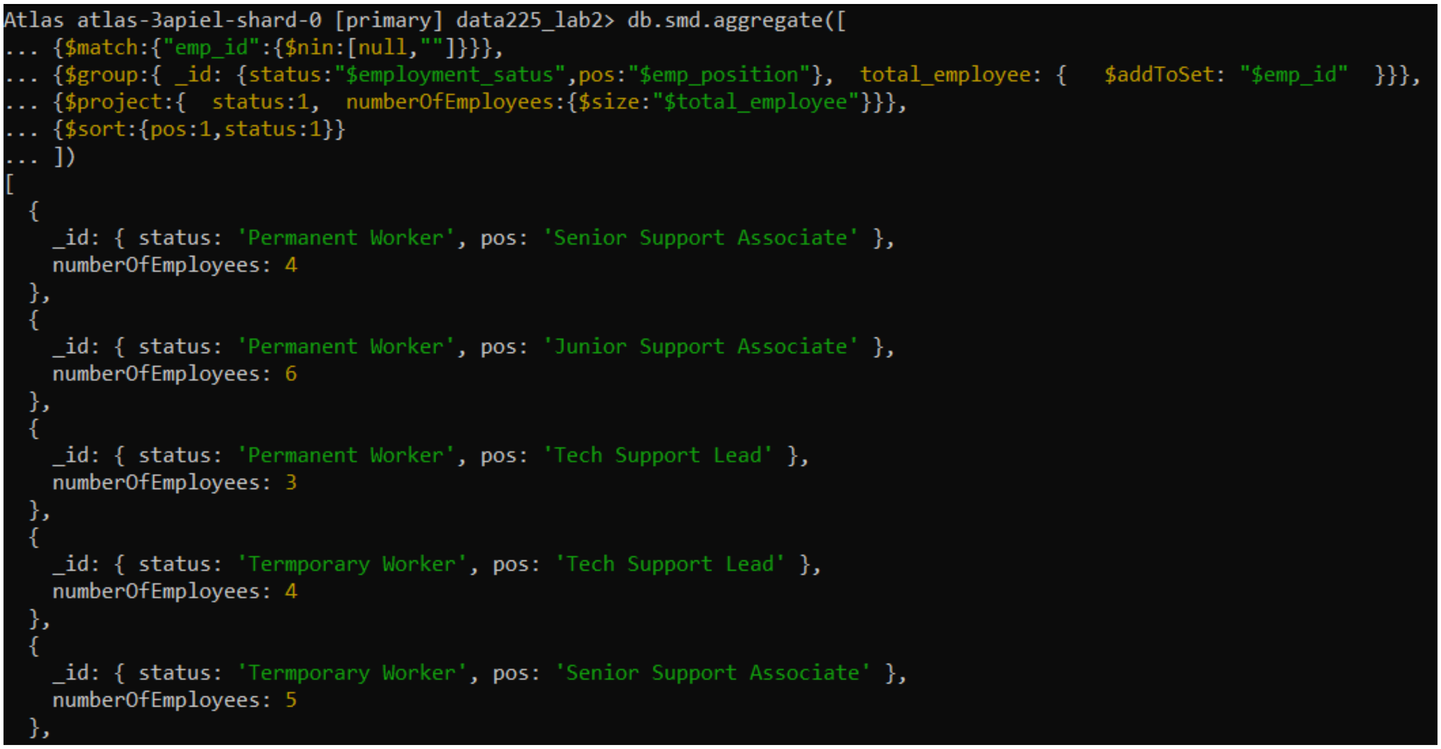
SCENARIO 13 - EMPLOYEES PER DESIGNATION BASED ON EMPLOYMENT STATUS

MONGO CHARTS:

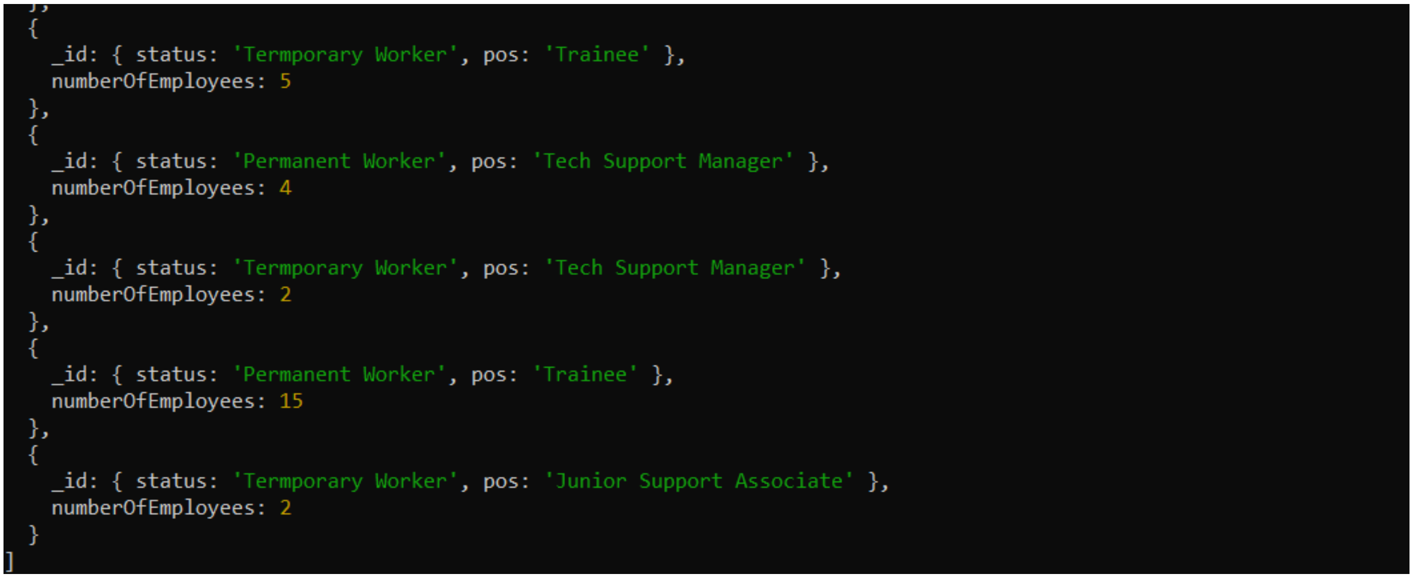


QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$match:{"emp_id":{$nin:[null,""]}}},
{$group:{ _id: {status:"$employment_satus",pos:"$emp_position"},  total_employee: {   $addToSet:
"$emp_id"  }}},
{$project:{  status:1,  numberOfEmployees:{$size:"$total_employee"}}},
{$sort:{pos:1,status:1}}
])
```



```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:{"emp_id":{$nin:[null,""]}}},
... {$group:{ _id: {status:"$employment_satus",pos:"$emp_position"},  total_employee: {   $addToSet: "$emp_id"  }}},
... {$project:{  status:1,  numberOfEmployees:{$size:"$total_employee"}}},
... {$sort:{pos:1,status:1}}
... ])
[
  {
    _id: { status: 'Permanent Worker', pos: 'Senior Support Associate' },
    numberOfEmployees: 4
  },
  {
    _id: { status: 'Permanent Worker', pos: 'Junior Support Associate' },
    numberOfEmployees: 6
  },
  {
    _id: { status: 'Permanent Worker', pos: 'Tech Support Lead' },
    numberOfEmployees: 3
  },
  {
    _id: { status: 'Termporary Worker', pos: 'Tech Support Lead' },
    numberOfEmployees: 4
  },
  {
    _id: { status: 'Termporary Worker', pos: 'Senior Support Associate' },
    numberOfEmployees: 5
  },
```

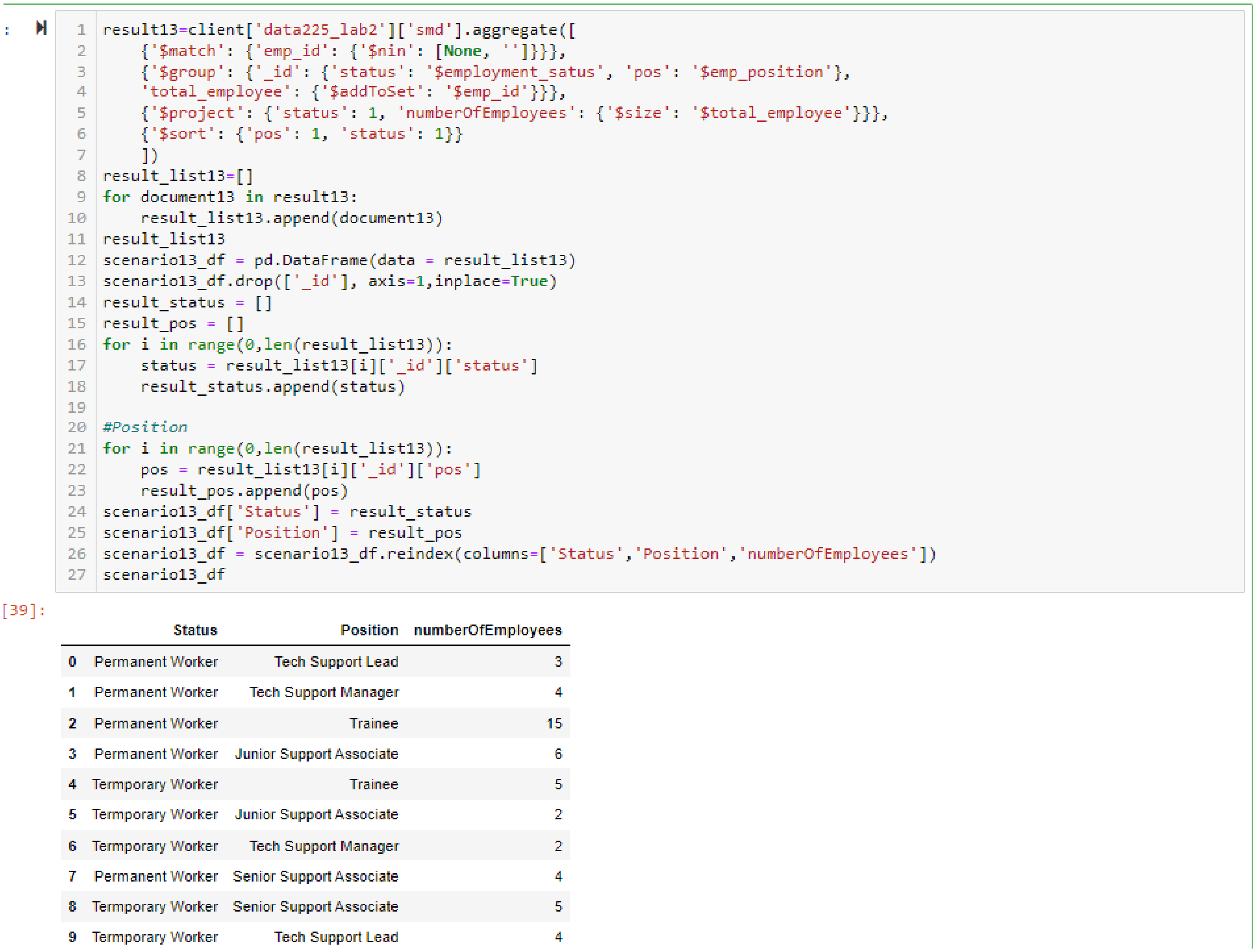


```
  },
  {
    _id: { status: 'Termporary Worker', pos: 'Trainee' },
    numberOfEmployees: 5
  },
  {
    _id: { status: 'Permanent Worker', pos: 'Tech Support Manager' },
    numberOfEmployees: 4
  },
  {
    _id: { status: 'Termporary Worker', pos: 'Tech Support Manager' },
    numberOfEmployees: 2
  },
  {
    _id: { status: 'Permanent Worker', pos: 'Trainee' },
    numberOfEmployees: 15
  },
  {
    _id: { status: 'Termporary Worker', pos: 'Junior Support Associate' },
    numberOfEmployees: 2
  }
]
```

QUERY & RESULT (PYTHON EXECUTION):

```
result13=client['data225_lab2']['smd'].aggregate([
{'$match': {'emp_id': {'$nin': [None, '']}}},
{'$group': {'_id': {'status': '$employment_satus', 'pos': '$emp_position'},
'total_employee': {'$addToSet': '$emp_id'}}},
{'$project': {'status': 1, 'numberOfEmployees': {'$size': '$total_employee'}}},
{'$sort': {'pos': 1, 'status': 1}}
])
result_list13=[]
for document13 in result13:
result_list13.append(document13)
result_list13
scenario13_df = pd.DataFrame(data = result_list13)
scenario13_df.drop(['_id'], axis=1,inplace=True)
result_status = []
result_pos = []
for i in range(0,len(result_list13)):
status = result_list13[i]['_id']['status']
result_status.append(status)
#Position
for i in range(0,len(result_list13)):
pos = result_list13[i]['_id']['pos']
result_pos.append(pos)
scenario13_df['Status'] = result_status
scenario13_df['Position'] = result_pos
scenario13_df = scenario13_df.reindex(columns=['Status','Position','numberOfEmployees'])
scenario13_df
```

SCENARIO 13 - EMPLOYEES PER DESIGNATION BASED ON EMPLOYMENT STATUS



```
result13=client['data225_lab2']['smd'].aggregate([
    {'$match': {'emp_id': {'$nin': [None, '']}}},
    {'$group': {'_id': {'status': '$employment_satus', 'pos': '$emp_position'},
    'total_employee': {'$addToSet': '$emp_id'}}},
    {'$project': {'status': 1, 'numberOfEmployees': {'$size': '$total_employee'}}},
    {'$sort': {'pos': 1, 'status': 1}}
    ])
result_list13=[]
for document13 in result13:
    result_list13.append(document13)
result_list13
scenario13_df = pd.DataFrame(data = result_list13)
scenario13_df.drop(['_id'], axis=1,inplace=True)
result_status = []
result_pos = []
for i in range(0,len(result_list13)):
    status = result_list13[i]['_id']['status']
    result_status.append(status)

#Position
for i in range(0,len(result_list13)):
    pos = result_list13[i]['_id']['pos']
    result_pos.append(pos)
scenario13_df['Status'] = result_status
scenario13_df['Position'] = result_pos
scenario13_df = scenario13_df.reindex(columns=['Status','Position','numberOfEmployees'])
scenario13_df
```

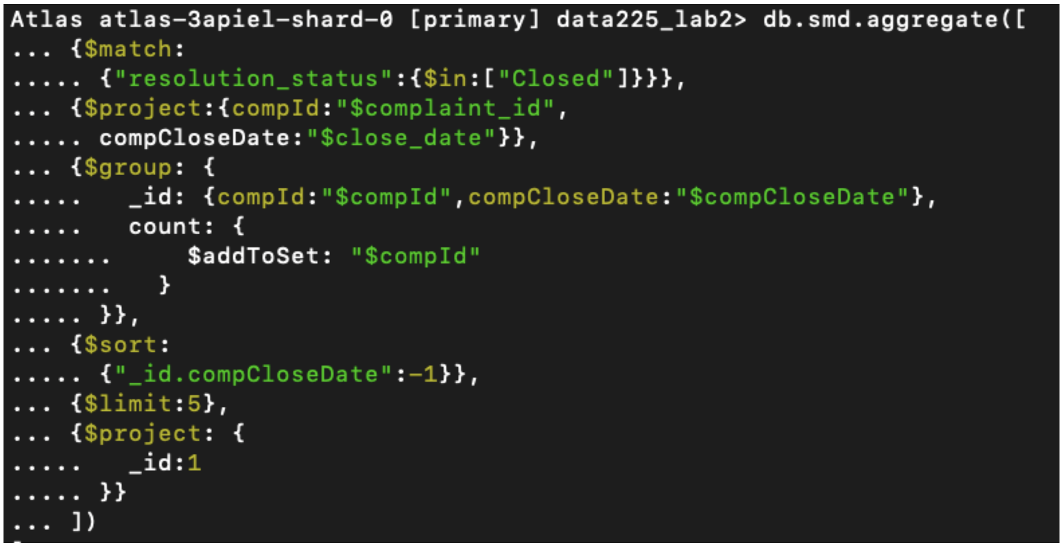
[39]:

	Status	Position	numberOfEmployees
0	Permanent Worker	Tech Support Lead	3
1	Permanent Worker	Tech Support Manager	4
2	Permanent Worker	Trainee	15
3	Permanent Worker	Junior Support Associate	6
4	Termporary Worker	Trainee	5
5	Termporary Worker	Junior Support Associate	2
6	Termporary Worker	Tech Support Manager	2
7	Permanent Worker	Senior Support Associate	4
8	Termporary Worker	Senior Support Associate	5
9	Termporary Worker	Tech Support Lead	4

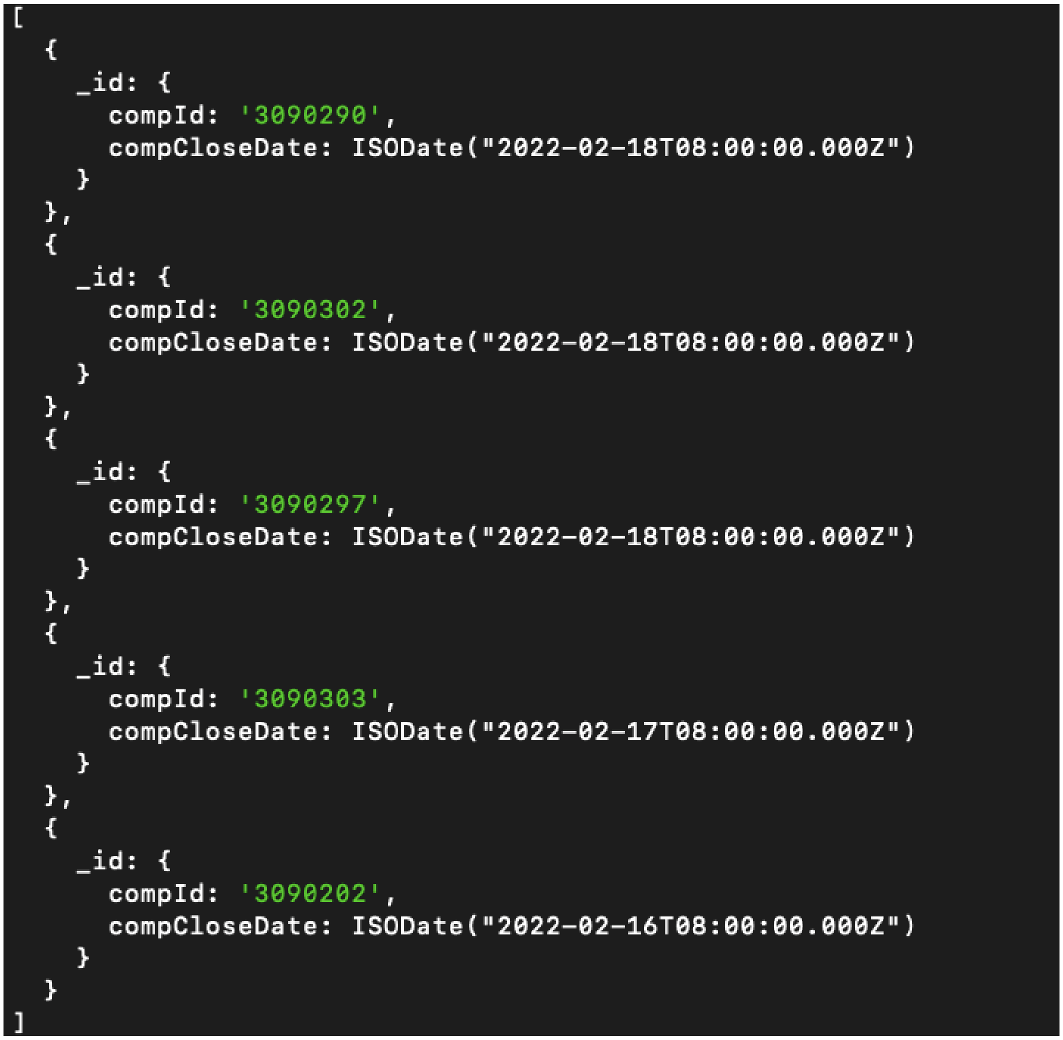
SCENARIO 14 - 5 LATEST CLOSED COMPLAINTS

QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$match:
{"resolution_status":{$in:["Closed"]}}},
{$project:{compId:"$complaint_id",
compCloseDate:"$close_date"}},
{$group: {
_id: {compId:"$compId",compCloseDate:"$compCloseDate"},
count: {
$addToSet: "$compId"
}
}},
{$sort:
{"_id.compCloseDate":-1}},
{$limit:5},
{$project: { _id:1}}])
```



```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:
..... {"resolution_status":{$in:["Closed"]}}},
... {$project:{compId:"$complaint_id",
..... compCloseDate:"$close_date"}},
... {$group: {
.....   _id: {compId:"$compId",compCloseDate:"$compCloseDate"},
.....   count: {
.......     $addToSet: "$compId"
.......   }
..... }},
... {$sort:
..... {"_id.compCloseDate":-1}},
... {$limit:5},
... {$project: {
.....   _id:1
..... }}
... ])
```

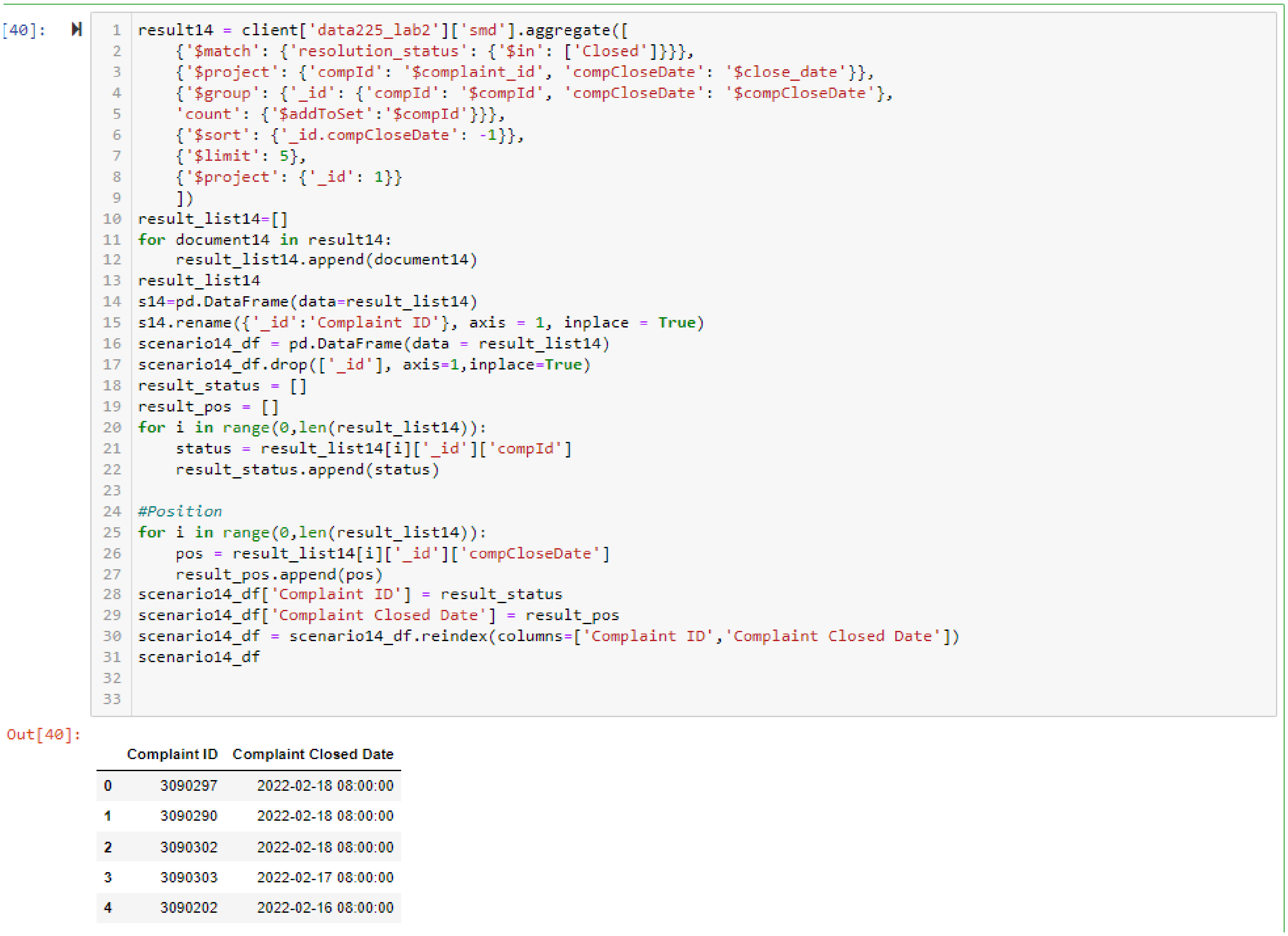


```
[
  {
    _id: {
      compId: '3090290',
      compCloseDate: ISODate("2022-02-18T08:00:00.000Z")
    }
  },
  {
    _id: {
      compId: '3090302',
      compCloseDate: ISODate("2022-02-18T08:00:00.000Z")
    }
  },
  {
    _id: {
      compId: '3090297',
      compCloseDate: ISODate("2022-02-18T08:00:00.000Z")
    }
  },
  {
    _id: {
      compId: '3090303',
      compCloseDate: ISODate("2022-02-17T08:00:00.000Z")
    }
  },
  {
    _id: {
      compId: '3090202',
      compCloseDate: ISODate("2022-02-16T08:00:00.000Z")
    }
  }
]
```

QUERY & RESULT (PYTHON EXECUTION):

```
result14 = client['data225_lab2']['smd'].aggregate([
{'$match': {'resolution_status': {'$in': ['Closed']}}},
{'$project': {'compId': '$complaint_id', 'compCloseDate': '$close_date'}},
{'$group': {'_id': {'compId': '$compId', 'compCloseDate': '$compCloseDate'},
'count': {'$addToSet':'$compId'}}},
{'$sort': {'_id.compCloseDate': -1}},
{'$limit': 5},
{'$project': {'_id': 1}}
])
result_list14=[]
for document14 in result14:
result_list14.append(document14)
result_list14
s14=pd.DataFrame(data=result_list14)
s14.rename({'_id':'Complaint ID'}, axis = 1, inplace = True)
scenario14_df = pd.DataFrame(data = result_list14)
scenario14_df.drop(['_id'], axis=1,inplace=True)
result_status = []
result_pos = []
for i in range(0,len(result_list14)):
status = result_list14[i]['_id']['compId']
result_status.append(status)

#Position
for i in range(0,len(result_list14)):
pos = result_list14[i]['_id']['compCloseDate']
result_pos.append(pos)
scenario14_df['Complaint ID'] = result_status
scenario14_df['Complaint Closed Date'] = result_pos
scenario14_df = scenario14_df.reindex(columns=['Complaint ID','Complaint Closed Date'])
scenario14_df
```

SCENARIO 14 - 5 LATEST CLOSED COMPLAINTS



[40]:

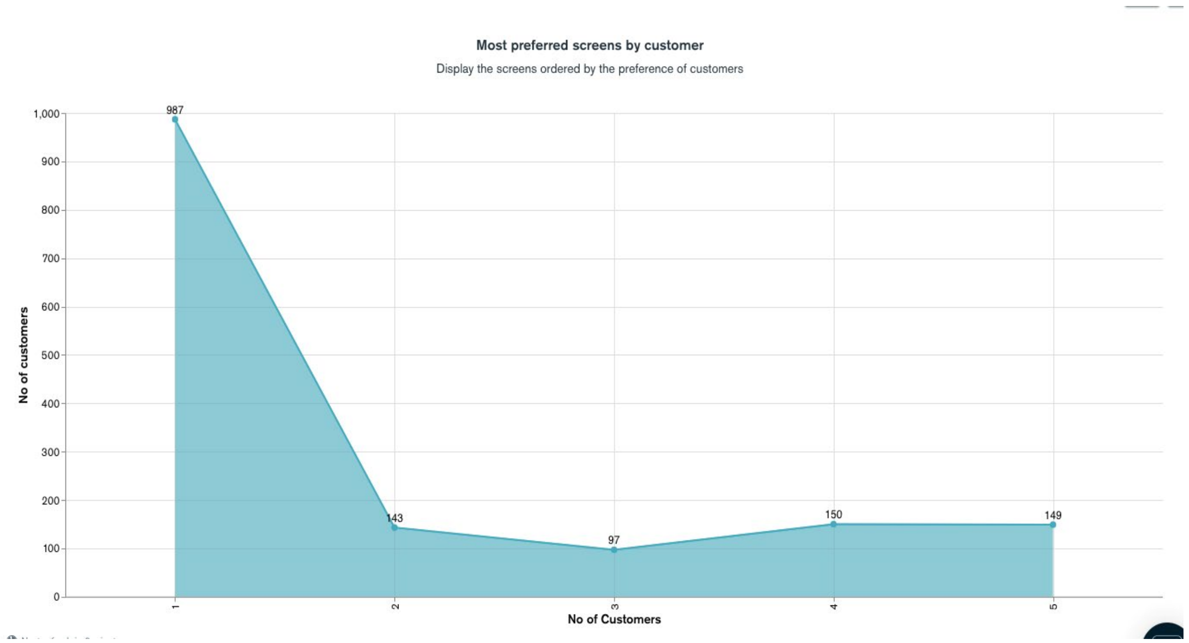
```
result14 = client['data225_lab2']['smd'].aggregate([
    {'$match': {'resolution_status': {'$in': ['Closed']}}},
    {'$project': {'compId': '$complaint_id', 'compCloseDate': '$close_date'}},
    {'$group': {'_id': {'compId': '$compId', 'compCloseDate': '$compCloseDate'},
    'count': {'$addToSet':'$compId'}}},
    {'$sort': {'_id.compCloseDate': -1}},
    {'$limit': 5},
    {'$project': {'_id': 1}}
    ])
result_list14=[]
for document14 in result14:
    result_list14.append(document14)
result_list14
s14=pd.DataFrame(data=result_list14)
s14.rename({'_id':'Complaint ID'}, axis = 1, inplace = True)
scenario14_df = pd.DataFrame(data = result_list14)
scenario14_df.drop(['_id'], axis=1,inplace=True)
result_status = []
result_pos = []
for i in range(0,len(result_list14)):
    status = result_list14[i]['_id']['compId']
    result_status.append(status)

#Position
for i in range(0,len(result_list14)):
    pos = result_list14[i]['_id']['compCloseDate']
    result_pos.append(pos)
scenario14_df['Complaint ID'] = result_status
scenario14_df['Complaint Closed Date'] = result_pos
scenario14_df = scenario14_df.reindex(columns=['Complaint ID','Complaint Closed Date'])
scenario14_df
```

Out[40]:

	Complaint ID	Complaint Closed Date
0	3090297	2022-02-18 08:00:00
1	3090290	2022-02-18 08:00:00
2	3090302	2022-02-18 08:00:00
3	3090303	2022-02-17 08:00:00
4	3090202	2022-02-16 08:00:00

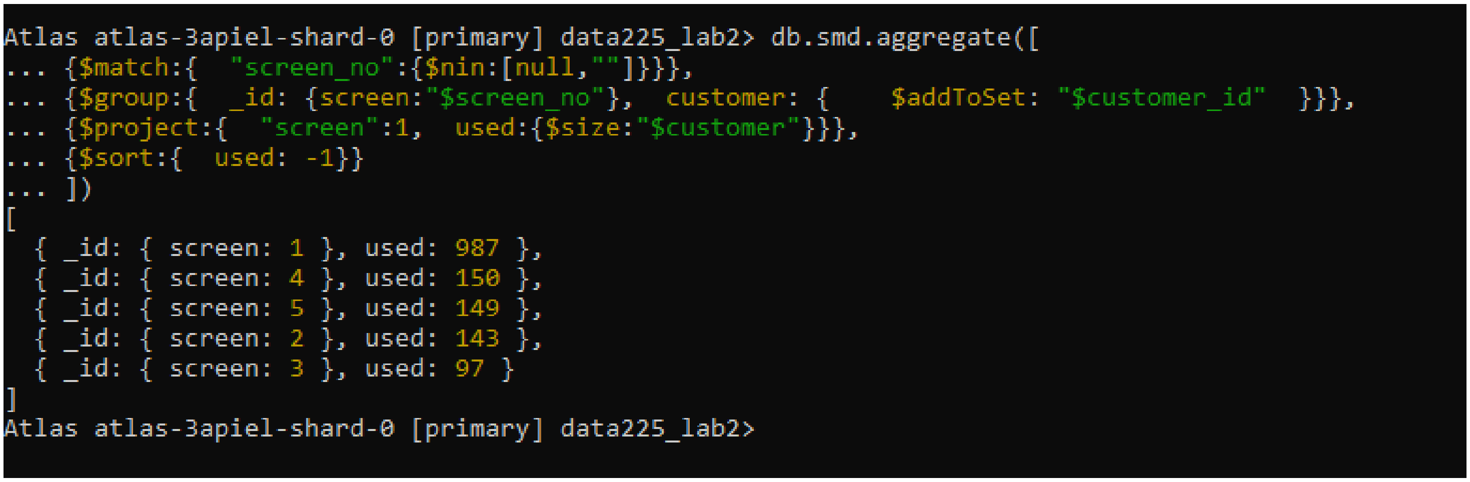
SCENARIO 15 – MOST PREFERRED SCREENS BY THE CUSTOMERS

MONGO CHARTS:



QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$match:{  "screen_no":{$nin:[null,""]}}},
{$group:{  _id: {screen:"$screen_no"},  customer: {    $addToSet: "$customer_id"  }}},
{$project:{  "screen":1,  used:{$size:"$customer"}}},
{$sort:{  used: -1}}
])
```

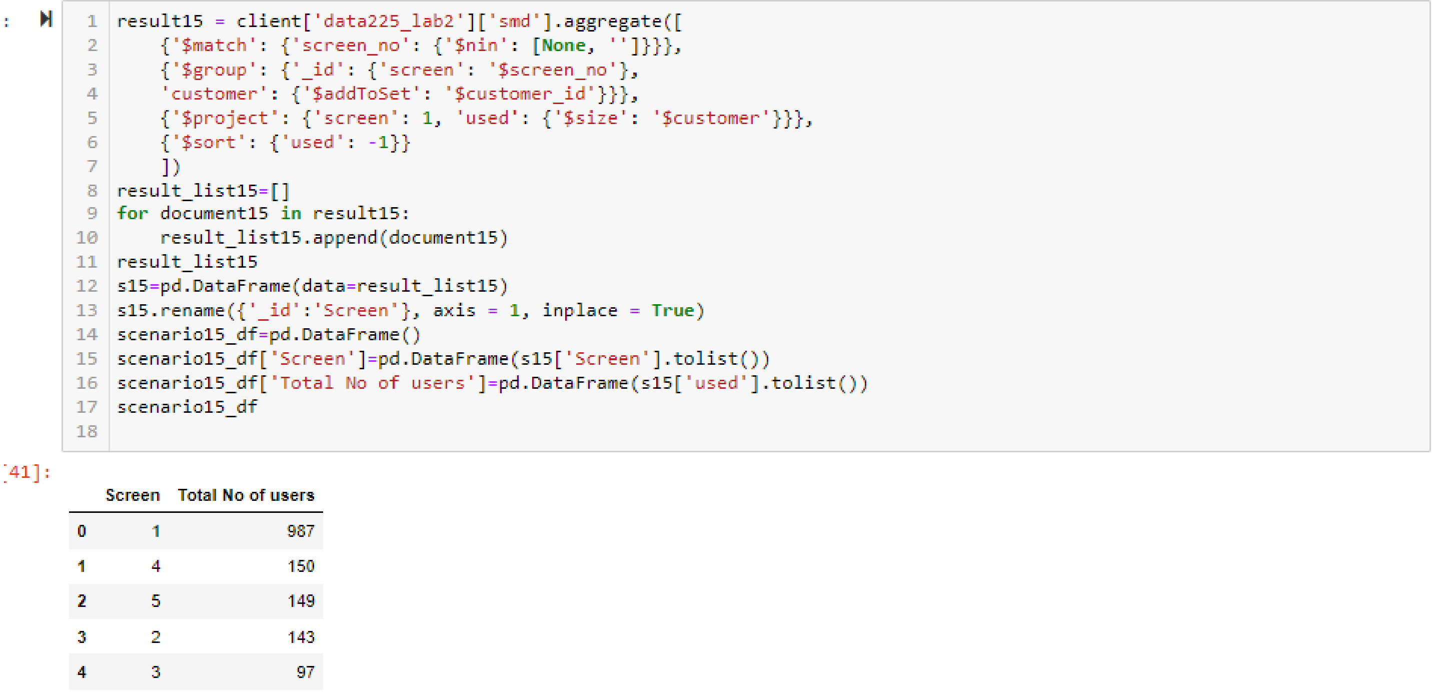


```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:{  "screen_no":{$nin:[null,""]}}},
... {$group:{  _id: {screen:"$screen_no"},  customer: {    $addToSet: "$customer_id"  }}},
... {$project:{  "screen":1,  used:{$size:"$customer"}}},
... {$sort:{  used: -1}}
... ])
[
  { _id: { screen: 1 }, used: 987 },
  { _id: { screen: 4 }, used: 150 },
  { _id: { screen: 5 }, used: 149 },
  { _id: { screen: 2 }, used: 143 },
  { _id: { screen: 3 }, used: 97 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2>
```

QUERY & RESULT (PYTHON EXECUTION):

```
result15 = client['data225_lab2']['smd'].aggregate([
{'$match': {'screen_no': {'$nin': [None, '']}}},
{'$group': {'_id': {'screen': '$screen_no'},
'customer': {'$addToSet': '$customer_id'}}},
{'$project': {'screen': 1, 'used': {'$size': '$customer'}}},
{'$sort': {'used': -1}}
])
result_list15=[]
for document15 in result15:
result_list15.append(document15)
result_list15
s15=pd.DataFrame(data=result_list15)
s15.rename({'_id':'Screen'}, axis = 1, inplace = True)
scenario15_df=pd.DataFrame()
scenario15_df['Screen']=pd.DataFrame(s15['Screen'].tolist())
scenario15_df['Total No of users']=pd.DataFrame(s15['used'].tolist())
scenario15_df
```

SCENARIO 15 – MOST PREFERRED SCREENS BY THE CUSTOMERS

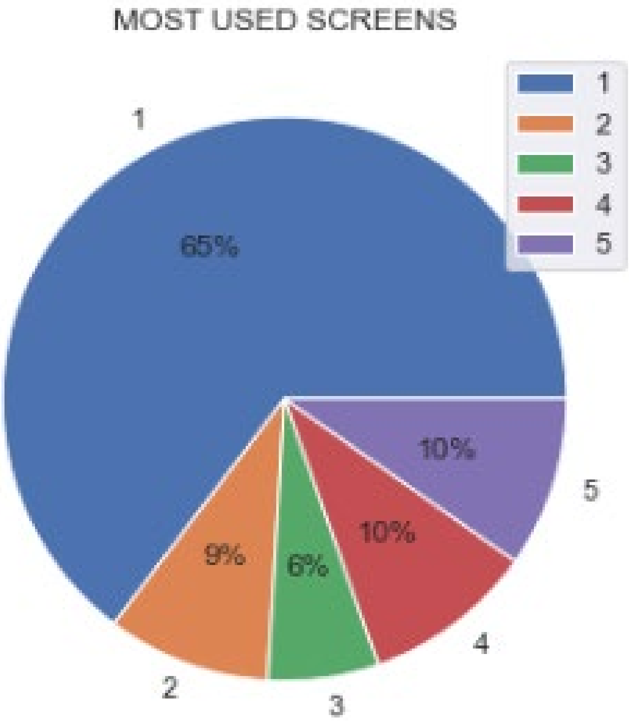


```
result15 = client['data225_lab2']['smd'].aggregate([
    {'$match': {'screen_no': {'$nin': [None, '']}}},
    {'$group': {'_id': {'screen': '$screen_no'},
    'customer': {'$addToSet': '$customer_id'}}},
    {'$project': {'screen': 1, 'used': {'$size': '$customer'}}},
    {'$sort': {'used': -1}}
    ])
result_list15=[]
for document15 in result15:
    result_list15.append(document15)
result_list15
s15=pd.DataFrame(data=result_list15)
s15.rename({'_id':'Screen'}, axis = 1, inplace = True)
scenario15_df=pd.DataFrame()
scenario15_df['Screen']=pd.DataFrame(s15['Screen'].tolist())
scenario15_df['Total No of users']=pd.DataFrame(s15['used'].tolist())
scenario15_df
```

[41]:

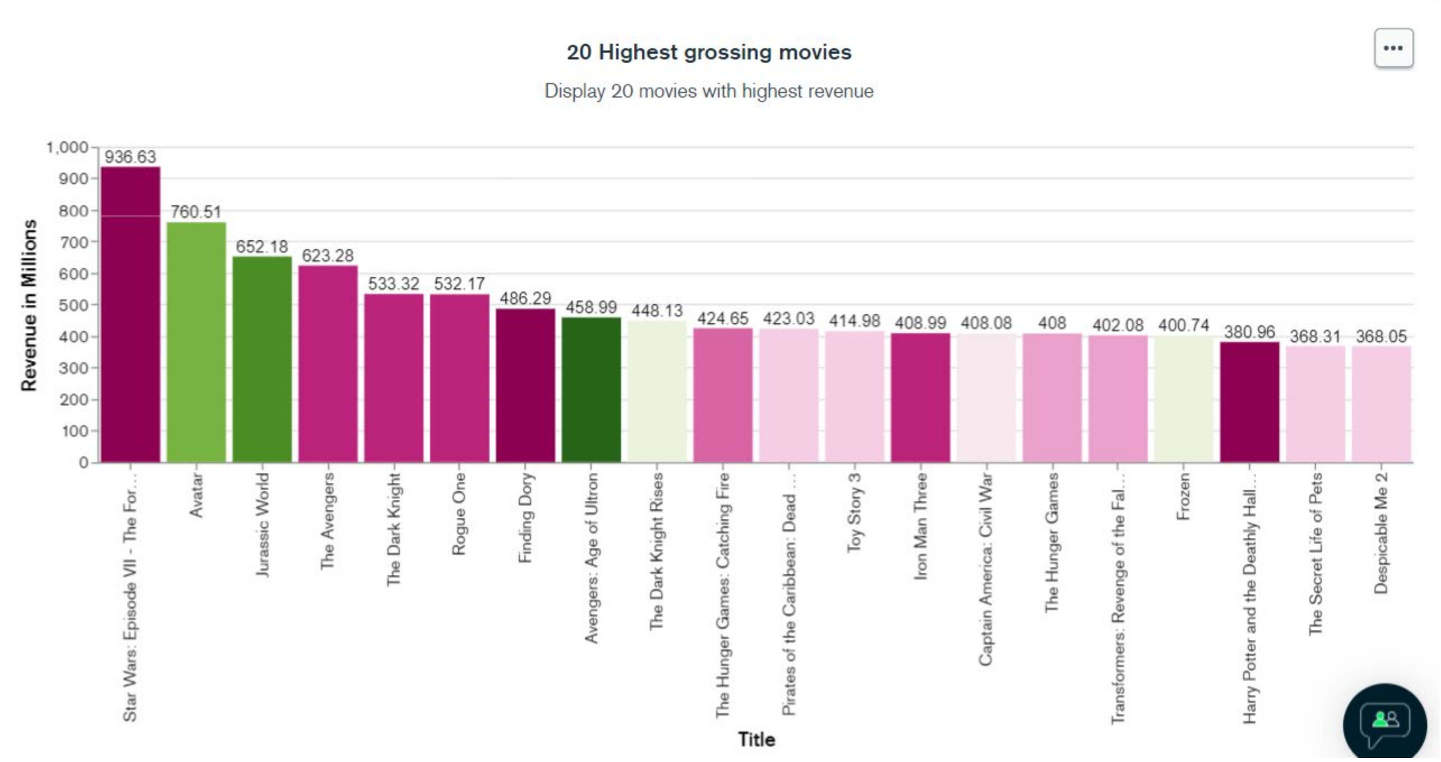
	Screen	Total No of users
0	1	987
1	4	150
2	5	149
3	2	143
4	3	97

JUPYTER VISUALIZATION:



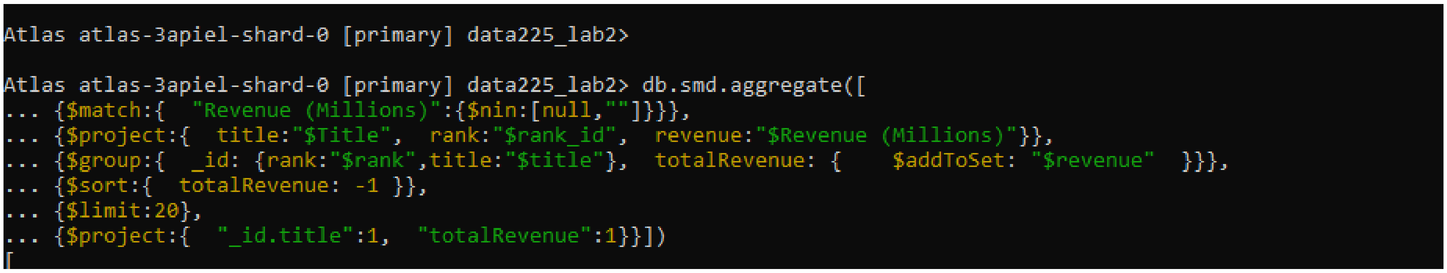
SCENARIO 16 – TOP HIGHEST GROSSING MOVIES

MONGO CHARTS:

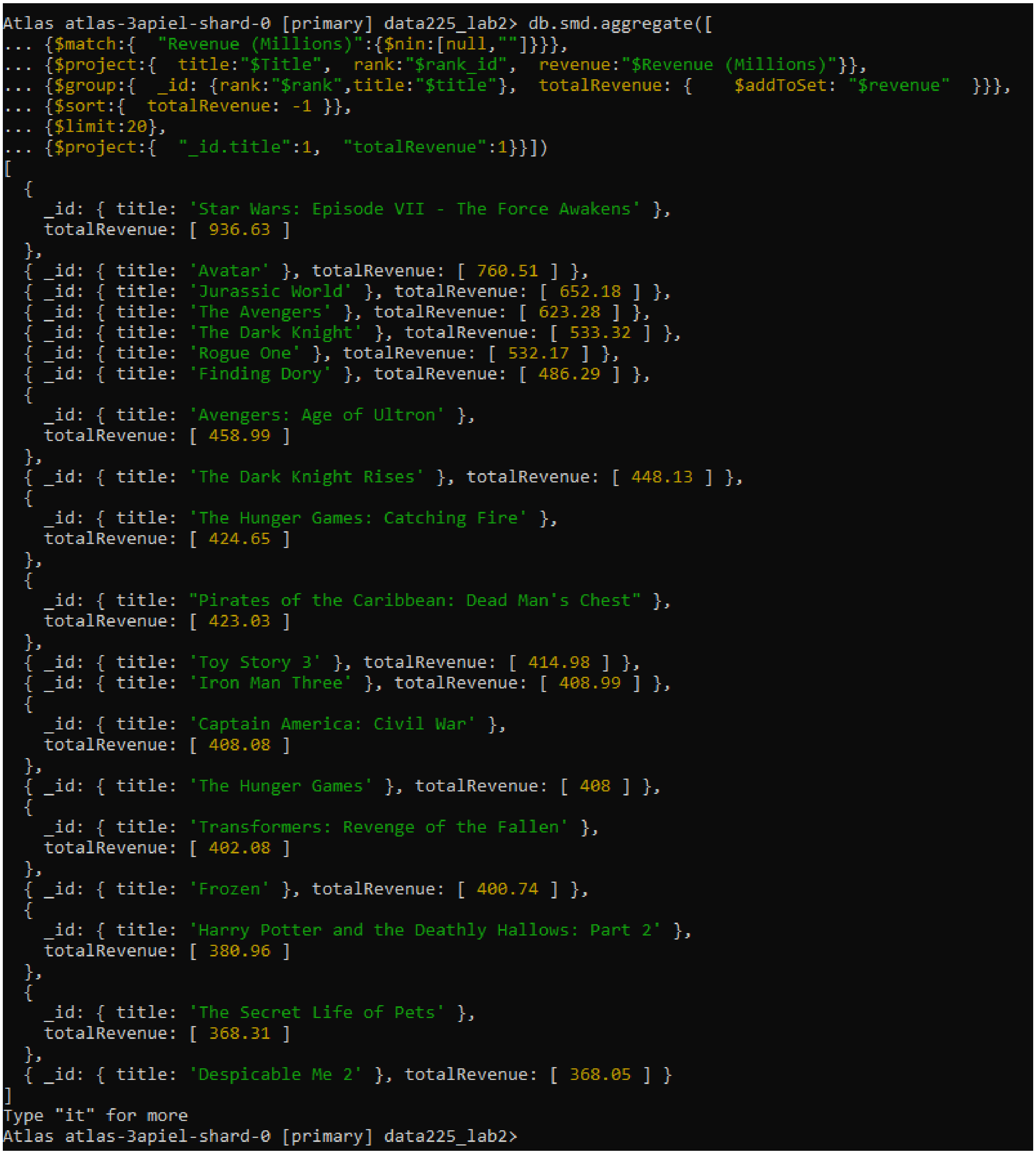


QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$match:{  "Revenue (Millions)":{$nin:[null,""]}}},
{$project:{  title:"$Title",  rank:"$rank_id",  revenue:"$Revenue (Millions)"}},
{$group:{  _id: {rank:"$rank",title:"$title"},  totalRevenue: {    $addToSet: "$revenue"  }}},
{$sort:{  totalRevenue: -1 }},
{$limit:20},
{$project:{  "_id.title":1,  "totalRevenue":1}}])
```



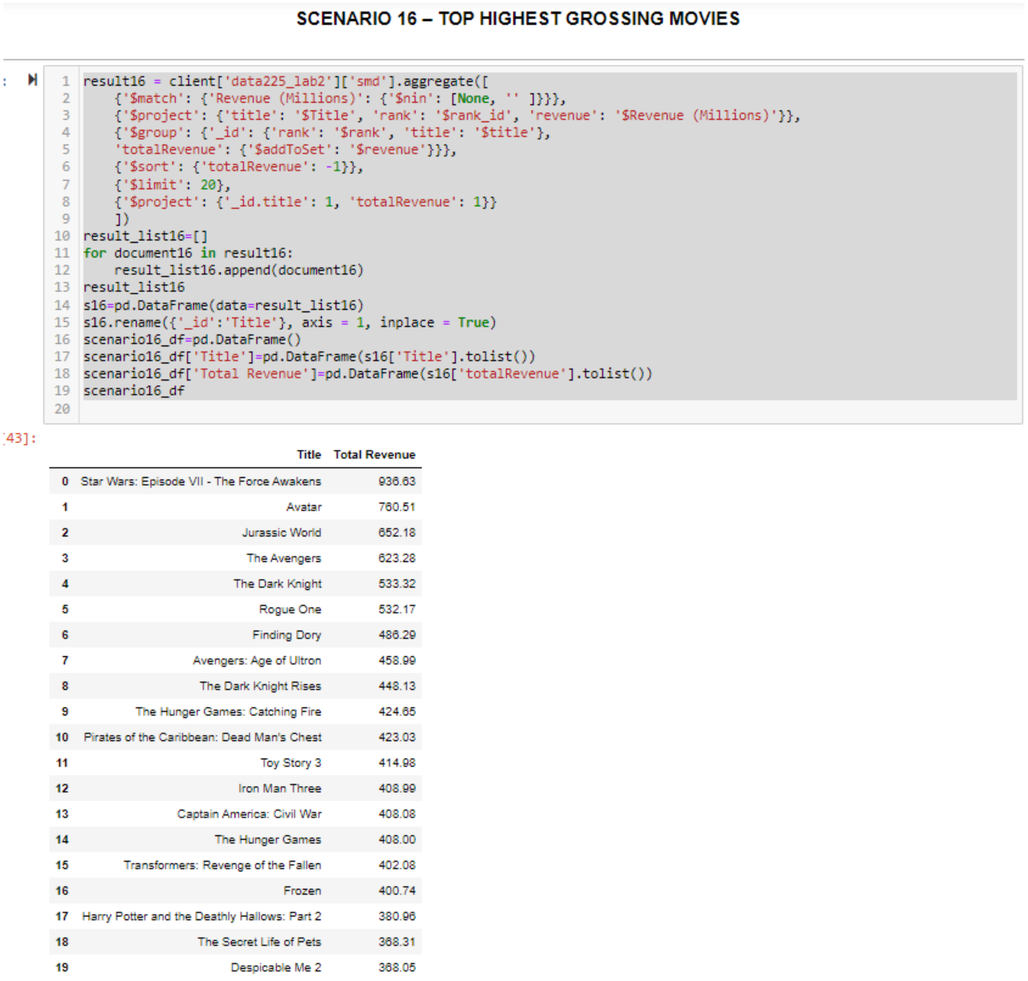
```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2>

Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:{  "Revenue (Millions)":{$nin:[null,""]}}},
... {$project:{  title:"$Title",  rank:"$rank_id",  revenue:"$Revenue (Millions)"}},
... {$group:{  _id: {rank:"$rank",title:"$title"},  totalRevenue: {    $addToSet:  "$revenue"  }}},
... {$sort:{  totalRevenue: -1 }},
... {$limit:20},
... {$project:{  "_id.title":1,  "totalRevenue":1}}])
[
```



```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:{  "Revenue (Millions)":{$nin:[null,""]}}},
... {$project:{  title:"$Title",  rank:"$rank_id",  revenue:"$Revenue (Millions)"}},
... {$group:{  _id: {rank:"$rank",title:"$title"},  totalRevenue: {    $addToSet: "$revenue"  }}},
... {$sort:{  totalRevenue: -1 }},
... {$limit:20},
... {$project:{  "_id.title":1,  "totalRevenue":1}}])
[
  {
    _id: { title: 'Star Wars: Episode VII - The Force Awakens' },
    totalRevenue: [ 936.63 ]
  },
  { _id: { title: 'Avatar' }, totalRevenue: [ 760.51 ] },
  { _id: { title: 'Jurassic World' }, totalRevenue: [ 652.18 ] },
  { _id: { title: 'The Avengers' }, totalRevenue: [ 623.28 ] },
  { _id: { title: 'The Dark Knight' }, totalRevenue: [ 533.32 ] },
  { _id: { title: 'Rogue One' }, totalRevenue: [ 532.17 ] },
  { _id: { title: 'Finding Dory' }, totalRevenue: [ 486.29 ] },
  {
    _id: { title: 'Avengers: Age of Ultron' },
    totalRevenue: [ 458.99 ]
  },
  { _id: { title: 'The Dark Knight Rises' }, totalRevenue: [ 448.13 ] },
  {
    _id: { title: 'The Hunger Games: Catching Fire' },
    totalRevenue: [ 424.65 ]
  },
  {
    _id: { title: "Pirates of the Caribbean: Dead Man's Chest" },
    totalRevenue: [ 423.03 ]
  },
  { _id: { title: 'Toy Story 3' }, totalRevenue: [ 414.98 ] },
  { _id: { title: 'Iron Man Three' }, totalRevenue: [ 408.99 ] },
  {
    _id: { title: 'Captain America: Civil War' },
    totalRevenue: [ 408.08 ]
  },
  { _id: { title: 'The Hunger Games' }, totalRevenue: [ 408 ] },
  {
    _id: { title: 'Transformers: Revenge of the Fallen' },
    totalRevenue: [ 402.08 ]
  },
  { _id: { title: 'Frozen' }, totalRevenue: [ 400.74 ] },
  {
    _id: { title: 'Harry Potter and the Deathly Hallows: Part 2' },
    totalRevenue: [ 380.96 ]
  },
  {
    _id: { title: 'The Secret Life of Pets' },
    totalRevenue: [ 368.31 ]
  },
  { _id: { title: 'Despicable Me 2' }, totalRevenue: [ 368.05 ] }
]
Type "it" for more
Atlas atlas-3apiel-shard-0 [primary] data225_lab2>
```

QUERY & RESULT (PYTHON EXECUTION):

```
result16 = client['data225_lab2']['smd'].aggregate([
{'$match': {'Revenue (Millions)': {'$nin': [None, '' ]}}},
{'$project': {'title': '$Title', 'rank': '$rank_id', 'revenue': '$Revenue (Millions)'}},
{'$group': {'_id': {'rank': '$rank', 'title': '$title'},
'totalRevenue': {'$addToSet': '$revenue'}}},
{'$sort': {'totalRevenue': -1}},
{'$limit': 20},
{'$project': {'_id.title': 1, 'totalRevenue': 1}}
])
result_list16=[]
for document16 in result16:
result_list16.append(document16)
result_list16
s16=pd.DataFrame(data=result_list16)
s16.rename({'_id':'Title'}, axis = 1, inplace = True)
scenario16_df=pd.DataFrame()
scenario16_df['Title']=pd.DataFrame(s16['Title'].tolist())
scenario16_df['Total Revenue']=pd.DataFrame(s16['totalRevenue'].tolist())
scenario16_df
```



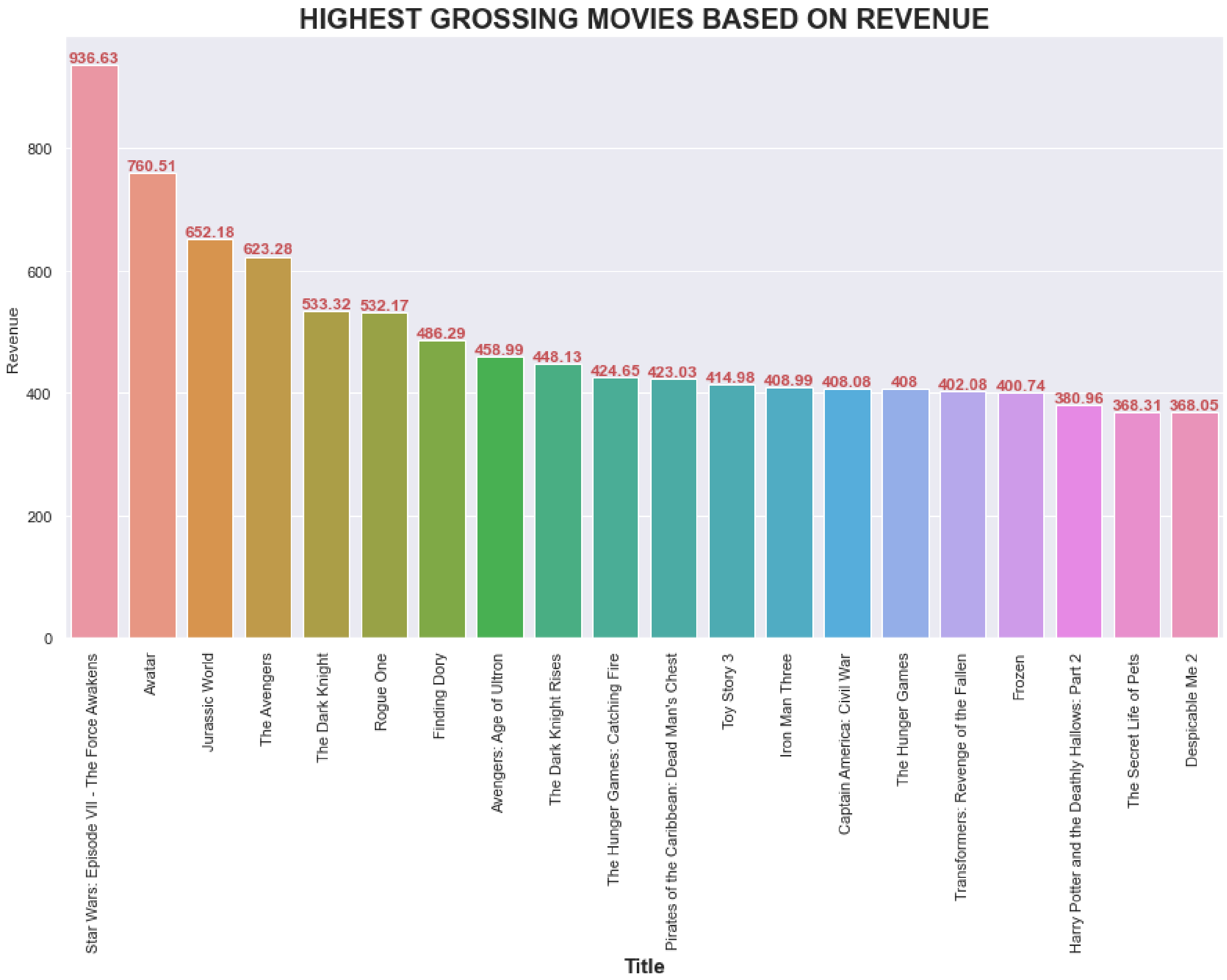
SCENARIO 16 – TOP HIGHEST GROSSING MOVIES

```
result16 = client['data225_lab2']['smd'].aggregate([
    {'$match': {'Revenue (Millions)': {'$nin': [None, '' ]}}},
    {'$project': {'title': '$Title', 'rank': '$rank_id', 'revenue': '$Revenue (Millions)'}},
    {'$group': {'_id': {'rank': '$rank', 'title': '$title'},
    'totalRevenue': {'$addToSet': '$revenue'}}},
    {'$sort': {'totalRevenue': -1}},
    {'$limit': 20},
    {'$project': {'_id.title': 1, 'totalRevenue': 1}}
    ])
result_list16=[]
for document16 in result16:
    result_list16.append(document16)
result_list16
s16=pd.DataFrame(data=result_list16)
s16.rename({'_id':'Title'}, axis = 1, inplace = True)
scenario16_df=pd.DataFrame()
scenario16_df['Title']=pd.DataFrame(s16['Title'].tolist())
scenario16_df['Total Revenue']=pd.DataFrame(s16['totalRevenue'].tolist())
scenario16_df
```

43]:

	Title	Total Revenue
0	Star Wars: Episode VII - The Force Awakens	936.63
1	Avatar	760.51
2	Jurassic World	652.18
3	The Avengers	623.28
4	The Dark Knight	533.32
5	Rogue One	532.17
6	Finding Dory	486.29
7	Avengers: Age of Ultron	458.99
8	The Dark Knight Rises	448.13
9	The Hunger Games: Catching Fire	424.65
10	Pirates of the Caribbean: Dead Man's Chest	423.03
11	Toy Story 3	414.98
12	Iron Man Three	408.99
13	Captain America: Civil War	408.08
14	The Hunger Games	408.00
15	Transformers: Revenge of the Fallen	402.08
16	Frozen	400.74
17	Harry Potter and the Deathly Hallows: Part 2	380.96
18	The Secret Life of Pets	368.31
19	Despicable Me 2	368.05

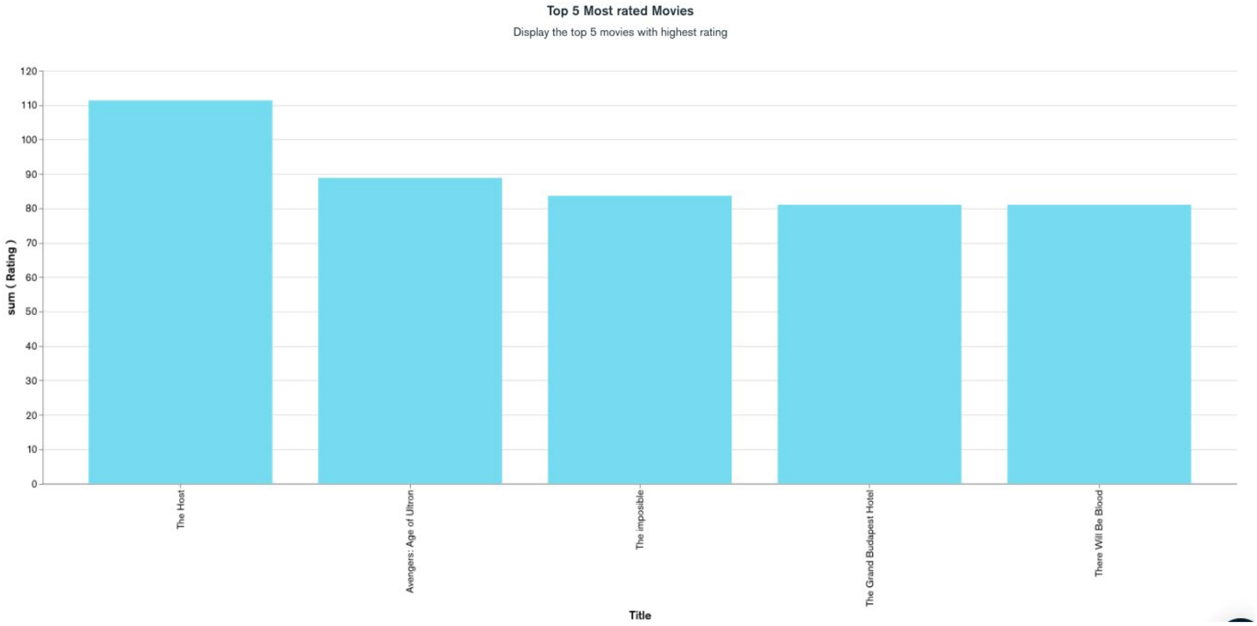
JUPYTER VISUALIZATION:



HIGHEST GROSSING MOVIES BASED ON REVENUE

SCENARIO 17 – TOP 5 MOST RATED MOVIES

MONGO CHARTS:

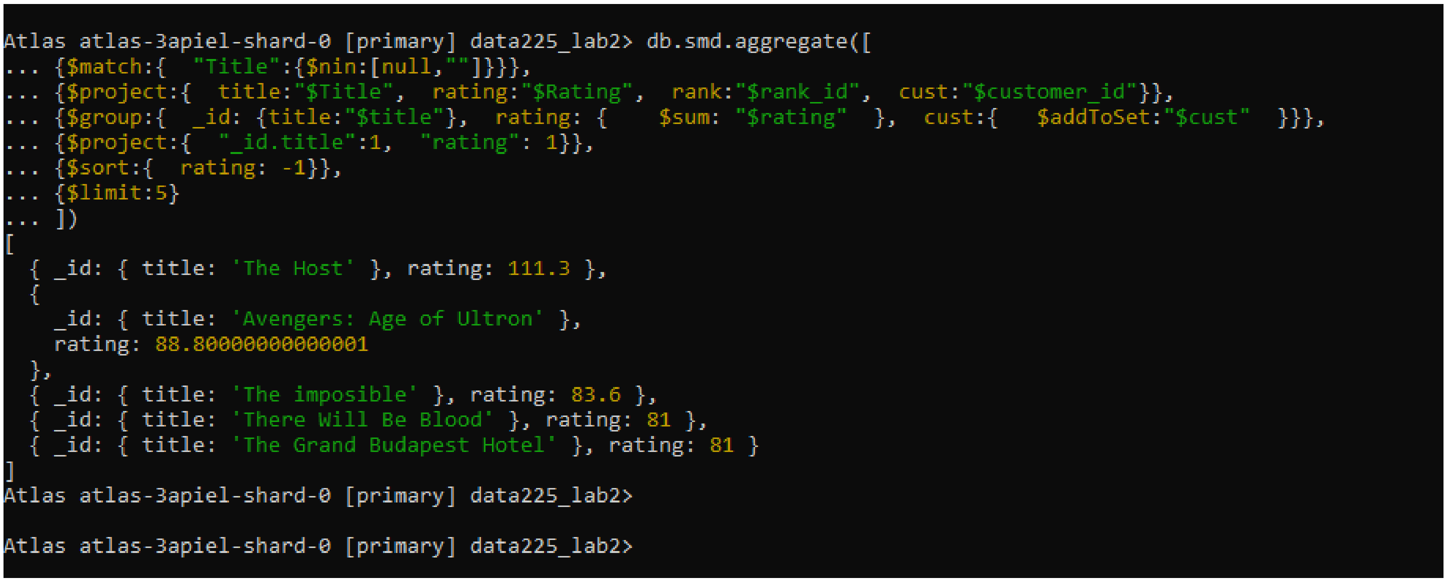


Top 5 Most rated Movies

Display the top 5 movies with highest rating

QUERY & RESULT (SHELL EXECUTION):

```
db.smd.aggregate([
{$match:{  "Title":{$nin:[null,""]}}},
{$project:{  title:"$Title",  rating:"$Rating",  rank:"$rank_id",  cust:"$customer_id"}},
{$group:{  _id: {title:"$title"},  rating: {   $sum: "$rating"  },  cust:{  $addToSet:"$cust"  }}},
{$project:{  "_id.title":1,  "rating": 1}},
{$sort:{  rating: -1}},
{$limit:5}])
```

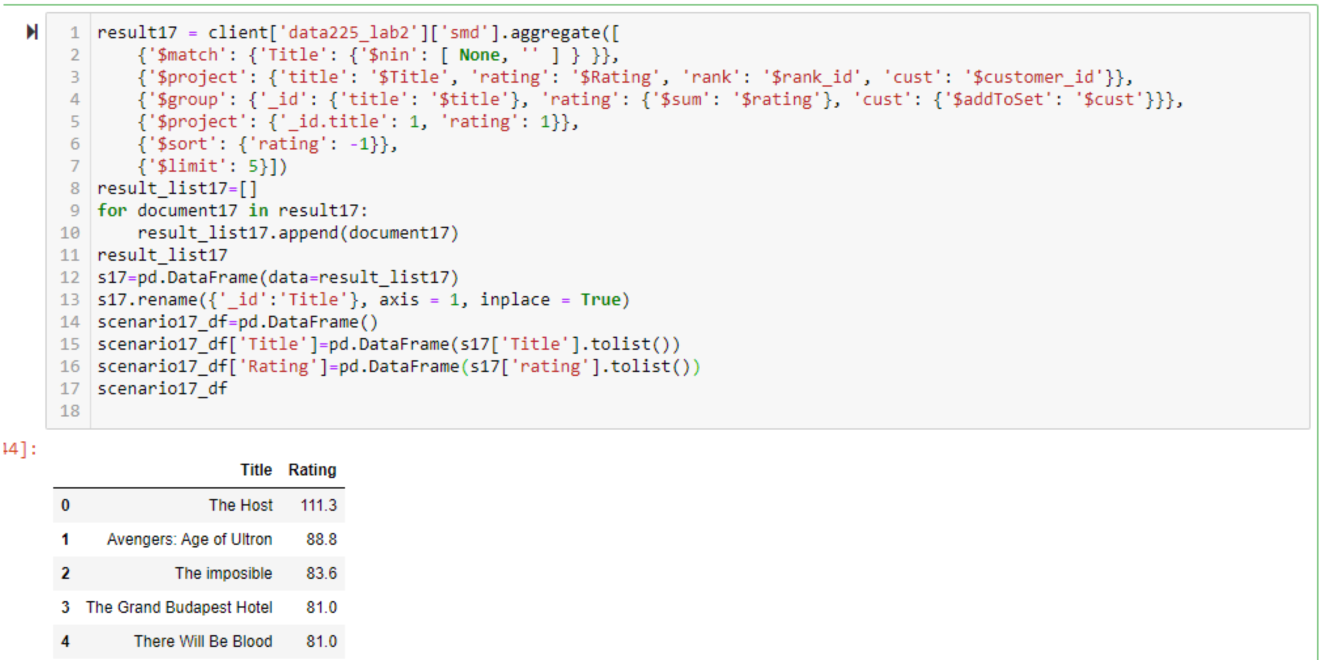


```
Atlas atlas-3apiel-shard-0 [primary] data225_lab2> db.smd.aggregate([
... {$match:{  "Title":{$nin:[null,""]}}},
... {$project:{  title:"$Title",  rating:"$Rating",  rank:"$rank_id",  cust:"$customer_id"}},
... {$group:{  _id: {title:"$title"},  rating: {    $sum: "$rating"  },  cust:{   $addToSet:"$cust"  }}},
... {$project:{  "_id.title":1,  "rating": 1}},
... {$sort:{  rating: -1}},
... {$limit:5}
... ])
[
  { _id: { title: 'The Host' }, rating: 111.3 },
  {
    _id: { title: 'Avengers: Age of Ultron' },
    rating: 88.80000000000001
  },
  { _id: { title: 'The imposible' }, rating: 83.6 },
  { _id: { title: 'There Will Be Blood' }, rating: 81 },
  { _id: { title: 'The Grand Budapest Hotel' }, rating: 81 }
]
Atlas atlas-3apiel-shard-0 [primary] data225_lab2>

Atlas atlas-3apiel-shard-0 [primary] data225_lab2>
```

QUERY & RESULT (PYTHON EXECUTION):

```
result17 = client['data225_lab2']['smd'].aggregate([
{'$match': {'Title': {'$nin': [ None, '' ] } }},
{'$project': {'title': '$Title', 'rating': '$Rating', 'rank': '$rank_id', 'cust': '$customer_id'}},
{'$group': {'_id': {'title': '$title'}, 'rating': {'$sum': '$rating'}, 'cust': {'$addToSet': '$cust'}}},
{'$project': {'_id.title': 1, 'rating': 1}},
{'$sort': {'rating': -1}},
{'$limit': 5}])
result_list17=[]
for document17 in result17:
result_list17.append(document17)
result_list17
s17=pd.DataFrame(data=result_list17)
s17.rename({'_id':'Title'}, axis = 1, inplace = True)
scenario17_df=pd.DataFrame()
scenario17_df['Title']=pd.DataFrame(s17['Title'].tolist())
scenario17_df['Rating']=pd.DataFrame(s17['rating'].tolist())
scenario17_df
```

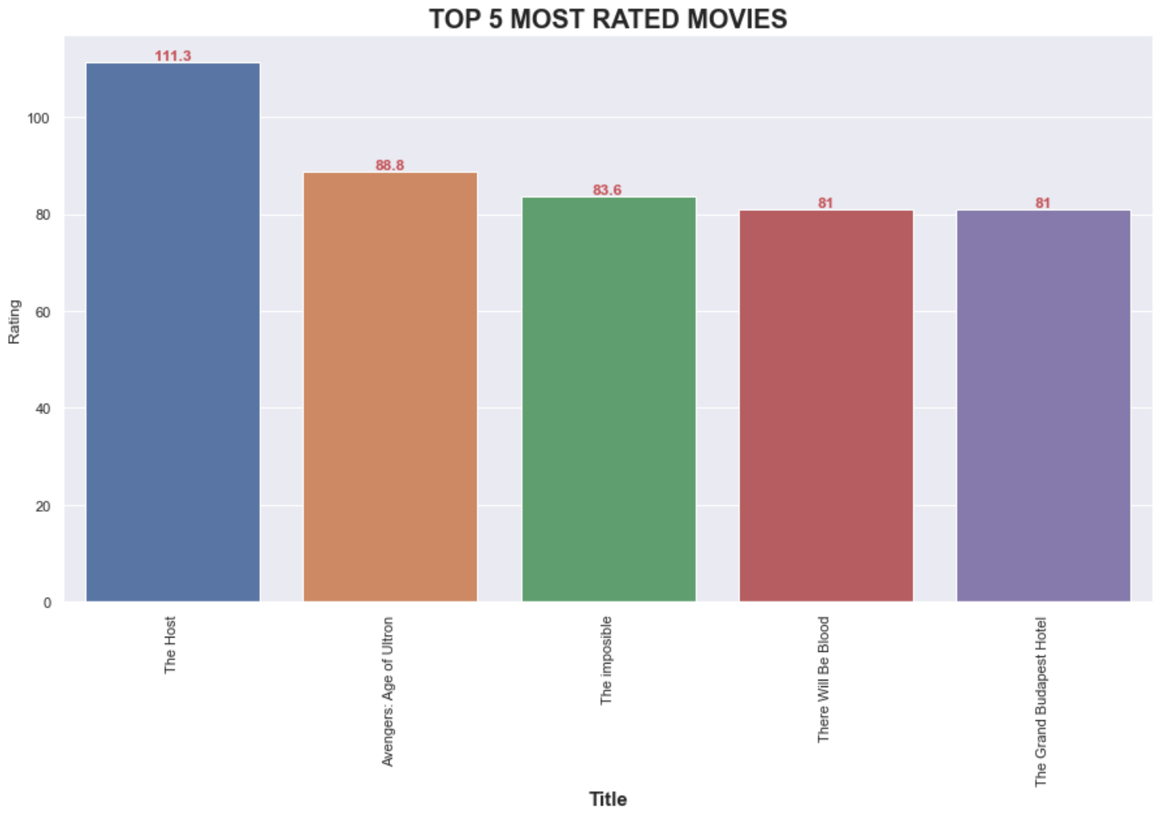
SCENARIO 17 – TOP 5 MOST RATED MOVIES



```python
result17 = client['data225_lab2']['smd'].aggregate([
    {'$match': {'Title': {'$nin': [ None, '' ] } }},
    {'$project': {'title': '$Title', 'rating': '$Rating', 'rank': '$rank_id', 'cust': '$customer_id'}},
    {'$group': {'_id': {'title': '$title'}, 'rating': {'$sum': '$rating'}, 'cust': {'$addToSet': '$cust'}}},
    {'$project': {'_id.title': 1, 'rating': 1}},
    {'$sort': {'rating': -1}},
    {'$limit': 5}])
result_list17=[]
for document17 in result17:
    result_list17.append(document17)
result_list17
s17=pd.DataFrame(data=result_list17)
s17.rename({'_id':'Title'}, axis = 1, inplace = True)
scenario17_df=pd.DataFrame()
scenario17_df['Title']=pd.DataFrame(s17['Title'].tolist())
scenario17_df['Rating']=pd.DataFrame(s17['rating'].tolist())
scenario17_df

```

	Title	Rating
0	The Host	111.3
1	Avengers: Age of Ultron	88.8
2	The imposible	83.6
3	The Grand Budapest Hotel	81.0
4	There Will Be Blood	81.0

JUPYTER VISUALIZATION:



<End of Document>

Submitted By: Gayathri Sundareshwar, Keerthana Gopikrishnan and Deepasha Jenamani

03rd May 2022