LAB 2 - MONGO DB WITH PYTHON QUERIES (Jupyter logs)

Submitted By : Gayathri Sundareshwar, Keerthana Gopikrishnan and Deepasha Jenamani

SCENARIO 1 - NO OF MOVIES WATCHED PER COUNTRY

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 2 - NO OF MOVIES RELEASED PER YEAR

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 3 - TOP 5 EMPLOYEES WHO HAVE RESPONDED TO MOST COMPLAINTS

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 4 - NO OF CUSTOMERS PER COUNTRY

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 5 - NO OF COMPLAINTS RECORDED PER COMPLAINT CATEGORY

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 6 - COMPLAINTS CLOSED PER MONTH

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 7 – COMPLAINTS CREATED PER MONTH

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 8 - PAYMENT MADE PER MONTH

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 9 - TOP 10 CUSTOMERS BASED ON MOVIES WATCHED

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 10 - TOP 10 DIRECTORS BASED ON USER WATCH HISTORY

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 11 - COMPLAINTS BASED ON SEVERITY

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 12 - AVERAGE SALARY BASED ON DESIGNATION

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 13 - EMPLOYEES PER DESIGNATION BASED ON EMPLOYMENT STATUS

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 14 - 5 LATEST CLOSED COMPLAINTS

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 15 – MOST PREFERRED SCREENS BY THE CUSTOMERS

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 16 – TOP HIGHEST GROSSING MOVIES

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 17 – TOP 5 MOST RATED MOVIES

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
```

<End of Document>

Submitted By: Gayathri Sundareshwar, Keerthana Gopikrishnan and Deepasha Jenamani

03rd May 2022