**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