

Athena Queries

Query1: To verify the data imported from S3

```
SELECT * FROM "group15project-db"."datafolder" limit 10
```

Completed									
Time in queue: 139 ms Run time: 750 ms Data scanned: 724.67 KB									
Results (10)									
<input type="text" value="Search rows"/>									
#	state	sex	generalhealth	physicalhealthdays	mentalhealthdays	lastcheckuptime	physicalactivities	sleep	sl
1	Michigan	Female	Fair	0.0	3.0	Within past year (anytime less than 12 months ago)	No	6.0	
2	Michigan	Female	Very good	3.0	0.0	Within past year (anytime less than 12 months ago)	No	8.0	
3	Michigan	Female	Fair	15.0	20.0	Within past year (anytime less than 12 months ago)	No	6.0	
4	Michigan	Female	Good	0.0	1.0	Within past year (anytime less than 12 months ago)	No	5.0	
5	Michigan	Female	Good	0.0	0.0	Within past year (anytime less than 12 months ago)	No	8.0	
6	Michigan	Male	Very good	15.0	0.0	Within past year (anytime less than 12 months ago)	Yes	8.0	
7	Michigan	Female	Fair	0.0	0.0	Within past year (anytime less than 12 months ago)	Yes	9.0	
8	Michigan	Female	Good	2.0	10.0	Within past year (anytime less than 12 months ago)	Yes	7.0	
9	Michigan	Male	Fair	3.0	10.0	Within past year (anytime less than 12 months ago)	Yes	8.0	
10	Michigan	Female	Good	10.0	1.0	Within past year (anytime less than 12 months ago)	Yes	8.0	

Query2: Query to understand the general health status distribution

```
SELECT GeneralHealth, PhysicalActivities, COUNT(*) as Total
FROM datafolder
GROUP BY GeneralHealth, PhysicalActivities
```

1	SELECT GeneralHealth, PhysicalActivities, COUNT(*) as Total
2	FROM datafolder
3	GROUP BY GeneralHealth, PhysicalActivities

SQL Ln 1, Col 1

Run again Explain Cancel Clear Create

Reuse query results up to 60 minutes ago

Query results Query stats

Completed			
Time in queue: 109 ms Run time: 1.064 sec Data scanned: 78.19 MB			
Results (17)			
<input type="text" value="Search rows"/>			
#	GeneralHealth	PhysicalActivities	Total
1	"Good"	"No"	2
2	Good	No	19654
3	Excellent	No	4479
4	"Very good"	"No"	1
5			68
6	Very good	Yes	74699
7	Excellent	Yes	37046
8	Very good	No	13200

Query3: Finding Out no of records for Heart attacks in the dataset

```
SELECT COUNT(*) as HeartAttackCount
FROM datafolder
WHERE HadHeartAttack = 'Yes'
```

The screenshot shows a SQL query execution interface. The query is: `SELECT COUNT(*) as HeartAttackCount FROM datafolder WHERE HadHeartAttack = 'Yes'`. The interface includes a "Run again" button, an "Explain" button, and a "Cancel" button. The query status is "Completed". The results are displayed in a table with one column, "HeartAttackCount", and one row with the value "10430".

#	HeartAttackCount
1	10430

Query4: Query to understand alcohol drinkers in each state

```
SELECT State, AlcoholDrinkers, COUNT(*) as Total
FROM datafolder
WHERE AlcoholDrinkers = 'Yes'
GROUP BY State, AlcoholDrinkers
```

The screenshot shows a SQL query execution interface. The query is: `SELECT State, AlcoholDrinkers, COUNT(*) as Total FROM datafolder WHERE AlcoholDrinkers = 'Yes' GROUP BY State, AlcoholDrinkers`. The interface includes a "Run again" button, an "Explain" button, and a "Cancel" button. The query status is "Completed". The results are displayed in a table with three columns: "State", "AlcoholDrinkers", and "Total".

#	State	AlcoholDrinkers	Total
1	Texas	Yes	866
2	Utah	Yes	220
3	Virginia	Yes	121
4	Wyoming	Yes	55
5	Nebraska	Yes	137
6	North Dakota	Yes	21
7	Oklahoma	Yes	91
8	Illinois	Yes	107

Query5: Query to understand heart attack trends in different age groups

```
SELECT AgeCategory, COUNT(*) as Total
FROM datafolder
WHERE HadHeartAttack = 'Yes'
GROUP BY AgeCategory
```

The screenshot shows a SQL query execution interface. The query is: `SELECT AgeCategory, COUNT(*) as Total FROM datafolder WHERE HadHeartAttack = 'Yes' GROUP BY AgeCategory`. The query is executed successfully, as indicated by the 'Completed' status. The results table shows the total count for each age category.

#	AgeCategory	Total
1	Age 25 to 29	10
2	Age 50 to 54	62
3	"Black only	20
4	Age 80 or older	36
5	Age 35 to 39	27
6	Age 60 to 64	95
7	Non-Hispanic"	9472

Other SQL Queries used for exploring in Athena:

```
SELECT SleepHours, COUNT(*) as NumberOfPeople
FROM your_table
GROUP BY SleepHours
ORDER BY SleepHours;
```

```
SELECT State, Sex, AVG(BMI) as AverageBMI
FROM your_table
GROUP BY State, Sex;
```

```
SELECT AgeCategory, LastCheckupTime, COUNT(*) as Total
FROM your_table
GROUP BY AgeCategory, LastCheckupTime;
```

```
SELECT PhysicalActivities, AVG(SleepHours) as AvgSleepHours
FROM your_table
GROUP BY PhysicalActivities;
```

```
SELECT State, COUNT(*) as Total
FROM your_table
WHERE BlindOrVisionDifficulty = 'Yes'
GROUP BY State;
```

```
SELECT COUNT(*) as Total
FROM your_table
WHERE GeneralHealth = 'Poor' AND MentalHealthDays > 5;
```