

SQL Project with Community Health Dataset from Kaggle

The goal of this project was to provide deeper evaluation into community health services and health quality based on biomechanical time series. See the summary below for how I conducted analysis on this community health dataset that was sourced from Kaggle.

Here are the hypothetical scenarios that I created to analyze the data, along with my solutions and results.

1. Joint Angle Variation by Gender and Age Group

```
--Joint angle variation by gender and age group

select
  gender
  , case
      when age < 30 then '18-29'
      when age between 30 and 49 then '30-49'
      when age between 50 and 65 then '50-65'
      else '66-70'
    end as age_group
  , round(avg(joint_angle), 2) as avg_joint_angle
  , count(*) as participant_count
from
  community_health
group by
  gender
  , age_group
order by
  gender
  , age_group
;
```

This query analyzes the average joint angle during movement, categorized by gender and age group, to explore differences in biomechanics across demographics.

	gender character (1)	age_group text	avg_joint_angle numeric	participant_count bigint
1	F	18-29	20.28	38
2	F	30-49	20.23	74
3	F	50-65	19.72	59
4	F	66-70	20.47	13
5	M	18-29	19.88	44
6	M	30-49	20.07	54
7	M	50-65	20.38	50
8	M	66-70	18.98	15

From the result set, we can see that the data suggests potential differences in joint angle measurements influenced by both age and gender, which could be valuable for targeted health interventions or further research into the impact of age and gender on joint health. The variation in participant counts across age groups highlights the importance of ensuring balanced representation in future studies to strengthen findings related to specific demographics.

2. Service Utilization and Visit Frequency

```
--Evaluate average patient satisfaction scores based on visit frequency and service type
--Understand how often participants utilize services
```

```
select
  service_type
  , visit_frequency
  , ROUND(avg(patient_satisfaction), 2) as avg_satisfaction
  , count(*) as participant_count
from
  community_health
group by
  service_type
  , visit_frequency
order by
  service_type
  , visit_frequency
;
```

Trying to understand how often participants utilize services, I evaluated the average patient satisfaction scores based on visit frequency and service type.

	service_type character varying (50) 🔒	visit_frequency character varying (50) 🔒	avg_satisfaction numeric 🔒	participant_count bigint 🔒
1	Consultation	Monthly	4.64	39
2	Consultation	Weekly	5.97	38
3	Consultation	Yearly	5.40	40
4	Preventive	Monthly	5.64	33
5	Preventive	Weekly	5.34	38
6	Preventive	Yearly	5.04	48
7	Rehab	Monthly	5.18	39
8	Rehab	Weekly	5.00	44
9	Rehab	Yearly	4.71	28

The data suggest that increased frequency of visits, for consultations, is associated with higher patient satisfaction, emphasizing the need for healthcare providers to encourage regular follow-ups. The variation in satisfaction levels among different service types and visit frequencies highlights areas for improvement.

3. Identifying Patterns in EMG Activity, Step Frequency, and Quality of Life

```
--Identify patterns in EMG activity levels, step frequency, and quality of life scores

select
    emg_activity
    , round(avg(quality_of_life_score), 2) as avg_quality_of_life
    , round(avg(step_frequency), 2) as avg_step_frequency
    , count(participant_id) as participant_count
from
    community_health
group by
    emg_activity
order by
    avg_quality_of_life desc
;
```

To reveal if there is any correlation on how muscle activity correlates with overall health perceptions from this data, I analyzed the average quality of life scores based on muscle activity as measured by electromyography (emg) and the average number of steps taken per minute.

	emg_activity character varying (50) 🔒	avg_quality_of_life numeric 🔒	avg_step_frequency numeric 🔒	participant_count bigint 🔒
1	Low	74.33	78.53	109
2	Moderate	74.16	80.91	121
3	High	74.11	80.79	117

Conversely to what I would have hypothesized, the group of participants that showed lower emg activity and a lower average step frequency produced the highest average quality of life scores. I wanted to explore further so I looked at the top 20 quality of life scores.

```
--Top 20 quality of life scores

select
    participant_id
    , quality_of_life_score
    , emg_activity
    , step_frequency
    , ses
from
    community_health
order by
    quality_of_life_score desc
limit
    20
;
```

I also included the socioeconomic status (ses) along with the emg activity and step frequency in this query to see if that was a factor in the quality-of-life scores.

	participant_id bigint	quality_of_life_score integer	emg_activity character varying (50)	step_frequency integer	ses integer
1	39	99	High	75	4
2	130	99	Low	70	1
3	132	99	Moderate	83	1
4	18	99	Low	98	2
5	166	99	High	86	4
6	153	98	Moderate	86	1
7	188	98	Moderate	86	4
8	281	98	High	86	2
9	5	98	High	73	3
10	42	98	Moderate	93	3
11	156	98	Low	65	1
12	233	97	Low	65	2
13	227	97	High	94	3
14	190	97	High	99	4
15	14	96	Moderate	89	2
16	214	96	Low	76	1
17	36	96	High	72	3
18	98	96	Moderate	68	4
19	201	96	Moderate	93	4
20	232	96	Low	92	2

I didn't find any big trends here. Interestingly, if you look at participant 156 on row 11, they had a low level of emg activity, a step frequency of 65 (one of only three participants that were 68 or below in this category), and are at the lowest socioeconomic ranking (ses) of 1 but still recorded the 11th highest quality-of-life score among all participants.

4. Patient Satisfaction by Service Type

```
--Patient satisfaction by service type

select
    service_type
    , round(avg(patient_satisfaction), 2) as avg_satisfaction
    , count(*) as participant_count
from
    community_health
group by
    service_type
order by
    avg_satisfaction desc
;
```

The last thing I looked closer at was the patient satisfaction levels based on the type of health service accessed, helping to identify which services were leading to higher satisfaction.

	service_type character varying (50) 🔒	avg_satisfaction numeric 🔒	participant_count bigint 🔒
1	Consultation	5.33	117
2	Preventive	5.30	119
3	Rehab	4.99	111

The participant count indicated a healthy level of engagement with each service type, allowing for more reliable averages in satisfaction scores. Given that rehab services have the lowest satisfaction score, healthcare providers might consider investigating participant feedback to identify specific areas for improvement. The results could also indicate that personalized care and proactive health management contribute positively to patient experiences as reflected in the higher satisfaction scores for consultation and preventive services.

This SQL project provided deeper evaluation into community health services and health quality based on biomechanical time series. Thank you for taking the time to explore.