ChatGPT vs Gemini Performance Analysis

vs

Introduction:

- **Challenge:** your challenge is to craft SQL queries to extract insights from the Gemini Vs ChatGPT database.

- **Tables:**

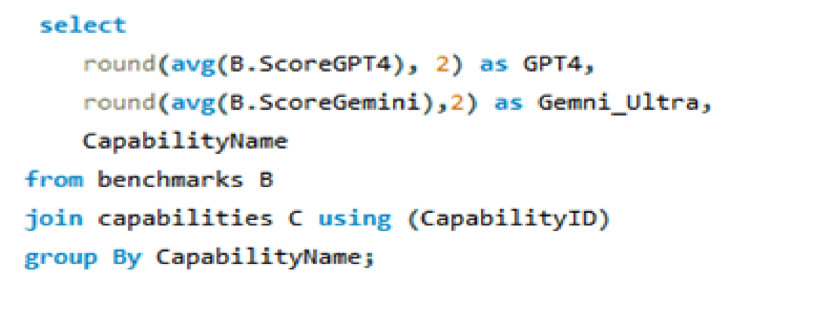
  1. Models

  2. Capabilities

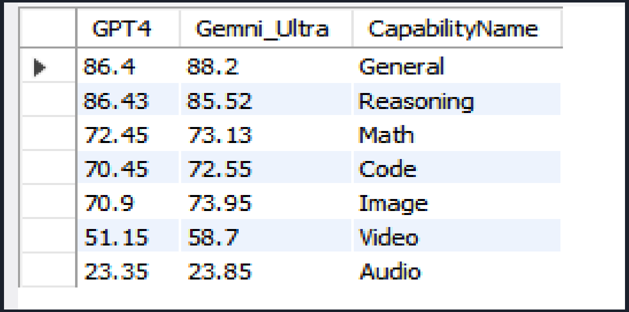
  3. Benchmarks

1. What are the average scores for each capability on both the Gemini Ultra and GPT-4 models?

Solution:



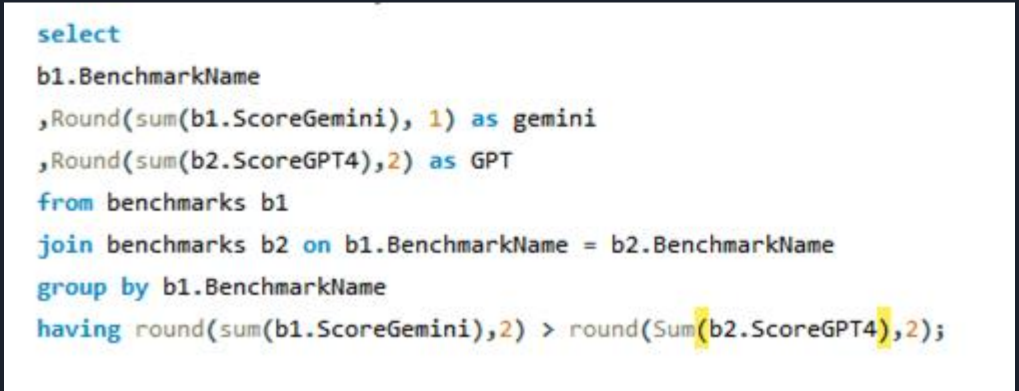
```
 select
    round(avg(B.ScoreGPT4), 2) as GPT4,
    round(avg(B.ScoreGemini),2) as Gemni_Ultra,
    CapabilityName
from benchmarks B
join capabilities C using (CapabilityID)
group By CapabilityName;
```



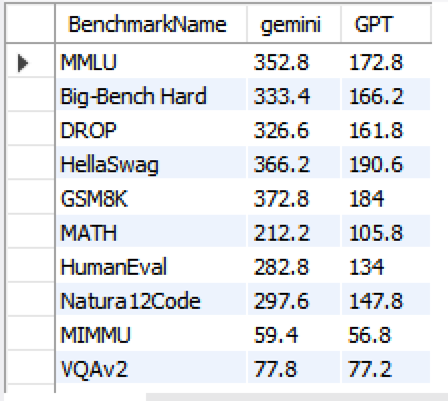
GPT4	Gemni_Ultra	CapabilityName
86.4	88.2	General
86.43	85.52	Reasoning
72.45	73.13	Math
70.45	72.55	Code
70.9	73.95	Image
51.15	58.7	Video
23.35	23.85	Audio

2. Which benchmarks does Gemini Ultra outperform GPT-4 in terms of scores?

Solution:



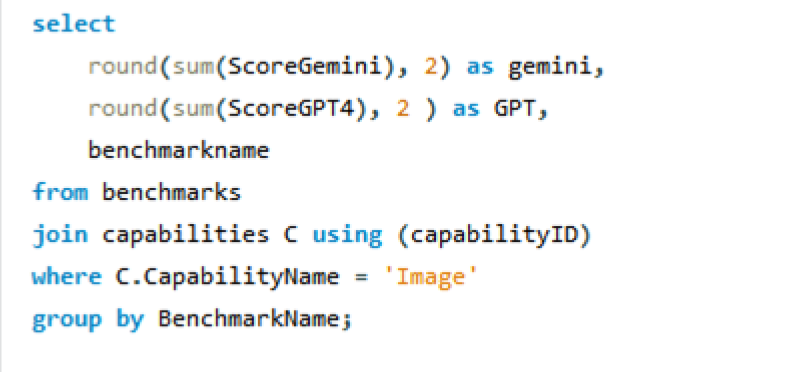
```sql
select
b1.BenchmarkName
,Round(sum(b1.ScoreGemini), 1) as gemini
,Round(sum(b2.ScoreGPT4),2) as GPT
from benchmarks b1
join benchmarks b2 on b1.BenchmarkName = b2.BenchmarkName
group by b1.BenchmarkName
having round(sum(b1.ScoreGemini),2) > round(Sum(b2.ScoreGPT4),2);
```



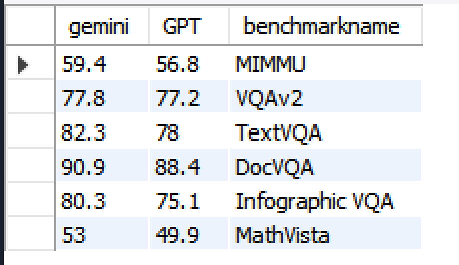
BenchmarkName	gemini	GPT
MMLU	352.8	172.8
Big-Bench Hard	333.4	166.2
DROP	326.6	161.8
HellaSwag	366.2	190.6
GSM8K	372.8	184
MATH	212.2	105.8
HumanEval	282.8	134
Natura12Code	297.6	147.8
MIMMU	59.4	56.8
VQAv2	77.8	77.2

3. What are the highest scores achieved by Gemini Ultra and GPT-4 for each benchmark in the Image capability ?

Solution:



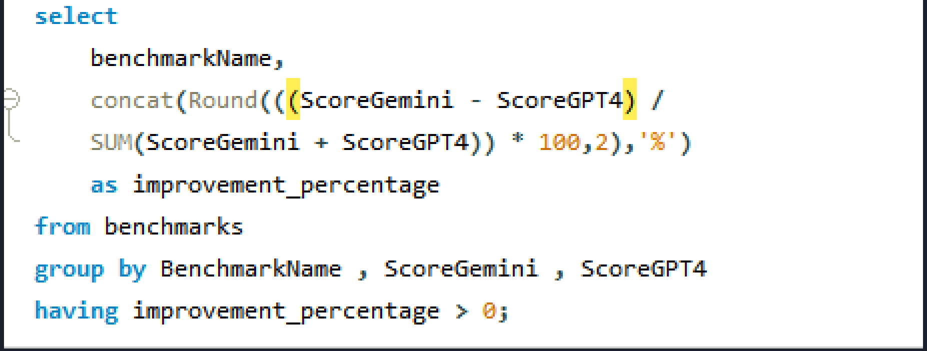
```sql
select
    round(sum(ScoreGemini), 2) as gemini,
    round(sum(ScoreGPT4), 2 ) as GPT,
    benchmarkname
from benchmarks
join capabilities C using (capabilityID)
where C.CapabilityName = 'Image'
group by BenchmarkName;
```



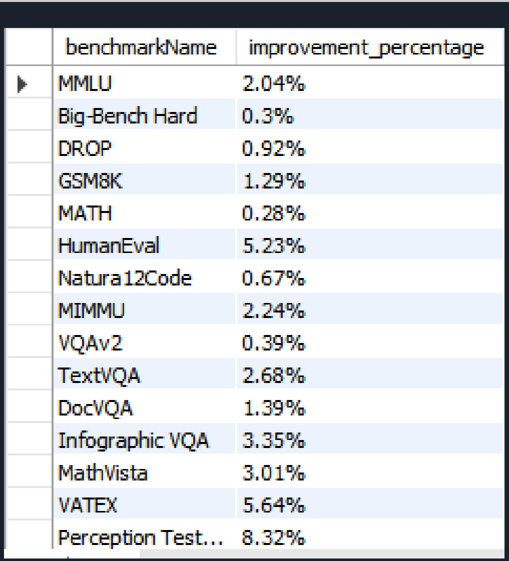
	gemini	GPT	benchmarkname
▶	59.4	56.8	MIMMU
	77.8	77.2	VQAv2
	82.3	78	TextVQA
	90.9	88.4	DocVQA
	80.3	75.1	Infographic VQA
	53	49.9	MathVista

4. Calculate the percentage improvement of Gemini Ultra over GPT-4 for each Benchmark?

Solution:



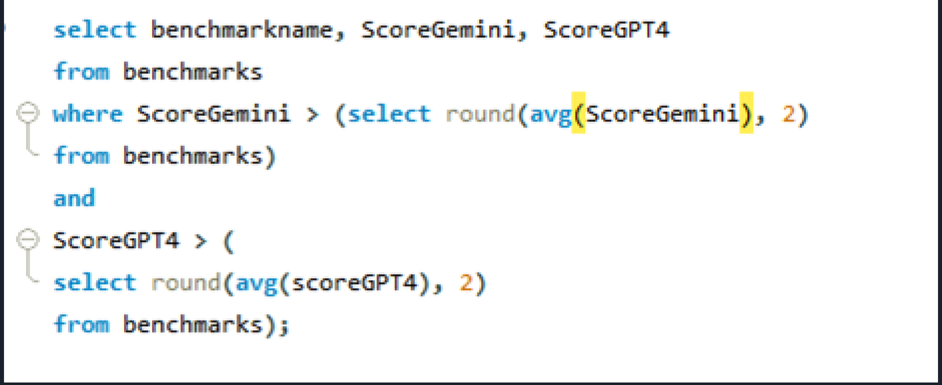
```
select
    benchmarkName,
    concat(Round(((ScoreGemini - ScoreGPT4) /
    SUM(ScoreGemini + ScoreGPT4)) * 100,2),'%')
    as improvement_percentage
from benchmarks
group by BenchmarkName , ScoreGemini , ScoreGPT4
having improvement_percentage > 0;
```



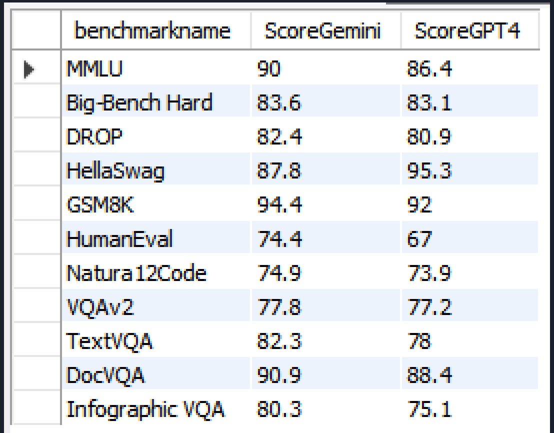
benchmarkName	improvement_percentage
MMLU	2.04%
Big-Bench Hard	0.3%
DROP	0.92%
GSM8K	1.29%
MATH	0.28%
HumanEval	5.23%
Natura12Code	0.67%
MIMMU	2.24%
VQAv2	0.39%
TextVQA	2.68%
DocVQA	1.39%
Infographic VQA	3.35%
MathVista	3.01%
VATEX	5.64%
Perception Test...	8.32%

5. Retrieve the benchmarks where both models scored above the average for their respective models?

Solution:



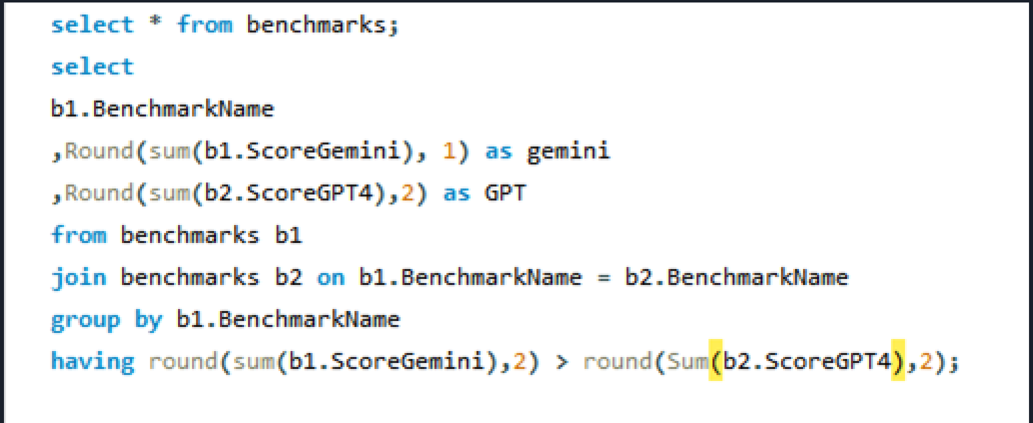
```
select benchmarkname, ScoreGemini, ScoreGPT4
from benchmarks
where ScoreGemini > (select round(avg(ScoreGemini), 2)
from benchmarks)
and
ScoreGPT4 > (
select round(avg(scoreGPT4), 2)
from benchmarks);
```



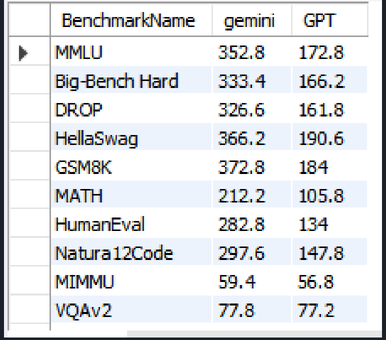
benchmarkname	ScoreGemini	ScoreGPT4
MMLU	90	86.4
Big-Bench Hard	83.6	83.1
DROP	82.4	80.9
HellaSwag	87.8	95.3
GSM8K	94.4	92
HumanEval	74.4	67
Natura12Code	74.9	73.9
VQAv2	77.8	77.2
TextVQA	82.3	78
DocVQA	90.9	88.4
Infographic VQA	80.3	75.1

6. Which benchmarks show that Gemini Ultra is expected to outperform GPT-4 based on the next score?

Solution:



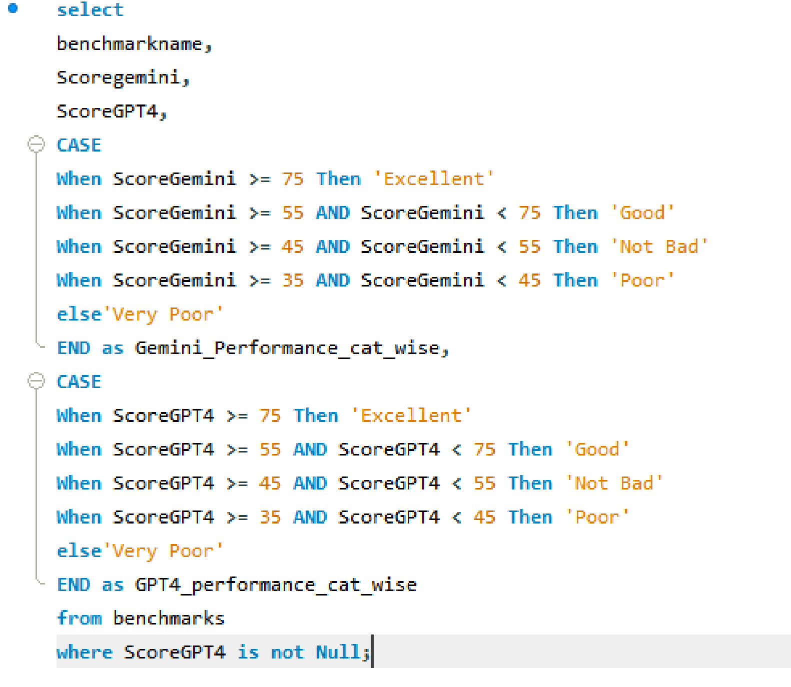
```
select * from benchmarks;
select
b1.BenchmarkName
,Round(sum(b1.ScoreGemini), 1) as gemini
,Round(sum(b2.ScoreGPT4),2) as GPT
from benchmarks b1
join benchmarks b2 on b1.BenchmarkName = b2.BenchmarkName
group by b1.BenchmarkName
having round(sum(b1.ScoreGemini),2) > round(Sum(b2.ScoreGPT4),2);
```



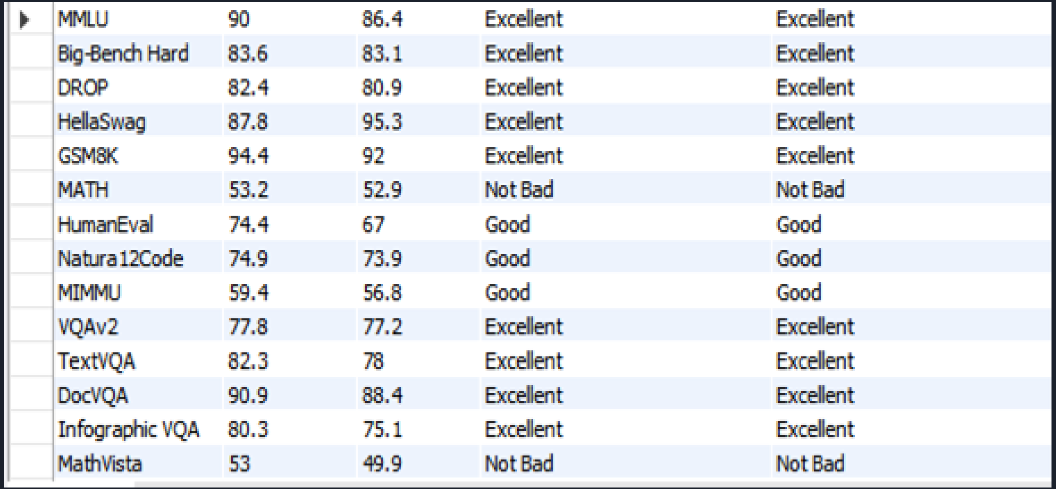
BenchmarkName	gemini	GPT
MMLU	352.8	172.8
Big-Bench Hard	333.4	166.2
DROP	326.6	161.8
HellaSwag	366.2	190.6
GSM8K	372.8	184
MATH	212.2	105.8
HumanEval	282.8	134
Natura12Code	297.6	147.8
MIMMU	59.4	56.8
VQAv2	77.8	77.2

7. Classify benchmarks into performance categories based on score ranges?

Solution:



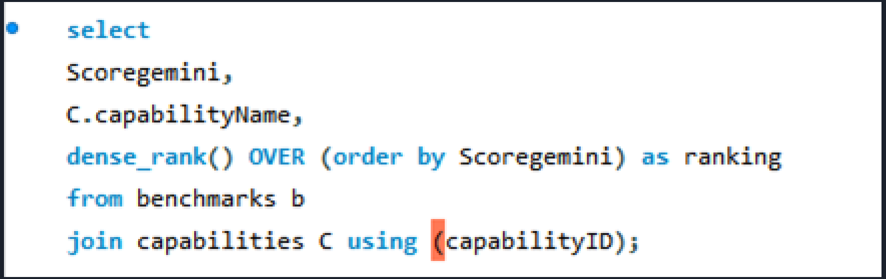
```
select
benchmarkname,
Scoregemini,
ScoreGPT4,
CASE
When ScoreGemini >= 75 Then 'Excellent'
When ScoreGemini >= 55 AND ScoreGemini < 75 Then 'Good'
When ScoreGemini >= 45 AND ScoreGemini < 55 Then 'Not Bad'
When ScoreGemini >= 35 AND ScoreGemini < 45 Then 'Poor'
else'Very Poor'
END as Gemini_Performance_cat_wise,
CASE
When ScoreGPT4 >= 75 Then 'Excellent'
When ScoreGPT4 >= 55 AND ScoreGPT4 < 75 Then 'Good'
When ScoreGPT4 >= 45 AND ScoreGPT4 < 55 Then 'Not Bad'
When ScoreGPT4 >= 35 AND ScoreGPT4 < 45 Then 'Poor'
else'Very Poor'
END as GPT4_performance_cat_wise
from benchmarks
where ScoreGPT4 is not Null;
```



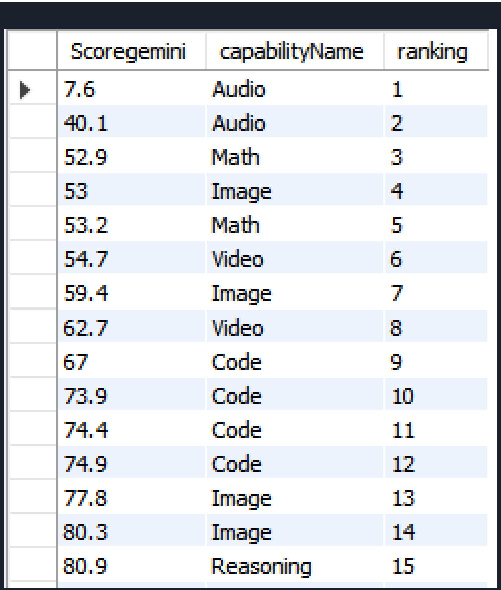
MMLU	90	86.4	Excellent	Excellent
Big-Bench Hard	83.6	83.1	Excellent	Excellent
DROP	82.4	80.9	Excellent	Excellent
HellaSwag	87.8	95.3	Excellent	Excellent
GSM8K	94.4	92	Excellent	Excellent
MATH	53.2	52.9	Not Bad	Not Bad
HumanEval	74.4	67	Good	Good
Natura12Code	74.9	73.9	Good	Good
MIMMU	59.4	56.8	Good	Good
VQAv2	77.8	77.2	Excellent	Excellent
TextVQA	82.3	78	Excellent	Excellent
DocVQA	90.9	88.4	Excellent	Excellent
Infographic VQA	80.3	75.1	Excellent	Excellent
MathVista	53	49.9	Not Bad	Not Bad

8. Retrieve the rankings for each capability based on Gemini Ultra scores?

Solution:



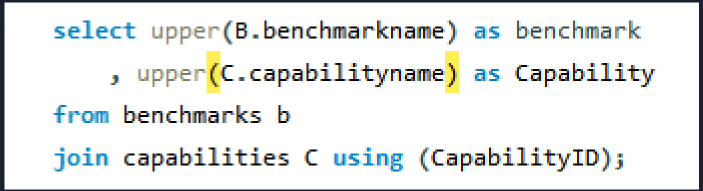
```
select
Scoregemini,
C.capabilityName,
dense_rank() OVER (order by Scoregemini) as ranking
from benchmarks b
join capabilities C using (capabilityID);
```



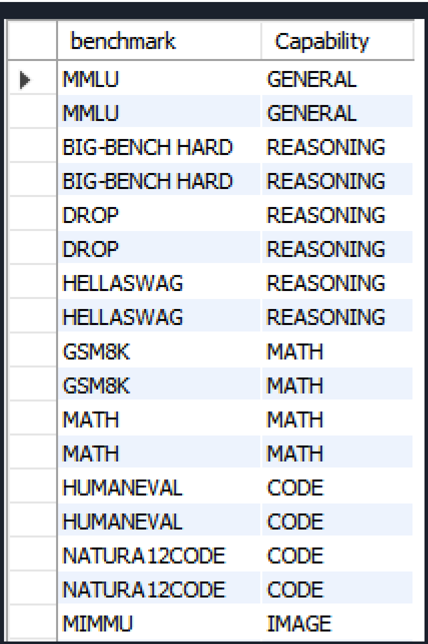
Scoregemini	capabilityName	ranking
7.6	Audio	1
40.1	Audio	2
52.9	Math	3
53	Image	4
53.2	Math	5
54.7	Video	6
59.4	Image	7
62.7	Video	8
67	Code	9
73.9	Code	10
74.4	Code	11
74.9	Code	12
77.8	Image	13
80.3	Image	14
80.9	Reasoning	15

9. Convert the Capability and Benchmark name to uppercase?

Solution:



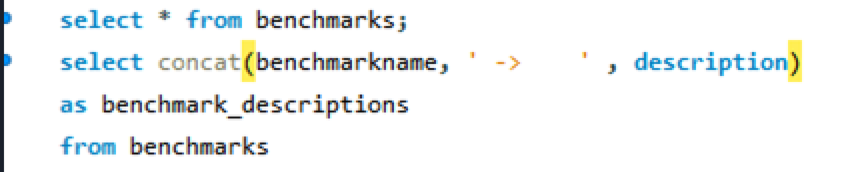
```
select upper(B.benchmarkname) as benchmark
    , upper(C.capabilityname) as Capability
from benchmarks b
join capabilities C using (CapabilityID);
```



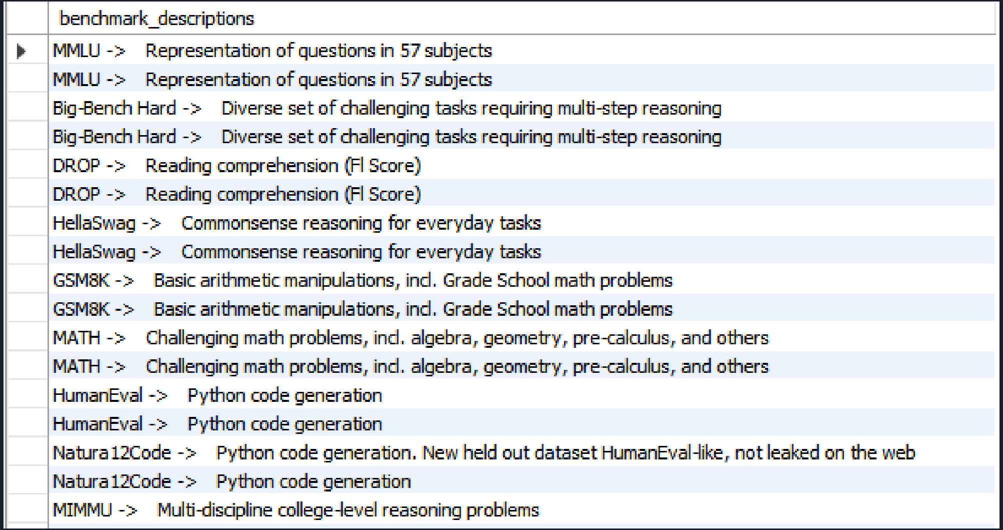
benchmark	Capability
MMLU	GENERAL
MMLU	GENERAL
BIG-BENCH HARD	REASONING
BIG-BENCH HARD	REASONING
DROP	REASONING
DROP	REASONING
HELLASWAG	REASONING
HELLASWAG	REASONING
GSM8K	MATH
GSM8K	MATH
MATH	MATH
MATH	MATH
HUMANEVAL	CODE
HUMANEVAL	CODE
NATURA12CODE	CODE
NATURA12CODE	CODE
MIMMU	IMAGE

10. Can you provide the benchmarks along with their descriptions in a concatenated format?

Solution:



```
select * from benchmarks;
select concat(benchmarkname, ' ->     ' , description)
as benchmark_descriptions
from benchmarks
```



benchmark_descriptions
MMLU -> Representation of questions in 57 subjects
MMLU -> Representation of questions in 57 subjects
Big-Bench Hard -> Diverse set of challenging tasks requiring multi-step reasoning
Big-Bench Hard -> Diverse set of challenging tasks requiring multi-step reasoning
DROP -> Reading comprehension (Fl Score)
DROP -> Reading comprehension (Fl Score)
HellaSwag -> Commonsense reasoning for everyday tasks
HellaSwag -> Commonsense reasoning for everyday tasks
GSM8K -> Basic arithmetic manipulations, incl. Grade School math problems
GSM8K -> Basic arithmetic manipulations, incl. Grade School math problems
MATH -> Challenging math problems, incl. algebra, geometry, pre-calculus, and others
MATH -> Challenging math problems, incl. algebra, geometry, pre-calculus, and others
HumanEval -> Python code generation
HumanEval -> Python code generation
Natura12Code -> Python code generation. New held out dataset HumanEval-like, not leaked on the web
Natura12Code -> Python code generation
MIMMU -> Multi-discipline college-level reasoning problems

Thankyou