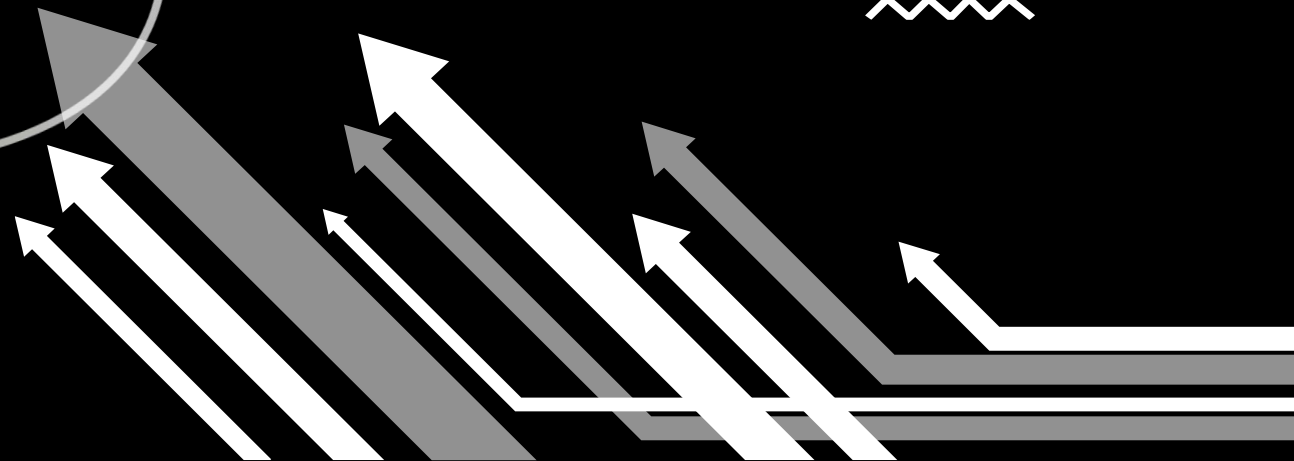








DS SALARY CASE STUDY

**MYSQL
PRESENTED BY-
YOGESH CHOUHAN**



- 
- 
- 
- 
- 
- 
- ✓ **WILD CARDS**
 - ✓ **SUBQURIES**
 - ✓ **WINDOW FUNCTIONS**
 - ✓ **CREATING VARIABLES**
 - ✓ **COMMON TABLE EXPRESSION**
 - ✓ **CREATE PIVOT TABLE**
 - ✓ **JOINS**
 - ✓ **ACCESS CONTROL**

TOPICS



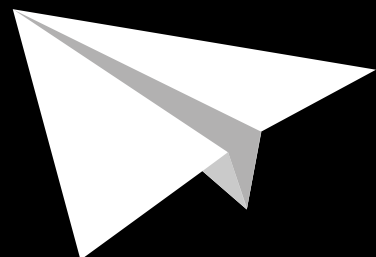


You're a Compensation analyst employed by a multinational corporation. Your Assignment is to Pinpoint Countries who give work fully remotely, for the title 'managers' Paying salaries Exceeding \$90,000 USD



```
SELECT DISTINCT
  company_location
FROM
  salary
WHERE
  job_title LIKE '%manager%'
  AND salary_in_usd > 90000
  AND remote_ratio = 100;
```

	company_location
▶	US
	MX
	AU
	FR



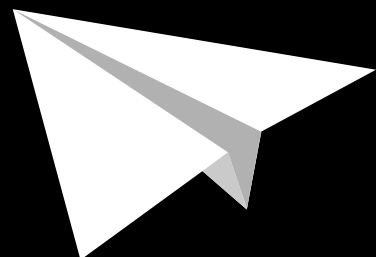


You're a Compensation analyst employed by a multinational corporation. Your Assignment is to Pinpoint Countries who give work fully remotely, for the title 'managers' Paying salaries Exceeding \$90,000 USD



```
SELECT DISTINCT
    company_location
FROM
    salary
WHERE
    job_title LIKE '%manager%'
    AND salary_in_usd > 90000
    AND remote_ratio = 100;
```

	company_location
▶	US
	MX
	AU
	FR





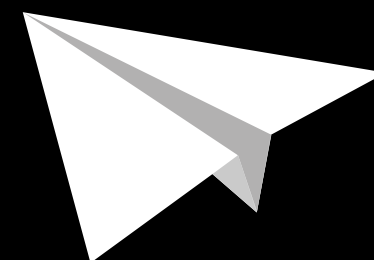
Picture yourself AS a data scientist Working for a workforce management platform. Your objective is to calculate the percentage of employees. Who enjoy fully remote roles WITH salaries Exceeding \$100,000 USD, Shedding light ON the attractiveness of high-paying remote positions IN today's job market.



```
• SELECT
  job_title,
  total,
  remote_worker,
  (remote_worker / total) * 100 AS 'r_work_percentage'
FROM
  (SELECT
    m.job_title, m.total, n.remote_worker
  FROM
    (SELECT
      job_title, COUNT(*) AS 'total'
    FROM
      salary
    GROUP BY job_title) m
  INNER JOIN (SELECT
    t.job_title, COUNT(*) AS 'remote_worker'
  FROM
    (SELECT
      *
    FROM
      salary
    WHERE
      salary_in_usd > 100000
      AND remote_ratio = 100) t
  GROUP BY job_title) n ON m.job_title = n.job_title) t;
```

job_title	total	remote_worker	r_work_percentage
AI Engineer	89	15	16.8539
Machine Learning Engineer	1466	347	23.6698
Business Intelligence Developer	49	7	14.2857
Data Engineer	3017	774	25.6546
Data Scientist	2874	842	29.2971
Cloud Database Engineer	7	7	100.0000
Research Engineer	260	18	6.9231
Data Analyst	2079	405	19.4805
Machine Learning Scientist	114	38	33.3333
Applied Scientist	367	102	27.7929
Data Science Manager	108	43	39.8148
Research Scientist	441	67	15.1927
Prompt Engineer	13	2	15.3846
Data Science	172	34	19.7674
Data Science Consultant	73	4	5.4795
Data Management Analyst	12	1	8.3333
Research Analyst	97	8	8.2474
Business Intelligence Analyst	121	26	21.4876
Analytics Engineer	387	157	40.5685
Data Architect	337	146	43.3234
Data Manager	176	9	5.1136
ML Engineer	157	45	28.6624
Machine Learning Researcher	21	5	23.8095
AI Architect	74	7	9.4667

Result 3 x





Picture yourself AS a data scientist Working for a workforce management platform. Your objective is to calculate the percentage of employees. Who enjoy fully remote roles WITH salaries Exceeding \$100,000 USD, Shedding light ON the attractiveness of high-paying remote positions IN today's job market.

```
53 • set @total = (select count(*) from salary where salary_in_usd>100000);
54 • set @r_total = (select count(*) from salary where salary_in_usd>100000 and remote_ratio=100);
55 • SELECT ((@r_total) / (@total)) * 100 AS 'percentage';
56
```

```
• SELECT
  job_title,
  total,
  remote_worker,
  (remote_worker / total) * 100 AS 'r_work_percentage'
FROM
  (SELECT
    m.job_title, m.total, n.remote_worker
  FROM
    (SELECT
      job_title, COUNT(*) AS 'total'
    FROM
      salary
    GROUP BY job_title) m
  INNER JOIN (SELECT
    t.job_title, COUNT(*) AS 'remote_worker'
  FROM
    (SELECT
      *
    FROM
      salary
    WHERE
      salary_in_usd > 100000
      AND remote_ratio = 100) t
  GROUP BY job_title) n ON m.job_title = n.job_title) t;
```

job_title	total	remote_worker	r_work_percentage
AI Engineer	89	15	16.8539
Machine Learning Engineer	1466	347	23.6698
Business Intelligence Developer	49	7	14.2857
Data Engineer	3017	774	25.6546
Data Scientist	2874	842	29.2971
Cloud Database Engineer	7	7	100.0000
Research Engineer	260	18	6.9231
Data Analyst	2079	405	19.4805
Machine Learning Scientist	114	38	33.3333
Applied Scientist	367	102	27.7929
Data Science Manager	108	43	39.8148
Research Scientist	441	67	15.1927
Prompt Engineer	13	2	15.3846
Data Science	172	34	19.7674
Data Science Consultant	73	4	5.4795
Data Management Analyst	12	1	8.3333
Research Analyst	97	8	8.2474
Business Intelligence Analyst	121	26	21.4876
Analytics Engineer	387	157	40.5685
Data Architect	337	146	43.3234
Data Manager	176	9	5.1136
ML Engineer	157	45	28.6624
Machine Learning Researcher	21	5	23.8095
AI Architect	24	7	29.1667

Result 3 x



Imagine you're a data analyst Working for a global recruitment agency. Your Task is to identify the Locations where entry-level average salaries exceed the average salary for that job title IN market for entry level, helping your agency guide candidates towards lucrative opportunities..



```
1 SELECT
2     company_location, job_title, c_avg_salary, a_avg_salary
3 FROM
4     (SELECT
5         m.company_location,
6         m.job_title,
7         m.c_avg_salary,
8         n.a_avg_salary
9     FROM
10         (SELECT
11             company_location,
12             job_title,
13             AVG(salary_in_usd) AS 'c_avg_salary'
14         FROM
15             salary
16         WHERE
17             experience_level = 'EN'
18         GROUP BY company_location , job_title) m
19     INNER JOIN (SELECT
20         job_title, AVG(salary_in_usd) AS 'a_avg_salary'
21     FROM
22         salary
23     WHERE
24         experience_level = 'EN'
25     GROUP BY job_title) n ON m.job_title = n.job_title) t
26 WHERE
27     c_avg_salary > a_avg_salary;
```

	company_location	job_title	c_avg_salary	a_avg_salary
▶	US	Data Analyst	89800.3519	84808.6361
	US	Analytics Engineer	110831.2500	96722.3000
	US	Data Engineer	106791.2584	92713.4701
	US	Research Analyst	110459.5385	107294.2143
	US	Machine Learning Engineer	126188.8529	110718.3778
	AU	Business Intelligence Analyst	91000.0000	76688.6333
	US	Business Intelligence Analyst	79243.9500	76688.6333
	US	Research Engineer	162668.1818	135467.7241
	CA	Machine Learning Research Engineer	80769.0000	55653.6667
	US	Data Scientist	104573.4333	87028.3731
	MX	Data Analyst	429950.0000	84808.6361
	US	Data Science	100450.0000	85511.7857
	IT	AI Engineer	44444.0000	33679.0000
	US	Research Scientist	168206.9231	158548.5517
	CA	Machine Learning Engineer	131600.0000	110718.3778
	LB	Data Quality Analyst	71750.0000	57407.0000
	US	ML Engineer	142000.0000	81663.1667
	GB	AI Research Engineer	36940.0000	28734.0000
	IN	AI Programmer	60207.0000	56858.8000
	DE	AI Engineer	35000.0000	33679.0000
	US	AI Developer	130000.0000	110119.5000
	US	Machine Learning Scientist	152500.0000	113224.0000
	CA	Data Integration Specialist	100000.0000	96605.0000
	US	BI Data Analyst	64250.0000	56134.0000
	CA	AI Programmer	74087.0000	56858.8000
	US	Compliance Data Analyst	60000.0000	45000.0000
	US	Computer Vision Engineer	145000.0000	115000.2500



You've been hired by a big HR Consultancy to look at how much people get paid IN different Countries. Your job is to Find out for each job title which. Country pays the maximum average salary.
This helps you to place your candidates IN those countries...

```
103 • select job_title,company_location,avg_salary from
104 (
105     select job_title,company_location,avg(salary_in_usd) as 'avg_salary',
106     dense_rank() over(partition by job_title order by avg(salary_in_usd) desc) as 'ranking'
107     from salary
108     group by job_title,company_location
109 ) t where ranking=1
110 ;
```

	job_title	company_location	avg_salary
▶	Admin & Data Analyst	IN	60000.0000
	AI Architect	CA	800000.0000
	AI Developer	CA	275000.0000
	AI Engineer	QA	300000.0000
	AI Product Manager	US	152650.0000
	AI Programmer	FR	120000.0000
	AI Research Engineer	US	175000.0000
	AI Research Scientist	DE	88888.0000
	AI Scientist	IL	417937.0000
	AI Software Engineer	EG	174100.0000
	Analytics Engineer	FR	188000.0000
	Analytics Engineerin...	GB	399880.0000
	Applied Data Scientist	US	238000.0000
	Applied Machine Lear...	US	177500.0000

Result 1 x

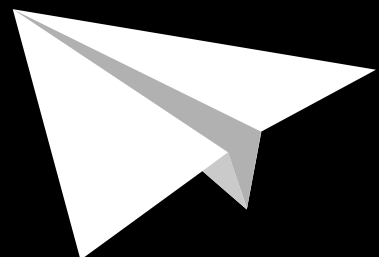


AS a data-driven Business consultant, you've been hired by a multinational corporation to analyze salary trends across different company Locations. Your goal is to Pinpoint Locations WHERE the average salary Has consistently Increased over the Past few years (Countries WHERE data is available for 3 years Only(present year and past two years) providing Insights into Locations experiencing Sustained salary growth.



```
WITH t AS
(
  SELECT * FROM salaries WHERE company_location IN
  (
    SELECT company_location FROM
    (
      SELECT company_location, AVG(salary_IN_usd) AS AVG_salary, COUNT(DISTINCT work_year) AS num_years FROM salaries WHERE work_year >= YEAR(CURRENT_DATE()) - 2
      GROUP BY company_location HAVING num_years = 3
    )m
  )
) -- step 4
-- SELECT company_location, work_year, AVG(salary_IN_usd) AS average FROM t GROUP BY company_location, work_year
SELECT
  company_location,
  MAX(CASE WHEN work_year = 2022 THEN average END) AS AVG_salary_2022,
  MAX(CASE WHEN work_year = 2023 THEN average END) AS AVG_salary_2023,
  MAX(CASE WHEN work_year = 2024 THEN average END) AS AVG_salary_2024
FROM
(
  SELECT company_location, work_year, AVG(salary_IN_usd) AS average FROM t GROUP BY company_location, work_year
)q GROUP BY company_location HAVING AVG_salary_2024 > AVG_salary_2023 AND AVG_salary_2023 > AVG_salary_2022 -- step 3 and having step 4.
```

	company_location
▶	AR
	CA
	ES
	FI
	FR
	HU
	IN
	PT





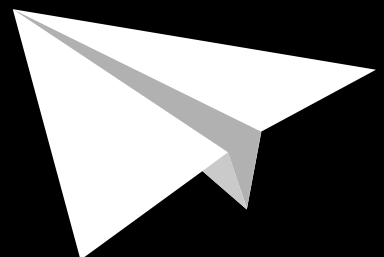
Picture yourself AS a workforce strategist employed by a global HR tech startup. Your Mission is to Determine the percentage of fully remote work for each experience level IN 2021 and compare it WITH the corresponding figures for 2024, Highlighting any significant Increases or decreases IN remote work Adoption over the years..



```
12 • SELECT
13 *
14 FROM
15 (SELECT
16     m.experience_level,
17     m.work_year,
18     (remote_total / total) * 100 AS 'percentage'
19 FROM
20     (SELECT
21         experience_level, work_year, COUNT(*) AS 'total'
22 FROM
23         salary
24 WHERE
25         work_year = 2021
26 GROUP BY experience_level) m
27 INNER JOIN (SELECT
28     experience_level, work_year, COUNT(*) AS 'remote_total'
29 FROM
30     salary
31 WHERE
32     work_year = 2021 AND remote_ratio = 100
33 GROUP BY experience_level) n ON m.experience_level = n.experience_level) x
34 INNER JOIN
35 (SELECT
36     m.experience_level,
37     m.work_year,
38     (remote_total / total) * 100 AS 'percentage'
```

```
163 GROUP BY experience_level) n ON m.experience_level = n.experience_level) x
164 INNER JOIN
165 (SELECT
166     m.experience_level,
167     m.work_year,
168     (remote_total / total) * 100 AS 'percentage'
169 FROM
170     (SELECT
171         experience_level, work_year, COUNT(*) AS 'total'
172 FROM
173         salary
174 WHERE
175         work_year = 2024
176 GROUP BY experience_level) m
177 INNER JOIN (SELECT
178     experience_level, work_year, COUNT(*) AS 'remote_total'
179 FROM
180     salary
181 WHERE
182     work_year = 2024 AND remote_ratio = 100
183 GROUP BY experience_level) n ON m.experience_level = n.experience_level) y ON x.experience_level = y.experience_level;
184
```

Result Grid						
Filter Rows:						
Export:						
Wrap Cell Content: IA						
	experience_level	work_year	percentage	experience_level	work_year	percentage
▶	SE	2021	58.6667	SE	2024	25.1563
	MI	2021	51.7241	MI	2024	20.5989
	EN	2021	47.8261	EN	2024	22.8346
	EX	2021	50.0000	EX	2024	33.0189





AS a compensation specialist at a Fortune 500 company, you're tasked WITH analyzing salary trends over time. Your objective is to calculate the average salary INcrease percentage for each experience level and job title between the years 2023 and 2024, helping the company stay competitive IN the talent market..



```
136 WITH t AS
137 (
138     SELECT experience_level, job_title, work_year, round(AVG(salary_in_usd),2) AS 'average' FROM salaries
139     WHERE work_year IN (2023,2024) GROUP BY experience_level, job_title, work_year
140 ) -- step 1
141
142
143 SELECT *,round((((AVG_salary_2024-AVG_salary_2023)/AVG_salary_2023)*100),2) AS changes
144 FROM
145 (
146     SELECT
147         experience_level, job_title,
148         MAX(CASE WHEN work_year = 2023 THEN average END) AS AVG_salary_2023,
149         MAX(CASE WHEN work_year = 2024 THEN average END) AS AVG_salary_2024
150     FROM t GROUP BY experience_level, job_title -- step 2
151 )a WHERE (((AVG_salary_2024-AVG_salary_2023)/AVG_salary_2023)*100) IS NOT NULL -- STEP 3
152
```

	experience_level	job_title	AVG_salary_2023	AVG_salary_2024	changes
▶	SE	AI Engineer	172245.94	180068.57	4.54
	SE	Machine Learning Engineer	196167.59	206863.44	5.45
	MI	Business Intelligence Developer	84032.00	83385.63	-0.77
	SE	Data Engineer	158309.32	161949.40	2.30
	SE	Data Scientist	173480.98	160234.25	-7.64
	SE	Cloud Database Engineer	141666.67	136437.50	-3.69
	MI	Data Engineer	124952.02	125574.88	0.50
	MI	Machine Learning Engineer	161348.47	162126.76	0.48
	MI	Research Engineer	149161.96	233809.50	56.75
	EN	Data Analyst	76922.90	94439.16	22.77
	SE	Machine Learning Scientist	183504.18	172770.29	-5.85
	SE	Research Engineer	193786.28	206686.04	6.66
	MI	Data Scientist	124881.49	135296.68	8.34
	MI	Applied Scientist	173830.71	181031.25	4.14
	SE	Data Analyst	120853.50	127195.41	5.25
	MI	Data Science Manager	149268.33	206256.25	38.18
	SE	Research Scientist	195332.03	234464.89	20.03
	SE	Prompt Engineer	166500.00	298750.00	79.43

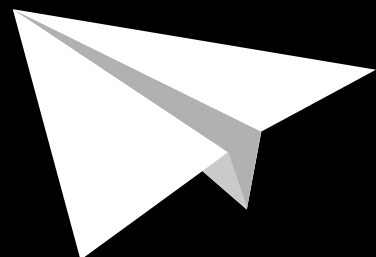
Result 6 x



You're a database administrator tasked with role-based access control for a company's employee database. Your goal is to implement a security measure where employees in different experience level (e.g. Entry Level, Senior level etc.) can only access details relevant to their respective experience_level, ensuring data confidentiality and minimizing the risk of unauthorized access..



```
166 CREATE USER 'Entry_level'@'%' IDENTIFIED BY 'EN';
167 • CREATE USER 'Junior_Mid_level'@'%' IDENTIFIED BY 'MI ';
168 • CREATE USER 'Intermediate_Senior_level'@'%' IDENTIFIED BY 'SE';
169 • CREATE USER 'Expert_Executive-level'@'%' IDENTIFIED BY 'EX ';
170
171
172 CREATE VIEW entry_level AS
173 SELECT * FROM salaries where experience_level='EN'
174
175 GRANT SELECT ON campusx.entry_level TO 'Entry_level'@'%'
176
177 UPDATE view entry_level set WORK_YEAR = 2025 WHERE EMPLOYMENT_TYPE='FT'
178
179
```



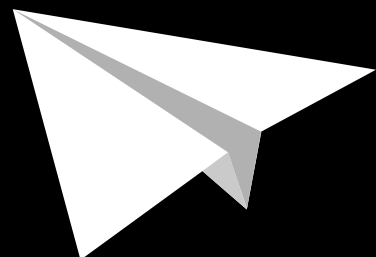


You are working with an consultancy firm, your client comes to you with certain data and preferences such as (their year of experience , their employment type, company location and company size) and want to make an transaction into different domain in data industry(like a person is working as a data analyst and want to move to some other domain such as data science or data engineering etc.)your work is to guide them to which domain they should switch to base on the input they provided, so that they can now update thier knowledge as per the suggestion/.. The Suggestion should be based on average salary..



```
263 DELIMITER //
264 • create PROCEDURE GetAverageSalary(IN exp_lev VARCHAR(2), IN emp_type VARCHAR(3), IN comp_loc VARCHAR(2), IN comp_size VARCHAR(2))
265 BEGIN
266     SELECT job_title, experience_level, company_location, company_size, employment_type, ROUND(AVG(salary), 2) AS avg_salary
267     FROM salary
268     WHERE experience_level = exp_lev AND company_location = comp_loc AND company_size = comp_size AND employment_type = emp_type
269     GROUP BY experience_level, employment_type, company_location, company_size, job_title order by avg_salary desc ;
270 END//
271 DELIMITER ;
272 • -- Deliminator By doing this, you're telling MySQL that statements within the block should be parsed as a single unit until the cus
273
274 call GetAverageSalary('EN','FT','AU','M')
275
```

	job_title	experience_level	company_location	company_size	employment_type	avg_salary
▶	Data Scientist	EN	AU	M	FT	120000.00
	Business Intelligence Analyst	EN	AU	M	FT	91000.00
	AI Programmer	EN	AU	M	FT	40000.00
	Machine Learning Developer	EN	AU	M	FT	40000.00
	Data Analyst	EN	AU	M	FT	36276.50

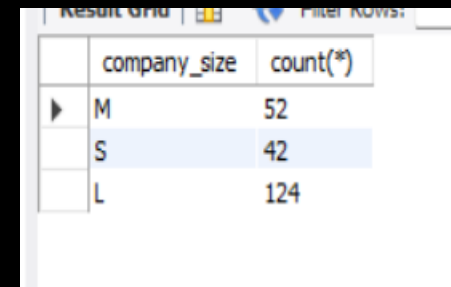




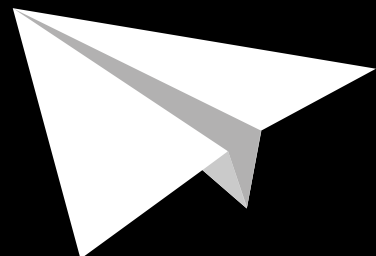
As a market researcher, your job is to Investigate the job market for a company that analyzes workforce data. Your Task is to know how many people were employed IN different types of companies AS per their size IN 2021..



```
6 • select company_size,count(*) from salary where work_year = 2021
7   group by company_size;
8
```



	company_size	count(*)
▶	M	52
	S	42
	L	124

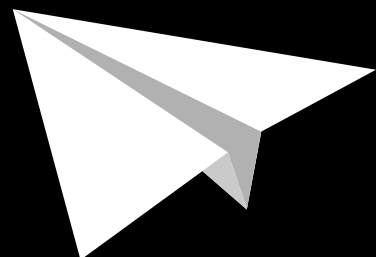




Imagine you are a talent Acquisition specialist Working for an International recruitment agency. Your Task is to identify the top 3 job titles that command the highest average salary Among part-time Positions Ensuring a robust sample size for your analysis

```
12 _
13 • select job_title, avg(salary_in_usd) from salary
14 where employment_type = "PT"
15 group by job_title
16 order by avg(salary_in_usd) desc limit 3;
17
```

Result Grid		Filter Rows:
	job_title	avg(salary_in_usd)
▶	Data Scientist	87767.0000
	Data Engineer	65168.5000
	Data Analyst	56916.6667



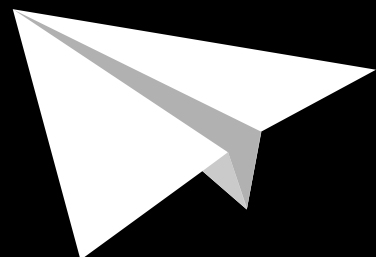


As a database analyst you have been assigned the task to Select Countries where average mid-level salary is higher than overall mid-level salary for the year 2023.



```
37 -- Set the delimiter for the stored procedure
38 DELIMITER //
39
40 • CREATE PROCEDURE GettSeniorSalaryStats()
41 BEGIN
42     -- Query to find the highest average salary for senior-level employees in 2023
43     SELECT company_location AS highest_location, AVG(salary_in_usd) AS highest_avg_salary
44     FROM salary
45     WHERE work_year = 2023 AND experience_level = 'SE'
46     GROUP BY company_location
47     ORDER BY highest_avg_salary DESC
48     LIMIT 1;
49
50     -- Query to find the lowest average salary for senior-level employees in 2023
51     SELECT company_location AS lowest_location, AVG(salary_in_usd) AS lowest_avg_salary
52     FROM salary
53     WHERE work_year = 2023 AND experience_level = 'SE'
54     GROUP BY company_location
55     ORDER BY lowest_avg_salary ASC
56     LIMIT 1;
57 END //
58
59 -- Reset the delimiter back to semicolon
60 DELIMITER ;
61
62 • -- Call the stored procedure to get the results
63 CALL GettSeniorSalaryStats();
```

	lowest_location	lowest_avg_salary
▶	TR	18381.0000

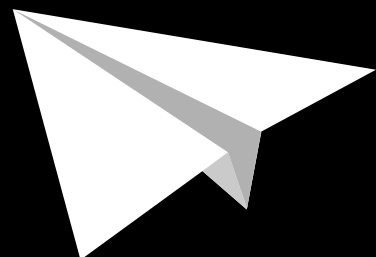




You've been hired by a global HR Consultancy to identify Countries experiencing significant salary growth for entry-level roles. Your task is to list the top three Countries with the highest salary growth rate FROM 2020 to 2023, helping multinational Corporations identify Emerging talent markets.

```
104 • select * from(  
105   select company_location, avg(salary_in_usd) as 'avg_2021' from salary  
106   where work_year=2021 and experience_level='EN'  
107   group by company_location) a  
108   inner join(  
109     select company_location, avg(salary_in_usd) as 'avg_2022' from salary  
110     where work_year=2022 and experience_level='EN'  
111     group by company_location) b  
112   on a.company_location = b.company_location  
113   inner join(  
114     select company_location, avg(salary_in_usd) as 'avg_2023' from salary  
115     where work_year=2023 and experience_level='EN'  
116     group by company_location) c  
117   on b.company_location = c.company_location
```

	company_location	avg_2021	company_location	avg_2022	company_location	avg_2023
▶	DE	65388.8333	DE	50046.1667	DE	69220.0000
	US	88617.6471	US	102347.9077	US	101592.8575
	GB	79318.3333	GB	43761.8889	GB	61925.0465
	AU	42028.0000	AU	78292.7500	AU	53089.3333
	IN	24407.1667	IN	20348.5000	IN	27344.1667
	FR	41347.7500	FR	59431.0000	FR	32390.0000



Thank You!



CREDITS

yogeshchouhan263@gmail.com