Group By and Having SQL Questions

Hacker Rank

1. **Weather Observation Station 20**
    - Link: https://www.hackerrank.com/challenges/weather-observation-station-20
2. **The Report**
    - Link: https://www.hackerrank.com/challenges/the-report
3. **Top Earners**
    - Link: https://www.hackerrank.com/challenges/earnings-of-employees
4. **Challenges**
    - Link: https://www.hackerrank.com/challenges/challenges
5. **SQL Project Planning**
    - Link: https://www.hackerrank.com/challenges/sql-projects

---

Leet Code

1. **Big Countries**
    - Link: https://leetcode.com/problems/big-countries/
2. **Second Highest Salary**
    - Link: https://leetcode.com/problems/second-highest-salary/
3. **Customers Who Bought All Products**
    - Link: https://leetcode.com/problems/customers-who-bought-all-products/
4. **Number of Transactions per Customer**
    - Link: https://leetcode.com/problems/number-of-transactions-per-customer/
5. **Average Time of Process per Machine**
    - Link: https://leetcode.com/problems/average-time-of-process-per-machine/
6. **SQL 50 Collection**
    - Link: https://leetcode.com/study-plan/sql/?progress=xo2ojlkm

- Description: LeetCode's SQL 50 study plan includes several problems requiring GROUP BY and HAVING. Notable ones:
  - **Problem 1693**: Daily Leads and Partners
    - Uses GROUP BY to count leads and HAVING to filter groups.
  - **Problem 1729**: Find Followers Count
    - Involves grouping by user and filtering with HAVING.
  - **Problem 1873**: Calculate Special Bonus
    - May use GROUP BY and HAVING in advanced solutions.

Chat GPT

1. Sample Table: Employee

emp_id	name	department	salary	age	city	hire_date	manager_id
1	John	IT	75000	28	New York	2020-01-15	5
2	Sarah	HR	65000	32	Chicago	2019-03-20	6
3	Mike	IT	80000	35	New York	2018-07-10	5
4	Lisa	Finance	70000	29	Boston	2021-02-28	7
5	David	IT	95000	40	New York	2017-05-12	NULL
6	Emma	HR	85000	38	Chicago	2016-11-08	NULL
7	Tom	Finance	90000	45	Boston	2015-09-14	NULL
8	Anna	Marketing	60000	26	Miami	2022-01-05	9
9	Chris	Marketing	72000	33	Miami	2020-06-18	NULL
10	Kate	IT	68000	27	Seattle	2021-08-22	5
11	James	Finance	66000	31	Boston	2020-12-01	7
12	Rachel	HR	58000	24	Chicago	2023-03-15	6
13	Mark	IT	77000	30	Seattle	2019-10-30	5
14	Sophie	Marketing	63000	28	Miami	2021-11-12	9
15	Alex	Finance	71000	34	Boston	2018-04-25	7

GROUP BY Questions (Basic Level)

1. Find the number of employees in each department.
2. Calculate the average salary for each department.
3. Find the total salary expenditure for each city.
4. Count how many employees are there in each city.

5. Find the maximum salary in each department.
6. Find the minimum age in each department.
7. Calculate the sum of salaries for each manager_id (including NULL).
8. Find the average age of employees in each city.
9. Count the number of employees hired each year.
10. Find the highest salary in each city.

GROUP BY Questions (Intermediate Level)

11. Find the department-wise count of employees along with their average salary.
12. Calculate the total and average salary for each department, ordered by average salary descending.
13. Find the city-wise distribution of employees by department.
14. Calculate the age range (max - min) for each department.
15. Find the number of employees and total salary for each manager.
16. Get the count of employees in each department for each city.
17. Find the earliest and latest hire dates for each department.
18. Calculate the salary variance for each department (max - min salary).
19. Find departments where the average employee age is above 30.
20. Get the median salary range for each city (you can use max and min as approximation).

HAVING Clause Questions (Basic Level)

21. Find departments that have more than 3 employees.
22. Find cities where the average salary is greater than 70000.
23. Find departments where the total salary expenditure exceeds 200000.
24. Find cities that have at least 2 employees.
25. Find departments where the maximum salary is greater than 80000.
26. Find managers who manage more than 2 employees.
27. Find departments where the minimum salary is less than 65000.

28. Find cities where the total number of employees is exactly 3.
29. Find departments where the average age is less than 32.
30. Find hire years that have more than 2 employees hired.

HAVING Clause Questions (Intermediate Level)

31. Find departments where the salary range (max - min) is greater than 15000.
32. Find cities where both the employee count is greater than 2 AND average salary is above 65000.
33. Find departments that have employees in more than 2 different cities.
34. Find managers who manage employees with a total salary sum exceeding 150000.
35. Find departments where the average salary is between 65000 and 80000.
36. Find cities where the youngest employee is older than 25.
37. Find departments with at least 3 employees AND maximum salary above 75000.
38. Find hire years where the average salary of hired employees exceeds 70000.
39. Find cities where the oldest employee is younger than 40.
40. Find departments where the standard deviation of salaries is high (use max-min > 20000 as approximation).

Complex GROUP BY + HAVING Questions

41. Find departments that have more than 2 employees in the same city.
42. Find cities where IT department has more than 1 employee.
43. Find departments where the average salary is above the overall company average.
44. Find managers who manage employees from more than 1 department.
45. Find departments where all employees earn more than 60000.
46. Find cities that have employees from at least 3 different departments.
47. Find departments where the newest employee was hired after 2020 AND the department has more than 2 employees.

48. Find cities where the HR department average salary exceeds 70000.
49. Find departments that have both junior (age < 30) and senior (age > 35) employees.
50. Find managers whose managed employees have an average age difference of more than 5 years from the manager's department average.

Advanced Challenge Questions

51. Find department-city combinations that have more than 1 employee.
52. Find the top 2 departments by average salary that have at least 3 employees.
53. Find cities where the salary gap between highest and lowest paid employee exceeds 15000.
54. Find departments where more than 50% of employees earn above 70000.
55. Find years where more employees were hired than the previous year (complex logic).
56. Find departments that have employees in all major cities (New York, Chicago, Boston).
57. Find managers who manage the highest number of employees (tie-breaking allowed).
58. Find department-city pairs where the local average salary exceeds the department's overall average.
59. Find cities that have the most diverse age range across all departments.
60. Find the most recently hired employee in each department for departments with more than 2 employees.

Notes for Practice:

- Start with basic GROUP BY questions (1-10)
- Move to HAVING clause questions (21-30)
- Combine both concepts (31-50)
- Challenge yourself with advanced queries (51-60)
- Pay attention to NULL handling in manager_id

- Consider using subqueries for complex conditions
- Practice with different aggregate functions (COUNT, SUM, AVG, MAX, MIN)

Case Study

customers

customer_id	name	email	registration_date	city	country	age_grou
C001	Alice Johnson	alice@email.com	2022-01-15	New York	USA	25-34
C002	Bob Smith	bob@email.com	2022-03-20	London	UK	35-44
C003	Carol Davis	carol@email.com	2021-11-10	Toronto	Canada	25-34
C004	David Wilson	david@email.com	2023-02-14	Sydney	Australia	45-54
C005	Emma Brown	emma@email.com	2022-08-05	Paris	France	18-24
C006	Frank Miller	frank@email.com	2021-12-30	Berlin	Germany	35-44

order_items

order_item_id	order_id	product_id	quantity	unit_price
OI001	O001	P003	2	120.00
OI002	O001	P004	1	85.00
OI003	O002	P005	1	150.00
OI004	O003	P001	1	999.00
OI005	O005	P002	1	1200.00
OI006	O007	P006	1	450.00
OI007	O009	P003	1	120.00
OI008	O009	P005	1	150.00

products

product_id	product_name	category	price	supplier_id	launch_date
P001	iPhone 14	Electronics	999.00	S001	2022-09-01
P002	Samsung TV	Electronics	1200.00	S002	2022-08-15
P003	Nike Shoes	Fashion	120.00	S003	2023-01-10
P004	Adidas Jacket	Fashion	85.00	S003	2022-11-20
P005	Coffee Maker	Home	150.00	S004	2022-10-05
P006	Dining Table	Home	450.00	S005	2023-02-12

orders

order_id	customer_id	order_date	total_amount	status	payment_method	shipping_
O001	C001	2023-01-10	250.00	delivered	credit_card	USA
O002	C001	2023-02-15	180.50	delivered	paypal	USA
O003	C002	2023-01-20	450.00	delivered	credit_card	UK
O004	C003	2023-03-05	320.75	cancelled	debit_card	Canada
O005	C002	2023-02-28	680.00	delivered	credit_card	UK
O006	C004	2023-04-12	150.25	pending	paypal	Australia
O007	C001	2023-03-18	520.00	delivered	credit_card	USA
O008	C005	2023-01-25	95.00	delivered	debit_card	France
O009	C003	2023-04-20	275.50	delivered	credit_card	Canada
O010	C006	2023-02-10	380.00	returned	paypal	Germany

employees

emp_id	name	department	salary	hire_date	performance_rating	manager_id
E001	John Doe	Sales	65000	2021-03-15	Excellent	E005
E002	Jane Smith	Marketing	58000	2020-07-22	Good	E006
E003	Mike Johnson	Sales	62000	2022-01-10	Average	E005
E004	Sarah Wilson	IT	75000	2019-11-30	Excellent	E007
E005	Tom Brown	Sales	85000	2018-05-12	Excellent	NULL
E006	Lisa Davis	Marketing	72000	2019-09-18	Good	NULL
E007	Chris Lee	IT	95000	2017-12-01	Excellent	NULL
E008	Amy Taylor	IT	68000	2021-08-25	Good	E007
E009	Kevin White	Marketing	55000	2022-04-14	Average	E006

Case Study Interview Questions

Case 1: E-commerce Business Intelligence

Scenario: You're a data analyst at an e-commerce company. The CEO wants insights for the quarterly board meeting.

1. **Customer Segmentation Analysis**: "Identify our most valuable customer segments. Show countries where customers have an average order value above $300 and have placed at least 2 orders."
2. **Payment Method Performance**: "The finance team is reviewing payment processing fees. Find payment methods that generate total revenue above $500 and show their usage frequency."
3. **Geographic Expansion Strategy**: "We're planning international expansion. Identify countries with total order value exceeding $400 and average order value above $200."

Case 2: Product Portfolio Management

Scenario: You're supporting the product management team for strategic decisions.

4. **Category Performance Review**: "The product team wants to discontinue underperforming categories. Find product categories where total sales are below the average category performance."
5. **Supplier Relationship Analysis**: "We're renegotiating supplier contracts. Identify suppliers whose products generate total revenue above $1000 and have been ordered more than twice."
6. **Launch Success Metrics**: "Evaluate product launch success. Find products launched in 2023 that have generated sales and show their performance metrics."

Case 3: Customer Retention & Churn Analysis

Scenario: The marketing team is designing retention campaigns.

7. **High-Value Customer Identification**: "Create a premium customer list. Find customers with total purchase amount above $400 who have never cancelled an order."
8. **Age Group Purchasing Behavior**: "Design age-targeted campaigns. Show age groups with total spending above $300 and average order frequency above 1.5 orders per customer."
9. **Country-wise Customer Loyalty**: "Identify markets with loyal customers. Find countries where customers have average order values above $250 and low cancellation rates."

Case 4: Sales Team Performance Management

Scenario: HR wants to evaluate sales team performance for bonuses and promotions.

10. **Department Efficiency Analysis**: "The board is reviewing department budgets. Find departments where average salary exceeds $65,000 and they have at least 2 employees."
11. **Performance-Based Compensation Review**: "Design performance bonuses. Identify departments with all employees having 'Excellent' or 'Good' ratings and average salary below $70,000."
12. **Team Size Optimization**: "HR is evaluating team structures. Find departments that have more than 2 employees and show their salary distribution."

Case 5: Operational Excellence

Scenario: Operations team needs insights for process improvements.

13. **Order Status Analysis**: "Improve order fulfillment. Find order statuses that occur more than once and have average order values above $200."

14. **Monthly Revenue Trends**: "The CFO needs monthly performance data. Show months with total revenue above $500 and more than 2 orders processed."
15. **Shipping Performance by Country**: "Optimize logistics. Find shipping countries with total order volume above $300 and average delivery success rate (non-cancelled orders)."

Case 6: Financial Planning & Analysis

Scenario: Finance team is preparing annual budgets and forecasts.

16. **Revenue Concentration Risk**: "Assess customer concentration risk. Find customers who contribute more than 20% of total revenue (assuming $2000+ total spending indicates high concentration)."
17. **Department ROI Analysis**: "Evaluate department ROI for budget allocation. Find departments with total salary costs above $120,000 and calculate their cost per employee."
18. **Payment Method Risk Assessment**: "Evaluate payment processing risks. Show payment methods with total transaction value above $400 and identify patterns."

Case 7: Market Research & Competitive Analysis

Scenario: Strategy team is analyzing market positioning.

19. **Product Category Dominance**: "Identify our strongest product categories. Find categories with more than 1 product sold and total revenue above average."
20. **Customer Registration Trends**: "Analyze customer acquisition patterns. Show registration years with more than 1 customer and their subsequent purchasing behavior."
21. **Geographic Market Penetration**: "Evaluate market penetration. Find countries with customers who have total spending above $200 and multiple orders."

Case 8: Data Quality & Business Rules

Scenario: Data governance team is establishing business rules.

22. **Order Validation Rules**: "Establish order validation rules. Find orders where customers have placed multiple orders totaling above $300."
23. **Customer Lifecycle Analysis**: "Define customer lifecycle stages. Group customers by registration year and show only years with customers having average spending above $150."
24. **Product Performance Thresholds**: "Set product performance benchmarks. Find products that appear in more than 1 order and generate significant revenue."

Case 9: Advanced Business Intelligence

Scenario: Senior management needs complex analytical insights.

25. **Cross-Category Customer Behavior**: "Analyze customer purchase diversity. Find customers who have purchased from multiple product categories (require joining multiple tables)."
26. **Seasonal Performance Analysis**: "Identify seasonal trends. Show quarters with total orders above $800 and consistent customer activity."
27. **Employee Performance vs Department Budget**: "Correlate individual and team performance. Find departments where high-performing employees (Excellent rating) work and department average salary exceeds $65,000."

Case 10: Strategic Decision Support

Scenario: C-suite executives need data for strategic decisions.

28. **Market Exit Analysis**: "Evaluate market exit decisions. Find countries with total customer spending below $200 or average order value below $150."
29. **Product Line Expansion**: "Support product line decisions. Identify successful categories (total sales > $400) that should receive additional investment."
30. **Customer Acquisition Cost Analysis**: "Optimize marketing spend. Find registration periods with customers whose lifetime value (total orders) exceeds $350."

Case 11: Risk Management

Scenario: Risk management team needs fraud and business continuity insights.

31. **Fraudulent Activity Detection**: "Identify potential fraud patterns. Find customers with unusually high order frequency (more than 3 orders) and large order values (average > $400)."
32. **Supplier Dependency Risk**: "Assess supplier risks. Find suppliers providing products worth more than $800 in total sales to evaluate dependency."
33. **Payment Failure Analysis**: "Reduce payment failures. Analyze order statuses by payment method where certain methods show higher failure rates."

Case 12: Customer Experience Optimization

Scenario: Customer experience team wants to improve satisfaction.

34. **Return/Cancellation Analysis**: "Minimize returns and cancellations. Find customer segments (by country/age group) with return rates and spending patterns."

35. **Premium Service Qualification**: "Design VIP customer service. Identify customers qualifying for premium support (high spending, multiple orders, good payment history)."
