Problem Statement

You have been hired by Online University to mine their extensive data and extract useful information. Following are some questions asked by the leadership team. Using MySQL queries, find solutions to the following queries:

1. Which region has the highest median score in the final exam?
2. Which course has the highest pass rate? What is the percentage of each activity_type within this course?
3. Which activity type is most exercised amongst the students who got the distinction?
4. During what day (Number of days from course-batch start) for each course-batch do we see maximum activity (clicks) on VLE?
5. What is the extent of activity (number of clicks) for course batch AAA 2013J above which 95% of students pass the course?
6. Compare the average result of batches 2013J and 2014J only for females who engaged with Virtual Learning Environment in 28th day of the course-batches for 1st and 2nd Tutor Marked Assignment (TMA) using combined weighted score.
7. Print the id of students who achieved full scores (100) for more than one assessment. Order your output in descending order by the total number of assessments in which the student earned a full score. If more than one student received full scores in the same number of assessments, then sort them by decreasing average score (weighted) across all assessments
8. (A) Print total number of unique students who made at least 20 clicks each week for course-batch AAA 2013J (starting from the first day of the course), and (B) find the student id of the student who made the maximum number of clicks each week. In case more than one such student has a maximum number of clicks, print the lowest student id.
9. Print top 5 id_sites for material from VLE in each course-batch for which the total number of clicks has the highest correlation (Pearson's correlation) with pass rate in decreasing order of correlation within each course-batch.
10. What are the primary factors you think are leading students to withdraw or fail? Which of these can be checked from the datasets provided? Illustrate using query results. (For only this question, participants are allowed to use other open source tools as well but only for visualizations)