

DSC 204A: Scalable Data Systems

Programming Assignment 1: Grading Scheme

For evaluation we will run your code on a hidden test dataset.

Accuracy (80)

Your solutions rounded to 2 decimal points must match exactly as our solutions, otherwise 0 will be rewarded. We will tolerate 1% error. There is no partial marking.

| Task No. | Task Description | Points |
|----------|---|--------|
| 1 | % Missing values for all columns in <i>reviews</i> table | 6 |
| 2 | % Missing values for all columns in <i>products</i> table | 6 |
| 3 | Pearson correlation between rating and price | 14 |
| 4 | Descriptive stats for price | 12 |
| 5 | Number of products for each super-category | 14 |
| 6 | Check for dangling reference of product ids from <i>reviews</i> table to <i>products</i> table | 14 |
| 7 | Check for dangling reference of product ids in the <i>related</i> column to <i>asin</i> column of <i>products</i> table | 14 |

Runtime (20)

We will run your functions thrice and take the average for getting the runtime measurement. If your accuracy points are greater than 40, the runtime evaluation will be automated and the grading is based on the following tables. If the accuracy points fall below 40 then you will get partial credit based on manual inspection by TAs. The partial credit percentage will be multiplied with your actual runtime score (based on the following tables). For instance, based on your effort if we assigned partial credit percentage to be 70% then the runtime score you get will be multiplied by 0.7.

| Absolute 4 node runtimes (10) | Points |
|-------------------------------|--------|
| Under 5 mins | 10 |
| Between 5 mins to 10 mins | 6 |
| Between 10 mins to 30 mins | 4 |
| Anything above 30 mins | 0 |

| Runtime speedup with 4 nodes (10) | Points |
|-----------------------------------|--------|
| Greater than 3.5x | 10 |
| Between 2.5x to 3.5x | 6 |
| Between 1x to 2.5x | 4 |
| Under 1x | 0 |

Extra Credit (10)

If you received full accuracy points and your absolute runtimes (mean of 3 runs) is less than or equal to our solution's runtime, you will be awarded with 10 extra points.