



Transport feedback scoring

Revision: 1.4

Problem statement

There are a multitude of things that influences the daily experience of commuters on public transport - congestion and disruptions being the major ones. We believe that there are other less visible influences that have significant impacts too.

WhereIsMyTransport strives to “make the invisible visible” through data, and we have collected some *anonymous route sentiment scores*. Your task is to summarise these scores and see if they can help us in identifying other influences.

Input / Output

The input is simple text files located in the same directory as the program, and named as follows:

- `reference-data.txt`
- `scores.txt`

The program should generate a single text output file `results.txt` - see the "Samples" section for more details.



Note

You should assume that the input will be well-formed - there is **no need** to add special handling for malformed input files.

Scoring rules

We need to calculate average scores for each route, per day of the week. In other words, all the eligible scores for a Monday for a particular route needs to be added together and the average sentiment score must be calculated. This value must be *rounded to 2 decimal places*.

We have a hypothesis that values at the extremes of the scoring range are obscuring the real sentiment from commuters, therefore entries with an associated score of either 0 or 10 should be omitted from *all* calculations.

Samples



These samples are for **illustration purposes only** - running your program on these inputs will not generate the same results as the sample output below.

Complete sample data sets can be downloaded from [this url](#).

Reference data

The reference data file (`reference-data.txt`) contains information about routes and the transport agencies who operate them. Each entry contains a route identifier, an agency identifier and the name of the route as it is commonly known.

More precisely:

- route identifiers always start with the prefix `route_` , and contain no spaces
- agency identifiers have no common prefix, and consists of only numbers (0-9), letters from the (Latin) Alphabet (A-Z), "_" or "-"

- route names however, may also contain whitespace and accented letters such as é, á and so on

```
route_f97ad4ff-0f3c-4ecb-88bc-ff2e35135bb2 METRO;Ciudad Azteca - Buenavista
route_cc4722b0-db3b-4732-9e5f-456325d59ffe METRO;Cuatro Caminos - Tasqueña
route_29eac9a5-0133-4f1e-b69e-6eb8046c0c71 METRO;El Rosario - Barranca del Muerto
```

Sentiment scores

The input file named `scores.txt` contains one or more feedback score entries, anonymously collected from commuters. Each entry contains a sentiment score, for a particular date and route.

More precisely:

- date values are formatted as *ccyy/mm/dd*, e.g. 2021/11/23
- route identifiers always start with the prefix `route_`, and contain no spaces
- sentiment scores are integer values in the range 0 to 10 (inclusive).

```
2021/11/19;route_44f46f7a-0da7-4bbc-ae28-747d92643385 8
2021/11/20;route_cc4722b0-db3b-4732-9e5f-456325d59ffe 4
2021/11/12;route_03abae1a-7662-4a67-92a5-6a9649749fee 8
2021/11/08;route_68a711e1-440d-426c-a6e9-a119d98502d2 5
```

Results

The calculated scores for each route should be output, from highest to lowest, along with the day of the week.

```
Pantitlán - Tacubaya Monday 7.89
Pantitlán - Tacubaya Sunday 5.97
Pantitlán - Tacubaya Tuesday 5.79
Pantitlán - Tacubaya Wednesday 2.97
Pantitlán - Tacubaya Thursday 2.0
```

Pantitlán - Tacubaya Friday 1.92
Pantitlán - Tacubaya Saturday 1.0
Cuatro Caminos - Tasqueña Monday 7.98
Cuatro Caminos - Tasqueña Sunday 5.97
Cuatro Caminos - Tasqueña Tuesday 5.94
Cuatro Caminos - Tasqueña Wednesday 2.92
Cuatro Caminos - Tasqueña Thursday 2.0
Cuatro Caminos - Tasqueña Friday 1.95
Cuatro Caminos - Tasqueña Saturday 1.0