



Bloom Credit Engineering Challenge

Bloom Credit provides credit data and insights via a modern, performant, and standards-compliant API. Much of the raw credit data comes from legacy sources and use outdated proprietary formats. The goal of this data engineering challenge is to process a batch of (simulated) data in a textual fixed-width format, store it in a relational database, and make it discoverable via a REST API.

We have placed a data file in an s3 bucket¹, the bucket contains simulated credit records in a fixed-width format with one consumer record per line.

The first line of the file is the header. The structure of the credit records is as follows:

- consumer name (string, width = 72)
- social security number (integer, width = 9)
- credit tags X0001 through X0200 (zero-padded integers, each of width = 9)
- Non-negative values of credit tags represent regular data
- While negative values indicate error conditions, e.g., no data available.

Using a programming language, database, and web framework of your choice, please write the code to perform the following tasks:

1. Define a data schema for the credit data in a relational database using SQL or an ORM of your choice.
2. Parse the `test.dat` file based on the above description and bulk insert the data into the database. Use generated UUIDs as consumer identifiers.
3. Create a REST API endpoint for retrieving the full set of credit tags for a given consumer by consumer id provided as a query-string parameter. The endpoint should return data in JSON format.
4. If time permits, create a second REST API endpoint for retrieving consumer statistics for a given credit tag. The endpoint should retrieve the mean, median, and standard deviation for a credit tag provided as a query-string parameter. The statistics should only include regular (non-negative) values of each credit tag. The endpoint should return data in JSON format.

Please send the final results back to us or upload to a public Github repo.

Good luck!

¹ <https://bloom-data-engineering-challenge.s3.amazonaws.com/test.dat.bz2>