Some notes about the execution:

- GCP products such as Bigquery, Dataflow, and Cloud Storage are used.
- The task is done by two methods.
  - First, using A Python script that gets the URLs and exports a combined CSV file using pandas.
  - Second, a python script that first makes a Dataflow (apache beam) pipeline for unifying and cleaning the loan CSV files and then passes it to the Bigquery engine for performing joins and exporting the final combined CSV table (can be run as a scheduled query). This method might not be the best way to do the task, as the Beam pipeline doesn't entirely do all the jobs, but I personally wanted to struggle with Beam on that.

- I aimed to make the final table informative from BI perspective. I thought BI analysts might probably want to see which campaigns or referrers have led to which loans (Outcome, Amount, …) or which customers (Age, Gender, City, …). So, as one customer might be landed to the website by different referrers or campaigns, I made a field named "campaign_referer" by concatenating the campaign and the referrer, each time a customer visits the website and then aggregating all the result for a single customer (e.g. "display1Google - display1Twitter - display3Twitter - display1Facebook"). In this way, all campaigns and referrers are searchable within this field for every customer. Otherwise, we would have multiple lines for each time a customer has visited the website with a simple join.
- Since loans information is questioned, the base table for joins is the loans table.