

README

run_analysis.R

P. A. Ortiz Otalvaro

21 August 2019

This is a script that cleans a data set on loans from Home Credit

Data set

Home Credit is an international non-bank consumer finance group founded in the Czech Republic. and as explained in Home Credit's website:

"Founded in 1997, Home Credit Group is an international consumer finance provider with operations in 9 countries. We focus on responsible lending primarily to people with little or no credit history. Our services are simple, easy and fast."

Source of data: <https://www.kaggle.com/c/home-credit-default-risk/overview>

Download data: <https://www.kaggle.com/c/home-credit-default-risk/data>

How to clean this data set using wrangling_homecredit.R

Note: You should have installed both dplyr and data.table in RStudio before following the next steps.

1. Download the data. **It should be inside a folder called *UCI HAR Dataset***
- 2. Download run_analysis.R and locate it in the parent folder of *UCI HAR Dataset***
- 3. Open run_analysis.R in RStudio. Set your working directory to the folder where run_analysis.R is located (i.e. the parent folder of *UCI HAR Dataset*).**
- 4. Run the script in RStudio**

Run `source("run_analysis.R")`

A new file called *har_clean.csv* should appear in your current working directory with the resulting data set after the wrangling process.