**Which Debts Are Worth the Bank's Effort?**

Play bank data scientist and use regression discontinuity to see which debts are worth collecting.

#### Project Description

After a debt has been legally declared "uncollectable" by a bank, the account is considered to be "charged-off." But that doesn't mean the bank simply **walks away** from the debt. They still want to collect some of the money they are owed. In this project, you will look at a situation where a bank assigned delinquent customers to different recovery strategies based on the expected amount the bank believed it would recover from the customer. The goal for the data scientist is to determine in this non-random assignment whether the incremental amount the bank earns exceeded the additional cost of assigning customers to a higher recovery strategy.

Threshold assignments like this also one occur in medicine (above a certain temperature you get medicine), education (above a certain test score students get admitted to a special class), other areas of finance (above a certain wealth customers get different levels of service), and public sector (below a certain income someone is eligible for housing benefits). Regression discontinuity is an intuitive and useful analysis method in any situation of a threshold assignment.

This project lets you apply the skills from [pandas Foundations](https://www.datacamp.com/courses/pandas-foundations) and [Manipulating DataFrames with pandas](https://www.datacamp.com/courses/manipulating-dataframes-with-pandas), including reading, exploring, filtering, and grouping data. This project also uses basic statistics, where an intro to statistics course like [Statistical Thinking in Python](https://www.datacamp.com/courses/statistical-thinking-in-python-part-1) is useful.

#### Project Tasks

* 1 Regression discontinuity: banking recovery
* 2 Graphical exploratory data analysis
* 3 Statistical test: age vs. expected recovery amount
* 4 Statistical test: sex vs. expected recovery amount
* 5 Exploratory graphical analysis: recovery amount
* 6 Statistical analysis: recovery amount
* 7 Regression modeling: no threshold
* 8 Regression modeling: adding true threshold
* 9 Regression modeling: adjusting the window