

## 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, we 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 be able to recover from the customer.

Now assigning costumers to higher recovery strategies require some operational costs on the bank side. Therefore, the goal for us is to determine in this non-random assignment whether the incremental amount the bank earns(recovered amount) exceeded the additional cost of assigning customers to a higher recovery strategy.

The bank has implemented different recovery strategies at different thresholds (\$1000, \$2000, \$3000 and \$5000) where the greater the Expected Recovery Amount, the more effort the bank puts into contacting the customer.

- Level 0: Expected recovery amounts  $> \$0$  and  $\leq \$1000$
- Level 1: Expected recovery amounts  $> \$1000$  and  $\leq \$2000$

The threshold of \$1000 separates Level 0 from Level 1

We found that there is no major jump in the average customer age just above and just below the \$1000 threshold by doing a **Kruskal-Wallis test**(statistical test) along with analyzing it with **Exploratory Data Analysis**. We did it to make sure that the expected recovery amount does not vary significantly with age just below and above the threshold level.

Then, the next step is examining the relationship between the actual recovery amount and the expected recovery amount, to develop a better visual.

We have used the Scatter plot and **Regression discontinuity model** in this project because it is an intuitive and useful analysis method in any situation of a threshold assignment. Here, the regression coefficient for the true threshold measures the size of the discontinuity for customers just above and just below the threshold.

If the higher recovery strategy helped recover more money, then the regression coefficient of the true threshold will be greater than zero. If the higher recovery

strategy did not help recover more money, then the regression coefficient will not be statistically significant.

In this project we found that the regression coefficient for the true threshold was statistically significant with an estimated impact of around \$278. This is much larger than the \$50 per customer needed to run this higher recovery strategy.

the incremental recovery amount at the higher recovery strategy is much greater than the \$50 per customer it costs for the higher recovery strategy. So we conclude that the higher recovery strategy is worth the extra cost of \$50 per customer.