

Case Study #1

Below is a data set that represents thousands of loans made through the Lending Club platform, which is a platform that allows individuals to lend to other individuals.

We would like you to perform the following using the language of your choice:

- Describe the dataset and any issues with it.
- Generate a minimum of 5 unique visualizations using the data and write a brief description of your observations. Additionally, all attempts should be made to make the visualizations visually appealing
- Create a feature set and create a model which predicts *interest_rate* using at least 2 algorithms. Describe any data cleansing that must be performed and analysis when examining the data.
- Visualize the test results and propose enhancements to the model, what would you do if you had more time. Also describe assumptions you made and your approach.

Dataset

https://www.openintro.org/data/index.php?data=loans_full_schema

Output

An HTML website hosting all visualizations and documenting all visualizations and descriptions. All code hosted on GitHub for viewing. Please provide URL's to both the output and the GitHub repo.

* If you submit a jupyter notebook, also submit the accompanying python file. You may use python(.py), R, and RMD(knit to HTML) files. Other languages are acceptable as well.

Case Study #2

There is 1 dataset(csv) with 3 years worth of customer orders. There are 4 columns in the csv dataset: index, CUSTOMER_EMAIL(unique identifier as hash), Net_Revenue, and Year.

For each year we need the following information:

- Total revenue for the current year
- New Customer Revenue **e.g. new customers not present in previous year only**
- Existing Customer Growth. To calculate this, use the Revenue of existing customers for current year –(minus) Revenue of existing customers from the previous year
- Revenue lost from attrition
- Existing Customer Revenue Current Year
- Existing Customer Revenue Prior Year
- Total Customers Current Year
- Total Customers Previous Year
- New Customers
- Lost Customers

Additionally, generate a few unique plots highlighting some information from the dataset. Are there any interesting observations?

Dataset

https://www.dropbox.com/sh/xhy2fzjdvg3ykhy/AADAVKH9tgD_dWh6TZtOd34ia?dl=0

customer_orders.csv

Output

An HTML website with the results of the data. Please highlight which year the calculations are for. All code should be hosted on GitHub for viewing. Please provide URL's to both the output and the GitHub repo.

* If you submit a jupyter notebook, also submit the accompanying python file. You may use python(.py), R, and RMD(knit to HTML) files. Other languages are acceptable as well.