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.

Case Study 2 NOTES:

Use Python and Pandas dataframe to join data to dataframe and then calculate and arrange the data as per the business intelligence objectives.

Export results as HTML Report

Print output to both the console and the output HTML file

Total Revenue per year: Sum of revenues that is grouped by year

New Customer Revenue: Sum of revenues from customers who purchased this year but not last year

Existing Customer Growth: Existing customer revenue from year 1 minus total revenue from year 0 (because all customers of year 0 who purchased again in year 1 would be an existing customer and all who did not purchase again represent a negative impact on existing customer growth)

Revenue lost from attrition: Sum of revenues from previous year of lost customers (or sum of revenues from customers last year who did not order this year)

Existing Customer Revenue Current/Previous Year: Sum of revenues of current year orders from customers who had purchased last year

Total Customers Current/Previous Year: Count of customer records that is grouped by year

New Customers: All customers from year 1 who are not in year 0

Lost Customers: All customers in year 0, who are not in year 1