**Case Study #1**

Below is a dataset containing synthetic transactions and some transactions are marked as fraudulent. 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 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 perform prediction of fraudulent transactions using at least 2 algorithms. Describe any data cleansing that must be performed.
* Visualize the test results and propose what could be done to improve results. Also describe assumptions you made and your approach.

**Dataset**

<https://www.kaggle.com/ntnu-testimon/paysim1>

**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.

**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, Revenue of existing customers for current year – Revenue of existing customers from existing 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

**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.