**Option C: Fraud Detection**

**Requirements:**

Your task is to analyze credit card information to predict whether or not a transaction was likely fraudulent. It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase. Use [this dataset](https://drive.google.com/file/d/1WMa_vYptM1H6MODfo8T9unc59ASRnsmT/view?usp=sharing) to predict whether or not the transaction was fraudulent.

Your task for this dataset is to:

* Clean up the dataset.
* Create at least 2 different types of visualizations to better understand the dataset.
* Deal with any categorical variables to prepare for machine learning.
* Determine which features to include in your model or how to best prepare this model for machine learning.
* Build a machine learning model(s) to predict fraud (`Class`). If you build more than one, you should choose which model you will use as your final model.
* You should generate predictions on a held-out testing dataset using the model you built and calculate the testing (or cross-validated) accuracy. You will be scored on your model's accuracy on your testing set.

Data Dictionary:

|  |  |  |
| --- | --- | --- |
| Variable | Description | Data Type |
| Time | The seconds elapsed between each transaction and the first transaction in the dataset | float |
| V1 - V28 | The principal components obtained with PCA (due to confidentiality issues, we cannot provide the original features and more background information about the data). | float |
| Amount | The transaction amount | float |
| Class | The target variable to predict: 1 in the case of fraud, 0 otherwise | int |

**Submission Requirements:**

* Follow all of the directions carefully and submit all code and solutions.

**Notes:**

* The data is available for download [here](https://drive.google.com/file/d/1WMa_vYptM1H6MODfo8T9unc59ASRnsmT/view?usp=sharing).
* To prevent other students from copying your work, please make sure code is not saved on GitHub.
* To minimize score deduction, please make sure you upload the complete requirements.
* If you are having trouble uploading files on the Exam App, please DO NOT unlock another exam. Please email the file to your instructor.
* You may use any resources you have available (class notes, Google, Stackoverflow etc.) - BUT YOU MUST CITE ANY SOURCES USED. You can cite sources in the form of a code comment or a text cell with links to any resources you used. Failure to cite sources will be considered plagiarism and result in failing the exam.
* You may NOT get or receive help from any other person on this code. This includes classmates, alumni, your instructor, etc. Collaboration with another person on this exam will be considered plagiarism and result in failing the exam.