Project Proposals Ideas- Vandit Narula

1. <https://www.kaggle.com/wendykan/lending-club-loan-data>

Lending club is a peer-to-peer lending company with $11.1 billion originated loans. Since a loan default will result in a loss of both principal and interest, the goal is to attempt to maximize returns by predict the probability of default to help avoid investments that are deemed risky. The data filetype contains loan data through 2007-2015. The file contains about 890,000 observations with 75 variables.

1. <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>

IBM data scientists have supplied a dummy dataset that contains employee attrition data. The goal of this to analyze the factors that lead to employee attrition and predict which employees are more likely to leave. This dataset contains a variable for whether an employee has left or not therefore supervised learning can be applied to this dataset.

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

A synthetic dataset generated by a simulator called PaySim which simulated mobile money transactions based real transactions. The goal of this problem is to detect fraudulent activity and build a fraud detection model. For example, somebody trying to transfer more than 200,000 in a single transaction.

Fraud is challenging, it involves over time. Popular frauds.