Capstone Two: Project Ideas

## Michael Ward

# Idea #1: Tampa Bay Real Estate Investment Predictor

## Description:

The real estate market in the Tampa Bay area is very active. Single family homes are selling quickly. Real estate investors in the Tampa Bay area need to be able to assess the value of homes coming onto the market quickly and accurately so that they can beat the competition in making a competitive offer. They need to be able to evaluate the listing price against the predicted sale price in order to identify properties that may be priced below market value and would make good investments.

## Data Acquisition:

Tampa is in Hillsborough County, Florida. Historical data on property sales can be acquired from the Hillsborough County Property Appraiser’s website below.

<https://downloads.hcpafl.org/> (allsales\_01\_28\_2022.zip)

The property descriptions can also be found on the Hillsborough County Property Appraiser’s website below.

<https://downloads.hcpafl.org/> (parcel\_01\_28\_2022.zip)

# Idea #2: Ethereum Cryptocurrency Blockchain Fraud Detection

## Description:

Ethereum is a decentralized, open-source blockchain with smart contract functionality. Ether is the native cryptocurrency of the platform. There is no central authority overseeing cryptocurrency transactions on blockchain networks. Although the transactions are traceable, they are partially anonymous. Due to these facts, criminals are inclined to leverage the Ethereum network for fraudulent transactions. The objective is to build a model to identify potentially fraudulent transactions on the Ethereum network.

## Data Acquisition:

The following Kaggle link provides a dataset of 9841 Ethereum transactions and 51 columns of data. The data contains a flag to indicate if the transaction was identified as fraudulent. A total of 2179 of the transactions in the dataset are flagged as fraudulent.

<https://www.kaggle.com/rupakroy/ethereum-fraud-detection/version/1>

# Idea #3: Predicting Credit Risk of Loan Default

## Description:

Financial institutions providing loans to individuals need to assess the risk that the individual will default on the loan. Loan defaults are very costly for financial institutions as they have to spend resources in trying to recover the outstanding balance. If they cannot recover the outstanding balance, they may sell off the loan to a credit agency for pennies on the dollar. There are several attributes that can factor into calculating credit risk including credit score, annual income, outstanding loans, loan amount, loan term, interest rate, etc. Having an understanding of risk exposure can help a financial institution make better decisions on actions to take on existing loans such as increasing interest rates, call for repayment, sell the loan to another institution. It can also help the financial institution to make decisions on future loans in order to keep a balanced risk portfolio.

## Data Acquisition:

The following Kaggle link provides a datasets of 100,000 loans and 18 columns of data. The data contains a status to indicate if the loan was fully paid or charged off. A total of 23% of the loans in the dataset have a status of “Charged Off”.

<https://www.kaggle.com/azeem1607/credit-prediction-interest-rate-environments>