# Project 3

Find a problem worth solving, analyzing, or visualizing.

## Ideas:

Perform a deep dive analysis using Kaggle data from Lending Club risk analysis and metrics

## Team Members

|  |
| --- |
| Alex Ledger |
| Daniel Murchison |
| Edith Jimenez Mendez |
| Jin Dai |
| Meg Peterson |

What are we going to analyze? Why? When? How?

## List of the Project Requirements

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Requirement** | | **Responsible** | **Included** | **Priority** | **Due Date** | **Progress %** | **Complete %** | **Observations** |
| Prepare Preliminary Proposal | | All |  |  |  |  |  |  |
| Prepare Final Scope | | All |  |  |  |  |  |  |
| Use Machine Learning in the context of technologies learned. | |  |  |  | 12/1/2020 |  |  |  |
| We must use: Scikit-Learn and/or another machine learning library. | |  |  |  |  |  |  |  |
| You must use at least two of the below: | |  |  |  |  |  |  |  |
| * Python Pandas - | X |  |  |  |  |  |  |  |
| * Python Matplotlib | x |  |  |  |  |  |  |  |
| * HTML/CSS/Bootstrap |  |  |  |  |  |  |  |  |
| * JavaScript Plotly | x |  |  |  |  |  |  |  |
| * JavaScript D3.js | x |  |  |  |  |  |  |  |
| * JavaScript Leaflet | x |  |  |  |  |  |  |  |
| * SQL Database | x |  |  |  |  |  |  |  |
| * MongoDB Database |  |  |  |  |  |  |  |  |
| * Google Cloud SQL |  |  |  |  |  |  |  |  |
| * Amazon AWS |  |  |  |  |  |  |  |  |
| * Tableau |  |  |  |  |  |  |  |  |
| Prepare a 15-minute data deep-dive or infrastructure walkthrough that shows machine learning in the context of what we’ve already learned. | | All |  |  |  |  |  |  |

## List of the Project Activities

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Activities to Success** | **Responsible** | **Included** | **Priority** | **Due Date** | **Progress %** | **Complete %** | **Observations** |
|  | | | | | | | |
| Search and select dataset | All |  |  |  |  | 100% | <https://www.kaggle.com/janiobachmann/lending-club-risk-analysis-and-metrics> |
| Carry out ETL database file that meet work requirement: | All |  |  |  |  |  |  |
| * write the Jupyter notebook with the process follow to clean the data and explain the reason to do it. | Meg |  |  |  | 50% |  |  |
| * ETL Prepare: | Meg |  |  |  | 50% |  |  |
| * ETL Sample | Alex |  |  |  | 100% |  |  |
| * ETL 2011 | Tosh |  |  |  | 100% |  | Loan Clean.csv |
| **Model Selection** |  |  |  |  |  |  |  |
| Test Analysis of Models - Exploration |  |  |  |  |  |  |  |
| * Model 1- Decision Tree | Andy |  |  |  |  |  |  |
| * Model 2 – Naïve | Tosh |  |  |  |  |  |  |
| * Model 3 - Logistic | Meg |  |  |  |  |  |  |
| * Model 4 - |  |  |  |  |  |  |  |
| Prepare Principal Component Analysis | Tosh |  |  |  | 100% |  |  |
| SQL Database: Prepare database.sqlite to postgress | Alex |  |  |  |  |  |  |
| Visualization Code / Explanations | All |  |  |  |  |  |  |
| * Chart 1 – Javascript/ Matplotlib/D3/Leaflet/Tableau | Tosh |  |  |  |  |  |  |
| * Chart 2 - Javascript/ Matplotlib/D3/Leaflet/Tableau | Jin |  |  |  |  |  |  |
| * Chart 3 - Javascript/ Matplotlib/D3/Leaflet/Tableau | Meg |  |  |  |  |  |  |
| * Chart 4 - Javascript/ Matplotlib/D3/Leaflet/Tableau | Alex |  |  |  |  |  |  |
| * Chart 5 - Javascript/ Matplotlib/D3/Leaflet/Tableau | Edith |  |  |  |  |  |  |
| * Chart 6 - Javascript/ Matplotlib/D3/Leaflet/Tableau | Andy |  |  |  |  |  |  |
| Index.html for front end. CSS file |  |  |  |  |  |  |  |
| * Prepare Basic HTML / CSS | Edith |  |  |  |  |  |  |
| * Identify Visualization |  |  |  |  |  |  |  |
| * Add graph to |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Prepare Book explaining results | All |  |  |  |  |  |  |
| Put all together for presentation | All |  |  |  |  |  |  |

## Presentation Schedule on the May 28th 2020

|  |  |  |
| --- | --- | --- |
| **Team Members** | **Presentation Time** | **Section** |
| Alex Ledger | 2 minutes |  |
| Daniel Murchison | 2 minutes |  |
| Edith Jimenez Mendez | 2 minutes |  |
| Jin Dai | 2 minutes |  |
| Meg Peterson | 2 minutes |  |
| Andy Shi | 2 minutes |  |

## Example projects:

1. Create a front-end interface that maps to an API to “smarten” the algorithm.
2. Perform a deep dive of existing data using machine learning.
3. Create a visualization that continues to learn where clusters lie based on ML. (Use D3 or Plotly to change the visualization.)
4. Create an idea with mock data that simulates how machine learning might be used.
5. Create an analysis of existing data to make a prediction, classification, or regression.

## Ideas project

1. Perform a deep dive analysis using Kaggle data from Lending Club risk analysis and metrics
   1. <https://www.kaggle.com/vanshjatana/applied-machine-learning>
   2. <https://www.kaggle.com/eraaz1/a-comprehensive-guide-to-titanic-machine-learning>
   3. <https://www.kaggle.com/gpiosenka/100-bird-species>
   4. <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset/kernels>
   5. <https://www.kaggle.com/janiobachmann/lending-club-risk-analysis-and-metrics>
   6. <https://www.kaggle.com/ionaskel/credit-risk-modelling-eda-classification>

NOTES

Explore some models

Accuracy is important.

Clean presentation