**Kickstarter Project Plan**

**Use Case:** (What is the problem you are trying to solve? What is its purpose and how does it provide value?)

* Provide a tool for donors to understand which projects will be successful to prioritize where/how to fund a project.

**Objectives/Goals:** (What is the desired end-state of the project?)

* A classification model that provides clear results on which projects are predicted to be successful.
* A completed proof-of-concept model and dashboard
* *For Penn’s benefit*, technical exposure and experience!

**Scope:** (What parameters are you working within to achieve the goal?)

* Pre-provided data that has already been collected and sourced
* Limit modeling to an XGBoost classification model

**Deliverables:** (What is going to be delivered?)

* Single Classification Model
* PowerBI Dashboard
* Readout Deck/Executive Summary of process and findings

**Resources**: (Identify all the necessary resources that must be used to create and deliver the final product, including technology and team members.)

* Jupyter Notebooks, Power BI (all local)
  + Packages for Model Building
    - Pandas
    - XGBoost
    - MLflow
* Brandon & Courtney – guidance, support, touchpoints, etc.

**Limitations/Dependencies:** (What factors will inhibit your ability to complete the project, and how do you mitigate those factors?)

* Prioritization of billable work
* Availability of Azure resources

**Timeline**: (Record all desired deadlines and checkpoints.)

* 4 Weeks, see project phases for deadlines

**Stakeholders:** (Who is your audience? Record how much time your stakeholder may need to review and approve any component of the project.)

* Brandon

**Success Criteria:** (How do we determine if this project was a success? What are we measuring?)

* The donor could feel more confident in their ability to fund successful projects.
  + Based on experience and baseline knowledge of the platform, aiming for .7 score, balancing both recall and precision
    - Improves value 20% since more informative than flipping a coin
    - Recall is more relevant because we want to focus on best potential investments
* Positive donor feedback and outcomes

**Communication Plan:** (What is the desired method and cadence of team communication?)

* Weekly update with Brandon

**Stakeholder FAQ**

* How can I use this day to day?
  + Productionization conversation
* How do I interpret the model?

**Win Condition**

* .7 precision, with improvements added if there is still time within the project timeline
* Client has found value and desires to continue engagement with next steps

**Project Phases**

Planning & Assessment – *1/2 Week*

* Discovery & Requirements Gathering
* Data Assessment – what data do we have/what data do we need to get?
* Technology Set-Up/Access
* Acquire & Load Data

Data Preparation – *1 Week*

* Data Profiling & Exploration
  + - Understand data distribution and features of the dataset
    - Profiling Package
* Data Cleaning
  + - Parse JSON
    - Assess Data Quality
    - Cleanse Data
* Model Prototyping Dataset
  + - Feature Importance
    - Feature Selection

Analysis & Modeling - *2 Weeks*

* Iterative Model Development
  + - Basic XGBoost classification model
* Selection of Best Performing Model
* Model Explainability
* Data Visualization
  + - PowerBI – connect to model results to visualize and build basic dashboard

Readout & Presentation – *1/2 Week*

* Summarize Modeling Results
* Deliver Findings & Output
* Archive Resources