**Proposal for Capstone**

*Introduction:*

Banning books in public schools has a long and controversial history in the United States and is not a practice that has been decreasing. According to the New York Times, “the American Library Association received an “unprecedented” 330 reports of book challenges last fall.”[[1]](#footnote-0) Does banning books protect the innocence of children or limit their ability to learn? Is there a connection between areas where the overall population is less educated and those that ban more books? While exploratory data analysis may not be able to fully answer these questions and their complex implications, exploring data for banned books, literacy rates, and education levels may be able to help us find useful patterns and connections. Finally, we can compile these findings to help individual school districts, administrators, and educators understand more about the practice of banning books and ideally to understand additional facts to consider as they make these choices in the future.

*Project description:*

This project is aimed at helping Client X create a more personalised customer experience. We will focus on two main objectives: (1) an analysis of banned books, literacy rates, and education levels and (2) the development of a presentation to help express this analysis and any relevant findings. These two branches of the project will combine to give school districts a better understanding of banning books and will allow them to consider these facts in the choices they make for future students.

In Objective 1, “Insights”, we will perform an analysis on data such as PEN America's Index of School Book Bans, National Center for Education Statistics, World bank world development indicator, and any other easily accessible, reliable, relevant datasets deemed necessary during exploratory analysis. We will first perform an in-depth analysis in which we will assess data quality and completeness and perform preliminary exploratory analyses, looking for clear relationships between various aspects. We will then select a subset of features that we feel have the most utility for meaningful, interpretable conclusions. To achieve this we will explore several methodologies before determining which we feel has the most promise for these particular datasets.

While performing this and future analysis, we will also consider problems or concerns with the data, its collection, or any other indications that we should be cautious of as we proceed. We may be unable to avoid any of these concerns, but we will monitor them as we analyze the data and report them during our final presentation as to avoid misleading conclusions. Before beginning analysis, we are considering aspects such as how literacy rates are determined, the potential for researcher and analyser individual bias in areas like worldview, politics, and personal experience,

In Objective 2, “Product recommendations”, we will focus on creating recommendations that can be used to advise school districts and others considering future book banning decisions. We plan to use the aspects identified in Objective 1 as the focus of these conclusions. However we may find through the course of researching that a different approach is more appropriate. We have broken this project into two discrete phases: research and building. The research phase will explore various approaches intended to gain enough knowledge to identify the most meaningful connections.

We outline the plan for this project below.

**Project Proposal: Insights and Recommendations[[2]](#footnote-1)**

*Objective: This project is aimed at developing recommendations that will employ innovative approaches to understanding any connections between banning books and literacy in individual school districts.*

**Objective 1: Data insights**

| **Milestone** | **Work Carried Out** | **Outcome/Deliverable** | **Time to complete** |
| --- | --- | --- | --- |
| Milestone 1 | * **Exploratory data analysis** * We will explore the data available to support the project. * We will ensure that data are accessible and will perform a preliminary analysis with a view to identifying a subset of features that best-suited for deeper analysis in Milestone 2. | * A presentation or report showing the initial findings of relationships between features. * A list of features to use for further research. | 6-8 hours |
| Milestone 2 | * **In depth analysis** * Using the learnings from Milestone 1, we will experiment with various algorithms to find any further insights from the data. * The most useful aspects will be selected for recommended improvements. | * A GitHub repository containing Python code with any useful models. * A brief report highlighting findings. | 4-6 hours |

**Objective 2: Recommendations**

| **Milestone** | **Work Carried Out** | **Outcome/Deliverable** | **Time to complete** |
| --- | --- | --- | --- |
| Milestone 3 | * **Conclusions** * We will complete and compile our findings to reach any concrete conclusions. * We will determine recommendations that could be useful for school districts. | * A brief report outlining conclusions and possible approaches to improve outcomes for school districts. | 4 hours |
| Milestone 4 | * **Report and present** * We will focus on how to communicate our findings and any recommendations in ways that are accessible to school board members, administrators, and educators. | * A presentation and report outlining experimental approach and detailed findings, as well as recommendations.. * A GitHub repo containing all code used. | 6 hours |

1. [Elizabeth A. Harris](https://www.nytimes.com/by/elizabeth-a-harris) and [Alexandra Alter](https://www.nytimes.com/by/alexandra-alter). “Book Ban Efforts Spread Across the U.S.” The New York Times. Published Jan. 30, 2022. Updated Feb. 8, 2022. Accessed Aug. 7, 2022. www.nytimes.com/2022/01/30/books/book-ban-us-schools.html. [↑](#footnote-ref-0)
2. The format and some language from this proposal follow an example by Jonathan Leslie and Neri Van Otten. “Designing and Building Data Science Solutions: 8 Example project plan.” Created Nov. 1, 2022. Accessed Aug. 7, 2022. datasciencedesign.com/example.html. [↑](#footnote-ref-1)