Protect PT
Executive Summary

Community Partner
James Cirilano
Team
Justin Oeni, Yang Pan, Mingxi Yan

Community Partner Background

Protect PT (Penn-Trafford) is a nonprofit advocacy organization based in Westmoreland and Allegheny counties in Western Pennsylvania. Founded in 2014, its mission is to ensure residents' constitutional rights to clean air, pure water, and a healthy environment by opposing environmentally harmful oil and gas development. Protect PT combines community education, legal advocacy, and environmental monitoring to empower residents in policy-making and litigation efforts. Their work includes newsletters on DEP permit activity, legal filings challenging unsafe fracking proposals, environmental data collection, and public education workshops.

Project Description

Project Opportunity

Fracking and other fossil fuel extraction activities in Western Pennsylvania continue to pose threats to public health and the environment. At the start of this engagement, Protect PT lacked an efficient and centralized system to collect, validate, and use data from the Department of Environmental Protection (DEP) and other external sources. Manual workflows were time consuming and error prone. There was an opportunity to streamline Protect PT's data collection and case preparation efforts with a technology-enabled solution that matched their legal and advocacy workflows.

Project Vision

Our vision was to enhance a previously initiated data infrastructure by implementing a centralized, queryable system that aggregates DEP violation data and links it with supporting documents. We aimed to deliver a usable, scalable, and maintainable Airtable based database that would allow Protect PT staff to focus less on data wrangling and more on advocacy. The system needed to be intuitive for staff without coding experience and flexible enough to grow with Protect PT's evolving use cases.

Project Outcomes

Our final product is a fully functional Airtable database that aggregates site violation data from the Pennsylvania DEP, allows PDF uploads tied to specific fracking sites, and enables Protect PT staff to quickly filter and search through records. On the people side, we successfully trained staff members to upload, filter, and search data independently. Their feedback confirmed that the tool aligns well with their workflows and increases their ability to respond to community concerns efficiently.

On the process front, we developed comprehensive Standard Operating Procedures (SOPs) to guide ongoing data entry and system use. These SOPs help ensure consistency across users and enable smooth handoffs in the event of staffing transitions. From a technology perspective, we implemented a semi-automated data pipeline using a Python script that pulls DEP data into Google Sheets and syncs it

into Airtable. We also built a normalized, relational schema in Airtable that links violation records with supporting documents, allowing users to view a complete site history with minimal effort.

Project Deliverables

The project deliverables include a working Airtable base with linked tables that store DEP violation records and associated PDF documents. We also provided a semi-automated Python script that extracts data from the DEP's online reports and pushes it into the Airtable system via a Google Sheets intermediary. To ensure sustainability, we delivered a detailed set of Standard Operating Procedures that explain how to use and maintain the system. Finally, we packaged all credentials, source code, and supporting documentation into a centralized Google Drive folder to facilitate long-term access and maintenance.

Recommendations

To ensure continued success and scalability, we recommend that Protect PT upgrade to Airtable Pro, which would allow for full automation of the data pipeline and remove the need for a manual Google Sheets intermediary. We also encourage regular maintenance and updating of the SOPs, and that these materials be incorporated into onboarding for any new staff or volunteers. In addition, we recommend developing public-facing dashboards using platforms like Softr or Tableau to share violation trends and increase community engagement. We also advise implementing basic error-logging and alert mechanisms to flag pipeline failures or data inconsistencies early.

About The Team

Mingxi Yan is a junior studying Information Systems and Computer Science at Carnegie Mellon University, looking for a career in software engineering or technical product management. Mingxi worked on the project as a technical developer, making technical decisions and documentation.

Justin Oeni is a senior studying Information Systems and Statistics and Machine Learning at Carnegie Mellon University, pursuing a career in Finance. Justin worked on the project as the project manager, handling the communication between the community partner and the faculty advisor. He was also in charge of assigning tasks to the team and finalizing deliverables.

Yang Pan is a junior studying Information Systems and Computer Science at Carnegie Mellon University, looking to pursue a career in software engineering. Yang primarily worked as the software developer for this project, focusing on syncing the various platforms together.