Module 6 Group Assignment

Roadmap for Power of Patients Project

Professor: Valerie Atherley

Group Number: 5

Submitted by:

Kumar Saransh

Maljha Fatima Syeda

Krishna Murari Sharma

Shlok Sharma

Ishan Srivastava

Northeastern University: College of Professional Studies

ALY 6080: Integrated Experiential Learning

21st May, 2025

Abstract

This roadmap outlines our team's strategy for the Power of Patients capstone project. It details the analytic methods, key milestones, team structure, risk management, success criteria, and presentation plan. We leverage Python for data ingestion, cleaning, exploratory analysis, predictive modeling, and visualization. This is supplemented by SQL for initial data queries and GitHub for version control.

Our milestones span six weeks—from data preparation through proof-of-concept delivery; ensuring measurable progress at each stage. Kumar Saransh leads overall execution, with Fatima, Ishaan, Slok, and Krishna providing specialized support. We identify potential risks (data quality, scope creep, model over fitting, integration delays) and prescribe mitigation strategies.

Success will be demonstrated via an interactive dashboard achieving ≥80% model accuracy and sub-200 ms visualization load times. The final deliverable comprises a live demo and concise slide deck, positioning stakeholders to advance to full implementation.

Introduction

Effective patient engagement analytics demand a structured approach that integrates robust data engineering, rigorous modeling, and intuitive visualizations. In partnership with Power of Patients, our objective is to build a proof-of-concept system that correlates environmental factors—such as humidity and temperature—with patient-reported symptom flare-ups.

This document presents a detailed roadmap for achieving that goal over an eight-week timeline. It begins by describing our analytic approach (data preprocessing, exploratory analysis, predictive modeling, visualization). Then it defines clear milestones, assigns responsibilities, highlights potential risks with mitigation plans, and establishes quantitative measures of success. Finally, we outline our method for presenting the proof of concept through a live dashboard demonstration and supporting slides.

Analytic Approach (6-Week Plan)

Weeks 1–2: Sponsor Onboarding & Compliance

- Conduct background research on Power of Patients' mission, workflows, and existing tool chain.
- Review and sign NDA; establish data governance and access protocols.

Week 3: Dataset Exploration & Requirements Alignment

- Run SQL queries to skim each provided sheet and catalog variable definitions.
- Map data tables to project requirements; document gaps and clarifying questions for the sponsor.

Week 4: Data Preparation

- Clean and anonymize patient and clinical records in Python (pandas).
- Standardize formats, handle missing values, and enforce controlled vocabularies.

Week 5: Individual-Dataset EDA & Visualization

- Generate summary statistics and single-source visualizations (histograms, time series, scatterplots) in Python (matplotlib/plotly).
- Identify key patterns within each dataset and produce standalone dashboards for review.

Week 6: Integrated EDA & Initial Modeling

- Merge all clinical sheets with NOAA weather data.
- Perform cross-dataset exploratory analysis to uncover weather–symptom correlations.
- Train a preliminary predictive model; document initial accuracy and feature insights for refinement in the next phase.

Milestones & Progress Checks

Timeline	Milestone	Deliverable
Weeks 1–2	Sponsor Onboarding & Compliance	Sponsor research summary; signed NDA; data-access protocols document
Week 3	Dataset Exploration & Requirements Alignment	SQL skim reports for each sheet; mapped variables to project requirements; list of clarifications
Week 4	Data Preparation	Cleaned/anonymized patient and clinical data frames; standardization log
Week 5	Individual-Dataset EDA & Visualization	Standalone EDA notebooks; Python-generated charts (histograms, time-series, scatterplots)
Week 6	Integrated EDA & Initial Prediction	Merged clinical + NOAA weather dataset; cross-dataset EDA summary; preliminary model results

Team Roles & Responsibilities

Teammates	Roles & Responsibilities
Kumar Saransh	• Led sponsor research, negotiated and signed the NDA, established data-access protocols • Oversaw SQL skims and Python data-cleaning pipelines; mapped tables to requirements • Coordinated individual and integrated EDA, directed preliminary model training, and reviewed all deliverables • Managed GitHub branching strategy, pull-request reviews, and merge approvals
Maljha Fatima Syeda	• Supported background research and compliance documentation • Authored and optimized SQL queries to inventory each dataset's structure • Assisted in data-anonymization and format standardization routines
Ishan Srivastava	• Designed and coded standalone EDA scripts and visualizations (histograms, time-series, scatterplots) in Python • Refined cross-dataset charts post-merge; validated interactive dashboard performance
Shlok Sharma	• Collaborated on cleaning routines and missing-value imputation in pandas • Helped merge clinical sheets with NOAA weather data and ran the initial predictive model
Krishna Murari Sharma	• Managed the technical merge of all sheets with NOAA data, resolved schema mismatches • Maintained GitHub issues and pull requests; kept README, data dictionaries, and technical notes up to date

Key Risks & Mitigation Strategies

- **Data Quality Issues:**
  Mitigation: Automated pandas assertions; peer review of SQL extractions.
- **Model Over fitting:**
  Mitigation: Stratified k-fold CV; hold-out validation; early stopping if needed.
- **Scope Creep:**
  Mitigation: Only implement high-priority features; freeze feature list after Week 2.
- **Performance Bottlenecks:**
  Mitigation: Profile Python scripts; down sample for prototype; optimize loops/queries.

Measures of Success

- **Dashboard Functionality:** Interactive Python-based dashboard loading key views in <200 ms.
- **Model Accuracy:** ≥ 80% on test set; ROC AUC ≥ 0.80.
- **Code Quality & Reproducibility:** All scripts pass linting; documented in GitHub.
- **Stakeholder Approval:** Formal sign-off at end of Week 6 demo.

Presentation & Delivery

- **Live Demo (10 min):** Showcase end-to-end flow
  raw data → cleaned data → model prediction → visualization.
- **Slide Deck (5 min Q&A):** 6–8 slides summarizing objectives, methods, results, risks, and next steps.
- **Supporting Materials:** Technical memo (1–2 pages) detailing model performance and data-quality findings.

Conclusion:

Our roadmap provides a comprehensive guide to deliver a high-impact proof of concept within the specified timeframe and resource constraints. By centering on Python-based workflows, SQL-driven data sampling, and GitHub-managed collaboration, we ensure technical consistency and traceability.

The division of roles facilitates parallel progress. Kumar Saransh steers core development while supporting team members enhance data preparation, visualization, modeling, and documentation. Risk controls and success metrics keep the project on track toward stakeholder approval.

The planned live demo and concise slide deck will effectively showcase our system's capabilities. This presentation will pave the way for full-scale implementation in patient engagement analytics.

References

- Power of Patients. (n.d.). Retrieved April 16, 2025, from https://www.powerofpatients.com
- Statology. (n.d.). Pandas wide to long. Retrieved April 16, 2025, from https://www.statology.org/pandas-wide-to-long/
- GeeksforGeeks. (n.d.). Exploratory data analysis in Python. Retrieved April 16, 2025, from https://www.geeksforgeeks.org/exploratory-data-analysis-in-python/
- Nussbaumer Knaflic, C. (2015). Storytelling with data: A data visualization guide for business professionals
- National Centers for Environmental Information. (n.d.). Climate data online search. Retrieved April 16, 2025, from https://www.ncei.noaa.gov/cdo-web/search