

Purva Prakash Kekan

Boston, MA | 857.869.4863 | purvakekan3@gmail.com | linkedin.com/in/purva-prakash-kekan | <https://purva-kekan.github.io/portfolio/>

EDUCATION

Masters of Science in Analytics | GPA 3.93/4.0 - Northeastern University | Boston, MA

Award: “CPS Scholars and Leaders Award” - Awarded to top 1% graduate students for exceptional academic achievement and leadership
Relevant Coursework: Big Data, Predictive Analytics, Enterprise Risk Management, Data Mining, Healthcare and Pharma Analytics

Bachelors of Science in Information Technology | GPA 3.70/4.0 - Mumbai University | Mumbai, India

Relevant Coursework: Database Management, Business Intelligence, Project Management, Python, Artificial Intelligence

WORK EXPERIENCE

Data Analyst, IpserLab | Remote, USA

May 2025 – Present

Early-stage AI startup incubator building CAVO, an inclusive trip-planning platform focused on traveler safety and accessibility

- **Led analytics** building data infrastructure from ground up, **processing 100K+ geospatial records via Google Places API**
- **Engineered 15+ Power BI dashboards** with DAX transformations, **improving stakeholder visibility into platform metrics by 30%**
- Crafted production ETL workflows with **automated quality checks, reducing ML model latency by 25%** across 5+ models
- Implemented data governance standards and documentation, **boosting data consistency and reliability by 20%**

Healthcare Data Analyst, Boston Public Health Commission | Boston, USA

Jan 2025 – Mar 2025

Boston’s \$300M+ municipal public health agency delivering healthcare, safety, and equity programs to 700K+ residents

- **Built end to end time series forecasting system with SARIMA models on 65K+ records, optimizing resource allocation by 35%**
- Processed location data across 12 city zones by **building geospatial pipelines with ArcGIS & Python**, boosting accuracy by 24%

Data Analyst, KARA - Kids At Risk Action | Boston, USA

Apr 2024 – Jun 2024

U.S. nonprofit advocacy group dedicated to strengthening child welfare systems and advancing national policy reforms

- **Refined prediction accuracy by 40%** on 1M+ child welfare service records using **Scikit learn machine learning classification model**
- **Automated 25%** of data cleaning and transformation processes for large datasets by **building scalable SQL based ETL pipelines**

Data Analyst, HDFC Bank Ltd | Mumbai, India

Jan 2023 – Jun 2023

India’s largest private sector bank with \$21B+ annual revenue, 173,000+ employees, and operations across 6,300+ branches in 18 countries

- **Optimized data transformation workflows** using SQL for credit risk modeling on 500K+ financial records, **improving 35% efficiency**
- Generated CI/CD pipelines using dbt & GitHub Actions, automating feature engineering and **reducing manual reporting by 40%**

Data Analyst, Acute Bioscience | Remote, India

Jul 2022 – Dec 2022

Indian life sciences distributor supporting 100+ healthcare and research organizations with lab consumables and equipment.

- **Developed scalable ETL pipelines via MongoDB & Python**, analyzing 50K+ transactions monthly and **boosting data accuracy by 25%**
- **Predicted product demand** by building R based time series forecasting model, **decreasing 20% stockouts and 12% in service delays**

TECHNICAL SKILLS

- **Programming and Databases:** Python, SQL, R, MongoDB
- **Data Engineering and DevOps:** ETL/ELT Pipelines, dbt, CI/CD (GitHub Actions), Git
- **Machine Learning and NLP:** Scikit-learn, SARIMA, BART Transformers, TF-IDF, Classification Models
- **Visualization & Analytics:** Power BI, DAX, Streamlit, PyVis, ArcGIS, Google Places API

PROJECTS

Atomic Habits: NLP Text Mining and Knowledge Graph System

Tech stack: PyVis, TF-IDF, Python, Streamlit

- **Created interactive Streamlit dashboard** with PyVis network graphs **visualizing 50+ concept relationships and behavioral patterns**
- Constructed NLP pipeline with BART transformers, TF-IDF for text summarization and keyword extraction, **processing 18 chapters**

NFL Play Type Prediction: ML Classification Analysis

Tech stack: Python, Scikit-learn, machine learning models, Pandas

- **Identified top 5 predictive game variables** by performing VIF analysis & engineered interaction features for multicollinearity detection
- **Compared 6ML models on 100K+ NFL plays** building binary classification pipeline, **achieving 67% accuracy** with SVM