Chronic Absenteeism and Attendance Patterns in NYC Schools (2013–2019)

1. Project Team

- Team Members: Nasim Aalemi and Shobha Panthi
- Title: Chronic Absenteeism and Attendance Patterns in NYC Schools (2013–2019)

2. Project Overview

This capstone project investigates student attendance trends in New York City public schools between 2013 and 2019 using a large dataset of 824,147 attendance records. The project seeks to uncover factors driving chronic absenteeism and forecast future attendance risk through combining exploratory data analysis (EDA) with predictive modeling. We are specifically interested in answering which demographic and school-level factors are most predictive of chronic absenteeism in NYC schools and how attendance patterns have changed from 2013 to 2019. The final outputs will include:

- A Power BI dashboard for interactive exploration of attendance trends
- A Streamlit web app to predict chronic absenteeism risk by student group or school profile

These tools will allow educators to identify schools and student groups at high risk for chronic absenteeism, supporting targeted attendance interventions.

3. Data Source

- Primary Dataset and Metadata Link: https://data.cityofnewyork.us/Education/2013-2019-Attendance-Results-School/vww9-qguh/about_data

4. Dataset Summary

The dataset includes annual school-level attendance metrics, disaggregated by grade, demographic category, and student group.

- School Identity & Structure
- Time Dimension
- Student Demographics
- Attendance Metrics
- Chronic Absenteeism Indicators

5. Modeling Approach and Analysis

Exploratory Data Analysis (EDA)

- Compare attendance rates across years, grades, and demographic groups
- Identify trends in chronic absenteeism by race, gender, and special populations
- Evaluate school-level variability and outliers
- Track impacts of time (policy shifts or socioeconomic changes)

Classification Model

- Goal: Predict whether a given subgroup will have a high chronic absenteeism rate
- Target: Binary label based on whether % Chronically Absent > 20%
- Features: Grade, demographic group, total days, school, year
- Models: Logistic Regression, Random Forest, XGBoost
- Metrics: Accuracy, F1-Score, ROC-AUC

Regression Model

- Goal: Predict % Chronically Absent as a continuous outcome
- Model: Linear Regression, Gradient Boosting, XGBoost
- Metric: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared (R^2 Score)

6. Frontend Implementation

Power BI Dashboard

- Interactive Filters: School year, grade, demographic group, school
- Visuals:
  - Time series of attendance and absenteeism rates
  - Heatmaps of chronic absenteeism by grade and year
  - Bar charts comparing subgroups (Female vs Male, ELL vs non-ELL)
- KPIs:
  - Citywide average attendance
  - Subgroup absenteeism gaps
  - Most improved schools or groups

Streamlit Web App or Interactive Prediction Tool

- User Inputs:
  - School, grade, year, demographic characteristics
- Model Outputs:
  - Predicted % Chronically Absent or high-risk flag
  - Suggestions for interventions (if applicable)

7. Expected Outcomes

- Data-driven insight into equity gaps and high-risk student groups
- A prediction tool for identifying future risk of chronic absenteeism
- Visualizations that support policy review and funding decisions
- A deeper understanding of how attendance patterns shift by population

8. Conclusion

This capstone will leverage NYC's rich attendance dataset to spotlight attendance disparities, develop practical forecasting models, and empower decision-makers with interactive dashboards. The insights generated will help improve school engagement and reduce chronic absenteeism at a systemic level.