

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green color. They are positioned diagonally, with the blue one partially covering the green one.

Classification of Poverty

Crystal Han

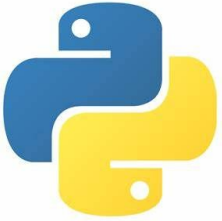


Introduction

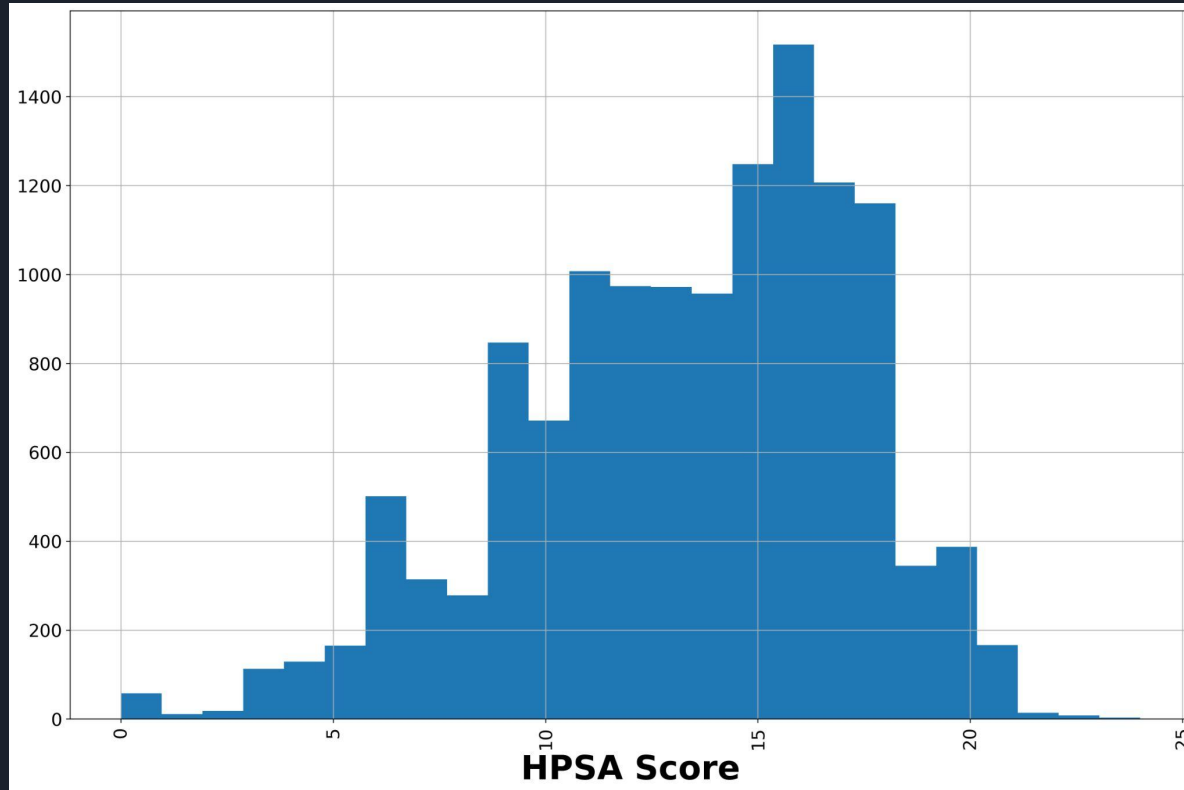
- HPSA (Health Professional Shortage Areas) Scores below 15 are considered in poverty (>10,000 datapoints)
- Government can send Healthcare workers to impoverished areas to help
- Target: Impoverished areas, places lacking healthcare workers
- **Agenda:** Classify areas as in poverty or not in poverty and provide a direction on where to allocate healthcare workers

Methodology

- Data
 - Health Professional Shortage Areas - Google Big Cloud Query
 - Supplementary Population County-state data - US Census Bureau
 - In poverty if HPSA < 15
 - Data filtered between 2010-2019

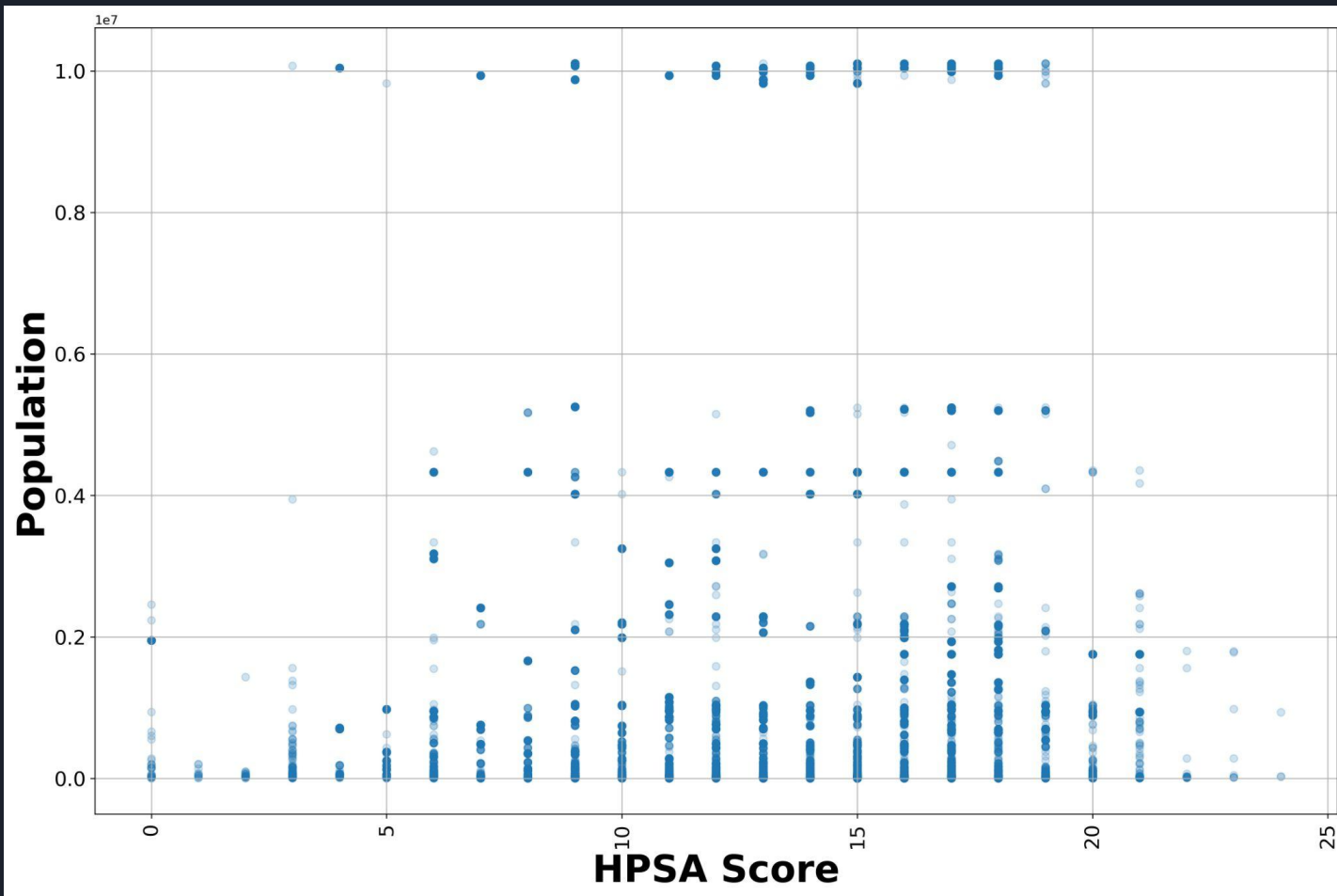


Histogram of HPSA Score



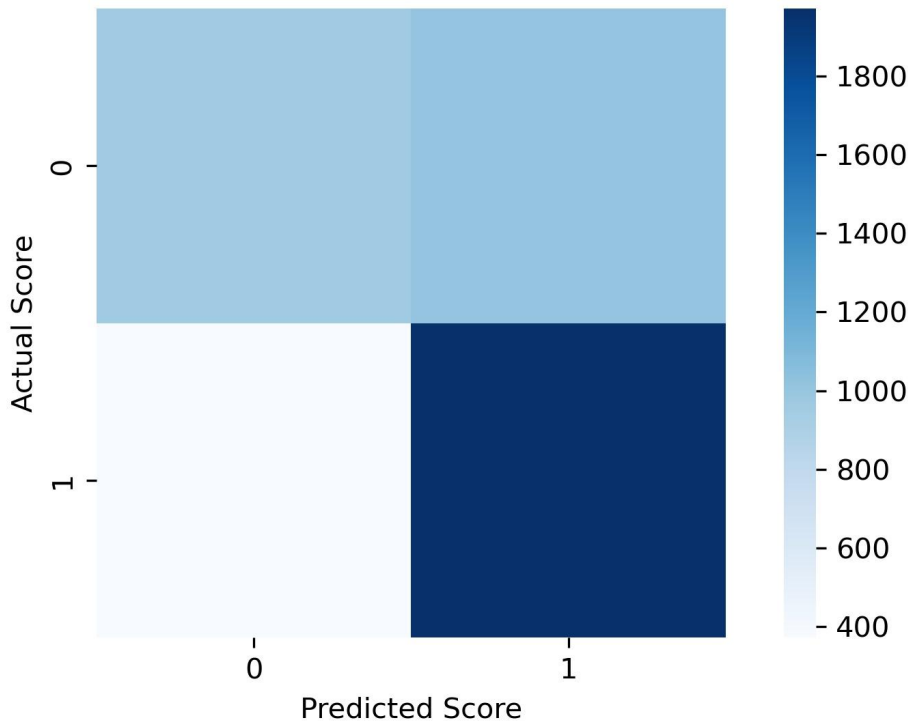


Population vs HPSA Score Scatter Plot

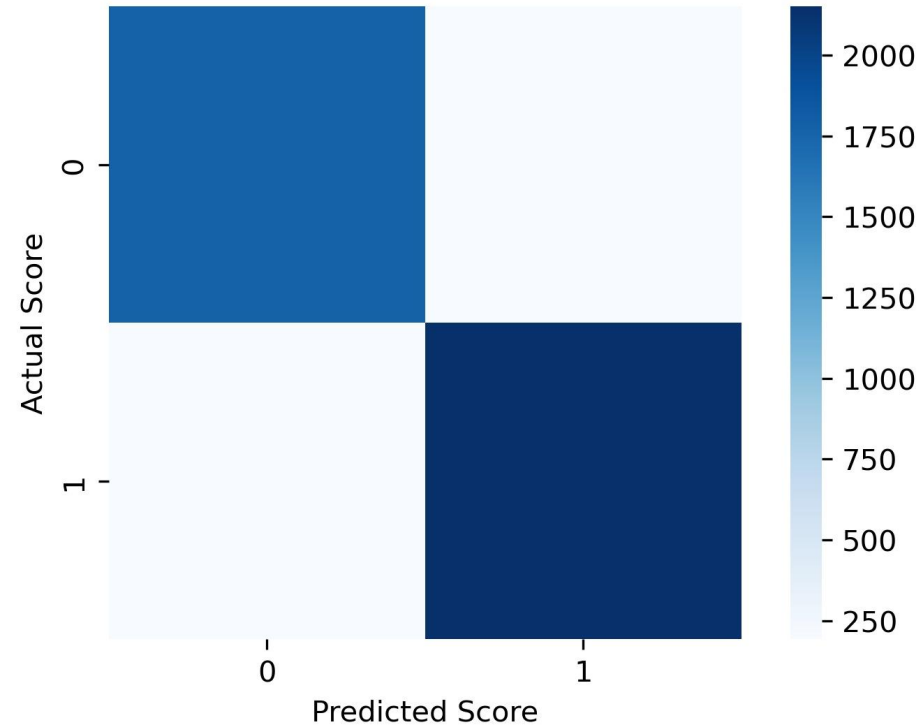


Confusion Matrix - Logistic vs KNN

Logistic Regression Confusion Matrix



KNN Confusion Matrix





Machine Learning Model Scoring

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.6785	0.6899	0.6630	0.6596
KNN	0.9094	0.9087	0.9086	0.9086



Future Work

- Using other machine learning models to predict and classify poverty
- Compare levels and scores between previous decade with the current decade
- Gather more supplementary and recent data to provide better classification and predictions



Conclusion

- Poverty levels seen in populations less than 50 million
- Other factors may cause poverty or shortage of healthcare workers
- Model was able to predict and classify most of the data



Appendix

- Time frame: 2010-2019
- Data cleaning: Fixing some outliers that do not present normal behavior, dropped rows that had nan values as supplementary data was not found for it
- Reason for selecting time frame:
 - More relatable with recent years as it workers can be allocated to those areas
 - Previous time frames had many more missing data and different templates of storing data that were difficult to handle or read

Thank you

