# Big Data for Crime Data Analytics

This project explores crime data using Python and data analytics techniques to identify patterns and classify regions based on violent crime rates. It includes data preprocessing, visualization, statistical analysis, and binary classification using the mean of ViolentCrimesPerPop.

## 📊 Project Highlights

* Load and clean crime data (.csv)
* Handle missing values and convert columns to numeric
* Calculate average violent crime rate
* Classify areas as high/low crime occurrence
* Perform grouping and analysis
* Prepare the data for future machine learning or dashboard use

## 🛠 Tools Used

* Python
* Pandas
* NumPy
* Matplotlib & Seaborn
* Jupyter Notebook

## 📂 Project Structure

Big-Data-Crime-Analytics/  
├── crimedata.csv # Raw crime data file  
├── crime\_analysis.ipynb # Jupyter notebook with complete analysis  
├── README.md # Project overview  
└── requirements.txt # Python package dependencies (optional)

## ⚙️ How to Run

### Clone the repository

git clone https://github.com/your-username/Big-Data-Crime-Analytics.git  
cd Big-Data-Crime-Analytics

### Install dependencies

pip install pandas numpy matplotlib seaborn

### Run the Jupyter Notebook

jupyter notebook crime\_analysis.ipynb

## 🔍 Sample Analysis Performed

* Converted ViolentCrimesPerPop to numeric
* Replaced ? values with NaN and dropped invalid rows
* Calculated mean crime rate
* Classified areas:
  + **1** → Above average violent crime
  + **0** → Below average
* Grouped data and calculated overall averages

## 📈 Example Output (Console)

population householdsize racepctblack ...   
violent\_crime\_occurrence   
0 0.56 2.50 0.12 ...  
1 0.72 2.80 0.32 ...

## 😋 Author

**Amaljith Pv**  
📧 [Email](mailto:your-email@example.com)  
🔗 [GitHub](https://github.com/your-username)

## 📄 License

This project is open-source and available under the [MIT License](LICENSE).