

# ZECHEN YANG

## # NYPD Complaint Data Processing Report

### ## Data Source

- Dataset: NYPD Complaint Data Historic
- Format: CSV
- Fields: 34 columns including complaint details, dates, location, and demographic information
- Key Fields:
  - \* CMPLNT\_NUM (Complaint Number)
  - \* CMPLNT\_FR\_DT (Complaint From Date)
  - \* OFNS\_DESC (Offense Description)
  - \* BORO\_NM (Borough)
  - \* Geographical coordinates (Latitude/Longitude)
  - \* Demographic information (Age groups, Race, Sex)

### ## Data Processing Steps

#### 1. Data Profiling

- Analyzed date formats and validity
- Profiled crime categories distribution
- Analyzed geographical coverage
- Assessed demographic data completeness

#### 2. Data Cleaning

- Standardized date formats to YYYY-MM-DD
- Normalized age group categories
- Validated and formatted geographical coordinates
- Standardized missing value handling

#### 3. Quality Improvements

- Standardized case for categorical fields
- Validated coordinate pairs
- Normalized borough names
- Handled missing values consistently