

THE DATABASE

by Philip Fabrelus & Jan Stein

**A story about lonely entities looking for
relationships**

PROJECT GOALS

Use our dataset to create:

- Relational DB
- NoSQL DB

Analyse the dataset

Create interactive Python environments for both

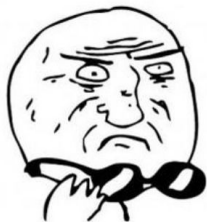
(The true goals are the relations we defined along the way)



DATASET

The Murder Accountability Project database contains U.S. homicide records from 1980 to 2014. It is the most complete database of homicides in the United States currently available

MAMMA MIA



THAT'S A LOT OF COLUMNS

Columns

Record ID, Agency Code, Agency Name, Agency Type, City, State, Year, Month, Incident, Crime Type, Crime Solved, Victim Sex, Victim Age, Victim Race, Victim Ethnicity, Perpetrator Sex, Perpetrator Age, Perpetrator Race, Perpetrator Ethnicity, Relationship, Weapon, Victim Count, Perpetrator Count, Record Source

columns: 24

ER DIAGRAM

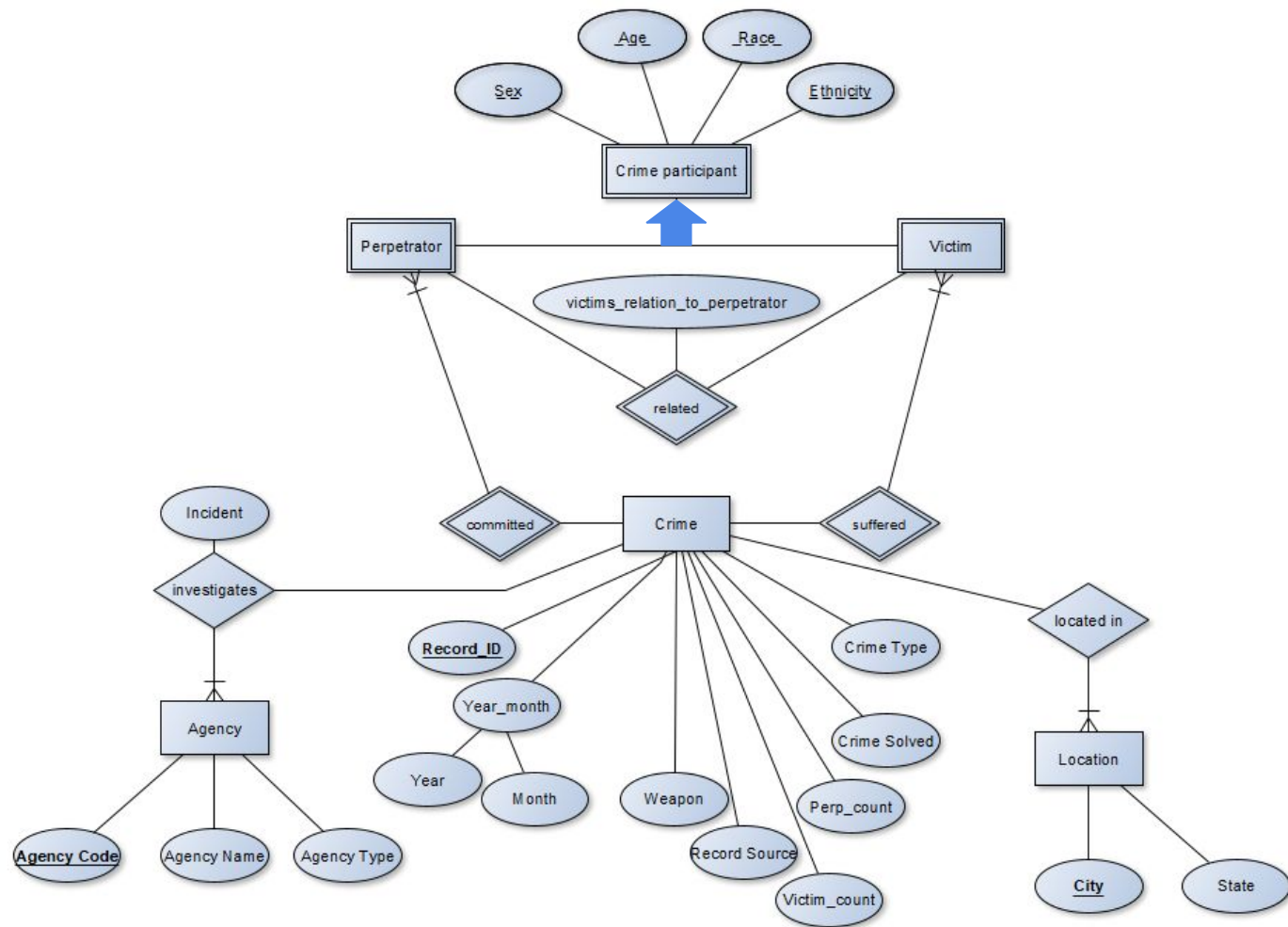
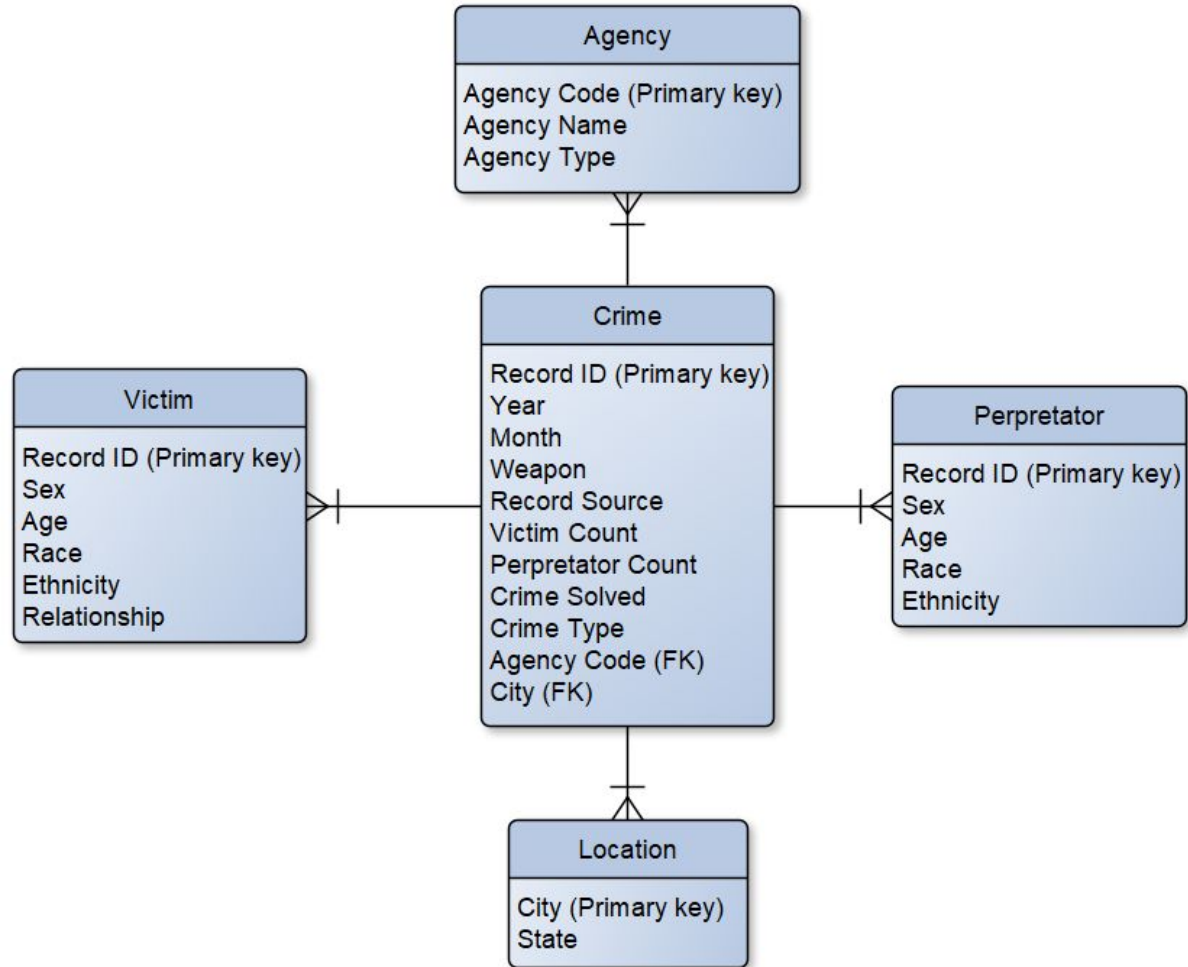


TABLE SCHEMA



DATA INTEGRITY & DATA CLEANING

After importing our dataset we investigated data integrity with SQL queries

Low and behold - problems were found

SQL UPDATE queries properly replaced these with NULL

Identified issues:

Victim sex - 'Unknown'
Victim age - 99, 998
Victim race - 'Unknown'
Victim ethnicity - 'Unknown'
Perpetrator sex - 'Unknown'
Perpetrator age - 0
Perpetrator race - 'Unknown'
Perpetrator ethnicity - 'Unknown'
Relationship - 'Unknown'
Weapon - 'Unknown'
Agency Name - 'Unknown'

Incident - inconsistent data

NoSQL

NoSQL DB in MongoDB

used Pymongo/MongoClient to create the DB in Python

Results:

1 collection

1 additional Python Application

5 queries

Fraction of time spent creating DB



ANALYSIS AND FINDINGS

Creating NoSQL DB was a lot faster and easier

There are a lot of murders in the US

Debugging is 90% of the work

Random datasets from the internet should not be trusted to be 'clean'

More common to murder your partner during the summer compared to during the winter.

The solve rate for murders with handguns are lower compared to other types of weapons, especially from year 1990 onwards.

More common to murder someone younger than yourself.