

[illegible]

The SFPD data was retrieved and cleaned by:

- Removing un-needed columns (such as the incident ID and date-time of each incident)
- Removing incident reports that don't have latitude and longitude included.
- Removing incident report types that don't relate to a crime (such as "Lost Property" and "Non-Criminal")

The city of San Francisco was then divided into a 20x20 grid, and each incident mapped to one of the grid squares. The total number of crime reports of each type were then calculated for each grid square.

Foursquare Venues Lists

For each grid square, the Foursquare API was called to retrieve the list of venues in that grid square across each of Foursquare's ten top level categories. The count of the venues of each type was thus calculated for each grid square.

Combined Dataset

The two datasets were then combined to produce a single table listing the number of incident reports (by type) and the number of Foursquare venues (by type) for each grid square.

Methodology

Results

Discussion

Conclusion