

Goal: Identify New York City Police precincts who are most unbiased (or biased) regarding the rate in which police “Stop/Question/Frisk” people.

Identify the procedural steps in providing an *analytics* solution for the goal stated above.

Step 1	Identify the best way to match the race levels between the two datasets
Step 2	Modify / delete race levels in each dataset so that levels match
Step 3	Identify which blocks / neighborhoods are patrolled by each precinct -- add precinct identifier for each block to census data
Step 4	Obtain the population proportion of each race for every precinct
Step 5	Obtain the total number of SQF instances for every precinct
Step 6	Compute the expected number SQF instances for each precinct
Step 7	<p>Compute a discrepancy measure for each precinct</p> $X^2 = \sum_{\text{races}} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$
Step 8	Rank diversity measure from smallest to largest – small values indicate less bias and large value indicate more bias

What elements of Task #1 are

Statistical in nature	Data Science in nature
<ul style="list-style-type: none"> • Matching race levels in such a way to minimize impact on analysis, i.e. reduction of bias, etc. • Obtain summaries for SQF and Census data to compute a discrepancy measure 	<ul style="list-style-type: none"> • Retrieve data • Incorporate precinct information (boundaries via shape files) into census data • Create necessary variables so that a diversity measure can be computed