Analysis Proposal
NYPD Criminal Summons, Historic

Question:

Over time, what does the composition of criminal court summons in NYC look like? Are

certain demographics issued more summons for certain violations? Have the number of

summons gone down over the last few years, or have they increased?

Dataset issues:

There are typos and null values in the dataset. On a value_counts(), I found that some of

the listed offense descriptions at the very end of the list (they only contained one value)

were either mistyped or incoherent, and there are quite a few NaNs scattered

throughout the dataset. Additionally, though the NYPD bucketed age categories,

someone began cataloguing "ages" that weren't ages at all, and were instead a random

string of numbers. Moreover, sometimes the person entering the data only entered

"OTHER ABC" or "ABC" as an offense type.

Plan:

Week of 3/23: Preliminary data cleaning and aggregation

Week of 3/30: Data analysis begins

Week of 4/6: Implementing statistical methods and choosing the best kind of data
manipulation

Week of 4/13: Research on context and best practices for this data

Week of 4/20: Continued analysis, begin preliminary visualizations

Week of 4/27: Visualization and narrative

Week of 5/4: Summary of insights, compiling documentation

Week of 5/11: Continue compiling documentation

Week of 5/18: Final report delivery