

 [casanave](#) / [time\\_series\\_stop\\_and\\_frisk](#) Public

☆ 0 stars    🍴 0 forks


☆ Star

👁 Unwatch ▼


- <> Code
- 🔍 Issues
- 🔗 Pull requests
- ▶ Actions
- 📁 Projects
- 📖 Wiki
- 🛡 Security
- 📈 Insights
- ⚙ Settings

🔗 main ▼

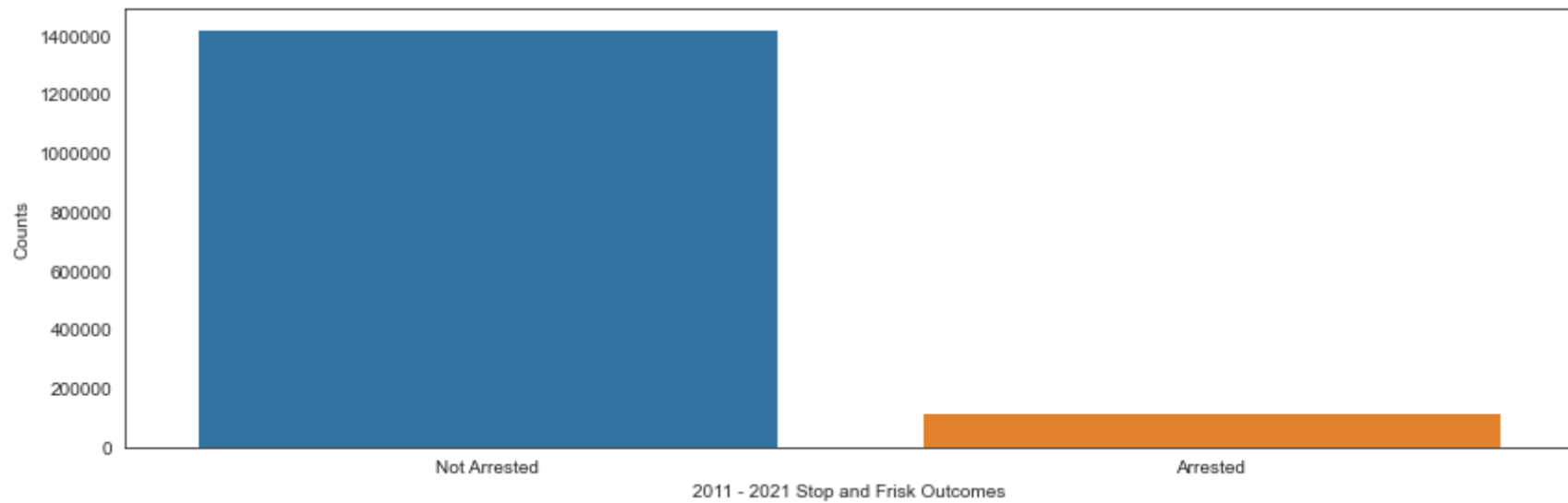
...

 **casanave** Delete 2011 - 2021 NYPD Stop and Frisks by Race, Precinct and Outcome...    ...    18 hours ago    ⌚ 31

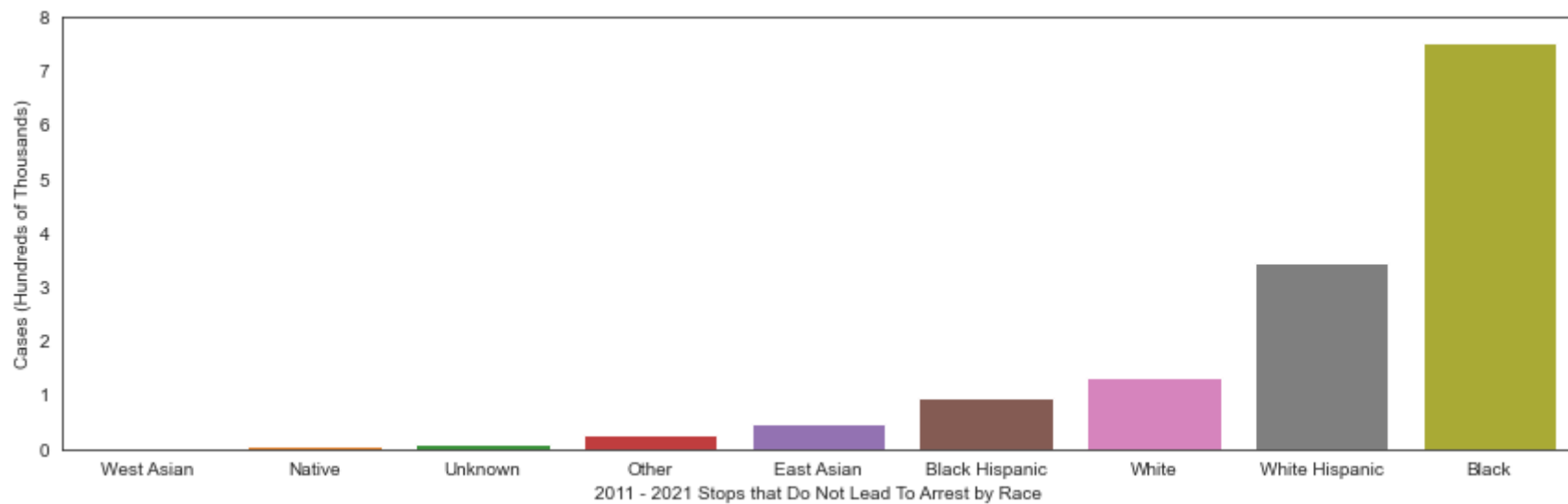
[View code](#)

README.md 

Problem : NYPD conducts too many stop-and-frisks that do not lead to an arrest. 92% of them in the last 10 years have not lead to arrest.



There's a racial disparity problem with stop-and-frisk, making it especially important that NYPD lowers the stops it does that do not lead to an arrest.



Hypothetical Business Case : NYPD, in partnership with the city council, has hired me, a data expert to do data analysis and use machine learning to try and limit stops that do not lead to an arrest. They are looking for actionable insights and recommendations.

Proposed Solution : A time series model which can predict seasonality of stop-and-frisks that do not lead to an arrest. NYPD can act by limiting stops during times when stops are unlikely to lead to arrest.

Why time series modeling with this data:

- Crime has provable seasonality[<https://pinkerton.com/our-insights/blog/the-seasonality-of-crime>]
- Time data doesn't contain demographic information. It's unethical to predict which demographics of people will commit future crimes.

Dashboards for Data Analysis:

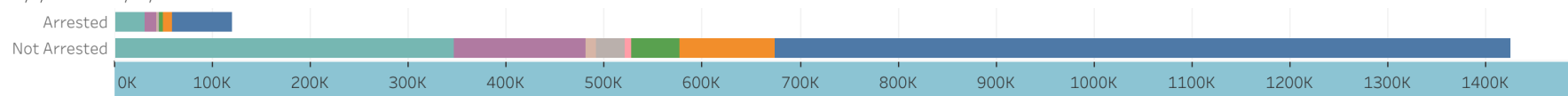
- 1,545,827 stops
- 10 Years Since 2011

Features: Precinct, Race, Outcome, Time

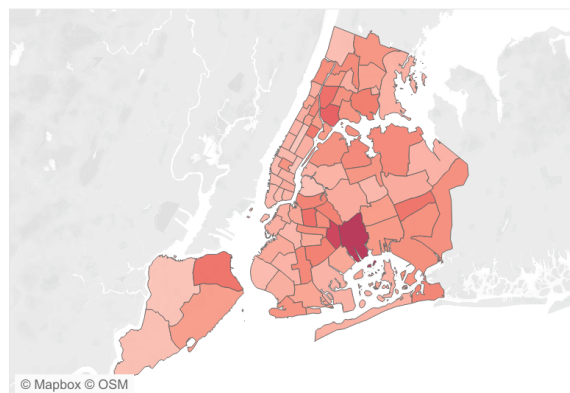
## Results:

Date

1/1/2011 to 12/31/2021

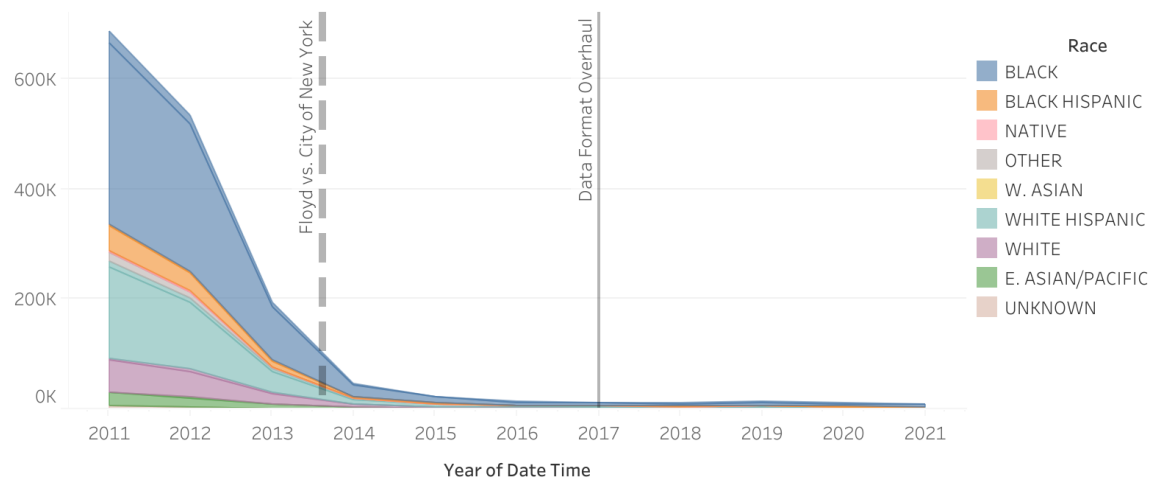


## NYPD Precinct



Cases

2,439 66,014



[[https://public.tableau.com/views/2011-2021NYPDStopandFrisksbyRacePrecinctandOutcome/2011-2021NYPDStopandFrisksbyRacePrecinctandOutcome?:language=en-US&:display\\_count=n&:origin=viz\\_share\\_link&:device=desktop](https://public.tableau.com/views/2011-2021NYPDStopandFrisksbyRacePrecinctandOutcome/2011-2021NYPDStopandFrisksbyRacePrecinctandOutcome?:language=en-US&:display_count=n&:origin=viz_share_link&:device=desktop)]

- 8,947 stops
- 2021 only

Features: Precinct, Race, Outcome, Time, Suspected Crime, Age, Physical Force Used

## Results:

## Race Description

All

## Suspect Sex

☒ FEMALE☒ MALE

## Physical Force Used

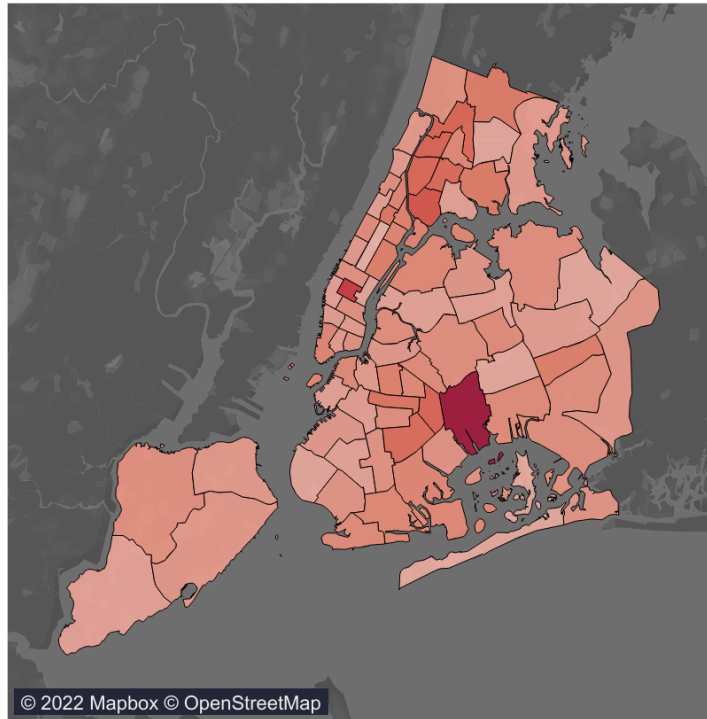
☒ UNKNOWN☒ TRUE

## Outcome

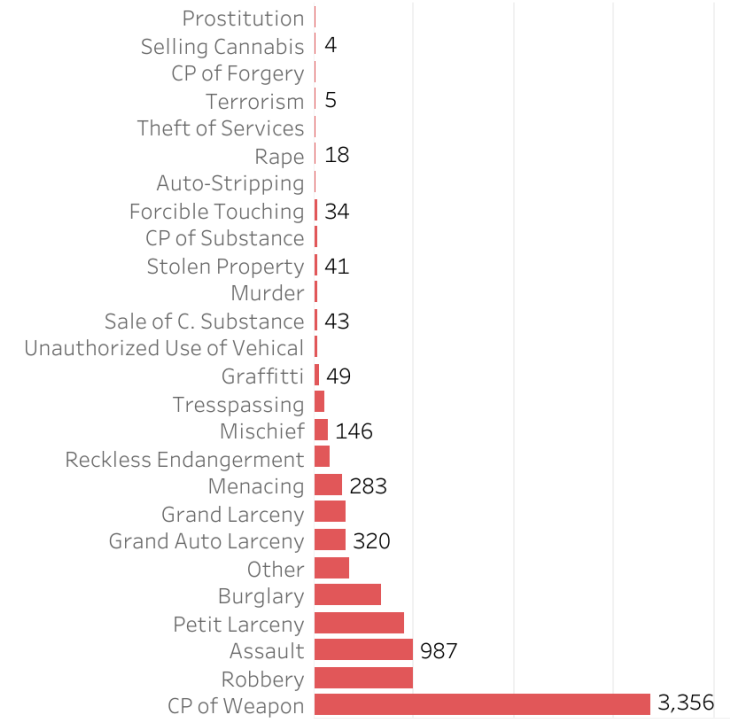
☒ ARRESTED☒ NOT ARRESTED\*

\* 'Not Arrested' contains data for both Not Arrested and Unknown Outcomes

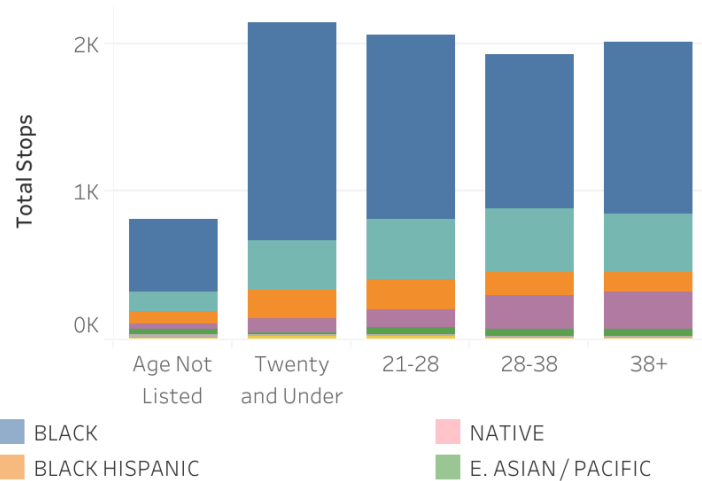
## Outcome by Precinct



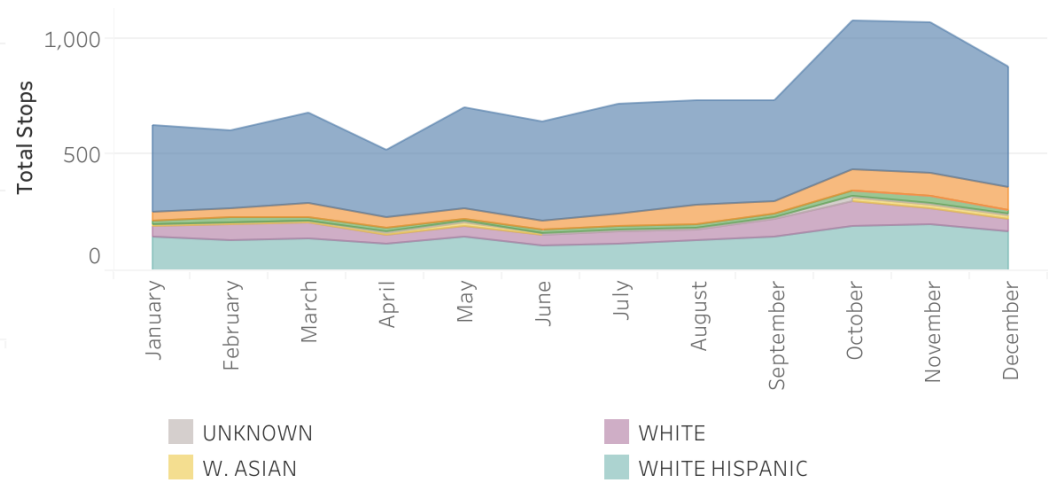
## Suspected Crime of Stopped



## Ages



## Racial Disparity Over Time



[[https://public.tableau.com/views/StopandFrisk2021NYPD/StopandFrisk2021DemographicOverview?:language=en-US&:display\\_count=n&:origin=viz\\_share\\_link&:device=desktop](https://public.tableau.com/views/StopandFrisk2021NYPD/StopandFrisk2021DemographicOverview?:language=en-US&:display_count=n&:origin=viz_share_link&:device=desktop)]

## MODELING METHODS:

- Predict 6 month's rate of stops that do not lead to an arrest based on monthly resample from 2014-2021
- Target: Last 6 months of data
- Machines ranked by MAE score

## Machines Used:

- Naive
- Random Walk
- ARI Model
- IMA Model
- SARIMA Model

## Tuning Methods:

- ACF and PCF Plots to hypothesize term maximum
- AutoArima to iterate over SARIMA model to test different combinations of terms

## MODELING RESULTS:

NAIVE MODEL: 2% off on average

RANDOM WALK MODEL: 10% off on average

ARI MODEL: 10% off on average

IMA MODEL: 10% off on average (rounded slightly lower than other non-naive models)

SARIMA MODEL: 10% off on average

## MODELING CONCLUSIONS:

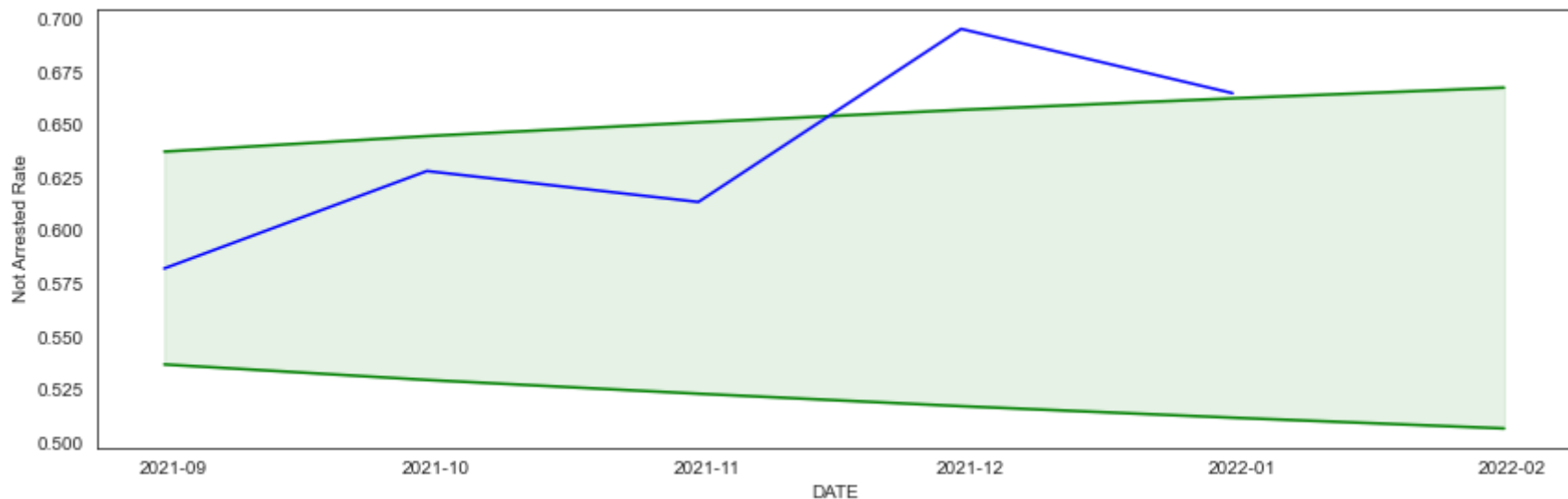
Sophisticated modeling does NOT offer predictive capabilities for spikes in not-arrest rates only based on timing of stop.

Sophisticated modeling DOES offer anomaly detection inferentially for when spikes in not-arrest rates are historically unprecedented.

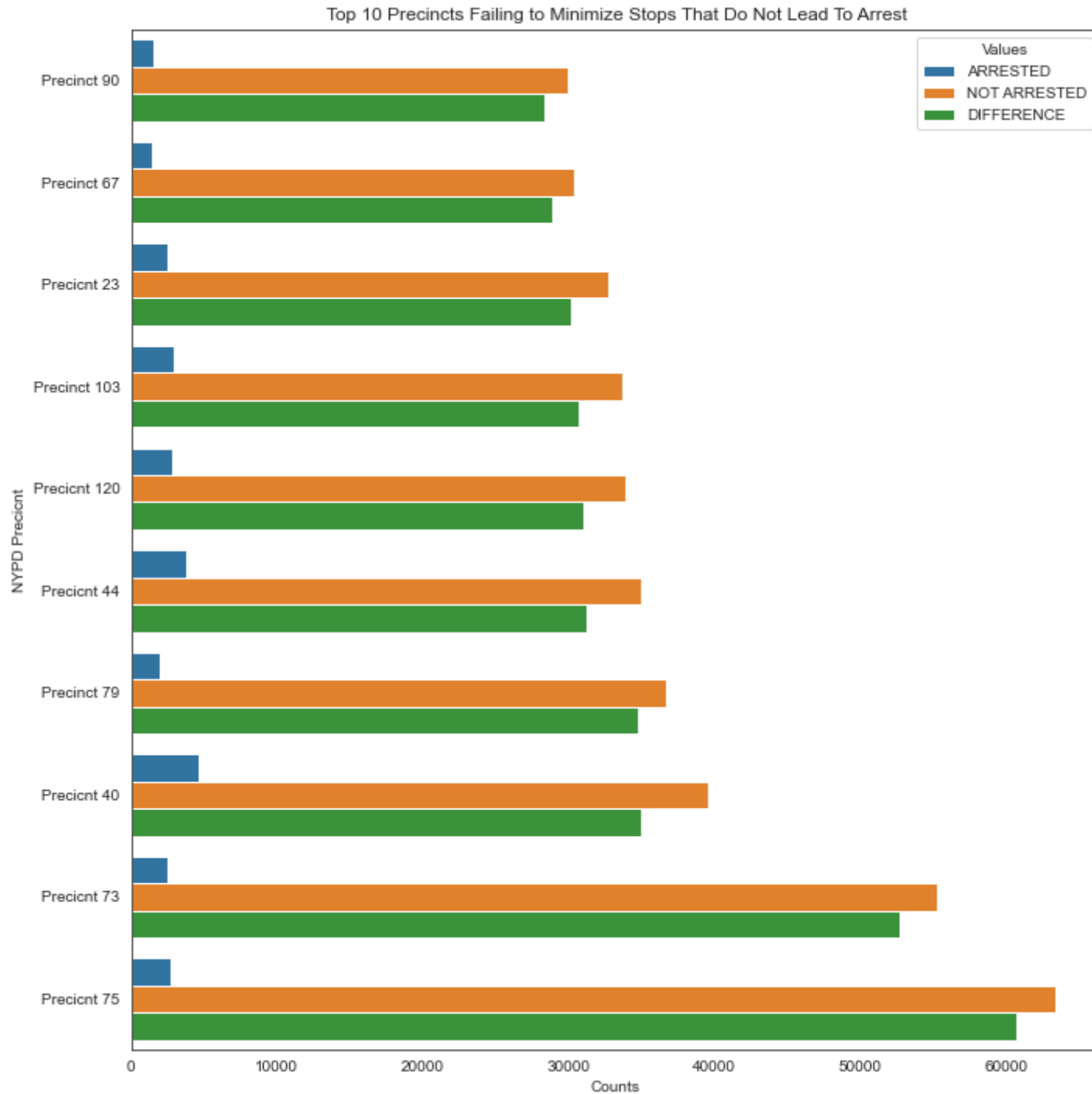
Naive modeling's success is plain evidence of NYPD needing to be more receptive to when last month's stops did not lead to arrest and do less stops during those months as a resonable reaction.

## RECOMMENDATIONS TO NYPD:

- Do less stops if not-arrested rates spiked last month
- Use IMA model to recognize outliers as is shown below



- Retrain the following 10 precinct where the difference is biggest between stops that do and don't lead to arrest:





**\*\*FUTURE WORK: \*\***

- Try adding more variables to make SARIMAX model (esp interested in Precinct and Suspected Crime)
- Try a categorical model without using time as a factor

Stop and Frisk Data: <https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page>

Precinct geojson file: <https://www1.nyc.gov/site/planning/data-maps/open-data/districts-download-metadata.page>

```
|— README.md
|— Stop_and_Frisk_Timeseries_Notebook_Monthly.ipynb      *** MAIN NOTEBOOK
|— SAF_Presentation.pdf                                  *** PRESENTATION SLIDES
|— images
|   |— NYPD_budget.webp
|   |— 2011_2021_SAF_Outcomes.png
|   |— NYPD_budget.webp
|   |— nypd_badge.png
|   |— Not_Arrested_By_Race.png
|   |— seasonal-bars-1200.webp
|   |— 2021_SAF_Overview.png
|   |— 2011-2021_SAF_Overview.png
|   |— Top_10_Precincts_Failing_To_Minimize.png
|   |— IMA_in_action.png
|— raw_data
|   |— SF_2011.csv.gz
|   |— SF_2012.csv.gz
|   |— SF_2013.csv.gz
|   |— SF_2014.csv.gz
|   |— SF_2015.csv.gz
|   |— SF_2016.csv.gz
|   |— SF_2017.csv.gz
|   |— SF_2018.csv.gz
|   |— SF_2019.csv.gz
|   |— SF_2020.csv.gz
|   |— SF_2021.csv.gz
```

```
├── tableau_data
│   ├── precincts.json
│   ├── 2021_analysis.ipynb
│   ├── 2021_data.csv
│   └── ten_years_tableau_data.csv
├── new_data
│   ├── stop_and_frisk_no_race.csv
│   └── stop_and_frisk_w_race.csv
├── non_main_notebooks
│   └── Crimes_Notebook
├── Stop_and_Frisk_Timeseries_Data_Prep.ipynb
├── Stop_and_Frisk_Timeseries_Notebook.ipynb
├── .DS_Store
├── .gitattributes
└── .gitignore
```

\*\*\* FURTHER WORK

\*\*\* DATA PREP NOTEBOOK

\*\*\* FURTHER WORK

## Releases

No releases published

[Create a new release](#)

## Packages

No packages published

[Publish your first package](#)

## Languages

● Jupyter Notebook 100.0%