Andrew Trick
Data Analyst, Project Lead

FORECASTING CHICAGO CRIME

Increasing Schedule Efficiency and Lowering Crime Through Predictive Policing and Data-Driven Resource Planning

BACKGROUND

Chicago PD personnel resources are projected to cost around $1.5b for 2019

Crime also costs taxpayers in legal, administration, and incarceration fees

Crime rates tend to occur in regular, consistent cycles every year

BUSINESS VALUE

Predicting crime for scheduling has several benefits

Reduce unit response time

Increase efficiency of PD Scheduling

Deter crime through predictive policing

Data-driven scheduling boost efficiency and saves money

Companies typically save between 5-10% in cost from traditional methods

Santa Cruz saw significant results from a similar project

19% drop in burglaries

4% decrease in motor theft

DATA OVERVIEW

CHICAGO DATA PORTAL

- Hosted and updated daily online through Chicago Data Portal
- Complete with no personal data or privacy concerns
- Salesforce API connectivity
- Over 7m entries w/ 24 features
- Each row is one reported incident

DATA WAREHOUSING

- MySQL DB for relational storage
- Secure, Private, and Safe
- Easy integration into AWS
- Open Source = Free

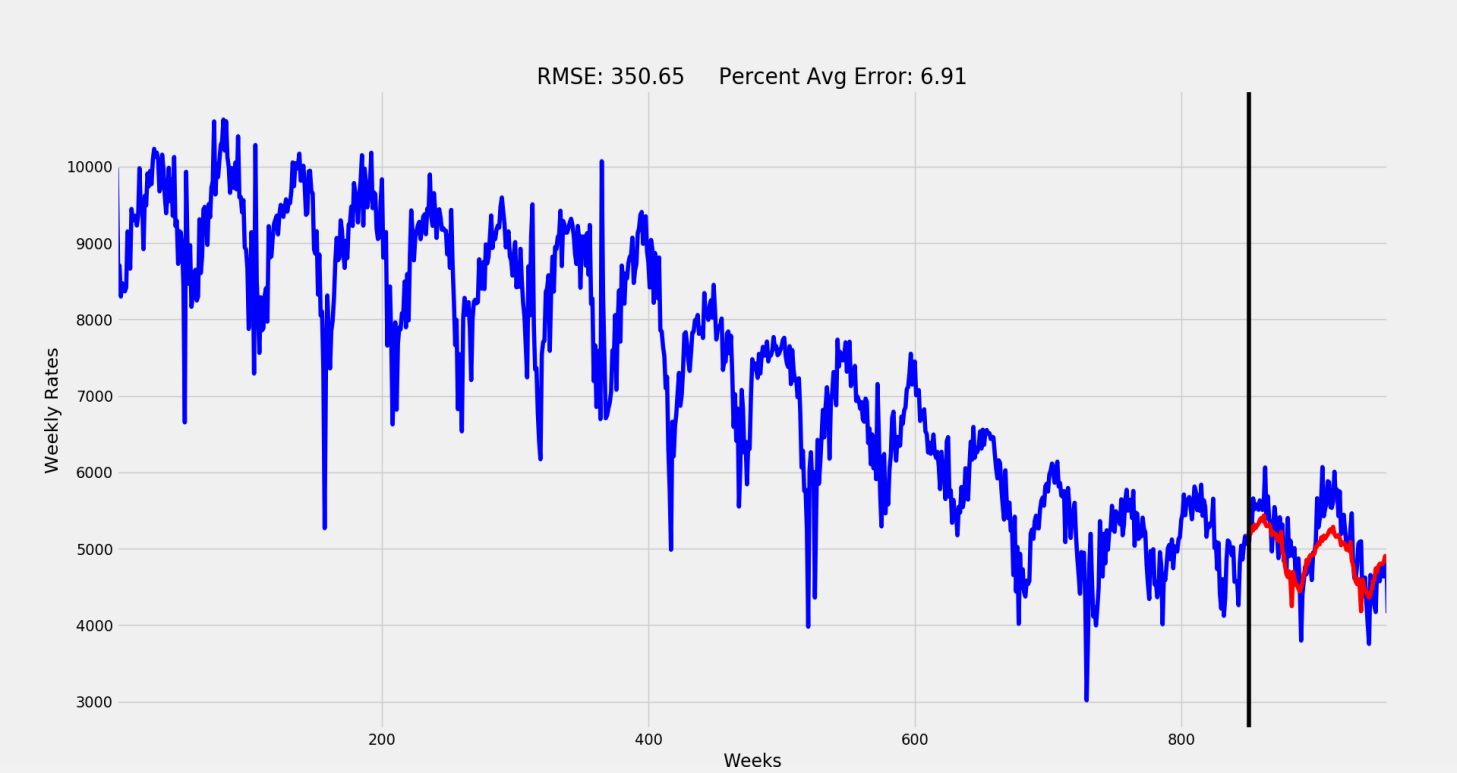
ANALYTIC MODEL

PROPOSED ANALYTIC MODEL

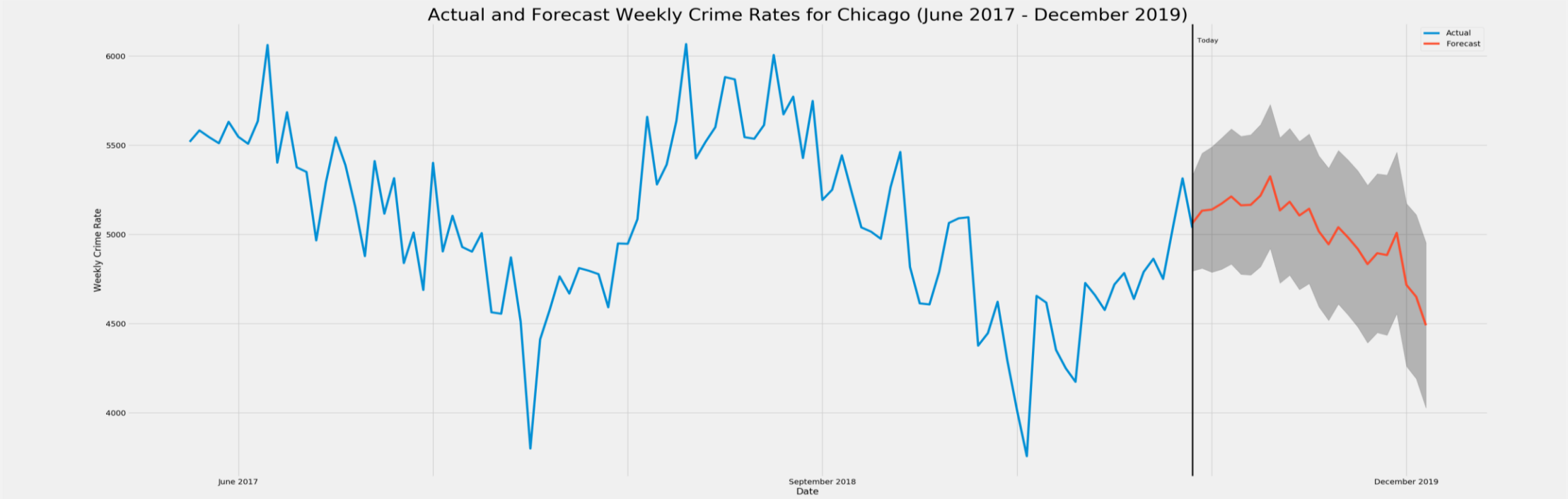
- Seasonal Auto-Regressive Integrate Moving Average (SARIMA) Model
- Best Fit: SARIMA(1, 1, 2)(1,0,1)52

  Average Error: 6.9%, RMSE: 351
- Best Model given Trend, Cycle, and Seasonality of data

Time-Series Trend and Forecast Results



MODEL: HOW TO USE

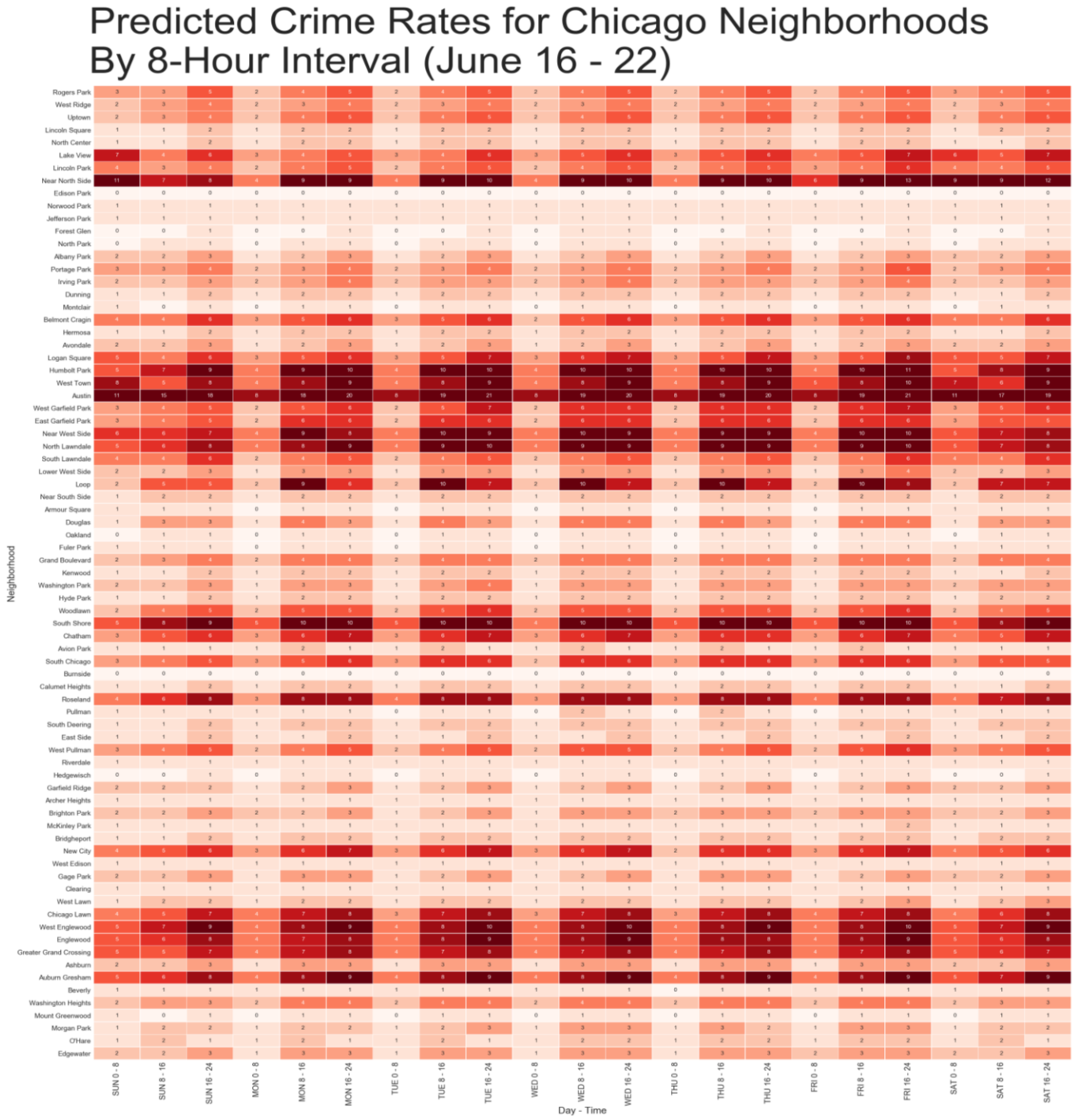


PILOT PROGRAM RESULTS: LONG-TERM

- Six month outlook for projected overall crime rates based on historic trends
- Time-series forecaster estimated overall weekly rates at 94.2% accuracy

PILOT PROGRAM RESULTS: NEAR-TERM

- Comprehensive look at weekly crime rates (neighborhood by time)
- Highlights peak crime segments at ~92% avg accurate per matrix cell

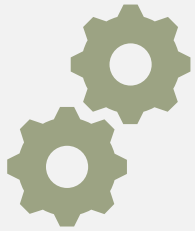


DEPLOYING THE SYSTEM

DEPLOYMENT PROCESS: OVERVIEW

Pilot program is full-fledged version and ready for implementation

Migrate pilot from local DB into MySQL warehouse in AWS RDS environment



Automated pipeline available for ETL, modeling, visualization, and updating

DEPLOYMENT PROCESS: TOOLS

Amazon RDS Warehousing

Salesforce API connectivity

Python 3 environment to run automated pipeline

Modules: Pandas, numpy, matplotlib, requests, seaborn, statsmodels, mysql-connector-python

DEPLOYMENT PROCESS: REQUIREMENTS

ONE DATA SCIENTIST

ONE WEB/DB DEVELOPER

ONE DOMAIN EXPERT

32GB AWS RDS MYSQL SERVER

INTERNAL APP / WEBSITE

DEPLOYMENT PROCESS: TIMELINE

Three Weeks to Deploy!

- One week to prepare AWS environment
- Two weeks to set up app/website
- One week for pipeline migration and testing`

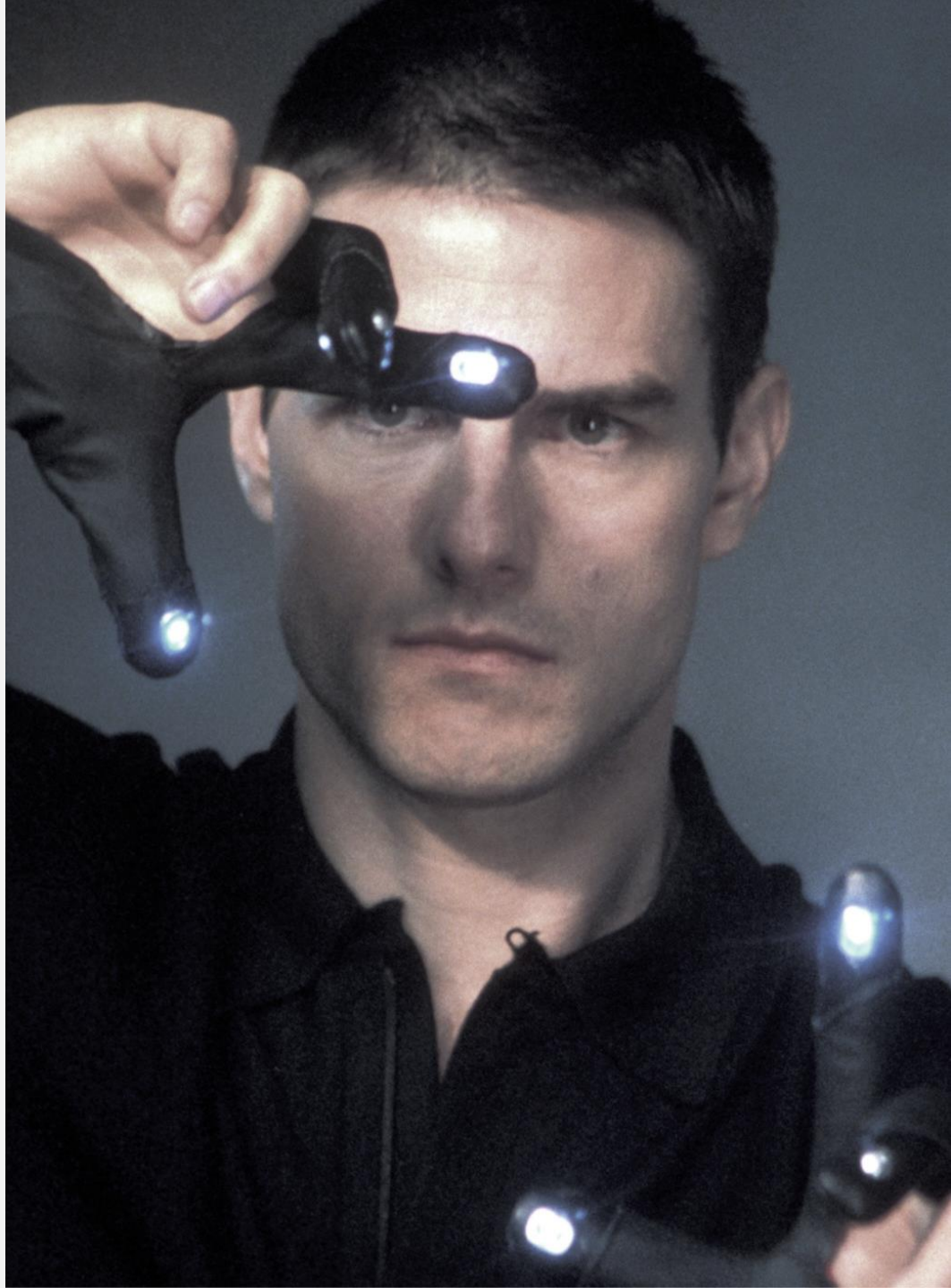
FINAL CONSIDERATIONS

RECOMMENDATIONS

Leverage data-driven insights to predict and deter crime

Plan personnel according to forecast

Refer to heatmap to schedule patrols



POSSIBLE RISKS

Difficulty migrating automated system into live environment

SOLUTION:
Ensure Data Scientist's understanding of automated pipeline

Changes to model terms needed to remain accurate

SOLUTION:
Evaluate and refit model every quarter.

Outliers abound and should be expected

SOLUTION:
"All models are wrong but some are useful"
– George E. P. Box

REASONS TO IMPLEMENT

Boost

Boost scheduling and patrol route efficiency

Discourage

Discourage crime through focused presence where and when needed

Save

Save taxpayers by:

- Lower PD personnel costs
- Fewer criminal trials and legal fees
- Less damage to city and personal property

RESOURCES

Live Pilot Program:
http://andrewtrick.com/pages/chicago-crime.html

Github Repo for Project:
https://github.com/leaflettuce/chicago_crime

PDF version of Presentation:
http://andrewtrick.com/pages/CC_pres.html

REFERENCES

- Amazon. (n.d.) Amazon RDS for MySQL. *AWS Products*. Retrieved from https://aws.amazon.com/rds/mysql/
- City of Chicago. (2019). 2019 Budget Ordinance. Retrieved from https://www.chicago.gov/content/dam/city/depts/obm/supp_info/2019Budget/2019BudgetOrdinance.pdf
- Friend, Zach. (2013). Predictive Policing: Using Technology to Reduce Crime. *FBI: Law Enforcement Bulletin*. Retrieved from https://leb.fbi.gov/articles/featured-articles/predictive-policing-using-technology-to-reduce-crime
- MySQL. (2019). MySQL Community Edition. *MySQL Products*. Retrieved from https://www.mysql.com/products/community/
- Nau, Robert. (n.d.) Identifying the Number of AR and MA terms in an ARIMA model. *Duke University*. Retrieved from https://people.duke.edu/~rnau/411arim3.htm
- Rotageek. (n.d.) Five Long Term Benefits of Data-Driven Scheduling. *Rotageek*. Retrieved from https://blog.rotageek.com/blog/five-long-term-benefits-of-data-driven-scheduling-1