LA Robbery Analysis & Time Series Prediction

Team Outliers: Jennifer Zhao, Charles Bi, Bowen Liu, Jingfan Lu

Robbery Data Quick Facts

Robbery is...

12th

Most frequent crime in Los Angeles among **134** crime types

2nd

Most frequent juvenile-committed crime and **2.5%** of robbery were committed by juvenile

1.58 Days

reporting delay, compared to **16.4** days delay for all crimes and **3.3** days for battery

Strong-Arm

51% of the robbery are committed using strong-arm, other methods include pepper spray & stick

7 - 9 pm

Peak hour for robbery and **3x** more than early morning

Saturday

Robbery peaks on weekend, which has **12%** more robbery cases than weekdays on average

Airport

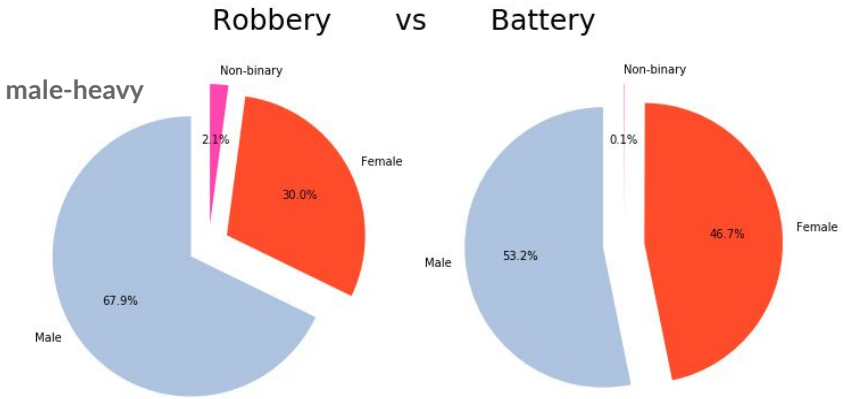
several robbery hubs are also areas close to airports (i.e., Burbank & Inglewood)

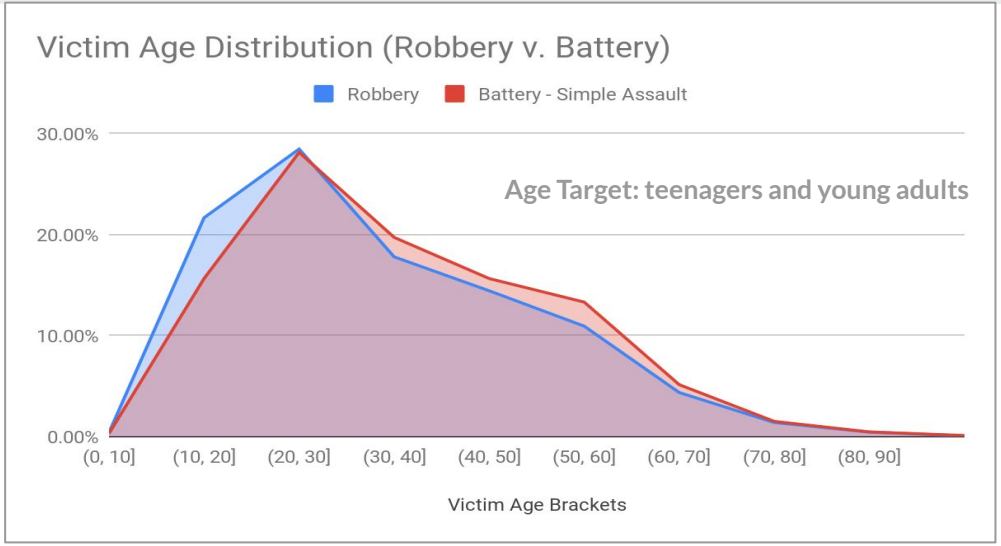
0.88

The occurrence of robbery and battery are highly correlated by area

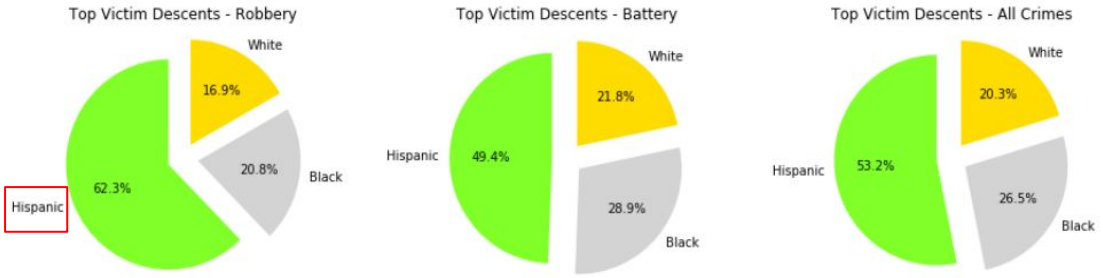
Robbery v. Battery - Victim Profile

Victim Sex Breakdown

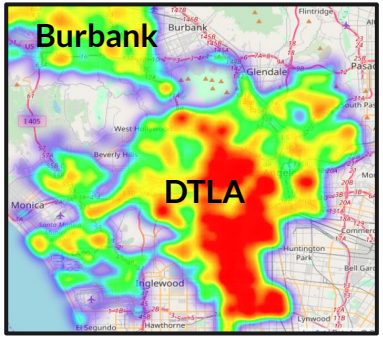




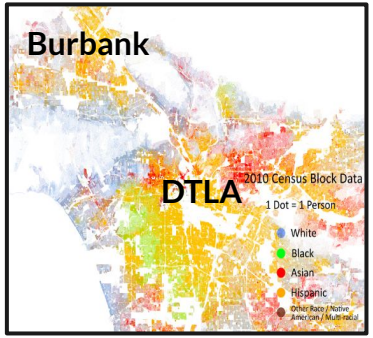
Descent Analysis: robbery 10% more victims of hispanic descent



Hispanic Victim Heatmap

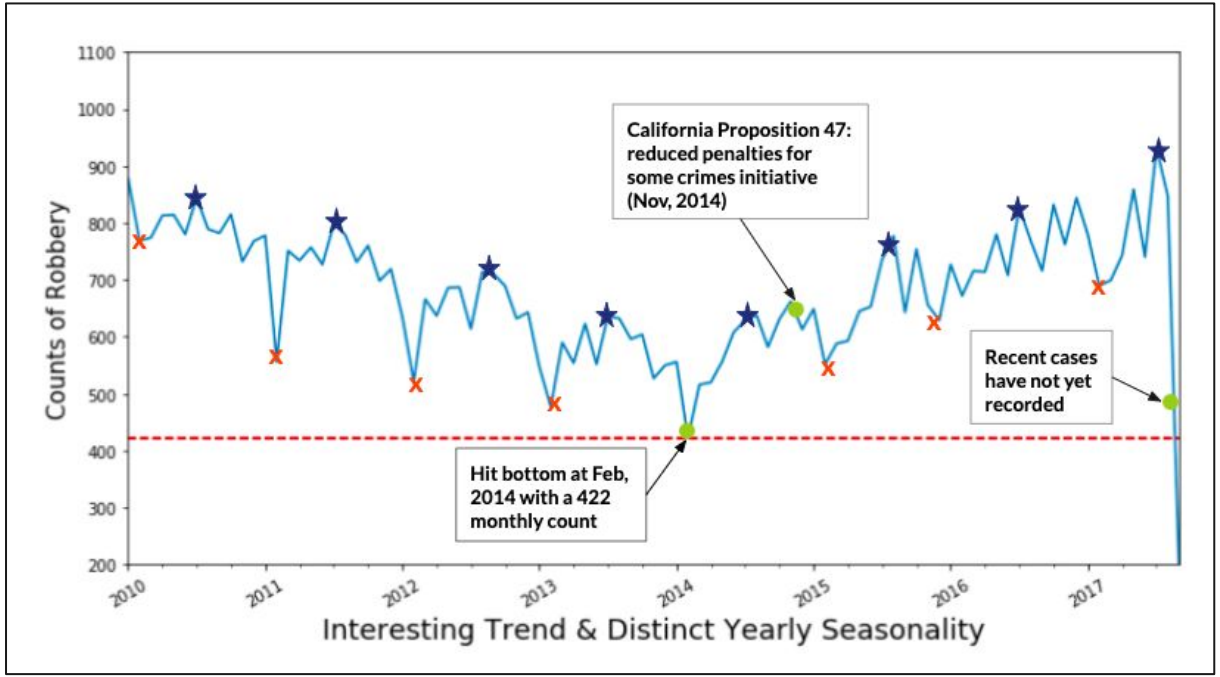


LA Demographics Map



Trend of Robbery 2010 - 2017

Throughout the years 2010 to 2017, the amount of robberies went down before year 2014; that may be due to the decrease in unemployment rate. However, the amount went up starting 2015 after California passed proposition 47.



Consistent Yearly Seasonality

★ : peak in a year (summer)

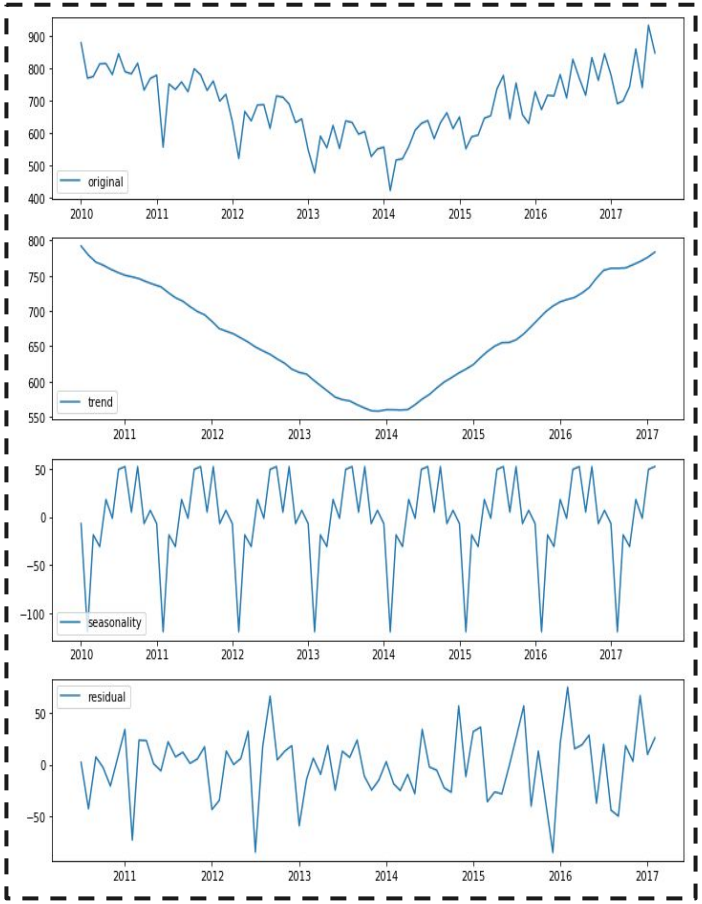
X : bottom in a year (winter)

Overall Trend & Probable Explanations

1. **2010 - 2014 drop**
   - > Unemployment rate decreased from 13% to 8%
   - > Unsheltered homeless population dropped by 15%

2. **2014 - 2017 rise**
   - > California Proposition 47 (decriminalize
   - > Unsheltered homeless increased 20k

Time Series Modeling (visualize & stationarise)

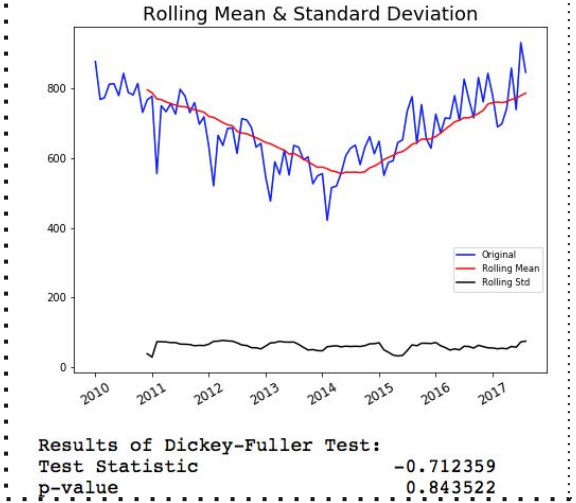


⬅️1️⃣ Visualize the time series:

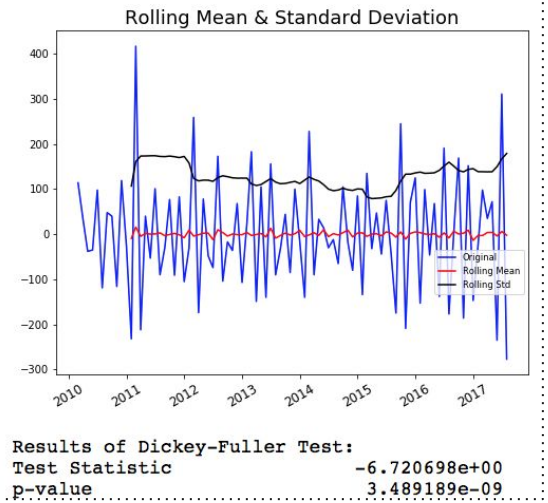
- V-trend in the data
- Consistent y early Seasonality

⬇️2️⃣ Stationarize the series:

- All its statistical properties(mean, variance) remain constant over time (the mean of the series should not be a function of time
- Stationarize by taking second differencing order

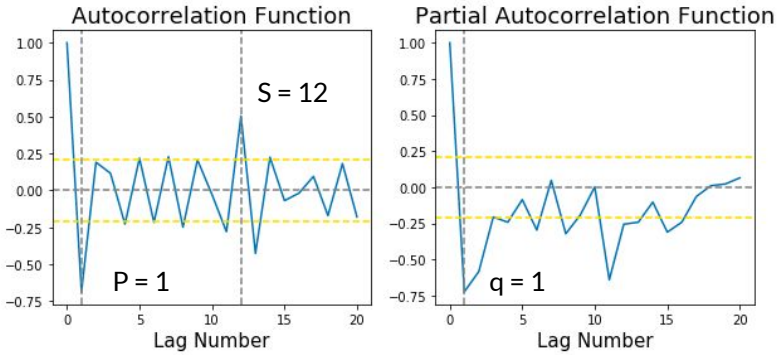


Before: non-stationary



After: stationary

Time Series Modeling (selection & prediction)



⬅️ 3️⃣ Parameter testing: plot ACF/PACF charts to find optimal parameters

⬇️ 4️⃣ Build the SARIMA model

```
model = sm.tsa.statespace.SARIMAX (timeseries, order = (1,2,1), seasonal_order = (1,0,0,12))
```

➡️ Prediction 5️⃣