Arrests Report to the Public Safety Committee

Analysis conducted by Lyle Lalunio

October 24, 2022

1 Introduction

The primary motivation of this paper is to respond to a request from the Council's Public Safety Committee. We were tasked to analyze NYPD arrests between 2018 and 2022 to better understand the general trends of arrest rate in New York City. Specifically:

1. Has the arrest rate been decreasing from 2018-2022?
2. What are the top 5 most frequent arrests in 2018-2022?
3. If we think of arrests as a sample of total crime, is there more crime in precinct 19 (Upper East Side) than precinct 73 (Brownsville)?
4. Given the available data, what model would you build to predict crime to better allocate NYPD resources? What challenges do you foresee?

In gaining a better understanding of our past, we can evaluate our present laws and policing measures to better inform future policy changes. Ultimately, it is our hope that the insights in this report help fulfill both the Public Safety Committee and NYPD's mission to create better, safer communities for the citizens of New York City.

2 Data Description

The data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning before being posted on the NYPD website. The full dataset is obtained from combining the historical dataset (2006-2021) with the year-to-date (2021-YTD) dataset. In total, it contains 5,590,004 arrests in NYC by the NYPD from January 1, 2006 to September 30, 2022. The data was last updated on October 19, 2022. For each arrest, there are 19 attributes, which can be placed in one of three categories:

- Information about the arrest (Arrest Key, Date, Borough, Latitude/Longitude)
- Descriptions and details of the committed crime (Text descriptions, Level of offense, Law Codes and Police Codes, Jurisdiction)
- Demographics of the perpetrator (Age, Sex, Race)

For the intents and purposes of this analysis, we restrict the time of arrests to fall between January 1, 2018 and September 30, 2022. We removed 45 records that lacked any information on the offense. After filtering and determining there were no duplicate records based on the ARREST_KEY field, our final dataset contained 897,376 arrests, with offense descriptions of 34 records labeled as "UNKNOWN." It is important to note that some laws have been repealed between 2018 and 2022, most significantly marijuana-related laws (NY Penal Law Article 221). In total, there are 16,500 arrests for repealed laws, which have been left in.

External datasets for population were obtained from the US Census Bureau. Monthly population estimates of NYC were interpolated using 1-year surveys from the American Community Survey (ACS). Officer data was obtained from publicly available datasets on the NYPD website. Special thanks to John Keefe for his processing of precinct-level data available on his GitHub.

3 Has the arrest rate been decreasing from 2018-2022?

The arrest rate for New York City on January 2018 was 0.00279, or 279 per 100,000. Since then, there has been a slight but steady decline, bottoming in July 2020 at 89 per 100,000 before slightly increasing back up to 160 per 100,000. Notably, between the February and March of 2022, there was a sharp 20% rise from 160 to 200, where it sits near today.

The overall downward trend is captured by the red trendline in Figure 1, which has a statistically significant negative slope (Table 1). When translated back into arrests per hundred thousand, it states an average annual decrease of 21.1 ($\pm$ 6.6).

The steady decrease in crime is also reflected on the yearly scale (Figure 2), with the arrest rate declining by approximately 31% since January 2018 to December 2021. It should be noted that the current annual arrest rate displayed for 2022 is only based on the first 9 months of the year, and uses the same population numbers from 2021.

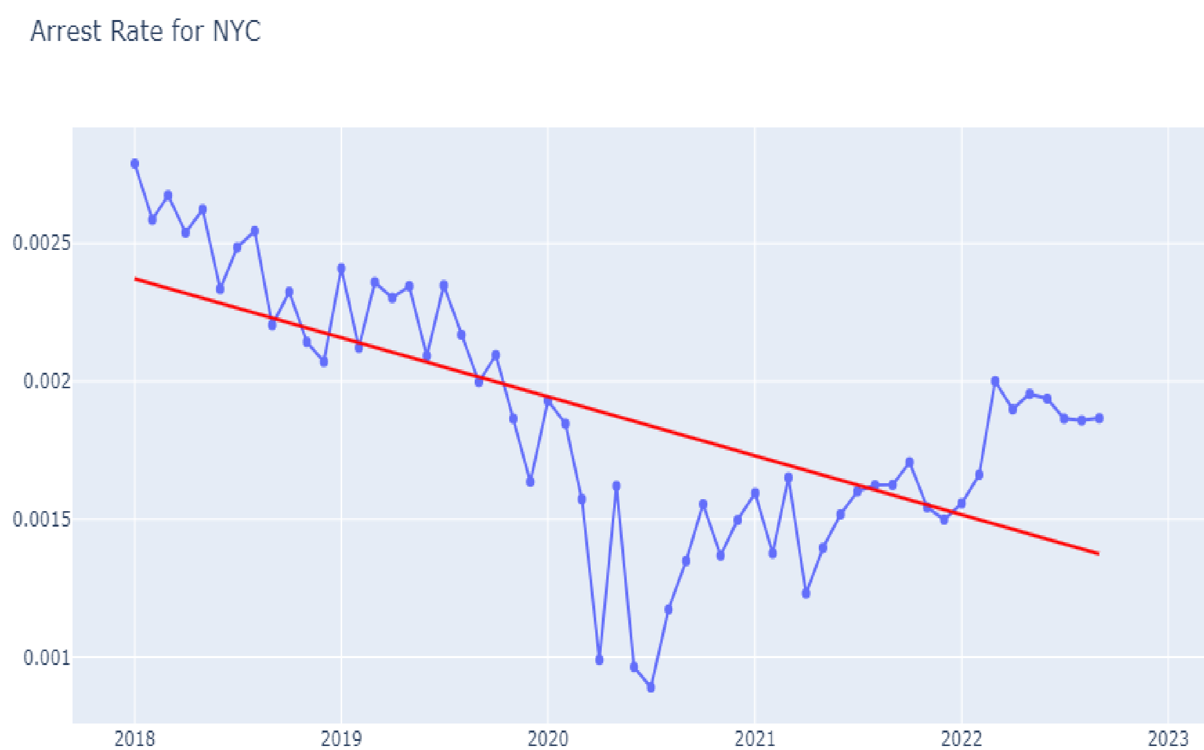

Figure 1. Arrest Rate Regressed on Time (monthly)

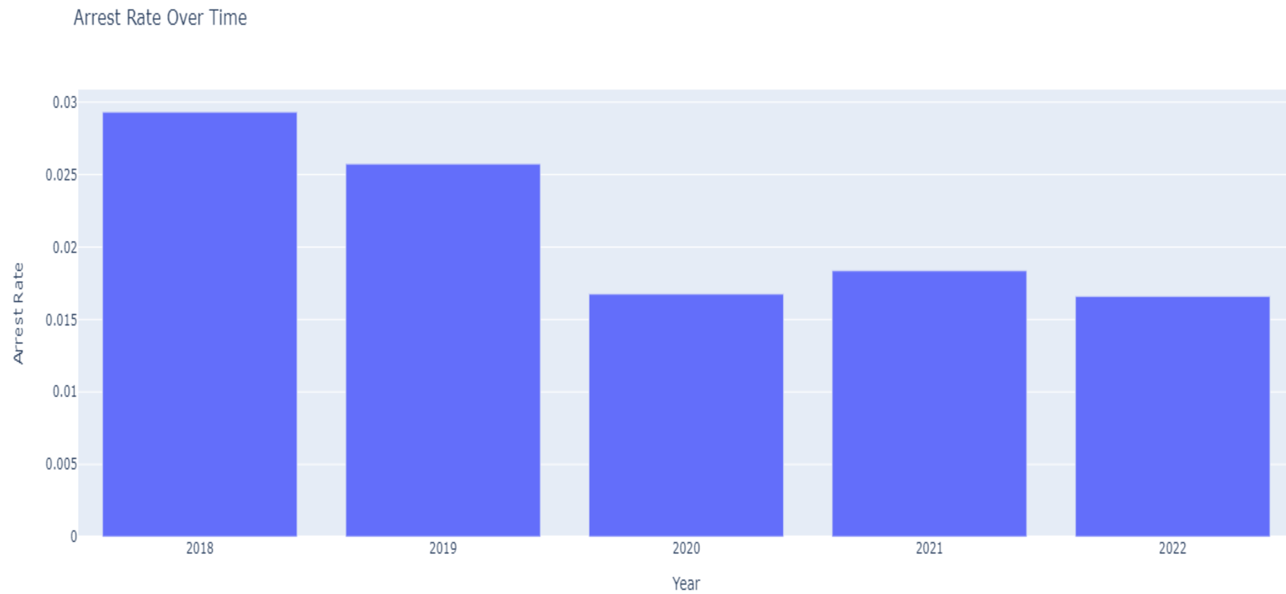

Figure 2. Arrest Rate (Annualized)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.418
Model:                            OLS   Adj. R-squared:                  0.408
Method:                 Least Squares   F-statistic:                     39.56
Date:                Tue, 25 Oct 2022   Prob (F-statistic):           5.45e-08
Time:                        02:34:38   Log-Likelihood:                 373.39
No. Observations:                  57   AIC:                            -742.8
Df Residuals:                      55   BIC:                            -738.7
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0126      0.002      7.382      0.000       0.009       0.016
x1         -6.778e-12   1.08e-12     -6.290      0.000   -8.94e-12   -4.62e-12
==============================================================================
Omnibus:                        7.703   Durbin-Watson:                   0.438
Prob(Omnibus):                  0.021   Jarque-Bera (JB):                6.903
Skew:                          -0.802   Prob(JB):                       0.0317
Kurtosis:                       3.575   Cond. No.                     5.83e+10
==============================================================================
```

Table 1. Arrest Rate Regressed on Time Summary Statistics

4 What are the top 5 most frequent arrests as described in the column 'pd_desc' from January 2018 to September 2022?

In order from most frequent to least frequent, it is 3rd degree Assault Misdemeanors, Petit Larceny from Open Areas, 1st/2nd/Other Types of Assault, Unclassified Traffic Misdemeanors, and Unclassified Robbery in open areas (Table 2). The term "Unclassified" often indicates an offense which violates more specific clauses within an article or section, although there are a few exceptions (Table 2). Note, aside from 76 cases of 2nd degree Menacing, all assault charges under the "1st/2nd/Other Types of Assault" description are felonies, i.e., more serious charges than Assault 3.

The 5 most frequent arrests mirror the general trend of arrests overall (Figure 3): substantial declines starting from 2018, bottoming in 2020, and slightly rising back up again in 2021 and 2022. Traffic misdemeanors saw the largest decline from 18,494 arrests in 2018 to just 4,908 in 2021. In that same time period, Petit Larceny declined over 50% from 23,378 to just 11,374. Assault misdemeanors declined a modest 25%, while more serious assault charges stayed essentially the same. In fact, 1st/2nd degree assault and robbery charges - which are mostly felonies - have seen the smallest changes over time.

	Offense Description	Count	Unique Law Codes
0	ASSAULT 3	105271	9
1	LARCENY,PETIT FROM OPEN AREAS,UNCLASSIFIED	84693	1
2	ASSAULT 2,1,UNCLASSIFIED	55421	93
3	TRAFFIC,UNCLASSIFIED MISDEMEANOR	45741	47
4	ROBBERY,UNCLASSIFIED,OPEN AREAS	41203	30

Table 2. Top 5 most frequent arrests, counts, and number of violated law codes

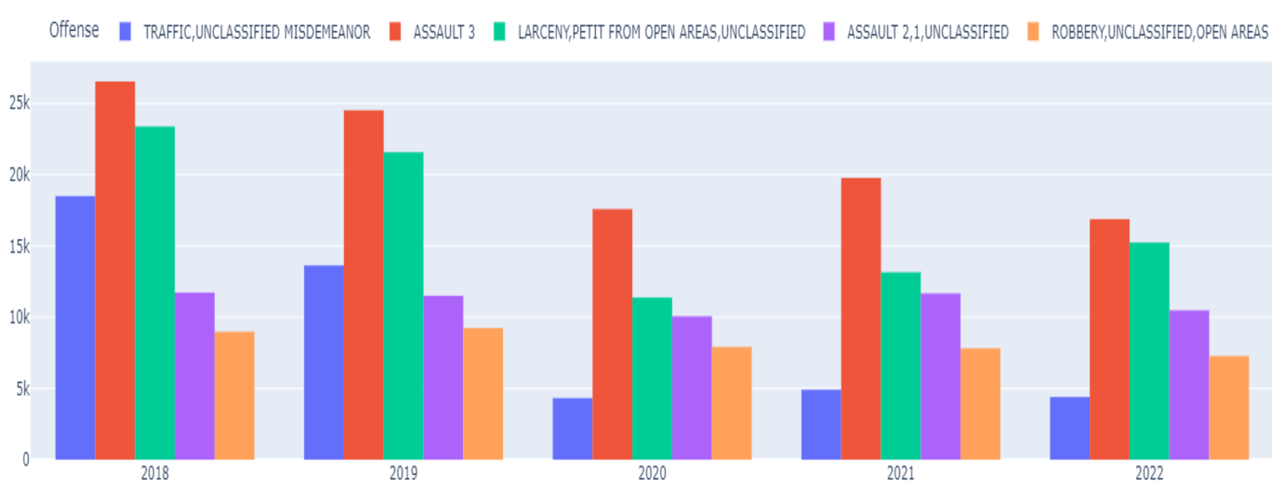

Figure 3. Top 5 most frequent arrests and counts over time

5 If we think of arrests as a sample of total crime, is there more crime in precinct 19 (Upper East Side) than precinct 73 (Brownsville)?

- Couple of ways to think about this:
- Absolute number of arrests
- Rate relative to population of precincts
- Number of arrests relative to number of law enforcement officers in precinct

There are nearly twice as many arrests in precinct 73 than 19, yet less than half of the population, explaining the widened disparity between the two lines (Table 3). Adjusting for the 60% more officers in Precinct 73 than precinct 19 shrinks, but does not eliminate the disparity in crime levels between the two especially in 2018 (Table 3, Figure 4c).

In terms of trends, both precincts experienced statistically significant arrest reductions between 2018 and 2021. However, Precinct 73 has reduced crime at nearly four times the rate of Precinct 19 (Chart 4a, 4b). Numerically, Precinct 19's trendline states an annual decrease in arrests of 52, and Precinct 73 near 13 arrests.

Both precincts reached their lowest levels of crime during the middle of 2020 (the start of the COVID-19 Pandemic). Since then, arrests have slowly climbed back up, with approximately 50% spikes in arrests for both precincts on August-September 2021 and February-March 2022 (Charts 1,2).

In terms of NYPD officer effectiveness, both precincts averaged above one arrest per officer prior to the COVID-19 pandemic, with Precinct 73 at times nearly double that of 19. However, there has been a steady decline in officer effectiveness over time, stagnating in 2021 at around 0.58-0.65 arrests per officer. While it again increases in 2022, the number of arrests per officer do not reach the same levels as prior years at similar crime rates (compare July 2019 and July 2022). This may be an indication that the current strategies employed by NYPD officers are losing its effectiveness.

	Total Arrests	Population	Officers
Precinct 19 Total Arrests (2018-2022)	9464	220261	241.0
Precinct 73 Total Arrests (2018-2022)	17983	98506	353.0

Table 3. Total arrests, population counts, & officer counts for precincts 19, 73

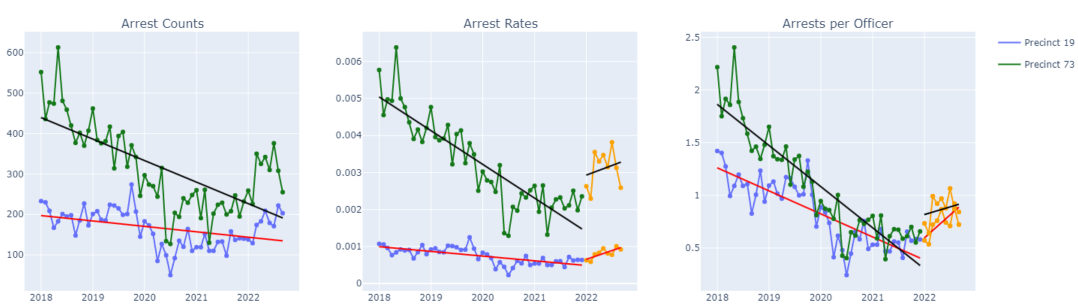

Figure 4. (a) Arrest Counts regressed on time (b) Arrest Rates regressed on time (c) Arrests per Officer regressed on time

6 Given the available data, what model would you build to predict crime to better allocate NYPD resources? What challenges do you foresee?

What first comes to mind is using a linear regression model (possibly combined with an ARIMA model). Important features include the time, hotspots within precincts (using the latitude/longitude coordinates), offense, and perhaps the arrestees' demographics. The target variable could be arrests or arrest rates, but these don't take into account the severity of the crime committed. Assuming data was available, I'd be interested at predicting how many total years were sentenced in a precinct. Then, with those predictions, we can inform the allocation of proportionate amounts of officers in those areas. Our preliminary ARIMA prediction model indicates that the previous two months were most informative of the next month's arrest counts (Figure 6).

Ideally, I envision a feedback loop between our predictions, NYPD's response, and our next batch of predictions. In this way, we can better evaluate the effectiveness of NYPD's strategies and resources, and potentially discover new ones. In a way, it resembles how portfolios and investment strategies are being increasingly algorithmically managed and discovered in the field of finance.

Glaring logistical challenges include convincing the NYPD to trust the model's predictions, and managing the redistribution of officers. Challenges in building the model include the usual suspects: data sourcing (due to availability or privacy issues), precise forecasting (Figure 5, Table 4), validating statistical assumptions (stationarity, normality, heteroskedasticity), and remaining relatively white-box.

Most importantly, we want to build a model that does not infringe on any individual's rights, nor reinforce societal biases (as AI has carelessly done in the past). Keeping models interpretable help us understand our models enough to keep this from happening, and operating on the precinct level is broad enough to avoid infringement on individuals yet granular enough to enact and effect change that people can perceive.

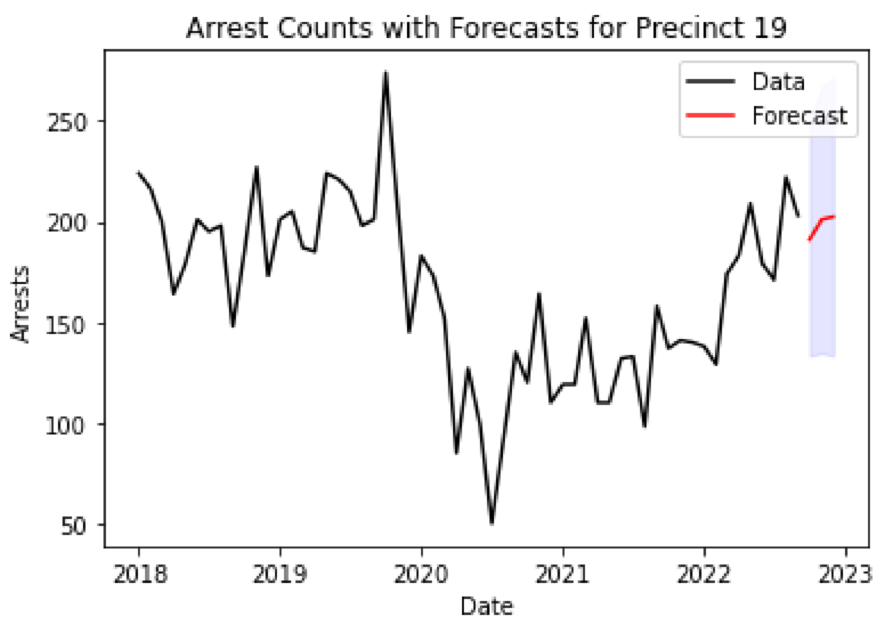

Figure 5. Arrest Counts for Precinct 19 charted with ARIMA(2,1,1) forecasts

7 Appendix

	lower arrests	upper arrests	forecast
2022-10-01	142.570861	239.430771	191.000816
2022-11-01	145.377554	256.504480	200.941017
2022-12-01	144.325065	260.241195	202.283130

Table 4. Out-of-sample predictions for Arrest Counts in 2022 4th Quarter

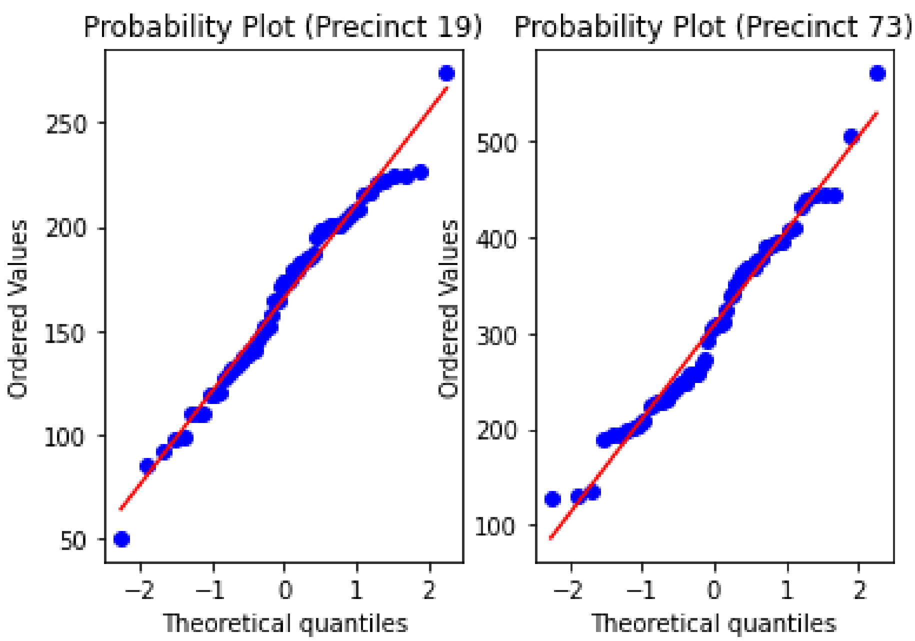

Figure 5. Q-Q Probability Plots of Precinct-level Arrest Counts

	p-values
Raw	0.46402
1-Differenced	0.00000
OLS Residuals	0.62650

Table 5. p-values of Augmented Dickey-Fuller stationarity tests

Actually, 1-differencing the data was sufficient to achieve stationarity. Will include a differencing=1 parameter in ARIMA model.

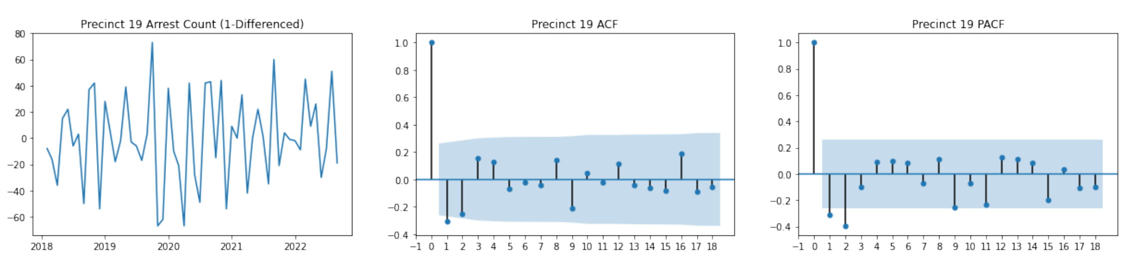

Figure 6. (a) Arrest Counts after 1-Differencing; (b) ACF Plot for 1-Differenced Arrest Counts; (c) PACF Plot for 1-Differenced Arrest Counts

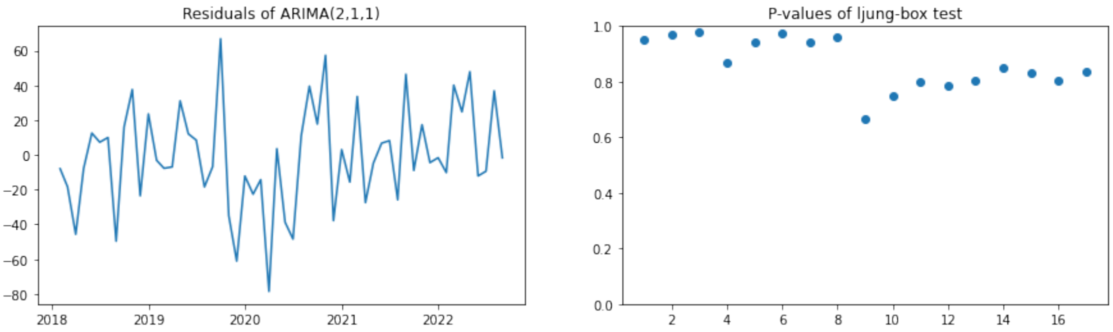

Figure 7. Model Diagnostics for Fitted ARIMA(2,1,1) Model