

# MCU movie prediction

Lilian Gao



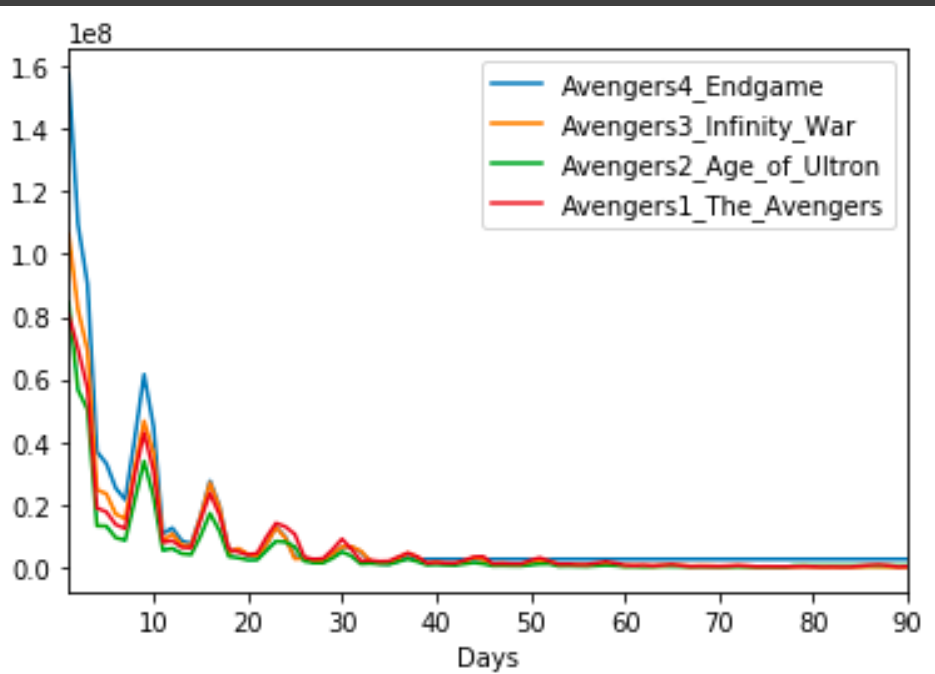
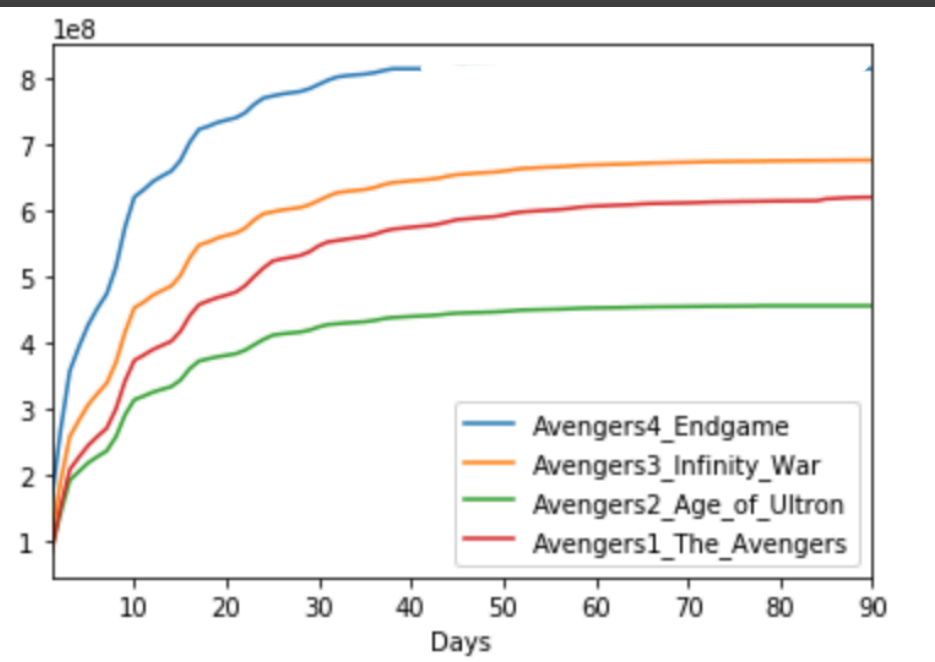
- Web scraping Html with Python & BeautifulSoup
- Daily box office performance of all MCU movies
  - Within 90 days after the opening
  - Domestic only
- Original table includes 8 columns

Date	Rank	Gross	% Change	# Theaters	Per Theater	Total gross	Days
------	------	-------	----------	------------	-------------	-------------	------

# Data Collection

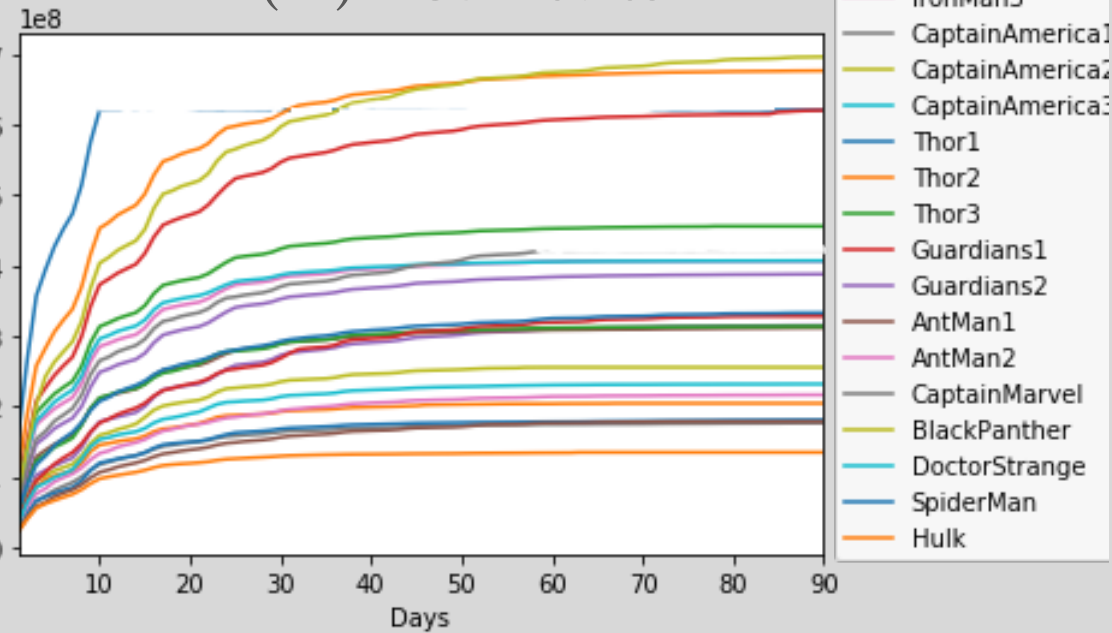
“<https://www.the-numbers.com/movies/franchise/Marvel-Cinematic-Universe>”

# Data Exploration

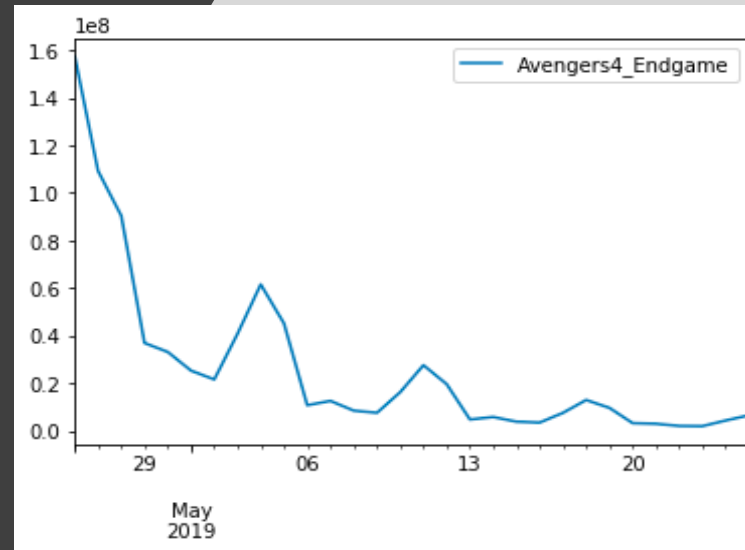
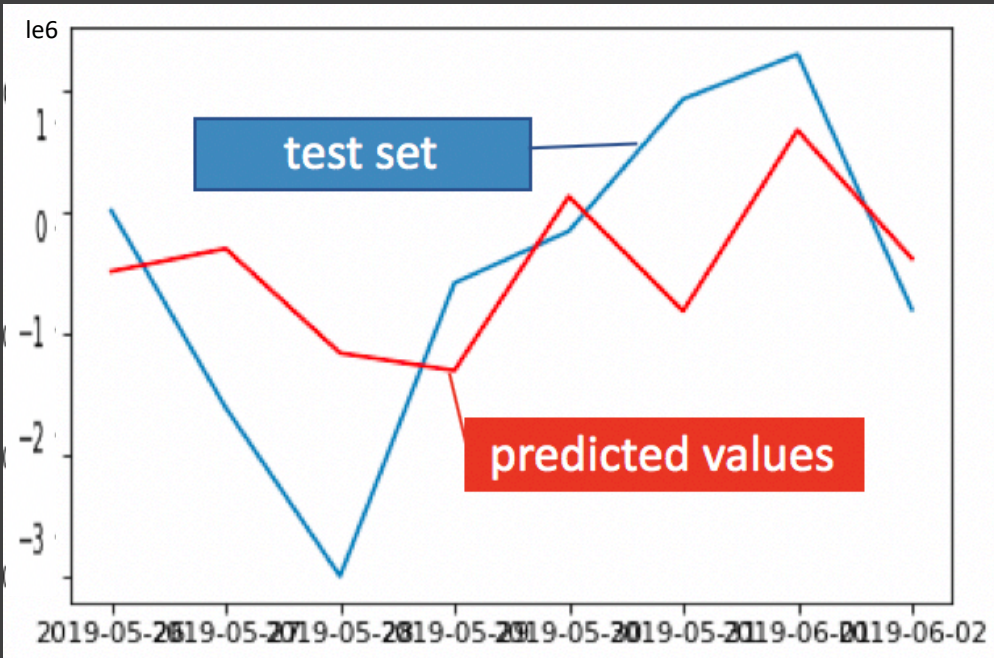
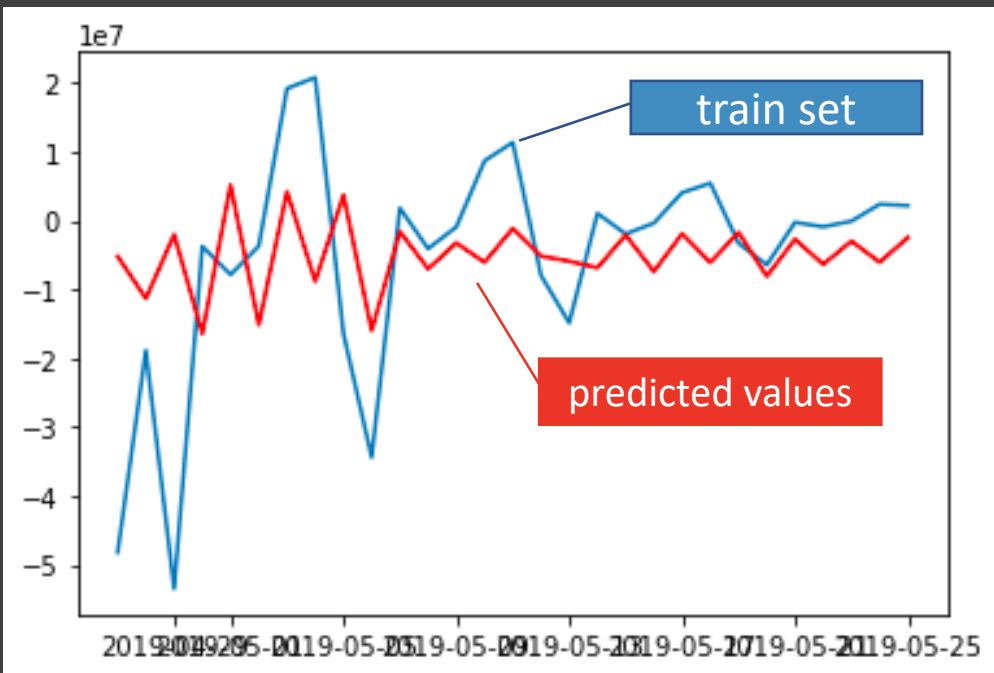


cumulative box office, domestic (M)

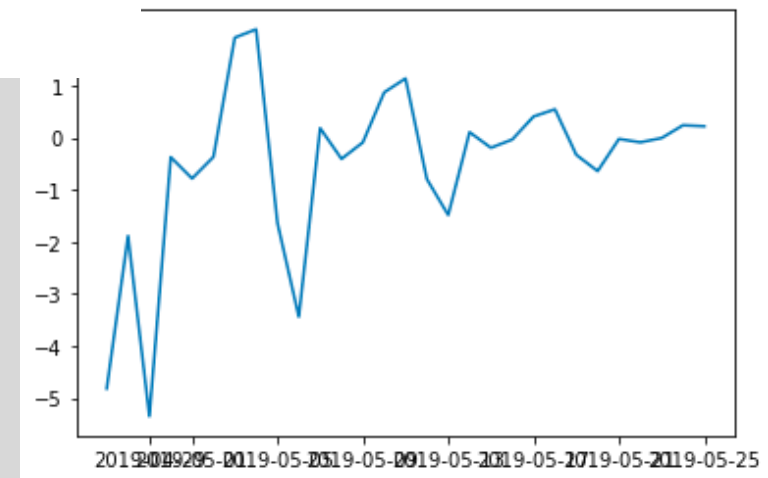
All (22) MCU movies



# Model Fitting: Time Series ARIMA (1,1,1)



Differencing:  
Remove trend  
& seasonality



- Amazon EC2
- Input: Days to forecast starting from 5/26/2019  
Output: Predicted daily gross
- ```
$ curl -H "Content-Type: application/json" -X POST -d '{ "days": "7" }' http://ec2-3-17-205-24.us-east-2.compute.amazonaws.com:5000/end_game
```
- ```
{ " dgross " : { "5968707.7", "5672830.9", " 4516931.9", "3216635.9", "3350855.1", "2541035.2", "3220791,1" }}
```

# Deployment

- Potential prediction models
  - More variables
  - Change of pattern after 30 days
- Performance of current model
  - On other previous MCU movies
  - On upcoming movies

# Future Improvements