Exploring U.S. National Parks

Time Series & Recommendations

By Jacky Lu

Project Motivation & Focus

1) Time Series Model

When do people visit National Parks? Can I predict National Park visitor counts?

2) Content-Based Recommender

If I like this national park, what other national parks are similar?

Workflow

National Parks website & API

Time series & Park activities

EDA on Yosemite

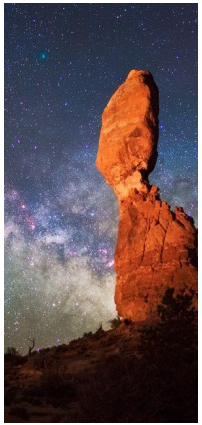
UI with model & recommendations

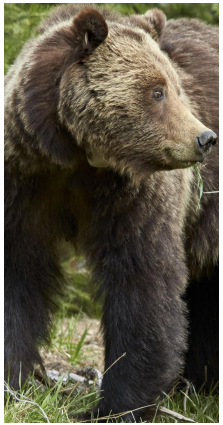
Tools: Beautiful Soup, jupyter, matplotlib, scikit learn, Streamlit

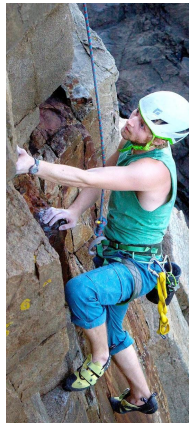
Data Collection

Monthly National Park Recreational Visits

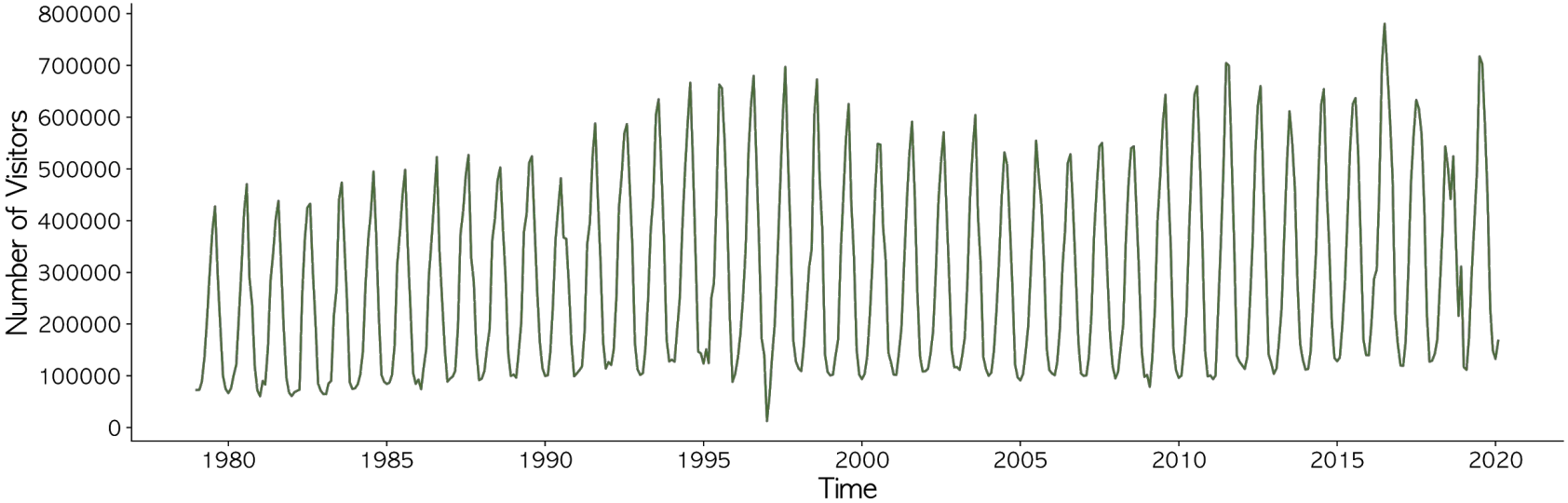
National Park Topics & Activities



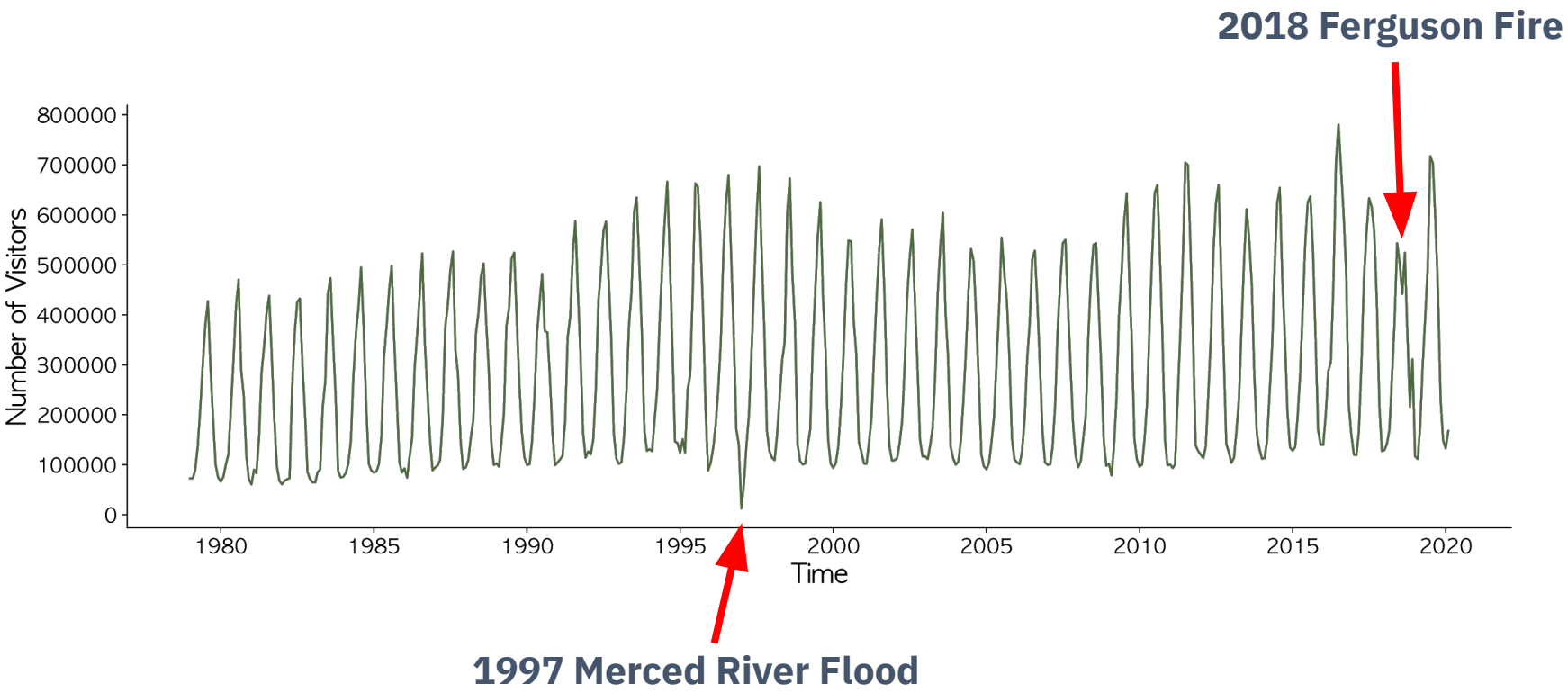




Yosemite Visitor Counts Show Seasonality

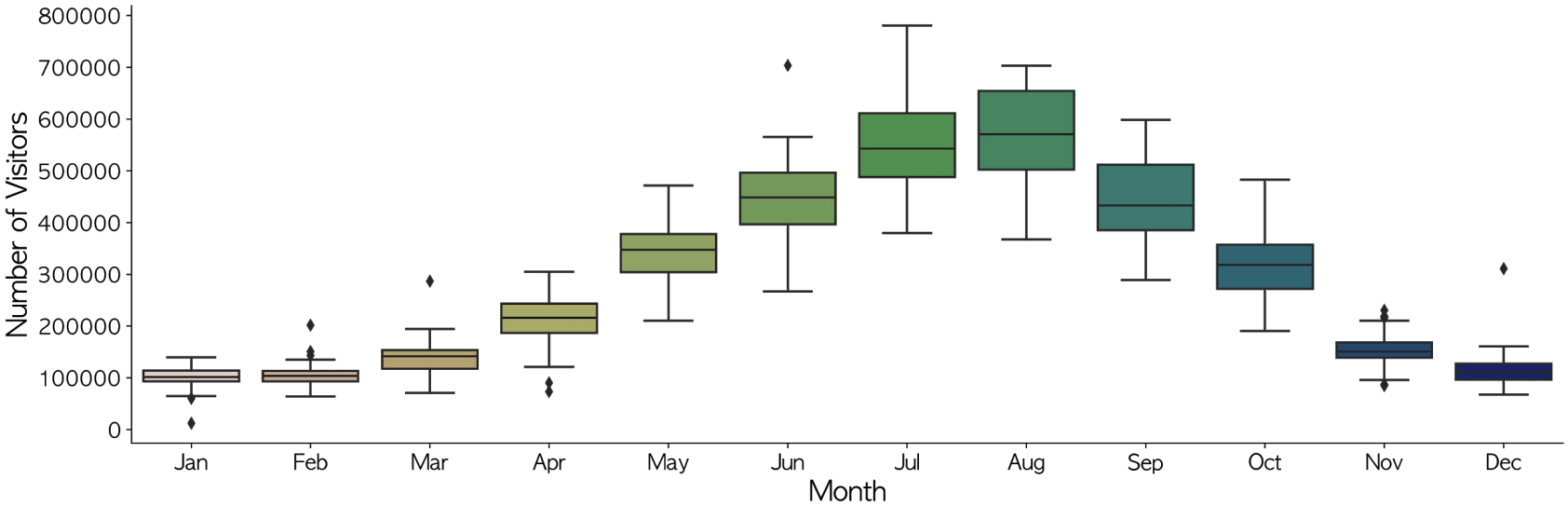


Investigating Dips in Visitors



Yosemite Visitor Monthly Trend

Yosemite has the most visitors in summer and fewest visitors in winter



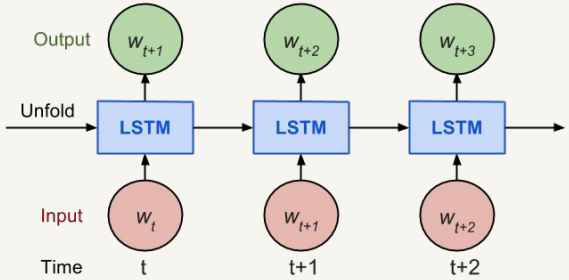
Time Series Modeling

Tools Used:

pmdarima 1.6.1

PROPHET

Keras



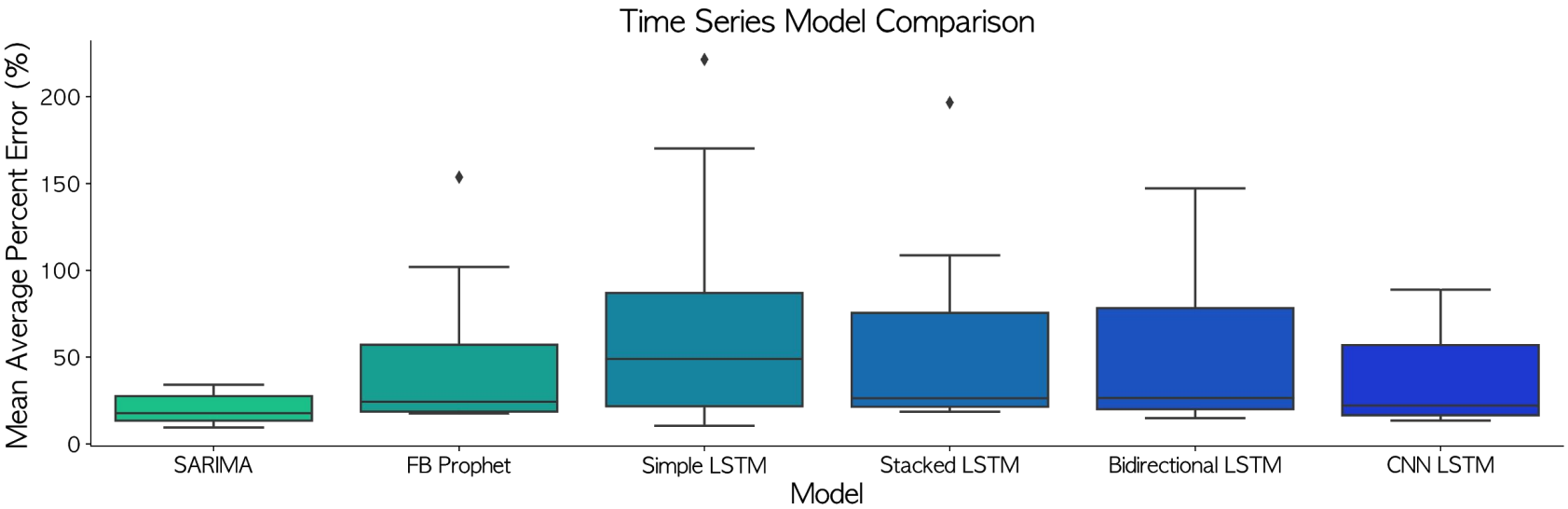
Time Series Models

- ARIMA/SARIMA
- FB Prophet

Neural Net Models

- LSTMs
  - Simple
  - Stacked
  - Bidirectional
  - CNN

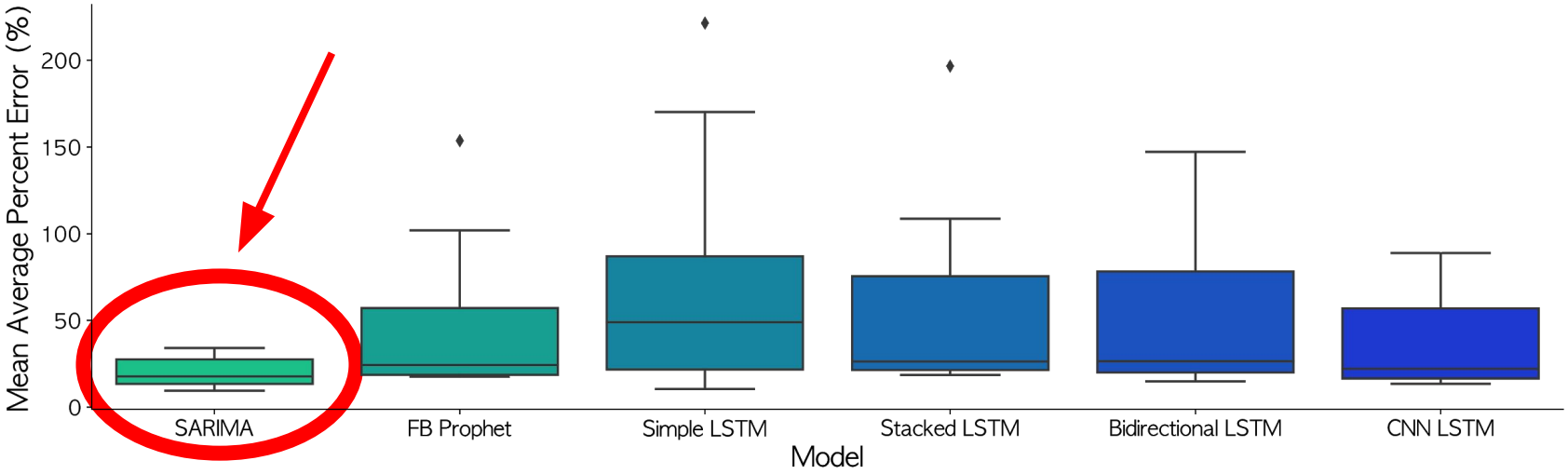
Comparing Models



Models compared using visitor data from 10 national parks

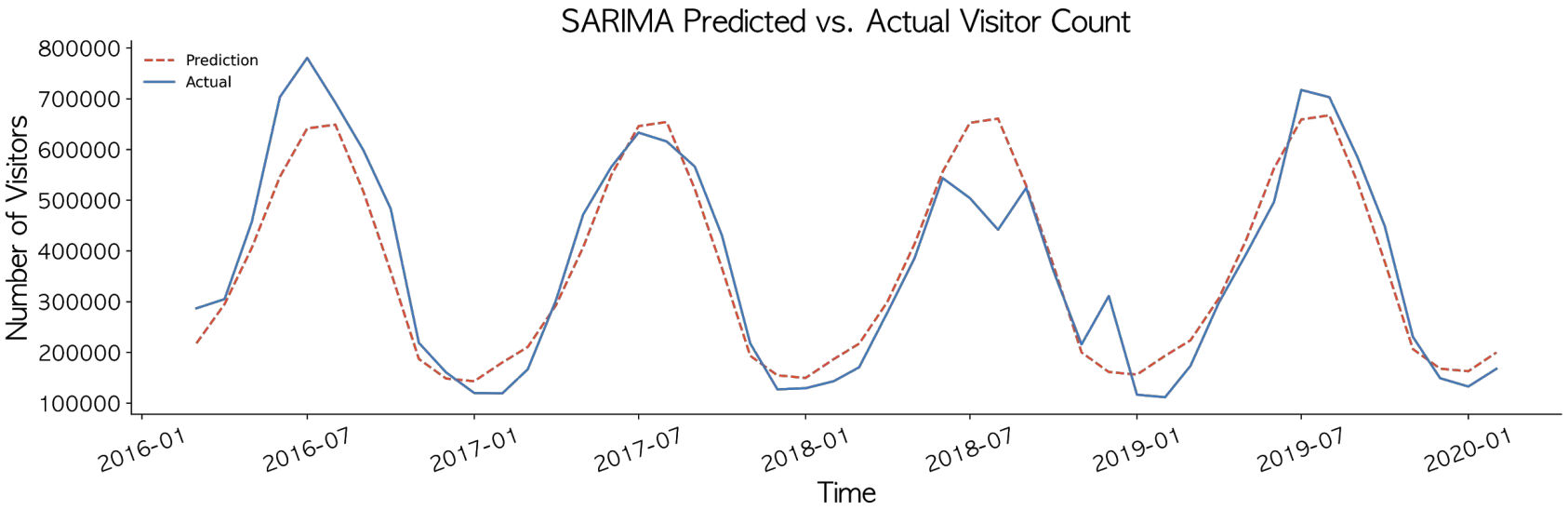
Comparing Models

SARIMA model has the lowest mean average percent error

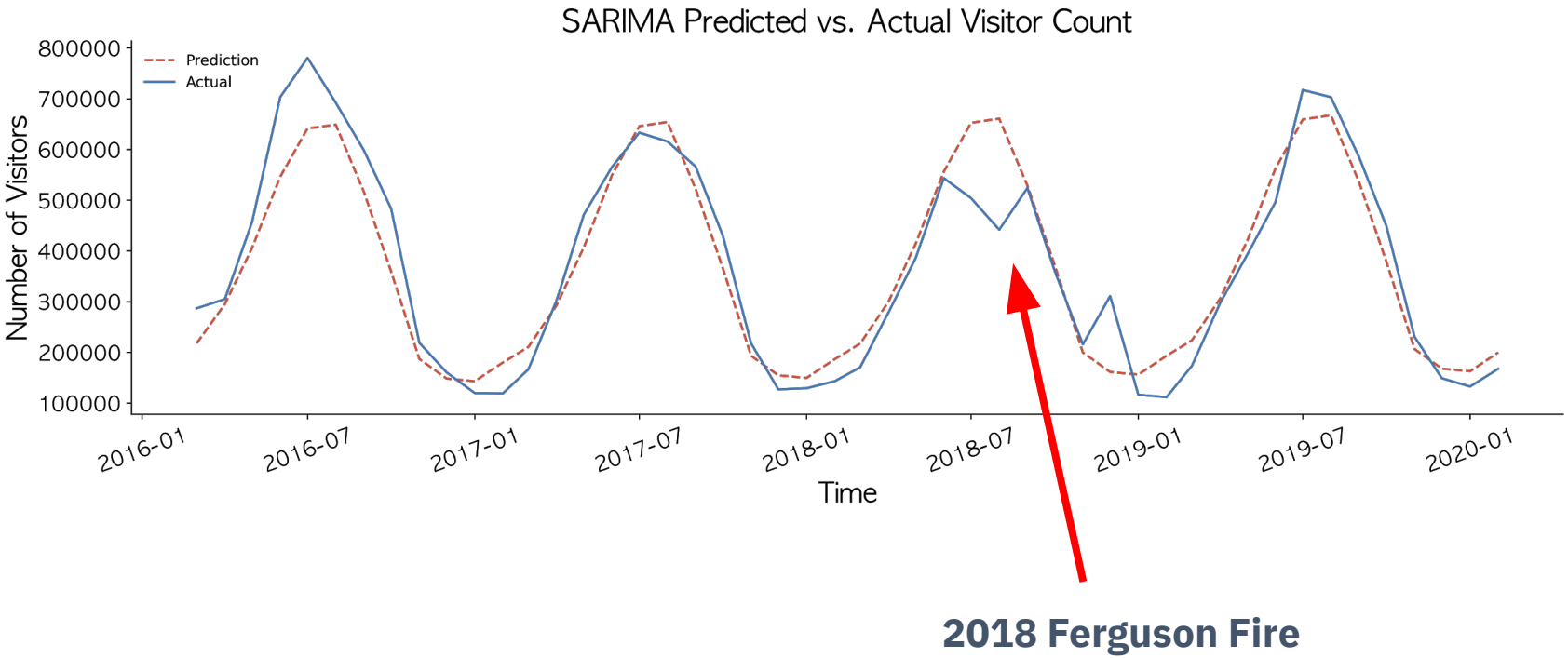


Models compared using visitor data from 10 national parks

SARIMA Forecast for Yosemite



SARIMA Forecast for Yosemite

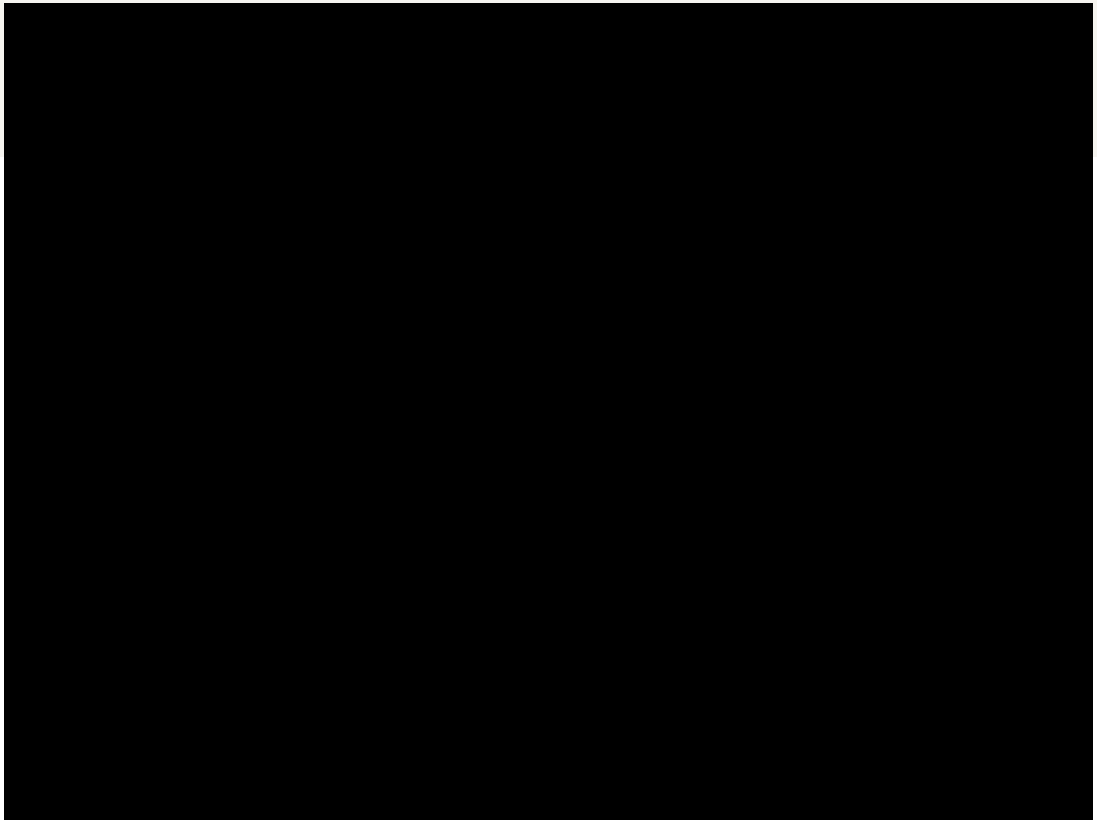


2018 Ferguson Fire

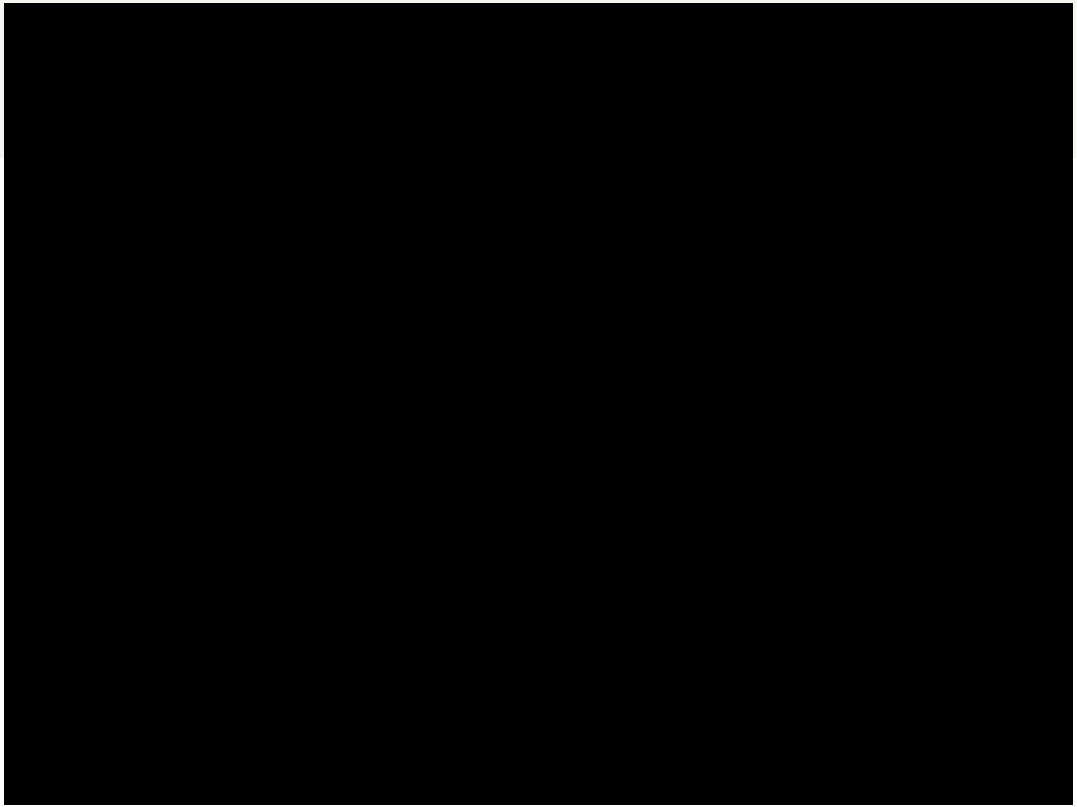
Recommender - Topic/Activity Cosine Distance

National Park	Astronomy	Downhill Skiing	Junior Ranger Program	Monument & Memorials	Voting Rights & Suffrage
Yosemite	1	0	1	0	0
Rocky Mountain	1	1	1	0	0
Gateway Arch	0	0	1	1	1

Web App - Visitor Count Prediction



Web App - Recommender



Future Work

1. Expand Predictions

Include national recreation areas, preserves, battlefields etc.

2. Examine Dips

Analyze how natural disasters such as fires and floods impact visitor counts

3. Collaborative Recs

Incorporate user review data into recommendations

THANKS!

Any questions?

https://www.linkedin.com/in/jackywlu
https://github.com/jackywlu

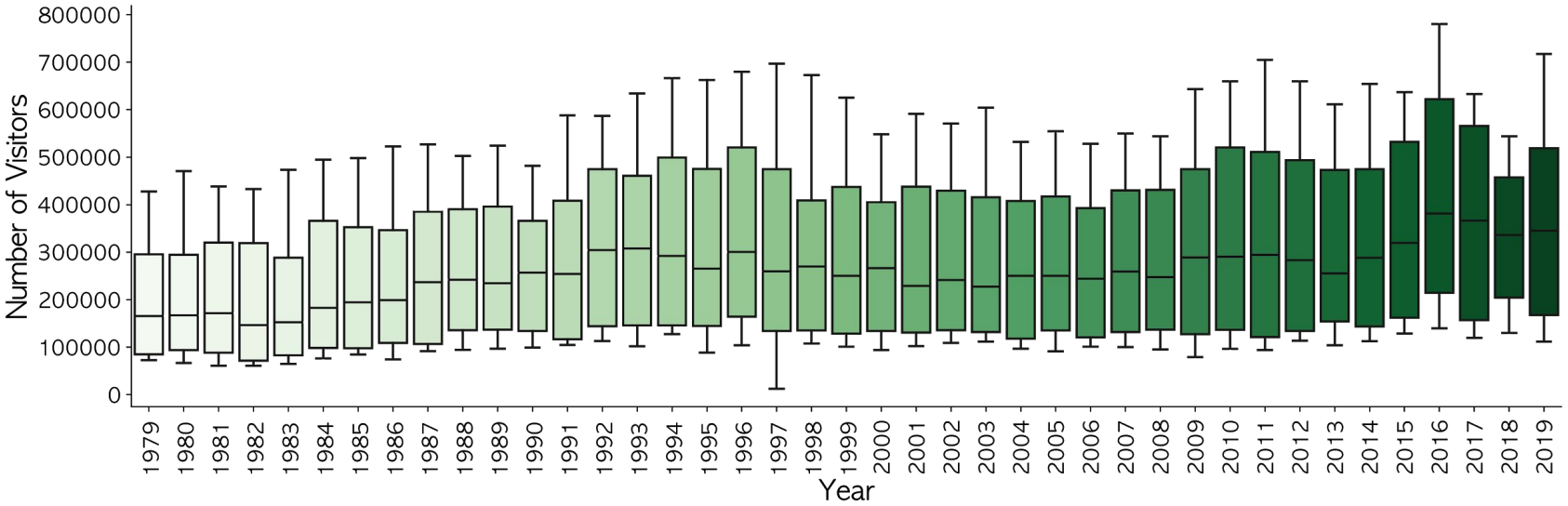
Image Credits

- https://wallpapersite.com/images/wallpapers/yosemite-national-park-4000x2700-el-capitan-yosemite-valley-4k-6680.jpg
- https://cdn.shopify.com/s/files/1/0735/7693/files/What_Makes_for_a_Great_Hiking_Boot_c348e7a8-731b-48e6-8711-b173d7f047f2.jpg?5880494476319946385
- https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcRqYPIHuMn1538P3zaHyI7zJXza_QsdFQnDZPzq8QYUKAHN4IoynsGapARzZuxAyeX2o8lwOlgyGOQMtTcpucGbjjyto0HPwD19cw&usqp=CAU&ec=45699844
- https://lilianweng.github.io/lil-log/assets/images/unrolled_RNN.png
- https://keras.io/img/logo.png
- https://miro.medium.com/max/964/0*tVCene42rgUTNv9Q.png
- https://miro.medium.com/max/3466/1*qkhefdsEERZdbn1kOcuPcQ.jpeg
- https://www.gannett-cdn.com/media/2018/01/17/USATODAY/USATODAY/636518022498866605-RockyMountainNPCBrindleNPS.jpg?width=2560
- https://funthon.files.wordpress.com/2017/05/bs.png
- https://upload.wikimedia.org/wikipedia/commons/thumb/3/38/Jupyter_logo.svg/1200px-Jupyter_logo.svg.png
- https://matplotlib.org/3.1.0/_images/sphx_glr_logos2_003.png
- https://media.cntraveler.com/photos/5e1380ce5a0316000864edde/master/pass/Bryce-Canyon-National-Park-GettyImages-921951796.jpg
- https://assets.website-files.com/5dc3b47ddc6c0c2a1af74ad0/5e18182db827fa0659541754_RGB_Logo_Vertical_Color_Light_Bg.png
- https://upload.wikimedia.org/wikipedia/commons/thumb/0/05/Scikit_learn_logo_small.svg/1280px-Scikit_learn_logo_small.svg.png
- https://cdn.mos.cms.futurecdn.net/fzWDnkZw8am545rjYpjMTT-1200-80.jp
- https://hips.hearstapps.com/ghk.h-cdn.co/assets/17/09/ghk-animals-grizzly-bear-yellowstone-national-park.jpg

Appendix

Yearly Trend

Between 1979 and 2019, there was a gradual increase in visitors to Yosemite



SARIMA Model

Seasonal ARIMA Model

- Individually fit to each national park's visitor count data
- Used pmdarima's auto_arima function to find SARIMA models with lowest AIC (Akaike Information Critera)
- Yosemite's Model: SARIMA (2, 0, 0)(0, 1, 1)[12]

LSTM Model Architecture

Simple

- 1 LSTM unit, 1 Dense unit

Stacked

- 2 LSTM units, 1 Dense unit

Bidirectional

- 1 Bidirectional LSTM unit, 1 Dense unit

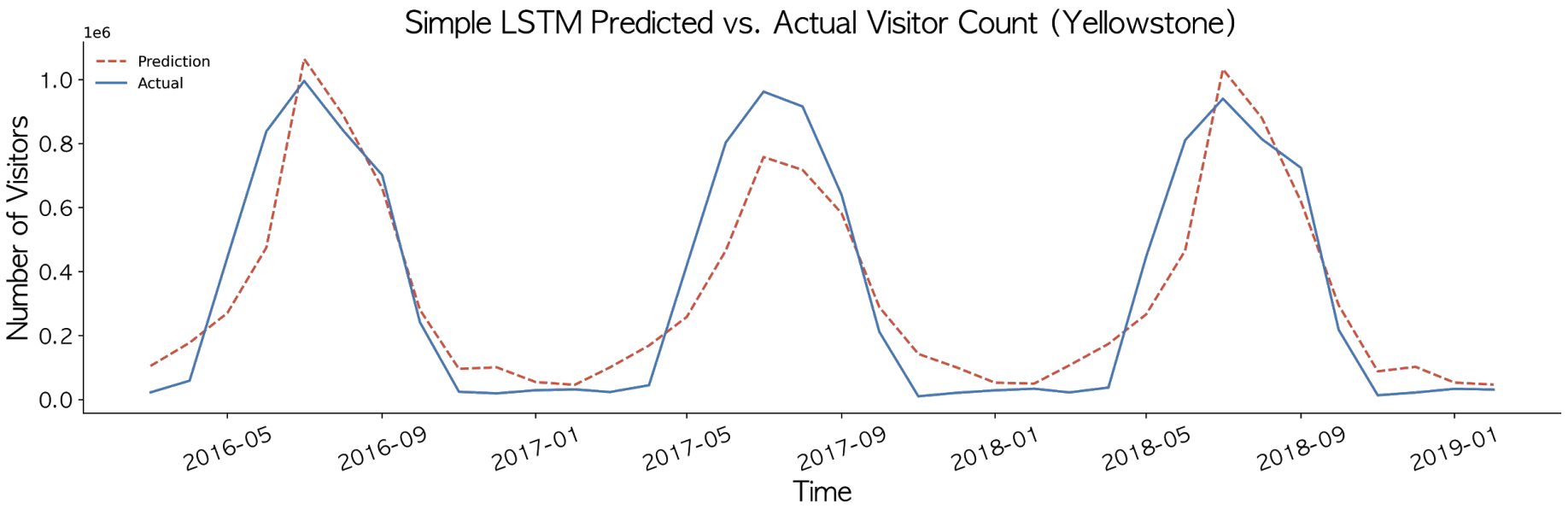
CNN LSTM

- TimeDistributed (Conv1D, MaxPooling1D, Flatten) ->1 LSTM unit, 1 Dense unit

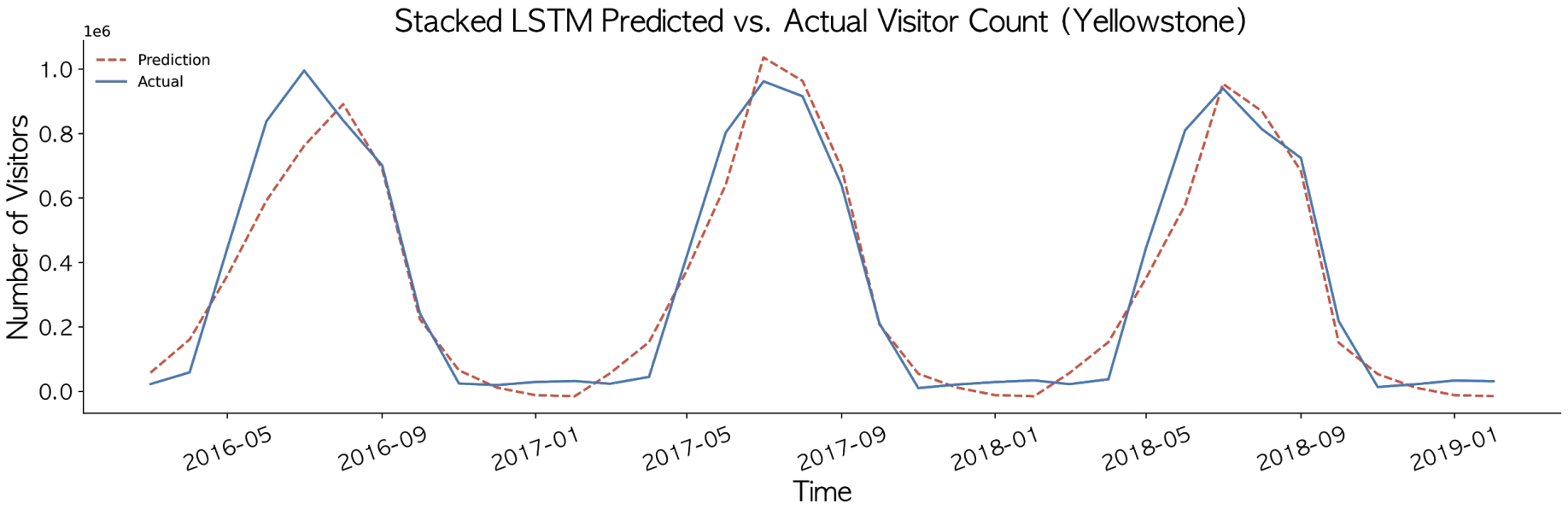
Optimizer: adam

Loss: MSE

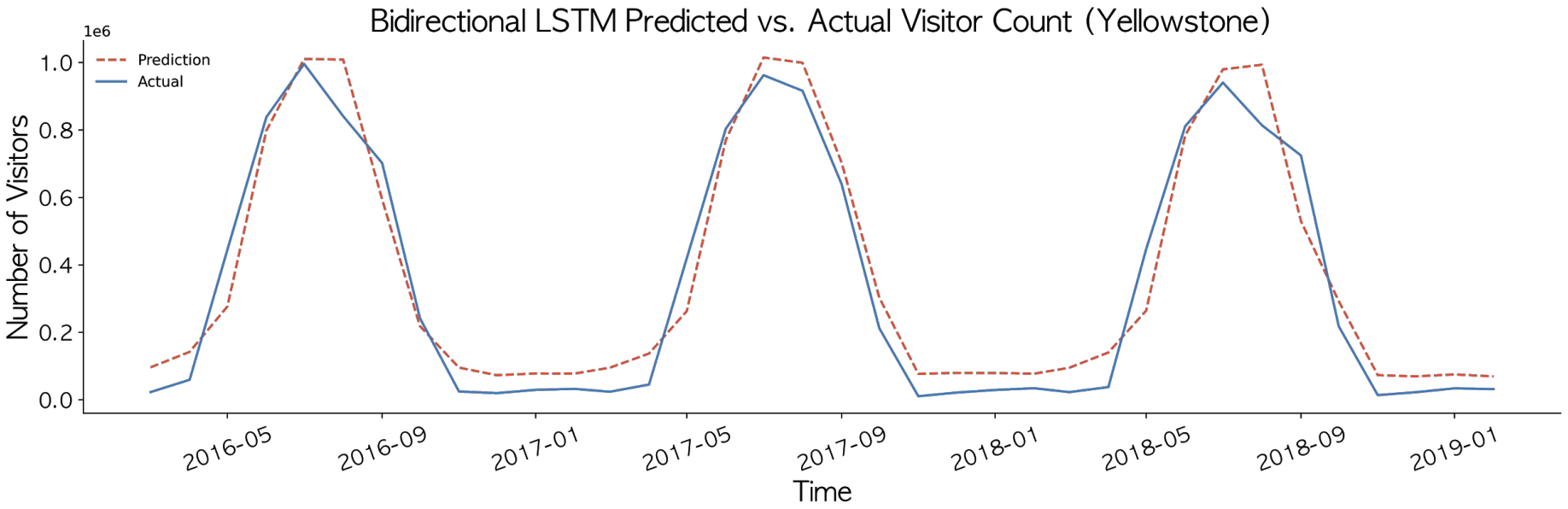
Simple LSTM Prediction vs Actual



Stacked LSTM Prediction vs Actual



Bidirectional LSTM Prediction vs Actual



CNN LSTM Prediction vs Actual