# Optimising Hospital Bed Occupancy through Machine Learning

**A presentation by FWSG**

**DSI - 41 - SG : Muhd Faaiz Khan, Sharifah Nurulhuda, Gabriel Tan, Tan Wei Chiong**

# Background Research

## Literature Review

- SG hospitals have faced the issue of lacking bed space for inpatient care since 2010. (1)
- This may be attributed to Singapore’s steadily increasing population, as well as the increasing proportion of its aging population (2).
- In recent years, 3 additional public hospitals have been opened to cater for this need. Nonetheless, Singapore’s public hospitals remain at risk of overcrowding, especially in lieu of the COVID epidemic


1. Lim, Jeremy (March 2010). "The Bed Crunch: A Systems Perspective" (PDF). SMA News. Singapore Medical Association. Archived (PDF) from the original on 11 February 2020. Retrieved 15 October 2018.
2. Ng Jing Ying (21 January 2014). "Ageing society contributes to hospital bed crunch: Gan". TODAYonline. MediaCorp Press. Archived from the original on 15 October 2018. Retrieved 15 October 2018.

# Background Research
## Literature Review

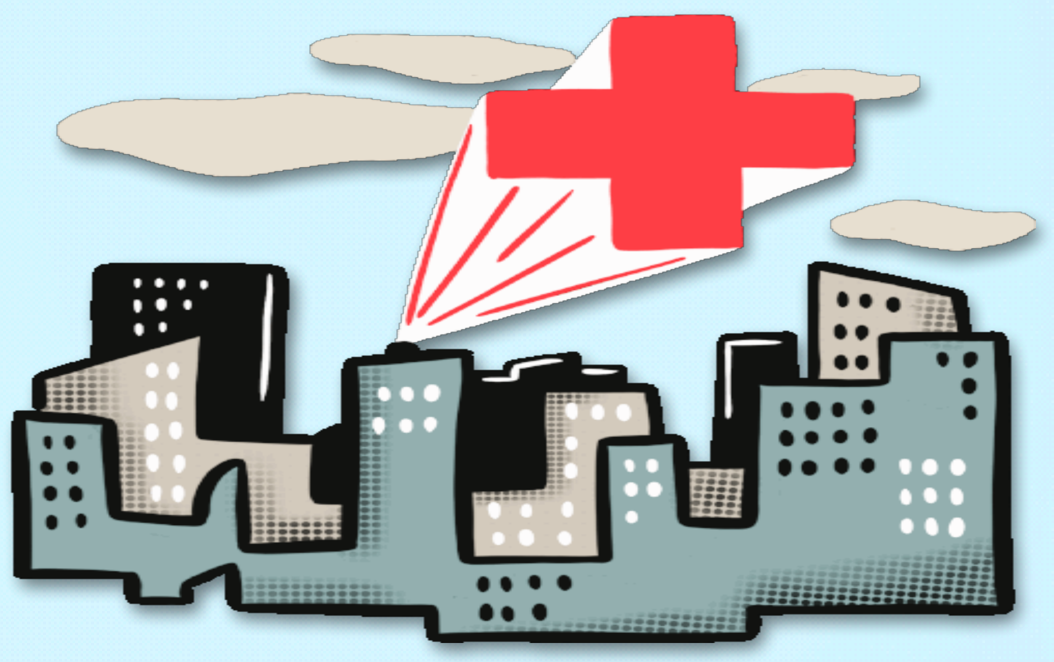

Problem of overcrowding in hospital:

- Greater costs associated with managing excess demand (3)
- Adversely affects hospital's ability to provide care (eg longer wait time, lower quality of treatment) (4)
- Increased risk of avoidable mortality

(3) Keeler TE, Yig JS. Hospital costs and excess bed capacity: a statisticalanalysis. Rev Econ Stat. 1996;78(3):470–81. https:// doi. org/ 10. 11436/ mssj.15. 250
(4) McConnell KJ, Richards CF, Daya M, Bernell SL, Weathers CC, Lowe RA. Effect of increased ICU capacity on emergency department length of stay and ambulance diversion. Ann Emerg Med. 2005;45(5):471–8. https:// doi.org/ 10. 1016/j. annem ergmed. 2004. 10. 032

# Background Research

## Domain Knowledge Introduction



| SingHealth | National Healthcare Group (NHG) | National University Health System (NUHS) |
| --- | --- | --- |
| Changi General Hospital | Khoo Teck Puat Hospital | National University Hospital |
| Sengkang General Hospital | Tan Tock Seng Hospital | Alexandra Hospital |
| Singapore General Hospital | | Ng Teng Fong General Hospital |

# Problem Statement

To accurately predict the
3-month bed occupancy rate of
National University Hospital (NUH)
so as to guide efficient allocation of
resources and manpower with
optimal use of public funds

# Meet Dr Satoshi ARIMA

Financial management and budgeting

Ensure the smooth operations of all departments within the hospital

Coordinate staff scheduling, supervision and training

Coordinate with medical governing boards

**Hospital Administrator at NUH**

# General Workflow

Data Collection

Data Cleaning

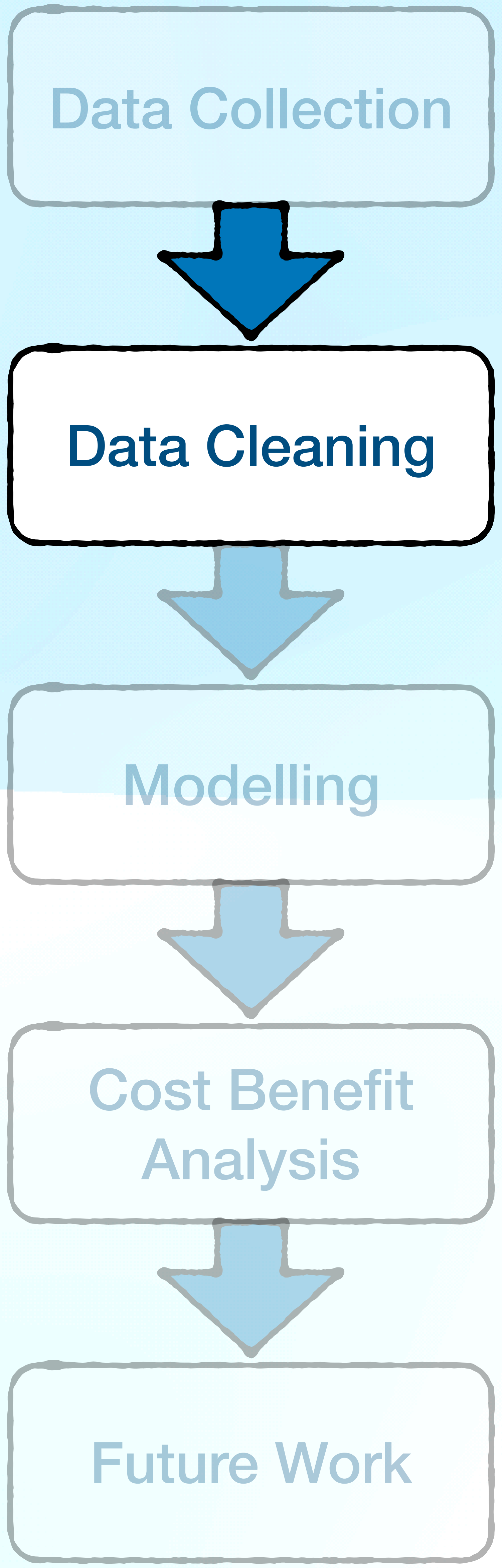
Modelling

Cost Benefit Analysis

Future Work

- Procurement of data through web scraping and other sources
- Weather data such as wet bulb temperature and rainfall
- Bed occupancy rates for eight public hospitals from 2018 to 2023
- Reported Number of Infectious Diseases cases
- Singapore Police Force Accident Data

# General Workflow




- Format standardisation to improve readability
- Managing null / missing values
- Removing unnecessary features
- Feature engineering (e.g. polynomial features)

# Missing Values

## Method: Imputing data using the mean of similar data

| Dataset on Infectious Diseases | | |
|---|---|---|
| Data structure in original dataset | Process of how it was managed | Data structure in final dataset |
| Weekly totals | 1. Separate each week into 7 days<br>2. Calculate the daily mean (average occurrences per day)<br>3. Fill in the missing cells with the imputed daily value | Daily totals |

# Removing Unnecessary Features

| Weather Data Set (scraped from weather.gov.sg and dowloaded from data.gov.sg) | | |
|---|---|---|
| **Description** | **Features Kept** | **Features Dropped** |
| Location | Changi station | all other stations |
| Date | year, month, day | |
| Rainfall | daily rainfall total (mm) | highest 30 min rainfall (mm),<br>highest 60 min rainfall (mm),<br>highest 120 min rainfall (mm) |
| Temperature | daily wet bulb temperature (°C) | mean temperature (°C),<br>maximum temperature (°C),<br>minimum temperature (°C) |
| Wind speed | | mean wind speed (km/h)<br>max wind speed (km/h) |

# Feature Engineering

## Creating new features to help make sense of or classify our existing data into categories

| date | day_of_week | is_holiday |
|---|---|---|
| 2020-08-08 | Sat | 0 |
| 2020-08-09 | Sun | 1 |
| 2020-08-10 | Mon | 1 |
| 2020-08-11 | Tue | 0 |
| 2020-08-12 | Wed | 0 |

new features!

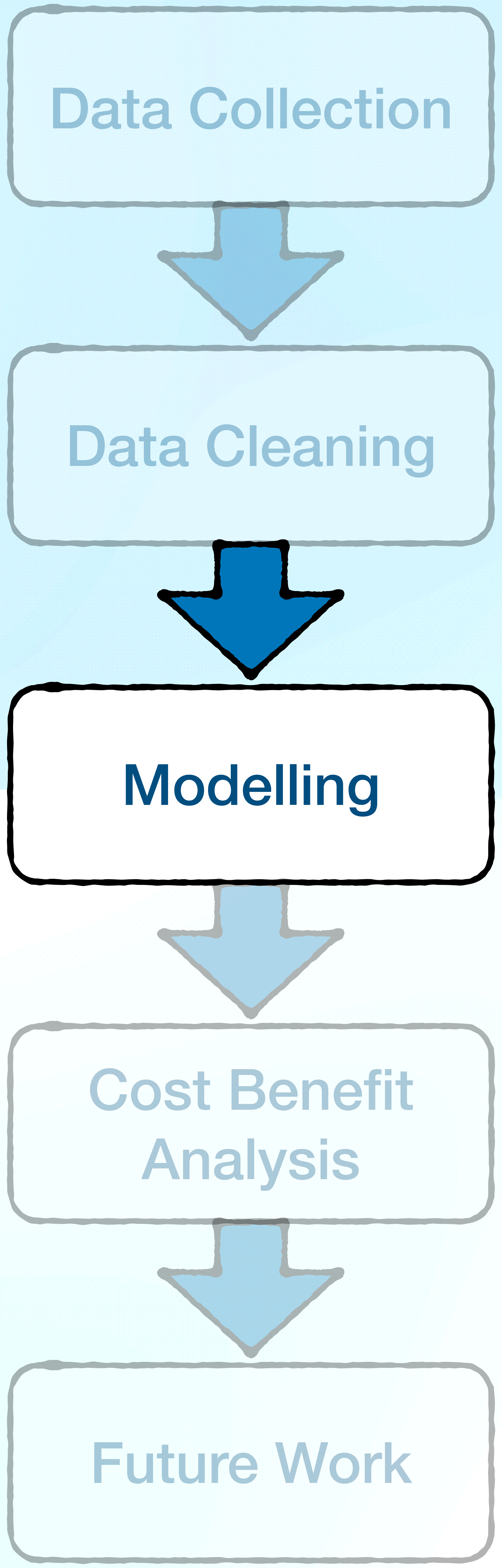
# Feature Engineering

## Dummifying Categorical Features

- Transforming categorical features into binary using OneHotEncoder

## Polynomial Features

- Transforming numerical features
- May create stronger predictors of our target variable




# General Workflow

- Linear Regression with RidgeCV and Lasso CV
- Random Forest

- Seasonal Auto Regressive Integrated Moving Averages (SARIMA)
- SARIMA with added exogenous features (SARIMAX)

- Long-Short Term Memory Recurrent Neural Network (LSTM RNN)

# Modelling

## Linear Regression with Ridge CV and Lasso CV

| Linear Regression | Ridge CV | Lasso CV |
|---|---|---|
| • Used to model the linear relationship between continuous variables<br>• Tries to minimise the discrepancies between predicted and actual values | • Regularisation technique to penalise large coefficients<br>• To correct for overfitting | • Regularisation technique to penalise for the sum of absolute values of coefficients<br>• To reduce the coefficients (sometimes to zero) |

| Metrics | Linear Regression | Ridge CV | Lasso CV |
|---|---|---|---|
| R2 Score | 0.3993 | 0.4296 | 0.4327 |
| RMSE | 3.0788 | 3.0002 | 2.9922 |

# Modelling : Insights

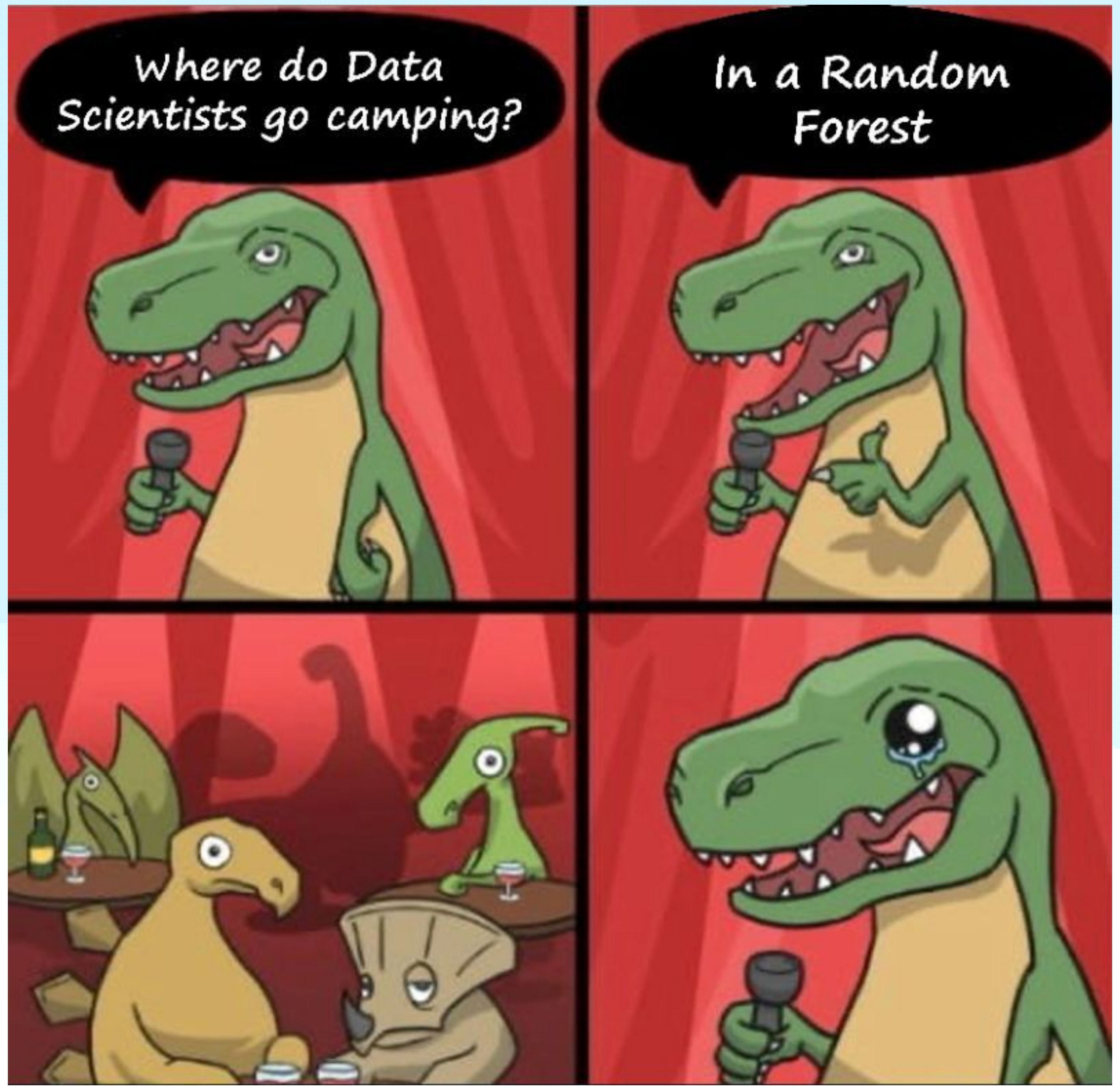
## Linear Regression with Ridge CV and Lasso CV

- All three models agree : time (specifically in months) affect the availability of hospital beds the most (has the largest coefficients)
- After regularisation,
  - May is the month with the largest positive impact on occupancy
  - January is the month with the largest negative impact on occupancy
  - Public holidays and weekends tend to have lower occupancy rates

# Modelling

## Random Forest with Grid Search CV

- takes a subset of the features at random to train a single prediction
- each prediction is then averaged to produce the final prediction
- Grid Search : searches for the best values for the random forest hyper parameters




## Metrics

| Training R2 score | Testing R2 score | RMSE |
|---|---|---|
| 0.8572 | 0.5363 | 2.7053 |

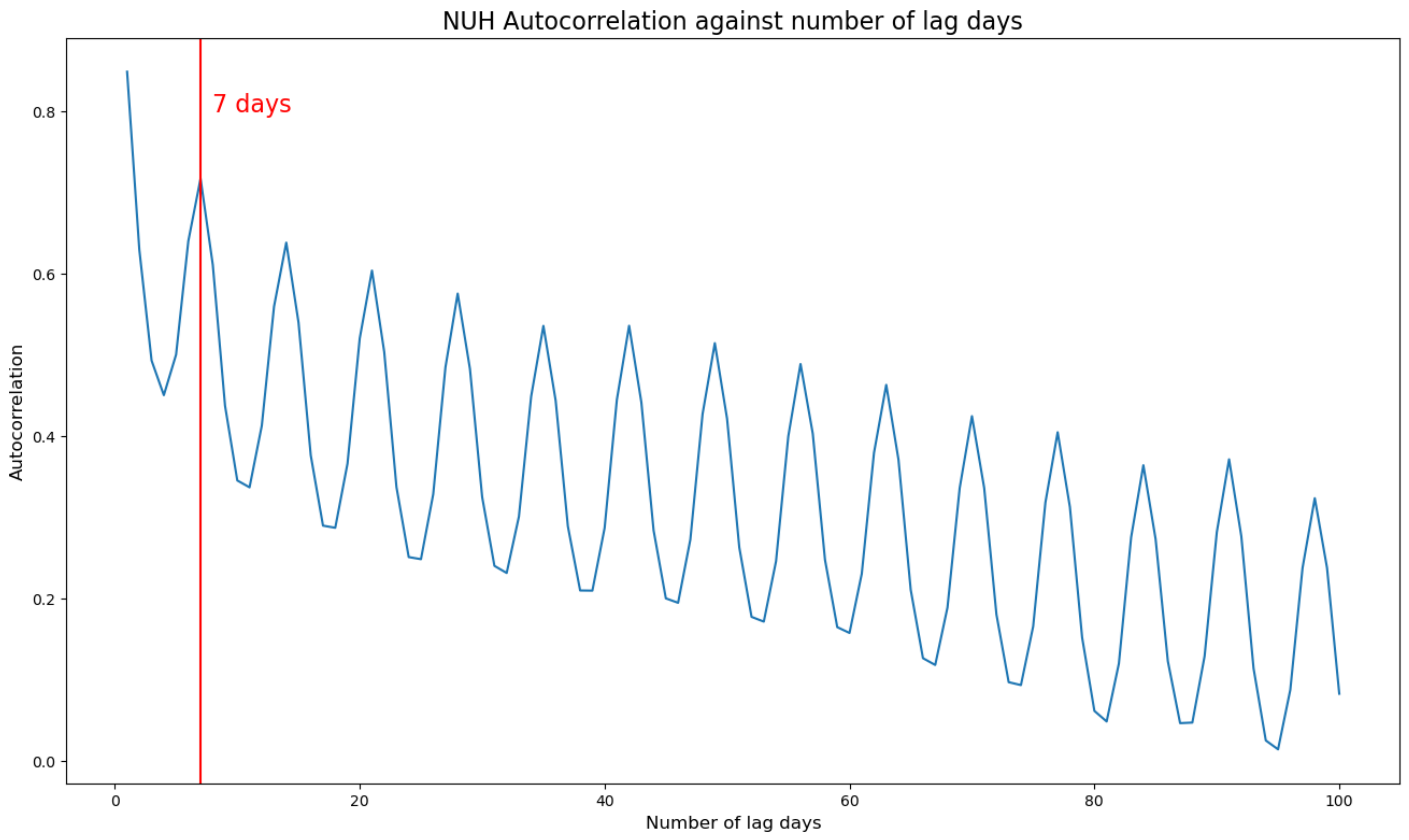
significant difference

| Grid Search best params | | |
|---|---|---|
| Max depth | Min Sample Split | N_estimators |
| 20 | 10 | 200 |

# Modelling : Time Series

| Model | What it is | What it's used for |
|---|---|---|
| SARIMA | • Stands for Seasonal Auto Regressive Integrated Moving Average<br>• Takes into account any seasonality patterns | • Used for predicting future points in a time series |
| SARIMAX | • Everything about SARIMA<br>• Includes exogenous (X) features that might contribute to a more accurate prediction | |

# Modelling : Time Series

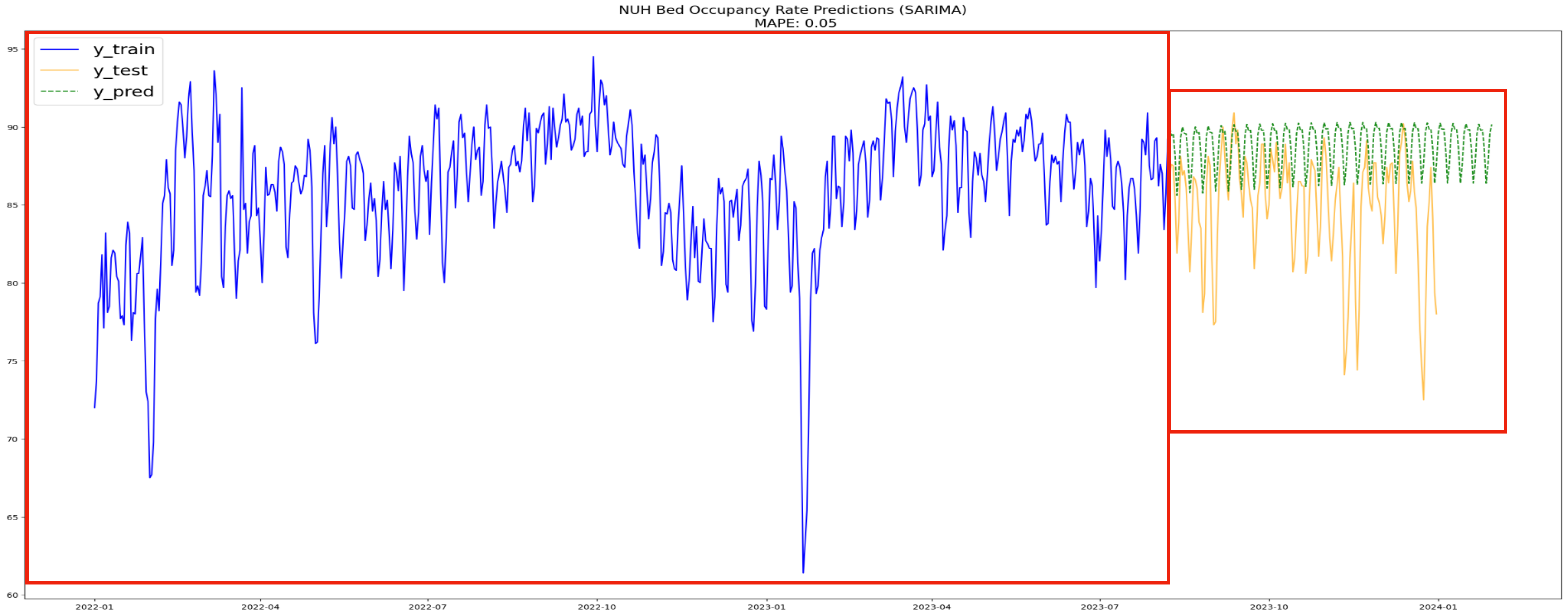



- Autocorrelation : explores the correlation between a series and a lagged version of itself
- Lag : shift of one observation relative to another observation, similar to a time difference
- Seasonality : repetitive and predictable variations in a series that are influenced by seasonal factors

# Modelling : SARIMA

MAPE : Mean Absolute Percentage Error
Measure of prediction accuracy of the forecast

MAPE = 0.05



# Modelling : Insights

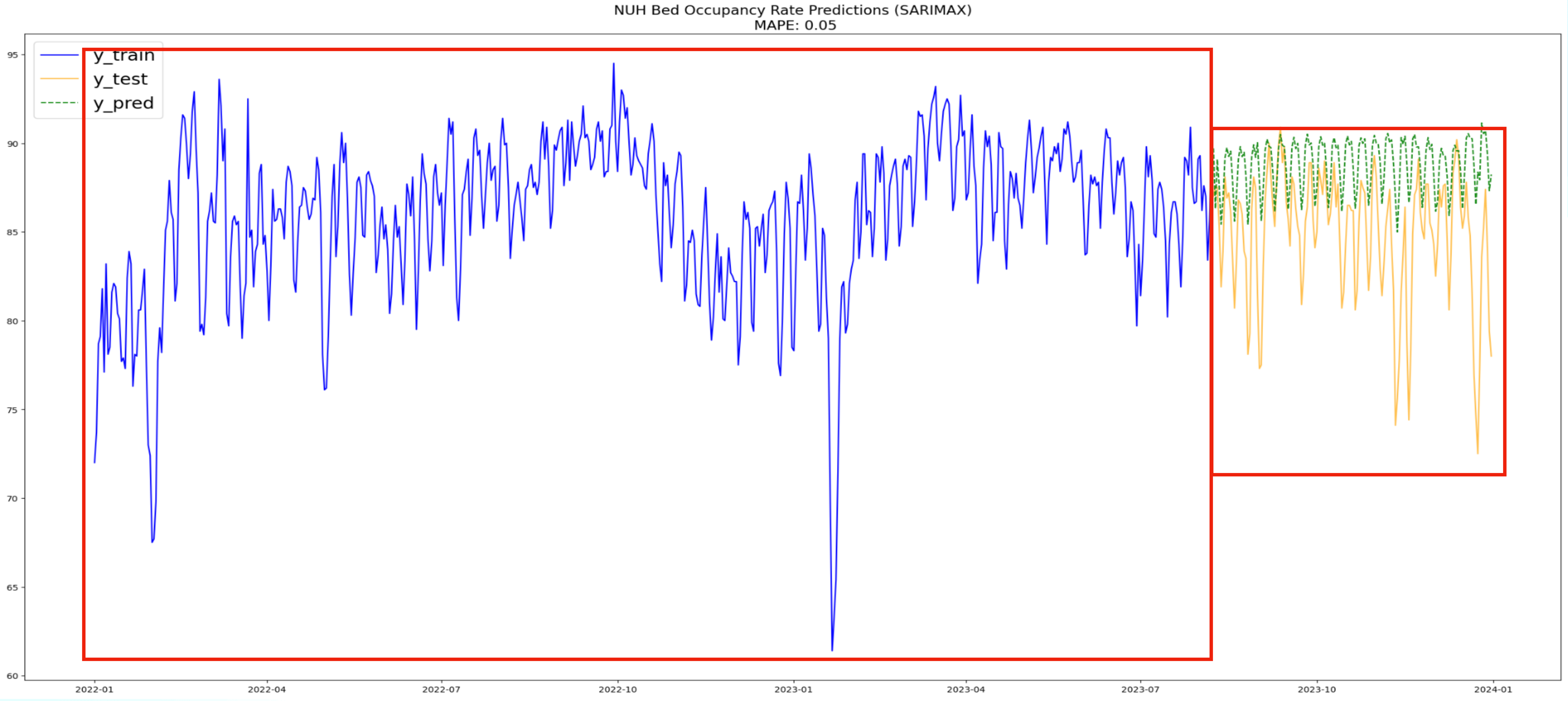
## SARIMA

- Predictions of test set matches closely with peaks of the actual test data
- Fairly uniform trend, likely a result of the autocorrelation. Trend carries on even after leaving the test set.
- Optimal hyperparameters are ARIMA(0,1,0)(2,0,2)[7] with intercept

# Modelling : SARIMAX

MAPE : Mean Absolute Percentage Error
Measure of prediction accuracy of the forecast

MAPE = 0.05



# Modelling : Insights
## SARIMAX

- Fits and looks almost identical to the model obtained by SARIMA
- Cannot make predictions outside of test set because of lack of X for those times
- Downsides to this approach: Errors propagate very quickly
- Seems like exogenous features are not too predictive in this context
- Model hyperparameters are ARIMA(1,1,0)(2,0,2)[7] with intercept
- Future work: Use SARIMA to individually predict each one of the X features past the test set, and then use those predicted X values as X input to SARIMAX

# Modelling : LSTM RNN

## Long-Short Term Memory Recurrent Neural Networks

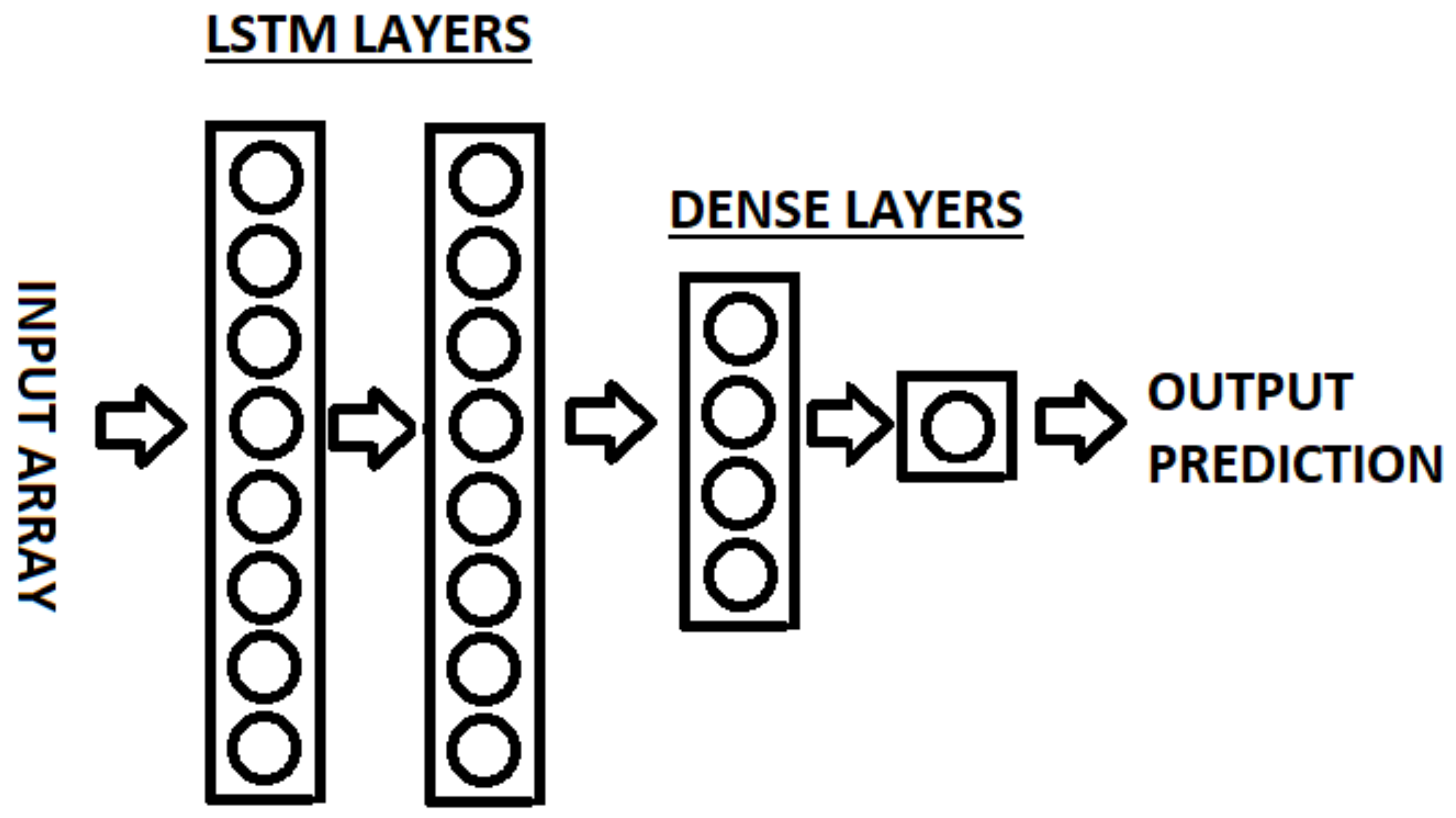
- Sequential neural network as a predictive model
- We will attempt to create an autoregressive model using LSTM.

**Features:**

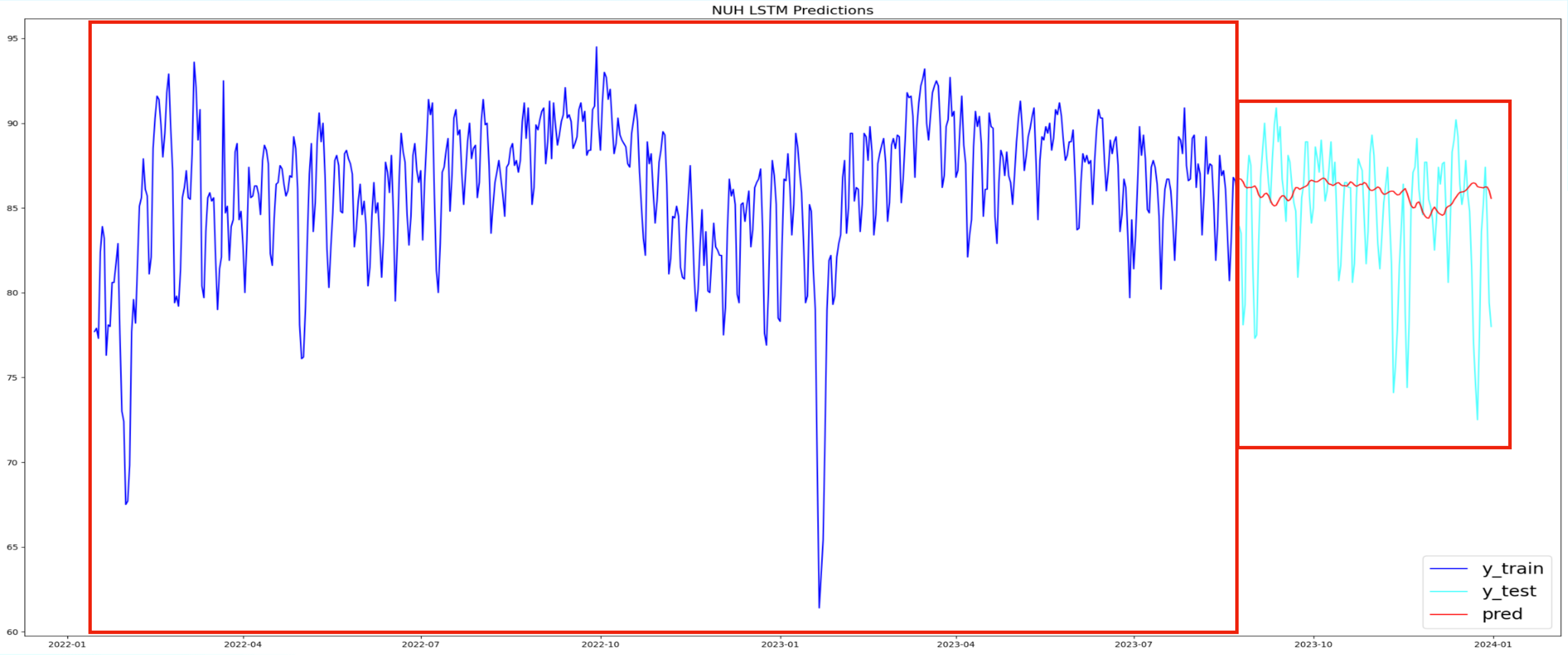
After multiple permutations, we have found the following features contribute to stronger predictive ability of our model:

- occupancy from 5 days prior
- occupancy from 6 days prior
- occupancy from 7 days prior

This aligns with our prior observations when performing SARIMA and SARIMAX, where strong seasonality was seen in the timeframe of approximately 1 week.



# Modelling : LSTM



# Modelling : Insights

## LSTM

As seen in the plot, our autoregressive LSTM model appears to perform well on the unseen testing data.

This model will face similar issues as the SARIMA model, where predictions beyond the near future may not be as accurate due to unstability when extrapolated further in time.

# General Workflow

Data Collection

Data Cleaning

Modelling

**Cost Benefit Analysis**

Future Work

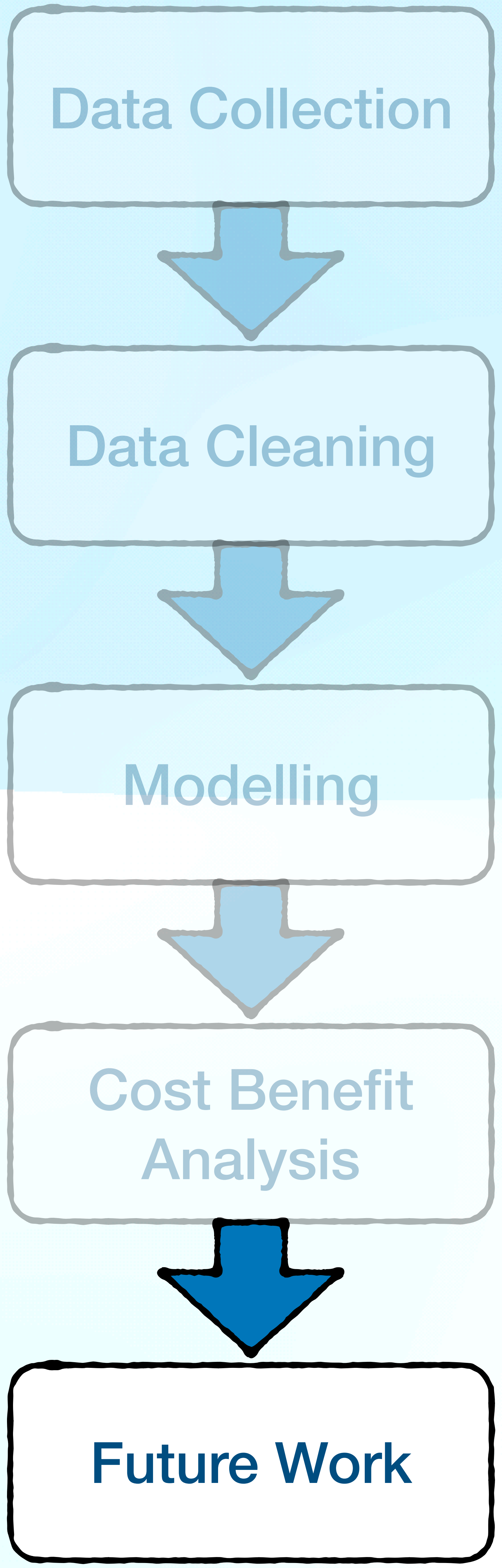
- What this means for NUH in terms of manhours and operating cost savings for the next 2 years

# Cost Benefit Analysis : Year 1

| | Item | | | |
|---|---|---|---|---|
| | | **CY 2024 : Q3 to Q4** | | |
| | | *Assuming tender for project awarded by Q2'24* | | |
| | | | | |
| **COSTS** | | **Setup (6 months)** | | |
| | 1 | Model development (data acquisition and cleaning, model training and development) | $ | 300,000.00 |
| | 2 | Development of dashboard app; implementation and integration with existing systems | $ | 200,000.00 |
| | 3 | Hardware and software acquisition | $ | 40,000.00 |
| | 4 | Coordination, staff training and education costs | $ | 20,000.00 |
| | | ***Sub-total for costs*** | $ | 560,000.00 |
| | | | | |
| | | **CY 2025 : Q1 to Q2** | | |
| | | Based on **3% reduction in sub-optimal bed utilisation,** i.e. ***36 beds in NUH***: | | |
| BENEFITS | | **Implementation (6 months)** | | |
| | | | | |
| | 5 | Manhour savings from staffing optimisation | $ | 176,000.00 |
| | | *Assuming:* | | |
| | | - Bed ratio of 1 senior nurse to 4 beds, i.e. 9 nurses x 4-hour overtime shift/day x 182.5 man-days | | |
| | | | | |
| | 6 | Cost savings from supply chain optimisation | $ | 262,800.00 |
| | | *Assuming:* | | |
| | | *- Savings through inventory/medical supply procurement processes; in-patient meal catering, etc.* | | |
| | | *- $40 cost savings per bed/day x 182.5 man-days* | | |
| | | | | |
| | | ***Sub-total for savings*** | $ | 438,800.00 |
| | | | | |
| | | ***NET BENEFIT for month 1 - 12*** *(Savings minus Costs)* | ***-$*** | ***121,200.00*** |

# Cost Benefit Analysis : Year 2

| | | | |
|---|---|---|---|
| | | **CY 2025 : Q3, Q4** to **CY 2025 - Q1, Q2** | |
| | | | |
| COSTS | | **Annual maintenance and updates (e.g. software licences, storage solutions)** | |
| | 7 | Maintenance cost and updates (e.g. software licences, data storage solutions) | $ 30,000.00 |
| | | ***Sub-total for costs*** | $ 30,000.00 |
| | | | |
| | | | |
| BENEFITS | | Based on **3% reduction in sub-optimal bed utilisation,** i.e. ***36 beds in NUH***: | |
| | | **Implementation (12 months)** | |
| | 8 | Manhour savings from staffing optimisation | $ 352,000.00 |
| | 9 | Cost savings from supply chain optimisation | $ 525,600.00 |
| | | ***Sub-total for savings*** | $ 937,600.00 |
| | | | |
| | | ***NET BENEFIT for month 13 - 24*** *(Savings minus Costs)* | **$ 907,600.00** |
| | | ***NET BENEFIT over 24 months (2 years)*** | **$ 786,400.00** |
| | | | |
| **Sources:** | | | |
| Nursing manhour costs (NUH) | https://www.hseu.org.sg/wps/wcm/connect/4bd241f4-d006-4edb-ba27-f7cec483954b/NUHS+Collective+Agreement+2022.pdf?MOD=AJPERES | | |
| Nurse-to-bed ratio | https://www.moh.gov.sg/news-highlights/details/ensuring-adequate-rest-for-nurses/#:~:text=The%20typical%20nurse%2Dto%2Dbed,for%20more%20complex%20l | | |




# General Workflow

Upon commissioning, future work for the modeling could include hospital data to be supplied by NUH, such as:

- - Emergency Department (ED) admissions data
- - Length of stay data for different patient types and disease mix
- - Discharge rates and patterns
- - Readmission rates for specific conditions or patient groups
- - Specific changes in work processes or innovations that could have had an impact patient care efficiency


# Thank you for listening!

A presentation by FWSG

DSI - 41 - SG : Muhd Faaiz Khan, Sharifah Nurulhuda, Gabriel Tan, Tan Wei Chiong