Oakland Airbnb Pricing Analysis

TEAM 5 - Evelyn (Wanyi) Dong, Jenny Shang, Qiongqiong Lin, Yoki (Lingrou) Wang

Agenda

airbnb

1. Introduction & Motivation
2. Data Details
3. Methodology
4. Models & Performance
5. Conclusion

Section 1
Introduction & Motivation

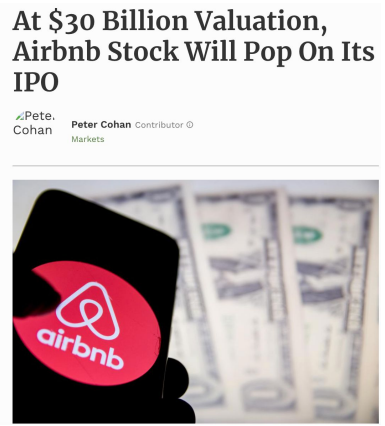
Our Achievements So Far

“ ‘Airbnb’ has become synonymous with one of a kind travel on a global scale.”

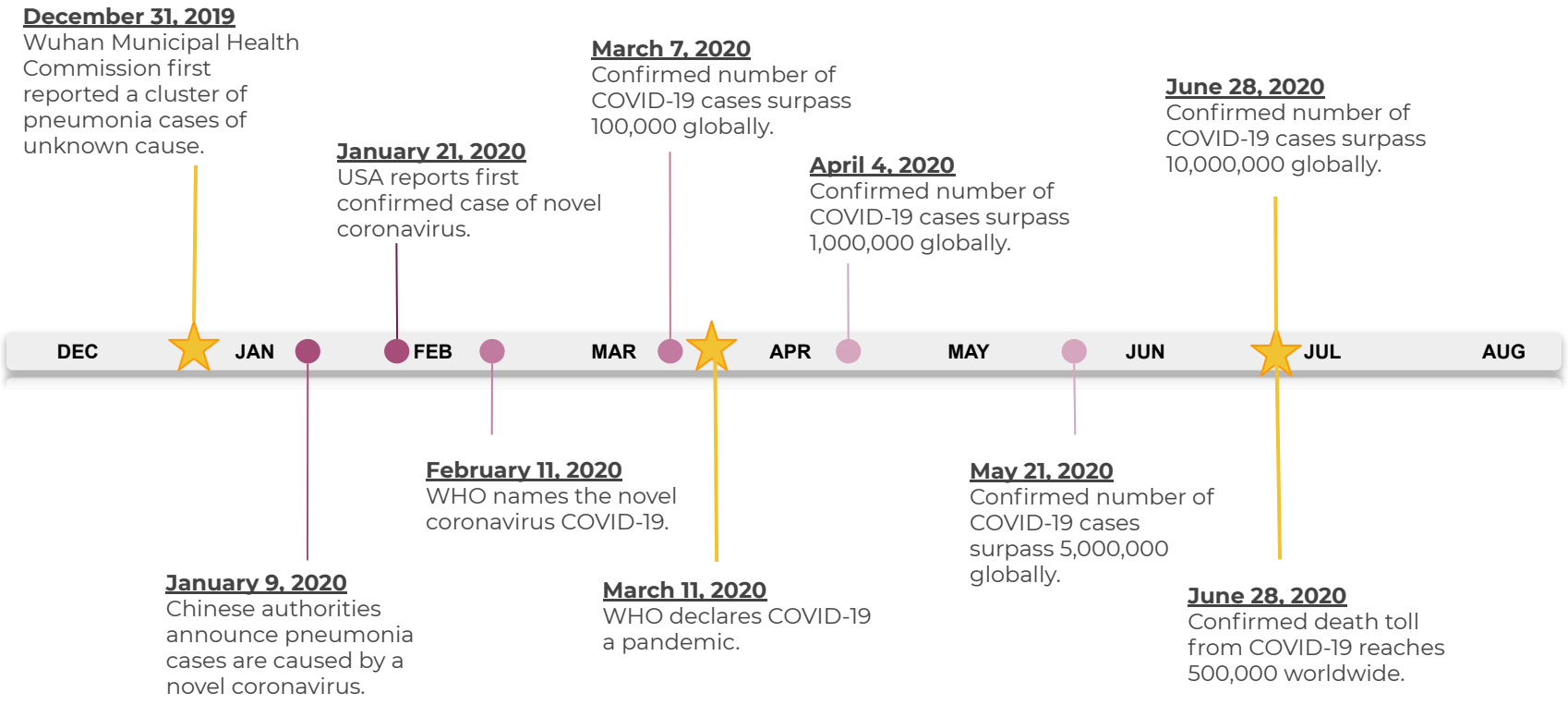
$38 Billion
in gross booking value (GBV) in 2019

4 Million+ hosts

+7.4 Million+ listings

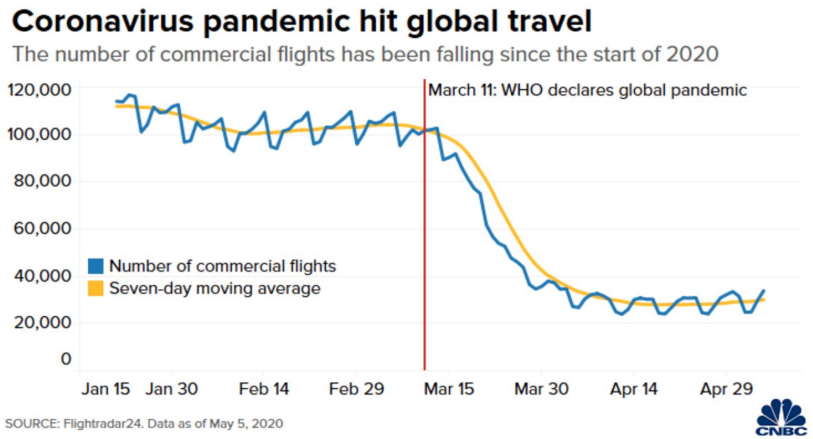


Timeline of COVID-19



Team Objective

1. Understand the **historical patterns** of the Airbnb average listing price
2. **Assess the impact of COVID-19** on Airbnb average listings price using various time-series methods
3. Generate a **30-day average price forecast**
4. Develop **insights and recommendations** that can help Airbnb stakeholders in planning and strategy



Section 2
Data Details

Data Details - Main Data Source

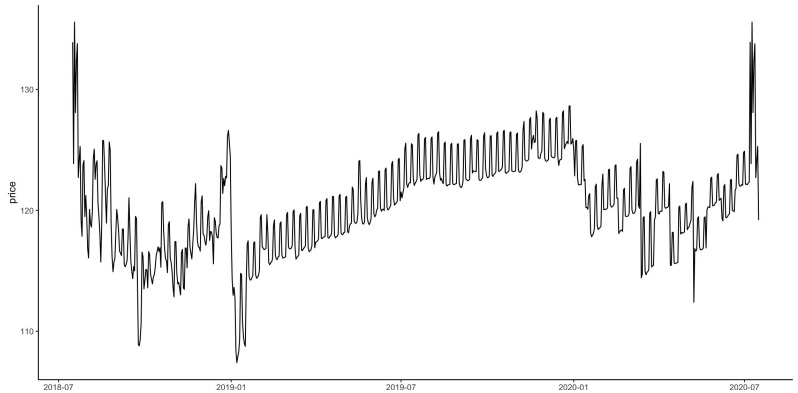
DATA DESCRIPTION

DATASET Inside Airbnb (7/16/2018 - 7/16/2020)

SOURCE http://insideairbnb.com/get-the-data.html

CONTEXT Inside Airbnb is an independent, non-commercial set of tools and data that allows you to explore how Airbnb is being used in cities around the world.

CONTENT Data frequency is on daily-basis from July 16, 2018 to July 16, 2020 (722 data points). Cleaned data includes average listing price of all available properties in Oakland, CA.]



Data Details - External Data Source

We considered **three** potential **alternative** data sources :

GDP
(Nationwide)

Unemployment
(Nationwide)

Housing Data
(Oakland, CA)

Section 3
Methodology

Data Staging

The year of 2020 has been a disruptive one because of the outbreak of coronavirus, COVID-19.

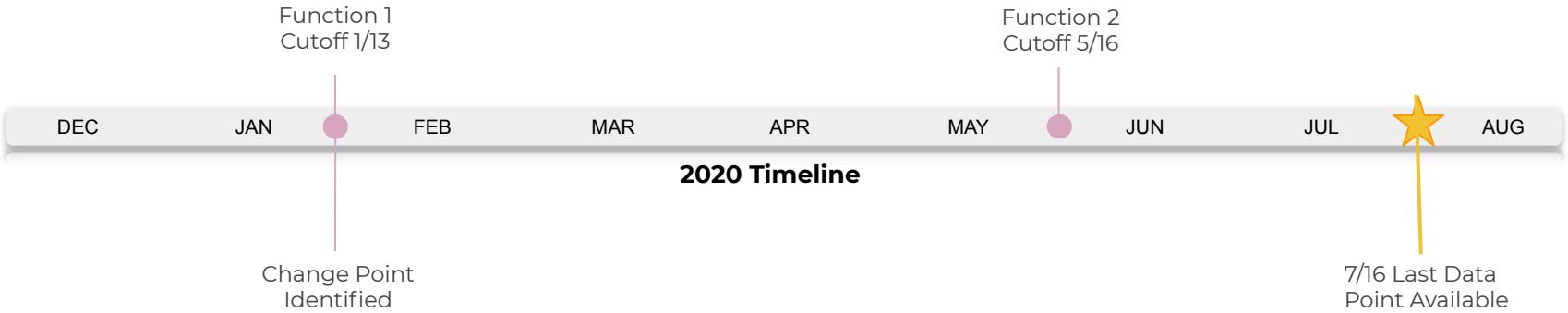
A test was performed to find the dropping point of the data in 2020, and the date was identified as 1/13/2020. To examine the impact of the pandemic, two functions are designed:

1. **Only using Pre-COVID Data**
   a. Data is identified to be data from 8/16/2018 to 1/13/2020

2. **Using all data available**
   a. Data is identified to be all data available from 8/16/2018 to 7/16/2020

For two functions, the same analysis method is applied:

- A 60 days validation window is reserved to slide and test model performance
- Forecast the next 30-day average listing prices as final deliverables



Details of Predictive Modeling

1

6 Methods Were Developed as Candidate Models

Naive Forecast (baseline)
Holt Winters Exponential Smoothing
ARIMA
Neural Net
Linear Regression
Decision Tree

2

Cross-Validation of Prediction Errors

A sliding window for cross validation

Enhance the credibility and reliability of the models

3

Deployment of the Predictive Models

Compare patterns of the two staging methods (pre-covid v.s. covid)

Occurrence of the emergency pandemic not aligns with previous historic patterns

Section 4
Models & Performance

Naive Model

Naive Model

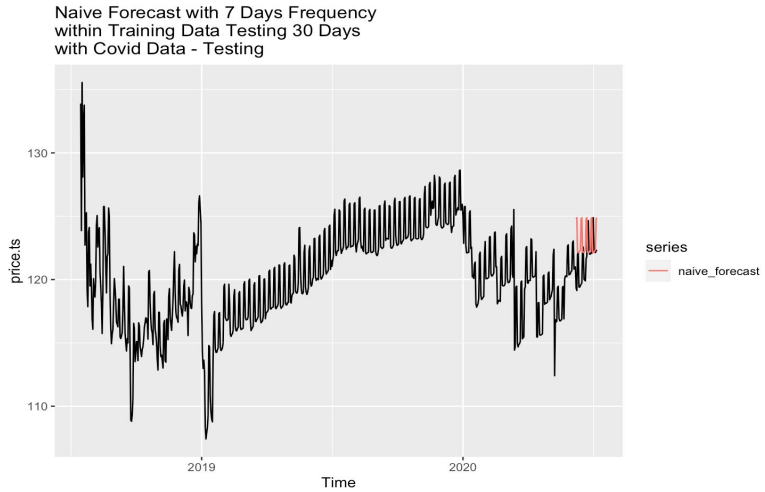
Uses previous actual data as the prediction of future prices. The rationale behind this method is that the future data will have a similar pattern as historical data.

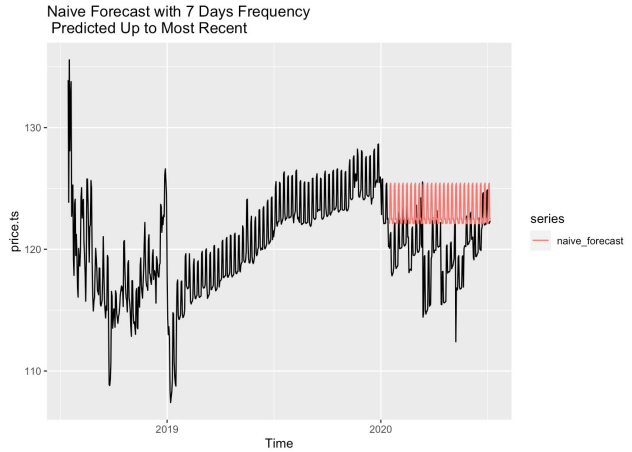
Only using Pre-COVID Data

Training MAPE - 1.8, Testing MAPE - 3.03

Using all data available

Training MAPE - 2.26, Testing MAPE - 1..77





Holt Winters Exponential Smoothing

Holt Winters

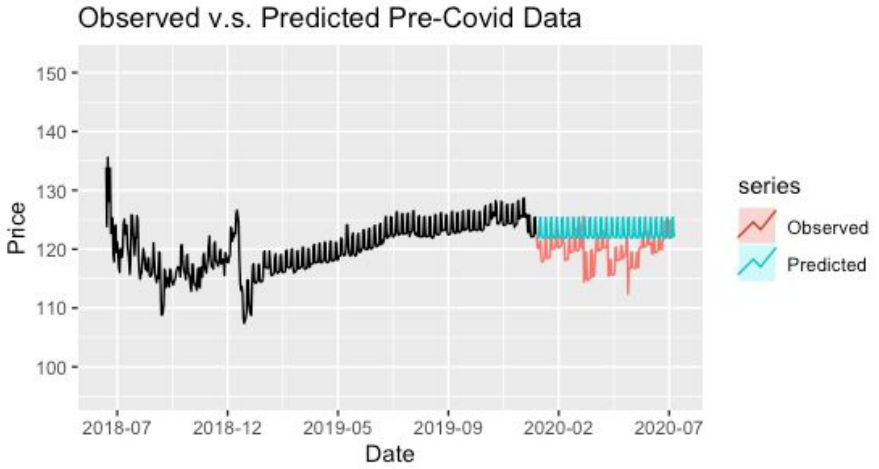
The most optimal one with the multiplicative error, a damped-additive trend, and additive seasonality (MAdA)

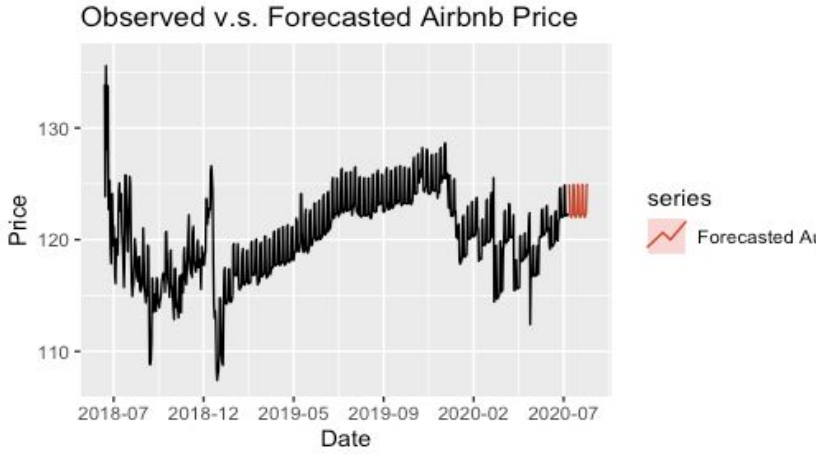
Only using Pre-COVID Data

Training MAPE - 0.52, Testing MAPE - 1.52

Using all data available

Training MAPE - 0.68, Testing MAPE - 1.91





Arima

Arima

The PACF graph cuts off and the ACF graph slowly tails off.
According to the ACF and PACF graph, we have tried:

Only using Pre-COVID Data

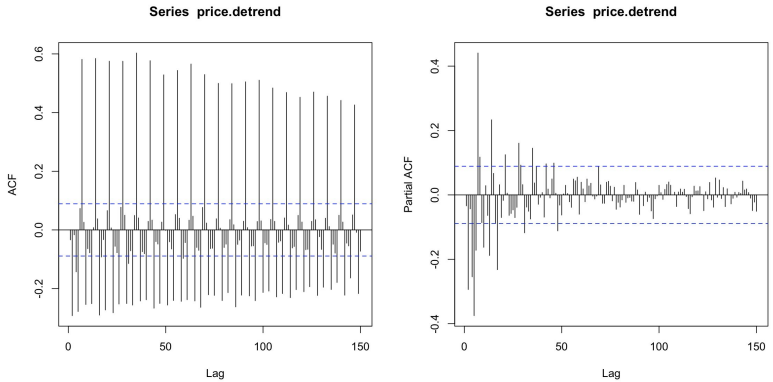
Training MAPE - 0.54, Testing MAPE - 1.51

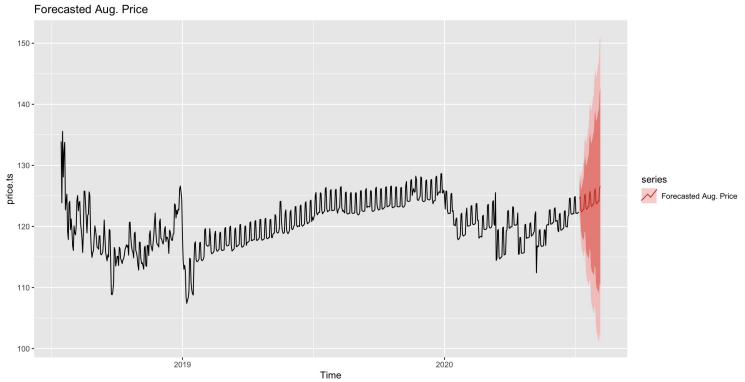
- **SARIMA(0,1,1)(0,1,1)[7]**
- SARIMA(0,1,0)(1,1,0)[7]
- SARIMA(0,1,0)(1,1,1)[7]
- SARIMA(1,1,1)(1,1,1)[7]

Using all data available

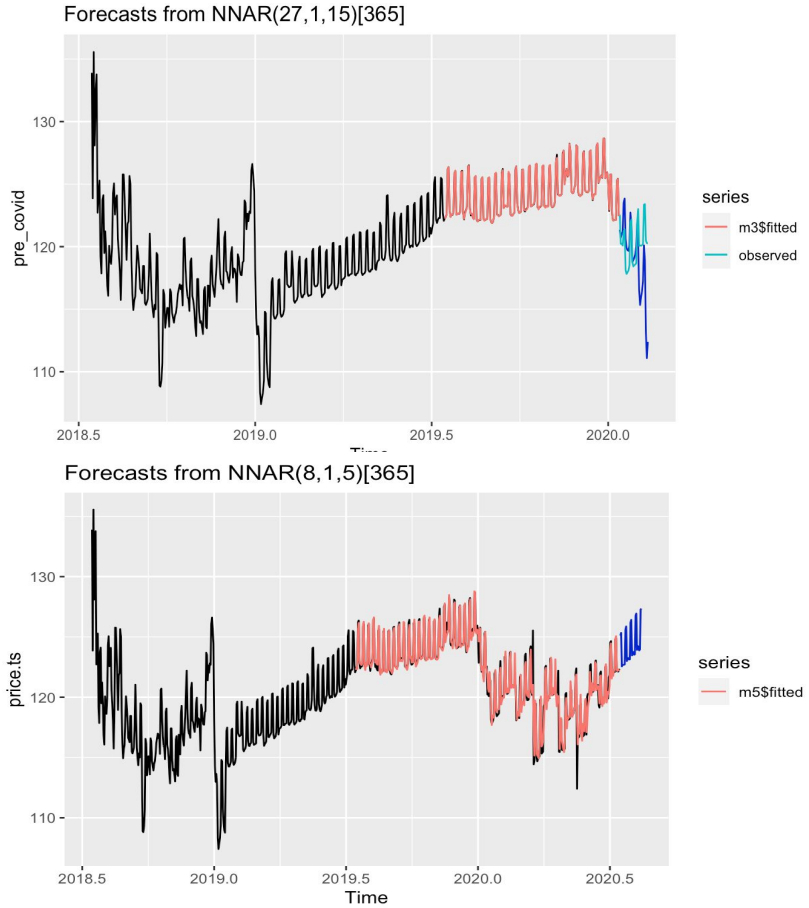
Training MAPE - 0.62, Testing MAPE - 1.16

- SARIMA(0,1,1)(0,1,1)[7]
- SARIMA(0,1,0)(1,1,0)[7]
- SARIMA(0,1,0)(1,1,1)[7]
- **SARIMA(1,1,1)(1,1,1)[7]**





NNAR



NNAR

NNAR is equivalent to an ARIMA model but without stationarity restrictions.

Only using Pre-COVID Data

Training MAPE - 0.1, Testing MAPE - 1.42

- NNAR(27,1,14)[365] without external variable
- **NNAR(27,1,15)[365]** with external variables national GDP
- NNAR(27,1,15)[365] with external variables unemployment rate
- NNAR(27,1,16)[365] with GDP and unemployment rate
- NNAR(8,1,5)[365]

Using all data available

Training MAPE - 0.3l, Testing MAPE - 4.77

- NNAR(28,1,15)[365] without external variable
- NNAR(28,1,16)[365] with external variables national GDP
- NNAR(28,1,16)[365] with external variables unemployment rate
- NNAR(28,1,16)[365] with GDP and unemployment rate
- **NNAR(8,1,5)[365]**

Linear Regression & Decision Tree

Linear Regression

Use external variables (GDP & Unemployment rate) as X to build linear regression model:

Only using Pre-COVID Data
MAPE - 2.15

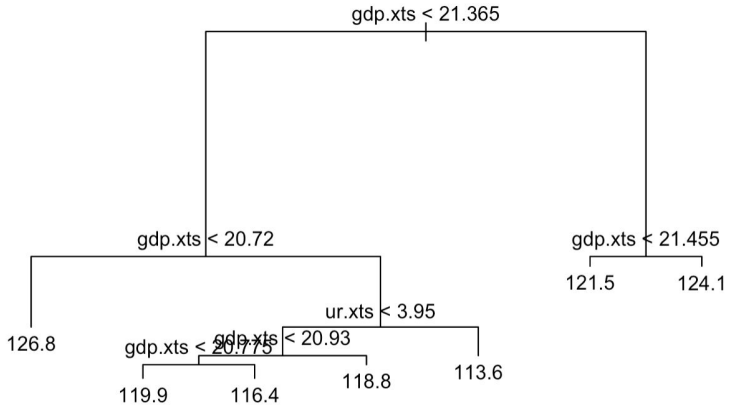
- **tslm(price~trend+ur+gdp)**
- tslm(price~ur+gdp)
- tslm(price~ur)
- tslm(price~gdp)

Using all data available
MAPE - 1.97

tslm(price~trend+ur+gdp)
slm(price~ur+gdp)
slm(price~ur)
tslm(price~gdp)

Decision Tree

Decision trees partition all possible values of the attributes into different subcategories or regions and each subcategory is assigned a value. By splitting variables we can predict the Airbnb price through a decision tree model.



Only using Pre-COVID Data
Training MAPE - 1.82
Testing MAPE - 2.07

Using all data available
Training MAPE - 1.69
Testing MAPE - 1.78

Section 5
Conclusion

Key Takeaways

Two functions
Assess the impact of COVID-19 pandemic & Predict future values on a 30-day forecasting horizon
Cross-validation to smooth out variability and give a more accurate evaluation

Multiple Methods
Price-only - Naive forecasting, Holt Winters, ARIMA
With external variables (GDP and Unemployment Rate) - NN, Linear Regression, Decision Tree

Best model for Forecasting
SARIMA(0,1,0)(1,1,1)[7]
A robust and accurate result with a training MAPE of 0.61% and a validation MAPE of 1.16%

Scalable, Valuable and Multipurpose

Externally

Collaboration with government

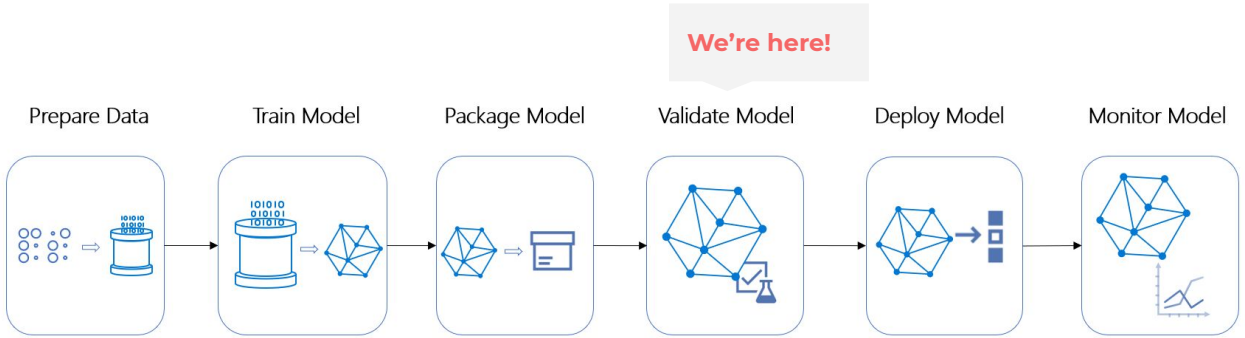
Listers & Travellers

Internally

Executives & Managers

Business Development

Next Steps



Deploying the forecasting model & **Architecturing** a pipeline

1	Data storage and retrieval
2	Frameworks and tools
3	Feedback and iteration

Thank You!