Data Analysis

CHRIS MOREHOUSE

Analysis Process

1. Business evaluation
   - Initial hypothesis that, in this business, there is seasonality around job requests (fewer jobs in colder months vs. warmer months)
2. Data exploration
   - Visually and programmatically inspect the data
3. Data preparation
   - Clean data
   - Filter data for relevant data points only
4. Data modeling
   - With limited data (no sales funnel, customer data, etc.) I will deploy a time series forecasting model
5. Data evaluation
   - Chart and interpret results
   - Test accuracy of forecasts and calculate errors in forecasts vs. actuals

Forecast Methodology

Model considerations: Facebook Prophet

- Prophet includes many different forecasting techniques (ARIMA, exponential smoothing, etc). Given the expertise required to choose the correct model, I chose Prophet for its scalability and ease of use
- Benefits
  - Model handles hourly, daily, or weekly observations with at least a few months of history
  - Handles historical trend changes, for instance due to product launches
  - There are smoothing parameters for seasonality that allow you to adjust how closely to fit historical cycles
  - For growth curves, you can manually specify “capacities” or the upper limit of the growth curve, allowing you to inject your own prior information about how your forecast will grow (or decline)
  - You can specify irregular holidays to model, like extreme weather events
- Limitations
  - Doesn't catch interactions between external features, which could improve the forecasting power of a model. In this case, these variables could be marketing spend or new channel partners
  - Fits and extrapolates on fixed trends

Assumptions and Additional Considerations

Assumptions:

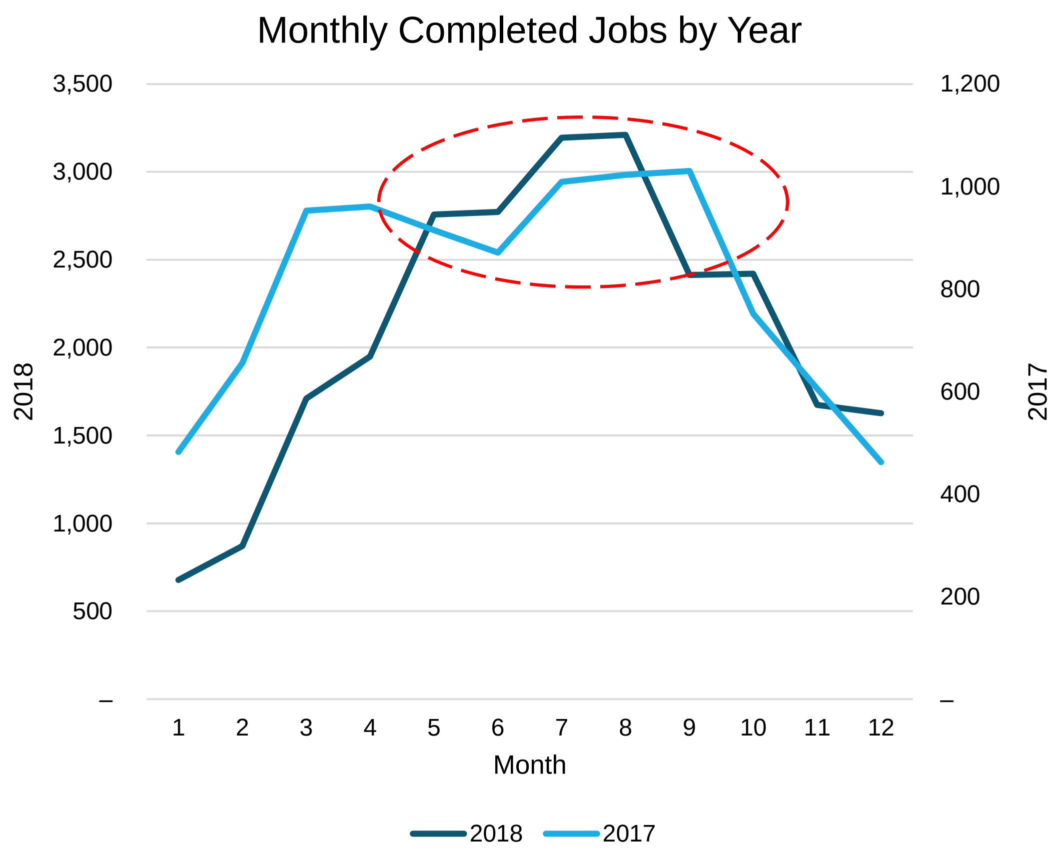
- Only storms where the description includes words such as 'home', 'building', 'property', 'roof', 'garage', 'wall', 'house', 'residence', or 'shingle' will be considered as a storm that caused damage that would require a job
- The storm's "Location" must take place in the same city/state as the job to be considered as a storm that caused damage that would require a job
- "Holidays" – storms and job windows are expected to occur within a 14 day +/- window, when compared to the previous year, in the future

Select Additional Considerations:

- Marketing spend and efforts
- Expected new channel partnerships and growth of existing channel partners
- Detailed customer information (at the organization level)
- Sales headcount expansion

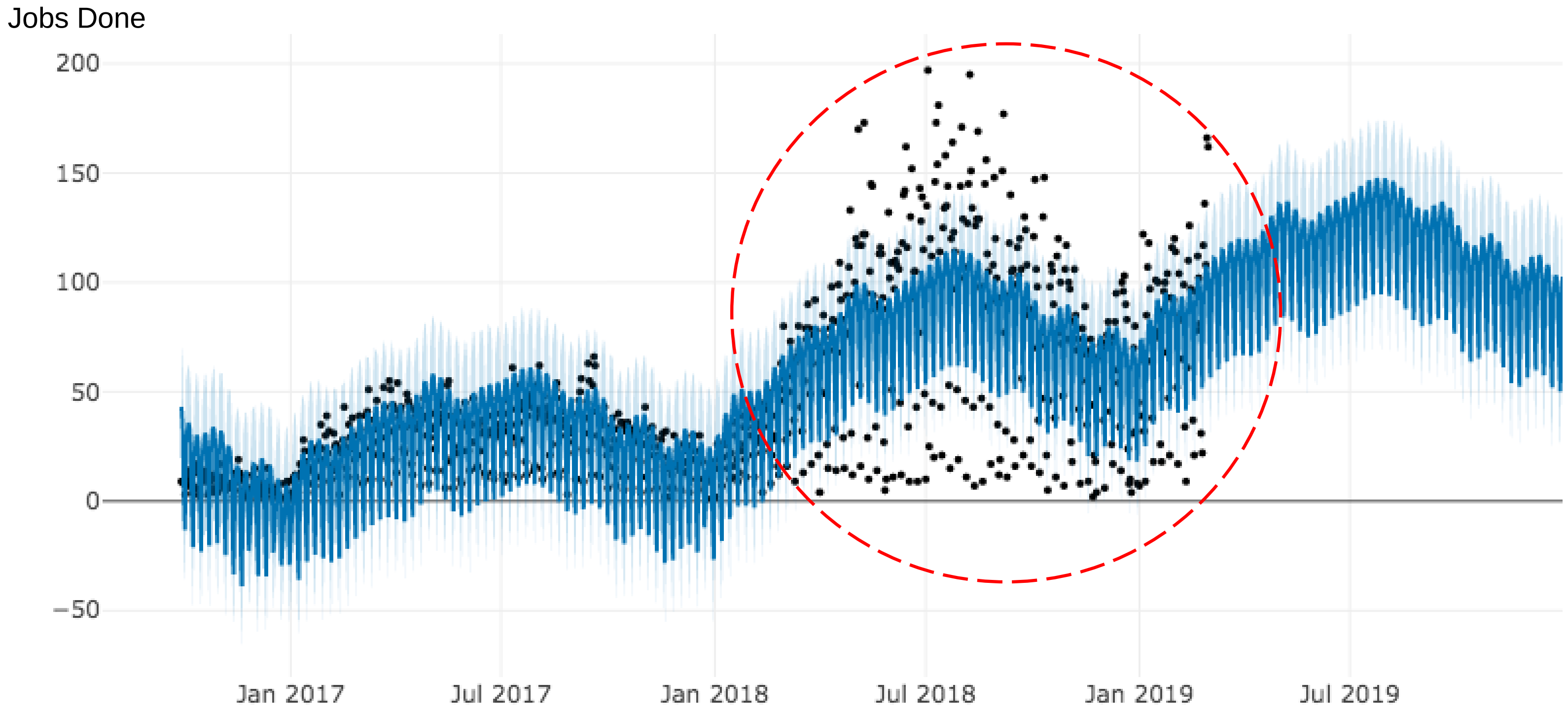
Overall Demand in 2019

Review of Data



- For the two years with complete data, there is a clear trend of seasonality
- The visual distribution of the data supports my initial hypothesis that individuals are less likely to seek contractor work during winter months
- Additionally, the data points to a significant uptick/acceleration of work in 2018

Forecast Outputs – Without Log Normalization

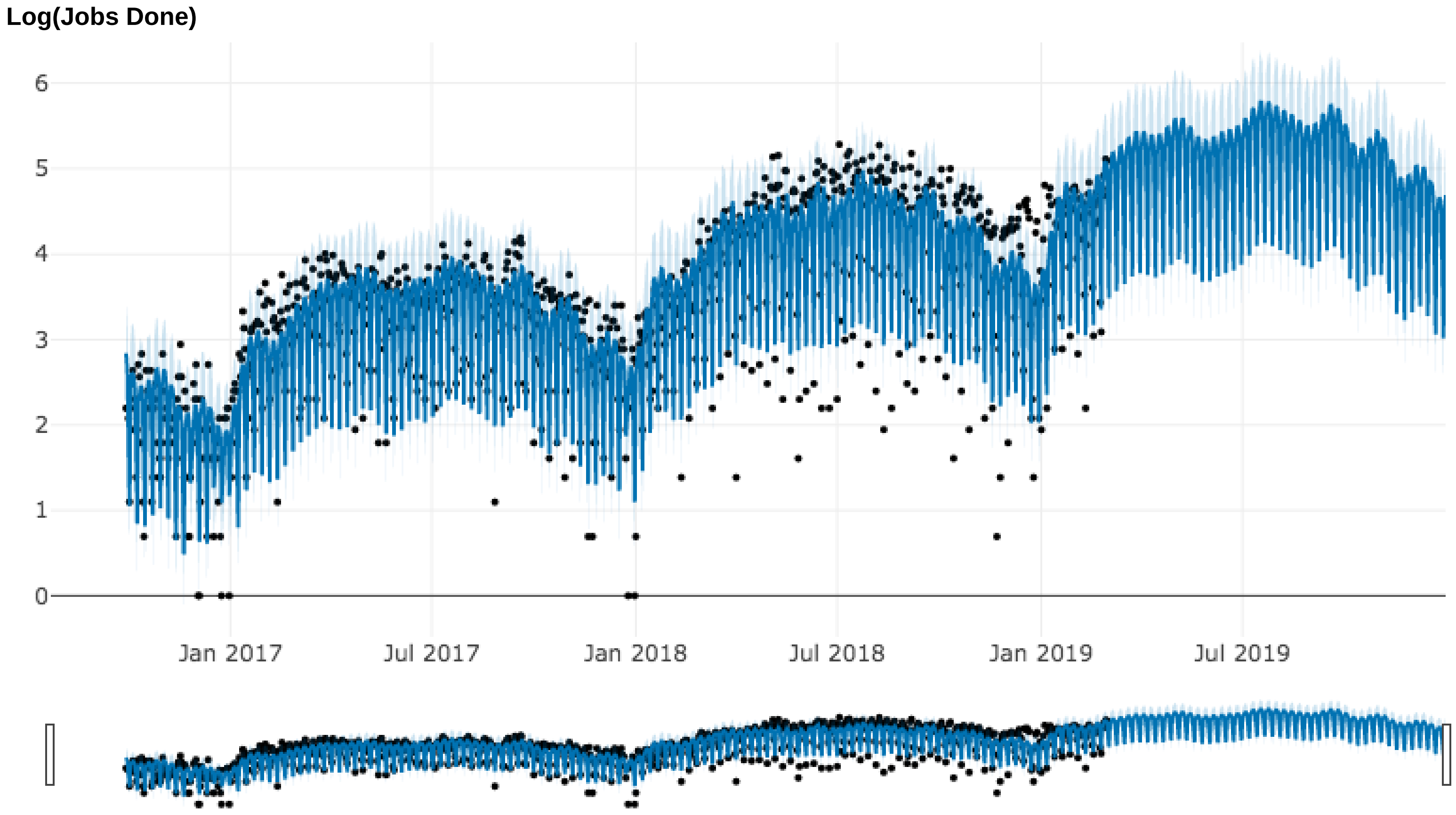


Note: black-circular points are actuals

Does not contain “holiday” events

- Root Mean Square Error: 21.09
- Mean daily jobs done in historical data is ~46
- Prior to 2018, the model does a relatively good job fitting to historical data
- Volatility in 2018 shows a divergence in model accuracy in 2018
- Next step is to reduce noise by normalizing the data with log

Forecast – Total Jobs (Log Normalization)



After fitting the initial model, I added weather events as holidays and looked to reduce volatility by taking the log of the data

Assumptions:

- Weather events that occurred in the same cities/states where Company A operates
- Applying weather events as holidays that will occur within a 2 week +/- window in 2019 as they did in historical events

Accuracy:

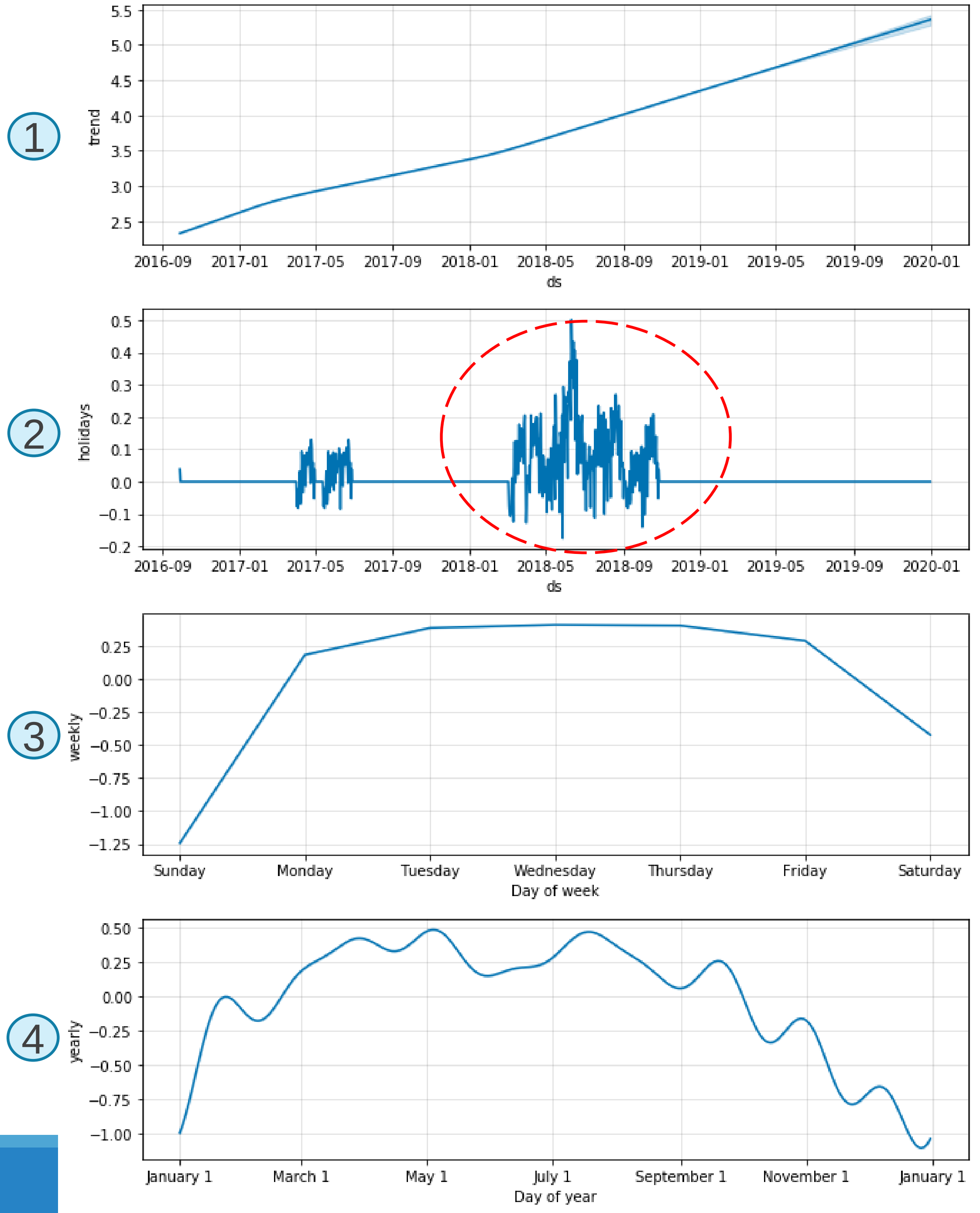
- Adding weather events (as holidays) and log normalizing the data yielded the most accurate forecast to date
- The log of the data reduces the noise and normalizes the data
- **Root Mean Square Error: 17.50**

Select Limitations:

- Assumed relative stability in timing of weather events from 2018 to 2019

Note: black-circular points are actuals

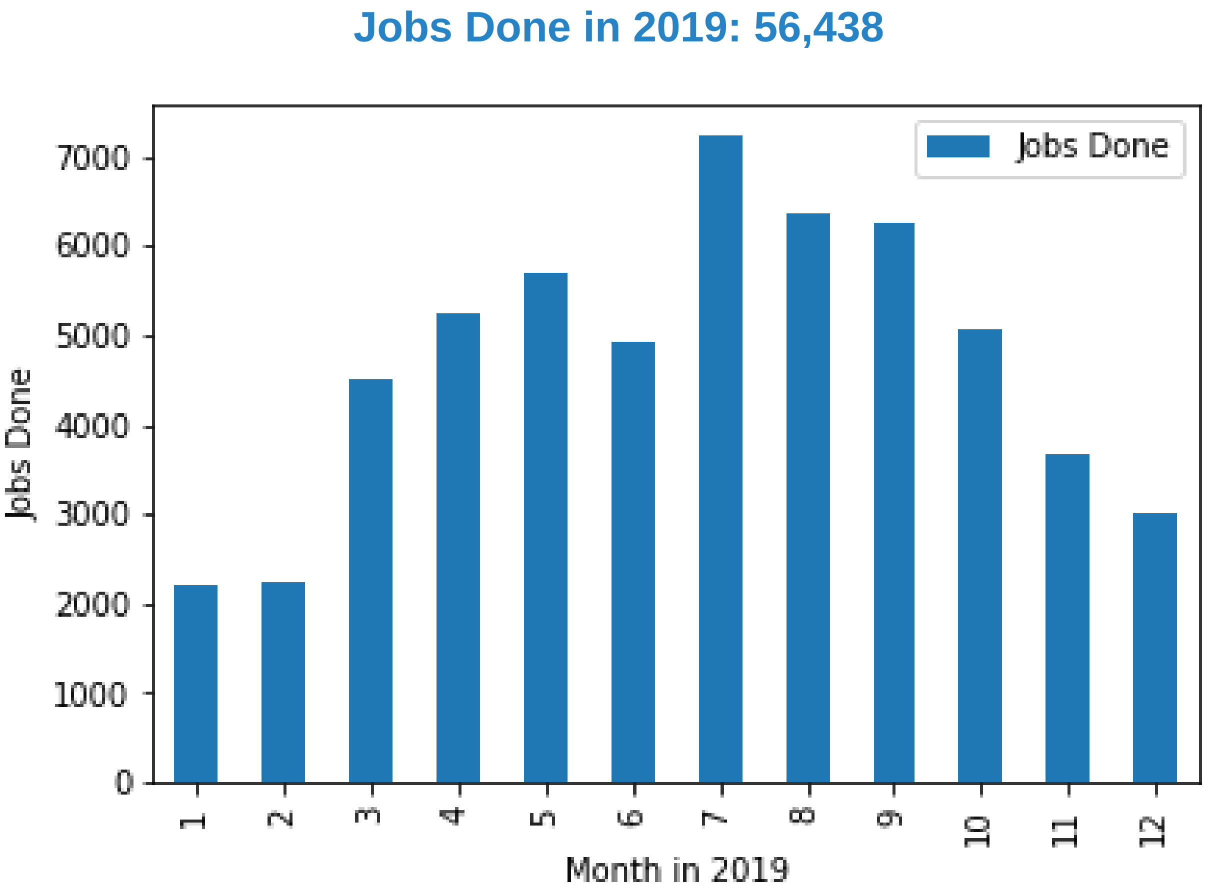
Population and housing sources: US Census bureau and Statista



Forecast Outputs – Trends and Seasonality

1. Linear growth trends for Total Jobs Done
2. Clear “holiday” impact of weather events on the data
3. Surprisingly, weekly seasonality indicates peak jobs on Wednesday. This is likely a function of contractors not operating on weekends rather than consumer demand
4. Prophet picked up similar monthly seasonal trends as the data on slide 6, with peak demand in the summer

Jobs Forecast 2019: Log Normalize and Weather Events



Output:

- **56,438 jobs in 2019** are expected when log normalizing and accounting for weather events

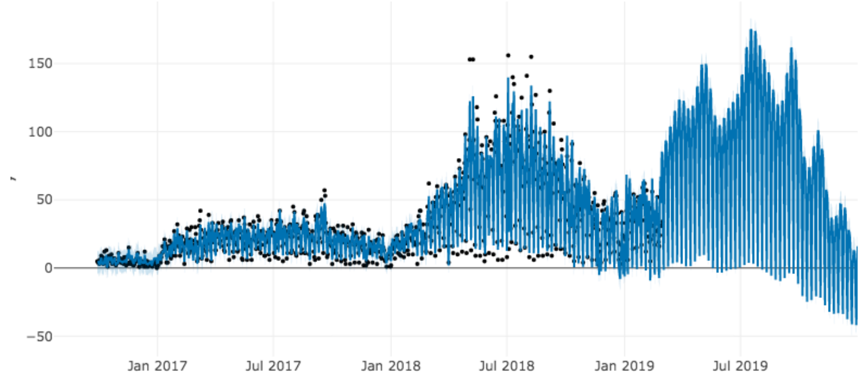
Additional Considerations:

- Fitting to a logistic growth curve for longer term forecasts with a capacity rate of the number of housing units in the state of operations

Roof vs. Complete Demand in 2019

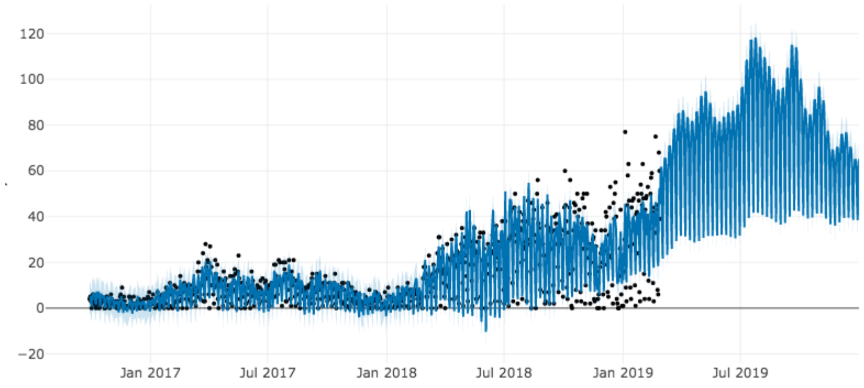
1

Complete Jobs



2

Roof Jobs



Roof and Complete Forecast

While normalizing the data for total jobs brought down the error rate, in both instances of sub-group jobs, the error rate was lowest without log data

1 **Complete**:

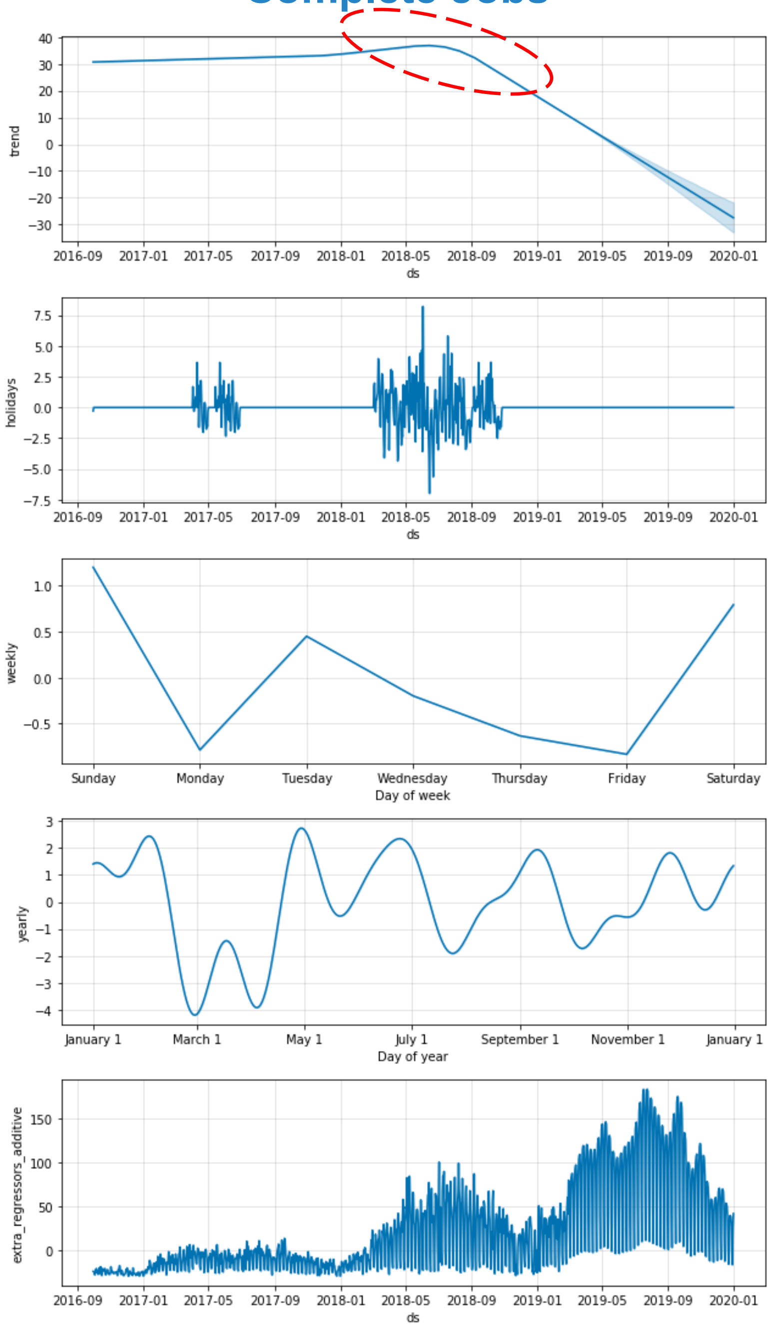
- Root Mean Square Error: 6.37
- On some days the Complete jobs forecast turns negative. Given that this is impossible, I set negative estimates to 0

2 **Roof:**

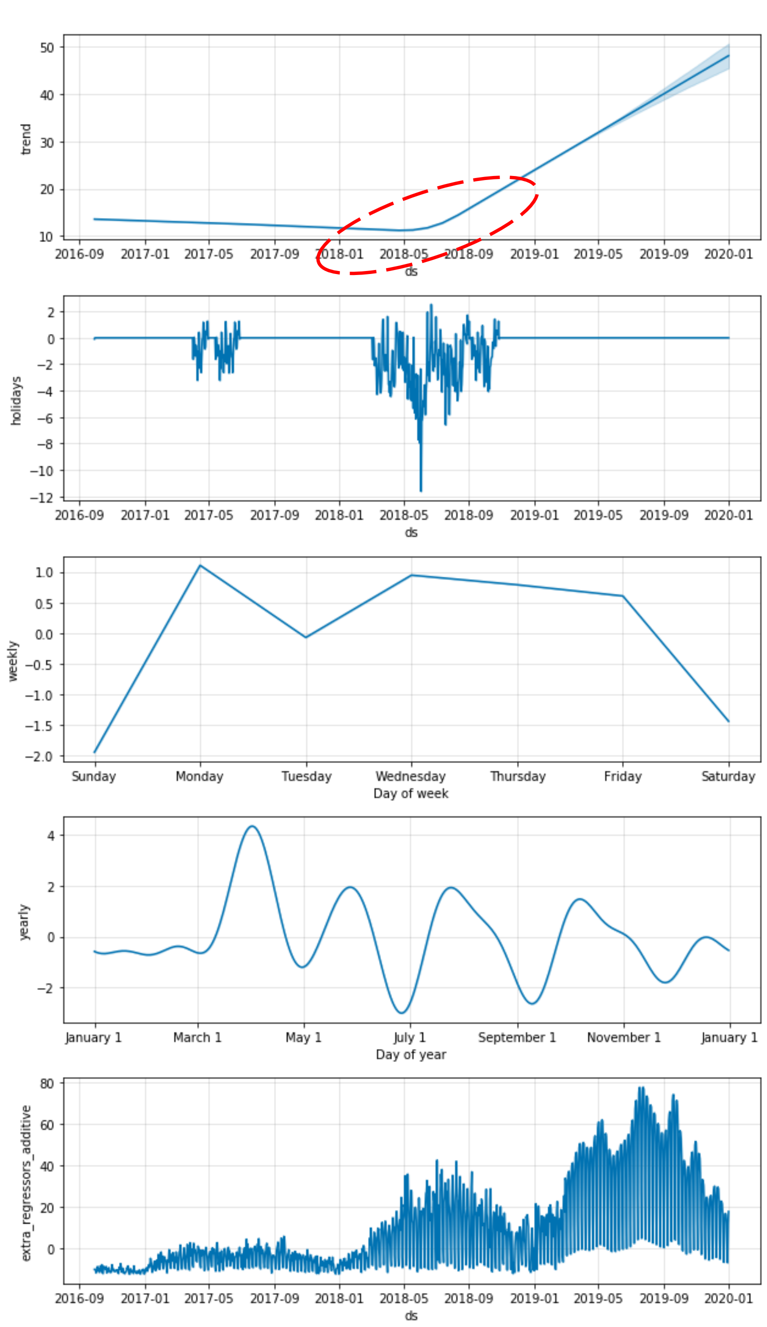
- Root Mean Square Error: 5.33

Diverging Linear Trends in Complete and Roof Jobs

Complete Jobs



Roof Jobs

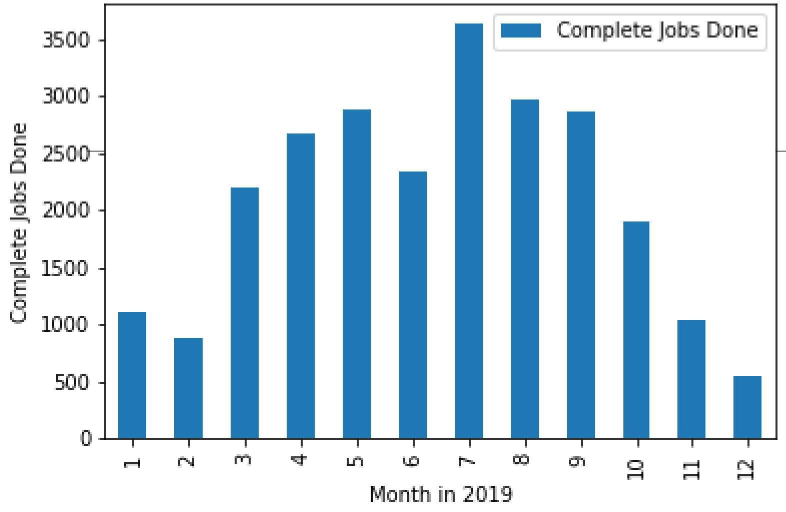


Observations:

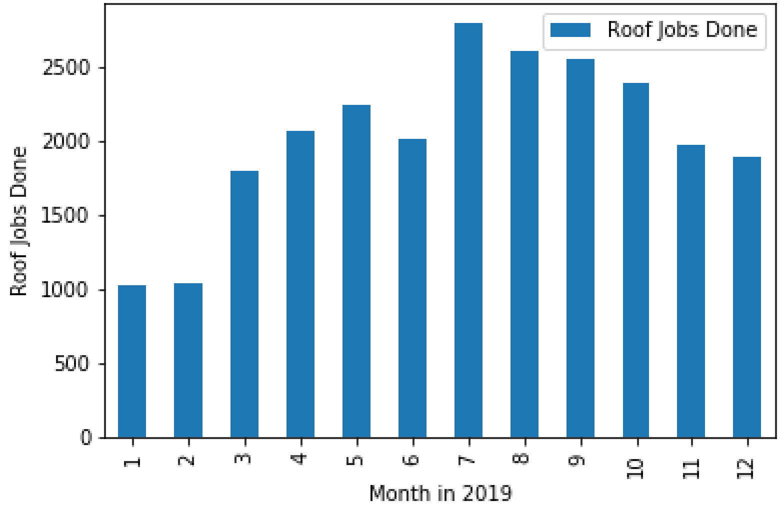
- In the data, Complete jobs are slowing down in growth more than Roof jobs
- The model is extrapolating a large downward trend in complete jobs, which for a longer term forecast is most likely an inaccurate fit given the overall growth of the business

2019 Jobs Done Forecast (By Type)

1 **Complete Jobs in 2019: 25,041**



2 **Roof Jobs in 2019: 24,385**



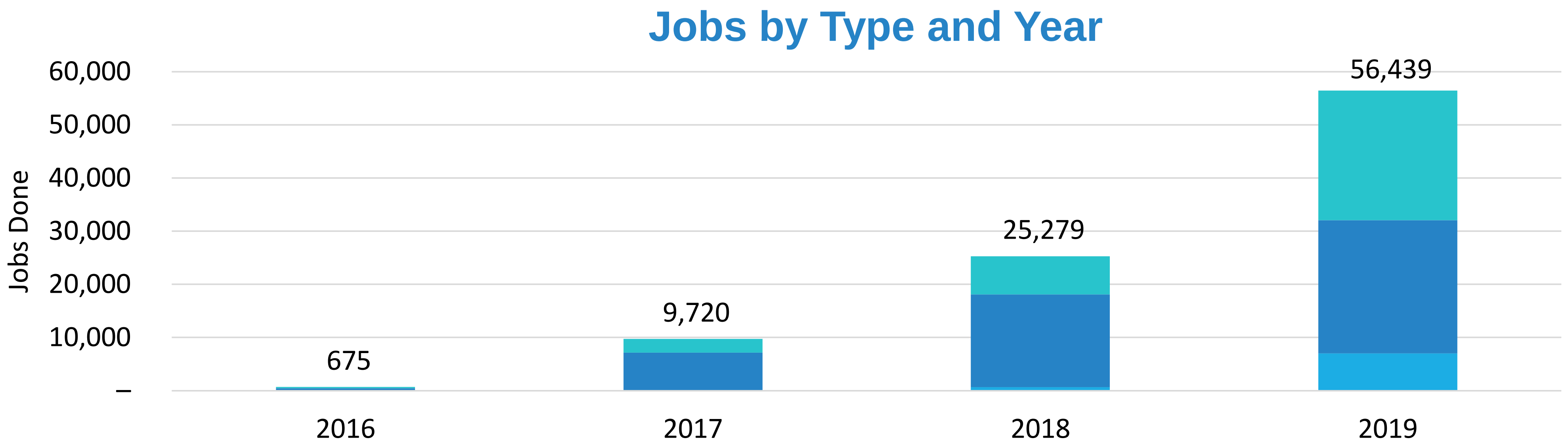
1 **Complete**:

- **25,041 jobs in 2019** are forecasted when adding the total jobs forecast and weather events to the model
- Select limitations – With a growing business, I find it strange that the model is projecting such a steep negative trend.
  - There might be additional regressors that need to be modeled to tell a complete story with Complete jobs

2 **Roof:**

- **24,385 jobs in 2019** are forecasted when adding the total jobs forecast and weather events to the model

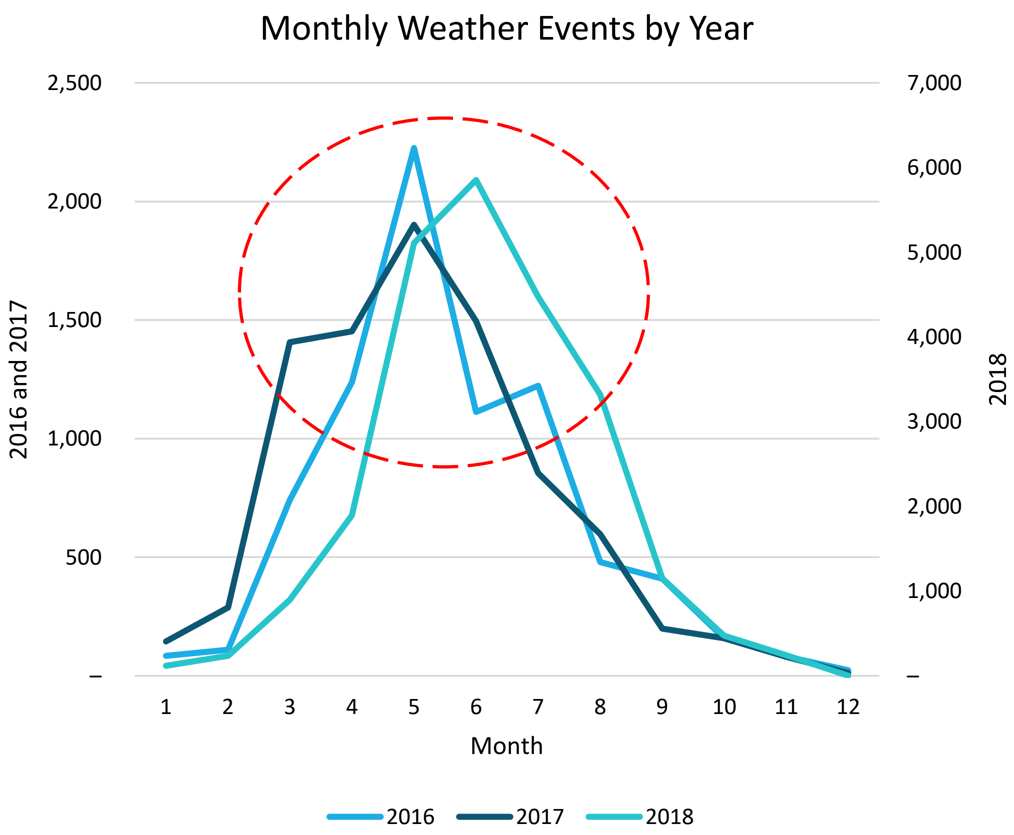
2016 to 2019 Annual Jobs Done



Growth	2017A	2018A	2019E
Total	1,340.0%	160.1%	123.3%
Roof	1,008.7%	179.5%	240.7%
Complete	1,504.7%	145.5%	43.1%
Other	NA	1,741.2%	1,020.0%

Demand From Weather

Review of Data and Assumptions

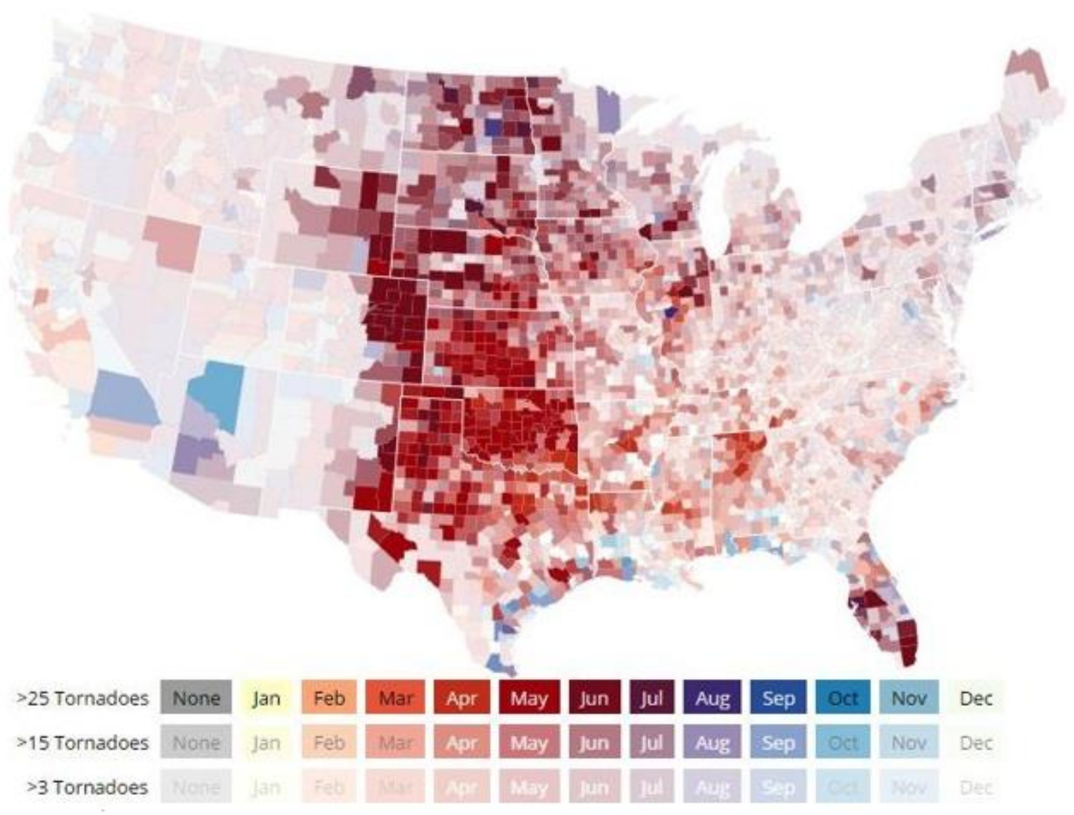


- Over the three years of complete data, there is a clear seasonal pattern in reported weather events
  - Peak reports happen in May of 2016-2017 and June of 2018
- 2018 saw a sharp uptick in the number of weather events
- Given global trends, severe weather events are likely to continue and are likely to rise

Assumptions:

- Weather events and job windows are calculated as +/- 14 days of historical weather events

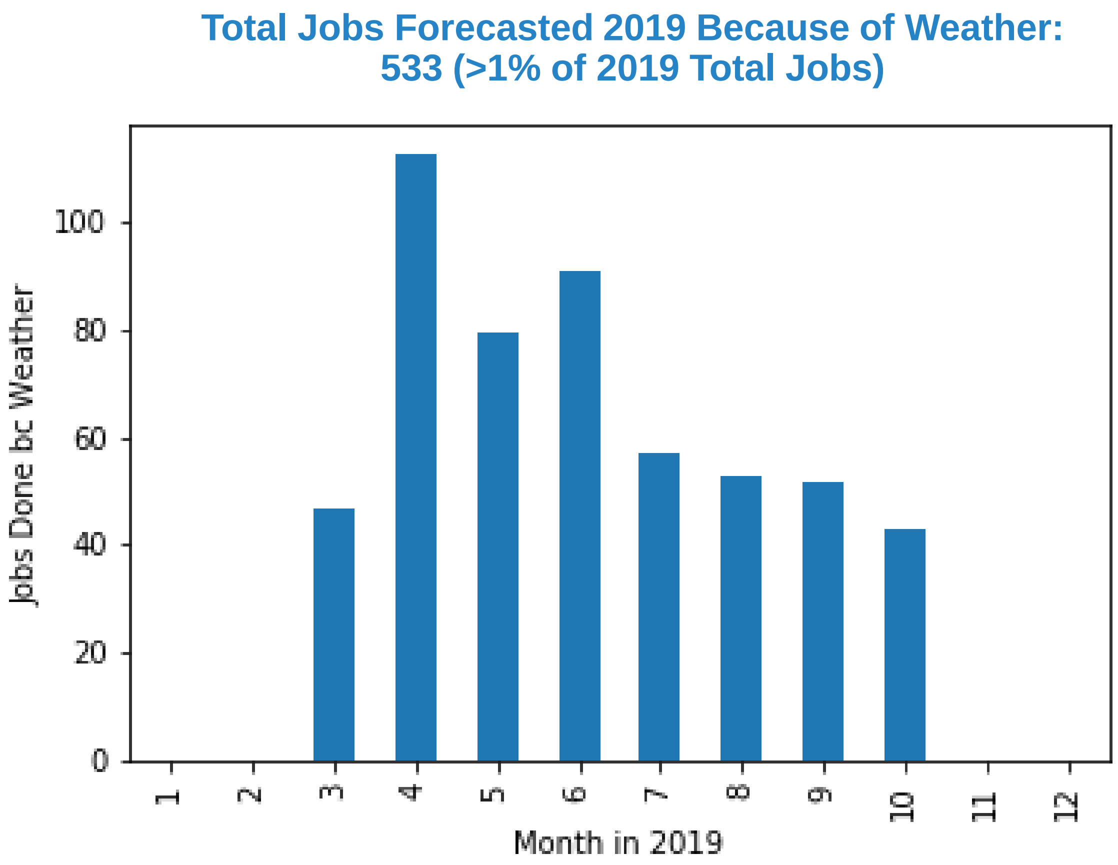
U.S. Tornado Frequency



- States of operation for Company A:
  - TX, NE, SD, KS, ND, OK
- All of the states of operation are in tornado territory and tend to suffer from extreme weather events
- The seasonal job and weather event data also corresponds with tornado season in the service areas
- April-June for most extreme weather events

Source: http://www.ustornadoes.com/2015/04/22/the-month-with-the-most-tornadoes-across-the-country/

Jobs because of Weather



- Using the “holiday” (weather event) impact on jobs done in 2016-2018**, jobs done because of weather is forecasted to be 533**
- The same monthly distribution of “holiday” impact from 2016-2018 is applied to 2019 to get a monthly distribution of weather jobs from the forecasted demand figures

2 Call out (2) from slide 9 shows the impact weather events had on demand from 2016-2018

- Manually and programmatically reviewing the data, it appeared that 161 weather events caused jobs in 2018 (>1% demand)

Additional Considerations

Geocoding:

- Attribution was only applied to weather events that occurred with a city/state match between jobs.csv and weather.csv
- Geocoding (or reverse geocoding) could utilize lon and lat coordinates to better identify the storm zone

Attribution Window:

- Window adjustment – More industry expertise could be used to adjust the window of when contractors typically handle a job after a storm
- Job type – A single attribution window was applied to all jobs, but there might be different windows for different job types. For example, roof claims may need to be addressed sooner than wall claims