

HOVER

Gustavo Oliveira

Job Prediction - 2019

Walk Through

- Introduction
- The Data
- Assumptions
- Analysis Strategy
- Feature Engineering
- Exploratory Analysis
- Predictive Modeling
- Results
- Key Takeaways
- Actionable Insights
- Future Analysis

Introduction

- Goal of the analysis is to answer:
 - How many jobs can Hover expect in 2019?
 - How many of these are complete jobs/ roof jobs?
 - What volume of jobs comes from weather events?

The Data

- Two Types of Data - Jobs and Weather
 - Both relatively clean data
- Jobs:
 - ~40k rows / 6 columns (2 numeric, 3 categorical, 1 date)
 - Important features:
 - Region: TX(Texas), NE(Nebraska), OK(Oklahoma), KS(Kansas), ND(North Dakota), SD(South Dakota)
 - Types of Jobs: Roof, Complete (main ones)
 - Date photos were uploaded for the first time.
- Weather
 - ~40k rows / 9 columns (5 categorical, 1on, lat, 2 date)
 - Important features: lon,lat,date,time
 - Comments interesting but would need more processing (nlp)

Assumptions

- Every type of weather event (independent of type) can spark a job
- Every job is considered unique (could evaluate if from same company, uploaded on the same day, etc.)
- Job Regions = States
- Weather event has a radius of 15miles (could have a decay)
- Photos can take up to a week to be uploaded depending on the damage from the event.

Analysis Strategy



Create variables that are relevant and might help understand and model the data

Visualize, analyze the data to find and understand patterns

Develop, test, evaluate and run models to forecast number of jobs

Feature Engineering

- Associate jobs to weather events.
 - If photos uploaded up to 1 week of event



Assumption:

- Jobs less than 15 miles from event AND that were uploaded up to a week from the event are related to that event

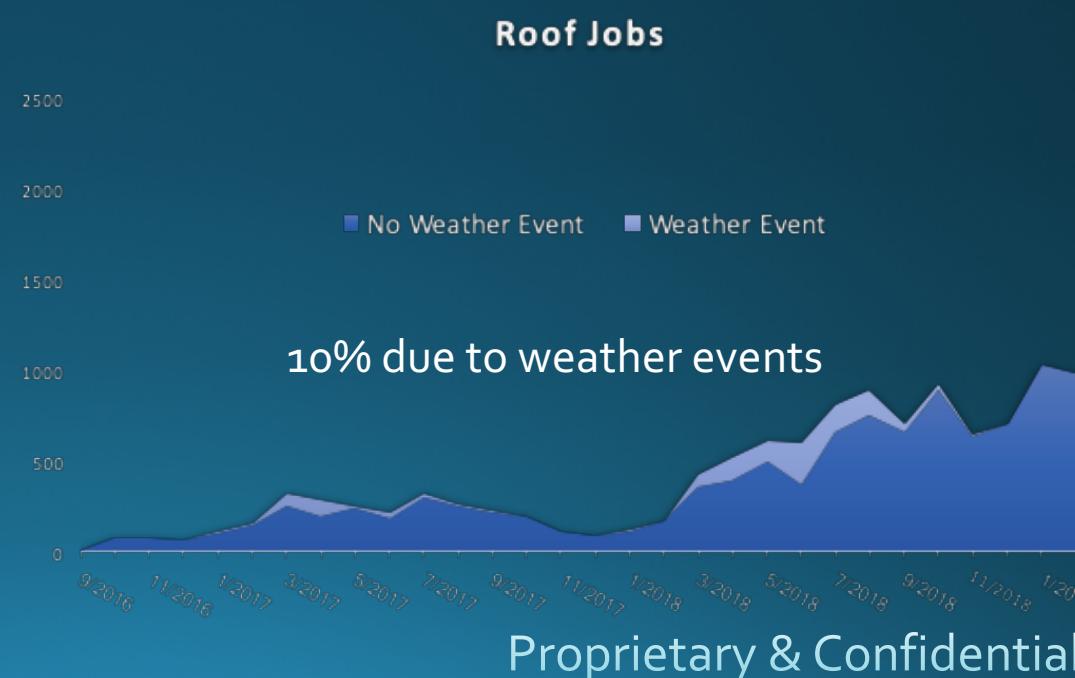
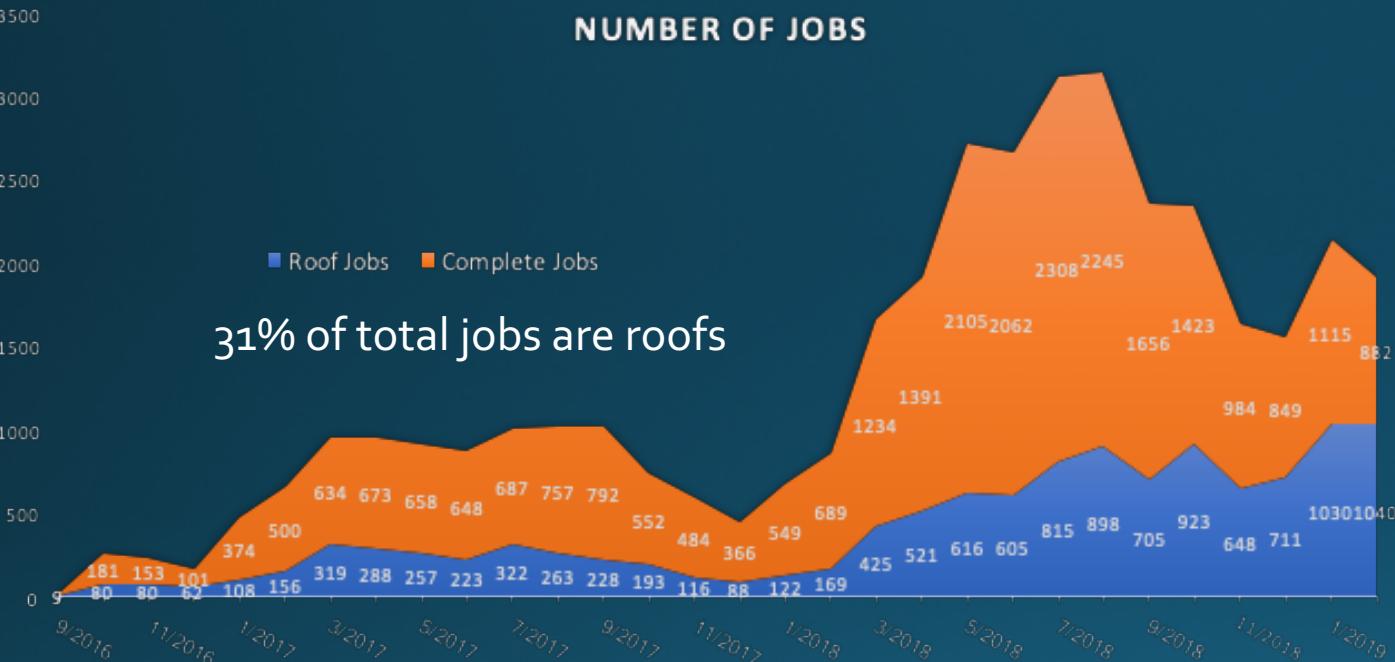
Feature Engineering



<https://simplemaps.com/data/us-cities>

Exploratory Analysis

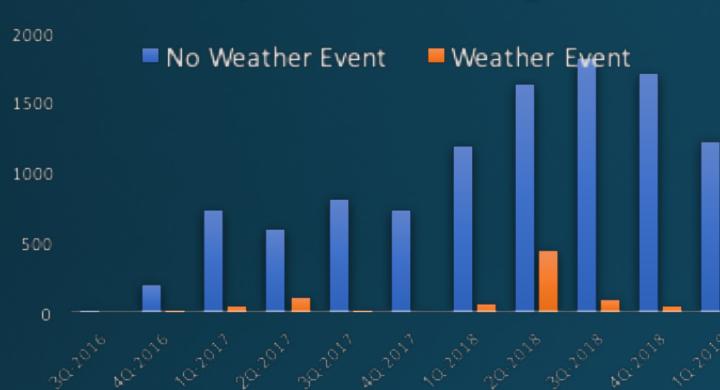
HOVER



Growth YoY	Roof		Complete	
	No Event	Event	No Event	Event
18/17	173%	217%	123%	371%
19/18	78%	835%	72%	381%

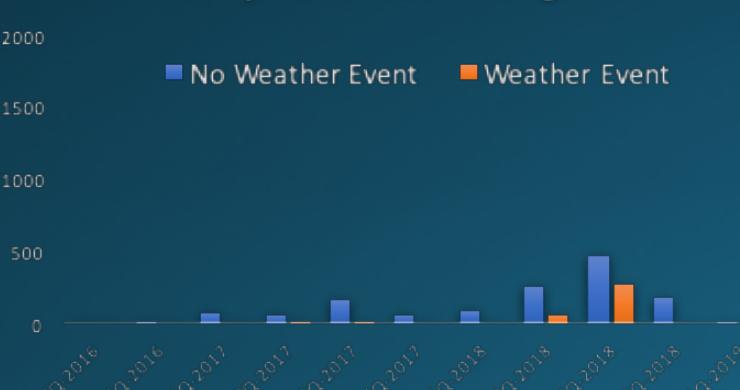
Exploratory Analysis – Complete by Region

Complete Jobs - TX Region



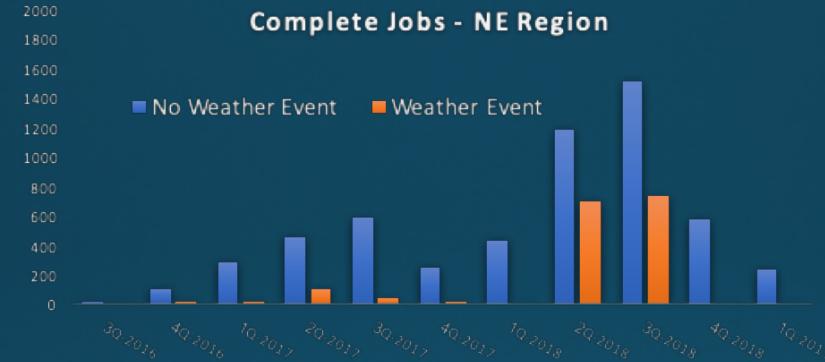
11374/7% (Total Jobs/Perc Weather)

Complete Jobs - SD Region



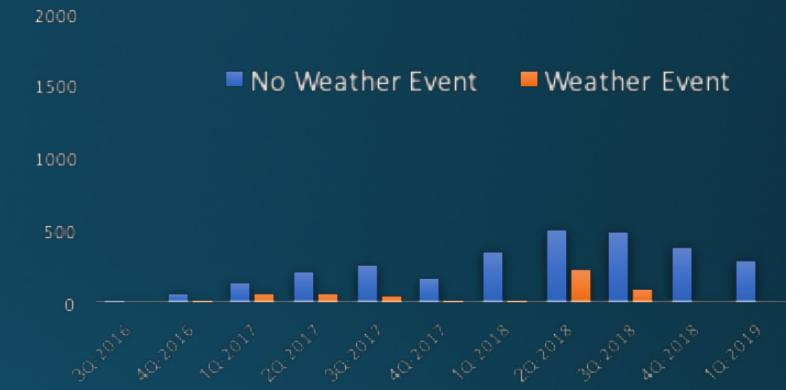
1728/21%

Complete Jobs - NE Region



702/22%

Complete Jobs - KS Region



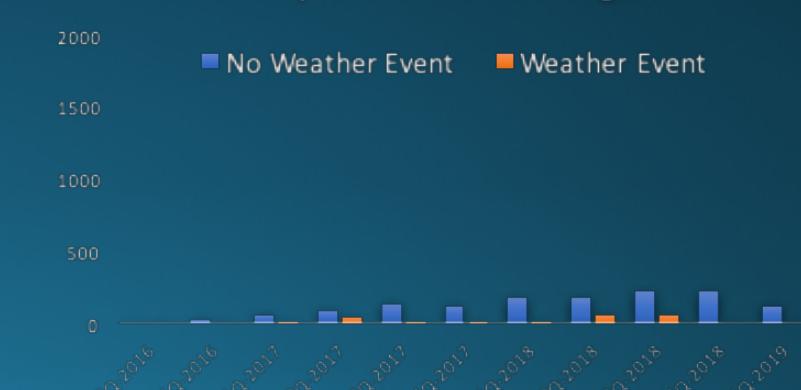
3328/15%

Complete Jobs - ND Region



1720/9%

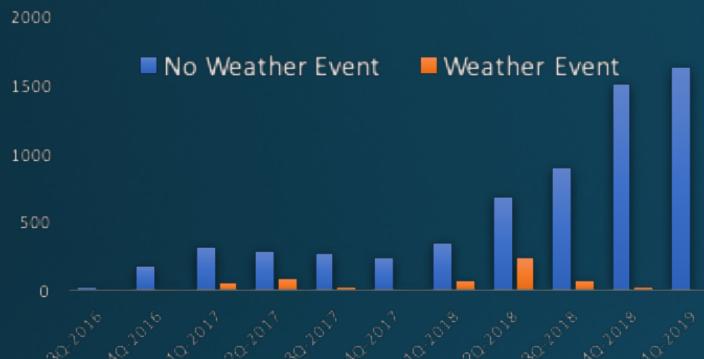
Complete Jobs - OK Region



1558/12%

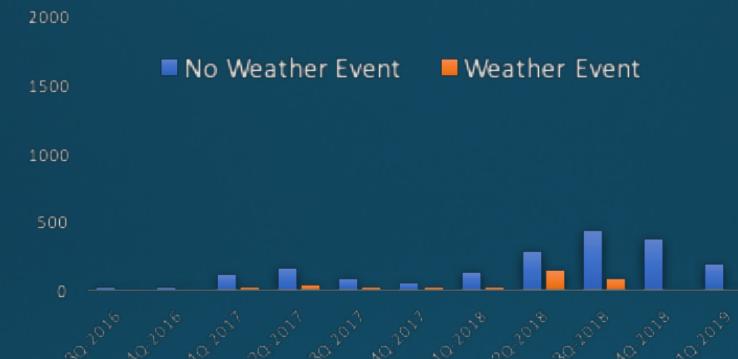
Exploratory Analysis – Roof by Region

Roof Jobs - TX Region



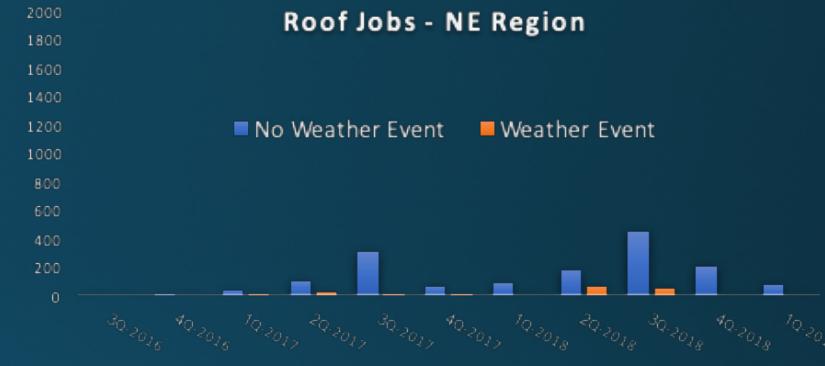
6819/8%

Roof Jobs - KS Region



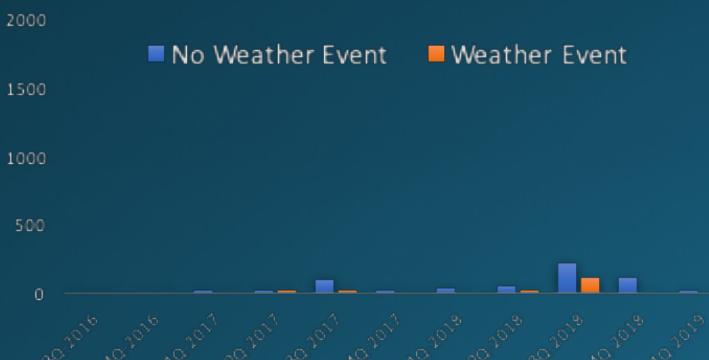
2135/14%

Roof Jobs - NE Region



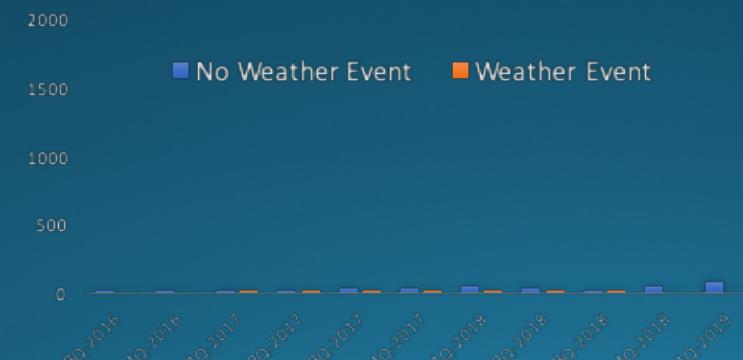
1646/9%

Roof Jobs - SD Region



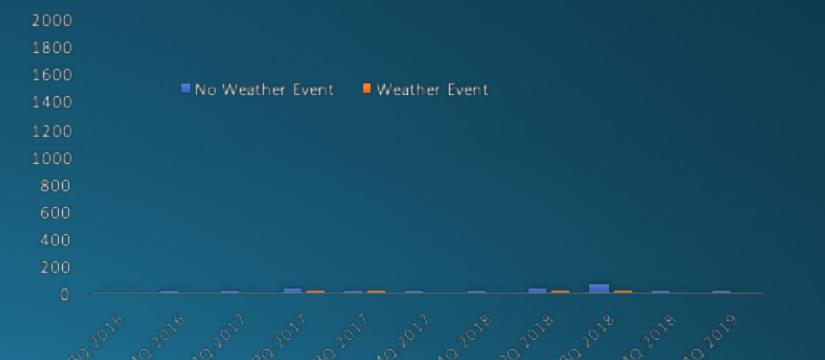
702/18%

Roof Jobs - OK Region



404/9%

Roof Jobs - ND Region

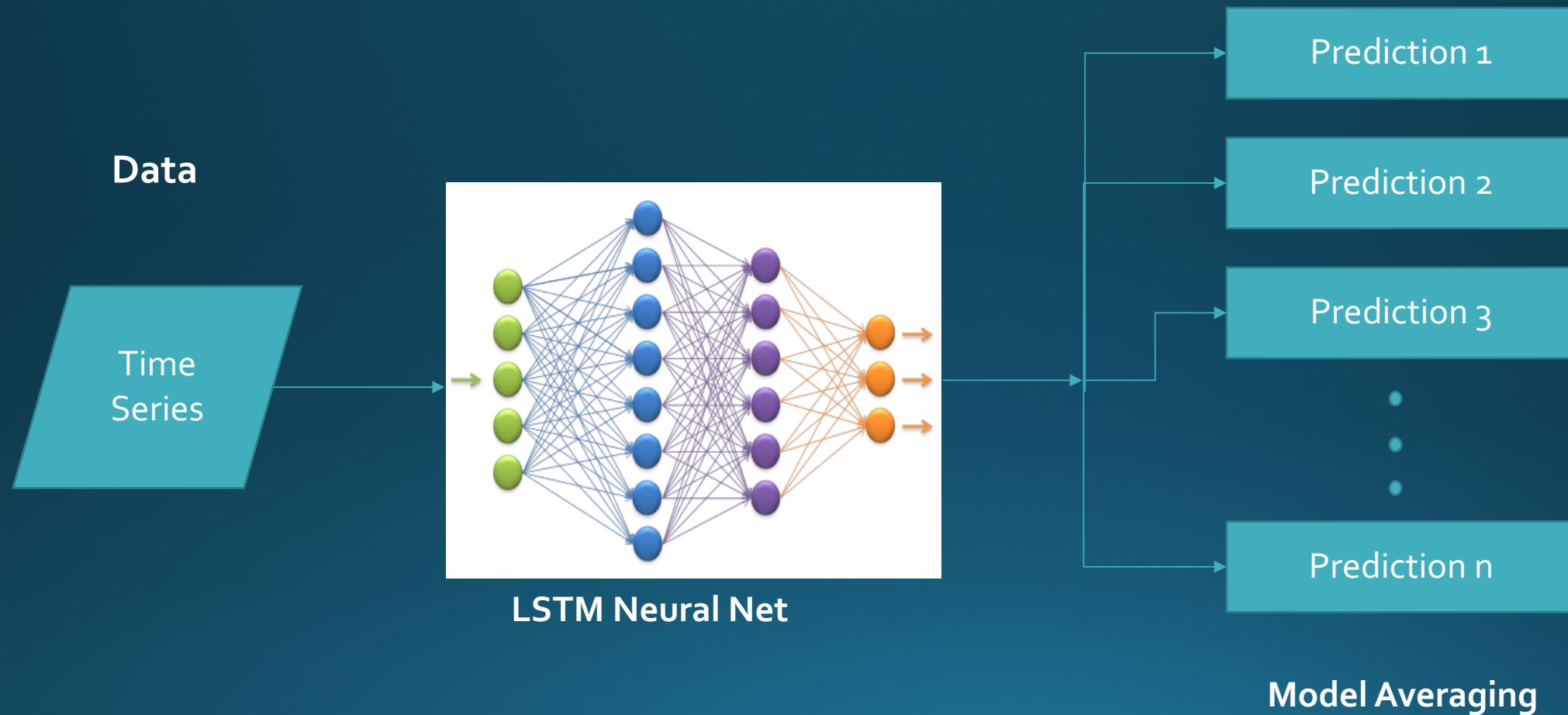


193/9% Proprietary & Confidential

Predictive Modeling

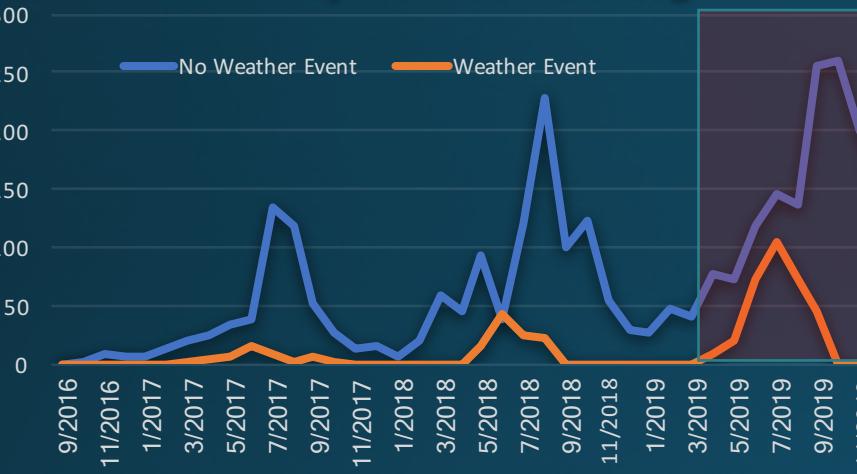
- Clearly an Univariate Autoregressive Problem
 - Predictive variable correlates with itself in certain lags
 - In other words – Number of jobs has a trend and/or seasonality (both)
- Possible Models:
 - ARIMA models – very laborious (Checking stationarity, differentiating, finding autocorrelation, etc.)
 - Linear Regression with Lagged Variables (Similar to ARIMA)
 - Long Short-Term Memory Neural Networks (model learns trend and/or seasonality – not transparent)
- Predict for Regions/Type of Job/Weather

Predictive Modeling - LSTM

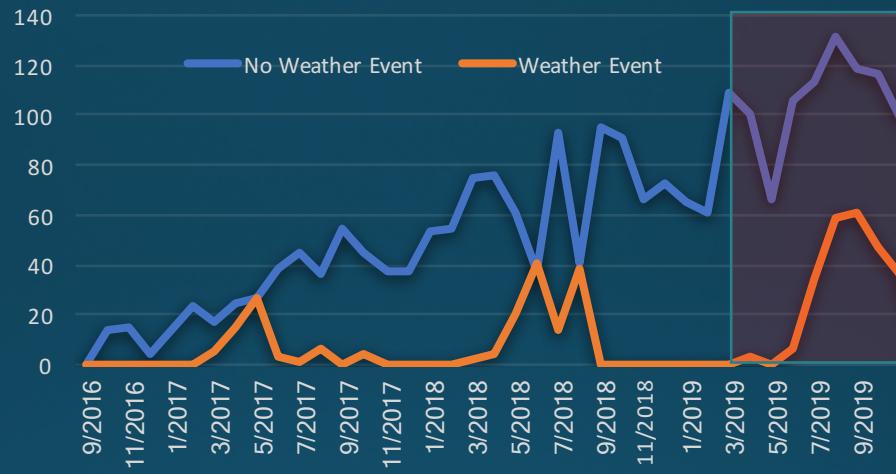


Results - Complete

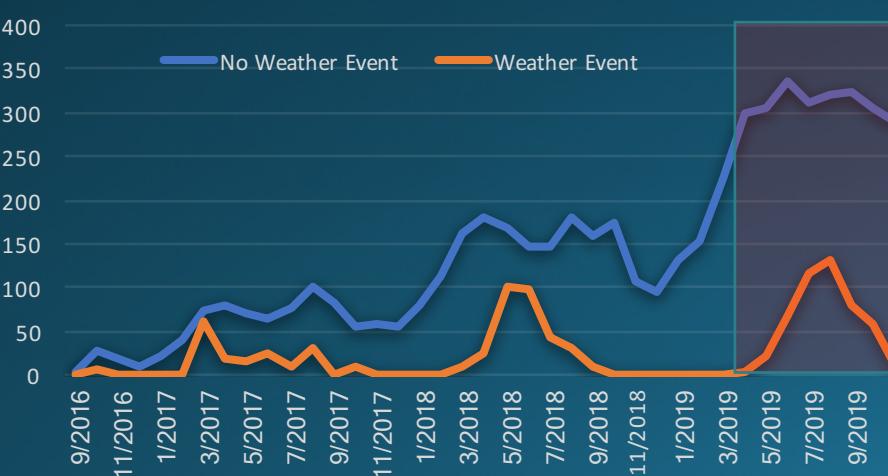
Complete Job - NE Region



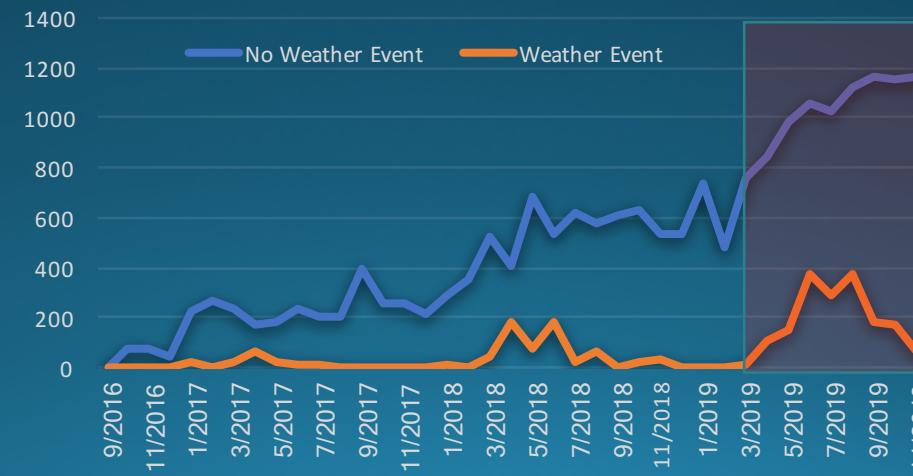
Complete Job - OK Region



Complete Job - KS Region

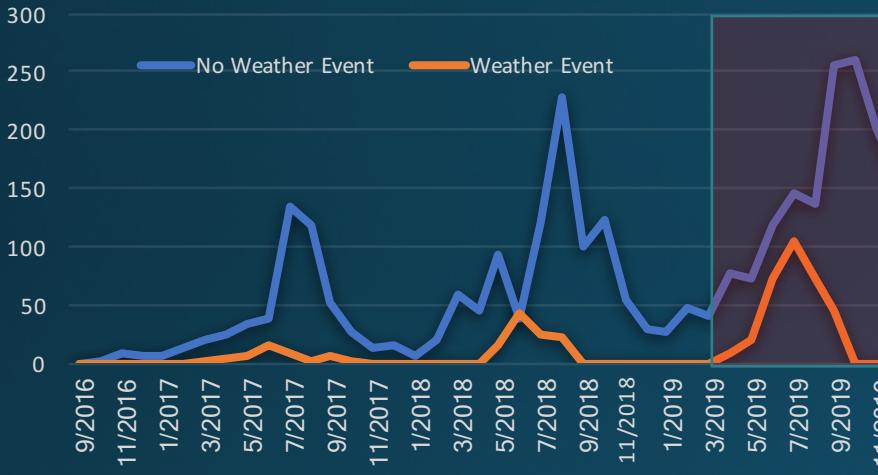


Complete Job - TX Region

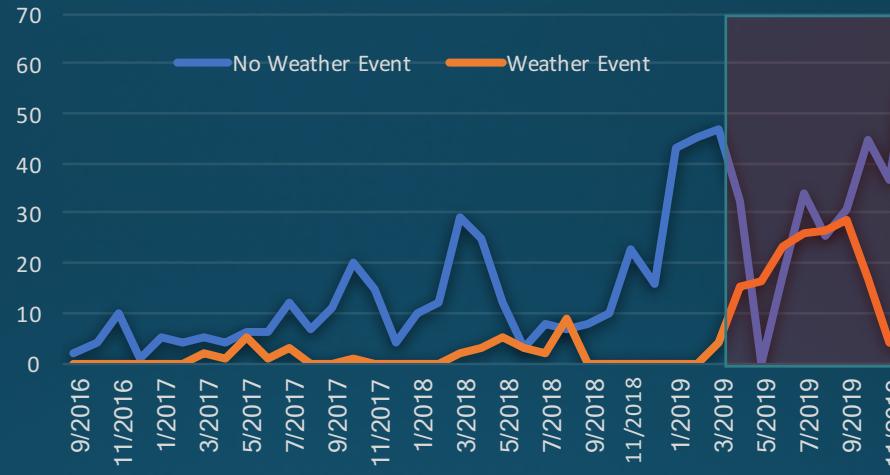


Results - Roof

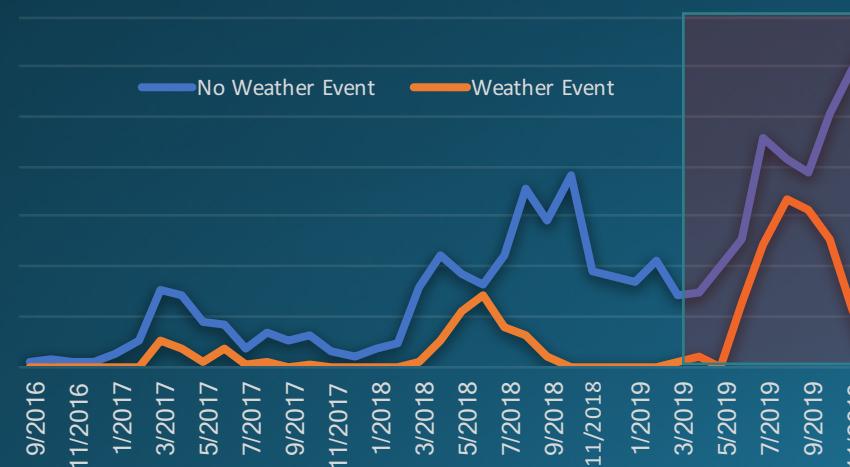
Roof Job - NE Region



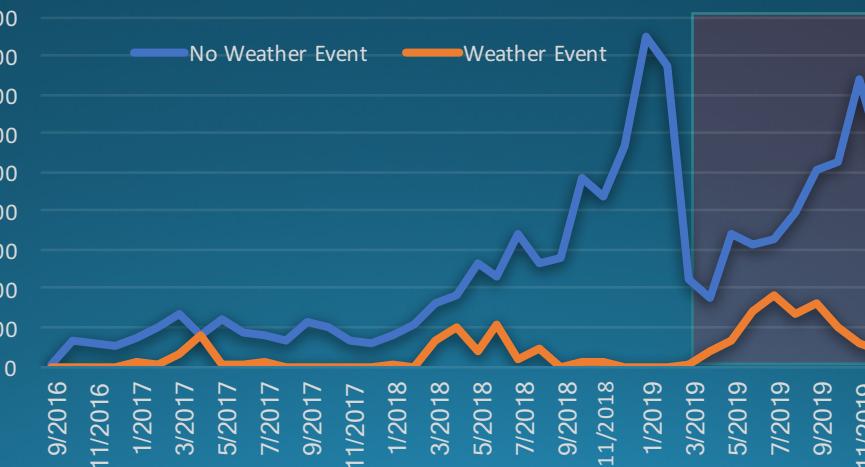
Roof Job - OK Region



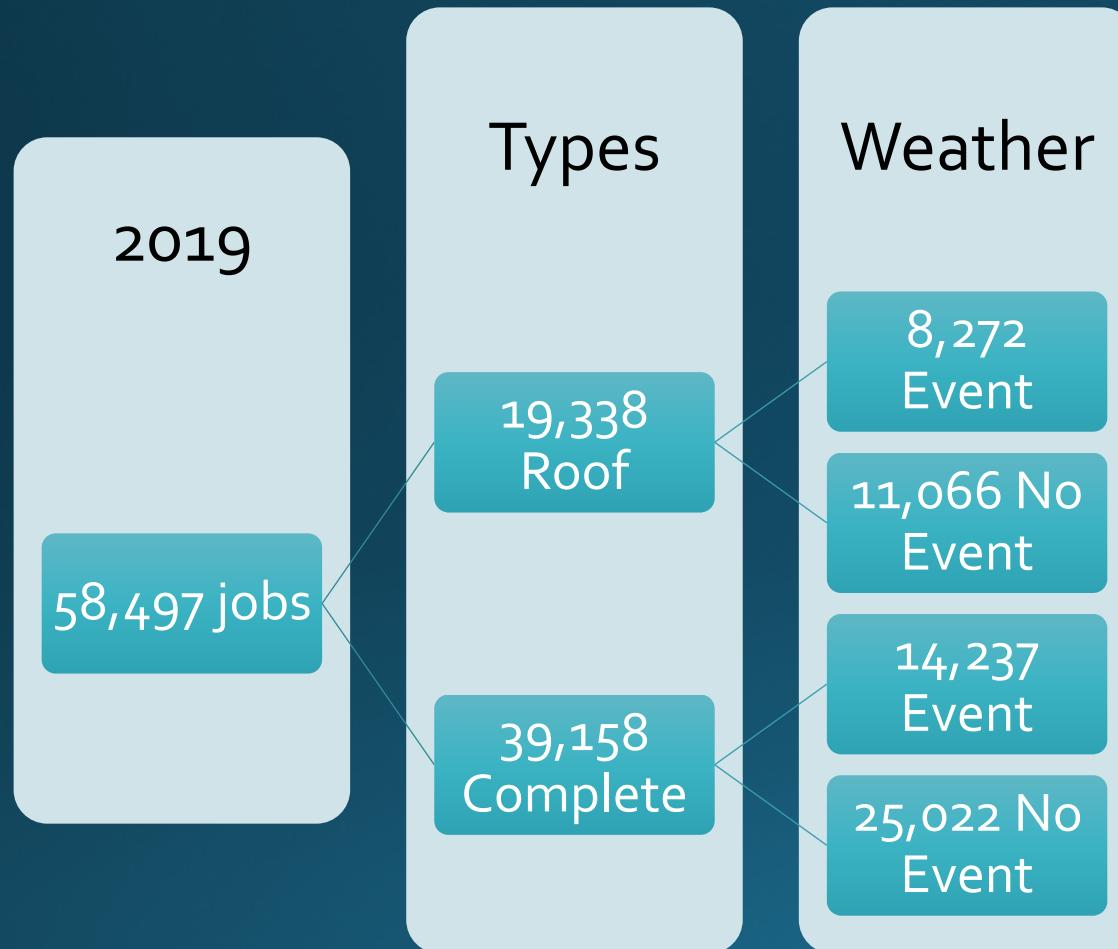
Roof Job - KS Region



Roof Job - TX Region



Results



In 2019, 38% of jobs will come from weather related events, increasing from 15% in 2018.

Jobs are expected to grow 138% in 2019 compared to 2018.

Errors on models range from 0.02 to 27 (jobs - MAE) or 0.04 to 37 (jobs – RMSE)

Numbers are estimates and have errors associated to them.

Key Takeaways

- There is a seasonality and trend which can be seen in the number of jobs. The increase in the summer seems to repeat itself.
- Complete Jobs are the ones most related to weather events (you'd think it would be roof!?)
- TX is largest and fastest growing region, but with the least weather event related jobs, proportionally – suggests organic market growth.
- Seasonality is mainly due to non-weather related jobs, although weather related are also seasonal.
- House buying season is April – July - <https://smartasset.com/mortgage/what-is-home-buying-season>
- Weather related jobs are growing faster than non related (insurance mkt).

Actionable Insights

- Aggressively market to regions that have the most weather event related jobs, during those periods. Find similar areas.
- TX beat records in 2018 for home sales -
<https://www.prnewswire.com/news-releases/texas-housing-market-breaks-record-for-home-sales-median-price-in-2018-300799910.html>
- Find regions similar to TX and NE in the US to expand or analyze go-to-market strategy.
- Why are other Regions so under developed? Slow market?

Future Analysis

- NLP on the weather comments to evaluate severity of event.
- Develop other types of models to compare errors and results.
- Use seasons (summer) as a variable for the prediction model. Use Housing market data as variables for prediction.
- Do a cluster analysis on the Regions to see if there are similar ones that can be aggregated, for a more robust forecast.
- Do a deeper analysis on weather events and understand how to use prediction of events for efficient marketing strategy.

Thank you!

Gustavo Oliveira
(302) 727-1599
goliveir@syr.edu