



# Predicting Wage-loss after a Natural Disaster

<b>Client:</b>	New Light Technologies
<b>Project Team:</b>	Elijah Curme Oliver Balbuena Hernandez Matthew B. Murrell
<b>Date:</b>	21 Feb 2020
<b>Location:</b>	General Assembly, Boston

C  
l  
i  
e  
n  
t  
  
B  
r  
i  
e  
f

3

P  
r  
o  
j  
e  
c  
t  
  
G  
o  
a  
l  
s

4

D  
a  
t  
a

8

S  
o  
l  
u  
t  
i  
o  
n

11

N  
e  
x  
t  
  
s  
t  
e  
p  
s

# Predicting Wage-Loss After a Natural Disaster

## Overview

# Client Brief

The problem  
statement ...

## Initial Point of Departure

Using Indeed or Glassdoor together with number and type of affected businesses to estimate the expected economic loss due to a disaster

## Problem Statement

This tool will rely on websites that provide employment information and sector-specific wage estimations (such as Glassdoor and Indeed) to project the economic loss (wage loss) due to a disaster. Based on the type of businesses and services in a given affected area and/or using supplementary demographic data (for example, from the Census Bureau of Statistics), the tool will provide an estimation about the projected economic loss in a given locality based on the reported or estimated wage loss in the locality.

# Project Goals

Creating a useful tool ...

- Predict lost wages resulting from natural disaster event(s)
- Deliver an interactive tool with which users can input factors and generate location-specific wage-loss estimates

# Data

What might wage interruption look like?

## Direct Impact(s)

- Job displacement – place of work damaged or destroyed
- Home displacement – home is damaged or destroyed
- Injury Injury/trauma – person sustains physical injury and/or psychological trauma
- Combinations of the above
- Etc. ...

## Indirect Impact(s)

- Loss of customers who were displaced
- Child care emergencies resulting from school closings
- Etc. ...

# Data

What information might help to predict wage loss as a result of natural disasters?

## In an ideal world ...

- Type of disaster
- Duration of disaster
- Magnitude of disaster
- Affected geographic area(s)
- Number of businesses damaged in affected area
- Number of homes damaged in affected area
- Degree of structural damage for a given property (e.g. damaged but usable, damaged but not usable, destroyed, undamaged)
- Number of wage-earners, by zip code, displaced by the event
- Duration of wage-earner displacement, by event, by zip code
- Salary information for affected wage-earners

# Data

Gathering available data...

## Existing Public Data Sources

- Disaster data very fragmented
  - much “knitting” is required
- No single source of truth for disaster data
- Disaster data inconsistent, e.g.:
  - location specificity
  - aggregate and component estimates of economic loss
  - Duration
  - Severity
  - Category/type
  - Etc.

## Wage Data

- Generally aggregated
- Nearly impossible to tie wages to both job location and domicile location

## Measurement Challenges

- Static vs dynamic effects
- Wage loss would not be at a constant rate
- How do we measure/account for change over time?

# Data

Sourcing the data  
required casting a  
wide net ...

## Sources Explored

glassdoor.com  
indeed.com  
census.gov  
bls.gov  
data.gov  
noaa.gov  
Internal-displacement.org  
simplemaps.com  
Fema.gov  
Google.com  
And many more ...

## Sources Used

Census  
NOAA  
Internal-Displacement.org  
simplemaps.com  
FEMA  
google

**Bottom line:** The complete data needed to drive this analysis does not currently exist or is not readily available, so work-arounds will be required.



# Solution

Working within  
the data  
limitations...

Given the data limitations, we opted to use a proxy, displaced population, for the initial prediction model ...

## *Independent Variables:*

- Severity
- Affected zip count (how many affected zip codes)
- Total population
- Disaster types (as dummies)

## *Dependent Variable:*

- Displaced people




# Solution

Putting it all  
together ...


1. Assemble data sets from variety of sources.




2. Estimate affected geographic area, given event type and magnitude.



3. Predict number of displaced individuals, given Event Type, Severity, Location, and Population in the affected area(s).



4. Estimate proportion of individuals who are wage earners; multiply that number by median daily income for the affected geographic area(s).



5. **Result:** Estimated total lost wages resulting from the offending event in affected area(s).

# Solution

Seeing it in  
action...

Start Demo



# Solution

Understanding the  
current  
limitations...

## Small Data Set:

Num. of Obs. = 20  
Features = 7  
Response Var = Number of  
Displaced People

## Features in Model:

- Severity
- Affected zips count
- Total pop
- Event-type dummies:
  - Earthquake
  - Tornado
  - Hurricane
  - Wildfire
  - Flood

## Model Used:

Random Forest Regressor

## Model Performance

R2 – Training Data	R2 – Test Data
0.82	0.42

# Next Steps

Data, data, data ...  
and some  
functional  
enhancements.

## Data Improvement

Some of the data can be obtained via scalable means, but some critical data may need to be manually assembled.

## Public/Private Partnership

Payroll data could be uniquely impactful for solving this problem. Warrants exploration of possible partnerships with payroll providers (e.g. Paychex, ADP, regional providers, etc.).

## Model Re-specification

Model will need to be refit if new data are added.

## Interactive Enhancement

Build out interface functionality to enable graphical input (e.g. click on map, etc.).