Preventing Workplace Injuries in Pennsylvania

Bofan Xue
Cole Thomas
Kelly McManus
Kirtiman Rai
Yitian Liu

pennsylvania
DEPARTMENT OF LABOR & INDUSTRY

5,500,000 people have a job in Pennsylvania

Source: QCEW Data, 2021

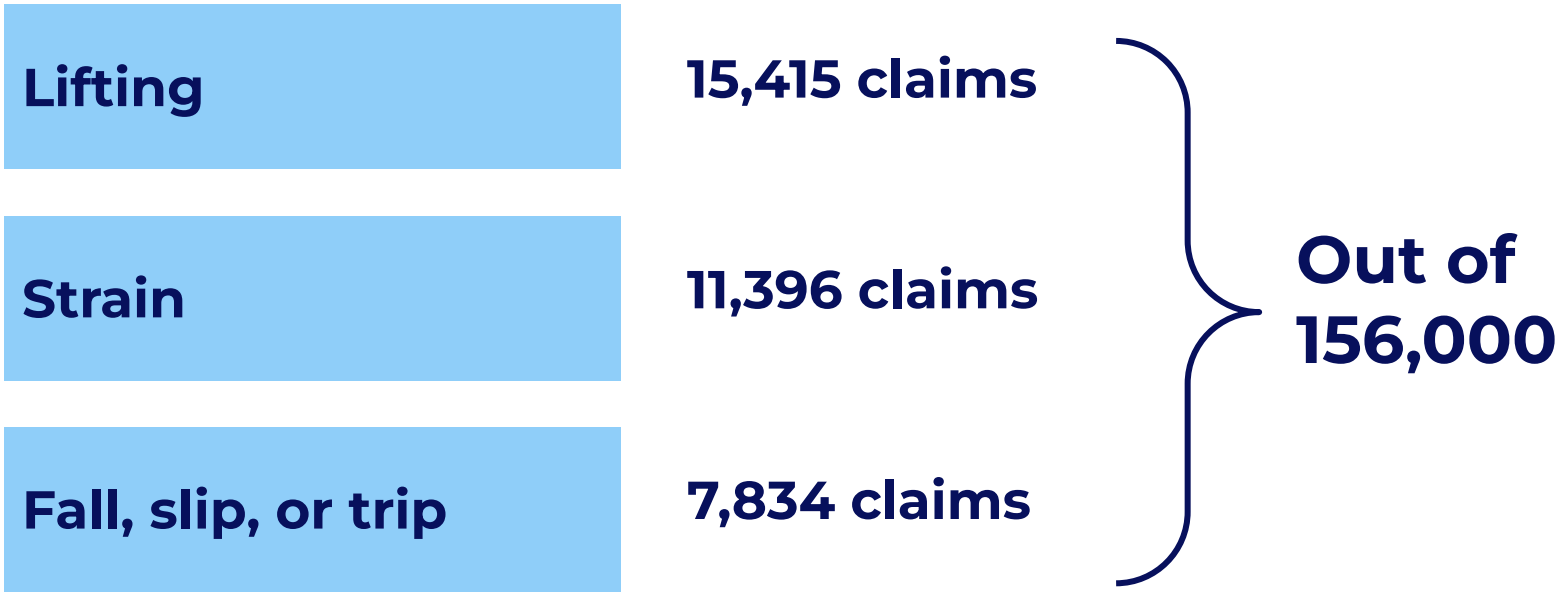
156,000 **workers are injured each year**

Source: WCAIS Data, 2021

1 in 50 workers get injured each year

Source: QCEW and WCAIS Data, 2021

Most common injury causes in 2021



Source: WCAIS Data, 2021

Some injuries are preventable with proper employee training and good health and safety programs

About the team

Cole

4 years of banking experience as an analytic consultant

Kelly

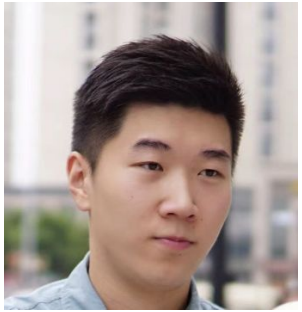
3 years working at a startup as data analyst

Kirtiman

Recent internship at Adobe surfacing insights to management

Grace

Recent internship at TikTok helping to define the product roadmap



Bofan

Recent internship at Deloitte consulting on tech strategy

Cole

BI Engineer & finance manager

Working to integrate medical costs into prediction and ensure we're using the best technique

Kelly

Project manager & data engineer

Defining project scope and integrating new data into the solution

Kirtiman

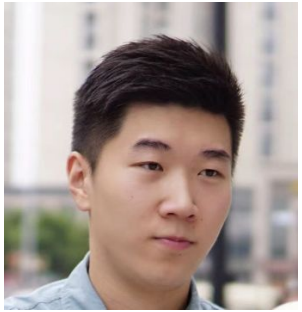
Data scientist

Extracting actionable insights from thousands of reported injury claims

Grace

Product manager

Breaking down our objectives into actionable product improvements



Bofan

Software engineer

Building out the web application to ensure predictions can run smoothly

Project timeline

Scope Sept-Oct	• Scope goal • Gather requirements • Onboard team members to current solutions 
Build Oct-Nov	• Exploratory data analysis • Re-define and implement severity using medical costs • Visualization for claims and employment data
Communicate Nov-Dec	• Client training and tutorial documentation • Recommendations for future iterations

Today’s goals

Walkthrough the **problem** of focusing resources

Outline the **objectives** for this iteration

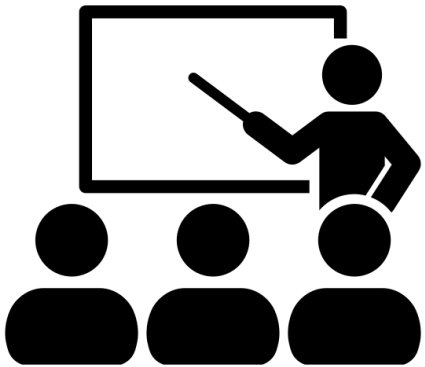
Look at what has been **accomplished**

Outline and progress on **deliverables**

Confident that we’re heading in the right direction

Walk-through the **problem** of focusing resources

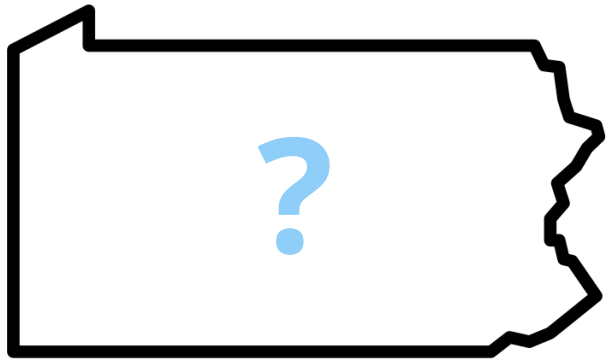
Pennsylvania Department of Labor & Industry



Provide trainings to
proactively prevent injuries

Source: Icon Master and Cuputo from the Noun Project

Where to focus resources?



Source: Andrejs Kirma from the Noun Project

Identify counties and industries that

will have **high injury rates**

to **proactively** allocate

resources and training

and **reduce injuries**

Source: CMU

Outline the **objectives** for this iteration

Current problems informed objectives

Inaccurate prediction, does not reflect severity of injuries

Technical challenges running the predictions and analysis

Current problems informed objectives

Inaccurate prediction, does not reflect severity of injuries

Technical challenges running the predictions and analysis

Current problems informed objectives

Integrate medical costs into current prediction model to account for injury severity

Technical challenges running the predictions and analysis

Integrate medical costs into current prediction model to account for injury severity

Leverage the Pennsylvania Compensation Rating Bureau (PCRB) data to obtain medical costs per claim

“Amount Paid” to inform a better definition of injury severity

Current problems informed objectives

Integrate medical costs into current prediction model to account for injury severity

Technical challenges running the predictions and analysis

Current problems informed objectives

Integrate medical costs into current prediction model to account for injury severity

Enable the DLI team to analyze prediction results for decision making

Enable the DLI team to analyze prediction results for decision making

Unblock current solution to work on infrastructure

Surface key counties and industries impacted with high injury rates

Integrate medical costs into current prediction model to account for injury severity

Enable the DLI team to analyze prediction results for decision making

Until these objectives are tackled, we’re not using all available resources to protect workers

Look at what has been **accomplished**

Current solution includes

Data

Web application

Prediction model

Report

Two types of data

Claims data (WCAIS) =

Individual injury claims submitted by employers across Pennsylvania

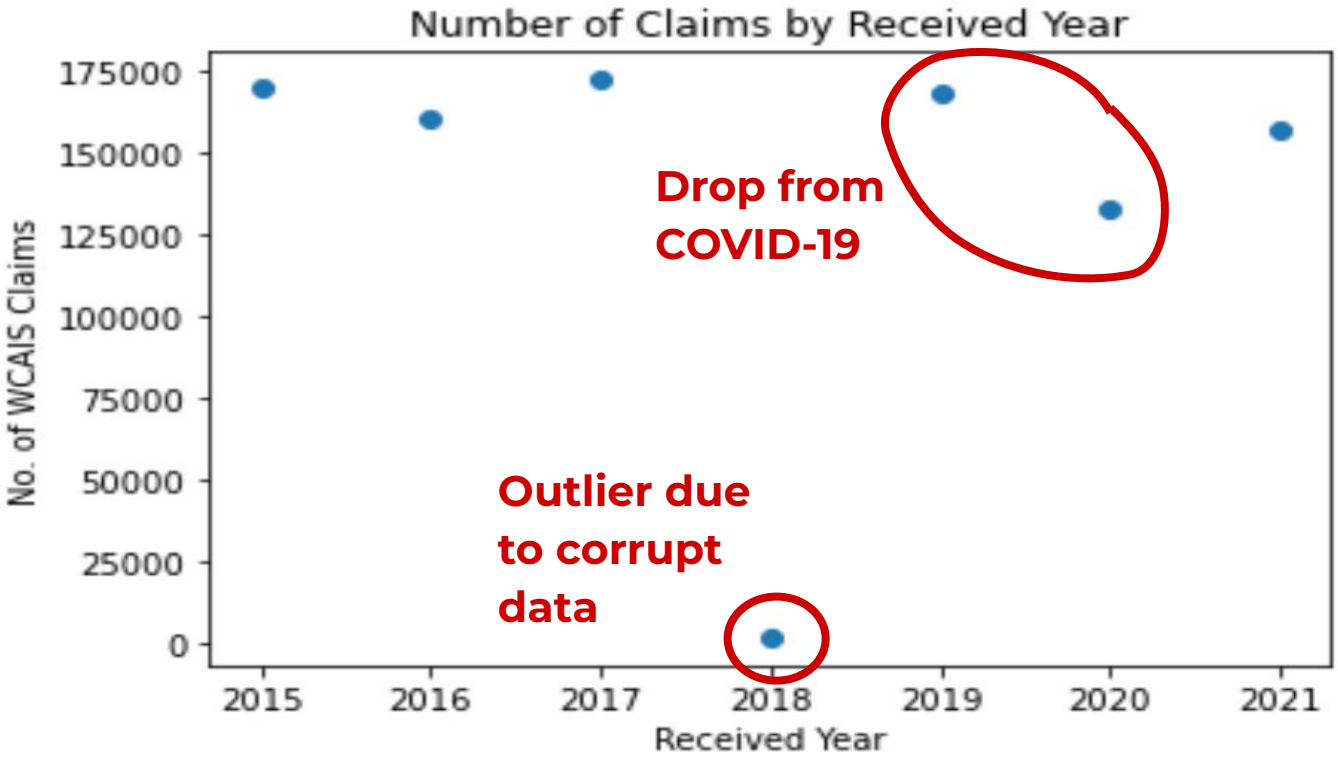
Employment data (QCEW) =

Monthly number of employees

Both are …

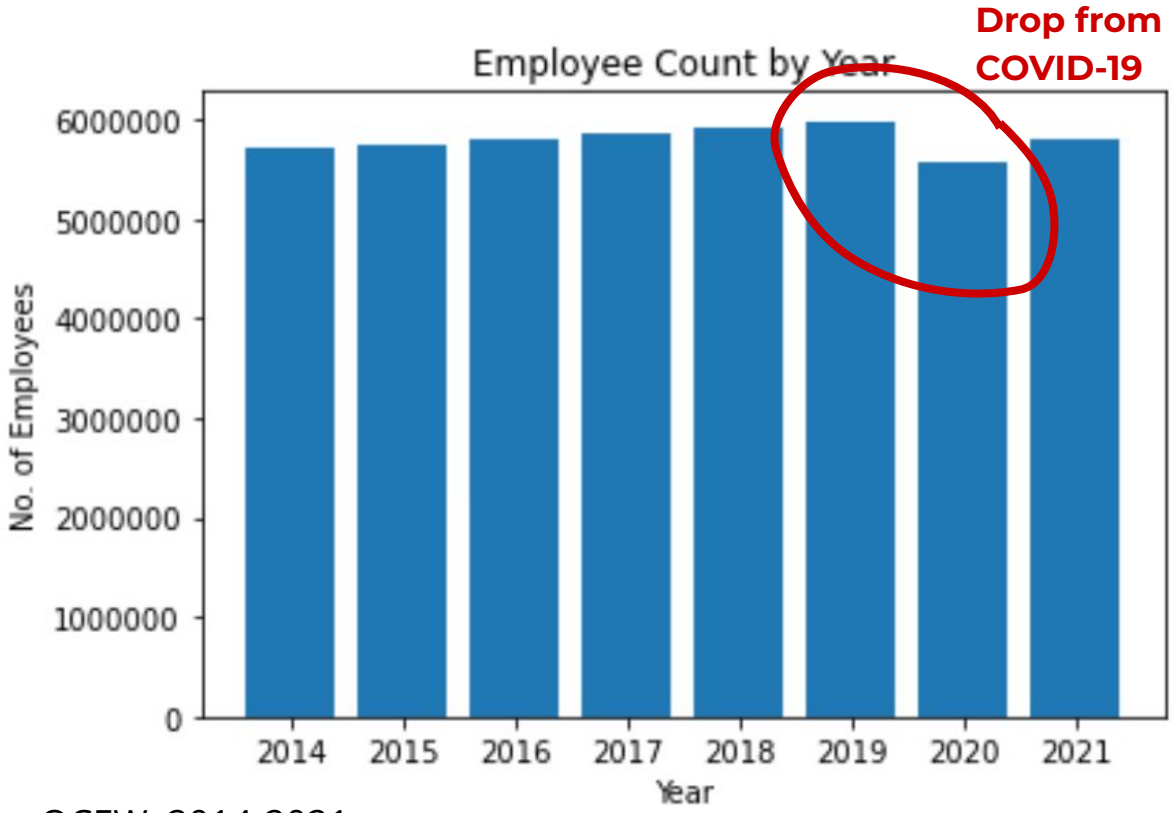
- Per county and industry (NAICS)
- 2014 - 2021

Around 150,000 injury claims per year



Source: WCAIS, 2015-2021

Around 5.5 million workers each year

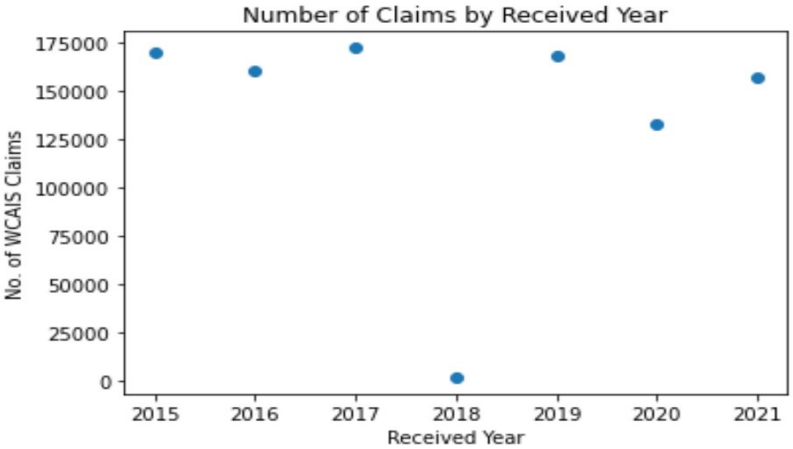


Year	Average number of employees
2014	5,717,464
2015	5,759,036
2016	5,801,422
2017	5,868,636
2018	5,936,389
2019	5,988,879
2020	5,573,212
2021	5,798,448

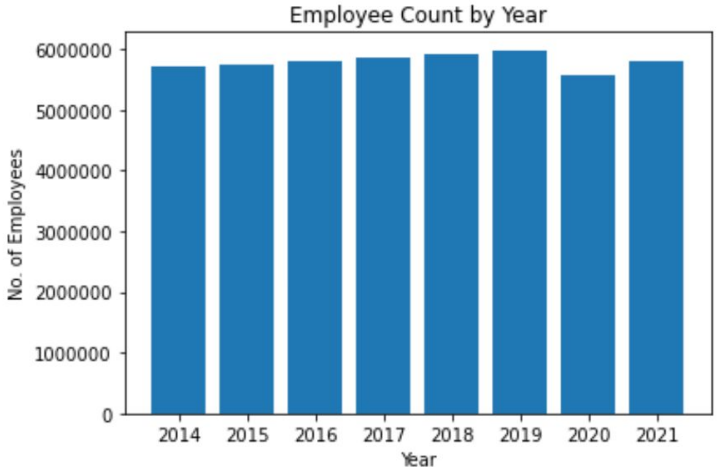
Source: QCEW, 2014-2021

Both datasets calculate injury rate

Claims



Employment

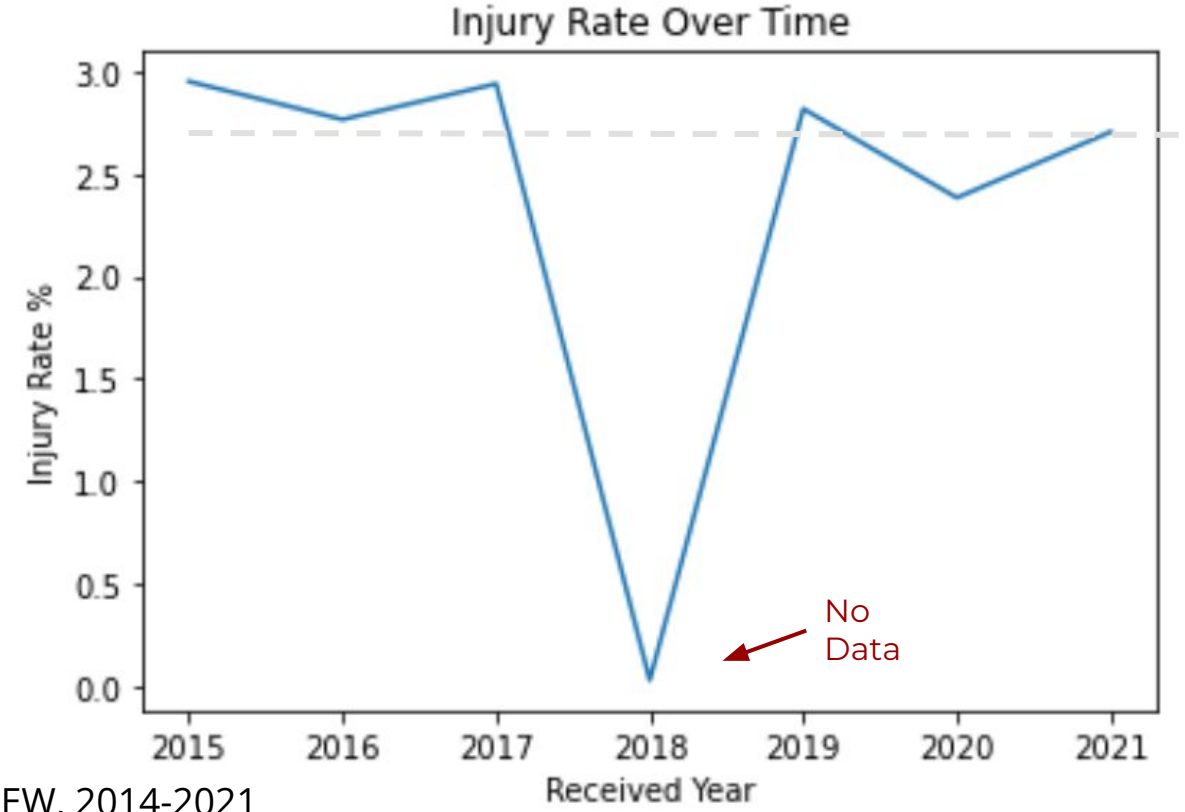


Source: WCAIS & QCEW, 2014-2021

$$\text{Injury rate} = \frac{\text{Number of claims}}{\text{Number of employees}}$$

Per **county**, **industry**, and **month**

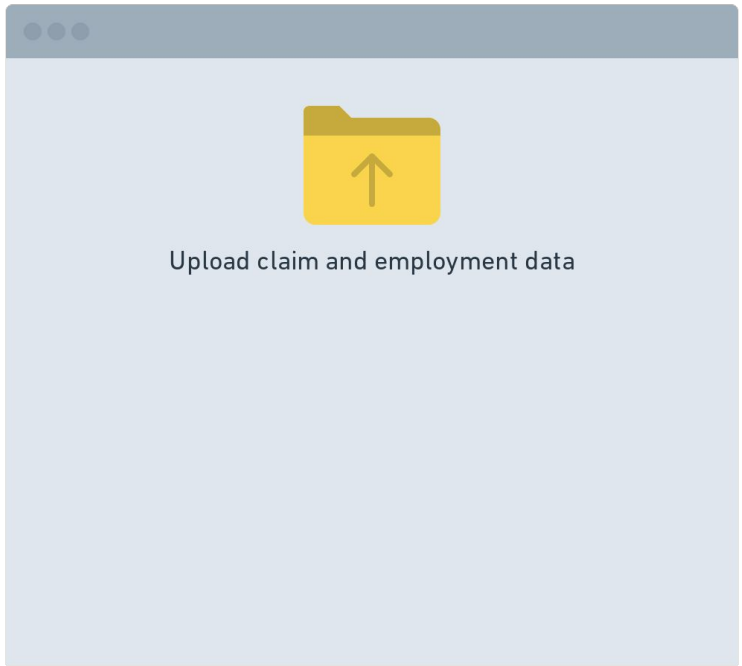
Around 2.6% injury rate in Pennsylvania



Source: WCAIS & QCEW, 2014-2021

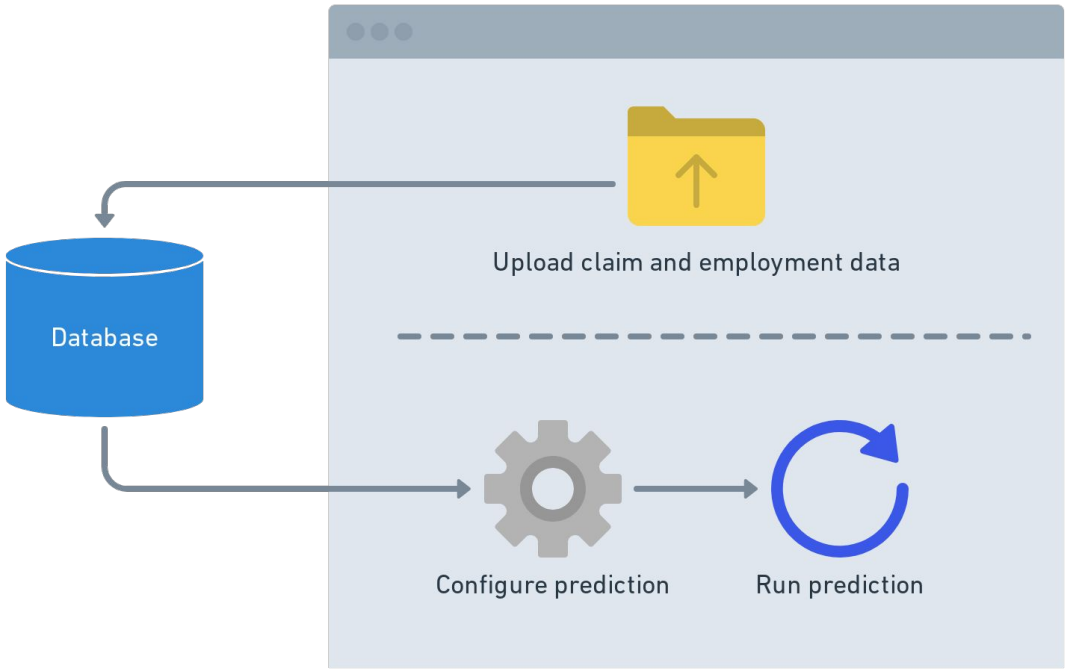
Use **claims and employment data** to predict injury rates in the **web application**

Upload data in web application

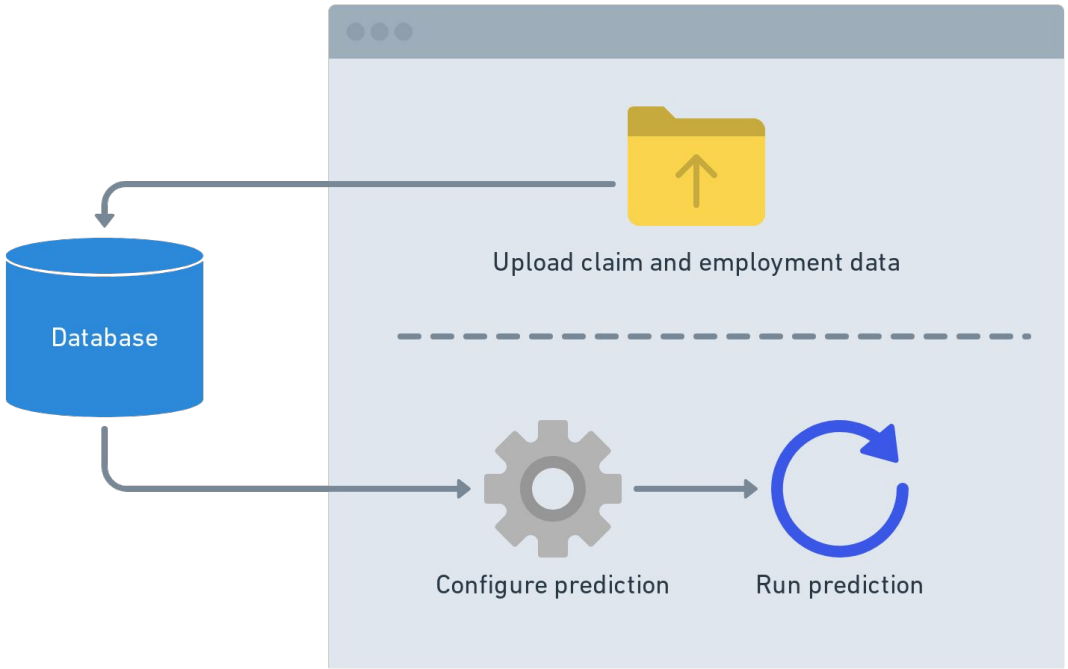


Clean and store in database

Configure and run injury rate prediction



Configure and run injury rate prediction



"Model"

- SARIMAX
- Prophet

SARIMAX by statsmodels

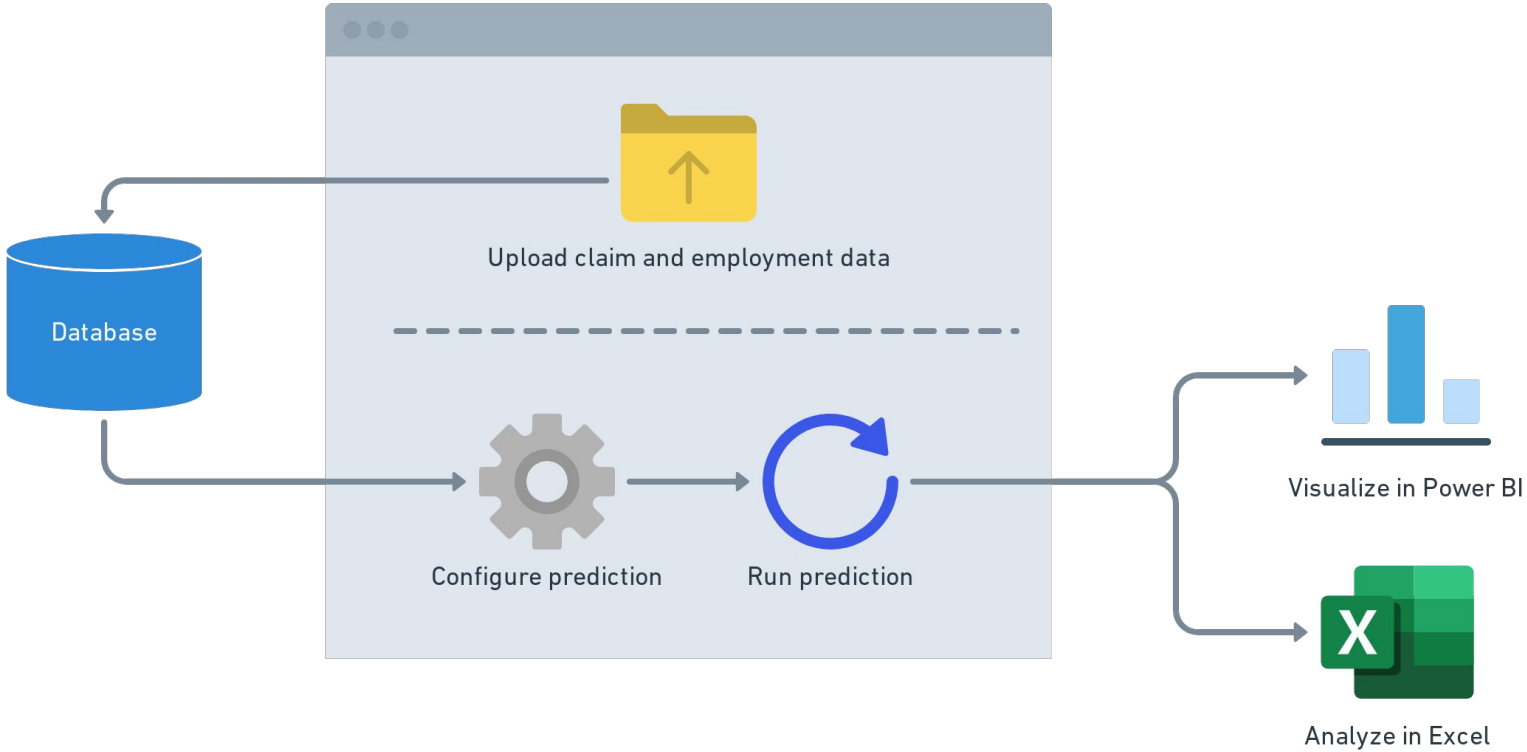
Predicts injury rate
Y: Injury Rate
Forward looking monthly injury rate by industry (NAICS) and county
Takes into account seasonal variations **Potential for additional variables**
Documentation Code

PROPHET by Meta

Predicts injury rate
X: Date, Y: Injury Rate
Forward looking monthly injury rate by industry (NAICS) and county
Takes into account seasonal variations **Robust to shifts in trends and outliers**
Documentation Code

Image: Facebook Github
Image: Statsmodels

Analyze predictions in report



Current solution status

Accomplishments

Nice baseline for predicting injury rates

Easily extendable to to add other prediction models beyond SARIMAX & Prophet

Team now able to use the solution!

Areas for improvement

Prediction currently does not take advantage of additional variables (injury nature, cause, medical costs)

Predictions are not split by severity

Predictions highly coupled into the Django web application architecture

Outline our **deliverables**

Recall our two objectives

Integrate medical costs into current prediction model to account for injury severity

Enable the DLI team to analyze prediction results for decision making

Our main deliverables

Upload medical costs to database through the web application

Analyze injury rates by severity, updated with medical costs

Surface prediction results in Excel

Visualize top 5 counties/industries injury rate and medical costs in Power BI

Our main deliverables

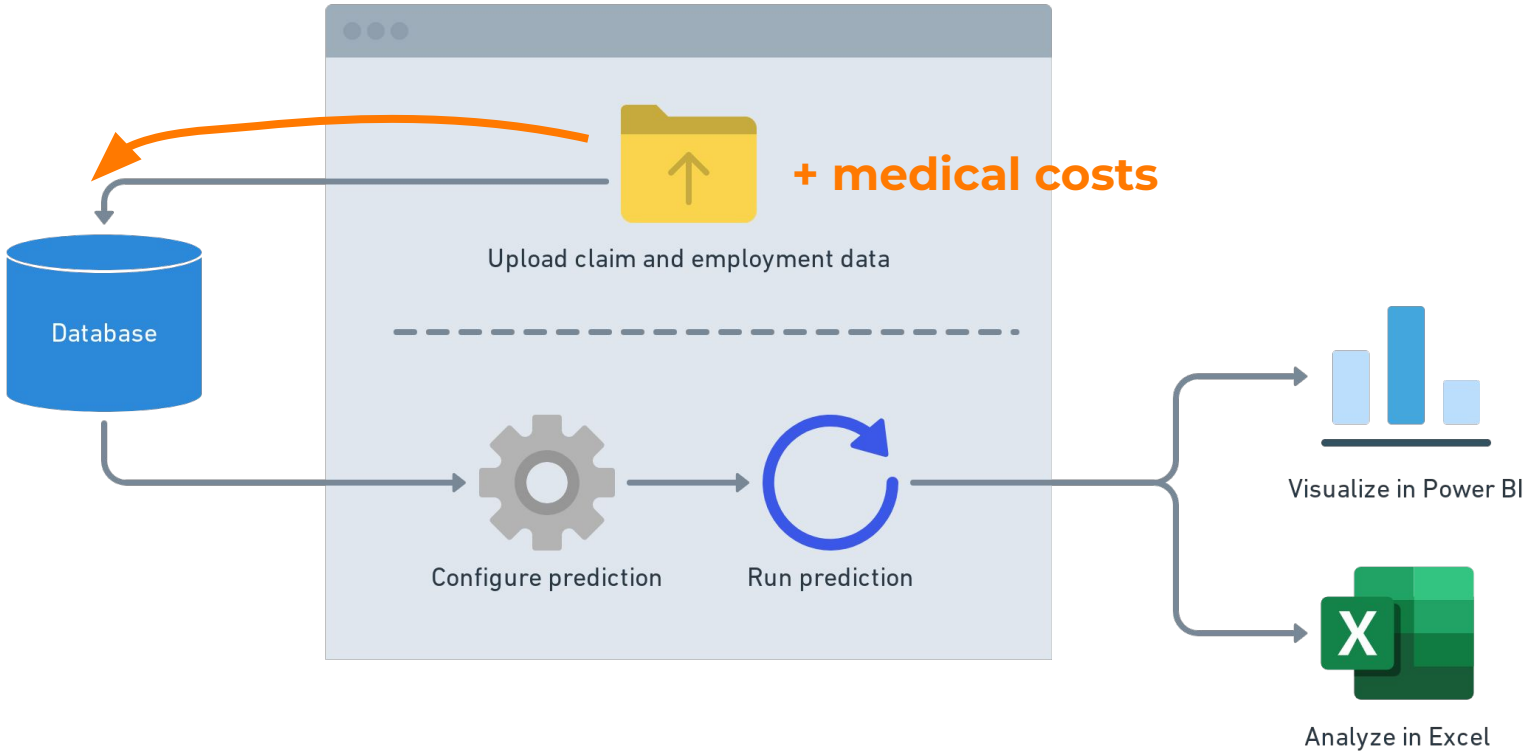
Upload medical costs to database through the web application

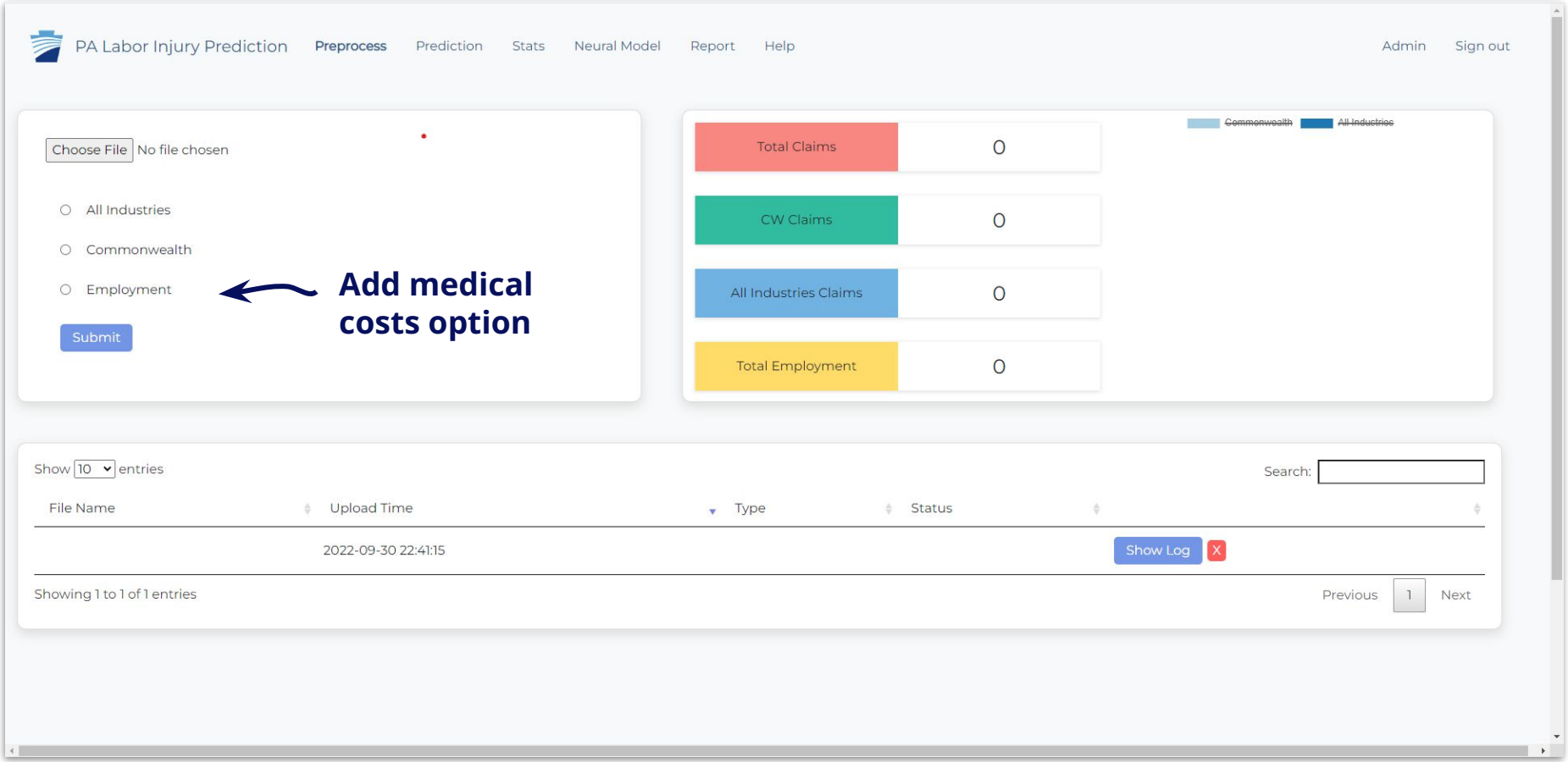
Analyze injury rates by severity, updated with medical costs

Surface prediction results in Excel

Visualize top 5 counties/industries injury rate and medical costs in Power BI

Upload medical costs




PA Labor Injury Prediction
Preprocess
Prediction
Stats
Neural Model
Report
Help
Admin
Sign out
Choose File
No file chosen
All Industries
Commonwealth
Employment
Submit
Add medical costs option
Total Claims
0
CW Claims
0
All Industries Claims
0
Total Employment
0
Commonwealth
All Industries
Show 10 entries
Search:
File Name
Upload Time
Type
Status
2022-09-30 22:41:15
Show Log
X
Showing 1 to 1 of 1 entries
Previous
1
Next

Our main deliverables

Upload medical costs to database through the web application

Analyze injury rates by severity, updated with medical costs

Surface prediction results in Excel

Visualize top 5 counties/industries injury rate and medical costs in Power BI

Prediction output in Excel

id	naics_code	county	year	month	date	severity	claim_number	emp_count	injury_rate	type	county_formatted	month_str	naics_level	commonwealth	task_id
27	62	Adams County	2020	9	9/1/2020	All	8	4959	0.001613228	Actual	Adams County PA	September	2	TRUE	11
28	62	Adams County	2020	10	10/1/2020	High	1	5069	0.000197278	Actual	Adams County PA	October	2	TRUE	11
29	62	Adams County	2020	10	10/1/2020	Low	7	5069	0.001380943	Actual	Adams County PA	October	2	TRUE	11
30	62	Adams County	2020	10	10/1/2020	Medium	3	5069	0.000591833	Actual	Adams County PA	October	2	TRUE	11
31	62	Adams County	2020	10	10/1/2020	All	11	5069	0.002170053	Actual	Adams County PA	October	2	TRUE	11
32	62	Adams County	2020	11	11/1/2020	High	1	5030	0.000198807	Actual	Adams County PA	November	2	TRUE	11
33	62	Adams County	2020	11	11/1/2020	Low	5	5030	0.000994036	Actual	Adams County PA	November	2	TRUE	11
34	62	Adams County	2020	11	11/1/2020	Medium	4	5030	0.000795229	Actual	Adams County PA	November	2	TRUE	11
35	62	Adams County	2020	11	11/1/2020	All	10	5030	0.001988072	Actual	Adams County PA	November	2	TRUE	11
36	62	Adams County	2020	12	12/1/2020	High	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
37	62	Adams County	2020	12	12/1/2020	Low	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
38	62	Adams County	2020	12	12/1/2020	Medium	3	5008	0.000599042	Actual	Adams County PA	December	2	TRUE	11
39	62	Adams County	2020	12	12/1/2020	All	5	5008	0.000998403	Actual	Adams County PA	December	2	TRUE	11
0	62	Adams County	2021	1	1/1/2021	All	4.880162184	5008	0.000974473	Predicted	Adams County PA	January	2	TRUE	11
1	62	Adams County	2021	2	2/1/2021	All	12.16736266	5008	0.002429585	Predicted	Adams County PA	February	2	TRUE	11
2	62	Adams County	2021	3	3/1/2021	All	1.814446778	5008	0.00036231	Predicted	Adams County PA	March	2	TRUE	11
3	62	Adams County	2021	4	4/1/2021	All	0.1705288	5008	3.41E-05	Predicted	Adams County PA	April	2	TRUE	11
4	62	Adams County	2021	5	5/1/2021	All	6.290760082	5008	0.001256142	Predicted	Adams County PA	May	2	TRUE	11
5	62	Adams County	2021	6	6/1/2021	All	7.815840807	5008	0.001560671	Predicted	Adams County PA	June	2	TRUE	11
6	62	Adams County	2021	7	7/1/2021	All	13.47827404	5008	0.002691349	Predicted	Adams County PA	July	2	TRUE	11
7	62	Adams County	2021	8	8/1/2021	All	9.678400805	5008	0.001932588	Predicted	Adams County PA	August	2	TRUE	11
8	62	Adams County	2021	9	9/1/2021	All	3.870799189	5008	0.000772923	Predicted	Adams County PA	September	2	TRUE	11
9	62	Adams County	2021	10	10/1/2021	All	10.11950012	5008	0.002020667	Predicted	Adams County PA	October	2	TRUE	11
10	62	Adams County	2021	11	11/1/2021	All	6.369047979	5008	0.001271775	Predicted	Adams County PA	November	2	TRUE	11
11	62	Adams County	2021	12	12/1/2021	All	3.629526048	5008	0.000724746	Predicted	Adams County PA	December	2	TRUE	11

Adams County, NAICS = 62, 2020

Historic values

id	naics_code	county	year	month	date	severity	claim_number	emp_count	injury_rate	type	county_formatted	month_str	naics_level	commonwealth	task_id
27	62	Adams County	2020	9	9/1/2020	All	8	4959	0.001613228	Actual	Adams County PA	September	2	TRUE	11
28	62	Adams County	2020	10	10/1/2020	High	1	5069	0.000197278	Actual	Adams County PA	October	2	TRUE	11
29	62	Adams County	2020	10	10/1/2020	Low	7	5069	0.001380943	Actual	Adams County PA	October	2	TRUE	11
30	62	Adams County	2020	10	10/1/2020	Medium	3	5069	0.000591833	Actual	Adams County PA	October	2	TRUE	11
31	62	Adams County	2020	10	10/1/2020	All	11	5069	0.002170053	Actual	Adams County PA	October	2	TRUE	11
32	62	Adams County	2020	11	11/1/2020	High	1	5030	0.000198807	Actual	Adams County PA	November	2	TRUE	11
33	62	Adams County	2020	11	11/1/2020	Low	5	5030	0.000994036	Actual	Adams County PA	November	2	TRUE	11
34	62	Adams County	2020	11	11/1/2020	Medium	4	5030	0.000795229	Actual	Adams County PA	November	2	TRUE	11
35	62	Adams County	2020	11	11/1/2020	All	10	5030	0.001988072	Actual	Adams County PA	November	2	TRUE	11
36	62	Adams County	2020	12	12/1/2020	High	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
37	62	Adams County	2020	12	12/1/2020	Low	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
38	62	Adams County	2020	12	12/1/2020	Medium	3	5008	0.000599042	Actual	Adams County PA	December	2	TRUE	11
39	62	Adams County	2020	12	12/1/2020	All	5	5008	0.000998403	Actual	Adams County PA	December	2	TRUE	11
0	62	Adams County	2021	1	1/1/2021	All	4.880162184	5008	0.000974473	Predicted	Adams County PA	January	2	TRUE	11
1	62	Adams County	2021	2	2/1/2021	All	12.16736266	5008	0.002429585	Predicted	Adams County PA	February	2	TRUE	11
2	62	Adams County	2021	3	3/1/2021	All	1.814446778	5008	0.00036231	Predicted	Adams County PA	March	2	TRUE	11
3	62	Adams County	2021	4	4/1/2021	All	0.1705288	5008	3.41E-05	Predicted	Adams County PA	April	2	TRUE	11
4	62	Adams County	2021	5	5/1/2021	All	6.290760082	5008	0.001256142	Predicted	Adams County PA	May	2	TRUE	11
5	62	Adams County	2021	6	6/1/2021	All	7.815840807	5008	0.001560671	Predicted	Adams County PA	June	2	TRUE	11
6	62	Adams County	2021	7	7/1/2021	All	13.47827404	5008	0.002691349	Predicted	Adams County PA	July	2	TRUE	11
7	62	Adams County	2021	8	8/1/2021	All	9.678400805	5008	0.001932588	Predicted	Adams County PA	August	2	TRUE	11
8	62	Adams County	2021	9	9/1/2021	All	3.870799189	5008	0.000772923	Predicted	Adams County PA	September	2	TRUE	11
9	62	Adams County	2021	10	10/1/2021	All	10.11950012	5008	0.002020667	Predicted	Adams County PA	October	2	TRUE	11
10	62	Adams County	2021	11	11/1/2021	All	6.369047979	5008	0.001271775	Predicted	Adams County PA	November	2	TRUE	11
11	62	Adams County	2021	12	12/1/2021	All	3.629526048	5008	0.000724746	Predicted	Adams County PA	December	2	TRUE	11

Historic values

id	naics_code	county	year	month	date	severity	claim_number	emp_count	injury_rate	type	county_formatted	month_str	naics_level	commonwealth	task_id
27	62	Adams County	2020	9	9/1/2020	All	8	4959	0.001613228	Actual	Adams County PA	September	2	TRUE	11
28	62	Adams County	2020	10	10/1/2020	High	1	5069	0.000197278	Actual	Adams County PA	October	2	TRUE	11
29	62	Adams County	2020	10	10/1/2020	Low	7	5069	0.001380943	Actual	Adams County PA	October	2	TRUE	11
30	62	Adams County	2020	10	10/1/2020	Medium	3	5069	0.000591833	Actual	Adams County PA	October	2	TRUE	11
31	62	Adams County	2020	10	10/1/2020	All	11	5069	0.002170053	Actual	Adams County PA	October	2	TRUE	11
32	62	Adams County	2020	11	11/1/2020	High	1	5030	0.000198807	Actual	Adams County PA	November	2	TRUE	11
33	62	Adams County	2020	11	11/1/2020	Low	5	5030	0.000994036	Actual	Adams County PA	November	2	TRUE	11
34	62	Adams County	2020	11	11/1/2020	Medium	4	5030	0.000795229	Actual	Adams County PA	November	2	TRUE	11
35	62	Adams County	2020	11	11/1/2020	All	10	5030	0.001988072	Actual	Adams County PA	November	2	TRUE	11
36	62	Adams County	2020	12	12/1/2020	High	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
37	62	Adams County	2020	12	12/1/2020	Low	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
38	62	Adams County	2020	12	12/1/2020	Medium	3	5008	0.000599042	Actual	Adams County PA	December	2	TRUE	11
39	62	Adams County	2020	12	12/1/2020	All	5	5008	0.000998403	Actual	Adams County PA	December	2	TRUE	11

Predicted values

0	62	Adams County	2021	1	1/1/2021	All	4.880162184	5008	0.000974473	Predicted	Adams County PA	January	2	TRUE	11
1	62	Adams County	2021	2	2/1/2021	All	12.16736266	5008	0.002429585	Predicted	Adams County PA	February	2	TRUE	11
2	62	Adams County	2021	3	3/1/2021	All	1.814446778	5008	0.00036231	Predicted	Adams County PA	March	2	TRUE	11
3	62	Adams County	2021	4	4/1/2021	All	0.1705288	5008	3.41E-05	Predicted	Adams County PA	April	2	TRUE	11
4	62	Adams County	2021	5	5/1/2021	All	6.290760082	5008	0.001256142	Predicted	Adams County PA	May	2	TRUE	11
5	62	Adams County	2021	6	6/1/2021	All	7.815840807	5008	0.001560671	Predicted	Adams County PA	June	2	TRUE	11
6	62	Adams County	2021	7	7/1/2021	All	13.47827404	5008	0.002691349	Predicted	Adams County PA	July	2	TRUE	11
7	62	Adams County	2021	8	8/1/2021	All	9.678400805	5008	0.001932588	Predicted	Adams County PA	August	2	TRUE	11
8	62	Adams County	2021	9	9/1/2021	All	3.870799189	5008	0.000772923	Predicted	Adams County PA	September	2	TRUE	11
9	62	Adams County	2021	10	10/1/2021	All	10.11950012	5008	0.002020667	Predicted	Adams County PA	October	2	TRUE	11
10	62	Adams County	2021	11	11/1/2021	All	6.369047979	5008	0.001271775	Predicted	Adams County PA	November	2	TRUE	11
11	62	Adams County	2021	12	12/1/2021	All	3.629526048	5008	0.000724746	Predicted	Adams County PA	December	2	TRUE	11

Severity

id	naics_code	county	year	month	date	severity	claim_number	emp_count	injury_rate	type	county_formatted	month_str	naics_level	commonwealth	task_id
27	62	Adams County	2020	9	9/1/2020	All	8	4959	0.001613228	Actual	Adams County PA	September	2	TRUE	11
28	62	Adams County	2020	10	10/1/2020	High	1	5069	0.000197278	Actual	Adams County PA	October	2	TRUE	11
29	62	Adams County	2020	10	10/1/2020	Low	7	5069	0.001380943	Actual	Adams County PA	October	2	TRUE	11
30	62	Adams County	2020	10	10/1/2020	Medium	3	5069	0.000591833	Actual	Adams County PA	October	2	TRUE	11
31	62	Adams County	2020	10	10/1/2020	All	11	5069	0.002170053	Actual	Adams County PA	October	2	TRUE	11
32	62	Adams County	2020	11	11/1/2020	High	1	5030	0.000198807	Actual	Adams County PA	November	2	TRUE	11
33	62	Adams County	2020	11	11/1/2020	Low	5	5030	0.000994036	Actual	Adams County PA	November	2	TRUE	11
34	62	Adams County	2020	11	11/1/2020	Medium	4	5030	0.000795229	Actual	Adams County PA	November	2	TRUE	11
35	62	Adams County	2020	11	11/1/2020	All	10	5030	0.001988072	Actual	Adams County PA	November	2	TRUE	11
36	62	Adams County	2020	12	12/1/2020	High	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
37	62	Adams County	2020	12	12/1/2020	Low	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
38	62	Adams County	2020	12	12/1/2020	Medium	3	5008	0.000599042	Actual	Adams County PA	December	2	TRUE	11
39	62	Adams County	2020	12	12/1/2020	All	5	5008	0.000998403	Actual	Adams County PA	December	2	TRUE	11
0	62	Adams County	2021	1	1/1/2021	All	4.880162184	5008	0.000974473	Predicted	Adams County PA	January	2	TRUE	11
1	62	Adams County	2021	2	2/1/2021	All	12.16736266	5008	0.002429585	Predicted	Adams County PA	February	2	TRUE	11
2	62	Adams County	2021	3	3/1/2021	All	1.814446778	5008	0.00036231	Predicted	Adams County PA	March	2	TRUE	11
3	62	Adams County	2021	4	4/1/2021	All	0.1705288	5008	3.41E-05	Predicted	Adams County PA	April	2	TRUE	11
4	62	Adams County	2021	5	5/1/2021	All	6.290760082	5008	0.001256142	Predicted	Adams County PA	May	2	TRUE	11
5	62	Adams County	2021	6	6/1/2021	All	7.815840807	5008	0.001560671	Predicted	Adams County PA	June	2	TRUE	11
6	62	Adams County	2021	7	7/1/2021	All	13.47827404	5008	0.002691349	Predicted	Adams County PA	July	2	TRUE	11
7	62	Adams County	2021	8	8/1/2021	All	9.678400805	5008	0.001932588	Predicted	Adams County PA	August	2	TRUE	11
8	62	Adams County	2021	9	9/1/2021	All	3.870799189	5008	0.000772923	Predicted	Adams County PA	September	2	TRUE	11
9	62	Adams County	2021	10	10/1/2021	All	10.11950012	5008	0.002020667	Predicted	Adams County PA	October	2	TRUE	11
10	62	Adams County	2021	11	11/1/2021	All	6.369047979	5008	0.001271775	Predicted	Adams County PA	November	2	TRUE	11
11	62	Adams County	2021	12	12/1/2021	All	3.629526048	5008	0.000724746	Predicted	Adams County PA	December	2	TRUE	11

Severity

	severity	claim_number	emp_count	injury_rate	type
0	All	8	4959	0.001613228	Actual
0	High	1	5069	0.000197278	Actual
0	Low	7	5069	0.001380943	Actual
0	Medium	3	5069	0.000591833	Actual
0	All	11	5069	0.002170053	Actual
0	High	1	5030	0.000198807	Actual
0	Low	5	5030	0.000994036	Actual
0	Medium	4	5030	0.000795229	Actual
0	All	10	5030	0.001988072	Actual
0	High	1	5008	0.000199681	Actual
0	Low	1	5008	0.000199681	Actual
0	Medium	3	5008	0.000599042	Actual
0	All	5	5008	0.000998403	Actual
1	All	4.880162184	5008	0.000974473	Predicted
1	All	12.16736266	5008	0.002429585	Predicted
1	All	1.814446778	5008	0.00036231	Predicted
1	All	0.1705288	5008	3.41E-05	Predicted
1	All	6.290760082	5008	0.001256142	Predicted
1	All	7.815840807	5008	0.001560671	Predicted
1	All	13.47827404	5008	0.002691349	Predicted
1	All	9.678400805	5008	0.001932588	Predicted
1	All	3.870799189	5008	0.000772923	Predicted
1	All	10.11950012	5008	0.002020667	Predicted
1	All	6.369047979	5008	0.001271775	Predicted
1	All	3.629526048	5008	0.000724746	Predicted

Severity = injury cause score (0 - 8)
+
injury nature score (0 - 8)

Low = 0 to 7

Medium = 8 to 13

High = 13 to 16

Severity

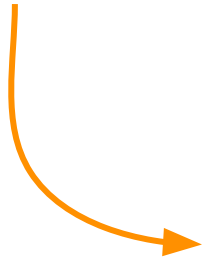
	severity	claim_number	emp_count	injury_rate	type
0	All	8	4959	0.001613228	Actual
0	High	1	5069	0.000197278	Actual
0	Low	7	5069	0.001380943	Actual
0	Medium	3	5069	0.000591833	Actual
0	All	11	5069	0.002170053	Actual
0	High	1	5030	0.000198807	Actual
0	Low	5	5030	0.000994036	Actual
0	Medium	4	5030	0.000795229	Actual
0	All	10	5030	0.001988072	Actual
0	High	1	5008	0.000199681	Actual
0	Low	1	5008	0.000199681	Actual
0	Medium	3	5008	0.000599042	Actual
0	All	5	5008	0.000998403	Actual
1	All	4.880162184	5008	0.000974473	Predicted
1	All	12.16736266	5008	0.002429585	Predicted
1	All	1.814446778	5008	0.00036231	Predicted
1	All	0.1705288	5008	3.41E-05	Predicted
1	All	6.290760082	5008	0.001256142	Predicted
1	All	7.815840807	5008	0.001560671	Predicted
1	All	13.47827404	5008	0.002691349	Predicted
1	All	9.678400805	5008	0.001932588	Predicted
1	All	3.870799189	5008	0.000772923	Predicted
1	All	10.11950012	5008	0.002020667	Predicted
1	All	6.369047979	5008	0.001271775	Predicted
1	All	3.629526048	5008	0.000724746	Predicted

Severity = injury cause score (0 - 8)
+
injury nature score (0 - 8)

Low = 0 to 7

Medium = 8 to 13

High = 13 to 16



Make this smarter with medical costs

Severity

	severity	claim_number	emp_count	injury_rate	type
0	All	8	4959	0.001613228	Actual
0	High	1	5069	0.000197278	Actual
0	Low	7	5069	0.001380943	Actual
0	Medium	3	5069	0.000591833	Actual
0	All	11	5069	0.002170053	Actual
0	High	1	5030	0.000198807	Actual
0	Low	5	5030	0.000994036	Actual
0	Medium	4	5030	0.000795229	Actual
0	All	10	5030	0.001988072	Actual
0	High	1	5008	0.000199681	Actual
0	Low	1	5008	0.000199681	Actual
0	Medium	3	5008	0.000599042	Actual
0	All	5	5008	0.000998403	Actual
1	All	4.880162184	5008	0.000974473	Predicted
1	All	12.16736266	5008	0.002429585	Predicted
1	All	1.814446778	5008	0.00036231	Predicted
1	All	0.1705288	5008	3.41E-05	Predicted
1	All	6.290760082	5008	0.001256142	Predicted
1	All	7.815840807	5008	0.001560671	Predicted
1	All	13.47827404	5008	0.002691349	Predicted
1	All	9.678400805	5008	0.001932588	Predicted
1	All	3.870799189	5008	0.000772923	Predicted
1	All	10.11950012	5008	0.002020667	Predicted
1	All	6.369047979	5008	0.001271775	Predicted
1	All	3.629526048	5008	0.000724746	Predicted

Methods to explore...

Clustering

Jenks natural breaks

Waterfall (if-else)

... we will discuss this together!

Severity

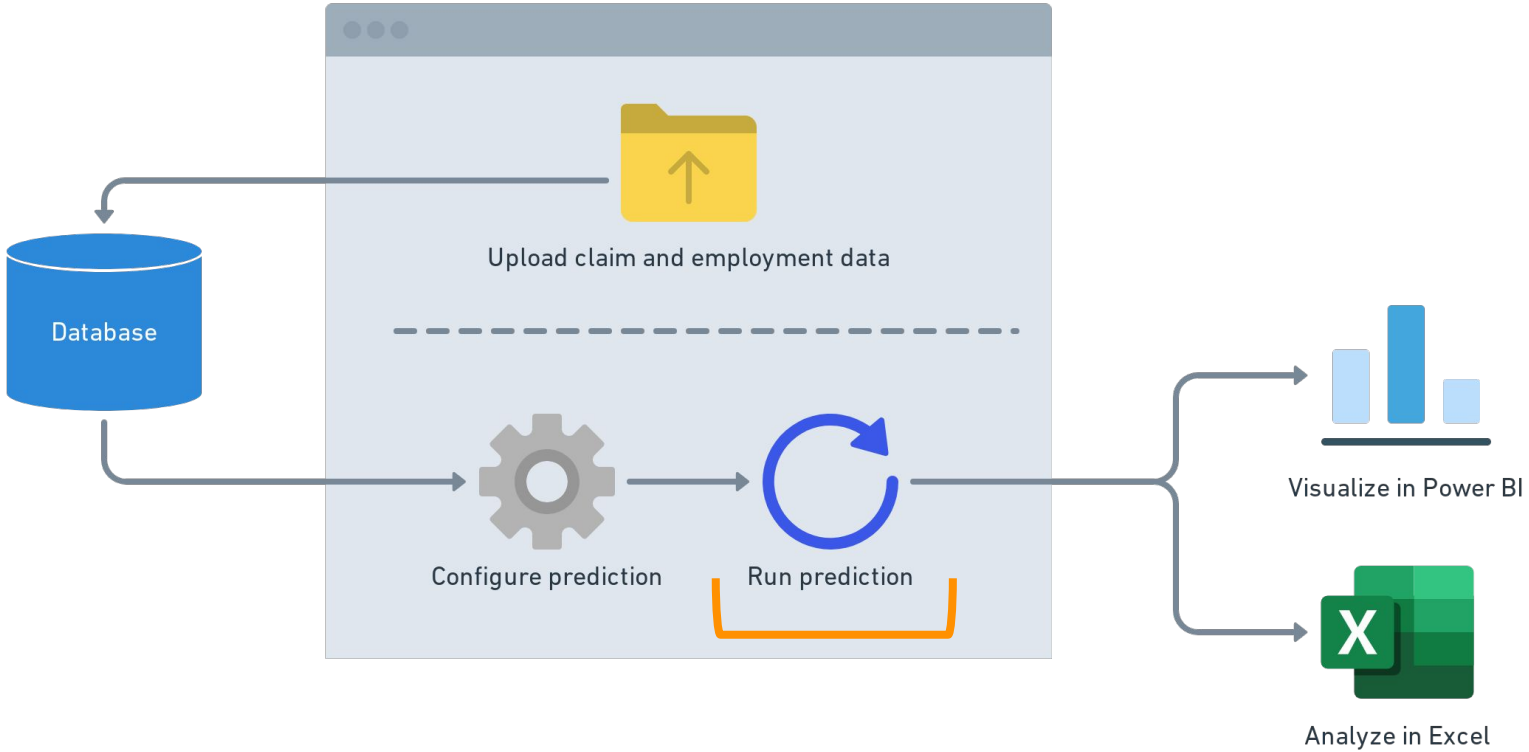
	severity	claim_number	emp_count	injury_rate	type
0	All	8	4959	0.001613228	Actual
0	High	1	5069	0.000197278	Actual
0	Low	7	5069	0.001380943	Actual
0	Medium	3	5069	0.000591833	Actual
0	All	11	5069	0.002170053	Actual
0	High	1	5030	0.000198807	Actual
0	Low	5	5030	0.000994036	Actual
0	Medium	4	5030	0.000795229	Actual
0	All	10	5030	0.001988072	Actual
0	High	1	5008	0.000199681	Actual
0	Low	1	5008	0.000199681	Actual
0	Medium	3	5008	0.000599042	Actual
0	All	5	5008	0.000998403	Actual
1	All	12.16736266	5008	0.002429585	Predicted
1	All	1.814446778	5008	0.00036231	Predicted
1	All	0.1705288	5008	3.41E-05	Predicted
1	All	6.290760082	5008	0.001256142	Predicted
1	All	7.815840807	5008	0.001560671	Predicted
1	All	13.47827404	5008	0.002691349	Predicted
1	All	9.678400805	5008	0.001932588	Predicted
1	All	3.870759189	5008	0.000772923	Predicted
1	All	10.11950012	5008	0.002020667	Predicted
1	All	6.369047979	5008	0.001271775	Predicted
1	All	3.629526048	5008	0.000724746	Predicted

Severity is not a factor in prediction model.

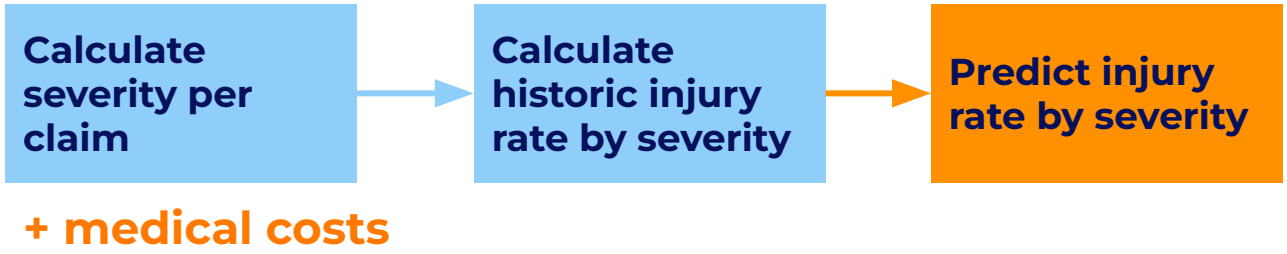
It is used to report on historic activity.

Predicted severity rates is something we are looking into extending.

Redefining severity with medical costs



Redefining severity with medical costs



Our main deliverables

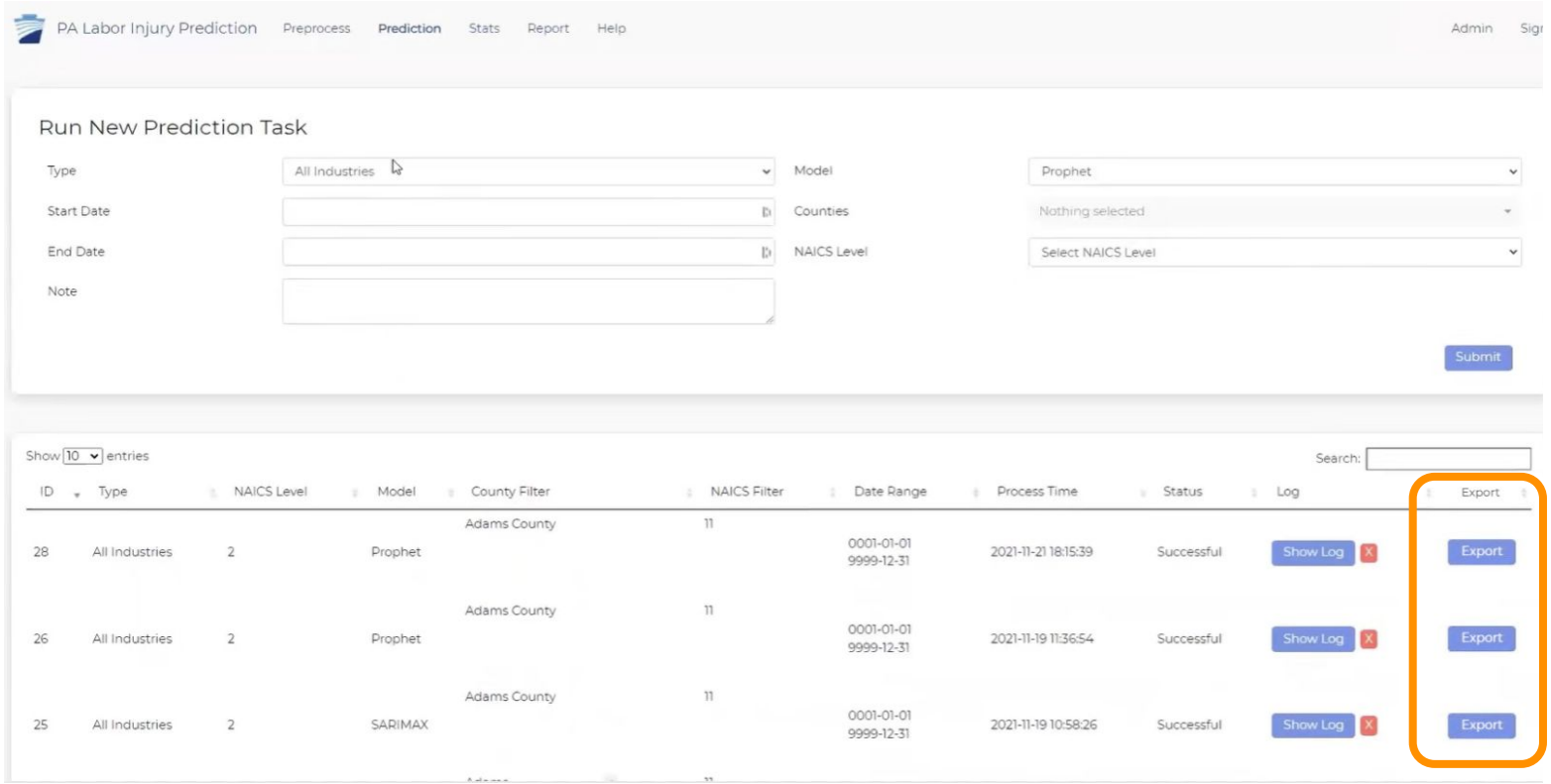
Upload medical costs to database through the web application

Analyze injury rates by severity, updated with medical costs

Surface prediction results in Excel

Visualize top 5 counties/industries injury rate and medical costs in Power BI

Surfacing prediction output in Excel



Our main deliverables

Upload medical costs to database through the web application

Analyze injury rates by severity, updated with medical costs

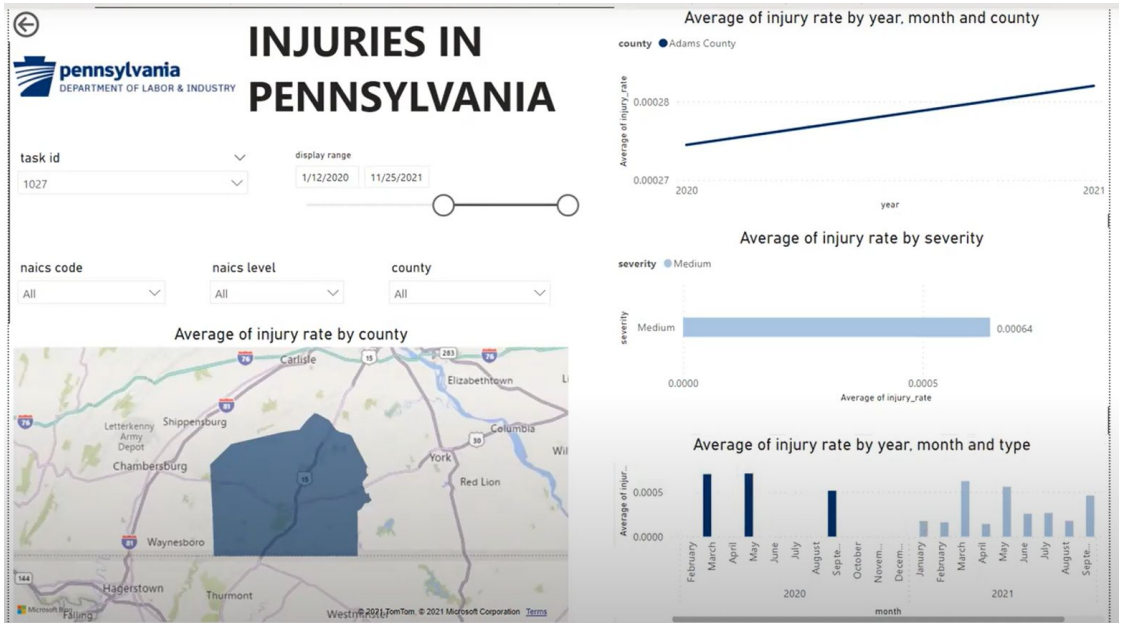
Surface prediction results in Excel

Visualize top 5 counties/industries injury rate and medical costs in Power BI

Enable analysis by medical costs

Add in **top 10** counties & industries with high injury rates

Add analysis of **medical costs**



Our main deliverables

Upload medical costs to database through the web application

Analyze injury rates by severity, updated with medical costs

Surface prediction results in Excel

Update code in Github repository

Visualize top 5 counties/industries injury rate and medical costs in Power BI

Update Power BI dashboard

& create overall support documentation

Another tool in toolbox to make informed, strategic decisions and reduce the number of injuries at work.

Progress
on deliverables

Progress on deliverables

Exploring the data

Setting up the application

Running predictions

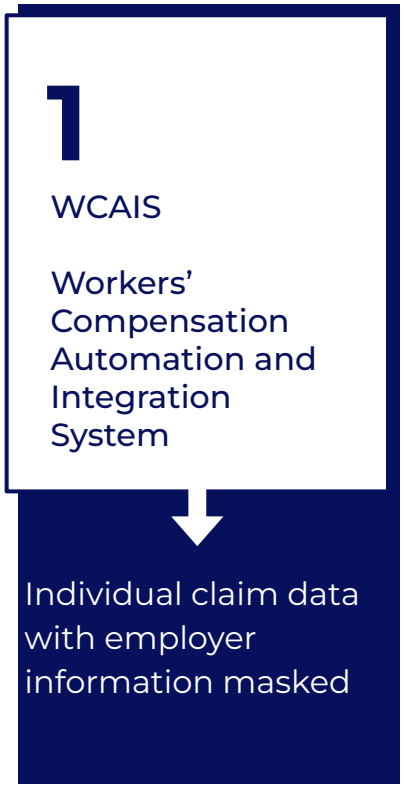
Progress on deliverables

Exploring the data

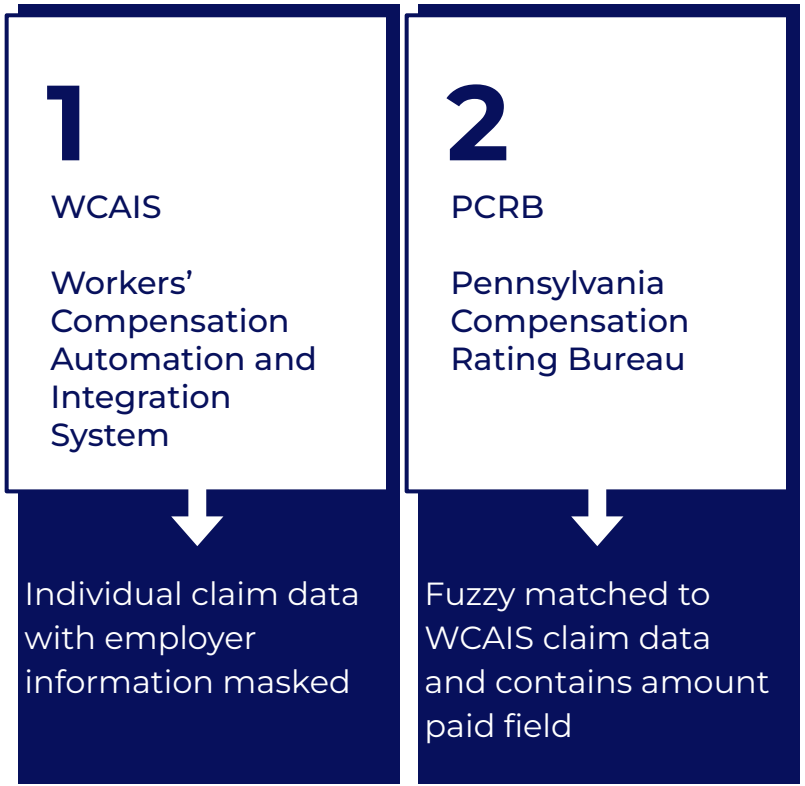
Setting up the application

Running predictions

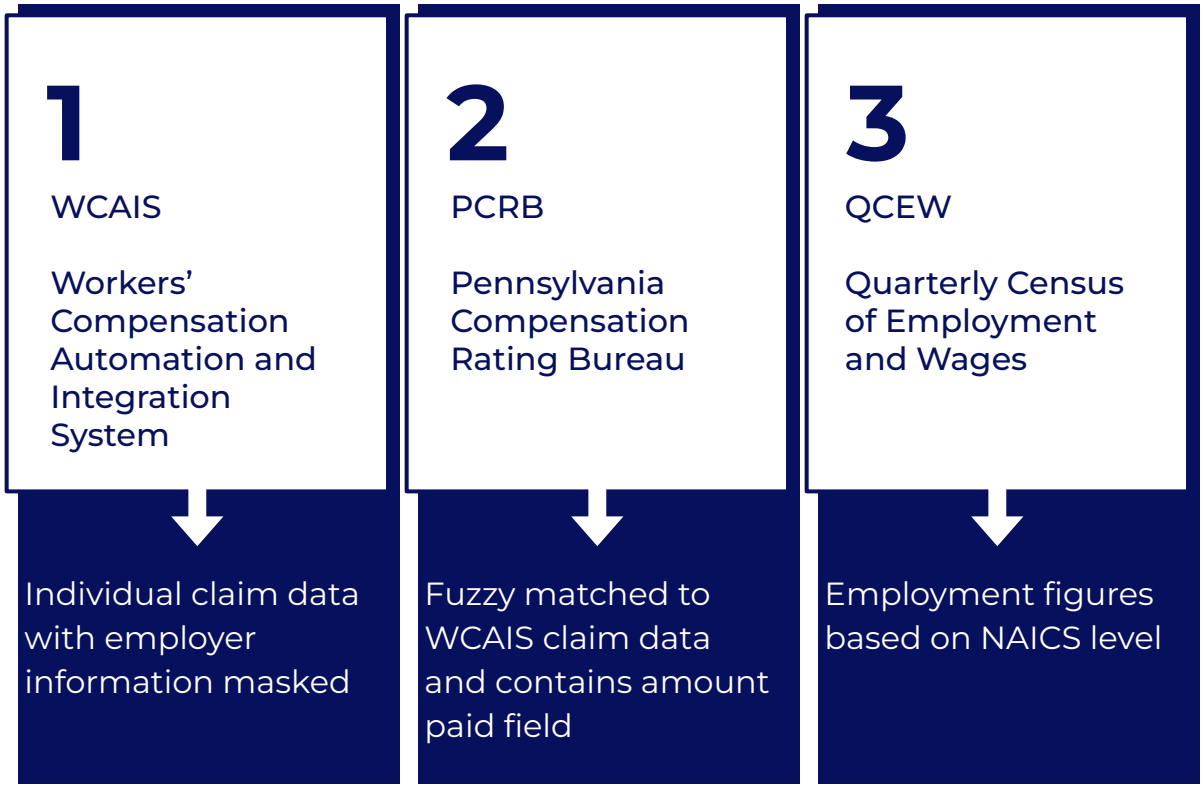
Data Source Overview



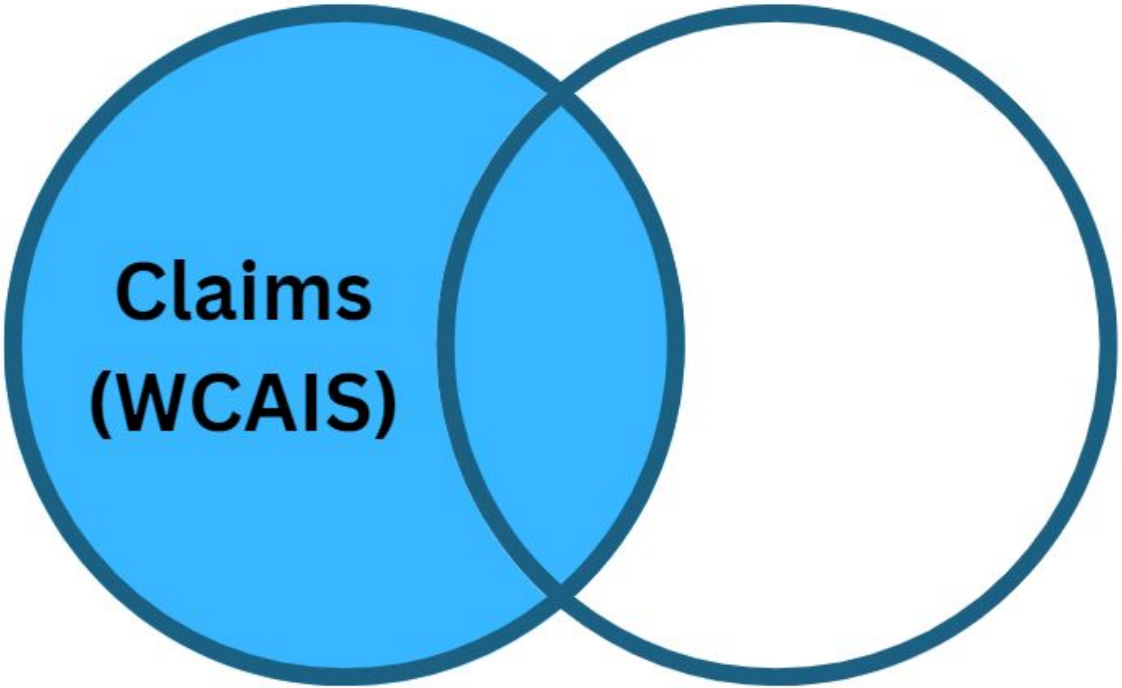
Data Source Overview



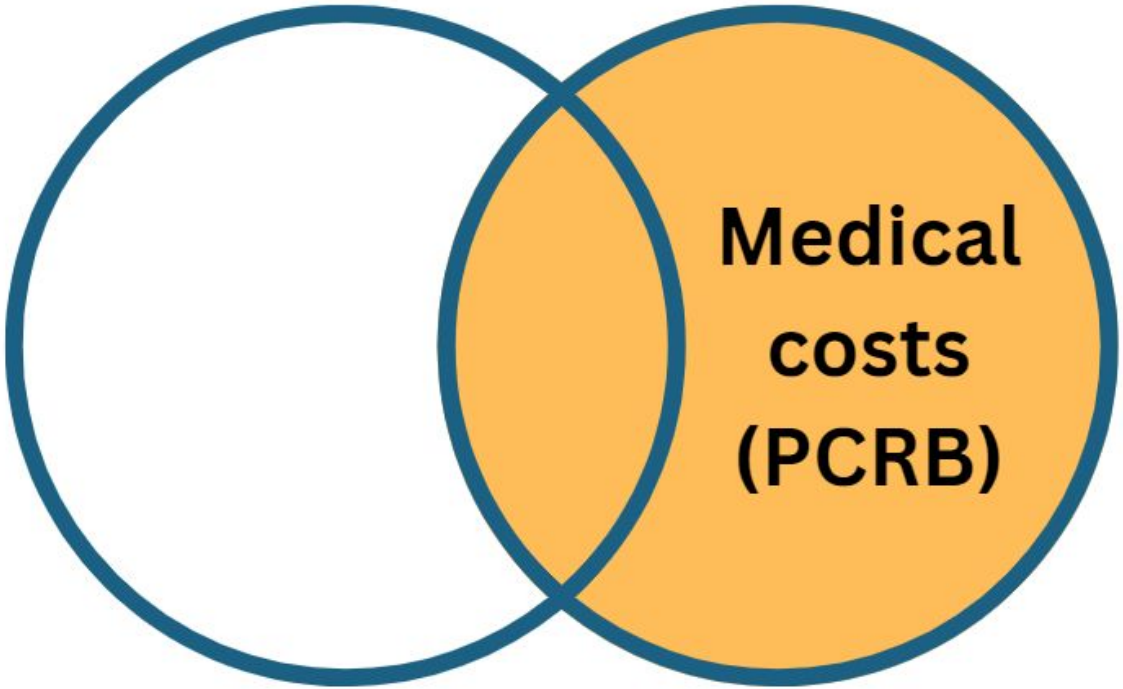
Data Source Overview



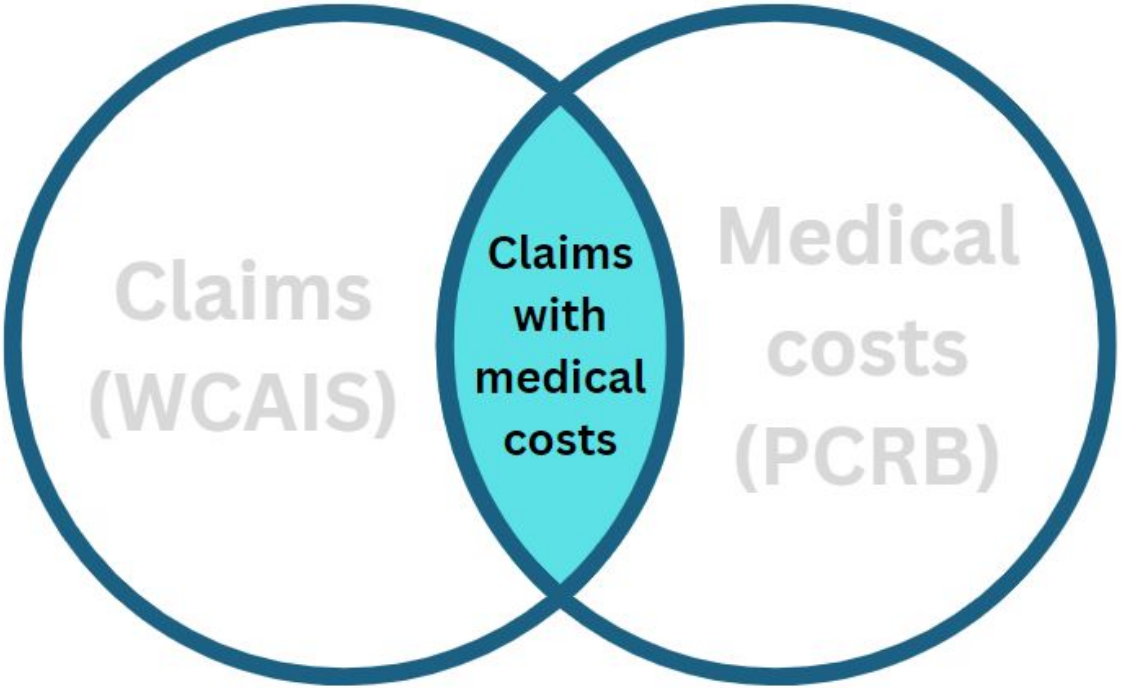
Integrating medical costs with claims



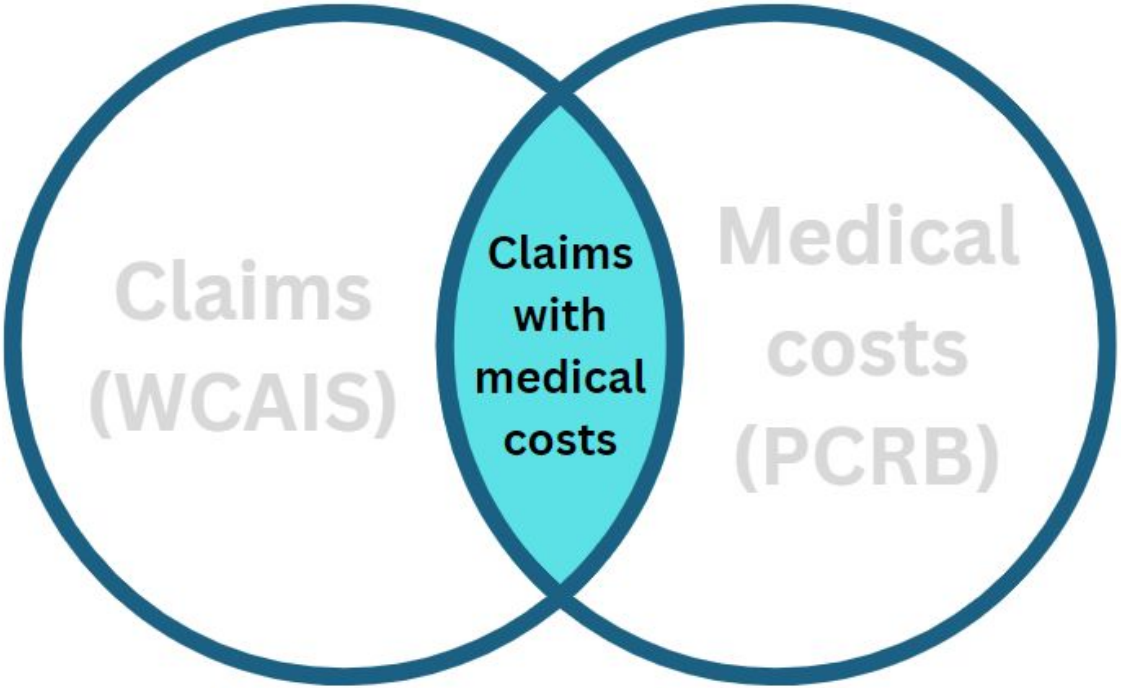
Integrating medical costs with claims



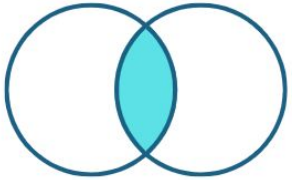
Integrating medical costs with claims



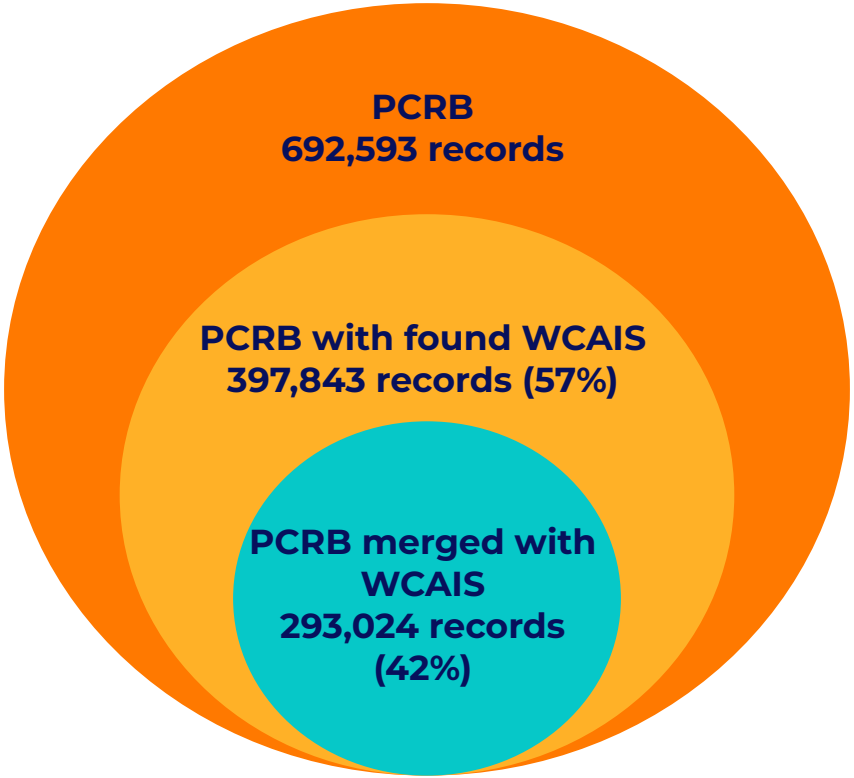
Integrating medical costs with claims



42% match, PCRB & WCAIS



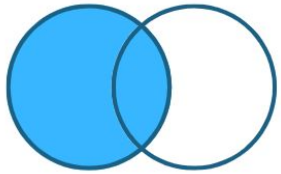
Claims with medical costs

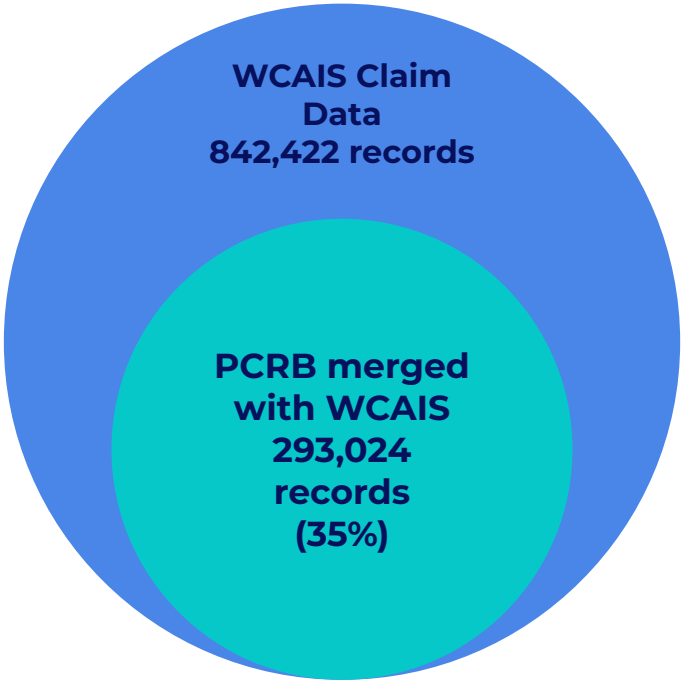


Source: WCAIS & PCRB

2018 excluded

35% match, WCAIS with costs

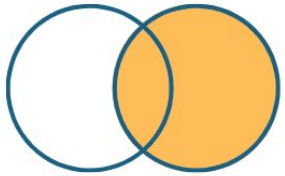




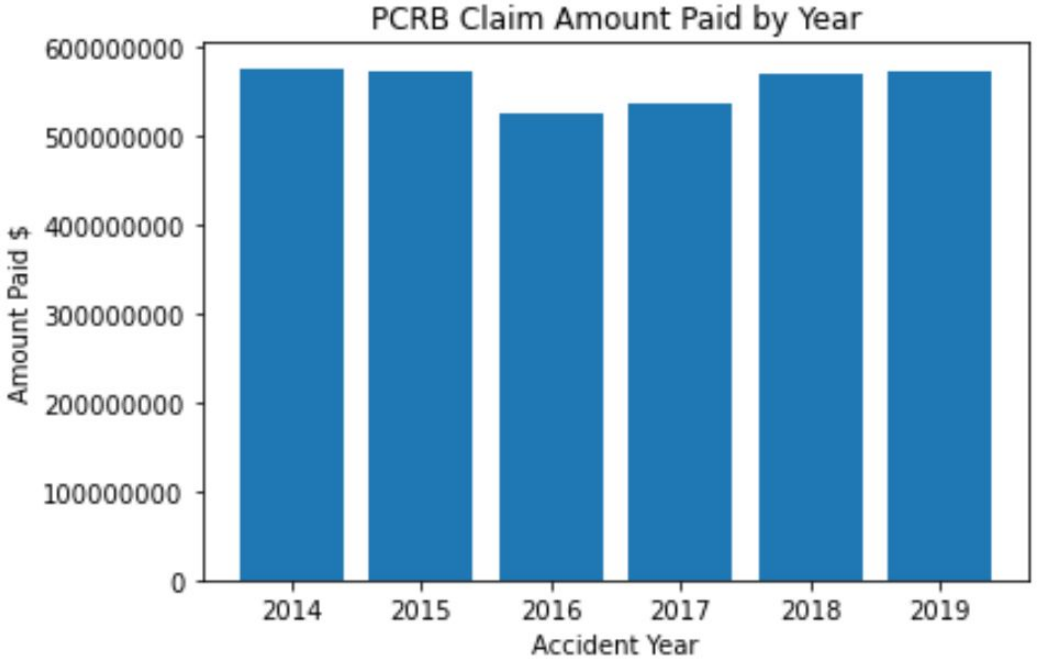
Source: WCAIS & PCRB

2018 excluded

Annual medical costs from injuries around $550 million

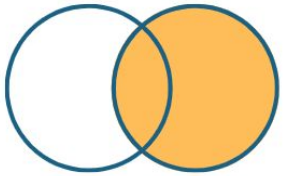


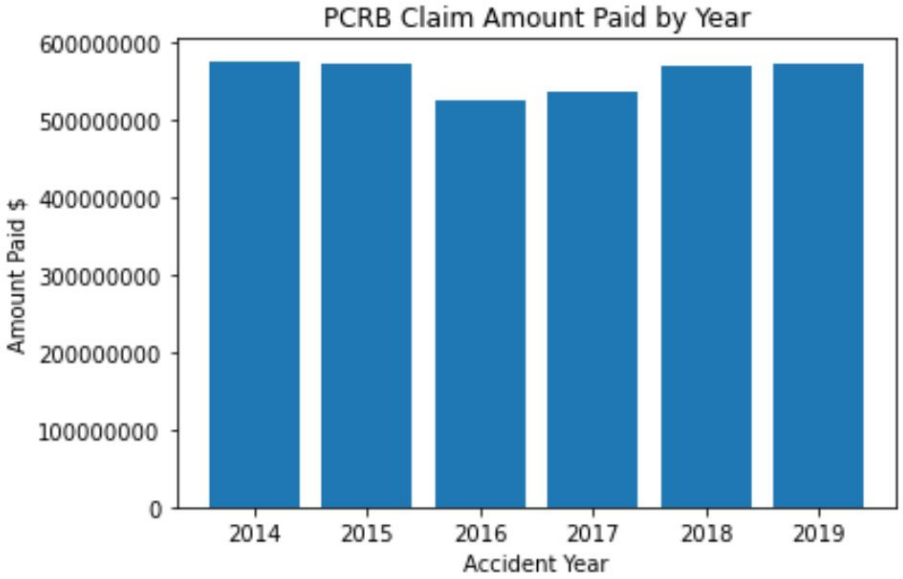
Medical costs (PCRB)

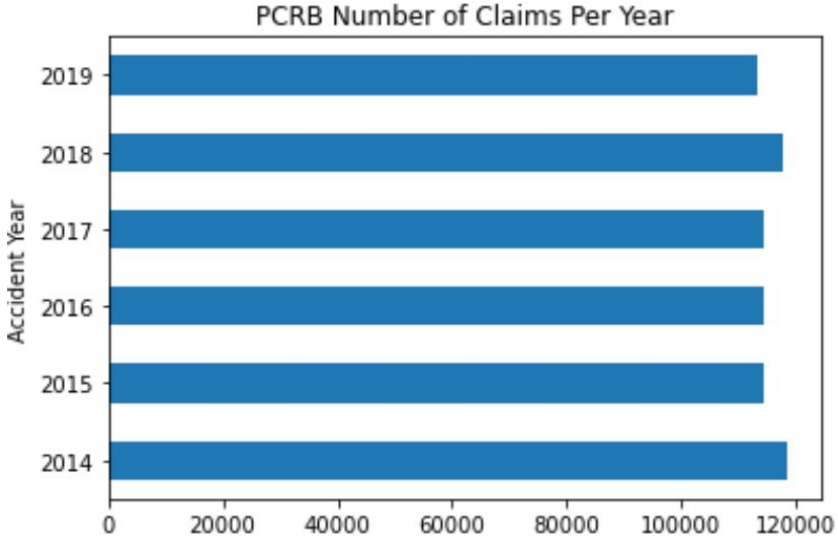


Accident Year	Amount Paid Tot
2014	$575,025,579.38
2015	$570,130,015.80
2016	$523,704,153.20
2017	$534,030,162.99
2018	$568,108,963.80
2019	$571,721,297.00

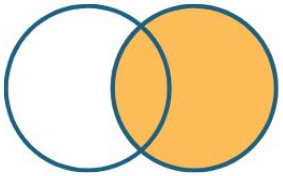
Annual medical costs and claims stable year over year

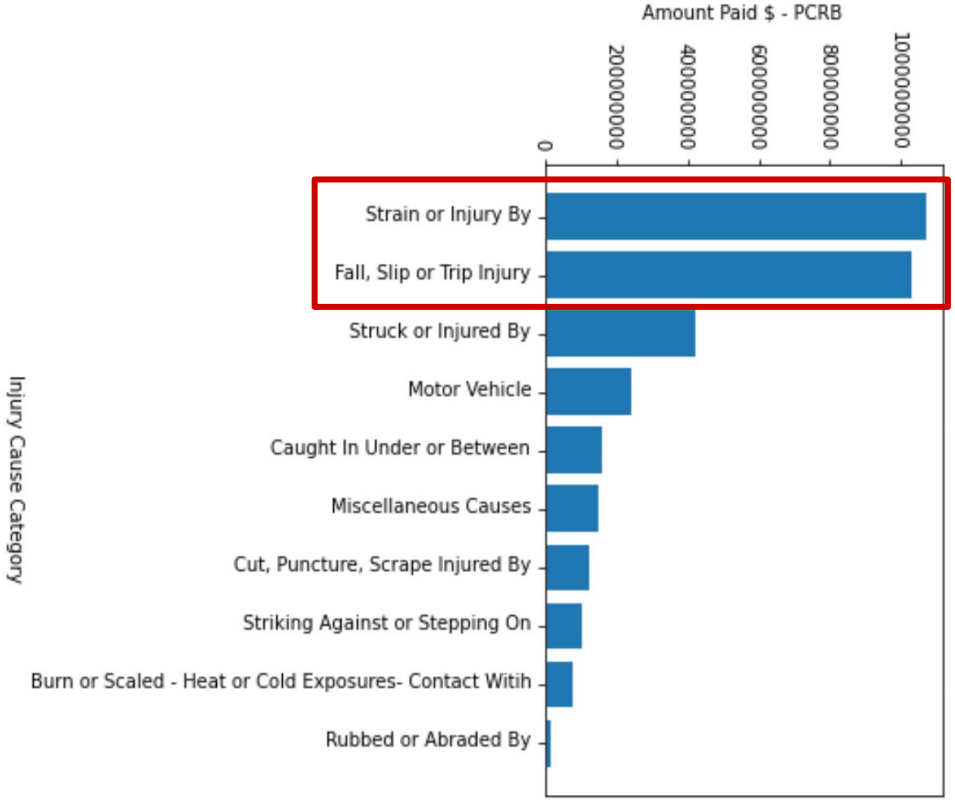






Strains and slips drive high costs

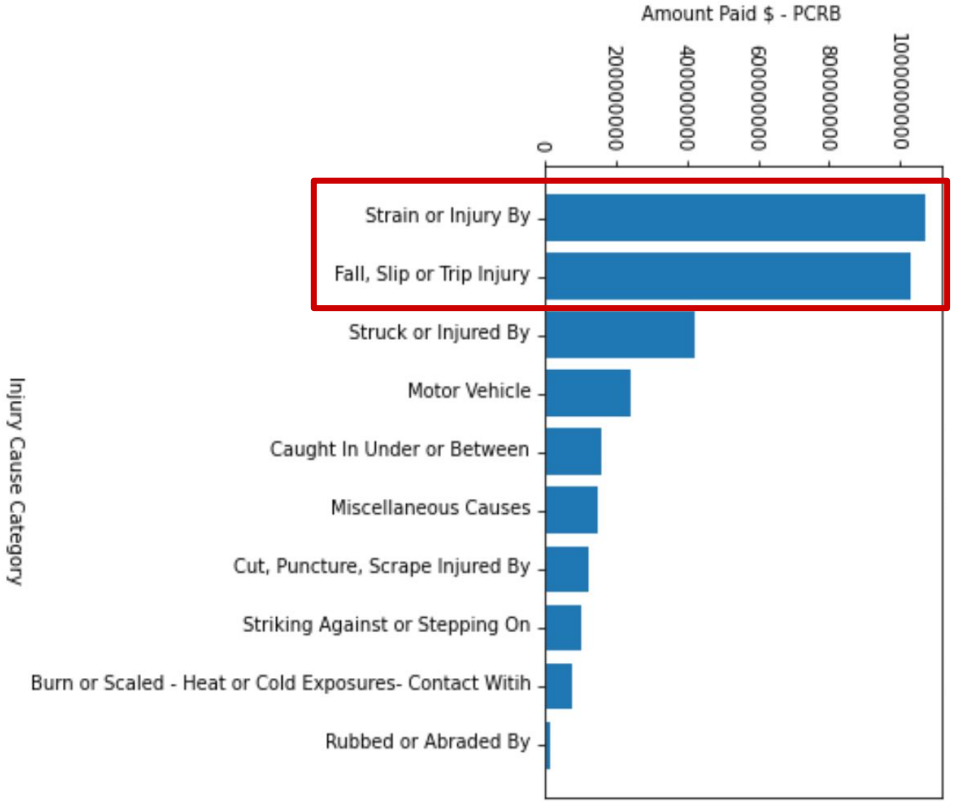




60% of total costs

2014-2019

Strains and slips driving high costs paid

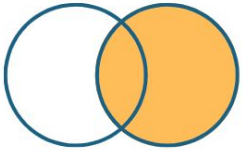


$2 billion

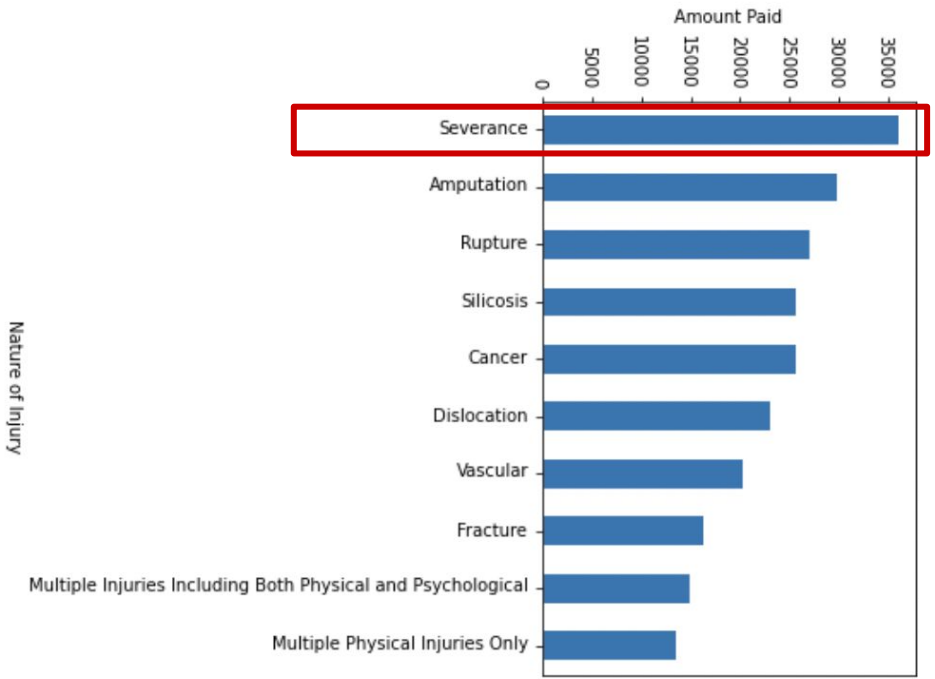
CAUSE_CAT	Amount Paid Tot	Cum Pct
Strain or Injury By	$1,066,026,477.82	31.89
Fall, Slip or Trip Injury	$1,022,280,263.91	62.47
Struck or Injured By	$415,057,333.44	74.89
Motor Vehicle	$238,776,769.06	82.03
Caught In Under or Between	$153,010,242.09	86.61
Miscellaneous Causes	$142,972,799.77	90.89
Cut, Puncture, Scrape Injured By	$119,348,388.30	94.46
Striking Against or Stepping On	$100,035,406.15	97.45
Burn or Scaled - Heat or Cold Exposures- Conta...	$73,916,522.15	99.66
Rubbed or Abraded By	$11,295,969.48	100.00

2014-2019

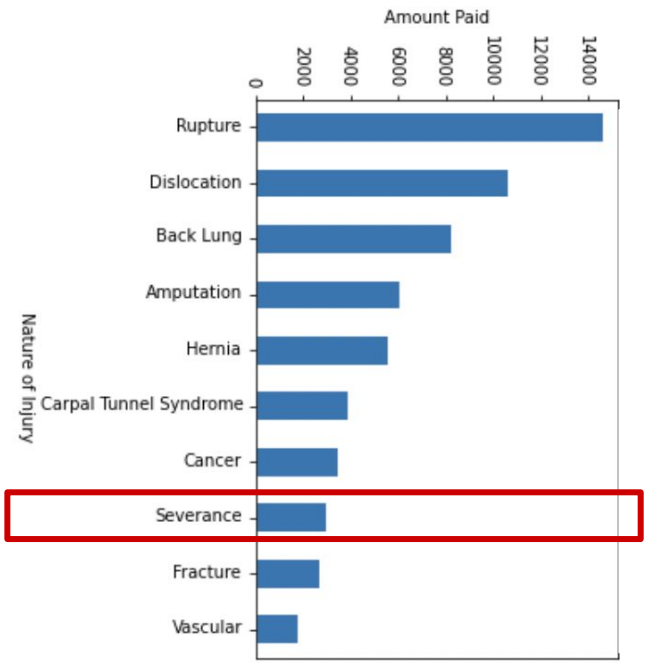
Medical costs data shows skew



Mean costs



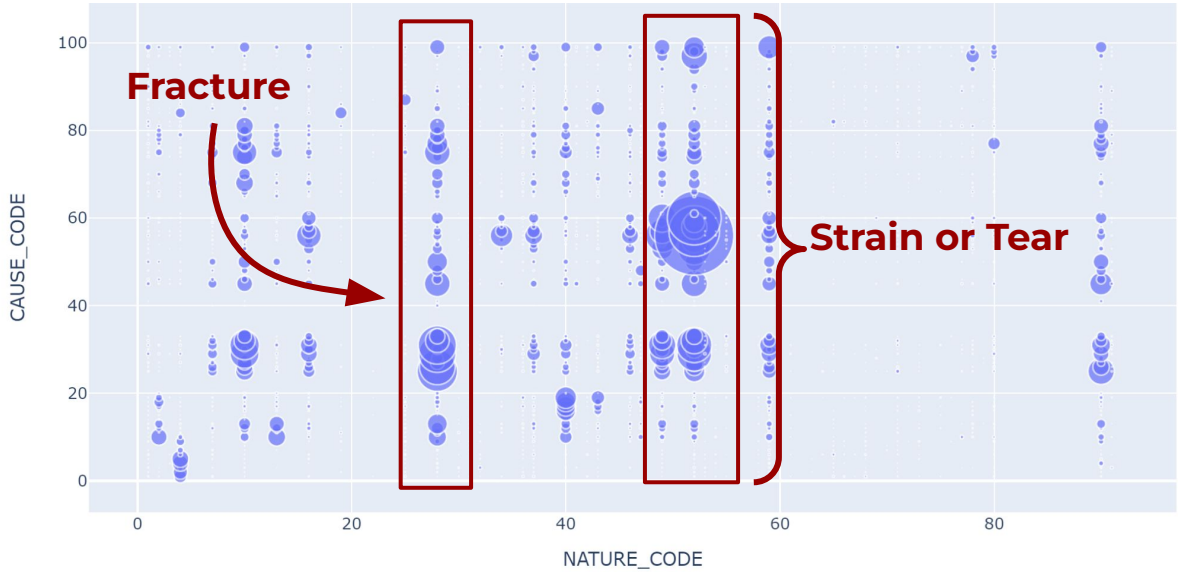
Median costs



2014-2019

Total costs by injury nature and cause

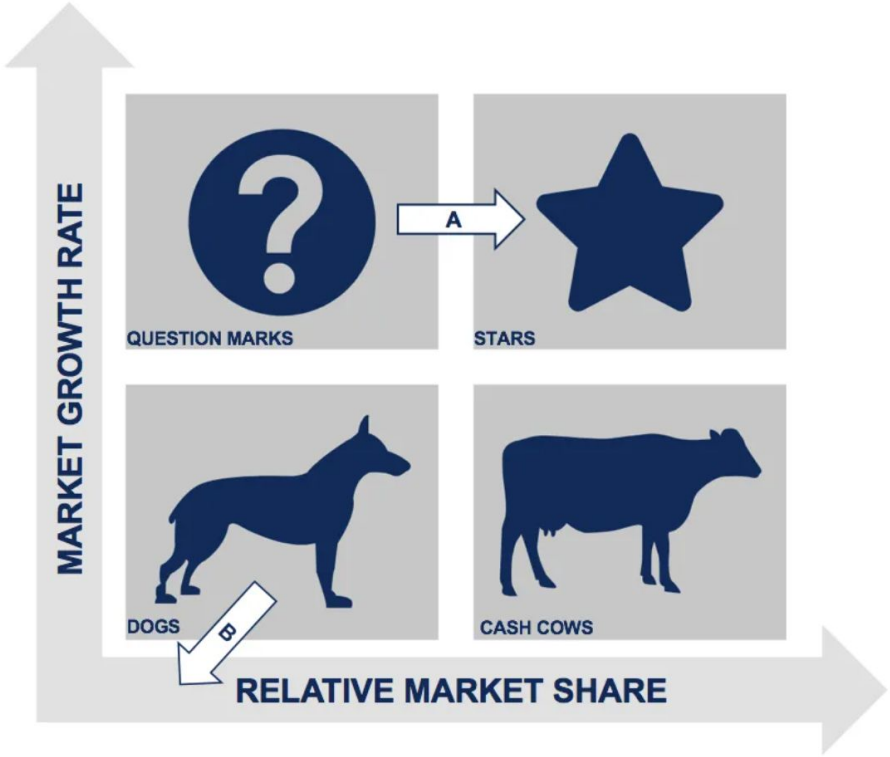
Medical costs (PCRB)



NATURE_DESC	NATURE_CODE	CAUSE_DESC	CAUSE_CODE	Amount Paid Tot	Cum Pct
Strain or Tear	52.0	Lifting	56	$259,612,806.36	7.77
Strain or Tear	52.0	Strain by Injury By, NOC	60	$119,615,452.80	11.34
Strain or Tear	52.0	Pushing or Pulling	57	$106,445,027.05	14.53
Fracture	28.0	From Different Level (Elevation)	25	$68,175,935.10	16.57
Fracture	28.0	Fall, Slip or Trip, NOC	31	$60,151,219.06	18.37
Fracture	28.0	On Same Level	29	$58,047,182.86	20.10
Fracture	28.0	From Ladder or Scaffolding	26	$52,333,230.12	21.67

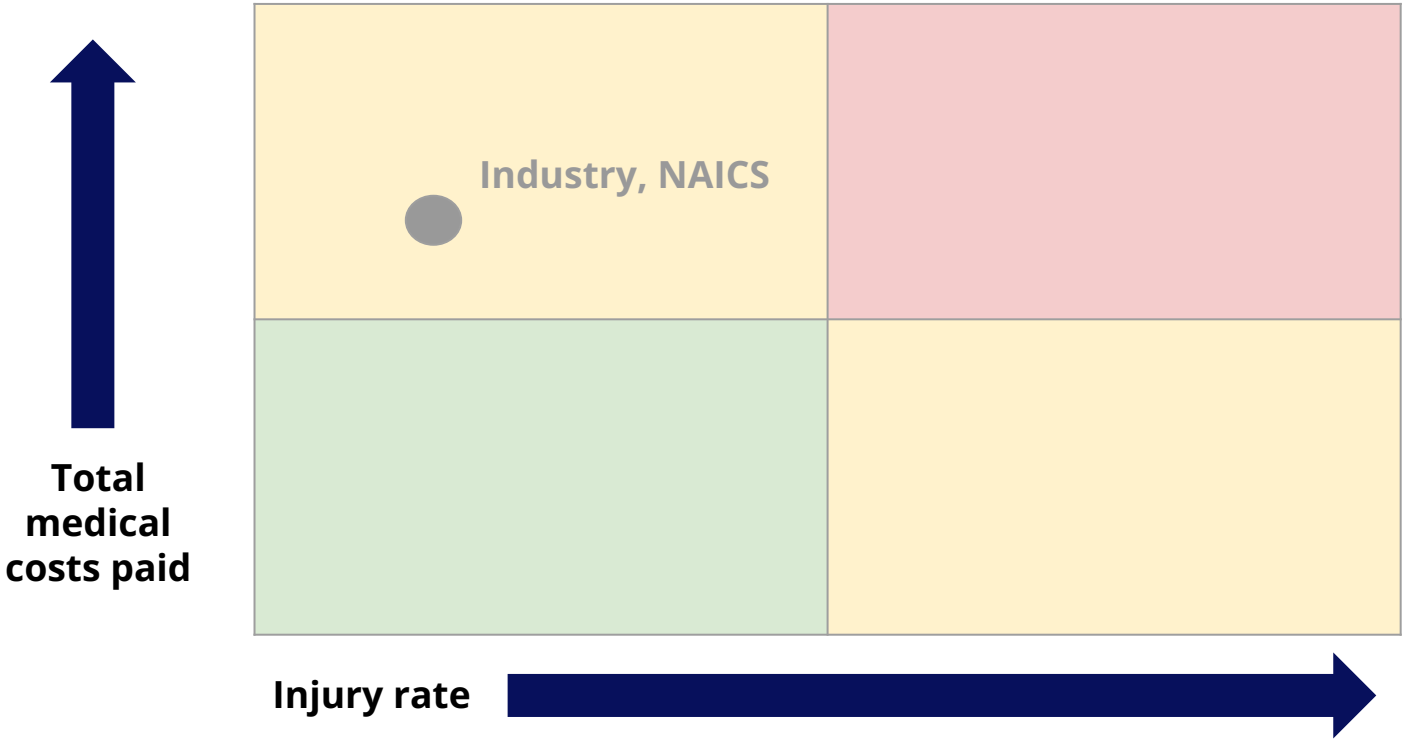
2014-2019

Boston Consulting Group Matrix

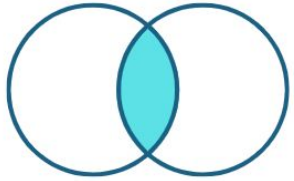


Source: Corporate Finance Institute (2022)

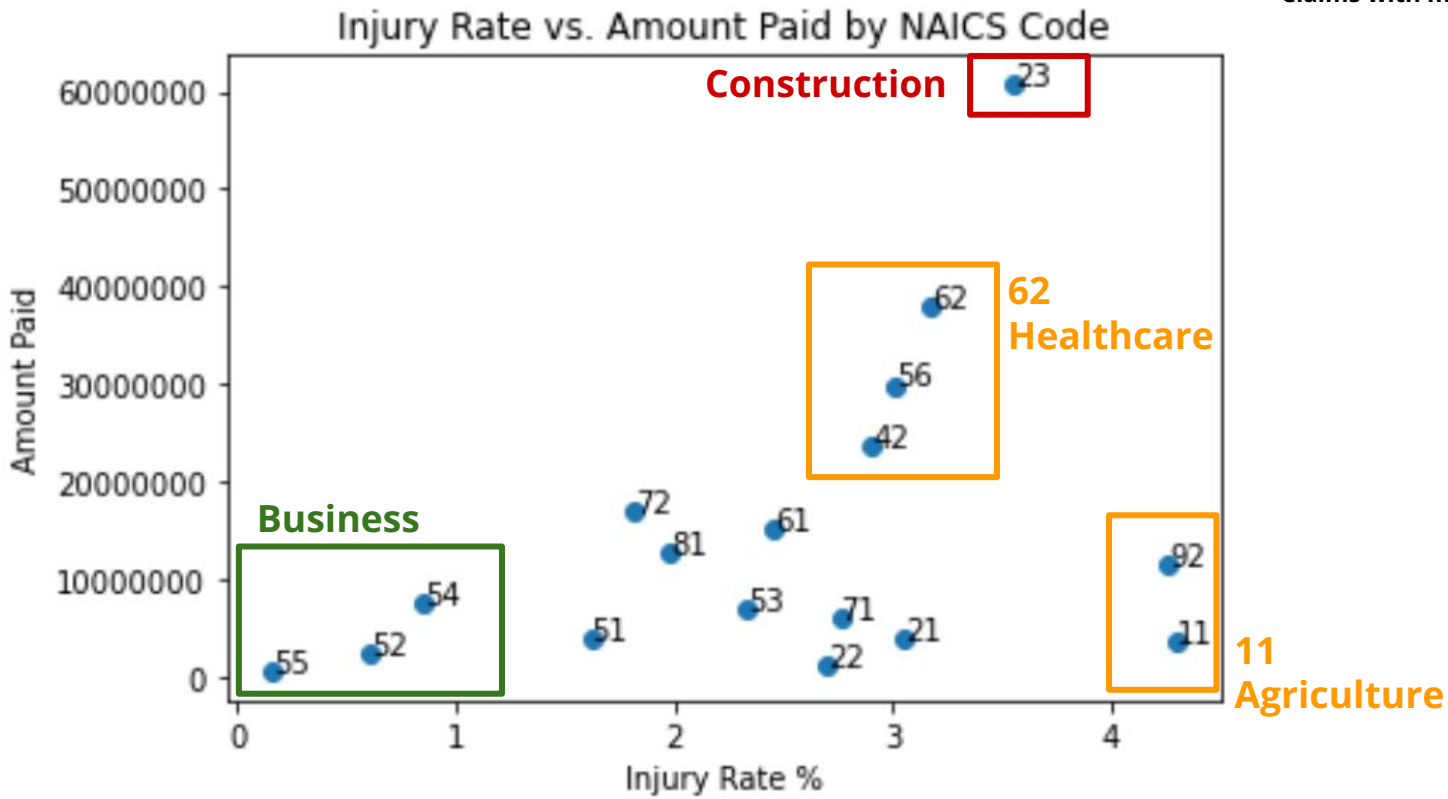
Identifying key areas



Identifying key areas in 2019



Claims with medical costs



Future analysis

Identify more relationships between industries, regions, and claims

Develop a method to incorporate these relationships into the severity calculation

Create intuitive visualizations to identify key areas to monitor

Progress on deliverables

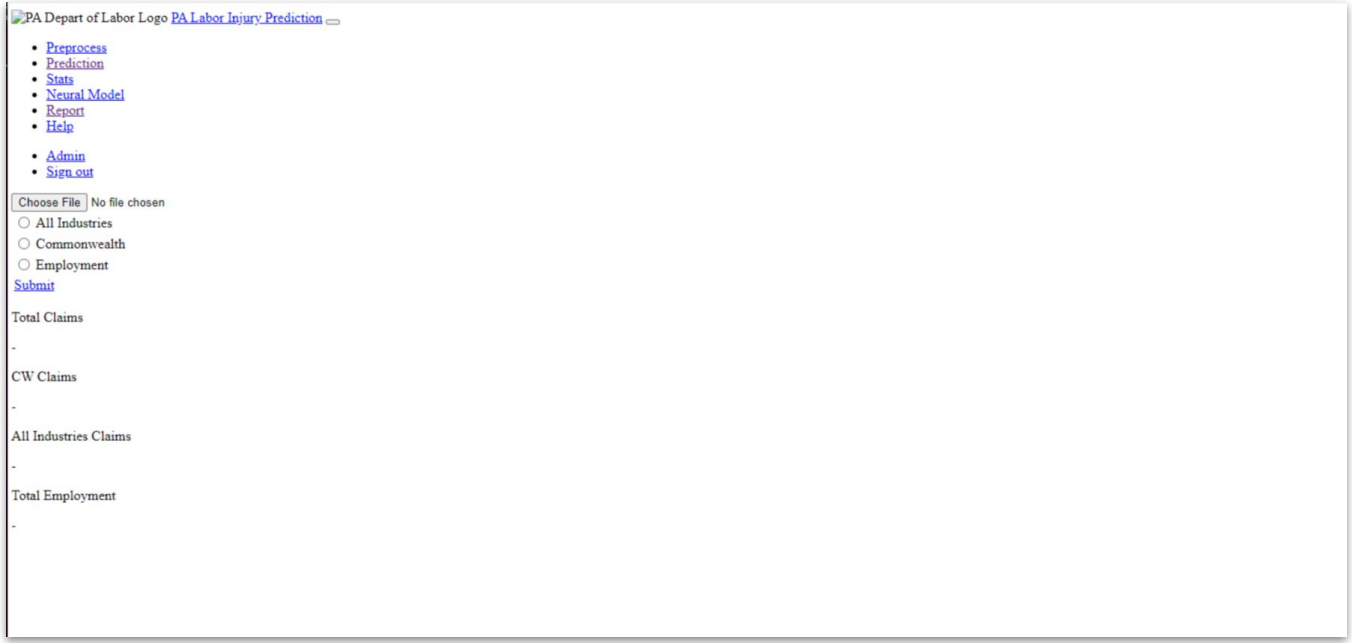
Exploring the data

Setting up the application

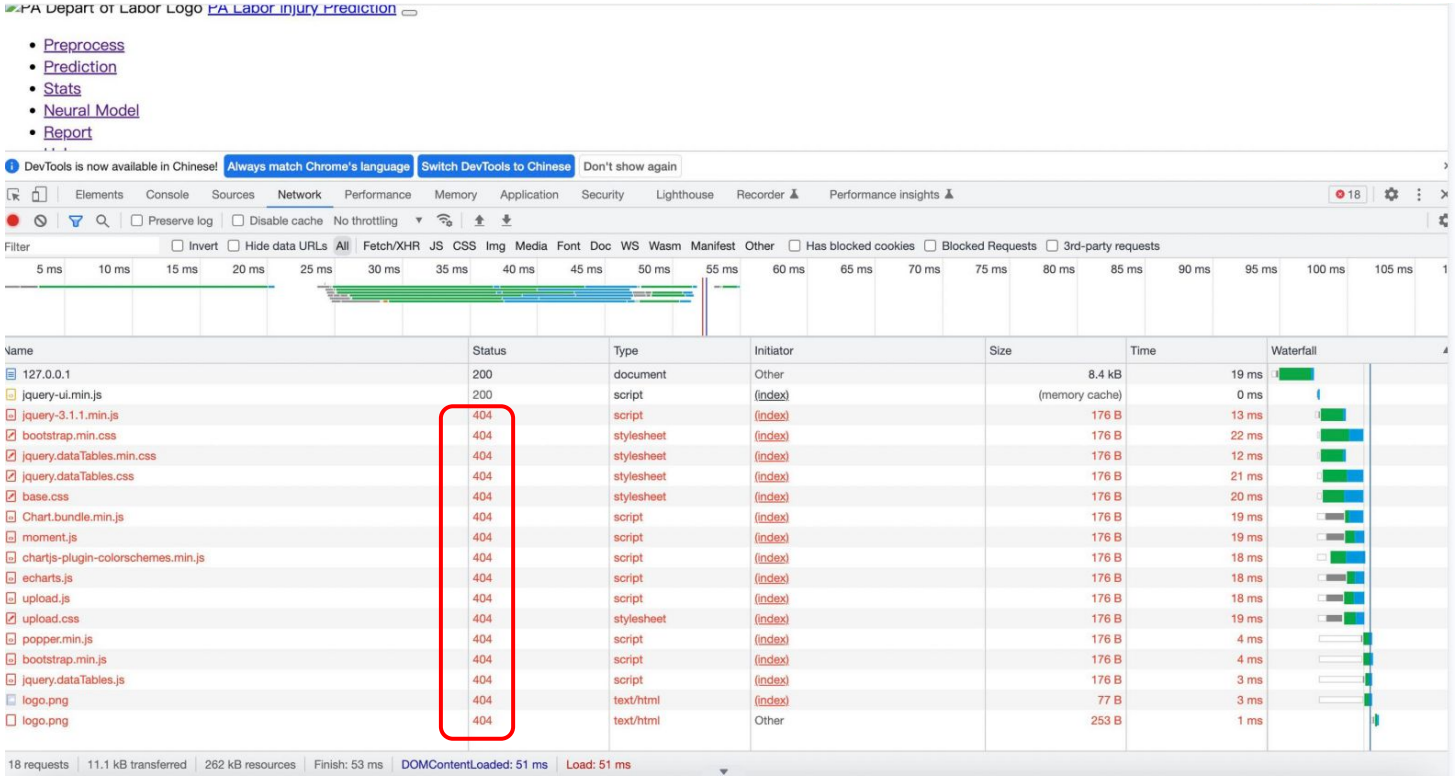
Running predictions

Setting up the application

Website components not loading correctly



Setting up the application



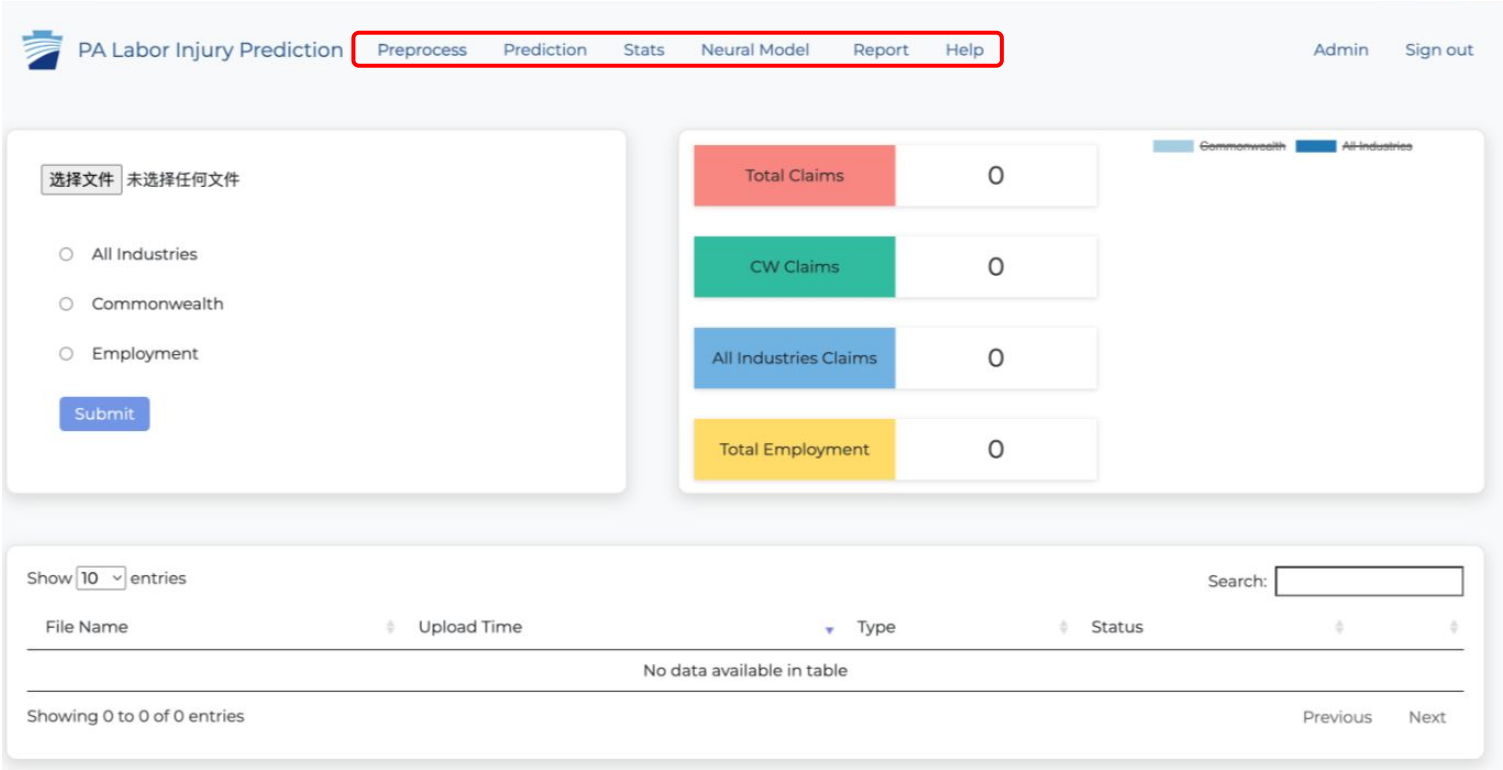
Setting up the application

Local files do exist ✓

Setting files correct ✗

Working to solve

Setting up the application



Onboarding takes a while

Took time for our team to get the application running on our computers

Faced issues of:

- Outdated software versions
- Compatibility issues

Clearer document onboarding steps

Took time for our team to get the application running on our computers

Faced issues of:

- Outdated software versions
- Compatibility issues

To help for future onboarding:

- Updating software requirements
- Adding detailed documentation to repository

Progress on deliverables

Exploring the data

Setting up the application

Running predictions

Ran predictions locally

Able to generate predictions by extracting model code and running it locally on a Jupyter notebook.

Ran prediction on a small dataset:

- Adams County
- NAICS code = 62 (Healthcare and Social Assistance Sector)
- 2020

Ran predictions locally

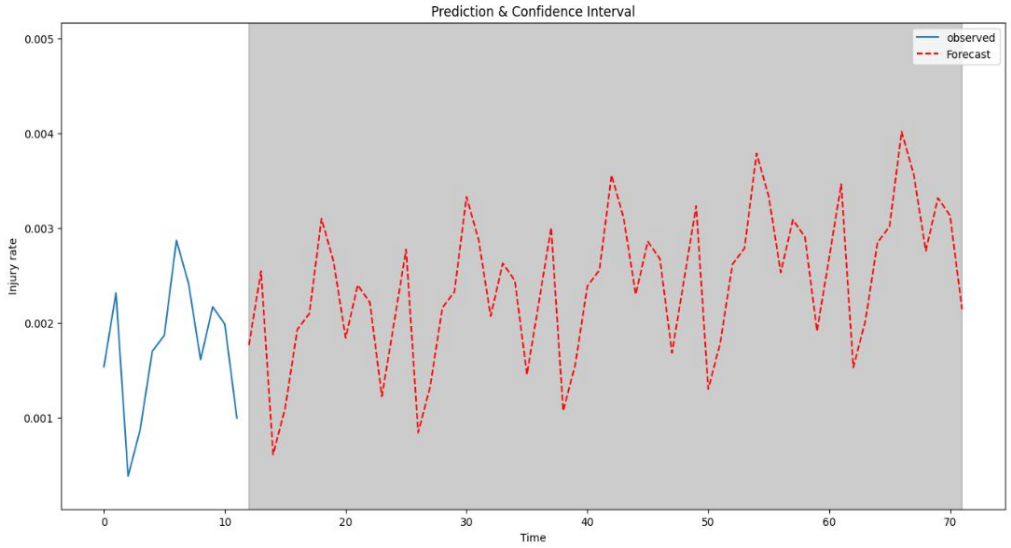
Able to generate predictions by extracting model code and running it locally on a Jupyter notebook.

Ran prediction on a small dataset:

- Adams County
- NAICS code = 62 (Healthcare and Social Assistance Sector)
- 2020

Produced forecasted injury rates for 2021

Jupyter notebook a tool for developing outside of the web application



Historic values

id	naics_code	county	year	month	date	severity	claim_number	emp_count	injury_rate	type	county_formatted	month_str	naics_level	commonwealth	task_id
27	62	Adams County	2020	9	9/1/2020	All	8	4959	0.001613228	Actual	Adams County PA	September	2	TRUE	11
28	62	Adams County	2020	10	10/1/2020	High	1	5069	0.000197278	Actual	Adams County PA	October	2	TRUE	11
29	62	Adams County	2020	10	10/1/2020	Low	7	5069	0.001380943	Actual	Adams County PA	October	2	TRUE	11
30	62	Adams County	2020	10	10/1/2020	Medium	3	5069	0.000591833	Actual	Adams County PA	October	2	TRUE	11
31	62	Adams County	2020	10	10/1/2020	All	11	5069	0.002170053	Actual	Adams County PA	October	2	TRUE	11
32	62	Adams County	2020	11	11/1/2020	High	1	5030	0.000198807	Actual	Adams County PA	November	2	TRUE	11
33	62	Adams County	2020	11	11/1/2020	Low	5	5030	0.000994036	Actual	Adams County PA	November	2	TRUE	11
34	62	Adams County	2020	11	11/1/2020	Medium	4	5030	0.000795229	Actual	Adams County PA	November	2	TRUE	11
35	62	Adams County	2020	11	11/1/2020	All	10	5030	0.001988072	Actual	Adams County PA	November	2	TRUE	11
36	62	Adams County	2020	12	12/1/2020	High	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
37	62	Adams County	2020	12	12/1/2020	Low	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
38	62	Adams County	2020	12	12/1/2020	Medium	3	5008	0.000599042	Actual	Adams County PA	December	2	TRUE	11
39	62	Adams County	2020	12	12/1/2020	All	5	5008	0.000998403	Actual	Adams County PA	December	2	TRUE	11

Predicted values

0	62	Adams County	2021	1	1/1/2021	All	4.880162184	5008	0.000974473	Predicted	Adams County PA	January	2	TRUE	11
1	62	Adams County	2021	2	2/1/2021	All	12.16736266	5008	0.002429585	Predicted	Adams County PA	February	2	TRUE	11
2	62	Adams County	2021	3	3/1/2021	All	1.814446778	5008	0.00036231	Predicted	Adams County PA	March	2	TRUE	11
3	62	Adams County	2021	4	4/1/2021	All	0.1705288	5008	3.41E-05	Predicted	Adams County PA	April	2	TRUE	11
4	62	Adams County	2021	5	5/1/2021	All	6.290760082	5008	0.001256142	Predicted	Adams County PA	May	2	TRUE	11
5	62	Adams County	2021	6	6/1/2021	All	7.815840807	5008	0.001560671	Predicted	Adams County PA	June	2	TRUE	11
6	62	Adams County	2021	7	7/1/2021	All	13.47827404	5008	0.002691349	Predicted	Adams County PA	July	2	TRUE	11
7	62	Adams County	2021	8	8/1/2021	All	9.678400805	5008	0.001932588	Predicted	Adams County PA	August	2	TRUE	11
8	62	Adams County	2021	9	9/1/2021	All	3.870799189	5008	0.000772923	Predicted	Adams County PA	September	2	TRUE	11
9	62	Adams County	2021	10	10/1/2021	All	10.11950012	5008	0.002020667	Predicted	Adams County PA	October	2	TRUE	11
10	62	Adams County	2021	11	11/1/2021	All	6.369047979	5008	0.001271775	Predicted	Adams County PA	November	2	TRUE	11
11	62	Adams County	2021	12	12/1/2021	All	3.629526048	5008	0.000724746	Predicted	Adams County PA	December	2	TRUE	11

Injury rates

id	naics_code	county	year	month	date	severity	claim_number	emp_coun	injury_rate	type	county_formatted	month_str	naics_level	commonwealth	task_id
27	62	Adams County	2020	9	9/1/2020	All	8	495	0.001613228	Actual	Adams County PA	September	2	TRUE	11
28	62	Adams County	2020	10	10/1/2020	High	1	506	0.000197278	Actual	Adams County PA	October	2	TRUE	11
29	62	Adams County	2020	10	10/1/2020	Low	7	506	0.001380943	Actual	Adams County PA	October	2	TRUE	11
30	62	Adams County	2020	10	10/1/2020	Medium	3	506	0.000591833	Actual	Adams County PA	October	2	TRUE	11
31	62	Adams County	2020	10	10/1/2020	All	11	506	0.002170053	Actual	Adams County PA	October	2	TRUE	11
32	62	Adams County	2020	11	11/1/2020	High	1	503	0.000198807	Actual	Adams County PA	November	2	TRUE	11
33	62	Adams County	2020	11	11/1/2020	Low	5	503	0.000994036	Actual	Adams County PA	November	2	TRUE	11
34	62	Adams County	2020	11	11/1/2020	Medium	4	503	0.000795229	Actual	Adams County PA	November	2	TRUE	11
35	62	Adams County	2020	11	11/1/2020	All	10	503	0.001988072	Actual	Adams County PA	November	2	TRUE	11
36	62	Adams County	2020	12	12/1/2020	High	1	500	0.000199681	Actual	Adams County PA	December	2	TRUE	11
37	62	Adams County	2020	12	12/1/2020	Low	1	500	0.000199681	Actual	Adams County PA	December	2	TRUE	11
38	62	Adams County	2020	12	12/1/2020	Medium	3	500	0.000599042	Actual	Adams County PA	December	2	TRUE	11
39	62	Adams County	2020	12	12/1/2020	All	5	500	0.000998403	Actual	Adams County PA	December	2	TRUE	11
0	62	Adams County	2021	1	1/1/2021	All	4.880162184	500	0.000974473	Predicted	Adams County PA	January	2	TRUE	11
1	62	Adams County	2021	2	2/1/2021	All	12.16736266	500	0.002429585	Predicted	Adams County PA	February	2	TRUE	11
2	62	Adams County	2021	3	3/1/2021	All	1.814446778	500	0.00036231	Predicted	Adams County PA	March	2	TRUE	11
3	62	Adams County	2021	4	4/1/2021	All	0.1705288	500	3.41E-05	Predicted	Adams County PA	April	2	TRUE	11
4	62	Adams County	2021	5	5/1/2021	All	6.290760082	500	0.001256142	Predicted	Adams County PA	May	2	TRUE	11
5	62	Adams County	2021	6	6/1/2021	All	7.815840807	500	0.001560671	Predicted	Adams County PA	June	2	TRUE	11
6	62	Adams County	2021	7	7/1/2021	All	13.47827404	500	0.002691349	Predicted	Adams County PA	July	2	TRUE	11
7	62	Adams County	2021	8	8/1/2021	All	9.678400805	500	0.001932588	Predicted	Adams County PA	August	2	TRUE	11
8	62	Adams County	2021	9	9/1/2021	All	3.870799189	500	0.000772923	Predicted	Adams County PA	September	2	TRUE	11
9	62	Adams County	2021	10	10/1/2021	All	10.11950012	500	0.002020667	Predicted	Adams County PA	October	2	TRUE	11
10	62	Adams County	2021	11	11/1/2021	All	6.369047979	500	0.001271775	Predicted	Adams County PA	November	2	TRUE	11
11	62	Adams County	2021	12	12/1/2021	All	3.629526048	500	0.000724746	Predicted	Adams County PA	December	2	TRUE	11

id	naics_code	county	year	month	date	severity	claim_number	emp_count	injury_rate	type	county_formatted	month_str	naics_level	commonwealth	task_id
27	62	Adams County	2020	9	9/1/2020	All	8	4959	0.001613228	Actual	Adams County PA	September	2	TRUE	11
28	62	Adams County	2020	10	10/1/2020	High	1	5069	0.000197278	Actual	Adams County PA	October	2	TRUE	11
29	62	Adams County	2020	10	10/1/2020	Low	7	5069	0.001380943	Actual	Adams County PA	October	2	TRUE	11
30	62	Adams County	2020	10	10/1/2020	Medium	3	5069	0.000591833	Actual	Adams County PA	October	2	TRUE	11
31	62	Adams County	2020	10	10/1/2020	All	11	5069	0.002170053	Actual	Adams County PA	October	2	TRUE	11
32	62	Adams County	2020	11	11/1/2020	High	1	5030	0.000198807	Actual	Adams County PA	November	2	TRUE	11
33	62	Adams County	2020	11	11/1/2020	Low	5	5030	0.000994036	Actual	Adams County PA	November	2	TRUE	11
34	62	Adams County	2020	11	11/1/2020	Medium	4	5030	0.000795229	Actual	Adams County PA	November	2	TRUE	11
35	62	Adams County	2020	11	11/1/2020	All	10	5030	0.001988072	Actual	Adams County PA	November	2	TRUE	11
36	62	Adams County	2020	12	12/1/2020	High	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
37	62	Adams County	2020	12	12/1/2020	Low	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	11
38	62	Adams County	2020	12	12/1/2020	Medium	3	5008	0.000599042	Actual	Adams County PA	December	2	TRUE	11
39	62	Adams County	2020	12	12/1/2020	All	5	5008	0.000998403	Actual	Adams County PA	December	2	TRUE	11
0	62	Adams County	2021	1	1/1/2021	All	4.880162184	5008	0.000974473	Predicted	Adams County PA	January	2	TRUE	11
1	62	Adams County	2021	2	2/1/2021	All	12.16736266	5008	0.002429585	Predicted	Adams County PA	February	2	TRUE	11
2	62	Adams County	2021	3	3/1/2021	All	1.814446778	5008	0.00036231	Predicted	Adams County PA	March	2	TRUE	11
3	62	Adams County	2021	4	4/1/2021	All	0.1705288	5008	3.41E-05	Predicted	Adams County PA	April	2	TRUE	11
4	62	Adams County	2021	5	5/1/2021	All	6.290760082	5008	0.001256142	Predicted	Adams County PA	May	2	TRUE	11
5	62	Adams County	2021	6	6/1/2021	All	7.815840807	5008	0.001560671	Predicted	Adams County PA	June	2	TRUE	11
6	62	Adams County	2021	7	7/1/2021	All	13.47827404	5008	0.002691349	Predicted	Adams County PA	July	2	TRUE	11
7	62	Adams County	2021	8	8/1/2021	All	9.678400805	5008	0.001932588	Predicted	Adams County PA	August	2	TRUE	11
8	62	Adams County	2021	9	9/1/2021	All	3.870799189	5008	0.000772923	Predicted	Adams County PA	September	2	TRUE	11
9	62	Adams County	2021	10	10/1/2021	All	10.11950012	5008	0.002020667	Predicted	Adams County PA	October	2	TRUE	11
10	62	Adams County	2021	11	11/1/2021	All	6.369047979	5008	0.001271775	Predicted	Adams County PA	November	2	TRUE	11
11	62	Adams County	2021	12	12/1/2021	All	3.629526048	5008	0.000724746	Predicted	Adams County PA	December	2	TRUE	11

Monthly predicted values split by county and industry code

Desired solution

Current

Injury rate predicted on **county** and **industry** level

Desired

Update **severity** with **medical costs**.
Predict injury rates by severity.

Progress on deliverables

Explored the data ✓

Set up the application ✓

Ran predictions ✓

How will we **measure success**

Measuring success of efforts

Analytical tools

Prediction tools (like this!)

Expert knowledge

Decide where to allocate resource for safety trainings

Measuring success of efforts

t-test = Statistical test to compare the difference of a **metric** between **two groups**

injury rate

Source: Penn State (2022)

Measuring success of efforts

Population of workers

Measuring success of efforts

Workers who have <u>not</u> received training

Workers who have received training

Measuring success of efforts

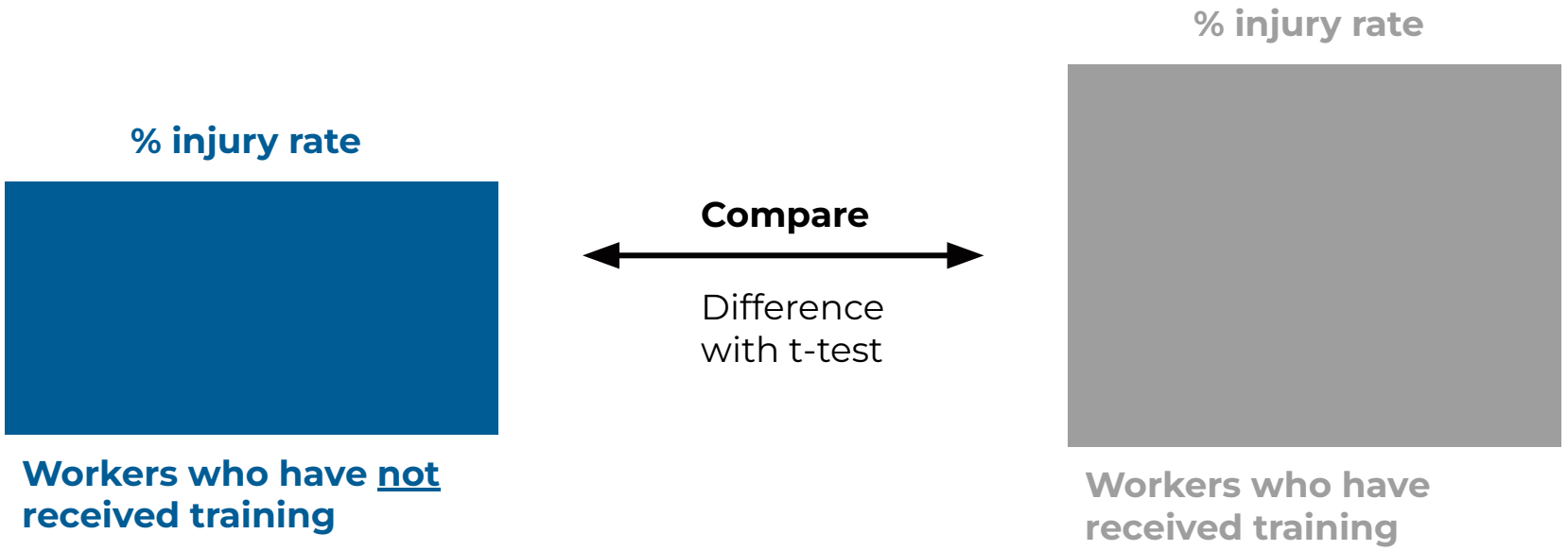
% injury rate

Workers who have not received training

% injury rate

Workers who have received training

Measuring success of efforts



Source: Penn State (2022)

Measuring success of efforts

% injury rate

Workers who have not received training

Uncover if allocating trainings are significantly impacting injury rate

% injury rate

Workers who have received training

Source: Penn State (2022)

These are long term measures that can’t be done in the next 3 months.

But, we’re excited to help move in the right direction.

Where are we going **next**

Next steps

Create medical costs upload functionality

Re-define and implement severity using medical costs

Look deeper into extending predictions by severity

Continue analysis to highlight key areas of action (high cost, high injury rates)

Develop recommendations for future iterations

Our main deliverables

Upload medical costs to database through the web application

Analyze injury rates by severity, updated with medical costs

Surface prediction results in Excel

Visualize top 5 counties/industries injury rate and medical costs in Power BI

Sources

Carnegie Mellon University. (2022). *Wordmarks, Lettermark, Unitmarks*. [Image]. https://www.cmu.edu/brand/brand-guidelines/visual-identity/logos.html

Corporate Finance Institute. (2022). *Boston Consulting Group (BCG) Matrix*. https://corporatefinanceinstitute.com/resources/knowledge/strategy/boston-consulting-group-bcg-matrix/

Pennsylvania Department of Labor & Industry. (2021). *Quarterly Census of Employment and Wages Data (QCEW), 2014-2021*.

Pennsylvania Department of Labor & Industry. (2021). *Workers' Compensation Automation and Integration System Data (WCAIS), 2014-2021*.

Pennsylvania Department of Labor & Industry. (2019). *Pennsylvania Compensation Rating Bureau (PCRB), 2014-2019*.

Sources

Pennsylvania Department of Labor & Industry. (2022). *Logo.* [Image]. https://www.dli.pa.gov/Style%20Library/Agency/img/logos/logo_mobile.png

Penn State. (2022). *Comparing two proportions.* https://online.stat.psu.edu/stat415/lesson/9/9.4

Facebook Github. (2022). *Prophet Quick Start*. https://facebook.github.io/prophet/docs/quick_start.html#python-api

Statsmodels. (2022). *statsmodels.tsa.statespace.sarimax.SARIMAX.* https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html

Slidesgo. (2022). *Slides Template.* https://slidesgo.com/

The Noun Project. (2022). https://thenounproject.com/