



Weekly Capstone Update Meeting

Nov 17 2022

PA Dept. of Labor & Industry + CMU

Goals for today

- Discuss completed this past week
- Discuss next few weeks

Completed this past week

November/December

23 Sprint 4

30 Sprint 5

6 Sprint 6

13 Sprint 7

20 Finalize Build & Break

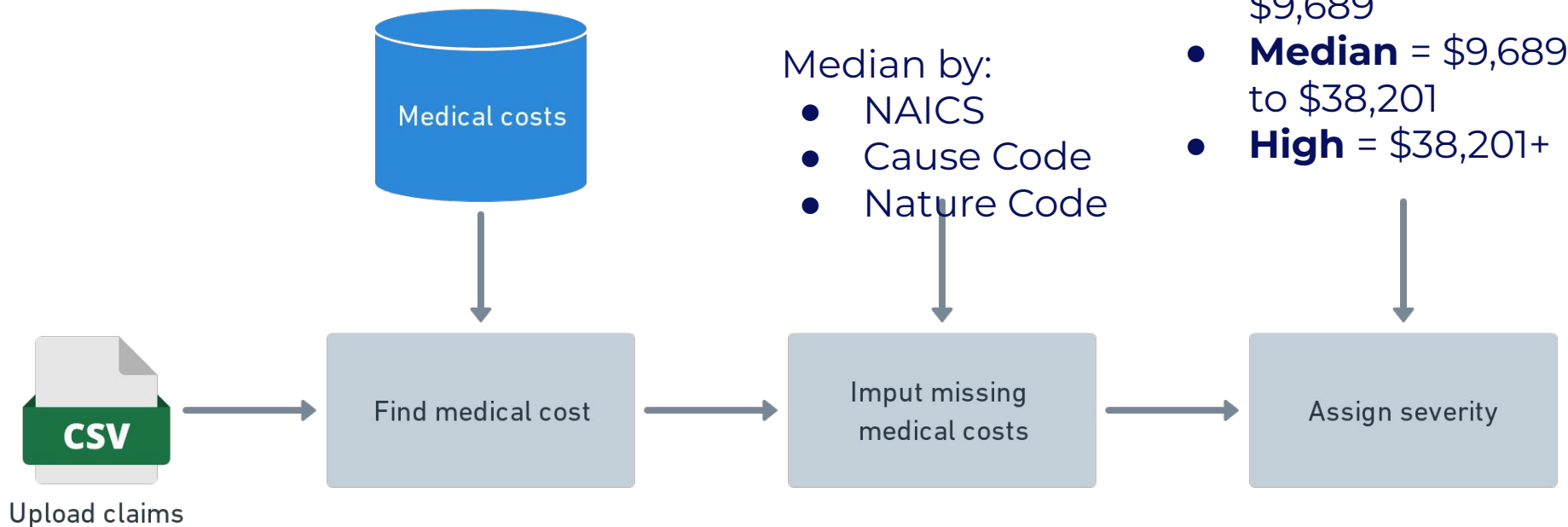


- Assigning severity
 - Implemented missing medical costs
 - Implemented new severity assignment using medical costs
- Work to implement exporting predictions
- Updated PowerBI Report



Assigning severity

New upload flow



Imputing with the median costs

	INJURY_CAUSE_CD	INJURY_NATURE_CD	Median Amount Paid
0	1	1	260.385
1	1	4	418.710
2	1	7	477.780
3	1	10	190.875
4	1	25	298.170
5	1	28	947.980
6	1	30	213.395
7	1	34	16891.160
8	1	36	760.945
9	1	37	286.265
10	1	40	271.070
11	1	42	355.700
12	1	43	676.840
13	1	52	127.615
14	1	54	1138.005
15	1	58	590.470
16	1	59	262.380
17	1	65	392.320
18	1	66	280.850

Cause = Chemicals
Nature = Fracture

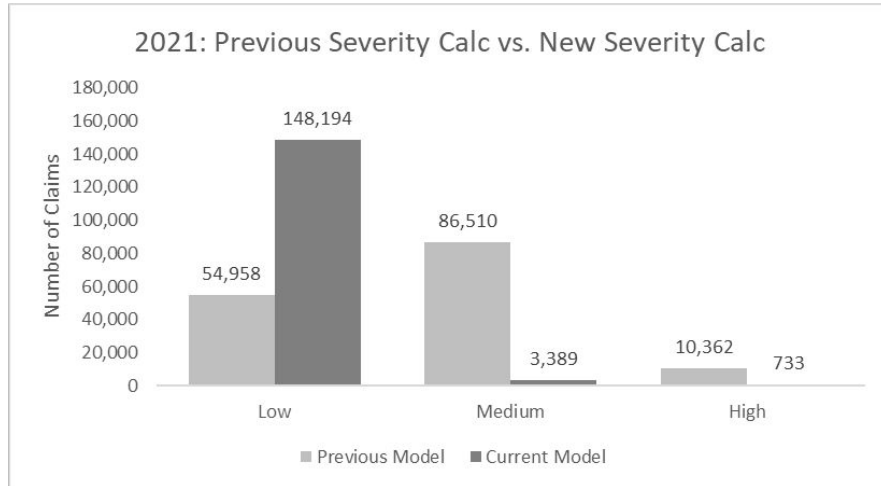
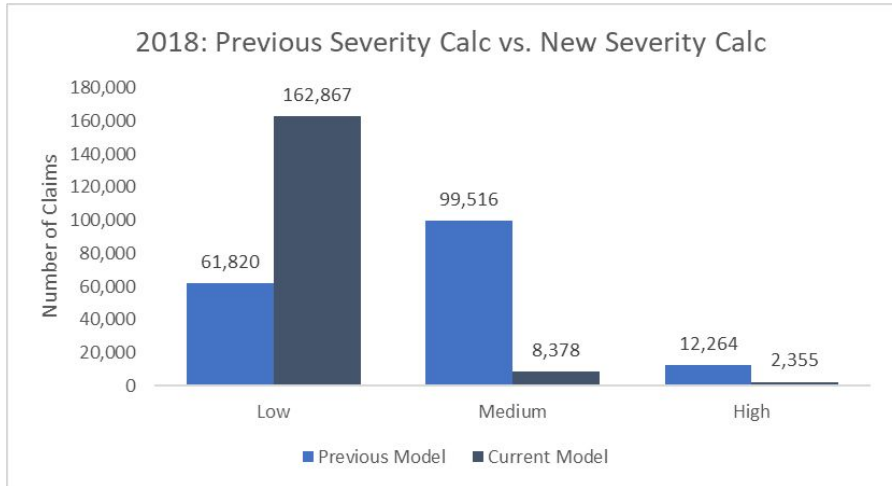
Missing value = \$947.98

Before vs. after

For both sample years:

The new severity calculation pushes claims away from “high” severity based on medical cost information

Will enable team to more effectively target “high” severity NAICS or regions



Select to predict injury rates by severity

PA Labor Injury Prediction

Preprocess

Prediction

Stats

Neural Model

Report

Help

Admin

Sign out

Run New Prediction Task

Type

All Industries

Model

Prophet

Start Date

Counties

Nothing selected

End Date

Severity

All

Note

NAICS Level

Submit

Show 10 entries

Search:

ID	Type	NAICS Level	Model	County Filter	NAICS Filter	Date Range	Process Time	Status	Log	Evaluation
83	Commonwealth	2	SARIMAX	Adams County;Allegheny County;Armstrong County;Beaver	11;21;22;23;42;51;52;53;54;55;56;61;62;71;72;81;92	2014-01-01 2022-01-01	2022-11-13 15:50:36	Successful	Show Log X	Evaluation
82	Commonwealth	2	SARIMAX	Allegheny County	62	2014-01-01 2022-01-01	2022-11-13 15:44:20	Successful	Show Log X	Evaluation



Exporting prediction results

Export prediction results

- Researched about the CSS syntax
- Added a new column for export button in prediction page
- Tested with the existing file directory
- Continuing next week



Updating PowerBI for easy analysis

Analyzing predicted injury rates



Predicted Injury Rate Analysis

Task ID
23

NAICS Level
All

NAICS Code
All

County
All

Severity
All

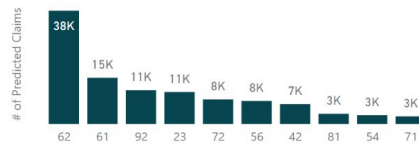
Previous Year Summary Stats (Actual)

2.48% 93,188 3,752,762
Injury Rate # of Claims # of Employees

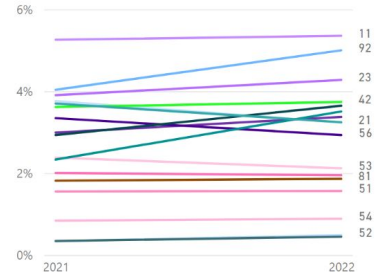
Next Year Summary Stats (Predicted)

2.86% 113,914 3,980,424
Injury Rate # of Claims # of Employees

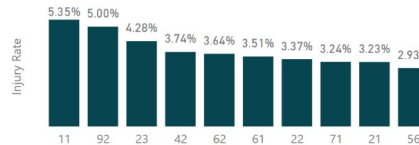
Top 10 NAICS Codes by Number of Claims (Predicted)



YoY Injury Rate Change by NAICS Code (Predicted & Actual)



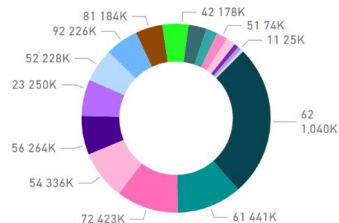
Top 10 NAICS Codes by Injury Rate (Predicted)



Heatmap of Injury Rate by County and NAICS (Predicted)

county	11	21	22	23	42	51	52	53	54	55	56	61	62	71	72	81	92
Montour County	4.53%			2.22%	1.56%		2.96%		1.05%				15.42%		1.43%	3.24%	4.03%
Perry County	0.97%			8.60%	4.30%	3.13%			9.58%	650.00%	0.88%	6.09%	11.93%		0.21%	3.09%	12.85%
Snyder County	6.33%			8.70%	2.10%		0.47%	3.48%	17.63%		1.92%		14.59%		1.87%	1.48%	2.50%
Forest County				22.39%									8.79%		0.90%	9.09%	4.79%
Cameron County				2.70%	9.09%		5.88%						17.37%		0.91%	1.64%	0.49%
Jefferson County	14.01%	0.87%	1.56%	11.88%	0.73%	2.65%		2.27%	0.90%	8.55%	0.87%	2.55%	9.40%	3.08%	0.53%	4.21%	2.65%
Total	5.35%	3.23%	3.37%	4.28%	3.74%	1.56%	0.48%	2.12%	0.88%	0.44%	2.93%	3.51%	3.64%	3.24%	1.95%	1.86%	5.00%

Employee Count By NAICS (Predicted) (NAICS, # Emps)



Problematic Data Points



Predicted Injury Rate Analysis

Task ID
23

NAICS Level
All

NAICS Code
All

County
All

Severity
All

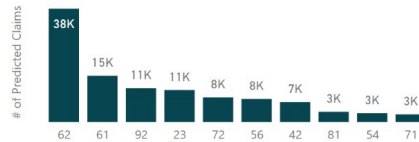
Previous Year Summary Stats (Actual)

2.48% 93,188 3,752,762
Injury Rate # of Claims # of Employees

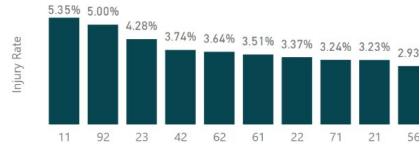
Next Year Summary Stats (Predicted)

2.86% 113,914 3,980,424
Injury Rate # of Claims # of Employees

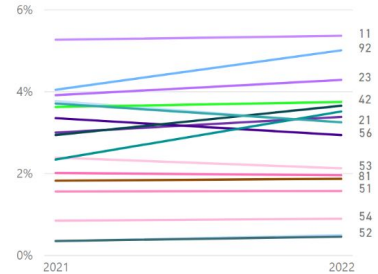
Top 10 NAICS Codes by Number of Claims (Predicted)



Top 10 NAICS Codes by Injury Rate (Predicted)



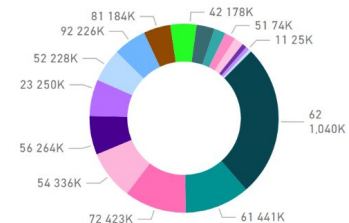
YoY Injury Rate Change by NAICS Code (Predicted & Actual)



Heatmap of Injury Rate by County and NAICS (Predicted)

county	11	21	22	23	42	51	52	53	54	55	56	61	62	71	72	81	92
Montour County	4.53%			2.22%	1.56%		2.96%		1.05%				15.42%		1.43%	3.24%	4.03%
Perry County	0.97%			8.60%	4.30%	3.13%			9.58%	650.00%	0.88%	6.09%	11.93%		0.21%	3.09%	12.85%
Snyder County	6.33%			8.70%	2.10%		0.47%	3.48%	17.63%		1.92%		14.59%		1.87%	1.48%	2.50%
Forest County				22.39%											8.79%	9.09%	4.79%
Cameron County				2.70%	9.09%		5.88%						17.37%		0.91%	1.64%	0.49%
Jefferson County	14.01%	0.87%	1.56%	11.88%	0.73%	2.65%		2.27%	0.90%	8.55%	0.87%	2.55%	9.40%	3.08%	0.53%	4.21%	2.65%
Total	5.35%	3.23%	3.37%	4.28%	3.74%	1.56%	0.48%	2.12%	0.88%	0.44%	2.93%	3.51%	3.64%	3.24%	1.95%	1.86%	5.00%

Employee Count By NAICS (Predicted) (NAICS, # Emps)



Problematic Data Points

Perry County, at year end, had 2 employees working in NAICS 55 according to QCEW Data

periodyea	perio	area	areaname	indcod	prelir	suppres	estab	avgemp	mnth1emp	mnth2emp	mnth3emp
2021	01	000099	Perry County	55	0	0		3	3	2	5
2021	02	000099	Perry County	55	0	0		3	3	2	5
2021	03	000099	Perry County	55	0	0		3	3	4	2
2021	04	000099	Perry County	55	0	0		3	2	2	2

Problematic Data Points

severity	All	
county	Perry County	
naics_code	55	
Column Labels		
2022		
Row Labels	Sum of emp_count	Sum of claim_number
Predicted	24	13
1	2	1
2	2	1
3	2	1
4	2	1
5	2	1
6	2	1
7	2	1
8	2	1
9	2	1
10	2	1
11	2	1
12	2	2

**Injury Rate =
Claims / # Employees**

In Perry County, NAICS 55:

- 2 Employees
- 13 Predicted Claims

Therefore:


- Injury Rate = 13 claims / 2 Employees
- Injury Rate = **650%**

Problematic Data Points

**Model Output Must be Evaluated
with Business Domain Knowledge**

Discussion

- Does it make sense to limit heatmap to where there are > 1,000 employees?
- What is your feedback on the data visualization for predicted values
 - Keep in mind that the fields available for predictions are:
 - NAICS Code
 - NAICS Level
 - County
 - Year / Month
 - # of Claims
 - # of Employees

A light blue abstract shape, resembling a stylized drop or a leaf, is positioned on the right side of a white background. The shape has a smooth, wavy edge on its left side, which curves towards the center of the image. The text is placed within this blue area.

**A look at the next
few weeks**

Nov. 20

Thanksgiving - No meeting

Nov. 27

Onboarding documentation

Dec. 4

Prepare project report & presentation

Dec. 11

Pass off & presentation

Final presentation: Wednesday, Dec. 14 @ 1:40pm

Next week

- Rest - Thanksgiving!
- Merge imputation and severity assignment into repository
- Update Power BI report
- Start onboarding documentation
- Update requirements