



# Weekly Capstone Update Meeting

Oct 3 2022

PA Dept. of Labor & Industry + CMU

# Goals for today

Propose four main project deliverables & get feedback

Discuss work completed this past week

Discuss timeline for next weeks

# Done this past week

October

**2** Sprint 2



**9** Midterm, Thurs. 10/13 @ 1:40pm ([Zoom](#))

**16** Break, no meeting

**23** Sprint 3

**30**

- Developed outline for project deliverables
- Ran predictions locally on Jupyter notebook – wahoo!
- Conducted exploratory data analysis on claim, employee, and medical cost datasets
- Continued research similar solutions around the US
- Updated repository with past team documents and latest requirements (in review)
- Prepared for midterm presentation

# Questions

What is the status of unblocking statisticians from viewing results in Power BI?

What tools do the statisticians use?

How do you measure success? Any target metric?



**Proposed  
project  
deliverables**

# Project deliverables

Upload medical costs to database through the interface

Analyze injury rates by severity – updated with medical costs

Run predictions outside of the web application

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

# Project deliverables

Upload medical costs to database through the interface

Analyze injury rates by severity – updated with medical costs


Run predictions outside of the web application

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

Injury Prediction

127.0.0.1:8000

SIOHealthConentHandshakeCanvasISLCCampus GroupsInternshipWorkdayRoche House Herita...TartanConnectDistributed Systems...DS TAsDS TA QueueCanvaBoxShorthandFlourishTrellobburns591/Capston...

PA Labor Injury Prediction

PreprocessPredictionStatsNeural ModelReportHelp

AdminSign out

Choose FileNo file chosen

All Industries

Commonwealth

Employment

Submit

Total Claims0

CW Claims0

All Industries Claims0

Total Employment0

Commonwealth

All Industries

Search:

Show LogX

Showing 1 to 1 of 1 entries

Previous1Next

2022-09-30 22:41:15

Showing 1 to 1 of 1 entries

Add medical costs option



# Project deliverables

Upload medical costs to database through the interface

Analyze injury rates by severity – updated with medical costs

Run predictions outside of the web application

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

## Prediction output in Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
		naics_code	county	year	month	date	severity	claim_number	emp_count	injury_rate	type	county_formatted	month_str	naics_level	commonwealth	task_id	
1	27	62	Adams County	2020	9	9/1/2020	All	8	4959	0.001613228	Actual	Adams County PA	September	2	TRUE	10	
2	28	62	Adams County	2020	10	10/1/2020	High	1	5069	0.000197278	Actual	Adams County PA	October	2	TRUE	10	
3	29	62	Adams County	2020	10	10/1/2020	Low	7	5069	0.001380943	Actual	Adams County PA	October	2	TRUE	10	
4	30	62	Adams County	2020	10	10/1/2020	Medium	3	5069	0.000591833	Actual	Adams County PA	October	2	TRUE	10	
5	31	62	Adams County	2020	10	10/1/2020	All	11	5069	0.002170053	Actual	Adams County PA	October	2	TRUE	10	
6	32	62	Adams County	2020	11	11/1/2020	High	1	5030	0.000198807	Actual	Adams County PA	November	2	TRUE	10	
7	33	62	Adams County	2020	11	11/1/2020	Low	5	5030	0.000994036	Actual	Adams County PA	November	2	TRUE	10	
8	34	62	Adams County	2020	11	11/1/2020	Medium	4	5030	0.000795229	Actual	Adams County PA	November	2	TRUE	10	
9	35	62	Adams County	2020	11	11/1/2020	All	10	5030	0.001988072	Actual	Adams County PA	November	2	TRUE	10	
10	36	62	Adams County	2020	12	12/1/2020	High	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	10	
11	37	62	Adams County	2020	12	12/1/2020	Low	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	10	
12	38	62	Adams County	2020	12	12/1/2020	Medium	3	5008	0.000599042	Actual	Adams County PA	December	2	TRUE	10	
13	39	62	Adams County	2020	12	12/1/2020	All	5	5008	0.000998403	Actual	Adams County PA	December	2	TRUE	10	
14	0	62	Adams County	2021	1	1/1/2021	All	8.851568932	5008	0.001767486	Predicted	Adams County PA	January	2	TRUE	10	
15	1	62	Adams County	2021	2	2/1/2021	All	12.74997932	5008	0.002545922	Predicted	Adams County PA	February	2	TRUE	10	
16	2	62	Adams County	2021	3	3/1/2021	All	3.075329879	5008	0.000614083	Predicted	Adams County PA	March	2	TRUE	10	
17	3	62	Adams County	2021	4	4/1/2021	All	5.53084577	5008	0.001104402	Predicted	Adams County PA	April	2	TRUE	10	
18	4	62	Adams County	2021	5	5/1/2021	All	9.663631519	5008	0.001929639	Predicted	Adams County PA	May	2	TRUE	10	
19	5	62	Adams County	2021	6	6/1/2021	All	10.51307236	5008	0.002099256	Predicted	Adams County PA	June	2	TRUE	10	
20	6	62	Adams County	2021	7	7/1/2021	All	15.52448577	5008	0.003099937	Predicted	Adams County PA	July	2	TRUE	10	
21	7	62	Adams County	2021	8	8/1/2021	All	13.25968038	5008	0.0026477	Predicted	Adams County PA	August	2	TRUE	10	
22	8	62	Adams County	2021	9	9/1/2021	All	9.227483115	5008	0.001842549	Predicted	Adams County PA	September	2	TRUE	10	
23	9	62	Adams County	2021	10	10/1/2021	All	12.01606167	5008	0.002399373	Predicted	Adams County PA	October	2	TRUE	10	
24	10	62	Adams County	2021	11	11/1/2021	All	11.10469734	5008	0.002217392	Predicted	Adams County PA	November	2	TRUE	10	
25	11	62	Adams County	2021	12	12/1/2021	All	6.148434919	5008	0.001227723	Predicted	Adams County PA	December	2	TRUE	10	

Adams County, NAICS = 62, 2020

## Historic values

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	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	10
28	62	Adams County	2020	10	10/1/2020	High	1	5069	0.000197278	Actual	Adams County PA	October	2	TRUE	10
29	62	Adams County	2020	10	10/1/2020	Low	7	5069	0.001380943	Actual	Adams County PA	October	2	TRUE	10
30	62	Adams County	2020	10	10/1/2020	Medium	3	5069	0.000591833	Actual	Adams County PA	October	2	TRUE	10
31	62	Adams County	2020	10	10/1/2020	All	11	5069	0.002170053	Actual	Adams County PA	October	2	TRUE	10
32	62	Adams County	2020	11	11/1/2020	High	1	5030	0.000198807	Actual	Adams County PA	November	2	TRUE	10
33	62	Adams County	2020	11	11/1/2020	Low	5	5030	0.000994036	Actual	Adams County PA	November	2	TRUE	10
34	62	Adams County	2020	11	11/1/2020	Medium	4	5030	0.000795229	Actual	Adams County PA	November	2	TRUE	10
35	62	Adams County	2020	11	11/1/2020	All	10	5030	0.001988072	Actual	Adams County PA	November	2	TRUE	10
36	62	Adams County	2020	12	12/1/2020	High	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	10
37	62	Adams County	2020	12	12/1/2020	Low	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	10
38	62	Adams County	2020	12	12/1/2020	Medium	3	5008	0.000599042	Actual	Adams County PA	December	2	TRUE	10
39	62	Adams County	2020	12	12/1/2020	All	5	5008	0.000998403	Actual	Adams County PA	December	2	TRUE	10
0	62	Adams County	2021	1	1/1/2021	All	8.851568932	5008	0.001767486	Predicted	Adams County PA	January	2	TRUE	10
1	62	Adams County	2021	2	2/1/2021	All	12.74997932	5008	0.002545922	Predicted	Adams County PA	February	2	TRUE	10
2	62	Adams County	2021	3	3/1/2021	All	3.075329879	5008	0.000614083	Predicted	Adams County PA	March	2	TRUE	10
3	62	Adams County	2021	4	4/1/2021	All	5.53084577	5008	0.001104402	Predicted	Adams County PA	April	2	TRUE	10
4	62	Adams County	2021	5	5/1/2021	All	9.663631519	5008	0.001929639	Predicted	Adams County PA	May	2	TRUE	10
5	62	Adams County	2021	6	6/1/2021	All	10.51307236	5008	0.002099256	Predicted	Adams County PA	June	2	TRUE	10
6	62	Adams County	2021	7	7/1/2021	All	15.52448577	5008	0.003099937	Predicted	Adams County PA	July	2	TRUE	10
7	62	Adams County	2021	8	8/1/2021	All	13.25968038	5008	0.0026477	Predicted	Adams County PA	August	2	TRUE	10
8	62	Adams County	2021	9	9/1/2021	All	9.227483115	5008	0.001842549	Predicted	Adams County PA	September	2	TRUE	10
9	62	Adams County	2021	10	10/1/2021	All	12.01606167	5008	0.002399373	Predicted	Adams County PA	October	2	TRUE	10
10	62	Adams County	2021	11	11/1/2021	All	11.10469734	5008	0.002217392	Predicted	Adams County PA	November	2	TRUE	10
11	62	Adams County	2021	12	12/1/2021	All	6.148434919	5008	0.001227723	Predicted	Adams County PA	December	2	TRUE	10



## Historic values

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	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	10
28	62	Adams County	2020	10	10/1/2020	High	1	5069	0.000197278	Actual	Adams County PA	October	2	TRUE	10
29	62	Adams County	2020	10	10/1/2020	Low	7	5069	0.001380943	Actual	Adams County PA	October	2	TRUE	10
30	62	Adams County	2020	10	10/1/2020	Medium	3	5069	0.000591833	Actual	Adams County PA	October	2	TRUE	10
31	62	Adams County	2020	10	10/1/2020	All	11	5069	0.002170053	Actual	Adams County PA	October	2	TRUE	10
32	62	Adams County	2020	11	11/1/2020	High	1	5030	0.000198807	Actual	Adams County PA	November	2	TRUE	10
33	62	Adams County	2020	11	11/1/2020	Low	5	5030	0.000994036	Actual	Adams County PA	November	2	TRUE	10
34	62	Adams County	2020	11	11/1/2020	Medium	4	5030	0.000795229	Actual	Adams County PA	November	2	TRUE	10
35	62	Adams County	2020	11	11/1/2020	All	10	5030	0.001988072	Actual	Adams County PA	November	2	TRUE	10
36	62	Adams County	2020	12	12/1/2020	High	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	10
37	62	Adams County	2020	12	12/1/2020	Low	1	5008	0.000199681	Actual	Adams County PA	December	2	TRUE	10
38	62	Adams County	2020	12	12/1/2020	Medium	3	5008	0.000599042	Actual	Adams County PA	December	2	TRUE	10
39	62	Adams County	2020	12	12/1/2020	All	5	5008	0.000998403	Actual	Adams County PA	December	2	TRUE	10
0	62	Adams County	2021	1	1/1/2021	All	8.851568922	5008	0.001767486	Predicted	Adams County PA	January	2	TRUE	10
1	62	Adams County	2021	2	2/1/2021	All	12.74997932	5008	0.002545922	Predicted	Adams County PA	February	2	TRUE	10
2	62	Adams County	2021	3	3/1/2021	All	3.075329879	5008	0.000614083	Predicted	Adams County PA	March	2	TRUE	10
3	62	Adams County	2021	4	4/1/2021	All	5.53084577	5008	0.001104402	Predicted	Adams County PA	April	2	TRUE	10
4	62	Adams County	2021	5	5/1/2021	All	9.663631519	5008	0.001929639	Predicted	Adams County PA	May	2	TRUE	10
5	62	Adams County	2021	6	6/1/2021	All	10.51307236	5008	0.002099256	Predicted	Adams County PA	June	2	TRUE	10
6	62	Adams County	2021	7	7/1/2021	All	15.52448577	5008	0.003099937	Predicted	Adams County PA	July	2	TRUE	10
7	62	Adams County	2021	8	8/1/2021	All	13.25968038	5008	0.0026477	Predicted	Adams County PA	August	2	TRUE	10
8	62	Adams County	2021	9	9/1/2021	All	9.227483115	5008	0.001842549	Predicted	Adams County PA	September	2	TRUE	10
9	62	Adams County	2021	10	10/1/2021	All	12.01606167	5008	0.002399373	Predicted	Adams County PA	October	2	TRUE	10
10	62	Adams County	2021	11	11/1/2021	All	11.10469734	5008	0.002217392	Predicted	Adams County PA	November	2	TRUE	10
11	62	Adams County	2021	12	12/1/2021	All	6.148434919	5008	0.001227723	Predicted	Adams County PA	December	2	TRUE	10

## Predicted values

## Severity

[illegible]

## Severity

	G	H	I	J	K
	severity	claim_number	emp_count	injury_rate	type
020	All	8	4959	0.001613228	Actual
020	High	1	5069	0.000197278	Actual
020	Low	7	5069	0.001380943	Actual
020	Medium	3	5069	0.000591833	Actual
020	All	11	5069	0.002170053	Actual
020	High	1	5030	0.000198807	Actual
020	Low	5	5030	0.000994036	Actual
020	Medium	4	5030	0.000795229	Actual
020	All	10	5030	0.001988072	Actual
020	High	1	5008	0.000199681	Actual
020	Low	1	5008	0.000199681	Actual
020	Medium	3	5008	0.000599042	Actual
020	All	5	5008	0.000998403	Actual
021	All	8.851568932	5008	0.001767486	Predicted
021	All	12.74997932	5008	0.002545922	Predicted
021	All	3.075329879	5008	0.000614083	Predicted
021	All	5.53084577	5008	0.001104402	Predicted
021	All	9.663631519	5008	0.001929639	Predicted
021	All	10.51307236	5008	0.002099256	Predicted
021	All	15.52448577	5008	0.003099937	Predicted
021	All	13.25968038	5008	0.0026477	Predicted
021	All	9.227483115	5008	0.001842549	Predicted
021	All	12.01606167	5008	0.002399373	Predicted
021	All	11.10469734	5008	0.002217392	Predicted
021	All	6.148434919	5008	0.001227723	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

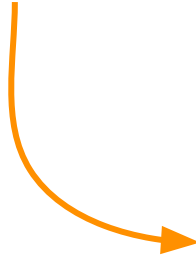
	G	H	I	J	K
	severity	claim_number	emp_count	injury_rate	type
020	All	8	4959	0.001613228	Actual
020	High	1	5069	0.000197278	Actual
020	Low	7	5069	0.001380943	Actual
020	Medium	3	5069	0.000591833	Actual
020	All	11	5069	0.002170053	Actual
020	High	1	5030	0.000198807	Actual
020	Low	5	5030	0.000994036	Actual
020	Medium	4	5030	0.000795229	Actual
020	All	10	5030	0.001988072	Actual
020	High	1	5008	0.000199681	Actual
020	Low	1	5008	0.000199681	Actual
020	Medium	3	5008	0.000599042	Actual
020	All	5	5008	0.000998403	Actual
021	All	8.851568932	5008	0.001767486	Predicted
021	All	12.74997932	5008	0.002545922	Predicted
021	All	3.075329879	5008	0.000614083	Predicted
021	All	5.53084577	5008	0.001104402	Predicted
021	All	9.663631519	5008	0.001929639	Predicted
021	All	10.51307236	5008	0.002099256	Predicted
021	All	15.52448577	5008	0.003099937	Predicted
021	All	13.25968038	5008	0.0026477	Predicted
021	All	9.227483115	5008	0.001842549	Predicted
021	All	12.01606167	5008	0.002399373	Predicted
021	All	11.10469734	5008	0.002217392	Predicted
021	All	6.148434919	5008	0.001227723	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

	G	H	I	J	K
	severity	claim_number	emp_count	injury_rate	type
120	All	8	4959	0.001613228	Actual
120	High	1	5069	0.000197278	Actual
120	Low	7	5069	0.001380943	Actual
120	Medium	3	5069	0.000591833	Actual
120	All	11	5069	0.002170053	Actual
120	High	1	5030	0.000198807	Actual
120	Low	5	5030	0.000994036	Actual
120	Medium	4	5030	0.000795229	Actual
120	All	10	5030	0.001988072	Actual
120	High	1	5008	0.000199681	Actual
120	Low	1	5008	0.000199681	Actual
120	Medium	3	5008	0.000599042	Actual
120	All	5	5008	0.000998403	Actual
121	All	8.851568932	5008	0.001767486	Predicted
121	All	12.74997932	5008	0.002545922	Predicted
121	All	3.075329879	5008	0.000614083	Predicted
121	All	5.5308457	5008	0.001104402	Predicted
121	All	1.66363	5008	0.001929639	Predicted
121	All	10.1	5008	0.002099256	Predicted
121	All	1.524485	5008	0.003099937	Predicted
121	All	13.25968038	5008	0.0026477	Predicted
121	All	9.227483115	5008	0.001842549	Predicted
121	All	12.01606167	5008	0.002399373	Predicted
121	All	11.10469734	5008	0.0017392	Predicted
121	All	6.148434919	5008	0.001227	Predicted

## Note –

Currently, severity only splits injury rates on **historic values**.

It **does not** contribute to predictions.

Predicted severity rates is something that future teams could explore.



## Severity

	G	H	I	J	K
	severity	claim_number	emp_count	injury_rate	type
120	All	8	4959	0.001613228	Actual
120	High	1	5069	0.000197278	Actual
120	Low	7	5069	0.001380943	Actual
120	Medium	3	5069	0.000591833	Actual
120	All	11	5069	0.002170053	Actual
120	High	1	5030	0.000198807	Actual
120	Low	5	5030	0.000994036	Actual
120	Medium	4	5030	0.000795229	Actual
120	All	10	5030	0.001988072	Actual
120	High	1	5008	0.000199681	Actual
120	Low	1	5008	0.000199681	Actual
120	Medium	3	5008	0.000599042	Actual
120	All	5	5008	0.000998403	Actual
121	All	8.851568932	5008	0.001767486	Predicted
121	All	12.74997932	5008	0.002545922	Predicted
121	All	3.075329879	5008	0.000614083	Predicted
121	All	5.53084577	5008	0.001104402	Predicted
121	All	9.663631519	5008	0.001929639	Predicted
121	All	10.51307236	5008	0.002099256	Predicted
121	All	15.52448577	5008	0.003099937	Predicted
121	All	13.25968038	5008	0.0026477	Predicted
121	All	9.227483115	5008	0.001842549	Predicted
121	All	12.01606167	5008	0.002399373	Predicted
121	All	11.10469734	5008	0.002217392	Predicted
121	All	6.148434919	5008	0.001227723	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

– Is this aligned with the goal?

# Project deliverables

Upload medical costs to database through the interface

Analyze injury rates by severity – updated with medical costs

Run predictions outside of the web application

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

Alternative option to web application

Specify parameters (county, NAICS)

Get results (Excel, charts)

Used for development, future teams

## Predict injury rates

This notebook manually predicts injury rates using the SARIMAX and Prophet models outside of the web application UI.

```
In [1]: from django.db.models import Model
```

```
In [2]: import pandas as pd
import utilities.etl_controller as etl_controller
import utilities.predict_controller as predict_controller
import prediction_module.models as predict_models
```

### Define parameters

```
In [276.. # Define the parameters to filter the data for your prediction run
```

### Import data

```
In [3]: pd.set_option('display.max_columns', None)
```

```
In [4]: claim_df = etl_controller.ETLController.read_file("../sample/2020 WCAIS data rev copy.xlsx")
```

```
In [5]: claim_df.head()
```

```
Out[5]:
```

	NAICS	CLAIM_NUMBER	CLAIM_ADMIN_CLAIM_NUMBER	FEIN_orig	EMPLR_NAME	EMPLR_PHY_PRIM_ADR	EMPLR_PHY_SECD_ADR	EMPLR_PHY_CITY	EMPLR_PHY_POSTAL_CD
0	622110	8432564	20525-PUH2020224911	132456789	M sr	123 Main St	.	Anytown	NaN
1	452990	8432619	201918710	132456789	lerc	123 Main St	.	Anytown	NaN
2	238212	8432663	10213442	132456789	GDL	123 Main St	.	Anytown	19606

Current draft

Alternative option to web application

Specify parameters (county, NAICS)

Get results (Excel, charts)

Used for development, future teams

→ Is this interesting?

→ Do you currently use notebook solutions?

→ Is this technically feasible?

## Predict injury rates

This notebook manually predicts injury rates using the SARIMAX and Prophet models outside of the web application UI.

```
In [1]: from django.db.models import Model
```

```
In [2]: import pandas as pd
import utilities.etl_controller as etl_controller
import utilities.predict_controller as predict_controller
import prediction_module.models as predict_models
```

### Define parameters

```
In [276... # Define the parameters to filter the data for your prediction run
```

### Import data

```
In [3]: pd.set_option('display.max_columns', None)
```

```
In [4]: claim_df = etl_controller.ETLController.read_file("../sample/2020 WCAIS data rev copy.xlsx")
```

```
In [5]: claim_df.head()
```

```
Out[5]:
```

	NAICS	CLAIM_NUMBER	CLAIM_ADMIN_CLAIM_NUMBER	FEIN_orig	EMPLR_NAME	EMPLR_PHY_PRIM_ADR	EMPLR_PHY_SECD_ADR	EMPLR_PHY_CITY	EMPLR_PHY_POSTAL_CD
0	622110	8432564	20525-PUH2020224911	132456789	M sr	123 Main St	.	Anytown	NaN
1	452990	8432619	201918710	132456789	lerc	123 Main St	.	Anytown	NaN
2	238212	8432663	10213442	132456789	GDL	123 Main St	.	Anytown	19606

Current draft

# Project deliverables

Upload medical costs to database through the interface

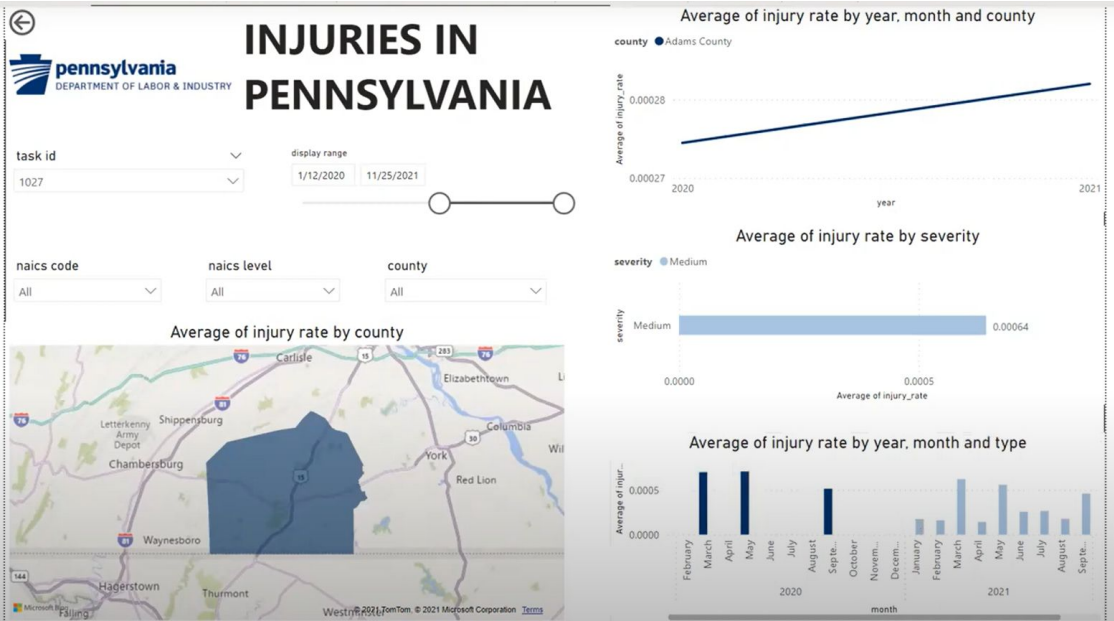
Analyze injury rates by severity – updated with medical costs

Run predictions outside of the web application

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

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

Add analysis of **medical costs**



# Project deliverables

Upload medical costs to database through the interface

Analyze injury rates by severity – updated with medical costs

Run predictions outside of the web application

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

# Project deliverables

Upload medical costs to database through the interface

Analyze injury rates by severity – updated with medical costs

Run predictions outside of the web application

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

**→ Does this align with your goals?**



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 wavy, organic edge on its left side, which points towards the center of the image. The text is placed within the white space to the left of this blue shape.

**More completed  
this week**

# Ran predictions locally

Ran prediction on small set

- Adams county
- NAICS = 62, Health Care and Social Assistance sector
- 2020

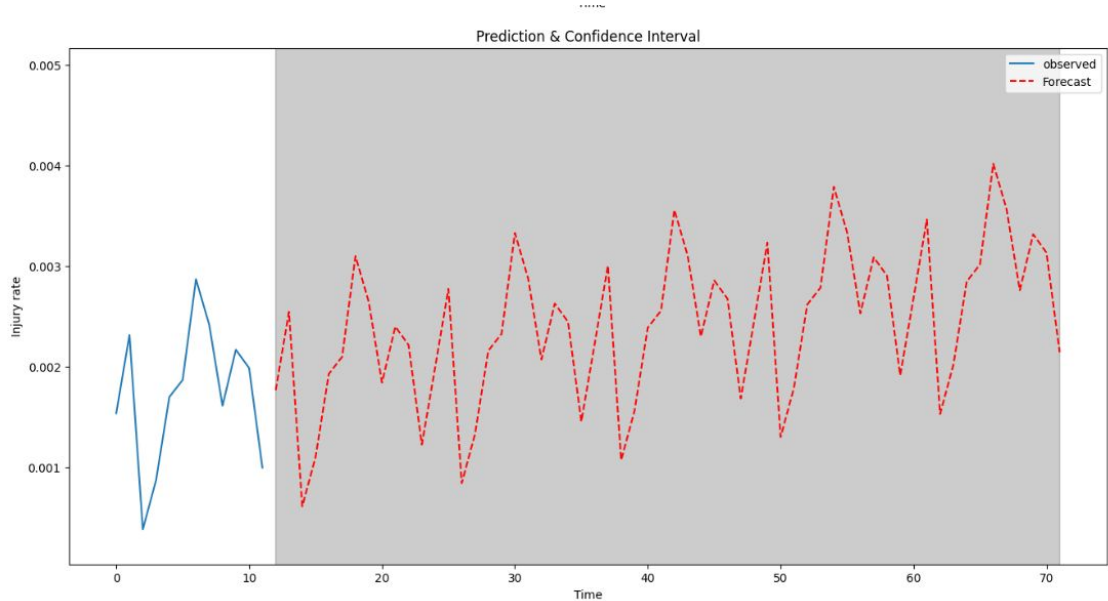
Predicted 2021 monthly values

Excel output (shown early)

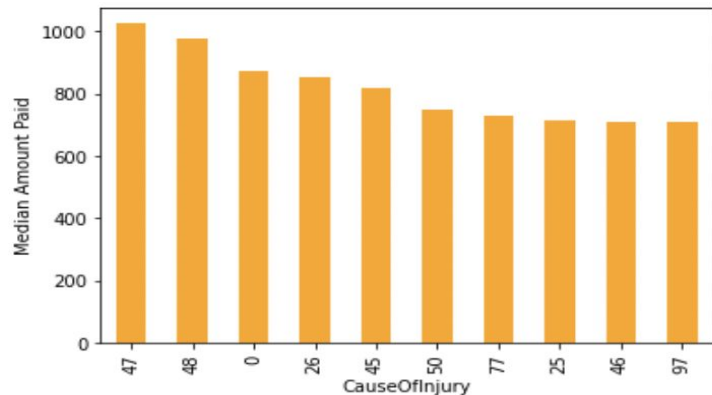
Displayed charts (not Power BI report)

Used Jupyter notebook, not web application

Will use for development



# Exploratory data analysis



Analyzed PCRB (medical costs) data

Calculated **top-10** injury causes, nature of injury and parts of body by count and median medical costs

Obtained a **42%** match rate on joining PCRB (medical costs) and WCAIS (claims) data

→ **2018 WCAIS corrupted – resend?**

Cause of injury:

47 = Crash of Airplane

48 = Vehicle Upset

56 = Lifting

60 = Strain by Injury

# Update repository with documentation

bburns591 / Capstone\_S2022 Private

<> Code Issues Pull requests 1 Actions Projects Security Insights

kmcmmanus-updat... 4 branches 0 tags Go to file Add file Code

This branch is 3 commits ahead of master. #1

kelmcm Add data files and update read me ✓ f6a553f yesterday 32 commits

.github/workflows	Version 11_30_2021 (#22)	10 months ago
.idea	feat:add timeseries model evaluation integration	5 months ago
Data	Add data files and update read me	yesterday
documentation	Add past team documentation	2 days ago
injury_predict_web	Merge pull request #4 from LansingC001/feat/nn	5 months ago
output	Merge branch 'feature/updated-db-model' of https://github.com/nnayd...	12 months ago
sample	feat: nerual network	5 months ago
.gitattributes	Initial commit	13 months ago
.gitignore	Version 11_30_2021 (#22)	10 months ago
.pre-commit-config.yaml	Version 11_30_2021 (#22)	10 months ago
FinalPowerBIReport.pbix	Version 11_30_2021 (#22)	10 months ago
README.md	Add data files and update read me	yesterday
collectstatic.bat	Version 11_30_2021 (#22)	10 months ago
install.bat	Version 11_30_2021 (#22)	10 months ago

Past team's documents

Our documents

Is it okay to upload full datasets?

# Next steps

## Next week

Conduct midterm presentation, Thursday, Oct. 13 @ 1:40pm ([Zoom Link](#))

Merge documentation to repository

## After break, Sprint 3

Create medical costs upload functionality

Re-define and implement severity using medical costs

Demo jupyter notebook to run predictions