Weekly Capstone Update Meeting

Oct 28 2022

PA Dept. of Labor & Industry + CMU

Goals for today

Discuss severity calculation

Discuss work completed this past week

Discuss timeline for next weeks

Done this past week

October / November

23	Sprint 4 
30	Sprint 5
6	Sprint 6
13	Sprint 7
20	Finalize Build & Break

- Explored methods to redefine severity calculation
- Created field on claims table in SQL Server to store medical costs
- Began working on export prediction button on web application
- Began working on imputation methods for claims data during model runs

Proposed project deliverables

Project deliverables (from midterm)

Upload medical costs to database through the interface

Analyze injury rates by severity – updated with medical costs

Surface prediction output Excel report

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

Project deliverables (from midterm)

Upload medical costs to database through the interface

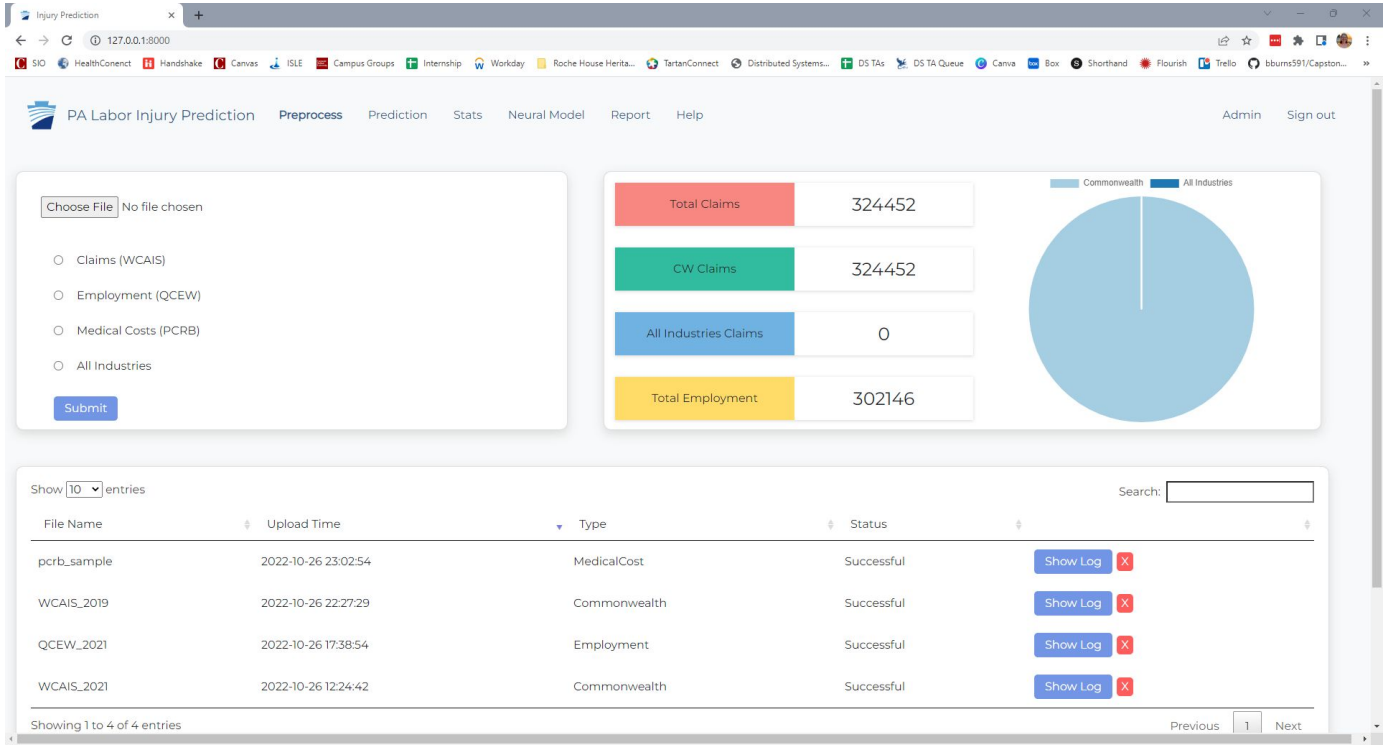
Analyze injury rates by severity – updated with medical costs

Surface prediction output Excel report

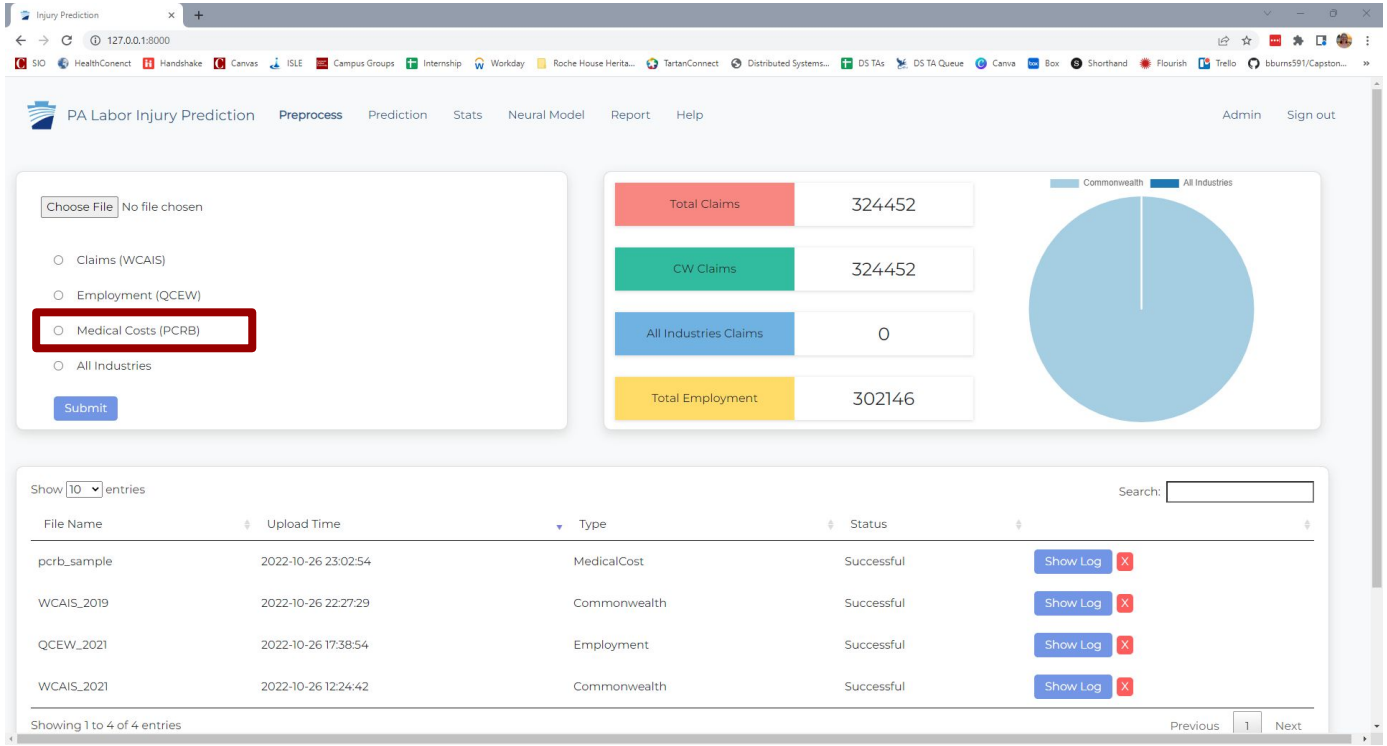
Visual top 5 counties/industries injury rate and medical costs in Power BI report

Uploading Medical Costs

Medical Cost Upload Functionality

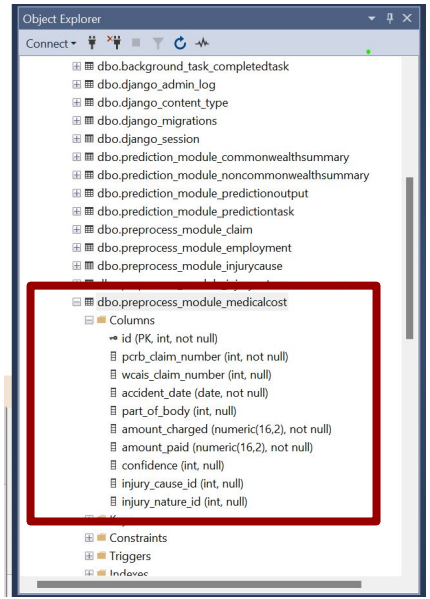


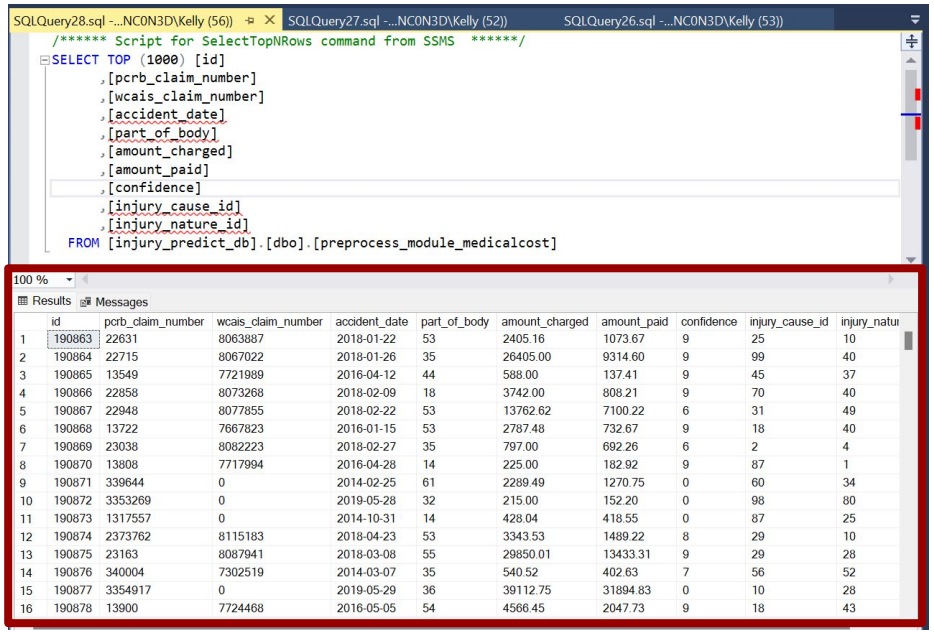
Medical Cost Upload Functionality



Medical Cost Upload Functionality

New SQL Server Tables are being fed data from the upload function





	id	pcrb_claim_number	wcais_claim_number	accident_date	part_of_body	amount_charged	amount_paid	confidence	injury_cause_id	injury_natur
1	190863	22631	8063887	2018-01-22	53	2405.16	1073.67	9	25	10
2	190864	22715	8067022	2018-01-26	35	26405.00	9314.60	9	99	40
3	190865	13549	7721989	2016-04-12	44	588.00	137.41	9	45	37
4	190866	22858	8073268	2018-02-09	18	3742.00	808.21	9	70	40
5	190867	22948	8077855	2018-02-22	53	13762.62	7100.22	6	31	49
6	190868	13722	7667823	2016-01-15	53	2787.48	732.67	9	18	40
7	190869	23038	8082223	2018-02-27	35	797.00	692.26	6	2	4
8	190870	13808	7717994	2016-04-28	14	225.00	182.92	9	87	1
9	190871	339644	0	2014-02-25	61	2289.49	1270.75	0	60	34
10	190872	3353269	0	2019-05-28	32	215.00	152.20	0	98	80
11	190873	1317557	0	2014-10-31	14	428.04	418.55	0	87	25
12	190874	2373762	8115183	2018-04-23	53	3343.53	1489.22	8	29	10
13	190875	23163	8087941	2018-03-08	55	29850.01	13433.31	9	29	28
14	190876	340004	7302519	2014-03-07	35	540.52	402.63	7	56	52
15	190877	3354917	0	2019-05-29	36	39112.75	31894.83	0	10	28
16	190878	13900	7724468	2016-05-05	54	4566.45	2047.73	9	18	43

Medical Cost Upload Functionality

Next Steps:

1. Merge our modified files into the main repository

2. Determine how to handle missingness on WCAIS Claims that can’t be linked to PCRB data

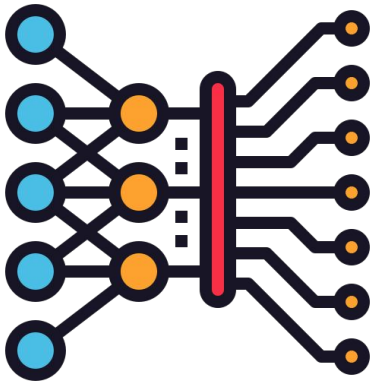
Imputing Missing Medical Cost Data

Currently determining best method to utilize

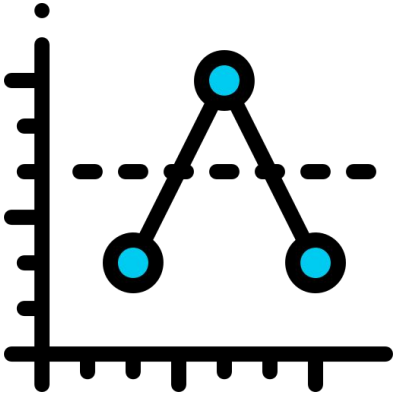
Imputing Missing Medical Cost Data

Currently determining best method to utilize

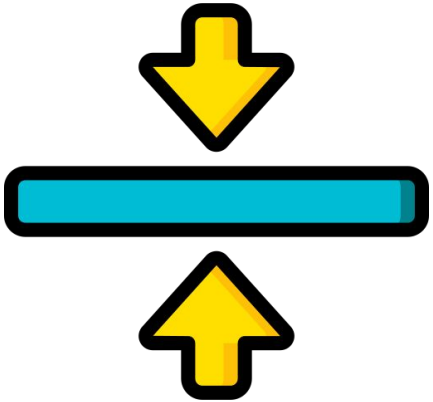
Model Based



Class Level Mean



Class Level Median



Severity Calculation

Severity Calculation - Current Method

	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 Calculation - Proposed Method 1

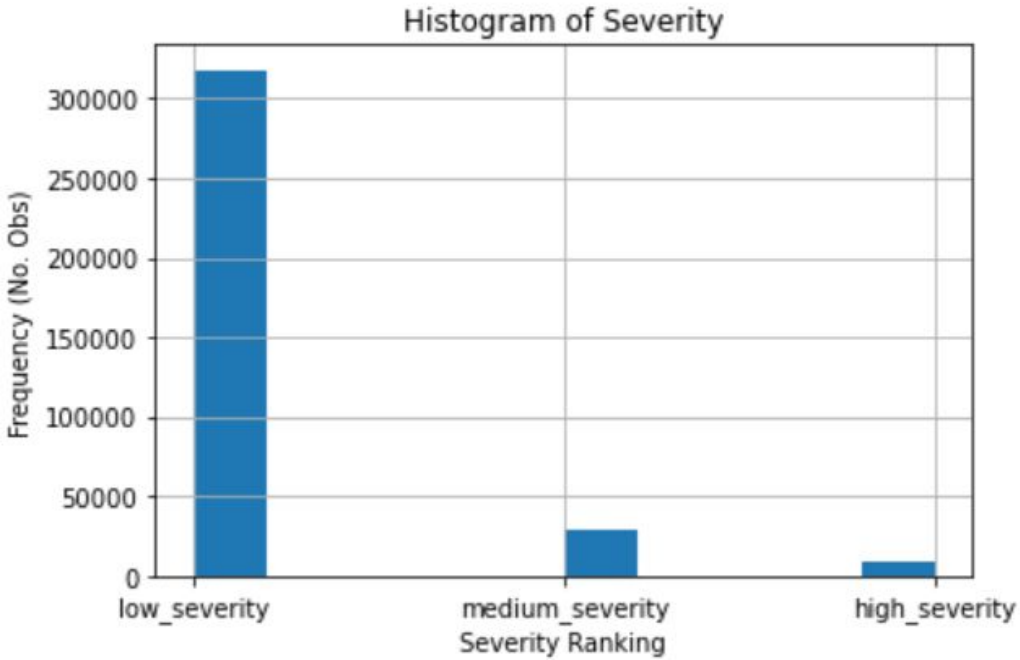
Minimum “Amount Paid” in PCRB dataset was 0

“High Severity” begins at $37K

“Medium Severity” bin begins at $10K

Maximum of dataset after outlier removal

Severity Calculation - Proposed Method 1



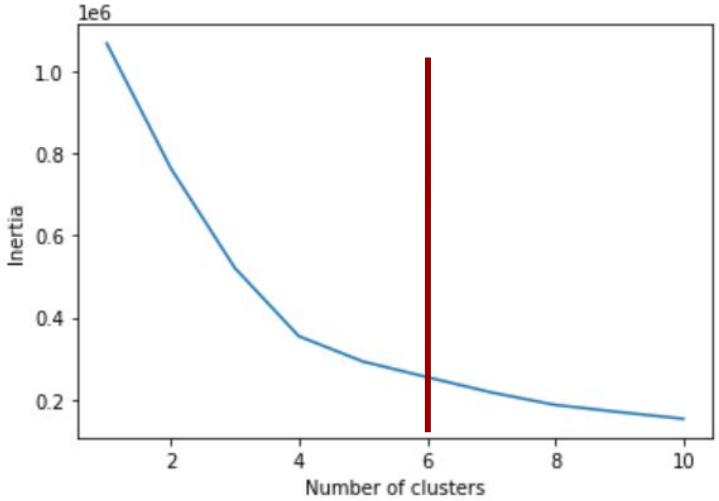
Compared to the current method, more observations will be bucketed into the “Low Severity” category

Is this your experience?

Severity Calculation - Proposed Method 2

Elbow method indicates ideal number of K-Means clusters ~5-6

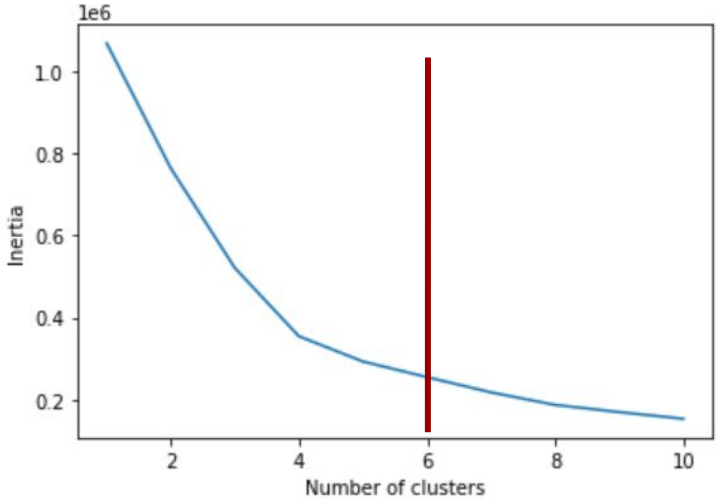
However, this results in poorly formed clusters with little differentiation

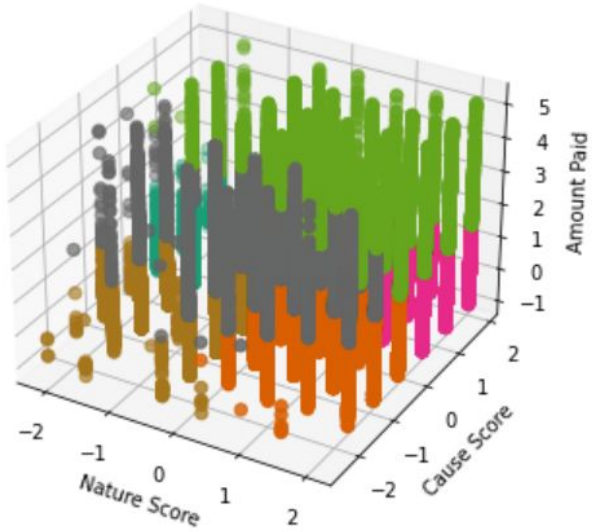


Severity Calculation - Proposed Method 2

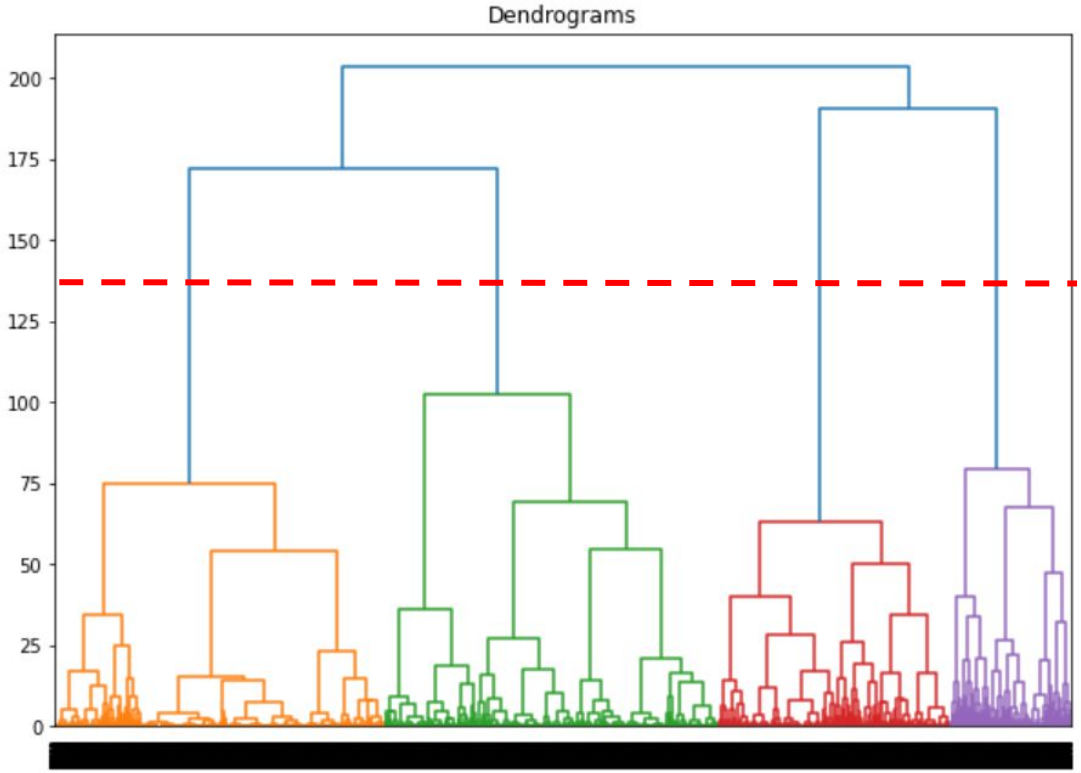
Elbow method indicates ideal number of K-Means clusters ~5-6

However, this results in poorly formed clusters with little differentiation





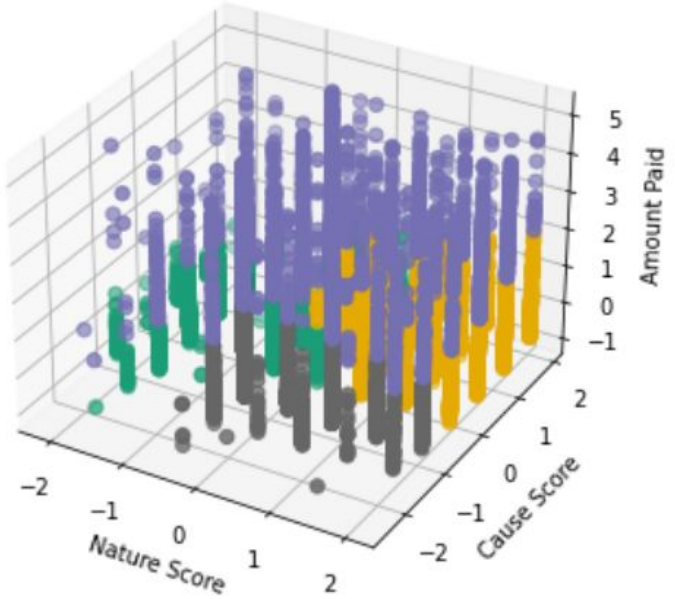
Severity Calculation - Proposed Method 3



The dendrogram indicates an ideal number of clusters is 4

Severity Calculation - Proposed Method 3

More natural clusters appear using hierarchical clustering



Pro: hierarchical clustering doesn't need circular clusters or pre-specified number of centroids

Con: large memory usage and slow algorithm

Severity Calculation - Discussion

Each method has benefits and drawbacks

Are there any other methods that our team should consider that we may have overlooked?

Are there any additional covariates that we should consider for a severity calculation?

Next Week’s Sprint

Week beginning on October 30th

Analysis on patterns in all medical cost data

Integrating medical costs into the severity calculation and model

Begin analyzing and formatting Power BI report for end users

Creating button to export model predictions