

Analysis of wage and hour case data from US state labor agencies

Data import and preparation

Data completeness

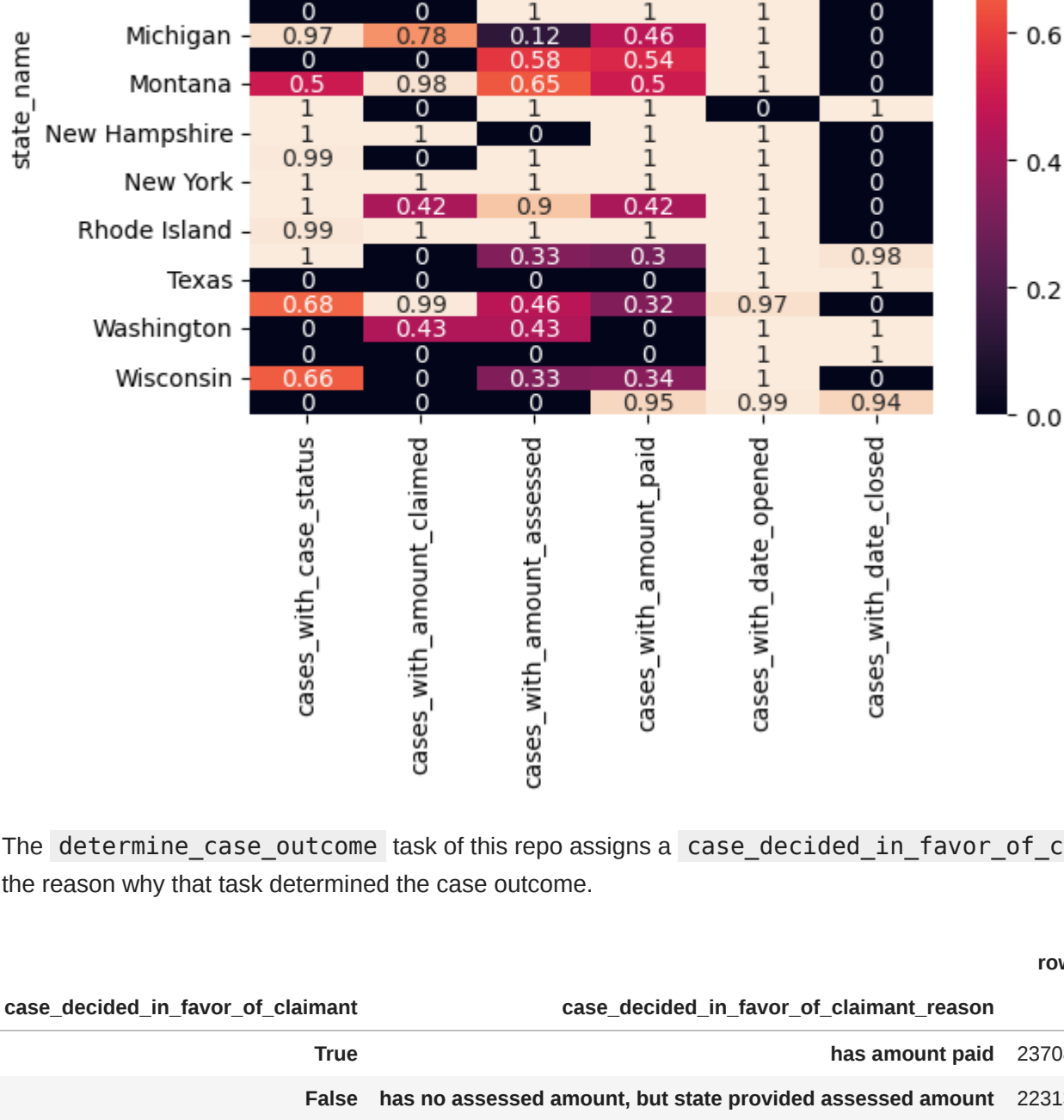
The data states provided in response to my public records requests was wildly inconsistent. I requested a long list of fields -- case statuses, start, end and paid dates, amounts claimed, assessed by the agency and paid -- but not all states possessed that data. Many states exported data directly from their case management systems, so lots of those fields were missing. This section will examine the completeness of various data points and explain how I am handling missing data.

NOTE: the `get_coverage_df` function outputs the completeness (% non-null) of key columns in the analysis. You can find it at `shared/src/shared_functions.py`.

The data overall is very incomplete and has a ton of missing values. This is largely because many of the cases sent from case management systems are not completed.

```
case_status      0.426
amount_claimed   0.521
amount_assessed  0.402
amount_paid      0.382
date_opened      0.985
date_closed      0.231
dtype: float64
```

This state breakdown heatmap shows which states have which data.

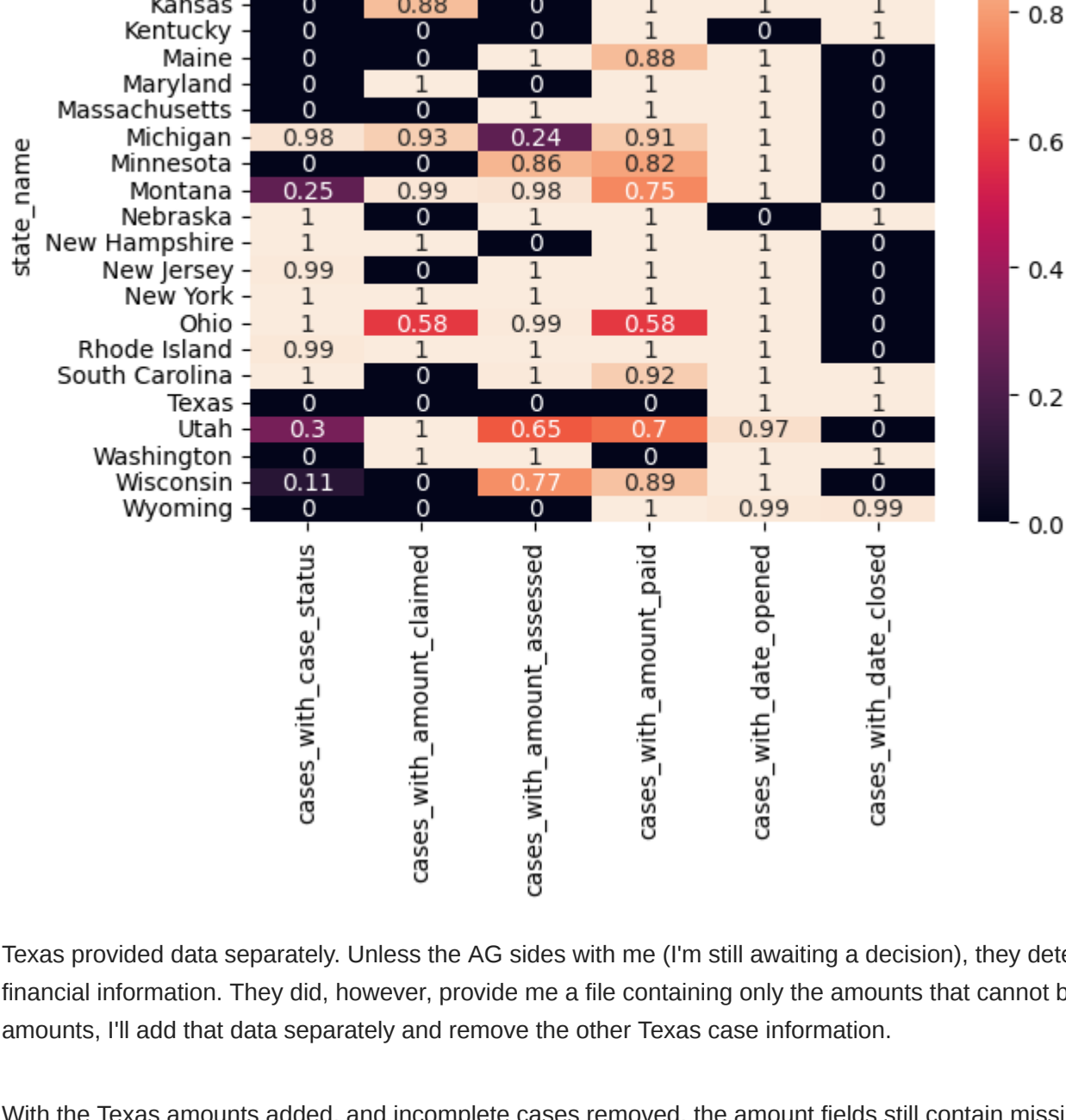


The `determine_case_outcome` task of this repo assigns a `case_decided_in_favor_of_claimant` column to the dataframe, as well as the reason why that task determined the case outcome.

			rows
case_decided_in_favor_of_claimant	case_decided_in_favor_of_claimant_reason		
	True	has amount paid	237082
False	has no assessed amount, but state provided assessed amount		223131
	state only provided closed cases		60683
True	has assessed amount greater than 0		44684
	has assessed amount less than or equal to 0		37868
False	has open or incomplete case status		28684
	has final case status		2947

When you select only cases marked completed, you get much better completeness.

```
case_status      0.442
amount_claimed   0.636
amount_assessed  0.613
amount_paid      0.687
date_opened      0.975
date_closed      0.286
dtype: float64
```



Texas provided data separately. Unless the AG sides with me (I'm still awaiting a decision), they determined that the amounts were private personal financial information. They did, however, provide me a file containing only the amounts that cannot be linked to the case data. For analyses of amounts, I'll add that data separately and remove the other Texas case information.

With the Texas amounts added, and incomplete cases removed, the amount fields still contain missing data but almost all cases contain at least one non-null amount.

396 rows (%0.11465100927630895) have no amounts

There are also indeterminate cases, in which I couldn't determine whether the case was completed

	case_uid	violation_uid	state_name	employer_name	employer_dba_name	employer_city	violation_category	case_status	date_opened	d
626545	234e0b0c-eaad-494a-b782-29740075ee94	ead79bb7-fecf-4a1f-988b-0b22b7f709e6	Indiana	Mission Complete Transport LLC		NaN	NaN	amount exceeds statutory limit	2019-09-14	
626713	03576319-6c6d-4960-a7c7-badbdc273871	d92945ae-a511-4575-80ab-d534894e320d	Indiana	Diverse Net		NaN	NaN	amount exceeds statutory limit	2019-11-04	
626749	af6b0f19-d6771-40ce-b040-ab6f6b604111	b0c68b50-961c-417f-9652-6b6176052b32	Indiana	Code red safety		NaN	NaN	amount exceeds statutory limit	2019-12-31	
626774	2b98e842-063d-4f3b-35a3-084c3e33db4e	527eece2-bb24-4997-9866-4777ca268808	Indiana	Morningview Assisted Living		NaN	NaN	amount exceeds statutory limit	2019-12-16	
626785	420dbe77-6703-4096-a2df-5a7ae0f10b6b	566d3738-4edf-4886-b635-961e62747db	Indiana	Hidden Valley Inn, Inc		NaN	NaN	amount exceeds statutory limit	2019-09-23	
...
647666	257ab6d5-12ae-4e3d-ab65-16ff8aedaa63a	662d72a-12b7-4721-b290-b787209950dc	Indiana	Creative Technic		NaN	NaN	amount exceeds statutory limit	2019-07-15	
647714	138255e0-0479-480e-94a2-f136f60d1692	985f1f3d-374f-4c5b-a4f9-f12b55f77db	Indiana	Aaa Mattress & Furniture		NaN	NaN	amount exceeds statutory limit	2019-06-18	
647742	34337b3d-c95b-40d5-9917-cd30c421c43d	5833c37f-cfa3-406f-9fbc-b802-95e4d766e514	Indiana	Trugreen Lawncare		NaN	NaN	amount exceeds statutory limit	2019-06-24	
647848	b2a04481-33f9-4d77-843c-c75acc0746ae	5201913f-f064-40b7-aab3-0e05062f531e	Indiana	XY Tool & Die		NaN	NaN	amount exceeds statutory limit	2019-07-15	
647940	a5949a38-67bc-4d09-8aae-455456404e4c	eb83f724-a702-4873-b635-791a99297745	Indiana	Real Alloy		NaN	NaN	amount exceeds statutory limit	2019-07-31	

325 rows × 16 columns

	case_uid	violation_uid	state_name	employer_name	employer_dba_name	employer_city	violation_category	case_status	date_opened	c
307280	dd75e9ae-f5f3-4513-b197-4247e4244b4e	3a93b9d4-ca78-449b-b240-4f587eb96954	Illinois	The Lucky Hot dog		NaN	NaN	NaN	closed	2022-09-28
307281	59b2b89b-7f33-4a7a-b6f5-24f222aa5f5e	40b952c8-406f-9fbc-b802-1e52f2f551f	Illinois	The Lucky Hot dog		NaN	NaN	NaN	closed	2022-09-28
307282	0a3f976-877c-4e06-96c5-575ae0cc542	ec306bfc-98b3-4c24-a0aa-d6fe427a94d2	Illinois	The Lucky Hot dog		NaN	NaN	NaN	closed	2022-09-27
307283	7203ba64-7c1c-469d-8cd7-ad4d10890997	2dd6c973-97f1-4434-80c5-2425e1d1d01	Illinois	The Lucky Hot dog		NaN	NaN	NaN	closed	2022-09-27
307284	b45ba73d-d18c-4a5f-b7c5-da0ef1d97571	c613a4cb-00b4-4efc-bc6a-c453482cd03d	Illinois	The Lucky Hot dog		NaN	NaN	NaN	closed	2022-09-27
...
626464	18919351-24b0-49c2-a48e-e53ac4c530cd0	8f4b4b4d9-0c71-431c-87d6-df34875cc95	West Virginia	KOONTZ CONTRACTING LLC		NaN	NaN	unpaid wages	NaN	2022-07-11
626465	0d019dd4-ac72-4015-8610-d09fa9a460dd	514fb270-94ef-4a69-9431-0c5320c36ac2	West Virginia	MLK STAFFING LLC		NaN	NaN	unpaid wages	NaN	2022-07-12
626466	906b9187-d9ea-45e5-b68e-18b168f02a7	ddfe80ac-66ec-4c51-be5a-837d14434ae3	West Virginia	MAXIM HEALTHCARE SERVICES INC		NaN	NaN	unpaid wages	NaN	2022-07-18
626467	fe9ec27c-f04b-4b88-bafc-be67100b966e	643cfc5b-214d-4542-947b-ftb29c919fa4	West Virginia	WENDYS INTERNATIONAL LLC		NaN	NaN	unpaid wages	NaN	2022-07-19
626468	b872de96-2b3b-4d51-b1a5-62a278e768fa	d05b3988-865c-4e98-baae-f1efb1f04671	West Virginia	SDR FOODS OF WV LLC		NaN	NaN	unpaid wages	NaN	2022-07-20

15373 rows × 16 columns

They contain very incomplete data, but all for one single reason: they are closed cases with no assessed amount and no paid amount in a state that didn't provide either.

If the state provided either an assessed or paid amount, I can use those to determine the case outcome. But if the state provided only claim amounts or no amounts at all AND no clear case disposition other than "closed," I can't be confident that those cases are all decided in favor of the claimants.

closed case with no assessed amount and no paid amount in a state that provided neither 15373

Name: case_decided_in_favor_of_claimant_reason, dtype: int64

```
case_status      0.724
amount_claimed   0.725
amount_assessed  0.000
amount_paid      0.000
date_opened      0.970
date_closed      0.275
dtype: float64
```

Those cases come only from three states -- Illinois, West Virginia and Colorado. All of those provided insufficient data to analyze

Illinois 11141

West Virginia 3778

Colorado 454

Name: state_name, dtype: int64

Analysis

Below you will find basic descriptive statistics and state breakdowns. I will expand on this in the coming days, including by making comparisons to average household income and consumer expenses. 1

I will focus on the following data points:

- overall amount assessed and the impacts that can have
- amount claimed or assessed by the state vs. the amount paid by the company
- the amount of time these cases take to resolve

All cases used in this analysis are decided in favor of the claimant unless otherwise specified.

Overall case amounts

```
count      346117.000
mean       5121.276
std        32016.237
min         0.000
25%        300.000
50%        867.680
75%        2847.700
max       5912348.830
```

Name: overall_case_amount, dtype: float64

	total_cases	pct_of_total_cases
overall_case_amount		
0–100	26634	0.079
100–500	90611	0.269
500–1,000	60395	0.179
1,000–2,500	66112	0.196
2,500–5,000	37225	0.111
5,000–10,000	25639	0.076
10,000–50,000	24902	0.074
50,000–100,000	3015	0.009
100,000–500,000	2174	0.006
500,000–1,000,000	117	0.000
\$1,000,000+	45	0.000

	count	mean	std	min	25%	50%	75%	max
state_name								
California	63442.000	10692.895	46977.871	0.010	924.820	3401.950	8658.000	5912348.830
Illinois	769.000	5479.789	11667.095	0.000	882.260	2300.000	5804.840	156500.000
Indiana	7832.000	5031.187	51612.102	0.000	245.642	590.550	1524.862	2297378.000
Iowa	7370.000	1174.898	1593.859	0.000	260.000	660.000	1500.000	45416.000
Kansas	4931.000	2131.209	29528.906	0.440	250.000	560.000	1280.500	2019015.000
Kentucky	7020.000	3495.740	16168.527	25.380	207.765	588.645	1164.223	125810.820
Maine	792.000	2996.591	8276.580	2.750	222.040	395.750	1251.312	153993.440
Maryland	8290.000	2429.502	6628.096	0.000	384.000	934.470	2200.000	287701.580
Massachusetts	2408.000	20130.521	75635.165	50.000	874.562	2596.640	9919.257	1235887.250
Michigan	41752.000	1468.392	8299.844	0.010	200.632	492.920	1140.000	1000000.000
Minnesota	743.000	1866.480	47384.888	2.000	298.000	900.000	4313.100	1115426.880
Montana	5361.000	2133.470	5969.388	0.230	280.000	720.000	1767.870	167750.180
Nebraska	563.000	903.941	2406.379	50.000	500.000	500.000	500.000	51000.000
New Hampshire	8200.000	5217.947	26228.588	0.000	342.000	920.000	2631.500	658292.000
New Jersey	25725.000	5676.761	25955.542	0.200	393.750	1101.190	3515.000	1083752.300
New York	62662.000	7895.793	45394.639	0.000	238.000	900.000	3502.043	4130507.510
Ohio	9567.000	4921.408	23184.341	0.000	0.000	712.190	1910.330	1225315.460
Rhode Island	2292.000	5343.196	31963.695	2.910	297.803	823.250	1275323.220	7053623.240
South Carolina	4804.000	1740.089	21510.529	4.000	200.000	498.575	1176.612	1424044.000
Texas	613158.000	1896.542	7704.863	0.980	300.000	710.500	1602.105	628000.000
Utah	10103.000	2038.912	3261.794	0.010	309.985	855.000	2259.655	30411.000
Washington	6365.000	2146.731	6924.469	0.010	280.000	723.000	1716.660	233785.980
Wisconsin	11004.000	3095.354	41166.491	1.000	215.220	554.630	1397.485	2031142.800
Wyoming	964.000	1125.615	1629.319	3.360	252.750	578.025	1216.500	13703.660

Amounts recovered

	pct_of_claimed_or_assessed_paid	unpaid_amount
count	255766.000	347871.000
mean	inf	1592.339
std	NaN	18142.105
min	0.000	0.000
25%	0.000	0.000
50%	1.000	0.000
75%	1.000	301.500
max	inf	3767316.730

	total_cases	pct_of_total_cases
funds recovered		
False	208808	0.733
True	75905	0.267

553928577.4067786

Texas	63158.000	1896.542	7704.863	0.980	300.000	710.500	1602.105	628000.000
Utah	10103.000	2038.912	3261.794	0.010	309.985	855.000	2259.655	30411.000
Washington	6365.000	2146.731	6924.469	0.010	280.000	723.000	1716.660	233785.980
Wisconsin	11004.000	3095.354	41166.491	1.000	215.220	554.630	1297.485	2031142.800
Wyoming	964.000	1125.615	1629.319	3.360	252.750	578.025	1216.500	13703.660