

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

NOTE: notes in red were added on January 11, 2023 and document how we're handling new information from the Texas Workforce Commission. See notebooks/texas/texas.ipynb for more details.

Data import and preparation

(652155, 18)

Data completeness

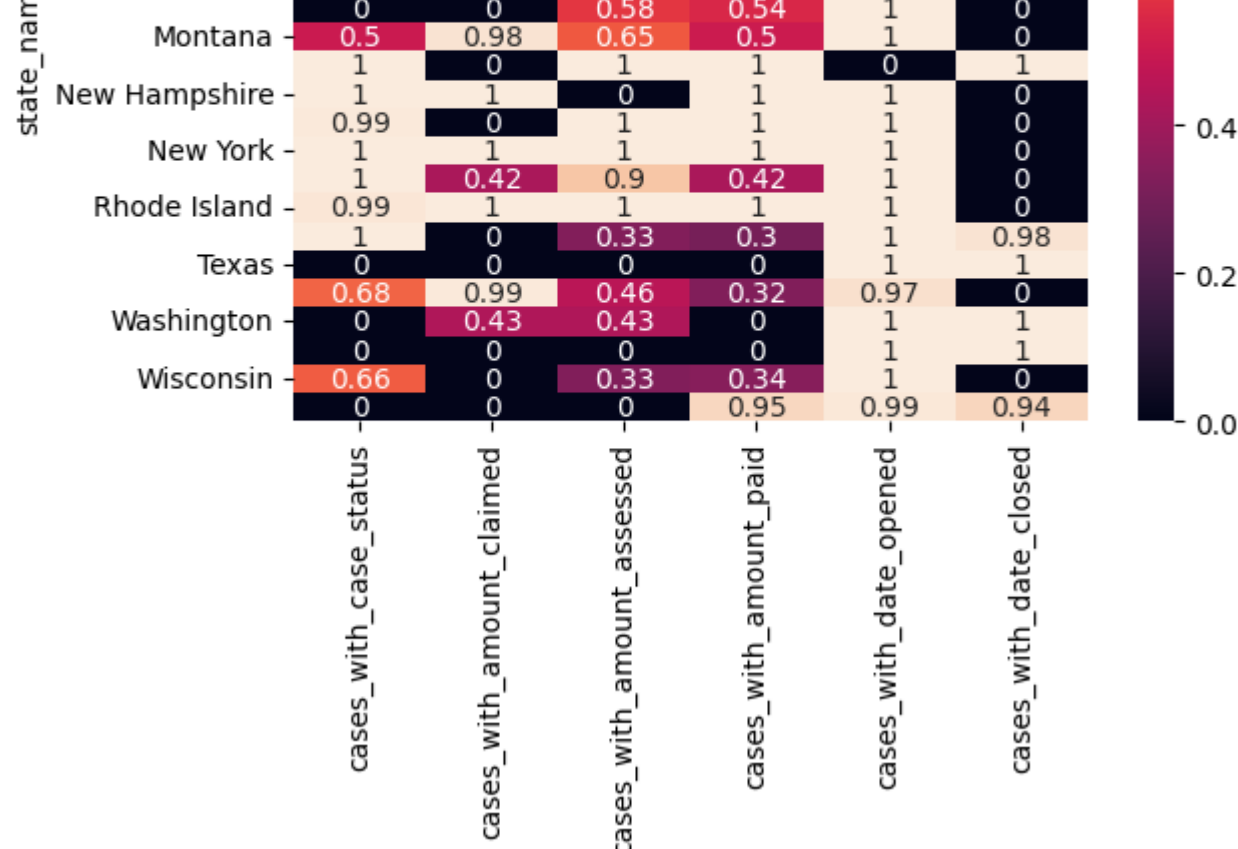
The data states provided in response to my public records requests was wildly inconsistent. I requested a long list of fields -- case statuses, start, end and paid dates, amounts claimed, assessed by the agency and paid -- but not all states possessed that data. Many states exported data directly from 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      0.425
amount_claimed   0.519
amount_assessed  0.401
amount_paid      0.381
date_opened      0.985
date_closed      0.233
dtype: float64
```

This state breakdown heatmap shows which states have which data.

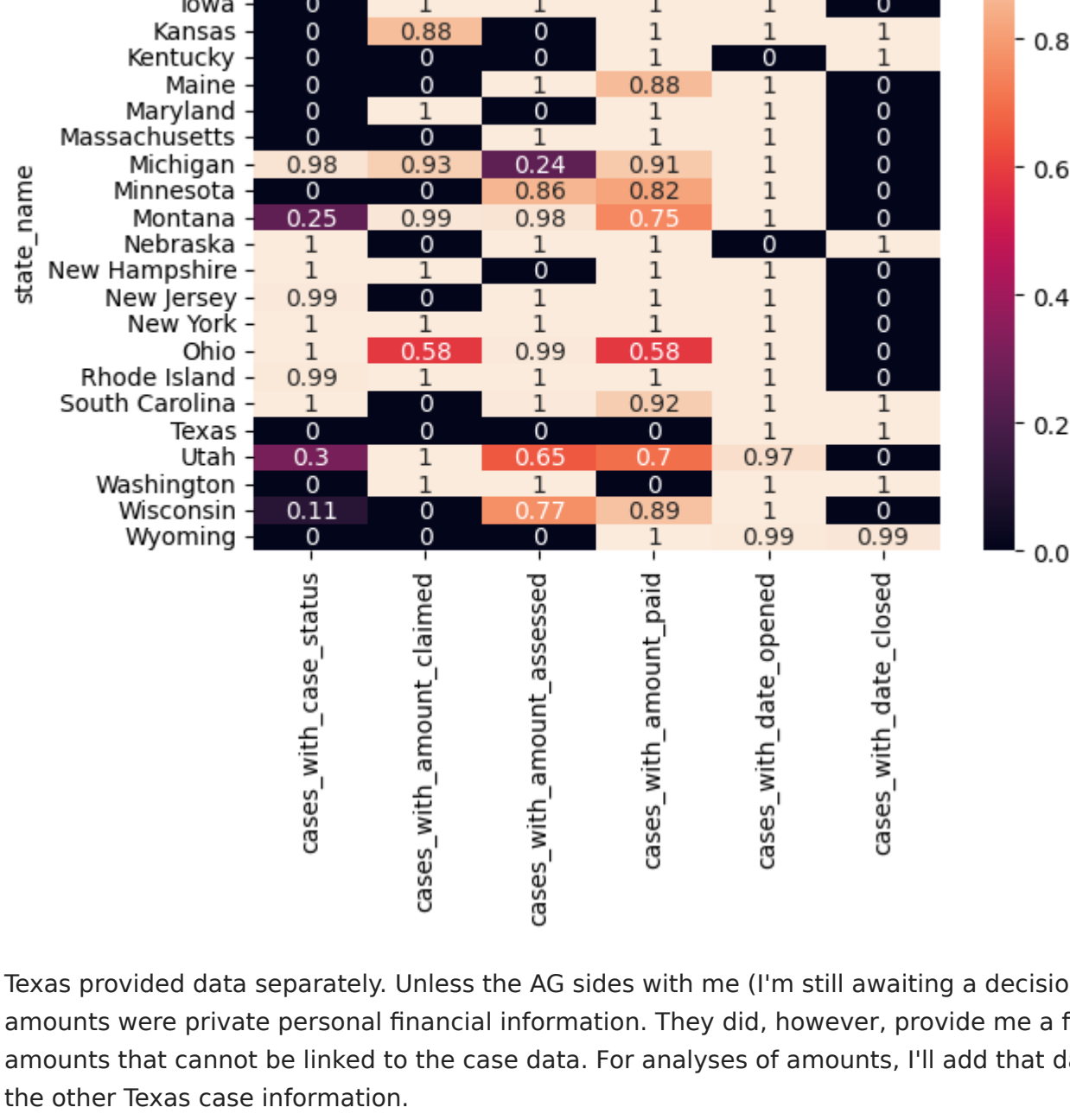


The `determine_case_outcome` task of this repo assigns a `case_decided_in_favor_of_claimant` column to the dataframe, as well as the reason why that task determined the case outcome.

case_decided_in_favor_of_claimant	rows	
	case_decided_in_favor_of_claimant_reason	
True	has amount paid	237082
False	has no amount values	178282
True	state only provided closed cases	62386
False	has no assessed amount, but state provided assessed amount	49081
True	has assessed amount greater than 0	44684
False	has assessed amount less than or equal to 0	37868
	has open or incomplete case status	28787
True	has final case status	13874

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

```
case_status      0.457
amount_claimed   0.644
amount_assessed  0.591
amount_paid      0.663
date_opened      0.976
date_closed      0.281
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.

See [shared_functions.py](#)

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 (%0.11060649226592482) have no amounts

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

5850

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

0.5489891206844999

```
count    357092.000
mean      4583.633
std       30836.330
min         0.000
25%        200.000
50%        720.000
75%       2400.000
max       5912348.830
Name: overall_case_amount, dtype: float64
```

Total amount across all states: 1,636,670,320

	total_cases	pct_of_total_cases
overall_case_amount		
0–100	25152	0.079
100–500	85504	0.269
500–1,000	57528	0.181
1,000–2,500	63470	0.200
2,500–5,000	34698	0.109
5,000–10,000	23790	0.075
10,000–50,000	22846	0.072
50,000–100,000	2752	0.009
100,000–500,000	2006	0.006
500,000–1,000,000	102	0.000
\$1,000,000+	45	0.000

	overall_case_amount	count	mean	std	min	25%	50%	75%	max
state_name									
California	558617653.636	63442.000	8805.171	44730.691	0.000	295.025	2070.010	6955.500	5912348.830
Illinois	54807091.760	11696.000	4685.969	38133.438	0.000	288.000	960.000	2700.000	1923000.000
Indiana	39404254.013	7832.000	5031.187	51612.102	0.000	245.642	590.550	1524.862	2297378.000
Iowa	853899.000	7370.000	115.861	625.433	0.000	0.000	0.000	0.000	7625.000
Kansas	10508991.876	4931.000	2131.209	29528.906	0.440	250.000	560.000	1280.500	2019015.000
Kentucky	23838093.630	7020.000	3395.740	16168.527	25.380	207.765	398.645	1164.223	125810.820
Maine	1977300.070	792.000	2496.591	8276.580	2.750	222.040	585.750	1551.312	153993.440
Maryland	20140568.600	8290.000	24038.982	6628.096	0.000	384.000	934.470	2200.000	287701.580
Massachusetts	48474294.560	2149.000	20038.501	75475.087	0.000	857.320	2560.000	9826.160	1325287.250
Michigan	61287422.210	41752.000	1467.892	8299.871	0.000	200.248	492.580	1137.010	1000000.000
Minnesota	6559247.360	744.000	8816.193	47355.421	0.000	289.175	886.500	4312.220	1115426.880
Montana	11437534.690	5361.000	2133.470	5969.388	0.230	280.000	720.000	1767.870	167750.180
Nebraska	508919.000	563.000	903.941	2406.379	50.000	500.000	500.000	500.000	51000.000
New Hampshire	42787166.000	8200.000	5217.947	26228.588	0.000	342.000	920.000	2631.500	658292.000
New Jersey	146034669.760	25725.000	5676.761	25955.542	0.200	393.750	1101.190	3515.000	1083752.300
New York	363780038.260	52662.000	6907.828	41563.036	0.000	0.000	600.000	2775.630	4130507.510
Ohio	41043963.840	9596.000	4277.195	23145.961	0.000	0.000	390.790	1851.245	1225315.440
Rhode Island	9288775.110	2292.000	4052.694	31058.217	0.000	0.000	0.000	414.430	703623.260
South Carolina	8358985.600	4810.000	1737.835	21497.200	0.000	200.000	497.815	1171.410	1424044.000
Texas	119781779.020	63158.000	1896.542	7704.863	0.980	300.000	710.500	1602.105	628000.000
Utah	18597232.020	10103.000	1840.763	3290.260	0.000	83.190	622.100	2014.460	30411.000
Washington	13663941.420	6365.000	2146.731	6924.469	0.010	280.000	723.000	1716.660	233785.980
Wisconsin	33833405.810	11005.000	3074.367	41164.701	0.000	204.310	540.000	1365.000	2031142.800
Wyoming	1085092.920	964.000	1125.615	1629.316	3.360	252.750	578.025	1216.500	13703.660

Amounts recovered

For Texas paid amounts, we have been told by the Texas Workforce Commission that they do not reliably track the amounts paid, but rather use a numeric status code that was included in the original data to designate if payment was made. For calculations of percentages of cases paid, we will use that status code to designate whether a case was paid, and for total amounts recovered we will use the actual paid amount field, even though it is incomplete.

paid	False	True	total	pct_true
n	25742	36539	62281	0.587

	total_cases	pct_of_total_cases
False	126551.000	0.383
True	204067.000	0.617

Total amount paid: \$631,838,076

	total_cases	cases_with_paid_amounts	total_amount_paid	pct_of_total_cases
state_name				
California	63442.000	26509.000	126853456.268	0.418
Indiana	8175.000	5541.000	3645489.950	0.678
Iowa	7370.000	3199.000	2617419.000	0.434
Kansas	4931.000	4931.000	10508991.876	1.000
Maine	792.000	690.000	1451891.010	0.871
Maryland	8290.000	5102.000	8346019.710	0.615
Massachusetts	2419.000	2408.000	48474294.560	0.995
Michigan	41787.000	37796.000	36834248.610	0.904
Minnesota	745.000	606.000	5319233.840	0.813
Montana	5361.000	4041.000	7312386.590	0.754
Nebraska	563.000	80.000	37324.000	0.142
New Hampshire	8200.000	3227.000	7670262.000	0.394
New Jersey	25725.000	20677.000	90118620.150	0.804
New York	52662.000	31109.000	219924360.450	0.591
Ohio	9651.000	967.000	1481679.000	0.100
Rhode Island	2292.000	1749.000	4571119.010	0.763
South Carolina	4818.000	2053.000	307040.600	0.426
Utah	10103.000	7036.000	7911044.940	0.696
Washington	11011.000	9807.000	25047136.630	0.891
Wisconsin	62281.000	36539.000	NaN	0.587

Case duration

```
count          92920
mean    310 days 16:15:04.139552204
std      410 days 08:37:45.687898968
min           0 days 00:00:00
25%          106 days 00:00:00
50%          196 days 00:00:00
75%          337 days 00:00:00
max         4585 days 00:00:00
Name: case_duration, dtype: object
```

	total_cases	pct_of_total_cases
duration_days		
0-30 days	4594	0.049
30-60 days	5104	0.055
60-90 days	8446	0.091
90-180 days	24617	0.265
180-360 days	29375	0.316
360-720 days	13297	0.143
720+ days	7387	0.080

	count	mean	std	min	25%	50%	75%	max
state_name								
California	12366.000	523.836	306.061	26.000	300.000	439.000	676.750	1951.000
Kansas	4917.000	158.355	306.790	0.000	47.000	99.000	168.000	3869.000
South Carolina	4807.000	65.342	48.877	0.000	35.000	57.000	84.000	826.000
Texas	62386.000	332.346	446.831	29.000	133.000	209.000	324.000	4585.000
Washington	6364.000	55.863	55.484	0.000	22.000	43.000	69.000	1197.000
Wyoming	2080.000	99.992	137.802	0.000	24.000	48.000	124.000	1114.000

Comparison to state felony theft thresholds

	total_cases	pct_of_total_cases
felony_theft		
False	475856	0.729
True	177071	0.271

felony_theft		False	True	total	pct
state_name	felony_theft_threshold				
California	950.0	150618	39609	190227	0.208
Colorado	2000.0	454	0	454	0.000
Illinois	500.0	7470	16339	23809	0.686
Indiana	750.0	12821	8690	21511	0.404
Iowa	1000.0	7090	280	7370	0.038
Kansas	1500.0	7171	3080	10251	0.300
Kentucky	500.0	3978	3159	7137	0.443
Maine	1000.0	520	272	792	0.343
Maryland	1500.0	5364	2926	8290	0.353
Massachusetts	1200.0	802	1618	2420	0.669
Michigan	1000.0	58677	24347	83024	0.293
Minnesota	1000.0	773	354	1127	0.314
Montana	1500.0	5336	2722	8058	0.338
Nebraska	1500.0	530	49	579	0.085
New Hampshire	1000.0	4319	4107	8426	0.487
New Jersey	200.0	3606	22119	25725	0.860
New York	1000.0	40212	21648	61860	0.350
Ohio	1000.0	14468	3328	17796	0.187
Rhode Island	1500.0	1955	341	2296	0.149
South Carolina	2000.0	13944	680	14624	0.046
Texas	2500.0	53394	9764	63158	0.155
Utah	1500.0	15258	6418	21676	0.296
Washington	750.0	34523	3127	37650	0.083
West Virginia	1000.0	3778	0	3778	0.000
Wisconsin	2500.0	26880	1787	28667	0.062
Wyoming	1000.0	1915	307	2222	0.138
total	total	475856	177071	652927	0.271