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. 0.426 case_status 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. <AxesSubplot: ylabel='state_name'> 1.0 0.21 California 0.076 0 Illinois 1 0.98 0 0.21 0.26 0.78 lowa 0 0 1 0.8 0.99 0.022 0 0.48 0.89 Kentucky 000 0 0 0 0 0.88 0 1 0 Maryland -1 0 0 1 0 0.97 1 0 Michigan 0.46 0.6 0 0 0.54 1 0 0 Montana 0.5 0.98 0.5 1 0 0 1 New Hampshire -0 0.99 1 0 0 1 0.4

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

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

Data import and preparation

Data completeness

lowa

Kansas

Maine

Kentucky

Maryland

Michigan

Minnesota

New Jersey -

Rhode Island -

Washington

Wisconsin

Wyoming -

South Carolina

non-null amount.

626545

626713

626749

626774

626785

647666

647714

647742

647848

647940

307280

307281

307282

307283

307284

626464

626465

626466

626467

626468

case_uuid

234ed0bc-

eaad-494a-

03576319-

6c6d-4960-

29740075ee94

badbdc273871

fabc0f19-de77-

4cce-b040-

2b98e842-

063d-4f3b-

420dbe77-

6703-4096-

257ab6d5-

12ae-4e3d-

1dff8aeda63a

138255e0-

0479-48be-

34337b3d-

c95b-40d5-

b2a04481-

33f9-4d77-

a5949a38-

67bc-4d09-

dd75e9ae-

f5f3-45f3-

59b2b89b-

7f33-4a7a-

4247e4244b4e

0a3ff976-877c-

575aefccc542

4e06-96c5-

7203ba64-

7c1c-469d-

b45ba738d18c-4a27-

18919351-

24b0-49c2-

0d019dd4-

ac72-4015-

906b9187-

d9ea-45e5-

18b1fd8f02a7

fe9ec27c-f04b-

be67100b966e

4b88-bafc-

b872de56-

2bab-4651-

62a278e768fa

15373 rows × 16 columns

didn't provide either.

case_status amount_claimed

amount_paid

date_opened

date_closed

Illinois

Colorado

dtype: float64

West Virginia

Analysis

amount_assessed

b1a5-

d09fa9a460bd

e53a4c530cd0

a48e-

8610-

b86e-

ad4d10890997

da0ef1d97571

8cd7-

b7c5-

b197-

bf66-

aaf5fe

cd30c421c43d

c75acc0746ae

455456404e4c

325 rows × 16 columns

f136f60d1692

a8e5-

9f4a-

9917-

843c-

8aae-

5a7ae0f10b6b

084c3e33db4e

95a3-

a2bf-

a6f6fb604111

b782-

a7c7-

New York -

Ohio -

Texas

Utah

Montana -Nebraska -

Massachusetts

New Hampshire -

0

0

0

0

0

0

0.98

0

0.25

1

1

0.99

1

1

0.99

0

0.3

0

0.11

0

cases with case status

396 rows (%0.11465100927630895) have no amounts

violation_uuid

ead79bb7-fecf-

0b22fbf7b9e6

4a1f-988b-

d92845ae-

a511-4575-

b0c68b50-

961c-417f-

527eece2-

bb24-4997-

566d3738-

4edf-4886-

662d7d2a-

b280-

12b7-4721-

7b8720990bdc

985f1f3d-374f-

f12b55f77dbc

5833c37f-cfa3-

95e4d766e514

5201913f-f064-

0e05062f331e

40b7-aa63-

eb83f724-

b635-

a702-4873-

791a99297745

violation_uuid

3a93b9d4-

ca78-449b-

40b952c8-

0d26-4e44-

ec306bfc-

98b3-4c24-

2dd6c973-

97f1-4434-

c613a4cb-

00b4-4efc-

8f4bb4d9-

0c71-431c-

df34f875cc95

514fb270-94ef-

0c5320c36ac2

4a69-9431-

ddfe80ac-

643cfc5b-

214d-4542-

ffb29c919fa4

d05b3f88-

865c-4e98-

f1efbf104671

be5a-

947b-

baee-

Name: case_decided_in_favor_of_claimant_reason, dtype: int64

0.724

0.725

0.000

0.000

0.970

0.275

11141

Name: state_name, dtype: int64

I will focus on the following data points:

Overall case amounts

count mean

std min

25%

50%

75%

max

overall case amount

346117.000

5121.276 32016.237

0.000

Name: overall_case_amount, dtype: float64

26634

90611

60395

66112

37225

25639

24902

3015

2174

117

45

mean

10692.895

5031.187

1174.898

2131.209

3395.740

2496.591

2429.502

20130.521

1468.392

8836.480

2133.470

903.941

5943.196

1740.089

1896.542

2038.912

2146.731

3095.354

1125.615

pct_of_claimed_or_assessed_paid unpaid_amount

255766.000

inf

NaN 0.000

0.000

1.000

1.000

total_cases pct_of_total_cases

208808

75905

71104025.493

951550.794

596069.000

4596031.530

300382.560

16176739.150

35710253.990

988249.230

1510725.360

471595.000

38291271.000

55916049.610

183733408.220

19678929.430

6989232.550

7638860.460

96994415.180

4509400.470 10103.000

7771388.380 11011.000

310 days 14:02:53.448449204

412 days 07:59:42.409025324

Name: case duration, dtype: object

4594

5102

8393

24256

28515

12934

7323

count

4917.000 158.355

12366.000

4807.000

60683.000

6364.000

2080.000

mean

523.836

65.342

332.816

55.863

sum

inf

count

63442.000

8175.000

7370.000

4931.000

792.000

8290.000

41787.000

745.000

5361.000

563.000

8200.000

25725.000

52662.000

9651.000

2292.000

4818.000

63158.000

0 days 00:00:00

105 days 00:00:00

195 days 00:00:00

337 days 00:00:00 4585 days 00:00:00

total_cases pct_of_total_cases

5217.947 26228.588

5676.761 25955.542

7895.793 45394.639

4321.408 23184.341

count

63442.000

769.000

7832.000

7370.000

4931.000

7020.000

792.000

8290.000

2408.000

743.000

5361.000

563.000

8200.000

9567.000

2292,000

4804.000

63158.000

10103.000

6365.000

11004.000

964.000

41752.000

total_cases pct_of_total_cases

300.000

867.680

2847.700 5912348.830

0 - 100

100 - 500

500 - 1,000

1,000-2,500

2,500 - 5,000

5,000 - 10,000

10,000 -**50,000**

50,000 -**100,000**

100,000-500,000

500,000 - 1,000,000

state_name

California

Illinois

Indiana

Kansas

Kentucky

Maryland

Michigan

Minnesota

Montana Nebraska

New Jersey 25725.000

Ohio

Texas Utah

New York 52662 000

Massachusetts

New Hampshire

Rhode Island

Washington

Wisconsin

Wyoming

count

mean

std

min 25%

50%

75%

max

funds recovered

False

True

553928577.4067786

state_name

California

Indiana

Kansas

Maine

Maryland

Michigan

Minnesota

Montana

Nebraska

New Hampshire

New Jersey

Rhode Island

South Carolina

New York

Ohio

Texas

Utah

Wisconsin

Case duration

count

mean

std

min

25% 50%

75%

max

duration_days

0-30 days

30-60 days

60-90 days

90-180 days

180-360 days

360-720 days

720+ days

state_name

South Carolina

Washington

Wyoming

California Kansas

Texas

Iowa

Amounts recovered

South Carolina

Maine

Iowa

\$1,000,000+

3778 454

average household income and consumer expenses. 1

3. the amount of time these cases take to resolve

1. overall amount assessed and the impacts that can have

2. amount claimed or assessed by the state vs. the amount paid by the company

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

0.079

0.269

0.179

0.196

0.111

0.076

0.074

0.009

0.006

0.000

0.000

min

0.010

0.000

0.000

0.000

0.440

25.380

2.750

0.000

50.000

0.010

2.000

0.230

50.000

0.000

0.000

2.910

4.000

0.980

0.010

0.010

1.000

3.360

347871.000

1592.339

18142.105

0.000

0.000

0.000

301.500

3767316.730

0.733

0.267

mean

116.398

80.878

932.069

379.271

1951.356

854.578

1326.509

281.799

837.647

4669.667

2173.607

3488.918

2039.056

3049.403

1585.484

1535.742

446.343

91217

0.050

0.056

0.092

0.266

0.313

0.142

0.080

std

306.061

306.790

48.877

450.520

55,484

99.992 137.802

min

26.000

0.000

0.000

29.000

0.000

0.000

25%

300.000

47.000

35.000

133.000

22.000

24.000

50%

99.000

57.000

208.000

43.000

439.000 676.750

75%

168.000

84.000

323.000

48.000 124.000 1114.000

max

1951.000

3869.000

826.000

4585.000

69.000 1197.000

1120.772 17926.635

std

46977.871

51612.102

1593.859

29528.906

16168.527

8276.580

6628.096

75635.165

8299.844

47384.888

5969.388

2406.379

31963.695

21510.529

7704.863

3261.794

6924.469

41166.491

1629.319

5479.789 11667.095

25%

924.820

882.260

245.642

260.000

250.000

207.765

222.040

384.000

874.562

200.632

298.000

280.000

500.000

342.000

0.000

297.803

200.000

300.000

309.985

280.000

215.220

252.750

0.200 393.750

0.000 238.000

50%

3401.950

2300.000

590.550

660.000

560.000

585.750

934.470

2596.640

492.920

900.000

500.000

823.190

498.575

710.500

855.000

723.000

554.630

578.025

75%

8658.000

5804.840

1524.862

1500.000

1280.500

1551.312

2200.000

9919.257

1140.000

4313.100

500.000

2381.997

1176.612

1602.105

2259.655

1716.660

1397.485

1216.500

50%

0.000

0.000

0.000

0.000

0.000

2.190

0.000

0.000

500.000

532.500

0.000

0.000

0.000

0.000

362.250

420.000

0.000

0.000

296.580

25%

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

69.812

32.000

0.000

0.000

500.000

std

2594.808

524.621

5669.344

5235.251

7979.102

13495.129

3299.363

2421.692

25789.510

20744.120

33241.377

20448.727

28793.288

21460.725

705.784 22237.170 0.000

7436.291 0.000

1543.811 0.000

4595.715 0.000

min

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

0.000

75%

0.000

29.830

0.000

0.000

173.725

1562.315

362.185

0.000

0.000

0.000

62.377

32.695

0.000

1009.523

1246.378

20.000

500.000

2100.000

max

3767316.730

231039.270

145000.000

143126.310

408010.100

250000.000

267256.140

167673.210

51000.000

658292.000

1083752.300

3257734.000

1225315.440

633503.260

628000.000

26205.580

0.000 1987997.000

7625.000

398.645 1164.223

720.000 1767.870

920.000 2631.500

1101.190 3515.000

900.000 3502.043

415.250 1910.330

max

5912348.830

156500.000

2297378.000

2019015.000

125810.820

153993.440

287701.580

1325287.250

1000000.000

1115426.880

167750.180

51000.000

658292.000

1083752.300

1225315.440

703623.260

1424044.000

628000.000

30411.000

233785.980

2031142.800

13703.660

45416.000

66ec-4c51-

837d14434ec9

87d6-

efc2455e1d01

c453482cf03d

d6fe427a94d2

4f587eb96954

b240-

bf02-

2f551ff

a0aa-

80c5-

bc6a-

406f-9fbc-

4c5b-a4f9-

477fca2688b8

961ef82747db

d534894e320d

6b6176b52b32

80ab-

9652-

8e66-

ab5d-

1

0.88

0

0

0

0.93

0.99

0

1

0

1 0.58

1

0

0

1

0

0

claimed

cases with amount

amounts, I'll add that data separately and remove the other Texas case information.

0

0

0.24

0.86

0.98

0

1 1

0.99

1

0

1

0

cases_with_amount_assessed

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

Indiana

state_name

Illinois

Illinois

Illinois

Illinois

Illinois

West

West

West

West

West

Virginia

Virginia

Virginia

Virginia

Virginia

New York -111 0 0.42 0.9 0.42 0 Rhode Island -0.99 0 1 0 0.33 0.98 0 0 Texas 0 0.2 0.97 0.99 0 0.46 0.32Washington 0.43 0.43 1 0 0 0 0.33 1 0 000 0 Wisconsin 0.34 0 0 0.95 0.99 0.94 cases with case status with amount claimed with amount assessed ses_with_amount_paid ses_with_date_opened cases with date closed 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 has amount paid 237082 True False has no assessed amount, but state provided assessed amount True state only provided closed cases 60683 has assessed amount greater than 0 44684 **False** has assessed amount less than or equal to 0 37868 has open or incomplete case status 28684 True 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 <AxesSubplot: ylabel='state_name'> 1.0 California -1 1 0.2 1 0 Illinois 1 0 0 0.94 Indiana 0.22 0.78 0 0.56

1

1

1

0.88

1

1

0.91

0.82

1

1

1

1

0.58

1

0.92

0

0

0.89

1

cases with amount paid

1

1

0

1

1

1

1

1

0

1

1

1

1

1

1

1

0.97

1

1

0.99

cases with date opened

Texas provided data separately. Unless the AG sides with me (I'm still awaiting a decision), they determined that the amounts were private personal

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

state_name employer_name employer_dba_name employer_city violation_category case_status

NaN

employer_name employer_dba_name employer_city violation_category

NaN

amount

exceeds

statutory

limit

limit

limit

limit

limit

closed

closed

closed

closed

closed

NaN

NaN

NaN

NaN

NaN

limit

limit

limit

2019-09-14

2019-11-04

2019-12-31

2019-12-16

2019-09-23

2019-07-15

2019-06-18

2019-06-24

2019-07-15

2019-07-31

2022-09-28

2022-09-28

2022-09-27

2022-09-27

2022-09-27

2022-07-11

2022-07-12

2022-07-18

2022-07-19

2022-07-20

NaN

unpaid wages

unpaid wages

unpaid wages

unpaid wages

unpaid wages

15373

financial information. They did, however, provide me a file containing only the amounts that cannot be linked to the case data. For analyses of

Mission

Complete

Transport LLC

Diverse Net

Code red safety

Morningview

Assisted Living

Hidden Valley

Creative Technic

Aaa Mattress &

Furniture

Trugreen

Lawncare

XY Tool & Die

Real Alloy

The Lucky Hot

KOONTZ

LLC

LLC

MAXIM

WENDYS

LLC

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

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.

CONTRACTING

MLK STAFFING

HEALTHCARE

SERVICES INC

INTERNATIONAL

SDR FOODS OF

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

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

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

dog

dog

dog

dog

Inn, Inc

0

1

0

0

0

0

0

0

0

0

0

0

0

0

0.99

cases with date closed

0.8

0.6

0.4

0.2