

Texas wage theft data details

We shared the results of our analysis with the Texas Workforce Commission so they could comment. After receiving no response for about two weeks, Brian New, our local reporter in Texas interviewed the director of the Commission. In that interview, the director told Brian that they gave us the wrong fields for several of the fields we requested. This notebook will document how we're handling this in our analysis. There are also additional notes in notebooks/national_analysis/national_analysis.ipynb.

They informed us there are duplicates in the data. These have also been addressed using a new processor in task 1

We were told that their data does not reliably track the amounts of money that claimants received, despite the fact that a field exists in their database and appears to be used some of the time. They informed us that these numeric status codes are used to designate cases that are paid.

These same figures are also now reflected in notebooks/national_analysis/national_analysis.ipynb.

```
-----
NameError                                Traceback (most recent call last)
Cell In[1], line 2
      1 (
----> 2     df.assign(paid=lambda x: x.FK_VCMLNT_STSCD.astype(int).isin(PAID_STATUS_CODES))
      3     .query("AWARDED == 'YES' | paid == True")
      4     .query("FK_VCMLNT_STSCD.isin(@DROP_STATUS_CODES) == False")
      5     .groupby("paid")
      6     .size()
      7     .to_frame("n")
      8     .transpose()
      9     .assign(total=lambda x: x[True] + x[False])
     10     .assign(pct_true=lambda x: x[True] / x["total"])
     11 )

NameError: name 'df' is not defined
```

For comparison, this is the figure you get when you use the paid amounts.

	False	True	pct_true
n	106124	13341	0.111673

The TWC also informed us that they provided the wrong date field for the case end date. They gave us the status date, which designates the most recent change to the case, such as filing an appeal or entering a lien. They told us that they would rather us use the date a preliminary decision was reached, which gives you a 65 day case duration rather than the > 200 day duration we had originally. However, we believe that the status date is fair to use. The preliminary decision does not mean a person was paid, and if a case is appealed or penalties are applied to a business, that still means that the case is functionally pending for the claimant.

They did not provide the decision date field.