**Data Source:**

1. Transactional data (TTD and TPD)
2. Claim basic info (dept, cause, etc.)
3. Los Days info (All lost days, All Restr days, Is Accommodate, Transitional Duty, etc)

**Filter**: on or after 2016/1/1

**Combine dataset**: all records within 5 year with TTD or TPD data.

**Problem1:** We have 489 records, 26 of them are missing loss days information.

Solution1: Fill those missing data (I will provide Claim Number)

**Problem2:** Around 50% missing in ‘Is Accommodate’ and ‘Transitional Duty’ column. We need those information to distinguish the difference between Lost days and Restr days.

Solution2:

* Assume all missing data as False
* Or clean them one by one

**Problem3:** TTD+TPD != Indemnity, is that ok?

Now let’s take 463 dataset to do some viz.



