

CIS 4400

Homework 1

Kelly Huang

## Business Requirements:

Workers' compensation non-subscriber employer information data set is about This data set includes basic information about employers that have opted out of Texas workers' compensation coverage over the past two years. I would like to analyst which company have most of people doesn't have Texas workers' compensation coverage

## Functional Requirements:

### Column:

filing\_id : Filing ID

filing\_date\_received: Filing Date Received

record\_type : Record Type separate by Parent / Location

company\_name : Company who out of Texas workers' compensation

address\_1 : Company's address

address\_2: Company's address apt number

city : Company's address city

state : Company's address state

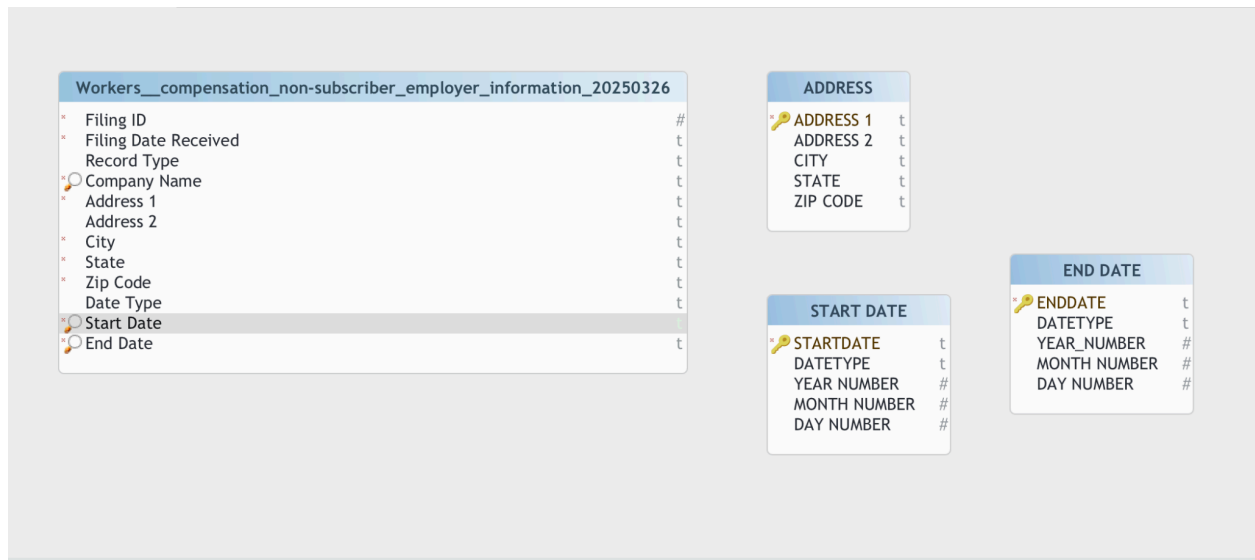
zip\_code : Company's address zip code

date\_type : Separate by filing Period and Location

start\_date : the start date for out of Texas workers' compensation

end\_date: the start date for out of Texas workers' compensation

Company Name , start\_date and end\_date are the primary key.



## Data Requirements

website link:

Workers' compensation non-subscriber employer information

[https://data.texas.gov/dataset/Workers-compensation-non-subscriber-employer-information/azae-8krr/data\\_preview](https://data.texas.gov/dataset/Workers-compensation-non-subscriber-employer-information/azae-8krr/data_preview)

## Data Sourcing

1. Remove the duplicate and clean the data
2. Sort and separate different state
3. Sum how many people are out of Texas workers' compensation on that company
4. Create a histogram chart combine and compare with all companies in every year
5. Find out which those are higher
6. Save the chart picture and data file