ETL Final Report:

Objective:

I want to study the average earnings, benefit offers, and turnover rates by the industry at the national scale. The idea is to study overall happiness of an employee by the industry.

Data Sources:

1. Bureau of Labor Statistics (https://data.bls.gov/)
2. NAICS from US Census (https://www.census.gov/eos/www/naics/)

Process:

Use API call to BLS to extract the data from the Bureau of Labor.

Plan to extract multiyear for:

Weekly Earning data by Industry Sectors

Turnover over by Industry Sectors

Holiday Benefits available by the Industry Sectors

Created a cvs file of the sector codes.

Extracted Weekly Earning data and created dictionaries for extracted data.

Uploaded into Mondo DB into four collections.

Planning to aggregate four collections into one collection.

Reason to choose sources:

BLS provides overall statistics for earning, benefit, and turnover rate.

Schema & Tables in Mongo DB.

Schema: nat\_earn\_ind (National Earnings Stats by Industry)

Tables:

Workers – All Employees

Hours – Average Weekly Hours Worked

OverTime – Average Weekly Overtime Hours Worked

WeeklyEarning – Average Weekly Earnings

Cases:

1. Which Industry expects higher overtime from employees?
2. Which Industry pay higher wages and better holiday benefits?
3. Which Industry has higher turnover rate and its relation to avg. weekly hours?