CASE STUDY 013

[Python]

US Working Visa Analysis

Disclaimer: the dataset of Labor Condition Application used in this case study was downloaded from United States Department of Labor website. The dataset has minors modifications from the website version to simplify the analysis in this case study, but it can be downloaded in its original form here.

The dataset containing NAICS (North America Industry Classification System) codes was downloaded from data.gov website.

And the populations of US states was extracted from this Wikipedia page and stored as a csv file.

You are a Data Scientist conducting a research of how US companies hire and apply for working visa for foreign workers. Your objective is to find interesting patterns in given data.

You have been supplied with three real datasets:

- The result of 2016 applications for H1B visas;

- NAICS codes;

- US States;

Your analysis must be able to address the following questions:

1) What are the top 10 occupations for H1B visas with certified status?

2) Which company hires more data scientists?

3) What is the average salary for data scientists? How it is compared to the mean of applications?

4) Which state has the higher deny proportionally to application. And proportionally to the population?

5) What industries are in the top 5 percentile of applications?

6) Does the month of the application play any role in to denied process? What about decision dates?

Good luck!

Difficulty note: this is a difficult assignment. Do not be surprised that there will be lots of nuances we have not covered off in the courses. But just like in the Real Life – there will be things training has not prepared you for and you will need to do research to find how to solve the problems at hand. If you get stuck, check the clues file.