Data Mining: H1B Applications

Roshan Thalal

Data Science and Informatics, CSIS, Northwest Missouri State University
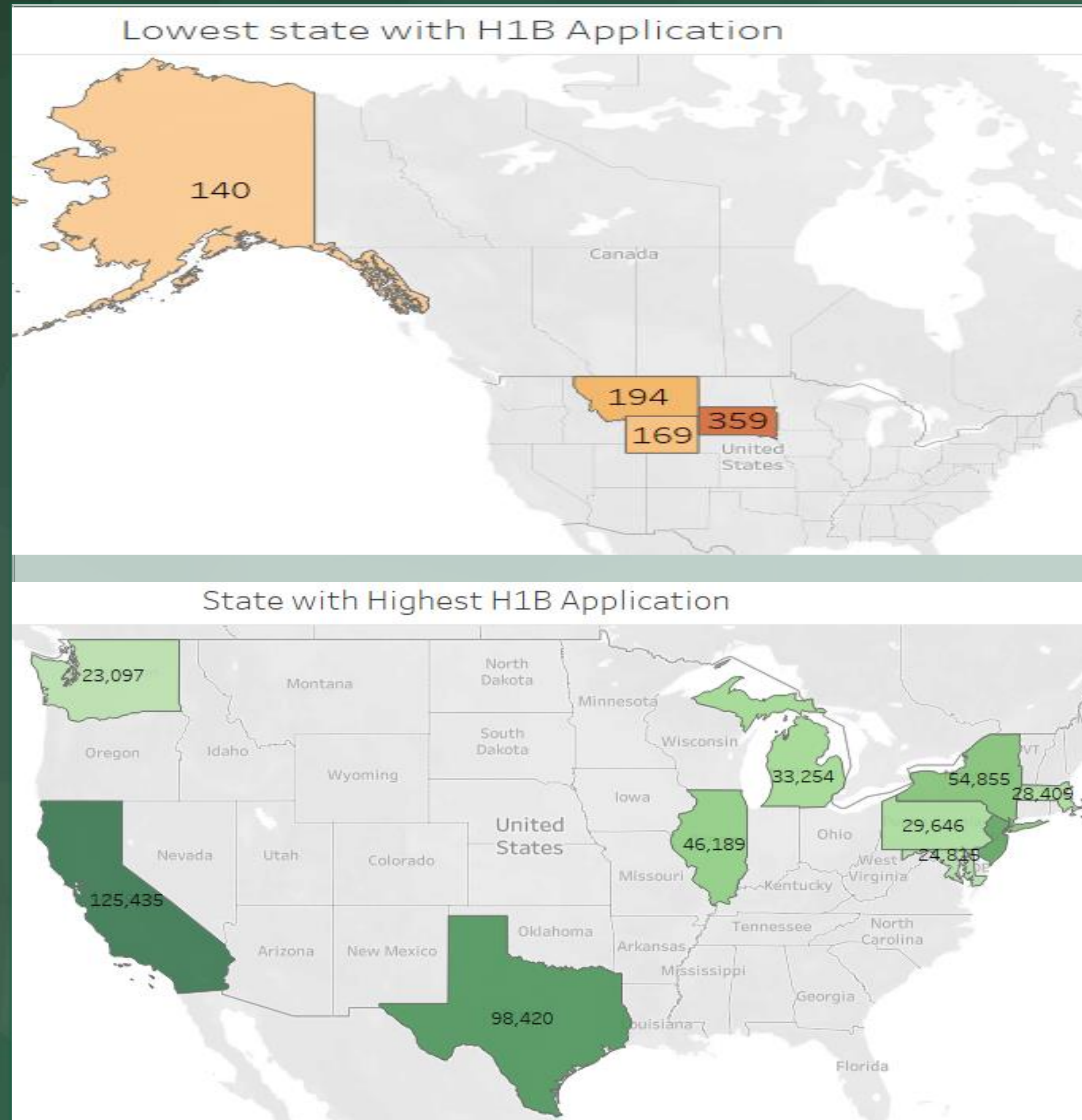
Introduction

- H1B application program is written to pursue total number of H1B applications from each states for five year duration, starting from 2012 till 2017.

Methodology

- Codes are written in python

- Program uses three separate files, mapper.py, reducer.py and sort.py

- Sorted data is then transferred to excel file from .txt format

- Tableau, one of the finest visualization tool, adds life to those number with one of its amazing features, Heat map.

Heat Map

Lowest state with H1B Application

State with Highest H1B Application

Conclusion

- H1B program runs H1B application data from 2012 to 2017, total of 1.5 million records. H1B program generates highest 10 and lowest 5, H1B application filing states within United States.

- According to the data, Alaska has filed total of 140 H1B Visa Applications from 2012 to 2017, securing the lowest position in this race.

- State of California manages to rule the top of this report with the total of 125, 435 applications over 5 year period

Resources Required

Data Source:https://www.kaggle.com/evangelize/h1b-visa-applications#H1B.csv
Python/ Pycharm
Tableau

Poster Repository

https://github.com/Roshanthalal13