

## DATA SCIENCE JOB POSTING ANALYSIS

## GROUP 3

Eben Ezer Sitinjak (PIC)  
Paskah Riverdi Manurung  
Delina Kezia  
Muhammad Farhan Firdaus  
Shafira Naya Aprisadiani  
Floean Alexander  
Ni Putu Putri Marinda  
Muhammad Farrel Savero

## INSIGHTS GAINED FROM:

- Relation between **Each of skills needed** and State
- Relation between **Salary** at Each State
- Relation between **Work Load** at Each State

## DATA MANIPULATION:

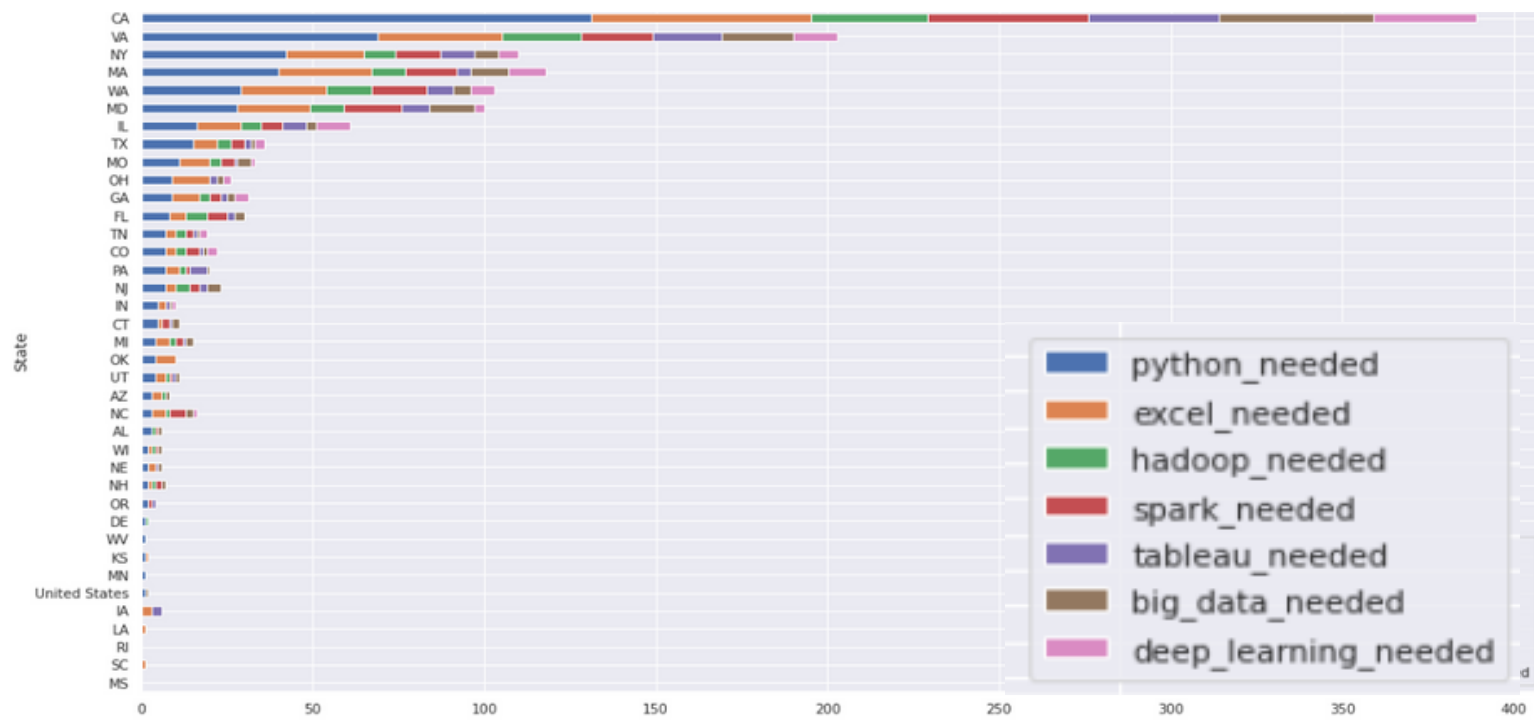
- **State** is defined from *Location* columns with second phrase after comma (*split by ', '*) with some modification

Ex: New York, **NY** → States

- **Workload** is defined as the length of the job description
- **Skill** needed is extracted from job description that contains at least these keywords: *python, excel, hadoop, spark, tableau, big data, deep learning*

# Relation between Each of skills needed and State

State with Each Need Skilled Employee Distribution

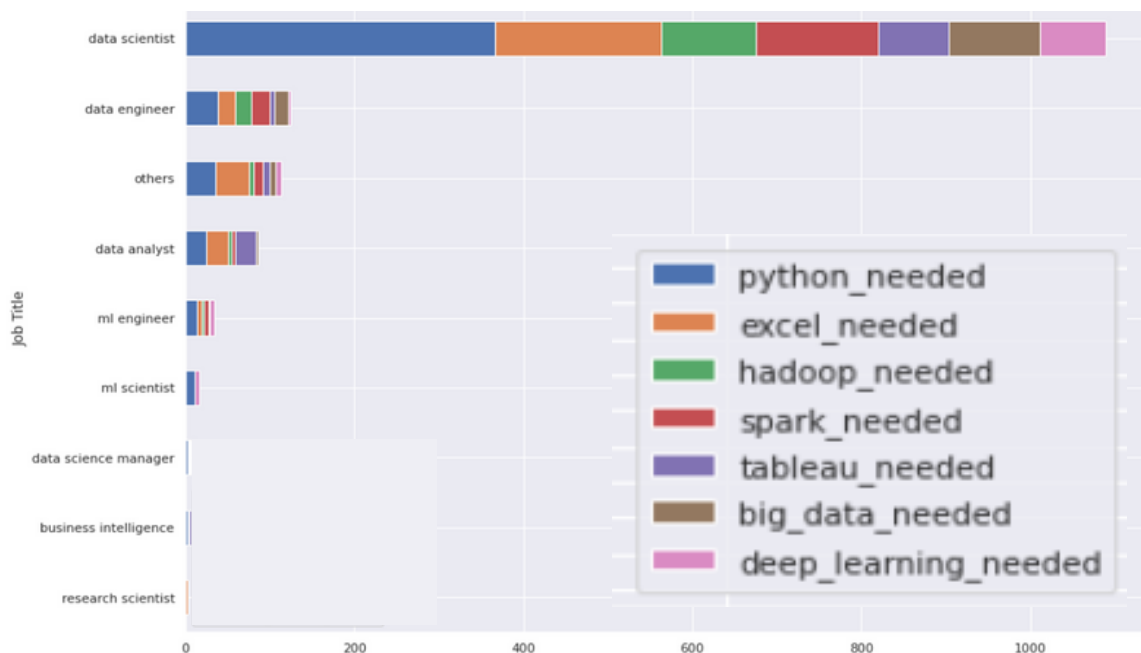


**TOP 5**  
States that need many/varied Skills

- 📍 CA
- 📍 VA
- 📍 NY
- 📍 MA
- 📍 WA

 *python™ is the most needed skill in each state*

Job Distribution that Needed Python Skilled Employee

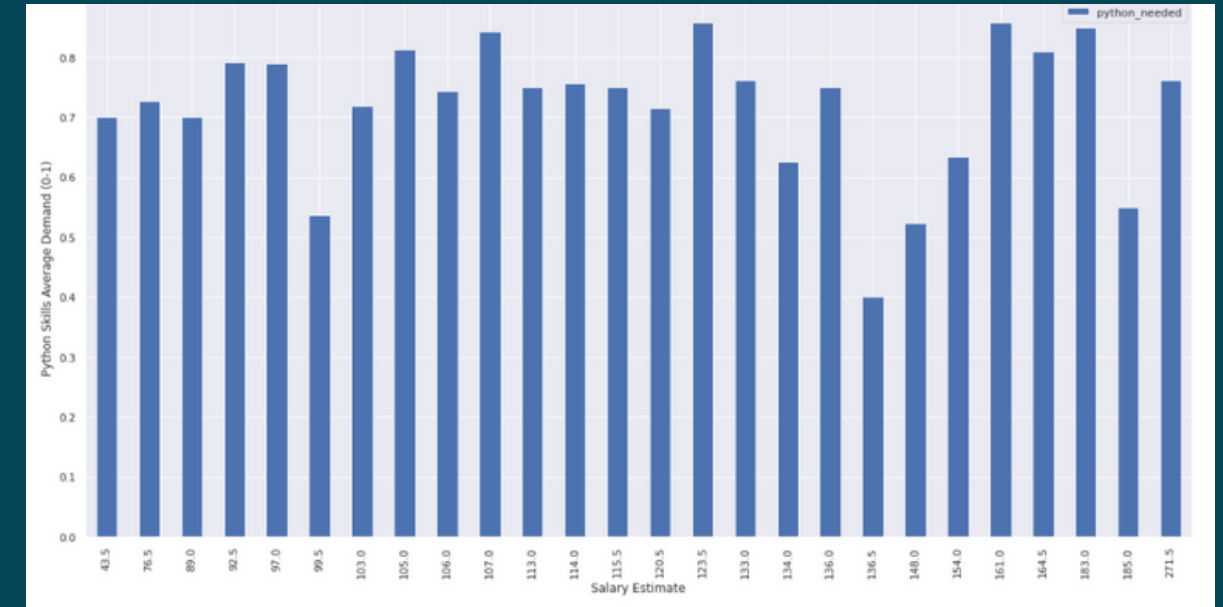


Almost all job types need **Python skill**



with **data scientist** become the job that need many/varied skills

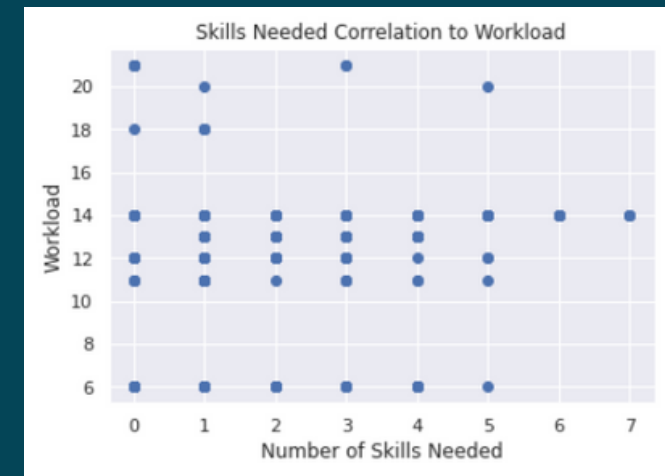
Relationship between *Salary Estimate* and *Python Skill Demand*



However, even though Python become the most needed skill, it doesn't mean you will get high salary if you master Python.

High Salary depends on the job, not because you master Python

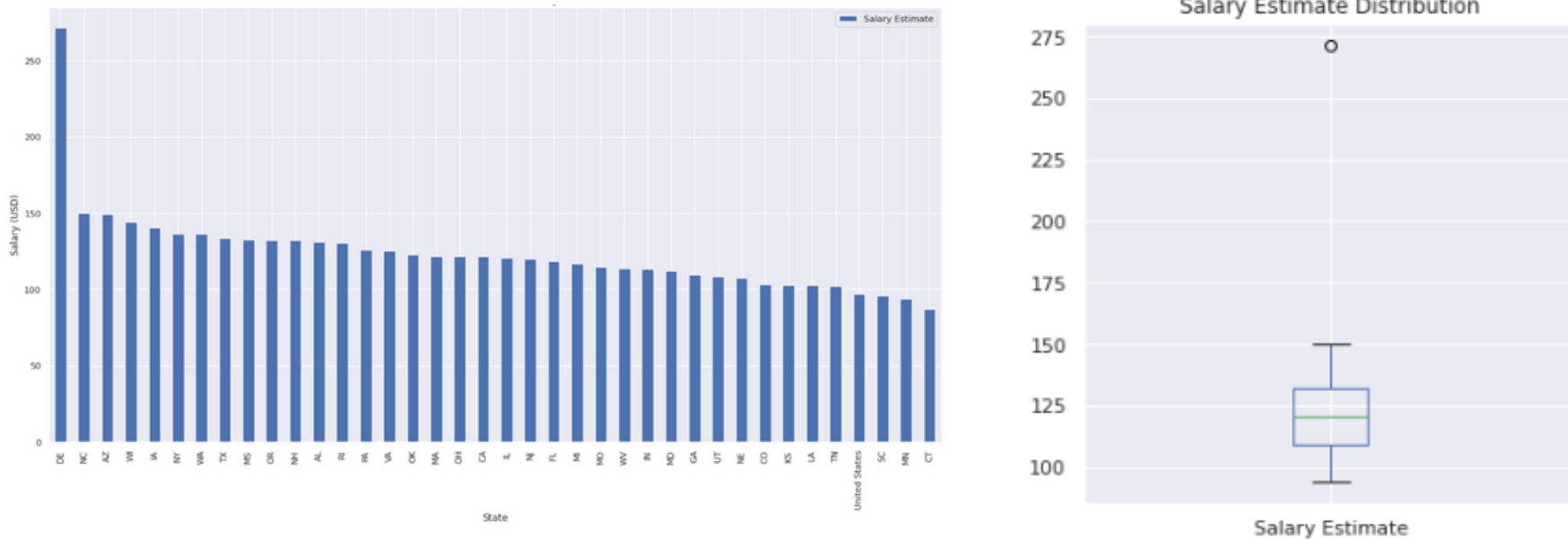
Relationship between *Salary Estimate*, *Workload* and *Skill Demand*



Number of skills needed also have no relationship with how many salary and workload you will get

# Relation between Salary at Each State

State Distribution Based on Salary



Salary for each state has a fairly good distribution with only one outlier with significantly high. It means that each state has almost same salary standard, except DE.

Biggest Salary **DE**      Lowest Salary **1st CT** **2nd MN**

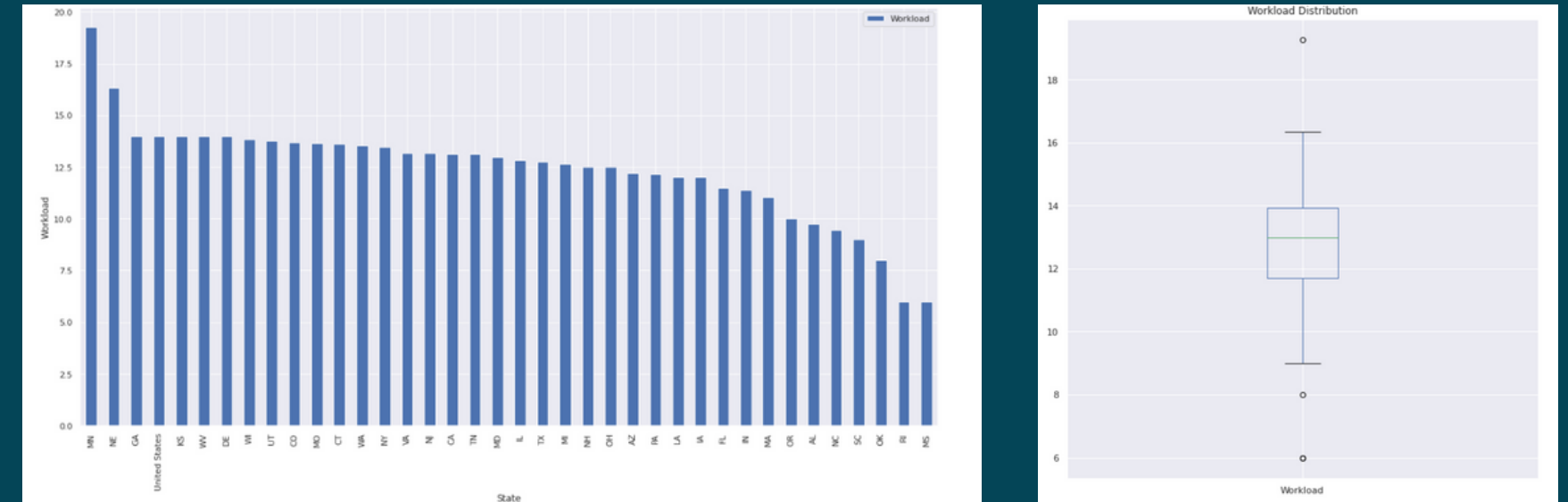
Salary Estimate Correlation to Workload



High Work Load doesn't mean high salary, as you can see from the graph there is a plot where even though 1 plot has only 6 work load, the salary reach up to more than \$250K.

# Relation between Work Load at Each State

State Distribution Based on Workload



Compared to salary distribution, workload distribution has worse distribution, with several outliers that is significantly high and low

Biggest Workload **MN**      Lowest Workload **MS**

One of the unique cases can also be seen when comparing the distribution of salary and workload that State MN has the largest workload, but the salary given is the second smallest.



It can also be proven if we develop relationship which is grouped by States which shows the regression results that **the larger the workload, the smaller the salary**. Although R-squared is not very large, this indicates that **some states provide salaries that are not commensurate with the workload**.

# State with High Demand of Jobs

From previous explanation, there are 5 state that has high demand of Job that deal with data problem as shown in map that is mostly in *the eastern of United States*. Then, Data Scientist become the most Job Type Vacancy with 455 demands.

However, from what we found high demand of jobs does not mean high salary. Instead in those states the salary distribution is at the range of \$ 125k to \$ 150k

## Unique Points for Each State with High Demand

### California (CA)

- The most varied skill needed
- Workload above average
- Python needed skill above average
- Salary below average

### Virginia (VA)

- Second most varied skill needed
- Workload above average
- Salary above average

### Maryland (MD)

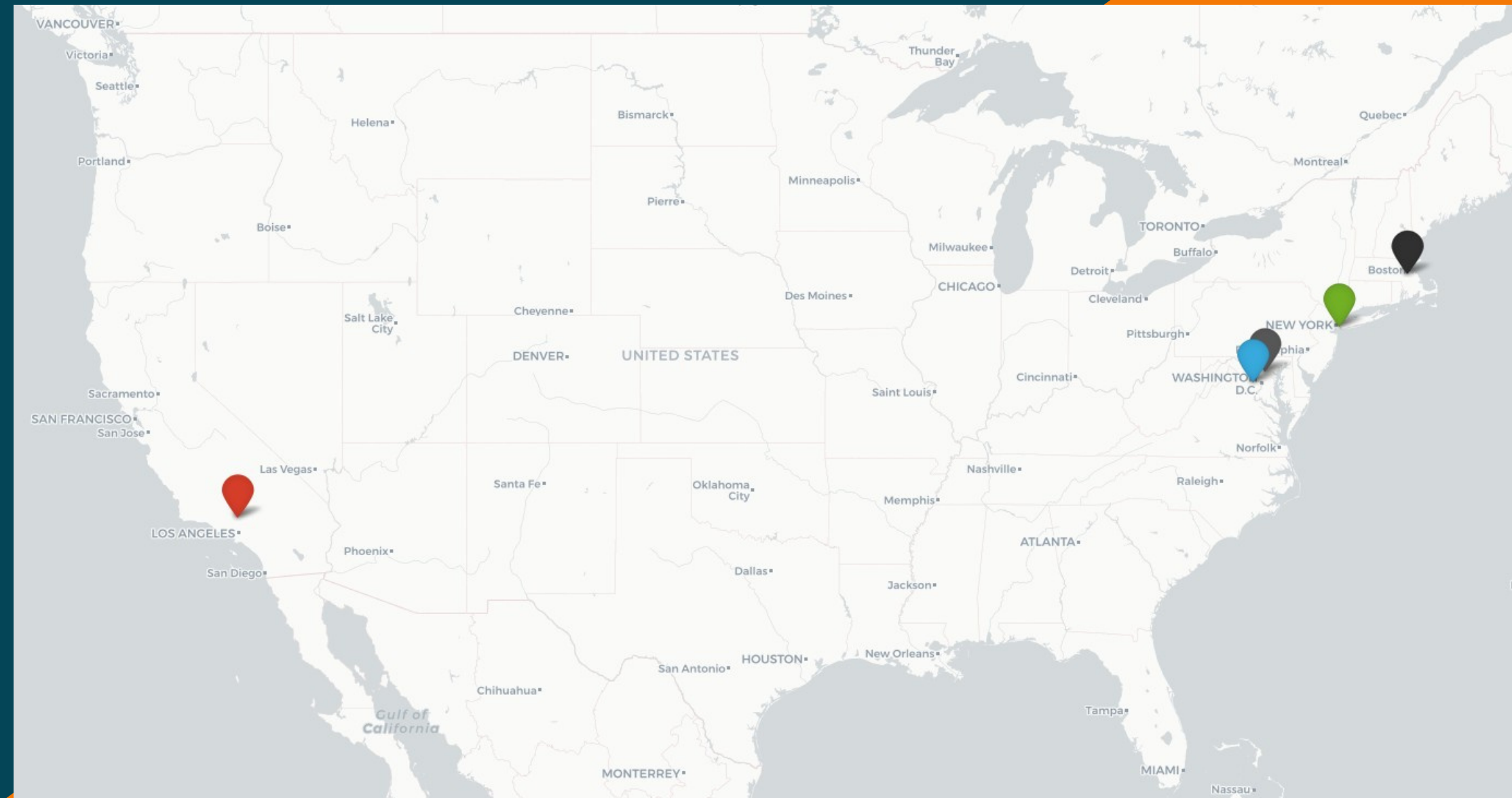
- Top 5 most spark needed skill
- Workload above average
- Salary below average

### New York (NY)

- Third most varied skill needed
- Workload above average
- Top 6 highest salary

### Massachusetts (MA)

- Top 4 most varied skill needed
- Workload above average
- Salary below average



 California  Virginia  Maryland  New York  Massachusetts

Those state are mostly located in east of USA

# CONCLUSION

---



In each state, **python is the most needed skill**. However, it doesn't guarantee you to get high salary.



In United States, most of the states provide **salaries which are not commensurate with the workload**.



Even though the **workload is high, it does not mean that the salary will be high**. This can be seen from the condition of MN and scatter plot in slide 3.



**California, Virginia, Maryland, New York, and Massachusetts** become the highest job demand and most varied skill needed

# RECOMMENDATION

---



It is better for new graduates in the United States **who want a career in the High Tech industry to prepare Python skill** before graduating.

**New York** become the most recommended state since it has high job demand in United State and provide high salary



---

**THANK YOU**

---