



Group Project

WORKING WITH MESSY DATA

Team Obelix

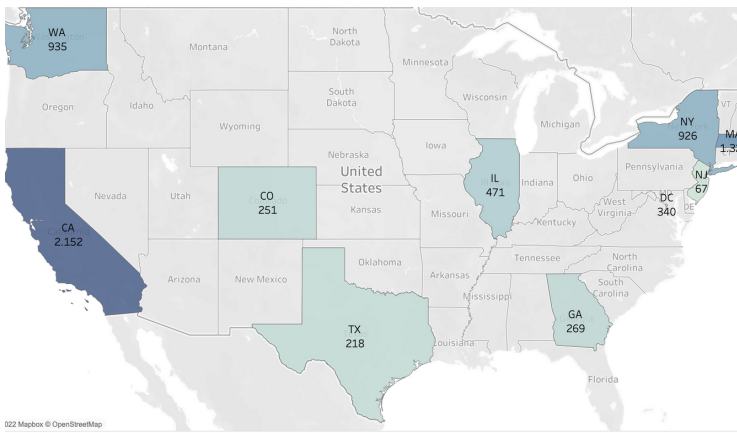
Josep | Lam | Angela

Data: job openings post from Indeed in the USA market during August 2018

Challenge: clean, wrangle and reshape the data to analyse and visualise



Which location has the most job opportunities?

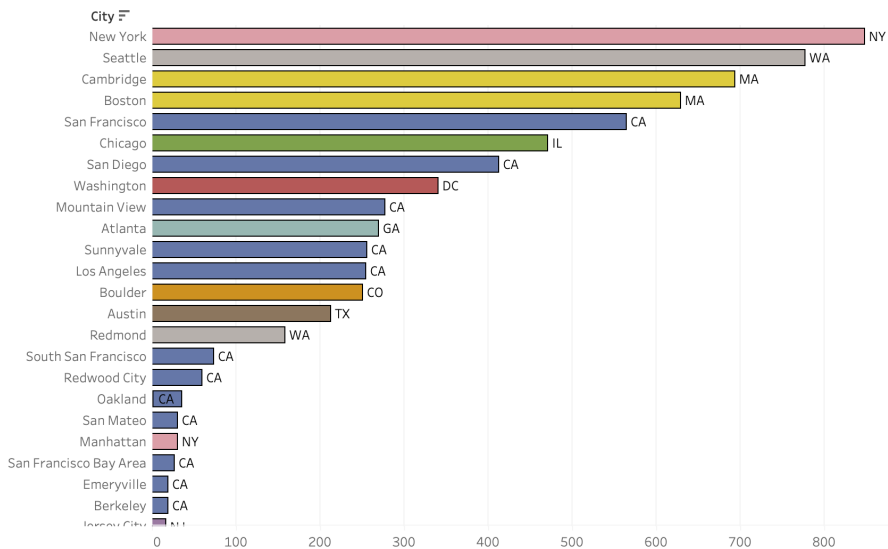


TOP 5 STATES

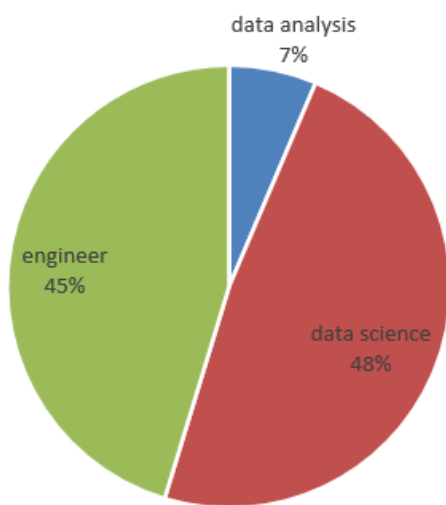
1. CALIFORNIA
2. MASSACHUSETTS
3. WASHINGTON
4. NEW YORK
5. NEW JERSEY

TOP 5 CITIES

1. NEW YORK
2. SEATTLE
3. CAMBRIDGE
4. BOSTON
5. SAN FRANCISCO



Who gets hired ???



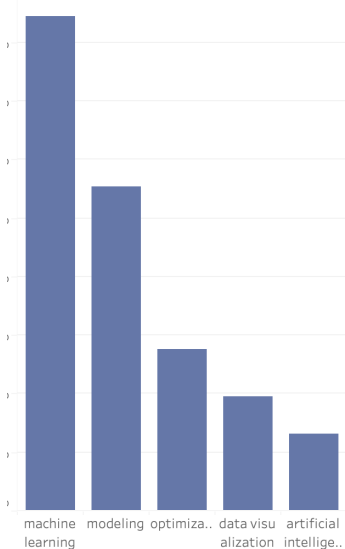
TOP TALENTS

1. DATA SCIENCE
2. ENGINEER
3. DATA ANALYSIS

The ideal data scientist

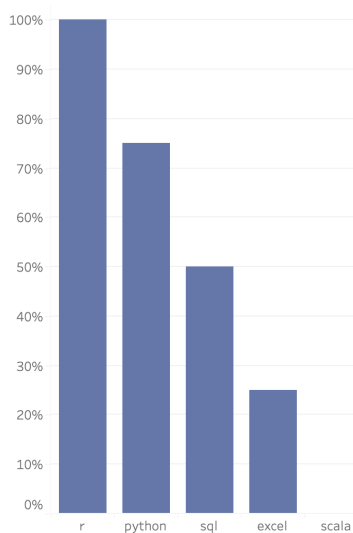
TOP 3 SKILLS

1. MACHINE LEARNING
2. MODELING
3. OPTIMIZATION



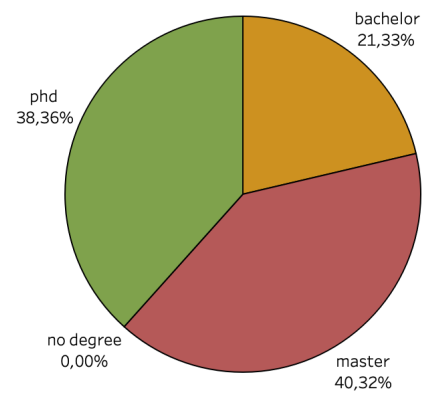
TOP 3 TOOLS

1. R
2. PYTHON
3. SQL



TOP 3 EDUCATION

1. MASTER
2. PHD
3. BACHELOR



METHOD

PYTHON

- Standardise and clean the data using string functions
- Apply map and for loops to check the True and False values for the defined keywords
- Counting true values to define subtotals.

TABLEAU

- Data aggregations and visualisations

HIGHLIGHTS

Python filter malfunctions, R

