

The Problems

#1

The data on the "description" column had relevant data with a lot of text.

We did try to approach it with Regex but with the time given we couldn't achieve it.

#2

The data on the "location" column had to be cleaned as some of the states had their pin codes alongside.

It was cleaned using excel.

#3

Majority of the "companies" just had 1 job posting and most of it was for Data Scientist / Research Data Scientist.

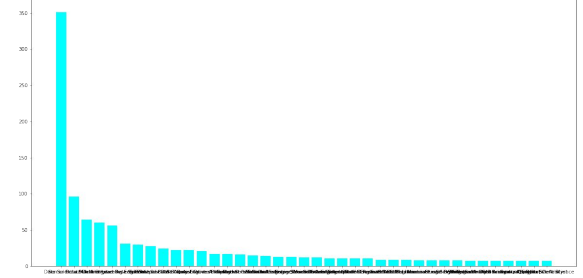
It was cleaned and removed companies that had job postings less than 15.

The Objectives

- Who gets hired?
- What kind of talent do employers want when they are hiring a data scientist?
- What skills, tools, degrees or majors do employers want the most for data scientists?
- Which location has the most opportunities?

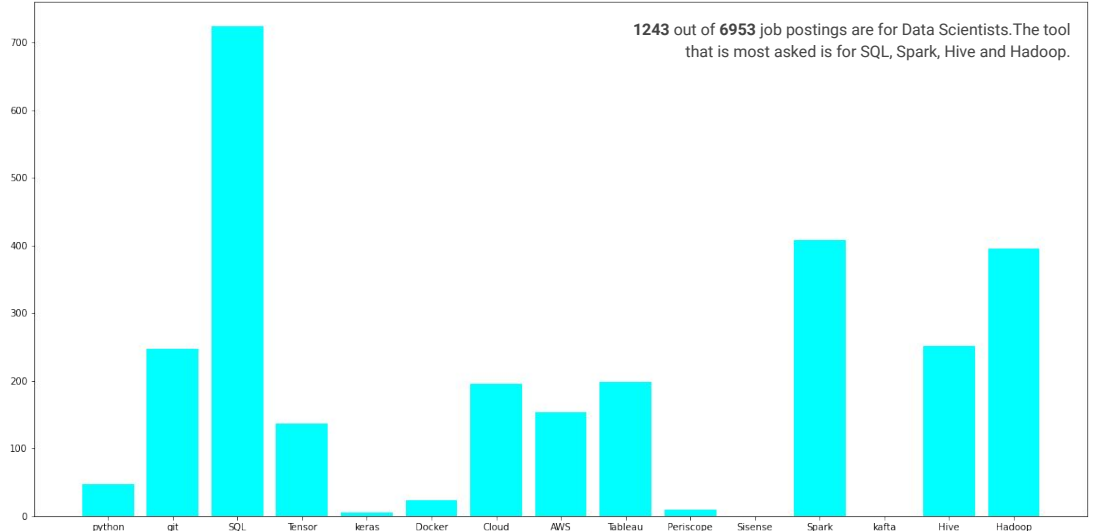
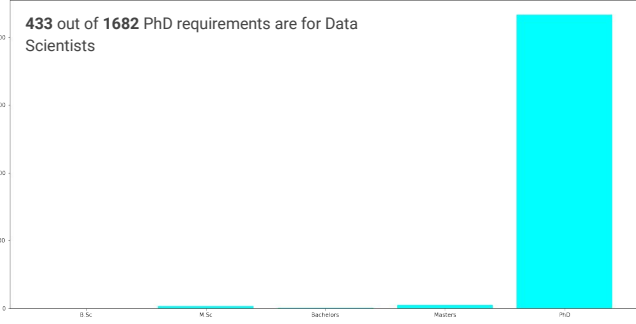
Who gets hired?

Data Scientists



What kind of talent do employers want when they are hiring a data scientist?

What skills, tools, degrees or majors do employers want the most for data scientists?



1243 out of **6953** job postings are for Data Scientists. The tool that is most asked for is SQL, Spark, Hive and Hadoop.