You live for crunching data?

Kaggle DS&ML Survey 2019

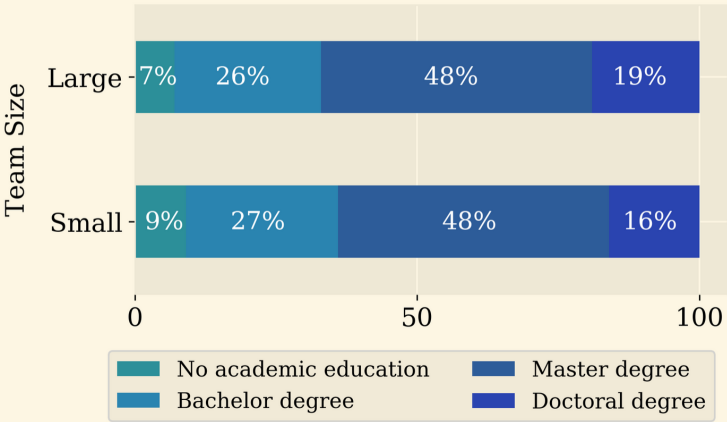### HERE'S WHAT YOU NEED TO KNOW ABOUT THE SIZE OF YOUR TEAM !

There are two sides regarding the size of the Data Science teams. Small companies are great environment to share ideas, knowledge, be flexible and probably to have better working conditions. On the other hand, big teams may work with bigger projects and there is always the right person for a new task. In this analysis, offered to you by Oculart, you will know more about the differences between small and big data teams. For this purpose, we will review the latest **Kaggle** survey from 11743 professional who work with data. From this survey 5324 responders work in team of 2 to 4 members and 6419 responders work to team of more than 4 members. The data were collected in 2019 from users all over the world.

> **Kaggle is a Data Science community where data scientist can publish datasets and build machine learning models. It is also a platform for data scientists to work with others to solve data science challenges.**
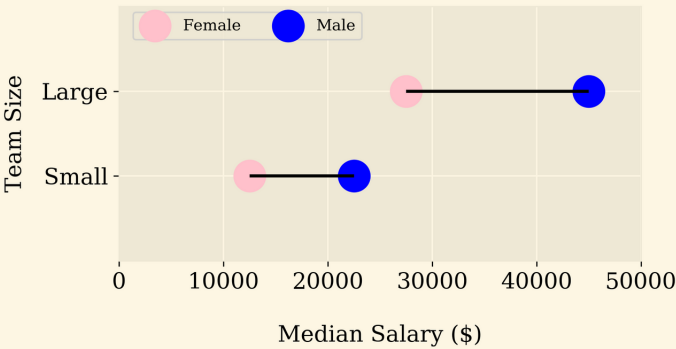
EDUCATION LEVEL

It is true that data teams may come with people from different academic backgrounds. Nowadays, the variety of Massive Open Online Courses (MooC) allows anyone to gain data analytical skills without the need to pursue postgraduate (and probably expensive) studies. From our analysis it seems that still the majority of data professionals have at least a master degree. However, there is no great difference in educational level across small and large teams.

**Employees of different education levels across teams of different size**

GENDER PAY GAP

Gender pay gap is now a hot topic. It is widely acknowledged that gender pay gap exists in many industries, so how about data related roles? As we can see in the plot, female data specialists tend to have a lower salary compared to male in both team size groups; however, this distance gets wider for respondents in large groups. It is clear that the gender pay gap it is greater in large teams, indicating that gender inequality might be more widespread in large data teams.

Gender-pay gap across different data science teams

WORKING EXPERIENCE IN ML

Many people might be also interested in the actual pay gap across as their career grows. In general, the median salary increases as years of working experience in Machine Learning increase in both groups. Moreover, the median salary in large teams outperforms that in small teams (i.e. around 10.000$ pay gap in 4 years of experience and 27.500$ pay gap in more than 20 years of experience).
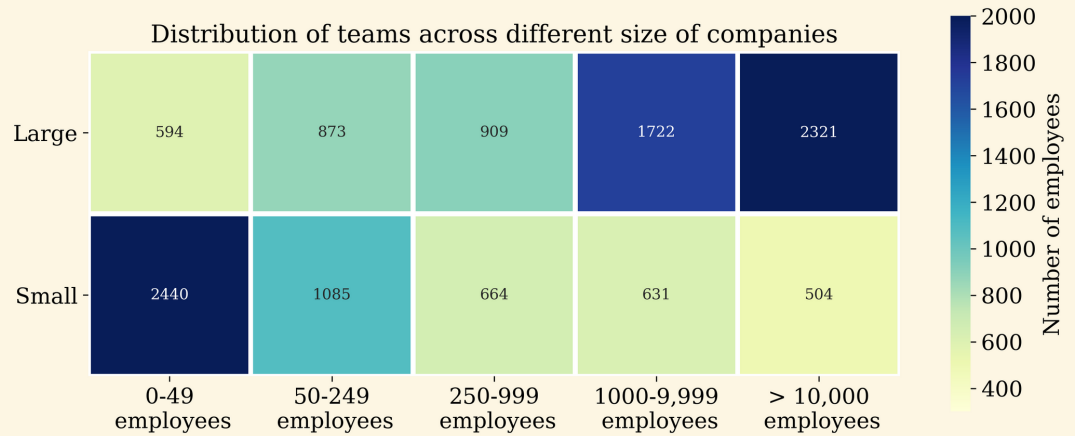
Comparison of Pay gaps in different size of teams
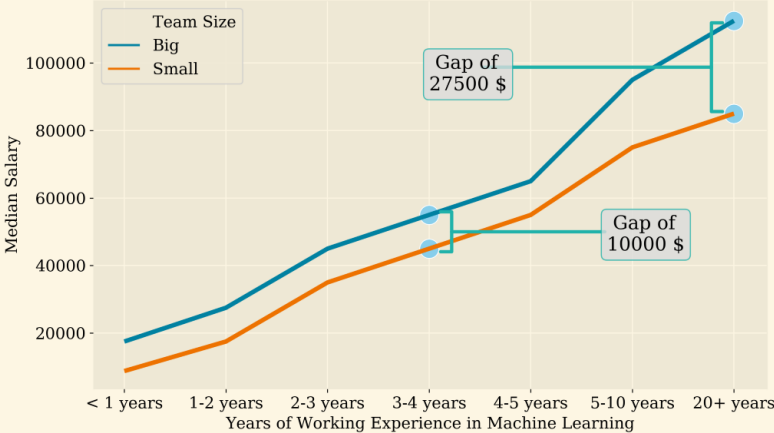
COMPANY SIZE

Does the size of a company relate with the size of a data team? Nowadays, more companies trying to exploit the potentials of their data, however not all of the companies have sufficiently large data teams. This heatmap explores the size of data teams for different size of companies. What is interesting is that large data teams may exist also in small teams (probably the most data-oriented ones). On the contrary, we can still find big companies with relatively small data teams (less than 4 people). In the second case, probably the companies are still developing their data teams, or their industry cannot rely to a large extend on data-driven decisions.

Click here to see interactive visualizations!

Distribution of teams across different size of companies

|	0-49 employees	50-249 employees	250-999 employees	1000-9,999 employees	> 10,000 employees
Large	594	873	909	1722	2321
Small	2440	1085	664	631	504

VISUALIZATION TOOL

Visualisation tools play a pivotal role into delivering knowledge to a company. The following plot explores employees' preferences in different visualisation tools. Both groups have almost the same preference of visualization tools. Moreover, it seems that Matplotlib, Seaborn (Python) and ggplot2 (R) are the top three visualization tools used by experts.

Preferences of DS tools in different group size
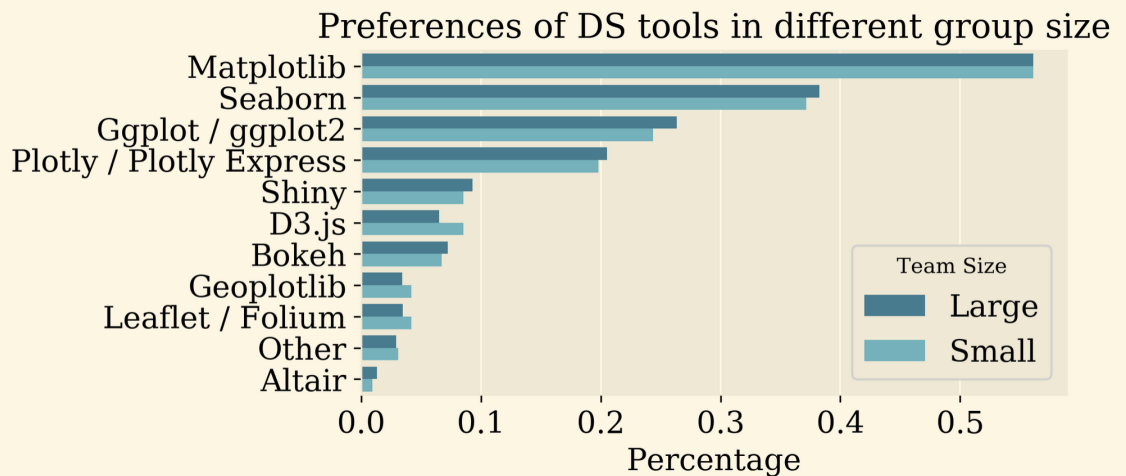
CLOUD TOOLS

But what about cloud products? Below we can find the percent of experts who use different cloud frameworks in small and large teams. Overall, it seems that Google BigQuery it's the most popular framework for both small and large teams (7.5%). Large teams show in general a greater preference to many of the cloud products. However, Google Cloud Dataflow and Microsoft Analysis Services are more popular on small teams. Finally, AWS Athena, AWS Kinesis and Google Cloud Pub/Sub have nearly the same popularity for both sizes of team.

Percentage of responders who use different cloud products