­

SMM635 — Data Visualisation

Final course project submission template

# Visualisation #1

## Chart Description automatically generated

The dataset surveys 9,094 professionals around the world working in the Data Science and Machine Learning (ML) field, conducted by Kaggle in 2022. It contains information on their background, technologies, and techniques.

This visualisation displays the distribution of gender and title distribution in each industry. The bar charts on the top and right side of the visualisation shows the number of professionals in each industry and job title, respectively, broken down by gender. The center grid illustrates the intersection, with the size of the pie charts scaled according to the count of the professionals, also broken down by gender.

The chart suggests that the most popular job titles are data scientists and data analysts, globally and for the majority of the industries. Unsurprisingly, in Academics/education, the most popular titles are teacher/professor and research scientist, while on the other hand, software engineers and data scientists are most popular in computer/technology.

Interestingly, the visualisation suggests that all fields and professions are dominated by male. The least male-dominated industries are Academic/Education and Medical/Pharmaceutical with the female proportion slightly exceeding 25%. This is consistent with the common knowledge that STEM is mostly a male-dominated field.

# Visualisation #2

Chart

Description automatically generated with medium confidence

This visualisation displays the state of Machine Learning (ML) adoption for each country and industry, calculated by averaging the ML adoption score (0 for not utilising ML and 4 for having used ML in production for more than 2 years) given by the respondents in each category.

In the first chart, the popularity of ML in each country was calculated by dividing the number of respondents by the country population. It indicates that ML is more popular and has high level of adoption in the USA, Europe and developed Asian countries such as Japan and Singapore. In contrast, most countries in Africa and Latin America have not adopted ML or still in the exploration phase and with low level of ML popularity. Interestingly, in the United Arab Emirates, ML adoption is still immature despite its popularity.

The second chart indicates significant variation in ML adoption by industry and company size. For instance, adoption in Academics/Education is lowest while Insurance/Risk assessment and Online/Internet service is highest. Generally, across all industries, bigger companies tend to adopt ML more than smaller ones. However, in Academics/Education and Shipping/ Transportation industries, small and medium companies are leading in the adoption process.

# Visualisation #3

Table

Description automatically generated with low confidence

## Chart Description automatically generated

**TODO: fix education legend**

The visualisation illustrates the academic background of respondents based on their roles. Education levels by role are displayed via the first stacked bar chart. In all positions, more than 80% of the professionals have at least a Bachelor’s Degree. The groups of Developer Advocate, Data Administrator, and Data Engineers have around 10 to 20% people with no degrees, which are higher than in other groups being less than 10%. Notably, for Data Scientists, Research Scientists, Statisticians, and Teachers/Professors, these numbers are around and under 5%, while the proportions of people with a Doctoral Degress are highest among all positions which are 17.99%, 53.97%, 23.01%, and 51.59% respectively.

The rates of professionals who involved Machine Learning in their published research are demonstrated via the parallel coordinated chart. Research Scientists, Teacher/Professors, ML/ MLops Engineer, Data Scientist have the highest rates of professionals with ML utilised in theoretical reacher which are around 29%, 25%, 19% and 15% respectively. These are also the positions with the highest rates of people with ML in empirical research, being 51%, 41%, 31% and 22% respectively. ML is used in applied research more than theoretical research. Engineers (non software) have the highest rates of people who have not used ML at 31%. Teacher/ Professor and Research Scientist also have a high rates of research without ML involvement being around 25%.

# Visualisation #4

Chart, scatter chart, bubble chart

Description automatically generated

This visualisation shows the average salary in USD by job title, coding experience in each industry, with the sizes of the bubbles scaled according to the number of respondents.

The global average coding experience and salary of the respondents are approximately 7.5 years and 63,000 USD respectively. The variation in experience and salary by title can be clearly noted as the titles are grouped into clusters. On average, data analysts have less than 5 years of experience and below average salary, while software engineers have noticeably more experience but not significantly higher salary. The extreme can be seen in software engineers working in the Government/Public service industry, with the most experience but below average pay. Data scientists and ML engineers all have close to average experience and pay.

Interestingly, managers on average have higher salary across almost all industry regardless of experience. This makes sense as managers also perform leadership duties besides technical contribution, therefore technical experience is not always strictly required.

Finally, research scientists appear to be outliers, with extremely high salaries in the marketing, internet service and finance industries, but earn the least in the academic/education industry.

# Visualisation #5

## Chart, radar chart Description automatically generated

Chart, radar chart

Description automatically generated

Chart, radar chart

Description automatically generated

## Companion description (max 200 words)