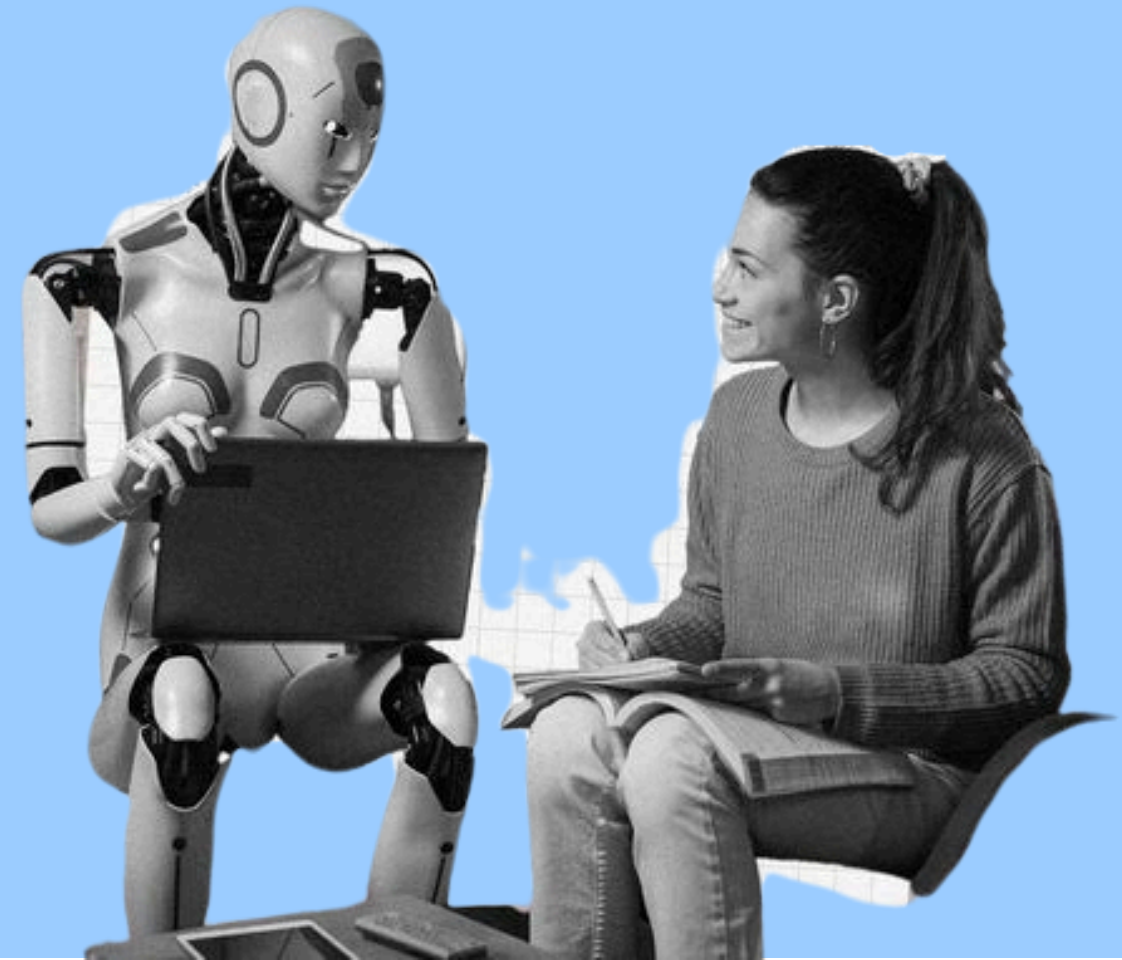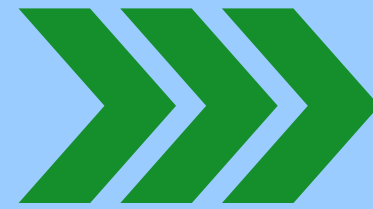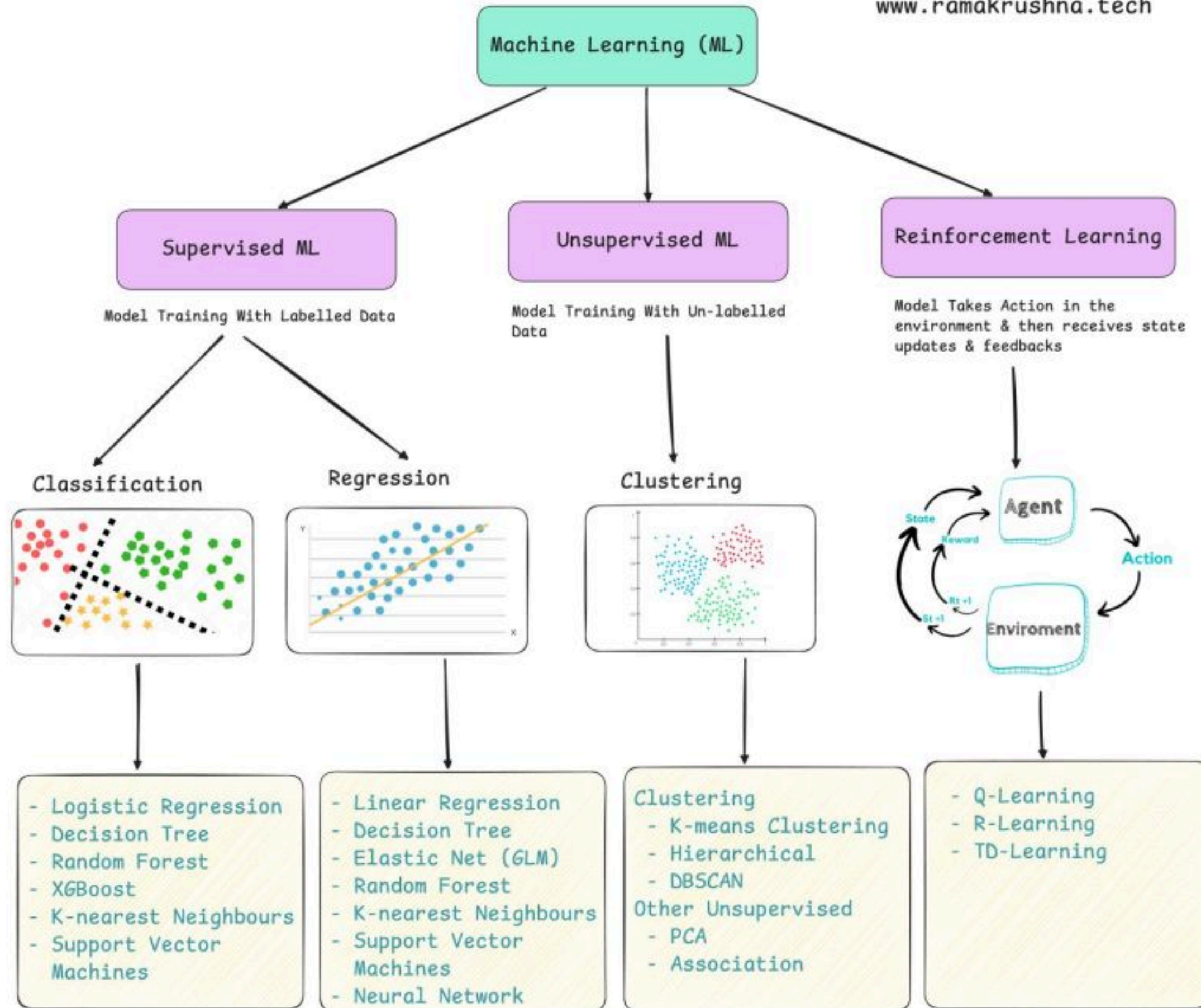# THE DATA WORLD

A GUIDE TO START YOUR CAREER JOURNEY

▲ 12 %
▼ 5 %

IT IS THE FUTURE OF DATA..

"The use of data dates back millions of years, shaping the progress of ancient civilizations from tracking the stars to building the pyramids. Today, with the rise of AI and advanced technologies, data is more powerful than ever. The future belongs to those who can understand, analyze, and harness it
THE DATA ARCHITECTS OF TOMORROW!"

Machine Learning (ML) taxonomy diagram

www.ramakrushna.tech

Machine Learning (ML)

- **Supervised ML** — Model Training With Labelled Data
  - Classification
    - Logistic Regression
    - Decision Tree
    - Random Forest
    - XGBoost
    - K-nearest Neighbours
    - Support Vector Machines
  - Regression
    - Linear Regression
    - Decision Tree
    - Elastic Net (GLM)
    - Random Forest
    - K-nearest Neighbours
    - Support Vector Machines
    - Neural Network
- **Unsupervised ML** — Model Training With Un-labelled Data
  - Clustering
    - Clustering
      - K-means Clustering
      - Hierarchical
      - DBSCAN
    - Other Unsupervised
      - PCA
      - Association
- **Reinforcement Learning** — Model Takes Action in the environment & then receives state updates & feedbacks
  - Agent — State, Reward, Action, Environment (Rt+1, St+1)
    - Q-Learning
    - R-Learning
    - TD-Learning

IT IS NOT ABOUT AUTOMATION, IT IS ABOUT THE INDIVIDUALS WHO AUTOMATE IT!

From 09/01/2021 to 10/04/2021, Kaggle conducted its 5th annual industry-wide survey in providing a truly comprehensive view of the state of data science and machine learning and related feilds.

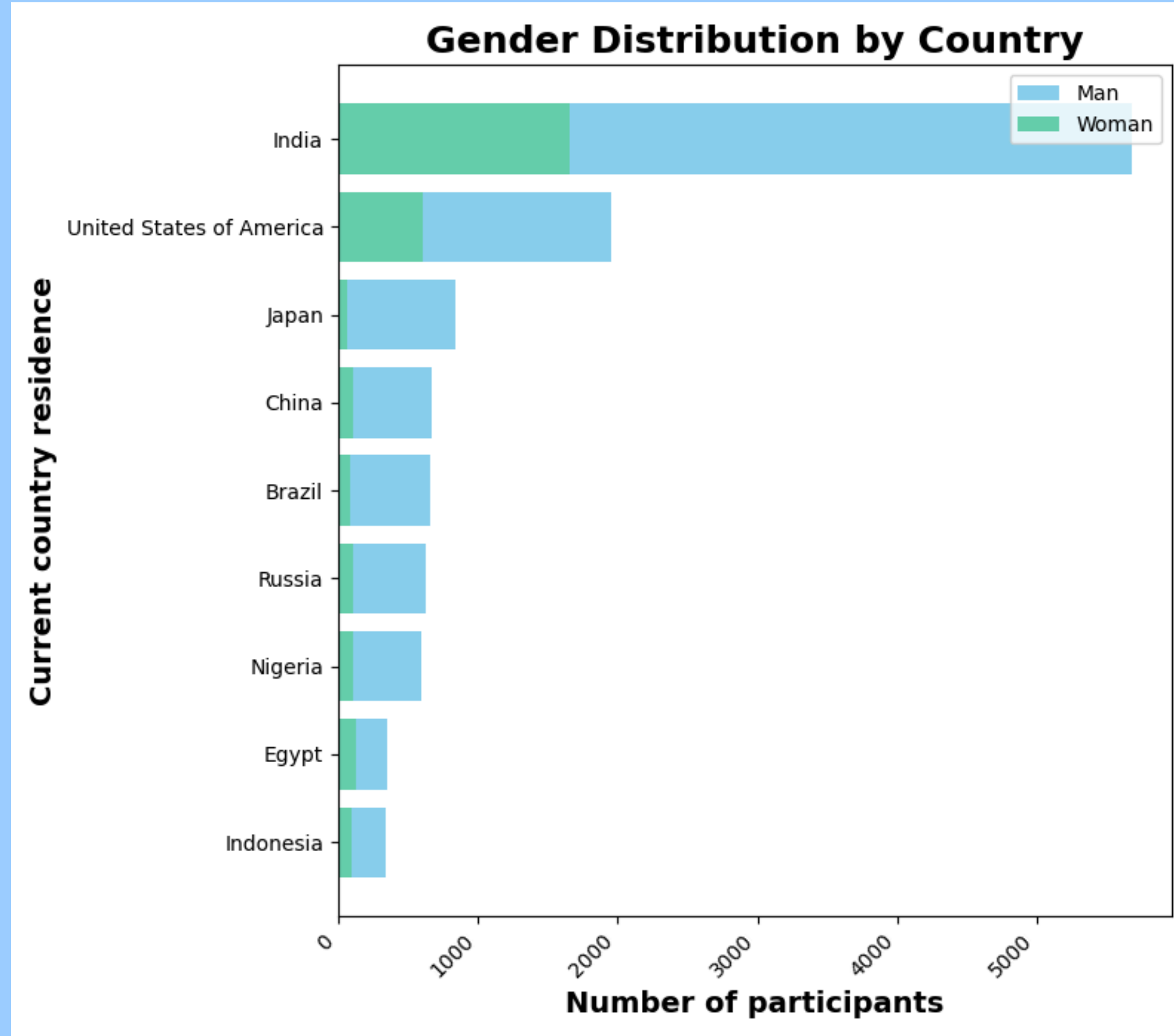# CONTENTS

KEY FIGURES

25,974

NUMBER OF RESPONDENTS

43

NUMBER OF QUESTIONS IN THE SURVEY

1/3

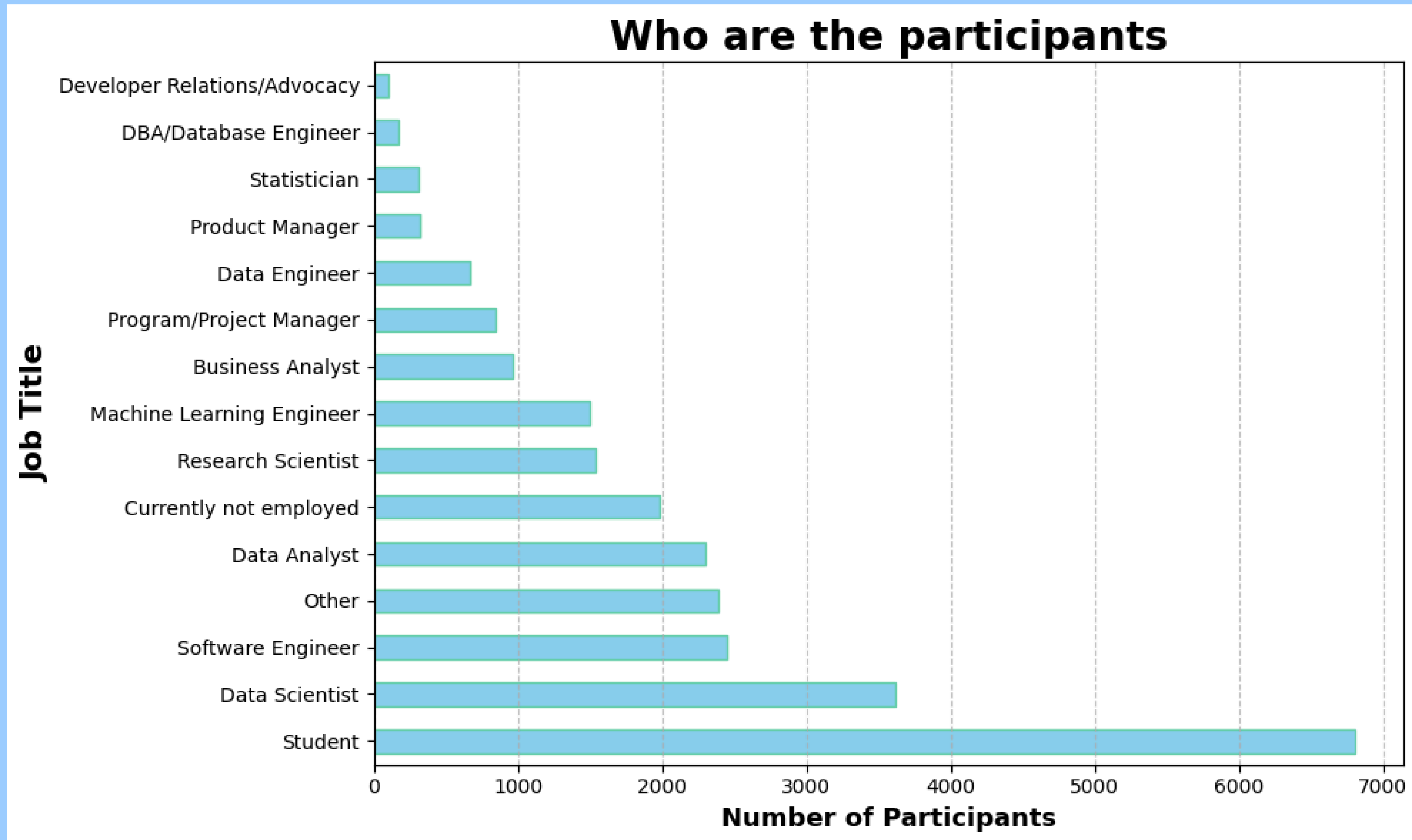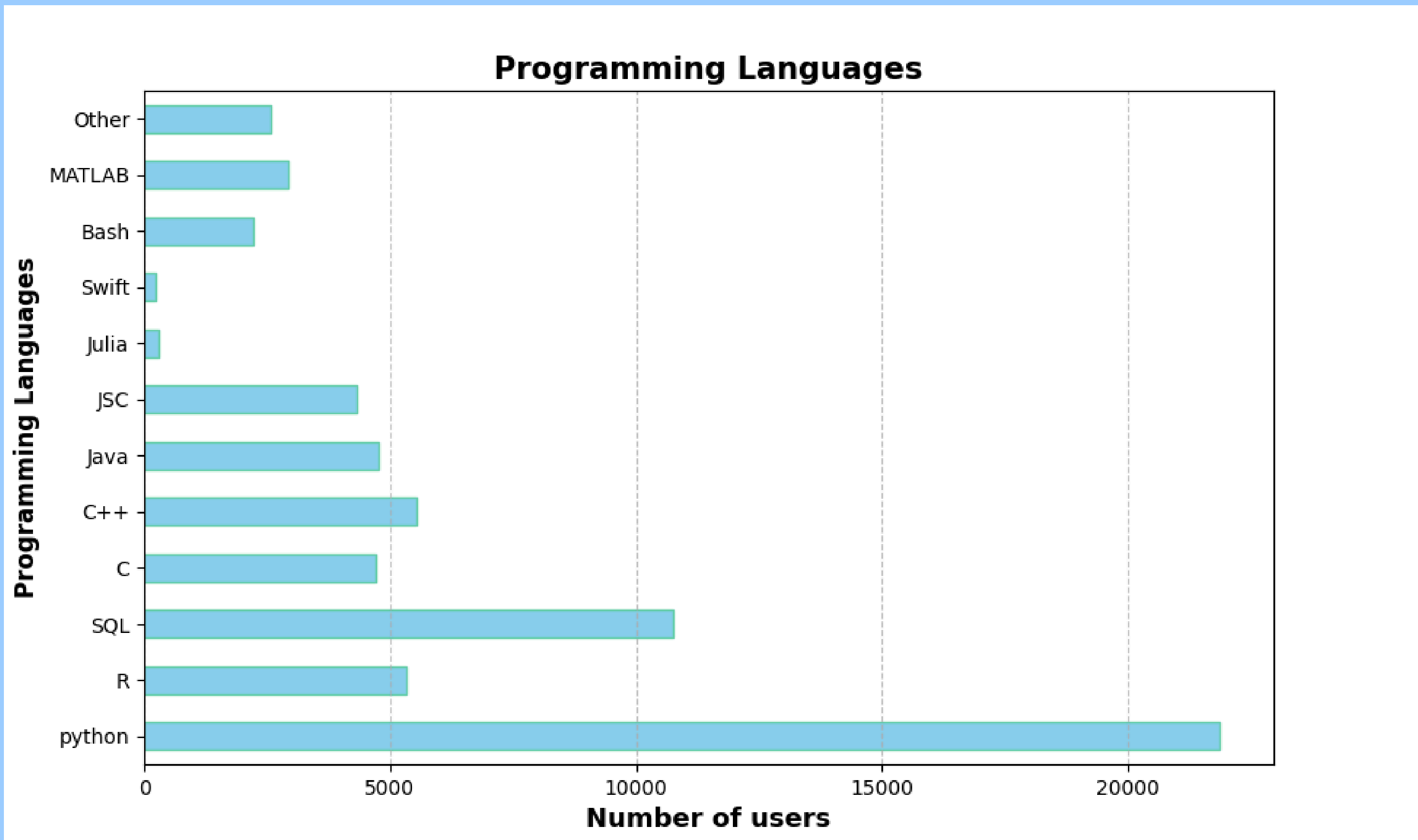**OF PARTICIPANT ARE DATA SCIENTIST AND MACHINE LEARNING ENGINEERS**

Gender Distribution by Country

Who are the participants

01.

Programming Languages

01.

Most Used Libraries for Data Visualization
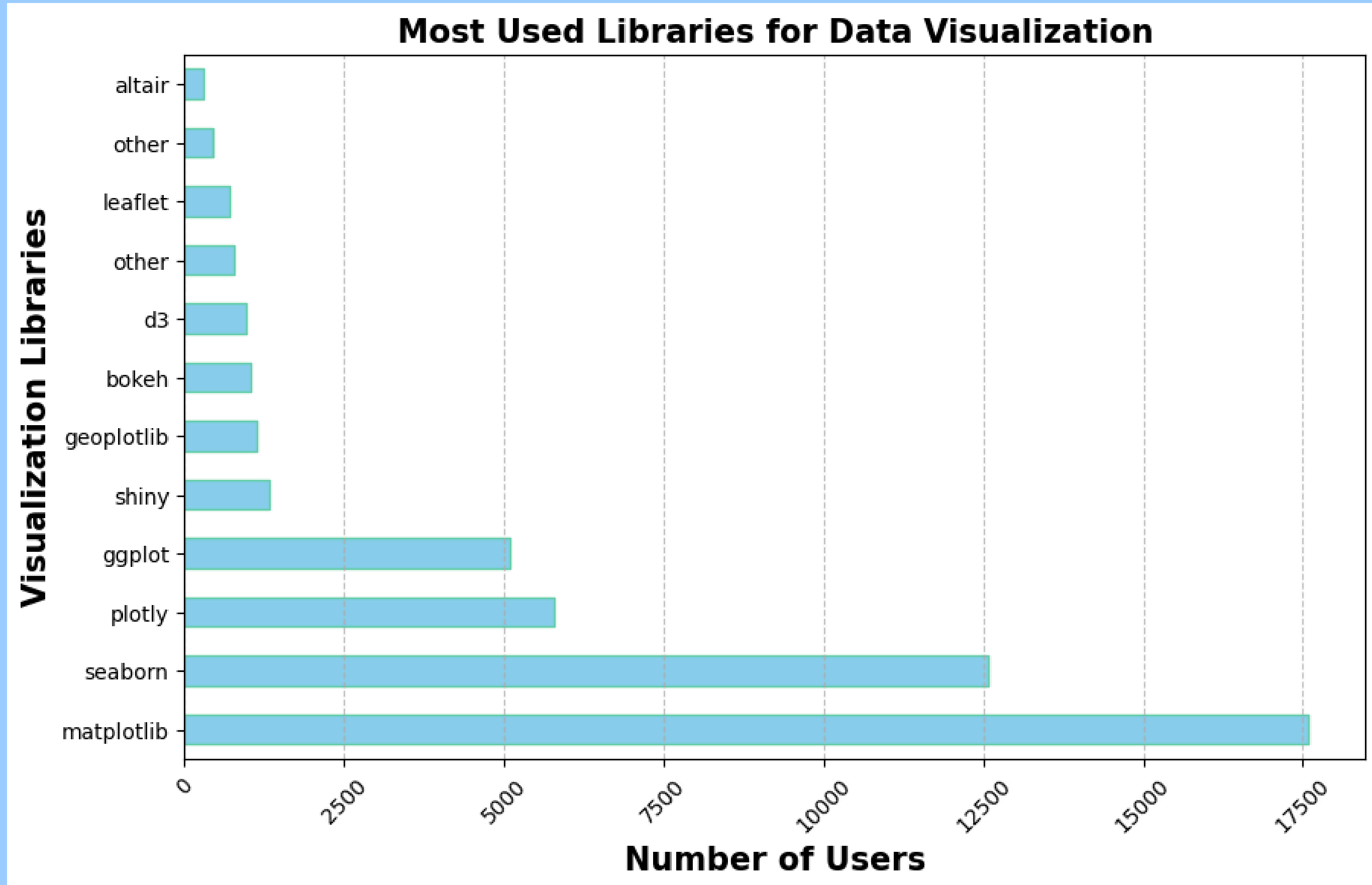
2.1 HOW MANY YEARS OF EXPERIENCE ARE REQUIRED FOR EACH JOB TITLE🎯
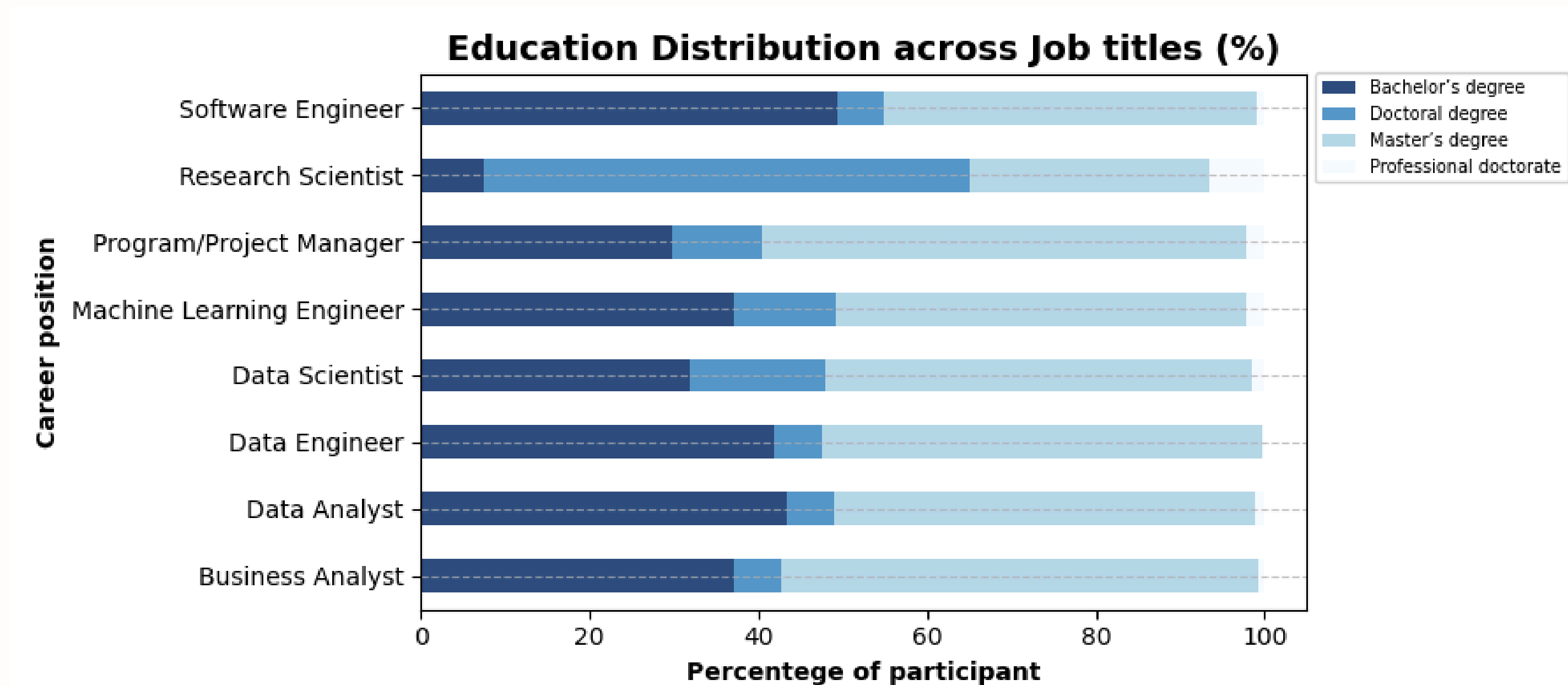
Coding experience Vs Job title
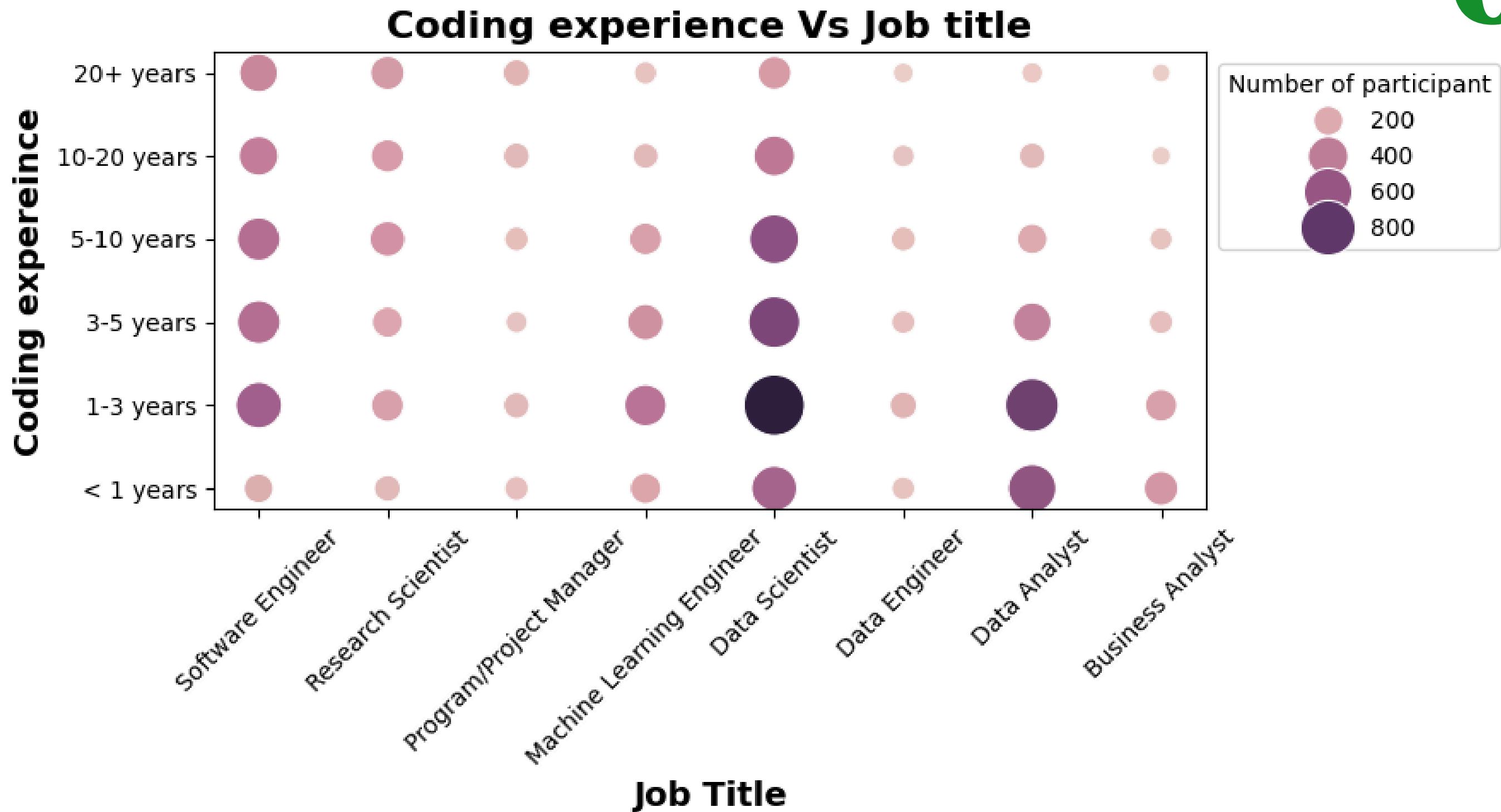
The majority of respondents have 1-3 year of experience,💡.
- Data science particpant's experience peaks at 3-10 years of coding experience
- Most of the Software Engineer participants are professionals in coding..

02.

Education Distribution across Job titles (%)

Legend:
- Bachelor's degree
- Doctoral degree
- Master's degree
- Professional doctorate

- Seems like becoming a machine learning engineer or a data scientist does not necessarily require a Master degree!
- Data engineer, Business analyst, and project manager jobs have large number of participants with masters degree, more than the bachelor's..
- If you want to become a research scientist, large opportunities are for ones who earned their doctorate!

02.

Yearly income Distribution Across Job Titles (%)

| Job Title	$0-999	1,000-1,999	10,000-14,999	15,000-19,999	100,000-124,999	125,000-149,999	150,000-199,999	2,000-2,999	20,000-24,999	200,000-249,999
Business Analyst	30.3	13.2	14.8	8.7	8.9	4.6	1.8	8.4	8.4	1.0
Data Analyst	42.7	15.4	12.7	6.9	6.6	1.3	0.9	6.6	6.6	0.4
Data Engineer	29.4	13.5	10.2	7.3	9.6	5.9	6.6	8.3	7.9	1.3
Data Scientist	36.4	11.0	8.4	6.0	10.4	6.2	7.1	5.2	5.9	2.9
Machine Learning Engineer	49.4	11.7	8.7	4.4	5.5	3.9	3.6	6.5	3.5	2.1
Program/Project Manager	22.3	9.2	10.3	7.5	15.1	7.8	8.1	6.1	8.1	4.7
Research Scientist	34.6	11.3	9.6	10.0	7.2	3.3	4.1	7.2	10.4	2.3
Software Engineer	37.9	11.0	13.8	7.3	6.2	4.5	3.4	7.4	7.3	1.0

Yearly income Range

Machine Learning Engineers tend to have the highest percentage of employees in lower salary ranges! This suggests that entry-level roles in these fields may be more common. 💼📊

- Product Managers show a relatively higher percentage in the mid salary ranges, indicating that this role may have fewer entry-level positions and a higher earning potential over time. Similarly, Data Scientists and Software Engineers exhibit a more balanced distribution across salary ranges.

# 2.4 HOW DO PROGRAMMING LANGUAGES PREFERENCES VARY BASED ON JOB TITLES🤔?

Preferred programming languages according to job titles
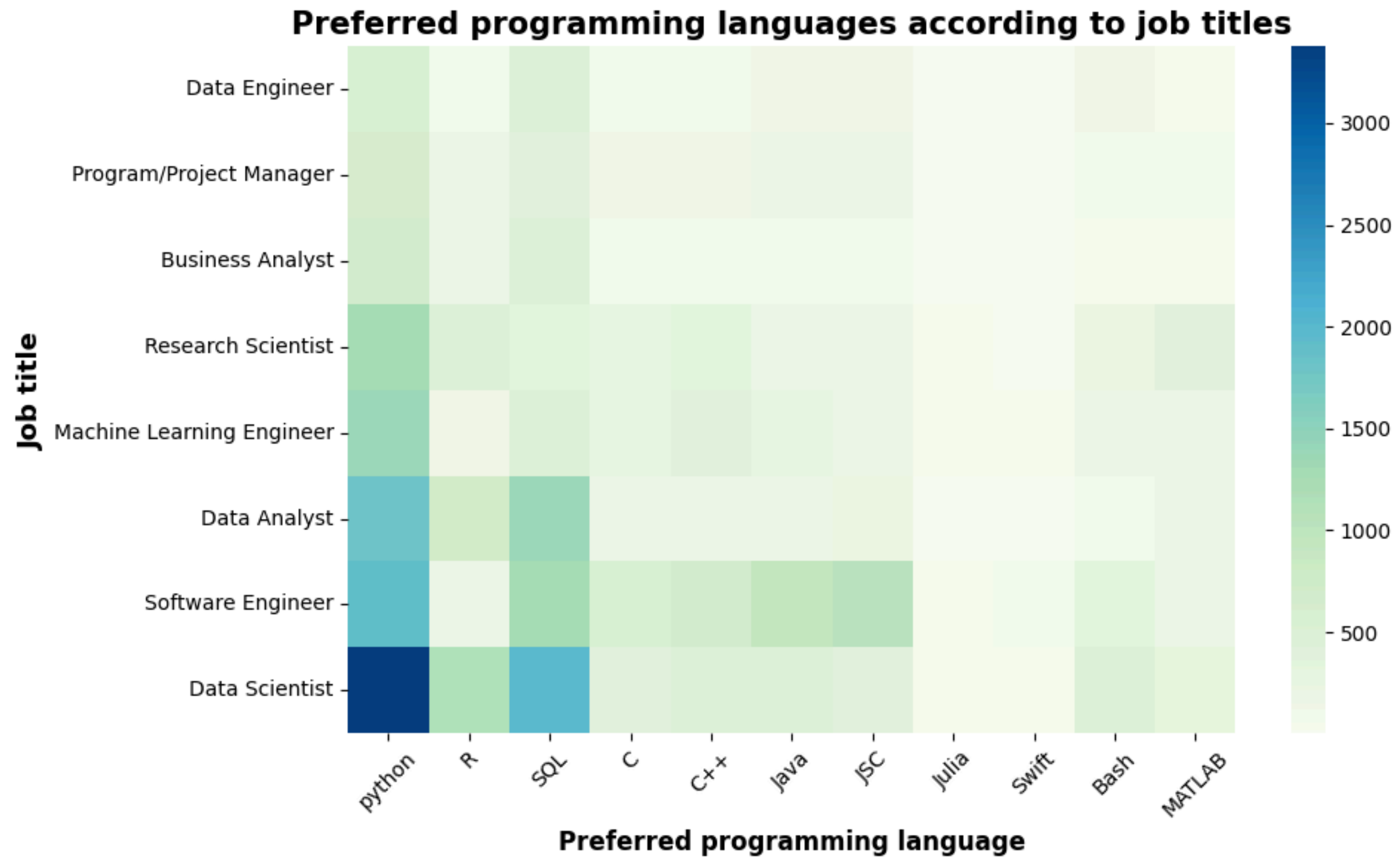
1. Python users are mostly (Students, Data Scientists, and Software Engineers).
2. SQL is mostly used by (Students, Data Scientists, and Data Analysts).
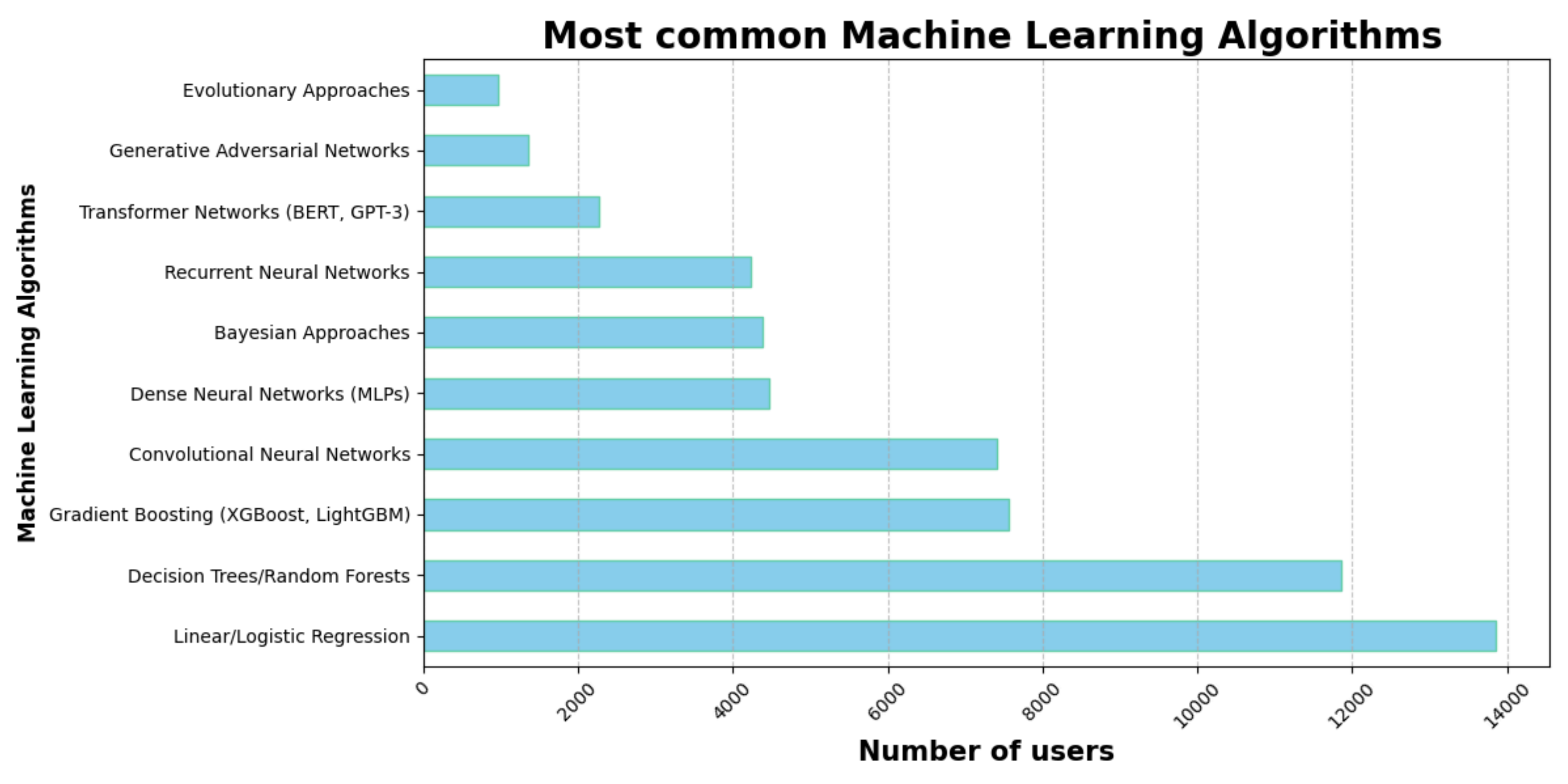3. C++ is also used mostly by Students.

Machine Learning frameworks used based on experience

3.2 MACHINE LEARNING ALGORITHMS
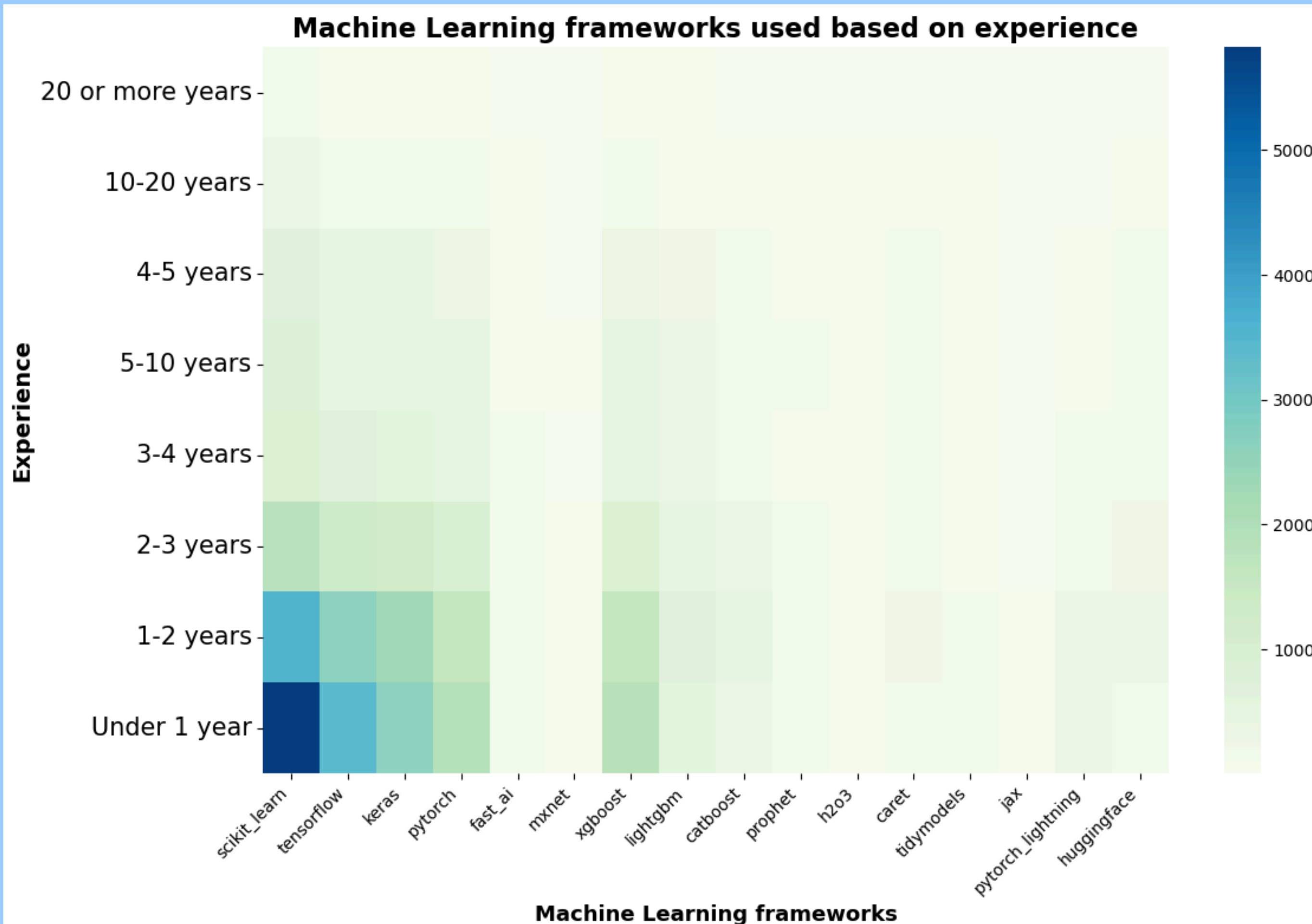
Most common Machine Learning Algorithms

CONCLUSION: REMARKS

- Our analysis provide an overview of the state of data science, machine learning feild in the year 2021.

- There is a huge gender gap with ratio of 20:80, where most of the participant are from US and india.
- Python users are mostly (Students, Data Scientists, and Software Engineers).

- SQL is mostly used by (Students, Data Scientists, and Data Analysts)

- Linear/ Regression is the mother of the Machine learning algorithms.
- The majority of respondents have 1-3 year of coding experience. Participants of different background experience vary from profession in coding to beginners.

FUTURE WORKS

adding more data from other surveys or reports to compare trends over time.

people's opinions on the frameworks and languages they use

analysis on specific job titles or exploring a specific subcategory