



Data Science & Big Data

A part-time course to train the new generation of data scientists.

DataScience@UB



DATA SCIENCE
PROGRAMS AND
COURSES AT
UNIVERSITY OF
BARCELONA



MASTER
OF
FOUNDATIONS
OF DATA
SCIENCE



DEEP
LEARNING
SUMMER
COURSE



DATA
SCIENCE
AND BIG
DATA
COURSE

jordi.vitria@ub.edu

Departament de Matemàtiques i Informàtica de la Universitat de Barcelona



Since 2007, I am a Full Professor at the Mathematics & Computer Science Department, **Universitat de Barcelona**, where I teach an introduction to Algorithms, Machine Learning and Data Science. Before that I spent 20 years on the faculty of the Computer Science Department at the **Universitat Autònoma de Barcelona**. I am the co-director of the **Data Science & Big Data Postgraduate Course** at UB and Director of the **Master in Foundations of Data Science**. I am the leader of the **DataScience@UB** group, whose objective is to promote technology transfer.

Instructors



Laura Igual



Petia Radeva



Oriol Pujol



Jordi Vitrià



Lluis Garrido



Albert Diaz-Guilera



Eloi Puertas



Sergio Escalera



Santi Seguí



Josep Perelló



Montse Guillen



Introduction

Universitat de Barcelona's **Data Science and Big Data** course offers students a program that covers the concepts and tools you will need throughout the entire data science pipeline: asking the right questions; wrangling and cleaning data; generating hypothesis; making inferences; visualizing data; assessing solutions; and building data products.

Schedule

~~Every Tuesday 18h 20h and Thursday 18h 20h~~

Workload

8 hours per week (including lectures and homework)

Location

Aula T1, Facultat de Matemàtiques i Informàtica,
Edifici Històric de la Universitat de Barcelona,
Gran Via de les Corts Catalanes 585, 08007, Barcelona.

Requirements:

The program is specially designed for students with a background in computer science, mathematics, and applied statistics, but other backgrounds can be considered.

We will require to follow lessons and complete class exercises **using personal laptops**. You will not be able to complete all your assignments in class if you rely solely on desktop equipment at home.

Raw Data

1. Processing: How I do clean and separate my data?

- Identification: filter data.
- Outliers.
- Imputation: missing value processing.
- Reduction: dimensionality reduction.
- Normalization: duplicates, ranges, format, coordinates, units, etc.
- Feature extraction.

3. Enrichment: How do I add more information to my data?

- Feature engineering.
- Search for additional data sources.

5. Analyze: How do I model my data?

- Variable selection (How do I determine important variables?)
- Probabilistic modeling (How are my variables related?)

7. Evaluate: Are the outcomes generic and robust?

- Statistical Testing.
- Model performance.

2. Acquire: How do I get my data?

- Web Scraping.
- Data Base queries.
- Access to bulk data stores.

4. Aggregation: How do I collect and summarize my data?

- Basic Statistics: mean, std, box plots, scatter plots, counts, etc.
- Distribution fitting.
- Feature aggregation.

6. Discover: What are the key relationships in my data?

- Clustering (How do I segment the data to find natural groupings?)
- Visualization (Are there unexpected relationships?)

8. Predict: What are the likely future outcomes?

- Regression (How do I predict the future?)
- Classification (How do I predict a category?)
- Recommendation (How do I predict relevant conditions?)

Data Science Path

Insights

Activities

twitter: @datascienceub

alumni @datascienceub

Evaluation

Capstone Project: An important part of the course is the IPython process notebook. This notebook details your steps in developing a solution to a real problem, including how you collected the data, alternative solutions you tried, describing statistical methods you used, and the insights you got.

Master Classes



<https://github.com/DataScienceUB/Postgrau>

The screenshot shows the GitHub repository page for 'DataScienceUB / Postgrau'. The repository has 18 commits, 1 branch, 0 releases, and 1 contributor. It contains files named 'algorismes', 'images', and 'README.md'. A large banner at the bottom promotes the 'Data Science and Big Data Postgraduate Course at UB'.

This repository

Search

Pull requests Issues Marketplace Explore

Watch 1 Star 0 Fork 0

DataScienceUB / Postgrau

Code Issues 0 Pull requests 0 Projects 0 Wiki Settings Insights

No description, website, or topics provided. Edit

Add topics

18 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

algorismes Update README.md Latest commit 5b812d6 24 seconds ago

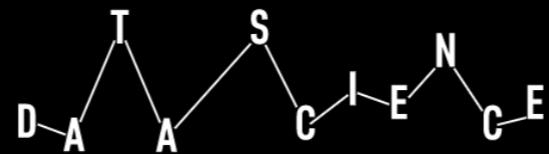
images Add files via upload 3 days ago

README.md Update README.md 23 seconds ago

README.md

Data Science and Big Data Postgraduate Course at UB

Data Science and Big Data course offers students a program that covers the concepts and tools you will need throughout the entire data science pipeline: asking the right questions; wrangling and cleaning data; generating hypothesis; making inferences; visualizing data; assessing solutions; and building data products.



Data Science & Big Data Postgraduate Course

2020-21

8a edición

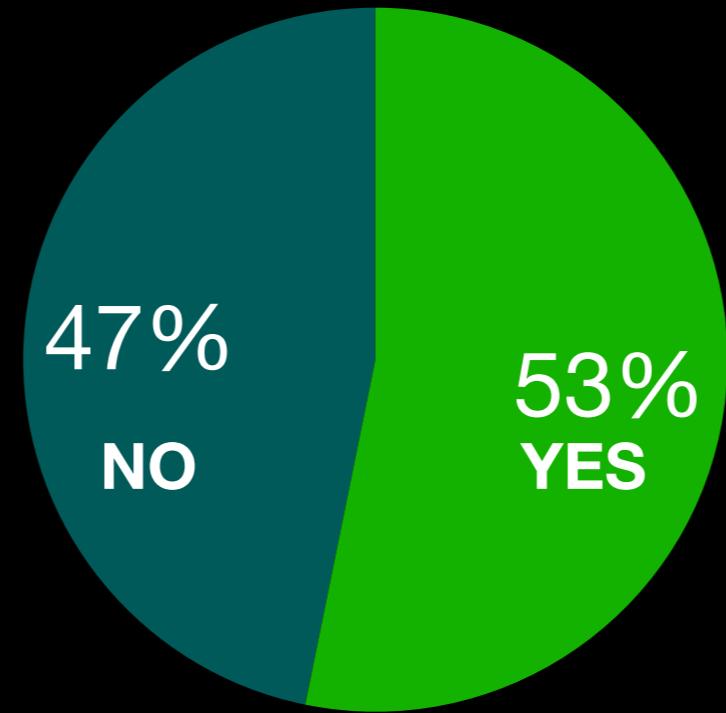
+250 alumni

Alta demanda (+200 solicitudes)



Results from pasts editions

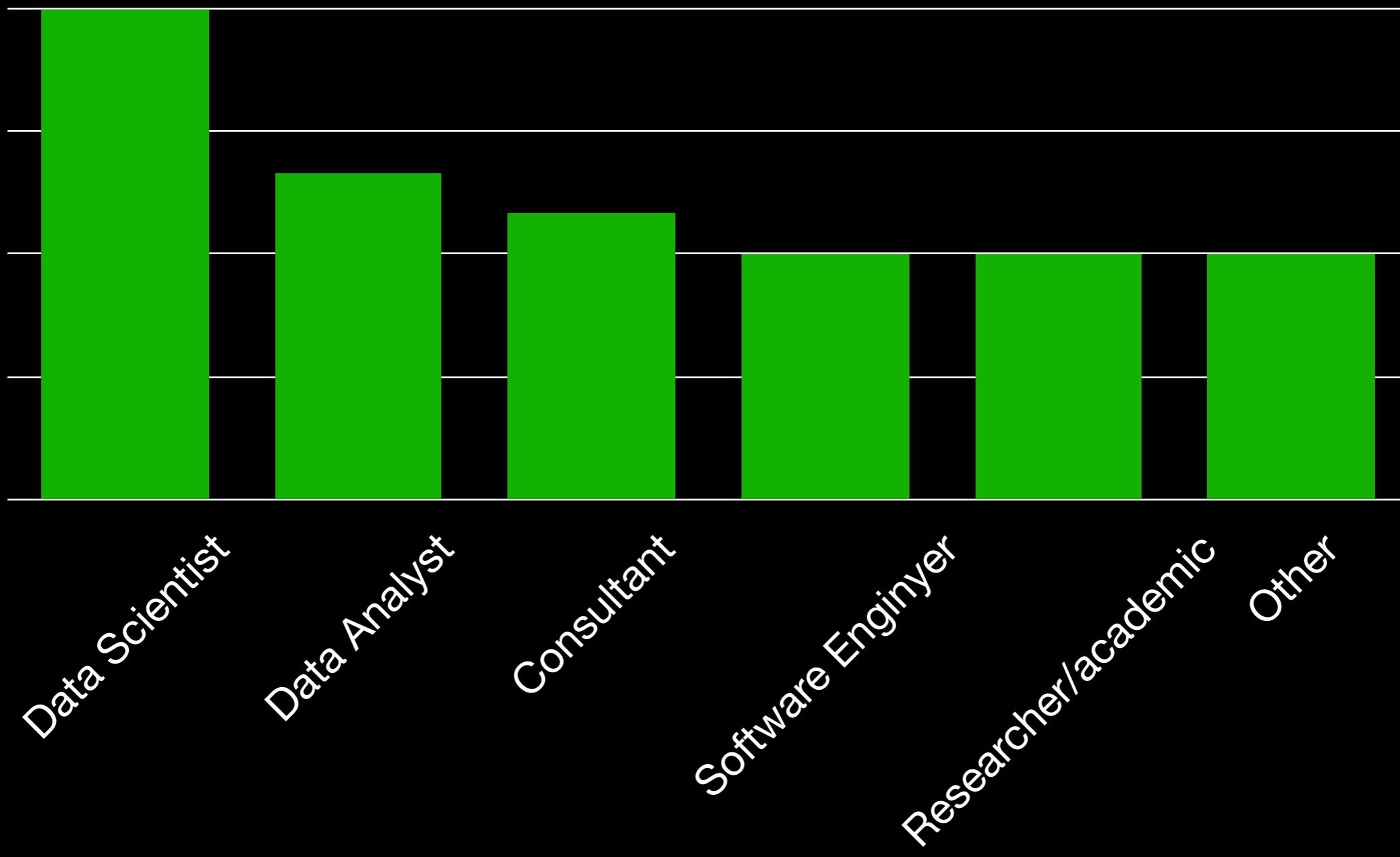
Did you get a **job**
thanks
(or partially thanks)
to the
DataScience course at **UB** ?



From a total of 37 responses

Results from past editions

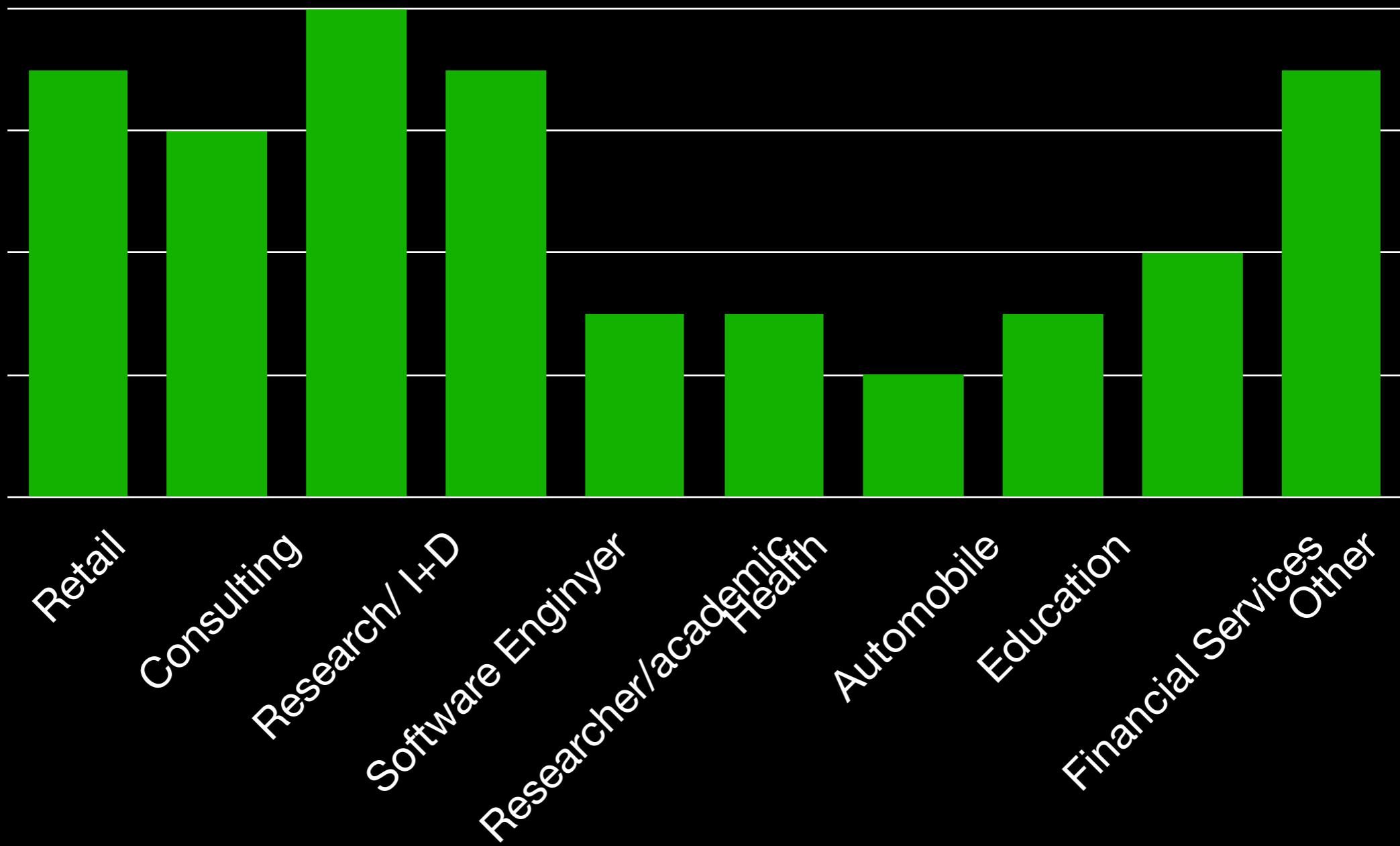
What is your Job Title?



From a total of 47 responses

Results from past editions

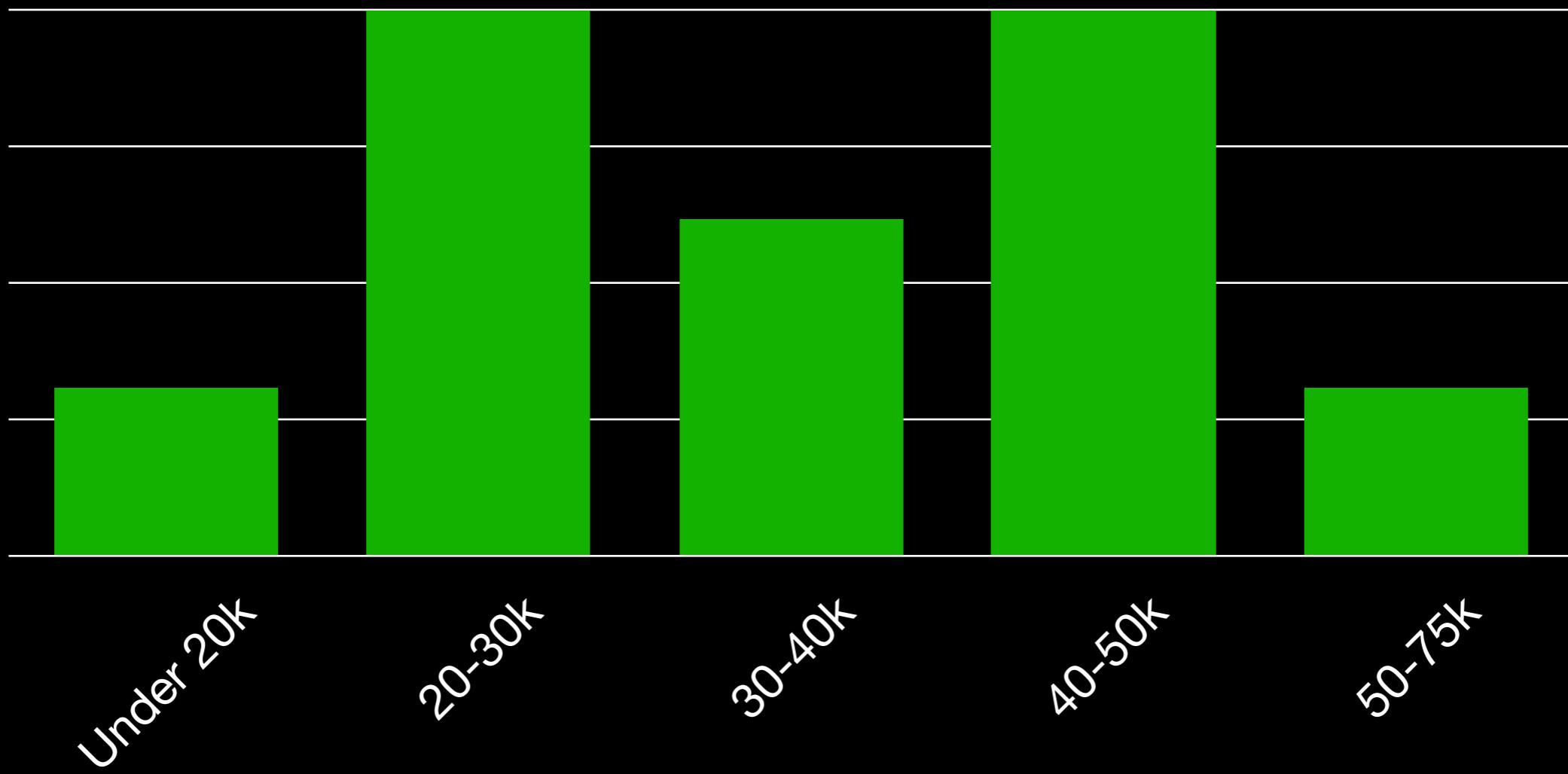
Which industry sector do you work in (or are most involved with)?



From a total of 37 responses

Results from past editions

Which **industry** sector do you **work in** (or are most involved with)?



From a total of 37 responses

Alumni

