Webinar Attendance:

A Data Science Project for 3D Heals

Background

3DHeals “aims to bridge the knowledge and experience gap between 3D printing, an emerging technology, and the established healthcare and life science innovation ecosystem.”

3dheals.com/about/

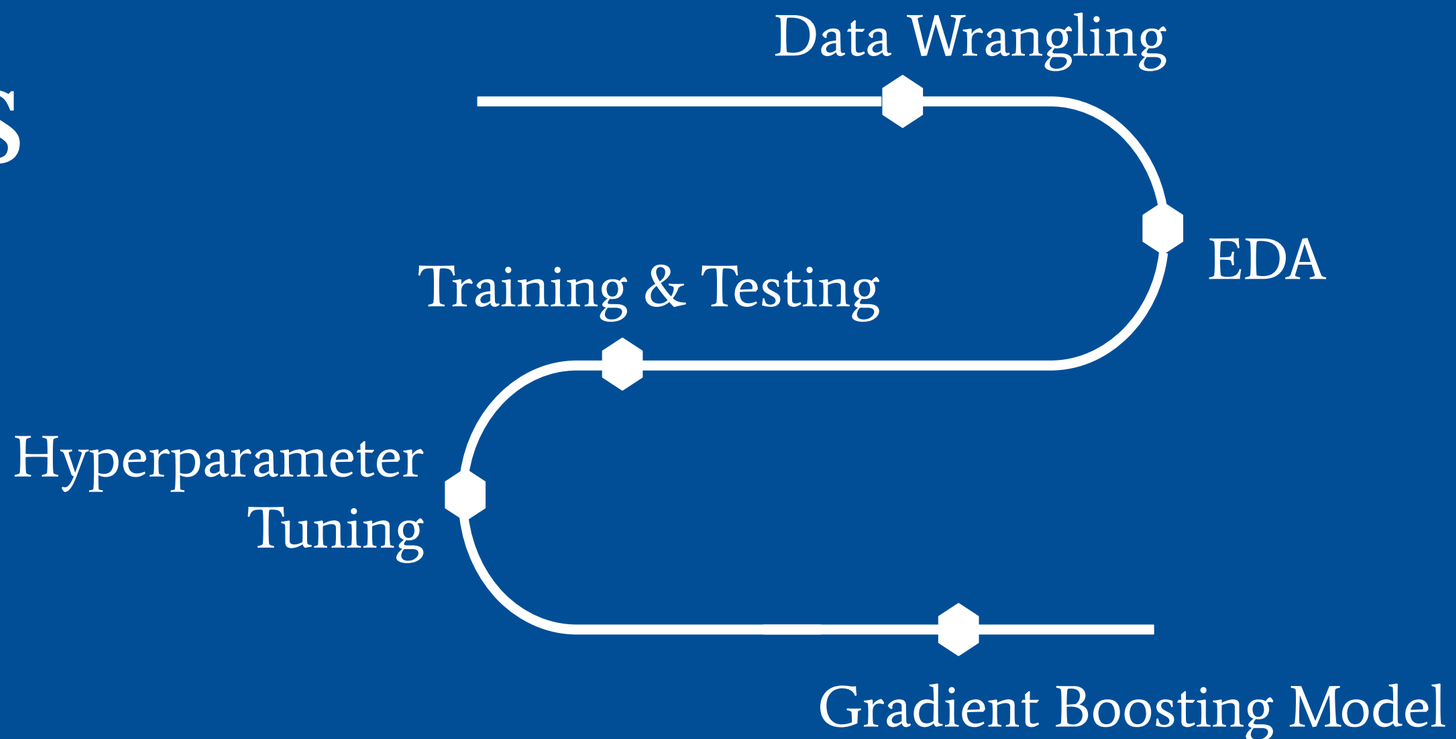
Background

3DHeals wants to know who is joining their free webinars in order to tailor the webinars to encourage more participants, as well as more engagement during the events.

The Project

The Question	The Data	The Deliverables
What can be developed that informs 3DHeals who is joining their webinars so that they can increase the number of participants?	The data is from Zoom of four different webinars put on by 3DHeals in 2019.	Colab notebooks of data wrangling through modeling, as well as this final report and presentation

The Process


Data Wrangling
EDA
Training & Testing
Hyperparameter Tuning
Gradient Boosting Model

Data Wrangling

We cleaned the data.

Imputing Missing Values

Condensing Job Titles

Aggregating Multiple Logins

Academic	93
C-Suite	68
Researcher	62
Manager, Director	57
Engineer	49
Clinician	22
Marketer, Salesperson	16
Designer	12
Business Developer	11
Consultant	9
Operator	2

Data Wrangling

We created new features.

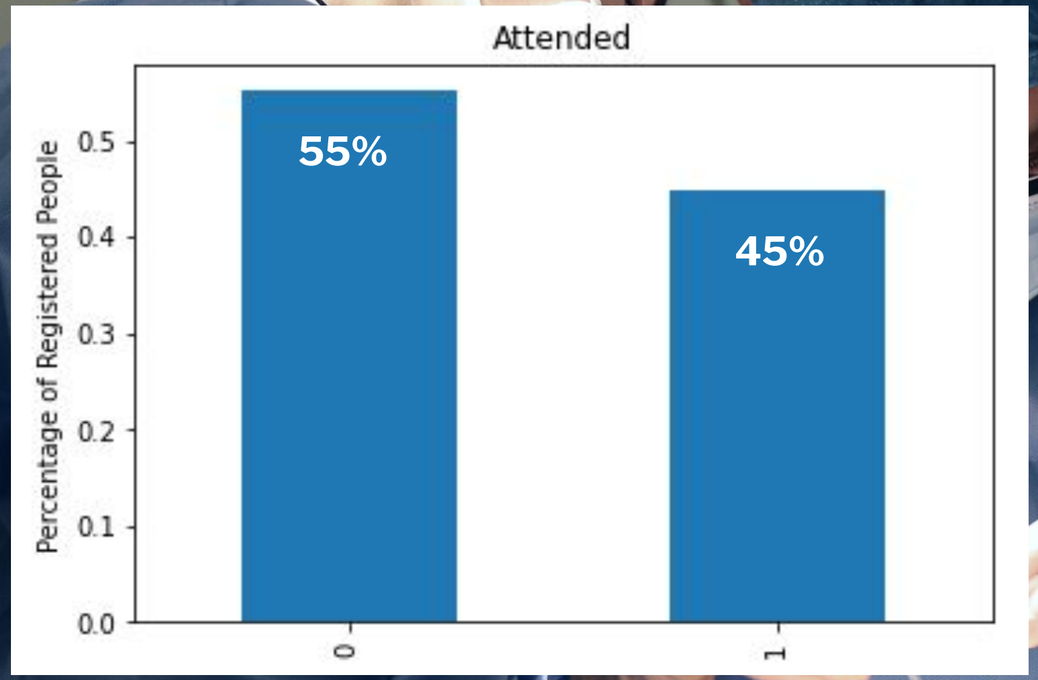
Number of Logins

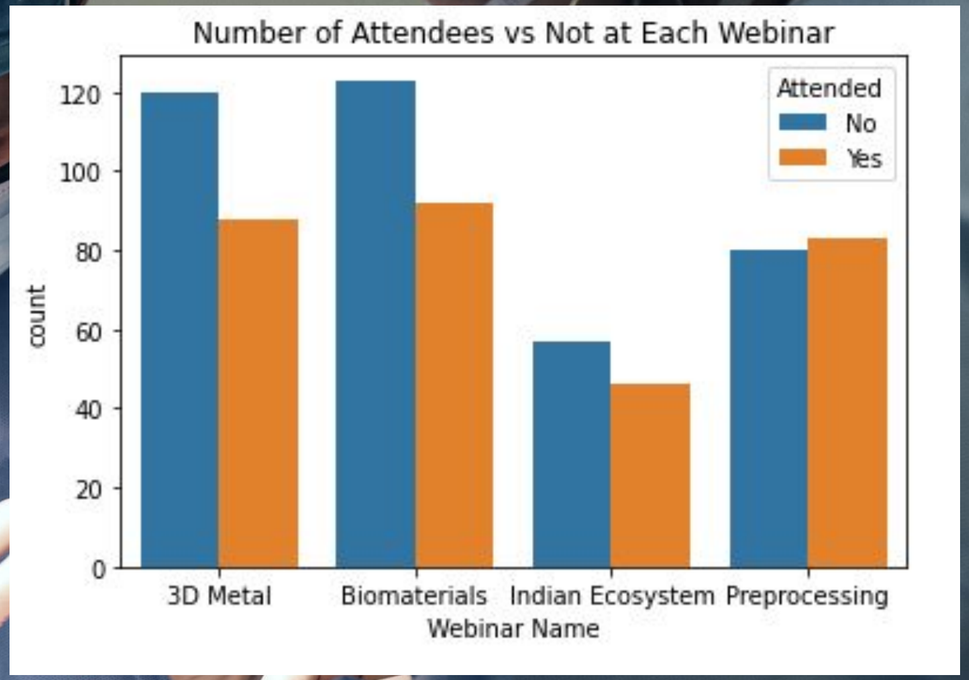
Time in Session as a Percent

Night time logins

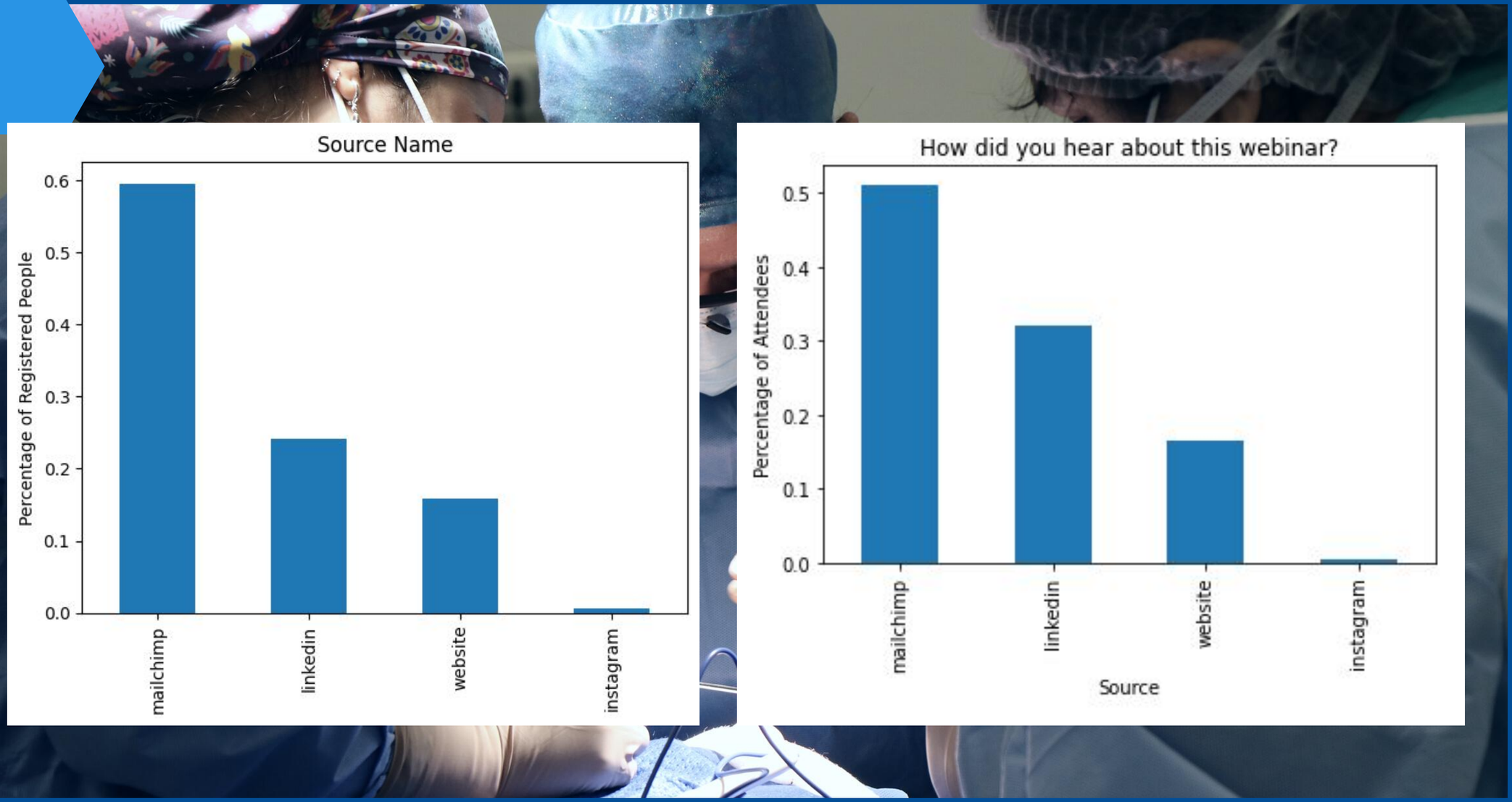
Number of Webinars Attended

Number of Webinars Registered


Attended
Percentage of Registered People
0.0
0.1
0.2
0.3
0.4
0.5
55%
45%
0
1


Number of Attendees vs Not at Each Webinar
Attended
No
Yes
count
0
20
40
60
80
100
120
3D Metal
Biomaterials
Indian Ecosystem
Preprocessing
Webinar Name

EDA Results


Source Name
Percentage of Registered People
0.6
0.5
0.4
0.3
0.2
0.1
0.0
mailchimp
linkedin
website
instagram
How did you hear about this webinar?
Percentage of Attendees
0.5
0.4
0.3
0.2
0.1
0.0
mailchimp
linkedin
website
instagram
Source

EDA Results

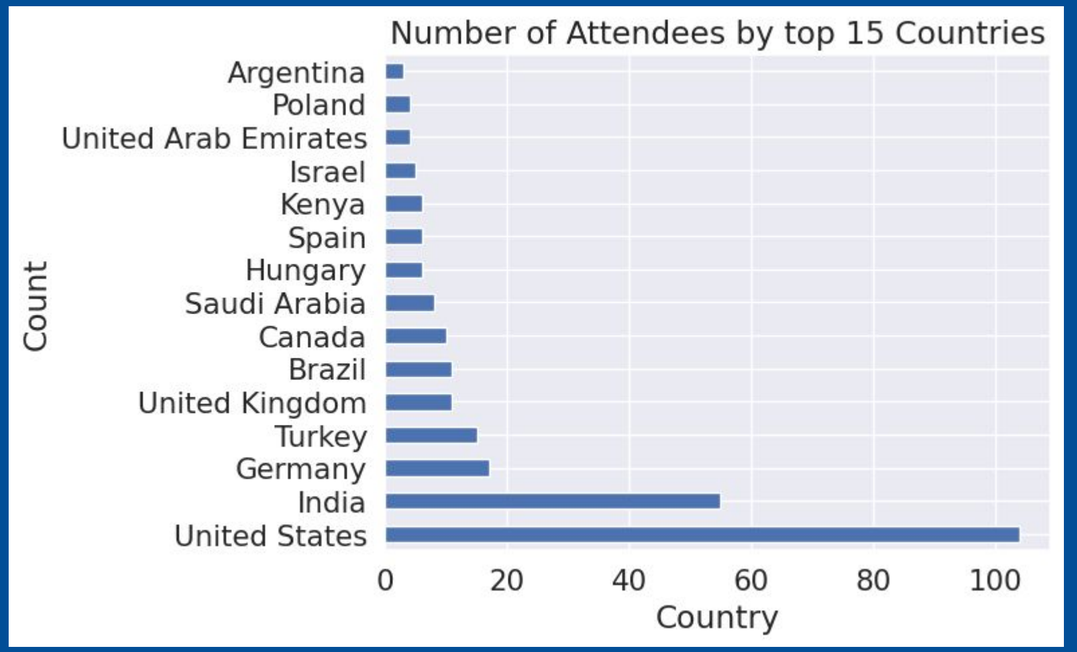

Average Time Spent in a Webinar by Job Title
Job Title
Operator
Clinician
Other
Manager, Director
Academic
Engineer
Business Developer
Consultant
C-Suite
None
Researcher
Marketer, Salesperson
Designer
0
20
40
60
80
Minutes

EDA Results


Average Time Spent in a Webinar by top 15 Industries
Industry
Accounting
High Tech - Hardware
High Tech - ISP
Hospital, Clinic, Doctor Office
Software - Healthcare
Energy, Chemical, Utilities
Medical, Pharma, Biotech
Education
Consulting
Manufacturing
High Tech - Other
Other
Telecommunications
Advertising/Marketing/PR
Software - Other
0
20
40
60
80
100
Minutes

EDA Results


Average Time Spent in a Webinar by top 15 Countries
United Kingdom
France
Hungary
Canada
Brazil
Vietnam
Germany
Bosnia and Herzegovina
Finland
Nigeria
Bolivia
Costa Rica
Guatemala
Colombia
Indonesia
Country
0
20
40
60
80
100
Minutes


Number of Attendees by top 15 Countries
Argentina
Poland
United Arab Emirates
Israel
Kenya
Spain
Hungary
Saudi Arabia
Canada
Brazil
United Kingdom
Turkey
Germany
India
United States
Count
0
20
40
60
80
100
Country

EDA Results

Gradient Boosting Regression Model

RMSE = 0.24

The model predicted the percent of time a participant will stay in the webinar (between o and 1).

After tuning, the root mean squared error (RMSE) was 0.24, which means on average our model was off by 24%.

Best Hyper-Parameters

01. n_estimators:
50

02. min_samples_split:
10

03. max_depth:
10

04. learning_rate:
0.1

Next Steps

- Include more webinars
- Collect more information on participants
- Explore clustering
- Deploy model

Credits

Thanks to the following:

- Jenny Chen
- Diana Golman
- Mentors: Mary Anne Thygesen, Kenneth Gil-Pasquel, Rahul Sagrolikar
- SlidesCarnival for the presentation template
- Pexels for the photos

Thank you!

All slides, reports, notebooks, functions, and models are available for future use in our shared Google Drive folder.

Andrea Ricci
Data Scientist
andreaeddy10@gmail.com
github.com/andreaeddyricci