



# Kaggle & GitHub

Build your Data Science profile

# Agenda

kaggle



**GitHub**



**Desktop**

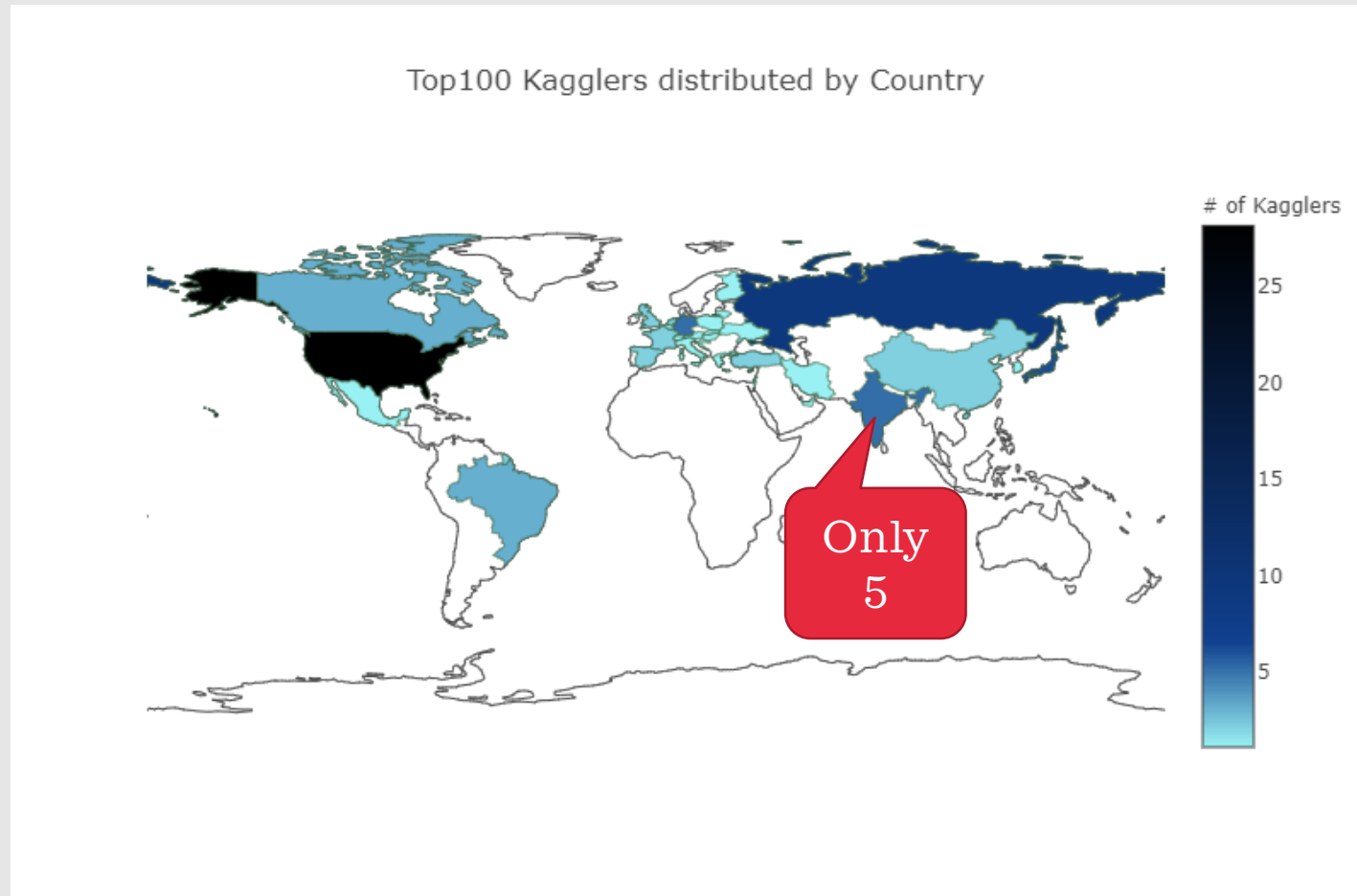


Kaggle

# Why Kaggle?

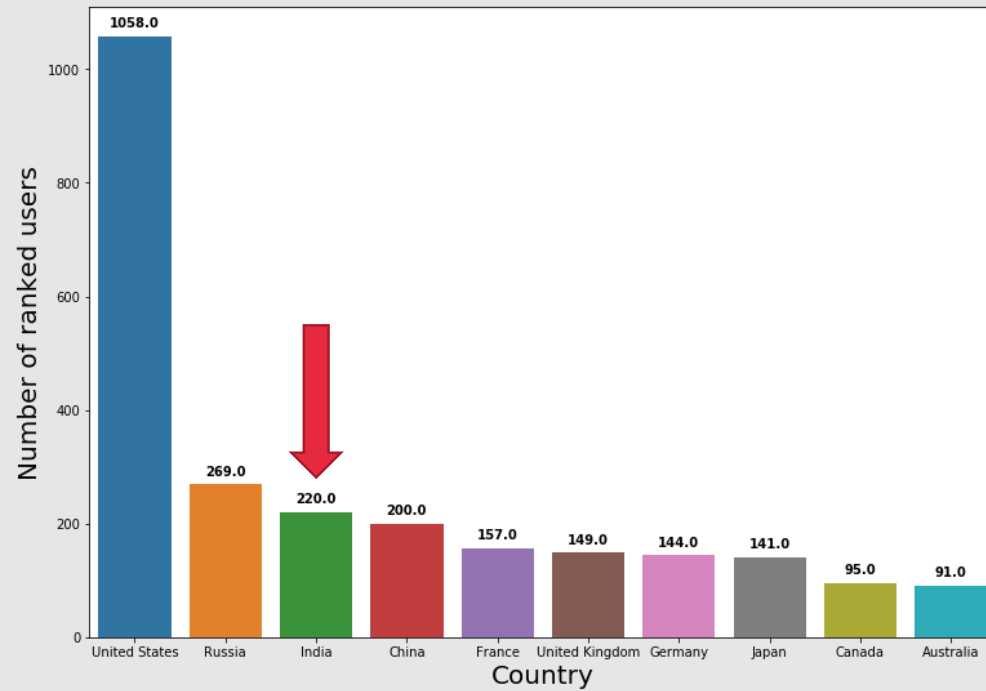
- Well defined problems statement with proper datasets for training
- Public and Private leaderboards help you get reality check on how good or bad your model performs
- Great discussion and sharing of starter codes
- Build portfolio of real-world projects (with luxury of prepared datasets) that can demonstrates your skills
- No matter which background you come from or what degree or course your skill will be tested here
- Most important one! If you love Data Science, you would naturally spend all the time here

# Fact of Kagglers's in India

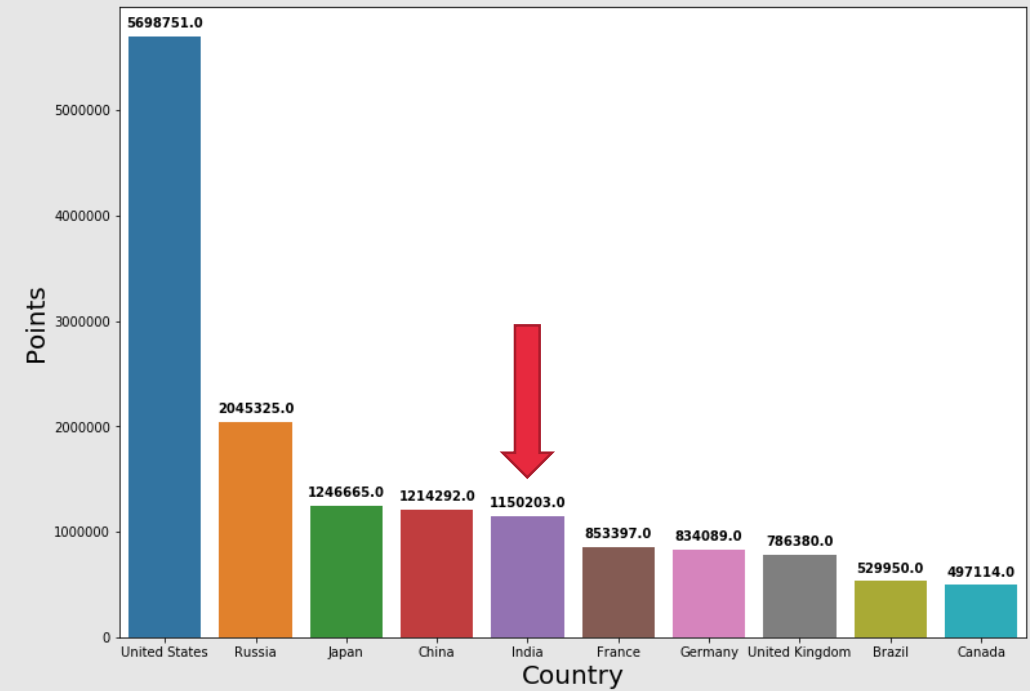


# Fact of Kaggler's in India

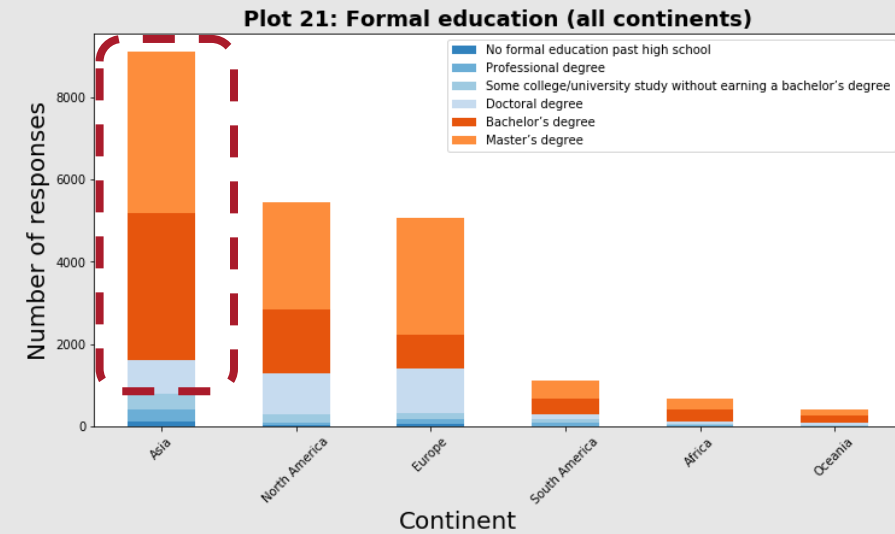
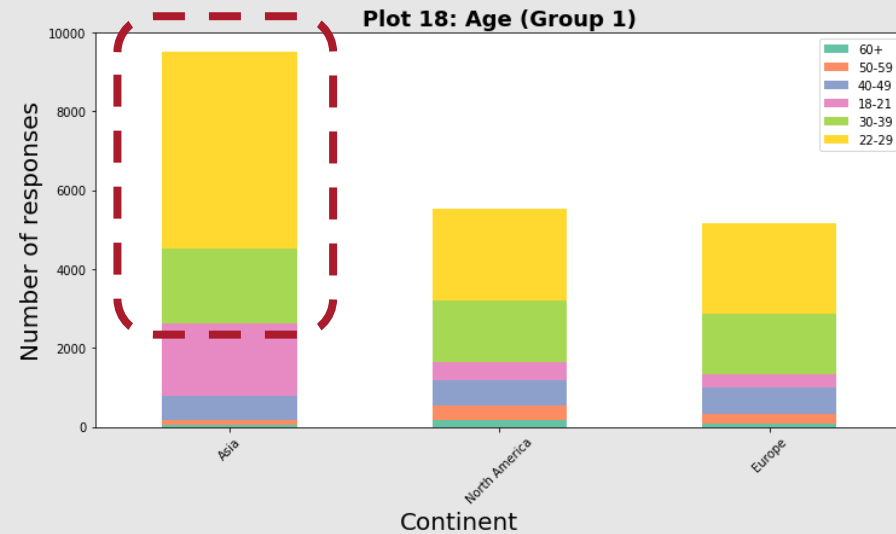
Plot 1: Top 10 countries by number of ranked users



Plot 11: Top 10 countries by total points



# Fact of Kagglers's in India



# What should be your Goal? – Kaggle Ranks



## **Kaggle Progression System**

Kaggle's Progression System uses performance tiers to track your growth as a data scientist on Kaggle. Along the way, you'll earn medals for your achievements and compete for data science glory on live leaderboards.



# Begin with



## Novice

You've joined the community.

- ☒ Register!



## Contributor

You've completed your profile, engaged with the community, and fully explored Kaggle's platform.

- ☒ Add your bio
- ☒ Add your location
- ☒ Add your occupation
- ☒ Add your organization
- ☒ SMS verify your account
- ☒ Run 1 script
- ☒ Make 1 competition submission
- ☒ Make 1 comment
- ☒ Cast 1 upvote

# Improve by Participating and contributing



## Expert

You've completed a significant body of work on Kaggle in one or more categories of expertise. Once you've reached the expert tier for a category, you will be entered into the site wide Kaggle Ranking for that category.

Competitions

☐ 2 bronze medals

Datasets

☐ 3 bronze medals

Kernels

☐ 5 bronze medals

Discussions

☐ 50 bronze medals



## Master

You've demonstrated excellence in one or more categories of expertise on Kaggle to reach this prestigious tier. Masters in the Competitions category are eligible for exclusive Master-Only competitions.

Competitions

☐ 1 gold medal

☐ 2 silver medals

Datasets

☐ 1 gold medal

☐ 4 silver medals

Kernels

☐ 10 silver medals

Discussions

☐ 50 silver medals

☐ 200 medals in total

# Being a Grand Master – Life long endeavor



## Grandmaster

You've consistently demonstrated outstanding performance in one or more categories of expertise on Kaggle to reach this pinnacle tier. You're the best of the best.

### Competitions

- ☐ 5 gold medals
- ☐ Solo gold medal

### Datasets

- ☐ 5 gold medals
- ☐ 5 silver medals

### Kernels

- ☐ 15 gold medals

### Discussions

- ☐ 50 gold medals
- ☐ 500 medals in total

# How to Rank?



## Competition Medals

Competition medals are awarded for top competition results. The number of medals awarded per competition varies depending on the size of the competition. Note that InClass, playground, and getting started competitions do not award medals.

	0-99 Teams	100-249 Teams	250-999 Teams	1000+ Teams
Bronze	Top 40%	Top 40%	Top 100	Top 10%
Silver	Top 20%	Top 20%	Top 50	Top 5%
Gold	Top 10%	Top 10	Top 10 + 0.2%*	Top 10 + 0.2%*

\* (Top 10 + 0.2%) means that an extra gold medal will be awarded for every 500 additional teams in the competition. For example, a competition with 500 teams will award gold medals to the top 11 teams and a competition with 5000 teams will award gold medals to the top 20 teams.

# How to Rank?



## Dataset Medals

Dataset Medals are awarded to popular public datasets published to the site, as measured by number of upvotes. Not all upvotes count towards medals: votes by novices are excluded from medal calculation.



 Bronze	5 Votes
 Silver	20 Votes
 Gold	50 Votes



## Kernel Medals

Kernel Medals are awarded to popular kernels, as measured by the number of upvotes a kernel receives. Not all upvotes count towards medals: self-votes, votes by novices, and votes on old posts are excluded from medal calculation.



 Bronze	5 Votes
 Silver	20 Votes
 Gold	50 Votes

# How to Rank?



## Discussion Medals

Discussion Medals are awarded to popular topics and comments posted across the site, as measured by net votes (upvotes minus downvotes). Not all upvotes count towards medals: votes by novices and votes on old posts are excluded from medal calculation.

 Bronze	1 Vote
 Silver	5 Votes
 Gold	10 Votes

# How?

- <https://elitedatascience.com/beginner-kaggle>
- <https://machinelearningmastery.com/get-started-with-kaggle/>

# The Business Context of Kaggle

- **Speed of execution** – A key factor in Business is to make progressive and iterative improvements versus making a one time perfect model
- **Getting projects done** – There is not straight forward ways to solve business problems. Kaggle prepares you for the uncertain by handling multiple projects
- **Feature Engineering** – Your models accuracy depends on how well you choose or create features. Kaggle community shows tremendous support in showcasing various ways to engineer features
- **Usage of Kaggle Notebooks** – You do not have to worry about running models on your local system. Kaggle provides you with a 30 hour runtime Server to run your projects.
- **Facing interviews better** – If you learn to have fun while solving Kaggle without any deadlines, it helps you be confident in front of interviews





# GitHub

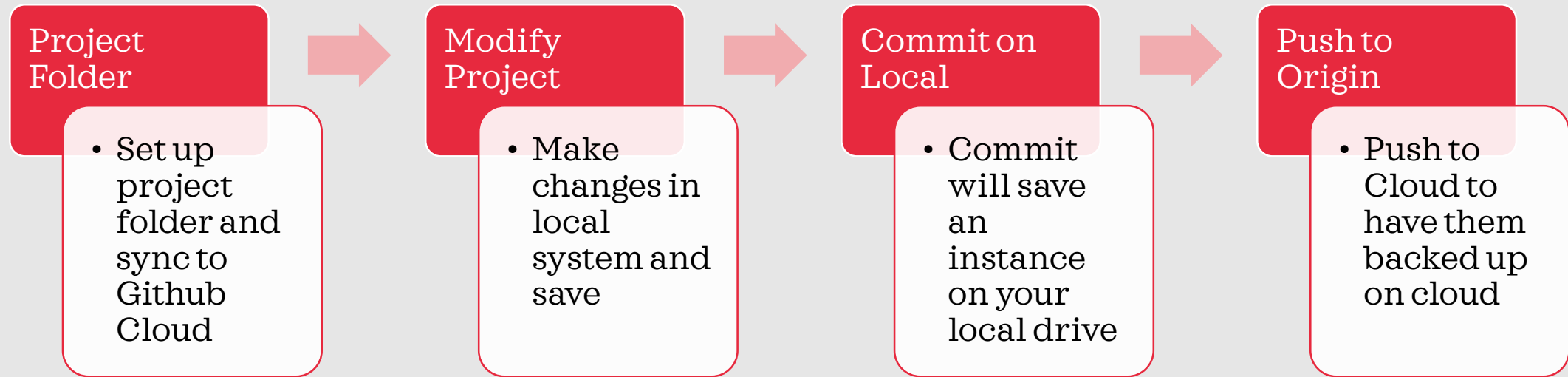
GitHub – Don't forget to save your work!

GitHub – On  
Cloud



GitHub Desktop  
- save your work  
to Cloud

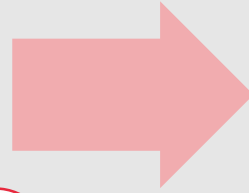
# Steps to Save on Cloud



# Retrieve from Cloud

## Select Project

- Choose the project folder on local you want to save the project



## Select Branch

- Choose a branch you want to revert to and pull

# The Business Context of GitHub

- **Record of all changes of project** – Keeping the a version that works is key to how you make progress in Data Science projects. GitHub helps you do it efficiently
- **Best for working in teams**– Combining work can be a tedious task when working with teams. GitHub helps teams work together and combine work easily
- **Restore working codes** – In case of mess up you can rely on GitHub to restore you to a point where things were working
- **System crash?** – No problem, GitHub backs it up on cloud and you can start where you left off!
- **For Resume** – Having a well organised GitHub profile helps you track your progress and show off your work to recruiters



# Lifecycle of Data Science projects

With Context of Kaggle and GitHub

# Broad points

- Start with a end Goal in mind – Always make a first draft
- Save the last working version – Keep iterating to improve accuracy or improve recommendations of analysis