



# Penn Data Science Group

## Spring 2017 Info Session

January, 26<sup>th</sup> 2017, Berger Auditorium



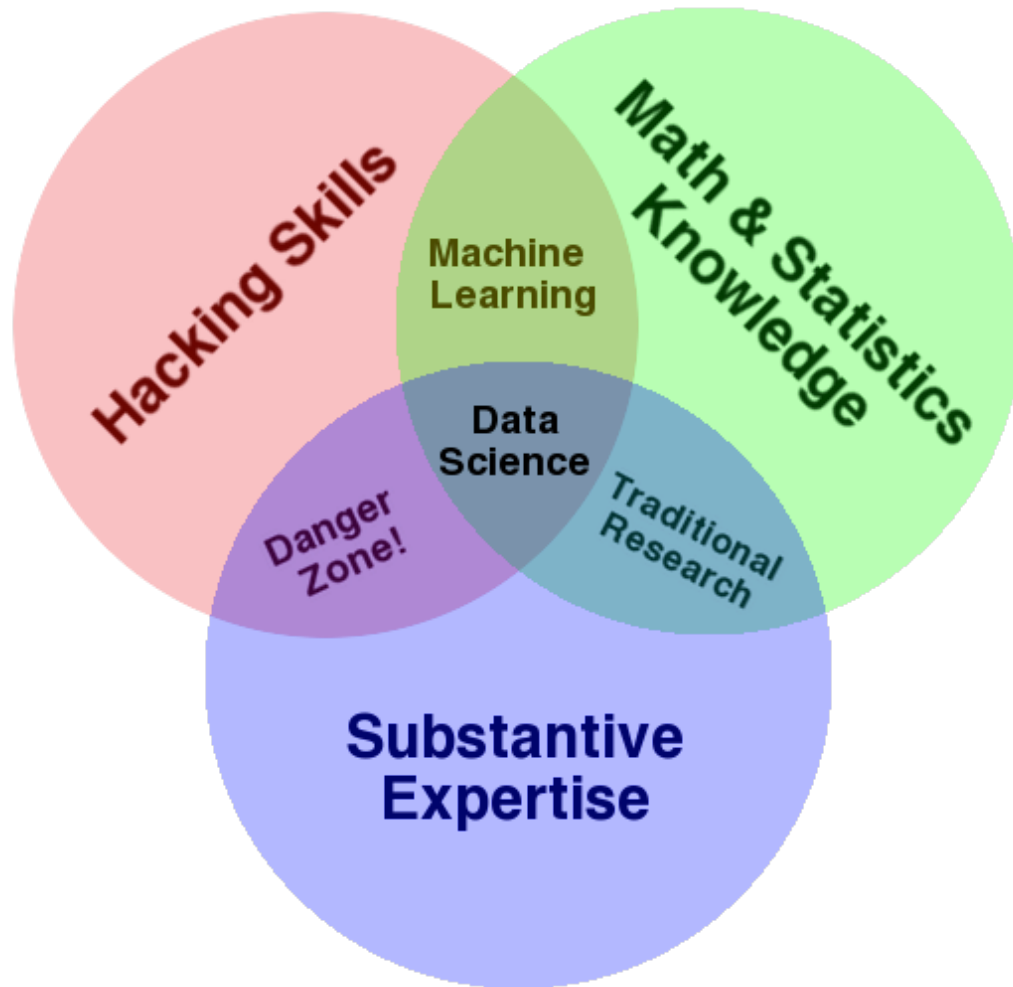
**gapsa**

GRADUATE *and* PROFESSIONAL STUDENT ASSEMBLY  
UNIVERSITY of PENNSYLVANIA



**GSEG**  
Graduate Student Engineering Group

# What is Data Science?



Extract insights from data using...

- Statistics
- Machine Learning
- Data Mining
- Predictive Analytics
- Data Visualization
- Probability Models
- ...and more

[https://en.wikipedia.org/wiki/Data\\_science](https://en.wikipedia.org/wiki/Data_science)

# Why Be a Data Scientist?



# Why Be a Data Scientist?

## Data Scientist Salaries

1,977 Salaries

Updated Jan 13, 2017

National Avg

**\$113,436**

Min

\$76k

Max

\$146k

Philadelphia, PA Area Avg

**\$102,935\***

Min

\$76k

Max

\$132k

# What PDSG Does

▶ Community Events Calendar

Resources Page

Workshops

Career Talks

Student Talks

Projects

# Community Events Calendar

Types of events we list:

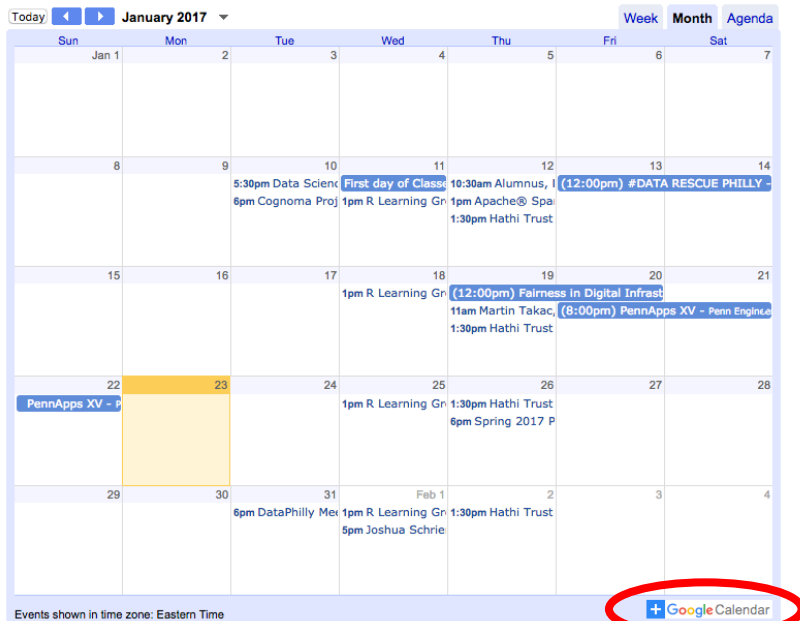
- DataPhilly Meetups ([meetup.com/DataPhilly](https://www.meetup.com/DataPhilly))
- Hackathons
- Data-related Academic Seminars
- Career Fairs
- Company Info Sessions
- Networking Events
- Workshops
- PDSG Events

# Community Events Calendar

Add to your  
calendar

[penndsg.com/events](http://penndsg.com/events)

Data Science Events Near Penn Campus



Get the weekly  
newsletter

[penndsg.com/join](http://penndsg.com/join)

comfortable

Would you like to receive weekly emails with a list of data science related events occurring that week (including PDSG events)? \*

If yes or not sure, we'll email you an invitation to a Google Group mailing list for you to accept or decline.

☒ Yes

☐ No

☐ Not Sure

Would you like to participate in a data science project on a small

Other events? Send info to:  
[board@penndsg.com](mailto:board@penndsg.com)

# What PDSG Does

Community Events Calendar

▶ Resources Page [penndsg.com/resources](https://penndsg.com/resources)

Workshops

Career Talks

Student Talks

Projects



# What PDSG Does

Community Events Calendar

Resources Page

▶ Workshops

Career Talks

Student Talks

Projects

# Workshops

Hands-on practice with data science tools and techniques

Workshop currently in planning:

**Predicting Titanic passenger outcomes  
using Jupyter/iPython Notebook**



Want to lead a workshop? Let us know: [board@penndsg.com](mailto:board@penndsg.com)

# What PDSG Does

Community Events Calendar

Resources Page

Workshops

► Career Talks

Student Talks

Projects

We want to hear from...

- Data Scientists
  - Data Engineers
  - Data Analysts
- etc.

Send referrals to:

[board@penndsg.com](mailto:board@penndsg.com)

# What PDSG Does

Community Events Calendar

Resources Page

Workshops

Career Talks

▶ Student Talks

Projects

We want to hear about your data-related...

- Research
- Internship Experience
- Side Project

To volunteer, email  
[board@penndsg.com](mailto:board@penndsg.com)

# What PDSG Does

Community Events Calendar

Resources Page

Workshops

Career Talks

Student Talks

► Projects

# Projects

Why do a project through PDSG?

- Build your data science portfolio
- Find like-minded people to work with you
- Build something you can be proud of
- Get help from other PDSG members by posting questions on Piazza

No really, why?

# Project Prizes

If you want prizes like \$50 Amazon gift cards, here's what to do:

1. Get a group of 2-4 people
2. Pick a data problem to explore
3. Tell us your team and project idea
4. Write a blog post style report on your project
5. Present your results in front of a panel of judges
6. Have one of the best projects
7. Collect gift cards (1 per team member, excluding board members)

# Raffle Prizes

At the end of the semester, we'll raffle off small prizes, like \$5 Amazon or Starbucks gift cards. You get 1 raffle ticket for each:

- Project report put up on the website
- Student talk you give
- Data professional referral who ends up giving a career talk

2 raffle tickets if you lead a workshop!



# Kaggle Competitions

If you don't know what kind of project to do, do a Kaggle competition

- Clean datasets
- Immediate model evaluation
- Scripts shared by other competitors

# Recommended Kaggle Competitions

## Titanic: Machine Learning from Disaster

- [kaggle.com/c/titanic](https://kaggle.com/c/titanic)
- Predict Titanic passenger outcome

## Digit Recognizer

- [kaggle.com/c/digit-recognizer](https://kaggle.com/c/digit-recognizer)
- Classify handwritten digits from famous MNIST data

## Shelter Animal Outcomes

- [kaggle.com/c/shelter-animal-outcomes](https://kaggle.com/c/shelter-animal-outcomes)
- Predict Austin Animal Center animal outcomes

## TalkingData Mobile User Demographics

- [kaggle.com/c/talkingdata-mobile-user-demographics](https://kaggle.com/c/talkingdata-mobile-user-demographics)
- Predict demographics characteristics from mobile phone data

## Predicting Facebook Check-ins

- [kaggle.com/c/facebook-v-predicting-check-ins](https://kaggle.com/c/facebook-v-predicting-check-ins)
- Predict where people like to check in on Facebook based on fictitious data similar to data Facebook employees use