# THE DATA MINE:

## EXPERIENTIAL INDUSTRY PRACTICUMS IN DATA SCIENCE

*Margaret Betz, Managing Director of Corporate Partnerships*

**PURDUE UNIVERSITY** | The Data Mine

# Overview of The Data Mine

- Living, learning community (many students live in Hillenbrand Hall)
- All backgrounds and majors welcome: **Data Science for All!**
- Supportive, active learning environment
- Learn data science skills in HPC, Python, R, SQL; no lectures; all project based
- Data sets at modern complexity and size, from all disciplines
- 47 Corporate Partners, offering 60 projects, during 2021-22 academic year
- 675 undergrad and 97 grad students (fall 2021); aiming for 1300 total for fall 2022
- **124 unique majors represented by students**



Burton D. Morgan Center for Entrepreneurship (MRGN)

PURDUE UNIVERSITY® | The Data Mine



Hillenbrand Hall

# Weekly Seminar (1 credit)



Above: Students meet in the Hillenbrand Dining Court Atrium for seminar on Mondays

the-examples-book.com

# Summary of student resources

- **Weekly seminar (1 credit) –** 4 (identical) weekly synchronous interactive seminars on Mondays
- **Weekly projects** – 1 weekly project where students gain experiences working with *real* datasets on HPC cluster using a variety of common tools like R, Python, bash, awk, SQL, web scraping, regular expressions, XML.
- **The Examples Book** – a resources for supplementary material for students' weekly projects (https://the-examples-book.com/)
- **Office hours** – 80 TA office hours each week; Mon-Fri 8 AM – 9 PM in the residence hall
- **Piazza.com** – an asynchronous platform for student interaction enable communication between instructors and students; **7,864** student interactions in AY 2021-22; average response time of **9 minutes**

# Corporate Partners

- Students work 8-10 hours per week for 3 credits per semester
- Data from industry partners and meaningful projects
- Mentoring by experts in industry and academia
- Full academic year long projects (August – May)
- Team leadership (TA) positions for 2nd, 3rd, 4th year
- Paths to internships and permanent employment
- Sponsored Research Contracts with 3 partners and expanding

|  | 2018-19 | 2019-20 | 2020-21 | 2021-22 | 2022-23 (predicted) |
|---|---|---|---|---|---|
| Number of Partners | 1 | 12 | 23 | 47 | 65 |
| Number of Projects | 1 | 16 | 31 | 60 | 85 |
| UG Students in CRP program | 10 | 141 | 310 | 407 | 700 |
| Grad Students in CRP program | 0 | 0 | 92 | 100 | 150 |

**PURDUE UNIVERSITY®** | The Data Mine

# 2021 – 2022 Corporate Partners

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| American First Finance | ATOM | Bayer Crop Sciences | Beck's Hybrids | Microsoft Minecraft | Purdue MFRI | MISO | Mursix |
| Caterpillar | CDC PROTECT | Clifton Larsen Allen | Corteva | Nuvve | Pebblst | PlaneEnglish | Purdue Athletics |
| Cummins | Delta Faucet | DORIS | Elanco | PCRD | Purdue RecWell | Raytheon Technologies | REACH Public Health |
| Farfetch | Ford | Gro Master | Halderman | Renzoe Box | Republic Airways | Sandia National Lab | Telemetry Sports |
| Helmer Scientific | Indiana FSSA | IU Health | Jobvite | TMap - Make My Move | UPS | USAA | USDA Forest Service |
| John Deere & Company | Lockheed Martin | MakuSafe | Merck | Viasat | Wabash | Webee | |

**PURDUE UNIVERSITY** | The Data Mine

# All Presentations from Symposium Available Online



Corporate Partners Symposium
(Wednesday, April 20, 2022)

# Indiana Data Mine (funded by Lilly Endowment)

- New statewide network of regional hubs

- Students, faculty, industry, community leaders

- Collaborate to solve regional challenges

- Help manufacturing companies adopt cutting-edge technologies in automation and data exchange

- Opportunities to learn data science skills through immersive engagement

- Indiana companies; lead to careers in Indiana.

- Expand opportunities for data science training for first-generation and Black UGs



**LEGEND**

- Purdue Campuses
- Polytechnic Institute Statewide Sites
- Purdue Polytechnic High Schools
- Potential Future Purdue Polytechnic High Schools
- Technology Parks
- Indiana Digital Crossroads (IDC) Hubs
- Potential Future Indiana Digital Crossroads (IDC) Hubs

**Note:** Every Indiana county has a Purdue Extension Office

PURDUE UNIVERSITY® | The Data Mine
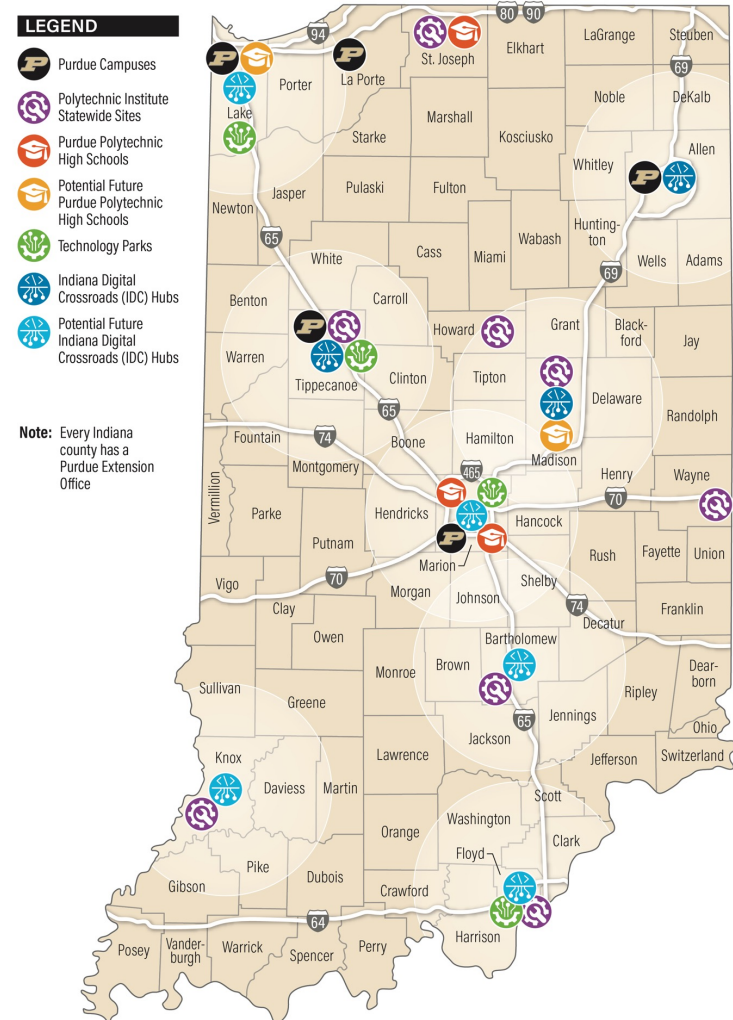
# National Data Mine Network (funded by NSF)

- New national data science experiential learning and research program

- Awarded to the American Statistical Association (Ward is PI)

- 300 undergraduate students at Minority Serving Institutions
- (including HBCUs, Hispanic Serving Institutions, Tribal Colleges and Universities)

- Research stipends ($5000 each) for 9-month projects

- Best practices for faculty who want to start university-industry partnerships

- Use high-performance computing to solve data-driven challenges

- Every sector of industry, including:

    - biomedical engineering, genomics, healthcare engineering, imaging, manufacturing, supply chain management, transportation, etc.

- https://beta.nsf.gov/science-matters/developing-21st-century-data-science-workforce

**PURDUE UNIVERSITY.** | The Data Mine

# Video: Cummins + The Data Mine



https://youtu.be/JabJjMzD6D0

# THANK YOU!

**Email:** datamine@purdue.edu
**Webpage:** datamine.purdue.edu

PURDUE UNIVERSITY® | The Data Mine