Data Science is all about the processes and methods to access and analyze data to gain insights for informed decision making. To promote the awareness and analytic technology of Big Data, IEEE COMPSAC and the IEEE Big Data Initiative are organizing a Data Analytics Competition.

Data
Data is all around us and everywhere. In particular, data that is relevant to the themes of the 41st COMPSAC, a major international forum for academia, industry, and government to discuss research results and advancements, emerging problems, and future trends in computer and software technologies and applications.


In this competition, participants will explore internet usage trends around the globe. By 2020, it is expected that 90% of the world’s population will be covered by mobile broadband networks and 70% of the population will be using smartphones. What impact will this have on our existing methods of internet usage and connectivity? Participants will examine today’s data to predict tomorrow’s developments.

In addition, we are providing U.S. Patent and Trademark Office (USPTO) application data from the PAIR Bulk Data database which contains over 9.4 million records. You may optionally select this dataset to analyze and present an interesting finding based on your analysis.

Datasets

Internet usage/Network data:

- Proportion of individuals using the Internet
  
  Source: Sustainable Development Goals Indicators | United Nations Statistics Division

- Fixed Internet broadband Subscriptions per 100 inhabitants
  
  Source: Sustainable Development Goals Indicators | United Nations Statistics Division
- Mobile-cellular telephone subscriptions per 100 inhabitants
  
  Source: World Telecommunication/ICT Indicators Database | International Telecommunications Union

U.S. Patent and Trademark Office application data:

- PAIR Bulk Data (full dataset)
  
  Source: United States Patent and Trademark Office

Instructions:

Questions
When analyzing the Internet usage/Network datasets, participants are encouraged to consider the following questions:
How has the trend of individuals accessing the internet changed over the past decade? How has it impacted the rate of broadband and mobile-cellular subscriptions? What areas of the world are undergoing the most drastic rate of change? What areas are experiencing the least change? Which countries contain the greatest divide between fixed internet broadband subscriptions and mobile-cellular subscriptions? What conclusions can you draw regarding the correlation of internet users per subscription? How does this differ between countries? What additional types of data would be useful in determining this? Based on this data, what near-term predictions would you make concerning the future of internet and mobile-cellular usage/adoption around the world? Long term predictions?

When analyzing the U.S. Patent and Trademark Office dataset (PAIR Bulk Data), participants are encouraged to consider the following questions:
What insights can you provide by analyzing the dataset? Your insights and suggestions are expected to be creative. Based on this dataset, what are the most common technology areas where patent development is occurring? How frequently are patents being filed with the same title? How would you improve this dataset to better distinguish unique patents with duplicate titles? What additional data / metadata would you include in this dataset to help researchers more efficiently locate relevant technology patents? What conclusions can you draw from this data? What trends, if any, have formed over the past decade? Where are the trends moving? What anomalies can you find in this data? Is there anything that affects the integrity of the data?

Participation
Participation is open to IEEE COMPSAC 2017 conference attendees only. To participate, please send an email to bdicompetition@ieee.org. You may participate individually or as part of a team.
Evaluation and Awards
You are expected to present your findings to a panel of judges. Presentations will be evaluated based on the following criteria:

- Selection criteria of the dataset
- Clarity and Relevance of Analysis
- Methodology
- Creative use of data
- Significance of findings and recommendations
- Delivery of findings and recommendations

Monetary awards will be given out to the top winning individuals or teams. Categories to include, 1) winning student or student team; 2) winning professional or professional team. A judging panel of practitioners will evaluate presentations.

The presentations should be formatted as a PowerPoint document that can be uploaded via IEEE DataPort using the “Submit an Analysis” button below. Please be prepared to discuss your presentation for roughly 8-10 minutes.

Timeline
- Open now
- Monday, 3 July 2017 – Registration closes
- Thursday, 6 July 2017 - Presentations
- Thursday, 6 July 2017 – Awards during the Dinner