HACKATHON DATA BOOKLET

Data Prompt: **Help Teachers Get What They Need**

Partner:
The Bill & Melinda Gates Foundation

Challenge Questions
Can we figure out what teachers need most based on teachers' responses on what digital tools are most essential to help their students be prepared for college and careers?

How do classrooms differ in their needs? Are there natural clusters of types of classrooms that have different needs in digital tools?

Data
Raw survey data of 3,100 educators on digital instruction tools

Datasets and details:
http://collegeready.gatesfoundation.org/article/teachers-know-best

Impact
As part of the College-Ready Education program, the Gates Foundation has dedicated $20 million in funding innovative next generation learning programs in recent years. By helping the Gates Foundation identify what teachers need most and how needs vary by types of classrooms, your hack can help teachers get access to the tools they need to succeed.

Data Prompt: **Fight Child Sexual Exploitation**

Partner:
Thorn – Digital Defenders of Children

Challenge Questions
Can we model / visualize demand for the escort market through new pricing information, and use this to gain insights into sex trafficking?

Can we pull out messy, unstructured pricing data from heavily masked escort ads?

Data
7M+ rows of data collected from three escort services sites in TSV format

Datasets and details:
https://s3-us-west-1.amazonaws.com/thorn-hackathon-escort/escort_all.tar.gz

Sample:
Rates: Massage rates start at 120 (Newbie Friendly) and Girlfriend rates (Must have references)start at 160
Add'l rates, pics, and time frames Available and listed on my Website
[[Everything]] YOU [[want]] and [[NEED]] right [[NOW]] $60 Specials – 22
~~~~THIS INCREDIBLE PACKAGE-- ONLY$80~~~~$50 FOR 30 MIN.
100 ONE HOUR...
*AmAZiNg CuRVeS .... SiNfuLLy DeLiCiOuS ___ * BL _O_ NDE * ___ Eye CaNdY ....*
$8O__12O__14O SPECIALS - 24

Impact
It is estimated that 100,000 to 300,000 children are at risk of commercial sexual exploitation in the United States and one million children are exploited by the global commercial sex trade each year. Price information, which is often posted within escort services ads, can help us get a better understanding of demand in the online sex advertising market. Determine how listed price has changed over time and differs from city to city, to help paint a picture of this marketplace and help investigators stop the sexual exploitation of children.

Data Prompt: **Reduce Accidents in Mines**

Partner:
US Department of Labor / Mine Safety & Health Administration

Challenge Questions
Can we leverage mine inspections reports to determine the likelihood that a mine will have an accident in the future?

Data
The Mine Safety and Health Administration data sets on Accident Injuries Data Set and Inspections Data Set

Datasets and details:
http://www.msha.gov/OpenGovernmentData/OGIMSHA.asp
Sample Data:
https://docs.google.com/spreadsheets/d/1OoWdDKiowIInmm26F8yiDYcz4poQ7LOVxfsj9lzeaek/edit?usp=sharing

Impact
Although the number of mining accidents is at a historical low, we can leverage data science to help make the jobs of miners even safer. There are 40+ mining fatalities annually, and with your hack, your team can help eliminate mining accidents by connecting inspection reports with both accidents and injuries in mines.

BAYES IMPACT

Partner:
Donors Choose

Challenge Questions
Can we see the effects of budget cuts across states / school districts on what teachers are requesting on donorschoose.org?

Data
1 million rows of transaction data on DonorsChoose.org

Datasets and details:
http://data.donorschoose.org/open-data/overview/
https://donorschoose.box.com/s/fkndrb5c9flkomzreqw5

Field	Sample Data
_resourceid	"ff04b9ee3a381a9d7b452e362f29b26f"
_projectid	"0af83109a29611e44b938c0a0d7fd879"
vendorid	
vendor_name	
project_resource_type	Technology
item_name	blank CD's
item_number	GBCDR8050
item_unit_price	15.95
item_quantity	1

Impact
As of December 2013, donors have contributed over $200 million to fund more than 400,000 classroom projects posted on the site, reaching 10 million students in public schools across the United States. Your team can help Donors Choose provide more resources for more teachers across the country.

BAYES IMPACT

Data Prompt: **Help Police Prevent Domestic Violence**

Partner:
High Point Police Department

Challenge Questions
Can we predict when a particular person prone to domestic violence will engage in abuse?

Can we identify factors of domestic violence recidivism and predict risk of reoffending?

Data
Calls for service and domestic violence arrest data from HPPD

Datasets and details:
https://www.dropbox.com/sh/67s6lvildk7rut5/AAB_THCeUvViZjuXdLT80JEFa?dl=0
https://docs.google.com/document/d/1_IMCZ0mGpd-Stx9BwfVnAsNW1J5GnBv_Vj8XHNHu2LY/edit#heading=h.fetx2h87t3y5

Impact
Statistically speaking, 1 in 4 women will experience domestic violence during her lifetime. Women experience more than 4 million physical assaults and rapes because of their partners, and men are victims of nearly 3 million physical assaults. Help the High Points Police Department and law enforcement everywhere to leverage data science to prevent domestic violence.

BAYES IMPACT

Data Prompt: **Parks and Rec**

Partner:
US Department of Interior

Challenge Questions
Can existing datasets help us identify activities OR serene locations in public land and water to help generate health benefits for people (e.g. PTSD alleviation for veterans)? Can we measure this health benefit?

Data
Mock patient data that allows you to simulate the interactions of a health care practitioner

Geospatial acoustic data from major national parks across the United States

Datasets and details:
https://drive.google.com/folderview?id=0BztyQrfwnsJHVWJnTUtNODhYNEk&usp=sharing
http://knowledge.amia.org/onc-ntdc/vista-1.379752
http://www.blm.gov/wo/st/en/prog/blm_special_areas/NLCS.html
http://bison.usgs.ornl.gov/#api
https://ridb.recreation.gov/

Impact
Millions of Americans live within close proximity to national lands and waters but don't realize the free benefits that exist by enjoying the great outdoors. This can be particularly compounded by the perception that visiting these areas can be difficult, expensive, or too time consuming. Difficulty in understanding regulations surrounding useage compounds the problem.

Furthermore, there are numerous public health challenges that can be mitigated through better exposure to the outdoors. Active lifestyles can contribute to a decrease of chronic illnesses like obesity and diabetes; experiencing the solitude of nature can lower stress and enhance overall mental wellbeing.

BAYES IMPACT

Data Prompt: **Ebola – Mining Available Data**

Partner:
Internews

Select Challenge Questions
Can we create a map that shows where certain rumors about Ebola are being spread?
Can we use a visualization that shows us the most reported/predominant rumors?

How many social mobilization projects regarding Ebola are mentioned and which areas
are they covering? Can we create a map that aggregates all the information about
Social Mobilization and that shows us the coverage of each region in Guinea, Liberia
and Sierra Leone? And overlay it with population, clinics location and media location?

What information about Ebola is circulating in the social media sphere? Who are the
most influential sources of information? How predominant (or not) is the diaspora in the
online conversation about Ebola?

Data
Various Ebola data sets

Dataset:
https://docs.google.com/spreadsheets/d/19hfIVeOlkaBTzTEzgpGuSCDypYgxB0VRhTd
gflmareA/edit?usp=sharing

Impact
Ebola is a rare and deadly disease caused by infection with a strain of Ebola virus. The
2014 Ebola epidemic is the largest in history, affecting multiple countries in West Africa.
The risk of an Ebola outbreak affecting multiple people in the U.S. is very low. Use the
mine of available data provided by Internews and help the fight against the Ebola
epidemic.

BAYES IMPACT

Data Prompt: **Prepare California for Elderly Health Needs**

Partner:
California HealthCare Foundation

Challenge Questions
How prepared are we to accommodate the growth in the elderly population in California? How can we determine what districts are lacking in nursing home services, and determine which districts are lacking in high-quality nursing home services?

Data
The data set provides quality scores for all hospitals in California (except children's hospitals), broken out into various categories: patient experience, readmissions, patient safety, heart conditions, mother and baby, and lung conditions. Participants may come up with their own analyses/research questions based on these data.

Dataset:
https://data.cacs.org/browse?category=Health&q=California%20Nursing%20Home&sortBy=relevance&utf8=%E2%9C%93

Impact
California is on the cusp of a substantial growth in its elderly population, which raises the question of whether nursing homes in California are located in areas of need, both now and based on projected future growth (that is, are nursing homes in locations with the fastest growing aging populations?). Your team can help CHCF determine now and future needs of nursing home care for the elderly population in California.

Data Prompt: **Predicting Climate Threats**

Partner:
National Oceanic and Atmospheric Administration

Challenge Questions
Can we determine which climate threat a particular town/city/region/neighborhood is most at risk of being affected by?

- *"Combine different climate-related datasets (drought prediction models, sea level rise, river forecasts, extreme weather, food shortages, etc.) to determine standardized risk levels for each type of threat for each zip code (or other regional breakdown)."*

Data
Various climate-related datasets

Datasets:
http://www.ncdc.noaa.gov/data-access
http://tidesandcurrents.noaa.gov/sltrends/sltrends.html
http://water.weather.gov/ahps/forecasts.php
http://www.cpc.ncep.noaa.gov/
http://www.ers.usda.gov/data-products/food-availability-(per-capita)-data-system/food-availability-documentation.aspx
http://droughtmonitor.unl.edu/MapsAndData/GISData.aspx
http://coast.noaa.gov/digitalcoast/data/sovi

Impact
In 2012, there were 905 natural disasters worldwide, 93% of which were weather-related disasters. Overall costs were US$170 billion and insured losses $70 billion. The human cost of this was substantial: a total of 9,655 people were killed and 124.5 million people become victims worldwide. Your team can help identify what climate-related threats are most likely to affect individual communities across the United States.

Data Prompt: **Assess Flood Risk of Roads and Airports**

Partner:
National Oceanic and Atmospheric Administration

Challenge Questions
What regional infrastructure (roads, utilities, airports, etc.) is most at risk of flooding due to sea level rise, storm surge, or river flooding?

Data
Various climate-related datasets

Datasets:
http://tidesandcurrents.noaa.gov/sltrends/sltrends.html
http://tidesandcurrents.noaa.gov/est/
http://water.weather.gov/ahps/forecasts.php
http://nationalmap.gov/
http://www.eia.gov/tools/faqs/faq.cfm?id=767&t=3
http://www.census.gov/geo/maps-data/data/tiger.html

Impact
Flood damage has increased in the United States, despite local efforts and federal encouragement to mitigate flood hazards and regulate development in flood-prone areas. To help researchers and policy makers assess national progress in reducing vulnerability to flood hazards, reasonably accurate assessments of flood damage risk are urgently needed. Your team can help determine the infrastructure projects that are most needed across the country due to flood damage risk.

Data Prompt: **Prevent Supertankers from Grounding**

Partner:
National Oceanic and Atmospheric Administration

Challenge Questions
What is the likelihood of a supertanker grounding given a particular track?
- *"Given bathymetry (submarine topographical) data, tide data, and wave data, identify points where a large ship is at most risk of grounding."*

Data
Various climate-related datasets

Datasets:
http://www.ngdc.noaa.gov/mgg/bathymetry/relief.html
http://maps.ngdc.noaa.gov/viewers/bathymetry/
http://tidesandcurrents.noaa.gov/tide_predictions.html
http://polar.ncep.noaa.gov/waves/index2.shtml
http://www.ndbc.noaa.gov/
http://polar.ncep.noaa.gov/waves/viewer.shtml?-multi_1-latest-hs-US_eastcoast-

BAYES IMPACT

Data Prompt: **Predict Whereabouts of Types of Fish**

Partner:
National Oceanic and Atmospheric Administration

Challenge Questions
Where should we look to find new stocks of a particular type of fish?

- *"Find temperature, salinity, oxygen, etc. levels that might be associated with places fish like to hang out; cross-reference with existing fishing reports and stock assessments to identify ones that were previously unknown."*

Data
Various climate-related datasets

Datasets:
http://www.st.nmfs.noaa.gov/commercial-fisheries/commercial-landings/annual-landings/
http://www.st.nmfs.noaa.gov/stock-assessment/index
https://www.st.nmfs.noaa.gov/appstech/data-portals/index
http://www.nodc.noaa.gov/General/temperature.html
http://www.nodc.noaa.gov/cwtg/
http://www.nodc.noaa.gov/General/oxygen.html

Data Prompt: **Visualize 911 Service Calls & Police Reports**

Partner:
San Francisco District Attorney's Office

Challenge Questions
Insight into emergency response times and various temporal patterns can be gained from times of 911 service calls and police incident reports. How can these events be best visualized along a timeline that allows filtering and selection according to other attributes of the event records?

Data
San Francisco Police Department Service Calls

San Francisco Police Department Incident Calls for 3 Months

Datasets:
https://s3-us-west-1.amazonaws.com/acs-sfpd-data/sfpd_service_calls.csv
https://data.sfgov.org/Public-Safety/SFPD-Incidents-Previous-Three-Months/tmnf-yvry