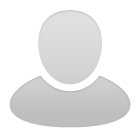
WEEK1

Posted : August 28, 2018 9:02:29 PM EDT  
Last edited : August 28, 2018 9:13:29 PM EDT posted 16 minutes ago (last edited 5 minutes ago)

**Anthony Pagan**

**Infant mortality rate as an indicator of population health**

[**Collapse**](https://bbhosted.cuny.edu/webapps/discussionboard/do/message?action=list_messages&course_id=_1608093_1&nav=discussion_board&conf_id=_1732969_1&forum_id=_1807667_1&message_id=_28521974_1)

Top of Form

**Overall Rating:**

* 1
* 2
* 3

4

* 5

In the article referenced below, the infant Mortality rate as an indicator of population health has its pluses and minuses. As a proxy measure it fall short as it focuses on a specific age range of the population and postulates that the proportional analysis would apply to the whole. In comparison, the disability adjusted life expectancy applies to the whole population age ranges. As a plus, IMR is a less costly analysis and may be the only go to metric available to lesser developed countries.

I agree with the article, the IMR may cause a country to appropriate funds to a smaller segment of the population, which will result in reducing the IMR while overlooking the health needs of other age groups. In the short term it would seem to fall short, but in the long term, as that age group itself ages, the long-term effect may result in positive health affect for the whole.

This begs a moral question. Would a country that can afford to use the costly DALE analysis be morally wrong to appropriate funds based on the IMR study that have the disadvantages of benefiting one age group while neglecting the rest. I think there is a fine line between proxy measures used for more evidence and proofs vs proxy measures used to gear results towards a specific interpretation and agenda.

Reference: <https://jech.bmj.com/content/57/5/344>

Bottom of Form

WEEK 2

This article touches on another chapter of human evolution. From agrarian to industrial, to technological we have always found a way to survive and advance. This new robotic age is another phase like any of the others. As the article mention, people skills will be important, but even more valuable will be people skills with technical know-how. This new phase will bring opportunities in engineering and maintenance of these advance technologies and equipment. Embracing changes is the best approach, doing otherwise stagnates progress.

WEEK3

Our team had a need to notify a list of 100+ DBA teams when their agents were misconfigured. The challenge was the agent aggregation details were on an application that was not integrated or linked to the SQL server inventory and contact information.

1. The solution was to extract the SQL server details with a SQL script to a Sharepoint location
2. Extract the agent details with a Powershell Script to same Sharepoint location. (The agent application has Powershell integration)
3. Create an automated orchestrator runbook from the extracted data that:
   1. Selected sever names, agent details and support group that matched the agent data and the SQL data, where agents were misconfigured and joined columns based on sever name.
   2. Emailed Report with failed agents to the linked DBA team
4. Built a Qlikview Dashboard report that pointed to the csv files at the sharepoint location and made it available on the Sharepoint site for further analysis.

WEEK4

Human nature wires us to try to find what is wrong and try to fix it. We look at the bad to make things good or better. The concept of this article goes against our natural instinct to fix what is wrong. It is difficult for many to grasp. The idea is to not focus what is wrong, but what is right and improve on that.

Data is important in uncovering bright spots that are all around us. In business, franchises will not be successfully replicated unless their initial launch is successful. Reviewing financial data on prospective franchise can help uncover trends and opportunities. Other bright spots that can be uncovered by data would be product sales. If a product is selling at a large profit and 3 other products are selling at a small profit, the data may show that doubling down and putting 100% focus on 1 product would be more profitable then selling 4 products.

Week 5

The link below has examples of bad data. Some issues include spaces in words, % and $ signs in the numbers, and other punctuation marks that would need to be removed. It also included an example with vertical numbers. All if these issues would need to be tidy’d up before being able to use for analysis.

<http://okfnlabs.org/bad-data/>

Week6

This Forbes article highlights the following points on collaboration studies:

* Teams who collaborated stuck to their tasks 67% longer
* Companies the promote collaborative work were 5 times a likely to be high performance
* The key to productive collaboration was purpose.
* Teams tend not to collaborate when there are no incentives
* Having the right culture is key continued collaboration in trying times

<https://www.forbes.com/sites/adigaskell/2017/06/22/new-study-finds-that-collaboration-drives-workplace-performance/#4b136f793d02>

To sum up the 2 articles in the assignment and this Forbes article I would say the key to successful collaboration is communication and purpose. As individuals we all want to contribute and achieve goals. In a group the purpose has to mean something to everyone involved and roles have to be defined. I’ve been on conference calls where there are 5 individuals with 5 agendas and nothing is accomplished. The only way for progress is to agree on the agenda collectively. Once agreed, real progress begins.

Week 7

<https://techbeacon.com/app-nirvana-when-internet-things-meets-api-economy>

In the techbeacon article, it discusses how Internet of Things (IoT) has resulted in an explosion of data. With smart devices that utilize Nest applications and smart watches that can help keep track of eating habits, heart rates and activity; there are APIs that allow these devices to communicate with each other and post data on cloud services and web-based URLs.

The addition of the data from these devices will make that world more connected and informed than it already is. This is good and bad. It is good when the data is used to better the collective, like medical research or to bring light to social issues. However, since business sole purpose is to be profitable, this data explosion very likely will require a review of current privacy laws.

Week 8

<https://trends.google.com/trends/explore>

As the largest search engine /data company in the world, this freely available public data can be an excellent resource to begin any data analysis. The google trends site give rank statistics of google searches. The searches can be broken down by:

* Country
* Date
* Category
* Web search Type

The site includes percent changes of trends and direction, which can be used to analyze behavior of large populations, giving a view of worldwide collective interests. Businesses can use this data to get a pulse on demand for products and services.

The gtrendsr R package can be used to scrape data from Google trends. The articles below provide some how-to examples.

<https://www.r-bloggers.com/analyzing-google-trends-data-in-r-2/>

<https://www.r-bloggers.com/download-and-plot-google-trends-data-with-r/>

Week 9

<http://besthistorysites.net/>

The best of history site target educators and students. The target user goal is to find a central source for historical documents and references.  The site breaks down history by region, historical era and type. It also includes maps, research, games and animations.

To cater to their users more, I would recommend the site add more interactive content and some social media capabilities.  It seems like it would attract educator and avid readers of history, but will miss the mark with students. Students are more accustom to interaction and multimedia content. Adding more interactive content may spark the students interests.

Week 10

<http://www.academia.edu/37446669/EPIDEMIC_OUTBREAK_PREDICTION_USING_ARTIFICIAL_INTELLIGENCE>

The academia link describes how an analysis from twitter APIs was used to predict epidemic outbreaks. Python was used to extract the data on a Mongo db residing on HDFS file system.

The approach was to extract tweets and use natural language processing and sentiment analysis to predict areas where epidemics might happen. They used 11 steps to pre-process data, which included removing stop-words, whitespaces, expressions, punctuation…etc. Part of preprocessing included comparing positive/negative/neutral tweets. After cleaning the data they used machine learning algorithm for classification of tweets. The end result was the ability to predict location of epidemics.

I would say a similar approach can be applied to any crowd think ideas:

* Consumer affinity of products and services
* Shift in voter sentiment
* Shifts in world views and tolerances

Week 11 Recommender Systems (Fake News)

<https://blog.dataiku.com/fake-news-and-filter-bubbles>

The dataiku link discusses how recommender systems for sites such as Google and Facebook, which are supposed to give users recommendations on news and products based on their behavior, are also creating filter bubbles.

Filter bubbles have a tendency to warpi user views and perspectives. It isolates users from other points of view outside of their bubble. This create silos of like-minded people who are fed biased information. Although the site intention is to cater to user preferences, it has been exploited by fake news sources.

Facebooks is a recent good example. Below are a few links on how Facebook recommender systems work at the low level

<https://code.fb.com/core-data/recommending-items-to-more-than-a-billion-people/>

<https://pdfs.semanticscholar.org/eb95/7789f53814a290bc0f8bb01dd01cbd0746cc.pdf>

At a high level, Facebook recommendation system can be described with the performance analysis

1. Who are the users?

* Facebooks user base is anyone on the planet that has a computer and can connect to the internet

1. What are their goals?
   * A Facebook user is looking to stay connected with friends and families and to make new friends. They want to share their experiences and let their friends and families know they are connected. Facebooks official mission is “**to give people the power to build community and bring the world closer together.**”
2. How can you help them accomplish their goals?

* This brings us back to the dataiku article. Facebook has been in the headlines the past few years. It’s inability to filter content of unverifiable news and allow false news sources to reach its users base has caused confusion. This miss information has divided online communities into silos of warped thinkers. This is damaging company’s image of a bringing the world together to one the build silos of clustered ideologies.

As a results, Facebook has been attempted to correct its shortfalls by using user surveys to were built so users can tell Facebook which news is good or and which news is bad.

<https://www.wired.com/story/facebooks-latest-fix-for-fake-news-ask-users-what-they-trust/>

This approach may fall short of expectation, as this explicit user survey may be misleading and fail to capture honest, truthful information. I would recommend two-pronged verification approach. They can stick to the surveys as one source, but should add a trusted third-party source as a combined verification. The trusted source would be similar to what SSL certs do for trusted internet sites.

The key here is to use multiple sources for verification and then use algorithms as an additional check to identify any anomalies. Sources should not be posted until the trifecta of checks have been completed.

Week11 Final Project Proposal

<https://data.world/cityofchicago/array-of-things-locations>

I am interested is analyzing IOT data and understanding how it can help better societal needs. The Array of Things project in Chicago is trying to use sensors to do exactly this. Their approach is to place sensors throughout the City to help understand the City’s health. The sensors can track air quality, road conditions, traffic patterns as well as other data. The goal of the Array of Things Project is to enhance the quality of life of its population.

For my final project I will explore this data via the site API to determine if it can be used as a proof of concept to help other communities in other parts of the US.

Week12

I found the video quite interesting. Although the idea is quite logical. Weather it is an entity or an individual, the company kept or connections made can mold who you become. The social network analysis approach is to find these connections, determine the behaviors at edges that were produced by the vertices at the endpoints, and extrapolate predictions based on the data.

I couldn’t help but think about the current state US politics. Would it be too much to ask the question, what are the connections that molded the person that is asking for your vote? Having a publicly available social network analysis of politicians before they run for office would be interesting. Just knowing politicians by what they say is not enough. Knowing who they know digs deeper and can help weed out the undesirables. Once that is accomplished, we can put focus back on the important things.

Week14

Code First vs Data First Design

There are plus and minuses when choosing Code first vs Data first. It all depends on what the project objectives are. If a project is user centric, where the interface needs to be easy to use and not a lot of data is being stored , going code first has its advantages of a quick turn-around. If the project is data centric, data first would be best approach. The first steps should always be to get all the requirements from the user then decide on the best methods.