

Assignment -1

Part 1: Answer the Following Questions:

1) What was the problem with the google flu detection algorithm?

Ans: The main problem was that it was predicting more than double the proportion of doctor visits for influenza like illness as it could only measure what people search for; it didn't analyze why they were searching for those words. By removing human input, and letting the raw data do the work, the model had to make its predictions using only search queries from the previous handful of years. GFT was built to ease the work of CDC, but it was not able to predict the chances of getting flu is high or not and this becomes creates issue.

2) What is big data hubris?

Ans: Big Data Hubris is assumption that big data are substitute for traditional data and analysis rather it is supplement. Enormous data quantity does not mean that one can overlook one foundational issues of measurement and make validity and reliability on that data.

3) What approach could have been used to improve the Google flu detection algorithm?

Ans: By combining GFT with other near-real-time health data we can obtain greater values. For instance, we can achieve significantly good performance by combining GFT and decreasing CDC data. Moreover, we can also do recalibration of GFT dynamically.

4) What is “algorithm dynamics?”

Ans: It is defined as optimization of system which means changes made by engineers to improve business opportunity and attract customer to use that system over the period.

5) What aspect of algorithm dynamics impacted the Google flu detection algorithm?

Ans: Few of the reasons because of which GFT was impacted by the algorithms dynamics are:

- Google was using google correlate which allows users to identify search data that correlate with given time frame. Although, it was limited to national data where GFT was built on correlation data of regional level.
- One more thing is there is interesting difference between CDC and GFT data. Like searches for treatments for the flu, searches for information on differentiating the cold from the flu track closely with GFT's error.

Part 2: S3 Bucket Creating Bucket


[Amazon S3](#) > Buckets

► **Account snapshot**


Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

[View Storage Lens dashboard](#)

Buckets (1) [Info](#)

 [Copy ARN](#) [Empty](#) [Delete](#) [Create bucket](#)

Buckets are containers for data stored in S3. [Learn more](#)

< 1 > 

	Name ▲	AWS Region ▼	Access ▼	Creation date ▼
<input type="radio"/>	a20502846-csp554	US East (N. Virginia) us-east-1	<u>Bucket and objects not public</u>	August 31, 2022, 12:22:36 (UTC-04:00)

Adding Files to the Bucket


[Amazon S3](#) > [Buckets](#) > [a20502846-csp554](#)

a20502846-csp554 [Info](#)


[Objects](#) | [Properties](#) | [Permissions](#) | [Metrics](#) | [Management](#) | [Access Points](#)


Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

 [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions ▼](#) [Create folder](#)

[Upload](#)

< 1 > 

<input type="checkbox"/>	Name ▲	Type ▼	Last modified ▼	Size ▼	Storage class ▼
<input type="checkbox"/>	 CSP 554 Assignment 1 v3.docx	docx	August 31, 2022, 12:24:00 (UTC-04:00)	24.4 KB	Standard