

# Assignment 1

Atharva Tanaji Kadam

A20467229

## Question 1:

The problem with Google Flu Trends(GFT) was that it was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC). The big data was overfitting the smaller number of cases. The two major issues that contributed to GFT's mistakes were big data hubris and algorithm dynamics.

## Question 2:

Big data hubris" is the often implicit assumption that big data is a substitute for, rather than a supplement to, traditional data collection and analysis.

## Question 3:

By combining GFT and lagged CDC data, as well as dynamically recalibrating GFT, we can substantially improve on the performance of GFT or the CDC alone. Combining the Google flu detection algorithm with other near-real-time data can also result in better performance of the GFT.

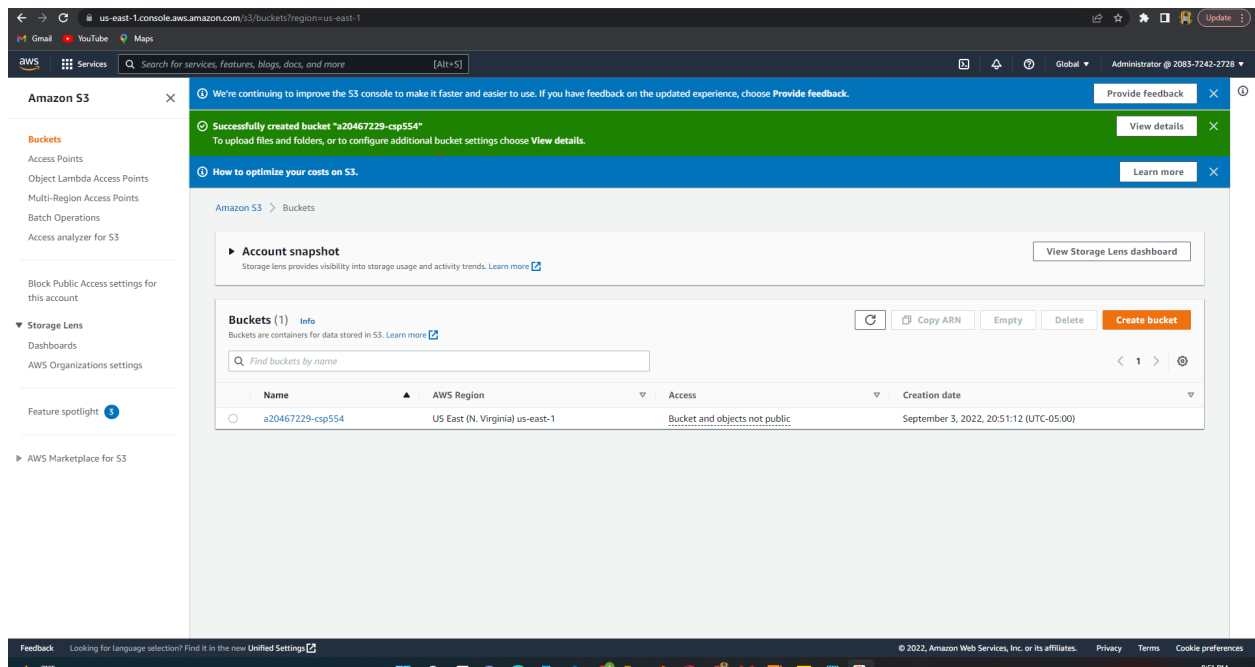
## Question 4:

Algorithm dynamics are the changes made by engineers to improve the commercial service and by consumers in using that service.

## Question 5:

Several changes in Google's search algorithm and user behavior likely affected GFT's tracking. Because GFT uses the relative prevalence of search terms in its model, improvements in the search algorithm can adversely affect GFT's estimates.

### Creating a Bucket:



### Uploading an object to that bucket:

