# Capstone Project Proposal

**“Is that the food talking?”**

**Insight:** We have always heard about the adverse effects that a food product has on humans. But still that popsicle and frito-lay never make a miss in your shopping list. This project shows how these products have caused effects on a person and which industries are one of the causes of it.

Let me introduce you to my dataset. This dataset is obtained from the U.S. Food and Drug Administration (FDA),Center for Food Safety and Applied Nutrition (CFSAN), Office of Analytics and Outreach (OAO. The CFSAN Adverse Event Reporting System has complied a dataset have complied the adverse effects that the food products have caused, their symptoms and what was the action taken. It is from a time of 2004 to 2016.

The data set has the following columns: "Report\_number","Report\_Created\_date","Event\_Occured\_Date","Product\_Role","Product\_Name","Industry\_Code","Industry\_Name","Age\_Event","Age\_Event\_Unit","Gender\_Customer","Action\_Taken","Symptoms" columns. A lot of them have missing values because consumers fail to report their demographics properly or would not have bothered to report to CFSAN. For all the food industries out there, As even the FDA says, there is no valid evidence that only these products caused the adverse effect but can also be a mere coincidence and cannot evaluate the risk associated with the product however we can have a clear understanding as to what symptoms are associated to which industry and there by the product.

The client that is my food industry can later make a detailed report on if the consumers faced the effects due to products or due to a previous medical condition, coincidence by introspecting the consumers. The client now has an idea as to what ask and report for so a thorough analysis can be done by the FDA and determine the root cause of the symptoms and how Industry plays a role.

I would like to see few explanatory analysis on the data, basically if male or female reported more, difference between the event date and the reported data, what age group the adverse effect has its impact more, what each industry produces and so on. It would be interesting to perform text classification on the symptoms data where given a symptom which industry is more likely to be associated with. It would be an interesting insight to have cluster analysis on the industry products and which industry has favored more towards the effects.

My project would typically include a detailed analysis of the subject, code, report and slides for perusal. Excited as it sounds, this project sharpens the learning curve of budding data scientists.