**Gerald Jean Baptiste**

1)Prostate cancer diagnosis using the Gleason grading system

I got My first idea after a doctor visit because I had checked my PSA level. Unfortunately, the level is a little bit high due to family history. Almost every male in my family has the PSA level raise at this age. So, with some curiosity to know more about the disease I decide to dive in deeper to understand what is going on with my PSA. The dataset I am considering deals with cancer tissue sample place on a grading scale called the Gleason grading system After the biopsy is assigned a Gleason score, it is converted into an ISUP grade on a 1-5 scale. The Gleason grading system is the most important prognostic marker for PCa, and the ISUP grade has a crucial role when deciding how a patient should be treated. There is both a risk of missing cancers and a large risk of over-grading resulting in unnecessary treatment. However, the system suffers from significant inter-observer variability between pathologists, limiting its usefulness for individual patients. This variability in ratings could lead to unnecessary treatment, or worse, missing a severe diagnosis. The dataset is in CSV format.

Data Source: [Prostate CANcer graDe Assessment (PANDA) Challenge | Kaggle](https://www.kaggle.com/competitions/prostate-cancer-grade-assessment/data)

2) How to improve restaurant rating using yelp user reviews

The Yelp.com challenge sparked the idea for this design. The challenge asks participants to identify the most often used words in both positive and negative reviews of various business categories (such as restaurants, cafes, and so on). I'd like to investigate further into the most used words in reviews, not just in company categories, but also in a few other categories such as neighborhood, time, and rating. Customers may expect a specific service in addition to food while dining at a very high-rated restaurant, and if the business fails to meet their expectations, they express it in their ratings; nevertheless, a customer may not expect something similar when dining at a low-rated restaurant. Furthermore, consumers from various communities are more likely to be diverse.

The data comes directly from Yelp.com and is grouped into six main JSON format data types (company, reviews, user, check in, and tips). All of the datasets' details are available at the following link: Source of data: https://www.yelp.com/dataset/documentation/json

The third concept is an examination and insights of the Ukrainian conflict. The Russo-Ukrainian War is an ongoing conflict between Russia and pro-Russian separatist groups in Ukraine. It was launched by Russia in February 2014, following the Ukrainian Dignity Revolution, and initially focused on the status of Crimea and the Donbas, both of which are internationally recognized as part of Ukraine. The Russian annexation of Crimea (2014) and the war in Donbas (2014-2022) between Ukraine and Russian-backed separatists, as well as naval incidents, cyberwarfare, and political tensions, characterized the first eight years of the conflict. Following a Russian military buildup on the Russia-Ukraine border beginning in late 2021, Russia launched a full-scale invasion of Ukraine on February 24, 2022. The invasion resulted in Europe's worst refugee crisis since World War II, with over 8.8 million Ukrainians leaving and a third of the population displaced. The invasion also resulted in worldwide food shortages. This data is in Jason format data types. Here is the link for the data:[2022 Russia Ukraine War | Kaggle](https://www.kaggle.com/datasets/piterfm/2022-ukraine-russian-war)