**Title:** How does smoking impact lung cancer?

**Team members:**

Achempim, George; Scarth, Nicolas; Mincucci, Nick; Moody, Madison

**Project Description**

In thisproject we aim to investigate demographics and other factors that surround smoking and lung cancer. We will use different graphs and statistics to glean insights from our dataset.

**Objectives**

1. Determine the distribution of demographics such as age and gender within our dataset
2. Evaluate how smoking affects the onset and outcome of lung cancer.
3. Assess factors that contribute to length of survival after treatment such as treatment type.

**Research Questions**

1. Are there any demographics within the dataset that demonstrate higher risk of lung cancer?
2. What impact does smoking have on the onset of lung cancer?
3. Does smoking affect tumor size?
4. Does smoking affect tumor location?
5. Does smoking affect the stage of cancer?
6. What is the average length of survival after treatment?
7. Which treatment has a higher survival rate?
8. Does smoking affect survival rate?

**Data Set to be Used**

We will be using dataset found on Kaggle. The data set is named Lung Cancer Prediction which has 38 columns and 23659 rows. It includes data on demographics, smoking history, treatment, tumor information, and survival length.

**Task Breakdown**

1. Use Pandas to clean data
2. Create Jupyter notebook to describe our process through the clean up and data exploration
3. Create Jupyter notebook to illustrate the final data analysis through different plots and graphs using Matplotlib
4. Save the visualizations in an appropriate format for class and instructional team
5. Write-up summary of major findings that includes headings for each question, a brief description of the findings, and any relevant plots.
6. Create slideshow for presentation.