First, we looked at personal characteristics, to see if any group was more susceptible to getting lung cancer. Using characteristics of the patients in the dataset, no one factor stands out as impactful on lung cancer rates. Ethnicity, gender, and family history are evenly distributed. Age is almost normally distributed with a mean of about 54 years, however there is a slight tick up past 75 years old.

Then, we evaluated the relationship between smoking status and lung cancer. The dataset broke down patients into three categories based on smoking status, current smokers, former smokers, and non-smokers. All three groups had a mean age of 54, meaning that smoking history does not impact the onset of lung cancer.

Furthermore, we analyzed the relationship between smoking pack years and tumor size. Smoking pack years is a way to determine how much a patient smoked, one pack a day equals one smoking pack year. The relationship between smoking pack years and tumor size has an r2 value of 0.0001, highlighting that the two variables are completely independent.

Next, we investigated the relationship between smoking and tumor location. There are three lobes where cancer could be found. Smoking also does not appear to affect where the tumors are located within the patient’s lungs. The distribution of lobes based on smoking pack years was almost perfectly even, showing that the two variables are not related.

Additionally, we explored the relationship between smoking pack years and stage of cancer. All four stages of lung cancer were evenly distributed based on smoking pack years, leading us to determine that smoking pack years has no correlation to the stage of cancer the patient has.

Then, we analyzed the investigated the average survival rate so we could look more into smoking’s impact on survival. The mean and median survival rate was 60 months after treatment for the patients in the study.

Next, we examined the relationship between treat type and survival rates. The study had four treatment types. The four treatment types had almost identical means, medians, and standard deviations. Due to this, we determined that treatment types did not have an impact on survival.

Lastly, we analyzed the relationship between smoking pack years and survival in months. The r2 value of this relationship was 3.36070175723838e-08. This means that smoking pack years and survival have no correlation, which means that smoking has no impact on survival after treatment.

Overall, we determined that smoking has no impact on lung cancer, survival rates, location of cancer, and cancer size. Our comprehensive analysis of the dataset on lung cancer indicates that smoking history and current smoking status are not relevant indicators for any of the metrics studied.