**LUNG CANCER RISK PREDICTION: A ML END-TO-END PROJECT**

*This project raises awareness of lung cancer and risk factors*

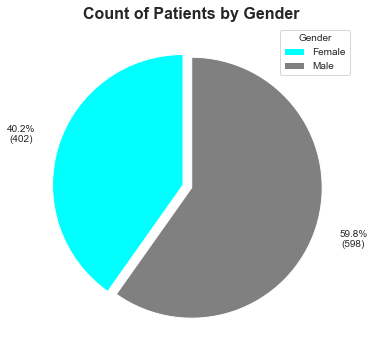
Lung cancer is the build-up of too many cells in the lungs which can lead to the blockage of major airways. Usually, cells in the body die at a certain stage in their life cycle to prevent this build-up, but there are abnormal cases where cells continue to grow and multiply leading to tumours that can either be benign (not cancerous/harmful) or cancerous (harmful and life threatening).

What makes cancerous tumours harmful is that they spread into, or invade, nearby tissues and can travel to distant places in the body to form new tumours. Benign tumours, on the other hand, do not spread from their origin, making them easier to remove from the body.

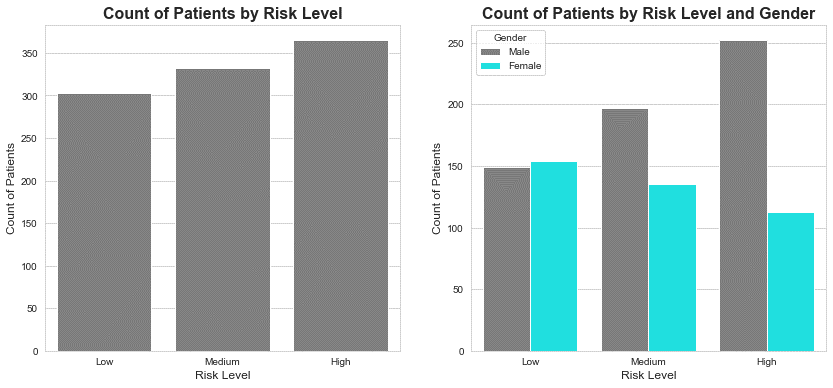
Cancerous tumours are often called cancer. Lung cancer is the second most common cancer in the world and the leading cause of cancer deaths worldwide The World Health Organization (WHO) recorded 1.80 million lung cancer deaths in 2020. Meanwhile, breast cancer- the most common cancer in the world- caused 685,000 deaths in 2020.

In many research studies, early detection of cancers has been linked to increased chances of survival. This is the aim of this project; to aid in the early detection of lung cancer using a trained machine learning model to predict the risk level of lung cancer in people. In this way, the number of deaths caused by lung cancer might be reduced.

Patient data from [data.world](https://data.world/cancerdatahp/lung-cancer-data) was used to train the model. It has 1000 rows and 24 columns. About 60% of the patients are males, while 40% are females.

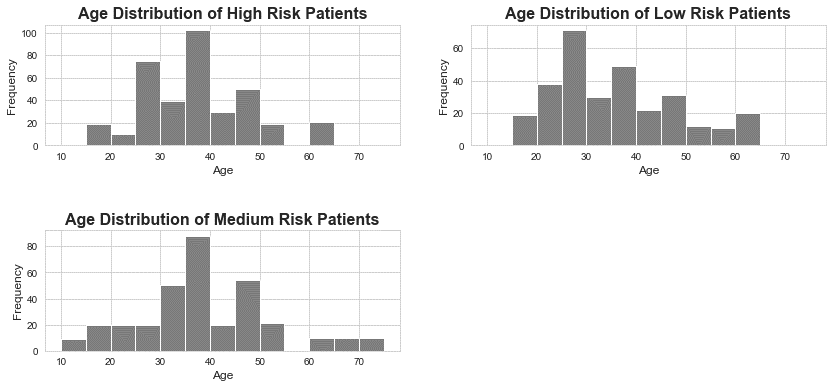


Most of the patients had high risk of lung cancer (365 patients). Approximately 250 (70%) high risk patients are males while about 100 (30%) are females.

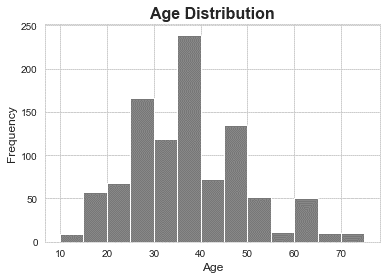


This goes in line with the fact that lung cancer is the most common cancer in men but the second most common cancer in women. Historically, lung cancer has been more prevalent in men than in women.

It is said that most people diagnosed with lung cancer are 65 years or older. But from the data, I saw that most patients with high risk of lung cancer were between 35 and 40 years of age.

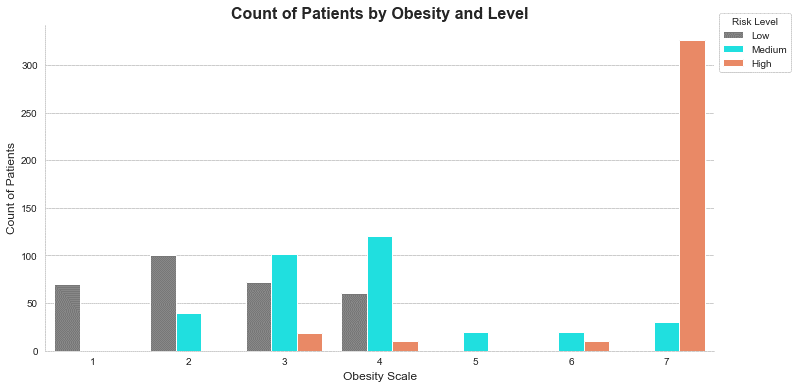


It should be noted, however, that the overall age distribution of patients revealed that most patients fell within the 35-40 years’ age bracket.



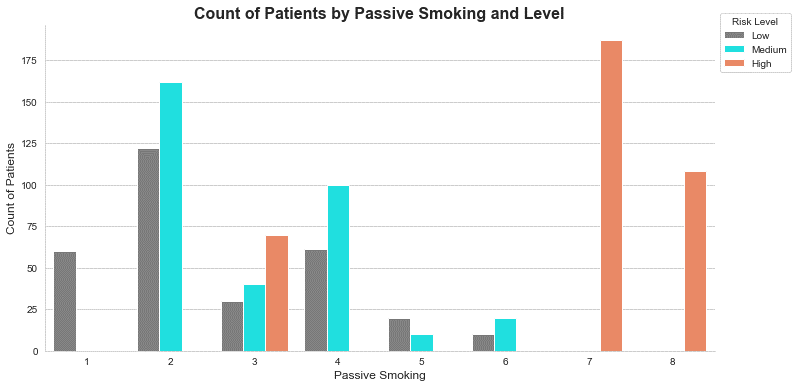
At this point, you might wonder what causes lung cancer or what increases the risk of cancer. Many factors contribute to the risk of developing lung cancer. Some of these factors include:

* Obesity: Obese people have a higher risk for cancer than other people. Research has found that obesity increases the risk for several cancers, including colo-rectal, breast, uterine, esophageal, kidney and pancreatic cancers. For lung cancer, however, the story is different. Being overweight has been associated with a decreased risk of lung cancer. However, in the patient data, most of the patients were greatly obese and had high risk of lung cancer.



This could be explained by the fact that other factors contribute to the risk of cancer, and not just obesity alone.

* Smoking: Smoking is the number one risk factor of lung cancer. The smoke released from burning tobacco is full of cancer-causing substances. When inhaled, these substances damage the cells that line the lungs. The body may attempt to repair this damage but with each repeated exposure, normal cells that line the lungs are increasingly damaged. Overtime, the damage causes the cells to act abnormally and eventually cancer develops. Smokers are not the only persons endangered by tobacco smoke. Anyone who inhales this smoke is prone to lung cancer. This is called passive smoking and is also a risk factor of lung cancer.



With increasing level of passive smoking, the risk level increased.

* Alcohol Consumption: It is not clear whether or not alcohol consumption increases the risk of lung cancer. It is generally suggested that alcohol consumption increases lung cancer risk, however, many researchers are still sceptical about this because of the correlation between alcohol consumption and smoking. Most people that smoke also drink alcohol, so the question is does alcohol really affect lung cancer risk or is it simply due to the associated smoking habit? In certain research studies, it has been found that in non-smokers, low and moderate drinking does not increase lung cancer risk but heavy consumption (particularly beer) increases the risk of lung cancer.

In the patient data, the male patients generally showed higher alcohol intake than the females.