DATA 607 – Final Project proposal

1. Project Objective

Identify the density of the correlation between Lung Cancer with Air Pollution, Tobacco Smoke, exposure to Radon and Radiation, Arsenic in drinking water.

2. Inspiration

My mother was diagnose with a lung cancer in her early 70's but still she live with continues Oxygen Support.
She had a very good healthy village life. No significant exposure to any strong risk factors. No family history of this type of cancer. It is a very rear type in SriLanka.
I understood only two risk factor might be possible. That is due to Arsenic in drinking water because of the chemicals use for agriculture, the drinking water system is getting polluted. The other possible risk identified is natural Radon and Radiation .I don't have any acceptable data sets or any tool to measure them in SriLanka but USA.

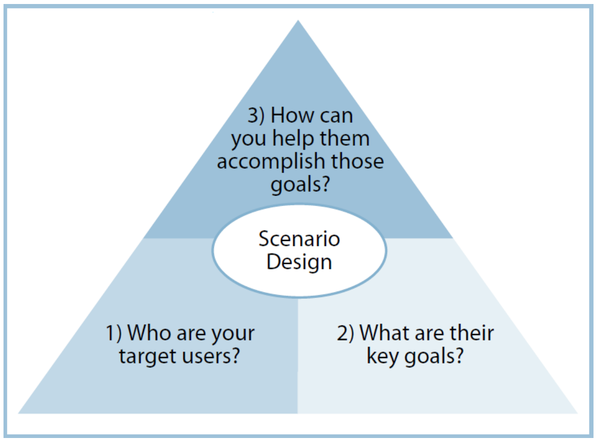
3. Proposed Sources

Stilling I am searching the best available data sources in USA for the risk factors selected and Lung Cancer statistics.

Lung Cancer statistics in the	**Maximum**	**Moderate**	**Minimum**
Cities with Air Pollution	√	√	√
Cities with Tobacco Smoke	√	√	√
Cities with Radon and Radiation	√	√	√
Cities with high Agriculture Industry	√	√	√

Risk Factor statistics in the	**Maximum**	**Moderate**	**Minimum**
Cities with Lung Cancer	√	√	√

4 Proposed Methodology


3) How can you help them accomplish those goals?
Scenario Design
1) Who are your target users?
2) What are their key goals?

After collecting enough data sources I think it is important to do some source validation and accuracy check. That leads to increase the accuracy and confidence of my final findings.

Data extraction, loading, transformation, modeling and storing will be done with the learnt R packages and MySql.
According to statistical analysis and different scenarios I will do the necessary data modeling.

Next with the analysis and plotting can identify the correlations and their densities.
Finally the conclusion.