1. Prepare a business case for the use of data analysis in an organization.

Objectives:

1.1 Identify relevant questions and objectives by engaging your client.

1.2 Select an appropriate open data source to analyze.

1.3 Prepare a statement of intent for the use of data analysis in an organization.

1. 2. Write a data analysis proposal.

Objectives:

2.1 Develop a project-appropriate data analysis proposal for the organization.

2.2 Demonstrate how data analytics can support business needs.

2.3 Describe data and information and how they interrelate.

Project Proposal – Heart Disease

Data analytics supports business needs by transforming data into information. Data by itself is not very useful; it is simply a large assortment of facts grouped together and as such, context that allows intuitive business decisions do not come easily. Through data analysis, data is structured, processed, interpreted, and presented such that it becomes meaningful facts – or information – thus allowing a more informed, more practical decision making process.

The intent of this piece is the model and analyze medical data pertaining to heart disease in hopes to predict the issue of a patient’s heart. Doctors typically rely on ECG graphs and visualizations from various tests to make informed decisions on a patients’ best interest based on their domain knowledge, which can be subjective and perhaps sometimes unreliable at the worst. The use of data analytics will provide them a suggestion based on medical data, allowing a decision making process that is based on objective predictive modeling.

Because machine learning models are such black boxes, using them to prescribe expensive treatments, such as surgeries, or the lack thereof, can be difficult to implement. Therefore, instead of using data to replace the normative process, data can be used in a more supportive way: by using them to supplement the diagnosis, the use of these models is easier to justify, and the decision itself is perhaps more useful; it reduces the subjectivity on a doctor’s interpretation of the data, allowing faster, and more importantly, more precise treatment of the patient.

While heart disease is caused by a myriad of environmental, hereditary, and self-induced factors such as stress, weight, family history, smoking, and more, this piece seeks to see whether or not predictions can be made solely on medical data available, and if these factors affect its predictability.

The dataset is a reduced set that comes from UC Irving and the Cleveland database. The main predictor for heart disease utilized in this set will be chest pain experienced, as different kinds of diseases will cause different kinds of pain. The “KPI’s”, or the indicators will be age, blood pressure, cholesterol, ECG results, maximum heart rate, exercise induced angina, ST depression and peak, number of major blood vessels, and thalassemia.

Acceptable/healthy ranges for ea:

* Blood pressure
  + Normal < 120
  + Elevated 120 – 129
  + High, stage 1 130 – 139
  + High, stage 2 > 140
  + Hypertensive > 180
* Cholesterol
  + Total 125 - 200
  + Non-HDL < 130
  + LDL (bad chol) < 100
  + HDL (good chol) > 45

<https://www.kaggle.com/ronitf/heart-disease-uci>

<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>