Topics to be covered:

**Context -** What is the opportunity or gap being addressed? Which audience does this project serve?

* Our team aims to identify and understand the various factors that contribute to low bone mineral density, which can be a significant predictor for fractures. This project serves healthcare professionals, policy makers, and individuals at risk, providing insights to prevent fractures, improve patient outcomes, and allocate healthcare resources more efficiently.

**Domain -** What are the central domain questions to answer and measurable project objectives? How will you measure success?

* The central domain questions include identifying the impact of lifestyle factors, diseases, biomarkers, and health history on bone mineral density and fracture risk. The project's success will be measured by the accuracy and comprehensiveness of the identified factors, the effectiveness of the predictive model, and the practical utility of the developed guidelines or recommendations in a real-world context.

**Data -** What data is driving or enabling (e.g., source, characteristics, provenance, governance, etc.) this investigation?

* Data will be driven by the MROS study, a large study (~6,000 male patients) that began in 2000 with 3 follow-ups per year. The MrOS datasets contain a variety of information, but we will be using baseline lifestyle factors, biomarkers, health history, and disease prevalence will be incorporated to analyze endpoint fracture incidence.

**Outcome -** What is the domain outcome being delivered based on analytical type (e.g., predictive analytics, prescriptive analytics, data mining/knowledge discovery)?

* Our group hopes to predict incidence of fractures in the adult male population based on baseline features.
* Other goals include risk factors, common trends, and exploratory analytics.

**Technology/Algorithms -** from a high-level perspective, which technologies and algorithms will be utilized? (e.g., Supervised/Unsupervised, NLP, ensembles, DL, DT, Regression/Classification, etc.)?

* We intend to use Supervised learning to perform classification
* Decision trees, Random Forest, XGBoost and LightGBM modeling will be explored as these models make for great classifiers and can also appropriately handle NAN values.
* Dimension reduction will need to be performed on the data to reduce dimensionality

**Desired Deliverable -** what is the planned and desired final deliverable data story/product?

* The end goal is to produce a proof-of-concept model that can accurately predict the likelihood of fractures. This model could then be utilized in healthcare settings to guide treatment plans and preventive measures.

**Overall Benefit Justification -** what is the business benefit (direct or indirect) and how can this project be justified in terms of time or performance, finance, reduced risk, or other direct business impact.

* Fractures can severely impact an individual's quality of life and increase their risk of mortality. By developing a model that can predict fractures, healthcare providers can intervene earlier, potentially preventing fractures from occurring. This could lead to improved patient outcomes, cost savings for the healthcare system, and a decrease in the trauma and disruption caused by these events. Additionally, using supplemental datasets will help justify the project's overall benefits from a social and economic perspective.