# Spin #2 Retrospective Document

**Refined Project Scope:**

Our project scope is identifying and understanding the effects of lifestyle factors, medical history, and specific biomarkers on the risk of bone fractures. The emphasis is on osteoporosis-related fractures, a health concern that impacts a growing number of adults in the US each year.

The core of this project involves building a predictive model that can anticipate the likelihood of fractures based on identified risk factors. In addition to predictive analytics, we also intend to incorporate:

Descriptive Analytics: Utilized to examine the general trends, patterns, and correlations in the data.

Diagnostic Analytics: After identifying patterns, we will delve deeper to ascertain the reasons behind these patterns.

Prescriptive Analytics: Once we have a reliable predictive model, we aim to offer prescriptive insights that can guide preventive measures, treatment plans, and policy decisions. This can involve recommending lifestyle changes or highlighting high-risk groups that may benefit from specific interventions.

**Refined Domain Questions**

What is the impact of lifestyle factors such as exercise, diet, and alcohol intake on bone mineral density and the risk of fractures?

What specific biomarkers can be indicative of low bone mineral density or increased fracture risk?

**Refined Statement of Project Goals, Measurables, And Constraints**

*Project Goals:*

Identify lifestyle factors, prevalent diseases, biomarkers, and aspects of health history that significantly contribute to low bone mineral density and the subsequent risk of fractures. The ultimate goal is to achieve high accuracy, precision, and recall in predicting fractures.

Establish a model that predicts fracture risk based on these identified factors.

Develop guidelines for preventive strategies based on these identified risk factors.

*Measurables:*

A comprehensive list of lifestyle factors, diseases, biomarkers, and health history aspects that contribute to low bone mineral density and fracture risk.

A predictive model with quantifiable accuracy and reliability metrics (e.g., accuracy, precision, recall, and AUC-ROC) for assessing fracture risk based on these factors.

*Constraints:*

Data Limitations: The project relies heavily on the MrOS dataset. There may be limitations due to missing or incomplete data. The dataset exclusively contains male records, meaning our findings may likely be limited to males.

Time Constraints: The project needs to be completed within eight weeks, potentially limiting the scope of analyses.

**Relevant KPIs**

* Determine which columns in the datasets are most relevant
* Reduce datasets to relevant columns
* Determine what factors to use to compare different datasets (this may include multiple comparisons)
* Combine reduced datasets to improve ease of visualization
* Create visualizations
* May need to balance dataset – resampling methods under review
* High recall may be of utmost importance for the final model. The ability to predict positive cases of fracture. Our dataset, like the real world, is imbalanced in the case of number of people that experienced a fracture. Dealing with this imbalanced data will be critical.

**Refined Vision for data story**

* **Audience:** Our intended audience includes healthcare professionals, policy makers, and individuals at risk of fractures.
  + Healthcare Professionals: We aim to provide them with a tool that can help them identify patients at high risk and guide them in implementing appropriate preventative measures.
  + Policy Makers: Our insights can help inform healthcare policies and resource allocation decisions aimed at reducing the prevalence of fractures and improving patient care.
  + At-risk Individuals: By making our findings accessible to the general public, we can raise awareness about the risks and prevention strategies associated with fractures, empowering individuals to take charge of their own health.
* **Product:** We will present the results of our analysis, using a mix of charts, graphs, and infographics to visualize the relationships between different risk factors and fracture risk. We will also discuss what our findings mean for different stakeholders, suggesting possible interventions and policies that could help reduce fracture risk.

**Summary of Project Literature Review**

While fracture risk and prediction has been extensively studied, rarely has such a large dataset been open for public use. The team has reviewed the MrOS list of publications and while they cover a wide breadth of topics the dataset is updated regularly which allows for the potential of newly discovered results or to confirm previously performed research using the MrOS dataset but updated for the year 2023 endpoints. The list of MrOS publications can be accessed here: <https://mrosonline.ucsf.edu/PublicFiles/MrOSPublicationsListing.pdf>

Factors such as lifestyle factors (i.e. exercise, diet, and alcohol intake), current diseases (i.e.diabetes, chronic obstructive pulmonary disease, and hyperthyroidism), and health history (i.e.history of previous fractures and stroke, bone mineral density) have been studied as risk factors for fractures, and we will attempt to address these factors in our analysis. Hypertension and arthritis are common results from fractures but do not cause them and therefore won’t be addressed in this analysis. Radiation treatment from cancer and smoking slow repair of fracture but do not cause them and, also, won’t be addressed in this analysis [1,2].

Literature was also reviewed to find decomposition techniques used for regional bone density scans. A study was found that performed PCA on bone mineral density results. The study was able to reduce results to a single principal component that accounted for 73% of the variation in density. The component loadings were approximately equal for each site [3].

**Changes in data storage, curation, and management**

The raw data flat files are stored in the team’s shared folder. From there, the raw data will be uploaded to a Postgres database where it will then be accessed for carpentry, analysis, and our data story. The process of moving data from raw files to Postgres has been completed although during EDA, if we identify the need for additionally supplemental datasets, we’ll return to this step for uploading the data.

**Changes to data provenance and assessment**

The MrOS (Osteoporotic Fractures in Men) dataset, which forms the basis of our project, is a comprehensive and detailed database specifically gathered to understand the factors influencing bone health in older men. This dataset was originally collected through a research study conducted from 2000-2002 at six clinical sites across the United States, and its primary aim was to determine how various lifestyle factors, diseases, and health histories impact bone health, particularly the risk of fractures in men.

The baseline examination of the study included a range of tests such as neuromuscular, visual, cognitive function tests; bone mineral density (BMD) measurements; x-rays of the spine; quantitative computed tomography (QCT) scans of the hip and spine, and the collection of biospecimens. Participants have been surveyed three times a year for follow-ups since the baseline visit.

Our project focuses on the original baseline visit in addition to the latest endpoint data. The MrOS dataset accurately reflects the state of the domain we're exploring – the risk factors contributing to bone fractures, particularly those related to osteoporosis, in older men.

However, the dataset is not without its limitations. Firstly, the data only covers male participants, so any results and models we derive from this data cannot be directly extrapolated to women. The gender-specific nature of the dataset is a known limitation that we're aware of and that we're taking into consideration in our study.

**Address any potential or identified weakness or biases to the data**

The MROS data was collected specifically on men, thus any results cannot be extrapolated to women. Additionally, a portion of the data is Null, and must be handled appropriately.

**Describe Data Carpentry**

* Data import, database creation
* Column headers changed
* Data types applied to database columns
* Nan values removed (as part of EDA)

**Describe EDA and Visualization Performed**

We began by exploring the variables in our dataset, understanding the type of each variable (numerical, categorical), checking for any inconsistencies, and examining their distributions. This provided us with a clearer understanding of our dataset's structure and potential anomalies, such as skewness or outliers.

Visualizations of PCA and Feature Selection were performed primarily with Seaborn. Through the use of histograms, scatterplots, and pair plots we were able to confirm what we had found through our literature review.

**Weekly individual team member accountability/contribution assessment/evaluation included**

Progress tracking takes place through weekly or twice weekly zoom meetings where the team breaks down action items, responsibilities, and progress. In addition to the regular Wednesday team meeting and mentor meeting, all team members have been communicating frequently and extensively on our group site on our Slack channel - casestudy\_su23\_group03.

Contributions for the week (not comprehensive, most areas may well have been worked on by all):

* Karen – SpIn Document and presentation, exploratory data analysis, lit review
* David – Postgres script, importing datasets, fixing data types, preparing “StartHere” file
* Josh -- Postgres script, SpIn Document, fixing data types, BMD exploratory analysis
* Tyler -- SpIn Document and presentation, exploratory data analysis V1 datset, lit review

Weekly Group Meeting notes can be found in our OneNote: [Group Meetings](onenote:https://mailmissouri-my.sharepoint.com/personal/jwj8c8_umsystem_edu1/Documents/SU23_DSA8080%20Casestudy/SU23_DSA8080%20Casestudy/Group%20Meetings.one#section-id={22F8CF55-8C19-4BDF-B116-5B29E397A512}&end) ([Web view](https://mailmissouri-my.sharepoint.com/personal/jwj8c8_umsystem_edu1/_layouts/OneNote.aspx?id=%2Fpersonal%2Fjwj8c8_umsystem_edu1%2FDocuments%2FSU23_DSA8080%20Casestudy%2FSU23_DSA8080%20Casestudy&wd=target%28Group%20Meetings.one%7C22F8CF55-8C19-4BDF-B116-5B29E397A512%2F%29))

**Project work success status evaluated and future work/tasks discussed**

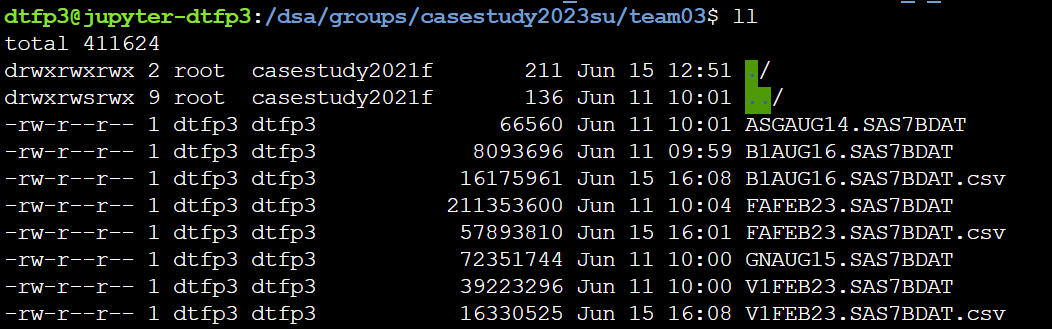
We have successfully completed the initial stages of our project, which include defining the project scope, setting goals and metrics, identifying relevant KPIs, refining our data story vision, and conducting a literature review. We have also addressed data storage and management, and made progress on data carpentry, which involved handling null values and exporting datasets into PostgreSQL. Preliminary EDA and visualizations have begun.

**Jupyter Notebook(s) fully internally documented**

Yes

**All paths to data files map to DSA team shared folders**

Yes



**Current project SpIn artifacts (notebooks) are located in the TeamArtifacts\SpIn\_2\_Artifacts folder (provide link (within Europa) to the first notebook in the pipeline)**

Yes

<https://europa.dsa.missouri.edu/user/ejm301/notebooks/su23CaseStudy_Team03/TeamArtifacts/SpIn_2_Artifacts/SpIn2-StartHere.ipynb>

**Jupyter Notebook(s) execute without exceptions**

Yes

**Link to Mentor recorded mentor meeting and key meeting takeaways provided**

[6-14-23\_Mentor\_Meeting\_Team03.mp4](https://mailmissouri-my.sharepoint.com/:v:/r/personal/jwj8c8_umsystem_edu1/Documents/SU23_DSA8080%20Casestudy/Zoom%20Meetings/6-14-23_Mentor_Meeting_Team03.mp4?csf=1&web=1&e=5RAV9H)

**Bibliography**

1. Unnanuntana, A., Gladnick, B. P., Donnelly, E., & Lane, J. M. (2010). Osteoporotic Fracture Risk Assessment. The Journal of Bone & Joint Surgery, 92(3), 743–753.
2. Lewiecki, E. M., Rosen, C. J., Schmader, K. E., & Rubinow, K. (n.d.). Osteoporotic Fracture Risk Assessment. Retrieved from https://www.uptodate.com/contents/osteoporotic-fracture-risk-assessment
3. Feingold, M., Nelson, D. A., & Parfitt, A. M. (1992). Composite index of skeletal mass: Principal components analysis of regional bone mineral densities. Journal of Bone and Mineral Research, 7(1), 89-96. https://doi.org/10.1002/jbmr.5650070113