1. Healthcare Sales Wizard

2. Owen Randolph, Erik Kreider, Ian Boen

3. The reason these topics interest us is that they can be used to help managers make wise decisions about how to direct a company sector.  Business is a very competitive field, so data-driven decisions are a key to success. Healthcare sales is a particularly interesting domain because there are so many unique variables that come into play when making sales decisions.

The user-facing feature of our project will be a dashboard that can be adjusted and give outputs based on features selected. Our project will be comprised of four ML and visualization features. The following shows the modules, their importance to the user, and the machine learning model used to give output data:

* Sales Lead Scoring and Prioritization
  + Help sales managers and sales reps prioritize and manage their leads to determine which to tackle when, which are most likely to convert, adjust marketing strategy, etc.
  + Model: Multinomial Logistic Regression (High, Medium, Low Priority)
* Sales Territory Optimization
  + Determine the maximum sales potential by territory. Which cities and regions are more likely to buy certain
  + Model: KNN or a more advance classification model like Random Forest
* Product Bundling
  + Determine which products are often ordered together and can more likely be sold together, perhaps in reduced price “bundles”, or for supply chain/logistic purposes
  + Model: clustering using product similarities
* Product Pricing
  + Used for pricing decision-making based on past sales, regions, times, and other factors
  + Model: a simple Random Forest model to determine adequate pricing utilizing product, company size, average cost to implement, etc., maybe and maybe multivariate regression

Web Application Platform: **Streamlit**. We are interested in using this library due to simple and lightweight nature, it’s fast prototyping capabilities, and we will be using static data.

Programming language: **Python.** As the whole team is more comfortable using Python than R, we will by writing our code using Python. We will be utilizing scikit-learn for building the ML models

Datasets: We have two primary datasets that we will be using to create our models: One for sales leads and one for Sales Territory optimization, product bundling and product pricing. These are synthetic datasets created using Python code with randomization features and realistic pattern features added to them.

Other processes we will undertake are data cleaning, data manipulation, and exploratory data analysis methods.