**Presentation Notes**

* **Selected topic** – **Jason**
  + Heart disease – utilizing data & machine learning to determine the main drivers of heart disease
* **Reason the topic is selected** – **Jason**
  + Potential to help people – heart disease is the number one cause of death globally. The ability to understand what factors drive heart disease will allow individuals to monitor their health and spot potential red flags
* **Description of the source of data** – **Dylan**
  + Our dataset was found on kaggle. This dataset was created by combining five other datasets based on eleven common features. The five datasets used were Cleveland, Hungarian, Switzerland, Long Beach (VA), and the Stalog heart dataset. This is currently the largest heart disease dataset currently suitable for analysis.
* **Questions the team hopes to answer with the data** – **Dylan**
  + Data contains 11 features – goal is to answer which features can be used to predict heart disease
  + End goal is for an individual to be able to take the insights from the ML model and input their own health metrics to determine if they should be concerned about heart disease
* **Description of the data exploration phase of the project** – **Yuta**
  + Data was cleaned prior to download. Data was checked for null values, duplicate rows, datatypes issues, etc. prior to being loaded into the model and was confirmed to have no cleaning necessary
* **Description of the analysis phase of the project** – **Yuta**
  + Classification models were considered for use
    - Logistic regression model found to have highest accuracy
    - Deep Learning and SVM were also tested but did not yield as accurate of results
* **Technologies, languages, tools, and algorithms used throughout the project** – **Hammad**
  + Jupyter
  + Python
  + Pandas
  + MongoDB
  + Tableau
* **Result of the analysis** – **Yuta**
  + The following features were found to be the most predictive in determining heart disease:
    - Chest pain type ASY
    - St\_slope flat
    - Fasting BS
* **Recommendation for future analysis** – **Hammad**
  + Expand dataset – more data will yield a more accurate model and allow the train/test groups to include more data in the training
  + More features will open additional options for which models can be used
  + Gender representation in the data – a large percentage of the data was male, obtaining data with a more accurate representation of the actual population would lead to better results
* **Anything the team would have done differently** – **Jason**
  + Expand data or verify results across datasets, add usability for individuals