The goal of my project is to design and develop a wearable fall detection device capable of identifying when a user has fallen and automatically sending an alert. The dataset I will be using for my project is the KFall dataset. I didn't create this dataset so I had to download this dataset from Kaggle. After downloading the dataset to my computer I investigated the data. The dataset consists of two main folders. One folder is called sensor_data and the other folder is called label_data. Within the sensor_data folder are 32 subfolders. Each subfolder is named by subject ID. For example, the first folder in the sensor_data folder is named SA06. SA06 means that this folder is for subject ID 06. Within each of the subfolders are the motion files for the subject. The motion files have a specific naming scheme. The name of a motion file has the subject ID, task ID, and trial ID. For example, S06T01R01 is a motion file found in the SA06 subfolder. S06 means the subject ID is 06, T01 means the task ID is 01, and R01 means the trial ID is 01. There are 34 different tasks. One of the tasks is "Forward fall when trying to get up". This task has a task ID of 23. Within each motion file contains the numerical data. There are 11 columns in each motion file. They are TimeStamp(s), FrameCounter, AccX, AccY, AccZ, GyrX, GyrY, GyrZ, EulerX, EulerY, and EulerZ. The TimeStamps(s) column contains the time in seconds from when the trial began. The FrameCounter column contains the frame number relative to the time in the TimeStamp(s) column. The AccX column contains the acceleration due to gravity along the x-axis at the time in the TimeStamp(s) column. The AccY column contains the acceleration due to gravity along the y-axis at the time in the TimeStamp(s) column. The AccZ column contains the acceleration due to gravity along the z-axis at the time in the TimeStamp(s) column. The GyrX contains the angular velocity in degrees per second along the x-axis at the time in the TimeStamp(s) column. The GyrY contains the angular velocity in degrees per second along the y-axis at the time in the TimeStamp(s) column. The GyrZ contains the angular velocity in degrees per second along the z-axis at the time in the TimeStamp(s) column. The EulerX column contains the euler angle in degrees along the x-axis at the time in the TimeStamp(s) column. The EulerY column contains the euler angle in degrees along the y-axis at the time in the TimeStamp(s) column. The EulerZ column contains the euler angle in degrees along the z-axis at the time in the TimeStamp(s) column. Within the label_data folder are 32 label files that are named by subject ID. For example, SA10_label is the name of the label file for subject ID 10. In each label file there are 5 columns. They are Task Code (Task ID), Description, Trial ID, Fall_onset_frame, and Fall_impact_frame. The Task Code (Task ID) column contains the task code along with the task ID. In each label file the task IDs included are 20-34. These tasks are all the tasks that include a fall. The Description column contains the description of the task related to the task code and task ID in the Task Code (Task ID) column. The Trial ID column contains the trial IDs. The Fall_onset_frame column contains the frame number where the subject begins to fall. The Fall_impact_frame contains the frame number where the subject impacts the ground. Now that I have investigated this dataset I have some things that I need to do in order to be able to work with this data. One thing I will need to do is add labels for the tasks where a fall isn't involved. A second thing I will need to do is ensure that my dataset can be used for binary classification instead of multiclass classification. Currently the dataset has multiple labels for

different types of falls. I will need to change the labels so the labels will be either “Fall” or “No Fall”. A third thing I will need to do is a create a Pandas DataFrame for this dataset. After completing these tasks I believe I will be able to work with the data. Overall I think this dataset provides me with a good starting point for my project, but I will need to make some adjustments so the dataset can help me achieve the goal of my project.