Included with this prompt is a dataset of 5000 pitches, 1000 of which are missing pitch classifications. The goal of this project is to fill in those missing pitch classifications with the information provided in the dataset. Each row in the dataset represents an individual pitch (defined by a pitcher_id and pitch_id combination; pitch_id's are not unique across pitchers), each containing the following columns:

- pitch_id – a unique identifier of the pitch within a given pitcher_id
- pitcher_id – a global unique identifier of the pitcher that threw the pitch
- pitcher_side – the hand with which the pitcher threw the pitch
- pitch_type – the provided classification of the pitch
- pitch_initial_speed_a – the velocity of the pitch as measured by System A
- break_x_a – the horizontal break of the pitch as measured by System A
- break_z_a – the vertical break of the pitch as measured by System A
- pitch_initial_speed_b – the velocity of the pitch as measured by System B
- spinrate _b– the spin rate of the pitch as measured by System B
- break_x_b – the horizontal break of the pitch as measured by System B
- break_z_b – the vertical break of the pitch as measured by System B

In general, we trust the measurements of System B more than we do System A. However, in some cases, the only data we have on a pitch comes from System A. For this exercise, please assume all provided pitch type classifications are accurate.

Please return a concise final write-up of your project, pitch type classifications for those 1000 pitches missing them, and any code you wrote along the way. Your project will be graded on:

- Critical thinking about the problem
- Implementation of various analytical methodologies
- Pitch type classification accuracy
- Clarity and efficiency of final write-up