In baseball, a pitcher's success is heavily determined on the outcomes of a batted ball, and a swing and miss. Finding ways to increase success can come in many forms, velocity, spin rate, horizontal, and vertical break all play a key role.

I used data from Baseball Savant (https://baseballsavant.mlb.com/) from the 2019, 2021 and the shortened 2020 season. All pitchers used had a minimum batters faced of 200 for the 2019 and 2021 data and in the 2020 data minimum of 75 batters faced due to a shortened season, pitch types used were 4 seam fastball, slider, curveball, changeup, and cutter. The variables of interest are all occurrences of batted ball outcomes and metrics such as evit velocity, launch angle, barrel percentage, poor/weak contact, in and out of strikezone swing and swing and miss, expected batting average, expected slugging percentage, expected on base plus slugging percentage. Expected means each batted ball is assigned an xBA based on how often comparable balls, in terms of exit velocity, launch angle and, on certain types of batted balls, Sprint Speed -- have become hits since Statcast was implemented Major League wide in 2015 (https://www.mlb.com/glossary/statcast/expected-batting-average). Imputation included 0 for pitch types that are not in that particular pitcher's arsenal.

The model I selected was an unsupervised cluster using K Means. I used Standard Scaler to standardize my data and used pca to narrow down features from 86 to 9. To see the optimal amount of clusters I used the elbow method using inertia and the result I got was 4. Cluster 0: Oldest group, majority likely starters, high ch %, solid command of the strike zone, highest cutter percentage, lowest 4seam velocity, mixes pitches well. Cluster 1: young group, majority likely relievers, heavy fastball and slider percentage and lowest changeup percentage, highest average 4seam velocity, least command. Cluster 2: older group, likely healthy mix of starters and relievers, no 4seam, heavy cutter, more LHP in this group, solid command of the strike zone. Cluster 3: youngest group, majority likely starters, 70% right handed, highest 4seam and ch percentage, mixes pitches well, high average 4seam velo 93.18, really good command of the strike zone. Clusters 0, 2 and probably going to be more crafty, cluster 1 is going to be your power pitchers, and cluster 3 is going to be well balanced.