Training Data Set Creation

The training data set used for this analysis is shown below (first 5 lines):
- Read of the following files game-clicks.csv and buy-clicks.csv
- Feature Selection of the selected columns
  - user_purchases = buyclicks_df[['userId','price']]
  - user_hits = gameclicks_df[['userId','isHit"]]
- Sum the target columns per User Id
  - hits_per_user = user_hits.groupby('userId').sum()
  - revenue_per_user = user_purchases.groupby('userId').sum()
- Merge the datasets into one for the analysis
  - combined_df = hits_per_user.merge(revenue_per_user, on='userId')

Dimensions of the final data set:
cluster_df.shape
(546, 2)

# of clusters created: 3