Econ 323 Final Project.

End Result:

Create a ML algorithm that can classify WR' prospects. (meaningful prediction)
Based off many measurables:

- Combine results
- college production
- School?

↳ Draft position
↳ Similar careers
↳ Grade: A, B, C ...

Issues:

- Hard to determine career 'success' when late round picks are at disadvantage
  → less opportunities, cut easier
- Hard to evaluate 'success' w/ respond to career length.
  solution: only evaluate based off rookie contracts

Data:

- Combine Data
- Pro-Day data?

  Issue: Pro-Day numbers are generally higher.
  → can lower the weight for Pro-Day measurements

- College stats

  Issue: Basic stats (ie. rec yards, TDs, receptions) are heavily usage/scheme dependant.
  Solution: Normalize.
  - Something like: Target share (Rec/QB completions) → measures 'usage' (high v. low volume pass catcher)
  - Something like: TD share (TDs/QB TDs or Total offensive TDs) → usage
  - Something like: Normalized yards (normalized on equal # of pass attempts) → production
  - Smth for YAC?? Hard without data on spoils reference

ML: using scikit-learn K-nearest neighbors

1. Split data into training & test. Stratify by what class?

→ Class: some sort of subjective production score
- Madden ratings
- PFF score
- Create new

} rating

2. Create Pipeline → preprocessor & classifier
↳ rating

3. Parameter grid: finding K value to use

4. GridSearchCV → estimate accuracy for range of K values.
→ pass pipeline & parameter grid as arguments

5. Execute grid search → passing training data into .fit of GridSearchCV

6. Pick K-value that yields high cross-validation accuracy estimate

7. Create new model for selected K-value. Retrain classifier

8. Evaluate estimated accuracy / precision / recall

* Find # parameters variables that minimize RMSE

Example of classifying a player:

Let's use ~~Jordan Jefferson~~ CeeDee Lamb

Why? • 4 NFL seasons (rookie contract)
• full combine data
• full collegiate stats

CD
JJ_combine:

player	Ht	Wt	Arm Length	Hand Size	40yd	10yd	Vert	Broad	B.P
CeeDee Lamb	6'2"	198	32 1/4"	9 1/4"	4.5	1.46	34 1/2"	124"	11

Missing Data: 3-cone, 20-yd shuttle

CD_stats:

class	Target share	TD share	* yards / 30 attempts
Fr	46/285	7/43	~~304~~ 838.88
SO	65/~~344~~ 260	11/42	1,240.5
JR	62/237	14/32	1,634.6

- - - - - - - - - - ->

yds / 30 QB attempts:

- This stat 'normalizes' rec yards by making passing attempts equal. For example, CeeDee's JR season: he had 1,377 rec yards, but his QB Jalen Hurts threw 340 passes in 14 games for 24.28 attempts per game.

→ 30 / 24.28 = 1.235, the amount we multiply CeeDee's yards by:

1,327 · 1.235 = 1,639 adjusted yards

Can merge CD_combine & CD_stats to get CD_profile. ~~[illegible]~~

→ Now, evaluate CeeDee's 4-year NFL career:

- very subjective, but things like stats, PFF ratings, and madden ratings help. Also pro-football-reference AV (approx. value)

CeeDee Madden ratings (end of year):

rookie:	85	• average is not telling enough:
2nd:	90	ratings carry over from each year.
3rd:	93	Some sort of weighted avg? TBD
4th:	96	

rolling avg. 93.125

CeeDee PFR AV (approx. value)

rookie:	9	• not carry over
2nd:	10	
3rd:	14	
4th:	20	

• Find some way to standardize rating & AV. Of course, this part is just rough ideas so far so will glance over for now and give CeeDee a career rating of:

≈ 95/100 • 'Major success'

↳ classifier: 90+ 'major success'
80-90 'success'
so on...
so on...

Thus, can test the model accuracy by seeing if:

• CD_profile will have a classifier of 'success'