

# Cloudy With a Chance of Football (Cohort 23): Capstone Project Data Architecture

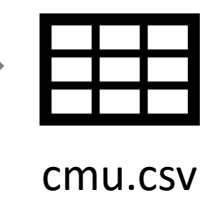
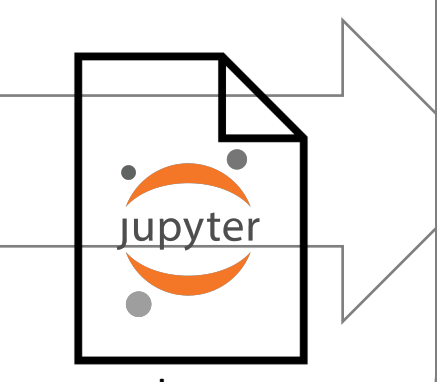
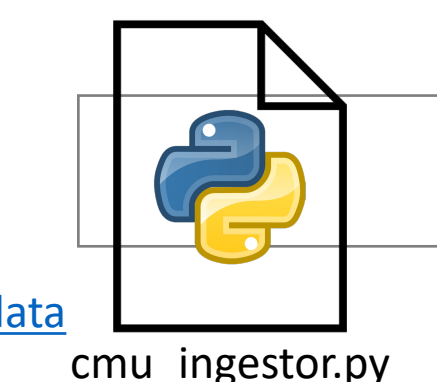
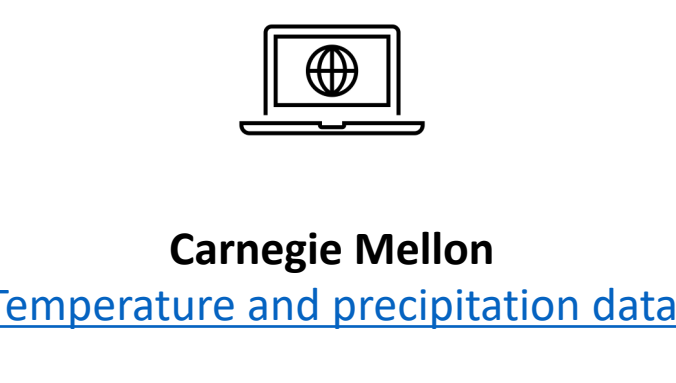
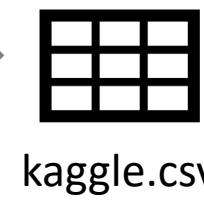
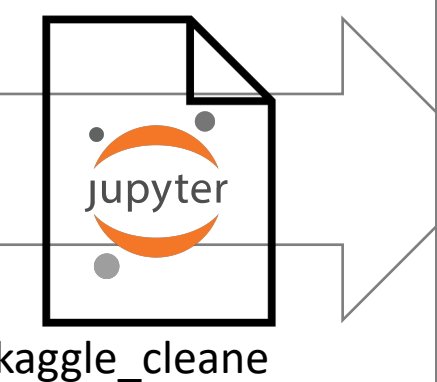
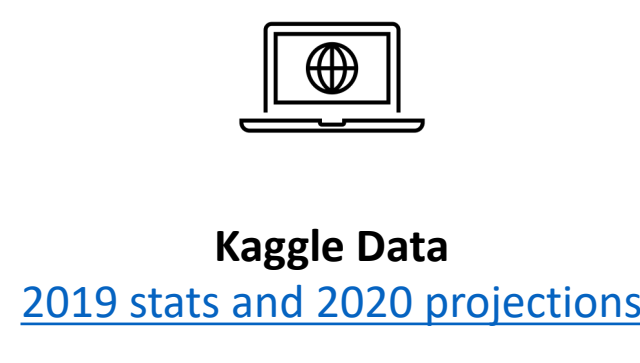
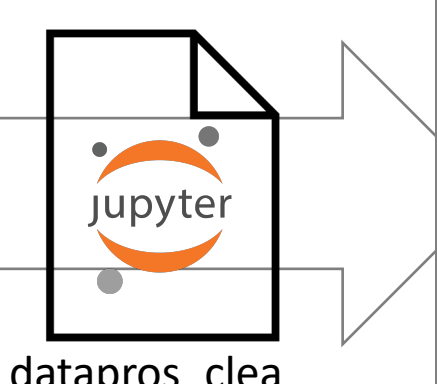
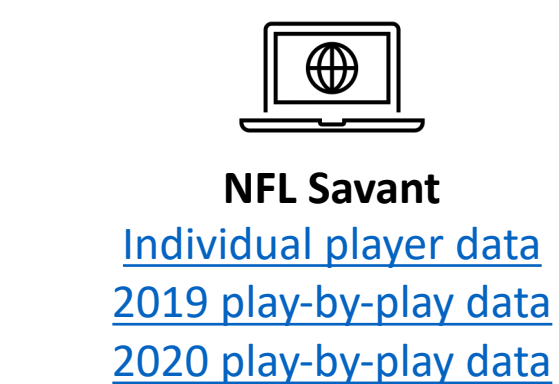
## Data Ingestion

### Data Limits

Years: 2019 and 2020

Positions: All offensive positions

Platform: ESPN



players		
player_id	unique_id	player_name
0	tombrady	Tom Brady
1	joshallen	Josh Allen

week_table			
year	week	week_start	week_id
2019	1	20190907	2019_1
2019	2	20190914	2019_2

positions	
player_id	position
0	qb
1	qb

player_age		
player_id	week_id	age
0	2019_1	42.1
1	2019_1	25.52

player_dob	
player_id	dob
0	19770803
1	19960521

player_team		
player_id	week_id	team_name
0	2019_1	NE
1	2019_1	BUF

injury_status		
player_id	week_id	injury_status
0	2019_1	0
1	2019_1	0

days_since_last_game		
team_name	week_id	days_since_last_game
NE	2019_2	6
BUF	2019_2	5

time_displacement_table		
week_id	team_name	hours_displaced
2019_1	NE	0
2019_2	BUF	0

weather				
week_id	team_name	temperature_min	temperature_max	precipitation
2019_1	NE	58	75	.1
2019_1	BUF	35	48	3

opponents table																						
week_id	team_name	game_date	ARI	ATL	BAL	BUF	CAR	CHI	CIN	CLE	...	NE	NO	NYG	NYJ	PHI	PIT	SF	SEA	TB	TEN	WAS
2019_1	NE	20190907	0	0	0	0	0	0	0	0	...	0	0	0	0	0	1	0	0	0	0	0
2019_2	BUF	20190906	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	0	0	0	0	0

stats_table														
player_id	week_id	PassingYds	PassingTD	Int	PassingAtt	Cmp	RushingAtt	RushingYds	RushingTD	Rec	Tgt	ReceivingYds	ReceivingTD	FL
0	2019_1	324	5	0	20	17	3	6	0	0	0	0	0	0
1	2019_1	405	4	0	32	25	4	12	0	0	0	0	0	0

team_home_time_zone		
team_name	week_id	time_zone
NE	2019_1	-5
BUF	2019_1	-5

game_location		
week_id	team_name	location
2019_1	NE	1
2019_1	BUF	0

performance				
player_id	week	predicted	actual	performance
0	2019_1	18.20	21.90	1
1	2019_1	29.35	27.10	0
0	2019_2	25.87	18.50	1
1	2019_2	19.93	28.98	1

Columnst		
Column Name	Data Type	Unique Variables or Description
player_id	integer	
unique_id	string	firstnamelastname
player_name	string	Firstname Lastname
year	integer	2019 2020
week	integer	
week_start	datetime	YYYY-MM-DD
week_id	string	YYYY_WW
position	string	qb, rb, te, wr, k
age	float	
dob	datetime	YYYY-MM-DD
time_zone	integer	Hours away from Greenwich mean time of home time zone
team_name	string	<a href="#">Official team abbreviation codes</a>
injury_status	integer	0 : no injury 1 : questionable 2 : out
days_since_last_game	integer	
hours_displaced	integer	Number of hours away from home time zone
temperature_min	integer	
temperature_max	integer	
precipitation	float	
PassingYds	float	
PassingTD	integer	
Int	integer	
PassingAtt	integer	
Cmp	integer	
RushingAtt	integer	
RushingYds	float	
RushingTD	integer	
Rec	integer	
Tgt	integer	
ReceivingYds	float	
ReceivingTD	integer	
FL	integer	
location	integer	0 : away 1 : home
predicted	float	ESPN prediction for player fantasy points
actual	float	Actual number of ESPN fantasy points achieved
performance	integer	0 : predicted < actual 1 : actual < predicted

## Machine Learning

Model 6: Kickers

Model 5: Wide Receivers

Model 4: Tight Ends

Model 3: Running Backs

Model 2: Quarterbacks

Model 1: All players



## Reporting and Visualization

Develop a simple web application using either Django or Flask



## Computation and Analysis

- Determine best features for model selection using, among other things, a correlation matrix
- Normalize all data from 0 to 1
- Convert string data to integer data
- Check on class imbalance
- Divide into a train (80%) test (20%) split of data

