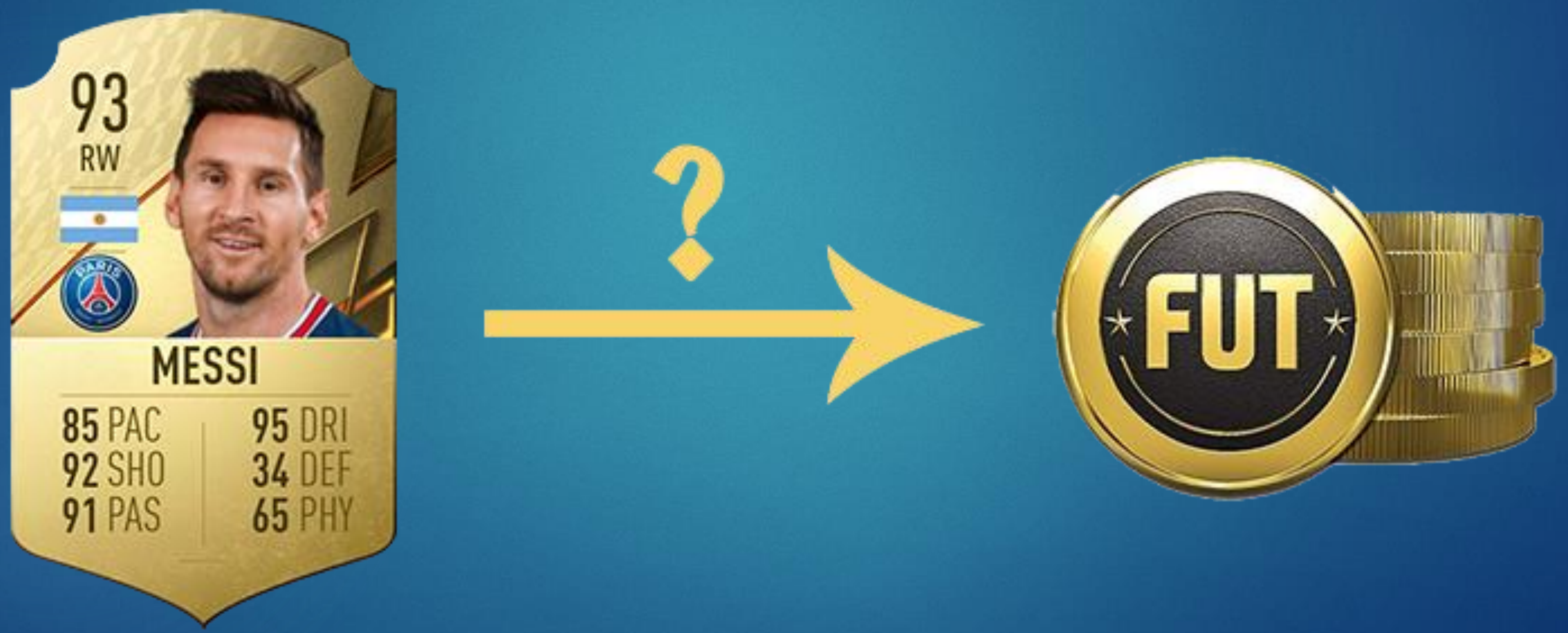
FIFA22

Price Prediction

BY SHAI GUETA

Research Question

- Can I predict prices of Fifa cards based on their stats?



But first, what's a fifa card?

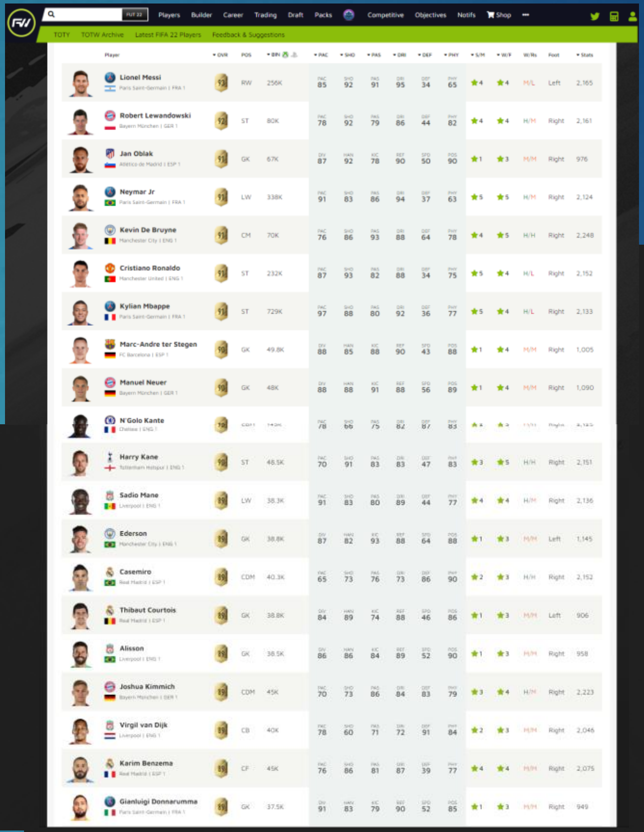
- A fifa card is an item you get in fifas game mode “Ultimate Team”
- A card represents a player (based on real football players) you can use in your team to build your “Ultimate Team” and shows the stats of that player suggesting how will the player play in game (will he be fast, strong etc.)
- To obtain this cards you can either “Pack” them from packs available on the game, or buy them of the FUT(fifa ultimate team) market using virtual currency called “fifa coins”
  (note: some cards are available to get as a reward via challenges within the game)
- Each player has a “Base Card” that is available in packs for the whole game circle and can also get “Special’ cards throughout the circle, special cards are available for a limited time and have increased stats

Base Card

Special Card

Finding And Scraping Data

- FUTWIZ.com
  the website I choose to get the
  stats and information of the fifa cards
  from
- beautifulsoup
  the python library I used to scrap the
  needed data from website
- I've scraped 447 pages of the website
  and made a data frame
  by using pythons "Pandas"



Data Frame

- The data frame

	Name	Rating	Position	Price	Version	Nation	Team	League	Pace	Shot	Pass	Drib	Deff	Phy	Skill_Mov	Weak_Foo	Work_Rat	Strong_Fo	In_Games
0	Pele	98	CAM	5.5M	icon	54	Icons	ICO	95	96	93	96	60	76	5	4	H/M	Right	2,415
1	DiegoMar	97	CAM	3.8M	icon	52	Icons	ICO	92	93	92	97	40	76	5	3	H/M	Left	2,297
2	ZinedineZ	96	CAM	3.5M	icon	18	Icons	ICO	85	92	96	95	75	86	5	5	M/M	Right	2,451
3	RonaldoN	96	ST	9.1M	icon	54	Icons	ICO	97	95	81	95	45	76	5	5	M/M	Right	2,252
4	Pele	95	CF	3.8M	icon	54	Icons	ICO	96	93	90	95	56	75	5	4	H/M	Right	2,357
5	DiegoMar	95	CAM	1.7M	icon	52	Icons	ICO	88	91	90	95	42	75	5	3	H/M	Left	2,262

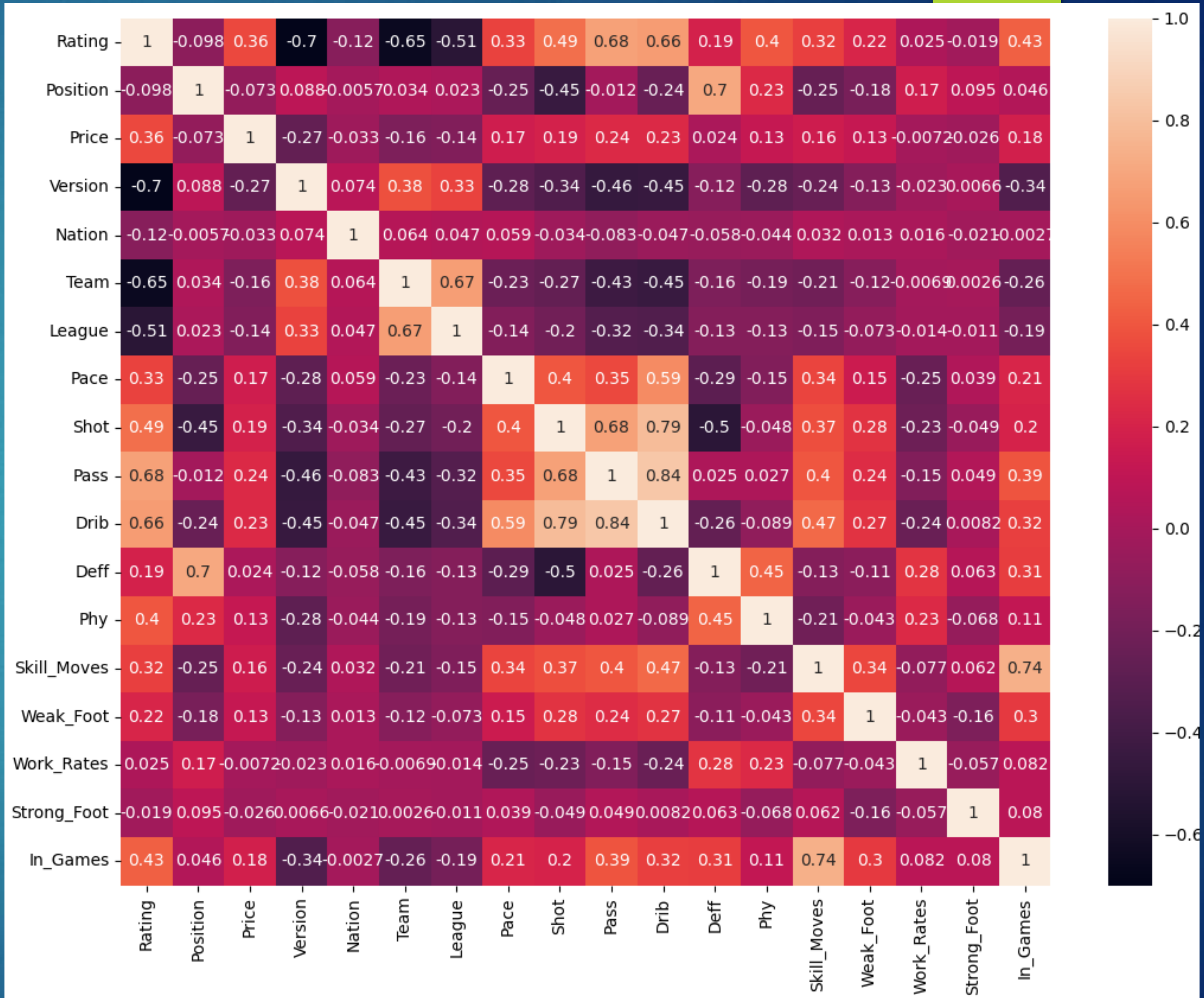
The data frame was created with 11,175 rows, a row per card and19 columns
With the following meanings:
Players **Name**, **Rating**, **Position**, **Price**, **Nation**, **Team**, **League**, **Strong Foot**
Are self explanatory
Pace = Pace stat
Shot = Shooting stat
Pass = Passing stat
Drib = Dribbling stat
Deff = Defending stat
Phy = Physical stat
Skill Move = what level of special tricks you can pull of with the card
Weak Foot = how well the card will preform while using his weaker foot
Work Rate = how often the card will push to the attack/defense
In games = a sum of the players in game stats

Data Clearing

- After scraping the data and creating a data frame
I found 3 missing values, the values were indeed missing in the website and I've manually inserted them into my data frame after checking the values on a similar site called "Futbin.com"
- Some card had a NULL value in their price column.
After going over them I found out that the majority of them were Reward cards that you unlock by completing in game tasks and therefore you cannot buy them with fifa coins.
The rest of the NULL prices were cards that weren't on the fifa market at the time of scraping.
Since they all were low rated and undesirable I've decided to drop them from my data frame.
- I've also dropped the name column and converted all of the values into numeric ones.

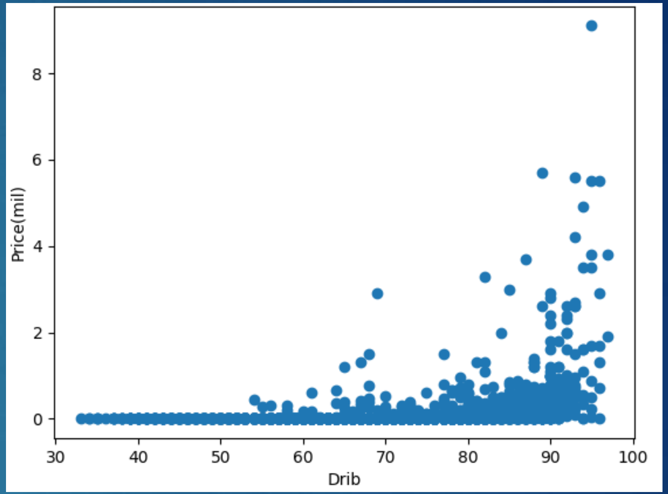
Visualization

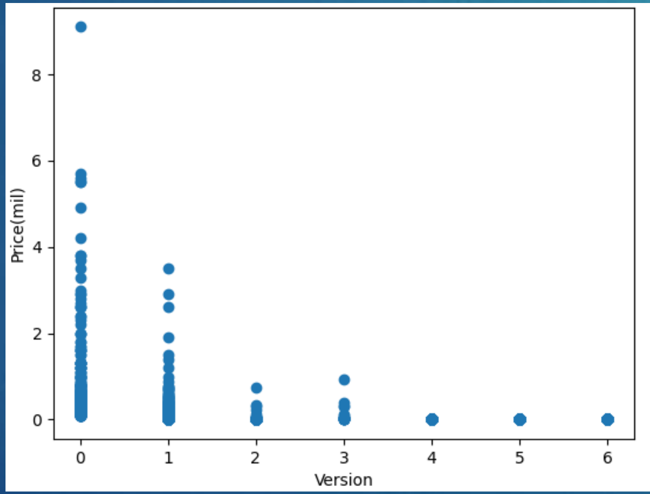
- First I made a correlation matrix to see how each column in my data frame affect the other using the “seaborn” library in python
- As we can see there is no one column that affects the price heavily.
  But we can also see that things like ‘Work Rates’, ’Strong Foot’ and ‘Position’ hardly affect the price, we will keep it in mind.

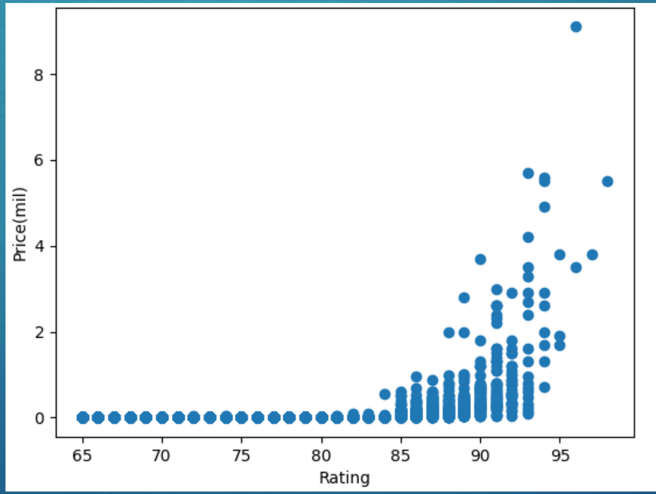


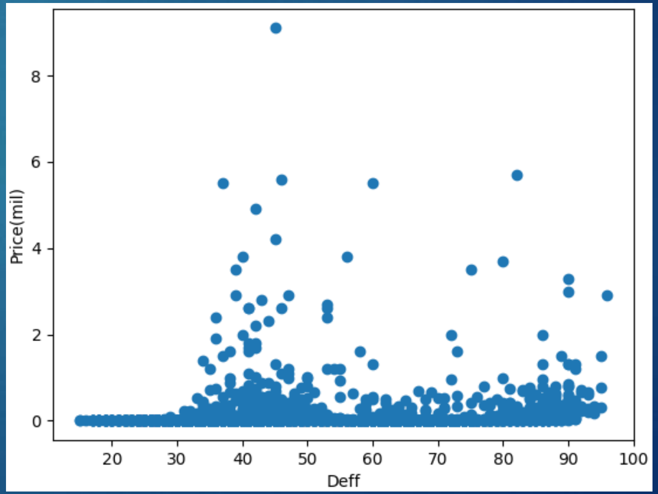
Visualization

- After looking at some graphs I noticed there are a lot of players at their minimum value (also known as “discards” on fifa).
  Also noticed that most of the non “discards” are getting to a max price of ≈4 million, and there are only a few anomalies that crossed that price point
- All that will be considered at the prediction phase



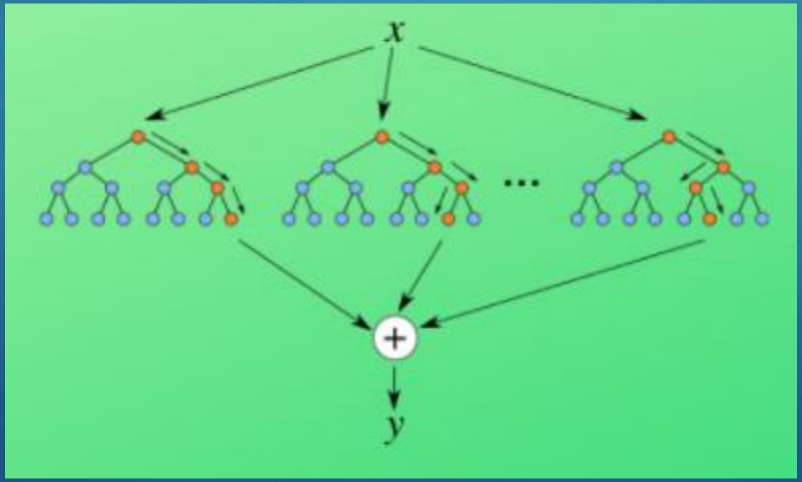






Picking An Algorithm

- To make my prediction I chose to use the "Random Forest Regressor" from the Sklearn library in python.
  Why? After trial and error with other options such as "Linear Regression" and "SGDClassifier".
  I got the best results from Random Forest and decided to continue with it.
- Also used "GridSearchCV" to help me find the optimal values for the "n_estimators" and "max_depth" variables in Random Forest



Prediction

- To estimate the model I've used the r2_score from the "sklearn.metrics" library
- With the data frame as it is I got a low score of ≈0.32
- With the information I got from the visualization phase I've narrowed my data frame to only cards that are above discard value and below the anomaly value.
  Putting my price range as such: 900 < price < 4 million
- With said price range adjustment the new r2_score got up to ≈0.80 a massive improvement of my previous score! And the highest score I managed to achieve.
- ***Side note****: to confirm what we saw on the correlation matrix*
  I tried to drop the: Work Rate, Strong Foot, Position, Nation columns.
  All but "Position" (who had a ≈0.01drop to the score) hardly had any affect on the final score.

Prediction Sum Up

Other algorithms	R2_Score
Linear Regression	≈0.14
SGDClassifier	≈0.13

Random Forest	R2_Score
“As is” data frame	≈0.32
Data frame with custom price range	≈0.80