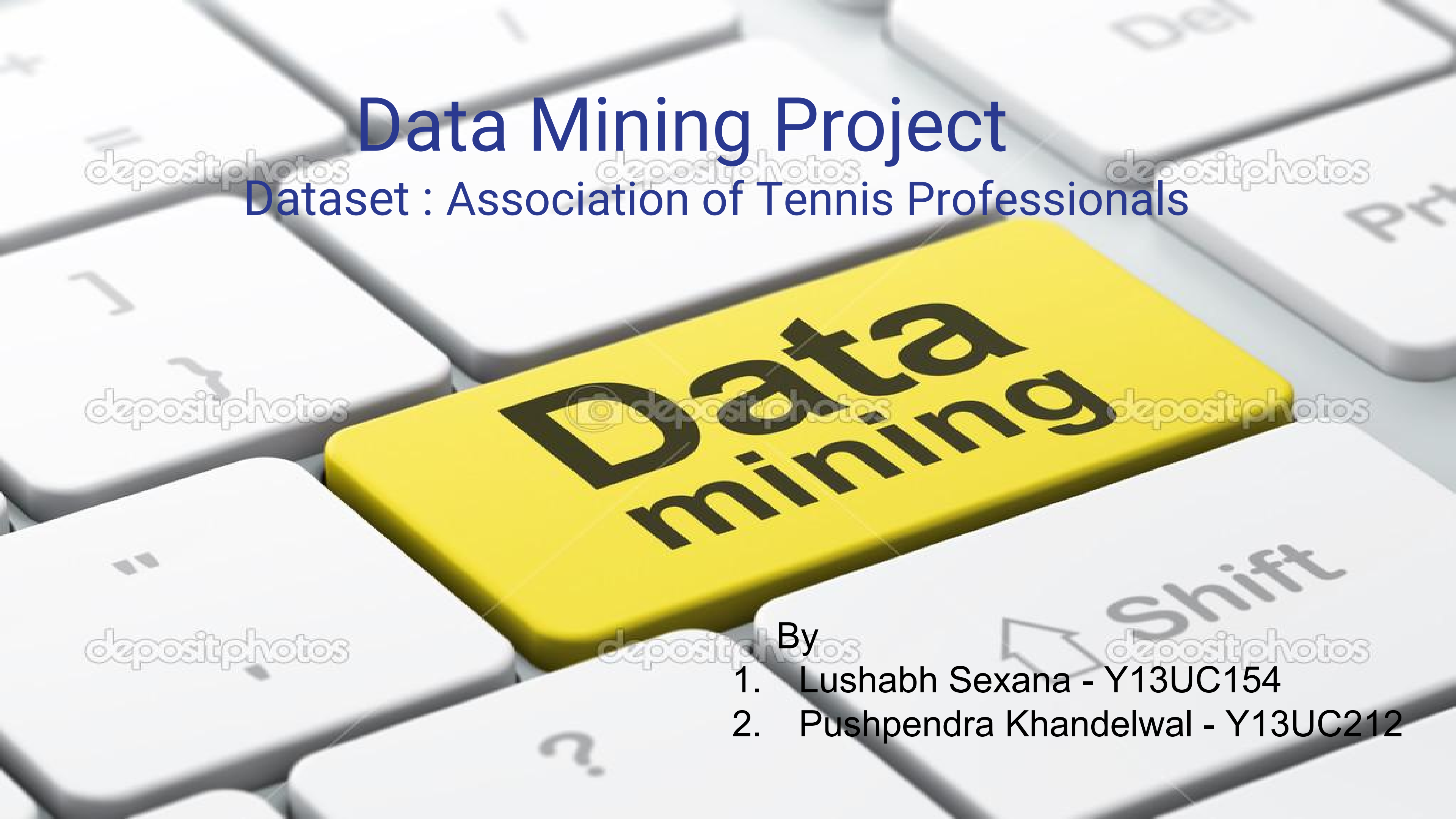
Data Mining Project

Dataset : Association of Tennis Professionals



depositphotos

By

1. Lushabh Sexana - Y13UC154
2. Pushpendra Khandelwal - Y13UC212

About Data

- Results for the men's ATP tour
- Series: Masters Series, Grand Slam, International Series competitions etc
- Surface on which match is being played
- Ranking of winner and loser of each set
- Grand slam consists of four important annual tennis events

Series	Surface	Best of	Winner	Loser	WRank	LRank	W1	L1	W2	L2	W3	L3	W4	L4	W5	L5	Wsets	Lset
International	Hard	3	Dosedel S.	Ljubicic I.	63	77	6	4	6	2							2	
International	Hard	3	Enqvist T.	Clement A.	5	56	6	3	6	3							2	
International	Hard	3	Escude N.	Baccanello P.	40	655	6	7	7	5	6	3					2	
Grand Slam	Clay	5	Ferrero J.C.	Kratochvil M.	3	85	7	5	7	5	6	4					3	
Grand Slam	Clay	5	Gaudio G.	Mathieu P.H.	22	44	7	5	6	3	6	3					3	
ATP250	Clay	5	Clement A.	Enqvist T.	36	100	6	3	6	4	6	2					3	
International	Hard	3	Grosjean S.	Ilie A.	26	51	6	2	6	1							2	
Grand Slam	Grass	5	Schalken S.	Morrison J.	23	98	6	4	7	6	6	0					3	
Grand Slam	Grass	5	Vacek J.	Voinea A.	80	63	6	1	4	6	6	3	6	4			3	
Grand Slam	Grass	5	Youzhny M.	Escude N.	61	33	6	2	1	6	6	3	6	3			3	
International	Carpet	3	Mirnyi M.	Carlsen K.	45	63	7	5	2	6	4	0					1	
International	Carpet	3	Nieminen J.	Kratochvil M.	37	80	6	4	3	6	6	1					2	
International	Hard	3	Petrovic D.	Huet S.	208	122	3	6	6	3	6	3					2	

Problem

1. Probability of Winning based on difference of rank
2. Surface wise performance of players
3. Clustering on player performance

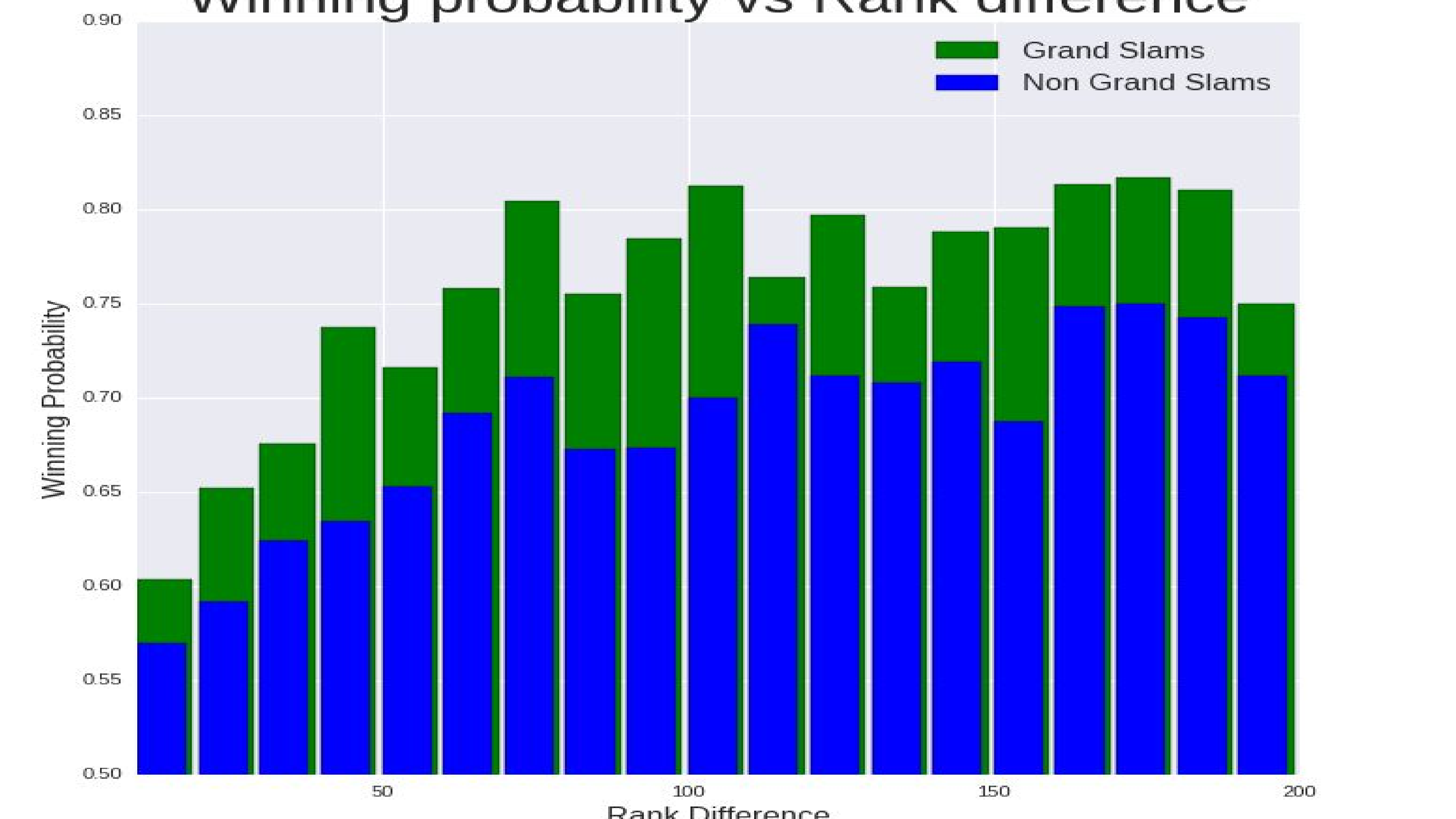
1. Winning Probability based on Difference of Rank

- Requires Pre-processing the data
    - Calculate difference of rank of Winner and Loser
    - Normalise the rank difference
    - Separate Grand Slam and Non-grand slam series data
    - Given a rank difference calculate winning probability for both Grand Slam and regular series

Probability = (# +ve Winning) / (#+ve winning + #-ve winning)

where
+ve winning - rank difference is positive
-ve winning - rank difference is negative

After this plot the graph b/w probability and rank difference

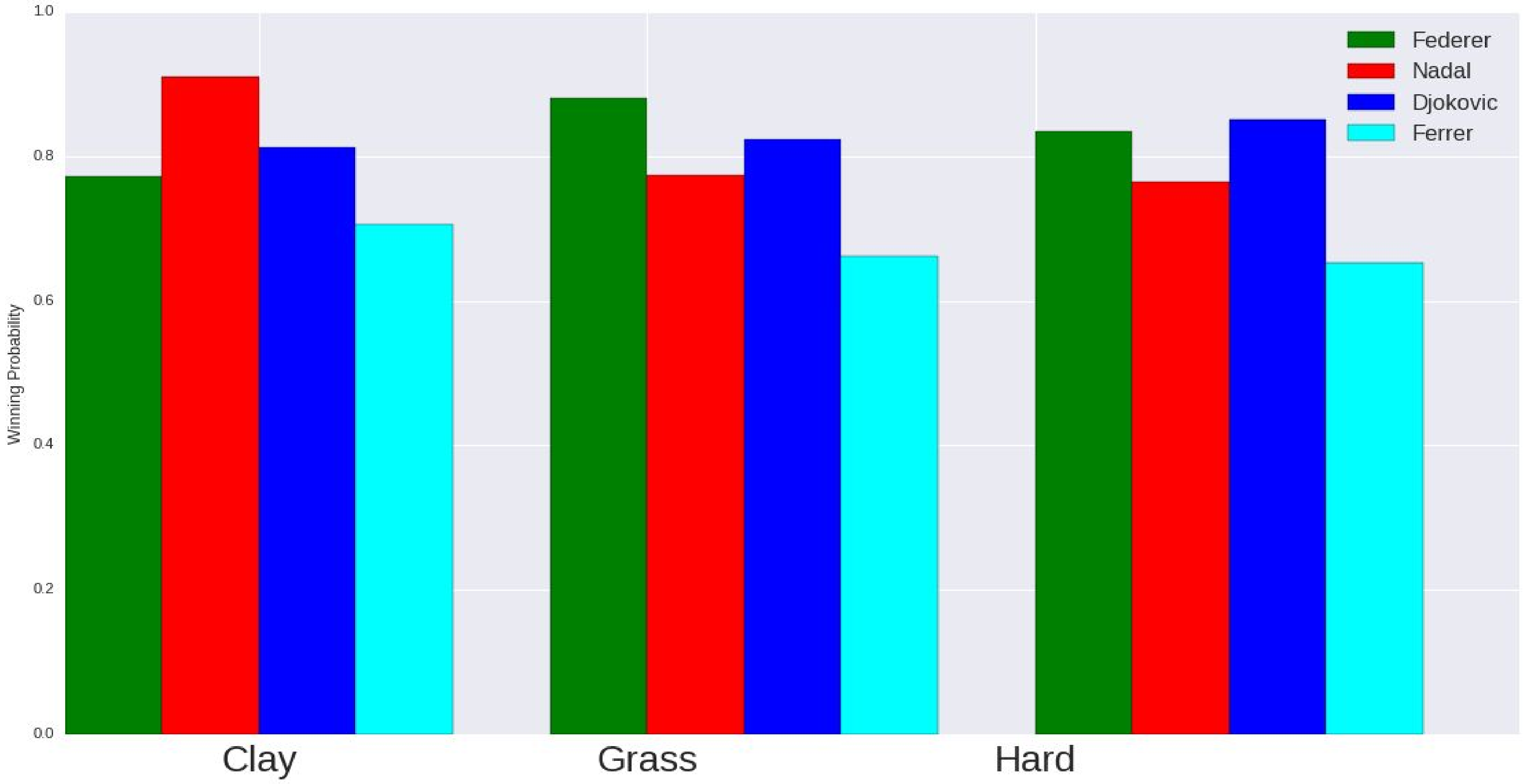

Winning probability vs Rank difference
Grand Slams
Non Grand Slams
Winning Probability
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
50
100
150
200
Rank Difference

Conclusion

- chances of winning increase as the rank difference does.
- However - this effect tends to saturate when the rank difference reaches 100 places. This saturation is for both Grand Slam and "Regular" (that is - non Grand Slam) tournaments.
- A favorite player has a ~80% chance of winning when facing person ranked 100 places below in a Grand Slam tournament, but only around 70% winning chances in a regular one.

2. Surface wise performance of players

- For each player, calculate the winning probability in Clay, Hard and Grass surface
- For analysis, extract data of players who have played more than 850 games
- Plot the bar chart of each player’s probability in all 3 surfaces


1.0
0.8
0.6
0.4
0.2
0.0
Winning Probability
Clay
Grass
Hard
🟩 Federer
🟥 Nadal
🟦 Djokovic
🟦 Ferrer