

SUPSI

Analisi statistica sull'evoluzione del gioco dell'Atalanta (2011 - 2019)

Studente/i	Relatore	Correlatore
Massimo De Santi	Alessandro Giusti	
Corso di laurea	Modulo / Codice Progetto	Anno
Ingegneria Informatica	Data Science	2020
Committente	Data	
SUPSI DTI	08.05.2020	



Presentazione squadra

- Atalanta Bergamo Calcio
- Anno fondazione 1907
- Città: Bergamo (~120'000 abitanti)
- Esperienze Internazionali:
 - Coppa delle Coppe: 1964, 1987
 - Europa League: 1990, 2017, 2018
 - Champions League: 2019
- Trofei: Coppa Italia 1962
- Stadio di proprietà dal 2017
- Negli ultimi 10 anni l'Atalanta è passata dalla **Serie B** ai quarti di **Champions League**
- Dal 2016 Mr. Gasperini ha sviluppato un gioco ultra offensivo

1	Season	Comp	LgRank	MP	W	D	L	GF	GA	GDiff	Pts	Spettatori	Notes
2	2019-2020	1. Serie A	4th	25	14	6	5	70	34	36	48	17,681	Stagione in corso, quarti finale Champions
3	2018-2019	1. Serie A	3rd	38	20	9	9	77	46	31	69	18,301	Europa League: eliminata ad agosto (playoff)
4	2017-2018	1. Serie A	7th	38	16	12	10	57	39	18	60	17,921	Europa League: eliminata a febbraio (sedicesimi)
5	2016-2017	1. Serie A	4th	38	21	9	8	62	41	21	72	16,948	Allenatore Gasperini
6	2015-2016	1. Serie A	13th	38	11	12	15	41	47	-6	45	15,945	
7	2014-2015	1. Serie A	17th	38	7	16	15	38	57	-19	37	15,160	Allenatore Reja
8	2013-2014	1. Serie A	11th	38	15	5	18	43	51	-8	50	14,194	
9	2012-2013	1. Serie A	15th	38	11	9	18	39	56	-17	40	15,396	
10	2011-2012	1. Serie A	12th	38	13	13	12	41	43	-2	46	15,648	
11	2010-2011	2. Serie B	1st	42	22	13	7	61	35	26	79	18,737	Promozione in serie A (Allenatore Colantuono)



Obiettivo analisi

- Obiettivo: Evidenziare l'evoluzione del gioco dell'Atalanta
- La Serie A è composta da 20 squadre che si incontrano in match di andata e ritorno:
 - $19 * 2 * 10 = 380$ incontri a stagione
- Incrociando le statistiche di gioco di tutti gli incontri in Serie A negli ultimi 8 anni:
 - Dalla stagione 11/12 (ritorno in serie A dell'Atalanta)
 - alla stagione 18/19 (ultima stagione completa; la stagione 19/20 non è terminata)
 - $8 \text{ stagioni} * 380 \text{ incontri} = \underline{\underline{3040 \text{ osservazioni}}}$
- Codice sorgente:
 - <https://github.com/code-mds/datascience/blob/master/progetto/AnalisiAtalanta.ipynb>
- Dataset provider:
 - <https://www.football-data.co.uk>
- Annotazioni complementari ricavate da:
 - <https://fbref.com/en/squads/922493f3/history/Atalanta-Stats>
 - https://it.wikipedia.org/wiki/Serie_A

Dataset

- Dati dettagliati delle partite di Serie A negli ultimi 8 anni:
 - Dalla stagione 11/12 (<https://www.football-data.co.uk/mmz4281/1112/l1.csv>)
 - ...
 - alla stagione 18/19 (<https://www.football-data.co.uk/mmz4281/1819/l1.csv>)
 - Descrizione dataset (<https://www.football-data.co.uk/notes.txt>)
- Utilizzo dei seguenti 12 attributi:
 - Goal fatti (**FTHG**) subiti (**FTAG**)
 - Tiri fatti (**HS**) subiti (**AS**)
 - Tiri (nello specchio) fatti (**HST**) subiti (**AST**)
 - Falli concessi (**HF**) subiti (**AF**)
 - Ammonizioni a favore (**HY**) subiti (**AY**)
 - Espulsioni a favore (**HR**) subiti (**AR**)

Data Wrangling

• Data Wrangling:

- 8 file separati da unire in un unico dataset
- Normalizzare informazioni (partite giocate in trasferta e partite giocate in casa)
- Calcolo dei punti
- Calcolo differenza reti

Full dataset

fulldf

	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HS	AS	HST	AST	...	AF	HC	AC	HY	AY	HR	AR
Season																		
11/12	2011-09-09	Milan	Lazio	2.0	2.0	D	18.0	12.0	8.0	5.0	...	15.0	3.0	1.0	2.0	2.0	0.0	0.0
11/12	2011-10-09	Cesena	Napoli	1.0	3.0	A	11.0	18.0	3.0	6.0	...	12.0	4.0	6.0	2.0	3.0	1.0	0.0
11/12	2011-11-09	Catania	Siena	0.0	0.0	D	9.0	4.0	1.0	2.0	...	25.0	4.0	4.0	2.0	4.0	0.0	0.0
11/12	2011-11-09	Chievo	Novara	2.0	2.0	D	11.0	14.0	4.0	4.0	...	17.0	0.0	6.0	1.0	1.0	1.0	0.0
11/12	2011-11-09	Fiorentina	Bologna	2.0	0.0	H	15.0	14.0	7.0	2.0	...	12.0	12.0	1.0	1.0	1.0	0.0	0.0
...
18/19	2019-05-26	Inter	Empoli	2.0	1.0	H	20.0	9.0	15.0	5.0	...	9.0	8.0	2.0	4.0	2.0	1.0	1.0

teams_dic[selected_team].my_normalized_df

	Goal	Shots	ShotsTarget	Corner	Fouls	YellowCard	RedCard	Point	GDiff
Season									
11/12	1.0	8.0	1.0	15.0	1.0	4.0	0.0	3.0	0
11/12	2.0	12.0	5.0	20.0	4.0	4.0	0.0	3.0	0
11/12	0.0	18.0	3.0	18.0	3.0	3.0	0.0	1.0	0
11/12	1.0	5.0	4.0	17.0	4.0	2.0	0.0	1.0	0
11/12	1.0	11.0	3.0	18.0	1.0	4.0	0.0	3.0	0
...
18/19	3.0	18.0	11.0	10.0	6.0	1.0	0.0	3.0	0
18/19	0.0	5.0	2.0	13.0	4.0	2.0	0.0	1.0	0

teams_dic[selected_team].my_tot_stats

	Goal	Shots	ShotsTarget	Corner	Fouls	YellowCard	RedCard	Point	GDiff
Season									
11/12	41.0	413.0	142.0	592.0	189.0	96.0	4.0	52.0	-2.0
12/13	39.0	409.0	129.0	585.0	180.0	105.0	13.0	42.0	-17.0
13/14	43.0	447.0	154.0	525.0	190.0	81.0	6.0	50.0	-8.0
14/15	38.0	463.0	132.0	629.0	193.0	103.0	9.0	37.0	-19.0
15/16	41.0	454.0	137.0	636.0	190.0	95.0	14.0	45.0	-6.0
16/17	62.0	534.0	201.0	603.0	229.0	85.0	3.0	72.0	21.0
17/18	57.0	573.0	197.0	491.0	229.0	67.0	4.0	60.0	18.0
18/19	77.0	475.0	254.0	429.0	242.0	61.0	3.0	69.0	31.0

Evoluzione bilancio gol

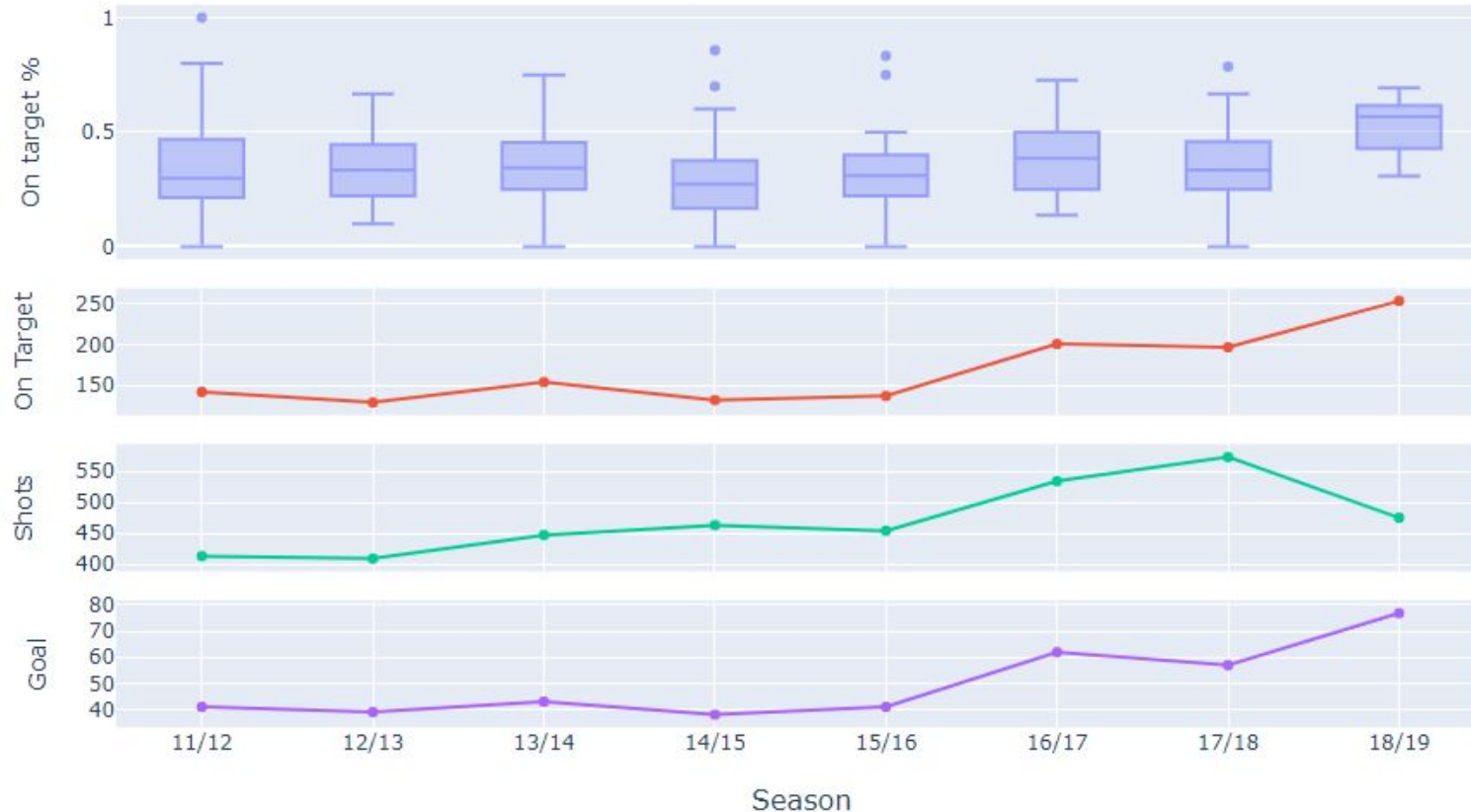
Positive Goal balance in last 3 seasons (Atalanta)



- Dalla stagione 16/17 la differenza reti stagionale passa da negativa a positiva.
- Ne consegue un marcato miglioramento sia in termini di punti che di posizionamento in classifica.

Evoluzione precisione al tiro

Improved shooting precision in last 3 seasons (Atalanta)



- Dalla stagione 16/17 la precisione al tiro è aumentata portando un incremento del numero di gol.
- Interessante il dato dell'ultima stagione dove il decremento dei tiri totali non abbia impattato il numero dei gol.

Evoluzione falli commessi

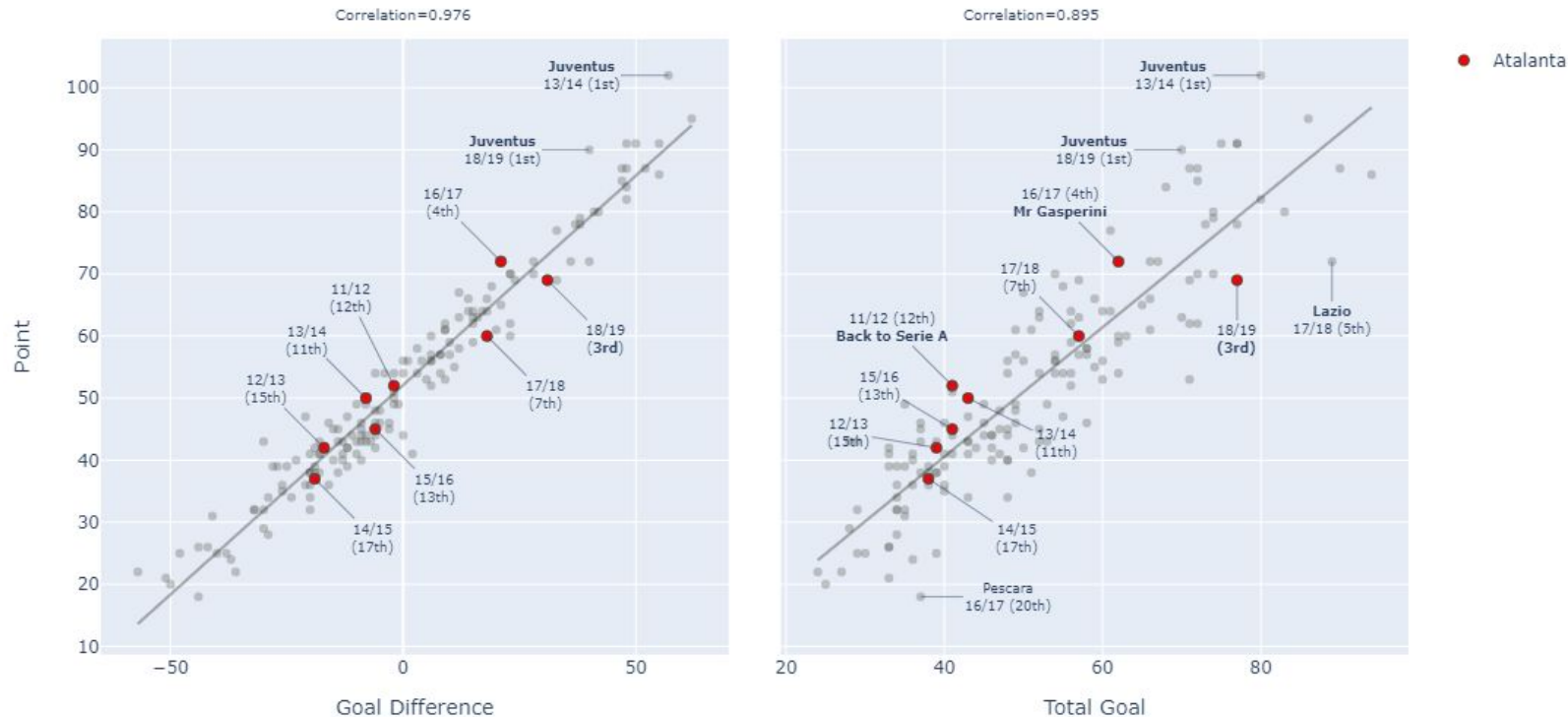
Increased number of committed fouls (Atalanta)



- Dalla stagione 16/17 è incrementato il numero di falli commessi.
- Sorprendentemente è diminuito il numero di cartellini gialli e rossi.
- Purtroppo i dati disponibili non chiariscono le possibili cause.

Correlazione tra Gol e Punti

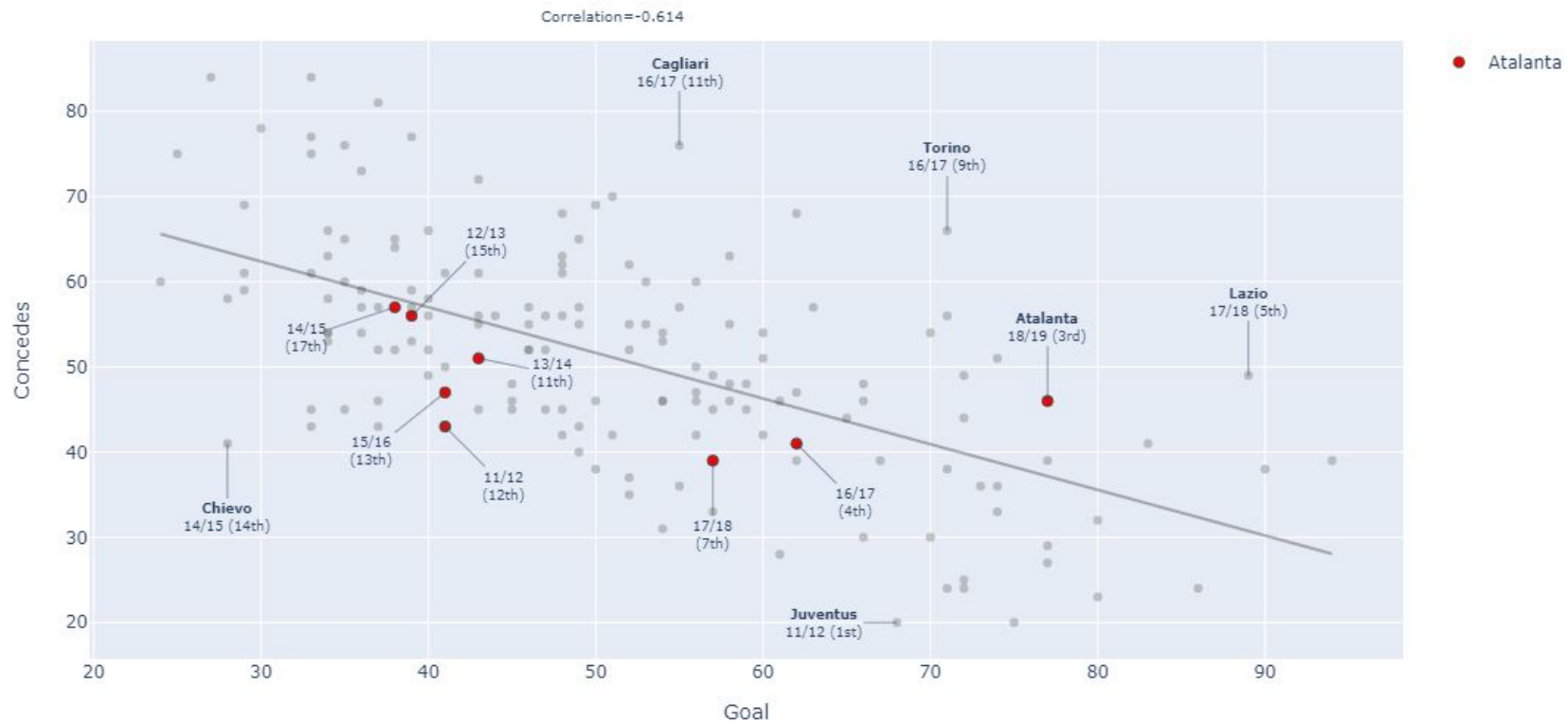
Strong correlations goal/point



- In questi grafici sono rappresentate tutte le partite di serie A dalla stagione 11/12 alla stagione 18/19.
- La correlazione **differenza reti / punti** è altissima (0.976) con poche squadre che si discostano dalla linea di regressione.
- Anche la correlazione **gol / punti** è alta (0.895) ma con più eccezioni.
- Interessante il caso della **Lazio 17/18** che ha segnato tantissimo (89 gol) ma è arrivata solo quinta. La **Juventus 13/14** segnando meno (80 gol) ha vinto il campionato.
- Considerando differenza reti e gol segnati, l'**Atalanta 18/19** avrebbe forse potuto ottenere più punti.

Correlazione tra Gol fatti e Gol subiti

Weaker correlation Concedes/Goal



- In questo grafico sono rappresentate tutte le partite di serie A dalla stagione 11/12 alla stagione 18/19.
- La correlazione **gol fatti / gol subiti** è meno marcata (-0.614) rispetto alle correlazioni viste precedentemente.
- Diverse squadre si discostano dalla linea di regressione (**Cagliari 16/17**, **Torino 16/17**, **Lazio 17/18**, ...)
- Per quanto riguarda l'**Atalanta**, si nota un miglioramento nelle ultime stagioni che segue la linea di regressione.
- Il numero di gol fatti è cresciuto pur mantenendo stabile il numero di gol subiti.