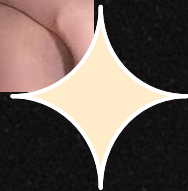




1992-2020 K-POP 新興韓團 火熱程度關聯分析 PROJECT PROPOSAL

歷史二 王學謙 / 心理五 李采蓉 / 地理碩一 賴郁升

2023/12/21 @NTULING





目錄



01.

研究動機

02.

資料選用

03.

研究流程

04.

描述統計結果

05.

關聯分析結果

06.

總結與討論

研究動機




- 不斷推陳出新的韓國流行音樂與團體組合
- K-pop是否存在著一套在競爭中成功的法則？
- 使用Kaggle資料集和Youtube數據進行分析



資料選用：主要資料集

- 採用Kaggle資料集：K-Pop Database (1992-2020)
- 上開資料整理自 K-Pop Database

 kpop_boy_groups.csv

Name	Short	Debut	Company	Members	Orig. Memb.	Fanclub Name	Active	Average_age
------	-------	-------	---------	---------	-------------	--------------	--------	-------------

這份資料共紀錄147個男性組成韓國流行樂團體

 kpop_girl_groups.csv

Name	Short	Debut	Company	Members	Orig. Memb.	Fanclub Name	Active	Average_age
------	-------	-------	---------	---------	-------------	--------------	--------	-------------

這份資料共紀錄152個女性組成韓國流行樂團體

 youtube.csv

Name	Likes	Views	Comments
------	-------	-------	----------

這份資料共紀錄3443筆韓國流行樂影片



資料選用：次要資料集

- 出道時的平均年齡
 1. 根據boy與girl_groups.csv的團體名稱欄位
- 2. 搜尋與計算該年出道的團員平均年齡
- 3. 將年齡資料存於.txt檔中，格式為[Group, Age]



研究流程

- 先整理團體成員之性別、出道時的平均年齡、出道時的團體人數、出道年份等五個資料欄位與整理。
- 進而分析上述因素與 *團體火熱程度 的關聯性。

*該資料以Youtube爬蟲方式取得，算法 = $\frac{\text{MV點閱率}}{(2020 - \text{出道年份})}$



爬蟲前：原始DataFrame

kpop_music_videos.csv (322.64 kB)

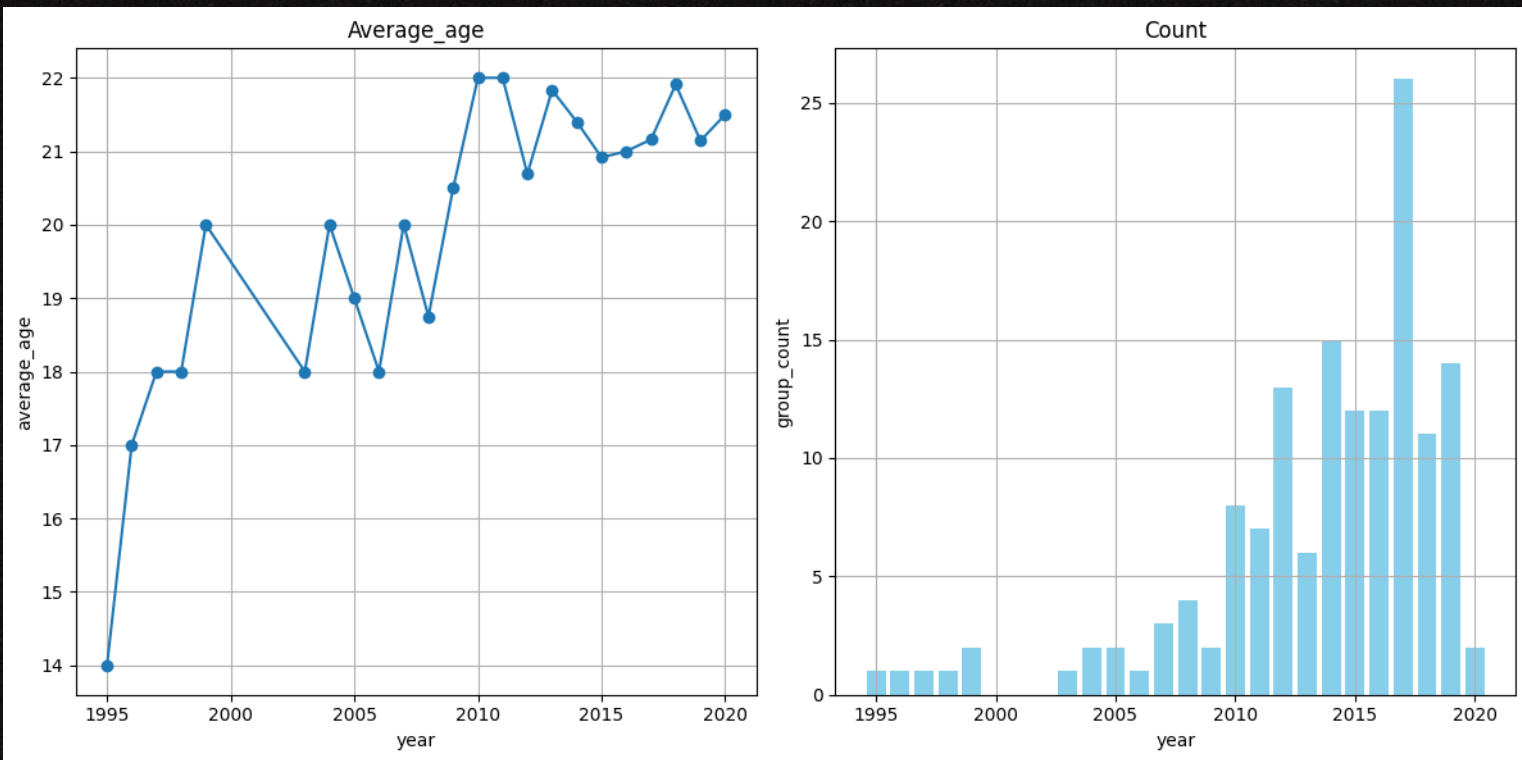
Detail Compact Column 8 of 8 columns

Date	Artist	Song Name	Korean Na...	Director	Video	Type	Release
2020-05-22	Agust D	Daechwita	대취타		https://youtu.be/qGjAWJ2zWWI	Boy Solo	Major
2020-05-21	Yubin	yaya (Me Time)	넵넵		https://youtu.be/BJD0arHF_5c	Girl Solo	Major
2020-05-21	OnlyOneOf	Angel (prod. GRAY)			https://youtu.be/vVTo8p72FxQ	Boy	Major
2020-05-20	Ryu Sujeong	Tiger Eyes			https://youtu.be/aE6curPGQRY	Girl Solo	Major
2020-05-20	Crush	Mayday feat. Joy (Red Velvet)	자나깨나		https://youtu.be/29ycT6fA-Rs	Boy Solo	Major
2020-05-19	NCT 127	Punch			https://youtu.be/U080S13V4po	Boy	Major
2020-05-19	SECRET NUMBER	Who Dis?			https://youtu.be/mkYwq_CKpyw	Girl	Major
2020-05-18	REDSQUARE	ColorFull			https://youtu.be/we1wqjcsYBU	Girl	Major

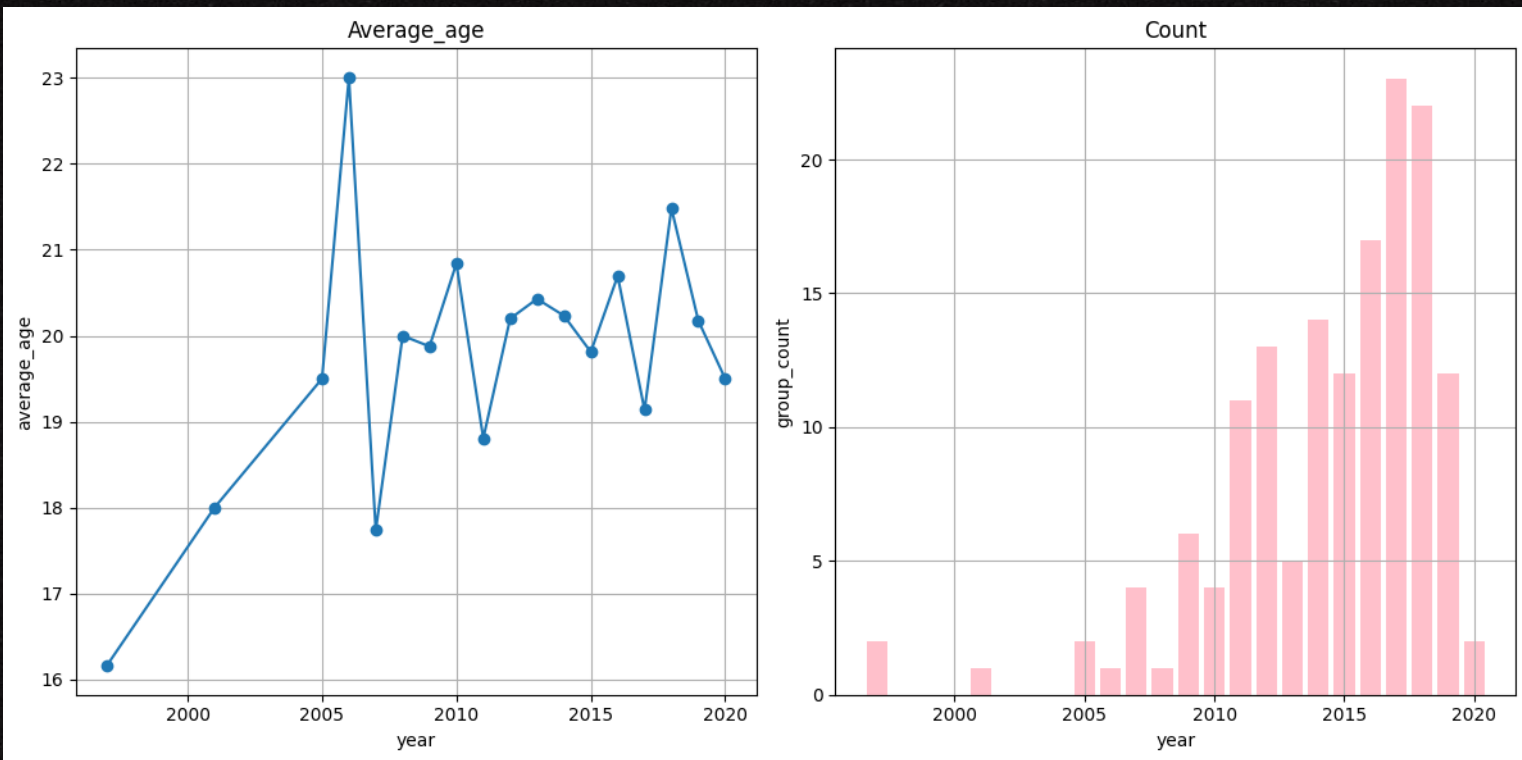
爬蟲結果：

	Name	Date	Likes	Views	Comments
0	Agust D	2020-05-22	14797047	441791136	1368655
1	Yubin	2020-05-21	53399	1184969	2722
2	OnlyOneOf	2020-05-21	84254	7569079	3995
3	Ryu Sujeong	2020-05-20	43251	2663847	3500
4	Crush	2020-05-20	663484	18084352	31972

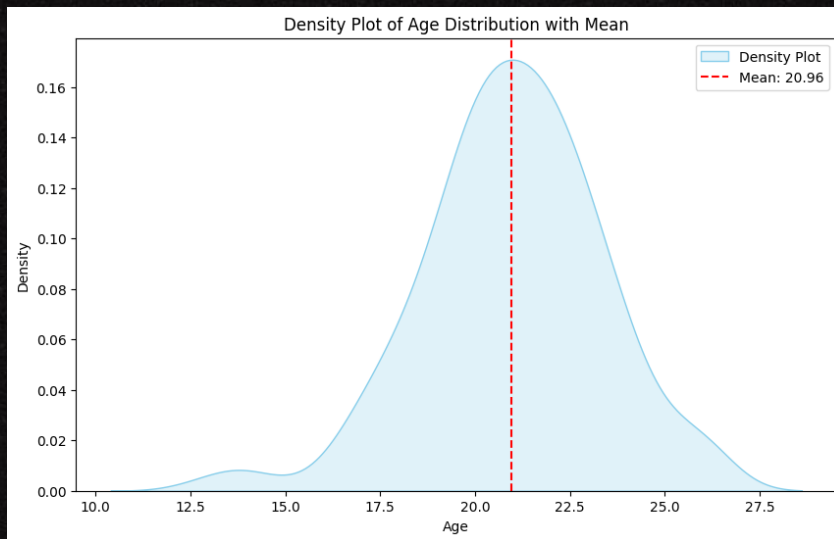
描述統計：男團平均出道年齡的時間趨勢



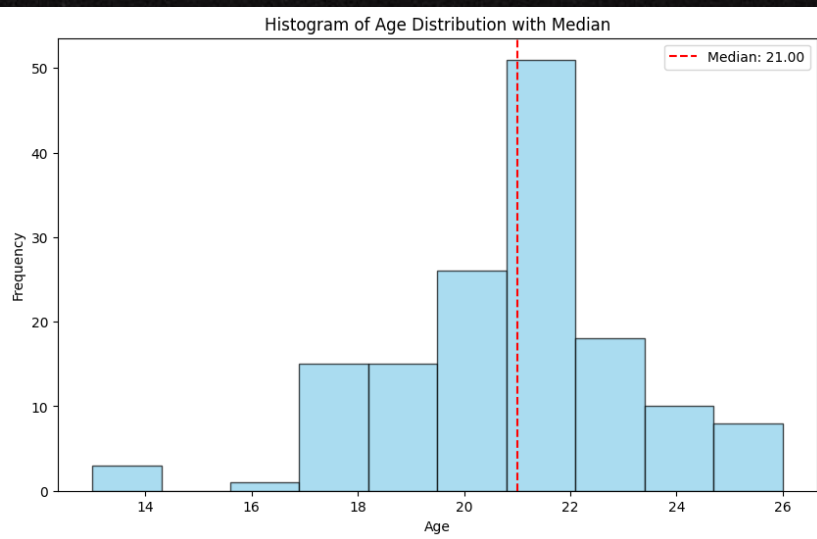
描述統計：女團平均出道年齡的時間趨勢



描述統計：男團平均出道年齡樣本數據中心趨勢

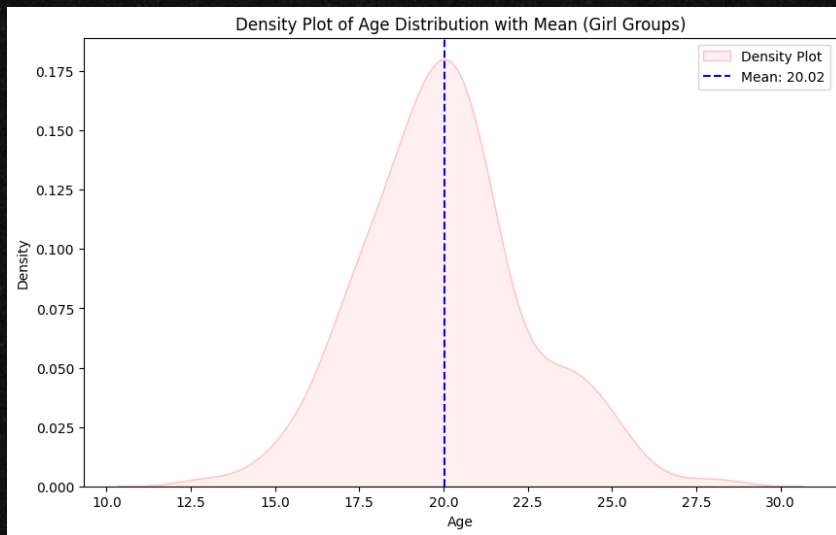


算術平均數為 20.96歲（四捨五入至小數點後第二位）

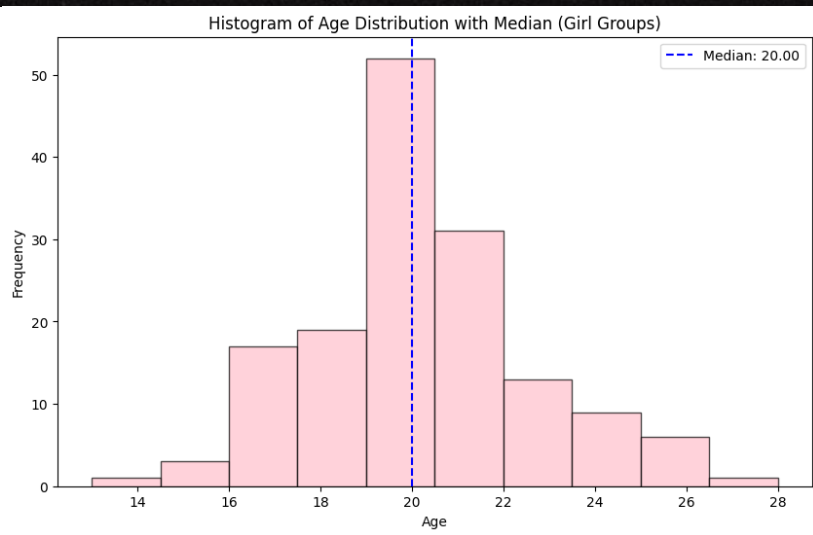


中位數為21歲

描述統計：女團平均出道年齡樣本數據中心趨勢

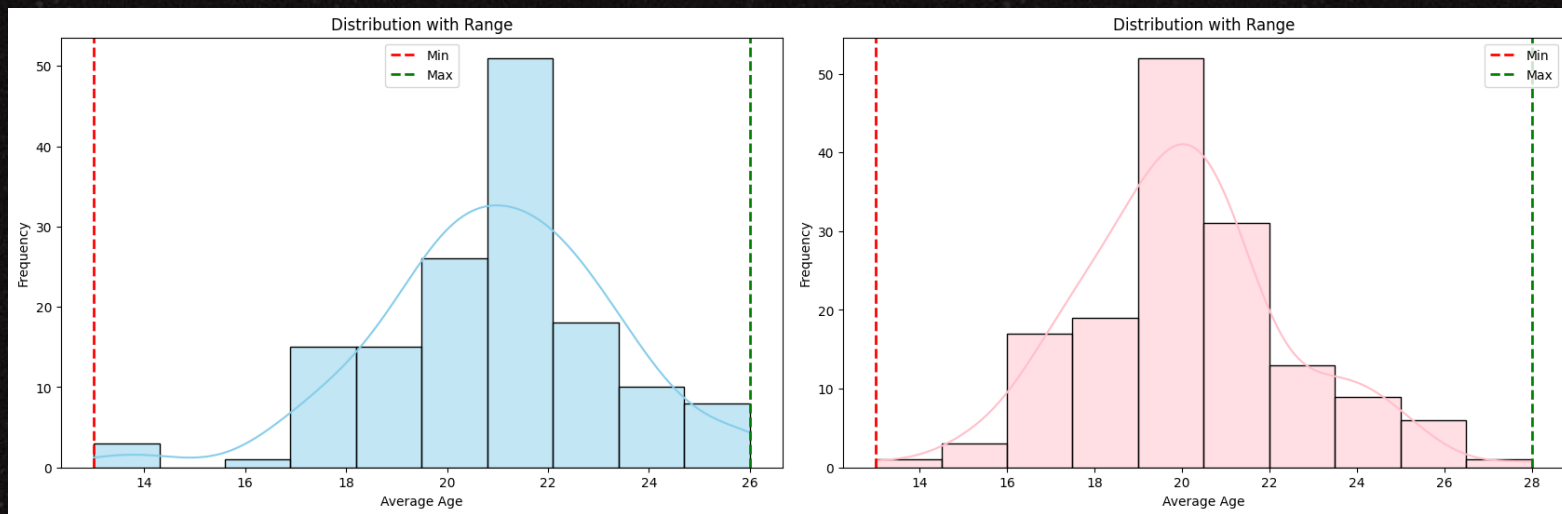


算術平均數為 20.02歲（四捨五入至小數點後第二位）



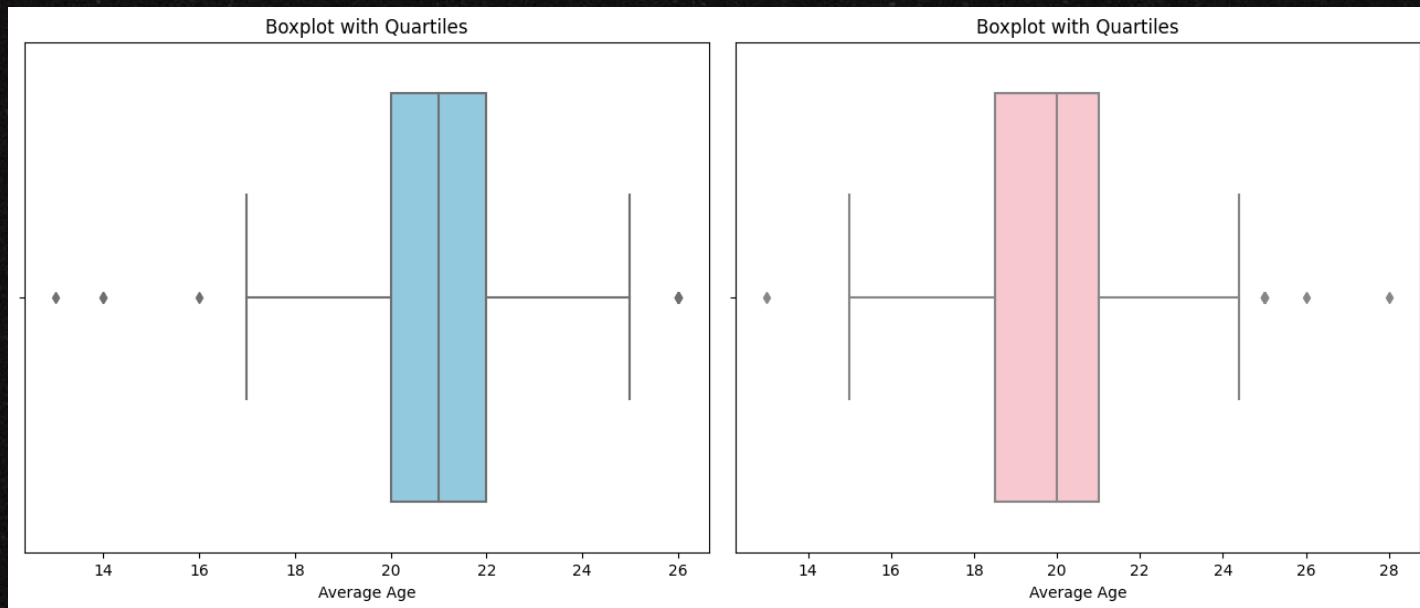
中位數為20歲

描述統計：男女團平均出道年齡樣本數據散佈度



男團平均出道年齡樣本數據範圍為13(歲)；女團平均出道年齡樣本數據範圍為15(歲)

描述統計：男女團平均出道年齡樣本數據散佈度



男團平均出道年齡樣本數據之Q1:20.0, Q2:21.0, Q3:22.0

女團平均出道年齡樣本數據之Q1:18.5, Q2:20.0, Q3:21.0

描述統計：男女團平均出道年齡樣本數據散佈度

- 標準差（四捨五入至小數點後第二位）：

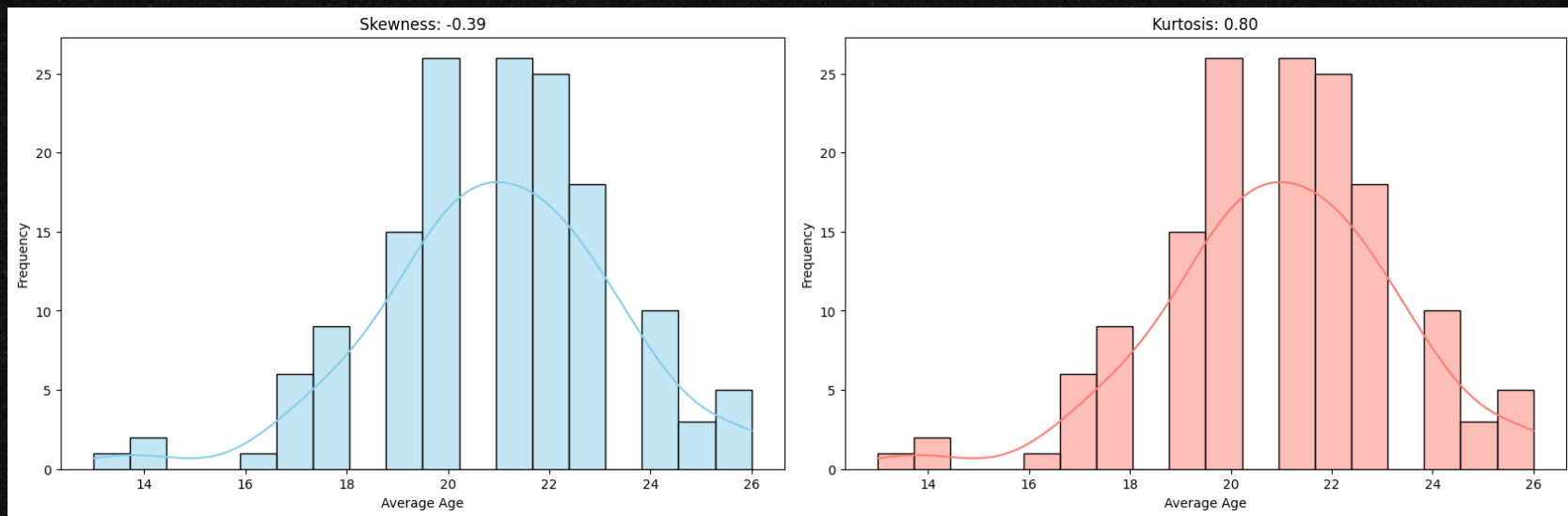
男團爲2.35；女團爲2.41

- 變異數（四捨五入至小數點後第二位）：

男團爲5.51；女團爲5.84



描述統計：男團平均出道年齡樣本數據分佈形狀



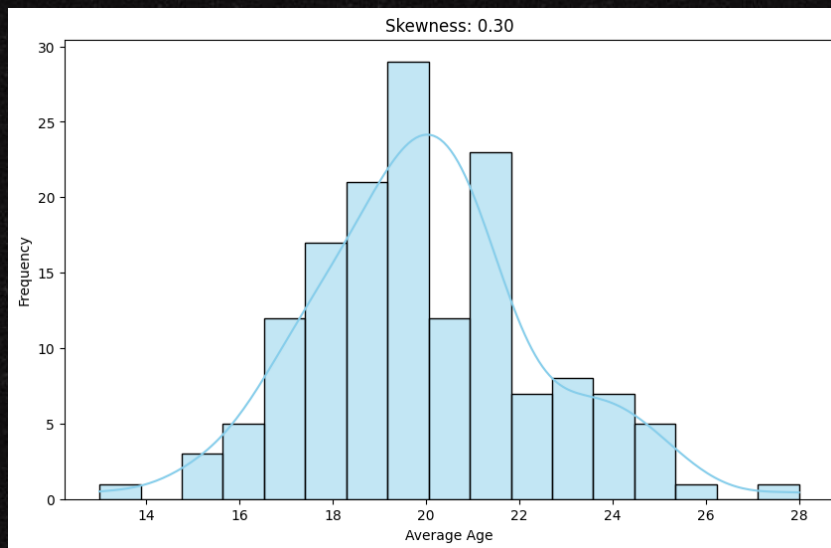
偏度：-0.39（四捨五入至小數點後第二位）

峰度：0.80（四捨五入至小數點後第二位）

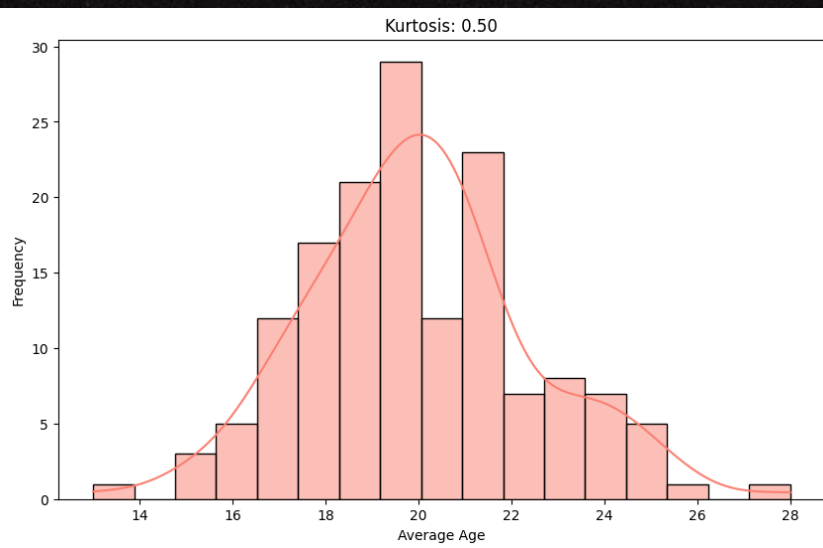




描述統計：女團平均出道年齡樣本數據分佈形狀



偏度：0.30（四捨五入至小數點後第二位）



峰度：0.50（四捨五入至小數點後第二位）





關聯分析：載入資料

	Name	Short	Debut	Company	Members	Orig. Memb.	Fanclub Name	Active	Average_age
0	100%	NaN	2012/9/18	TOP Media	4	7	Perfection	Yes	20.0
1	14U	NaN	2017/4/17	BG	14	14	NaN	Yes	22.0
2	1the9	NaN	2019/2/9	MBK	9	9	NaN	Yes	19.0
3	24K	NaN	2012/9/6	Choeun	8	6	24U	Yes	22.0
4	2AM	NaN	2008/6/21	JYP, Big Hit	4	4	I Am	No	20.0

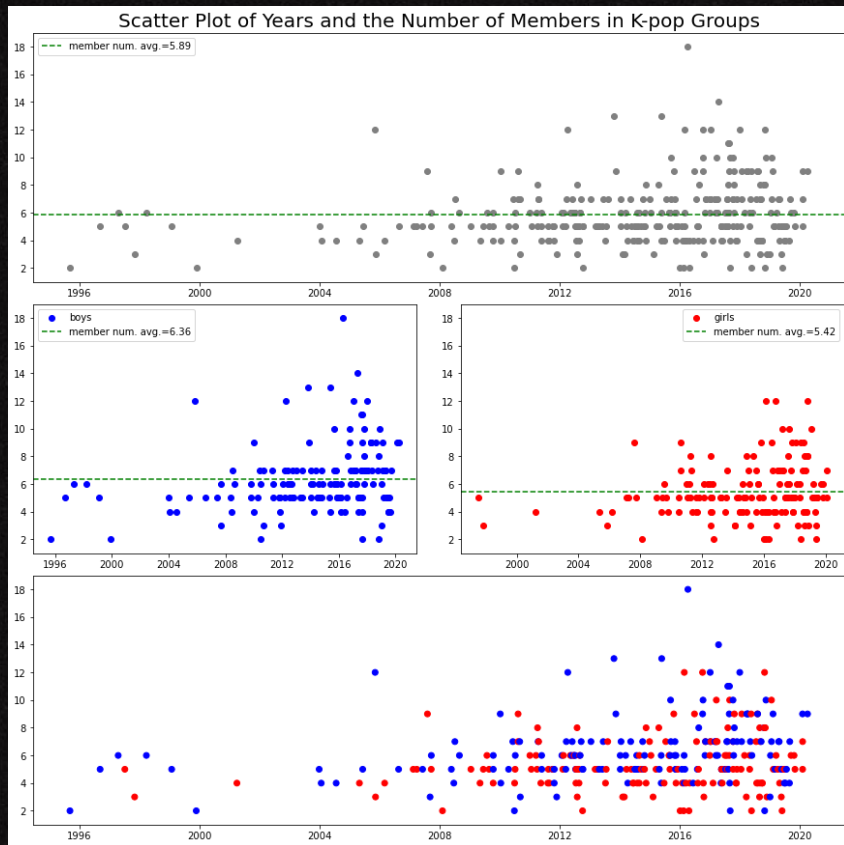
	Name	Short	Debut	Company	Members	Orig. Memb.	Fanclub Name	Active	Average_age
0	(G)I-DLE	NaN	2018/5/2	Cube	6	6	NaN	Yes	20.0
1	15&	NaN	2012/10/5	JYP	2	2	NaN	No	15.0
2	2EYES	NaN	2013/7/20	SidusHQ	4	5	NaN	No	19.0
3	2NE1	NaN	2009/5/6	YG	4	4	Blackjack	No	21.0
4	3YE	NaN	2019/5/21	GH	3	3	NaN	Yes	21.0



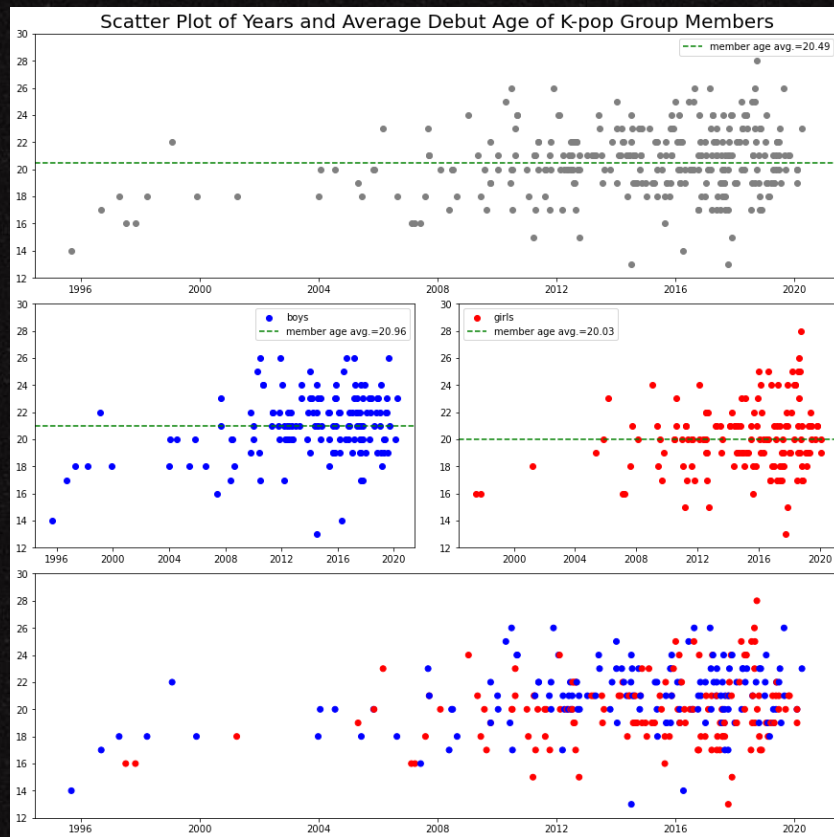
關聯分析：資料預處理

	Debut	Company	Orig. Memb.	Average_age	Gender	Company Code
Name						
Turbo	1995-09-06	MBK, Turbo Co.	2	14.0	M	0
H.O.T	1996-09-07	SM	5	17.0	M	1
SECHKIES	1997-04-15	DSP, YG	6	18.0	M	2
Baby V.O.X	1997-07-05	DR Music	5	16.0	F	109
S.E.S	1997-11-01	SM	3	16.0	F	1

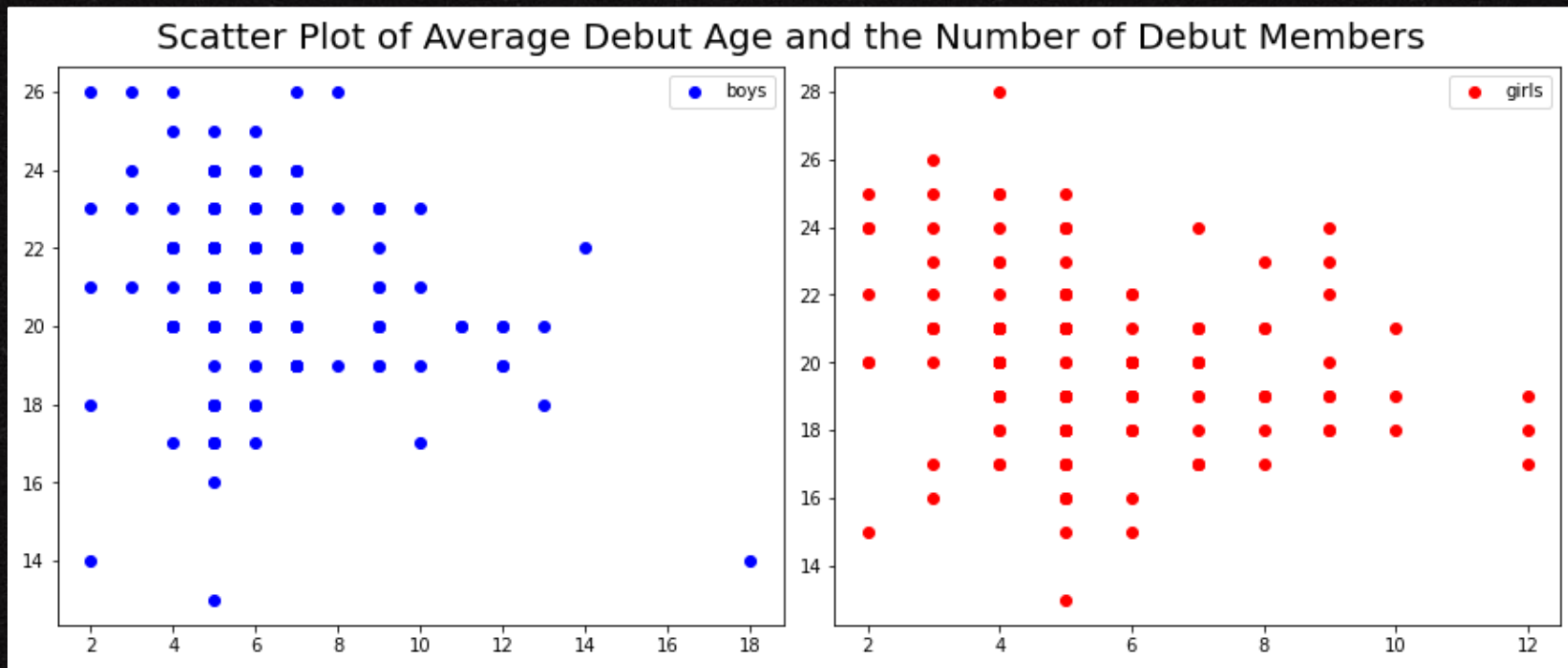
關聯分析：年份與出道團員數量散布圖



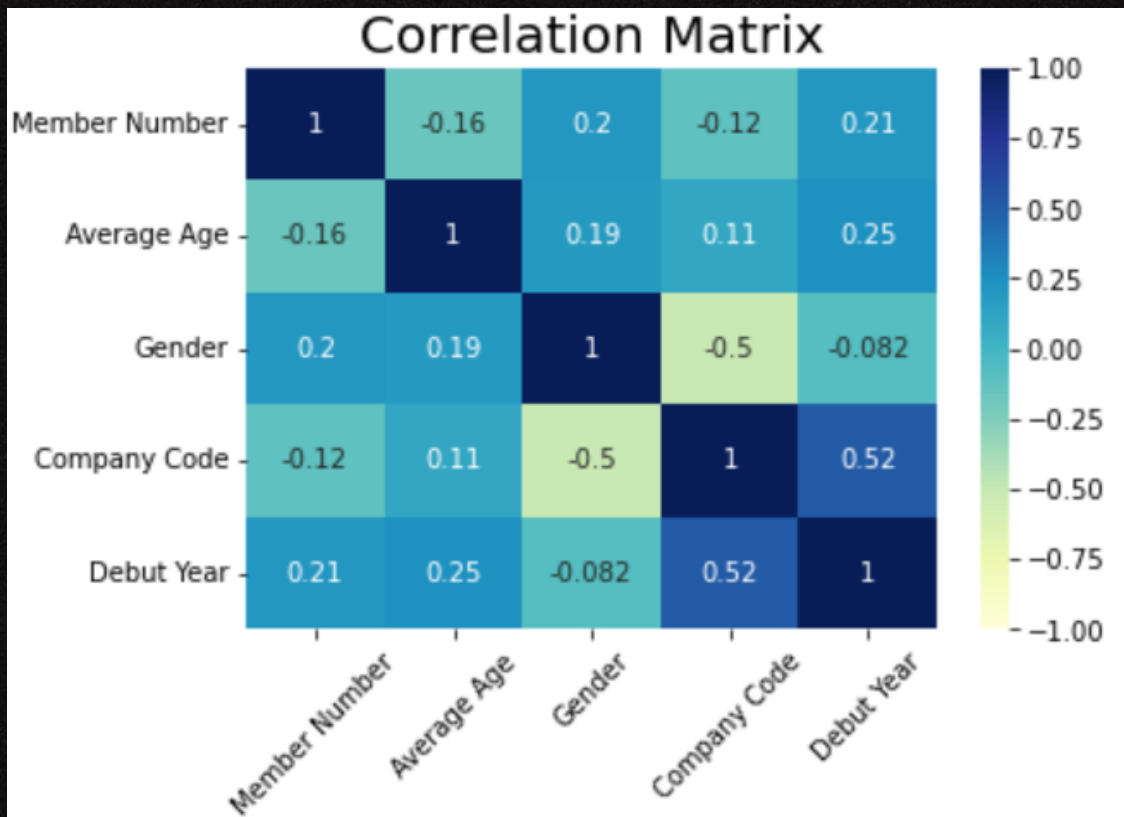
關聯分析：年份與平均出道年齡散布圖



關聯分析：平均出道年齡與出道團員數量散布圖



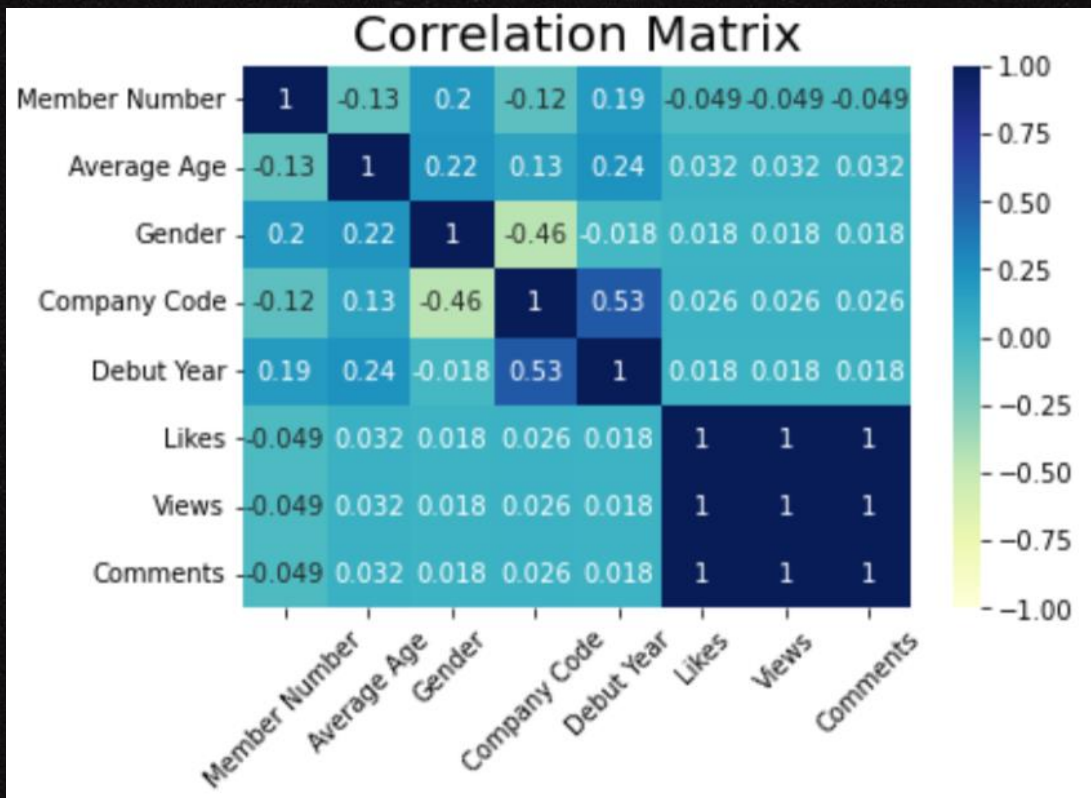
關聯分析：相關矩陣



關聯分析：多元迴歸分析

	Debut	Company	Member Number	Average Age	Gender	Company Code	Debut Year	Likes	Views	Comments
Name										
H.O.T	1996-09-07	SM	5	17.0	1	1	1996	393	90719	54
S.E.S	1997-11-01	SM	3	16.0	0	1	1997	5455	140624	249
Shinhwa	1998-03-24	SM, Good, Shinhwa	6	18.0	1	3	1998	1334	101984	104
Jewelry	2001-03-30	Star Empire	4	18.0	0	17	2001	422	73554	40
TVXQ!	2003-12-26	SM	5	18.0	1	1	2003	13966	1284981	1518

關聯分析：多元迴歸分析





關聯分析：多元迴歸分析

- 讚數

Score: 0.004734814692184175

- 觀看數

Score: 0.004734814691657152

- 留言數

Score: 0.004734814692189837





總結與討論





感謝您的聆聽與指教!

