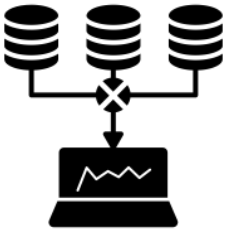
Data Description & Problem Statement

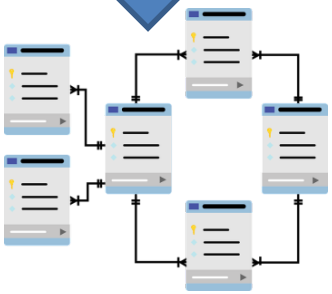
Data Source:
https://cricsheet.org/

```
    total: 0
- 3.4:
    batsman: DJ Thornely
    bowler: Z Khan
    non_striker: ST Jayasuriya
    runs:
      batsman: 0
      extras: 0
      total: 0
    wicket:
      kind: retired hurt
      player_out: DJ Thornely
- 3.5:
    batsman: RV Uthappa
    bowler: Z Khan
    non_striker: ST Jayasuriya
    runs:
      batsman: 2
      extras: 0
      total: 2
```

Ball by Ball Data attributes of every IPL match



Match wise Data aggregation



Relational DB creation

2019

2019 session's data

Extraction of Player level features of whole session

EDA and Statistical Analysis

Final finding & conclusion

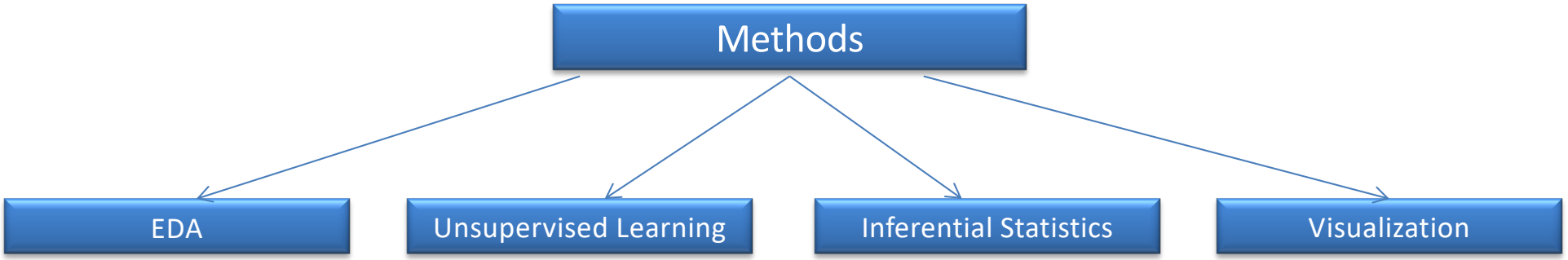
What are the impacts of the following factors on a session's outcome?

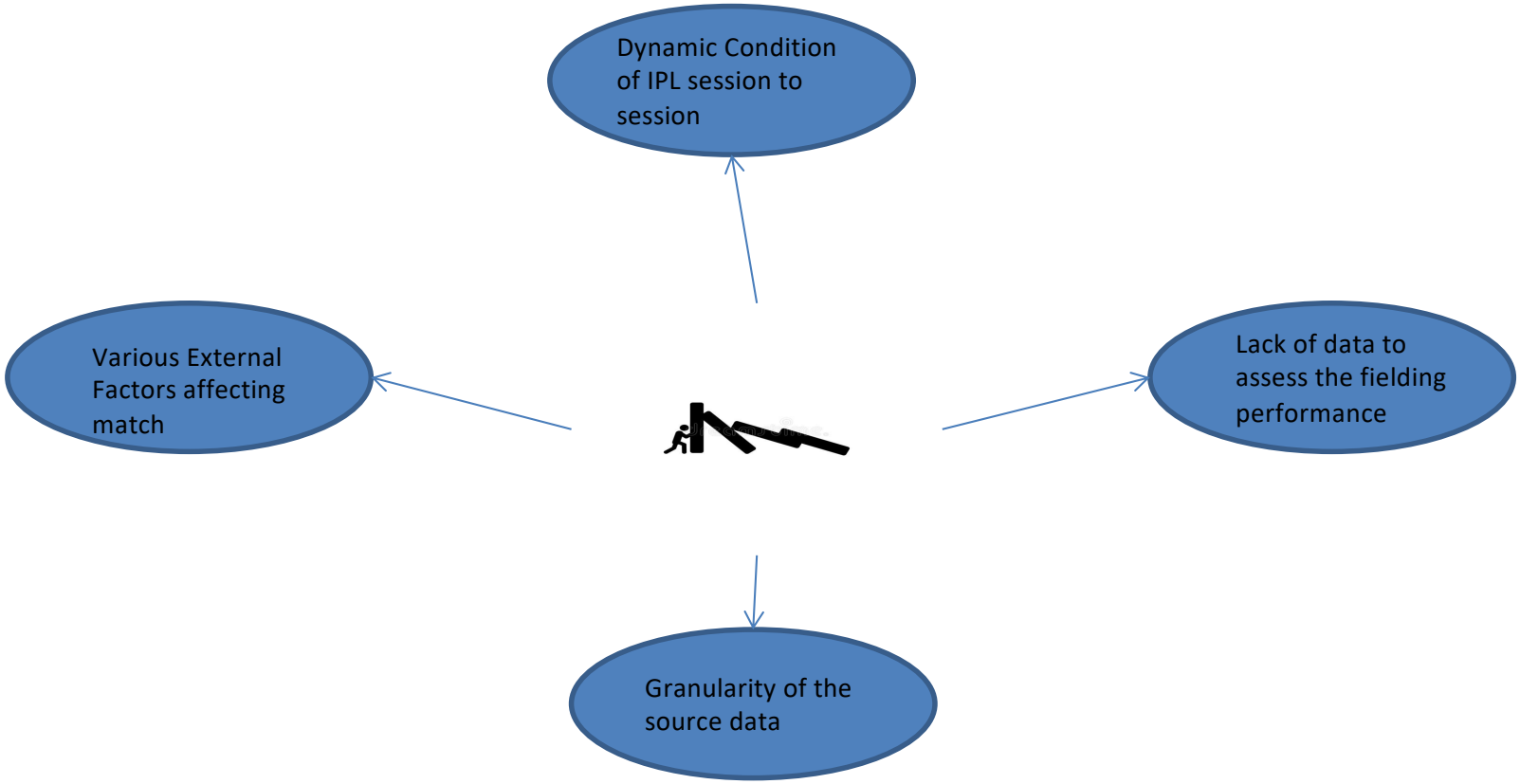
Toss

Batting Performance

Bowling Performance

Methodology & Challenges faced

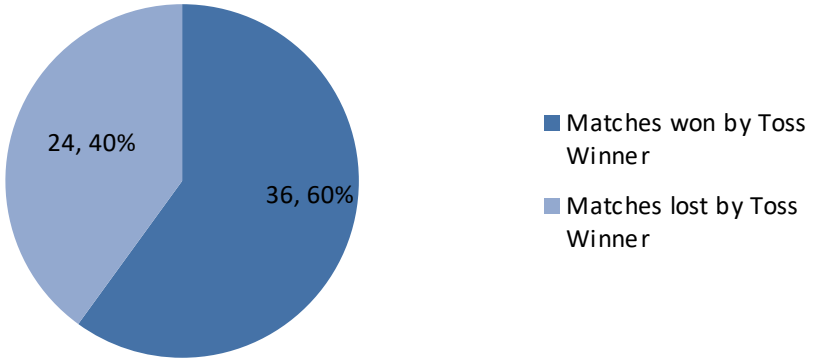




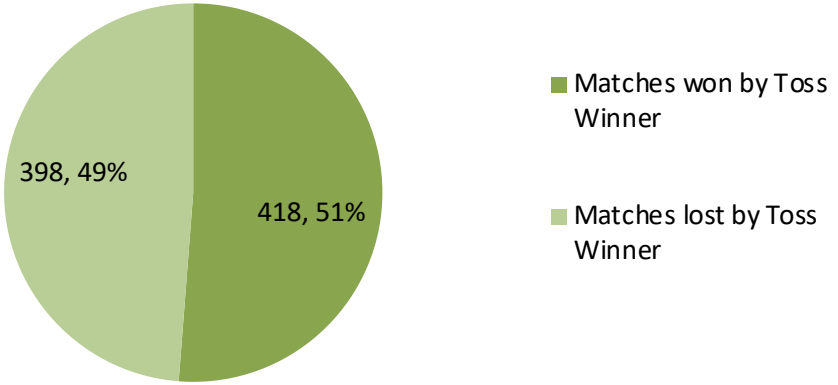
Impact of Toss in 2019

Does winning the toss actually have positive impact in winning the match?

Pie Chart Showing the Effect of toss on match outcome in 2019



Pie Chart Showing the effect of toss on match outcome in all sessions



Can not draw inference on the impact of toss!!

Statistical Hypothesis Testing

Let p be probability of winning match by toss winner.

H0: p=0.5 vs H1: p>0.5

f be number matches won by toss winner, n be total number of matches

Test statistic $Z=\frac{\left(\frac{f}{n}\right)-0.5}{\sqrt{0.5*(1-0.5)}}\sqrt{n}$ which under H0 follows N(0,1) aproximately. For this test, z=1.549193

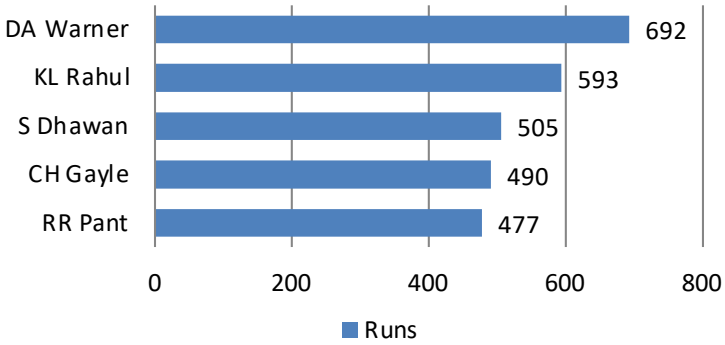
P value of the test =p=0.06066767 > 0.05

At 95% confidence the null hypothesis can not be rejected.

Failed to conclude that winning the toss have positive impact in winning a match!!!!!!

Batting Performance Analysis in IPL 2019

Top 5 Batsmen scoring Most Runs in 2019 IPL



Top 5 Players with Highest Strike rate in IPL 2019

How to choose the best batsmen ?

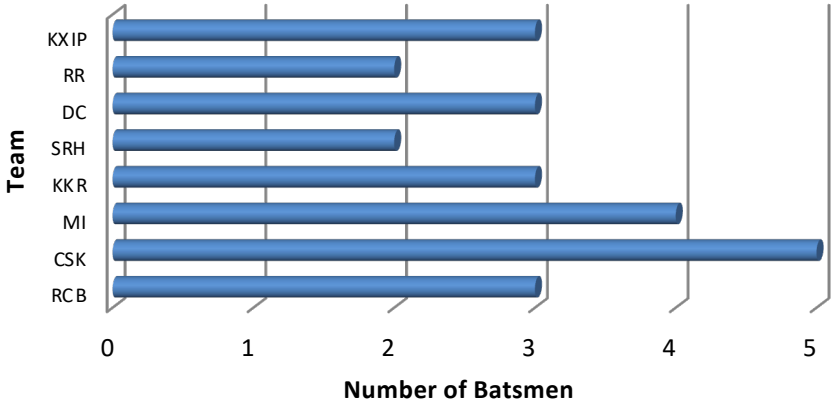
Clustering

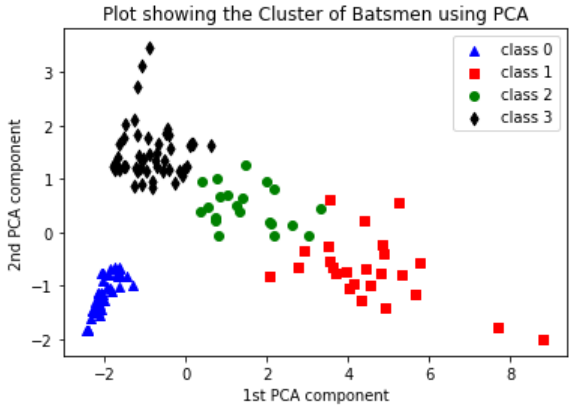
How many clusters to create? → From Elbow Plot

Table showing cluster wise average values of Batting features of players

Features	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Runs	6.40	**423.16**	216.19	50.32
Fours	0.36	**38.72**	18.95	4.09
Sixes	0.16	**18.12**	8.52	2.36
Balls	8.86	**299.40**	161.76	37.98
# Fifty	0	**3.08**	0.95	0.08
# Hundred	0	**0.20**	0.05	0
Strike Rate	52.34	**143.18**	134.58	147.30
Out	1.58	**11.52**	7.38	2.81
Average	3.67	**39.18**	32.31	17.69
No. of Players	50	**25**	21	47

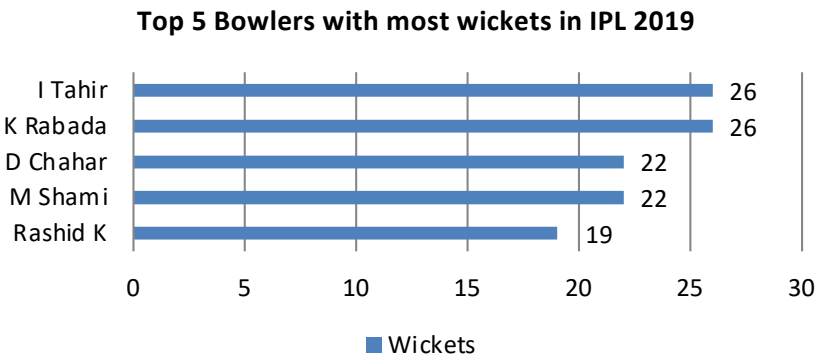
Team wise distribution of Batsmen in premium cluster





MI and CSK were more dominating in the batting performance

Bowling Performance Analysis in IPL 2019



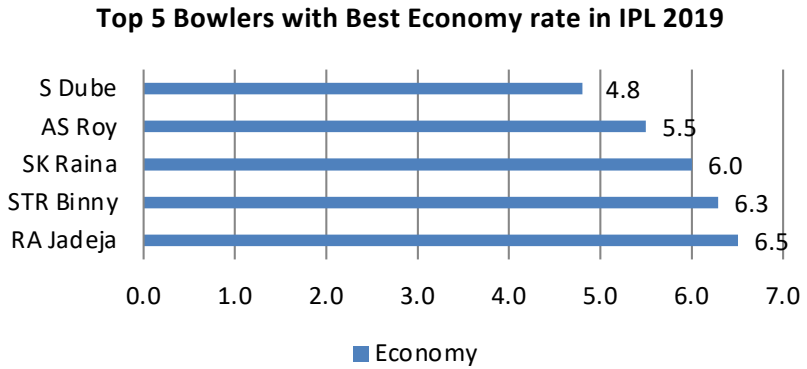
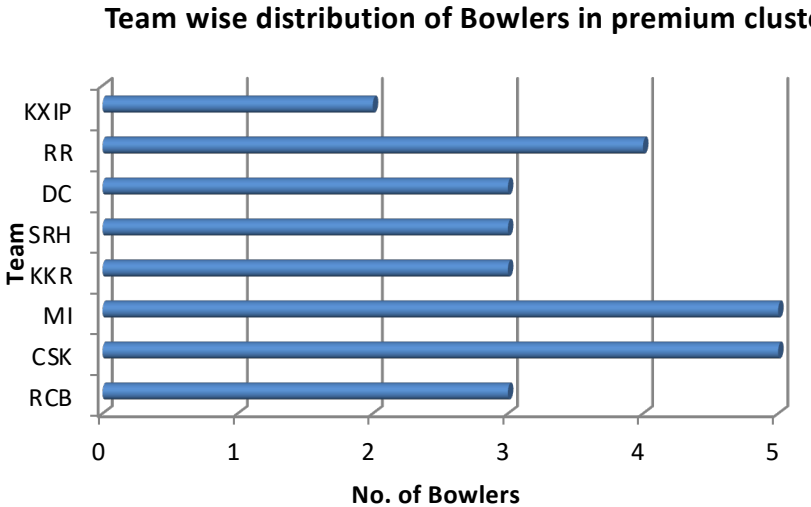
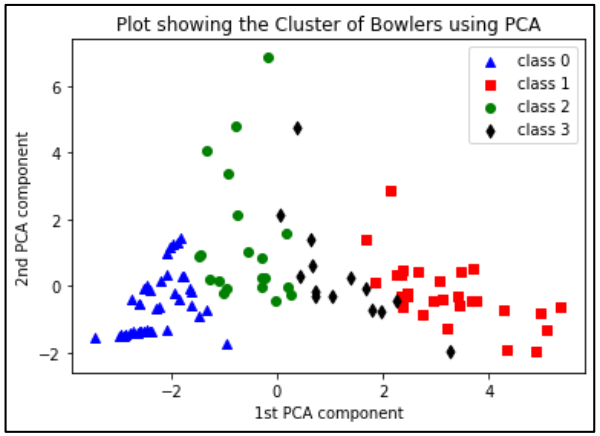


Table showing cluster wise average values of Bowling features of players

Features	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Balls	26.40	**276.36**	93.40	171.07
Wickets	1.17	**14.68**	3.50	8.43
Extra	1.57	**15.29**	7.60	10.93
Runs	41.81	**371.79**	142.55	256.14
# Four	3.29	**31.39**	11.95	22.86
# Six	2	**14.61**	6.15	9.64
Economy Rate	10.06	**8.21**	9.46	9.09
Average	18.16	**29.39**	58.92	40.07
Three	0.05	**1.36**	0.15	1
Five	0.02	**0.04**	0	0
Strike rate	12.15	**21.08**	37.00	26.79
No. of Players	42	**28**	20	14





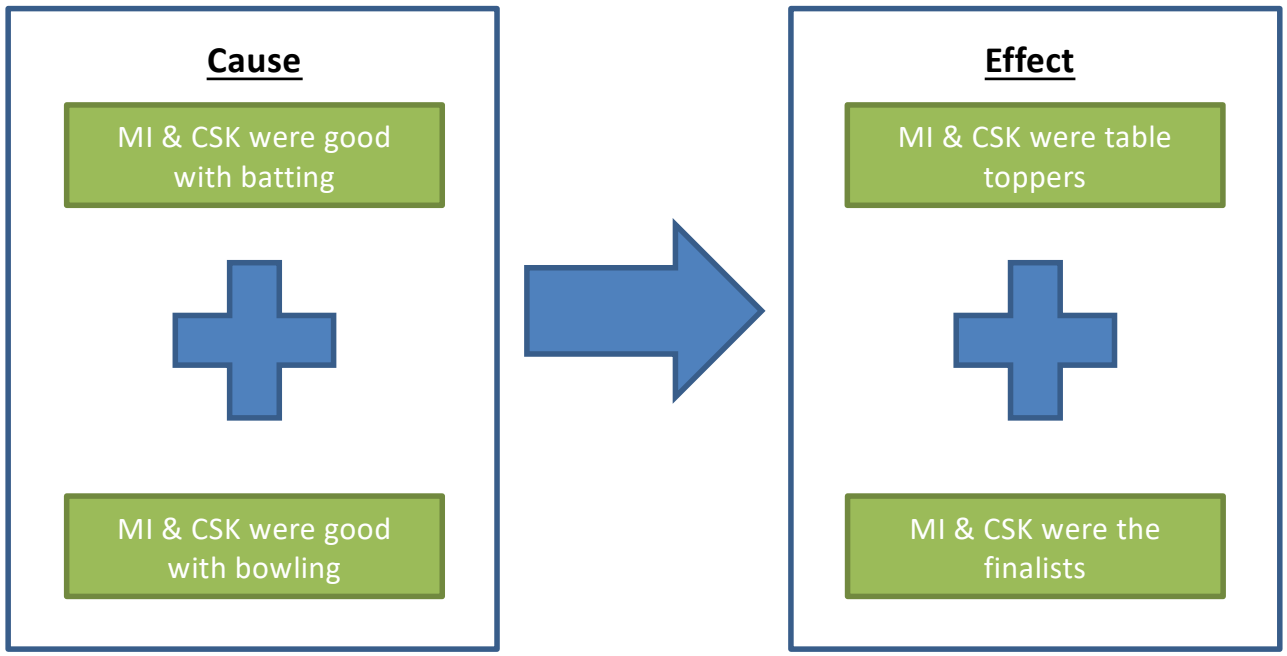
MI and CSK were dominating in Bowling performance also

Conclusion

Can not conclude that winning toss helps in winning match.

CSK and MI had the most number of splendid batsmen.

In bowling department also CSK and MI have outperformed other teams.



Performing well in both batting and bowling is the key to have a good IPL session!!!!!

Individual Contribution

Aditya Patel : Project Idea, Data Aggregation, Database Creation

Rahul Kumar Dev : Project Idea, Data Pre processing, Feature Engineering

Soumalya Nandi : Project Idea, Statistical Data Analysis, Visualization

Appendix

Data Snippet in csv format

over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	total_runs	r
6	5	RT Ponting	BB McCullum	AA Noffke	1	0	1	
6	6	BB McCullum	RT Ponting	AA Noffke	1	0	1	
7	1	BB McCullum	RT Ponting	Z Khan	0	0	0	
7	2	BB McCullum	RT Ponting	Z Khan	1	0	1	
7	3	RT Ponting	BB McCullum	Z Khan	1	0	1	
7	4	BB McCullum	RT Ponting	Z Khan	1	0	1	
7	5	RT Ponting	BB McCullum	Z Khan	1	0	1	
7	6	BB McCullum	RT Ponting	Z Khan	1	0	1	
8	1	BB McCullum	RT Ponting	JH Kallis	0	0	0	
8	2	BB McCullum	RT Ponting	JH Kallis	0	0	0	
8	3	BB McCullum	RT Ponting	JH Kallis	0	0	0	
8	4	BB McCullum	RT Ponting	JH Kallis	1	0	1	
8	5	RT Ponting	BB McCullum	JH Kallis	1	0	1	

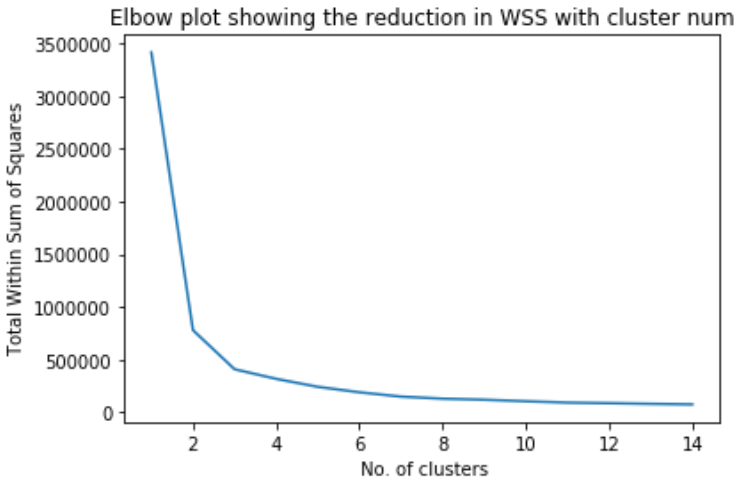
Variability explained ratio by PCA for bowler

1st PCA : 0.51786309
2nd PCA: 0.1909745

Variability explained ratio by PCA for batsman

1st PCA : 0.62904852
2nd PCA: 0.16531433

Elbow Plot for Bowlers



Elbow Plot for Batsmen