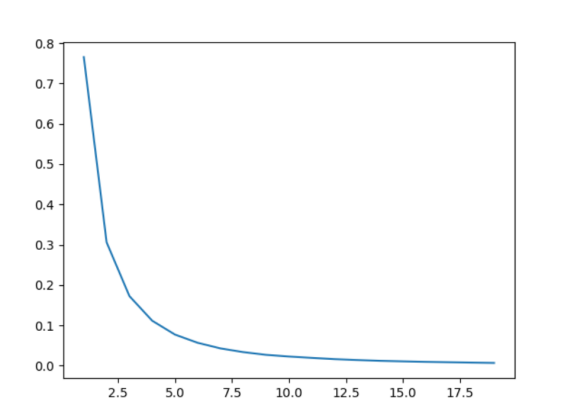
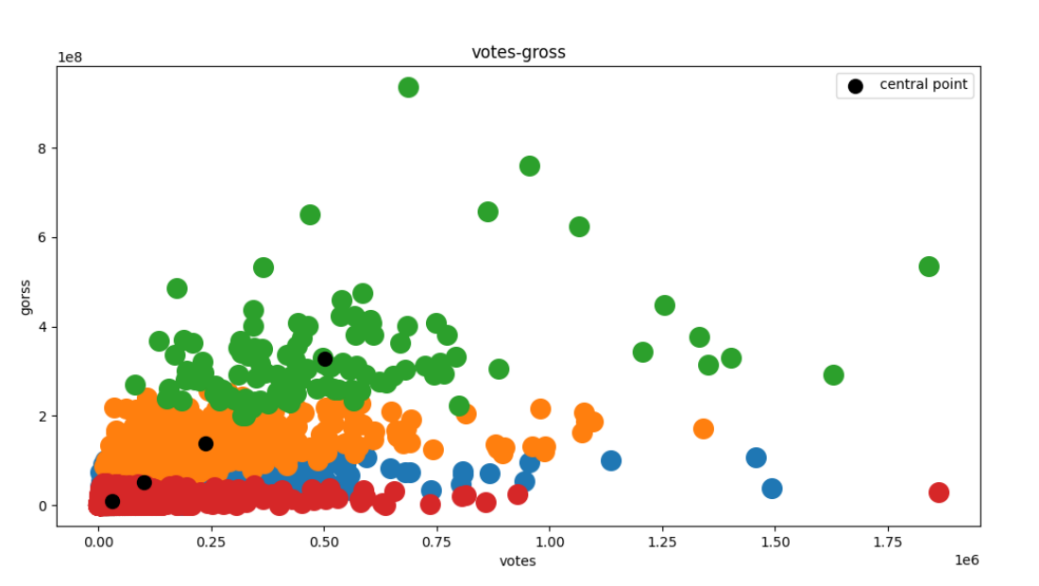
**Homework 3**

1.

a.



According to the calculation, the optimal value of K is 4.



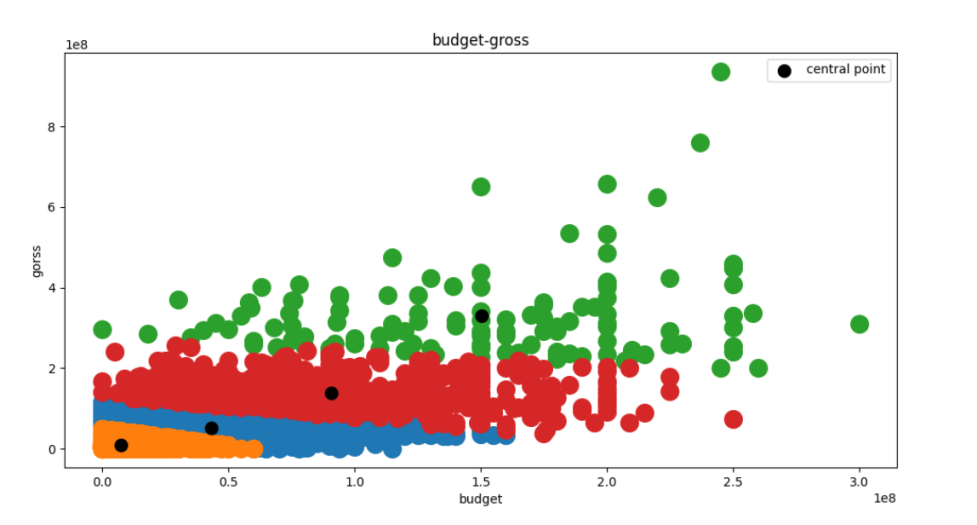
RED: Most of the values of vote are clustered in the range of 0-0.75, and only a very few exceed 0.75. The values of gross are below 1

BLUE: The values of the blue clusters overlap well with the red and orange clusters. More vote values over 0.75 compared to red clustering.

ORANGE: Most of the values of gross are concentrated between 1-2. The overall vote value is higher than the blue and red clusters

GREEN: The values of gross are more dispersed compared to other clusters. Although the values of votes are still mostly concentrated between 0 and 0.75, the gross values are significantly larger than the other three clusters

The higher the number of votes, the higher the gross



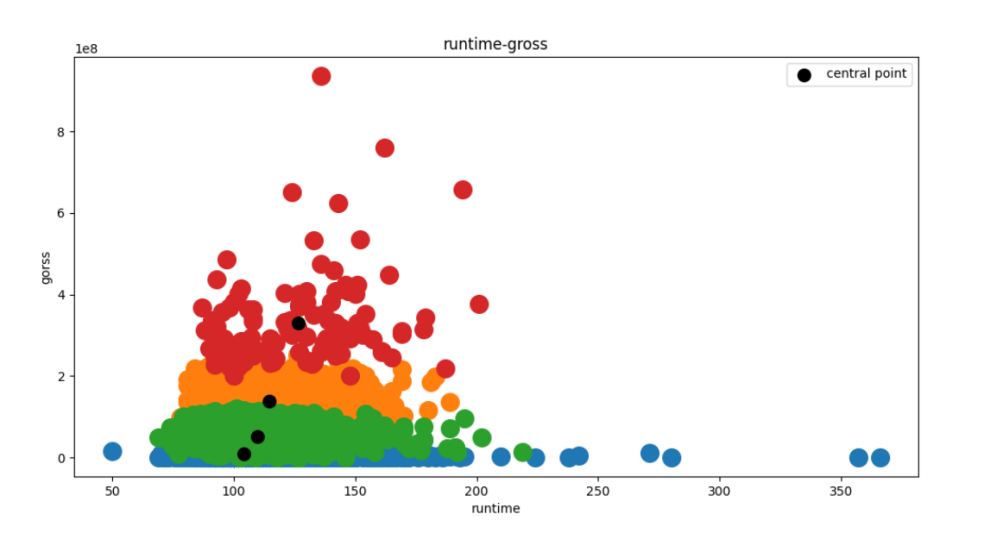
ORANGE: The values of budget are generally concentrated in the range of 0-0.5 and the values of gross are below 1. Compared with other clusters, the values of this cluster are more concentrated

BLUE: The values of budget are generally concentrated in the range of 0-1.5. The values of gross are distributed between 0 and 1, but are generally higher than the orange clusters and lower than the red clusters.

RED: The red clusters have mostly higher gross values than the blue clusters. budget values start to disperse relatively

GREEN: The green clusters have the most dispersed values. All the gross values are greater than the other three clusters

Budget low movie cluster’s gross is generally lower, budget high movie cluster’s gross is more likely to be high



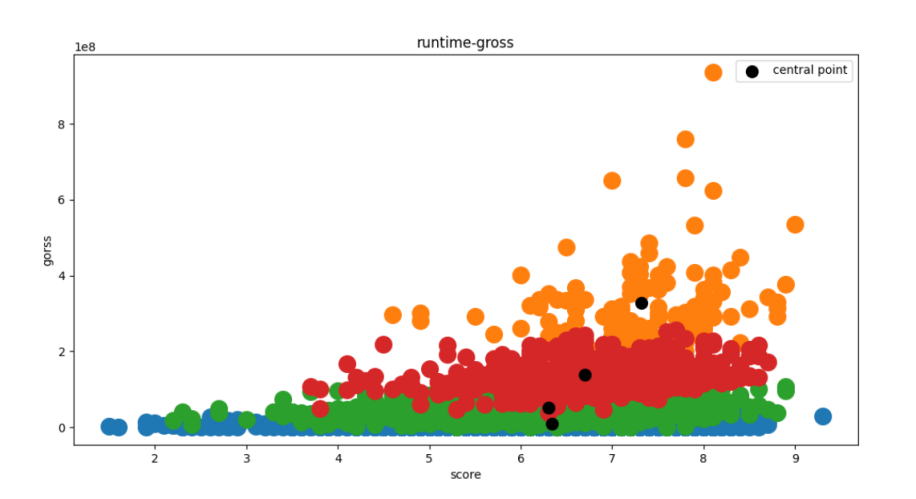
RED: The red clusters have the most dispersed values. The overall gross values are greater than the other three clusters

BLUE: The blue clusters have the most dispersed runtime values. There is no clear demarcation with the gross values of the green clusters.

ORANGE: The overlap with the red and green clusters is high. The cross values of orange clusters are generally larger than those of green and blue clusters

GREEN: The green clusters have a high overlap with the blue and orange clusters. runtime values are almost all concentrated between 50 and 200

Green, blue and orange clusters are not clearly demarcated. The higher the runtime the more likely the clustering gross is to be high



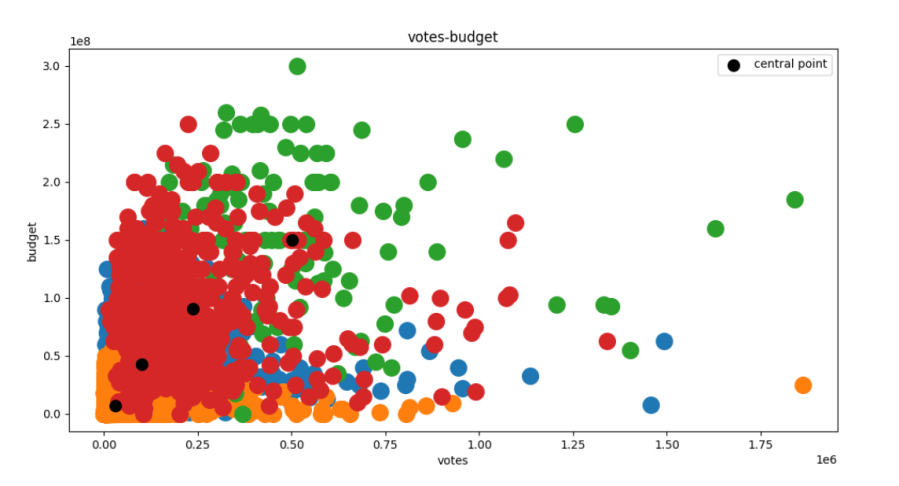
RED: The value distribution of the red clusters is more concentrated. There is a high degree of overlap with the green clusters and there is no clear division with the orange clusters

BLUE: The blue clusters have the most dispersed score distribution and the lowest average gross value. There is a high degree of overlap with the green clusters

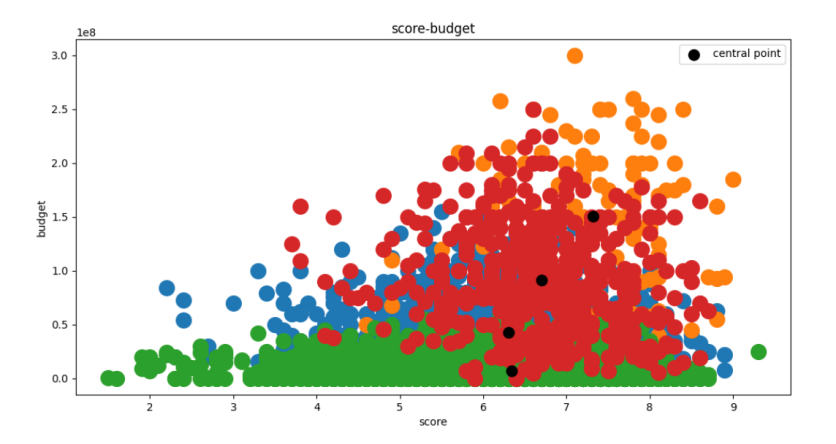
ORANGE: The distribution of the orange clustering values is the most dispersed, and the values of score and gross are the highest compared to the other clusters

GREEN: The distribution of the scores of the green clusters is also relatively scattered and has a high overlap with the blue and red clusters

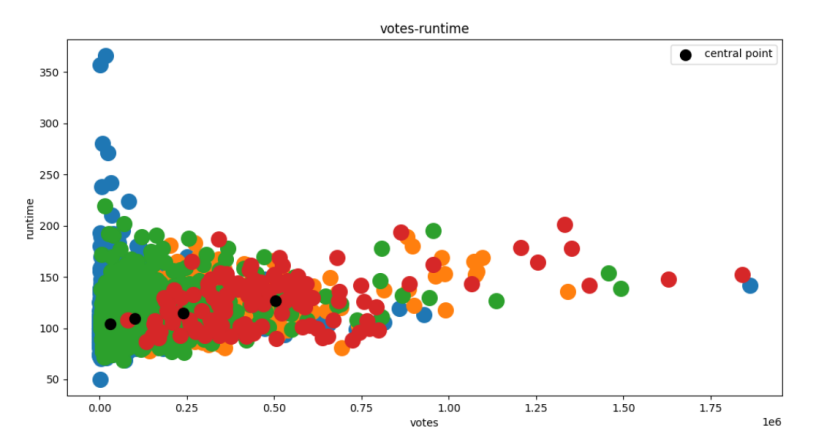
In clusters with gross values below 1, there is no significant difference in the scores. in clusters with gross values above 1, the higher the score the higher the gross



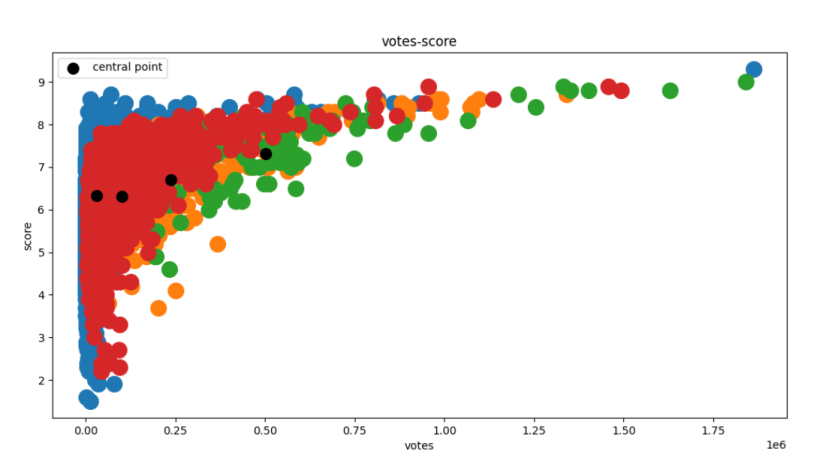
The higher the score the higher the general budget of the clusters. The four clusters have a high degree of overlap and unclear classification. Green clustering has the highest dispersion, followed by red clustering.



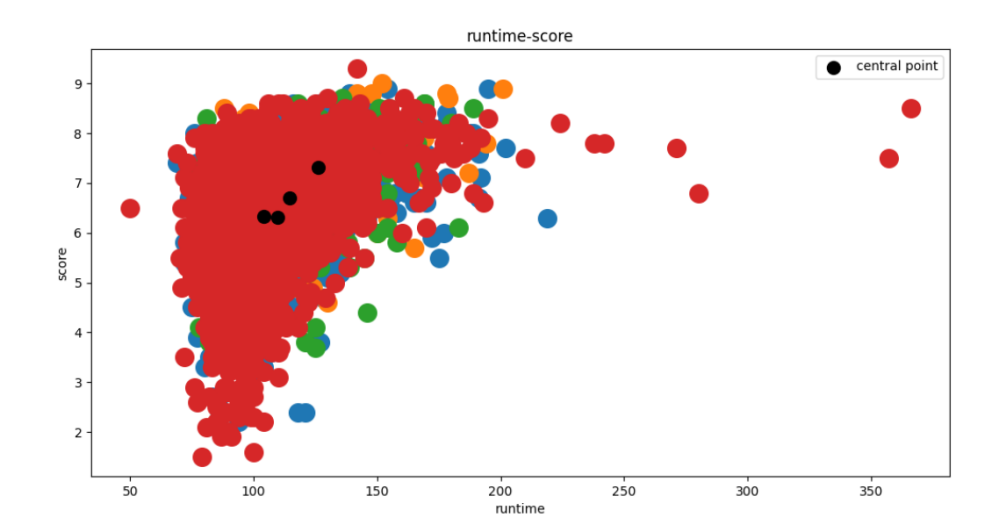
The higher the score, the higher the budget of the clusters in general. But the data of the four clusters overlap more. Green clusters have the highest dispersion of score values. Orange clusters have the highest dispersion of budget values.



The data of the four clusters cover almost one area. But the higher the runtime value the higher the value of the votes. The red clusters have the highest dispersion of votes values and the blue clusters have the highest dispersion of runtime values. But the centrists of the four clusters are not significantly different in the vertical coordinates.

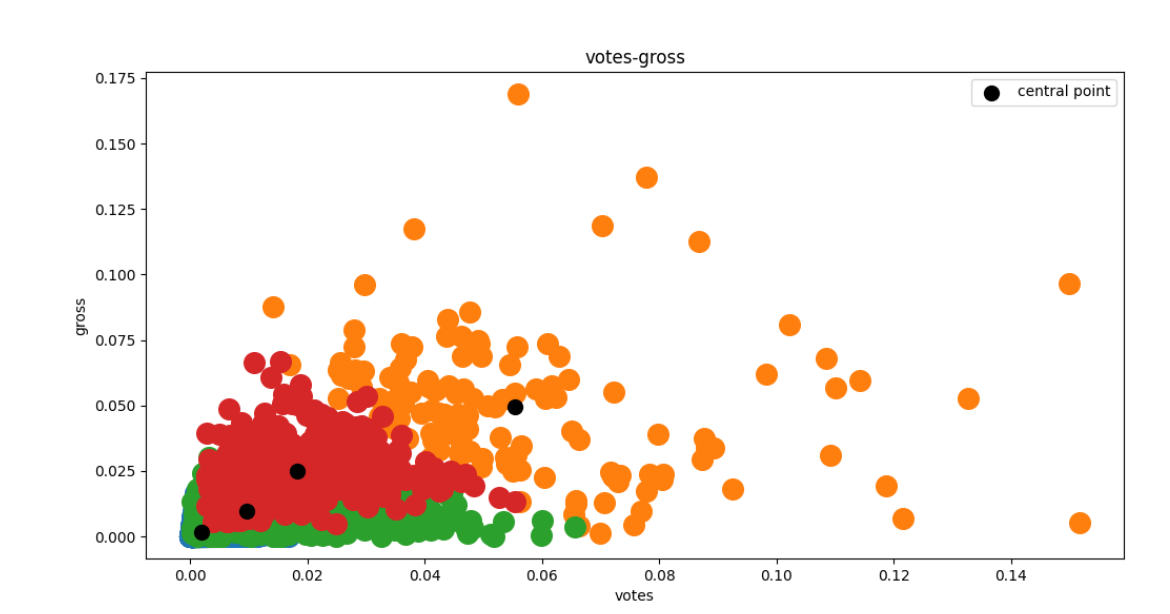
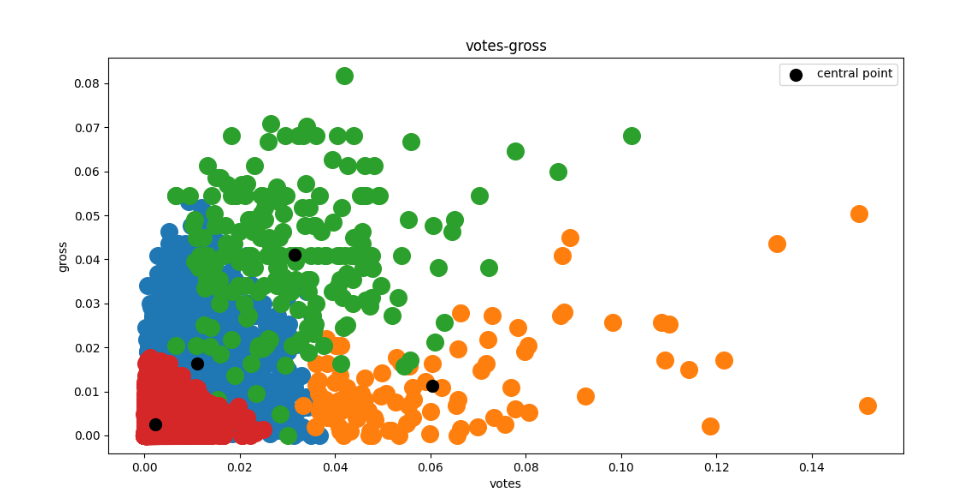
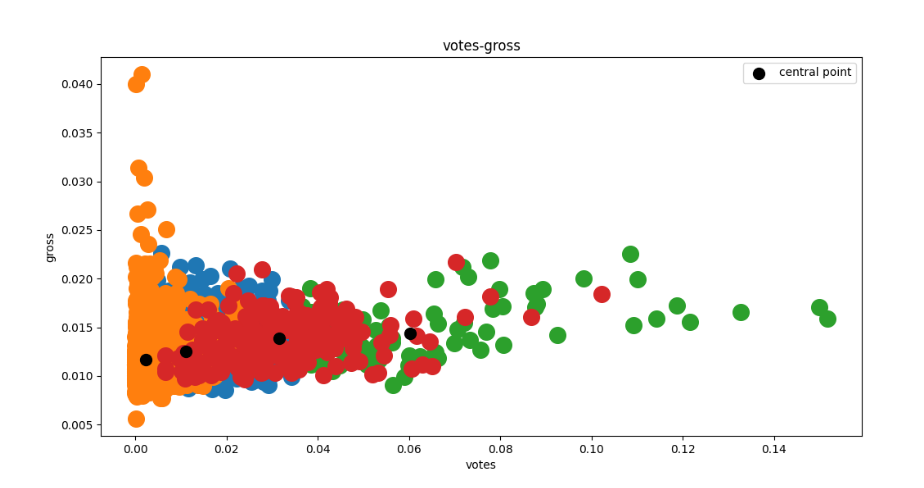


The distribution of the four clustering values is almost identical, with no clear classification nor a clear trend. Blue clusters have the highest dispersion of score values and green clusters have the highest dispersion of votes values.



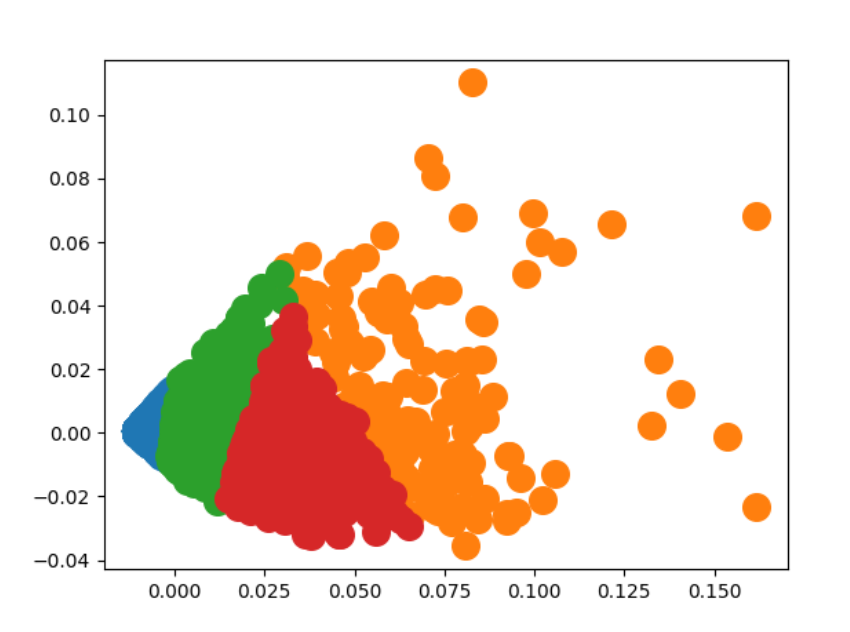
All values of the four clusters overlap almost entirely and the classification is not clear. The values of the four clusters almost overlap in the same area, and it is not possible to completely separate each cluster.

B.



The results are significantly better after regularization than the previous results. The classification of the results is much clearer

C.



The grouping of the four clusters is clearly divided into plants.

A list [0,0,0,0,0,0,0] is used to represent the parking spaces. Randomly insert four ‘1’ into the list to represent each of the four cars. Simulate the experiment 1,000,000 times and calculate the probability that parallel parking is not required. After several experiments, the results were calculated as 61.22%, 61.52%, 61.58% and 61.43%.

**Thus, the probability that have to parallel park is 38% - 39%**