Due to high number of stops in the bus route using dynamic programming would lead to very long computational time as the number of sub-problems grows exponentially with the number of stops. Hence the stops were classified into 3 major clusters. So that the dynamic programming can be performed in each small cluster and with reduced complexity.

1.0 Dividing the stops into 3 different clusters

1.10 K-means clustering algorithm

K-means clustering algorithm, "*It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.*"[1] Distance based measurements are used to identify the similarities between the data points.

The algorithm involves defining the value of "k" the number of clusters to be formed by the user. After number of clusters is decided the algorithm randomly selects "k" number of centroids. Using the Euclidean distance formula each data set is associated with a cluster. This process of associating the data set with a cluster is repeated until optimum solution is found.

2.0 Process of k-means algorithm

1. Each stop was given a co-ordinate, this was done by the co-ordinate plotting system of Google Maps.

Stops	**Co-ordinates**	
	X	**Y**
1	22.28822779	70.7602492
2	22.29037206	70.78024775
3	22.28906167	70.77968985
4	22.28691738	70.77999026
5	22.28648058	70.78153521
6	22.28635036	70.78251657
7	22.28703269	70.78407489
8	22.28743179	70.7855219
9	22.28632461	70.78504884
10	22.28598988	70.78617584
11	22.28535904	70.78716371
12	22.28481832	70.78557756
13	22.28395573	70.78460361
14	22.28367249	70.78447839
15	22.28297727	70.78432534

Table 1: Coordinates of stops

2. 3 random points (C1, C2, C3) were selected to be set as centroids.

C1		**C2**		**C3**	
X	**Y**	**X**	**Y**	**X**	**Y**
22.28815022	70.78148298	22.28659521	70.78433731	22.28498171	70.78434266

Table 2: Iteration 1

3. Distances of each point from the centroids were measured using the euclidian distance formula : $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. Each point was assigned to the centroid closest to it.

[1] https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning

Stops	Co-ordinates		Distances			Cluster
	X	Y	C1	C2	C3	
1	22.28822779	70.7602492	0.021233922	0.024143369	0.02431115	C1
2	22.29037206	70.78024775	0.002542119	0.005566785	0.006769352	C1
3	22.28906167	70.77968985	0.002011484	0.0052614	0.006188274	C1
4	22.28691738	70.77999026	0.001936002	0.004358973	0.004763428	C1
5	22.28648058	70.78153521	0.001670457	0.002804441	0.003182513	C1
6	22.28635036	70.78251657	0.002075527	0.001837123	0.002282058	C2
7	22.28703269	70.78407489	0.002822568	0.00051015	0.002068381	C2
8	22.28743179	70.7855219	0.004102325	0.001450223	0.002719095	C2
9	22.28632461	70.78504884	0.004006025	0.000761254	0.001517253	C2
10	22.28598988	70.78617584	0.005166241	0.001935624	0.002092115	C2
11	22.28535904	70.78716371	0.006329405	0.003084908	0.002846167	C3
12	22.28481832	70.78557756	0.005278936	0.002166926	0.001245661	C3
13	22.28395573	70.78460361	0.005228008	0.002652876	0.00105865	C3
14	22.28367249	70.78447839	0.005387254	0.002926118	0.00131624	C3
15	22.28297727	70.78432534	0.005902408	0.003617958	0.002004522	C3

Table 3: k-means iteration - 1

4. The new centroids were calculated. The x co-ordinate of new centroid is the average of all the x co-ordinates of stops assigned to that particular centroid. The y co-ordinate of new centroid is the average of all the y co-ordinates of stops assigned to that particular centroid. This was done using 'Average' function in Microsoft Excel spread sheet.

C1		C2		C3	
X	Y	X	Y	X	Y
22.2882119	70.77634245	22.28662586	70.78466761	22.28415657	70.78522972

Table 4: Iteration 2

5. Now the distance of each stop from new centroids is calculated using the Euclidean distance formula. Each point is assigned to the centroid closest to it.

6. The new cluster is compared to the old cluster. If there is no change, the cluster is optimum. If there is a change, steps 4 and 5 needs to be repeated.

The calculations of all other iterations are given in appendix. The final solution provided by the k-means algorithms is shown in the table 5

Stops	Stops		Distances			Cluster
	X	Y	C1	C2	C3	
1	22.28822779	70.7602492	0	0.020550456	0.025142554	C1
2	22.29037206	70.78024775	0.020113177	0.002594227	0.007112953	C2
3	22.28906167	70.77968985	0.019458527	0.001650661	0.006695994	C2
4	22.28691738	70.77999026	0.019784502	0.001222177	0.005477639	C2
5	22.28648058	70.78153521	0.021357598	0.001544286	0.003872961	C2
6	22.28635036	70.78251657	0.02234638	0.002273538	0.002904877	C2
7	22.28703269	70.78407489	0.023855649	0.003376031	0.002089117	C3
8	22.28743179	70.7855219	0.025285239	0.004743268	0.002168419	C3
9	22.28632461	70.78504884	0.024872565	0.004513629	0.001053776	C3
10	22.28598988	70.78617584	0.026023051	0.005687986	0.001188738	C3
11	22.28535904	70.78716371	0.027066964	0.006832716	0.001946231	C3
12	22.28481832	70.78557756	0.025556807	0.005654459	0.000588304	C3
13	22.28395573	70.78460361	0.024726258	0.005436739	0.001464445	C3
14	22.28367249	70.78447839	0.024653688	0.005558663	0.001774091	C3
15	22.28297727	70.78432534	0.024642006	0.006005665	0.002474357	C3

Table 5: k-means iteration - 7