Apurba Pokharel 11627243

Requirement One

The optimal policy generated is:

17.9	20	16.1	13.38	11.94
15.81	16.81	0	10.46	10.38
13.85	14.25	11.11	9.66	9.11
12.07	Car	10.52	9.09	8.09
10.76	10.49	9.34	8.26	7.33

Requirement Two

The optimal policy generated is:

18.06	20.18	24.2	30	26.9
15.75	15.84	0	25.27	23.82
13.88	14.68	16.81	21.49	20.92
12.39	Car	15.99	18.46	18.28
11.29	12.66	14.26	15.94	16.35

Requirement Three

The optimal policy generated is:

17.9	20	22.4	30	26.9
15.81	16.81	0	23.62	23.52
13.93	14.68	15.77	20.17	20.46
12.29	Car	15.04	17.37	17.76
10.96	11.94	13.44	15.05	15.88

Requirement Four

a. Showing the optimal policy for 1000 iterations is done in the code output below. Kindly refer there.

b. The episodes ran for 1000 times.
Since a second request is only generated 60% of the time we have two requests for 600 episodes.
Out of these 600 episodes there is only a 30% of probability when a premium customer is generated i.e 180 times.

So,
There was a premium and a regular customer generated 179 times.
The fraction of selection of premium over regular was 105 / 179 , i.e 58% of the times.
(Best result I got, in the code pdf below I have 51% result)