Report - HCHC-DS

The clustering results are subject to fluctuations owing to the randomness in the initial choice of cluster centres.

The following Silhouette scores were found for k-means clustering:

```
k = 3, silhouette score = 0.6720396255846293
k = 4, silhouette score = 0.6614674946757273
k = 5, silhouette score = 0.6473525454270854
k = 6, silhouette score = 0.6265072985572774
```

Hence the optimal number of clusters is **3** which has the highest Silhouette score.

Some other statistics commonly found in the final clustering in several runs:

```
Cluster centers                                                 Cluster sizes
(52847.41453515896, 9259.463547241472, 8004.959898347217)       59418
(29816.412478651575, 9753.918246149158, 8517.38490805729)       76352
(17422.709037552417, 10552.820842645277, 9496.101458508969)     27295
```

The time taken to execute k-means was recorded as follows:

```
Total running time: 630.2164981365204 seconds
```

Due to a large number of data points (~160k), executing the O(n^2) divisive hierarchical clustering algorithm is impractically slow, hence several optimizations need to be applied in order to get results in a reasonable time.