Zuha Ahmad

Dr. Jaume

DS-670-01-17SPSG: Capstone: Big Data & Bus Analytics

29 January 2017

**Assignment 2 – Data Summary**

In the Capstone research course, we have to select a dataset from the CityPulse EU Data Collection. The categorical dataset in which I am tackling on is the social events dataset. The social event dataset derives from the municipality of Surrey in the United Kingdom. The timeline of the dataset is taken for two years from June 20, 2012 to July 30, 2014. The raw dataset is in the CSV which has 8,783 bytes file. However, the annotated dataset is distributed in the TTL format which has 23,658 bytes. The structure of the dataset is unstructured because it has to deal with data streams and web cast. Thus, the data is not heavily numerical. In the CSV file, there are no headers on the top of the fields. From the first field, it displays a list of governmental committees. The committees that are mentioned and noted are Planning and Regulatory Committee, Communities Select Committee, Cabinet, Council, Guildford Local Committee, Woking Joing Committee, Mole Valley Local Committee, Surrey Police and Crime Panel, and Planning & Regulatory Committee – Eco Park. In the second field, there is a list of websites that leads to Surrey’s Public TV through a webcast portal. There are list of 28 URL sites that are displayed out of a list of 30. The third field shows the committees, date and time, and the location of the broadcasts. The majority of the webcast takes place in the Ashcombe Suite County Hall. In the fourth column, it shows the same podcast URL’s as the second column. Finally, in the fifth column, it shows clearly the time and date similar to the third column. Additional information is also provided from the Surrey County Website. It has a Webcast library where there is a collection of video regarding committee meetings. Overall, this data provides information and links regarding video podcast.