Asuncion Los Banos

Assignment 2

DS-670

The cultural event dataset from CityPulse was provided by the city of Aarhus in Denmark. It is a dataset to explore cultural event announcements from May 5th, 2015 to January 25th, 2015. Consumers were able to purchase their ticket through a website called Billetlugen. Billetlugen is like Ticketmaster where they act as an agent and sells the tickets that the client makes available to them. All cultural events were held at the Musikhuset Aarhus (Aarhus Concert Hall) the largest concert hall in Scandinavia. Musikhuset Aarhus is located in the city center and it is Aarhus’ main venue for music that seats more than 3,600 people in total.

The dataset characteristic is multivariate. There is 100 number of instances and 19 number of attributes. Upon a quick glance of the dataset’s CSV there is no header information to communicate what the attributes are, therefore some deduction had to be done based off the information from the CSV and Turtle File.

**Attribute information for Cultural Dataset:**

**1 – status (numeric: 0, 1, 2, 4)** The first attribute was able to be derived from the Library events in CityPulse. The Library events were also from Aarhus and in the dataset contained header information in the CSV file. Based off the information available the first attribute in the Cultural Events Dataset and the status attribute in the Library Events seemed to be similar.

**2 – location (nominal: Aarhus C)** The second attribute were all Aarhus C. Aarhus C is a postal district in the city of Aarhus, Denmark, consisting of the Inner city, Vesterbro, University of Aarhus, Frederiksbjerg, Langenæs and Aarhus Ø, with postal code 8000.

**3 – title (nominal: multiple different types of titles)** The third attribute is the titled of the cultural events. Some of the titles are SPOT Festival 2014, KAMMERKONCERT, 8000 Comedy, Better than you think - Anders Haglund and Sofie Spanget Rasmussen - prÃ¦debut, cello.

**4 – urlBilletlugen (nominal: multiple different urls)** The fourth attribute is the url of the link to Billetlugen. However, since these cultural events passed already the urls no longer lead to the event page to purchase tickets.

**5 – price (numeric: prices varied depending on the event)** The fifth attribute is the price of the cultural events. Many of the events are Gratis (free), but some have ranging prices depending on the seat you pick at the venue. Such as the event Jenufa, the lowest ticket was set at 335.00 DKK and the highest ticket was 675 DKK.

**6 – unixtime (numeric)** The sixth attribute is really unknown, however due to past experience with datasets it seems this attribute is a unix timestamp. The unix timestamp differs however the time is always 7:00 AM.

**7 – zipcode (numeric)** The seventh attribute is the postal code 8000 of Aarhus C. Every single instance in the dataset has the same zip code information.

**8 – longitude (numeric)**  The eight attribute is the longitude, 10.19887, of Musikhuset Aarhus. Every single instance in the dataset has the same longitude information.

**9 – urlBilletlugenExtension (numeric)**  The ninth attribute is the ending extension of the url to the Billetlugen website. An example is if the urlBilletlugen is: “http://www.billetlugen.dk/referer/?r=266abe1b7fab4562a5c2531d0ae62171&p=/koeb/billetter/33506/47143” then the urlBilletlugenExtension is 47143.

**10 – teaser (nominal)** The tenth attribute is a short teaser of what the cultural event is about. It has information as to who is playing in the Koncert or who is performing in the Teater.

**11 – street (nominal)** The eleventh attribute is the street location, Thomas Jensens Allé, of Musikhuset Aarhus.

**12 – MusikhusetAarhusHall (nominal)** The twelfth attribute is the hall where the cultural event will take place in the Musikhuset Aarhus. Musikhuset Aarhus has 9 stages and 6 halls. The halls are: Symfonisk Sal, Rytmisk Sal, Lille Sal, Kammermusiksalen, Foyer, Filuren, The Café stage, Falken, Lokale 222, Amfiscenen, and Balkongalleriet.

**13 – eventdate (nominal)** The thirteenth attribute is the date the cultural event took place.

**14 - latitude (numeric)** The fourteenth attribute is the latitude, 56.1519158, of Musikhuset Aarhus. Every single instance in the dataset has the same latitude information.

**15 – calendar (nominal)** The fifteenth attribute is the url of the cultural event’s in the Musikhuset Aarhus calendar. The calendar was probably to show what event was going on certain days/months. All of the url leads to a “Page Does Not Exist” page in the Musikhuset Aarhus website. One of the calendar url is: http://www.musikhusetaarhus.dk/kalender/34821/

**16 – totalofattendees (numeric)** The sixteenth attribute, is a surmise, and it is the total number of attendees that attended the cultural event. This column of data following the cultural event can be seen as the number of attendees since the range varies widely within each event. The highest attendance was for an event titled “Hist & Her – Sanseudstilling” and it had 985 attendees. It’s eventcategory was Andet and the eventtype was Familie/bÃ¸børn,Gratis. The event with the lowest attendance was titled “Sanne Salomonsen Hjem 2014” with 230 people. It’s eventcategory was Musik and the eventtype was Rock, Pop.

**17 – eventcategory (nominal)** The seventeenth attribute is the category of cultural event whether it is Musik, Andet, or Teater.

**18 – imageurl (nominal)** The eighteenth attribute is the cultural event’s image url. An example of an image url is: <http://static.billetlugen.dk/images/events/b/29048.jpg>. The image is hosted in the Billetlugen website. To date the url of the images are still working.

**19 – eventtype (nominal)** The nineteenth attribute is the type of cultural event such as Klassisk, Gratis, Rock,Pop, Folk, Standup/Comedy, Klassisk,Gratis, Koncert,Gratis, Musical,Opera/Operette,Gratis, Revy or Rock.

Although the attribute names were not given exactly for the cultural events dataset through some deduction and research one can determine what the data is trying to represent and reflect and it is enough to come up with a good analysis especially with attribute numbers 3, 5, 13, 16, 17, and 19. The attributes were able to be concluded because as an example for the seventeenth attribute, eventcategory, given the data in this column the data represents a type of gathering due to the various cultural data events. Also, for the nineteenth attribute, eventype, according to the data represented, it is determined that there are different event types.