**Cultural Events**

**Introduction:**

Diverse, forefront and fun, Aarhus is a champion among a delighted urban zones on Earth. It moreover happens to be the European Capital of Culture 2017. 13% of Aarhus' people are understudies, making Aarhus the most energetic city in Denmark. For the most part, in any case, it's a standout amongst the most settled. 315,000 people live in Aarhus and 1.2 million people live in the more significant Aarhus area, so it's the perfect size for an end of the week break or family getaway. Being young on an essential level infers that the city can offer you an extensive variety of social experiences, from nightlife, social events and shopping to the best restaurants in the region. Aarhus is moreover home to predominant festivals, for instance, food festival, spot festival, and NorthSide Music Festival.

The DataSet for “Cultural Events” shows the set of cultural event that was announced and is collected as the stream from one of the municipalities, known as Aarhus. This proposal looks at how the cultural data collected can be used to interpret the impact of the cultural events happening in Aarhus and its influence on the city. This would be critical in making Aarhus a smart city.

**Data description/structure:**

The data provided is a set of cultural event announcements provided as a datastream from the municipality of Aarhus for a duration from May 5th, 2014 - January 25th, 2015. The dataset is available in both as “CSV” and “TTL” format to interpret. The data seems to have to have 19 attributes describing the cultural events in that municipality, as described later in the document later on. The cultural events are described by the type of concert, postal code, pricing, etc. All the data collected is in the Danish language so we would need to convert in English to read, interpret and analyze it better.

**Data processing:**

CityPulse has provided the online repository of cultural data set in both “.csv” and “.ttl” format which can be used for data analysis. For this project, I’m going to use the “.csv” format only as it’s easy to manipulate and analyze the data. The data analysis will be done in Apache Zeppelin environment that is being taught in the class.

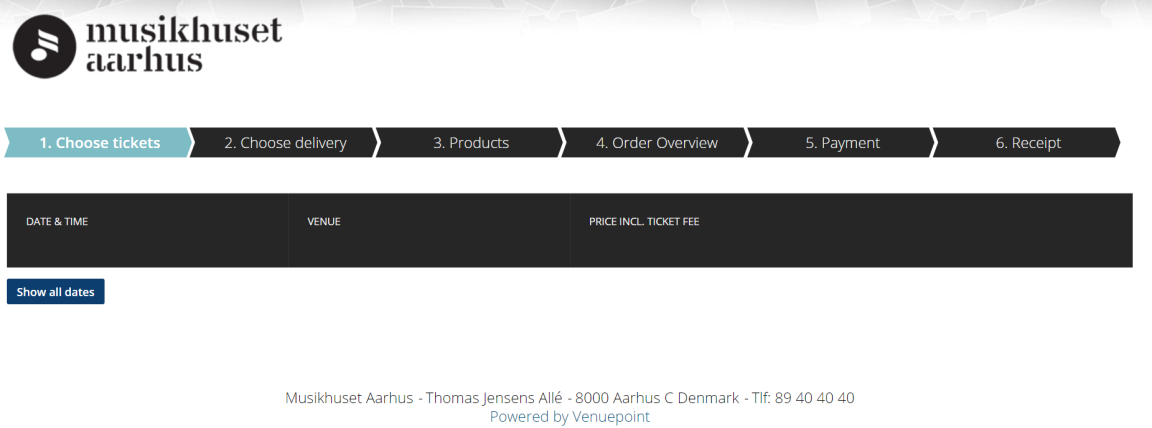
**Variables:**

1st variable: The first column is by all accounts a binary code 0 or 1, which implies the appropriate response is a "yes" or "no". The data is in integer format.

2nd variable: The second column seems to have the attribute “Aarhus C” which is a postal district in the city of Aarhus in Denmark that means all the attributes in that column represent that the cultural events took place in the same postal code of Aarhus C. The data is in string format.

3rd variable: This column seems to tell us about the type of concert. For example, “Chamber Concert”, or “the History of Jonny Cash,” or “8000 comedy”, etc which would be the data collected for that particular cultural event. The data is in string format.

4th variable: This column seems to be the “Pricing” or “Shopping Cart”, as shown below, to show the number of tickets being purchased by an individual user. The data is in string format.



5th Variable: This column has the price of the concert, represented in “Danish Krone”. In some instances, the pricing seems to be free or through a ticket. The data is in integer format.

6th Variable: This column seems to represent the unique “transaction id” for the concert. This must be to keep a track of each concern and the relative purchase. The data is in integer format.

7th Variable: This column seems to have the maximum capacity of the concern hall. The data is in double format.

8th Variable: This column seems to represent the “Longitude” of the concern hall to pin-point the exact location on the map. The data is in integer format.

9th Variable: This could be the number of people in the city at the time of the concert. The data is in string format.

10th Variable: This column has the details of the concert like the date, time, etc. and the price of different individuals like “adults” or “students”. The data is in string format.

11th Variable: This column has the same information “Thomas Jensens AllÃ©” which seems to the copyright of that individual. The data is in string format.

12th Variable: This column represents the hall in which the concert happened. For example, “symphonic Hall” or “Great Hall”, etc. The data is in string format.

13th Variable: This column has the specific date and time of the different concerts. The data is in string format.

14th Variable: This column contains the Latitude of the concert. The data is in string format.

15th Variable: This column contains the calendar dates of the cultural events from the website. The data is in string format.

16th Variable: This column could possibly represent the number of people attending the cultural concert. The data is in string format.

17th Variable: This column has two categories, representing the kind of concert, either “Music” or “Other”. The data is in string format.

18th Variable: This column seems to contain the “.jpg” image of the concert. The data is in string format.

19th Variable: This column seems to tell the category of the music event like “Classic”, “Pop”, “Musical”, “Opera”, etc. The data is in string format.

**Interpretation:**

We could interpret the correlation between the cultural events and the impact on other external events (like traffic or parking, etc.) to figure out how to avoid congestion. As the data is about the smart city, this data could be very useful in interpreting and making smart choices for the city. We could also determine which cultural event has the biggest impact on the cities and the time it impacts the most, which would be quite useful for the city planning. The cultural events are possibly described by the time, location, performers, the category of the concert, etc, so all the data would need to be looked at closely and see which variable has the biggest impact on the correlation.

**Results:**

To be determined by the end of the class.