**Annual festival donation and ticket sales analysis**

**EMIS 7357 Final Project**

The festival organizers would like to understand the following:

1. **Preliminary analysis (3 points)**
   1. Donors (1 points)
      1. How many total donors (people who contributed a positive amount any year) are there? What about donors who contributed at least $100? $500? $1000?
      2. How many donors were there each year in the data set? What about donors who contributed at least $100? $500? $1000?
      3. Plot the histograms of donations for the entire time period, for 2014 and for 2015. Comment.
      4. Plot a scatterplot with the donations for 2014 and 2015 (one dot = one row, removing [for this question only] instances with no donation either year). Comment.
   2. Ticketholders (1 point)
      1. How many total ticketholders are there? (note that some donors may never have bought tickets)
      2. How many ticketholders were there each year in the data set?
      3. Plot the histograms of ticket revenues for the entire time period, for 2014 and for 2015. Comment.
      4. Plot a scatterplot with the ticket purchases for 2014 and 2015 (one dot = one row, removing [for this question only] instances with no ticket purchase either year). Comment.
   3. Joint analysis (1 point)
      1. Plot total donations and total ticket revenue for each patron on a graph (one dot = one record or row of the data set).
      2. Comment on the general trend in your graph and the outliers.
2. **Donation patterns over time (8 points)**
   1. (3 points) Cluster the donors according to their patterns of donation over time up to 2014.
      1. Justify your number of clusters, the method used, and provide a description of each cluster suitable for a manager.
      2. Plot the donation pattern over time for each cluster.
   2. (3 points) Create a model to predict donations or donation groups for 2014 using the previous five years of data. You can use the method of your choice (linear regression, logistic regression, classification trees…)
   3. (1 point) Test your model on 2015 data. Comment.
   4. (1 point) Make predictions for donations in 2016.
3. **Ticket sales patterns over time (8 points)**
   1. (3 points) Cluster the ticketholders according to their patterns of donation over time up to 2014. You can use hierarchical clustering or k-means.
      1. Justify your number of clusters, the method used, and provide a description of each cluster suitable for a manager.
      2. Plot the ticket sales pattern over time for each cluster.
   2. (3 points) Create a model to predict ticket revenues or ticket revenue groups for 2014 using the previous five years of data. You can use the model of your choice (linear regression, logistic regression, classification trees…)
   3. (1 point) Test your model on 2015 data. Comment.
   4. (1 point) Make predictions for donations in 2016.
4. **Joint analysis (3 points)**
   1. (2 points) Cluster instances in the data set across both ticket revenue and donations for every year.
      1. Justify your number of clusters, the method used, and provide a description of each cluster suitable for a manager.
      2. Provide a visualization of the clusters.
   2. (1 point) Adjust your method so that the most recent donations or ticket purchases count more than those further in the past.
5. **Conclusions (3 points)**

Develop a strategy for the marketing team.

* 1. (1 point) Who are donors the festival organizers should cultivate for 2016?
     1. Donors who did not donate as much as expected in 2015,
     2. Donors who donate a lot and whom the festival can’t afford to lose.
  2. (1 point) Who are ticketholders the festival organizers should cultivate for 2016?
     1. Ticketholders who underperformed their 2015 estimate,
     2. Ticketholders who play a key role in overall revenue year after year.
  3. (1 point) From the aggregate viewpoint of total revenue (ticket purchases + donations), who are the patrons the festival organizers should cultivate for 2016?

**Report quality: 2 points**

Note: Some large ticket sales are group sales, especially from school districts. As usual, you may want to remove such outliers or treat them separately. The festival has children programming so some patrons or donors may contribute only as long as their children go to the shows. The average patron base is also rather old, so some patrons may stop buying tickets because they passed away.

The report must be suitable to be shared with (and understood by) a company. The report must look professional, which includes:

* a cover page with SMU logo, project title, names of students,
* all pages numbered after cover page
* executive summary
* table of contents
* list of figures
* list of tables
* each question addressed in its specific section, and the answers must be well-written
* conclusions
* any source file should be included separately in the zip file. Do not copy and paste code in the main body of your report.