Challenge #1 - Improving Employee Retention

Top of Form

|  |  |
| --- | --- |
| Times  The challenge will occur **July 31** through **August 31**.  Dataset  [https://oursites.myngc.com/ENT/analytics/Shared%20Documents/Analytics%20Competitions/Challenge1/analytics-competition-poster-preview.png](https://oursites.myngc.com/ENT/analytics/Shared%20Documents/Analytics%20Competitions/Challenge1/analytics-competition-poster.pdf)  The dataset to be used is 1,500 rows and 35 columns of completely fictitious employee data that was obtained, with permission, from an external source. The dataset in no way, shape, or form represents Northrop Grumman or any other company. The dataset was created using simulation software to identify interesting fields that may be used to predict attrition. As with any data science project, both artistic and scientific skills will be exercised to determine the best model. Therefore, the identification of key features that predict attrition will be required to develop the winning model.  Instructions  When the challenge commences, a link to download a zip file will be sent to each participant containing a training dataset (80% of data) and a testing set (20% of data) with 34 columns (e.g., minus the predictor column). Participants will use whatever language, tool, or platform they desire to predict attrition.  Submissions  Upon completion of a model, participants will submit a CSV file back that contains two fields: EmployeeNumber and Attrition (the attrition prediction). The area under the ROC curve of predicted versus actuals will be used to score the best model. Participation may occur individually or by team. While Northrop Grumman will not use any code, solution, or model that was developed, an opportunity to discuss and share the techniques with the analytics community will be available and possibly requested.  Scoring  The Enterprise Analytics team will score the results once per day at between 8 am and 10 am. Only one file per participant will be scored per day. Once scored, you will be sent an automated email with your last scored result and your highest scored result.  Please visit the [Challenge FAQ](https://oursites.myngc.com/ENT/analytics/SitePages/Challenge%20-%20FAQ.aspx) page to view frequent scoring errors.  How to Begin  Get started by clicking the button below. You'll be asked to agree to the terms and conditions and we'll be collecting your contact information so we can send further instructions and information about the challenge.  [Sign-up for the Challenge](https://oursites.myngc.com/ENT/analytics/Lists/Participants/Item/newifs.aspx?Source=https%3a//oursites.myngc.com/ENT/analytics/)  Challenge Tips and Help  This site will be used for collaboration, information sharing, and learning about the many enterprise and sector analytics activities. Additionally, an ability to opt-in for email notifications, blogs, threaded discussions, and lunch-and-learns will be offered. This site can be used to collaborate on techniques, tools, and ideas used in the [NG Analytics Challenge Series](https://oursites.myngc.com/ENT/analytics/SitePages/Challenges.aspx). While any tool or technology may be used, a few can be found in the [Enterprise Software Store](http://softwarestore.northgrum.com/esd), such as R, RStudio, Python, and many others.  For issues with tools, feel free to contact [analytics@ngc.com](mailto:analytics@ngc.com). |  |

false,false,2

Bottom of Form