R Programming for Business

Group Project

The purpose of this group project is to put to work the tools and knowledge that you gain throughout this course. This provides you with multiple benefits.

1. It will provide you with more experience using data wrangling tools on real life data sets.
2. It helps you become a self-directed learner. A large part of your job is to self-direct your learning and interests to find unique and creative ways to find insights in data.
3. It starts to build your analytics portfolio. This is a great way to show potential employers your ability to work with data.

The principal goal of this project is to import a real-life data set, clean the data, and perform basic exploratory data analysis.

You will need to find one data set. You may even need to learn new skills not taught to accomplish your mission. These include working with:

- Multiple data types (numerics, characters, dates, etc) ;
- Non-normalized characteristics (may contain punctuations, upper and lowercase letters, etc);
- Datasets that need to be merged;
- Unclean data (missing values, values that do not align to the data dictionary);
- Variables that need to be created (i.e. the data may contain income and expense variables but you want to analyze savings such that you need to create a savings variable out of the income and expense variables);
- Data that needs to be filtered out.

You will write a pdf report (Not more than 15 pages). You will need to import, assess, clean the data, and then come up with your own research questions that you would like to answer from the data by performing exploratory data analysis (if you'd like to perform a predictive model to answer your hypothesis that is fine but it is not required). Some thoughts to help you:

- Your project should be a logical, cohesive story–not simply a bunch of graphs created for the sake of making them. The story may change as you dive deeper into the data and find insights, but a storyboard gives you direction and purpose for developing insights. Clear writing means a clear mind, and a storyboard is vital to producing a good story;
- Speaking of insights, keep in mind that your project should follow the chain of data -> insights -> actions;

- Simple descriptive statistics can (and usually) yield more of an immediate impact than a complicated model;
- Do subgroups matter in your data?
- Why are data missing?
- Are trends over time important?

Try to be creative in your analysis. Creativity is an essential ingredient for a good data scientist!

Possible sources of datasets:

- Google Dataset Search: https://datasetsearch.research.google.com/
- Kaggle data science platform: https://www.kaggle.com/
- French OpenData initiative: https://www.data.gouv.fr/fr/
- World Bank Open Data - global development data: https://data.worldbank.org/
- EU Open Data portal : https://data.europa.eu/euodp/data/
- UC Irvine : https://archive.ics.uci.edu/