**Project Proposal Frederic TWAHIRWA**

**Foundations data science February 2017**

**Customer default payment**

The ability to predict what is likely to happen in the future is probably one of the most important issues for a manager. Strategically, being able to anticipate the results of its actions allows a business manager to identify the efficient strategies to adapt.

Financial crimes along with default payment are increasing at a high rate every year and financial institutions have to adopt processes to safeguard their reputations as well as their customers, especially after the global crisis on 2008.

Credit risk is the potential loss a financial service will suffer in a borrower fail to meet its obligation. The borrower’s repayment ability is probably the first criteria in risk management. Since, for years, a number of approach and methods have been developed and implemented to reduce the risk default.

The aim of this project is to identify models and tools for credit risk assessment and apply them trough real data sets. This project will focus on:

* Predictive data analytics tools;
* How data analytics tools work;
* The business issues that can be solved leveraging data;
* Interpret results with a business mindset.

The dataset is about customers default payment in Taiwan and provided by UCI Machine Learning Repository: (http://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients)

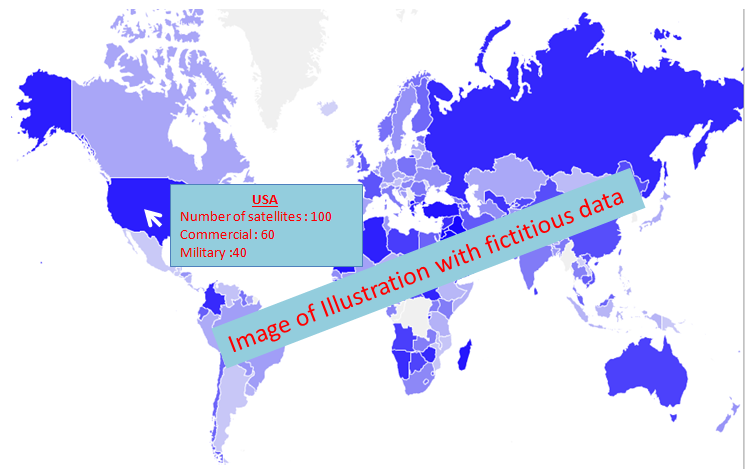
In addition, this project will provide a case to assess comparative effectiveness for predicting the probability accuracy of default among different techniques regarding particular data set or attributes set.

**Active satellites in Orbit around Earth**

If the data scientist wants that the many hours invested in the analysis to be fruitful, its conclusions must be understood and accepted.

This project aims to classify and identify the observations. Here, for example, data analysis may identify which country has the most satellites in orbit, or what are they used for.

This project will also allow reporting results and findings in a convincing and impactful way by using advanced tools such as topographical graphs. We will analyse data and try different visualisations to tell the story most accurately. For example set up an interactive map with different colours according to the density.



**Figure 1: Example of presentation**

The Satellite Database is a listing of active satellites currently in orbit around the Earth. The database includes basic information about the satellites and their orbits but does not contain the detailed information necessary to locate individual satellites

This project will focus on:

* Classification data analytics tools;
* interpret results with a business mindset;
* how to report and communicate findings efficiently regardless of the analytics tools used

Dataset is from is from kaggle : https://www.kaggle.com/ucsusa/active-satellites

# Student alcohol consumption

# A manager has to be able to anticipate what will be the consequence of its actions, in order to maximise its effectiveness and efficiency. Then, one of the main reasons to use business analytics is to understand the relationship between causes and consequences.

The student alcohol consumption project aims to highlight what makes a student more or less likely to consume alcohol. Here we can investigate what’s driving student to alcohol if this consumption impacts negatively theirs results or not as well as correlation between student behaviour

Since this project allows using statistical methods to understand relationships between different events which are crucial in business intelligence.

This project will focus on:

* Predictive data analytics tools;
* Understand the relation between causes and consequences;
* The business issues that can be solved leveraging data;
* Interpret results with a business mindset.

Dataset is from UCI Machine Learning Repository: <http://archive.ics.uci.edu/ml/machine-learning-databases/00356/>