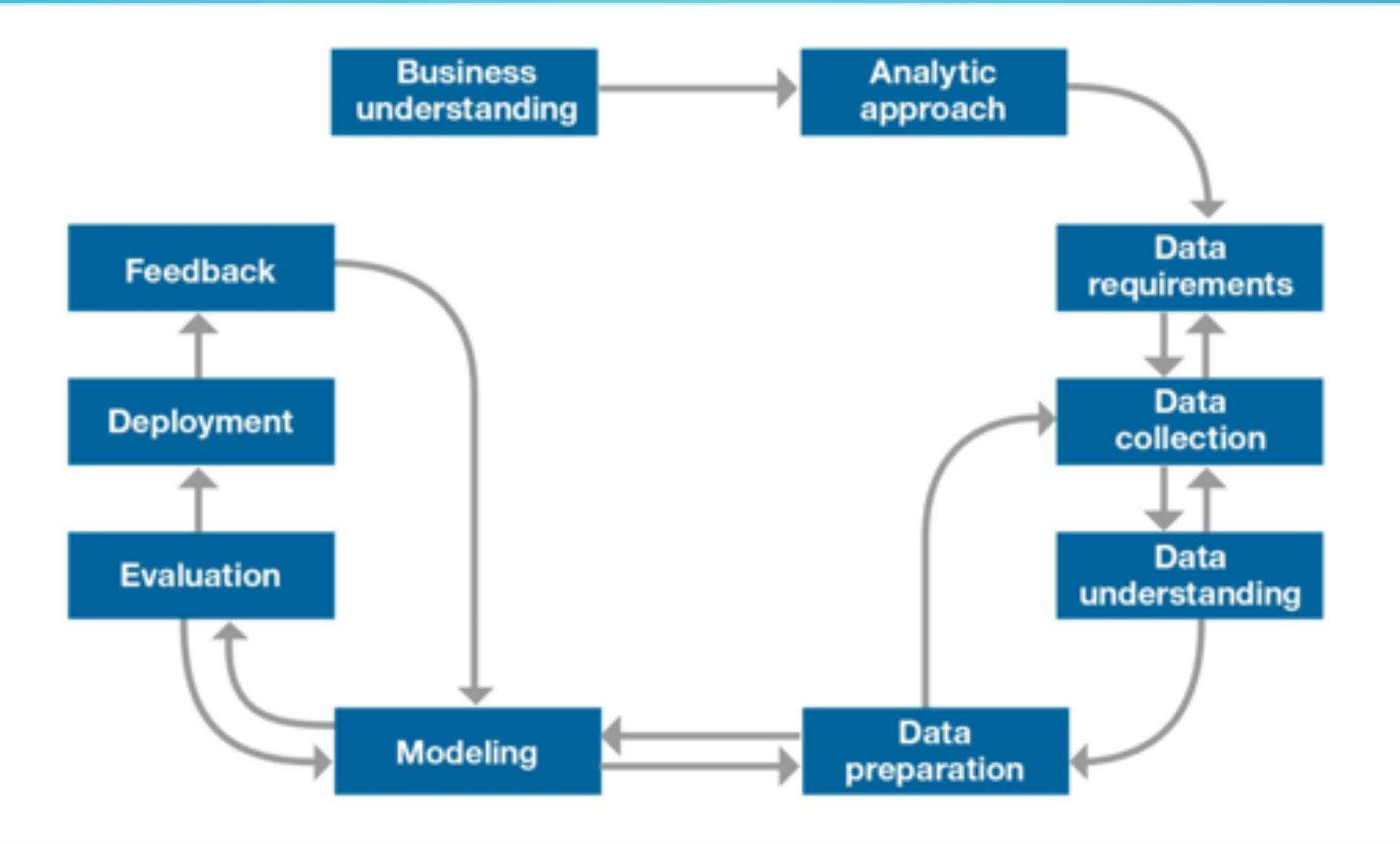


METHODOLOGY

COURSERA FINAL EXAM

STRUCTURES MY PROJECT

METHODOLOGY



BUSINESS UNDERSTANDING

THIS IS FIRST STEP FOR ANY DATA SCIENCE METHODOLOGY,
AND ANSWER THE QUESTION: **WHAT IS PROBLEM YOU TRYING TO SOLVE?**



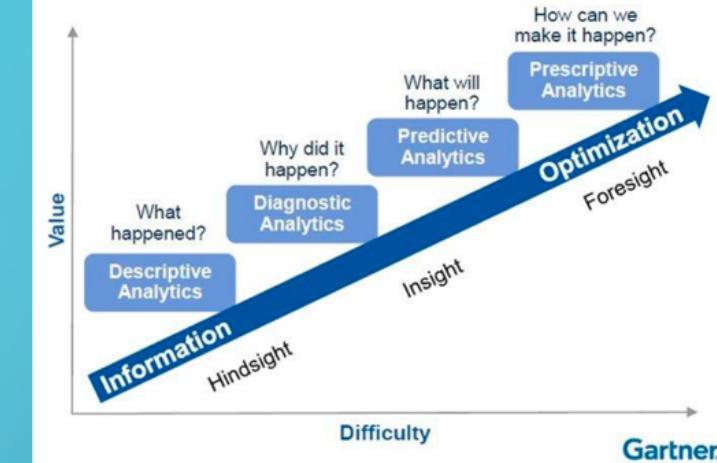
THEORY: “Every project, whatever its size, begins with the understanding of the business that forms the basis of an effective solution to the business problem. Business partners who need the analytics solution play a critical role in this phase by defining the problem, the project objectives, and the solution requirements from a business perspective.”

MY BUSINESS PROBLEM:

HOW MUCH INCREASE OF BANKS PROFIT, IF YOU GIVE CREDIT CARDS TO CONSUMERS
WITH AN AGE BETWEEN 12 AND 16 YEARS?

ANALYTIC APPROACH

THIS IS SECOND STEP FOR ANY DATA SCIENCE METHODOLOGY,
AND ANSWER THE QUESTION: **HOW CAN YOU USE THE DATA TO ANSWER THE QUESTION?**



THEORY: “Once a business problem has been clearly identified, the Data Scientist can define the analytical approach. To do this, the problem must be expressed in the context of statistical learning and machine learning techniques so that the Data Scientist can identify the techniques to achieve the desired result.”

MY ANALYTIC APPROCH IS :

PREDICTIVE

DATA REQUIREMENTS

THIS IS THIRD STEP FOR ANY DATA SCIENCE METHODOLOGY,
AND ANSWER THE QUESTION: **WHAT DATA DO YOU NEED TO ANSWER THE QUESTION?**



THEORY: “Analytic approach determines the data requirements because the methods of analysis to be used require specific content, formats, and data representations, based on domain knowledge.”

MY DATA REQUIREMENTS ARE:

I NEED SOCIAL ECONOMIC DATA, WITH POPULATION DIVIDED FOR AGE AND INCOME

MAYBE STUDY TITLE

DATA COLLECTION

THIS IS fourth STEP FOR ANY DATA SCIENCE METHODOLOGY,
AND ANSWER THE QUESTION: **WHERE IS THE DATA COMING FROM (IDENTIFY ALL
SOURCES) AND HOW WILL YOU GET IT?**



MY DATA COLLECTION COME FROM:

WIKIPEDIA

DATA UNDERSTANDING

THIS IS FIFTH STEP FOR ANY DATA SCIENCE METHODOLOGY,
AND ANSWER THE QUESTION: **IS THE DATA THAT YOU COLLECTED REPRESENTATIVE
OF THE PROBLEM TO BE SOLVED?**



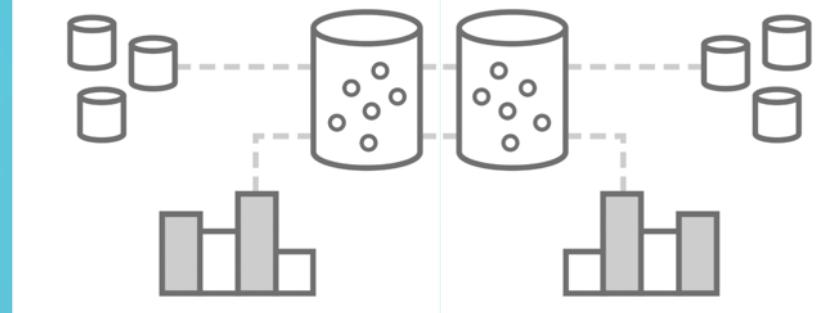
THEORY: Descriptive statistics and visualization techniques can help a data scientist understand the content of the data, assess its quality, and obtain initial information about the data. A recovery from the previous step, data collection, may be necessary to fill the gaps in understanding.

MY DATA UNDERSTANDING:

I NEED AN INTERVIEW WITH A BANKER AND PARENTS THAT HAVE CHILDREN IN THAT RANGE
OF AGE

DATA PREPARATION

THIS IS SIXTH STEP FOR ANY DATA SCIENCE METHODOLOGY,
AND ANSWER THE QUESTION: **WHAT ADDITIONAL WORK IS REQUIRED TO MANIPULATE
AND WORK WITH THE DATA?**



THEORY: "The Data preparation step includes all the activities used to create the data set used during the modeling phase. This includes cleansing data, combining data from multiple sources, and transforming data into more useful variables. In addition, feature engineering and text analysis can be used to derive new structured variables to enrich all predictors and improve model accuracy. The Data preparation phase is the longest. Although I have seen that it represents 90% of the total duration of the project, this figure is usually 70%. However, it can go down as much as 50% if the data resources are well managed, well integrated, and analytically clean, not just storage. Automating some phases of Data preparation can further reduce the percentage:

MY DATA PREPARATION:

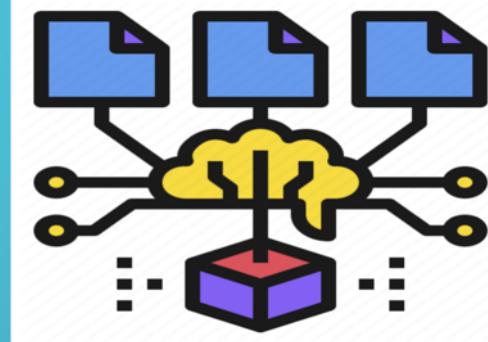
THIS IS A WORK FOR ALL THE TEAM

TO QUANTIFY THE WORK THE TEAM CAN CREATE A PIVOT TABLE WITH TYPE OF FORMATION AND
TYPE OF DATA

MODEL TRAINING

THIS IS SEVENTH STEP FOR ANY DATA SCIENCE METHODOLOGY,
AND ANSWER THE QUESTION:

IN WHAT WAY CAN THE DATA BE VISUALIZED TO GET THE ANSWER THAT IS REQUIRED?



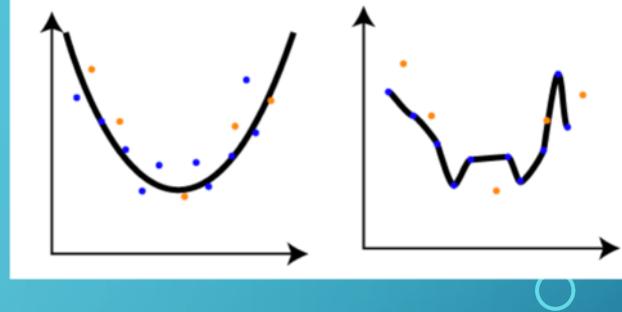
THEORY: "From the first version of the prepared data set, Data scientists use a Training data set(historical data in which the desired result is known) to develop predictive or descriptive models using the described analytical approach previously. The modeling process is very iterative. It may be vary with different situation as per problem."

MY MODEL TRAINING:

- GRAFICI A TORTA
- ISTOGRAMMI
- PIVOT DI TABELLE

MODEL EVALUATION

THIS IS EIGHTH STEP FOR ANY DATA SCIENCE METHODOLOGY,
AND ANSWER THE QUESTION: **DOES THE MODEL USED REALLY ANSWER THE INITIAL
QUESTION OR DOES IT NEED TO BE ADJUSTED?**



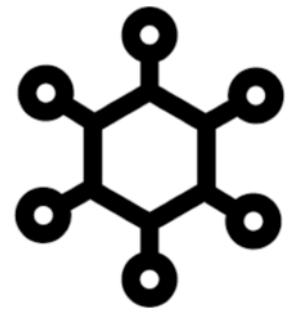
THEORY: “The Data Scientist evaluates the quality of the model and verifies that the business problem is handled in a complete and adequate manner. To do this, several diagnostic measures and other results, such as tables and graphs, must be calculated using a set of predictive model tests.”

MY MODEL EVALUATION:

I CAN TAKE A SAMPLE POPULATION TO BEGIN TO VISUALIZE THE DATA AND SUBMIT THE
RESULTS TO THE CUSTOMER

DEPLOYMENT

THIS IS NINTH STEP FOR ANY DATA SCIENCE METHODOLOGY,
AND ANSWER THE QUESTION: **CAN YOU PUT THE MODEL INTO PRACTICE?**



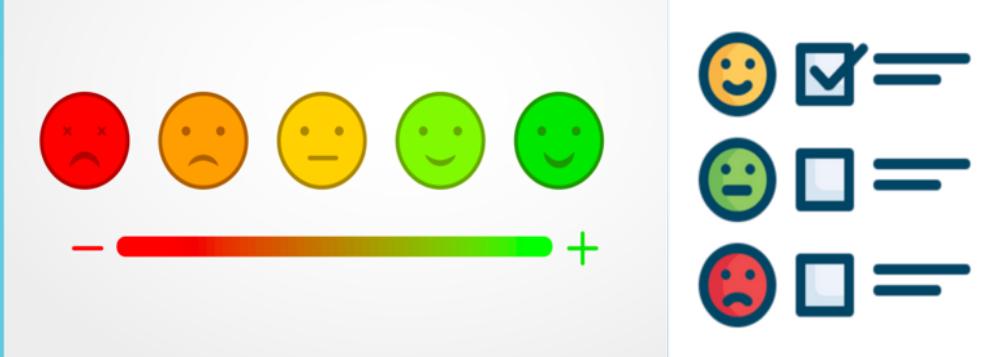
THEORY: “Once a satisfactory model has been developed and approved by commercial sponsors, it will be implemented in the production environment or in a comparable test environment. Such deployment is often initially limited to allow for performance evaluation. Implementing a model in an operational business process generally involves multiple groups, capabilities, and technologies.”

MY DEPLOYMENT:

I THINK I NEED 6 MONTHS OF WORKING AND A TEAM OF 15 PEOPLE AND MONEY TO
DEVELOP THIS MODEL!

FEEDBACK

THIS IS LAST STEP FOR ANY DATA SCIENCE METHODOLOGY,
AND ANSWER THE QUESTION: **CAN YOU GET CONSTRUCTIVE FEEDBACK INTO
ANSWERING THE QUESTION?**



THEORY: “By collecting the results of the implemented model, the organization receives feedback on the performance of the model and its impact on the implementation environment. By analyzing this information, the data scientist can refine the model, increasing its accuracy and, therefore, its utility.

This phase, often neglected, can have significant additional benefits when carried out as part of the overall process. The flow of this methodology illustrates the iterative nature of the problem-solving process.”

MY FEEDBACK:

CAN YOU GET CONSTRUCTIVE FEEDBACK INTO ANSWERING THE QUESTION?

YES

THANKS FOR READING...!!!

References :

- 1.<https://www.coursera.org/learn/data-science-methodology>
- 2.element61.be/en/competence/data-science-methodology