Sergio Cordoba (MSc. Physics, MSc. Mathematical modeling)
https://www.linkedin.com/in/sergiocordoba

DATA | ANALYTICS | AI

data scientist | **+10 years experience** in differents sectors | **oriented to business** processes | **task & timings** management | **agile** methodologies
end-to-end data (science) pipelines design, building and operationalization | **monitoring** business-KPIs and model performance

1

Projects about ML/AI, usually, born from an **idea around some business needed or improvement. Advising on the viability of the business goal, evaluating available resources**, linking with data and systems areas, **to propose an approach is my first goal.**

4

I perform **Exploratory Data Analysis (EDA)** to get data knowledge and business insights

- **Feature descriptives** give us knowledge about data at business (understanding, interpreting) and technical level (**modeling features, missings, datatypes, values, univariate analysis, outliers, skewness, scaling or transformation, etc**).
- **variables interaction** (specifically, with target in supervised models), correlations, PCA
- **statistical testing** for population **samples comparison (A-B tests)**, hypothesis checks

In according with each project or business case, one or the other is suitable

business understanding → data understanding → data preparation → modeling → evaluation → deployment

DATA SCIENCE WORKFLOW

2

From “my side” key points are:
- Define **target variables and features**, and their availability.
- Data sources and volume
- Systems and **cloud platform (my core have been AWS but i have worked in projects with GCP and Azure)**

3

To build the data inputs i’m **proficiency on SQL.** I have worked with several “dialects” (postgres, HiveQL, sparkSQL), storages, fileformats, and technologies,

- **core in AWSRedshift, AWSS3 (+7 years)**
- respect to **BigData technologies** i have worked with Hive, Hue, hdfs, parquet, spark,
- **NoSQL** storage and querying over ELK, AzureSearch, Solr
- **Input text** processing, parsing, chunking, embedding for GenAI projects.
- Also, i have worked with **graph-databases**; OWL, RDF, SparQL

** project steps based on CRISP-DM work-flow

Sergio Cordoba (MSc. Physics, MSc. Mathematical modeling)
https://www.linkedin.com/in/sergiocordoba

DATA | ANALYTICS | AI

data scientist | **+10 years experience** in differents sectors | **oriented to business** processes | **task & timings** management | **agile** methodologies
end-to-end data (science) pipelines design, building and operationalization | **monitoring** business-KPIs and model performance

---

To model or data analysis, i use python (last 6 years) and R (+4 years).

(NOTE:...i like much more R capabilities for data handling but python to deploy)

Usually, I use pandas, numpy, sklearn. But, i have used too many packages in according with the project goal and neede. Also, some knowledge in tensorflow and keras

Regarding R, my preferences were datatable VS dply and caret (pycaret exists) was a good model package (among others)

business understanding → data understanding → data preparation → modeling → evaluation → deployment

DATA SCIENCE WORKFLOW

A bit about business cases;
- CLTV
- BL

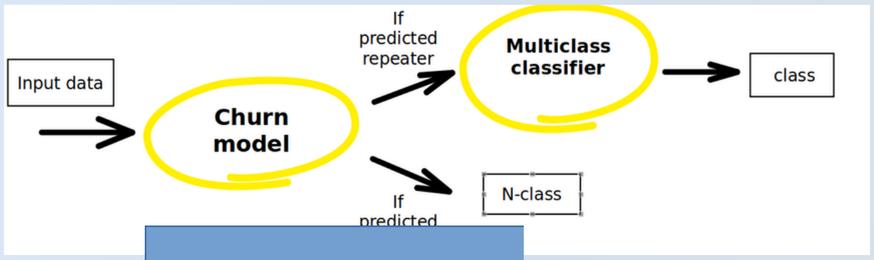
** project steps based on CRISP-DM work-flow

Sergio Cordoba (MSc. Physics, MSc. Mathematical modeling)
https://www.linkedin.com/in/sergiocordoba

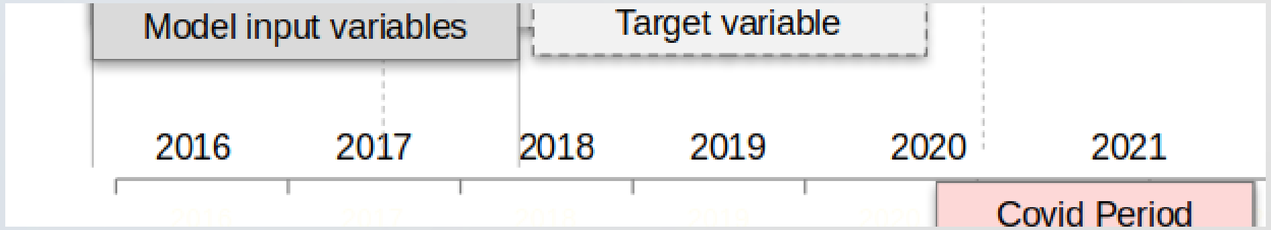
BUSINESS CASES EXAMPLES

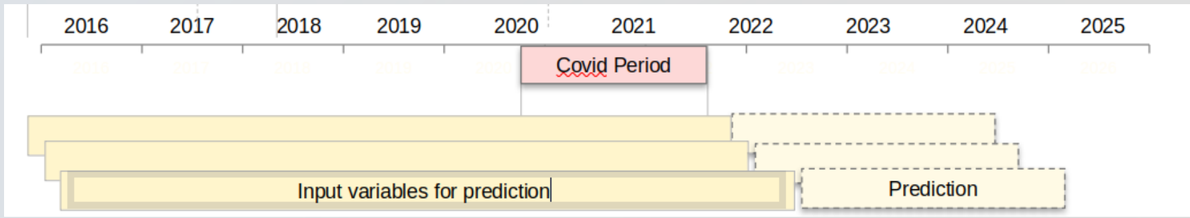
data scientist | **+10 years experience** in differents sectors | **oriented to business** processes | **task & timings** management | **agile** methodologies
end-to-end data (science) pipelines design, building and operationalization | **monitoring** business-KPIs and model performance

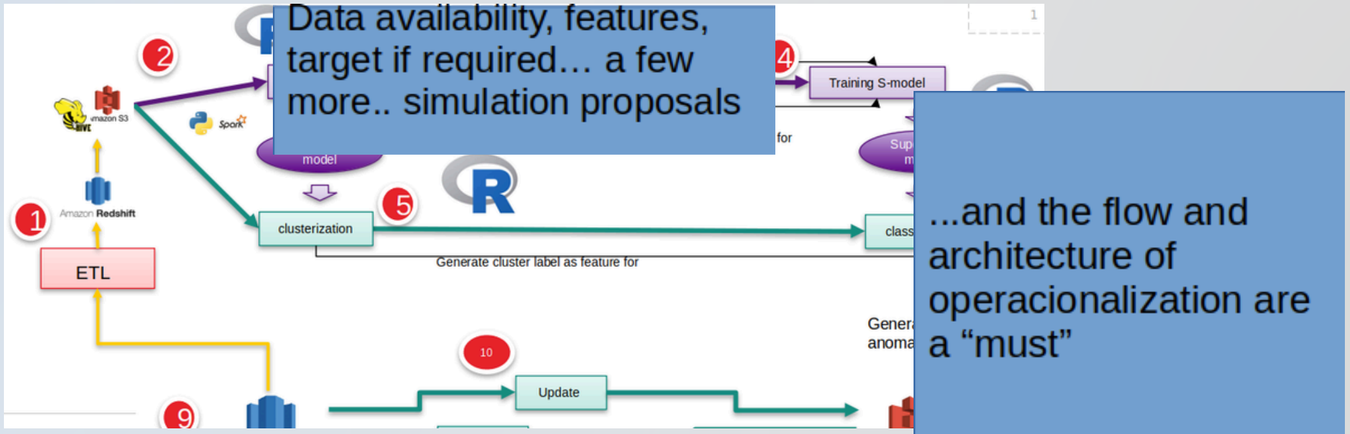
Understanding and definition in several senses...



...design and review the (model) approach from an analytical perspective...







...and the flow and architecture of operacionalization are a “must”

** project steps based on CRISP-DM work-flow

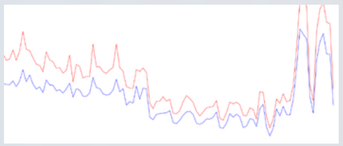
Sergio Cordoba (MSc. Physics, MSc. Mathematical modeling)
https://www.linkedin.com/in/sergiocordoba

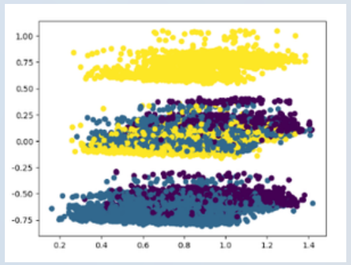
BUSINESS CASES EXAMPLES

data scientist | **+10 years experience** in differents sectors | **oriented to business** processes | **task & timings** management | **agile** methodologies
end-to-end data (science) pipelines design, building and operationalization | **monitoring** business-KPIs and model performance

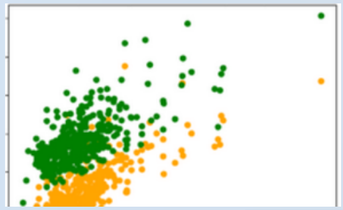
---

Data preparation and exploration...





Different graphs could be used in according with we want to show...

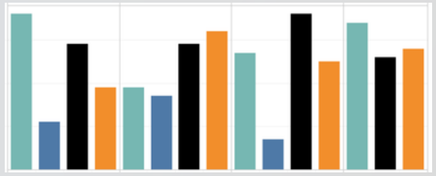


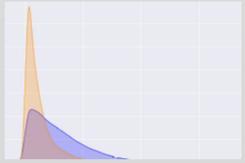
We can review, variable evolution in some sense;

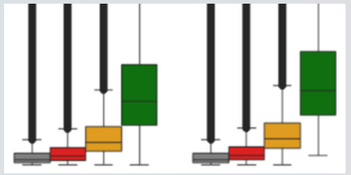
- target in training VS predict sets or trendencies along

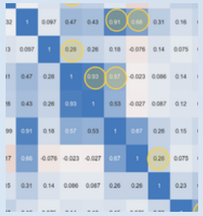
Even we could check different sources or queries and find differences...

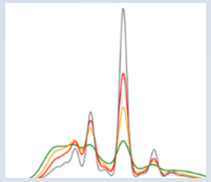
...the evolucion of a price along the time

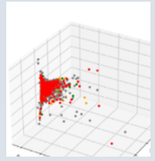


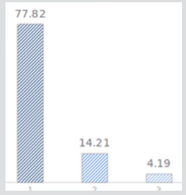


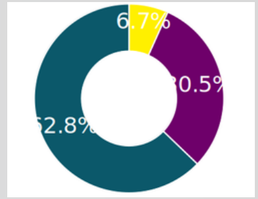












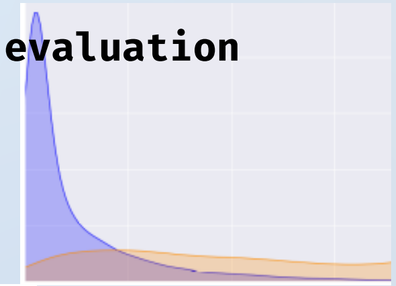
** project steps based on CRISP-DM work-flow

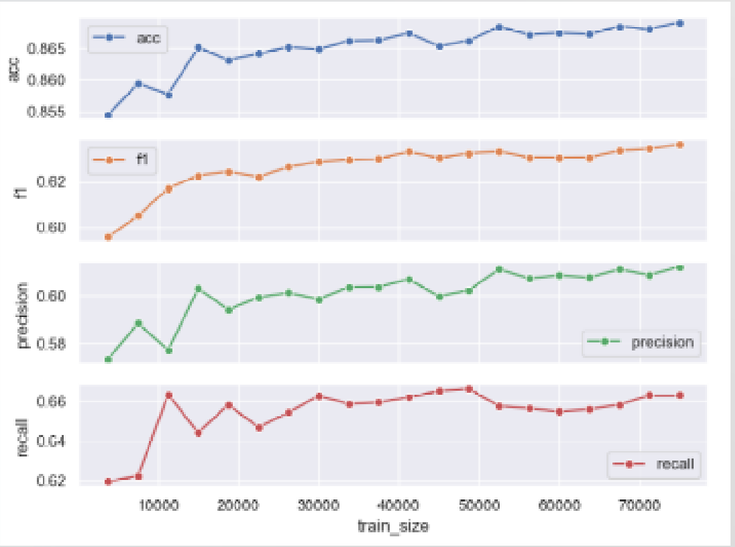
Sergio Cordoba (MSc. Physics, MSc. Mathematical modeling)
https://www.linkedin.com/in/sergiocordoba

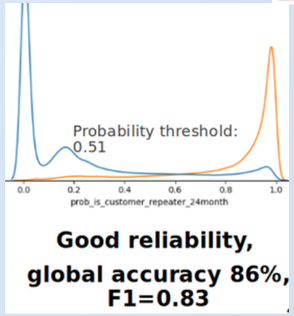
BUSINESS CASES EXAMPLES

data scientist | **+10 years experience** in differents sectors | **oriented to business** processes | **task & timings** management | **agile** methodologies
end-to-end data (science) pipelines design, building and operationalization | **monitoring** business-KPIs and model performance

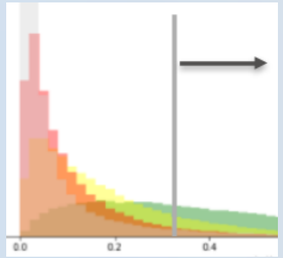
Modeling and evaluation



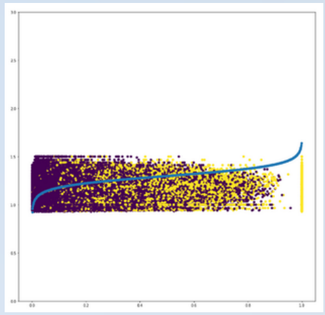




Good reliability,
global accuracy 86%,
F1=0.83



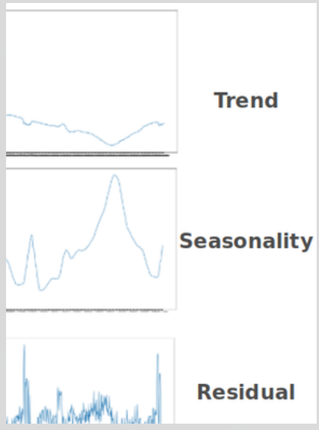
Metric	1.High Recall	2.Optimum F1 (reference)	3.High precision
probability threshold	**0.2**	**0.3**	**0.4**
F1	0.49	**0.51**	0.48
Precision	**0.38**	0.47	**0.54**
Recall (sensitivity)	**0.71**	0.55	**0.43**
Accuracy (global of classes)	0.51	0.53	0.53



Variables ciclicas se observan y se pueden descomponer (time-series, forecasst)

Clasifiers calibration can be analyzed to select the best performance

Or the more suitable fitting can be found



** project steps based on CRISP-DM work-flow

Sergio Cordoba (MSc. Physics, MSc. Mathematical modeling)
https://www.linkedin.com/in/sergiocordoba

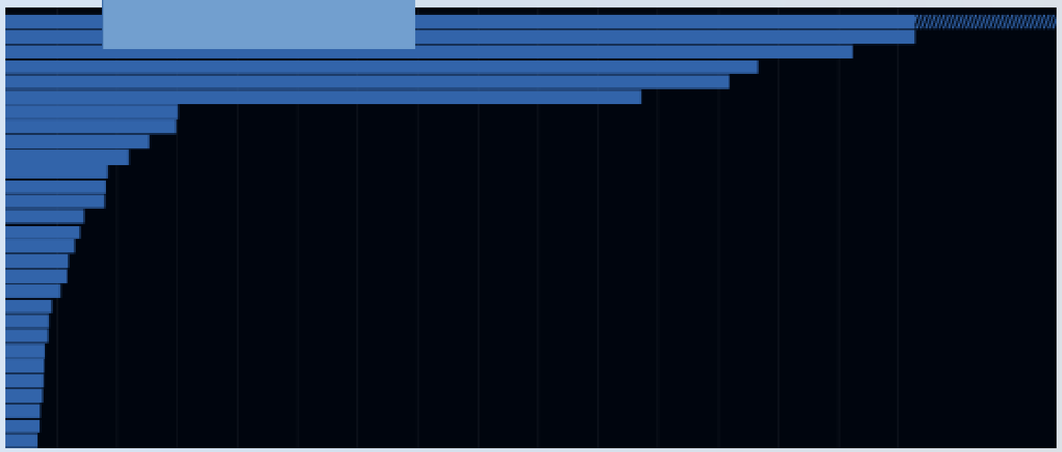
BUSINESS CASES EXAMPLES

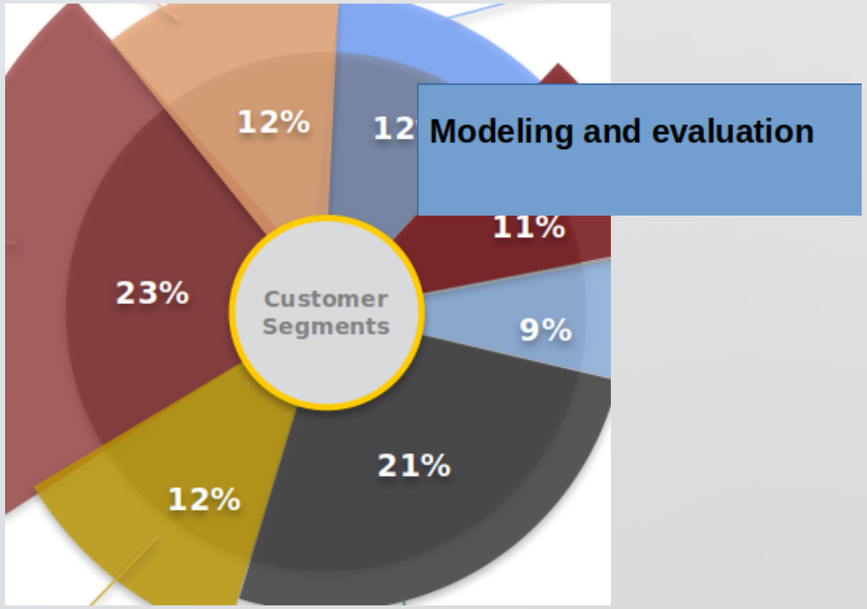
data scientist | **+10 years experience** in differents sectors | **oriented to business** processes | **task & timings** management | **agile** methodologies
end-to-end data (science) pipelines design, building and operationalization | **monitoring** business-KPIs and model performance

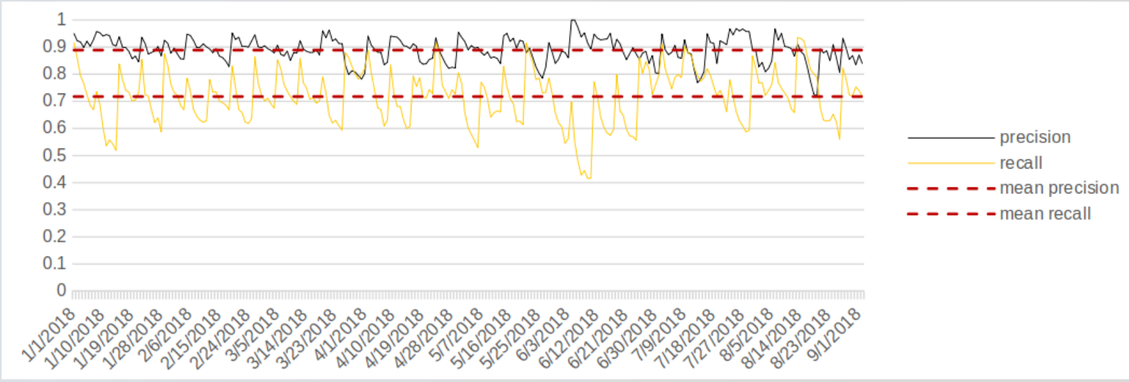
Modeling and evaluation

Metrics can be tracked along the time

Cycles or lost of performances in some sense, could be







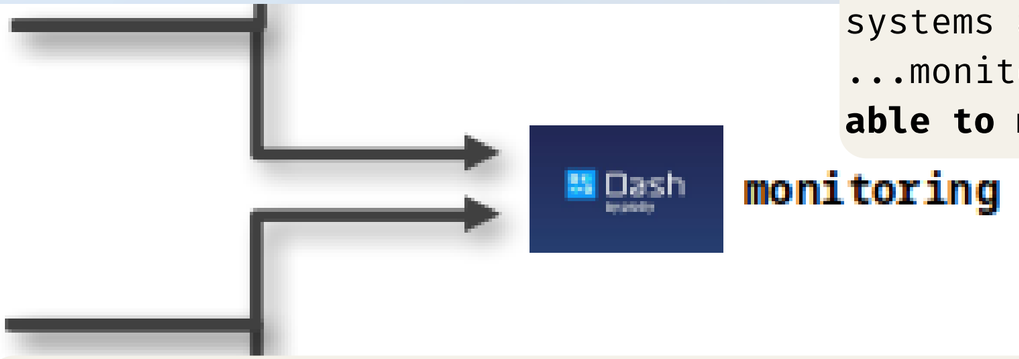
** project steps based on CRISP-DM work-flow

Sergio Cordoba (MSc. Physics, MSc. Mathematical modeling)
https://www.linkedin.com/in/sergiocordoba

BUSINESS CASES EXAMPLES

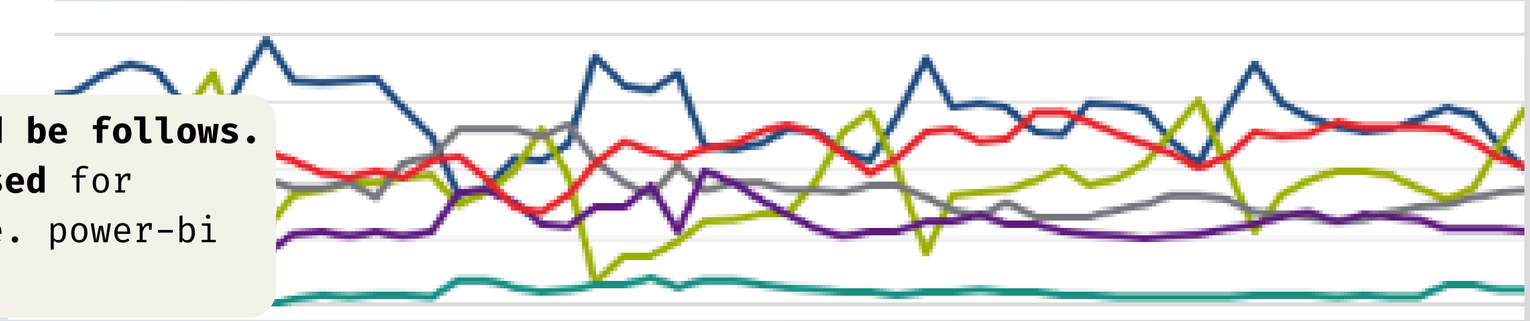
data scientist | **+10 years experience** in differents sectors | **oriented to business** processes | **task & timings** management | **agile** methodologies
end-to-end data (science) pipelines design, building and operationalization | **monitoring** business-KPIs and model performance

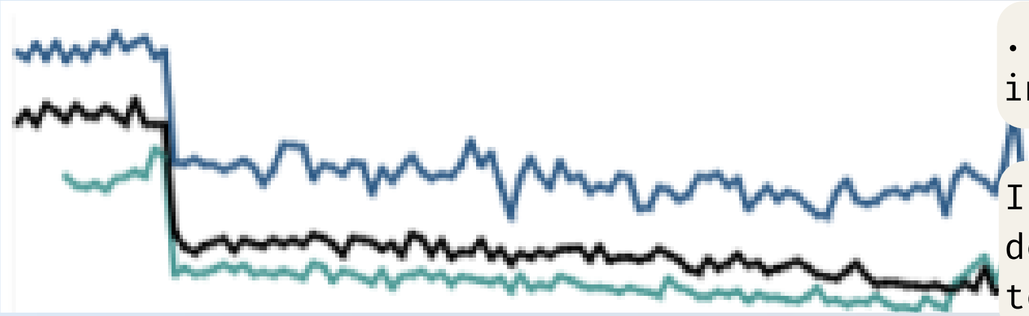
Deploy and monitoring



Productive models must be associated to control processes, in analytical, systems and business terms...
...monitoring along the time **becomes mandatory (and beautiful), also, to be able to monetize the project and track the performance and behaviour**

...different (technical or KPIS) **metrics could be follows. In this context, visualization tools can be used** for developers (i.e. grafana) and stakeholder (i.e. power-bi or tableau)...





...give us the capability to detect cycles, drastic changes, alerts, increments, goals...

I have been involved around projects standarization, visualization deployment tasks with IT and MLOps departments. Cross-functional teams.

** project steps based on CRISP-DM work-flow