MY NAME : SAGAR SADASHIV SATPUTE

DATA SCIENCE BATCH 3. ROLL NO: 2021248

PROJECT NAME : **Identify premium pricing attributes for home insurance using R**

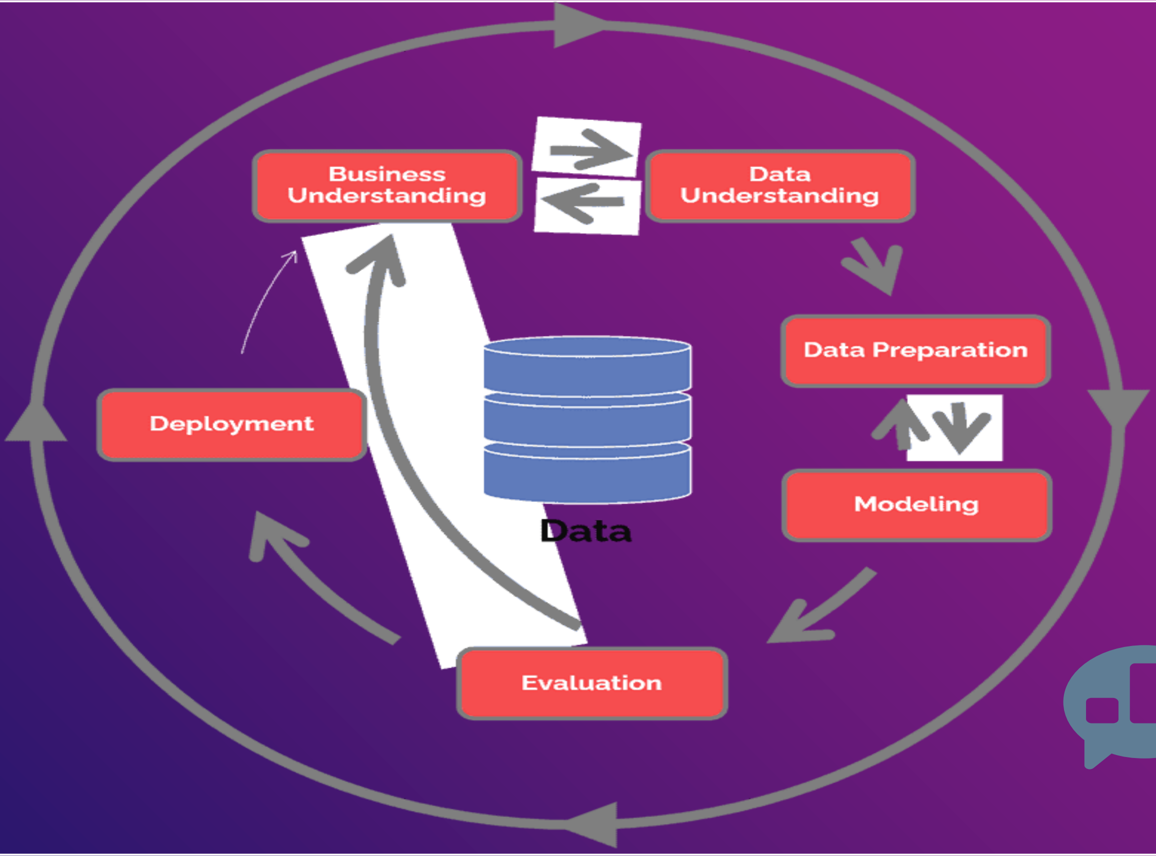
TASK : Research CRISP - DM

WHAT IS CRISP DM

INDRODUCTION

The CRoss Industry Standard Process for Data Mining (*CRISP-DM*) is a process model that serves as the base for a data science process. It has six sequential phases:

1. Business understanding – What does the business need?
2. Data understanding – What data do we have / need? Is it clean?
3. Data preparation – How do we organize the data for modeling?
4. Modeling – What modeling techniques should we apply?
5. Evaluation – Which model best meets the business objectives?
6. Deployment – How do stakeholders access the results?


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data

HISTORY

Published in 1999 to regularize data
mining processes across diligence, it has since come the
most common methodology for data mining, analytics, and
data science projects.

Data science teams that combine a loose perpetration of CRISP-
DM with overarching team-
 based nimble project management approaches will probably see t
he best results.

TABLE OF CONTENT

01

What are the 6 CRISP-DM Phases?

1.BUSINESS UNDERSTANDING

The *Business Understanding* phase focuses on understanding the objectives and requirements of the project. Aside from the third task, the three other tasks in this phase are foundational project management activities that are universal to most projects:

1. **Determine business objectives**
2. **Assess situation**
3. **Determine data mining goals**
4. **Produce project plan**

2.DATA UNDERSTANDING

Next is the *Data Understanding* phase. Adding to the foundation of *Business Understanding*, it drives the focus to identify, collect, and analyze the data sets that can help you accomplish the project goals. This phase also has four tasks:

1. **Collect initial data**
2. **Describe data**
3. **Explore data**
4. **Verify data quality**

3.DATA PREPARATION

This phase, which is often referred to as “data munging”, prepares the final data set(s) for modeling. It has five tasks:

1. Select data
2. Clean data
3. Construct data
4. Integrate data
5. Format data

4.MODELING

Here you'll likely build and assess various models based on several different modeling techniques. This phase has four tasks:

1. **Select modeling techniques**
2. **Generate test design**
3. **Build model**
4. **Assess model**

5. EVALUATION

Whereas the *Assess Model* task of the *Modeling* phase focuses on technical model assessment, the *Evaluation* phase looks more broadly at which model best meets the business and what to do next. This phase has three tasks:

1. **Evaluate results**
2. **Review process**
3. **Determine next steps**

6. DEPLOYMENT

A model is not particularly useful unless the customer can access its results. The complexity of this phase varies widely. This final phase has four tasks:

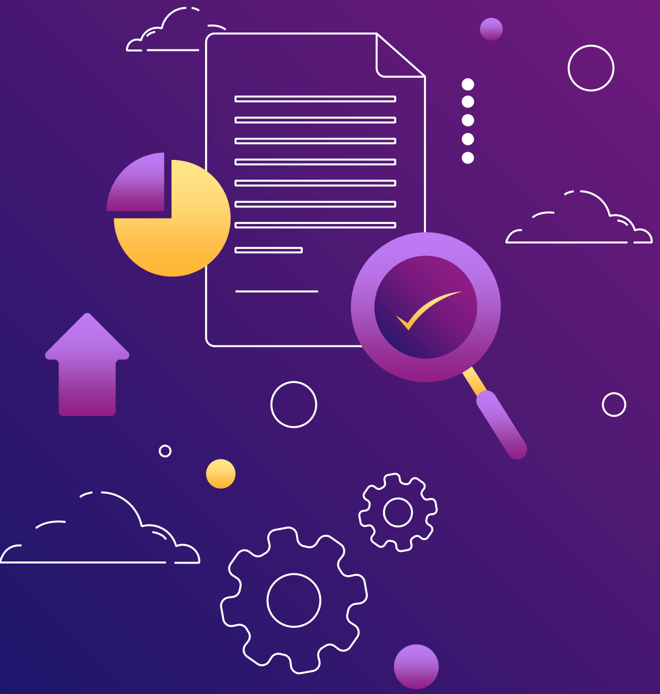
1. Plan deployment
2. Plan monitoring and maintenance
3. Produce final report
4. Review project

02

AGILE OR WATERFALL

CRISP-DM WATERFALL: HORIZONTAL SLICING

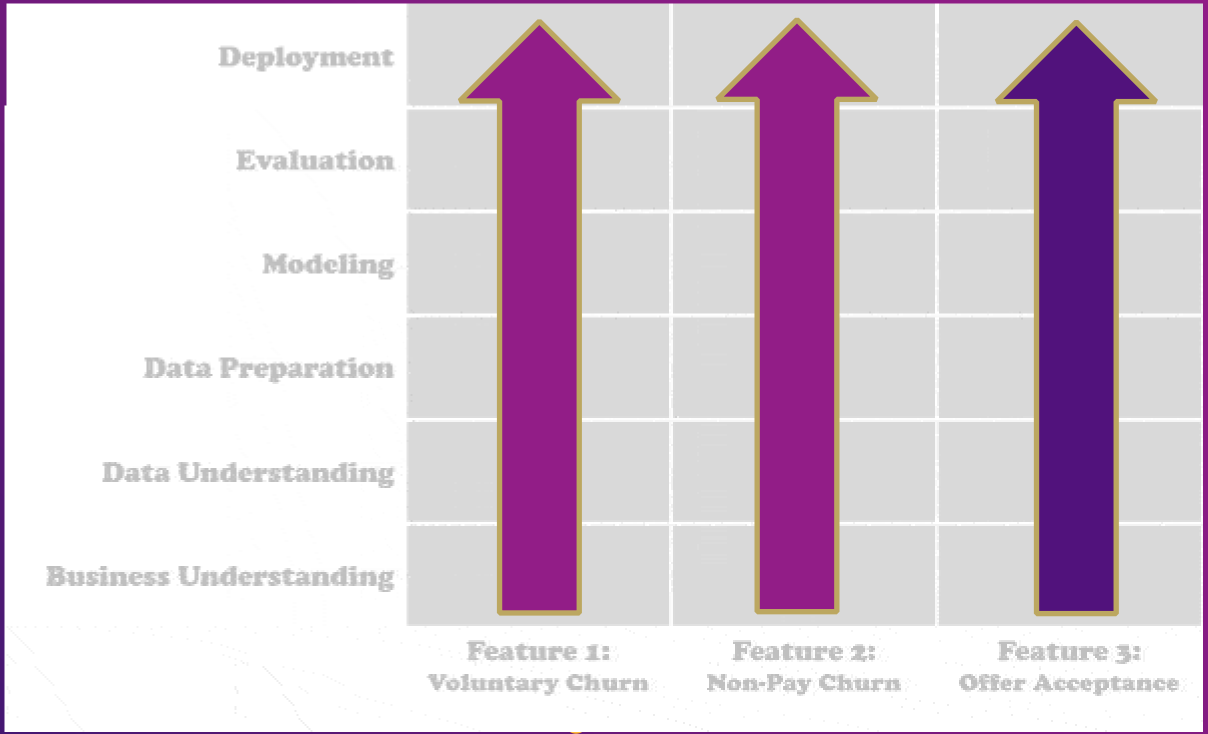
In a waterfall- style perpetration,
the team’s work would exhaustively and
horizontally measure across each deliverable as shown below.
The team might rarely loop back to a lower horizontal layer only if critically demanded. One “ big bang ” deliverable is delivered at the end of the project.




Deployment
Evaluation
Modeling
Data Preparation
Data Understanding
Business Understanding
Feature 1:
Voluntary Churn
Feature 2:
Non-Pay Churn
Feature 3:
Offer Acceptance

CRISP-DM AGILE: VERTICAL SLICING

Alternately, in an agile perpetration of CRISP- DM,
the team would hardly concentrate on snappily delivering one verti
cal slice up the value chain at a time as shown below. They
would deliver multiple smaller vertical releases and often solicit fe
edback along the way.


Deployment
Evaluation
Modeling
Data Preparation
Data Understanding
Business Understanding
Feature 1:
Voluntary Churn
Feature 2:
Non-Pay Churn
Feature 3:
Offer Acceptance

WHICH IS BETTER

When possible, take an agile approach and slice vertically so that:

1. Stakeholders get value sooner
2. Stakeholders can provide meaningful feedback
3. The data scientists can assess model performance earlier
4. The project team can adjust the plan based on stakeholder feedback

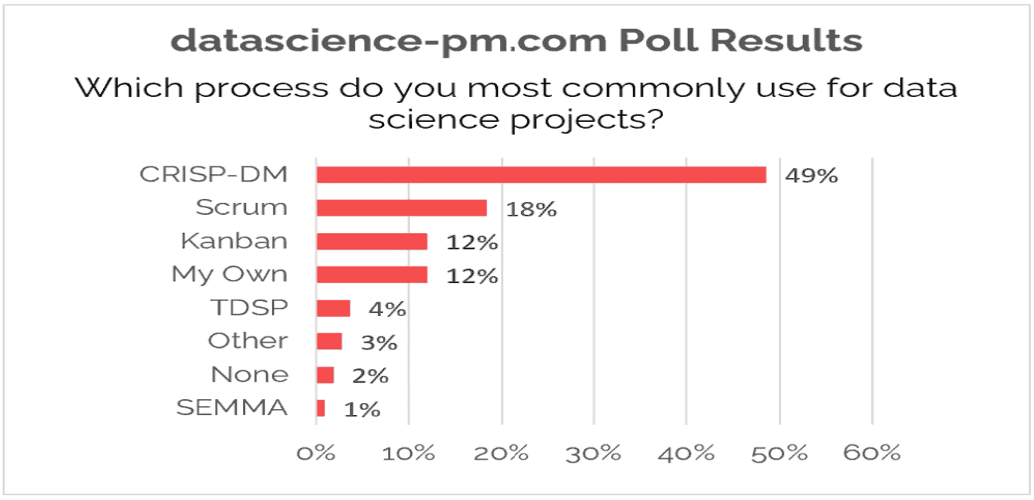
03

HOW POPULAR IS CRISP-DM

2020 POLL

For a more current look into the popularity of various approaches, we conducted our own poll on this site in August and September 2020.

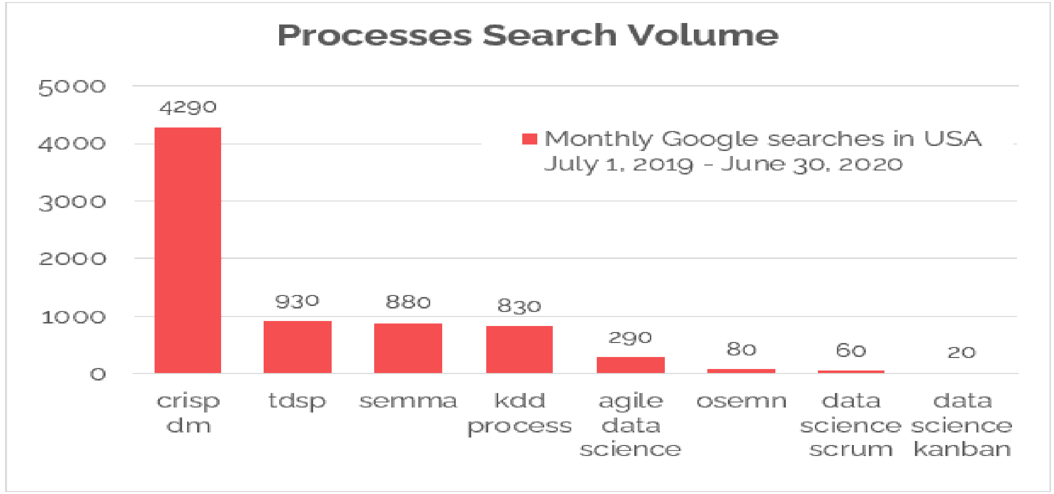
CRISP-DM was the clear winner, garnering nearly half of the 109 votes.



GOOGLE SEARCHES

For yet third view into CRISP-DM, we turned to Google Keyword Planner tool which provided the average monthly search volumes in the USA for select key search terms and related terms (e.g. “crispdm” or “crisp dm data science”). Clearly irrelevant searches like “tdsp electrical charges” or “semma both aagatha” were then removed

- CRISP-DM yet again reigned as king, and this time with a much broader margin.



04 BENEFITS & WEAKNESSES

BENEFITS

1. Common Sense
2. Generalize-able
3. Adopt-able
4. Right Start
5. Strong Finish
6. Flexible

WEAKNESSES & CHALLENGES

1. Rigid
2. Documentation Heavy
3. Not Modern
4. Not a Project Management Approach

05 CRISP-DM ALTERNATIVES

1.SEMMA*(Sample, Explore, Modify, Model, and Assess)*

Compared to CRISP- DM, SEMMA
is indeed more hardly focused on the specialized way of data mining. It skips over
the initial Business Understanding phase from CRISP- DM
and rather starts with data sampling processes.
SEMMA likewise doesn't cover the final Deployment aspects. els e, its phases kindly image the middle four phases of CRISP-DM. Although potentially useful as a process to follow data mining way, SEMMA shouldn't be viewed as
a comprehensive project management approach.

2. KDD AND KDDS

KDD:
Knowledge Discovery in Database(KDD) is the general process of discovering knowledge in data through data mining, or the birth of patterns and information from large datasets using machine literacy, statistics, and database systems.

KDDS (*Knowledge Discovery in Data Science)*:
KDDS can be a useful expansion of CRISP-DM for big data teams. However, KDDS only addresses some of the shortcomings of CRISP-DM. For example, it is not clear how a team should iterate when using KDDS. In addition, its combination of phases and processes is less straight-forward. Adoption of KDDS outside of SAIC is not known.

BIBLIOGRAPHY:

Website: https://www.datascience-pm.com/crisp-dm-2/

Author : **NICK HOTZ**

CRISP-DM Diagram. Inspired by Wikipedia

Template: https://slidesgo.com/

THANK YOU