DATA SCIENCE

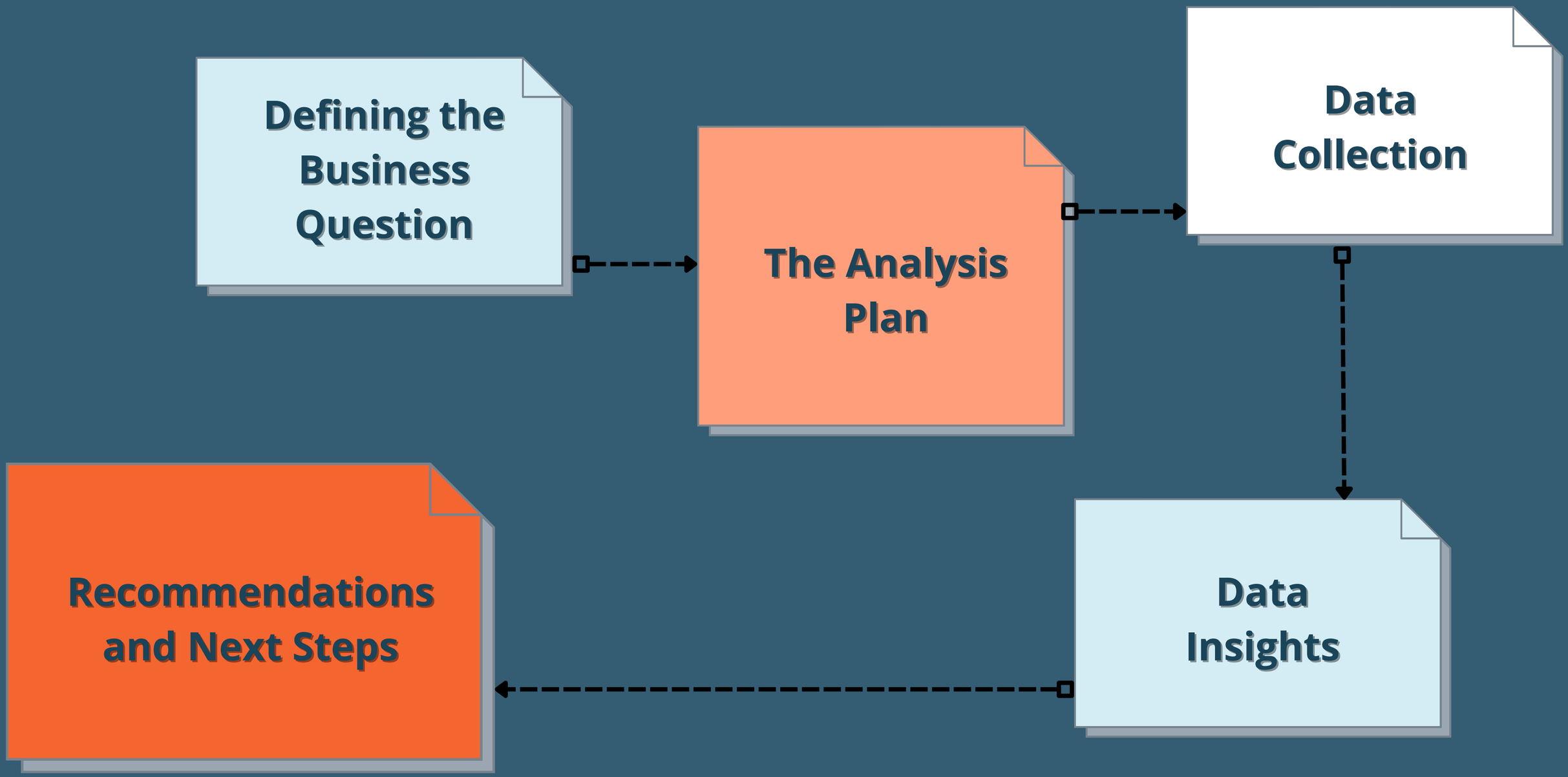
Framework Report

THE GOAL . . . . DEFINED

The goal of this project is to utilize data science to address the critical business issue of increased customer defaults which threatens the business of Credit One.

To Construct a clear and effective data science process, the BADIR framework (Jain and Sharma, Behind Every Good Decision, chapter 4) will be used as a guideline.

PROPOSED PROCESS



THE BUSINESS QUESTION

Credit One needs a much better way to understand how much credit to allow someone to use or, at the very least, if someone should be approved or not.

The Context

Over the past year or so Credit One has seen an increase in the number of customers who have defaulted on loans they have secured from various partners.

Impacted Segment

As the credit scoring service, Credit One could risk losing business if the problem is not solved right away.

Business Considerations

The Data Science team has been given full authority to solve this problem with whatever tools and methods needed.

ANALYSIS PLAN

ANALYSIS GOAL

To examine current customer demographics to better understand what traits might relate to whether or not a customer is likely to default on their current credit obligations.

HYPOTHESES

Which customer attributes relate significantly to customer default rates and can a predictive model be used to better classify potential customers as being ‘at-risk’?

REQUIRED DATA

- Customer Demographics
- Payment History
- Credit Limit
- Default Status

MYTHODOLOGY

- Data descriptives and correlations.
- Machine learning regression and classification methods in Python

DATA SOURCES

From Credit One Database

Given Credit Amount

Customer Demographics

Gender

Age

Marital Status

Education

Payment History

Payment Amount

Bill Statement Amount

Default? Yes/No

The data will be cleansed and validated using data mining and Exporatory Data Analysis to identify and resolve any issues such as data types and missing data.

DATA MANAGEMENT

Some initial issues identified

Data types are all object and some need to be converted to numerical.

Some data seems to be duplicated.

Seems there are no missing values

To resolve this, appropriate data types will be changed to numerical and duplicate rows will be removed.

INSIGHTS

Preliminary insight from data shows that payment history has a relationship with default status.

1. Present patterns seen in the data.
2. Present provability of the hypotheses.
3. Provide level of confidence stakeholders should place in the results.
4. Rank findings in terms of impact on the business.