***Text Analysis of <<Name Dataset(s)>> for <<Name Purpose>>***

***(Or something more impactful)***

*By*

*Author(s)… who you are*

*date*

**Executive Summary**

Insert here your executive summary… one page… max

This is usually a one-page abstract of the entire report, and may be all that a busy business executive reads.

It should:

* explain the terms of reference, the purpose and scope of the report,
* state the key methods and approach used… CRISP-DM approach to analysis
* list the main conclusions,
* list the key recommendations.

You usually do this last. It is not the same as the introduction.





# Introduction

First Paragraph:

1. The purpose of the report; name of the commissioner of the report;
2. Type of report … Proposal for Text Analytics Task on
3. Scope (to identify, to propose, to recommend…) using the CRISP-DM format

Next Paragraphs:

* Ideally, you want your readers engaged from the first sentence. Make your report as purposeful and useful as possible. This should start out with a hook… such as… “Artificial Intelligence is seen as…”
* Use some client/company/industry background information to get your readers involved. If you have any stark facts that shed some light on why the issue needs to be addressed and addressed immediately, use those here.
* Describe contents of the **Business Understanding**, **Data Understanding** and **Plan of Work** sections ahead:

# Methodology

## CRISP-DM Approach

Describe the stages of CRISP-DM and show how it informs the basis for your document.



Figure 1: CRISP-DM Processes

Focus on the contents of the proposal and your focus on Business Understanding of the current position, Data Understanding of the data you have been given, and the plan for doing Data Preparation, Modelling and Evaluation.

You will be presenting the results of your analysis to allow them to make deployment decisions.

# Business Understanding

## Business Objectives

### Current Business Situation

Record all relevant information that is known about the organization’s business situation at the beginning of the project. Be comprehensive and explain what is being proposed and what impact it will have.

### Primary and Secondary Objectives

Describe the customer’s primary objective(s), from a business perspective. In addition to the primary business objective, there are typically other related objectives. Be clear about the tasks you are addressing.

### Business Success Criteria

Describe the criteria for a successful or useful outcome to the project from the business situation the company is in, and from the business point of view. Any objective described should be specific, measurable, achievable, and realistic and time-limited (SMART).

## Data-Mining Goals

Introduce the key objectives you will aim at. Text analytics data-mining goal needs project objectives to be stated in appropriate technical terms.

### Describe the Descriptive Analytics Goals

* What you will do in each of the areas
* What tools you will use
* Give an example for each of the 3 of a tool in operation and show some preliminary results

### Describe the Predictive Analytics Goals

* What you will do in predictive analytics
* What tools you will use
* Give an example and show some preliminary results

You might also show images of workflows that you will use in Orange.

Make sure any image you use is labelled.

Describe the intended outputs of the analysis that will come as a consequence of doing the whole project and how the results will enable the achievement of the business objectives.

### Define Data-Mining Success Criteria

Define the criteria for a successful outcome to the project in technical terms—for example, a certain level of descriptive accuracy in describing the contents of the data. Or elements of a plan to employ predictive methods to discover a significant response and how to make it cope with it. Describe how you will seek to evaluate the work by sharing it with professionals in your client’s company.

# Data Understanding

## **Initial data collection report**

Describe the data that has been acquired (the spreadsheet materials and the sustainability report) including its format, its quantity (for example, the number of records and fields in each table), the identities of the fields and any other surface features which have been discovered. Evaluate whether the data acquired satisfies your requirements.

## Explore data

During this stage you’ll address data mining questions using querying, data visualization and reporting techniques.

### How you will Explore and Analyse all Datasets

* What’s in the data
* The relationship between some questions and unstructured text
* Simple descriptive analyses

Pick out key questions that relate to the proposal you will be making… for example, if it was about people’s beliefs, you may choose a particular question and then represent an answer to it using word cloud or concordance or data provided in a list or table.

Describe what ways to will look into the text data. Describe the results of your data exploration, including first findings or initial hypothesis and their impact on the remainder of the project.

Make sure that in your examples and preliminary work you can use impactful images, such as any word cloud you may wish to use. (Please note that this is a template … the above question is not related to the word cloud below… please ensure that you replace everything with appropriate materials).

If appropriate you could include graphs and plots here to indicate data characteristics that suggest further examination of interesting data subsets.

## Verify data quality

Examine the quality of the data, addressing questions such as:

* Is the data in your datasets complete (does it cover all the cases required)?
* Is the data correct, or does it contain errors and, if there are errors, how common are they?
* Are there missing values in the data? If so, how are they represented, where do they occur, and how common are they?

# Data Preparation

In this stage, perform Data Preprocessing/cleaning

Preprocess the given dataset to clean the text data. Include steps such as:

Removing irrelevant characters, numbers, and special symbols.

Converting text to lowercase.

Tokenization and stop-word removal.

Lemmatization or stemming.

Document the preprocessing steps and rationale behind each step.

# Modelling

In this section, present your results and findings from text analytics techniques word cloud, concordance, sentiment analysis, topic modeling, etc.

# Note: You could incorporate the four recommendations here using the storytelling framework based on the findings from modeling (i.e., analysis).

# Summary of Proposal

Summarize your findings in this section.

# References

Add references

# APPENDICES

Any appendices you wish to add