



# FINAL PROJECT PLAN

Light House Labs Q1 2021

## Abstract

NLP sentiment analysis of consumer goods with research company data to help optimize business metrics.

Stacey McLennan-Waldal  
Stacey.Waldal@gmail.com

## Executive Summary

This is the project management document for the Light House Labs Final Project for Q1 2021. This document serves as a guide for how the project will be executed.

As the consumer landscape evolves, companies are attempting to better understand their customer, their needs, their wants, and their experience with the product. Data science can help companies by using NLP (natural language processing) to gain insights into feedback provided with language (as opposed to numerical assessments) and insights into the consumer psychology. These insights are one tool which can help companies tailor their products to achieve their desired business goals and continuously improve their products.

[NLP has advanced significantly since 2018](#) to assess customer feedback and sentiment towards consumer goods. [BERT](#) is a state-of-the-art model that has been developed to obtain insights on this type of data and base models are available to be used as a starting point for new data sets.

This project will use BERT to assess customer feedback obtained during a research trial to evaluate brands. The outcome of the model will be able to assess the sentiment of the feedback provided as natural language data. From this, insights into business areas of focus can be obtained. Further, this project will serve as a proof of concept for the research company to be able to use machine learning/artificial intelligence to provide more detailed insights into their client's product studies.

## Table of Contents

Executive Summary.....	1
Project Description.....	0
Title .....	0
Goals/Outcomes .....	0
Description .....	0
Questions to Explore.....	0
Objectives & Success Criteria.....	1
Assumptions, Constraints/Dependencies .....	1
High-Level Risks.....	1
Project Setup.....	2
Git Repo .....	2
Access to Data .....	2
Communication.....	2
Reports.....	2
Project Meetings.....	2
Change Management.....	2
Deliverables.....	2
Work Breakdown Structure .....	3
Quality.....	3
Summary Milestones .....	4
Schedule.....	4
Stakeholders .....	0
Project Team .....	0

# Project Description

## Title

Sentiment Analysis for Consumer Products using NLP (BERT).

## Goals/Outcomes

To assess the sentiment of a consumer's experience with a product based on interview-style feedback.

## Description

As the consumer landscape evolves, companies are attempting to better understand their customer, their needs, their wants, and their experience with the product. Data science can help companies by using NLP (natural language processing) to gain insights into feedback provided with language (as opposed to numerical assessments) and insights into the consumer psychology. These insights are one tool which can help companies tailor their products to achieve their desired business goals and continuously improve their products.

[NLP has advanced significantly since 2018](#) to assess customer feedback and sentiment towards consumer goods. [BERT](#) is a state-of-the-art model that has been developed to obtain insights on this type of data and base models are available to be used as a starting point for new data sets.

This project will serve as a proof of concept for the research company to be able to use machine learning/artificial intelligence to provide more detailed insights into their client's product studies.

Questions to Explore:

1. What insights are in the feedback of the products based on:
  - a. What (language) was said and?
  - b. How (sentiment) was it said?

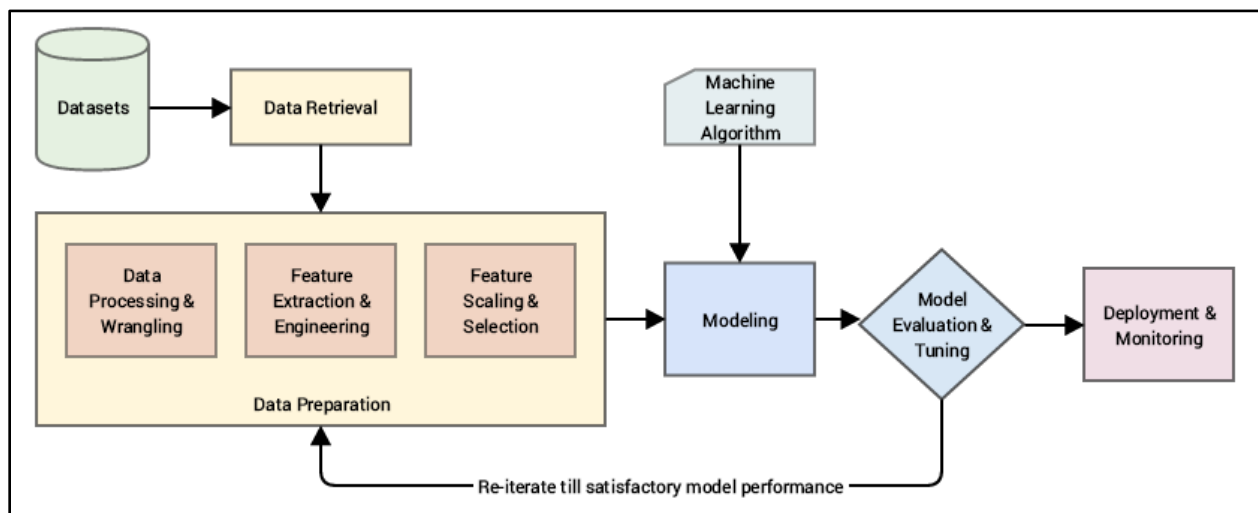


Figure 1: <https://towardsdatascience.com/understanding-feature-engineering-part-1-continuous-numeric-data-da4e47099a7b>

## Objectives & Success Criteria

Objectives	Success Criteria
<b>Data processing</b>	Clean dataframe with relevant features
<b>Modelling</b>	Use state-of-the-art model (BERT)
<b>Results</b>	Highest accuracy compared to base model
<b>Presentation</b>	Client and LHL are happy with the way the results are presented
<b>Complete project</b>	Client is happy with being involved in project LHL is happy with quality (challenge) of project

## Assumptions, Constraints/Dependencies, Requirements

Assumptions
<b>Data cleaning will take large amount of project time</b>
<b>BERT can be applied to data set</b>
<b>Mentor help available to get through stuck points</b>

Constraints/Dependencies
<b>Mentor availability and knowledge of using BERT</b>

Requirements
<b>Data</b>
<b>Computer hardware and software (Google Colab)</b>

## High-Level Risks

Risk	Level	Mitigation
<b>Communication breakdown</b>	Medium	Project Plan Progress Report Dry runs scheduled ahead of Demo Day
<b>Algorithm challenges (getting the model to work)</b>	Medium	Time dedicated to research Mentors available to help
<b>Project runs out of time</b>	Low	This project plan helps identify key work packages and timelines so that decisions to move forward when necessary can be made

## Project Setup

### Git Repo

[https://github.com/waldals/Final\\_Project](https://github.com/waldals/Final_Project)

### Access to Data

Data has been sent via email from Client to Stacey and will be stored securely on Stacey's private laptop. Original data will not be made available on GitHub.

## Communication

### Reports

- A project update will be sent on Fri Mar 19
- A final report will be created as legacy documentation and sent on Fri Mar 26

### Project Meetings

Due to the short nature of this project and resource intensiveness required to complete the model, no project meetings are scheduled.

## Change Management

Change management will be dealt with internally due to the nature of this project being a single team member. Project changes that impact the Client will be communicated at immediate onset, however none are anticipated before the presentation dry run.

## Deliverables

- Project Plan
- Timeline
- Work Packages/[WBS](#)
- Presentation
- Report

## Work Breakdown Structure

Table 1: Work Break Down Structure (WBS) of Tasks

Tasks	Est Completion Date	Notes, Dependencies
<b>Research</b>	Mar 11 – Mar 12	Research BERT model and applications
<b>Data Exploration</b>	Mar 12 – Mar 14	What do we have?
<b>Data Cleaning</b>	Mar 12 – Mar 15	Put data into table format, remove nulls, empty columns, encode, clean text, tokenize
<b>Baseline Model (MVP)</b>	Mar 16	Obtain baseline accuracy
<b>Iterate models</b>	Mar 16 – Mar 18	Obtain model with highest accuracy
<b>Select best model</b>	Mar 21	Model with highest accuracy from iteration should be tuned to find accuracy > 95%
<b>Make business observations</b>	Mar 16 – Mar 22	How does this model help the business?
<b>Create presentation</b>	Mar 21 – Mar 23	As per the guidelines
<b>Submit file</b>	Mar 26	As per the guidelines
<b>Prepare final report</b>	Mar 22 – Mar 26	Detail for Client approach and findings, identify areas for future projects

## Quality

How will we decide what good looks like?

- Accuracy performance increases over base model.
- Business insights using NLP on customer feedback obtained.
- Future areas of project work identified to build on initial findings.

How will we know when we are complete?

- When we get a model that can predict above base accuracy.
- Business insights found via output of the model.
- Presentation given on findings.
- Project report created/received.

## Summary Milestones

Table 2: Milestones

Date	Item	Notes
Mar 11	Project Kick Off Meeting	
Mar 15	Data Cleaning/Exploration completed	
Mar 16	MVP created	Initial accuracy
Mar 19	Progress Report	Send an update on how it is going
Mar 21	Finalize model	
Mar 23	Client Dry Run Presentation	Opportunity for client to give feedback on presentation
Mar 24	LHL Dry Run Presentation	Opportunity for LHL to give feedback on presentation
Mar 25	Demo Day Presentation	
Mar 26	Send Final Report Project Close	

## Schedule

Table 3: Schedule

Final Project for LHL		Status	Timeline	People
Project Planning		Working on it	Mar 11 - 12	SM
Research		Working on it	Mar 11 - 12	SM
Data Exploration		Working on it	Mar 12 - 14	SM
Data Cleaning		Working on it	Mar 12 - 15	SM
Build MVP		Up next	Mar 13 - 16	SM
MVP		Planned	Mar 16	SM
Iterate Models		Planned	Mar 16 - 18	SM
Build Final Model		Planned	Mar 18 - 21	SM
Final Model		Planned	Mar 21	SM
Prepare Presentation		Planned	Mar 21 - 23	SM
Client Presentation		Planned	Mar 23	SM
LHL Presentation		Planned	Mar 24	SM
Incorporate Presentation Fee...		Planned	Mar 23 - 24	SM
Demo Day Presentation		Planned	Mar 25	SM
Prepare Final Report		Planned	Mar 22 - 26	SM
Final Report		Planned	Mar 26	SM
Project Close Out		Planned	Mar 26	SM
			Mar 11 - 26	SM



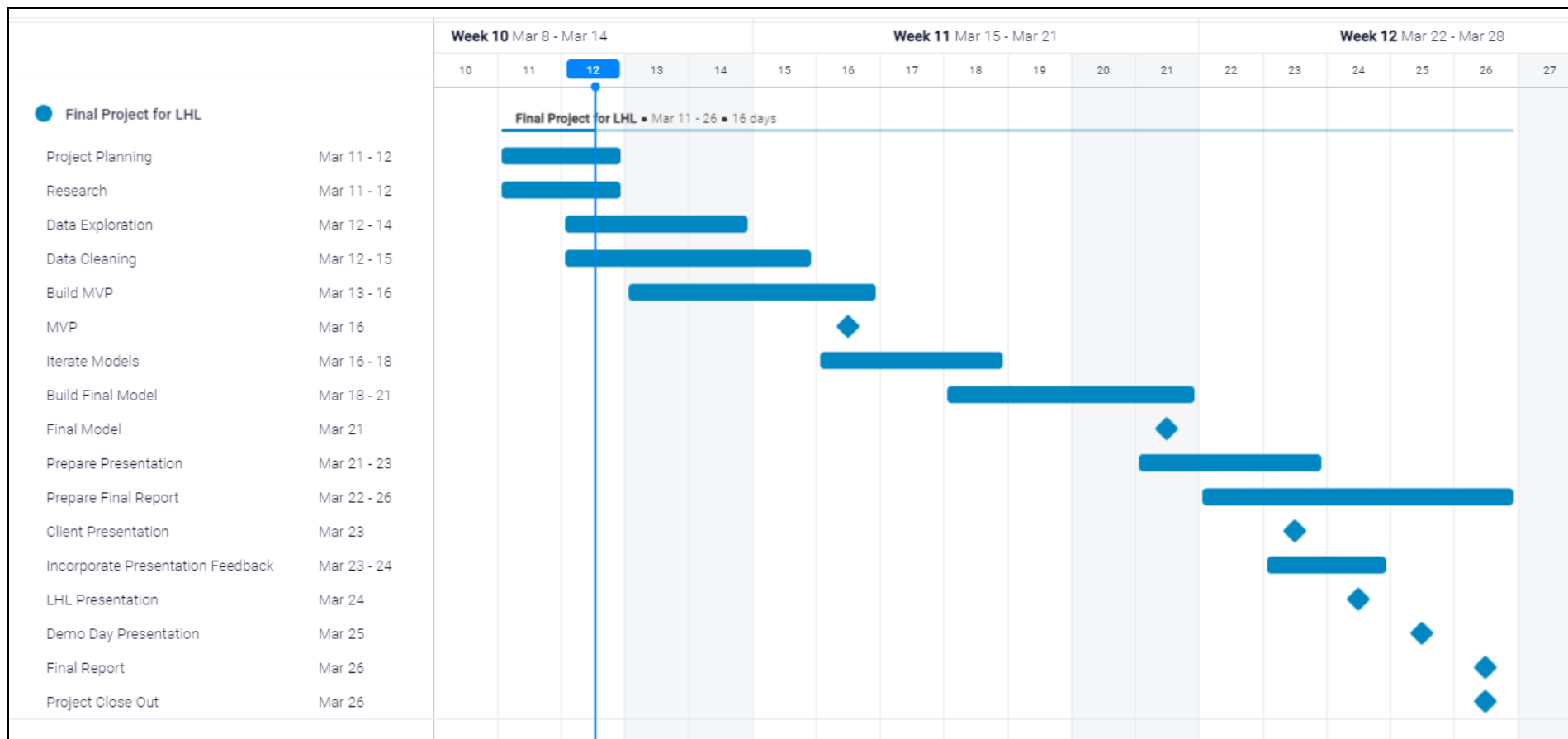


Figure 2: Gantt Chart Schedule View

## Stakeholders

Name	Position	How are they connected
<b>Clients</b>	Client	The data has been collected by their research company and has been provided to Stacey for the purposes of this Project. All personal information has been scrubbed and will not be used in the assessment.
<b>Client's Study Participants</b>	Study Participants	The data is collected based on their feedback and experience with the products. Their information has been scrubbed prior to receiving the data and will not be used in the assessment.
<b>Brands</b>	Client's client	The data is collected based on their products however all mention of Brand Name is scrubbed and will not be included in assessment. Need to be aware that they are intended to be unknown.
<b>Mentors</b>	Mentor	Advice on project execution, model evaluation, output/presentation.
<b>Light House Labs</b>	School	Provide assessment of project, host Demo day
<b>Demo Day Audience</b>	Audience	Audience for final Demo Day presentations. Potential employers will be in attendance so the presentation will focus on the data science approach of this project and the findings (without mention of any brands or names).
<b>LHL Cohort</b>	Classmates	Advice and comparison of approaches, shared learnings.

## Project Team

Name	Phone Number	Email
<b>Stacey McLennan-Waldal</b>	-	<a href="mailto:Stacey.waldal@gmail.com">Stacey.waldal@gmail.com</a>
<b>Clients</b>	-	-