

**Project: Data Collection Pipeline (Data Acquisition to Story Telling)**

**Week 7:** Deliverables

**Name:** Salih Eren Yüzbaşıoğlu

**Email:** algosalih@gmail.com

**Group Name:** Initial Group

**Country:** Turkey

**Specialization:** Data Analyst

**Batch Code:** LISUM20

**Date:** 19 May 2023

**Submitted to:** Data Glacier

**Table of Contents:**

1. Project Plan

1. Problem Statement

**1. Project Plan**

|  |  |  |
| --- | --- | --- |
| **Week** | **Date** | **Plan** |
| Week 7 | 19 May 2023 | **Perform Data acquisition. (Generate data)** |
| Week 8 | 26 May 2023 | **Collecting all Datasets into master data using** |
|  |  | **script.** |
| Week 9 | 02 June 2023 | **Clean the data and perform dedup check.** |
| Week 10 | 09 June 2023 | **Visualize the data into Dashboard.** |
| Week 11 | 16 June 2023 | **Create a batch which will at specified time** |
|  |  | **and dump the data into master file** |
| Week 12 | 23 June 2023 | **Document the challenges encountered** |
|  |  | **during this implementation.** |
| Week 13 | 30 June 2023 | **Final Project Report and Code** |

**2. Problem Statement**

**XYZ company is collecting the data customer using google forms and they have floated n number of forms on the web.**

**Company wants to create a pipeline which will collect all the data of these google forms and visualize the data in the dashboard.**

**The dataset needs to be clean and if there is any data issue present in the data then it should be treated by this pipeline (duplicate data or junk data. Dedup check should be preformed on the email id of the customer.**

3