Week 9 Report

Team Member's details

Group Name: The Potent self

Name: Abdulrahman Afifi

Email: abdulrahmanafifi33@gmail.com

Country: Turkey

College/Company: Bahcesehir University

Specialization: Data Analyst

Problem Description

XYZ firm is gathering client data using Google Forms/Survey Monkey and has published a variety of n forms on the web.

The company want to build a pipeline that would collect all of the data from these Google forms/survey surveys and visualise it on the dashboard.

The company needs clean data, and if there are any data issues in the data, they should be addressed via this process (duplicate data or junk data). A dedup check should be run on the customer's email address.

Data cleansing and transformation

I used some regex to extract email ID of each customer and then deleted any duplicated email ID that occurred in our data frame. Also, I extracted the country code from phone number of each customer and hour, minute, and second from timestamp date time object has been extracted for more details in analysis. There is no outliers and NA values detected in our dataset.

GitHub Repo link:

https://github.com/AfifiGhost2000/DataCollectionPipelineProject.git