Data Collection Pipeline

Group name: Initial Group

Name: Ahmet Metin Zengin

Email: ahmetmzengin@icloud.com

Country: Turkey

Specialization: Data Analyst

Submission Date: 19-May-2023

Submitted to: Data Glacier

1. Problem description

XYZ company is a global organization that provides various products and services to customers worldwide. As part of their efforts to enhance customer experience and satisfaction, the company collects customer feedback and preferences through online surveys conducted using Google Forms and Survey Monkey.

However, managing and analyzing the collected survey data poses a significant challenge for the company. The data received from these surveys often contain inconsistencies, such as duplicate entries, junk data, or missing values, which can affect the accuracy and reliability of the analysis.

To address this challenge, XYZ company aims to develop a data pipeline that efficiently collects, cleans, and processes the survey data. The pipeline will perform data validation and quality checks to ensure the data's integrity and eliminate any inconsistencies. One critical aspect of the pipeline is to identify and eliminate duplicate entries based on customers' unique email IDs.

Additionally, the company wants to visualize the collected data in a dashboard to gain valuable insights into customer preferences, satisfaction levels, and demographics. The dashboard will provide visual representations and interactive tools to explore the data and derive actionable insights for the company's decision-making processes.

By implementing this data pipeline and dashboard, XYZ company aims to streamline the data collection and analysis process, ensuring the availability of clean and reliable data for making informed business decisions. The pipeline's ability to identify and handle data issues will improve data accuracy and allow the company to gain valuable insights into customer behavior, preferences, and satisfaction levels, ultimately leading to enhanced customer satisfaction and improved business performance.

2. Project lifecycle

Week 7: Data Acquisition

During Week 7, in the Data Acquisition phase, we focus on gathering the necessary data for the analysis. This step involves identifying and obtaining the relevant data sources that are required to address the project objectives.

Week 8: Data Understanding

In Week 8, during the Data Understanding phase, we focus on gaining a deeper understanding of the data we have acquired for analysis. This involves examining the data types and identifying any problems or issues present in the dataset.

Week 9: Data Cleaning and Transformation

During Week 9, in the Data Cleaning and Transformation phase, we focus on preparing the acquired data for further analysis. This step involves identifying and addressing any data quality issues, performing necessary transformations, and ensuring that the data is in a suitable format for the analysis.

Week 10: Final Recommendation

During Week 10, in the Final Recommendation phase, we focus on conducting the analysis and generating actionable insights based on the cleaned and transformed data. This step involves applying appropriate analytical techniques, interpreting the results, and formulating final recommendations or conclusions for the project.

Week 11: EDA Presentation

During Week 11, in the EDA (Exploratory Data Analysis) Presentation phase, we focus on preparing and delivering a presentation that showcases the key findings and insights from the data analysis. The goal of this step is to effectively communicate the results of the analysis to stakeholders and provide them with a comprehensive understanding of the data.

Week 12: Dashboard

During Week 12, in the Dashboard phase, we focus on creating a data dashboard that presents the key findings and insights from the analysis in an interactive and visually appealing manner. The goal of this step is to provide stakeholders with a dynamic and user-friendly interface to explore the data and gain insights.

Week 13: Project Delivery

During Week 13, in the Project Delivery phase, we focus on finalizing and delivering the data analysis project to stakeholders. This step involves ensuring that all components of the project are complete, conducting a final review and validation, and delivering the project outputs.