Project Proposal: Customer Behavior Analysis

Introduction: The goal of this project is to analyze customer behavior by using the Google Analytics Customer Revenue Prediction dataset (<https://www.kaggle.com/c/ga-customer-revenue-prediction>) to better understand how customers interact with a website and how their behavior contributes to revenue.

Problem Statement: The problem is to predict the natural logarithm of the sum of all transactions per customer, which is a proxy for revenue. This is important because it can help businesses understand which factors are most important in driving revenue and make data-driven decisions to improve their websites and increase revenue.

Motivation for Solution: The motivation for this solution is to provide businesses with insights into customer behavior and how it contributes to revenue. By understanding these relationships, businesses can make data-driven decisions to improve their websites and increase revenue.

Methodology: The methodology for this project will be to first clean and preprocess the dataset, including handling missing values and dealing with outliers. Next, exploratory data analysis will be performed to understand the relationships between different variables and the target variable. Finally, machine learning algorithms will be applied to build a predictive model.

Technology: The technology to be used in this project will be Python programming language and its libraries such as pandas, numpy, matplotlib, and scikit-learn. These libraries provide a variety of tools for data cleaning, preprocessing, visualization, and modeling.

References:

1. Kaggle, (2021). Google Analytics Customer Revenue Prediction. <https://www.kaggle.com/c/ga-customer-revenue-prediction>
2. Scikit-learn, (2021). Machine Learning in Python. <https://scikit-learn.org/stable/>
3. Pandas, (2021). Data Analysis and Data Structures. <https://pandas.pydata.org/>
4. Matplotlib, (2021). Plotting Library. <https://matplotlib.org/>
5. Numpy, (2021). Scientific Computing with Python. <https://numpy.org/>