**Executive Summary**

In this report, an analysis of clickstream data from an online retail shop is conducted to provide insights for optimizing website traffic and revenue generation. The dataset provided, "e-shop clothing 20081.txt" and "shop clothing infor 2008.txt," was analyzed using appropriate data analytics tools to fulfill the tasks outlined. The report includes data cleaning, visualization of sales and revenue trends, and descriptive analysis of another dataset using R. Additionally, insights from the "starwars" dataset using dplyr in R are provided. The findings aim to guide decision-making for enhancing the online retail shop's performance.

**Introduction**

The objective of this report is to leverage data analytics techniques to derive actionable insights from the clickstream dataset of an online retail shop. By examining sales patterns, revenue trends, and conducting descriptive analysis, this report aims to support strategic decisions aimed at increasing traffic and revenue generation for the website.

**Data Cleaning and Preparation**

The initial step involved cleaning and preparing the dataset for analysis. This included handling missing values, correcting data formats, and ensuring data integrity. The columns relevant to the analysis were identified based on their significance in understanding sales, revenue, and user behavior on the website.

**Task A: Extracting Relevant Columns**

The columns extracted for analysis were chosen based on their relevance to understanding sales and revenue metrics. Key columns included product ID, transaction date, sales amount, and customer demographics where available. These columns provide insights into customer purchasing behavior and revenue generation trends.

**Task B: Sales Analysis by Month**

Using appropriate charts, the sales trends by month were visualized. This analysis helps in understanding seasonal variations in sales and identifying peak months for revenue generation. The insights gained contribute to planning marketing campaigns and optimizing inventory based on demand fluctuations.

**Task C: Revenue Analysis by Month**

Revenue trends were analyzed by aggregating sales data for each month and plotting total revenue over time. This visualization highlights revenue growth patterns and identifies opportunities for increasing revenue through targeted strategies such as promotional offers during peak sales months.

**Task D: Descriptive Analysis Using R**

A separate dataset ("input.csv") was analyzed using R to compute descriptive statistics such as mean, median, mode, standard deviation, and variance of the salary column. These statistics provide a comprehensive overview of the salary distribution within the dataset, aiding in workforce planning and salary benchmarking.

**Task E: Analysis Using the Starwars Dataset**

The "starwars" dataset from the dplyr package in R was utilized to perform specific analyses:

* **Part i:** Actors whose eye color is not black and height is over 150 cm were extracted. This segmentation helps in identifying distinct groups within the dataset based on physical attributes.
* **Part ii:** A new column for Body Mass Index (BMI) was added to the dataset using the formula BMI=mass(height100)2BMI = \frac{mass}{(\frac{height}{100})^2}BMI=(100height​)2mass​. A scatter plot of height against BMI was created to visualize the relationship between these variables. This visualization aids in understanding the distribution of BMI across different heights among the actors.