# DSA-210 FINAL PROJECT REPORT

Name: Karya Sert

# Project Title: The Cost of My Screen Time – Does More Screen Time Mean More Spending?

## WHAT'S IN THIS REPORT?

This report examines the correlation between daily digital behaviors and personal spending habits. It investigates whether behavioral patterns such as screen time, time spent on shopping/social media apps, and exposure to advertisements can help predict not only the amount of money spent but also the nature of the purchase. The report is based on a self-collected dataset spanning several weeks and utilizes statistical analysis, data visualization, supervised machine learning, and regression modeling to extract insights.

## INTRODUCTION

In an era dominated by mobile screens and algorithmically targeted advertising, understanding the factors influencing spending behavior is essential. This study aims to bridge the gap between screen engagement and economic decisions by analyzing how digital activities relate to both the type and amount of personal spending.

The digitization of modern life, especially via smartphones, has led to increasingly algorithm-driven consumer behavior. Advertisements are no longer generic—they are personalized, behaviorally targeted, and delivered at moments of peak engagement.

This project seeks to model the interplay between three major behavioral patterns:

1. **Screen Time**
2. **App Usage (Shopping and Social Media)**
3. **Ad Exposure Before Purchase**

By doing so, we attempt to predict not only **how much is spent** (regression) but also **what type of item is bought**(classification).

## METHODOLOGY

### Data Collection

The dataset used in this study was self-collected on a daily basis over a designated observation period. Each row in the dataset corresponds to a single day and includes both behavioral variables (e.g., screen time and app usage) and contextual variables (e.g., ad exposure and day of the week). Data entry was conducted manually using a structured format to ensure consistency and minimize measurement error.

The primary variables recorded include:

* **Total Screen Time (minutes)**: Total time spent on the mobile device per day.
* **Time on Shopping Applications (minutes)**: Time specifically allocated to apps used for browsing or purchasing goods.
* **Time on Social Media Applications (minutes)**: Time spent on social networking platforms.
* **Money Spent (TL)**: Total expenditure for the day.
* **Purchase Type**: Categorical variable indicating the nature of the item purchased (e.g., food, beverage, shopping).
* **Ad Exposure**: Binary indicator specifying whether the user was exposed to an online advertisement prior to purchase.
* **Time of Purchase**: Categorical variable capturing the general time window of the transaction (morning, afternoon, evening).

### Data Preprocessing

The dataset was subjected to standard preprocessing steps to enhance data quality and prepare for model input. These steps included:

* **Missing Value Handling**: All rows containing missing values in critical input or target columns were removed using listwise deletion.
* **Feature Engineering**:
  + Spending Rate was calculated by dividing daily expenditure by total screen time (TL/min).
  + Weekend Indicator was derived from date information to differentiate between weekdays and weekends.
  + Log-Transformed Spending was computed as the natural logarithm of daily expenditure to normalize the distribution for regression modeling.
* **Encoding**:
  + The Purchase Type variable was transformed into numerical labels using scikit-learn’s LabelEncoder.
  + The Time of Purchase feature was one-hot encoded to create binary indicators for morning, afternoon, and evening.

### Model Preparation

After cleaning and engineering features, the dataset was partitioned into training and testing subsets. An 80/20 stratified split was applied to ensure proportional representation of purchase types in both subsets. Features were scaled using StandardScaler for models sensitive to feature magnitudes.

### Machine Learning Models

Two types of supervised learning models were implemented to explore both classification and regression tasks:

* **Classification Task**:
  + Goal: Predict the categorical outcome (Purchase Type) based on behavioral features.
  + Models Used: Logistic Regression, Random Forest Classifier, Gradient Boosting Classifier.
  + Evaluation Metrics: Accuracy, macro-averaged Precision, Recall, and F1 Score. Confusion matrices were visualized for model validation.
* **Regression Task**:
  + Goal: Predict the continuous outcome (Money Spent), using both original and log-transformed targets.
  + Models Used: Linear Regression (baseline), Random Forest Regressor.
  + Evaluation Metrics: R-squared (R²), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE). Actual vs. predicted plots were included to visually assess model performance.

## KEY FINDINGS

- Higher overall screen time, especially on shopping apps, correlates with higher spending.  
- Ad exposure is linked with both increased spending and a higher chance of a shopping-related purchase.  
- Spending is generally higher on weekends.  
- Social media usage has a moderate positive correlation with money spent.

## MACHINE LEARNING: CLASSIFICATION OF PURCHASE TYPE

The classification task aimed to predict the **type of purchase**—categorized as food, beverage, or shopping—based on digital behavior features including app usage, advertisement exposure, and time-related factors.

### Algorithms Implemented

* Logistic Regression
* Random Forest Classifier
* Gradient Boosting Classifier

### Evaluation Metrics

Model performance was assessed using:

* **Accuracy**
* **Macro-Averaged Precision, Recall, and F1 Score**
* **Confusion Matrix Visualization**

### Performance Overview

* **Random Forest Classifier** produced stable results and allowed for interpretation of feature importances.
* **Logistic Regression**, while less complex, offered insight into linear separability of categories.

Feature importance analysis showed that **Advertisement Exposure** consistently ranked among the top predictors, while **Time on Social Media** had limited predictive power relative to Shopping App usage.

## MACHINE LEARNING: REGRESSION ON SPENDING AMOUNT

This regression task aimed to estimate the **daily amount of money spent** using behavioral features such as screen time, ad exposure, and log-transformed expenditure data.

### Models Implemented

* Linear Regression
* Random Forest Regressor

### Evaluation Criteria

The models were assessed based on:

* **R² Score (Coefficient of Determination)**
* **Mean Absolute Error (MAE)**
* **Root Mean Squared Error (RMSE)**

### Model Results

| **Model** | **R²** | **MAE** | **RMSE** |
| --- | --- | --- | --- |
| Linear Regression | 0.87 | ~12.3 | ~16.2 |
| Random Forest | **0.94** | **~7.8** | **~10.6** |
|  |  |  |  |

The **Random Forest Regressor** significantly outperformed the linear baseline, demonstrating its capacity to capture non-linear relationships and feature interactions. The log transformation of the target variable contributed to greater model stability and predictive accuracy.

Scatter plots comparing predicted vs. actual values indicated tighter clustering around the diagonal line for the Random Forest model, reinforcing its effectiveness.

## Insights and Interpretation

The machine learning analysis provides several actionable insights:

* **Advertisement Exposure** serves as a strong predictor of shopping-related purchases and higher expenditure.
* **Shopping App Screen Time** is a more influential factor than general social media usage in determining spending behavior.
* **Ensemble Models** such as Random Forest and Gradient Boosting consistently outperformed linear models in both classification and regression contexts.
* **Temporal Factors** (e.g., weekends) showed a clear influence on both purchase frequency and spending volume, aligning with established behavioral economics literature.

## LIMITATIONS

While the study yielded compelling results, several limitations must be acknowledged:

* **Single-Participant Dataset**: The data reflects the behavior of one individual, limiting generalizability.
* **Manual Logging of Ad Exposure**: This introduces potential recall bias and variability in data reliability.
* **Short Observation Period**: A longer timeframe may reveal seasonality or more stable trends.

## FINAL THOUGHTS

Our machine learning results support the hypothesis that digital behavior—especially shopping app usage and ad exposure—is a strong predictor of purchasing type and amount. The classification models were particularly successful in categorizing user behavior based on app engagement patterns.

The regression task was more challenging, indicating the complexity of financial decision-making, which may require more contextual or emotional features for better performance.

This study demonstrates that machine learning models, when combined with thoughtful feature engineering and behavioral data, can provide meaningful insights into consumer spending. While classification models yielded robust performance, future work could focus on improving regression accuracy by incorporating external datasets (e.g., income levels or sentiment analysis from social media).

This analysis holds value for marketers, app developers, and researchers in behavioral economics aiming to predict and influence spending behavior through digital engagement patterns.