**Project Report: Advertising Data Analysis**

**Introduction**

This project focuses on analyzing advertising data to evaluate and predict the Return on Investment (ROI) based on metrics such as impressions, likes, and runtime. Additionally, it determines the percentage increase in performance over time. The analysis combines statistical techniques and machine learning to derive actionable insights.

**Objective**

The main objectives of this project are:

1. To calculate the ROI based on advertising metrics.
2. To analyze and visualize the trends in impressions, likes, and runtime.
3. To determine the percentage increase in performance and identify key factors driving improvements.

**Tools and Libraries**

The following libraries were used for data analysis and visualization:

* **NumPy**: For numerical computations.
* **Pandas**: For data manipulation and cleaning.
* **Matplotlib & Seaborn**: For static data visualization.
* **Plotly**: For interactive visualizations.
* **Scikit-learn**: For preprocessing, training, and evaluation of machine learning models.

**Methodology**

**1. Data Collection**

The dataset was imported using Pandas, ensuring a structured and tabular format suitable for analysis. Data cleaning and preprocessing steps included handling missing values, ensuring consistent formats, and scaling numerical data for better model performance.

**2. Data Preprocessing**

* **Feature Scaling**: Min-Max Scaling was applied to normalize the data.
* **Data Splitting**: The dataset was split into training and testing sets using an 80-20 ratio to evaluate the model's performance effectively.

**3. Exploratory Data Analysis (EDA)**

Key visualizations were created to understand the dataset:

* Distribution plots of impressions, likes, and runtime.
* Correlation heatmap to identify relationships between variables.
* Time-series analysis to observe trends and changes over time.

**4. Modeling**

A linear regression model was implemented to predict ROI based on the input features. The steps included:

* Training the model on the training set.
* Evaluating the model's performance using Mean Squared Error (MSE).
* Tuning the model to improve accuracy.

**5. Visualization**

Interactive visualizations were created using Plotly to:

* Highlight trends in impressions, likes, and runtime.
* Demonstrate the relationship between features and ROI.

**6. ROI and Performance Increase Calculation**

The ROI was calculated based on the predicted output from the model. The percentage increase in performance was determined by comparing baseline metrics to updated values, providing insights into the effectiveness of the advertising campaigns.

**Results**

1. The multiple linear regression model achieved an acceptable accuracy level with minimal error in predictions.
2. Key insights from the data:
   * **Impressions and likes** had a strong positive correlation with ROI.
   * **Runtime** played a critical role in determining the effectiveness of campaigns.
3. Visualizations showed consistent improvement in ROI over time, aligning with increased impressions and likes.

**Conclusion**

This project successfully demonstrated how data analysis and machine learning can evaluate and predict advertising performance. The findings provide valuable insights for optimizing future campaigns and maximizing ROI.

**Future Work**

1. Incorporate additional features such as demographics and ad placement.
2. Explore advanced machine learning models like Random Forest or Gradient Boosting for better prediction accuracy.
3. Automate the analysis process to provide real-time insights into campaign performance.