



MOBILE USERS BEHAVIOR DATA MANAGEMENT AND ANALYSIS

BY SOAD ATEF AND HAGER AHMED



www.reallygreatsite.com





INTRODUCTION

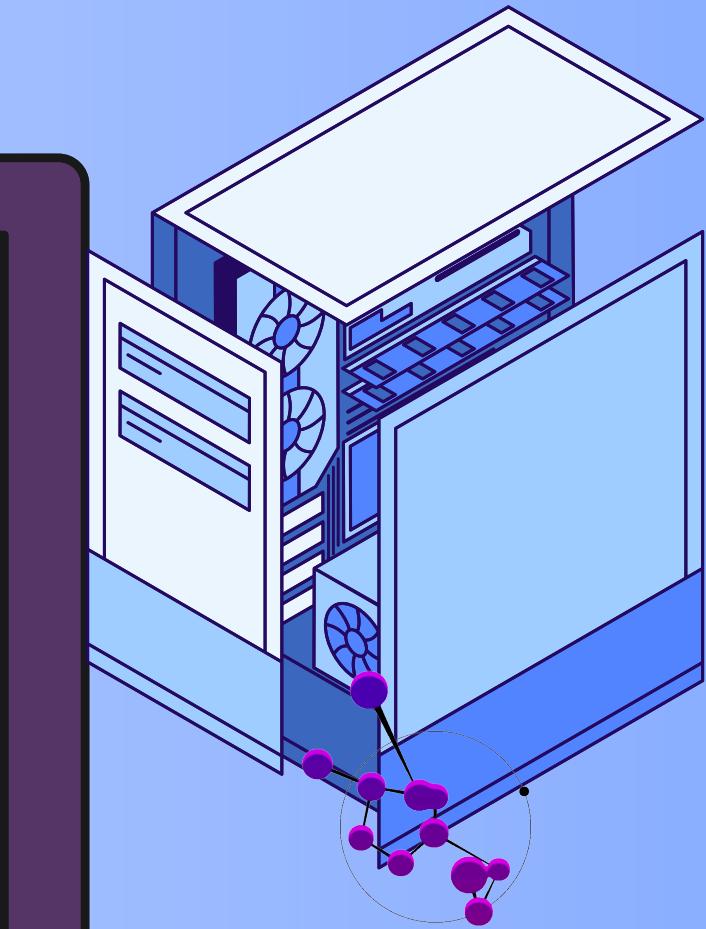
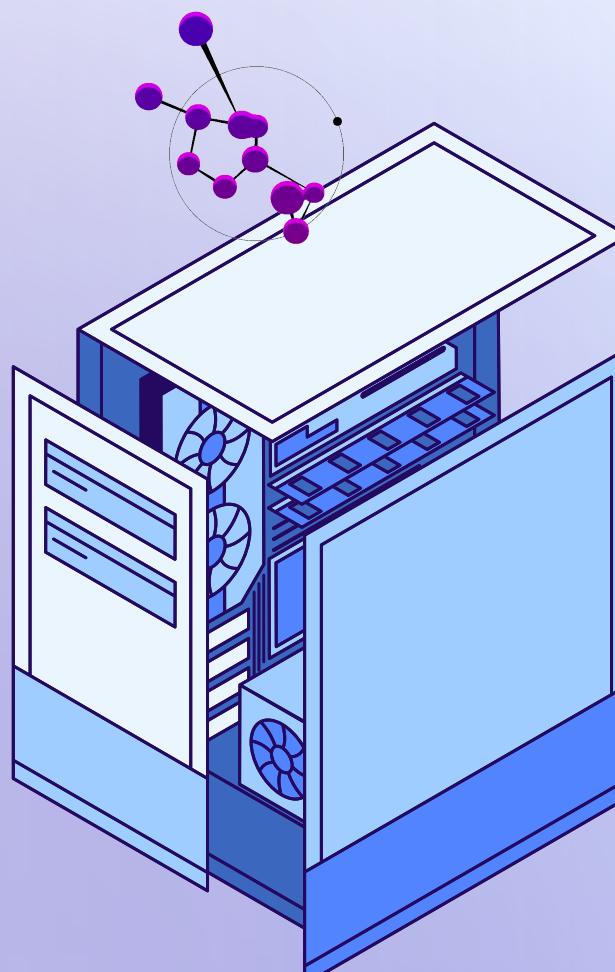
-

What is the project about?

- The project revolves around managing, cleaning, and analyzing mobile users' behavioral data. It combines SQL, SSIS (SQL Server Integration Services), and Python to handle data efficiently.

What tools were used?

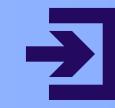
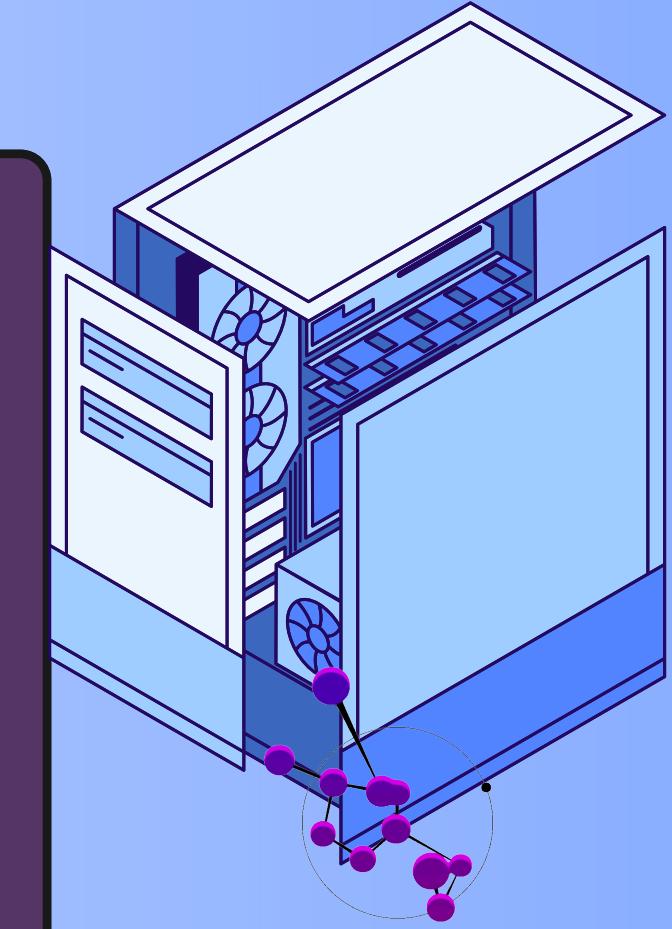
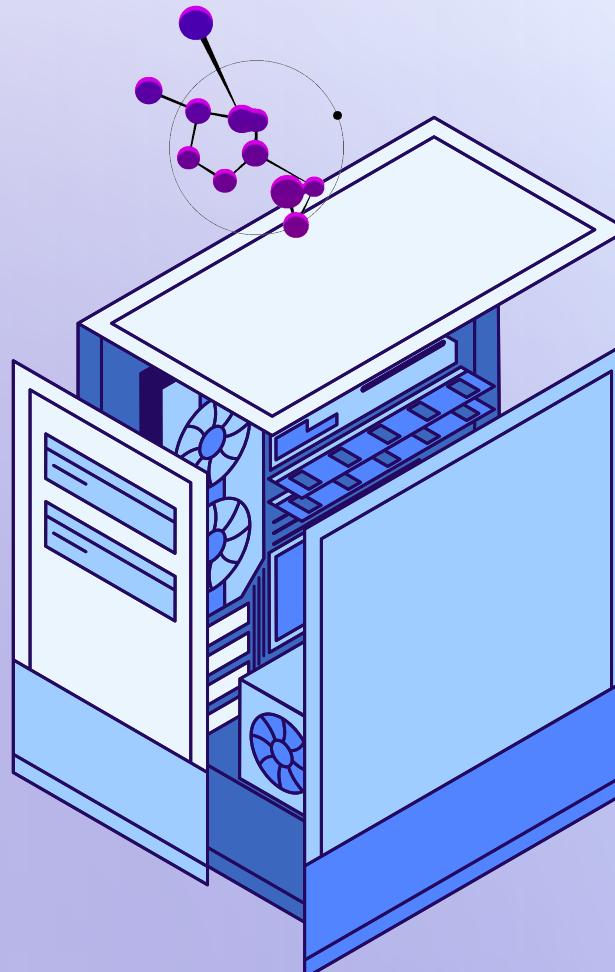
- SQL for database setup and management.
- SSIS for data extraction, transformation, and loading (ETL).
- Python for data analysis and model building.





INTRODUCTION

- Mobile Users Data Management and Analysis focuses on creating a system that allows for structured data management and insightful analysis of mobile user behaviors.
- We will utilize SQL for database management, SQL Server Integration Services (ssis) for ETL processes, and Python for deeper analysis.





KEY CONCEPTS



- SQL Database to manage user data efficiently.
- SSIS for loading and transforming large datasets.
- Python for further data cleaning and advanced analysis.





GOAL AND IMPORTANCE

- **Goal:** The goal of this project is to collect, clean, analyze, and store mobile user behavior data to gain insights that help improve services, marketing strategies, and user satisfaction.
- **Importance:**
 - Informed decision-making based on data.
 - Insights into user trends, preferences, and patterns.
 - Support for product improvement and customer satisfaction strategies.
 - Data-driven decisions are critical for competitive business strategies.
 - The ability to handle large datasets is essential for mobile user behavior analysis.
 - Combining SQL, SSIS, and Python provides an efficient solution for comprehensive data management and analysis



DATA OVERVIEW



- **Dataset:** Mobile Users Behavior Dataset
- **Source:** The dataset was sourced from user interactions across various mobile platforms.
- **Number of Records:** Over 701 user behavior records.
- **Main Features:**
 - User ID
 - Device Model
 - App Usage
 - Operating_System
 - Device ID
 - App_Usage_Time
 - Screen_On_Time
 - Battery_Drain
 - Number_of_Apps_Installed
 - Data_Usage
 - Age
 - Gender
 - User_Behavior_Class
 - OS_ID



DATA CLEANING

Handled Missing Data:

- Identified and managed missing values in columns such as app usage and interaction type. Methods like mean imputation and removal of records were used.

Removed Duplicates:

- Ensured data integrity by identifying and removing duplicate records.





FEATURE SELECTION

- **Why Feature Selection?** To remove irrelevant data and focus on features that would significantly impact the analysis.
- **Key Features Selected:**
 - App Usage Time
 - Battery Drain
 - Data Usage





MODEL SELECTION AND TRAINING

- **What kind of models?** Machine learning models were developed using Python for predictive analytics. **Models tested included:**
- Logistic Regression for user retention prediction.
- Linear Regression
- Lasso
- Ridge
- **Training Process:** The dataset was split into training and test sets. The models were trained using the training data and validated against the test data.





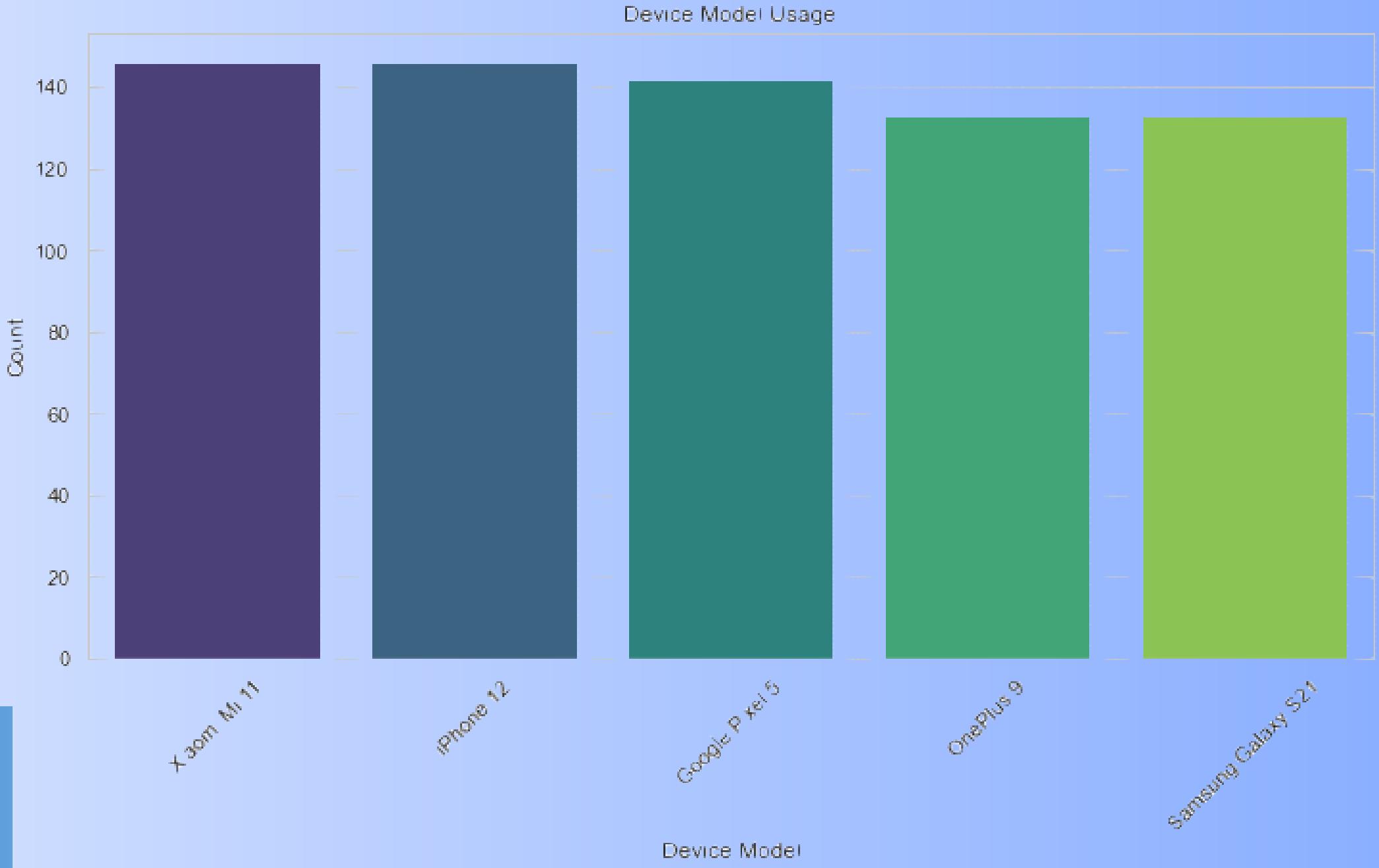
DATA VISUALIZATION

- **Why Visualization?** To gain better insights from the data through graphical representation.
- **Visualizations Created:**
- Bar Charts
- HeatMap
- Box Plot
- Histogram
- Line Graph





DATA VISUALIZATION



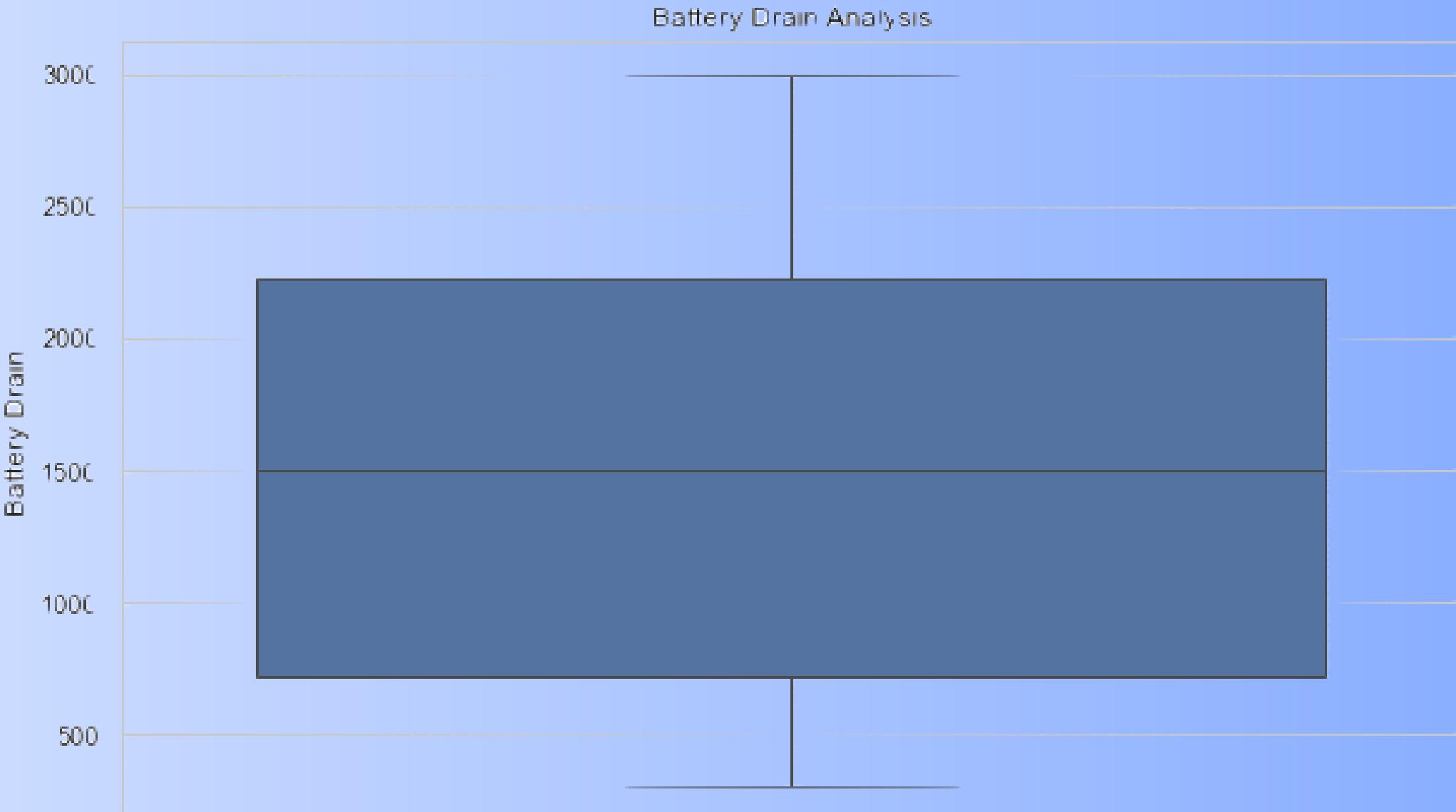
Device Model Usage

Type: Bar Chart

Purpose: Display the number of users for each device model to identify popular devices among users



DATA VISUALIZATION



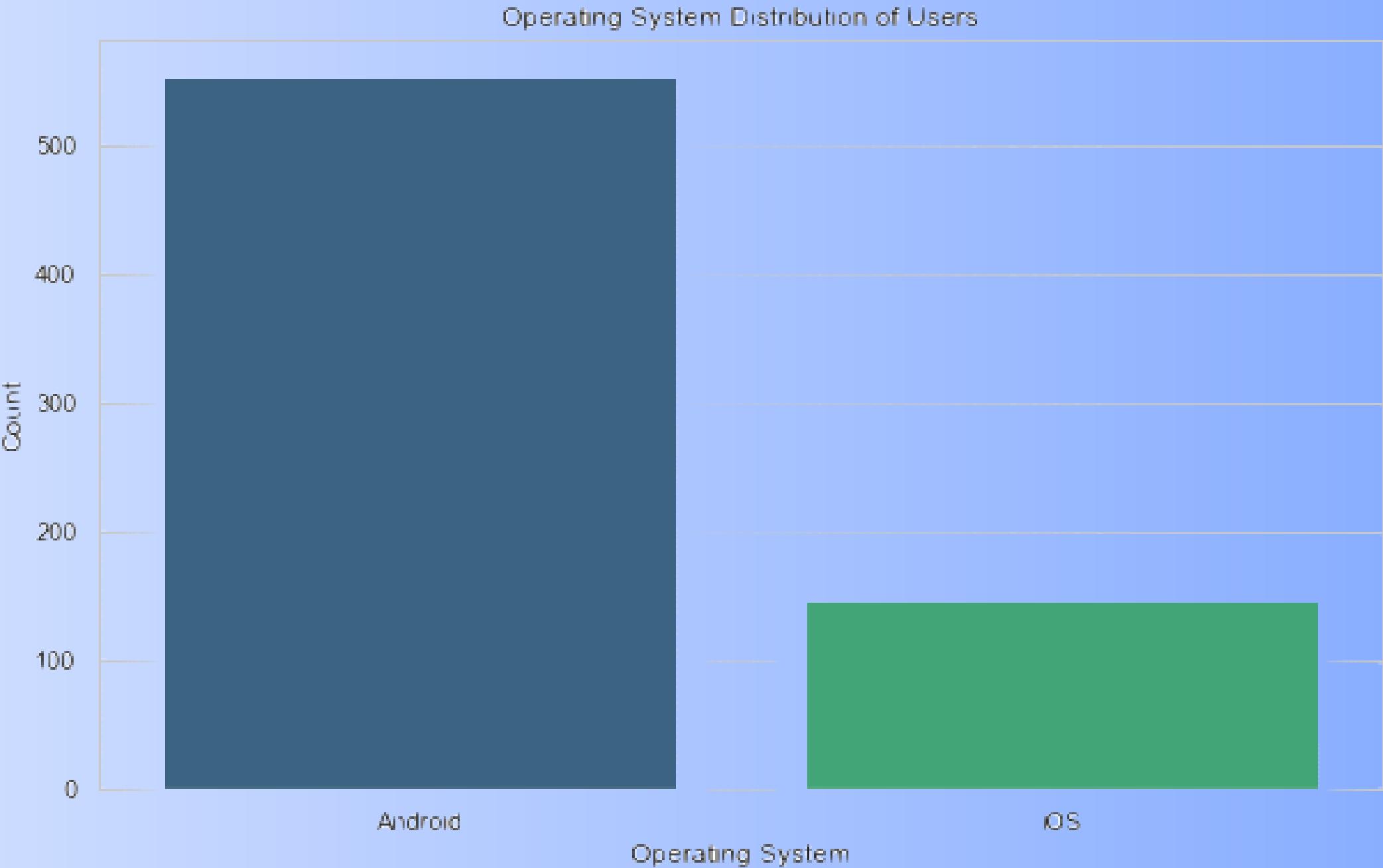
Battery Drain Analysis

Type: Box Plot

Purpose: Show the distribution of battery drain values to identify outliers and the general trend of battery usage among users.



DATA VISUALIZATION



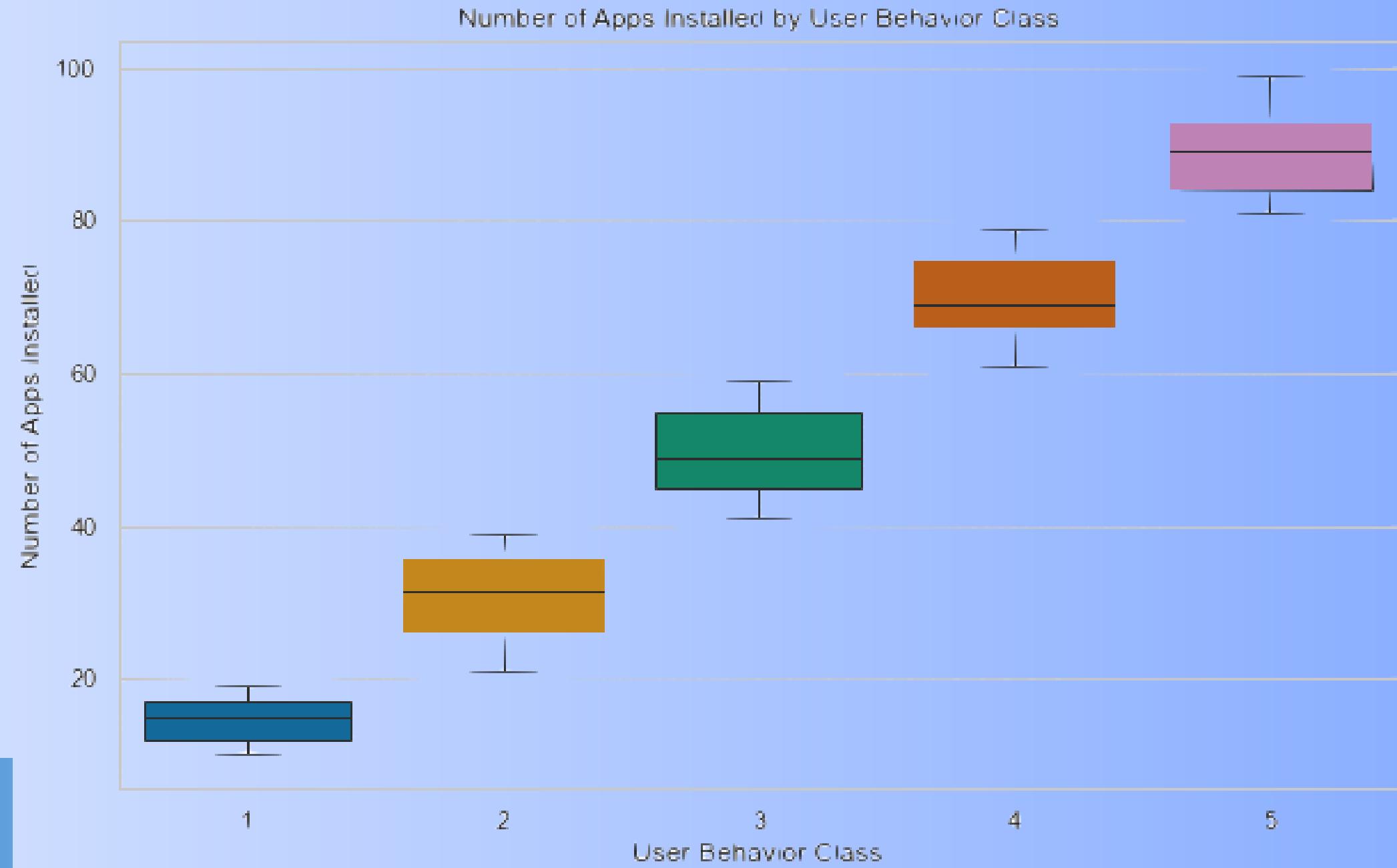
Operating System Distribution

Type: Bar Chart

Purpose: Illustrate the distribution of users across different operating systems (e.g., iOS, Android).



DATA VISUALIZATION



Number of Apps Installed vs. User Behavior Class

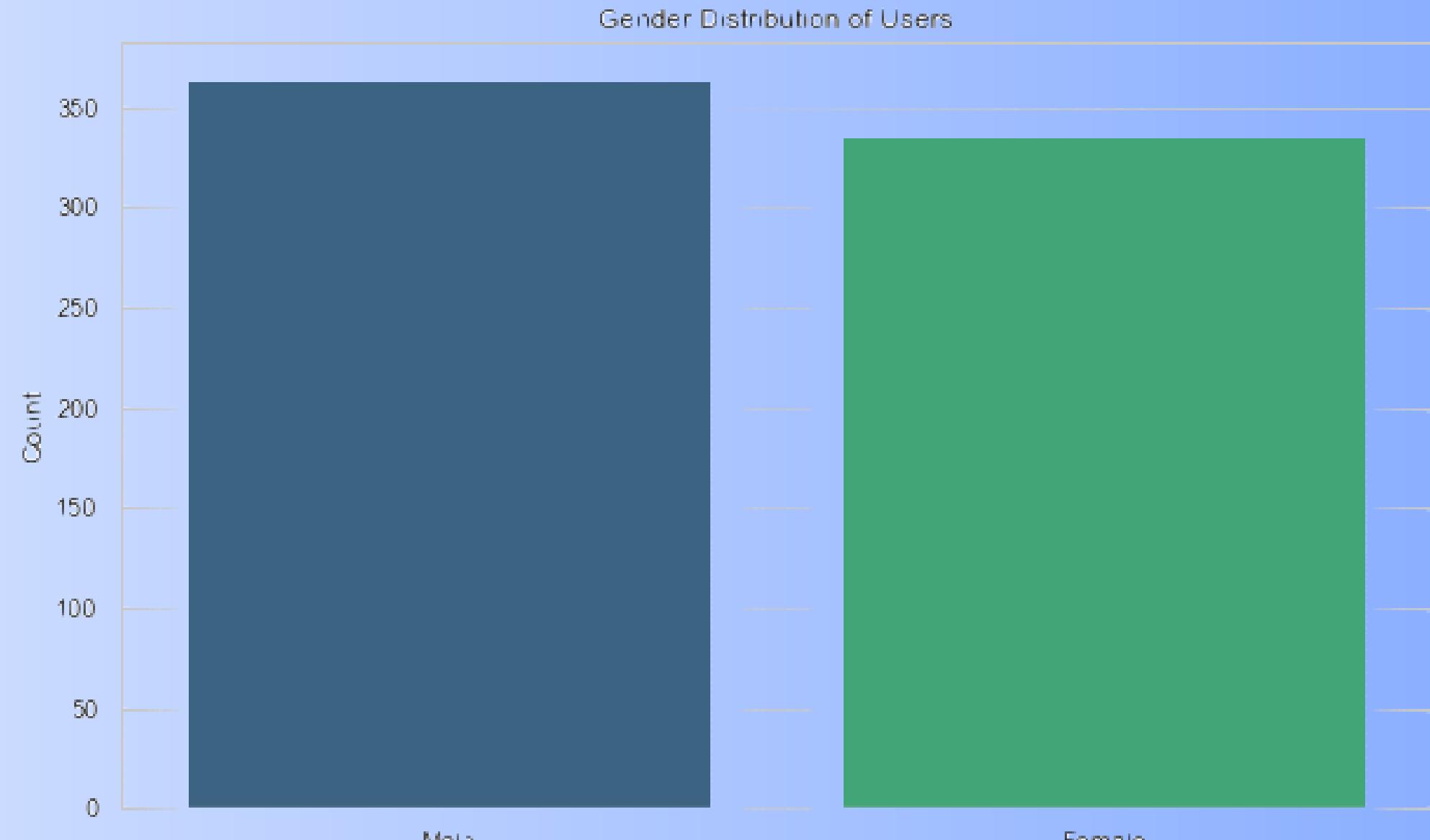
Type: Box Plot

Purpose: Compare the number of apps installed across different user behavior classes to identify any trends.



رواد مصر الرقمية

DATA VISUALIZATION



Gender Distribution

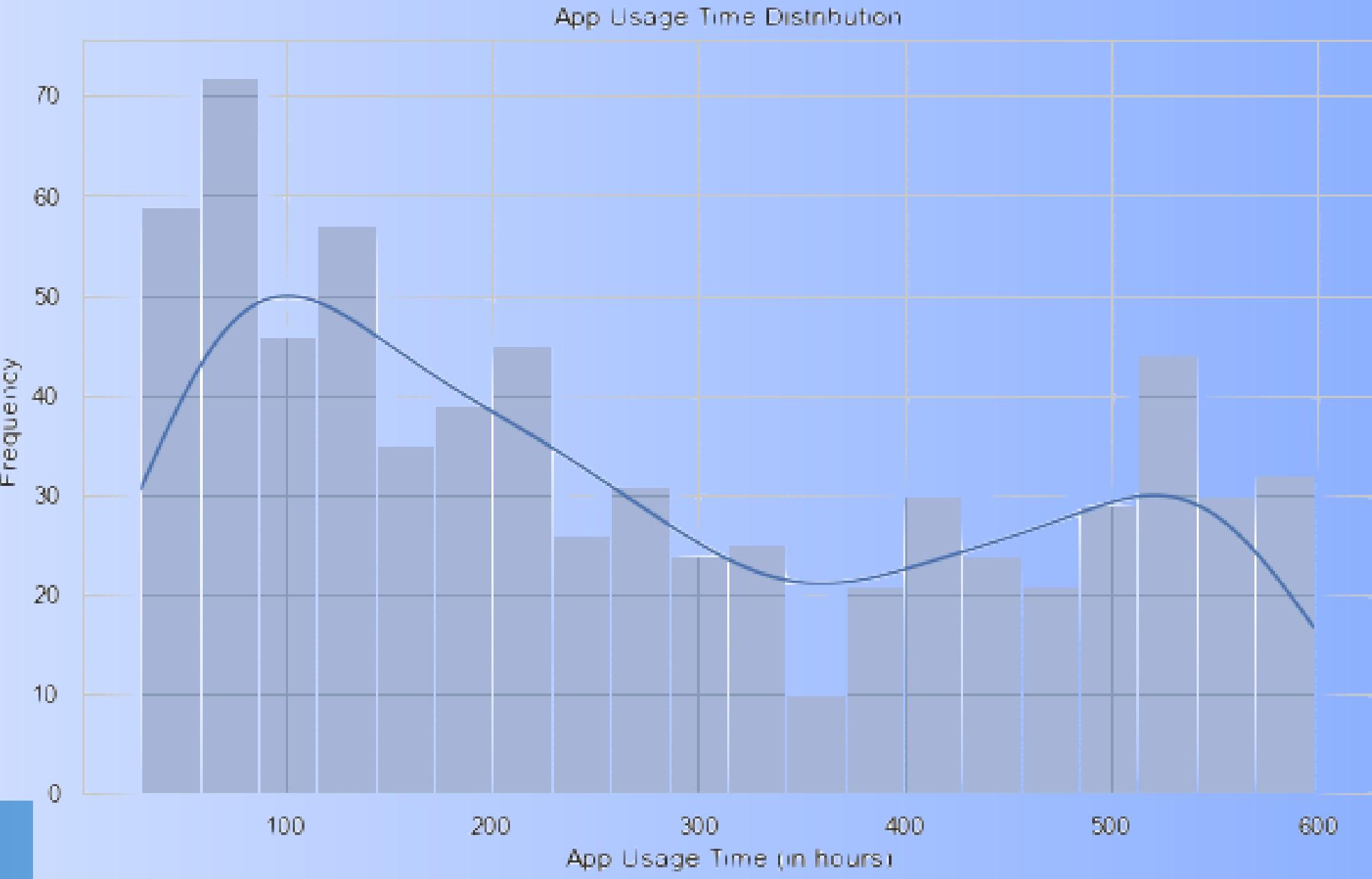
Type: Bar Chart

Purpose: Visualize the proportion of male and female users to assess gender representation in the dataset.



DATA

VISUALIZATION



App Usage Time distribution

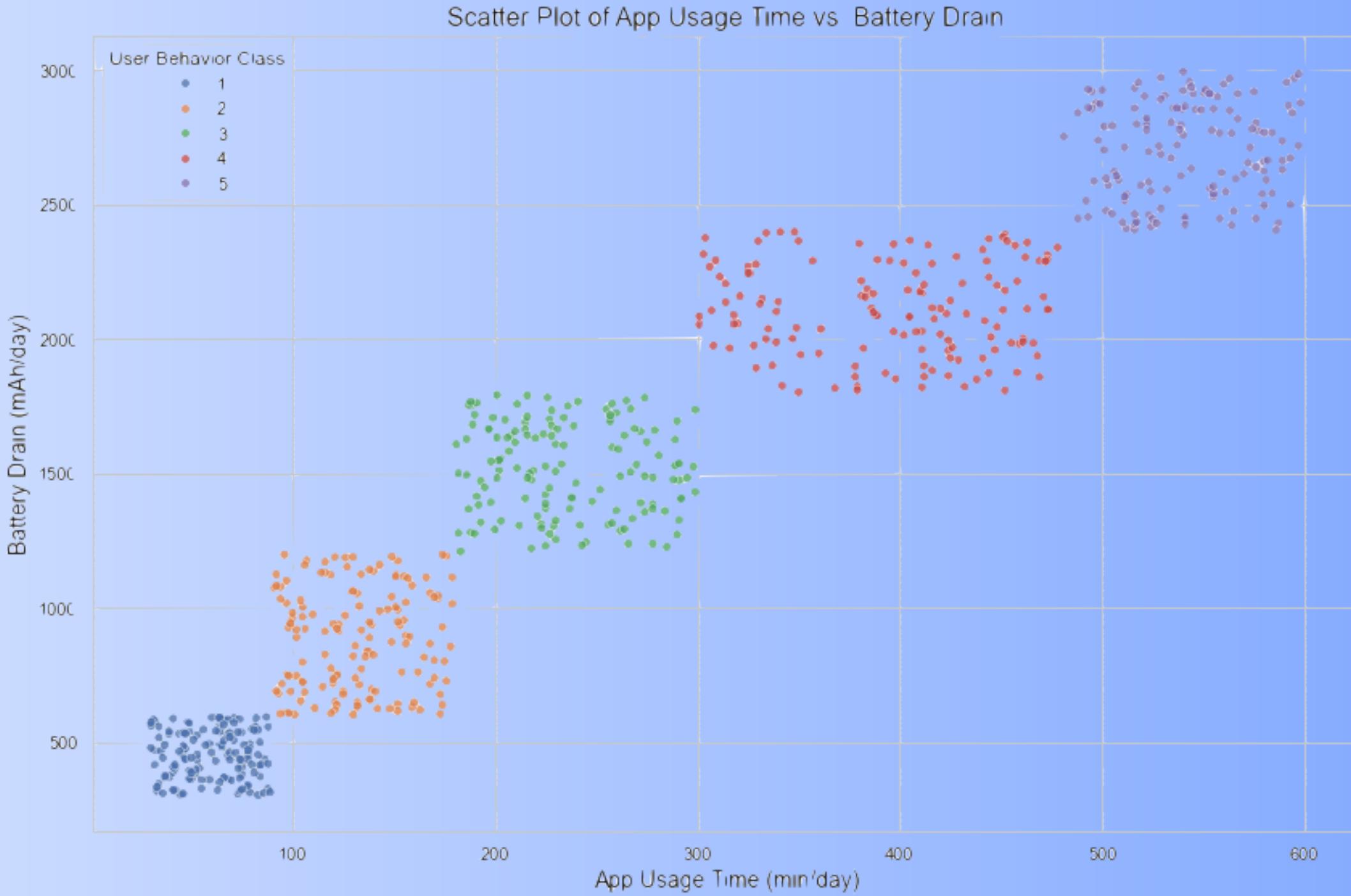
Type: Histogram

Purpose: Analyze the distribution of app usage time to see how much time users spend on their devices.



DATA

VISUALIZATION



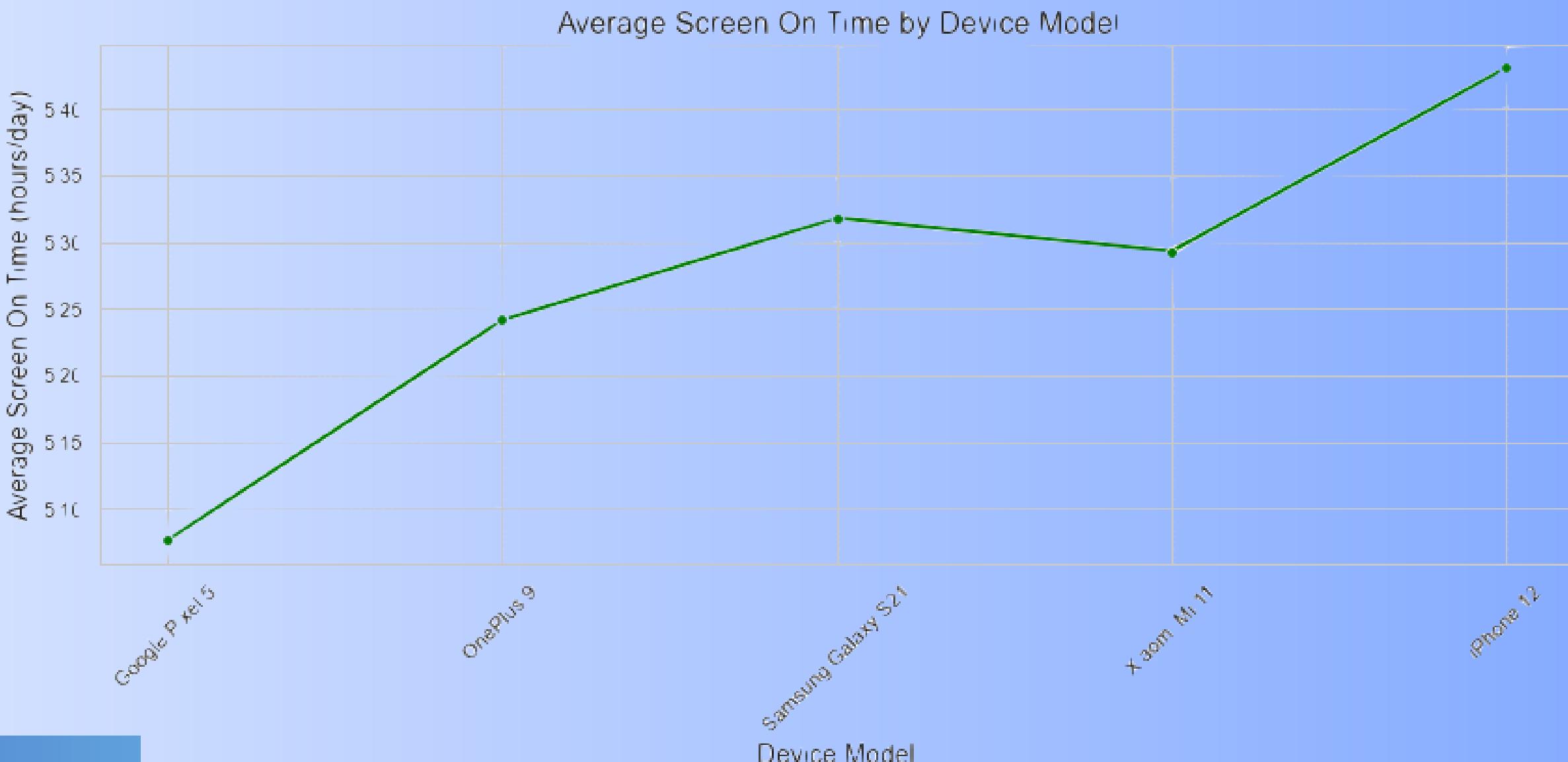
App Usage Time VS Battery Drain

Type: Scatter Plot

Purpose: Analyze the distribution of app usage time to see how much battery is drained



DATA VISUALIZATION



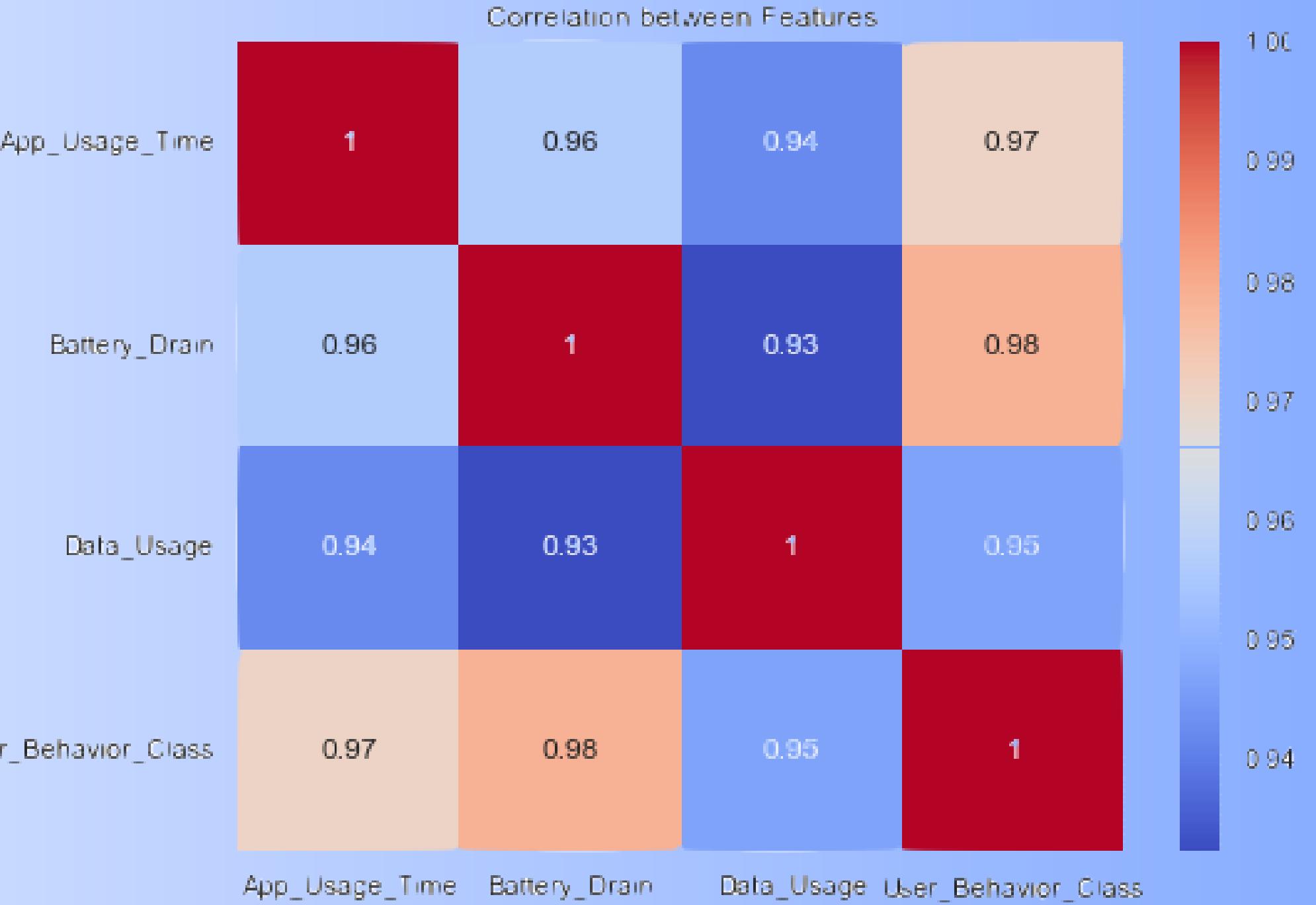
Average Screen On Time by Device Model

Type: Line Graph

Purpose: Analyze the average screen time



DATA VISUALIZATION



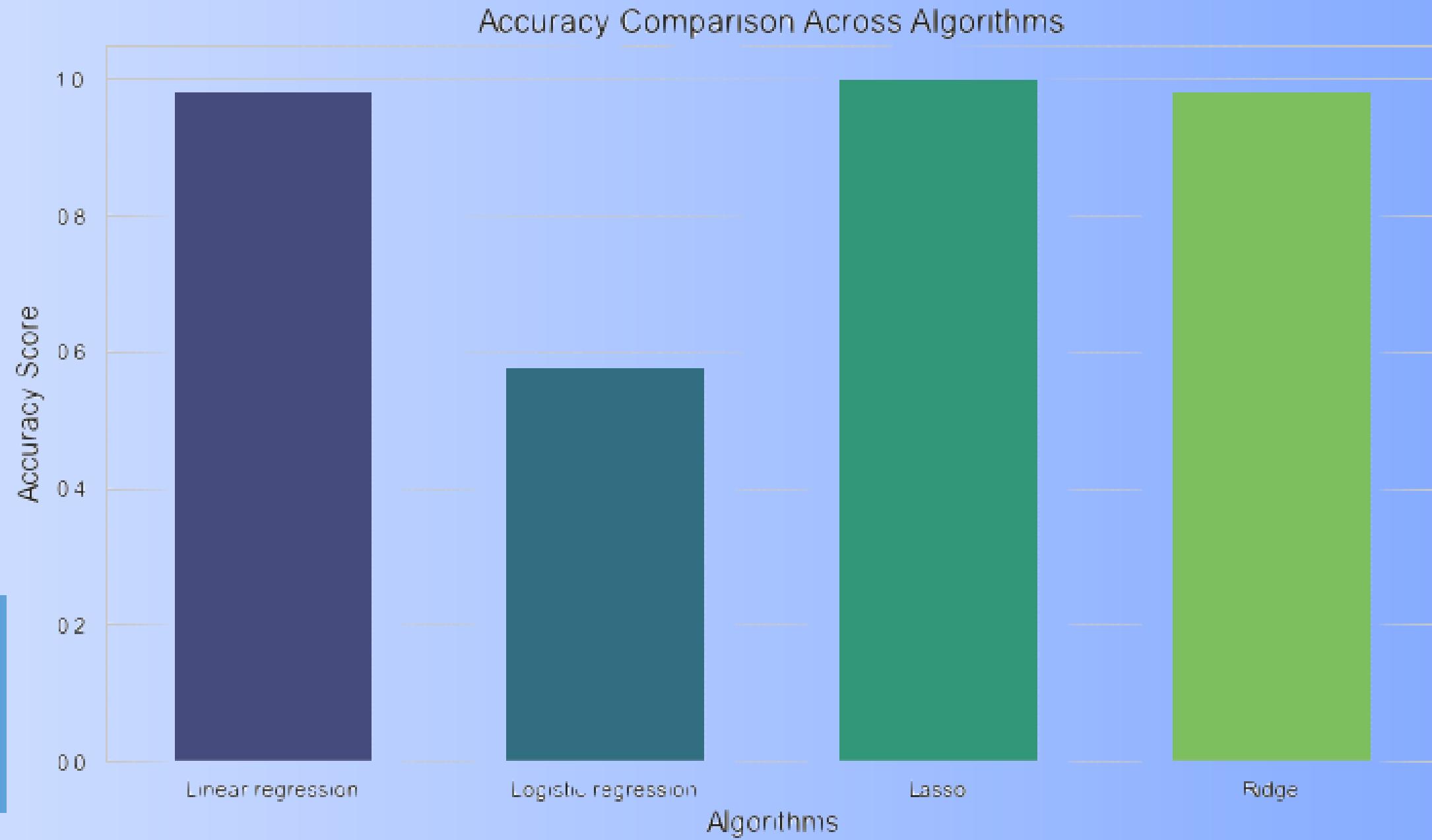
Correlation Between Features

Type: Table



رواد مصر الرقمية

DATA VISUALIZATION



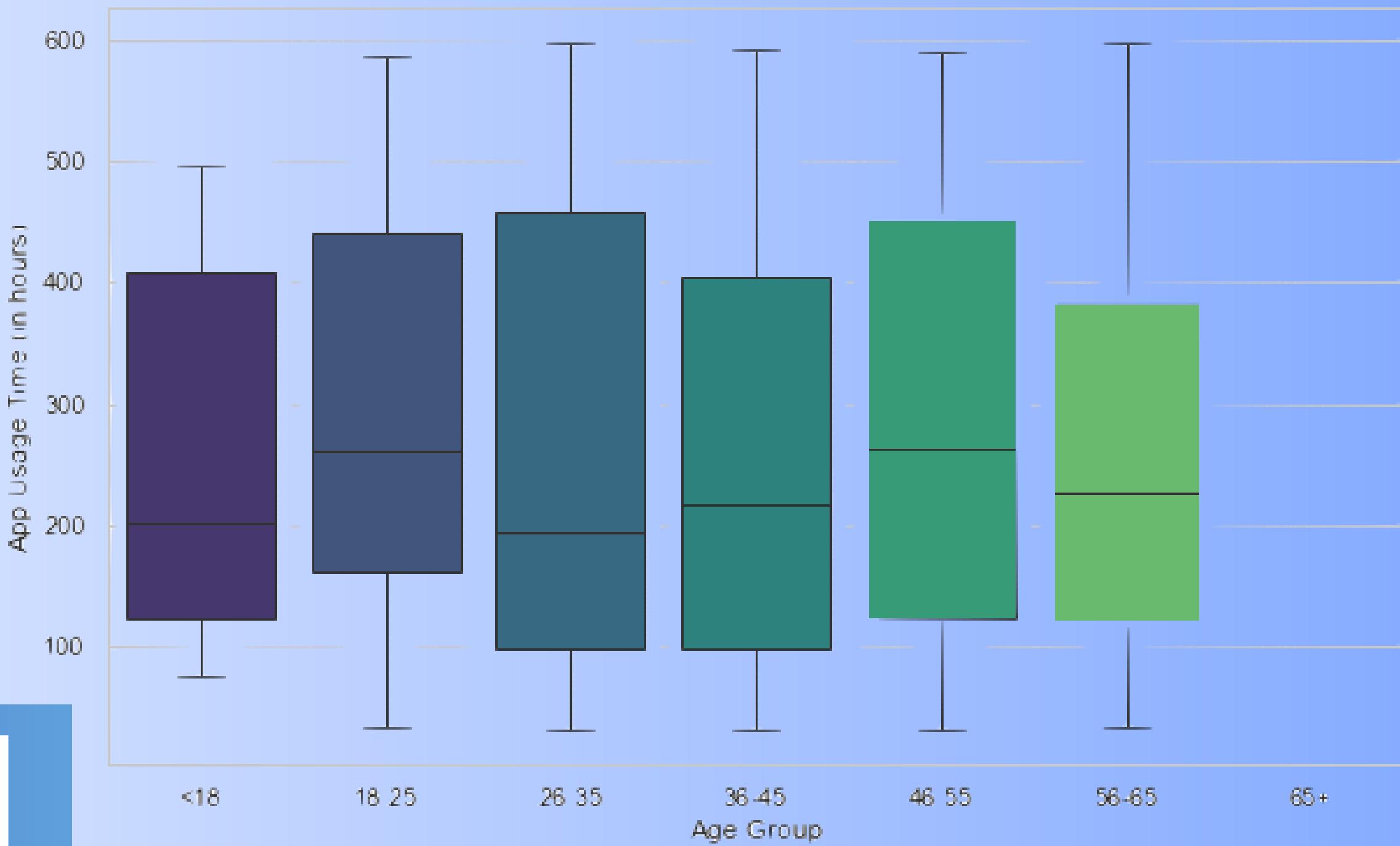
Accuracy Comparison Across Algorithms
Type: Bar Graph



DATA VISUALIZATION



App Usage Time by Age Group



Accuracy Comparison Across Algorithms

Type: Bar Graph



RESULTS

Insights Gained:

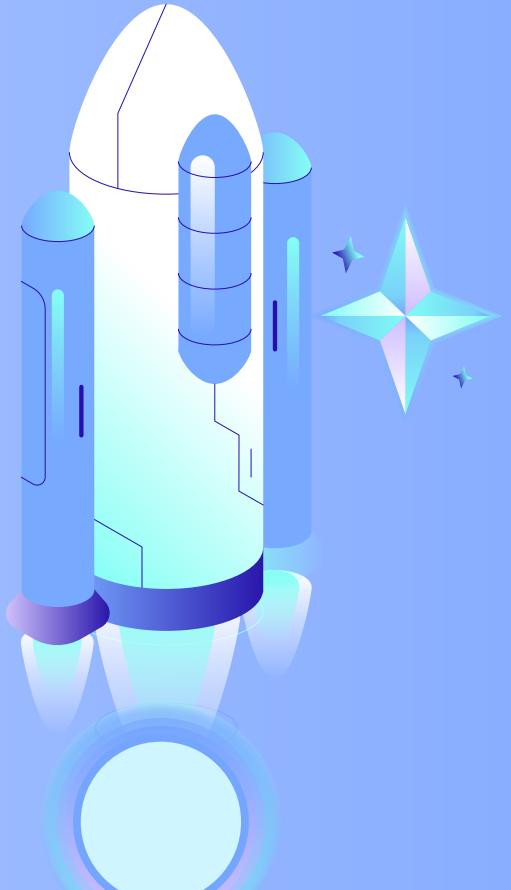
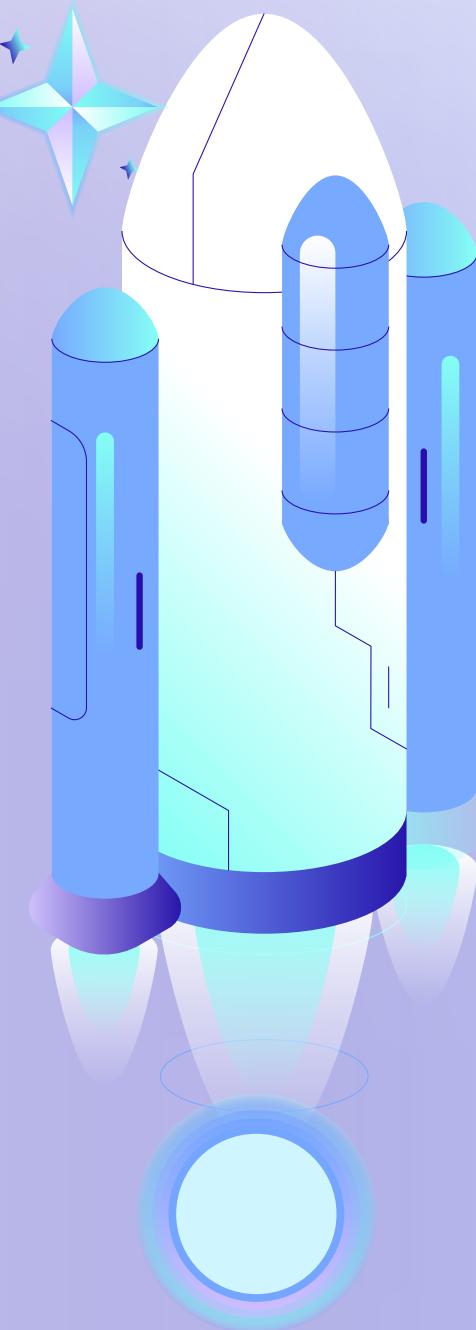
The most used apps are entertainment-based and are used by men with the age between 26-35

Device type significantly influences app usage patterns.
Peak usage times are between 6 PM and 9 PM.

Model Performance:

Lasso achieved 99% accuracy in predicting user retention.
Logistic Regression had an accuracy of 57% in classifying app behavior





CONCLUSION

Key Takeaways:

- Data management and cleaning are crucial steps that determine the quality of analysis.
- Machine learning models provide actionable insights into user behavior.
- Visualizations are an effective tool to communicate findings.

Future Work: Implement additional models to predict user behavior more accurately and optimize the database for faster processing.





THANK YOU!



FROM: HAGER AHMED AND SOAD ATEF