KIET
GROUP OF INSTITUTIONS
Connecting Life with Learning

INTERNET USAGE REPORT

REPORT

Student Name : ABHISHEK TRIPATHI

Roll Number : 202401100300010

Course : AI

Date : 22/04/2025

INTRODUCTION

In this analysis, we aim to segment internet users based on their behavioral data. By understanding usage patterns, such as the number of hours spent online daily, the variety of site categories visited, and the frequency of sessions per day, we can group users into distinct categories. This segmentation can be valuable for targeted marketing, user experience improvements, and product development strategies.

Methodology

Data Overview

The dataset comprises 100 entries with the following features:

- **daily_usage_hours: Average hours spent online per day (float).**
- **site_categories_visited: Number of distinct categories of websites visited (integer).**
- **sessions_per_day: Number of browsing sessions initiated daily (integer).**

Data Preprocessing

1. **Standardization: All features were standardized using StandardScaler from scikit-learn to bring them to the same scale.**

Clustering Technique

2. **K-Means Clustering:**
    - **We used the Elbow Method to determine the optimal number of clusters.**

- **Based on the elbow plot, 3 clusters were selected.**
- **KMeans clustering was then applied with n_clusters=3.**

Visualization

3. **Scatter Matrix:**
    - **A scatter matrix was used to visualize the distribution of users in each cluster across all three features.**

3. Code

Here is the Python code used to process the data and generate the visualizations.

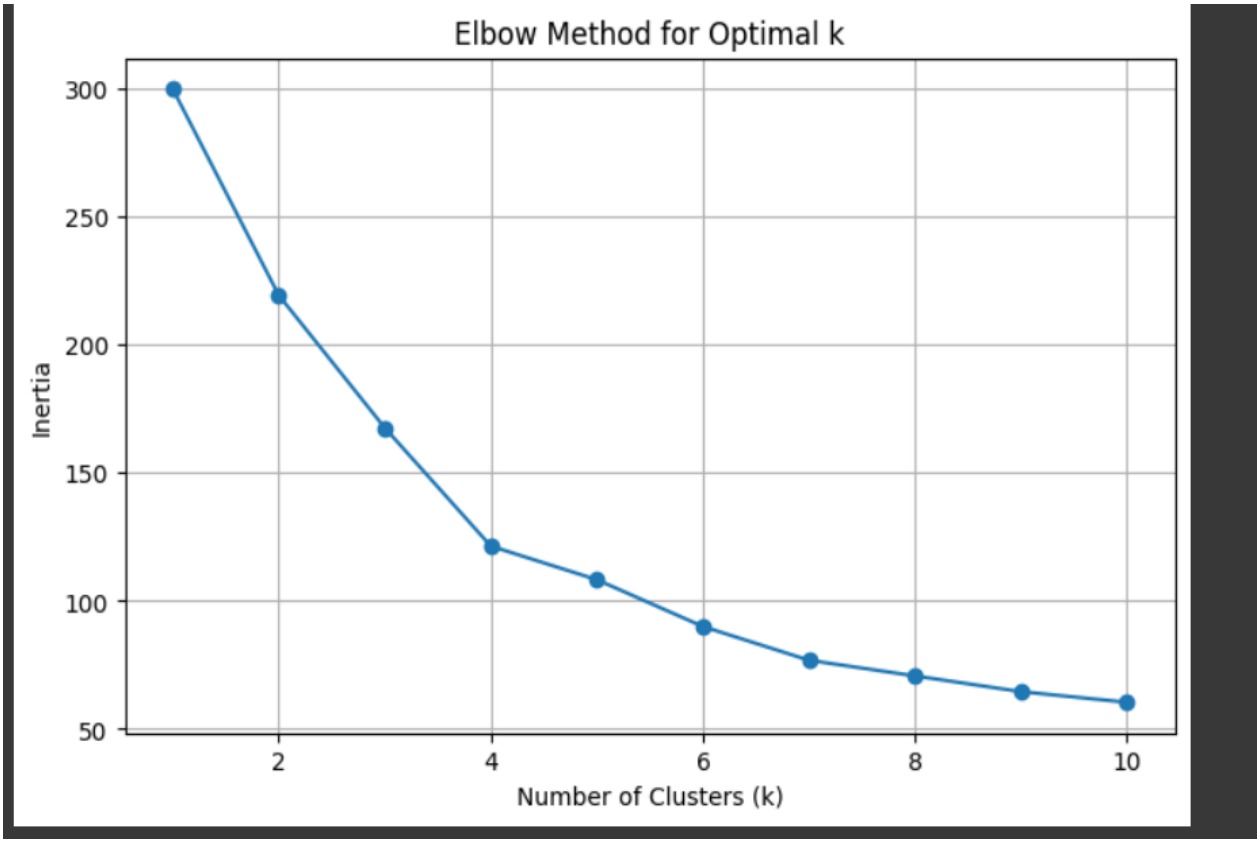
```
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix
# Load and preprocess data
df = pd.read_csv('internet_usage.csv')
scaler = StandardScaler()
X_scaled = scaler.fit_transform(df)
# Determine optimal clusters using Elbow Method
inertia = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(X_scaled)
```

```
    inertia.append(kmeans.inertia_)
plt.plot(range(1, 11), inertia, marker='o')
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('Inertia')
plt.show()
# Apply KMeans with optimal k=3
kmeans = KMeans(n_clusters=3,
random_state=42)
df['cluster'] = kmeans.fit_predict(X_scaled)
# Visualize clusters
features = ['daily_usage_hours',
'site_categories_visited', 'sessions_per_day']
color_map = df['cluster'].map({0: 'red', 1: 'green',
2: 'blue'})
scatter_matrix(df[features], figsize=(10, 8),
diagonal='hist', color=color_map)
plt.suptitle('Cluster Visualization of Internet
Users', y=1.02)
plt.show()
```

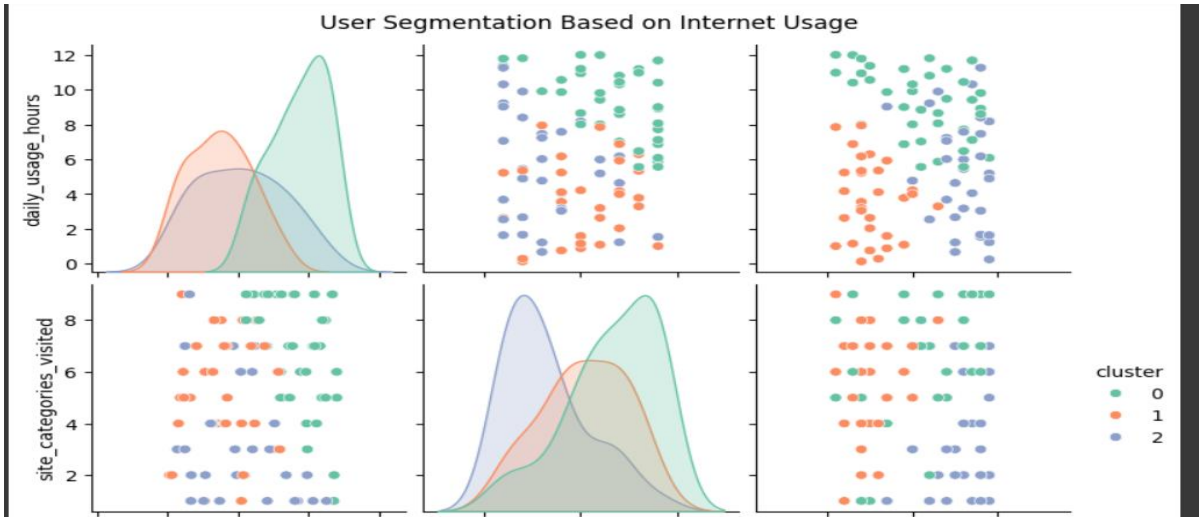
Results and Interpretation

The K-Means clustering algorithm segmented users into three distinct clusters:

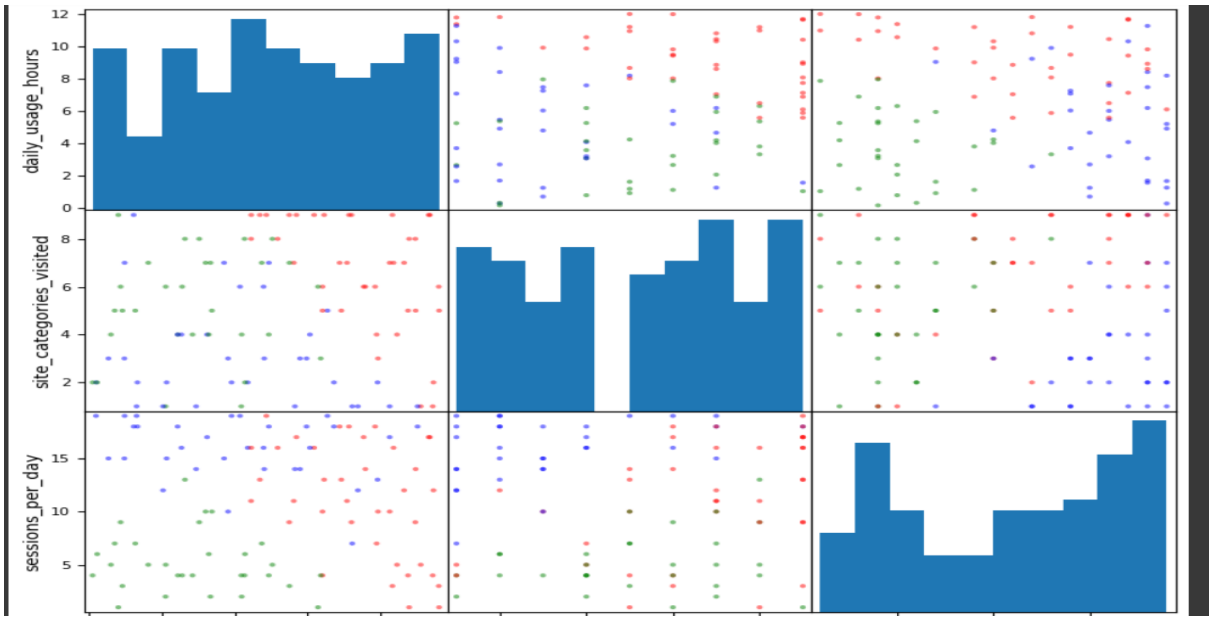
- **Cluster 0 (e.g., red): High daily usage and frequent sessions, likely representing power users.**



- **Cluster 1 (e.g., green): Moderate usage across all features, possibly regular users.**



Cluster 2 (e.g., blue): Lower usage, infrequent sessions—indicative of casual or occasional users.



This segmentation can help tailor digital experiences, services, or advertisements specific to each group.

Conclusion

Through clustering analysis, we've successfully grouped internet users based on their usage patterns. These insights can support data-driven decision-making for customer engagement strategies. Further refinements, like integrating demographic or contextual data, could yield even more actionable segments.