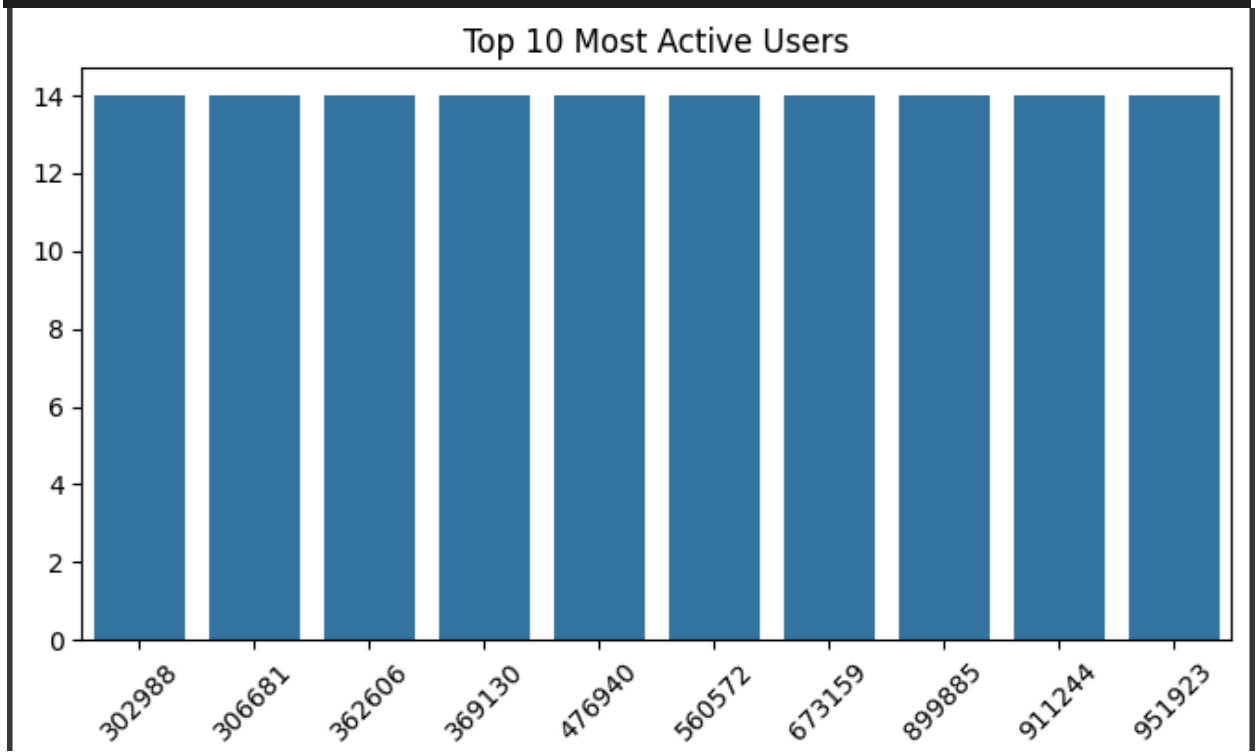
Visualization Report – Automatic Zoom Recommendation System

1. Top 10 Most Active Users

This chart highlights the users with the highest number of interactions in the dataset.

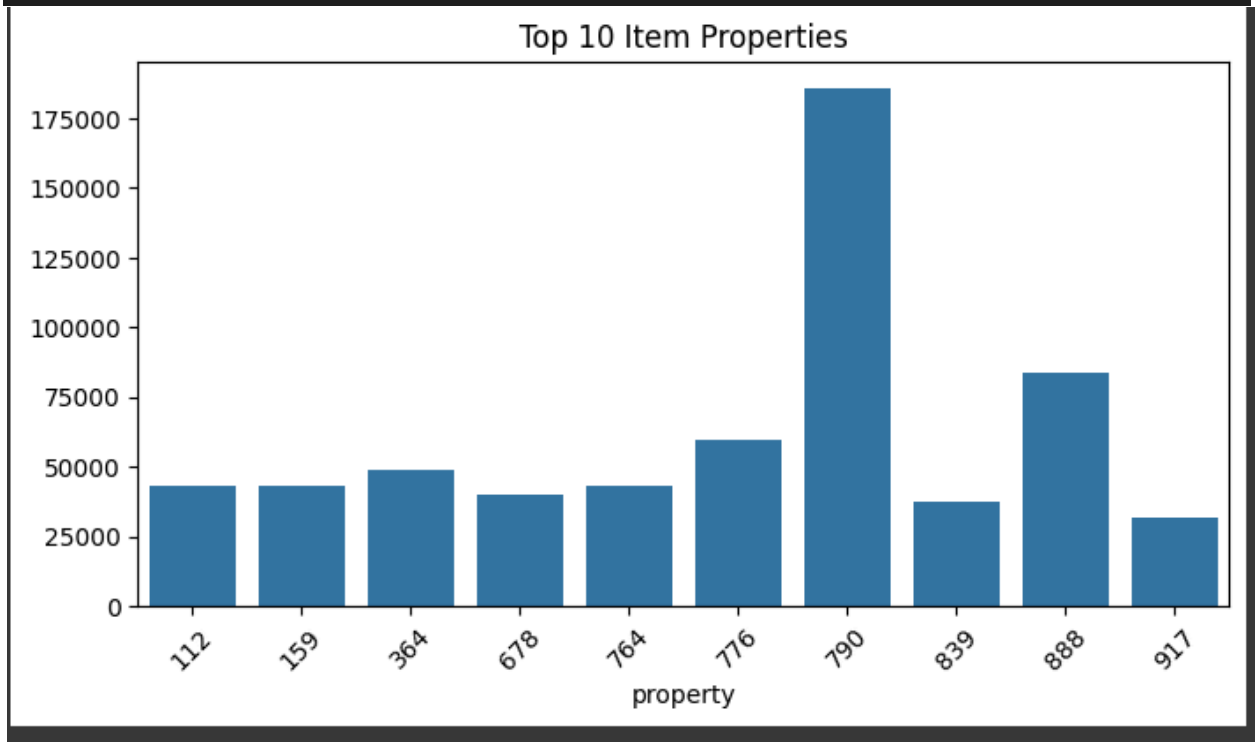
```
# Top 10 most active users
import matplotlib.pyplot as plt
import seaborn as sns

top_users = events["visitorid"].value_counts().head(10)
plt.figure(figsize=(8,4))
sns.barplot(x=top_users.index, y=top_users.values)
plt.xticks(rotation=45)
plt.title("Top 10 Most Active Users")
plt.show()
```



2. Top 10 Item Properties

This chart shows the most frequent item attributes (properties) in the dataset.

```
top_props = item_props["property"].value_counts().head(10)
plt.figure(figsize=(8,4))
sns.barplot(x=top_props.index, y=top_props.values)
plt.xticks(rotation=45)
plt.title("Top 10 Item Properties")
plt.show()
```

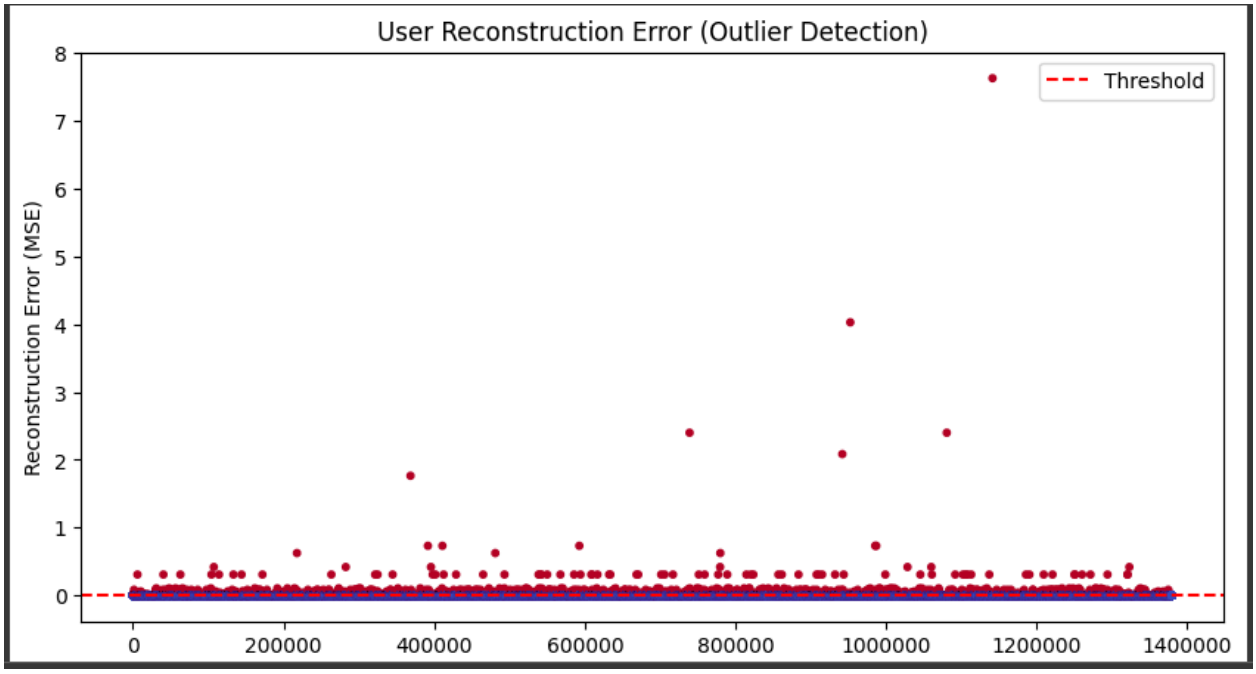


3. Reconstruction Error Distribution (Anomaly Detection)

This histogram shows the distribution of **reconstruction errors** from the CNN Autoencoder.
High errors indicate **abnormal users** (e.g., fraud, bots).

```
# --- Step 6: Visualization (No Scientific Notation) ---
plt.figure(figsize=(10,5))
sns.histplot(mse, bins=50, kde=True, color='blue')
```

```
plt.axvline(threshold, color='red', linestyle='--', label=f'Threshold
({threshold:.4f})')

plt.title("Reconstruction Error Distribution")
plt.xlabel("Reconstruction Error (MSE)")
plt.ylabel("Number of Users")

# Disable scientific notation on both axes
plt.ticklabel_format(style='plain', axis='x')
plt.ticklabel_format(style='plain', axis='y')

plt.legend()
plt.show()
```

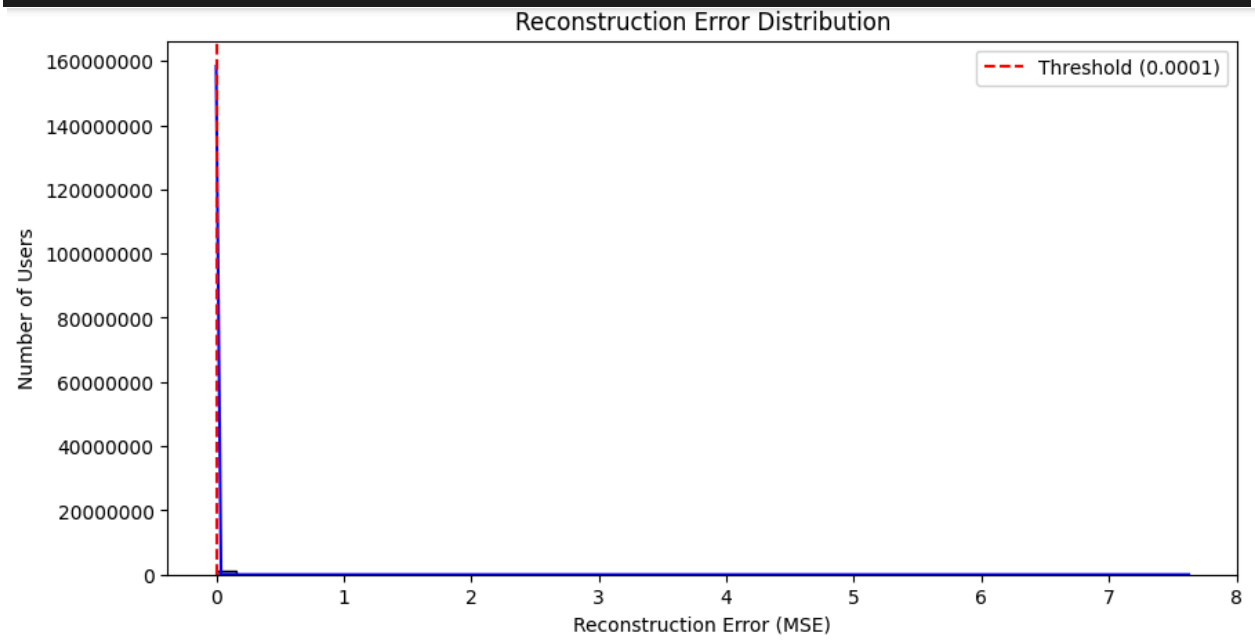


4. User Reconstruction Errors (Scatter Plot)

This chart shows each user's reconstruction error, with anomalies highlighted.

```
# --- Step 6: Visualization (No Scientific Notation) ---
plt.figure(figsize=(10,5))
sns.histplot(mse, bins=50, kde=True, color='blue')
```

```
plt.axvline(threshold, color='red', linestyle='--', label=f'Threshold
({threshold:.4f})')

plt.title("Reconstruction Error Distribution")
plt.xlabel("Reconstruction Error (MSE)")
plt.ylabel("Number of Users")

# Disable scientific notation on both axes
plt.ticklabel_format(style='plain', axis='x')
plt.ticklabel_format(style='plain', axis='y')

plt.legend()
plt.show()
```



Author

Kezia Agyemang-Saahene