

analysis

October 5, 2025

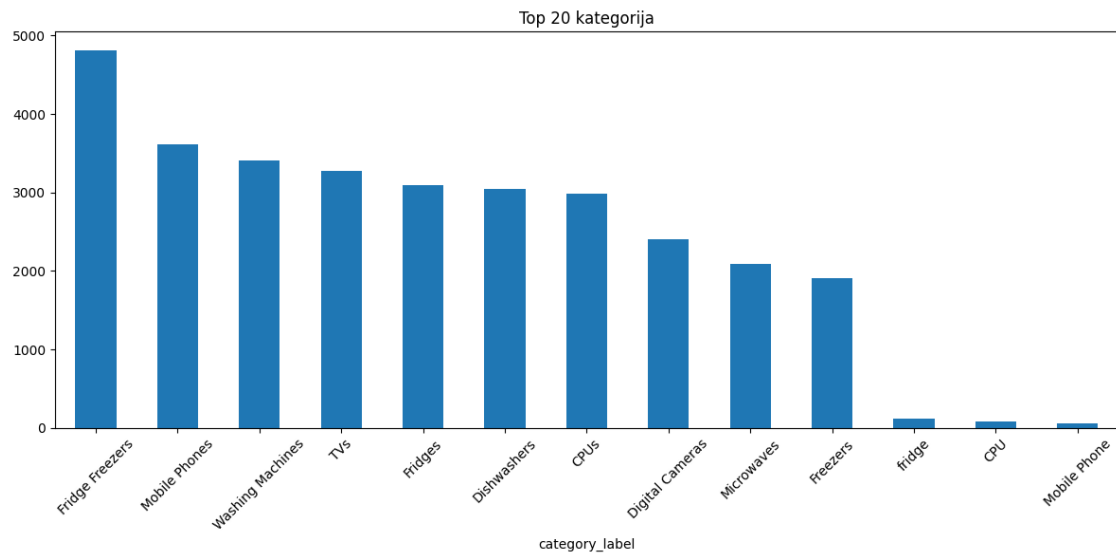
```
[2]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Učitavanje i čišćenje
df = pd.read_csv("../data/products.csv")
df.columns = df.columns.str.strip().str.lower().str.replace(" ", "_")

# Uklanjanje praznih vrednosti i duplikata
df = df.dropna(subset=["product_title", "category_label"])
df = df.drop_duplicates(subset=["product_title", "category_label"])

# Feature inženjering
df["title_length"] = df["product_title"].apply(lambda x: len(str(x).split()))
df["digit_count"] = df["product_title"].apply(lambda x: sum(c.isdigit() for c_
    ↪in str(x)))

# Vizualizacija kategorija
plt.figure(figsize=(12,6))
df["category_label"].value_counts().head(20).plot(kind="bar")
plt.title("Top 20 kategorija")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



[]: