import streamlit as st

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

# Load dataset

df = pd.read\_csv('unicorns.csv')

df.columns = df.columns.str.strip()

df['Valuation ($B)'] = df['Valuation ($B)'].astype(str).str.replace('[^0-9.]', '', regex=True)

df['Valuation ($B)'] = pd.to\_numeric(df['Valuation ($B)'], errors='coerce')

df['Date Joined'] = pd.to\_datetime(df['Date Joined'], errors='coerce')

df['Year Joined'] = df['Date Joined'].dt.year

st.title("🦄 Unicorn Startups Analysis")

# 1. Histogram

st.subheader("1. Distribution of Unicorn Valuations")

filtered\_val = df[df['Valuation ($B)'] < 200]

fig1, ax1 = plt.subplots()

sns.histplot(data=filtered\_val, x='Valuation ($B)', bins=30, kde=True, color='skyblue', ax=ax1)

ax1.set\_title("Distribution of Valuations")

st.pyplot(fig1)

# 2. Top Countries

st.subheader("2. Top 10 Countries by Number of Unicorns")

top\_countries = df['Country'].value\_counts().head(10)

fig2, ax2 = plt.subplots()

sns.barplot(x=top\_countries.values, y=top\_countries.index, palette='viridis', ax=ax2)

ax2.set\_title("Top Countries")

st.pyplot(fig2)

# 3. Boxplot by Industry

st.subheader("3. Valuation by Industry")

top\_industries = df['Industry'].value\_counts().head(10).index

filtered\_df = df[df['Industry'].isin(top\_industries)].copy()

order = filtered\_df.groupby('Industry')['Valuation ($B)'].median().sort\_values(ascending=False).index

fig3, ax3 = plt.subplots(figsize=(12, 6))

sns.boxplot(data=filtered\_df, x='Industry', y='Valuation ($B)', order=order, ax=ax3, showfliers=False)

plt.xticks(rotation=30, ha='right')

st.pyplot(fig3)

# 4. Lineplot by Year

st.subheader("4. Unicorns per Year")

year\_counts = df['Year Joined'].value\_counts().sort\_index()

fig4, ax4 = plt.subplots()

sns.lineplot(x=year\_counts.index, y=year\_counts.values, marker='o', ax=ax4)

ax4.set\_title("Unicorns Founded Per Year")

st.pyplot(fig4)