# Projekt – Analýza finančných správ

Meno študenta: Marek Tatár  
Trieda: 3. ročník  
Dátum: 16.9.2025

## Úvod

Cieľom projektu je analyzovať dataset s finančnými správami, zistiť, aký je celkový tón (sentiment) článkov a ktoré slová sa najčastejšie vyskytujú.

## Dataset

Dataset: Financial News & Market Events (2025)  
Link: https://www.kaggle.com/datasets/  
  
Stĺpce obsahujú napríklad: dátum, nadpis, text článku, sektor spoločnosti.

## Krok 1 – Načítanie dát

import pandas as pd  
df = pd.read\_csv('/kaggle/input/financial-news/financial\_news.csv')  
print(df.shape)  
print(df.columns)  
df.head()  
  
Počet riadkov a stĺpcov: (3024, 12)

## Krok 2 – Čistenie textu

df['text'] = df['headline'].fillna('') + ' ' + df['content'].fillna('')  
df['text'] = df['text'].str.lower()  
df['text'] = df['text'].str.replace('[^a-z ]','',regex=True)  
df['word\_count'] = df['text'].str.split().apply(len)  
  
→ Vytvor graf histogram počtu slov v článkoch.

## Krok 3 – Analýza slov a sentimentu

from wordcloud import WordCloud  
import matplotlib.pyplot as plt  
from textblob import TextBlob  
  
# WordCloud  
text = ' '.join(df['text'])  
wc = WordCloud(max\_words=100).generate(text)  
plt.imshow(wc, interpolation='bilinear')  
plt.axis('off')  
plt.show()  
  
# Sentiment  
df['sentiment'] = df['text'].apply(lambda x: TextBlob(x).sentiment.polarity)  
df['sentiment\_category'] = pd.cut(df['sentiment'], bins=[-1,-0.1,0.1,1], labels=['neg','neutral','pos'])  
df['sentiment\_category'].value\_counts().plot(kind='bar')  
plt.title('Počet článkov podľa sentimentu')

plt.show()  
  
→ Vlož do dokumentu graf WordCloud a graf sentimentu.

## Krok 4 – Záver

Napíš krátky text, čo si zistil. Napríklad:  
- Väčšina článkov bola neutrálna  
- Pozitívne správy boli najmä zo sektora technológií  
- Model by sa dal rozšíriť o predikciu cien akcií

