

In [8]:

```
# Практика 2
import pandas as pd
import numpy as np
from sklearn.decomposition import PCA
from scipy.stats import pearsonr

prices = pd.read_csv('close_prices.csv')
pca_10 = PCA(n_components=10)

X = prices.iloc[:, 1:]
pca_10.fit(X)
explained_variance_ratio = pca_10.explained_variance_ratio_
cumulative_variance_ratio = np.cumsum(explained_variance_ratio)
n_components_90 = np.argmax(cumulative_variance_ratio >= 0.9) + 1

first_component = X_pca[:, 0]

djia = pd.read_csv('djia_index.csv')

djia_column_name = "^DJI"
correlation, p_value = pearsonr(first_component, djia[djia_column_name])

first_component_weights = pca_10.components_[0]
company_names = X.columns
weights_df = pd.DataFrame({
    'Company': company_names,
    'Weight': first_component_weights
})
max_weight_idx = np.argmax(np.abs(first_component_weights))
company_with_max_weight = company_names[max_weight_idx]
max_weight = first_component_weights[max_weight_idx]

print(f"1. Количество компонент для объяснения 90% дисперсии: {n_components_90}")
print(f"2. Корреляция Пирсона между первой компонентой и индексом Доу-Джонса: {correlation}")
print(f"3. Компания с наибольшим весом в первой компоненте: {company_with_max_weight}")
```

1. Количество компонент для объяснения 90% дисперсии: 4
2. Корреляция Пирсона между первой компонентой и индексом Доу-Джонса: 0.91
3. Компания с наибольшим весом в первой компоненте: V

In [ ]: