EXCERPT

THE WORLD'S LARGEST PUBLIC COMPANIES

The Forbes Global 2000 releases the annual ranking of the top 2000 public companies in the world, distributed by Forbes magazine. "The Global 2000" annual ranking is determined by a weighted evaluation of four key criteria: sales, profit, assets, and market value.

OBJECTIVE

Evaluate the performance of top-performing industries from 2008 to 2022 to identify successful operational strategies using key performance indicators (KPIs) and gain insights into the impact of COVID-19 on business operations.

PROJECT & DATA

- Project Brief
- Datasets: 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022 | *Open Source from data.world.*
- Click here for Forbes' Global 2000 methodology

LIMITATIONS

- Data contains records from 2008 – 2022.
- The global largest public companies are limited to the top 2000 performers from various industry types.

TECHNIQUES APPLIED

- Data Sourcing & Data Cleaning: Wrangling, Subsetting, and Consistency Checks
- Data manipulation: Deriving New Variables, Aggregating, and Grouping Data
- Exploratory Visual Analysis: Linear Regression, Geospatial
- Linear Regression (Unsupervised Machine Learning Model)
- K-means Clustering (Supervised Machine Learning Model)
- Time Series Analysis
- Data Dashboard (Tableau Storyboard)

TOOLS

DATA MANAGEMENT

DATA SOURCING & DATA CLEANING
Source datasets that align with project objectives for advanced analytics, apply necessary preparatory methods to optimize diagnostic outcomes, and define questions to explore the data content.

DATA MANIPULATION & EXPLORATORY ANALYSIS
Utilize Excel and Python functions to aggregate and group data, perform basic statistical analysis, create new columns, and optimize the process through exploratory visual analysis using techniques such as geospatial analysis, scatterplots, histograms, and more.

ADVANCED ANALYTICAL TECHNIQUES
Employing a combination of techniques, such as supervised and unsupervised machine learning models, along with time series analysis, valuable insights can be conveyed regarding potential developments in regions and industry types.

DATA DASHBOARD
Forming a Tableau storyboard that presents curated significant findings of the analysis in an interactive format.

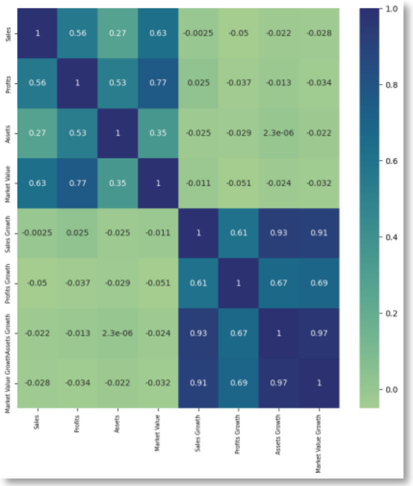
EXPLORATORY VISUAL ANALYSIS

python jupyter pandas NumPy seaborn matplotlib Folium

ANALYTICAL & VISUALIZATION TOOLS APPLICATION

To analyze the relationships between quantitative variables like sales, profits, assets, and market values, a **correlation heat map** was generated. The heat map helps identify initial connections between these variables.

The analysis revealed a **strong positive** correlation between **sales** and **profits**.

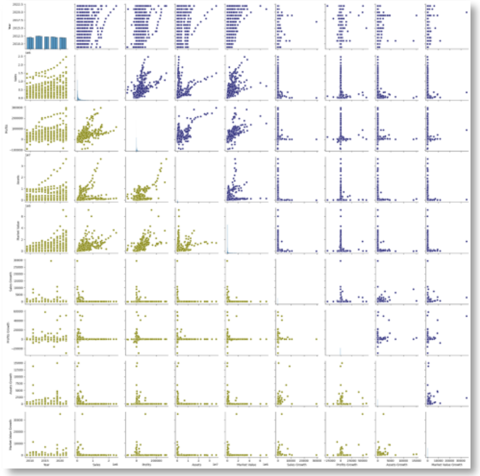


```
# GLOBAL
# Create a subplot with matplotlib

dfgwth_hm1,ax = plt.subplots(figsize = (9,9))
plt.xticks(range(dfgwth_hm.shape[1]), dfgwth_hm.columns, fontsize = 7, rotation = 90) # x axis labels
plt.yticks(range(dfgwth_hm.shape[1]), dfgwth_hm.columns, fontsize = 7) # y axis labels

# Create the correlation heatmap in seaborn by applying a heatmap onto the correlation matrix
glbl_hm2 = sns.heatmap(usa_hm.corr(), cmap = 'crest', annot = True, ax = ax)
```

FIG. 6a

Pair plots, similar to heat maps, depict relationships and directionality between variables. They help determine the potential effects of one variable on another.

In the pair plot analysis, variables related to **sales** show positive growth trends, indicating that sales could be a **significant predictor** of performance.

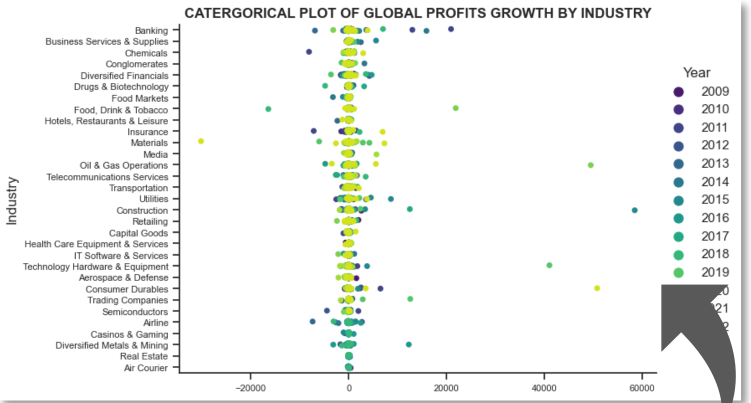


```
# GLOBAL
# Create a first Pair Plots

glbl_pp = sns.pairplot(glblpp)
glbl_pp.map_upper(sns.scatterplot, color = 'midnightblue')
glbl_pp.map_lower(sns.scatterplot, color = 'olive')
```

FIG. 6b

Categorical plots provide valuable insights into non-linear relationships, complementing the information from pair plots.

Analyzing **industry-specific** developments over time reveals an overview of **performance trends**.




```
# Set the figure size
plt.figure(figsize = (50,10))

# Creating a Categorical plot
sns.set(style = 'ticks')
cp_glbl = sns.catplot(x = 'Profits Growth',  y = 'Industry', hue = 'Year', palette = 'viridis', aspect

# Adding annotations to the graph
plt.xlabel('')
plt.yticks(fontsize = 8)
plt.xticks(fontsize = 8)
plt.title('CATERGORICAL PLOT OF GLOBAL PROFITS GROWTH BY INDUSTRY', fontweight = 'bold')
```

FIG. 6c

VIEW FULL REPORTS BELOW

Python Codes | Interactive Dashboard

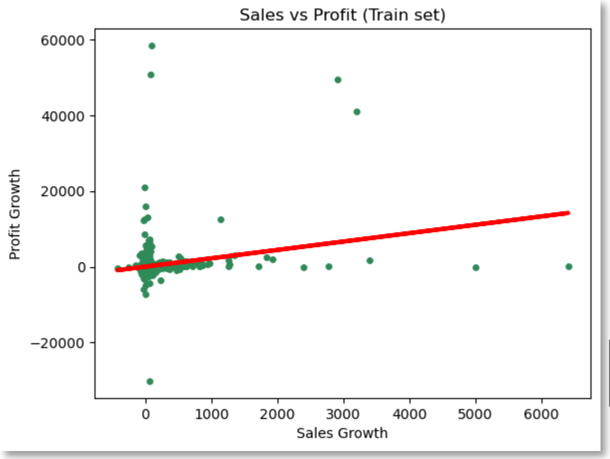
ADVANCED ANALYTICAL TECHNIQUES

python | jupyter | pandas | NumPy | seaborn | matplotlib | scikit learn machine learning in Python

ANALYTICAL & VISUALIZATION TOOLS APPLICATION

The **linear regression** analysis identifies a strong correlation between sales and profits, indicating a trend.

However, the statistical testing yields **poor results**, suggesting that there is not enough evidence to depict the presence of the trends.

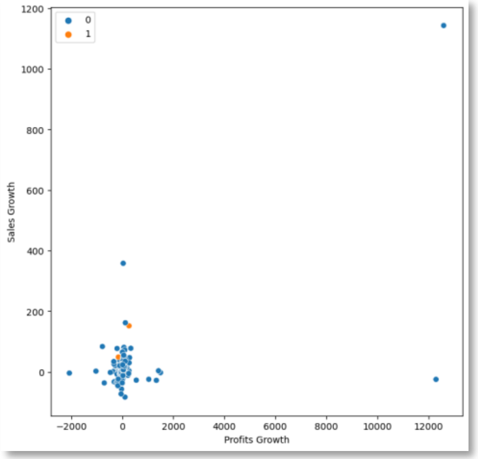



```
# Visualizing the training set results.

plot_test = plt
plot_test.scatter(X_train, y_train, color ='seagreen', s = 15)
plot_test.plot(X_train, y_predicted_train, color = 'red', linewidth = 3)
plot_test.title('Sales vs Profit (Train set)')
plot_test.xlabel('Sales Growth')
plot_test.ylabel('Profit Growth')
plot_test.show()
```

FIG. 6d

Further testing was conducted using the **K-means algorithm**, which groups data points with similar traits into clusters.

However, the scatterplot reveals overlapping data points without clear groupings, indicating an **inadequate fit** to the prediction model.

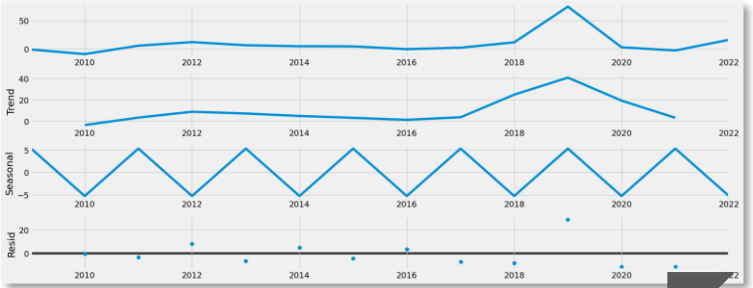



```
# Plot the clusters

plt.figure(figsize = (8,8))
ax1 = sns.scatterplot(x = usa2['Profits Growth'], y = usa2['Sales Growth'], hue
# Here, you're subsetting `X` for the x and y arguments to avoid using their la
# `hue` takes the value of the attribute `kmeans.labels_`, which is the result
# `s` represents the size of the points you want to see in the plot.

ax1.grid(False) # This removes the grid from the background.
plt.show()
```

FIG. 6e

The **decomposition** of the time series allows for the assessment of individual components, such as seasonality, to identify possible trends.

It is evident that **sales fluctuate** over time, with a significant decline in 2019, coinciding with the onset of the pandemic.




```
# Decompose the time series using an additive model

decomposition1 = sm.tsa.seasonal_decompose(usa2, model = 'additive', period = 2)

from pylab import rcParams # This will define a fixed size for all special charts.

rcParams['figure.figsize'] = 18, 7

# Plot the separate components

decomposition1.plot()
plt.show()
```

FIG. 6f

VIEW FULL REPORTS BELOW

Python Codes | Interactive Dashboard

RECOMMENDATIONS

NEXT STEPS

QUALITATIVE RESEARCH

Supplement quantitative analysis with qualitative research to gain a holistic view of a company's prospects. Consider non-financial factors like industry dynamics, competitive landscape, market trends, and overall economic conditions (GDP growth, inflation rates, labor market condition, consumer spending patterns) to understand the broader context in which financial metrics operate.

COMPARATIVE ANALYSIS

Compare the sales, profit, assets, and market value of the company with its competitors in the same industry or sector. Identify relative strengths and weaknesses, assess market positioning, and spot any significant divergences.

INVESTOR SENTIMENT ANALYSIS

Analyze stock price movements, market reactions to earnings releases or product launches, and current events. Provide insights into how investors perceive and react to company developments, which can influence market value and investor sentiment.

FINANCIAL RATIO ANALYSIS

Analyze the ratios over time and/or compare them to industry benchmarks, investors and analysts can gain insights into a company's financial health, efficiency, and market valuation.

VIEW FULL REPORTS BELOW

Python Codes | Interactive Dashboard