Introduction to Data Science

with Python

Alexis Bogroff

11 mai 2022

Presenter

Alexis Bogroff
Lecturer and Mentor in Data Science
at Paris 1 Panthéon-Sorbonne, ESILV,
Openclassrooms, EM-Lyon

Presenter

Alexis Bogroff
Lecturer and Mentor in Data Science
at Paris 1 Panthéon-Sorbonne, ESILV,
Openclassrooms, EM-Lyon

- 4 years Teaching Assistant and lecturer in VBA, Python for finance, SQL, Data Analysis and Data Science
- 9 months Researcher Assistant at Paris 1 Panthéon-Sorbonne within H2020 European Project
- 1 year Data Scientist at Pléiade Asset Management

Why using vizualizations

- Quick understanding simple patterns (trend line plot, groups scatter plot)
- Better intuition on complex patterns (CNN weights maps)
- Reporting

Graphs types

- Univariate Analysis
  - Histograms (distributions)
  - Line plots (Time series)
  - Lorentz Curve (inegalities)
- Multivariate Analysis
  - Scatter plots
  - Heatmaps
    - Correlations
    - Confusion matrices

Libraries

- Matplotlib (.pyplot)
- Seaborn for nice default graphs
- Plotly (Dash) for interactive graphs

Pandas

- Objects
  - pd.DataFrame
  - pd.Series
- Masks / Filters
- Methods (max, info, describe)
- Apply, vectorial operations
- Useful functions
  - Graphs (.plot, .scatter.plot, .plot.bar, .hist)
  - value_counts
  - .isna, .fillna
  - groupby
  - crosstab
- merging DataFrames the right method (outer, indicator=True)
- Pandas profiling