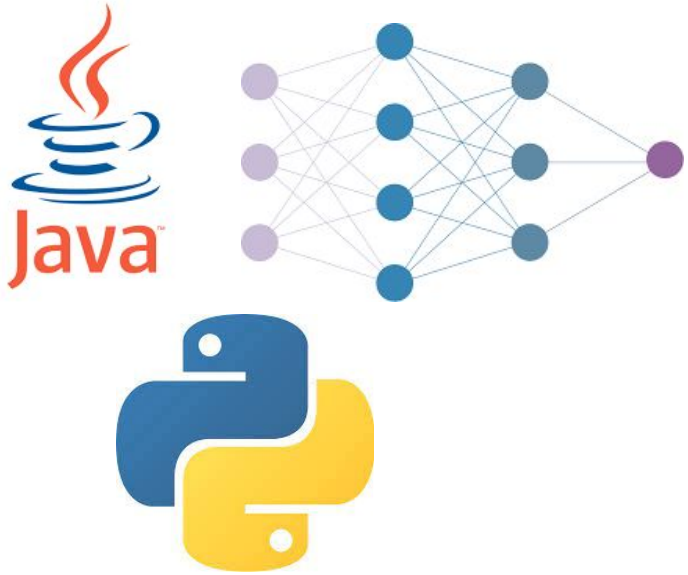
# Playing with time series data in python




Nele Custers
nele.custers@pxl.be

| Year | Breakthroughs in AI | Datasets (First Available) | Algorithms (First Proposed) |
|---|---|---|---|
| 1994 | Human-level spontaneous speech recognition | Spoken Wall Street Journal articles and other texts (1991) | Hidden Markov Model (1984) |
| 1997 | IBM Deep Blue defeated Garry Kasparov | 700,000 Grandmaster chess games, aka "The Extended Book" (1991) | Negascout planning algorithm (1983) |
| 2005 | Google's Arabic- and Chinese-to-English translation | 1.8 trillion tokens from Google Web and News pages (collected in 2005) | Statistical machine translation algorithm (1988) |
| 2011 | IBM Watson became the world Jeopardy! champion | 8.6 million documents from Wikipedia, Wiktionary, Wikiquote, and Project Gutenberg (updated in 2010) | Mixture-of-Experts algorithm (1991) |
| 2014 | Google's GoogLeNet object classification at near-human performance | ImageNet corpus of 1.5 million labeled images and 1,000 object categories (2010) | Convolution neural network algorithm (1989) |
| 2015 | Google's Deepmind achieved human parity in playing 29 Atari games by learning general control from video | Arcade Learning Environment dataset of over 50 Atari games (2013) | Q-learning algorithm (1992) |
| **Average No. of Years to Breakthrough:** | | **3 years** | **18 years** |

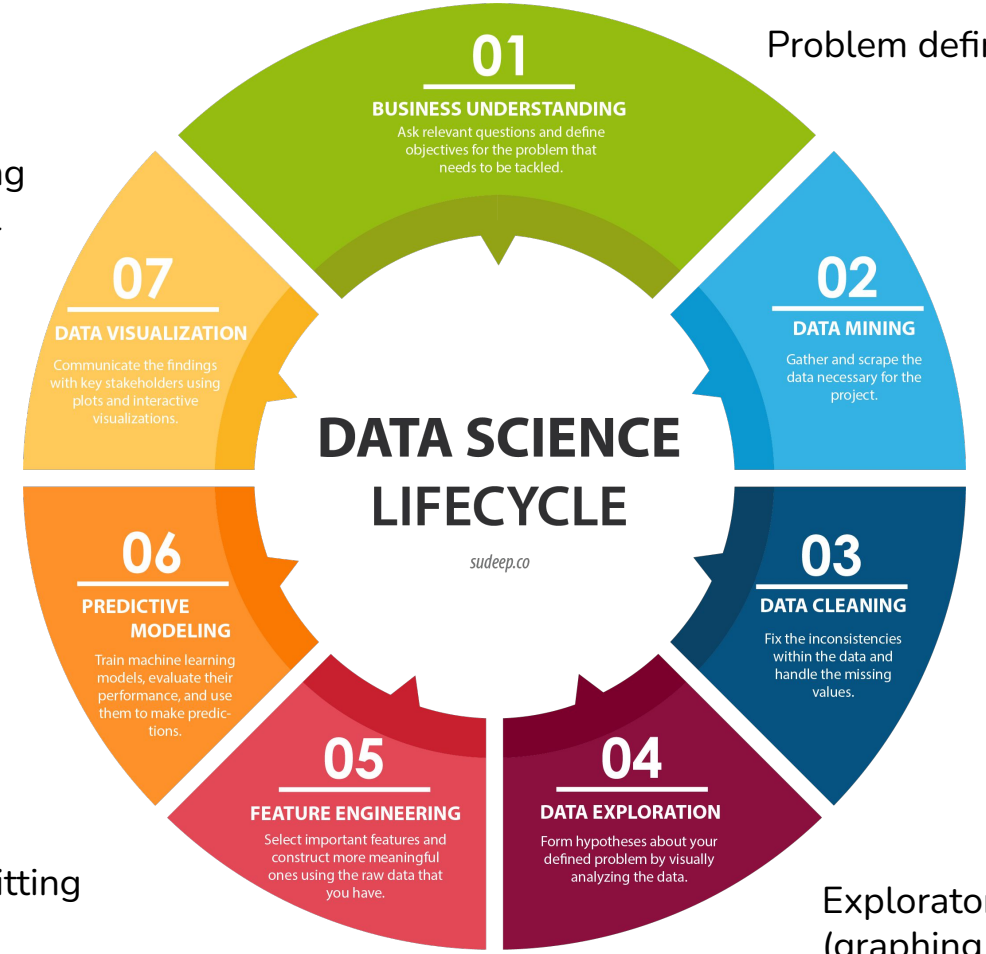
Wissner-Gross, Alexander (2016). Datasets Over Algorithms.

# What is a Time Series?

***Time series*** *is a sequence of observations recorded at regular time intervals.*

Problem definition

Using and evaluating a forecasting model




Gathering information

Choosing and fitting models

Exploratory analysis (graphing the data)

# Time series patterns

***Trend***

A *trend* exists when there is a long-term increase or decrease in the data.

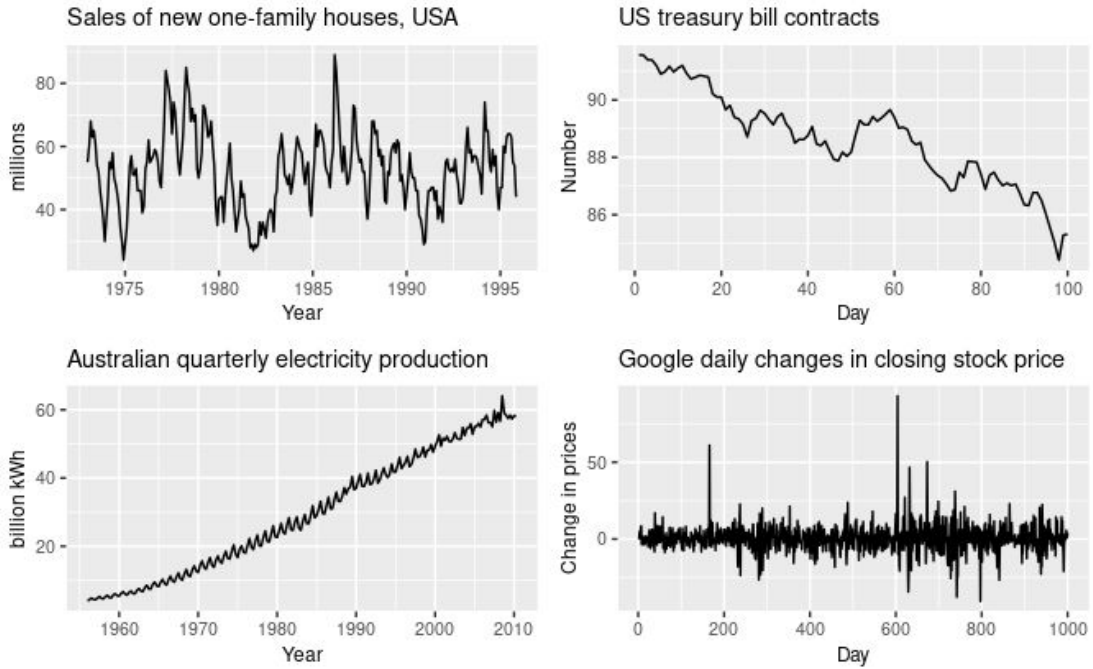
***Seasonal***

A *seasonal* pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week.

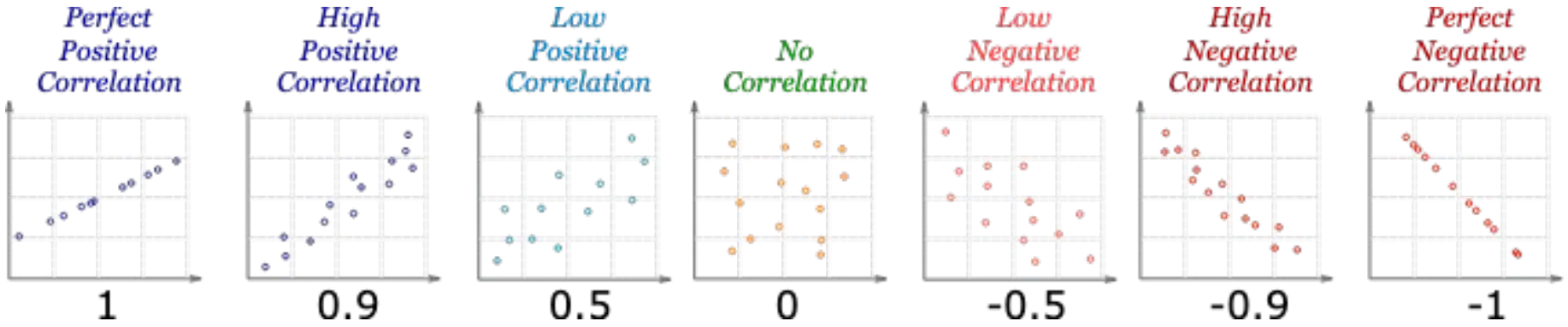
***Cyclic***

A *cycle* occurs when the data exhibit rises and falls that are not of a fixed frequency.

# Time series patterns



# Exploratory analysis



# Python libraries

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

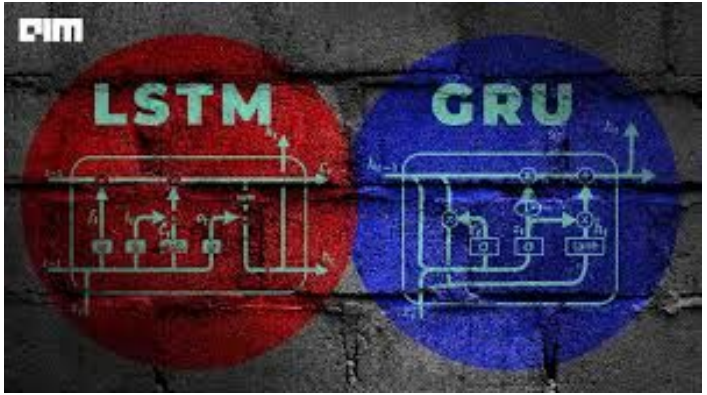
pandas provides data structures for in-memory analytics (which makes **using pandas to analyze datasets that are larger than memory datasets somewhat tricky)**.

matplotlib offers static, animated and interactive visualizations in Python.

seaborn is a Python data visualization library based on matplotlib.

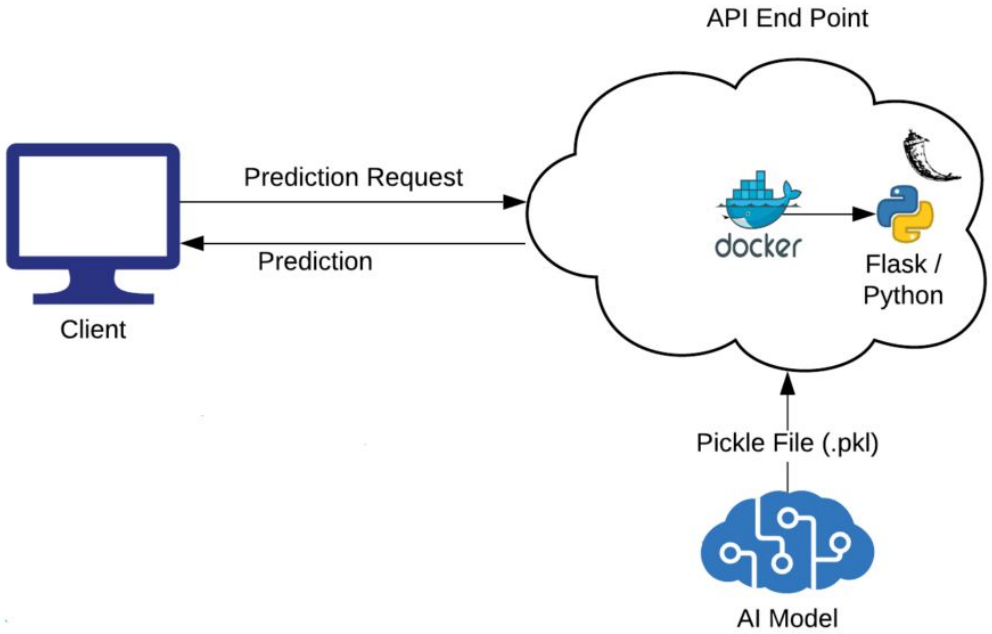
numpy is the standard for working with numerical data in Python.

# Several solutions for forecasting




PROPHET

# Deploying


API End Point
Prediction Request
Prediction
Client
docker
Flask /
Python
Pickle File (.pkl)
AI Model

# Resources

Forecasting principles and practice: https://otexts.com/fpp2/

https://facebook.github.io/prophet/

https://towardsdatascience.com/playing-with-time-series-data-in-python-959e2485bff8

https://towardsdatascience.com/time-series-analysis-with-theory-plots-and-code-part-1-dd3ea417d8c4

https://www.digitalocean.com/community/tutorials/how-to-build-and-deploy-a-flask-application-using-docker-on-ubuntu-20-04

https://blog.jetbrains.com/pycharm/2017/03/docker-compose-getting-flask-up-and-running/

http://sudeep.co/data-science/Understanding-the-Data-Science-Lifecycle/

https://www.vantage-ai.com/en/blog/4-strategies-how-to-deal-with-large-datasets-in-pandas