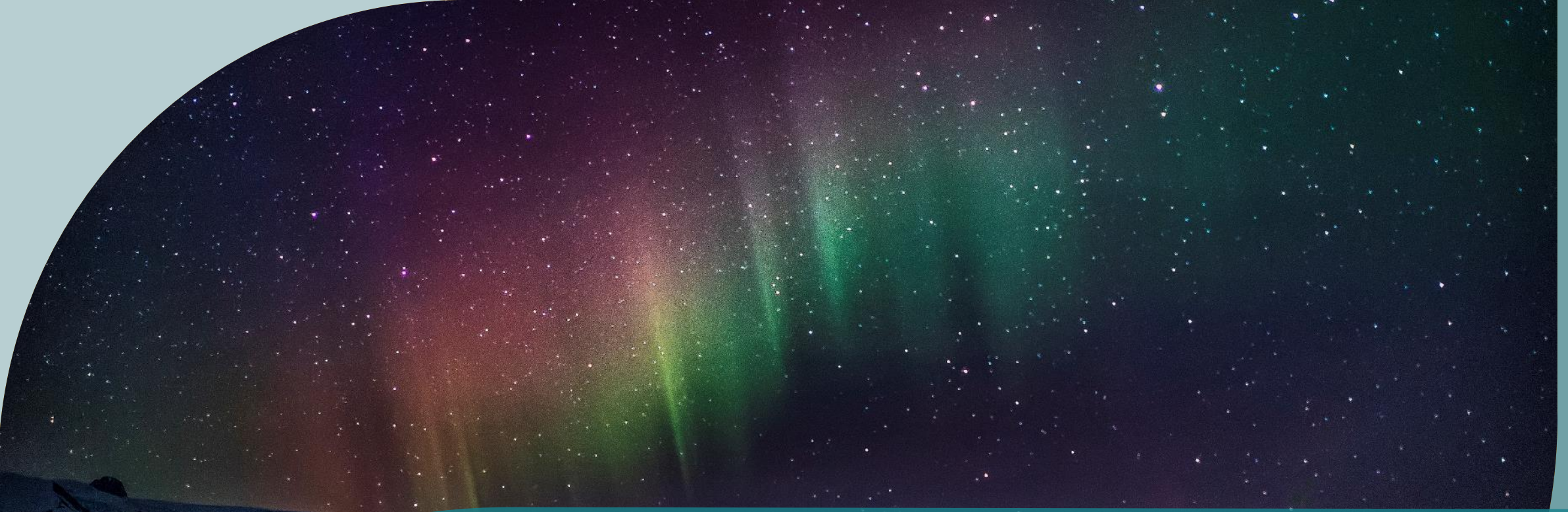


Boosting Your Data Analysis Power with Python

Part 1

Ray Chen

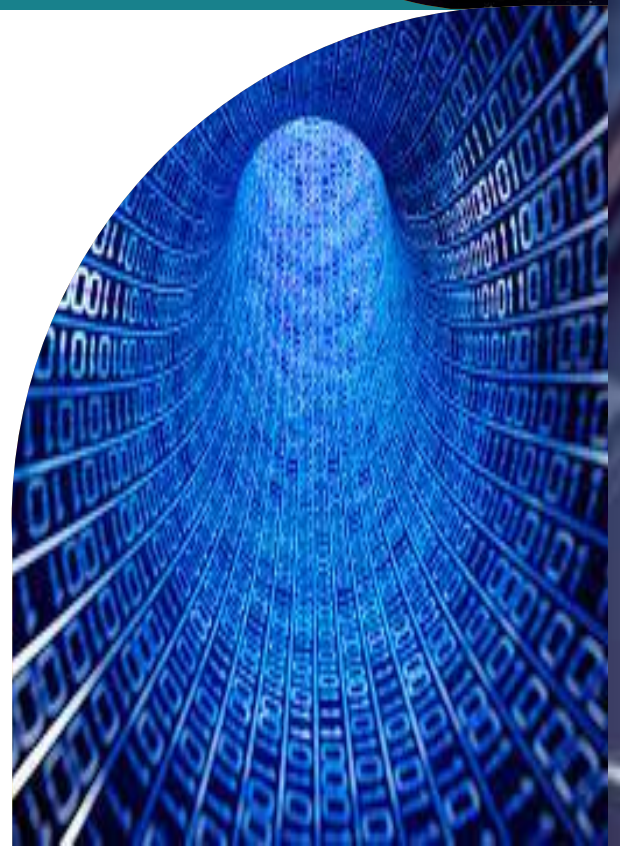
raychen_email@yahoo.com

Agenda

Python

Data Analysis in Quality

Demo



Python

Python is a programming language created by Guido Von Rossum and first released in 1991.

High level

Interpreted

Open Source

Data Analysis

- NumPy (Numerical Python) - numeric and scientific computing
- Pandas - data manipulation and analysis
- Matplotlib - plots and graphs
- Scikit-learn, TensorFlow, and PyTorch - machine learning and deep learning.
- PySpark - distributed computing for big data analytics tasks.

  *Integration with Apache Spark and Hadoop for processing analyzing large-scale datasets.

Integrated Development Environment (IDE)

Write and run code

PyCharm

- Excellent support for NumPy, Pandas, and Matplotlib
- Advanced debugging capabilities
- Work with Jupyter Notebook files

Integrated Development Environment (IDE)

Jupyter Notebook/JupyterLab

- Web-based
- Notebook style - add explanatory text, visualizations, and equations to your project
- Support multiple programming languages including Python & R

Integrated Development Environment (IDE)

Spyder

- Open-source
- Variable explorer
- Debug
- Integration with NumPy and Pandas

Distribution Package

Anaconda

- A collection of powerful data science libraries and IDEs including Python, PyCharm, Spyder and Jupyter Notebook
- Simplifies the setup and management of Python environments

Google Colab

Google Colab/ Colabortory

- Similar to Jupyter Notebook
- Cloud-based, no installation is needed
- Free access to GPUs

Correlation

- Identify Relationships - identifying and measuring the strength and direction of relationships between two or more variables
- Quality Improvement - identify potential factors that have a significant influence on the quality of a product or potential root causes of quality issues

Regression

- Model Relationship - model the relationship between a response (quality outcome) and one or more predictors (process inputs or factors), understanding the quality outcome based on the predictors
- Optimization - identify the optimal values or ranges for the predictors that will result in the desired quality level. Additonally, identifying control limits for effective quality control and monitoring.

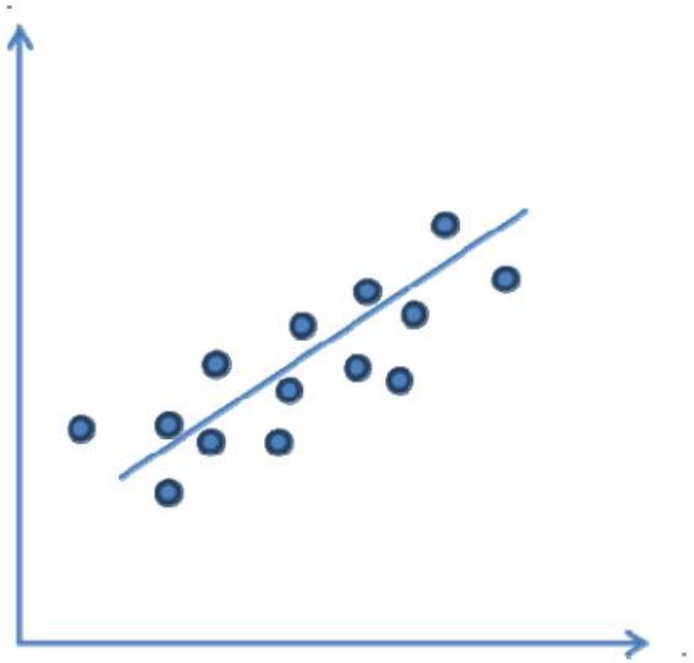
Correlation and Regression Example

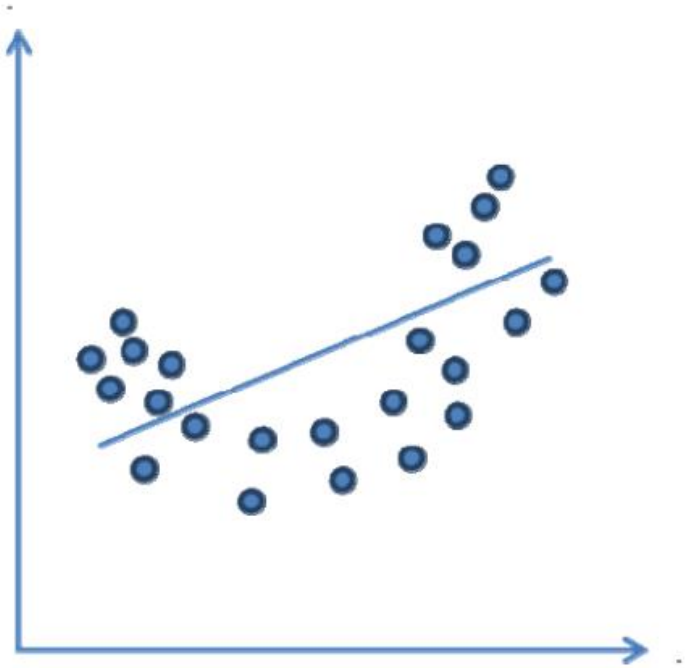
- Aluminum casted product
- Variables
  - Strength
  - Porosity
  - Impurity

Regression with Categorical Predictor Example

- Plastic product
- Response:
  - Strength
- Predictors:
  - Temperature (continuous)
  - Machine (categorical)

Regression with Categorical Predictor Example





Capability Analysis

- Process Performance - evaluate the performance of processes in meeting customer specifications. A measure of whether a process is capable of consistently performing within the desired tolerance limits.
- Process Variability - identify sources of process variability by analyzing the process spread, pinpointing areas of improvements that can reduce variation and enhance consistency.

Capability Analysis

$$CPU = \frac{(USL - \mu)}{(3 * \sigma_{Within})}$$

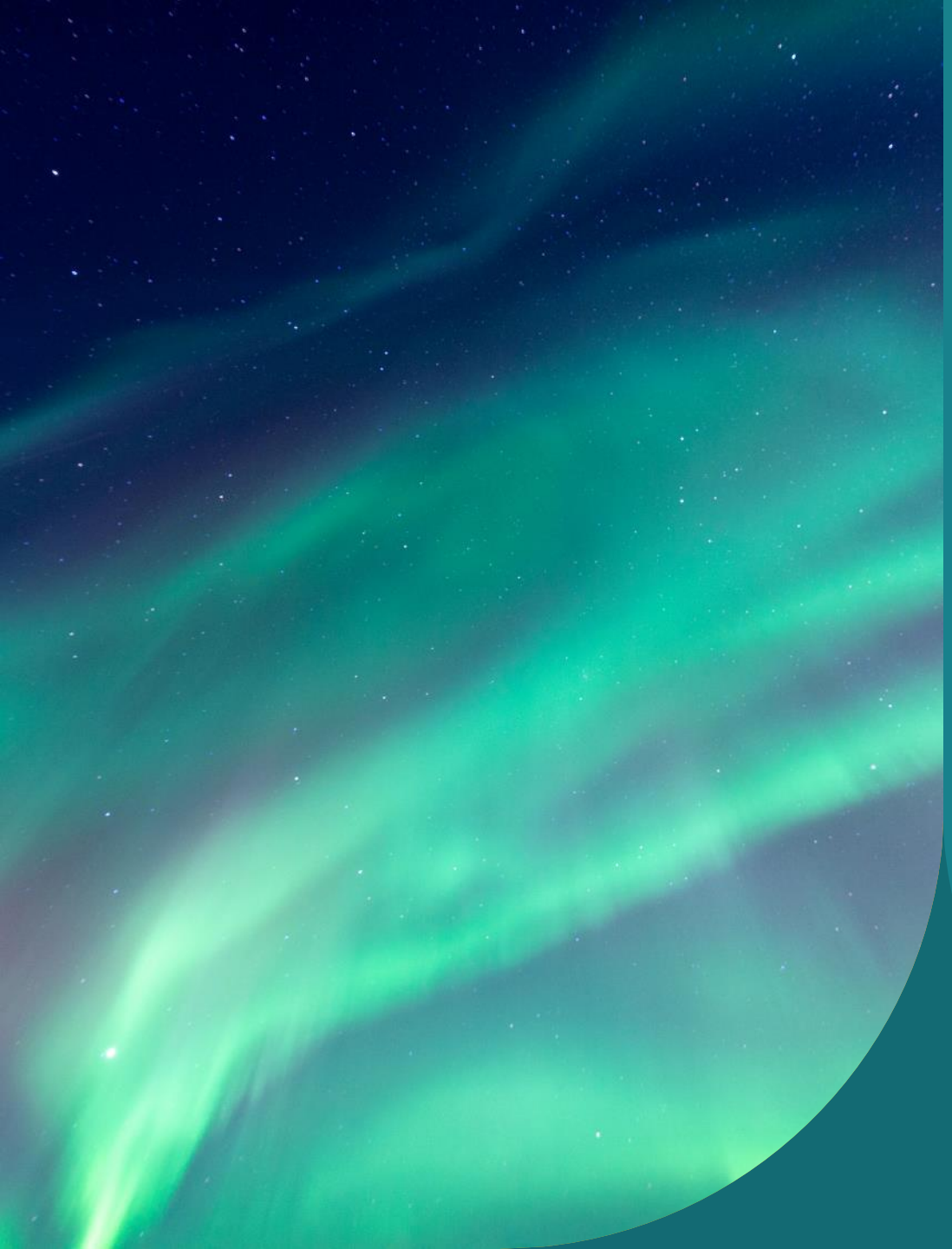
$$CPL = \frac{(\mu - LSL)}{(3 * \sigma_{Within})}$$

$$Cpk = \text{minimum}\{\text{CPU, CPL}\}$$

$$PPU = \frac{(USL - \mu)}{3 * \sigma_{Overall}}$$

$$PPL = \frac{(\mu - LSL)}{3 * \sigma_{Overall}}$$

$$Ppk = \text{minimum}\{\text{PPU, PPL}\}$$



Thank You