

# 1、基础知识(Fundamentals)

## 1.1 基础(Basic)

1.1.1 矩阵和线性代数基础(Matrices & Linear Algebra Fundamentals)

1.1.2 数据库基础(Database Basics)

1.1.3 表格数据(Tabular Data)

1.1.4 Pandas数据结构(Dataframes & Series)

1.1.5 数据仓库技术(ETL,Extract,Transform,Load)

1.1.6 业务分析与商业分析(Reporting vs BI vs Analytics)

1.1.7 数据格式(Data Formats)

(1) JSON

(2) XML

(3) CSV

(4) TXT

1.1.8 正则表达式(RegEx)

## 1.2 Python编程(Python Programming)

1.2.1 Python基础(Basics)

(1) 表达式(Expressions)

(2) 变量(Variables)

(3) 数据结构(Data Structures)

(4) 函数(Functions)

(5) 安装包(via pip,conda or similar)

(6) 代码风格(CodeStyles,e.g.PEP8)

1.2.2 重要库(Important Libraries)

(1) NumPy

(2) Pandas

1.2.3 虚拟环境(Virtual Environments)

(1) miniconda

(2) anaconda

1.2.4 Jupyter Notebook/Lab

## 1.3 数据源(Data Sources)

1.3.1 数据挖掘(Data mining)

1.3.2 网页抓取(Web Scraping)

1.3.3 开源数据集(Public Datasets)

1.3.4 数据竞赛平台(e.g Kaggle)

## 1.4 数据处理与分析(Exploratory Data Analysis)

1.4.1 主成分分析(PCA,Principal Component Analysis)

1.4.2 降维及数值约简(Dimensionality & Numerosity Reduction)

1.4.3 归一化(Normalization)

1.4.4 数据清理, 处理缺失值(Data Scrubbing, Handling Miss Value)

1.4.5 无偏估计(Unbiased Estimators)

1.4.6 稀疏值分箱(Binning Sparse Value)

1.4.7 特征提取(Feature Extraction)

1.4.8 降噪(Denoising)

1.4.9 采样(Sampling)