

Python数据处理编程实验指导书

实验目的：

通过使用Python语言的基本语法知识和相关库(如Numpy, Pandas, Matplotlib), 对给定的实验数据进行分析, 通过图表的形式展示分析结果, 得出分析结论, 掌握如何使用Python语言对数据进行处理分析。

实验1：气象数据分析

数据来源: NOAA (<https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND>)

时间尺度: 2010-1-1至2010-12-31

数据覆盖区域: 美国8个离海距离各不相同的城市

数据含义: 有两个文件, 其中一个文件为weatherData.xlsx, 文件中包含了这些城市1年内的天气数据(气温, 露点, 气压, 平均风速, 风向等), 观测时间间隔为1小时。另一个文件为distance.txt, 则表示这些城市与海洋的距离。

实验目标: 分析靠海对气候的影响, **包括但不限于**

- (1) 分析城市气温与海洋距离的关系, 并画图显示
- (2) 分析城市露点和海洋距离的关系, 并画图显示
- (3) 分析每个城市的风速和风向的分布情况, 并画图显示
- (4) 分析气温和气压的关系, 并画图显示。

实验2：共享单车租用影响因素分析

实验数据: Kaggle网站(Bike Sharing Demand | Kaggle)

时间尺度: 2011-01-01 00:00:00至2012-12-19 23:00:00

数据含义: 该数据集文件格式为csv格式, 该数据集包括了租车日期, 租车季节, 租车气温, 租车空气湿度等数据。该数据集的具体数据特征描述如下:

- datetime - 日期+时间
- season - 1=春天; 2=夏天; 3=秋天; 4=冬天
- holiday - 是否是节假日
- workingday -
 - 1=工作日
 - 0=周末
- weather -

- 1: 晴天, 多云
- 2: 雾天, 阴天
- 3: 小雪, 小雨
- 4: 大雨, 大雪, 大雾

- temp - 气温摄氏度
- atemp - 体感温度
- humidity - 湿度
- windspeed - 风速
- casual - 非注册用户个数
- registered - 注册用户个数
- count - 给定日期时间（每小时）总租车人数

实验目标: 分析影响共享单车的租用情况的因素, **包括但不限于** (1) 天气, (2) 温度, (3) 时间, (4) 风速。

实验3、Python课程辅助教学平台优化和数据分析

实验4、自选数据分析题目

自定义目标, 撰写申请书, 包括实验目标, 数据来源, 主要功能等, 于4月7日之前提交, 通过**老师批准**之后可开展实验.

实验提交

- (1) 以上实验任选其一。
- (2) 使用Jupyter编写代码, 并用文字详细描述数据分析的过程, 运行代码, 显示绘图结果。
- (3) 将整个Jupyter页面保存成为一个ipynb格式的文件和一个PDF文件, PDF中要显示代码的运行结果和绘图。
- (4) 将原始数据, ipynb文件以及PDF文件放在一个文件夹内, 并要保证该文件夹内的ipynb文件可以运行, 能够读取本文件夹中的数据。(注意文件路径问题)

参考评价标准

- 1、工作量。是否全面的考虑了从不同的维度分析了各种因素, 是否深入的探讨了各

种因素的影响。

- 2、分析结论能否自圆其说。是否得出了明确有意义的结论，是否有充分地证据支撑结论，结论的推导过程是否逻辑清晰。
- 3、实验报告的写作,包括但不限于图表规范(是否有横坐标与纵坐标的标签与刻度，是否有相应的图注)、排版、语言准确等因素。
- 4、代码是否结构清晰、核心代码处是否有注释等。