
泰迪科技数据分析项目实训

2020 东京奥运会奖牌数据分析

培训解决方案

广东泰迪智能科技股份有限公司 版权所有

地址：广州市经济技术开发区开泰大道 36 号 1 栋 212

网址：<http://www.tipdm.com>

邮箱：services@tipdm.com

邮编：510663

联系人：

电话：

目录

1 项目介绍.....	3
1.1 项目背景.....	3
1.2 项目目标.....	4
1.3 项目数据.....	4
1.4 项目周期及时间安排.....	5
1.5 项目难度.....	5
2 项目任务.....	6
3 项目流程.....	6
4 核心技术.....	6
5 实现工具.....	7
6 实训对象.....	7
7 前置知识.....	7
8 项目提交规范.....	7
8.1 项目交付内容要求.....	7
8.2 项目交付命名规范.....	7
8.3 项目提交邮件格式.....	8
9 实训对应的就业岗位.....	9
9.1 就业岗位.....	9
9.2 岗位分析.....	9
10 项目收获.....	9
11 项目评价及积分.....	10
12 附件一 工作室管理制度.....	12
13 附件二 前置课程课表.....	12

1 项目介绍

1.1 项目背景

奥林匹克运动会已成为当今世界无与伦比的文化现象与文明载体，而建立在鲜明人文精神基础上的奥林匹克价值观得到了全世界广泛的认同。体育是文化的重要组成部分，也是提升国家软实力建设的重要载体。2008 年北京奥运会，中国体育健儿的矫健英姿频频在国际体育舞台上涌现，先后有创造了多次世界纪录，总获得 100 多块奥运金牌，这次奥运会改变了中国的国际形象和国际地位。



第 32 届夏季奥林匹克运动会（Games of the XXXII Olympiad），即 2020 年东京奥运会于 2021 年 7 月 23 日开幕。本次奥运会共有 204 个国家和地区、俄罗斯奥运队和奥林匹克难民代表团 2 个参赛队伍参赛，各国运动员在这一个多月进行比拼争夺，在此期间，美国、日本和中国多次交替出现在奖牌榜第一名。东京奥运会的竞争非常激烈，金牌榜的排名不断

变更，而且本次奥运会新增了滑板、冲浪、竞技攀岩、棒球和空手道 5 个赛事项目。部分国家在某些赛事项目充分地突出了自身的优势，获取了多枚金牌。在 8 月 8 日，本届奥运会落下帷幕，中国共获得 38 枚金牌、32 枚银，18 枚铜牌，奖牌数共 88 枚，而美国以一枚金牌“全胜”中国，位居世界第一。

本项目旨在利用数据分析挖掘技术，探索并清洗东京奥运会奖牌榜和各奖牌赛程记录等数据，分析 2020 东京奥运会奖牌分布情况和奖牌变化情况，并且利用 Pyecharts 将分析结果可视化。

1.2 项目目标

本项目旨在利用数据分析挖掘技术，清洗并分析东京奥运会奖牌数据，运用 Pyecharts 库可视化奥运会数据分析结果。

1.3 项目数据

本项目的数据来源于央视网 2020 东京奥运会，一共有 4 份数据，分别是东京奥运会奖牌榜数据（national_medal.xlsx）、各赛事项目金牌得主数据（golds.xlsx）、各赛事项目银牌得主数据（silvers.xlsx）和各赛事项目铜牌得主数据（bronzes.xlsx）。

东京奥运会奖牌榜数据的具体数据字段说明如表 1-1 所示。

表 1-1 东京奥运会奖牌榜数据字段说明

字段名	说明
排名	2020 东京奥运会奖牌榜排名
国家	参加项目赛事得奖的国家和地区
金牌	获取金牌数量
银牌	获取银牌数量
铜牌	获取铜牌数量
总数	获取奖牌总数量

nationid	国家英文缩写（非标准国家缩写，仅用于本项目数据）
----------	--------------------------

3 份各赛事项目奖牌得主数据的数据字段一致，具体数据字段说明如表 1-2 所示。

表 1-2 各赛事项目奖牌得主数据字段说明

字段名	说明
日期	奥运会赛事项目确定名次日期
项目	奥运会赛事项目
获得者	奖牌获得者
名次	奥运会赛事项目获奖名次（金牌得主数据名次为 1，银牌得主数据名次为 2，铜牌得主数据名次为 3）
countryid	国家英文缩写（非标准国家缩写，仅用于本项目数据）

1.4项目周期及时间安排

注：以下周期仅为参考，工作室可在选择该项目后自行选择项目开始和结束时间，尽量控制在规定周期内完成，以确保学生可以达到训练目的。

（1）项目周期：2 周

（2）时间安排

任务安排	内容	提交场所	提交时间
第一周	任务 1 明确项目需求与目标	邮箱	Day2
	任务 2 环境准备		
	任务 3 数据读取与探索		
	任务 4 数据预处理		
	任务 5 分析奥运会奖牌分布情况并可视化		Day5
	任务 6 分析主要国家奖牌分布情况并可视化		
第二周	任务 7 分析我国奖牌分布与变化情况并可视化		Day2
	任务 8 比较我国与竞争国家奖牌情况并可视化		Day4
	任务 9 完成项目报告		Day5
在线答疑		QQ 群	

1.5项目难度

★★★

2 项目任务

任务 1 明确项目需求与目标。

任务 2 环境准备：NumPy、Pandas、Pyecharts 等。

任务 3 数据读取与探索，了解数据的基本情况。

任务 4 数据预处理。

任务 5 分析奥运会奖牌在世界国家的分布情况并运用 Pyecharts 可视化分析结果。

任务 6 分析主要国家奖牌分布情况并运用 Pyecharts 可视化分析结果。

任务 7 分析我国奖牌分布与变化情况并运用 Pyecharts 可视化分析结果。

任务 8 比较我国与竞争国家奖牌情况并运用 Pyecharts 可视化分析结果。

任务 9 完成项目报告。

3 项目流程

- (1) 明确项目目标：阅读项目任务书，明确项目交付内容。
- (2) 学习前置知识：提供项目所需知识的云课堂课程，快速掌握项目前置知识。
- (3) 项目实践：获取项目需求，动手做项目。
- (4) 项目验收：对项目成果进行验收，指导改进项目成果。
- (5) 信息入库：参训学生信息录入泰迪人才库，为优秀结业生引荐合作公司。

4 核心技术

- (1) Python 数据分析

(2) Python 数据可视化

5 实现工具

NumPy、Pandas、Pyecharts 等。

6 实训对象

数学、统计学、计算机等相关专业学生。

7 前置知识

第 1 模块：Python 编程基础

第 2 模块：Python 数据分析与应用

第 3 模块：Python 数据可视化

8 项目提交规范

8.1 项目交付内容要求

- (1) 请在“**项目名称（模板）.docx**”基础上进行报告的编写。
- (2) 请在“**项目名称（模板）.pptx**”基础上进行 PPT 的编写。
- (3) 具体的内容以及格式要求请查看课程【工作室项目交付要求讲解】。

注：请按照课程【工作室项目交付要求讲解】中提及的要求进行文档的编写，若不按要求进行规范编写，提交的项目文件将被返回修改。

8.2 项目交付命名规范

-
- (1) 项目交付文件包括中间数据（如有）、代码、报告、PPT。
 - (2) 中间数据需放入“01.中间数据（如有）”文件夹中，各中间数据文件命名为“任务 x-x”，如“任务 1-1”。
 - (3) 代码需放入“02.代码”文件夹中，各代码文件命名为“任务几 任务名称”，如“任务 1 预处理航空客户数据”。
 - (4) 报告需放入“03.报告”文件夹中，报告命名为“项目名称”，如“航空公司客户价值分析”。
 - (5) 报告需放入“04.PPT”文件夹中，PPT 命名为“项目名称”，如“航空公司客户价值分析”。
 - (6) 最终把所有文件夹放入总文件夹中，总文件夹命名为“项目名称--院校工作室名称--小组成员名称（用顿号隔开）--上交日期（如 2021.1.23）”，如“航空公司客户价值分析--重理工智能工作室--刘备、关羽、张飞--2021.1.23”。
 - (7) 最后仅需把(6)提及的总文件夹的压缩包提交至邮箱 taididingyuehao@tipdm.com。

注：具体也可查看课程【工作室项目交付要求讲解】中的项目交付模板。同时，请严格按照以上方式进行上缴，如上缴方式不正确，可能影响评级。

8.3 项目提交邮件格式

- (1) 邮件收件人为：taididingyuehao@tipdm.com
- (2) 邮件主题名称为：院校工作室名称
- (3) 邮件正文内容为：项目名称-小组成员名称
- (4) 邮件附件为：“8.2 项目交付命名规范”提及的项目总文件夹压缩包

项目提交邮件格式具体参考如下。

收件人
taididingyuehao@tipdm.com;

添加抄送 - 添加密送 | 分别发送

主题
T0001泰迪·黔南师院数统智能工作室

继续添加
插入正文
表情
正文模版
截屏
样式1

正文
B I U
T
A
ab
三
三
三
“
”
<HTML>

你好：

本周完成

- 1.Python供应链商品销售数据分析-成员为张三、李四
- 2.Python供应链商品销售数据分析-成员为小青、小白
- 3.Python新冠疫情数据分析-成员为小红、小明

详见附件，请查收

此致！

9 实训对应的就业岗位

9.1 就业岗位

数据分析师

9.2 岗位分析

序号	岗位	主要业务工作	所需技能	相应课程设置
1	数据分析师	数据处理、分析建模、 撰写分析报告	Python 数据分析 分析报告	Python 语言设计 Python 数据分析与应用 数据库

10 项目收获

- （1）足不出校门即可获得实战技能。
- （2）了解人工智能相关岗位目前的就业形式和前景，了解需要掌握的技能。
- （3）掌握一定的挖掘技能和工具，体验一个实际项目的全过程。

(4) 参训学生信息录入泰迪人才库，为优秀结业生引荐合作公司。

11 项目评价及积分

(1) 公司将按照工作室对该项目的完成情况给予评分，项目评分由低到高依次为 D、C、B、B+、A、A+，B 为及格，B 以下不及格。

(2) 企业将根据工作室学员的项目评分给予工作室对应的积分，积分达到一定水平，可以获取更多泰迪资源。

(3) 可获得的资源包括但不限于

A、由泰迪科技颁发的 CBDA 证书



B、由泰迪科技颁发的实习证明



C、云课堂课程资源

(4) 评分对应积分如下

项目评分	积分
项目评分为 B	+50
项目评分为 B+	+60
项目评分为 A	+70
项目评分为 A+	+80

(5) 工作室制度详见附件一。

12 附件一 工作室管理制度

查看网页：<https://kdocs.cn/l/ckdaaSEuHp0y>

13 附件二 前置课程课表

Python 编程基础	Python 数据分析与应用
1 准备工作	1 Python 数据分析概述
1.1 认识 Python	1.1 认识数据分析
1.2 搭建 Python 环境	1.2 熟悉 Python 数据分析的工具
1.3 安装 PyCharm	1.3 安装 anaconda 与启动 jupyter notebook
1.4 PyCharm 使用入门	1.4 掌握 Jupyter Notebook 常用功能
2 列表操作	2 NumPy 数值计算基础
2.1 第一个 Python 程序	2.1.1 NumPy 简介
2.2 Python 固定数据类型介绍	2.1.2 数组创建及基础属性
2.3 列表构建及索引操作	2.1.3 初识数组的特点
2.4 列表元素的增删改查操作	2.1.4 创建常用数组
2.5 列表推导式	2.1.5 数组数据类型
3 程序流程控制语句	2.1.6 生成随机数
3.1 Python 常用操作符	2.1.7 一维数组的索引
3.2 Python 条件判定语句	2.1.8 逻辑型索引
4 字符串操作	2.1.9 多维数组的索引
4.1 字符串及其索引&切片	2.1.10 求解距离矩阵
4.2 字符串的常见方法	2.1.11 变化数组 shape
4.3 字典的创建及索引	2.2.1 Numpy 矩阵介绍
4.4 字典常用操作	2.2.2 Numpy 通用函数介绍
4.5 字典推导式	2.2.3 通用函数的广播机制
5 Python 文件读取操作	2.3.1 Numpy 读写二进制文件
5.1 Python 读取文件	2.3.2 Numpy 读写 txt 文件
6 函数	2.3.3 利用 Numpy 对数据进行简单统计分析
6.1 Python 函数自定义	3 Matplotlib 数据可视化基础

7 面向对象与模块	3.1.1Matplotlib 介绍
7.1 Python 方法与函数对比介绍	3.1.2 基础图形绘制
7.2 Python 面向对象示例	3.1.3 常用参数设置
7.3 Python 模块使用	3.2.1 绘制散点图
7.4 第三方库的安装与调用	3.2.2 散点图参数设置
8 注意事项	3.2.3 绘制折线图
8.1 Python 工作路径说明	3.3.1 绘制直方图
8.2 模块命名及存放路径的注意事项	3.3.2 绘制饼图
8.3 结语	3.3.3 绘制箱线图
	4 Pandas 统计分析基础
	4.1Pandas 简介
	4.2.1Pandas 读取文本数据
	4.2.2 存储数据框
	4.2.3Pandas 读取 excel 文件
	4.2.4 将数据框存储为 excel 文件
	4.3.1 构建数据框
	4.3.2 查看数据框的常用属性
	4.3.3 按行列顺序访问数据框中的元素
	4.3.4 按行列名称访问数据框中的元素
	4.3.5 修改数据框中的元素
	4.3.6 删除数据框中的元素
	4.3.7 描述分析数据框中的元素
	4.4.1 转换成时间类型数据
	4.4.2 时间类型数据的常用操作
	4.5.1groupby 分组操作
	4.5.2agg 聚合操作
	4.6.1 生成透视表
	4.6.2 生成交叉表
	5 使用 Pandas 进行数据预处理
	5.1.1 表堆叠
	5.1.2 主键合并
	5.1.3 重叠合并

	<p>5.2.1 检测与处理重复值</p> <p>5.2.2 检测与处理缺失值</p> <p>5.2.3 检测与处理异常值</p> <p>5.3 标准化数据</p> <p>5.4.1 哑变量处理</p> <p>5.4.2 离散化连续型数据</p> <p>6 使用 scikit-learn 构建模型</p> <p>6.1.1scikit-learn 简介</p> <p>6.1.2 获取及认识 datasets 中的数据</p> <p>6.1.3 将数据集划分训练集和测试集</p> <p>6.1.4 利用转化器进行数据转化操作</p> <p>6.2 构建并评价聚类模型</p> <p>6.3 构建并评价分类模型</p> <p>6.4.1 构建回归模型</p> <p>6.4.2 评价回归模型</p>
Python 数据可视化	
<p>1 准备工作环境</p> <p>1.1Pyhon 数据可视化概述</p> <p>1.2Matplotlib 绘图库介绍</p> <p>1.3Matplotlib 散点图绘制</p> <p>1.4Matplotlib 绘制折线图</p> <p>1.5Matplotlib 绘制柱状图</p> <p>1.6Matplotlib 绘制饼图</p> <p>2 绘制并定制化图表</p> <p>2.1Matplotlib 例子的背景介绍</p> <p>2.2Matplotlib 例子：预处理</p> <p>2.3Matplotlib 例子：销售额随时间变化的可视化</p> <p>2.4Matplotlib 例子：星期与销售额关系</p> <p>2.5Matplotlib 例子：时间销售额与订单量的关系分析</p> <p>3 学习更多图表和定制化</p>	

3.1Pyecharts 简介与绘图逻辑说明	
3.2Pyecharts 绘制散点图	
3.3Pyecharts 绘制线图	
3.4Pyecharts 绘制饼图	
3.5Pyecharts 绘制柱状图	
3.6Pyecharts 图形组合	
4 创建 3D 可视化图表	
4.1Pyecharts 地理图表介绍	
4.2Pyecharts 绘制地理散点图	
4.3Pyecharts 绘制地理迁徙图	
4.4Pyecharts 绘制广东区域图	
5 用图像和地图绘制图表	
5.1 绘制微信好友性别分布饼图	
5.2 绘制微信好友地区分布地理图	

工作室邀请函

为了适应大数据与人工智能及发展的需求，顺应教育部提倡的深化校企合作的号召，更好的服务于广大数据分析爱好者，广东泰迪智能科技股份有限公司诚邀各高校相关专业老师、相关协会学会、俱乐部等组织合作成立“泰迪·智能工作室”，工作室以独立的模式运营，并以学生为中心成立，受泰迪科技监督且由其免费提供各种工作室所需资源的创新型数据智能工作室。



工作室旨在通过教育与产业之间的联动，实行“引进来，走出去”模式，引导学生学习数据科学与人工智能方法为导向，通过与企业的联系、合作、实践，激发学生的数据分析思维，全面推进数据分析与人工智能发展，提高大学生的数据分析素质，激发学生的创新创业精神，以实现创新型数据智能创业人才为培养目标。