

2023 年度“楚怡杯”湖南省职业院校技能竞赛

高职组电子信息类

“Python 程序开发”赛项

[时量：240 分钟，试卷号：A]

---

# 竞赛任务书

场次号：1 工位号：

2023 年 2 月

---

## 一、竞赛任务概述

本赛项包括“网络爬虫”、“数据清洗”、“数据分析与可视化”、“机器学习”4个竞赛任务，各任务分值分别为15分、30分、35分、20分，本赛项满分为100分。

## 二、注意事项

1. 请根据大赛所提供的竞赛环境，检查所列的硬件设备、软件清单、材料清单是否齐全，计算机设备是否能正常。

2. 竞赛结束前，在竞赛平台提供的虚拟机中，根据赛题将各试题代码进行完善整合，并运行；除了编写文档中提及到的函数，选手也可根据需要自定义函数实现功能。在竞赛平台左侧的答题区进行答题，并根据题目对运行代码及结果进行截图。

3. 竞赛结束时，请将答题区的答题报告进行提交操作，答题报告在竞赛结束前可重复提交；**未提交答题报告一律按0分处理。**

4. 选手进入赛场严禁携带移动存储器材，严禁携带手机等通讯工具，违者取消竞赛资格。

5. 在竞赛期间，选手扰乱赛场秩序，干扰裁判正常工作扣10分，情节严重者，经执委会批准，由裁判长宣布，取消竞赛资格。

## 三、竞赛环境

1. PC机：系统已安装Python相关环境、MySQL数据库、用户名密码分别为：root/123456。

2. 根据竞赛任务书说明，从竞赛平台虚拟机桌面获取项目工程代码包。桌面的工程代码可以直接使用虚拟机中的Pycharm导入、编译、运行和发布。

## 四、软件组件

1. Python编程语言及相关开发环境（Python、PyCharm）
2. Web框架（Django）
3. Python爬虫组件（Requests、lxml、BeautifulSoup）
4. Python数据分析组件（NumPy、Pandas）
5. Python可视化组件（Matplotlib、Pycharts）
6. Python机器学习组件（Scikit\_Learn）

## 五、赛题

# 模块一：网络爬虫（15 分）

## 第 1 题：抓取天气数据（15 分）

### 【任务说明】

互联网中的数据是海量的，通过人力操作进行数据采集低效繁琐，如何高效地获取数据源成为首要问题。本项目使用网络爬虫技术对数据信息进行采集，从“天气信息查询网站”中抓取天气数据，并将数据进行存储。

### 【任务要求】

在“TianTq”项目中的“appl”应用下的“spider”包中存在 crawl\_tq.py 文件，编写该文件用于从“天气信息查询网站”中抓取天气数据，该网站中包含青岛、开封、苏州、扬州、烟台、丽江、桂林、三亚、厦门、大理共 10 个城市的历史天气数据。

从网站中抓取 10 个城市的每日天气数据并分别保存到 CSV 文件中，CSV 文件存储到“TianTq”项目中的【appl/spider/day\_data/】目录下，若目录不存在，则需自行创建目录。

抓取的每日天气数据表字段见表 1。

表 1：每日天气数据说明表

表名	列名
城市名_day.csv (如“青岛_day.csv”)	城市、日期、最高气温、最低气温、天气、风向

### 【操作说明】

(1) 使用虚拟机中的谷歌浏览器访问“天气信息查询网站”，网站访问地址为【http://127.0.0.1:5000】，网站首页效果图如下：



图 1：“天气信息查询网站” 首页

(2) 点击城市标签跳转到天气历史记录页面。以“青岛”为例，“青岛历史天气” 页面展示如下图：

青岛历史天气				
<div>3°C</div> <div>平均高温</div>		<div>-1°C</div> <div>平均低温</div>		<div>9°C</div> <div>极端高温</div>
<div>100.0</div> <div>平均空气质量指数</div>		<div>24.0</div> <div>空气最好(01/05)</div>		<div>-14°C</div> <div>极端低温</div>
日期	最高气温	最低气温	天气	风向
2020-01-01 星期三	4°C	-3°C	晴转多云	西南风 2级
2020-01-02 星期四	6°C	1°C	晴	西北风 2级
2020-01-03 星期五	6°C	1°C	晴转多云	西北风 3级
2020-01-04 星期六	7°C	2°C	多云	西北风 2级
2020-01-05 星期日	7°C	3°C	多云转雨	东风 3级

图 2：“青岛历史天气” 页面

(3) 使用 PyCharm 打开桌面上的“TianTq” 项目进行编码。  
注意：实现功能后将结果截图粘贴到答题报告相应区域。

## 模块二：数据清洗（30 分）

### 第 2 题：清洗日表数据（30 分）

#### 【任务说明】

本项目用到的原始数据存储在 day.csv 文件中，该文件中包含了多个城市 2011 年 1 月 1 日至 2022 年 4 月 30 日的天气数据。原始数据中存在脏数据，不能直接用于后续数据分析，现需要对该表数据进行清洗操作，并将清洗之后的数据进行保存。

#### 【任务要求】

在“TianTq”项目中的“appl”应用下的“data\_clean”包中存在 clean.py 文件，编写该文件中的 deal\_day\_data() 函数用于实现日表数据清洗主体逻辑；日表数据 day.csv 文件存放在“TianTq”项目的【appl/data/】目录。

day.csv 数据表中包含字段见表 2。

表 2：日表数据字段

字段	字段含义
id	id 编号
city	城市名
date_week	日期及星期
hightest_tem	最高温度
lowest_tem	最低温度
weather	天气
wind	风向及风力级别

日表数据清洗要求如下：

（1）表中 date\_week、wind 字段都显示了两类数据，需要进行数据切分操作：

- date\_week 切分后列名设置为 date 和 weekday；
- wind 切分后列名设置为 wind\_direction、wind\_level。

（2）切分完成后分析数据表中的每个字段，字段的数据类型应符合该字段的实际意义。

（3）对数据表中的脏数据进行清洗操作，对于不同类型的脏数据处理方式不同，具体表现在：

- 重复值处理：每个城市同一天不能拥有重复的数据，如果有重复数据，只保留第一条数据。
  - 缺失值处理：每个城市不一定具有 2011 年 1 月 1 日至 2022 年 4 月 30 日中每一天的数据，清洗后必须满足每个城市都有全部天数的数据。对于数据中任意列的缺失值需要使用填充策略。填充策略如下：某城市某一天数据缺失，需要使用该城市其他年份该天的数据来填充。
    - 其中数字型数据采用平均值填充，数据四舍五入保留 2 位小数；
    - 类别型数据采用众数填充；
- (4) 清洗后的数据，每个城市需要按日期倒序排序，即最新的日期排在前面。
- (5) 数据清洗完成后保存到 CSV 文件中，该文件命名为 `clean_day.csv` 并存储到【`appl/data/cleaned_data/`】目录下，若目录不存在需自行创建。

## 【操作说明】

使用 PyCharm 打开桌面上的“TianTq”项目进行编码。

注意：实现功能后将结果截图粘贴到答题报告相应区域。

# 模块三：数据分析与可视化（35 分）

## 第 3 题：将数据写入 MySQL 数据库（5 分）

### 【任务说明】

将包含了多个城市的天气日表数据 `day_data.csv` 和月表数据 `month_data.csv`，通过 ORM 插入到本项目的数据库中。

### 【任务要求】

在“TianTq”项目中“appl”应用下的“create\_data”包中存在 `insert_data_to_db.py` 文件，编写该文件的 `insert_data()` 函数用于实现将数据插入数据库的主体逻辑。用于迁移的日表数据 `day_data.csv` 文件和月表数据 `month_data.csv` 文件存放在【`appl/data/`】目录下。

日表数据 `day_data.csv` 包含字段见表 3。

表 3: day\_data.csv 字段

字段	字段含义
city	城市名
highest_tem	最高温度
lowest_tem	最低温度
weather	天气
wind_direction	风向
wind_level	风力级别
weekday	星期

月表数据 month\_data.csv 包含字段见表 4。

表 4: month\_data.csv 字段

字段	字段含义
city	城市名
month	月份
avg_high_temp	平均高温
avg_low_tem	平均低温
extreme_high_tem	极端高温
extreme_low_tem	极端低温
avg_air_quality	平均空气质量指数
best_air	空气最好
best_air_date	空气最好日期
worst_air	空气最差
worst_air_date	空气最差日期

(1) 数据库中的数据表需要根据 day\_data.csv 和 month\_data.csv 文件中的字段创建模型类，配置数据库，然后迁移到 MySQL 的 tq 数据库中（若 tq 数据库不存在，需自行创建该数据库，后续题目如有用到数据库也是该数据库）。day\_data.csv 文件对应的表名为 day\_data，month\_data.csv 文件对应的表名为 month\_data。数据库用户名为 root，密码为 123456。

(2) 编码完成后直接运行 insert\_data\_to\_db.py 可将数据插入到对应的表中。

## 【操作说明】

使用 PyCharm 打开桌面上的“TianTq”项目进行编码。

注意：实现功能后将结果截图粘贴到答题报告相应区域。

## 第 4 题：绘制城市气温折线图（15 分）

### 【任务说明】

本题需要用到第 3 题写入数据库中的数据，对城市历年最高温度、最低温度走势分析，通过折线图进行可视化展示。

### 【任务要求】

在“TianTq”项目“app1”应用下的 views.py 中，编写 analysis\_city() 视图函数用于实现城市基本天气分析的功能，编写 index() 视图函数用于显示首页以及向首页传递必要的参数。

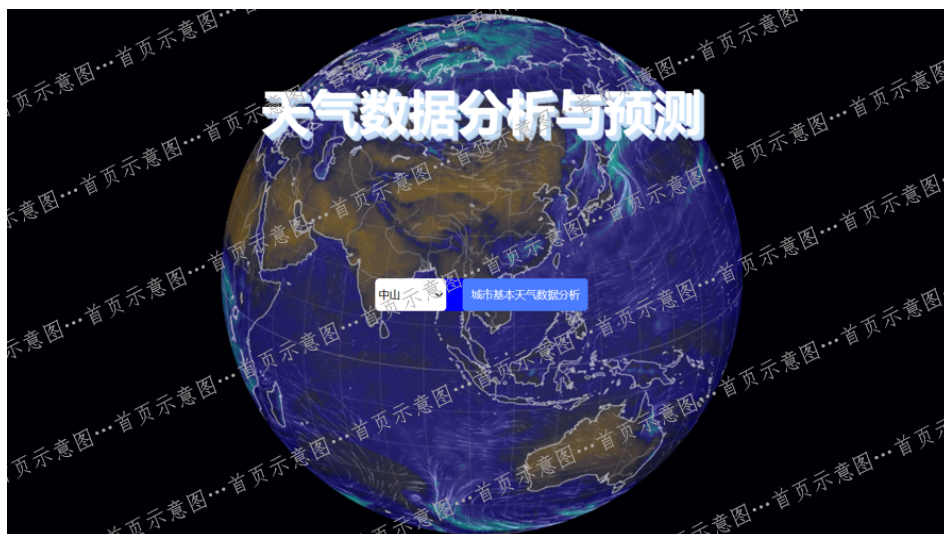


图 3：首页效果示意图

在“app1”应用下的“data\_visual”包内，编写 tq\_plot.py 文件进行封装绘图函数以供 analysis\_city() 函数调用。

(1) 分析某城市 2011 年 1 月 1 日至 2022 年 4 月 30 日来的最高温和最低温走势，并绘制折线图。

- 需要根据传递的城市参数，获取该城市的日期作为 x 轴数据，日期按升序排序（即最新日期排在最后），获取最高温和最低温作为 y 轴数据。
- 在 tq\_plot.py 中封装绘制折线图的函数，绘图功能需要通过 pyecharts 库来实现，pyecharts 相关的模板文件存放在【app1



- /templates/】目录下，可用于和 Django 框架进行整合。
- 需要将最高温和最低温折线图绘制到同一张图中。

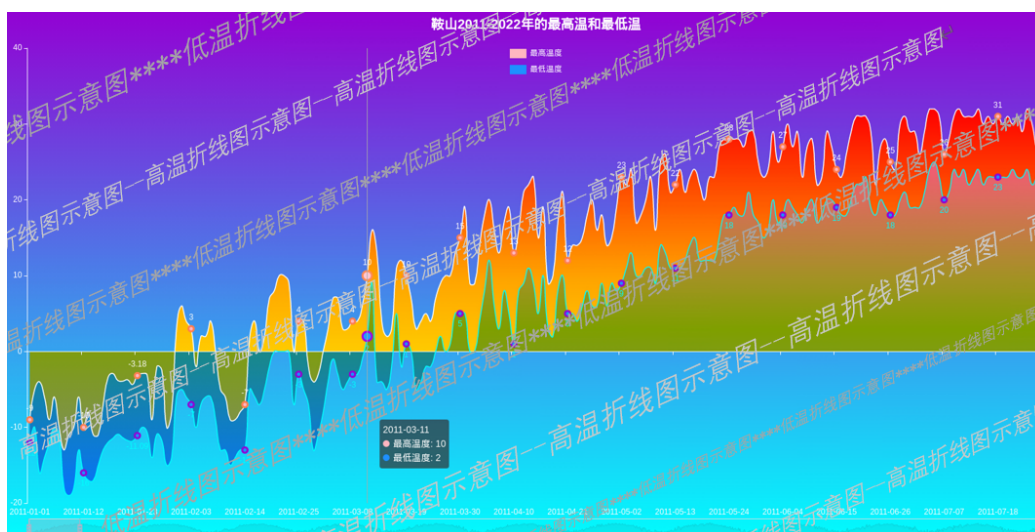


图 4：高温低温折线示意图

#### 绘制折线图的具体参数要求：

- ① 颜色设置：
  - 图的背景色采用渐变色从上至下为 darkviolet、cyan；
  - 高温折线填充渐变色从上至下为 red、orange、yellow；
  - 低温折线填充渐变色从上至下为 violet、green、blue；
- ② 初始配置：
  - 设置图片的宽度为 100%，高度 800px；
  - 网页标题设置为“某城市天气分析”，其中“某城市”根据选择城市自动匹配更新；
- ③ 最高温折线设置：
  - 显示数据标记点、标记设置为实心圆、标记大小为 8；
  - 标记点颜色设置为 lightpink，点的边框颜色设置为 coral，点的边框宽度设置为 3；
  - 线光滑、线颜色设置为白色，线宽 1.5；
  - 显示数据标签，数据标签颜色设置为白色，显示在数据上方；
- ④ 最低温折线设置：
  - 显示数据标记点、标记设置为实心圆、标记大小为 8；
  - 标记点颜色设置为 dodgerblue，点的边框颜色设置为 darkviolet，点的边框宽度设置为 3；
  - 线光滑、线颜色设置为 cyan，线宽 1.5；
  - 显示数据标签，数据标签颜色设置为 cyan，显示在数据底部；
- ⑤ 标题设置：
  - 标题设置为“某城市 2011-2022 年的最高温和最低温”，其中“某城市”根据选择城市自动匹配更新；
  - 标题显示为左右居中、距顶部 10px；
  - 标题字体大小设置为 20px、加粗显示；

- 标题字体颜色设置为白色;
- ⑥ 图例设置:
  - 图例设置为“最高温度”和“最低温度”;
  - 图例显示位置距顶部 7%, 垂直对齐;
  - 图例样式设置为方块;
  - 图例字体颜色设置为白色;
- ⑦ 坐标轴设置:
  - x 轴和 y 轴坐标线颜色为白色, 坐标刻度颜色为白色;
- ⑧ 区域缩放配置:
  - 缩放区域样式设置为滑块;
  - 数据起始百分比设置为 0%, 终止百分比设置为 5%;
- ⑨ 提示框组件配置:
  - 提示框显示信息设置为坐标轴触发;

(2) 网页标题为“某城市天气分析”, 其中“某城市”根据选择城市自动匹配更新, 每个图例之间的间隔为 1, 布局配置采用 Simple PageLayout。

(3) 编写 views.py 首页视图中的函数 index(), 实现从数据库中获取所有城市, 并进行其它必要配置(如首页模板文件路径和静态文件路径), 其中首页模板文件和依赖的静态文件放在项目下的 templates 目录和 static 目录中。

(4) 运行服务, 在虚拟机的谷歌浏览器中输入“http://127.0.0.1:8000/”访问首页。首页“城市基本天气分析”按钮前面的下拉菜单能显示所分析的所有城市, 选中任意一个城市, 然后点击“城市基本天气分析”按钮, 可跳转到一个新的页面, 该页面展示该城市气温折线图。

## 【操作说明】

使用 PyCharm 打开桌面上的“TianTq”项目进行编码。

注意: 实现功能后将结果截图粘贴到答题报告相应区域。

## 第 5 题: 绘制城市气温条形图 (15 分)

### 【任务说明】

本题需要用到第 3 题写入数据库中的数据, 对城市历年温度最高月份、温度最低月份、温度最高日期、温度最低日期等情况进行分析, 然后通过条形图进行可视化展示。

### 【任务要求】

在 tq\_plot.py 文件编写封装条形图的绘图函数。编写 views.py 的 analysis\_city() 视图函数用于实现城市基本天气分析功能, 调用 tq\_plot.py

中封装好的绘图函数绘制条形图。

(1) 分析某城市 2011 年 1 月 1 日至 2022 年 4 月 30 日来温度最高月份 Top10、温度最低月份 Top10、温度最高日期 Top10、温度最低日期 Top10 并绘制条形图。

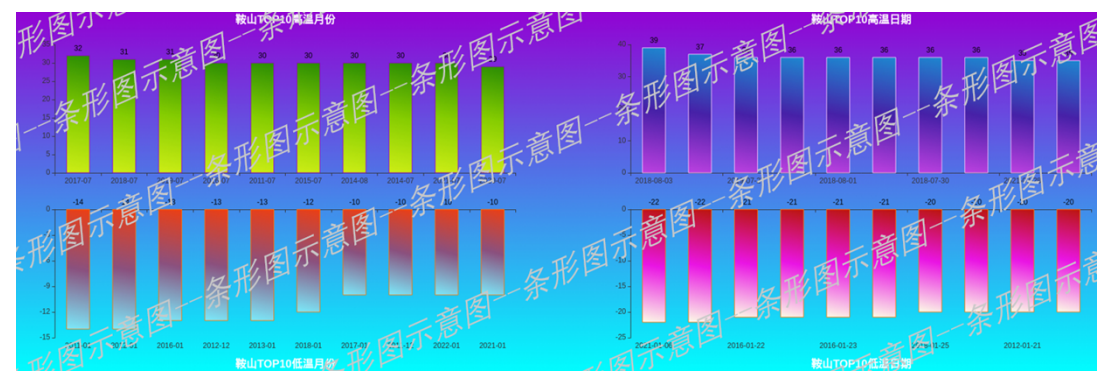


图 5：条形图示意图

- 通过并行多图将温度最高月份 Top10 条形图与温度最低月份 Top10 条形图组合起来；
- 将温度最高日期 Top10 条形图与温度最低日期 Top10 条形图组合起来。
- 需要根据传递的城市参数，获取该城市的 Top10 高温月份及其对应的温度、Top10 低温月份及其对应的温度、Top10 高温日期及其对应的温度、Top10 低温日期及其对应的温度。
- 设置月份和日期分别作为条形图的 x 轴数据，对应的温度作为条形图的 y 轴数据；
  - 通过平均高温字段分析高温月份；
  - 通过平均低温字段分析低温月份；
  - 通过最高温度字段分析高温日期；
  - 通过最低温度字段分析低温日期；
- 在 tq\_plot.py 中封装绘制条形图的函数，绘图功能需要通过 pyecharts 库来实现。

条形图具体要求如下：

- ① Top10 高温月份条形图参数如下：
  - 初始设置：宽 900px，高 500px；
  - 柱条设置：系列名为“最热月份”，柱子宽度设置为 50%，标签颜色设置为黑色；
  - 标题设置：标题为“某城市 TOP10 高温月份”，其中“某城市”根据选择城市自动匹配更新。标题位置设置为左右居中，距离顶部位置为 1%，标题颜色设置为白色、加粗显示；
  - 柱条样式设置：柱条颜色采用从左下至右上的渐变色为“#ceee16”、“#86cd01”、“#2a8c04”，柱条边框颜色 darkviolet，边框宽度为 1；
- ② Top10 低温月份条形图参数如下：
  - 初始设置：宽 900px，高 500px，

- 柱条设置：系列名为“最冷月份”，柱子宽度设置为 50%，标签颜色设置为黑色；
  - 标题设置：标题为“某城市 TOP10 低温月份”，其中“某城市”根据选择城市自动匹配更新。左右居中，距离顶部 95%，标题颜色白色、加粗；
  - 图例设置：不显示图例；
  - 柱条样式设置：柱条颜色采用从左下至右上的渐变色为“#7cecf0”、“#8b517f”、“#ed3f14”，柱条边框颜色“#ff7600”，边框宽度为 1；
- ③ 将 Top10 高温月份条形图与 Top10 低温月份条形图通过并行多图组合在一起，组合图形参数如下：
- 初始设置：宽 50%，高 600px，背景色采用从上至下渐变色为 darkviolet、cyan；
  - 高温月份条形图网格配置：底部 55%；
  - 低温月份条形图网格配置：顶部 55%；
- ④ Top10 高温日期条形图参数如下：
- 初始设置：宽 900px，高 500px，
  - 柱条设置：系列名为“最热日期”，条宽 50%，标签颜色为黑色；
  - 标题设置：标题为“某城市 TOP10 高温日期”，其中“某城市”根据选择城市自动匹配更新。左右居中，顶部 1%，标题颜色白色、加粗；
  - 图例设置：不显示图例；
  - 柱条样式设置：柱条颜色采用从左下至右上的渐变色为“#bd41e1”、“#4721a8”、“#1e88d1”，柱条边框颜色“#fccafd”，边框宽度为 1；
- ⑤ Top10 低温日期条形图参数如下：
- 初始设置：宽 900px，高 500px，
  - 柱条设置：系列名为“最冷日期”，条宽 50%，标签颜色为黑色；
  - 标题设置：标题为“某城市 TOP10 低温日期”，其中“某城市”根据选择城市自动匹配更新。左右居中，顶部 95%，标题颜色白色、加粗；
  - 图例设置：不显示图例；
  - 柱条样式设置：柱条颜色采用从左下至右上的渐变色为“#fdfee7”、“#eal5e4”、“#bb160c”，柱条边框颜色“#ff7600”，边框宽度为 1；
- ⑥ 将 Top10 高温日期条形图与 Top10 低温日期条形图通过并行多图组合在一起，组合图形参数如下：
- 初始设置：宽 50%，高 600px，背景色采用从上至下渐变色为 darkviolet、cyan；
  - 高温月份条形图网格配置：底部 55%；
  - 低温月份条形图网格配置：顶部 55%；

(2) 将第 4 题绘制的折线图、第 5 题绘制的 2 个组合的条形图通过顺序多图组合起来，组合后的图形网页标题仍为“某城市天气分析”，其中“某城市”

根据选择城市自动匹配更新，每个图例之间的间隔为 1，布局配置采用 Simple PageLayout。

(3) 完善首页视图函数 `index()` 获取所有城市，并进行其它必要配置，其中首页模板文件和依赖的静态文件已经放在项目下的 `templates` 目录和 `static` 目录中。

(4) 运行服务，在虚拟机的谷歌浏览器中输入 “`http://127.0.0.1:8000/`” 访问首页。首页 “城市基本天气分析” 按钮前面的下拉菜单能显示所分析的所有城市，选中任意一个城市，然后点击 “城市基本天气分析” 按钮，可跳转到一个新的页面，该页面显示的是我们前面通过顺序多图组合的折线图和条形。

## 【操作说明】

使用 PyCharm 打开桌面上的 “TianTq” 项目进行编码。

注意：实现功能后将结果截图粘贴到答题报告相应区域。

# 模块四：机器学习（20 分）

## 第 6 题：特征提取（8 分）

### 【任务说明】

用于做机器学习的数据存储在 `machine_data.csv` 文件，该数据包含了几个城市 2014 年 1 月 1 日至 2022 年 4 月 30 日的天气数据，但不是所有特征用于训练机器学习模型都能取得好的效果，现在需要从该表中按要求提取相应的特征，以供后续建立机器学习算法模型使用。

### 【任务要求】

在 “TianTq” 项目 “`apl`” 应用下的 “`machine_learning`” 包中存在 `train_model.py` 文件，编写该文件的 `feature_extraction()` 函数实现提取数据特征功能，用于机器学习建模的数据文件 `machine_data.csv` 存放在

【`apl/data/`】目录。

- (1) 需要根据传递的城市参数，提取该城市用于机器学习建模的特征。
- (2) 进行数据特征提取时：
  - 输入为前面七天的最高温、最低温、天气、风向、所属月份。
  - 输出为该日的最高温、最低温、天气。
  - 除此之外，还需要一列城市(`city`)作为数据的标志；

举例说明：假设当前日期为 2 月 10 日，则输入为其前七天（2 月 9 日-2

月 3 日) 对应的最高温、最低温、天气、风向、所属月份, 特征命名按时间倒序排序, 对应输入特征命名参考见表 5。

表 5: 列名命名规范

数据类型	日期	特征 (列名)
特征值	2 月 9 日	day1_high_tem、day1_low_tem、day1_weather、day1_wind、day1_month
	2 月 8 日	day2_high_tem、day2_low_tem、day2_weather、day2_wind、day2_month
	2 月 7 日	day3_high_tem、day3_low_tem、day3_weather、day3_wind、day3_month
	2 月 6 日	day4_high_tem、day4_low_tem、day4_weather、day4_wind、day4_month
	2 月 5 日	day5_high_tem、day5_low_tem、day5_weather、day5_wind、day5_month
	2 月 4 日	day6_high_tem、day6_low_tem、day6_weather、day6_wind、day6_month
	2 月 3 日	day7_high_tem、day7_low_tem、day7_weather、day7_wind、day7_month
标签值	2 月 10 日	cur_high_tem、cur_low_tem、cur_weather
城市	/	city

所以提取特征之后的数据是一个:

[ ‘city’, ’ day1\_high\_tem’, ’ day1\_low\_tem, ……’, ‘cur\_high\_tem’, ‘cur\_low\_tem’, ‘cur\_weather’ ]共 39 列的 DataFrame, 该数据将用于后续的机器学习;

注意: 提取的数据顺序必须先提取最新日期的, 比如长沙市的最新日期是 2022 年 4 月 30 日, 则提取特征之后 DataFrame 中的第一条数据必须对应 2022 年 4 月 30 日这天的数据 (包括当天的温度天气以及它过去 7 天的温度、天气等信息)。

(3) 在提取特征之后, 将 “city” 列为 “长沙” 的 DataFrame 的 info()、describe() 打印到控制台。

(4) 并将 DataFrame 的内容保存到 CSV 文件中, 命名为 “changsha\_feature. csv”, 存储到【appl/data/feature/】目录, 若目录不存在需自行创建。

## 【操作说明】

使用 PyCharm 打开桌面上的“TianTq”项目进行编码。

注意：实现功能后将结果截图粘贴到答题报告相应区域。

## 第 7 题：模型训练（12 分）

### 【任务说明】

在提取特征之后，需要构建机器学习算法模型对最高温度、最低温度进行预测。由于每个城市的温度差异很大，没有足够的规律性可言，因此不能通过一个统一的模型对多个城市进行预测，每个城市只能单独去训练模型；

又由于每个需要预测的目标值影响它们的因素有差异，所以不能使用一个模型同时预测这两个目标值，而应该分别创建模型进行预测。在创建模型的时候，每种模型所需要的特征也是不一样的。现需要基于第 6 题的结果来训练机器学习算法模型。

### 【任务要求】

第 6 题实现功能后，使用提取到的特征数据（共 39 个特征）构建模型进行训练。在“appl”应用下的“machine\_learning”包中存在 train\_model.py 文件，编写该文件中的 train\_all\_model() 函数用于训练一个城市的最高温、最低温的机器学习模型。

#### （1）训练高温预测模型

- 训练高温预测模型需要的特征数据为：前七天的最高温、天气、风力、所属月份，总共  $4 \times 7 = 28$  列特征值，目标值为当天的最高温。高温预测模型特征字段见表 6。

表 6：高温预测模型特征与目标值

数据类型	特征（列名）
特征值	day1_high_tem、day1_weather、day1_wind、day1_month、day2_high_tem、day2_weather、day2_wind、day2_month、day3_high_tem、day3_weather、day3_wind、day3_month、day4_high_tem、day4_weather、day4_wind、day4_month、day5_high_tem、day5_weather、day5_wind、day5_month、day6_high_tem、day6_weather、day6_wind、day6_month、day7_high_tem、day7_weather、day7_wind、day7_month
标签值	cur_high_tem

- 由于数据中既有类别型数据，又有数字型数据，因此需要在输入机



器学习模型之前对数据进行编码操作，每一列应具有一个独立的编码器。编码规则如下：数据为整数或者整数型字符串直接编码为整数；字符串则通过标签编码器进行编码；

- 数据集划分方式：编码完成之后，需要划分训练集和测试集，测试集比例为 0.2，随机种子选择 7。
- 模型训练规则：自己选择模型（模型种类不限），使用训练集对模型进行训练，使用测试集 r2\_score 分数对模型进行验证，保存最优模型。
- 模型训练完成后，将最终模型进行保存，模型保存到【app1/machine\_learning/model\_ckpt/】目录下（若目录不存在需自行创建目录），命名为“城市名\_high\_model.pkl”，例如长沙的高温模型名为“长沙\_high\_model.pkl”。

## （2）训练低温预测模型

- 训练低温预测模型需要的特征数据为：前七天的最低温、天气、风力、所属月份，总共 28 列，目标值为当天的最低温。低温预测模型特征字段见表 7。

表 7：低温预测模型特征与目标值

数据类型	特征（列名）
特征值	day1_low_tem、day1_weather、day1_wind、day1_month、day2_low_tem、day2_weather、day2_wind、day2_month、day3_low_tem、day3_weather、day3_wind、day3_month、day4_low_tem、day4_weather、day4_wind、day4_month、day5_low_tem、day5_weather、day5_wind、day5_month、day6_low_tem、day6_weather、day6_wind、day6_month、day7_low_tem、day7_weather、day7_wind、day7_month
标签值	cur_low_tem

- 编码规则、训练集划分方式以及模型训练规则与高温模型的方式一致。
- 训练模型后需要保存最终模型，模型保存到【app1/machine\_learning/model\_ckpt/】目录下，命名为“城市名\_low\_model.pkl”，例如长沙的低温模型名为“长沙\_low\_model.pkl”。

## 【操作说明】

使用 PyCharm 打开桌面上的“TianTq”项目进行编码。

注意：实现功能后将结果截图粘贴到答题报告相应区域。