***题目***

雾霾的监控和预测是现代工业社会越来越重视的一个问题，我们希望能够从大量的pm2.5监控数据来得到一个模型，使得它能够根据某些条件来判断pm2.5的大小，下面我们给出北京2010年到2015年的pm2.5观测数据（见数据包）

数据包中train.csv是带有PM\_US Post属性的训练集，test.csv则不带PM\_US Post，我们的模型需要在test.csv上做出预测，预测出相应数据条目的PM\_US Post 属性

我们的目标是训练出一个能够根据某些条件（位置，湿度，风向等等）来判断pm2.5值的模型

***数据字典***

*Year 年份*

*Month 月份*

*Day 日期*

*Hour 小时*

*Season 季节*

*PM\_Dongsi 东四pm（ug/m^3）*

*PM\_Dongsihuan 东四环pm（ug/m^3）*

*PM\_Nongzhanguan, 农展馆pm（ug/m^3）*

*DEWP 露点温度（Celsius）*

*HUMI 湿度（%）*

*PRES 气压（hpa）*

*TEMP 温度（Celsius）*

*Cbwd 结合风向*

*Iws 累计风速（m/s）*

*Precipitation 降水（mm/hour）*

*Iprec 累计降水（mm）*

*PM\_US Post PM发布值（要预测的属性）（ug/m^3）*

*No 编号*

**提交结果**

提交各式见example.csv

**比赛流程**

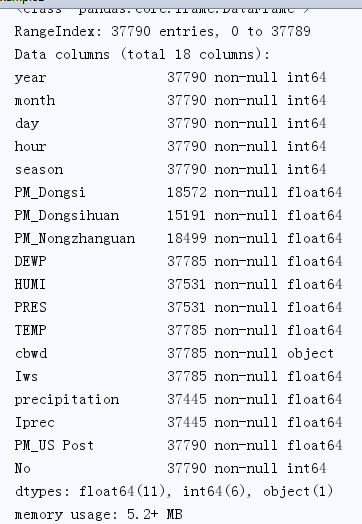
1. 报名记录在档，领取数据包（可以组队也可以个人）
2. 在比赛期间，可以多次提交结果（每天不超过三次），并给出评分，按评分排名
3. 比赛结束后将请前三名进行答辩

***暂定截止日期为2018年一月7号***

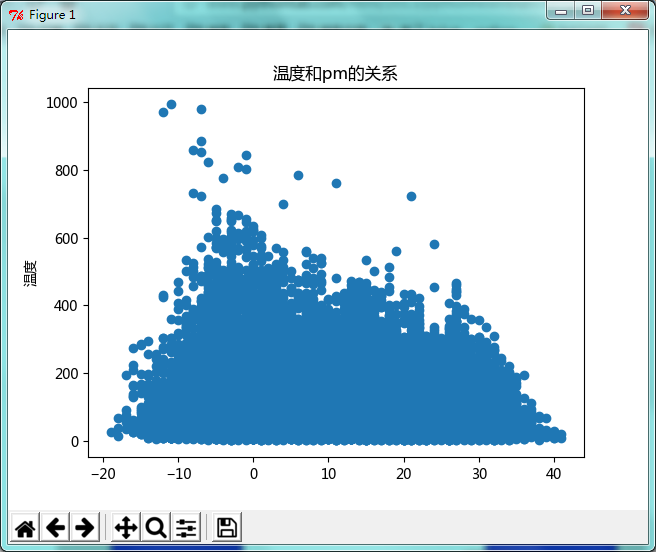
**比赛须知**

1. 严禁作弊，一旦作弊取消资格
2. 严禁利用外部数据源，一旦利用取消资格
3. 科技中心会根据评分异常变动进行质询，解释不了的将视作弊处理

统计信息：

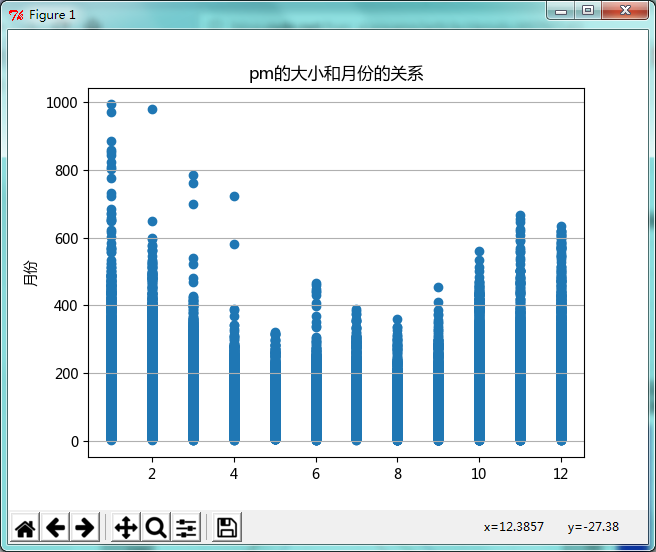


室内温度和pm的关系

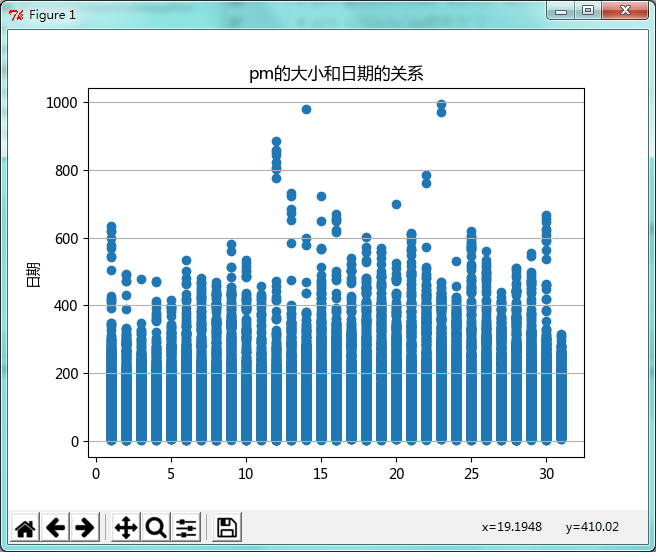




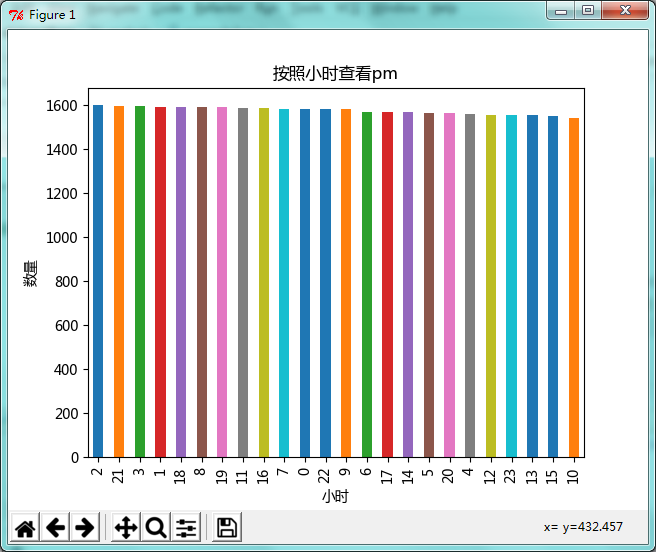
月份和pm的关系

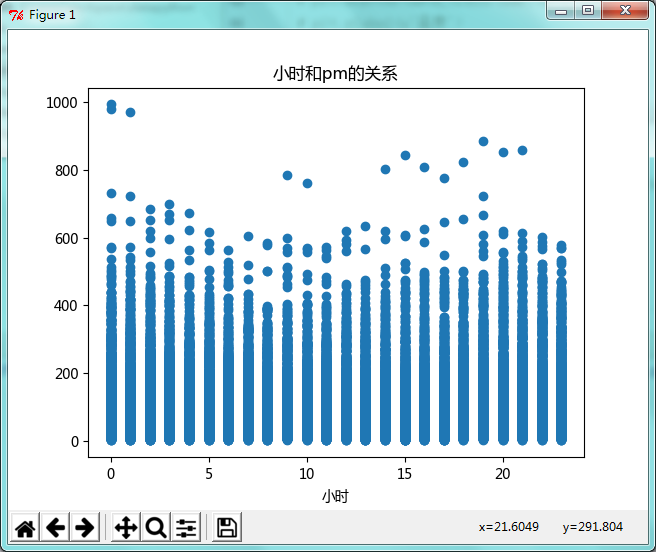


日期和pm的关系



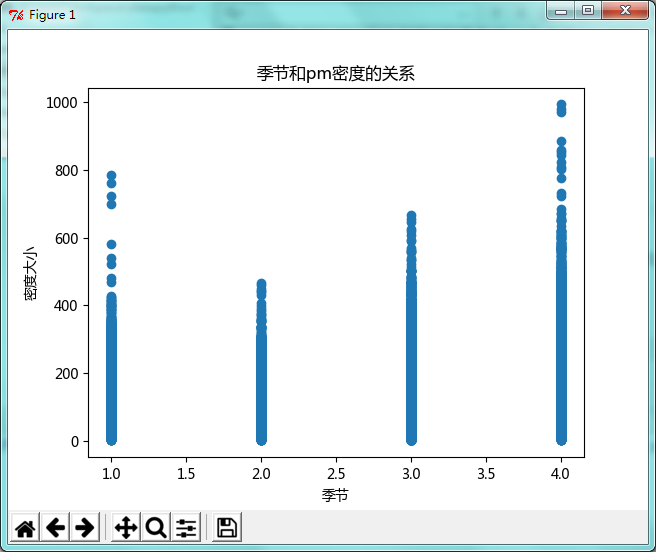
小时与pm的关系：（发现没多大关系在数量方面，反而在大小方面有点点）



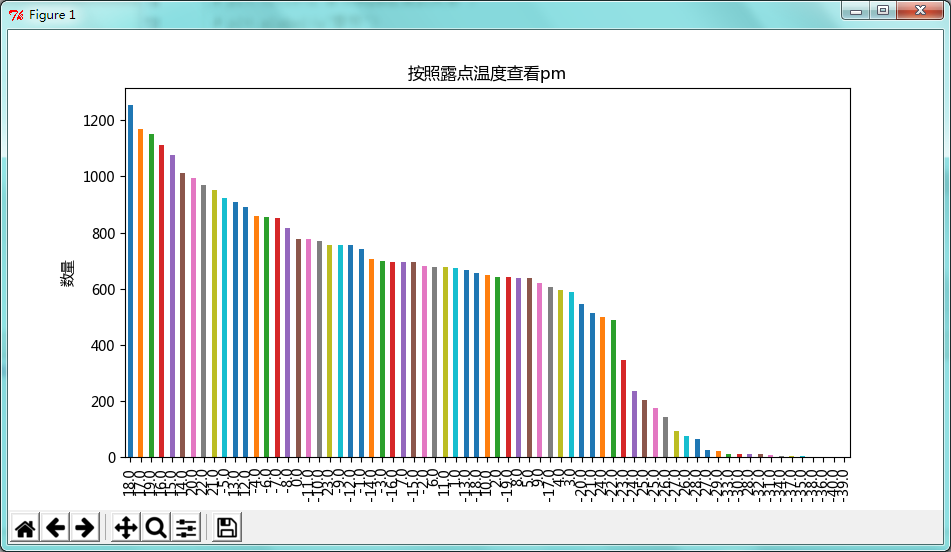


季节与pm数量没什么关系

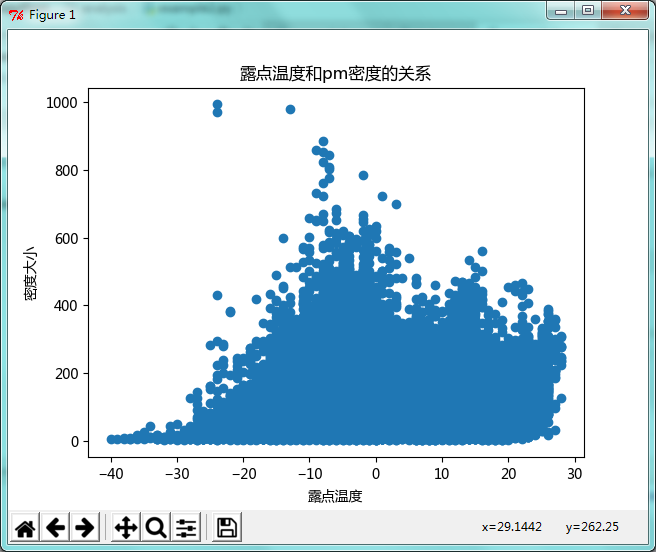
与密达大小有关系



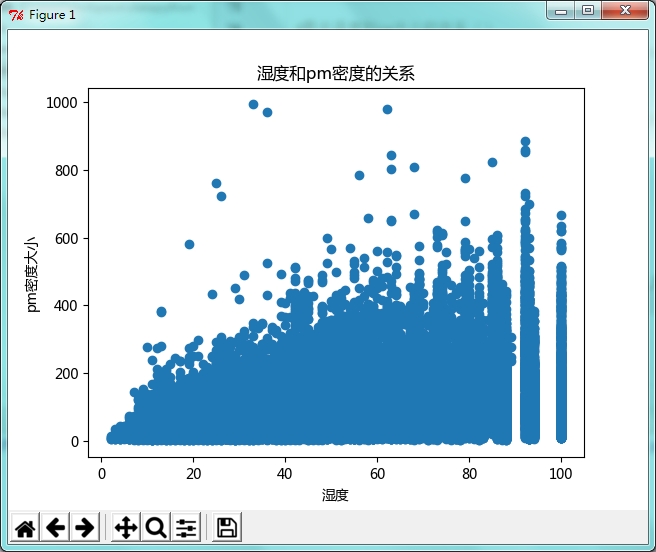
露点温度（*DEWP*）和数量的关系（发现温度越高，数量越多）



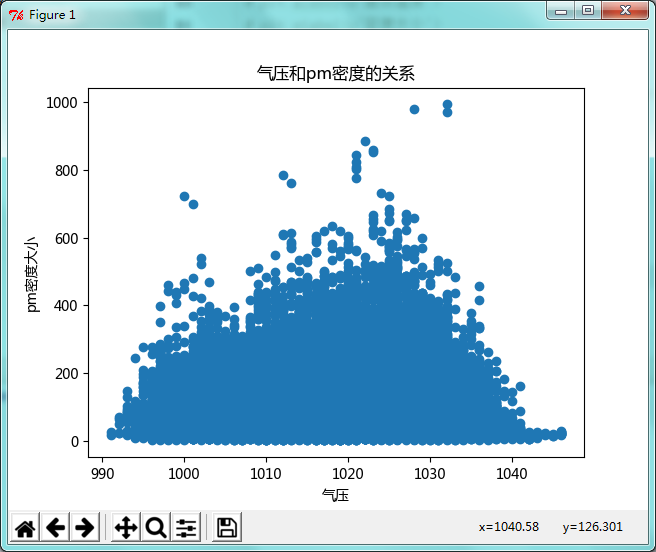
露点温度(*DEWP*)和pm大小的关系（-10度的pm密度最大了）



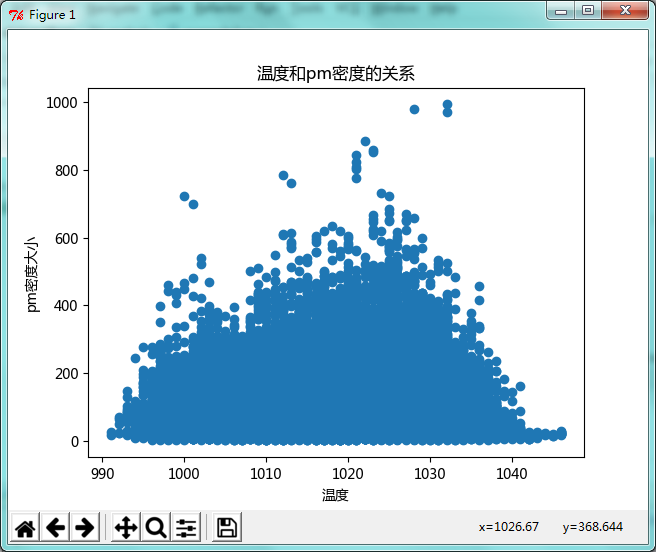
湿度（HUMI）和pm密度的关系（大致一看，还是有关系的）



气压（PRES）和pm密度的关系（有关系的）

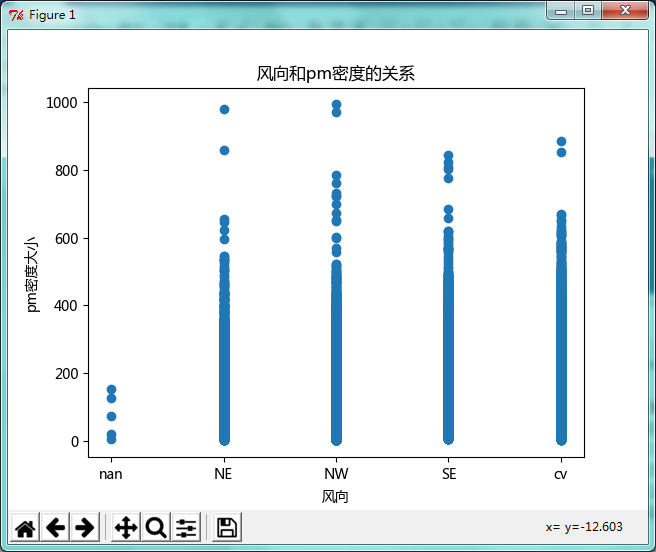


温度（TEMP）和pm密度大小的关系

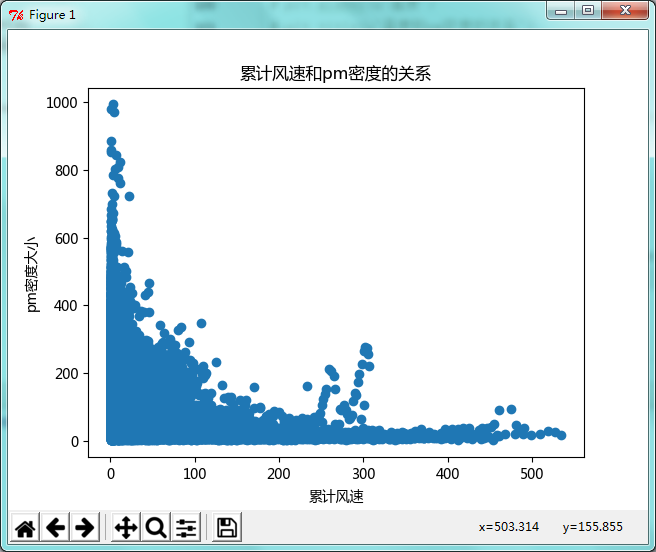


风向（cbwd）和pm密度大小关系

解释：在[气象观测](https://www.baidu.com/s?wd=%E6%B0%94%E8%B1%A1%E8%A7%82%E6%B5%8B&tn=44039180_cpr&fenlei=mv6quAkxTZn0IZRqIHckPjm4nH00T1Y3uW7bnvwBmWIBPWnLP1P-0ZwV5Hcvrjm3rH6sPfKWUMw85HfYnjn4nH6sgvPsT6KdThsqpZwYTjCEQLGCpyw9Uz4Bmy-bIi4WUvYETgN-TLwGUv3EPj6sPW0snjns)中，风向是指风的来向，地面风向用十六个方位表示，在气象服务过程中一般用汉语表述，在专业领域用英文字母表示，分别是：北（N）、东北东（NNE）、东北（NE）、东东北（ENE）、东（E）、东东南（ESE）、东南（SE）南东南（SSE）、南（S）、南西南（SSW）、西南（SW）、西西南（WSW）、西（W）、西西北（WNW）、西北（NW）、北西北（NNW）。根据风向的定义，从方位角在337.5°±11.25°范围内吹来的风的风向都记为NNW。

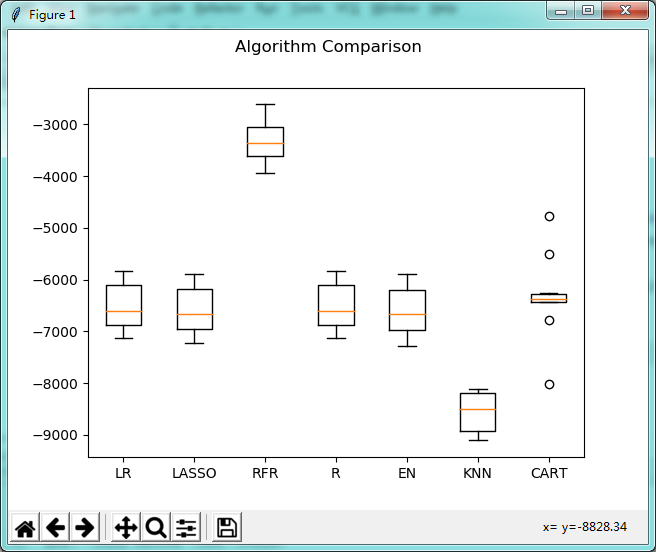


累计风速（Iws）和pm关系（关系很大的）



最高分0.8327

算法评估：

、

LR: -6506.592447 (433.642027)

LASSO: -6578.228075 (442.046748)

RFR: -3328.985998 (385.734026)

R: -6506.614502 (433.756322)

EN: -6592.754206 (445.635767)

KNN: -8567.623211 (389.177018)

CART: -6321.760254 (787.967718)

分别是准确度和标准差

正态化后的结果：

ScalerLR: -6506.592447 (433.642027)

ScalerSASSO: -6528.055346 (442.459249)

ScalerRFR: -3410.630120 (337.027993)

ScalerR: -6506.561337 (433.651710)

ScalerRN: -6861.057217 (491.597953)

ScalerKNN: -4962.232203 (355.337263)

ScalerCART: -6296.513511 (959.541584)