注：

1. 具体开始时间和提交时间可由学工那边决定，比赛时间尽量长一些，个人建议15-20天）
2. 如果需要，我可以在比赛开始前或比赛开始的前两天为参赛队伍做一个有关数据分析和比赛思路的报告。

## “共享单车”大数据分析比赛

1. 比赛类别

共享单车是一种新型的交通工具租赁业务-自行车租赁业务，其主要依靠载体为自行车，用户可以很容易的的从特定位置租借一辆自行车并在另一位置归还。由于共享单车可以充分利用城市因快速的经济发展而带来的自行车出行萎靡状况，最大化利用公共道路通过率，同时起到健康身体的作用，因此存在着巨大的效益。而与此同时，共享单车的的投放量、投放位置等问题也成为数据分析研究领域的一个热点。

本次比赛以美国华盛顿2011和2012两年的共享单车系统历史记录数据为样本，分析共享单车租赁数与季节、气候等因素的相关性，并对test测试集中的样本进行共享单车租赁数的预测（包括临时用户数、注册用户数及总用户数）。

1. 参赛要求
2. 竞赛接受单位组队参赛，每队由1-3人组成，可设指导教师一名。
3. 参赛过程必须遵守我国相关的法律、法规以及道德规范。
4. 各队需基于train训练数据集构建预测模型，并使用构建的模型对test测试集中的三类共享单车租赁数进行预测（包括临时用户租赁数、注册用户租赁数及总数）。比赛过程中的特征选择方法、预测模型算法、编程语言均不限。
5. 训练及测试数据集格式参见readme.txt文件。
6. 在比赛规定时间内将包含三类共享单车租赁数预测结果的test\_day.csv和test\_hour.csv文件提交到指定地点，同时准备PPT进行汇报答辩。
7. 评审原则

本着公开、公平、公正的原则评审参赛作品。

1. 模型预测成绩评定

根据每队提交的test\_day.csv和test\_hour.csv文件中对三类共享单车租赁数的预测值，依据预测值与真实值之间的均方误差MSE（Mean Squared Error）指标对各队成绩进行评定。

1. 汇报答辩成绩评定

每队在提交预测结果后，还需以ppt现场答辩方式对比赛过程中的数据分析思路进行介绍，由评委根据每队现场表现、讲解以及团队合作的情况给予评分。

1. 总成绩评定

最终根据模型预测成绩和汇报答辩成绩以4:6的权重确定比赛总成绩并予以公布。