尊敬的各位老师：

大家好！我是黄驰，下面由我来对我们的创新实践项目进行答辩。

我们的项目名称是基于出租车GPS数据的城市交通统计研究。

我将从以下六个方面进行答辩，分别是选题意义，当前进度，部分数据处理结果与可视化，论文初稿撰写，后期工作安排，创新模型。

1. 选题意义。随着城市的发展，交通数据的规模越来越庞大，如何利用它们成为重要问题。对北京出租车GPS数据进行归纳分析，可以为城市交通规划提供依据。通过分析，可以构建医疗推荐系统。我们建立的创新模型，可以广泛应用到其他类似问题上。
2. 当前进度。从项目开展到现在：我们查阅并学习了大量相关资料，研读领域专家论文数十篇，并将学习成果汇总；与北京城市规划院进行了合作，获取了充足精准的权威数据，我们有过去几年内北京市所有出租车GPS数据，为我们项目的顺利进行做好了数据准备。目前我们用部分数据进行试验，该数据条数共计14380000，预计使用正式数据时，数据的数量级将达到十亿以上，结果将会更精确。 下面是对数据的总体统计分析。我们已熟练掌握Python语言以及Numpy、Pandas和Matplotlib等工具，目前数据处理的代码量已接近千行。
3. 部分结果展示。获取数据初始，每一列代表的意义需要我们自己去判断。我们采用了多种方法并结合实际，比如绘制了散点图，分布图。例如下图是一天当中速度的散点图。然后对原始数据进行过滤，得到了多个不同筛选条件下的数据集。然后得到数据总量、车辆数等统计量。再分析得到所有的OD点，下面将会基于这个步骤得到的数据集进行。接下来对OD点进行了可视化处理。下面展现部分本阶段绘制的图表。
4. 论文初稿撰写。组员之间每周一次研讨并向老师汇报，在充分地探讨与沟通中，构建了论文框架，经过不断地完善、修改错误、增加创新点，已经形成了思路清晰、较为完善的论文体系。以下是我们的中英文论文摘要与提纲。
5. 后期工作安排。我们编写的程序能够在任何规模的数据集上运行，在对正式数据分析后就能完成论文的撰写。我们还创新地提出了一个基于概率的模型，目前数学推导已经完成，相关的代码正在编写中。之后我们将对模型的有效性进行分析和探讨并撰写相关论文。我们还初步提出了一个更加创新的基于强化学习的模型，但这个模型暂时具有缺陷，还需要进一步的优化，有关这部分的探究将放所有其他工作完成之后再进行。
6. 最后我简单介绍一下模型。这个模型的目标是分析去某个医疗机构的所耗时间，下面叙述该模型的大致流程。我们将这个耗时看作一个随机变量，挑选出会影响耗时的各个因素，将这些因素组合成为一个向量，作为单辆车的特征参数。首先定义了分布函数如下，然后可以概率分布律:

我们可以有很多种方法来求解其中的未知参数。我们将分别编写程序并进行对比。当前所有的模型和算法都已经设计完毕，程序正在编写中，运行后就能完成。

在项目进行过程中，指导老师认真负责，组员分工清晰明确，团队有较强的执行力和学习能力，项目进行过程中按时且高质量的完成了每个阶段的任务。认为可以按照预期计划进度和任务完成要求。我的汇报完毕，请各位老师批评指正，谢谢！（鞠躬）