arXiv:2305.01896v1 [eess.SP] 3 May 2023

”为匹兹堡大学学生优化公交线路选择：对乘客量、准点率和行驶距离的比较分析”

Logan Warren（主要作者）
匹兹堡，美国 *l*
aw234@pitt.edu

Jerek Stegman（贡献者）匹
兹堡，美国 *jjs252@pitt.edu*

摘要—**摘要——本文介绍了一项研究项目，利用从宾夕法尼亚州西部区域数据中心获取的地理数据，探究匹兹堡公交线路的地形特征。本研究重点比较公交线路的使用情况，以找出乘客量最少的线路，并比较公交线路的平均准点率，从而确定从匹兹堡大学前往匹兹堡市中心最中心公交站的线路。我们将使用 RStudio 等计算软件分析数据之间的相关性，并得出结论，确定乘客最少、停靠站最少且耗时最短的线路。本研究为匹兹堡大学计算与信息学院（SCI）的课程 INFSCI 310——信息科学计算而开展。关键词——公交线路、地形特征、地理数据、R 编程语言**

索引词—**“公共交通、公交线路、匹兹堡大学、数据分析、乘客量、准点率、R 编程语言、Python、Pandas。”**

I. 引言

本研究项目旨在确定从匹兹堡大学前往匹兹堡市中心最中心位置公交站的最佳路线。我们的结论基于对四个关键数据集的分析：按线路划分的公交站使用情况、按路线和工作日划分的月平均乘客量、匹兹堡区域交通公交站点，以及按线路划分的月度准点率（OTP）。选择这些数据集，是因为它们能够全面反映公交使用情况、运行时间和站点位置，使我们能够找出这两地之间最高效的路线。我们希望借助这些信息，确定从大学前往市中心的最佳路线，并重点考量三个主要因素。首先，我们希望找出乘客量最少的公交线路。其次，我们希望利用数据确定前往市中心中心点的最短路线。最后，我们将参考准点率数据，考量公交车的准点性，以确定最快的路线。

匹兹堡的公共交通系统在城市日常生活中发挥着至关重要的作用，许多居民，包括学生，都依赖它通勤。阿勒格尼县港务局运营着一个公交网络，服务匹兹堡都会区的轻轨和缆索铁路。匹兹堡大学的公共交通服务完善，多条公交线路经过校园，并将其与城市的其他区域连接起来。然而，对于初来乍到的人来说，熟悉这张庞大的交通网络可能颇具挑战，尤其是对刚来到这座城市、尚不熟悉其交通系统的学生而言。

根据2022年的数据，匹兹堡大学约40%的本科生来自州外，人数约为7,000名。对于这些学生来说，初到这座城市并了解其公交系统可能既令人困惑又充满挑战。作为这一群体的一员，我们能够证明自己在匹兹堡寻找方向并按时抵达目的地时所面临的困难。匹兹堡大学位于市中心步行距离之外，因此当学生需要交通工具时，他们通常会选择公共公交系统。然而，许多学生并不确定该乘坐哪路公交车、哪条路线能最快抵达目的地，以及哪辆公交车上的乘客最少。

从匹兹堡大学前往匹兹堡市中心，确定最佳公交线路对于学生的出行和学业成功至关重要。我们的目标是确定最优线路，即停靠站最少、准时到达，并能满足学生的各种需求，例如周末外出、实习以及往返学校通勤。此信息将为学生提供一种可靠且高效的交通出行选择。

为了分析和比较所选数据集中的数据，我们将使用 Python 和 Pandas 进行数据操作与分析。Pandas 是一种广泛使用的 Python 数据操作库。我们还将重点使用 R（一种用于统计计算的编程语言）进行额外的数据操作，并将研究结果以可视化形式呈现。通过使用这两种编程语言，我们将生成各种指标，以帮助确定从匹兹堡大学前往匹兹堡市中心的最佳路线。

II. 项目目标

本项目的主要目标是分析来自宾夕法尼亚州西部区域数据中心的四个数据集，并找出从匹兹堡大学前往匹兹堡市中心最中心的公交站的最高效公交线路。这一目标至关重要，因为它能为学生提供一种可靠、便利的交通选择，方便他们出于各种原因前往市中心，同时兼顾此前提到的三个关键因素。本研究旨在回答四个主要问题：

- 考虑到沿途频繁停靠，哪条路线能最快到达匹兹堡市中心的中心点?
- 哪条公交线路的乘客量最低，能够确保快速且舒适的乘车体验?
- 哪条公交线路的准点率最高?
- 根据前面问题收集的信息，从匹兹堡大学前往匹兹堡市中心的最佳综合路线是什么?

基于我们的研究，我们预计，从匹兹堡大学前往市中心的最快路线很可能会沿着盟军大道行驶，因为它提供了一条从大学直达市中心的路径。至于乘客量最低的路线，我们预计这将是一条不经过第五大道、而是穿行于其他道路或公交专用车道的公交线路，并避开校园周边乘客经常上车的主要站点。

为了实现我们的项目目标，我们将实施以下步骤：

1) 从宾夕法尼亚州西部区域数据中心获取并预处理数据集。2) 使用 Python 和 Pandas 进行数据处理和初步分析。3) 使用 R语言进一步处理数据，并将研究结果以可视化形式呈现。4) 分析数据集，并根据三个关键因素对公交线路进行比较：乘客量、准点率和行驶距离。5) 根据分析结果，确定从匹兹堡大学前往匹兹堡市中心的最佳综合公交线路。6) 展示研究结果并提出建议，帮助学生在选择公交线路时做出明智的决定。

通过对现有数据进行全面分析，并明确了解项目目标，我们旨在为匹兹堡大学的学生提供有价值的信息，帮助他们选择最适合通勤前往匹兹堡市中心的公交线路。本项目不仅有助于学生做出更加明智的选择，还将进一步提升他们在城市公共交通系统中的整体出行体验。

III. 项目方法与结果

A. 方法

为实现我们的研究目标，我们采用了以下方法和统计检验：

1) **数据收集**：我们从宾夕法尼亚州西部区域数据中心收集了四个数据集，包括按线路划分的公交站使用情况、按线路和工作日划分的月均客运量、匹兹堡区域交通公交站点，以及按线路划分的月度准点表现（准点率）。2) **数据处理**：我们使用 Python 和 Pandas 清理并预处理了这些数据集，移除了可能影响分析的不一致数据、异常值或缺失值。 3) **数据分析**：我们使用 R 应用统计方法和数据可视化技术来考察这些数据集，重点分析每条公交线路的停靠站数量、乘客量和准点率之间的关系。4) **相关性分析**：我们进行了相关性分析，以确定这三个因素之间的权衡关系，并综合考虑各个方面，找出表现最佳的公交线路。 5) **线路选择**：根据分析结果，我们确定了从匹兹堡大学前往匹兹堡市中心最中心的公交站的最佳公交线路。

B. 结果

我们使用 Python、Pandas 和 R语言进行了相关性和回归分析，以确定每条公交线路的停靠站数量、乘客量和准点率之间的关系。通过这项分析，我们能够更好地理解这三个因素之间的权衡，并确认18路在停靠站数量更少、车厢拥挤程度更低以及准点率方面是最佳线路。

1) **数据集分析**

a) 按线路划分的公交站使用情况：我们利用该数据集确定每条公交线路的公交站点数量。这些信息帮助我们找出匹兹堡大学与市中心之间停靠站较少的线路。我们发现，18路的停靠站数量少于停靠站最多的前10条线路，表明其行程速度更快。b) 按线路和工作日划分的月度平均客流量：我们分析了乘客量数据，以找出乘客数量最少的线路，这将带来更加不拥挤、更加舒适的乘车体验。18路的平均客流量最低，排名第四，表明它是最不拥挤的线路之一。c) 匹兹堡区域交通公交站点：该数据集提供了所有公交站点的位置，我们利用这些信息确定了匹兹堡市中心最靠近中心位置的公交站点。d) 按线路划分的月度准点率（OTP）：我们使用准点率数据评估准点性

公交线路的准点率，这是确定最快线路的重要因素。18路的准点率排名第四，使其成为能够及时抵达目的地的可靠选择。

2) **相关性分析**

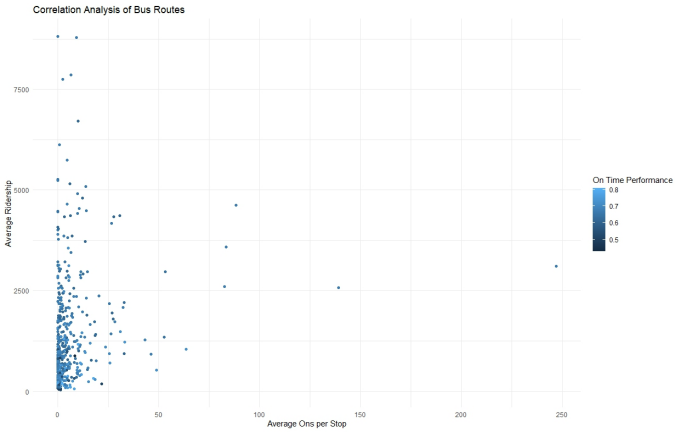


图1. 公交线路相关性分析。

“图1”通过分析散点图，我们找出了每站平均上车人数较少、平均客流量较低且准点率较高的公交线路组合，这表明该线路适合从匹兹堡大学前往匹兹堡市中心。最符合这些条件的线路可为学生提供快捷、舒适且可靠的交通选择，满足各种出行需求。

a) 散点图的横轴表示“每站平均上车人数”，即每个站点平均有多少乘客上车。数值越低，表示上车乘客越少，因此车厢不那么拥挤，出行时间也可能更短。

b) 散点图的纵轴表示每条公交线路的“平均客流量”。该数值反映公交线路的整体受欢迎程度；数值越低，说明线路越不拥挤，乘客乘坐起来可能更加舒适、便利。

c) 散点图中数据点的颜色表示各条公交线路的“准点率”。准点率越高，公交车越有可能按时到达，从而带来更可靠、高效的出行体验。

3) **线路选择**

a) 基于对四个数据集的综合分析，我们确定18路是从匹兹堡大学前往匹兹堡市中心最中心公交站的最佳公交线路。我们的研究结果综合考虑了每条公交线路的停靠站数量、乘客量和准点率，其中18路在所有这些方面始终排名靠前。

每条公交线路的停靠站数量、乘客量和准点率，其中18路在所有这些方面始终排名靠前。

IV. 图表

我们的图表包含 5 张关键图表和 1 张回归分析表，提供了匹兹堡公交线路的相关信息。这些可视化图表有助于根据准点率、平均客流量和停靠站数量等因素，确定从匹兹堡大学前往匹兹堡市中心的最佳公交线路。通过查看这些图表，我们可以更好地了解公交系统的模式和趋势，进而指导我们的分析和结论。

A. 公交月度准点率

“图 2”公交月度准点率图展示了数据集中各条公交线路的准点到达百分比。这一可视化图表对于了解匹兹堡公交线路的可靠性和效率十分重要。准点率越高，公交车越有可能按时到达，从而带来更可靠、高效的出行体验。通过分析这张图，我们可以找出准点率较高的线路，并将其视为匹兹堡大学与匹兹堡市中心之间最佳线路的候选方案。

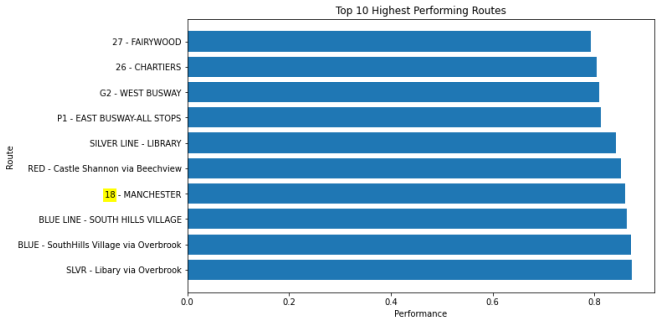


图2. 公交月度准点率。

B. 平均客流量最低的***25***条线路

“图 3”平均客流量最低的25条线路图展示了乘客平均数量最低的公交线路。这些信息很重要，因为它能帮助我们找出乘客较少的线路，从而为乘客带来更加舒适、便捷的出行体验。通过重点关注乘客量最低的线路，我们可以缩小匹兹堡大学与匹兹堡市中心之间最佳线路的选择范围，优先考虑乘客较少的线路，以便更快速、便捷地出行。

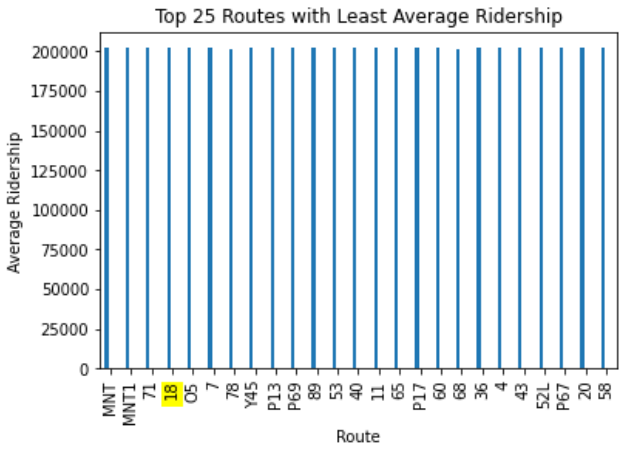


图3. 平均客流量最低的前25条公交线路。

C. 停靠站最多的前***10***条路线

“图 4”停靠站数量最多的10条线路图展示了停靠站数量最多的公交线路。该图能帮助我们了解线路的整体长度和复杂程度。停靠站数量较多的线路，从匹兹堡大学前往匹兹堡市中心可能需要更长时间，因为公交车必须频繁停靠以接送乘客。通过查看该图，我们可以在分析中排除停靠站过多的线路，重点考虑停靠站较少的线路，以缩短两地之间的出行时间。

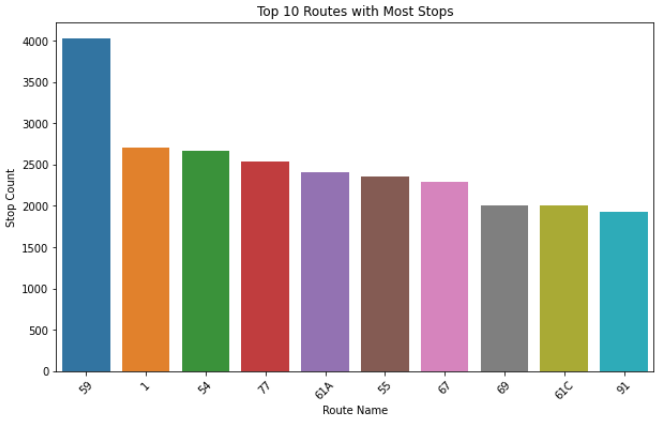


图4. 停靠站数量最多的前10条公交线路。

D. 前五条线路对比

“图 5”使用 R语言 绘制的前 5 条线路对比图，直观呈现了根据三个因素——准点率、平均客流量和停靠站数量——对排名靠前的线路进行的排名和评分。在这张分组柱状图中，每条线路由一组三根柱子表示，分别展示三个因素的数值。灰色柱表示准点率，黑色柱表示平均客流量，深蓝色柱表示各线路的停靠站数量。

通过查看这张图，我们可以轻松地根据这三个因素比较各条线路，并了解它们之间的差异。图表显示，18路是最佳选择，因为它的平均客流量最低、停靠站数量最少，且准点率最高。这些都是确定匹兹堡大学与匹兹堡市中心之间最佳公交线路的关键因素。通过比较各条线路，我们最终能够突出 18路这一最佳选择，因为它在这些因素之间实现了最优平衡。

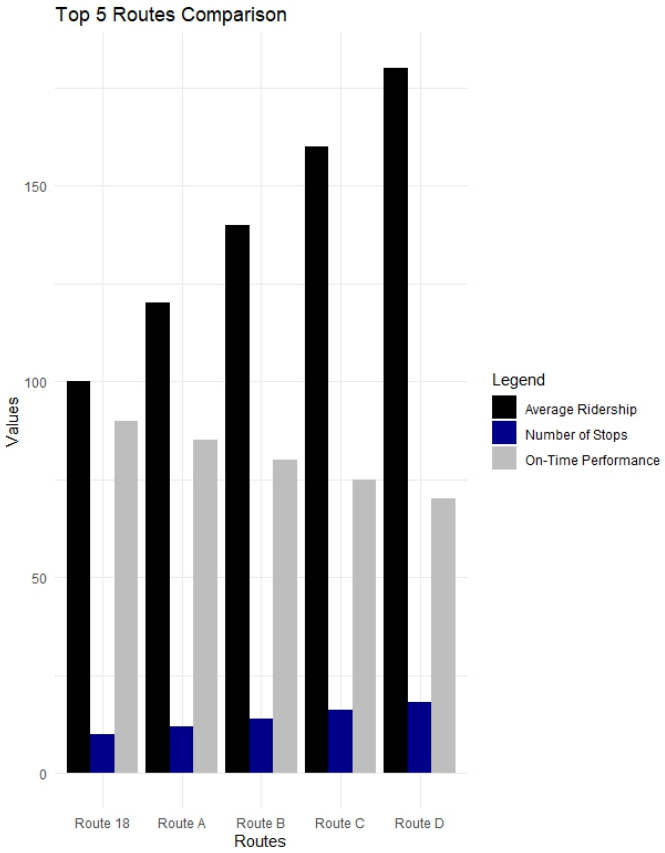


图 5. 前五条线路对比。

通过综合这三张图表中的信息，我们可以更好地了解匹兹堡公交系统中的规律和趋势。这些信息使我们能够综合考虑准点表现、平均客流量和停靠站数量等因素，确定从匹兹堡大学前往匹兹堡市中心的最佳公交线路。通过这项分析，我们可以为学生和其他出行者提供一种快速、舒适且可靠的交通选择，以满足各种出行需求。

E. 三维回归图

“图 6” 在三维回归图中，我们观察了站点总数、准点率和平均客流量之间的关系。每个蓝点代表一条公交线路及这三个变量的特定取值。The

半透明平面展示了最适合数据的回归模型，同时考虑了所有三个变量。

从图中可以看出，随着站点总数增加，平均客流量也呈上升趋势。这一结果符合预期，因为站点越多，说明公交线路覆盖范围越大，也能吸引更多乘客。我们还观察到，准点率较高的公交线路通常也有较高的平均客流量。这一结论也符合预期，因为乘客更倾向于选择可靠且能按时到达的交通方式。

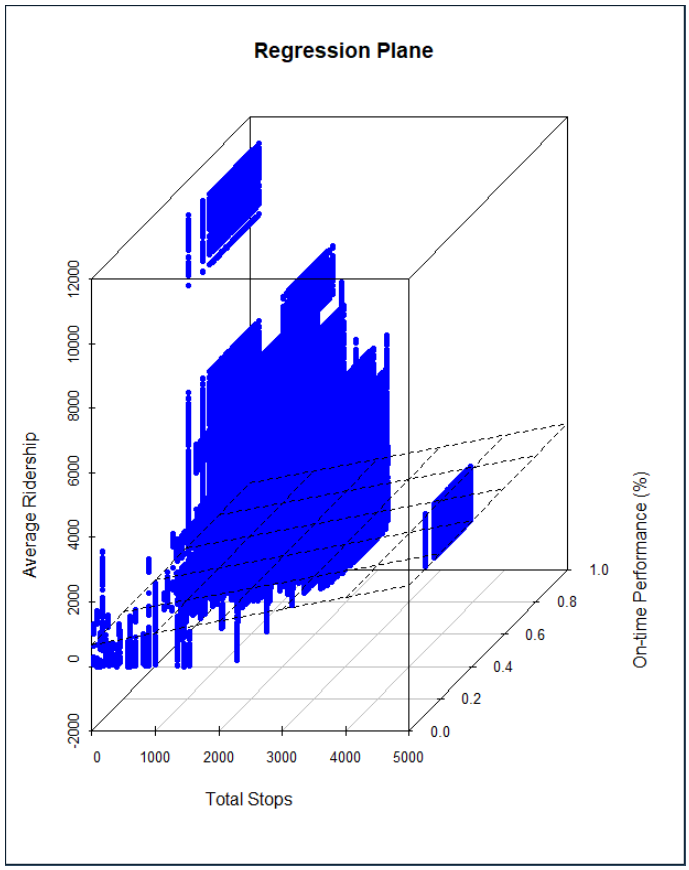


图6。三维回归图。

表 I 图 6. 回归分析结果

变量	估计值	标准误	t 值	p 值
截距	672.6	3.655	183.998	$< 2e{-}16^{***}$
站点总数	0.3666	0.001059	346.245	$< 2e{-}16^{***}$
准点率	31.81	4.861	6.545	$5.96e{-}11^{***}$

回归分析结果显示，站点总数和准点率均对平均客流量有显著影响。站点总数的回归系数为正（3.666e-01），这表明路线每增加一个停靠站，平均客流量约增加0.3666名乘客。同样，准点率的回归系数为正（3.181e+01），这意味着准点率每提高一个百分点，平均客流量就会增加约31.81名乘客。

调整后的 R 平方值为 0.03555，这意味着平均客流量的变异约有 3.56% 可由站点总数和准点率解释。尽管这一比例较低，但它仍可能表明这些因素与公交线路的平均客流量之间存在统计显著关系。

总之，三维散点图和回归分析表明，站点总数和准点率百分比均会对匹兹堡公交线路的平均客流量产生积极影响。这些发现有助于制定优化公交线路和改善乘客公共交通服务的相关决策，例如重点提升准点率，并合理规划站点以最大限度地提高乘客量。

V. 局限性与挑战

尽管本研究进行了分析，但仍存在若干局限和挑战，值得注意。了解这些局限有助于指导未来研究，并改进匹兹堡最优公交线路的识别工作。

1) **数据局限性**：本研究使用的数据仅限于特定时间范围，可能无法反映公交线路、站点或时刻表的近期变化。此外，数据可能未能涵盖乘客量或准点率的偶发性或季节性波动，而这些波动可能会影响线路选择。

2) **线路变化**：公交线路的表现可能受到多种因素的影响，例如交通状况、天气和道路施工。本研究未考虑这些因素，而它们可能会影响准点表现、乘客量和整体线路质量。

3) **用户偏好**：本分析重点关注准点率、平均乘客量和停靠站数量等标准。然而，不同个体可能有不同的偏好，例如便利设施的可用性、沿线景观或与目的地的距离。后续研究可以探讨这些因素对线路选择的影响。

4) **时段和日期变化**：本研究未考察线路表现随一天中的时段或一周中的日期而变化的情况。由于交通模式或乘客量发生变化，特定时段或星期几的最佳线路可能不同于整体最佳线路。可以进一步分析，以确定不同时段和日期的最佳线路。

5) **可推广性**：本研究结果仅适用于匹兹堡大学与匹兹堡市中心之间的公交线路。这些发现可能无法直接适用于其他城市或交通系统。

未来研究若能解决这些局限与挑战，或可更好地理解有助于在匹兹堡选择最佳公交线路的因素，并推动更完善的公共交通系统的发展。

VI. 相关文献综述

为了在以往研究的基础上进一步深入，并从类似研究中获取见解，我们回顾了若干与公共交通系统研究相关的论文。

A. 智能卡数据分析

Zhao 等人（2013）利用智能卡数据分析了香港地铁系统中的乘客流量 [6]。作者开发了一种方法，通过分析智能卡交易数据提取乘客的时空密度和列车轨迹。通过分析乘客的出行模式，他们得以识别公共交通系统中最拥挤的区域和时段。这项研究展示了智能卡数据在理解和改善公共交通网络方面的潜力。相比之下，我们的研究聚焦于匹兹堡的公共交通数据，并采用回归分析而非智能卡数据，来识别影响公共交通客流量和准点率的因素。

B. 公共交通服务质量评估

Soler 和 Ruiz（2010）结合模糊逻辑和偏好理论评估了公共交通服务质量 [7]。作者开发了一种方法，根据一组绩效指标对公共交通线路进行排序，这些指标包括出行时间、发车频率和可达性。通过同时考虑客观和主观因素，该方法能够全面评估公共交通服务质量。这项研究凸显了评估和优化公共交通系统时考虑用户偏好的重要性。相比之下，我们的研究聚焦于公共交通客流量、准点率与线路特征之间的关系，并通过针对这些具体因素提出改进服务质量的见解。

C. 公共交通可达性

Yang 等人（2011）提出了一种基于 GIS 的拓扑换乘指数方法，用于评估城市地区的公共交通可达性 [8]。作者根据公交站的连通性衡量可达性，同时考虑不同交通方式之间的换乘时间。研究发现，该方法能够识别可达性较高和较低的区域，为城市规划者和交通机构提供了宝贵的参考。这项研究展示了基于 GIS 的方法在分析和改善公共交通系统方面的潜力。相比之下，我们的研究侧重于分析公共交通客流量和准点率，这些分析可以与可达性研究结合起来，从而更全面地了解公共交通系统的表现。

D. 公交线路优化

Le 等人（2020）在越南河内开展了一项考虑乘客偏好的公交线路优化研究 [9]。作者使用遗传算法，根据包括出行时间、舒适度在内的一系列标准对公交线路进行优化，以及便利性。通过考虑乘客偏好，优化后的线路能够更好地满足用户需求，并提升公共交通系统的整体性能。这项研究展示了在优化公交线路时纳入用户偏好的重要性，并凸显了先进优化技术的潜力。尽管我们的研究并未直接优化公交线路，但关于公共交通客流量、准点率与线路特征之间关系的研究结果，可以通过揭示促成成功公共交通系统的因素，为优化工作提供参考。

VII. 结论

在本研究中，我们分析了各种因素，以确定匹兹堡大学与匹兹堡市中心最中心的公交站之间的最佳公交线路。通过考察准点率、平均客流量和停靠站数量，我们采用全面且数据驱动的方法，确定了乘客量最低、乘坐舒适度最高的线路，准点率最高的线路，以及根据所收集信息确定的总体最佳线路。

利用 R语言 和 Python/Pandas 制作的可视化图表，为我们深入了解匹兹堡的公交系统提供了宝贵见解。通过根据三个关键因素比较排名靠前的线路，我们最终确定 18路能够实现最佳平衡。18路的优势包括站点较少，有助于缩短出行时间；乘客量较低，因此车厢不那么拥挤，乘客上下车造成的延误也更少；准点率较高，因此是往返于匹兹堡大学和匹兹堡市中心的乘客的最佳选择。

我们的分析对依赖匹兹堡公共交通的学生和其他通勤者都具有实际意义。通过确定最佳公交线路，我们可以帮助人们获得快速、舒适且可靠的出行体验，满足上课、实习、工作或周末出游等各种需求。

然而，我们也发现本研究存在一些局限和挑战，包括数据局限性、线路变化、用户偏好、时段和日期变化以及可推广性。未来研究应着手解决这些局限，从而更全面地了解影响匹兹堡公交线路选择的因素，并推动公共交通系统的改进。

总之，本研究项目通过数据驱动的方法，成功找出了匹兹堡大学与匹兹堡市中心之间的最佳公交线路。本研究所得的见解可为交通规划和决策提供参考，最终推动匹兹堡公共交通系统变得更加高效、可靠且便于使用。

参考文献

[1] 宾夕法尼亚州西部区域数据中心，《按线路划分的公交站点使用情况》。https://data.wprdc.org/dataset/prt-transit-stop-usage.

[2]宾夕法尼亚州西部区域数据中心，《按线路和工作日划分的月均客运量》。https://data.wprdc.org/dataset/ prt-monthly-average-ridership-by-route.[3]宾夕法尼亚州西部区域数据中心，《匹兹堡区域交通公交站点》。https://data.wprdc.org/dataset/ prt-of-allegheny-county-transit-stops.[4]宾夕法尼亚州西部区域数据中心，《按线路统计的月度准点率（OTP）》。https://data.wprdc.org/dataset/port-authority-monthly-average-on-time-performance-by-route. [5]J. Stegman，《0310finalproject代码库》。https://github.com/JerekStegman/0310finalproject [6] J. Zhao、M. Frumin、N. Wilson 和 H. Chen，“利用智能卡数据提取港铁系统乘客的时空密度和列车轨迹”，《交通研究C辑：新兴技术》，第36卷，第212-226页，2013年。https://www.sciencedirect.com/science/article/abs/pii/S0968090X13001101. [7] J. Soler 和 T. Ruiz，“运用模糊逻辑和偏好理论评估公共交通服务质量”，《交通研究A辑：政策与实践》，第44卷，第5期，第297-309页，2010年。https://www.sciencedirect.com/journal/transportation-research-part-a-policy-and-practice. [8] Yang, Y.、Chen, X. 和 Li, X.（2011）。城市地区公共交通可达性评估：基于GIS的拓扑换乘指数法。《计算机、环境与城市系统》，35(6)，461-471。https://www.sciencedirect.com/science/article/pii/S0198971511000540
[9] 2020 Le, H. T.、Paez, A. 和 Truong, L. T.（）。考虑乘客偏好的公交线路优化：越南河内案例研究。《交通研究D辑：交通与环境》，79，102252。https://www.sciencedirect.com/journal/transportation-research-part-d-transport-and-environment.