

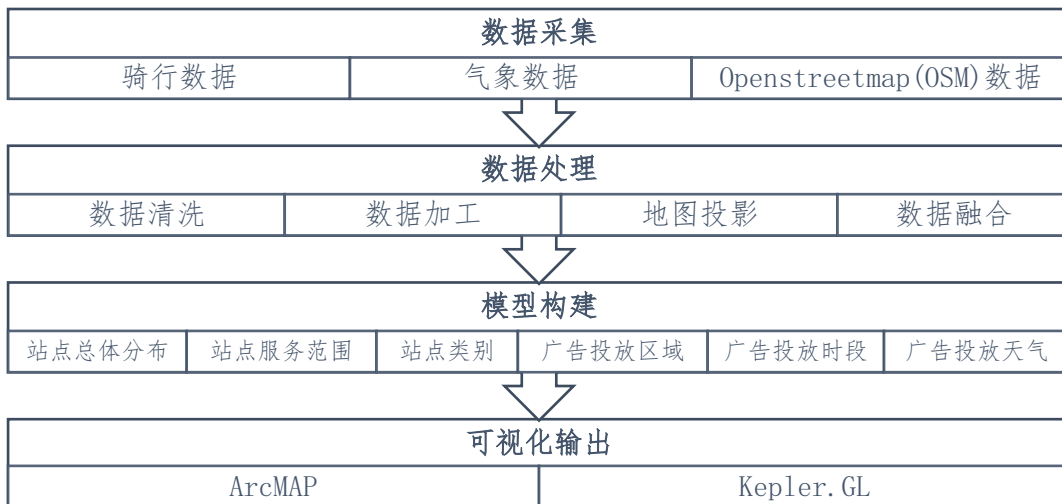
基于 FME 的共享单车租赁站点布局、广告精准投放分析

1 分析背景及内容

提到共享单车，不难想到它主要是用来接驳地铁与公交站点，然而共享单车对城市规划的影响远不止如此，它还能城市交通政策、道路规划、公交线网规划、非机动车停车设施规划和绿地系统规划提供新的参考思路。此外，共享单车本身就是一个很好的移动广告载体，如何结合空间位置与用户信息，给予用户个性化广告推荐，进行商业化变现，是一个值得深入探讨的问题。

为此，项目采集了 CitiBike(是纽约私有营利公共自行车系统，是美国最大的公共自行车系统，由花旗银行赞助并命名为花旗单车) 2019 年骑行记录数据及区域内 OpenStreetMap(简称 OSM) 数据，通过 FME 软件进行数据清洗、加工、融合与建模，完成了一些有趣的相关分析，如用户规模、站点分布、骑行规律等，以期为城市规划、广告投放提供新的思路和视角，同时展现 FME 软件在数据治理、洞察、挖掘等领域的实用价值。

2 总体思路



3 数据采集与处理

骑行数据 (CSV 格式)、OSM 数据最简单的采集方式是直接下载，当然通过 FME 工具也可批量完成，但访问该国外网站其网络很不稳定，极易中断。气象数据可通过 HttpCaller 工具进行爬取。需注意的是，在爬取数据过程中，为防止高频率向服务器发出请求遭拒，比较实用的做法是配合使用

RandomNumberGenerator、Decelerator 工具，以便让每次请求产生随机时间延迟。

预处理工作主要是通过StringSearch工具中的正则表达式检索出建模过程中用到的属性信息，通过AttributeManager工具完成属性类型转换（如字符型转换为日期型、数字型等）、规范化处理，通过Tester工具完成异常值剔除（如骑行时长仅有1~2分钟），最后调用FeatureMerger工具完成骑行数据、气象数据融合，调用EsriReprojector工具完成OSM中的行政区划、城市公园、POI等多源数据地图投影转换等。采集及处理的程序如图1所示。

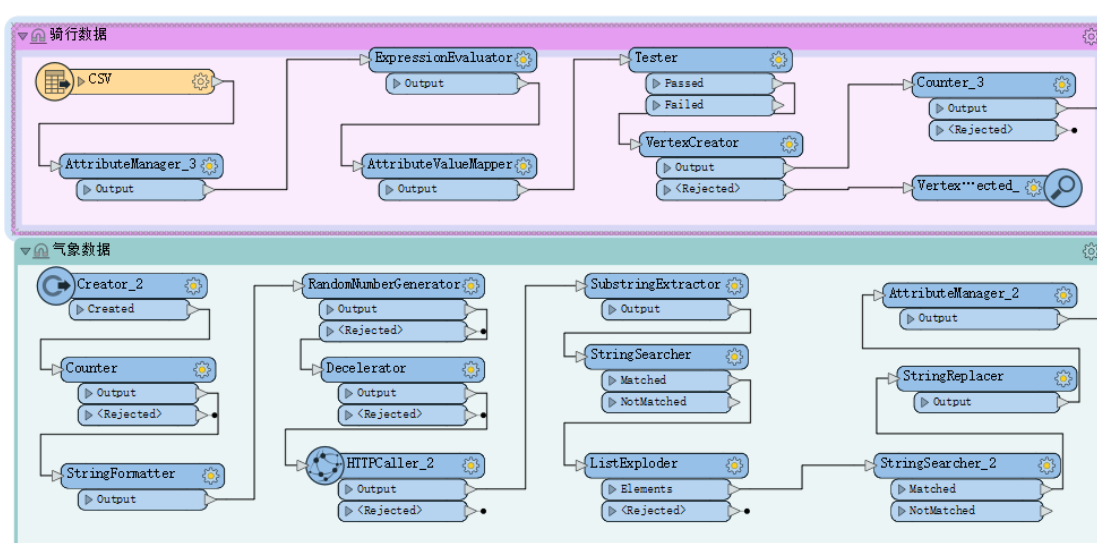


图1 数据采集及处理成果（部分）

4 租赁站点布局分析

与出租车不同的是花旗单车不能在城市中的任意地点上下车，骑手只能在城市中数量有限的租赁点上下车。对于公共自行车系统而言，租赁点就是其最基本的服务单元，所以，租赁点布局是否合理将直接影响公共自行车系统运营的流畅性，进而影响城市交通网络规划，为此，我们接下来我们通过FME软件进行租赁站点相关分析。

4.1 整体分布情况

分析数据可发现每条骑行记录都有标识用户起、止站点的经纬度信息，对于站点分析来说，无疑存在大量重复，可通过DuplicateFilter工具进行去重操作，再将行政区划与站点数据进行叠加，即可统计出各区站点分布数量。此外，

每个站点在不同时间段出行人数不一，为反映各站点冷热程度，可以以天为统计单元，按站点进行分组聚合，求出各站点每天平均总出行量。为增强可视化表达效果，笔者将各站点的出行量用高度表示，如图 2 所示。

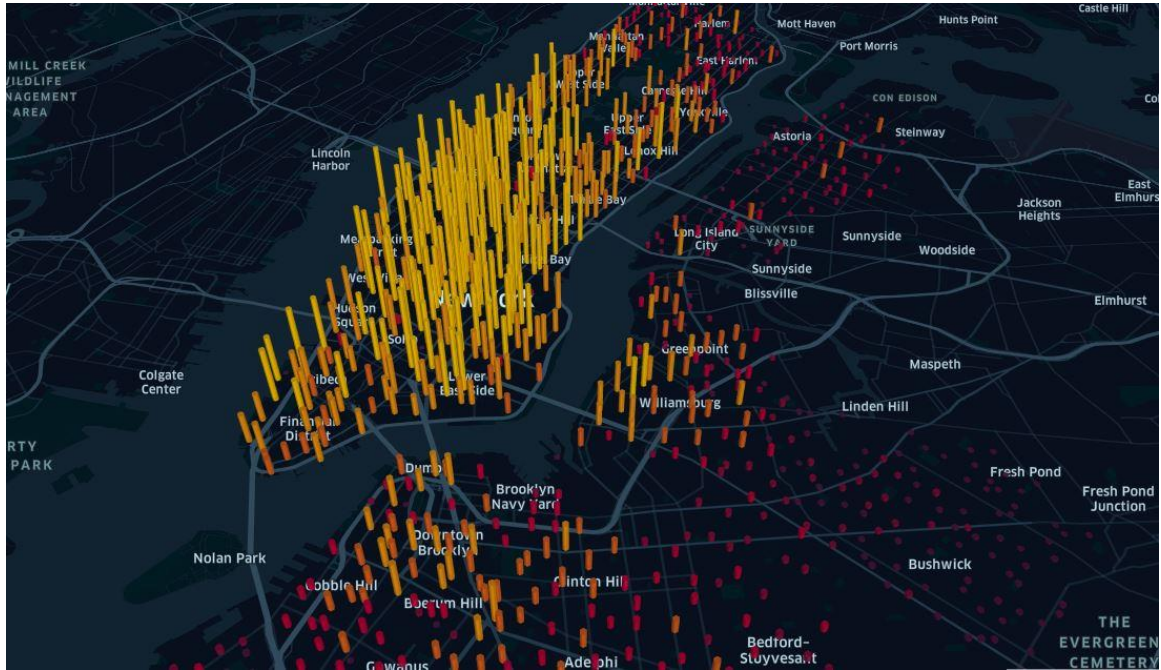


图 2 租赁站点与出行量分布图

分析结论：自行车停靠站点数共计 934 个，曼哈顿 471 个，布鲁克林 358 个，皇后区 104 个，热门出行站点主要分布在曼哈顿中央片区。此外，也得出 2019 年骑行总记录为 20520187 次，共享自行车共计 13816 辆，整年每辆自行车租借频率 1485.25，每天每辆自行车租借频率 4.07。

4.2 站点服务范围分析

据统计，通常情况使用公共自行车出行的距离都在 300m 以上，300m 以内可通过步行完成，很少有人会在如此短的出行距离选择公共自行车。倘若用户选择骑行，骑行的路线长度也是各不相同，我们不妨定义大多数用户骑行的平均距离作站点的服务范围。

(1) 平均骑行时长计算。以站点为分析对象，求解所有骑手从当前站点出发至行程结束所用的时间。原始数据中骑行时长都是精确到秒级别，分析过程中可通过 FME 工具将单位改正至分钟，并以 1 分钟时间间隔为统计单元，进行出行量统计。在分析过程中，尤其是面对区间统计时，FME 软件的便捷性充分显现，通过 Creator 工具实例化需要定义的区间个数，Counter 工具进行区间编号，

FeatureMerge 工具进行区间与统计数据挂接即可实现。依据统计结果，可形成图 3 骑行时长热力图与统计图。



图 3 骑行时长分布

分析结论：骑行时间较长区域分布在曼哈顿中央片区及大桥的两侧。形态满足正态分布，大部分用户的使用时间为 2~9 分钟，可选取这部分用户的骑行路线长度平均值作为租赁站点服务范围。

(2) 租赁点服务范围计算。原始数据中并不存在用户骑行的路线及路程信息，仅有起始点、终止点位置信息，笔者采用了 HttpCaller 工具调用 OpenRouteService 网站的 Direction API 接口(原本打算采用 Google 接口服务，但无法访问)，出行模式选择骑行，可生成 GeoJson 格式的规划路线。值得说明的是，笔者也对比了用户实际骑行时长与 API 返回的规划时长，发现两者基本一致，这说明这种做法是行之有效地。此外，该网站 API 接口的调用和百度接口调用类似，用户只需申请 Key 后方即可使用。将 GeoJson 数据可视化后可形成图 4 效果，不难发现，它和实际路网很是贴合。

分析结论：通过计算路线平均距离可得站点服务范围为 1738 米。



图 4 骑行路线（部分）

4.3 站点类别分析

结合实际情况可知有的租赁点是为了解决车站与目的地之间“最后一公里”的问题，有的是为了方便附近的居民上下班，日常生活出行等。根据租赁站点所发挥的作用不同，项目尝试租赁站点类别定义与分类。

(1) 类别定义。将地铁、公交站附近的租赁点定义为换乘点；旅游景点、城市公园附近的租赁点定义为休闲（娱乐）点；大型商场、超市、餐饮娱乐附近租赁点定义为商业点；一些大型社区和居住的小区附近租赁点定义为居住点等。

(2) 装载 OSM 中所涉及到的上述数据，采用空间叠加、缓冲区、数组去重排序、属性连接等工具综合分析各租赁点所属类别。为结合周边基础设施分布情况，分析过程中我们也引入泰森多边形，通过租赁站点生成泰森多边形，并挂接综合类别属性信息，可形成图 5 可视化效果。

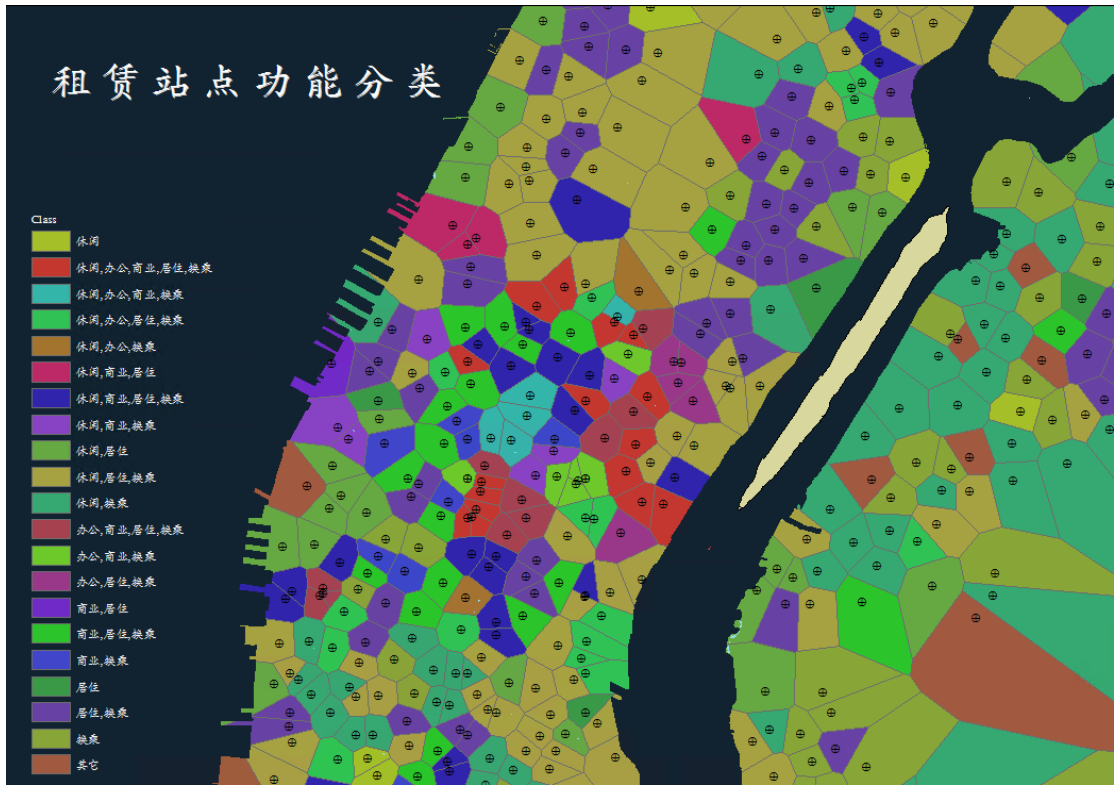


图5 租赁站点功能分类

分析结论：曼哈顿中央片区的大多数租赁站点主要是为完成地铁、公交接驳用途。整个分析过程用到的FME程序如图6所示：

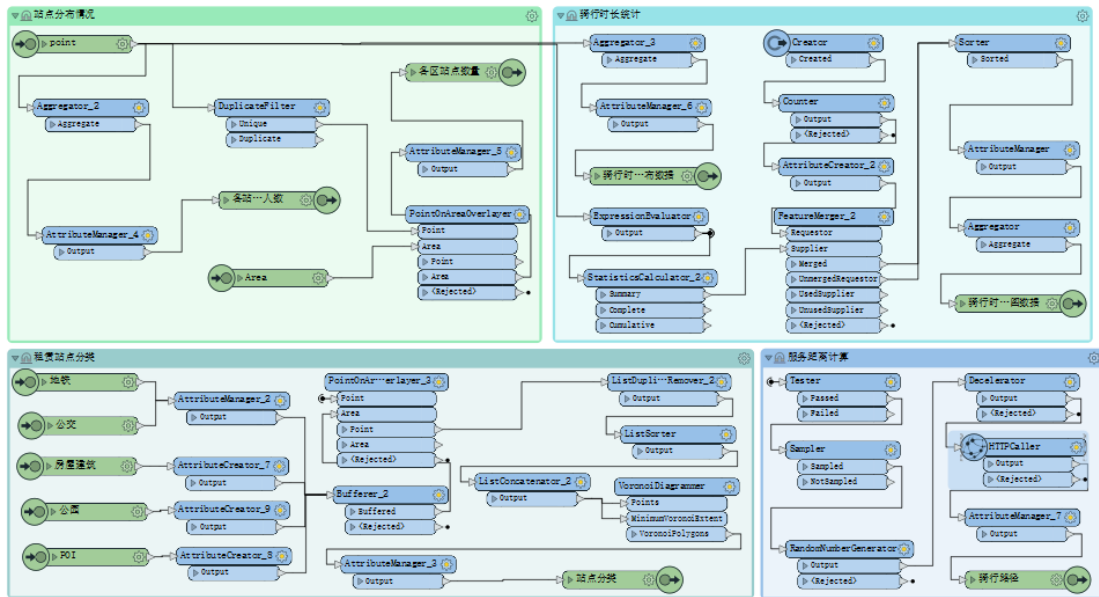


图6 租赁站点布局分析程序

5 广告精准投放分析

考虑到线上的流量成本越来越贵，很多共享单车选择在线下做推广，把车

尽可能多地投放到中央商务区上去，充当免费活体广告。那么通过 FME 软件我们如何为广告投放公司提供参考性意见。这里我们假设一个场景，某一新兴化妆品品牌想在纽约街区投放广告，哪些区域？哪些时间段？进行广告投放所发挥的功效最大。

5.1 空间维度分析

首先，定义目标人群。我们不妨将潜在客户锁定为青年、中年女性骑手。骑行数据中有骑手性别及出生日期字段，通过简单的分析可得出一些有趣的结论：2019 年总出行中，79%为男性骑手，15%为女性骑手，另外还有 6%的骑手性别未知。男性骑手比女性骑手多的原因可能归因于男性和女性的交通偏好。

接下来，进行年龄结构分析。将骑手年龄分为四组，0~18 岁为少年组，18~30 岁为青年组，30~50 岁为中年组，50 岁以上为老年组，通过 FME 工具过滤、出青年、中年女性骑手，将其定义为潜在用户数据集，进行分类统计。依据分析结果可知，租赁站点全天候最多的有 1223 个潜在用户出行，最少的仅有 1 个（近 50 个租赁站点）用户出行。为进一步表达其分布方向，我们引入了标准差椭圆。不妨取每日潜在用户数大于 700 人的站点，可形成图 7 所示的分布图。



图 7 潜在用户分布图

分析结论：潜在用户热点区域分布在 W 21 St & 6 Ave、E 13 St & Avenue A、E 17 St & Broadway，日均潜在用户量为 1223 人、1107 人、1065 人。

5.2 时间维度分析

通过潜在用户分析可以基本锁定广告投放的空间位置，接下来我们还需进一步分析最佳投放星期及时间段。

依据基姆拉尔森计算公式，将具体日期修正为星期，然后，按星期进行分类，以小时为记录单元进行 24 小时骑行数据聚合，求解出各站点各时段人流量。为直观表现不同星期不同时间段人流差异，笔者采用了三维柱状图进行可视化展示，如图 8 所示。

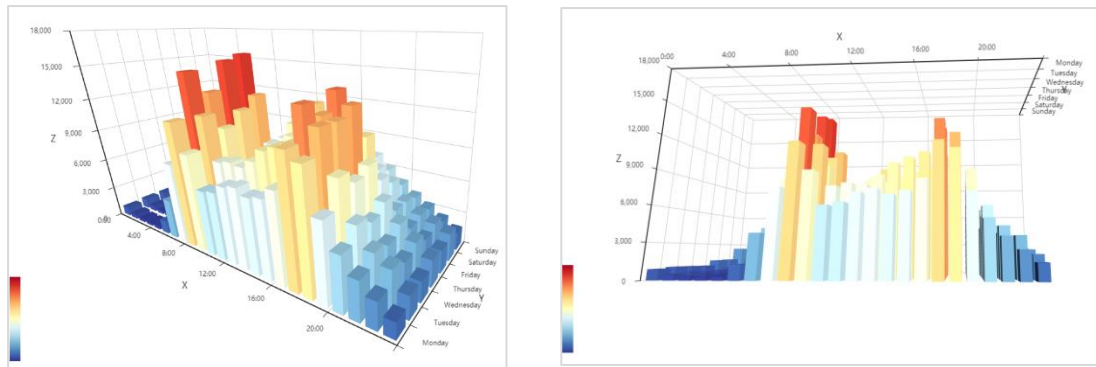


图 8 出行时间段统计图

分析结论：最佳星期为星期五（星期四次之，星期二再次之）；最佳时段为上午 8 点（17 点次之）。此外，不难发现图形中的两个波峰正是上下班高峰期。

5.3 天气维度分析

共享单车作为一种暴露在交通环境中的出行方式，环境中出现的所有因素都可能会对用户出行产生一定的影响，而让用户最能切身感受到的正是天气因素。暴风骤雨、寒冬腊月肯定少有用户出行，碧空如洗、晴空万里肯定会有诸多用户出行。那么建立两者之间关系模型势必对广告商预测用户出行量具有一定的价值。

确定自变量与因变量，自变量取当天最高、最低温度，因变量取每日用户出行量。按照时间序列，进行出行数据、温度数据聚合、汇总处理，可形成图 9 可视化成果。

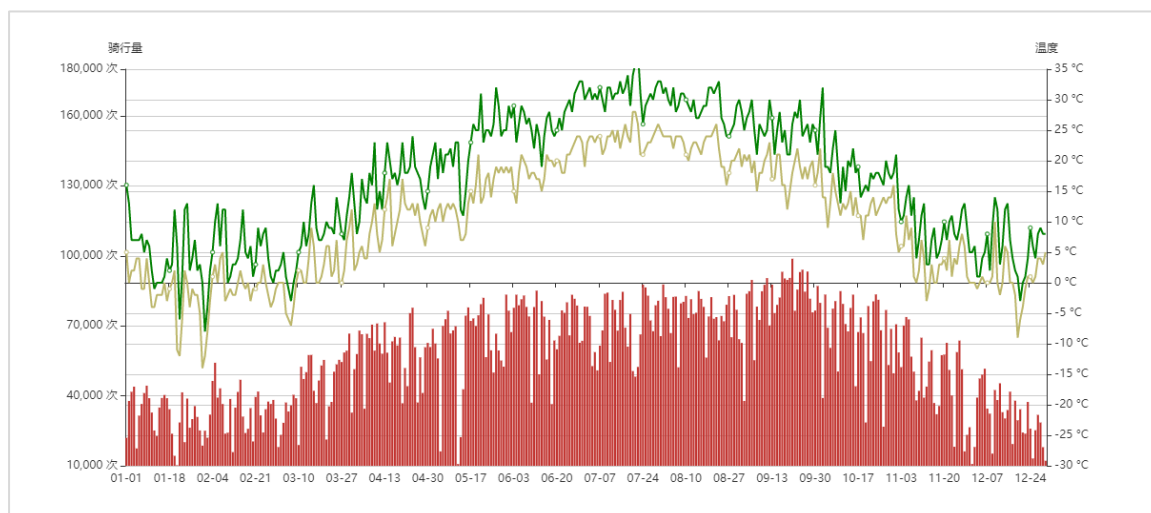


图 9 气温变化与骑行量分布

分析结论：最适宜投放广告的天气温度为 15°C~25°C。当然，以上分析仅从温度角度进行了分析，除次之外道路通行能力、非机动车道环境、道路级别等因素可能对共享单车出行情况也会存在不同程度的影响，值得进一步深入研究。

5.4 流向维度分析

线下广告投放一种方式是将其放置在自行车设备上，一种方式是扫街式推广。若放置在自行车设备上，广告商可选择潜在用户出没地及自行车上即可。若采用扫街式推广，我们还需进一步分析潜在用户流向的区域。

因每条骑行记录数据中有起、止位置信息，不难想到可进行 OD 分析，以便挖掘用户流向的区域。起始站点我们不妨选取潜在用户居多的站点位置，终止点选择潜在用户停靠的站点，通过 FME 工具求解流往至各地人流量，最终以不同粗细线条进行可视化展示，如图 10 所示。

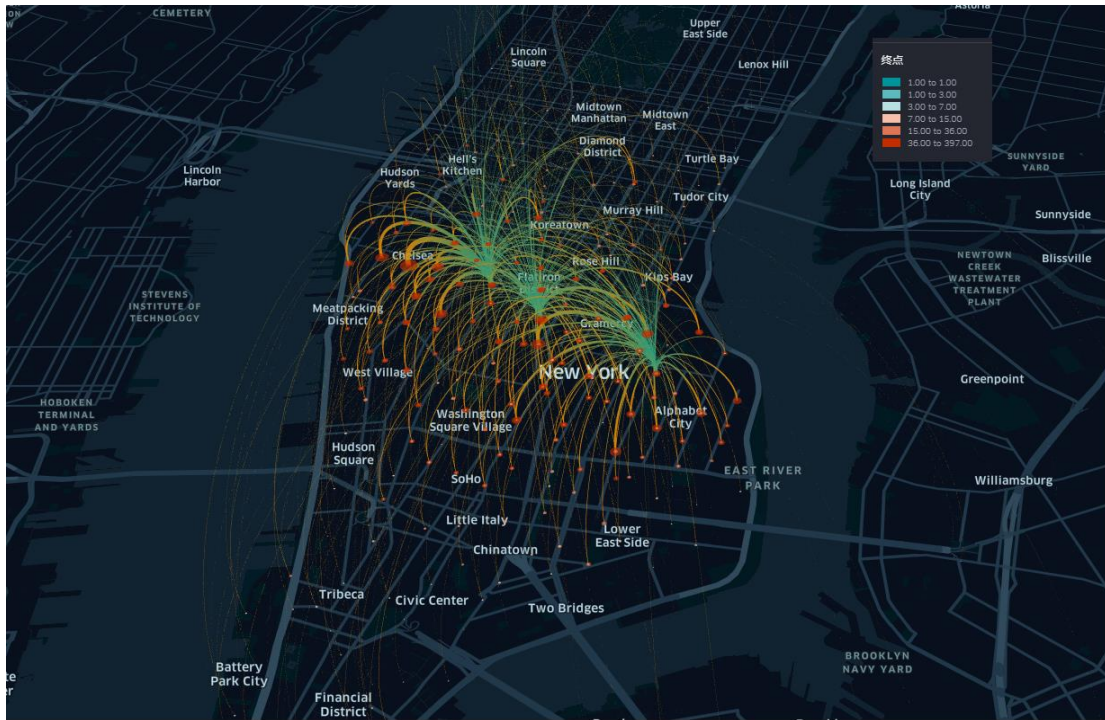


图 10 OD 分析图

分析结论：流向区域比较分散，密度较大区域主要分布在曼哈顿核心区域，广告运营商可在该区域进行流动式广告营销。整个分析过程用到的 FME 程序如下图所示 11 所示。

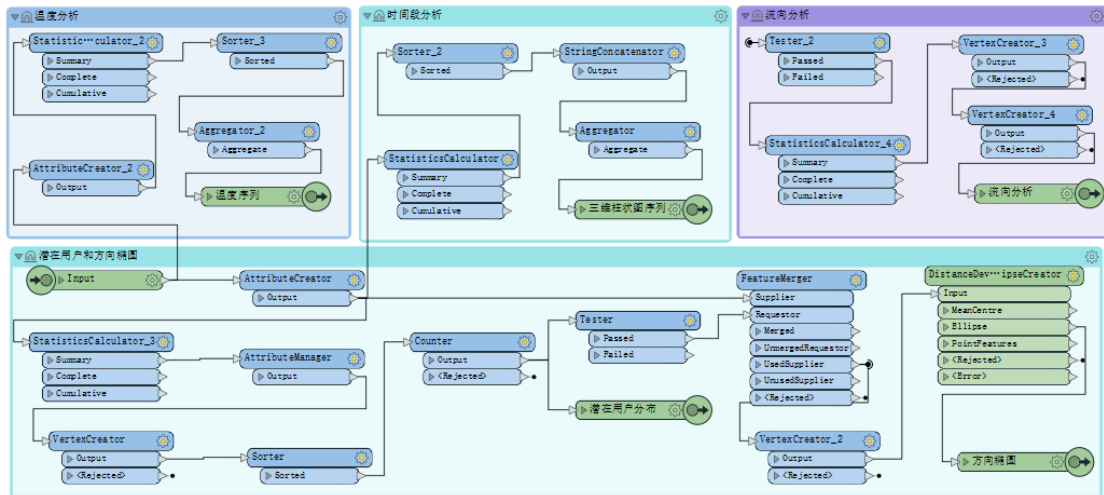


图 11 广告精准投放程序

5 总结

近几年，共享单车已成为了一种潮流，遍地的共享单车铺面而至，为解决最后一公里出行做出了突出的贡献。笔者原本打算采集国内单车骑行数据进行

若干分析，但因水平有限，所采集的数据时间跨度窄且难度大，索性选择了国外该开放的数据源。虽然笔者仅进行了以上两个方向分析，但骑行数据的价值远不止如此，还可以进一步深入挖掘，比如自行车投放量评估、租赁站点增减对交通流量的影响等。整个分析、建模过程中，让我深深的体会到了 FME 功能之强大，不仅有非常丰富、灵活的 ETL 组件，而且能够无缝集成多源数据模型及第三方处理工具。最后，非常感谢北京世纪安图公司人员及志同道合的 FME 朋友们，漫长的数据旅途有你相伴，纵然“人海”茫茫我也不彷徨！