

西安市美团平台的数据挖掘

1. 对西安市所有美团美食的店铺信息进行数据挖掘

1.1 聚类分析

首先通过爬虫获取美团美食平台上的所有店铺的名字、得分、平均价格、地址、以及评论数量。数据来源: (<https://xa.meituan.com/meishi/pn0/>), 爬取工具为 selenium, webdriver, BeautifulSoup。通过模仿人机登录, 进入界面, 破解美团加密的 CSS 字体, 使用绘图工具描绘出字体的大致形状, 利用 ORC 技术来识别对应的字体。爬虫只是获取数据工具, 不予以详述。

选取得分、平均价格、评论数量作为样本的特征进行聚类分析, 将所有店家分为 3 类 (好、中、差)。此处使用 KMeans 算法。样本集为 $D = \{x_1, x_2, \dots, x_{976}\}$, 聚

类簇的划分为 $C = \{C_1, C_2, C_3\}$, 最终目标函数为 $\min \sum_{i=1}^3 \sum_{x \in C_i} (x - \mu_i)^2$, 求取此目标

函数的最小值被证明是 NP 难问题, 故采用贪心算法, 通过每一轮的迭代, 不断优化目标函数。

算法具体过程为:

第一步: 随机选择 K 个样本作为初始均值向量, 设为 $\{\mu_1, \mu_2, \mu_3\}$, 并将这 K 个样本构成 K 簇。

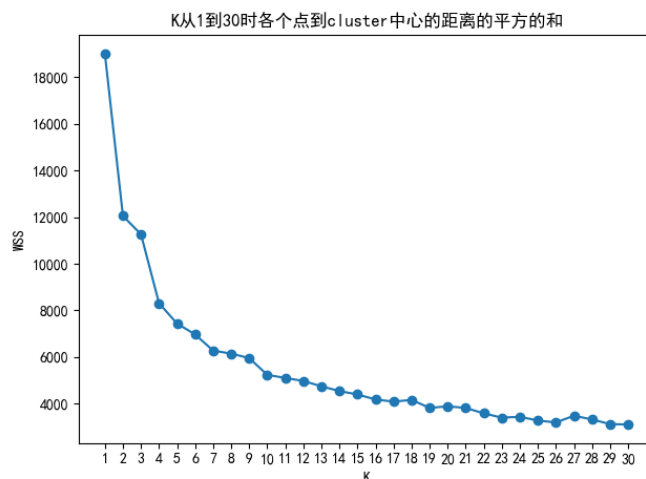
第二步: 随机选择一个簇, 将其设为空集。

第三步: 计算每个样本点到均值向量的距离。选择距离最近的均值向量所在簇作为此样本点所对应的簇。

第四步: 更新每个簇的均值向量 $\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$, 若均值向量发生改变, 则

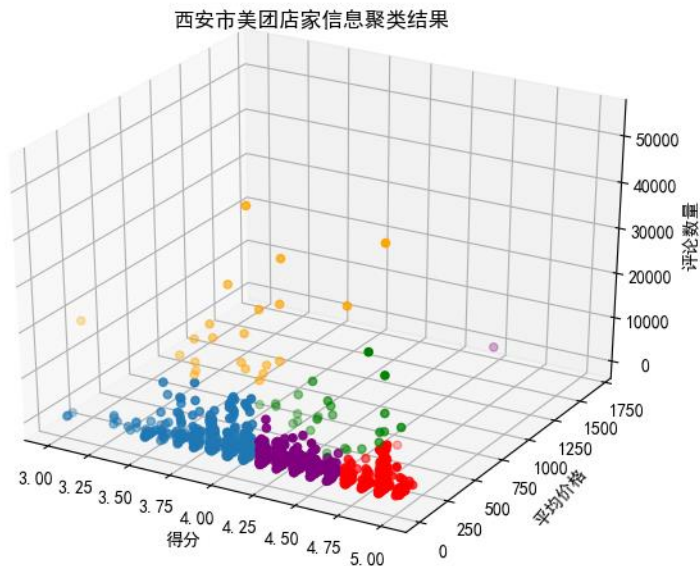
跳转第二步。若均值向量不发生改变, 则结束循环, 返回簇划分情况。

通过利用经验法则手肘法, 设 $K=1, 2, \dots, 30$, 分别得出对应 K 值情况下的 WSS 分布图。不难发现当 $K=5$ 时, WSS 的取值变化微小, 认为 $K=5$ 是 K 的最佳取值。



将 K 设为 5，进行聚类分析，得到了下面的聚类结果图，可以看出店家被分为五类，这五类的普遍特征分别为

- 1. 得分较高和平均价格较低。
- 2. 得分中等和平均价格较低。
- 3. 得分较低和平均价格较低
- 4. 得分较高、平均价格较高和评论数量较高。
- 5. 得分较低、平均价格较高和评论数量较高。

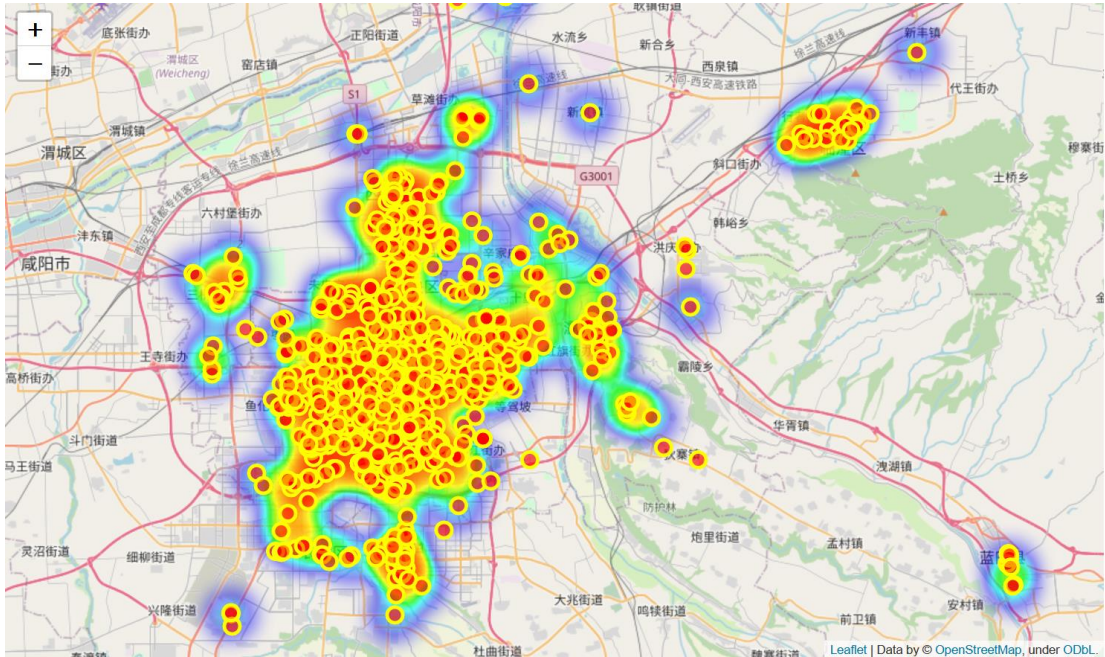


不妨将 5 类特征进行划分为五个等级如下表所示，分类方法比较符合常理，可以认为聚类效果较好。

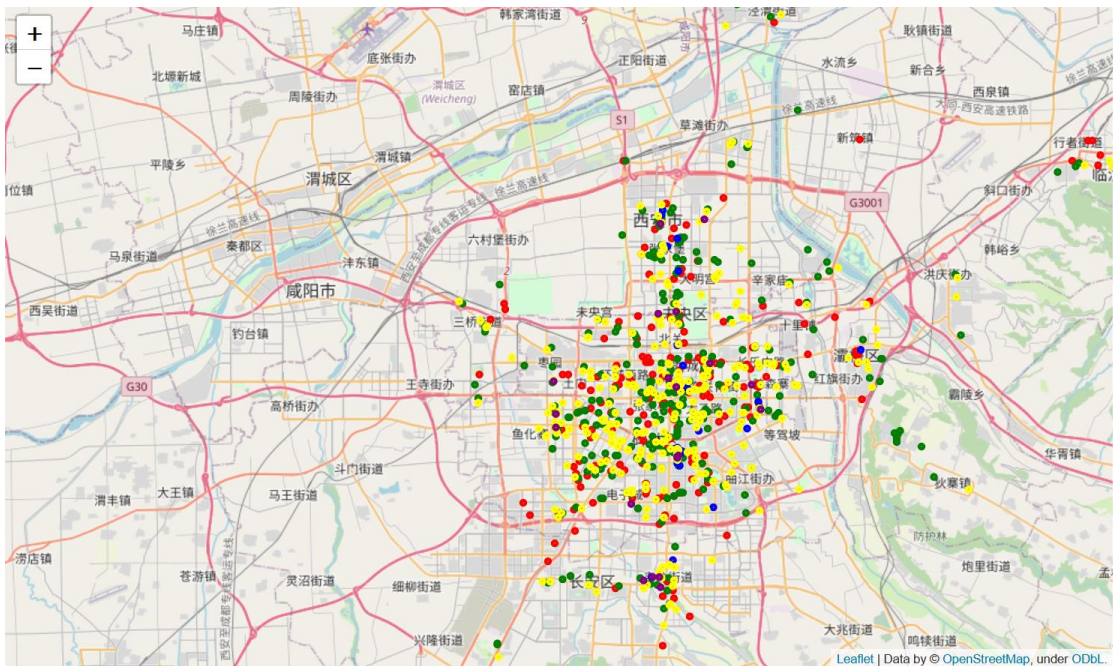
得分较高和平均价格较低。	A 类
得分中等和平均价格较低。	B 类
得分较高、平均价格较高和评论数量较高。	C 类
得分较低和平均价格较低。	D 类
得分较低、平均价格较高和评论数量较高。	E 类

2.2 基于地理位置对聚类结果进行分析

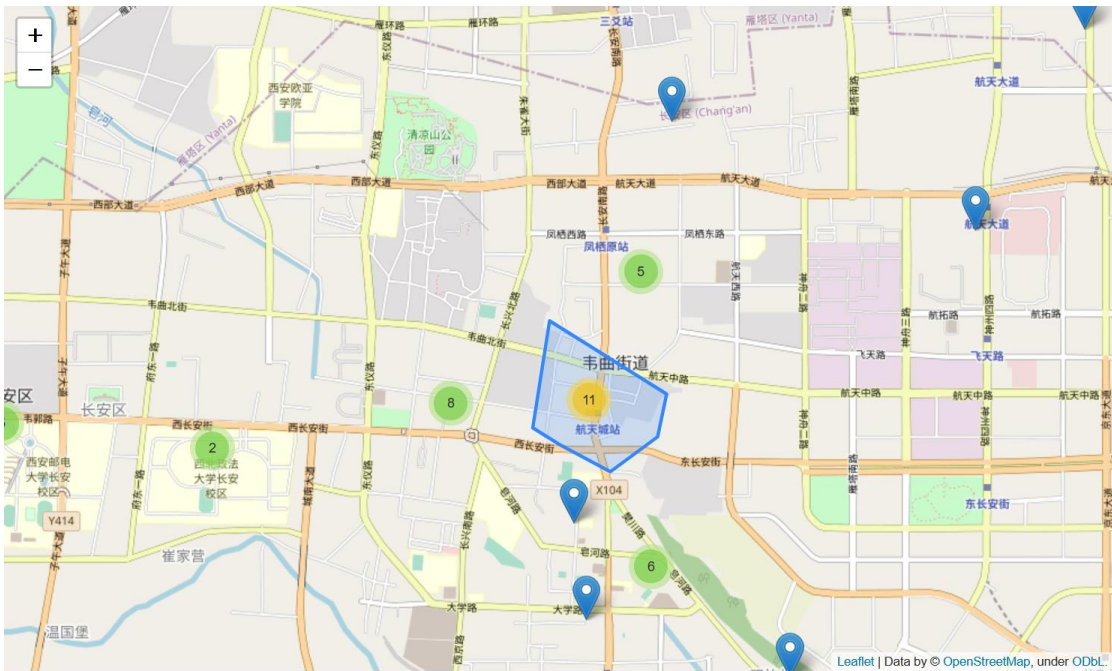
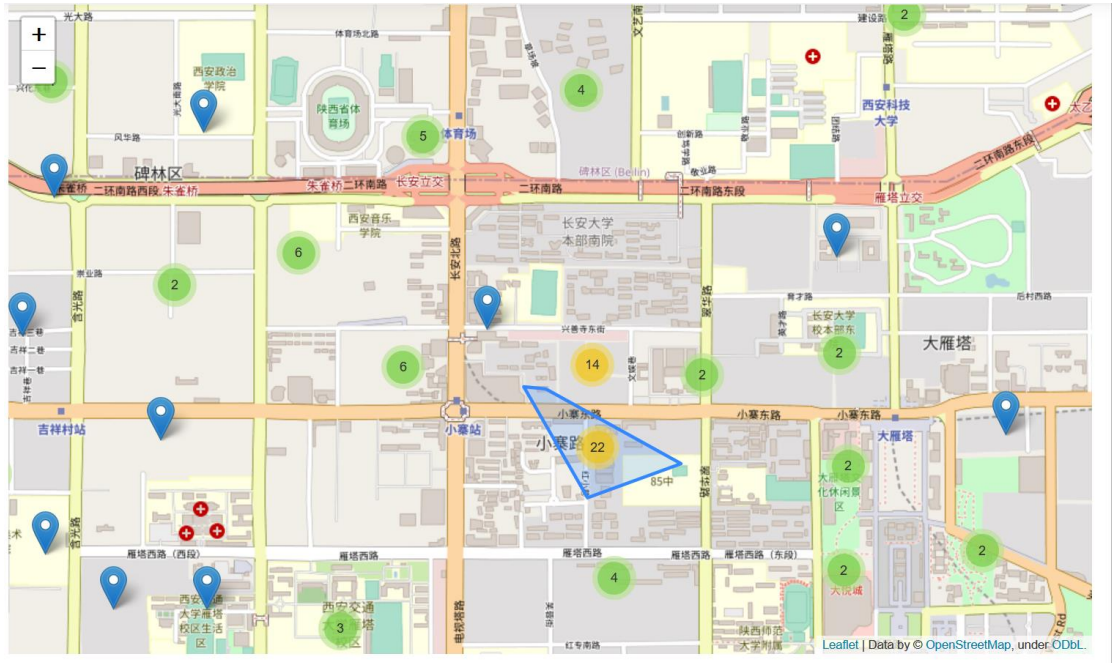
将获取到的店铺的位置信息，通过调用高德地图 API 进行逆地理编码，得到每个店铺所对应的经纬度。调用 folium 包，对西安的美团美食店铺的位置进行可视化处理，其中每个红色的圆圈代表在美团美食注册过的的餐饮店铺。（由于逆地理编码以及调用 folium 包的过程中难免出现具体精度不准确的问题，但不会影响店铺的大致位置）

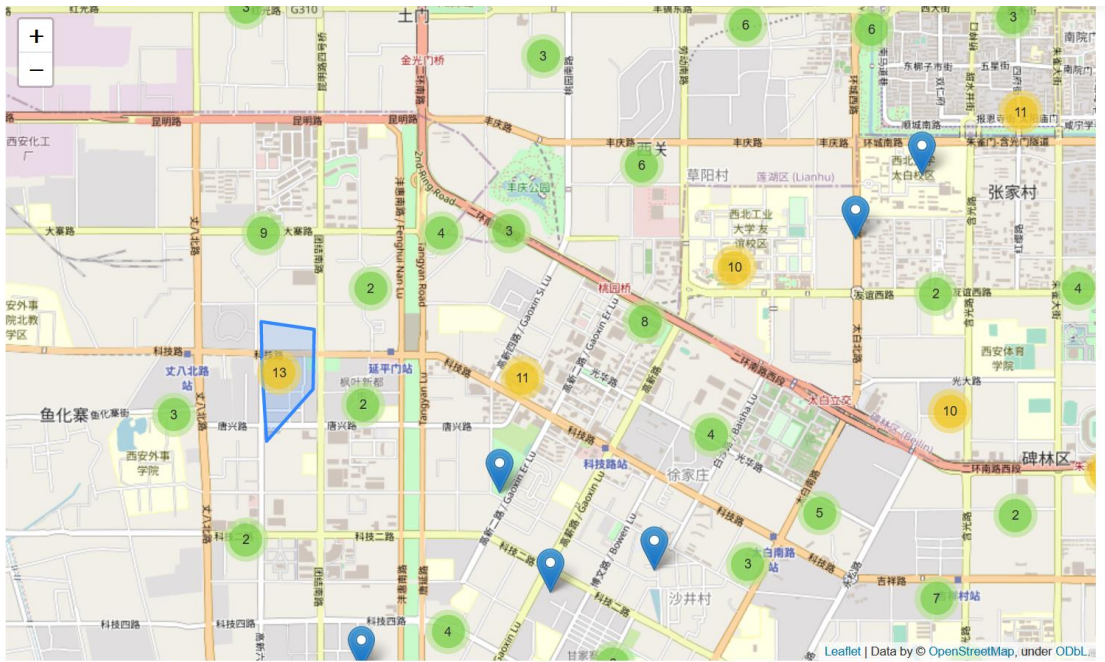


将聚类的分类结果作为变量加入到美团店铺位置可视化地图之中，如下图所示。每一个点表示一家餐饮店铺，其中红色 A 类，绿色 B 类，蓝色 C 类，黄色 D 类，紫色为 E 类。不难看出 B 类和 E 类餐厅相对较少，通过缩放地图来观察聚集点的分布特征，可以看出美团店铺的普遍集中于以下地铁口附近：1. 小寨站附近。2. 航天城站附近。3. 钟楼站附近。4. 永宁门站附近。5. 延平门站附近。



将美团店铺的数量按照区域的位置进行划分，通过缩放地图来调整地区的大小，进一步得出每个区域美团店铺的数量。继续不断细分区域，分别得出小寨站附近、钟楼站与永宁门站附近、航天城站附近、延平门站的店铺分布图。

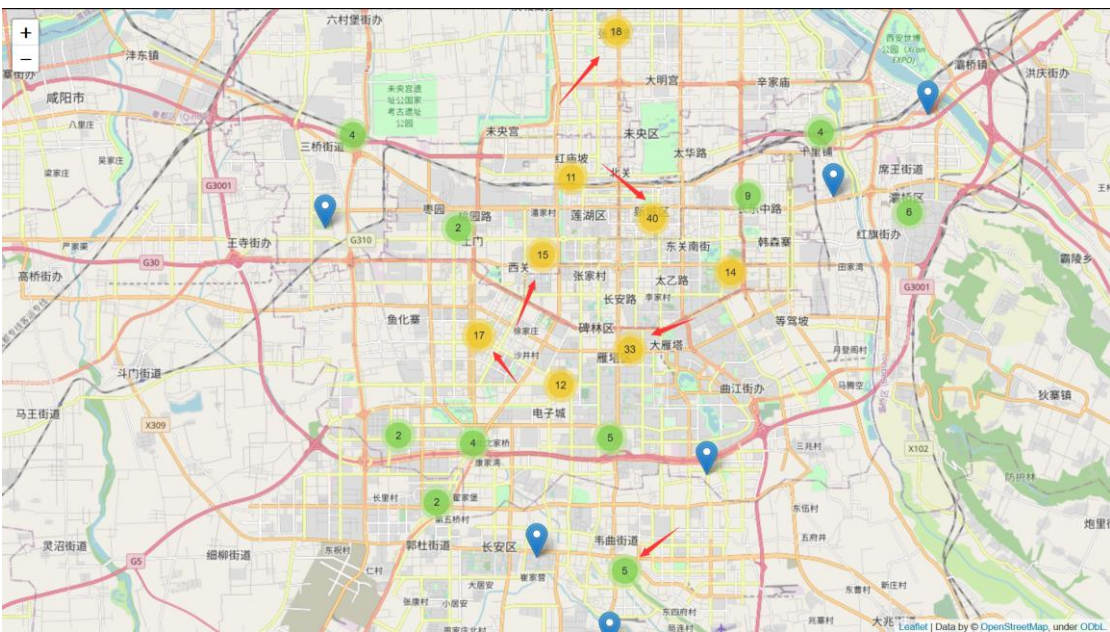


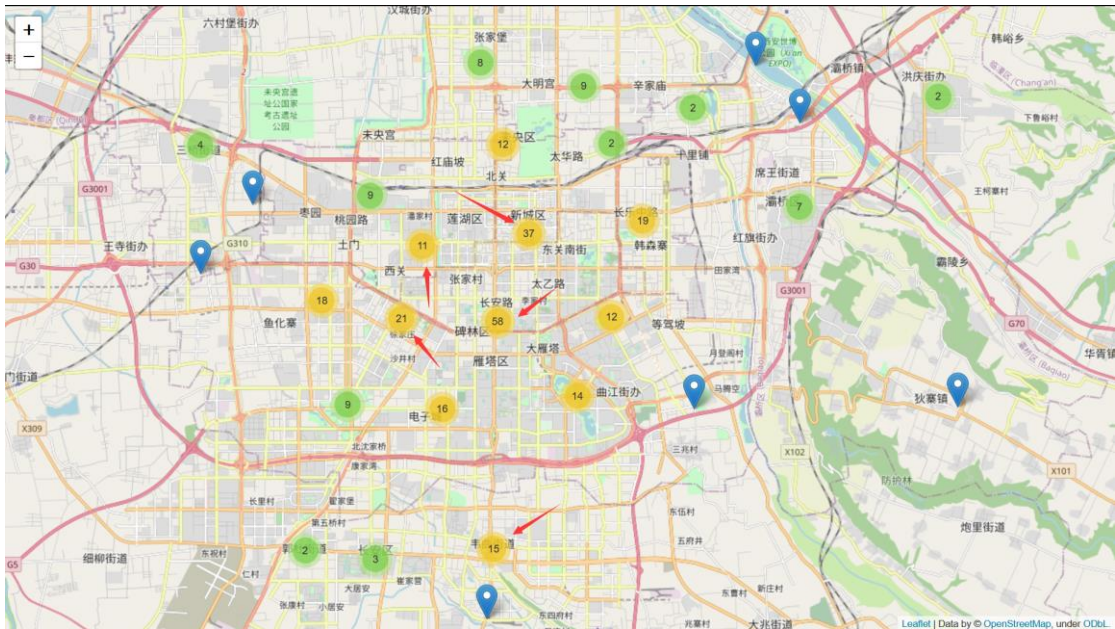
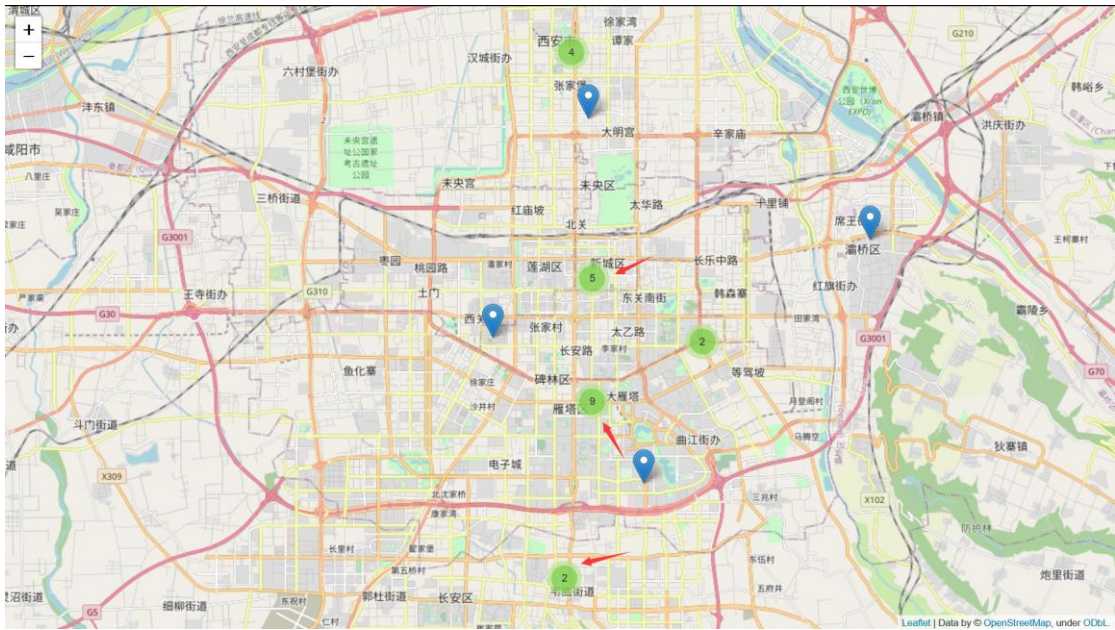


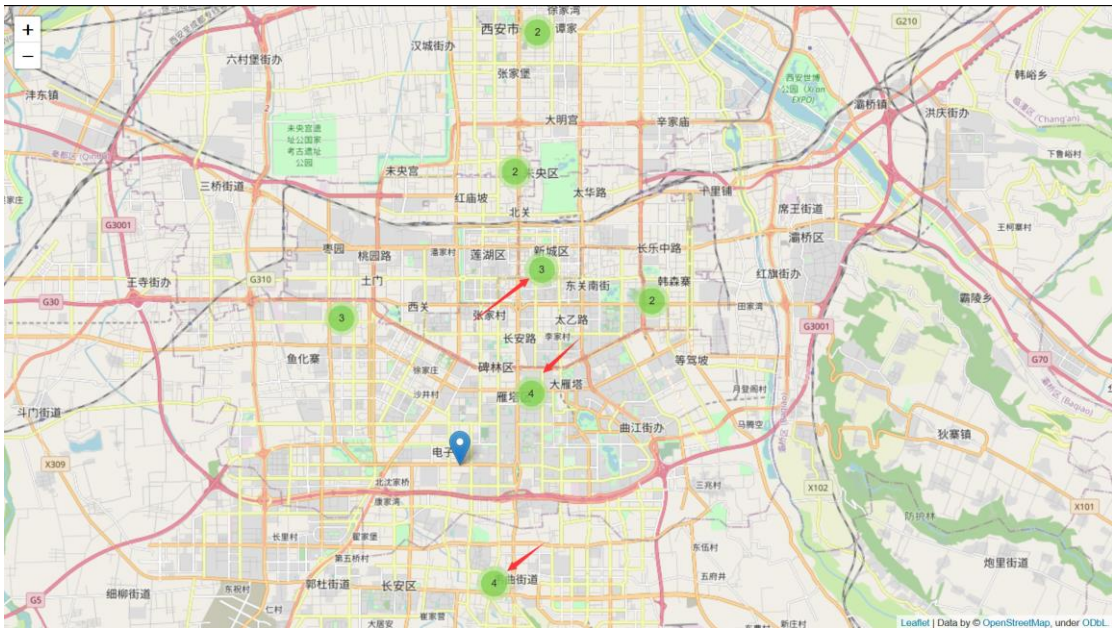
可以看出这些地铁口的餐饮店铺数量远大于周边地方的餐饮店铺的数量。其原因有以下几个方面：

1. 据统计，小寨区域有人口四十万，每日人流量约五十万人，是西安最密集的商圈之一，每天都有许多人去赛格国际购物中心进行购物，客流量极大。其次，小寨周边有着众多的高校，如长安大学、西安财经大学、西安交通大学、西安音乐学院，拥有着非常广阔的学生市场。
2. 钟楼作为西安市的中心，作为西安的地标性建筑，同样也是西安的另一大商圈，拥有着许多流行服装店，如 ZARA、名创优品、优衣库等，同样也有着许多高科技产品实体店，如 apple store。因此每天也拥有着极大的客流量。钟楼站与永宁门站之间的距离相对较近，会给永宁门站附近带来一定的客流量。
3. 延平门站位于西安高新区，高新区是西安极具特色的高新技术园区，成为许多市民购买房子的不二之选，观察地图不难看出延平门站附近被居民区所包围，因此有着非常大的餐饮需求。
4. 航天城地铁口位于地铁二号线的倒数第二站，由于地铁二号线并不能直达大学城，因此航天城站成为了与大学城进行连接的桥梁，陕西师范大学、西北大学、西北政法大学、西京大学等等的学生想要乘坐地铁去别的地方，航天城将会是学生的第一选择。

通过将 A 类、B 类、C 类、D 类、E 类五种类型的店铺与所有店铺的分布特征进行比较，发现 A 类、B 类、C 类、D 类、E 类五种类型店铺的集中分布情况与所有店铺的集中分布情况大致相吻合。下图对应 A 类、B 类、C 类、D 类、E 类五类店铺的分布情况。







就以小寨站与航天城站为例，列举出小寨站与航天城站附近的各种类型店铺：

航天城站

A 类	B 类	C 类	D 类	E
辣条印象	福临北京烤鸭（韦曲西街店）	味都西饼店（华美店）	刘大饼香辣土豆片夹馍（长安店）	韩味坊牛排自助（长安中央广场店）
星期八工坊	红透天自助涮烤（韦曲店）	和福顺养生焖锅（长百店）	臻膳轩自助涮烤城	李想大虾火锅花园（长百新市店）
欧味轩艺术蛋糕	王府臻品火锅（长安街店）		艾米客蛋糕	红透天自助涮烤（韦曲店）
川渝小渔哥	面道			

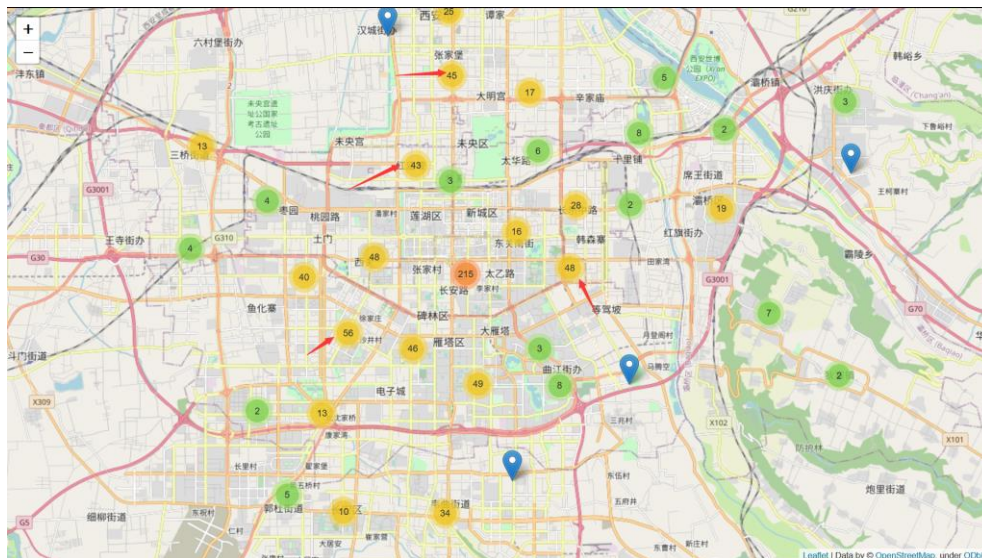
小寨站附近店铺

A 类	B 类	C 类	D 类	E 类 ‘
佛伦萨·古典火炉披萨（小寨店）s	蘑菇爱上饭（小寨店）	珍味林饺子馆（华旗国际广场店）	巷子火锅	千家粗粮王（小寨店）
猫堂小站猫咪主题馆	CoCo 都可（小寨赛格店）	福气焖锅烤肉	5 号酒馆	82°C 魔力焖锅（赛格国际店）
小肥羊（小寨金莎国际店）	长安大牌档之长安集市（赛格旗舰店）	泰熙家（飞炫店）	大自在火锅（金莎国际店）	拉菲达牛排自助（赛格国际店）
猫咪餐厅（小寨	京御煌三汁焖	赵家腊汁肉（小	米多多烘焙屋	瑞可爺爺的店

店)	锅 (赛格国际店)	寨店)		(赛格店)
阿姨奶茶专卖 (小寨店)	百富烤霸 (小寨汉唐店)	三姊妹香辣土豆片夹馍 (海港城店)	小哥过桥米线 (海港城店)	
太食獸泰式茶餐厅 (赛格国际店)		和記丸子專買	0057 香辣虾 (小寨店)	
M12 铁板餐厅 (赛格国际店)		重庆鸡公煲 (盛世商都店)		
洪氏嗨捞·新派猪肚鸡 (赛格国际店)				

营销。

4. 通过对比两地的 D 类和 E 类店铺的种类情况，不难看出自助餐饮和火锅类型餐厅的综合评价普遍不高。具有较大的自助和火锅市场，与此同时，也拥有着较大的风险。如何能让消费者满意，是市场进入者必须深思熟虑的问题。



除此之外，还有一些区域美团美食店铺数量相对较高，如上图所示。其原因可以归结为以下几个方面：

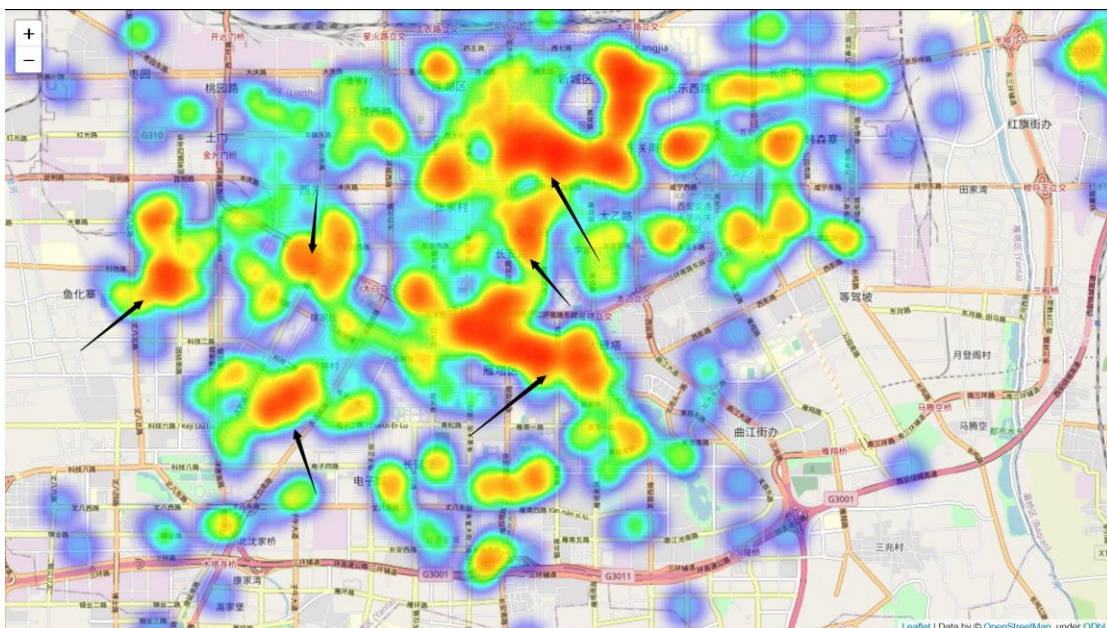
1. 科技路站附近地处高新区，经济发展形势较其他地方来说较为突出，同时也被学校和居民区所包围起来，具有较高的人流量。因此，餐饮业的基本收入有一定的保障，就会吸引不少餐饮业进入该市场。

2. 老城根商业街位于红庙坡，于 2016 年开业以来，为红庙坡吸引了一大批的人前来消费，为红庙坡的发展制造了一定的流量。与此同时，红庙坡附近有着较为聚集的居民区，具有很大的餐饮业市场。

3. 张家堡位于郊区地带，周围却具有非常多的美团美食店铺，不太符合常理。通过搜集张家堡相关的资料，得知西安市人民政府于 2015 年驻地由西安市莲湖区西华门大街迁至西安市未央区凤城八路，为张家堡附近发展吸引了一大批投资者，以及居民区向着张家堡方向转移，餐饮业的发展得到一定的保障，就会吸引餐饮业店铺进入张家堡市场。

4. 咸宁西路到咸宁东路延线上有着很多的美团美食店铺，通过放大地图，可以看出咸宁路两侧美团美食店铺的数量大致呈现出对称分布，具有较为聚积且呈现出一定对称分布的大学城，居民区也较为集中。因此餐饮业具有着一定的市场。

接下来，通过对所有美团店铺的分布做热力图看出每个区域美团店家的覆盖范围，同样也可以验证出店铺集中分布区域为商圈、景区、居民集中住宅区、大学城附近。



美团美食店铺的分布一般符合“人流量越大的地方，美团店铺就会越多”。但还是能找到一些人流量大的离群点，如火车站，汽车站，高铁站等。下图为火车站旁边美团餐饮店铺的分布图：



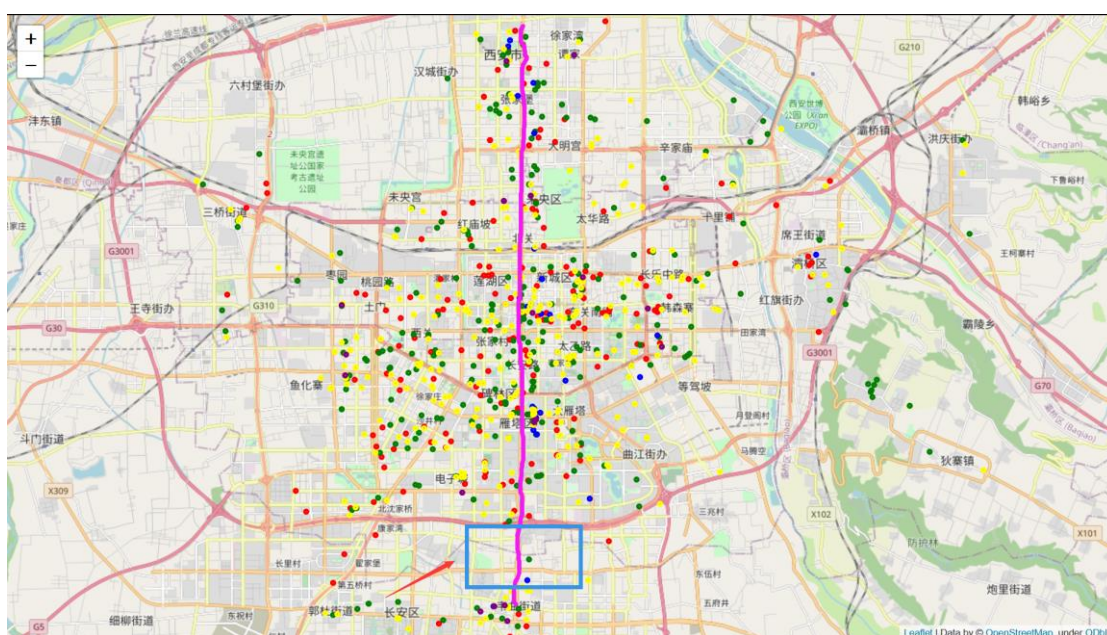
火车站和汽车站作为客流量极大的两个场所，但从图中可以看出火车站与汽车站旁边几乎没有美团店铺。这与“人流量越大的地方，美团店铺就会越多”的“道理”相违背。其原因有以下几点：

1. 火车站和汽车站附近的店铺与顾客的关系一般为一次性的，顾客基本上不会消费第二次。因此这些餐饮店铺为了获取更高的利益，就存在着“抬价”、“产品存在质量问题”的不良行为。若店铺开通美团服务的话，顾客不满意的话，就会给差评，影响潜在客户。
2. 火车站和汽车站周围的餐饮店铺凭借着地理位置的优势，经过火车站

的人都会看到店铺，因此只需雇人在门口吆喝拉客即可，不需要额外地给美团美食付平台费。

仔细观察店铺集中分布特征不难发现在未央路—北大街—长安北路—长安南路—航天城站这条主干道沿线基本上都有聚堆的美团店铺，但会发现中间有一段出现明显的短缺。通过查询长安南路相关资料，得出两大原因：

1. 长安南路周围在几年前处于未开发阶段，近些年才开始开发建设，附近并没有大规模的居民区，因此餐饮业相较于该主干道上的其他地方较为落后。
2. 长安南路基本上是汽车、摩托车修理厂和家居类产品的店铺，即使人流量也很大，但人均停留率相对极低，除非有修理汽车或者家居类产品的需求，否则人们在这条路上很难会选择停留下来，因此餐饮业很难会有市场。



2. 美团外卖平台

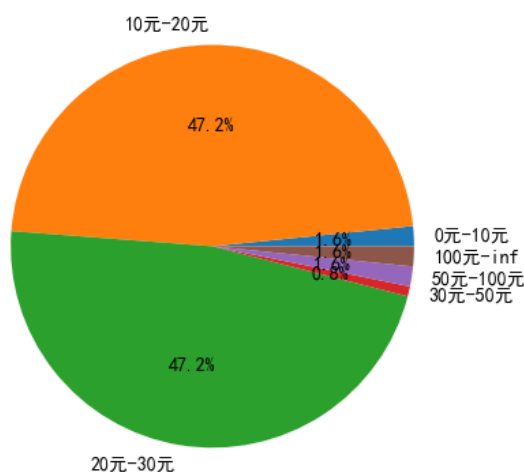
数据来源为 <https://h5.waimai.meituan.com/waimai/mindex/home>，爬取工具为 selenium, webdriver, BeautifulSoup。通过模仿人机登录，进入界面，破解美团加密的 CSS 字体，使用绘图工具描绘出字体的大致形状，利用 ORC 技术来识别对应的字体。爬虫只是获取数据工具，不予以详述。

2.1 数据分布特征

2.1.1 按照起送价分类：

观察饼状图不难看出西北大学长安校区周边外卖的起送价集中分布在 10-30 元，这很符合学生的行为。学生在校期间平均每顿饭在 15 元左右，同学之间也存在着拼单的现象，因此外卖的起送费集中在 10-30 元之间是学生可以接受的。若起送费过高，就会使得学生不点该外卖，而选择起送费相对较低的外卖，从而不利于店家的盈利。

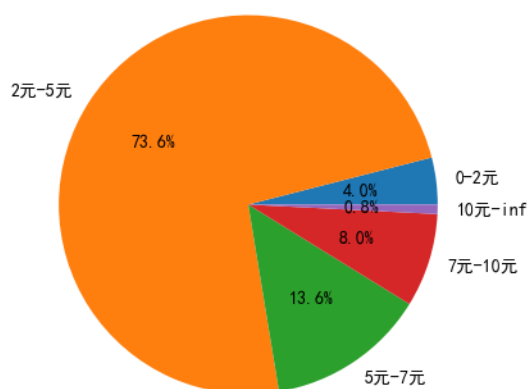
西北大学长安校区周边外卖起送价饼状图



2.1.2 按照配送费用和人均费用分类：

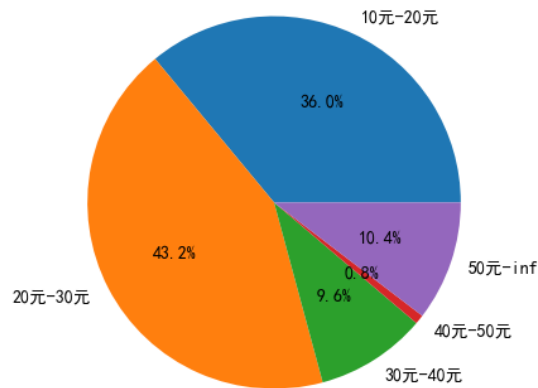
通过观察西北大学长安校区外卖配送费用饼状图，看出配送费用集中分布于2元-5元，这就很符合学生心理。若配送费用过高，就会导致学生选择去餐厅吃或者到店餐食，而不会选择点外卖。对于想吃学校外面美食而又不想出去的学生来说，出门既要花费时间，又有可能在交通工具上花费一定的钱，还不如点外卖，因此，2元-5元的外卖费是在可以接受范围内的。

西北大学长安校区周边外卖配送费用饼状图



大学生没有太多的收入来源，经济来源集中于家长给的生活费。吃饭是大学生生活支出中一大部分，每个人不可能在钱是有限的情况下，同时吃得好，穿的好，生活质量高。因此学生在吃饭的时候还考虑到每顿饭的价格，每顿饭10-20元是大多数学生日常每顿饭的价格，偶尔可以提升下生活质量，点一下比较贵的外卖，故外卖的人均价格集中在10-30元之间是比较合理的。

西北大学长安校区周边外卖人均价格饼状图

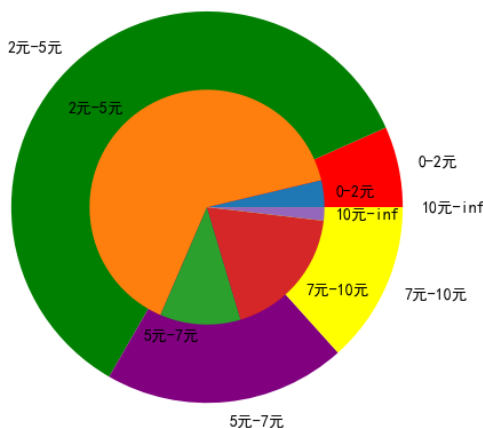


通过调用 matplotlib 制作出人均价格在 10-20 元和 20-30 元的外卖配送费的分布图，其中最外圈表示的人均价格在 10-20 元的外卖配送费分布情况，最内圈表示的是人均价格在 20-30 元的外卖配送费分布情况。

当人均价格在 10-20 元时，配送费集中分布于 2 元-5 元，不存在配送费在 10 元以上的店家。配送费 0-2 元的店家的比例为 6.7% 大于平均水平 4%，因为消费者很难购买配送费占比特别高的外卖，因此可以推断这些人均费用低的店家在学校附近，通过雇佣店员送外卖，收取极低或不收配送费。配送费 5-7 元所占的比例分别为 20.0% 大于平均水平 13.6%，同样根据消费者很难会选择购买配送费占比特别高的外卖，学生之间可以通过拼单来降低平均配送费用，故可以推断这些外卖的类型为黄焖鸡，炒米粉，汉堡，炸鸡等午餐、晚餐。

当人均价格在 20-30 元时，由于人均价格地提升，因此出现了配送费大于 10 元的店家。相对于 10-20 元的小吃、午餐等来说，人均价格提升相对店家来说没必要来自己雇佣员工来配送外卖，更加注重于食材，降低雇佣更多员工的费用，因此配送费 0-2 元的店家就会减少。配送费为 7-10 元比例为 18.5%，大于人均价格在 10-20 元时配送费为 7-10 元的比例 13.3%，佐证了人们很难会选择购买配送费用占特别高的外卖。

人均价格在10-20元和20元-30元的配送费分布图



2.2. 聚类分析

先利用 PCA 分析将月销售量、预计送达时间、距离、起送价、配送费、人均价格、得分这 7 个特征进行降维处理。其目的是避免维度灾难，舍弃掉不重要信息，从而极大地降低样本密度。与此同时，受到噪声影响的时候，较低的特征值的特征向量往往与噪声有关，舍弃掉较低特征值的特征向量在一定程度上可以达到降噪的效果。

样本集为 $D = \{x_1, x_2, \dots, x_{125}\}$ ，降低到的维度为 d' 。第一步，对所有样本进行中心化处理，即为 $x_i = x_i - \frac{1}{125} \sum_{k=1}^{125} x_i$ 。第二步，计算出样本的协方差矩阵 XX^T 。第三步，对协方差矩阵进行特征值分解。第四步，选择最大的 d' 个特征值所对应的特征向量作为投影矩阵，记为 $W = (w_1, w_2, \dots, w_{d'})$ 。

在本次实验中计算出特征值与特征矩阵分别为

[0.08080557 0.06147234 0.03054168 0.02265606 0.01754174 0.00696421 0.00317877] 和

[[9.99998853e-01 6.58868370e-04 1.12265664e-03 3.57043739e-04
-5.65657731e-04 -3.89782187e-04 1.30396446e-05]

[3.73506123e-04 4.24987575e-01 -7.77776782e-01 4.50984675e-01
-1.00428079e-01 -3.65849498e-03 3.10215862e-02]

[3.48424298e-04 8.07363106e-02 -4.82004447e-02 2.57180359e-02
5.79124859e-01 4.76319910e-02 -8.07986645e-01]

[-1.10304230e-03 8.85201372e-01 4.56882296e-01 -6.05332490e-02
-5.93276657e-02 1.35640246e-02 1.75457578e-02]

[4.44455706e-04 6.74877767e-02 -4.83595140e-02 -8.07735387e-03
8.05951879e-01 -4.23956629e-02 5.84538701e-01]

[6.51945177e-04 1.57206581e-01 -4.25241952e-01 -8.89132595e-01
-3.77127815e-02 -4.67603489e-02 -1.70115238e-02]

[4.40264142e-04 -4.09846858e-03 -2.87739799e-02 -4.08042693e-02
5.27465607e-03 9.96769658e-01 6.25500177e-02]]。我们前三项特

征值占比之和为 0.7744188467705035，在便于可视化的立场，本次实验中将原始数据投影的月销售量、预计送达时间、距离所对应的特征向量所构成的超平面中。

由于数据量较小，故采用 Gap Statistic 方法进行确定 K 值，具体公式为：

$Gap(K) = E(\log D_k) - \log D_k$ 。通过计算出不同 K 值下的 Gap 所对应的值，取最大值的索引作为该聚类 K 值。最终计算结果如下：

[-1.4210854715202004e-14, -0.0273180666809667, 0.0321432370756618,
-0.01023390275139846, 0.0016083358128637926, 0.02428196955558448,
-0.002149156671380048, 0.006404960372462298, 0.01969022814458654,
-0.030369389271641012]，其中 K=3 时，对应的 Gap 值最大，不妨将聚类划分成

3 类。

对数据进行处理后，采用 kMeans 算法来进行聚类。我们将西北大学长安校区附近的外卖分为三类（好、中、差）。

样本集为 $D = \{x_1, x_2, \dots, x_{125}\}$ ，聚类簇的划分为 $C = \{C_1, C_2, C_3\}$ ，最终目标函数

为 $\min \sum_{i=1}^3 \sum_{x \in C_i} (x - \mu_i)^2$ ，求取此目标函数的被证明是 NP 难问题，故采用贪心算法，

通过每一轮的迭代，不断优化目标函数。

算法具体过程为：

第一步：随机选择 3 个两个样本作为初始均值向量，设为 $\{\mu_1, \mu_2, \mu_3\}$ ，并将这两个样本构成两簇。

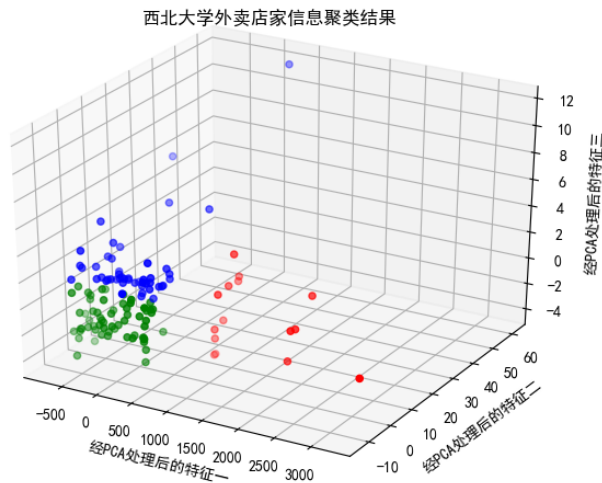
第二步：随机选择一个簇，将其设为空集。

第三步：计算每个样本点到均值向量的距离。选择距离最近的均值向量所在簇作为此样本点所对应的簇。

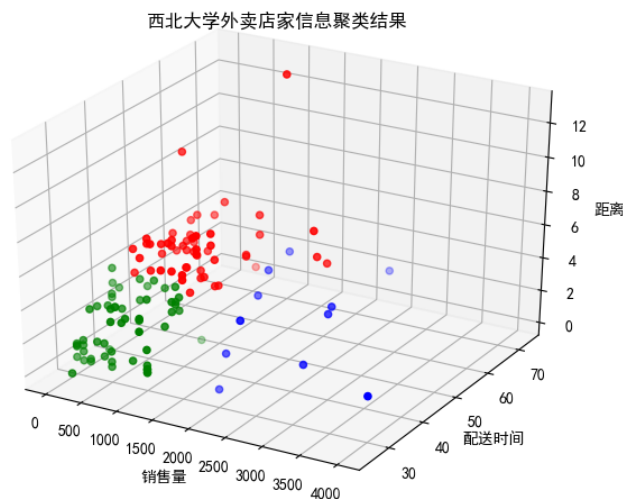
第四步：更新每个簇的均值向量 $\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$ ，若均值向量发生改变，则

跳转第二步。若均值向量不发生改变，则结束循环，返回簇划分情况。

按照以上算法，将西北大学长安校区外卖服务分为 3 类，一类为慕玛披萨（西安书香路店）、外婆印象（万科店）、小六汤包（韦郭路店）、德克士（长安三店）、泰熙家（樱花广场店）、益阳味道、老米家泡馍（万科润园店）、阿香米线（西安阳光天地餐厅）等等，更加偏向于小吃，快餐便食，普遍特点为相对人均价格较高。第二类为王菊面馆（郭杜旗舰店）、东勋干锅肉煎饭、西贝莜面村（西安阳光城店）、泰熙家（阳光天地店）、锋味私房菜（紫薇店）、刘大饼土豆片夹馍店、陕记老潼关肉夹馍旗舰店等等，这些店铺普遍特点为销量极高，评价好。第三类为肯德基宅急送（万科润园店）、湖南土菜馆（大仁店）、小鑫家煲仔饭·烤肉饭（西北大学店）、重庆鸡公煲（西北大店）、拴马桩擀面皮（师大店）、麦乐香汉堡（西北大学店）、金饭、重庆菜馆、邢老三糊辣汤、潮汕砂锅粥、飞鸿肉夹馍（樱花广场店）等等，这类的普遍特点为预期配送时间相对较快，价格较为适中，月销售量较为适中。对聚类结果进行可视化分析，得到下图，观察图形，可以分类情况较好。



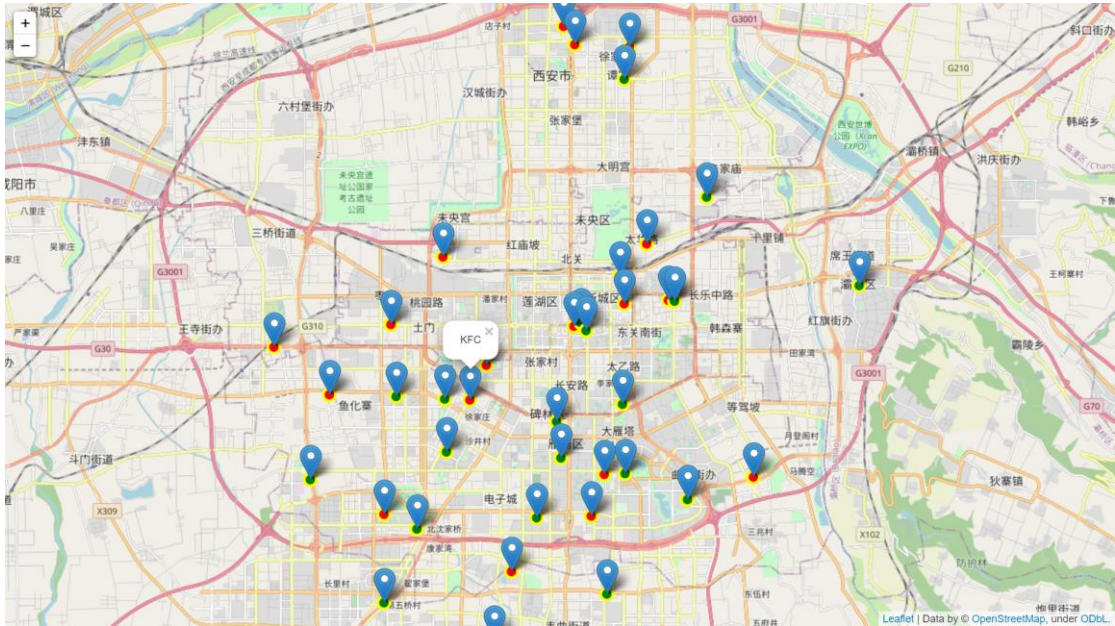
若不进行 PCA 降维处理，只选择特征值最高的三个特征作为样本特征进行聚类分析，将结果进行可视化处理后，发现每一簇中的样本之间相对较为离散，认为聚类效果并不理想，因此，经过 PCA 降维处理之后，将 7 维数据投影到一个 3 维的超平面上，不仅降低了样本的密度，还起到了一定的降噪作用。



3. 快餐类型分布情况：

在爬取美团外卖数据的过程中不难发现，肯德基、麦当劳这两家店铺的销量、得分远远超过其他店铺。因此，接下来对肯德基和麦当劳这两家店铺的空间分布特征进行分析。

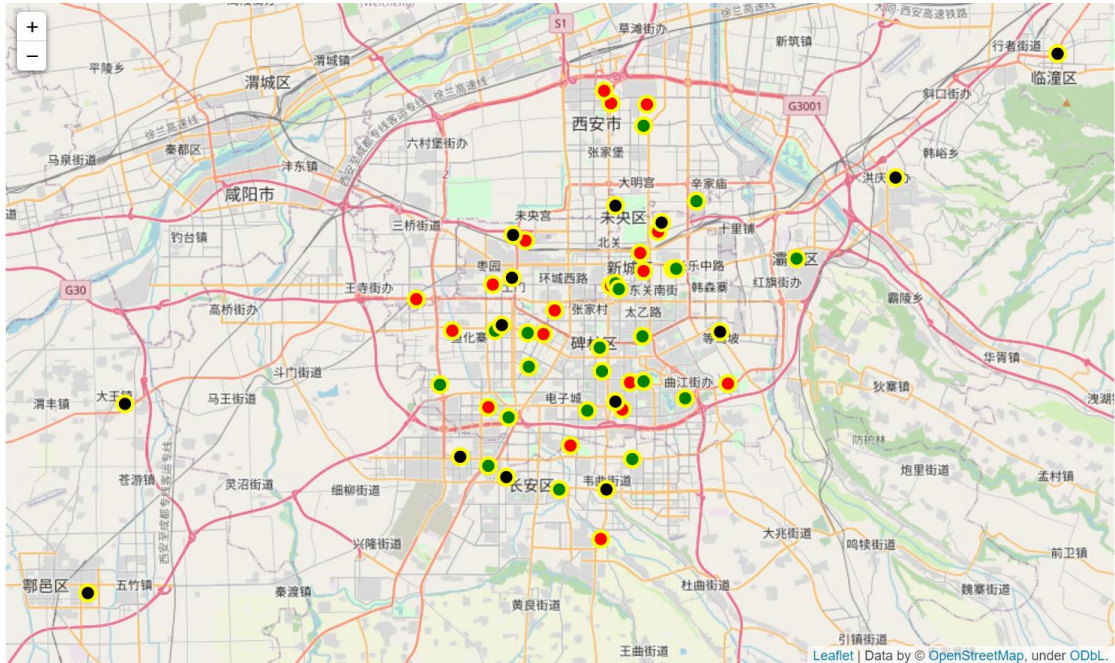
通过调用高德地图 API，获取到西安所有的肯德基和麦当劳店铺的经纬度，并调用 folium 包绘制出美团各家肯德基和麦当劳的分布情况。如下图所示：



观察地图不难看出，肯德基和麦当劳都集中分布于人流量大的商圈、地铁口、居民聚集区等，可以为店铺带来更好的收益。除此之外，大体上看麦当劳和肯德基店铺之间的位置相对较近。

通过经纬度距离计算公式，分别计算出任意两家麦当劳与肯德基之间的距离，公式为 $\hat{AB} = R \cdot \arccos(\cos A_\omega \cdot \cos B_\omega \cdot \cos(B_j - A_j) + \sin A_\omega \cdot \sin B_\omega)$ ，其中 A_j ， A_ω 代表 A 点的经度和纬度， B_j ， B_ω 代表 B 点的经纬度。将肯德基和麦当劳的数据代入到公式中，分别计算出不同店家之间的距离。得出在西安的 20 家店铺中，有 14 家肯德基周围 2.5km 之内至少有一家麦当劳。

再通过绘制麦当劳、肯德基和小众汉堡品牌（如华莱士、麦乐香、叫了只炸鸡、希多美炸鸡汉堡）的分布（红色代表肯德基，绿色代表麦当劳，黑色代表一些小众品牌），不难看出这小众品牌的店铺与肯德基和麦当劳的位置存在着较远的距离，这些店铺一般集中分布于附近没有麦当劳或肯德基的区域，且大多数位于郊区。



这些现象均可以通过博弈论进行解释。首先解释麦当劳附近大概率会至少有一家肯德基的现象，可以通过霍特林的空间竞争模型。

假设麦当劳先在一条商业街上开一家店，在这条商业街上的各个地方人流量均相等，由于肯德基和麦当劳所制作的食物具有同质性以及产品价格相同，顾客选择肯德基还是麦当劳完全取决于距离。这时麦当劳为了获取更大的收益，就会把店铺放在这条商业街的中心，从而使得店铺能够吸引更大范围的顾客。这时，肯德基也发现了这条商业街的商业价值，成为市场进入者，打破市场上麦当劳垄断的局面。为了更好的满足消费者，覆盖到更大范围的消费者，对于肯德基和麦当劳来说，就会选择两个 4 分位点作为店铺的位置，因此肯德基和麦当劳都可以覆盖 50% 的店铺。此时为社会最优的解决方案，消费者会根据距离的远近来选择去哪家店铺进行消费。第二天，麦当劳为了获取更大的利益，而向着肯德基的方向移动了总长度的 10%，肯德基并没有改变位置，这时麦当劳的覆盖范围就变成了 $50\% + 10\% / 2 = 55\%$ 。第三天，肯德基为了获取更大的收益，也向前移动了 10%，麦当劳并没有改变位置。此时，麦当劳和肯德基又出现了同时覆盖 50% 的店铺的情况。经过了若干天之后，肯德基和麦当劳都移动到商业街中心的时候，此时达到了纳什均衡，两家店无法通过移动位置来获取更高的收益。对于消费者来说并不是件好事，与两家店铺都位于 4 分位点的情况相比，消费者就需要多余走一些路程，才能进行消费。数学推导过程如下：

假设商业街的长度为 1，消费者在商业街内均匀分布，且总密度为 1，麦当劳在 $X=0$ 出开设商店，商品售价为 P_1 ，肯德基在 $X=1$ 出开设商店，商品售价为 P_2 。

成本均相等设为 C ，消费者需要为之付出商品的价格成本 C_1 以及走路的所花费的成本 C_2 ，走路所花费的成本与距离成正相关，设走路花费的单位成本为 t 。因此在 $[0, 1]$ 之内必能找到一个点 x ，使得位于 x 左边的消费者都去麦当劳，而 x 右

侧的消费者都去肯德基。得到方程： $p_1 + tx = p_2 + t(1-x)$ ，进行求解得到

$$x = \frac{P_2 - P_1 + t}{2t}。$$

同时也可以分布得到肯德基和麦当劳的商品和收益方程需求方程为：

$$\begin{cases} D_1 = x = \frac{P_2 - P_1 + t}{2t} \\ D_2 = 1 - x = \frac{P_1 - P_2 + t}{2t} \\ \pi_1 = D_1(P_1 - C) \\ \pi_2 = D_2(P_2 - C) \end{cases}$$

为了得出最大收益 $\max(\pi_1 + \pi_2)$ ，得出最优价格为 $P_1^* = P_2^* = C + t$ ，均衡利润为

$$\pi_1 = \pi_2 = \frac{t}{2}。$$

但经过肯德基和麦当劳经过上述的博弈过程，最终都到达了 $x=0.5$ 的位置，此时 $p_1 = p_2 = C, \pi_1 = \pi_2 = 0$ 。

接下来说明为什么小众品牌的店铺一般与肯德基、麦当劳不在一个区域，甚至在郊区。假设小众品牌可以做到与肯德基和麦当劳产品的无差别，会是在 $x=0$ 处开设一家小众品牌的店铺，并假设由于小众品牌的出现使得 $[0, x']$ 的消费者都会选择小众品牌进行消费。但由于小众品牌的规模限制，使得成本为麦当劳的 M 倍 ($M>1$)，此时 $p_3 + tx' = p_1 + t(0.5 - x')$

此时的需求与收益方程为

$$\begin{cases} D_3 = y = \frac{P_1 - P_3}{2t} + \frac{1}{4} \\ D_1 = 1 - y = \frac{P_3 - P_1}{2t} + \frac{3}{4} \\ \pi_1 = \frac{1}{4t}(2P_3 - 2P_1 + 3t)(P_1 - C) \\ \pi_3 = \frac{1}{4t}(2P_1 - 2P_3 + t)(P_3 - MC) \end{cases}$$

为 方 便 计 算 不 妨 令 $M=1.2$ ，为了得出最大收益 $\max(\pi_1 + \pi_3)$

$$P_1^* = \frac{7}{6}t + c, P_3^* = \frac{5t}{6} + c，此时 y = \frac{5}{12}。$$

这就很好地解释了为什么郊区的小众品牌多，小众品牌为了获得更大收益，吸收更多的用户，小众品牌必须向在市区进行偏移，但由于自身成本会大于麦当劳和肯德基，就会导致不能无休止向市区偏移，因此选择附近没有肯德基、麦当劳等大品牌的区域进行开店，使得附近的人

花宁愿花更多的钱或者花同样的钱吃品质不高的食物来降低走路带来的成本。