程序说明

- py
  - task1_1_multi_thread.ipynb

    调用高德api/调用hmap接口, 利用python线程池进行多线程调用Hmap-API，计算实际距离；

    获取的距离修改到`ods_july_only_cut_cal_0701_0707_0826_1.csv`中

  - task1_2_draw_dist_diff.ipynb

    筛选出 zke_lch_jyqk - Hmap_dist > 1 （考虑了Hmap与高德Amap本身的导航距离差异）的里程差异分布图（司机比导航多开的路程），画**箱线图**并统计数值特征；

    输出`task1_ods_july_diff_with_hmap_describe.csv`

	(3, 8]	(8, 13]	(13, 18]	(18, 23]	(23, 28]	(28, 33]	(33, 38]	(38, 43]	(43, 48]	(48, 53]	(53, 58]	(58, 63]
count	63601.000000	28443.000000	15580.000000	8674.000000	4948.000000	2763.000000	1559.000000	999.000000	751.000000	585.000000	382.000000	178.000000
mean	0.940102	1.285063	1.756655	2.246233	2.814924	3.794460	5.074239	5.438425	5.868609	5.089475	6.823353	10.709886
std	0.856781	1.325033	1.881974	2.535030	3.249465	4.229400	5.555130	6.501895	7.977701	7.544191	10.635034	9.487961
min	0.000034	0.000128	0.000356	0.000704	0.000871	0.000472	0.008368	0.006588	0.013207	0.043172	0.020630	0.186361
25%	0.294923	0.398623	0.566892	0.682551	0.830380	1.113007	1.469747	1.563013	1.685592	1.231687	1.080569	4.552842
50%	0.702028	0.926649	1.241414	1.544221	1.916451	2.578507	3.827732	3.512017	3.202317	2.889178	3.519510	8.468242
75%	1.339262	1.734287	2.310195	2.919757	3.587362	4.940042	6.658485	7.148673	6.421596	5.517716	7.685782	13.820713
max	7.859346	12.591456	17.771199	22.249807	27.911622	31.879625	37.620013	41.795241	46.717506	51.709087	57.610649	58.938212
upper_thresh	2.905771	3.737782	4.925150	6.275566	7.722834	10.680593	14.441590	15.527163	13.525601	11.946759	17.593603	27.722520

  - task2_1_geohash_combine.ipynb

    输入`k_top_router_0902.csv`

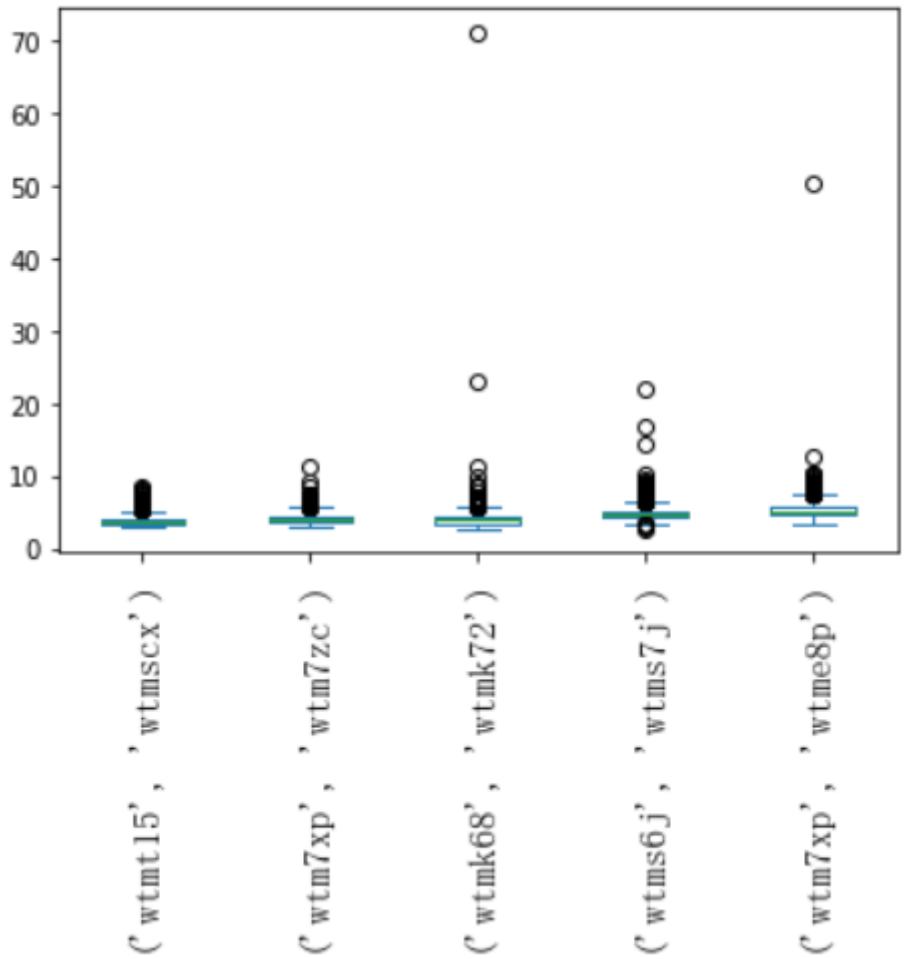
    处理geohash对， 为 A-B 和 B-A 都生成 A_B 的pairs字段，并存放到`all_pairs_hash.csv`

```
,pair_name,from,to
"('wthqgb', 'wthqs7')",wthqgb_wthqs7,wthqgb,wthqs7
"('wthqs7', 'wthqse')",wthqs7_wthqse,wthqs7,wthqse
"('wthqu0', 'wthqsp')",wthqu0_wthqsp,wthqu0,wthqsp
```

    并统计坐标对之间的订单实际里程分布关系

  - task2_2_plot_k_top.ipynb

    具体分析某坐标对之间的topk路线，直接画到图上,并保存为html



	('wtmt15', 'wtmscx')	('wtm7xp', 'wtm7zc')	('wtmk68', 'wtmk72')	('wtms6j', 'wtms7j')	('wtm7xp', 'wtme8p')
count	2400.0	2171.0	2023.0	1952.0	1914.0
mean	4	4	4	5	5
std	0.668451	0.680579	1.718908	0.896871	1.34363
min	3.03	3.04	3.01	3.01	3.71
25%	3.6	3.8	3.65	4.68	4.8
50%	3.945	4.18	4.2	5.025	5.29
75%	4.34	4.61	4.565	5.4	5.9
max	8.8	11.5	71.2	22.0	50.4
hmap_dist	5.101006	4.784046	4.576899	5.64056	5.521792
upper_thresh	5.45	5.825	5.9375	6.48	7.55

- sql
  - task_0_correct.sql

    修正异常经纬度数据，直接除以 1000000 即可...
  - task_0_filter_date.sql

    筛出某日期的记录
  - task_1_geohash.sql

    计算异常经纬度的geohash
  - task_1_rename_pair.sql

    根据 task2_1_geohash_combine.ipynb 得到的 all_pairs_hash.csv")

    导入PG new_hash
  - task3_2_K-top-router.sql

    没什么大用...