第三周

作业一：

目标：获取百度旅游目的地攻略--->目的地--->主题游的全部目的地景点详情信息（包括基本介绍、详情页面中的评论信息、以及全部行程路线的第一页的数据，获取完成后保存在mongodb）

具体页面流程如下

1.首先从这里进入



2.找到主题游的更多菜单点击进入景点的列表页面



3.筛选条件如下，获取如下条件对应的所有景点目的地详情信息

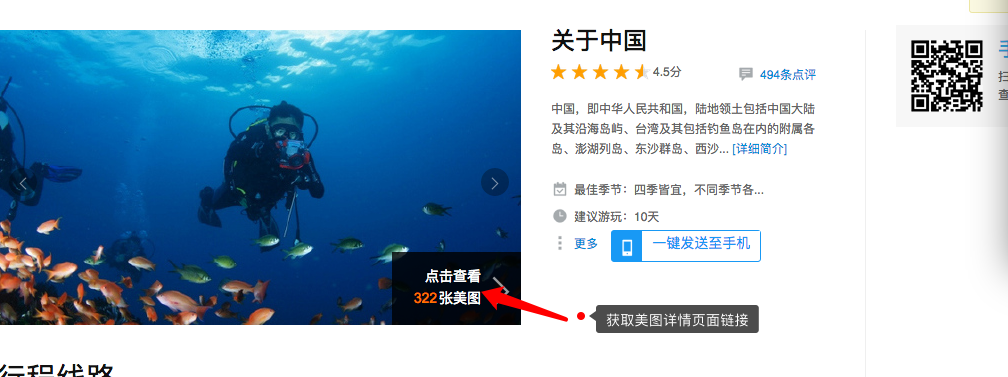




4示例：点击任意一个目的地进入详情，如下所示：

获取箭头所示的数据（名称、评分、描述、提示建议、完整的美图的详情页的跳转链接）



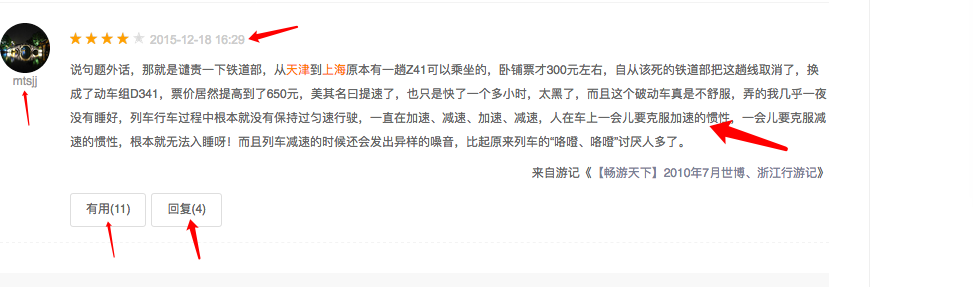


5.点击行共3000个行程字样，进入行程列表页面，如下所示，获取第一页的数据（箭头所指标题、关键字、路线1、路线2）





6.最后获取当前（目的地）详情页面中的所有评论信息内容如下（姓名、时间、评论内容、有用数、回复数）



7.将上面所有获取的数据保存在一个item下，并将这个item转换为字典存放在mongodb数据库中要求：

数据库名称：自己名称简写+baidulvyou

集合名称：自己名称简写+lvyou

例如：

数据库名称：ljh\_baidulvyou

集合名称：ljh\_lvyou

集合中的数据格式如下：

以下json数据的字段不分先后

{

'评分': '',

'目的地名称': ''

'美图页面完整链接': ''

'简介': ' ',

'行程总数': ' ',

'路线列表':（第一页的路线列表数据） [{'daydest1': '->->沱江->->->沱江跳岩->->->虹桥->->->回龙阁->->->万名塔->',

'daydest2': '->->虹桥->->->凤凰万寿宫->->->沈从文故居->->->熊希龄故居->->->凤凰广场->',

'keyword': '未知',

'title': '凤凰3日游',

'url': 'https://lvyou.baidu.com/plan/36ee73dbcbc3e4bee2abb689?request\_id=3560740672&idx=0'},

{‘同上’}

{‘同上’}

],

'评论总数': ' ',

'评论列表': （当前详情页中的所有评论）[{'available': '有用(16)',

'content': '\n'

'利用清明小长假\n',

'name': '顺时针1986',

'publishtime': '\n2016-04-07 21:08\n',

'replynums': '回复(87)'},

{'available': '有用(26)',

'content': ''很赞的地方，一定要去看看，和丽江古城是不一样的风格，主要是有一条长河，这样显得晚上特',

'name': 'super晋小辣椒',

'publishtime': '\n2015-05-19 18:52\n',

'replynums': '回复(50)'},

{结构同上}，

{结构同上}，

{结构同上}，

]

}