```
'''
全部信息  //div[@class="textPart_CDl999"]
    店名  //div[@class="nameLine_h53bsq"]
    信息    //div[@class="infoLeft_MNSTSp"]
        月售  //div[@class="infoCount_SzLoSC"]
        起送    //div[@class="infoCount_SzLoSC"]
        配送费  //div[@class="infoCount_SzLoSC"]
    营业信息  //div[@class="recommendLine_ghO92c"]
        24小时营业信息  //div[@class="recommendItem_RZbLVG"]/img
        建议信息  //div[@class="recommendItem_RZbLVG"]
    优惠标签  //div[@class="d_sublabel-block"]
        优惠内容    //div[@class="d_sublabel"]
最终要保存到excel的DateFrame对象的结构
data = [
    {
    '店名': 'xxxx',
    '月售': 44.0,
    '起送': 28.0,
    '配送费': 0.0,
    '建议信息1': '24小时营业',
    '建议信息...': '刚刚有用户看过',
    '优惠内容1': '30减1',
    '优惠内容...': '60减2',
    },
    {...},
]
'''
from lxml import etree
from tqdm import tqdm
import datetime, pandas, json

# 初始化相关xpath变量
xpath1 = {
    'all_info':'//div[@class="textPart_CDl999"]',
    'store_name':'//div[@class="nameLine_h53bsq"]',
    'base_info':'//div[@class="infoCount_SzLoSC"]',
    'suggestion_info':'//div[@class="recommendItem_RZbLVG"]',
    'discount_content':'//div[@class="d_sublabel"]',
    'sheet_name':'美团h5数据',
    'path':'C:\\Users\\Administrator\\Desktop\\美团爬虫\\',
    'html_name':'美团外卖.mht',
    'excel_name':'美团数据',
    'format_time':'%Y-%m-%d-%H-%M-%S'
}

# 配置文件路径，方便网站更新后，直接修改配置文件即可
json_url = 'config.json'
# 读取配置文件
try:
    with open(json_url, 'r', encoding='utf-8') as f:  # 确保文件编码为utf-8
        config = json.load(f)
```

```
        xpath1 = config
        print(f'配置文件{json_url}已成功读取\n正在加载数据')
except Exception as e:
    print(f'配置文件{json_url}不存在,已使用默认配置\n错误代码：{e}')
    try:
        with open('config.json', 'w', encoding='utf-8') as f:
            json.dump(xpath1, f, ensure_ascii=False, indent=4)
    except Exception as e:
        print("创建默认配置文件config.json发生了异常\n错误代码：{e}")



all_info, store_name, base_info, suggestion_info, discount_content, sheet_name, path,
html_name, excel_name, format_time = xpath1.values()
# 获取当前时间，定义最终保存的文件名
excel_name = f"{path +
excel_name}_{datetime.datetime.now().strftime(format_time)}.xlsx"
# 创建解析器对象
parser = etree.HTMLParser()

# 使用解析器对象
try:
    tree = etree.parse(path + html_name, parser=parser)
except Exception as e:
    input(f"使用解析器对象解析{path + html_name}发生了异常\n错误代码：{e}")
    exit()
# 即将保存的数据data
data = []
elements = tree.xpath(all_info)
print(f"共{len(elements)}条数据")
for element in tqdm(elements):
    d1 = {}
    d1["店名"] = element.xpath("." + store_name + "//text()")[0]
    d1["月售"] = float(element.xpath("." + base_info + "//text()")[0].replace('+', ''))
    d1["起送"] = float(element.xpath("." + base_info + "//text()")[1])
    d1["配送费"] = float(element.xpath("." + base_info + "//text()")[2])

    # 添加建议内容
    for i, suggestion in enumerate(element.xpath("." + suggestion_info + "//text()"),
1):
        d1[f"建议信息{i}"] = suggestion
    # 添加优惠内容
    for i, discount in enumerate(element.xpath("." + discount_content + "//text()"),
1):
        d1[f"优惠内容{i}"] = discount
    data.append(d1)
print(f"数据已成功爬取，共{len(data)}条")
# 将数据保存到excel文件中
df = pandas.DataFrame(data)  # 将data转换为DataFrame对象
# 将 DataFrame 写入 Excel
df.to_excel(excel_name, sheet_name=xpath1['sheet_name'], index=False)
input(f"数据已成功保存到{excel_name}\n按回车键退出")
```