This file is made by Tongze Wang (twang141 / 1702666)

File Explaination

- `/data/stations` contains the schedule for a specific train station (scraped from ip138.com using `trainList.py`)
- `/data/trains` contains the schedule for a specific train (scraped from ip138.com using `trainSchedule.py`)
- `/data/tripsInfo` contains the cleaned version of all the train data needed in this project (generated using `cleanData.py`)
- `/data/tripsInfoRaw` contains the raw version of all the detailed train data (scraped from 12306.cn using `trainDataRaw.py`)
- `transfer.py` contains the beta version of the train tranfer search algorithm
- `jiangsu_english.csv` contains all the english names of the train stations located within Jiangsu Province
- `zhejiang_english.csv` contains all the english names of the train stations located within Zhejiang Province
- `shanghai_english.csv` contains all the english names of the train stations located within Shanghai City
- `requirements.txt` contains all pip3 dependencies needed in this project

Libraries Used

Python3: beautifulsoup4, JSON, requests, re, numpy, pandas, selenium

```
In [ ]:  #!/bin/bash
         pip3 install -r requirements.txt
```

1. Scrape the Data

All the scraped data is already saved in this repo. You may use the following scripts to scrape the lastest data at your own risk as it may fail in your network environment.

You can use `trainList.py` to scrape the schedule of a spcific train station. The result will be saved to `/data/stations` folder.

```
In [ ]:  #!/bin/bash
         python3 trainList.py
```

You can use `trainSchedule.py` to scrape the schedule of a spcific train. The result will be saved to `/data/trains` folder.

```bash
In [ ]:  #!/bin/bash
         python3 trainSchedule.py
```

You can use `trainDataRaw.py` to scrape the raw version of all the detailed train data. The result will be saved to `/data/tripsInfo` folder.

```bash
In [ ]:  #!/bin/bash
         python3 trainDataRaw.py
```

> **`trainDataRaw.py` will send about 17,000 requests to 12306.cn to scrape the data. It will take hours to complete!!! We also saved the lastest data in `data/tripsInfoRaw.zip`. You can unzip it and so that it creates a folder called `data/tripsInfoRaw/`**

2. Clean the Data

You can use `cleanData.py` to clean all the data we scrapted from other websites. It will generate the `all.json` file under the `data/tripsInfo` folder as the result.

```bash
In [ ]:  #!/bin/bash
         python3 cleanData.py
```

3. Try the Train Transfer Search Algorithm (Extra Part)

> Note: this is the beta version of the train transfer search algorithm. This algorithm is not incorportated in our web vistuallization page since it's an extra part that is not required in the project due to its difficulty to build the code. This part is here because we want to show our extra efforts commited to this project. There might be bugs.

Currently, the search algorithm only consider tranfer trips with one stop. Some smaller train stations may cause error in this algorithm as it doesn't have enrough trains connected to bigger trains stations. The total time needed may not be displayed correctly for all the time.

We could use the following code to try to get the top 5 cheapest tranfer route and top 5 fastest tranfer route that could be used to travel from Shanghai Hongqiao to Hangzhou:

In [23]:
```python
from transfer import getTransfer

start_station = 'Shanghai Hongqiao'
end_station = 'Hangzhou'

# also try these two:
#
# start_station = 'Nanjing'
# end_station = 'Shanghai Hongqiao'
#
# start_station = 'Ningbo'
# end_station = 'Hangzhou'

getTransfer(start_station, end_station)
```

Shanghai Hongqiao can reach 77 stations without transfer
There are 31 possible stops that can be used to reach Hangzhou from Sh
anghai Hongqiao in 1 transfer

Shanghai Hongqiao ----> Hangzhou with 1 transfer
====== Top 5 Cheapest Transfer Route ======
¥ 72.5 / 04 hr 49 min : Shanghai Hongqiao (07:34) ===<D2212>===> (08:0
9) Suzhou (08:11) ===<Z281>==> Hangzhou (12:23)
¥ 73.5 / 05 hr 47 min : Shanghai Hongqiao (07:02) ===<D2287>===> (07:2
0) Jinshan North (11:54) ===<G7351>==> Hangzhou (12:49)
¥ 76.0 / 01 hr 43 min : Shanghai Hongqiao (06:50) ===<G7501>===> (07:1
3) Jiashan South (07:52) ===<G7395>==> Hangzhou (08:33)
¥ 76.0 / 01 hr 38 min : Shanghai Hongqiao (06:55) ===<G1651>===> (07:2
1) Jiashan South (07:52) ===<G7395>==> Hangzhou (08:33)
¥ 77.0 / 05 hr 17 min : Shanghai Hongqiao (07:32) ===<G1341>===> (07:5
0) Jinshan North (11:54) ===<G7351>==> Hangzhou (12:49)

====== Top 5 Fastest Transfer Route ======
01 hr 38 min / ¥ 76.0 : Shanghai Hongqiao (06:55) ===<G1651>===> (07:2
1) Jiashan South (07:52) ===<G7395>==> Hangzhou (08:33)
01 hr 43 min / ¥ 76.0 : Shanghai Hongqiao (06:50) ===<G7501>===> (07:1
3) Jiashan South (07:52) ===<G7395>==> Hangzhou (08:33)
01 hr 45 min / ¥ 77.0 : Shanghai Hongqiao (11:04) ===<G7461>===> (11:2
2) Jinshan North (11:54) ===<G7351>==> Hangzhou (12:49)
02 hr 56 min / ¥ 125.0 : Shanghai Hongqiao (06:55) ===<G1651>===> (08:
26) Zhuji (08:43) ===<T7786>==> Hangzhou (09:51)
03 hr 03 min / ¥ 126.0 : Shanghai Hongqiao (08:48) ===<G7104>===> (09:
16) Suzhou (09:17) ===<Z85>==> Hangzhou (11:51)

You may also choose any `start_station` and `end_station` from the following lists.

```
In [9]:  srtName_SH = ['Anting North',
                       'Jinshan North',
                       'Nanxiang North',
                       'Shanghai',
                       'Shanghai Hongqiao',
                       'Shanghai South',
                       'Shanghai West',
                       'Songjiang Railway',
                       'Songjiang South',
                       ]
         # All the station name in Shanghai City
         # 安亭北 金山北    南翔北 上海 上海虹桥 上海南 上海西 松江 松江南


         # All the station short code in Zhejiang Province

         staName_ZJ = ['Cangnan',
                       'Changshan',
                       'Changxing',
                       'Changxing South',
                       'Deqing',
                       'Deqing West',
                       'Fenghua',
                       'Fuyang',
                       'Haining',
                       'Haining West',
                       'Hangzhou',
                       'Hangzhou East',
                       'Huzhou',
                       'Jiande',
                       'Jiangshan',
                       'Jiashan',
                       'Jiashan South',
                       'Jiaxing',
                       'Jiaxing South',
                       'Jinhua',
                       'Jinhua South',
                       'Jinyun',
                       'Jinyun West',
                       'Kaihua',
                       'Lanxi',
                       'Yueqing',
                       'Linhai',
                       'Lishui',
                       'Longyou',
                       'Ningbo',
                       'Ninghai',
                       'Pinghu',
                       'Pingyang',
                       'Qiandaohu',
                       'Qingtian',
                       'Zhangzhou',
                       'Ruian',
                       'Sanmen County',
                       'Sanyang',
                       'Shangyu',
```

```python
                        'Shaoxing',
                        'Shaoxing North',
                        'Shaoxing East',
                        'Weifang',
                        'Taizhou',
                        'Tonglu',
                        'Tongxiang',
                        'Wenling',
                        'Wenzhou',
                        'Wenzhou South',
                        'Wuyi',
                        'Wuyi North',
                        'Yandangshan',
                        'Yiwu',
                        'Yongjia',
                        'Yongkang',
                        'Yongkang South',
                        'Yuhang',
                        'Yuyao',
                        'Yuyao North',
                        'Zhuangqiao',
                        'Zhuji'
                        ]

# All the station name in Zhejiang Province
# 苍南    常山      长兴      长兴南    德清
# 德清西  奉化      富阳      海宁      海宁西
# 杭州    杭州东    湖州      建德      江山
# 嘉善    嘉善南    嘉兴      嘉兴南    金华
# 金华南  缙云      缙云西    开化      兰溪
# 乐清    临海      丽水      龙游      宁波
# 宁海    平湖      平阳      千岛湖    青田
# 衢州    瑞安      三门县    三阳      上虞
# 绍兴    绍兴北    绍兴东    绅坊      台州
# 桐庐    桐乡      温岭      温州      温州南
# 武义    武义北    雁荡山    义乌      永嘉
# 永康    永康南    余杭      余姚      余姚北
# 庄桥    诸暨

staName_JS = ['Baohuashan',
                        'Bencha',
                        'Binhai Port',
                        'Changzhou',
                        'Changzhou North',
                        'Danyang',
                        'Danyang North',
                        'Donghai County',
                        'Suining',
                        'Suining East',
                        'Ganyu',
                        'Haian',
                        'Haimen',
                        'Huaian',
                        'Huaqiao',
                        'Huishan',
                        'Jiangdu',
                        'Jiangning',
```

```python
                'Jiangning West',
                'Jiangyan',
                'Jurong West',
                'Kunshan',
                'Kunshan South',
                'Lianyungang',
                'Lianyungang East',
                'Fushui',
                'Fuyang',
                'Nanjing',
                'Nanjing South',
                'Nantong',
                'Zhangzhou',
                'Qidong',
                'Qishuyan',
                'Rudong',
                'Rugao',
                'Sheyang',
                'Shuyang',
                'Siyang',
                'Suzhou',
                'Suzhou North',
                'Suzhou New District',
                'Suzhou Park',
                'Taizhou',
                'Wawushan',
                'Wuxi',
                'Wuxi East',
                'Wuxi New District',
                'Xiangshui County',
                'Xianlin',
                'Xinyi',
                'Xuzhou',
                'Xuzhou East',
                'Yancheng',
                'Yancheng North',
                'Yangcheng Lake',
                'Yanghe',
                'Yangzhou',
                'Yixing',
                'Zhenjiang',
                'Zhenjiang South'
                ]

# All the station name in Jiangsu Province
#  宝华山   栟茶      滨海港     常州      常州北
#  丹阳     丹阳北     东海县     阜宁      阜宁东
#  赣榆     海安      海门      淮安      花桥
#  惠山     江都      江宁      江宁西     姜堰
#  句容西   昆山      昆山南     连云港     连云港东
#  溧水     溧阳      南京      南京南     南通
#  邳州     启东      戚墅堰     如东      如皋
#  射阳     沭阳      泗阳      苏州      苏州北
#  苏州新区          苏州园区    泰州      瓦屋山      无锡
#  无锡东   无锡新区   响水县     仙林      新沂
#  徐州     徐州东     盐城      盐城北     阳澄湖
#  洋河     扬州      宜兴      镇江      镇江南
```