## 预测电商快递包裹数—记录文档

## 一、项目背景

常规报告中新增了<电商快递包裹数>这一指标，因此需要根据公开信息推算各地域的快递包裹数量。由于时间有限，本次只计算了合作地域的快递包裹数量。

## 二、项目时间

2019-03-07 -- 2019-03-28

## 三、指标定义

此处计算的**电商快递包裹数**是指：映潮监控体系内的电子商务经营者根据顾客提交的线上订单，借助快递物流公司将商品从发货地发往收货地时产生的快递包裹件数的总和。

## 四、推算方法

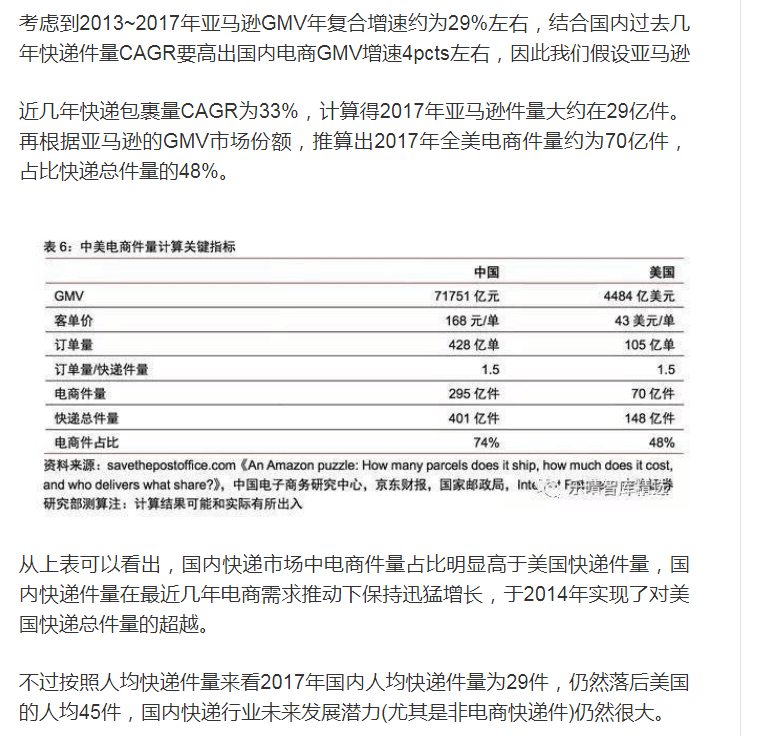
### 1.公开信息

（1）各省市邮政管理局公开数据报告

如：成都市邮政管理局---[成都市邮政管理局公布2019年2月邮政业发展情况](http://sccd.spb.gov.cn/xytj_9154/tjxx/201903/t20190320_1797115.html)

报告中会定期（月度）公布快递服务企业业务量（发出量）

（2）相关报告中比较权威的电商件占比数据，如下：



### 2.基本思路

（1）以统计局公布的数据为权威参考数据，即计算出的电商件数不应高于统计局公布的快递包裹总数。

（2）根据2017电商快递件数占比总快递件数的比例，推算2017月度电商件数量

（3）认为电商快递件数的月度环比趋势与映潮实物型网络零售额月度环比趋势有强相关关系。

### 3.模型选择

预测模型采用Facebook开源的时间序列预测框架prophet，目前支持R语言和python语言。

（一）模型原理



其中 g(t)表示趋势项，它表示时间序列在非周期上面的变化趋势； s(t) 表示周期项，或者称为季节项，一般来说是以周或者年为单位； h(t) 表示节假日项，表示在当天是否存在节假日；εt表示误差项或者称为剩余项。Prophet 算法就是通过拟合这几项，然后最后把它们累加起来就得到了时间序列的预测值y(t)。

具体原理可参见：[Facebook 时间序列预测算法 Prophet 的研究](https://zr9558.com/2018/11/30/timeseriespredictionfbprophet/)

（二）环境搭建

（1）首先需要配置python环境，可通过安装anaconda安装python3。

（2）prophet基于PyStan计算平台，因此需要先安装Pystan。步骤：

1. 安装依赖项：C ++编译器，在cmd中运行：

conda install libpython m2w64-toolchain -c msys2

2.安装依赖项numpy 和 cython，在cmd中运行：

conda install numpy cython -c conda-forge

3.安装PyStan，可选：

pip安装，运行：pip install pystan

或者

conda安装，运行：conda install pystan -c conda-forge

安装过程参见：[Installing PyStan](https://pystan.readthedocs.io/en/latest/windows.html)

（3）安装prophet：pip install fbprophet

安装及使用说明参见：<https://github.com/facebook/prophet/tree/master/python>

### 4.推算过程

1.预测下一季度快递业务总量（统计局公布的快递服务企业业务量），为当地发出快递件数的上限。

2.选取2017年7、8月份，根据电商件占比总量数据求解电商件数量。

选则2017年7/8月的原因：

（1）7月、8月节假日少，无重要电商节，因此电商件数增长变化受节假日因素影响小，月度变化小

（2）据抽样观察，各地统计局公布的快递包裹数在七月、八月的变量幅度最小，同时实物型网络零售额变化幅度也较小，可侧面证明电商快递件数月度波动较小，在总快递包裹数中占比相对稳定；在进行非线性规划问题求解时，最优解的区间相对较小，可降低求解计算的次数，节省时间。

3.假设电商快递件数变化趋势与当地实物型网络零售额变化趋势基本相同，即电商快递件数月环比是实物型网络零售额月环比的n%，选取多个n的取值，计算201707—201903的电商快递包裹数，最后根据计算结果判断n的取值在[89.9,92.5]区间内较为合理。因此将电商快递件数的月度环比设为n%,n为[89.9,92.5]内的随机值。

### 5.推算结果

以成都市和北京市为样本进行方法测试，预测出的快递包裹总数与电商快递件数的图像如下所示。



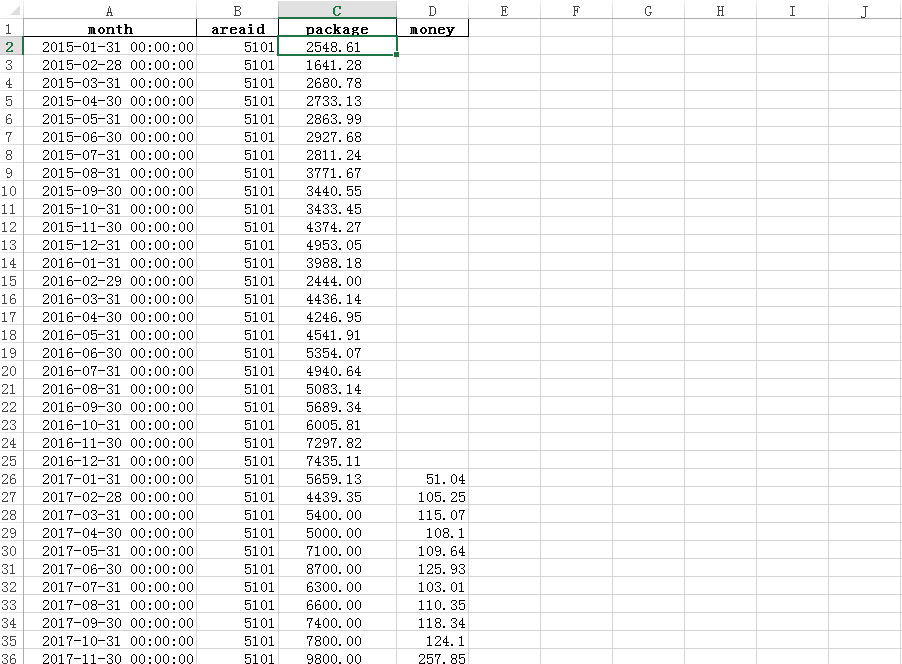
此推算过程也可参考文件:预测各省市快递包裹数.xlsx

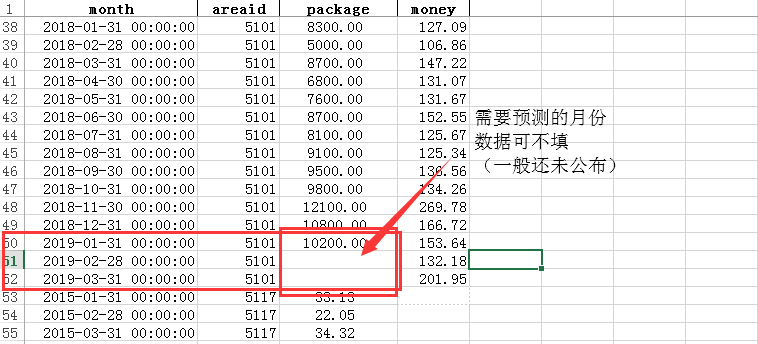
## 五、使用说明

### 1.excel数据表格式介绍

（1）数据表命名：excel数据表必须命名为：package\_data.xlsx

（2）数据表结构：如下图所示，数据表中共有四个字段，其中，month为月份，日期为每月最后一天；areaid为需要预测的地域id；package为该地区邮政管理局公布的**快递服务企业业务量**，其中需要预测的月份（也就是末尾的三个月份，最后一个季度）的数据可不填，一般此时邮政管理局还未公布；money为该地域对应月份的实物型网络零售额，从综合管理平台查询获得，其中2017年以前的数据可为空。





### 2.python程序文件介绍

（1）python程序文件：predict\_package.py，需要和package\_data.xlsx放在同一文件夹下。

### 3.使用步骤

**step 0. 前置条件**

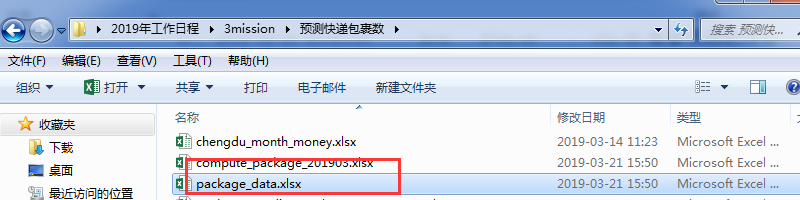
1.配置环境首先需要按4.3节配置环境。

2.准备数据：各省市的快递包裹数数据。此数据由数源部采集，保存在数据库中，因此可以从数据库中查询，并按照统一格式将新数据新增进excel数据表中；也可自己手动到邮政管理局网站复制粘贴，但要注意将单位统一为万件。

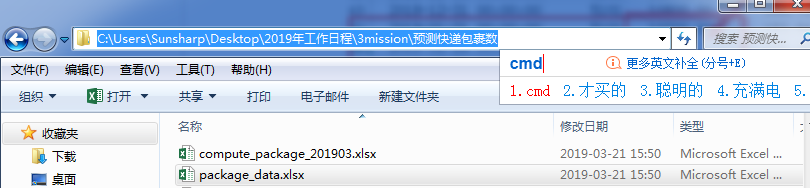
注意：第一次采集数据质量很差，已经将问题反馈给数源部，并手动修改数据。再次使用数据时要仔细检查；如有问题，及时反馈给数源部改善。

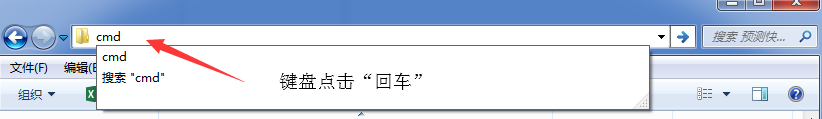
**step 1. 开始预测**

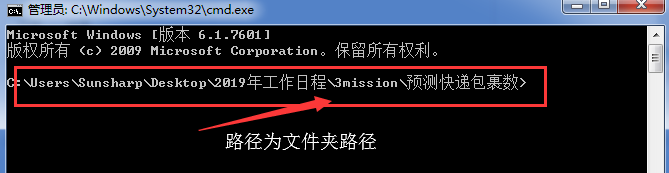
打开存放数据表package\_data.xlsx的文件夹，文件夹下必须有package\_data.xlsx文件。我的文件夹路径为C:\Users\Sunsharp\Desktop\2019年工作日程\3mission\预测快递包裹数



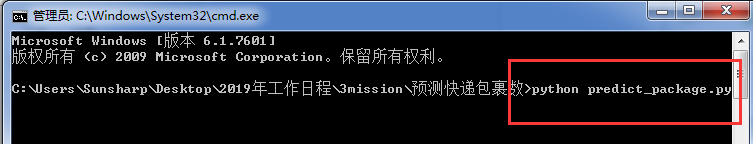
**step 2.** 清除文件夹路径并输入cmd，点击 回车 ，启动命令提示符



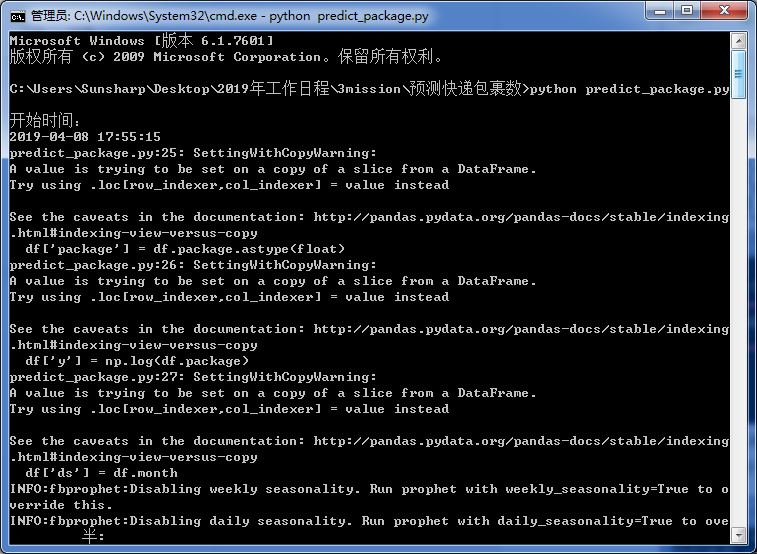




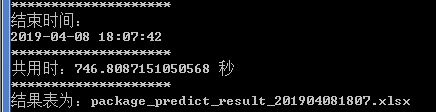
**step 3. 输入法在英文状态下，**在命令提示符(cmd)中输入 python predict\_package.py点击回车后程序开始运行。程序开始运行会显示开始时间，运行中会输出多行提示，可忽略；\*\*\*星号\*\*\*中的内容显示的是程序现在处理的地域id，可据此查看进度。







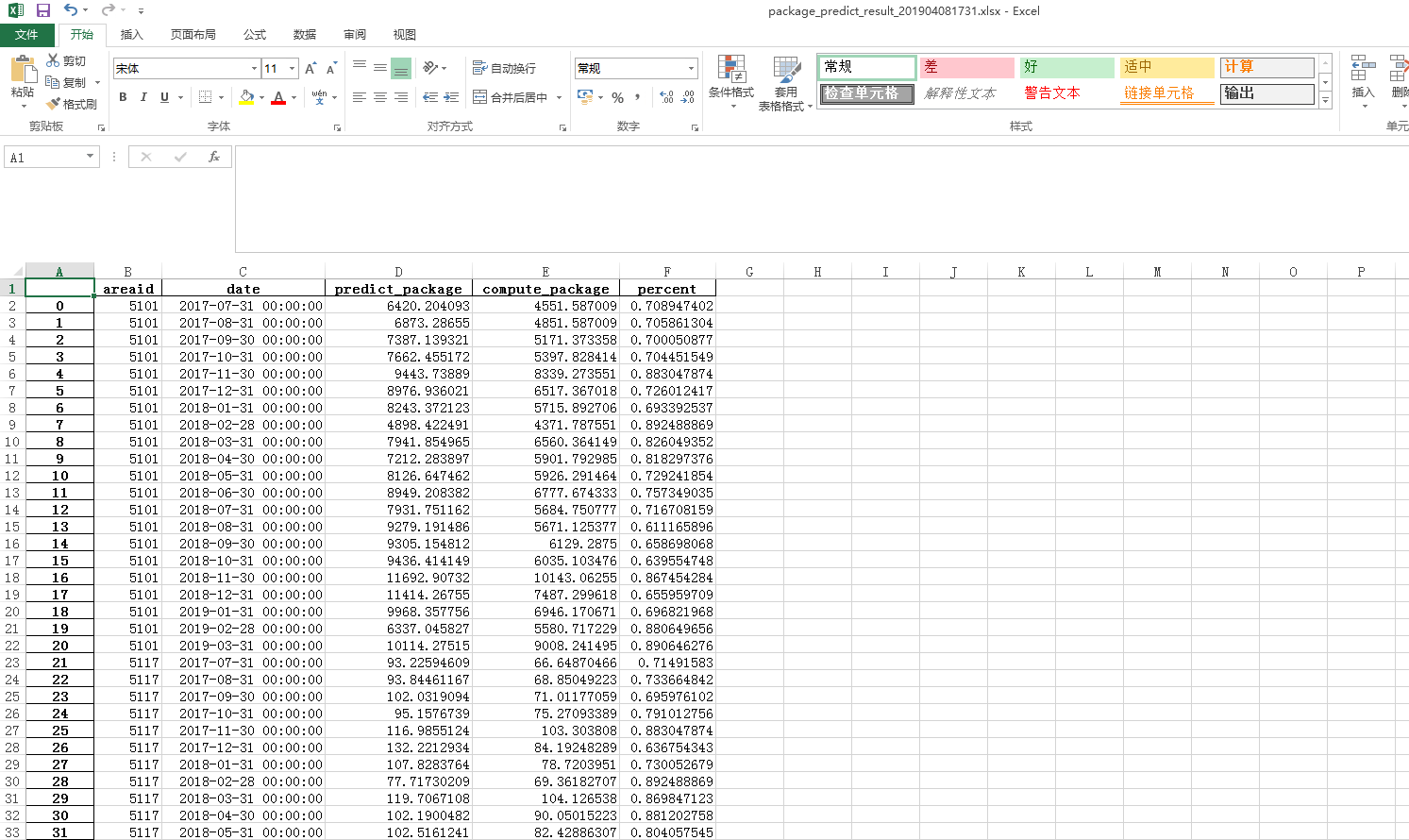
**step 4.**程序运行结束后会输出结束时间和总用时，预测结果会保存到同一文件夹下的excel文件中，如package\_predict\_result\_201904081731.xlsx，其后缀为程序运行结束时间（年月日时分）。

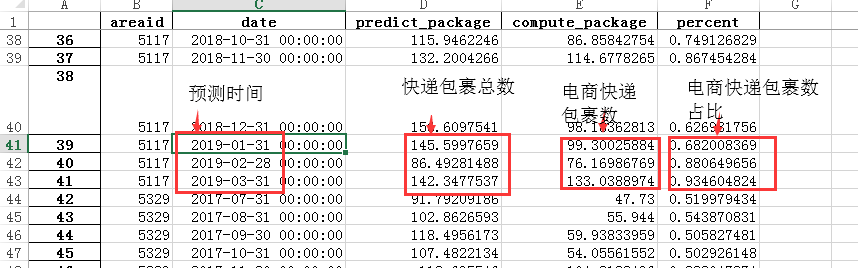




### 4.结果表介绍

结果表共有6列，第一列A列是默认索引，可以忽略；第二列B列字段areaid为地域id；第三列C列字段date为时间，从2017年7月31日开始，到预测时间（2019年3月31日）为止，2019年1月、2月、3月为此次的预测期；第四列D列字段predict\_package为预测的快递包裹总数（即预测的邮政管理局公布的数据）；第五列E列compute\_package为预测的电商快递包裹数，理论上说其值不能超过D列值，即不能超过预测的邮政管理局公布的数据；第六列F列percent为电商快递包裹数在包裹总数中的占比，可为参考用。

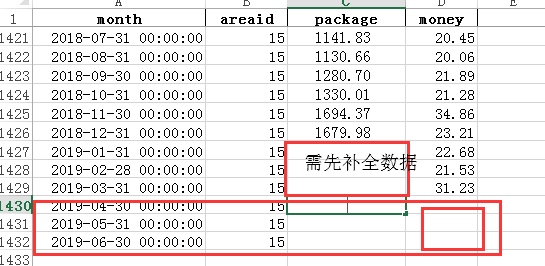




### 5.新增数据的方法

以上演示了预测2019年一季度电商快递包裹数的方法。如果下一期要预测二季度的电商快递包裹数，需要向excel数据表（package\_data.xlsx）里新增数据。新增数据建议参考以下方法：

（1）例如下图，将所有地域的新数据直接复制粘贴到数据表最后（并不需要按时间排序），并记得将上一期（此处指2019一季度）的空缺值补全。注意只有package列最后三个值可以空缺，其他位置不能空缺



### 6.手动校验与调整

得到预测结果后，需要根据客单价、同比等指标对结果进行验证，如果验证指标比较合理，则通过验证；如果不合理，需要手动调整快递包裹数数值。

示例文件为：合作地域快递包裹数-校对、调整.xlsx。