2021 年“泰迪杯”数据分析技能赛(A 题)

——通讯产品销售和盈利能力分析

分析报告

队伍名称：飞机炸弹

学校：中国地质大学（武汉）

队伍成员：叶宇涛，李亚骋，任帅

# 目录

[一.问题分析 3](#_Toc87729153)

[1. 背景介绍 3](#_Toc87729154)

[2. 分析目标 3](#_Toc87729155)

[二． 数据预处理 3](#_Toc87729156)

[1.数据检验 3](#_Toc87729157)

[三． 数据基本处理 5](#_Toc87729158)

[任务一.数据分析与预测 5](#_Toc87729159)

任务[1.1 5](#_Toc87729160)

[任务1.2 10](#_Toc87729161)

[任务1.3 13](#_Toc87729162)

[任务1.4 15](#_Toc87729163)

[任务二.可视化展示和撰写分析报告 19](#_Toc87729164)

[任务 2.1 19](#_Toc87729165)

[任务 2.2 28](#_Toc87729166)

[任务 2.3 30](#_Toc87729167)

[任务 2.4 31](#_Toc87729168)

[任务 2.5 32](#_Toc87729169)

[任务 2.6 33](#_Toc87729170)

[三、建议 34](#_Toc87729173)

# 一.问题分析

### 背景介绍

进入本世纪以来，我国通讯产品得到了飞速发展，其技术先进，价格便宜，

深受世界各国和地区尤其是非洲国家的欢迎。某通讯公司在非洲的多个国家深耕

多，产品与服务遍布整个非洲大陆。为了更好地了解公司的销售情况，采用产

品的销售额和利润数据，对其盈利能力进行分析和预测，给决策人员提供分析报

告，以便为非洲各国提供更好的产品销售策略和服务。

### 分析目标

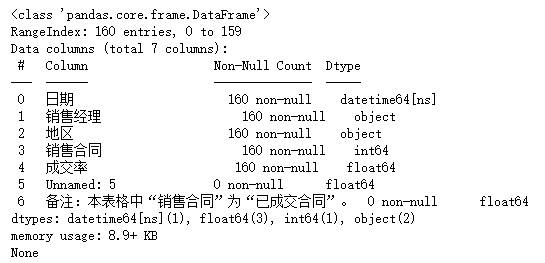
1. 统计产品在当地的销售数据，预测未来的销售情况。

2. 设计可视化数字大屏，展示产品的销售情况，分析产品的盈利能力。

# 数据预处理

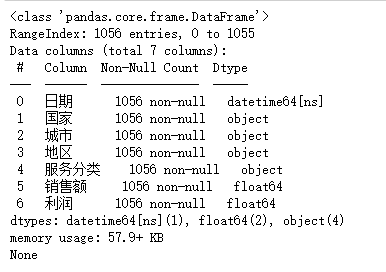
### 1.数据检验

数据预处理是数据分析中必不可少的关键一步，更是进行数据分析前的准备工作.通过pandas库我们对源数据中的两个表格进行了数据校验。

**SalespersonData:**

**SalespersonData**数据表信息表述

**SalesData:**



**SalesData**数据表信息表述

在本次数据分析的过程中，未存在缺失数据，异常数据以及重复数据，可以得出数据正确无需清洗。

# 数据基本处理

## 任务一.数据分析与预测

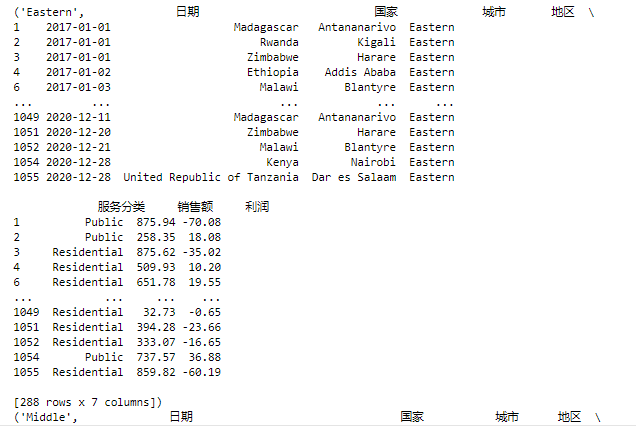
### 1.1

#### 1.1.1任务要求：

统计各个年度/季度中，地区、国家、服务分类的销售额和利润数据，并计算各国、各服务分类销售额和利润的同比增长率。

#### 1.1.2完成思路：

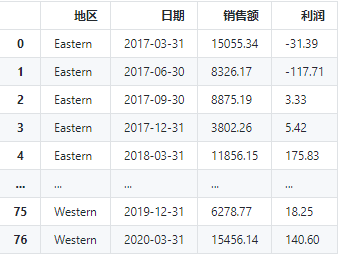
对于本道题，要求统计每个年度、季度中不同地区，国家以及服务分类的销售额和利润数据。因此，我们首先利用pandas中的groupby函数，依据国家，服务分类以及地区进行划分，结果如下:



按地区分类数据集

由于数据很多，只给出一部分数据进行展示。

因为需要对每个地区，国家以及服务分类按季度，年度划分后，进行销售额的累加求和，我们采用pandas中的resample函数对数据进行按季度、年度重采样，再进行agg操作，对销售额和利润进行加和，结果如下：



按地区分类季度数据集

下表中给出2020年度销售额前三名的国家：



对于同比增长率，给出下列定义:

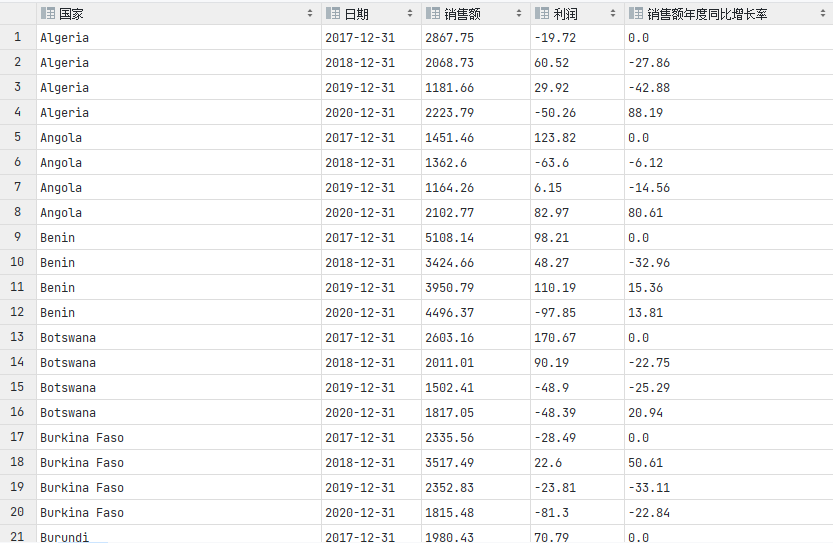
**同比增长率=（当年的指标值-去年同期的值）÷去年同期的值\*100%**

而对于销售额的同比增长率，经过检验，不存在销售额小于0的数据，因此，只要对年度以及季度的销售额，套用上述公式计算即可，具体实现过程如下：

**对于年度：**

**销售额：**

国家和服务分类的年度销售额增长率可以根据国家和服务分类的年度销售额数据集得到，我们在对数据groupby分组后，采用pandas的diff函数进行一阶差分计算。得出的一阶差分的值就是同比增长率的分子值。而对于年度同比增长率来说，2017年的同比增长率是和2016年相比较得出的，由于2016年的值不存在，因此，2017年的值为空，采用当前值减去一阶差分后的值，就是去年同期的值，即分母值。两者相除乘以100%即为当期同比增长率。给出图如下：



各个国家销售额年度同比增长率

**利润：**

国家和服务分类的年度利润增长率需要注意如下事项：

1.某些国家的去年同期利润可能为负值，而今年利润为正值，在计算增长率的时候，如果套用公式就会变成负数，反而不符合“扭亏为盈”的现状。因此，这种情况一般是无意义的，但是为了数据需要，我们决定采用取绝对值的方法，也就是将这种情况下的同比增长率取绝对值。

2.某些国家的去年同期利润可能为0,这种情况下，讨论增长率是没有意义的，因为没有相比较的数据，得出的值是INF。因此，我们将这种情况下同比增长率设置为空。

计算方法和上面计算销售额方法一样，加了判断条件后可以得出下图：



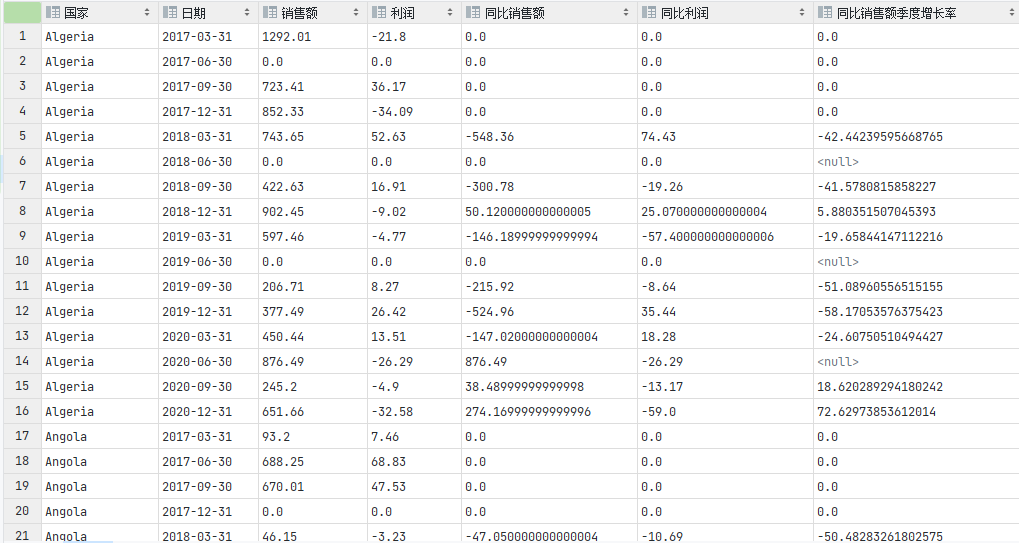
各个国家利润年度同比增长率

**季度：**

**销售额：**

国家的每个季度的销售额数据存在问题：缺少了2020年第四季度的值，导致无法通过遍历完成操作。因此，我们首先对数据集进行了处理，将每个缺少第四季度的国家进行数据填充，销售额和利润置为0.

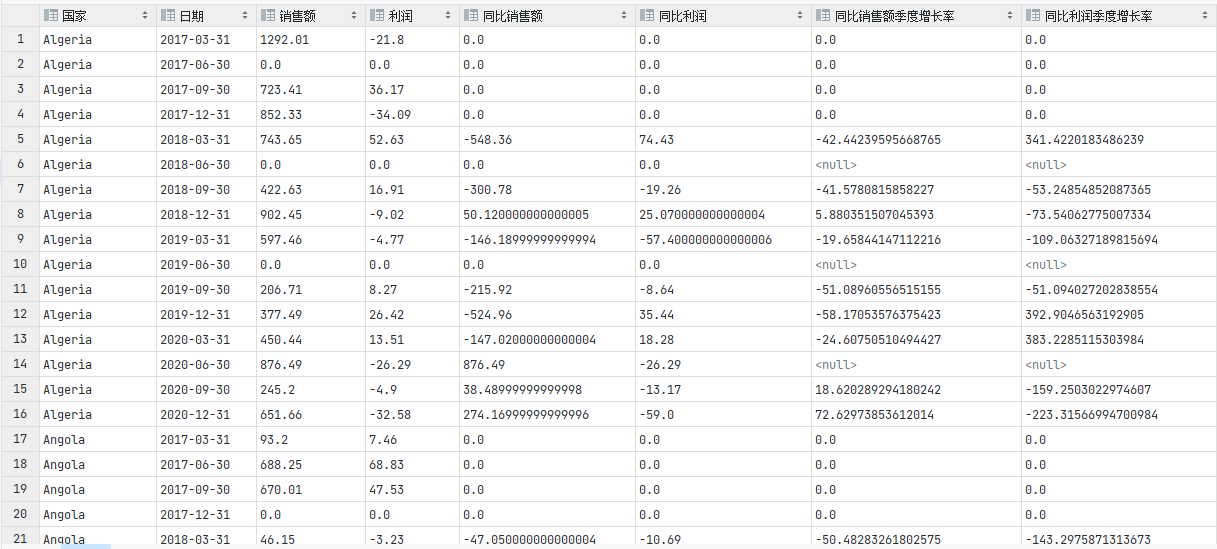
接着进行同比增长率的计算，采用pandas的diff函数进行四阶差分计算。得出的四阶差分的值就是同比增长率的分子值。而对于季度同比增长率来说，2017年四个季度的同比增长率是和2016年四个季度相比较得出的，由于2016年的值不存在，因此，2017年四个季度的同比增长率都为0，采用当前值减去四阶差分后的值，就是去年同期的值，即分母值。两者相除乘以100%即为当期同比增长率。给出图如下：



各个国家销售额季度同比增长率

**利润:**

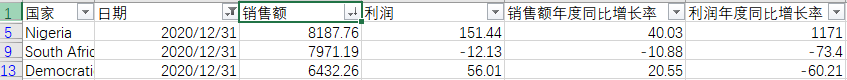
国家的季度增长率也同样存在年度增长率一样的问题，进行条件判断后，采取和季度销售额增长率一样的方法进行计算。



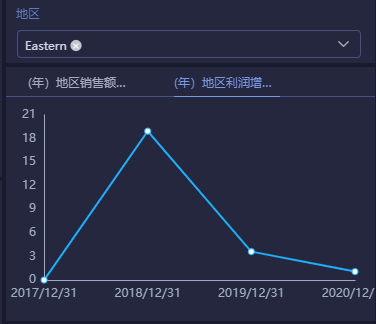
各个国家利润季度同比增长率

#### 1.1.4 任务结果

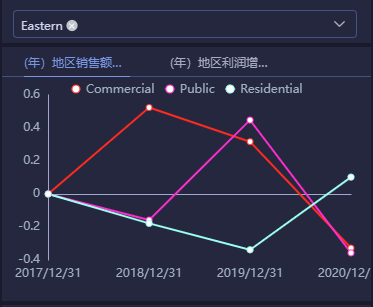
下表中给出2020年度销售额前三名的国家和年同比增长率：



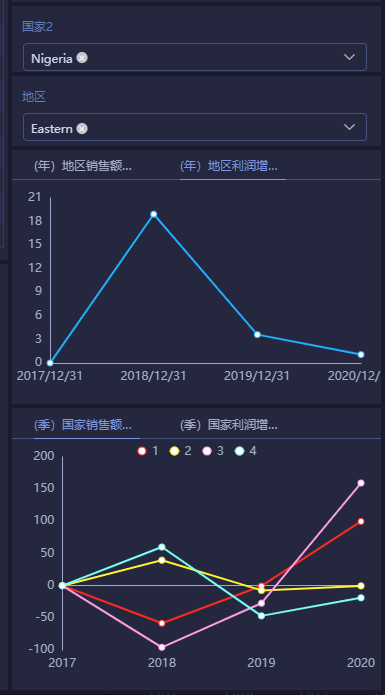
2020年度销售额前三名的国家和年同比增长率



东部地区年度利润增长率折线图



东部地区年销售额增长率折线图



Nigeria各个季度同比增长率

### 任务1.2

#### 1.2.1任务要求

统计各地区、国家有关服务分类销售额和利润数据。

#### 1.2.2完成思路

利用python通过pandas读取表，转化为dataFrame对象，由题意，需要统计个地区国家有关服务分类和销售额数据，只需求对所有地区相同，服务相同或者国家相同服务相同的记录的销售额和利润数据进行求和获得新的统计过的项，对每一种地区和每一种服务分类的组合或者每一种国家和每一种服务分类的组合进行相同的思路，再将结果保存到excel文件中即可。

#### 1.2.3操作步骤

先打开SalseData这一张数据表，使用groupby函数，利用”国家”和”服务分类”对数据进行分组，再利用sum()这一函数对”销售额”这一字段进行求和工作,再通过reset\_index()将得到的数据编程dateFrame格式最后保存到文件中，可以随时观看。对于”地区”和”服务分类”进行同样的操作即可。

任务结果及分析：



表1:Eastern地区和服务分类对应的销售额



表2:Eastern地区和服务分类对应的销售额



表3:Northern地区和服务分类对应的销售额



表4:Southern地区和服务分类对应的销售额



表5:Western地区和服务分类对应的销售额

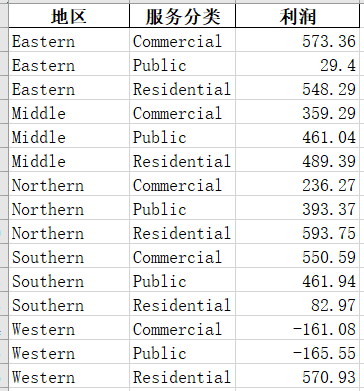


表6:各个地区地区和服务分类对应的利润

由图可知，各地区各服务分类均有销售额，其中Western地区和Eastern地区的各服务分类均有较高的销售额，说明这两个地区的通讯和经济比较发达，通讯产品的销量跟多，销售额更多，Northern地区的各服务分类销售额最少，说明此地经济和通讯最不发达。且根据各个区域来看public的服务分类相对来说较少，其他两项较多。

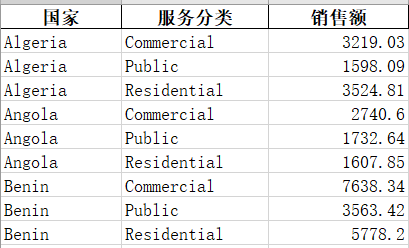


表7部分城市和服务分类对应的销售额

城市过多，仅列出部分城市用以参考。

### 任务1.3

#### 1.3.1 任务要求

统计各个销售经理的成交合同数和成交率。

#### 1.3.2 完成思路

利用python通过pandas读取张表，转化为dataFrame对象，据题意需要统计各个销售经理的成交合同数和成交率，其中合同数比较容易求得，只需将每个经理的各自的成交合同数相加即可得到每个经理总的成交合同数。成交率的计统计则较为复杂，不能单纯的对成交率做平均计算，而需要通过经理对应的每一项的的成交合同数和成交率进行计算，算的每一项的总合同数，最后相加获得每位经理对应的总的合同数，并将每位经理总的成交合同数除以总的合同数，即为每位经理对应的成交率。

#### 1.3.3操作步骤

成交合同数的统计，先打开SalespersonData这张表，通过groupby函数对

”销售经理”做划分，同时使用sum 函数对”销售经理”相同的记录对”销售合同”列做加法，最后得到了关于销售经理和成交合同数的dateFrame，最后将dateFrame保存成文件。

成交率的统计，先使用pandans打开SalespersonData这张表，通过groupby函数，对”销售经理”进行分组，得到每位经理和其对应的所有的成交率和成交合同数，并通过list强制转化为list型，最后通过循环，对每一位经理，对其每一项成交率和成交合同数，进行成交合同数/成交率的计算，最后将每一项加起来，即是每位经理对应的总的合同数，最后通过总成交合同数/总合同数，得到这位经理的成交率的统计，后续只需对每一位经理进行相同操作就行了。最后保存到excel文件中即可。

#### 1.3.4任务结果及分析

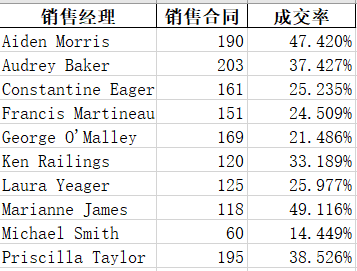


表1:销售经理销售合同成交率统计表

图中展示了每一位销售经理的销售合同数，和成交率，并不能非常直观显示每个人的销售能力，因此为了进一步的进行分析我们将对销售合同和成交率按降序进行排序，并分别得到前三名的的销售经理，方便分析销售经理的能力，为后续工作，如褒奖有能力的经理做准备。

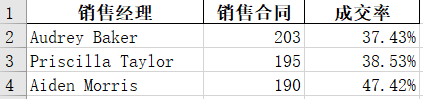


表2:销售经理销售合同前三名

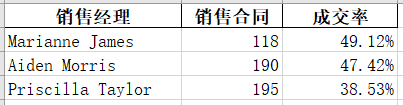


表3:销售经理成交率前三名

由图可知Aiden Morris和Priscilla Taylor 这两个人，是销售合同数和成交率的前三名，无疑这两个人的的销售能力非常好，Audrey Baker虽然销售合同很多，但成交率还是稍微逊色一点。

### 任务1.4

#### 1.4.1任务要求

分别预测各个地区、国家、服务分类 2021 年第一季度销售额和利润。

#### 1.4.2完成思路

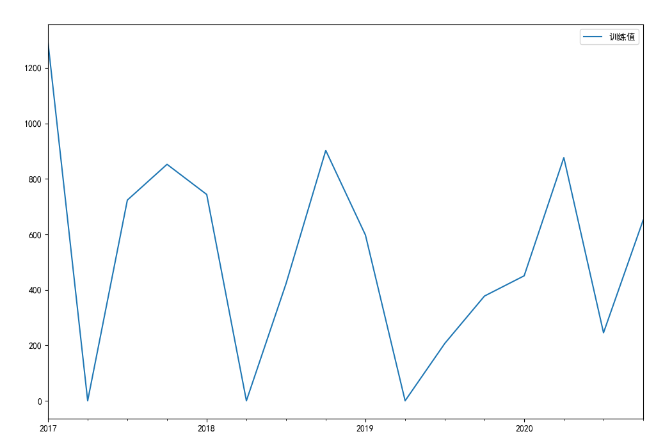
由于是存在时间序列预测，以及销售额随季节变化起伏比较大，我们决定采用ARIMA模型进行预测。

ARIMA模型，差分整合移动平均自回归模型，又称整合移动平均自回归模型（移动也可称作滑动），是时间序列预测分析方法之一。ARIMA(p，d，q)中，AR是“自回归”，p为自回归项数；MA为“滑动平均”，q为滑动平均项数，d为使之成为平稳序列所做的差分次数（阶数）。“差分”是关键步骤。

首先我们检查了数据是否存在突兀点，缺失点。进行了人为检查以及填充数据之后，我们队时间序列进行预处理：

我们选取第一个国家Algeria进行举例。

由于是随季度变化的时间序列，肯定是平稳非白噪声序列，分析得出如下图。



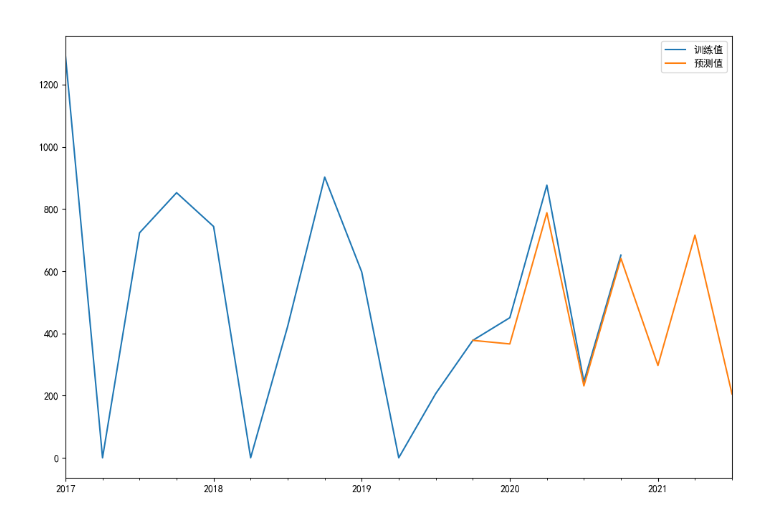
Algeria销售额随季度变化图

可以看出，销售额随季度的变化十分大，因此我们对其进行1阶差分，直到图像显示平稳为止。模型定阶方法采用AIC进行。

超参数设置如下：

model = pm.auto\_arima(dta, start\_p=1, start\_q=1,  
 max\_p=8, max\_q=8, m=2,  
 start\_P=0, seasonal=True,  
 max\_d=3, trace=True,  
 information\_criterion=**'aic'**,  
 error\_action=**'ignore'**,  
 suppress\_warnings=True,  
 stepwise=False)

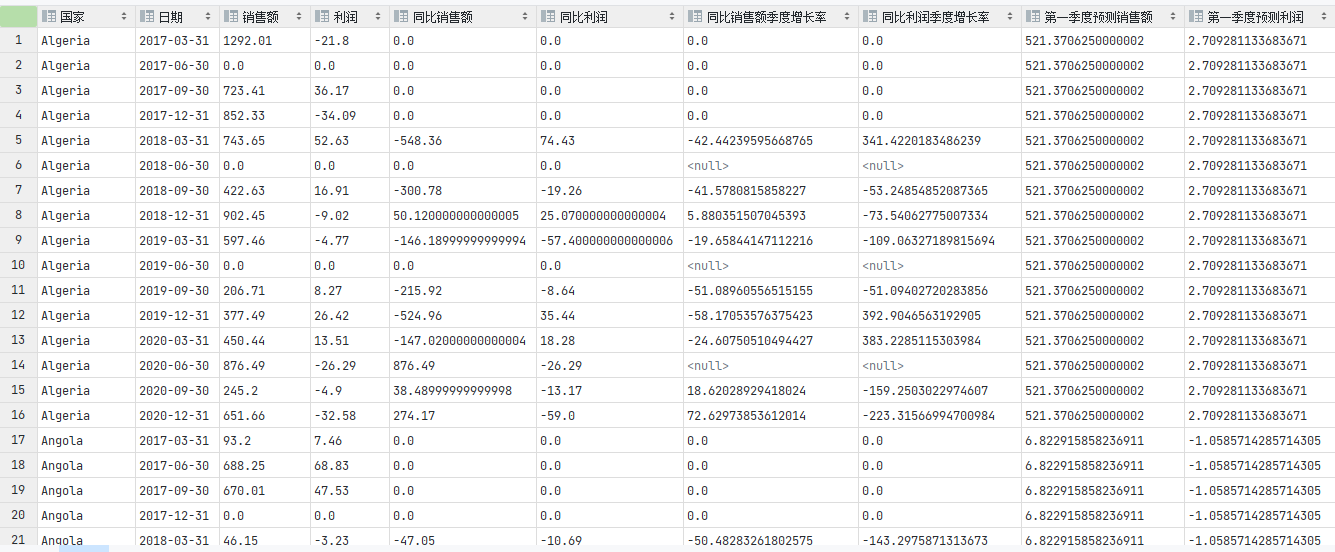
将数据前12个分为训练集，后4个为测试集，进行模型训练测试后得出下图：



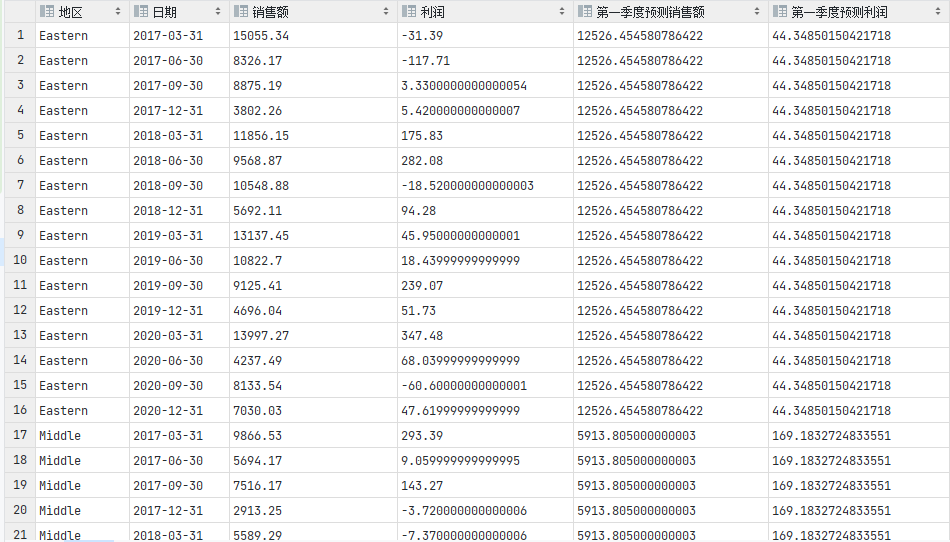
模型训练后效果对比图

可以看出该模型符合实际值，但是由于验证集过小，无法继续进行验证，于是我们直接进行了预测。

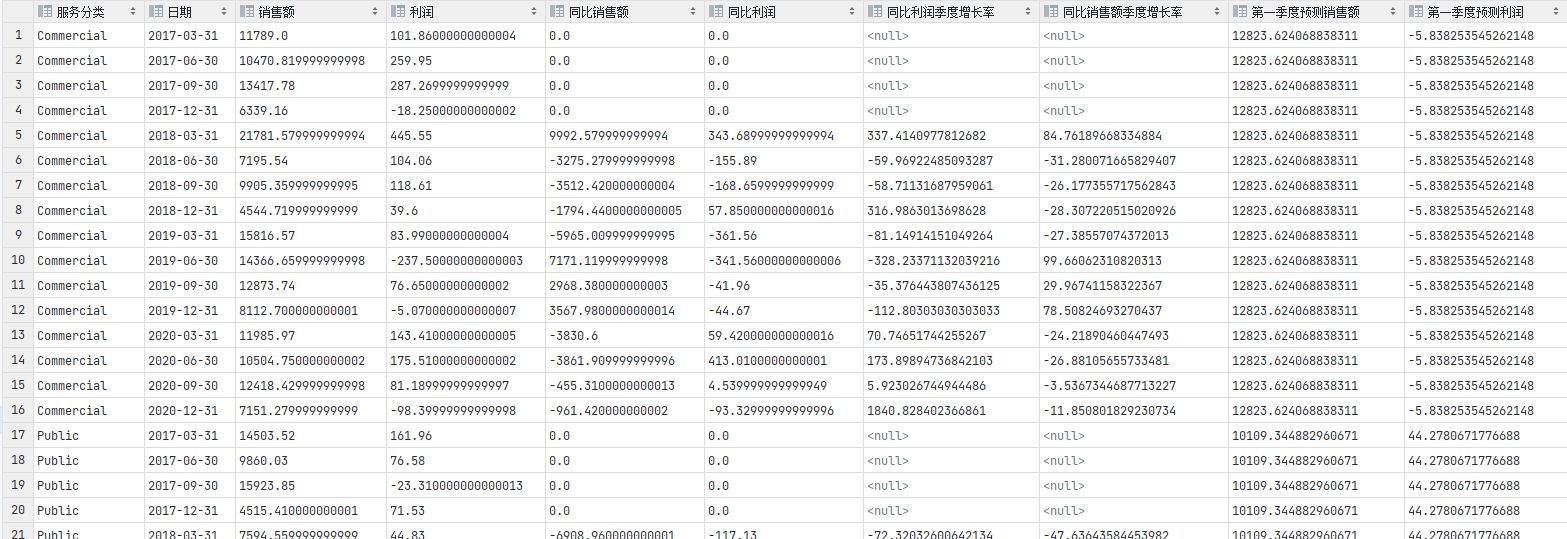
模型训练完成之后，我们进行各地区、国家、服务分类的预测，可以得出下图：



各国家分类的第一季度预测图



各地区分类的第一季度预测图



各服务分类的第一季度预测图

#### 1.4.3任务结果

下图给出地区、国家、服务分类预测值最大的一条数据：



各地区的第一季度最大销售额预测图



各国家的第一季度最大销售额预测图



各服务分类的第一季度最大销售额预测图

## 任务二.可视化展示和撰写分析报告

**任务目标：**设计一个数字大屏，根据目标，合理布局，展示能够代表产品销售情况和盈

利能力的数据指标和可视化图表等。

大赛提供了TipDM-BI数据分析和可视化平台（后简称BI平台）



自助仪表盘大图

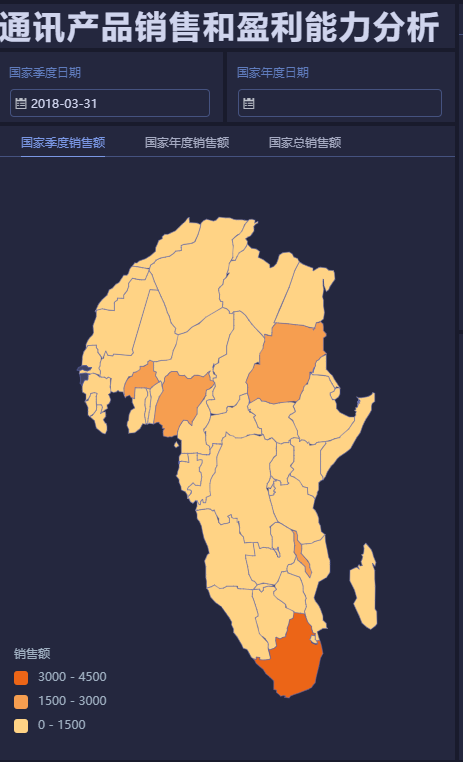
### 任务 2.1

#### 2.1.1任务要求

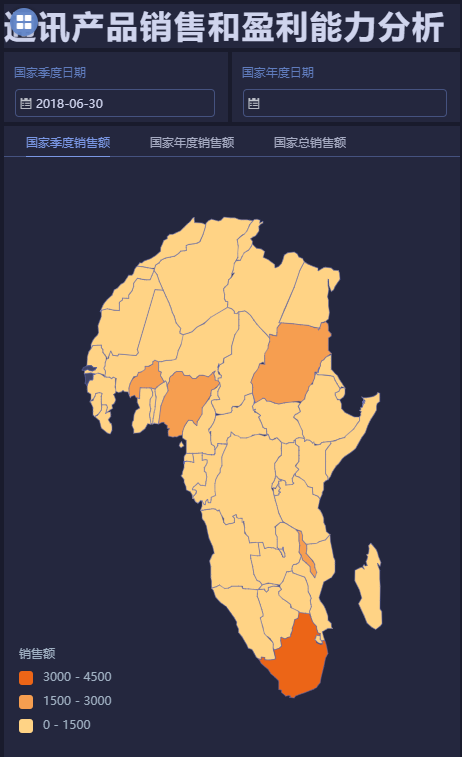
绘制非洲各国产品的销售地图，并能够查看该国的销售额和利润。根据销售额的降序排列，绘制非洲各国产品销售额和利润数据的图表。

#### 2.1.2实现结果

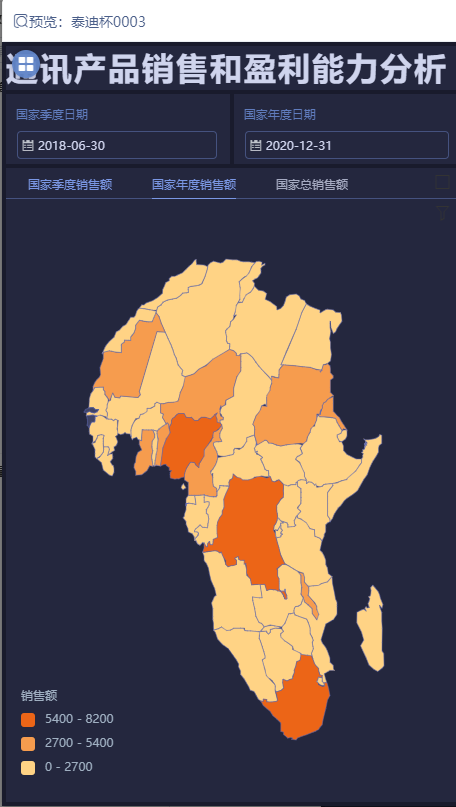
通过TipDM-BI数据分析和可视化平台，上传数据集，对数据集进行处理后，通过自助仪表盘可得出下图：



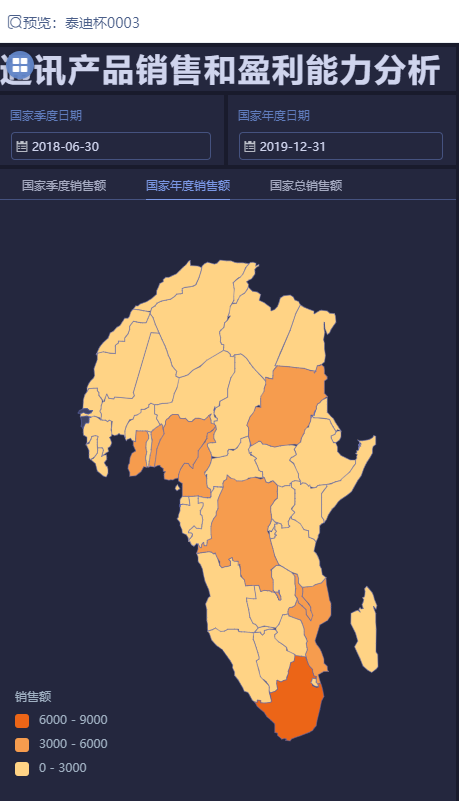
2018第一季度



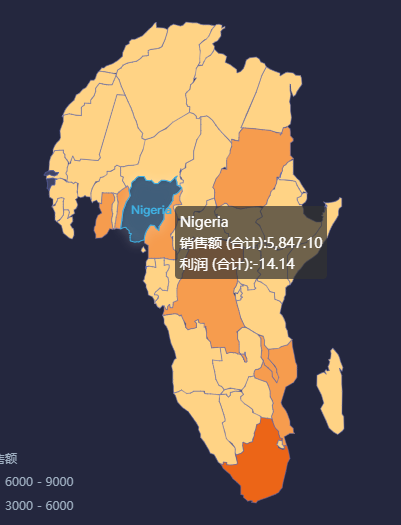
2018第四季度



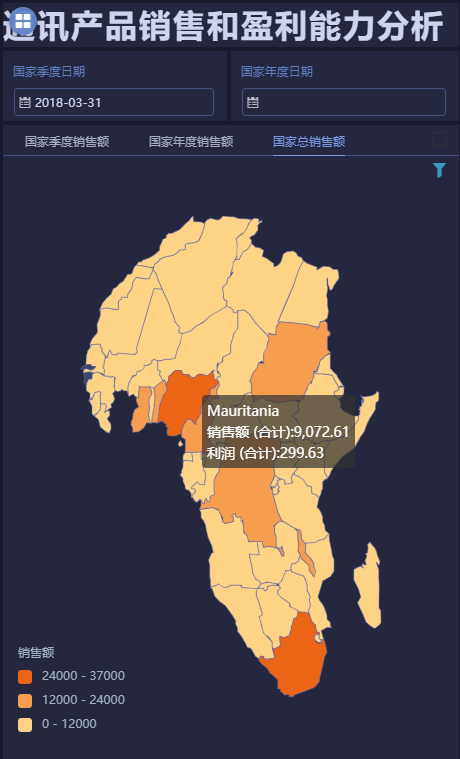
2020年年度销售额



2019年年度销售额

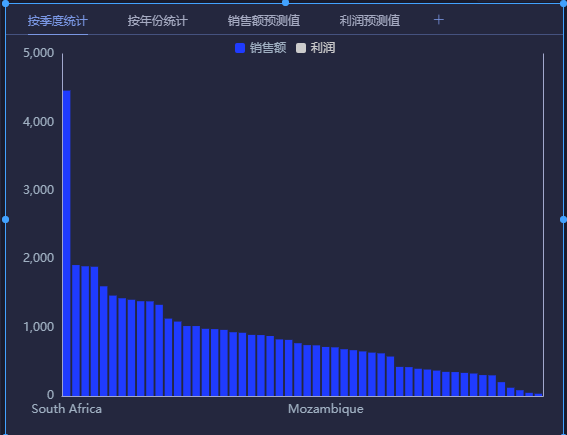


Nigeria 2019年度销售额以及利润合计

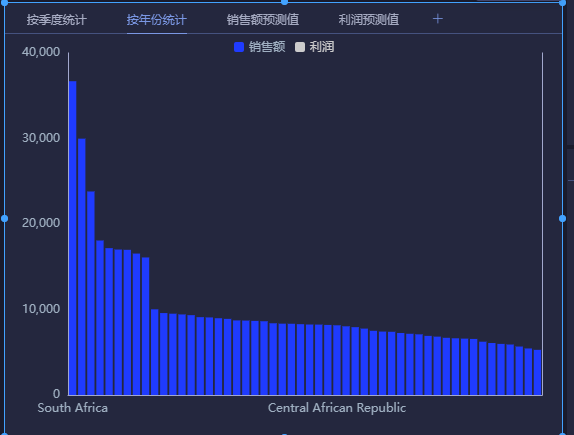


各个国家总销售额

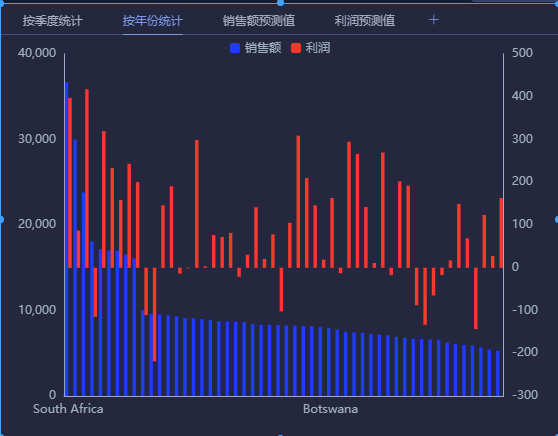
其中，可以查看各个国家的季度销售额，年度销售额，总销售额。都根据不同颜色进行了标记完成。



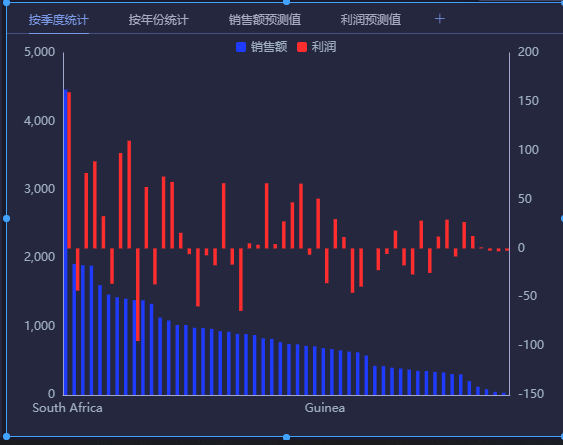
非洲各个国家季度销售额降序排列图表



非洲各个国家年度销售额降序排列图表



非洲各个国家年度总和图表



非洲各个国家季度总和图表

### 任务 2.2

#### 2.2.1任务要求

根据地区、国家等维度，绘制各服务分类的销售额和利润的年增长率及各季度同比增长率的图表。

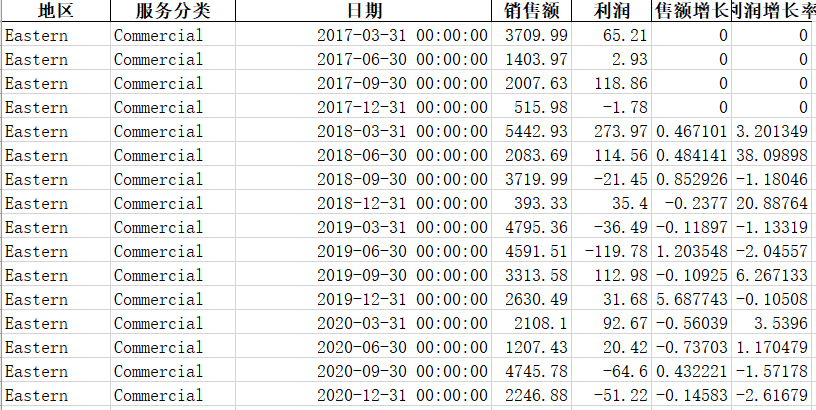
#### 2.2.2设计思路

由于同比增长率是与上一年同期对比，所以2017年的四个季度的增长率全部置零，然后通过pandas库先将源数据进行整合处理，得到2018及以后每个季度的地区/国家的各服务分类的销售额和利润的年增长率。

#### 2.2.3实现过程

(1)由题意，该题需要根据地区，国家等维度，绘制各服务分类的销售额和历任的年增长率及各种季度同比增长率的图标，所以需要生成以地区或者国家、服务分类、年度或者季度为维度，销售额，利润以及年度或季度增长率为度量的excel表。这里拿地区，各服务分类的季度同比增长率为例。

先用pandans读取数据，再同时用groupby对地区和服务分类进行分组，再用resample函数对日期进行季度的聚类，得到以地区，服务分类，季度为维度，销量和利润为度量的数据表，再对此数据进行遍历，由于每年的季度相差4个索引，在一个区域内时，同比季度增量为，当前销量减去四个索引前的销量除以四个索引前的销量。此时对于2017年的各个季度，以及四个索引前的值为0的情况，我们将此季度的同比季度增长率置为0，其他类似，最后将数据表导入到平台中。



地区服务分类销售额及利润增长率的部分截图

(2)在数据集中打开此表做进一步处理然后保存。

(3)在自主仪表盘中打开此数据集进行画图。

实际处理过程中季度销售额会存在0值的情况，我们选择用特殊值进行标记，表示该增长率异常。与此同时可视化通过下拉框选择地区或者国家以及显示的服务分类，随后绘出折线图，由此表现出增长率的变化。



季度国家、地区、服务分类销售额增长率图

### 任务 2.3

#### 2.3.1任务要求

根据地区、国家等维度，绘制 2021 年第一季度各服务分类的销售额和利润预测值的图表。

#### 2.3.2实现过程

由任务1.4得到的地区/国家的各服务分类预测值。在该任务中用油量表（走马灯形式）流动展示，并且可以通过点击2.1的地图选取国家（筛选器）来控制2.3的显示内容。



苏丹第一季度所有服务类别预测销售量总和



Nigeria第一季度预测销售量总和

### 任务 2.4

#### 2.4.1任务要求

绘制销售经理的销售合同数前 5 名排行榜。

2.4.2完成思路

大赛提供了 TipDM-BI 数据分析和可视化平台（简称 BI 平台），其中 有一整套完备的构建数字大屏方案，我们只需简单的几步即可搭建数字大屏框架。

其中的数据导入、数据连接、大屏组件设计部分等是我们根据需要进行自行组合 设置的。搭建数据大屏主要包含三步：

（1）数据连接-Excel 导入

（2）数据准备-自助数据集

（3）分析展示-自主仪表盘

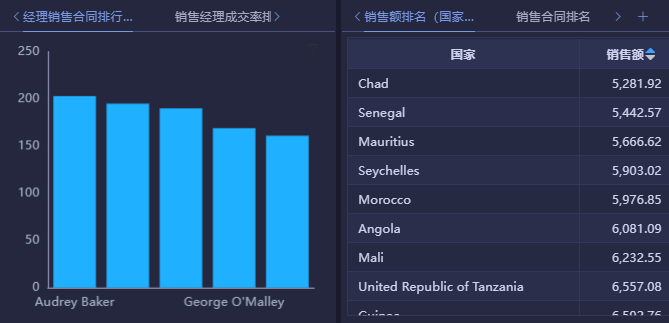
#### 2.4.2设计思路

(1)为了绘制销售经理销售合同的前五名，我们需要用到，销售经理和销售合同数的表格，该表格已由任务1.3给出并将该excel导入。

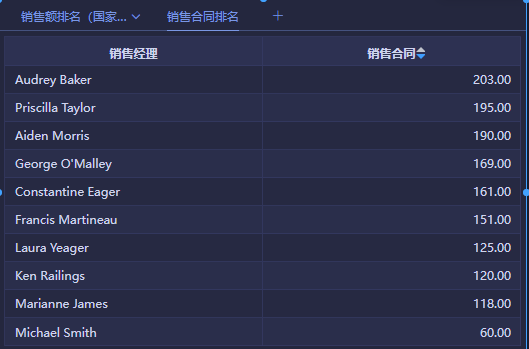
(2)将该表导入到自主数据集中作进一步的处理,并保存。

(3)将数据集中的”销售经理”作为维度，销售量作为度量，画出柱状图，并设置组件，

显示其最高的五项，于是得到销售经理销售合同数前五名，同时画出表格。



销售经理前五名柱状图



销售经理销售合同排名

### 任务 2.5

#### 2.5.1任务要求

绘制销售额后 10 名的国家排行榜。

#### 2.5.2设计思路

同2.4，可以查看升序或者降序排名。



销售额后 10 名的国家排行榜图

任务 2.6

#### **2.6.1****任务要求**

分析数字大屏的指标数据和图表，撰写公司产品的销售情况和盈利能力的分析报告。

#### **2.6.2分析**

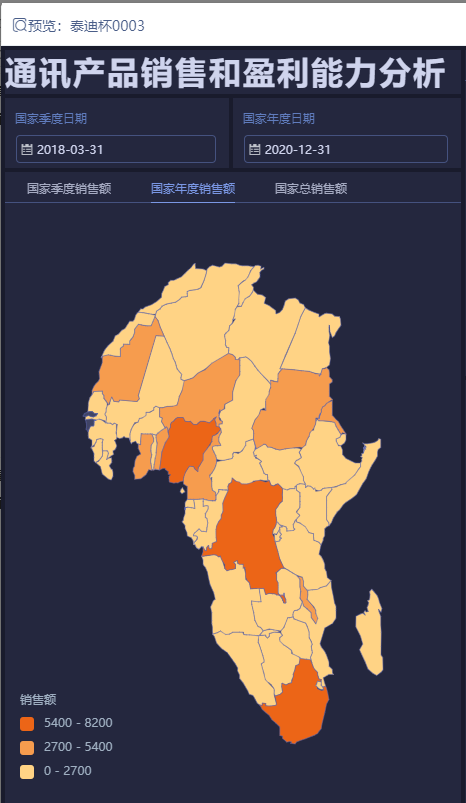
根据各地区和各国家在相应服务分类的销售额和来看Western地区和Eastern地区的各服务分类均有较高的销售额，说明这两个地区的通讯和经济比较发达，通讯产品的销量跟多，销售额更多，Northern地区的各服务分类销售额最少，说明此地经济和通讯最不发达。且根据各个区域来看public的服务分类相对来说较少，其他两项较多。所以不仅应增加Western和Eastern地区的商业投入，同时要不断拓宽Northern地区的市场.  
 与此同时，根据预测值，Commercial分类的利润将会为-5万元,而Public,Residential类别的收益将会达到近44万和150万，由此可以得出下个季度该公司应该加大public和Residential的投入力度，并把控好Commercial的成本支出，将亏损限定在一个可接受的范围。

在地区方面，预测值表明下个季度的Western可能会有大的亏损，达到了72万元，以及Northern也有2万的小型亏损，这两个地区需要改变市场策略来扭转亏损的局面，而其他三个地区都有比较好的预测收益，可以适量加大投入。

根据各地区和个国家在相应服务分类的利润来看，western地区虽然销售额多，但是盈利为负，所以要改变营销策略，在保住当地的销售额的同时，同时增加盈利。  
   除此之外为了增加公司的营销能力需要招聘更强的销售经理，对经理的成交合同数和成交率进行排名，激励好的经理，刺激营销能力差的经理，从而造成良性循环，激发经理潜质，增加盈利能力。

# 三、建议

**3.1对于该公司在各个地区的销售额分布图：**



各个地区的销售额分布图

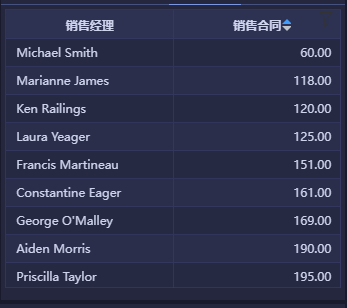
可以看出，分布比较不均匀，并没有完全遍布整个非洲的国家，因此，存在一些国家有隐形销售额的情况。公司可以根据不同的国家提供不同的服务，这样才能更好提高销售额以及利润。

**3.2 对于该企业的利润分布:**

存在不合理的地方，比如利润为负的地区，应该提高销售额或者干脆减少该服务类别的输出，可能是由于非洲贫困分布不均匀，对于比较贫困的国家不应该提供过多的产品，因为这样提供的产品并没有足够的人民来支付，导致利润为负数。

**3.3对于该企业的人员部门：**

存在不合理的地方，比如销售经理的销售合同量倒序查看如下：



销售经理的销量合同倒序查看图

可以看出一些销售经理存在业务能力不合格的情况，可以适当设置一定的门槛，进行淘汰制度，将销售额不合格的经理进行警告。