



华北理工大学

《数据可视化》课程报告

中国旅游业发展状况探究及前景预测报告

姓 名 _____

学 号 _____

2024 年 6 月 25 日

中国旅游业发展状况探究及前景预测

【摘要】

自疫情过后全球经济逐渐复苏，民众旅游消费热情高涨，我国旅游业作为重要支柱产业被寄予厚望。据此，本报告深入探究了中国旅游业的发展状况并对未来前景进行合理预测。

首先，通过运用可视化技术分别对中国旅游业总体发展情况、国内旅游及入境旅游情况进行了全面地描述性统计分析，展示了中国旅游业自 2004 年以来的**稳步上升趋势**。在国内旅游中，城乡居民仍存在较高的消费水平差异但有缩减趋势；在入境旅游中，亚洲、欧洲国家为主要入境旅游消费群体，长途交通在外汇收入中占比最高。

而后，报告中建立中国旅游业发展预测模型——利用 **K-means 聚类分析**，将历年旅游业发展划分为缓慢增长期、平稳过渡期和快速上升期；而后综合三个阶段使用 **Arima 模型**对未来八年的国内旅游花费和国际旅游收入进行了预测，为了进一步提高预测的准确性，报告继续选用 **LSTM 深度学习模型**进行时间序列预测，结果均显示未来一段时期中国旅游业将继续呈现稳步上升态势。

关键词：中国旅游业；K-means 聚类分析；Arima 模型；LSTM 深度学习模型

目录

一、绪论.....	5
1. 研究背景.....	5
2. 研究意义.....	5
二、变量选取与数据来源.....	6
1. 变量选取.....	6
2. 数据来源.....	6
三、描述性分析.....	8
◆总体情况介绍.....	8
◆国内旅游情况.....	9
◆入境旅游情况.....	11
四、模型建立与分析.....	15
1. K-means 聚类分析.....	15
2. Arima 模型预测.....	16
3. LSTM 预测.....	18
六、结论与建议.....	20
1. 结论.....	20
2. 建议.....	20
六、参考文献.....	21

图表目录

图表 1 我国旅行社与星级饭店总数堆叠面积图	8
图表 2 国内游客数量变化折线图	9
图表 3 城乡游客数量及旅游人均花费小提琴图	10
图表 4 国内旅游人均花费堆积图	10
图表 5 各国入境游客数量热力地图	11
图表 6 各国入境游客数量环形内向柱状图.....	11
图表 7 各国入境游客数量棒棒图	12
图表 8 国际旅游收入及构成南丁格尔玫瑰图	12
图表 9 长途交通收入小提琴图	13
图表 10 各项国际旅游收入热力系数图	14
图表 11 星型词云图	14
图表 12 发展阶段 3D 聚类图	15
图表 13 ARIMA 预测过程.....	16
图表 14 一、二阶差分图	17
图表 15 ARIMA 预测结果.....	17
图表 16 LSTM 原理图.....	18
图表 17 国际旅游收入 LOSS 迭代图	19
图表 18 国内旅游花费 LOSS 迭代图	19
图表 19 LSTM 预测结果.....	19
表格 1 变量选取	6
表格 2 国内旅游情况	7
表格 3 国际旅游收入及构成	8
表格 4 不同发展阶段	16

一、绪论

1. 研究背景

近年来，随着中国经济的高速增长和人民生活水平的不断提升，旅游业作为国民经济的重要组成部分，得到了前所未有的发展机遇。



在国务院印发的《“十四五”旅游业发展规划》中提到，着力推动旅游业高质量发展，积极推进旅游业进一步融入国家战略体系。“十四五”时期，我国将全面进入大众旅游时代，旅游业发展仍处于重要战略机遇期，但机遇和挑战都有新的发展变化。

2. 研究意义

本研究有助于全面了解中国旅游业的发展现状，揭示其存在的问题和面临的挑战，为旅游业的未来发展提供科学的决策依据。同时，研究成果还可以为旅游企业制定市场策略、优化产品结构、提升服务质量提供指导，促进旅游业的健康发展。

理论方面，本研究可以丰富和发展旅游学的理论体系，推动旅游学科的发展。通过对中国旅游业发展状况的探究和前景预测，可以揭示旅游业发展的规律和趋势，为旅游学科的研究提供新的视角和方法。

社会方面，旅游业的发展对于促进就业、增加收入、推动区域经济发展等方面具有积极作用。本研究成果可以为政府制定旅游业发展政策提供参考，促进旅游业与其他产业的融合发展，推动社会经济的全面发展。

二、变量选取与数据来源

1. 变量选取

基于报告所用数据及模型要求，设定以下变量：

变量	含义
X	国内旅游总花费(亿元)
Y	国际旅游收入(亿美元)
Z	旅行社数(个)
P	长途交通国际旅游收入(亿美元)
p_1	民航国际旅游收入(亿美元)
p_2	铁路国际旅游收入(亿美元)
p_3	汽车国际旅游收入(亿美元)
p_4	轮船国际旅游收入(亿美元)

表格 1 变量选取

2. 数据来源

本报告所选用数据皆摘取自“国家统计局国家数据官网-年度数据”。

根据报告研究内容分为四个部分：旅游业发展情况、国内旅游情况、国际旅游收入及构成、各国入境游客数量。

其中，我们主要针对国内旅游情况、国际旅游收入及构成两方面进行建模分析研究。

展示数据如下：

●国内旅游情况（近 20 年）

指标	国内游客(百万人次)	城镇居民国内游客(百万人次)	农村居民国内游客(百万人次)	国内旅游总花费(亿元)	城镇居民国内旅游总花费(亿元)	农村居民国内旅游总花费(亿元)	国内旅游人均花费(元)	城镇居民国内旅游人均花费(元)	农村居民国内旅游人均花费(元)
2023 年	4891	3758	1133	49133.1	41780.5	7352.6	1004.6	1111.8	649

2022 年	2530	1928	601	20444	16881.3	3562.7	808.1	875.6	592.8
2021 年	3246	2342	904	29190.7	23644.2	5546.6	899.3	1009.6	613.6
2020 年	2879	2065	814	22286.3	17966.5	4319.8	774.1	870.3	530.5
2019 年	6006	4471	1535	57250.9	47509	9741.9	953.3	1062.6	634.7
2018 年	5539	4119	1420	51278.3	42590	8688.3	925.8	1034	611.9
2017 年	5001	3677	1324	45660.8	37673	7987.7	913	1024.6	603.3
2016 年	4435	3195	1240	39389.8	32241.9	7147.9	888.2	1009.1	576.4
2015 年	3990	2802	1188	34195.1	27610.9	6584.2	857	985.5	554.2
2014 年	3611	2483	1128	30311.9	24219.8	6092.1	839.7	975.4	540.2
2013 年	3262	2186	1076	26276.1	20692.6	5583.5	805.5	946.6	518.9
2012 年	2957	1933	1024	22706.2	17678	5028.2	767.9	914.5	491
2011 年	2641	1687	954	19305.4	14808.6	4496.8	731	877.8	471.4
2010 年	2103	1065	1038	12579.8	9403.8	3176	598.2	883	306
2009 年	1902	903	999	10183.7	7233.8	2949.9	535.4	801.1	295.3
2008 年	1712	703	1009	8749.3	5971.7	2777.6	511	849.4	275.3
2007 年	1610	612	998	7770.6	5550.4	2220.2	482.6	906.9	222.5
2006 年	1394	576	818	6229.7	4414.7	1815	446.9	766.4	221.9
2005 年	1212	496	716	5285.9	3656.1	1629.7	436.1	737.1	227.6
2004 年	1102	459	643	4710.7	3359	1351.7	427.5	731.8	210.2

表格 2 国内旅游情况

●国际旅游收入及构成（近 5 年）

指标	2019 年	2018 年	2017 年	2016 年	2015 年
国际旅游收入(亿美元)	1312.54	1271.03	1234.17	1200	1136.5
长途交通国际旅游收入(亿美元)	401.91	366.31	449.46	446.5	448.5
民航国际旅游收入(亿美元)	369.02	333.53	304.87	290.6	294.8
铁路国际旅游收入(亿美元)	14.1	13.52	49.52	53.2	43.2
汽车国际旅游收入(亿美元)	15.93	13.72	29.43	31.6	32.5
轮船国际旅游收入(亿美元)	2.85	5.54	65.65	71	78
游览国际旅游外汇收入(亿美元)	58.66	53.71	65.04	67.1	44.8
住宿国际旅游外汇收入(亿美元)	200.49	181.09	122.08	116.3	132.9
餐饮国际旅游外汇收入(亿美元)	160.41	142.55	103.07	96.2	82.6
商品销售国际旅游外汇收入(亿美元)	302.97	327.61	229.95	209.5	209

娱乐国际旅游外汇收入(亿美元)	44.21	45.82	74.16	77.1	53.9
邮电通讯国际旅游外汇收入(亿美元)	7.47	11.62	27.57	28.9	23.9
市内交通国际旅游外汇收入(亿美元)	34.53	27.76	39.2	40.4	22.4
其他服务国际旅游外汇收入(亿美元)	101.89	114.54	123.64	118	118.6

表格 3 国际旅游收入及构成

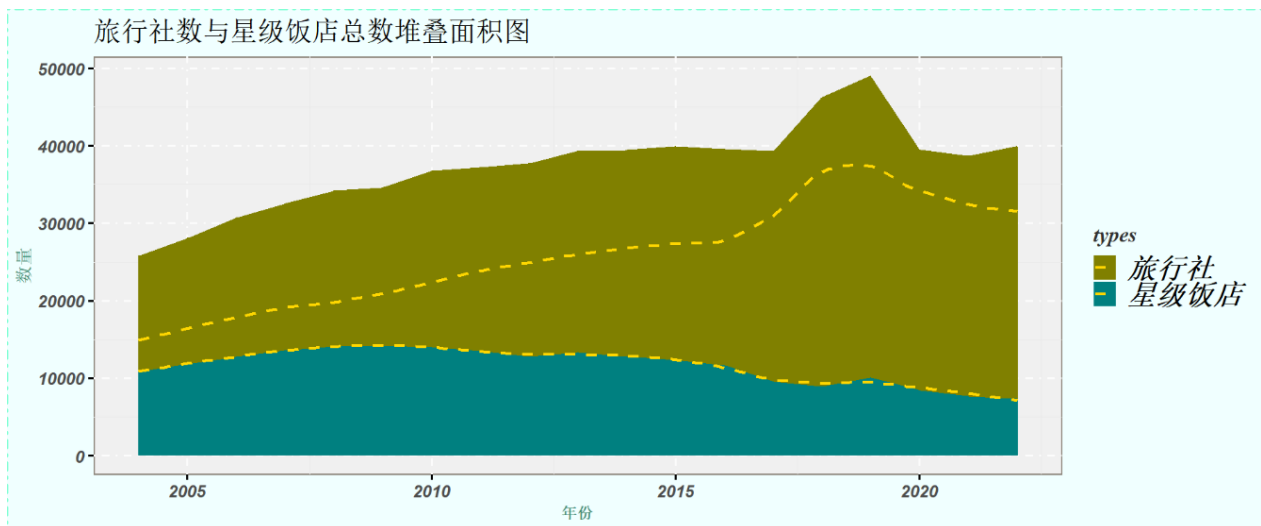
以上则为本报告所用的两部分重要数据。

三、描述性分析

本报告的描述性统计可以分为三大类，首先总体介绍中国旅游业的发展情况；而后，据游客主体，可以将中国旅游业发展状况分为国内外两大类。

◆总体情况介绍

对于一个国家来说，旅行社与星级饭店的数量变化趋势可以直接反映其旅游业发展状况。为更直观、详细地表述，采用堆叠面积图来进行可视化展示，如下为我国旅行社与星级饭店总数堆叠面积图。



图表 1 我国旅行社与星级饭店总数堆叠面积图

不难看出，自 2004 年我国旅游社同星级饭店数量均呈平稳上升趋势，其中，2009 年将我国星级饭店行业划分成两大阶段：快速发展阶段和优胜劣汰阶段，由于市场竞争激烈等原因，星级饭店持续减少。

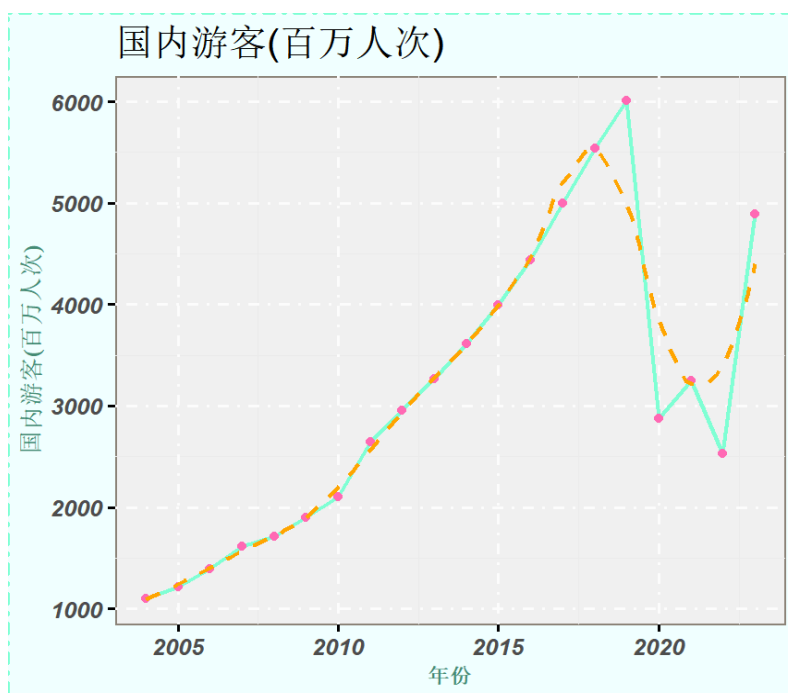
但是值得注意的是，尽管 2019 年末新冠疫情爆发，扑灭了刚刚燃起的高峰期，但

二者总数总体上仍保持上升趋势，尤其是旅行社的发展。这表明我国旅游业在经历疫情过后仍有较强活力。

◆国内旅游情况

国内旅游方面，我们先将国内游客当作研究对象，从国内游客总数量上直观展示其发展趋势，之后进一步剖析：将国内游客群体划分为城镇居民游客与农村居民游客，将二者的游客数量、人均花费进行比较并可视化展示。

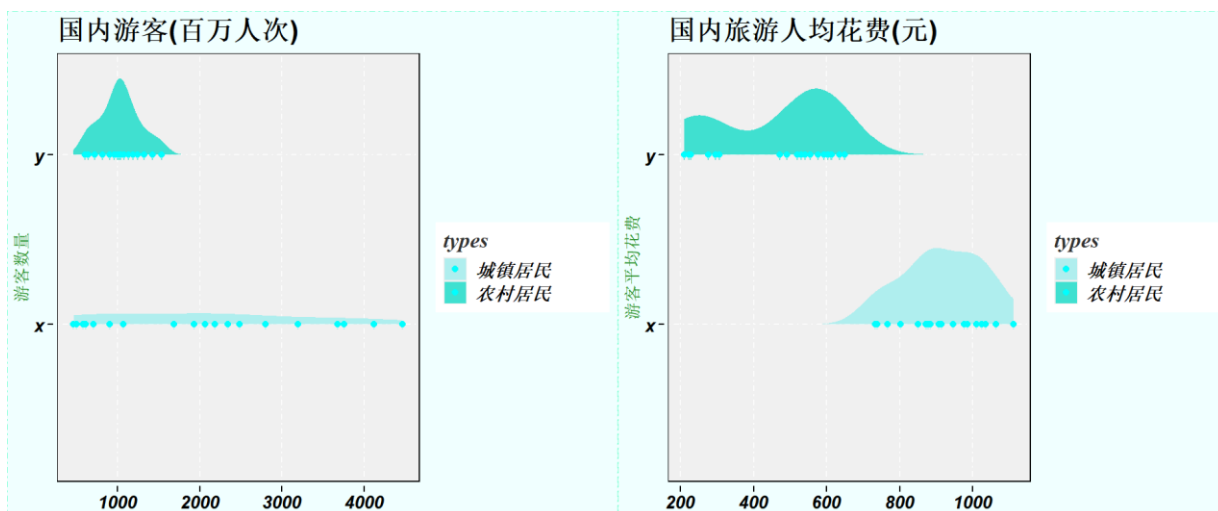
首先观察国内游客数量自 2004 至 2023 年间发展变化趋势：



图表 2 国内游客数量变化折线图

自 2004 年以来，我国国内游客数量持续保持快速上升。尽管在 2019 年末的新冠疫情中受到了显著影响，但近两年迅速恢复并呈现出强劲增长势头。

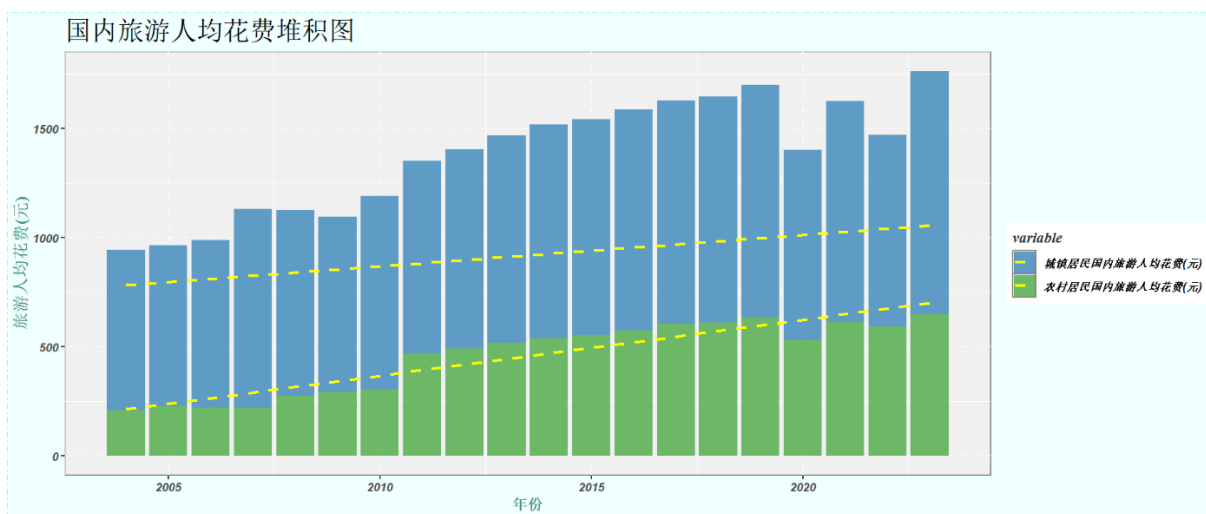
接下来，我们将研究对象细化为城镇居民与农村居民，首先分别对游客数量与人均花费进行分析，我们截取近 20 年的相关数据进行核密度分析：



图表 3 城乡游客数量及旅游人均花费小提琴图

在图中，我们可以清楚地看到农村居民的游客数量及旅游人均花费普遍低于城镇居民，这可能与城乡居民人均收入差距有密切联系。此外，在游客数量方面，农村居民的分布较为稳定，而城镇居民有显著差异，这很可能与人口城镇化有关；在旅游人均花费方面，农村居民的“云”似乎有向右飘的趋势。

为进一步探究，我们绘制了国内旅游人均花费堆积图：



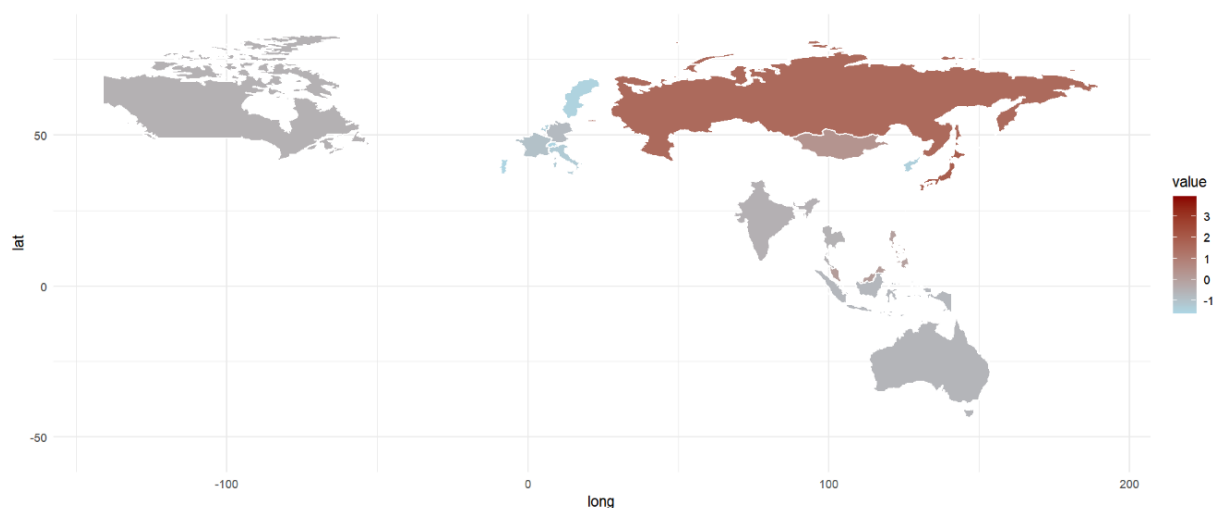
图表 4 国内旅游人均花费堆积图

这幅图回答了我们前文的疑问：在 2004 至 2023 年间，不仅城镇居民游客旅游人均花费稳步增长，农村居民同样在增长，且斜率更大，增长趋势更为可观。在这幅图中，我们还可以看出：国内旅游人均花费总体呈稳步上升趋势。

综合图 x 与图 y，我们清楚地了解到国内旅游游客数量、人均花费均呈稳步上升趋势。这代表着我国国内旅游方面发展态势良好。

◆入境旅游情况

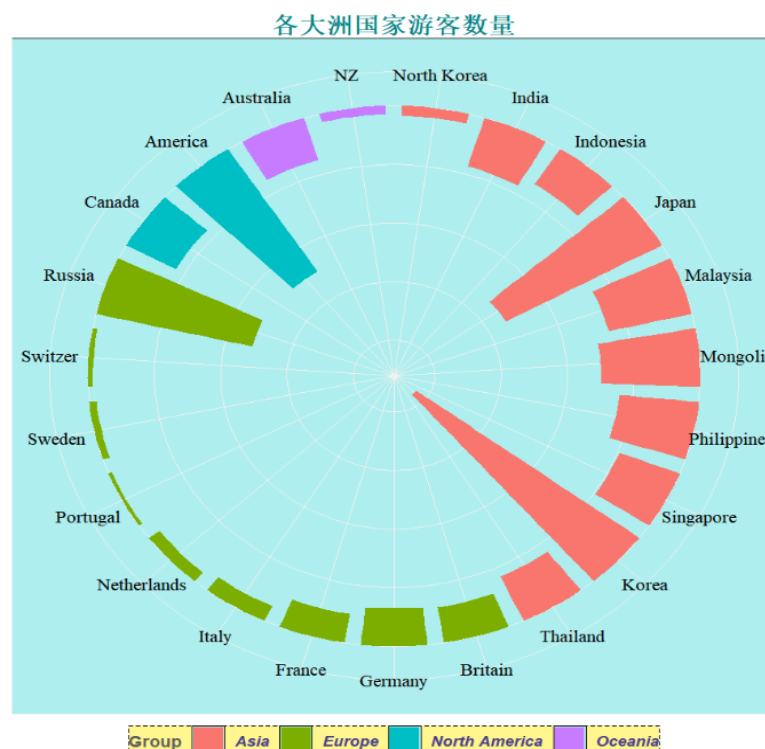
入境旅游方面，本报告首先对各国入境游客数量进行了详细统计，直观地分析了各国入境旅游的热度差异及主要入境旅游国家。



图表 5 各国入境游客数量热力地图

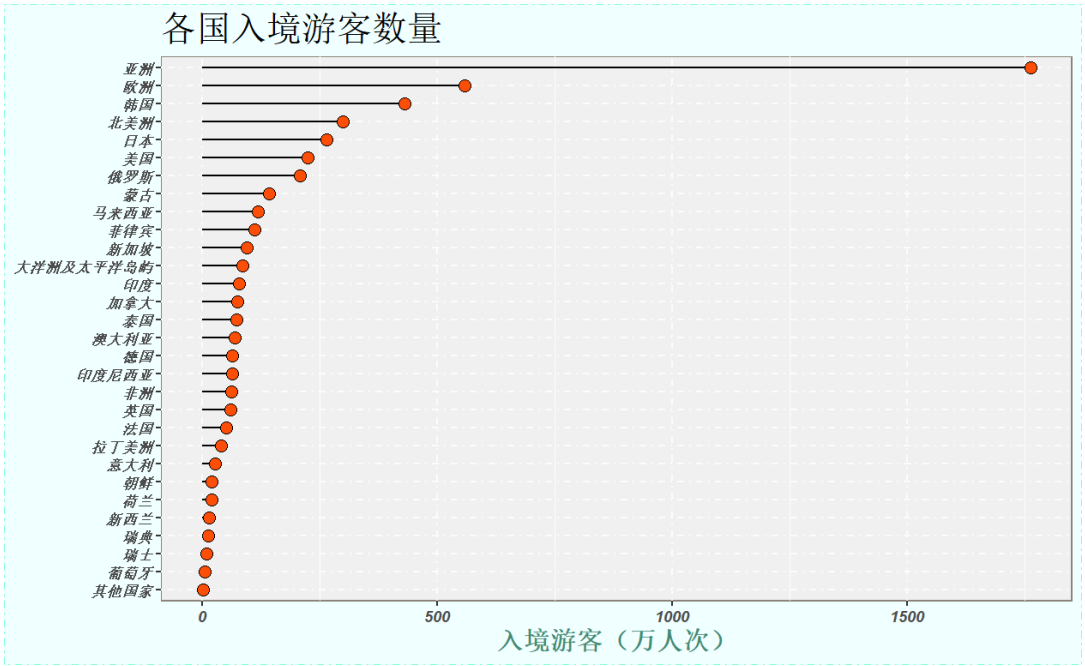
从热力图中我们可以直观地看出：入境游客数量中，亚洲国家所占热度最高，而大洋洲、北美洲等地热度则明显较低。

为了更详细地观察各国具体游客数量，本报告对其做了具体展示，首先我们根据大洲对各国家入境游客数量做出细化：



图表 6 各国入境游客数量环形内向柱状图

之后，我们采取更直观地方式对数值进行对比：

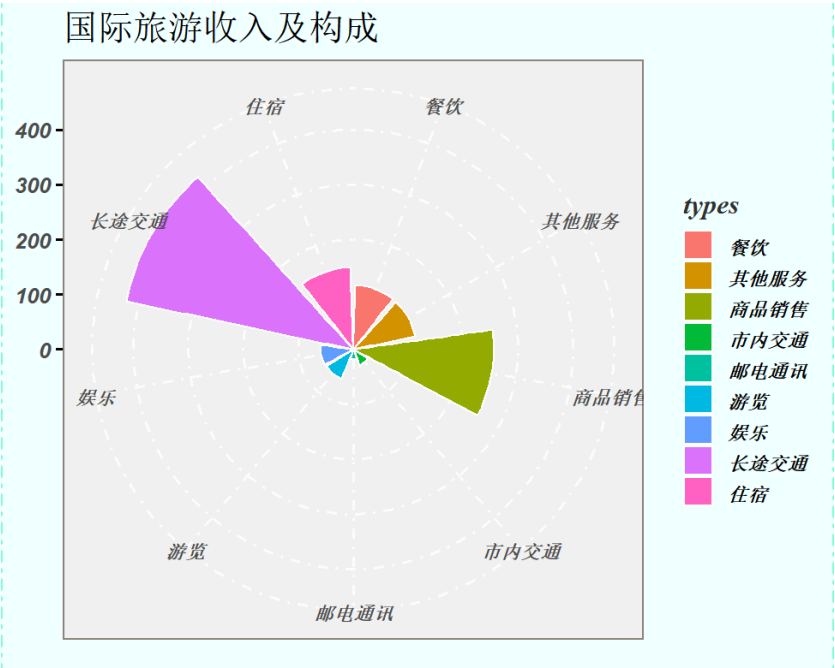


图表 7 各国入境游客数量棒棒图

该棒棒图并未区分各大洲与国家，旨在凸显国家及大洲的数量差异。

由该棒棒图可以直观地看出：在各国入境游客数量中，亚洲确实以断崖式水平位居榜首，其中不乏有距离近、旅行方便等客观因素；欧洲位居第二；而第三位的韩国竟高于北美洲国家，这在一定程度上反映了文化吸引力与文化认同感。

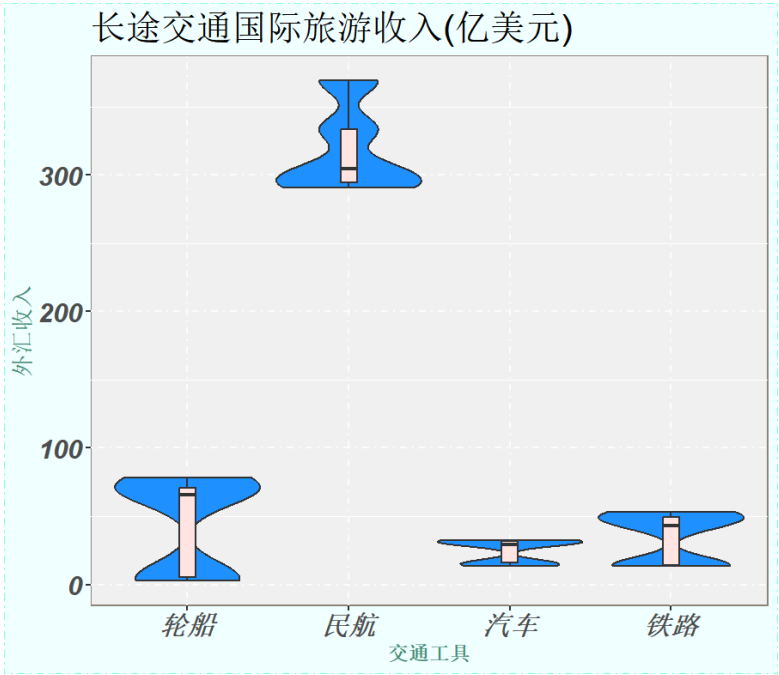
此外，为进一步分析入境旅游消费，同时为国内旅游业针对入境游客服务提升提供可靠依据，本报告对国际旅游收入及构成做了详细划分及比较：



图表 8 国际旅游收入及构成南丁格尔玫瑰图

在该玫瑰图中，长途交通国际旅游外汇收入（亿美元）花瓣最大，即其收入占比最大，在总外汇收入中占据极重要的地位；其它花瓣则较为均匀，其中邮电通讯与市内交通花瓣极小，这类花费本质决定了其不会创造较高的外汇收入，但娱乐和游览也并不突出，这可能是国内旅游业未来一段时期的发展改进方向。

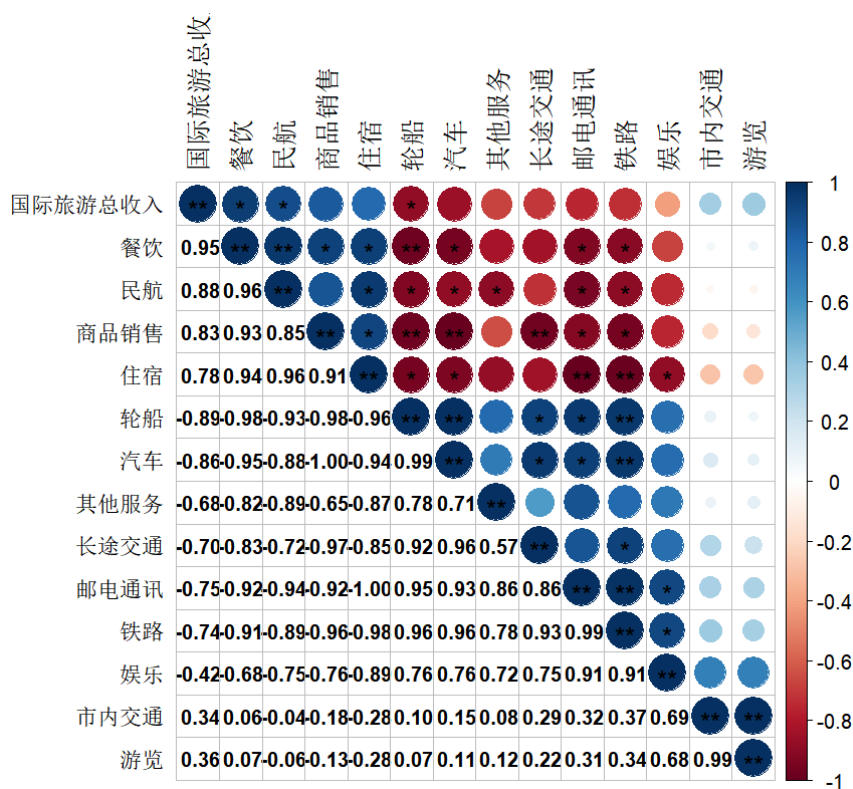
由于长途交通收入的突出占比，本报告进一步对各类长途交通方式进行分析，包括轮船、民航、汽车及铁路四种交通方式，对其进行数据范围及核密度的确定：



图表 9 长途交通收入小提琴图

在这张小提琴图中，我们可以清楚地看到：民航在四种交通工具中的外汇收入最高，这也是各国入境旅游的主要交通方式，快捷、必要且花费较多；而其余三类（轮船、汽车及铁路）则相差不大，但中位数都较高。

本报告还对各项国际旅游收入之间进行了相关性分析并绘制热力系数图，以探明各项收入之间的依赖关系，为政府企业决策提供可靠依据：



图表 10 各项国际旅游收入热力系数图

在这幅热力系数图中，不仅可以看到各个项目之间的依赖关系，同时还可以看出各项收入与总外汇收入之间的相关性。其中，与总收入相关度较高的依次有餐饮、各项长途交通、商品销售等，政府及企业对于这些方面可以针对性制定提升方案，以提高国际旅游外汇总收入。

最后，本报告使用 python 爬虫技术爬取了近期网络热度较高且与旅游相关的词语，并绘制了星型词云图：



图表 11 星型词云图

四、模型建立与分析

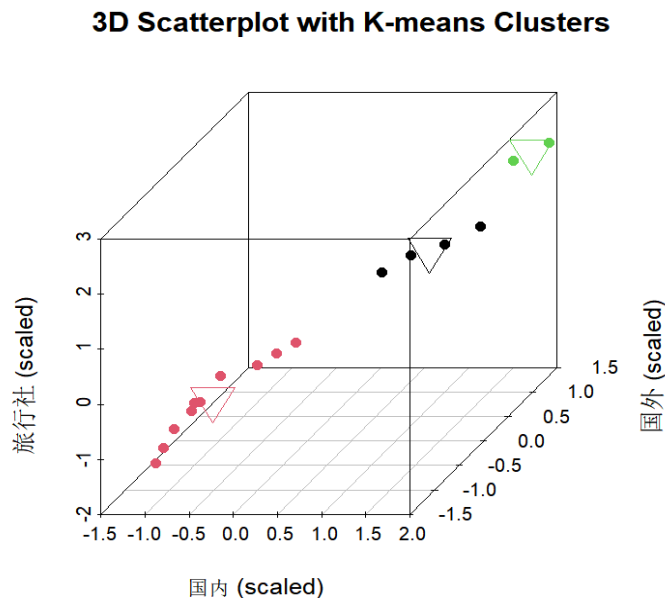
为对中国旅游业发展状况进一步探究，本报告决定就其发展前景进行预测，建立中国旅游业发展预测模型。

首先为更加科学、全面地进行预测，使用 K-means 聚类分析方法将中国旅游业自 2004 至 2019 年时期（2019 年末疫情爆发，本报告选择建立普适模型）据其综合发展指标分为三个阶段：快速上升期、平稳过渡期和缓慢增长期；之后，综合三个阶段采用 Arima 时间序列预测模型对中国旅游业发展（国内、国际两大方面）进行预测；最后通过 LSTM 深度学习模型对其预测结果进一步验证分析。

1. K-means 聚类分析

K-means 算法是一种迭代求解的聚类分析算法，其核心思想是将数据集中的 n 个对象划分为 K 个聚类，使得每个对象到其所属聚类的中心（或称为均值点、质心）的距离之和最小。

在本报告中，以 2004 至 2019 年作为研究对象，选取归一化后的国内旅游总花费（亿元）、国际旅游收入（亿美元）及旅行社数（个）作为综合发展指标进行三维聚类分析，得到如下 3D 图：



图表 12 发展阶段 3D 聚类图

进而得到下表：

阶段	年	经历时期
2004-2013		缓慢增长期
2014-2017		平稳过渡期
2018-2019		快速上升期

表格 4 不同发展阶段

注：这里的总体趋势始终上升，受疫情影响前，2018-2019 年正值高峰火热期。

2. Arima 模型预测

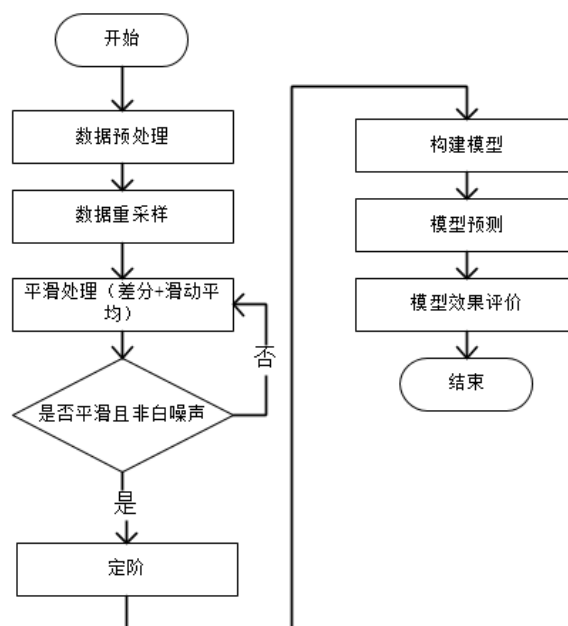
在得到 K-means 聚类分析对历史的结论——中国旅游业经历的三个时期后，我们进一步探索未来，分别对未来 8 年的国内旅游总花费(亿元)、国际旅游收入(亿美元)进行预测。

Arima 模型基本原理：将数据通过差分转化为平稳数据，再将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型。

$$y_t = \mu + \sum_{i=1}^p r_i y_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

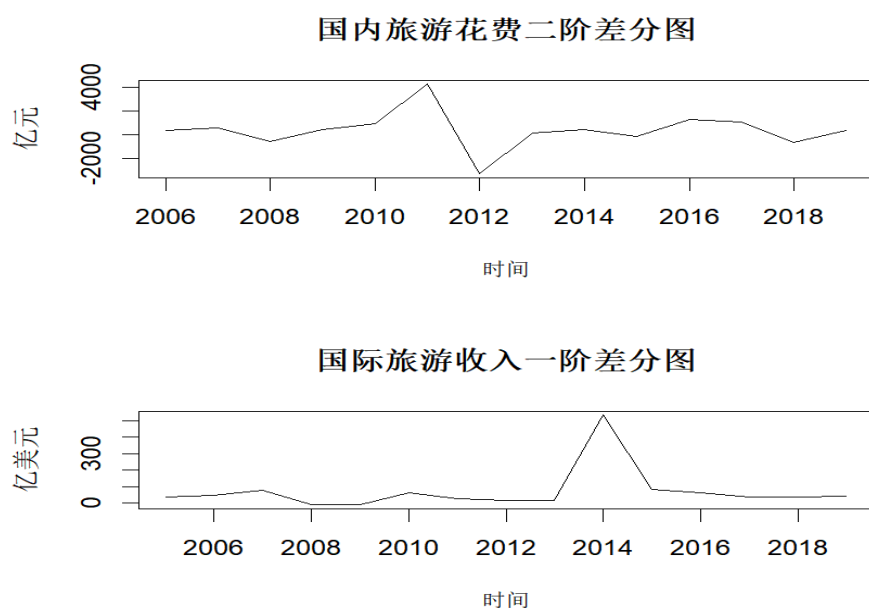
AR 是自回归，p 为自回归项；MA 为移动平均，q 为移动平均项数，d 为时间序列成为平稳时所做的差分次数，一般做一阶差分，很少做二阶差分。

一般步骤如下图所示：



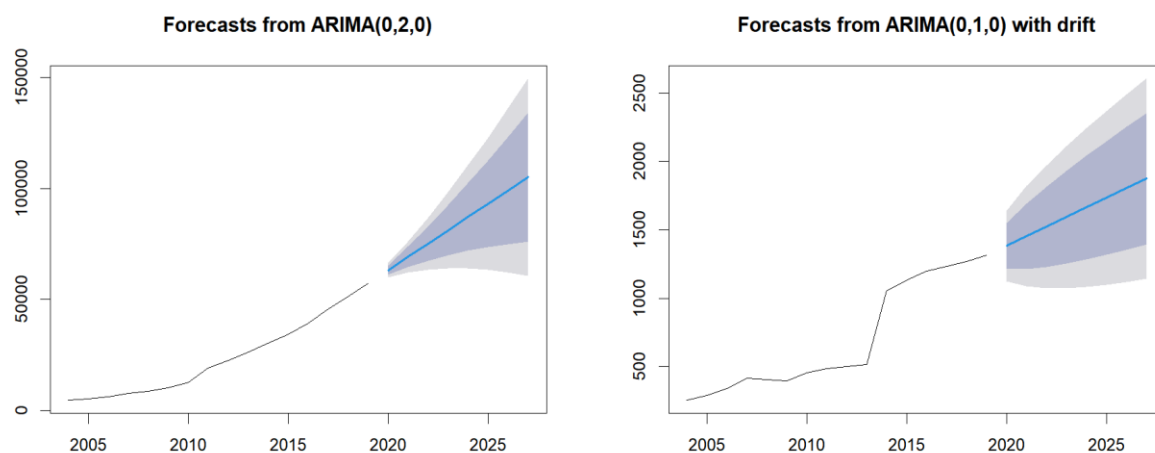
图表 13 Arima 预测过程

该 Arima 模型所用到的两组数据：国内旅游花费（亿元）、国际旅游收入（亿美元）均经过预处理确保无缺失值、异常值，之后分别进行二阶、一阶差分做平稳化处理：



图表 14 一、二阶差分图

在平稳化后，使用 R 提供的 `auto.arima` 自动对两个模型进行判别及定阶并得出未来 8 年的预测结果：



图表 15 Arima 预测结果

在 Arima 自动预测模型可以看到：国内旅游花费（左侧）、国际旅游收入（右侧）均成良好发展态势，但同时我们也可以观察到，Arima 的 p 阶、 q 阶均为 0。此类模型就是一个简单的随机游走模型（Random Walk）加上一个常数项。

即模型形式为：

$$y_t = y_{t-1} + \varepsilon_t + c$$

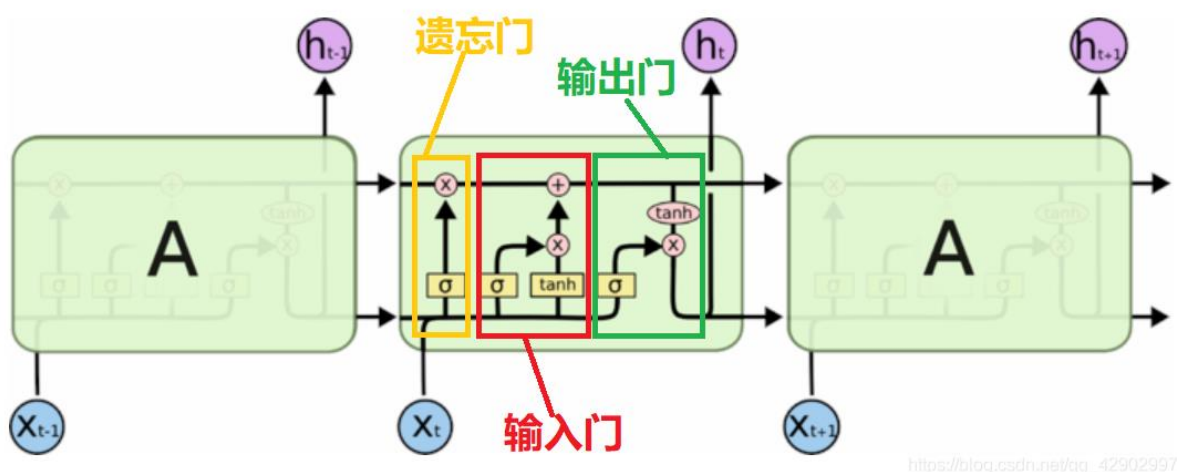
其中 ε_t 噪声， c 常数。

这种情况下，模型对序列的依赖主要在于前一个时间点的值，并且没有考虑更长期的历史数据的自回归结构和移动平均结构。

3. LSTM 预测

因此，我们考虑继续使用深度学习模型——LSTM 进行时间序列预测。

LSTM 是一种改进的循环神经网络（RNN），专门设计用于处理和预测具有长期依赖关系的序列数据。它通过引入遗忘门、输入门和输出门等结构来有效地管理记忆和遗忘信息，从而更好地捕捉时间序列中的长期模式和趋势。如下图所示：

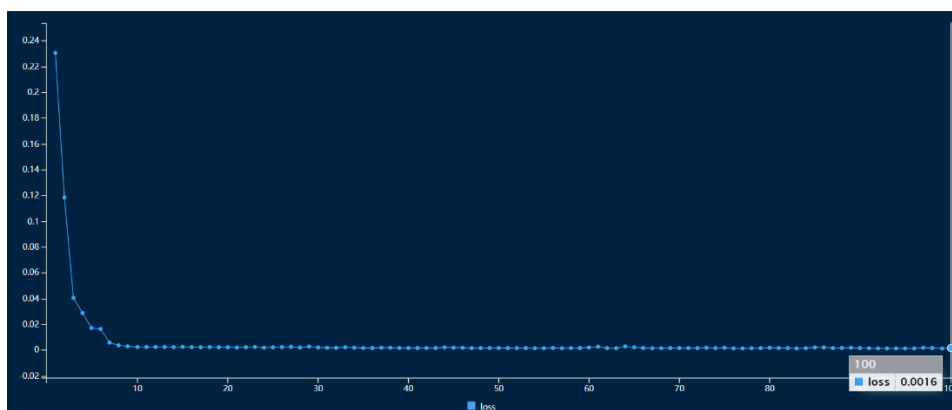


图表 16 LSTM 原理图

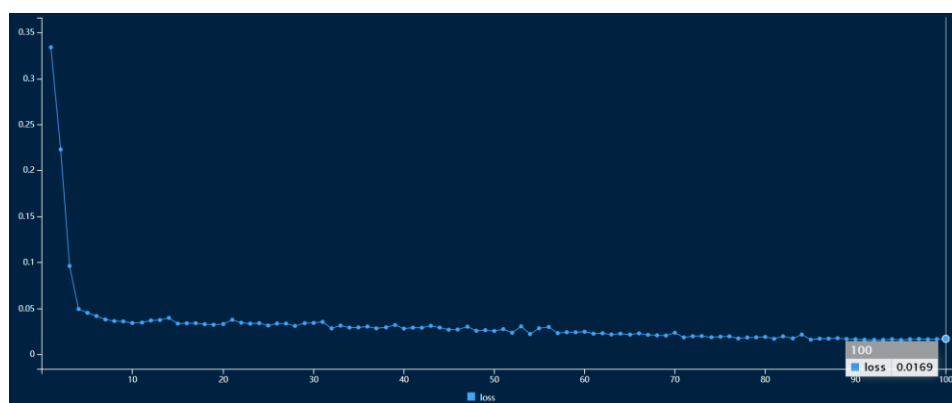
据此，我们进行 LSTM 预测：

在使用 LSTM 进行国际旅游收入和国内旅游花费预测时，我们需要关注模型训练过程中的损失函数（Loss function）的变化。损失函数通常用于衡量模型预测值与真实值之间的差异。

下面的两幅图展示了损失函数随训练迭代次数的变化，这些图表往往显示出损失函数逐渐减小，反映了模型在学习和优化过程中的改善。

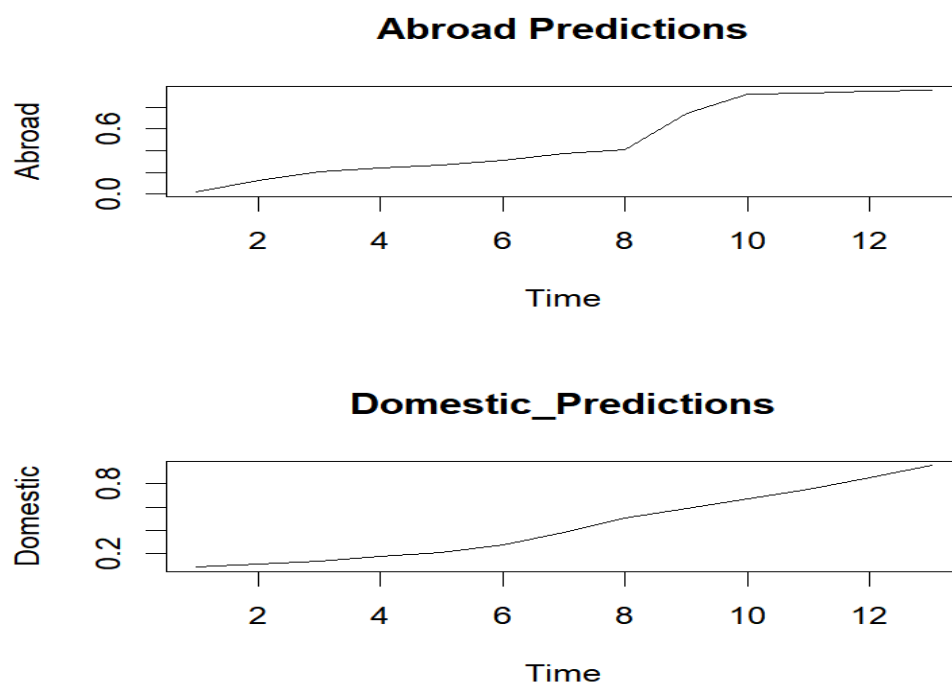


图表 17 国际旅游收入 Loss 迭代图



图表 18 国内旅游花费 Loss 迭代图

接下来就是国际旅游收入和国内旅游花费的预测结果：



图表 19 LSTM 预测结果

从预测结果可以看出：国外和国内旅游预测结果依旧均显示**上升趋势**，同 Arima 模型预测十分相似。这种趋势反映了预测模型对未来中国旅游业发展的乐观预期，这些信息对于政策制定和市场策略制定具有重要意义。

六、结论与建议

通过中国旅游业发展状况的深入探究以及发展前景的模型预测，我们得出一系列结论，包括针对国内游客及国际入境游客，同时基于所得结论为政府及企业提供相应建议，助力中国旅游业在疫情过后经济复苏的大环境下愈发蓬勃。

1. 结论

◆**总体发展情况**：中国旅游业自 2004 年以来呈现出稳步上升的趋势，尽管受到 2019 年末新冠疫情的冲击，但整体上仍保持较强的活力和增长势头。旅行社和星级饭店的数量变化反映了旅游业的发展动态，其中星级饭店行业经历了由快速发展到优胜劣汰的转变。

◆**国内旅游情况**：国内游客数量持续增长，特别是城镇居民游客数量增长显著，反映了人口城镇化的趋势。农村居民游客数量虽稳定但人均花费较低，显示出城乡居民在旅游消费上的差异。整体而言，国内旅游人均花费稳步上升，显示出良好的发展态势。

◆**入境旅游情况**：入境游客数量中，亚洲国家占据主导地位，欧洲次之，反映出地理和文化因素对入境旅游的影响。长途交通是入境旅游外汇收入的主要来源，其中民航占比最高。此外，国际旅游收入各构成部分之间存在一定的依赖关系，餐饮、长途交通和商品销售与总收入相关性较高。

2. 建议

- (1) **加强旅游业基础设施建设**：继续提升旅行社和星级饭店的服务质量，优化旅游环境，以满足日益增长的旅游需求。
- (2) **促进城乡旅游均衡发展**：针对农村居民游客数量稳定但人均花费较低的情况，可以开发更多适合农村居民的旅游产品，提高农村居民的旅游消费能力。
- (3) **加强国际旅游市场开拓**：针对入境旅游市场，可以进一步加大对亚洲和欧洲等主要

客源市场的宣传和推广力度，同时探索新的客源市场。

- (4) **优化旅游消费结构**：针对国际旅游收入构成，可以加强餐饮、长途交通和商品销售等领域的服务质量和产品创新，提高这些领域的收入贡献率。
- (5) **加强数据分析和预测**：合理有效利用机器学习、深度学习等各类数据挖掘分析方法，加强对旅游业发展趋势的预测和分析，为政府和企业决策提供科学依据。

六、参考文献

- [1] 秦宏瑶, 胡丹, 杜晓希. 后疫情时代四川入境旅游趋势与渐进式应对 [Z]. 北方经贸, 2023 (2) :150-152.
- [2] 林文凯, 夏会琴, 胡海胜. 改革开放以来中国入境旅游周期波动的阶段性变迁分析 [Z]. 江西财经大学学报, 2023 (1) :53-64.
- [3] 杨劲松. 中国入境旅游再出发的期待和行动 [Z]. 中外文化交流, 2023 (3) :49-51.
- [4] 尚林, 秦伟良. 基于 ARIMA 和 EGARCH 模型的中国入境旅游收汇预测分析 [J]. 统计教育, 2007 (4) :46-48.
- [5] 刘昌龙. 国内旅游需求研究综述 [Z]. 中文科技期刊数据库 (文摘版) 社会科学, 2024 (1) :0154-0157.
- [6] 彭淑贞, 刘桂菊, 梁薇. 旅游业高质量发展评价指标体系构建研究 [Z]. 中国商论, 2024 (4) :149-152.
- [7] 李文媛. 中国入境旅游如何实现高质量发展——基于中国入境旅游的客源结构分析 [Z]. 大陆桥视野, 2024 (1) :32-33.
- [8] Biyun Chen, Wenqiang Li. Prediction of Wordle Scores Based on ARIMA and LSTM Models [Z]. JOURNAL OF APPLIED MATHEMATICS AND PHYSICS, 2024, 12 (2) :543-553.