Tourism Regression Analysis

DS6021 - Linear Models

Group 5: Gabby Cordelli, Disha Dubey, Anjali Goel, John Hope, Kurt Samuels

01

Dataset Overview

Data Sources

UN Tourism

Annual tourism data collected from countries through a series of yearly questionnaires. Contains key statistics, such as tourism income, number of visitors, tourism infrastructure, etc.

U.S. Department of Agriculture

Annual historical macroeconomic data for countries. Chosen indicators were real GDP, real GDP per capita, real exchange rates, and consumer price indexes

World Bank

Annual historical unemployment rates for countries

UN Development Programme

Annual historical Human Development Index (HDI) metrics for countries

Research Questions

01

How well are countries' macroeconomic indicators able to predict their tourism income?

02

How well are countries' tourism infrastructure able to predict their tourism income?

03

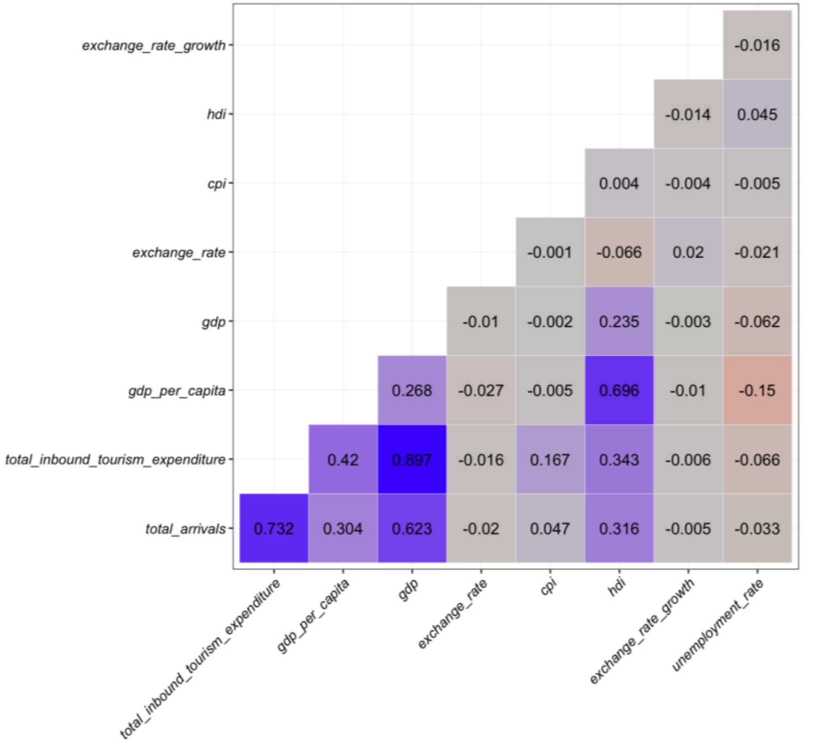
How well are we able to predict whether or not a country grew their tourism income based on the annual infrastructure and macroeconomic growth indicators?

02

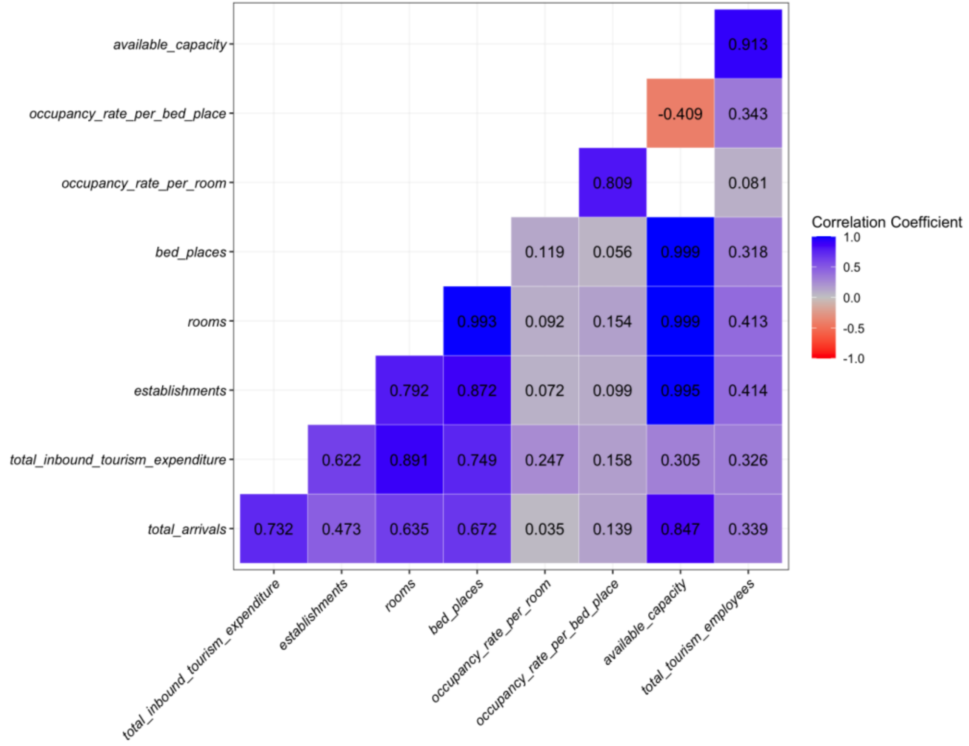
Exploratory Data Analysis

Correlation Plots

Economic Features

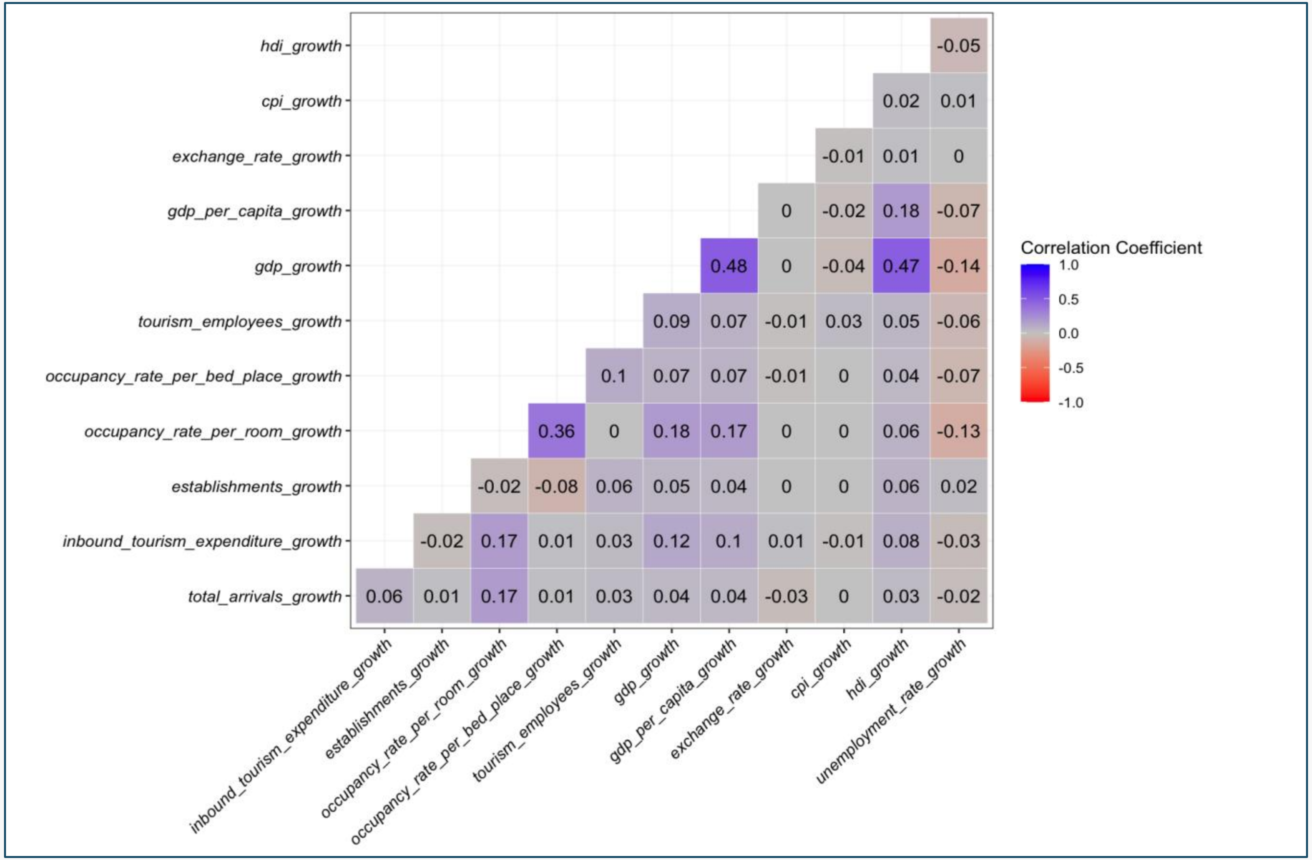


Infrastructure Features



Correlation Plots

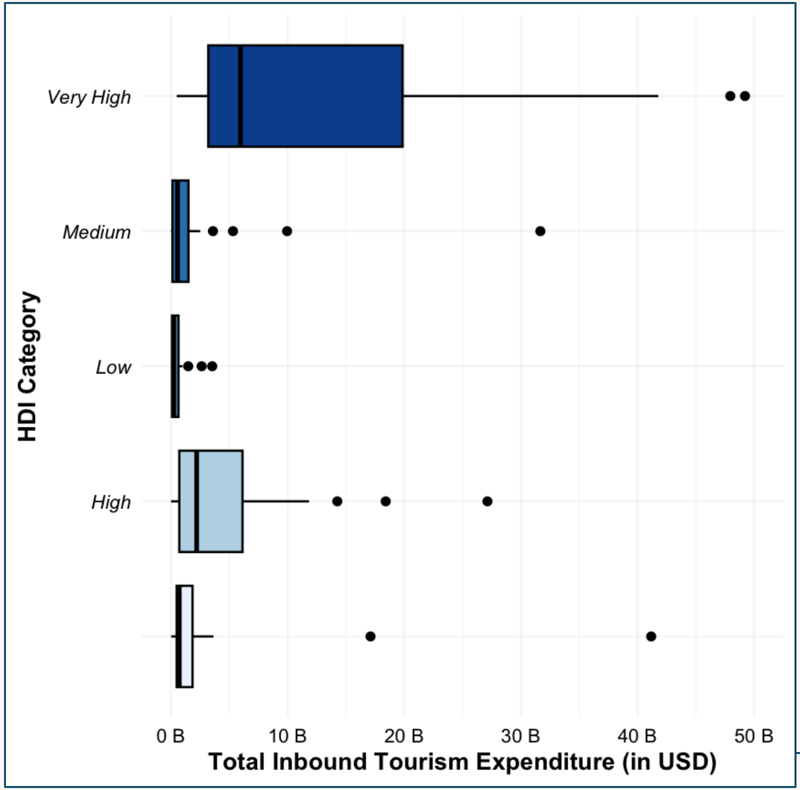
Growth Features



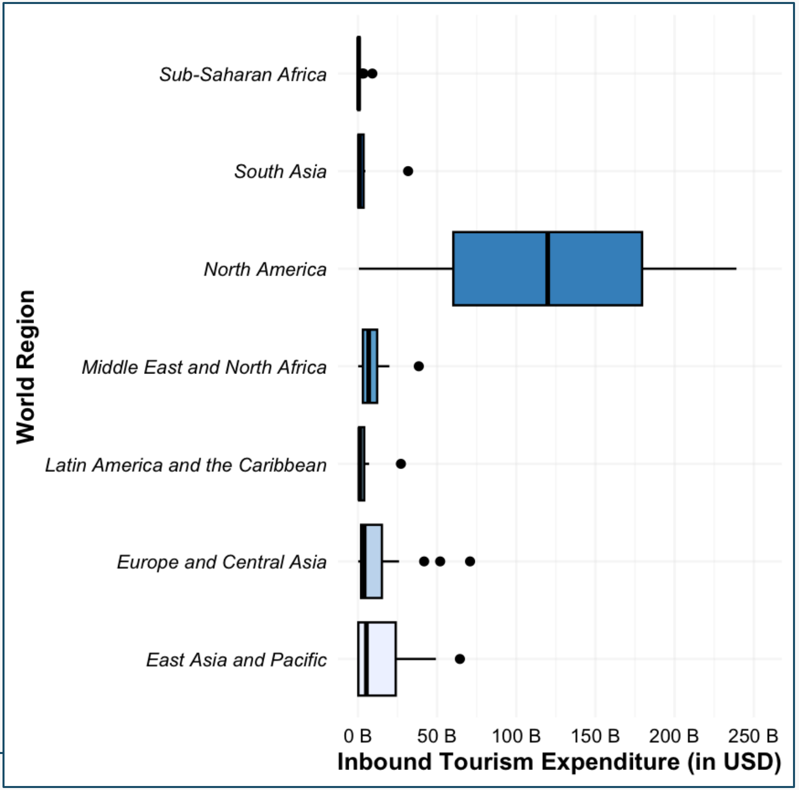
	inbound_tourism_expenditure_growth	establishments_growth	occupancy_rate_per_room_growth	occupancy_rate_per_bed_place_growth	tourism_employees_growth	gdp_growth	gdp_per_capita_growth	exchange_rate_growth	cpi_growth	hdi_growth	unemployment_rate_growth
hdi_growth											-0.05
cpi_growth										0.02	0.01
exchange_rate_growth									-0.01	0.01	0
gdp_per_capita_growth								0	-0.02	0.18	-0.07
gdp_growth							0.48	0	-0.04	0.47	-0.14
tourism_employees_growth						0.09	0.07	-0.01	0.03	0.05	-0.06
occupancy_rate_per_bed_place_growth					0.1	0.07	0.07	-0.01	0	0.04	-0.07
occupancy_rate_per_room_growth				0.36	0	0.18	0.17	0	0	0.06	-0.13
establishments_growth			-0.02	-0.08	0.06	0.05	0.04	0	0	0.06	0.02
inbound_tourism_expenditure_growth		-0.02	0.17	0.01	0.03	0.12	0.1	0.01	-0.01	0.08	-0.03
total_arrivals_growth	0.06	0.01	0.17	0.01	0.03	0.04	0.04	-0.03	0	0.03	-0.02

2019 Tourism Income

HDI Category



World Region



03

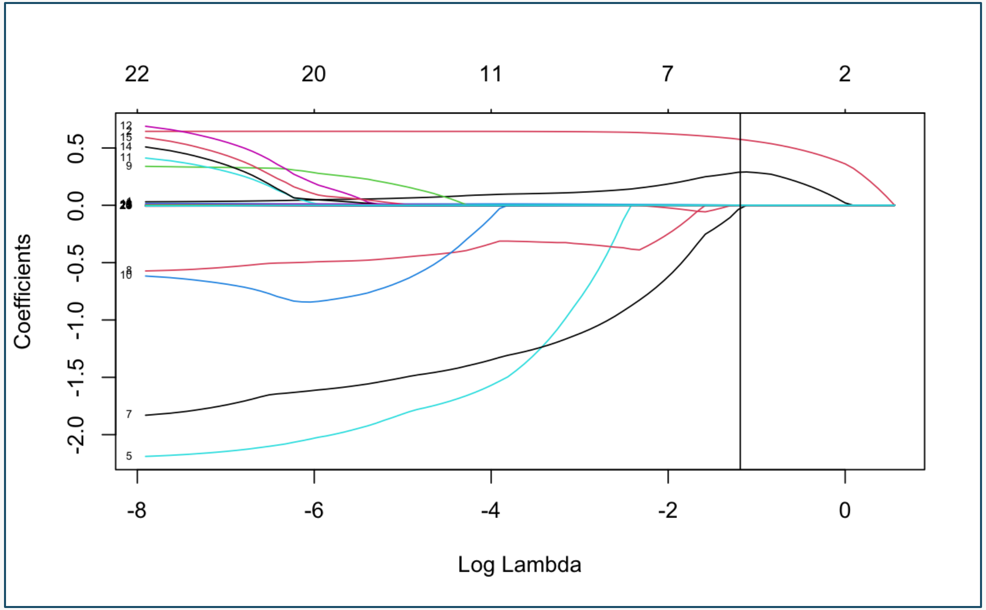
Multiple Linear Regression

Model 1 – Model Building

Goal Predict annual tourism income by macroeconomic indicators

Lasso Regression

- Introduces a penalty to the model to get sparse estimates
- Lasso only kept GDP per capita, GDP, CPI, HDI, and HDI code
- Outperformed Ridge & AIC



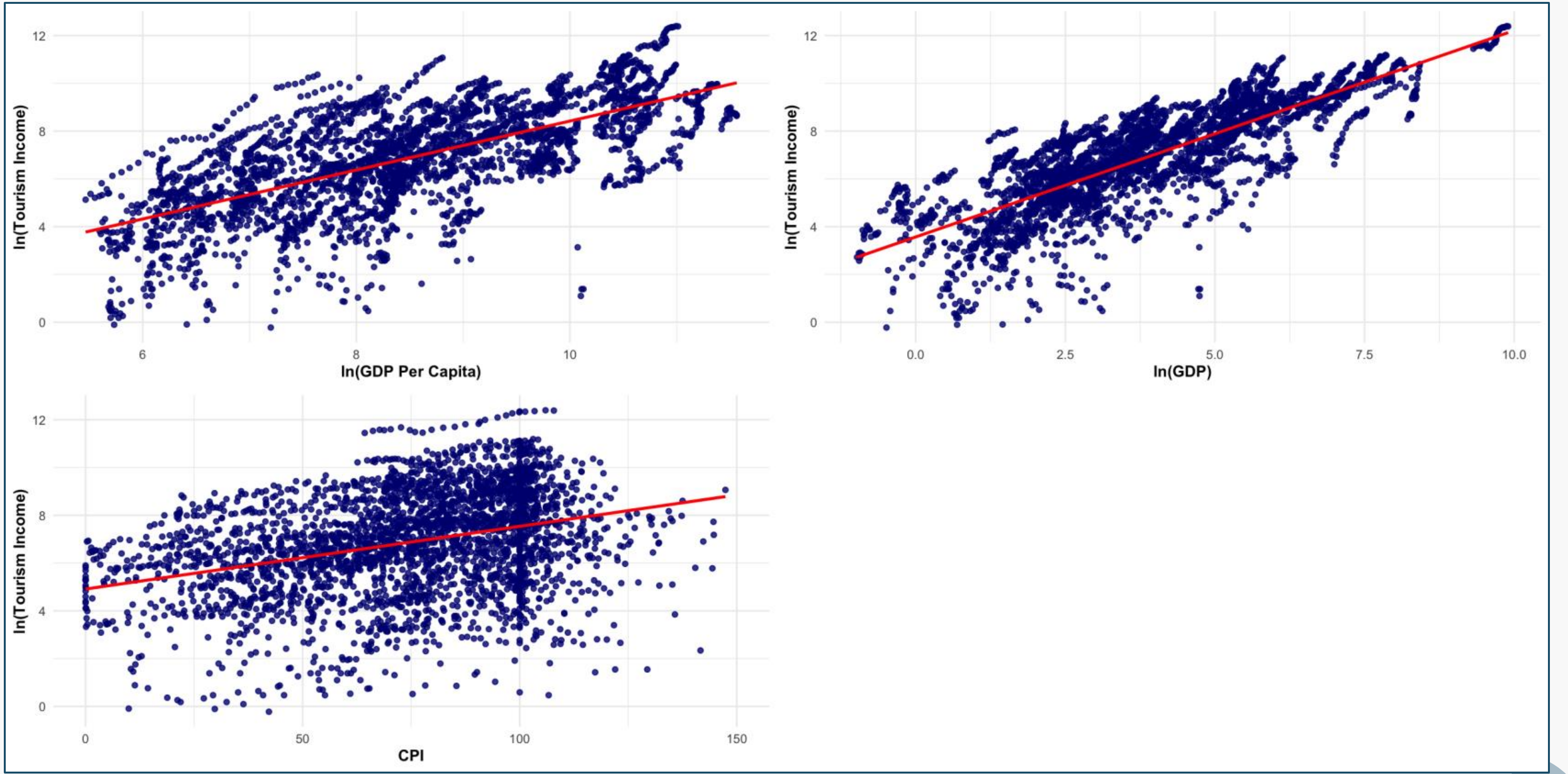
Multicollinearity

- Removed HDI (number) from the model due to high VIF(> 10)

Final Model

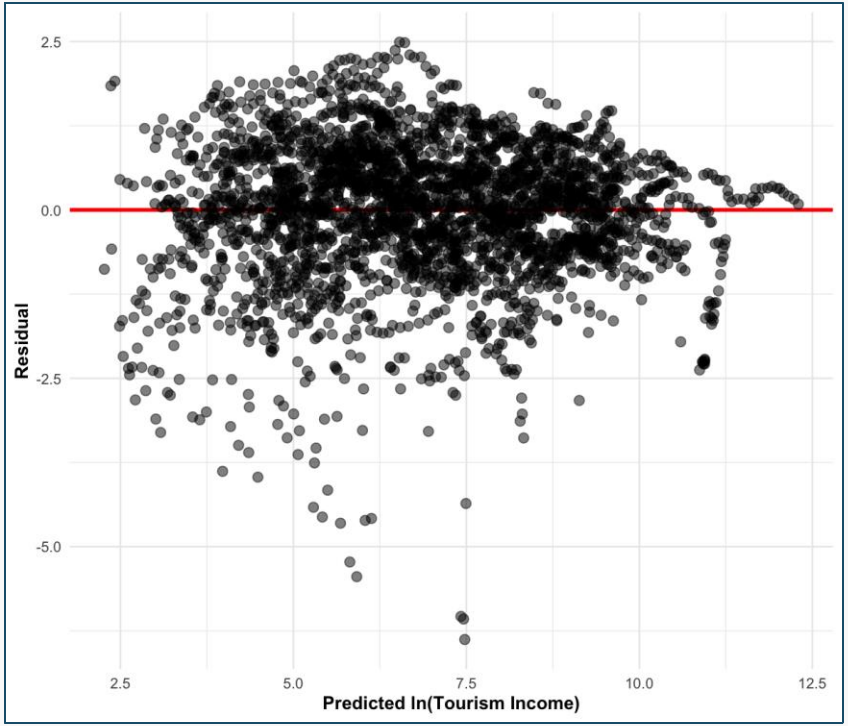
$$\ln(\text{Tourism Income}) = 2.92+0.09*\ln(\text{GDP per Capita})+0.67*\ln(\text{GDP})+0.01*\text{CPI} -1.28*I_{\text{HDI\_Low}} -0.73*I_{\text{HDI\_Med}} + 0.27*I_{\text{HDI\_VeryHigh}}$$

Model 1 – Model Assumptions

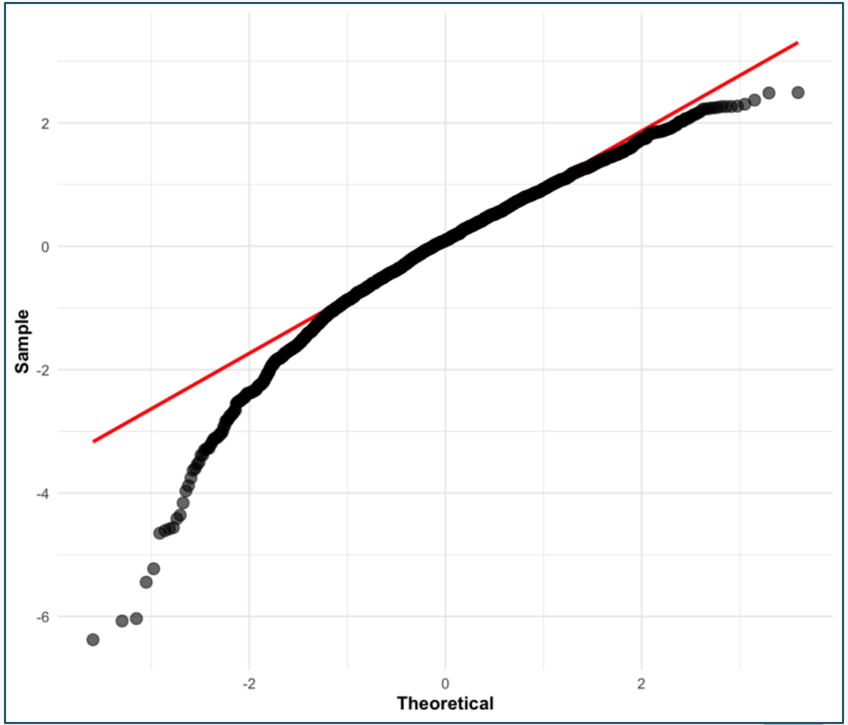


Model 1 – Model Assumptions

Residuals



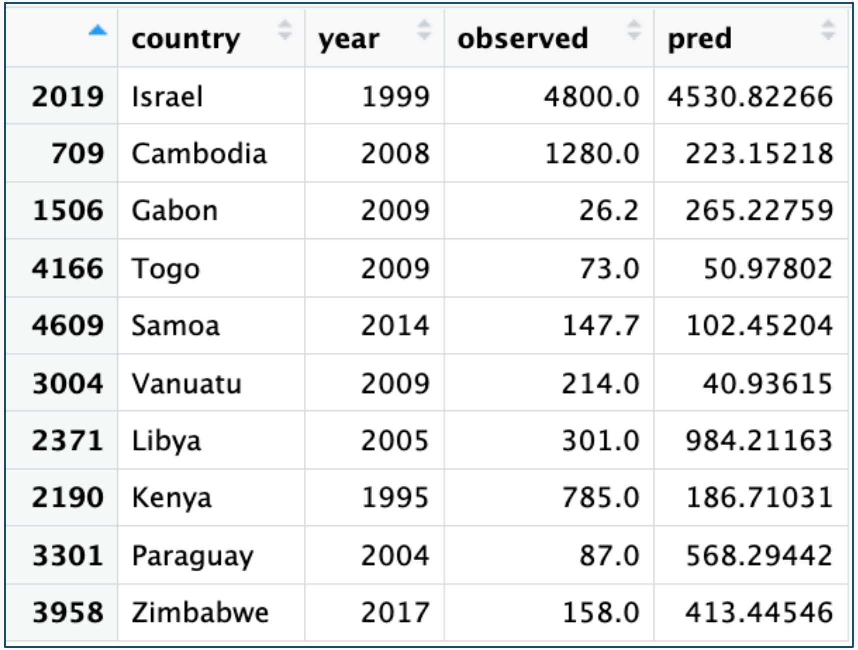
QQ Plot



Model 1 – Predictions & Assessments

		ln(Tourism Income) Predictions
Measures	RMSE	1.063
	MAE	0.800
	R-Squared	0.758

		Tourism Income (in millions) Predictions
Measures	RMSE	8,743.062
	MAE	2,820.360
	MAPE	244.318%



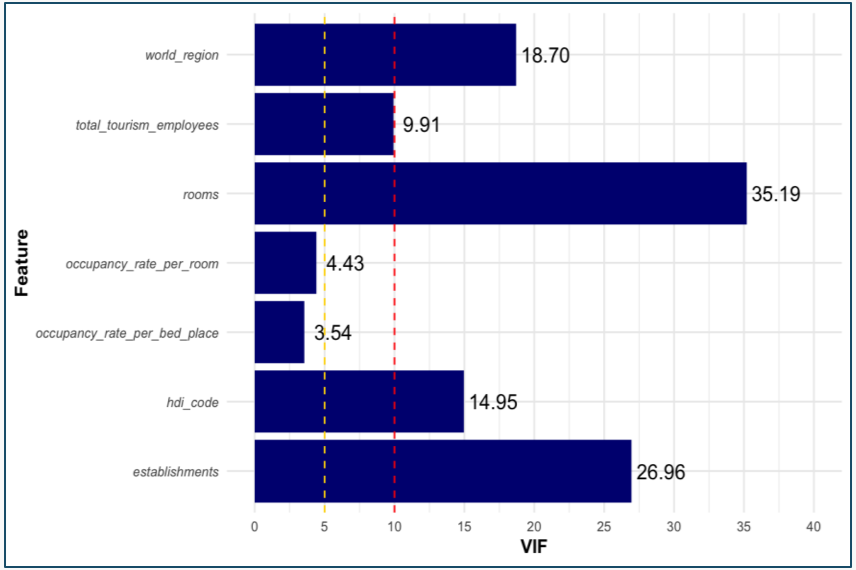
	country	year	observed	pred
2019	Israel	1999	4800.0	4530.82266
709	Cambodia	2008	1280.0	223.15218
1506	Gabon	2009	26.2	265.22759
4166	Togo	2009	73.0	50.97802
4609	Samoa	2014	147.7	102.45204
3004	Vanuatu	2009	214.0	40.93615
2371	Libya	2005	301.0	984.21163
2190	Kenya	1995	785.0	186.71031
3301	Paraguay	2004	87.0	568.29442
3958	Zimbabwe	2017	158.0	413.44546

Model 2 – Model Building

Goal Predict annual tourism income by infrastructure indicators

Initial Model

- Initial model contained all infrastructure features of interest
- Returned insignificant coefficients and high VIF values, indicating correlations between predictors



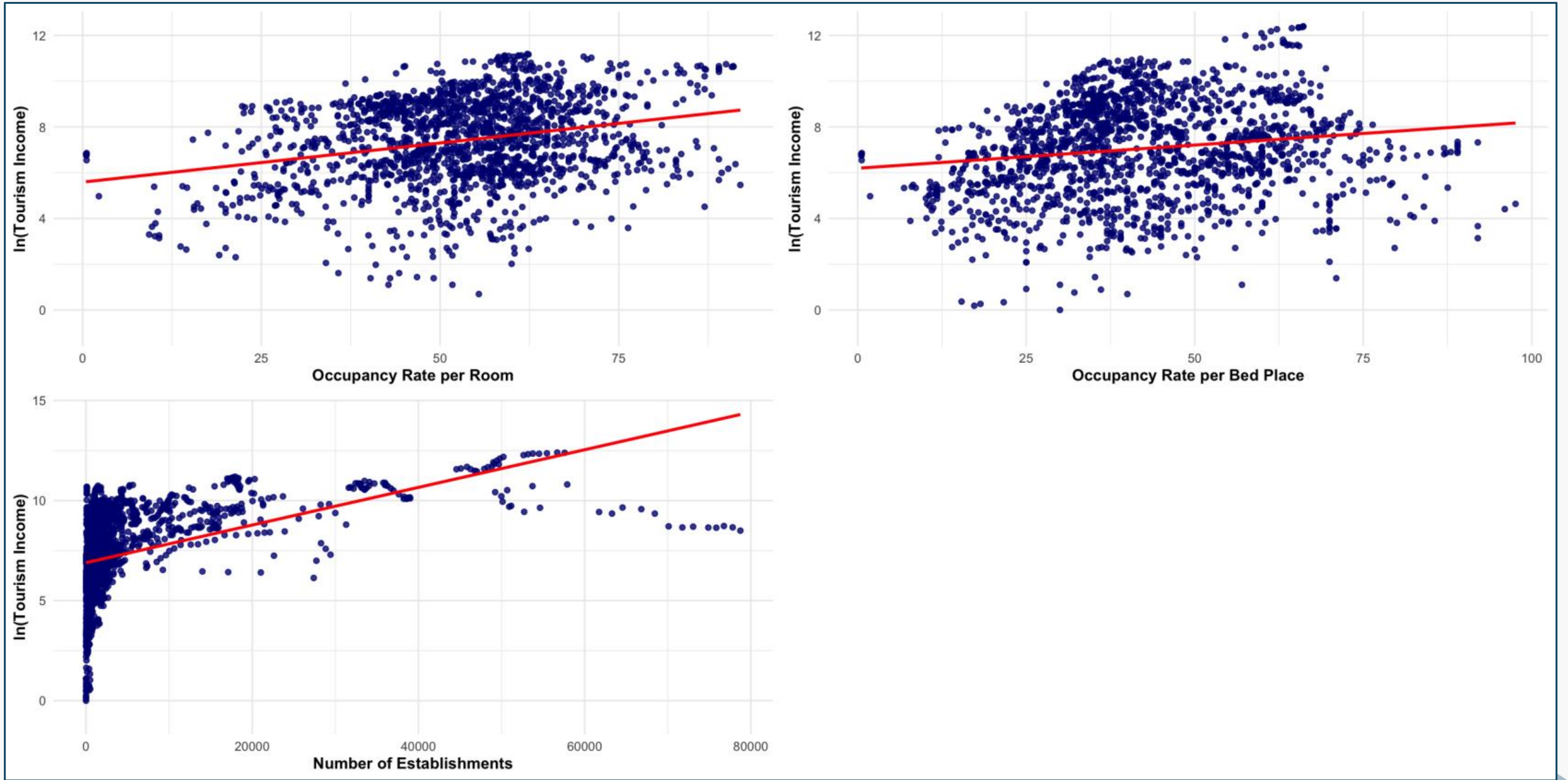
Step-Wise AIC

- Used AIC for variable selection and removed high VIF variables

Final Model

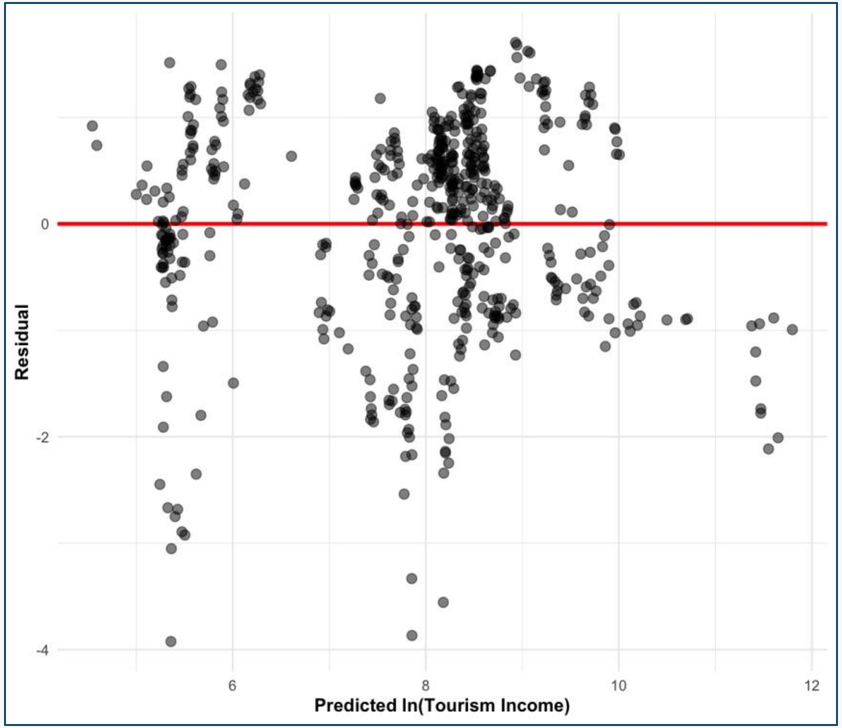
$$\ln(\text{Tourism Income}) = 9.09 - .02*\text{OccRateRoom} + .03*\text{OccRateBed} + 0.00005*\text{Establishments} - 0.57*I_{\text{HDI\_High}} - 0.93*I_{\text{HDI\_Low}} + 0.27*I_{\text{HDI\_Medium}} + 0.03*I_{\text{HDI\_VeryHigh}} - 1.05*I_{\text{ECA}} - 1.82*I_{\text{LAC}} - 0.15*I_{\text{MENA}} - 2.75*I_{\text{SSA}}$$

Model 2 – Model Assumptions

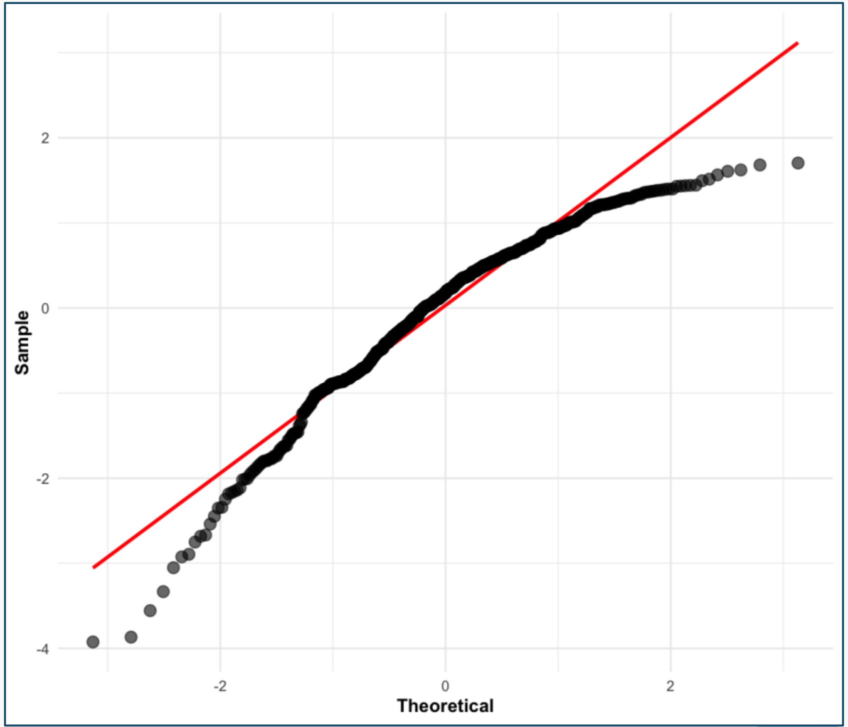


Model 2 – Model Assumptions

Residuals



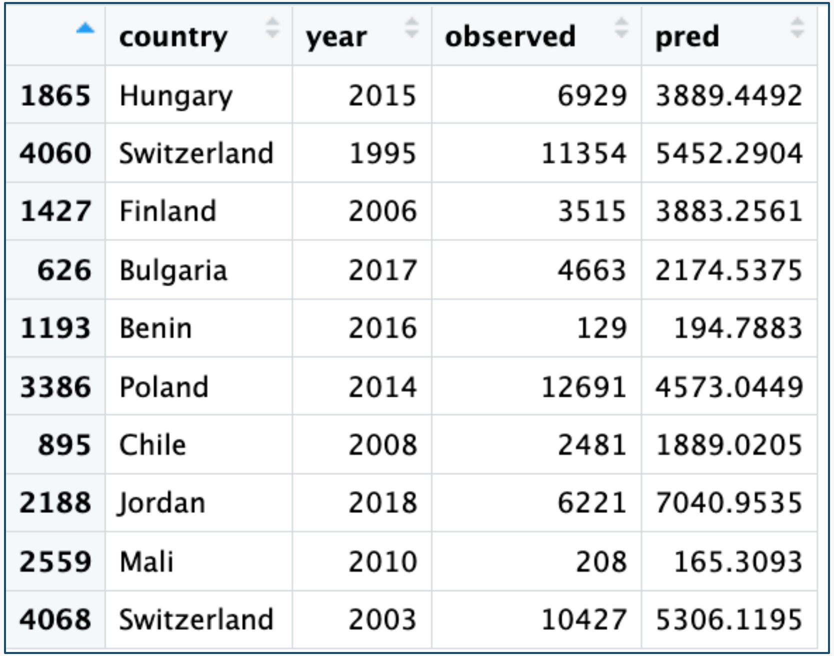
QQ Plot



Model 2 – Predictions & Assessments

		ln(Tourism Income) Predictions
Measures	RMSE	0.999
	MAE	0.798
	R-Squared	0.663

		Tourism Income (in millions) Predictions
Measures	RMSE	12,910.232
	MAE	5,573.397
	MAPE	143.344%



▲	country	year	observed	pred
1865	Hungary	2015	6929	3889.4492
4060	Switzerland	1995	11354	5452.2904
1427	Finland	2006	3515	3883.2561
626	Bulgaria	2017	4663	2174.5375
1193	Benin	2016	129	194.7883
3386	Poland	2014	12691	4573.0449
895	Chile	2008	2481	1889.0205
2188	Jordan	2018	6221	7040.9535
2559	Mali	2010	208	165.3093
4068	Switzerland	2003	10427	5306.1195

04

Logistic Regression

Model 3 – Model Building

Goal Predict whether or not country's tourism income grew at least 5% compared the previous year

Initial Model

- Originally included all prior year growth variables for visitor, economic, and infrastructure features

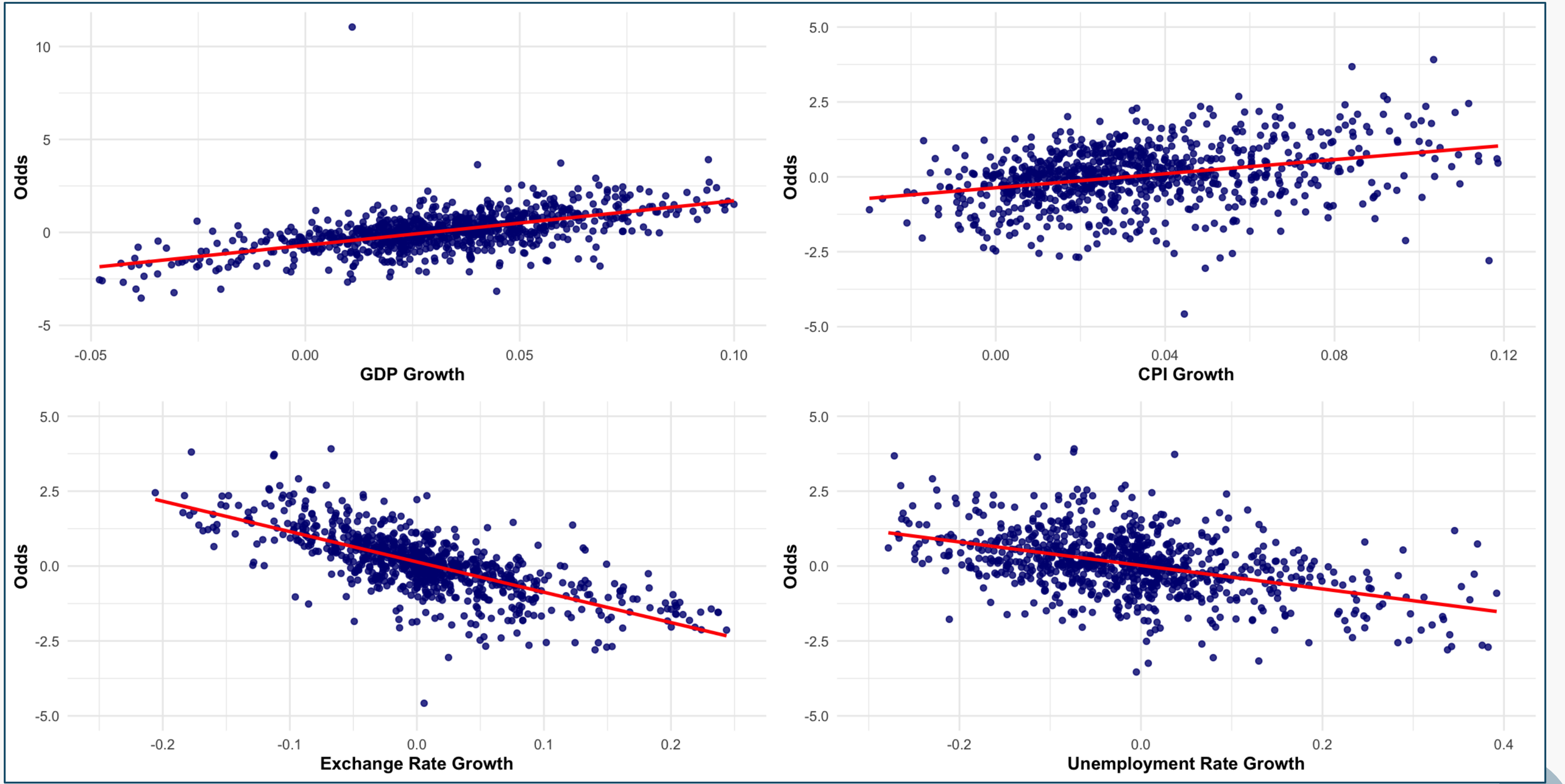
Step-Wise AIC

- Removed insignificant variables, only keeping the prior year growths for the following: total arrivals, GDP, GDP per capita, Consumer Price Index, unemployment rate

Final Model

OddsTourismGrew = -0.60+2.20*ArrivalsGrowth+13.94*GDPGrowth+3.09*CPIGrowth-7.11*ExchangeGrowth
-1.08*UnempGrowth

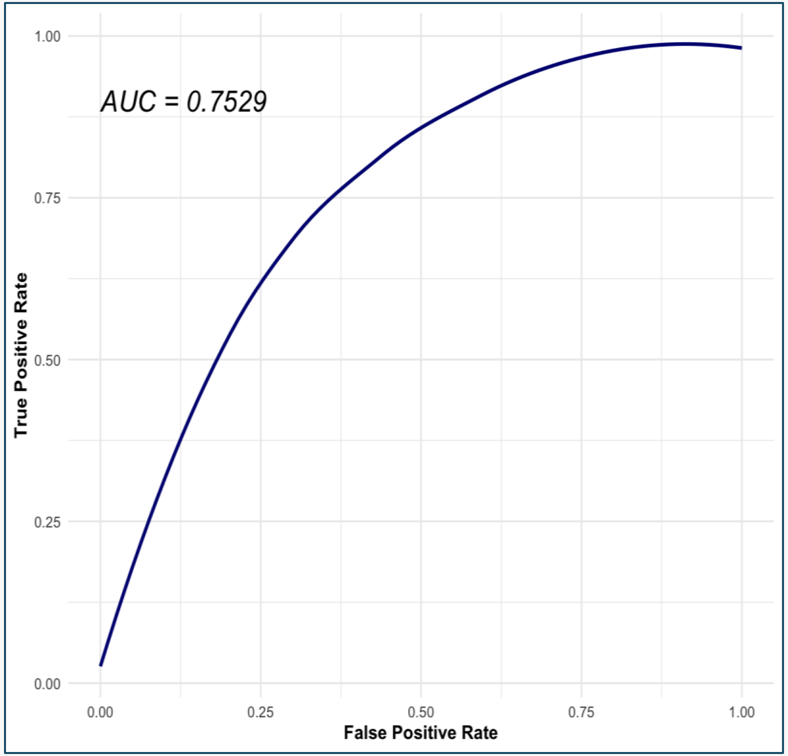
Model 3 – Model Assumptions



Model 3 – Predictions & Assessments

		Actual	
	Tourism Grew 5%?	Yes	No
Predict	Yes	300	130
	No	128	291

Measures	Accuracy	69.61%
	Sensitivity	70.09%
	Specificity	69.12%
	Precision	69.78%
	F-1 Score	0.6993



References

[1] https://www.unwto.org/tourism-statistics/key-tourism-statistics

[2] https://ers.usda.gov/data-products/international-macroeconomic-data-set.aspx

[3] https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS

[4] https://hdr.undp.org/data-center/documentation-and-downloads

Thank you!

GitHub Repository:

https://github.com/johnhope829/tourism_regression_analysis