Durham E-Theses

Insights from Tweets: Analysing Destination Topics and Sentiments, and Predicting Tourist Arrivals

YULEI LI

Appendix

Tables

Table 1. Categories with topics and keywords (Sydney)

Category	Topics with important keywords
Economic	*flights; sale airfare; fly Sydney; fares*
Exogenous	*sydney; sydney morning; sydney news; news aus*
	jobs Sydney; job Australia; sydney jobs; job jobs
	sydney property; real estate; house prices; property market
	sydney traffic; trafficnetwork; sydtraffic trafficnetwork; trafficnetwork sydneys
	Building; sydney architecture; building Sydney; house sydney
	sydney cat; dog Sydney; pets; dog lovers
	Iraq; Syria; iraq Sydney; syria world
	asian shares; sydney reuters; reuters Asian; betting china
	sydney Anglicans; archbishop Sydney; sydney Anglican; church sydney
	australian shares; australia shares; trading Sydney; australian stocks
	Cloud; aws; sydney cloud; cloudcentral
	aud asx; open aud; market open; sydney stock
	gay marriage; gay sydney; sydney gay; marriage sydney
Internal	*sleeps; sydney sleep; sleep sydney; rough sleepers*
Attraction	*opera house; sydney opera; house sydney; opera sydney*
	biennale; biennale sydney; sydney art; art sydney
	harbour bridge; bridge sydney; bridge climb; sydney bridge
	zoo; taronga zoo; zoo sydney; sydney zoo
	gardens; botanic gardens; botanical gardens; gardens sydney
	sydney fish; fish market; fish markets; sydney salmon
Geographic	*raining sydney; sydney rain; rain sydney; sydney raining*
	bondi beach; beach sydney; bondi sydney; sydney bondi
	kmh humidity; humidity; clouds wind; sydney weather

	atis; sydney atis; aest temperature; overhead sydney
	temp; temp increased; australia current; current temp
Tourism facilities	*hotels; hotel sydney; hotels sydney; sydney hotels*
	zayn sydney; sydney airport; airport sydney; zayn airport
	hospital sydney; medical; hospitals; health proposal
	qantas; london flights; direct sydney; qantas plan
	denpasarsydney delay; indonesia denpasarsydney; weather forecast; challenge bowling
	gym; fitness sydney; sydney gym; personal trainer
	telstra; free wifi; law twitter; telstra lays
Events and activities	*concert; sydney festival; sydney concert; concert sydney*
	conference sydney; workshop; sydney university; seminar
	new years; christmas; year sydney; sydney christmas
	fires; bushfire; reduction fires; bushfires
	shows; sydney shows; sydney; shows sydney
	fireworks; sydney fireworks; fireworks sydney; fireworks tonight
	sydney roosters; nrl; roosters vs; roosters sydney
	sydney fashion; fashion week; fashion sydney; fashion festival
	sydney writers; writers festival; books; sydney poetry
	film festival; sydney film; cinema sydney; film sydney
	swans; sydney swans; afl sydney; swans afl
	sydney surf; swimming sydney; city surf; pool sydney
	massage; salon; pamper package; clinic
	cycling sydney; sydney cyclist; bikelane skeptic; sydney cycling
	sydney wedding; wedding photographer; wedding sydney; wedding photography
	sydney dance; dancer sydney; dancing sydney; dance sydney
	justin bieber; meets justin; sydney meets; sydney bieber
	royal easter; sydney easter; easter sydney; easter happy
	ufc; sydney ufc; ufc sydney; ufc brisbane
	sydney olympic; olympic games; sydney olympics; 2000 olympics

	racing; jockey; sydney horse; race finish
	motorshow; motor motorshow; sydneymotorshow; motorshow sydneymotorshow
	party sydney; sydney party; party tonight; carry partying
Food	sydney; sydneys; breakfast sydney; sydney food
	coffee sydney; sydney coffee; morning coffee; cafe sydney
Symbolic	wallaby; wallaby way; 42 wallaby; calle wallaby
	tickets sydney; sydney tickets; tickets sale; ticket sydney
	liked video; video sydney; sydney liked; playlist sydney
	trending sydney; backstagepingpong trending; sydney torrential; sydney centrist
	tour sydney; touring sydney; sydney tour; australian tour
	social media; socialmedia; socialmedia crisis; blogthe social

Table 2. Descriptive summary of Sydney arrivals and overall sentiment (Sydney)

	Arrivals	Overall Sentiment
Mean	30715.6	1.143
Std	3254.78	0.04
Min	25048.0	1.093
25%	29207.0	1.12
50%	30127.5	1.13
75%	32500.75	1.173
Max	36675.0	1.21

Table 3. Descriptive summary of sentiment scores of the topics (Sydney)

	Attraction	Events and activities	Exogenous	Food	Geographic	Social	Symbolic	Tourism facilities	Travel costs
mean	1.23	1.27	1.07	1.3	1.08	0.93	1.22	1.15	1.01
std	0.08	0.07	0.04	0.07	0.06	0.13	0.05	0.08	0.12
min	1.03	1.16	1.03	1.16	0.99	0.68	1.14	1.05	0.75
25%	1.21	1.22	1.04	1.26	1.02	0.85	1.2	1.1	0.94
50%	1.25	1.26	1.06	1.3	1.09	0.91	1.23	1.13	1.05
75%	1.28	1.31	1.09	1.32	1.13	1.05	1.25	1.22	1.1

max	1.3	1.37	1.15	1.42	1.16	1.07	1.33	1.26	1.15

Table 4. Leader board of best performing models for Sydney

Model	RMSE	MAE
* DeepLearning_grid_1_AutoML_2_20220525_85253_model_114	2197.47	1593.71
DeepLearning_grid_3_AutoML_2_20220525_85253_model_25	2339.34	2235.71
DeepLearning_grid_2_AutoML_2_20220525_85253_model_11	2479.59	2038.13
DeepLearning_grid_1_AutoML_2_20220525_85253_model_341	2682.59	1967.63
DeepLearning_grid_1_AutoML_2_20220525_85253_model_63	2764.65	2144.91

Table 5. Comparison between H2O model and benchmark models

Model	MAE	MAPE
Fast Fourier Transformation	8803.0	24.95%
Facebook Prophet	2494.0	6.99%
Temporal Convolutional Network	4231.5	11.82%
LSTM	3394.5	9.47%
H2O AutoML	**2385.0**	**6.68%**

Table 6. Ten categories with keywords (London)

Category	Topics with important keywords
Economic	*expensive; money; pay; cash*
	deals; groupon; off at; daily
Exogenous	*jobs; job; manager; united*
	mayor; boris; brexit; labour
	nhs; health; nurse; hospital
	iran; syria; pakistan; palestine
	fire; fireworks; firefighters; grenfell
	paris; london paris; to paris; paris london
	muslim; muslims; islamic; islam
	gay; women; lgbt; womens
	buhari; nigeria; nigerian; president

		riots; protest; london riots; march
		stock; exchange; london stock; stock exchange
		startup; tech; startups; business
		khan; sadiq; sadiq khan; mayor
		homeless; rough; homelessness; the homeless
		accent; london accent; accents; cockney
		migrants; merkel; migrant; eu
		russian; russia; steele; the russian
		dublin; irish; ireland; belfast
		awards; award; awards in; congratulations
		war; churchill; memorial; ww2
		missing; missing from; help find; help
		kate; prince; duchess; duchess of
		charity; donate; fundraising; fundraiser
Internal		*birthday; happy birthday; happy; my birthday*
		packing; pack; packing for; packed
Attraction		*bridge; tower; london bridge; tower bridge*
		zoo; london zoo; wildlife; animals
		eye; london eye; eyes; eye london
		theatre; london theatre; theatre london; theatre in
		museum; museums; natural history; history museum
Geographic		*weather; cold; rain; the weather*
		sunset; sky; lights; light
		snow; snowing; snow in; the snow
		pollution; air; low; smog
		climate; climate change; extinction; change
Tourism facilities		*flight; airport; flights; heathrow*
		uber; taxi; cab; london taxi
		flat; bedroom; rent; bed
		hotel; hotels; hotel london; london hotel

	underground; tube; london underground; map
	shopping; shop; store; store in
	fitness; gym; workout; trainer
	bus; london bus; buses; transport
	train; trains; the train; train to
	hair; salon; barber; beauty
Events and activities	*tickets; going to; tomorrow; cant*
	music; gig; tour; radio
	event; teacher; conference; school'
	fashion; fashion week; london fashion; lfw
	marathon; london marathon; running; running the
	art; exhibition; gallery; painting
	cycling; bike; ride; cycle
	show; london show; shows; show in
	film; film festival; festival; london film
	book; books; read; london book
	christmas; xmas; merry; merry christmas
	escort; sexy; massage; services
	harry; potter; harry potter; london today
	niall; liam; louis; zayn
	church; prayer; pray; worship
	boxing; fight; vs; conor
	wimbledon; tennis; murray; finals
	party; party in; to party; the party
	wedding; royal wedding; london wedding; weddinghour
	justin; performing; he; beliebers
	orchestra; symphony; london symphony; symphony orchestra
	dance; dancing; ballet; dancers
	comic; comic con; con; london comic
	knights; london knights; game; score

	american; an american; warren; horror
	play; playing; playing in; to play
Food	*food; restaurant; tea; coffee*
Symbolic	*photo; photography; photos; photographer*
	video; liked; liked video; video from
	blog; blog post; new blog; bloggers
Social	*None*

Table 7. Descriptive summary of London arrivals and overall sentiment

	sentiment	arrivals
mean	1.19	4447.25
std	0.05	751.28
min	1.07	2975.0
25%	1.16	3982.25
50%	1.20	4504.0
75%	1.22	4988.5
max	1.29	6122.0

Table 8. Descriptive summary of sentiment scores of the topics

	Attraction	Economic	Events and activities	Exogenous	Food	Geographic factors	Internal	Symbolic	Tourism facilities	arrivals
mean	1.19	0.96	1.34	0.95	1.41	1.06	1.54	1.36	1.1	4447.55
std	0.06	0.12	0.03	0.07	0.06	0.11	0.11	0.05	0.07	751.28
min	1.04	0.64	1.3	0.84	1.26	0.84	1.3	1.24	0.95	2975.0
25%	1.13	0.87	1.32	0.89	1.38	0.99	1.46	1.33	1.06	3982.25
50%	1.19	0.97	1.34	0.95	1.41	1.05	1.55	1.35	1.11	4504.0
75%	1.23	1.06	1.36	1.01	1.45	1.15	1.61	1.39	1.14	4988.5
max	1.32	1.15	1.4	1.07	1.54	1.25	1.82	1.53	1.29	6122.0

Table 9. Leaderboard of best performing models

Model	RMSE	MAE
*	504.02	350.25
DeepLearning_grid_3_AutoML_1_20220701_93140_model_7	3	
6		

DeepLearning_grid_3_AutoML_1_20220701_93140_model_28	509.169	389.023
XGBoost_grid_1_AutoML_1_20220701_93140_model_9	519.929	406.477
XGBoost_grid_1_AutoML_1_20220701_93140_model_4	524.847	394.713
DeepLearning_grid_3_AutoML_1_20220701_93140_model_569	525.416	387.353

Table 10. Comparison between H2O model and benchmark models

Model	MAE	MAPE
Fast Fourier Transform	1277.5	24.74%
Facebook Prophet	318.125	6.4%
Temporal Convolutional Network	447.375	8.63%
LSTM	706.75	13.57%
H2O	**350.5**	**6.36%**

Figures

Figure 1. Overall sentiment for Sydney from 2008 to 2019
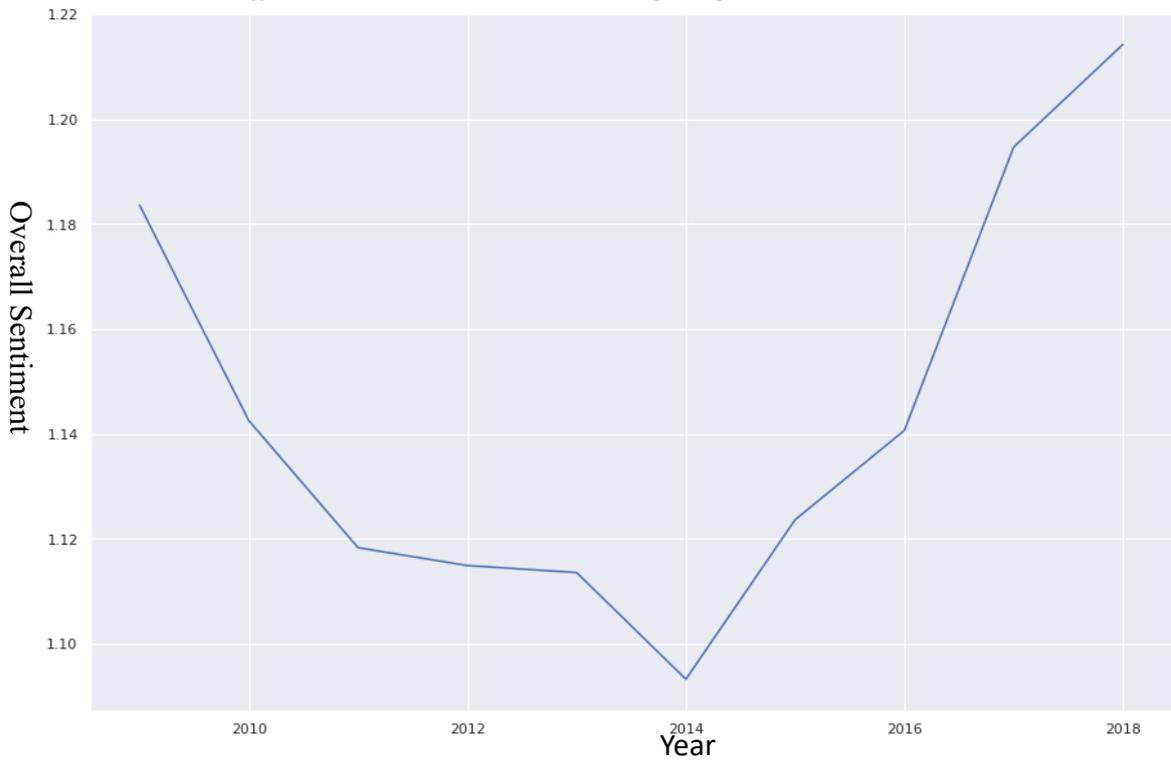
Figure 2. Variable importance for Sydney (overall sentiments)
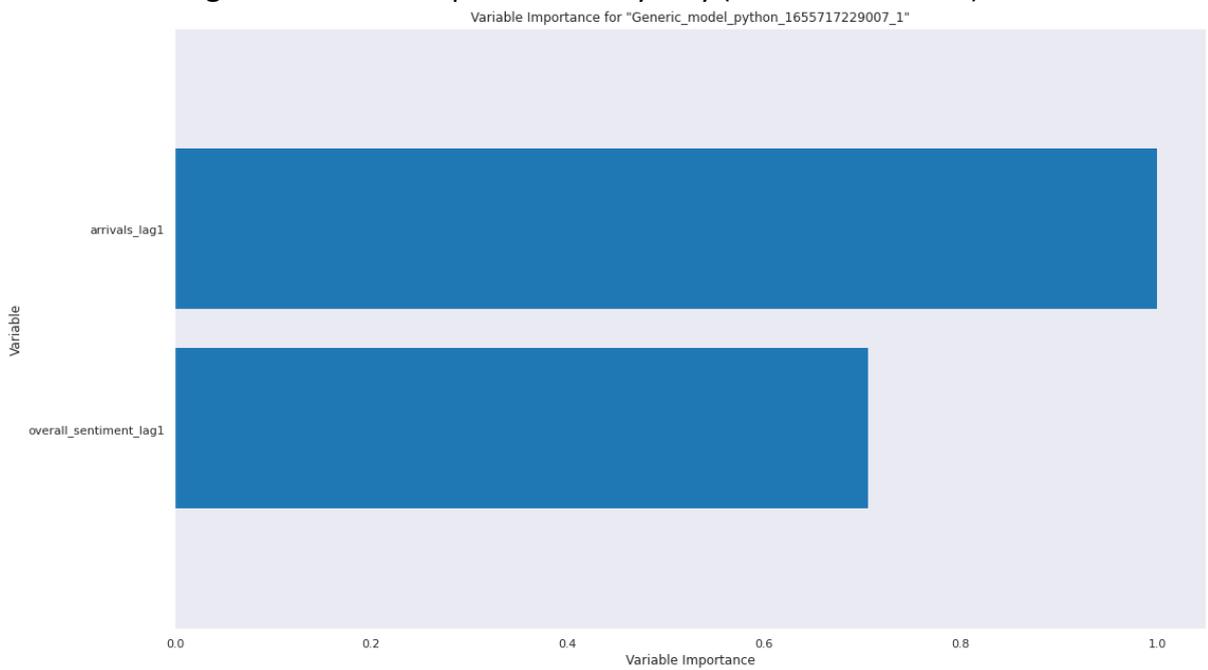
Figure 3. Partial Dependence plot for arrivals at lag (1) (Sydney)

Partial Dependence plot for "arrivals_lag1"
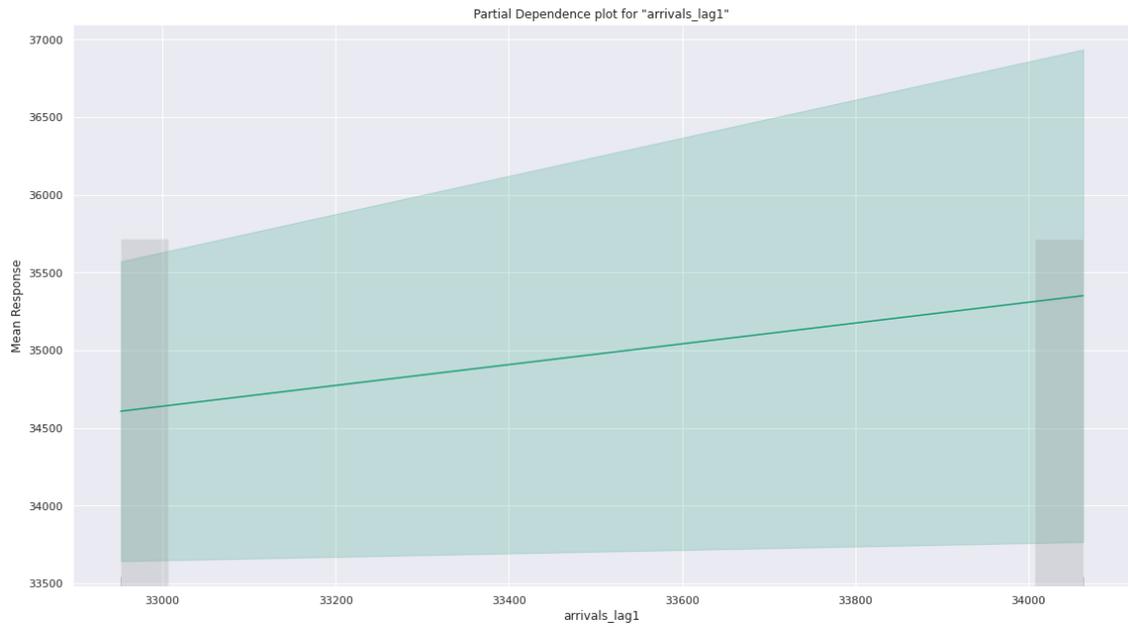
Figure 4. Partial Dependence plot for overall sentiment at lag (1) (Sydney)

Partial Dependence plot for "overall_sentiment_lag1"
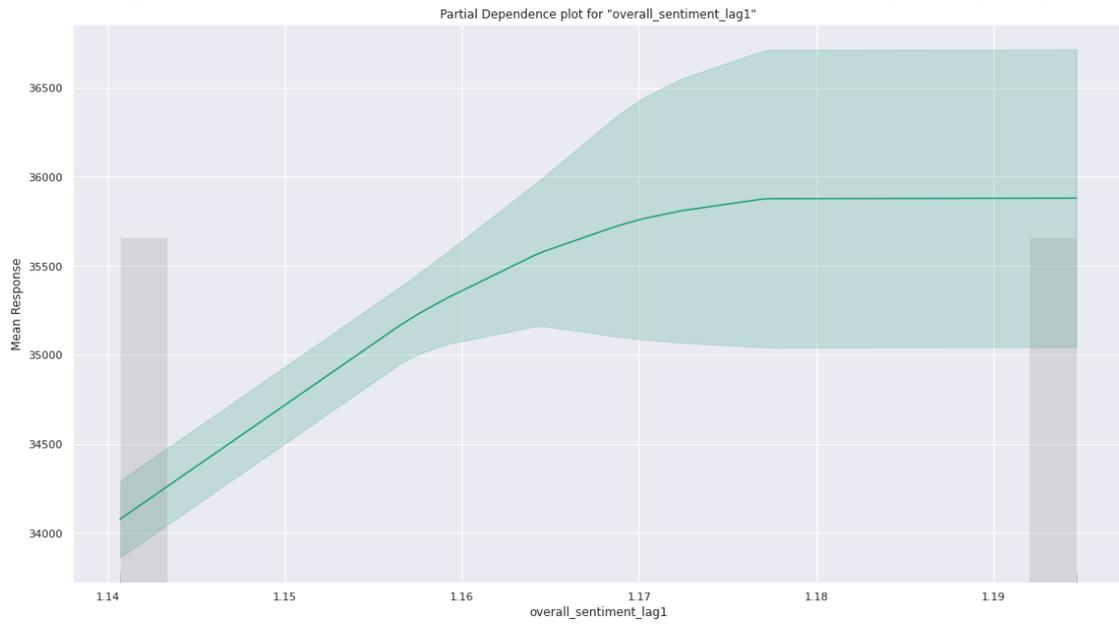
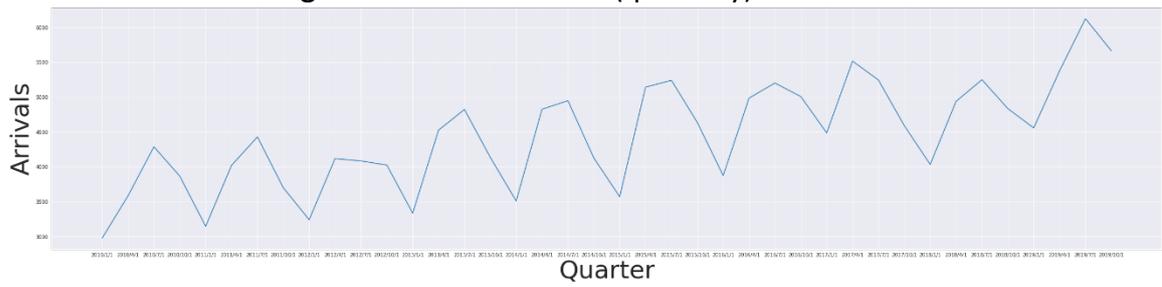**Figure 5.** Tourist arrivals (annually) to Sydney
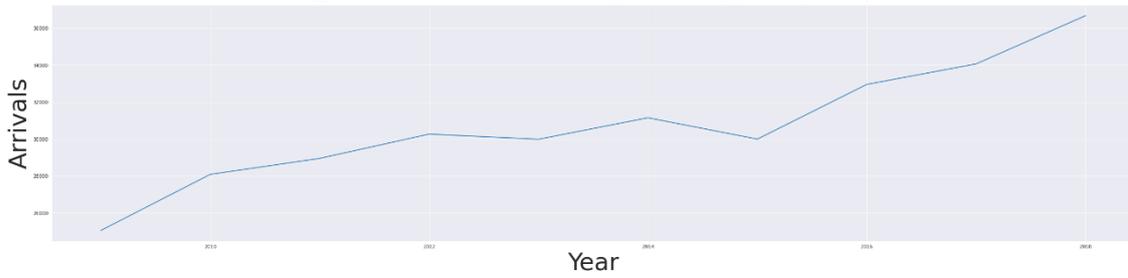
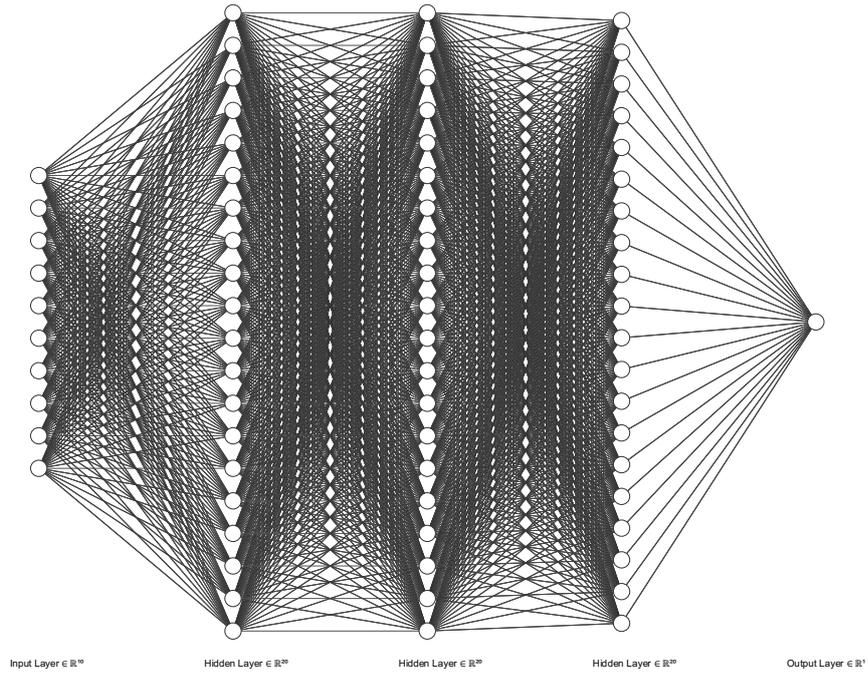**Figure 6.** Diagram for the leading model
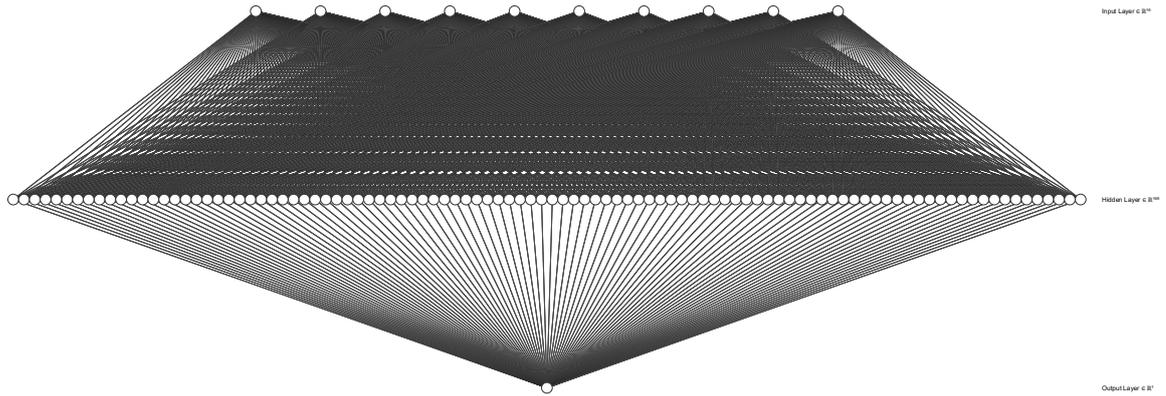
Figure 7. Comparison between the leading model and other baseline models
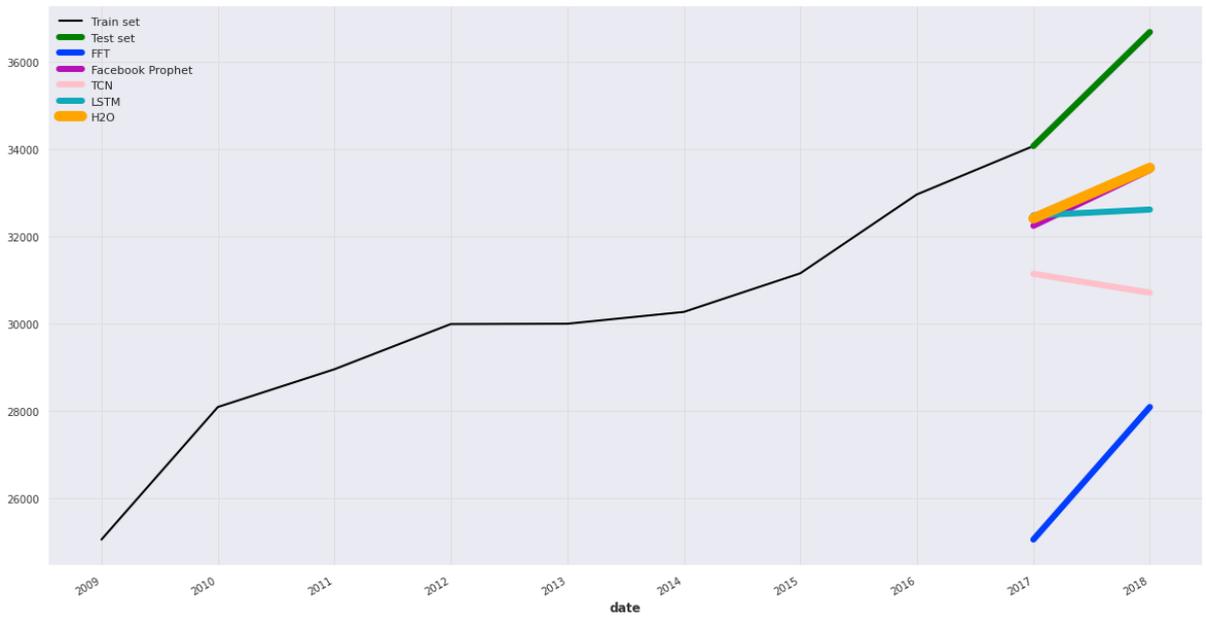
Figure 8. Variable importance
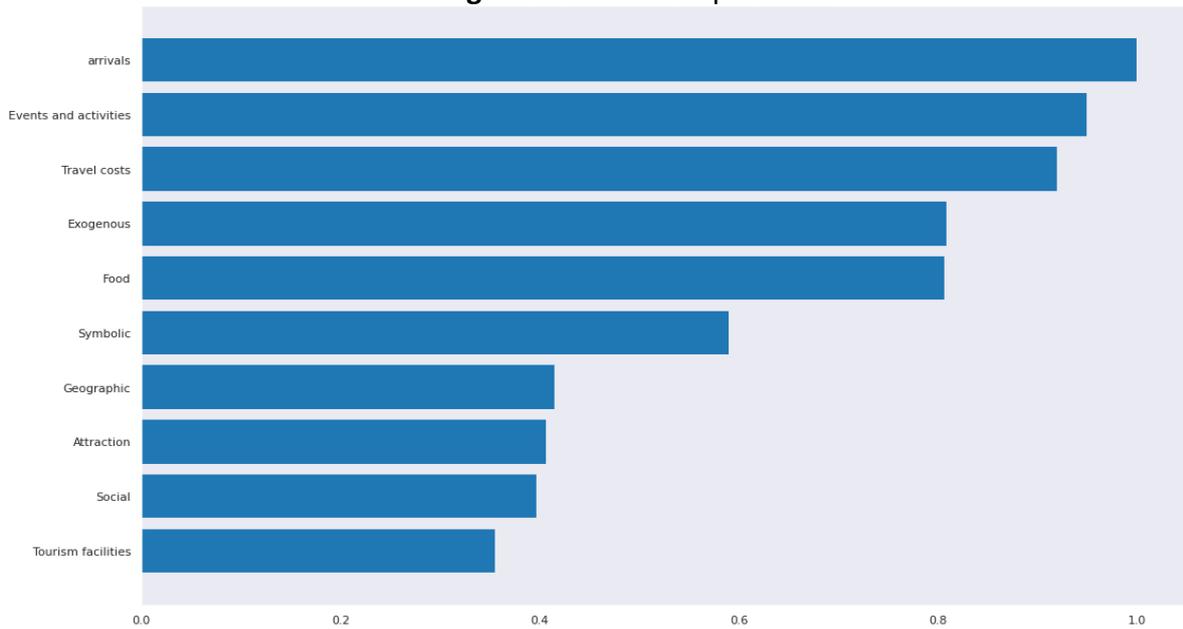
Figure 9. Partial dependence plot for the events and activities (Sydney)
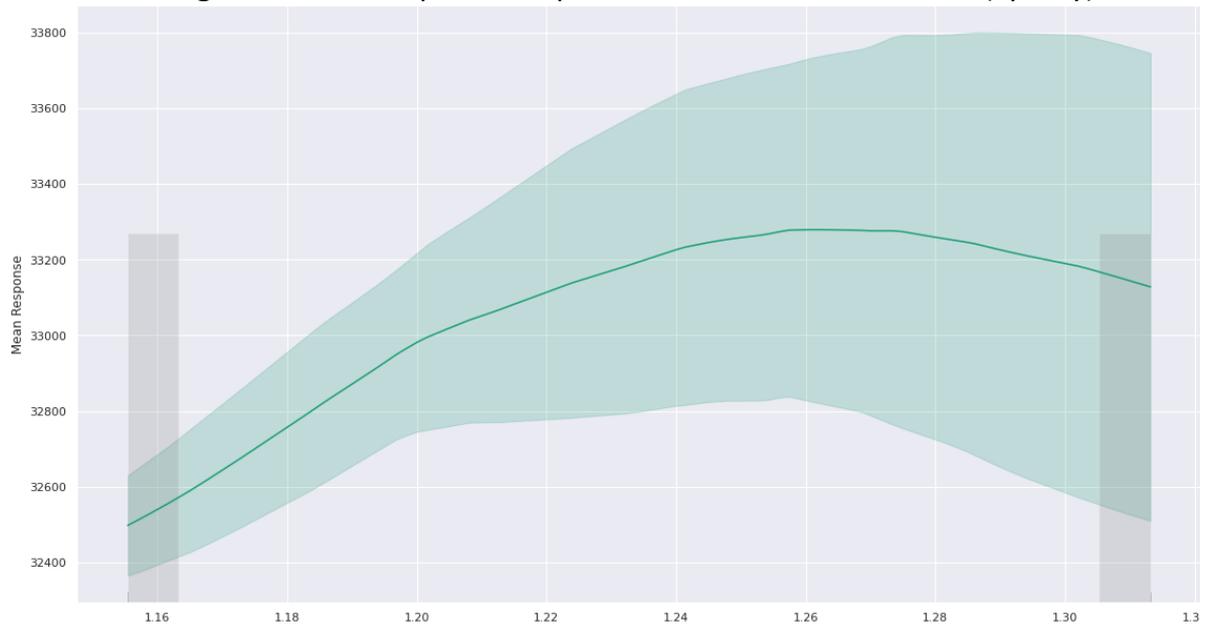
Figure 10. PDP for the sentiment of travel costs (Sydney)
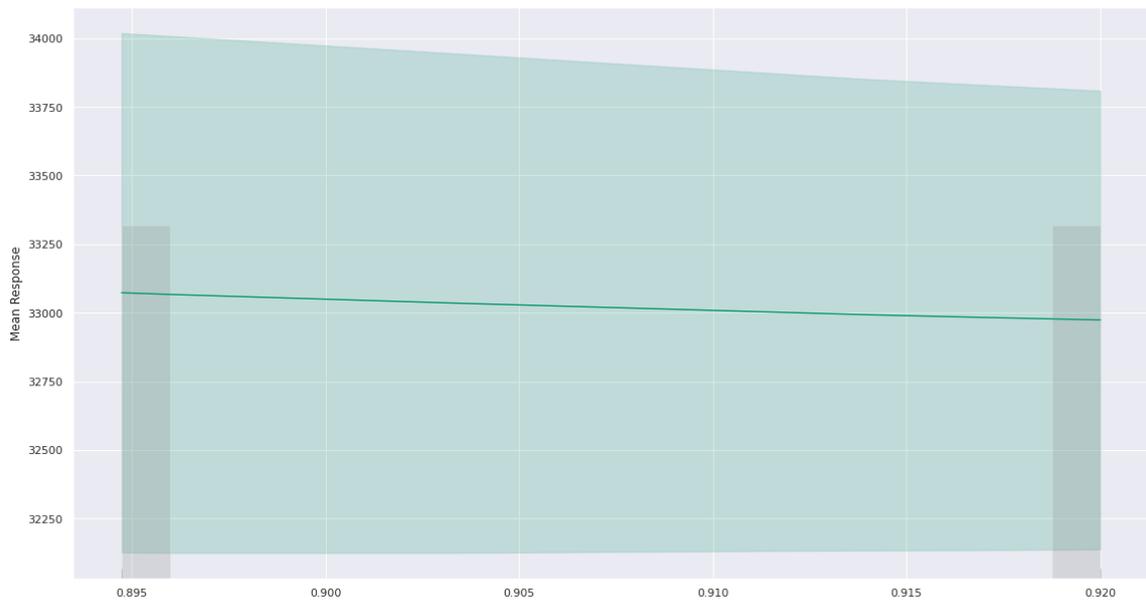
Figure 11. PDP for the sentiment of exogenous factors (Sydney)
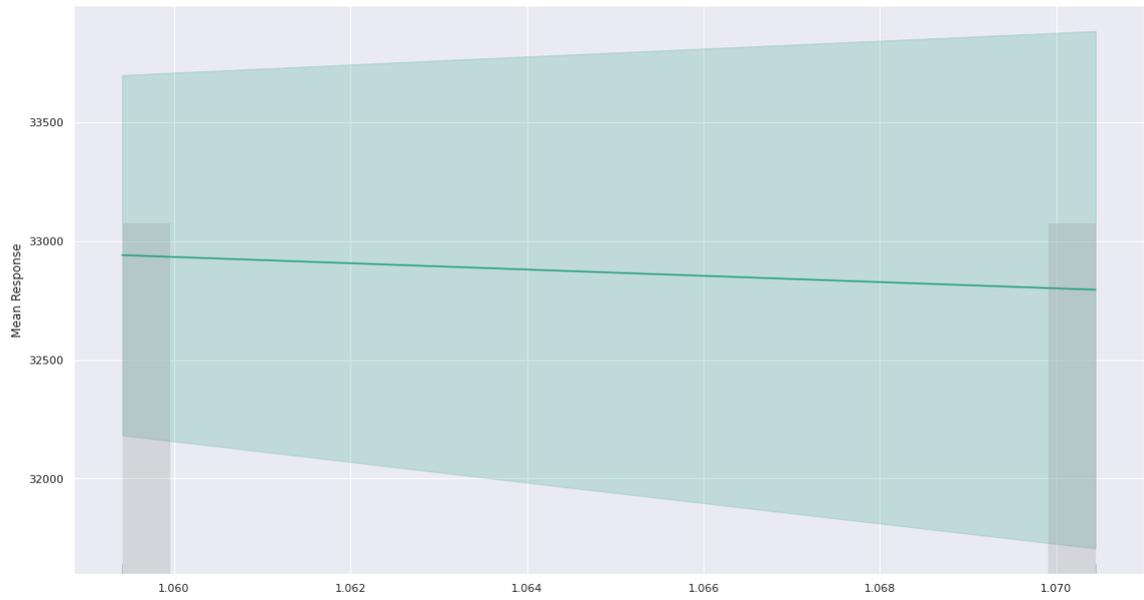
Figure 12. PDP for the sentiment of food (Sydney)
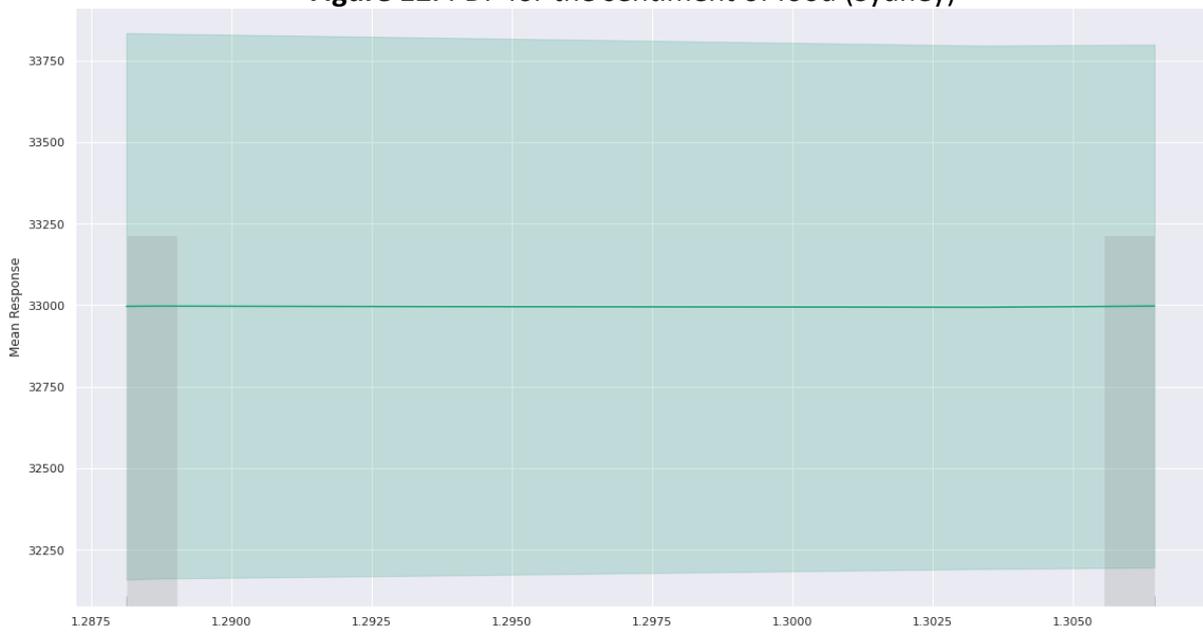
Figure 13. PDP for symbolic factors (Sydney)
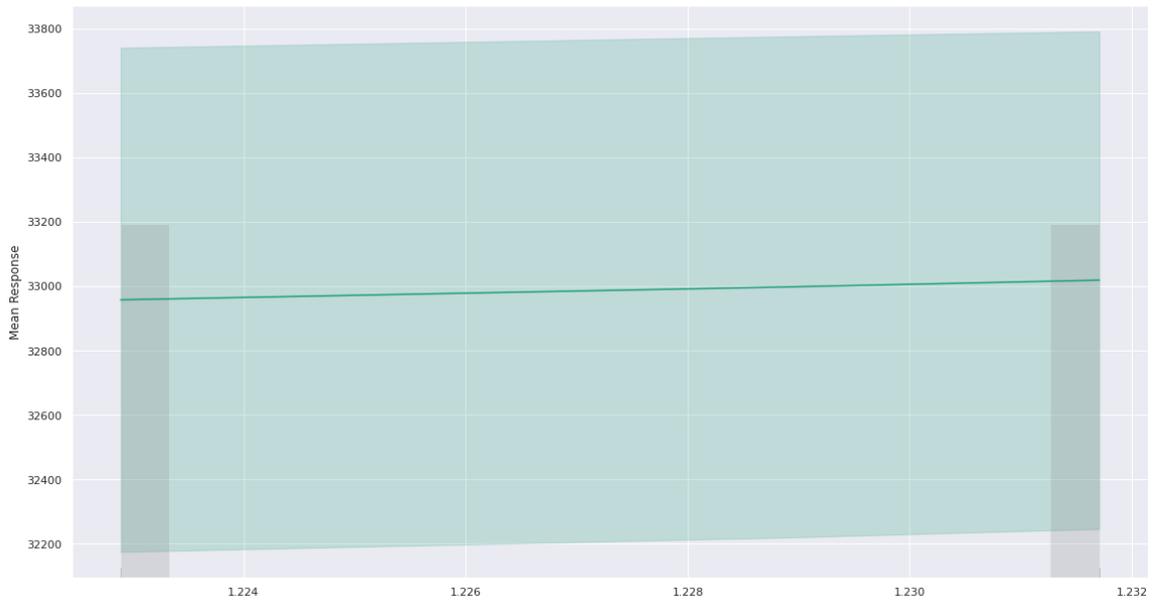
Figure 14. PDP for arrivals at lag (1) (Sydney)

Partial Dependence plot for "arrivals"
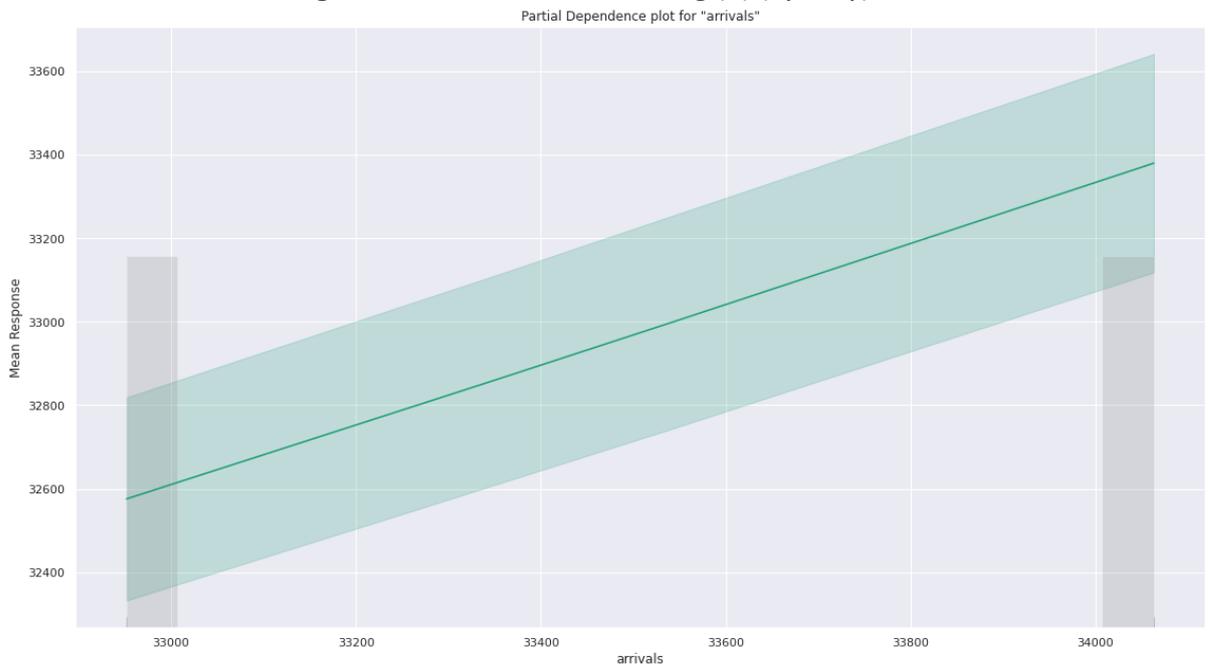
Figure 15. Overall sentiment for London from 2010 to 2019
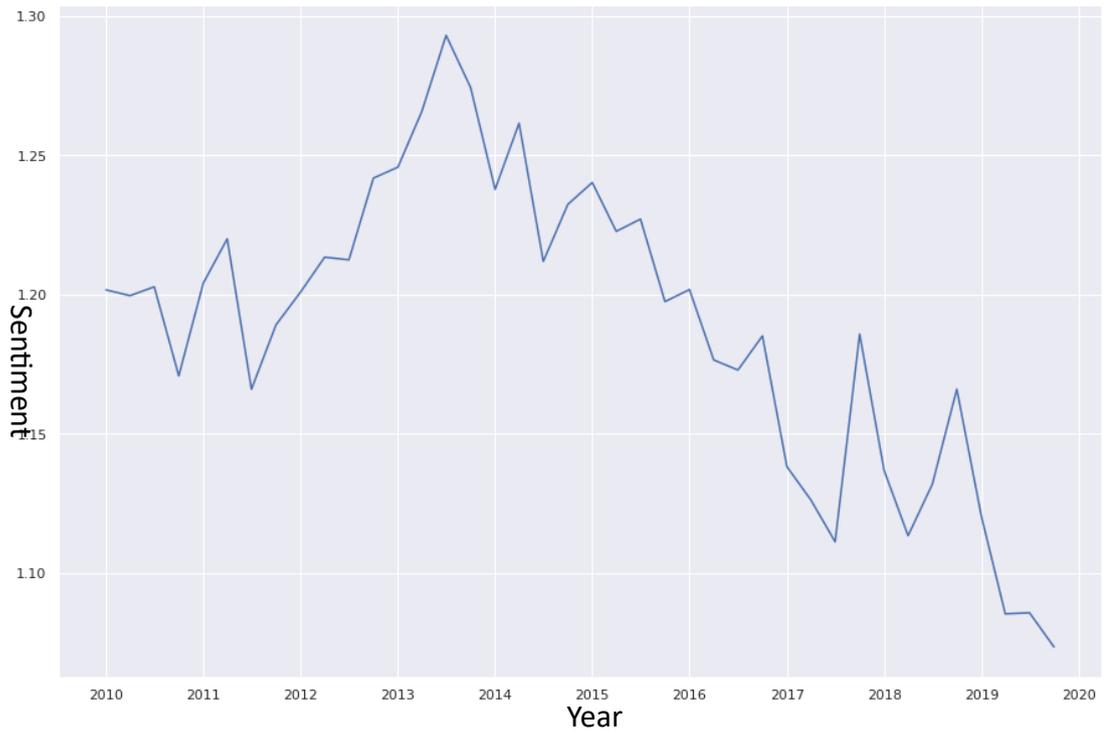
Figure 16. Partial Dependence plot for overall sentiment at lags (4)

Partial Dependence plot for "sentiment_lag4"
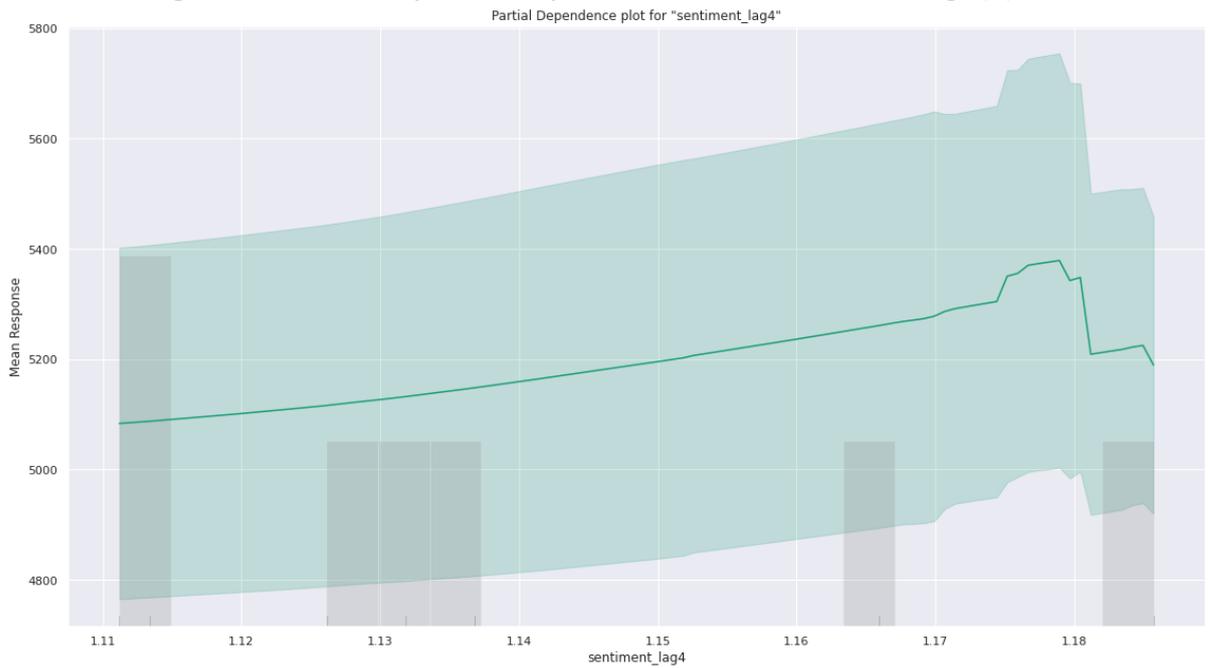
Figure 17. Partial Dependence plot for arrivals at lags (4)
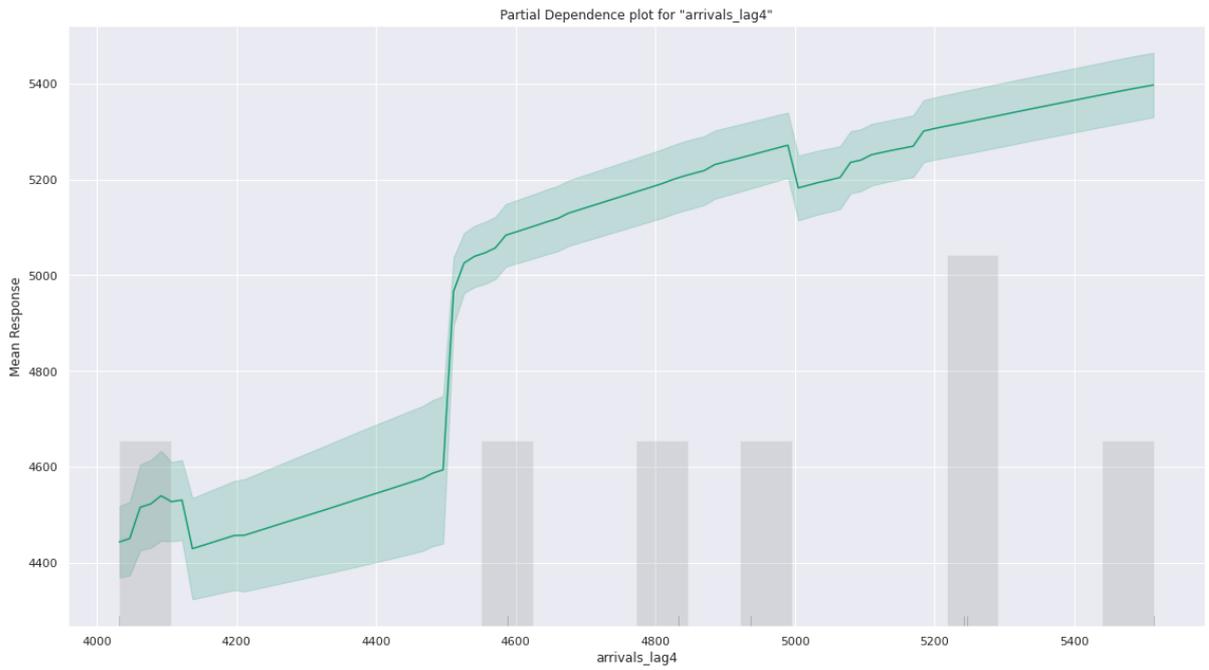
Figure 18. Tourist arrivals (quarterly) to London

Figure 19. Diagram for the leading neural network model

Input Layer ∈ ℝ¹⁰   Hidden Layer ∈ ℝ²⁰   Hidden Layer ∈ ℝ²⁰   Hidden Layer ∈ ℝ²⁰   Output Layer ∈ ℝ¹

Figure 20. Comparison between the leading model and baseline models
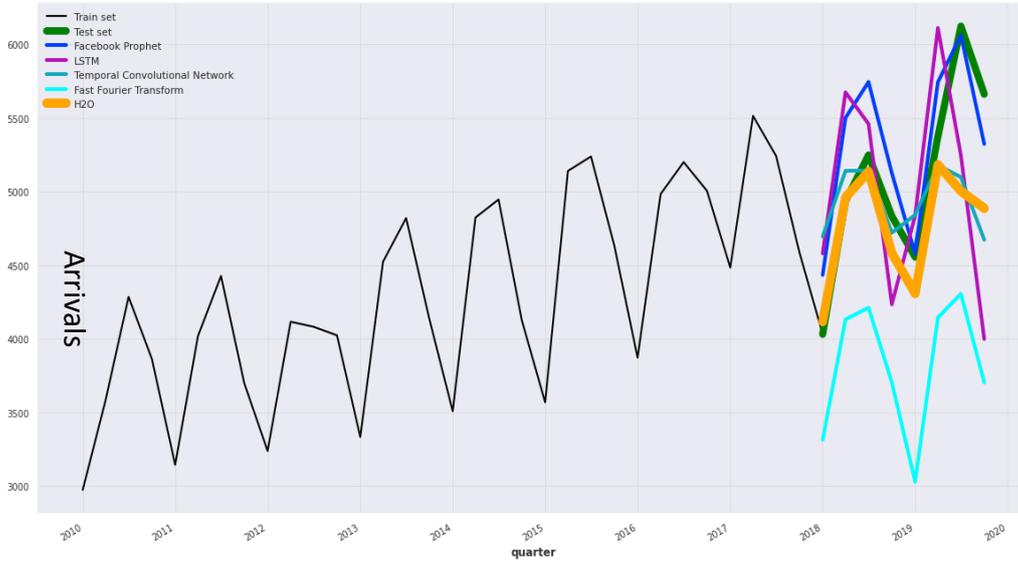
Figure 21. Variable importance

Variable Importance for "Generic_model_python_1657635874441_1"
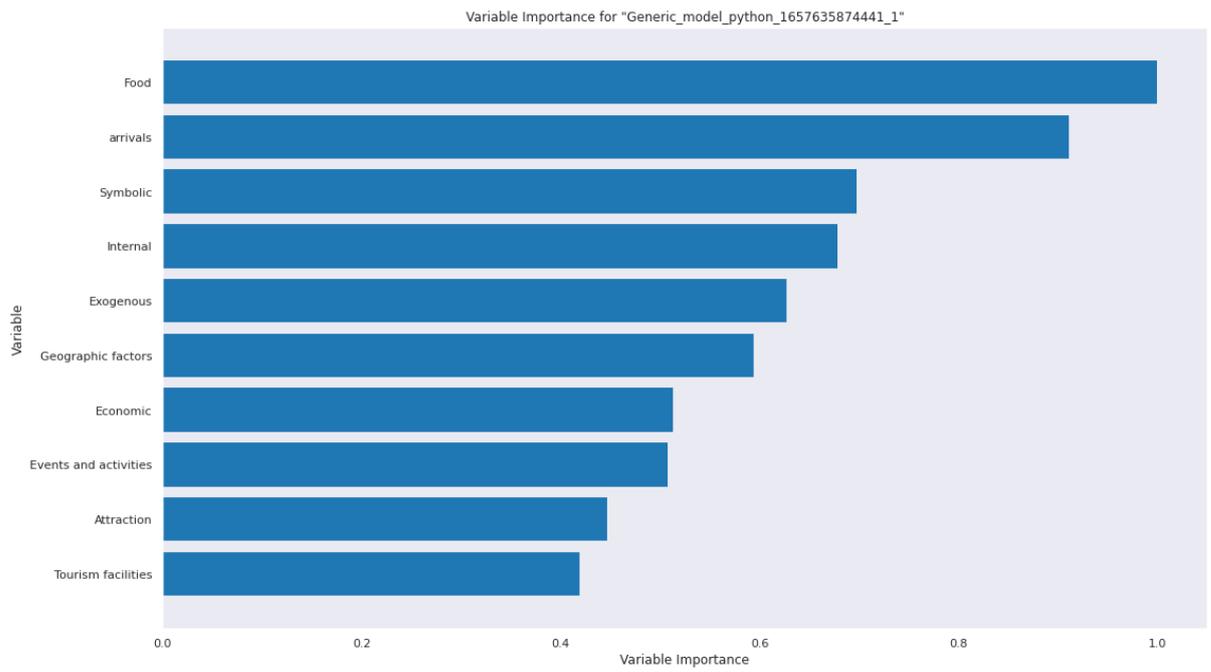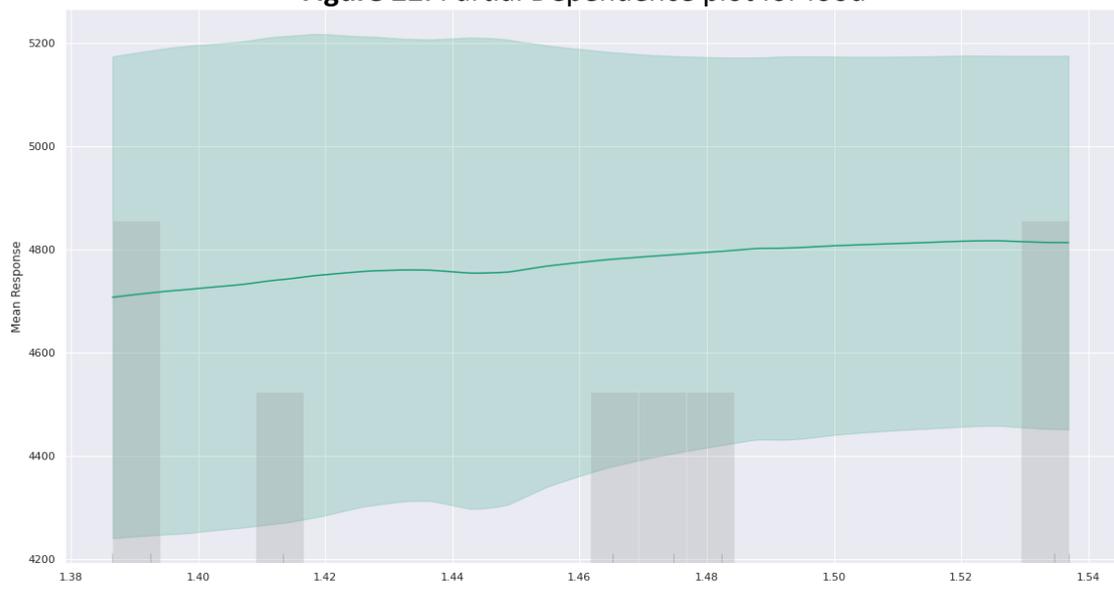
Figure 22. Partial Dependence plot for food

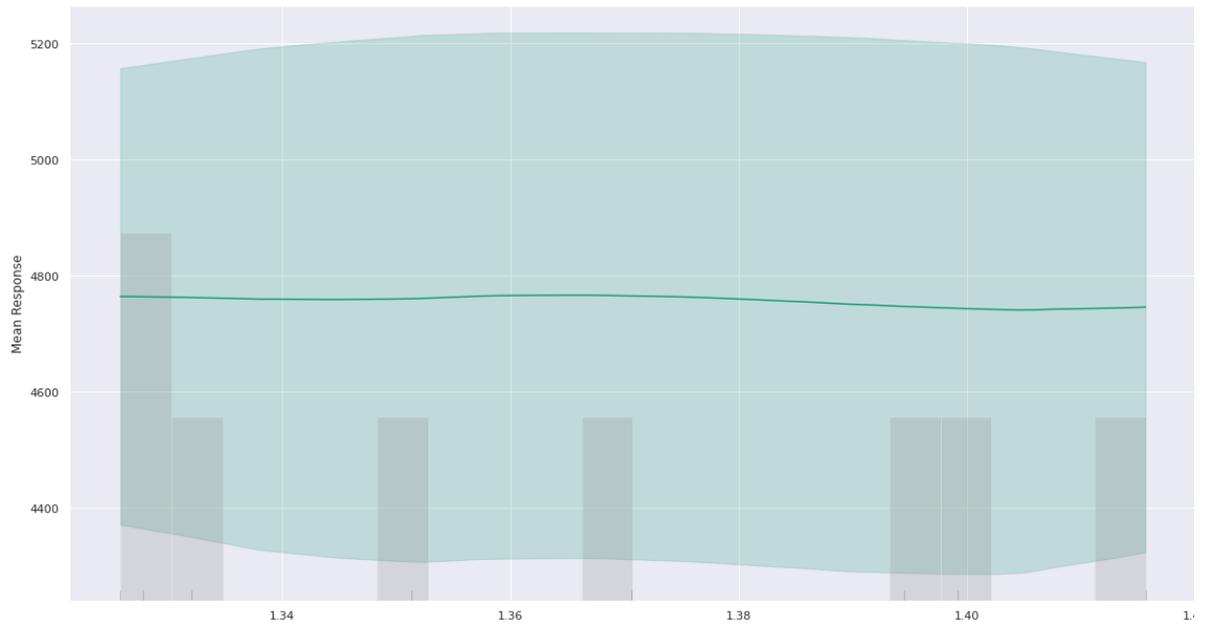**Figure 23.** Partial Dependence plot for symbolic factors

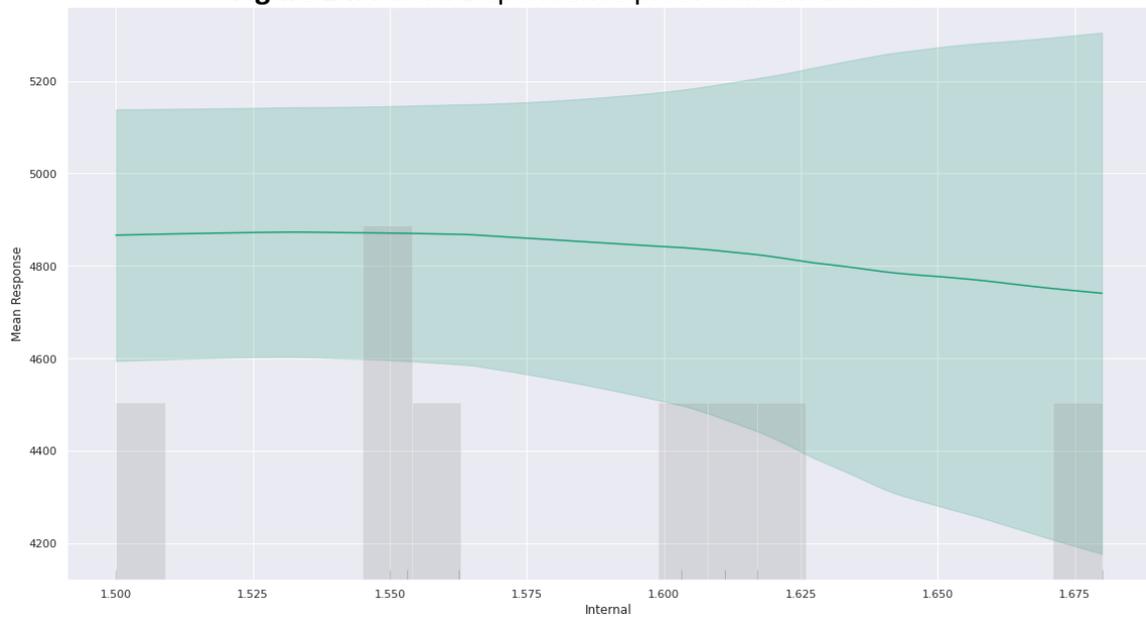**Figure 24.** Partial Dependence plot for internal factors

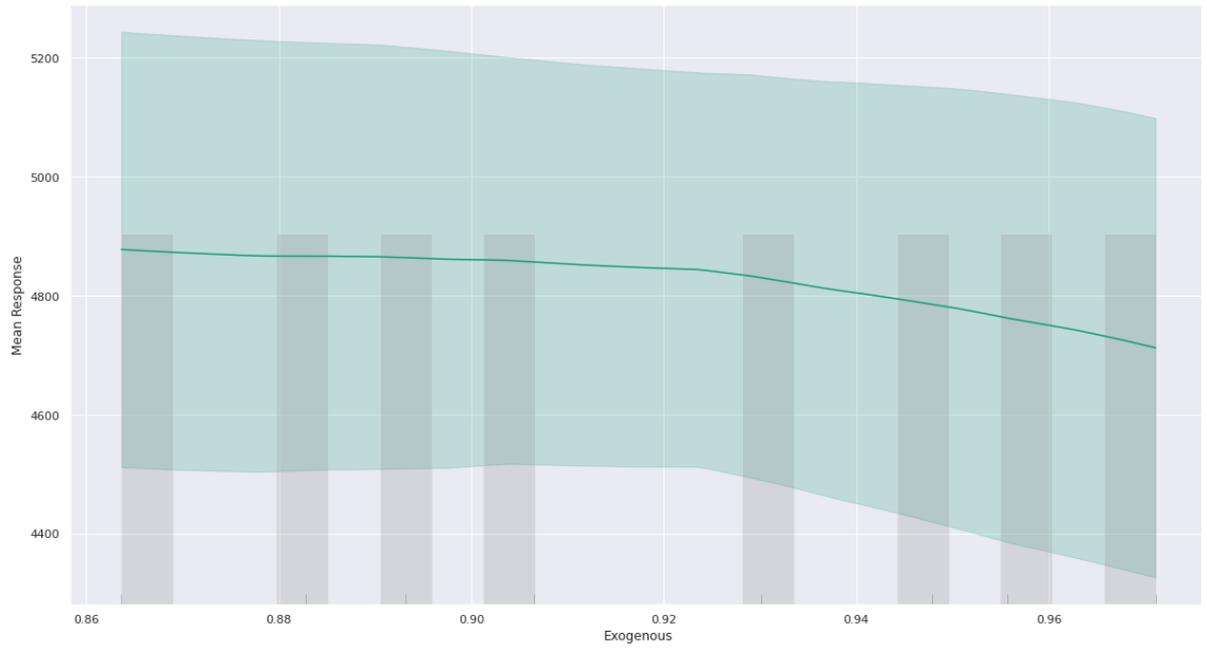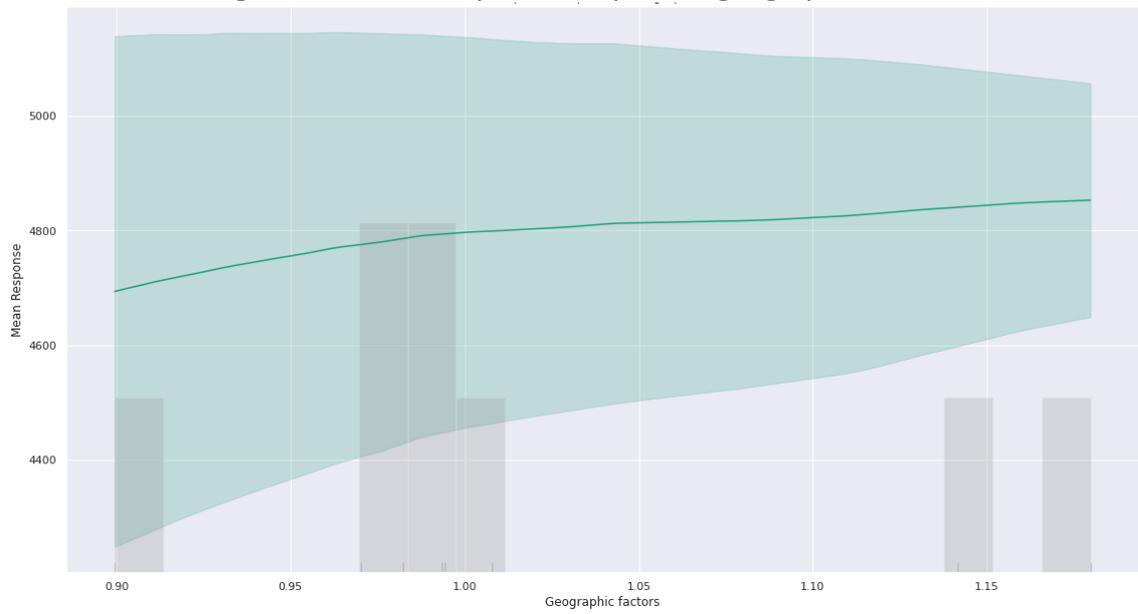**Figure 25.** Partial Dependence plot for exogenous factors

Figure 26. Partial Dependence plot for geographic factors

Abbreviations

ADLM: Autoregressive Distributed Lag Model
ANN: Artificial Neural Network
API: Application Programming Interface
ARIMA: Autoregressive Integrated Moving Average
BERT: Bidirectional Encoder Representation from Transformers
BERTopic: a topic modelling algorithm utilising advanced text embedding.
COVID-19: Coronavirus disease
CPI: Consumer Price Index
DI: Destination Image
DMO: Destination Marketing Organisations
DNN: Deep Neural Network
FFT: Fast Fourier Transform
GBM: Gradient Boosting Machine
GBRT: Gradient Boosted Regression Tree
GDP: Gross Domestic Product
GLM: Generalise Linear Models
HDBSCAN: Hierarchical Density-Based Clustering
IPI: Industrial Production Index
LDA: Latent Dirichlet Allocation
LOV: List of Values
LSTM: Long Short-Term Memory Networks
MAE: Mean Absolute Error
MAPE: Mean Absolute Percentage Error
MECE: Mutually Exclusive Collectively Exhaustive
NLP: Natural Language Processing
OECD: Organisation for Economic Co-operation and Development
PDP: Partial Dependence Plot
RF: Random Forest
RMSE: Root Mean Squared Error
RP: Relative Price
RVS: Rokeach Value Survey
SP: Substitute Price
TCI: Tourism Climate Index
TCL: Travel Career Ladder
TCN: Temporal Convolutional Network
TF-IDF: Term Frequency-Inverse Document Frequency
UGC: User-generated content
UMAP: Uniform Manifold Approximation and Projection
VAL: Value and Lifestyles
VAR: Vector Regressive
WOM: Words of mouth
WHSs: World Heritage Sites
XGBoost: Extreme Gradient Boosting