

Durham E-Theses

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

LI, YULEI

How to cite:

LI, YULEI (2023) *Insights from Tweets: Analysing Destination Topics and Sentiments, and Predicting Tourist Arrivals*, Durham theses, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/15088/>

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

Appendix

Tables

Table 1. Categories with topics and keywords (Sydney)

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

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

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

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
*	2197.47	1593.71
DeepLearning_grid_1_AutoML_2_20220525_85253_model_114		
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	<i>expensive; money; pay; cash</i>
	<i>deals;groupon; off at; daily</i>
Exogenous	<i>jobs; job; manager; united</i>
	<i>mayor; boris; brexit; labour</i>
	<i>nhs; health; nurse; hospital</i>
	<i>iran; syria; pakistan; palestine</i>
	<i>fire; fireworks; firefighters; grenfell</i>
	<i>paris; london paris; to paris; paris london</i>
	<i>muslim; muslims; islamic; islam</i>
	<i>gay; women; lgbt; womens</i>
	<i>buhari; nigeria; nigerian; president</i>

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

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

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

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

DeepLearning_grid_3_AutoML_1_20220701_93140_mode 1_28	509.16	389.02
XGBoost_grid_1_AutoML_1_20220701_93140_model_9	519.92	406.47
XGBoost_grid_1_AutoML_1_20220701_93140_model_4	524.84	394.71
DeepLearning_grid_3_AutoML_1_20220701_93140_mode 1_569	525.41	387.35

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

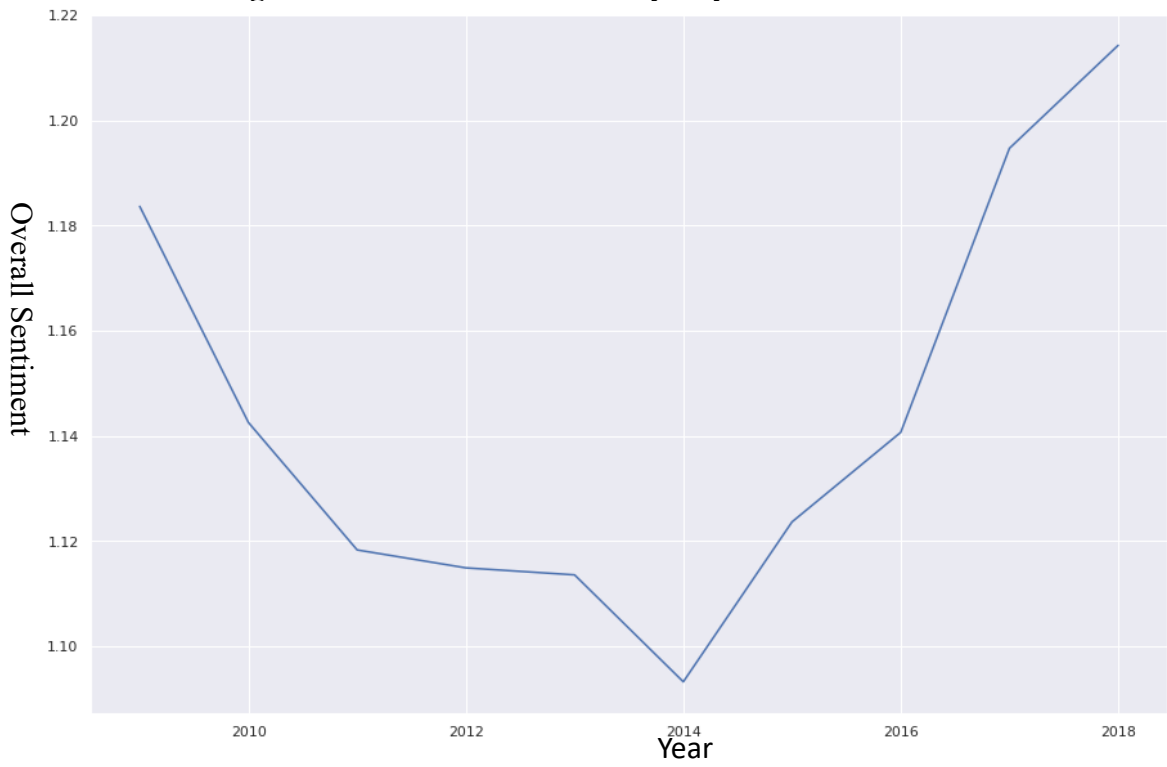


Figure 2. Variable importance for Sydney (overall sentiments)

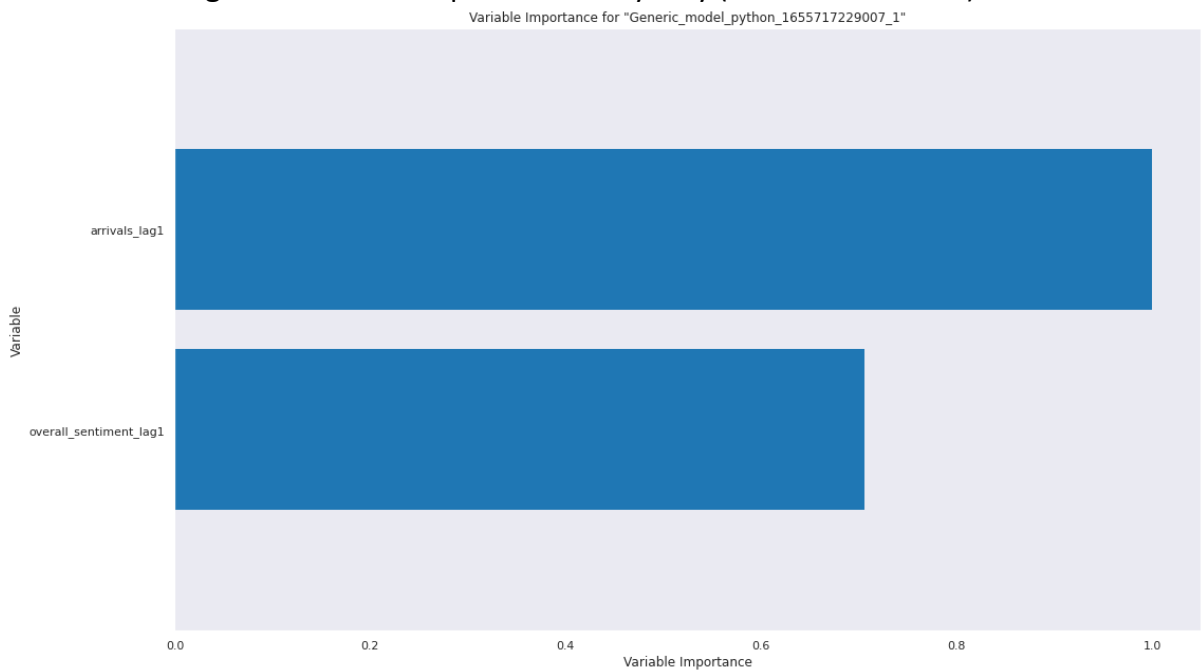


Figure 3. Partial Dependence plot for arrivals at lag (1) (Sydney)

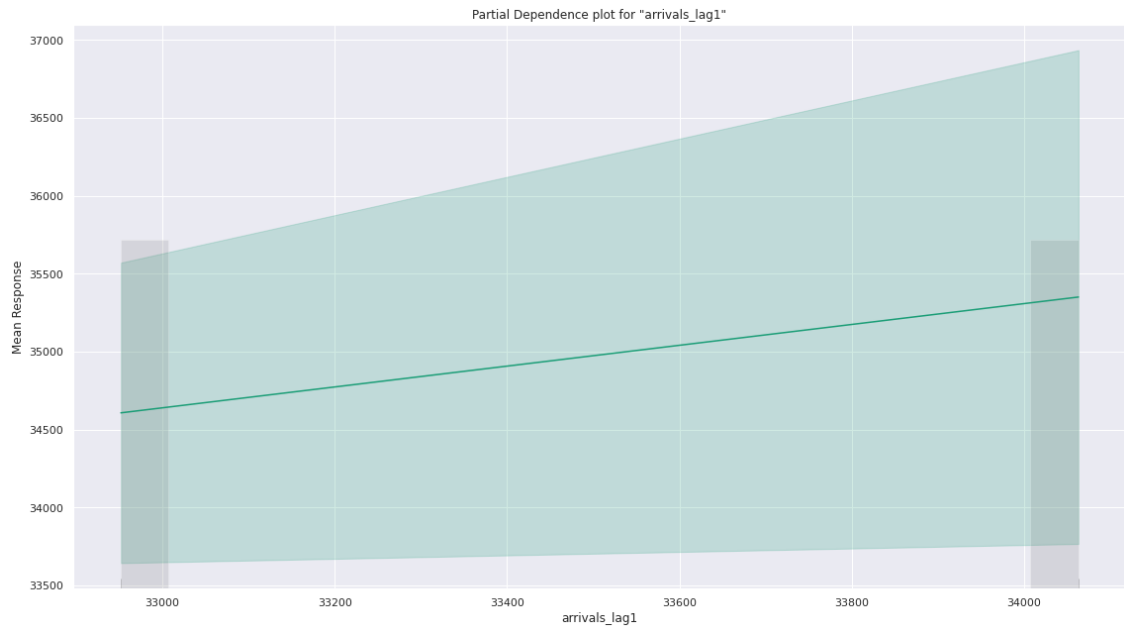


Figure 4. Partial Dependence plot for overall sentiment at lag (1) (Sydney)

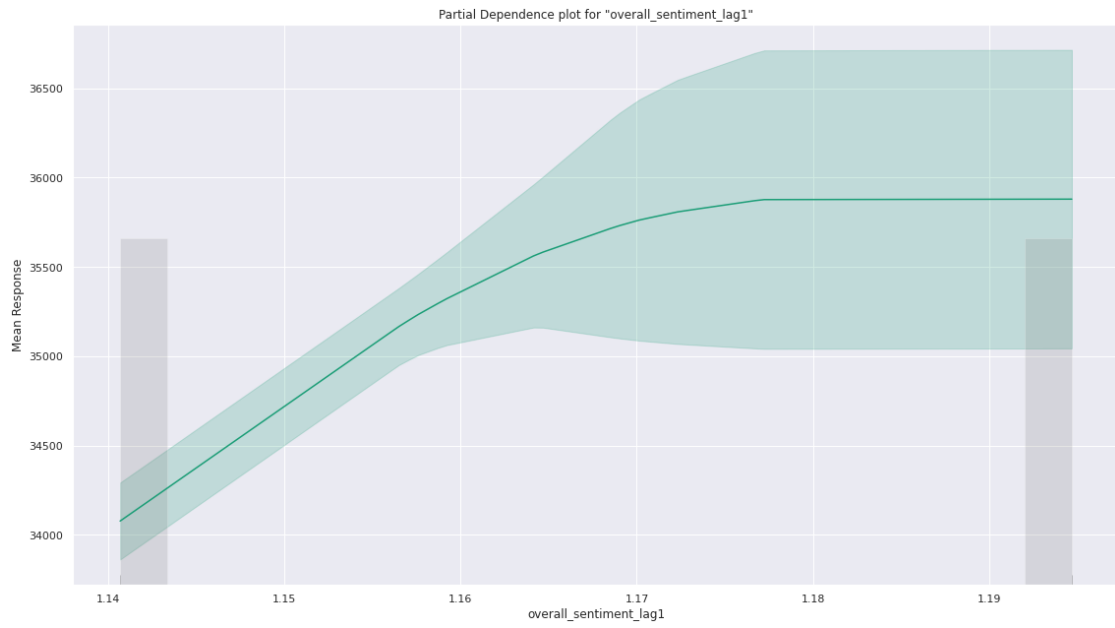


Figure 5. Tourist arrivals (annually) to Sydney

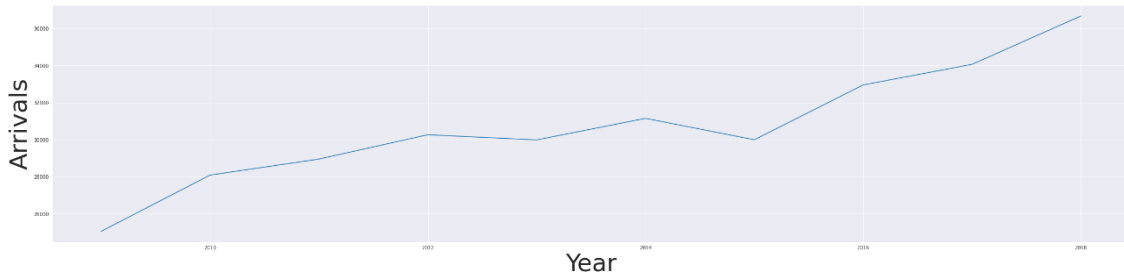


Figure 6. Diagram for the leading model

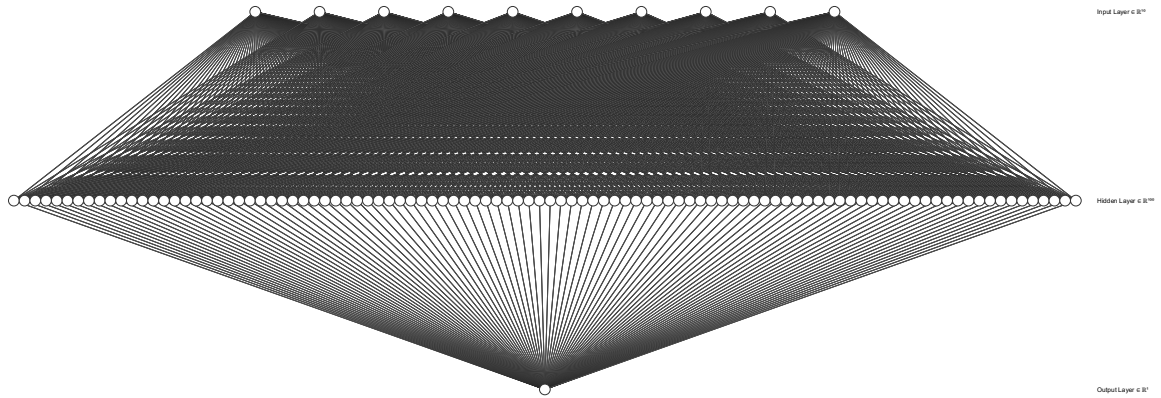


Figure 7. Comparison between the leading model and other baseline models

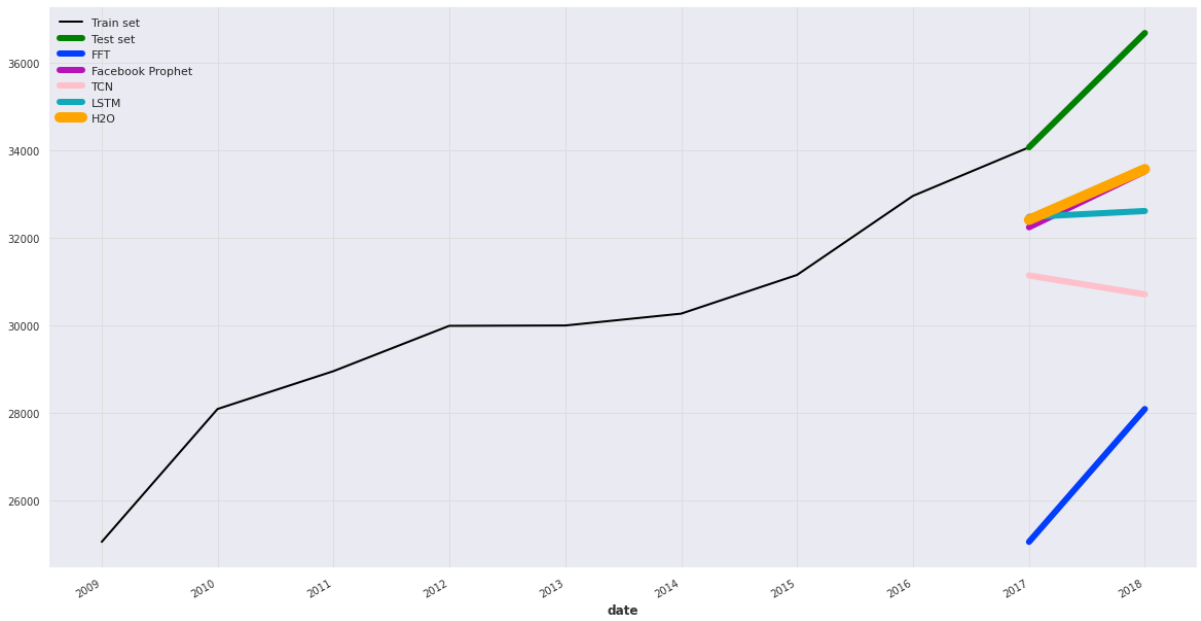


Figure 8. Variable importance

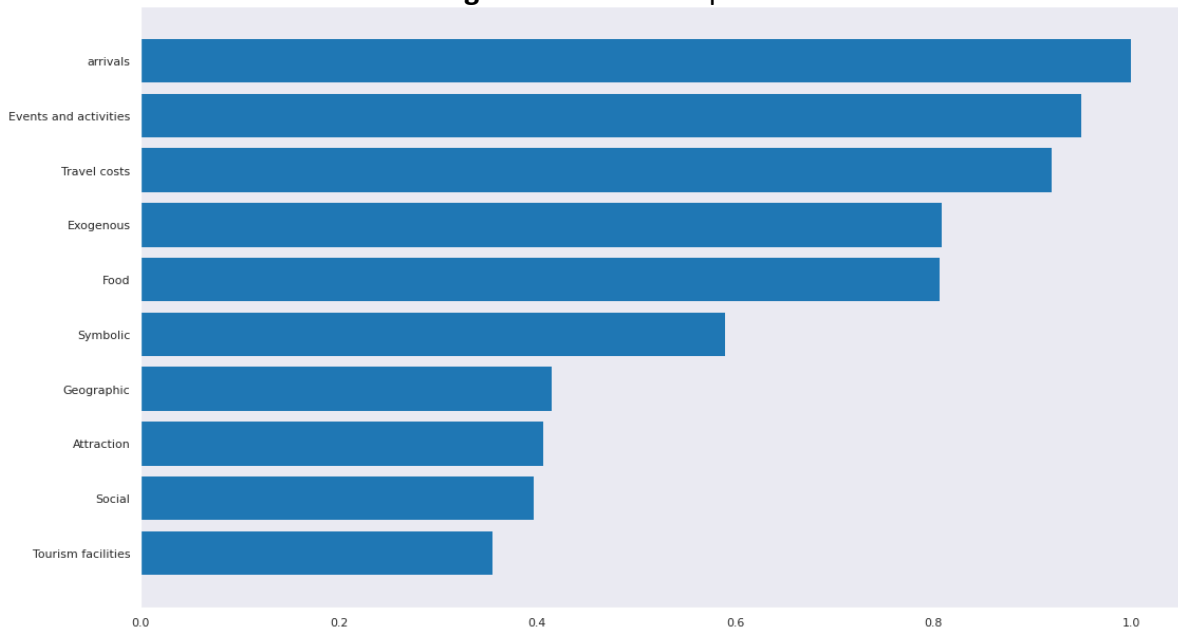


Figure 9. Partial dependence plot for the events and activities (Sydney)

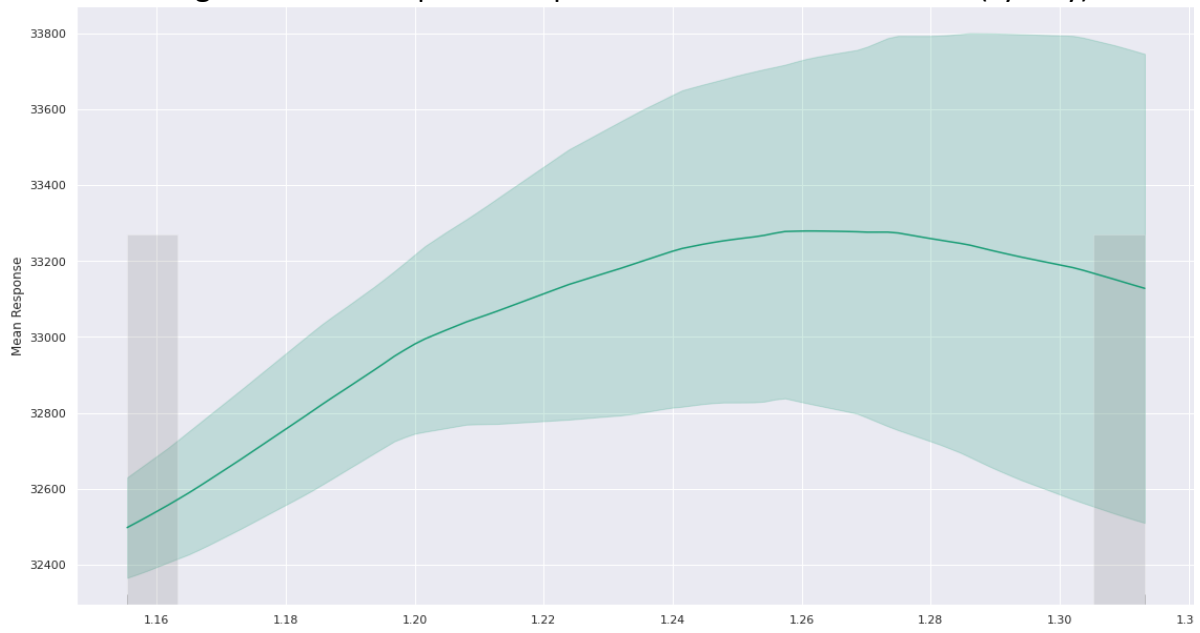


Figure 10. PDP for the sentiment of travel costs (Sydney)

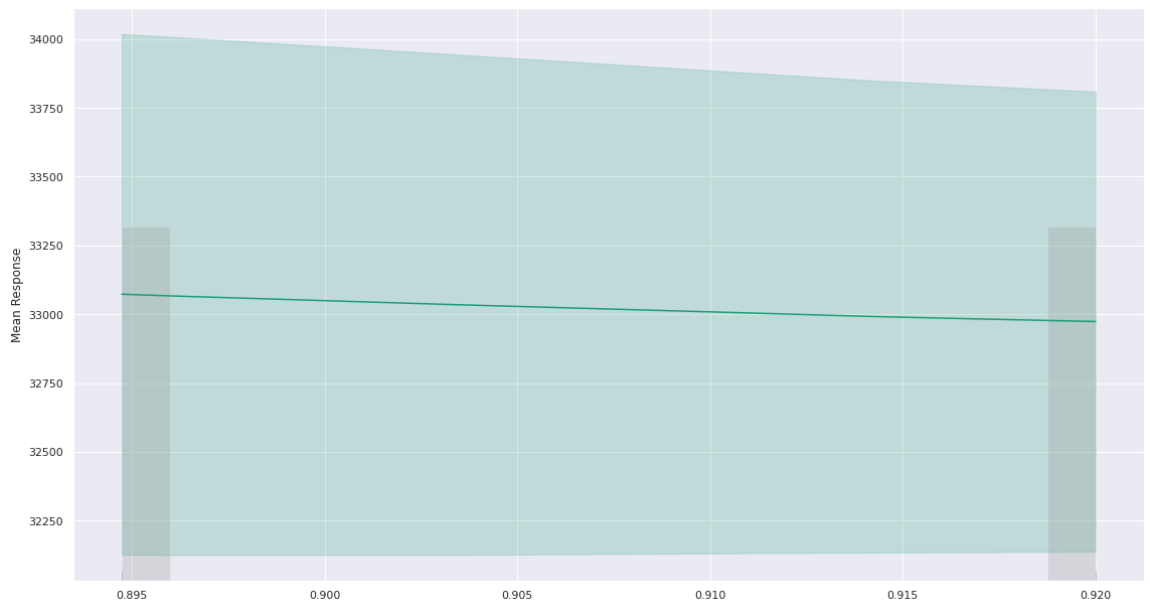


Figure 11. PDP for the sentiment of exogenous factors (Sydney)

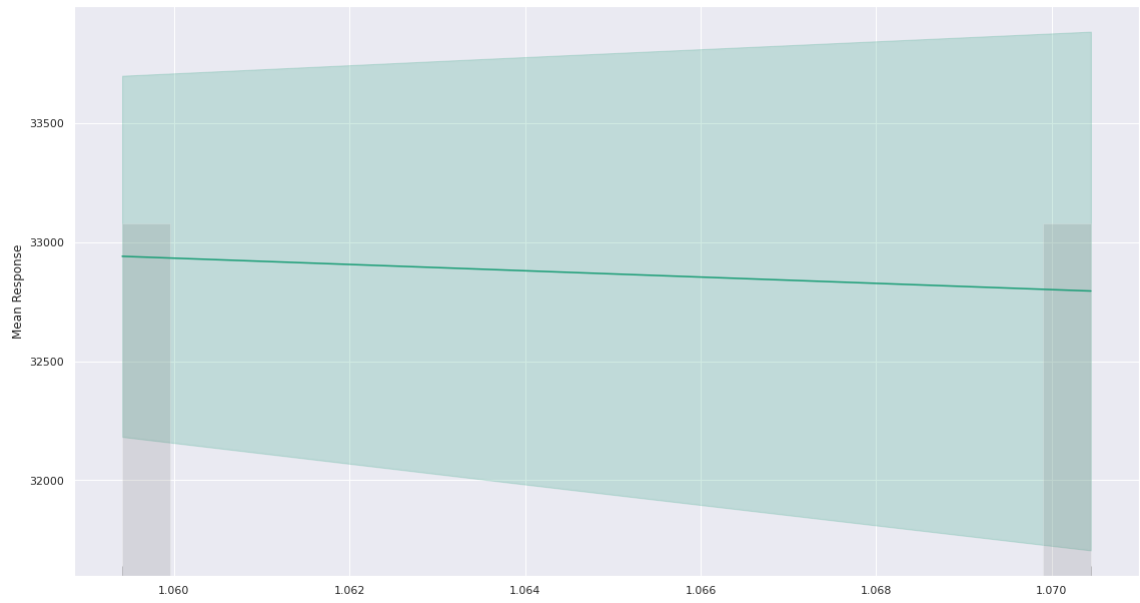


Figure 12. PDP for the sentiment of food (Sydney)

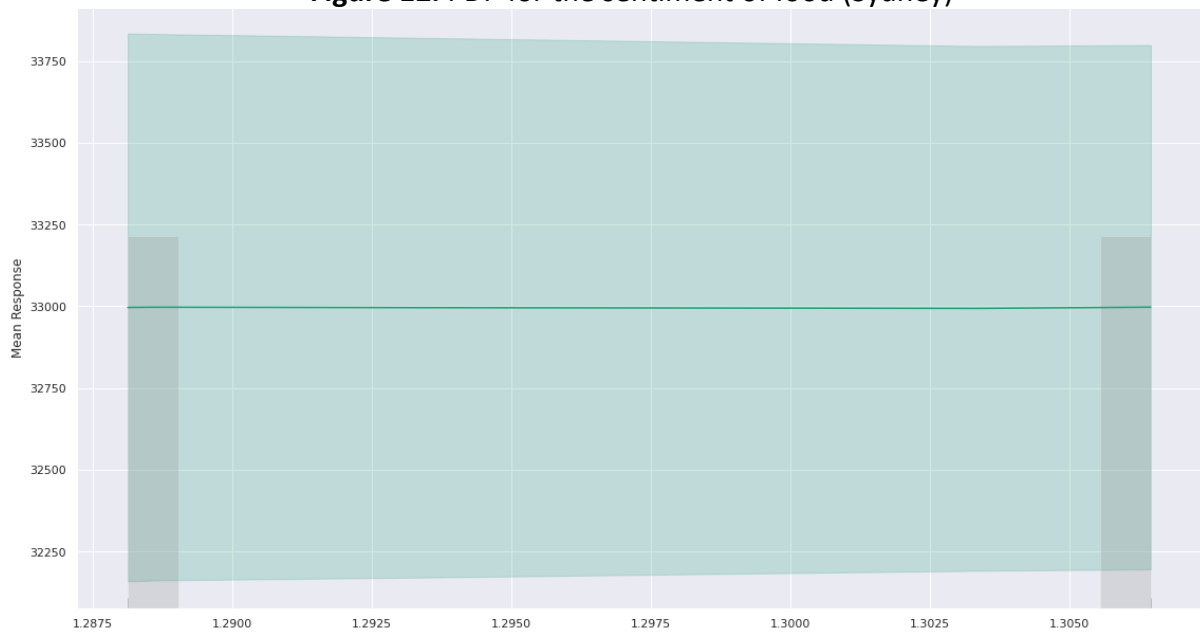


Figure 13. PDP for symbolic factors (Sydney)

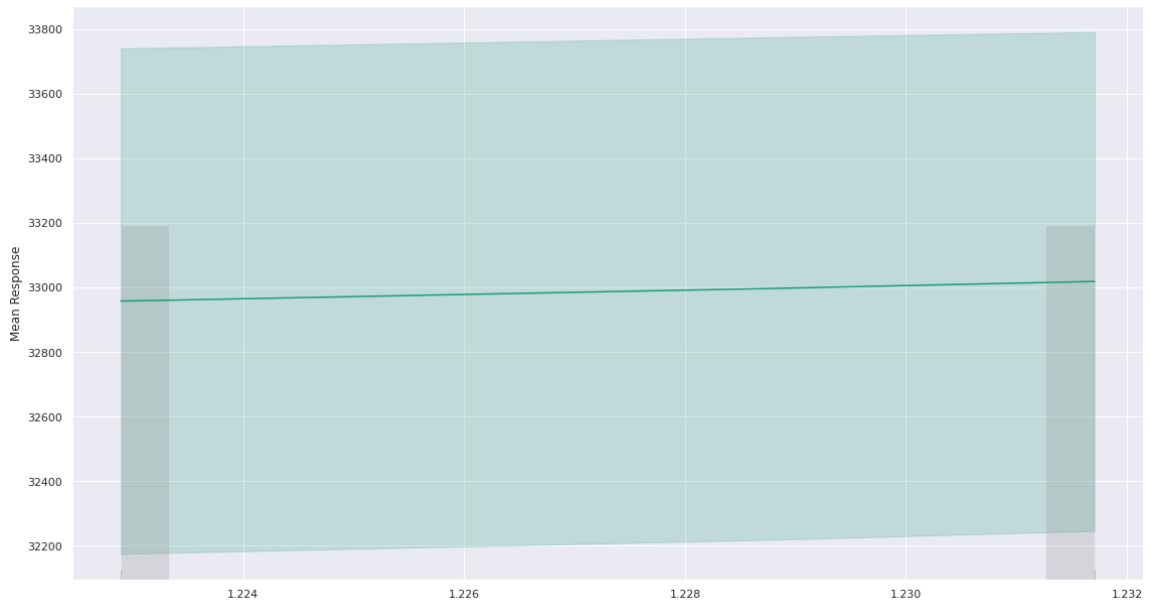


Figure 14. PDP for arrivals at lag (1) (Sydney)

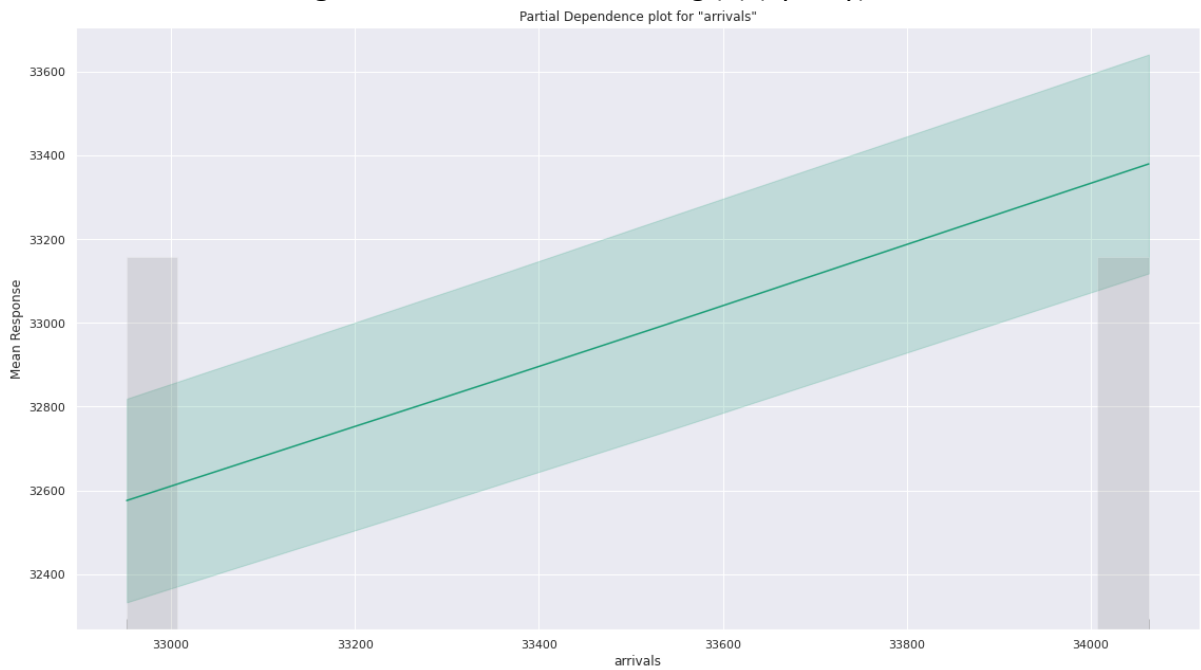


Figure 15. Overall sentiment for London from 2010 to 2019

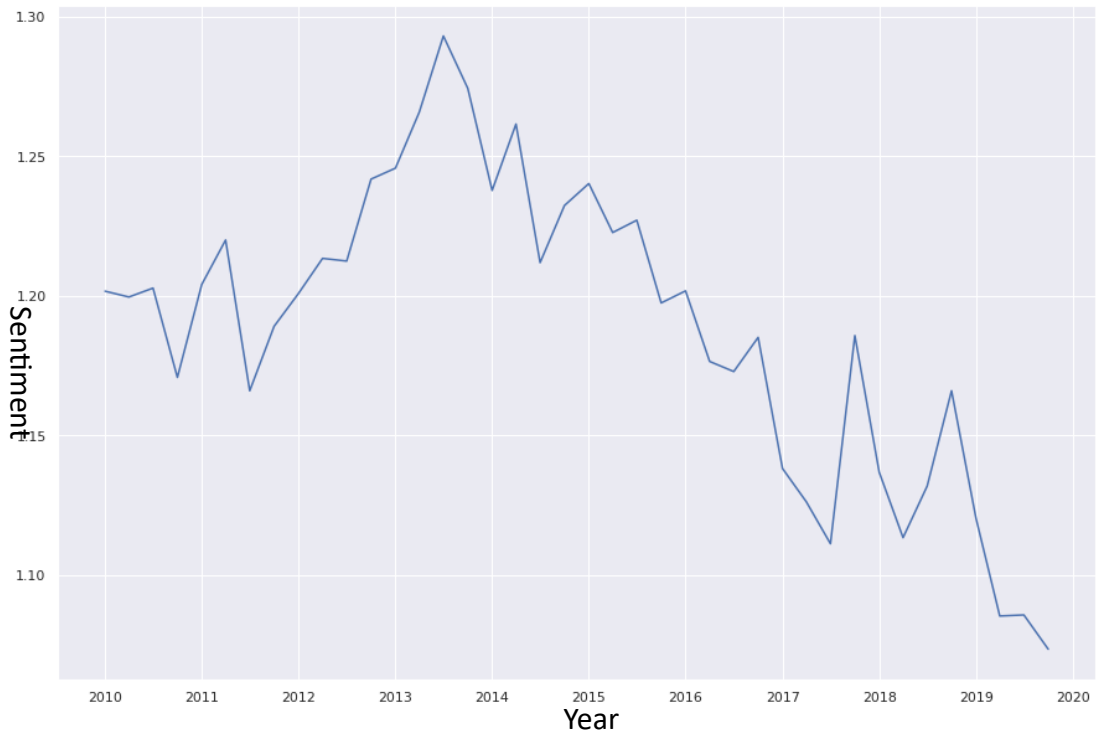


Figure 16. Partial Dependence plot for overall sentiment at lags (4)

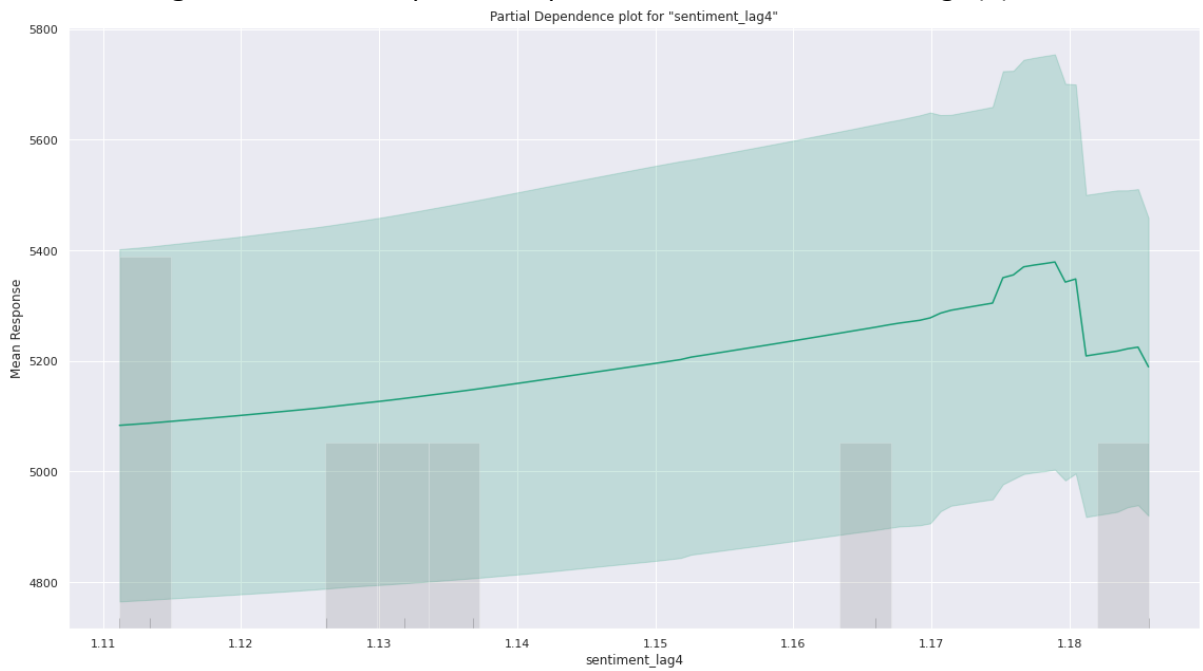


Figure 17. Partial Dependence plot for arrivals at lags (4)

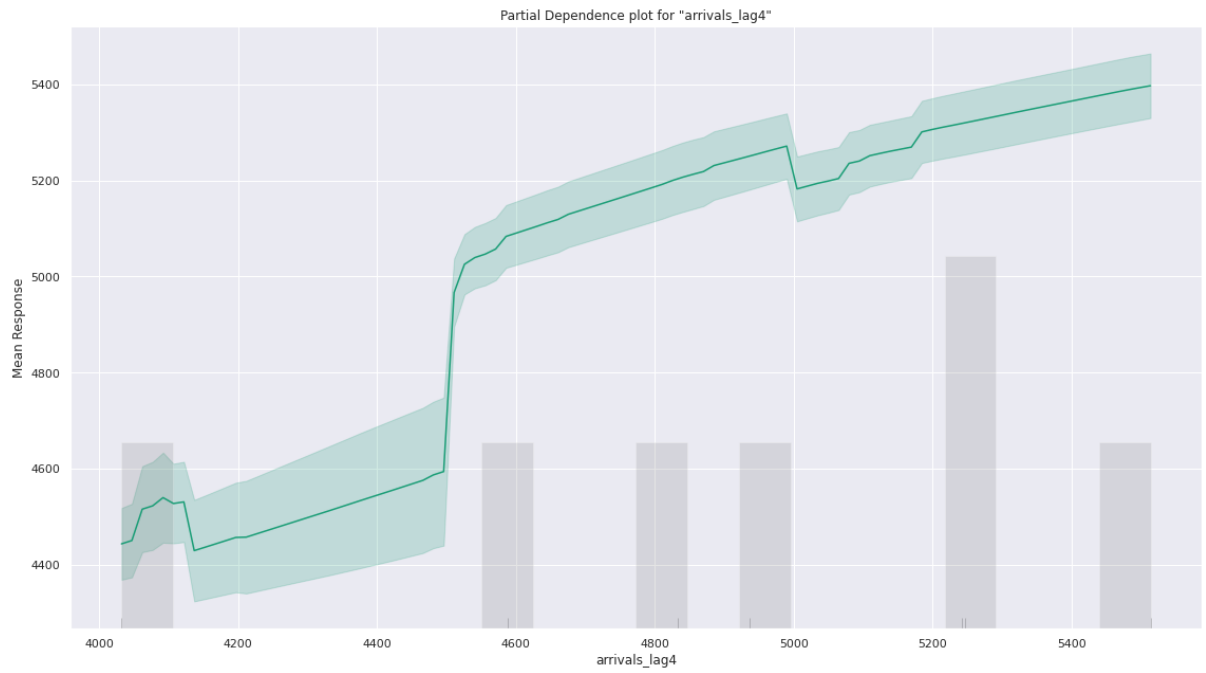


Figure 18. Tourist arrivals (quarterly) to London

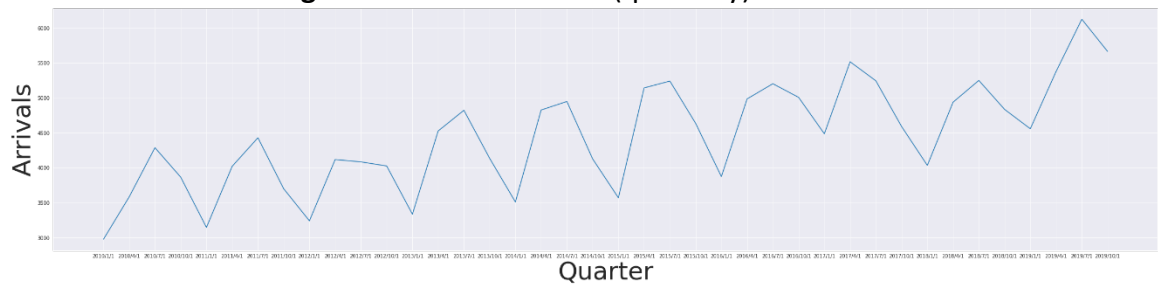


Figure 19. Diagram for the leading neural network model

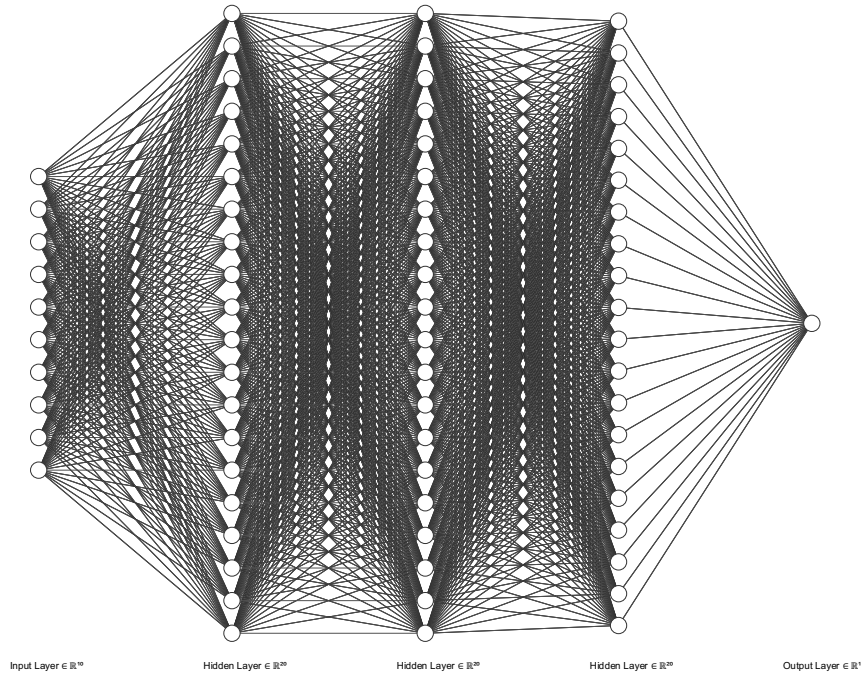


Figure 20. Comparison between the leading model and baseline models

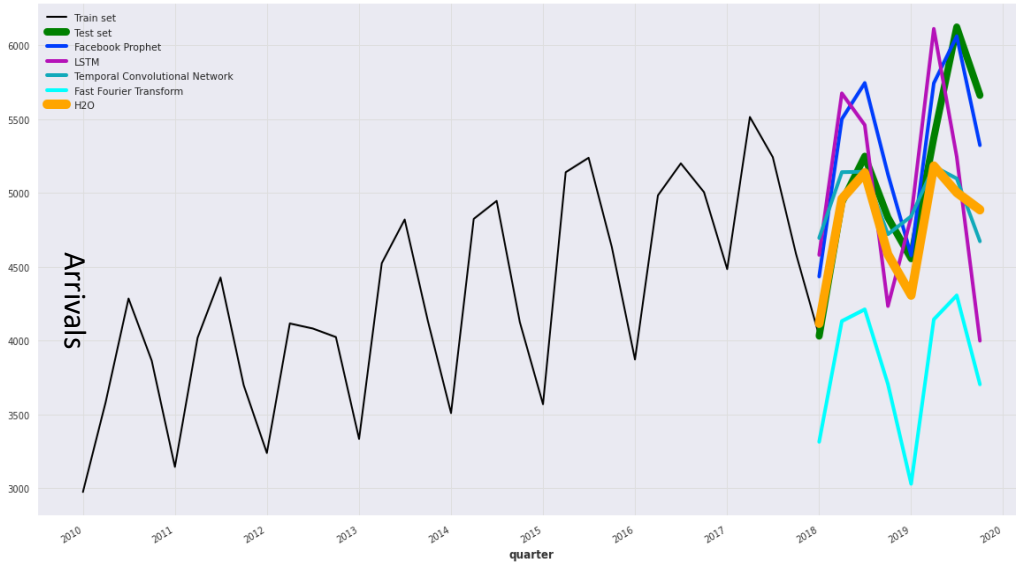


Figure 21. Variable importance

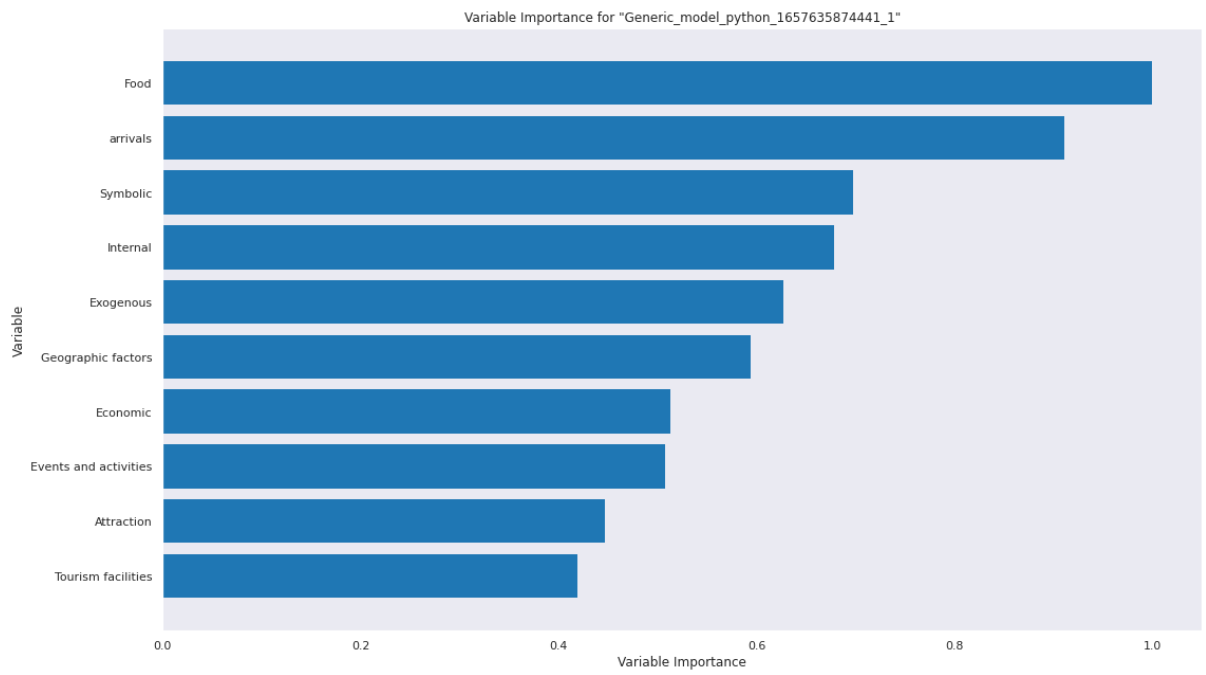


Figure 22. Partial Dependence plot for food

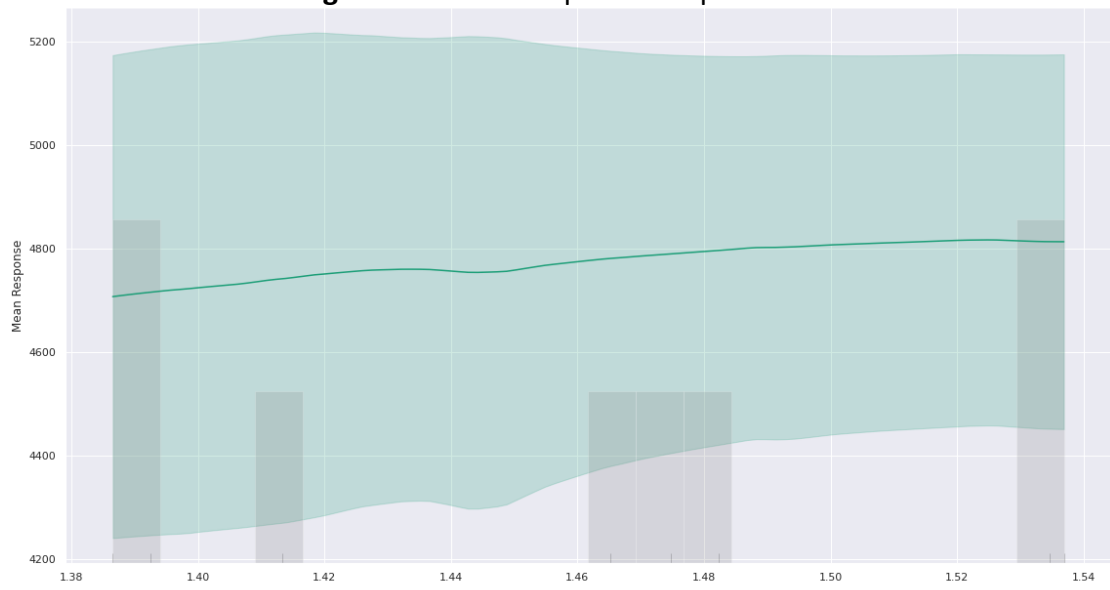


Figure 23. Partial Dependence plot for symbolic factors

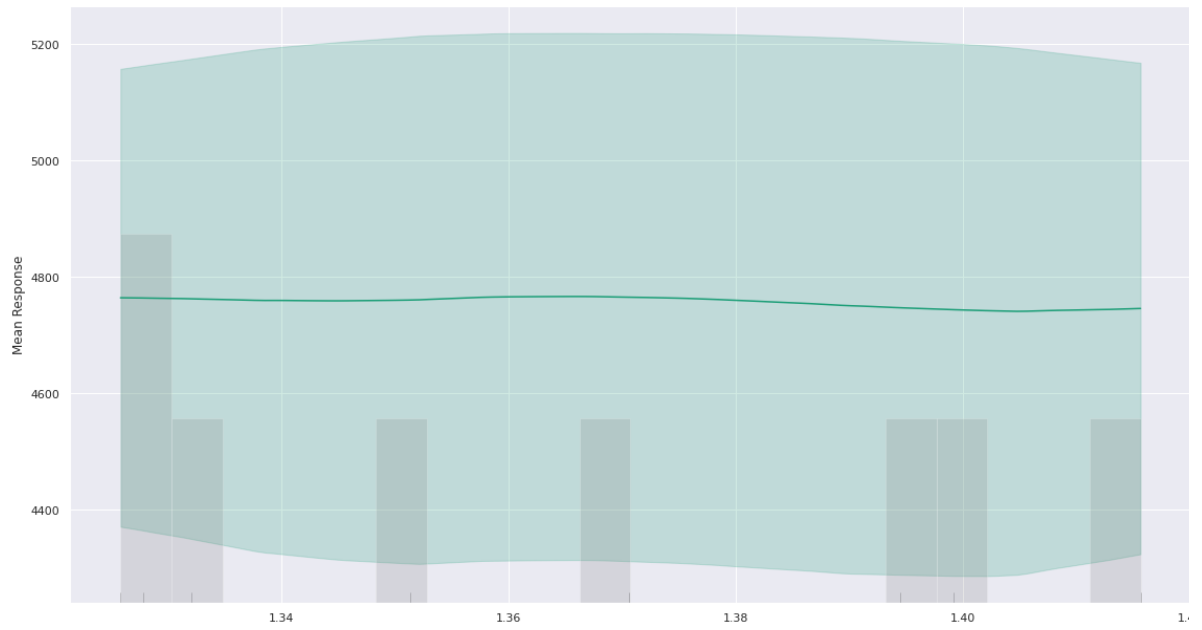


Figure 24. Partial Dependence plot for internal factors

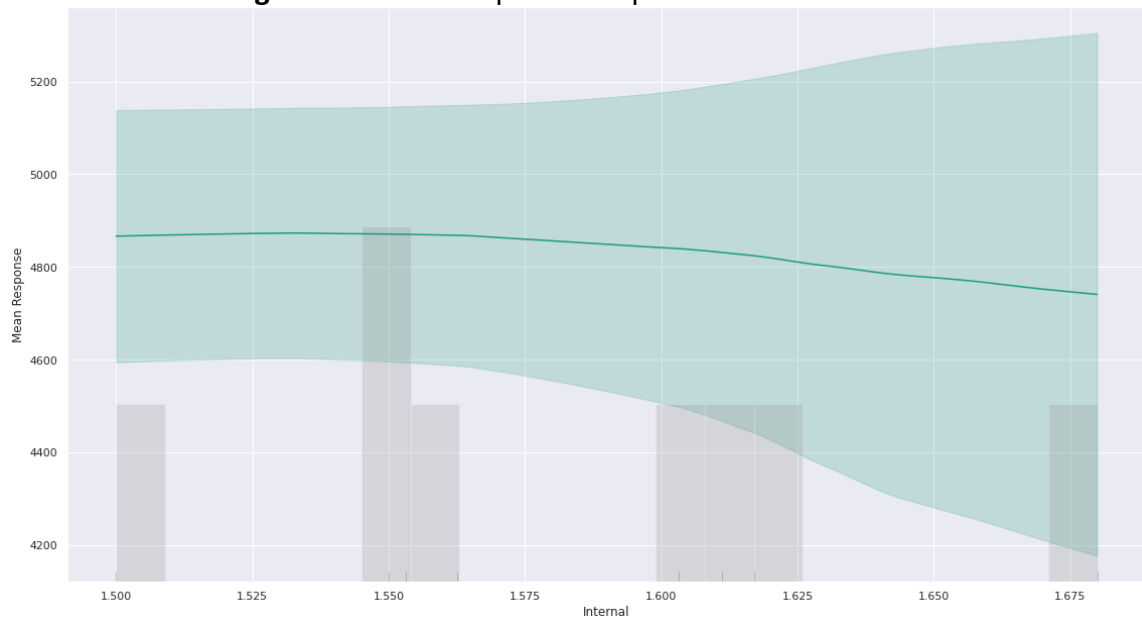


Figure 25. Partial Dependence plot for exogenous factors

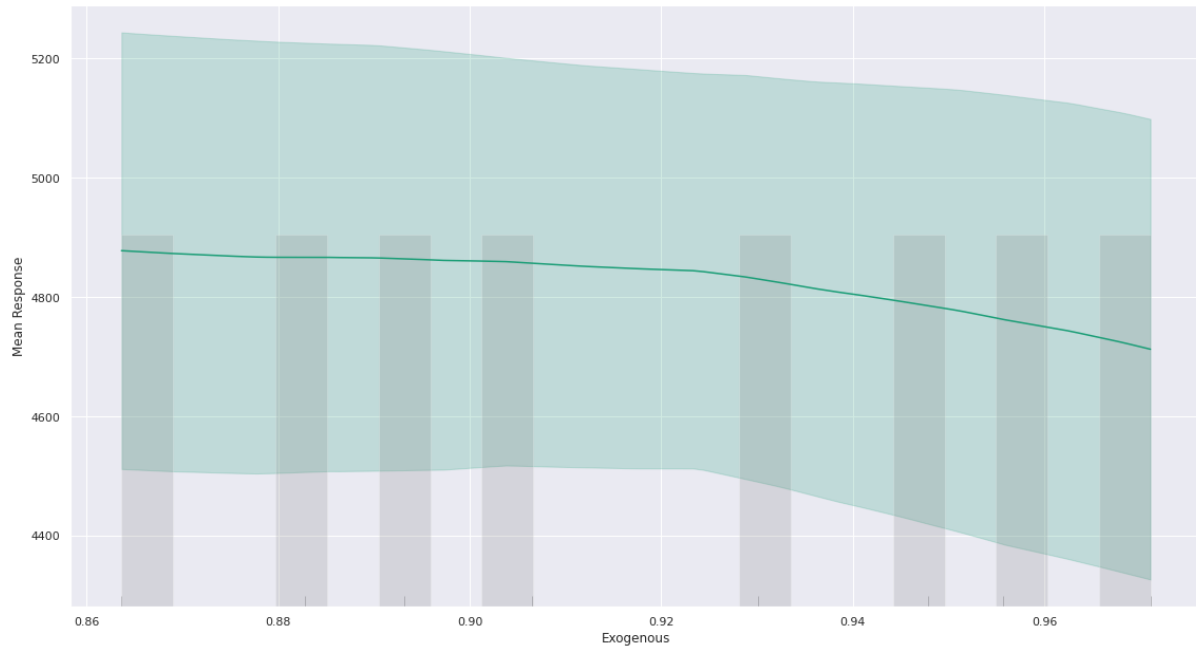
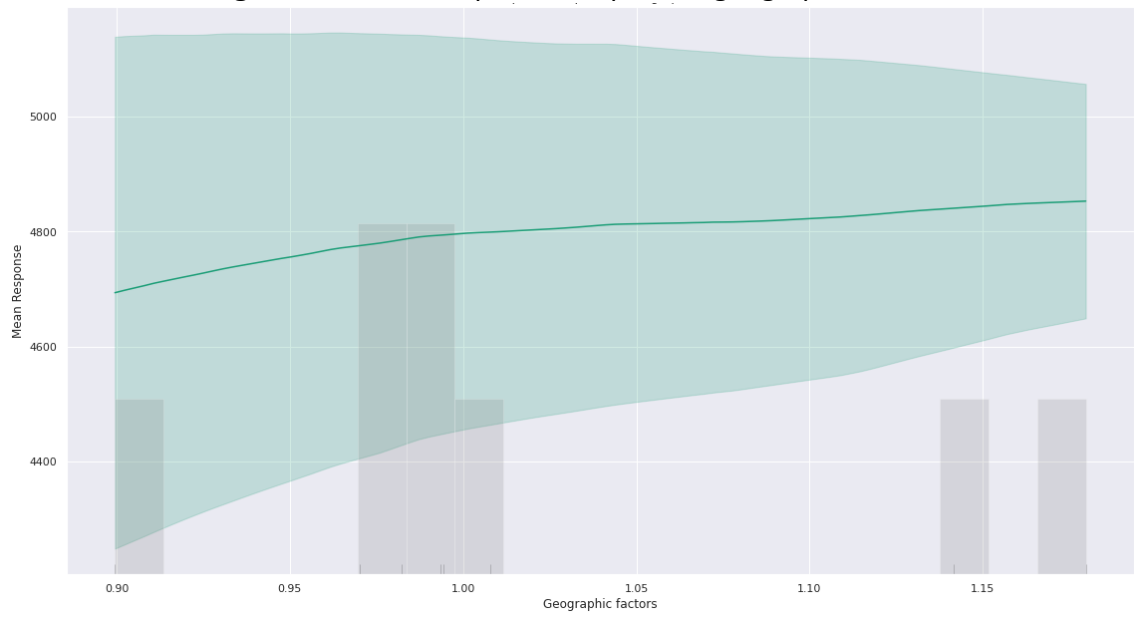


Figure 26. Partial Dependence plot for geographic factors



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
WHs: World Heritage Sites
XGBoost: Extreme Gradient Boosting