SCORES OF HUNGER : UTILIZING GLOBAL DATASETS TO PREDICT REGIONAL FOOD INSECURITY

- Group No 10
- Alok Singh (50545018)
- Rohan Sharma (50545225)
- Shriram Sankaranarayanan (50548133)

CONTENTS:

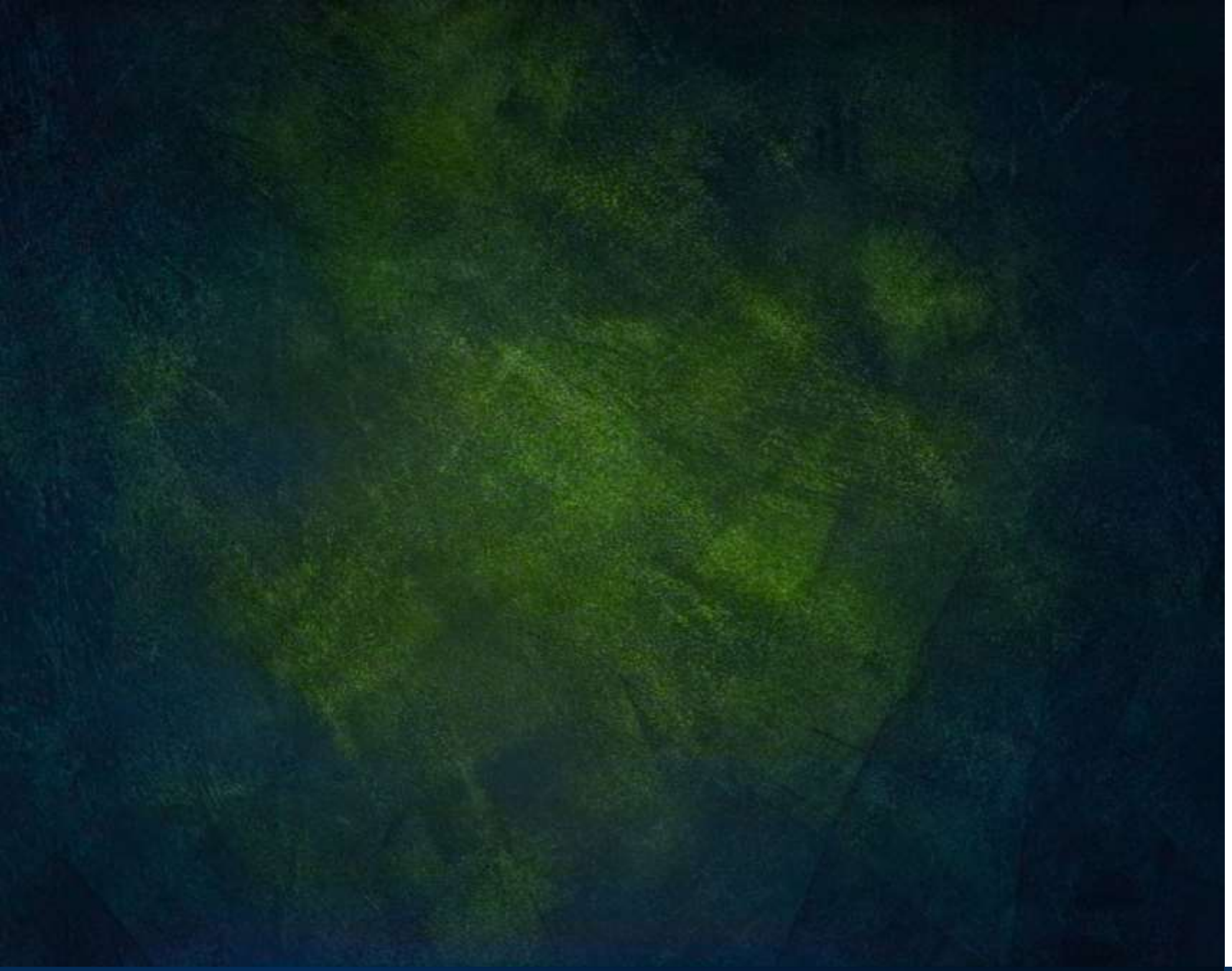
Motivation:

- Currently, there are approximately 870 million people globally lacking access to adequate, nutritious, and safe food supplies.
- These project can help Assists governments and NGOs in better allocating resources and designing effective policies for food security.
- Food security reduces poverty levels as access to adequate food improves livelihoods, particularly in agricultural-dependent communities.
- This can help UN mission of supports the achievement of zero hunger.

Food Security

Introduction:

- Food security is defined when all people, at all times, have physical and economic access to sufficient safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life.
- The Global Food Security Index (GFSI) is an annual assessment that measures the state of food security in countries around the world. It is designed to evaluate the underlying factors and key risks affecting food security in each country and across regions.
- Understanding and optimizing these factors can improve food security of any country.
- These factors are corruption, availability to adequate water supply,
- Trade Freedom, Presence of food safety net program, etc.



Methodology:

- Data sources and collection methods
- Data cleaning and pre-processing

- Selecting relevant predictors for model
- Feature engineering

- Applying Regression models

- Calculating performance metrics (R^2)

- Interpretation of results
- Comparison of model performance

Data Collection and Processing:

Data was sourced from a global food security index, FAO and world data bank.

We pre-processed the data by removing the missing values from Score.

Data was split into 80% training set and 20% testing set.

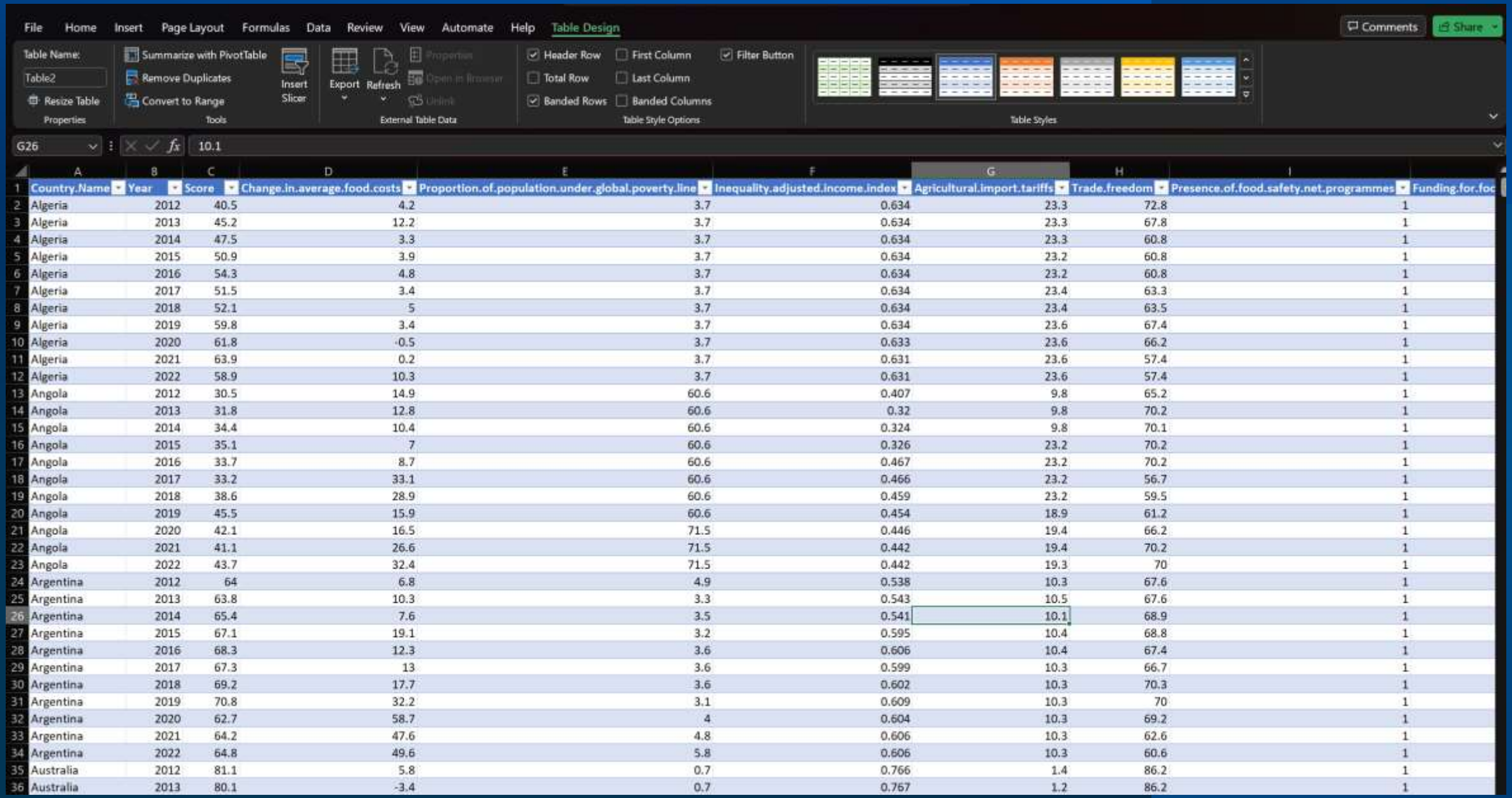
We had 67 predictors, we performed feature selection by visually inspecting the scatterplots and boxplots of the dependent variable vs the individual predictors.

Data collection and processing cont.

We removed the outliers from all numeric variables in our data by removing the top and bottom 2.5% of observations.

Then, we scaled the numeric variables except the target variable in our data using minmax scaler before beginning the modeling phase.

After that we split our data into 80% train and 20% test sets which will be used for all our subsequent models.



File Home Insert Page Layout Formulas Data Review View Automate Help Table Design

Comments Share

Table Name: Table2
Resize Table
Properties

Summarize with PivotTable
Remove Duplicates
Convert to Range
Tools

Insert Slicer

Export Refresh
Properties
Open in Browser
Unlink
External Table Data

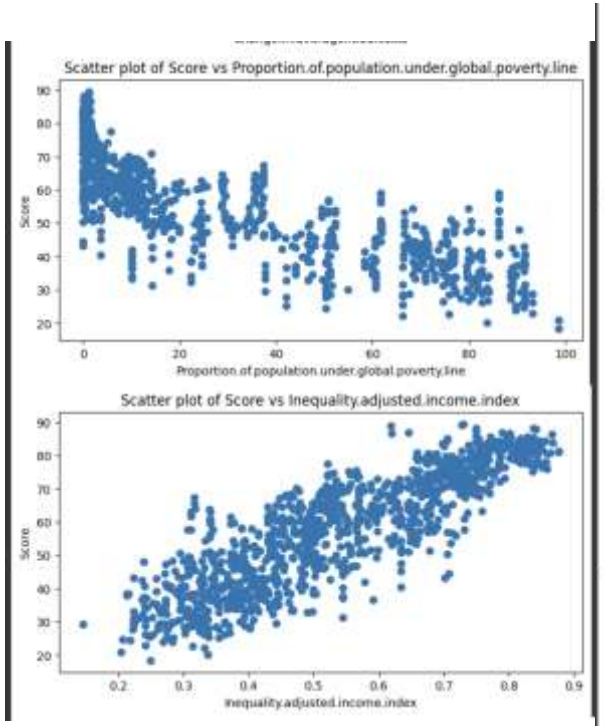
Header Row
Total Row
Banded Rows
First Column
Last Column
Banded Columns
Filter Button
Table Style Options

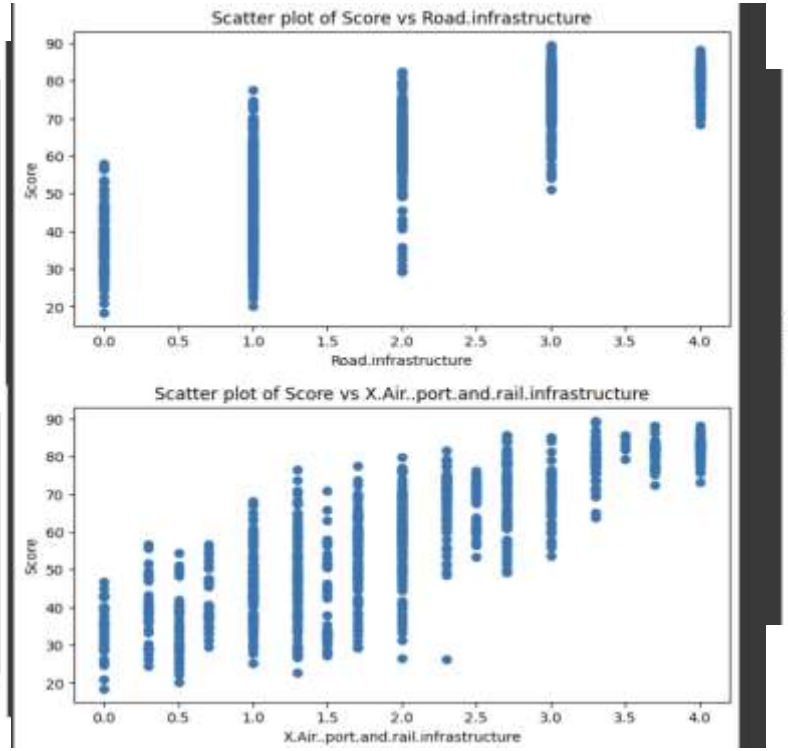
Table Styles

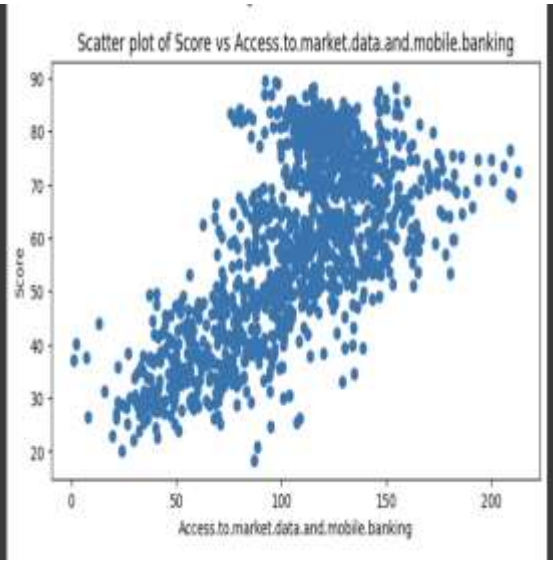
G26 10.1

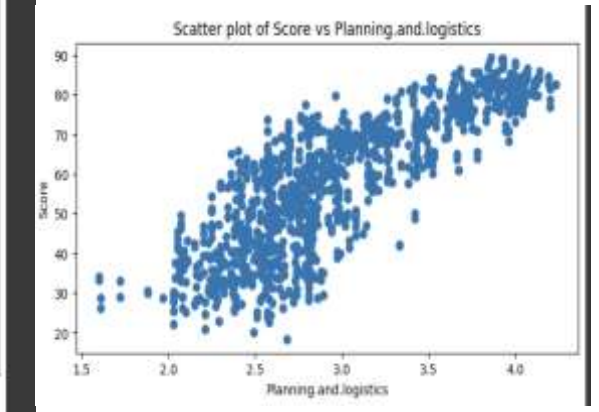
	A	B	C	D	E	F	G	H	I	
1	Country.Name	Year	Score	Change.in.average.food.costs	Proportion.of.population.under.global.poverty.line	Inequality.adjusted.income.index	Agricultural.import.tariffs	Trade.freedom	Presence.of.food.safety.net.programmes	Funding.for.foo
2	Algeria	2012	40.5	4.2	3.7	0.634	23.3	72.8	1	
3	Algeria	2013	45.2	12.2	3.7	0.634	23.3	67.8	1	
4	Algeria	2014	47.5	3.3	3.7	0.634	23.3	60.8	1	
5	Algeria	2015	50.9	3.9	3.7	0.634	23.2	60.8	1	
6	Algeria	2016	54.3	4.8	3.7	0.634	23.2	60.8	1	
7	Algeria	2017	51.5	3.4	3.7	0.634	23.4	63.3	1	
8	Algeria	2018	52.1	5	3.7	0.634	23.4	63.5	1	
9	Algeria	2019	59.8	3.4	3.7	0.634	23.6	67.4	1	
10	Algeria	2020	61.8	-0.5	3.7	0.633	23.6	66.2	1	
11	Algeria	2021	63.9	0.2	3.7	0.631	23.6	57.4	1	
12	Algeria	2022	58.9	10.3	3.7	0.631	23.6	57.4	1	
13	Angola	2012	30.5	14.9	60.6	0.407	9.8	65.2	1	
14	Angola	2013	31.8	12.8	60.6	0.32	9.8	70.2	1	
15	Angola	2014	34.4	10.4	60.6	0.324	9.8	70.1	1	
16	Angola	2015	35.1	7	60.6	0.326	23.2	70.2	1	
17	Angola	2016	33.7	8.7	60.6	0.467	23.2	70.2	1	
18	Angola	2017	33.2	33.1	60.6	0.466	23.2	56.7	1	
19	Angola	2018	38.6	28.9	60.6	0.459	23.2	59.5	1	
20	Angola	2019	45.5	15.9	60.6	0.454	18.9	61.2	1	
21	Angola	2020	42.1	16.5	71.5	0.446	19.4	66.2	1	
22	Angola	2021	41.1	26.6	71.5	0.442	19.4	70.2	1	
23	Angola	2022	43.7	32.4	71.5	0.442	19.3	70	1	
24	Argentina	2012	64	6.8	4.9	0.538	10.3	67.6	1	
25	Argentina	2013	63.8	10.3	3.3	0.543	10.5	67.6	1	
26	Argentina	2014	65.4	7.6	3.5	0.541	10.1	68.9	1	
27	Argentina	2015	67.1	19.1	3.2	0.595	10.4	68.8	1	
28	Argentina	2016	68.3	12.3	3.6	0.606	10.4	67.4	1	
29	Argentina	2017	67.3	13	3.6	0.599	10.3	66.7	1	
30	Argentina	2018	69.2	17.7	3.6	0.602	10.3	70.3	1	
31	Argentina	2019	70.8	32.2	3.1	0.609	10.3	70	1	
32	Argentina	2020	62.7	58.7	4	0.604	10.3	69.2	1	
33	Argentina	2021	64.2	47.6	4.8	0.606	10.3	62.6	1	
34	Argentina	2022	64.8	49.6	5.8	0.606	10.3	60.6	1	
35	Australia	2012	81.1	5.8	0.7	0.766	1.4	86.2	1	
36	Australia	2013	80.1	-3.4	0.7	0.767	1.2	86.2	1	

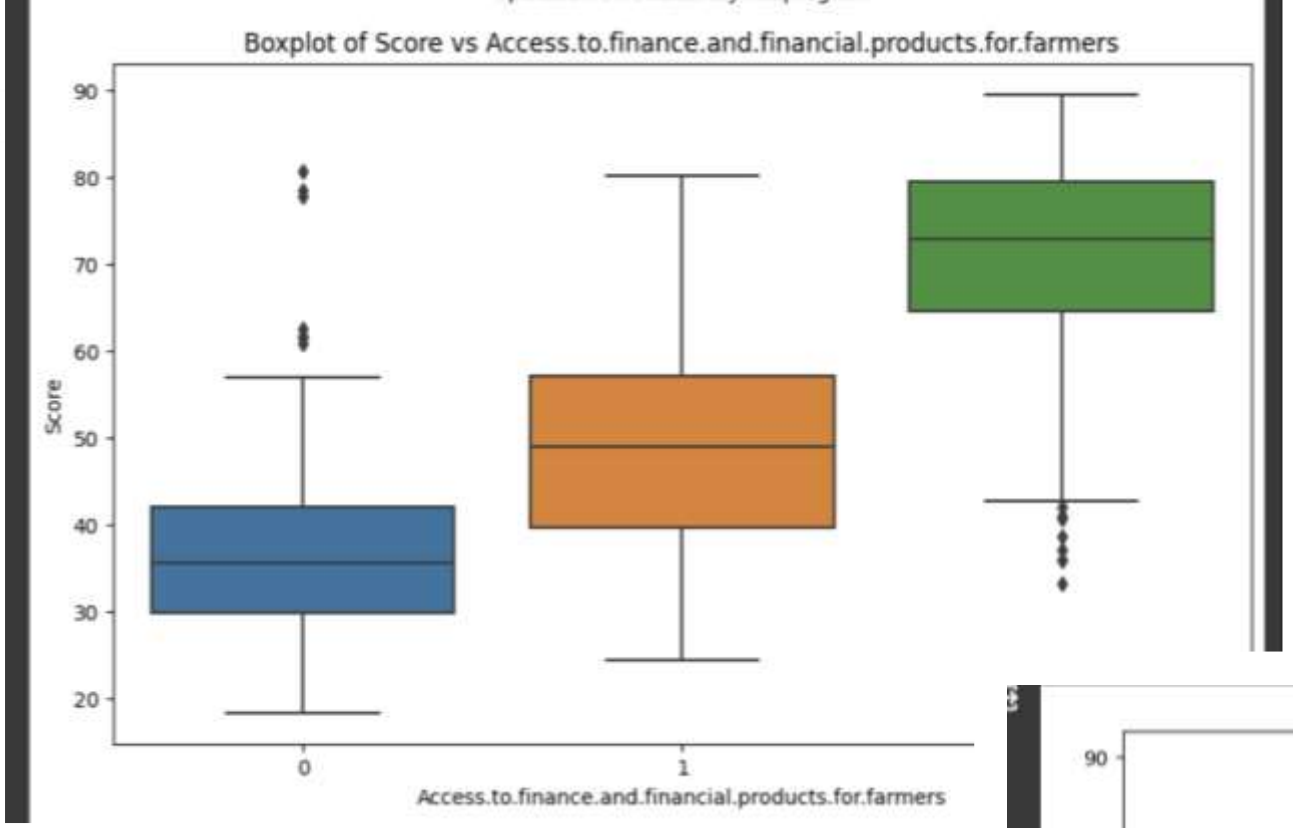
Exploratory Data Analysis: Feature Selection

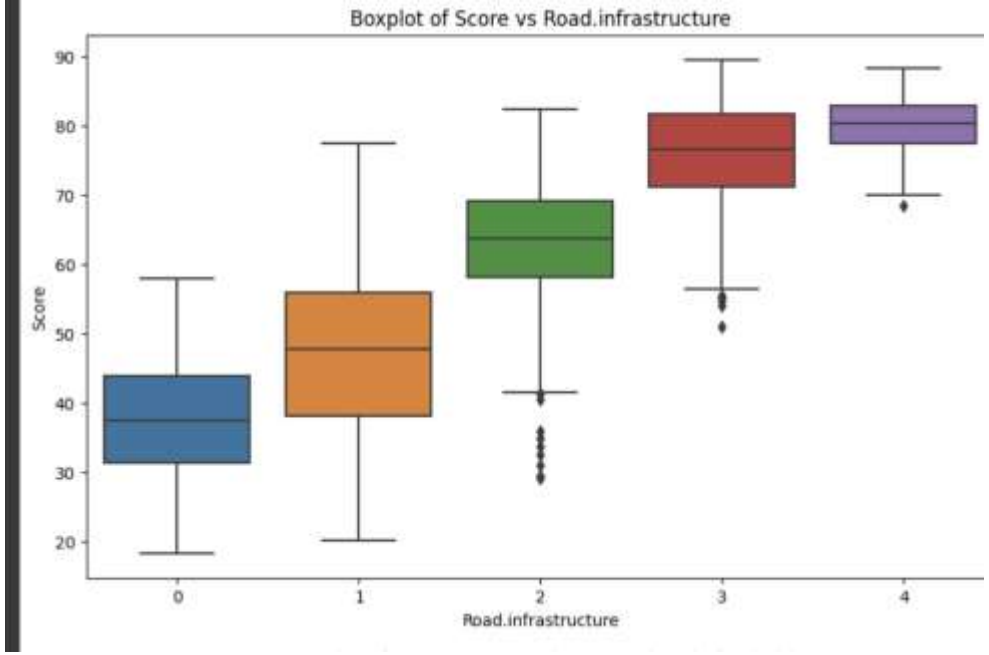


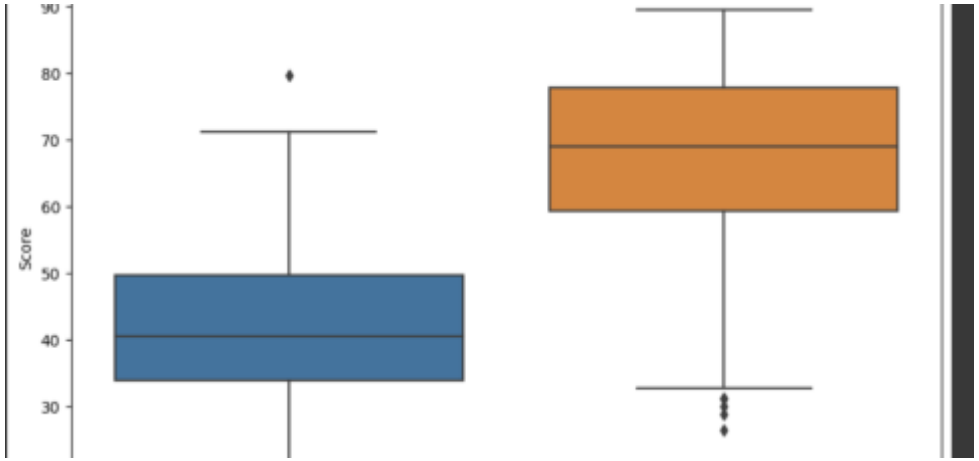


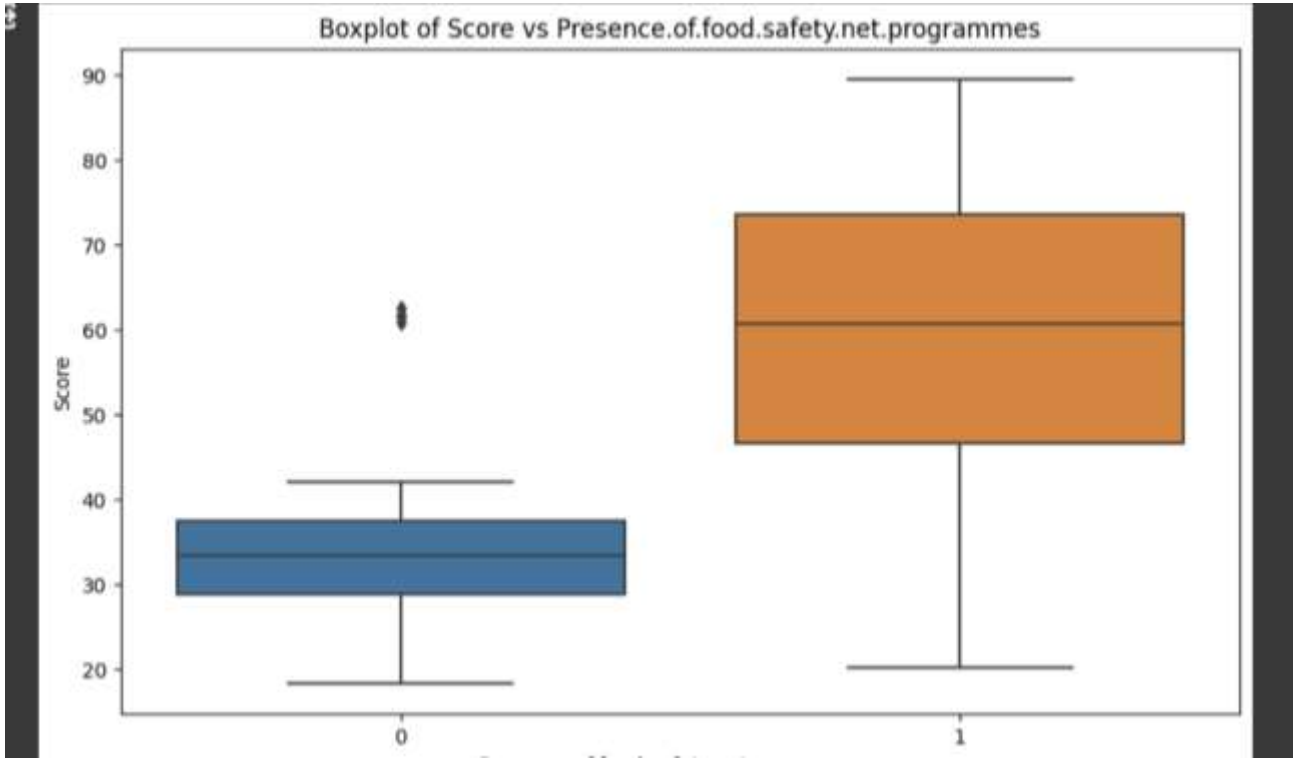





Boxplot of Score vs Access.to.finance.and.financial.products.for.farmers
Score
Access.to.finance.and.financial.products.for.farmers


Boxplot of Score vs Road.infrastructure
Score
Road.infrastructure


Score

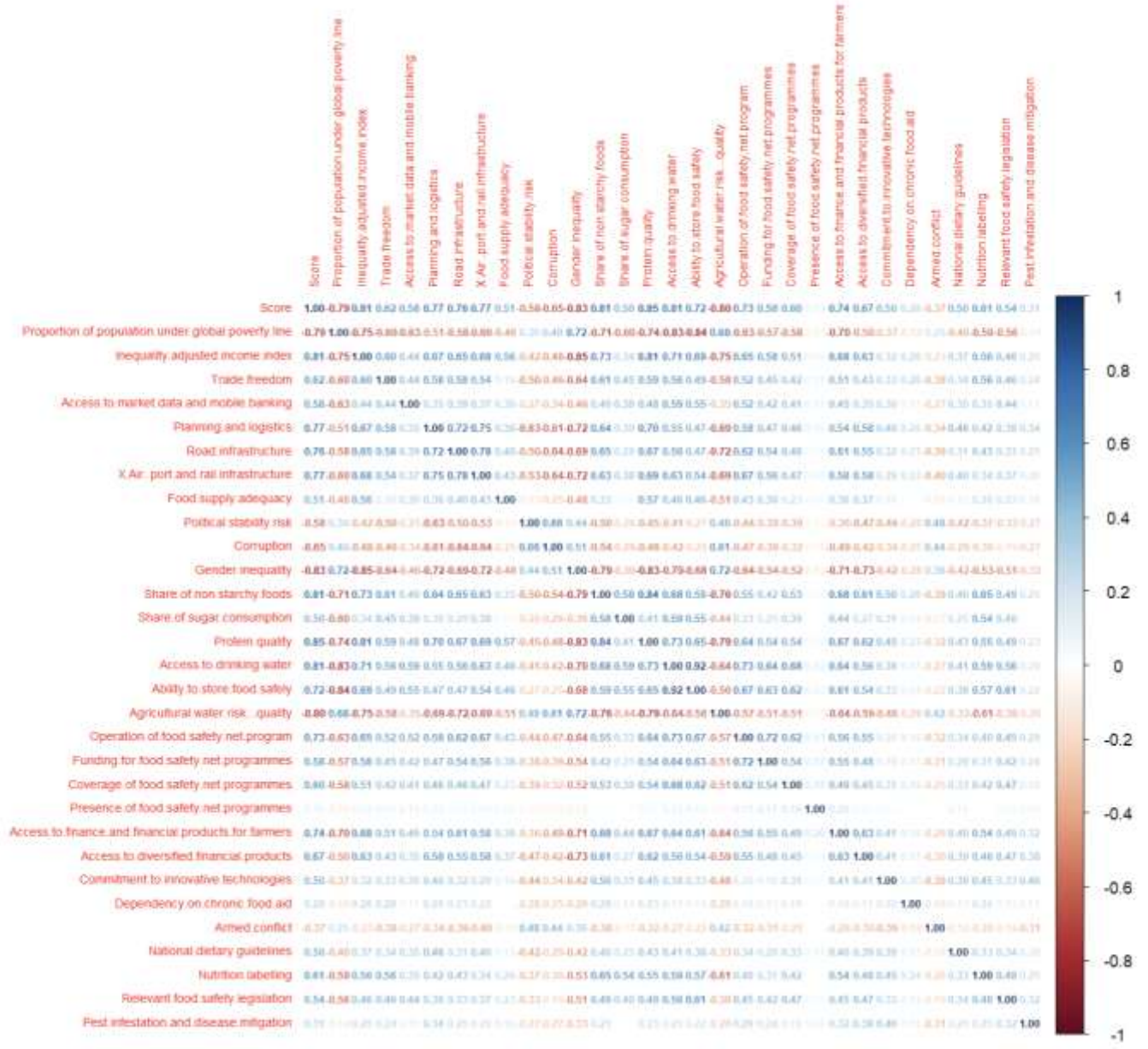

Boxplot of Score vs Presence.of.food.safety.net.programmes
Score

Predictors Selected

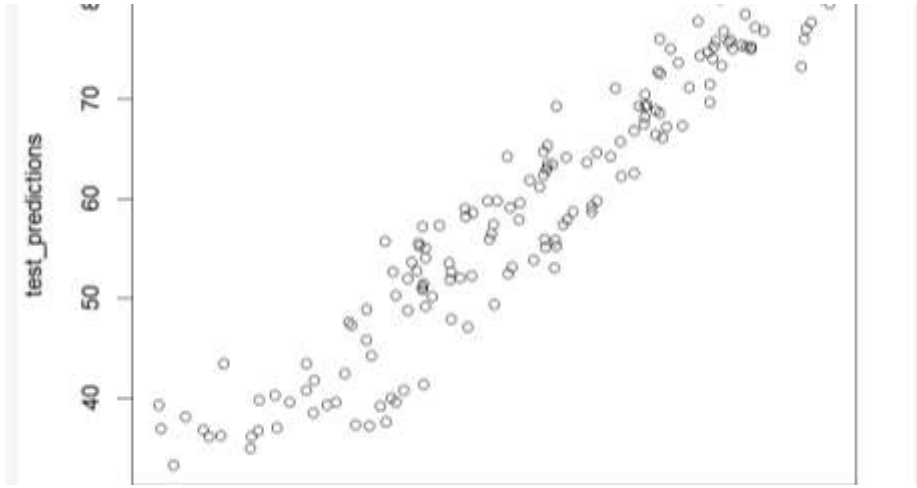
proportion of population under global poverty
inequality adjusted income index
Trade freedom
Acess to market data and mobile banking
planning and logistics
Road infrastructure
Air port and rail infrastructure
food supply adequacy
political stability risk
corruption
gender inequality
share of non starchy foods
share of sugar consuption
protein quality
Acess to drinking water
ability to store food safely
agriculture water risk quality

categorical
Food safety net program
funding for food safety programmes
coverage of food safety net program
operation of food safety net program
acess to finance and fianancial product for farmers
acess to diversified financial products
commitment to innovative technologies
dependency on chronic food
armed conflict
national dietary guidline
nutrition labelling
relevant food legislation
pest infestation and disease mitigation

Correlation plot

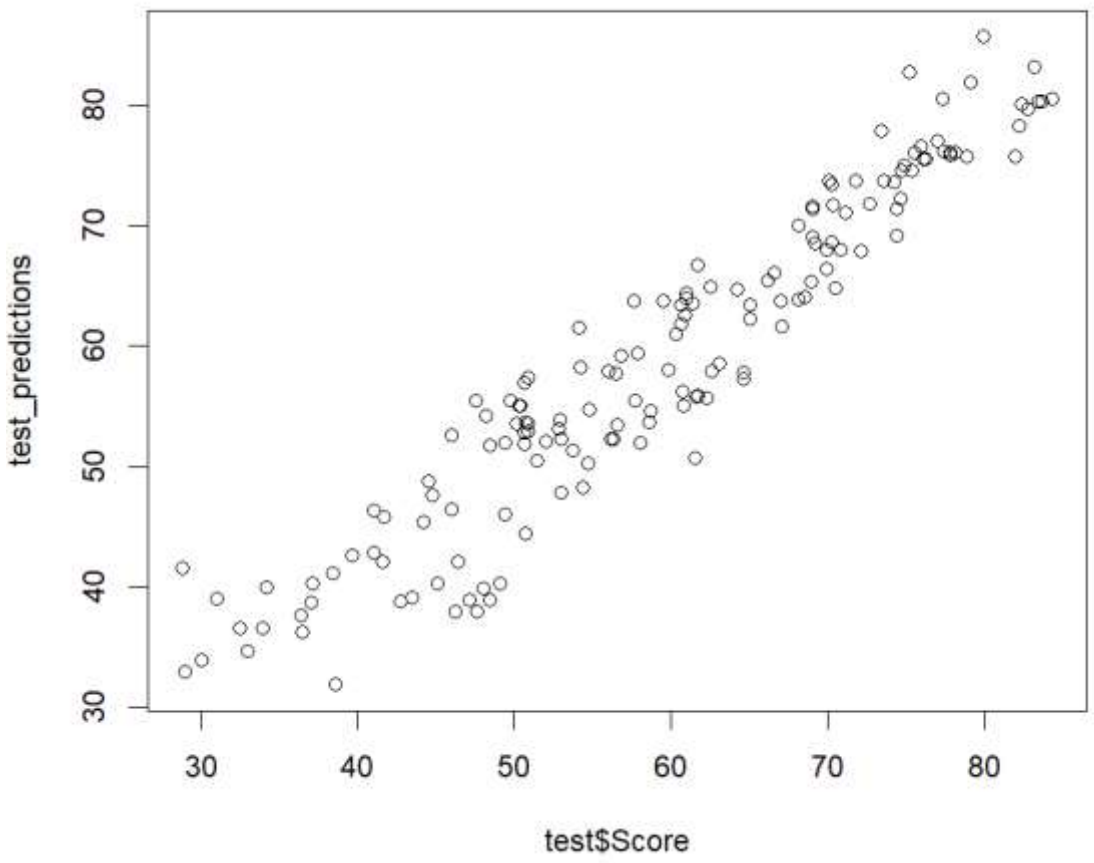


Models




```
(Intercept)                                                  ***
Proportion.of.population.under.global.poverty.line           *
Inequality.adjusted.income.index
Trade.freedom                                                **
Access.to.market.data.and.mobile.banking                     ***
Planning.and.logistics                                       ***
Road.infrastructure                                          ***
X.Air..port.and.rail.infrastructure
Food.supply.adequacy                                         ***
Political.stability.risk                                     .
Corruption                                                   ***
Gender.inequality                                            .
Share.of.non.starchy.foods                                   *
Share.of.sugar.consumption
Protein.quality                                              **
```

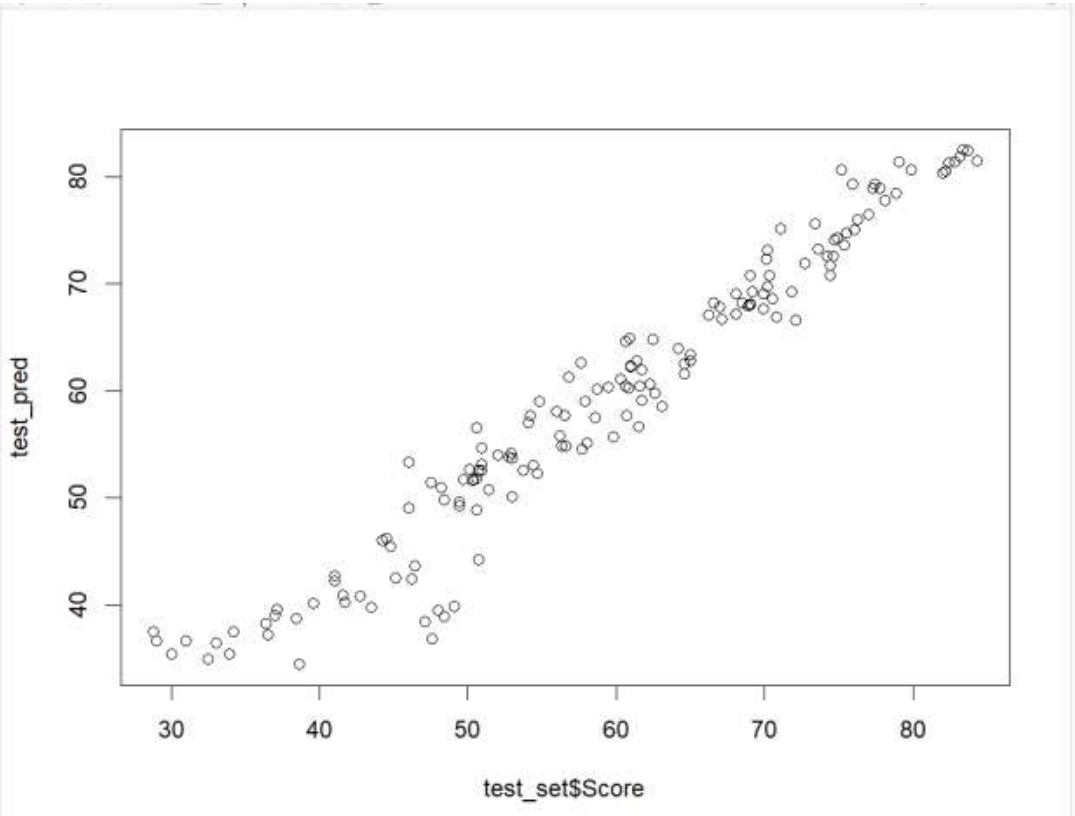
- Linear Regression- We fit the model on the training set and used it to make predictions on the test set.
- Results : R-squared for Training Set: 0.9194423
- R-squared for Testing Set: 0.908228

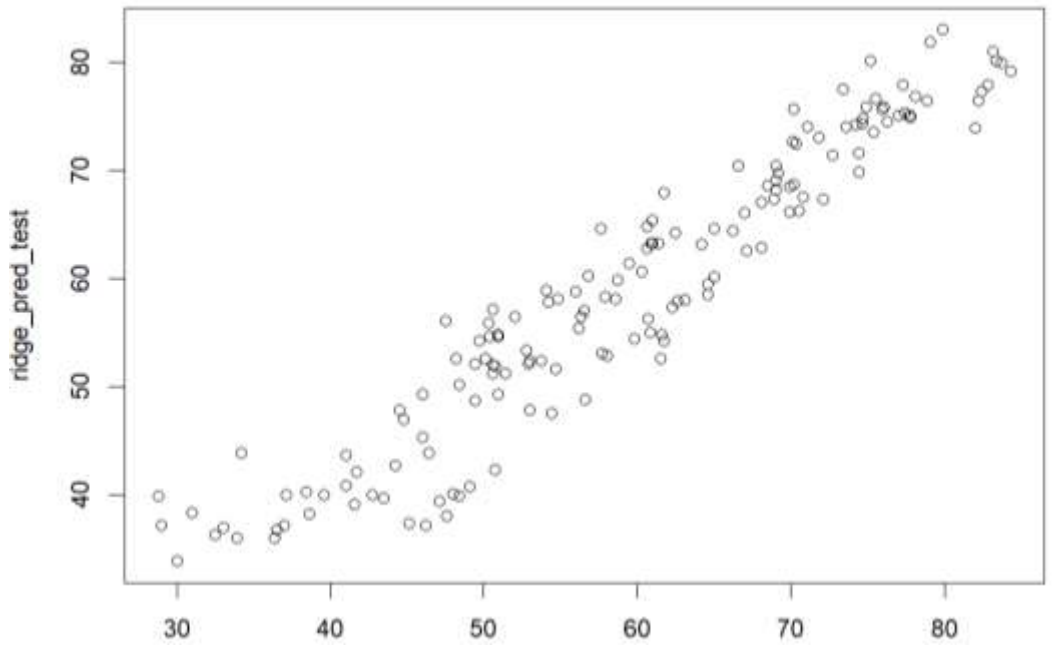


Mars Model
Results:
R-squared for Training Set: 0.9098528
R-squared for Testing Set: 0.9057779
For MARS model, we used cross validation to select our model parameters degree and nprune.

Lasso and Ridge Model
Lasso:
Results
Lasso R-squared for Training Set: 0.9188376
Lasso R-squared for Testing Set: 0.9088209

Ridge:
Results:
Ridge R-squared for Training Set: 0.9183641
Ridge R-squared for Testing Set: 0.9099601

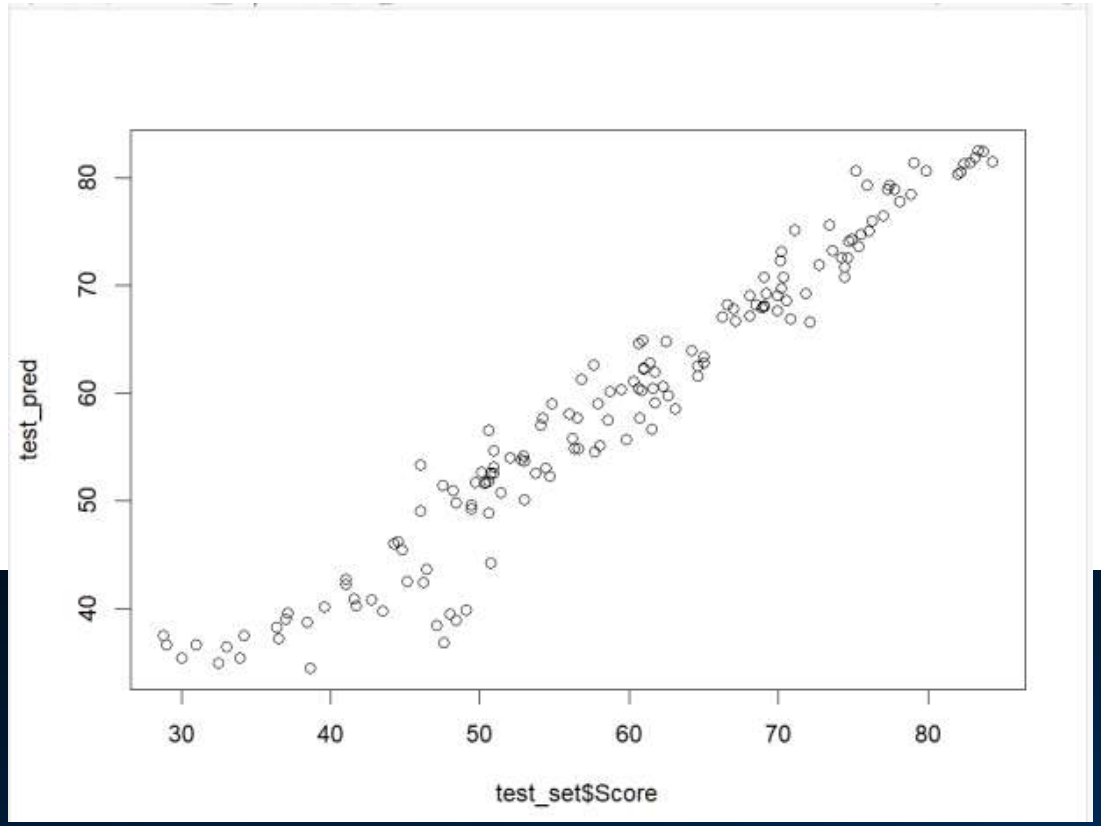




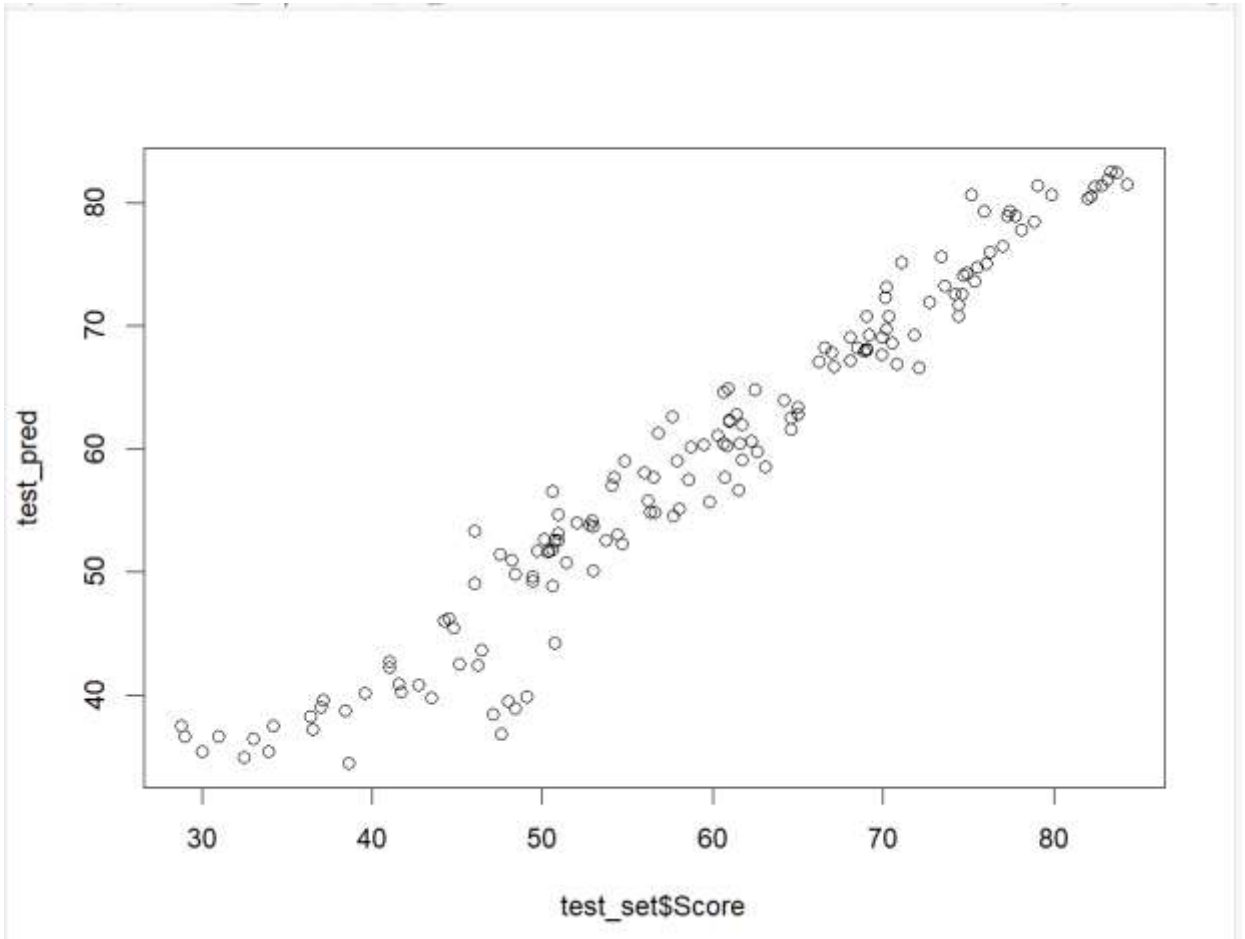
Random Forest Results:

Random Forest R-squared for Training Set: 0.9897387

Random Forest R-squared for Testing Set: 0.9495487

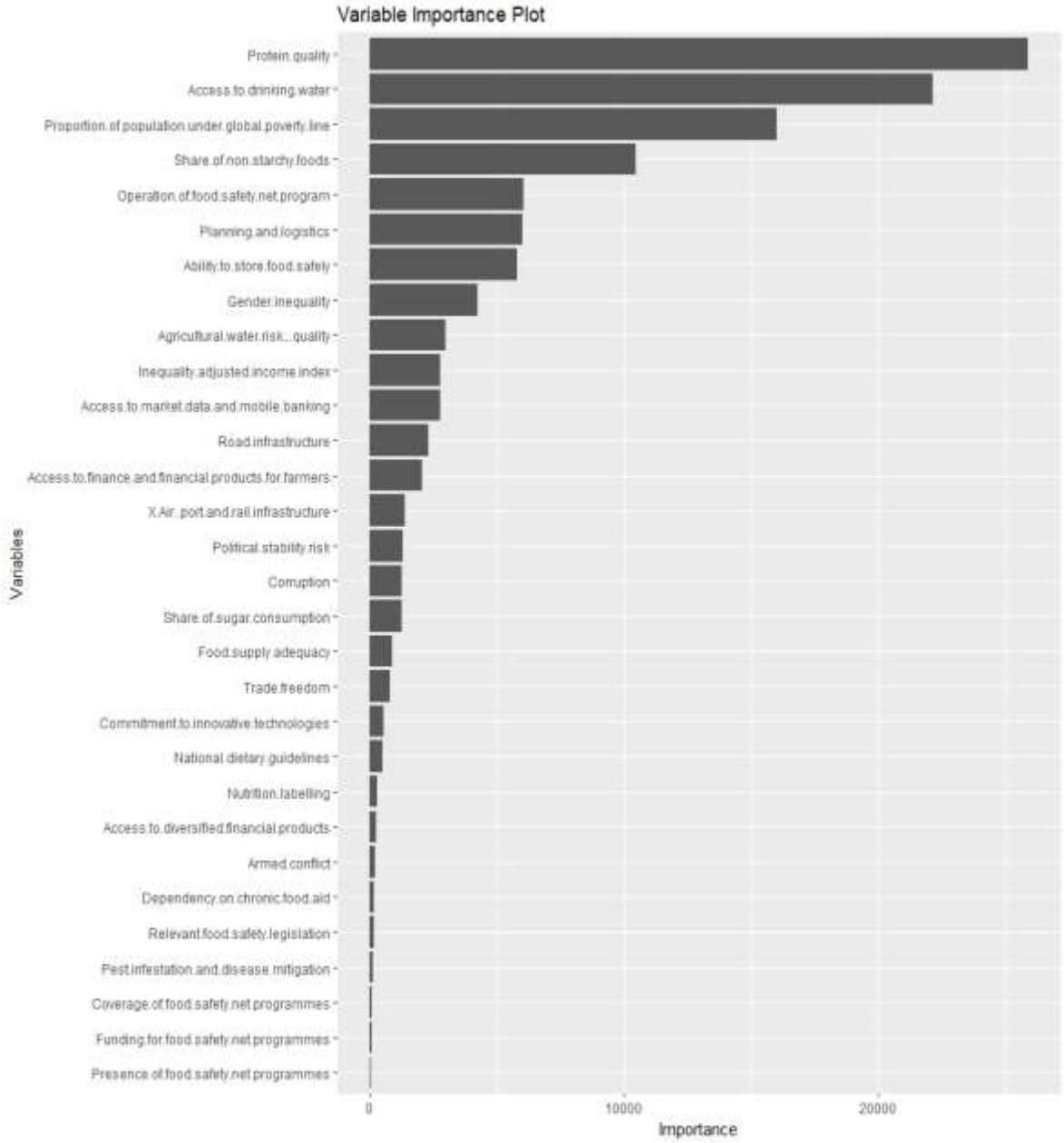


XG Boost:
Results:
XGBoost R-squared for Training Set: 0.9999228
XGBoost R-squared for Testing Set: 0.9476693

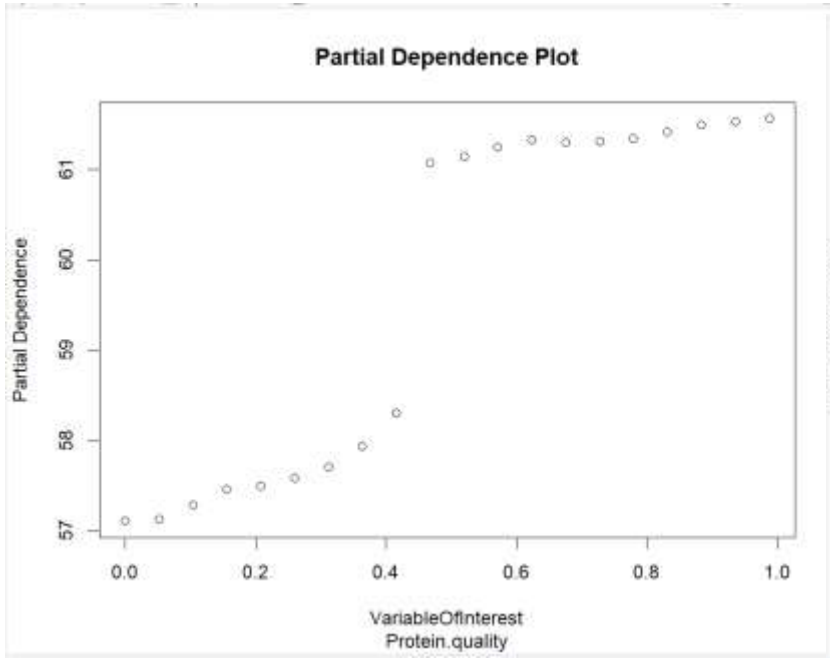


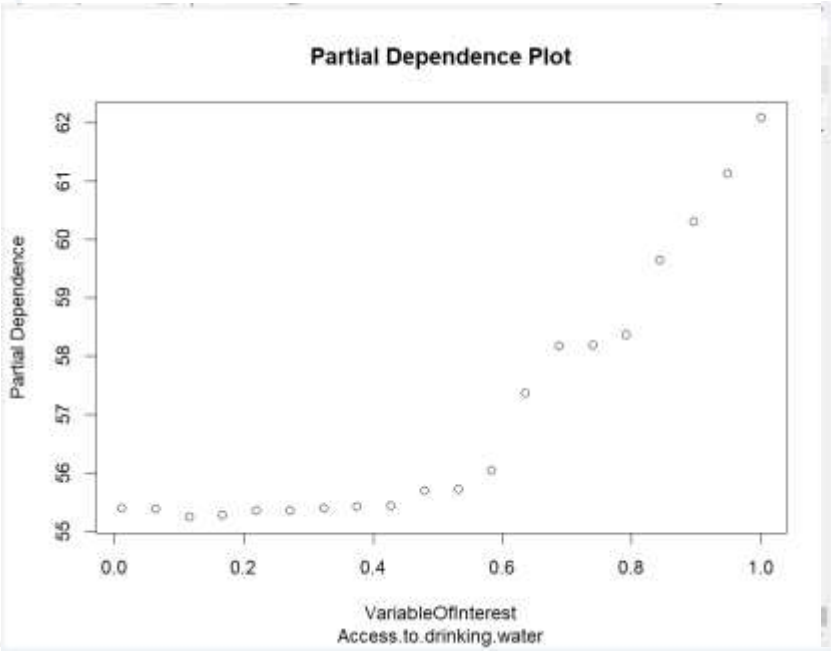
Results and Conclusion

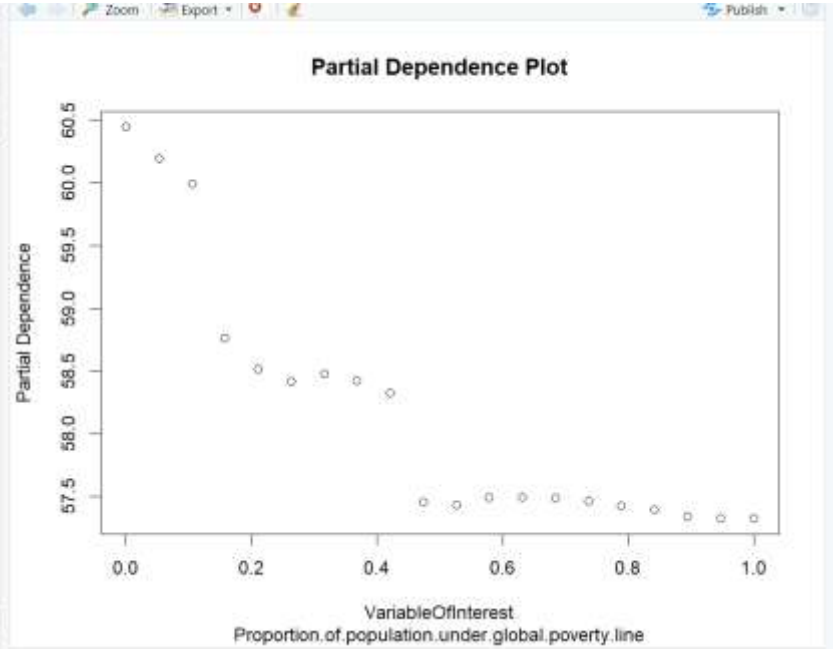
- Variable Importance plot from random forest model.

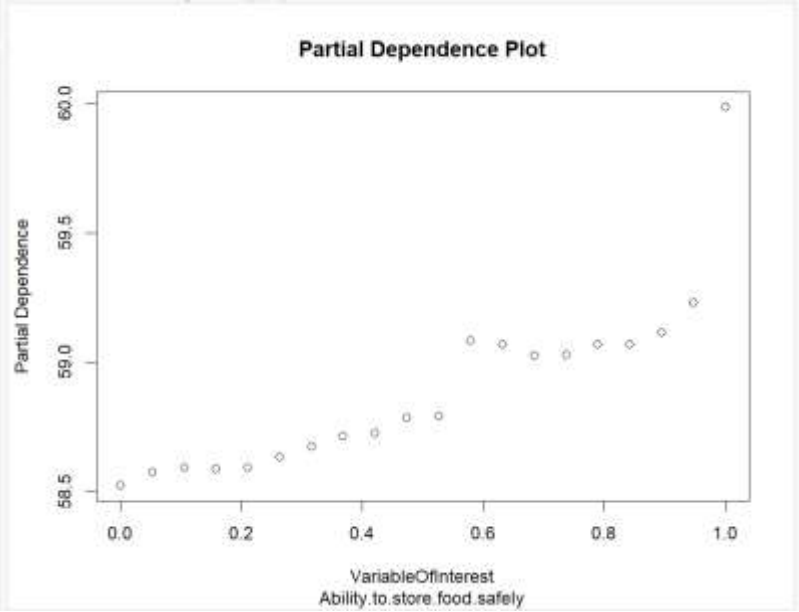


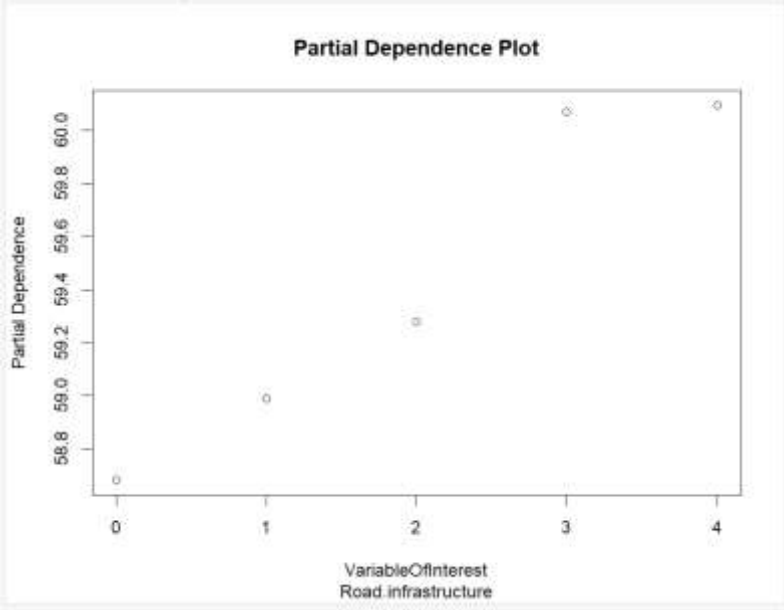
Partial dependence plots

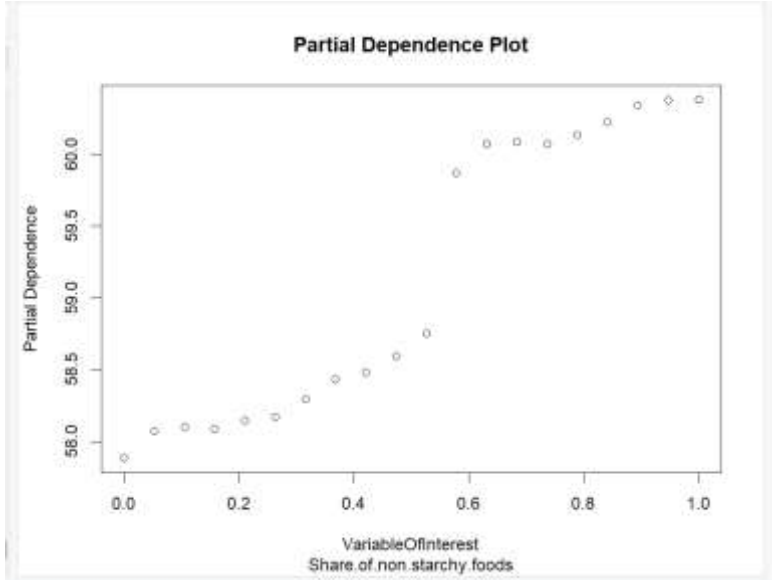


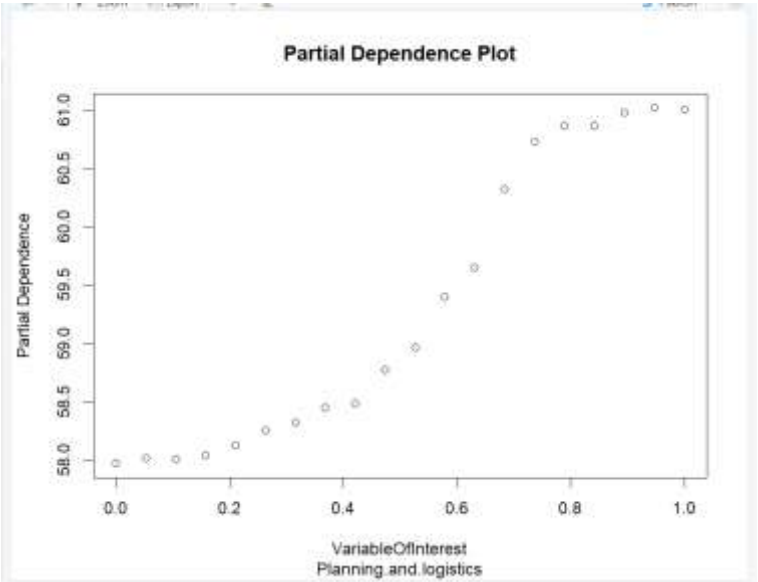





Partial Dependence Plot
Partial Dependence
60.0
59.5
59.0
58.5
0.0
0.2
0.4
0.6
0.8
1.0
VariableOfInterest
Ability.to.store.food.safely


Partial Dependence Plot
Partial Dependence
60.0
59.8
59.6
59.4
59.2
59.0
58.8
0
1
2
3
4
VariableOfInterest
Road.infrastructure


Partial Dependence Plot
Partial Dependence
60.0
59.5
59.0
58.5
58.0
0.0
0.2
0.4
0.6
0.8
1.0
VariableOfInterest
Share.of.non.starchy.foods


Partial Dependence Plot
Partial Dependence
61.0
60.5
60.0
59.5
59.0
58.5
58.0
0.0
0.2
0.4
0.6
0.8
1.0
VariableOfInterest
Planning.and.logistics

Comparison

Model	Training_R2	Testing_R2
Linear Model	0.9194423	0.9082280
Mars Model	0.9098528	0.9057779
Lasso	0.9188376	0.9088209
Ridge	0.9183641	0.9099601
Random Forest	0.9897387	0.9495487
XG Boost	0.9999228	0.9476693

Conclusion:

- We see that Random Forest Model gives us the best results.
- From the variable importance plots we saw which variables are the most significant in determining the Food security score of countries.
- We can infer from our results that areas such as Access to drinking water, Protein quality, Percentage of people living under global poverty line, Planning and Logistics, and Infrastructure for storage and transport are the ones that require focus from governments or organizations that aim at improving food security for a region.

Thank You