Data Analysis Report

Raw Data Overview

Road Traffic Accidents and Related Variables

year_month	total_road_traffic_accidents	composite_consumer_price_index	year_on_year_percent_change	consumer_price_index_a	year_on_
2018-07	1397	96.9	2.4	97.1	2.5
2018-08	1419	97.0	2.3	97.2	2.5
2018-09	1271	97.2	2.7	97.9	3.3
2018-10	1439	97.4	2.7	98.1	3.2
2018-11	1375	97.6	2.6	98.1	3.0
2018-12	1304	97.9	2.5	98.4	3.0

Summary Statistics

Total Road Traffic Accidents

Statistic	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Count	833	1317	1394	1374	1474	1615

Composite Consumer Price Index

Statistic	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Count	96.9	100.2	101.6	102.2	105.0	107.9

Traffic Density

Statistic	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Count	1326	8717	10858	9706	11737	13515

Principal Component Analysis (PCA)

Importance of Components

Component	Standard Deviation	Proportion of Variance	Cumulative Proportion
PC1	2.9246	0.5031	0.5031
PC2	1.8875	0.2096	0.7127
PC3	1.2193	0.08745	0.80017
PC4	1.0448	0.06421	0.86438
PC5	0.9757	0.0560	0.9204

Loadings

PC1

- Composite Consumer Price Index: **0.327**
- Consumer Price Index B: **0.329**
- Consumer Price Index C: **0.329**

PC2

- Year-on-Year Percent Change: **0.506**
- Year-on-Year Percent Change B: **0.498**

PC3

- Traffic Density: **0.447**
- Accidents Per Thousand: **-0.471**

Regression Analysis

Linear Regression

Formula:

```
total_road_traffic_accidents ~ composite_consumer_price_index + year_on_year_percent_change + average_rainfall + traffic_density
```

Summary:

- **R-squared**: 0.2096
- **Adjusted R-squared**: 0.1423
- **RMSE**: 185.69
- **Significant Variables**:
  - Composite Consumer Price Index (*p = 0.01872*)
  - Traffic Density (*p = 0.00639*)

Coefficients:

Variable	Estimate	Std. Error	t-value	p-value
Intercept	-936.7681	887.3080	-1.056	0.29648
Composite Consumer Price Index	20.3156	8.3413	2.436	0.01872 *
Year-on-Year Percent Change	-14.3306	16.1300	-0.888	0.37883
Average Rainfall	4.0625	2.5149	1.615	0.11292
Traffic Density	0.7009	0.2455	2.855	0.00639 **

Logistic Regression

Formula:

```
high_risk ~ composite_consumer_price_index + year_on_year_percent_change + average_rainfall + traffic_density
```

Summary:

- **AUC-ROC**: 0.68
- **Accuracy**: 0.65
- **Precision**: 0.33
- **Recall**: 0.33

Coefficients:

Variable	Estimate	Std. Error	z-value	p-value
Intercept	-27.065444	19.950934	-1.357	0.1749
Composite Consumer Price Index	0.252512	0.187348	1.348	0.1777
Year-on-Year Percent Change	-0.056262	0.318621	-0.177	0.8598
Average Rainfall	0.078217	0.046698	1.675	0.0939 .
Traffic Density	-0.001079	0.004332	-0.249	0.8032

Confusion Matrix:

Predicted	Actual 0	Actual 1
0	13	4
1	4	2

Predictions

Increase Percentage	Predicted Accidents	Required Officers
0	1382.012	691.0061
10	1384.989	692.4943
20	1387.965	693.9824
30	1390.941	695.4706

Conclusion

The analysis highlights key patterns and relationships among variables, such as the significant impact of **Composite Consumer Price Index** and **Traffic Density** on road traffic accidents. Logistic regression provides limited classification accuracy, while linear regression explains 21% of the variance in accidents. PCA reveals that 5 components explain 90% of the dataset's variance.