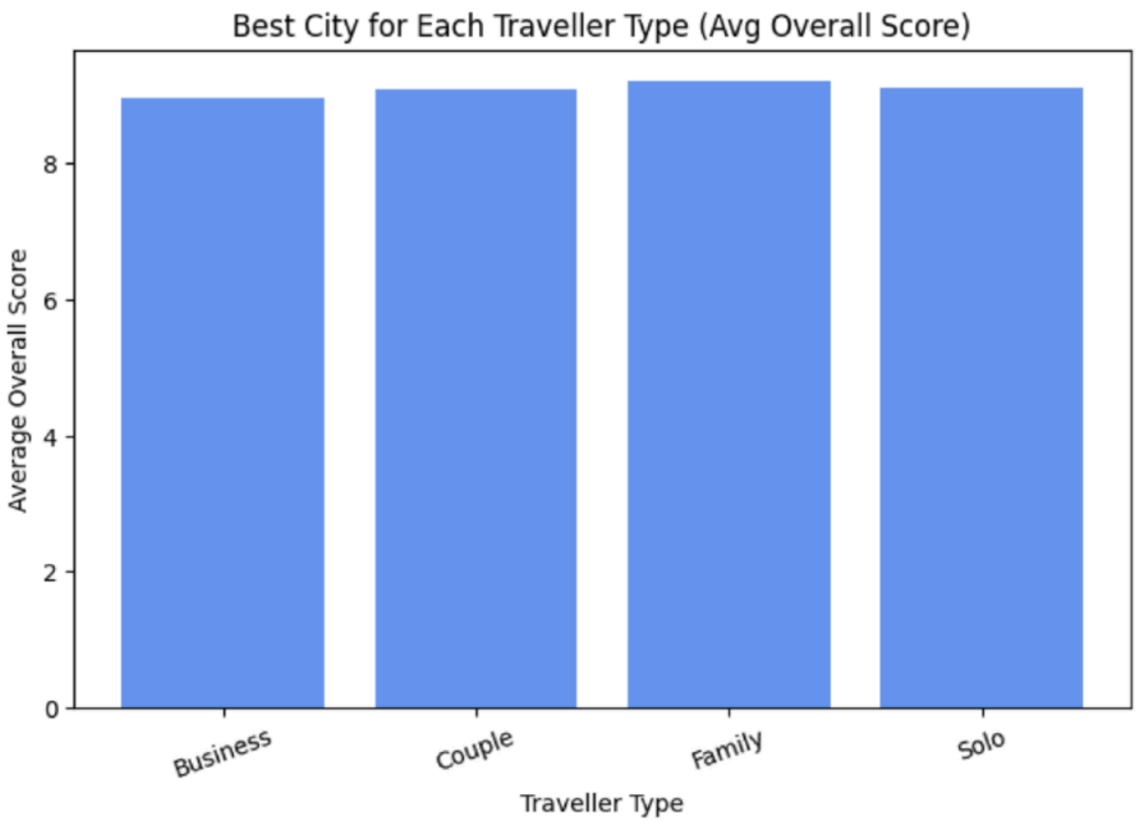
Milestone 1 report

Question 1

	traveller_type	city	score_overall
57	Family	Dubai	9.214381
75	Solo	Amsterdam	9.108454
25	Couple	Amsterdam	9.096989
7	Business	Dubai	8.965668



Question 2

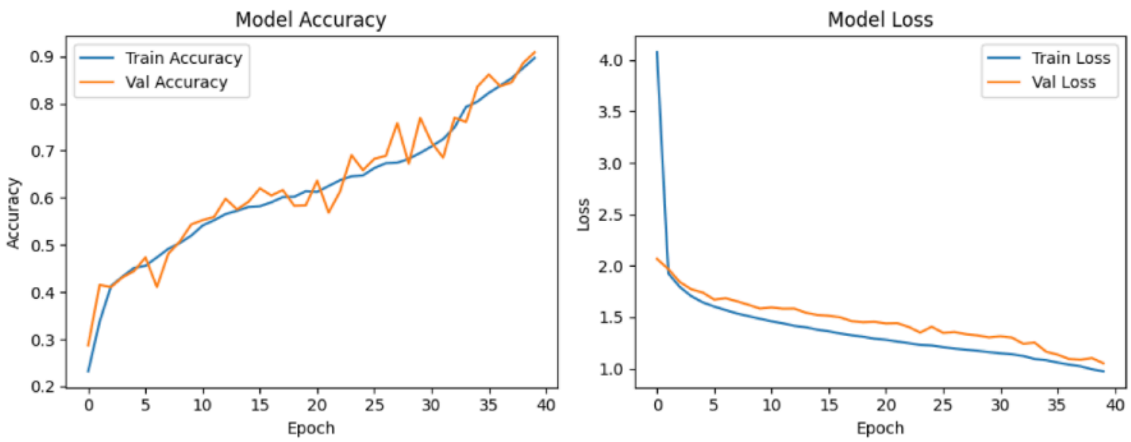
	age_group	country_x	score_value_for_money
5	18-24	Egypt	9.007317
19	18-24	Spain	8.768132
0	18-24	Argentina	8.689000
44	25-34	Spain	8.733259
43	25-34	South Korea	8.632800
37	25-34	Netherlands	8.542157
66	35-44	Singapore	8.795385
50	35-44	Argentina	8.696519
63	35-44	New Zealand	8.641079
94	45-54	Turkey	8.641778
89	45-54	Singapore	8.555914
79	45-54	China	8.547538
111	55+	New Zealand	8.686264
101	55+	Canada	8.621084
108	55+	Japan	8.524918

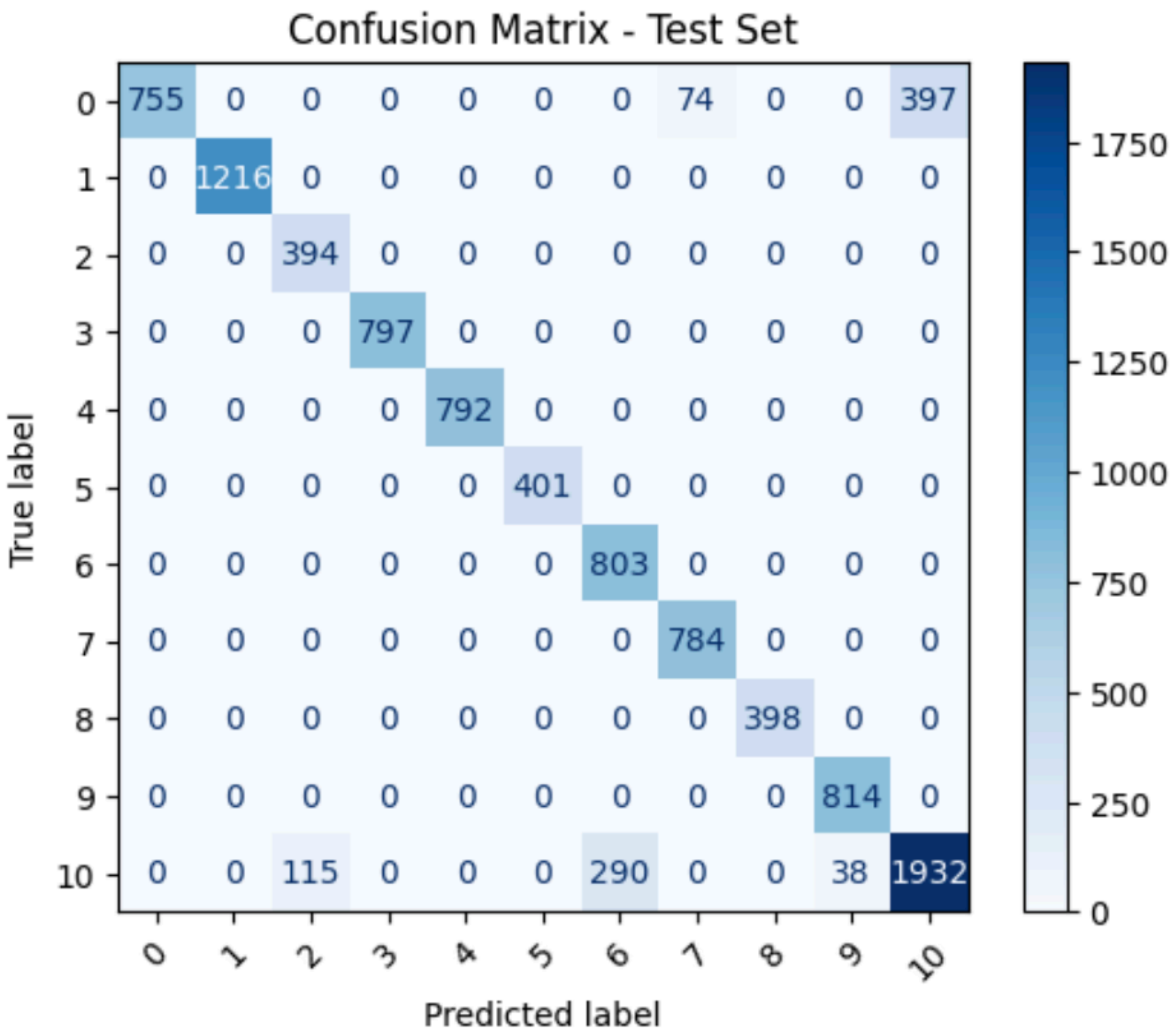
Model info

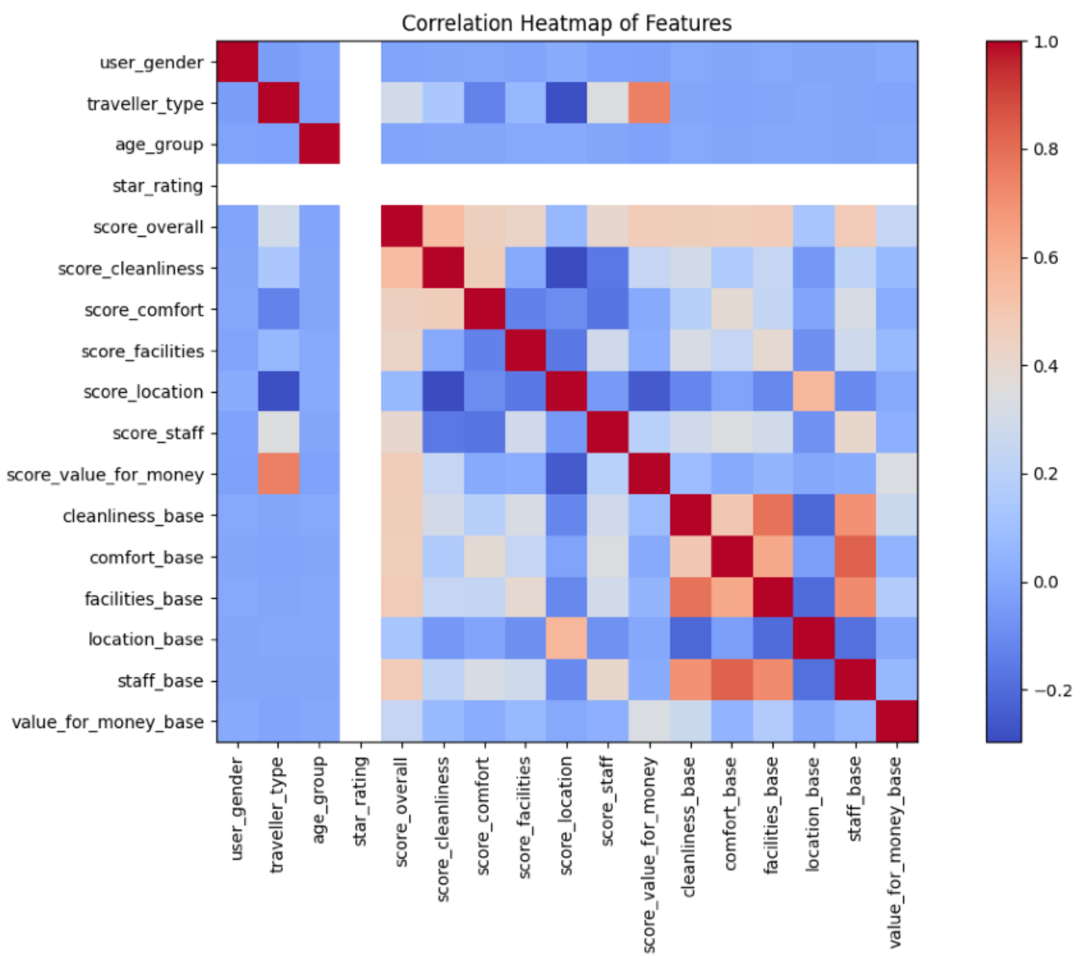
Model Architecture:

```
model_A = Sequential([
    Dense(256, activation='relu',
          input_shape=(X_train.shape[1],),
          kernel_regularizer=regularizers.l2(l2_lambda)),

    Dense(128, activation='relu',
          kernel_regularizer=regularizers.l2(l2_lambda)),

    Dense(64, activation='relu',
          kernel_regularizer=regularizers.l2(l2_lambda)),

    Dense(num_classes, activation='softmax')  # Output layer
])
```

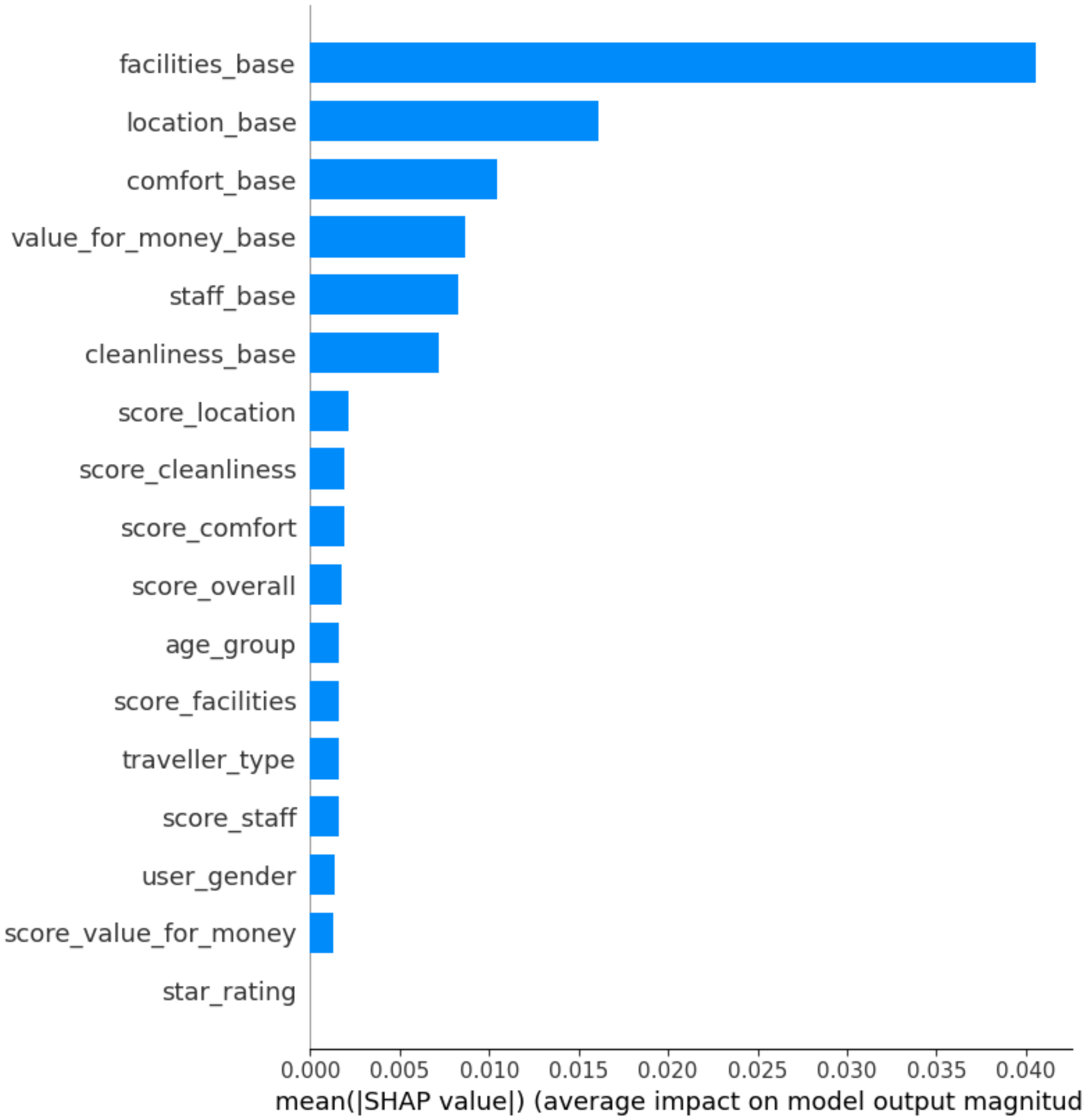
```
# Compile the model
model_A.compile(
    optimizer='adam',
    loss='categorical_crossentropy',
    metrics=['accuracy']
)
```

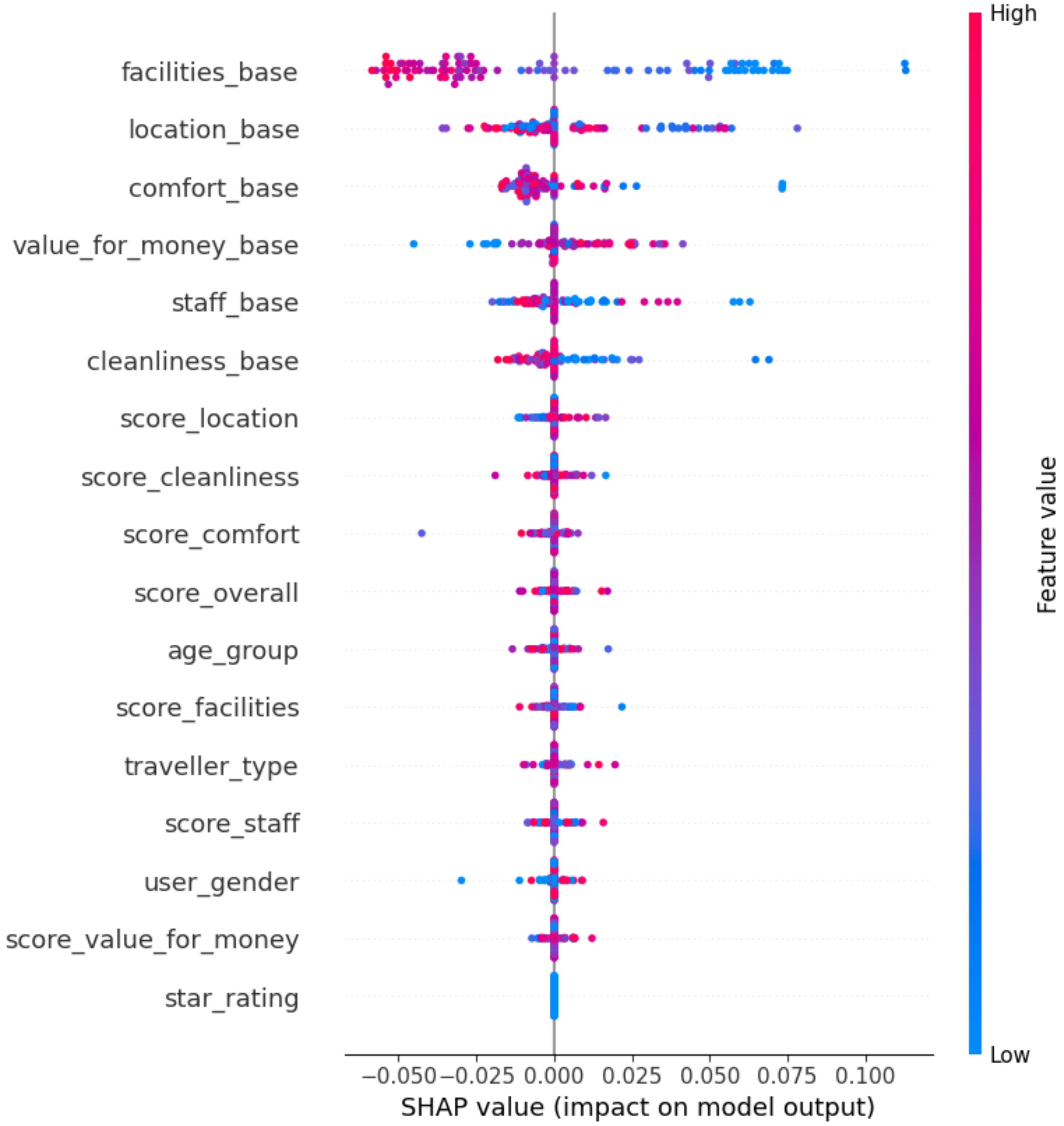



Confusion Matrix - Test Set
True label
Predicted label
755
1216
394
797
792
401
803
784
398
814
1932
74
397
115
290
38
1750
1500
1250
1000
750
500
250
0


Correlation Heatmap of Features
user_gender
traveller_type
age_group
star_rating
score_overall
score_cleanliness
score_comfort
score_facilities
score_location
score_staff
score_value_for_money
cleanliness_base
comfort_base
facilities_base
location_base
staff_base
value_for_money_base
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

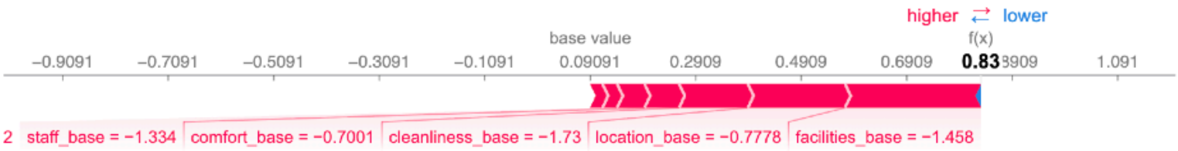
SHAP summary plot



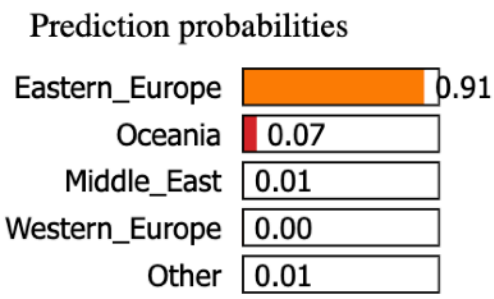


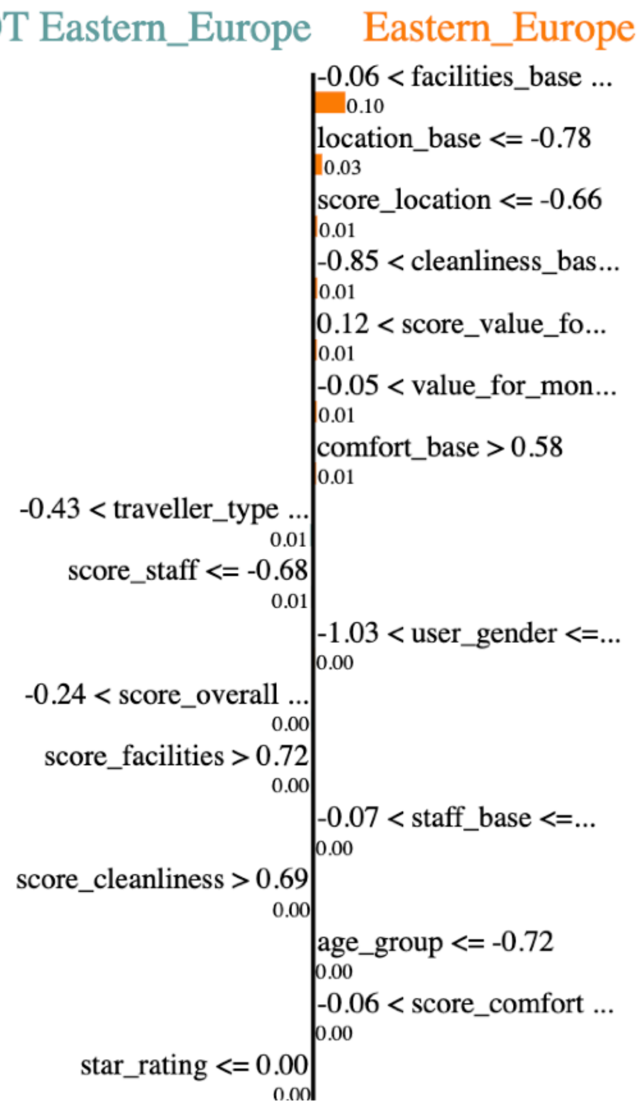

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 36ms/step
Predicted Class: North_America_Mexico
4/4 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step
```

Out[23]:



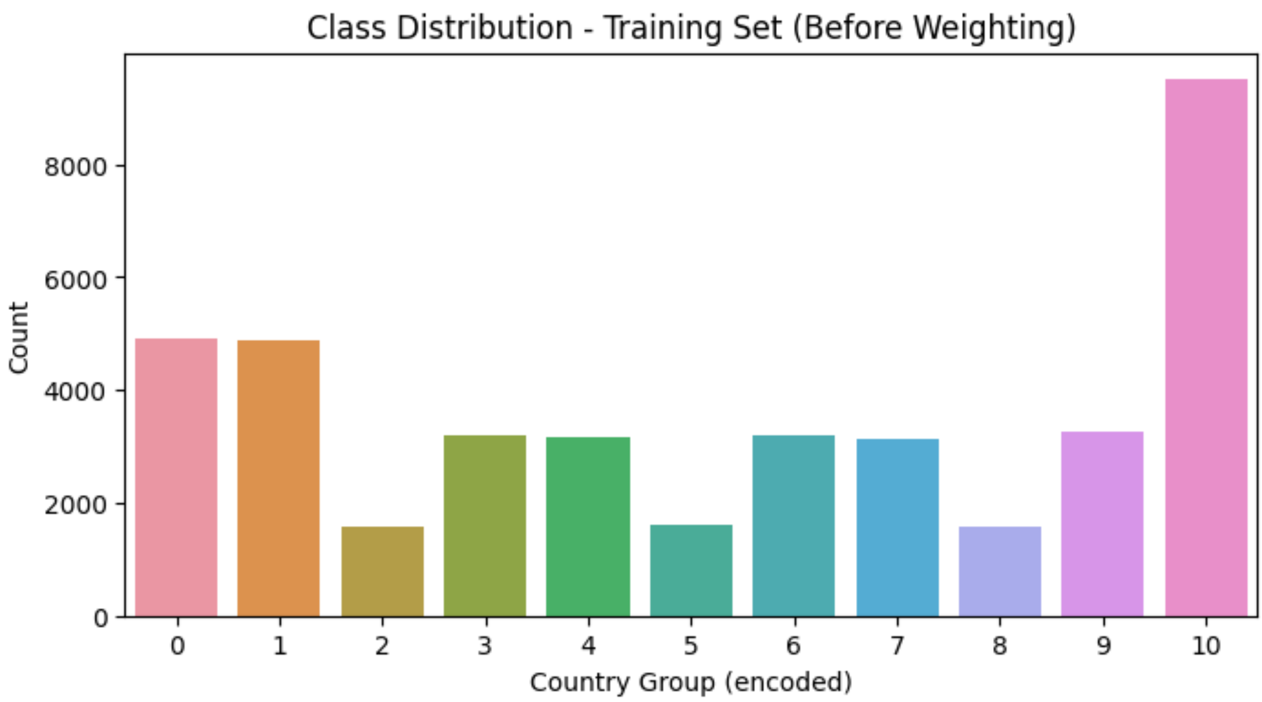
Lime





Feature	Value
facilities_base	0.29
location_base	-0.78
score_location	-1.84
cleanliness_base	0.04
score_value_for_money	0.69
value_for_money_base	0.34
comfort_base	1.01
traveller_type	0.53
score_staff	-0.68
user_gender	0.53
score_overall	0.31
score_facilities	1.72
staff_base	0.36

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 34ms/step
✅ Actual class index: 2
✅ Actual class name:  Eastern_Europe
🤖 Predicted class index: 2
🤖 Predicted class name:  Eastern_Europe
```



Features used:

user_gender traveller_type age_group star_rating score_overall score_cleanliness score_comfort score_facilities score_location score_staff score_value_for_money cleanliness_base comfort_base facilities_base location_base staff_base value_for_money_base

Why these features:

User Information Features

These reflect who the user is, which affects what regions they prefer.

Feature	Why it affects country_group
user_gender	Males vs females show different region preferences
traveller_type	Families prefer different regions than business travelers
age_group	Younger travelers tend to visit cheaper/hotter regions

Hotel Star Rating

Feature	Reason
star_rating	Luxury hotels are common in Europe/Middle East, budget in Asia/Africa

It strongly indicates economic development of the region.

Hotel Quality Scores (Base Features)

Features:
cleanliness_base, comfort_base, facilities_base, location_base, staff_base, value_for_money_base

They represent true hotel quality differences by region
Western Europe/USA → high comfort
Southeast Asia → high value for money
Africa → lower average scores
User Review Scores (Personal Experience)

Features:
score_overall, score_cleanliness, score_comfort, score_facilities, score_location, score_staff, score_value_for_money

These show how the user felt about that region:

Certain travelers consistently review certain regions high/low
Helps model learn where the user tends to stay