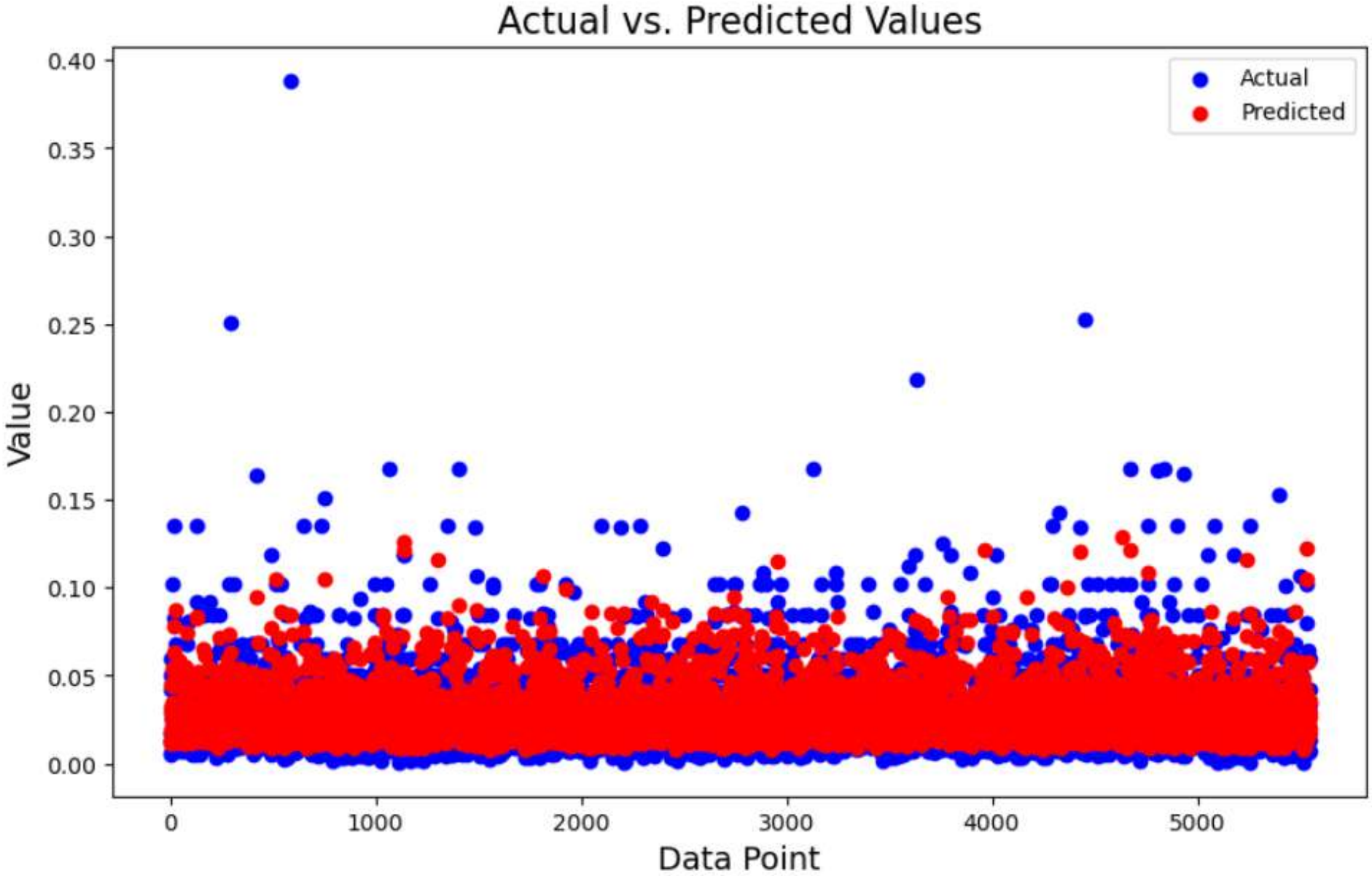
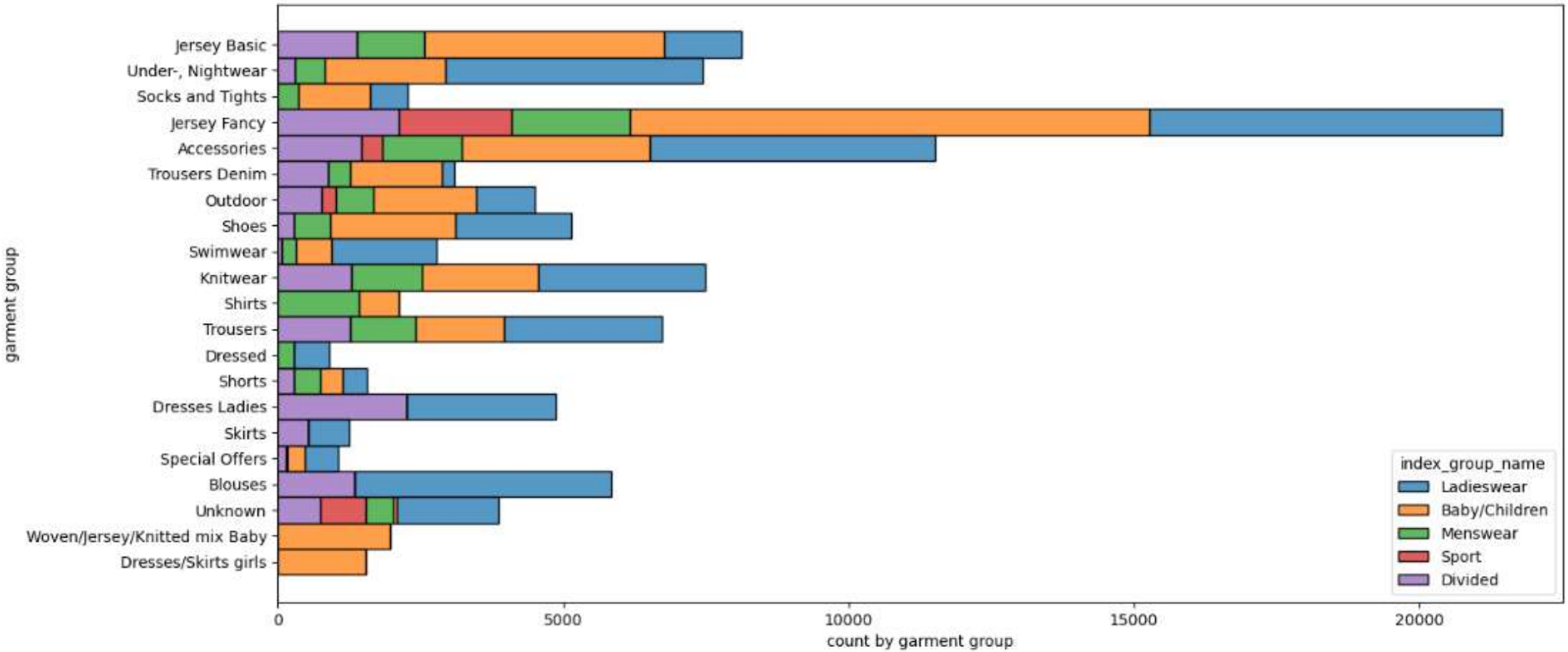
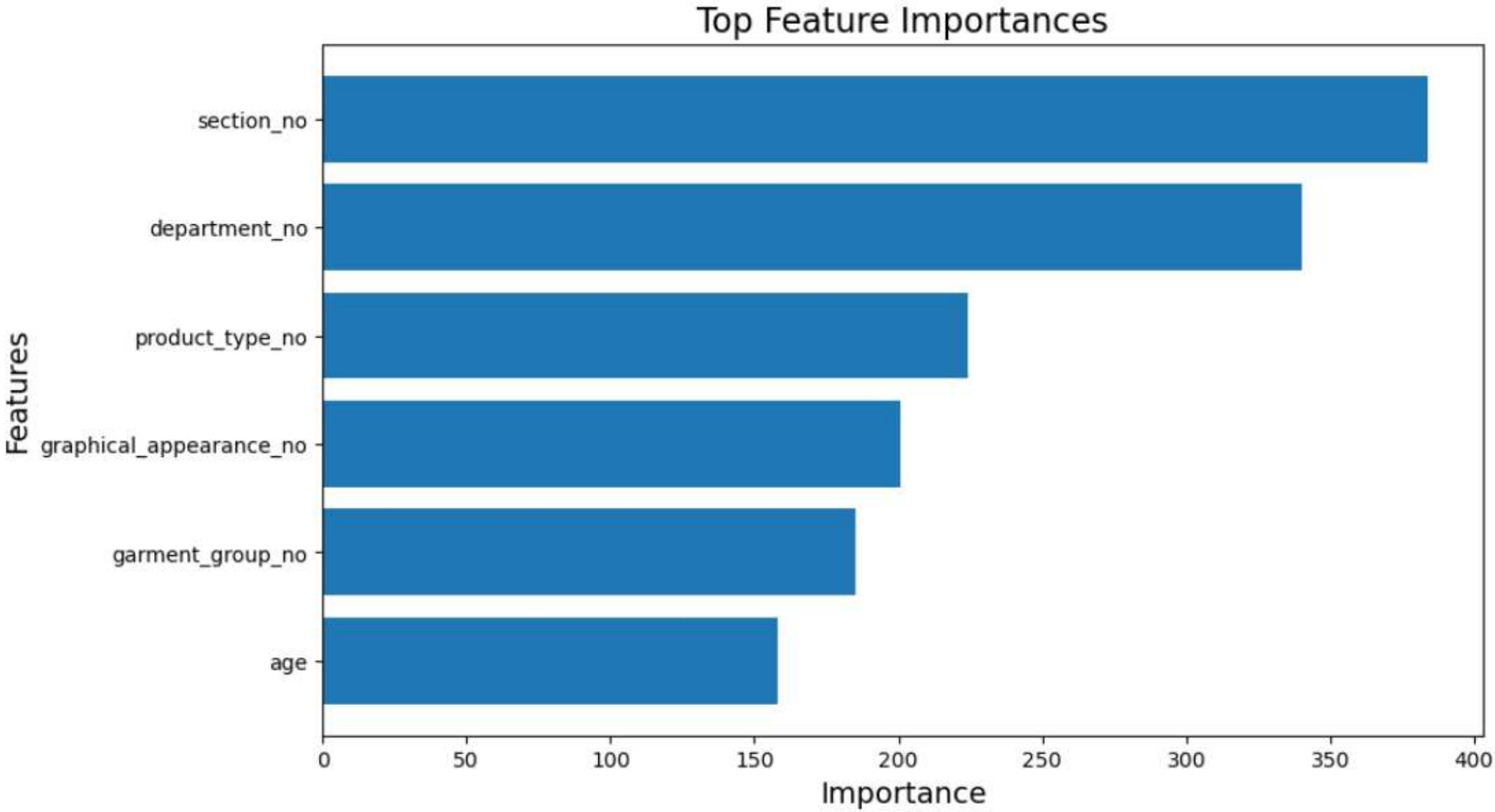
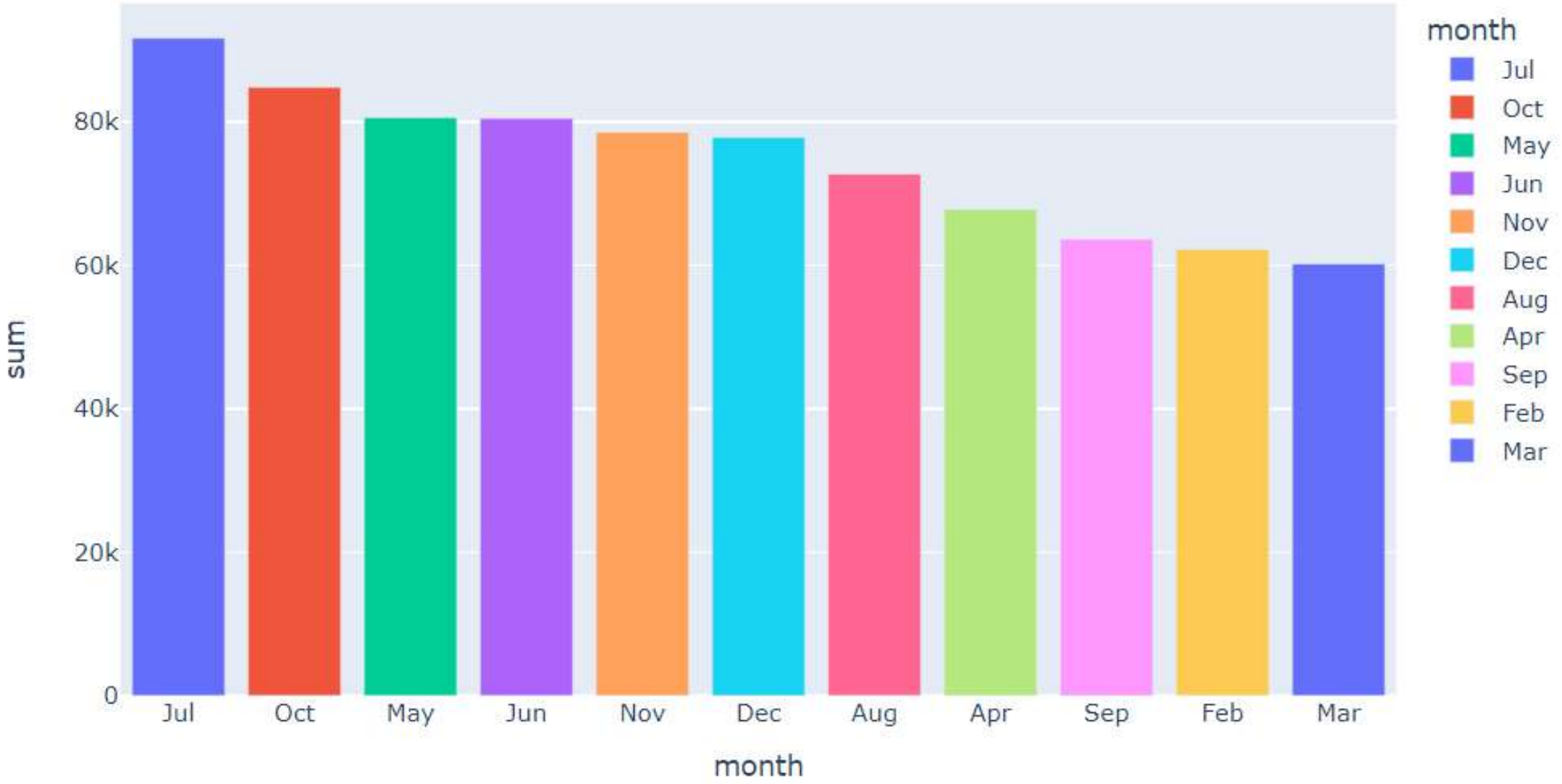

Actual vs. Predicted Values
Actual
Predicted
Value
Data Point
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
1000
2000
3000
4000
5000


Jersey Basic
Under-, Nightwear
Socks and Tights
Jersey Fancy
Accessories
Trousers Denim
Outdoor
Shoes
Swimwear
Knitwear
Shirts
Trousers
Dressed
Shorts
Dresses Ladies
Skirts
Special Offers
Blouses
Unknown
Woven/Jersey/Knitted mix Baby
Dresses/Skirts girls
garment group
0
5000
10000
15000
20000
count by garment group
index_group_name
Ladieswear
Baby/Children
Menswear
Sport
Divided

```
# Convert categorical columns to numeric using one-hot encoding
categorical_cols = ['product_code', 'product_type_name', 'colour_group_name', 'department_name', 'FN']
X_encoded = pd.get_dummies(X_encoded, columns=categorical_cols)

# Drop remaining columns with object data type
X_encoded = X_encoded.select_dtypes(exclude=['object'])

train_ratio = 0.8  # 80% for training, 20% for testing
train_size = int(train_ratio * len(merged_df))
X_train, X_test = X_encoded[:train_size], X_encoded[train_size:]
y_train, y_test = y[:train_size].values, y[train_size:].values  # Convert to NumPy arrays

# Define the LightGBM dataset
train_data = lgb.Dataset(X_train, label=y_train)

# Set the hyperparameters for LightGBM
params = {
    'objective': 'regression',
    'metric': 'rmse',
    'boosting_type': 'gbdt',
    'num_leaves': 31,
    'learning_rate': 0.05,
    'feature_fraction': 0.9,
    'bagging_fraction': 0.8,
    'bagging_freq': 5,
    'verbose': 0
}
# Train the LightGBM model
model = lgb.train(params, train_data, num_boost_round=100)

# Predict on the test set
y_pred = model.predict(X_test)

# Access valuable information from the trained model
feature importances = model.feature importance()
```


Top Feature Importances
section_no
department_no
product_type_no
graphical_appearance_no
garment_group_no
age
Features
0
50
100
150
200
250
300
350
400
Importance


Monthly sales
sum
80k
60k
40k
20k
0
Jul
Oct
May
Jun
Nov
Dec
Aug
Apr
Sep
Feb
Mar
month
month
Jul
Oct
May
Jun
Nov
Dec
Aug
Apr
Sep
Feb
Mar