Maximizing Profitability: Analyzing App vs. Website Performance for Customer Spending in an Online Clothing Store

Christopher Ereforokuma.
Evangel George.

Agenda

- Introduction
- Tools Used
- Data Preprocessing
- Exploratory Data Analysis
- Recommendations
- Building the Regression Model
- Coefficients and Intercept of the Model
- Model evaluation
- Prototyping with Streamlit
- User Interface
- Conclusion
- Thankyou

Introduction

This project aims to analyze customer behavior and preferences in an online clothing store with personalized in-store style and clothing advice sessions. The goal is to help the company determine whether to prioritize their mobile app or website to improve the customer experience and maximize profitability.

By analyzing the dataset, conducting predictive modeling using techniques like linear regression, and prototyping an interactive web application, valuable insights can be obtained to guide decision-making and resource allocation within the company.

Tools Used

- XLSTAT was used for exploratory data analysis and correlation analysis.
- Python and Scikit-learn were utilized for developing a robust linear regression model.
- Streamlit was employed for prototyping an interactive web application for real-time spending predictions.
- Joblib facilitated model serialization and deserialization for efficient deployment.
- The project integrated data analysis, modeling, and interactive prototyping to enable informed decision-making in the online clothing store.

Data Pre-processing

Usually, meticulous steps would be taken to ensure a clean and well-organized dataset, including checks for duplicates, missing values, and inconsistencies. However, the data, sourced from Kaggle, was clean as a whistle.

Train-Test Split:

The dataset was divided into training and testing sets to enable the implementation of linear regression using scikit-learn.

This split allows evaluating the model's performance on unseen data, enhancing its ability to generalize and predict customer spending accurately.

```
from sklearn.model_selection import train_test_split
X = df[['Avg. Session Length', 'Time on App', 'Time on Website', 'Length of Membership']]
y = df['Yearly Amount Spent']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.3, random_state=0)
```

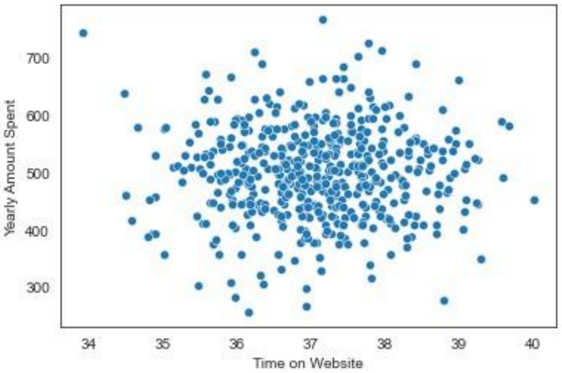
Exploratory Data Analysis

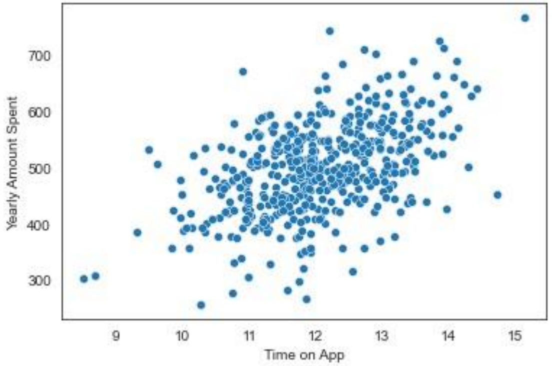
In the exploratory data analysis conducted for this project, a focus was placed on examining the correlation between different features of the dataset. Specifically, the correlations between time spent on the app, time spent on the website, length of membership, and the yearly amount spent by customers were analyzed.

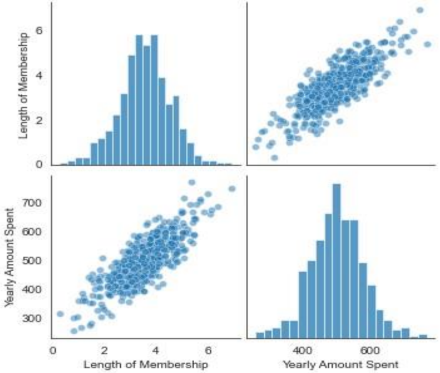
	Avg. Session Length	Time on App	Time on Website	Length of Membership	Yearly Amount Spent
Avg. Session Length	1.000000	-0.027826	-0.034987	0.060247	0.355088
Time on App	-0.027826	1.000000	0.082388	0.029143	0.499328
Time on Website	-0.034987	0.082388	1.000000	-0.047582	-0.002641
Length of Membership	0.060247	0.029143	-0.047582	1.000000	0.809084
Yearly Amount Spent	0.355088	0.499328	-0.002641	0.809084	1.000000

Exploratory Data Analysis

Time spent on the app shows a positive correlation with yearly spending, indicating that higher app usage is associated with increased spending.
There is negligible correlation between time spent on the website and yearly spending, suggesting that website usage has minimal impact on customer spending behavior.
A positive relationship exists between length of membership and yearly spending, highlighting the influence of customer loyalty and engagement on higher spending levels.







Recommendations

1.Prioritize the Mobile App: Enhance the mobile app experience to increase customer engagement and spending. Improve user interface, personalization features, and checkout process to provide a convenient and enjoyable shopping experience.

2. **Foster Customer Loyalty**: Implement strategies to enhance customer retention and loyalty. Offer exclusive perks, rewards programs, personalized recommendations, and excellent customer service to encourage repeat purchases and increase customer lifetime value.

Building the Regression Model

During the regression model building phase, the Scikit-Learn library was utilized to create an effective predictive model. The focus was on predicting customer spending based on various features.

```
from sklearn.linear_model import LinearRegression
```

```
lm = LinearRegression()
lm.fit(X_train, y_train)
```

```
LinearRegression()
```

The code snippet demonstrates the use of the Linear Regression class from Scikit-Learn, where the model was trained using the training data.

Coefficients and Intercept of the Model

The estimated coefficients and intercept provided insights into the relationships and biases within the data, forming the foundation of the predictive model.

```
coef_df = pd.DataFrame(lm.coef_, X.columns, columns=['Coeffients'])
coef_df
```

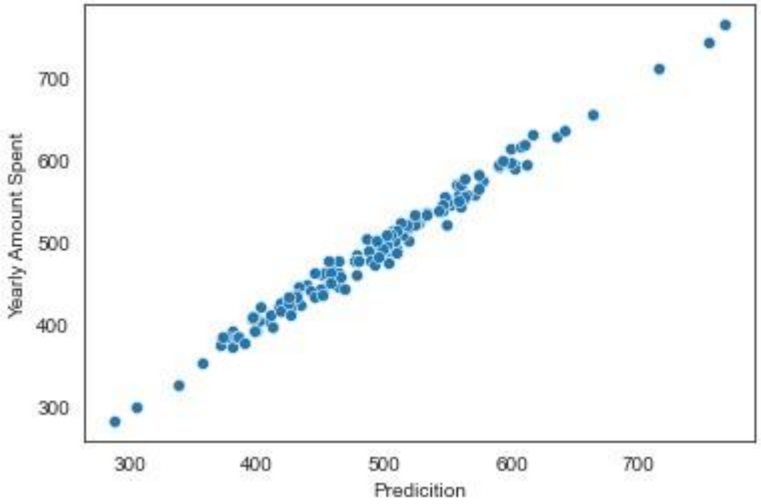
	Coeffients
Avg. Session Length	25.767530
Time on App	38.800394
Time on Website	-0.018041
Length of Membership	61.852568

Model Evaluation

The model evaluation shows strong predictive performance with low errors (MAE: 7.85, MSE: 94.56, RMSE: 9.72) and a high explanatory power (R-squared: 0.98). It effectively predicts customer spending, making it a valuable tool for decision-making.

```python
from sklearn.metrics import mean_squared_error as mse
from sklearn.metrics import mean_absolute_error as mae
from sklearn.metrics import r2_score

print('Mean Absolute Error:', mae(y_test, predictions))
print('Mean Squared Error:', mse(y_test, predictions))
print('Root Mean Squared Error:', mse(y_test, predictions, squared=False))
print('R-squared:', r2_score(y_test, predictions))
```

```
Mean Absolute Error: 7.851377170861448
Mean Squared Error: 94.55779479273275
Root Mean Squared Error: 9.72408323662096
R-squared: 0.9849262667370623
```



Prototyping with Streamlit

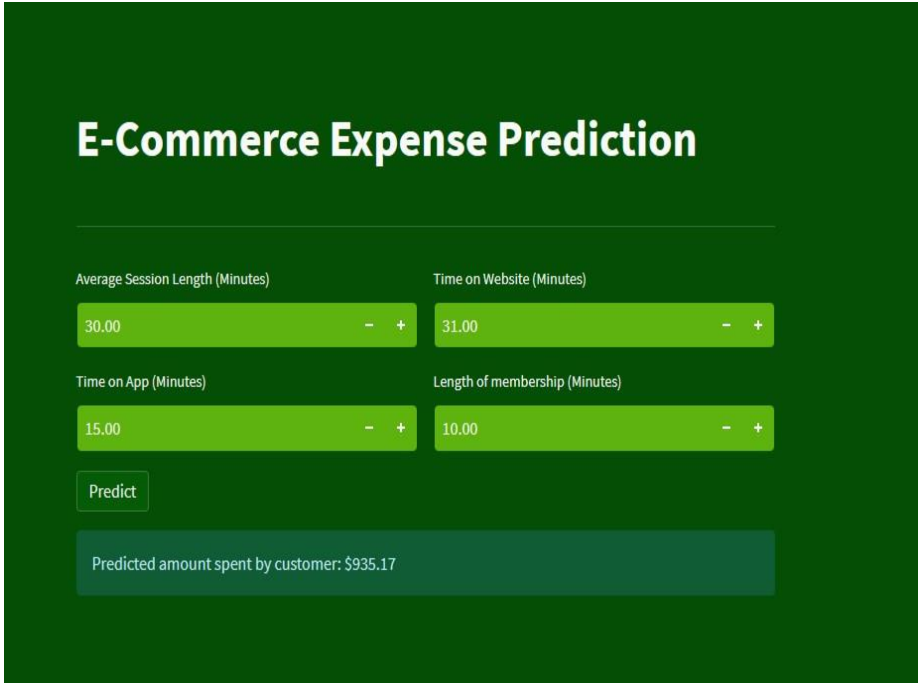
The machine learning model was prototyped using Streamlit, a Python library for creating interactive web applications.

To begin, the trained linear regression model was integrated into a Streamlit application. Streamlit’s simple syntax and built-in components allowed for easy incorporation of the model’s functionalities.

Streamlit's user-friendly interface allowed for easy integration of the trained linear regression model. The application accepts user input for session duration, app/website usage time, and membership length.

```python
import streamlit as st
import joblib
import numpy as np
from sklearn.linear_model import LinearRegression

model = joblib.load('model.joblib')

st.markdown(
    """
    <style>
    .reportview-container {
        background: url("https://images.app.goo.gl/LFCobouKtT7oZ7Qv7")
    }

   .sidebar .sidebar-content {
        background: url("https://images.app.goo.gl/LFCobouKtT7oZ7Qv7")
    }
    </style>
    """,
    unsafe_allow_html=True
)

st.markdown('# E-Commerce Expense Prediction')
st.markdown('----')
col1, col2 = st.columns(2)

with col1:
    sess = st.number_input('Average Session Length (Minutes)')
    app_time = st.number_input('Time on App (Minutes)')

with col2:
    web_time = st.number_input('Time on Website (Minutes)')
    mem_length = st.number_input('Length of membership (Minutes)')

if st.button('Predict'):
    sample = np.array([sess, app_time, web_time, mem_length]).reshape(1, -1)
    prediction = model.predict(sample)[0]
    prediction = f'${prediction:.2f}'
    st.info(f'Predicted amount spent by customer: {prediction}')
```

User Interface



Streamlit's interactive features, such as sliders, dropdown menus, and text input fields, made it possible to create an engaging user interface for exploring different scenarios and obtaining immediate predictions. The real-time nature of the application allows users to experiment with various input values and observe the corresponding changes in the predicted results.

Conclusion

The project successfully analyzed customer behavior and preferences in an online clothing store and provided valuable insights for decision-making. The findings revealed that the mobile app played a significant role in driving higher spending, indicating the importance of enhancing the app experience. Additionally, customer loyalty, as reflected in the length of membership, showed a positive correlation with annual spending, emphasizing the need for strategies to foster loyalty. By investing in the mobile app and implementing customer retention initiatives, the company can increase customer engagement, drive higher spending, and ultimately boost profitability.

Furthermore, the prototype of the machine learning model using Streamlit provided an interactive web application that allows users to input their session duration, app/website usage time, and membership length to obtain real-time predictions of their yearly spending. The user-friendly interface of Streamlit enabled easy integration of the model's functionalities, providing an intuitive and engaging experience for exploring different scenarios.

Thank you

Christopher Ereforokuma
cereforokuma@gmail.com
https://github.com/fero-chris94

Evangel George
georgeevangel39@gmail.com
https://github.com/EvangelGeorge