

LOGISTIC REGRESSION LEAD SCORE CASE STUDY ASSIGNMENT

By JALPA CHAUHAN

Assignment-based Content

- ❑ **Problem statement**
- ❑ **Problem approach**
- ❑ **EDA**

- ❑ **Correlations**
- ❑ **Model Evaluation**
- ❑ **Observations**
- ❑ **Conclusion**

Problem Statement

- **X Education, an education company, offers online courses to professionals. On any given day, many professionals interested in these courses visit their website and fill out a form, after which they are considered as leads by the company.**
- **After acquiring these leads, the sales team starts reaching out through calls, emails, etc. However, only a small percentage of these leads get converted into actual customers, with most remaining unconverted.**
- **Typically, X Education experiences a lead conversion rate of around 30%. This means that for every 100 leads acquired, only about 30 will end up converting into customers. To improve this process, the company aims to identify the leads with the highest potential, known as "Hot Leads."**
- **By successfully identifying these Hot Leads, the sales team will be able to focus their efforts on the most promising prospects, leading to a higher conversion rate, rather than wasting time contacting all leads equally.**

Business Objectives

X Education wants us to build a model that assigns a lead score from 0 to 100 for each lead, helping them identify "Hot Leads" and improve their conversion rate. The CEO aims to achieve a target conversion rate of 80%, up from the current 30%.

The model should:

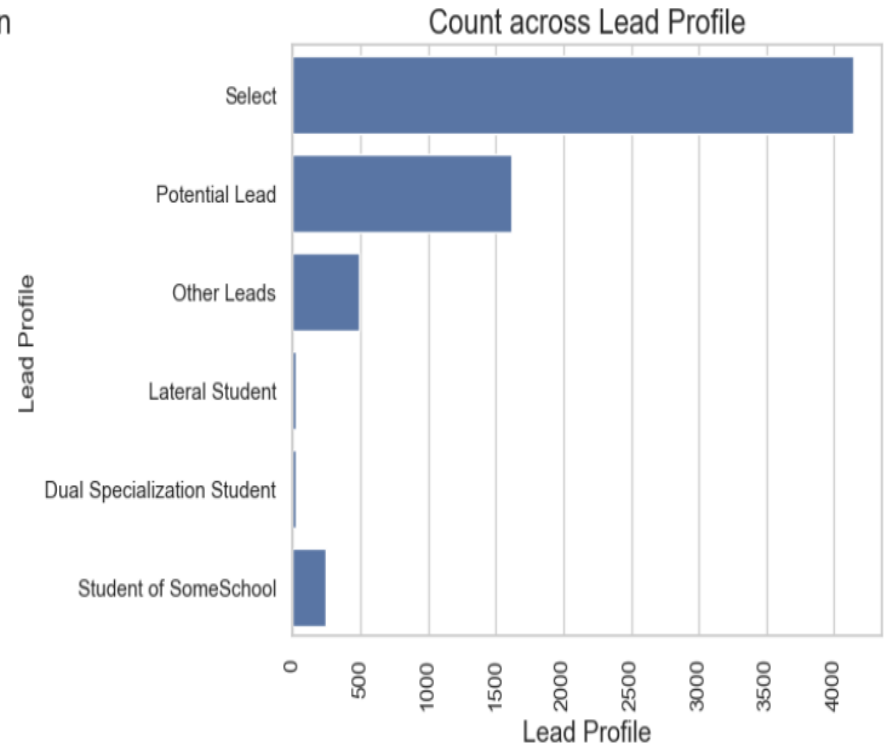
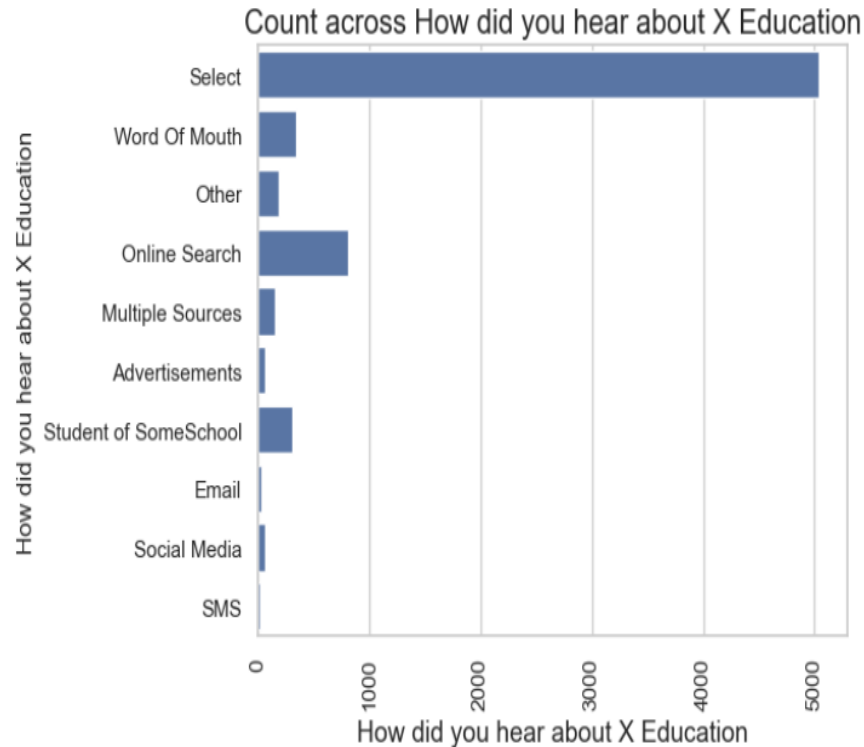
- Predict the likelihood of lead conversion and assign a score.
- Help prioritize resources, especially during peak times, to maximize efficiency.
- Offer strategies for sustaining or improving conversion rates after reaching the 80% goal.

The approach includes collecting lead data, creating relevant features, using machine learning to predict conversions, and evaluating the model's performance to ensure scalability and adaptability to future needs.

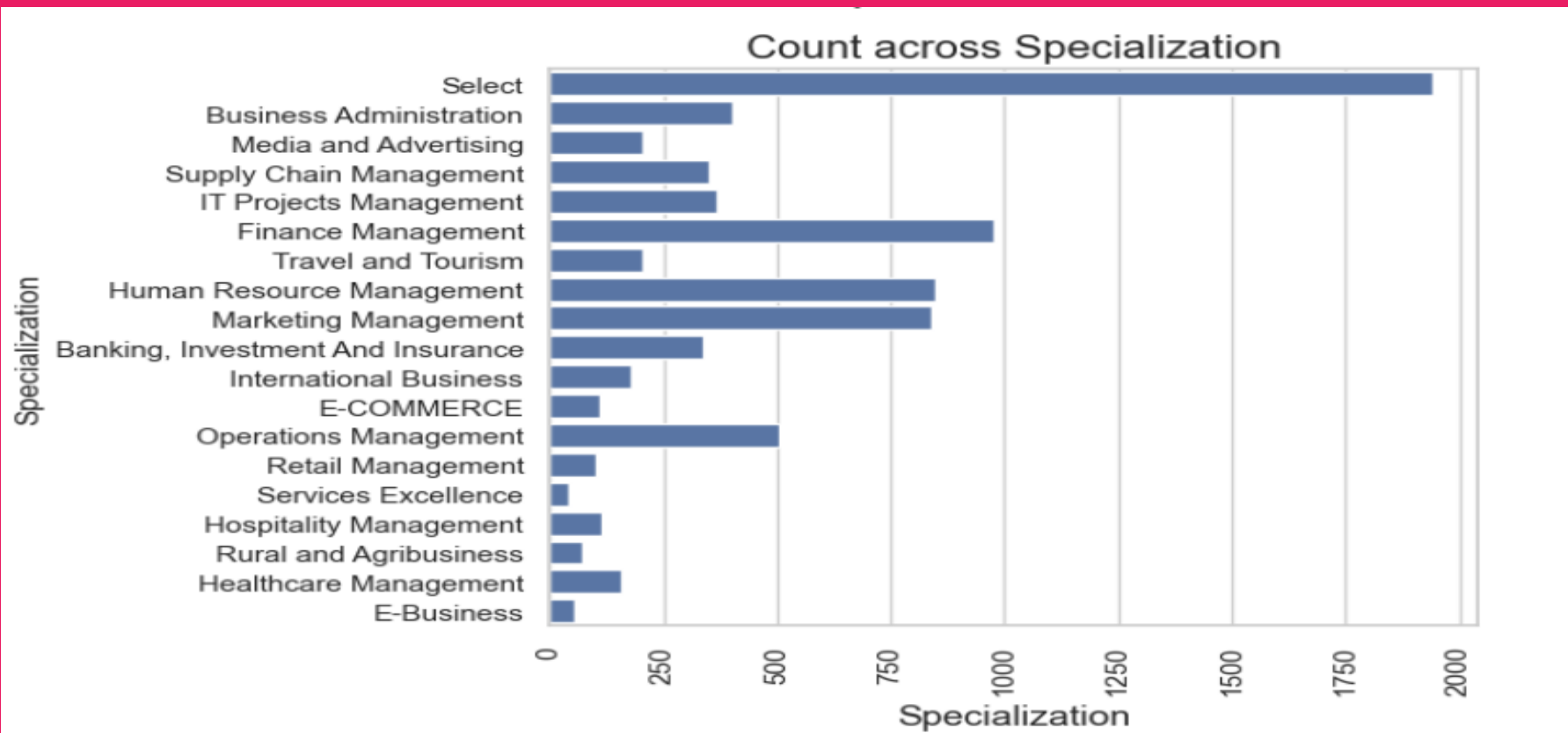
Problem Approach

1. Read and understand the data
2. Clean the data
3. Prepare the data for Model Building
4. Model Building
5. Model Evaluation
6. Making Predictions on the Test Set

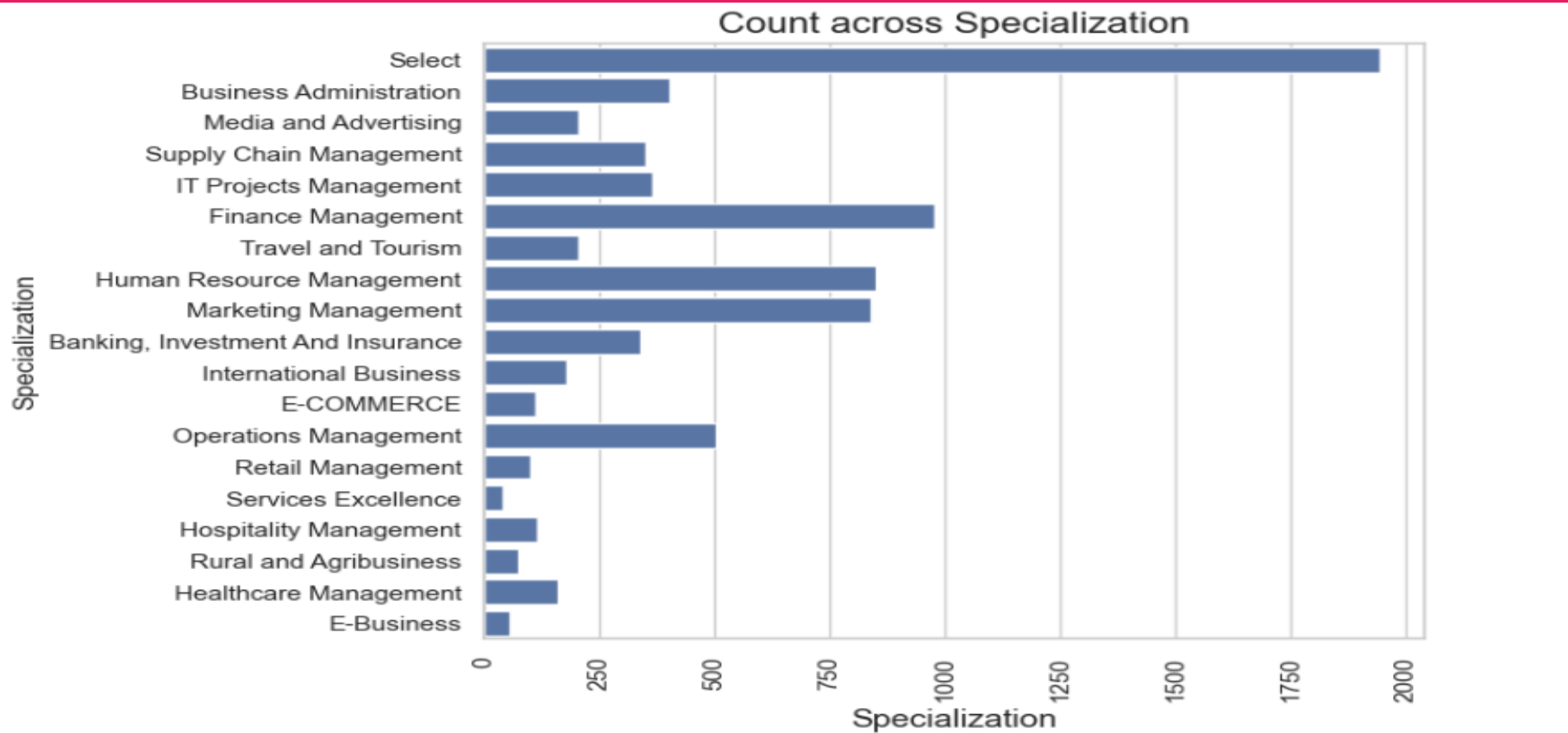
EDA – Data Cleaning



EDA – Data Cleaning

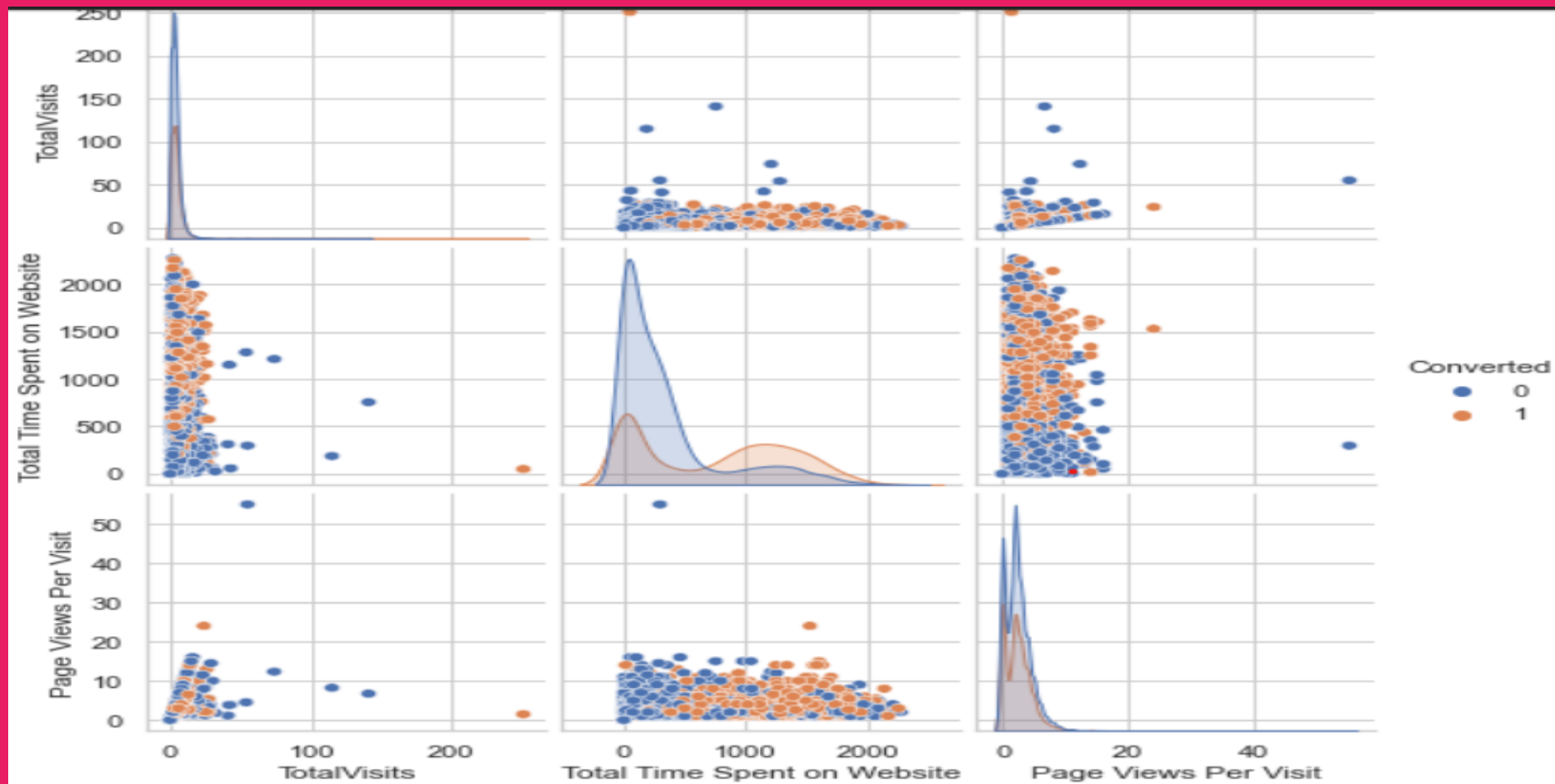


Specialization

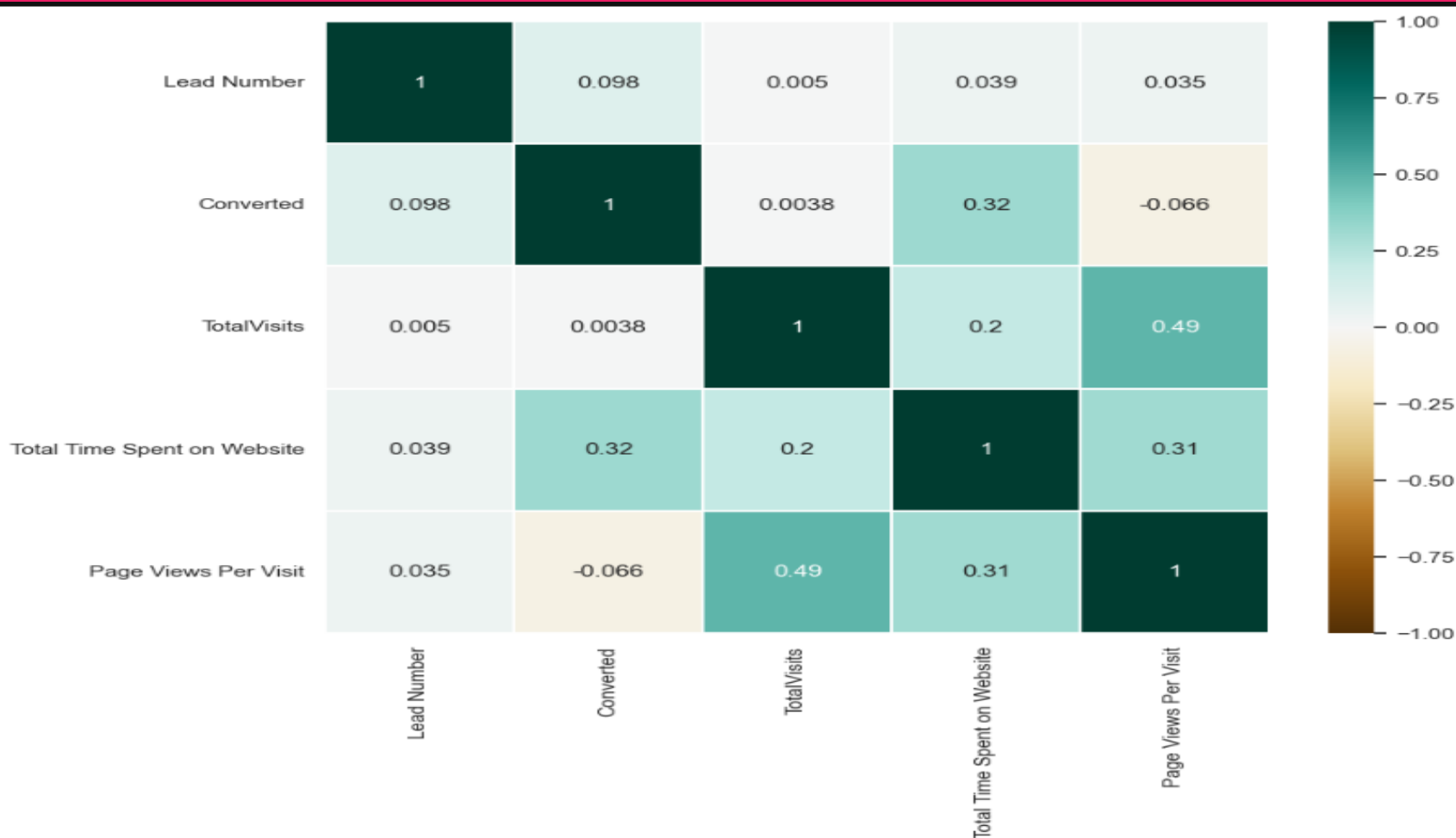


Visualizing Features

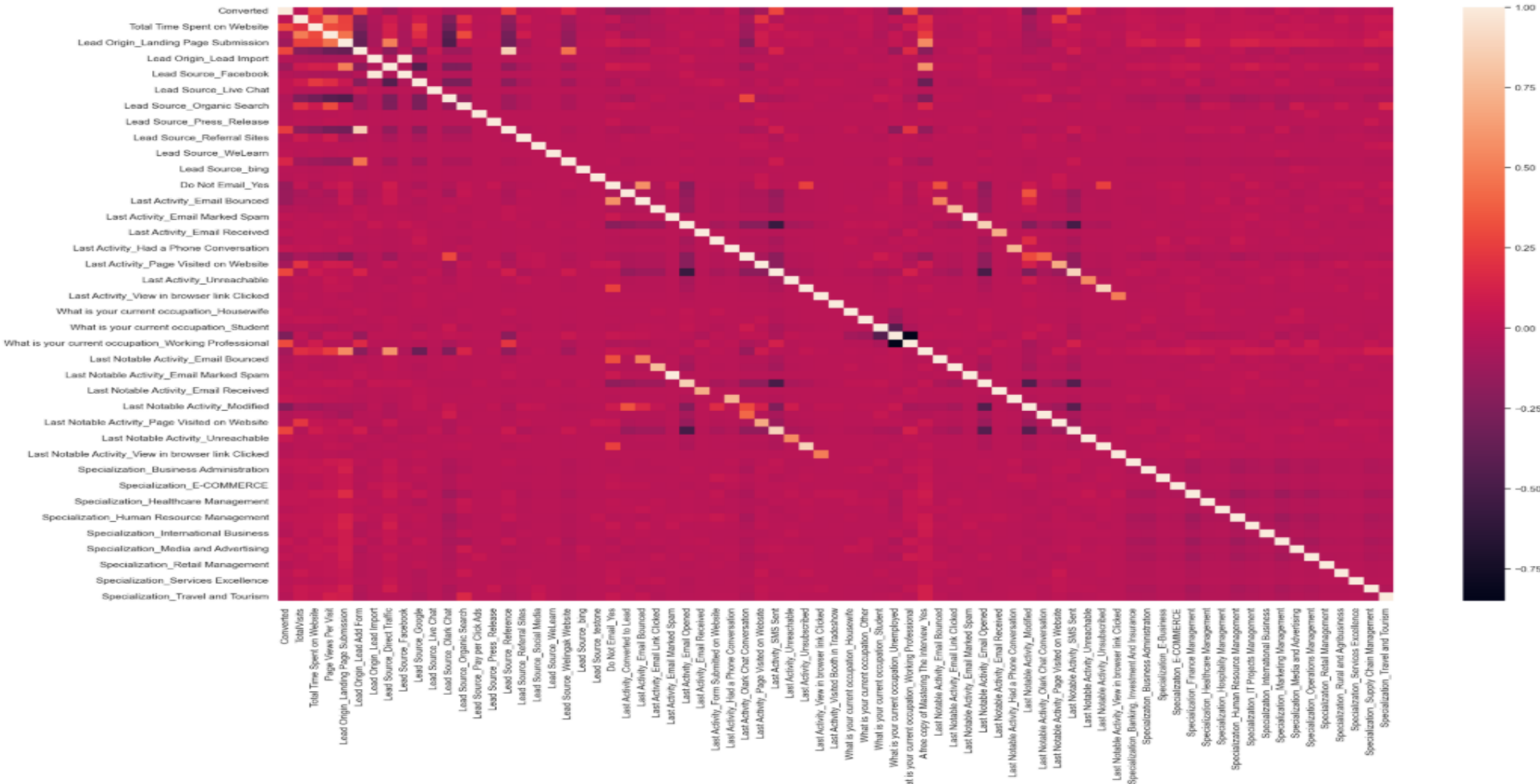
Lead Source & Lead origin



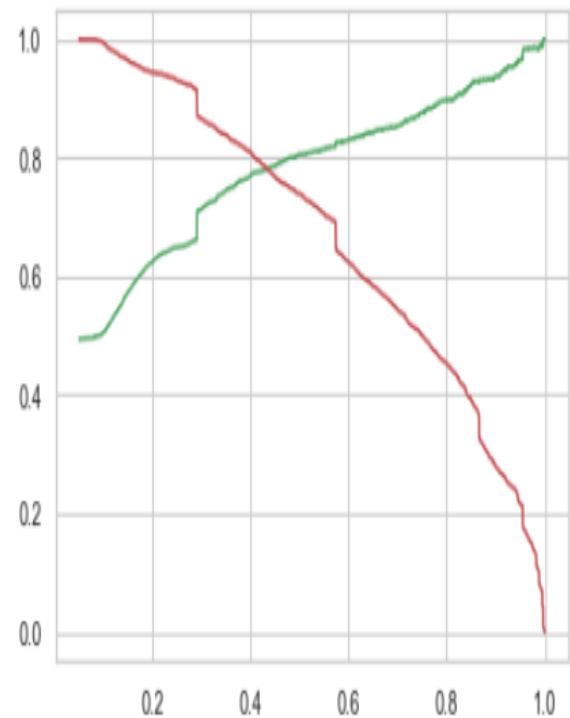
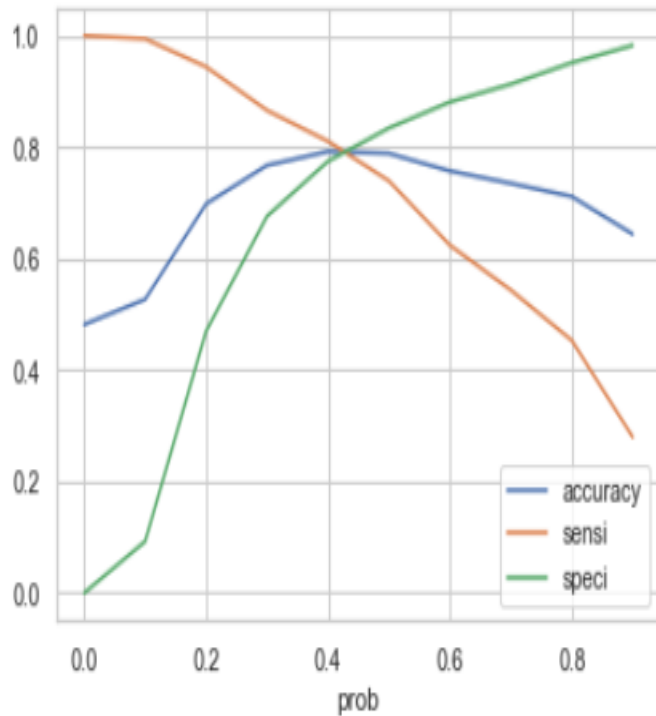
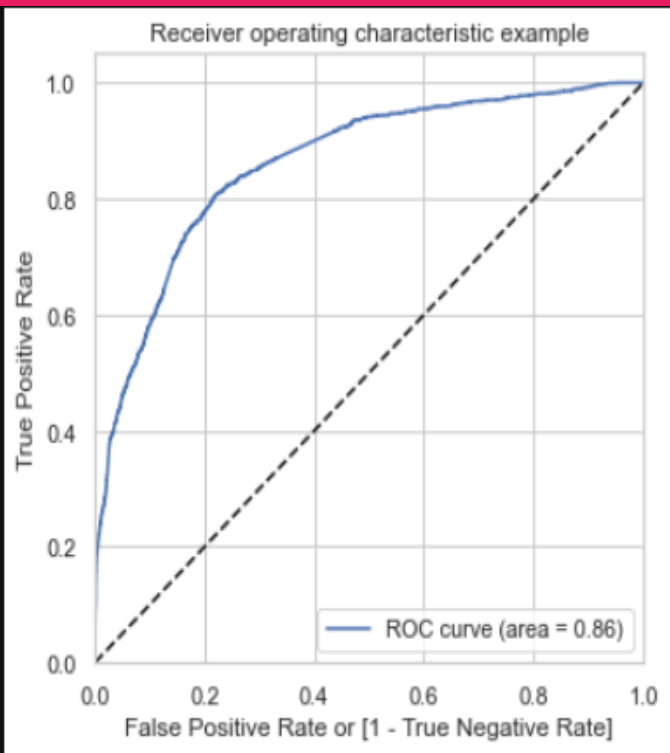
Correlation



Correlation



Model Evaluation - ROC Curve



Performance Metrics:

- **Accuracy:** 80%
- **Sensitivity:** 77%
- **Specificity:** 80%

Final Features Used in the Model:

1. **Lead Source_Olark Chat:** Indicates if the lead came from a chat interaction on the website.
2. **Specialization_Others:** Specifies the lead's specialization (if they fall into the 'Others' category).
3. **Lead Origin_Lead Add Form:** Indicates if the lead came from the "Lead Add Form" submission.
4. **Lead Source_Welingak Website:** Identifies whether the lead came from the Welingak website.
5. **Total Time Spent on Website:** Tracks how much time the lead spent on the website.
6. **Lead Origin_Landing Page Submission:** Indicates whether the lead originated from a landing page submission.
7. **What is your current occupation_Working Professionals:** Identifies the lead's occupation (whether they are a working professional).
8. **Do Not Email:** A feature indicating whether the lead opted out of receiving emails.



THANK
YOU

The image is a rectangular card with a white border. In the center is a dark maroon oval. Inside the oval, the words 'THANK' and 'YOU' are written in a white, serif, all-caps font, stacked vertically. Surrounding the oval is a delicate floral wreath. The wreath includes several large, light pink roses with yellow centers, a smaller dark pink rose, and clusters of tiny purple flowers. Green leaves and stems are interspersed throughout the floral arrangement. The background of the card is a soft, light pink with subtle watercolor-style textures and faint yellow speckles.