

LEAD SCORING PRESENTATION

**TEAM -1 VAISHNAVI SHARMA
2. SHASHANK NAINWAL
3 SHASHIKIRAN
PEETHALA**



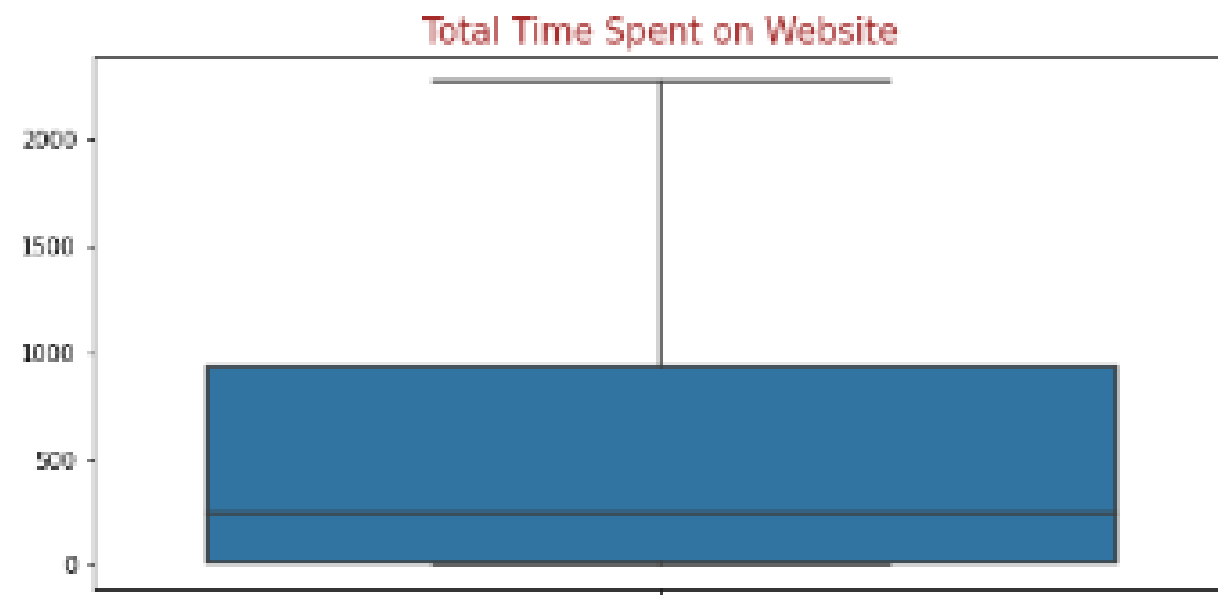
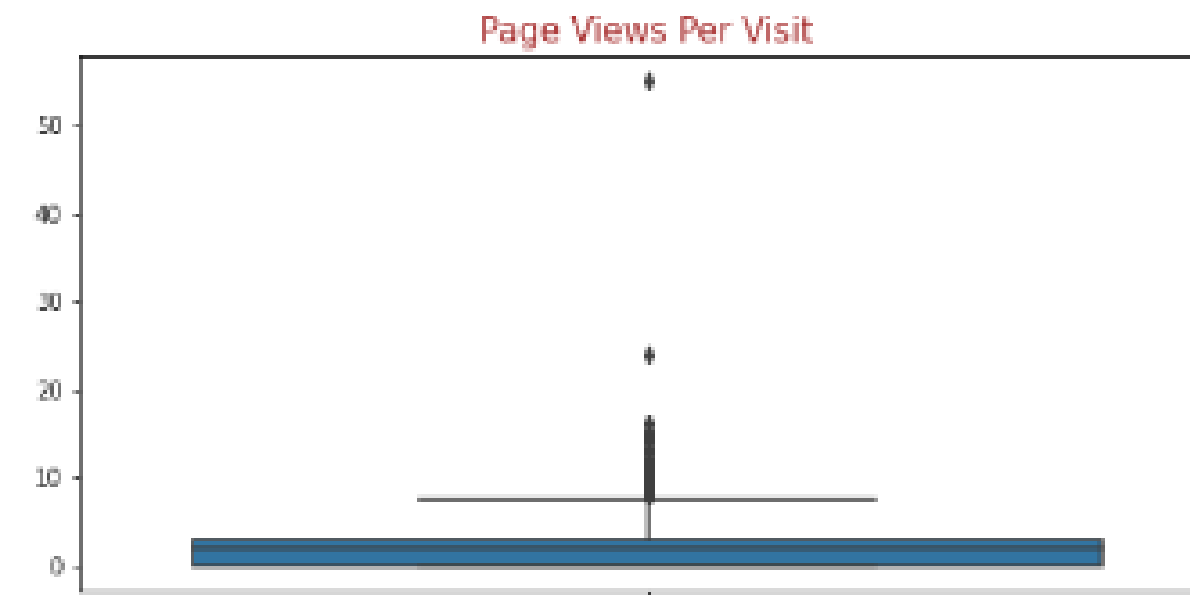
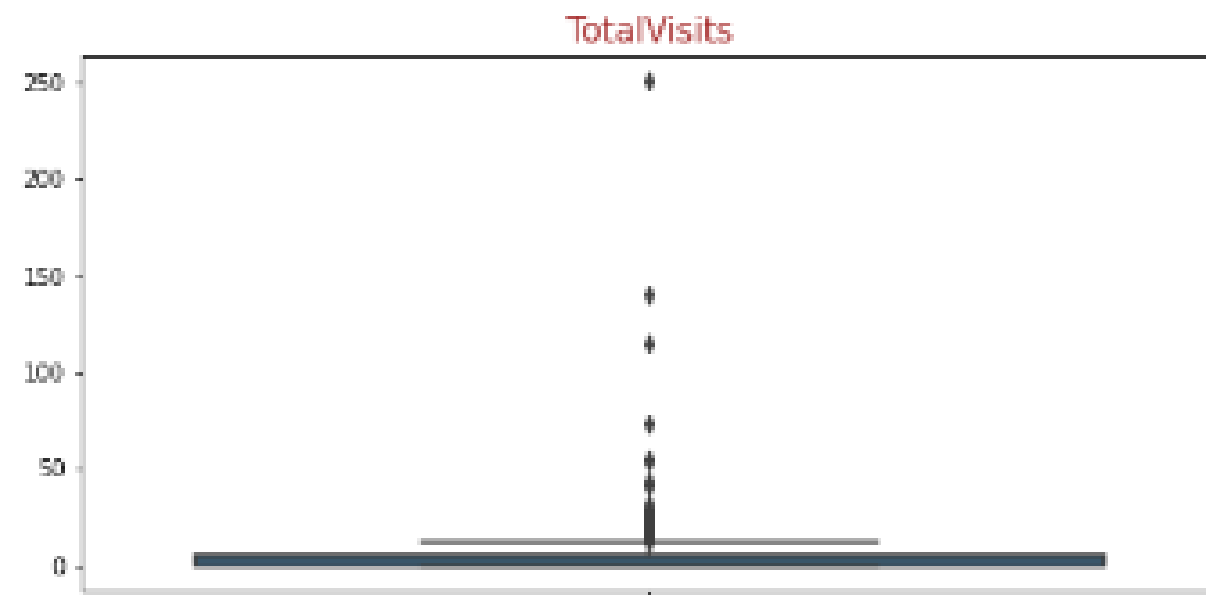
INSIGHTS FROM DATA

Insights:

- **City:** City has 39.71 % missing values. Imputing missing values with Mumbai will make the data more skewed. Skewness will later cause bias in the model. Hence City column can be dropped.
- **Specialization:** Specialization has 36.58 % missing values. The specialization selected is evenly distributed. Hence imputation or dropping is not a good choice. We need to create additional category called 'Others'.
- **Tags:** Tags has 36.29 % missing values. Tags are assigned to customers indicating the current status of the lead. Since this is current status, this column will not be useful for modeling. Hence it can be dropped.
- **What matters most to you in choosing a course:** This variable has 29.32 % missing values. 99.95% customers have selected 'better career prospects'. This is massively skewed and will not provide any insight.
- **What is your current occupation:** We can impute the missing values with 'Unemployed' as it has the most values. This seems to be an important variable from business context, since X Education sells online courses and unemployed people might take this course to increase their chances of getting employed.
- **Country:** X Education sells online courses and appx 96% of the customers are from India. Does not make business sense right now to impute missing values with India. Hence Country column can be dropped.
- **Last Activity:** "Email Opened" is having highest number of values and overall missing values in this column is just 1.11%, hence we will impute the missing values with label 'Email Opened'.
- **Lead Source:** "Google" is having highest number of occurrences and overall nulls in this column is just 0.39%, hence we will impute the missing values with label 'Google'.

OUTLIER ANALYSIS

Checking Outliers using Boxplot

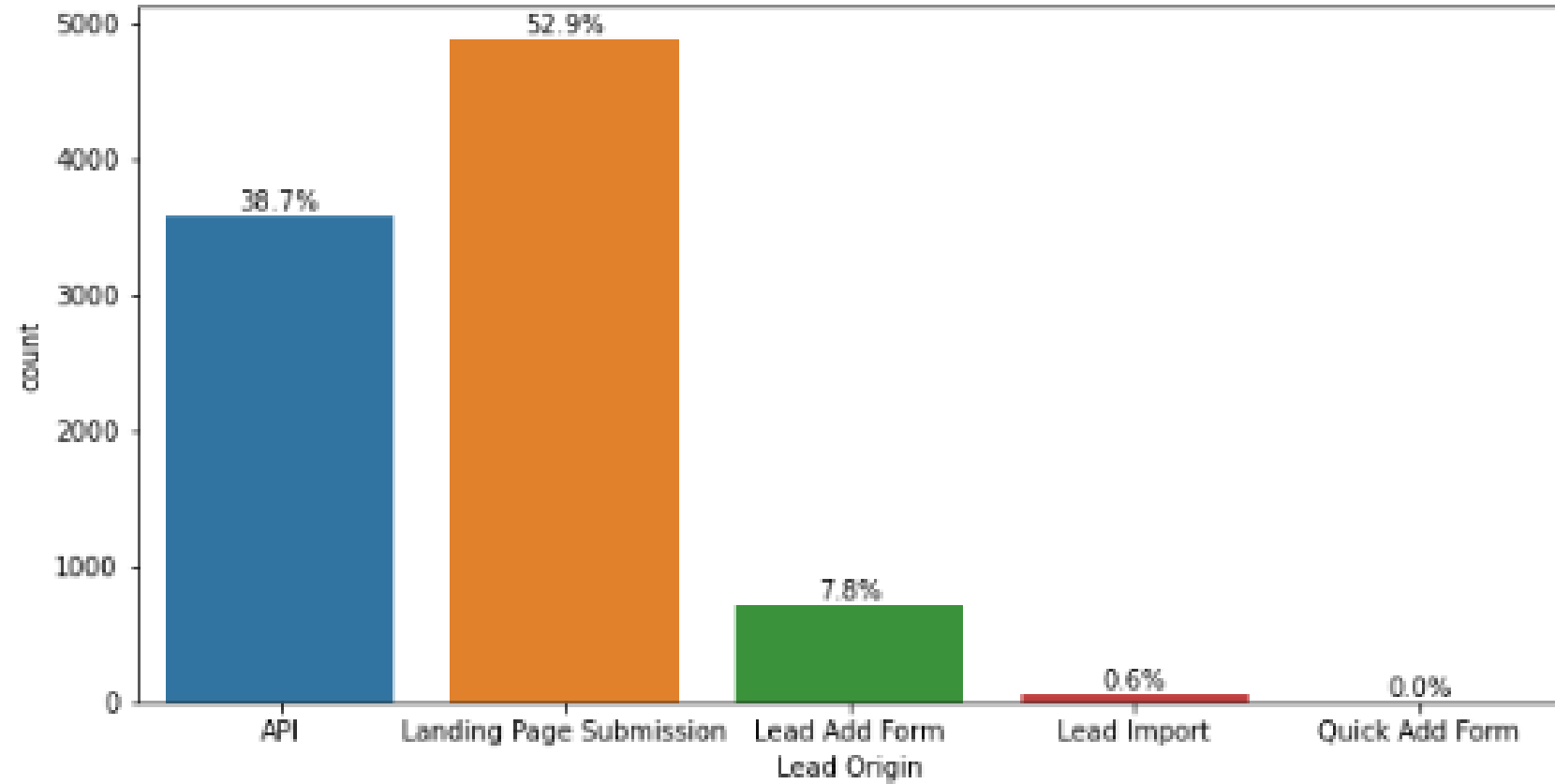


"TotalVisits", "Page Views Per Visit": Both these variables contain outliers as can be seen in the boxplot So, These outliers needs to be treated for these variables

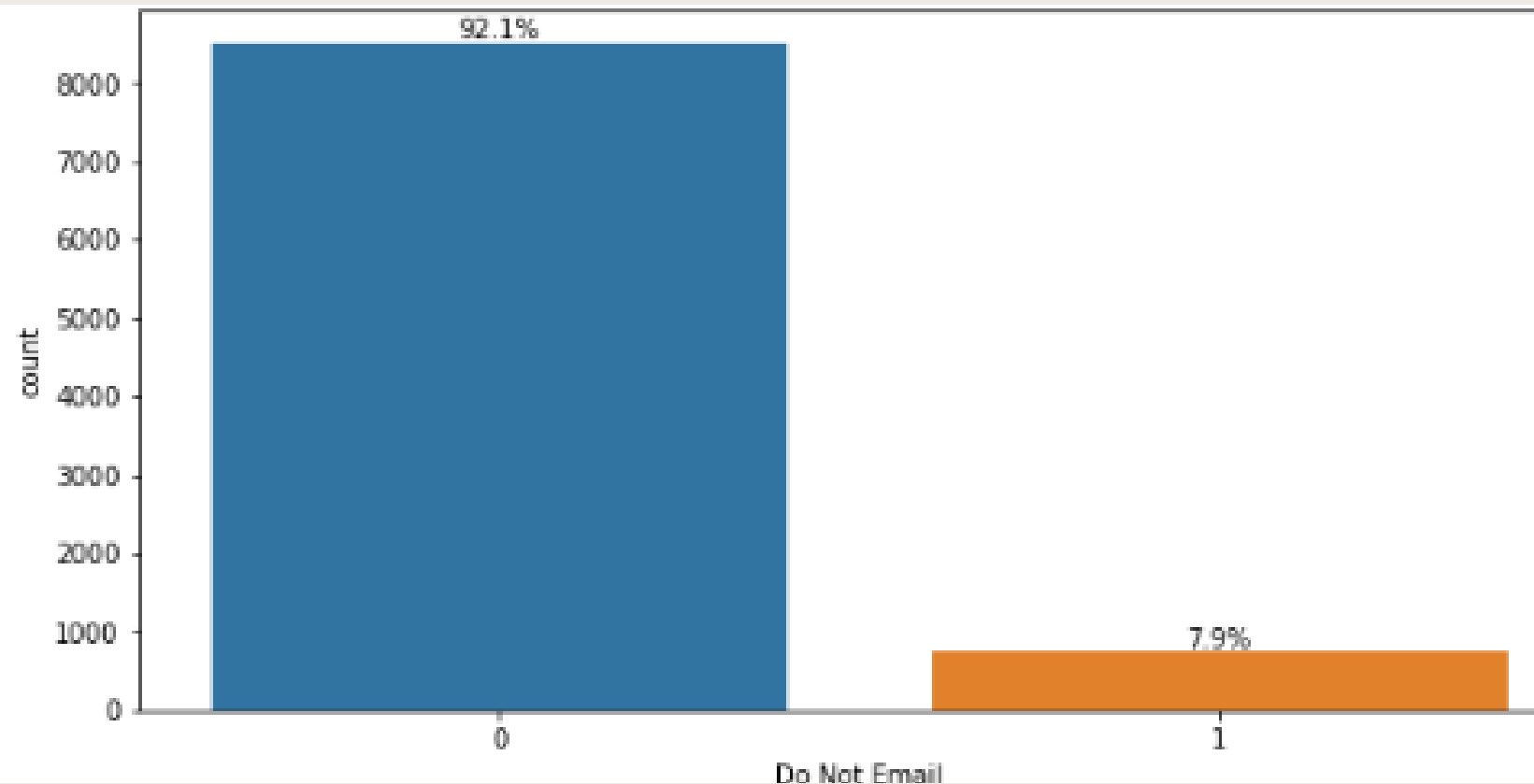
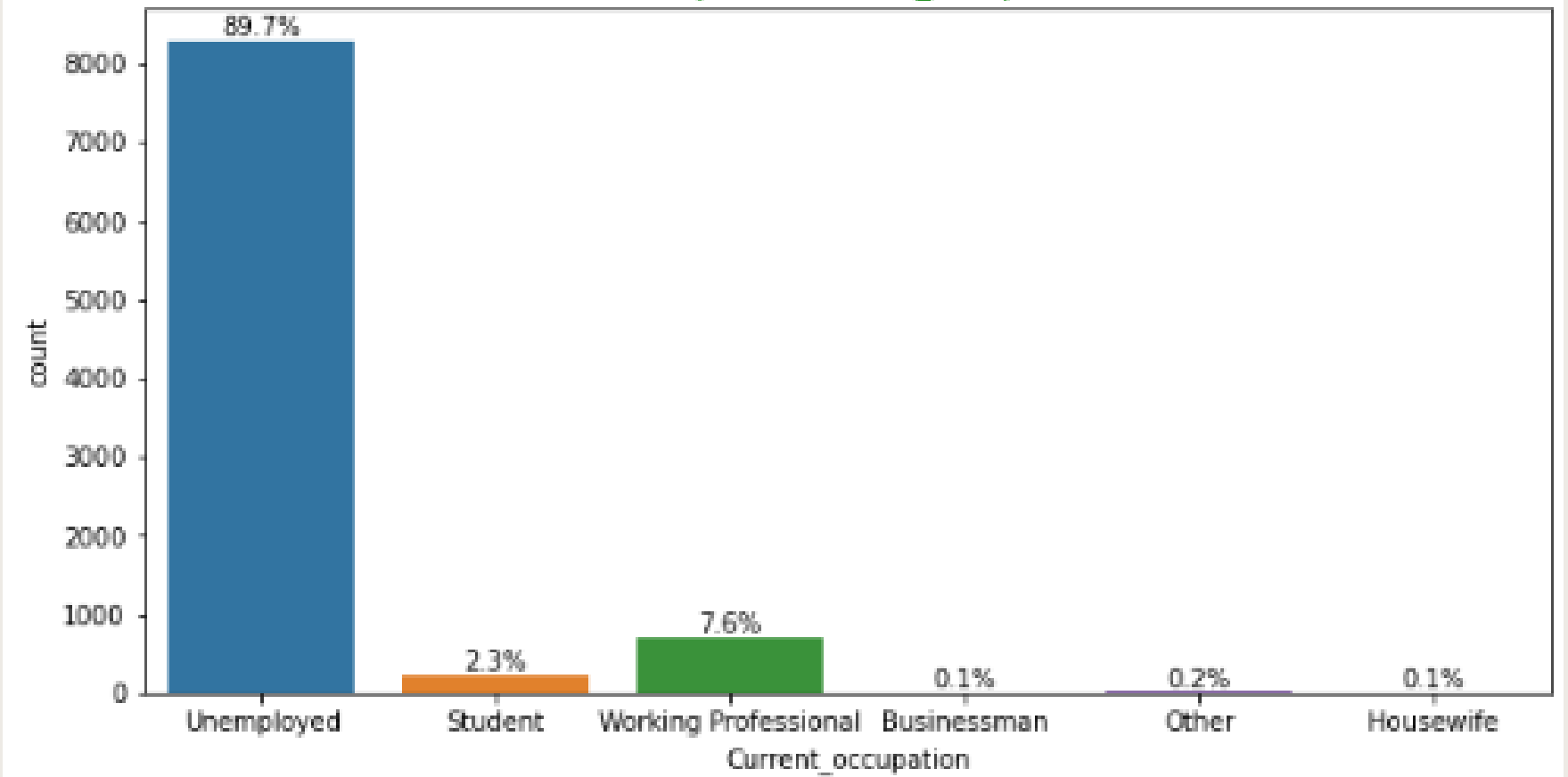


UNIVARIATE ANALYSIS

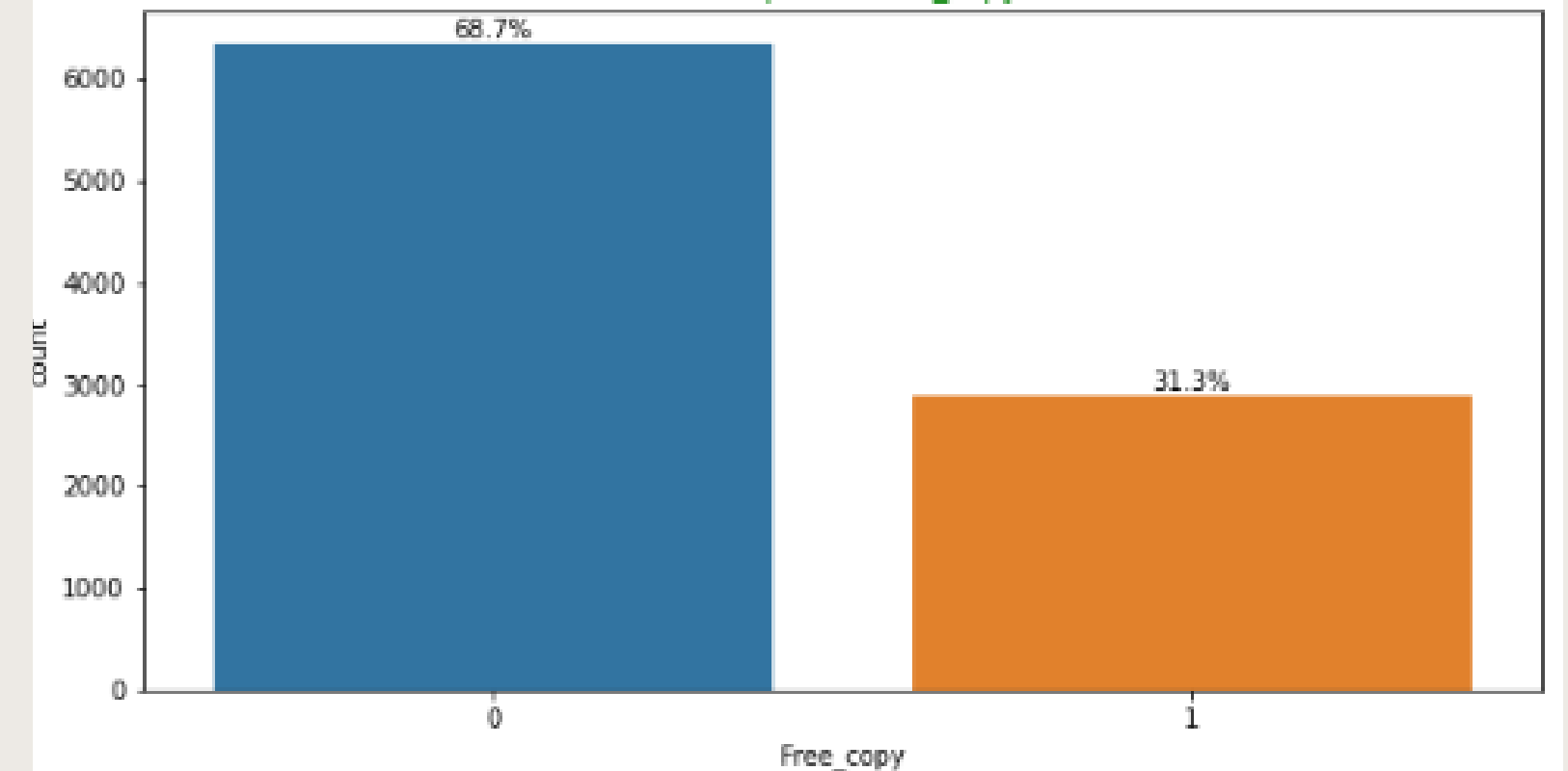
Count plot of Lead Origin



Count plot of Current_occupation



Count plot of Free_copy



Insights:

Lead Origin: Around 52% of all leads originated from "Landing Page Submission" with a lead conversion rate (LCR) of 36%. The "API" identified approximately 39% of customers with a lead conversion rate (LCR) of 31%.

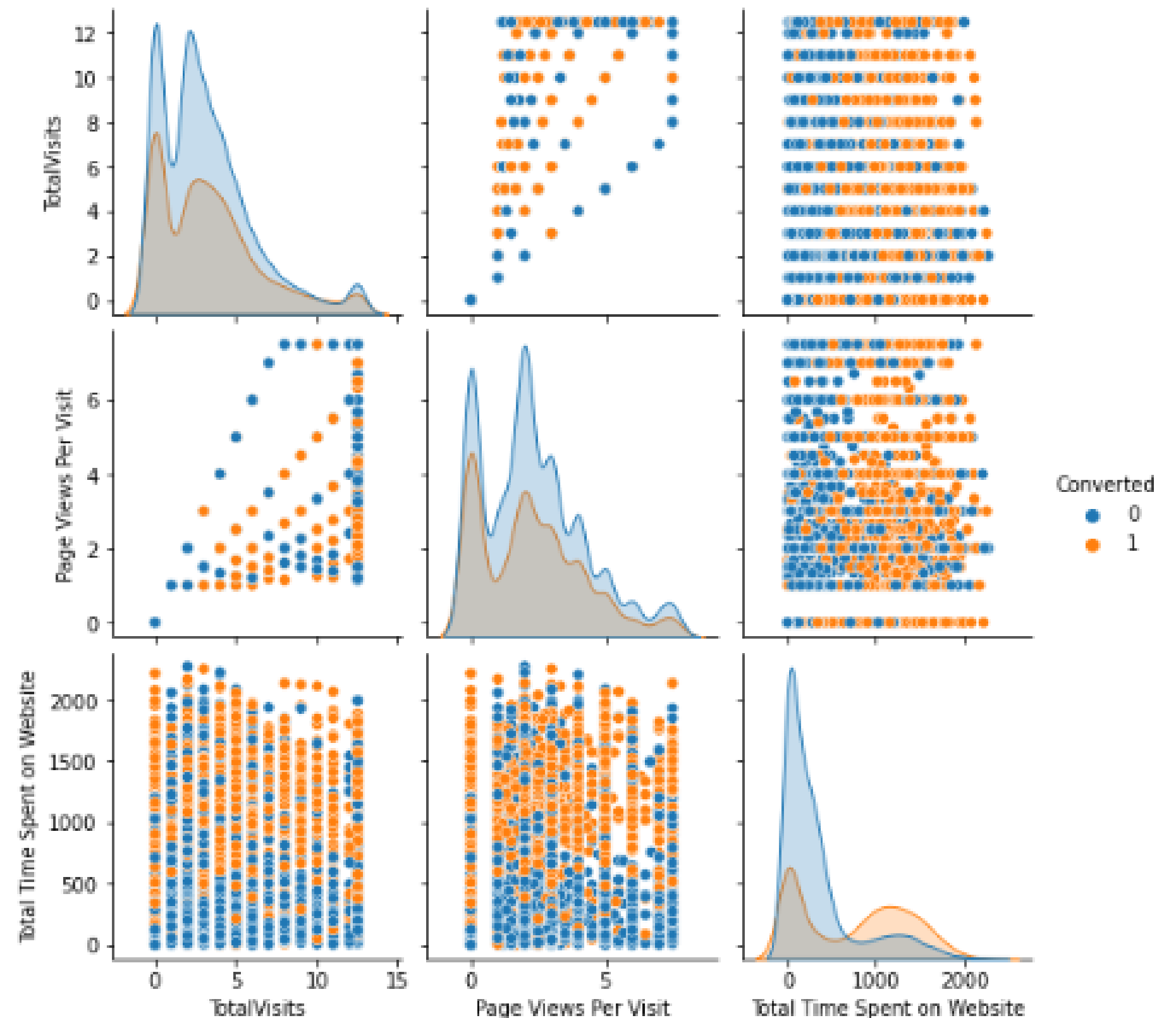
Current_occupation: Around 90% of the customers are Unemployed with lead conversion rate (LCR) of 34%. While Working Professional contribute only 7.6% of total customers with almost 92% lead conversion rate (LCR).

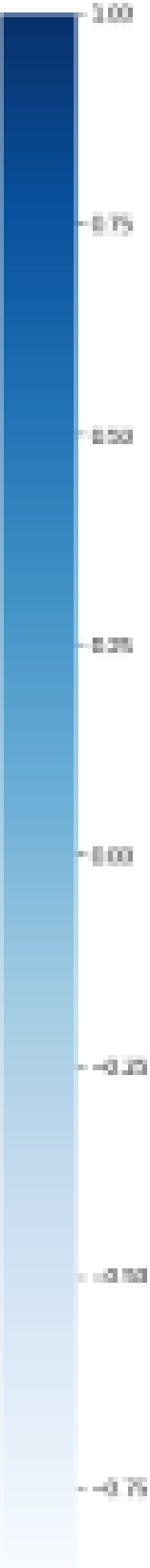
Do Not Email: 92% of the people has opted that they dont want to be emailed about the course.

Note: We have assumed LCR as Lead Conversion Rate in short form.

BIVARIATE ANALYSIS

<Figure size 1152x288 with 0 Axes>





CORRELATIONS



Recommendations

To increase our Lead Conversion Rates:

1. Targeted Marketing Strategies:

- **Focus on features with positive coefficients for effective marketing strategies.**

2. Attract High-Quality Leads:

- **Develop strategies to attract high-quality leads from top-performing sources.**

3. Tailored Messaging:

- **Engage working professionals with personalized messaging.**

4. Optimize Communication:

- **Enhance communication channels based on their impact on lead engagement.**

5. Website Advertising:

- **Allocate more budget for advertising on the Welingak website.**
- 



thank
you