

Lead Scoring Case Study

BY PHANEESH N & RUCHIR AGARWAL

OCT 2022

Index

- Problem Statement
- Assumptions
- EDA
- Overall Approach
- Result & Conclusion
- Recommendation

Problem Statement

Help X Education to select the most promising leads (the leads that are most likely to convert into paying customers) so that target lead conversion rate could be around 80%.

Utilizing past 9000 data points build a model and assign a lead score (between 0 to 100) to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

Assumptions

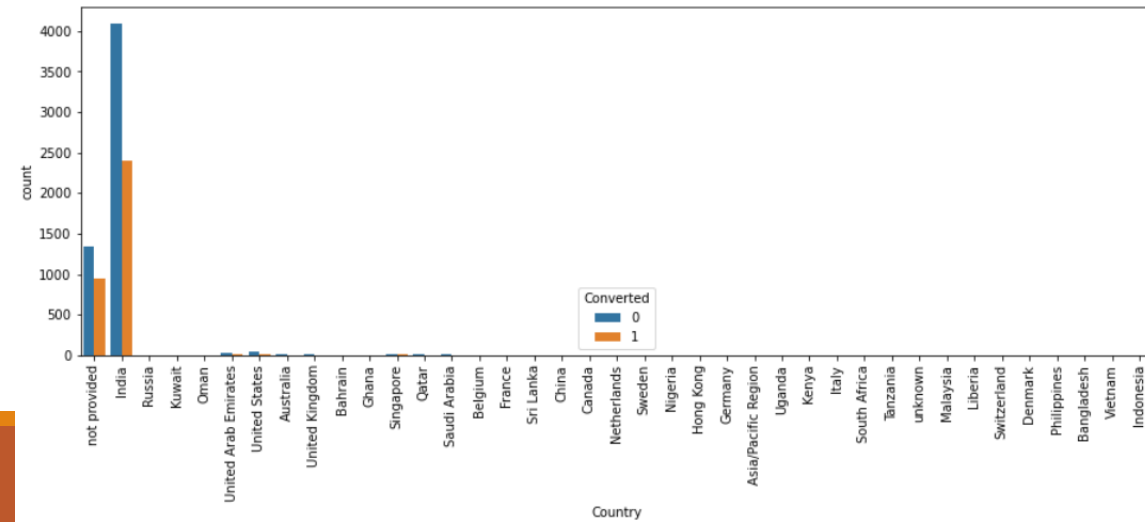
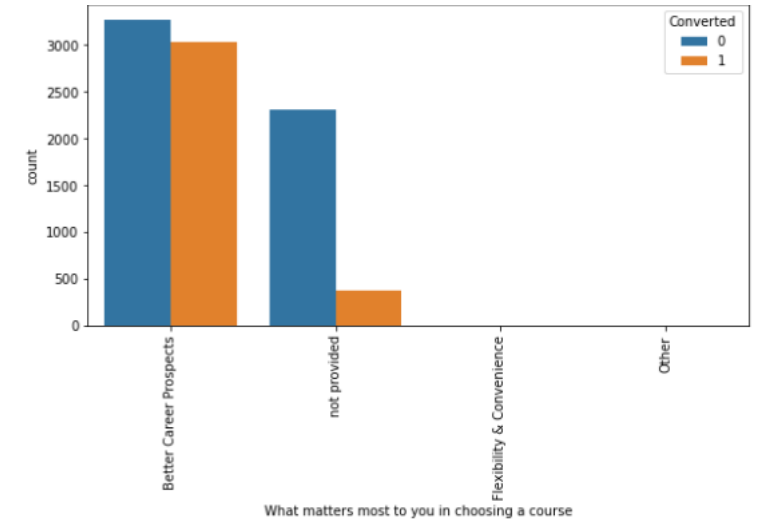
- Any parameters with more than 45% Null values are irrelevant for the model
- Values having data imbalance with single value more than 85% are removed. Example: City, Search, Newspaper Article, Newspaper, Digital Advertisement, etc.
- Bucketing the values which have less percentage of value_count as compared to other values in the parameter. Example: For “Last Activity” below conversion was carried out
 - Others → 'Unreachable', 'Unsubscribed', 'Had a Phone Conversation', 'Approached upfront', 'View in browser link Clicked', 'Email Marked Spam', 'Email Received', 'Resubscribed to emails', 'Visited Booth in Tradeshow

EDA – Data Cleansing

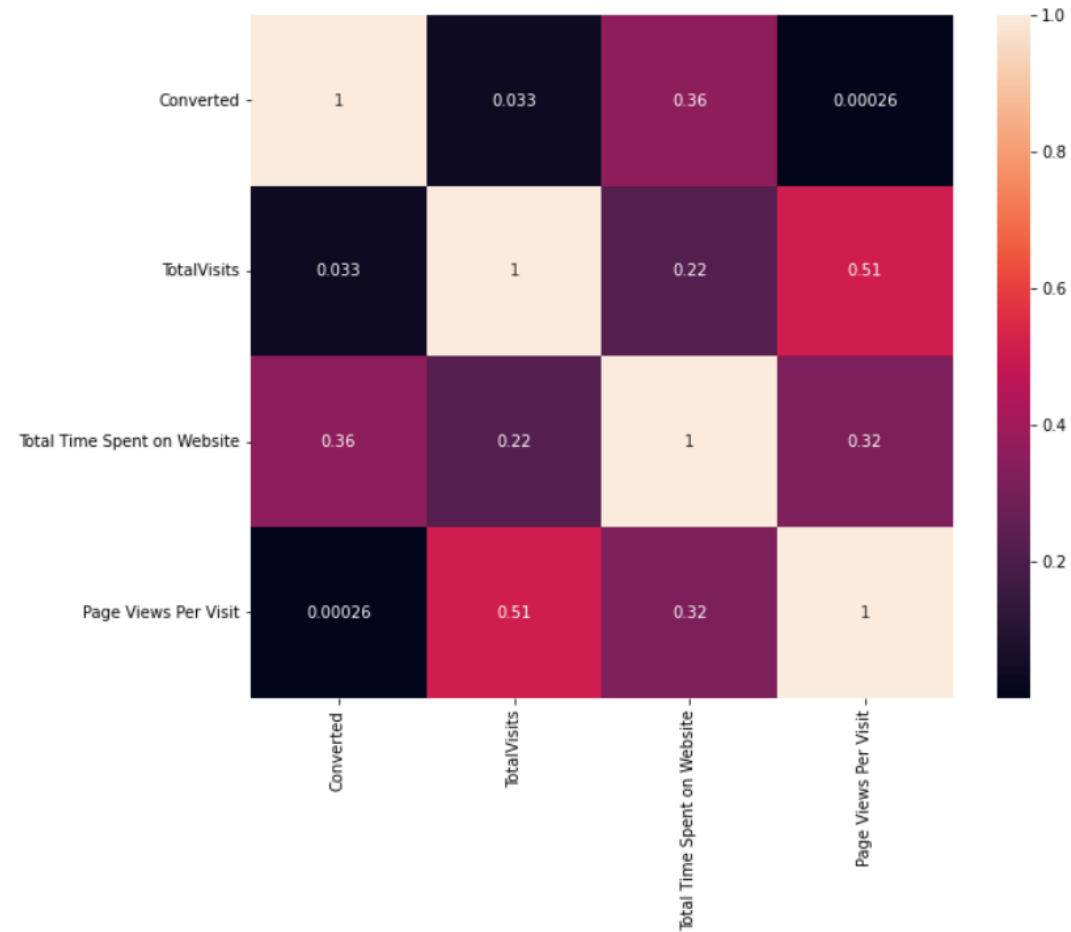
- Sourcing – Read datasets into Dataframes
 - Leads.csv
- Impute “Select” as “NaN”
- Remove Parameters which have $\geq 45\%$ null values
 - How did you hear about X Education
 - Asymmetrique Activity Index
 - Asymmetrique Profile Index
 - Asymmetrique Activity Score
 - Asymmetrique Profile Score
 - Lead Quality
 - Lead Profile
- Replacing NaN values for other Categorical parameters to mode value
- Bucketing certain Parameters into groups for ease of analysis

EDA – Data Cleansing

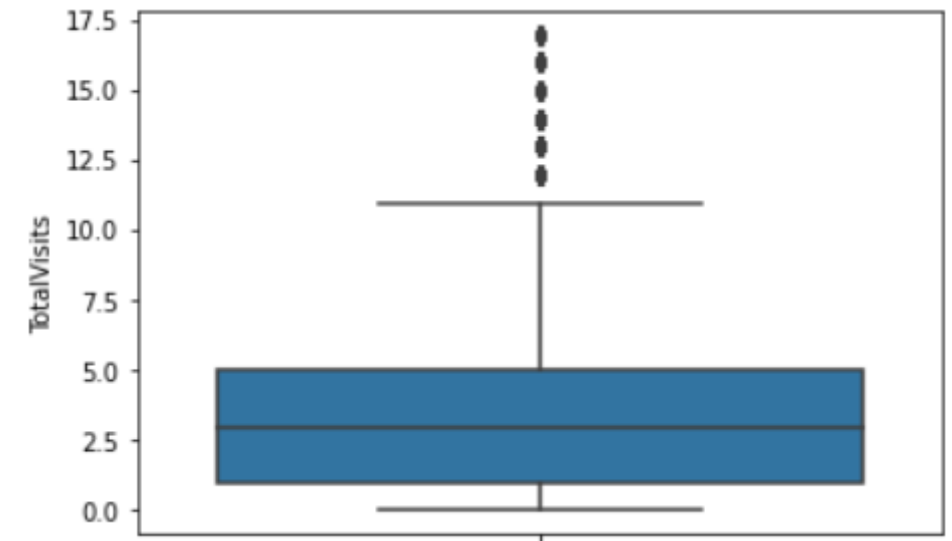
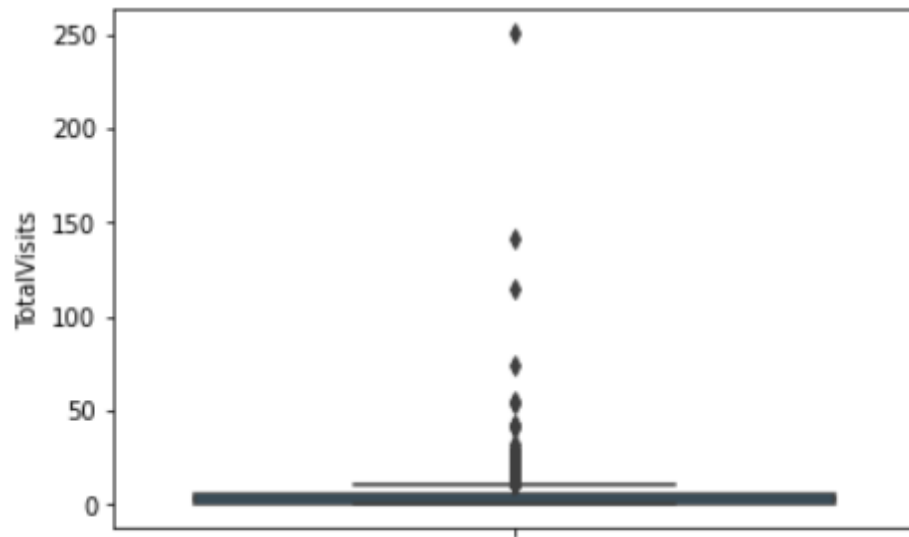
- Data Imbalance - Remove Parameters with Dominating Single values >85%
- Country
- What matters most to you in choosing a course
- Do Not Call
- Search
- Magazine
- Newspaper Article
- X Education Forums
- Newspaper
- Digital Advertisement
- Through Recommendations
- Receive More Updates About Our Courses
- Update me on Supply Chain Content
- Get updates on DM Content
- I agree to pay the amount through cheque



Correlation between Numeric Values



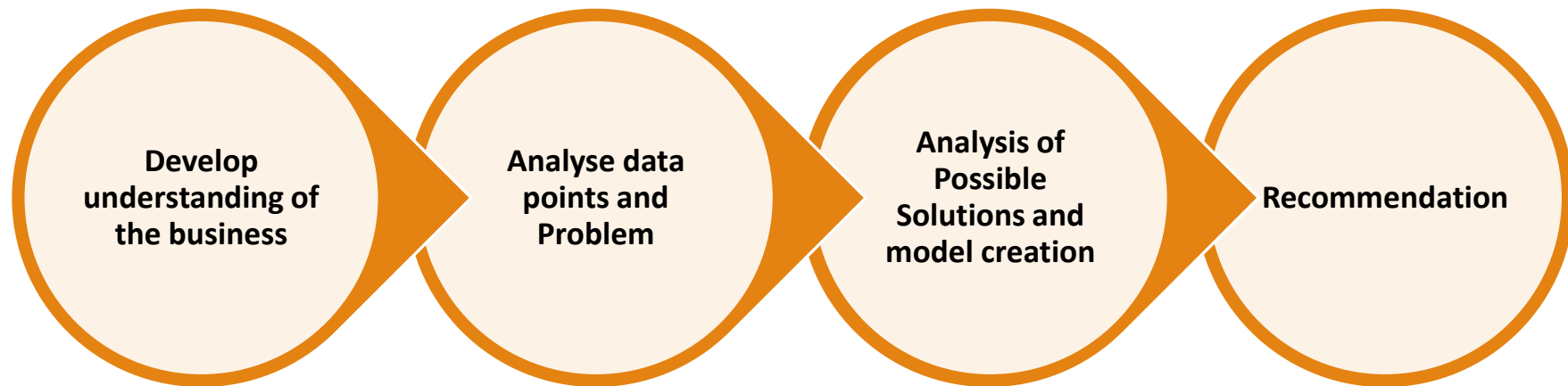
Outlier Treatment



EDA – Data Cleansing

- After Data Cleansing we have 8991 rows and 21 parameters remaining for building the model
- 97.31% of Data retained after Data Cleansing
- 37.80% of Data has “Converted” value as 1

Overall Approach

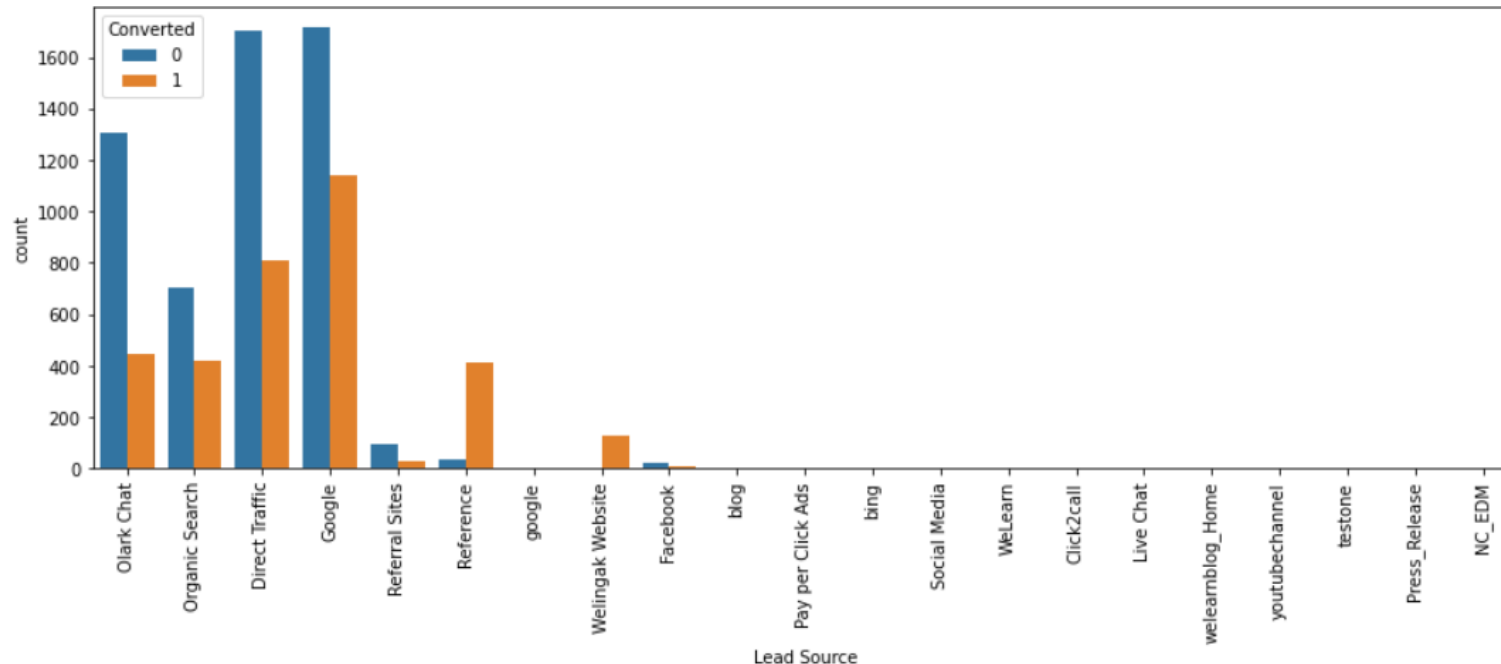


Business Understanding

- X Education sells online courses to industry professionals
- Company markets its courses on several websites and search engines like Google
- Lead is defined as any person filling up the form and providing their email address or phone number on the “X Education” website
- “X Education” executives then reach out to these leads for converting them into paying customers

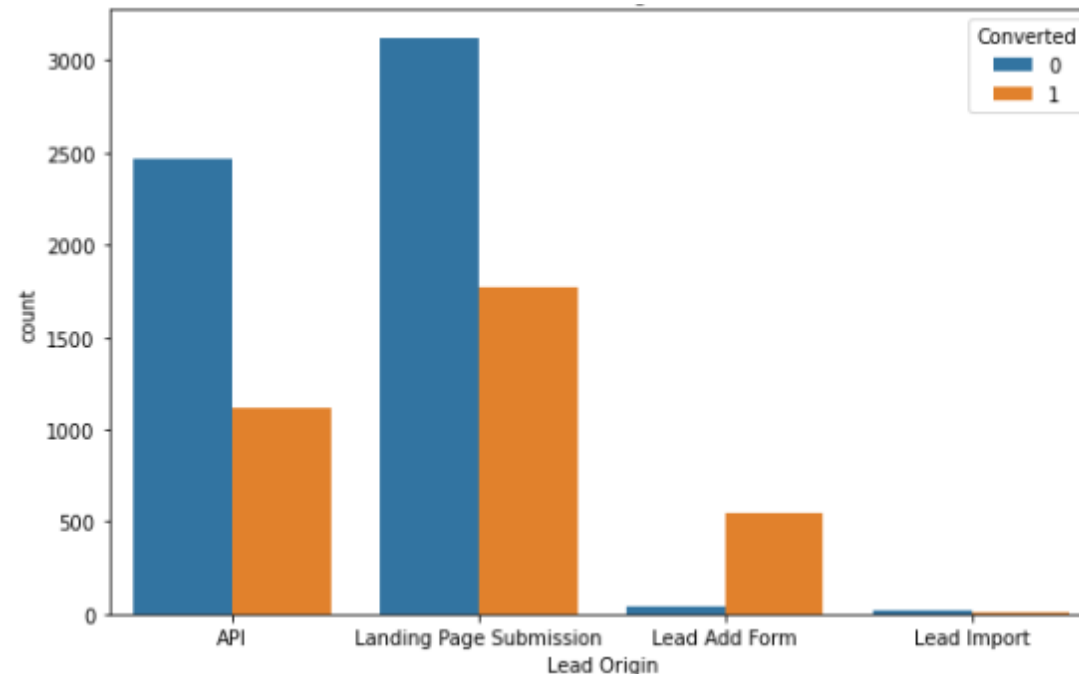
Data Understanding and Preparation

- Target Variable = Converted
- Maximum leads are generated by Google and Direct traffic To improve overall conversion rate, focus should be on improving lead conversion from olark chat, organic search, direct traffic, google leads and generate more leads from reference and welingak website



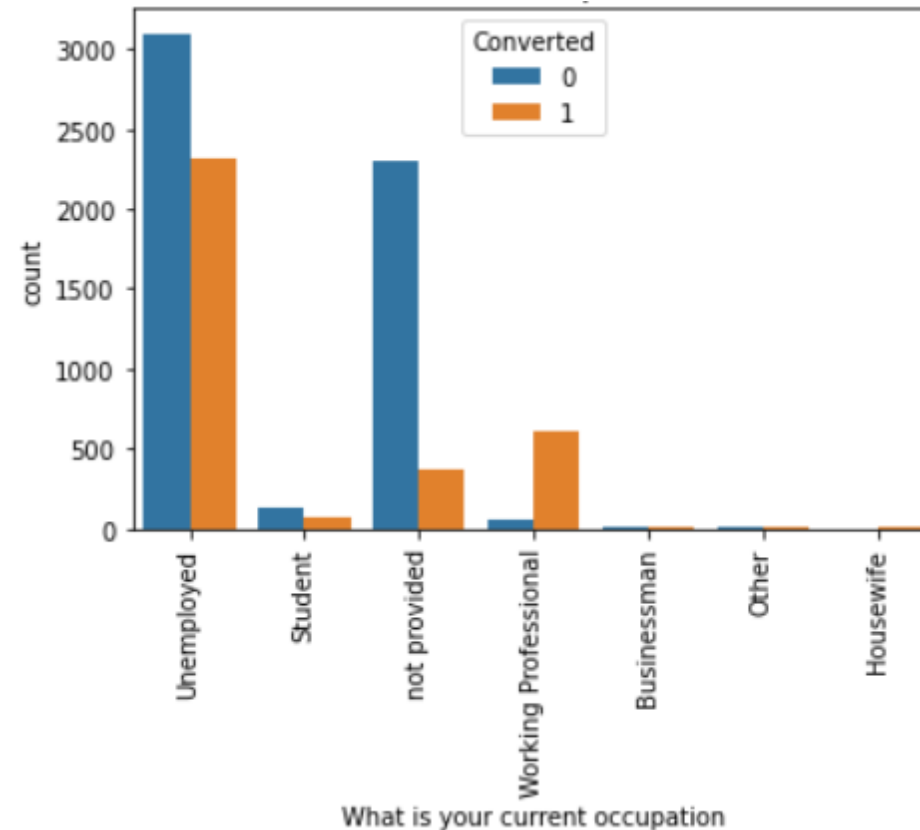
Data Understanding and Preparation

- API and Landing Page Submission bring higher number of leads as well as conversion. Lead Add Form has a very high conversion rate but count of leads are not very high. Lead Import and Quick Add Form get very few leads. In order to improve overall lead conversion rate, need to improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.



Data Understanding and Preparation

- Working Professionals are the most converted after unemployed



Data Understanding and Preparation

- Converting Categorical variables to binary values or Dummy variables:
 - Lead Origin
 - Lead Source
 - Do Not Email
 - Last Activity
 - Specialization
 - What is your current occupation
 - A free copy of Mastering The Interview
 - Last Notable Activity

Model Building

- Using RFE kept 15 parameters
- Optimal Cutoff is 0.35
- Built 3 models to achieve below stats on train data (with 13 parameters):
 - Precision : 72.09%
 - Recall : 80.02%
 - Accuracy: 80.80%
 - Specificity : 81.12%
- Stats on test data:
 - Precision : 72.82%
 - Recall : 80.31%
 - Accuracy : 81.20%
 - Specificity : 81.75%

Result & Conclusion

- Optimal cutoff is 0.35
- Comparing the train and test results:
 - Train Data:
 - Precision : 72.09%
 - Recall : 80.02%
 - Accuracy: 80.80%
 - Specificity : 81.12%
 - Test Data:
 - Precision : 72.82%
 - Recall : 80.31%
 - Accuracy : 81.20%
 - Specificity : 81.75%
- Below are the top 3 parameters (including dummy variables) that will impact the Lead conversion:
 - What is your current occupation_Working Professional
 - Lead Origin_Lead Add Form
 - Lead Source_Welingak Website

Recommendation

- Top 3 parameters to be considered are :
 - What is your current Occupation
 - Lead Origin
 - Lead Source
- X Education should focus on Working professionals, Leads from Add forms and Welingak Websites to have better conversions
- Based on the model, X Education should contact industry professionals whose "Lead Score" is equal to or greater than 80. They can be termed as 'Hot Lead'.

THANK YOU
