Lead Score Case Study

Submitted by:
Jyothi Suvarna
Bharani Kumar
Priyanka Odhar

Case Study for X Education

Problem Statement :

- X Education sells online courses to industry professionals.
- If people fill up a form providing their email address or phone number, then are classified to be a lead.
- Sales team use these information and reach out to these leads.
- Some of the leads get converted while some may not.
- The typical lead conversion rate at X education is around 30%.

Business Goal:

- X Education needs help in identifying the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- Build a model – which assigns a lead score 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.
- Ballpark of the target lead conversion rate need to be around 80%.

Steps Followed

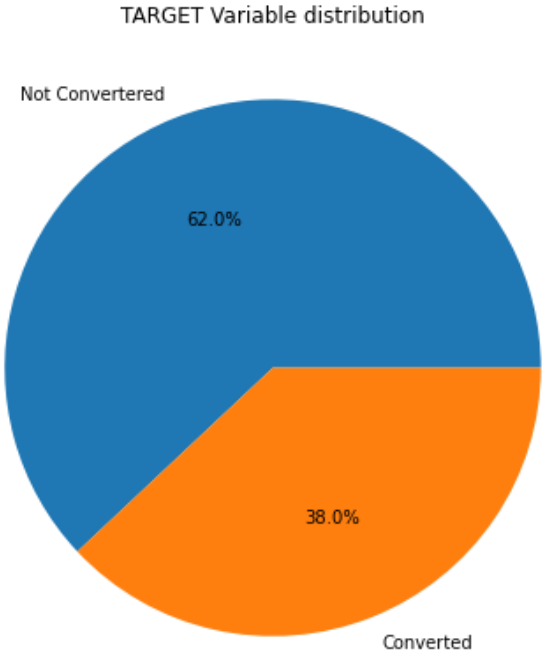
1. Data Cleaning and Exploratory Data Analysis
    a. Read the source data provided by X Education
    b. Missing value imputation
    c. Univariate and Bivariate analysis on categorical and numerical data
    d. Outlier Treatment

2. Data Preparation
    a. Create dummy variables for categorical variables
    b. Splitting data into train and test set
    c. Scaling of numeric data using standard scaler

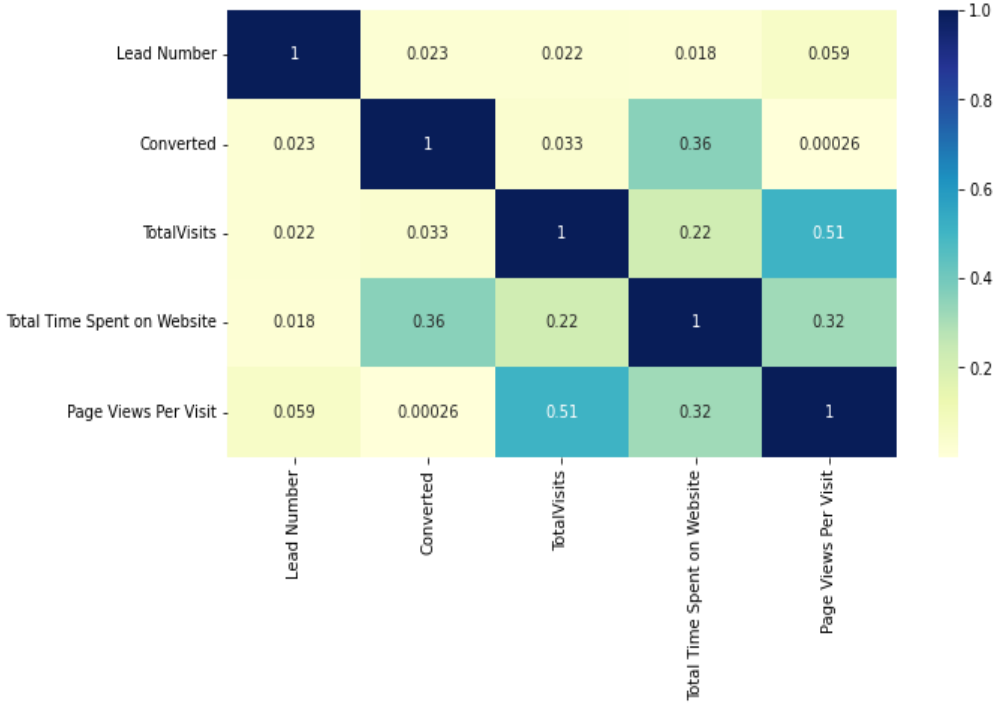
3. Modelling:
    1. Feature Selection using RFE
    2. Building Logistic Regression Model
    3. Model Evaluation using various metrics like sensitivity, specificity, precision, recall
    4. Finding optimal cut-off using ROC Curve
    5. Applying the best model to the Test data
    6. Evaluate the prediction on the test set using cut off threshold from ROC curve
    7. Calculating Lead Score using 85% conversion rate.

Exploratory Data Analysis

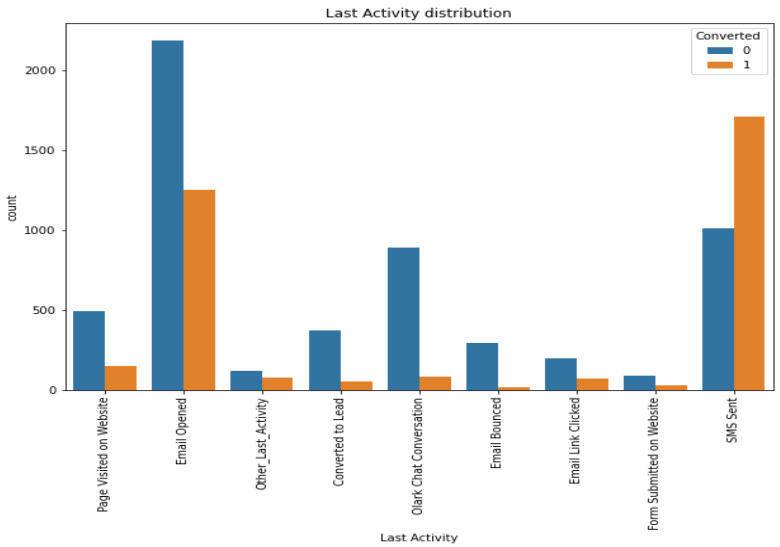
Company has 38% conversion rate



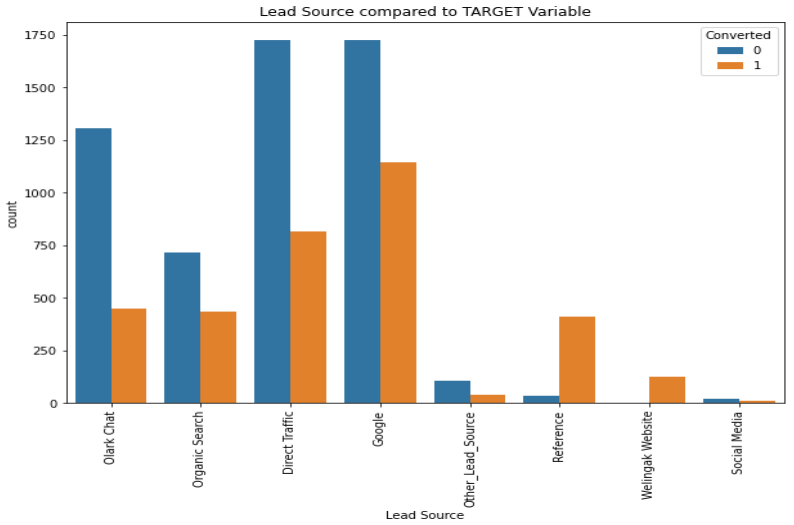
There is good correlation between conversion rate and Total visit to the web page



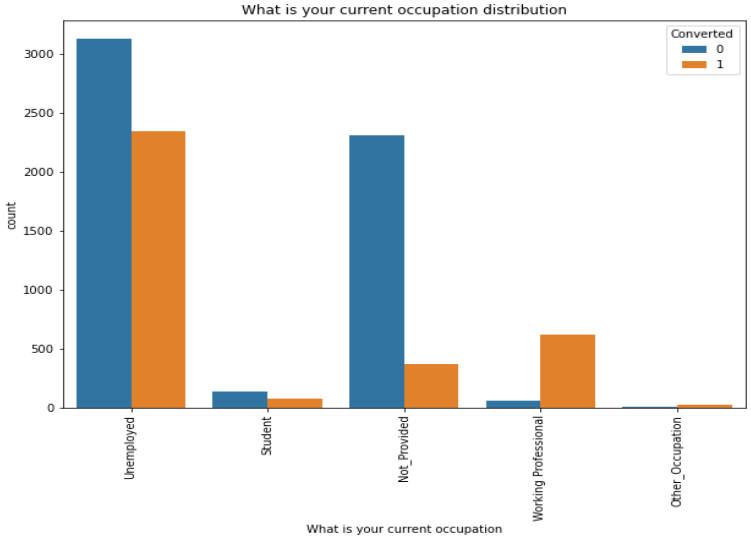
Customers with last activity of SMS sent has higher conversion rate



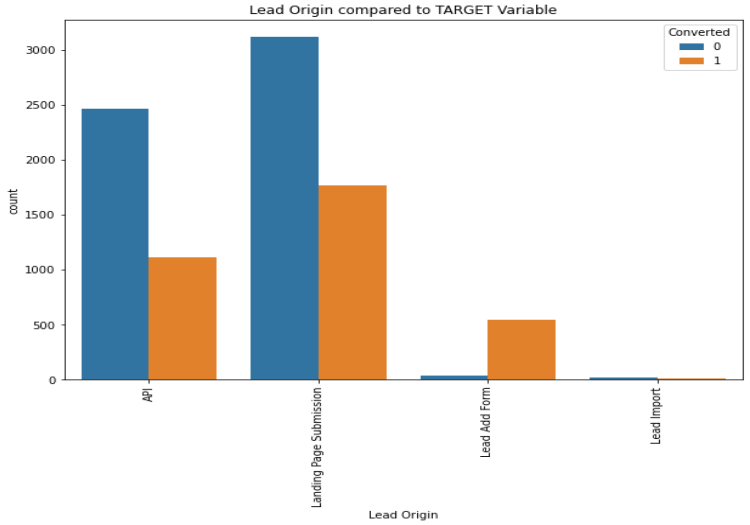
Search engine Google results in higher conversion rate



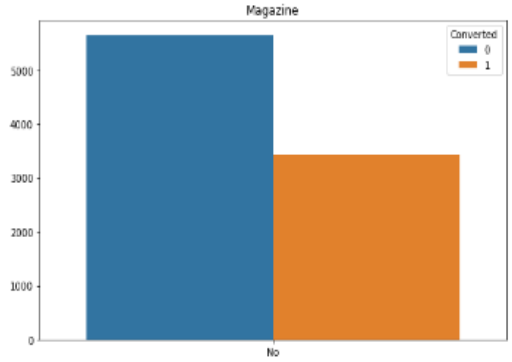
Unemployed people result in higher conversion rate

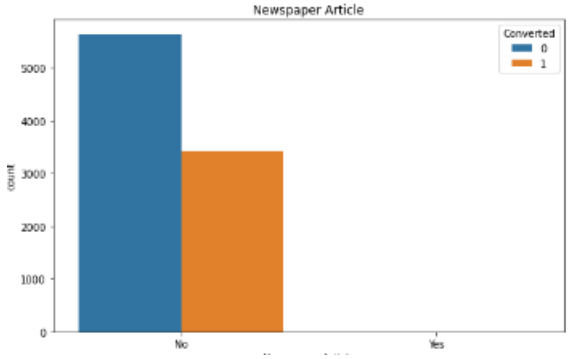


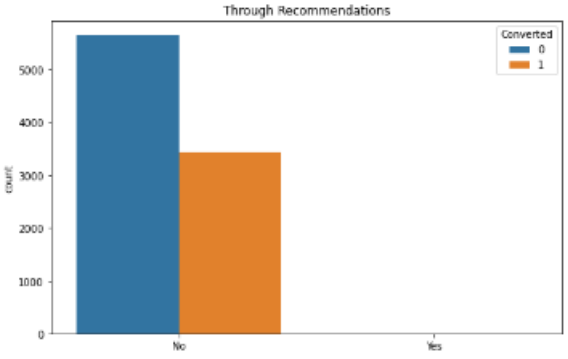
Landing page submission results in higher conversion rate

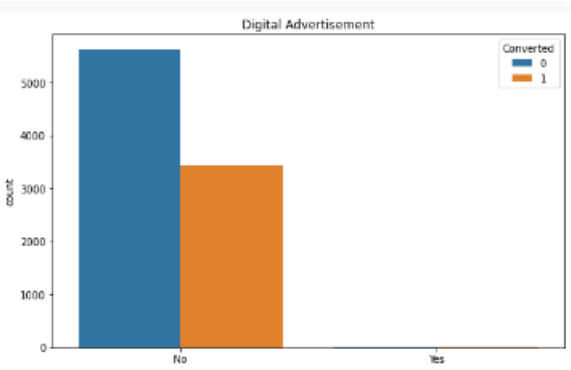


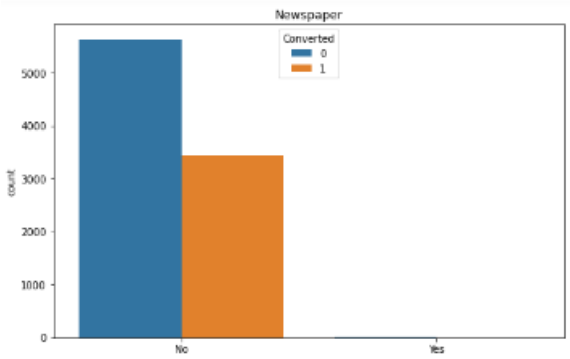
No impact on conversion rate from these variables

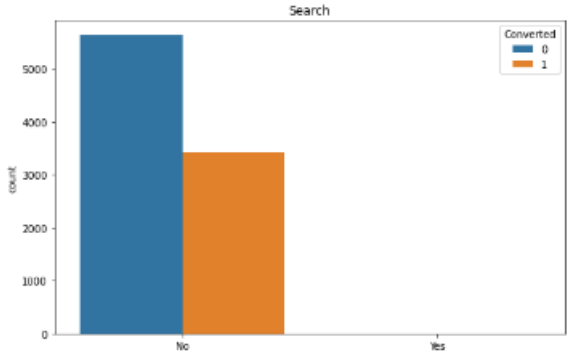












Final Model

Generalized Linear Model Regression Results

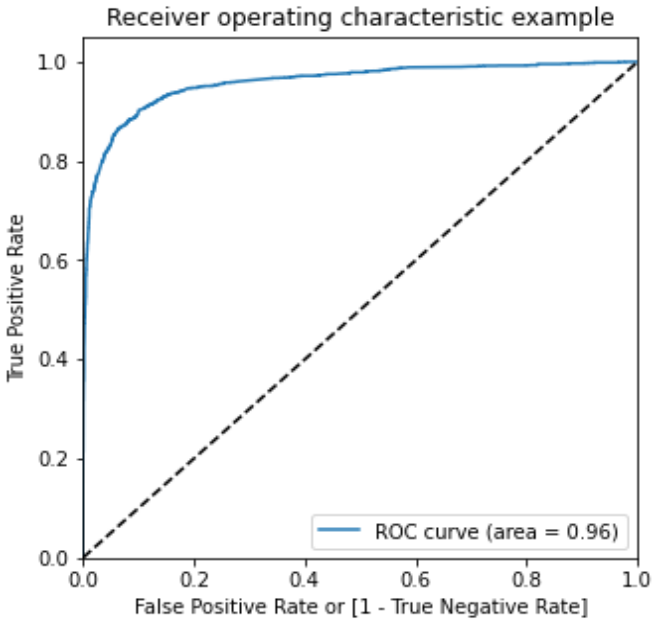
Dep. Variable:	Converted	**No. Observations:**	6057
Model:	GLM	**Df Residuals:**	6040
Model Family:	Binomial	**Df Model:**	16
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-1449.9
Date:	Mon, 14 Nov 2022	**Deviance:**	2899.9
Time:	14:18:32	**Pearson chi2:**	8.91e+03
No. Iterations:	8	**Pseudo R-squ. (CS):**	0.5705
Covariance Type:	nonrobust		

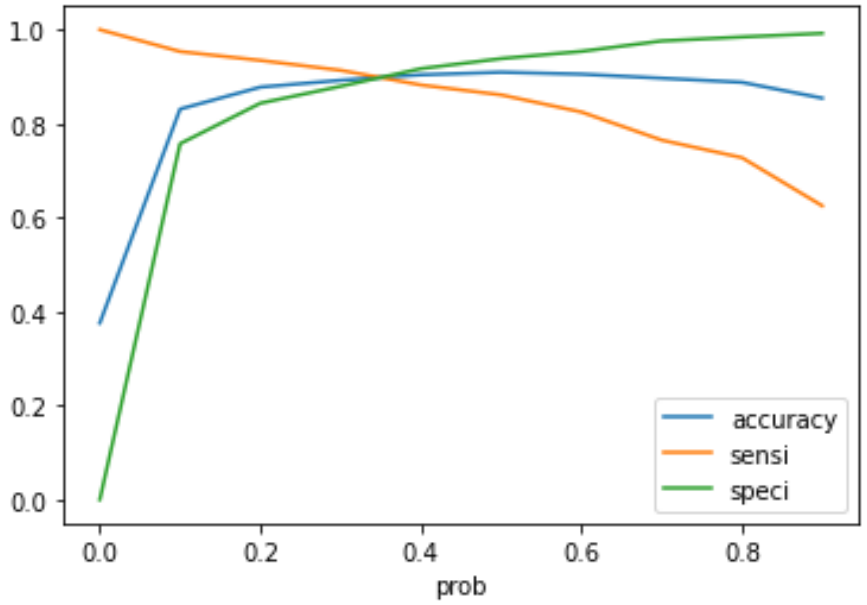
	coef	std err	z	P>|z|	[0.025	0.975]
const	-6.9447	0.279	-24.899	0.000	-7.491	-6.398
Do Not Email	-1.6806	0.210	-7.994	0.000	-2.093	-1.269
Total Time Spent on Website	1.0645	0.055	19.421	0.000	0.957	1.172
Lead Origin_Landing Page Submission	-0.7340	0.182	-4.032	0.000	-1.091	-0.377
Lead Source_Olark Chat	0.9214	0.165	5.591	0.000	0.598	1.244
Lead Source_Reference	2.5673	0.391	6.563	0.000	1.801	3.334
Lead Source_Welingak Website	3.8189	0.768	4.970	0.000	2.313	5.325
Last Activity_Olark Chat Conversation	-1.4127	0.208	-6.805	0.000	-1.820	-1.006
Specialization_Not Specified_Specialization	-0.9467	0.173	-5.471	0.000	-1.286	-0.608
What is your current occupation_Other_Occupation	2.6285	0.709	3.709	0.000	1.240	4.017
What is your current occupation_Student	3.3096	0.407	8.132	0.000	2.512	4.107
What is your current occupation_Unemployed	3.1769	0.123	25.845	0.000	2.936	3.418
What is your current occupation_Working Professional	4.4451	0.287	15.482	0.000	3.882	5.008
Tags_Closed by Horizzon	8.8481	1.028	8.603	0.000	6.832	10.864
Tags_Other_Tags	2.9692	0.196	15.164	0.000	2.585	3.353
Tags_Will revert after reading the email	5.5022	0.195	28.249	0.000	5.120	5.884
Last Notable Activity_SMS Sent	1.7835	0.115	15.490	0.000	1.558	2.009

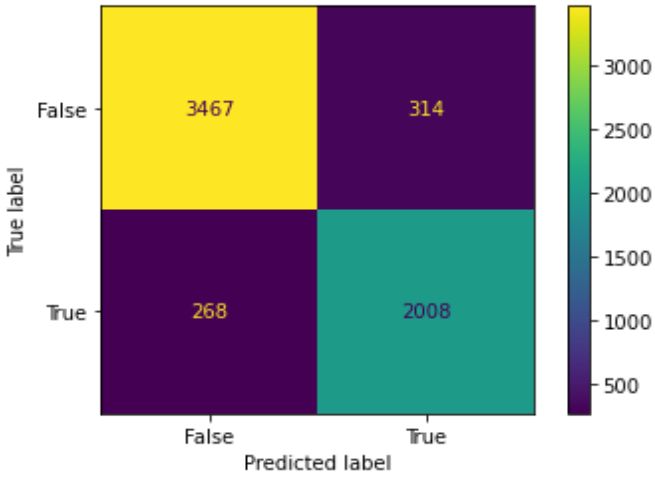
Model-10 is final model: p-values of all variables is 0 and VIF values are less than 5

	Features	VIF
2	Lead Origin_Landing Page Submission	3.72
14	Tags_Will revert after reading the email	3.41
10	What is your current occupation_Unemployed	3.12
7	Specialization_Not Specified_Specialization	3.02
3	Lead Source_Olark Chat	2.35
4	Lead Source_Reference	1.66
15	Last Notable Activity_SMS Sent	1.55
6	Last Activity_Olark Chat Conversation	1.47
13	Tags_Other_Tags	1.44
1	Total Time Spent on Website	1.43
11	What is your current occupation_Working Profes...	1.35
12	Tags_Closed by Horizzon	1.28
5	Lead Source_Welingak Website	1.15
0	Do Not Email	1.12
9	What is your current occupation_Student	1.09
8	What is your current occupation_Other_Occupation	1.01

Model Evaluation on Train data



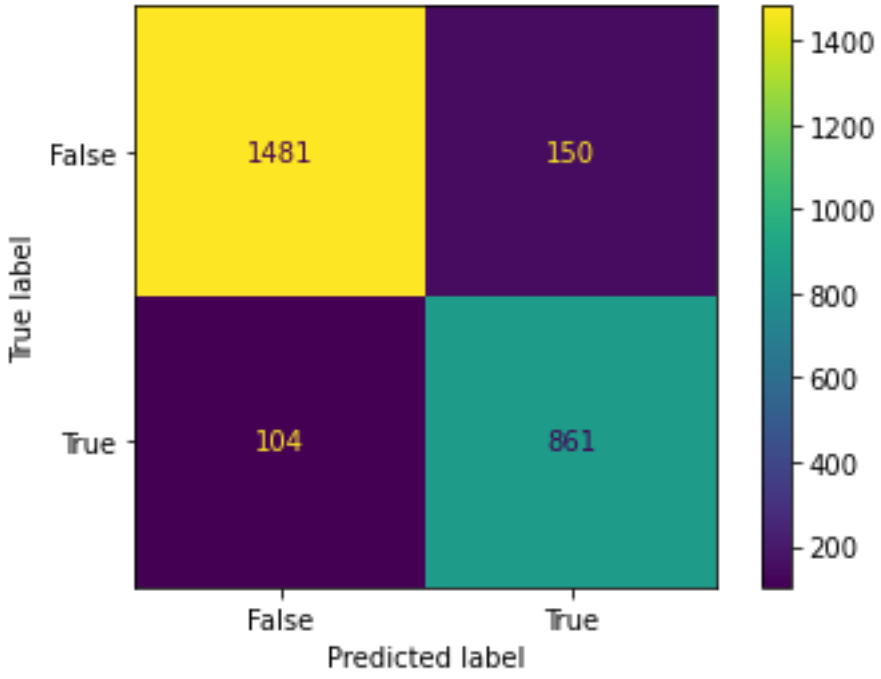




- **Findings:**
  - **Optimal cut-off is ~0.4 as per above**
- **Metrics:**
  - **Accuracy : 90.39%**
  - **Sensitivity : 88.22%**
  - **Specificity : 91.7%**
  - **Recall : 86.12%**
  - **Precision : 89.38%**

Model Evaluation on Test Data

- **Using the cut-off f 0.4 below are the Model Metrics on Test Data:**
  - **Accuracy : 90.39%**
  - **Sensitivity : 88.22%**
  - **Specificity : 91.7%**
  - **Recall : 86.12%**
  - **Precision : 89.38%**



Recommendation

- Company should make calls to the leads with below category;
  - Tags with Closed by Horizon
  - Tags with Will revert after reading the email
  - Leads who are Working Professional, Student and Unemployed
  - Lead Source from Welingak Website

- Company should not make calls to the leads with below category;
  - Lead Origin with Landing Page Submission
  - Specialization Not Specified
  - Lead opted for Do Not Email

HOT Leads

- Treated leas with 85% above lead score as Hot leads
- There are 708 leads which can be contacted and have a high chance of conversion
- Below are 10 records of hot leads for example;

	Prospect ID	Converted	Converted_prob	final_predicted	Lead_Score
0	2629	1	0.967268	1	97
4	2198	1	0.969447	1	97
5	6255	1	0.929363	1	93
6	1511	1	0.959480	1	96
7	4168	1	0.982468	1	98
10	804	1	0.920392	1	92
12	3437	1	0.920763	1	92
15	2868	1	0.962547	1	96
28	4151	1	0.986205	1	99
29	7220	1	0.963268	1	96