Introduction:

I have a dataset of Easyvisa. Easy visa is a travel agency. My task is to Analyze and fetch information from the data to Facilitate the process of visa approvals and Recommend a suitable profile for the applicants for whom the visa should be certified or denied based on the drivers that significantly influence the case status. Data Science is helping people in various socioeconomic and health sectors. Therefore, we realize the need for data and data scientists to help the world become a better place. We need to learn Data Science to create better solutions for real-world problems that people face today. So here we are going to find out our answers with the help of data science.

What I am going to do?

I will do data analysis into two segments. I will slice the data by the case_status column which is the target column, I will separate the data frame into the "Certified" and "Denied" datasets. Then I will visualize the features one by one and give the conclusions on plotted Graphs. Then I will do Data preprocessing and clean the dataset, Then based on the dataset I will build the model. So I will do this project in two parts first I will visualize the features and then I will train the model.

Features

1) case_id: ID of each visa application.
2) continent: Information of continent the employee.
3) education_of_employee: Information of education of the employee.
4) has_job_experience: Does the employee has any job experience? Y= Yes; N = No.
5) requires_job_training: Does the employee require any job training? Y = Yes; N = No.
6) no_of_employees: Number of employees in the employer's company.
7) yr_of_estab: Year in which the employer's company was established.
8) region_of_employment: Information of foreign worker's intended region of employment in the US.
9) prevailing_wage: Average wage paid to similarly employed workers in a specific occupation in the area of intended employment. The purpose of the prevailing wage is to ensure that the foreign worker is not underpaid compared to other workers offering the same or similar service in the same area of employment.
10) unit_of_wage: Unit of prevailing wage. Values include Hourly,

Weekly, Monthly, and Yearly.
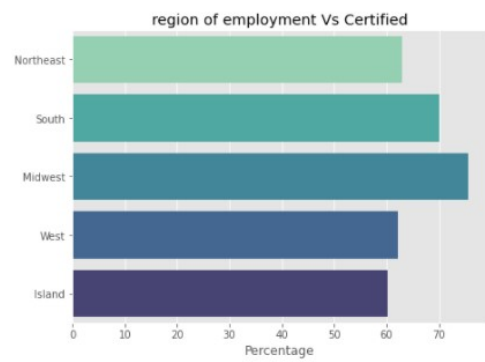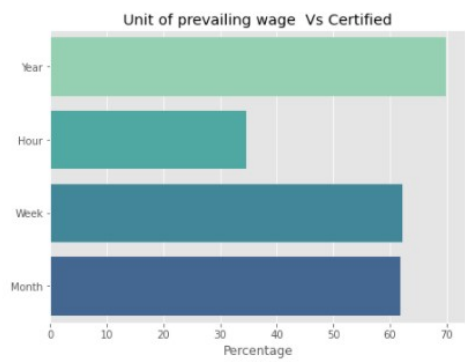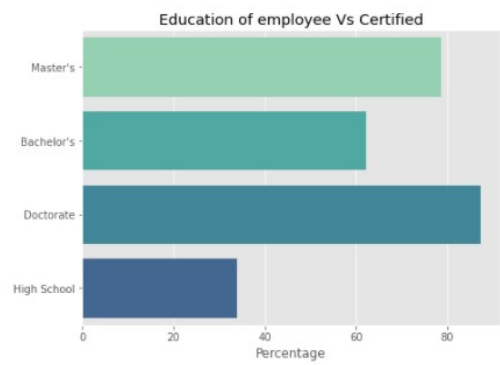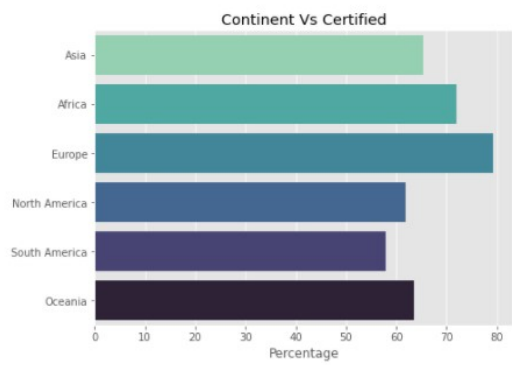11) full_time_position: Is the position of work full-time? Y = Full-Time Position; N = Part-Time Position.
12) case_status: Flag indicating if the Visa was certified or denied (Target column).

Dataset:

continent	education_of_empl	has_job_experience	requires_job_training	no_of_employees	yr_of_estab	region_of_employm	prevailing_wage	unit_of_wage	full_time_position	case_status
Asia	High School	N	N	14513	2007	West	592.2029	Hour	Y	Denied
Asia	Master's	Y	N	2412	2002	Northeast	83425.65	Year	Y	Certified
Asia	Bachelor's	N	Y	44444	2008	West	122996.86	Year	Y	Denied
Asia	Bachelor's	N	N	98	1897	West	83434.03	Year	Y	Denied
Africa	Master's	Y	N	1082	2005	South	149907.39	Year	Y	Certified
Asia	Master's	Y	N	2339	2012	South	78252.14	Year	Y	Certified
Asia	Bachelor's	N	N	4985	1994	South	53635.39	Year	Y	Certified
North America	Bachelor's	Y	N	3035	1924	West	418.2298	Hour	Y	Denied
Asia	Bachelor's	N	N	4810	2012	Midwest	74362.19	Year	Y	Certified
Europe	Doctorate	Y	N	2251	1995	South	67514.76	Year	Y	Certified
Asia	Master's	N	N	2465	2004	Midwest	83588.56	Year	Y	Certified
Asia	High School	Y	N	4069	2005	Northeast	70813.09	Year	Y	Denied
Asia	Bachelor's	Y	N	123876	1963	Northeast	28663.05	Year	Y	Certified
Asia	Bachelor's	Y	N	1652	2007	Midwest	107196.03	Year	Y	Certified
Asia	Master's	Y	Y	15756	2006	South	220081.73	Year	Y	Certified
Asia	High School	Y	N	4897	1987	West	74108.02	Year	Y	Denied
Europe	Master's	Y	N	76638	1991	Midwest	3706.79	Year	Y	Certified

Plots of Certified percentage rate of features

Certified Ratio with different features

Continent Vs Certified

Education of employee Vs Certified

Unit of prevailing wage Vs Certified

region of employment Vs Certified

Visa Success Rate by continent :

- Asia success rate is 65.31%.
- Europe success rate is 79.23%.
- Africa success rate is 72.05%.
- Oceniasuccess rate is 63.54.
- South America success rate is 57.86%.
- North America success rate is 61.88%.

Visa Success Rate by Education level:

- Master's level success rate is 78.63%.
- Bachelor's level success rate is 62.21%.
- Doctorate level rate is 87.23%.
- High school level success rate is 34.04%.

Visa Success Rate by Unit of prevailing wage :

- Yearly success rate is 69.88%.
- Monthly success rate is 61.79%.
- Weekly success rate is 62.13%.
- Hourly success rate is 34.63%.

Visa Success Rate by region of employment:

- Nourtheast success rate is 62.90%.
- South success rate is 70.02%.
- midwest success rate is 75.53%.
- Island success rate is 62.25%.

Note :

- Eupropeans have the high ratio of getting visa.
- If you have Doctorate , masters and bachelors then the success ratio of getting visa will increase.
- The Yearly prevailing wage has high success ratio.
- If your company region is in midwest then you have high success ratio.
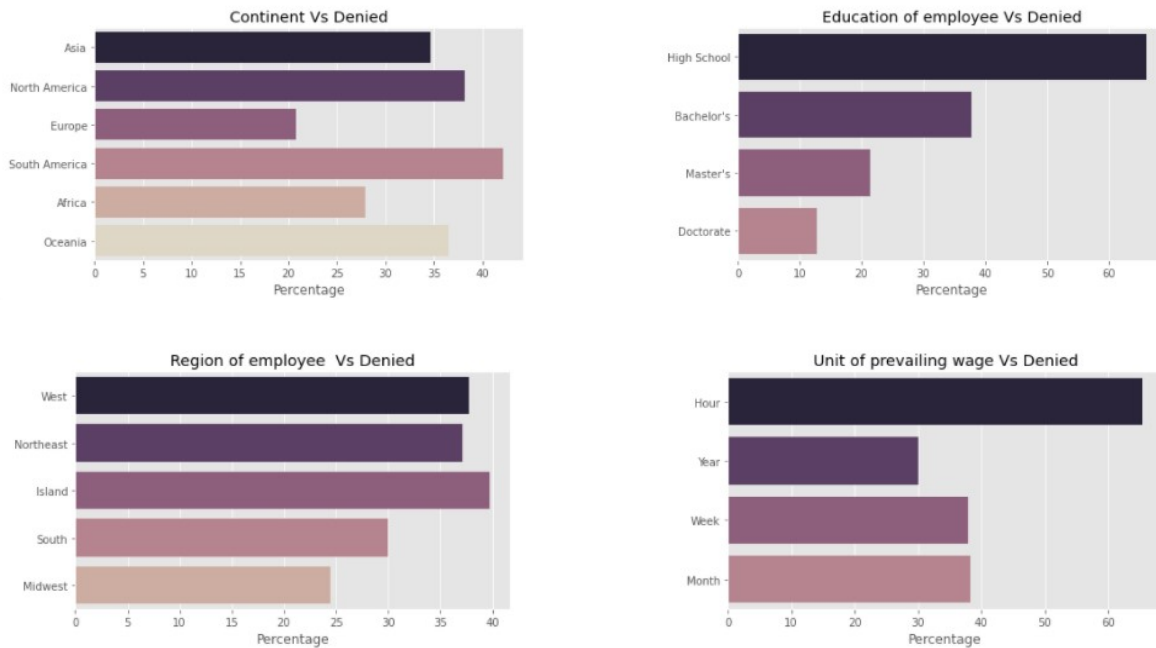
has_job_experience	Percentage of Certified	Require_job_training	Percentage of Certified	Full time positive	Percentage of Certified
0	Y	74.476422	0	N	66.645949	0	Y	66.583235
1	N	56.134108	1	Y	67.884941	1	N	68.526044

Note :

If you have good job experience then the chance of getting visa will increase

Plots of Visa rejection Rate
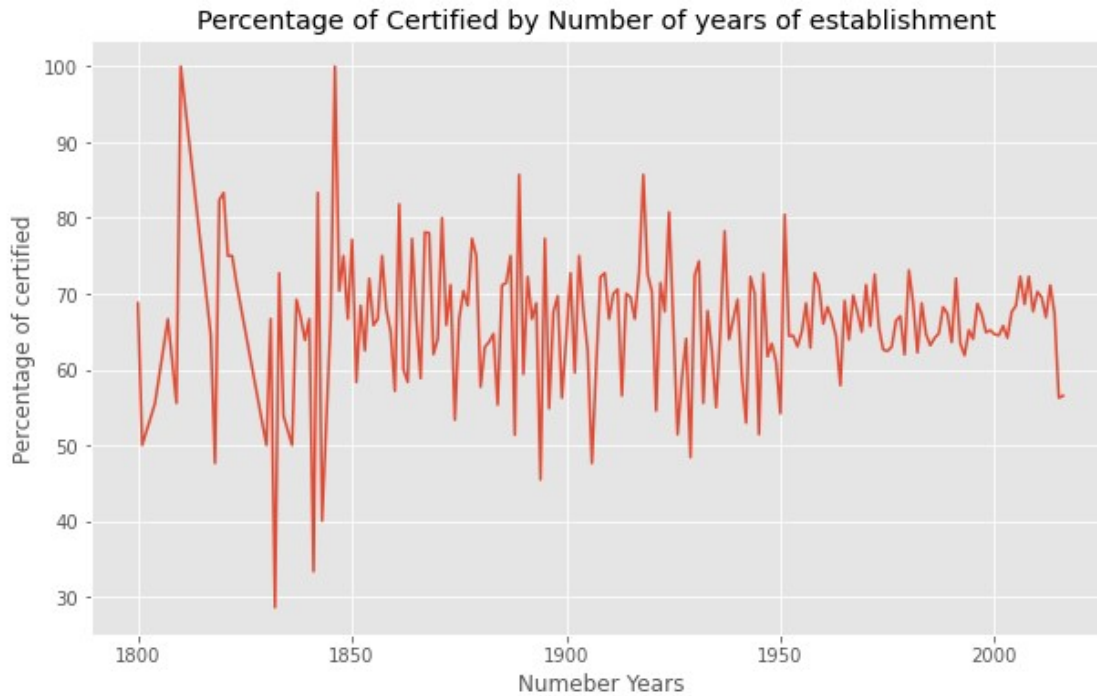
Denied Ratio with different features

Note:

- South Americans has high ratio of rejection.
- If you are a high school student then the rejection ratio is very high.
- If your company region is on island then the rejection ratio is very high.
- The hourly prevailing wage has high rejection ratio.

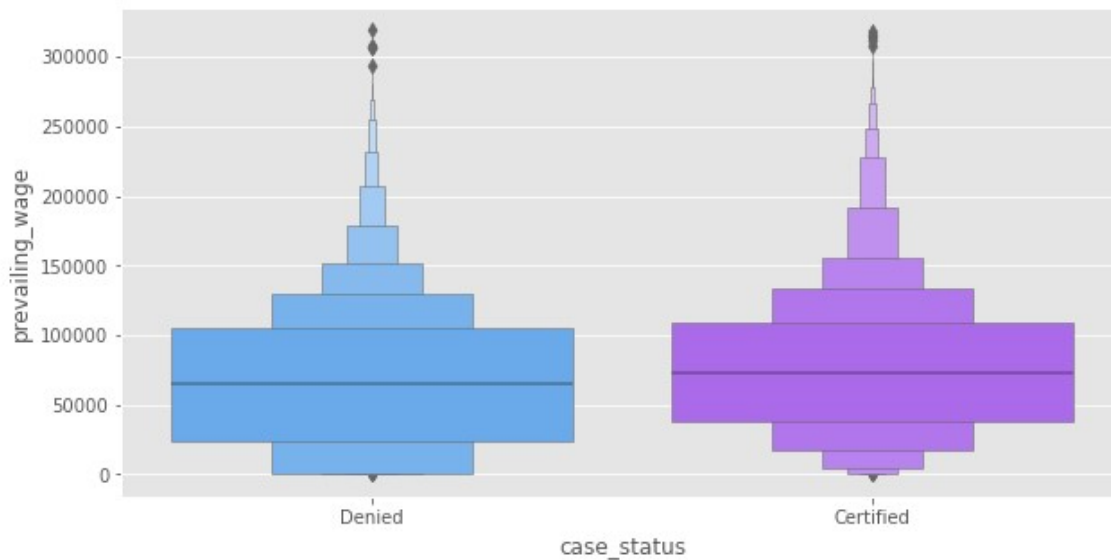|	has_job_experience	Percentage of Denied	Require_job_training	Percentage of Denied	Full time positive	Percentage of Denied
0	N	43.865892	0 N	33.354051	0 Y	33.416765
1	Y	25.523578	1 Y	32.115059	1 N	31.473956

Note :

If you no job experience then your success ratio will decrease and other features like require job training and full time postive does not have any major impact on case status.
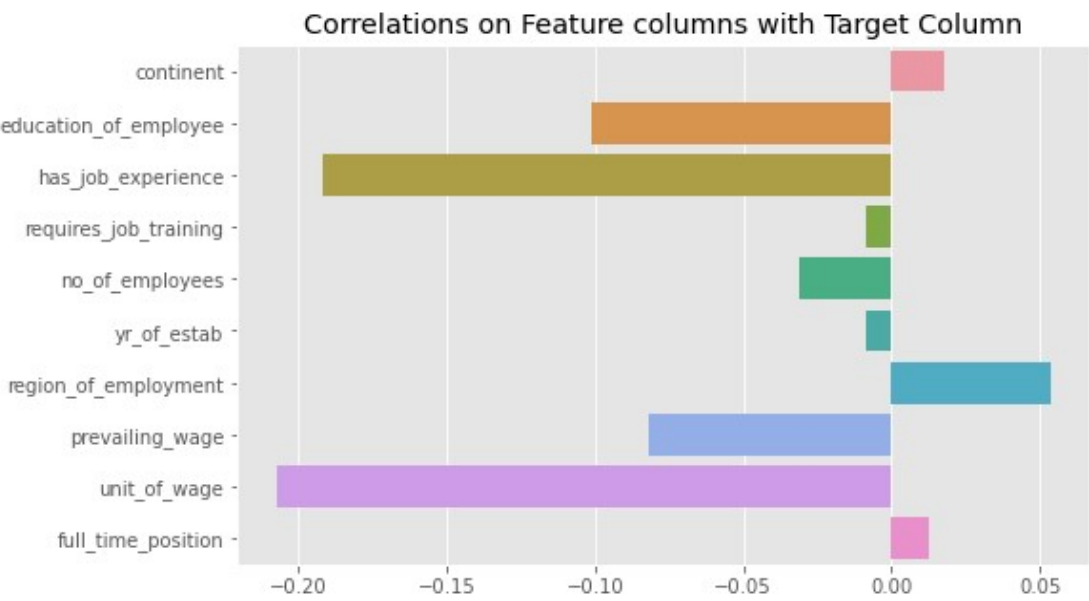
Percentage of Certified by Number of years of establishment

Results:

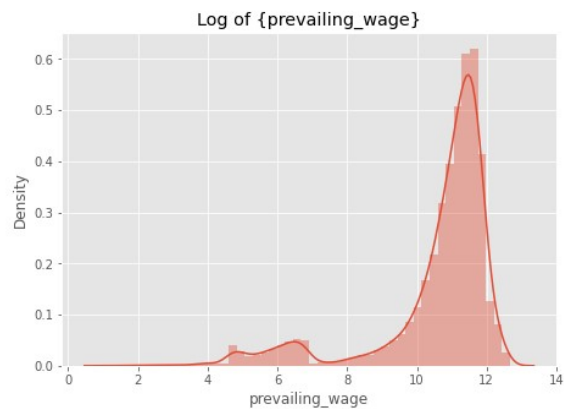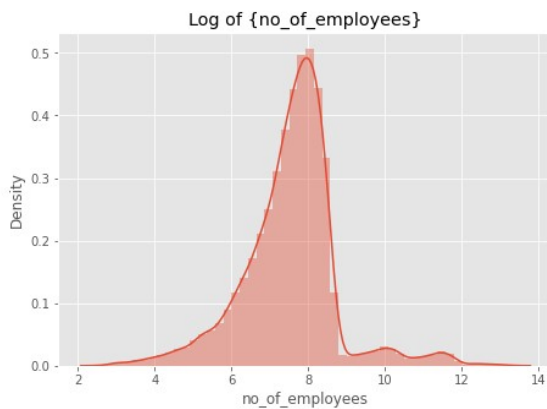As you can see if the company is very old then the success ratio of getting visa will increase.

Note: There is no Great impact of prevailing_wage on case status however if the the salary is high then visa succusse rate will increase

Correlation Graph
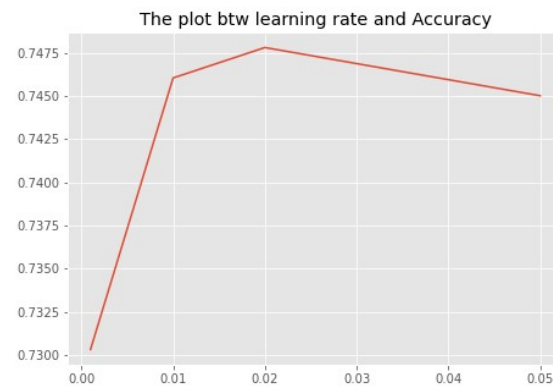

Correlations on Feature columns with Target Column

Predictive Analysis:

In Order to train the model. I have to preprocessing first,In preprocessing i have done a lot of feature engeneering ,I did Log transformation to distribute the data remove the outliers , Normalize the data and balance the classes of target column etc.

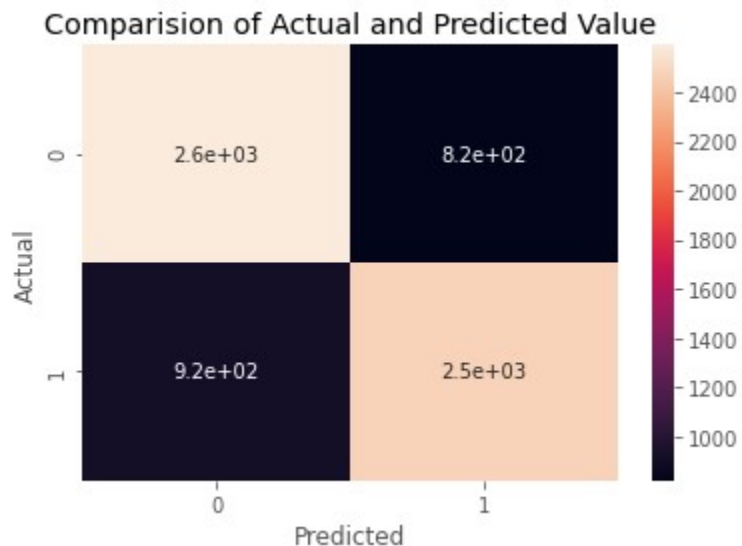Now as you can see the dataset is normally distributed.

Gradient Boosting Model Results:

Classification report:

	precision	recall	f1-score	support
0	0.74	0.76	0.75	3413
1	0.75	0.73	0.74	3395
accuracy			0.75	6808
macro avg	0.75	0.74	0.74	6808
weighted avg	0.75	0.75	0.74	6808

Confusion matrix:

Comparision of Actual and Predicted Value

As you can see that this model shows the accuracy of 75% and the confusion matrix shows that this model correctly predicted 2600 0 class and 2500 1 class out of 6000 values which is good

Conclusion:

I have done the data visualization on Easyvisa. I have found many interesting key points to increase the visa success rate. There were very few features to see the outcomes. If there were information about financial support, Documents and Country then I could find much more accurate results but still not bad. I have found a lot of useful information about the visa success rate.After that i trained the gradient boosting model

How to increase chances of getting a Visa approval?

1) If you have education level more than high school then i will increase the visa approval rate.
2) If the Unit of prevailing wage is Yearly then the approval rate will increase.
3) If you have a job experience then it will much better.
4) IF The year of company establishment is high then the chance of approval will increase. So make sure you have selected the right company.
5) Make sure that the company is located in midwest,south and northeast not on any island.

Model Results:

- Model accuracy is greater than 75%.
- Model precision is greater than 75%.
- recall score is 74%.
- f1 score is 73%.
- Column Has_job_exp is higly correlated with target (case_status) column.

END