CAPSTONE 2

Exploring the factors that influence engagement of online learners

By *Nalini Suresh*

DATASET DESCRIPTION

In 2012, MIT launched Massive Online Open Courses(MOOCs) to students in many countries.

This dataset captures several features about the learners and how they interact with the learning platform.

Some of the features of the series includes but not limited to *the age of the student, level of education, gender, motivation for taking the class, country or region, grade, exploration of the class, and whether they completed the course or not.*

DATASET

- Dataset was downloaded from Kaggle.com

- Course analyzed: 6.00x – Intro to Computer Science & Programming (MITx)

- Year and Semester: 2012 Fall

- Dataset contains 59,279 records, one per student and 31 rows

GOAL OF THE PROJECT

Perform Exploratory Data Analysis on MITx Online Learning Data set.

Determine the correlation of predictor variables and the target variables to help understand the variables that will help students stay engaged and complete the online course.

The understanding of the predictor variables and their correlation to the target variable can help providers of online learning platforms such as edX to improve learning experiences to keep their learners engaged and successful.

DATASET VARIABLES

INPUT VARIABLES

Userid_DI : Student UserID (dtype=object)

Institute: array(['MITx'], dtype=object)

Course_id: (dtype=object)

Year: Year the course was taken (dtype('int64')

Semester: Semester when the class was taken (dtype=object)

Explored: The class material was explored in detail

Country_region: Country of the student (dtype=object)

Degree: education level of the student - array(["Bachelor's", 'Secondary', "Master's", 'Doctorate', 'Less than Secondary'], dtype=object)

Gender: Gender of the student - array(['m', 'f', 'gender not disclosed'], dtype=object)

Age: Age of the student (dtype('int64'))

Motivation_Label: Reason for taking the class (array(['Course Sampling', 'Certification Earning'], dtype=object))

<u>Predicted Variable(Desired Target) </u> - **Completed_or_Not**: Whether Course was completed or not (dtype('int64'))

Key Insights:

Total number of students either viewed, explored, pursued or completed 6.00x – Intro to Computer Science & Programming (MITx) online program recorded in this dataset is 59,279.

- 97% of all students who viewed or explored the program did not complete the program.

- About 3% of all students completed the program.

- It is important to investigate and understand what variables in the dataset will improve the probability of completion among students.

Overwhelming Percentage of students did not complete the online course (Introduction to Computer Science) offered by MITx compared to percentage of student who completed the course (97.06% vs 2.94%).
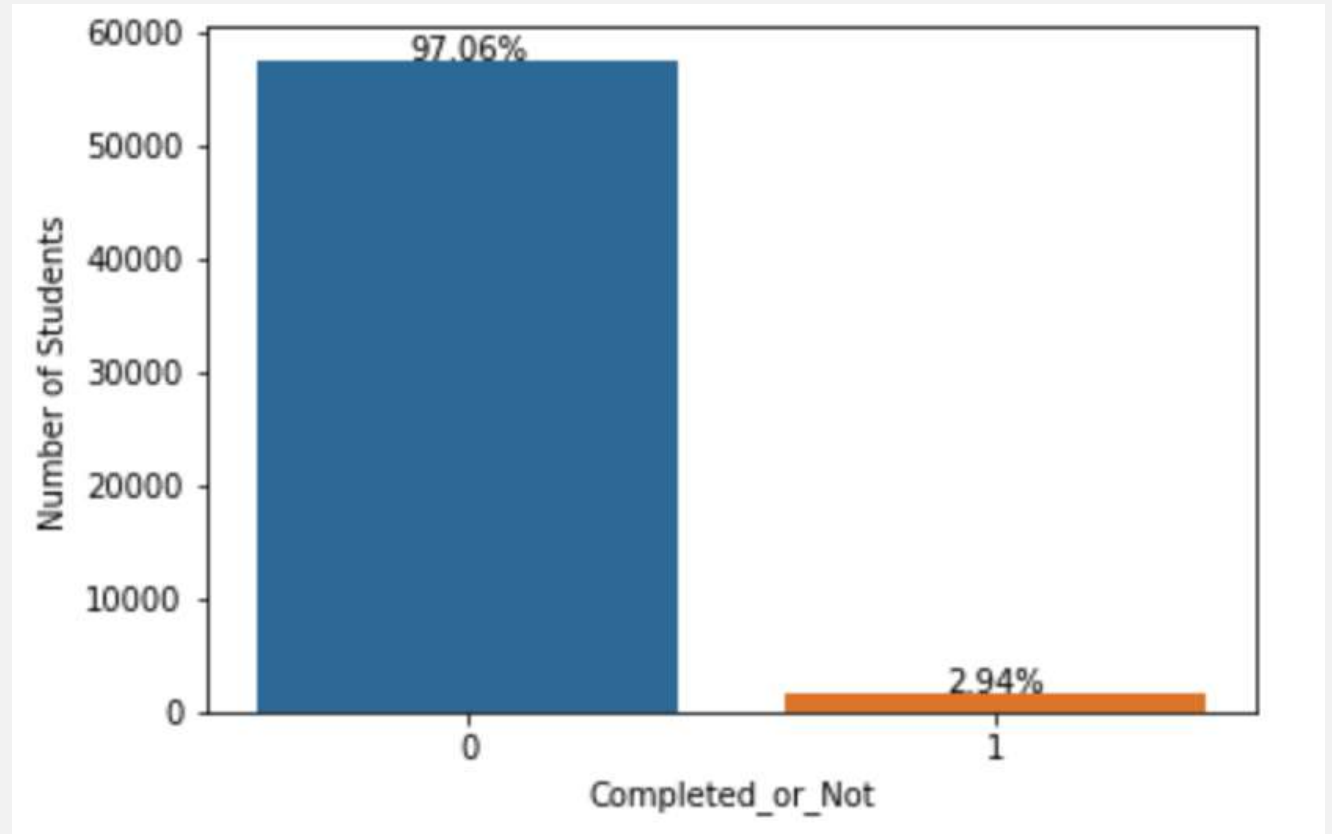
Key Insights:

- There fewer older students and fewer younger students below the age of 20.

- There are more students around the age of 20.

- The higher number of students were around the age of 30.

The average age of the students is between 20 and 30 years old.
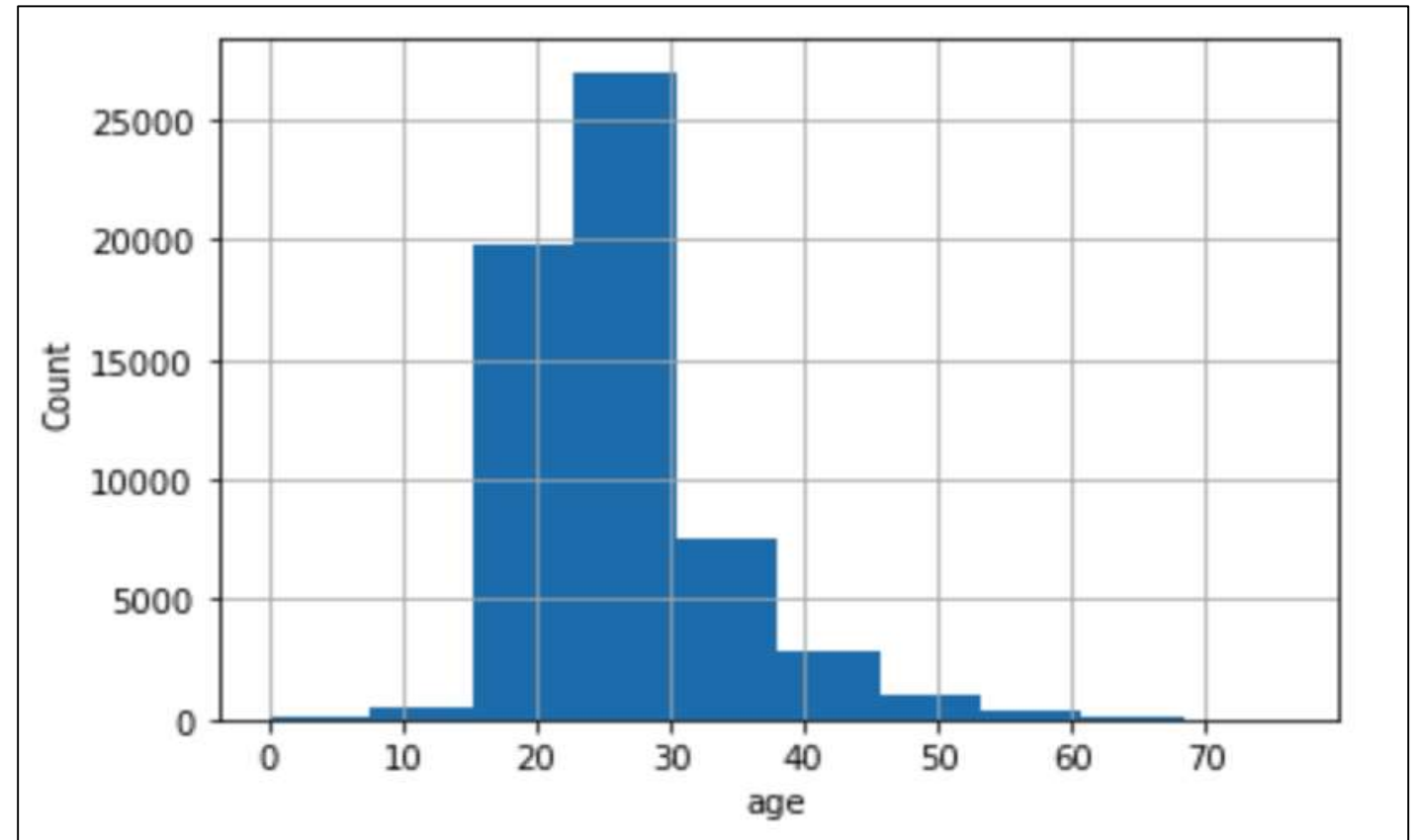
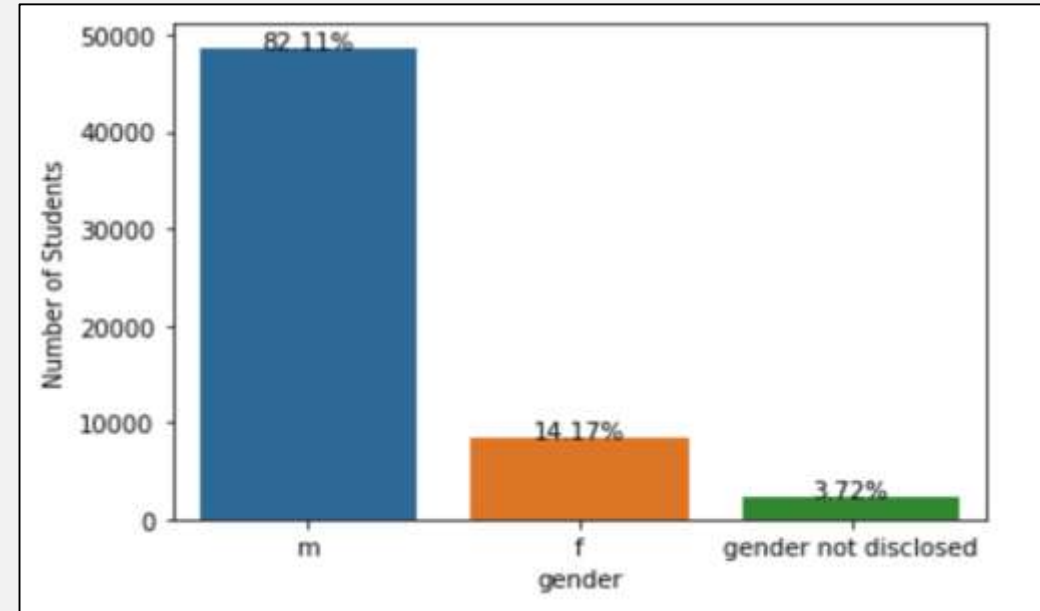## Key Insights:

- The average age for both male and female students is around 25 years of age.

- The maximum and minimum age is also the same for both genders.

- However, there are more male students above the age of 60 taking the class compared to female students.

Male students were more likely to take the online course (Introduction to Computer Science) offered by MITx than female students
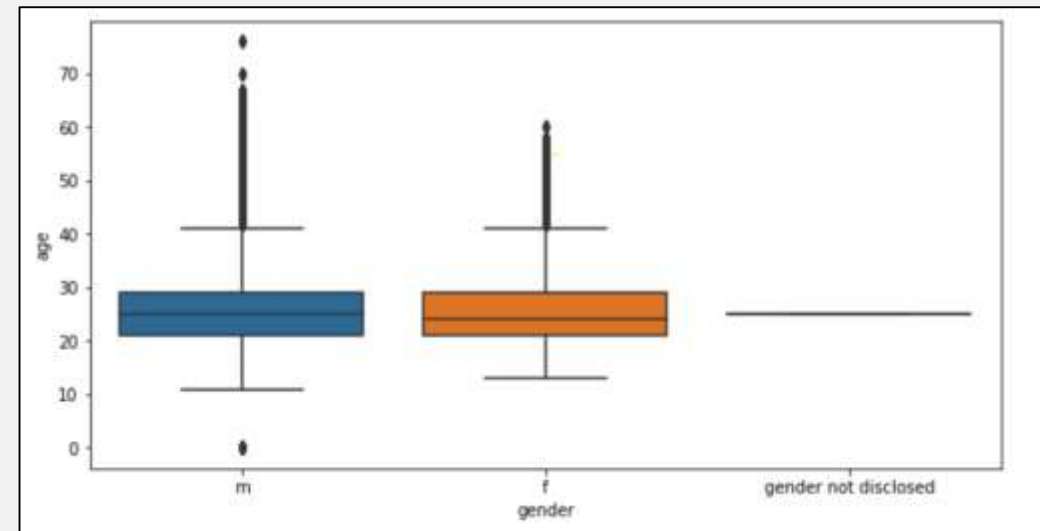
Key Insights:

- Also, male students were overwhelmingly more likely to not complete the course than female students (79.72% vs 13.76%).

- Of 48,674 males, about 79% of males did not complete the program while about 2% of males did complete the program.

- Of 8,399 females, about 14% of females did not complete the program while only about less than 0.5% of females completed the programs.

- Of 2,206 students who did not disclose their gender, about 4% did not complete the program while about 0.1% of them complete the program.

Male students were more likely to complete the online course (Introduction to Computer Science) offered by MITx than female students (2.39% vs 0.41%), which is about 6 times more). Also, male students were overwhelmingly more likely to not complete the course than female students (79.72% vs 13.76%).
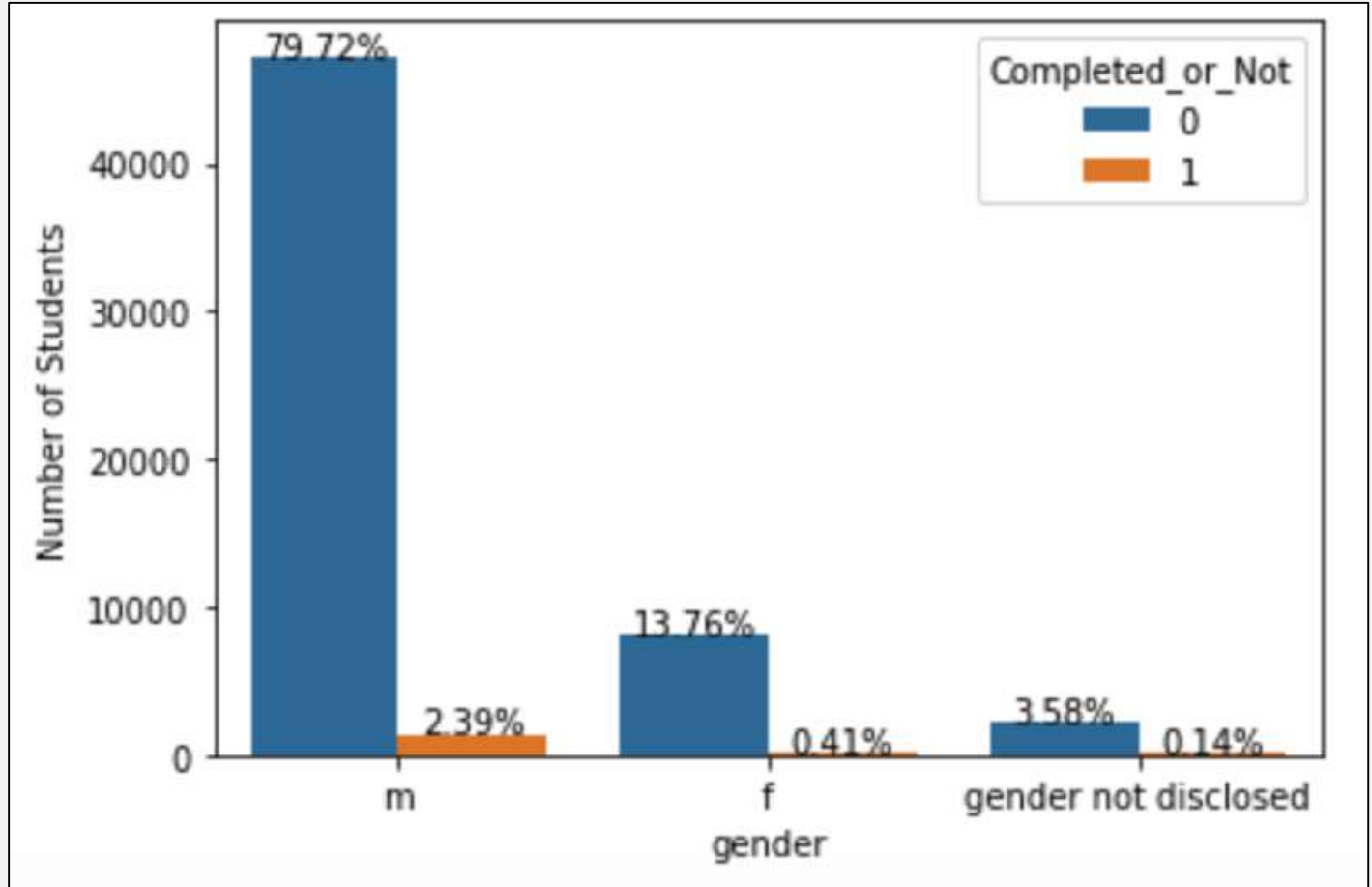
Key Insights:

- About 46% of students have bachelor's degree which is a total of 27,841 students.

- About 32% of students have higher secondary school education which is a total of 18668 students.

- About 18% of students have Master's degree which is a total of 10,820 students.

- About 1% of students have doctorial degree which is a total of 1,226 students.

- About 2% of students have less than a secondary education which is a total of 724 students.

More students with bachelor degree seem to be interested or actively pursuing this online course at MIT across all countries surveyed.
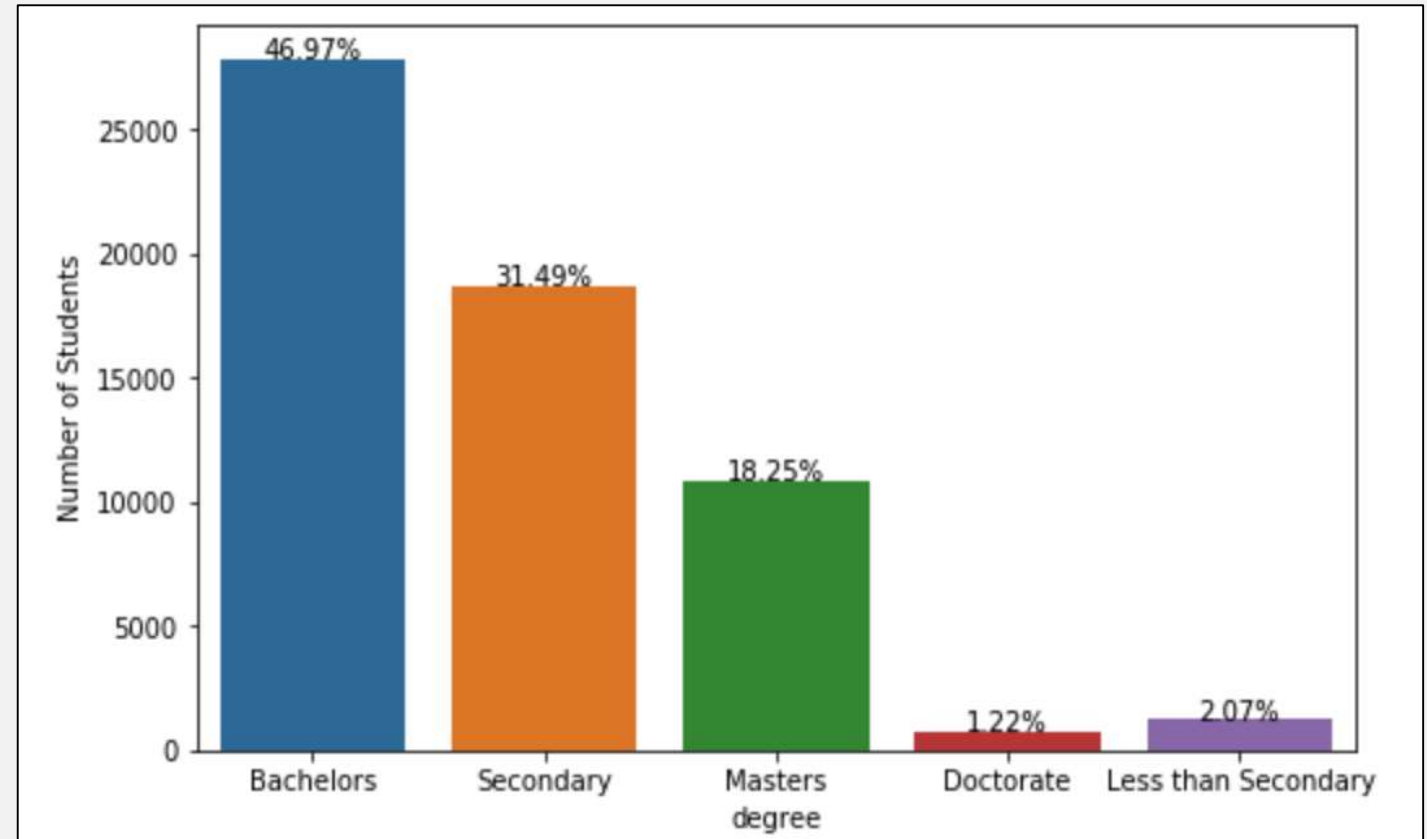
Key Insights:

- About 46% of students have bachelor's degree which is a total of 27,841 students.

- The average age of students completing and not completing the course with bachelor degree is around 30.

More students with bachelor degree seem to be interested this online course at MIT, but did not complete the course. They are also more likely to complete the course.
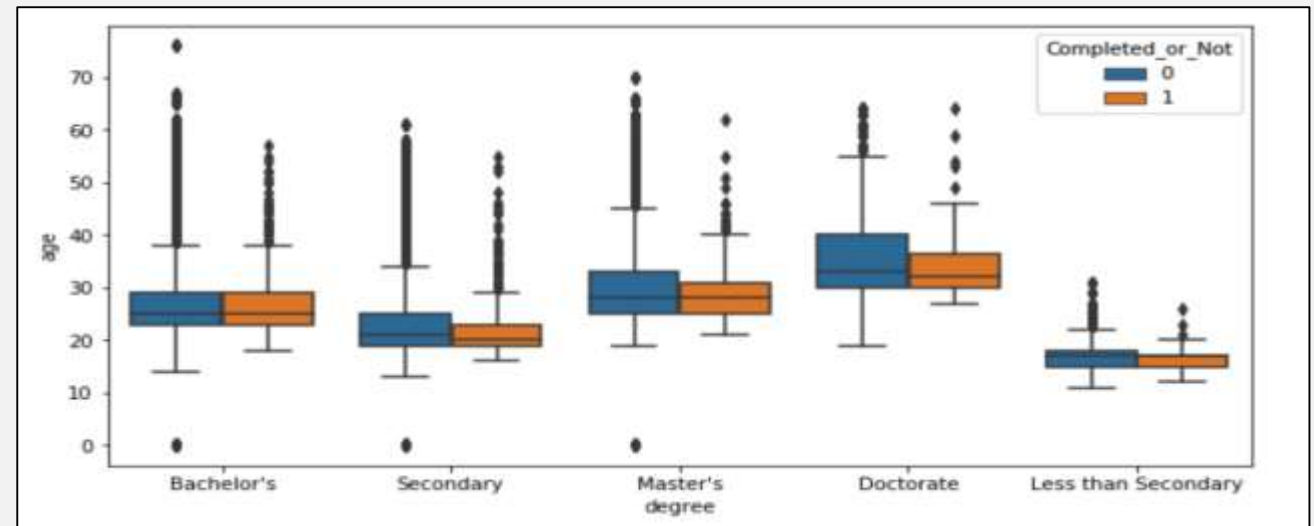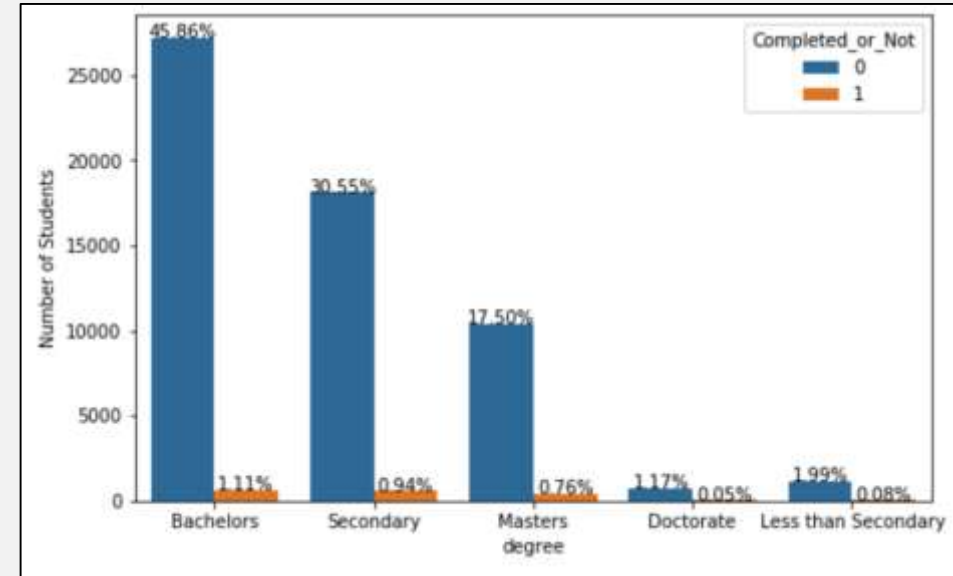
Key Insights:

- About 2.5 percent of the total students who viewed the online course explored the course.

- However about 3% of the students completed the course.

- It also shows about 94% of students viewed the course but did not explore it.

Students who explored the online class were successful in completing the course rather than students who simply viewed the course.
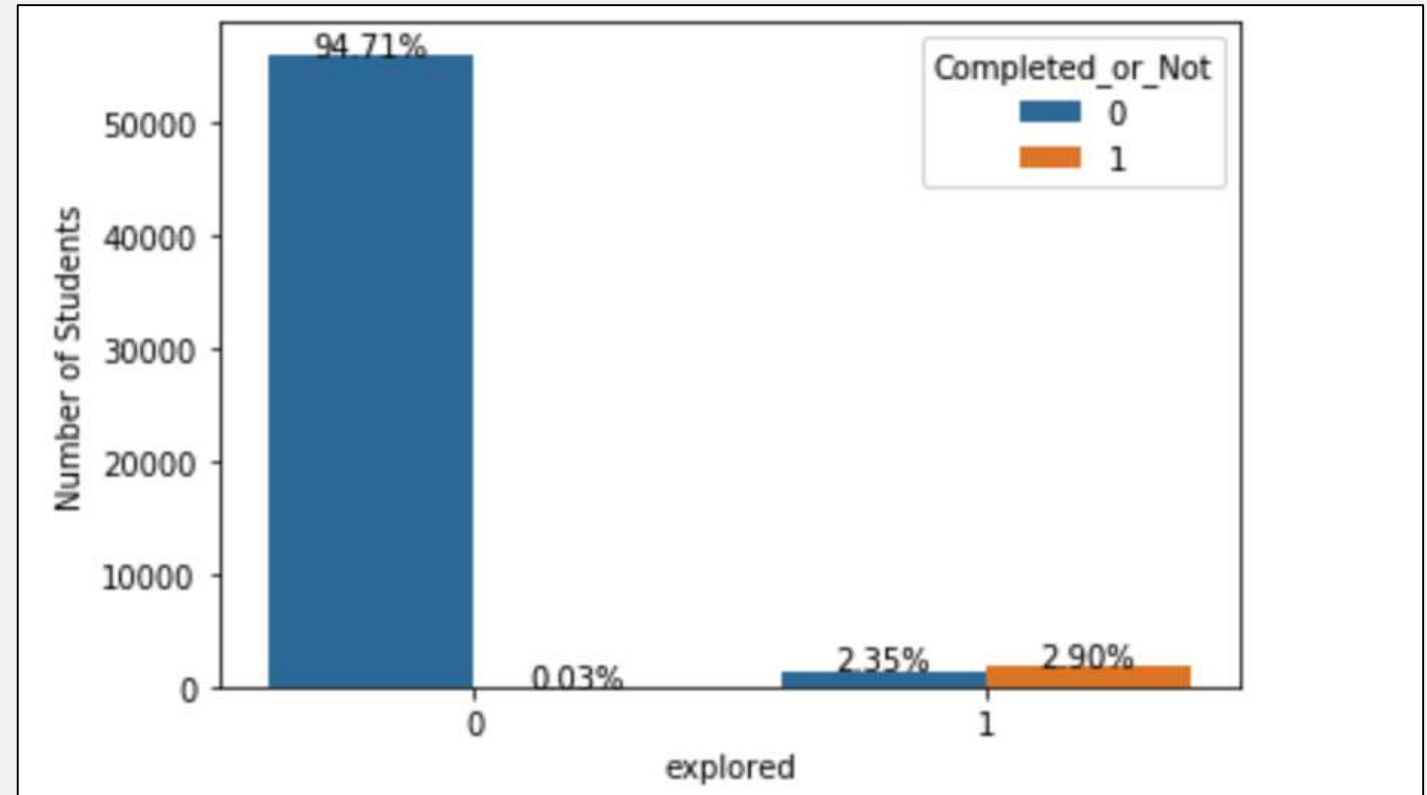
Key Insights:

- Two countries, United States and India have high percent participation in the course across all education levels.

- United states has higher percentage of students participating compared to India across all education levels.

United states and India have the high percent participation in the course across all education levels.

| country_region \ degree	Bachelors	Doctorate	Less than Secondary	Masters	Secondary	All
Australia	0.881543	0.000000	0.035262	0.293113	0.493664	1.703581
Brazil	1.996694	0.000000	0.088154	0.279890	1.674931	4.039669
Canada	1.207713	0.000000	0.072727	0.204959	0.764738	2.250138
Germany	0.478237	0.066116	0.000000	0.822039	0.385675	1.752066
Greece	0.286501	0.000000	0.000000	0.114601	0.198347	0.599449
India	9.458953	0.039669	0.297521	2.261157	7.601102	19.658402
Indonesia	0.566391	0.000000	0.011019	0.022039	0.599449	1.198898
Other Africa	3.687052	0.061708	0.063912	0.923416	1.899725	6.635813
Other Europe	2.794490	0.088154	0.346006	2.109091	3.537190	8.874931
Other South America	1.115152	0.000000	0.000000	0.392287	0.800000	2.307438
Poland	0.169697	0.000000	0.028650	0.306336	0.465014	0.969697
Russian Federation	1.090909	0.024242	0.198347	2.142149	1.256198	4.711846
Spain	0.861708	0.057300	0.000000	1.395041	0.359229	2.673278
Ukraine	0.504683	0.000000	0.070523	0.976309	0.407713	1.959229
United Kingdom	2.166391	0.110193	0.024242	1.260606	1.287052	4.848485
United States	16.863912	1.062259	1.223140	5.890909	10.776860	35.817080
All	44.130028	1.509642	2.459504	19.393939	32.506887	100.000000

Key Insights:

- Two countries, United States and India have high percent participation in the course across all genders.

- United states has higher percentage of both male and female students participating compared to India.

- Also, in both countries, percentage of male students are much higher

United states and India have the high percentage of participation in the course across all both male and female students.

gender / country_region	f	gender not disclosed	m	All
Australia	0.130028	0.092562	1.480992	1.703581
Brazil	0.337190	0.059504	3.642975	4.039669
Canada	0.182920	0.132231	1.934986	2.250138
Germany	0.143251	0.156474	1.452342	1.752066
Greece	0.011019	0.063912	0.524518	0.599449
India	2.884848	0.310744	16.462810	19.658402
Indonesia	0.299725	0.011019	0.888154	1.198898
Other Africa	0.599449	0.116804	5.919559	6.635813
Other Europe	1.240771	0.442975	7.191185	8.874931
Other South America	0.163085	0.101377	2.042975	2.307438
Poland	0.066116	0.066116	0.837466	0.969697
Russian Federation	0.588430	0.229201	3.894215	4.711846
Spain	0.394490	0.092562	2.186226	2.673278
Ukraine	0.301928	0.169697	1.487603	1.959229
United Kingdom	0.539945	0.154270	4.154270	4.848485
United States	7.325620	1.648485	26.842975	35.817080
All	15.208815	3.847934	80.943251	100.000000
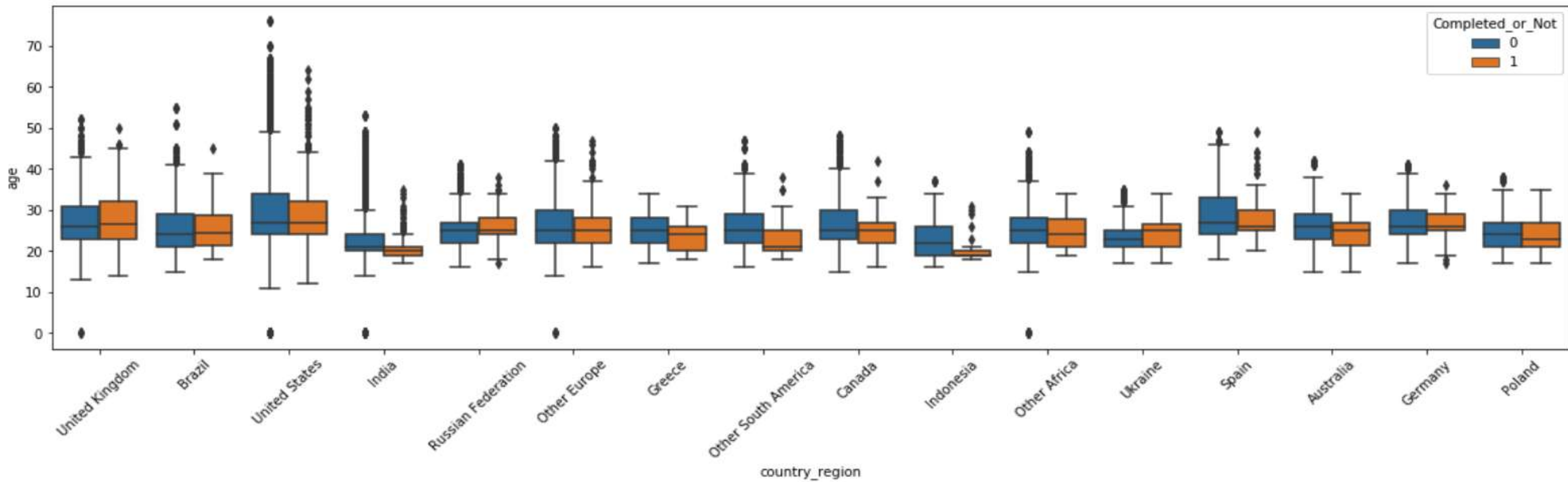
COURSE COMPLETION IN TOP 16
COUNTRIES BY AGE

Key Insights:

- Average age of students completing and not completing the class are significantly different for United States and India. Average age in the US is about 28 years and for India it is about 20 years.
- Also, there are more students above the age of 50 years in the US compared to other states.

Key Insights:

- "Explored" has the highest correlation of whether or not a student completes the online course.

- This shows that after viewing the class, their interest has peaked to a point where they thought is is worth completing the class.

KEY INSIGHTS:

- "Explored" has the highest correlation of whether or not a student completes the online course.

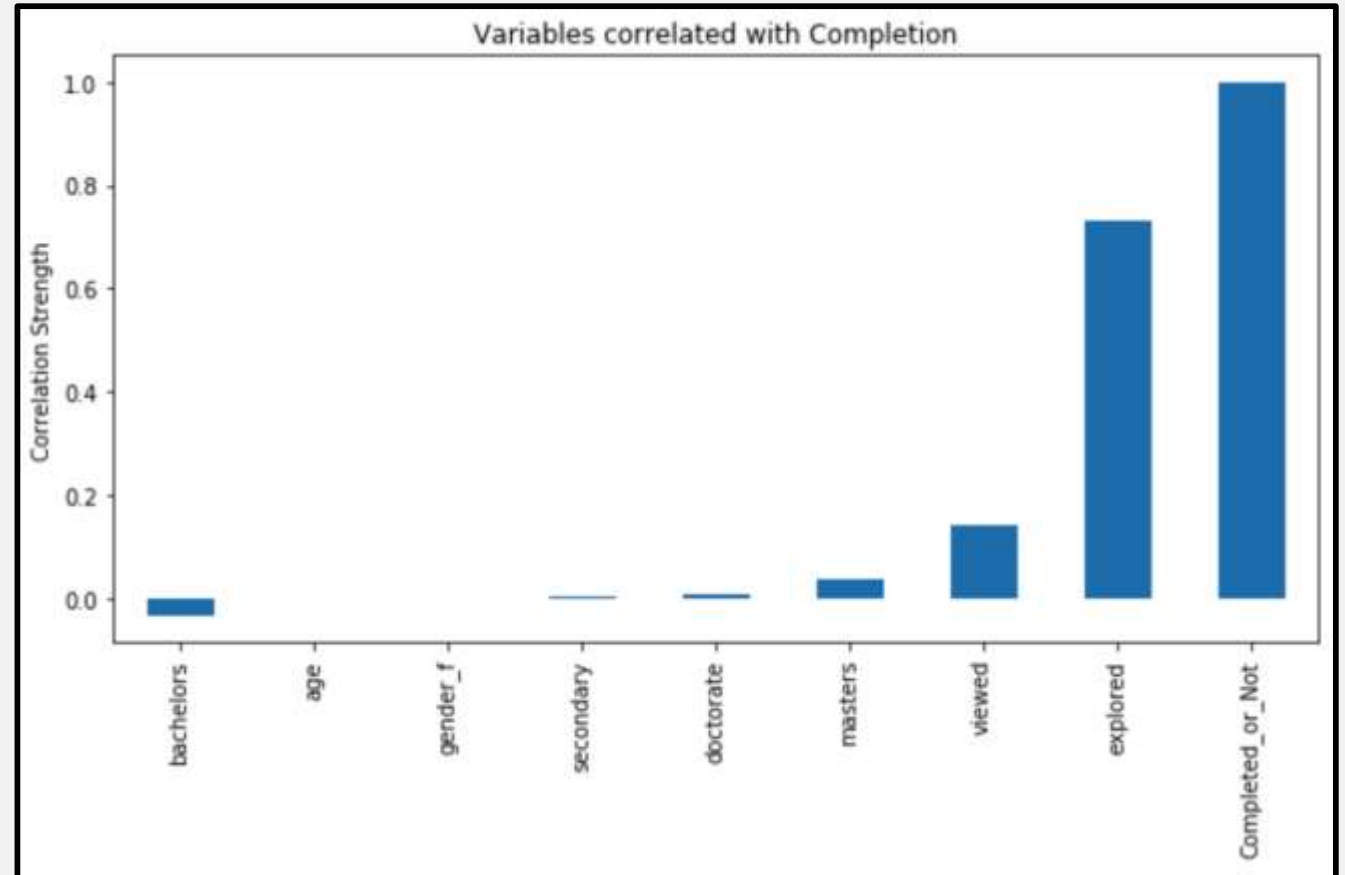- This shows that after viewing the class, their interest has peaked to a point where they thought is is worth completing the class.

Variables correlated with Completion
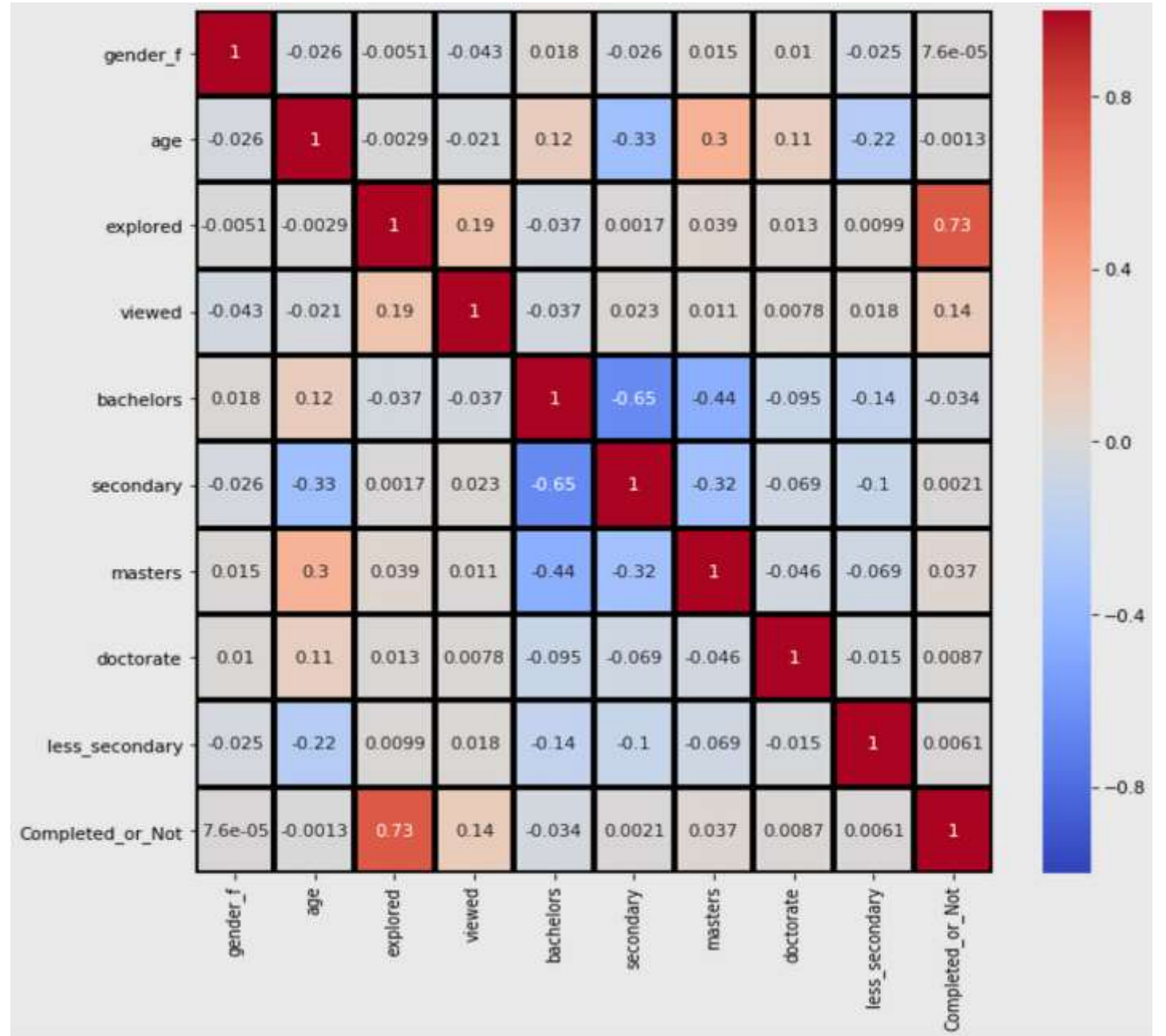
Inference

Approximately, little more than half of the observed variation can be explained by the model's inputs. Also, a p-value indicator suggests that students completion rate was higher when they *explored* the course.

```
                           OLS Regression Results
==============================================================================
Dep. Variable:        Completed_or_Not   R-squared (uncentered):           0.550
Model:                             OLS   Adj. R-squared (uncentered):      0.550
Method:                  Least Squares   F-statistic:                      7673.
Date:                 Mon, 08 Jun 2020   Prob (F-statistic):                0.00
Time:                         12:00:22   Log-Likelihood:                  41751.
No. Observations:                56586   AIC:                          -8.348e+04
Df Residuals:                    56577   BIC:                          -8.340e+04
Df Model:                            9
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
gender_f       0.0057      0.002      2.337      0.019       0.001       0.010
gender_m       0.0041      0.002      1.936      0.053   -5.18e-05       0.008
age         -9.303e-05   9.44e-05     -0.985      0.325      -0.000    9.21e-05
explored       0.5550      0.002    250.501      0.000       0.551       0.559
viewed         0.0005      0.001      0.458      0.647      -0.002       0.002
bachelors     -0.0028      0.003     -1.008      0.314      -0.008       0.003
secondary     -0.0018      0.003     -0.695      0.487      -0.007       0.003
masters        0.0015      0.003      0.487      0.626      -0.005       0.008
doctorate     -0.0027      0.006     -0.471      0.638      -0.014       0.009
==============================================================================
```

FINDINGS & IMPLICATIONS

RESULT OF ANALYSIS

Students who actively explored the course, completed the course.

In many countries, the trend in the data shows that male students explored and completed the online classes at a higher rate than female students.

The average age of students who explored and completed the course was around 28 years of age.

The highest percentage of students who completed the course had bachelor's degree.

Percentage of students who completed the online course was only 3% compared to 97% of students who viewed or explored the course.

Percentage of students who explored and completed the class was highest in the United States followed by India. Participation in other countries was moderate.

APPENDIX

FURTHER ANALYSIS

- Please check the following Tableau link for visualization for analysis by country.

- https://public.tableau.com/views/Capstone2FINALProject/GeoMap?:display_count=y&:origin=viz_share_link