

# Salary Prediction using AMCAT dataset

## 1.0 Importing packages

In [392]:

```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn import preprocessing
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import GridSearchCV
from scipy.sparse import hstack
import warnings
warnings.filterwarnings("ignore")
```

## 1.1 Importing data

In [369]:

```
train = pd.read_excel('Data/train.xlsx')
test = pd.read_excel('Data/test.xlsx')
```

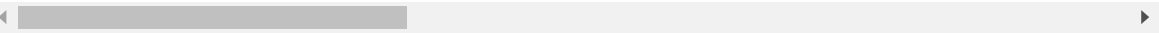
In [6]:

```
train.head()
```

Out[6]:

	Unnamed: 0	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10pe
0	train	203097	420000	2012-06-01	present	senior quality engineer	Bangalore	f	1990-02-19	
1	train	579905	500000	2013-09-01	present	assistant manager	Indore	m	1989-10-04	
2	train	810601	325000	2014-06-01	present	systems engineer	Chennai	f	1992-08-03	
3	train	267447	1100000	2011-07-01	present	senior software engineer	Gurgaon	m	1989-12-05	
4	train	343523	200000	2014-03-01	2015-03-01 00:00:00	get	Manesar	m	1991-02-27	

5 rows × 39 columns



In [7]:

```
test.head()
```

Out[7]:

	Unnamed: 0	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percenta
0	test	664736	?	?	?	?	?	m	1992-01-16	75
1	test	1123290	?	?	?	?	?	m	1992-06-05	83
2	test	1062444	?	?	?	?	?	f	1992-11-22	85
3	test	1072028	?	?	?	?	?	f	1990-10-17	81
4	test	267259	?	?	?	?	?	m	1990-03-20	78

5 rows × 39 columns



# 1.2 Understanding high level details about data

## 1.2.1 Missing values

In [9]:

```
train.isnull().sum()
```

Out[9]:

Unnamed: 0	0
ID	0
Salary	0
DOJ	0
DOL	0
Designation	0
JobCity	0
Gender	0
DOB	0
10percentage	0
10board	0
12graduation	0
12percentage	0
12board	0
CollegeID	0
CollegeTier	0
Degree	0
Specialization	0
collegeGPA	0
CollegeCityID	0
CollegeCityTier	0
CollegeState	0
GraduationYear	0
English	0
Logical	0
Quant	0
Domain	0
ComputerProgramming	0
ElectronicsAndSemicon	0
ComputerScience	0
MechanicalEngg	0
ElectricalEngg	0
TelecomEngg	0
CivilEngg	0
conscientiousness	0
agreeableness	0
extraversion	0
neuroticism	0
openness_to_experience	0
dtype: int64	

In [10]:

```
test.isnull().sum()
```

Out[10]:

Unnamed: 0	0
ID	0
Salary	0
DOJ	0
DOL	0
Designation	0
JobCity	0
Gender	0
DOB	0
10percentage	0
10board	0
12graduation	0
12percentage	0
12board	0
CollegeID	0
CollegeTier	0
Degree	0
Specialization	0
collegeGPA	0
CollegeCityID	0
CollegeCityTier	0
CollegeState	0
GraduationYear	0
English	0
Logical	0
Quant	0
Domain	0
ComputerProgramming	0
ElectronicsAndSemicon	0
ComputerScience	0
MechanicalEngg	0
ElectricalEngg	0
TelecomEngg	0
CivilEngg	0
conscientiousness	0
agreeableness	0
extraversion	0
neuroticism	0
openness_to_experience	0
dtype: int64	

There are no missing values in the training data and testing data.

### 1.2.2 Other information

In [11]:

```
train.describe().T
```

Out[11]:

	count	mean	std	min	25%	75%
ID	3998.0	663794.540520	363218.245829	11244.0000	334284.250000	663794.540520
Salary	3998.0	307699.849925	212737.499957	35000.0000	180000.000000	307699.849925
10percentage	3998.0	77.925443	9.850162	43.0000	71.680000	83.160000
12graduation	3998.0	2008.087544	1.653599	1995.0000	2007.000000	2008.087544
12percentage	3998.0	74.466366	10.999933	40.0000	66.000000	82.873333
CollegeID	3998.0	5156.851426	4802.261482	2.0000	494.000000	5156.851426
CollegeTier	3998.0	1.925713	0.262270	1.0000	2.000000	1.925713
collegeGPA	3998.0	71.486171	8.167338	6.4500	66.407500	71.486171
CollegeCityID	3998.0	5156.851426	4802.261482	2.0000	494.000000	5156.851426
CollegeCityTier	3998.0	0.300400	0.458489	0.0000	0.000000	0.300400
GraduationYear	3998.0	2012.105803	31.857271	0.0000	2012.000000	2012.105803
English	3998.0	501.649075	104.940021	180.0000	425.000000	501.649075
Logical	3998.0	501.598799	86.783297	195.0000	445.000000	501.598799
Quant	3998.0	513.378189	122.302332	120.0000	430.000000	513.378189
Domain	3998.0	0.510490	0.468671	-1.0000	0.342315	0.510490
ComputerProgramming	3998.0	353.102801	205.355519	-1.0000	295.000000	353.102801
ElectronicsAndSemicon	3998.0	95.328414	158.241218	-1.0000	-1.000000	95.328414
ComputerScience	3998.0	90.742371	175.273083	-1.0000	-1.000000	90.742371
MechanicalEngg	3998.0	22.974737	98.123311	-1.0000	-1.000000	22.974737
ElectricalEngg	3998.0	16.478739	87.585634	-1.0000	-1.000000	16.478739
TelecomEngg	3998.0	31.851176	104.852845	-1.0000	-1.000000	31.851176
CivilEngg	3998.0	2.683842	36.658505	-1.0000	-1.000000	2.683842
conscientiousness	3998.0	-0.037831	1.028666	-4.1267	-0.713525	-0.037831
agreeableness	3998.0	0.146496	0.941782	-5.7816	-0.287100	0.146496
extraversion	3998.0	0.002763	0.951471	-4.6009	-0.604800	0.002763
neuroticism	3998.0	-0.169033	1.007580	-2.6430	-0.868200	-0.169033
openness_to_experience	3998.0	-0.138110	1.008075	-7.3757	-0.669200	-0.138110

In [12]:

```
test.describe().T
```

Out[12]:

	count	mean	std	min	25%	75%
ID	1500.0	665286.300667	360532.421901	7474.0000	335062.500000	641000.000000
10percentage	1500.0	78.384553	9.565983	43.0800	72.000000	84.000000
12graduation	1500.0	2008.122000	1.588542	2000.0000	2007.000000	2009.000000
12percentage	1500.0	74.947040	10.632432	40.8300	67.000000	83.000000
CollegeID	1500.0	5202.454667	4750.131676	2.0000	830.000000	3300.000000
CollegeTier	1500.0	1.926667	0.260770	1.0000	2.000000	2.000000
collegeGPA	1500.0	71.615147	8.747405	7.0000	67.147500	77.000000
CollegeCityID	1500.0	5202.454667	4750.131676	2.0000	830.000000	3300.000000
CollegeCityTier	1500.0	0.278000	0.448163	0.0000	0.000000	0.000000
GraduationYear	1500.0	2012.621333	1.288559	2008.0000	2012.000000	2013.000000
English	1500.0	501.883333	103.976105	195.0000	430.000000	570.000000
Logical	1500.0	501.210000	87.208808	215.0000	445.000000	560.000000
Quant	1500.0	514.505333	118.143311	120.0000	433.750000	570.000000
Domain	1500.0	0.509663	0.460305	-1.0000	0.352981	0.657019
ComputerProgramming	1500.0	357.774667	202.022875	-1.0000	305.000000	410.000000
ElectronicsAndSemicon	1500.0	93.853333	155.940551	-1.0000	-1.000000	187.000000
ComputerScience	1500.0	84.992000	171.721189	-1.0000	-1.000000	170.000000
MechanicalEngg	1500.0	22.992667	99.572364	-1.0000	-1.000000	115.000000
ElectricalEngg	1500.0	20.673333	98.467198	-1.0000	-1.000000	110.000000
TelecomEngg	1500.0	34.525333	106.704076	-1.0000	-1.000000	120.000000
CivilEngg	1500.0	3.997333	43.220335	-1.0000	-1.000000	50.000000
conscientiousness	1500.0	-0.038361	1.021743	-3.5085	-0.733500	0.656778
agreeableness	1500.0	0.183612	0.858094	-4.2831	-0.287100	0.656778
extraversion	1500.0	0.048418	0.919562	-3.3713	-0.598000	0.656778
neuroticism	1500.0	-0.091588	1.010601	-2.6430	-0.868200	0.656778
openness_to_experience	1500.0	-0.102384	0.908886	-5.8428	-0.669200	0.656778

## 1.3 Exploratory Data analysis

### 1.3.1 Visualization

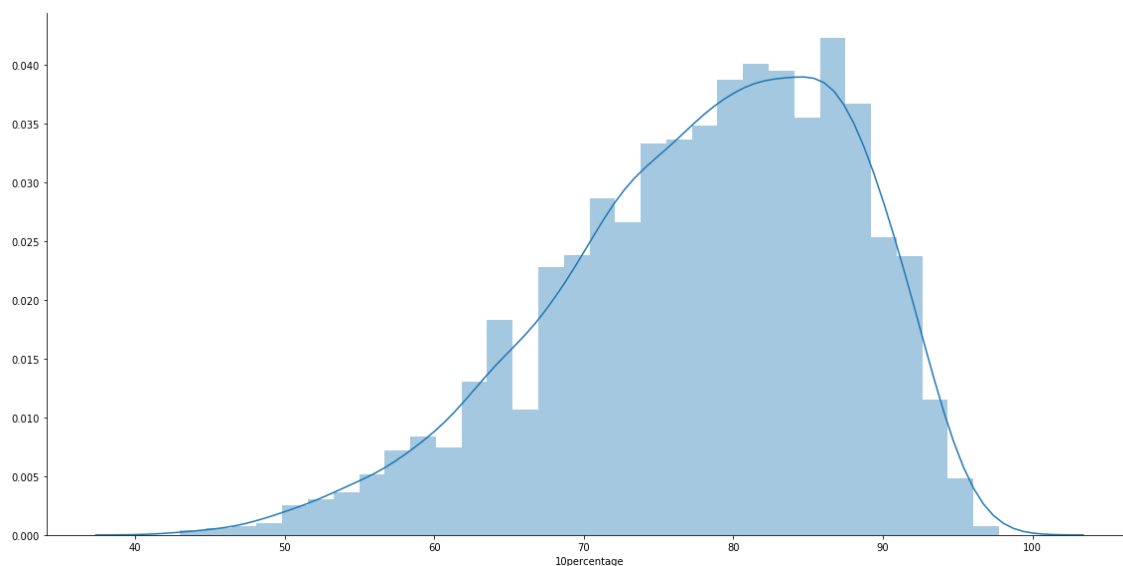
#### 1.3.1.1 '10'nth percentage attribute

In [17]:

```
sns.FacetGrid(data=train,height=8,aspect=2).map(sns.distplot,"10percentage").add_
_legend()
```

Out[17]:

<seaborn.axisgrid.FacetGrid at 0x7f22a83f73c8>

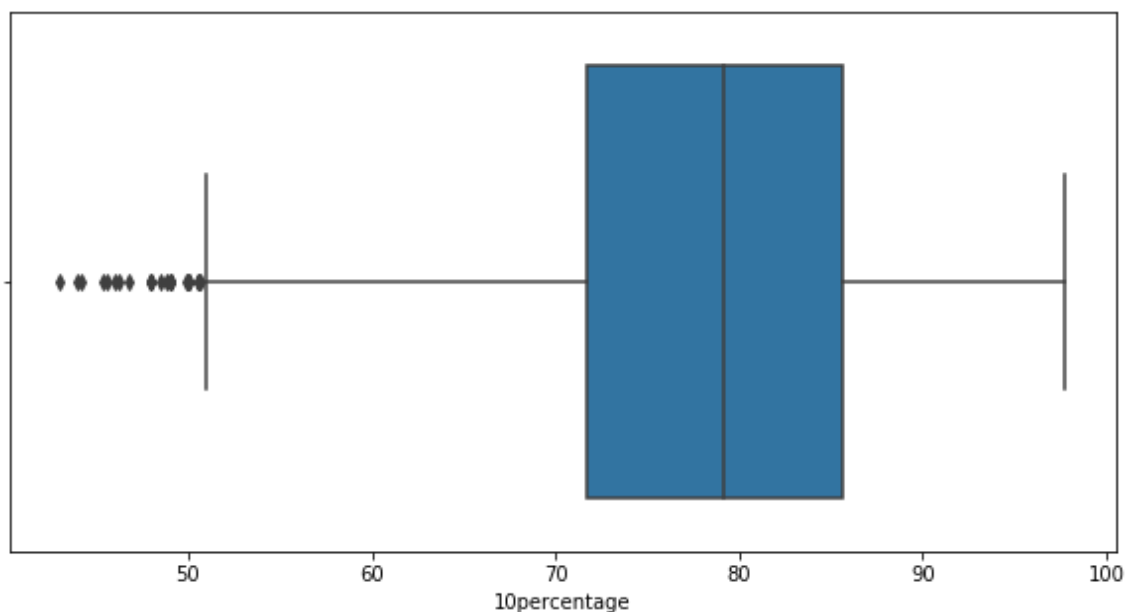


In [24]:

```
plt.figure(figsize=(10,5))
sns.boxplot(train['10percentage'])
```

Out[24]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f229ea528d0>



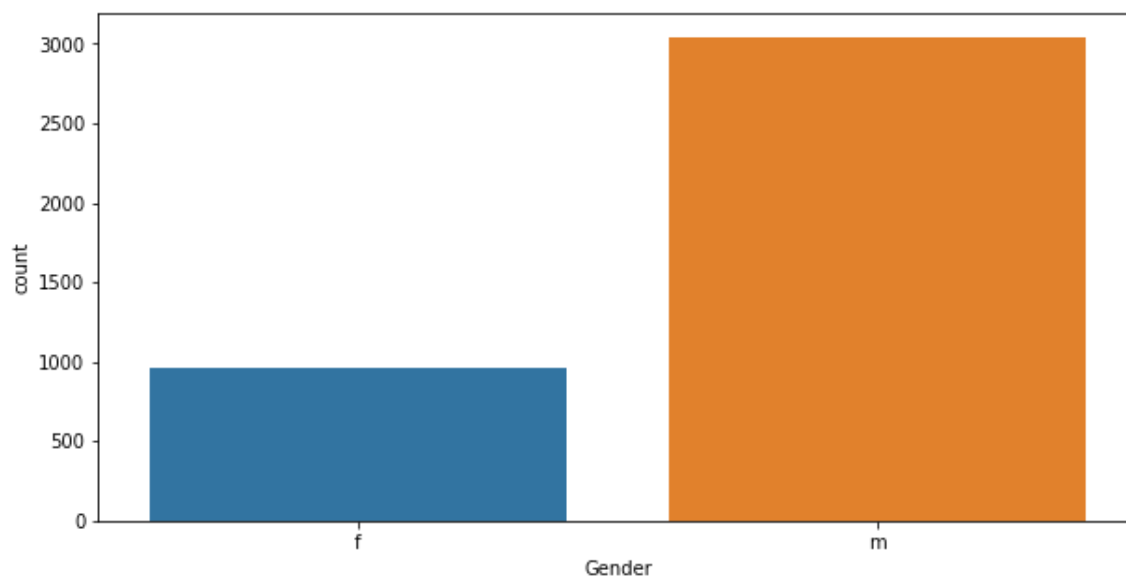
### 1.3.1.2 Gender attribute

In [29]:

```
plt.figure(figsize=(10,5))  
sns.countplot(train['Gender'])
```

Out[29]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f229e8ff550>



### 1.3.1.3 '10'nth Board attribute

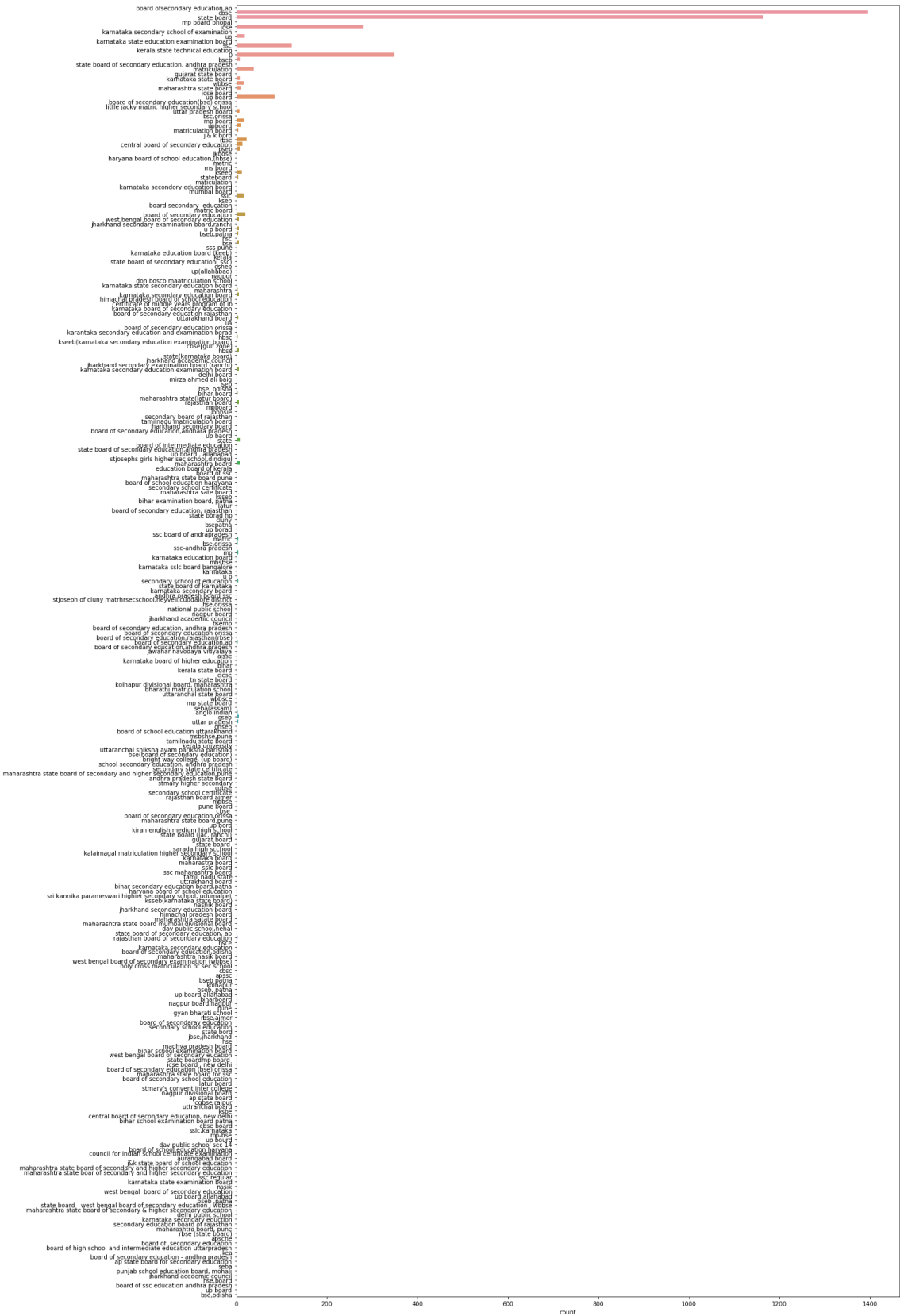


In [86]:

```
plt.figure(figsize=(20,40))  
sns.countplot(y=train['10board'])
```

Out[86]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f2296e0d2e8>





In [41]:

```
train['10board'].unique()
```

Out[41]:

```
array(['board ofsecondary education,ap', 'cbse', 'state board',
      'mp board bhopal', 'icse',
      'karnataka secondary school of examination', 'up',
      'karnataka state education examination board', 'ssc',
      'kerala state technical education', 0, 'bseb',
      'state board of secondary education, andhra pradesh',
      'matriculation', 'gujarat state board', 'karnataka state boar
d',
      'wbbsc', 'maharashtra state board', 'icse board', 'up board',
      'board of secondary education(bse) orissa',
      'little jacky matric higher secondary school',
      'uttar pradesh board', 'bsc,orissa', 'mp board', 'upboard',
      'matriculation board', 'j & k bord', 'rbse',
      'central board of secondary education', 'pseb', 'jkbose',
      'haryana board of school education,(hbse)', 'metric', 'ms boar
rd',
      'kseeb', 'stateboard', 'maticulation',
      'karnataka secondary education board', 'mumbai board', 'ssl
c',
      'kseeb', 'board secondary education', 'matric board',
      'board of secondary education',
      'west bengal board of secondary education',
      'jharkhand secondary examination board,ranchi', 'u p board',
      'bseb,patna', 'hsc', 'bse', 'sss pune',
      'karnataka education board (keeb)', 'kerala',
      'state board of secondary education( ssc)', 'gsheb',
      'up(allahabad)', 'nagpur', 'don bosco maatriculation school',
      'karnataka state secondary education board', 'maharashtra',
      'karnataka secondary education board',
      'himachal pradesh board of school education',
      'certificate of middle years program of ib',
      'karnataka board of secondary education',
      'board of secondary education rajasthan', 'uttarakhand boar
d',
      'ua', 'board of secendary education orissa',
      'karantaka secondary education and examination borad', 'hbs
c',
      'kseeb(karnataka secondary education examination board)',
      'cbse[gulf zone]', 'hbse', 'state(karnataka board)',
      'jharkhand accademic council',
      'jharkhand secondary examination board (ranchi)',
      'karnataka secondary education examination board', 'delhi boa
rd',
      'mirza ahmed ali baig', 'jseb', 'bse, odisha', 'bihar board',
      'maharashtra state(latur board)', 'rajasthan board', 'mpboar
d',
      'upbhsie', 'secondary board of rajasthan',
      'tamilnadu matriculation board', 'jharkhand secondary board',
      'board of secondary education,andhara pradesh', 'up baord',
      'state', 'board of intermediate education',
      'state board of secondary education,andhra pradesh',
      'up board , allahabad',
      'stjosephs girls higher sec school,dindigul', 'maharashtra bo
ard',
      'education board of kerala', 'board of ssc',
      'maharashtra state board pune',
      'board of school education harayana',
      'secondary school cerfificate', 'maharashtra sate board', 'ks
seb',
```

'bihar examination board, patna', 'latur',  
 'board of secondary education, rajasthan', 'state borad hp',  
 'cluny', 'bsepatna', 'up borad', 'ssc board of andrapradesh',  
 'matric', 'bse,orissa', 'ssc-andhra pradesh', 'mp',  
 'karnataka education board', 'mhsbse',  
 'karnataka sslc board bangalore', 'karnataka', 'u p',  
 'secondary school of education', 'state board of karnataka',  
 'karnataka secondary board', 'andhra pradesh board ssc',  
 'stjoseph of cluny matrhrsecschool,neyveli,cuddalore distric  
 t',  
 'hse,orissa', 'national public school', 'nagpur board',  
 'jharkhand academic council', 'bsemp',  
 'board of secondary education, andhra pradesh',  
 'board of secondary education orissa',  
 'board of secondary education,rajasthan(rbse)',  
 'board of secondary education,ap',  
 'board of secondary education,andhra pradesh',  
 'jawahar navodaya vidyalaya', 'aisse',  
 'karnataka board of higher education', 'bihar',  
 'kerala state board', 'cicse', 'tn state board',  
 'kolhapur divisional board, maharashtra',  
 'bharathi matriculation school', 'uttaranchal state board',  
 'wbbsce', 'mp state board', 'seba(assam)', 'anglo indian', 'g  
 seb',  
 'uttar pradesh', 'ghseb', 'board of school education uttarakh  
 and',  
 'msbshse,pune', 'tamilnadu state board', 'kerala university',  
 'uttaranchal shiksha avam pariksha parishad',  
 'bse(board of secondary education)',  
 'bright way college, (up board)',  
 'school secondary education, andhra pradesh',  
 'secondary state certificate',  
 'maharashtra state board of secondary and higher secondary ed  
 ucation,pune',  
 'andhra pradesh state board', 'stmary higher secondary', 'cgb  
 se',  
 'secondary school certificate', 'rajasthan board ajmer', 'mpb  
 se',  
 'pune board', 'cbse ', 'board of secondary education,orissa',  
 'maharashtra state board,pune', 'up bord',  
 'kiran english medium high school', 'state board (jac, ranch  
 i)',  
 'gujarat board', 'state board ', 'sarada high scschool',  
 'kalaimagal matriculation higher secondary school',  
 'karnataka board', 'maharashtra board', 'sslc board',  
 'ssc maharashtra board', 'tamil nadu state', 'uttrakhand boar  
 d',  
 'bihar secondary education board,patna',  
 'haryana board of school education',  
 'sri kannika parameswari highier secondary school, udumalpe  
 t',  
 'ksseb(karnataka state board)', 'nashik board',  
 'jharkhand secondary education board', 'himachal pradesh boar  
 d',  
 'maharashtra satate board',  
 'maharashtra state board mumbai divisional board',  
 'dav public school,hehal',  
 'state board of secondary education, ap',  
 'rajasthan board of secondary education', 'hsce',  
 'karnataka secondary education',  
 'board of secondary education,odisha', 'maharashtra nasik boa

```

rd',
    'west bengal board of secondary examination (wbbse)',
    'holy cross matriculation hr sec school', 'cbse', 'apssc',
    'bseb patna', 'kolhapur', 'bseb, patna', 'up board allahaba
d',
    'biharboard', 'nagpur board,nagpur', 'pune', 'gyan bharati sc
hool',
    'rbse,ajmer', 'board of secondaray education',
    'secondary school education', 'state bord', 'jbse,jharkhand',
    'hse', 'madhya pradesh board', 'bihar school examination boar
d',
    'west bengal board of secondary eucation', 'state boardmp bo
rd ',
    'icse board , new delhi',
    'board of secondary education (bse) orissa',
    'maharashtra state board for ssc',
    'board of secondary school education', 'latur board',
    'stmary's convent inter college', 'nagpur divisional board',
    'ap state board', 'cgbse raipur', 'uttranchal board', 'ksbe',
    'central board of secondary education, new delhi',
    'bihar school examination board patna', 'cbse board',
    'sslc,karnataka', 'mp-bse', 'up bourd', 'dav public school se
c 14',
    'board of school education haryana',
    'council for indian school certificate examination',
    'aurangabad board', 'j&k state board of school education',
    'maharashtra state board of secondary and higher secondary ed
ucation',
    'maharashtra state boar of secondary and higher secondary edu
cation',
    'ssc regular', 'karnataka state examination board', 'nasik',
    'west bengal board of secondary education', 'up board,allaha
bad',
    'bseb ,patna',
    'state board - west bengal board of secondary education : wbb
se',
    'maharashtra state board of secondary & higher secondary educ
ation',
    'delhi public school', 'karnataka secondary eduction',
    'secondary education board of rajasthan',
    'maharashtra board, pune', 'rbse (state board)', 'apsche',
    'board of secondary education',
    'board of high school and intermediate education uttarprades
h',
    'kea', 'board of secondary education - andhra pradesh',
    'ap state board for secondary education', 'seba',
    'punjab school education board, mohali',
    'jharkhand acedemic council', 'hse,board',
    'board of ssc education andhra pradesh', 'up-board', 'bse,odi
sha'],
dtype=object)

```

#### 1.3.1.4 Year of 12th graduation attribute

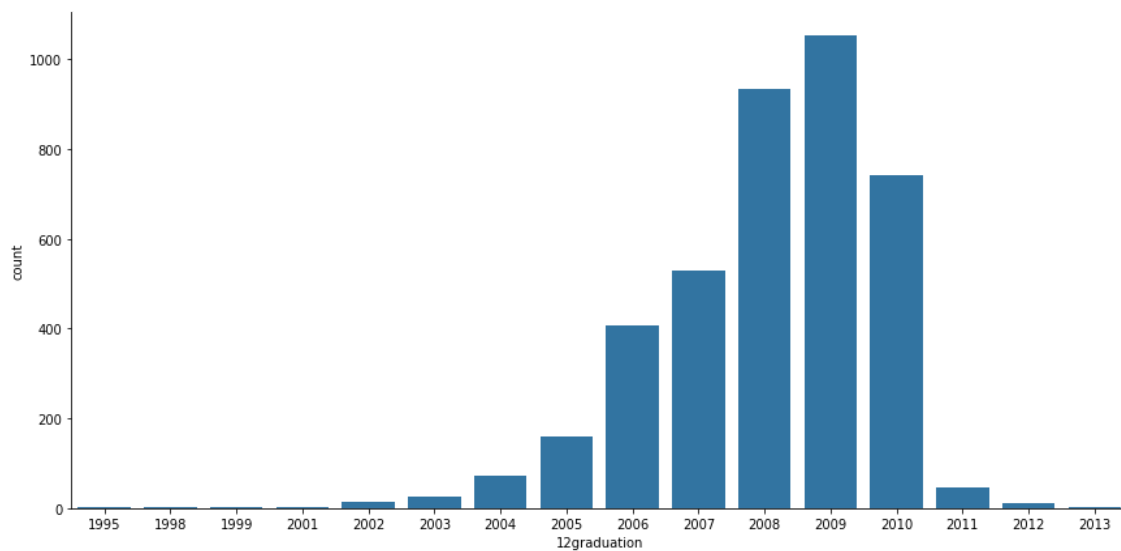


In [49]:

```
sns.FacetGrid(data=train,height=6,aspect=2).map(sns.countplot,"12graduation").add_legend()  
plt.ylabel("count")
```

Out[49]:

Text(7.197453703703708, 0.5, 'count')



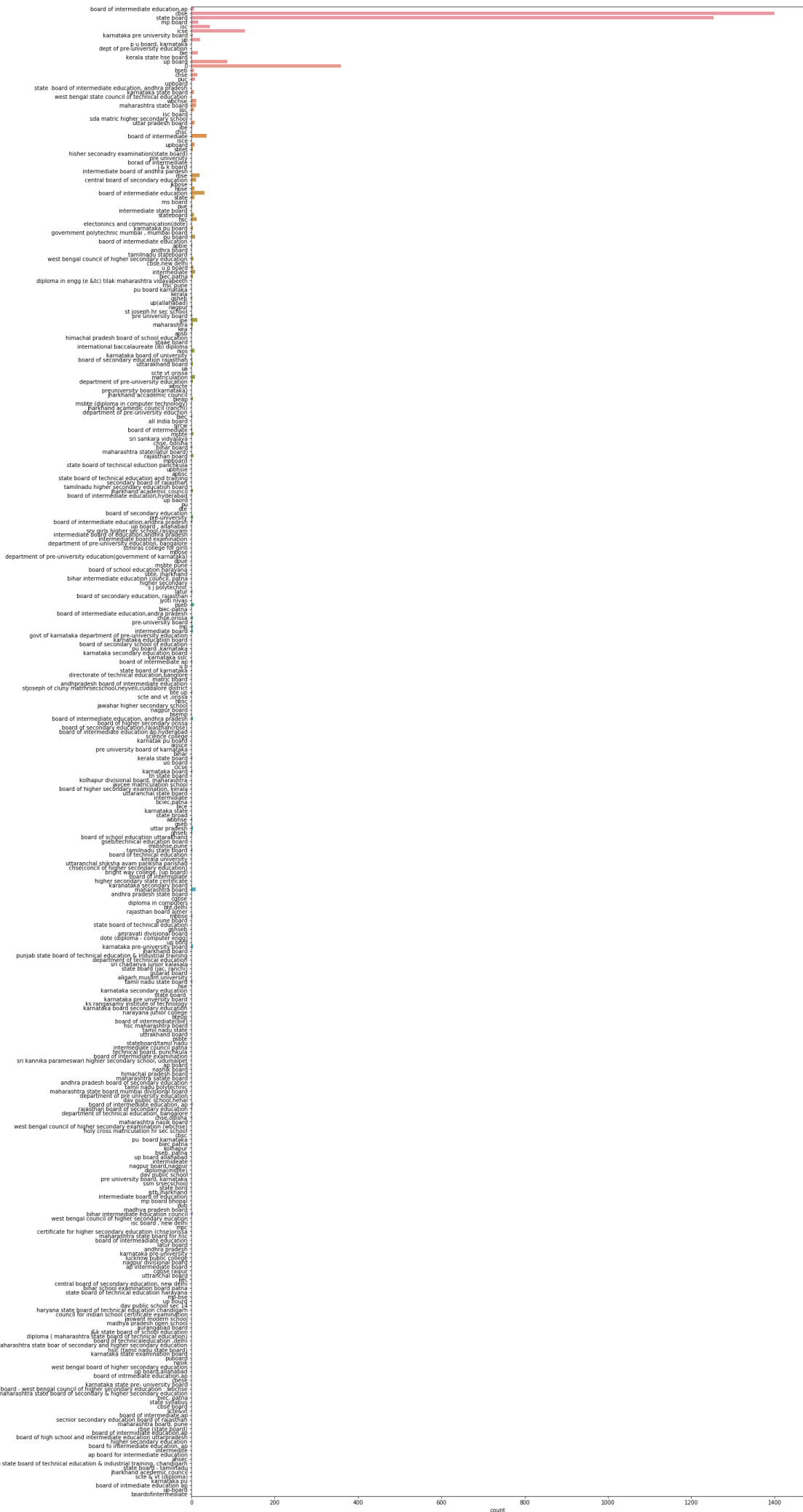
#### 1.3.1.5 '12'th board attribute

In [84]:

```
plt.figure(figsize=(20,50))  
sns.countplot(y=train['12board'])
```

Out[84]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f22962f6f60>





In [87]:

```
print(train['12board'].unique())
```

['board of intermediate education,ap' 'cbse' 'state board' 'mp board'  
 'isc' 'icse' 'karnataka pre university board' 'up' 'p u board, karnataka'  
 'dept of pre-university education' 'bie' 'kerala state hse board'  
 'up board' 0 'bseb' 'chse' 'puc' 'upboard'  
 'state board of intermediate education, andhra pradesh'  
 'karnataka state board'  
 'west bengal state council of technical education' 'wbchse'  
 'maharashtra state board' 'ssc' 'isc board'  
 'sda matric higher secondary school' 'uttar pradesh board' 'ibe' 'c  
 hsc'  
 'board of intermediate' 'isce' 'upboard' 'sbtet'  
 'higher secondary examination(state board)' 'pre university'  
 'board of intermediate' 'j & k board'  
 'intermediate board of andhra pradesh' 'rbse'  
 'central board of secondary education' 'jkbose' 'hbse'  
 'board of intermediate education' 'state' 'ms board' 'pue'  
 'intermediate state board' 'stateboard' 'hsc'  
 'electronics and communication(dote)' 'karnataka pu board'  
 'government polytechnic mumbai , mumbai board' 'pu board'  
 'board of intermediate education' 'apbie' 'andhra board'  
 'tamilnadu stateboard'  
 'west bengal council of higher secondary education' 'cbse,new delhi'  
 'u p board' 'intermediate' 'biec,patna'  
 'diploma in engg (e &tc) tilak maharashtra vidyapeeth' 'hsc pune'  
 'pu board karnataka' 'kerala' 'gsheb' 'up(allahabad)' 'nagpur'  
 'st joseph hr sec school' 'pre university board' 'ipe' 'maharashtra'  
 'kea' 'apsb' 'himachal pradesh board of school education' 'staate board'  
 'international baccalaureate (ib) diploma' 'nios'  
 'karnataka board of university' 'board of secondary education rajasthan'  
 'uttarakhand board' 'ua' 'scte vt orissa' 'matriculation'  
 'department of pre-university education' 'wbscte'  
 'preuniversity board(karnataka)' 'jharkhand academic council' 'bie  
 ap'  
 'msbte (diploma in computer technology)'  
 'jharkhand academic council (ranchi)'  
 'department of pre-university education' 'biec' 'all india board' 's  
 jrcw'  
 'board of intermediate' 'msbte' 'sri sankara vidyalaya' 'chse, odisha'  
 'bihar board' 'maharashtra state(latur board)' 'rajasthan board'  
 'mpboard' 'state board of technical education panchkula' 'upbhsie'  
 'apbsc'  
 'state board of technical education and training'  
 'secondary board of rajasthan'  
 'tamilnadu higher secondary education board' 'jharkhand academic council'  
 'board of intermediate education,hyderabad' 'up board' 'pu' 'dte'  
 'board of secondary education' 'pre-university'  
 'board of intermediate education, andhra pradesh' 'up board , allahabad'  
 'srvgirls higher sec school,rasipuram'  
 'intermediate board of education, andhra pradesh'  
 'intermediate board examination'  
 'department of pre-university education, bangalore'  
 'stmiras college for girls' 'mbose'

'department of pre-university education(government of karnataka)'  
 'dpue'  
 'msbte pune' 'board of school education harayana' 'sbte, jharkhand'  
 'bihar intermediate education council, patna' 'higher secondary'  
 's j polytechnic' 'latur' 'board of secondary education, rajasthan'  
 'jyoti nivas' 'pseb' 'biec-patna'  
 'board of intermediate education,andra pradesh' 'chse,orissa'  
 'pre-university board' 'mp' 'intermediate board'  
 'govt of karnataka department of pre-university education'  
 'karnataka education board' 'board of secondary school of educatio  
 n'  
 'pu board ,karnataka' 'karnataka secondary education board'  
 'karnataka sslc' 'board of intermediate ap' 'u p'  
 'state board of karnataka' 'directorate of technical education,bang  
 lore'  
 'matric board' 'andhpradesh board of intermediate education'  
 'stjoseph of cluny matrhrsecschool,neyveli,cuddalore district' 'bte  
 up'  
 'scte and vt ,orissa' 'hbse' 'jawahar higher secondary school'  
 'nagpur board' 'bsemp' 'board of intermediate education, andhra pra  
 desh'  
 'board of higher secondary orissa'  
 'board of secondary education,rajasthan(rbse)'  
 'board of intermediate education:ap,hyderabad' 'science college'  
 'karnatak pu board' 'aissce' 'pre university board of karnataka' 'b  
 ihar'  
 'kerala state board' 'uo board' 'cicse' 'karnataka board'  
 'tn state board' 'kolhapur divisional board, maharashtra'  
 'jaycee matriculation school'  
 'board of higher secondary examination, kerala' 'uttaranchal state  
 board'  
 'intermidiate' 'bciec,patna' 'bice' 'karnataka state' 'state broad'  
 'wbbhse' 'gseb' 'uttar pradesh' 'ghseb'  
 'board of school education uttarakhnad' 'gseb/technical education b  
 oard'  
 'msbshse,pune' 'tamilnadu state board' 'board of technical educatio  
 n'  
 'kerala university' 'uttaranchal shiksha avam pariksha parishad'  
 'chse(concil of higher secondary education)'  
 'bright way college, (up board)' 'board of intermidiate'  
 'higher secondary state certificate' 'karanataka secondary board'  
 'maharashtra board' 'andhra pradesh state board' 'cgbse'  
 'diploma in computers' 'bte,delhi' 'rajasthan board ajmer' 'mpbse'  
 'pune board' 'state board of technical education' 'gshseb'  
 'amravati divisional board' 'dote (diploma - computer engg)' 'up bo  
 rd'  
 'karnataka pre-university board' 'jharkhand board'  
 'punjab state board of technical education & industrial training'  
 'department of technical education' 'sri chaitanya junior kalasala'  
 'state board (jac, ranchi)' 'gujarat board' 'aligarh muslim univers  
 ity'  
 'tamil nadu state board' 'hse' 'karnataka secondary education'  
 'state board ' 'karnataka pre unversity board'  
 'ks rangasamy institute of technology'  
 'karnataka board secondary education' 'narayana junior college' 'bt  
 eup'  
 'board of intermediate(bie)' 'hsc maharashtra board' 'tamil nadu st  
 ate'  
 'uttrakhand board' 'psbte' 'stateboard/tamil nadu'  
 'intermediate council patna' 'technical board, punchkula'  
 'board of intermidiate examination'



'sri kannika parameswari highier secondary school, udumalpet' 'ap b  
oard'  
'nashik board' 'himachal pradesh board' 'maharashtra satate board'  
'andhra pradesh board of secondary education' 'tamil nadu polytechn  
ic'  
'maharashtra state board mumbai divisional board'  
'department of pre university education' 'dav public school,hehal'  
'board of intermediate education, ap'  
'rajasthan board of secondary education'  
'department of technical education, bangalore' 'chse,odisha'  
'maharashtra nasik board'  
'west bengal council of higher secondary examination (wbchse)'  
'holy cross matriculation hr sec school' 'cbse' 'pu board karnatak  
a'  
'biec patna' 'kolhapur' 'bseb, patna' 'up board allahabad' 'intermi  
deate'  
'nagpur board,nagpur' 'diploma(msbte)' 'dav public school'  
'pre university board, karnataka' 'ssm srsecschool' 'state bord'  
'jstb,jharkhand' 'intermediate board of education' 'mp board bhopa  
l'  
'pub' 'madhya pradesh board' 'bihar intermediate education council'  
'west bengal council of higher secondary eucation'  
'isc board , new delhi' 'mpc'  
'certificate for higher secondary education (chse)orissa'  
'maharashtra state board for hsc' 'board of intermeadiate educatio  
n'  
'latur board' 'andhra pradesh' 'karnataka pre-university'  
'lucknow public college' 'nagpur divisional board'  
'ap intermediate board' 'cgbse raipur' 'uttranchal board' 'jiec'  
'central board of secondary education, new delhi'  
'bihar school examination board patna'  
'state board of technical education harayana' 'mp-bse' 'up bourd'  
'dav public school sec 14'  
'haryana state board of technical education chandigarh'  
'council for indian school certificate examination'  
'jaswant modern school' 'madhya pradesh open school' 'aurangabad bo  
ard'  
'j&k state board of school education'  
'diploma ( maharashtra state board of technical education)'  
'board of technicaleducation ,delhi'  
'maharashtra state boar of secondary and higher secondary educatio  
n'  
'hslc (tamil nadu state board)' 'karnataka state examination board'  
'puboard' 'nasik' 'west bengal board of higher secondary education'  
'up board,allahabad' 'board of intrmediate education,ap' 'cbese'  
'karnataka state pre- university board'  
'state board - west bengal council of higher secondary education :  
wbchse'  
'maharashtra state board of secondary & higher secondary education'  
'biec, patna' 'state syllabus' 'cbse board' 'scte&vt'  
'board of intermediate,ap'  
'secnior secondary education board of rajasthan'  
'maharashtra board, pune' 'rbse (state board)'  
'board of intermidiate education,ap'  
'board of high school and intermediate education uttarpradesh'  
'higher secondary education' 'board fo intermediate education, ap'  
'intermedite' 'ap board for intermediate education' 'ahsec'  
'punjab state board of technical education & industrial training, c  
handigarh'  
'state board - tamilnadu' 'jharkhand acedemic council'  
'scte & vt (diploma)' 'karnataka pu' 'board of intmediate education

```
ap'  
'up-board' 'boardofintermediate']
```

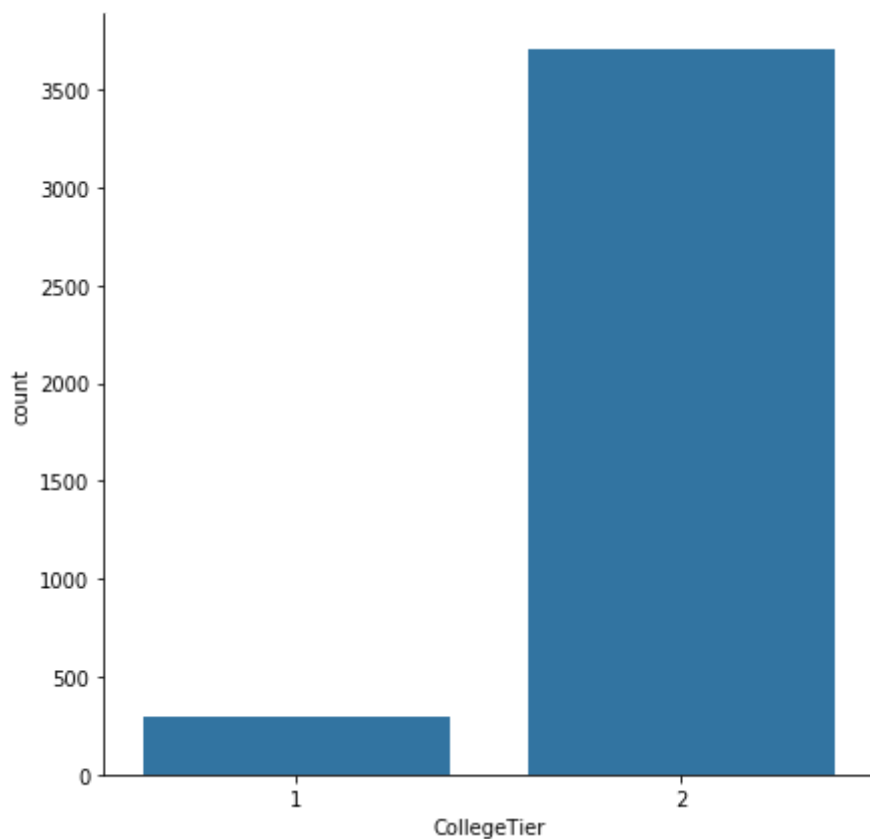
#### 1.3.1.6 College tier

In [54]:

```
sns.FacetGrid(data=train,height=6).map(sns.countplot,"CollegeTier").add_legend()  
plt.ylabel("count")
```

Out[54]:

```
Text(7.597222222222285, 0.5, 'count')
```



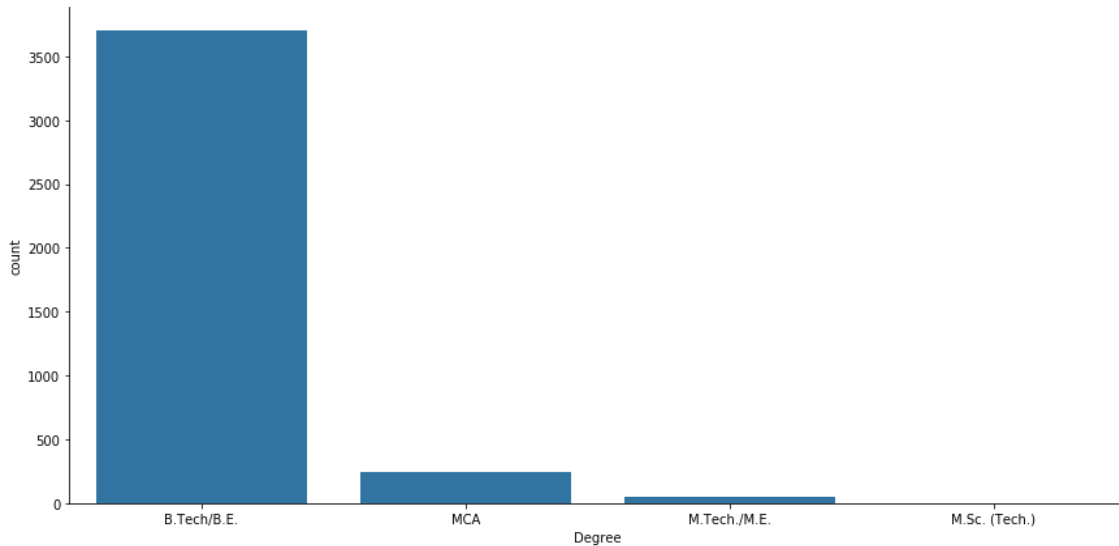
#### 1.3.1.7 Degree attribute

In [57]:

```
sns.FacetGrid(data=train,height=6,aspect=2).map(sns.countplot,"Degree").add_legend()  
plt.ylabel("count")
```

Out[57]:

Text(7.198611111111113, 0.5, 'count')



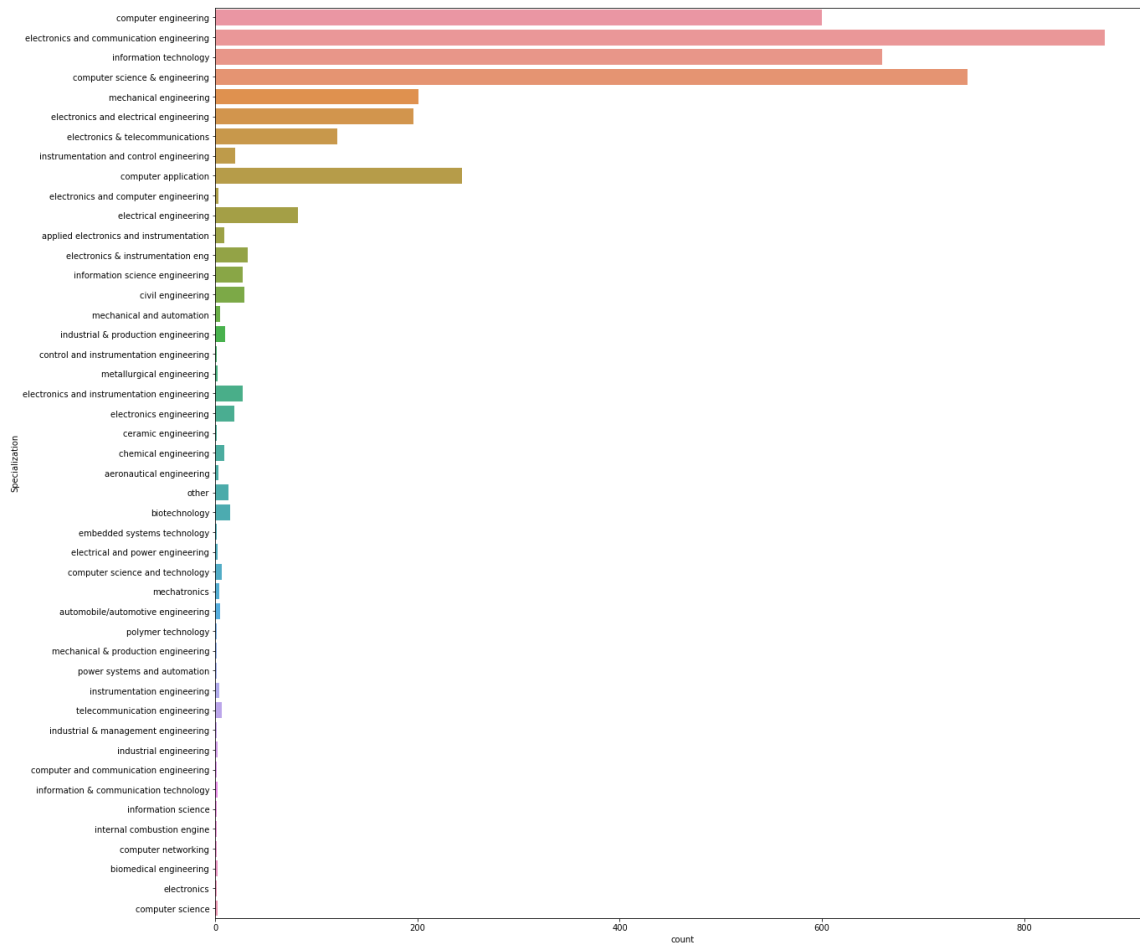
### 1.3.1.8 Specalization attribute

In [81]:

```
plt.figure(figsize=(20,20))  
sns.countplot(y=train['Specialization'])
```

Out[81]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f229611a9e8>



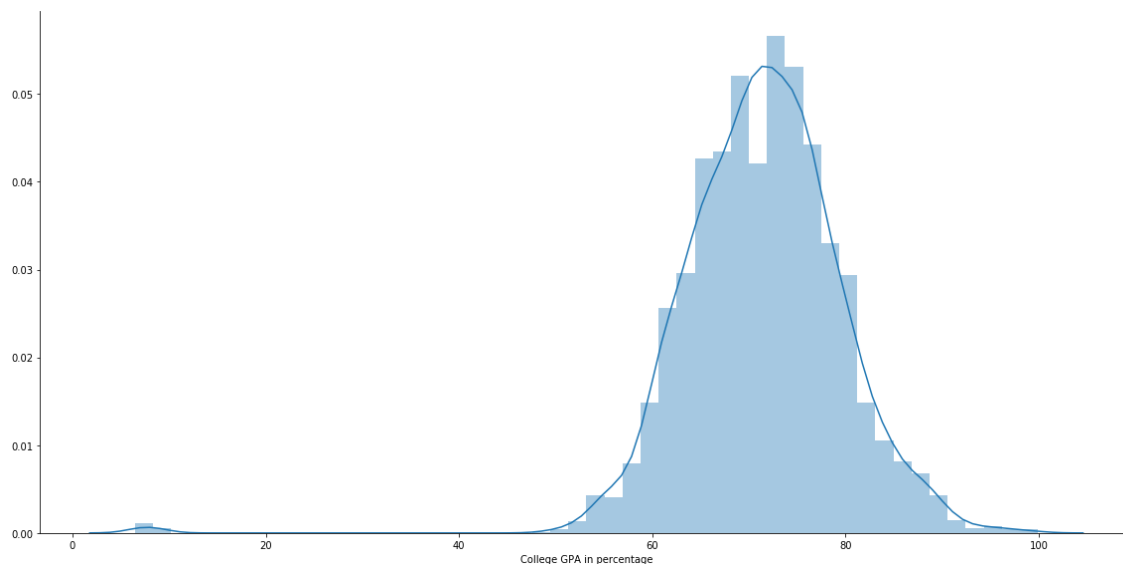
### 1.3.1.9 Graduation GPA (in percentage) attribute

In [68]:

```
sns.FacetGrid(data=train,height=8,aspect=2).map(sns.distplot,"collegeGPA").add_  
legend()  
plt.xlabel("College GPA in percentage")
```

Out[68]:

Text(0.5, 20.800000000000001, 'College GPA in percentage')



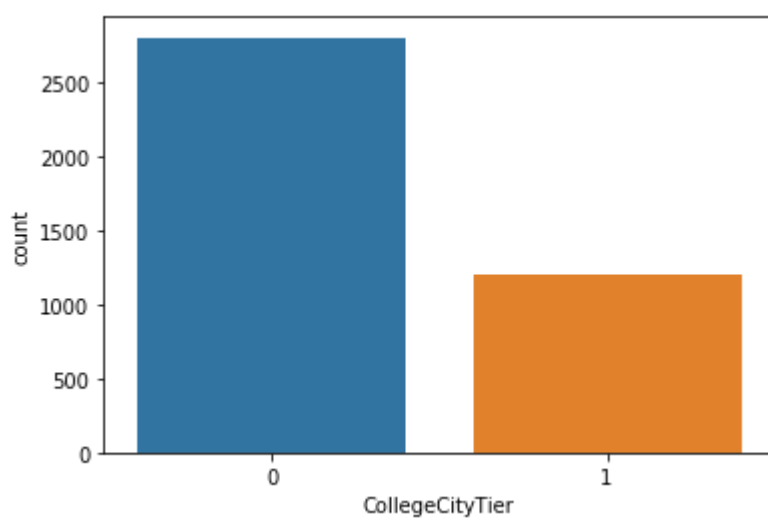
### 1.3.1.10 College City Tier attribute

In [74]:

```
sns.countplot(train['CollegeCityTier'])
```

Out[74]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f2296807a90>



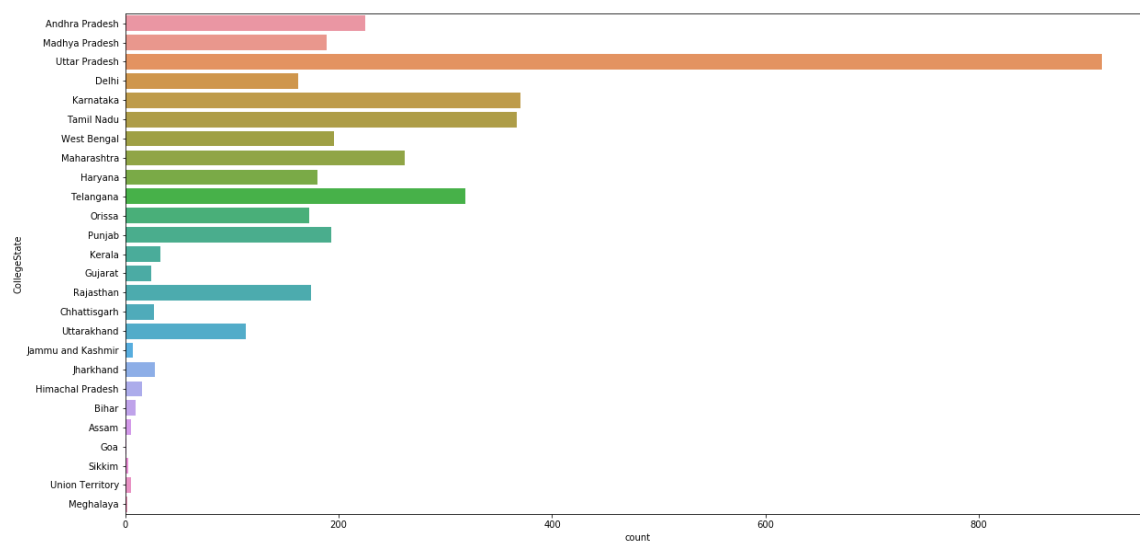
### 1.3.1.11 College state

In [89]:

```
plt.figure(figsize=(20,10))
sns.countplot(y=train['CollegeState'])
```

Out[89]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f22931436d8>



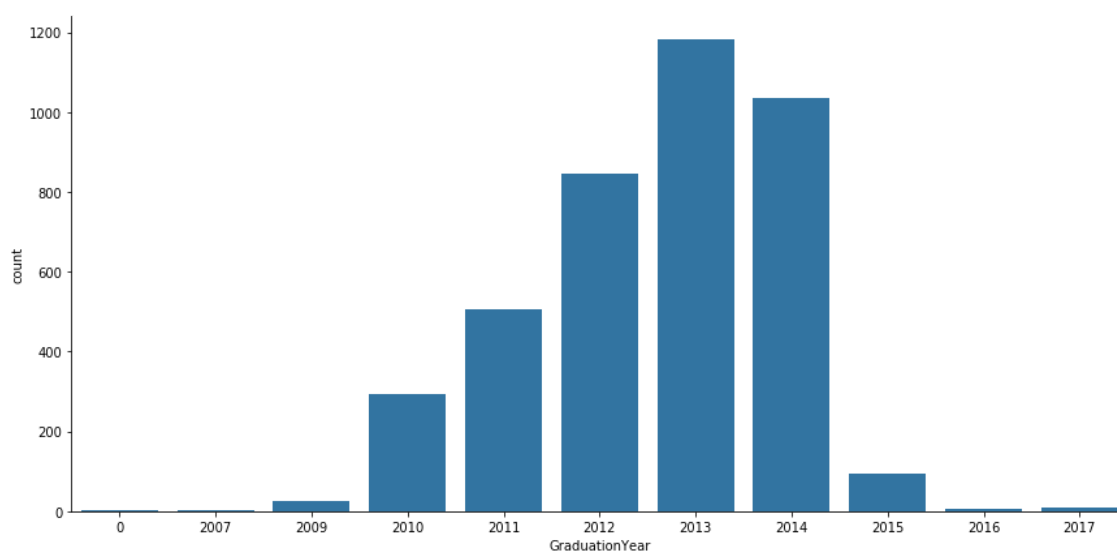
### 1.3.1.12 Year of bachelors degree graduation

In [90]:

```
sns.FacetGrid(data=train,height=6,aspect=2).map(sns.countplot,"GraduationYear").
add_legend()
plt.ylabel("count")
```

Out[90]:

Text(7.198611111111113, 0.5, 'count')



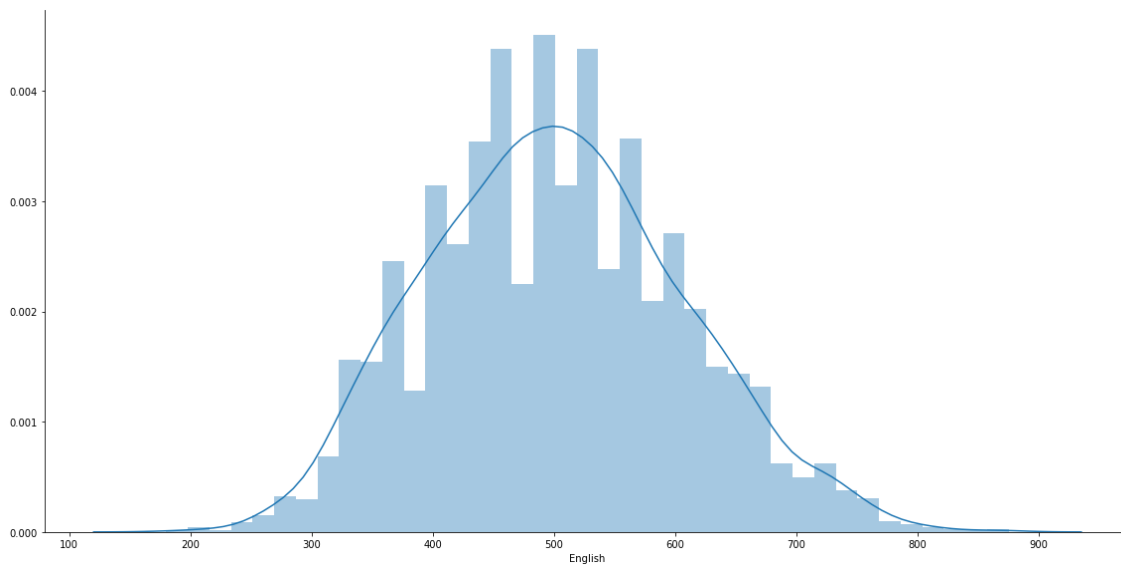
### 1.3.1.13 Amcat English section score attribute

In [91]:

```
sns.FacetGrid(data=train,height=8,aspect=2).map(sns.distplot,"English").add_legend()
```

Out[91]:

<seaborn.axisgrid.FacetGrid at 0x7f2292e9ff98>



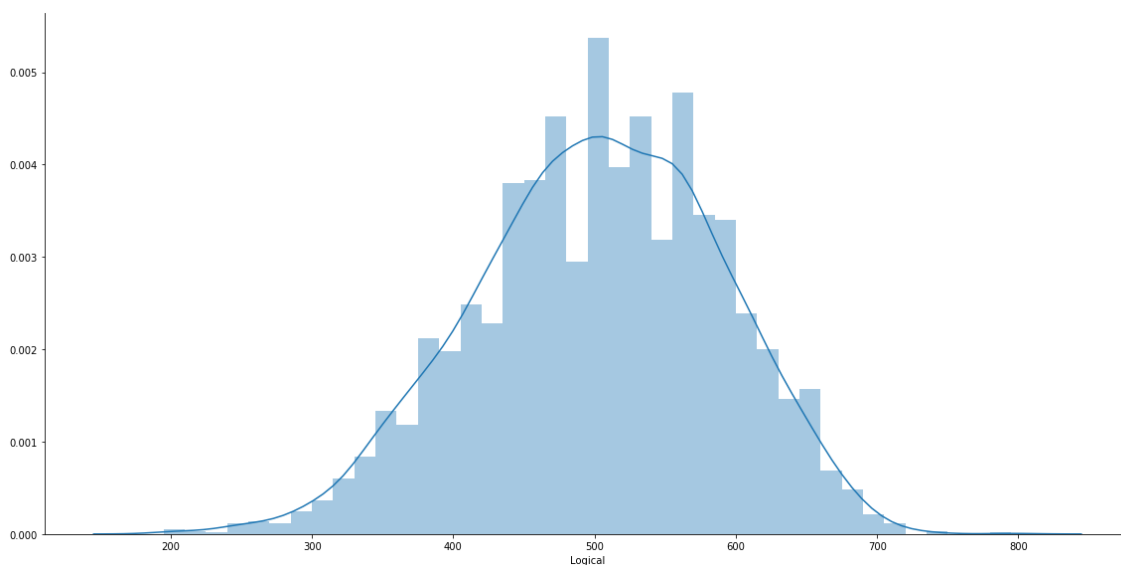
#### 1.3.1.14 Amcat Logical ability section score attribute

In [92]:

```
sns.FacetGrid(data=train,height=8,aspect=2).map(sns.distplot,"Logical").add_legend()
```

Out[92]:

<seaborn.axisgrid.FacetGrid at 0x7f2292d9f748>



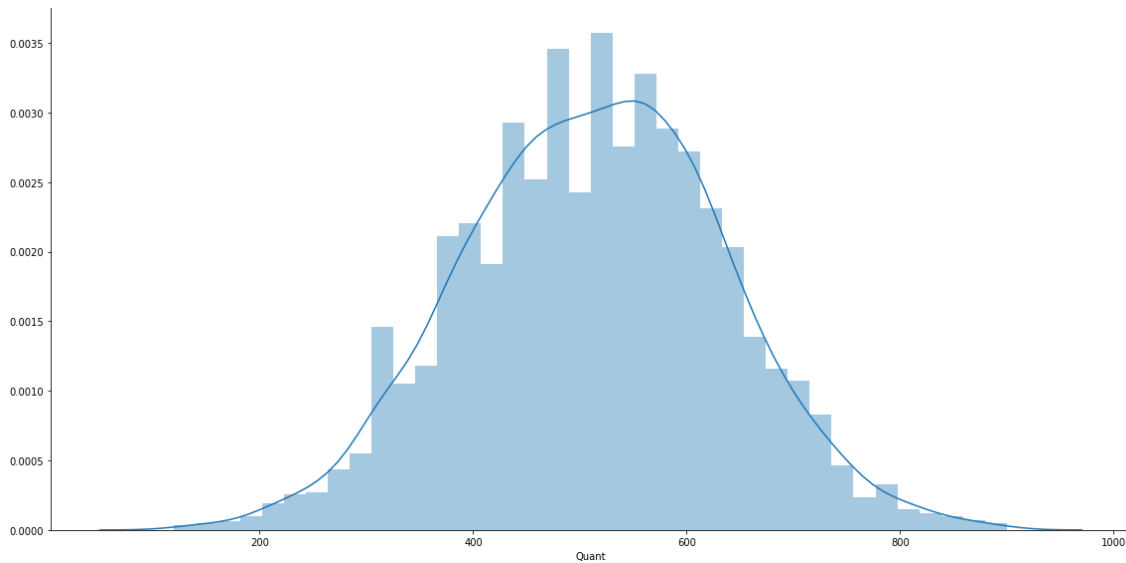
#### 1.3.1.15 Amcat Quants section score attribute

In [94]:

```
sns.FacetGrid(data=train,height=8,aspect=2).map(sns.distplot,"Quant").add_legend()
```

Out[94]:

<seaborn.axisgrid.FacetGrid at 0x7f2292c62048>



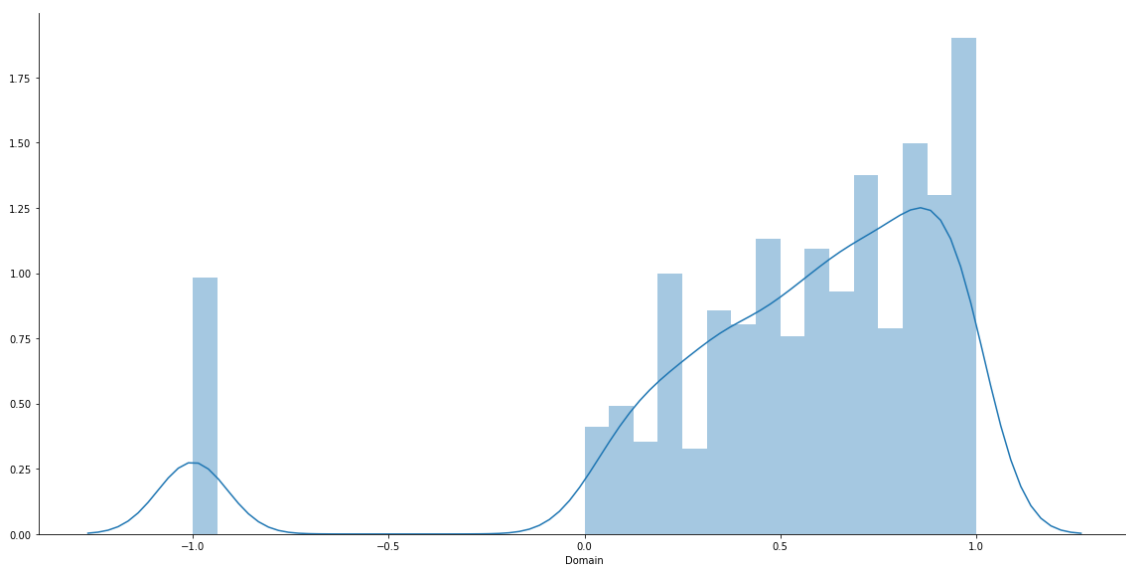
### 1.3.1.16 Amcat domain score

In [100]:

```
sns.FacetGrid(data=train,height=8,aspect=2).map(sns.distplot,"Domain").add_legend()
```

Out[100]:

<seaborn.axisgrid.FacetGrid at 0x7f228f6c7908>



### 1.3.1.17 Personality test conscientiousness score attribute

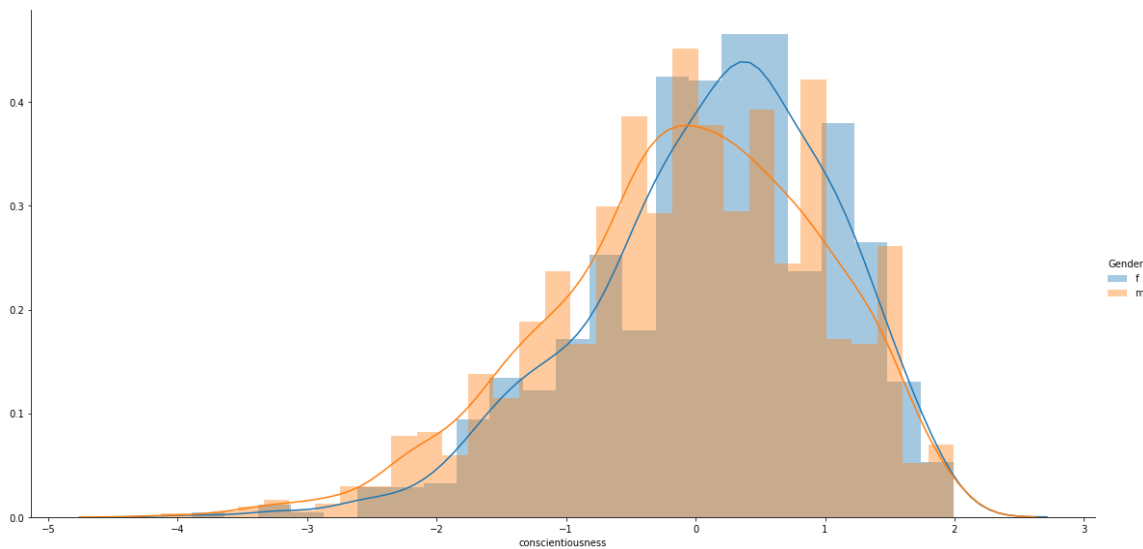


In [112]:

```
sns.FacetGrid(data=train,hue="Gender",height=8,aspect=2).map(sns.distplot,"conscientiousness").add_legend()
```

Out[112]:

<seaborn.axisgrid.FacetGrid at 0x7f22930966a0>



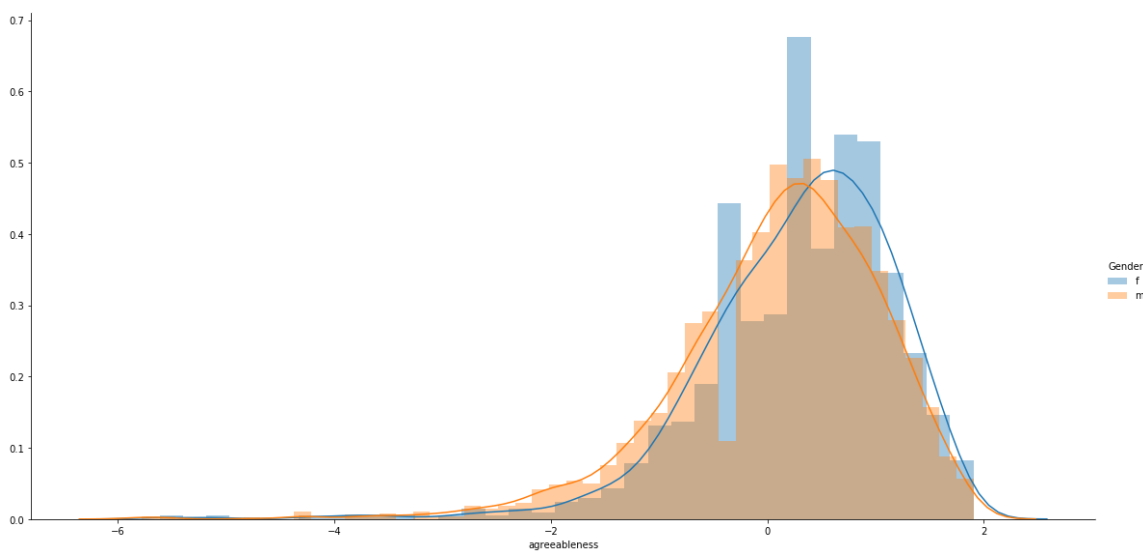
#### 1.3.1.18 Personality test agreeableness score attribute

In [111]:

```
sns.FacetGrid(data=train,hue="Gender",height=8,aspect=2).map(sns.distplot,"agreeableness").add_legend()
```

Out[111]:

<seaborn.axisgrid.FacetGrid at 0x7f2291a14a20>



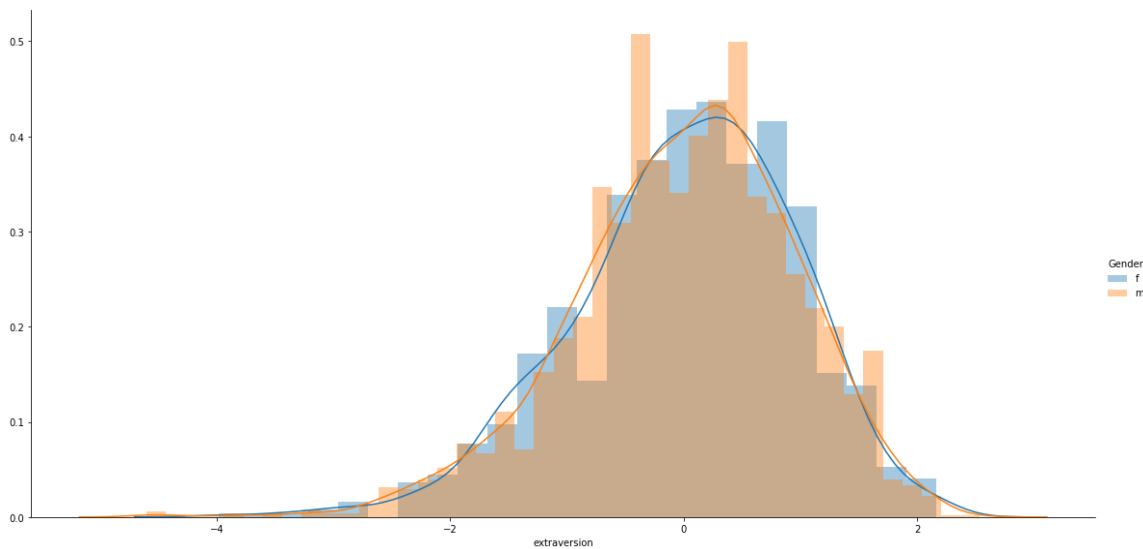
#### 1.3.1.19 Personality test extraversion score attribute

In [110]:

```
sns.FacetGrid(data=train,hue="Gender",height=8,aspect=2).map(sns.distplot,"extra  
version").add_legend()
```

Out[110]:

<seaborn.axisgrid.FacetGrid at 0x7f22924e5be0>



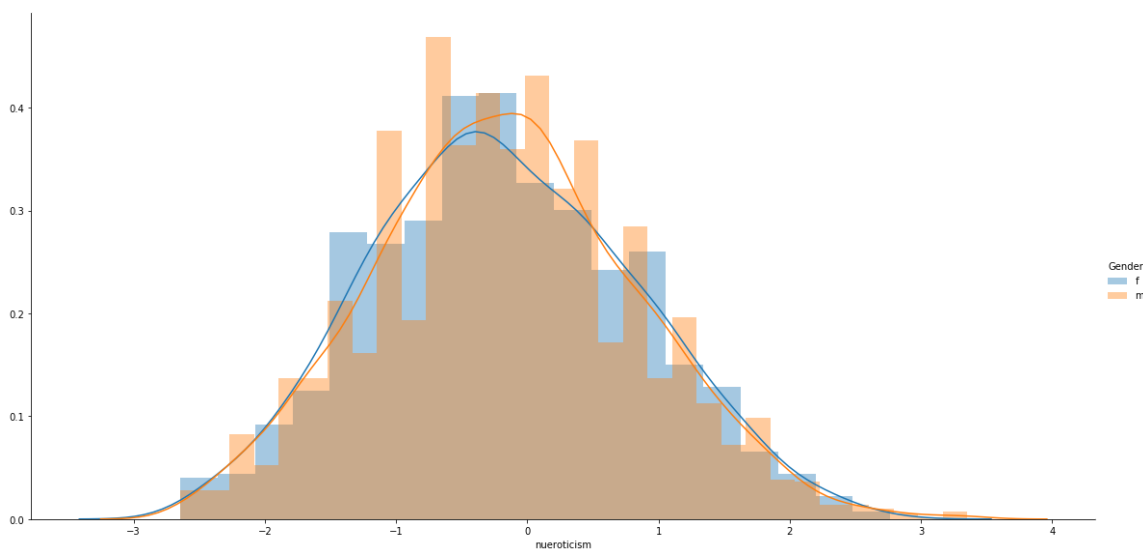
### 1.3.1.20 Personality test nueroticism score attribute

In [108]:

```
sns.FacetGrid(data=train,hue="Gender",height=8,aspect=2).map(sns.distplot,"nuero  
ticism").add_legend()
```

Out[108]:

<seaborn.axisgrid.FacetGrid at 0x7f22919e3b38>



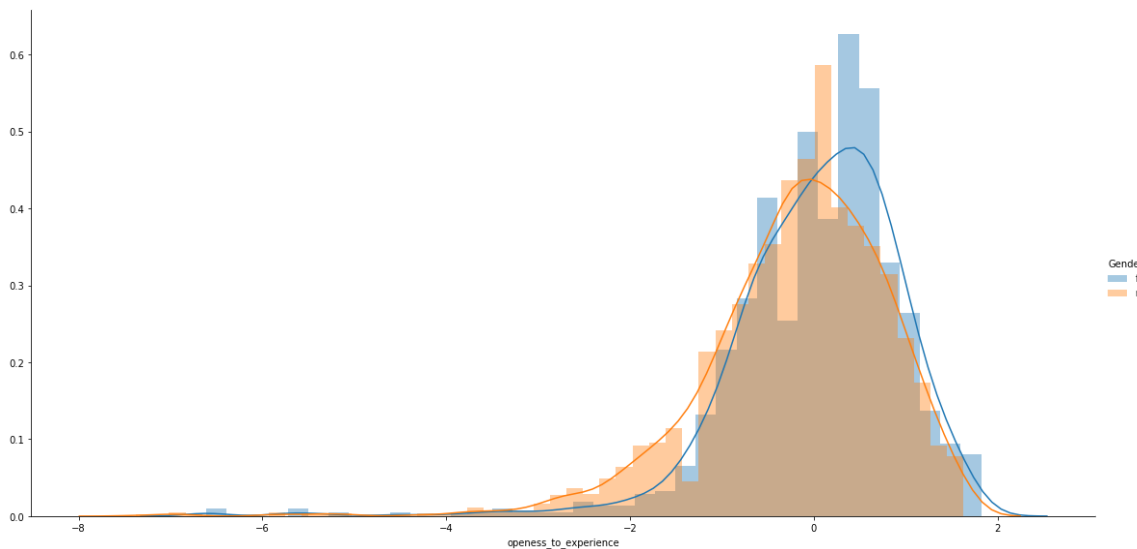
### 1.3.1.21 Personality test openness\_to\_experience score attribute

In [113]:

```
sns.FacetGrid(data=train,hue="Gender",height=8,aspect=2).map(sns.distplot,"openess_to_experience").add_legend()
```

Out[113]:

<seaborn.axisgrid.FacetGrid at 0x7f22970d0dd8>



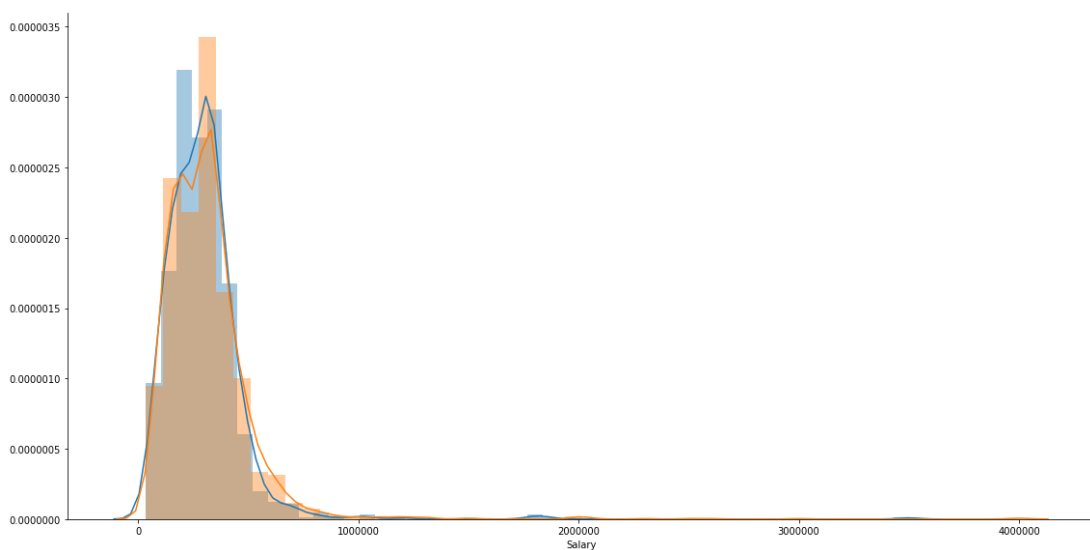
### 1.3.1.22 Analyzing the target (Salary) Variable by gender

In [115]:

```
sns.FacetGrid(data=train,hue="Gender",height=8,aspect=2).map(sns.distplot,"Salary").add_legend()
```

Out[115]:

<seaborn.axisgrid.FacetGrid at 0x7f2290756f98>



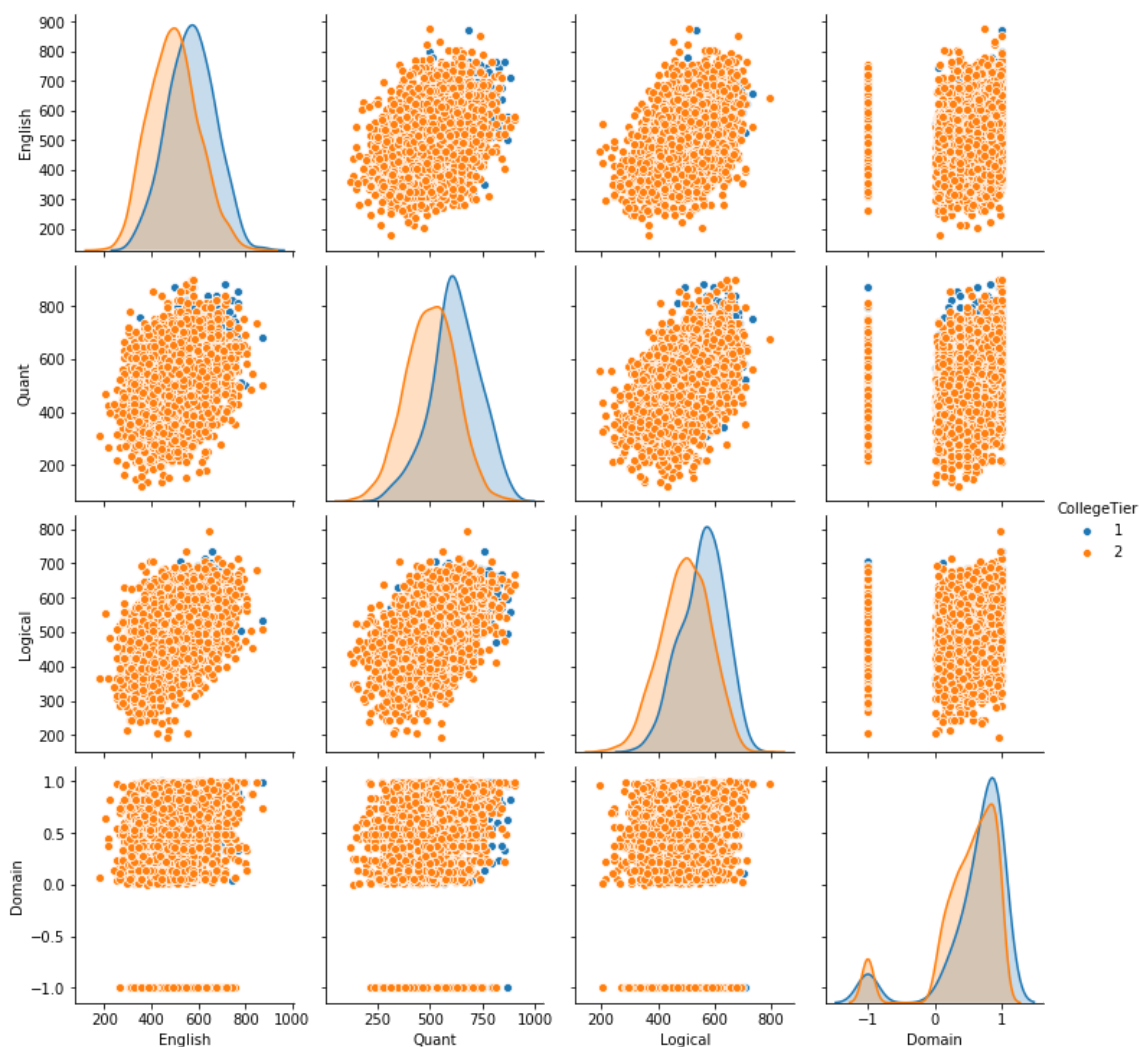
### 1.3.1.23 Pair plot of amcat scores by College Tier

In [130]:

```
amcatScores = train[['English', 'Quant', 'Logical', 'Domain', 'CollegeTier']]
sns.pairplot(data=amcatScores, hue="CollegeTier")
```

Out[130]:

<seaborn.axisgrid.PairGrid at 0x7f228c1f0940>



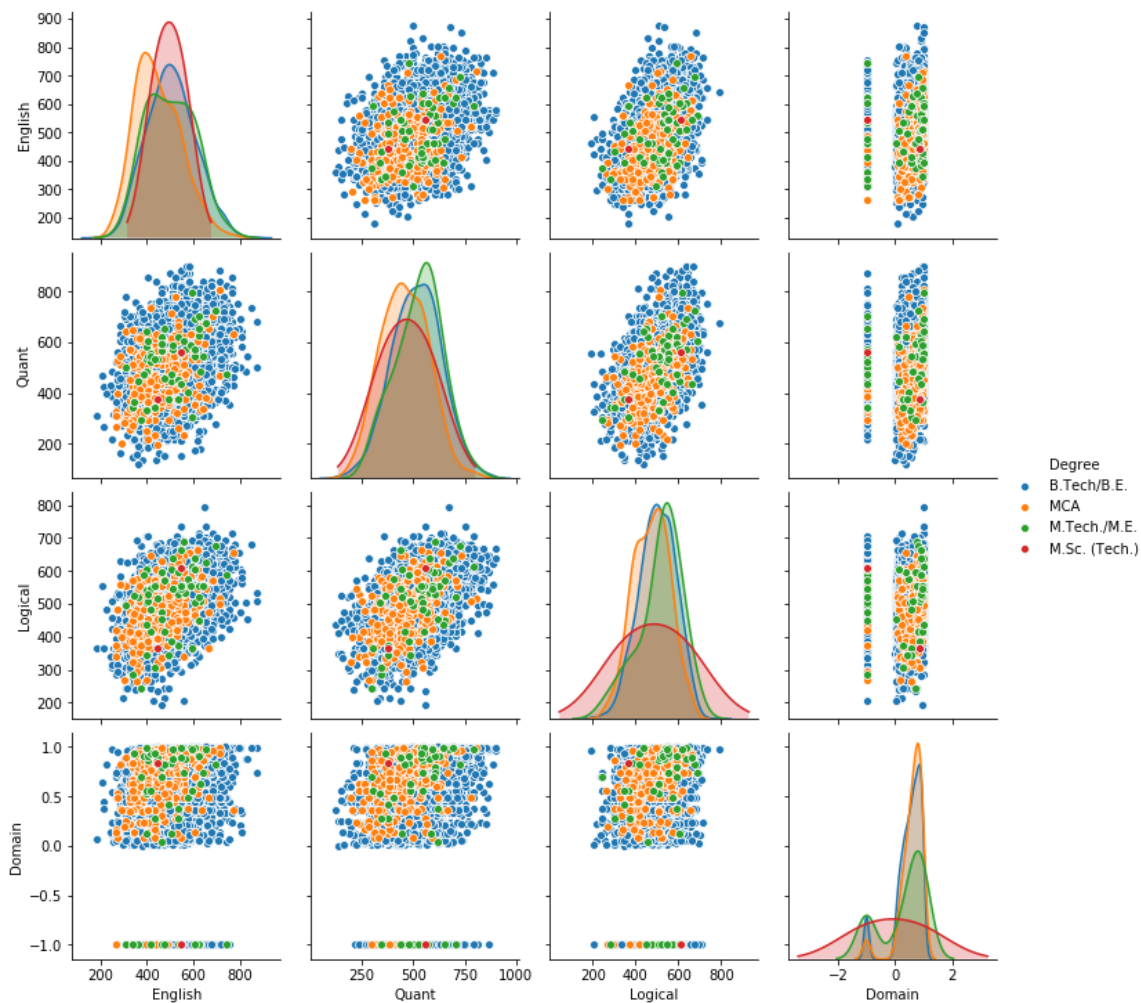
### 1.3.1.24 Pair plot of amcat scores by Degree

In [132]:

```
amcatScores1 = train[['English', 'Quant', 'Logical', 'Domain', 'Degree']]
sns.pairplot(data=amcatScores1, hue="Degree")
```

Out[132]:

<seaborn.axisgrid.PairGrid at 0x7f2288ba9ba8>



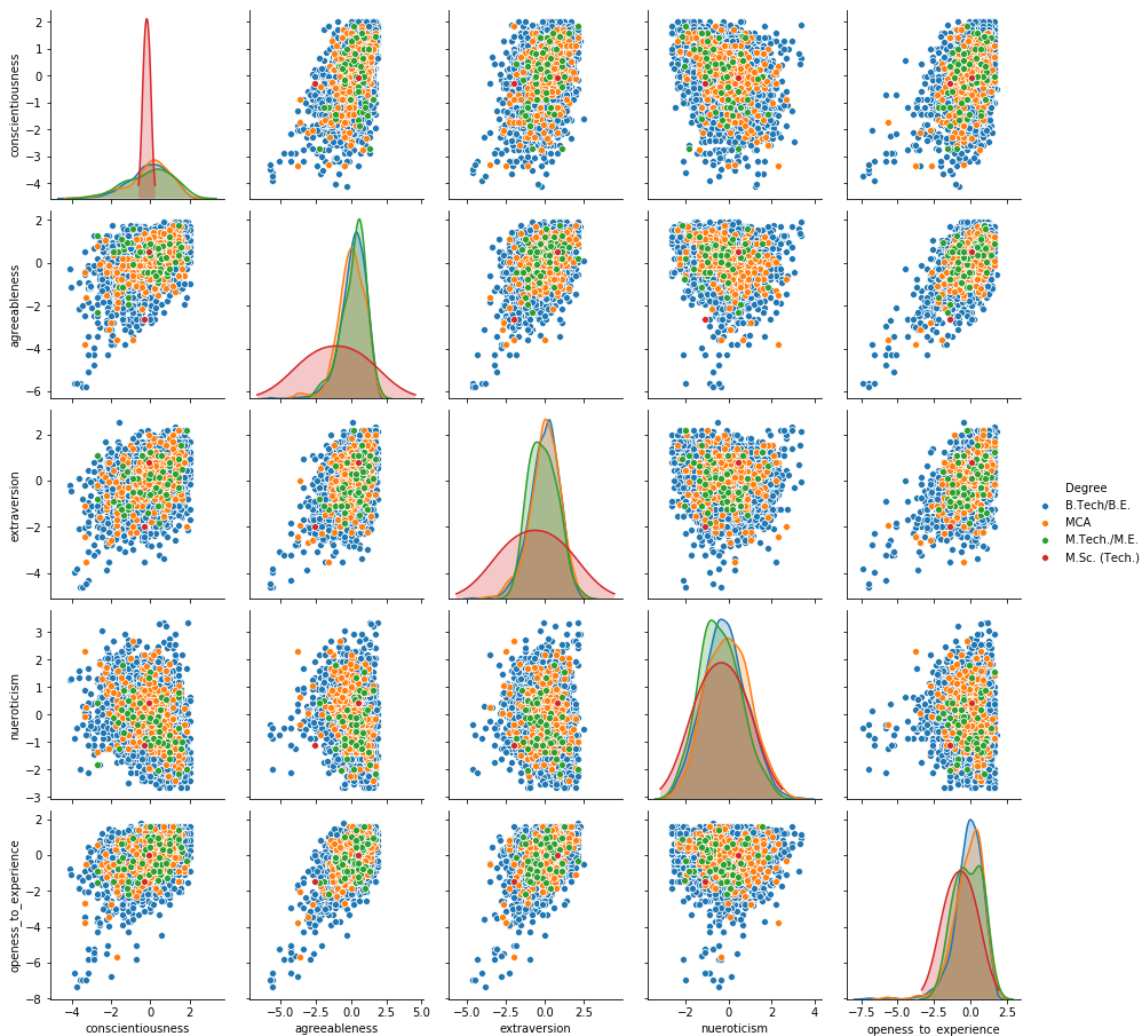
### 1.3.1.25 Pair plot of personality scores by Degree

In [133]:

```
personalityScores = train[['conscientiousness', 'agreeableness', 'extraversion', 'neuroticism', 'openness_to_experience', 'Degree']]
sns.pairplot(data=personalityScores, hue='Degree')
```

Out[133]:

<seaborn.axisgrid.PairGrid at 0x7f229002cef0>



## 1.4 Making necessary Assumptions and changes accordingly

- 1. As our Objective in this project is to Predict the salary given the details, We don't need the DOJ(Date of joining),DOL(Data of leaving),JobDesignation and JobCity attributes. As they are dependent variables we can simply drop them and use salary as our only dependent variable.
- 2. We don't need Collegeld and CollegeCityID attributes. And it is perfectly safe to drop them. Because we already have a unique ID for every candidate.
- 3. Instead of considering the specific score for each domain, We can simply consider the percentile of the domain score (Which is also a feature given).
- 4. We would encode the numerical attributes and categorical attributes and stack them to create our feature vector.

Dropping the unnecessary features

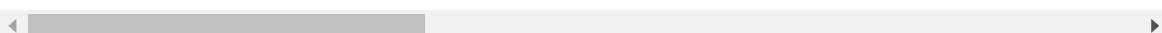
In [370]:

```
features = train.drop(['CollegeID','CollegeCityID','ComputerProgramming','ElectronicsAndSemicon','ComputerScience','ElectricalEngg','TelecomEngg','CivilEngg','MechanicalEngg','DOJ','DOL','Designation','JobCity','Unnamed: 0'],axis=1)
candidate_id = features['ID']
features = features.drop(['ID'],axis=1)
features.head()
```

Out[370]:

	Salary	Gender	DOB	10percentage	10board	12graduation	12percentage	12bo
0	420000	f	1990-02-19	84.3	board ofsecondary education,ap	2007	95.8	boar intermed education
1	500000	m	1989-10-04	85.4	cbse	2007	85.0	c
2	325000	f	1992-08-03	85.0	cbse	2010	68.2	c
3	1100000	m	1989-12-05	85.6	cbse	2007	83.6	c
4	200000	m	1991-02-27	78.0	cbse	2008	76.8	c

5 rows × 24 columns



Dividing features into dependent variables and independent variables

In [371]:

```
x = features.drop(['Salary'],axis=1)
y = features['Salary'].values
```

## 1.5 Dividing the data into Train,Test and Cross validation

In [372]:

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=0)
x_train, x_cv, y_train, y_cv = train_test_split(x_train, y_train, test_size=0.3, random_state=0)
print("Shape of x_train : ",x_train.shape)
print("Shape of x_cv : ",x_cv.shape)
print("Shape of x_test : ",x_test.shape)
```

```
Shape of x_train : (1958, 23)
Shape of x_cv : (840, 23)
Shape of x_test : (1200, 23)
```

## 1.6 Encoding the features

### 1.6.1 Encoding the Gender attribute

In [373]:

```
le = preprocessing.LabelEncoder()
le.fit(x_train['Gender'])
x_train_gender = le.transform(x_train['Gender']).reshape(1,-1).T
x_cv_gender = le.transform(x_cv['Gender']).reshape(1,-1).T
x_test_gender = le.transform(x_test['Gender']).reshape(1,-1).T
print(x_train_gender.shape)
print(x_cv_gender.shape)
print(x_test_gender.shape)
```

```
(1958, 1)
(840, 1)
(1200, 1)
```

### 1.6.2 Encoding the date of birth

Replacing the whole date of birth by only year

In [374]:

```
x_train_dob = x_train['DOB'].map(lambda x: x.year).values.reshape(1,-1).T
x_cv_dob = x_cv['DOB'].map(lambda x: x.year).values.reshape(1,-1).T
x_test_dob = x_test['DOB'].map(lambda x: x.year).values.reshape(1,-1).T
print(x_train_dob.shape)
print(x_cv_dob.shape)
print(x_test_dob.shape)
```

```
(1958, 1)
(840, 1)
(1200, 1)
```



### 1.6.3 Encoding the 10nth percentage

In [375]:

```
normalizer = preprocessing.Normalizer()
normalizer.fit(x_train['10percentage'].values.reshape(1,-1))
x_train_10percentage = normalizer.transform(x_train['10percentage'].values.reshape(1,-1)).T
x_cv_10percentage = normalizer.transform(x_cv['10percentage'].values.reshape(1,-1)).T
x_test_10percentage = normalizer.transform(x_test['10percentage'].values.reshape(1,-1)).T
print(x_train_10percentage.shape)
print(x_cv_10percentage.shape)
print(x_test_10percentage.shape)
```

```
(1958, 1)
(840, 1)
(1200, 1)
```

### 1.6.4 Encoding the 10nth board

In [376]:

```
x_train['10board'] = x_train['10board'].replace(0, 'no')
x_cv['10board'] = x_cv['10board'].replace(0, 'no')
x_test['10board'] = x_test['10board'].replace(0, 'no')
```

In [377]:

```
vectorizer = CountVectorizer()
vectorizer.fit(x_train['10board'])
x_train_10board = vectorizer.transform(x_train['10board'])
x_cv_10board = vectorizer.transform(x_cv['10board'])
x_test_10board = vectorizer.transform(x_test['10board'])
print(x_train_10board.shape)
print(x_cv_10board.shape)
print(x_test_10board.shape)
```

```
(1958, 137)
(840, 137)
(1200, 137)
```

### 1.6.5 Encoding the 12th percentage

In [378]:

```
normalizer = preprocessing.Normalizer()
normalizer.fit(x_train['12percentage'].values.reshape(1,-1))
x_train_12percentage = normalizer.transform(x_train['12percentage'].values.reshape(1,-1)).T
x_cv_12percentage = normalizer.transform(x_cv['12percentage'].values.reshape(1,-1)).T
x_test_12percentage = normalizer.transform(x_test['12percentage'].values.reshape(1,-1)).T
print(x_train_12percentage.shape)
print(x_cv_12percentage.shape)
print(x_test_12percentage.shape)
```

```
(1958, 1)
(840, 1)
(1200, 1)
```

## 1.6.6 Encoding the 12th board

In [379]:

```
x_train['12board'] = x_train['12board'].replace(0, 'no')
x_cv['12board'] = x_cv['12board'].replace(0, 'no')
x_test['12board'] = x_test['12board'].replace(0, 'no')
```

In [380]:

```
vectorizer = CountVectorizer()
vectorizer.fit(x_train['12board'])
x_train_12board = vectorizer.transform(x_train['12board'])
x_cv_12board = vectorizer.transform(x_cv['12board'])
x_test_12board = vectorizer.transform(x_test['12board'])
print(x_train_12board.shape)
print(x_cv_12board.shape)
print(x_test_12board.shape)
```

```
(1958, 183)
(840, 183)
(1200, 183)
```

## 1.6.7 Encoding the Degree

In [381]:

```
vectorizer = CountVectorizer()
vectorizer.fit(x_train['Degree'])
x_train_degree = vectorizer.transform(x_train['Degree'])
x_cv_degree = vectorizer.transform(x_cv['Degree'])
x_test_degree = vectorizer.transform(x_test['Degree'])
print(x_train_degree.shape)
print(x_cv_degree.shape)
print(x_test_degree.shape)
```

```
(1958, 3)
(840, 3)
(1200, 3)
```

## 1.6.8 Encoding the Specialization

In [382]:

```
vectorizer = CountVectorizer()
vectorizer.fit(x_train['Specialization'])
x_train_specialization = vectorizer.transform(x_train['Specialization'])
x_cv_specialization = vectorizer.transform(x_cv['Specialization'])
x_test_specialization = vectorizer.transform(x_test['Specialization'])
print(x_train_specialization.shape)
print(x_cv_specialization.shape)
print(x_test_specialization.shape)
```

```
(1958, 33)
(840, 33)
(1200, 33)
```

## 1.6.9 Encoding the college GPA

In [383]:

```
normalizer = preprocessing.Normalizer()
normalizer.fit(x_train['collegeGPA'].values.reshape(1, -1))
x_train_collegegpa = normalizer.transform(x_train['collegeGPA'].values.reshape(1, -1)).T
x_cv_collegegpa = normalizer.transform(x_cv['collegeGPA'].values.reshape(1, -1)).T
x_test_collegegpa = normalizer.transform(x_test['collegeGPA'].values.reshape(1, -1)).T
print(x_train_collegegpa.shape)
print(x_cv_collegegpa.shape)
print(x_test_collegegpa.shape)
```

```
(1958, 1)
(840, 1)
(1200, 1)
```

## 1.6.10 Encoding the college state

In [384]:

```
vectorizer = CountVectorizer()
vectorizer.fit(x_train['CollegeState'])
x_train_collegestate = vectorizer.transform(x_train['CollegeState'])
x_cv_collegestate = vectorizer.transform(x_cv['CollegeState'])
x_test_collegestate = vectorizer.transform(x_test['CollegeState'])
print(x_train_collegestate.shape)
print(x_cv_collegestate.shape)
print(x_test_collegestate.shape)
```

```
(1958, 32)
(840, 32)
(1200, 32)
```

### 1.6.11 Encoding the English score

In [385]:

```
normalizer = preprocessing.Normalizer()
normalizer.fit(x_train['English'].values.reshape(1,-1))
x_train_english = normalizer.transform(x_train['English'].values.reshape(1,-1)).T
x_cv_english = normalizer.transform(x_cv['English'].values.reshape(1,-1)).T
x_test_english = normalizer.transform(x_test['English'].values.reshape(1,-1)).T
print(x_train_english.shape)
print(x_cv_english.shape)
print(x_test_english.shape)
```

```
(1958, 1)
(840, 1)
(1200, 1)
```

### 1.6.12 Encoding the Logical score

In [386]:

```
normalizer = preprocessing.Normalizer()
normalizer.fit(x_train['Logical'].values.reshape(1,-1))
x_train_logical = normalizer.transform(x_train['Logical'].values.reshape(1,-1)).T
x_cv_logical = normalizer.transform(x_cv['Logical'].values.reshape(1,-1)).T
x_test_logical = normalizer.transform(x_test['Logical'].values.reshape(1,-1)).T
print(x_train_logical.shape)
print(x_cv_logical.shape)
print(x_test_logical.shape)
```

```
(1958, 1)
(840, 1)
(1200, 1)
```

### 1.6.13 Encoding the Quants score

In [387]:

```
normalizer = preprocessing.Normalizer()
normalizer.fit(x_train['Quant'].values.reshape(1,-1))
x_train_quant = normalizer.transform(x_train['Quant'].values.reshape(1,-1)).T
x_cv_quant = normalizer.transform(x_cv['Quant'].values.reshape(1,-1)).T
x_test_quant = normalizer.transform(x_test['Quant'].values.reshape(1,-1)).T
print(x_train_quant.shape)
print(x_cv_quant.shape)
print(x_test_quant.shape)
```

```
(1958, 1)
(840, 1)
(1200, 1)
```

### 1.6.14 Encoding the Domain\_score

In [388]:

```
x_train['Domain'] = x_train['Domain'].replace(-1, 0)
x_cv['Domain'] = x_cv['Domain'].replace(-1, 0)
x_test['Domain'] = x_test['Domain'].replace(-1, 0)
```

Leaving the personality test scores as is, because they are sampled from a distribution with mean 0 and standard deviation 1.

## 1.7 Creating feature sets

In [389]:

```
x_tr = hstack((x_train_gender,x_train_dob,x_train_10percentage,x_train_10board,x_train_12graduation'].values.reshape(1,-1).T,x_train_12percentage,x_train_12board,x_train_12CollegeTier'].values.reshape(1,-1).T,x_train_degree,x_train_specialization,x_train_collegegpa,x_train_12CollegeCityTier'].values.reshape(1,-1).T,x_train_collegestate,x_train_12GraduationYear'].values.reshape(1,-1).T,x_train_english,x_train_logical,x_train_quant,x_train_12Domain'].values.reshape(1,-1).T,x_train_12conscientiousness'].values.reshape(1,-1).T,x_train_12agreeableness'].values.reshape(1,-1).T,x_train_12extraversion'].values.reshape(1,-1).T,x_train_12nueroticism'].values.reshape(1,-1).T,x_train_12openess_to_experience'].values.reshape(1,-1).T)).tocsr()
x_cv = hstack((x_cv_gender,x_cv_dob,x_cv_10percentage,x_cv_10board,x_cv_12graduation'].values.reshape(1,-1).T,x_cv_12percentage,x_cv_12board,x_cv_12CollegeTier'].values.reshape(1,-1).T,x_cv_degree,x_cv_specialization,x_cv_collegegpa,x_cv_12CollegeCityTier'].values.reshape(1,-1).T,x_cv_collegestate,x_cv_12GraduationYear'].values.reshape(1,-1).T,x_cv_english,x_cv_logical,x_cv_quant,x_cv_12Domain'].values.reshape(1,-1).T,x_cv_12conscientiousness'].values.reshape(1,-1).T,x_cv_12agreeableness'].values.reshape(1,-1).T,x_cv_12extraversion'].values.reshape(1,-1).T,x_cv_12nueroticism'].values.reshape(1,-1).T,x_cv_12openess_to_experience'].values.reshape(1,-1).T)).tocsr()
x_ts = hstack((x_test_gender,x_test_dob,x_test_10percentage,x_test_10board,x_test_12graduation'].values.reshape(1,-1).T,x_test_12percentage,x_test_12board,x_test_12CollegeTier'].values.reshape(1,-1).T,x_test_degree,x_test_specialization,x_test_collegegpa,x_test_12CollegeCityTier'].values.reshape(1,-1).T,x_test_collegestate,x_test_12GraduationYear'].values.reshape(1,-1).T,x_test_english,x_test_logical,x_test_quant,x_test_12Domain'].values.reshape(1,-1).T,x_test_12conscientiousness'].values.reshape(1,-1).T,x_test_12agreeableness'].values.reshape(1,-1).T,x_test_12extraversion'].values.reshape(1,-1).T,x_test_12nueroticism'].values.reshape(1,-1).T,x_test_12openess_to_experience'].values.reshape(1,-1).T)).tocsr()
```

## 1.8 Hyperparameter tuning using Gridsearch

In [393]:

```
estimator = RandomForestRegressor()
param_grid = {
    "n_estimators"      : [10,20,30],
    "max_features"      : ["auto", "sqrt", "log2"],
    "min_samples_split" : [2,4,8],
    "bootstrap": [True, False],
}
grid = GridSearchCV(estimator, param_grid, n_jobs=-1)
grid.fit(x_tr, y_train)
```

Out[393]:

```
GridSearchCV(cv=None, error_score=nan,
             estimator=RandomForestRegressor(bootstrap=True, ccp_alpha=0.0,
             criterion='mse', max_depth=None,
             max_features='auto',
             max_leaf_nodes=None,
             max_samples=None,
             min_impurity_decrease=0.0,
             min_impurity_split=None,
             min_samples_leaf=1,
             min_samples_split=2,
             min_weight_fraction_leaf=0.0,
             n_estimators=100, n_jobs=-1,
             oob_score=False, random_state=None,
             verbose=0, warm_start=False),
             iid='deprecated', n_jobs=-1,
             param_grid={'bootstrap': [True, False],
                         'max_features': ['auto', 'sqrt', 'log2'],
                         'min_samples_split': [2, 4, 8],
                         'n_estimators': [10, 20, 30]},
             pre_dispatch='2*n_jobs', refit=True, return_train_score=False,
             scoring=None, verbose=0)
```

In [394]:

```
grid.best_params_
```

Out[394]:

```
{'bootstrap': False,
 'max_features': 'sqrt',
 'min_samples_split': 4,
 'n_estimators': 30}
```

## 1.9 Training Random forest regressor with best hyper parameters found

In [396]:

```
rf = RandomForestRegressor(n_estimators=30,min_samples_split=4,max_features='sqrt',bootstrap=False)
rf.fit(x_tr,y_train)
```

Out[396]:

```
RandomForestRegressor(bootstrap=False, ccp_alpha=0.0, criterion='mse',
                       max_depth=None, max_features='sqrt', max_leaf_
nodes=None,
                       max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=4, min_weight_fraction_leaf=
0.0,
                       n_estimators=30, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False)
```

In [398]:

```
print(rf.score(x_cv,y_cv))
print(rf.score(x_ts,y_test))
```

```
-1.1461096149719303
-0.12015566501889308
```