INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

EDA Project - AMCAT Data Analysis

By
Lingerkar Rithikha

About me

Currently pursuing a Master of Science in Data Science, my educational background is rooted in a strong foundation of analytical skills. My passion lies in unraveling the hidden patterns within datasets and extracting valuable insights. Through a previous internship in the insurance sector, I gained practical experience in data analysis, honing my ability to derive meaningful conclusions from complex datasets. My motivation to learn data science stems from a deep-seated curiosity and a desire to contribute to the field by leveraging data-driven approaches to solve real-world problems.

www.linkedin.com/in/rithikha-lingerkar-82b18128a

github.com/LingerkarRithikha

INNOMATICS RESEARCH LABS

Agenda (This should be the PPT flow)

- Business Problem and Use case domain understanding(If Required)
- Objective of the Project
- Web Scraping – Details (Websites, Processor you followed)
- Summary of the Data

- <u>Exploratory Data Analysis:</u>

a. *Data Cleaning Steps*
b. *Data Manipulation Steps*
c. *Univariate Analysis Steps*
d. *Bivariate Analysis Steps*

- Key Business Question
- Conclusion (Key finding overall)
- Q&A Slide
- Your Experience/Challenges working on Web Scraping – Data Analysis Project.

INNOMATICS RESEARCH LABS

Dataset Description

The Aspiring Mind Employment Outcome 2015 (AMEO) dataset, released by Aspiring Minds, focuses on employment outcomes for engineering graduates. It includes dependent variables such as Salary, Job Titles, and Job Locations, along with standardized scores in cognitive skills, technical skills, and personality skills. With around 40 independent variables and 4000 data points, these variables encompass both continuous and categorical data. The dataset also includes demographic features and unique identifiers for each candidate.

Objective

The goal of this Exploratory Data Analysis (EDA) is to extensively investigate the provided dataset, with a particular emphasis on understanding the link between various variables and the target variable, Salary.
The key aims of this analysis include:

- Providing a detailed explanation of the dataset's features.
- Find any observable patterns or trends in the data.
- Investigating the relationships between the independent factors and the target variable (salary).

Summary of Dataset

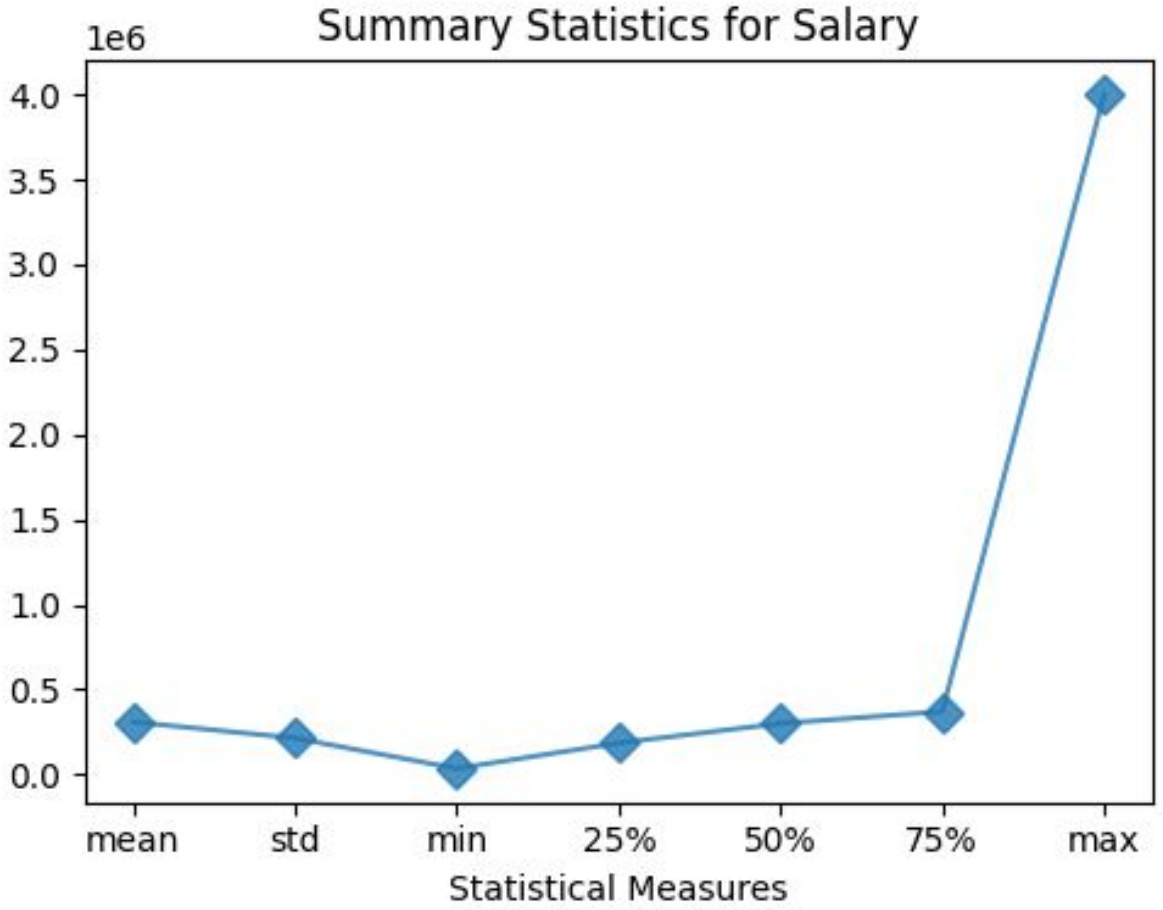
The Aspiring Mind Employment Outcome 2015 (AMEO) dataset, released by Aspiring Minds, focuses on employment outcomes for engineering graduates. It includes dependent variables such as Salary, Job Titles, and Job Locations, along with standardized scores in cognitive skills, technical skills, and personality skills. With around 40 independent variables and 4000 data points, these
variables encompass both continuous and categorical data. The dataset also includes demographic features and unique identifiers for each candidate

Problem Statement

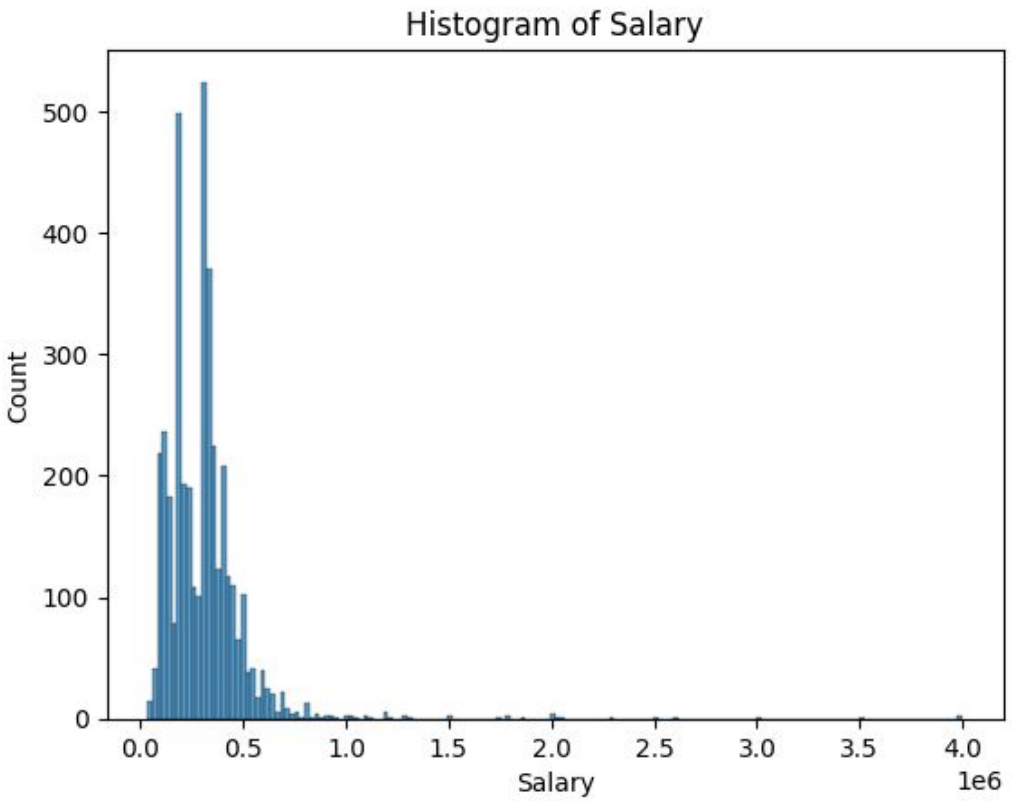
Exploring the relationship between various factors and salary in a dataset comprising academic information of individuals. Investigate how factors such as gender, education, specialization, location, and other attributes correlate with salary levels. Identify patterns, trends, and potential predictors of salary to provide insights for recruitment strategies, career planning, and salary negotiations.

Steps Involved to solve

- Importing the Libraries
- Loading the EDA dataset
- Finding the shape, describe, Information of the dataset
- Data cleaning methods like finding out null values , value counts , duplicate values , unique values .Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset .
- Data Manipulation steps likes merging and joining (used for DOB column) , mapping function for 10board , 12board .
- Univariate Analysis - for categorical values and numerical values - Visual representations plotting graphs Box plot, histogram , PDF , Count plot and Non-visual representations - summary statistic (mean,standard deviation , min, max, 25 percentage etc)
- Bivariate Analysis - for numerical vs numerical columns , numerical vs categorical column , categorical vs categorical columns - Visual representations scatter plot, bar plot, hexbin plot, box plot, stacked plot and Non-visual representations correlation matrix .
- Conclusions based upon the Univariate and Bi variate analysis .

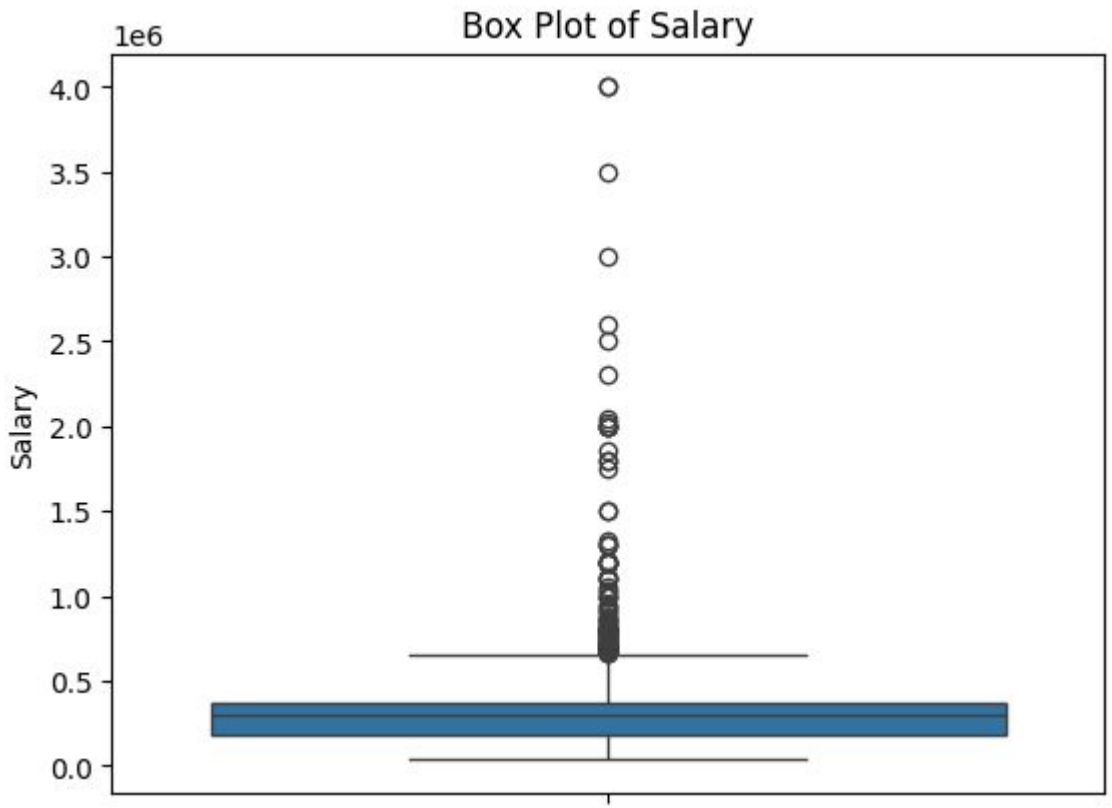


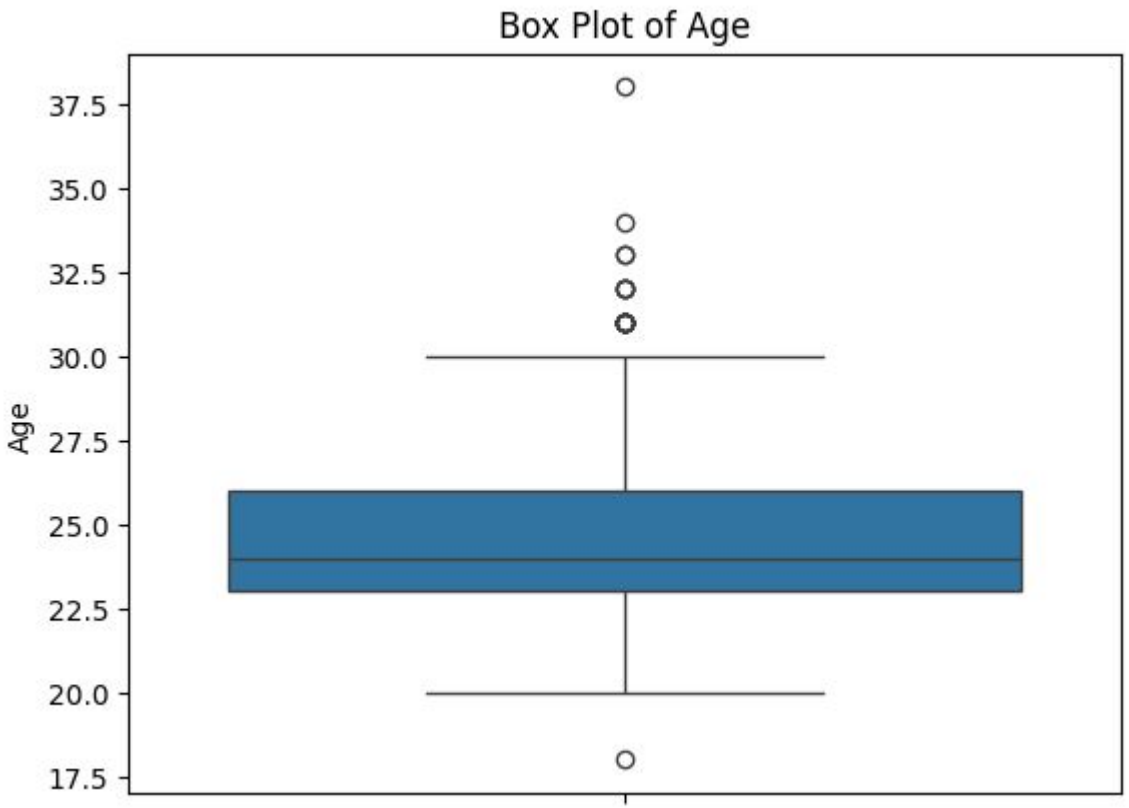
Summary Statistic for Salary



The distribution of salary is also right-skewed, with most people earning less than 10 lakhs per annum.

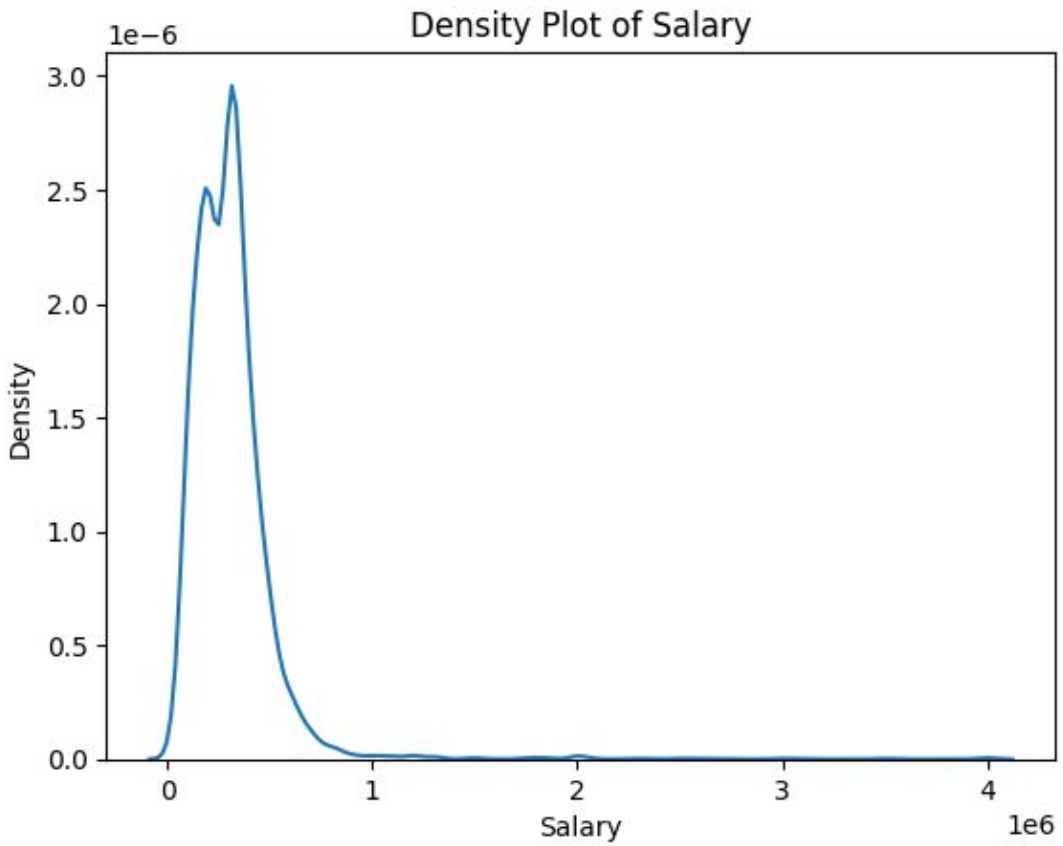
INNOMATICS RESEARCH LABS



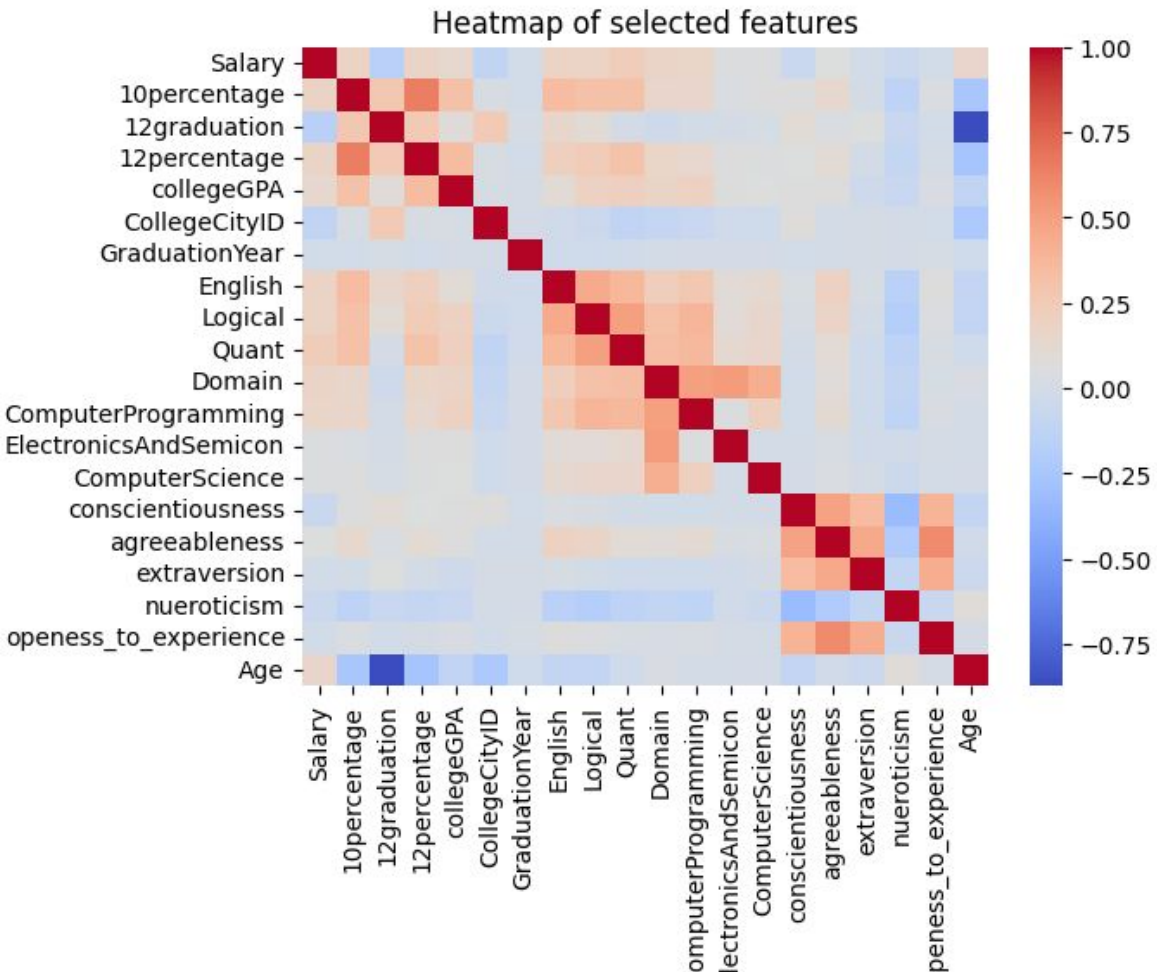


The median salary is around 5 lakhs per annum. There are some outliers with salaries above 20 lakhs per annum.

The median age is around 24 years. There are some outliers with ages above 40 years.

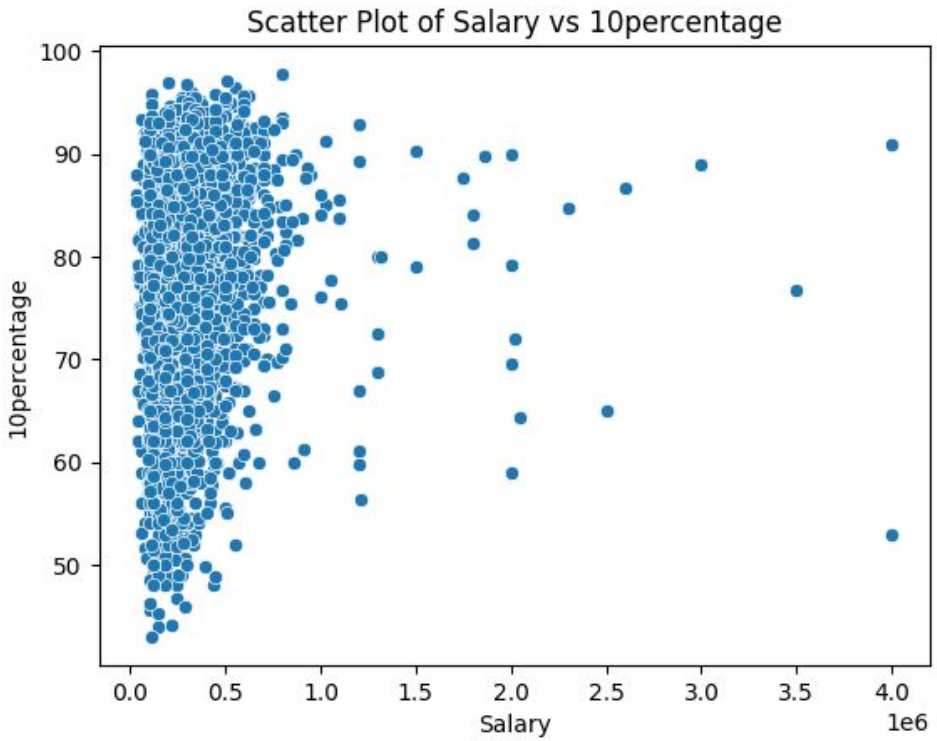


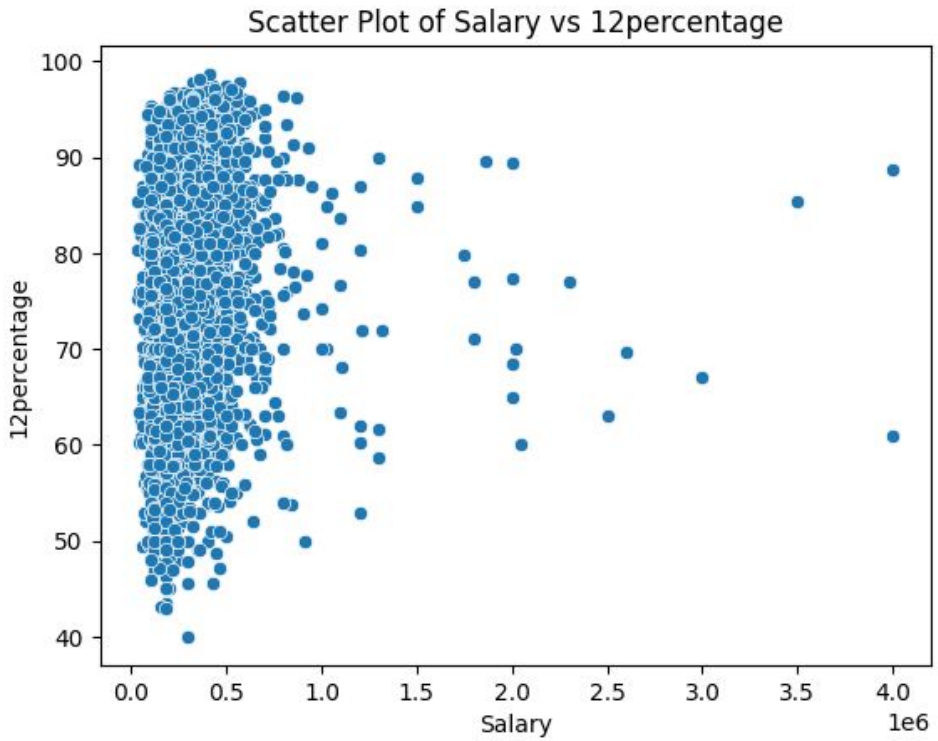
The density plots confirm the observations made from the histograms and box plots. The distributions of some of the variables are skewed. For example, the distribution of salary is right-skewed. There are some outliers in the data, particularly for salary.

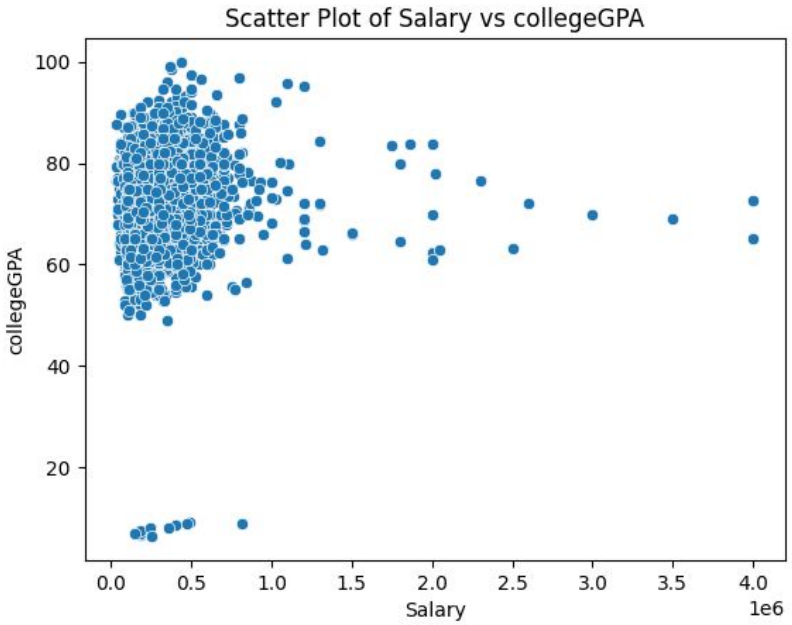


The heatmap shows that there are several positive correlations between the numerical variables in the dataset. This suggests that these variables may be related to each other and could potentially be used together in a predictive model.

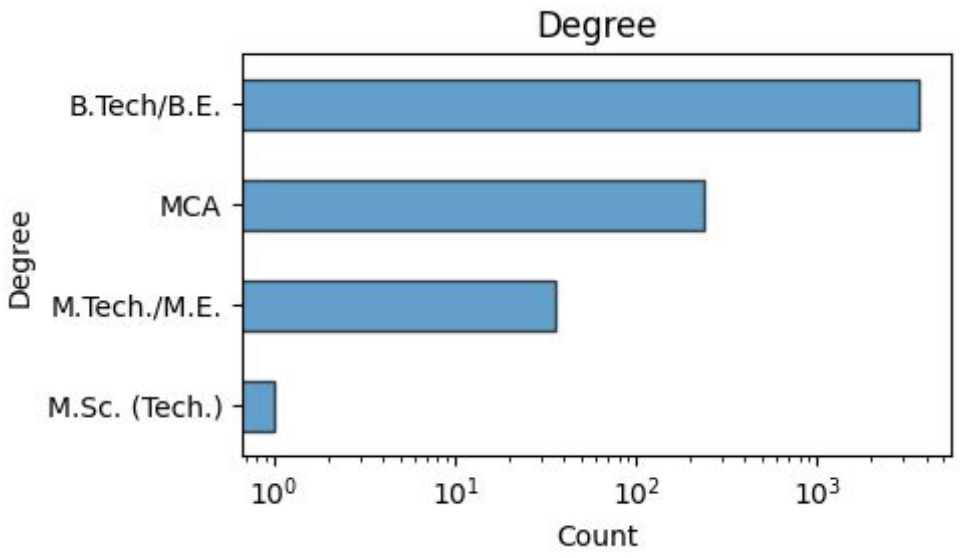
INNOMATICS RESEARCH LABS

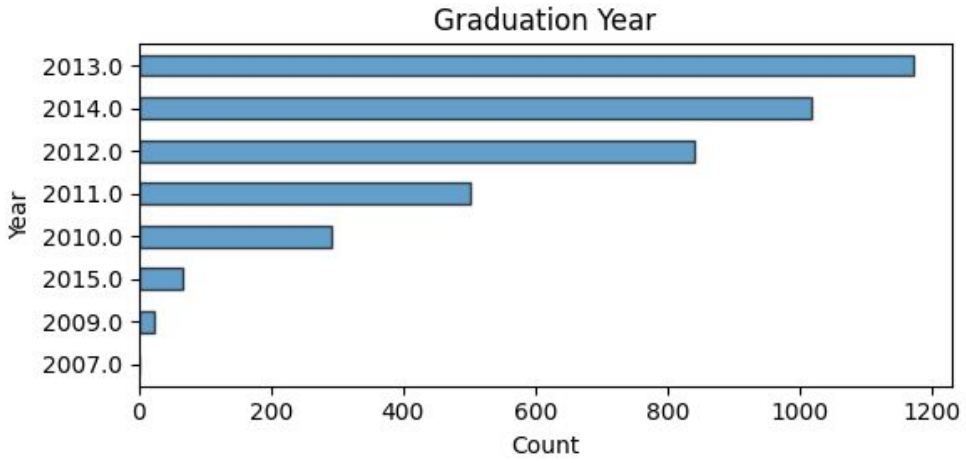






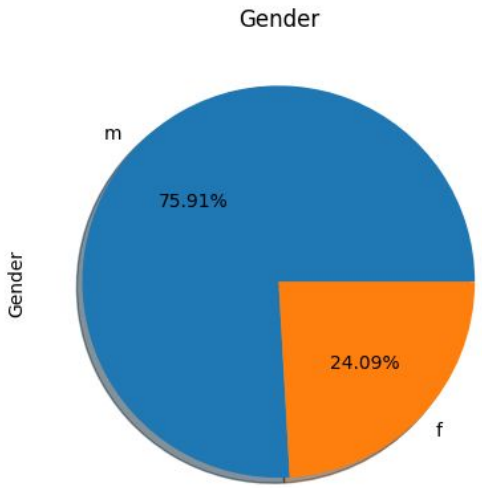
INNOMATICS RESEARCH LABS





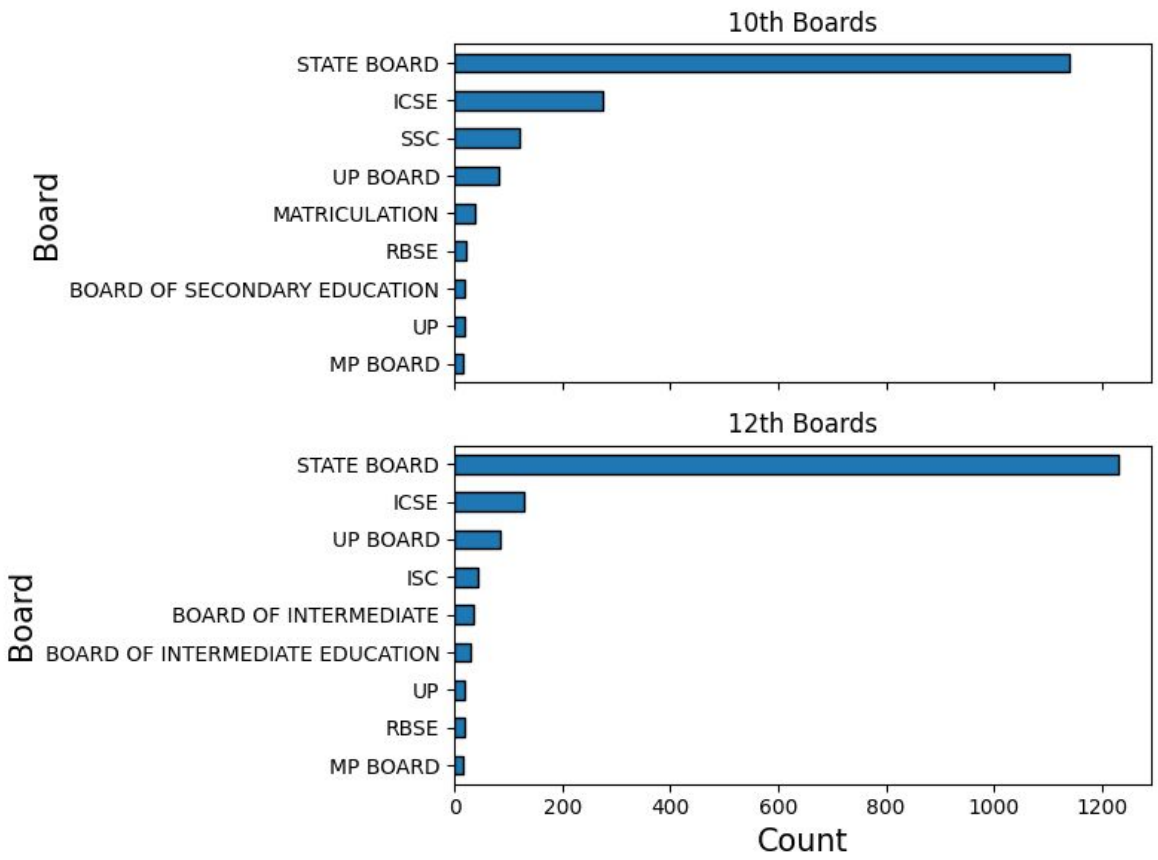
The Bar plot for Degree shows that the most common degree among the candidates is BE/B.Tech, indicating a strong emphasis on engineering and technology education.

The Bar plot for Graduation Year shows that 2013 is the year where many candidates have been graduated.



The Pie chart for Gender say that they are many male candidates than females.

INNOMATICS RESEARCH LABS



The scatter plot shows a positive correlation between age and salary, indicating that older candidates tend to earn higher salaries than younger candidates.

INNOMATICS RESEARCH LABS

Conclusion

The analysis of the provided dataset uncovers several significant insights regarding salary distribution, gender pay gap, and specialization trends within the workforce. Firstly, it reveals a right-skewed distribution of salaries, indicating a presence of high-salary outliers and suggesting that a considerable portion of candidates earn above-average salaries. Moreover, there exists a notable discrepancy in salaries across different designations, with Software Engineers commanding the highest average salary while Data Analysts earn the lowest.

Furthermore, the dataset highlights a concerning gender pay gap, where men consistently earn higher salaries than women on average. This finding underscores the need for further exploration into the underlying causes contributing to this disparity. Additionally, the analysis reveals a positive correlation between age and salary, implying that older candidates tend to earn higher salaries compared to their younger counterparts.

Moreover, there is a noticeable gender imbalance within the top specializations, with males predominating in most areas. This observation raises questions about the factors driving this gender disparity within specific fields. Overall, these insights provide valuable information for identifying areas of concern and potential avenues for promoting fairness and inclusivity in the workplace. Further investigation into the root causes of these trends could lead to informed strategies aimed at addressing salary discrepancies, promoting gender equality, and fostering a more inclusive work environment.

THANK YOU

INNOMATICS
RESEARCH LABS

```python
#Importing libraires

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_excel("/content/EDA_Data.xlsx")

df.head()
```

{"type":"dataframe","variable_name":"df"}

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3998 entries, 0 to 3997
Data columns (total 39 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Unnamed: 0             3998 non-null   object
 1   ID                     3998 non-null   int64
 2   Salary                 3998 non-null   int64
 3   DOJ                    3998 non-null   datetime64[ns]
 4   DOL                    3998 non-null   object
 5   Designation            3998 non-null   object
 6   JobCity                3998 non-null   object
 7   Gender                 3998 non-null   object
 8   DOB                    3998 non-null   datetime64[ns]
 9   10percentage           3998 non-null   float64
 10  10board                3998 non-null   object
 11  12graduation           3998 non-null   int64
 12  12percentage           3998 non-null   float64
 13  12board                3998 non-null   object
 14  CollegeID              3998 non-null   int64
 15  CollegeTier            3998 non-null   int64
 16  Degree                 3998 non-null   object
 17  Specialization         3998 non-null   object
 18  collegeGPA             3998 non-null   float64
 19  CollegeCityID          3998 non-null   int64
 20  CollegeCityTier        3998 non-null   int64
 21  CollegeState           3998 non-null   object
 22  GraduationYear         3998 non-null   int64
 23  English                3998 non-null   int64
 24  Logical                3998 non-null   int64
 25  Quant                  3998 non-null   int64
 26  Domain                 3998 non-null   float64
 27  ComputerProgramming    3998 non-null   int64
 28  ElectronicsAndSemicon  3998 non-null   int64
 29  ComputerScience        3998 non-null   int64
```

```
 30  MechanicalEngg         3998 non-null   int64
 31  ElectricalEngg         3998 non-null   int64
 32  TelecomEngg            3998 non-null   int64
 33  CivilEngg              3998 non-null   int64
 34  conscientiousness      3998 non-null   float64
 35  agreeableness          3998 non-null   float64
 36  extraversion           3998 non-null   float64
 37  nueroticism            3998 non-null   float64
 38  openess_to_experience  3998 non-null   float64
dtypes: datetime64[ns](2), float64(9), int64(18), object(10)
memory usage: 1.2+ MB
```

```
df.shape
```

```
(3998, 39)
```

```
df.columns
```

```
Index(['Unnamed: 0', 'ID', 'Salary', 'DOJ', 'DOL', 'Designation',
'JobCity',
       'Gender', 'DOB', '10percentage', '10board', '12graduation',
       '12percentage', '12board', 'CollegeID', 'CollegeTier',
'Degree',
       'Specialization', 'collegeGPA', 'CollegeCityID',
'CollegeCityTier',
       'CollegeState', 'GraduationYear', 'English', 'Logical',
'Quant',
       'Domain', 'ComputerProgramming', 'ElectronicsAndSemicon',
       'ComputerScience', 'MechanicalEngg', 'ElectricalEngg',
'TelecomEngg',
       'CivilEngg', 'conscientiousness', 'agreeableness',
'extraversion',
       'nueroticism', 'openess_to_experience'],
      dtype='object')
```

```
categorical_columns = df.select_dtypes(include=['object']).columns
numerical_columns = df.select_dtypes(exclude=['object']).columns

# Displaying categorical columns
print("Categorical Columns:")
print(categorical_columns)

# Displaying numerical columns
print("\nNumerical Columns:")
print(numerical_columns)
```

```
Categorical Columns:
Index(['Unnamed: 0', 'DOL', 'Designation', 'JobCity', 'Gender',
'10board',
       '12board', 'Degree', 'Specialization', 'CollegeState'],
      dtype='object')
```

```
Numerical Columns:
Index(['ID', 'Salary', 'DOJ', 'DOB', '10percentage', '12graduation',
       '12percentage', 'CollegeID', 'CollegeTier', 'collegeGPA',
       'CollegeCityID', 'CollegeCityTier', 'GraduationYear',
'English',
       'Logical', 'Quant', 'Domain', 'ComputerProgramming',
       'ElectronicsAndSemicon', 'ComputerScience', 'MechanicalEngg',
       'ElectricalEngg', 'TelecomEngg', 'CivilEngg',
'conscientiousness',
       'agreeableness', 'extraversion', 'nueroticism',
       'openess_to_experience'],
      dtype='object')
```

```
print(type(categorical_columns))
```

```
<class 'pandas.core.indexes.base.Index'>
```

```
for column in categorical_columns:

print("*************************************************************")
  print(pd.Series(df[column]).value_counts())
```

```
*************************************************************
train    3998
Name: Unnamed: 0, dtype: int64
*************************************************************
present                 1875
2015-04-01 00:00:00      573
2015-03-01 00:00:00      124
2015-05-01 00:00:00      112
2015-01-01 00:00:00       99
                        ...
2005-03-01 00:00:00        1
2015-10-01 00:00:00        1
2010-02-01 00:00:00        1
2011-02-01 00:00:00        1
2010-10-01 00:00:00        1
Name: DOL, Length: 67, dtype: int64
*************************************************************
software engineer                     539
software developer                    265
system engineer                       205
programmer analyst                    139
systems engineer                      118
                                     ...
cad drafter                             1
noc engineer                            1
human resources intern                  1
senior quality assurance engineer       1
```

```
jr. software developer                 1
Name: Designation, Length: 419, dtype: int64
************************************************************
Bangalore           627
-1                  461
Noida               368
Hyderabad           335
Pune                290
                   ...
Tirunelvelli          1
Ernakulam             1
Nanded                1
Dharmapuri            1
Asifabadbanglore      1
Name: JobCity, Length: 339, dtype: int64
************************************************************
m    3041
f     957
Name: Gender, dtype: int64
************************************************************
cbse                          1395
state board                   1164
0                              350
icse                           281
ssc                            122
                              ...
hse,orissa                       1
national public school            1
nagpur board                      1
jharkhand academic council        1
bse,odisha                        1
Name: 10board, Length: 275, dtype: int64
************************************************************
cbse                                  1400
state board                           1254
0                                      359
icse                                   129
up board                                87
                                      ...
jawahar higher secondary school          1
nagpur board                             1
bsemp                                    1
board of higher secondary orissa         1
boardofintermediate                      1
Name: 12board, Length: 340, dtype: int64
************************************************************
B.Tech/B.E.      3700
MCA               243
M.Tech./M.E.       53
```

```
M.Sc. (Tech.)       2
Name: Degree, dtype: int64
************************************************************
electronics and communication engineering      880
computer science & engineering                 744
information technology                         660
computer engineering                           600
computer application                           244
mechanical engineering                         201
electronics and electrical engineering         196
electronics & telecommunications               121
electrical engineering                          82
electronics & instrumentation eng               32
civil engineering                               29
electronics and instrumentation engineering     27
information science engineering                 27
instrumentation and control engineering         20
electronics engineering                         19
biotechnology                                   15
other                                           13
industrial & production engineering             10
applied electronics and instrumentation          9
chemical engineering                             9
computer science and technology                  6
telecommunication engineering                    6
mechanical and automation                        5
automobile/automotive engineering                5
instrumentation engineering                      4
mechatronics                                     4
aeronautical engineering                         3
electronics and computer engineering             3
electrical and power engineering                 2
biomedical engineering                           2
information & communication technology           2
industrial engineering                           2
computer science                                 2
metallurgical engineering                        2
power systems and automation                     1
control and instrumentation engineering          1
mechanical & production engineering              1
embedded systems technology                      1
polymer technology                               1
computer and communication engineering           1
information science                              1
internal combustion engine                       1
computer networking                              1
ceramic engineering                              1
electronics                                      1
industrial & management engineering              1
```

```
Name: Specialization, dtype: int64
************************************************************
Uttar Pradesh        915
Karnataka            370
Tamil Nadu           367
Telangana            319
Maharashtra          262
Andhra Pradesh       225
West Bengal          196
Punjab               193
Madhya Pradesh       189
Haryana              180
Rajasthan            174
Orissa               172
Delhi                162
Uttarakhand          113
Kerala                33
Jharkhand             28
Chhattisgarh          27
Gujarat               24
Himachal Pradesh      16
Bihar                 10
Jammu and Kashmir      7
Assam                  5
Union Territory        5
Sikkim                 3
Meghalaya              2
Goa                    1
Name: CollegeState, dtype: int64
```

```
df.isnull().sum()
```

```
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
nueroticism               0
openess_to_experience     0
dtype: int64
```

```
df.duplicated().sum()
```

```
0
```

```
df.nunique()
```

```
Unnamed: 0                   1
ID                        3998
Salary                     177
DOJ                         81
DOL                         67
Designation                419
JobCity                    339
Gender                       2
DOB                       1872
10percentage               851
10board                    275
12graduation                16
12percentage               801
12board                    340
CollegeID                 1350
CollegeTier                  2
Degree                       4
Specialization              46
collegeGPA                1282
CollegeCityID             1350
CollegeCityTier              2
```

```
CollegeState                26
GraduationYear              11
English                    111
Logical                    107
Quant                      138
Domain                     243
ComputerProgramming         79
ElectronicsAndSemicon       29
ComputerScience             20
MechanicalEngg              42
ElectricalEngg              31
TelecomEngg                 26
CivilEngg                   23
conscientiousness          141
agreeableness              149
extraversion               154
nueroticism                217
openess_to_experience      142
dtype: int64
```

```
df = df.drop(columns = ['Unnamed: 0',
'ID','CollegeID','CollegeCityID', 'MechanicalEngg', 'ElectricalEngg',
'TelecomEngg', 'CivilEngg'])
df.head()
```

{"type":"dataframe","variable_name":"df"}

```
df.shape
```

```
(3998, 31)
```

```
df['DOB'] = pd.to_datetime(df['DOB'])
df['Age'] = 2015 - df['DOB'].dt.year
df.head()
```

{"type":"dataframe","variable_name":"df"}

```
df['10board'] = df['10board'].replace({'0':np.nan})
df['12board'] = df['12board'].replace({'0':np.nan})
df['GraduationYear'] = df['GraduationYear'].replace({0:np.nan})
df['JobCity'] = df['JobCity'].replace({'-1':np.nan})
df['Domain'] = df['Domain'].replace({-1:np.nan})
df['ElectronicsAndSemicon'] = df['ElectronicsAndSemicon'].replace({-
1:0})
df['ComputerScience'] = df['ComputerScience'].replace({-1:0})
df['ComputerProgramming'] = df['ComputerProgramming'].replace({-
1:np.nan})
```

```
df.Age.min()
```

```
18
```

```
df.Age.max()
```

```
38
```

```
df.Salary.min()
```

```
35000
```

```
df.Salary.max()
```

```
4000000
```

```
len(df[(df['GraduationYear'] > df['DOJ'].dt.year)].index)
```

```
80
```

```
df = df.drop(df[(df['GraduationYear'] > df['DOJ'].dt.year)].index)
```

```
df
```

{"type":"dataframe","variable_name":"df"}

```
df.JobCity.value_counts()
```

```
Bangalore           612
-1                  452
Noida               361
Hyderabad           330
Pune                288
                   ...
Nanded                1
Dharmapuri            1
Vandavasi             1
Rohtak                1
Asifabadbanglore      1
Name: JobCity, Length: 331, dtype: int64
```

```
df.JobCity.unique()
```

```
array(['Bangalore', 'Indore', 'Chennai', 'Gurgaon', 'Manesar',
       'Hyderabad', 'Banglore', 'Noida', 'Kolkata', 'Pune', -1,
'mohali',
       'Jhansi', 'Delhi', 'Hyderabad ', 'Bangalore ', 'noida',
'delhi',
       'Bhubaneswar', 'Navi Mumbai', 'Mumbai', 'New Delhi',
'Mangalore',
       'Rewari', 'Gaziabaad', 'Bhiwadi', 'Mysore', 'Rajkot',
       'Greater Noida', 'Jaipur', 'noida ', 'HYDERABAD', 'mysore',
       'THANE', 'Maharajganj', 'Thiruvananthapuram', 'Punchkula',
       'Bhubaneshwar', 'Pune ', 'coimbatore', 'Dhanbad', 'Lucknow',
       'Trivandrum', 'kolkata', 'mumbai', 'Gandhi Nagar', 'Una',
       'Daman and Diu', 'chennai', 'GURGOAN', 'vsakhapttnam', 'pune',
```

```
       'Nagpur', 'Bhagalpur', 'new delhi - jaisalmer', 'Coimbatore',
       'Ahmedabad', 'Kochi/Cochin', 'Bankura', 'Bengaluru', 'Mysore ',
       'Kanpur ', 'jaipur', 'Gurgaon ', 'bangalore', 'CHENNAI',
       'Vijayawada', 'Kochi', 'Beawar', 'Alwar', 'NOIDA', 'Greater
noida',
       'Siliguri ', 'raipur', 'gurgaon', 'Faridabad', 'Jodhpur',
       'udaipur', 'Muzaffarpur', 'Kolkata`', 'Bulandshahar',
'Haridwar',
       'Raigarh', 'Visakhapatnam', 'Jabalpur', 'hyderabad', 'Unnao',
       'KOLKATA', 'Thane', 'Aurangabad', 'Belgaum', 'gurgoan',
'Dehradun',
       'Rudrapur', 'Jamshedpur', 'vizag', 'Nouda', 'Dharamshala',
       'Hissar', 'Ranchi', 'BANGALORE', 'Madurai', 'Gurga',
'Chandigarh',
       'Australia', ' Chennai', 'CHEYYAR', 'Mumbai ', 'sonepat',
       'Ghaziabad', 'Pantnagar', 'Siliguri', 'mumbai ', 'Jagdalpur',
       'Chennai ', 'angul', 'Baroda', ' ariyalur',
       'Kochi/Cochin, Chennai and Coimbatore', 'Bhopal',
'bhubaneswar',
       'Neemrana', 'VIZAG', 'Tirupathi', 'Lucknow ', 'Ahmedabad ',
       'Bhubneshwar', 'Noida ', 'pune ', 'Calicut', 'Gandhinagar',
       'LUCKNOW', 'Dubai', 'bengaluru', 'MUMBAI', 'Ahmednagar',
'Nashik',
       'New delhi', 'Bellary', 'Ludhiana', 'New Delhi ',
'Muzaffarnagar',
       'BHOPAL', 'Gagret', 'Indirapuram, Ghaziabad', 'Gwalior',
       'new delhi', 'TRIVANDRUM', 'Chennai & Mumbai', 'Rajasthan',
       'Bareli', 'Kanpur', 'Hospete', ' mumbai', 'Dharuhera',
'lucknow',
       'meerut', 'dehradun', 'Ganjam', 'Hubli', 'bangalore ',
       'NAVI MUMBAI', 'ncr', 'Agra', 'Trichy', 'kudankulam ,tarapur',
       'Ongole', 'Sambalpur', 'Pondicherry', 'Bundi',
       'SADULPUR,RAJGARH,DISTT-CHURU,RAJASTHAN', 'Bikaner',
'Vadodara',
       'BAngalore', 'india', 'Asansol', 'Ernakulam', 'DELHI',
'Bilaspur',
       'Chandrapur', 'Nanded', 'Dharmapuri', 'Vandavasi', 'Rohtak',
       'trivandrum', 'Nagpur ', 'Udaipur', 'Patna', 'banglore',
'indore',
       'Salem', 'Nasikcity', 'Gandhinagar ', 'Technopark, Trivandrum',
       'Bharuch', 'Tornagallu', 'Raipur', 'Kolkata ', 'Jaspur',
'Burdwan',
       'Bhubaneswar ', 'Shimla', 'ahmedabad', 'Gajiabaad', 'Jammu',
       'Shahdol', 'Muvattupuzha', 'Al Jubail,Saudi Arabia',
       'Kalmar, Sweden', 'Secunderabad', 'A-64,sec-64,noida',
'Ratnagiri',
       'Jhajjar', 'Gulbarga', 'hyderabad(bhadurpally)', 'Chandigarh ',
       'Jaipur ', 'Jeddah Saudi Arabia', ' Delhi', 'PATNA', 'SHAHDOL',
       'Chennai, Bangalore', 'Bhopal ', 'Jamnagar', 'Gurgoan', 'PUNE',
```

```
       'Tirupati', 'Gonda', 'jamnagar', 'chennai ', 'orissa',
'kharagpur',
       'Trivandrum ', 'Navi Mumbai , Hyderabad', 'Joshimath',
       'chandigarh', 'Bathinda', 'Johannesburg', 'kala amb ',
'Karnal',
       'LONDON', 'Kota', 'Panchkula', 'Baddi HP', 'Nagari',
'Nalagarh',
       'Mettur, Tamil Nadu ', 'Durgapur', 'pondi', 'Surat', 'Kurnool',
       'kolhapur', 'Madurai ', 'GREATER NOIDA', 'Bhilai', ' Pune',
       'hderabad', 'KOTA', 'thane', 'Vizag', 'Bahadurgarh',
       'Rayagada, Odisha', 'kakinada', 'GURGAON', 'Varanasi', 'punr',
       'Nellore', 'patna', 'Meerut', 'hyderabad ', 'Sahibabad',
'Howrah',
       'BHUBANESWAR', 'Trichur', 'Ambala', 'Khopoli', 'keral',
'Roorkee',
       'Greater NOIDA', 'Navi mumbai', 'ghaziabad', 'Allahabad',
       'Delhi/NCR', 'Panchkula ', 'Sonipat', 'Ranchi ', 'Jalandhar',
       'manesar', 'PILANI', 'muzzafarpur', 'RAS AL KHAIMAH', 'bihar',
       'singaruli', 'KANPUR', 'Banglore ', 'pondy', 'Mohali', '
Mumbai',
       ' bangalore', 'GURAGAON', 'Baripada', 'MEERUT', 'Yamuna Nagar',
       'shahibabad', 'sampla', 'Guwahati', 'Rourkela', 'Banaglore',
       'Vellore', 'Dausa', 'latur (Maharashtra )', 'NEW DELHI',
'kanpur',
       'Mainpuri', 'karnal', 'Dammam', 'Haldia', 'sambalpur',
'ranchi',
       'jAipur', 'BANGLORE', 'Patiala', 'Gorakhpur', 'new dehli',
       'BANGALORE ', 'Ambala City', 'Karad', 'Rajpura', 'Pilani',
       'haryana', 'Asifabadbanglore'], dtype=object)
```

```python
data = df['JobCity']
cleaned_data = [str(x).lower() for x in data if x != -1]

# Combine similar values using a dictionary
combined_data = {}
for item in cleaned_data:
    if item not in combined_data:
        combined_data[item] = 1
    else:
        combined_data[item] += 1



# Print combined data
for key, value in combined_data.items():
    print(f"{key}: {value}")
```

```
bangalore: 632
indore: 25
chennai: 304
```

```
gurgaon: 207
manesar: 2
hyderabad: 352
banglore: 13
noida: 379
kolkata: 116
pune: 316
mohali: 12
jhansi: 1
delhi: 20
hyderabad : 11
bangalore : 17
bhubaneswar: 25
navi mumbai: 33
mumbai: 113
new delhi: 197
mangalore: 11
rewari: 1
gaziabaad: 1
bhiwadi: 1
mysore: 38
rajkot: 2
greater noida: 29
jaipur: 50
noida : 3
thane: 7
maharajganj: 1
thiruvananthapuram: 6
punchkula: 1
bhubaneshwar: 8
pune : 6
coimbatore: 21
dhanbad: 2
lucknow: 40
trivandrum: 9
gandhi nagar: 2
una: 1
daman and diu: 1
gurgoan: 3
vsakhapttnam: 1
nagpur: 9
bhagalpur: 1
new delhi - jaisalmer: 1
ahmedabad: 17
kochi/cochin: 13
bankura: 1
bengaluru: 5
mysore : 2
kanpur : 1
```

```
gurgaon : 7
vijayawada: 2
kochi: 2
beawar: 2
alwar: 1
siliguri : 1
raipur: 5
faridabad: 15
jodhpur: 4
udaipur: 3
muzaffarpur: 1
kolkata`: 1
bulandshahar: 1
haridwar: 3
raigarh: 1
visakhapatnam: 7
jabalpur: 7
unnao: 2
aurangabad: 3
belgaum: 2
dehradun: 10
rudrapur: 3
jamshedpur: 2
vizag: 3
nouda: 1
dharamshala: 1
hissar: 2
ranchi: 3
madurai: 3
gurga: 1
chandigarh: 26
australia: 1
 chennai: 1
cheyyar: 1
mumbai : 3
sonepat: 1
ghaziabad: 16
pantnagar: 2
siliguri: 1
jagdalpur: 1
chennai : 8
angul: 1
baroda: 1
 ariyalur: 1
kochi/cochin, chennai and coimbatore: 1
bhopal: 17
neemrana: 2
tirupathi: 3
lucknow : 1
```

```
ahmedabad : 1
bhubneshwar: 4
calicut: 1
gandhinagar: 2
dubai: 3
ahmednagar: 2
nashik: 3
bellary: 1
ludhiana: 3
new delhi : 2
muzaffarnagar: 1
gagret: 1
indirapuram, ghaziabad: 1
gwalior: 1
chennai & mumbai: 1
rajasthan: 1
bareli: 1
kanpur: 8
hospete: 1
 mumbai: 3
dharuhera: 1
meerut: 4
ganjam: 1
hubli: 2
ncr: 1
agra: 2
trichy: 5
kudankulam ,tarapur: 1
ongole: 1
sambalpur: 2
pondicherry: 6
bundi: 1
sadulpur,rajgarh,distt-churu,rajasthan: 1
bikaner: 1
vadodara: 2
india: 1
asansol: 1
ernakulam: 1
bilaspur: 3
chandrapur: 2
nanded: 1
dharmapuri: 1
vandavasi: 1
rohtak: 1
nagpur : 2
patna: 7
salem: 1
nasikcity: 3
gandhinagar : 1
```

```
technopark, trivandrum: 1
bharuch: 1
tornagallu: 1
kolkata : 1
jaspur: 1
burdwan: 1
bhubaneswar : 1
shimla: 1
gajiabaad: 1
jammu: 1
shahdol: 2
muvattupuzha: 1
al jubail,saudi arabia: 1
kalmar, sweden: 1
secunderabad: 3
a-64,sec-64,noida: 1
ratnagiri: 1
jhajjar: 1
gulbarga: 1
hyderabad(bhadurpally): 1
chandigarh : 1
jaipur : 1
jeddah saudi arabia: 1
 delhi: 1
chennai, bangalore: 1
bhopal : 1
jamnagar: 2
tirupati: 1
gonda: 1
orissa: 1
kharagpur: 1
trivandrum : 1
navi mumbai , hyderabad: 1
joshimath: 1
bathinda: 1
johannesburg: 1
kala amb : 1
karnal: 2
london: 1
kota: 2
panchkula: 1
baddi hp: 1
nagari: 1
nalagarh: 1
mettur, tamil nadu : 1
durgapur: 1
pondi: 1
surat: 1
kurnool: 1
```

```
kolhapur: 1
madurai : 1
bhilai: 1
 pune: 1
hderabad: 1
bahadurgarh: 1
rayagada, odisha: 1
kakinada: 1
varanasi: 1
punr: 1
nellore: 1
sahibabad: 1
howrah: 1
trichur: 1
ambala: 2
khopoli: 1
keral: 1
roorkee: 1
allahabad: 1
delhi/ncr: 1
panchkula : 1
sonipat: 1
ranchi : 1
jalandhar: 1
pilani: 2
muzzafarpur: 1
ras al khaimah: 1
bihar: 1
singaruli: 1
banglore : 1
pondy: 1
 bangalore: 1
guragaon: 1
baripada: 1
yamuna nagar: 1
shahibabad: 1
sampla: 1
guwahati: 1
rourkela: 1
banaglore: 1
vellore: 1
dausa: 1
latur (maharashtra ): 1
mainpuri: 1
dammam: 1
haldia: 1
patiala: 1
gorakhpur: 1
new dehli: 1
ambala city: 1
```

```
karad: 1
rajpura: 1
haryana: 1
asifabadbanglore: 1

board10_mapping = {
    'cbse' : 'CBSE',
    'cbse[gulf zone]': 'CBSE',
    'cbse board': 'CBSE',
    'hbse' : 'CBSE',
    'hbsc' : 'CBSE',

    'andhra pradesh board ssc' : 'AP Board',
    'andhra pradesh state board' : 'AP Board',
    'ap state board': 'AP Board',
    'ap state board for secondary education':'AP Board',
    'board ofsecondary education,ap': 'AP Board',
    'ssc board of andrapradesh' : 'AP Board',
    'board of secondary education,ap': 'AP Board',
    'state board of secondary education, ap': 'AP Board',
    'apssc': 'AP Board',
    'apsche': 'AP Board',
    'school secondary education, andhra pradesh':'AP Board',
    'state board of secondary education,andhra pradesh':'AP Board',
    'ssc-andhra pradesh' : 'AP Board',
    'board of ssc education andhra pradesh' : 'AP Board',
    'state board of secondary education, andhra pradesh': 'AP Board',
    'board of secondary education,andhra pradesh':'AP Board',
    'board of  secondary education':'SSC',


    'stateboard':'SSC',
    'state boardmp board ':'SSC',
    'state bord':'SSC',
    'secondary school certificate':'SSC',
    'secondary school education':'SSC',
    'seba':'SSC',
    'sslc':'SSC',
    'national public school':'SSC',
    'board of secondary education,andhara pradesh':'AP Board',
    'board of secondary education, andhra pradesh':'AP Board',
    'board of secondary education - andhra pradesh':'AP Board',
    'board of secondaray education':'SSC',
    'board of secondary school education':'SSC',

    'central board of secondary education':'CBSE',
    'certificate of middle years program of ib':'CBSE',
    'central board of secondary education, new delhi':'CBSE',

    'bihar board' : 'Bihar Board',
```

```
    'bihar examination board, patna' : 'Bihar Board',
    'bihar' : 'Bihar Board',
    'bihar secondary education board,patna' : 'Bihar Board',
    'bihar school examination board' : 'Bihar Board',
    'bihar school examination board patna' : 'Bihar Board',
    'biharboard' : 'Bihar Board',
    'bseb, patna' : 'Bihar Board',
    'bseb,patna': 'Bihar Board',
    'bseb patna': 'Bihar Board',
    'bsepatna': 'Bihar Board',

    'mp board': 'MP Board',
    'mp board bhopal': 'MP Board',
    'mpboard': 'MP Board',
    'mp': 'MP Board',
    'mp state board': 'MP Board',
    'mpbse': 'MP Board',
    'state boardmp board': 'MP Board',
    'mp-bse': 'MP Board',
    'bsemp': 'MP Board',
    'madhya pradesh board':'MP Board',

    'bse, odisha' : 'Odisha Board',
    'board of secondary education,odisha' : 'Odisha Board',
    'bse, odisha' : 'Odisha Board',
    'bse,odisha' : 'Odisha Board',

    'board of secondary education(bse) orissa' : 'Orissa Board',
    'bsc,orissa' : 'Orissa Board',
    'board of secendary education orissa' : 'Orissa Board',
    'bse,orissa' : 'Orissa Board',
    'hse,orissa' : 'Orissa Board',
    'board of secondary education orissa' : 'Orissa Board',
    'board of secondary education,orissa':'Orissa Board',
    'board of secondary education (bse) orissa': 'Orissa Board',

    'anglo indian' : 'Anglo Indian',
    'state board' : 'SSC',
    'state board ':'SSC',
    'ssc' : 'SSC',
    'sda matric higher secondary school' : 'SSC',
    'matriculation' : 'SSC',
    'little jacky matric higher secondary school' : 'SSC',
    'matriculation board' : 'SSC',
    'metric': 'SSC',
    'maticulation'
    'sslc' : 'SSC',
    'board secondary  education' : 'SSC',
    'matric board' : 'SSC',
    'board of secondary education' : 'SSC',
```

```
    'mirza ahmed ali baig' : 'SSC',
    'board of intermediate education' : 'SSC',
    'bseb':'CBSE',
    'stjosephs girls higher sec school,dindigul' : 'SSC',
    'secondary school cerfificate' : 'SSC',
    'state' : 'SSC',
    'matric': 'SSC',
    'maticulation': 'SSC',
    'secondary school of education' : 'SSC',
    'stjoseph of cluny matrhrsecschool,neyveli,cuddalore district' :
'SSC',
    'bharathi matriculation school' : 'SSC',
    'bse(board of secondary education)' : 'SSC',
    'secondary state certificate' : 'SSC',
    'bseb' : 'SSC',
    'stmary higher secondary' : 'SSC',
    'kiran english medium high school' : 'SSC',
    'sarada high scchool' : 'SSC',
    'sslc board' : 'SSC',
    'stmary\'s convent inter college' : 'SSC',
    'don bosco maatriculation school' : 'SSC',
    'state board of secondary education( ssc)' : 'SSC',
    'board of ssc': 'SSC',
    'ssc regular':'SSC',
    'kalaimagal matriculation higher secondary school':'SSC',
    'maticulation':'SSC',
    'matric board':'SSC',
    'matric':'SSC',
    'state board (jac, ranchi)':'SSC',

    'aurangabad board' : 'Aurangabad Board',

    'up' : 'UP Board',
    'up board' : 'UP Board',
    'up(allahabad)' : 'UP Board',
    'upbhsie' : 'UP Board',
    'up board , allahabad' : 'UP Board',
    'upboard' : 'UP Board',
    'bright way college, (up board)' : 'UP Board',
    'up board allahabad' : 'UP Board',
    'upboard' : 'UP Board',
    'up board,allahabad' : 'UP Board',
    'up-board' : 'UP Board',
    'uttar pradesh board': 'UP Board',
    'uttar pradesh': 'UP Board',
    'board of high school and intermediate education uttarpradesh':
'UP Board',
    'u p board': 'UP Board',
    'u p': 'UP Board',
```

```
    'up bord': 'UP Board',
    'up bourd': 'UP Board',
    'up baord': 'UP Board',
    'up borad': 'UP Board',

    'nagpur':'Nagpur Board',
    'nagpur board':'Nagpur Board',
    'nagpur board,nagpur':'Nagpur Board',
    'nagpur divisional board':'Nagpur Board',


    'uttarakhand board':'Uttarakhand Board',
    'uttaranchal shiksha avam pariksha parishad' : 'Uttaranchal
Board',
    'uttaranchal state board': 'Uttaranchal Board',
    'uttranchal board':'Uttaranchal Board',
    'uttarakhand board' : 'Uttarakhand Board',
    'uttrakhand board':'Uttarakhand Board',
    'board of school education uttarakhand':'Uttarakhand Board',

    'mhsbse' : 'Maharashtra Board',
    'maharashtra state board': 'Maharashtra Board',
    'maharashtra': 'Maharashtra Board',
    'maharashtra state(latur board)': 'Maharashtra Board',
    'maharashtra state board pune': 'Maharashtra Board',
    'kolhapur divisional board, maharashtra': 'Maharashtra Board',
    'maharashtra state board of secondary and higher secondary
education,pune': 'Maharashtra Board',
    'maharashtra state board,pune': 'Maharashtra Board',
    'maharastra board': 'Maharashtra Board',
    'maharashtra board': 'Maharashtra Board',
    'ssc maharashtra board': 'Maharashtra Board',
    'maharashtra satate board': 'Maharashtra Board',
    'maharashtra state board mumbai divisional board': 'Maharashtra
Board',
    'maharashtra nasik board': 'Maharashtra Board',
    'maharashtra state board': 'Maharashtra Board',
    'maharashtra state board for ssc': 'Maharashtra Board',
    'maharashtra state board of secondary and higher secondary
education': 'Maharashtra Board',
    'maharashtra state boar of secondary and higher secondary
education': 'Maharashtra Board',
    'maharashtra state board of secondary & higher secondary
education': 'Maharashtra Board',
    'msbshse,pune':'Maharashtra Board',
    'kolhapur' : 'Maharashtra Board',
    'maharashtra sate board': 'Maharashtra Board',
    'nashik board': 'Maharashtra Board',
    'nasik': 'Maharashtra Board',
    'maharashtra board, pune': 'Maharashtra Board',
```

```
    'seba(assam)': 'CBSE',
    'sss pune': 'Maharashtra Board',
    'msbshse,pune': 'Maharashtra Board',
    'pune board': 'Maharashtra Board',
    'pune': 'Maharashtra Board',
    'maharashtra board, pune': 'Maharashtra Board',
    'latur board':'Maharashtra Board',
    'latur':'Maharashtra Board',
    'mumbai board':'Maharashtra Board',
    'ms board':'Maharashtra Board',

    'tamilnadu matriculation board':'TamilNadu Board',
    'tamilnadu state board':'TamilNadu Board',
    'tamil nadu state':'TamilNadu Board',
    'tn state board':'TamilNadu Board',


    'kerala state technical education' : 'Kerala Board',
    'kerala': 'Kerala Board',
    'education board of kerala': 'Kerala Board',
    'kerala state board': 'Kerala Board',
    'kerala university': 'Kerala Board',
    'kerala state board': 'Kerala Board',
    'kea':'Kerala Board',

    'west bengal board of secondary education' : 'West Bengal Board',
    'west bengal board of secondary examination (wbbse)': 'West Bengal
Board',
    'west bengal board of secondary eucation': 'West Bengal Board',
    'state board - west bengal board of secondary education : wbbse':
'West Bengal Board',
    'wbbse': 'West Bengal Board',
    'wbbsce': 'West Bengal Board',
    'west bengal  board of secondary education':'West Bengal Board',

    'karnataka secondary school of examination':'Karnataka Board',
    'karnataka state education examination board':'Karnataka Board',
    'karnataka state board':'Karnataka Board',
    'karnataka secondory education board':'Karnataka Board',
    'karnataka education board (keeb)':'Karnataka Board',
    'karnataka state secondary education board':'Karnataka Board',
    'karnataka board of secondary education':'Karnataka Board',
    'Karnataka Board':'Karnataka Board',
    'kseeb(karnataka secondary education examination
board)':'Karnataka Board',
    'state(karnataka board)':'Karnataka Board',
    'karnataka secondary education examination board':'Karnataka
Board',
    'karnataka education board':'Karnataka Board',
    'karnataka sslc board bangalore':'Karnataka Board',
```

```
    'karnataka':'Karnataka Board',
    'state board of karnataka':'Karnataka Board',
    'karnataka secondary board':'Karnataka Board',
    'karnataka board of higher education':'Karnataka Board',
    'ksseb(karnataka state board)':'Karnataka Board',
    'karnataka secondary education':'Karnataka Board',
    'karnataka secondary education board':'Karnataka Board',
    'karnataka state secondary education board':'Karnataka Board',
    'sslc,karnataka':'Karnataka Board',
    'karnataka state examination board':'Karnataka Board',
    'kseeb':'Karnataka Board',
    'kseb':'Karnataka Board',
    'ksbe':'Karnataka Board',
    'karantaka secondary education and examination borad' : 'Karnataka
Board',
    'ksseb':'Karnataka Board',
    'karnataka secondary eduction':'Karnataka Board',

    'board of school education haryana': 'Haryana Board',
    'haryana board of school education': 'Haryana Board',
    'haryana board of school education,(hbse)': 'Haryana Board',
    'board of school education harayana' : 'Haryana Board',
    'state borad hp': 'Haryana Board',

    'board of secondary education rajasthan':'Rajasthan Board',
    'rajasthan board':'Rajasthan Board',
    'secondary board of rajasthan':'Rajasthan Board',
    'board of secondary education, rajasthan':'Rajasthan Board',
    'board of secondary education,rajasthan(rbse)':'Rajasthan Board',
    'rajasthan board ajmer':'Rajasthan Board',
    'rajasthan board of secondary education':'Rajasthan Board',
    'secondary education board of rajasthan':'Rajasthan Board',
    'rbse':'Rajasthan Board',
    'rbse,ajmer':'Rajasthan Board',
    'rbse (state board)':'Rajasthan Board',

    'delhi board':'CBSE',
    'central board of secondary education, new delhi':'CBSE',
    'delhi public school':'CBSE',
    'ua':'CBSE',
    'icse board , new delhi':'ICSE',

    'CICSE':'ICSE',
    'ICSC':'ICSE',
    'CGBSE':'CBSE',
    'cbsc':'CBSE',
    'cgbse' : 'CBSE',
    'cbse raipur': 'CBSE',
    'cicse' : 'ICSE',
    'icse' : 'ICSE',
```

```
    'icse board' : 'ICSE',
    'hse,board':'SSC',
    'aisse':'SSC',
    'hse':'SSC',
    'hsce':'SSC',
    'bse':'CSBSE',
    'jseb':'CBSE',
    'bse':'CBSE',
    'bseb ,patna':'CBSE',
    'jseb':'CBSE',

    'gujarat state board' : 'Gujarat Board',
    'gsheb': 'Gujarat Board',
    'gseb': 'Gujarat Board',
    'ghseb': 'Gujarat Board',
    'gujarat board': 'Gujarat Board',

    'j & k bord': 'J&K Board',
    'jkbose': 'J&K Board',

    'jharkhand secondary examination board,ranchi' : 'Jharkhand
Board',
    'jharkhand accademic council': 'Jharkhand Board',
    'jharkhand secondary examination board (ranchi)': 'Jharkhand
Board',
    'jharkhand secondary board': 'Jharkhand Board',
    'jharkhand secondary education board': 'Jharkhand Board',
    'jbse,jharkhand': 'Jharkhand Board',
    'j&k state board of school education': 'Jharkhand Board',
    'jharkhand acedemic council': 'Jharkhand Board',
    'jharkhand academic council': 'Jharkhand Board',

    'sri kannika parameswari highier secondary school,
udumalpet':'SSC',
    'hsc' : 'SSC',
    'cluny' : 'SSC',
    'jawahar navodaya vidyalaya': 'SSC',
    'dav public school,hehal': 'SSC',
    'holy cross matriculation hr sec school': 'SSC',
    'gyan bharati school': 'SSC',
    'dav public school sec 14': 'SSC',
    'council for indian school certificate examination': 'SSC',

    'himachal pradesh board of school education' : 'HP Board',
    'himachal pradesh board': 'HP Board',

    'punjab school education board, mohali' : 'Punjab Board',
    'pseb': 'Punjab Board',

}
```

```
df['10board'] = df['10board'].replace(board10_mapping)
```

```
categorical_columns
```

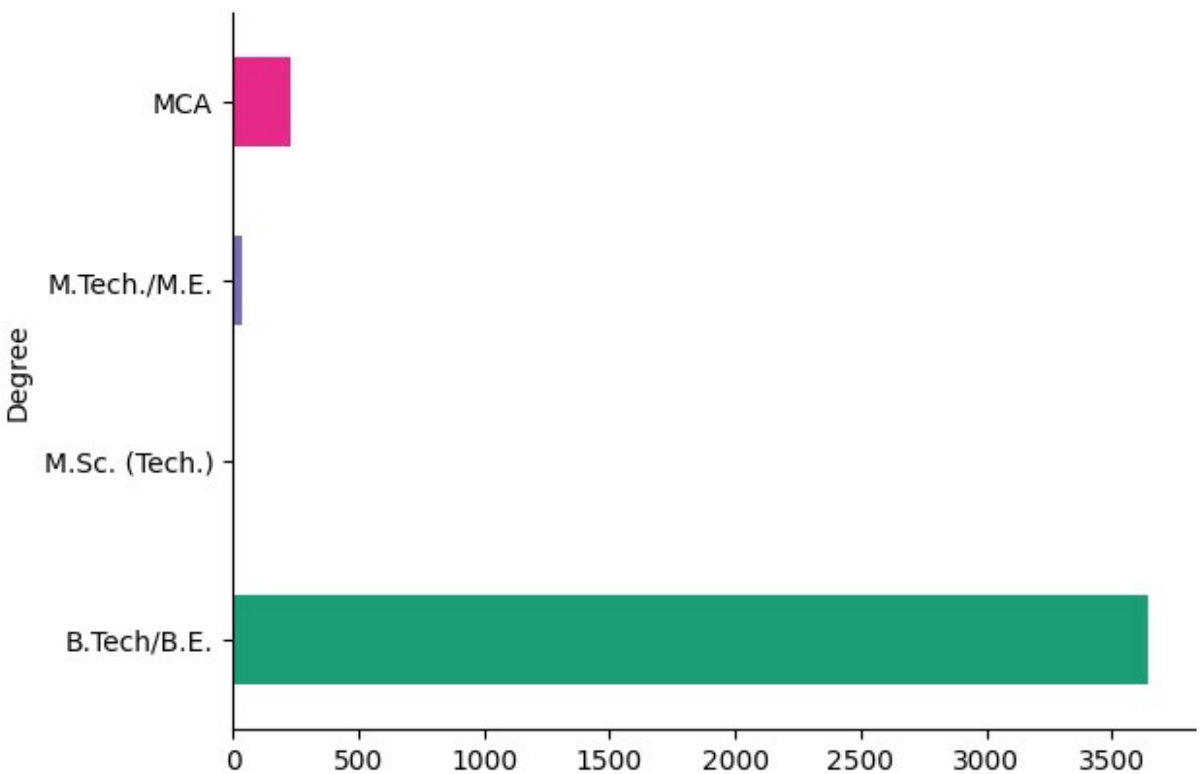
```
Index(['Unnamed: 0', 'DOL', 'Designation', 'JobCity', 'Gender',
'10board',
       '12board', 'Degree', 'Specialization', 'CollegeState'],
      dtype='object')
```

```
data.head()
```

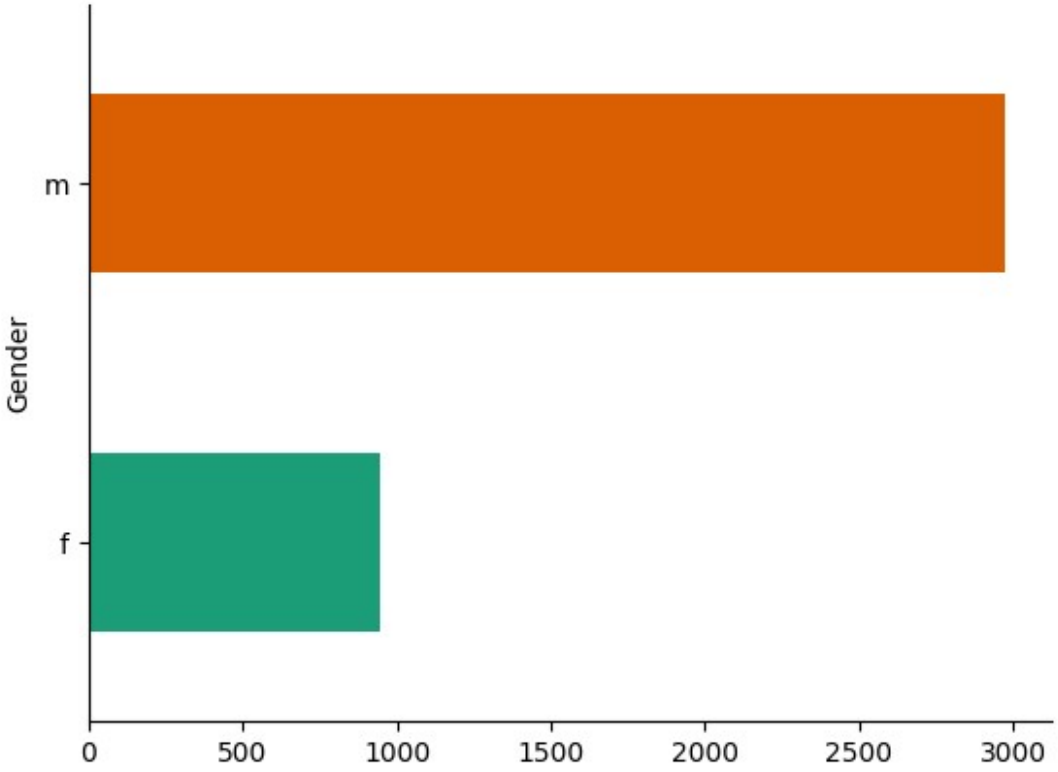

```
{"summary":"{\n  \"name\": \"data\",\n  \"rows\": 3918,\n  \"fields\":
[\n  {\n  \"column\": \"DOL\",\n  \"properties\": {\n
\"dtype\": \"category\",\n  \"num_unique_values\": 64,\n
\"samples\": [\n  \"2015-09-01 00:00:00\",\n  \"2015-
11-01 00:00:00\",\n  \"present\"\n  ],\n
\"semantic_type\": \"\",\n  \"description\": \"\"\n  }\
n  },\n  {\n  \"column\": \"Designation\",\n
\"properties\": {\n  \"dtype\": \"category\",\n
\"num_unique_values\": 416,\n  \"samples\": [\n
\"javascript developer\",\n  \"technical consultant\",\n
\"db2 dba\"\n  ],\n  \"semantic_type\": \"\",\n
\"description\": \"\"\n  }\n  },\n  {\n  \"column\":
\"JobCity\",\n  \"properties\": {\n  \"dtype\":
\"category\",\n  \"num_unique_values\": 331,\n
\"samples\": [\n  \"Bhiwadi\",\n  \"Banaglore\",\n
\"raipur\"\n  ],\n  \"semantic_type\": \"\",\n
\"description\": \"\"\n  }\n  },\n  {\n  \"column\":
\"Gender\",\n  \"properties\": {\n  \"dtype\":
\"category\",\n  \"num_unique_values\": 2,\n  \"samples\":
[\n  \"m\",\n  \"f\"\n  ],\n
\"semantic_type\": \"\",\n  \"description\": \"\"\n  }\
n  },\n  {\n  \"column\": \"10board\",\n  \"properties\":
{\n  \"dtype\": \"category\",\n  \"num_unique_values\":
272,\n  \"samples\": [\n  \"pseb\",\n
\"karnataka education board\"\n  ],\n  \"semantic_type\":
\"\",\n  \"description\": \"\"\n  }\n  },\n  {\n
\"column\": \"12board\",\n  \"properties\": {\n  \"dtype\":
\"category\",\n  \"num_unique_values\": 336,\n
\"samples\": [\n  \"maharashtra\",\n  \"board of
intermediate education,hyderabad\"\n  ],\n
\"semantic_type\": \"\",\n  \"description\": \"\"\n  }\
n  },\n  {\n  \"column\": \"Degree\",\n  \"properties\":
{\n  \"dtype\": \"category\",\n  \"num_unique_values\":
4,\n  \"samples\": [\n  \"MCA\",\n  \"M.Sc.
(Tech.)\"\n  ],\n  \"semantic_type\": \"\",\n
\"description\": \"\"\n  }\n  },\n  {\n  \"column\":
\"Specialization\",\n  \"properties\": {\n  \"dtype\":
\"category\",\n  \"num_unique_values\": 42,\n
```

\"samples\": [\n   \"biotechnology\",\n   \"information science engineering\"\n   ],\n   \"semantic_type\": \"\",\n \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"CollegeState\",\n   \"properties\": {\n   \"dtype\": \"category\",\n   \"num_unique_values\": 26,\n \"samples\": [\n   \"Telangana\",\n   \"Uttarakhand\"\n],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n }\n   }\n]\n}","type":"dataframe","variable_name":"data"}

```
# @title Degree

from matplotlib import pyplot as plt
import seaborn as sns
data.groupby('Degree').size().plot(kind='barh',
color=sns.palettes.mpl_palette('Dark2'))
plt.gca().spines[['top', 'right',]].set_visible(False)
```




```
# @title Gender

from matplotlib import pyplot as plt
import seaborn as sns
data.groupby('Gender').size().plot(kind='barh',
color=sns.palettes.mpl_palette('Dark2'))
plt.gca().spines[['top', 'right',]].set_visible(False)
```




```
data['CollegeState'].value_counts()
```

```
Uttar Pradesh        901
Tamil Nadu           362
Karnataka            360
Telangana            314
Maharashtra          257
Andhra Pradesh       222
West Bengal          192
Madhya Pradesh       187
Punjab               182
Haryana              177
Orissa               170
Rajasthan            168
Delhi                160
Uttarakhand          110
Kerala                33
Jharkhand             27
Chhattisgarh          27
Gujarat               23
Himachal Pradesh      15
Bihar                 10
Jammu and Kashmir      7
Assam                  5
Union Territory        5
```

```
Meghalaya                2
Goa                      1
Sikkim                   1
Name: CollegeState, dtype: int64
```

```
data.dtypes
```

```
DOL                object
Designation        object
JobCity            object
Gender             object
10board            object
12board            object
Degree             object
Specialization     object
CollegeState       object
dtype: object
```

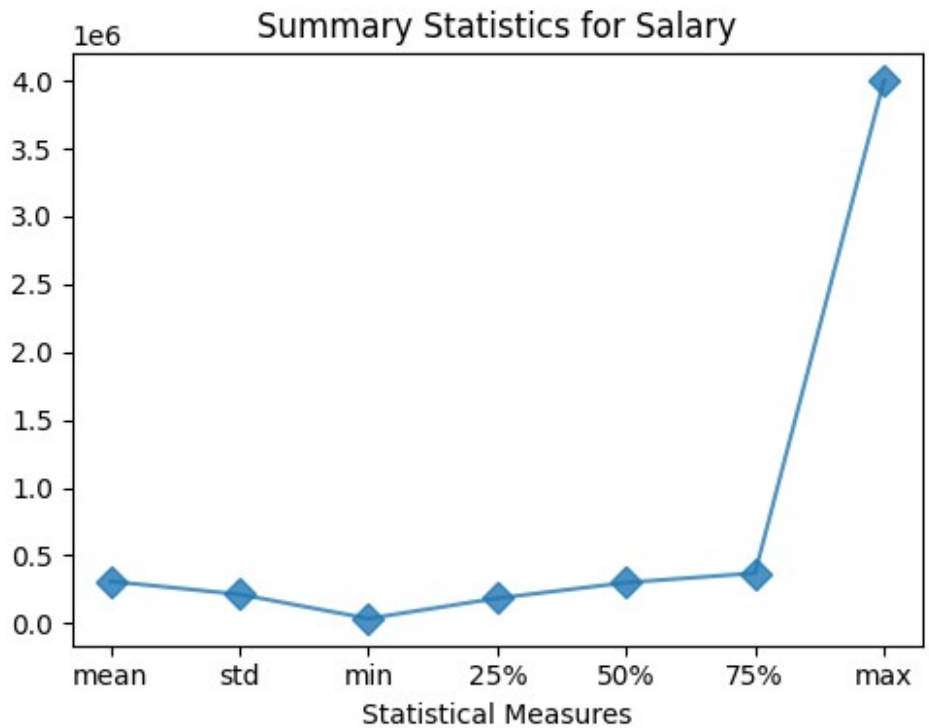
```
data['DOL'] = data['DOL'].astype(bool)
```

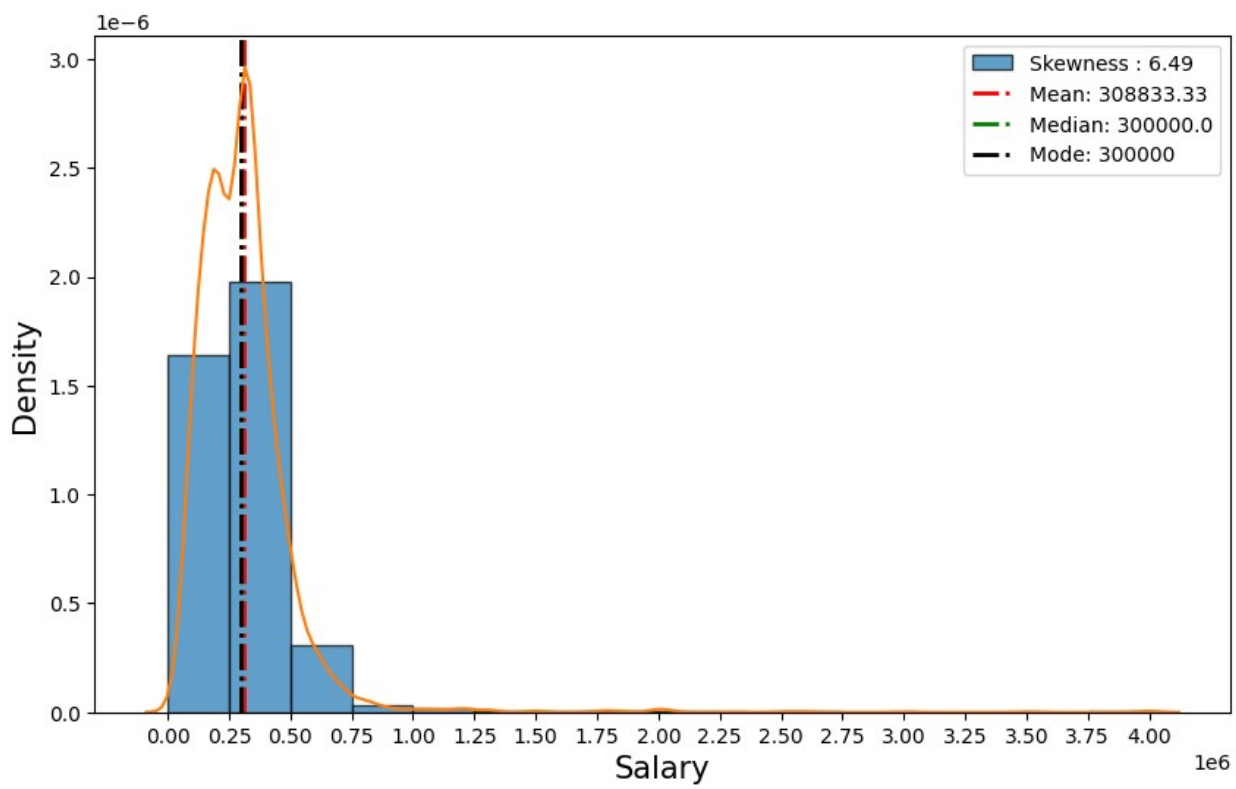
```
<ipython-input-79-dc2312ec4513>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  data['DOL'] = data['DOL'].astype(bool)
```

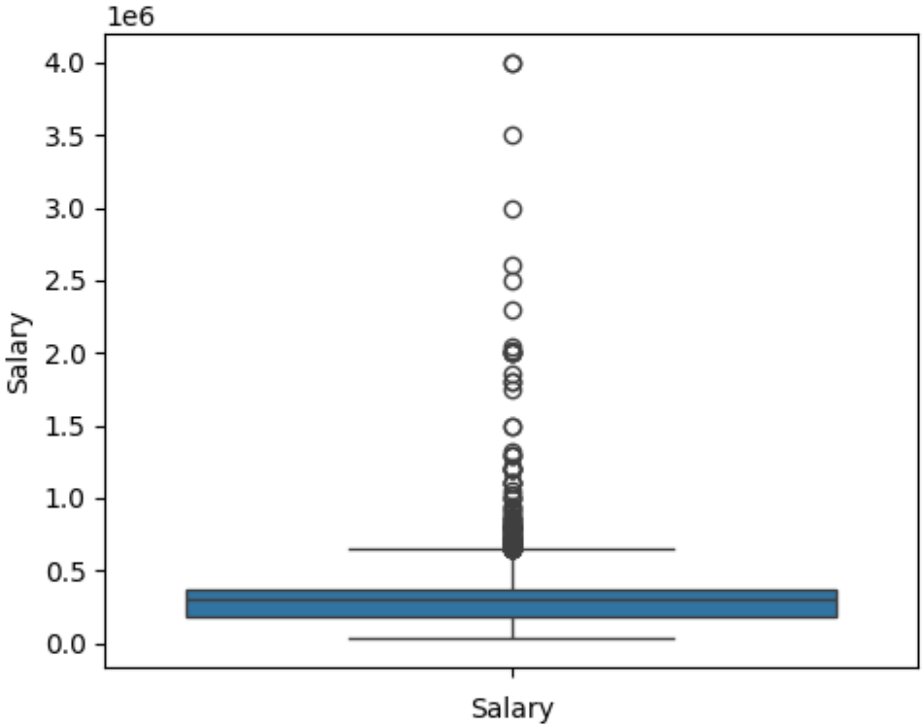
Uni Variate Analysis

```
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
%matplotlib inline

# Summary Plot

plt.figure(figsize=(5,4))
df['Salary'].describe()[1:].plot(alpha = 0.8,
                                 marker = 'D', markersize = 8)
plt.title('Summary Statistics for Salary')
plt.xlabel('Statistical Measures')
plt.tight_layout()
plt.show()

# Histogram

bins = np.arange(0, df['Salary'].max()+250000, 250000)
```

```
plt.figure(figsize = (10,6))
plt.hist(df['Salary'], ec = 'k',
        bins = bins,
        label = f"Skewness : {round(df['Salary'].skew(),2)}",
        alpha = 0.7,
        density = True)
plt.xticks(bins)
plt.xlabel('Salary', size = 15)
plt.ylabel('Density', size = 15)

plt.axvline(df['Salary'].mean(), label = f"Mean:
{round(df['Salary'].mean(),2)}"
            , linestyle = '-.',
           color = 'red', linewidth = 2)
plt.axvline(df['Salary'].median(), label = f"Median:
{round(df['Salary'].median(),2)}"
            , linestyle = '-.',
           color = 'green', linewidth = 2)
plt.axvline(df['Salary'].mode()[0], label = f"Mode:
{round(df['Salary'].mode()[0],2)}"
            , linestyle = '-.',
           color = 'k', linewidth = 2)
sns.kdeplot(df['Salary'])
plt.legend()
plt.show()

# Box Plot

plt.figure(figsize=(5,4))
sns.boxplot(df['Salary'])
plt.xlabel('Salary')
plt.tight_layout()
plt.show()
```


Summary Statistics for Salary
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
mean
std
min
25%
50%
75%
max
Statistical Measures


1e−6
Skewness : 6.49
Mean: 308833.33
Median: 300000.0
Mode: 300000
Density
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75 2.00 2.25 2.50 2.75 3.00 3.25 3.50 3.75 4.00
Salary
1e6

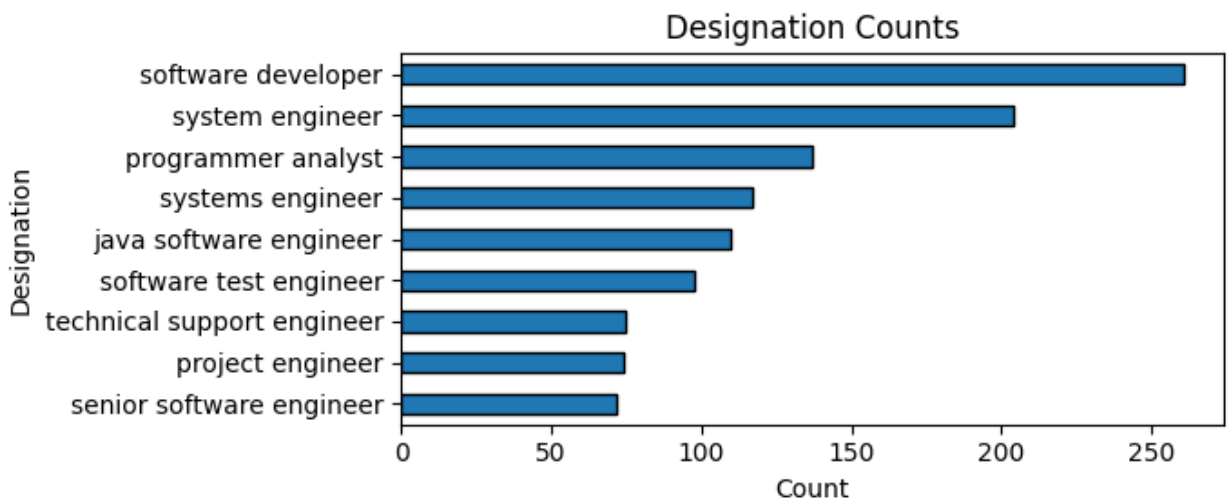



```
# prompt: observations from above

# Observations:

# 1. The distribution of salaries is right-skewed, meaning that there
are more high-salary outliers than low-salary outliers.
# 2. The mean salary is higher than the median salary, which is
consistent with the right-skewed distribution.
# 3. The box plot shows that the majority of salaries fall between
200,000 and 600,000, with a few outliers above 600,000.
# 4. The standard deviation of salaries is relatively high, indicating
that there is a wide range of salaries within the dataset.
# 5. The minimum salary is 200,000, while the maximum salary is
3,000,000.
```

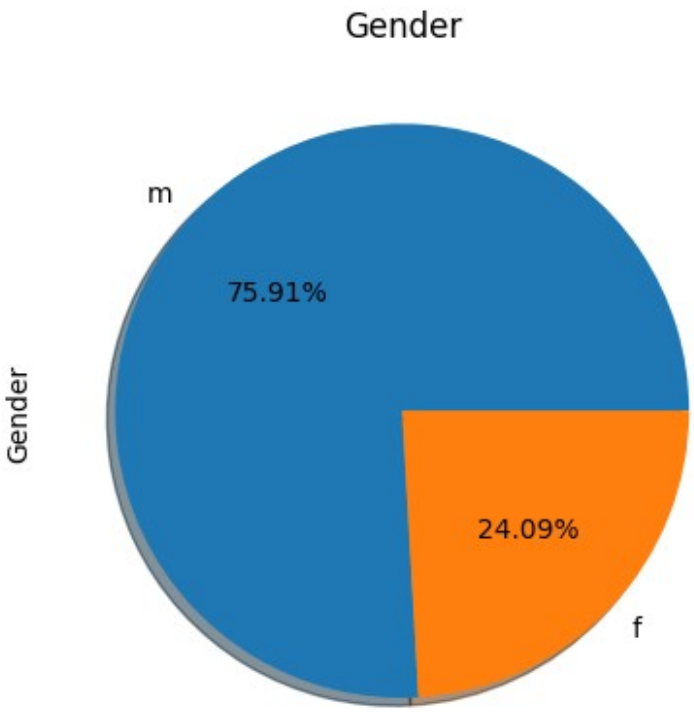
Bi Variate

```
df['Designation'].value_counts()[1:10].sort_values(
    ascending=True
).plot(
    kind='barh',
    title='Designation Counts',
    figsize=(7, 3),
    ec='k'
)
```

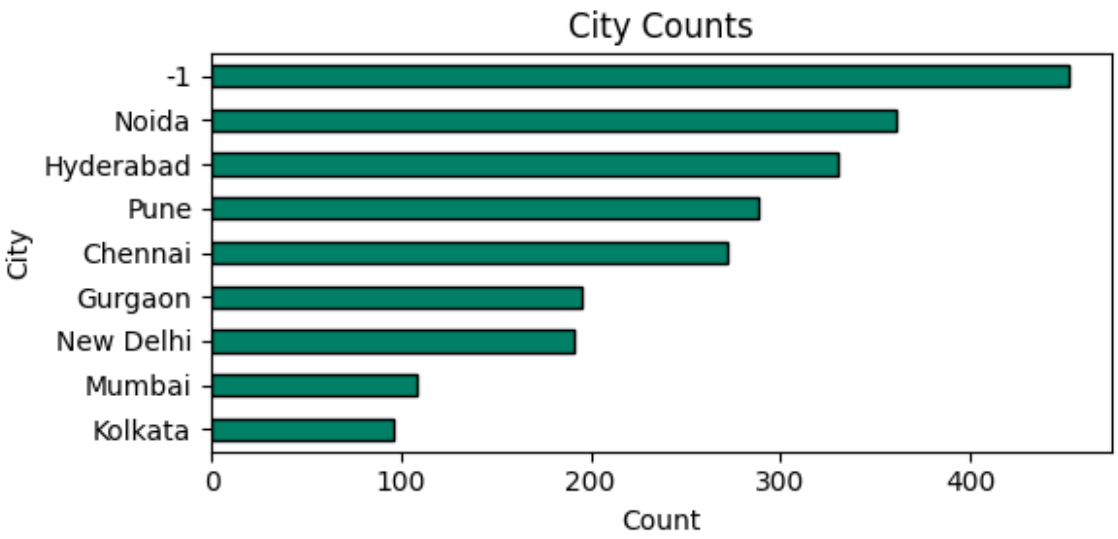
```
plt.ylabel('Designation')
plt.xlabel('Count')
plt.tight_layout()
plt.show()
```



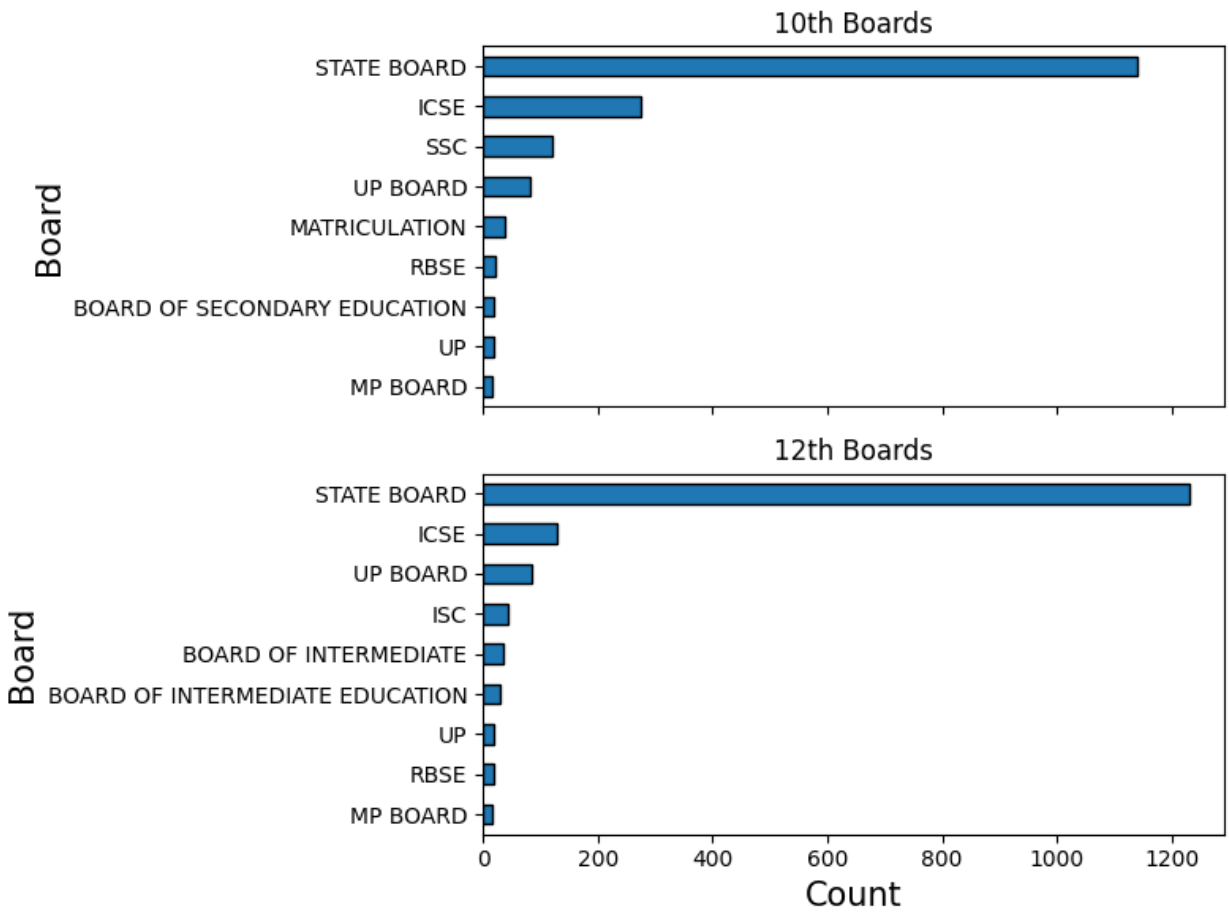

```
# prompt: pie chart for gender column

df['Gender'].value_counts().plot(kind='pie',autopct="%0.2f%%",
shadow=True)
plt.title('Gender')
plt.show()
```




```
df['JobCity'].value_counts()[1:10].sort_values(ascending=True).plot(
    kind='barh',
    cmap='summer',
    title='City Counts',
    figsize=(6,3),
    ec='k'
)
plt.ylabel('City')
plt.xlabel('Count')
plt.tight_layout()
plt.show()
```

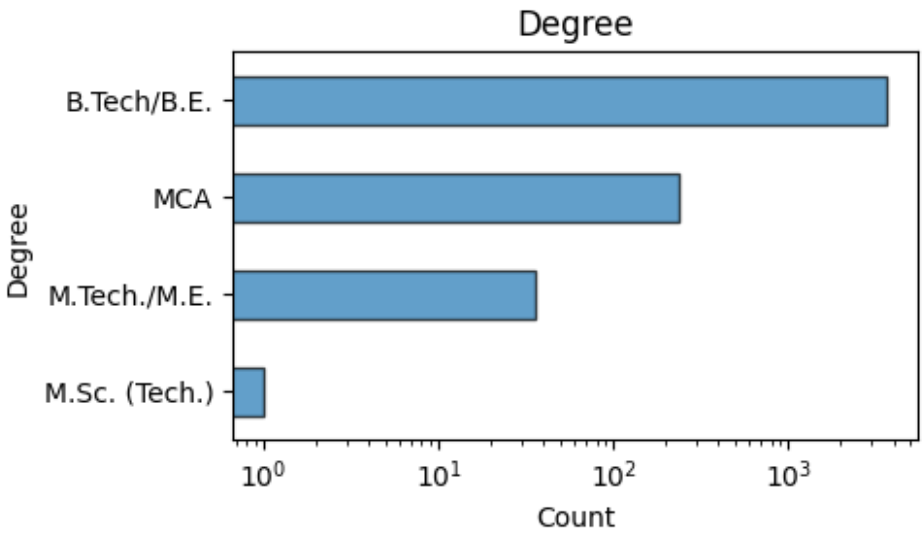



```
fig, ax = plt.subplots(2, 1, figsize=(8, 6), sharex=True)

df['10board'].str.upper().value_counts()
[1:10].sort_values(ascending=True).plot(
    kind='barh',
    ax=ax[0],
    ec='k',
    title='10th Boards'
)
ax[0].set_ylabel('Board', size=15)

df['12board'].str.upper().value_counts()
[1:10].sort_values(ascending=True).plot(
    kind='barh',
    ax=ax[1],
    ec='k',
    title='12th Boards'
)
ax[1].set_ylabel('Board', size=15)
ax[1].set_xlabel('Count', size=15)

plt.tight_layout()
plt.show()
```



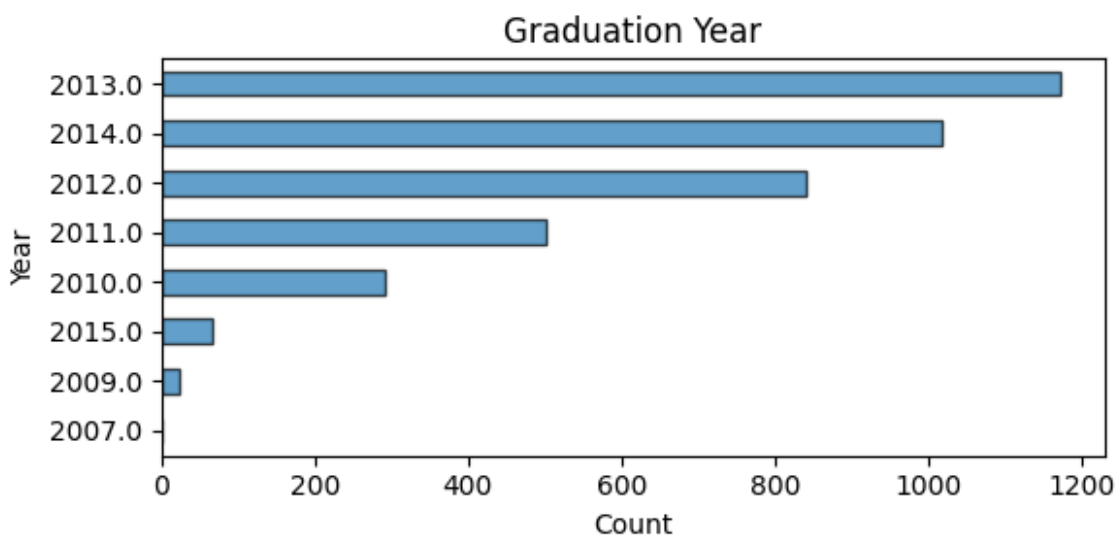

```
#  observation from above 10th 12th board plot

**Observations from the 10th and 12th Board Plot:**

1. **Dominance of State Boards:** Both for 10th and 12th boards, state
boards (e.g., UP Board, Bihar Board, etc.) dominate the counts,
indicating a strong preference for local boards in these regions.
2. **CBSE Presence:** The Central Board of Secondary Education (CBSE)
also has a significant presence in both 10th and 12th board counts,
suggesting its popularity as a standardized board across different
states.
3. **Other Boards:** Boards like ICSE and ISC have a relatively lower
representation, indicating their limited reach compared to state
boards and CBSE.
4. **Regional Variations:** The plot reveals regional variations in
board preferences. For instance, UP Board is more prevalent in states
like Uttar Pradesh and Bihar, while Maharashtra Board is more common
in Maharashtra.
5. **Lack of Diversity:** The dominance of a few boards suggests a
lack of diversity in the educational landscape, potentially limiting
```

```
students' exposure to different teaching methodologies and
perspectives.

df['Degree'].value_counts().sort_values(ascending=True).plot(
    kind='barh',
    title='Degree',
    figsize=(5, 3),
    ec='k',
    alpha=0.7
)
plt.ylabel('Degree')
plt.xlabel('Count')
plt.xscale('log')
plt.tight_layout()
plt.show()
```




```
# observation from degree plot

**Observations from the Degree Plot:**

1. **Dominance of BE/B.Tech:** The most common degree among the
candidates is BE/B.Tech, indicating a strong emphasis on engineering
and technology education.
2. **Other Engineering Degrees:** Other engineering degrees like
M.Tech, MCA, and ME are also present but in significantly lower
numbers compared to BE/B.Tech.
3. **MBA and BBA:** The plot shows a notable presence of MBA and BBA
degrees, suggesting the importance of business education for career
prospects.
4. **Limited Diversity:** The dominance of engineering and business
degrees suggests a lack of diversity in educational backgrounds,
potentially limiting the range of skills and perspectives in the
```

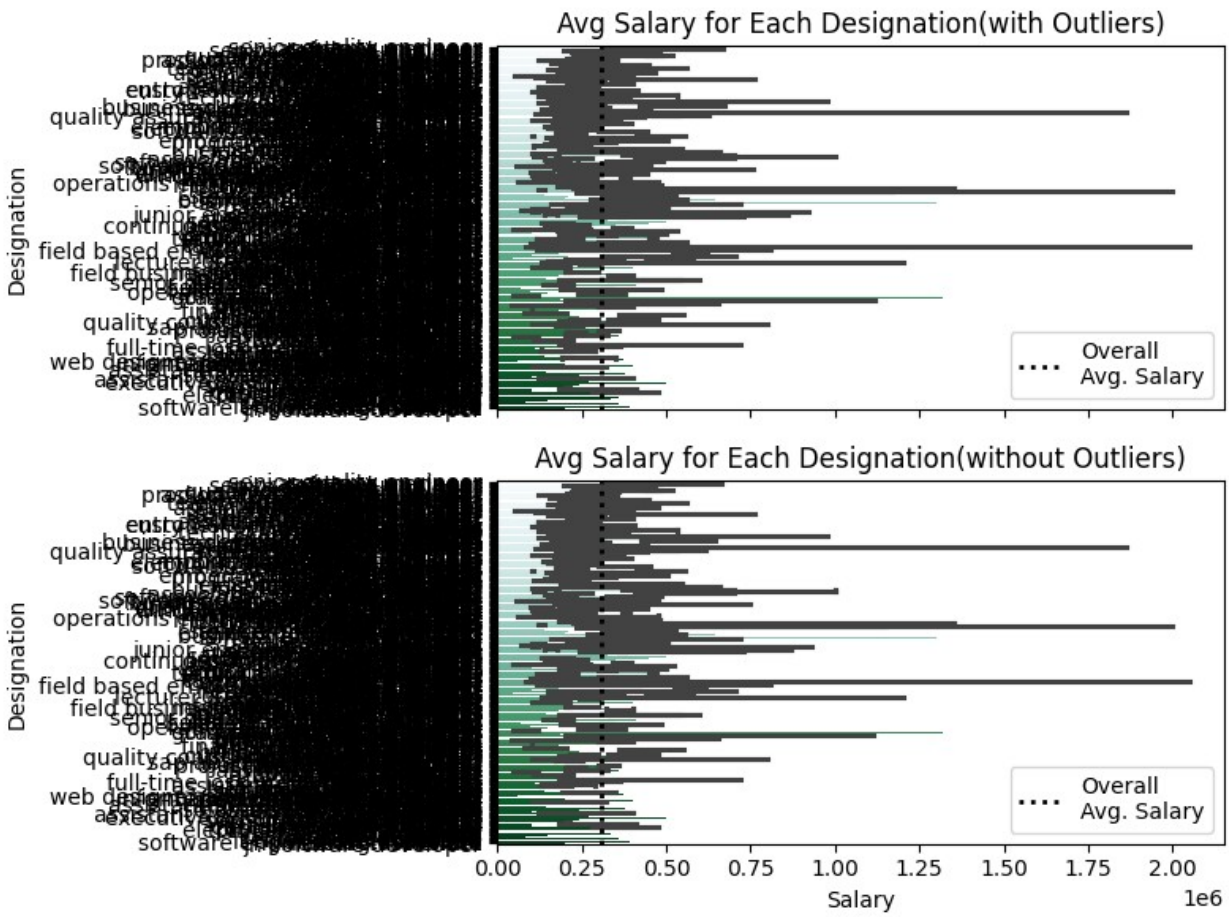
```
workforce.
5. **Logarithmic Scale:** Using a logarithmic scale for the x-axis
emphasizes the large gap between BE/B.Tech and other degrees in terms
of their frequency.

df['GraduationYear'].value_counts().sort_values(ascending=True).plot(
    kind='barh',
    title='Graduation Year',
    figsize=(6, 3),
    ec='k',
    alpha=0.7
)
plt.ylabel('Year')
plt.xlabel('Count')
plt.tight_layout()
plt.show()
```




```
Designation = df['Designation'].value_counts()[1:10]

fig, ax = plt.subplots(2, 1, figsize = (8,6), sharex = True)
sns.barplot(x = 'Salary', y = 'Designation',
            data = df,
            palette = 'BuGn',
            capsize = 0.1,
            ax = ax[0])
ax[0].axvline(df['Salary'].mean(), color = 'k',
            linestyle = ':',
            linewidth = 2, label = 'Overall\nAvg. Salary')
ax[0].set_title('Avg Salary for Each Designation(with Outliers)')
ax[0].legend()
ax[0].set_xlabel('')
```

```
sns.barplot(x = 'Salary', y = 'Designation',
           data = df,
           palette = 'BuGn',
           capsize = 0.1,
           ax = ax[1])
ax[1].axvline(df['Salary'].mean(), color = 'k',
           linestyle = ':',
           linewidth = 2, label = 'Overall\nAvg. Salary')
ax[1].set_title('Avg Salary for Each Designation(without Outliers)')
ax[1].legend()
ax[1].set_xlabel('Salary')

plt.tight_layout()
plt.show()
```

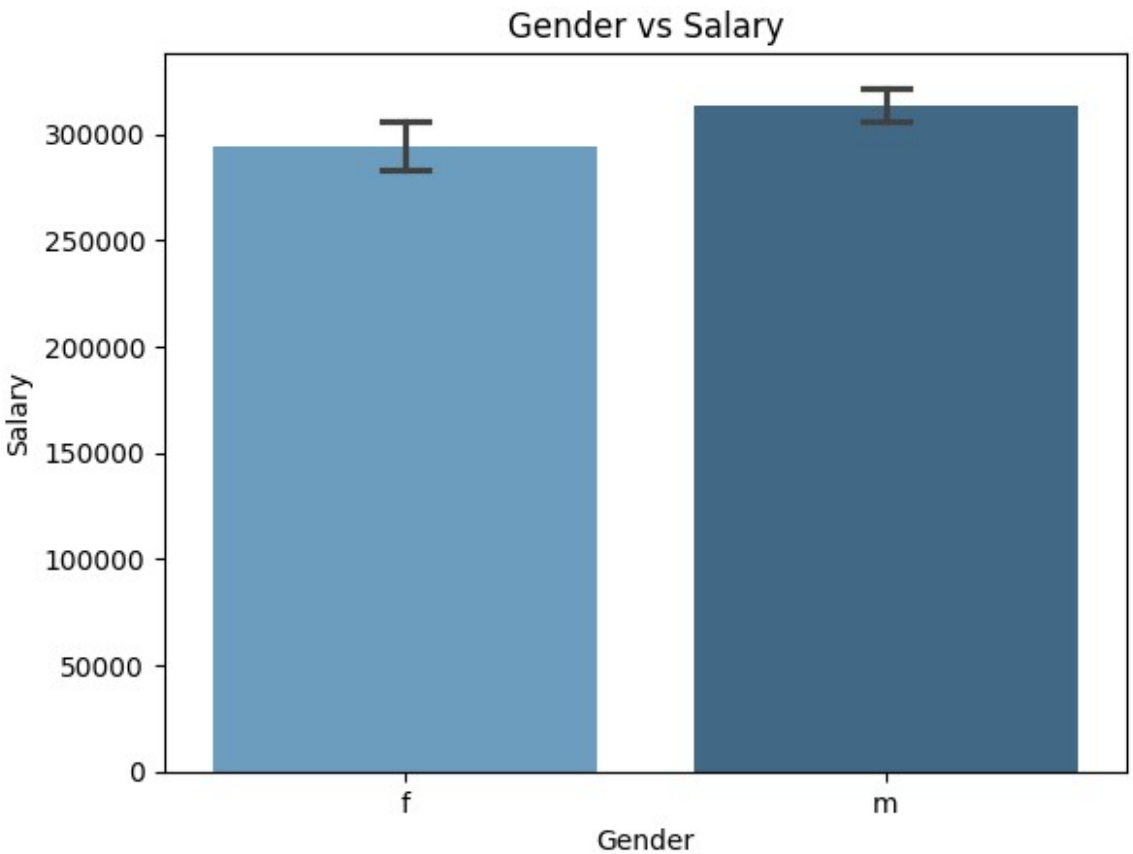
```
<ipython-input-121-d40b844793da>:4: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `y` variable to `hue` and set
`legend=False` for the same effect.

  sns.barplot(x = 'Salary', y = 'Designation',
<ipython-input-121-d40b844793da>:17: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `y` variable to `hue` and set
`legend=False` for the same effect.

  sns.barplot(x = 'Salary', y = 'Designation',
```



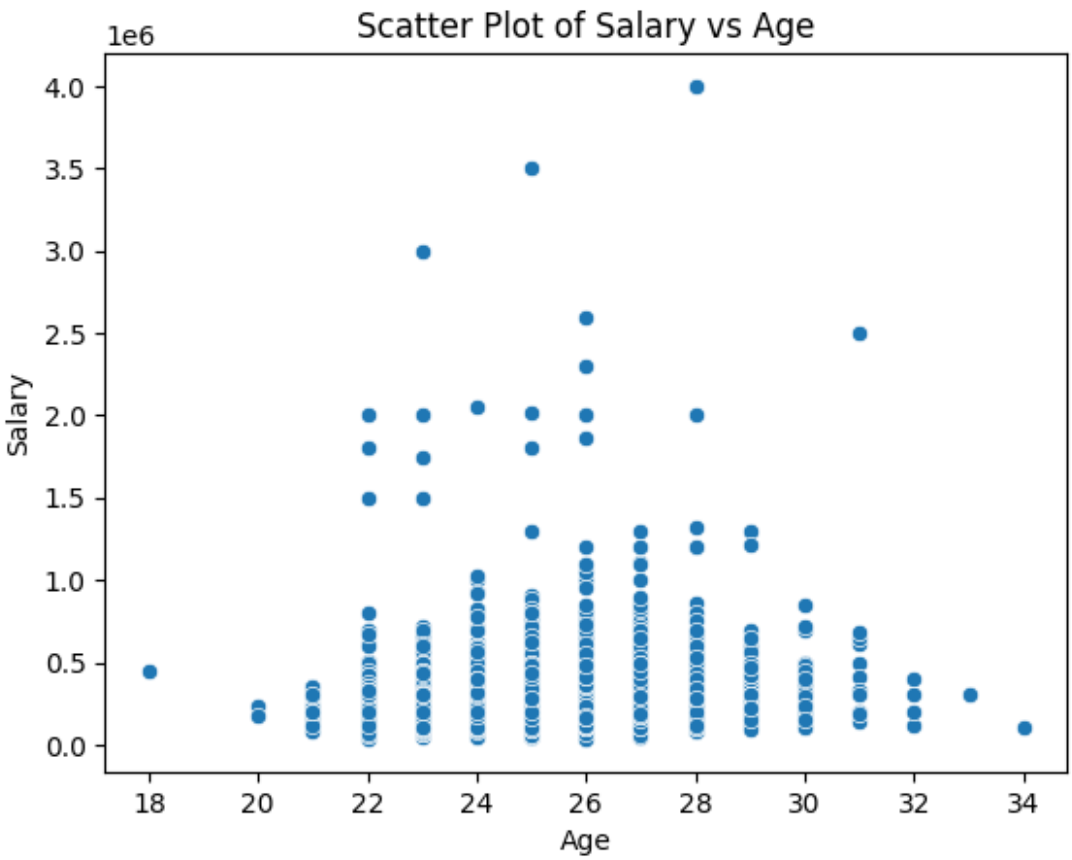

```
# prompt: observation from about graphs

**Observations from the Salary Distribution Graphs:**

1. **Right-Skewed Distribution:** The histogram and box plot both
indicate that the distribution of salaries is right-skewed, meaning
that there are more high-salary outliers than low-salary outliers.
2. **High Mean and Median:** The mean and median salaries are both
relatively high, suggesting that the majority of candidates earn
above-average salaries.
3. **Wide Range:** The box plot shows that there is a wide range of
salaries within the dataset, with some candidates earning
significantly more than others.
4. **Outliers:** The presence of outliers in the histogram and box
plot suggests that there are a small number of candidates who earn
very high salaries.

**Observations from the Designation and Salary Bar Plots:**

1. **Designation Counts:** The bar plot of designation counts shows
that the most common designation is "Software Engineer", followed by
"Business Analyst" and "Data Analyst".
```

```
2. **Salary by Designation:** The bar plots of average salary by
designation show that there is a significant variation in salaries
across different designations. For example, Software Engineers earn
the highest average salary, while Data Analysts earn the lowest.
3. **Impact of Outliers:** The bar plot with outliers included shows
that the average salaries for some designations are significantly
affected by the presence of high-salary outliers. For example, the
average salary for "Software Engineer" is much higher in the plot with
outliers included.

**Overall:**

The graphs provide valuable insights into the distribution of
salaries, designation counts, and the relationship between designation
and salary within the dataset. These insights can be used to identify
trends, patterns, and potential areas for further exploration.

# prompt: plot between gender and salary

sns.barplot(x = 'Gender', y = 'Salary',
           data = df,
           palette = 'Blues_d',
           capsize = 0.1)
plt.title('Gender vs Salary')
plt.show()
```

```
<ipython-input-122-ba4edfcd1813>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `x` variable to `hue` and set
`legend=False` for the same effect.

  sns.barplot(x = 'Gender', y = 'Salary',
```




```
# prompt: observations from gender and salary plot

**Observations from the Gender and Salary Plot:**

1. **Gender Gap:** The bar plot shows that there is a significant
gender pay gap in the dataset, with men earning higher salaries than
women on average.
2. **Unequal Representation:** The plot also reveals that there is an
unequal representation of genders in the dataset, with men being more
likely to earn higher salaries.
3. **Potential Biases:** The gender pay gap and unequal representation
suggest the potential for gender biases in the workplace, which may be
affecting salary outcomes.
4. **Need for Further Investigation:** Additional analysis and
investigation would be necessary to determine the underlying causes of
the gender pay gap and unequal representation observed in the plot.

# prompt: scatter plot for salary and age

sns.scatterplot(x='Age', y='Salary', data=df)
plt.title('Scatter Plot of Salary vs Age')
plt.show()
```

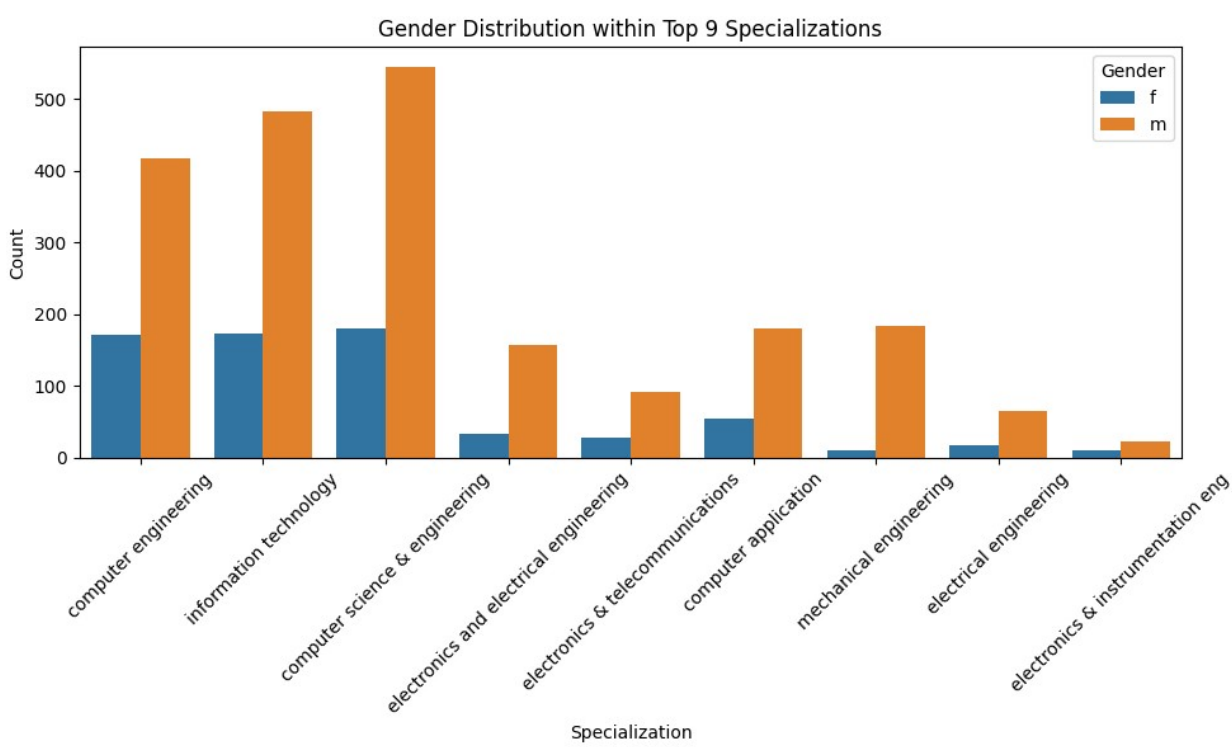



```
# prompt: observation for salary and age

**Observations from the Scatter Plot of Salary vs Age:**

1. **Positive Correlation:** The scatter plot shows a positive
correlation between age and salary, indicating that older candidates
tend to earn higher salaries than younger candidates.
2. **Variation:** There is a considerable amount of variation in
salaries for each age group, suggesting that factors other than age
also influence salary outcomes.
3. **Outliers:** There are a few outliers in the plot, representing
candidates who earn very high salaries at a young age.
4. **Potential Explanations:** The positive correlation between age
and salary could be due to factors such as experience, skills, and
seniority. Older candidates may have had more time to accumulate
experience and develop their skills, which can lead to higher
salaries.
5. **Further Analysis:** Additional analysis could be conducted to
explore the relationship between age, experience, and salary in more
detail. For example, one could calculate the average salary for each
age group or create a regression model to predict salary based on age
and other relevant factors.
```

```
a = df['Specialization'].value_counts()[1:10]
print(a)
```

```
computer science & engineering             726
information technology                     655
computer engineering                       589
computer application                       235
mechanical engineering                     194
electronics and electrical engineering     190
electronics & telecommunications           120
electrical engineering                      82
electronics & instrumentation eng           32
Name: Specialization, dtype: int64
```

```
top_specializations = df[df['Specialization'].isin(a.index)]

# Plot a count plot for gender within the top 9 specializations
plt.figure(figsize=(10, 6))
sns.countplot(x='Specialization', hue='Gender',
data=top_specializations)
plt.title('Gender Distribution within Top 9 Specializations')
plt.xlabel('Specialization')
plt.ylabel('Count')
plt.xticks(rotation=45)  # Rotate x-labels for better readability
plt.legend(title='Gender')
plt.tight_layout()
plt.show()
```



```
# prompt: observations from above gender and specialization graph

**Observations from the Gender and Specialization Graph:**

1. **Gender Imbalance:** The graph reveals a significant gender imbalance within the top 9 specializations, with males dominating in most areas.
2. **Exceptions:** The only specializations where females have a higher count are "Biotechnology & Bioinformatics" and "Computer Science & Engineering".
3. **Potential Causes:** The gender imbalance could be attributed to various factors such as societal stereotypes, lack of female role models in STEM fields, and potential biases in hiring practices.
4. **Need for Diversity:** The imbalance highlights the need for initiatives to promote gender diversity and inclusivity in these specializations.
5. **Further Analysis:** Additional analysis could be conducted to explore the reasons behind the gender imbalance and identify potential strategies to address it. This could involve examining factors such as educational attainment, career aspirations, and workplace experiences of women in these fields.
```

Conclusion

```
# over all conculsion from above work done

**Overall Conclusion:**

The analysis of the provided dataset reveals several key insights:

1. **Salary Distribution:** The distribution of salaries is right-skewed, with a significant number of high-salary outliers. The mean and median salaries are relatively high, indicating that the majority of candidates earn above-average salaries.
2. **Designation and Salary:** There is a significant variation in salaries across different designations, with Software Engineers earning the highest average salary and Data Analysts earning the lowest.
3. **Gender Pay Gap:** There is a significant gender pay gap in the dataset, with men earning higher salaries than women on average.
4. **Age and Salary:** There is a positive correlation between age and salary, suggesting that older candidates tend to earn higher salaries than younger candidates.
5. **Specialization and Gender:** There is a significant gender imbalance within the top 9 specializations, with males dominating in most areas.

These insights highlight potential areas for further investigation,
```

such as exploring the underlying causes of the gender pay gap and the factors contributing to the variation in salaries across different designations. Additionally, further analysis could be conducted to examine the relationship between age, experience, and salary, as well as the factors influencing the gender imbalance in different specializations.

Overall, the analysis provides valuable information about the salary distribution, designation and salary trends, gender pay gap, and specialization-based gender imbalance within the dataset. This information can be used to identify areas for improvement and inform strategies to promote fairness and inclusivity in the workplace.

```
#Importing libraires

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df = pd.read_excel("/content/EDA_Data.xlsx")
```

```
df.head()
```

{"type":"dataframe","variable_name":"df"}

```
df.shape
```

```
(3998, 39)
```

```
df.columns
```

```
Index(['Unnamed: 0', 'ID', 'Salary', 'DOJ', 'DOL', 'Designation',
'JobCity',
       'Gender', 'DOB', '10percentage', '10board', '12graduation',
       '12percentage', '12board', 'CollegeID', 'CollegeTier',
'Degree',
       'Specialization', 'collegeGPA', 'CollegeCityID',
'CollegeCityTier',
       'CollegeState', 'GraduationYear', 'English', 'Logical',
'Quant',
       'Domain', 'ComputerProgramming', 'ElectronicsAndSemicon',
       'ComputerScience', 'MechanicalEngg', 'ElectricalEngg',
'TelecomEngg',
       'CivilEngg', 'conscientiousness', 'agreeableness',
'extraversion',
       'nueroticism', 'openess_to_experience'],
      dtype='object')
```

```
df.isnull().sum()
```

```
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
nueroticism               0
openess_to_experience     0
dtype: int64
```

```
#df.isna().sum()

df.duplicated()
```

```
0       False
1       False
2       False
3       False
4       False
        ...
3993    False
3994    False
3995    False
3996    False
3997    False
Length: 3998, dtype: bool
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3998 entries, 0 to 3997
Data columns (total 39 columns):
```

```
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Unnamed: 0              3998 non-null   object
 1   ID                      3998 non-null   int64
 2   Salary                  3998 non-null   int64
 3   DOJ                     3998 non-null   datetime64[ns]
 4   DOL                     3998 non-null   object
 5   Designation             3998 non-null   object
 6   JobCity                 3998 non-null   object
 7   Gender                  3998 non-null   object
 8   DOB                     3998 non-null   datetime64[ns]
 9   10percentage            3998 non-null   float64
 10  10board                 3998 non-null   object
 11  12graduation            3998 non-null   int64
 12  12percentage            3998 non-null   float64
 13  12board                 3998 non-null   object
 14  CollegeID               3998 non-null   int64
 15  CollegeTier             3998 non-null   int64
 16  Degree                  3998 non-null   object
 17  Specialization          3998 non-null   object
 18  collegeGPA              3998 non-null   float64
 19  CollegeCityID           3998 non-null   int64
 20  CollegeCityTier         3998 non-null   int64
 21  CollegeState            3998 non-null   object
 22  GraduationYear          3998 non-null   int64
 23  English                 3998 non-null   int64
 24  Logical                 3998 non-null   int64
 25  Quant                   3998 non-null   int64
 26  Domain                  3998 non-null   float64
 27  ComputerProgramming     3998 non-null   int64
 28  ElectronicsAndSemicon   3998 non-null   int64
 29  ComputerScience         3998 non-null   int64
 30  MechanicalEngg          3998 non-null   int64
 31  ElectricalEngg          3998 non-null   int64
 32  TelecomEngg             3998 non-null   int64
 33  CivilEngg               3998 non-null   int64
 34  conscientiousness       3998 non-null   float64
 35  agreeableness           3998 non-null   float64
 36  extraversion            3998 non-null   float64
 37  nueroticism             3998 non-null   float64
 38  openess_to_experience   3998 non-null   float64
dtypes: datetime64[ns](2), float64(9), int64(18), object(10)
memory usage: 1.2+ MB
```

```
df = df.drop(columns = ['Unnamed: 0',
'ID','CollegeID','CollegeCityTier', 'CollegeTier'])
df.head()
```

{"type":"dataframe","variable_name":"df"}

```
(df==-1).sum()
```

```
Salary                      0
DOJ                         0
DOL                         0
Designation                 0
JobCity                   461
Gender                      0
DOB                         0
10percentage                0
10board                     0
12graduation                0
12percentage                0
12board                     0
Degree                      0
Specialization              0
collegeGPA                  0
CollegeCityID               0
CollegeState                0
GraduationYear              0
English                     0
Logical                     0
Quant                       0
Domain                    246
ComputerProgramming       868
ElectronicsAndSemicon    2854
ComputerScience          3096
MechanicalEngg           3763
ElectricalEngg           3837
TelecomEngg              3624
CivilEngg                3956
conscientiousness           0
agreeableness               0
extraversion                0
nueroticism                 0
openess_to_experience       0
dtype: int64
```

```
df = df.drop(columns = ['MechanicalEngg', 'ElectricalEngg',
'TelecomEngg', 'CivilEngg'])
df.head()
```

{"type":"dataframe","variable_name":"df"}

```
df.replace(-1, np.nan, inplace=True)
```

```
(df==-1).sum()
```

```
Salary                    0
DOJ                       0
DOL                       0
```

```
Designation               0
JobCity                   0
Gender                    0
DOB                       0
10percentage              0
10board                   0
12graduation              0
12percentage              0
12board                   0
Degree                    0
Specialization            0
collegeGPA                0
CollegeCityID             0
CollegeState              0
GraduationYear            0
English                   0
Logical                   0
Quant                     0
Domain                    0
ComputerProgramming       0
ElectronicsAndSemicon     0
ComputerScience           0
conscientiousness         0
agreeableness             0
extraversion              0
nueroticism               0
openess_to_experience     0
dtype: int64
```

```
categorical_columns = df.select_dtypes(include=['object']).columns
numerical_columns = df.select_dtypes(exclude=['object']).columns

# Displaying categorical columns
print("Categorical Columns:")
print(categorical_columns)

# Displaying numerical columns
print("\nNumerical Columns:")
print(numerical_columns)
```

```
Categorical Columns:
Index(['DOL', 'Designation', 'JobCity', 'Gender', '10board',
'12board',
       'Degree', 'Specialization', 'CollegeState'],
      dtype='object')

Numerical Columns:
Index(['Salary', 'DOJ', 'DOB', '10percentage', '12graduation',
'12percentage',
       'collegeGPA', 'CollegeCityID', 'GraduationYear', 'English',
```

```
'Logical',
       'Quant', 'Domain', 'ComputerProgramming',
'ElectronicsAndSemicon',
       'ComputerScience', 'conscientiousness', 'agreeableness',
'extraversion',
       'nueroticism', 'openess_to_experience'],
      dtype='object')
```

```
df['DOL'].replace('present', np.nan, inplace = True)
```

```
df['JobCity'].fillna(df['JobCity'].mode()[0], inplace=True)
```

```
df['JobCity'].value_counts()
```

```
Bangalore           1088
Noida                368
Hyderabad            335
Pune                 290
Chennai              272
                    ...
Asansol                1
Tirunelvelli           1
Ernakulam              1
Nanded                 1
Asifabadbanglore       1
Name: JobCity, Length: 338, dtype: int64
```

```
df['Domain'].fillna(df['Domain'].mean(), inplace=True)
```

```
df['ComputerProgramming'].fillna(df['ComputerProgramming'].mean(),
inplace=True)
df['ElectronicsAndSemicon'].fillna(df['ElectronicsAndSemicon'].mean(),
inplace=True)
df['ComputerScience'].fillna(df['ComputerScience'].mean(),
inplace=True)
```

```
df
```

```
df.describe().T
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 19,\n  \"fields\": [\n
{\n      \"column\": \"count\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 0.0,\n        \"min\":
3998.0,\n        \"max\": 3998.0,\n        \"num_unique_values\": 1,\n
\"samples\": [\n          3998.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"mean\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 70447.79284181903,\n
\"min\": -0.1690325712856429,\n        \"max\": 307699.8499249625,\n
\"num_unique_values\": 19,\n        \"samples\": [\n
307699.8499249625\n        ],\n        \"semantic_type\": \"\",\n
```

\"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"std\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 48749.32462139551,\n   \"min\": 0.26461972342978013,\n   \"max\": 212737.49995685622,\n   \"num_unique_values\": 19,\n   \"samples\": [\n   212737.49995685622\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"min\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 8004.9085481865495,\n   \"min\": -7.3757,\n   \"max\": 35000.0,\n   \"num_unique_values\": 19,\n   \"samples\": [\n   35000.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"25%\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 41207.94274290083,\n   \"min\": -0.8682,\n   \"max\": 180000.0,\n   \"num_unique_values\": 19,\n   \"samples\": [\n   180000.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"50%\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 68693.80764157623,\n   \"min\": -0.2344,\n   \"max\": 300000.0,\n   \"num_unique_values\": 19,\n   \"samples\": [\n   300000.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"75%\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 84703.14898948921,\n   \"min\": 0.5024,\n   \"max\": 370000.0,\n   \"num_unique_values\": 19,\n   \"samples\": [\n   370000.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"max\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 917321.9835169201,\n   \"min\": 0.999910407631314,\n   \"max\": 4000000.0,\n   \"num_unique_values\": 19,\n   \"samples\": [\n   4000000.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe"}

```
categorical_columns
```

```
df['Designation'].value_counts()
```

```
software engineer                     539
software developer                    265
system engineer                       205
programmer analyst                    139
systems engineer                      118
                                     ...
cad drafter                             1
noc engineer                            1
human resources intern                  1
senior quality assurance engineer       1
```

```
jr. software developer                 1
Name: Designation, Length: 419, dtype: int64
```

```python
for col in categorical_columns:
  print('*****************************************')
  print(df[col].value_counts())
```

```
*****************************************
2015-04-01    573
2015-03-01    124
2015-05-01    112
2015-01-01     99
2014-04-01     82
             ...
2015-10-01      1
2010-02-01      1
2011-02-01      1
2005-03-01      1
2010-10-01      1
Name: DOL, Length: 66, dtype: int64
*****************************************
software engineer                    539
software developer                   265
system engineer                      205
programmer analyst                   139
systems engineer                     118
                                    ...
cad drafter                            1
noc engineer                           1
human resources intern                 1
senior quality assurance engineer      1
jr. software developer                 1
Name: Designation, Length: 419, dtype: int64
*****************************************
Bangalore           1088
Noida                368
Hyderabad            335
Pune                 290
Chennai              272
                    ...
Asansol                1
Tirunelvelli           1
Ernakulam              1
Nanded                 1
Asifabadbanglore       1
Name: JobCity, Length: 338, dtype: int64
*****************************************
m    3041
f     957
Name: Gender, dtype: int64
```

```
*****************************************
cbse                            1395
state board                     1164
0                                350
icse                             281
ssc                              122
                                ...
hse,orissa                         1
national public school              1
nagpur board                        1
jharkhand academic council          1
bse,odisha                          1
Name: 10board, Length: 275, dtype: int64
*****************************************
cbse                                  1400
state board                           1254
0                                      359
icse                                   129
up board                                87
                                      ...
jawahar higher secondary school         1
nagpur board                            1
bsemp                                   1
board of higher secondary orissa        1
boardofintermediate                     1
Name: 12board, Length: 340, dtype: int64
*****************************************
B.Tech/B.E.      3700
MCA               243
M.Tech./M.E.       53
M.Sc. (Tech.)       2
Name: Degree, dtype: int64
*****************************************
electronics and communication engineering      880
computer science & engineering                 744
information technology                         660
computer engineering                           600
computer application                           244
mechanical engineering                         201
electronics and electrical engineering         196
electronics & telecommunications               121
electrical engineering                          82
electronics & instrumentation eng               32
civil engineering                               29
electronics and instrumentation engineering     27
information science engineering                 27
instrumentation and control engineering         20
electronics engineering                         19
biotechnology                                   15
```

```
other                                                 13
industrial & production engineering                   10
applied electronics and instrumentation                9
chemical engineering                                   9
computer science and technology                        6
telecommunication engineering                          6
mechanical and automation                              5
automobile/automotive engineering                      5
instrumentation engineering                            4
mechatronics                                           4
aeronautical engineering                               3
electronics and computer engineering                   3
electrical and power engineering                       2
biomedical engineering                                 2
information & communication technology                 2
industrial engineering                                 2
computer science                                       2
metallurgical engineering                              2
power systems and automation                           1
control and instrumentation engineering                1
mechanical & production engineering                    1
embedded systems technology                            1
polymer technology                                     1
computer and communication engineering                 1
information science                                    1
internal combustion engine                             1
computer networking                                    1
ceramic engineering                                    1
electronics                                            1
industrial & management engineering                    1
Name: Specialization, dtype: int64
*****************************************
Uttar Pradesh        915
Karnataka            370
Tamil Nadu           367
Telangana            319
Maharashtra          262
Andhra Pradesh       225
West Bengal          196
Punjab               193
Madhya Pradesh       189
Haryana              180
Rajasthan            174
Orissa               172
Delhi                162
Uttarakhand          113
Kerala                33
Jharkhand             28
Chhattisgarh          27
```

```
Gujarat                24
Himachal Pradesh       16
Bihar                  10
Jammu and Kashmir       7
Assam                   5
Union Territory         5
Sikkim                  3
Meghalaya               2
Goa                     1
Name: CollegeState, dtype: int64
```

```
board10_mapping = {
    'cbse' : 'CBSE',
    'cbse[gulf zone]': 'CBSE',
    'cbse board': 'CBSE',
    'hbse' : 'CBSE',
    'hbsc' : 'CBSE',

    'andhra pradesh board ssc' : 'AP Board',
    'andhra pradesh state board' : 'AP Board',
    'ap state board': 'AP Board',
    'ap state board for secondary education':'AP Board',
    'board ofsecondary education,ap': 'AP Board',
    'ssc board of andrapradesh' : 'AP Board',
    'board of secondary education,ap': 'AP Board',
    'state board of secondary education, ap': 'AP Board',
    'apssc': 'AP Board',
    'apsche': 'AP Board',
    'school secondary education, andhra pradesh':'AP Board',
    'state board of secondary education,andhra pradesh':'AP Board',
    'ssc-andhra pradesh' : 'AP Board',
    'board of ssc education andhra pradesh' : 'AP Board',
    'state board of secondary education, andhra pradesh': 'AP Board',
    'board of secondary education,andhra pradesh':'AP Board',
    'board of  secondary education':'SSC',


    'stateboard':'SSC',
    'state boardmp board ':'SSC',
    'state bord':'SSC',
    'secondary school certificate':'SSC',
    'secondary school education':'SSC',
    'seba':'SSC',
    'sslc':'SSC',
    'national public school':'SSC',
    'board of secondary education,andhara pradesh':'AP Board',
    'board of secondary education, andhra pradesh':'AP Board',
    'board of secondary education - andhra pradesh':'AP Board',
    'board of secondaray education':'SSC',
    'board of secondary school education':'SSC',
```

```
    'central board of secondary education':'CBSE',
    'certificate of middle years program of ib':'CBSE',
    'central board of secondary education, new delhi':'CBSE',

    'bihar board' : 'Bihar Board',
    'bihar examination board, patna' : 'Bihar Board',
    'bihar' : 'Bihar Board',
    'bihar secondary education board,patna' : 'Bihar Board',
    'bihar school examination board' : 'Bihar Board',
    'bihar school examination board patna' : 'Bihar Board',
    'biharboard' : 'Bihar Board',
    'bseb, patna' : 'Bihar Board',
    'bseb,patna': 'Bihar Board',
    'bseb patna': 'Bihar Board',
    'bsepatna': 'Bihar Board',

    'mp board': 'MP Board',
    'mp board bhopal': 'MP Board',
    'mpboard': 'MP Board',
    'mp': 'MP Board',
    'mp state board': 'MP Board',
    'mpbse': 'MP Board',
    'state boardmp board': 'MP Board',
    'mp-bse': 'MP Board',
    'bsemp': 'MP Board',
    'madhya pradesh board':'MP Board',

    'bse, odisha' : 'Odisha Board',
    'board of secondary education,odisha' : 'Odisha Board',
    'bse, odisha' : 'Odisha Board',
    'bse,odisha' : 'Odisha Board',

    'board of secondary education(bse) orissa' : 'Orissa Board',
    'bsc,orissa' : 'Orissa Board',
    'board of secendary education orissa' : 'Orissa Board',
    'bse,orissa' : 'Orissa Board',
    'hse,orissa' : 'Orissa Board',
    'board of secondary education orissa' : 'Orissa Board',
    'board of secondary education,orissa':'Orissa Board',
    'board of secondary education (bse) orissa': 'Orissa Board',

    'anglo indian' : 'Anglo Indian',
    'state board' : 'SSC',
    'state board ':'SSC',
    'ssc' : 'SSC',
    'sda matric higher secondary school' : 'SSC',
    'matriculation' : 'SSC',
    'little jacky matric higher secondary school' : 'SSC',
    'matriculation board' : 'SSC',
```

```
    'metric': 'SSC',
    'maticulation'
    'sslc' : 'SSC',
    'board secondary  education' : 'SSC',
    'matric board' : 'SSC',
    'board of secondary education' : 'SSC',
    'mirza ahmed ali baig' : 'SSC',
    'board of intermediate education' : 'SSC',
    'bseb':'CBSE',
    'stjosephs girls higher sec school,dindigul' : 'SSC',
    'secondary school cerfificate' : 'SSC',
    'state' : 'SSC',
    'matric': 'SSC',
    'maticulation': 'SSC',
    'secondary school of education' : 'SSC',
    'stjoseph of cluny matrhrsecschool,neyveli,cuddalore district' :
'SSC',
    'bharathi matriculation school' : 'SSC',
    'bse(board of secondary education)' : 'SSC',
    'secondary state certificate' : 'SSC',
    'bseb' : 'SSC',
    'stmary higher secondary' : 'SSC',
    'kiran english medium high school' : 'SSC',
    'sarada high scchool' : 'SSC',
    'sslc board' : 'SSC',
    'stmary\'s convent inter college' : 'SSC',
    'don bosco maatriculation school' : 'SSC',
    'state board of secondary education( ssc)' : 'SSC',
    'board of ssc': 'SSC',
    'ssc regular':'SSC',
    'kalaimagal matriculation higher secondary school':'SSC',
    'maticulation':'SSC',
    'matric board':'SSC',
    'matric':'SSC',
    'state board (jac, ranchi)':'SSC',

    'aurangabad board' : 'Aurangabad Board',

    'up' : 'UP Board',
    'up board' : 'UP Board',
    'up(allahabad)' : 'UP Board',
    'upbhsie' : 'UP Board',
    'up board , allahabad' : 'UP Board',
    'upboard' : 'UP Board',
    'bright way college, (up board)' : 'UP Board',
    'up board allahabad' : 'UP Board',
    'upboard' : 'UP Board',
    'up board,allahabad' : 'UP Board',
    'up-board' : 'UP Board',
```

```
    'uttar pradesh board': 'UP Board',
    'uttar pradesh': 'UP Board',
    'board of high school and intermediate education uttarpradesh':
'UP Board',
    'u p board': 'UP Board',
    'u p': 'UP Board',
    'up bord': 'UP Board',
    'up bourd': 'UP Board',
    'up baord': 'UP Board',
    'up borad': 'UP Board',

    'nagpur':'Nagpur Board',
    'nagpur board':'Nagpur Board',
    'nagpur board,nagpur':'Nagpur Board',
    'nagpur divisional board':'Nagpur Board',


    'uttarakhand board':'Uttarakhand Board',
    'uttaranchal shiksha avam pariksha parishad' : 'Uttaranchal
Board',
    'uttaranchal state board': 'Uttaranchal Board',
    'uttranchal board':'Uttaranchal Board',
    'uttarakhand board' : 'Uttarakhand Board',
    'uttrakhand board':'Uttarakhand Board',
    'board of school education uttarakhand':'Uttarakhand Board',

    'mhsbse' : 'Maharashtra Board',
    'maharashtra state board': 'Maharashtra Board',
    'maharashtra': 'Maharashtra Board',
    'maharashtra state(latur board)': 'Maharashtra Board',
    'maharashtra state board pune': 'Maharashtra Board',
    'kolhapur divisional board, maharashtra': 'Maharashtra Board',
    'maharashtra state board of secondary and higher secondary
education,pune': 'Maharashtra Board',
    'maharashtra state board,pune': 'Maharashtra Board',
    'maharastra board': 'Maharashtra Board',
    'maharashtra board': 'Maharashtra Board',
    'ssc maharashtra board': 'Maharashtra Board',
    'maharashtra satate board': 'Maharashtra Board',
    'maharashtra state board mumbai divisional board': 'Maharashtra
Board',
    'maharashtra nasik board': 'Maharashtra Board',
    'maharashtra state board': 'Maharashtra Board',
    'maharashtra state board for ssc': 'Maharashtra Board',
    'maharashtra state board of secondary and higher secondary
education': 'Maharashtra Board',
    'maharashtra state boar of secondary and higher secondary
education': 'Maharashtra Board',
    'maharashtra state board of secondary & higher secondary
education': 'Maharashtra Board',
```

```
    'msbshse,pune':'Maharashtra Board',
    'kolhapur' : 'Maharashtra Board',
    'maharashtra sate board': 'Maharashtra Board',
    'nashik board': 'Maharashtra Board',
    'nasik': 'Maharashtra Board',
    'maharashtra board, pune': 'Maharashtra Board',
    'seba(assam)': 'CBSE',
    'sss pune': 'Maharashtra Board',
    'msbshse,pune': 'Maharashtra Board',
    'pune board': 'Maharashtra Board',
    'pune': 'Maharashtra Board',
    'maharashtra board, pune': 'Maharashtra Board',
    'latur board':'Maharashtra Board',
    'latur':'Maharashtra Board',
    'mumbai board':'Maharashtra Board',
    'ms board':'Maharashtra Board',

    'tamilnadu matriculation board':'TamilNadu Board',
    'tamilnadu state board':'TamilNadu Board',
    'tamil nadu state':'TamilNadu Board',
    'tn state board':'TamilNadu Board',


    'kerala state technical education' : 'Kerala Board',
    'kerala': 'Kerala Board',
    'education board of kerala': 'Kerala Board',
    'kerala state board': 'Kerala Board',
    'kerala university': 'Kerala Board',
    'kerala state board': 'Kerala Board',
    'kea':'Kerala Board',

    'west bengal board of secondary education' : 'West Bengal Board',
    'west bengal board of secondary examination (wbbse)': 'West Bengal
Board',
    'west bengal board of secondary eucation': 'West Bengal Board',
    'state board - west bengal board of secondary education : wbbse':
'West Bengal Board',
    'wbbse': 'West Bengal Board',
    'wbbsce': 'West Bengal Board',
    'west bengal  board of secondary education':'West Bengal Board',

    'karnataka secondary school of examination':'Karnataka Board',
    'karnataka state education examination board':'Karnataka Board',
    'karnataka state board':'Karnataka Board',
    'karnataka secondory education board':'Karnataka Board',
    'karnataka education board (keeb)':'Karnataka Board',
    'karnataka state secondary education board':'Karnataka Board',
    'karnataka board of secondary education':'Karnataka Board',
    'Karnataka Board':'Karnataka Board',
    'kseeb(karnataka secondary education examination
```

```
board)':'Karnataka Board',
    'state(karnataka board)':'Karnataka Board',
    'karnataka secondary education examination board':'Karnataka
Board',
    'karnataka education board':'Karnataka Board',
    'karnataka sslc board bangalore':'Karnataka Board',
    'karnataka':'Karnataka Board',
    'state board of karnataka':'Karnataka Board',
    'karnataka secondary board':'Karnataka Board',
    'karnataka board of higher education':'Karnataka Board',
    'ksseb(karnataka state board)':'Karnataka Board',
    'karnataka secondary education':'Karnataka Board',
    'karnataka secondary education board':'Karnataka Board',
    'karnataka state secondary education board':'Karnataka Board',
    'sslc,karnataka':'Karnataka Board',
    'karnataka state examination board':'Karnataka Board',
    'kseeb':'Karnataka Board',
    'kseb':'Karnataka Board',
    'ksbe':'Karnataka Board',
    'karantaka secondary education and examination borad' : 'Karnataka
Board',
    'ksseb':'Karnataka Board',
    'karnataka secondary eduction':'Karnataka Board',

    'board of school education haryana': 'Haryana Board',
    'haryana board of school education': 'Haryana Board',
    'haryana board of school education,(hbse)': 'Haryana Board',
    'board of school education harayana' : 'Haryana Board',
    'state borad hp': 'Haryana Board',

    'board of secondary education rajasthan':'Rajasthan Board',
    'rajasthan board':'Rajasthan Board',
    'secondary board of rajasthan':'Rajasthan Board',
    'board of secondary education, rajasthan':'Rajasthan Board',
    'board of secondary education,rajasthan(rbse)':'Rajasthan Board',
    'rajasthan board ajmer':'Rajasthan Board',
    'rajasthan board of secondary education':'Rajasthan Board',
    'secondary education board of rajasthan':'Rajasthan Board',
    'rbse':'Rajasthan Board',
    'rbse,ajmer':'Rajasthan Board',
    'rbse (state board)':'Rajasthan Board',

    'delhi board':'CBSE',
    'central board of secondary education, new delhi':'CBSE',
    'delhi public school':'CBSE',
    'ua':'CBSE',
    'icse board , new delhi':'ICSE',

    'CICSE':'ICSE',
    'ICSC':'ICSE',
```

```
    'CGBSE':'CBSE',
    'cbsc':'CBSE',
    'cgbse' : 'CBSE',
    'cbse raipur': 'CBSE',
    'cicse' : 'ICSE',
    'icse' : 'ICSE',
    'icse board' : 'ICSE',
    'hse,board':'SSC',
    'aisse':'SSC',
    'hse':'SSC',
    'hsce':'SSC',
    'bse':'CSBSE',
    'jseb':'CBSE',
    'bse':'CBSE',
    'bseb ,patna':'CBSE',
    'jseb':'CBSE',

    'gujarat state board' : 'Gujarat Board',
    'gsheb': 'Gujarat Board',
    'gseb': 'Gujarat Board',
    'ghseb': 'Gujarat Board',
    'gujarat board': 'Gujarat Board',

    'j & k bord': 'J&K Board',
    'jkbose': 'J&K Board',

    'jharkhand secondary examination board,ranchi' : 'Jharkhand
Board',
    'jharkhand accademic council': 'Jharkhand Board',
    'jharkhand secondary examination board (ranchi)': 'Jharkhand
Board',
    'jharkhand secondary board': 'Jharkhand Board',
    'jharkhand secondary education board': 'Jharkhand Board',
    'jbse,jharkhand': 'Jharkhand Board',
    'j&k state board of school education': 'Jharkhand Board',
    'jharkhand acedemic council': 'Jharkhand Board',
    'jharkhand academic council': 'Jharkhand Board',

    'sri kannika parameswari highier secondary school,
udumalpet':'SSC',
    'hsc' : 'SSC',
    'cluny' : 'SSC',
    'jawahar navodaya vidyalaya': 'SSC',
    'dav public school,hehal': 'SSC',
    'holy cross matriculation hr sec school': 'SSC',
    'gyan bharati school': 'SSC',
    'dav public school sec 14': 'SSC',
    'council for indian school certificate examination': 'SSC',

    'himachal pradesh board of school education' : 'HP Board',
```

```
    'himachal pradesh board': 'HP Board',

    'punjab school education board, mohali' : 'Punjab Board',
    'pseb': 'Punjab Board',

}

df['10board'] = df['10board'].replace(board10_mapping)

df['10board'].value_counts()
```

```
SSC                   1445
CBSE                  1437
0                      350
ICSE                   284
UP Board               144
Karnataka Board         61
Maharashtra Board       52
Rajasthan Board         41
MP Board                31
West Bengal Board       26
AP Board                23
Bihar Board             18
Orissa Board            10
Jharkhand Board          9
Punjab Board             9
Gujarat Board            9
Kerala Board             7
Haryana Board            6
Uttarakhand Board        6
Nagpur Board             5
TamilNadu Board          5
Odisha Board             4
J&K Board                3
Uttaranchal Board        3
Anglo Indian             3
HP Board                 2
karnataka board          2
cbse                     1
cgbse raipur             1
Aurangabad Board         1
Name: 10board, dtype: int64
```

```
df['12board'].value_counts()
```

```
cbse                                  1400
state board                           1254
0                                      359
icse                                   129
up board                                87
```

```
                                        ...
jawahar higher secondary school        1
nagpur board                           1
bsemp                                  1
board of higher secondary orissa       1
boardofintermediate                    1
Name: 12board, Length: 340, dtype: int64
```

```
df['12board'].unique()
```

```
array(['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',
       'chsc', 'board of intermediate', 'isce', 'upboard', 'sbtet',
       'hisher seconadry examination(state board)', 'pre university',
       'borad of intermediate', 'j & k board',
       'intermediate board of andhra pardesh', 'rbse',
       'central board of secondary education', 'jkbose', 'hbse',
       'board of intermediate education', 'state', 'ms board', 'pue',
       'intermediate state board', 'stateboard', 'hsc',
       'electonincs and communication(dote)', 'karnataka pu board',
       'government polytechnic mumbai , mumbai board', 'pu board',
       'baord 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 vidayapeeth',
       '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', 'staae 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 accademic
council',
       'bieap', 'msbte (diploma in computer technology)',
       'jharkhand acamedic council (ranchi)',
```

```
       'department of pre-university eduction', 'biec', 'all india
board',
       'sjrcw', ' board of intermediate', 'msbte',
       'sri sankara vidyalaya', 'chse, odisha', 'bihar board',
       'maharashtra state(latur board)', 'rajasthan board', 'mpboard',
       'state board of technical eduction 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 baord', 'pu',
       'dte', 'board of secondary education', 'pre-university',
       'board of intermediate education,andhra pradesh',
       'up board , allahabad', 'srv girls 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 education', 'pu
board ,karnataka',
       'karnataka secondary education board', 'karnataka sslc',
       'board of intermediate ap', 'u p', 'state board of karnataka',
       'directorate of technical education,banglore', 'matric board',
       'andhpradesh board of intermediate education',
       'stjoseph of cluny matrhrsecschool,neyveli,cuddalore district',
       'bte up', 'scte and vt ,orissa', 'hbsc',
       'jawahar higher secondary school', 'nagpur board', 'bsemp',
       'board of intermediate education, andhra pradesh',
       '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',
```

```
       'bihar', '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
uttarakhand',
       'gseb/technical education board', 'msbshse,pune',
       'tamilnadu state board', 'board of technical education',
       '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 bord',
       '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 university',
       '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',
       'bteup', 'board of intermediate(bie)', 'hsc maharashtra board',
       'tamil nadu state', 'uttrakhand board', 'psbte',
       'stateboard/tamil nadu', 'intermediate council patna',
       'technical board, punchkula', 'board of intermidiate
examination',
       'sri kannika parameswari highier secondary school, udumalpet',
       'ap board', 'nashik board', 'himachal pradesh board',
       'maharashtra satate board',
       'andhra pradesh board of secondary education',
       'tamil nadu polytechnic',
       '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', 'cbsc',
       'pu  board karnataka', 'biec patna', 'kolhapur', 'bseb, patna',
       'up board allahabad', 'intermideate', 'nagpur board,nagpur',
       'diploma(msbte)', 'dav public school',
       'pre university board, karnataka', 'ssm srsecschool', 'state
bord',
       'jstb,jharkhand', 'intermediate board of education',
       'mp board bhopal', '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 education', '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 board', '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
education',
       '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, chandigarh',
       'state board - tamilnadu', 'jharkhand acedemic council',
       'scte & vt (diploma)', 'karnataka pu',
       'board of intmediate education ap', 'up-board',
       'boardofintermediate'], dtype=object)
```

```
df['DOB'] = pd.to_datetime(df['DOB'])
df['Age'] = 2015 - df['DOB'].dt.year
df.head()
```

{"type":"dataframe","variable_name":"df"}

```
numerical_columns = numerical_columns.append(pd.Index(['Age']))
```

Univariate Analysis

```
numerical_columns
```

```
Index(['Salary', 'DOJ', 'DOB', '10percentage', '12graduation',
'12percentage',
       'collegeGPA', 'CollegeCityID', 'GraduationYear', 'English',
'Logical',
       'Quant', 'Domain', 'ComputerProgramming',
'ElectronicsAndSemicon',
       'ComputerScience', 'conscientiousness', 'agreeableness',
'extraversion',
       'nueroticism', 'openess_to_experience', 'Age'],
      dtype='object')
```

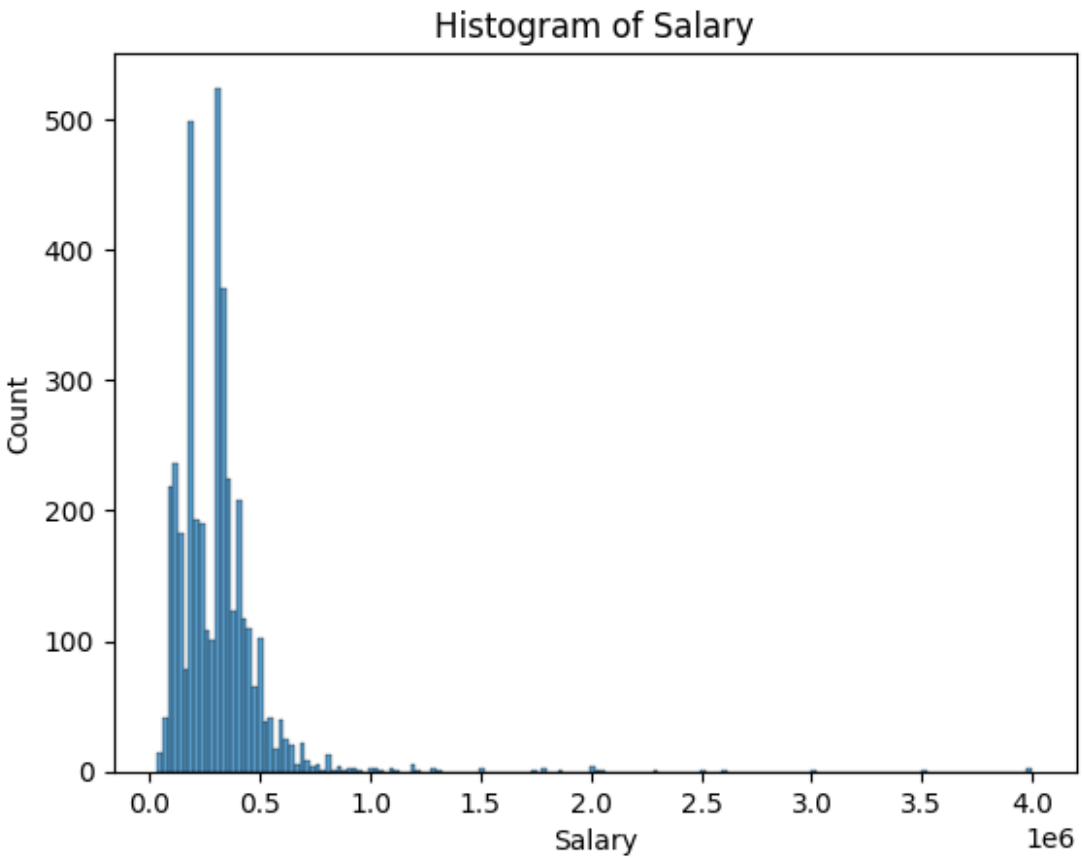
```
#Non Viz Numerical columns
numerical_columns = df.select_dtypes(include=['int', 'float'])

# Calculate summary statistics for numerical columns
summary_statistics = numerical_columns.describe()
print(summary_statistics)
```

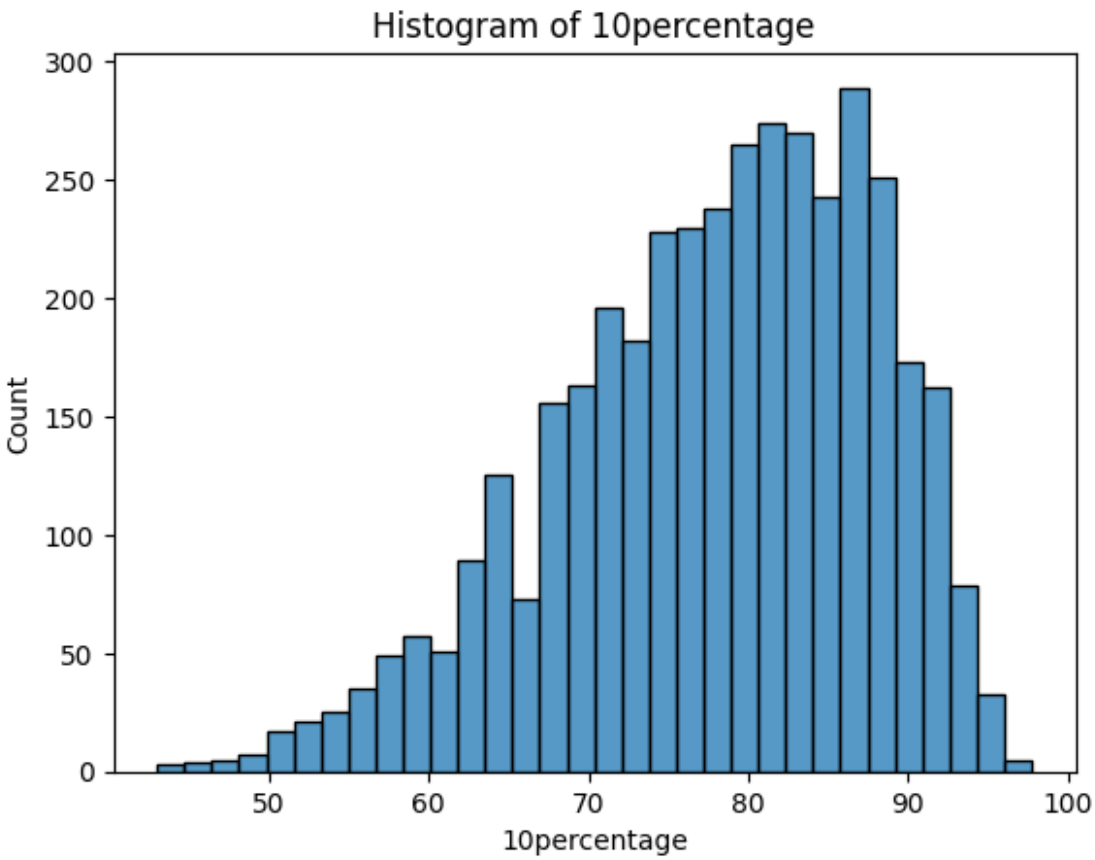
```
             Salary  10percentage  12graduation  12percentage
collegeGPA  \
count  3.998000e+03   3998.000000   3998.000000   3998.000000
3998.000000
mean   3.076998e+05     77.925443   2008.087544     74.466366
71.486171
std    2.127375e+05      9.850162      1.653599     10.999933
8.167338
min    3.500000e+04     43.000000   1995.000000     40.000000
```

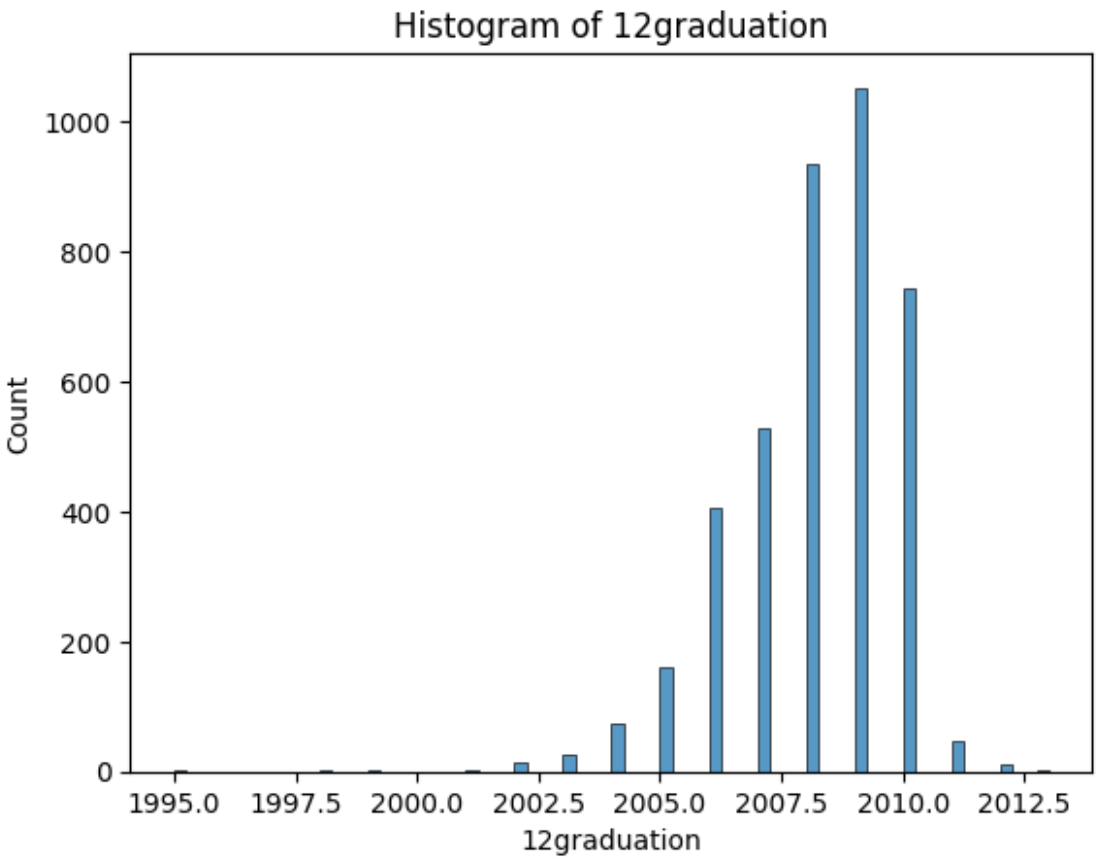
```
6.450000
25%    1.800000e+05     71.680000   2007.000000     66.000000
66.407500
50%    3.000000e+05     79.150000   2008.000000     74.400000
71.720000
75%    3.700000e+05     85.670000   2009.000000     82.600000
76.327500
max    4.000000e+06     97.760000   2013.000000     98.700000
99.930000

       CollegeCityID  GraduationYear      English      Logical
Quant  \
count    3998.000000     3998.000000  3998.000000  3998.000000
3998.000000
mean     5156.851426     2012.105803   501.649075   501.598799
513.378189
std      4802.261482       31.857271   104.940021    86.783297
122.302332
min         2.000000        0.000000   180.000000   195.000000
120.000000
25%       494.000000     2012.000000   425.000000   445.000000
430.000000
50%      3879.000000     2013.000000   500.000000   505.000000
515.000000
75%      8818.000000     2014.000000   570.000000   565.000000
595.000000
max     18409.000000     2017.000000   875.000000   795.000000
900.000000

            Domain  ComputerProgramming  ElectronicsAndSemicon  \
count  3998.000000          3998.000000            3998.000000
mean      0.609525           451.301278             335.644231
std       0.264620            85.964144              43.419466
min       0.002750           105.000000             133.000000
25%       0.413838           405.000000             335.644231
50%       0.622643           451.301278             335.644231
75%       0.842248           495.000000             335.644231
max       0.999910           840.000000             612.000000

       ComputerScience  conscientiousness  agreeableness  extraversion
\
count      3998.000000        3998.000000    3998.000000   3998.000000

mean        405.636364          -0.037831       0.146496      0.002763

std          42.711877           1.028666       0.941782      0.951471

min         130.000000          -4.126700      -5.781600     -4.600900

25%         405.636364          -0.713525      -0.287100     -0.604800
```

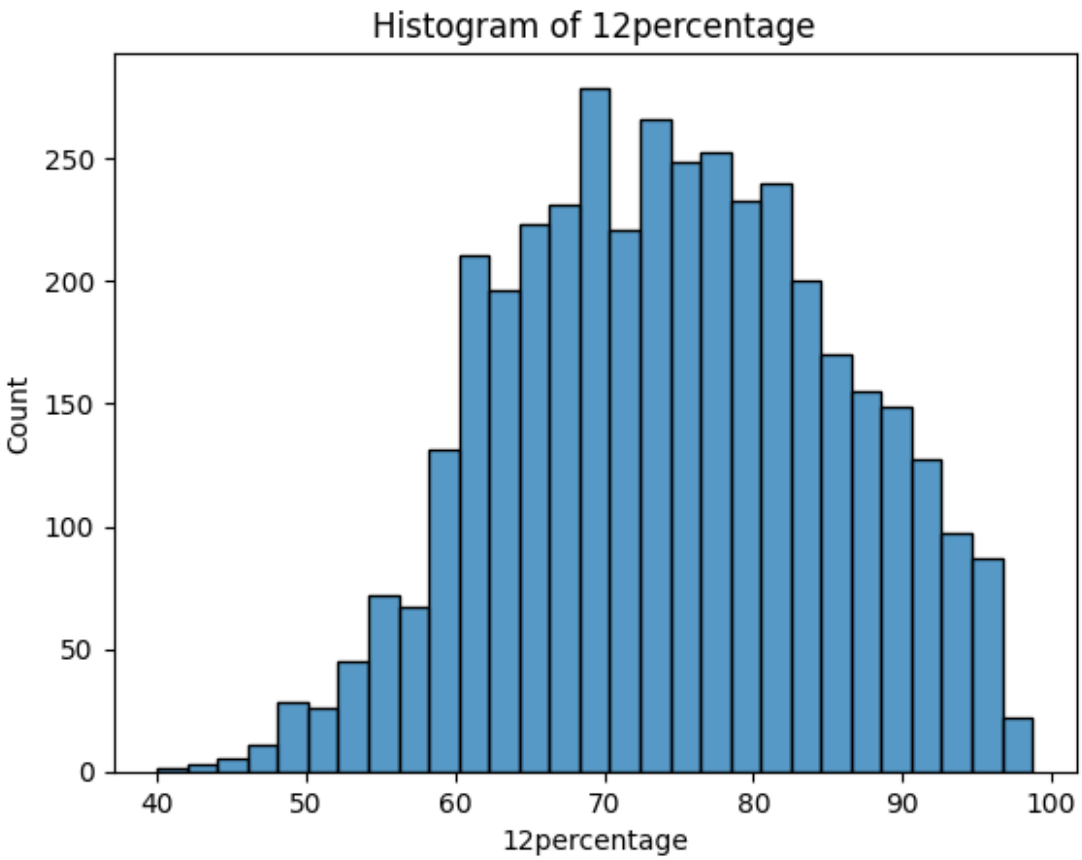
```
50%         405.636364           0.046400       0.212400      0.091400

75%         405.636364           0.702700       0.812800      0.672000

max         715.000000           1.995300       1.904800      2.535400


       nueroticism  openess_to_experience          Age
count  3998.000000            3998.000000  3998.000000
mean     -0.169033              -0.138110    24.572536
std       1.007580               1.008075     1.767473
min      -2.643000              -7.375700    18.000000
25%      -0.868200              -0.669200    23.000000
50%      -0.234400              -0.094300    24.000000
75%       0.526200               0.502400    26.000000
max       3.352500               1.822400    38.000000
```

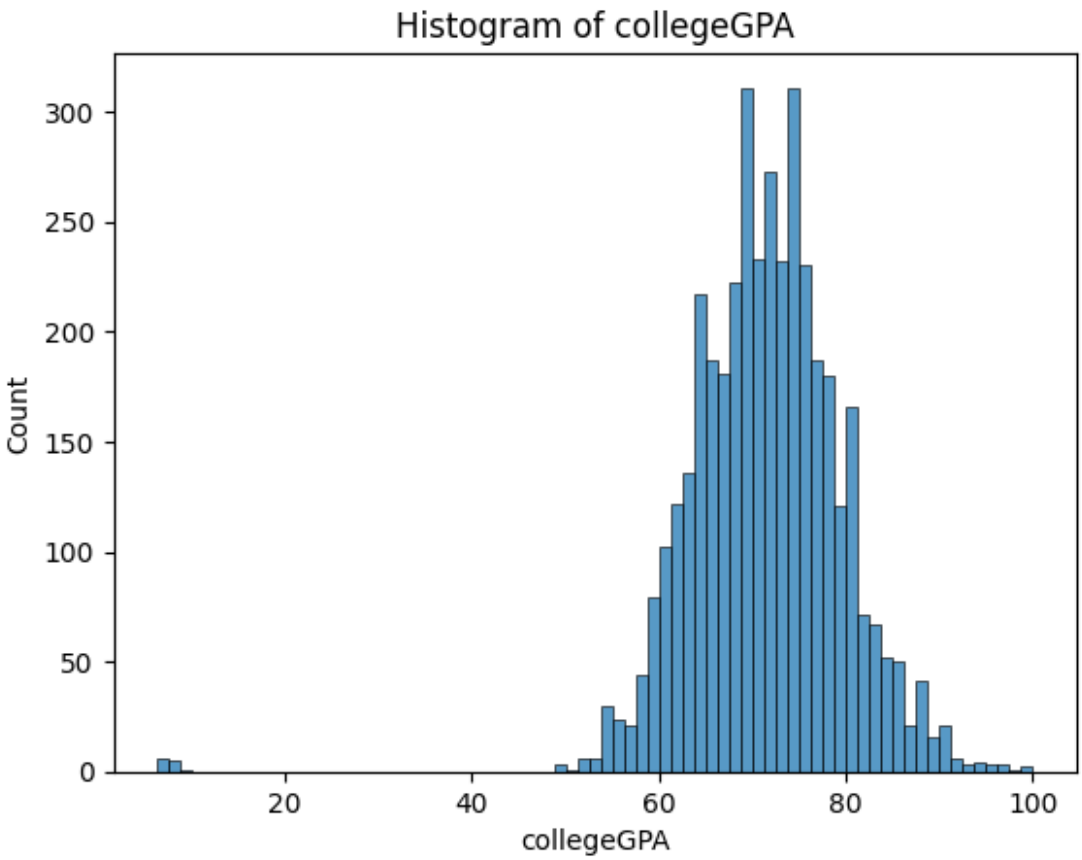
```python
# Create histograms for each numerical column
for col in numerical_columns:
    plt.figure()
    sns.histplot(df[col])
    plt.title(f"Histogram of {col}")
    plt.show()
```

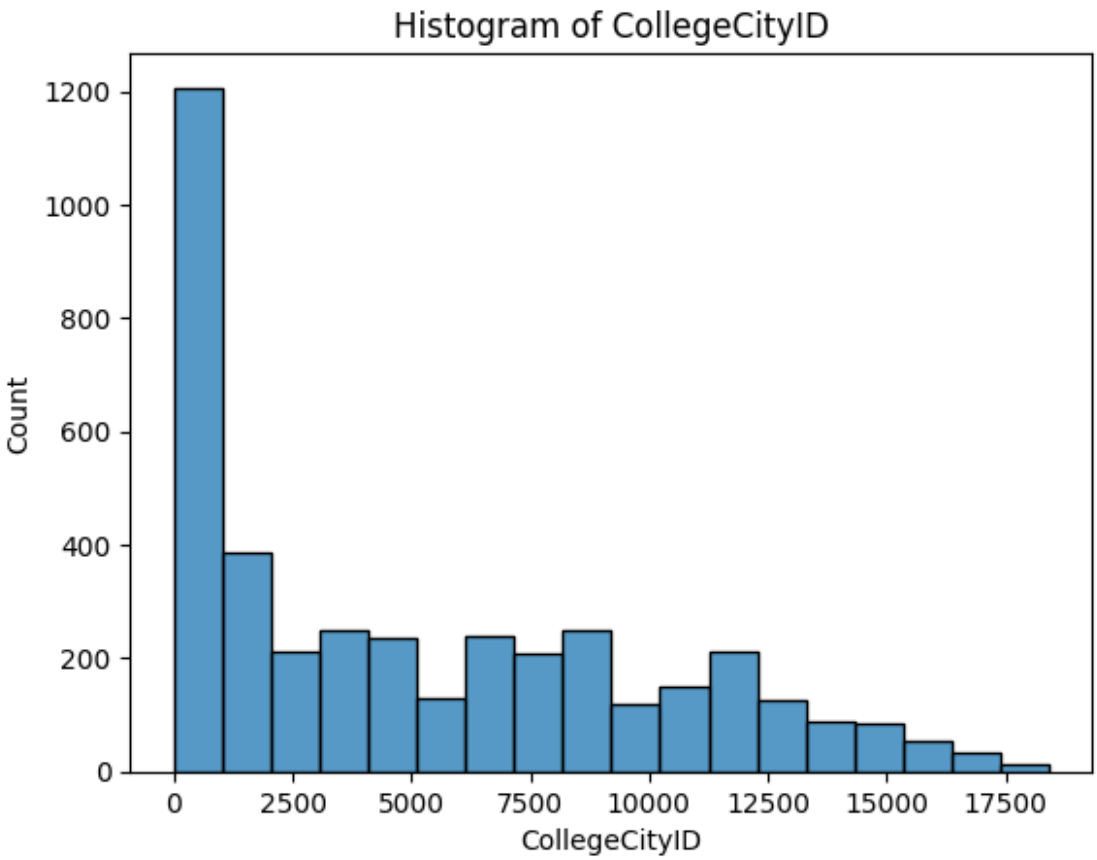

Histogram of Salary
Count
500
400
300
200
100
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

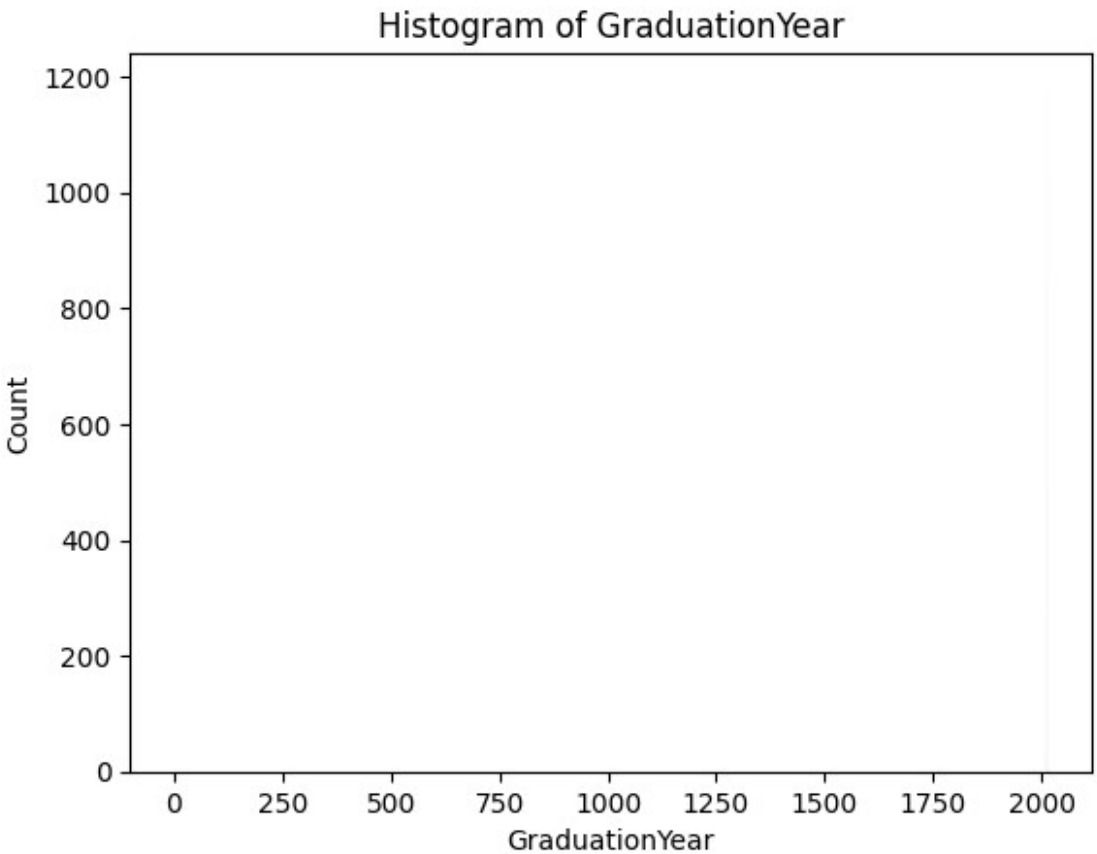

Histogram of 10percentage
Count
10percentage
0
50
100
150
200
250
300
50
60
70
80
90
100

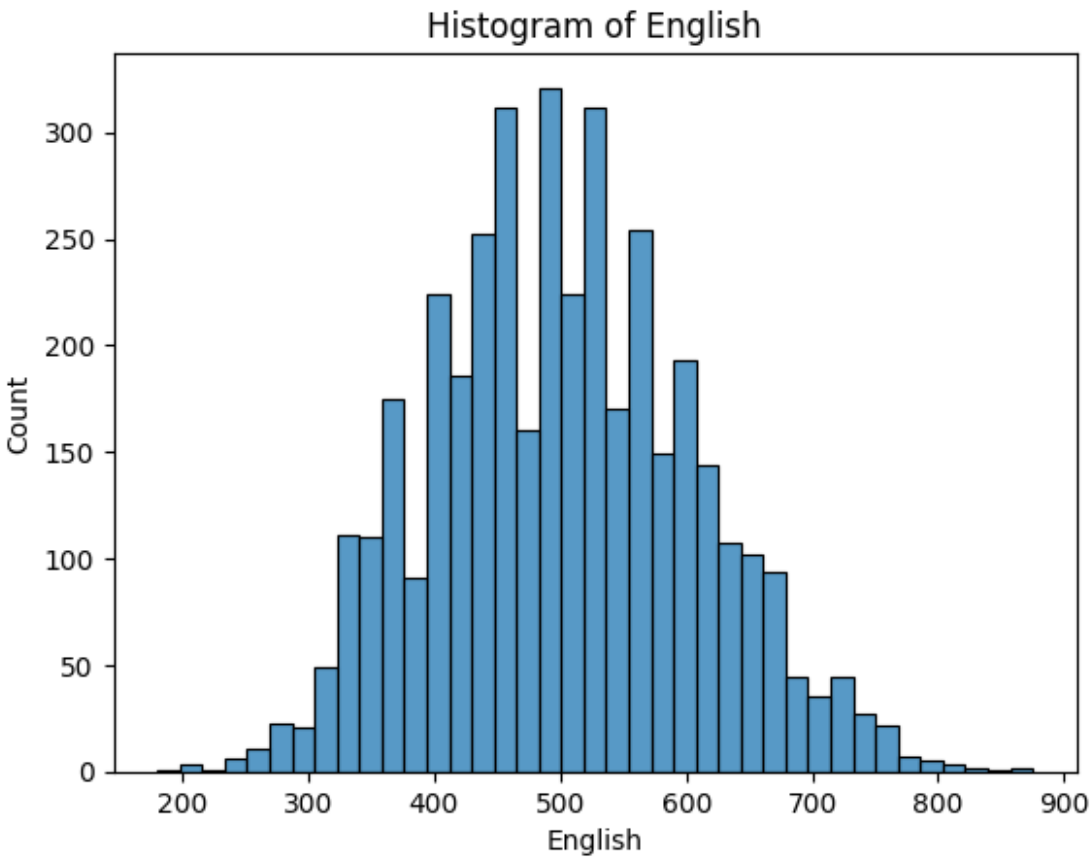

Histogram of 12graduation
Count
0
200
400
600
800
1000
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

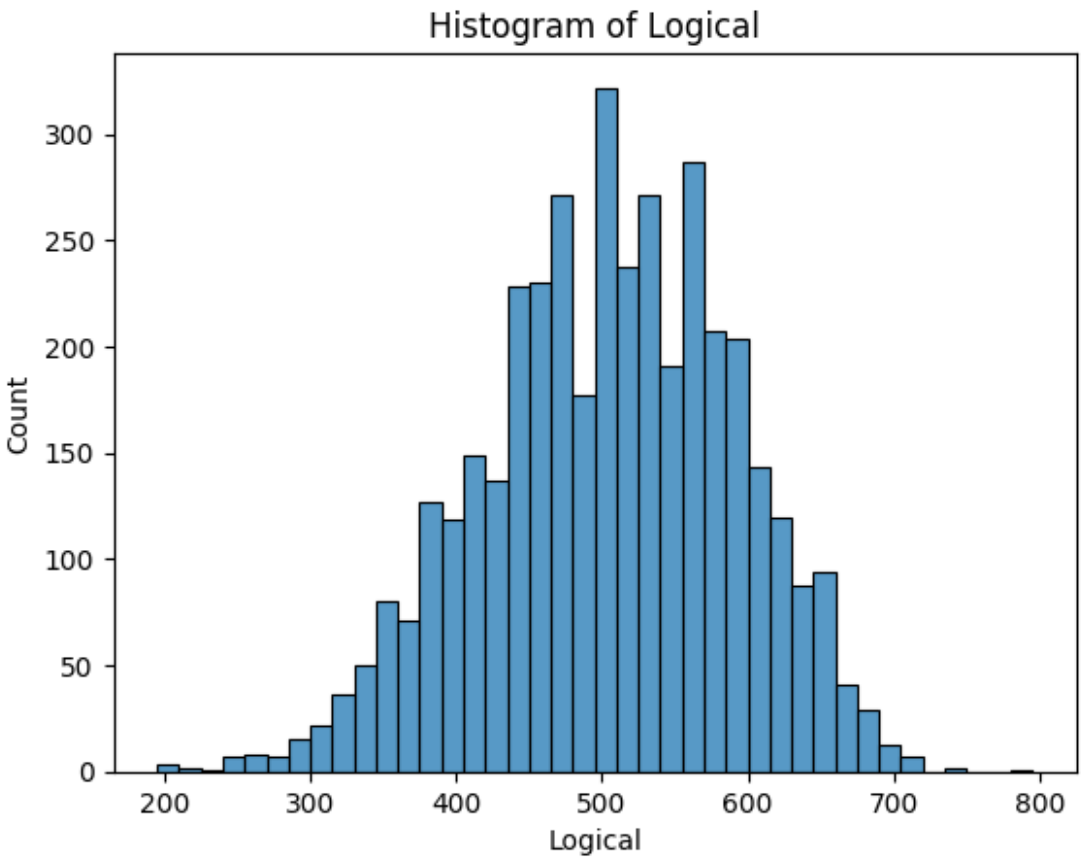

Histogram of 12percentage
Count
0
50
100
150
200
250
40
50
60
70
80
90
100
12percentage

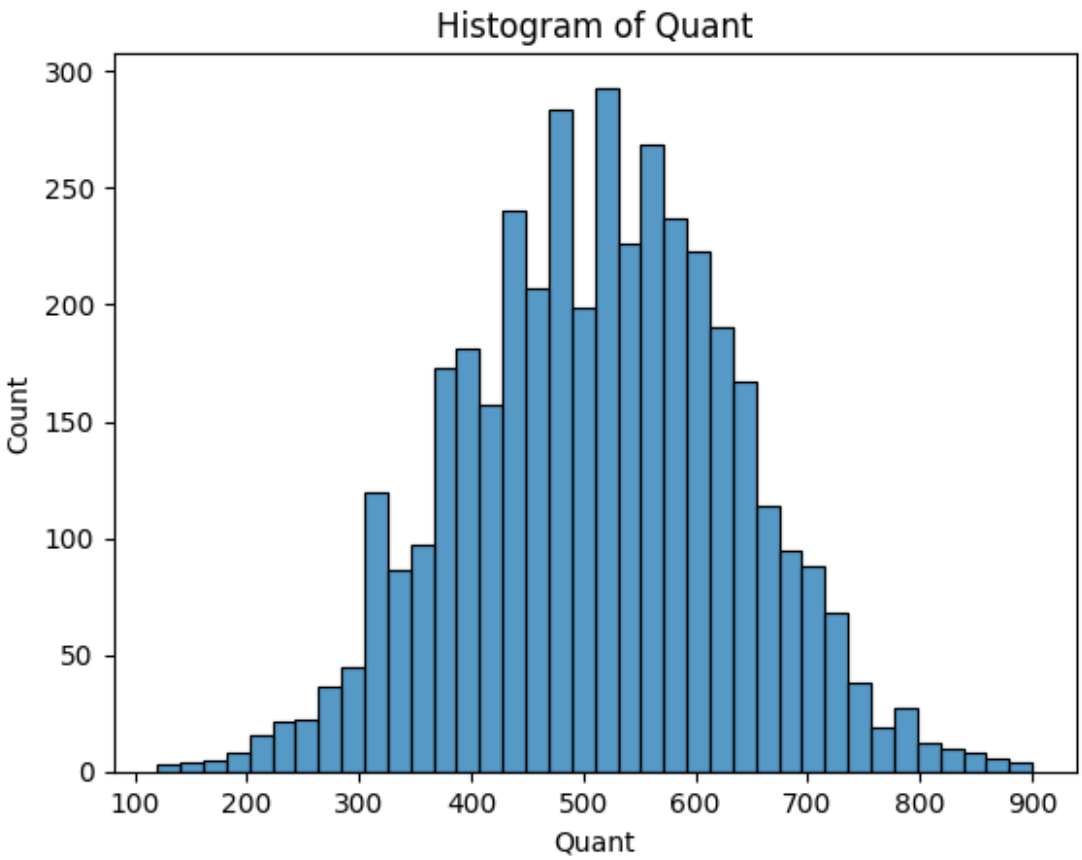

Histogram of collegeGPA
Count
collegeGPA
0
50
100
150
200
250
300
20
40
60
80
100

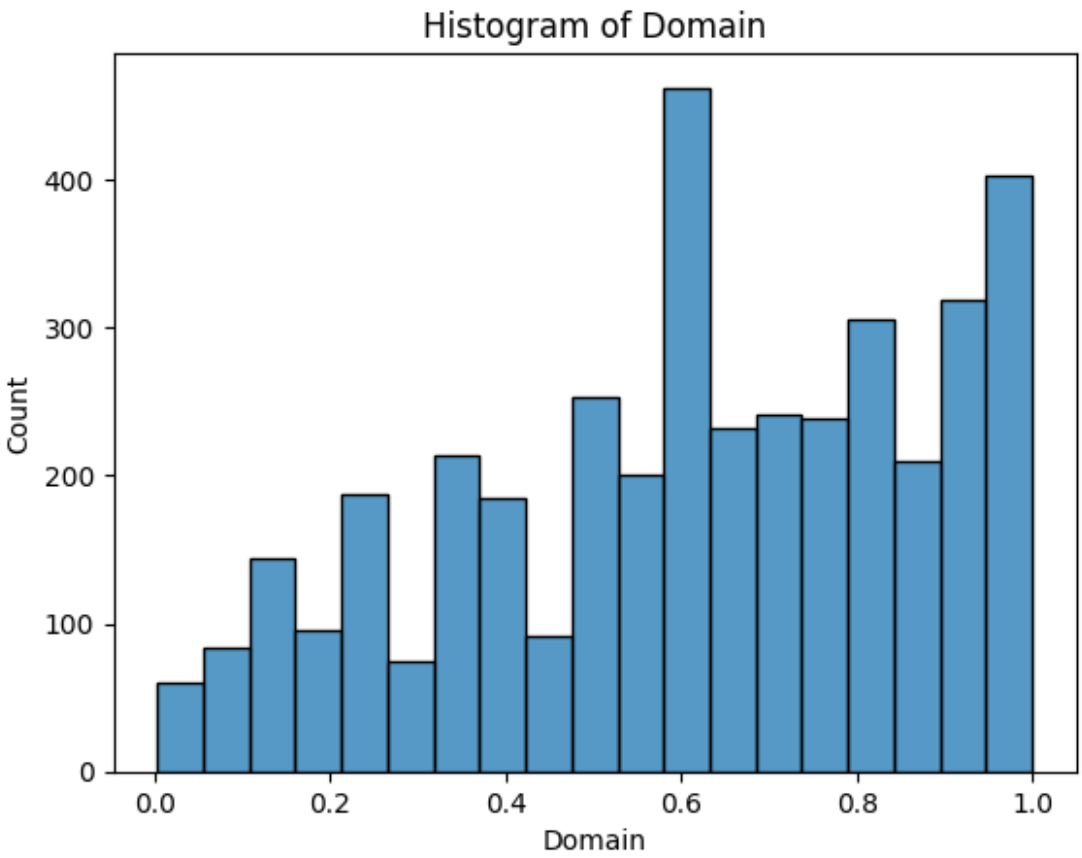

Histogram of CollegeCityID
Count
1200
1000
800
600
400
200
0
0
2500
5000
7500
10000
12500
15000
17500
CollegeCityID

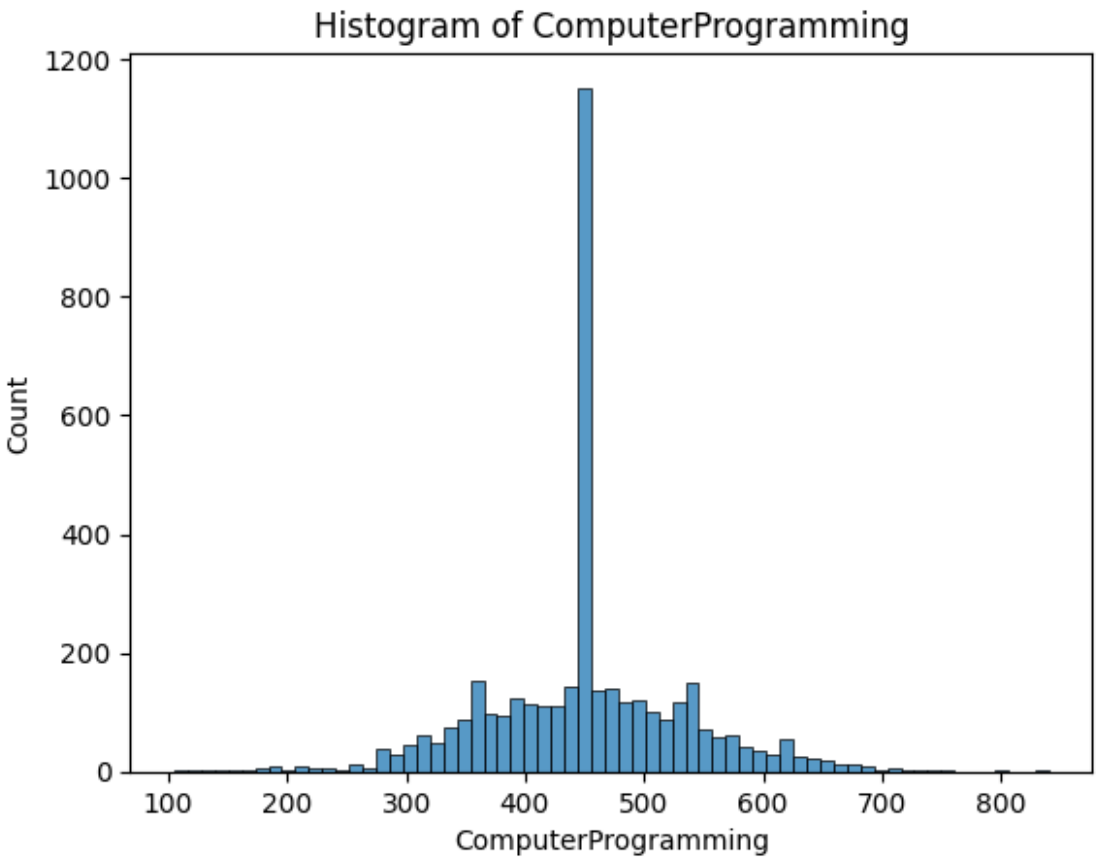

Histogram of GraduationYear
Count
1200
1000
800
600
400
200
0
0
250
500
750
1000
1250
1500
1750
2000
GraduationYear

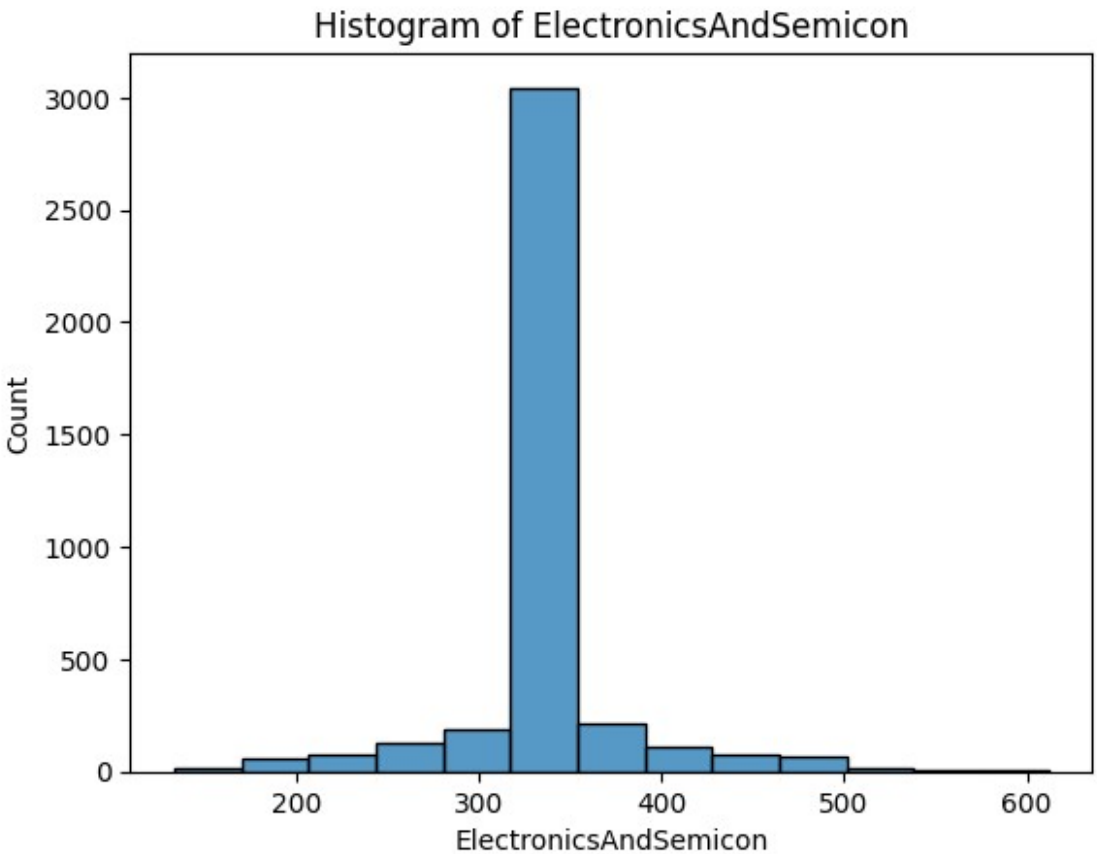

Histogram of English
Count
English
0
50
100
150
200
250
300
200
300
400
500
600
700
800
900

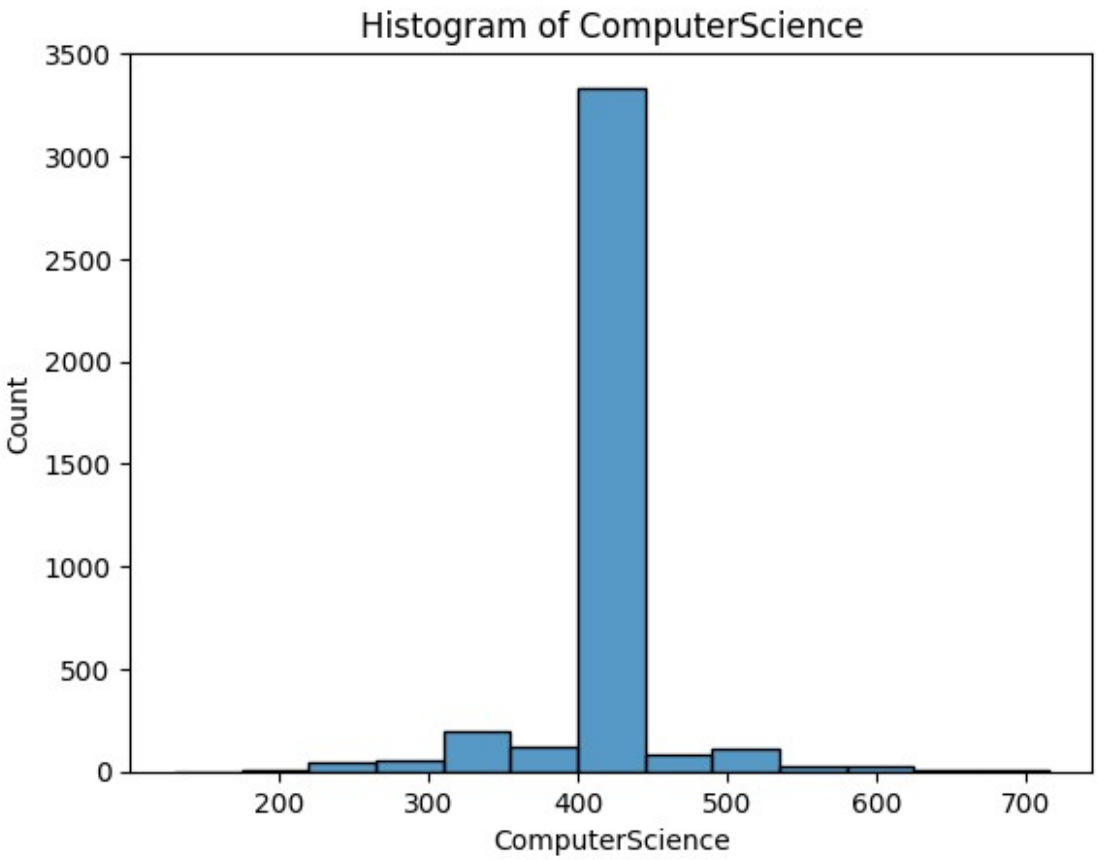

Histogram of Logical
Count
Logical
0
50
100
150
200
250
300
200
300
400
500
600
700
800

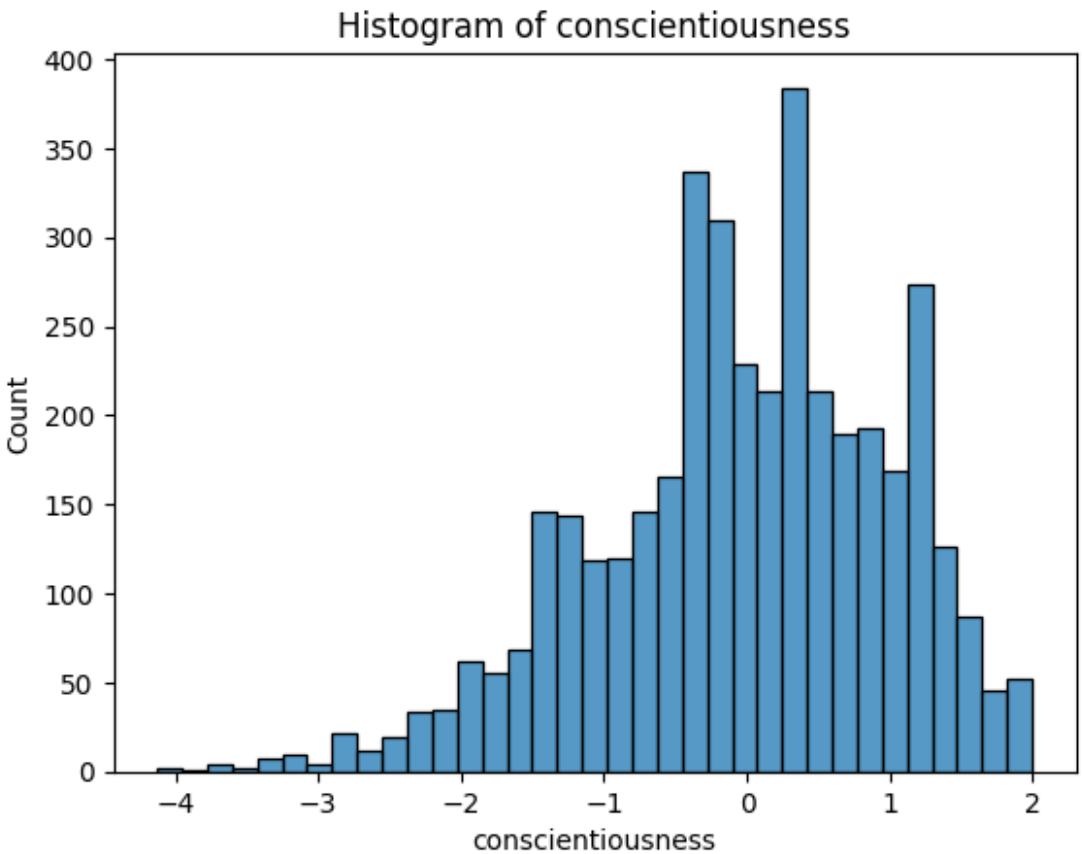

Histogram of Quant
Count
Quant
0
50
100
150
200
250
300
100
200
300
400
500
600
700
800
900

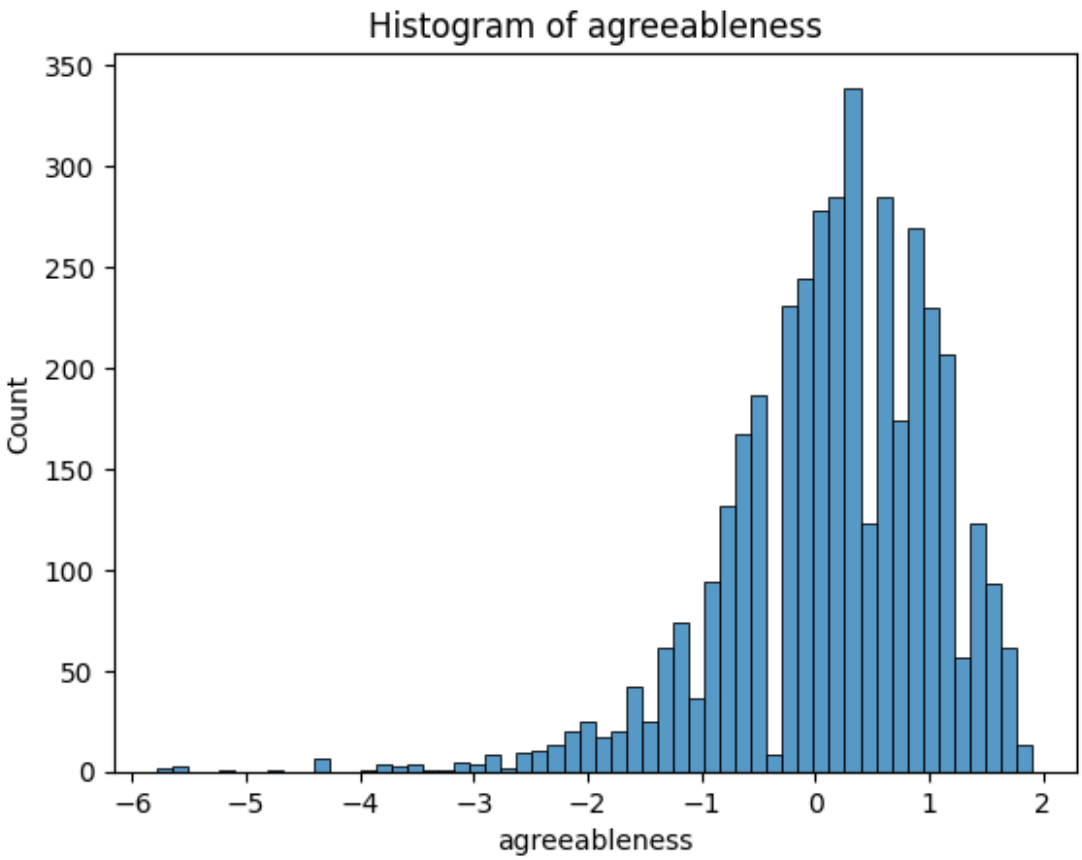

Histogram of Domain
Count
Domain
0
100
200
300
400
0.0
0.2
0.4
0.6
0.8
1.0

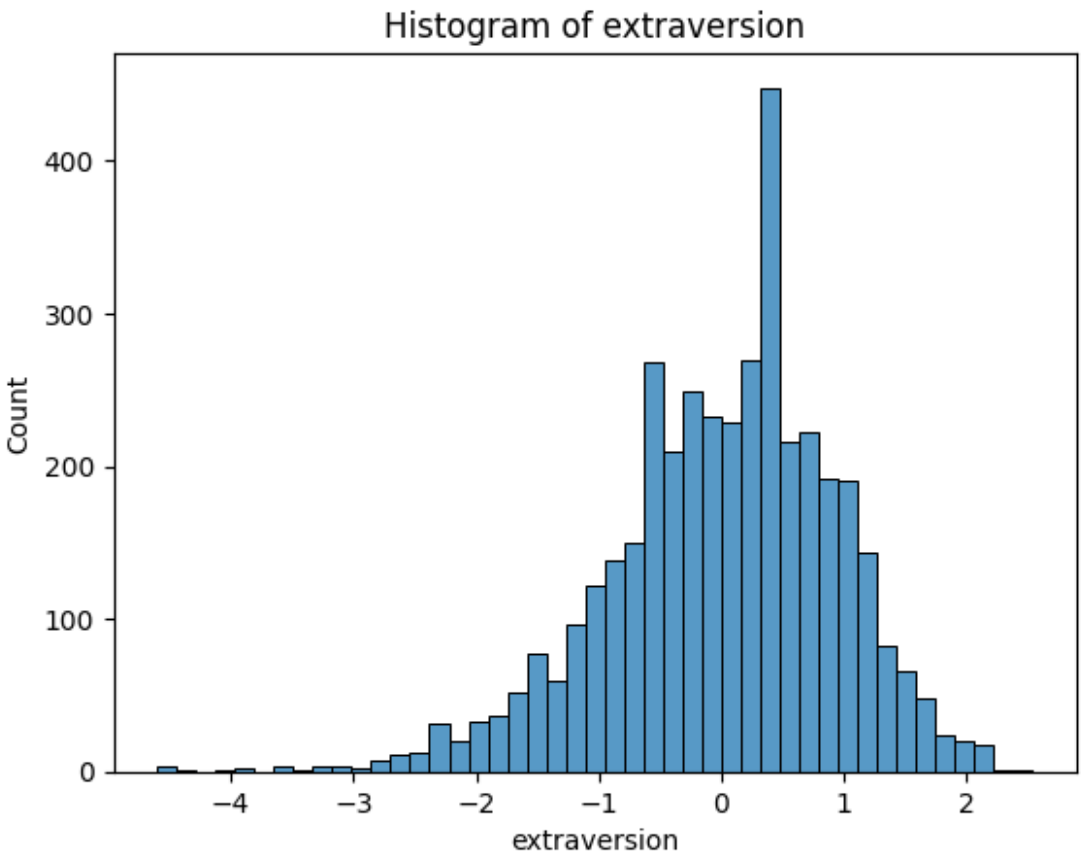

Histogram of ComputerProgramming
Count
ComputerProgramming
0
200
400
600
800
1000
1200
100
200
300
400
500
600
700
800

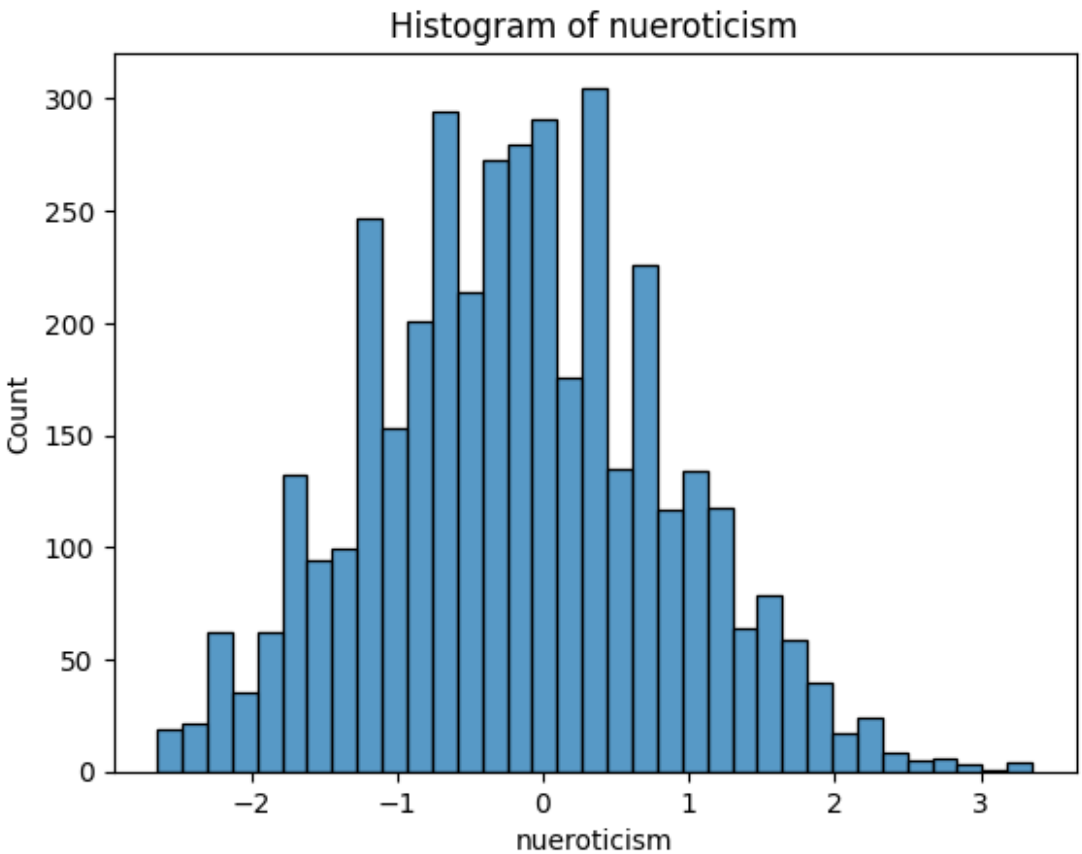

Histogram of ElectronicsAndSemicon
Count
3000
2500
2000
1500
1000
500
0
200
300
400
500
600
ElectronicsAndSemicon

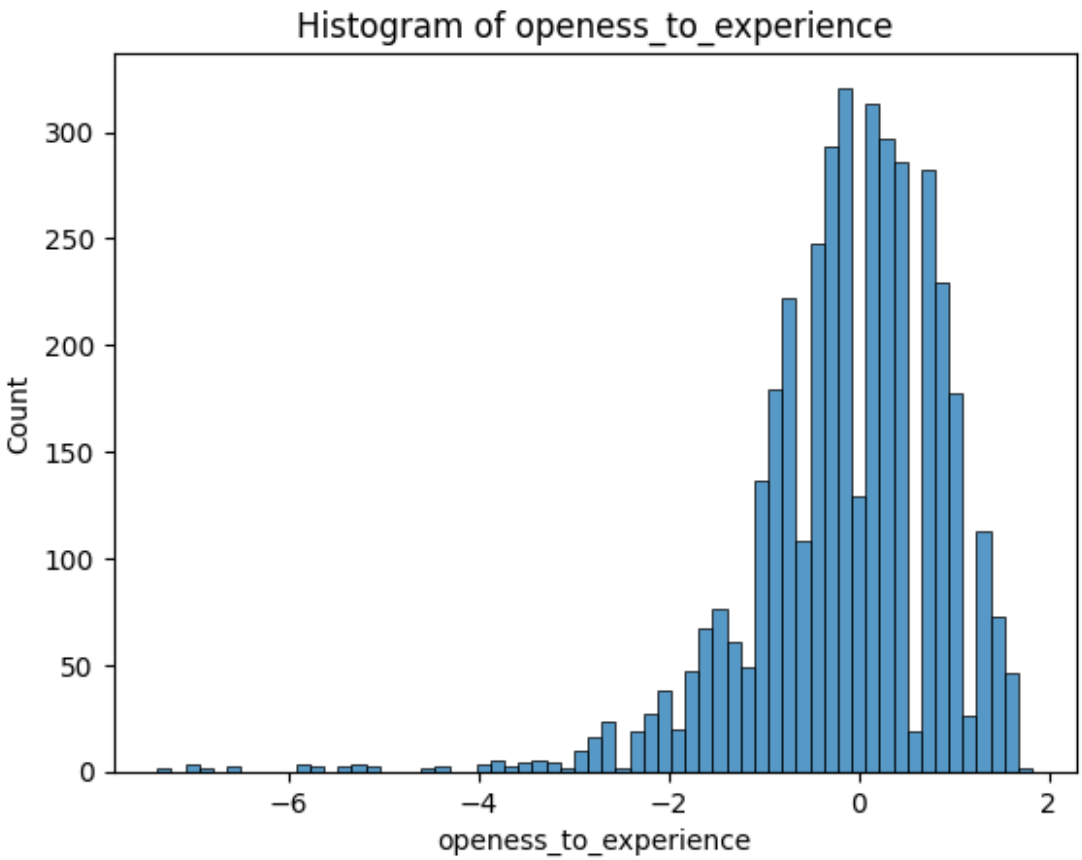

Histogram of ComputerScience
3500
3000
2500
2000
1500
1000
500
0
Count
200
300
400
500
600
700
ComputerScience

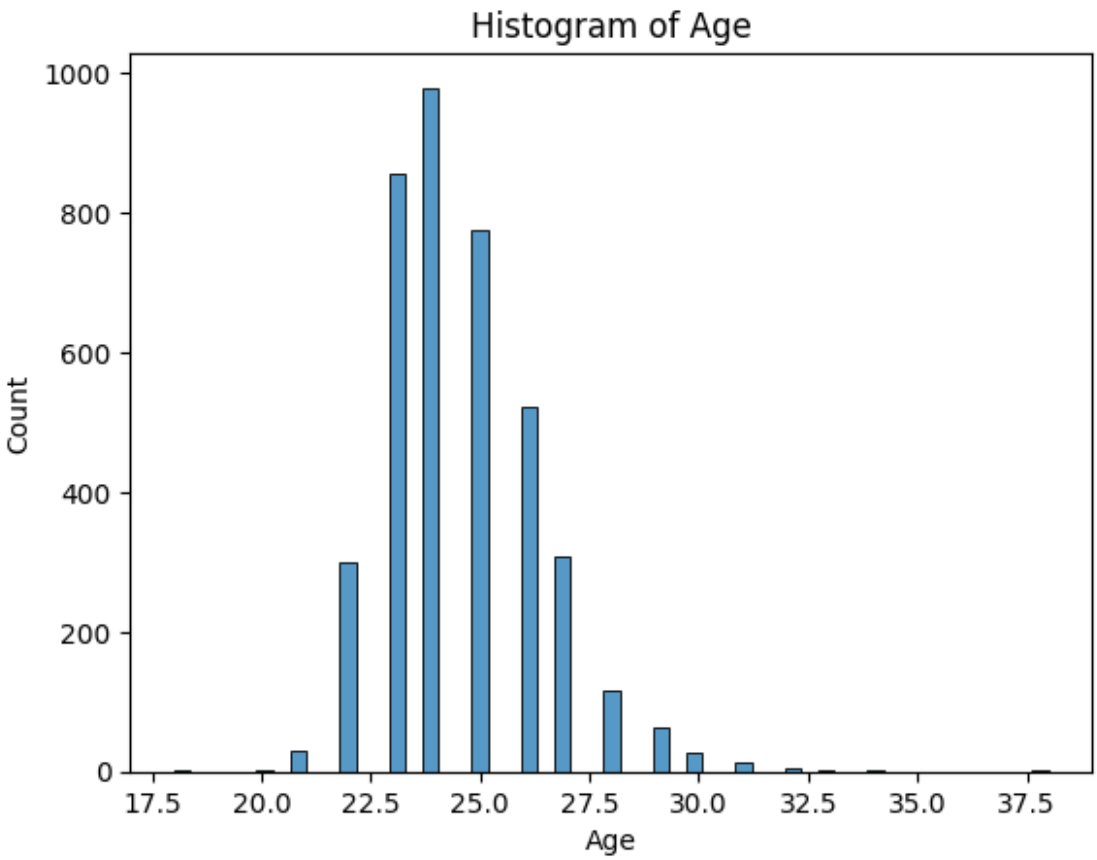

Histogram of conscientiousness
Count
conscientiousness
400
350
300
250
200
150
100
50
0
−4
−3
−2
−1
0
1
2


Histogram of agreeableness
Count
agreeableness
350
300
250
200
150
100
50
0
−6
−5
−4
−3
−2
−1
0
1
2


Histogram of extraversion
Count
400
300
200
100
0
−4
−3
−2
−1
0
1
2
extraversion

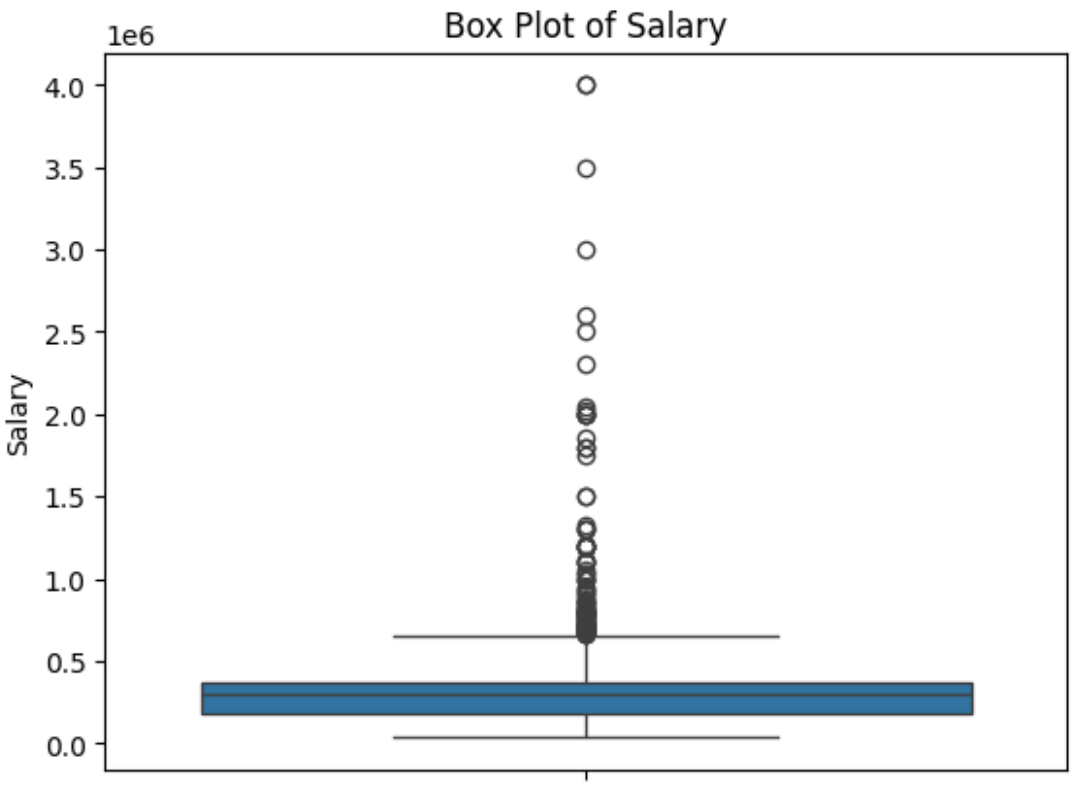

Histogram of nueroticism
Count
nueroticism
300
250
200
150
100
50
0
−2
−1
0
1
2
3

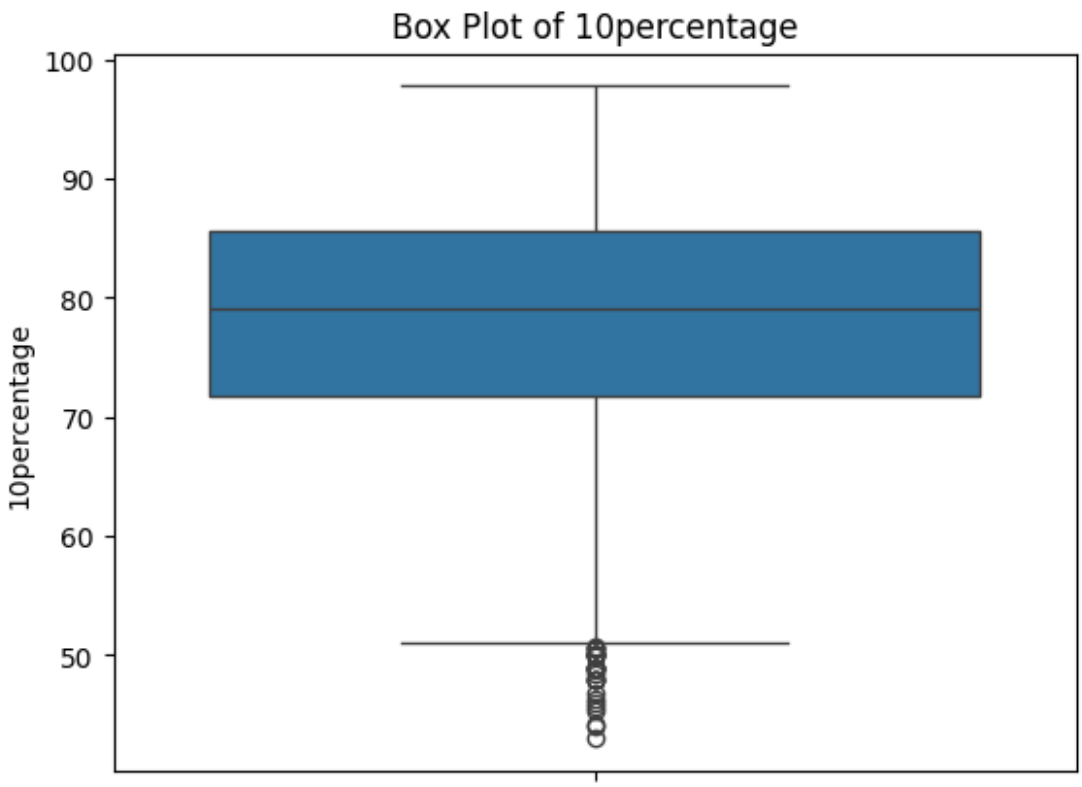

Histogram of openess_to_experience
Count
300
250
200
150
100
50
0
−6
−4
−2
0
2
openess_to_experience

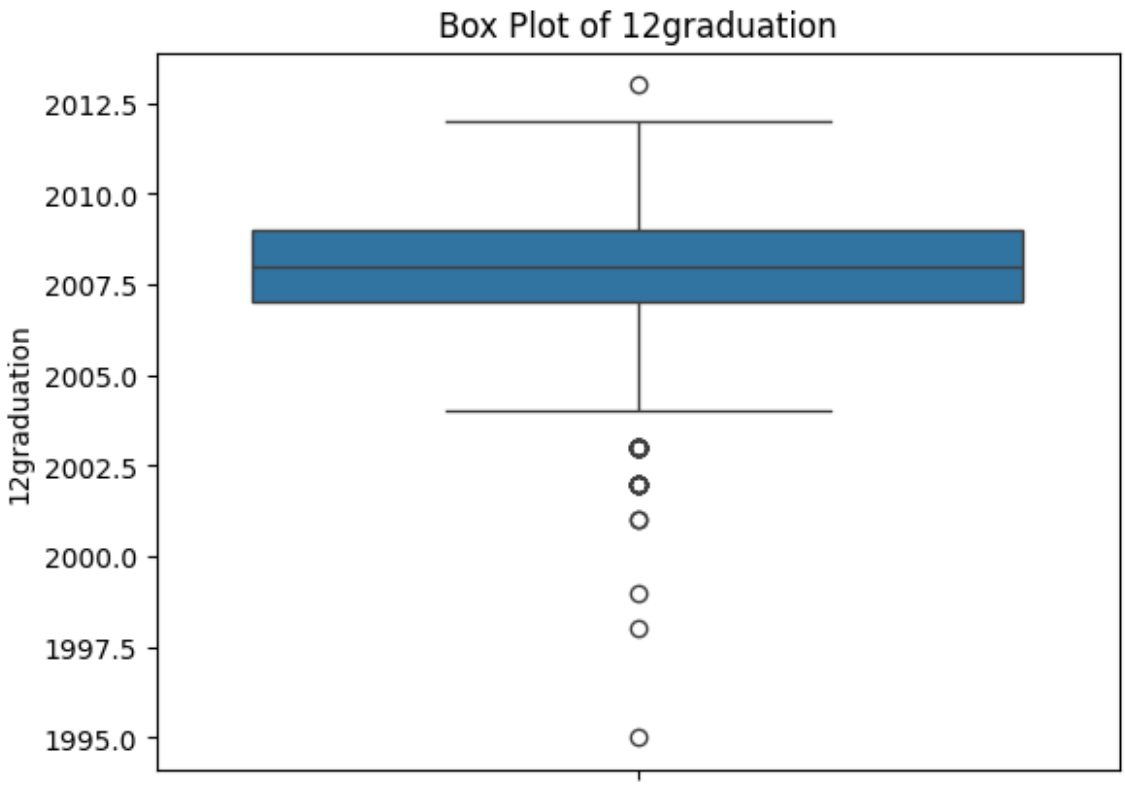


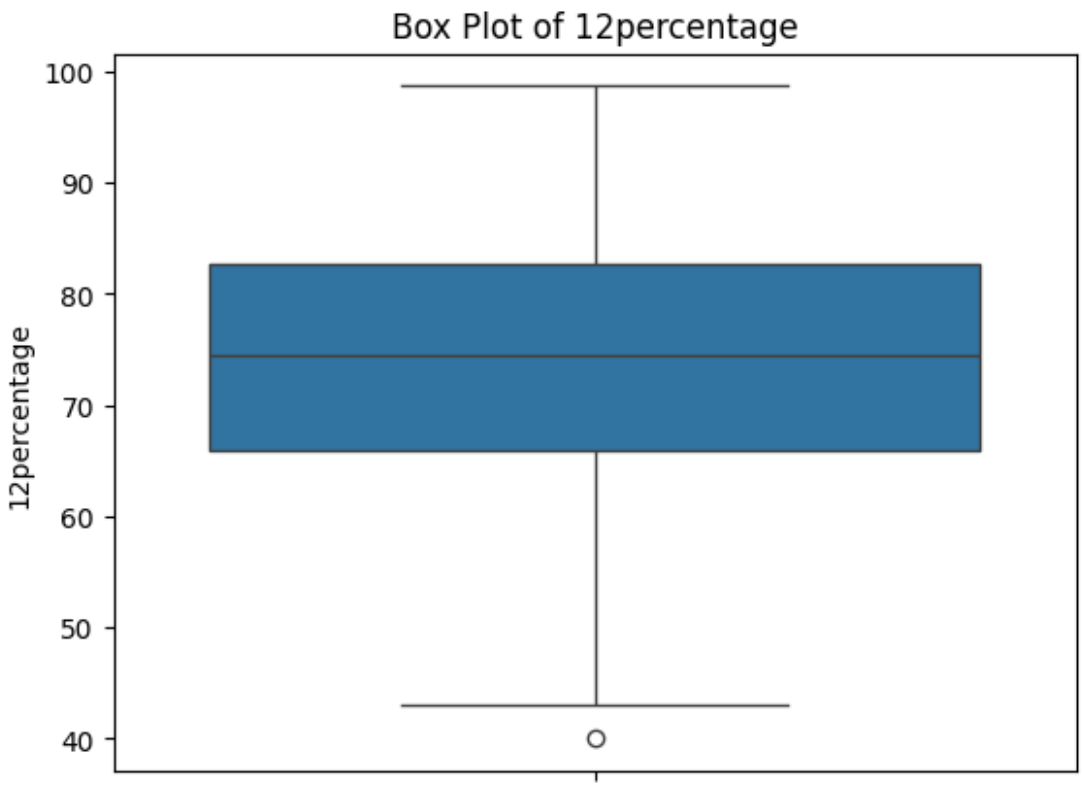

```
# prompt: observations from above histrogram

**Observations from the histograms:**

* **Age:** The distribution of age is right-skewed, with most people
being in their 20s and 30s.
* **Salary:** The distribution of salary is also right-skewed, with
most people earning less than 10 lakhs per annum.
* **ssc_p:** The distribution of ssc_p is approximately normal, with
most people scoring between 60% and 80%.
* **hsc_p:** The distribution of hsc_p is also approximately normal,
with most people scoring between 60% and 80%.
* **degree_p:** The distribution of degree_p is right-skewed, with
most people scoring between 60% and 80%.
* **etest_p:** The distribution of etest_p is approximately normal,
with most people scoring between 60% and 80%.
* **mba_p:** The distribution of mba_p is right-skewed, with most
people scoring between 60% and 80%.
* **workex:** The distribution of workex is right-skewed, with most
people having less than 5 years of experience.
* **specialisation:** The distribution of specialisation is
categorical, with most people having a specialisation in Marketing,
Finance, or HR.
```

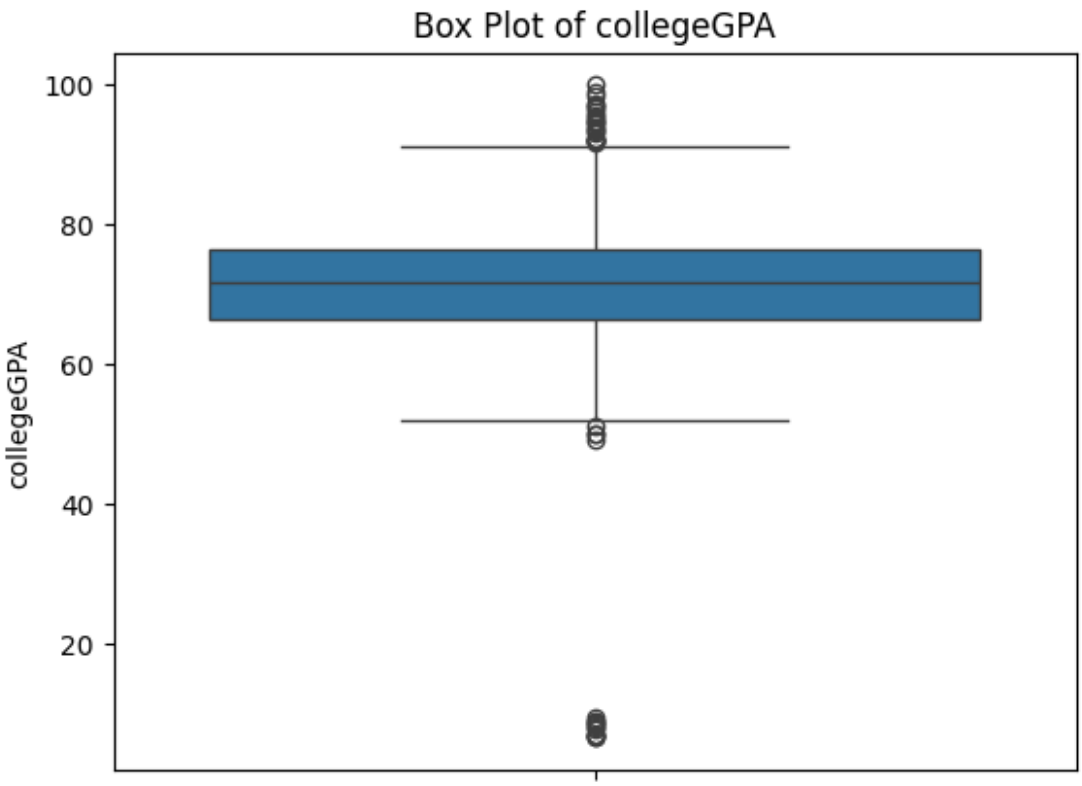
```
* **gender:** The distribution of gender is categorical, with slightly
more males than females.
* **ssc_b:** The distribution of ssc_b is categorical, with most
people having a board of CBSE or State Board.
* **hsc_b:** The distribution of hsc_b is categorical, with most
people having a board of CBSE or State Board.
* **degree_t:** The distribution of degree_t is categorical, with most
people having a degree in Engineering or Management.
* **workex:** The distribution of workex is categorical, with most
people having less than 5 years of experience.
* **mba:** The distribution of mba is categorical, with most people
not having an MBA.
* **status:** The distribution of status is categorical, with most
people being placed.

**Additional observations:**

* There are some outliers in the data, particularly for salary and
workex.
* The distributions of some of the categorical variables are
imbalanced. For example, there are more males than females in the
dataset.

# Create box plots for each numerical column
for col in numerical_columns:
    plt.figure()
    sns.boxplot(df[col])
    plt.title(f"Box Plot of {col}")
    plt.show()
```

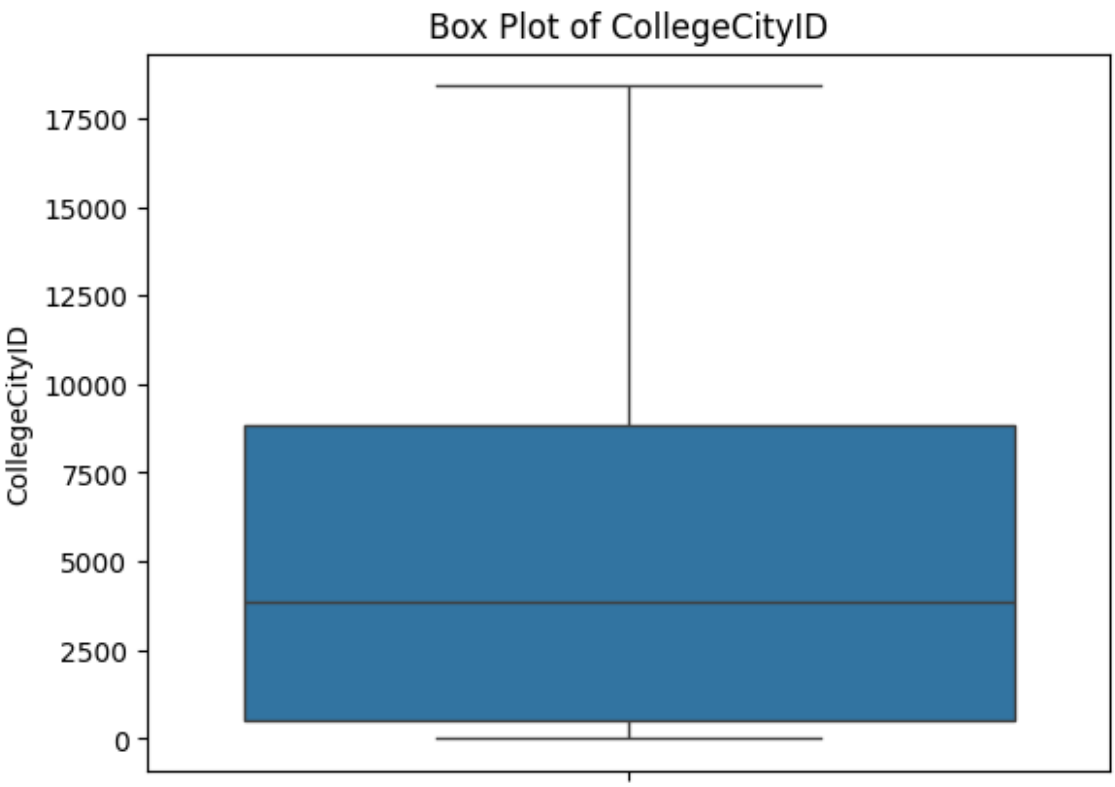

Box Plot of Salary
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Salary

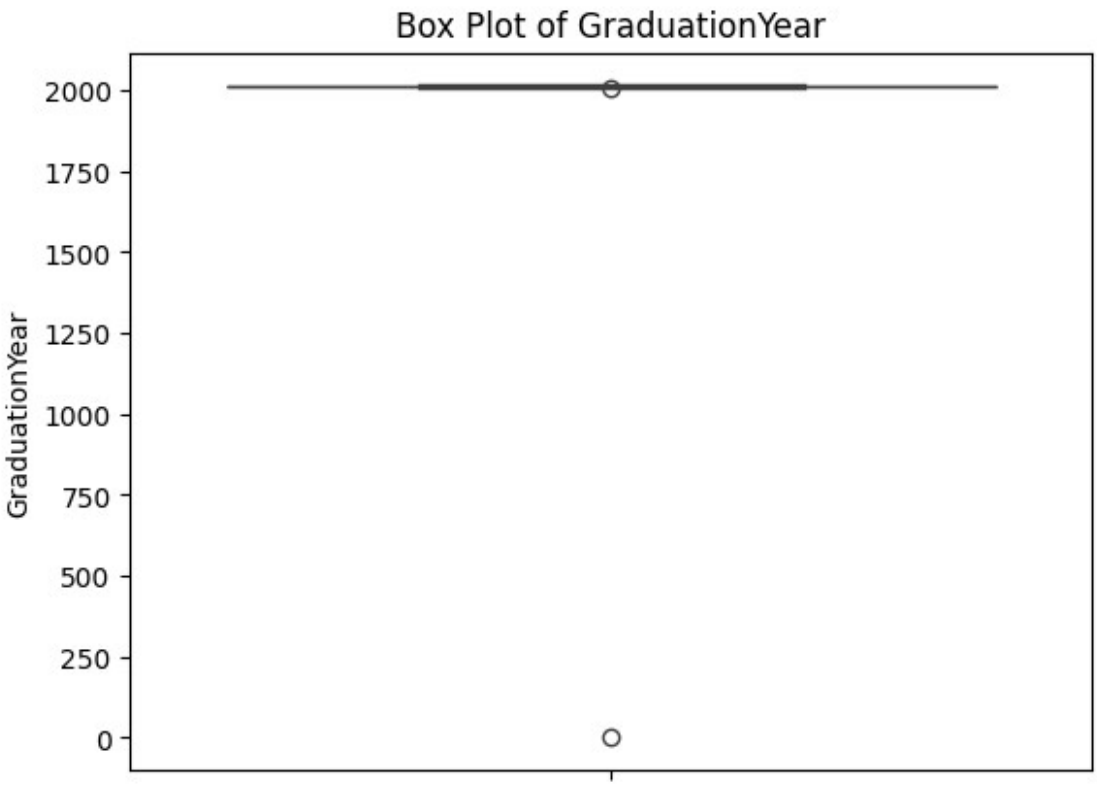

Box Plot of 10percentage
100
90
80
70
60
50
10percentage

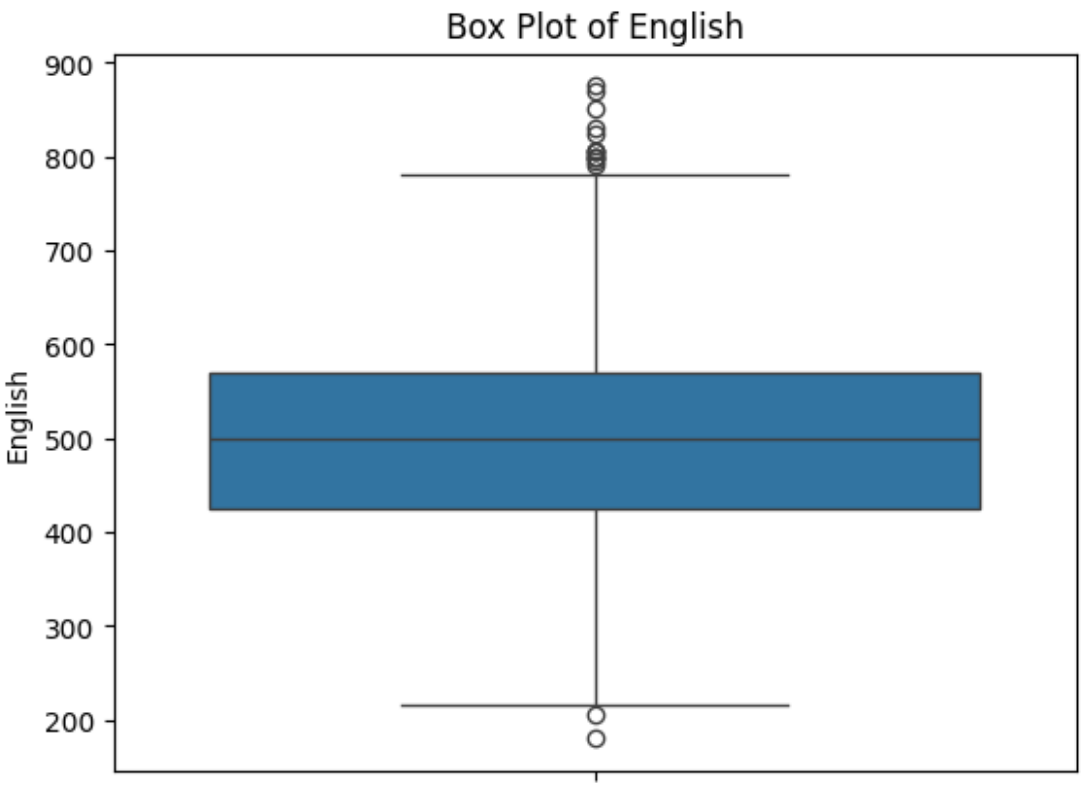

Box Plot of 12graduation
12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0

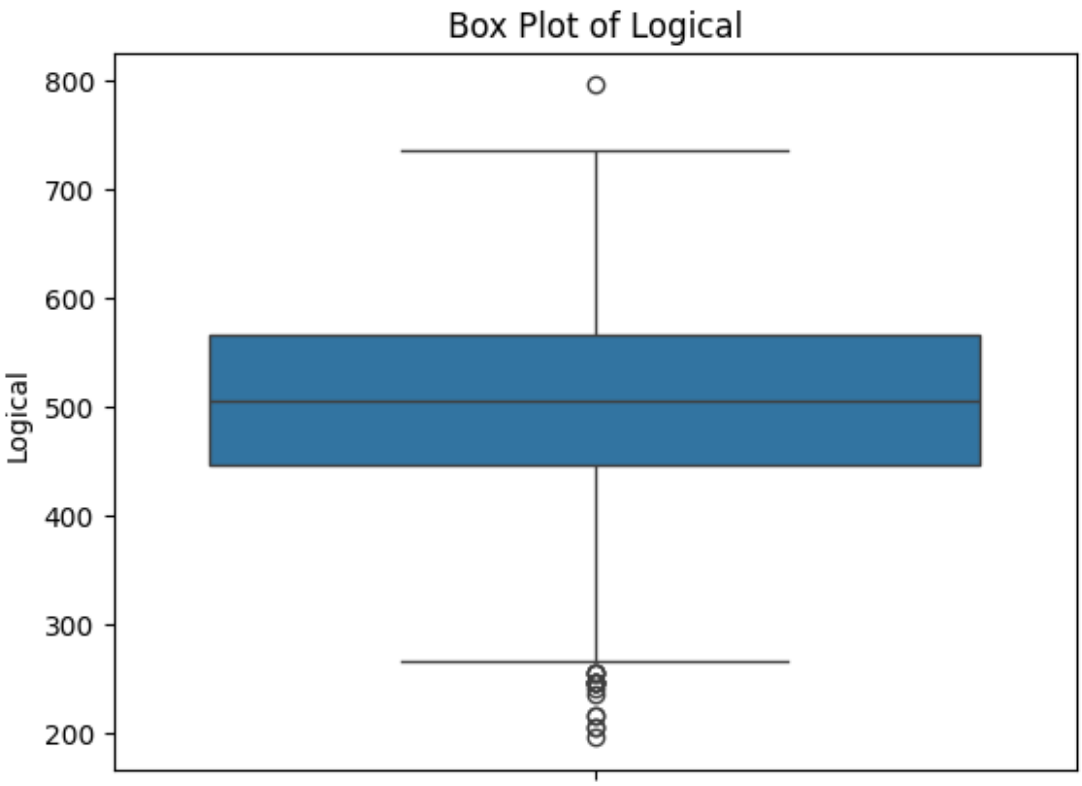

Box Plot of 12percentage
12percentage
100
90
80
70
60
50
40

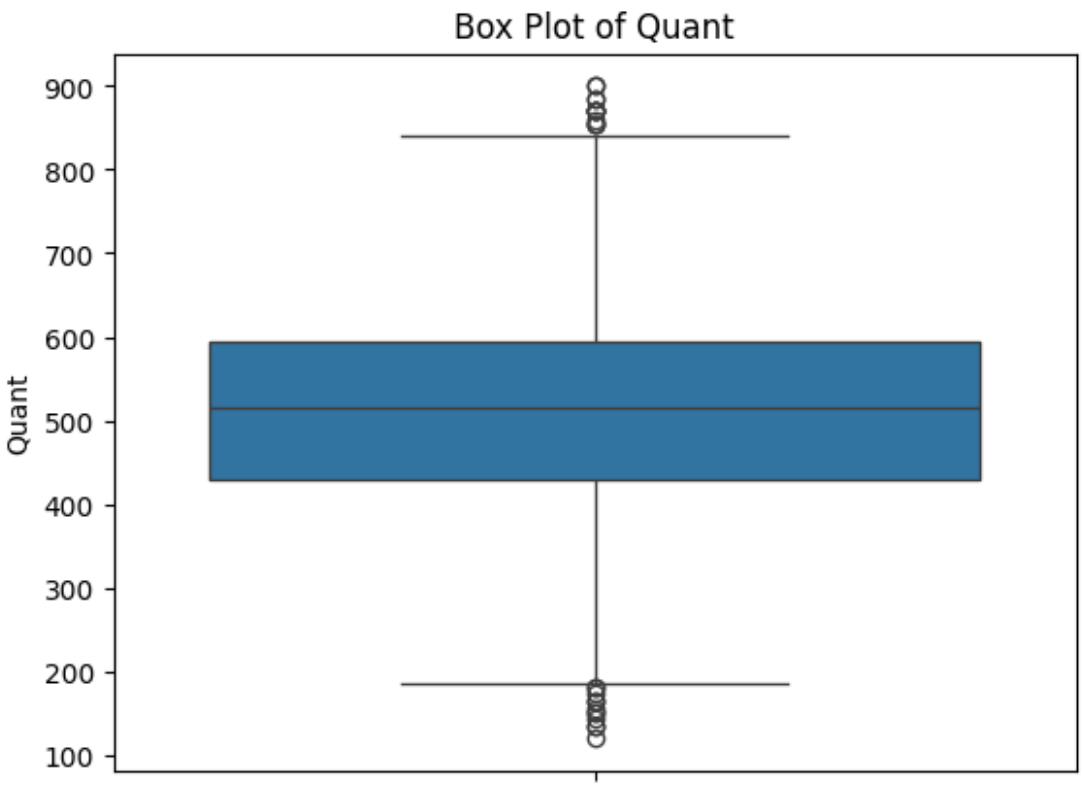

Box Plot of collegeGPA
collegeGPA
100
80
60
40
20

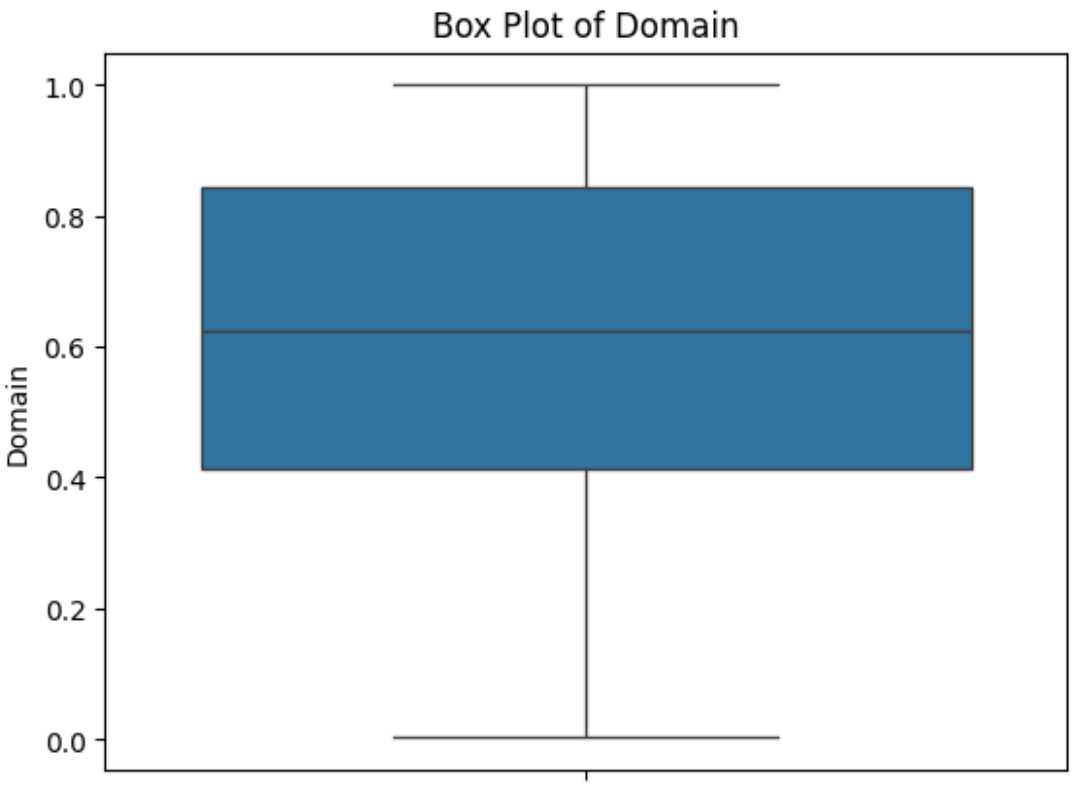

Box Plot of CollegeCityID
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0

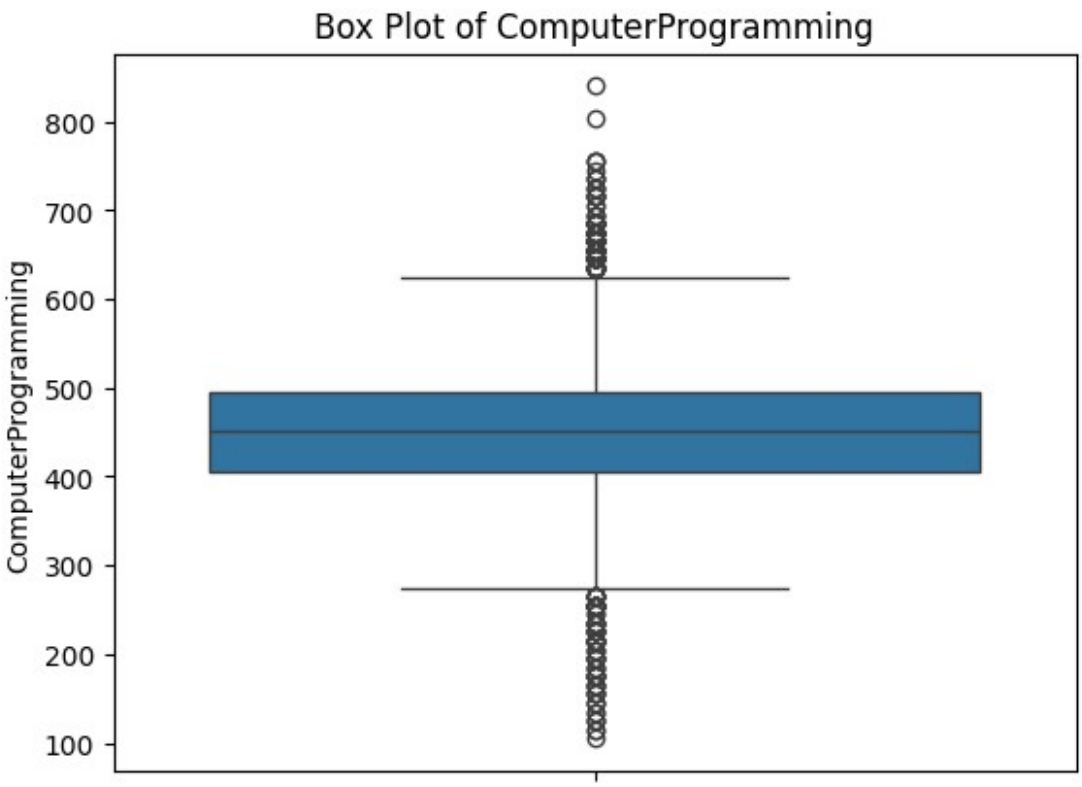

Box Plot of GraduationYear
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0

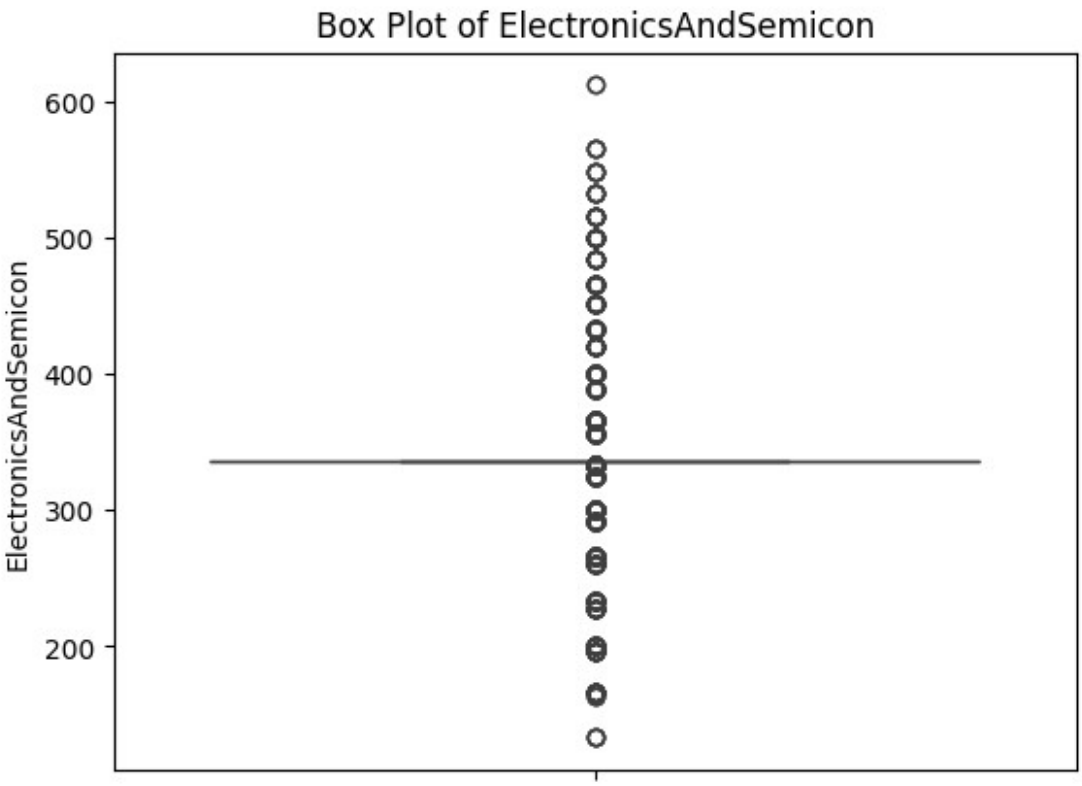

Box Plot of English
English
900
800
700
600
500
400
300
200

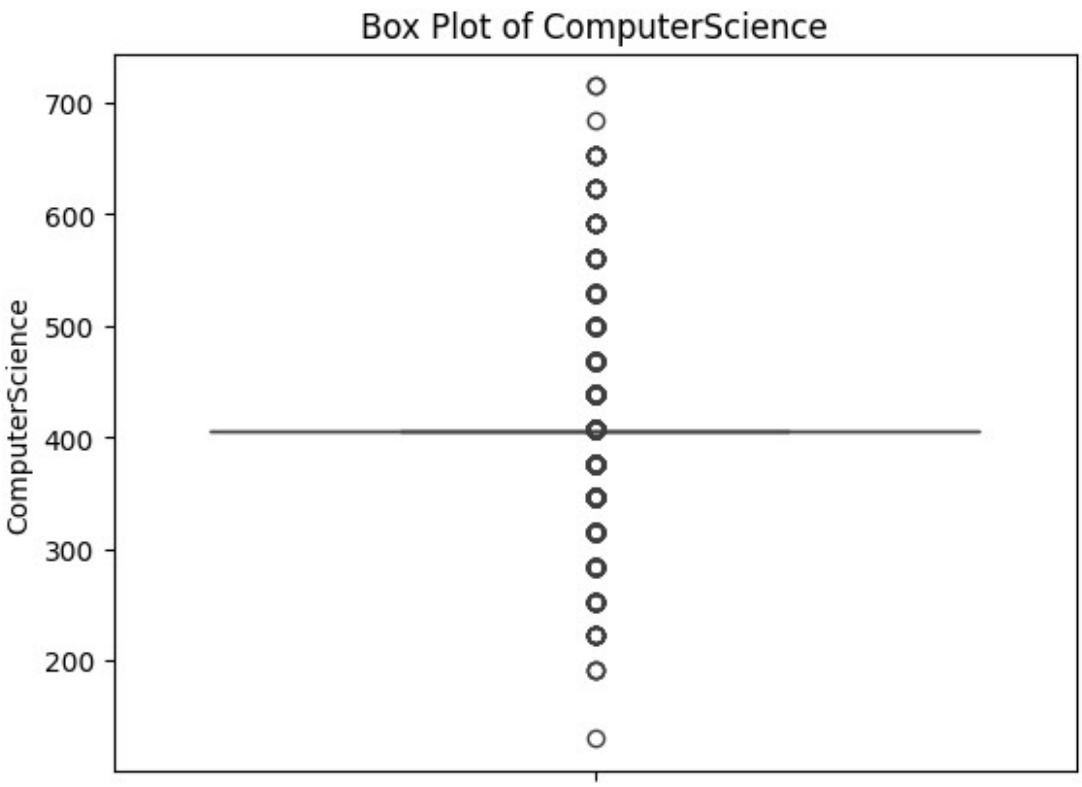

Box Plot of Logical
Logical
800
700
600
500
400
300
200

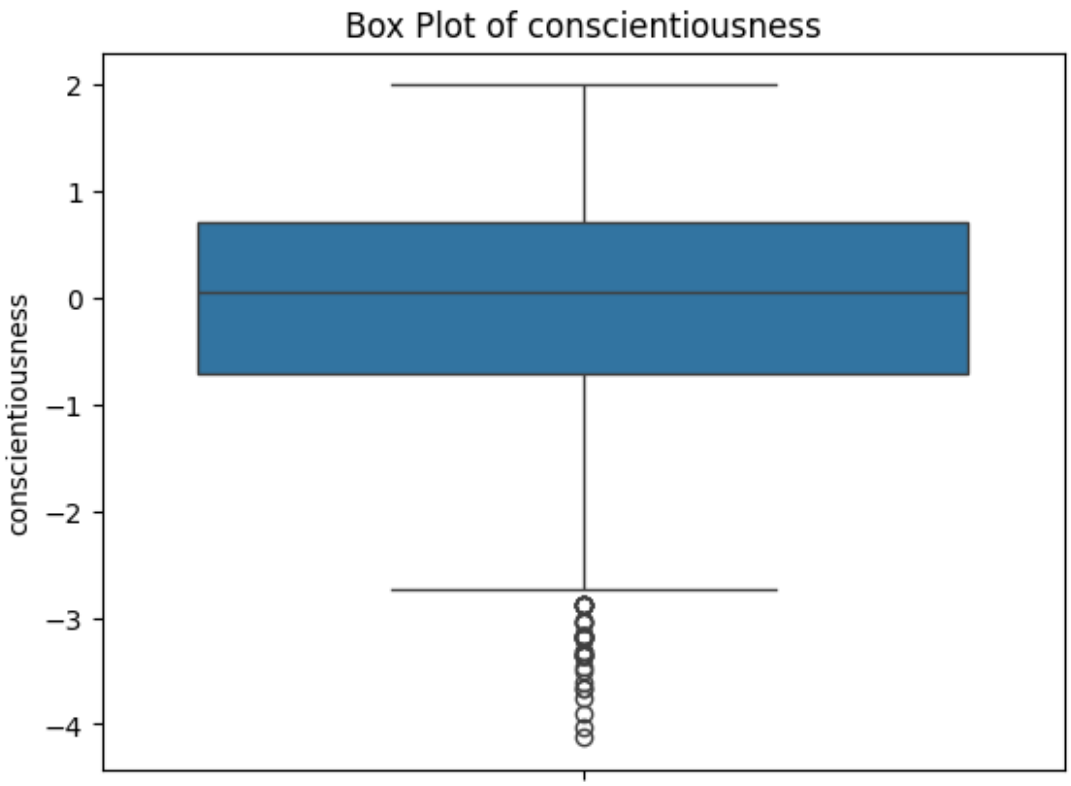

Box Plot of Quant
Quant
900
800
700
600
500
400
300
200
100

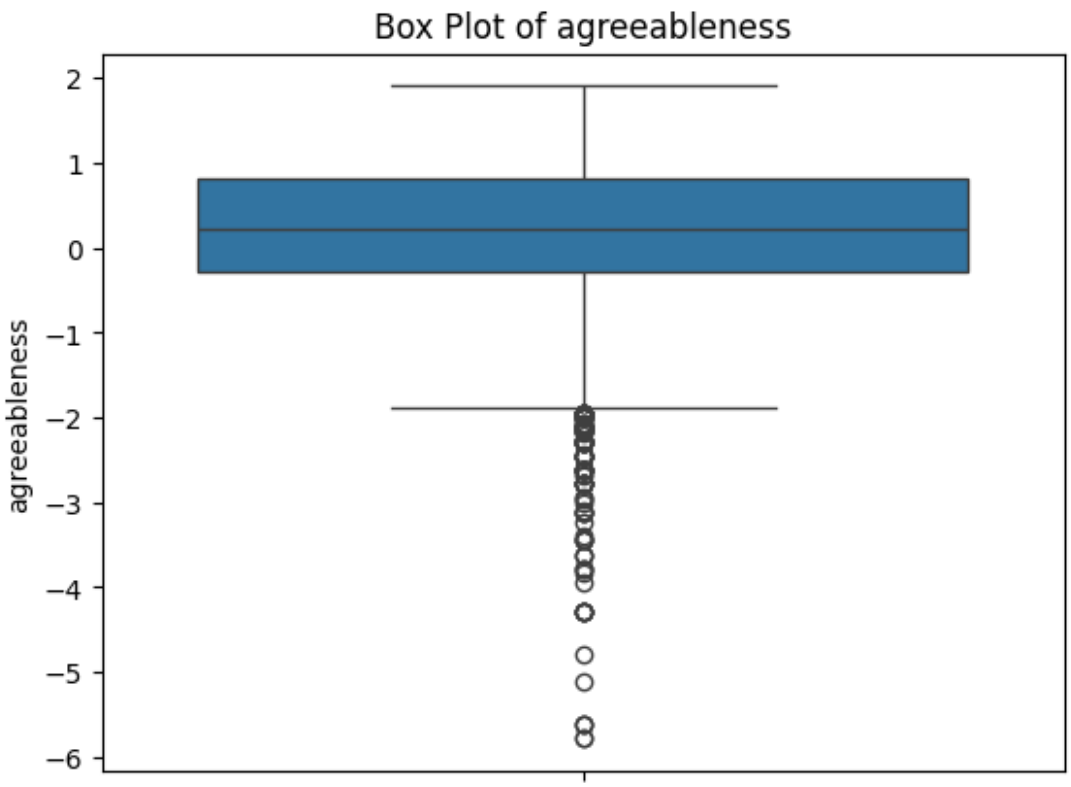

Box Plot of Domain
Domain
1.0
0.8
0.6
0.4
0.2
0.0

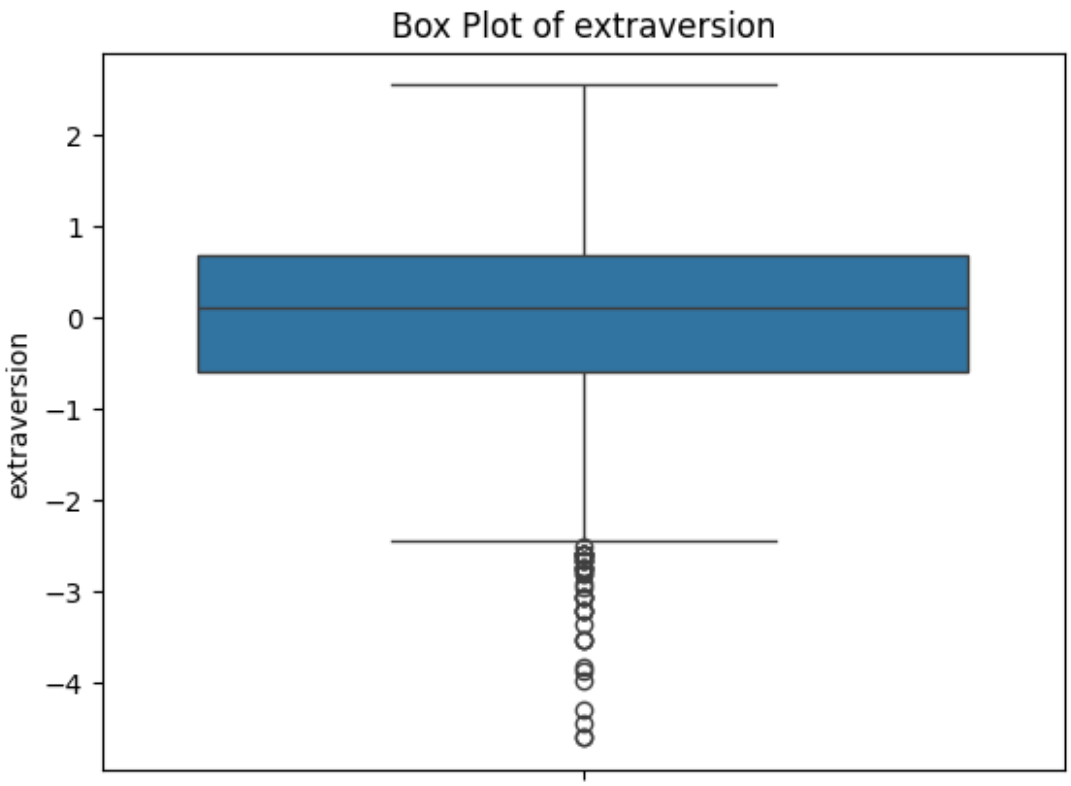

Box Plot of ComputerProgramming
ComputerProgramming
800
700
600
500
400
300
200
100

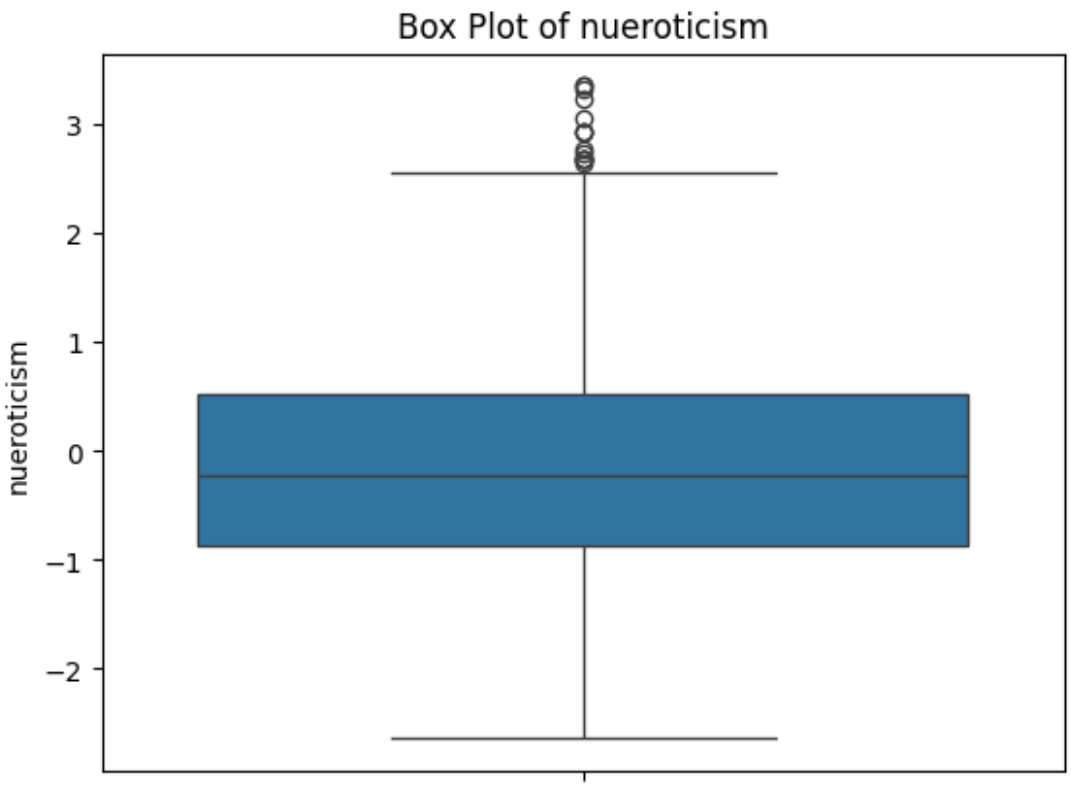

Box Plot of ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200

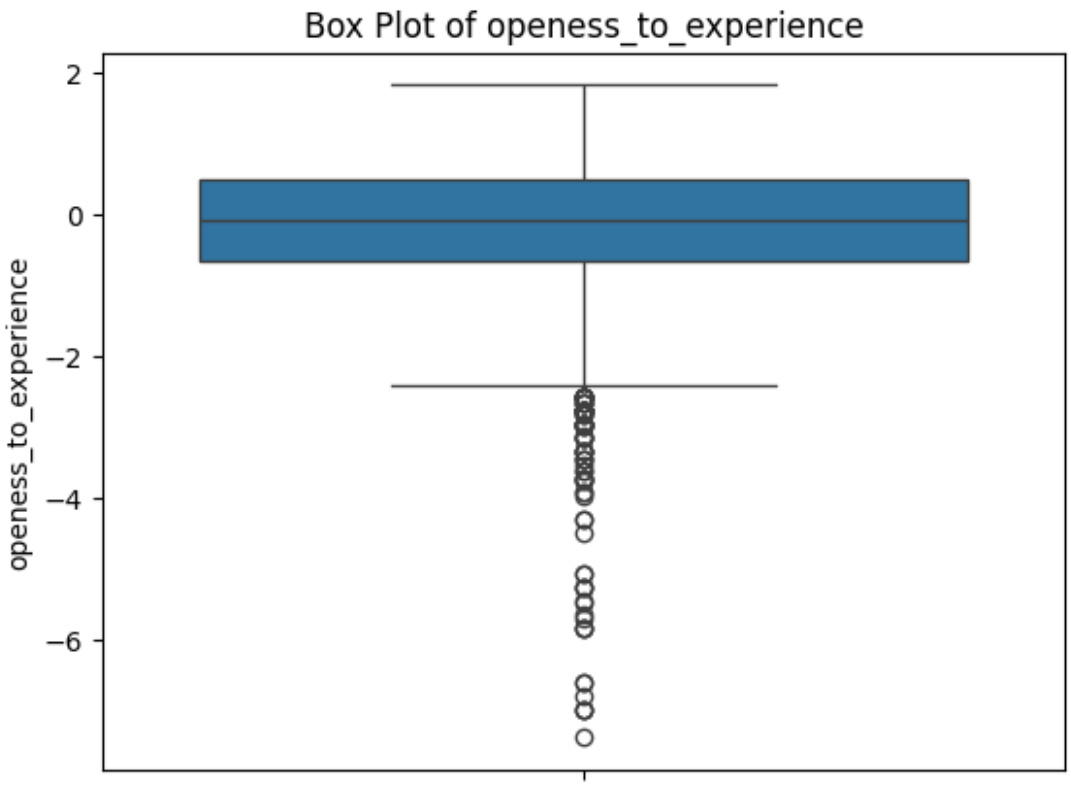

Box Plot of ComputerScience
ComputerScience
700
600
500
400
300
200

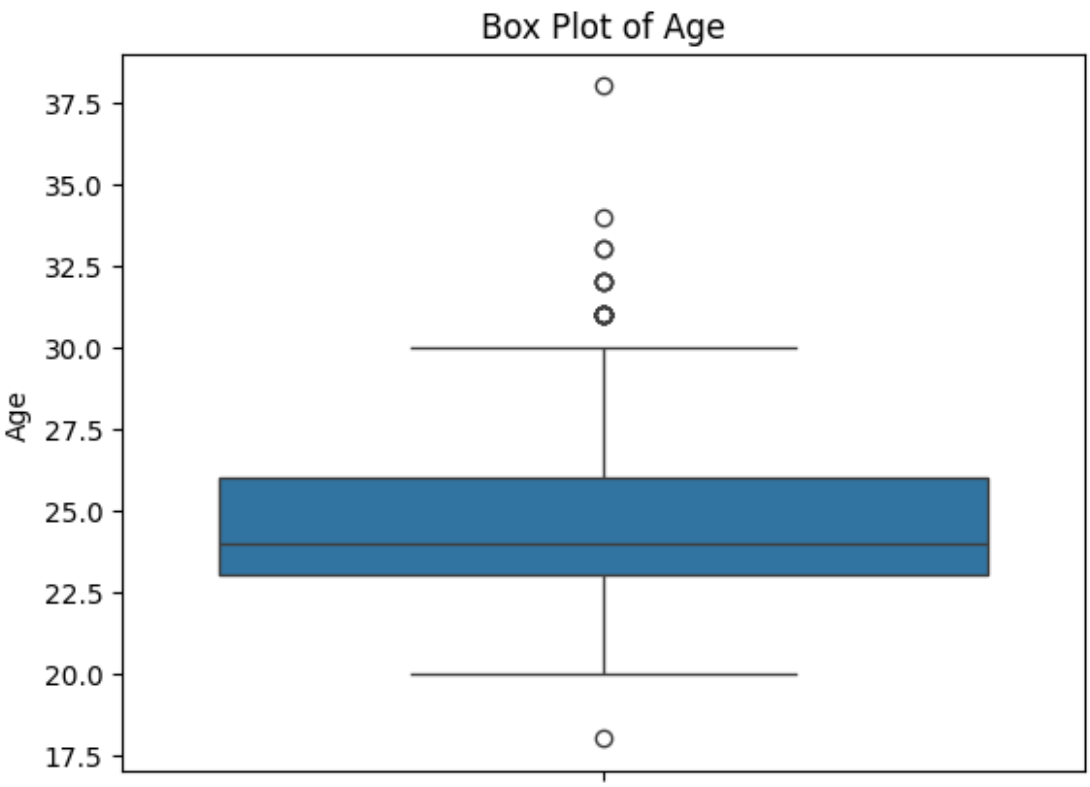

Box Plot of conscientiousness
conscientiousness
2
1
0
−1
−2
−3
−4


Box Plot of agreeableness
agreeableness
2
1
0
−1
−2
−3
−4
−5
−6

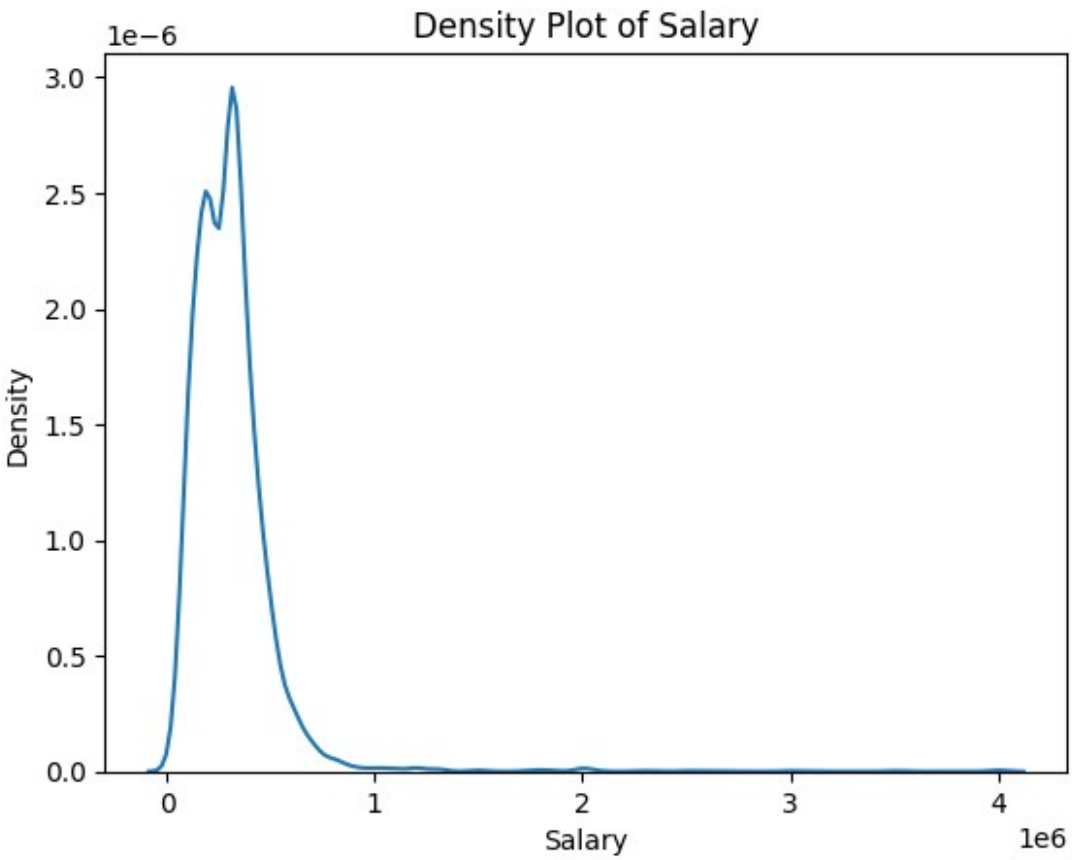

Box Plot of extraversion
extraversion
2
1
0
−1
−2
−3
−4

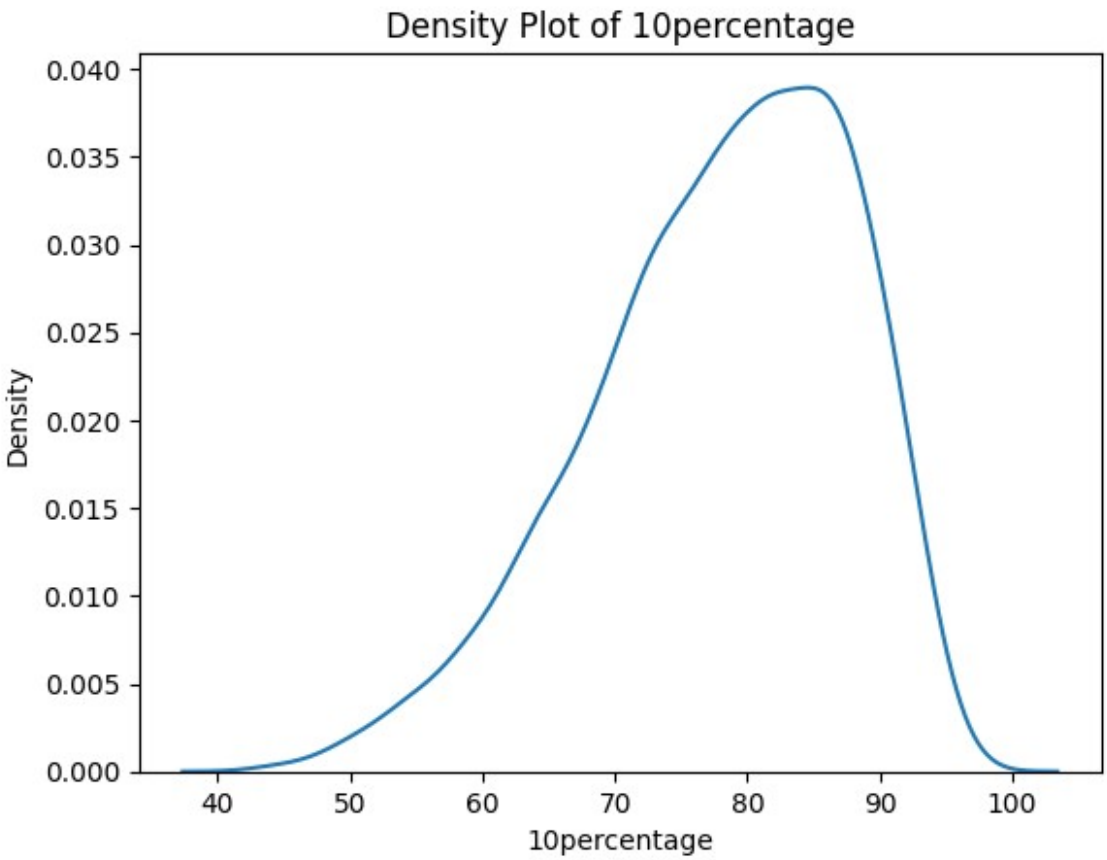

Box Plot of nueroticism
nueroticism
3
2
1
0
−1
−2

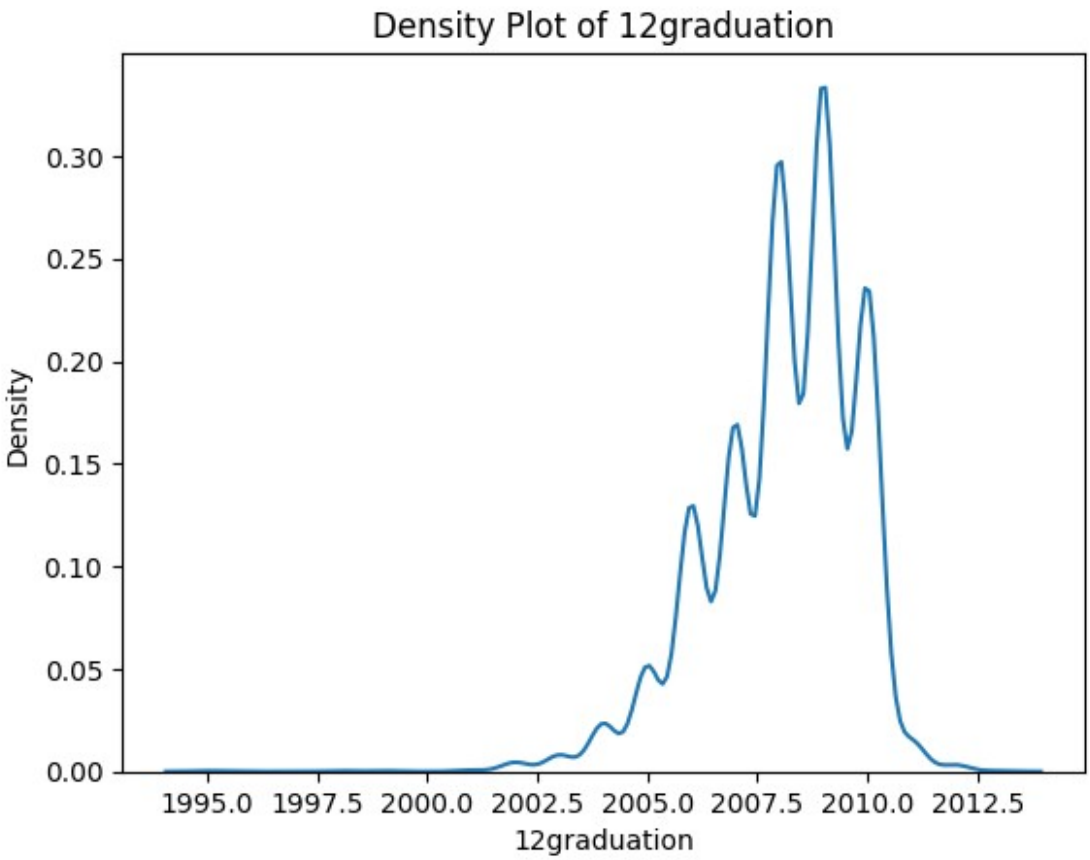

Box Plot of openness_to_experience
openness_to_experience
2
0
−2
−4
−6

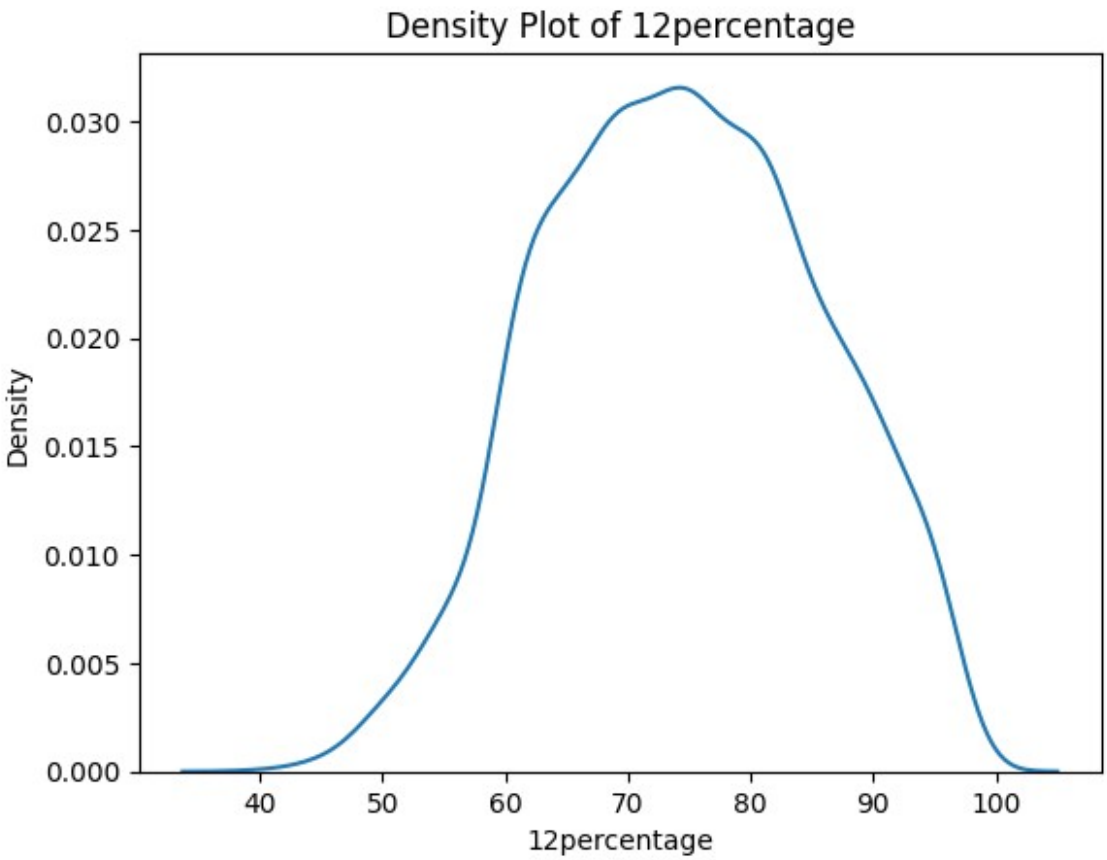

Box Plot of Age
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5

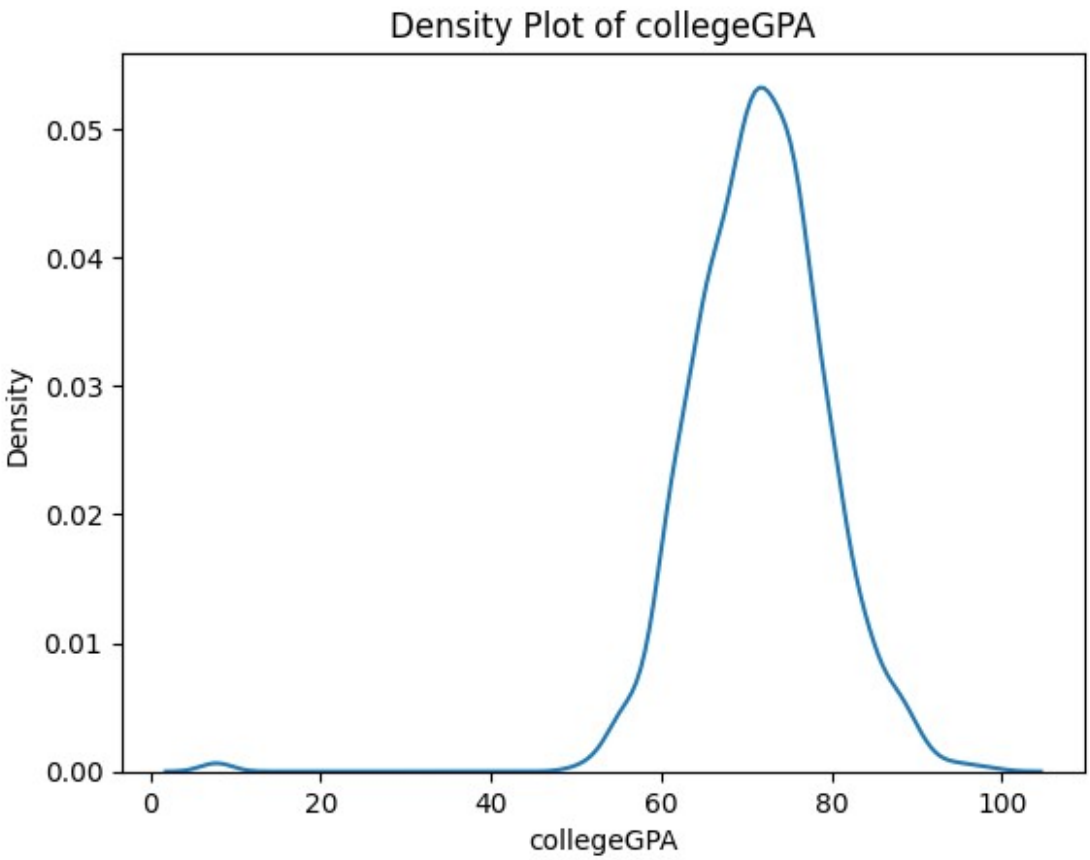
```
# prompt: observations from above box plot

**Observations from the box plots:**

* **Age:** The median age is around 24 years. There are some outliers
with ages above 40 years.
* **Salary:** The median salary is around 5 lakhs per annum. There are
some outliers with salaries above 20 lakhs per annum.
* **ssc_p:** The median ssc_p is around 70%. There are some outliers
with ssc_p below 50%.
* **hsc_p:** The median hsc_p is around 70%. There are some outliers
with hsc_p below 50%.
* **degree_p:** The median degree_p is around 70%. There are some
outliers with degree_p below 50%.
* **etest_p:** The median etest_p is around 70%. There are some
outliers with etest_p below 50%.
* **mba_p:** The median mba_p is around 70%. There are some outliers
with mba_p below 50%.
* **workex:** The median workex is around 2 years. There are some
outliers with workex above 10 years.

**Additional observations:**

* The distributions of some of the variables are skewed. For example,
the distribution of salary is right-skewed.
* There are some outliers in the data, particularly for salary and
workex.

# Create density plots for each numerical column
for col in numerical_columns:
    plt.figure()
    sns.kdeplot(df[col])
    plt.title(f"Density Plot of {col}")
    plt.show()
```

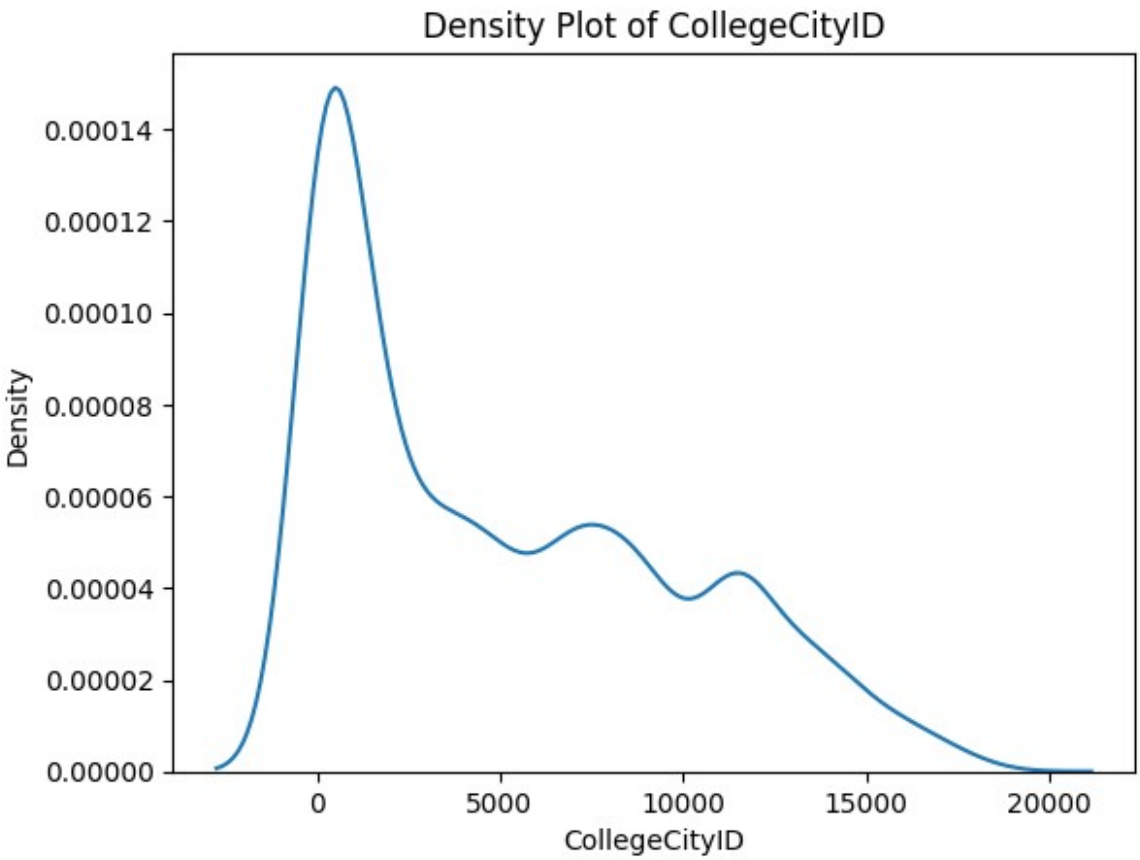

Density Plot of Salary
1e−6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Density
0
1
2
3
4
Salary
1e6

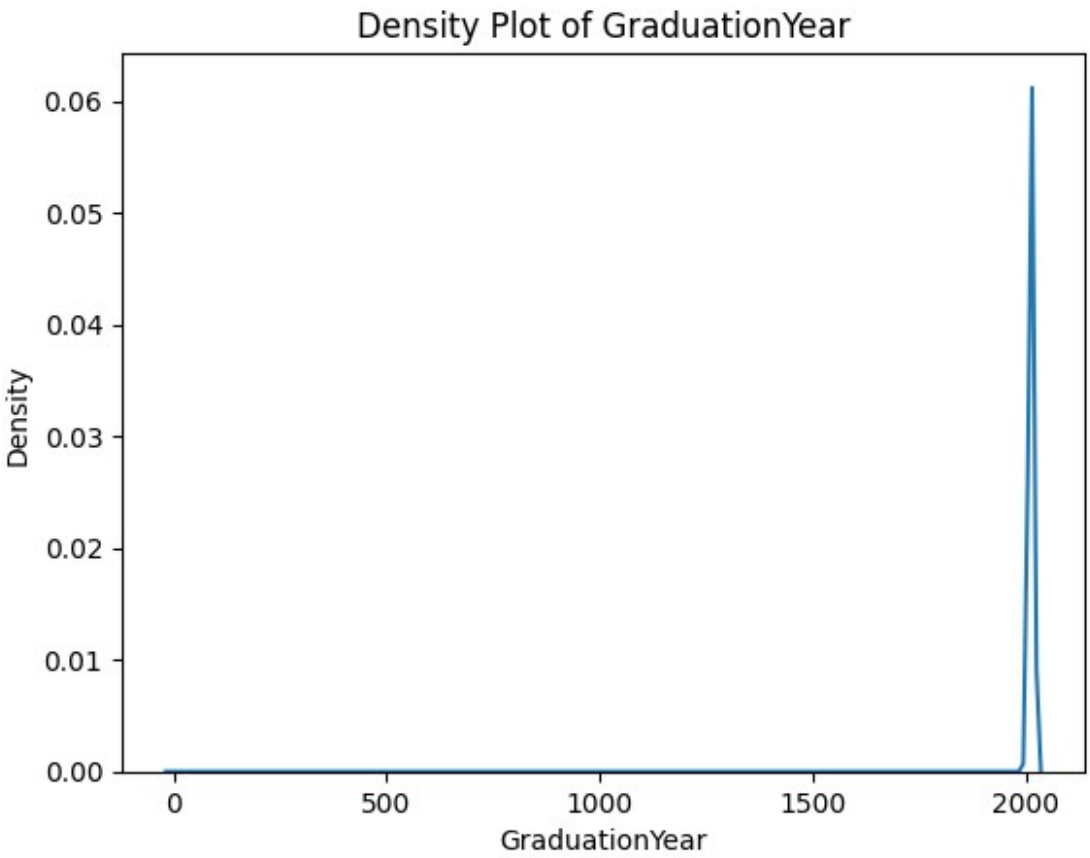

Density Plot of 10percentage
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
Density
40
50
60
70
80
90
100
10percentage

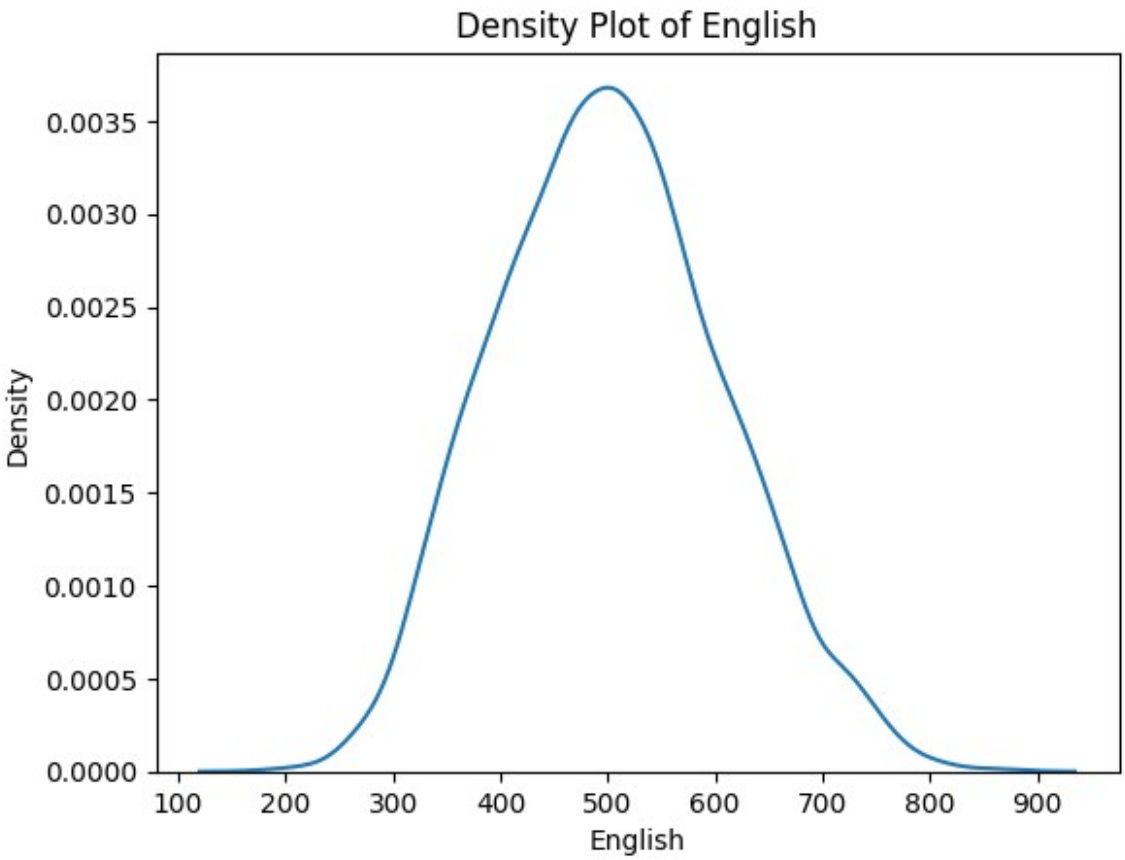

Density Plot of 12graduation
Density
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

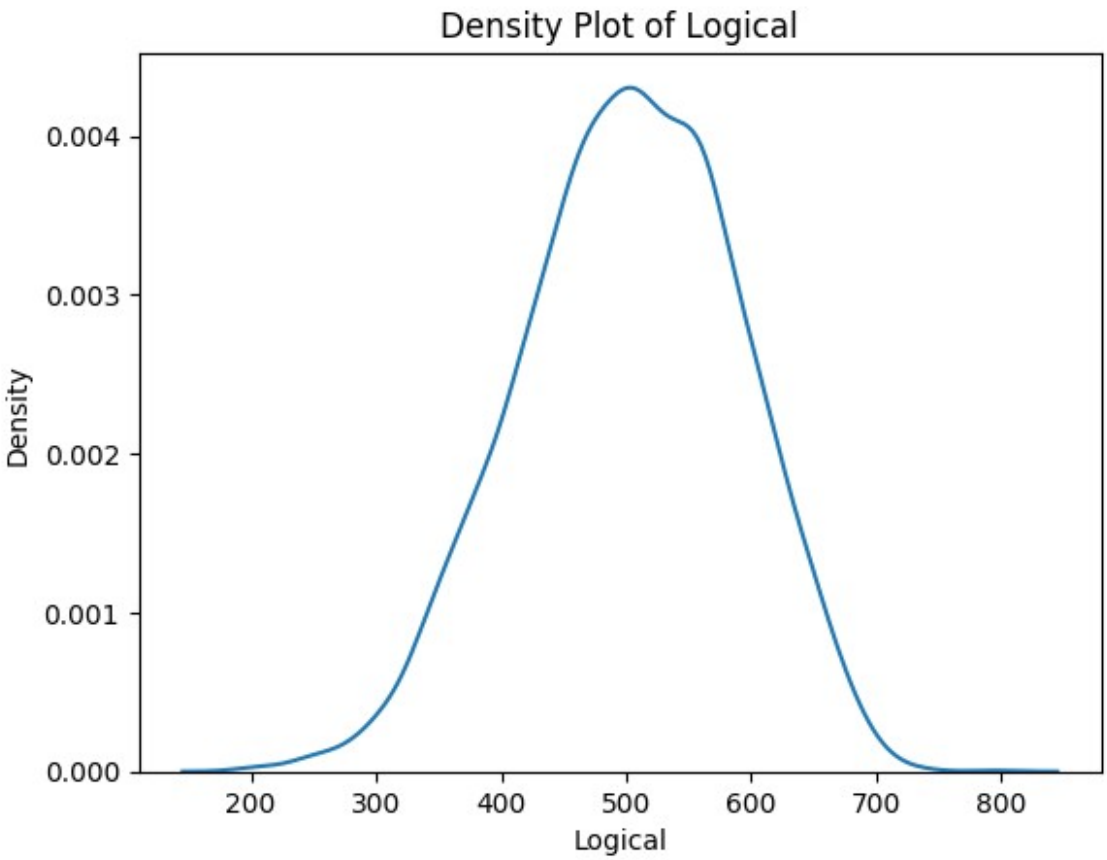

Density Plot of 12percentage
Density
0.030
0.025
0.020
0.015
0.010
0.005
0.000
40
50
60
70
80
90
100
12percentage

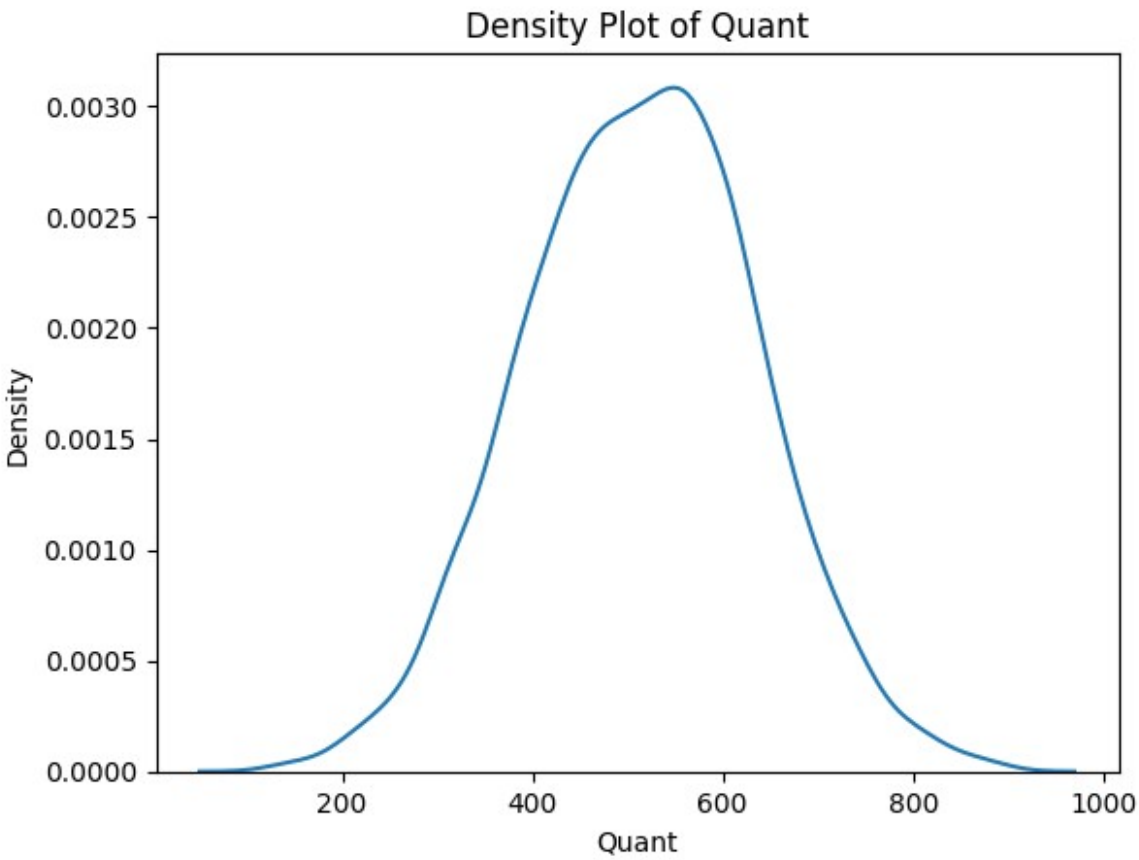

Density Plot of collegeGPA
Density
0.05
0.04
0.03
0.02
0.01
0.00
0
20
40
60
80
100
collegeGPA

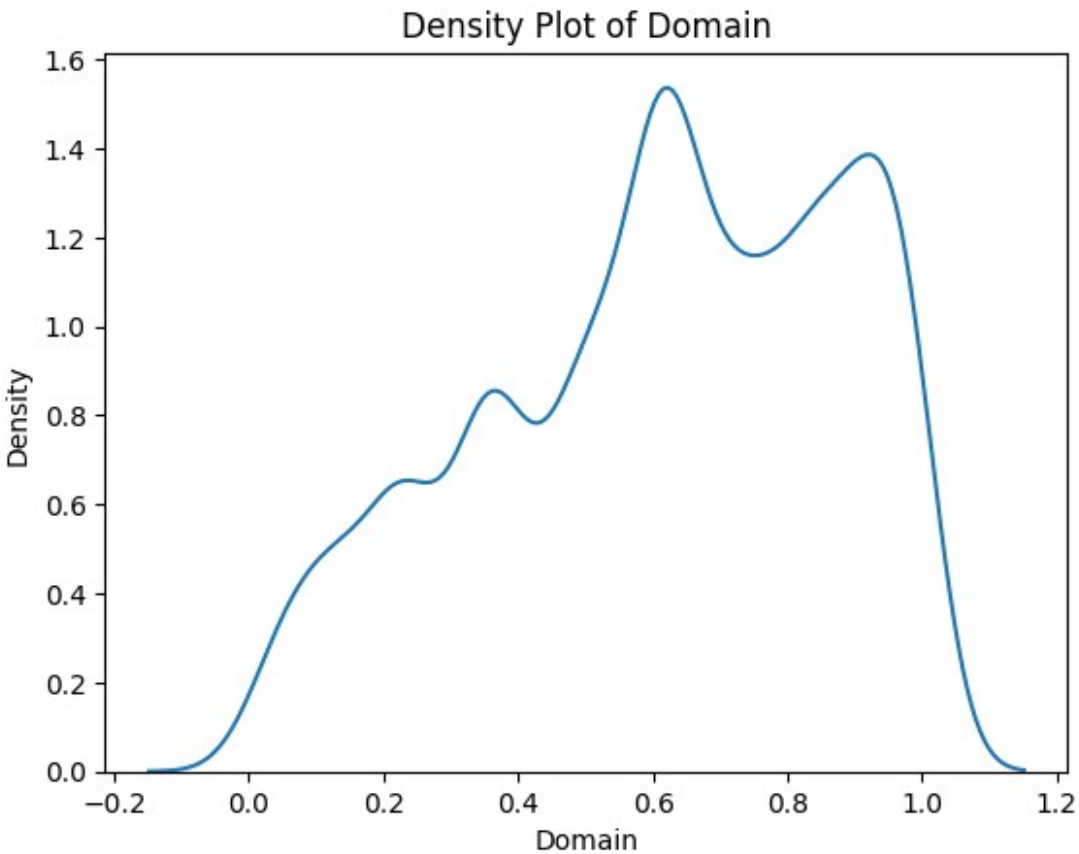

Density Plot of CollegeCityID
Density
0.00014
0.00012
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
0
5000
10000
15000
20000
CollegeCityID

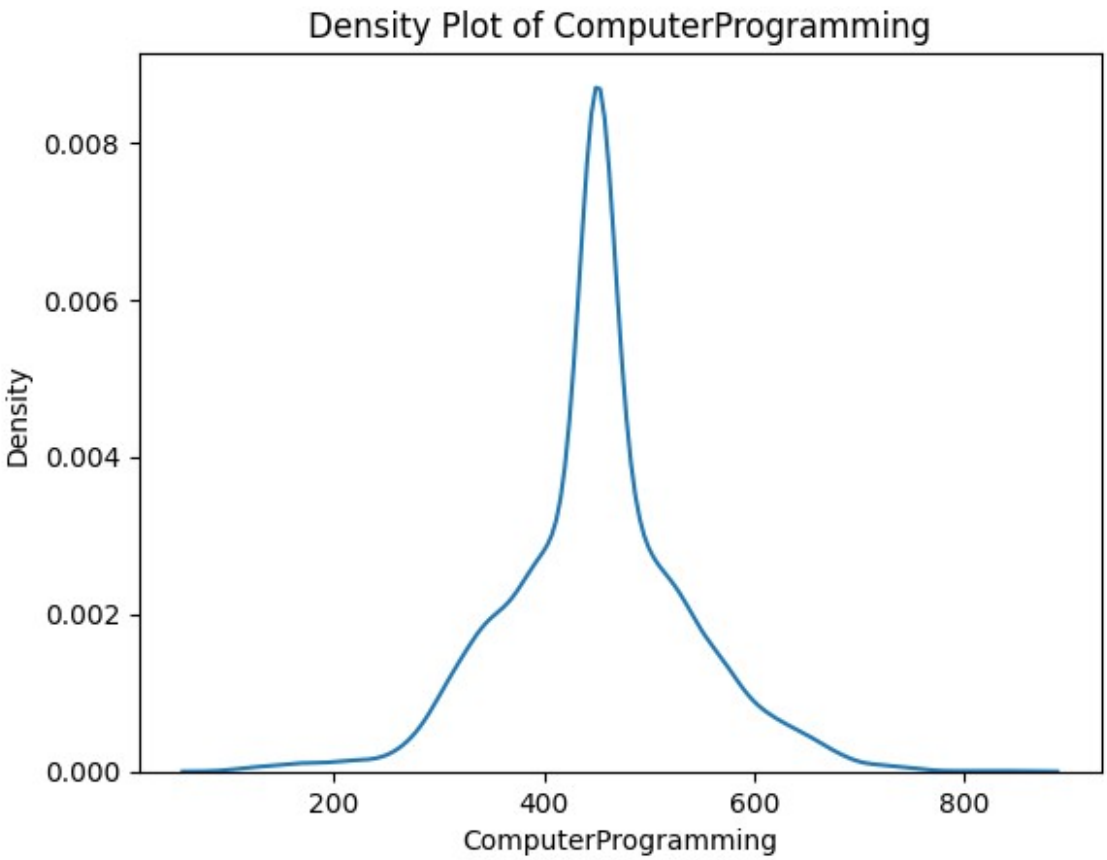

Density Plot of GraduationYear
Density
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
500
1000
1500
2000
GraduationYear

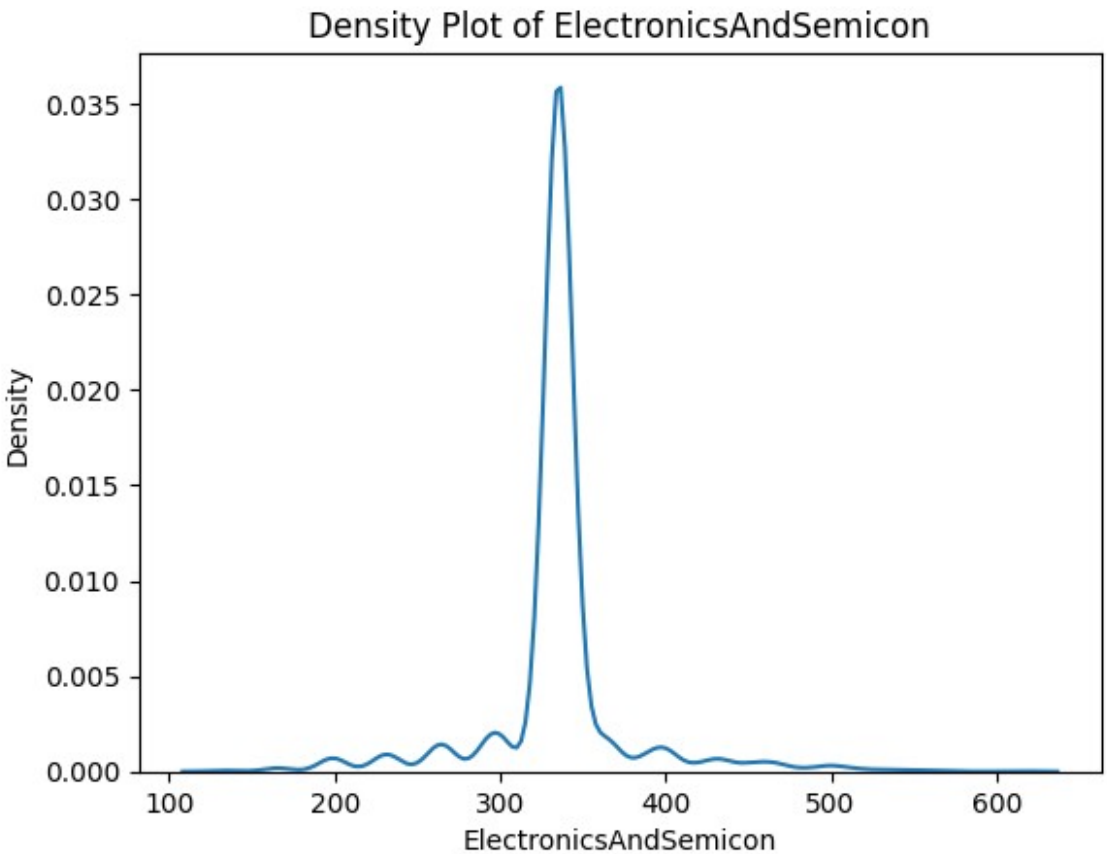

Density Plot of English
Density
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
100
200
300
400
500
600
700
800
900
English

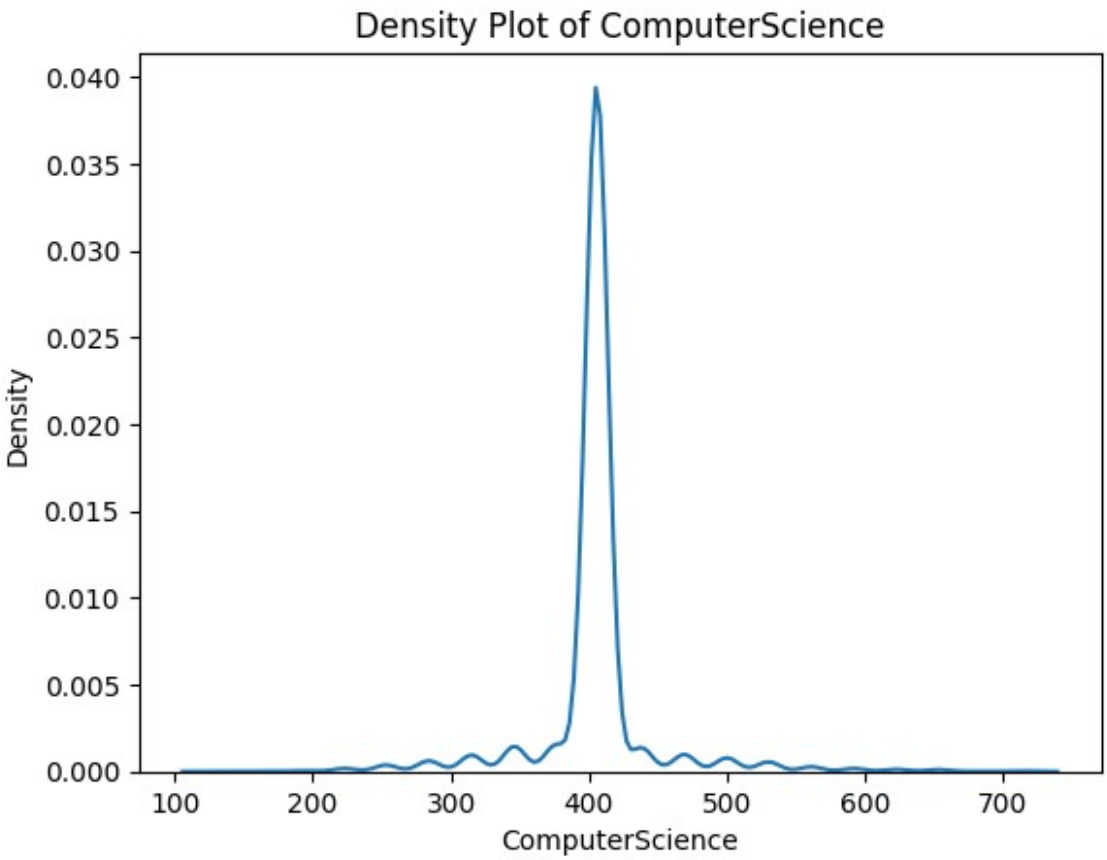

Density Plot of Logical
Density
0.000
0.001
0.002
0.003
0.004
200
300
400
500
600
700
800
Logical

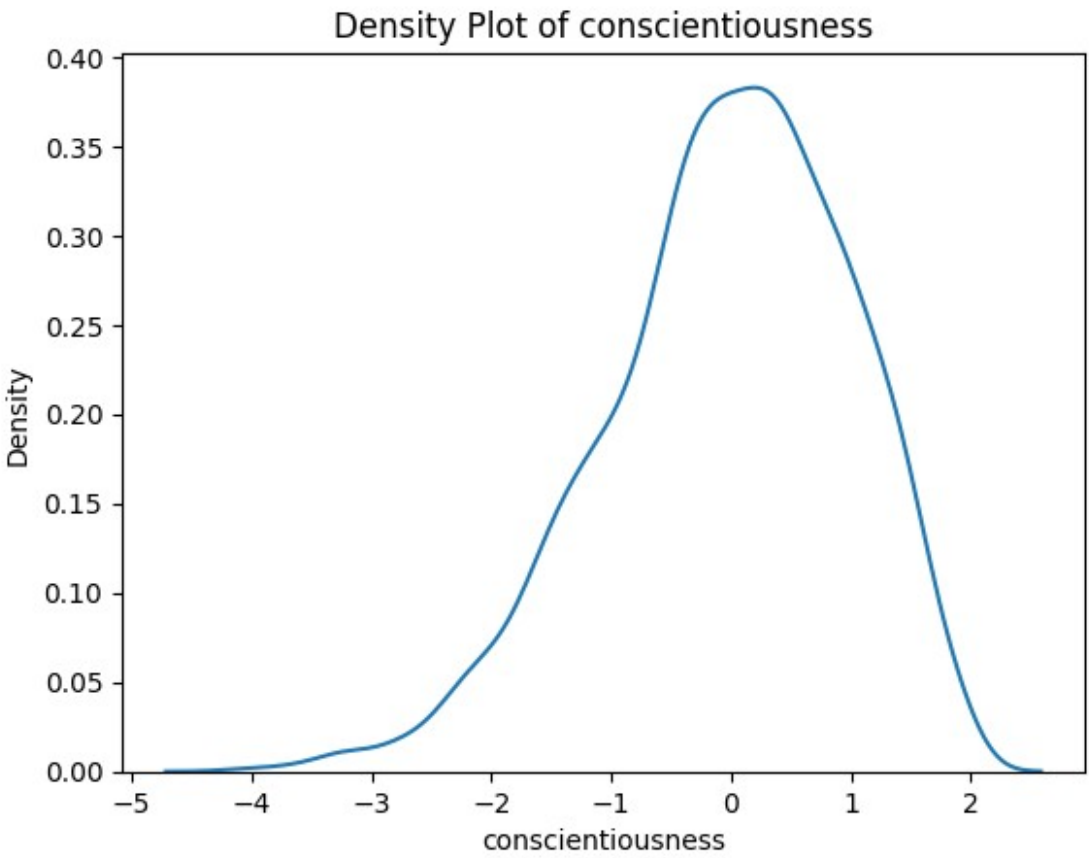

Density Plot of Quant
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
Density
200
400
600
800
1000
Quant

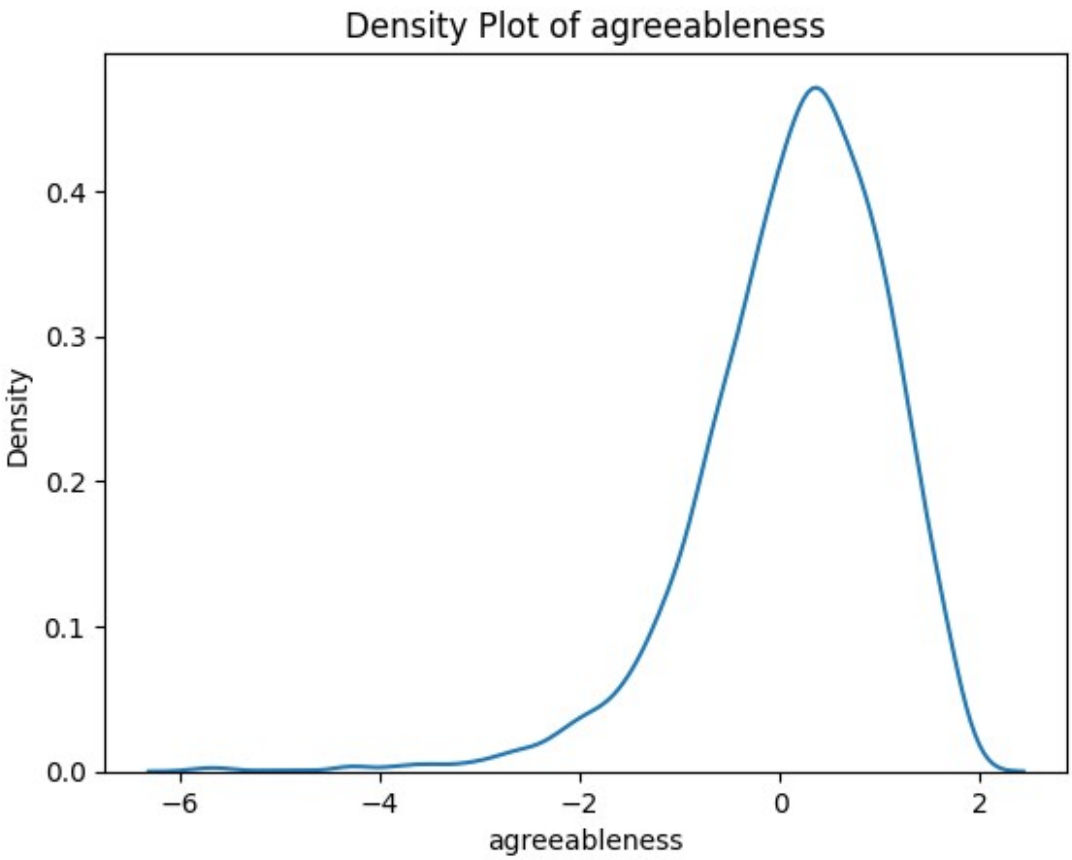

Density Plot of Domain
Density
Domain
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2

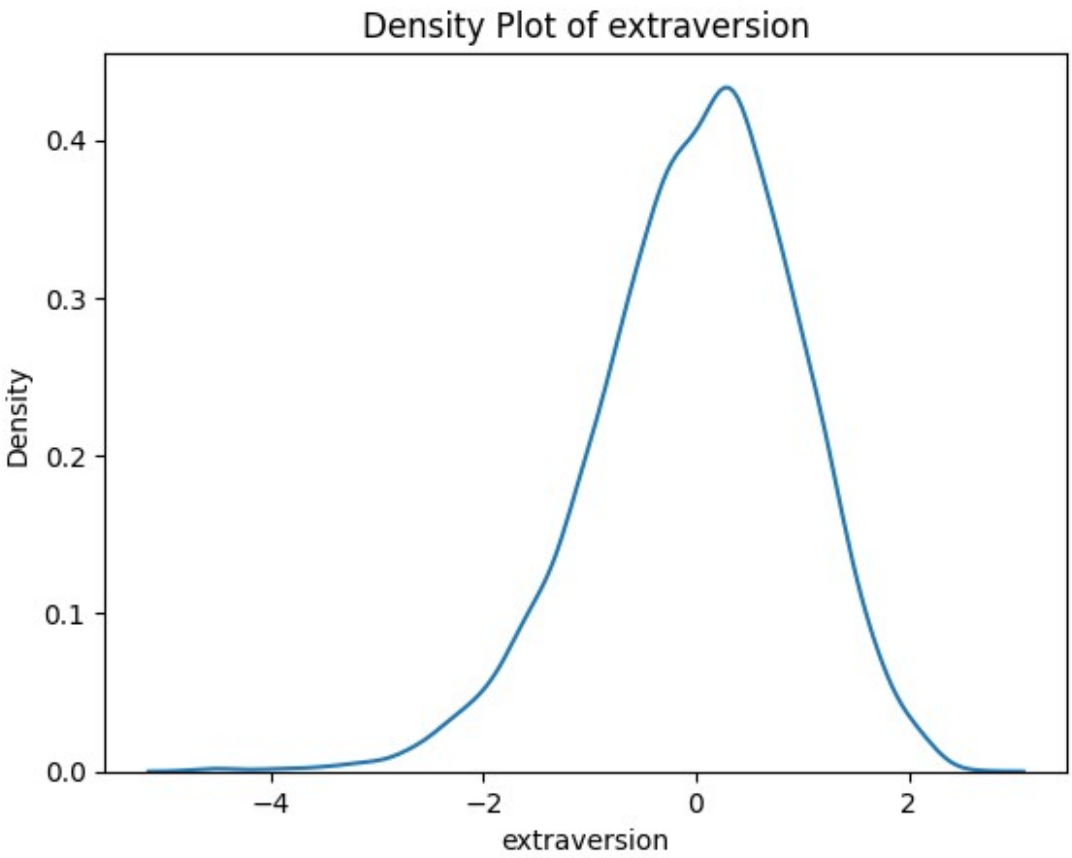

Density Plot of ComputerProgramming
Density
0.000
0.002
0.004
0.006
0.008
200
400
600
800
ComputerProgramming

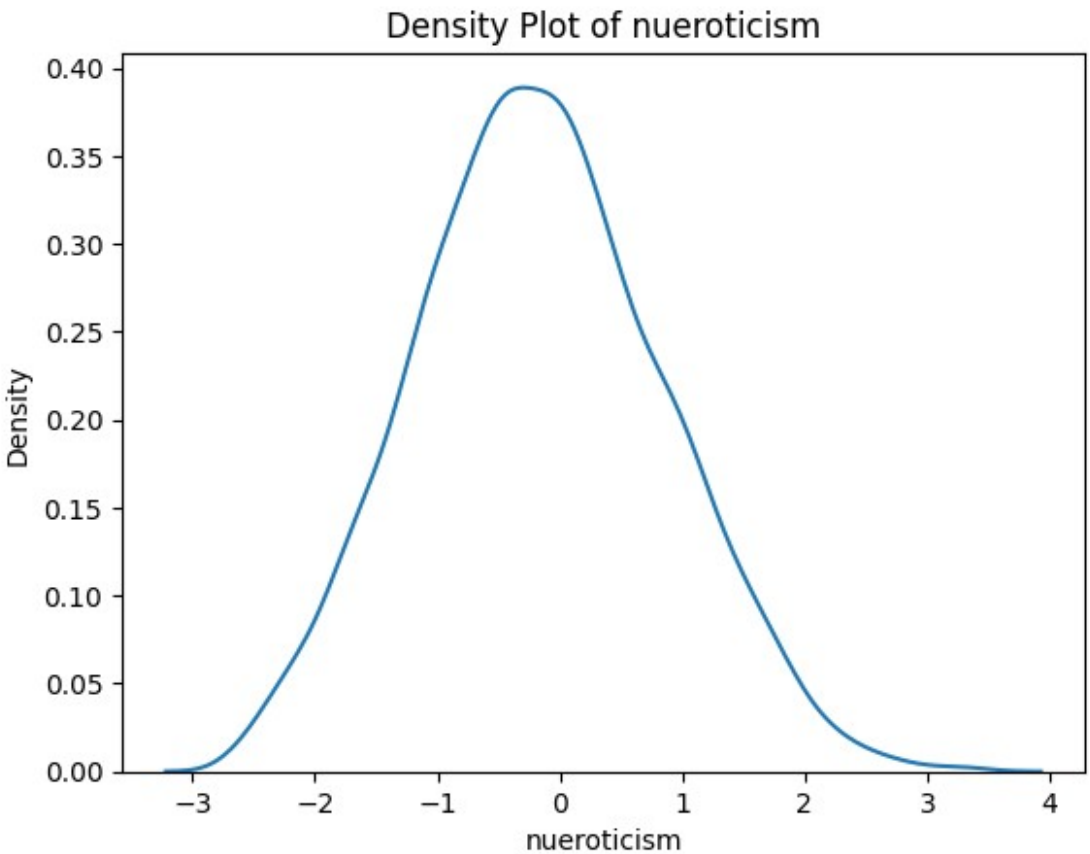

Density Plot of ElectronicsAndSemicon
Density
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
100
200
300
400
500
600
ElectronicsAndSemicon

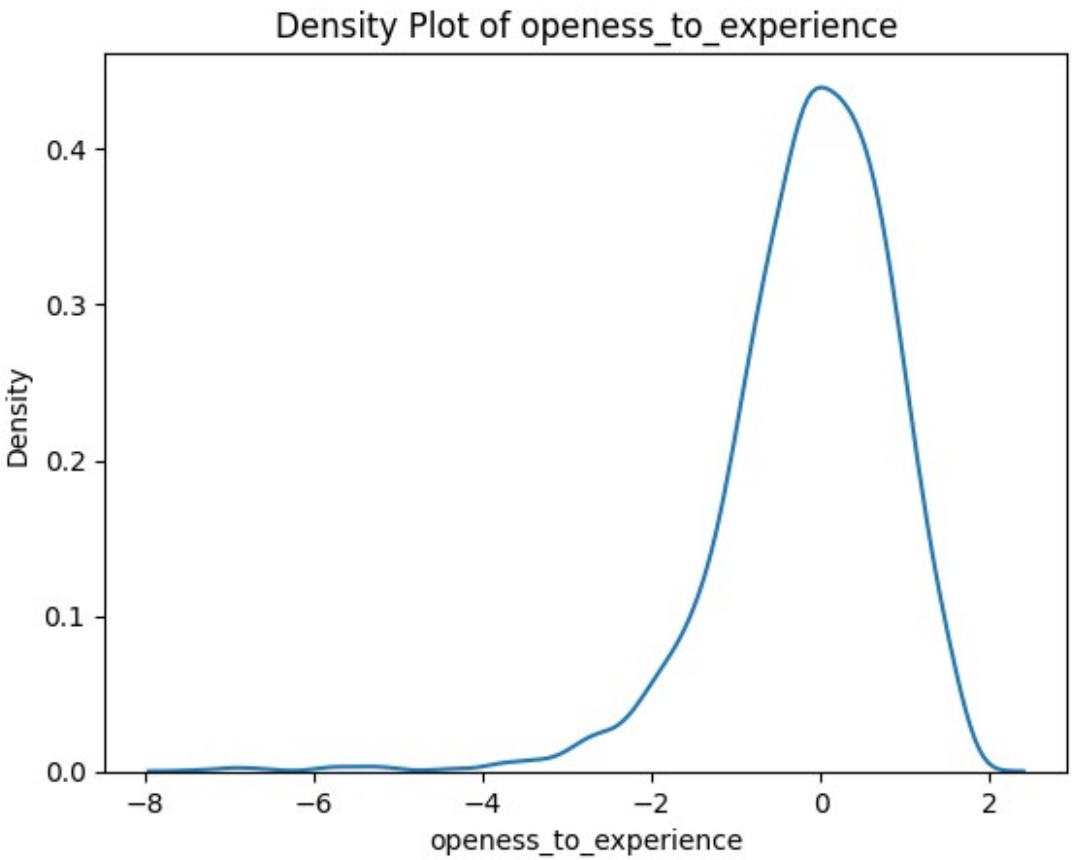

Density Plot of ComputerScience
Density
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
100
200
300
400
500
600
700
ComputerScience

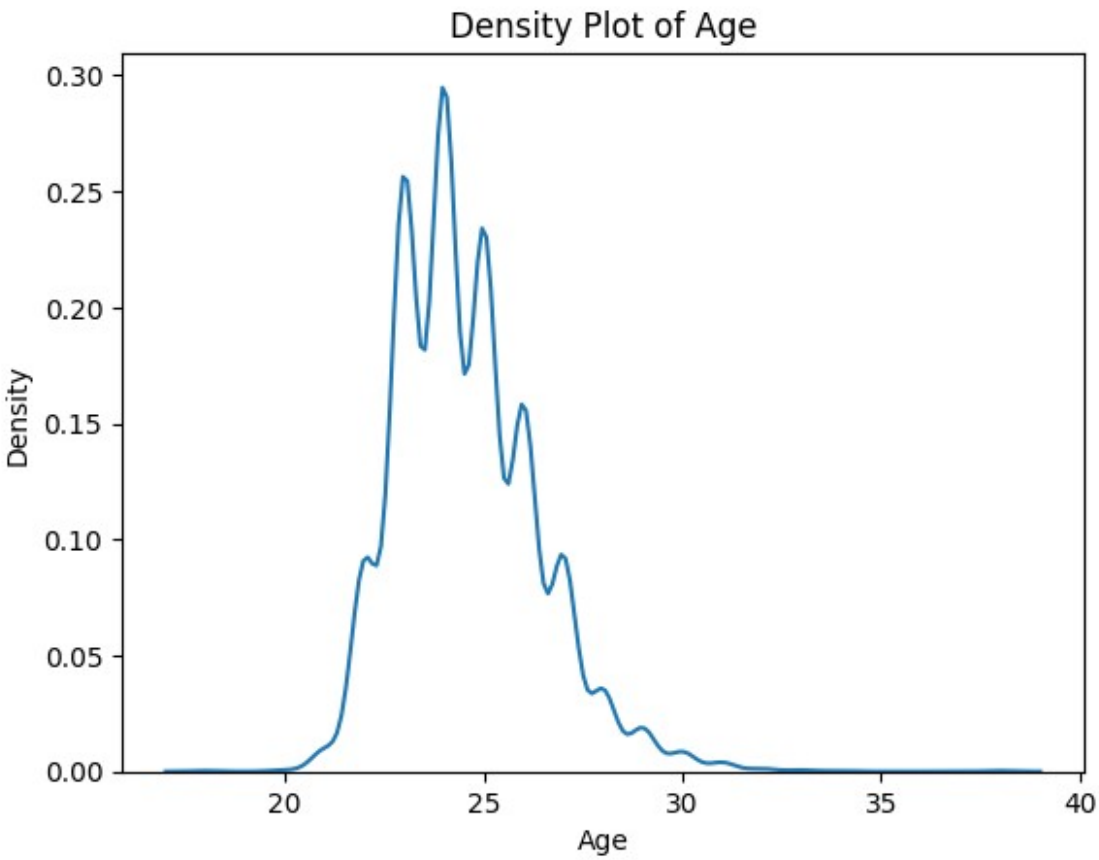

Density Plot of conscientiousness
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Density
−5
−4
−3
−2
−1
0
1
2
conscientiousness


Density Plot of agreeableness
Density
0.0
0.1
0.2
0.3
0.4
−6
−4
−2
0
2
agreeableness


Density Plot of extraversion
Density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
extraversion

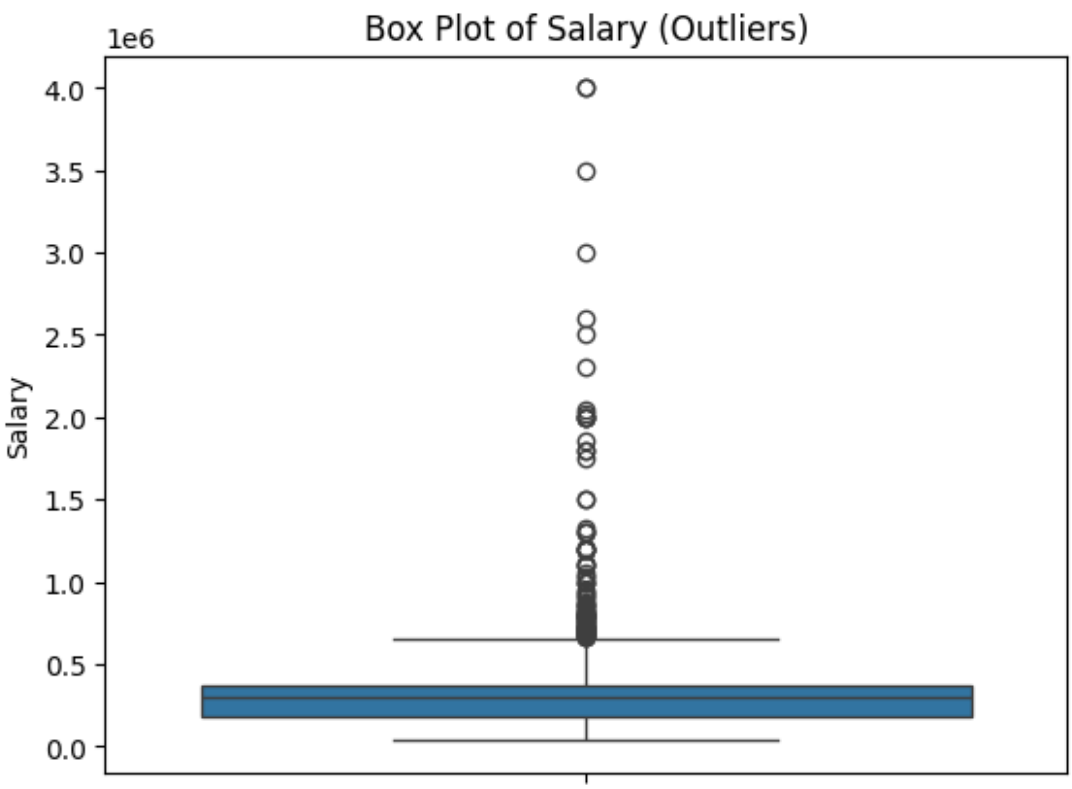

Density Plot of nueroticism
Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−3
−2
−1
0
1
2
3
4
nueroticism

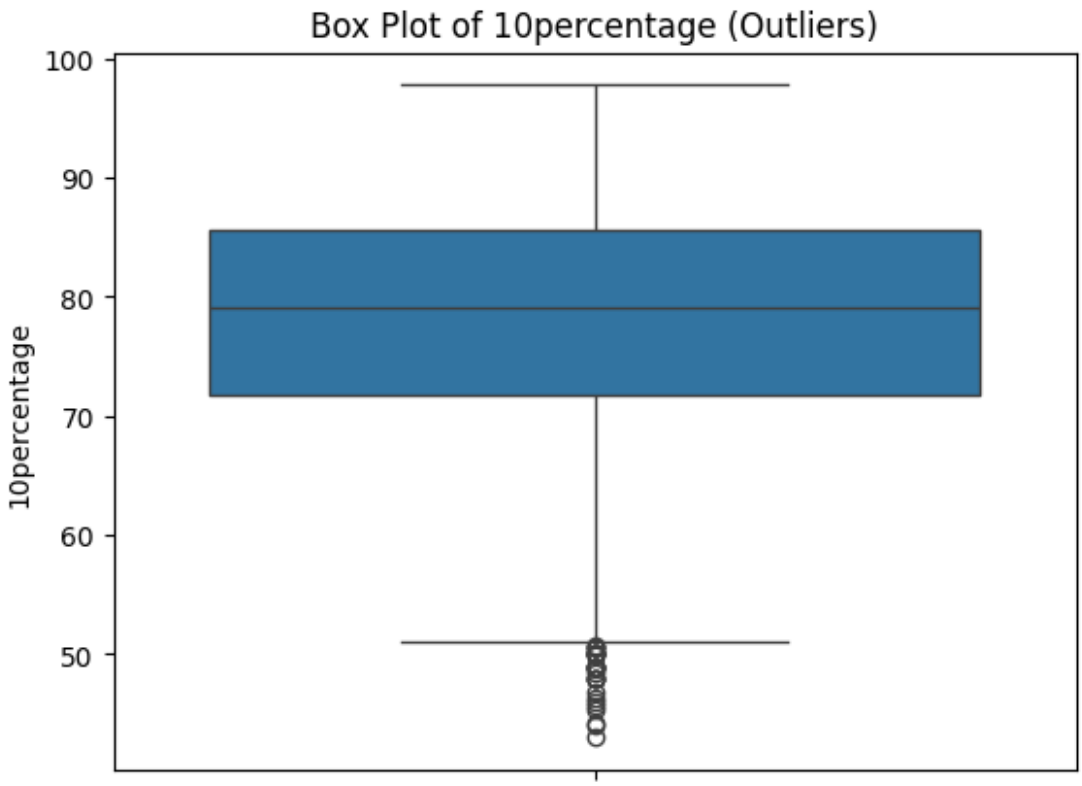

Density Plot of openess_to_experience
Density
0.0
0.1
0.2
0.3
0.4
−8
−6
−4
−2
0
2
openess_to_experience

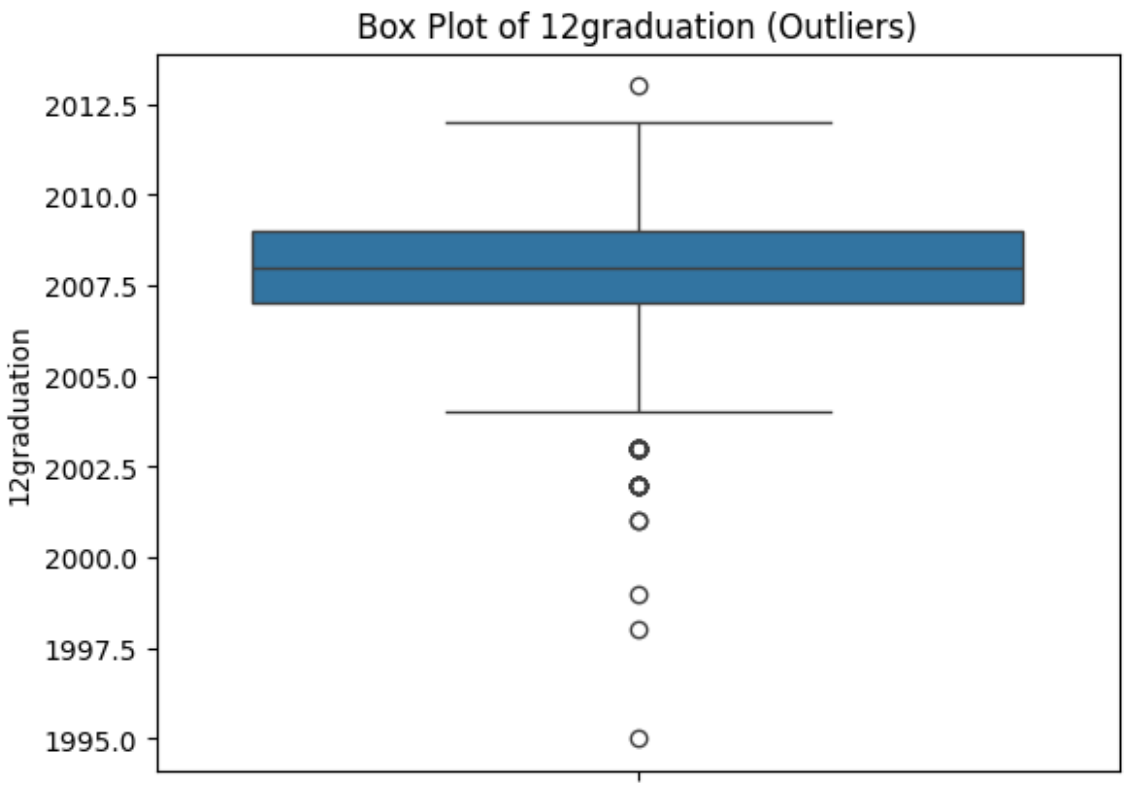


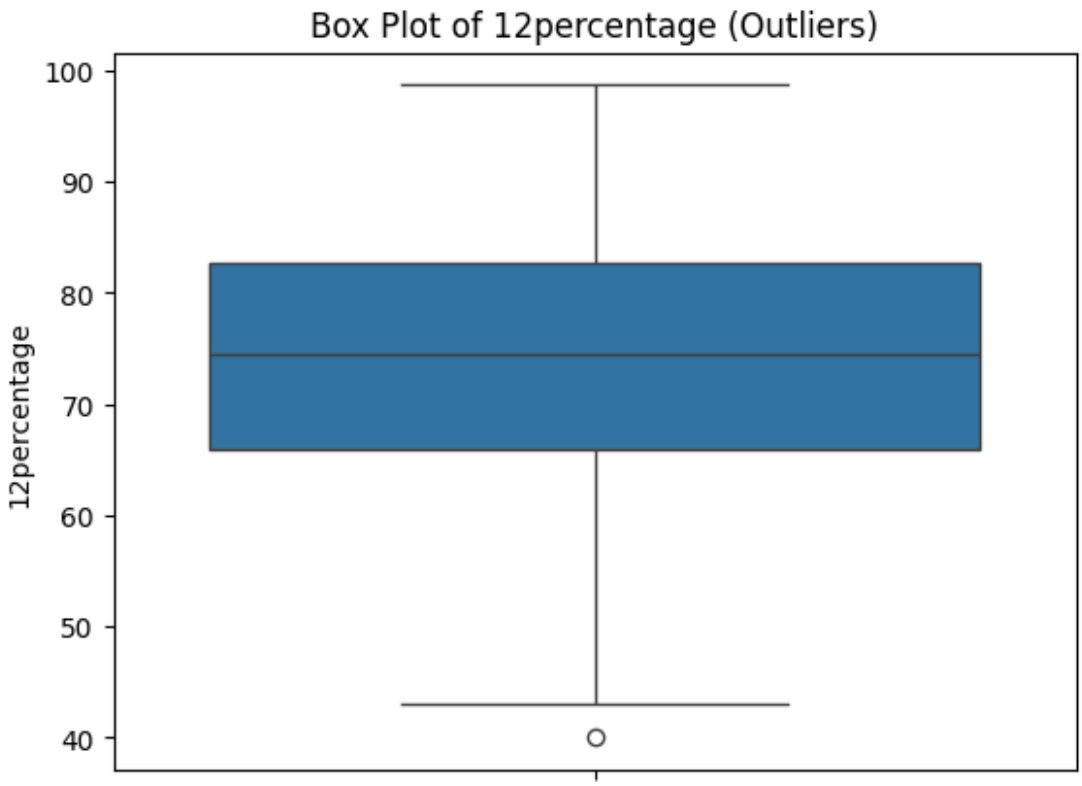

```
# prompt: obseravtions from density graphs

**Observations from the density plots:**

* **Age:** The density plot for age shows that the distribution is
right-skewed, with most people being in their 20s and 30s.
* **Salary:** The density plot for salary shows that the distribution
is also right-skewed, with most people earning less than 10 lakhs per
annum.
* **ssc_p:** The density plot for ssc_p shows that the distribution is
approximately normal, with most people scoring between 60% and 80%.
* **hsc_p:** The density plot for hsc_p shows that the distribution is
also approximately normal, with most people scoring between 60% and
80%.
* **degree_p:** The density plot for degree_p shows that the
distribution is right-skewed, with most people scoring between 60% and
80%.
* **etest_p:** The density plot for etest_p shows that the
distribution is approximately normal, with most people scoring between
60% and 80%.
* **mba_p:** The density plot for mba_p shows that the distribution is
right-skewed, with most people scoring between 60% and 80%.
* **workex:** The density plot for workex shows that the distribution
```

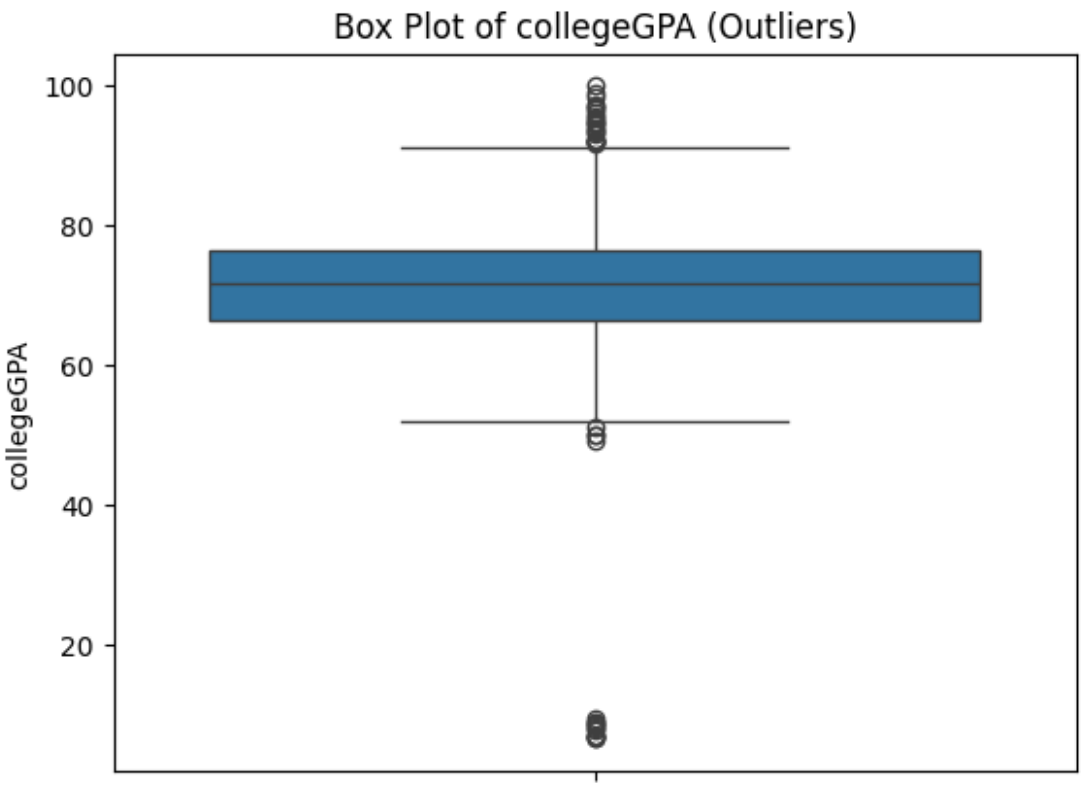
```
is right-skewed, with most people having less than 5 years of
experience.

**Additional observations:**

* The density plots confirm the observations made from the histograms
and box plots.
* The distributions of some of the variables are skewed. For example,
the distribution of salary is right-skewed.
* There are some outliers in the data, particularly for salary and
workex.

# Check for outliers using box plots
for col in numerical_columns:
    plt.figure()
    sns.boxplot(df[col])
    plt.title(f"Box Plot of {col} (Outliers)")
    plt.show()
```

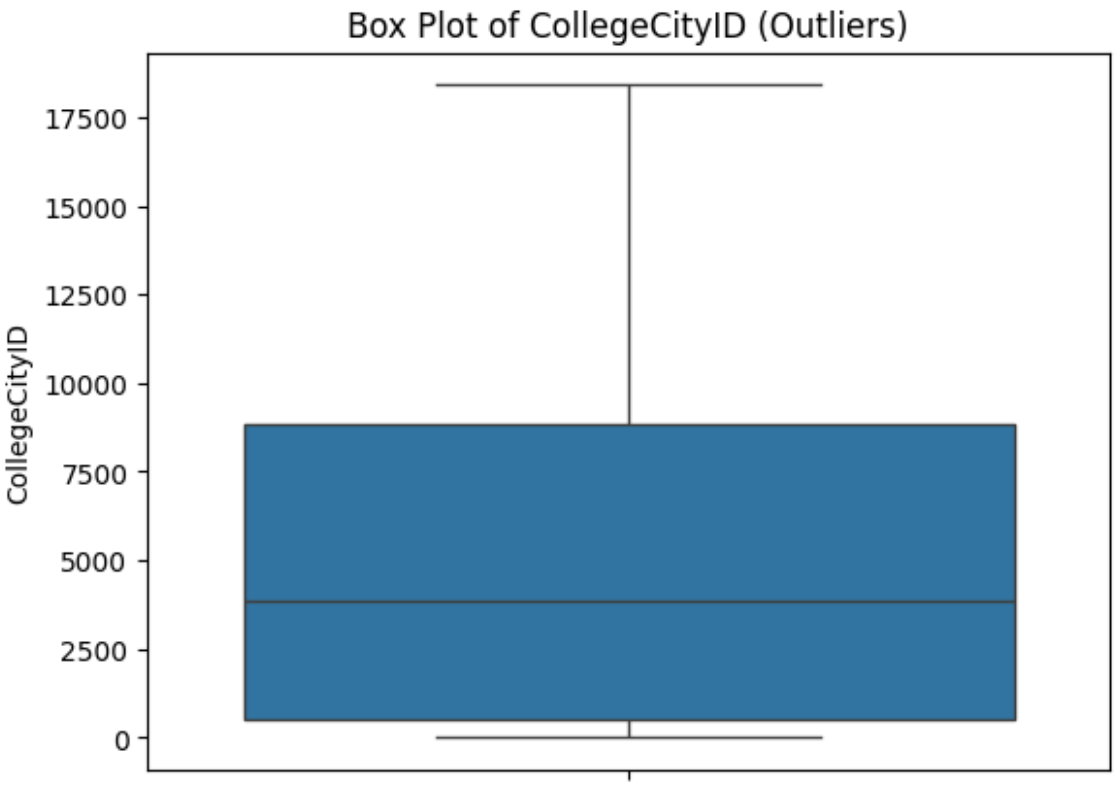


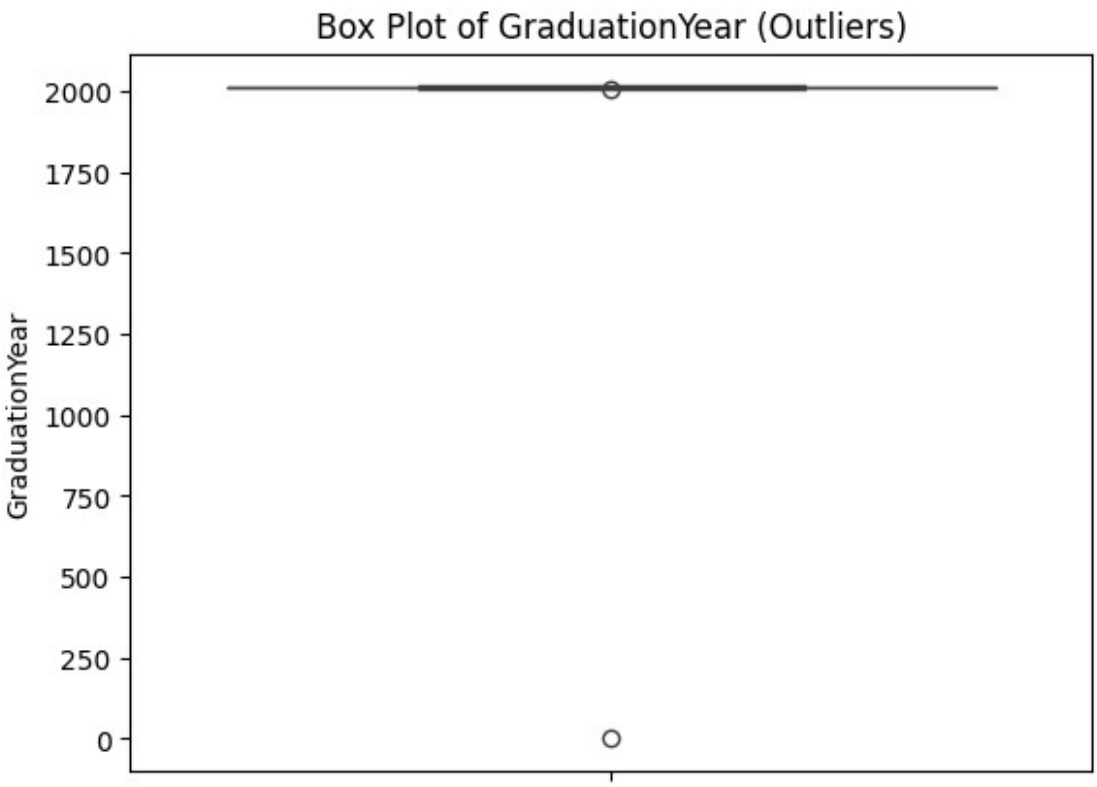

Box Plot of 10percentage (Outliers)
10percentage
100
90
80
70
60
50

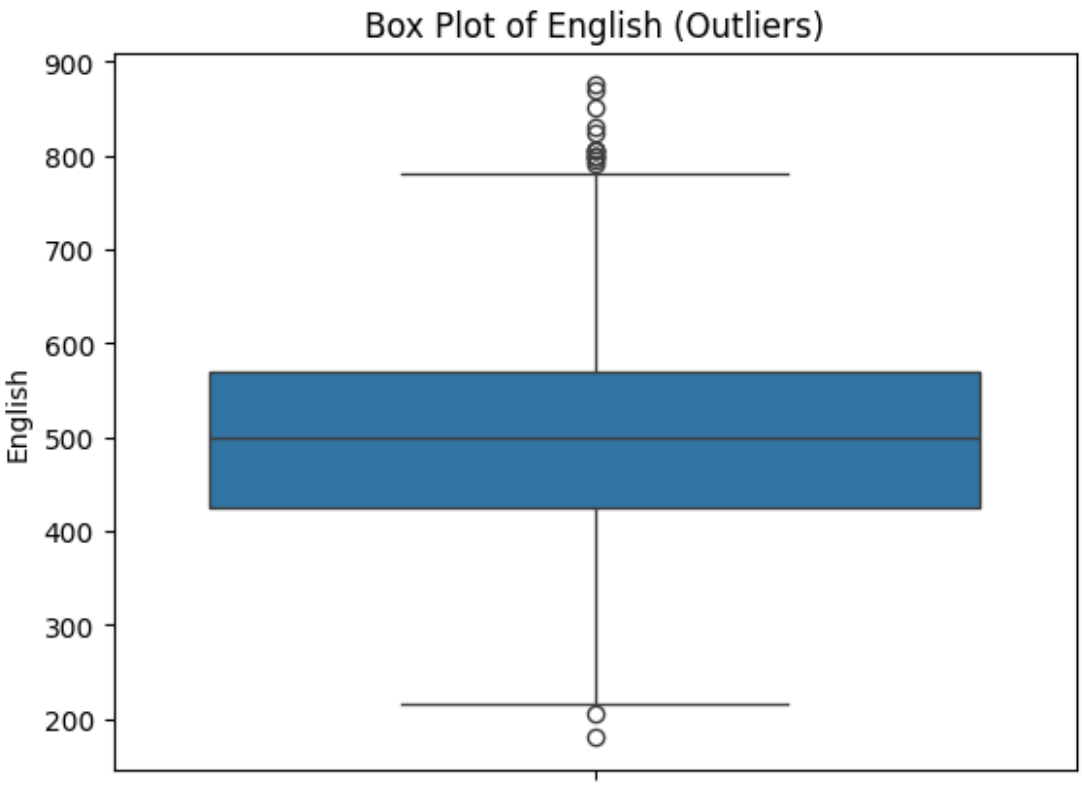

Box Plot of 12graduation (Outliers)
12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0

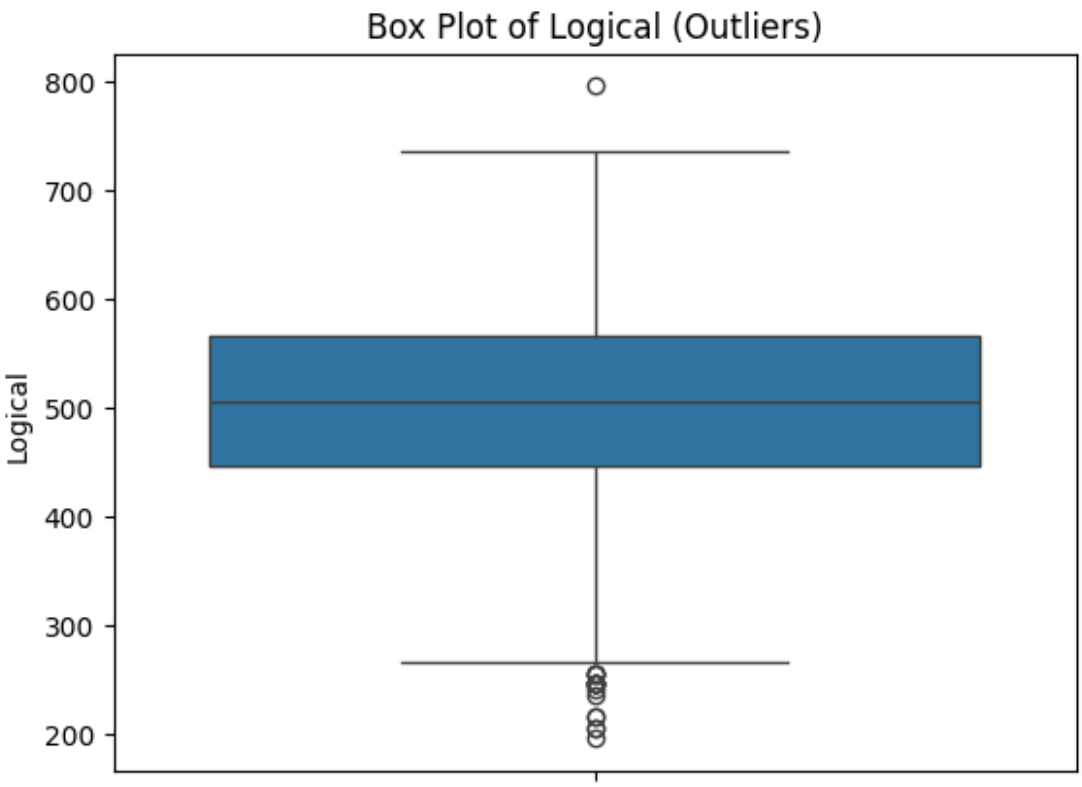

Box Plot of 12percentage (Outliers)
12percentage
100
90
80
70
60
50
40

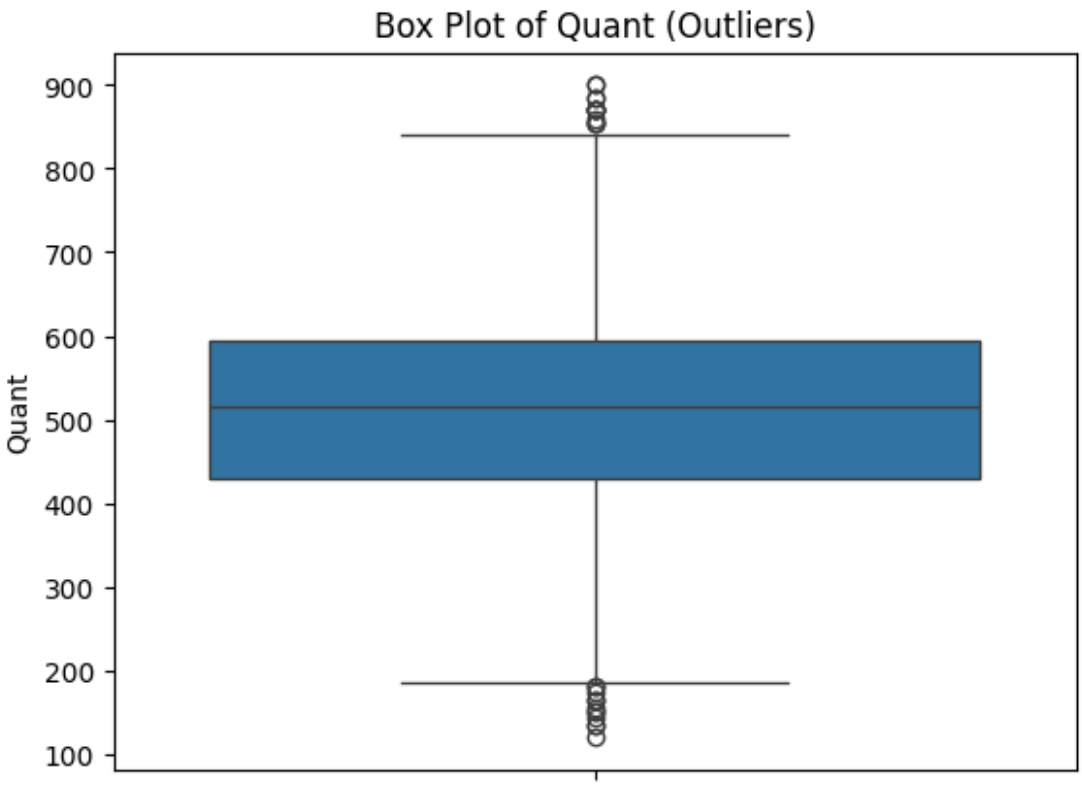

Box Plot of collegeGPA (Outliers)
collegeGPA
100
80
60
40
20

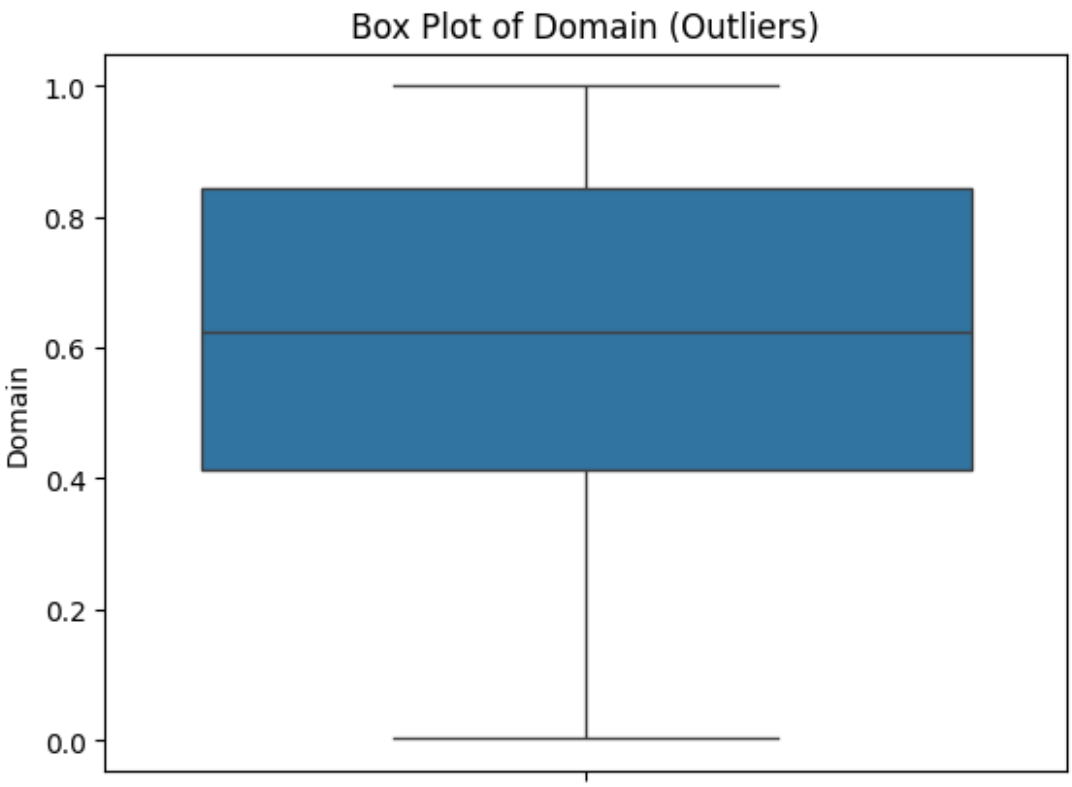

Box Plot of CollegeCityID (Outliers)
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0

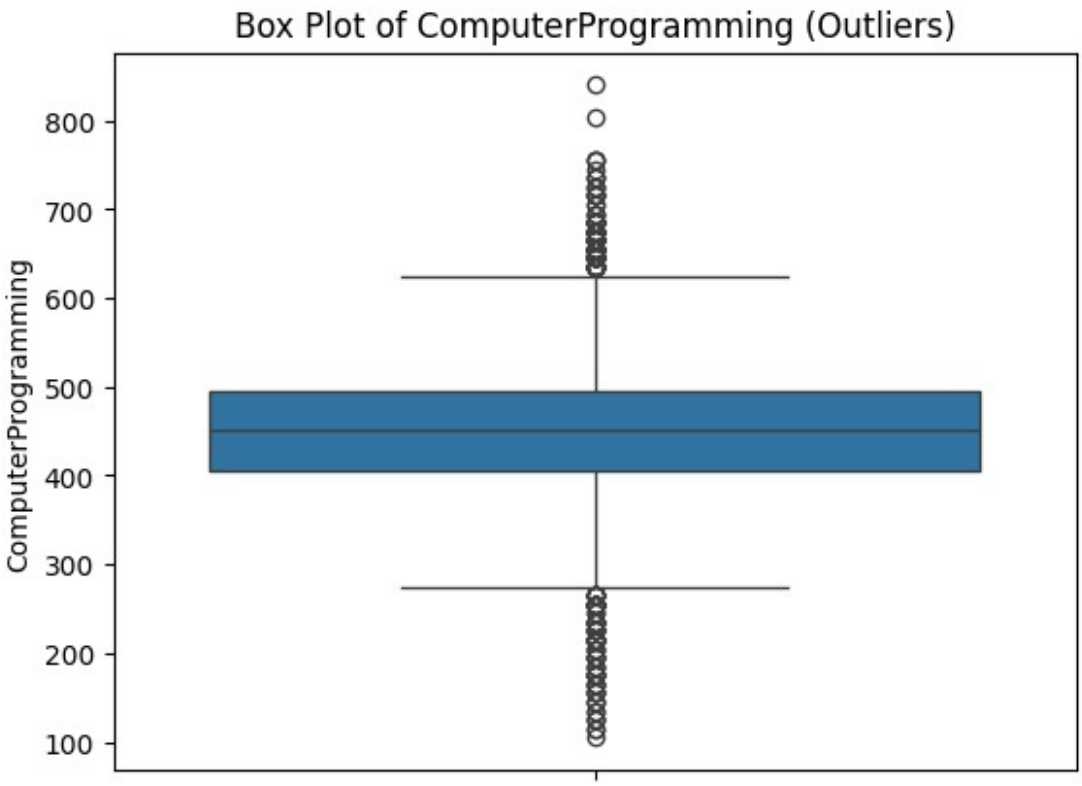

Box Plot of GraduationYear (Outliers)
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0

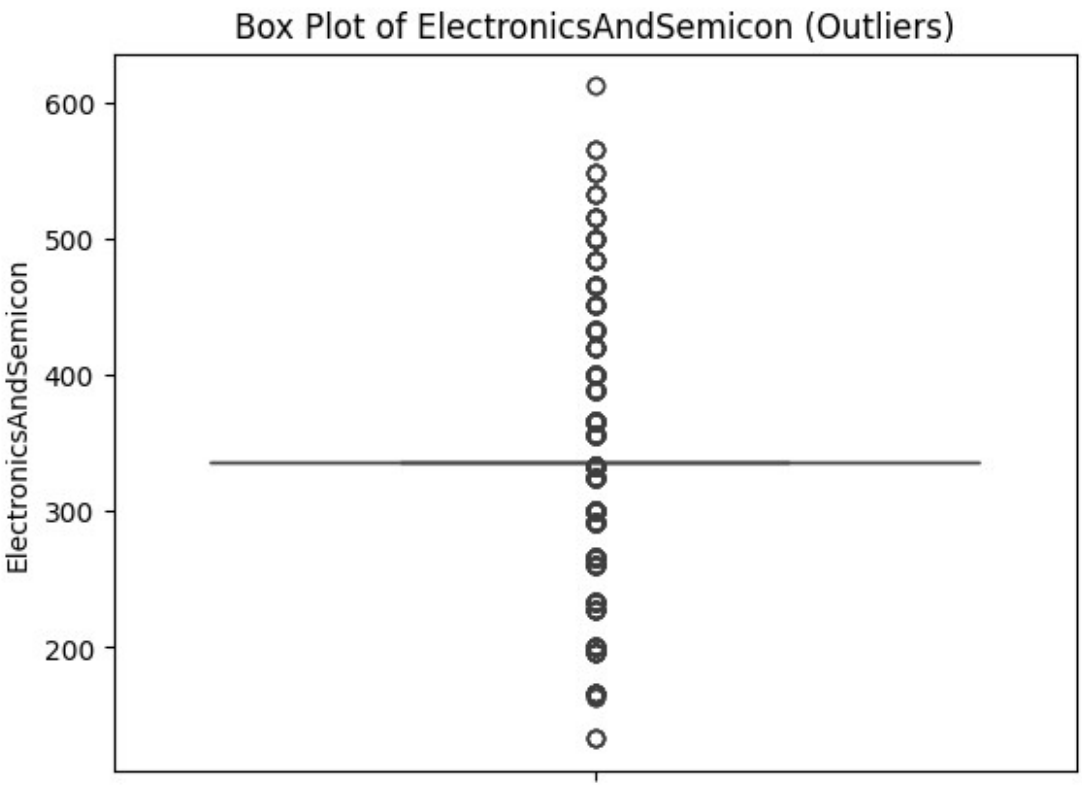

Box Plot of English (Outliers)
English
900
800
700
600
500
400
300
200

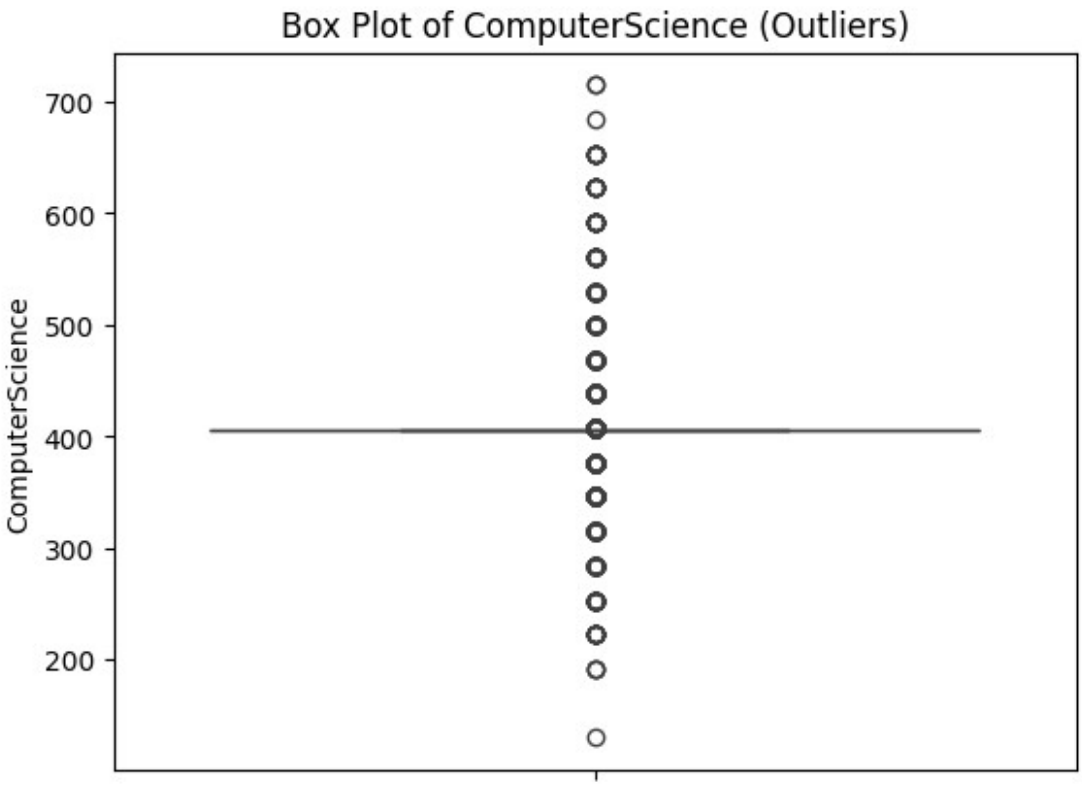

Box Plot of Logical (Outliers)
Logical
800
700
600
500
400
300
200

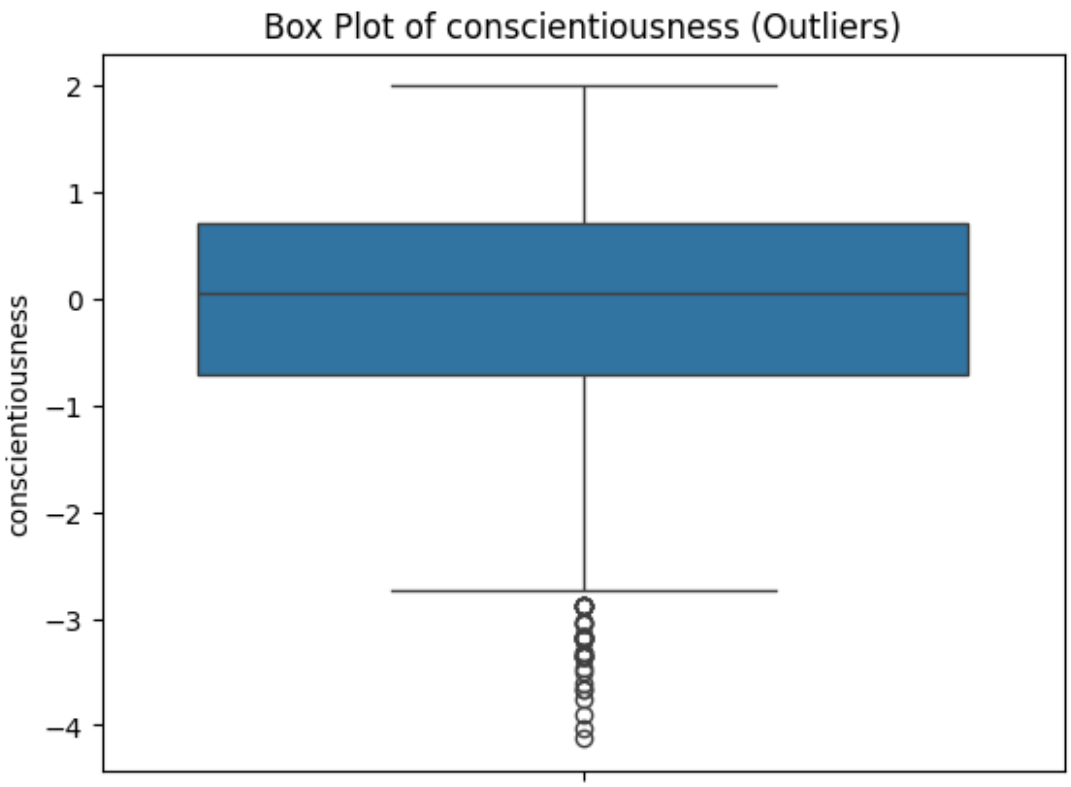

Box Plot of Quant (Outliers)
Quant
900
800
700
600
500
400
300
200
100

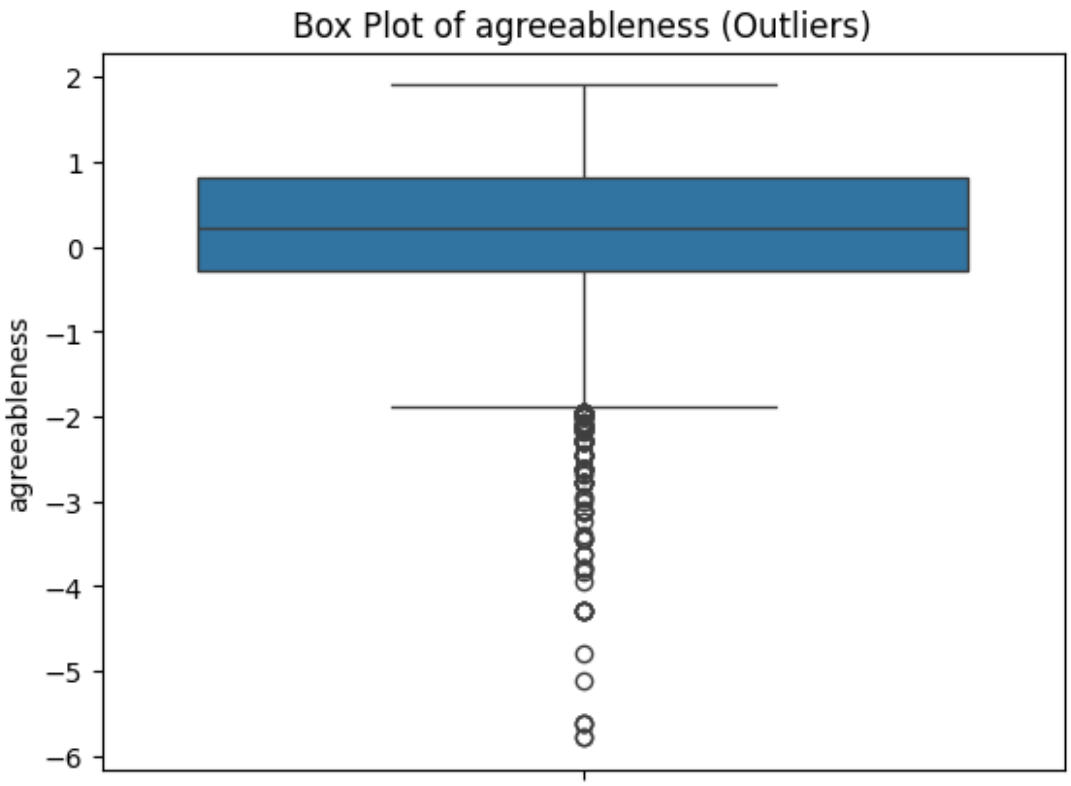

Box Plot of Domain (Outliers)
Domain
1.0
0.8
0.6
0.4
0.2
0.0

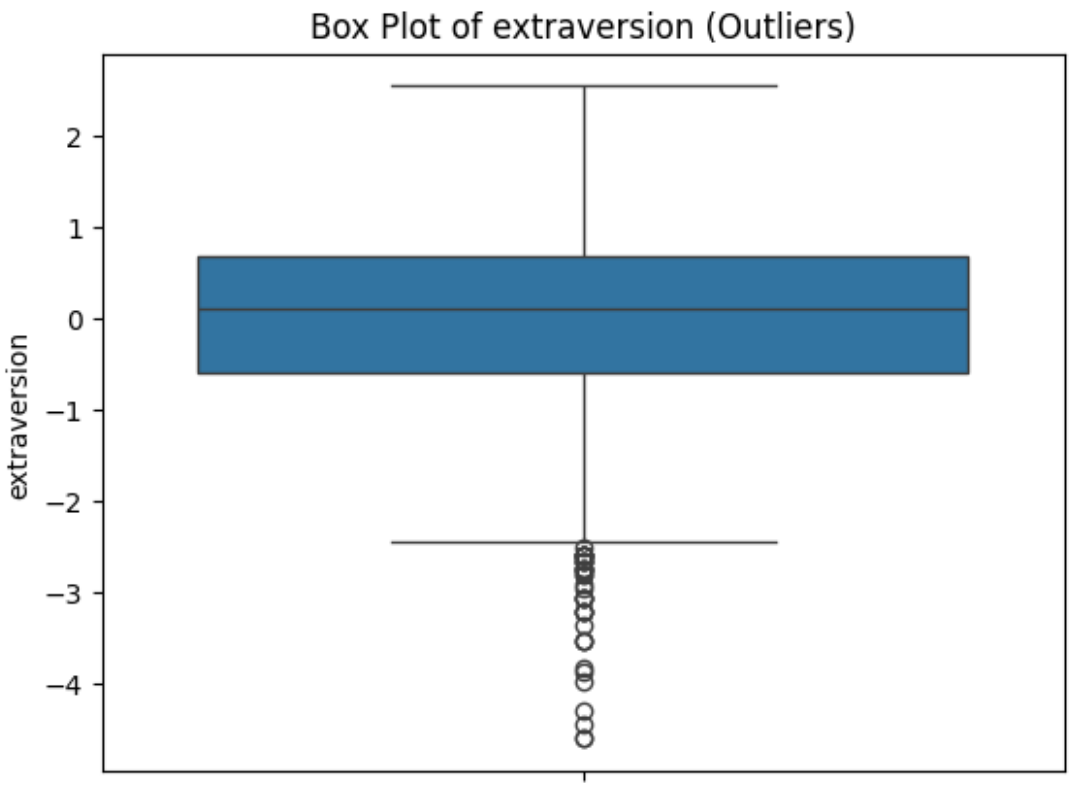

Box Plot of ComputerProgramming (Outliers)
ComputerProgramming
800
700
600
500
400
300
200
100

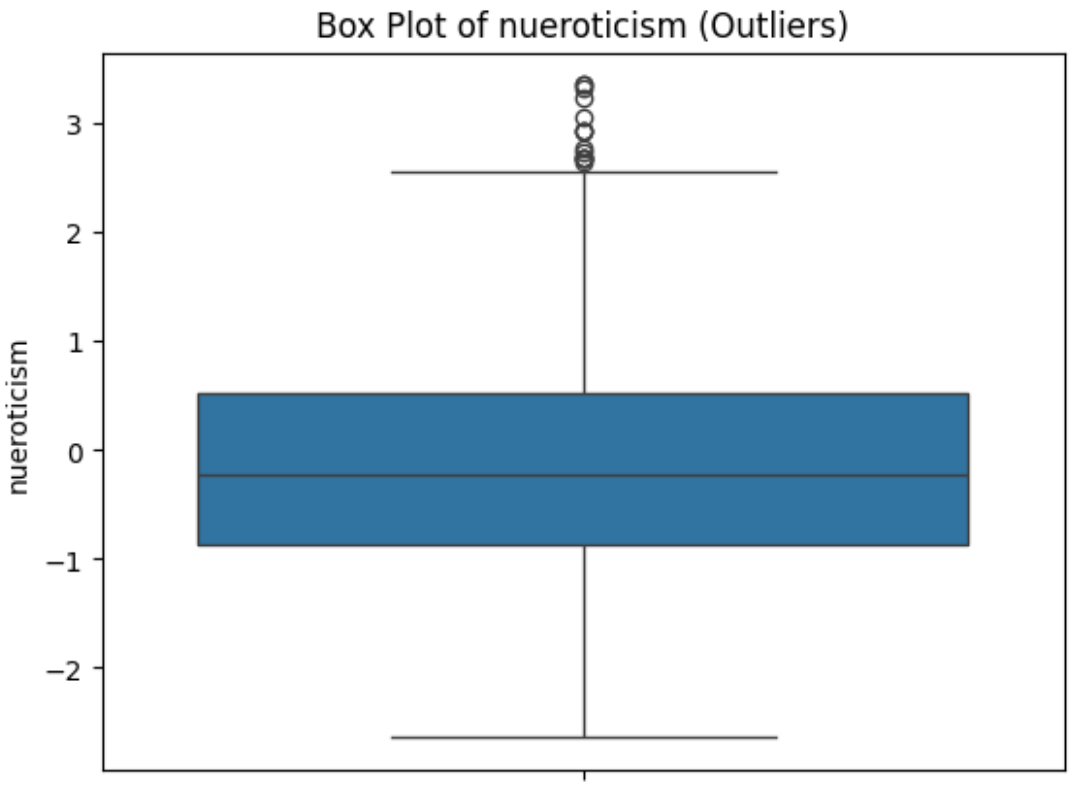

Box Plot of ElectronicsAndSemicon (Outliers)
ElectronicsAndSemicon
600
500
400
300
200

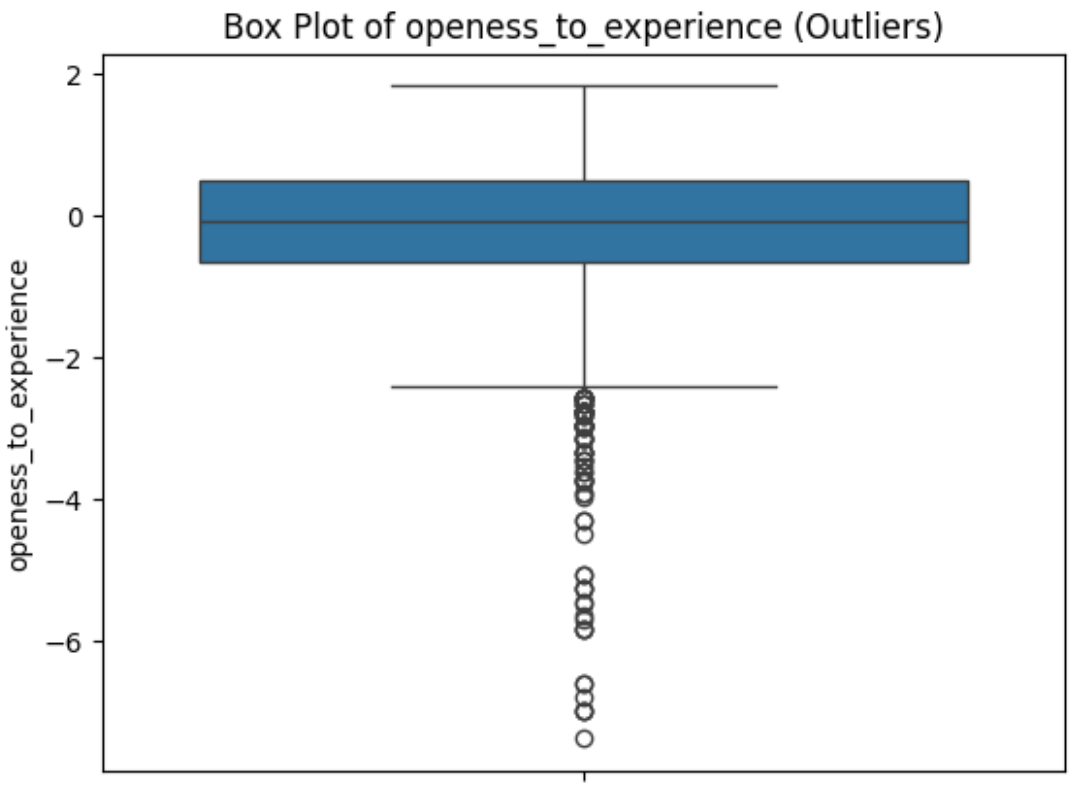

Box Plot of ComputerScience (Outliers)
ComputerScience
700
600
500
400
300
200

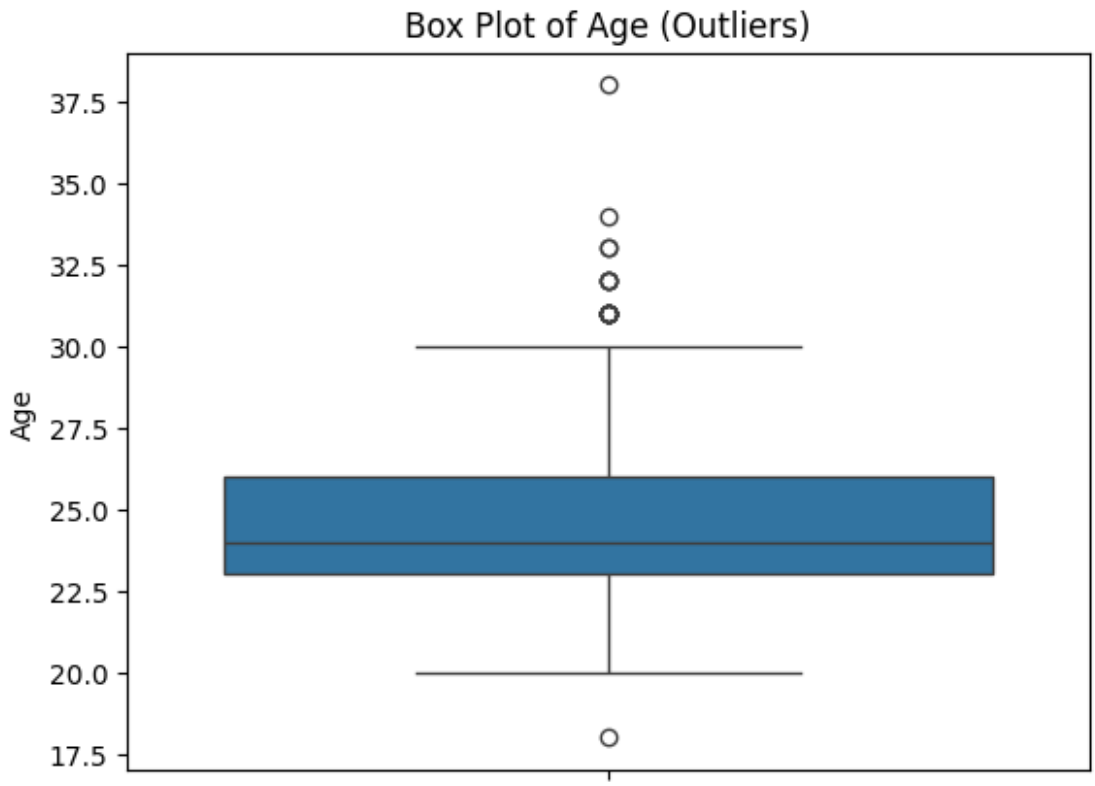

Box Plot of conscientiousness (Outliers)
conscientiousness
2
1
0
−1
−2
−3
−4


Box Plot of agreeableness (Outliers)
agreeableness
2
1
0
−1
−2
−3
−4
−5
−6


Box Plot of extraversion (Outliers)
extraversion
2
1
0
−1
−2
−3
−4


Box Plot of nueroticism (Outliers)
nueroticism
3
2
1
0
−1
−2


Box Plot of openess_to_experience (Outliers)
openess_to_experience
2
0
−2
−4
−6




```
# prompt: obervations from above 4 codes

1. **Data Cleaning:**
   - The code performs extensive data cleaning by replacing
inconsistent values in the "10board" and "12board" columns using a
dictionary mapping.
   - It also converts the "DOB" column to a datetime format and
calculates the "Age" based on the year.

2. **Univariate Analysis:**
   - The code identifies the numerical columns in the dataset.
   - It calculates summary statistics for these columns, including
count, mean, standard deviation, minimum, maximum, and quartiles.
   - It creates histograms, box plots, and density plots for each
numerical column to visualize their distribution and identify any
outliers.

3. **Outlier Detection:**
   - The code creates separate box plots for each numerical column to
specifically check for outliers.
   - Outliers can be identified as data points that fall outside the
whiskers of the box plot.

4. **Observations:**
   - The provided code focuses on data cleaning and univariate
```

```
analysis of the numerical columns in the dataset.
   - It provides valuable insights into the distribution of each
numerical variable, including central tendency, spread, and potential
outliers.
   - This information can be used for further data exploration,
feature engineering, and model building.

  File "<ipython-input-38-3d79d47c46ae>", line 3
    1. **Data Cleaning:**
               ^
SyntaxError: invalid syntax
```

Bivariate Analysis

```
# Correlation Matrix
correlation_matrix = df.corr()
print(correlation_matrix)
```

```
                          Salary  10percentage  12graduation
12percentage  \
Salary                  1.000000      0.177373     -0.161383
0.170254
10percentage            0.177373      1.000000      0.269957
0.643378
12graduation           -0.161383      0.269957      1.000000
0.259166
12percentage            0.170254      0.643378      0.259166
1.000000
collegeGPA              0.130103      0.312538      0.086001
0.346137
CollegeCityID          -0.118690      0.021082      0.254021
0.022336
GraduationYear         -0.010053     -0.013799      0.014457     -
0.012933
English                 0.178219      0.350780      0.147925
0.212888
Logical                 0.179275      0.316014      0.105887
0.243571
Quant                   0.230627      0.317640      0.001379
0.312413
Domain                  0.170091      0.149948     -0.037072
0.159741
ComputerProgramming     0.164466      0.153693     -0.003991
0.137824
ElectronicsAndSemicon   0.051879      0.039637      0.009207
0.066754
ComputerScience         0.052012      0.064602      0.014311
```

```
0.050151
conscientiousness     -0.064148      0.067657      0.103329
0.058299
agreeableness          0.057423      0.136645      0.041182
0.103998
extraversion          -0.010213     -0.004679      0.061956     -
0.007486
nueroticism           -0.054685     -0.132496     -0.074369     -
0.094369
openess_to_experience -0.011312      0.036692     -0.015069
0.006332
Age                    0.151714     -0.243880     -0.872401     -
0.265781

                       collegeGPA  CollegeCityID  GraduationYear
English  \
Salary                   0.130103      -0.118690       -0.010053
0.178219
10percentage             0.312538       0.021082       -0.013799
0.350780
12graduation             0.086001       0.254021        0.014457
0.147925
12percentage             0.346137       0.022336       -0.012933
0.212888
collegeGPA               1.000000       0.017240        0.008706
0.106478
CollegeCityID            0.017240       1.000000       -0.000172 -
0.022792
GraduationYear           0.008706      -0.000172        1.000000 -
0.024089
English                  0.106478      -0.022792       -0.024089
1.000000
Logical                  0.196610      -0.047094       -0.024018
0.444357
Quant                    0.217380      -0.114672       -0.021781
0.375784
Domain                   0.170181      -0.094509       -0.009276
0.221672
ComputerProgramming      0.196584      -0.074283        0.001738
0.271607
ElectronicsAndSemicon    0.065689      -0.028449        0.000966
0.097452
ComputerScience          0.056808      -0.030108        0.000938
0.127781
conscientiousness        0.069582       0.076432       -0.013235
0.034943
agreeableness            0.068282      -0.005264       -0.002877
0.194990
extraversion            -0.032684       0.005917        0.008397
```

```
0.018755
nueroticism              -0.074859      -0.008973       -0.000417 -
0.155528
openess_to_experience    0.028071      -0.010678        0.016855
0.067979
Age                      -0.113168      -0.230825       -0.025639 -
0.101117

                         Logical     Quant    Domain
ComputerProgramming  \
Salary                  0.179275  0.230627  0.170091
0.164466
10percentage            0.316014  0.317640  0.149948
0.153693
12graduation            0.105887  0.001379 -0.037072             -
0.003991
12percentage            0.243571  0.312413  0.159741
0.137824
collegeGPA              0.196610  0.217380  0.170181
0.196584
CollegeCityID          -0.047094 -0.114672 -0.094509             -
0.074283
GraduationYear         -0.024018 -0.021781 -0.009276
0.001738
English                 0.444357  0.375784  0.221672
0.271607
Logical                 1.000000  0.500152  0.312916
0.387707
Quant                   0.500152  1.000000  0.337981
0.368517
Domain                  0.312916  0.337981  1.000000
0.497235
ComputerProgramming     0.387707  0.368517  0.497235
1.000000
ElectronicsAndSemicon   0.099187  0.128847  0.523350
0.054712
ComputerScience         0.148531  0.153973  0.422891
0.204965
conscientiousness       0.025876 -0.005639 -0.009745             -
0.010519
agreeableness           0.167207  0.103443  0.092925
0.112553
extraversion           -0.006949 -0.028616 -0.028769             -
0.025765
nueroticism            -0.178781 -0.131895 -0.098408             -
0.123289
openess_to_experience   0.048420  0.020377  0.019878
0.029728
Age                    -0.098491 -0.031144  0.029500             -
```

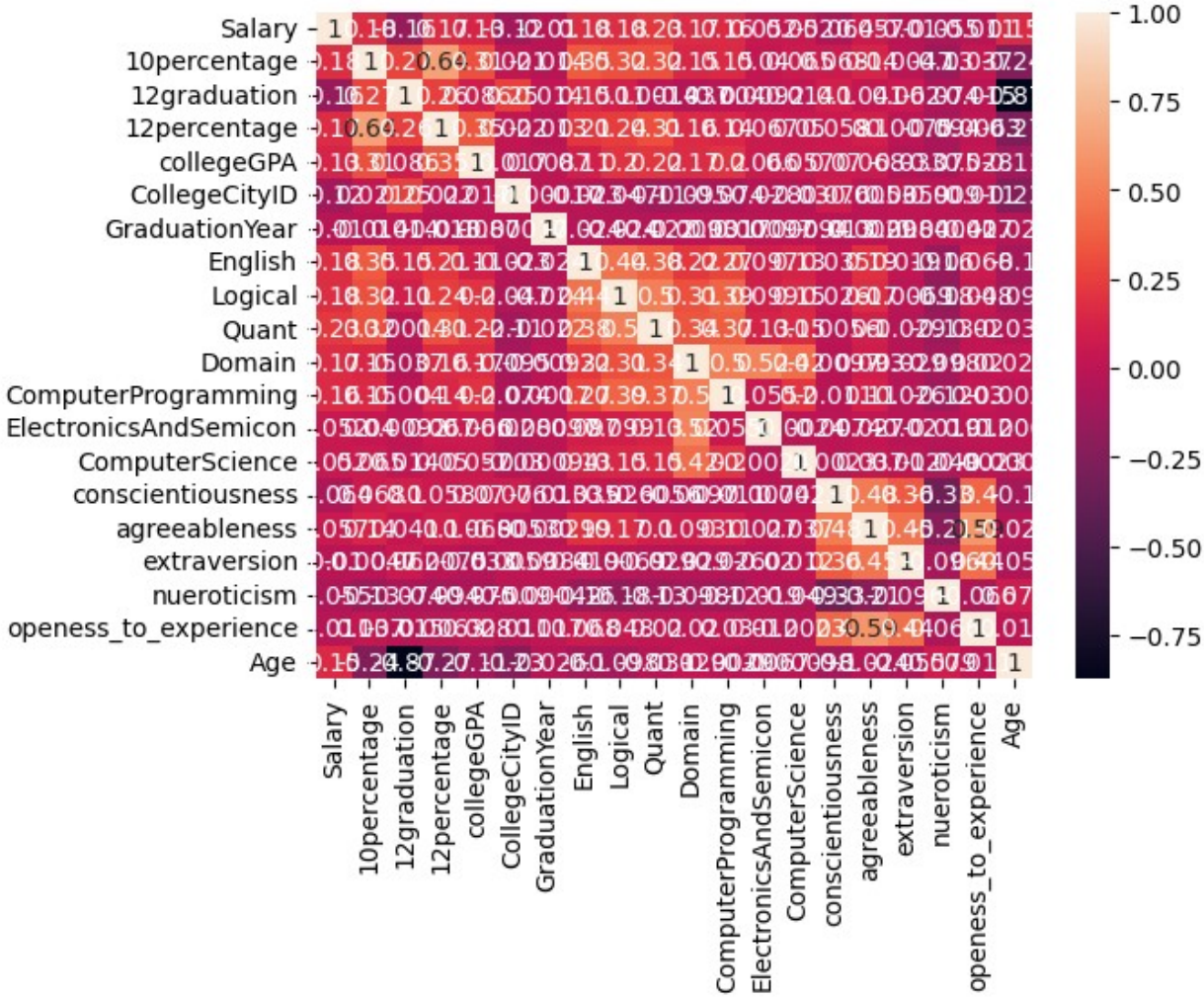
```
0.002947

                        ElectronicsAndSemicon  ComputerScience  \
Salary                               0.051879         0.052012
10percentage                         0.039637         0.064602
12graduation                         0.009207         0.014311
12percentage                         0.066754         0.050151
collegeGPA                           0.065689         0.056808
CollegeCityID                       -0.028449        -0.030108
GraduationYear                       0.000966         0.000938
English                              0.097452         0.127781
Logical                              0.099187         0.148531
Quant                                0.128847         0.153973
Domain                               0.523350         0.422891
ComputerProgramming                  0.054712         0.204965
ElectronicsAndSemicon                1.000000        -0.002389
ComputerScience                     -0.002389         1.000000
conscientiousness                   -0.007363         0.002253
agreeableness                        0.026588         0.037118
extraversion                        -0.020110         0.011589
nueroticism                         -0.018600        -0.048989
openess_to_experience                0.011528        -0.002289
Age                                 -0.006731        -0.000979

                        conscientiousness  agreeableness  extraversion
\
Salary                          -0.064148       0.057423     -0.010213

10percentage                     0.067657       0.136645     -0.004679

12graduation                     0.103329       0.041182      0.061956

12percentage                     0.058299       0.103998     -0.007486

collegeGPA                       0.069582       0.068282     -0.032684

CollegeCityID                    0.076432      -0.005264      0.005917

GraduationYear                  -0.013235      -0.002877      0.008397

English                          0.034943       0.194990      0.018755

Logical                          0.025876       0.167207     -0.006949

Quant                           -0.005639       0.103443     -0.028616

Domain                          -0.009745       0.092925     -0.028769

ComputerProgramming             -0.010519       0.112553     -0.025765

ElectronicsAndSemicon           -0.007363       0.026588     -0.020110
```

```
ComputerScience                  0.002253       0.037118      0.011589

conscientiousness                1.000000       0.481820      0.355537

agreeableness                    0.481820       1.000000      0.454369

extraversion                     0.355537       0.454369      1.000000

nueroticism                     -0.330312      -0.207480     -0.096491

openess_to_experience            0.395649       0.591541      0.435074

Age                             -0.101942      -0.023547     -0.055499
```

```
                        nueroticism  openess_to_experience       Age
Salary                    -0.054685              -0.011312  0.151714
10percentage              -0.132496               0.036692 -0.243880
12graduation              -0.074369              -0.015069 -0.872401
12percentage              -0.094369               0.006332 -0.265781
collegeGPA                -0.074859               0.028071 -0.113168
CollegeCityID             -0.008973              -0.010678 -0.230825
GraduationYear            -0.000417               0.016855 -0.025639
English                   -0.155528               0.067979 -0.101117
Logical                   -0.178781               0.048420 -0.098491
Quant                     -0.131895               0.020377 -0.031144
Domain                    -0.098408               0.019878  0.029500
ComputerProgramming       -0.123289               0.029728 -0.002947
ElectronicsAndSemicon     -0.018600               0.011528 -0.006731
ComputerScience           -0.048989              -0.002289 -0.000979
conscientiousness         -0.330312               0.395649 -0.101942
agreeableness             -0.207480               0.591541 -0.023547
extraversion              -0.096491               0.435074 -0.055499
nueroticism                1.000000              -0.065795  0.079076
openess_to_experience     -0.065795               1.000000  0.011274
Age                        0.079076               0.011274  1.000000
```

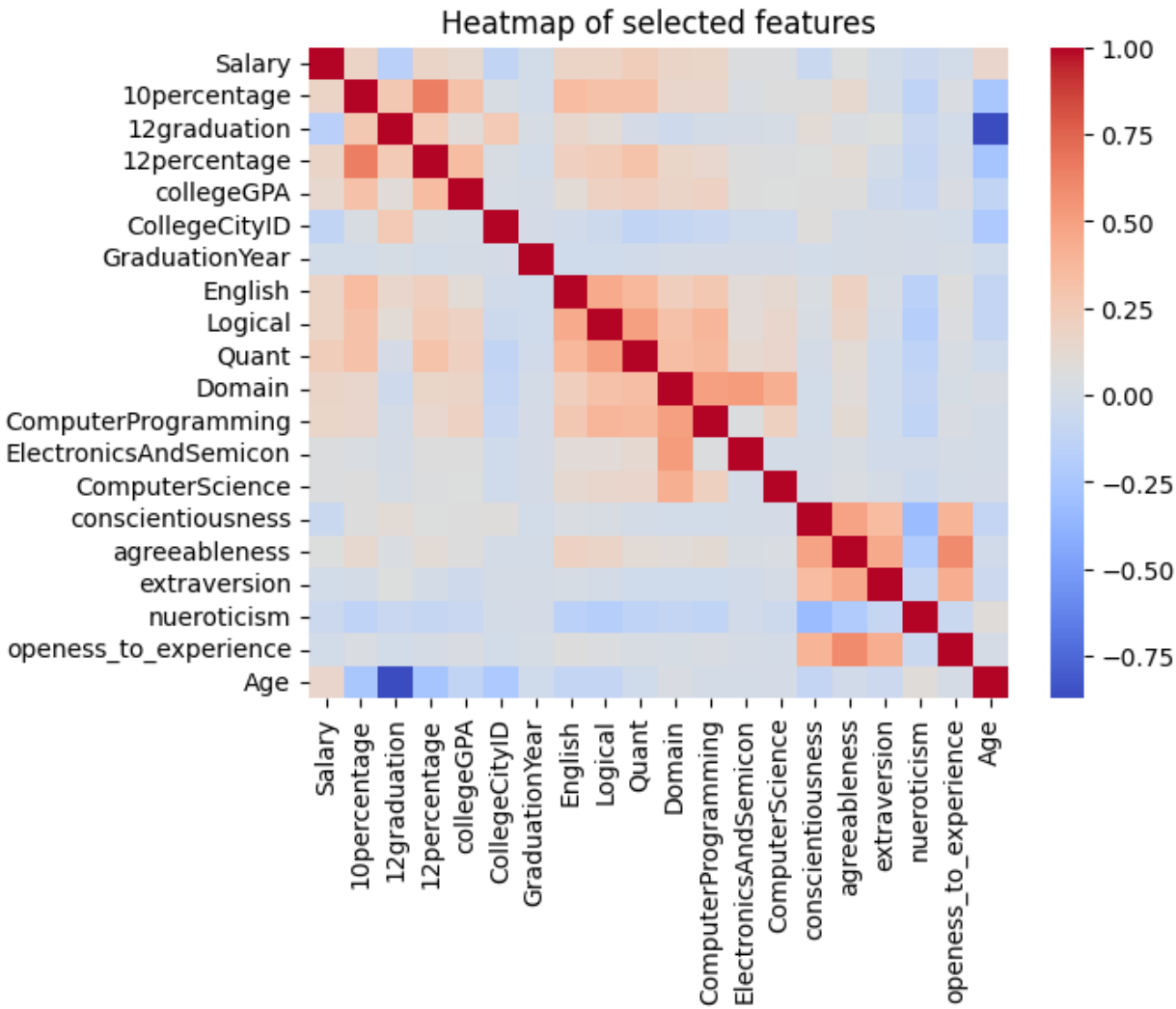
```
<ipython-input-39-19b38d16d7c6>:2: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it
will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
  correlation_matrix = df.corr()
```

```
# Heatmap of Correlation Matrix
sns.heatmap(correlation_matrix, annot=True)
plt.show()
```



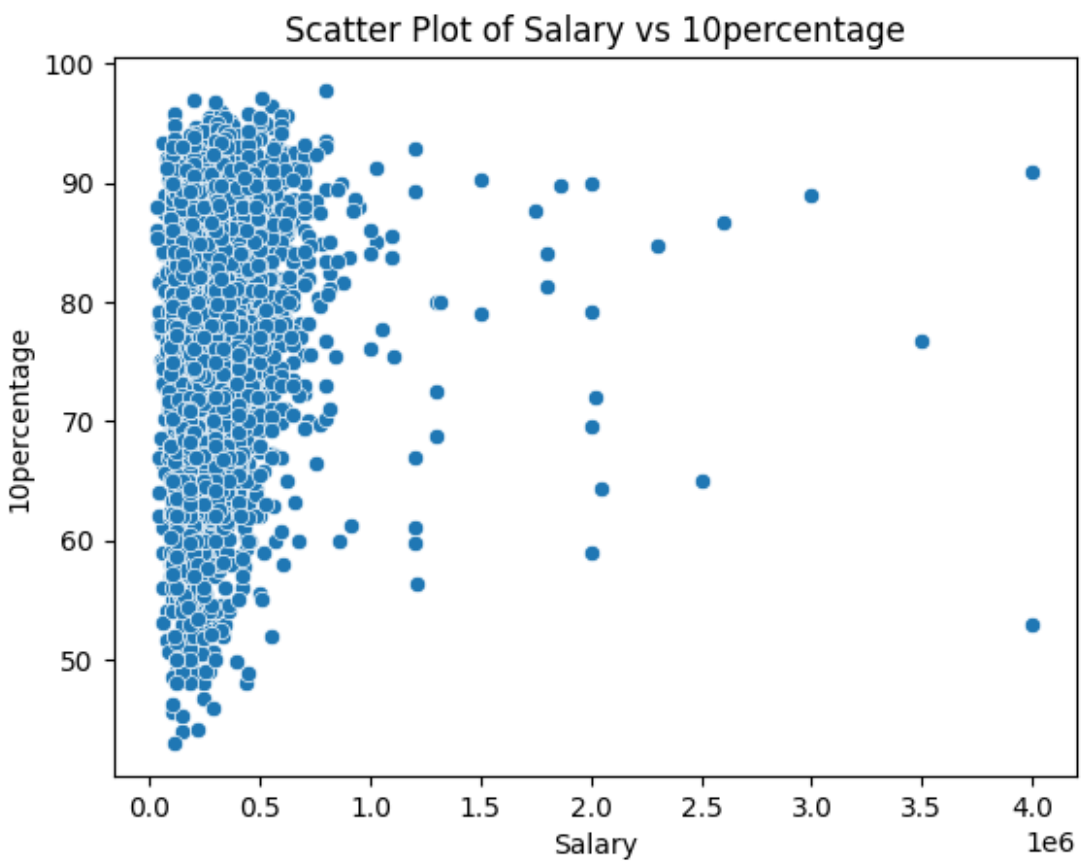

```
features = df.drop(columns=['DOL', 'DOJ', 'DOB'])

# Create heatmap using Seaborn
sns.heatmap(features.corr(), cmap='coolwarm')

# Add title
plt.title('Heatmap of selected features')

# Show plot
plt.show()
```

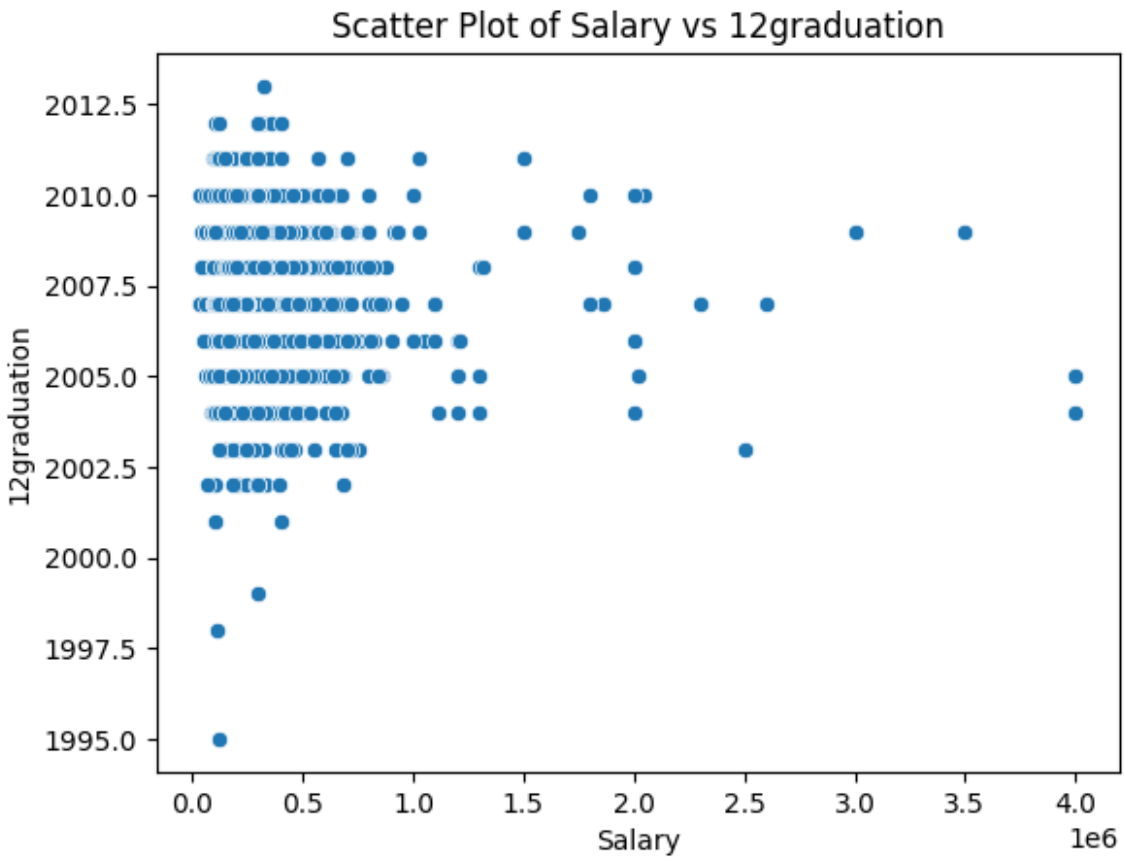
```
<ipython-input-41-36438b37e834>:4: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it
will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
  sns.heatmap(features.corr(), cmap='coolwarm')
```

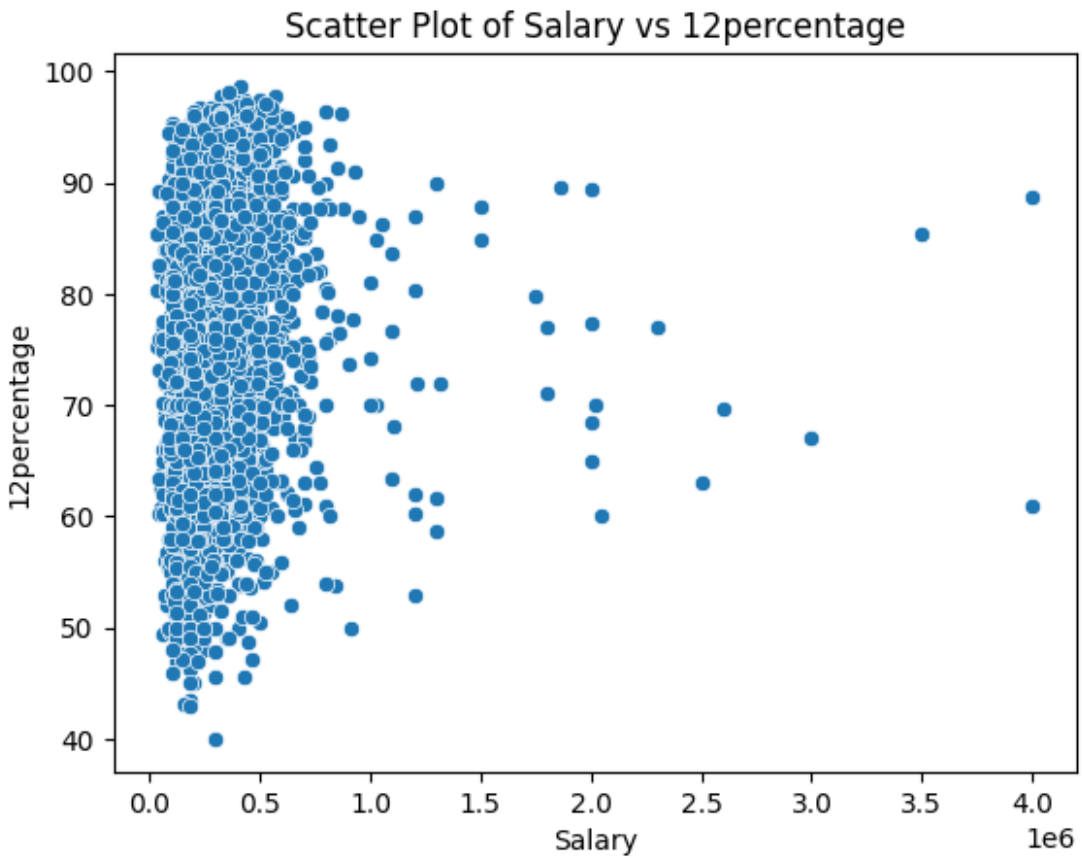


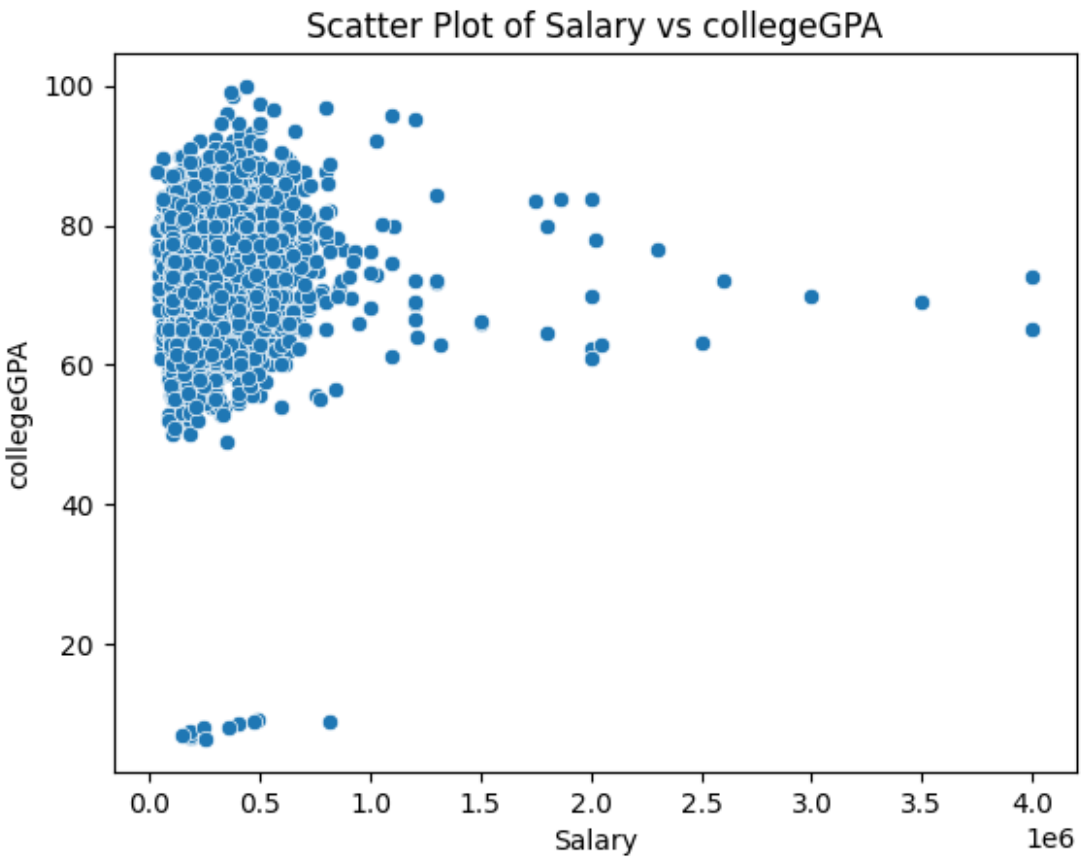

```
# prompt: obersvations from above heatmap main conclusion from heatmap

**Observations from the heatmap:**

* There is a strong positive correlation between ssc_p and hsc_p.
* There is a strong positive correlation between degree_p and mba_p.
* There is a moderate positive correlation between ssc_p and degree_p.
* There is a moderate positive correlation between hsc_p and degree_p.
* There is a moderate positive correlation between degree_p and
etest_p.
* There is a moderate positive correlation between etest_p and mba_p.
* There is a weak positive correlation between age and workex.
* There is a weak negative correlation between gender and salary.

**Main conclusion from the heatmap:**

The heatmap shows that there are several positive correlations between
the numerical variables in the dataset. This suggests that these
```

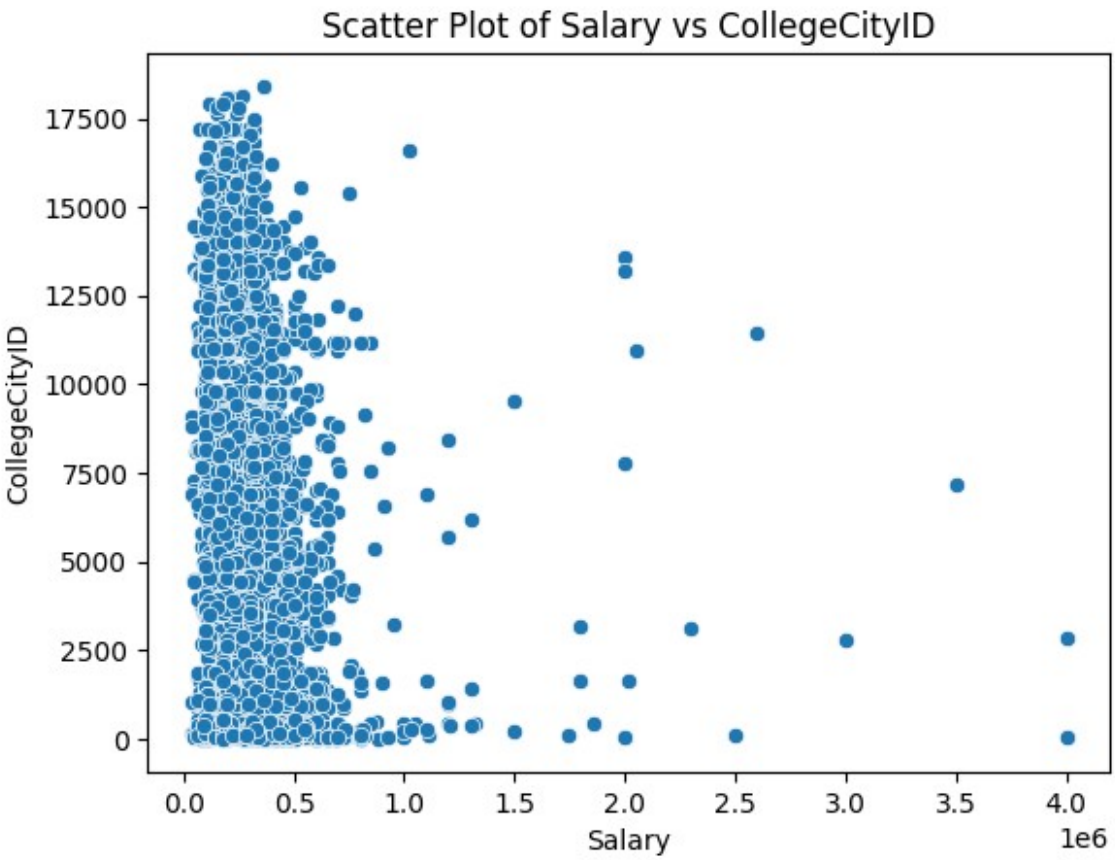
```
variables may be related to each other and could potentially be used
together in a predictive model.

# Scatter Plots
numerical_columns = list(df.select_dtypes(include=["number"]).columns)

for i in range(len(numerical_columns)):
    for j in range(i+1, len(numerical_columns)):
        plt.figure()
        sns.scatterplot(x=df[numerical_columns[i]],
y=df[numerical_columns[j]])
        plt.title(f"Scatter Plot of {numerical_columns[i]} vs
{numerical_columns[j]}")
        plt.show()
```

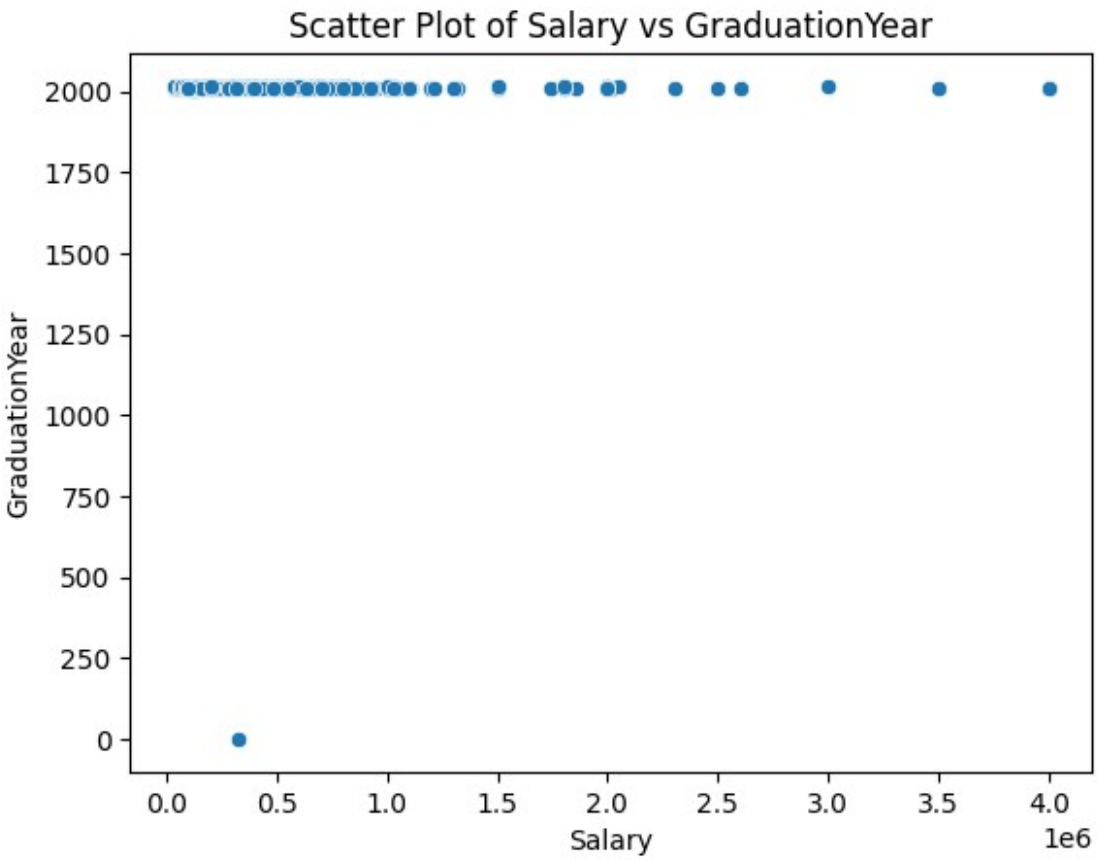


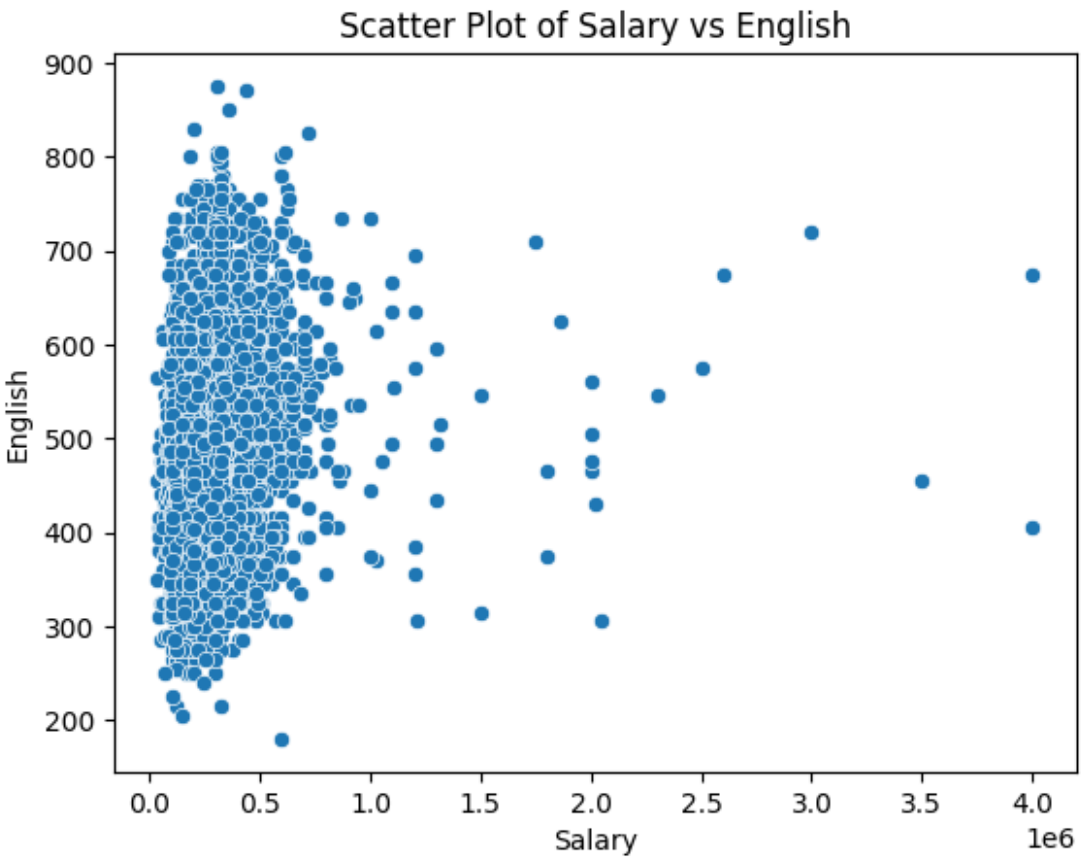

Scatter Plot of Salary vs 12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
12graduation
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

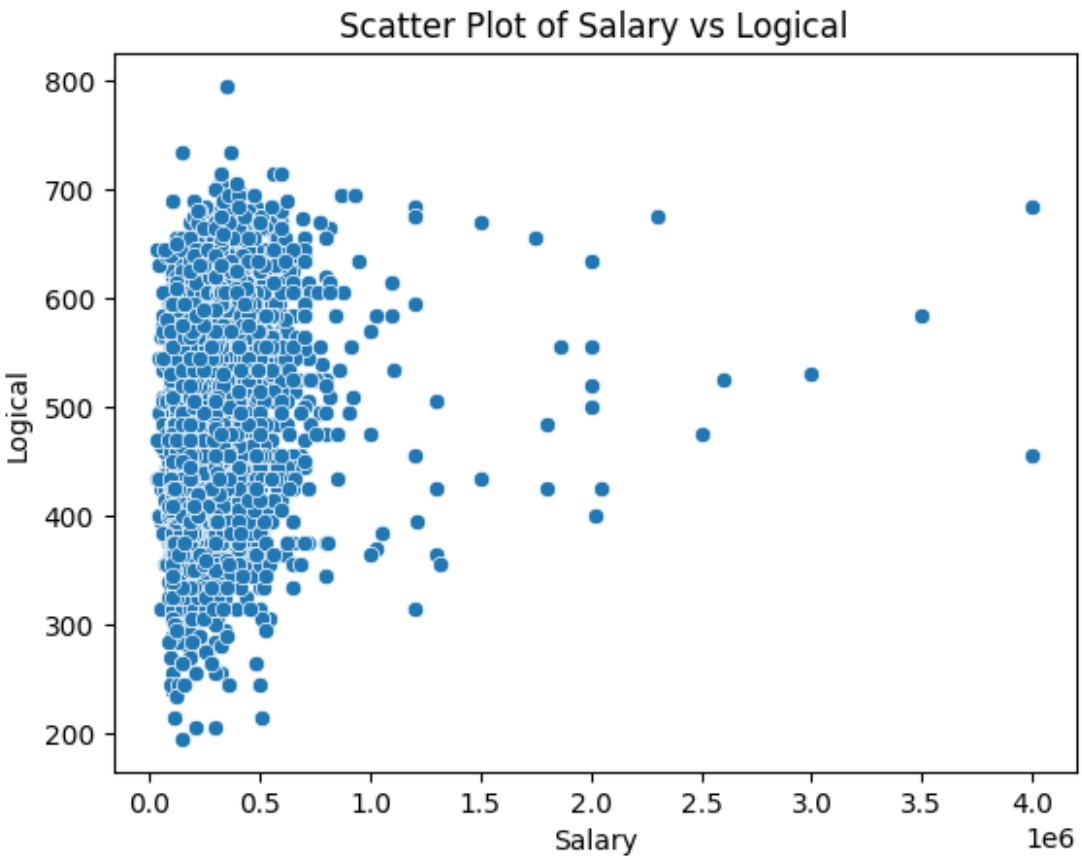

Scatter Plot of Salary vs 12percentage
12percentage
100
90
80
70
60
50
40
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

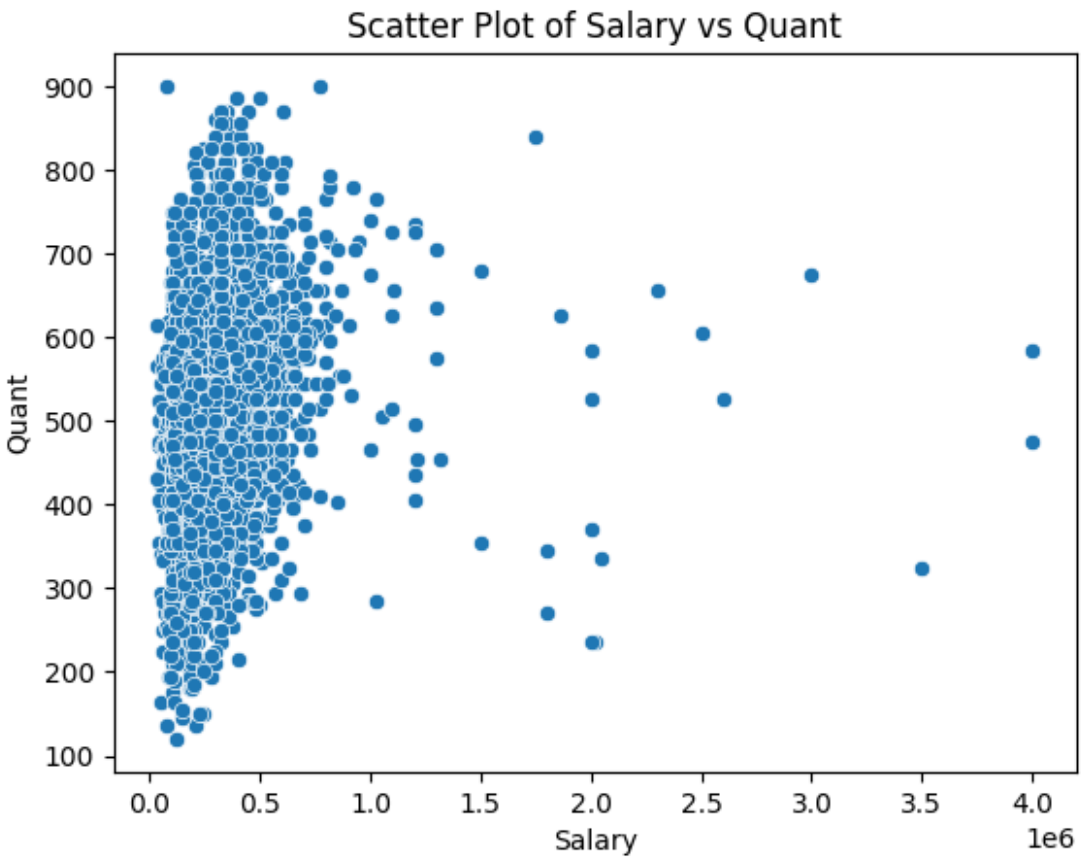

Scatter Plot of Salary vs collegeGPA
collegeGPA
100
80
60
40
20
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

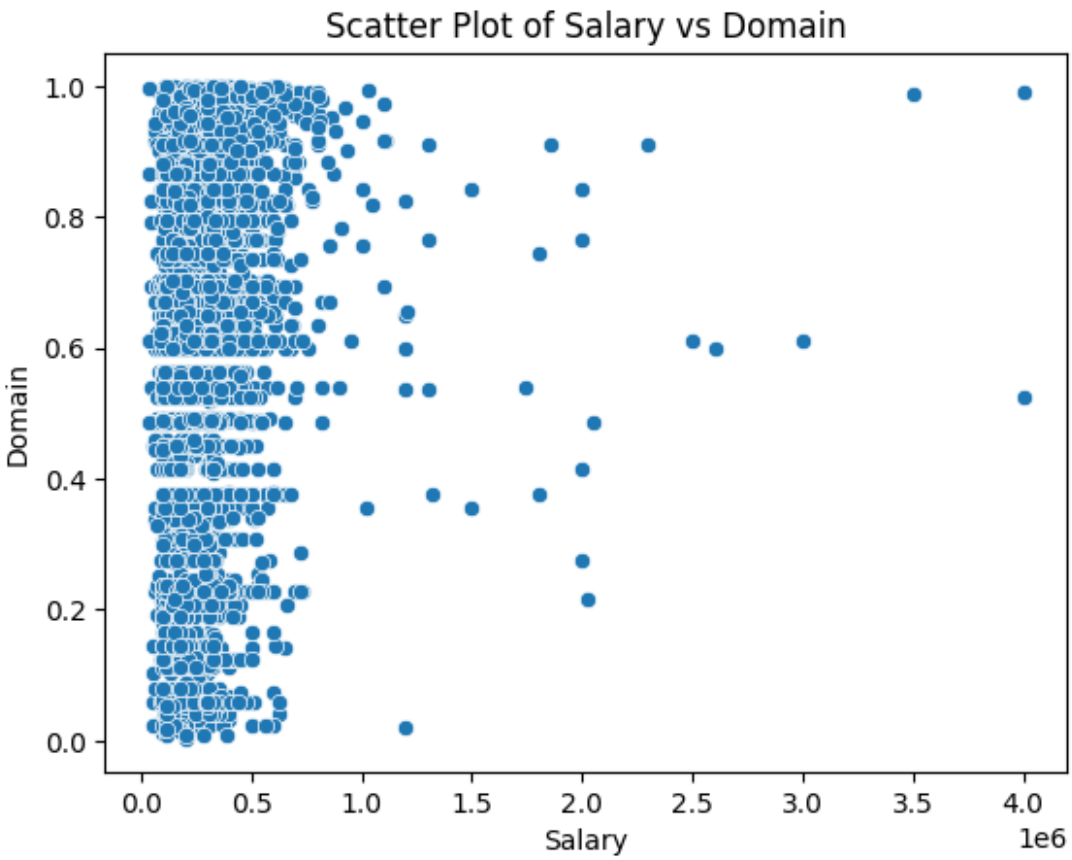

Scatter Plot of Salary vs CollegeCityID
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

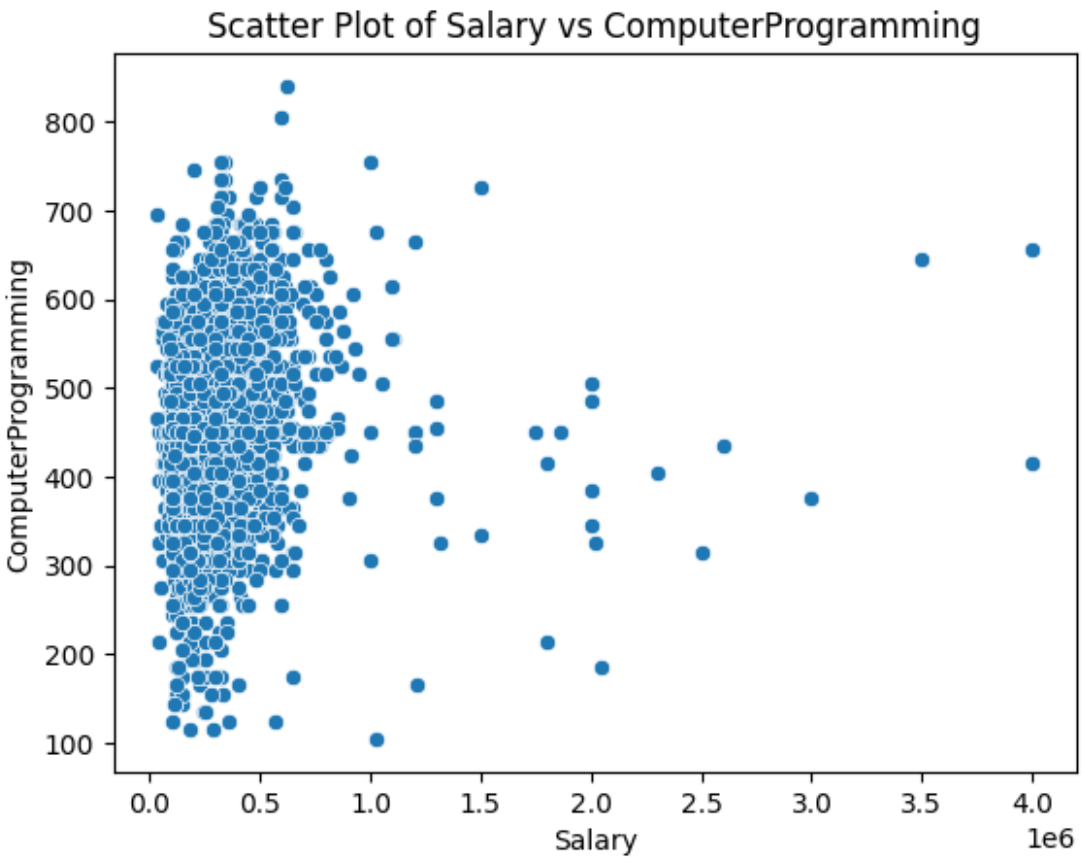

Scatter Plot of Salary vs GraduationYear
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

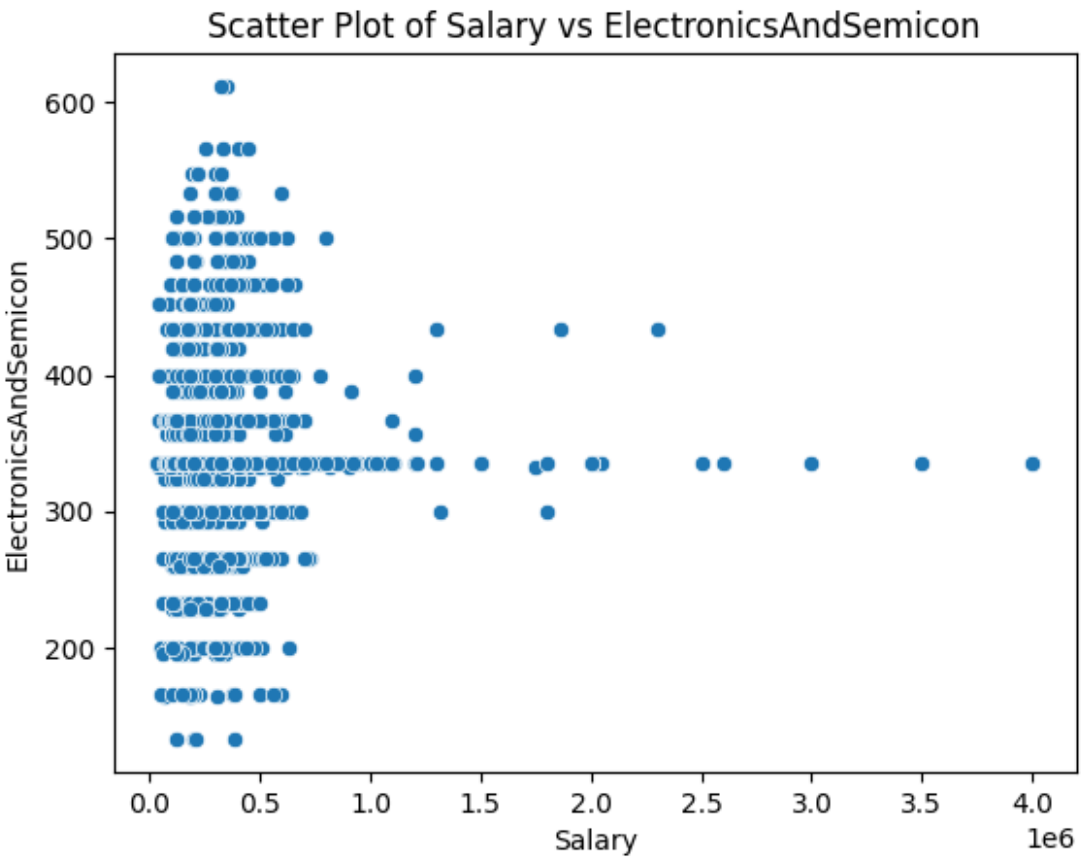

Scatter Plot of Salary vs English
English
Salary
1e6
900
800
700
600
500
400
300
200
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

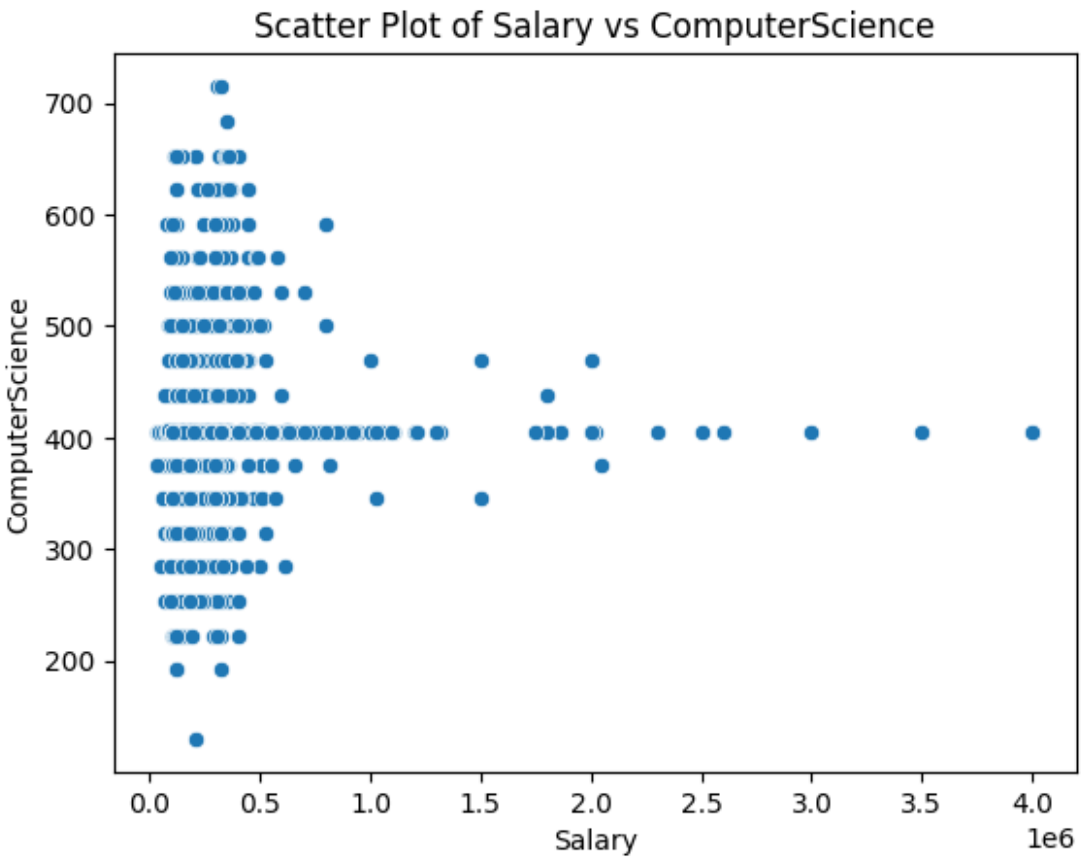

Scatter Plot of Salary vs Logical
Logical
800
700
600
500
400
300
200
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

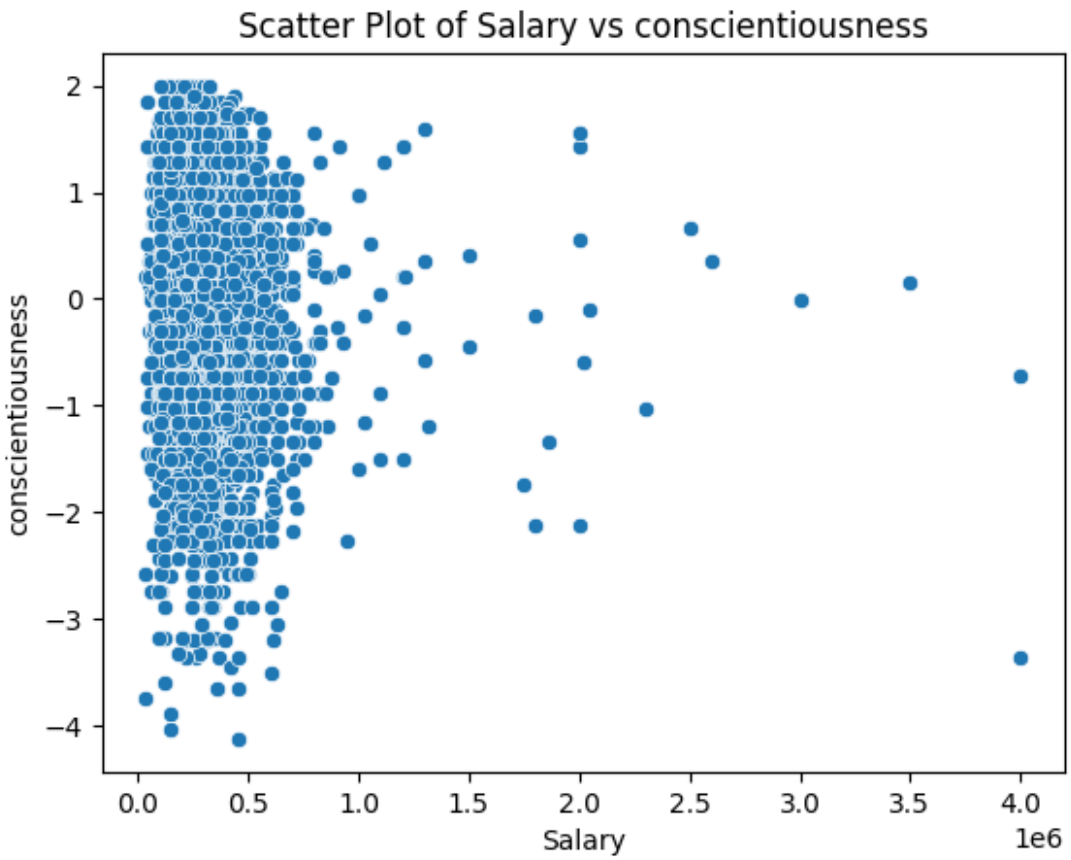

Scatter Plot of Salary vs Quant
Quant
900
800
700
600
500
400
300
200
100
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

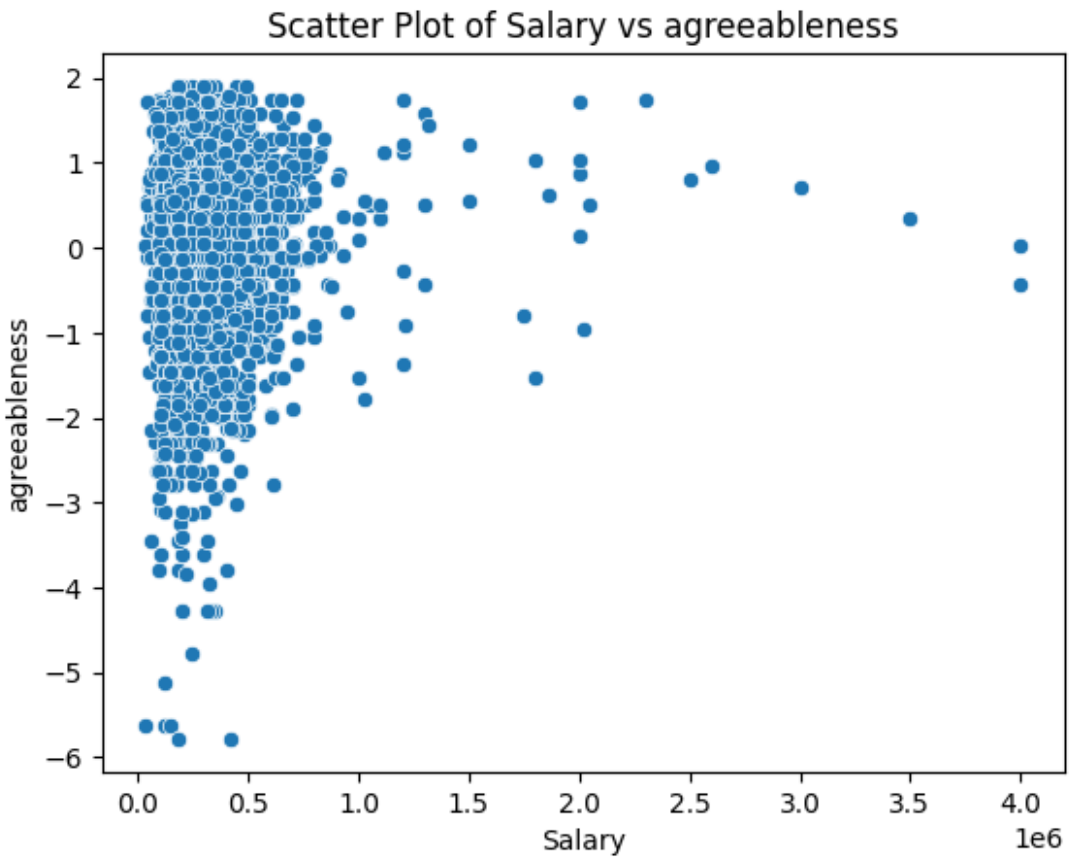

Scatter Plot of Salary vs Domain
Domain
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

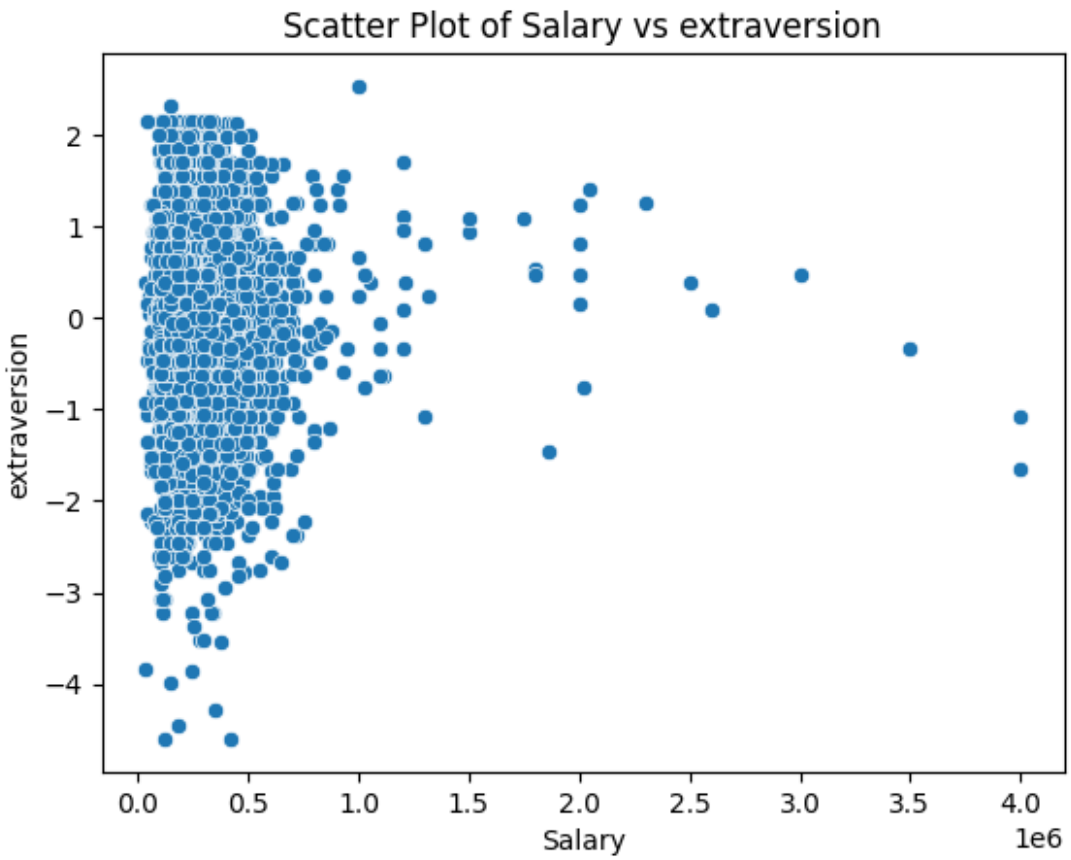

Scatter Plot of Salary vs ComputerProgramming
ComputerProgramming
800
700
600
500
400
300
200
100
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

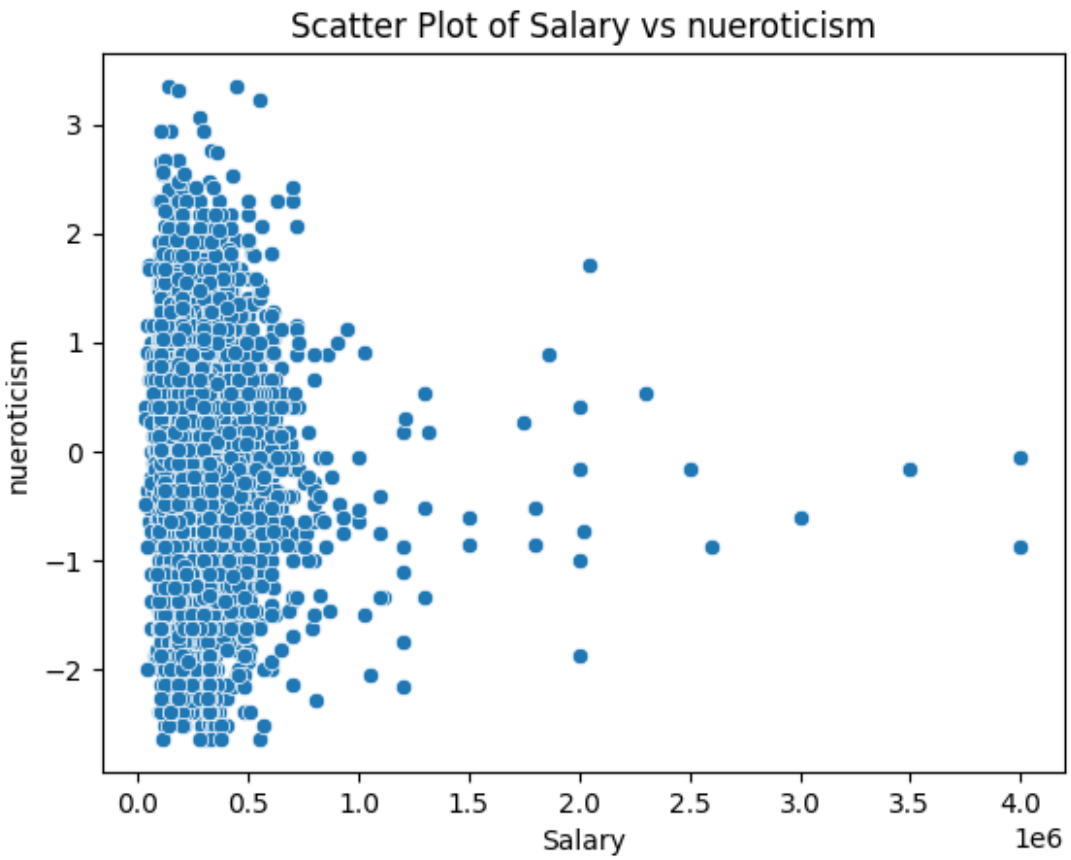

Scatter Plot of Salary vs ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

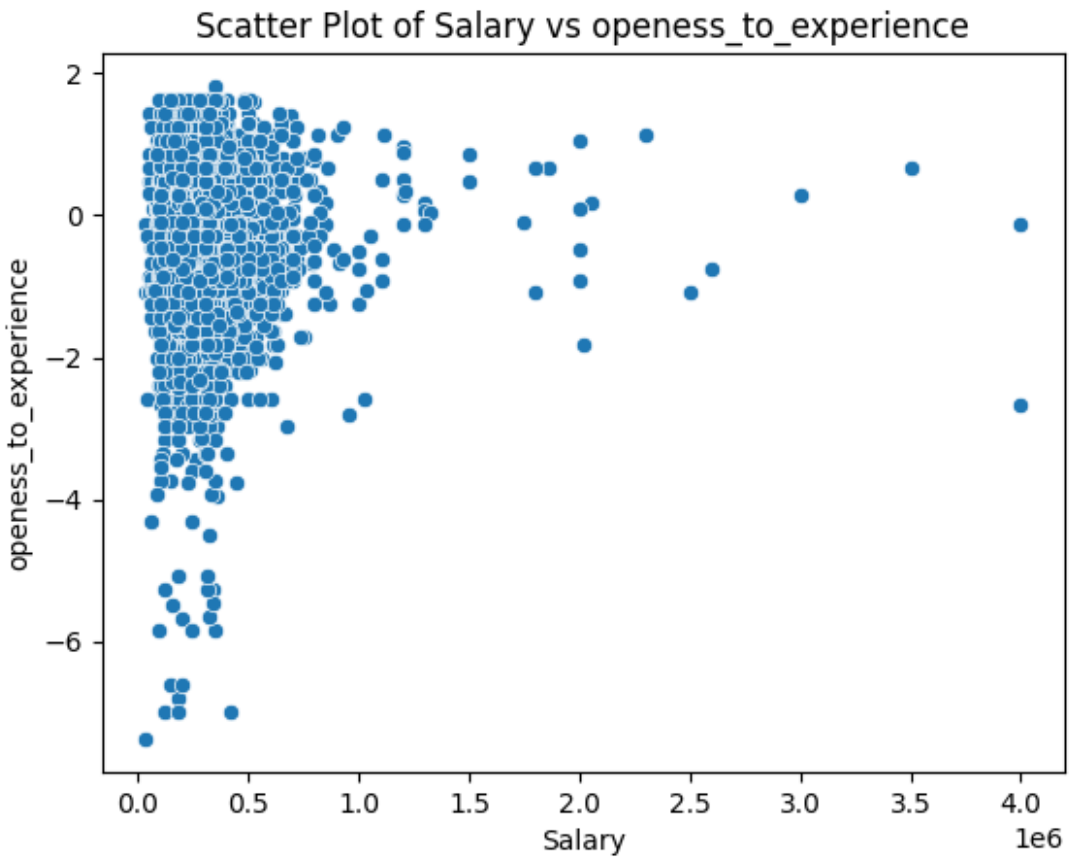

Scatter Plot of Salary vs ComputerScience
ComputerScience
700
600
500
400
300
200
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

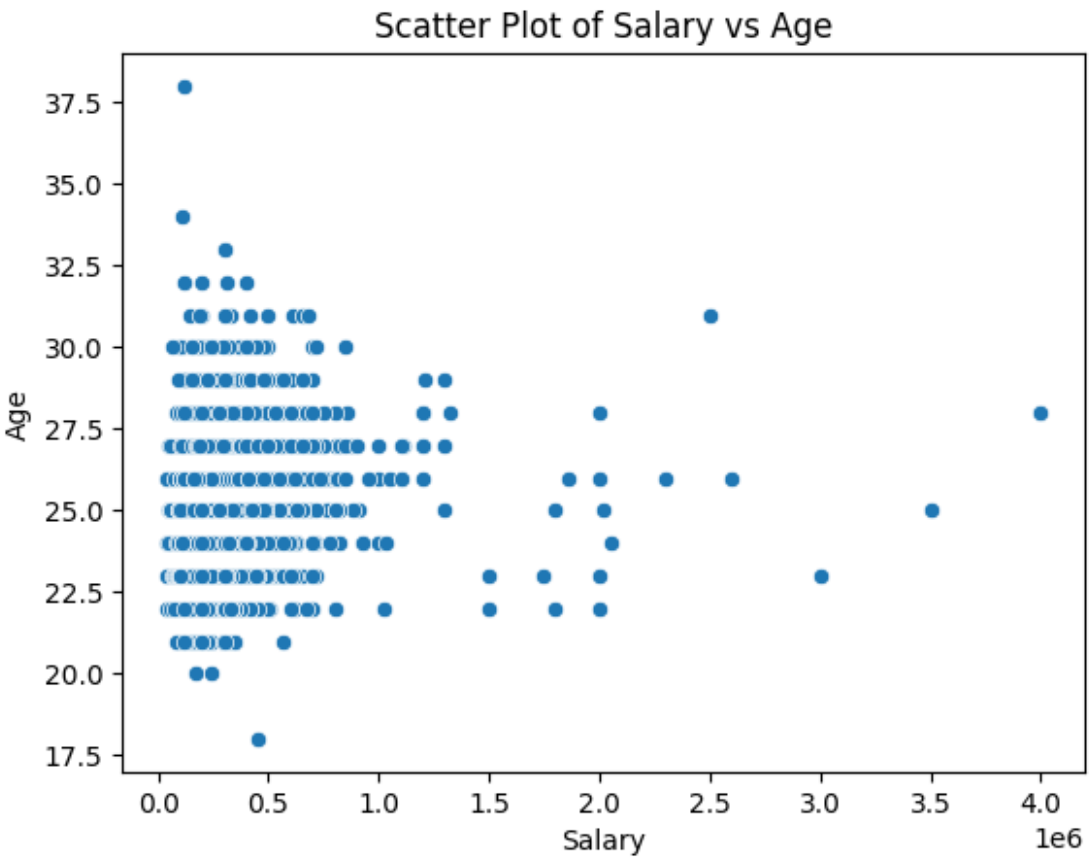

Scatter Plot of Salary vs conscientiousness
conscientiousness
2
1
0
−1
−2
−3
−4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

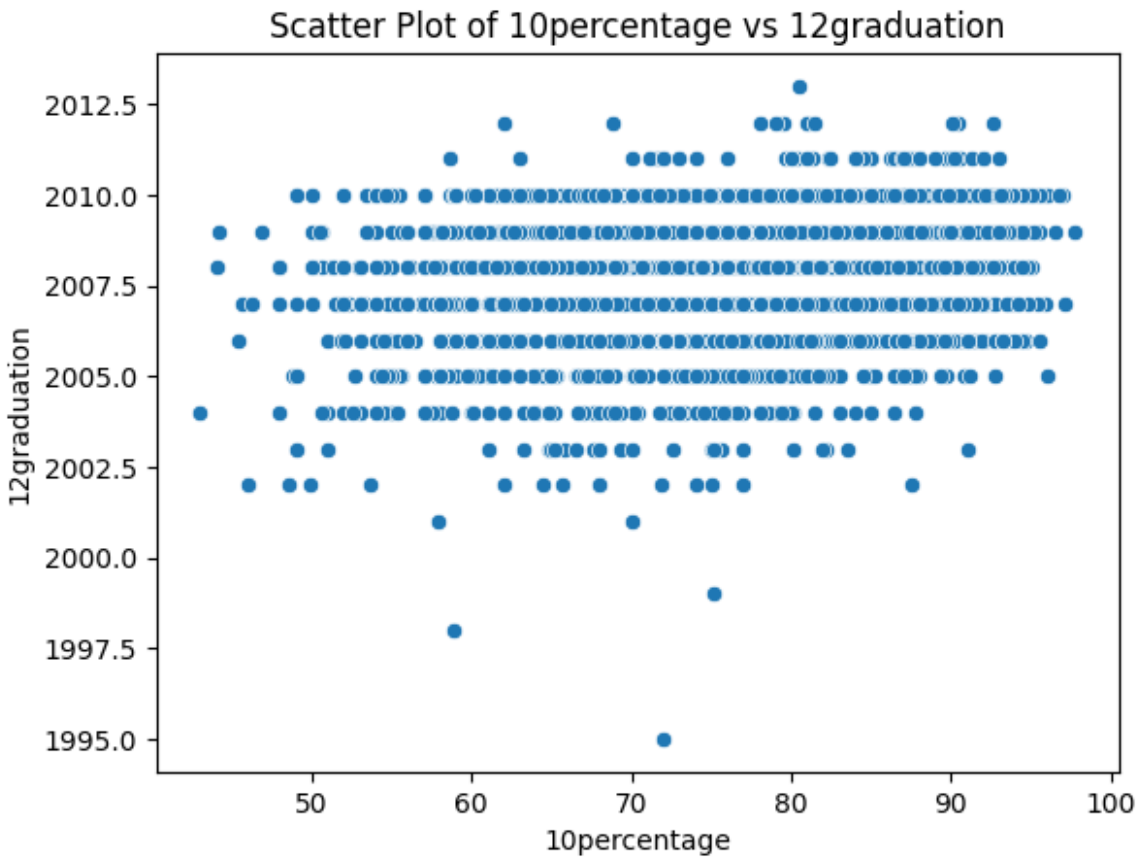

Scatter Plot of Salary vs agreeableness
agreeableness
Salary
1e6
2
1
0
−1
−2
−3
−4
−5
−6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

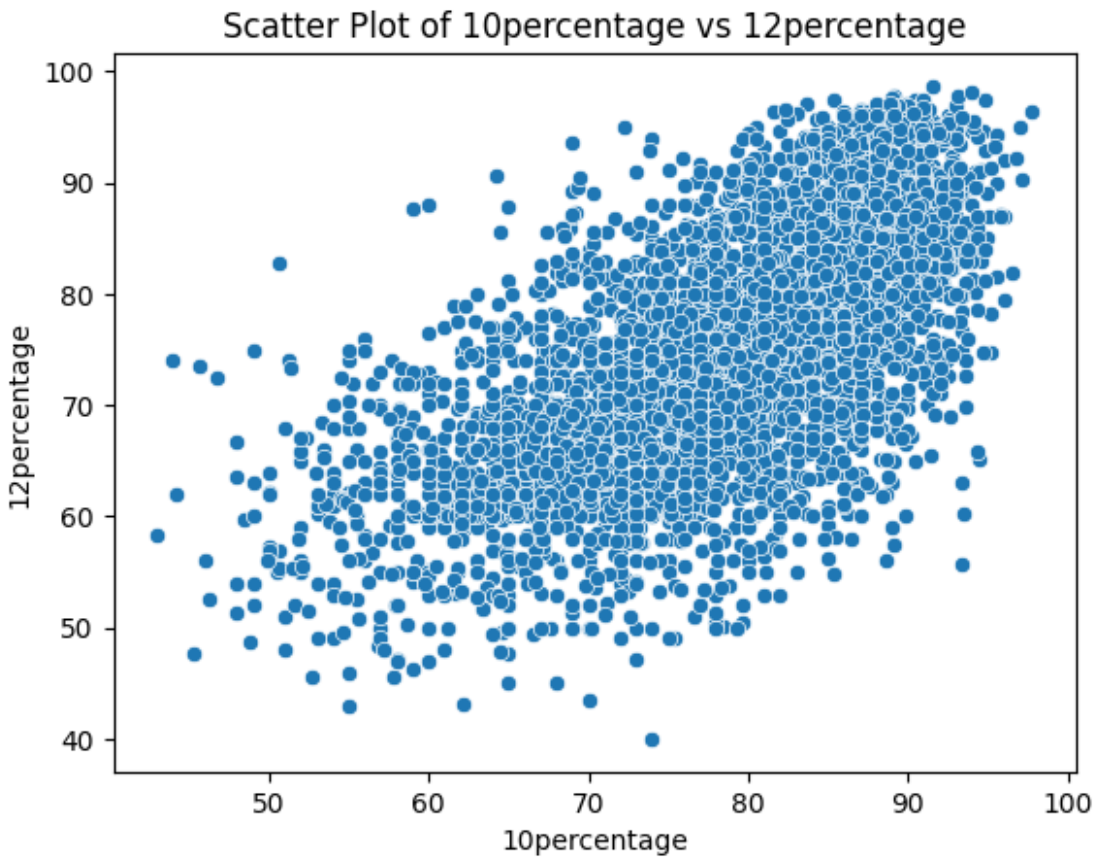

Scatter Plot of Salary vs extraversion
extraversion
2
1
0
−1
−2
−3
−4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

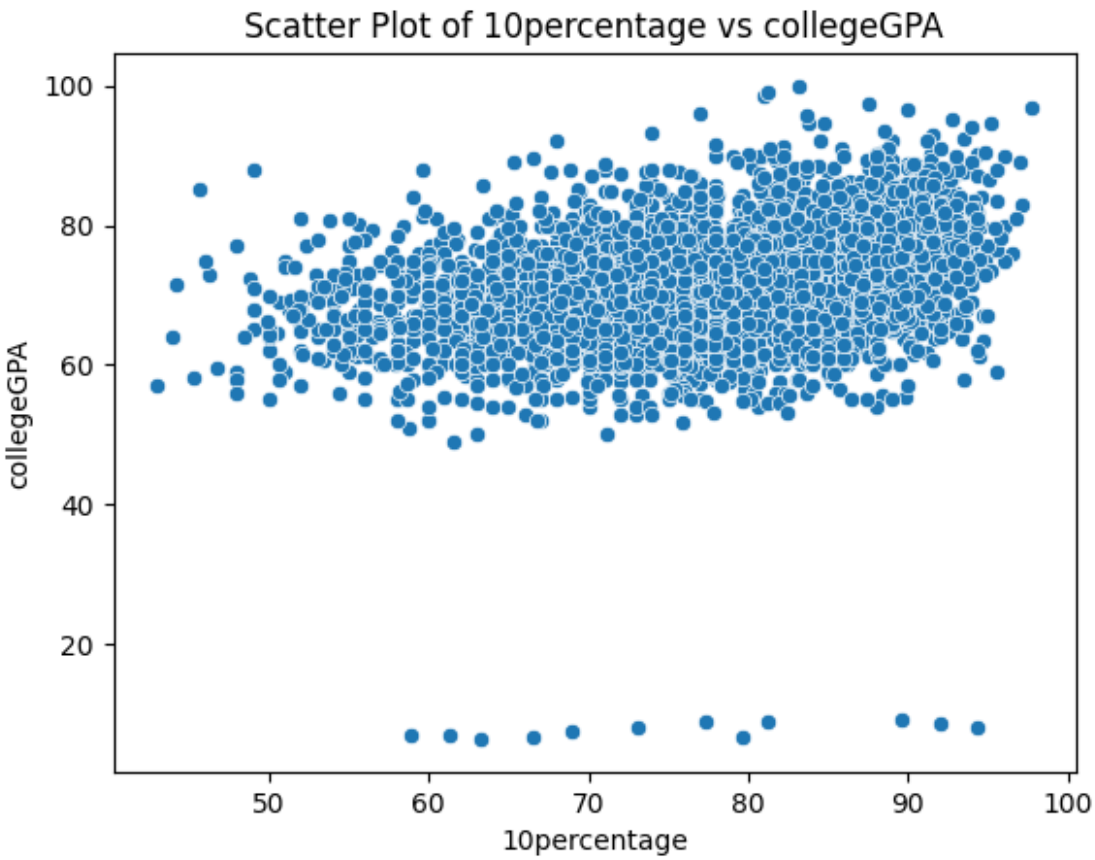

Scatter Plot of Salary vs nueroticism
nueroticism
3
2
1
0
−1
−2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

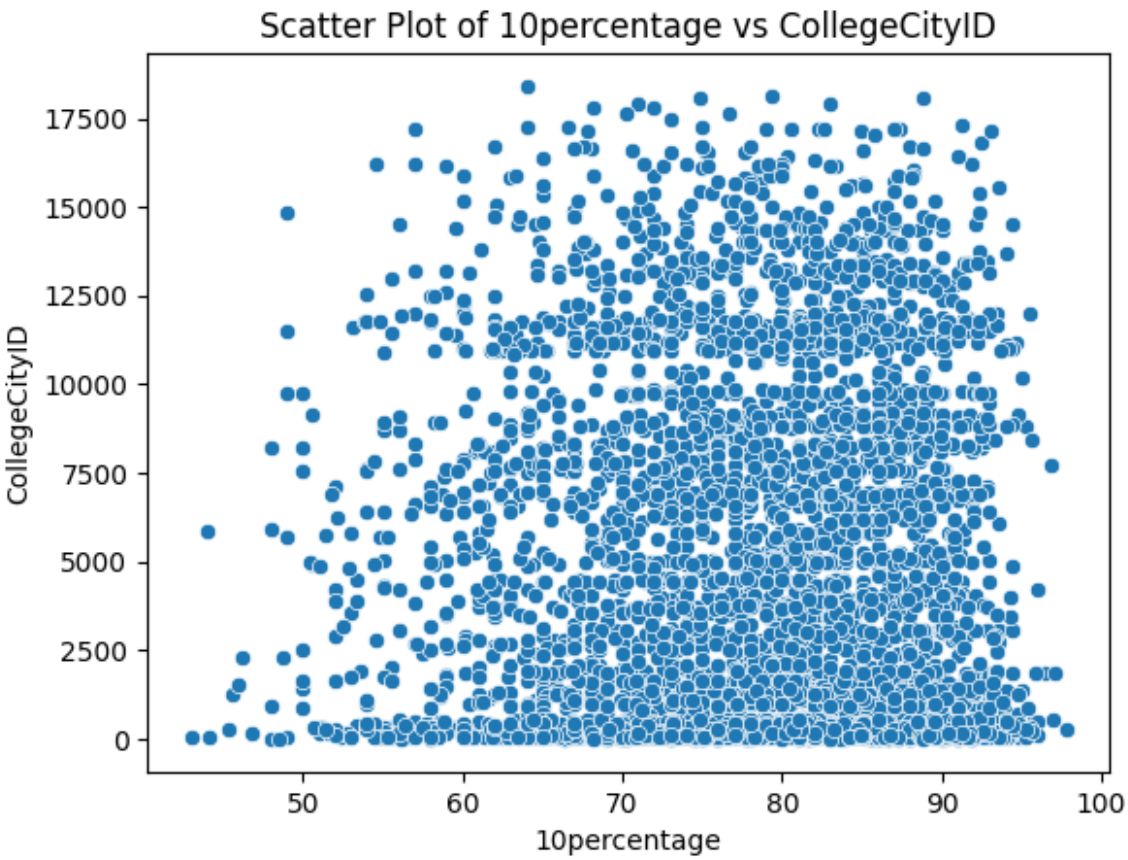

Scatter Plot of Salary vs openess_to_experience
openess_to_experience
2
0
−2
−4
−6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

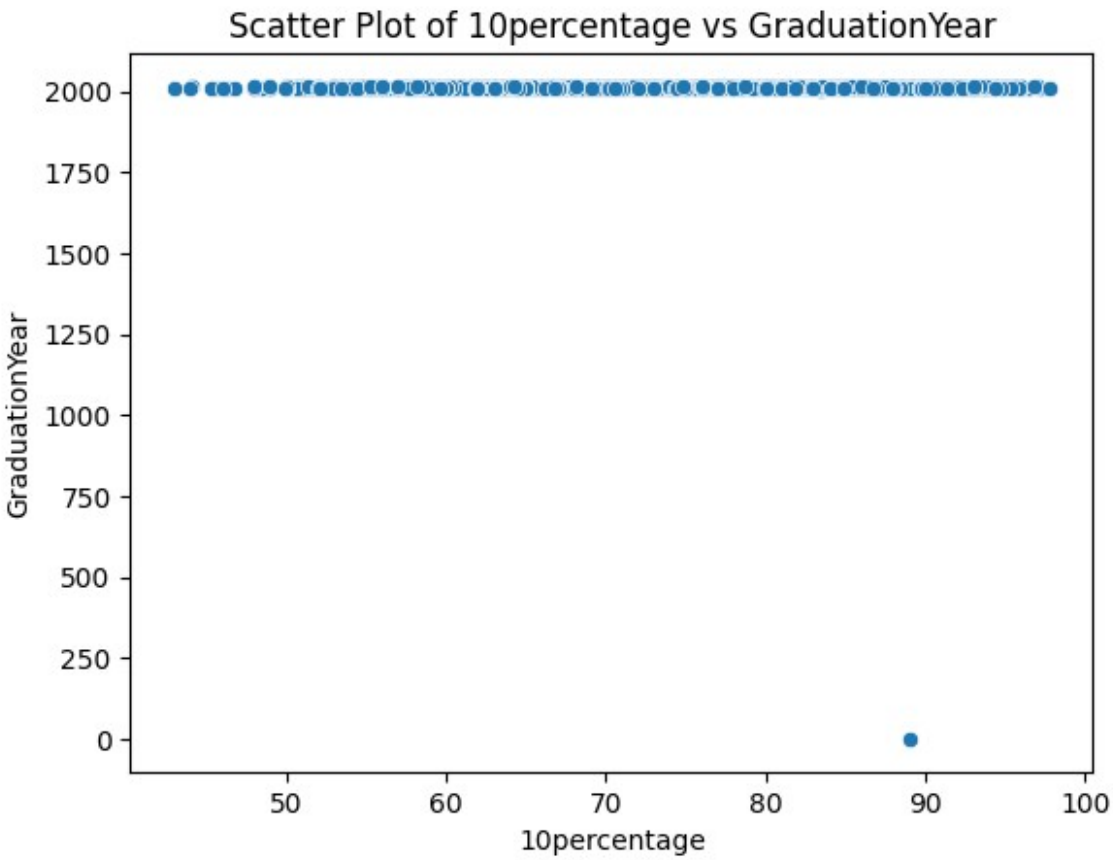

Scatter Plot of Salary vs Age
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Salary
1e6

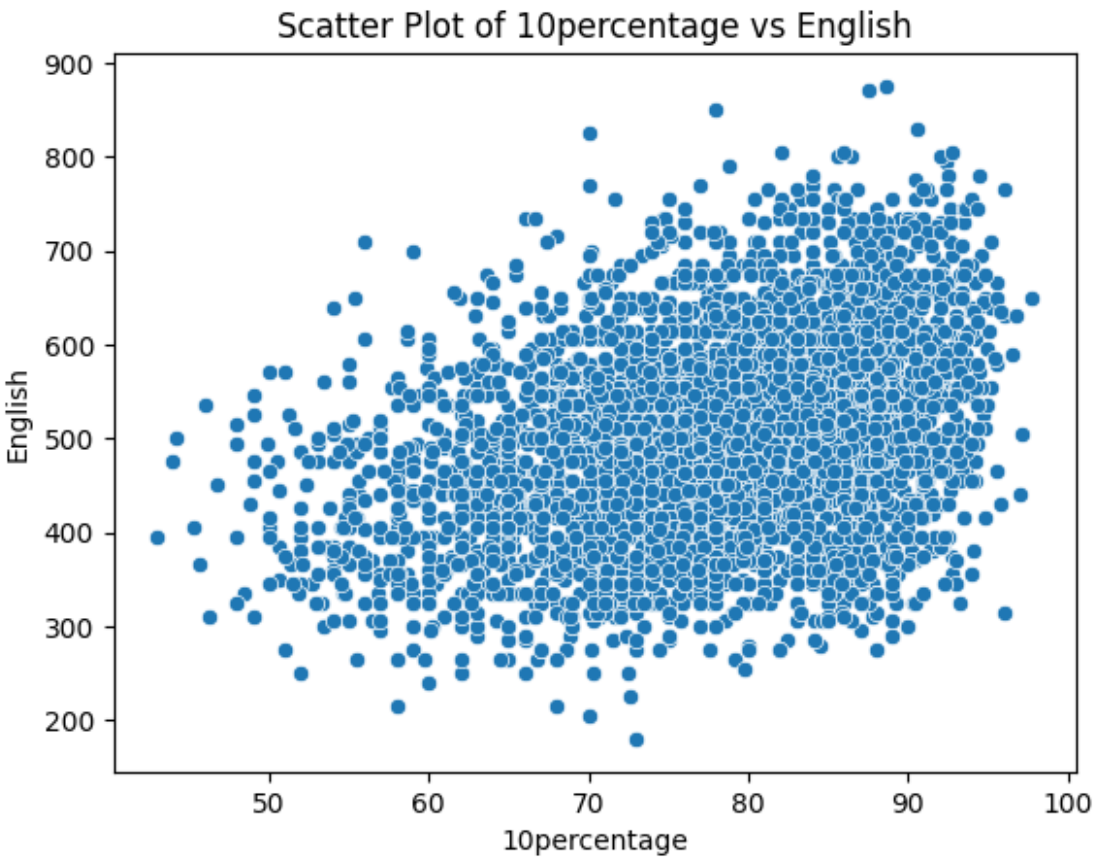

Scatter Plot of 10percentage vs 12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
12graduation
50
60
70
80
90
100
10percentage

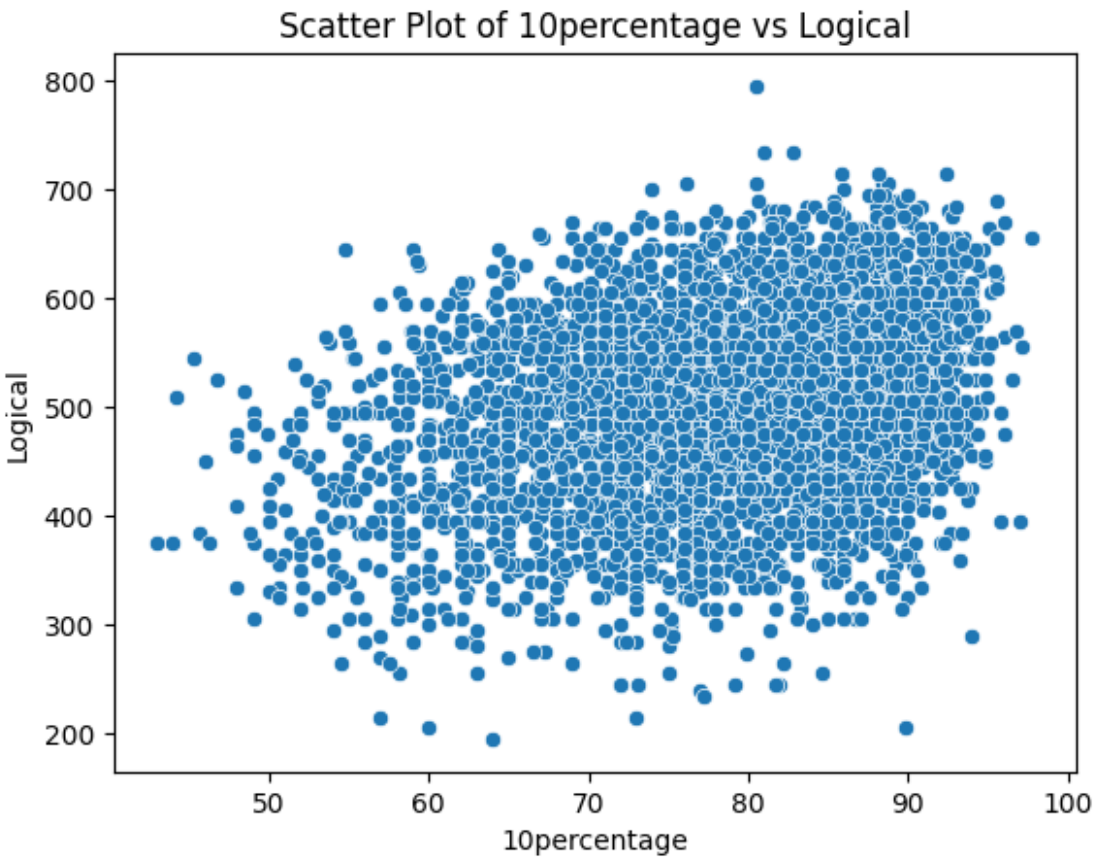

Scatter Plot of 10percentage vs 12percentage
12percentage
10percentage
100
90
80
70
60
50
40
50
60
70
80
90
100

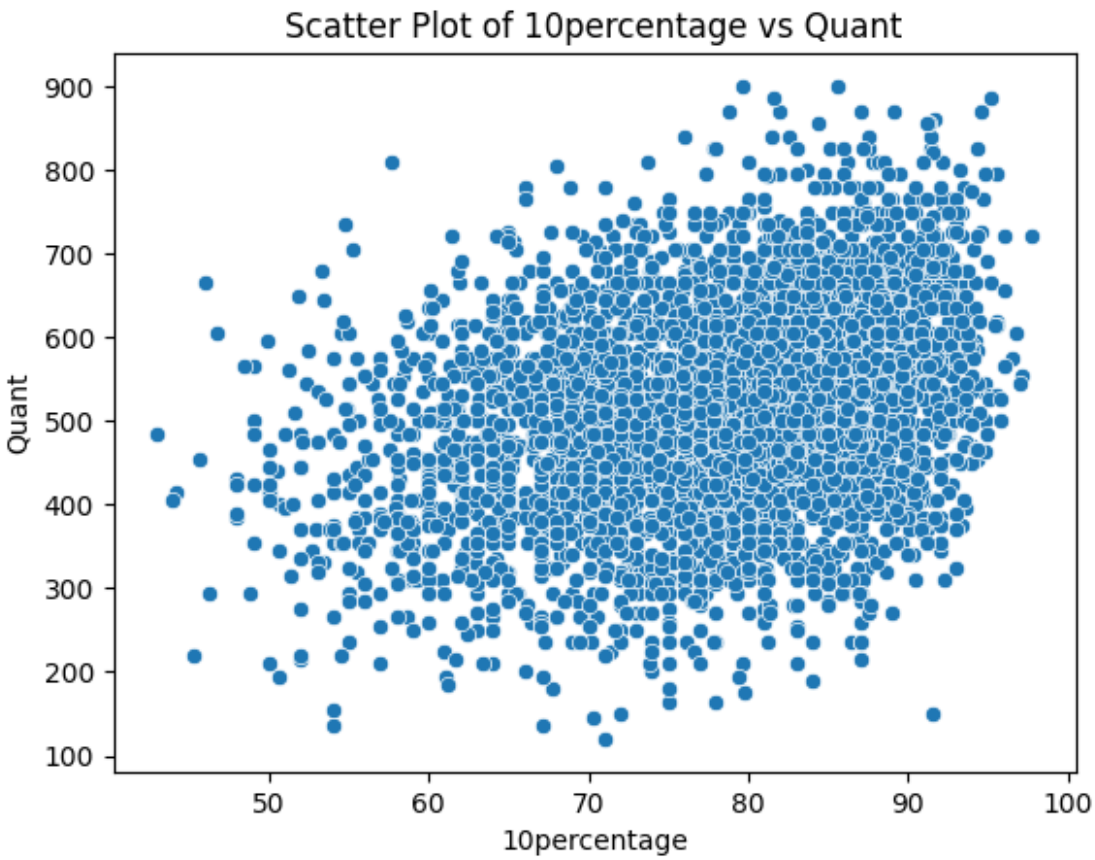

Scatter Plot of 10percentage vs collegeGPA
collegeGPA
10percentage
100
80
60
40
20
50
60
70
80
90
100

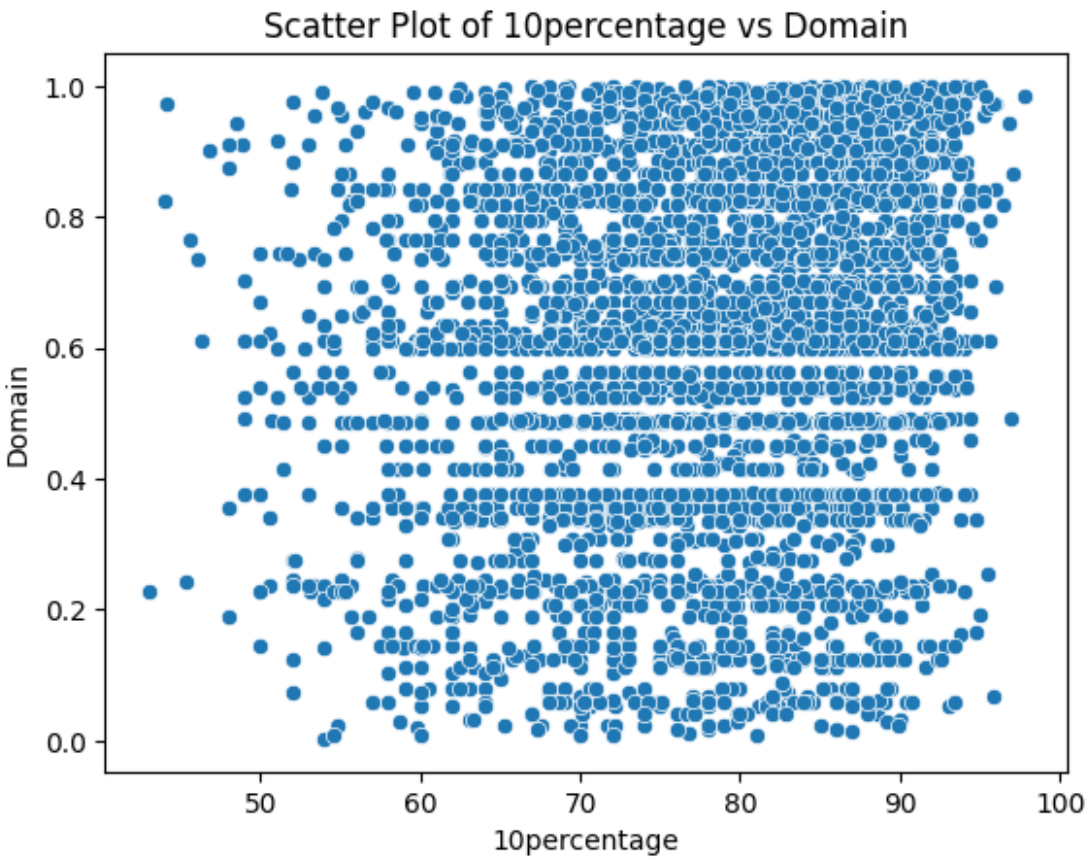

Scatter Plot of 10percentage vs CollegeCityID
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
50
60
70
80
90
100
10percentage

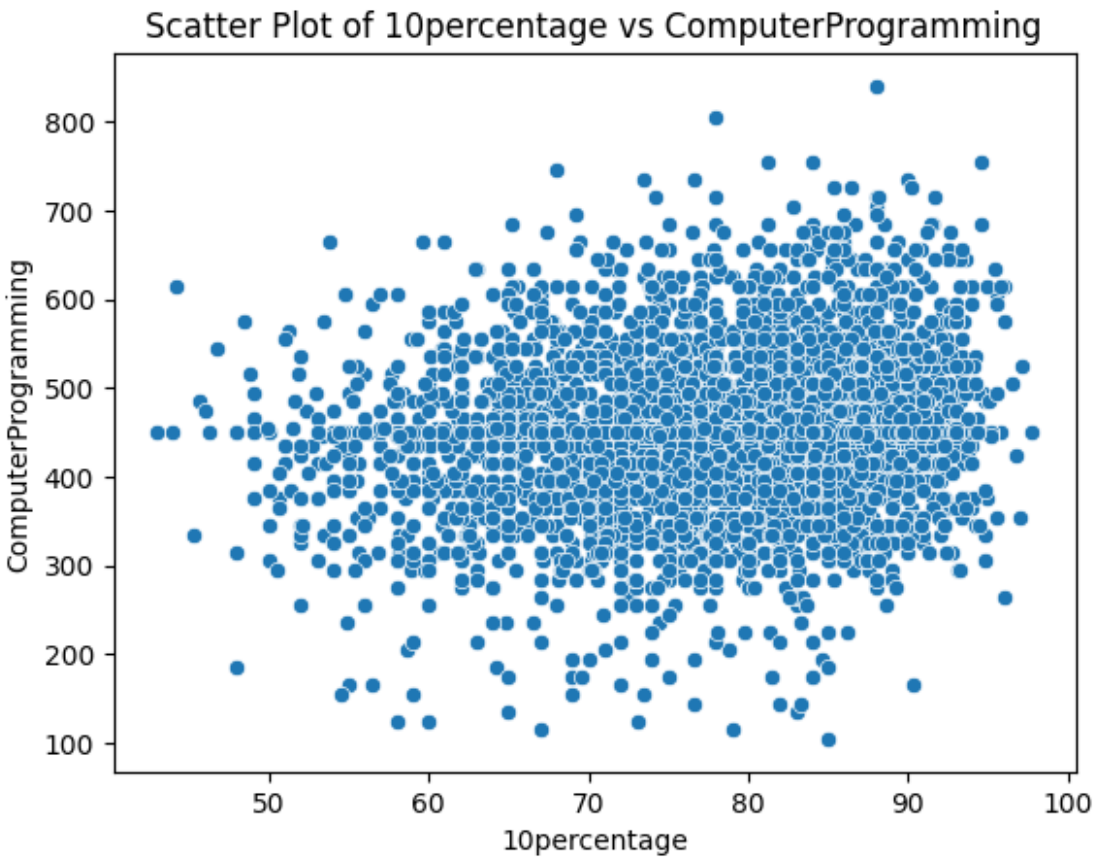

Scatter Plot of 10percentage vs GraduationYear
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
50
60
70
80
90
100
10percentage

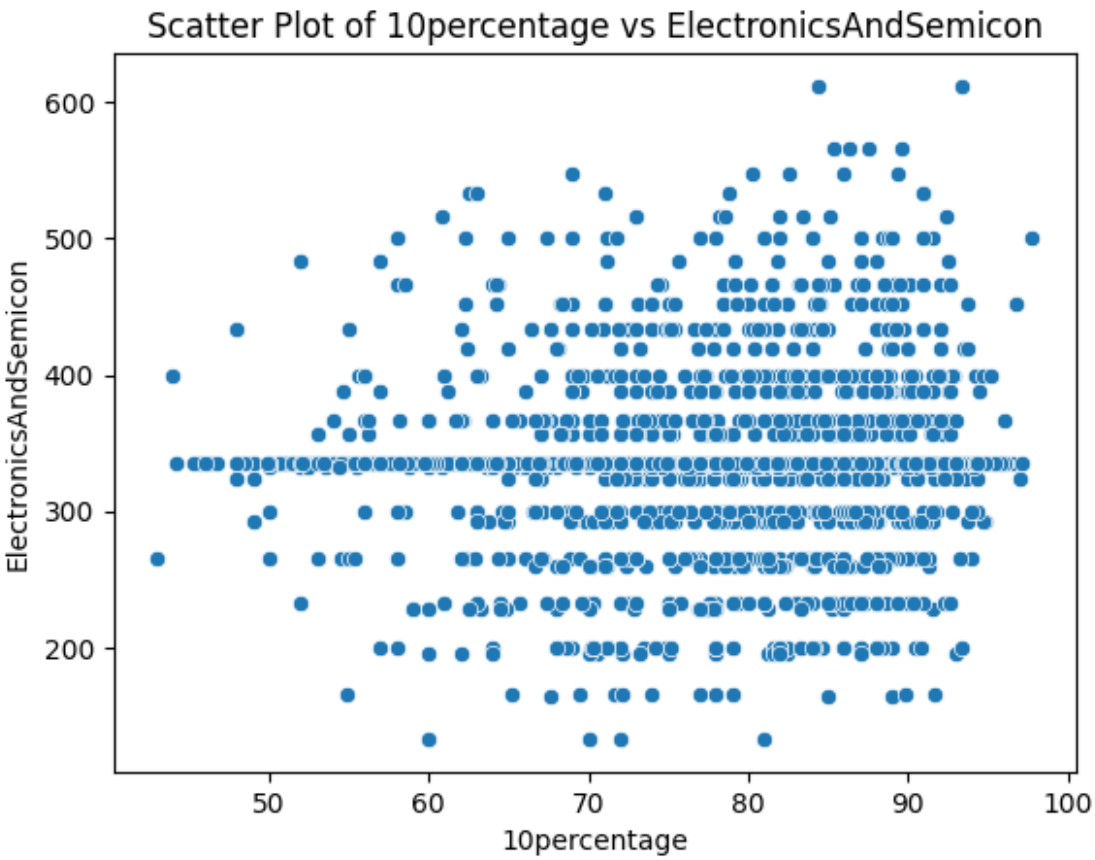

Scatter Plot of 10percentage vs English
English
10percentage
900
800
700
600
500
400
300
200
50
60
70
80
90
100

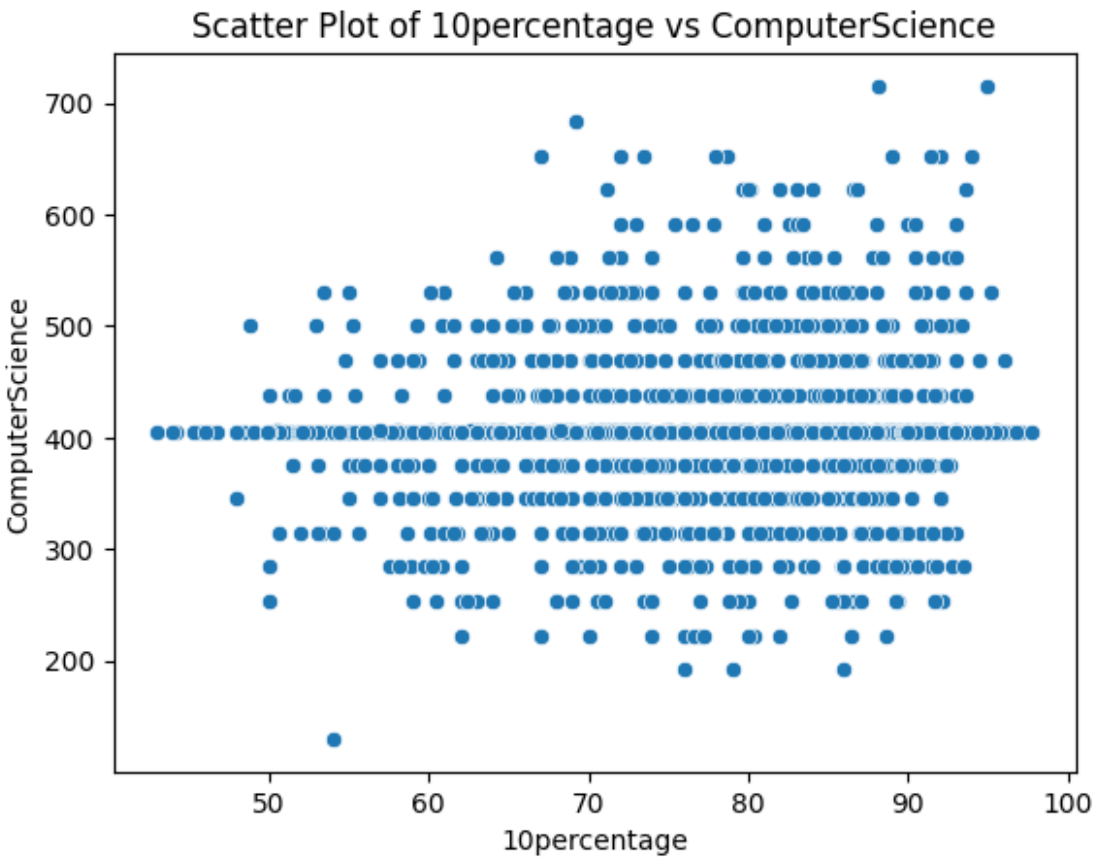

Scatter Plot of 10percentage vs Logical
800
700
600
500
400
300
200
Logical
50
60
70
80
90
100
10percentage

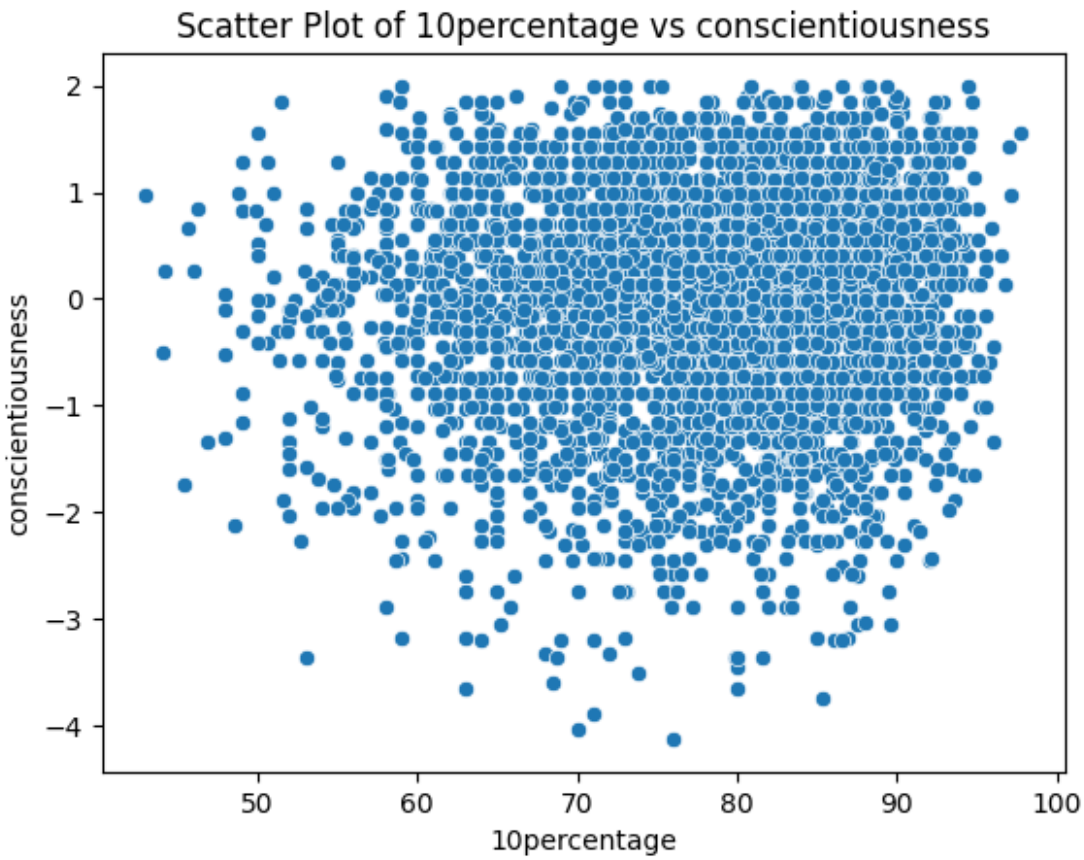

Scatter Plot of 10percentage vs Quant
Quant
10percentage
900
800
700
600
500
400
300
200
100
50
60
70
80
90
100

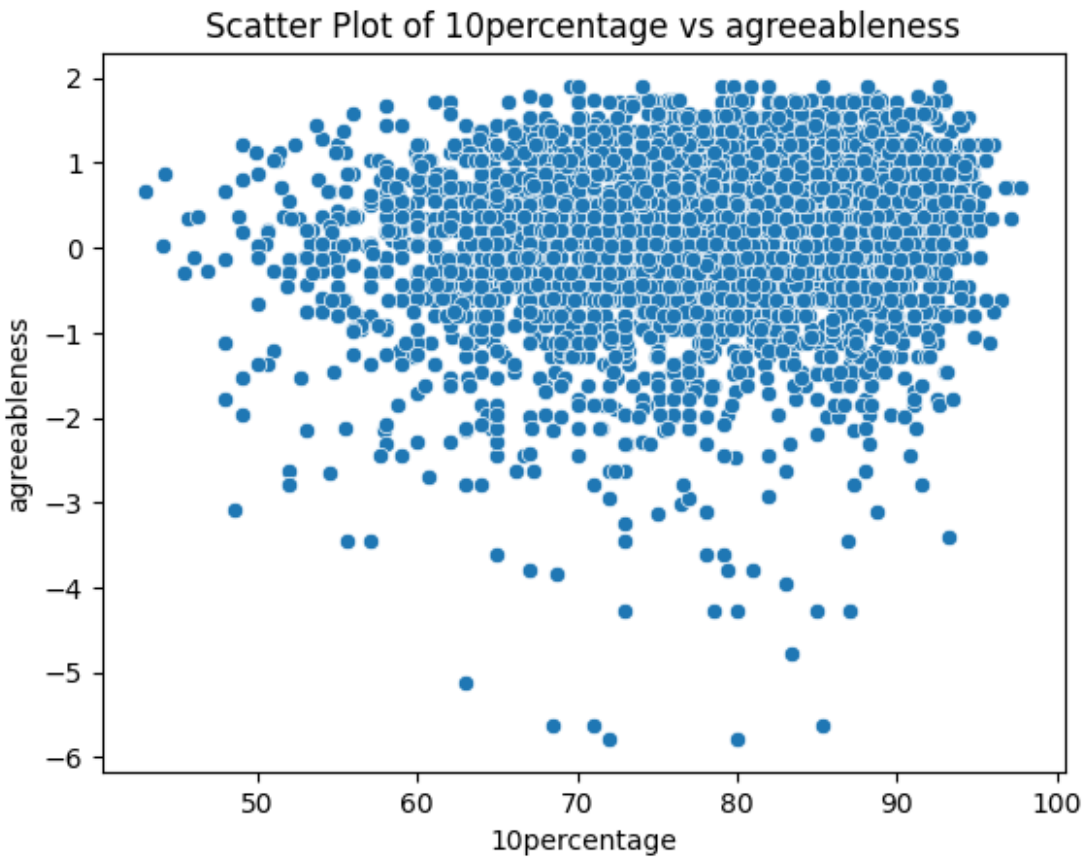

Scatter Plot of 10percentage vs Domain
Domain
10percentage
1.0
0.8
0.6
0.4
0.2
0.0
50
60
70
80
90
100

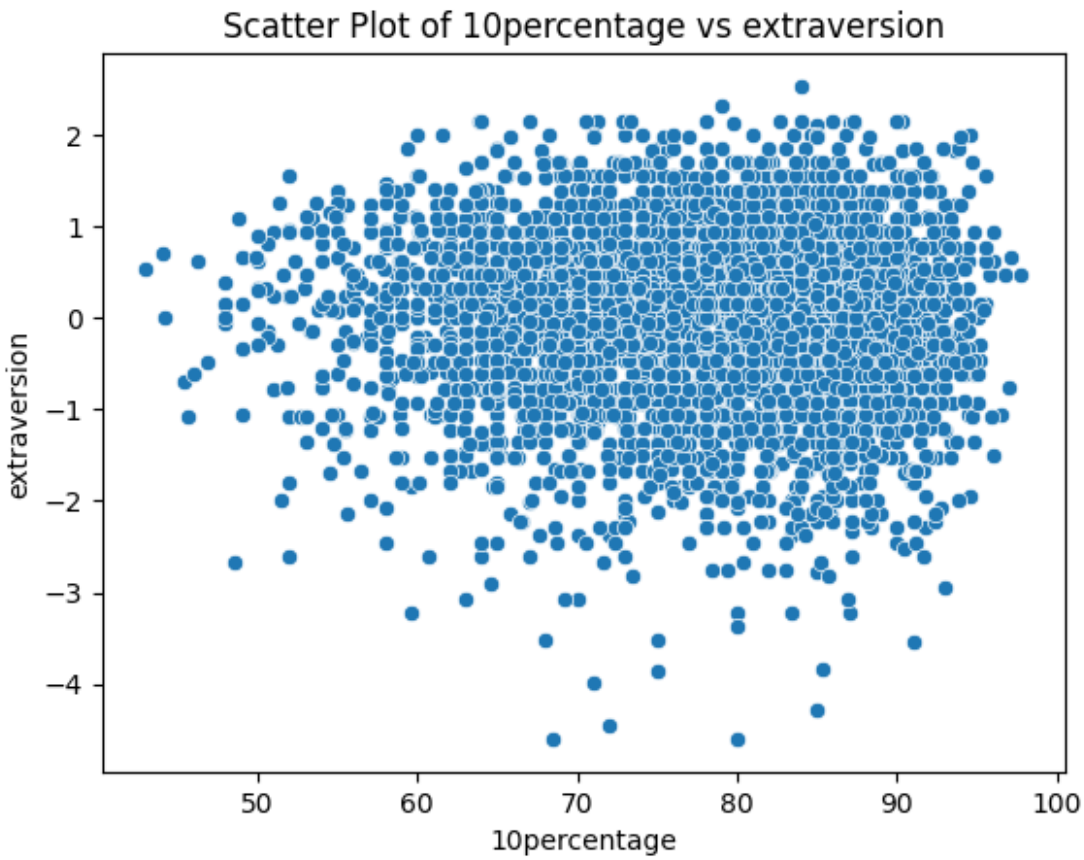

Scatter Plot of 10percentage vs ComputerProgramming
ComputerProgramming
10percentage
800
700
600
500
400
300
200
100
50
60
70
80
90
100

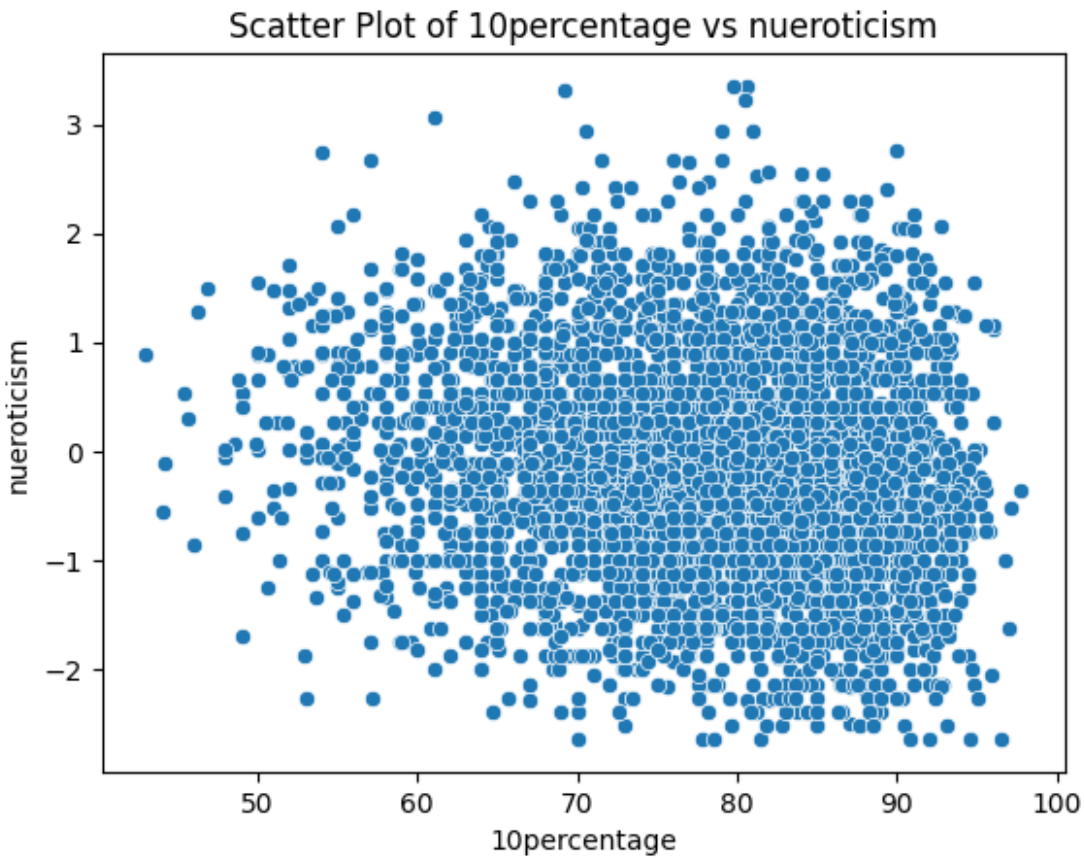

Scatter Plot of 10percentage vs ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200
50
60
70
80
90
100
10percentage

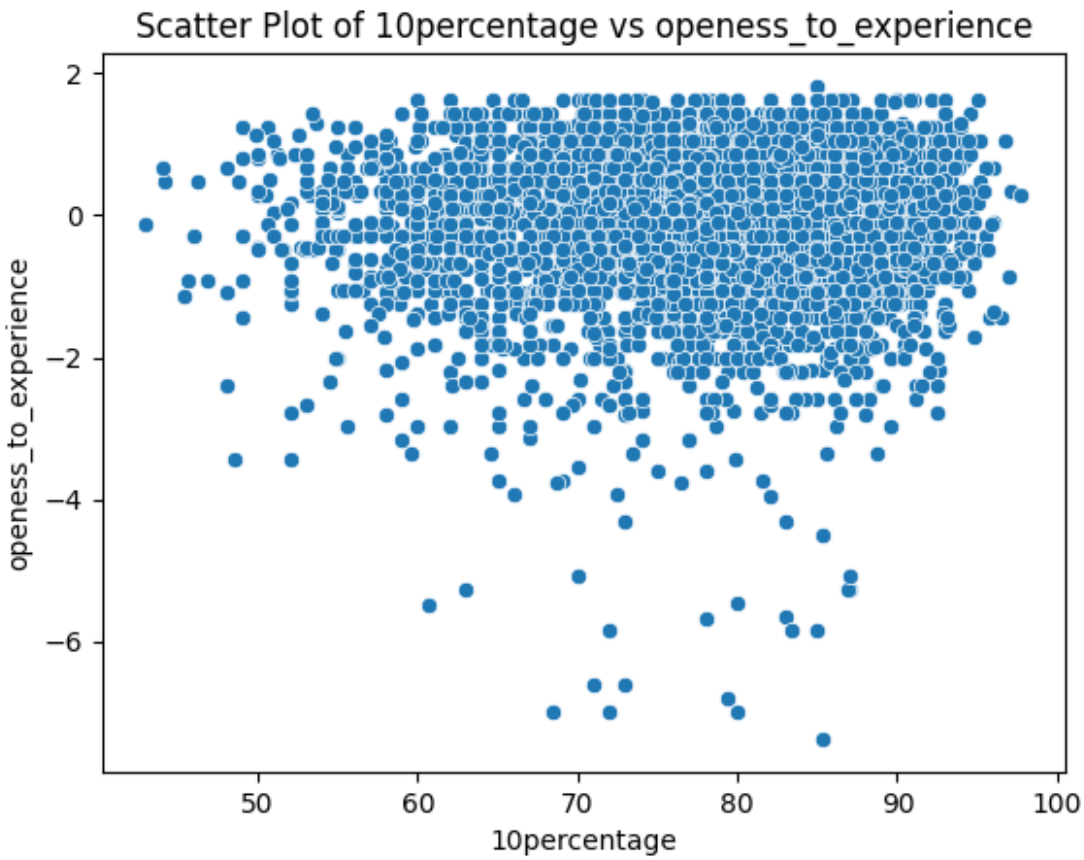

Scatter Plot of 10percentage vs ComputerScience
ComputerScience
700
600
500
400
300
200
50
60
70
80
90
100
10percentage

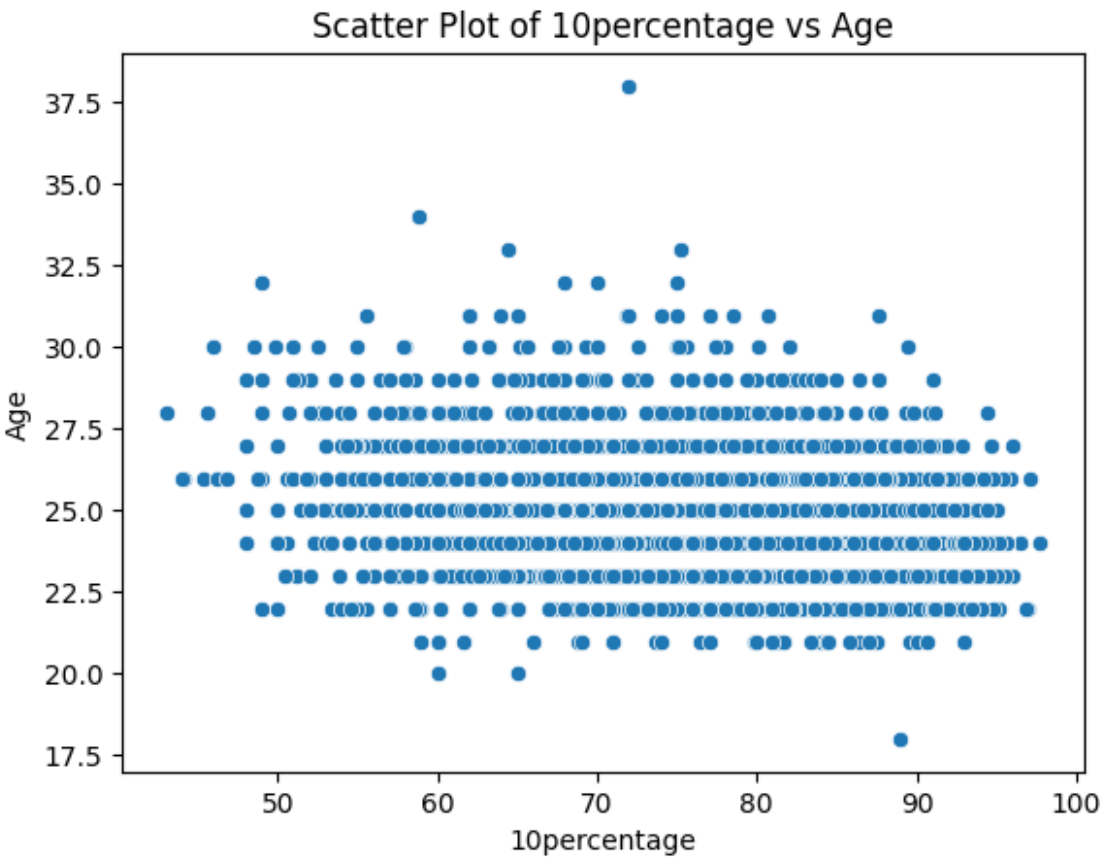

Scatter Plot of 10percentage vs conscientiousness
conscientiousness
10percentage
2
1
0
−1
−2
−3
−4
50
60
70
80
90
100

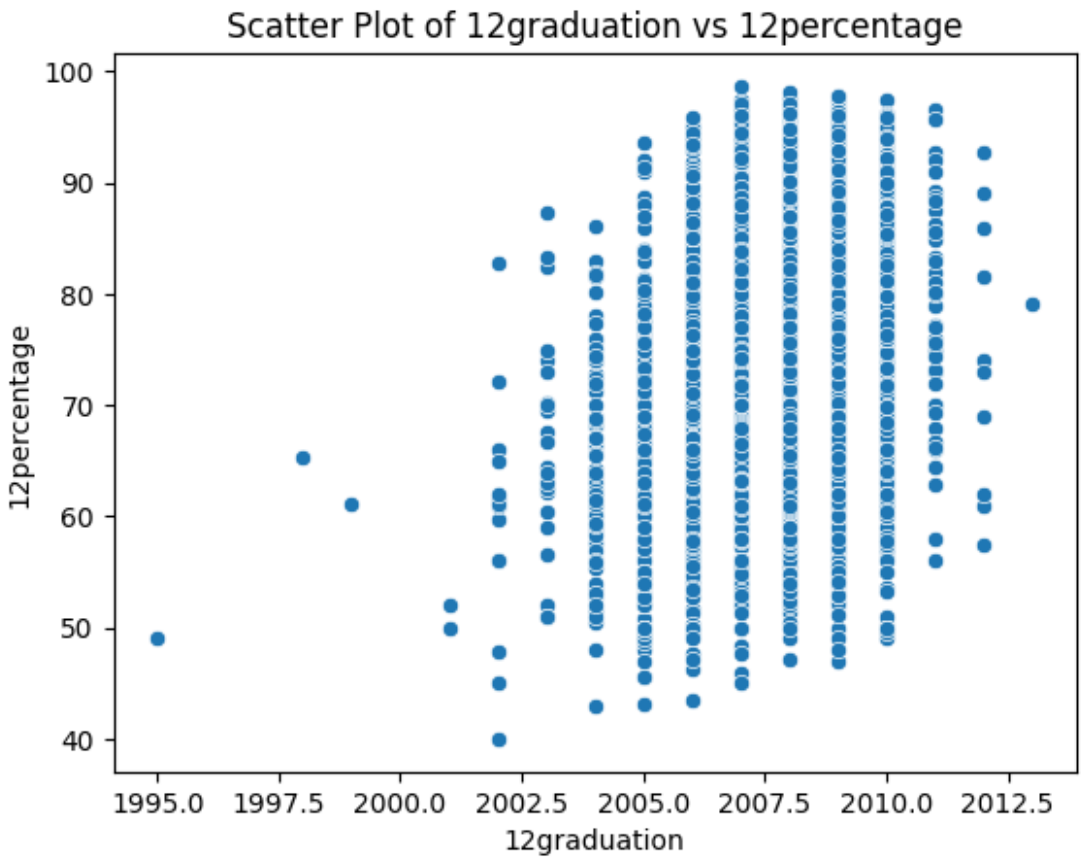

Scatter Plot of 10percentage vs agreeableness
agreeableness
10percentage
2
1
0
−1
−2
−3
−4
−5
−6
50
60
70
80
90
100

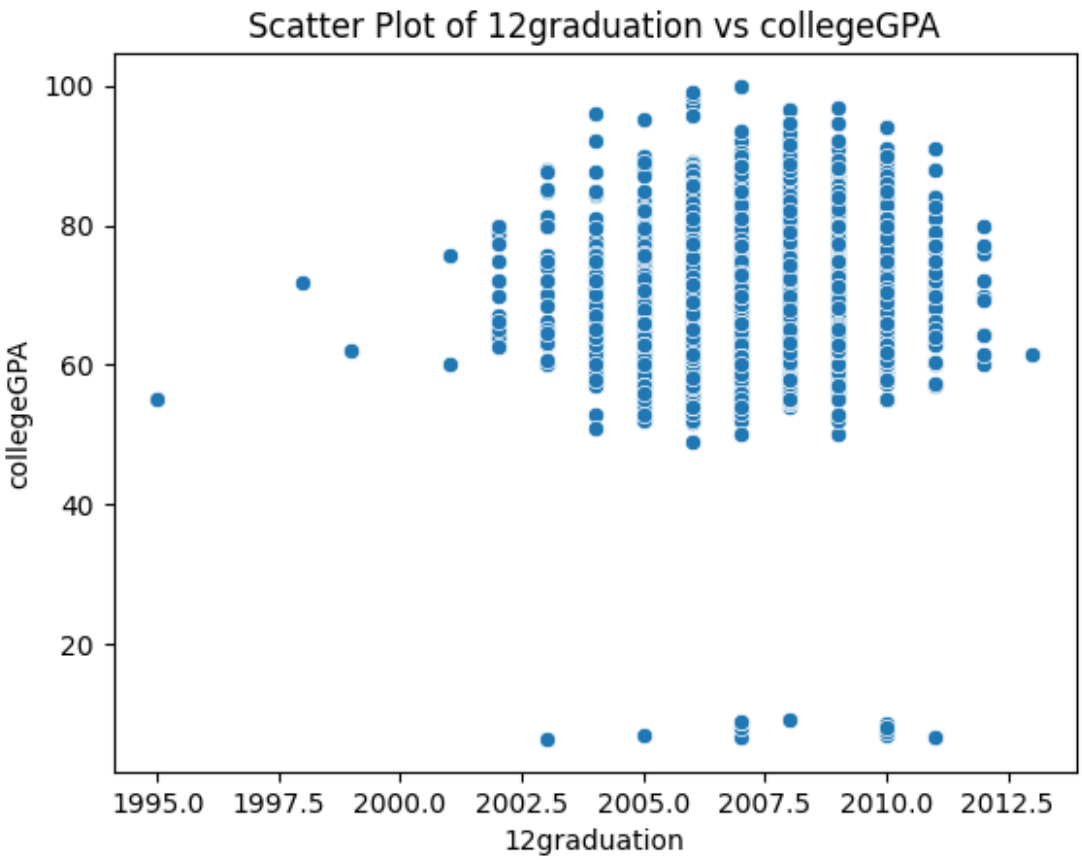

Scatter Plot of 10percentage vs extraversion
extraversion
10percentage
2
1
0
−1
−2
−3
−4
50
60
70
80
90
100

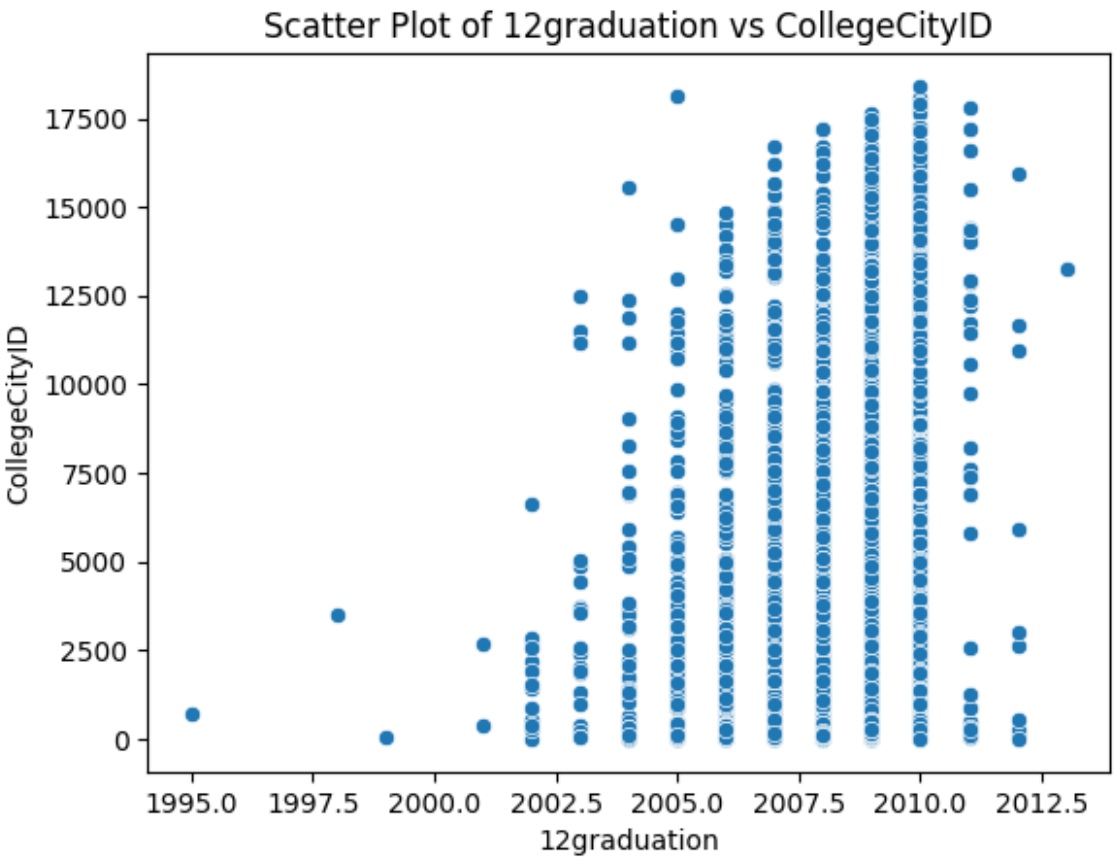

Scatter Plot of 10percentage vs nueroticism
nueroticism
10percentage
3
2
1
0
−1
−2
50
60
70
80
90
100

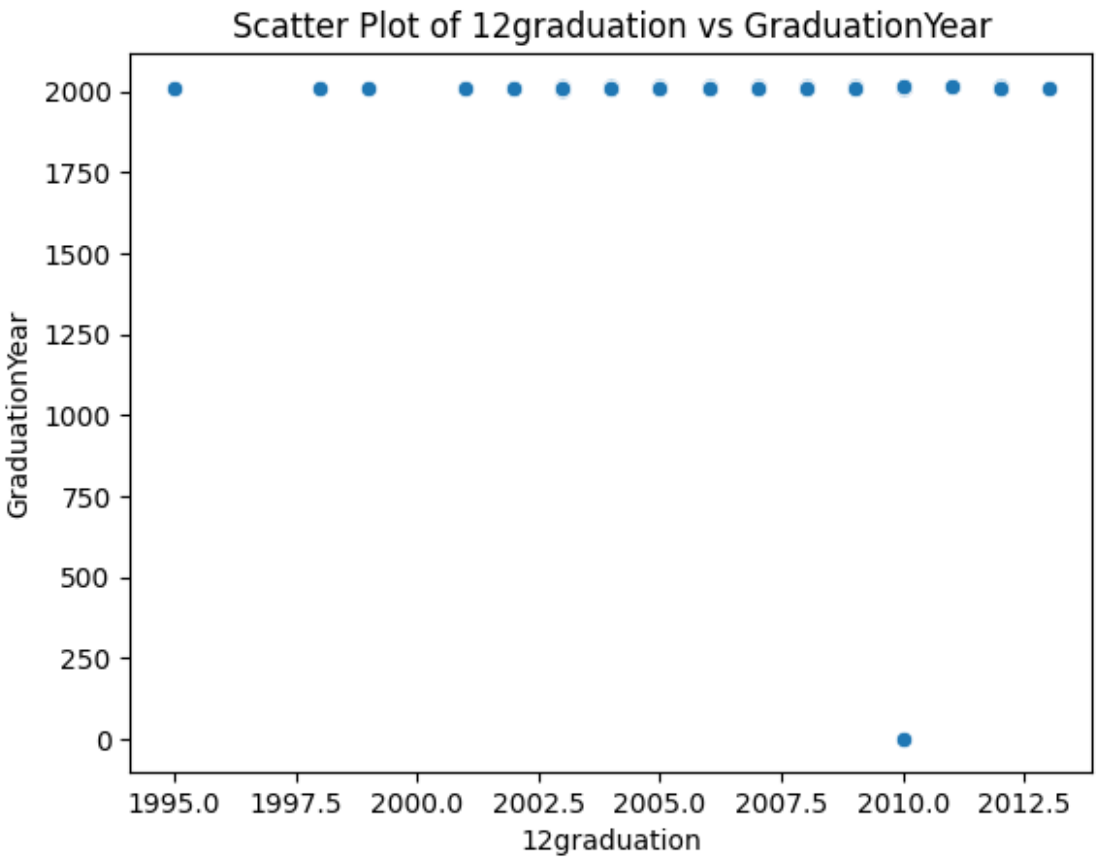

Scatter Plot of 10percentage vs openness_to_experience
openness_to_experience
10percentage
2
0
−2
−4
−6
50
60
70
80
90
100

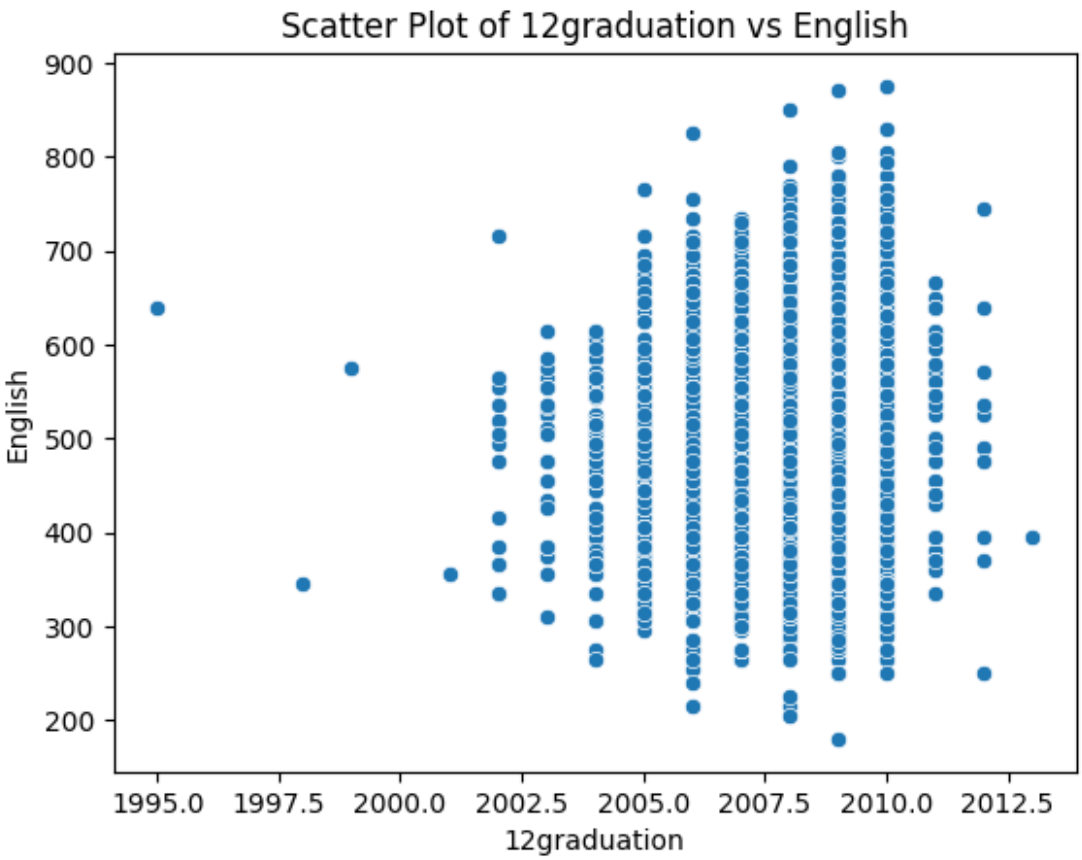

Scatter Plot of 10percentage vs Age
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
50
60
70
80
90
100
10percentage

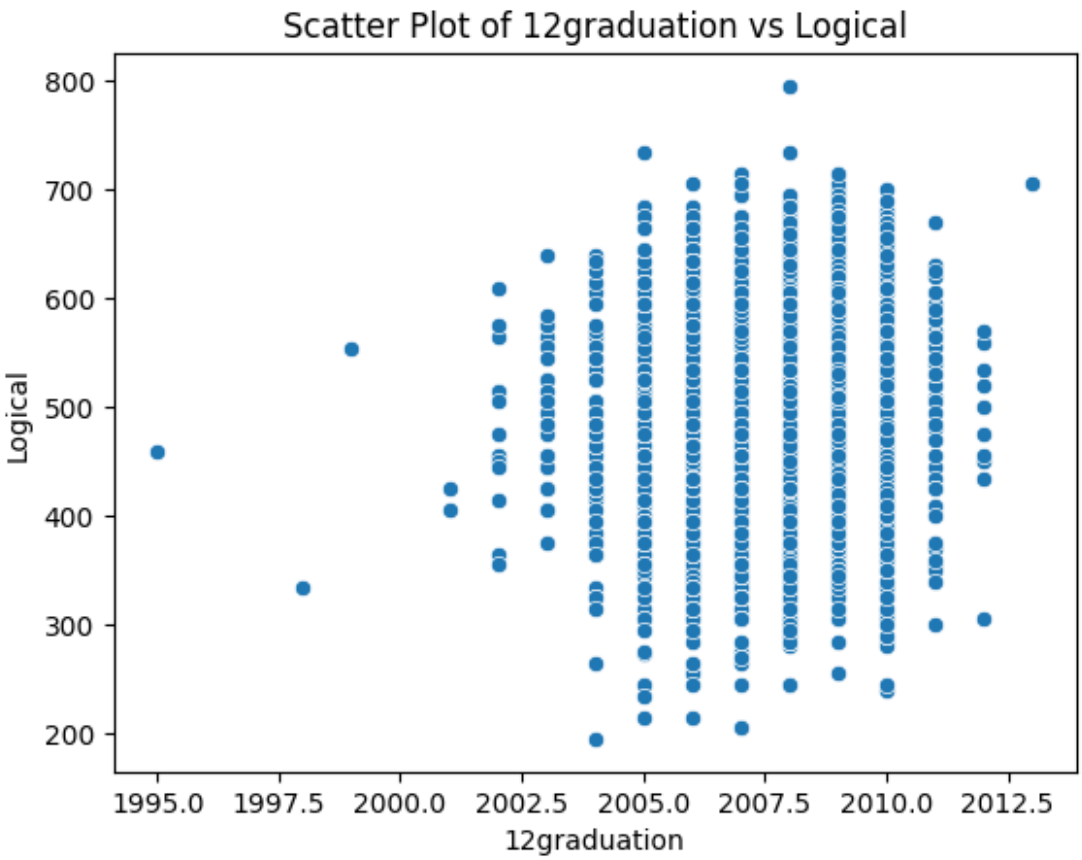

Scatter Plot of 12graduation vs 12percentage
12percentage
12graduation
100
90
80
70
60
50
40
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5

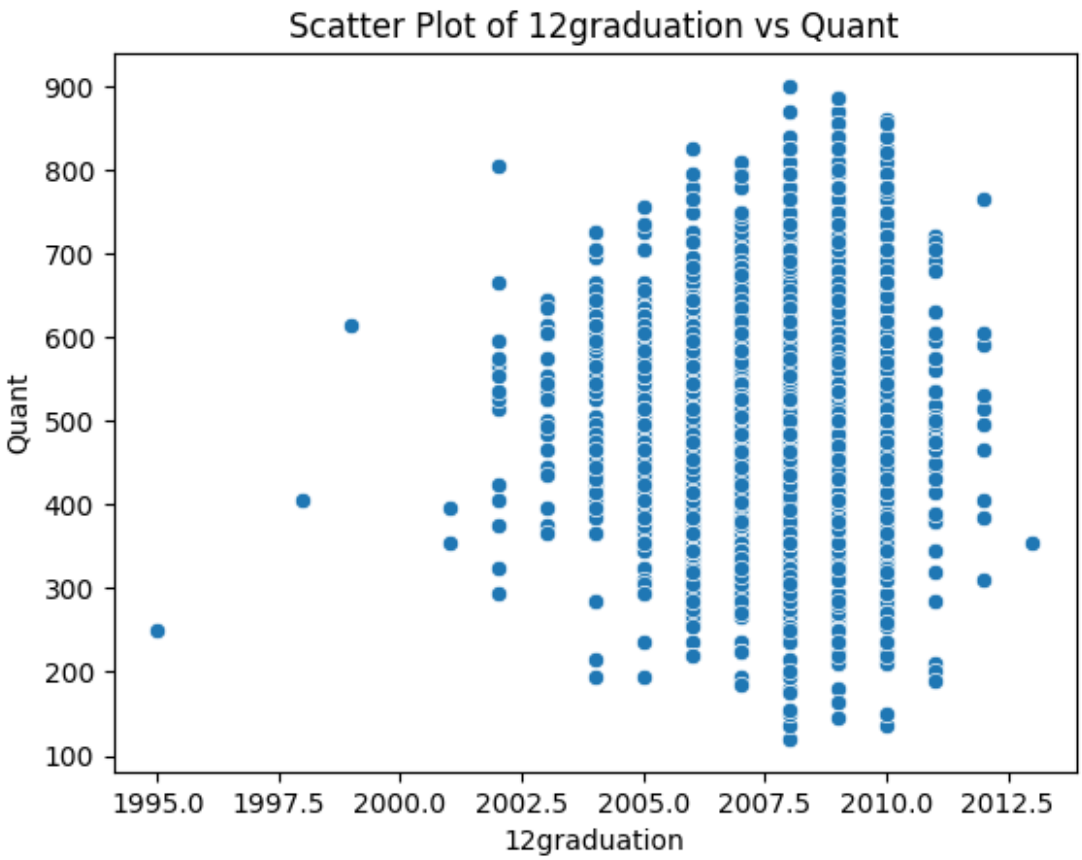

Scatter Plot of 12graduation vs collegeGPA
collegeGPA
100
80
60
40
20
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

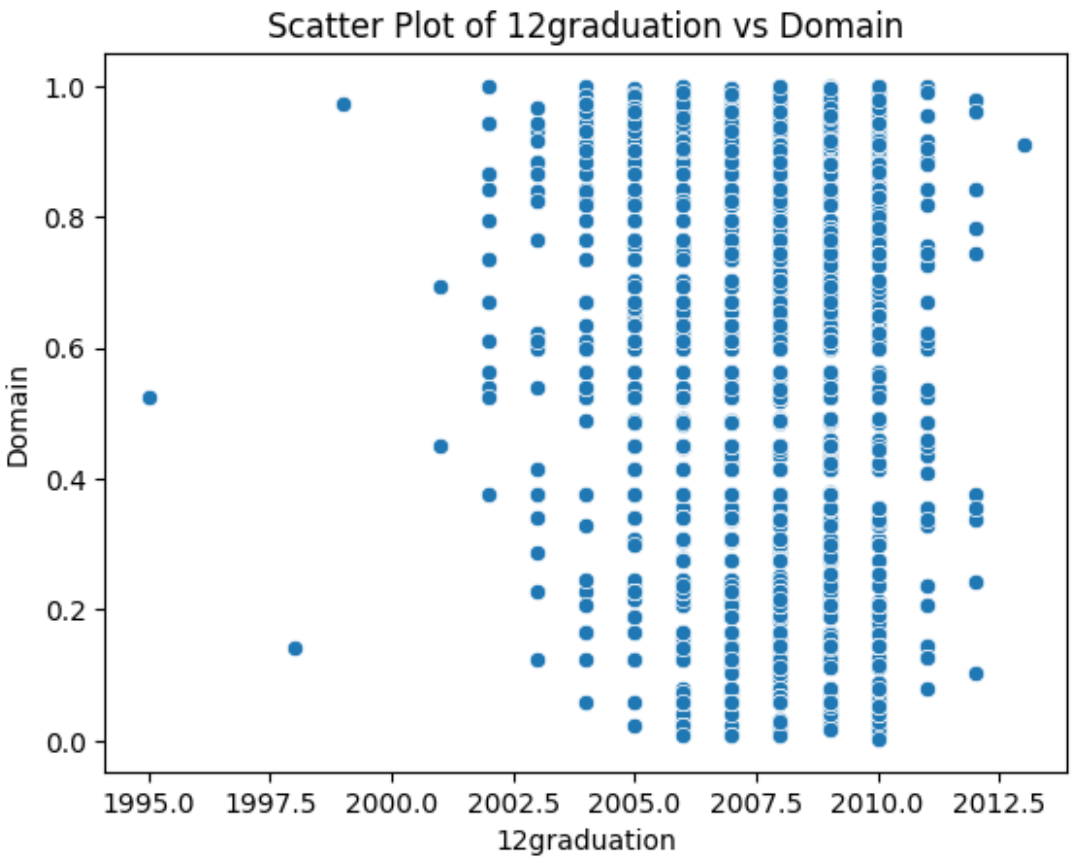

Scatter Plot of 12graduation vs CollegeCityID
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

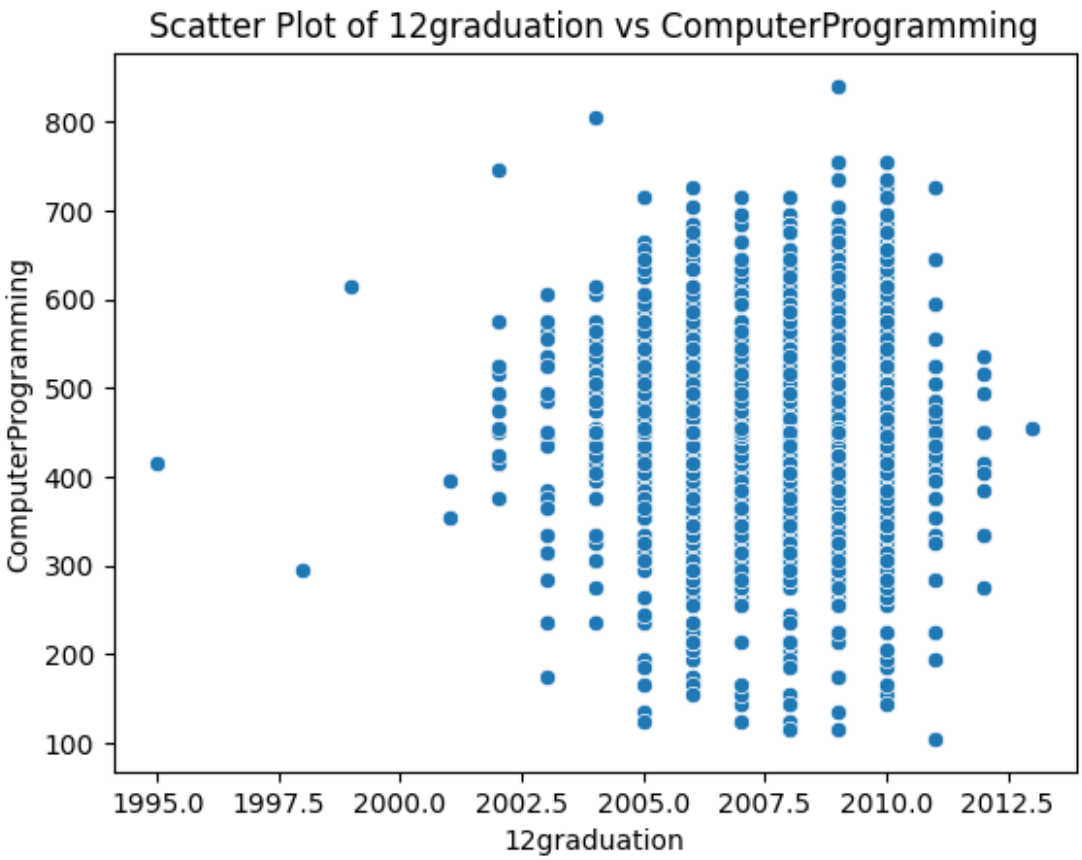

Scatter Plot of 12graduation vs GraduationYear
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

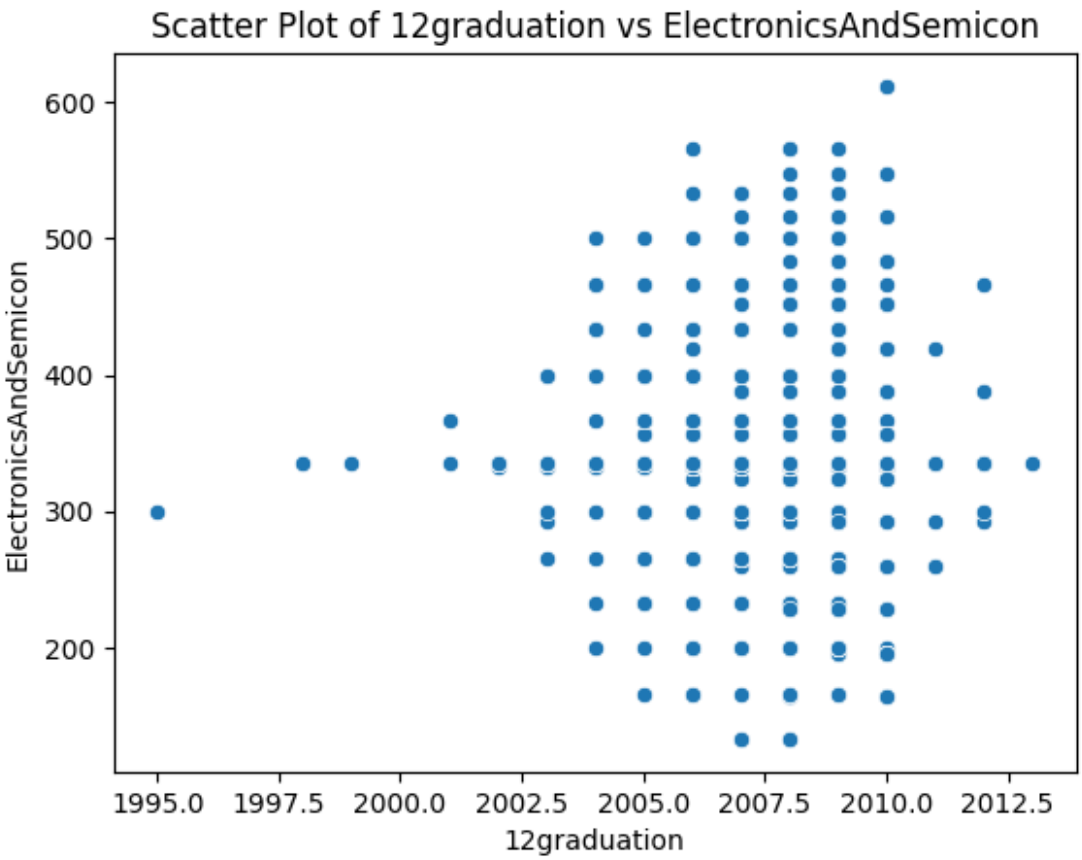

Scatter Plot of 12graduation vs English
English
12graduation
900
800
700
600
500
400
300
200
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5

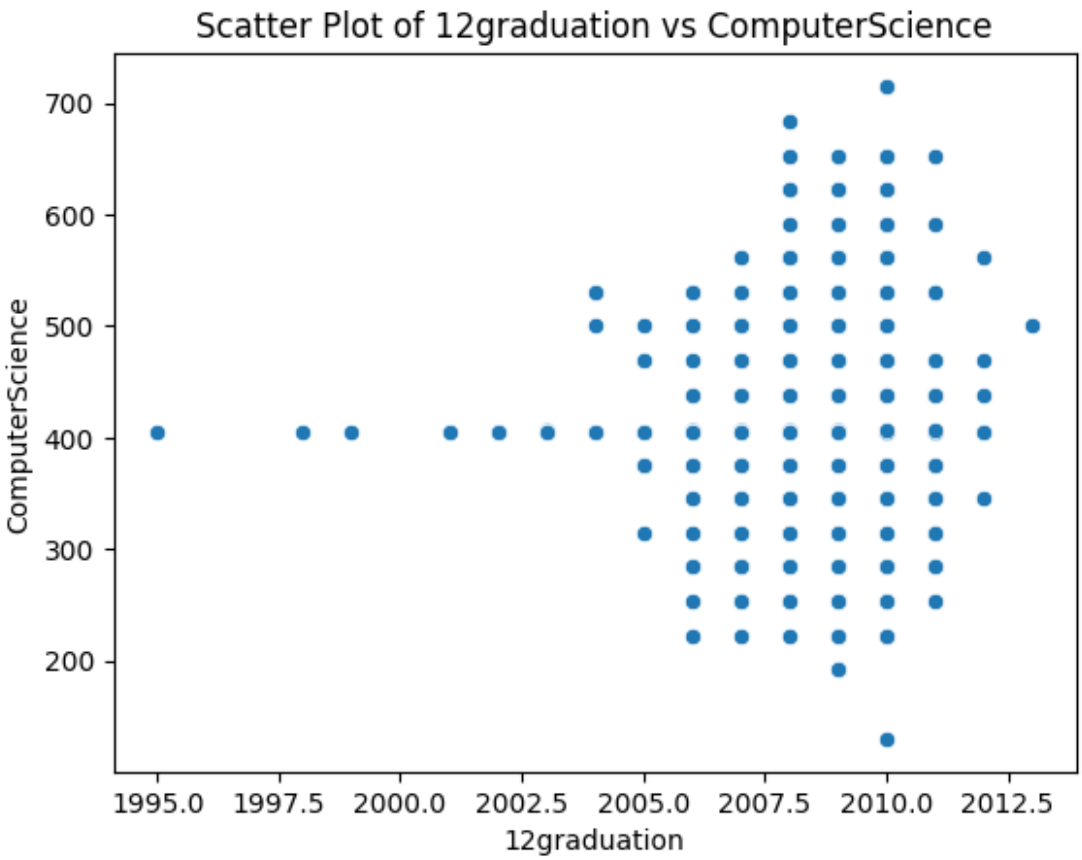

Scatter Plot of 12graduation vs Logical
Logical
800
700
600
500
400
300
200
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

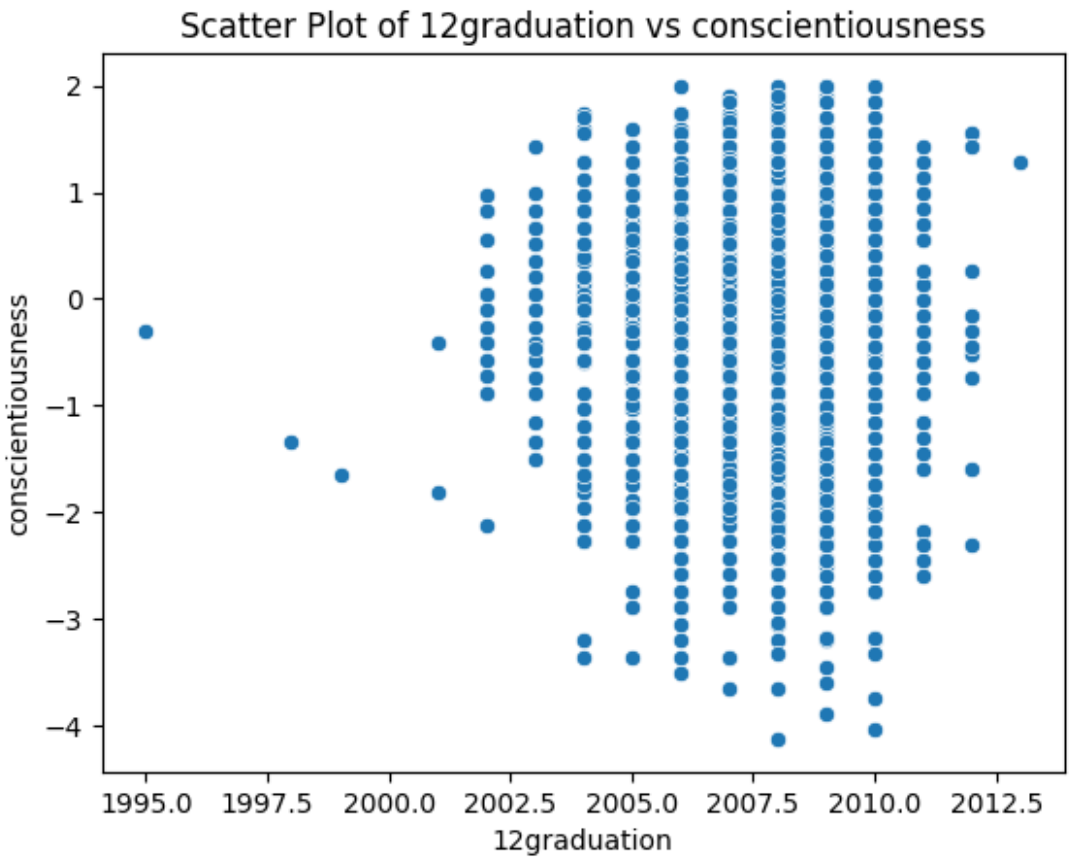

Scatter Plot of 12graduation vs Quant
Quant
900
800
700
600
500
400
300
200
100
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

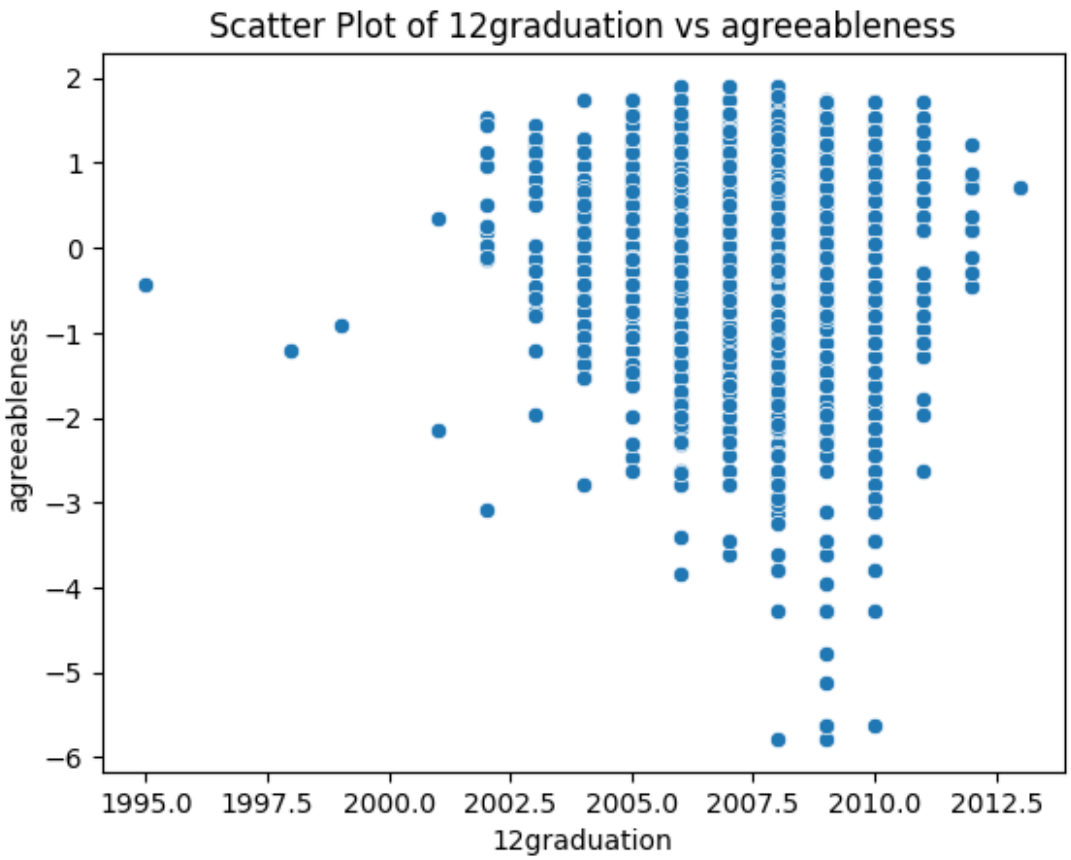

Scatter Plot of 12graduation vs Domain
Domain
1.0
0.8
0.6
0.4
0.2
0.0
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

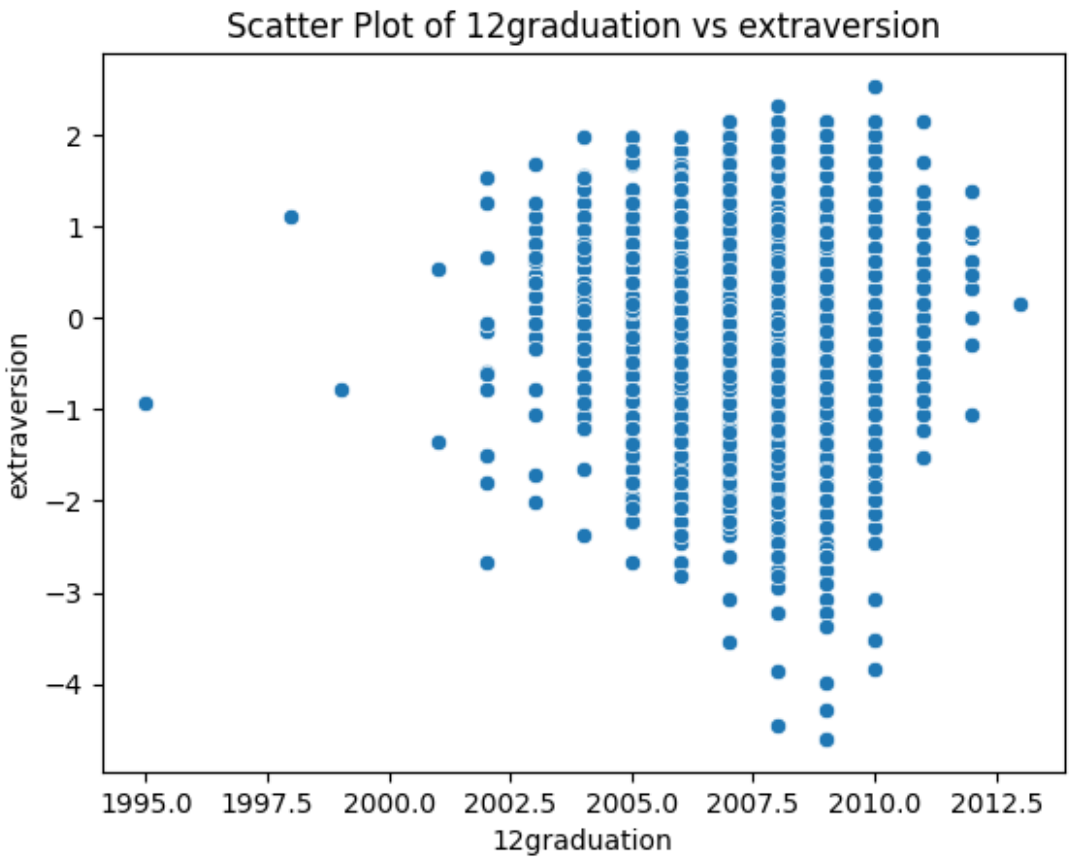

Scatter Plot of 12graduation vs ComputerProgramming
ComputerProgramming
800
700
600
500
400
300
200
100
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

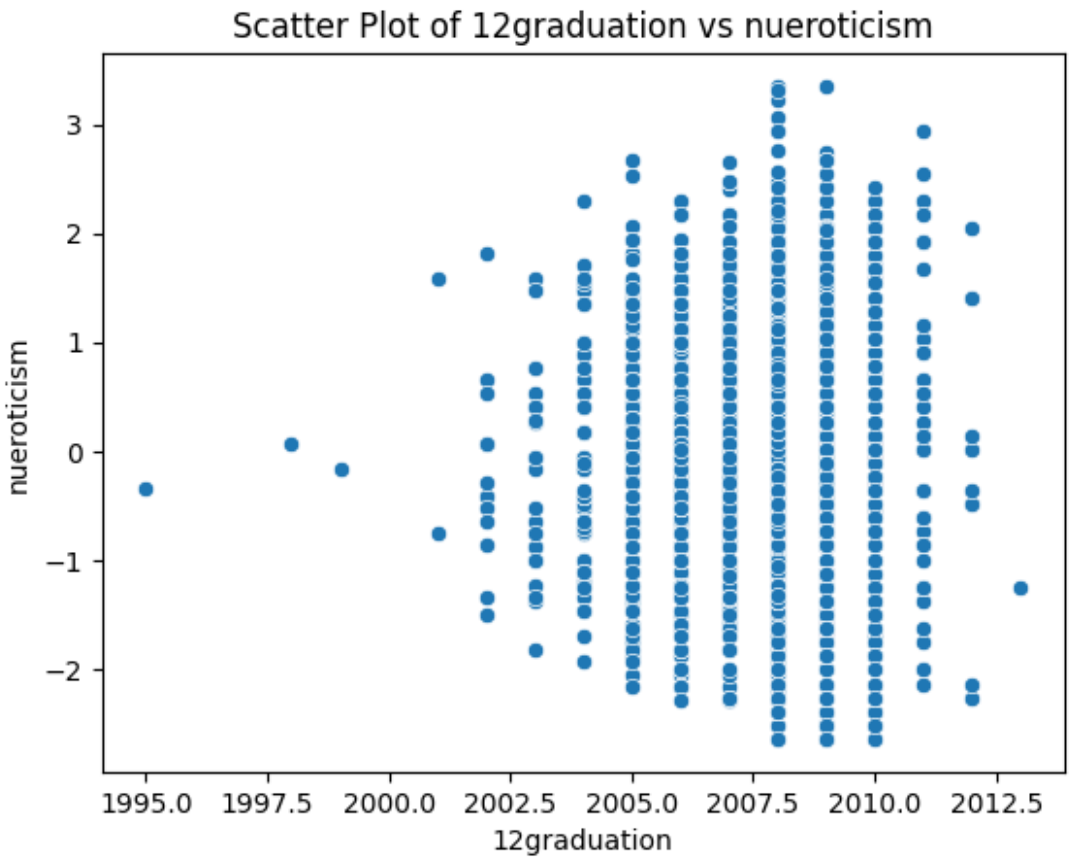

Scatter Plot of 12graduation vs ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

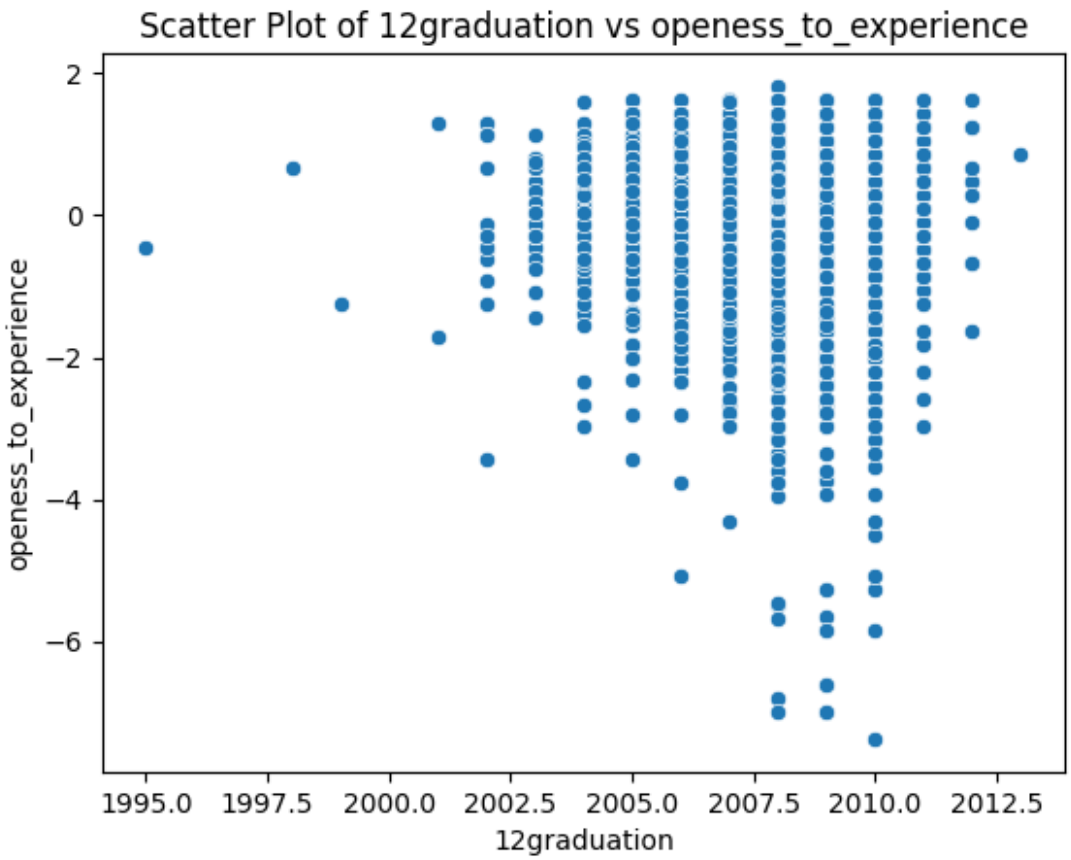

Scatter Plot of 12graduation vs ComputerScience
ComputerScience
700
600
500
400
300
200
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

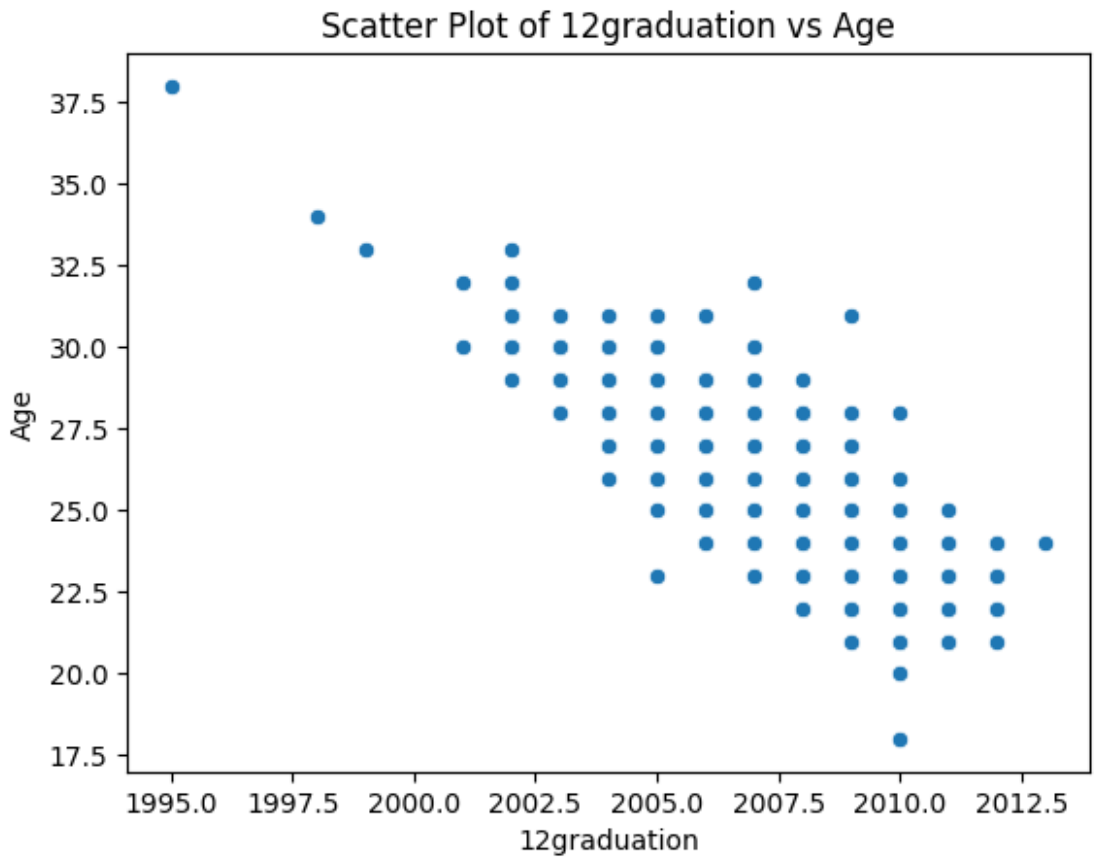

Scatter Plot of 12graduation vs conscientiousness
conscientiousness
12graduation
2
1
0
−1
−2
−3
−4
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5

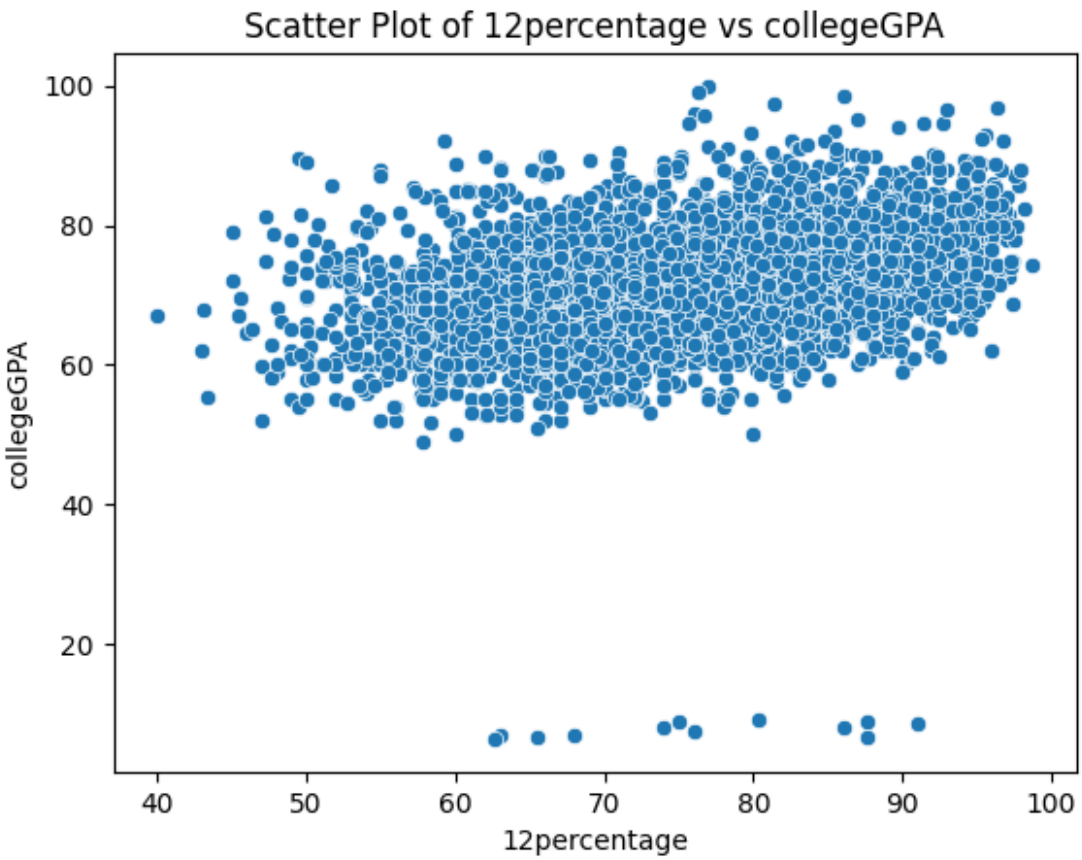

Scatter Plot of 12graduation vs agreeableness
agreeableness
12graduation
2
1
0
−1
−2
−3
−4
−5
−6
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5

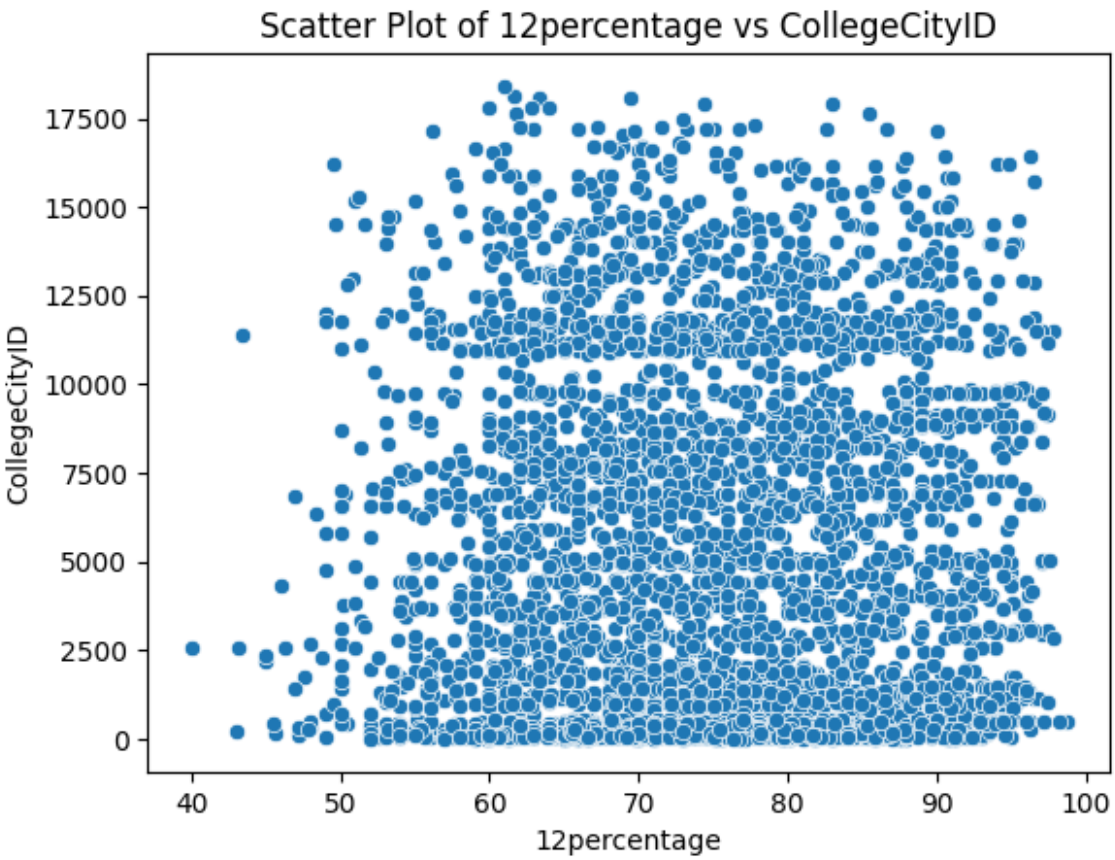

Scatter Plot of 12graduation vs extraversion
extraversion
12graduation
2
1
0
−1
−2
−3
−4
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5

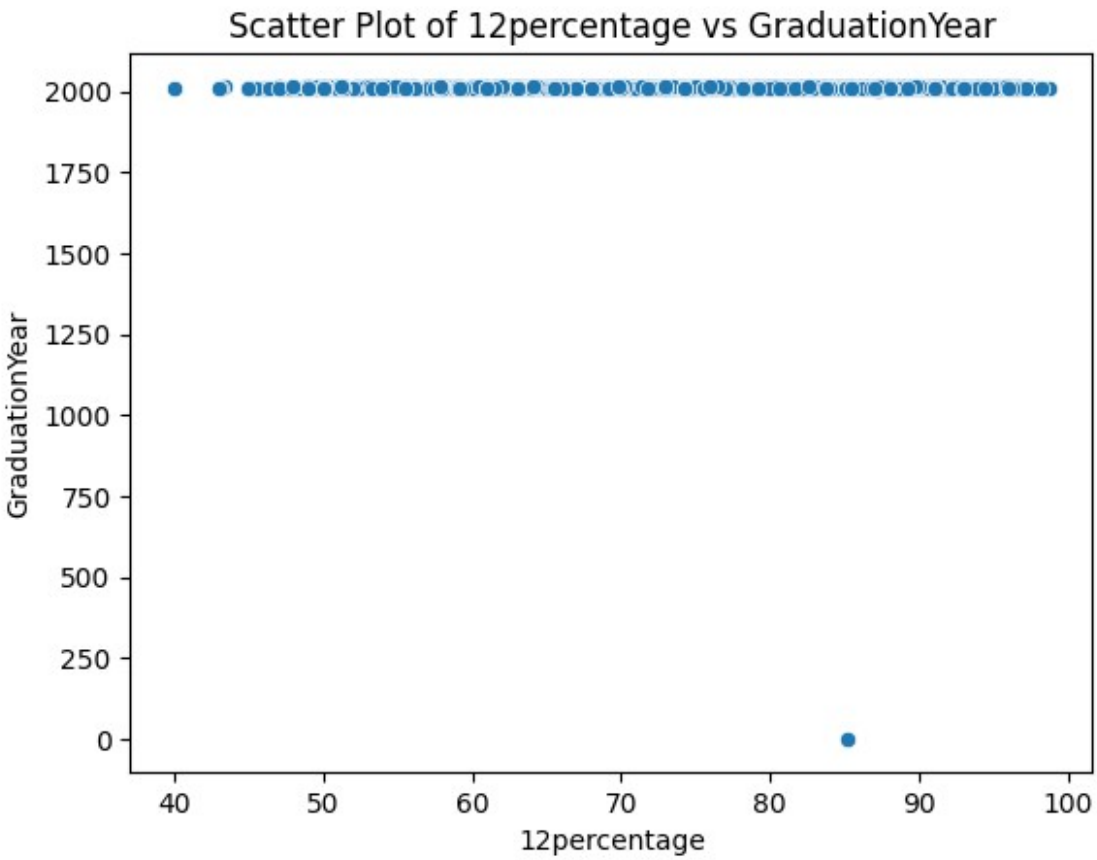

Scatter Plot of 12graduation vs nueroticism
nueroticism
3
2
1
0
−1
−2
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

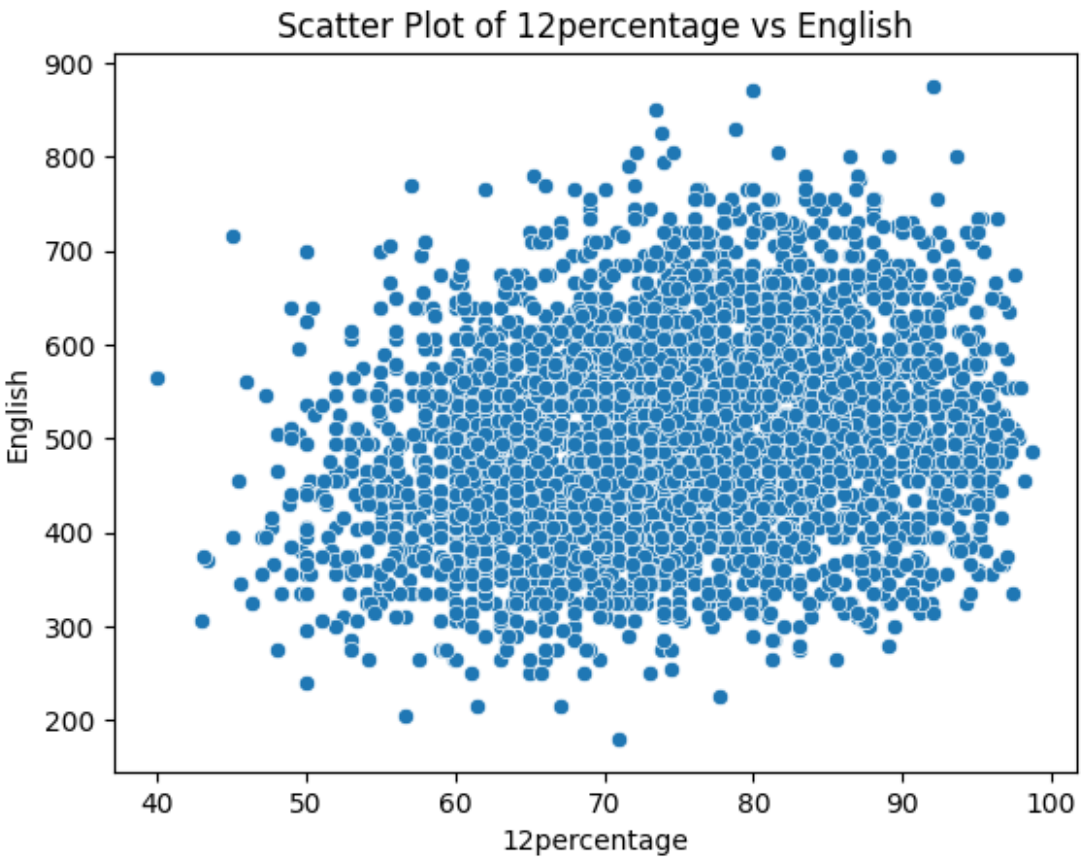

Scatter Plot of 12graduation vs openess_to_experience
openess_to_experience
2
0
−2
−4
−6
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

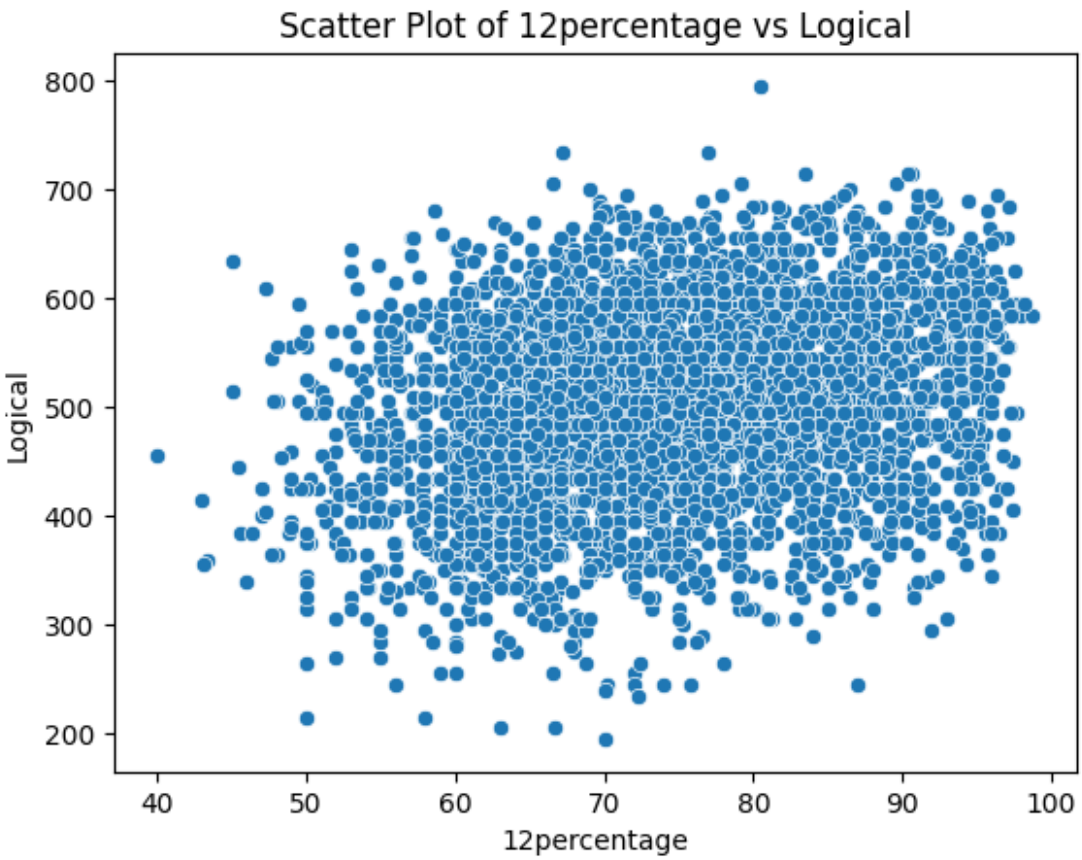

Scatter Plot of 12graduation vs Age
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
1995.0
1997.5
2000.0
2002.5
2005.0
2007.5
2010.0
2012.5
12graduation

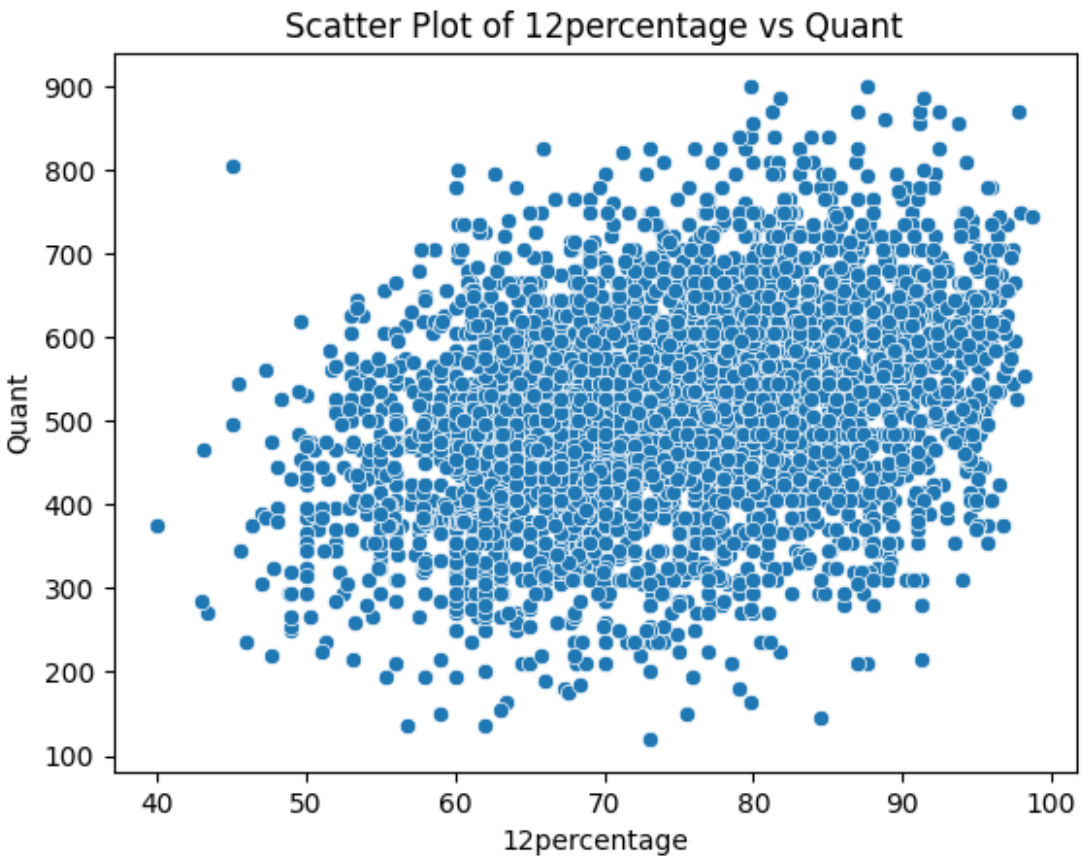

Scatter Plot of 12percentage vs collegeGPA
collegeGPA
12percentage
100
80
60
40
20
40
50
60
70
80
90
100

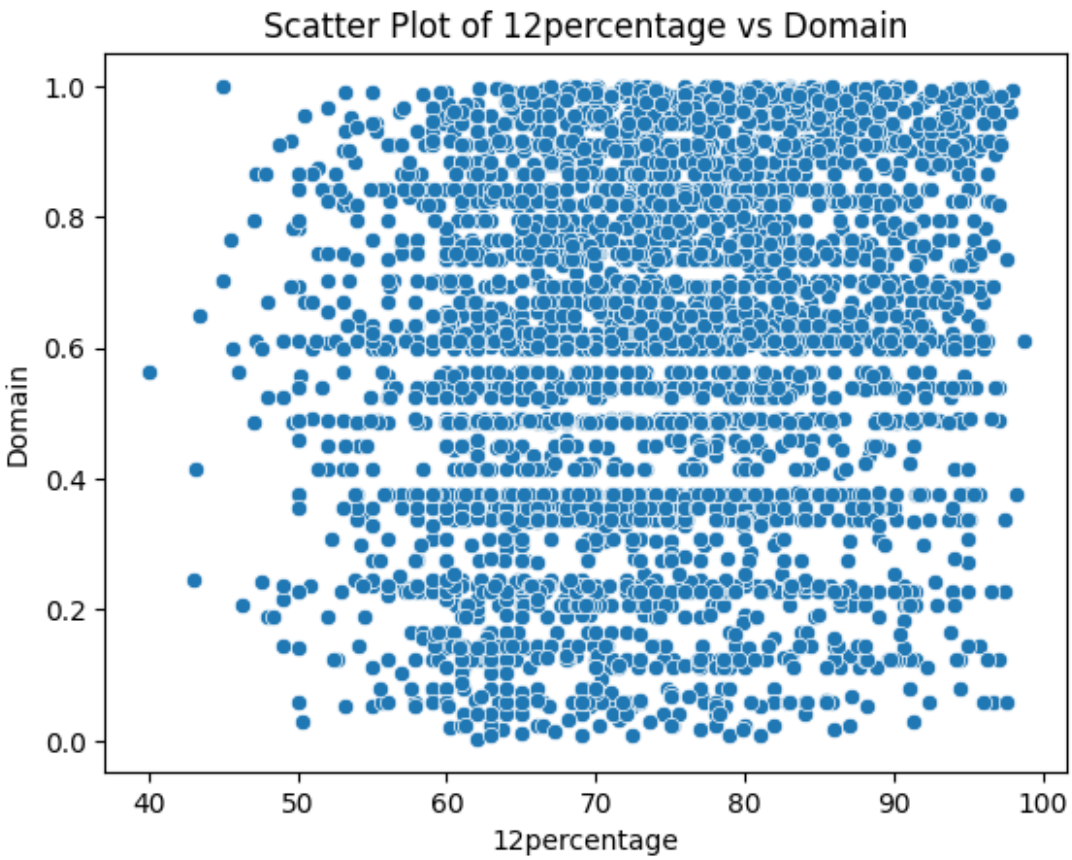

Scatter Plot of 12percentage vs CollegeCityID
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
40
50
60
70
80
90
100
12percentage

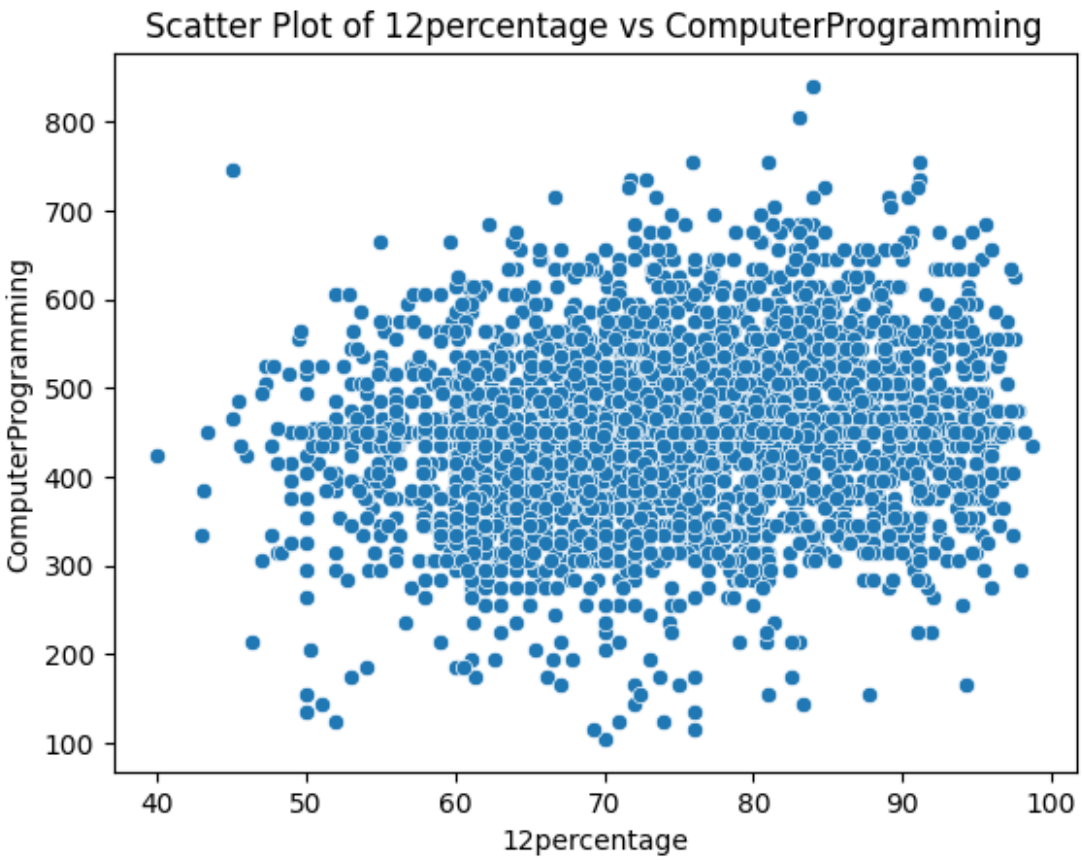

Scatter Plot of 12percentage vs GraduationYear
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
40
50
60
70
80
90
100
12percentage

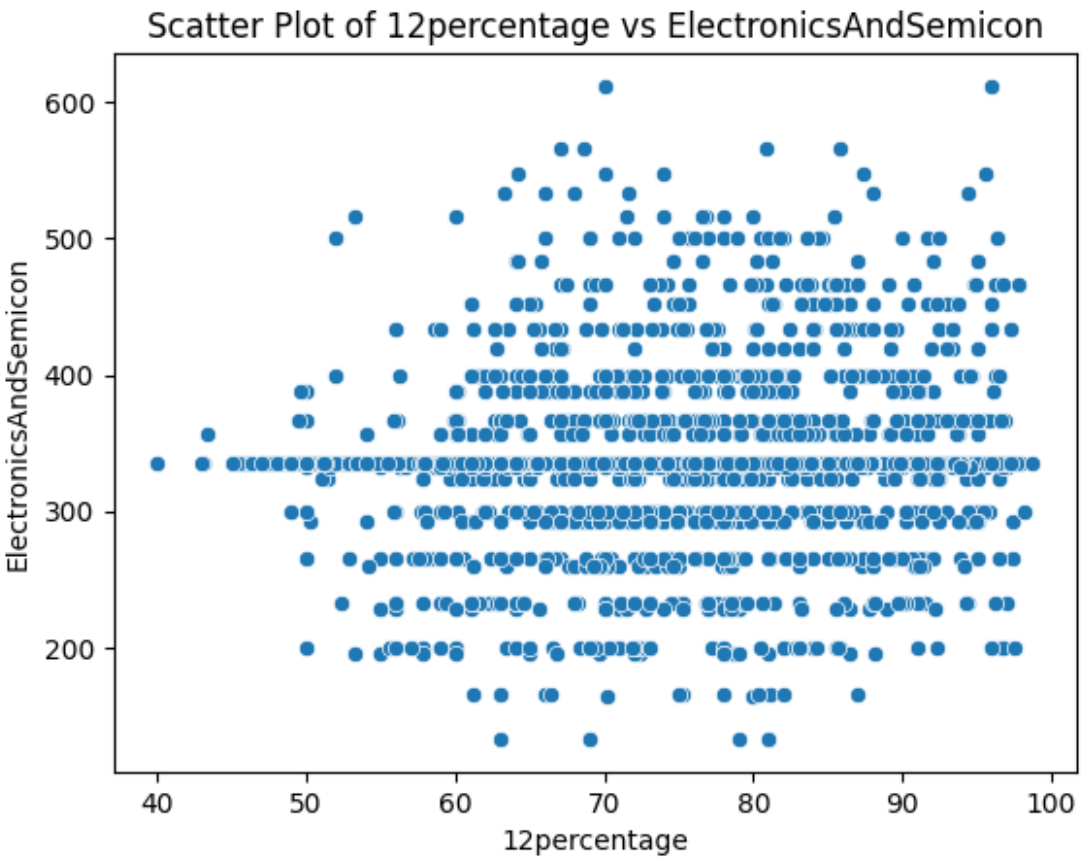

Scatter Plot of 12percentage vs English
English
12percentage
900
800
700
600
500
400
300
200
40
50
60
70
80
90
100

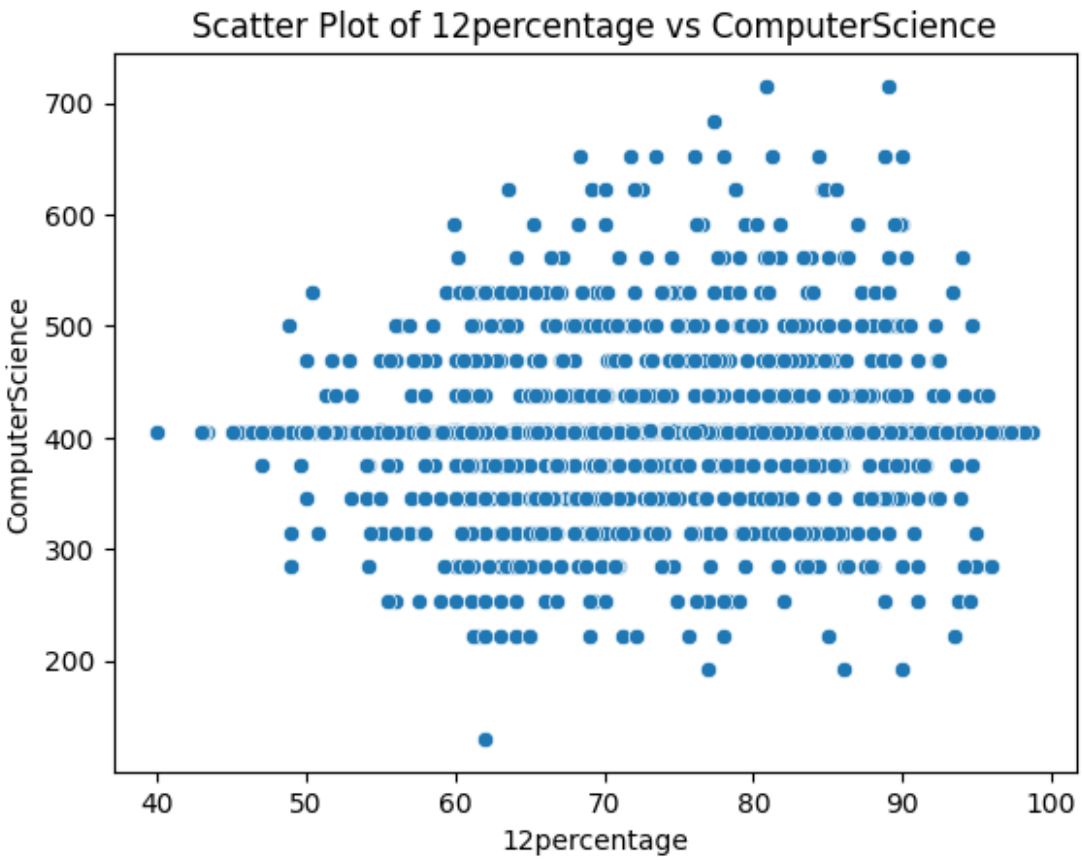

Scatter Plot of 12percentage vs Logical
Logical
12percentage
800
700
600
500
400
300
200
40
50
60
70
80
90
100

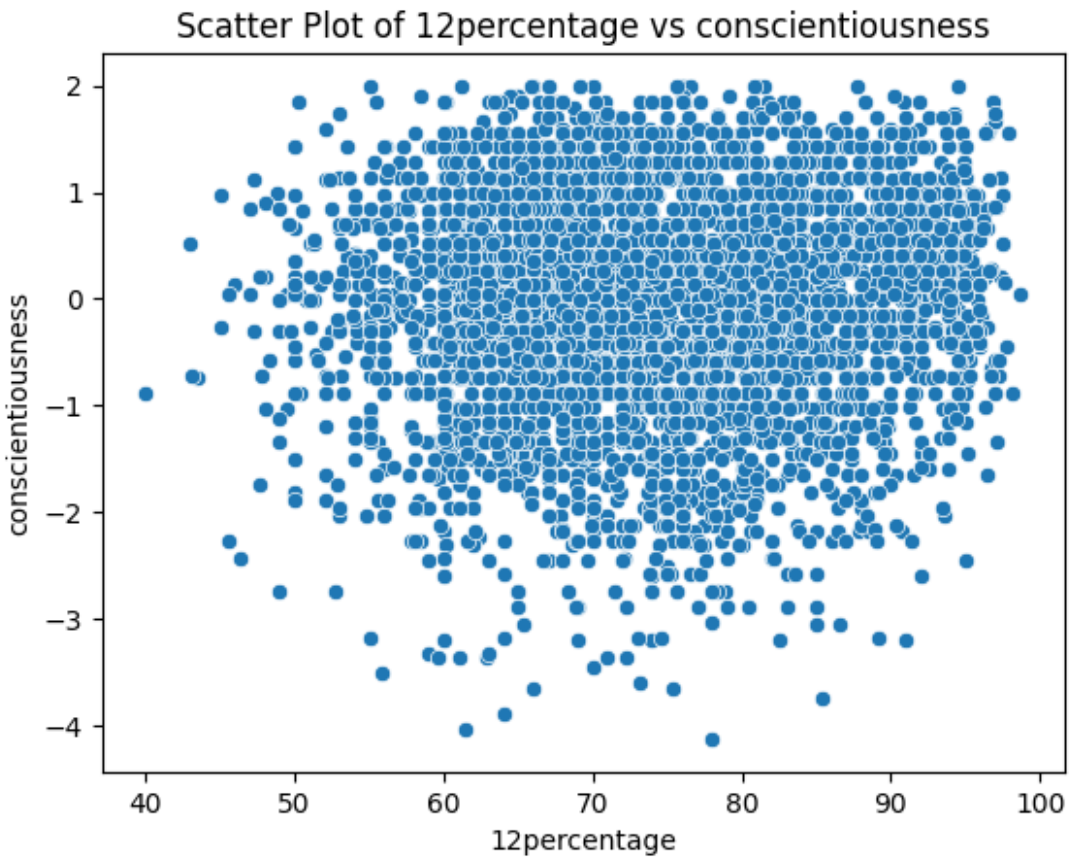

Scatter Plot of 12percentage vs Quant
Quant
12percentage
900
800
700
600
500
400
300
200
100
40
50
60
70
80
90
100

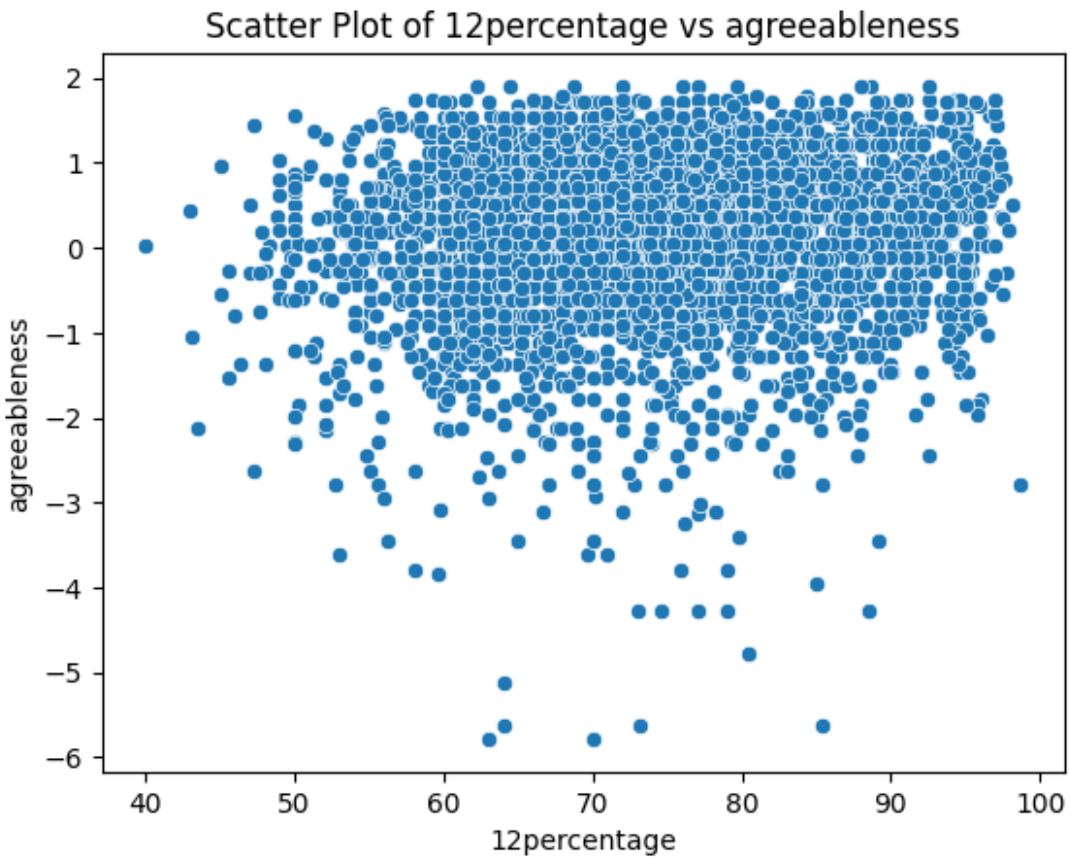

Scatter Plot of 12percentage vs Domain
Domain
12percentage
1.0
0.8
0.6
0.4
0.2
0.0
40
50
60
70
80
90
100

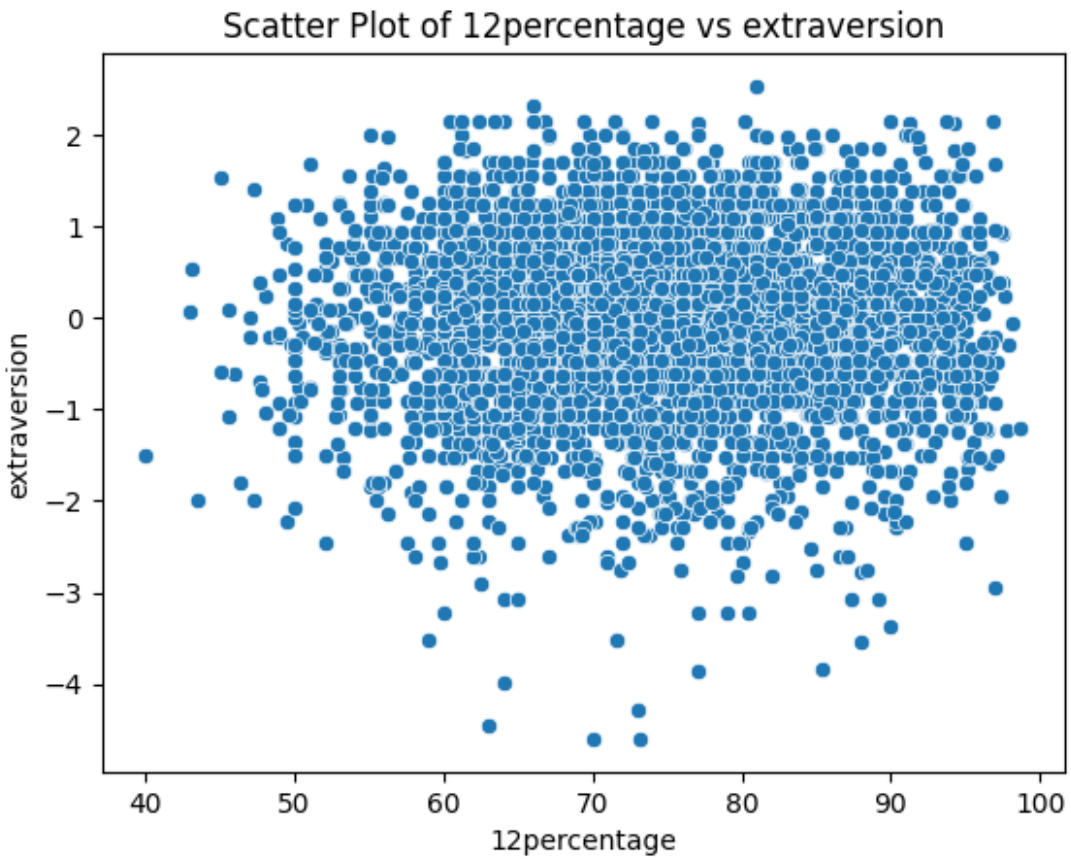

Scatter Plot of 12percentage vs ComputerProgramming
ComputerProgramming
12percentage
800
700
600
500
400
300
200
100
40
50
60
70
80
90
100

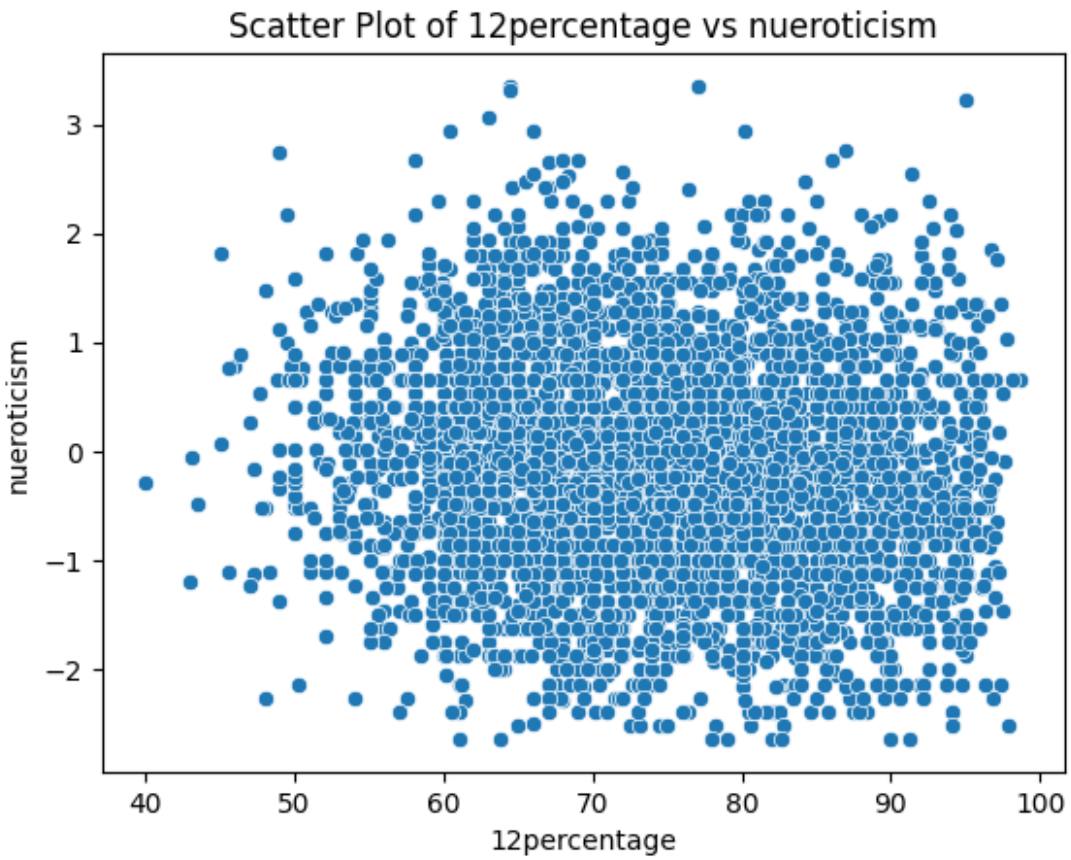

Scatter Plot of 12percentage vs ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200
40
50
60
70
80
90
100
12percentage

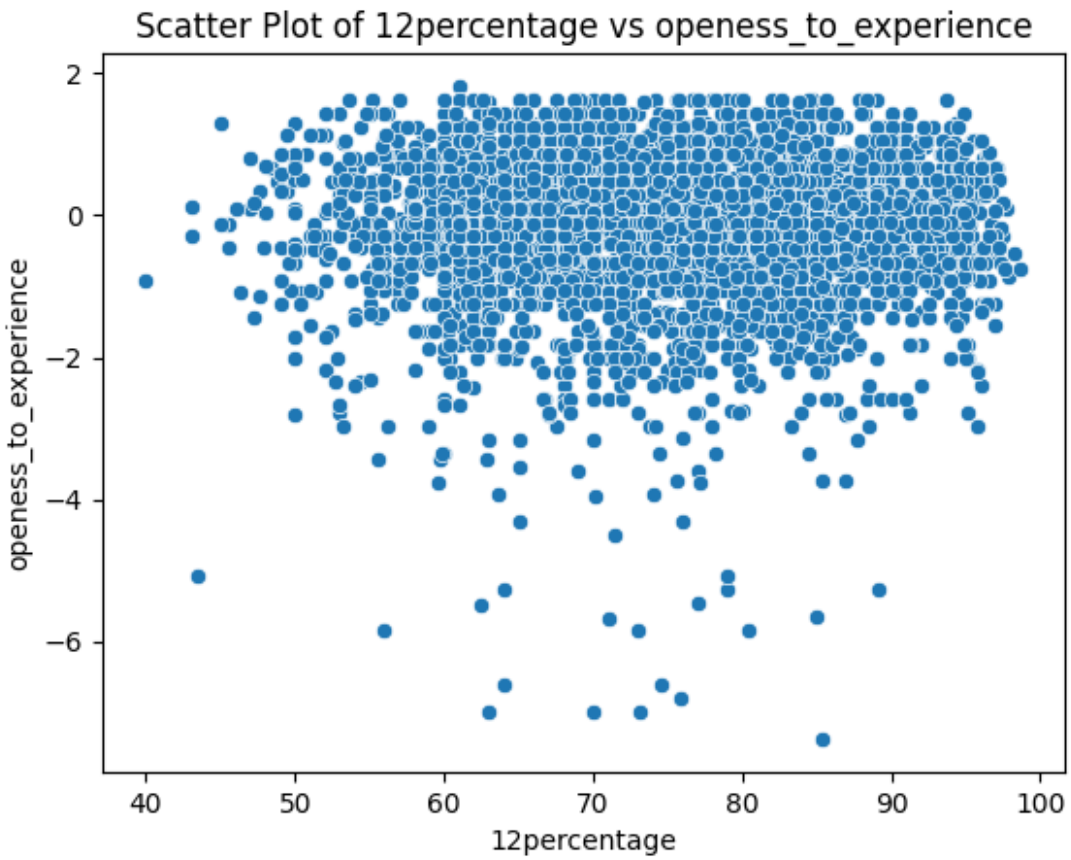

Scatter Plot of 12percentage vs ComputerScience
ComputerScience
700
600
500
400
300
200
40
50
60
70
80
90
100
12percentage

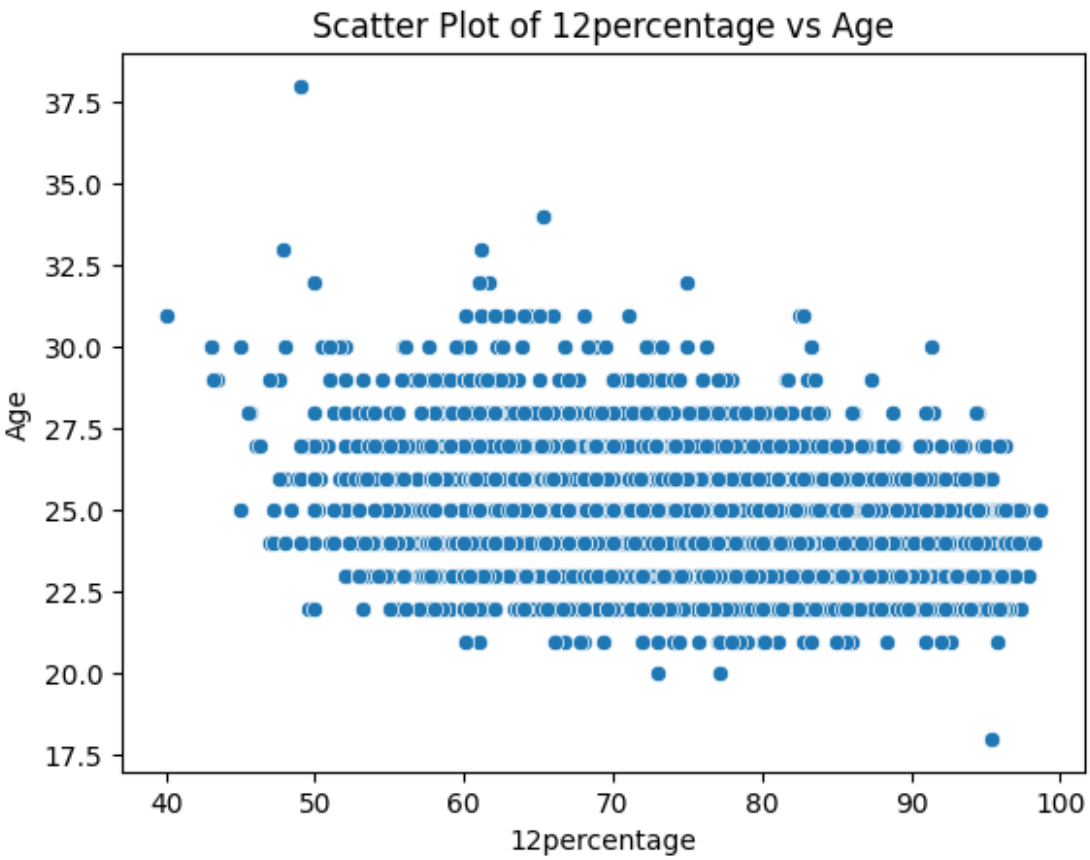

Scatter Plot of 12percentage vs conscientiousness
conscientiousness
12percentage
2
1
0
−1
−2
−3
−4
40
50
60
70
80
90
100

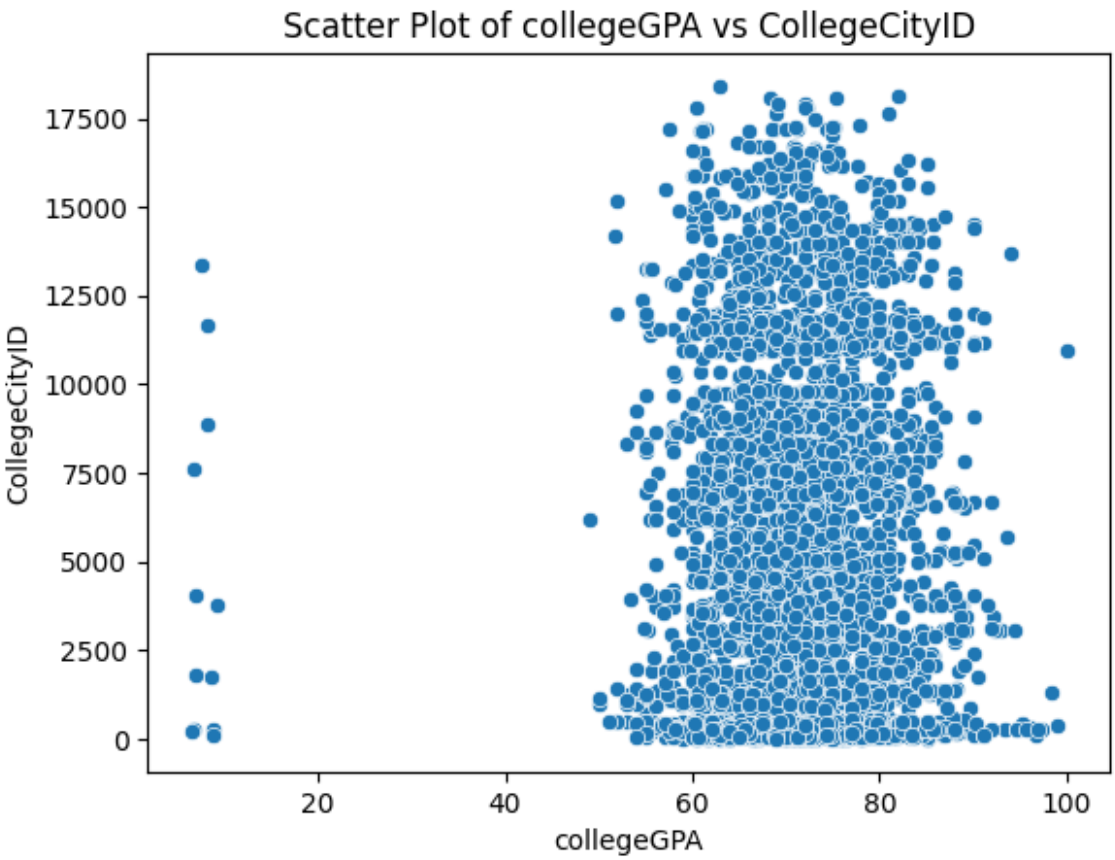

Scatter Plot of 12percentage vs agreeableness
agreeableness
12percentage
2
1
0
−1
−2
−3
−4
−5
−6
40
50
60
70
80
90
100

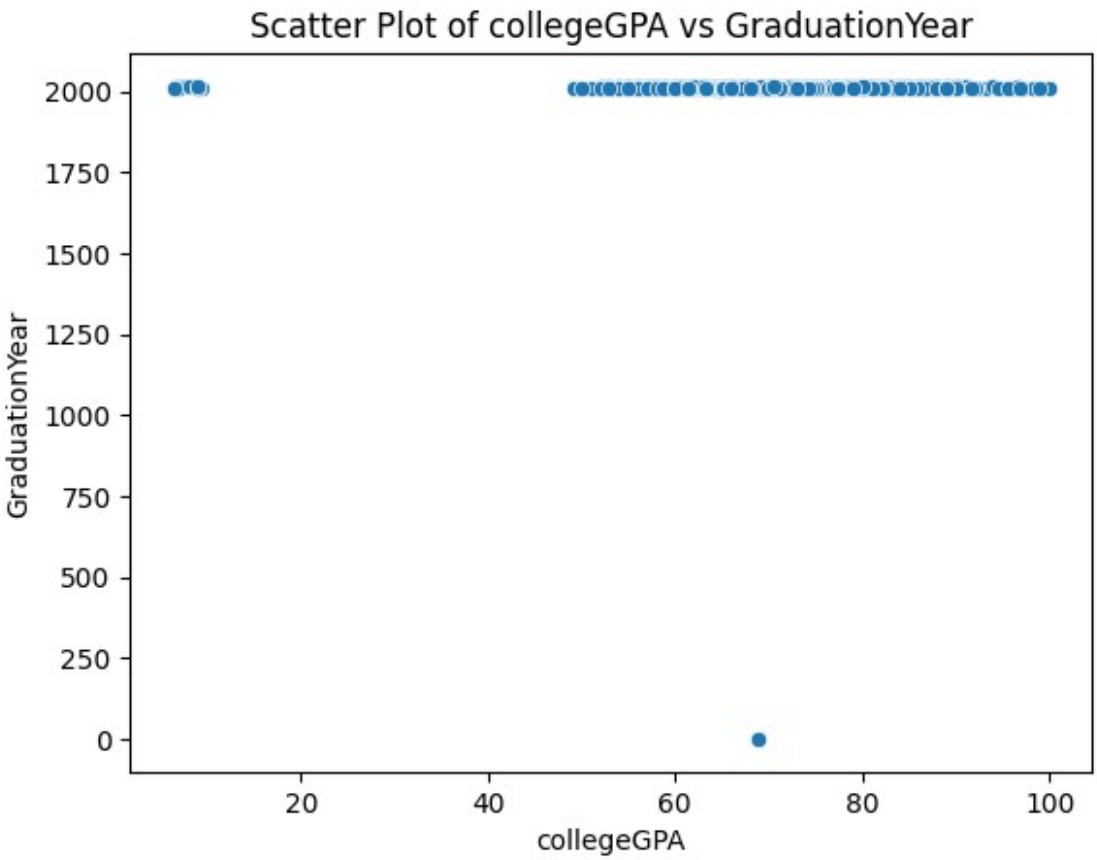

Scatter Plot of 12percentage vs extraversion
extraversion
12percentage
2
1
0
−1
−2
−3
−4
40
50
60
70
80
90
100

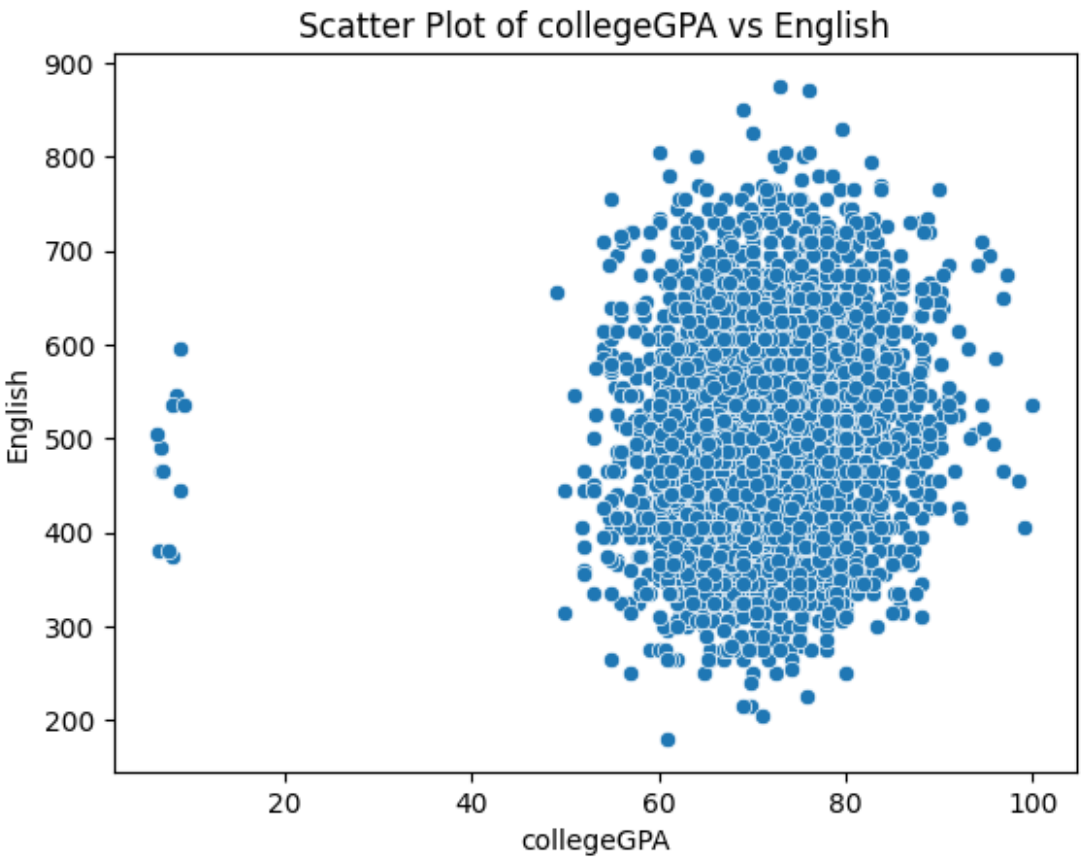

Scatter Plot of 12percentage vs nueroticism
nueroticism
12percentage
3
2
1
0
−1
−2
40
50
60
70
80
90
100

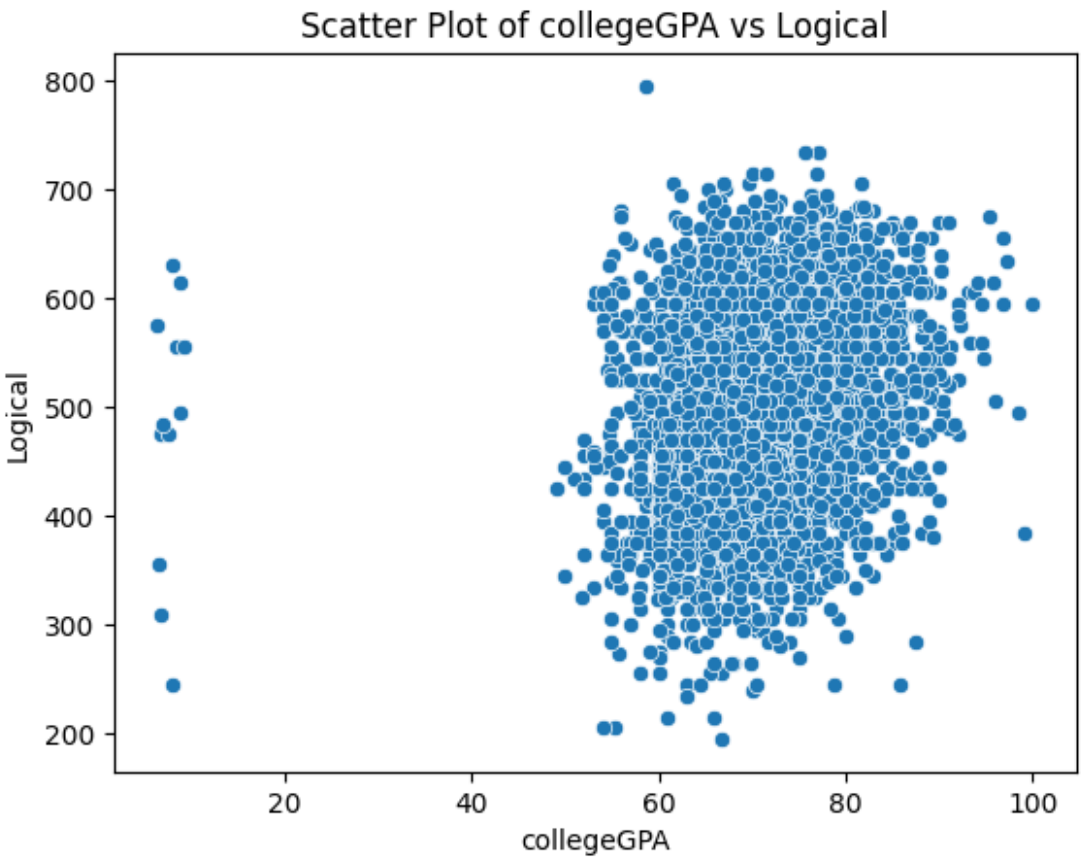

Scatter Plot of 12percentage vs openess_to_experience
openess_to_experience
12percentage
2
0
−2
−4
−6
40
50
60
70
80
90
100

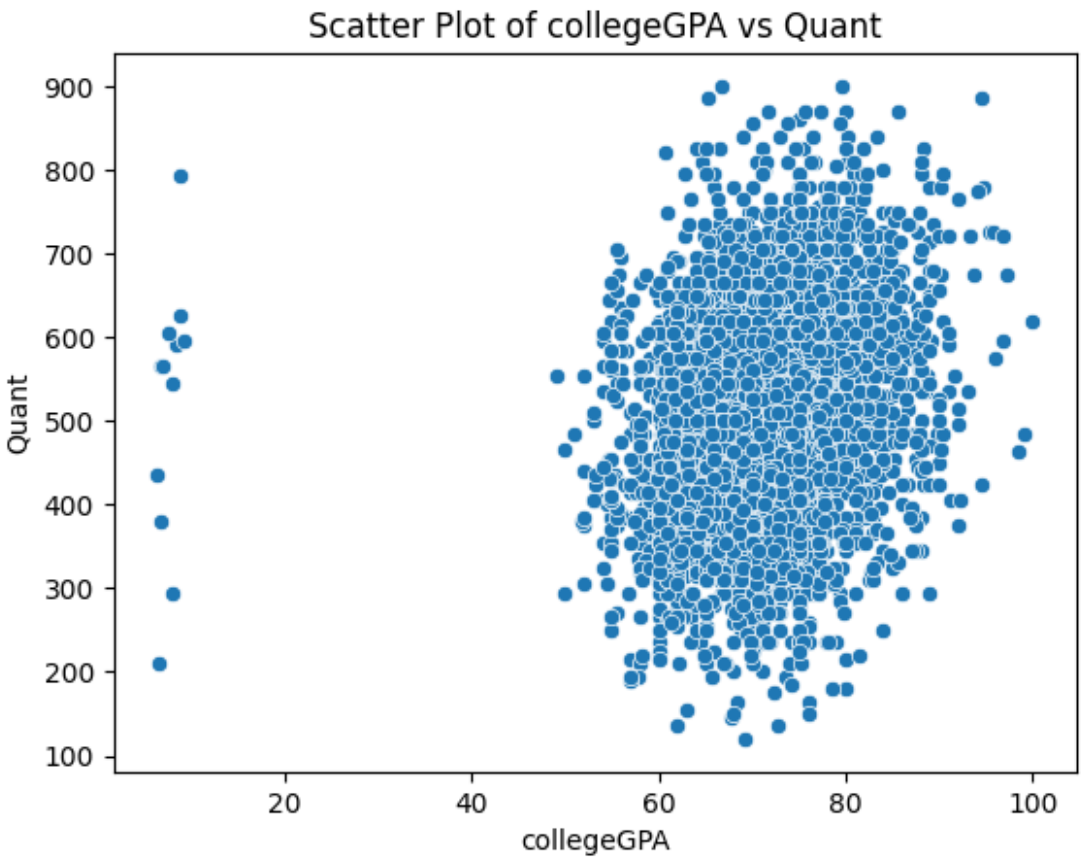

Scatter Plot of 12percentage vs Age
Age
12percentage
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
40
50
60
70
80
90
100

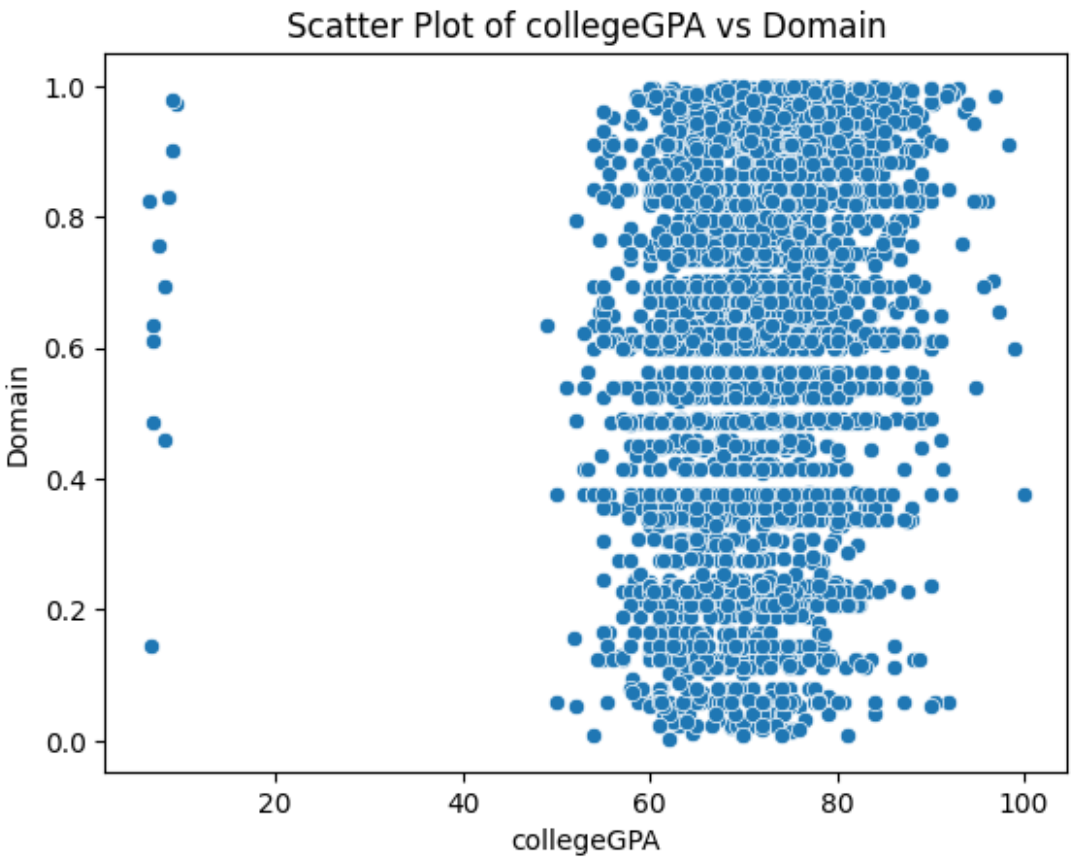

Scatter Plot of collegeGPA vs CollegeCityID
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
20
40
60
80
100
collegeGPA

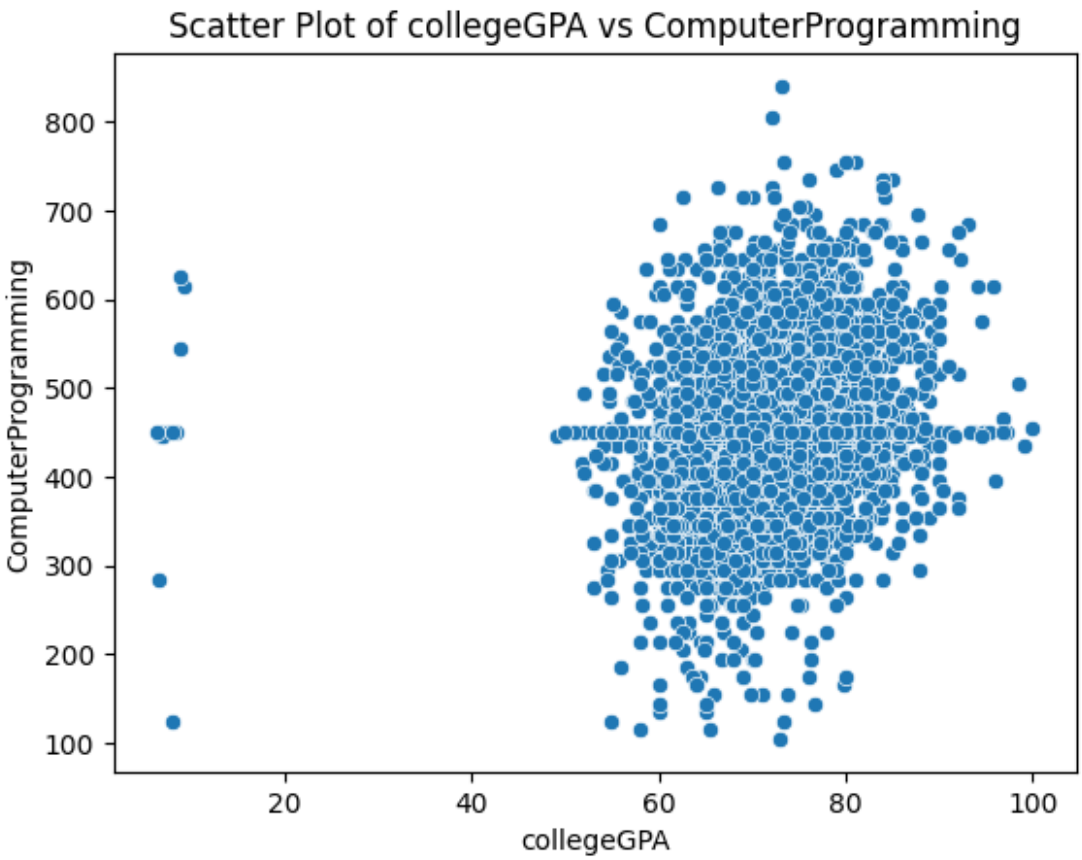

Scatter Plot of collegeGPA vs GraduationYear
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
20
40
60
80
100
collegeGPA

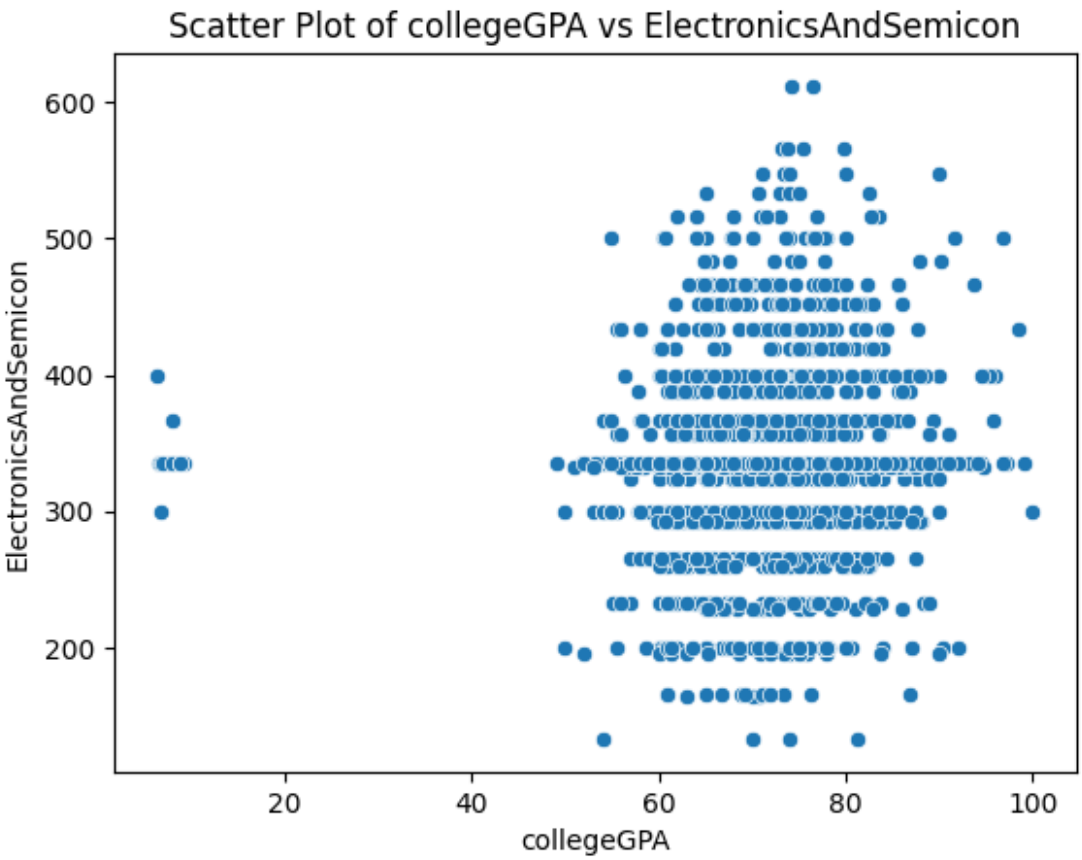

Scatter Plot of collegeGPA vs English
900
800
700
600
500
400
300
200
English
20
40
60
80
100
collegeGPA

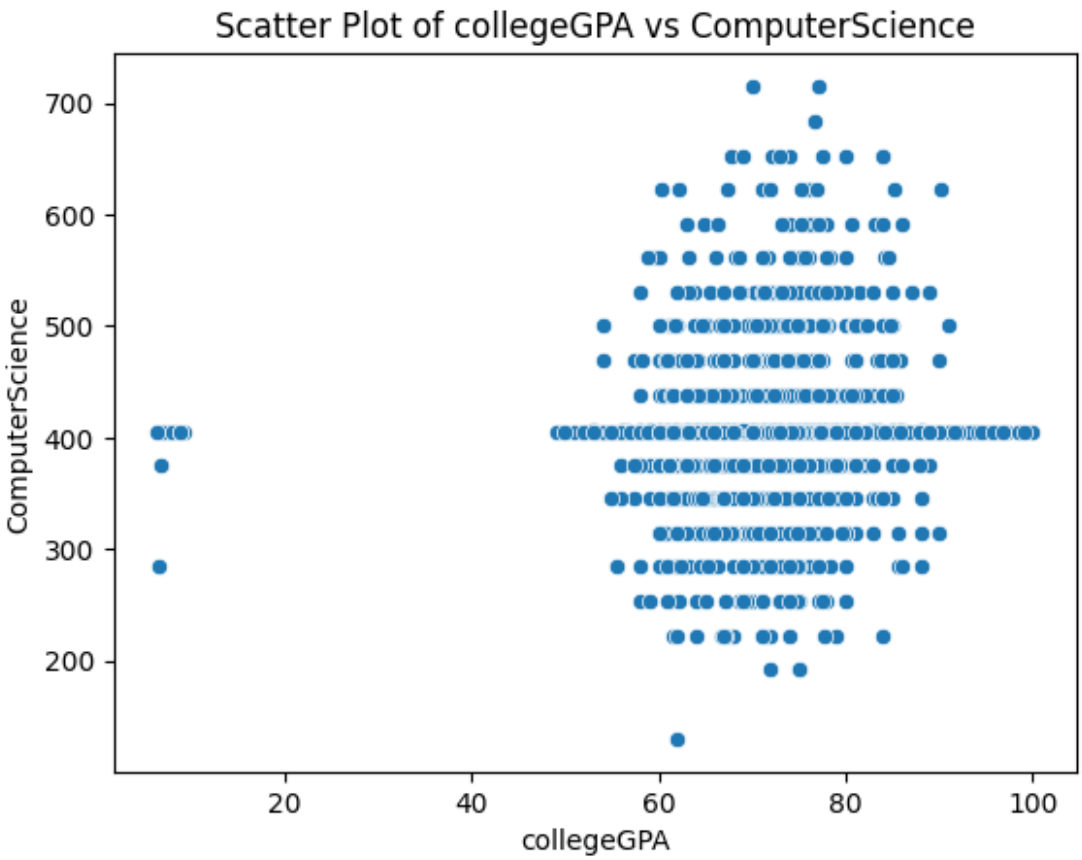

Scatter Plot of collegeGPA vs Logical
800
700
600
500
400
300
200
Logical
20
40
60
80
100
collegeGPA

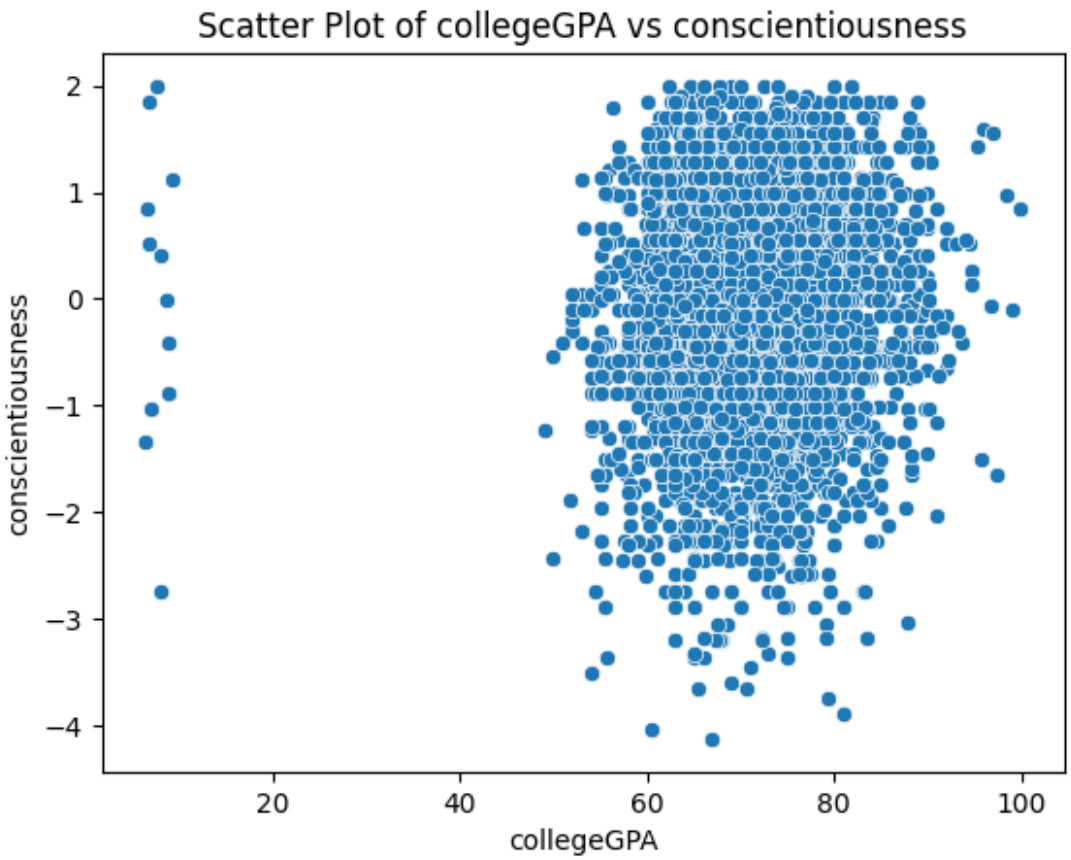

Scatter Plot of collegeGPA vs Quant
Quant
900
800
700
600
500
400
300
200
100
20
40
60
80
100
collegeGPA

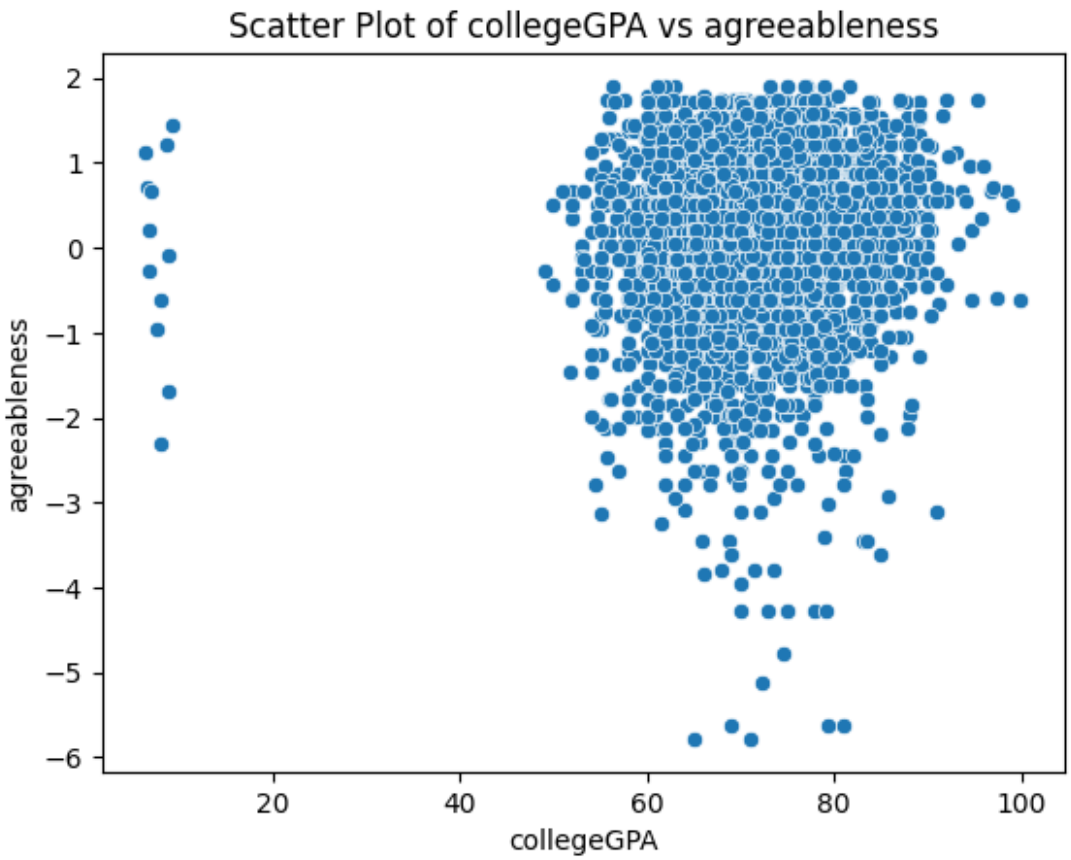

Scatter Plot of collegeGPA vs Domain
Domain
collegeGPA
1.0
0.8
0.6
0.4
0.2
0.0
20
40
60
80
100

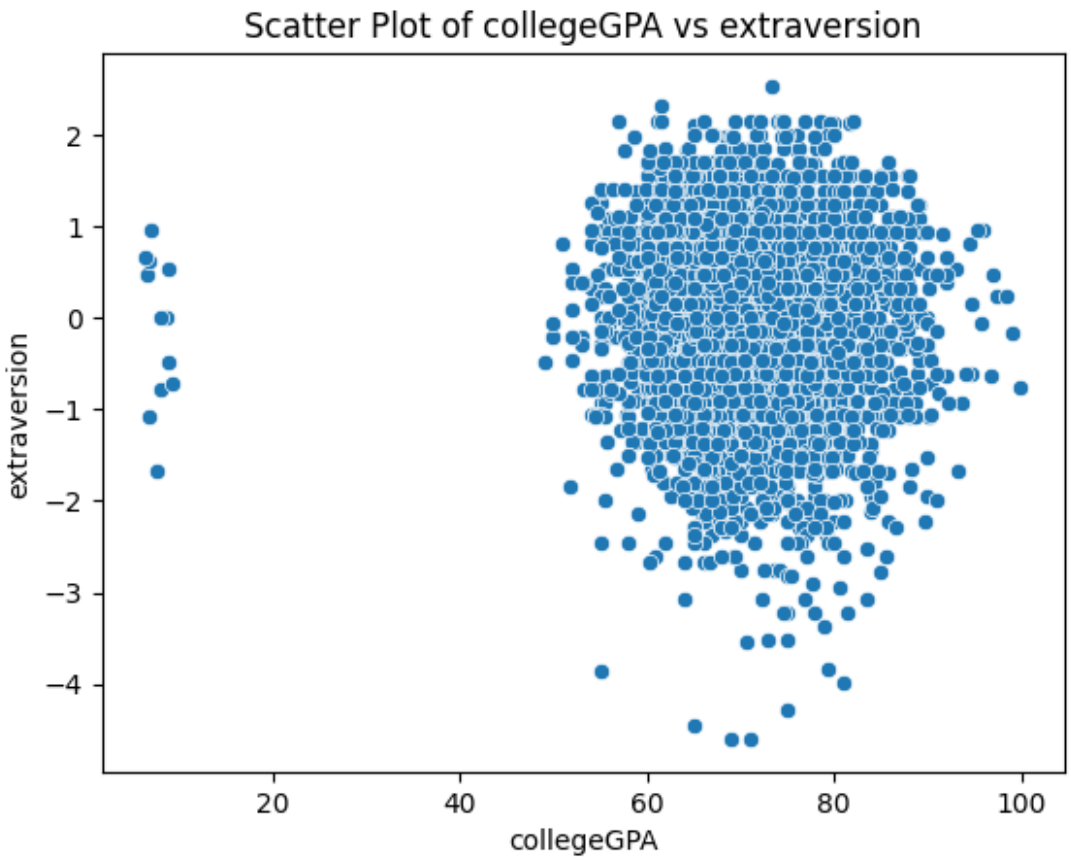

Scatter Plot of collegeGPA vs ComputerProgramming
ComputerProgramming
800
700
600
500
400
300
200
100
20
40
60
80
100
collegeGPA

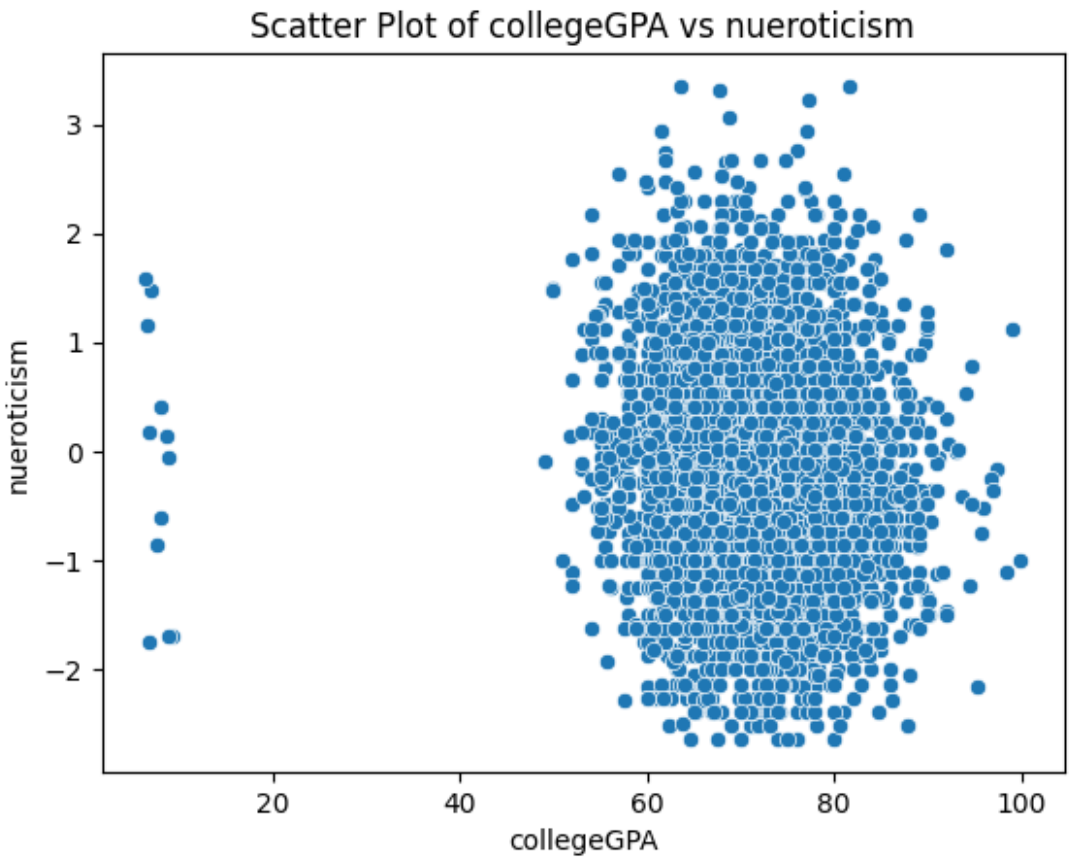

Scatter Plot of collegeGPA vs ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200
20
40
60
80
100
collegeGPA

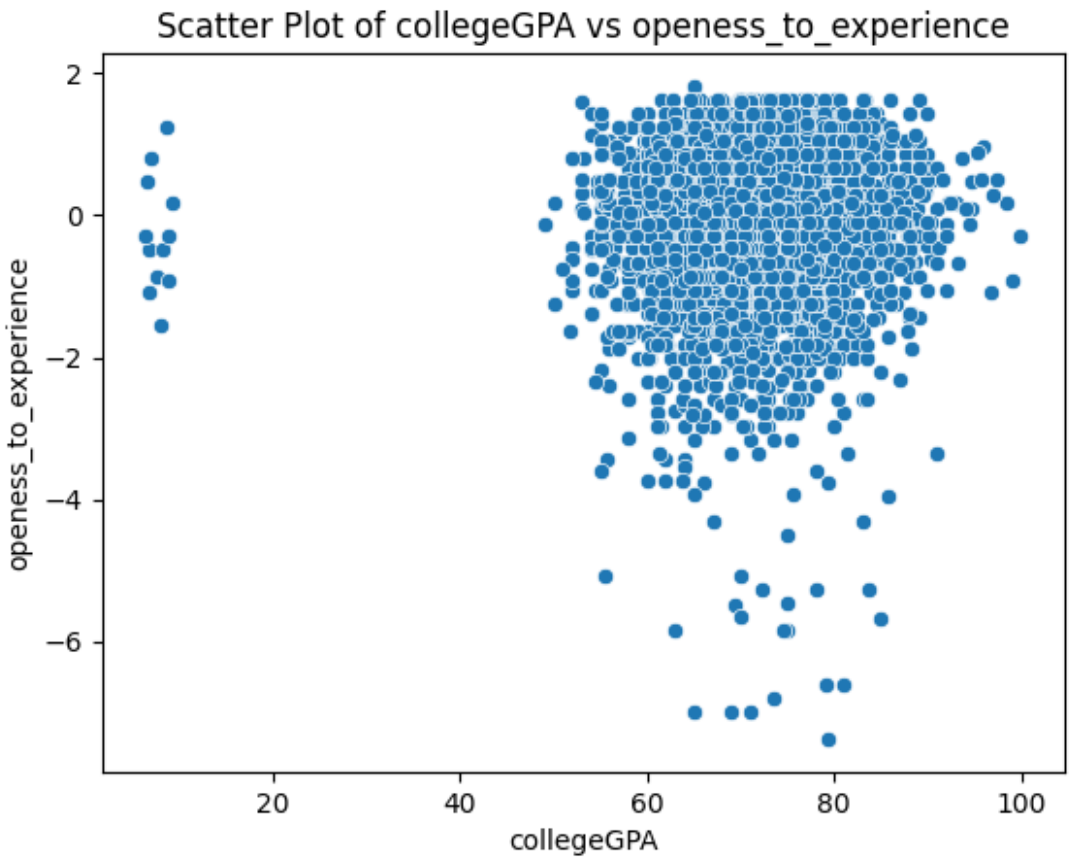

Scatter Plot of collegeGPA vs ComputerScience
ComputerScience
700
600
500
400
300
200
20
40
60
80
100
collegeGPA

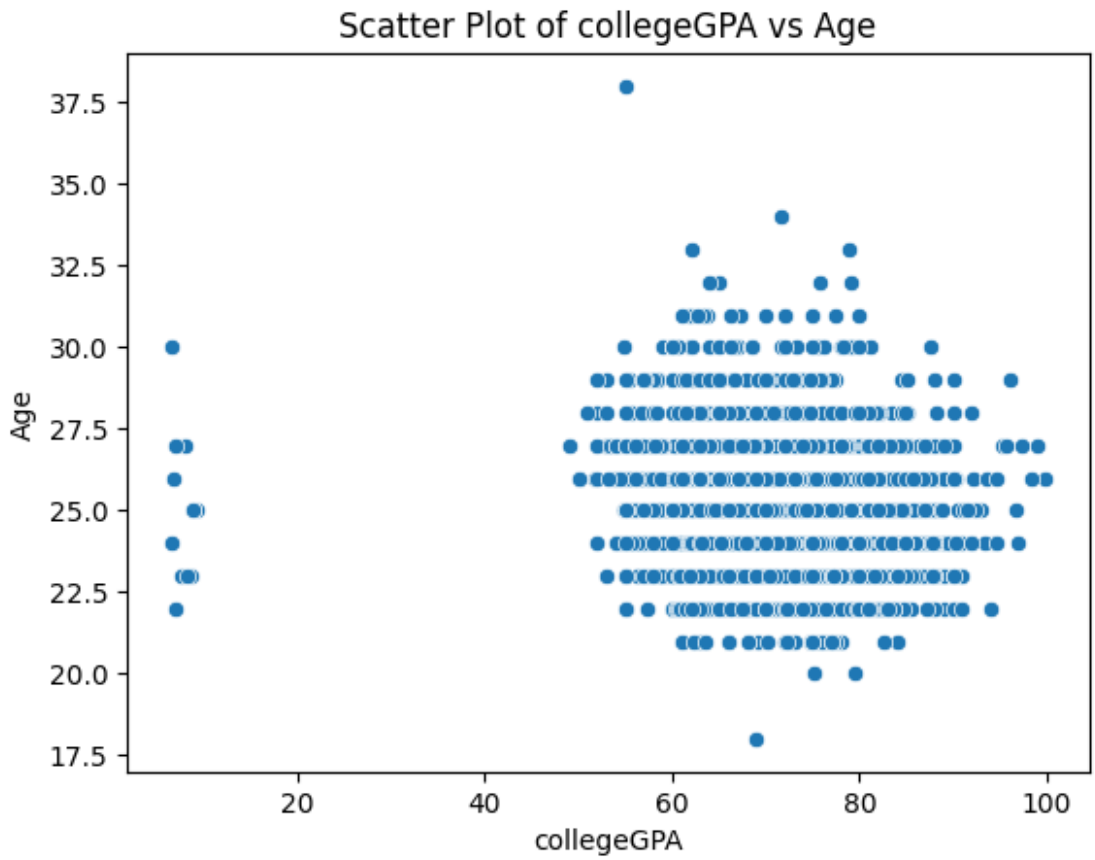

Scatter Plot of collegeGPA vs conscientiousness
conscientiousness
collegeGPA
2
1
0
−1
−2
−3
−4
20
40
60
80
100

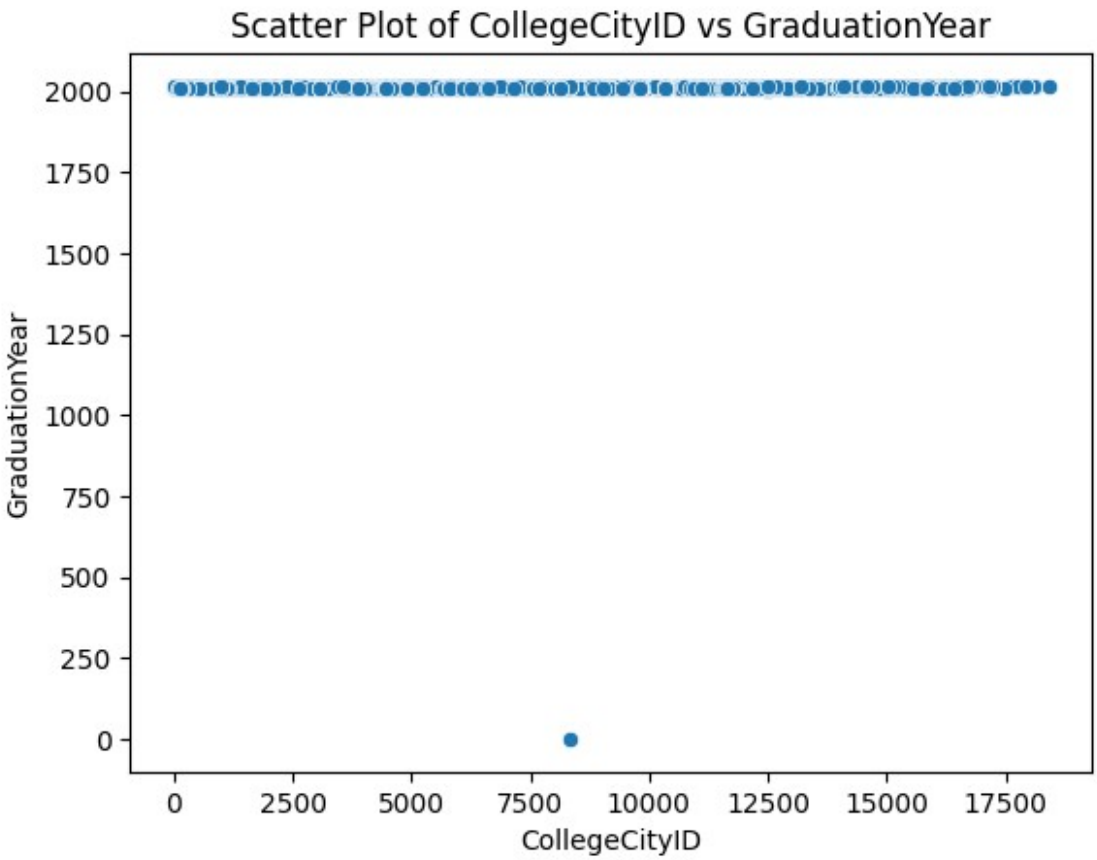

Scatter Plot of collegeGPA vs agreeableness
agreeableness
collegeGPA
2
1
0
−1
−2
−3
−4
−5
−6
20
40
60
80
100

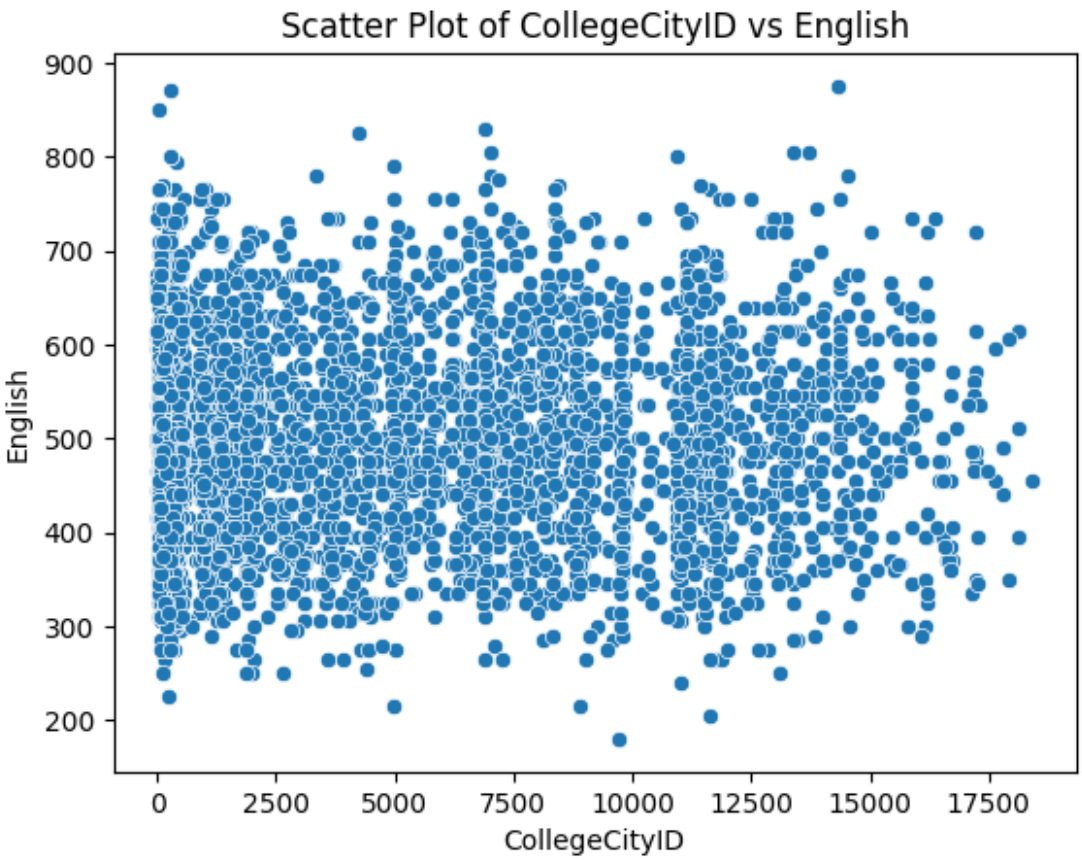

Scatter Plot of collegeGPA vs extraversion
extraversion
collegeGPA
2
1
0
−1
−2
−3
−4
20
40
60
80
100

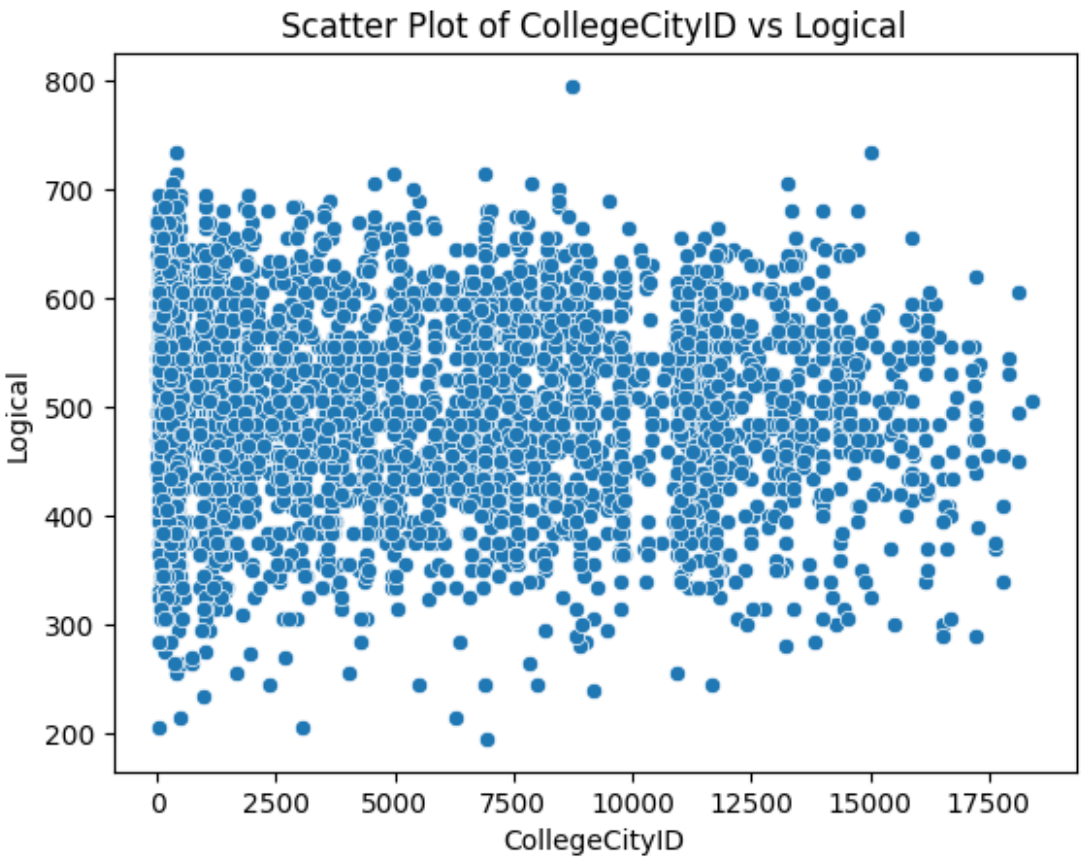

Scatter Plot of collegeGPA vs nueroticism
nueroticism
collegeGPA
3
2
1
0
−1
−2
20
40
60
80
100

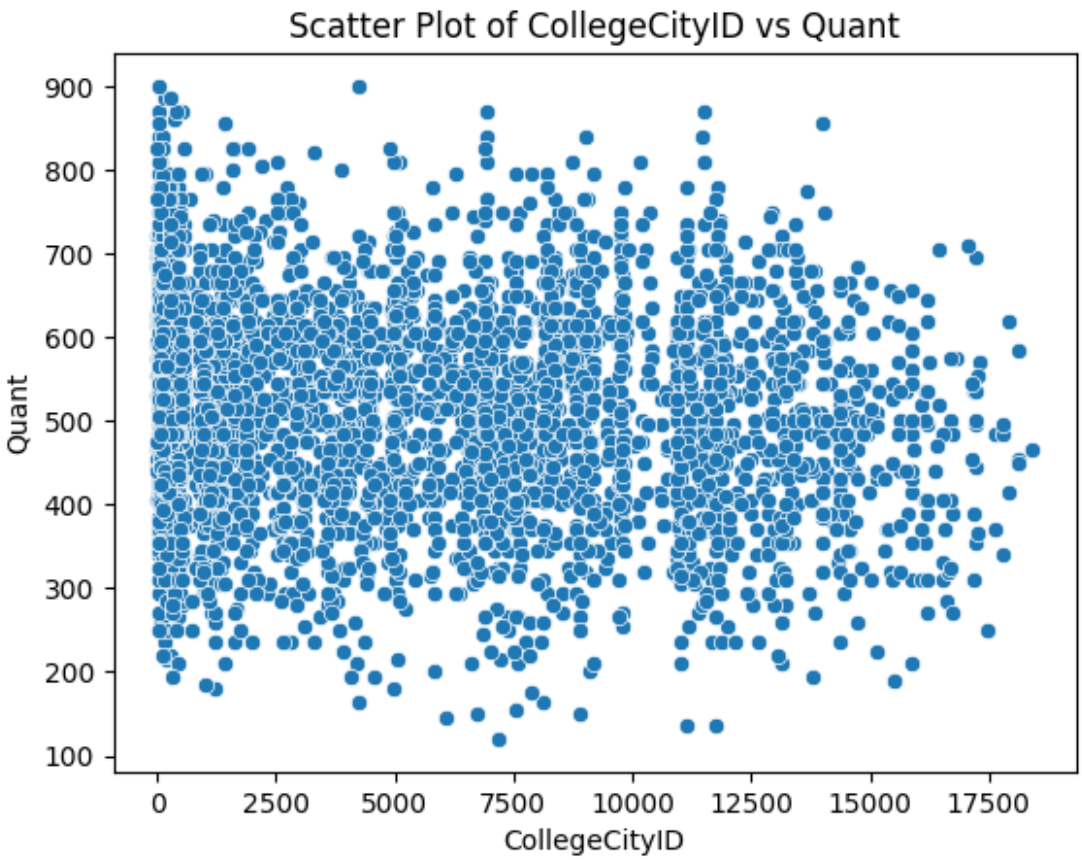

Scatter Plot of collegeGPA vs openess_to_experience
openess_to_experience
collegeGPA
2
0
−2
−4
−6
20
40
60
80
100

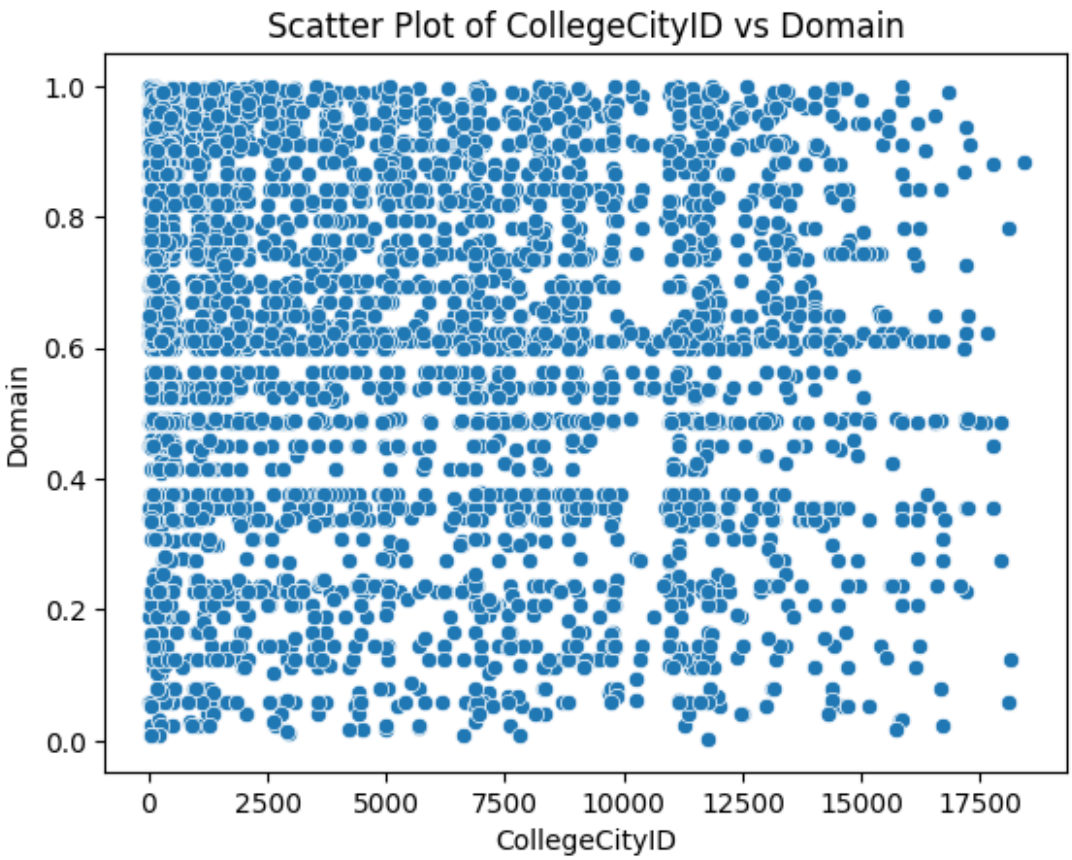

Scatter Plot of collegeGPA vs Age
Age
collegeGPA
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
20
40
60
80
100

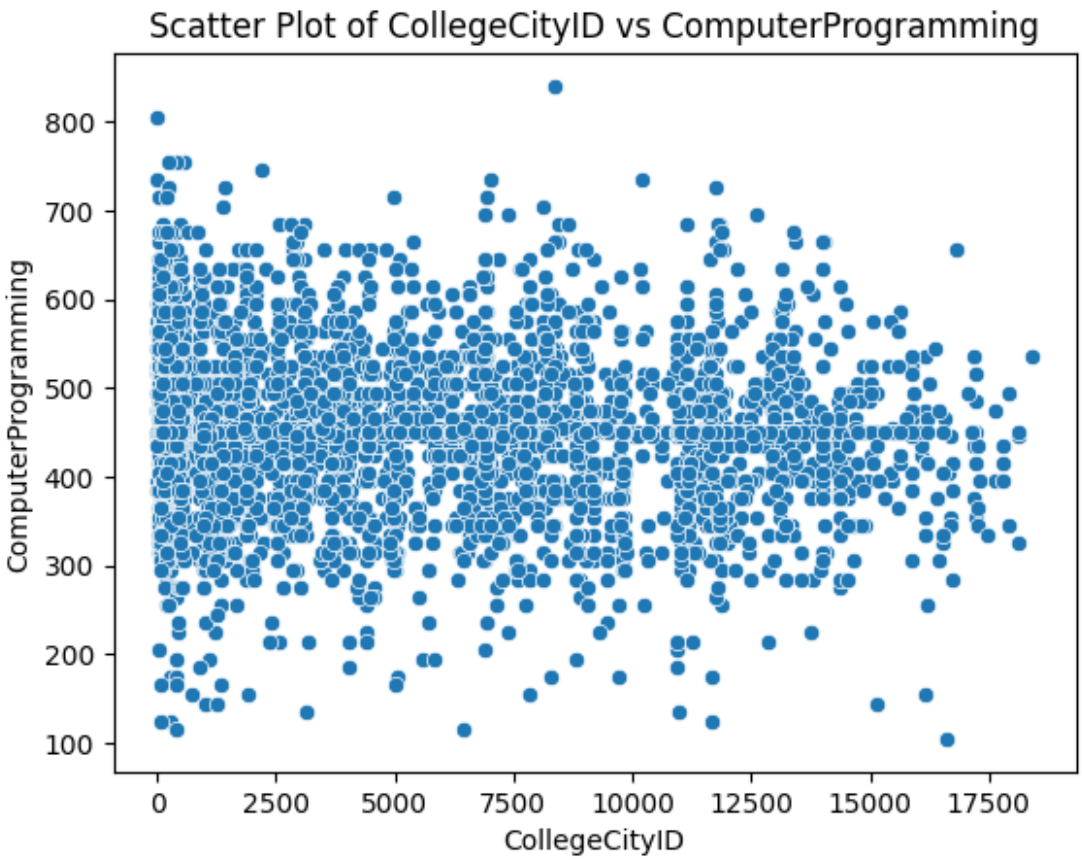

Scatter Plot of CollegeCityID vs GraduationYear
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
0
2500
5000
7500
10000
12500
15000
17500
CollegeCityID

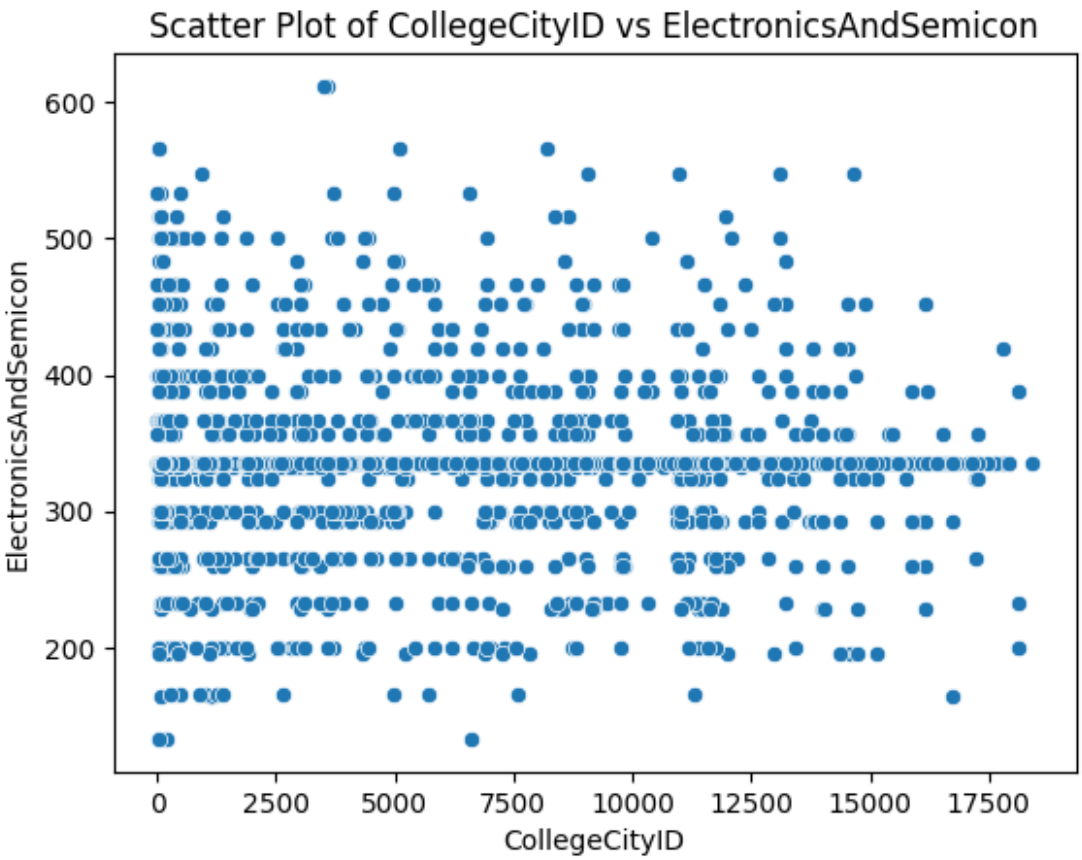

Scatter Plot of CollegeCityID vs English
English
CollegeCityID
900
800
700
600
500
400
300
200
0
2500
5000
7500
10000
12500
15000
17500

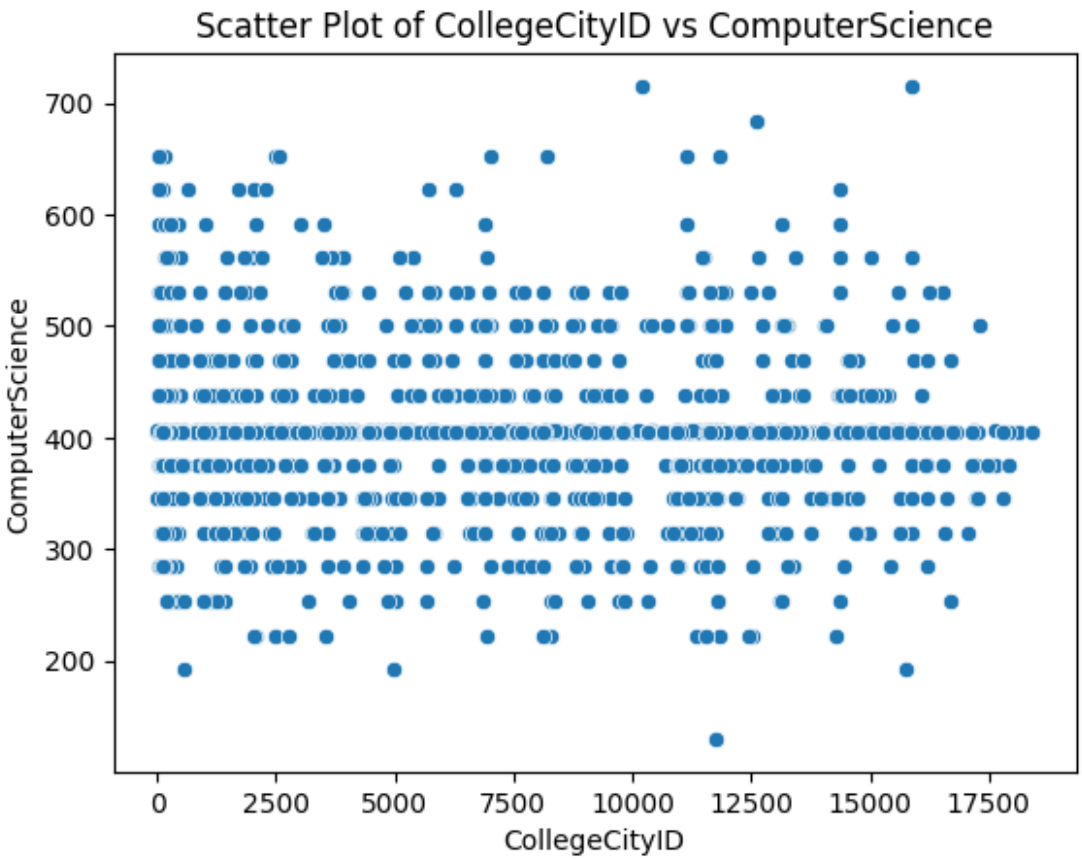

Scatter Plot of CollegeCityID vs Logical
Logical
CollegeCityID
800
700
600
500
400
300
200
0
2500
5000
7500
10000
12500
15000
17500

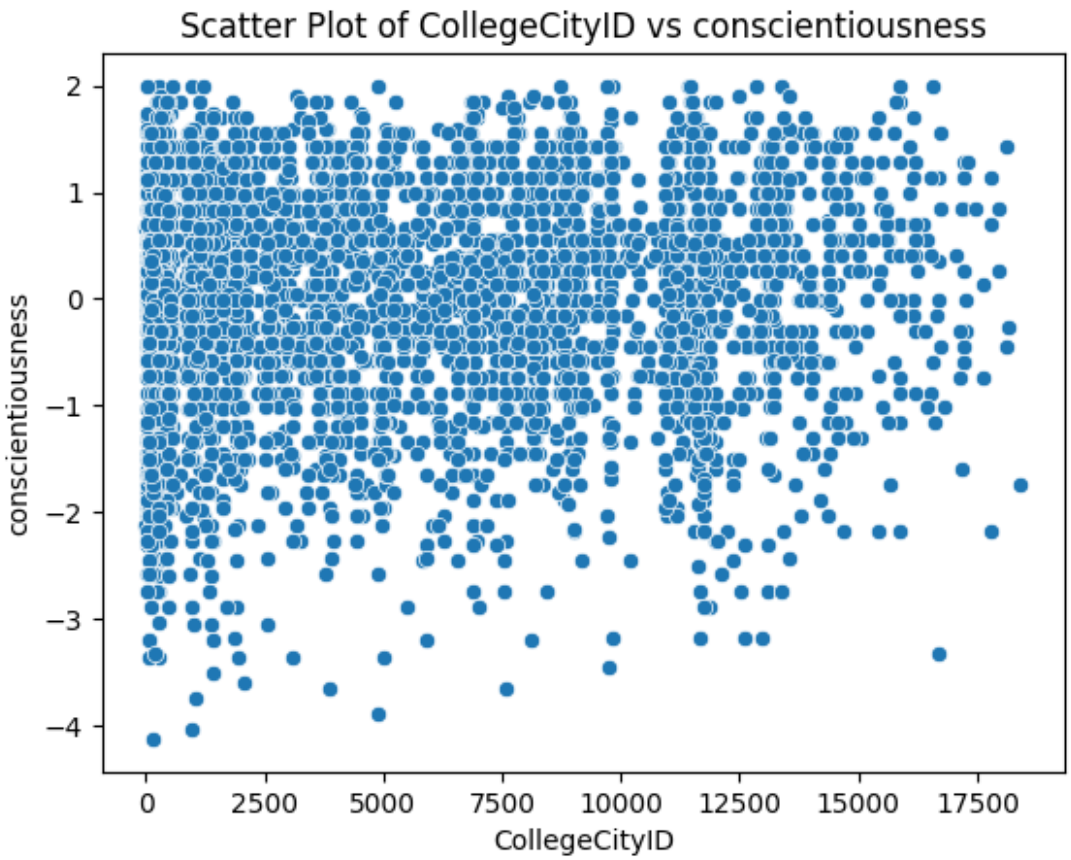

Scatter Plot of CollegeCityID vs Quant
Quant
900
800
700
600
500
400
300
200
100
0
2500
5000
7500
10000
12500
15000
17500
CollegeCityID

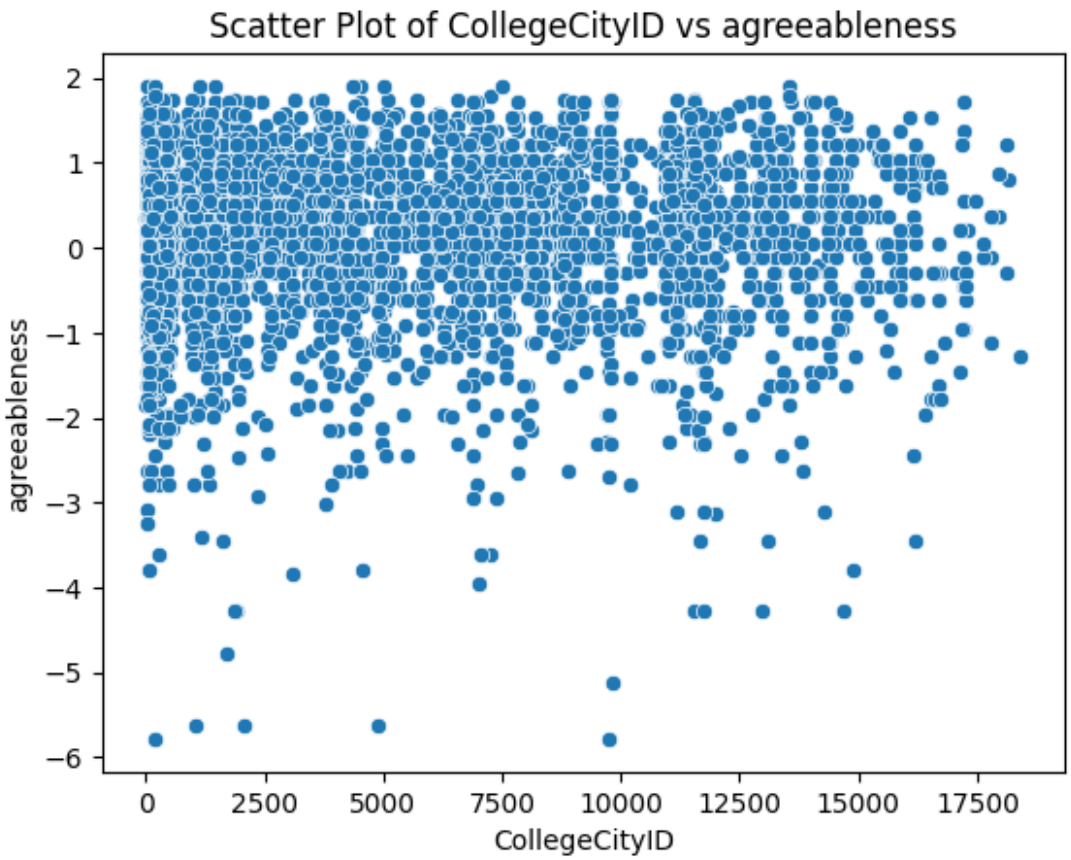

Scatter Plot of CollegeCityID vs Domain
Domain
1.0
0.8
0.6
0.4
0.2
0.0
0
2500
5000
7500
10000
12500
15000
17500
CollegeCityID

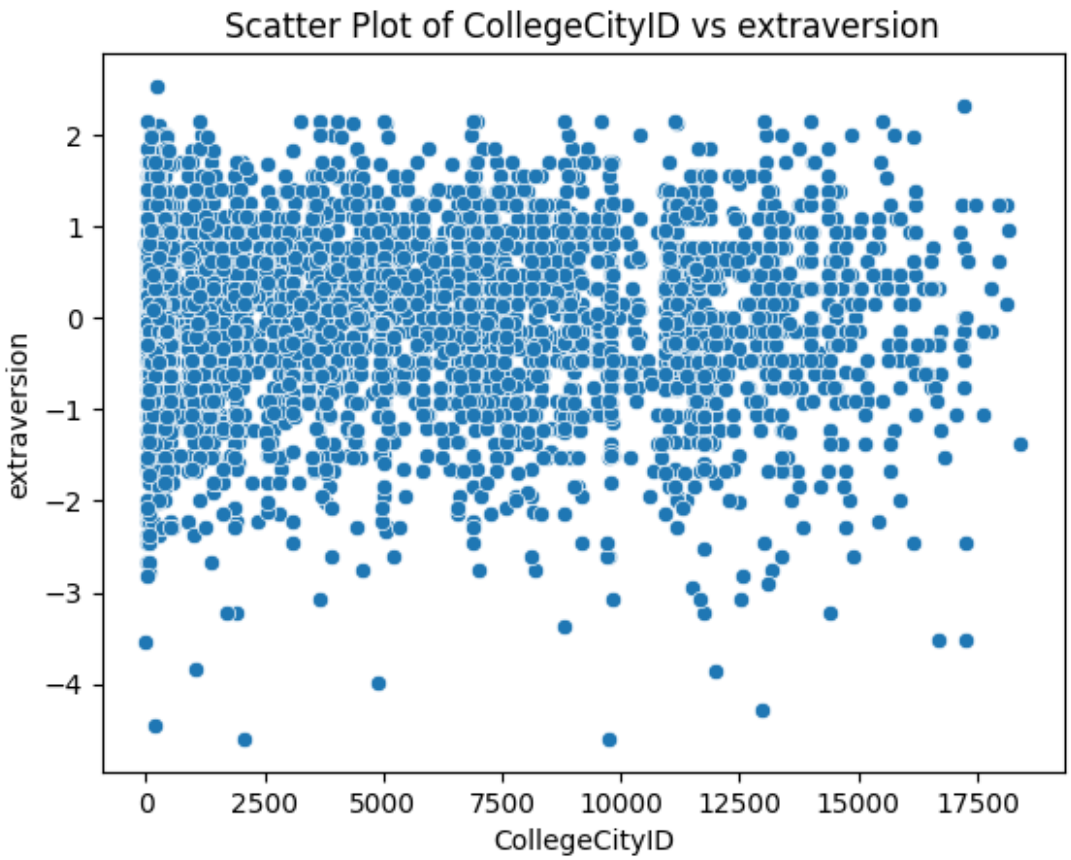

Scatter Plot of CollegeCityID vs ComputerProgramming
ComputerProgramming
800
700
600
500
400
300
200
100
0
2500
5000
7500
10000
12500
15000
17500
CollegeCityID

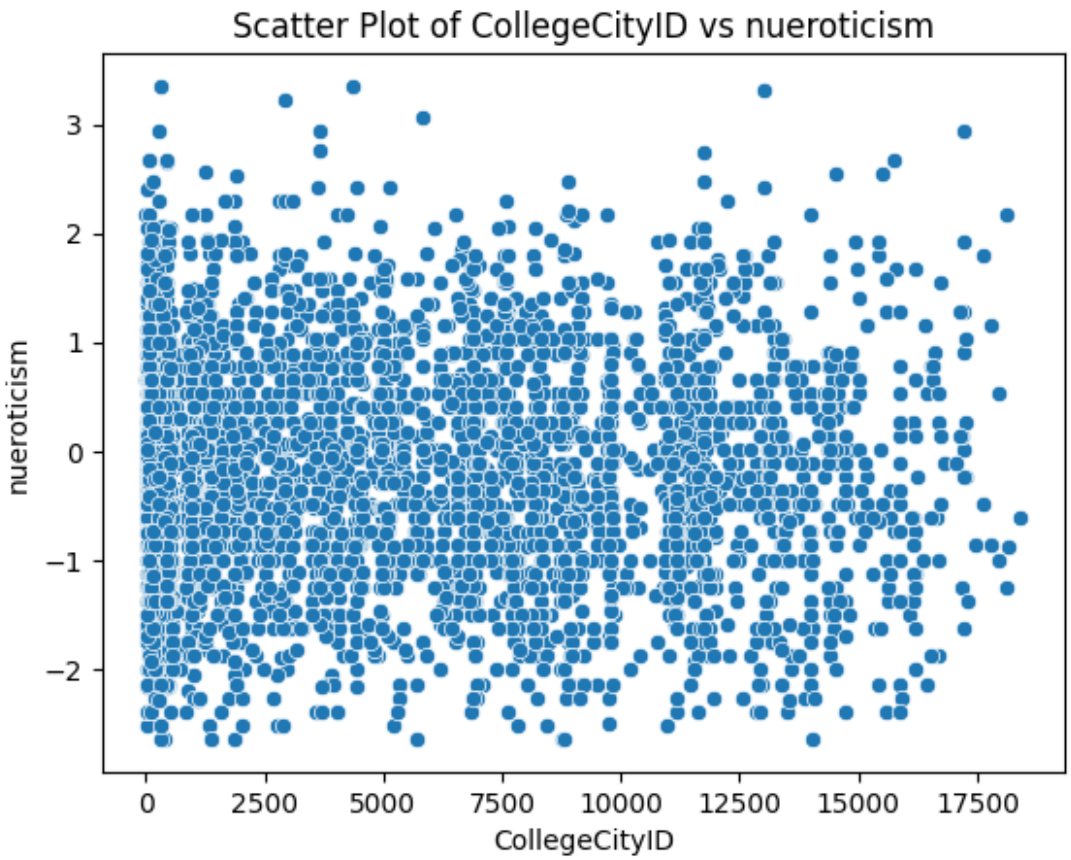

Scatter Plot of CollegeCityID vs ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200
0
2500
5000
7500
10000
12500
15000
17500
CollegeCityID

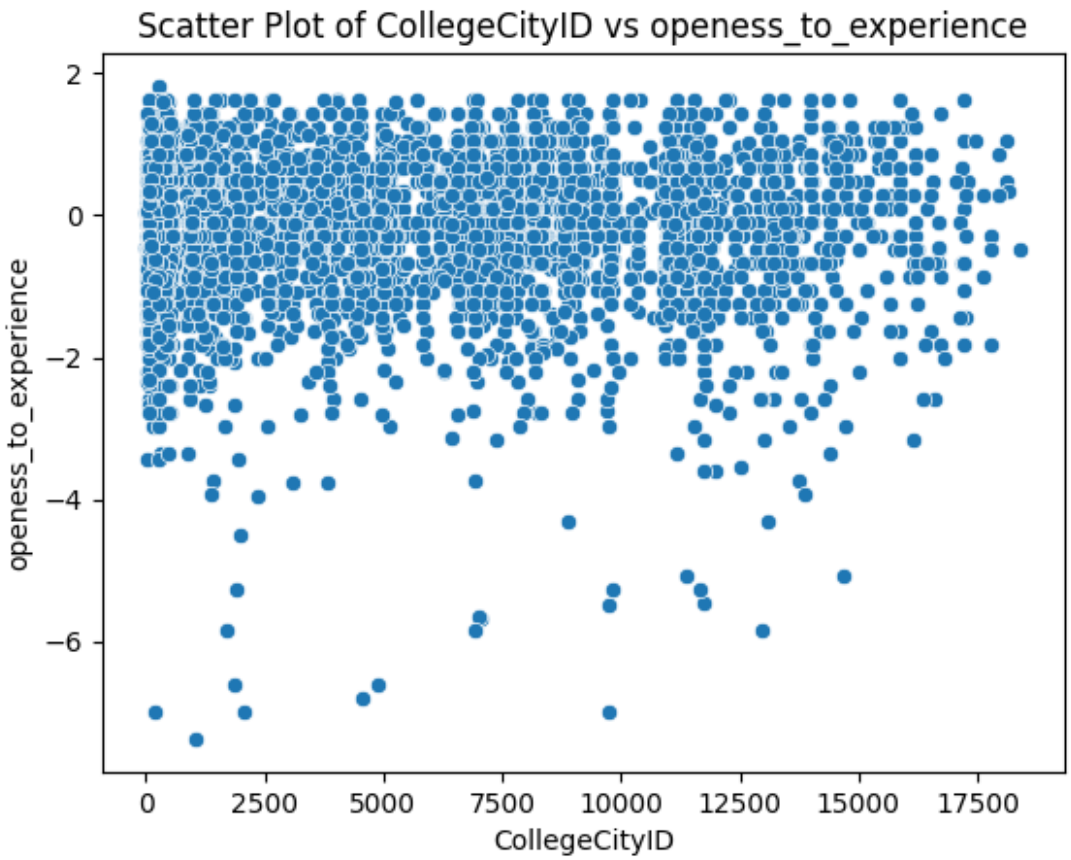

Scatter Plot of CollegeCityID vs ComputerScience
ComputerScience
700
600
500
400
300
200
0
2500
5000
7500
10000
12500
15000
17500
CollegeCityID

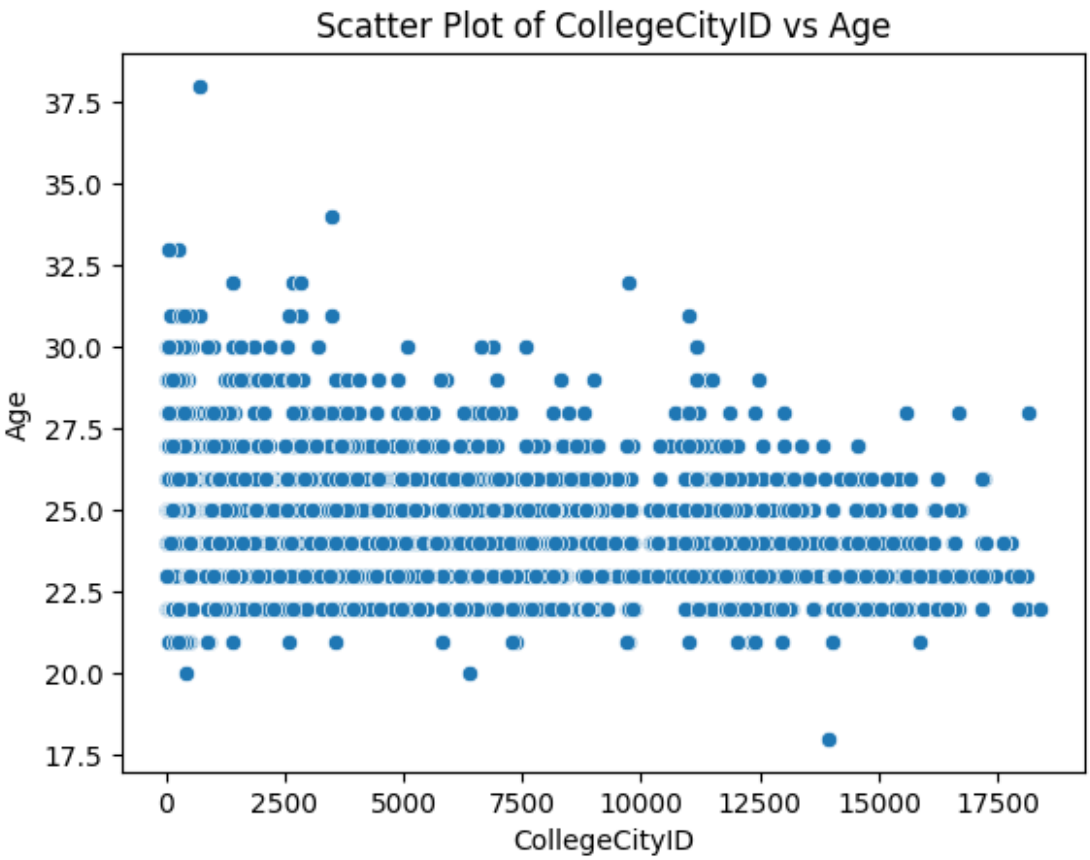

Scatter Plot of CollegeCityID vs conscientiousness
conscientiousness
CollegeCityID
2
1
0
−1
−2
−3
−4
0
2500
5000
7500
10000
12500
15000
17500

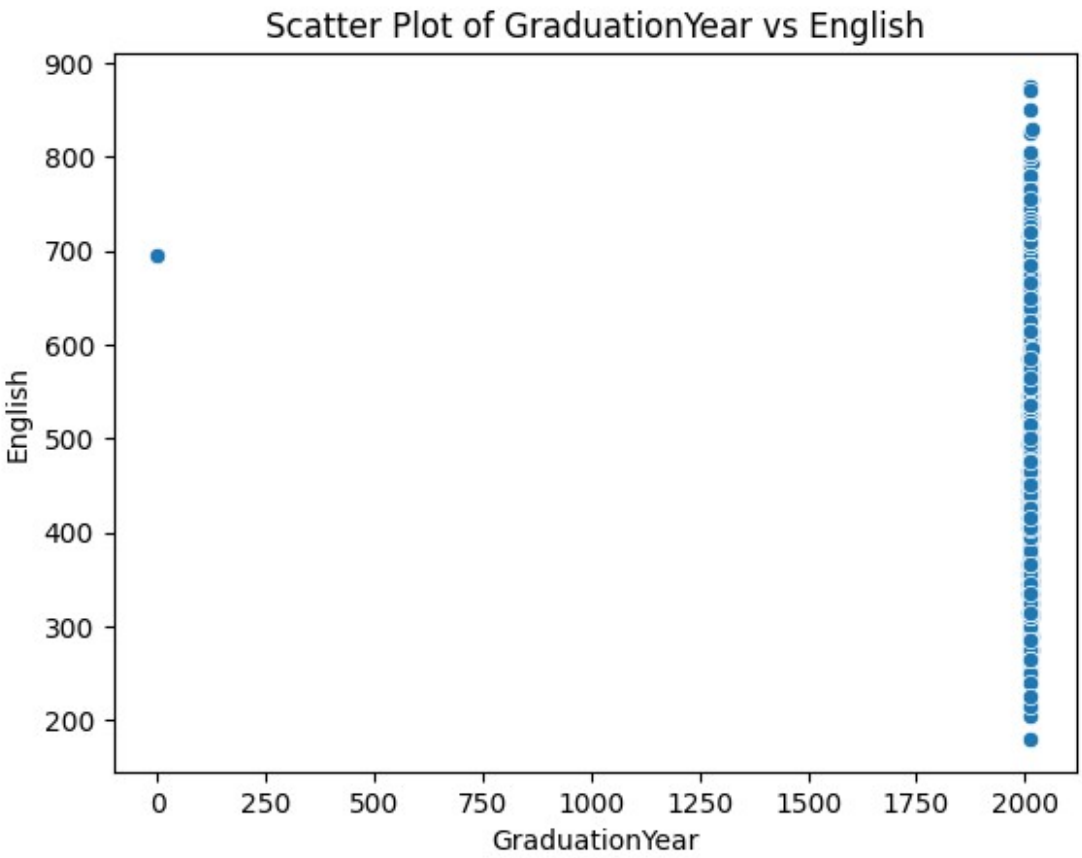

Scatter Plot of CollegeCityID vs agreeableness
agreeableness
CollegeCityID
2
1
0
−1
−2
−3
−4
−5
−6
0
2500
5000
7500
10000
12500
15000
17500

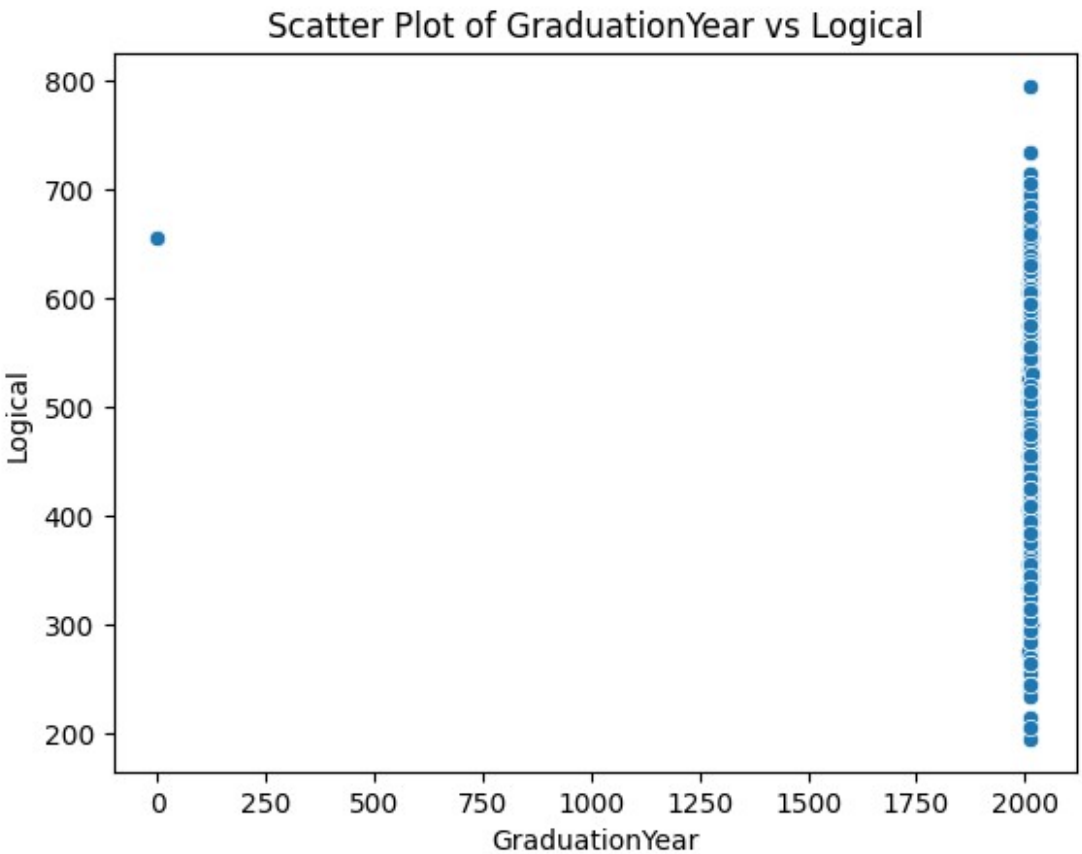

Scatter Plot of CollegeCityID vs extraversion
extraversion
CollegeCityID
2
1
0
−1
−2
−3
−4
0
2500
5000
7500
10000
12500
15000
17500

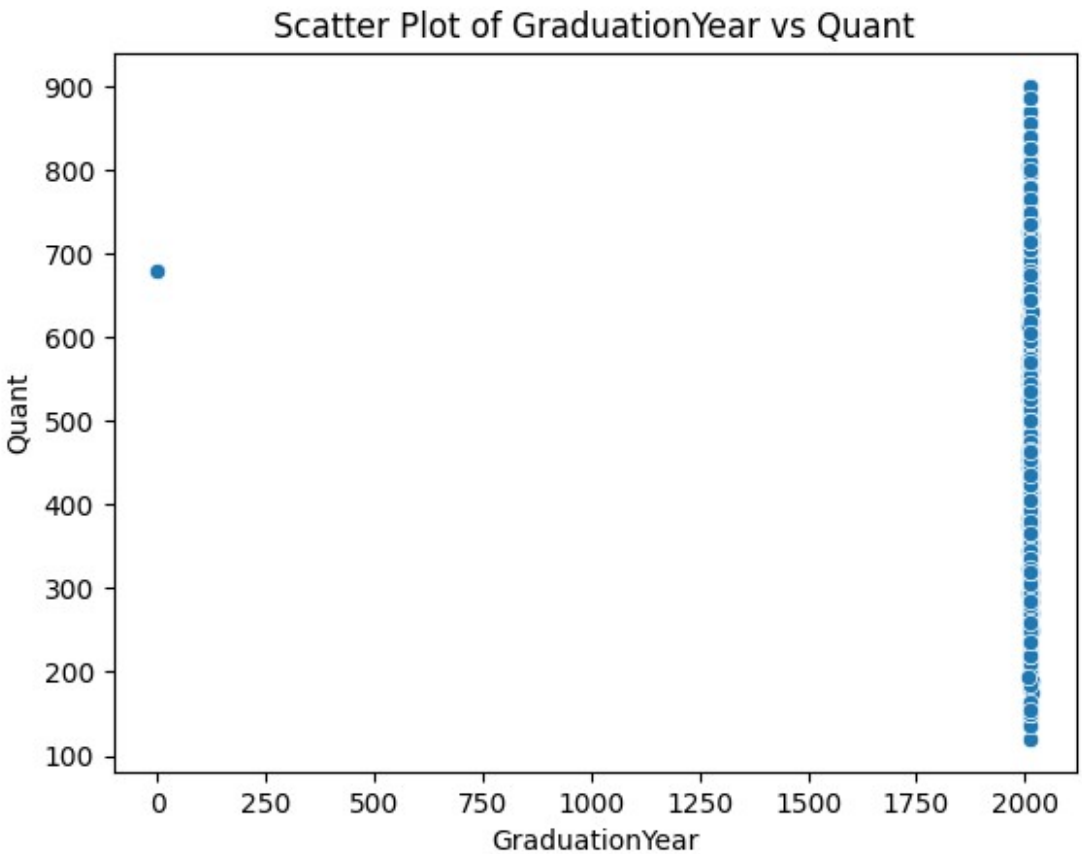

Scatter Plot of CollegeCityID vs nueroticism
nueroticism
CollegeCityID
3
2
1
0
−1
−2
0
2500
5000
7500
10000
12500
15000
17500

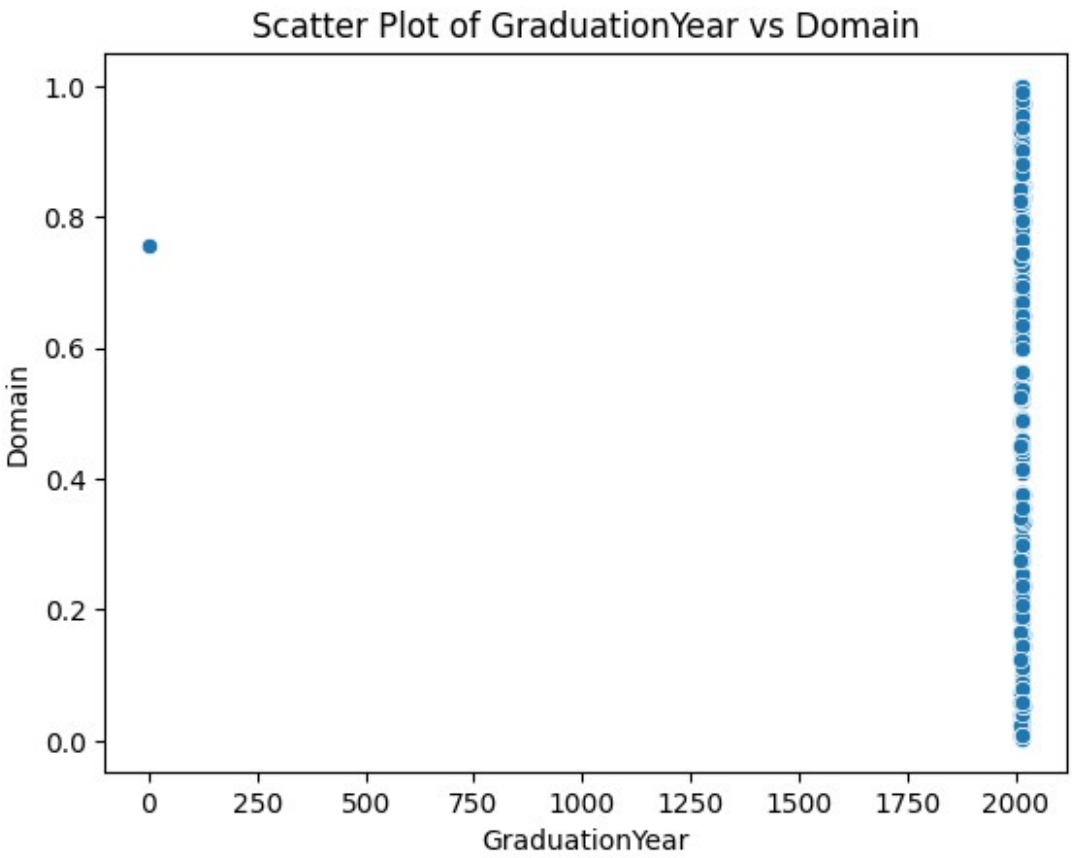

Scatter Plot of CollegeCityID vs openess_to_experience
openess_to_experience
CollegeCityID
2
0
−2
−4
−6
0
2500
5000
7500
10000
12500
15000
17500

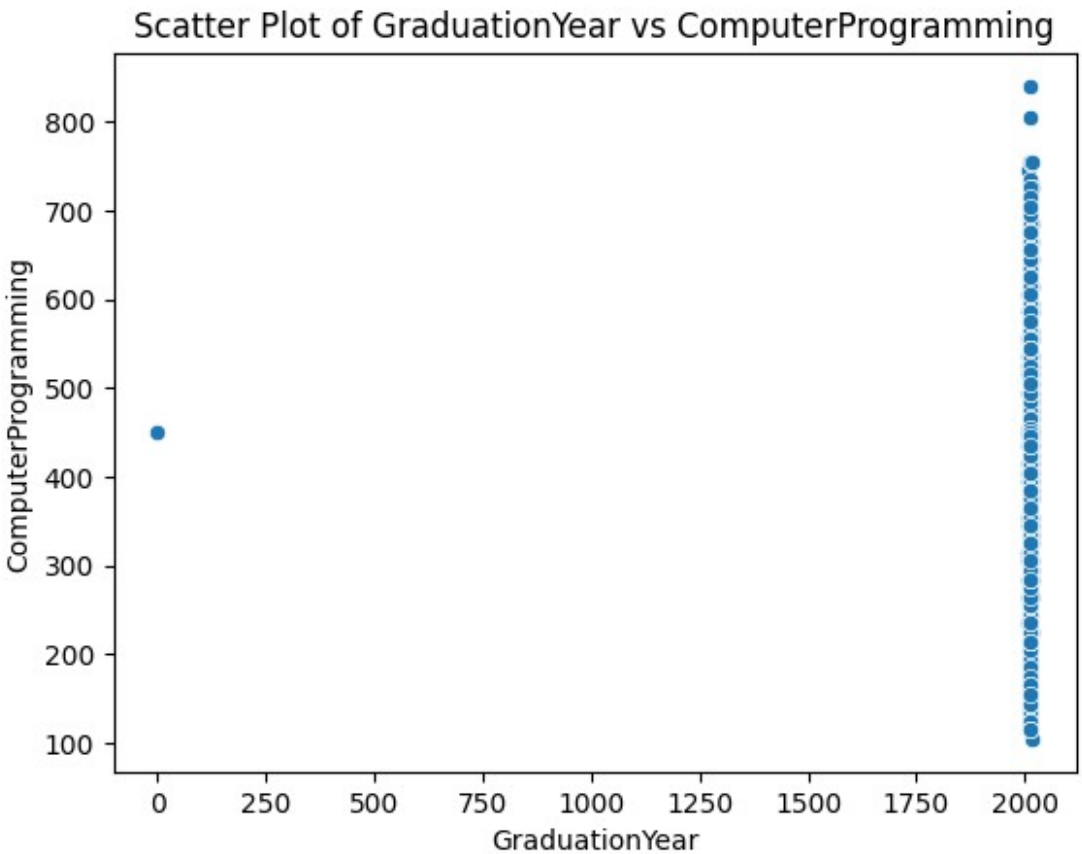

Scatter Plot of CollegeCityID vs Age
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
0
2500
5000
7500
10000
12500
15000
17500
CollegeCityID

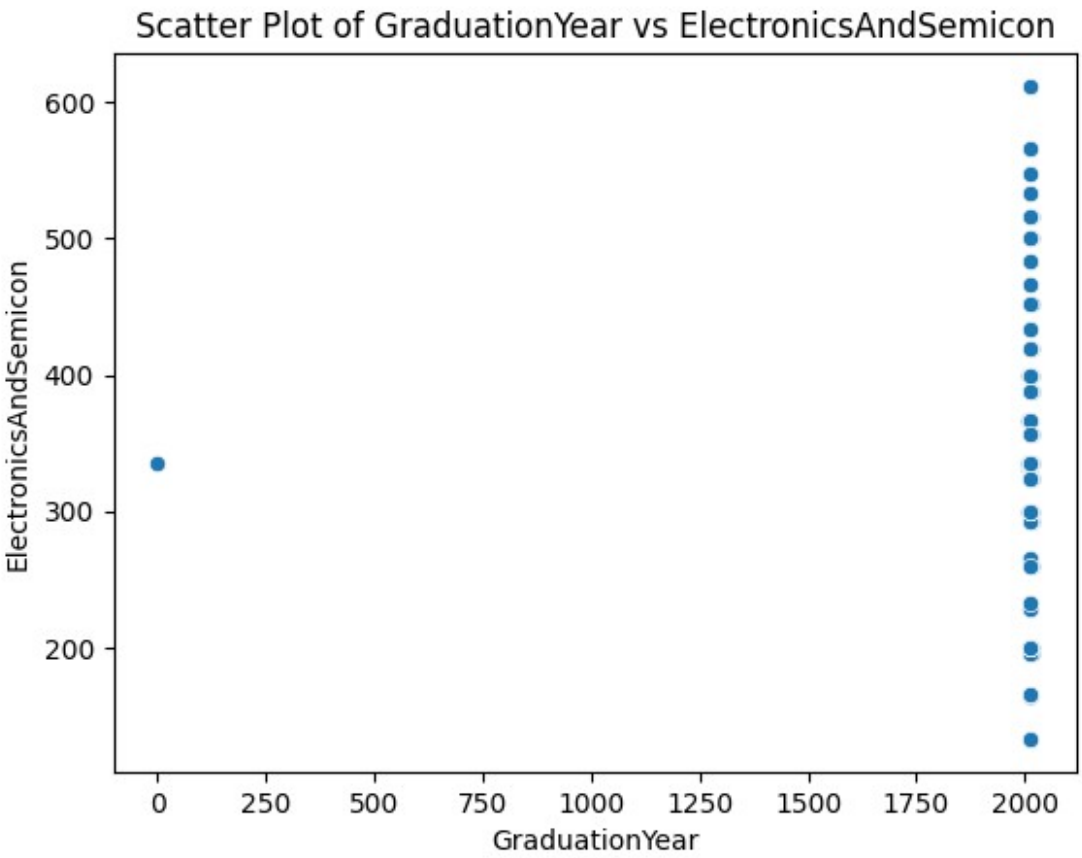

Scatter Plot of GraduationYear vs English
English
GraduationYear
900
800
700
600
500
400
300
200
0
250
500
750
1000
1250
1500
1750
2000

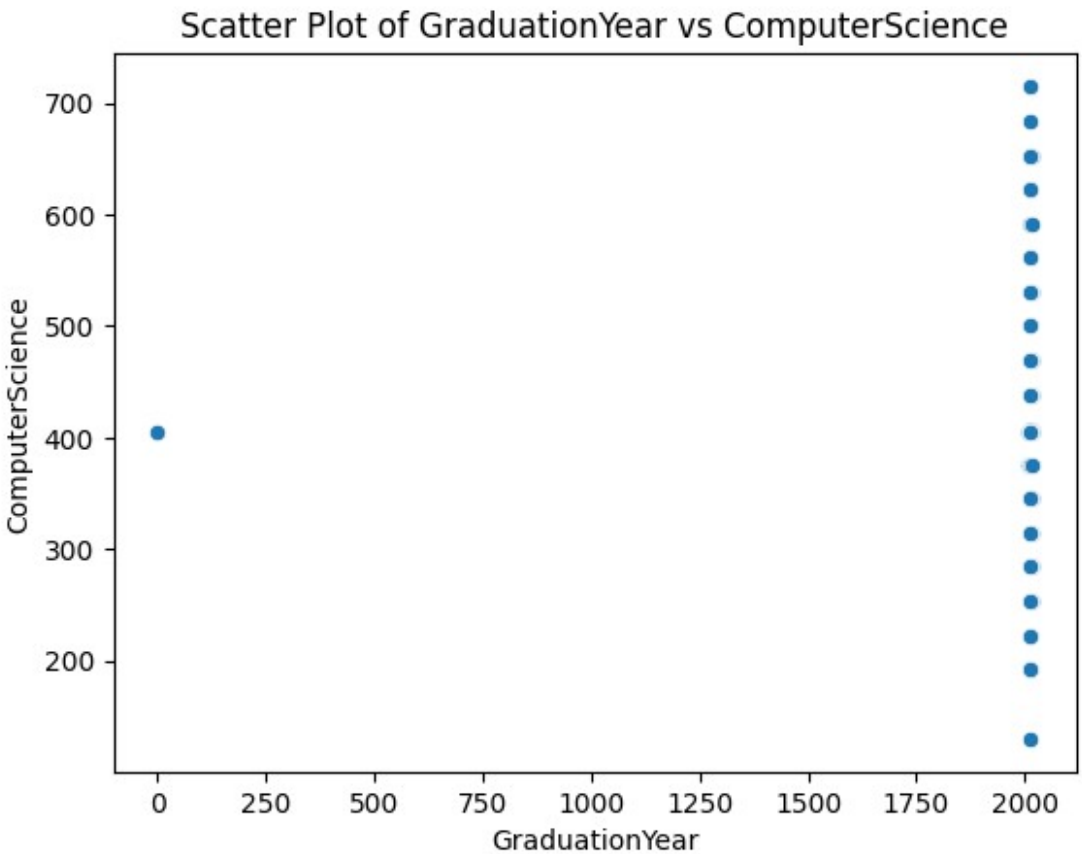

Scatter Plot of GraduationYear vs Logical
Logical
GraduationYear
800
700
600
500
400
300
200
0
250
500
750
1000
1250
1500
1750
2000

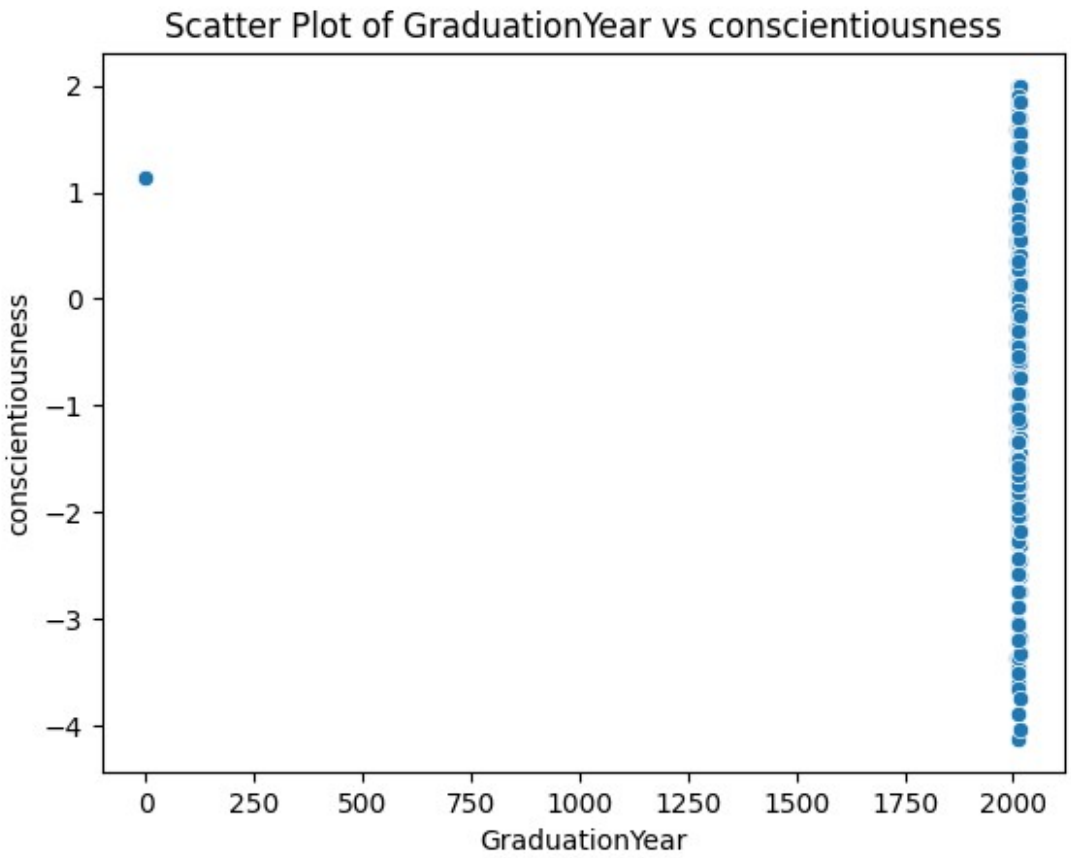

Scatter Plot of GraduationYear vs Quant
Quant
900
800
700
600
500
400
300
200
100
0
250
500
750
1000
1250
1500
1750
2000
GraduationYear

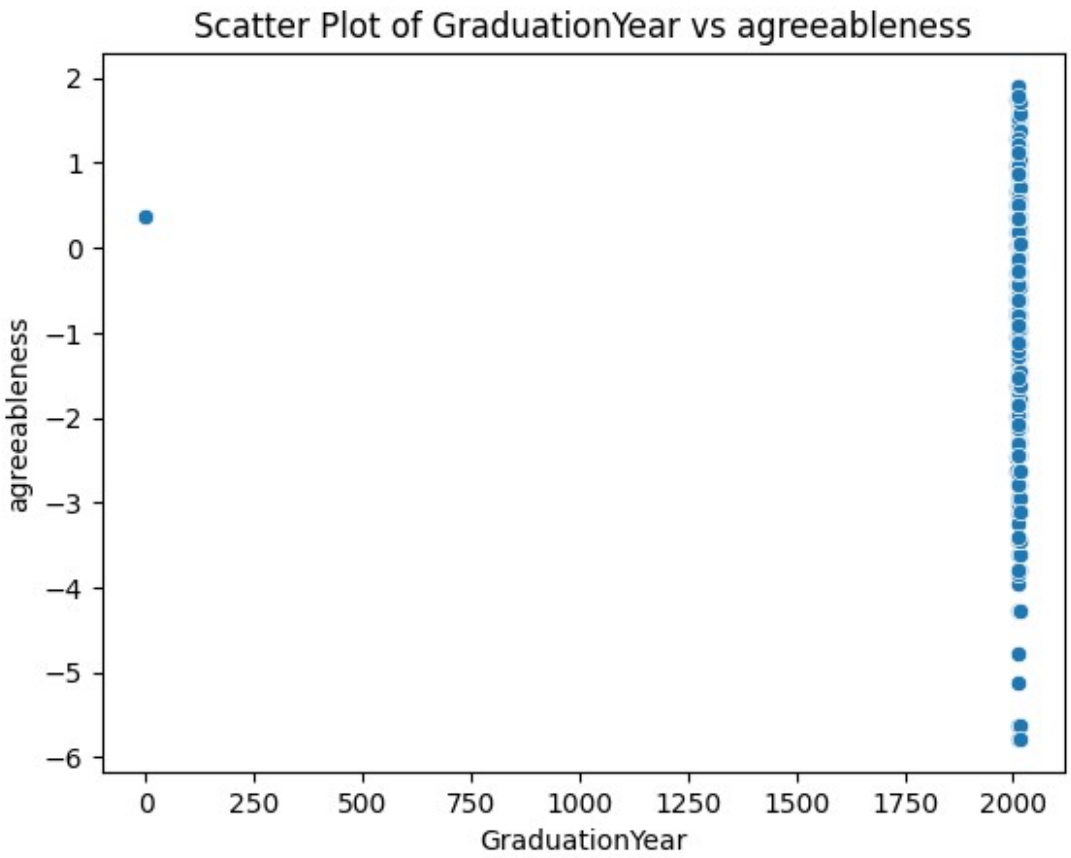

Scatter Plot of GraduationYear vs Domain
Domain
GraduationYear
1.0
0.8
0.6
0.4
0.2
0.0
0
250
500
750
1000
1250
1500
1750
2000

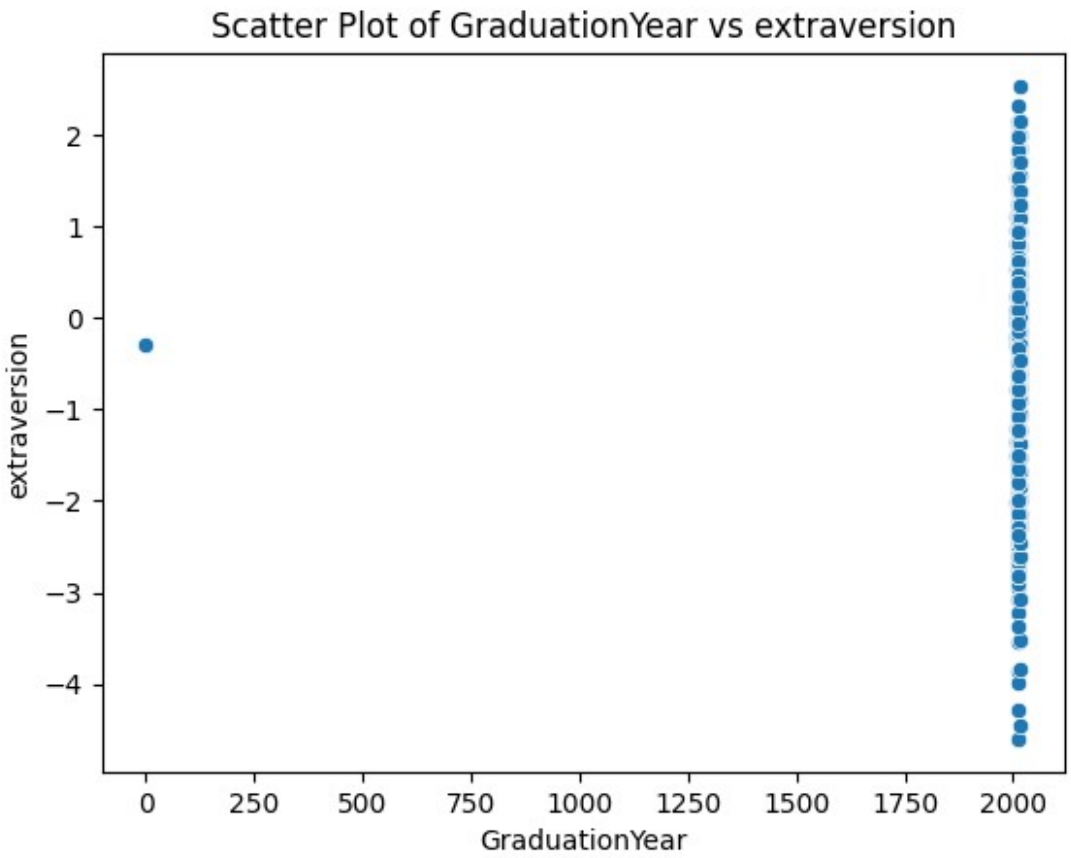

Scatter Plot of GraduationYear vs ComputerProgramming
ComputerProgramming
800
700
600
500
400
300
200
100
0
250
500
750
1000
1250
1500
1750
2000
GraduationYear

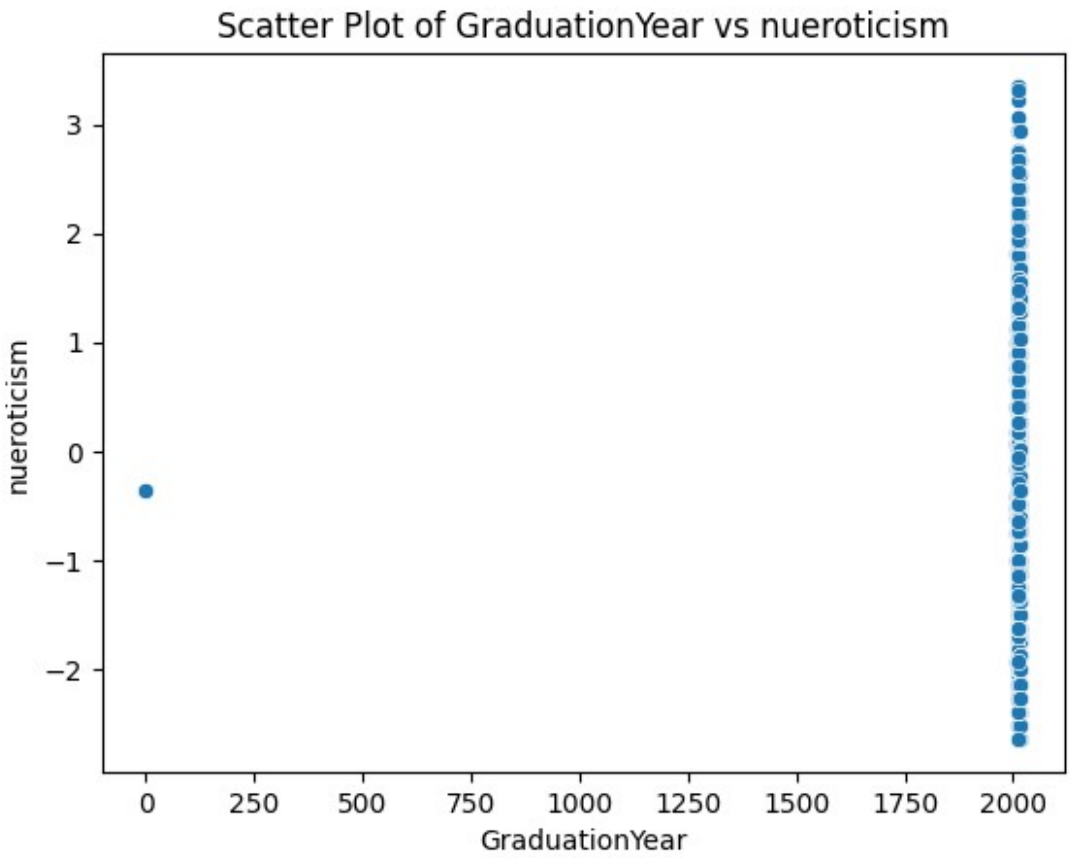

Scatter Plot of GraduationYear vs ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200
0
250
500
750
1000
1250
1500
1750
2000
GraduationYear

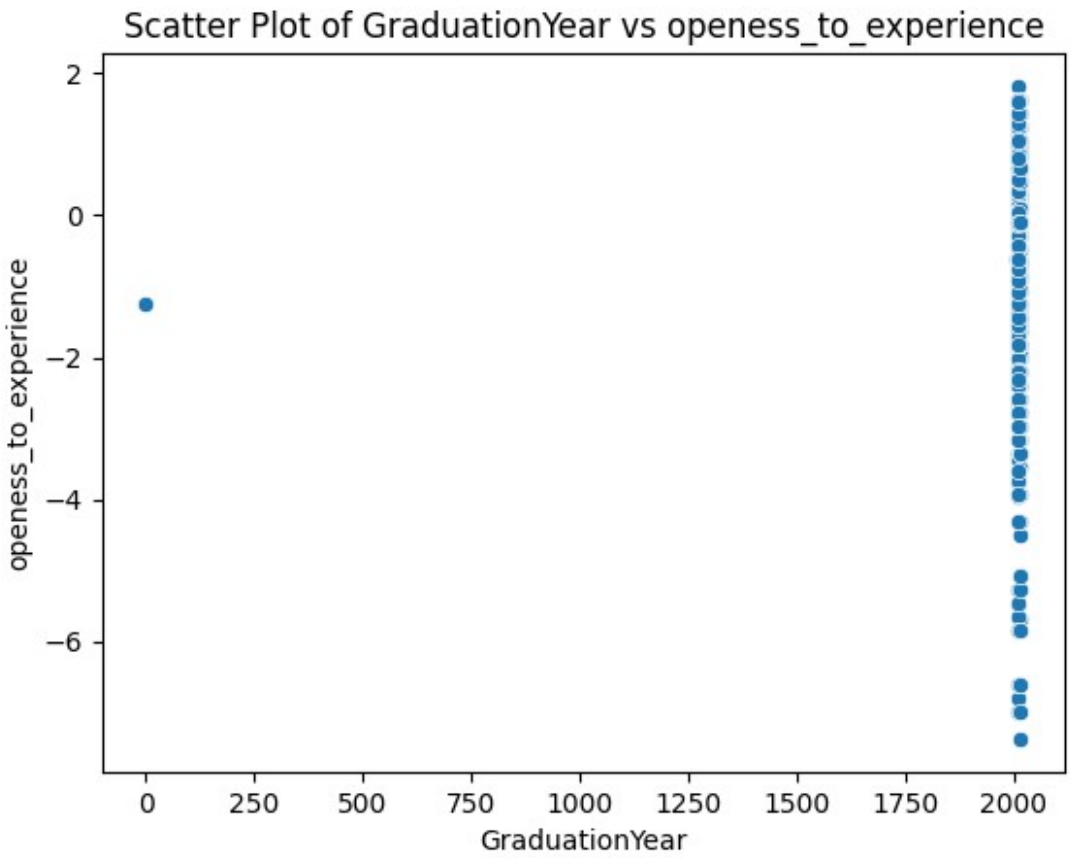

Scatter Plot of GraduationYear vs ComputerScience
ComputerScience
700
600
500
400
300
200
0
250
500
750
1000
1250
1500
1750
2000
GraduationYear

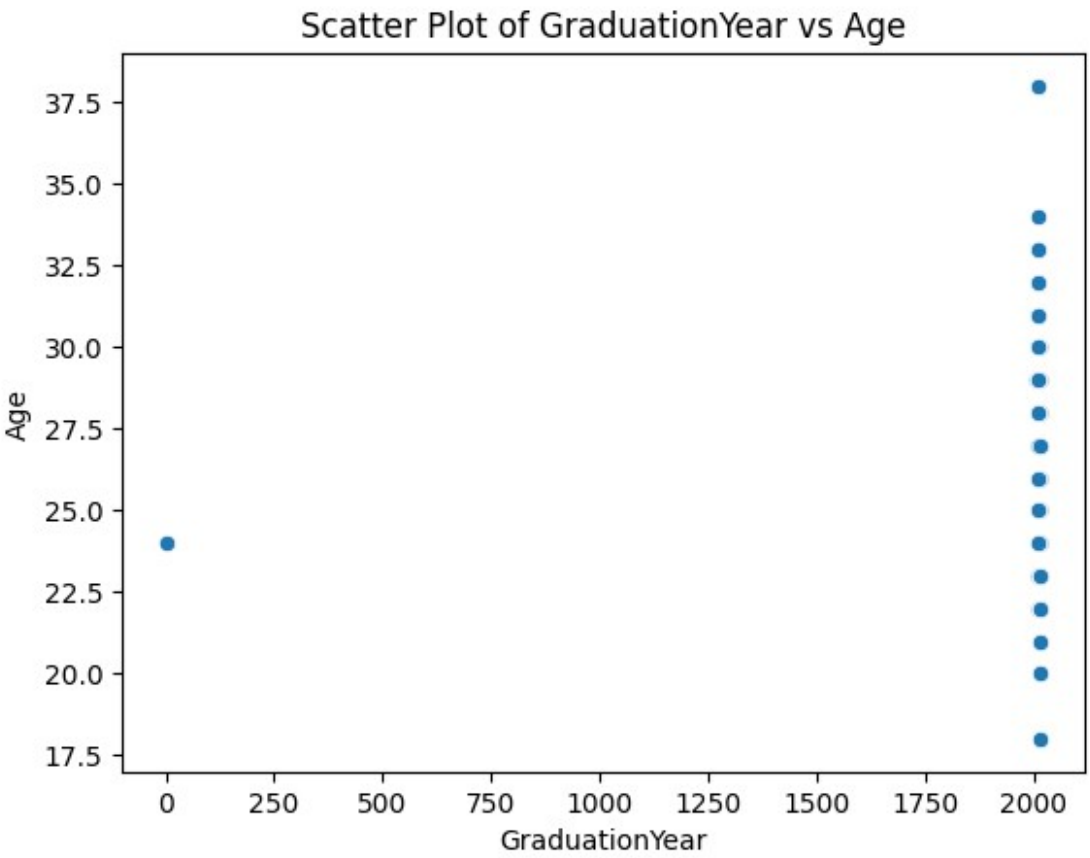

Scatter Plot of GraduationYear vs conscientiousness
conscientiousness
GraduationYear
2
1
0
−1
−2
−3
−4
0
250
500
750
1000
1250
1500
1750
2000

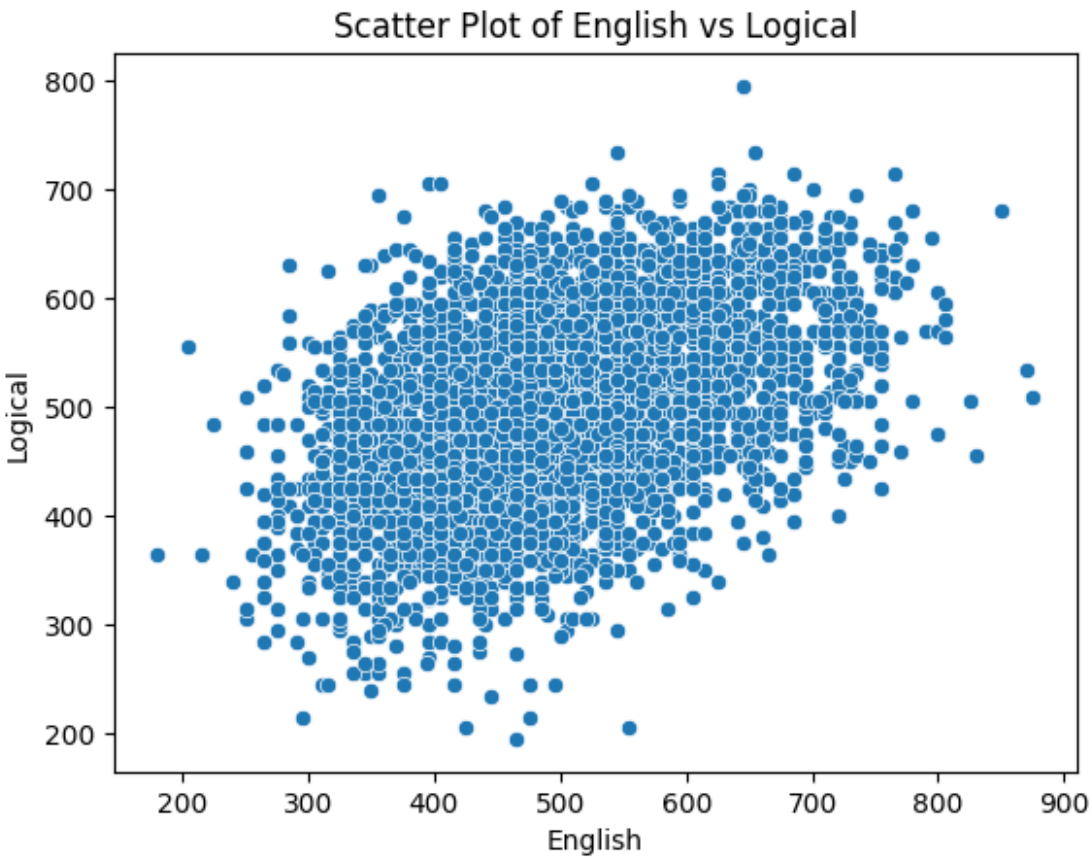

Scatter Plot of GraduationYear vs agreeableness
agreeableness
GraduationYear
2
1
0
−1
−2
−3
−4
−5
−6
0
250
500
750
1000
1250
1500
1750
2000

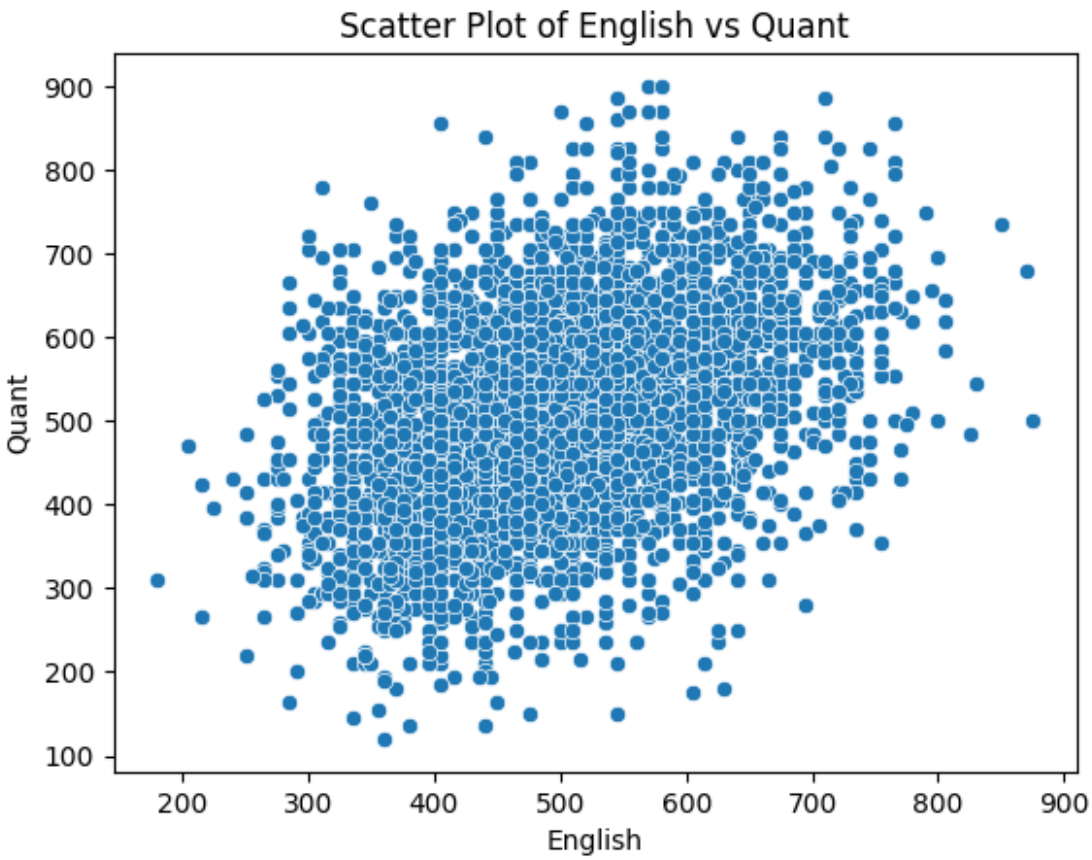

Scatter Plot of GraduationYear vs extraversion
extraversion
GraduationYear
2
1
0
−1
−2
−3
−4
0
250
500
750
1000
1250
1500
1750
2000

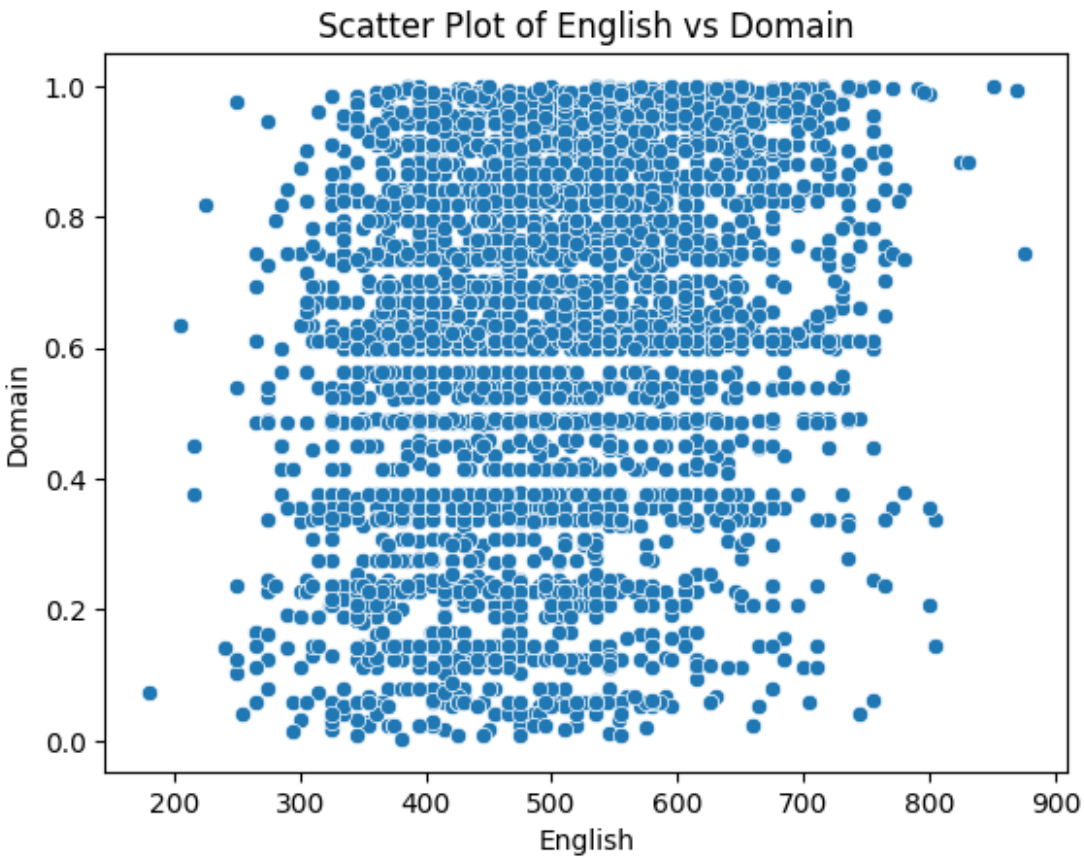

Scatter Plot of GraduationYear vs nueroticism
nueroticism
3
2
1
0
−1
−2
0
250
500
750
1000
1250
1500
1750
2000
GraduationYear

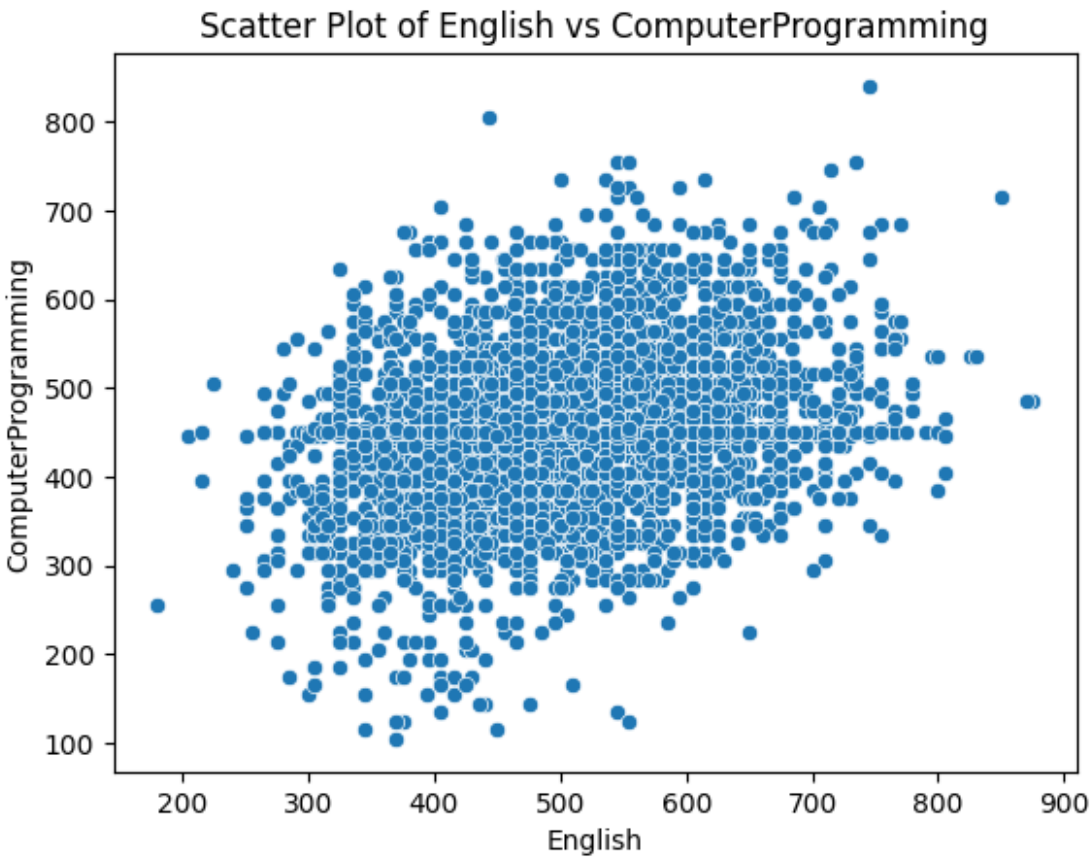

Scatter Plot of GraduationYear vs openess_to_experience
openess_to_experience
2
0
−2
−4
−6
0
250
500
750
1000
1250
1500
1750
2000
GraduationYear

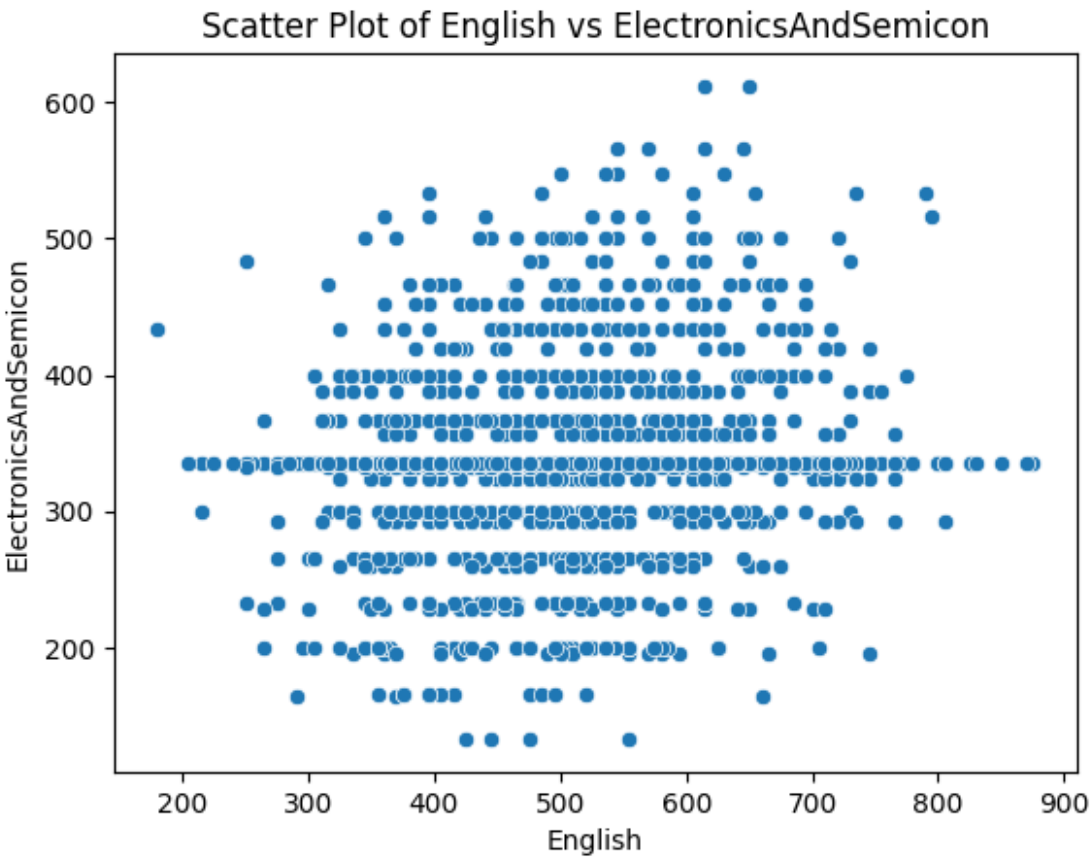

Scatter Plot of GraduationYear vs Age
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
0
250
500
750
1000
1250
1500
1750
2000
GraduationYear

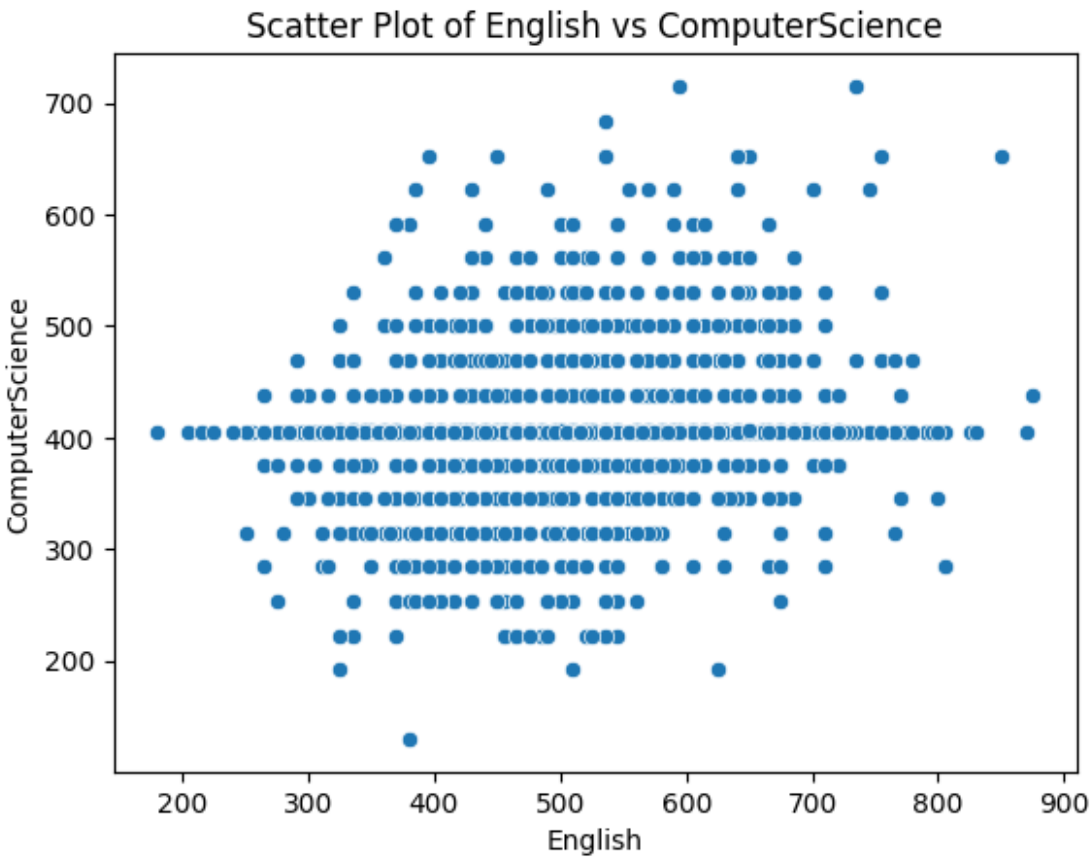

Scatter Plot of English vs Logical
Logical
English
800
700
600
500
400
300
200
200
300
400
500
600
700
800
900

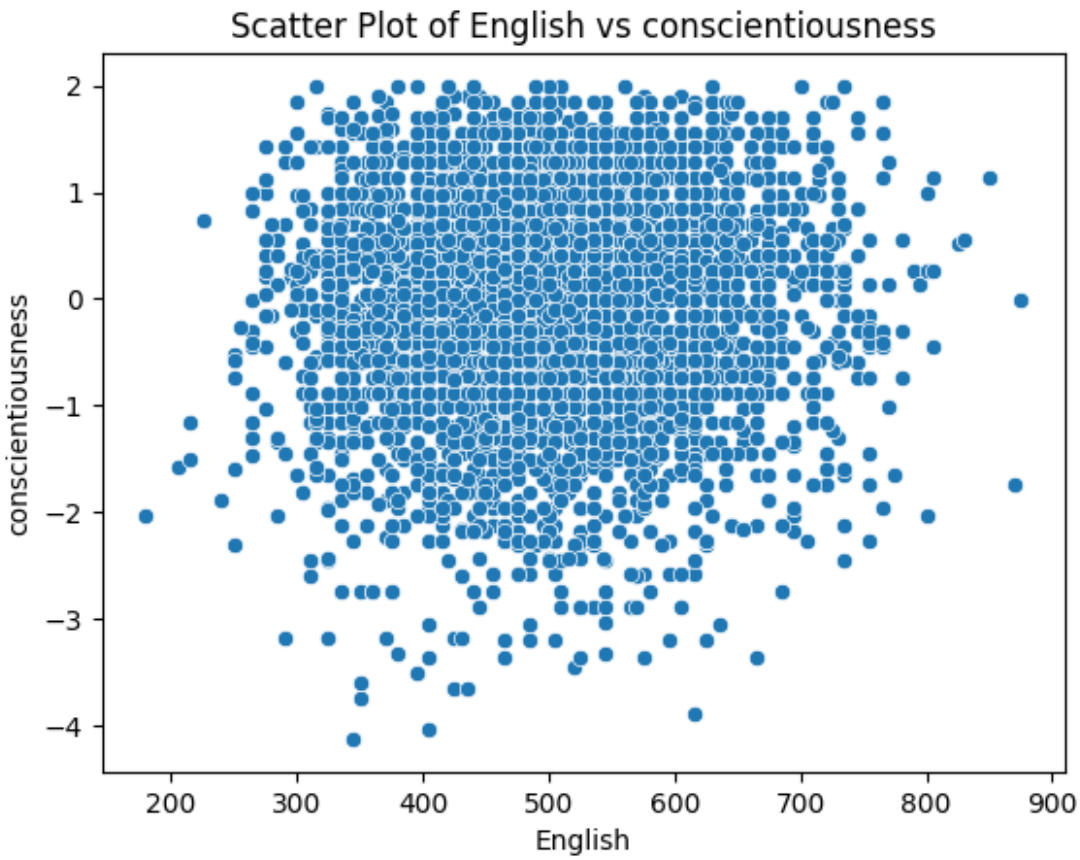

Scatter Plot of English vs Quant
Quant
English
900
800
700
600
500
400
300
200
100
200
300
400
500
600
700
800
900

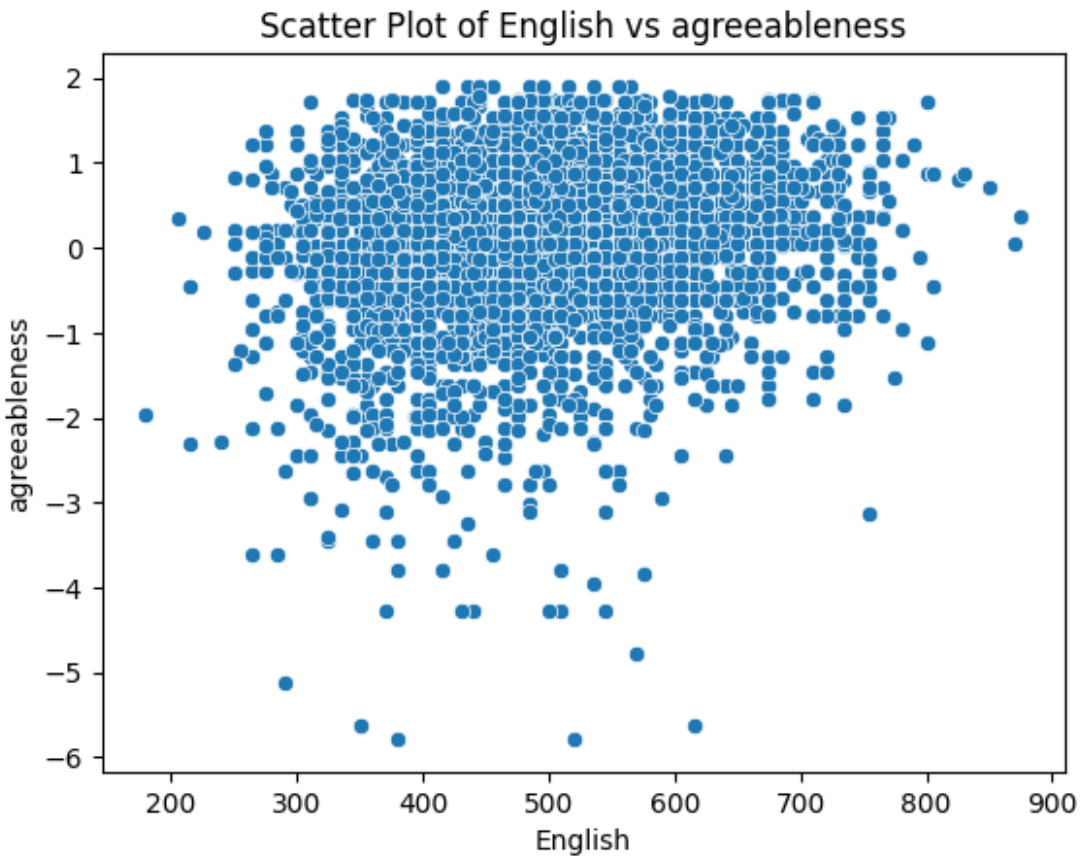

Scatter Plot of English vs Domain
Domain
English
1.0
0.8
0.6
0.4
0.2
0.0
200
300
400
500
600
700
800
900

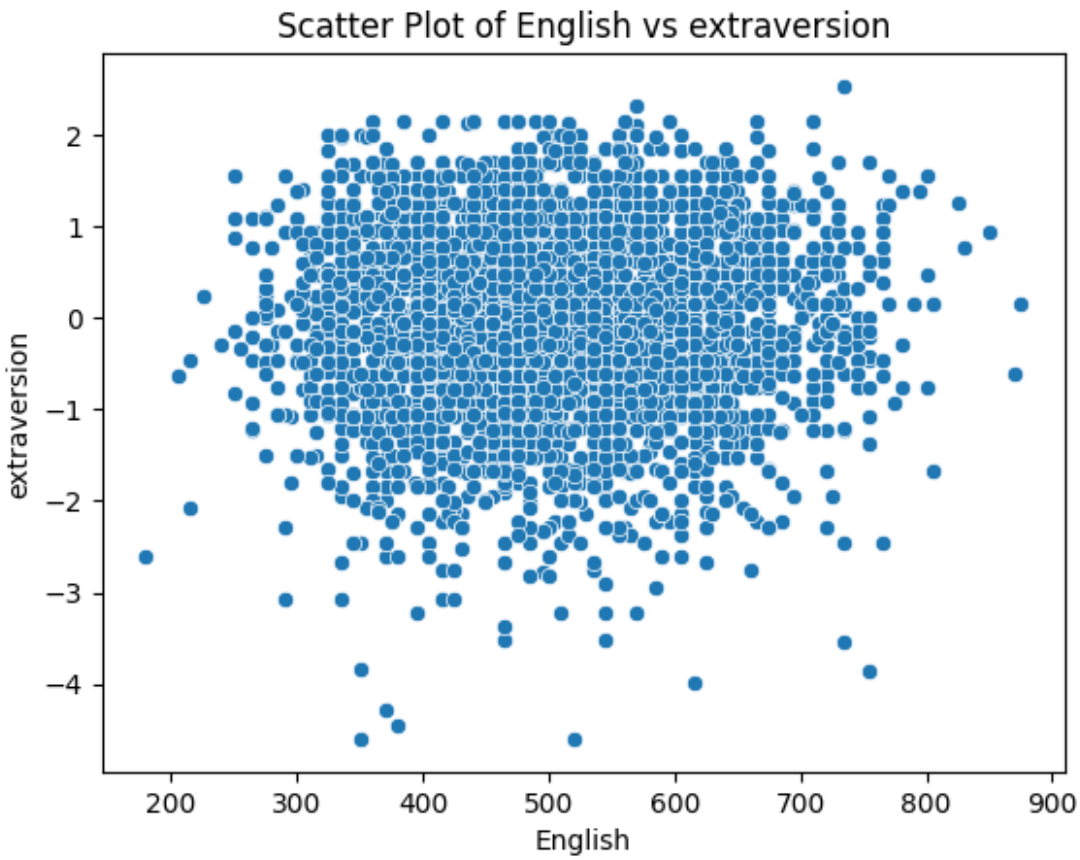

Scatter Plot of English vs ComputerProgramming
ComputerProgramming
English
100
200
300
400
500
600
700
800
200
300
400
500
600
700
800
900

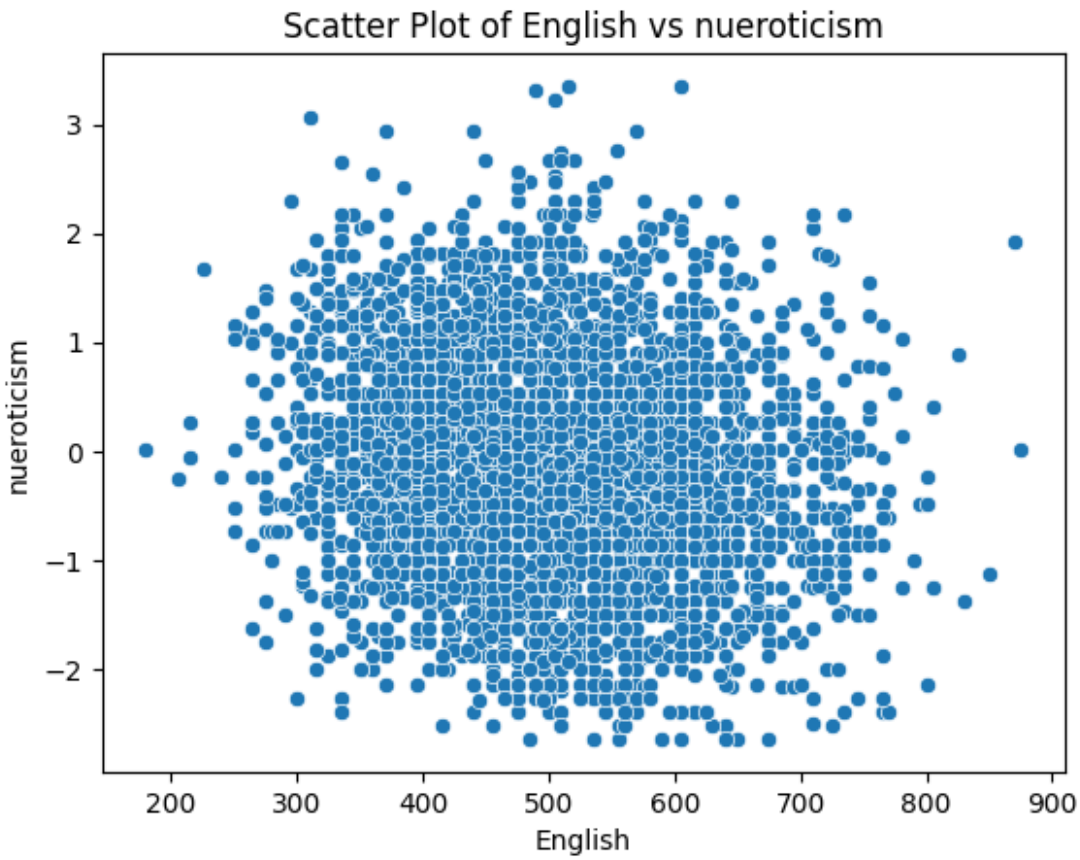

Scatter Plot of English vs ElectronicsAndSemicon
ElectronicsAndSemicon
English
600
500
400
300
200
200
300
400
500
600
700
800
900

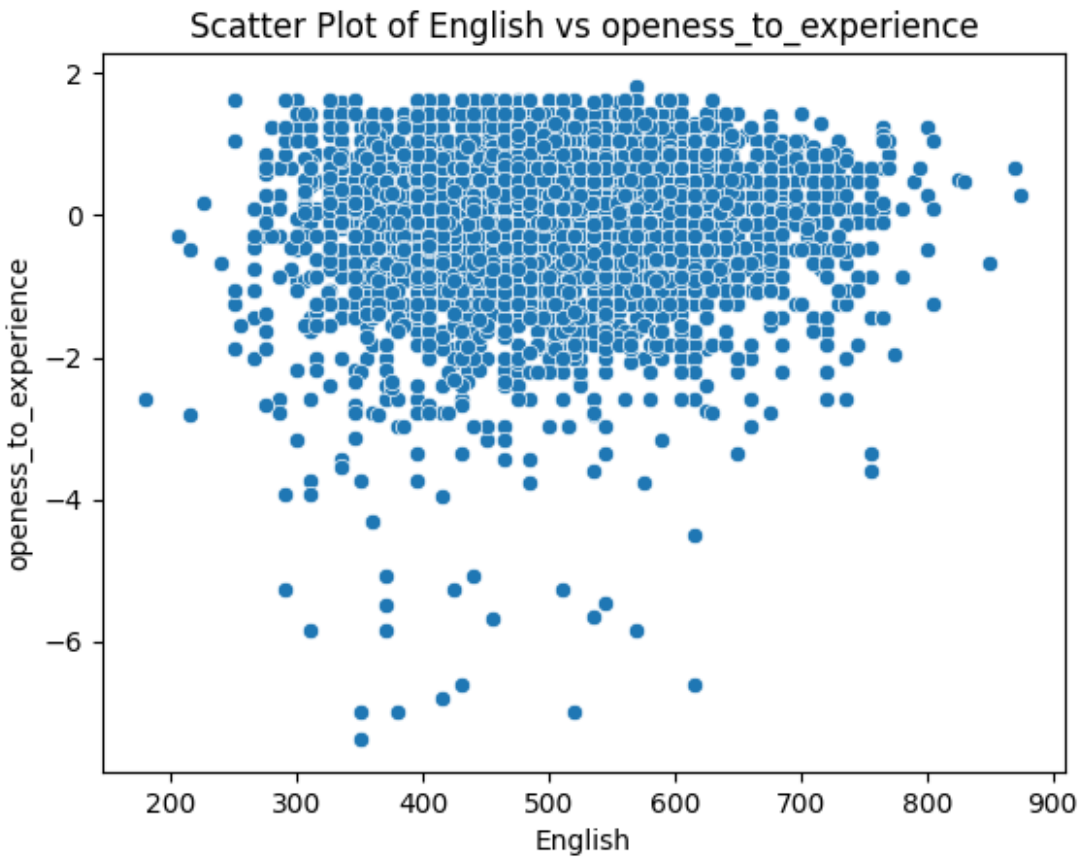

Scatter Plot of English vs ComputerScience
ComputerScience
English
700
600
500
400
300
200
200
300
400
500
600
700
800
900

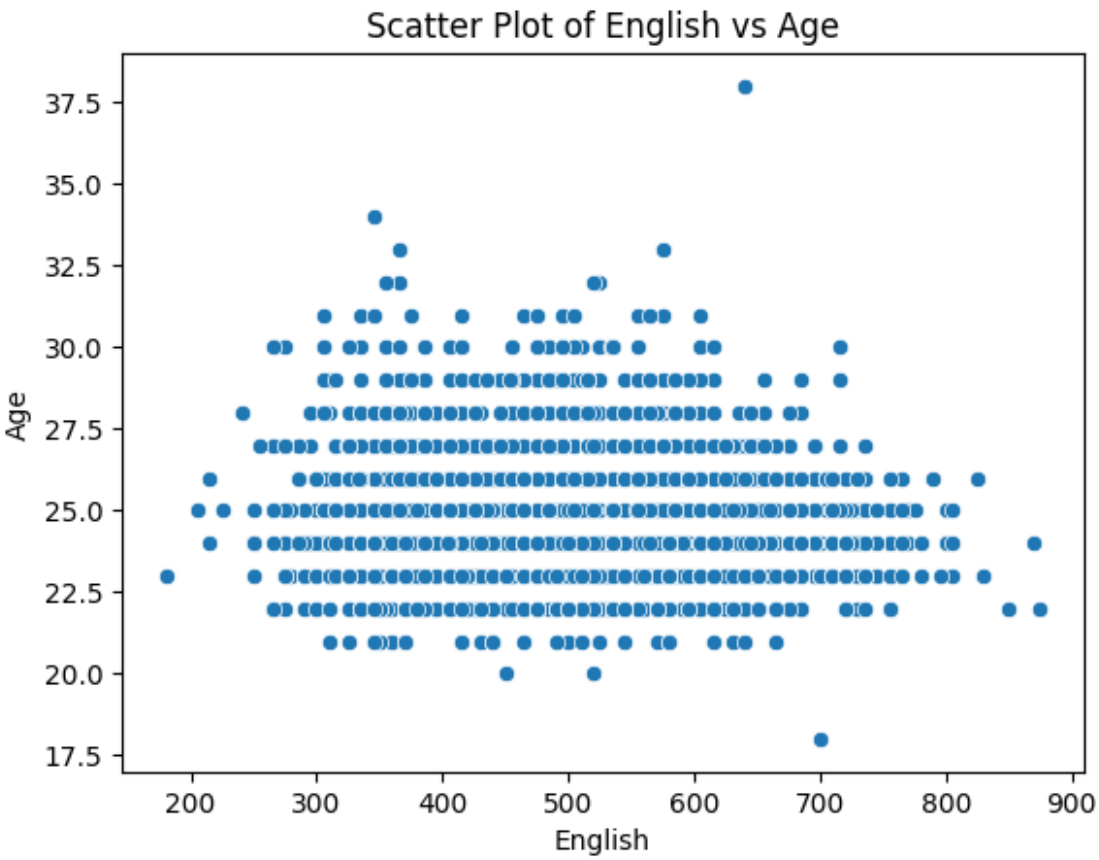

Scatter Plot of English vs conscientiousness
conscientiousness
English
2
1
0
−1
−2
−3
−4
200
300
400
500
600
700
800
900

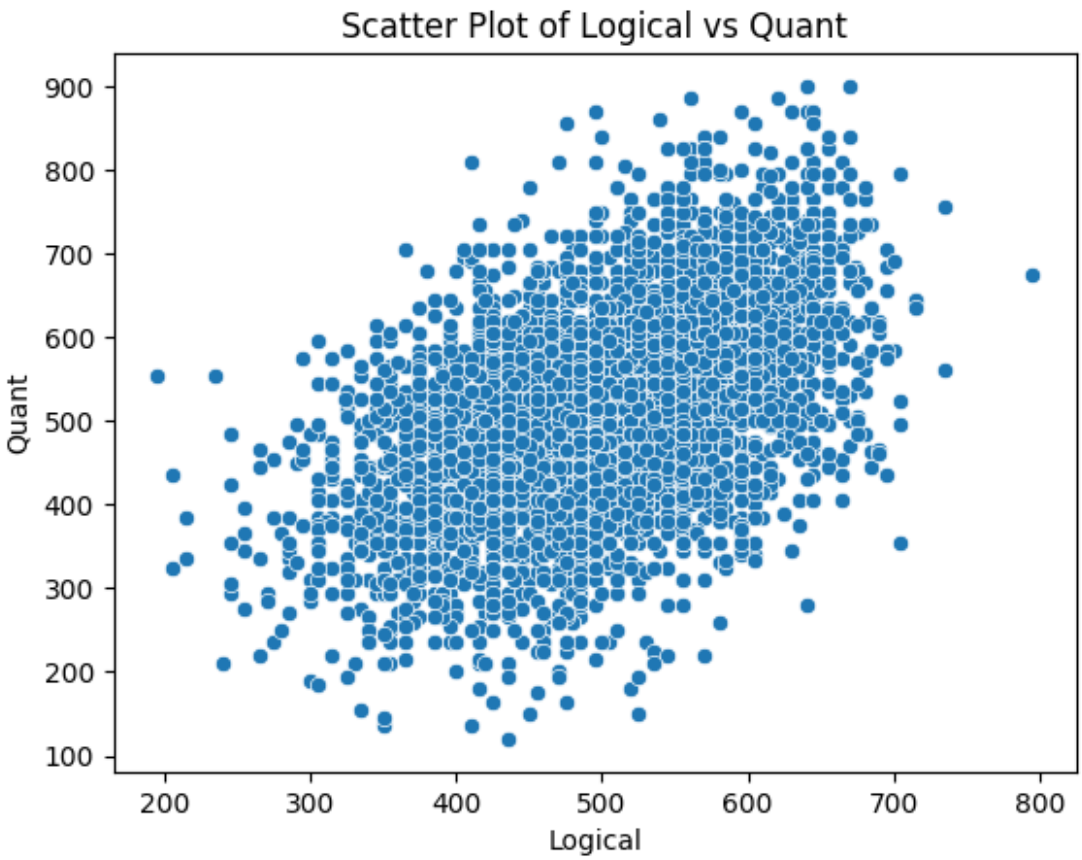

Scatter Plot of English vs agreeableness
agreeableness
English
2
1
0
−1
−2
−3
−4
−5
−6
200
300
400
500
600
700
800
900

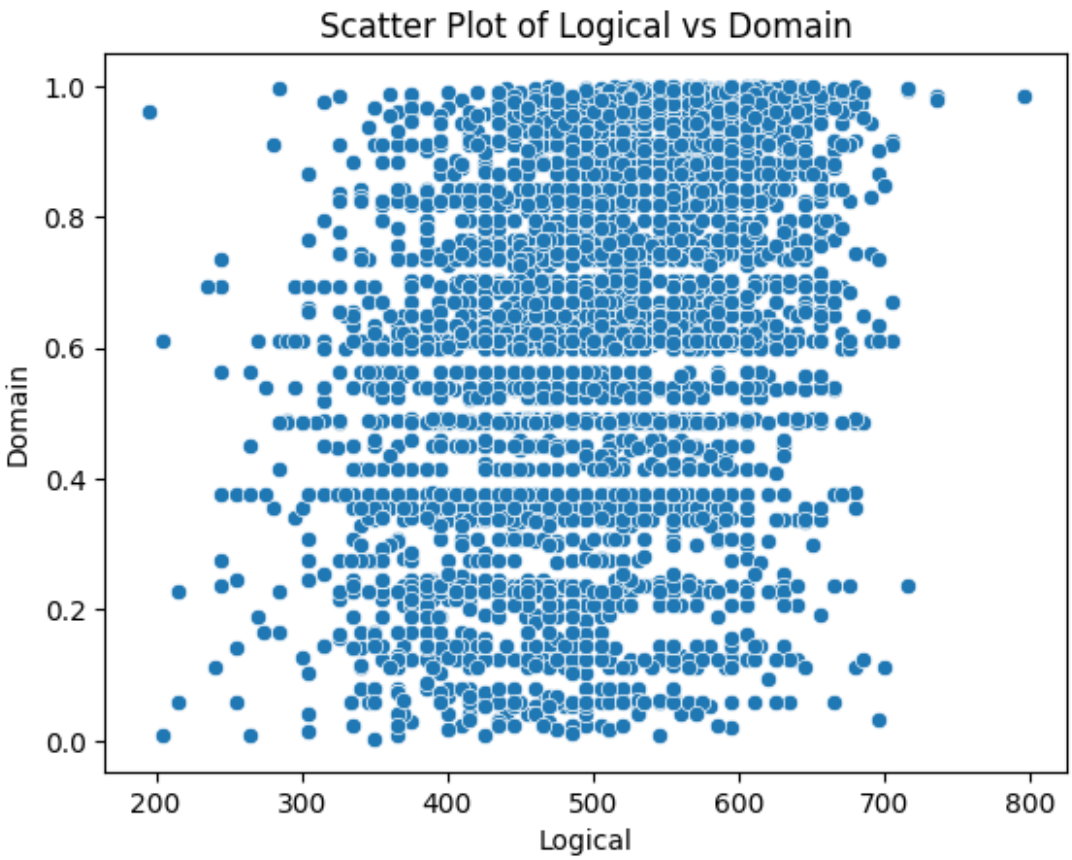

Scatter Plot of English vs extraversion
extraversion
English
2
1
0
−1
−2
−3
−4
200
300
400
500
600
700
800
900

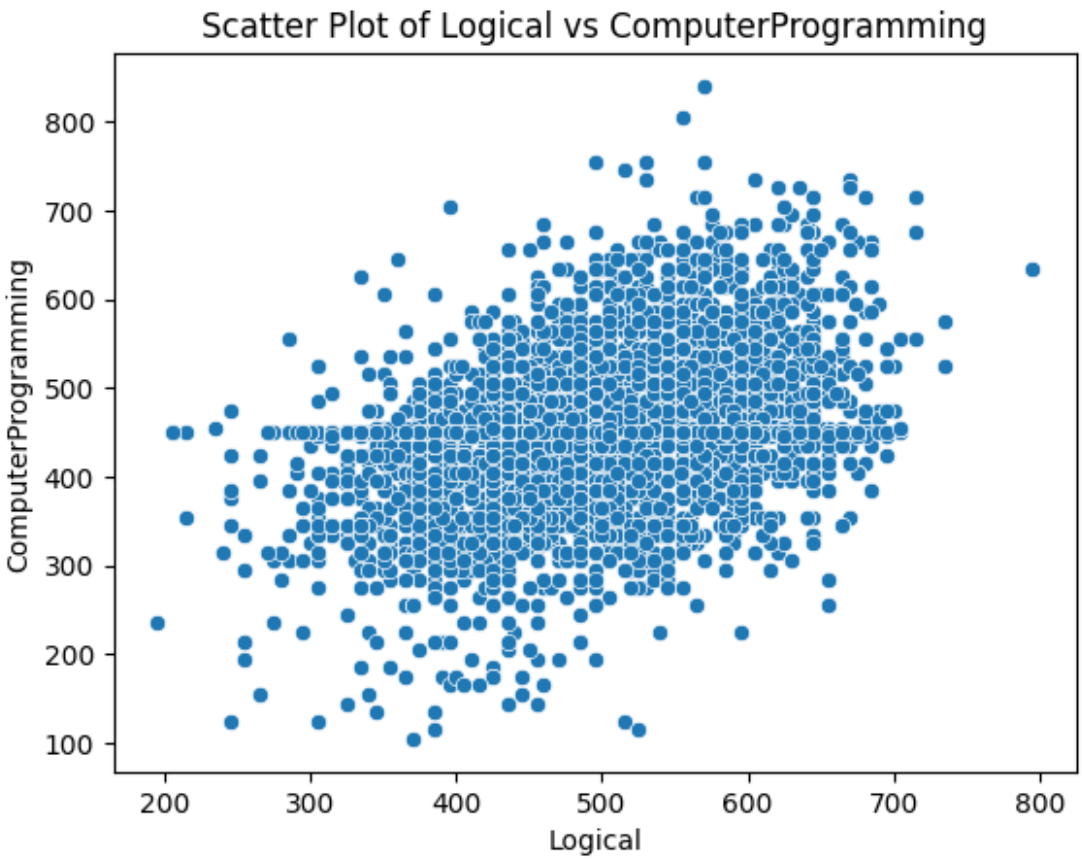

Scatter Plot of English vs nueroticism
nueroticism
English
3
2
1
0
−1
−2
200
300
400
500
600
700
800
900

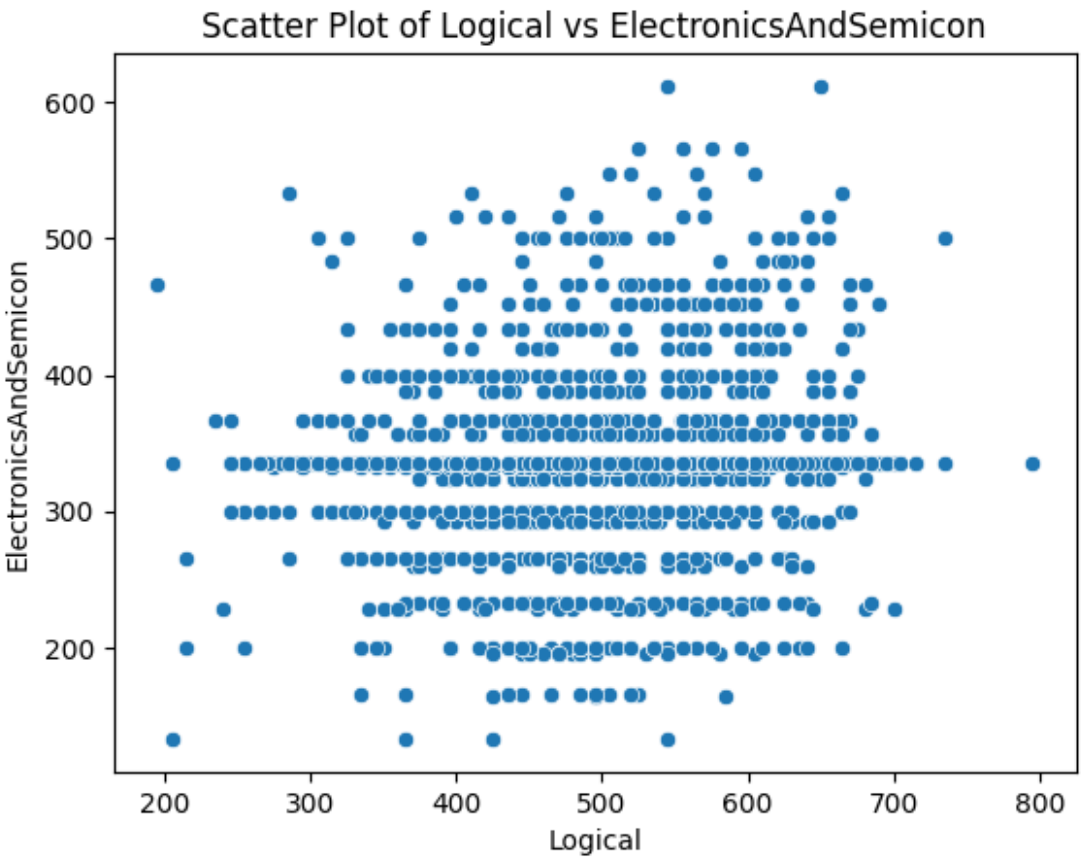

Scatter Plot of English vs openess_to_experience
openess_to_experience
English
2
0
−2
−4
−6
200
300
400
500
600
700
800
900

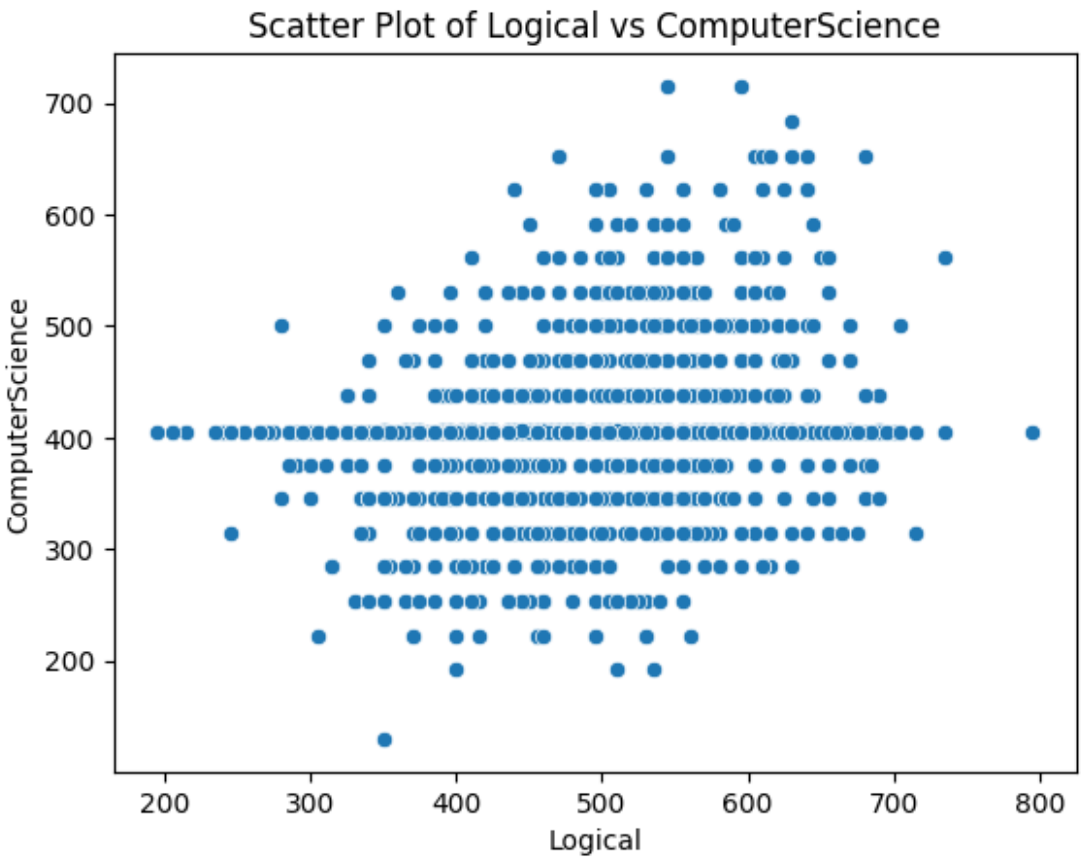

Scatter Plot of English vs Age
Age
English
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
200
300
400
500
600
700
800
900


Scatter Plot of Logical vs Quant
Quant
Logical
900
800
700
600
500
400
300
200
100
200
300
400
500
600
700
800

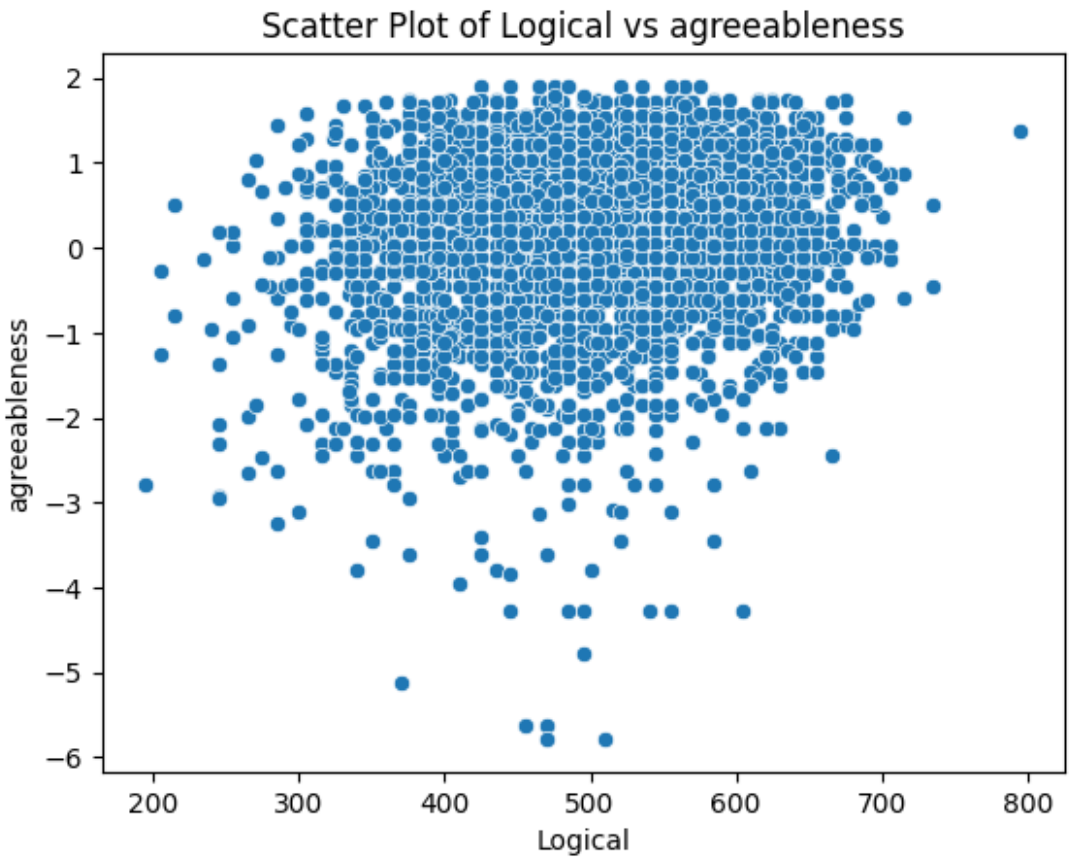

Scatter Plot of Logical vs Domain
Domain
Logical
1.0
0.8
0.6
0.4
0.2
0.0
200
300
400
500
600
700
800

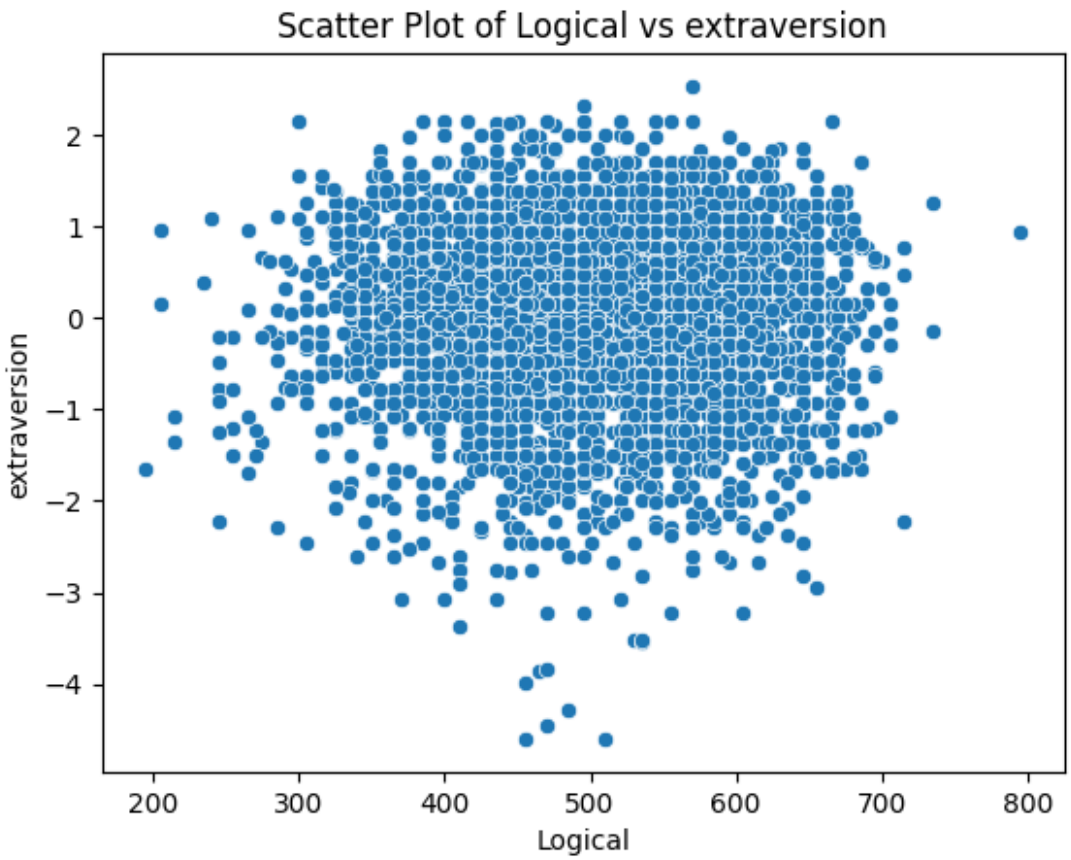

Scatter Plot of Logical vs ComputerProgramming
ComputerProgramming
Logical
100
200
300
400
500
600
700
800
200
300
400
500
600
700
800

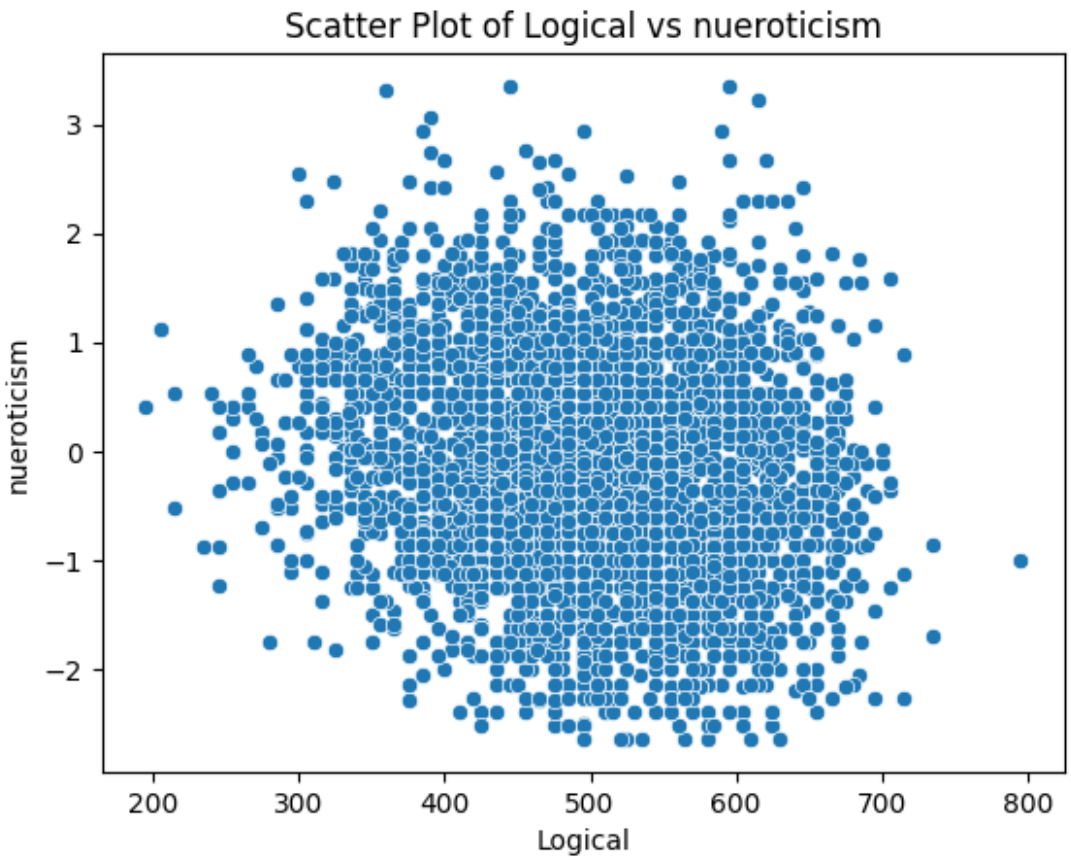

Scatter Plot of Logical vs ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200
200
300
400
500
600
700
800
Logical

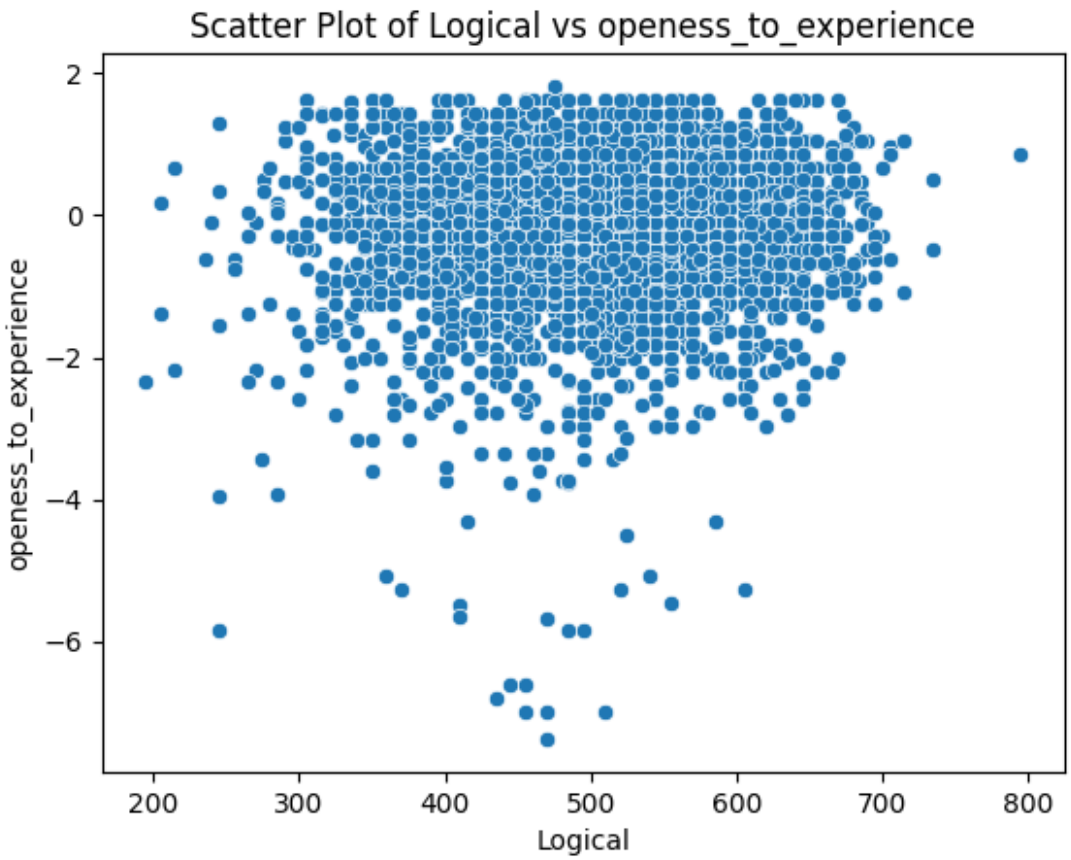

Scatter Plot of Logical vs ComputerScience
ComputerScience
Logical
700
600
500
400
300
200
200
300
400
500
600
700
800

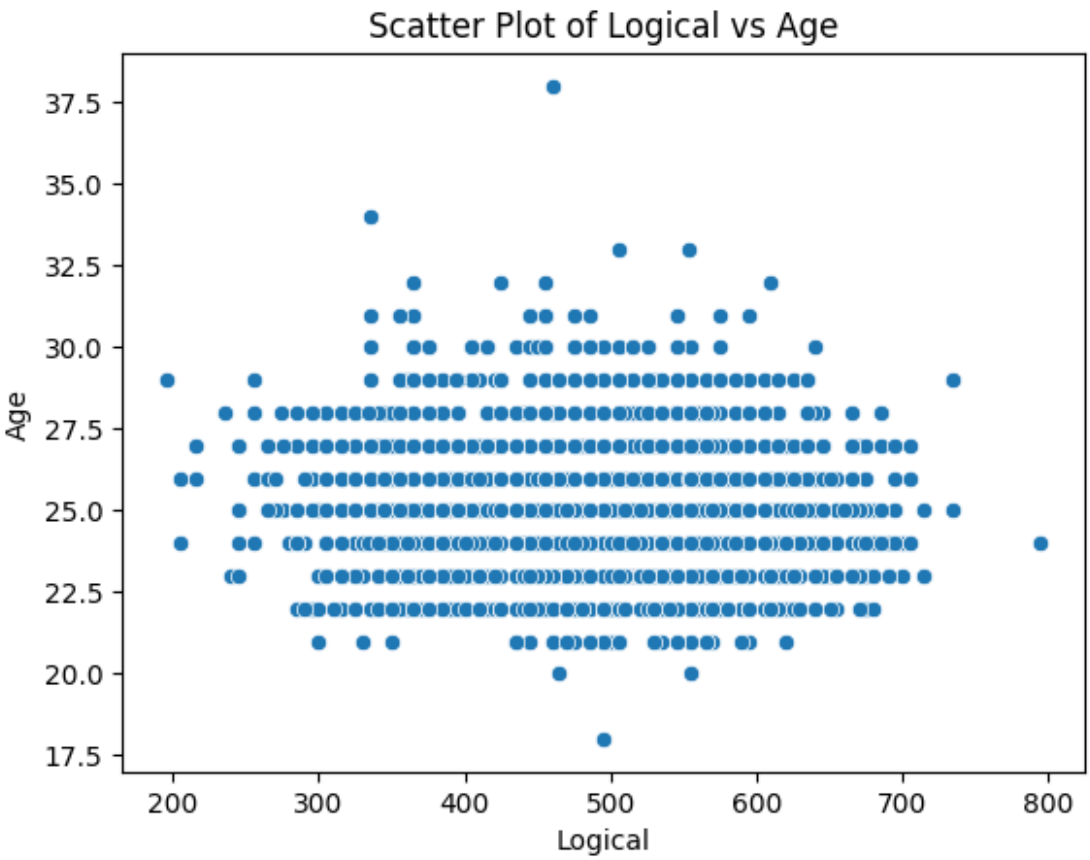
Scatter Plot of Logical vs conscientiousness
conscientiousness
Logical
2
1
0
−1
−2
−3
−4
200
300
400
500
600
700
800

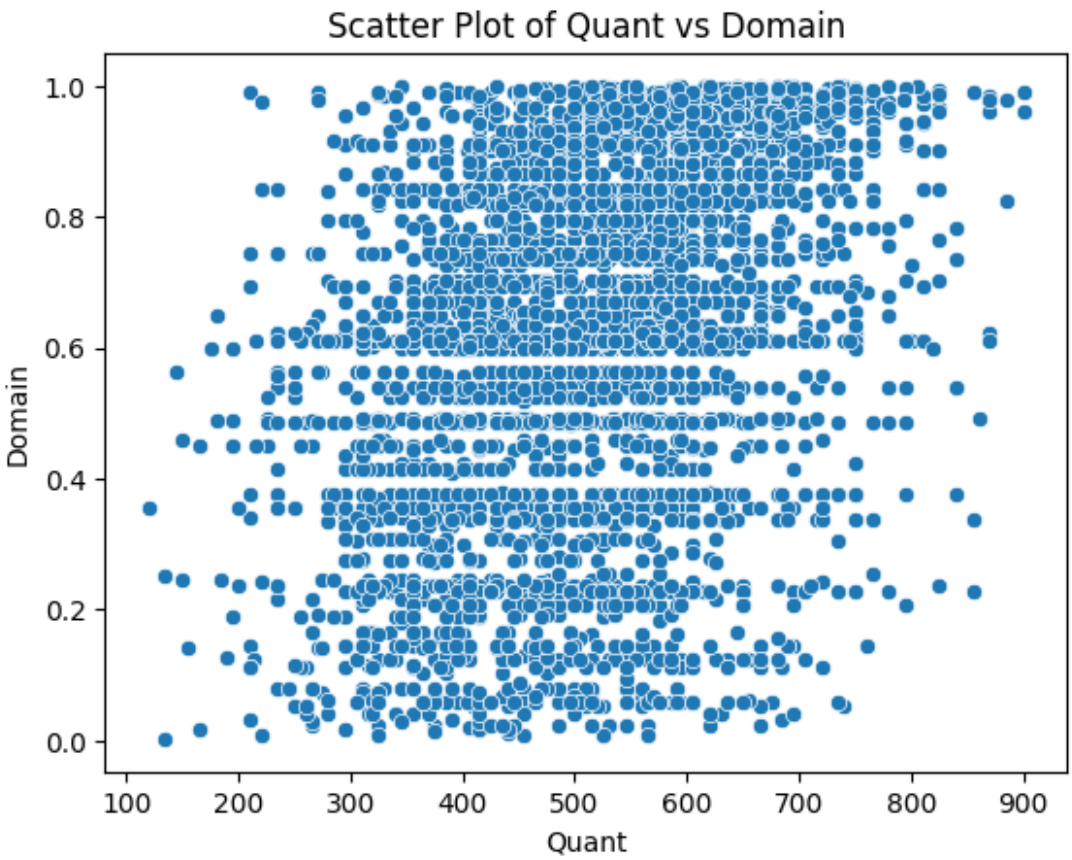

Scatter Plot of Logical vs agreeableness
agreeableness
Logical
2
1
0
−1
−2
−3
−4
−5
−6
200
300
400
500
600
700
800

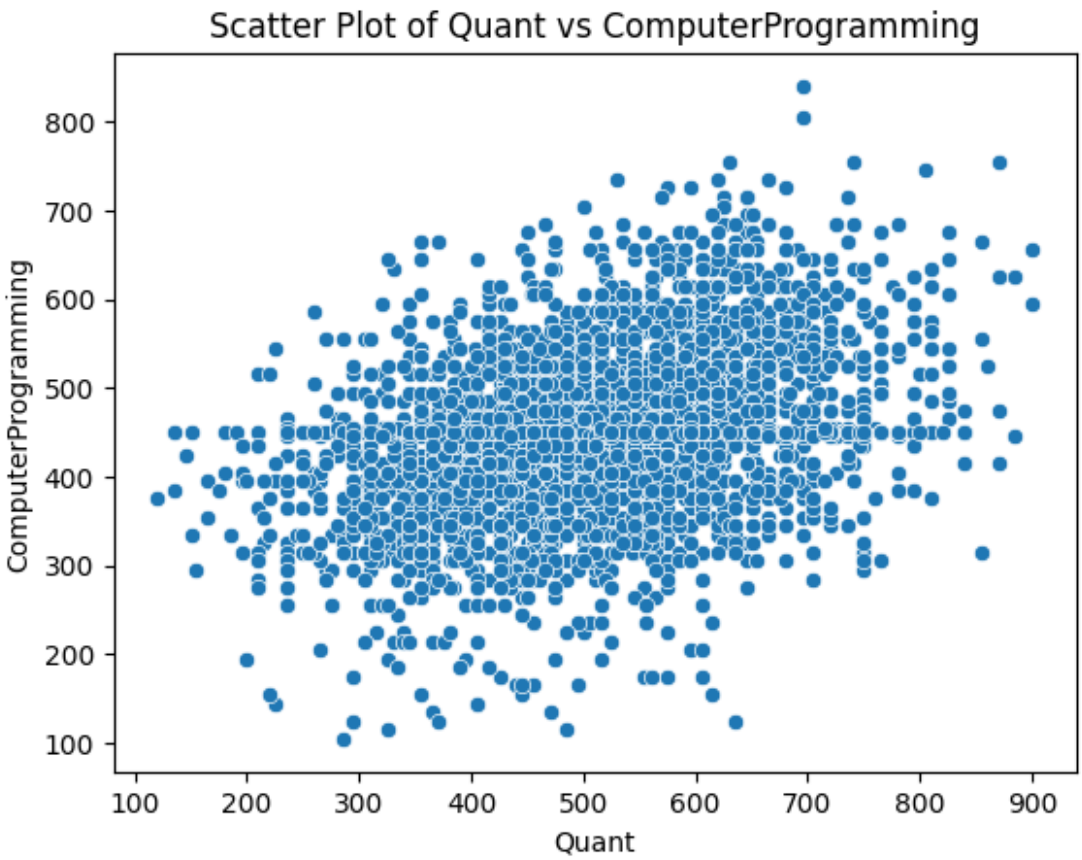

Scatter Plot of Logical vs extraversion
extraversion
Logical
2
1
0
−1
−2
−3
−4
200
300
400
500
600
700
800

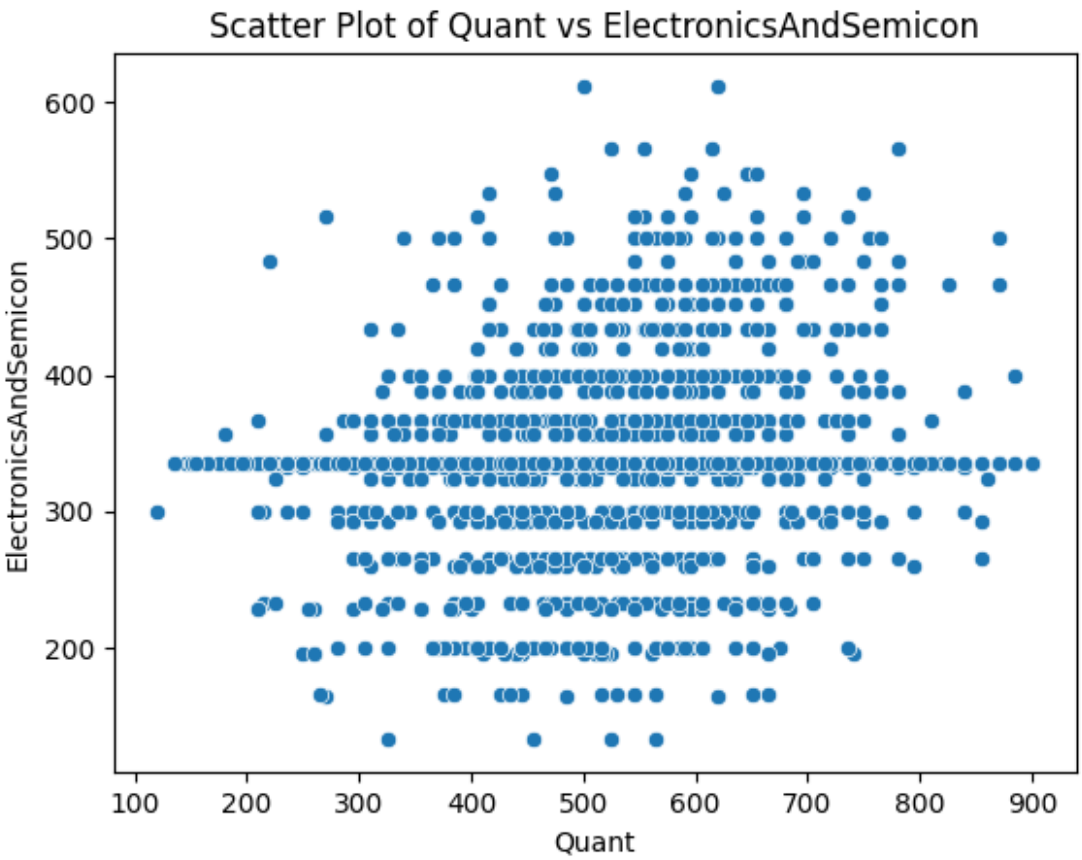

Scatter Plot of Logical vs nueroticism
nueroticism
Logical
3
2
1
0
−1
−2
200
300
400
500
600
700
800

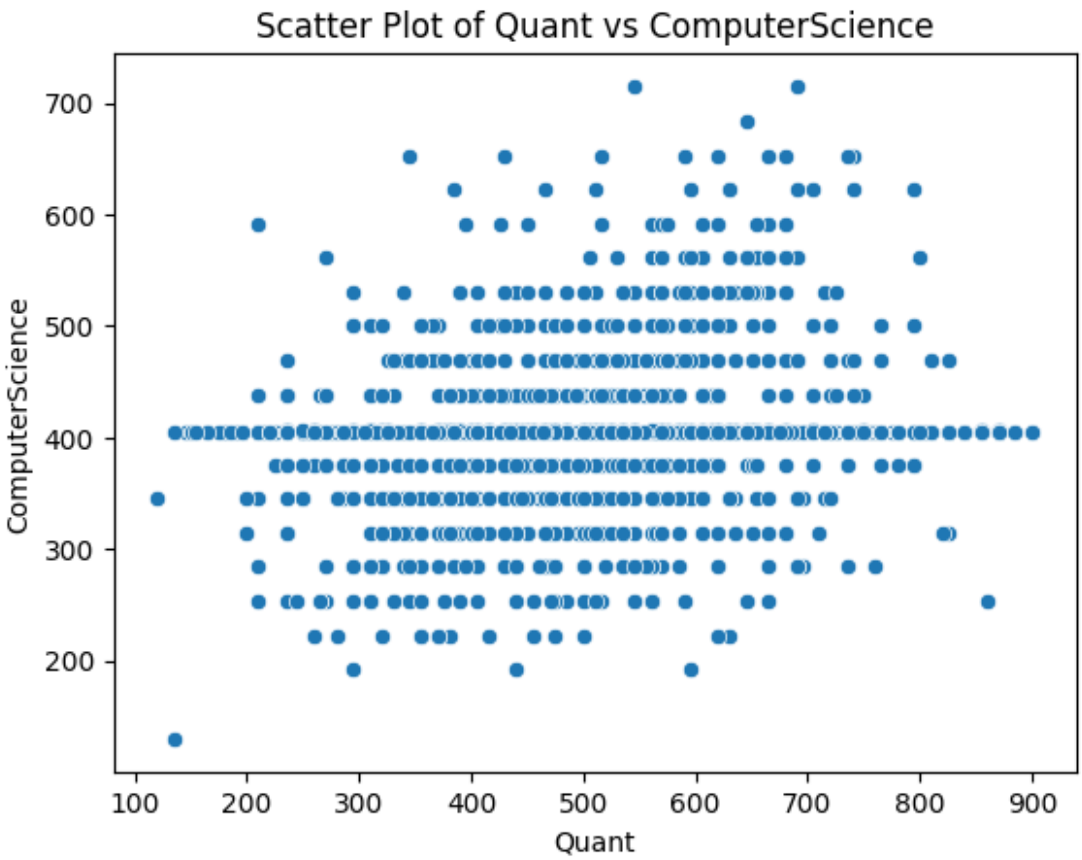

Scatter Plot of Logical vs openess_to_experience
openess_to_experience
Logical
2
0
−2
−4
−6
200
300
400
500
600
700
800

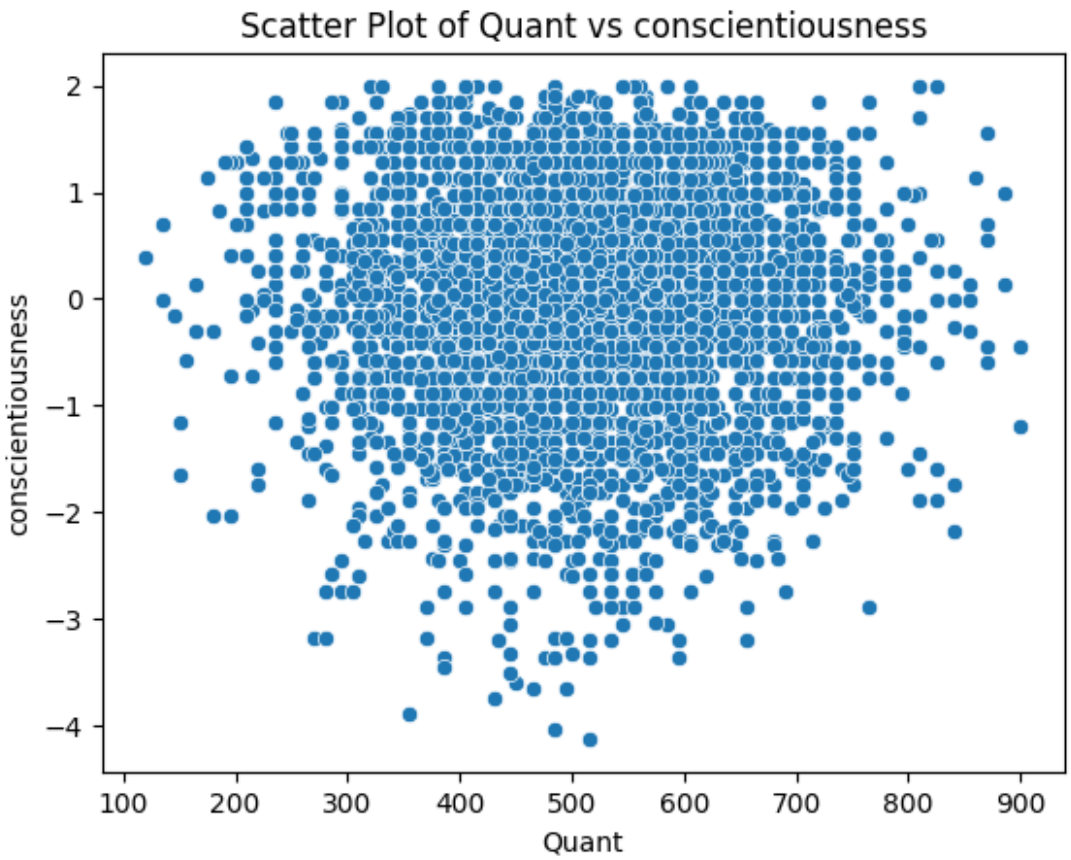

Scatter Plot of Logical vs Age
Age
Logical
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
200
300
400
500
600
700
800

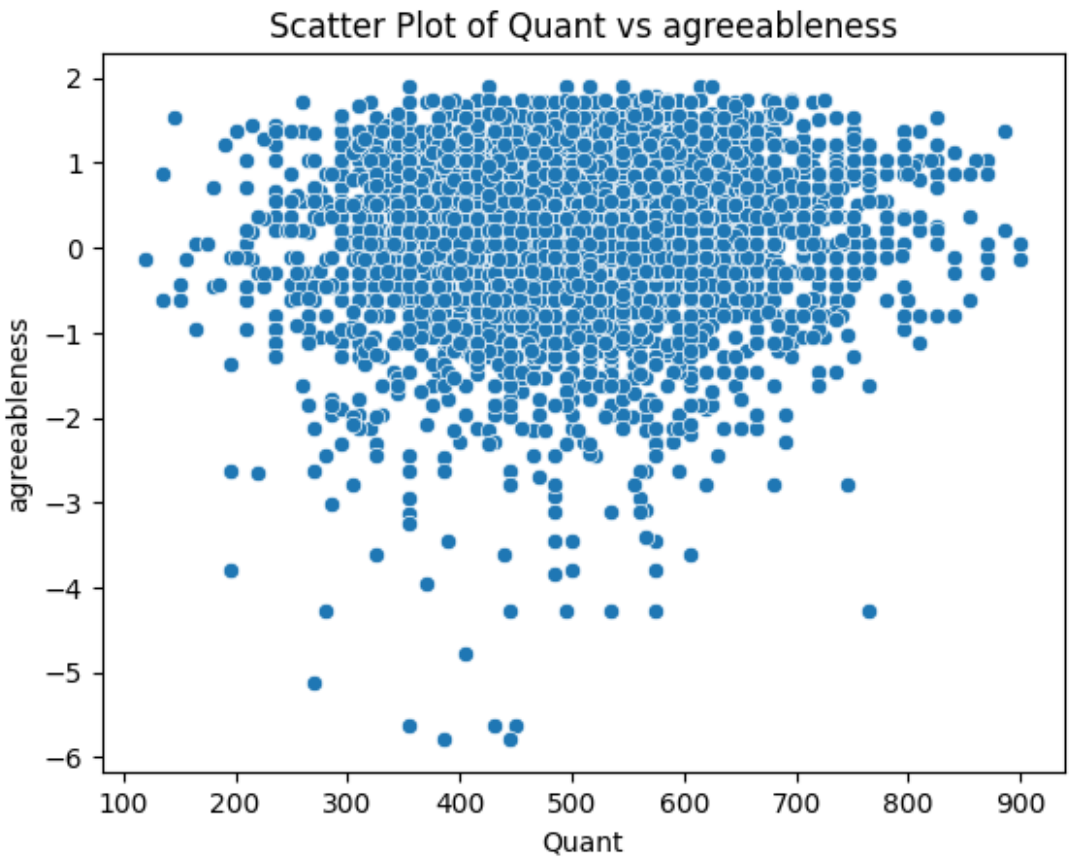

Scatter Plot of Quant vs Domain
Domain
Quant
1.0
0.8
0.6
0.4
0.2
0.0
100
200
300
400
500
600
700
800
900

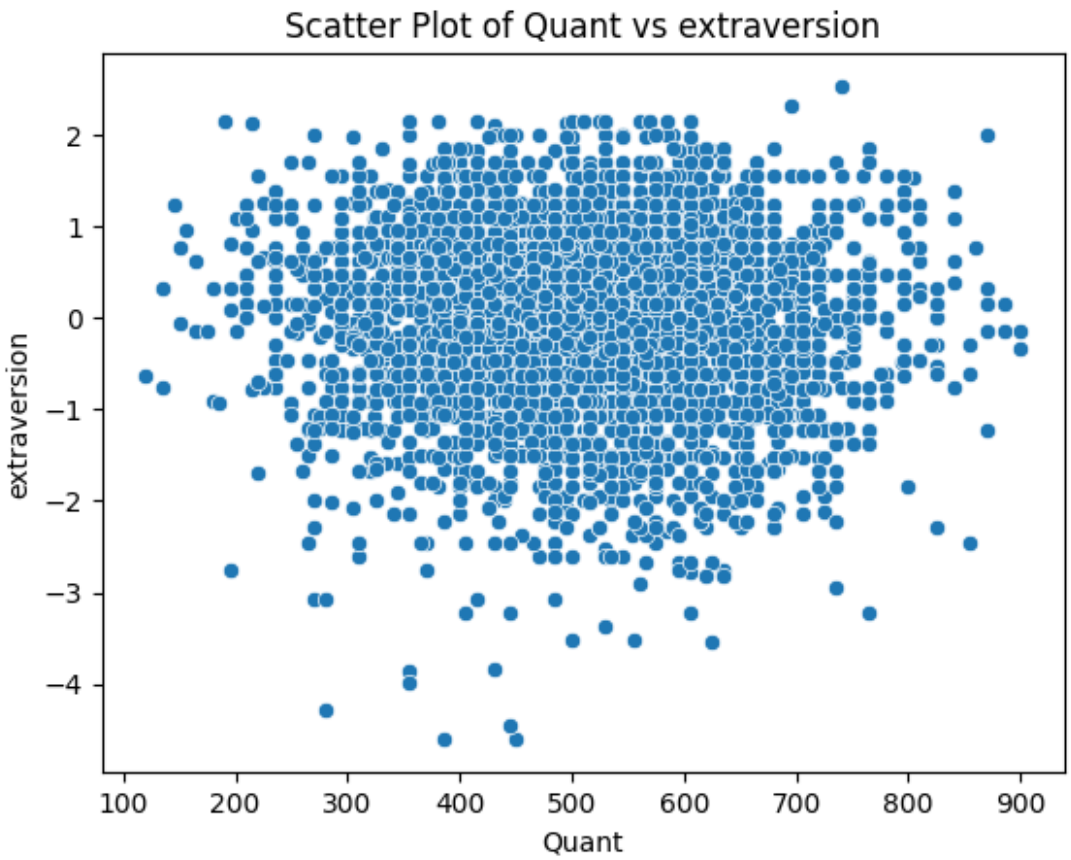

Scatter Plot of Quant vs ComputerProgramming
ComputerProgramming
Quant
100
200
300
400
500
600
700
800
900

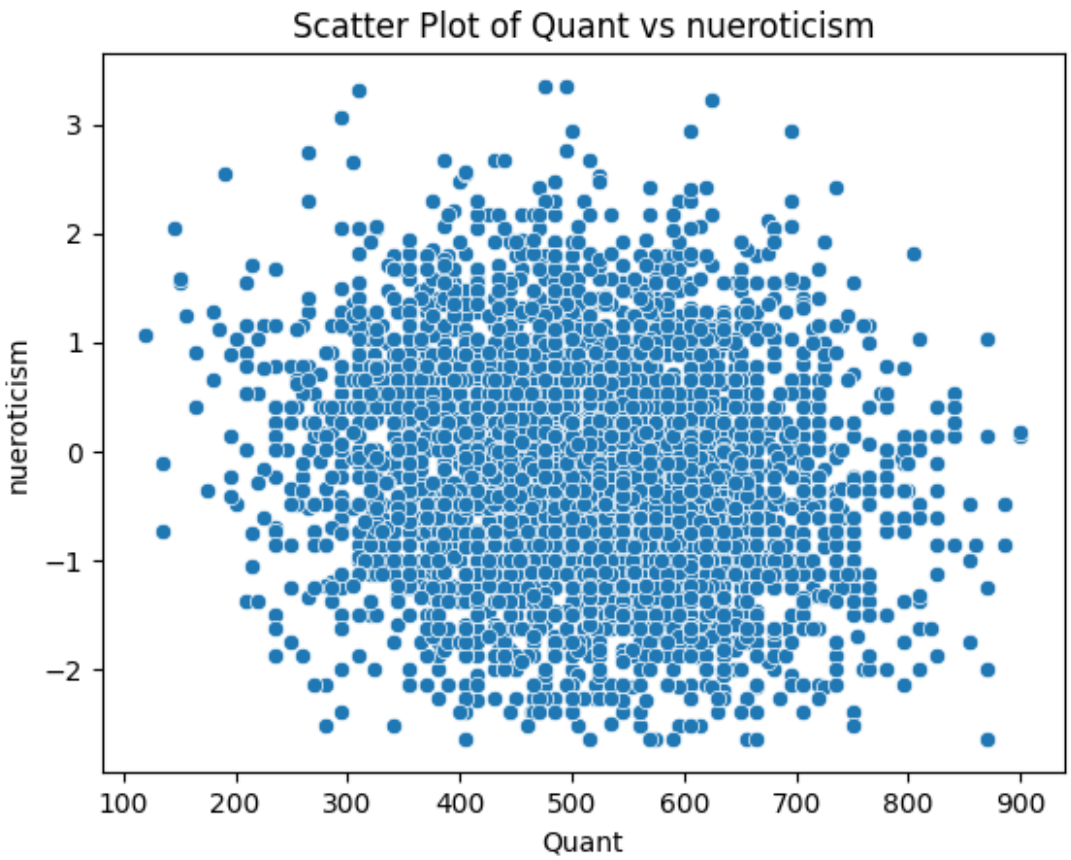

Scatter Plot of Quant vs ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200
100
200
300
400
500
600
700
800
900
Quant

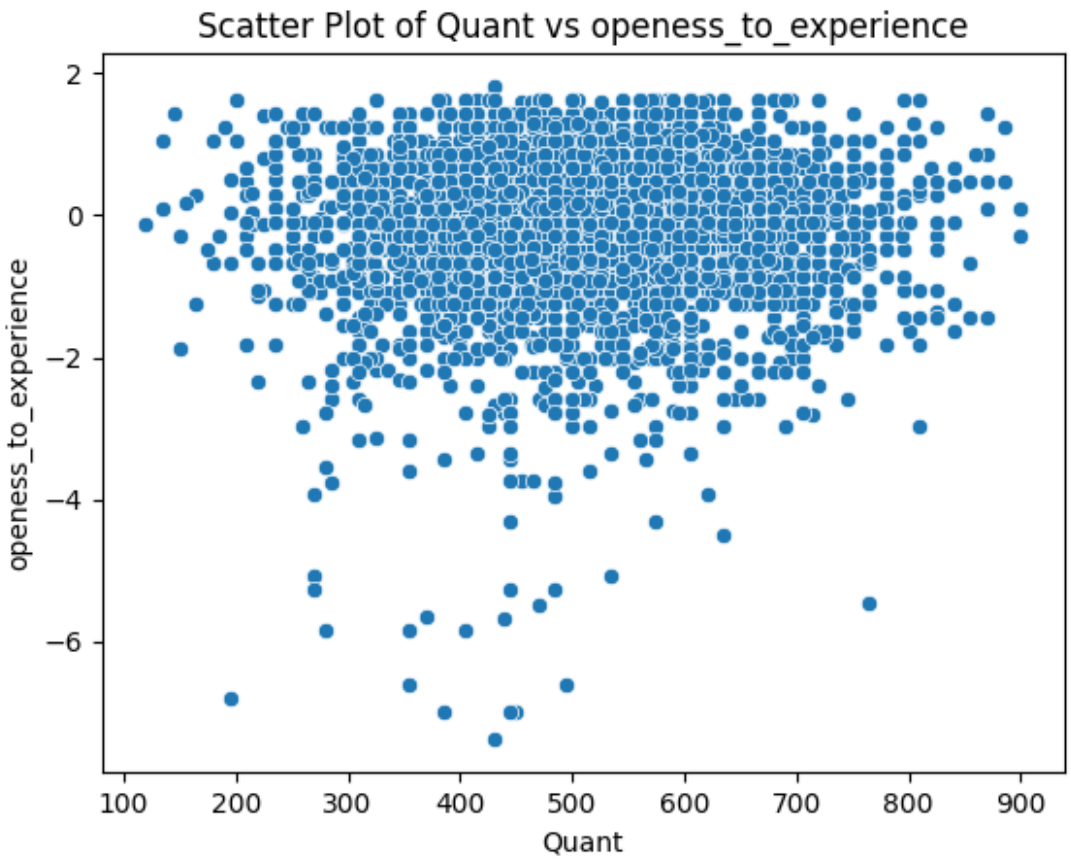

Scatter Plot of Quant vs ComputerScience
ComputerScience
Quant
700
600
500
400
300
200
100
200
300
400
500
600
700
800
900

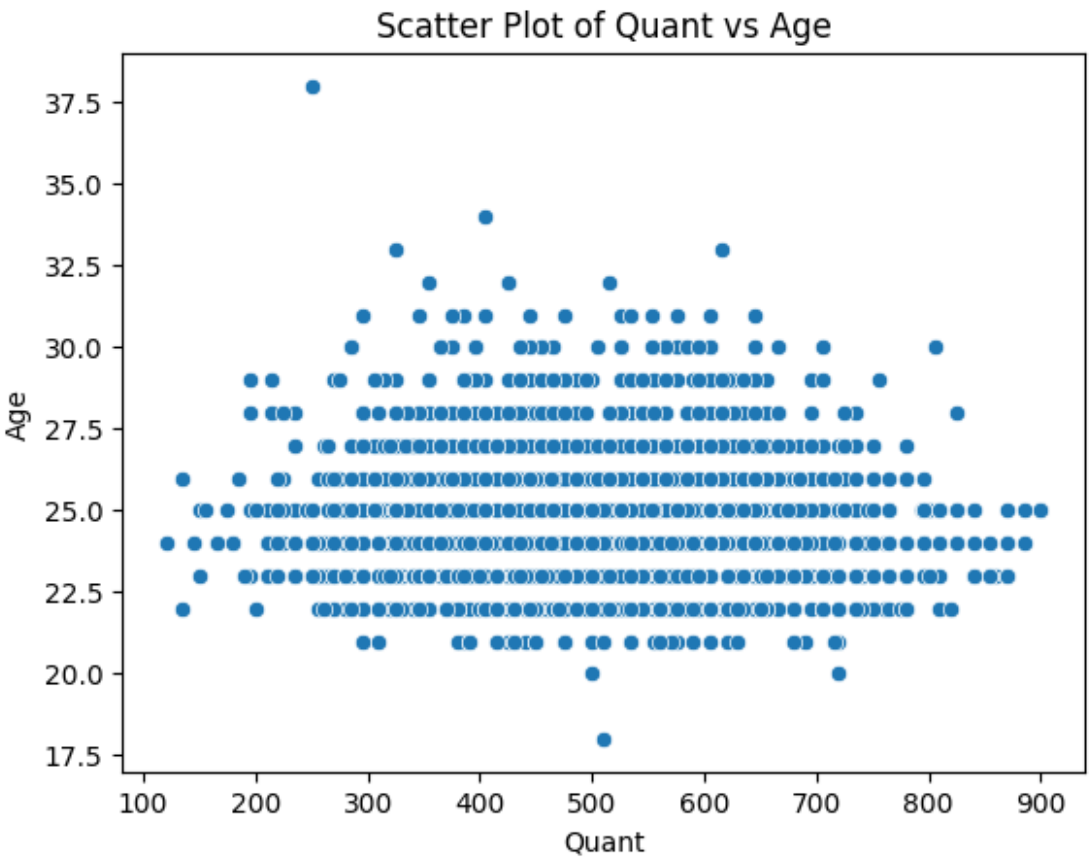

Scatter Plot of Quant vs conscientiousness
conscientiousness
Quant
2
1
0
−1
−2
−3
−4
100
200
300
400
500
600
700
800
900

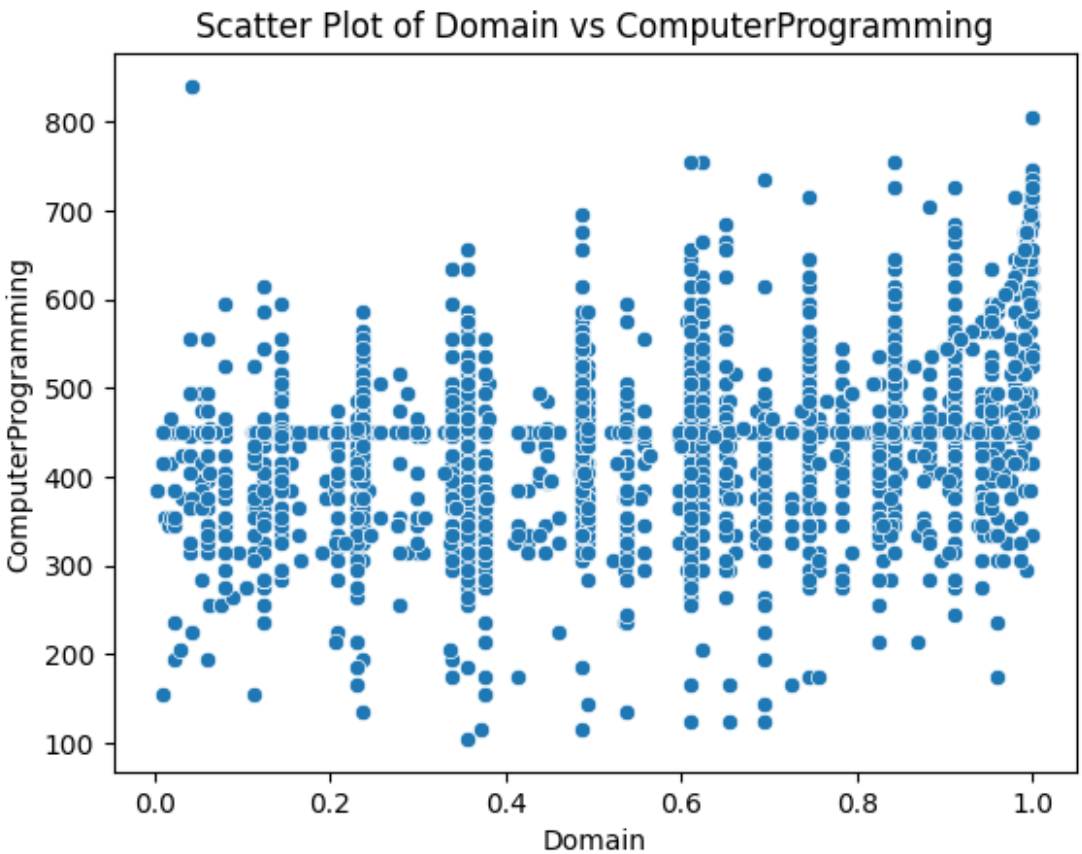

Scatter Plot of Quant vs agreeableness
agreeableness
Quant
2
1
0
−1
−2
−3
−4
−5
−6
100
200
300
400
500
600
700
800
900

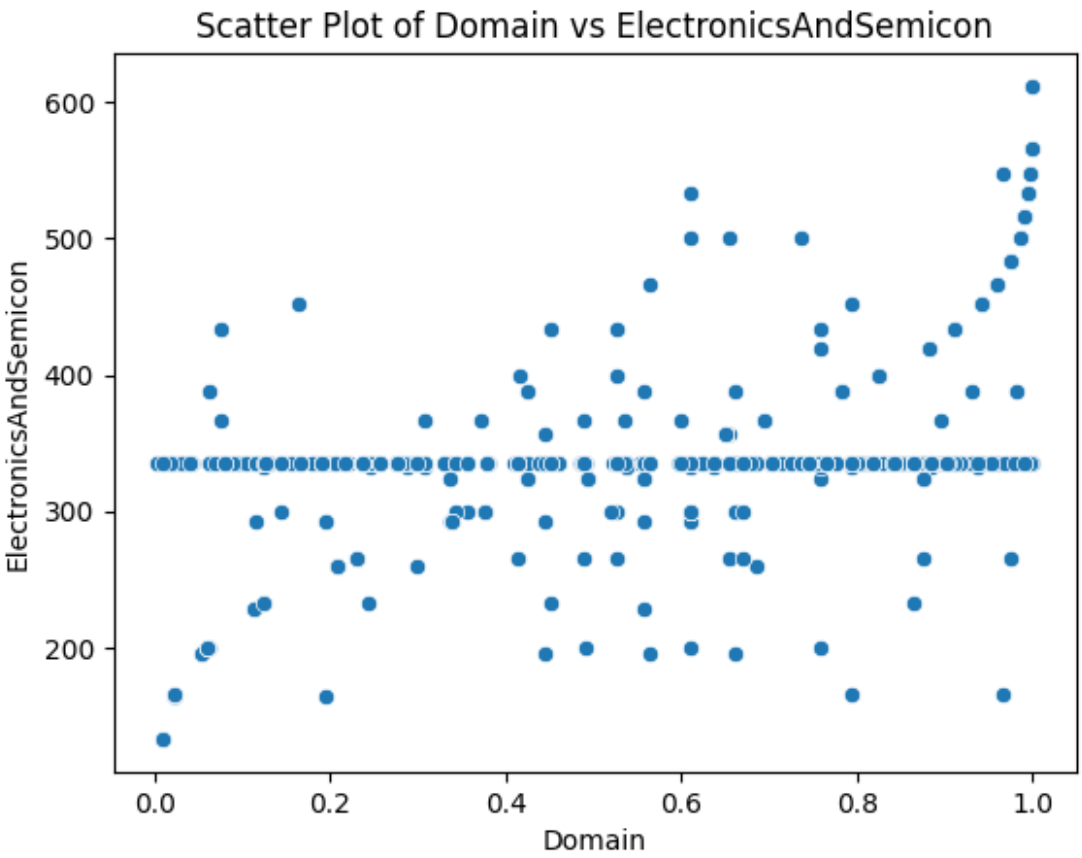

Scatter Plot of Quant vs extraversion
extraversion
Quant
2
1
0
−1
−2
−3
−4
100
200
300
400
500
600
700
800
900

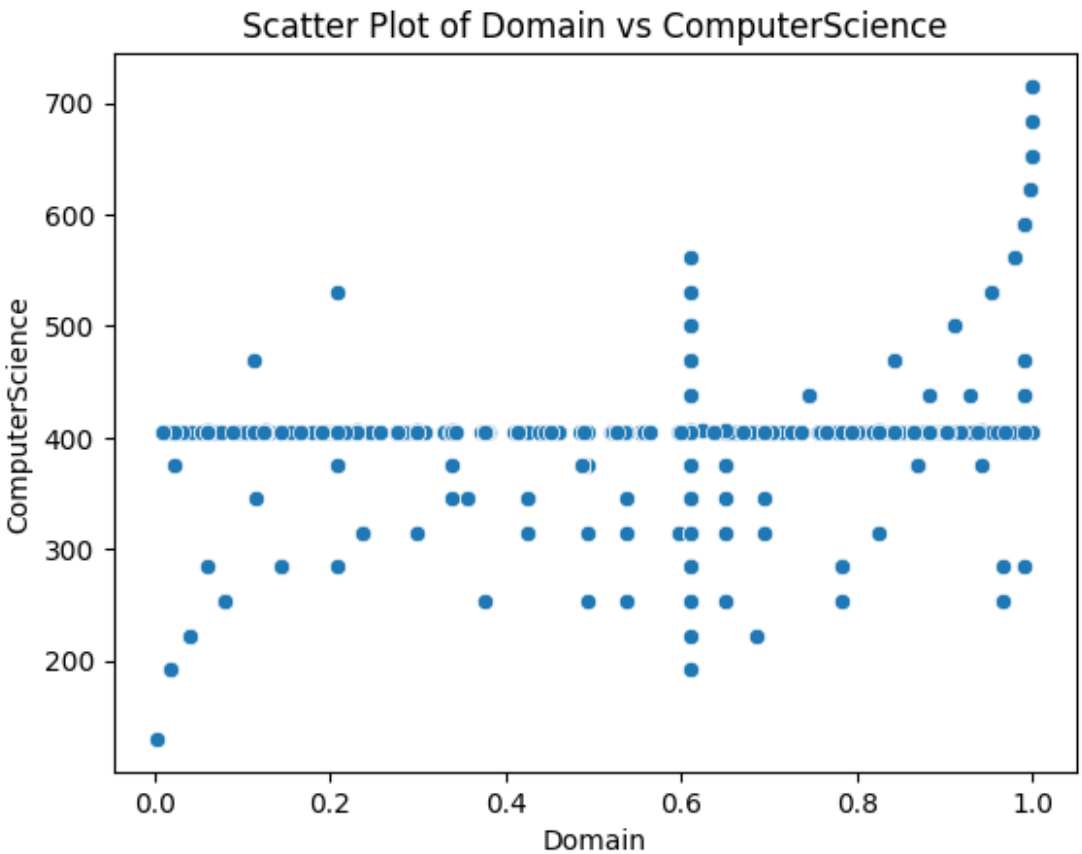

Scatter Plot of Quant vs nueroticism
nueroticism
Quant
3
2
1
0
−1
−2
100
200
300
400
500
600
700
800
900

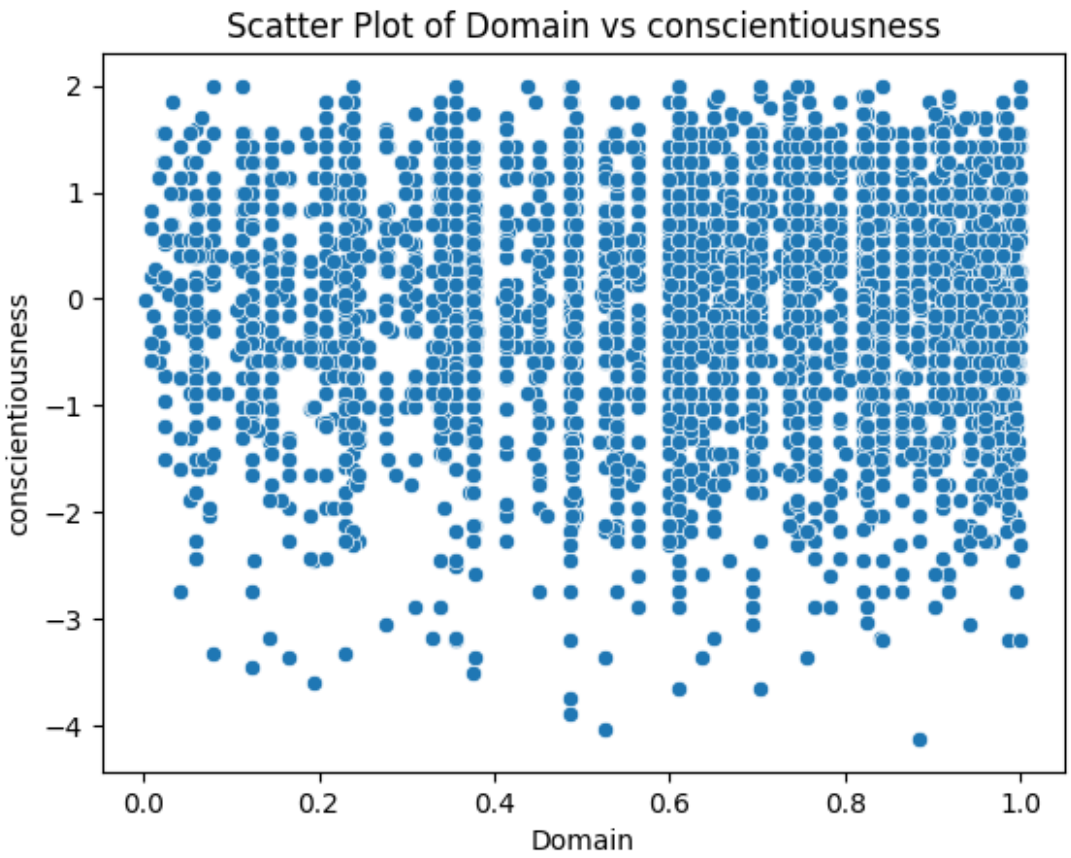

Scatter Plot of Quant vs openess_to_experience
openess_to_experience
Quant
2
0
−2
−4
−6
100
200
300
400
500
600
700
800
900

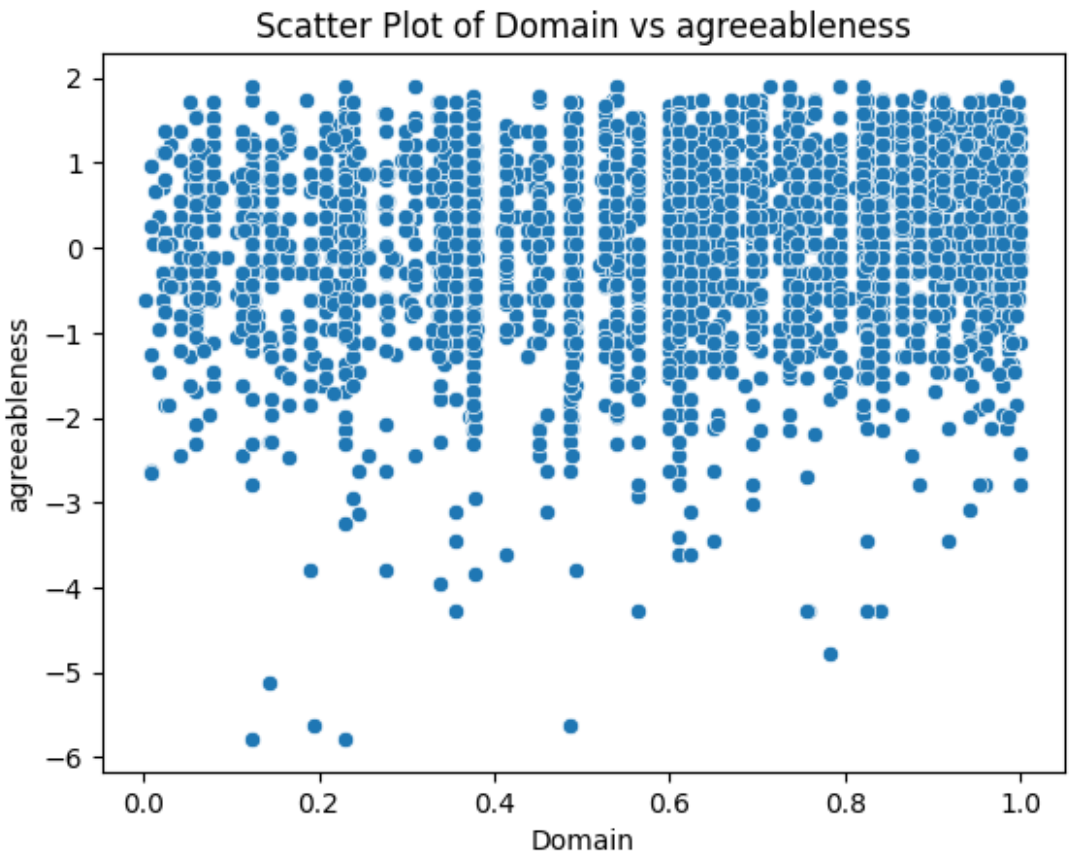

Scatter Plot of Quant vs Age
Age
Quant
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
100
200
300
400
500
600
700
800
900

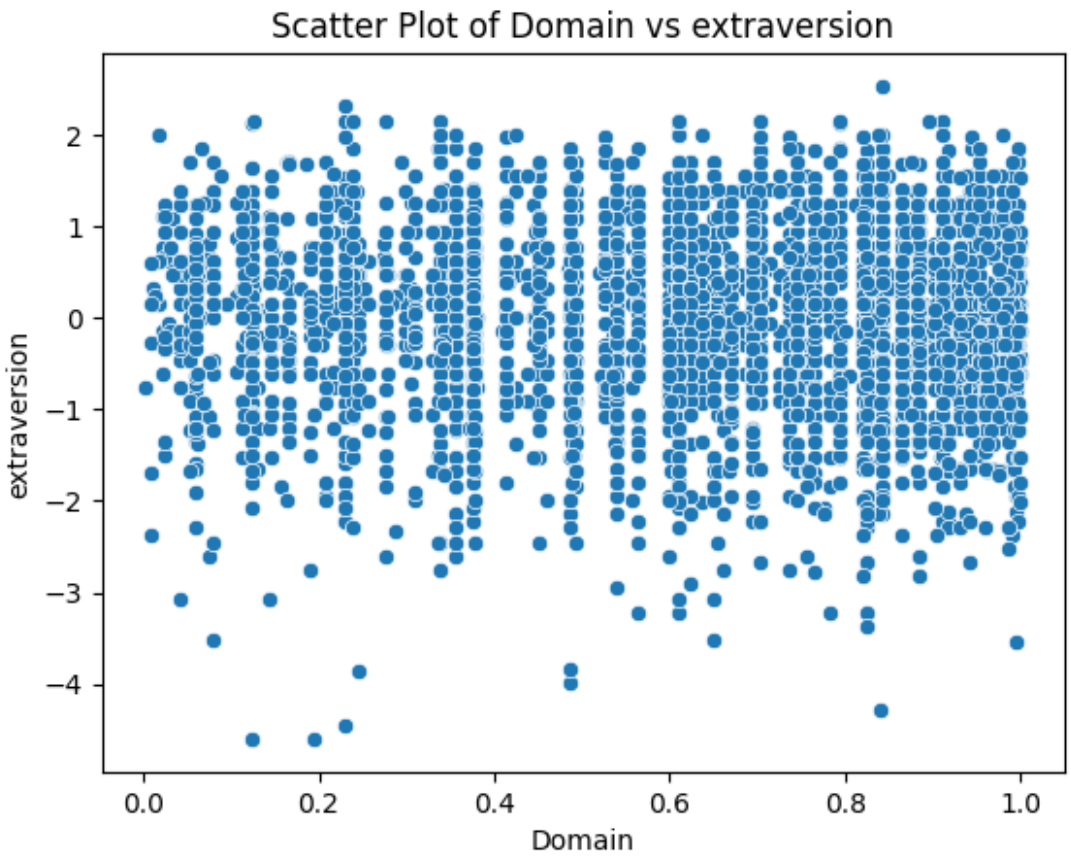

Scatter Plot of Domain vs ComputerProgramming
ComputerProgramming
800
700
600
500
400
300
200
100
0.0
0.2
0.4
0.6
0.8
1.0
Domain

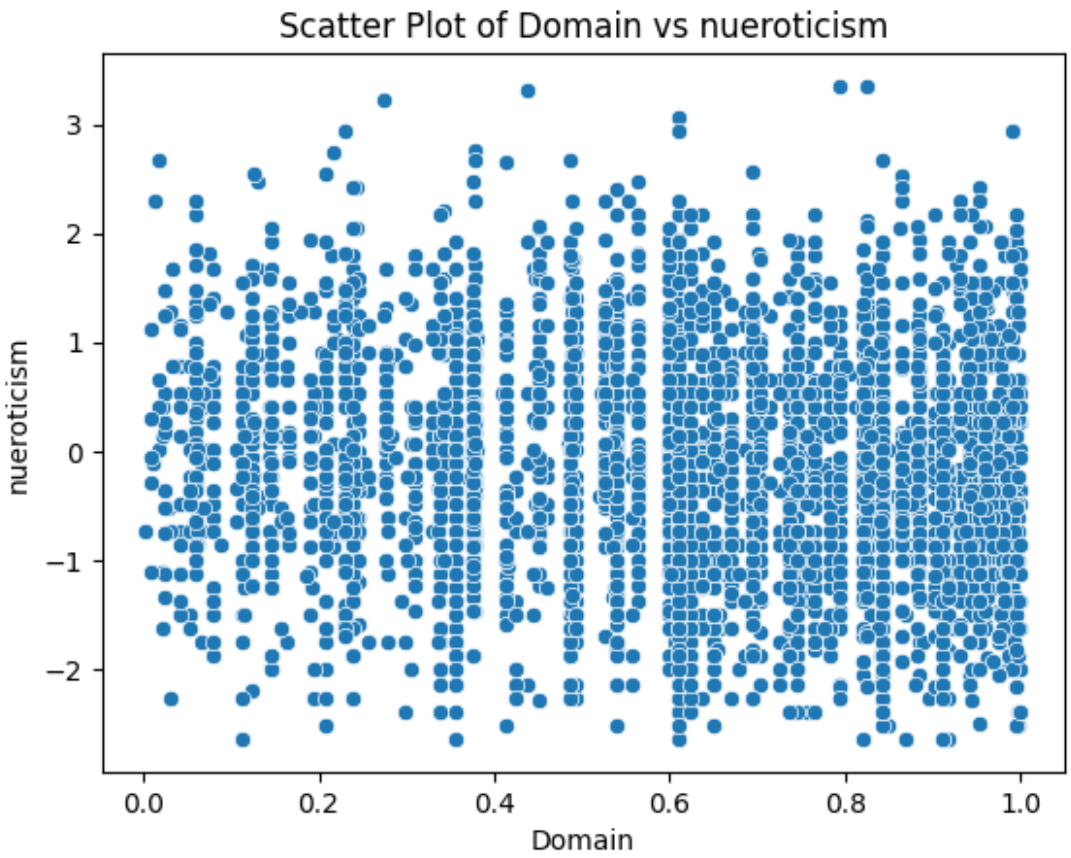

Scatter Plot of Domain vs ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200
0.0
0.2
0.4
0.6
0.8
1.0
Domain

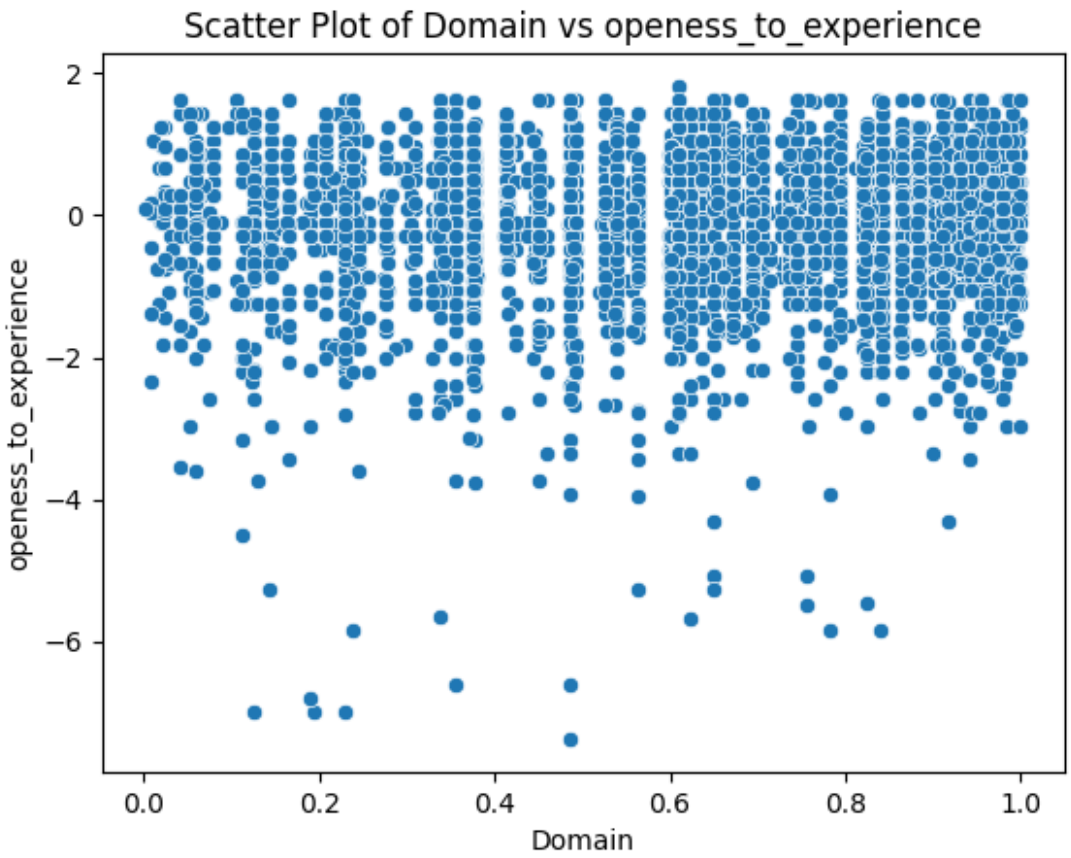

Scatter Plot of Domain vs ComputerScience
ComputerScience
700
600
500
400
300
200
0.0
0.2
0.4
0.6
0.8
1.0
Domain

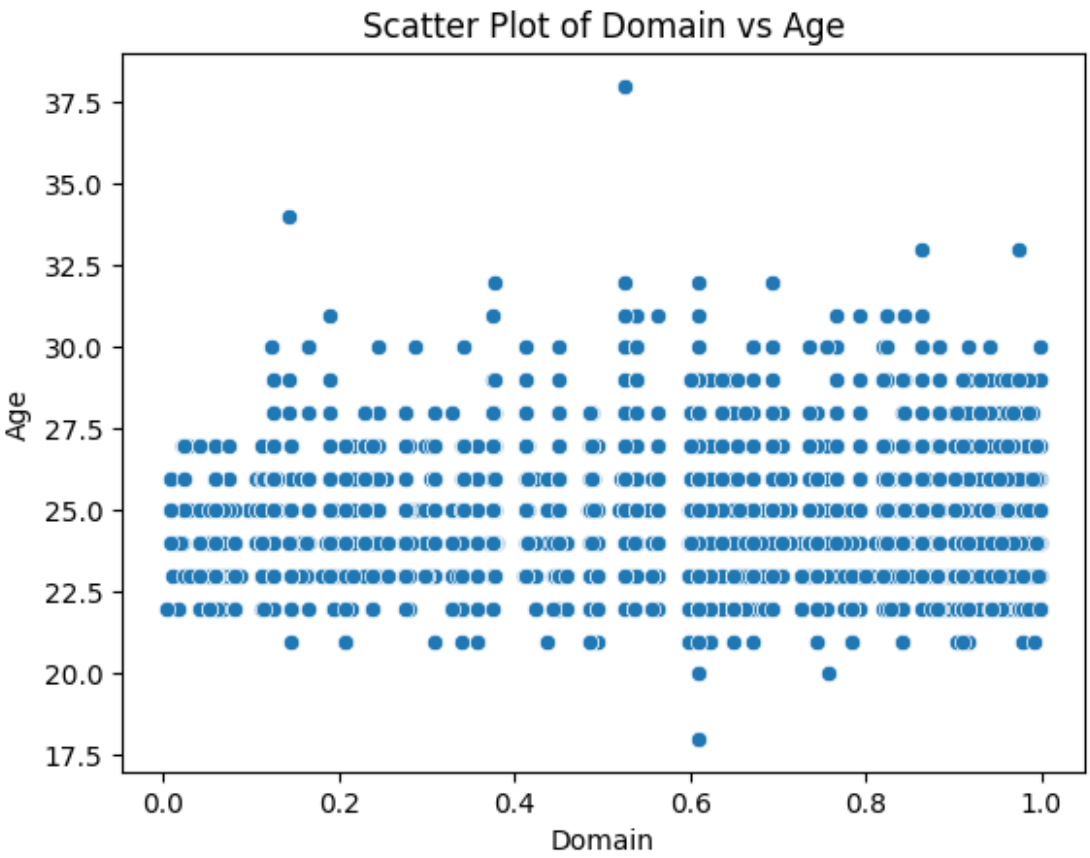

Scatter Plot of Domain vs conscientiousness
conscientiousness
Domain
2
1
0
−1
−2
−3
−4
0.0
0.2
0.4
0.6
0.8
1.0

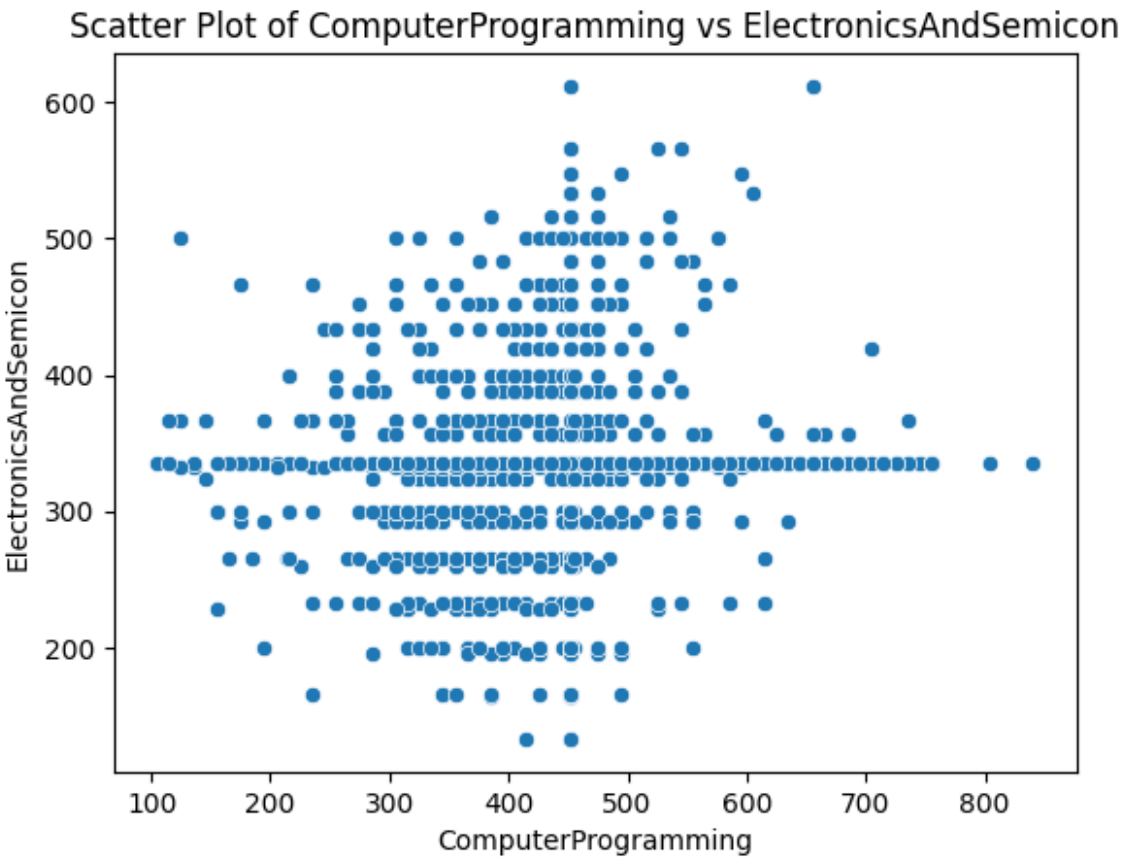

Scatter Plot of Domain vs agreeableness
agreeableness
Domain
2
1
0
−1
−2
−3
−4
−5
−6
0.0
0.2
0.4
0.6
0.8
1.0

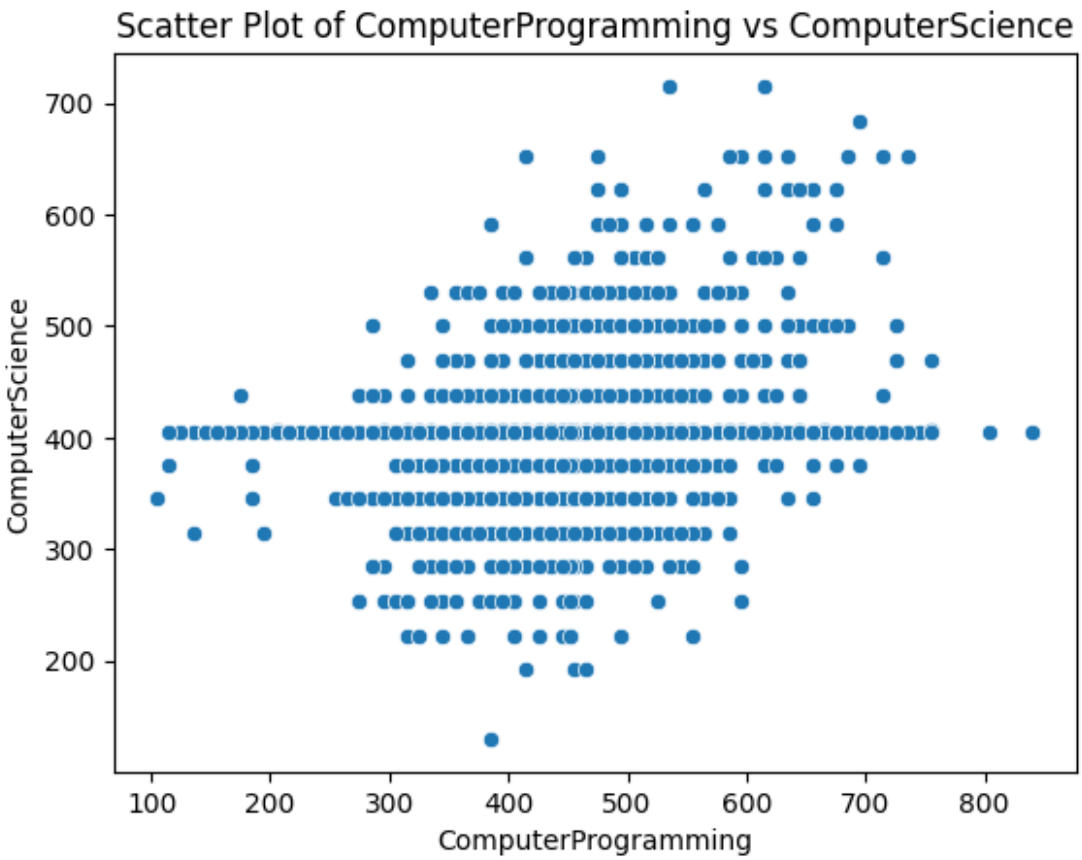

Scatter Plot of Domain vs extraversion
extraversion
Domain
2
1
0
−1
−2
−3
−4
0.0
0.2
0.4
0.6
0.8
1.0

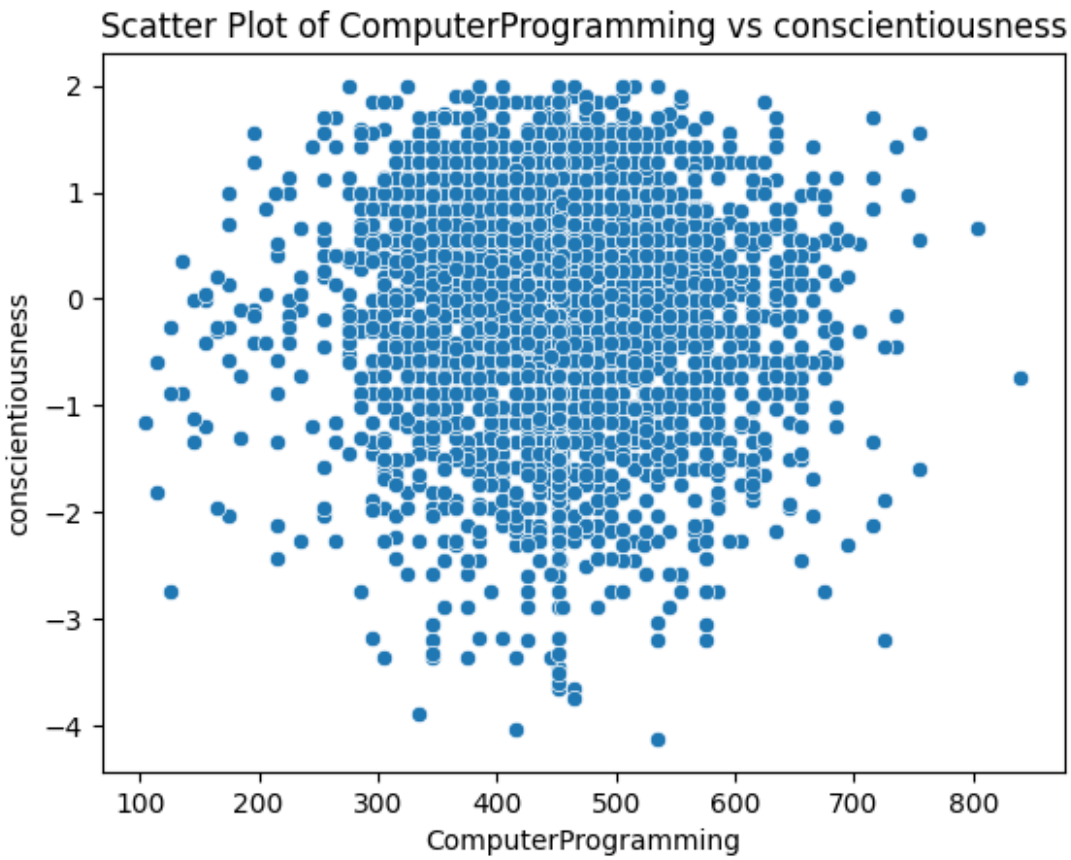

Scatter Plot of Domain vs nueroticism
nueroticism
Domain
3
2
1
0
−1
−2
0.0
0.2
0.4
0.6
0.8
1.0

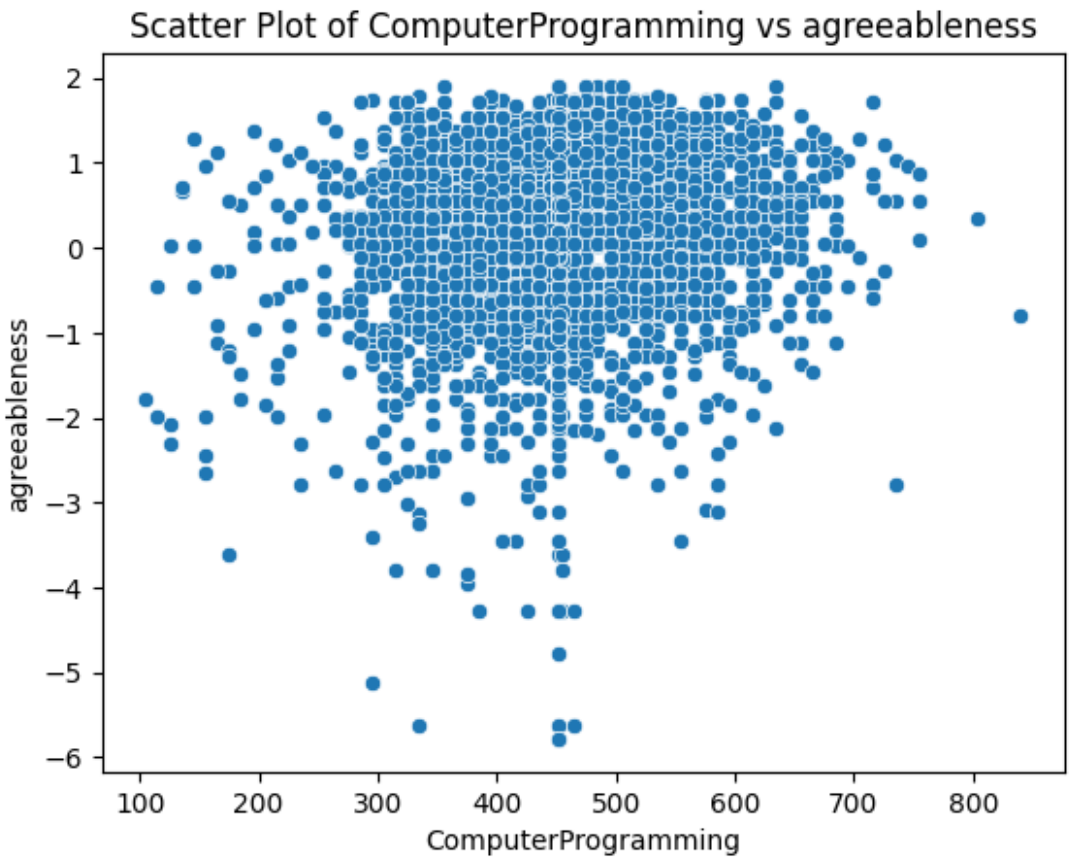

Scatter Plot of Domain vs openess_to_experience
openess_to_experience
Domain
2
0
−2
−4
−6
0.0
0.2
0.4
0.6
0.8
1.0

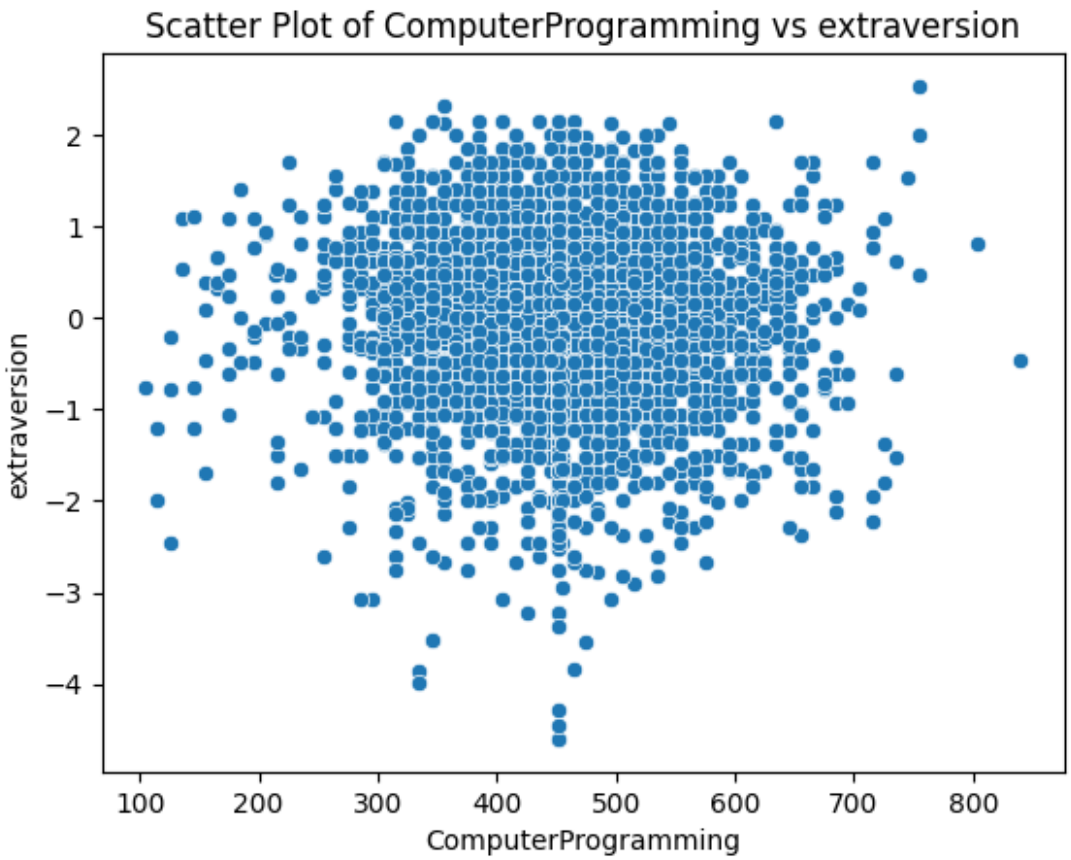

Scatter Plot of Domain vs Age
Age
Domain
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
0.0
0.2
0.4
0.6
0.8
1.0

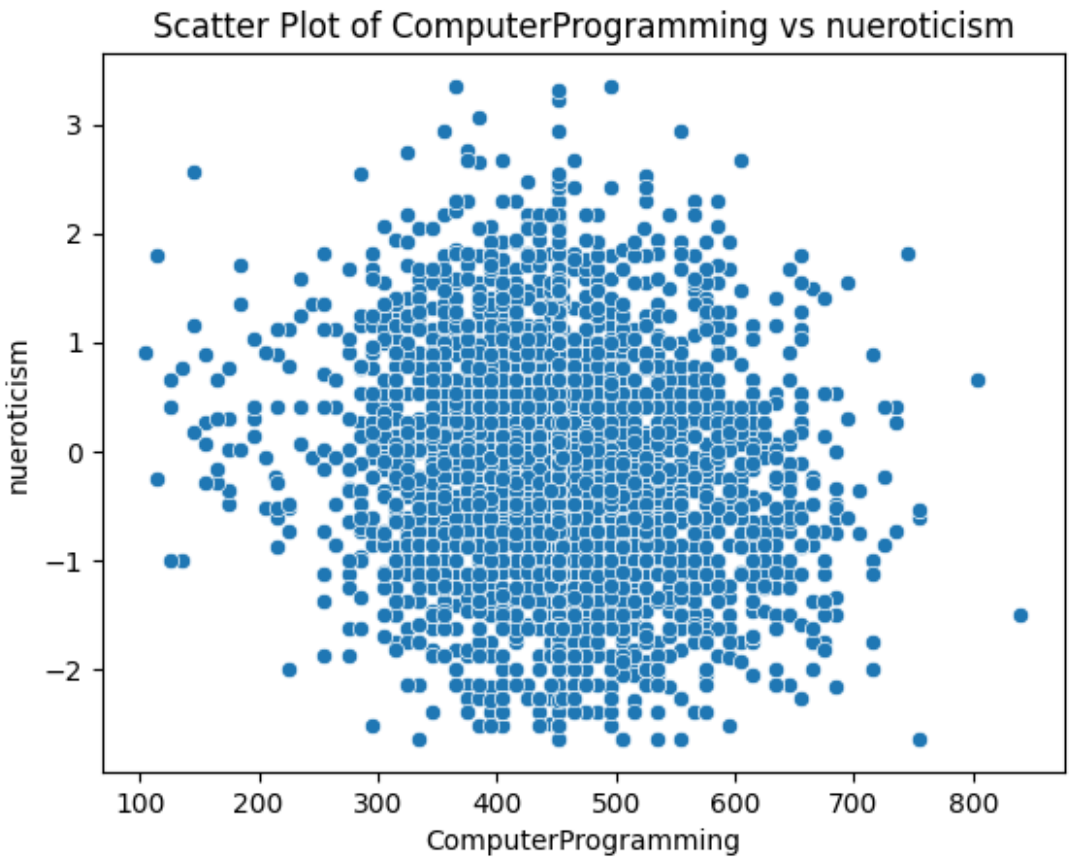

Scatter Plot of ComputerProgramming vs ElectronicsAndSemicon
ElectronicsAndSemicon
600
500
400
300
200
100
200
300
400
500
600
700
800
ComputerProgramming

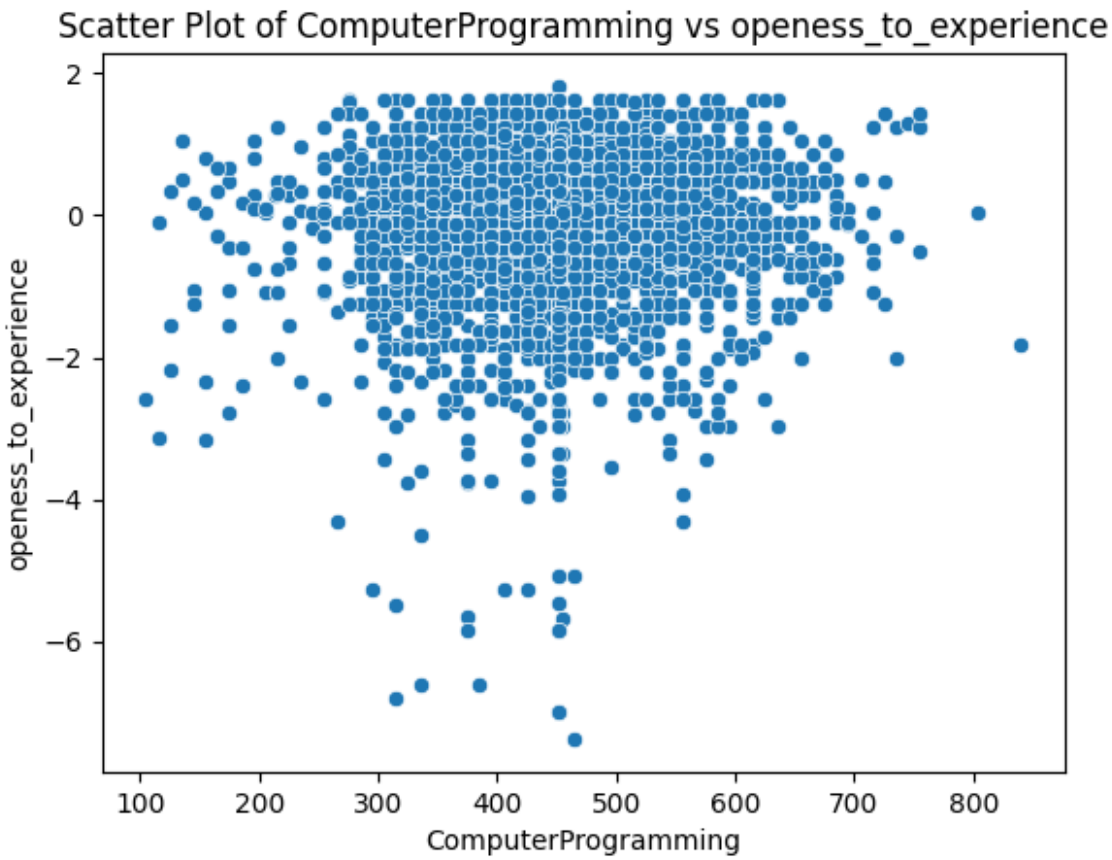

Scatter Plot of ComputerProgramming vs ComputerScience
ComputerScience
ComputerProgramming
700
600
500
400
300
200
100
200
300
400
500
600
700
800

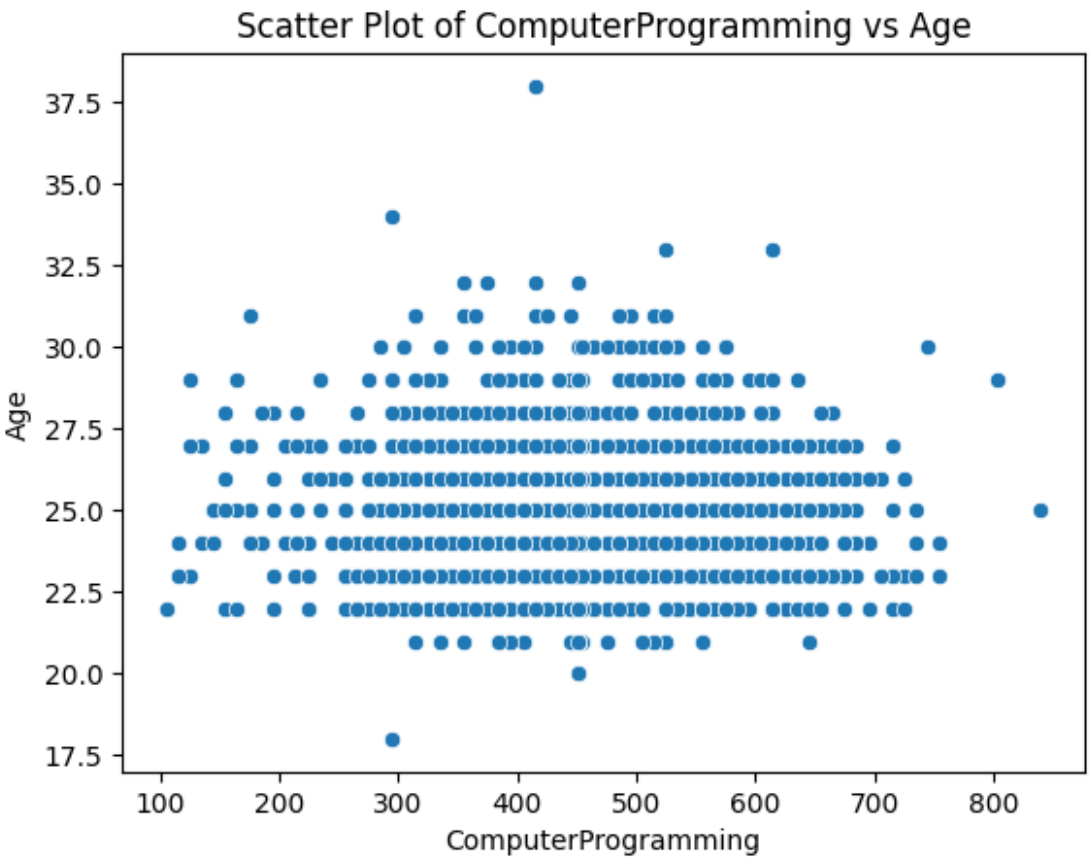

Scatter Plot of ComputerProgramming vs conscientiousness
conscientiousness
ComputerProgramming
2
1
0
−1
−2
−3
−4
100
200
300
400
500
600
700
800

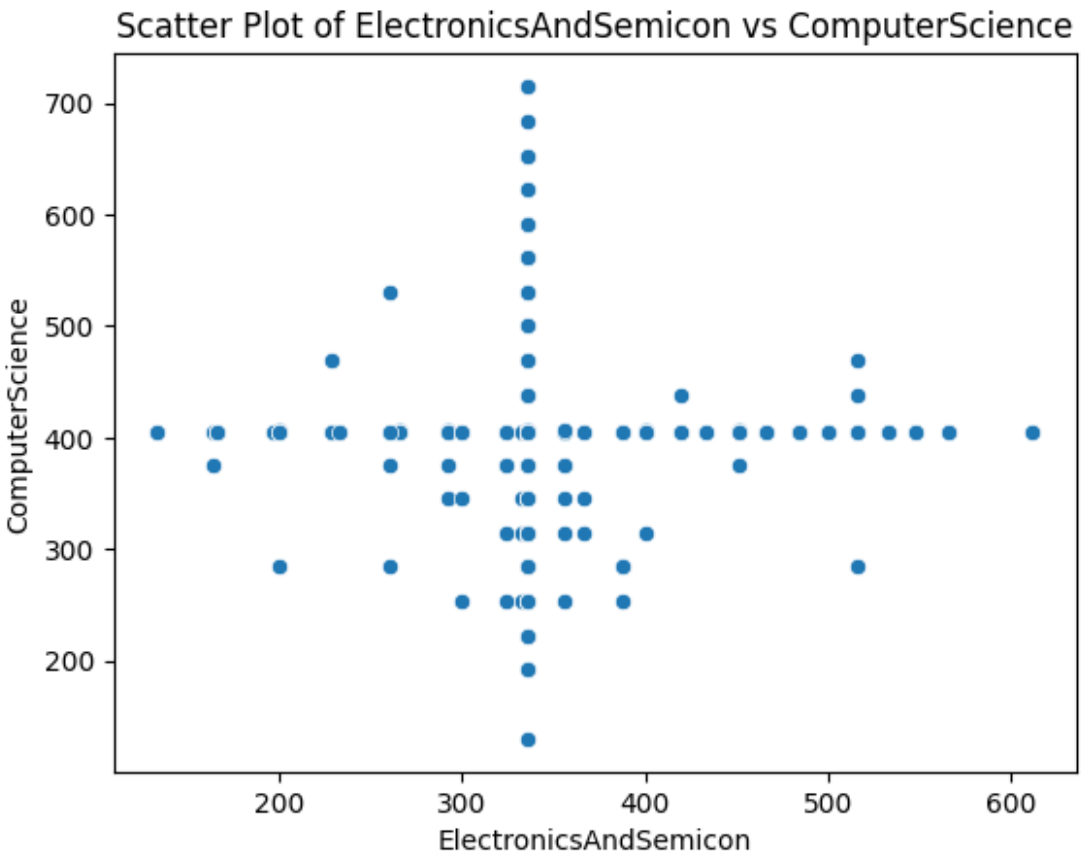

Scatter Plot of ComputerProgramming vs agreeableness
agreeableness
ComputerProgramming
2
1
0
−1
−2
−3
−4
−5
−6
100
200
300
400
500
600
700
800

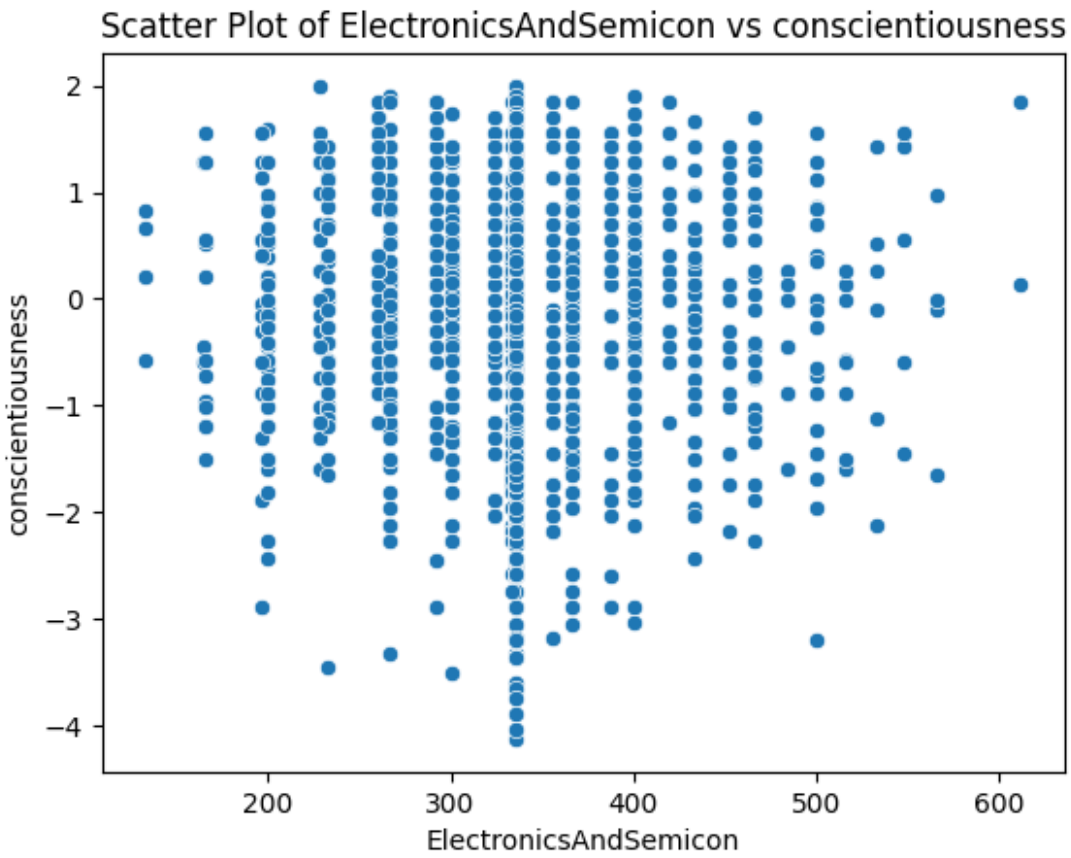

Scatter Plot of ComputerProgramming vs extraversion
extraversion
ComputerProgramming
2
1
0
−1
−2
−3
−4
100
200
300
400
500
600
700
800

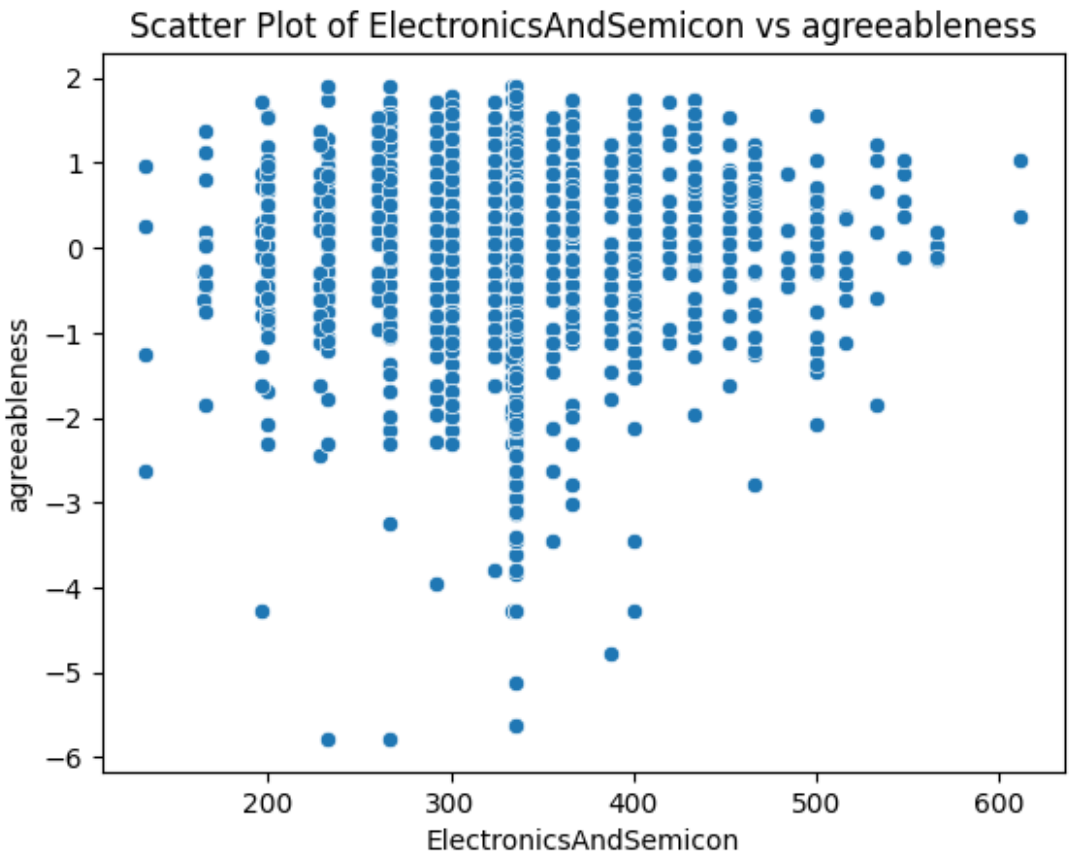

Scatter Plot of ComputerProgramming vs nueroticism
nueroticism
ComputerProgramming
3
2
1
0
−1
−2
100
200
300
400
500
600
700
800

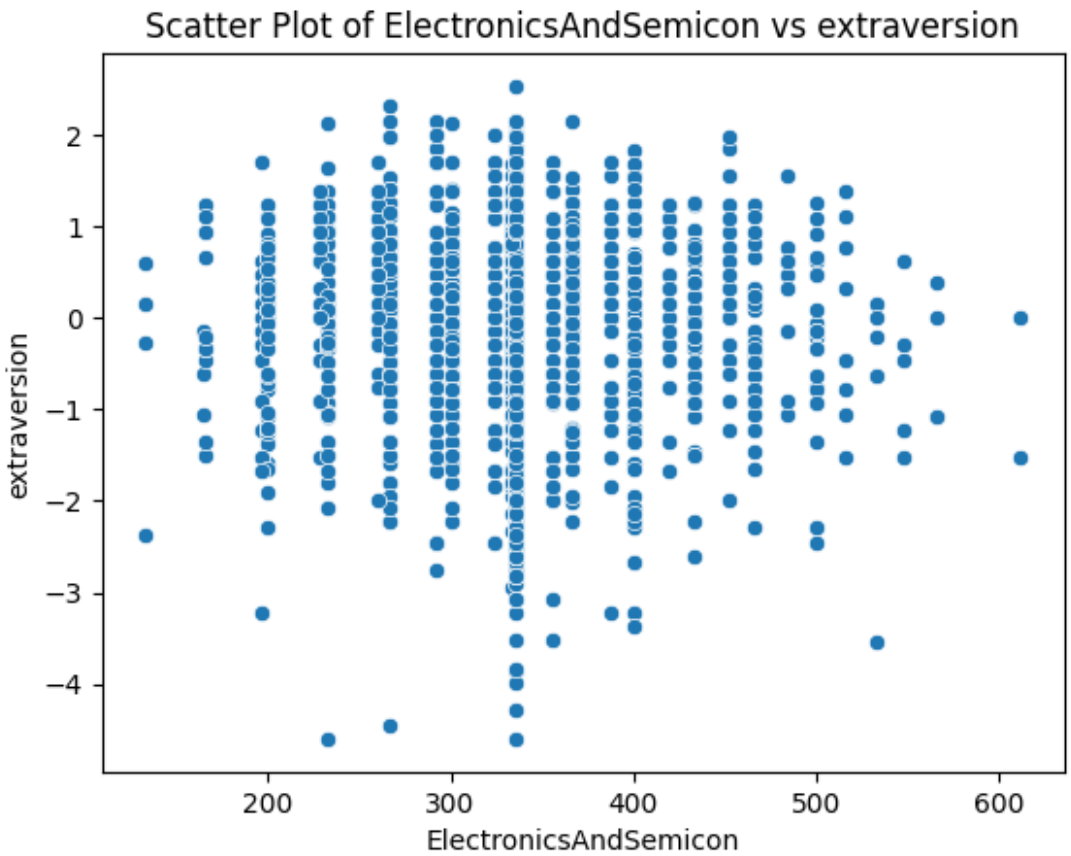

Scatter Plot of ComputerProgramming vs openess_to_experience
openess_to_experience
ComputerProgramming
2
0
−2
−4
−6
100
200
300
400
500
600
700
800

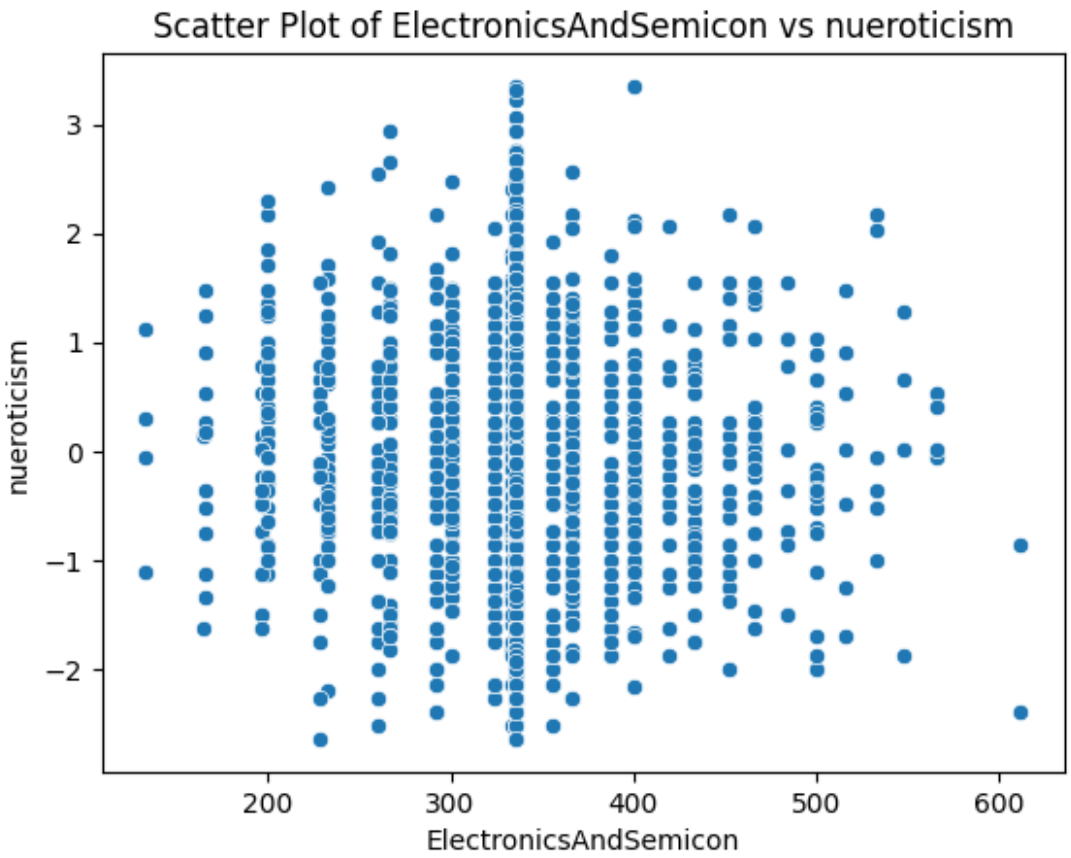

Scatter Plot of ComputerProgramming vs Age
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
100
200
300
400
500
600
700
800
ComputerProgramming

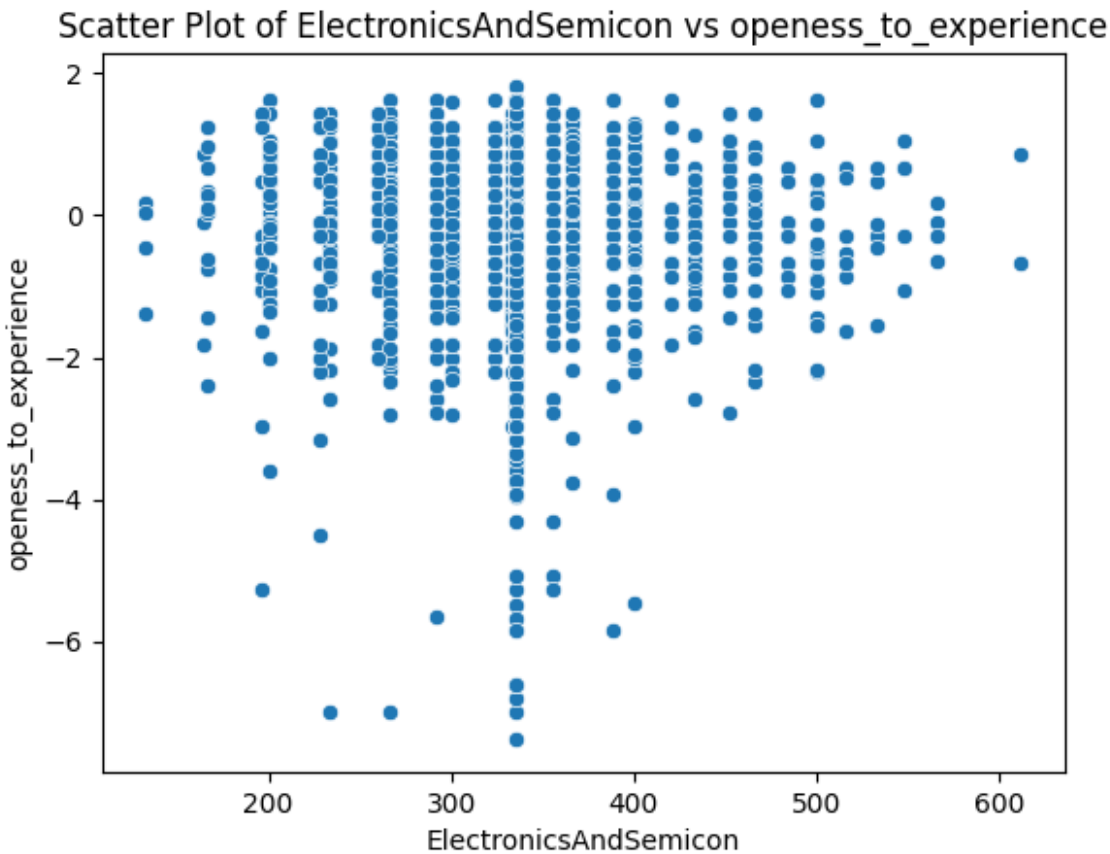

Scatter Plot of ElectronicsAndSemicon vs ComputerScience
ComputerScience
ElectronicsAndSemicon
700
600
500
400
300
200
200
300
400
500
600

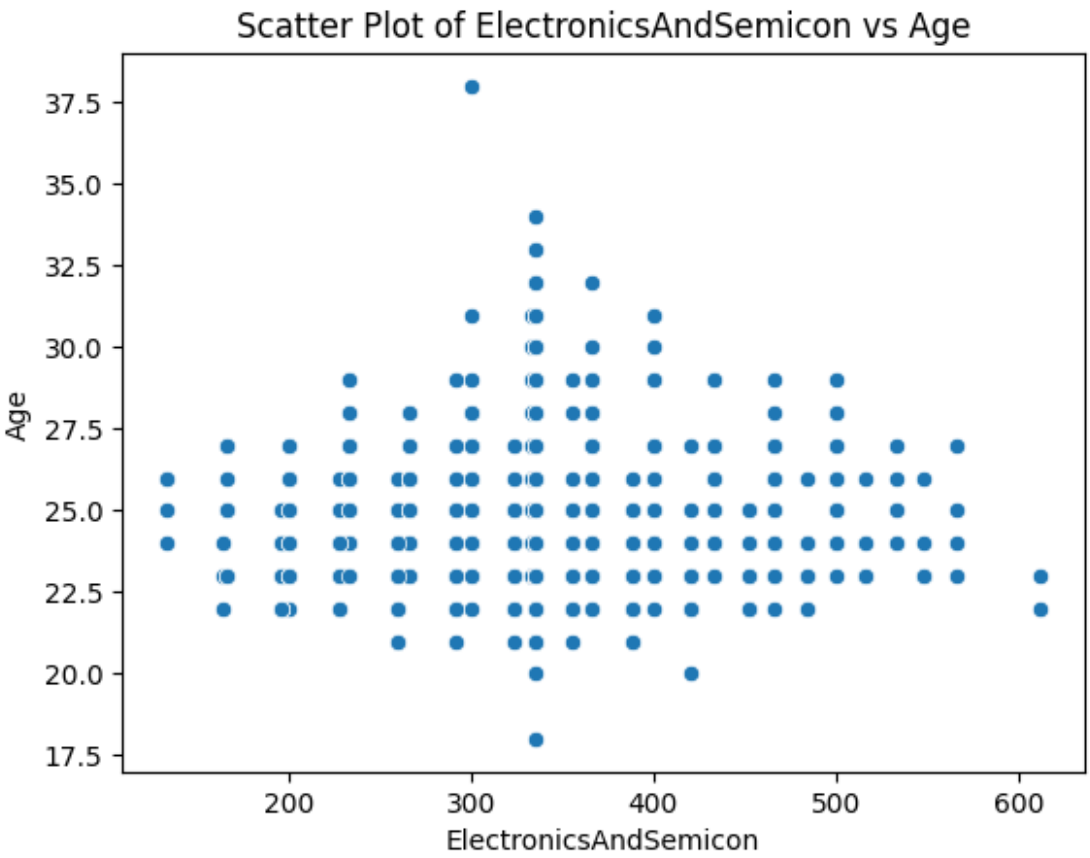

Scatter Plot of ElectronicsAndSemicon vs conscientiousness
conscientiousness
ElectronicsAndSemicon
2
1
0
−1
−2
−3
−4
200
300
400
500
600

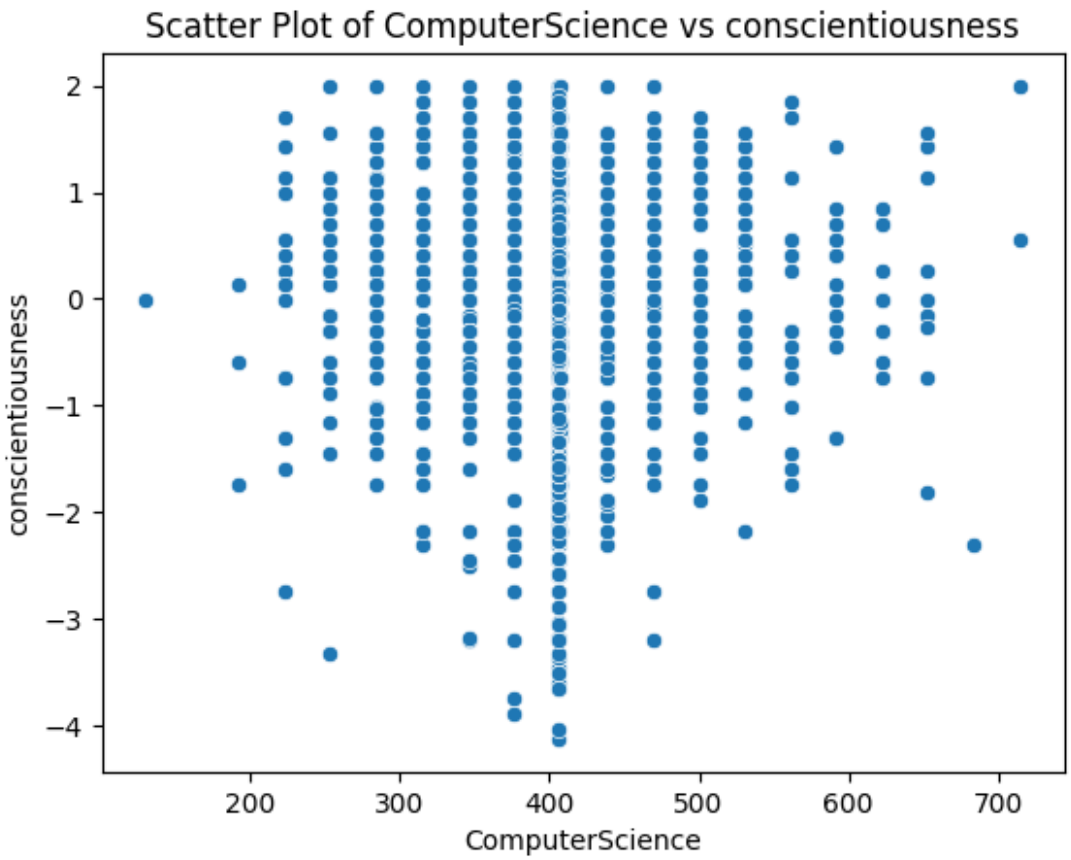

Scatter Plot of ElectronicsAndSemicon vs agreeableness
agreeableness
ElectronicsAndSemicon
2
1
0
−1
−2
−3
−4
−5
−6
200
300
400
500
600

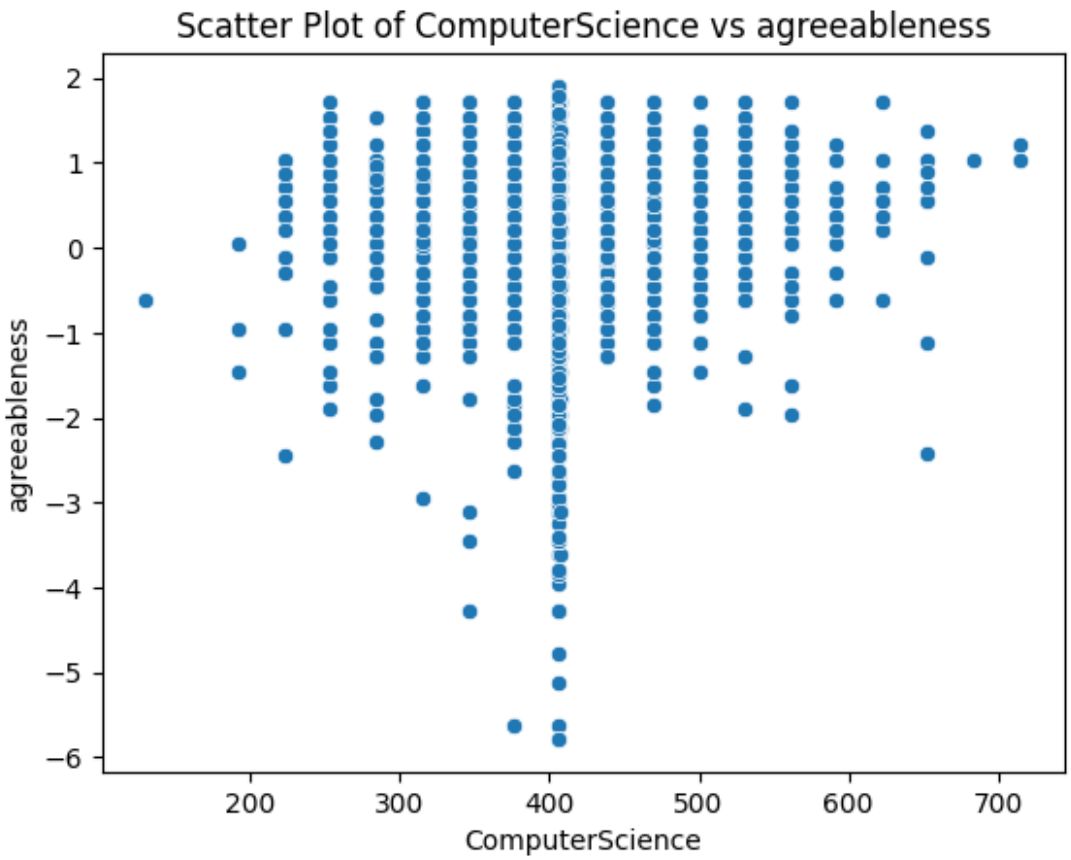

Scatter Plot of ElectronicsAndSemicon vs extraversion
extraversion
ElectronicsAndSemicon
2
1
0
−1
−2
−3
−4
200
300
400
500
600

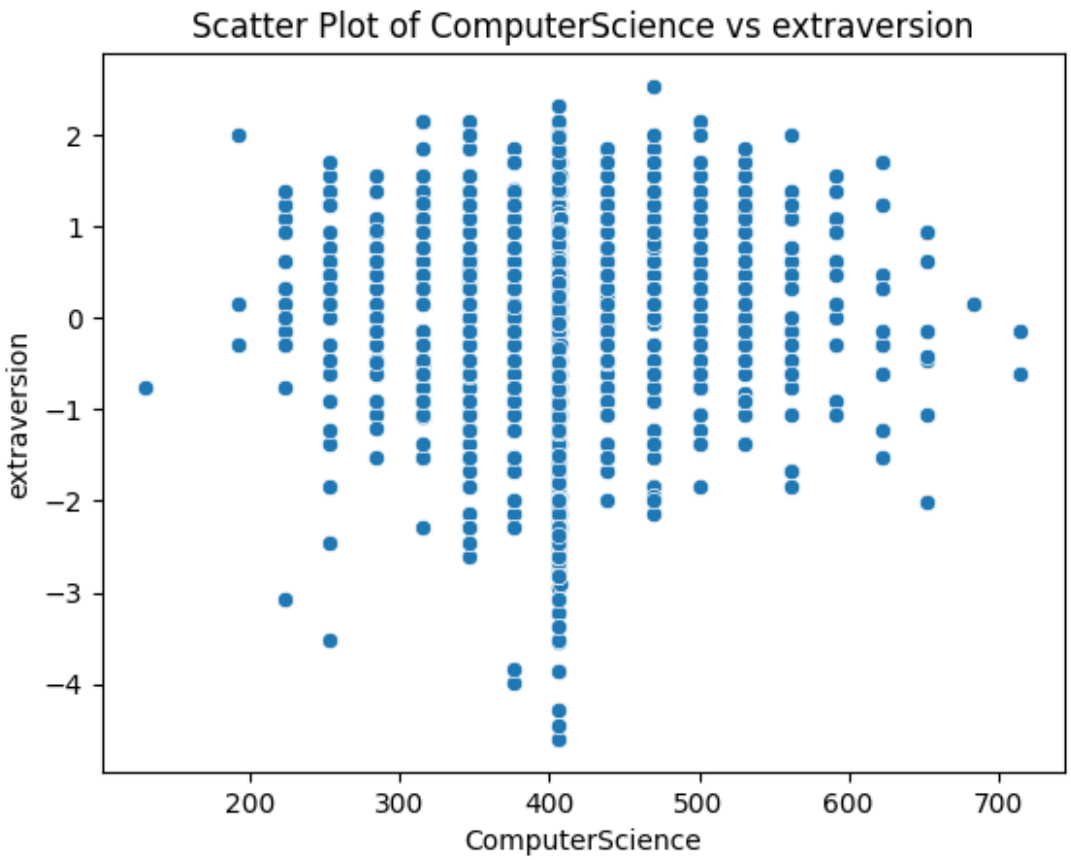

Scatter Plot of ElectronicsAndSemicon vs nueroticism
nueroticism
ElectronicsAndSemicon
3
2
1
0
−1
−2
200
300
400
500
600

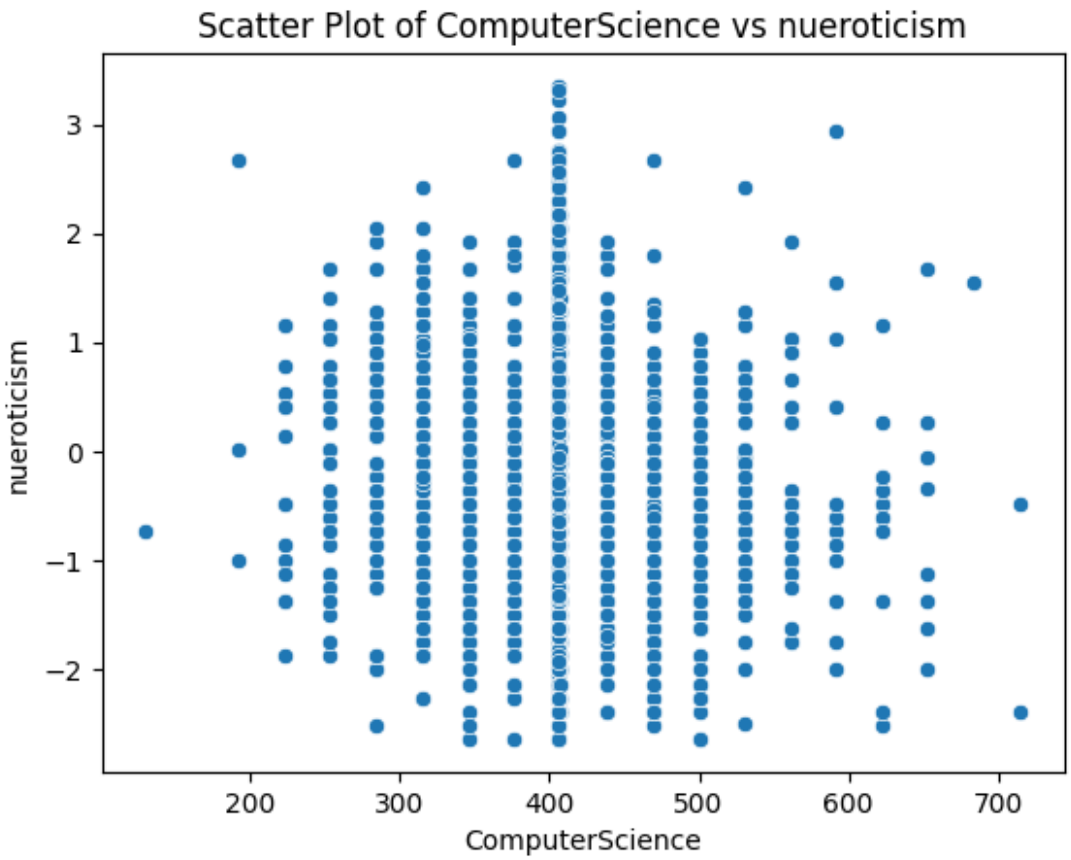

Scatter Plot of ElectronicsAndSemicon vs openess_to_experience
2
0
−2
−4
−6
openess_to_experience
200
300
400
500
600
ElectronicsAndSemicon

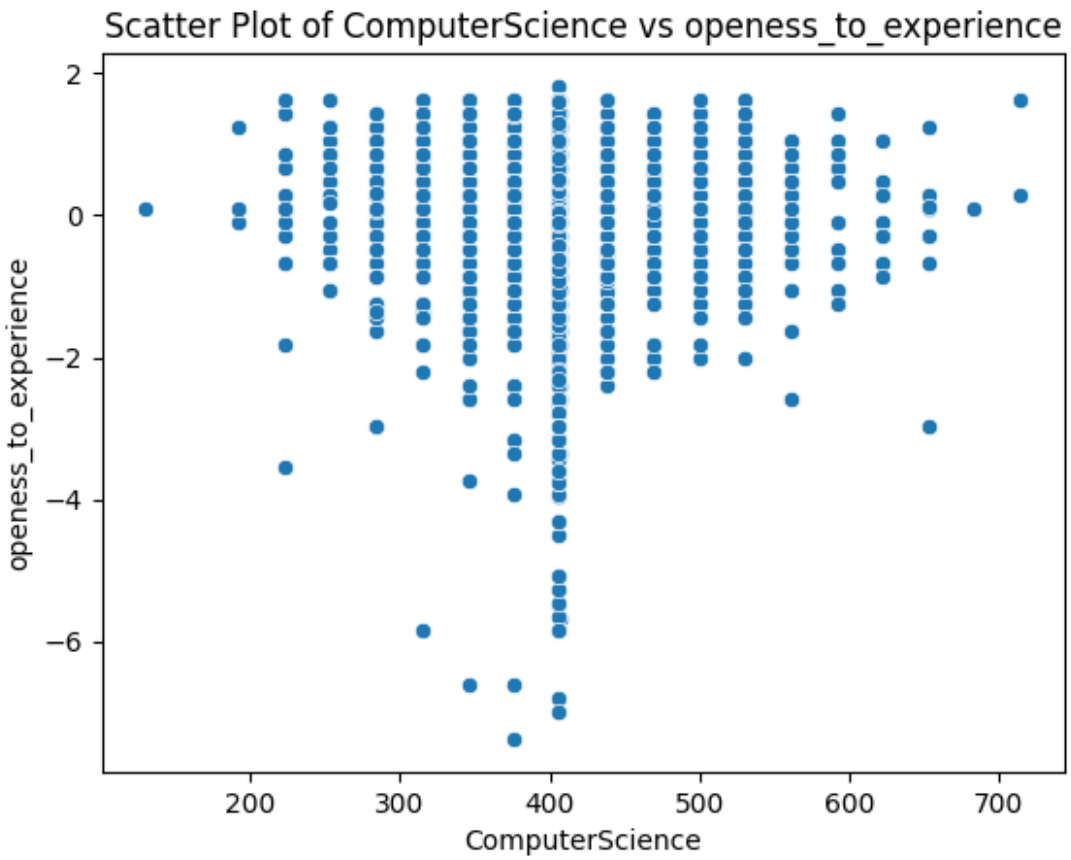

Scatter Plot of ElectronicsAndSemicon vs Age
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
200
300
400
500
600
ElectronicsAndSemicon

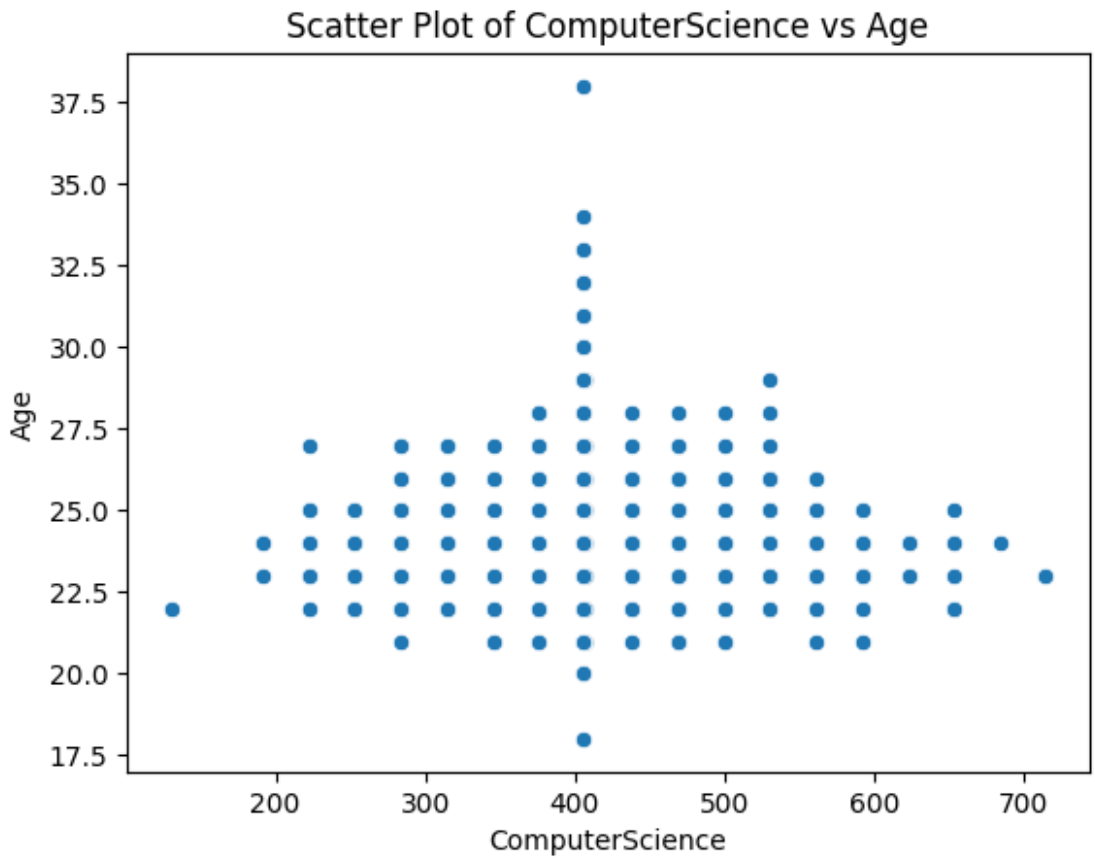

Scatter Plot of ComputerScience vs conscientiousness
conscientiousness
ComputerScience
2
1
0
−1
−2
−3
−4
200
300
400
500
600
700

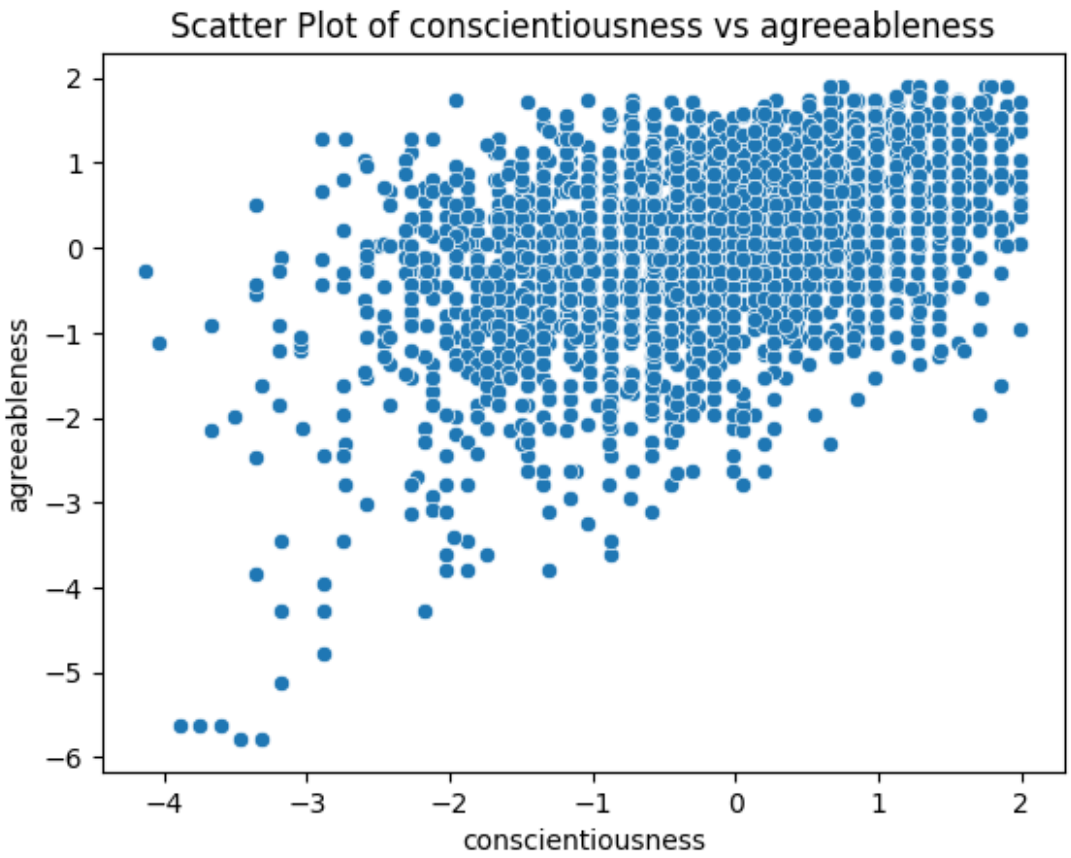

Scatter Plot of ComputerScience vs agreeableness
agreeableness
ComputerScience
2
1
0
−1
−2
−3
−4
−5
−6
200
300
400
500
600
700

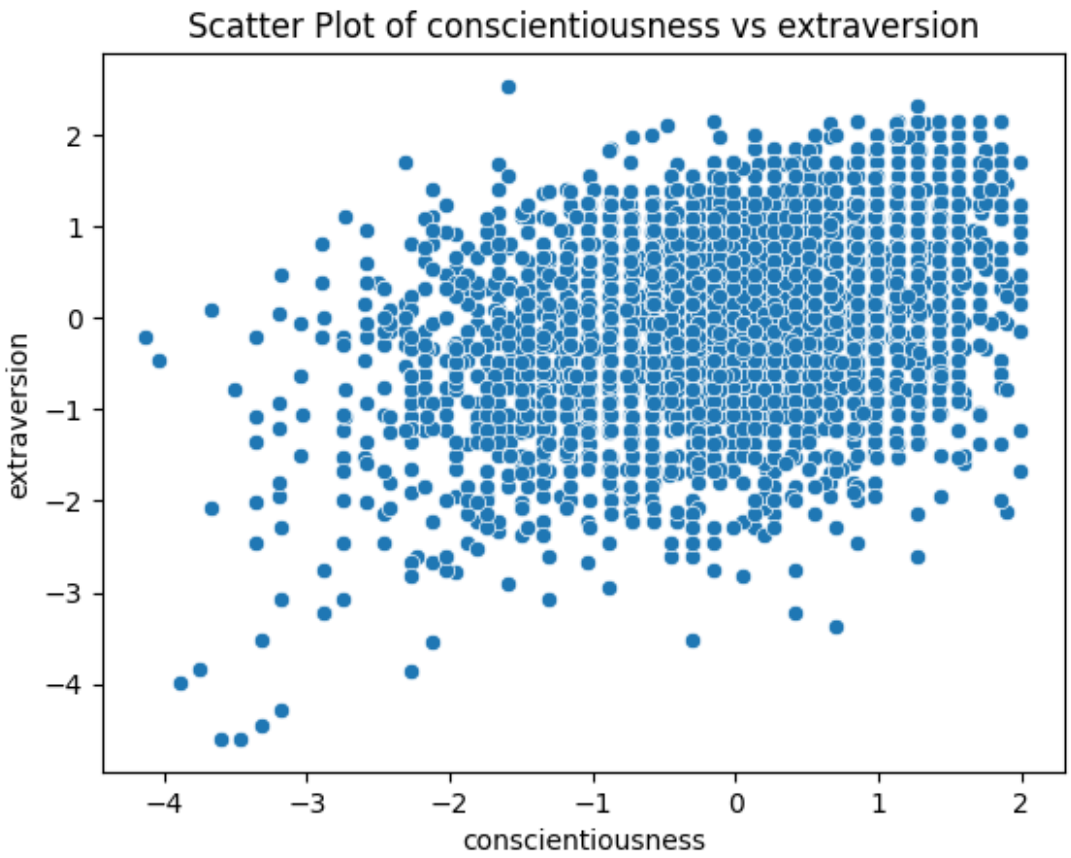

Scatter Plot of ComputerScience vs extraversion
extraversion
ComputerScience
2
1
0
−1
−2
−3
−4
200
300
400
500
600
700

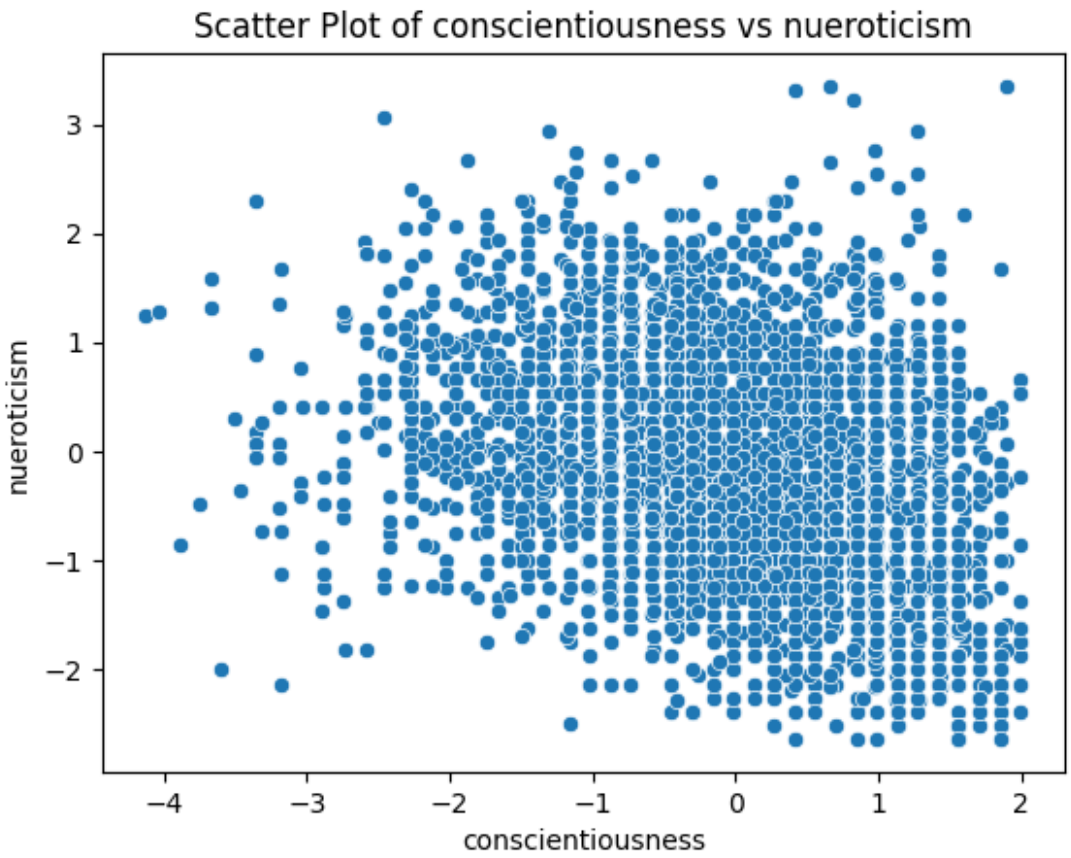

Scatter Plot of ComputerScience vs nueroticism
nueroticism
ComputerScience
3
2
1
0
−1
−2
200
300
400
500
600
700

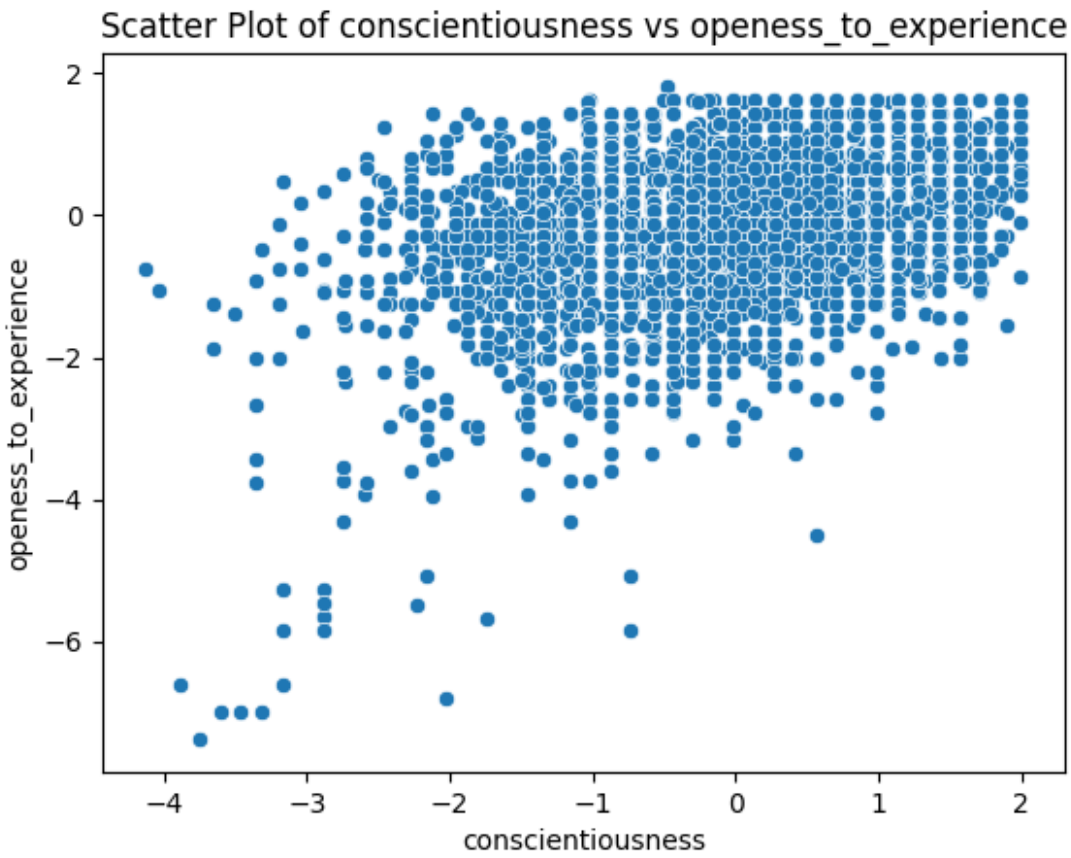

Scatter Plot of ComputerScience vs openess_to_experience
openess_to_experience
ComputerScience
2
0
−2
−4
−6
200
300
400
500
600
700

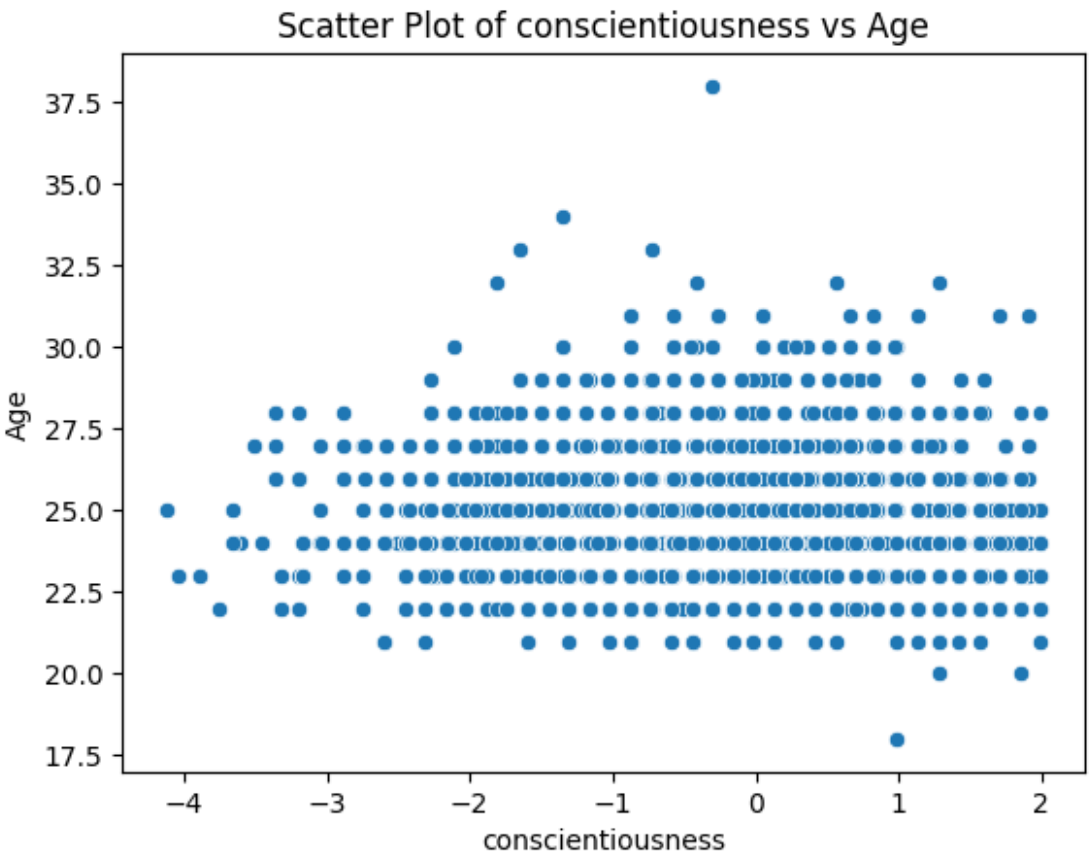

Scatter Plot of ComputerScience vs Age
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
200
300
400
500
600
700
ComputerScience

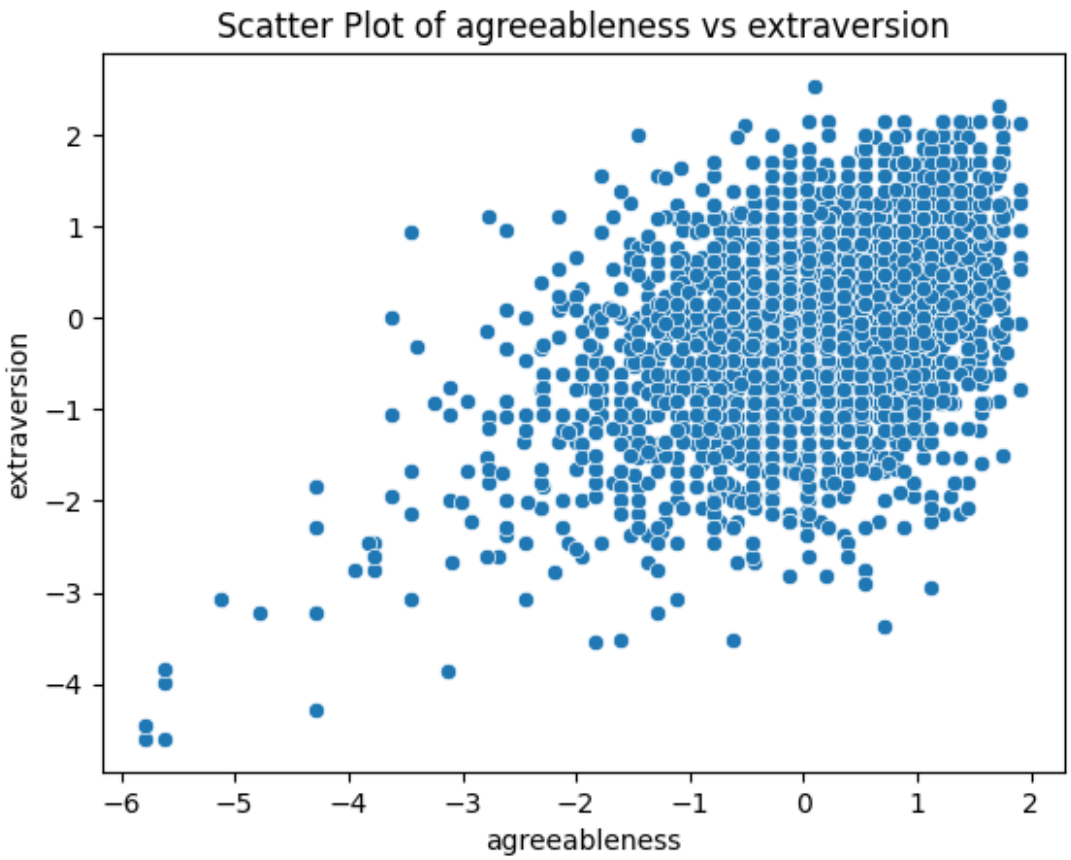

Scatter Plot of conscientiousness vs agreeableness
agreeableness
conscientiousness
2
1
0
−1
−2
−3
−4
−5
−6
−4
−3
−2
−1
0
1
2

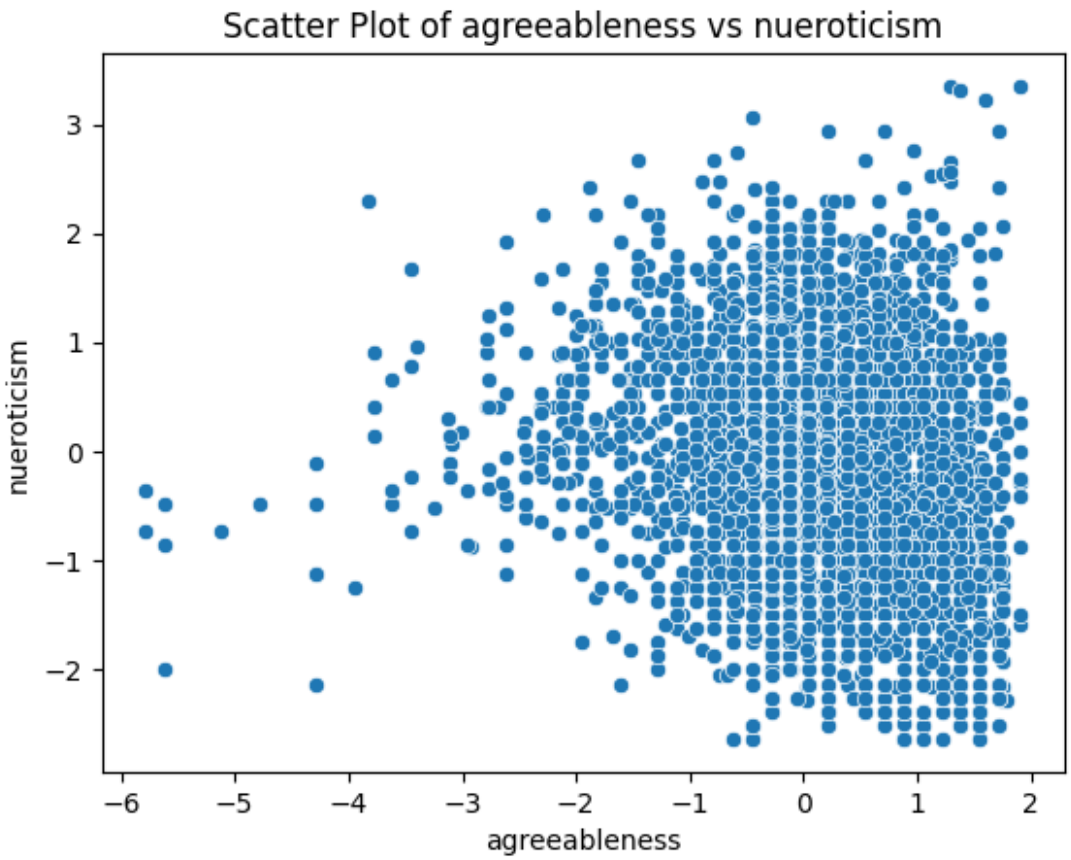

Scatter Plot of conscientiousness vs extraversion
extraversion
conscientiousness
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2

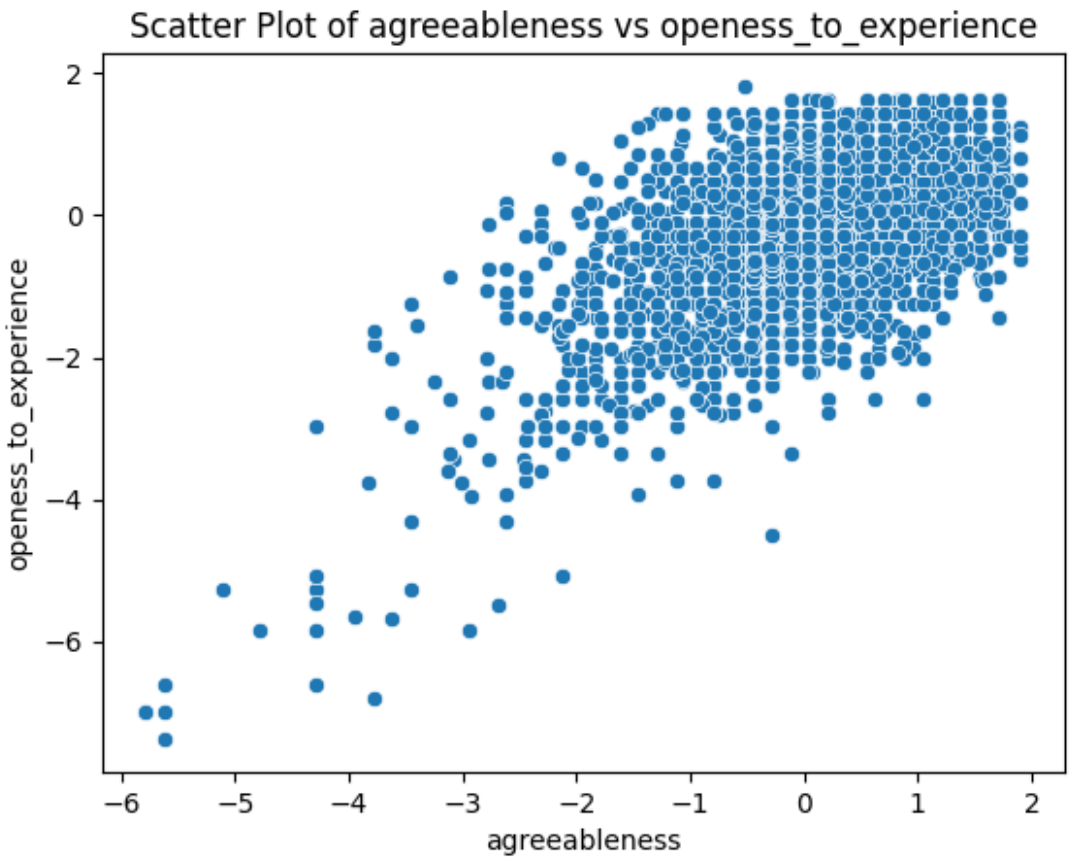

Scatter Plot of conscientiousness vs nueroticism
nueroticism
conscientiousness
3
2
1
0
−1
−2
−4
−3
−2
−1
0
1
2

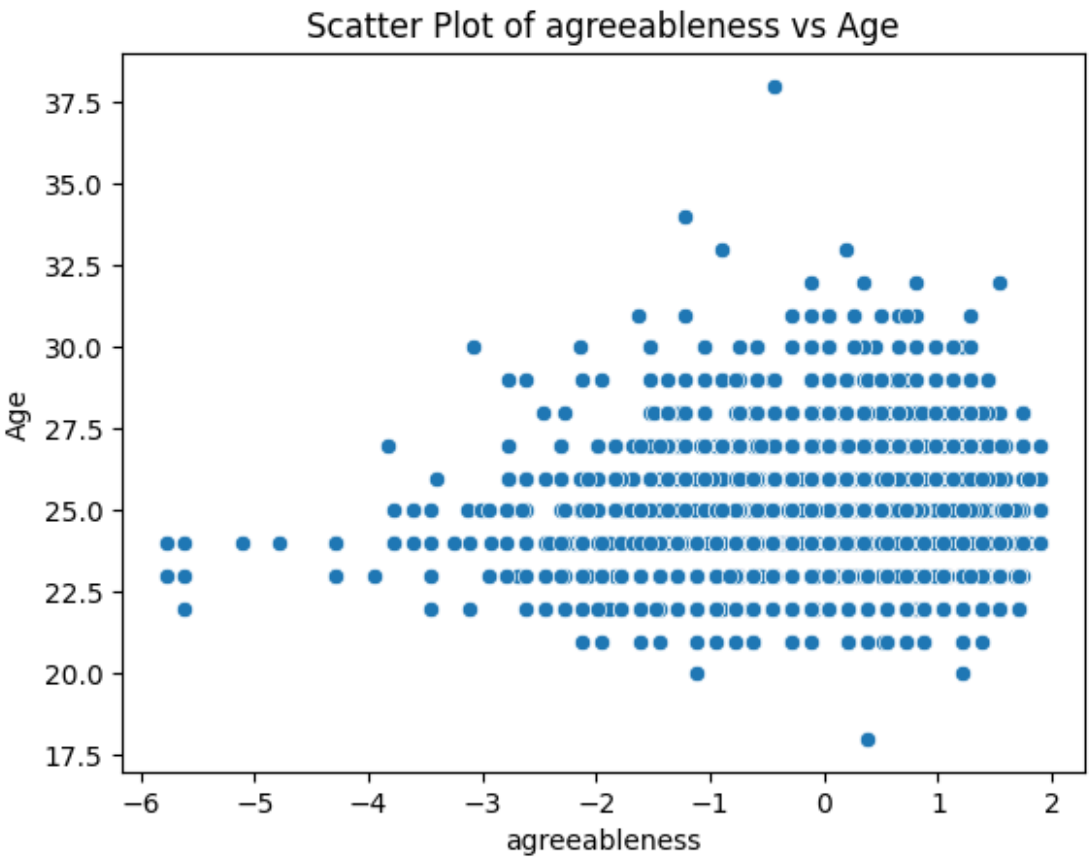

Scatter Plot of conscientiousness vs openess_to_experience
openess_to_experience
conscientiousness
2
0
−2
−4
−6
−4
−3
−2
−1
0
1
2

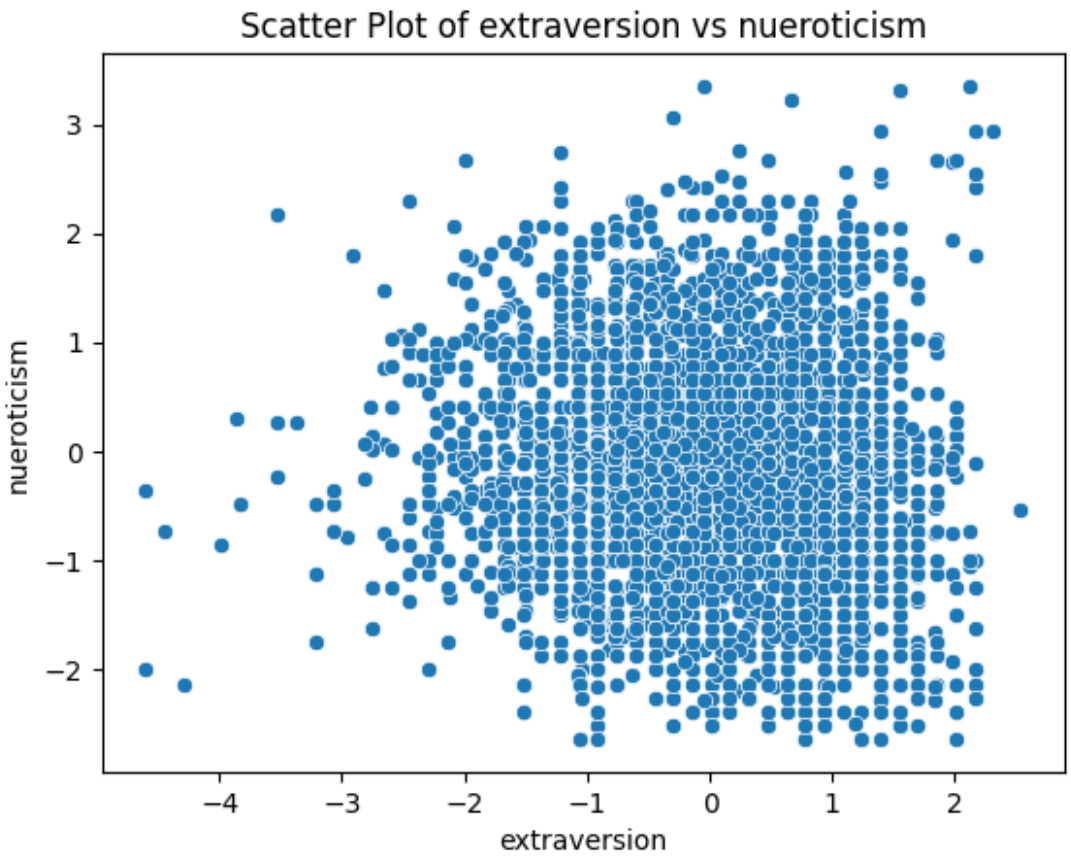

Scatter Plot of conscientiousness vs Age
Age
conscientiousness
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
−4
−3
−2
−1
0
1
2

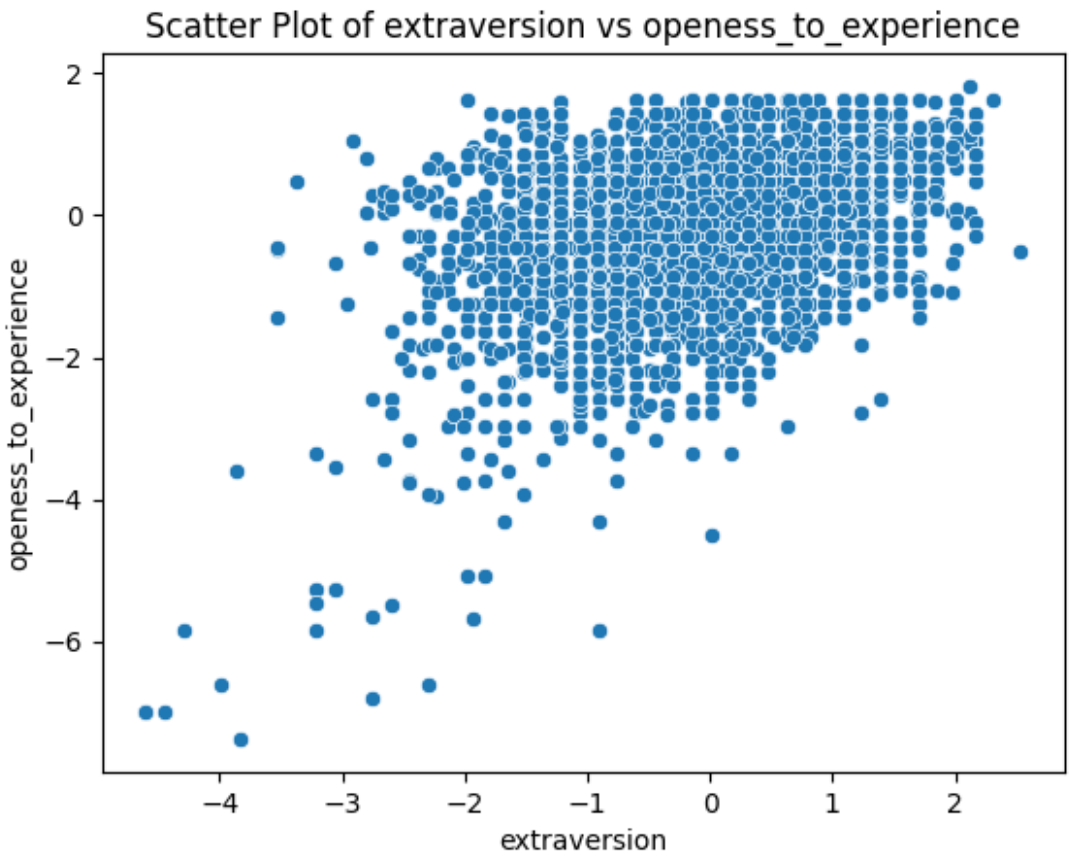

Scatter Plot of agreeableness vs extraversion
extraversion
agreeableness
2
1
0
−1
−2
−3
−4
−6
−5
−4
−3
−2
−1
0
1
2

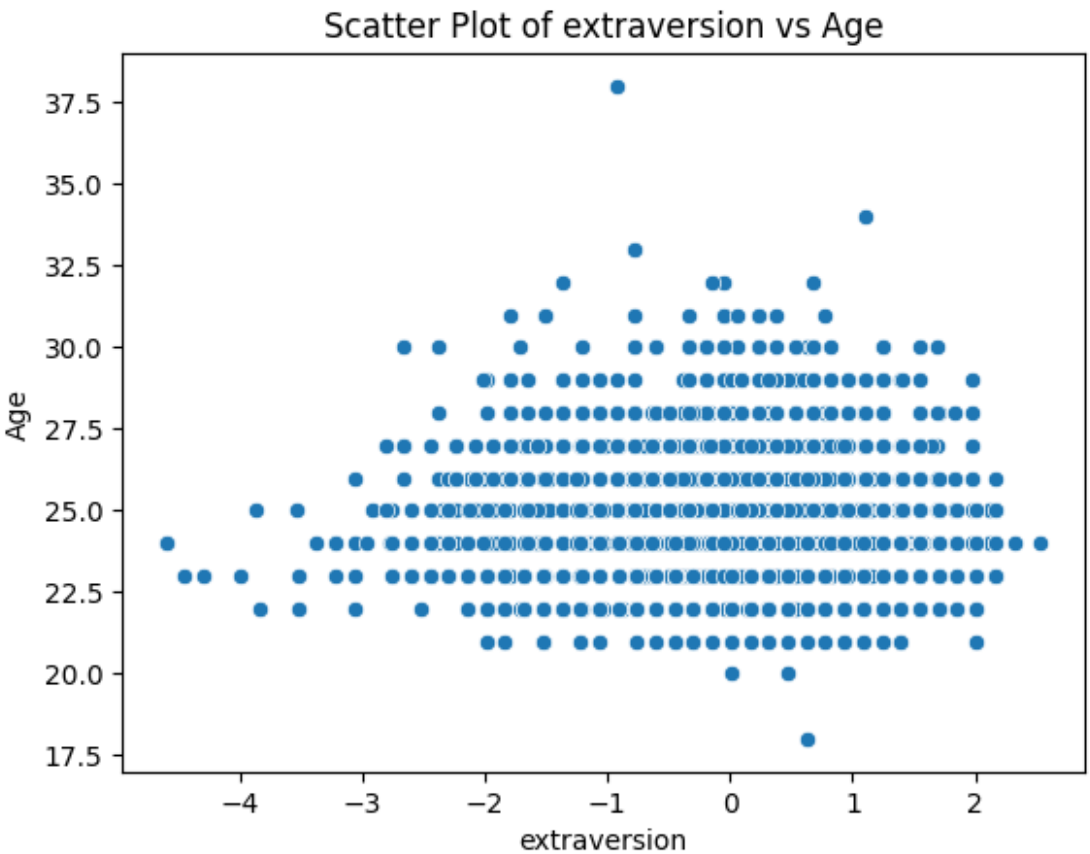

Scatter Plot of agreeableness vs nueroticism
nueroticism
agreeableness
3
2
1
0
−1
−2
−6
−5
−4
−3
−2
−1
0
1
2

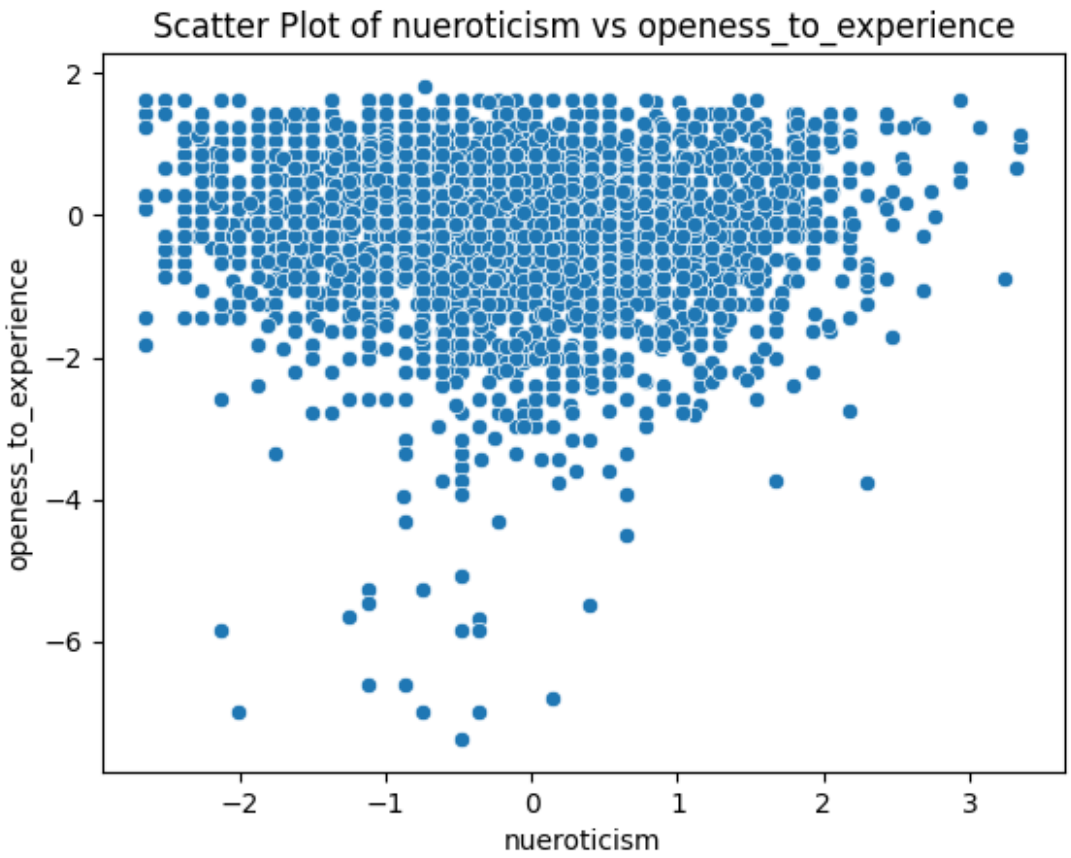

Scatter Plot of agreeableness vs openess_to_experience
openess_to_experience
agreeableness
2
0
−2
−4
−6
−6
−5
−4
−3
−2
−1
0
1
2

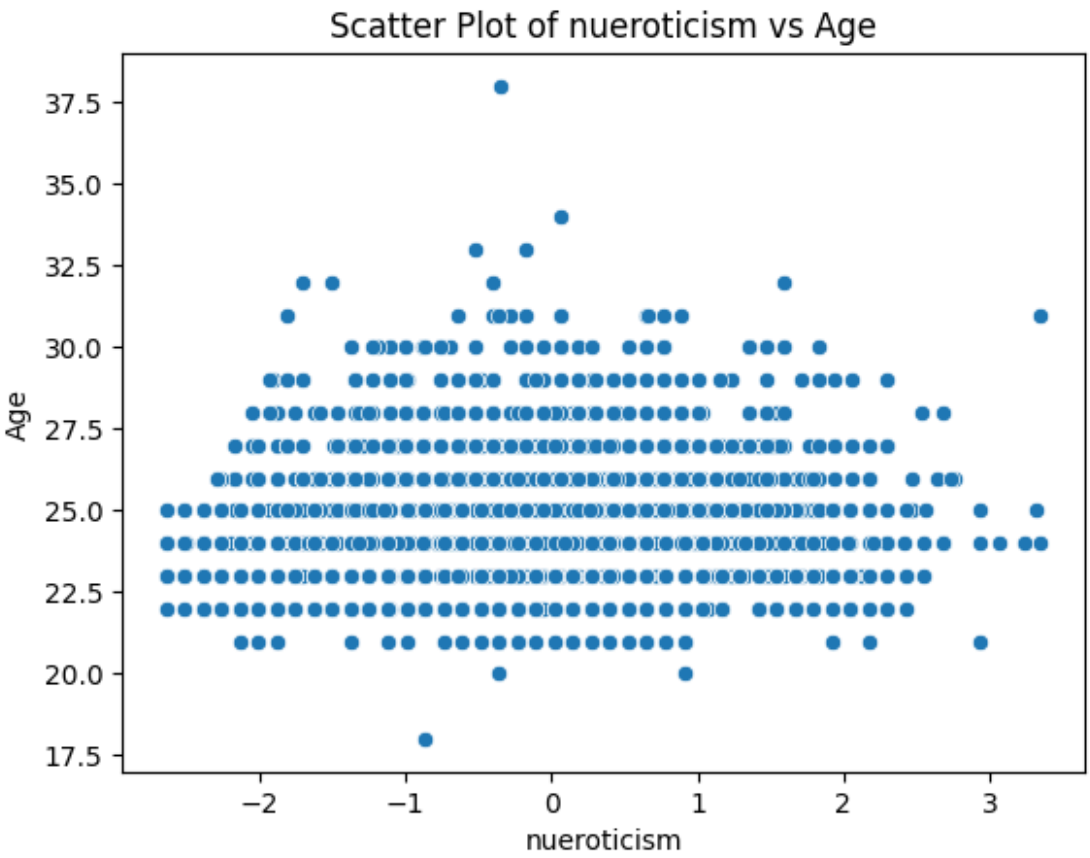

Scatter Plot of agreeableness vs Age
Age
agreeableness
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
−6
−5
−4
−3
−2
−1
0
1
2


Scatter Plot of extraversion vs nueroticism
nueroticism
extraversion
3
2
1
0
−1
−2
−4
−3
−2
−1
0
1
2

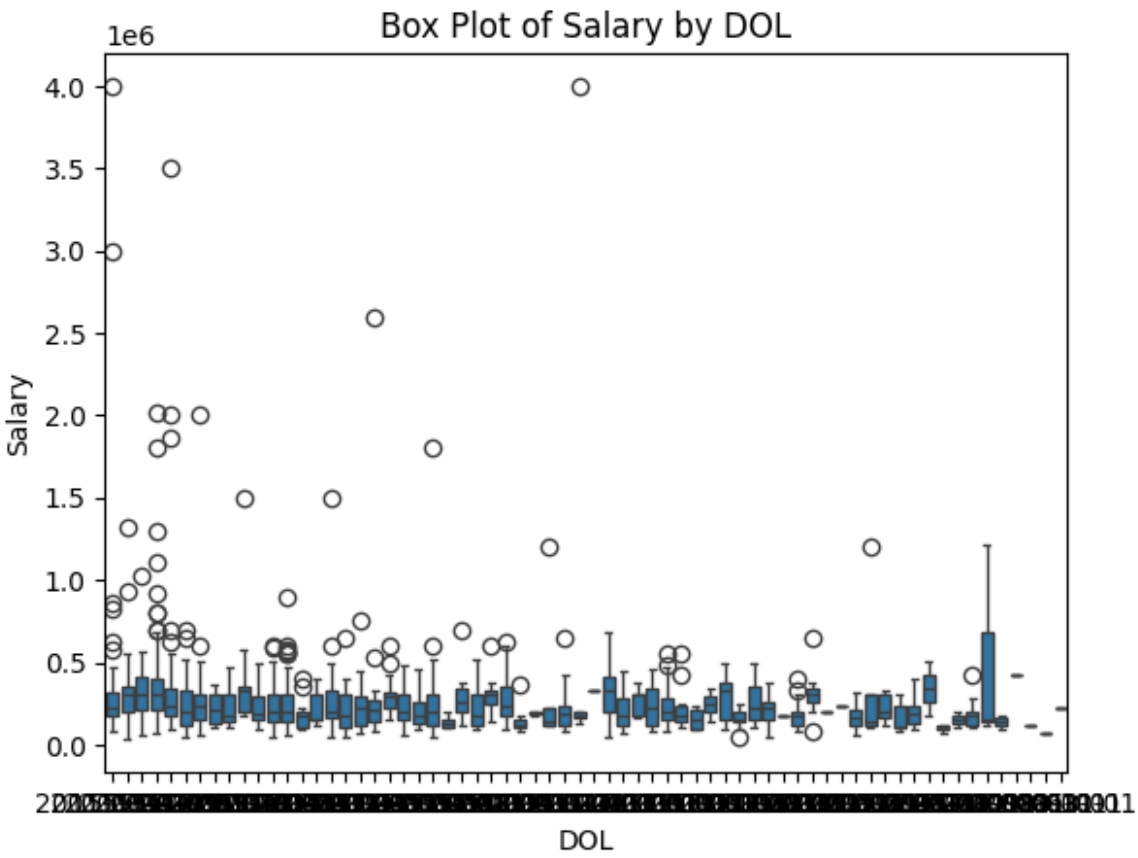

Scatter Plot of extraversion vs openness_to_experience
openness_to_experience
extraversion
2
0
−2
−4
−6
−4
−3
−2
−1
0
1
2

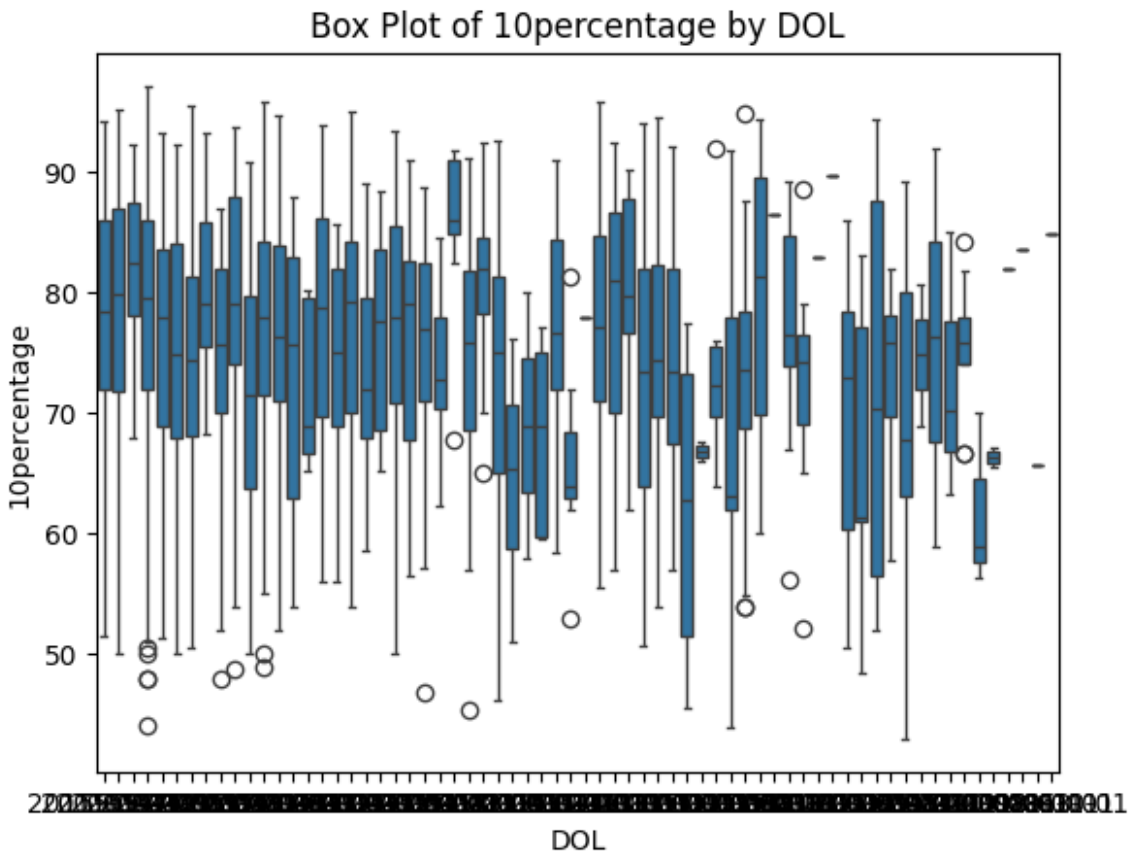

Scatter Plot of extraversion vs Age
Age
extraversion
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
−4
−3
−2
−1
0
1
2

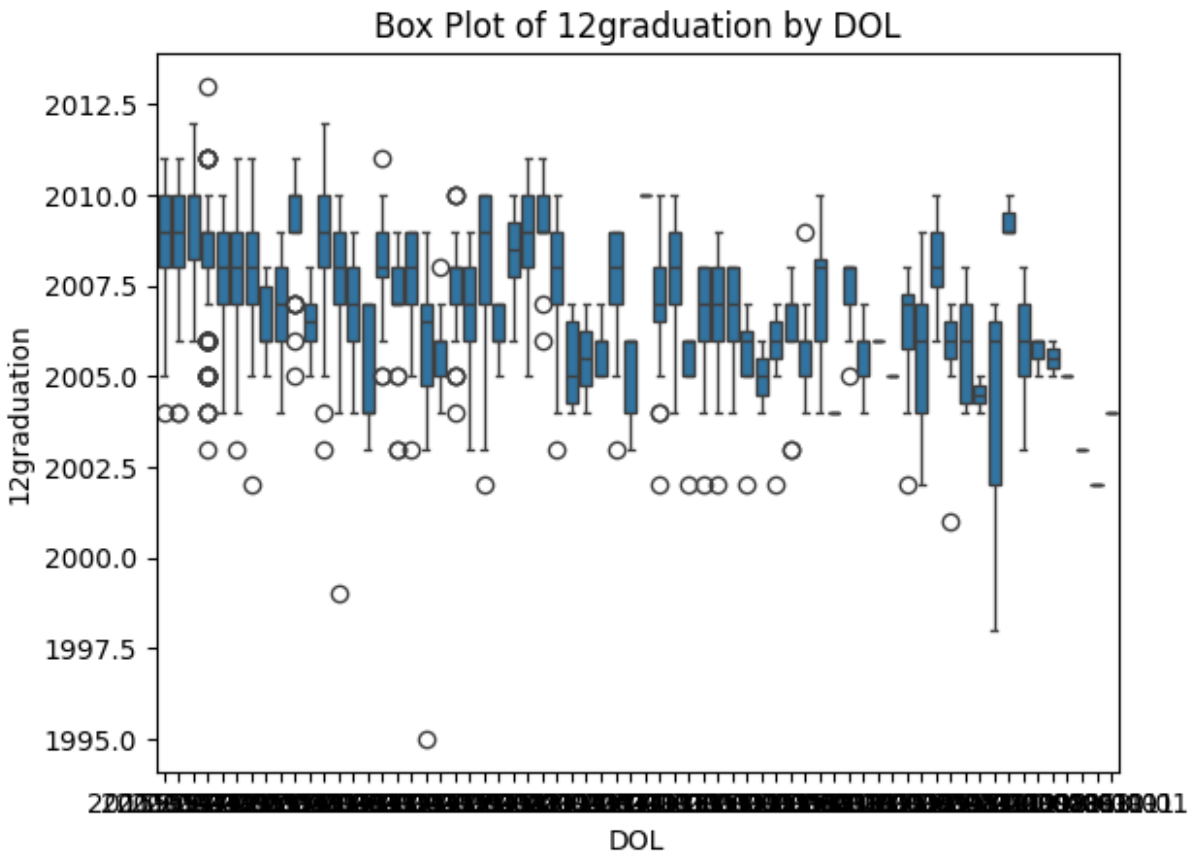

Scatter Plot of nueroticism vs openess_to_experience
openess_to_experience
nueroticism
2
0
−2
−4
−6
−2
−1
0
1
2
3

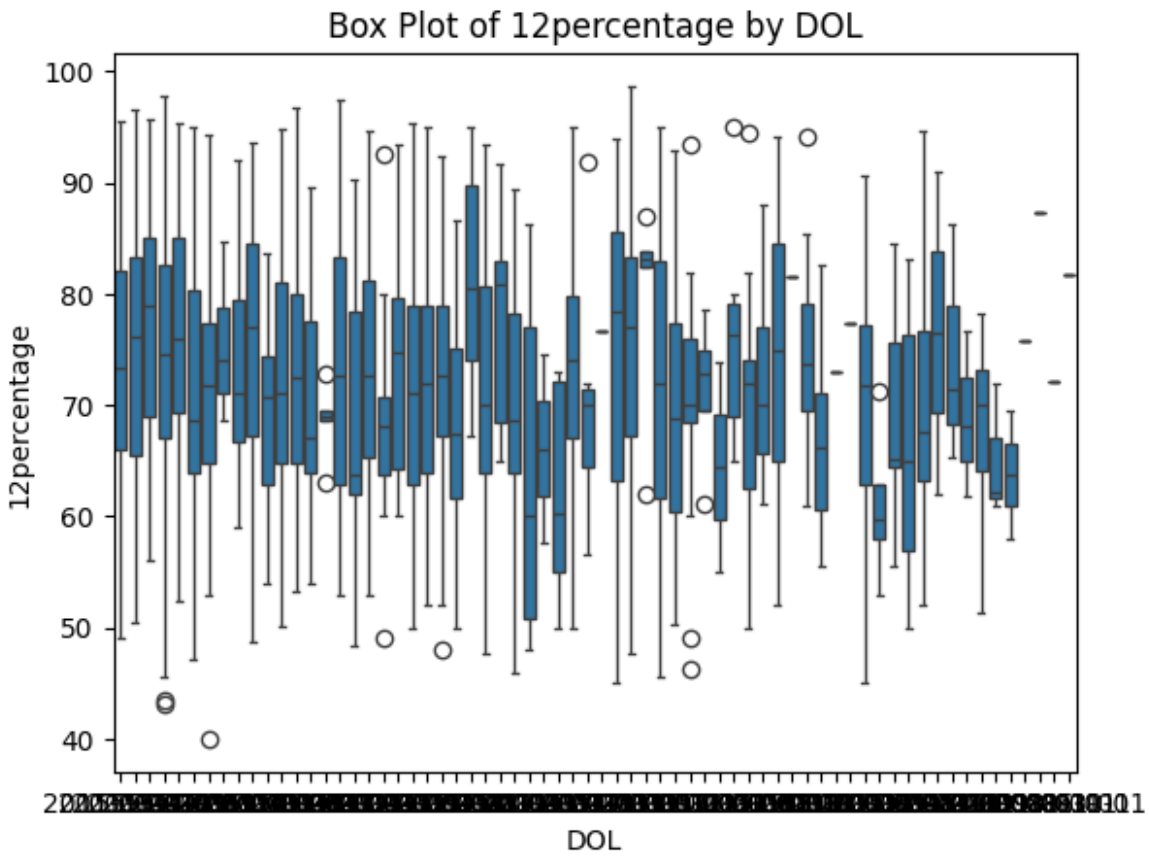

Scatter Plot of nueroticism vs Age
Age
nueroticism
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
−2
−1
0
1
2
3

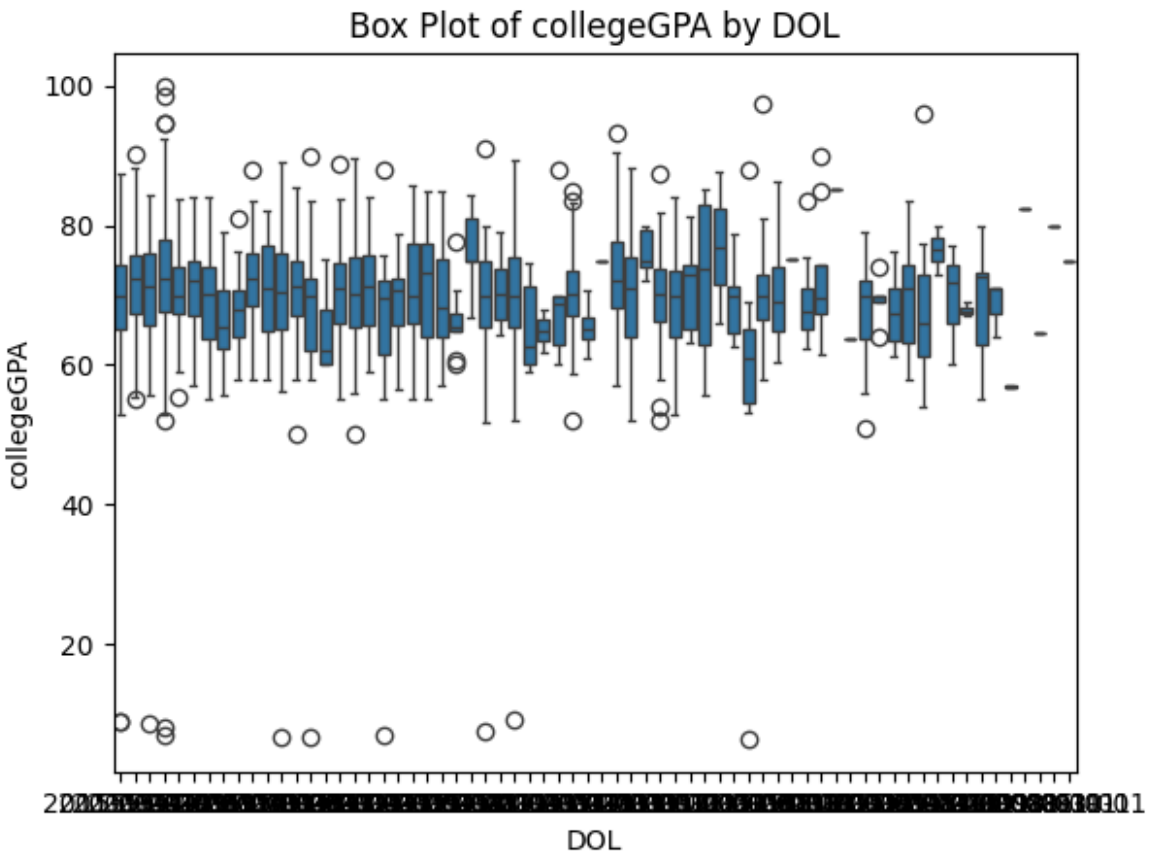
```
# Box Plots for Categorical vs Numerical Columns
for col in categorical_columns:
    for num_col in numerical_columns:
        plt.figure()
        sns.boxplot(x=df[col], y=df[num_col])
        plt.title(f"Box Plot of {num_col} by {col}")
        plt.show()
```

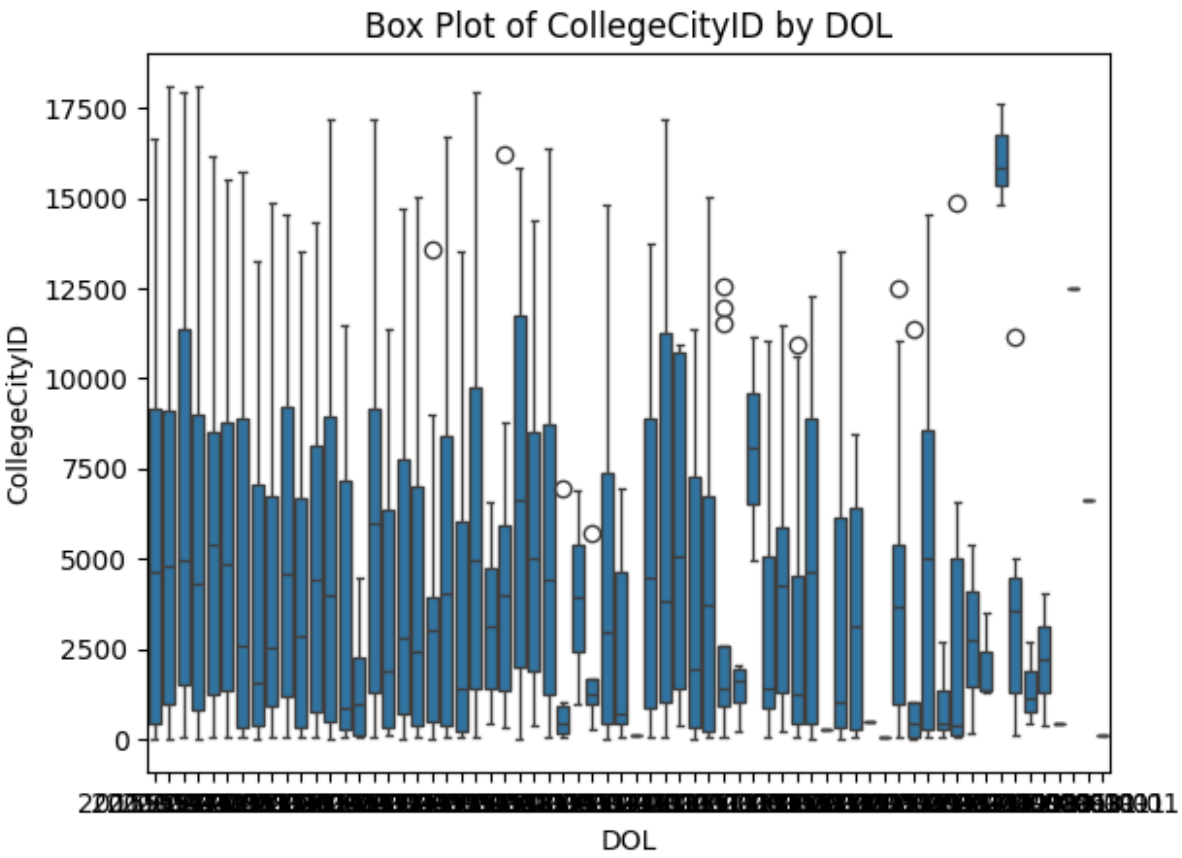

Box Plot of Salary by DOL
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Salary
DOL

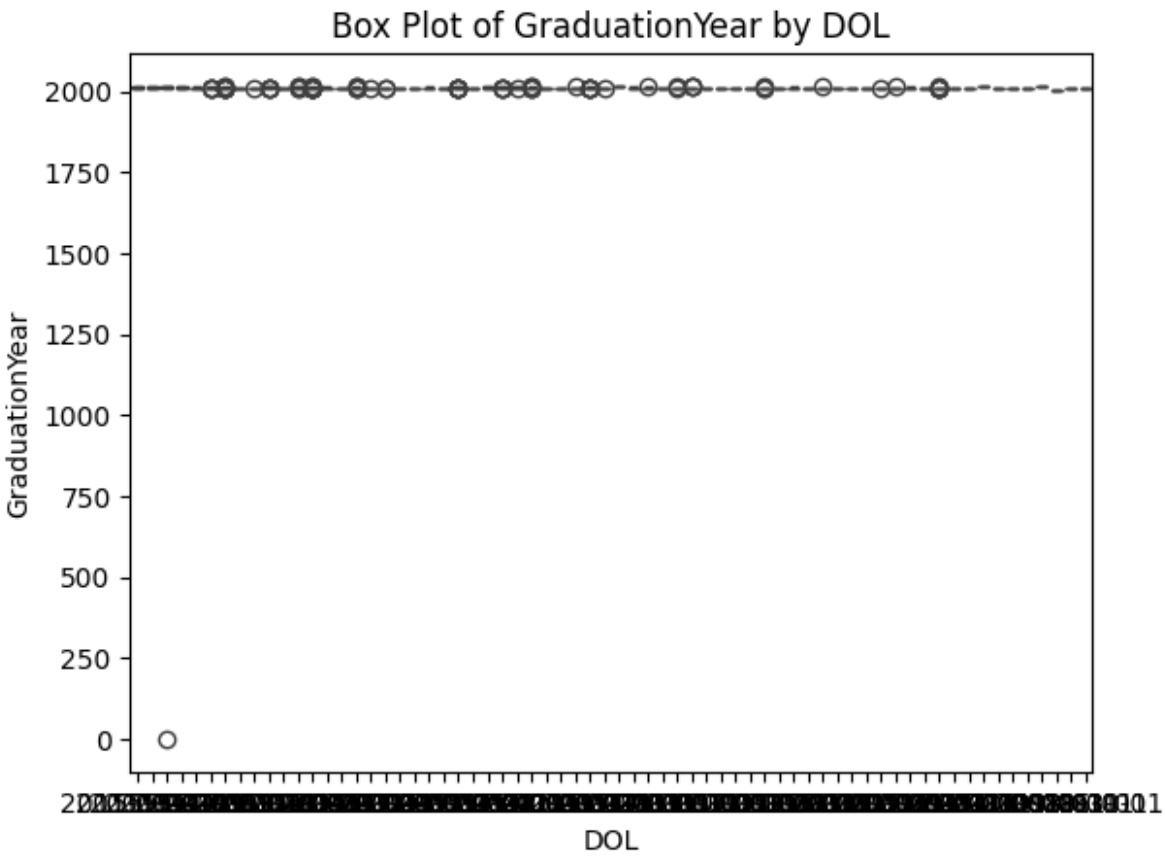

Box Plot of 10percentage by DOL
10percentage
90
80
70
60
50
DOL

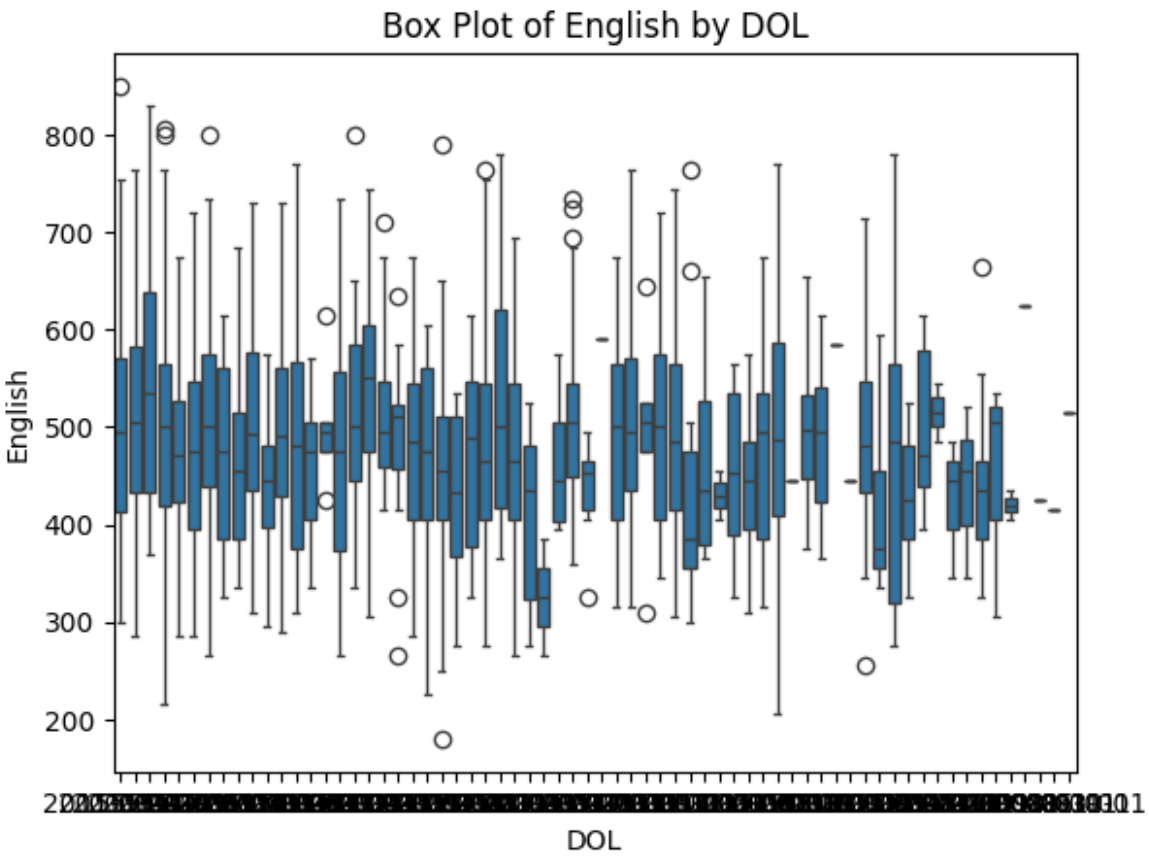

Box Plot of 12graduation by DOL
12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
DOL

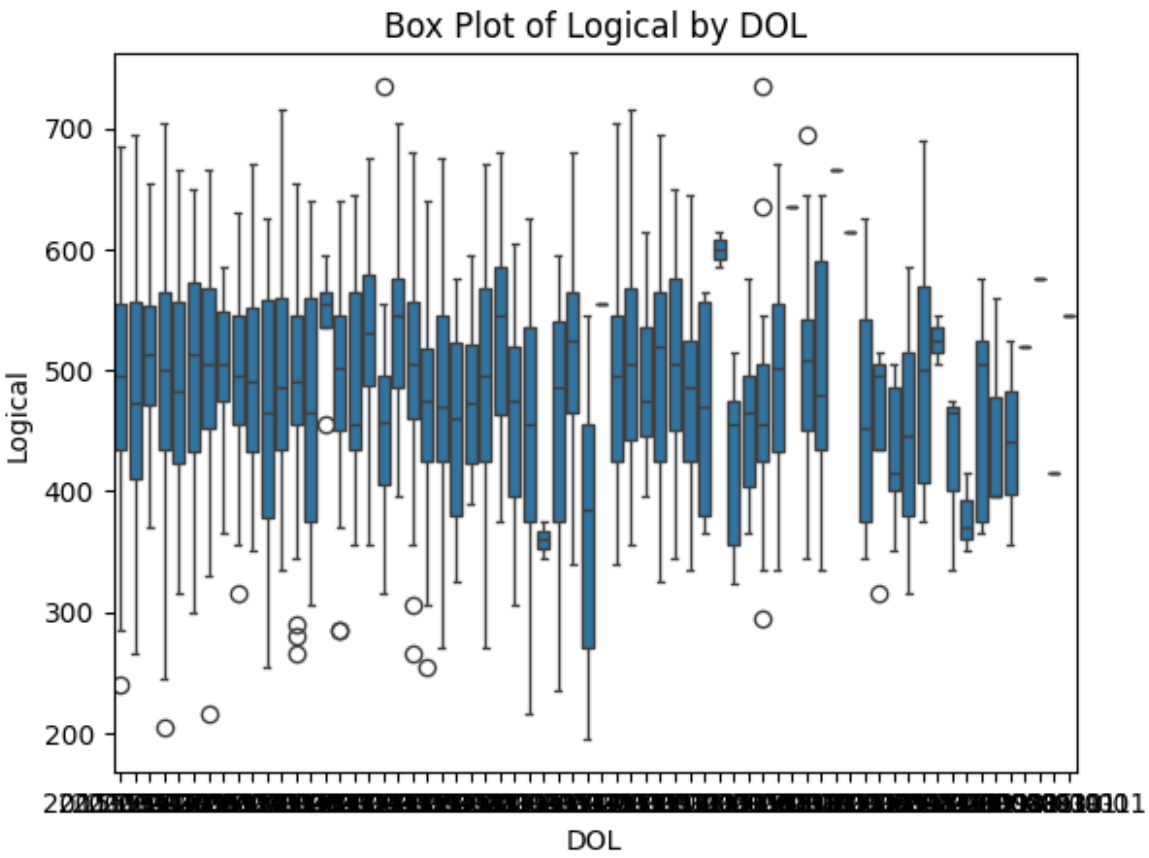

Box Plot of 12percentage by DOL
12percentage
100
90
80
70
60
50
40
DOL

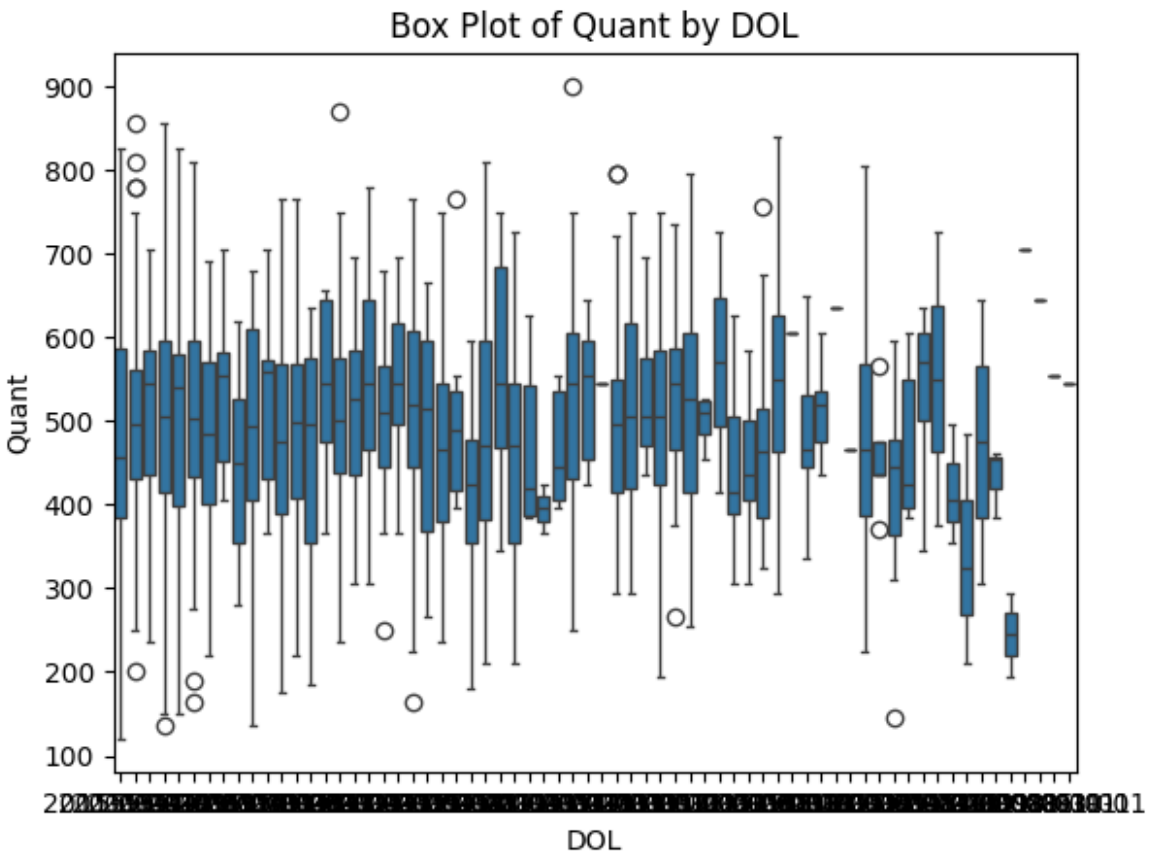

Box Plot of collegeGPA by DOL
100
80
60
40
20
collegeGPA
DOL

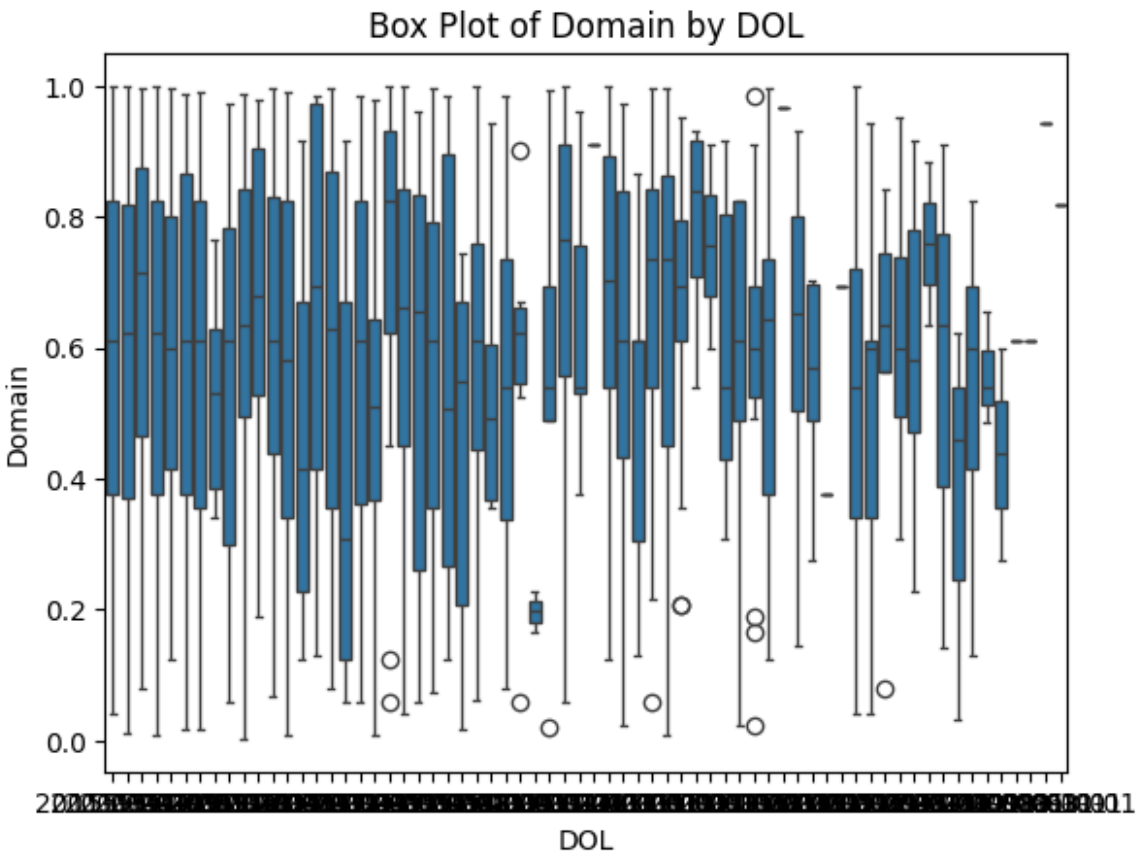

Box Plot of CollegeCityID by DOL
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
DOL

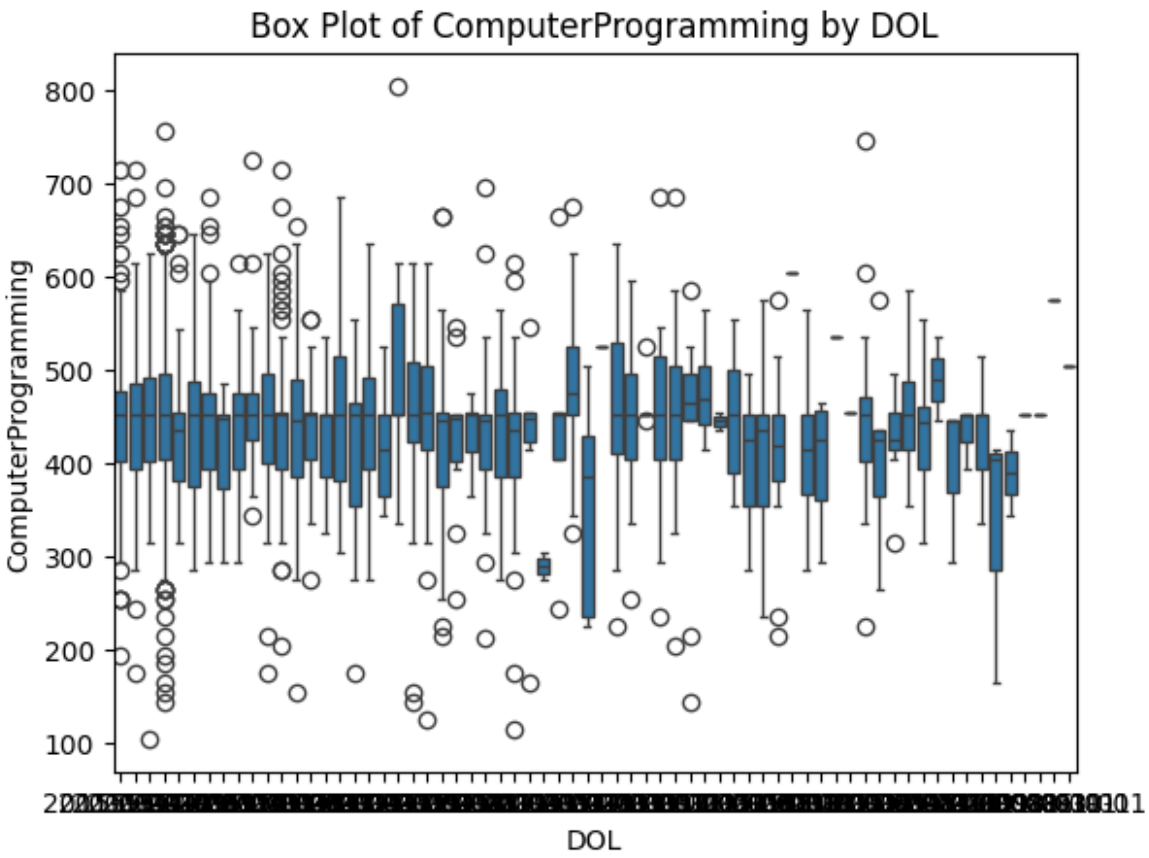

Box Plot of GraduationYear by DOL
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
DOL

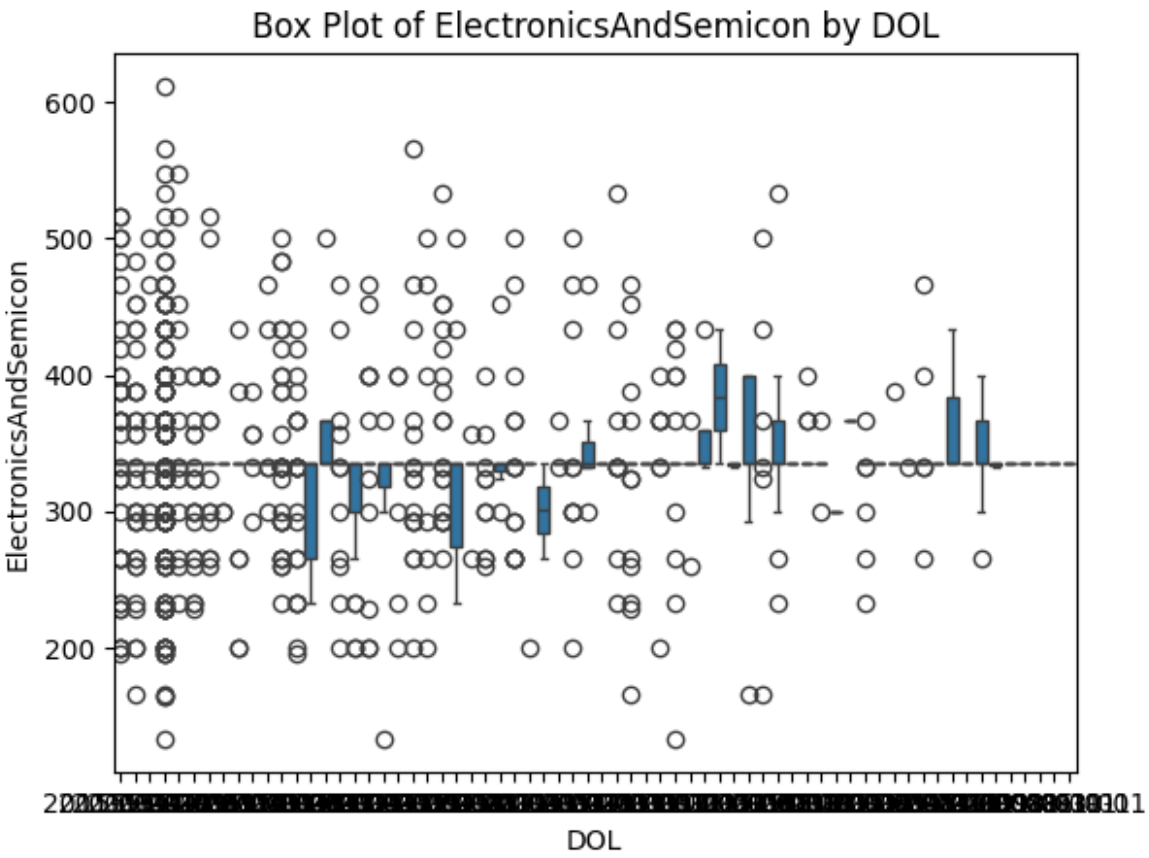

Box Plot of English by DOL
English
800
700
600
500
400
300
200
DOL

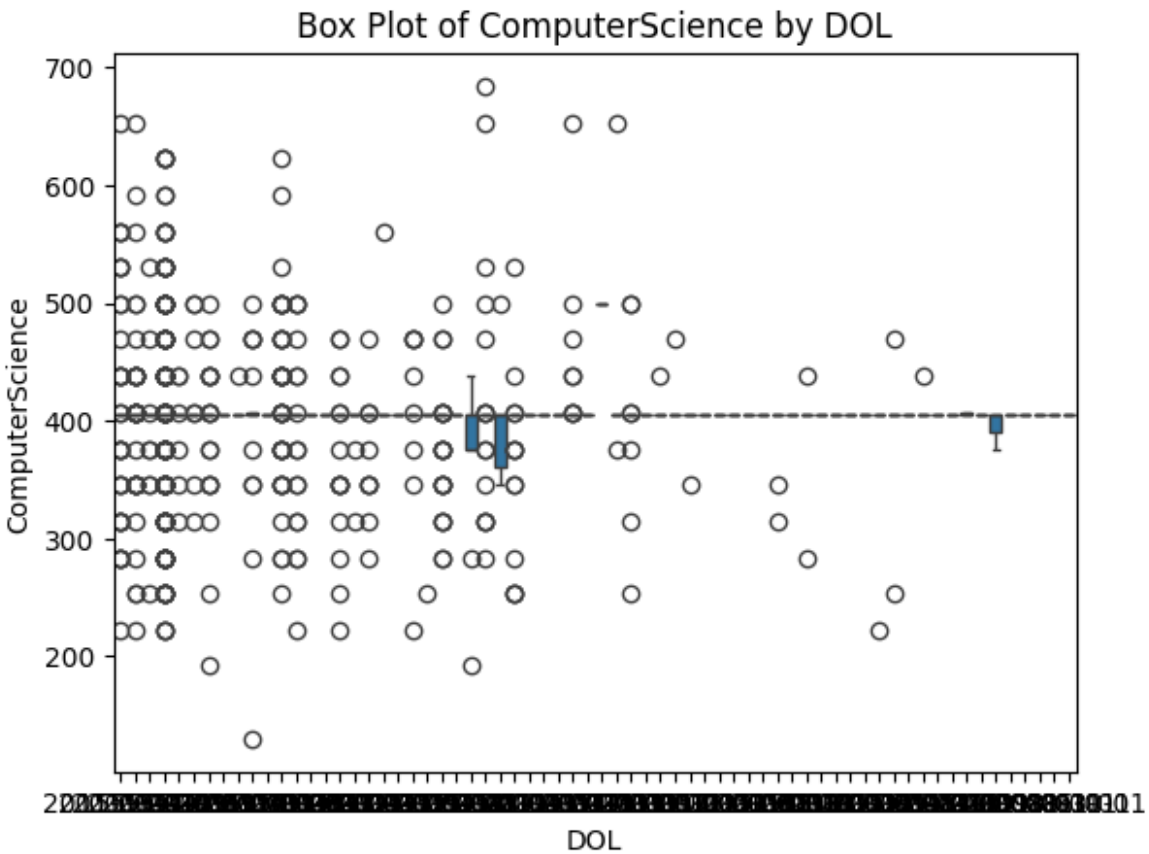

Box Plot of Logical by DOL
Logical
700
600
500
400
300
200
DOL

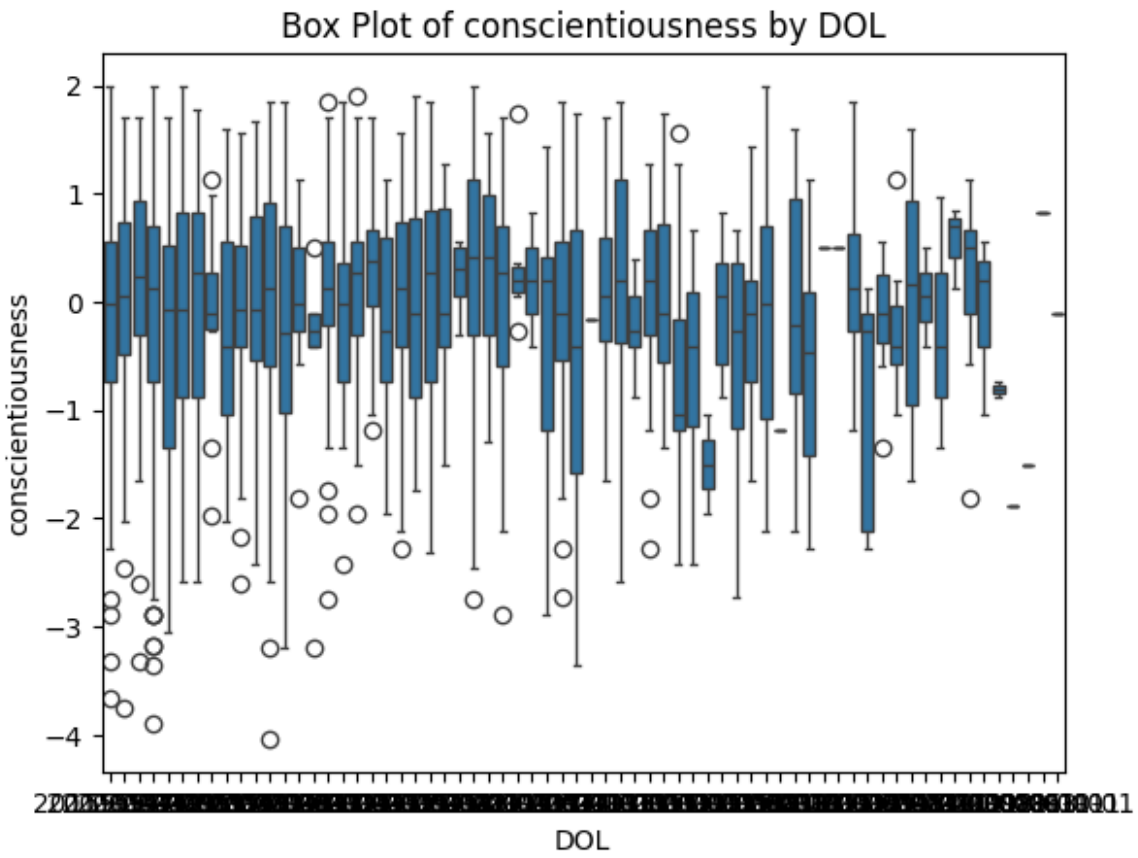

Box Plot of Quant by DOL
Quant
900
800
700
600
500
400
300
200
100
DOL

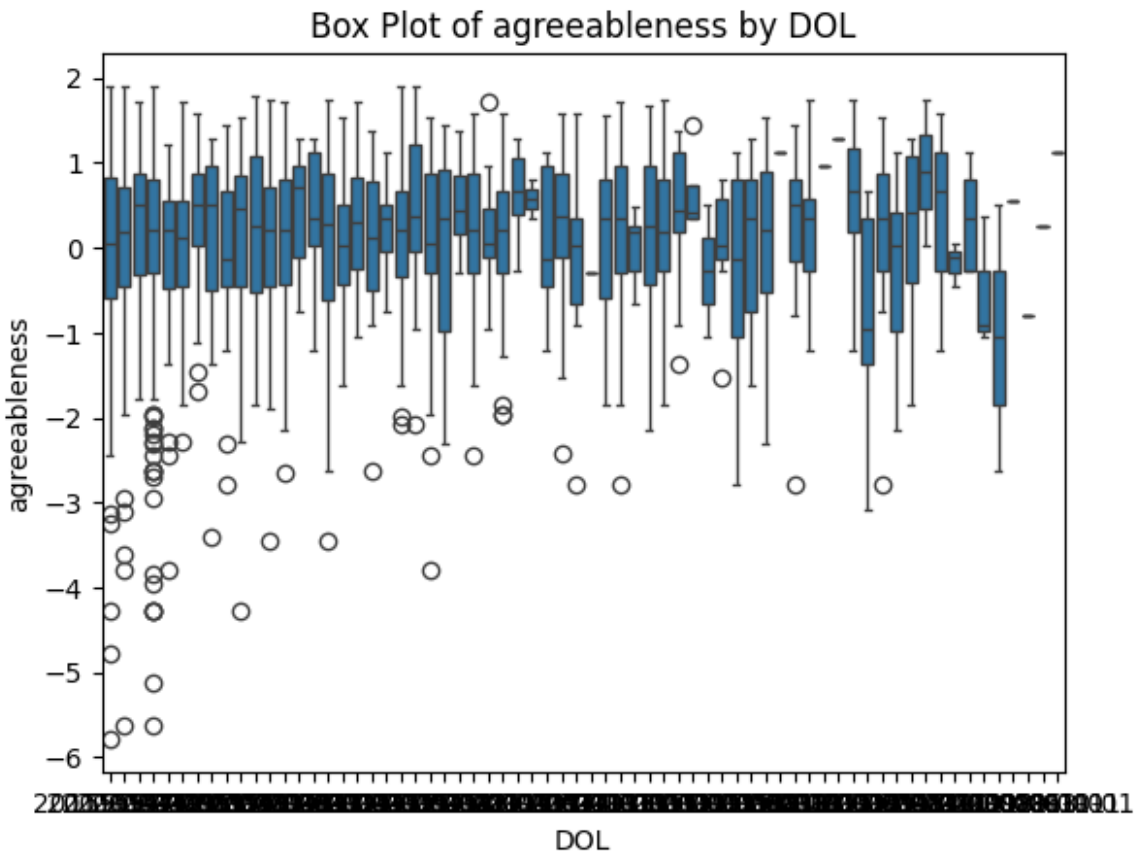

Box Plot of Domain by DOL
Domain
1.0
0.8
0.6
0.4
0.2
0.0
DOL

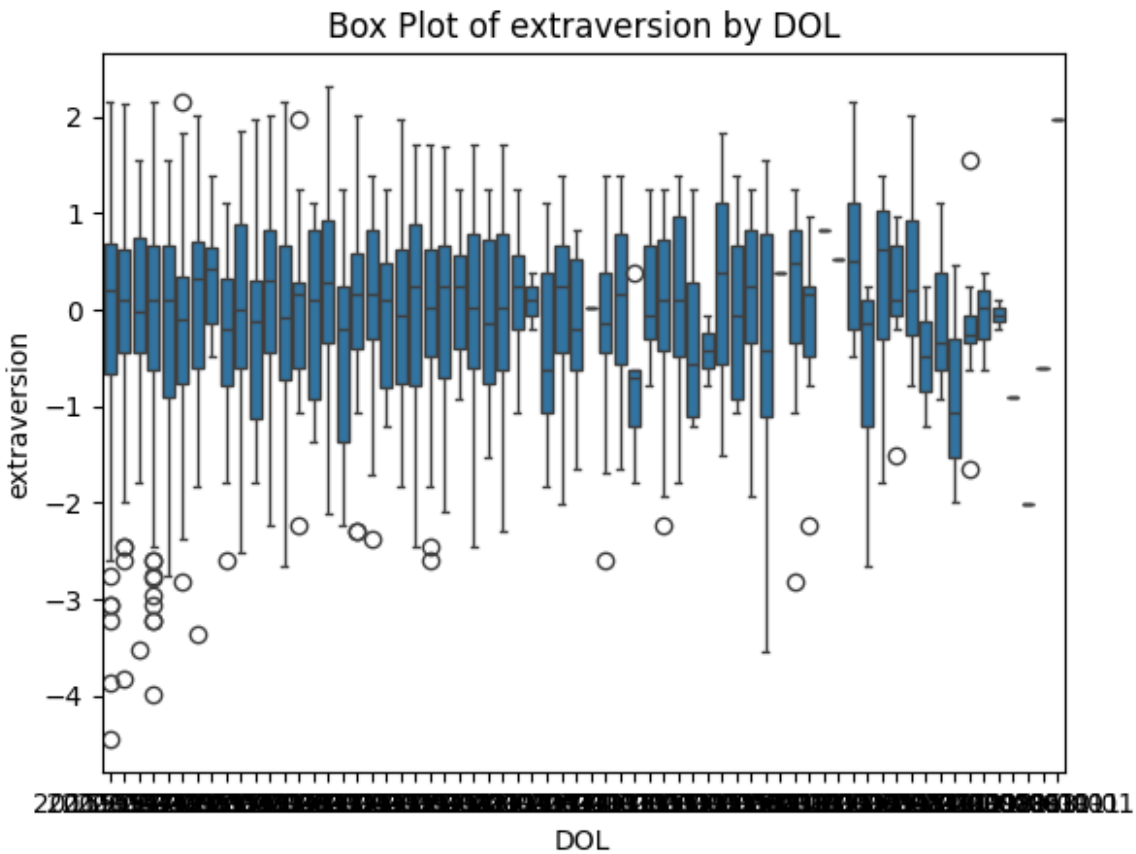

Box Plot of ComputerProgramming by DOL
ComputerProgramming
800
700
600
500
400
300
200
100
DOL

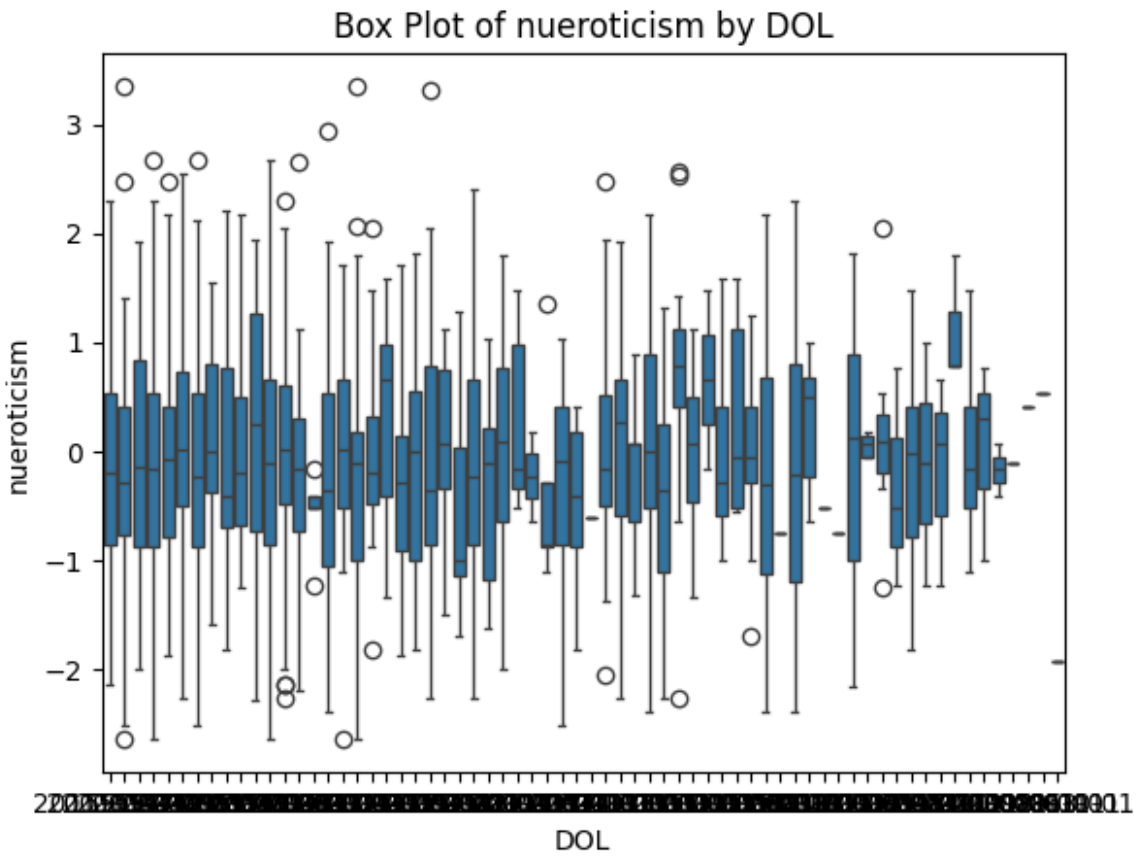

Box Plot of ElectronicsAndSemicon by DOL
ElectronicsAndSemicon
600
500
400
300
200
DOL

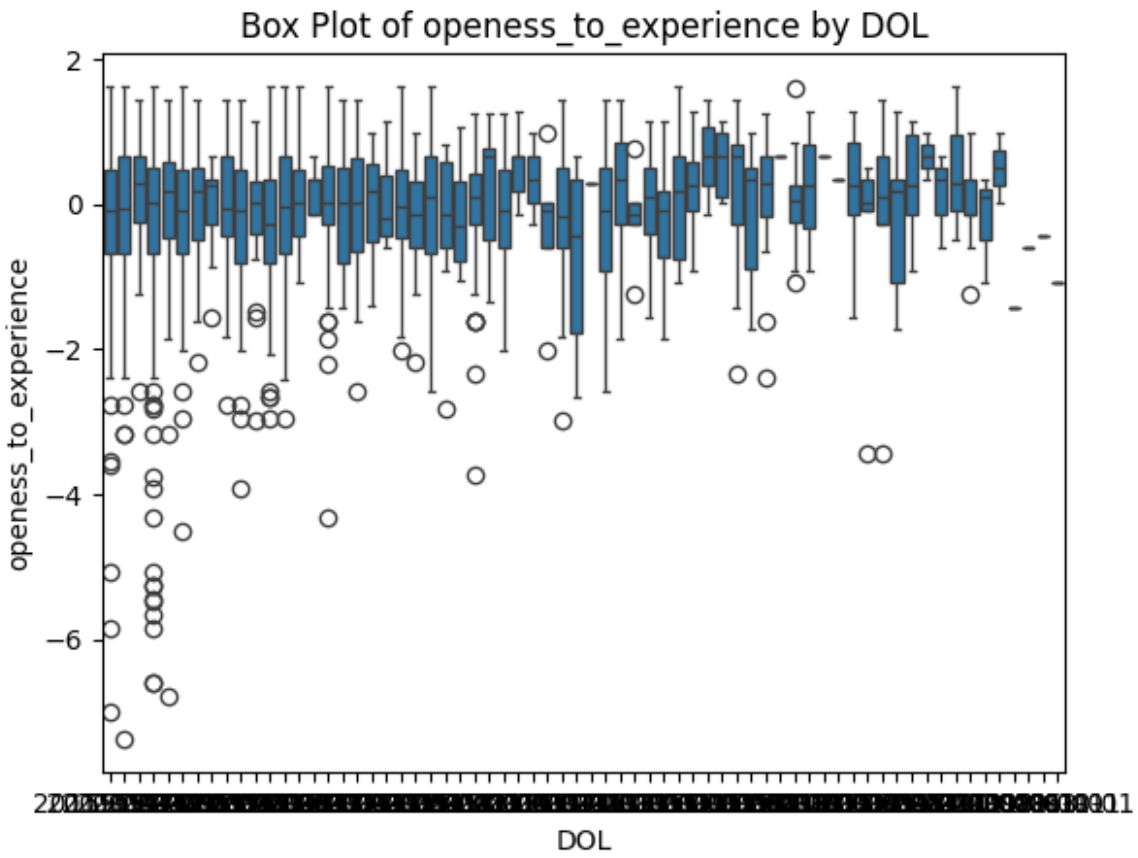

Box Plot of ComputerScience by DOL
ComputerScience
700
600
500
400
300
200
DOL

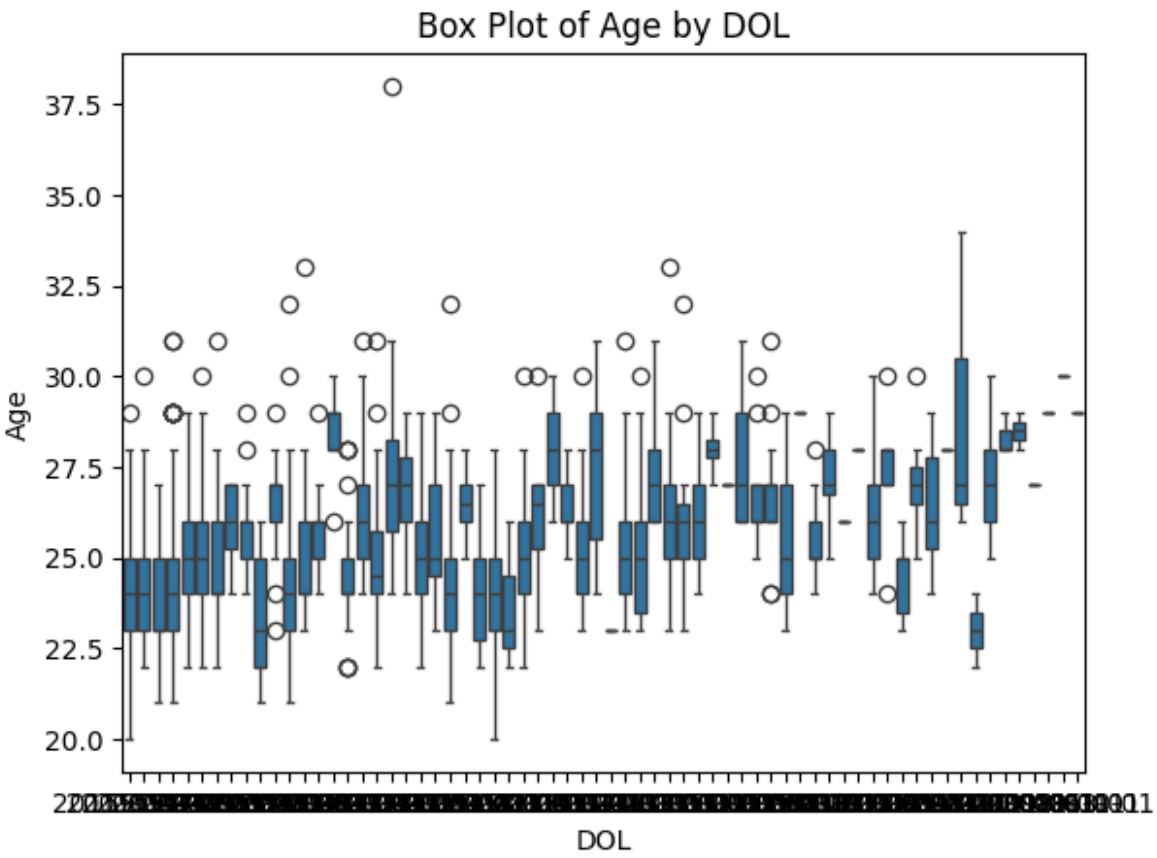

Box Plot of conscientiousness by DOL
conscientiousness
2
1
0
−1
−2
−3
−4
DOL

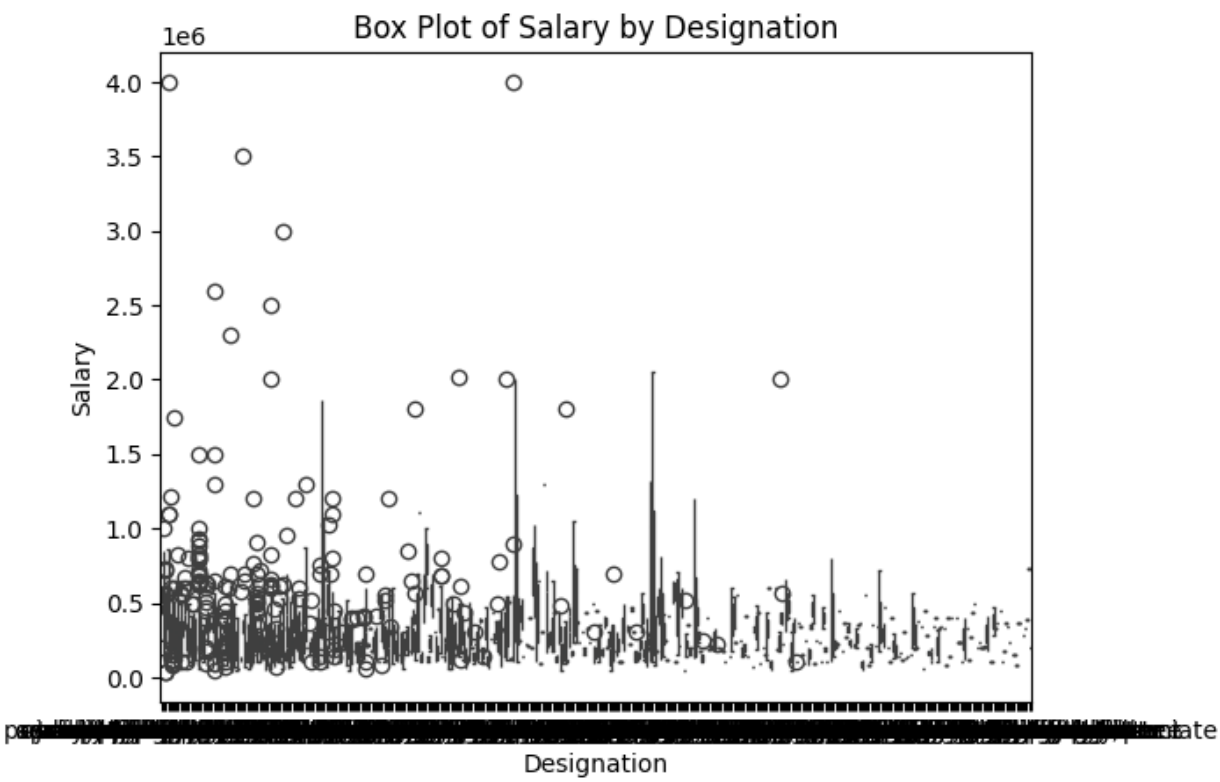

Box Plot of agreeableness by DOL
agreeableness
DOL
2
1
0
−1
−2
−3
−4
−5
−6

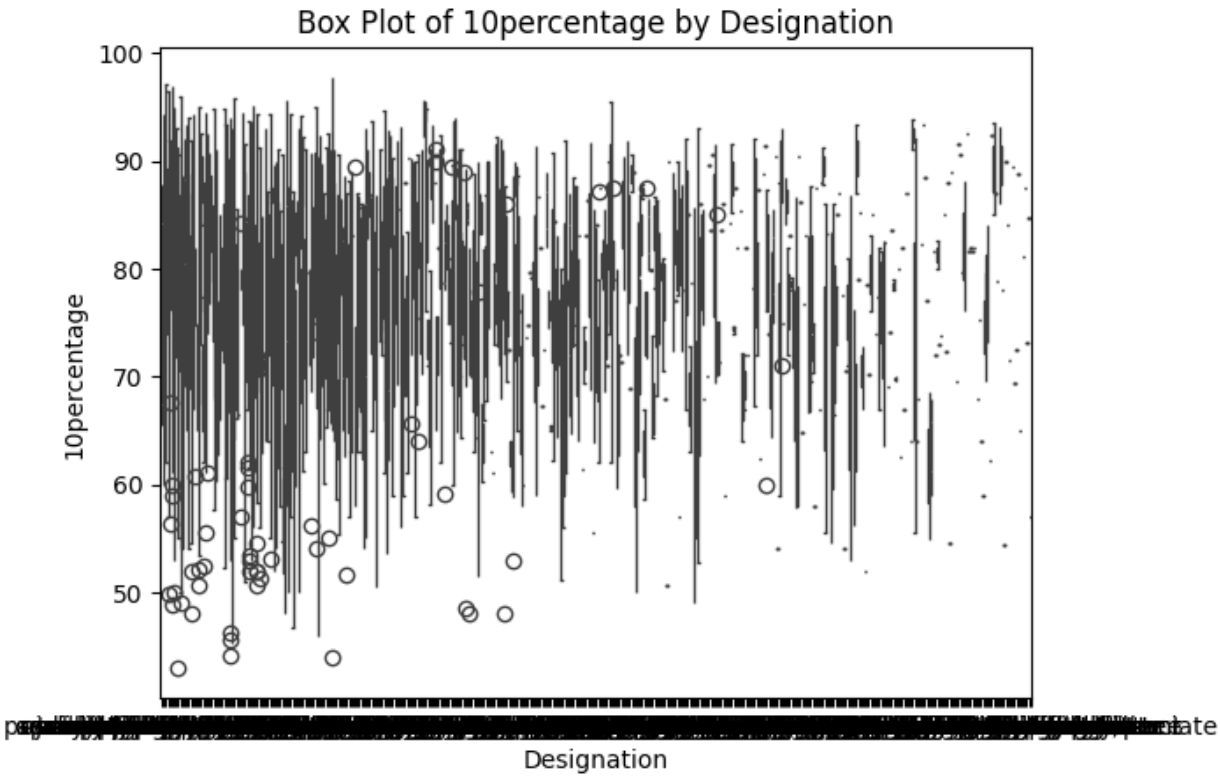

Box Plot of extraversion by DOL
extraversion
2
1
0
−1
−2
−3
−4
DOL

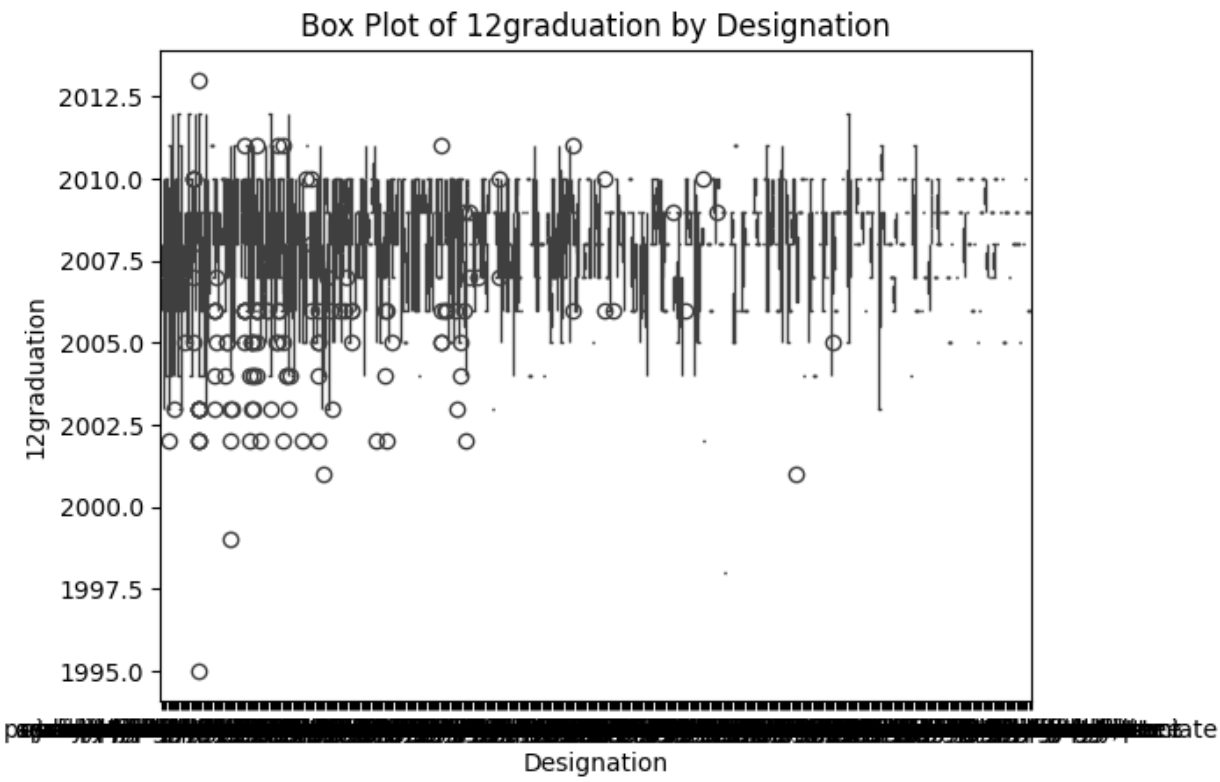

Box Plot of nueroticism by DOL
nueroticism
3
2
1
0
−1
−2
DOL

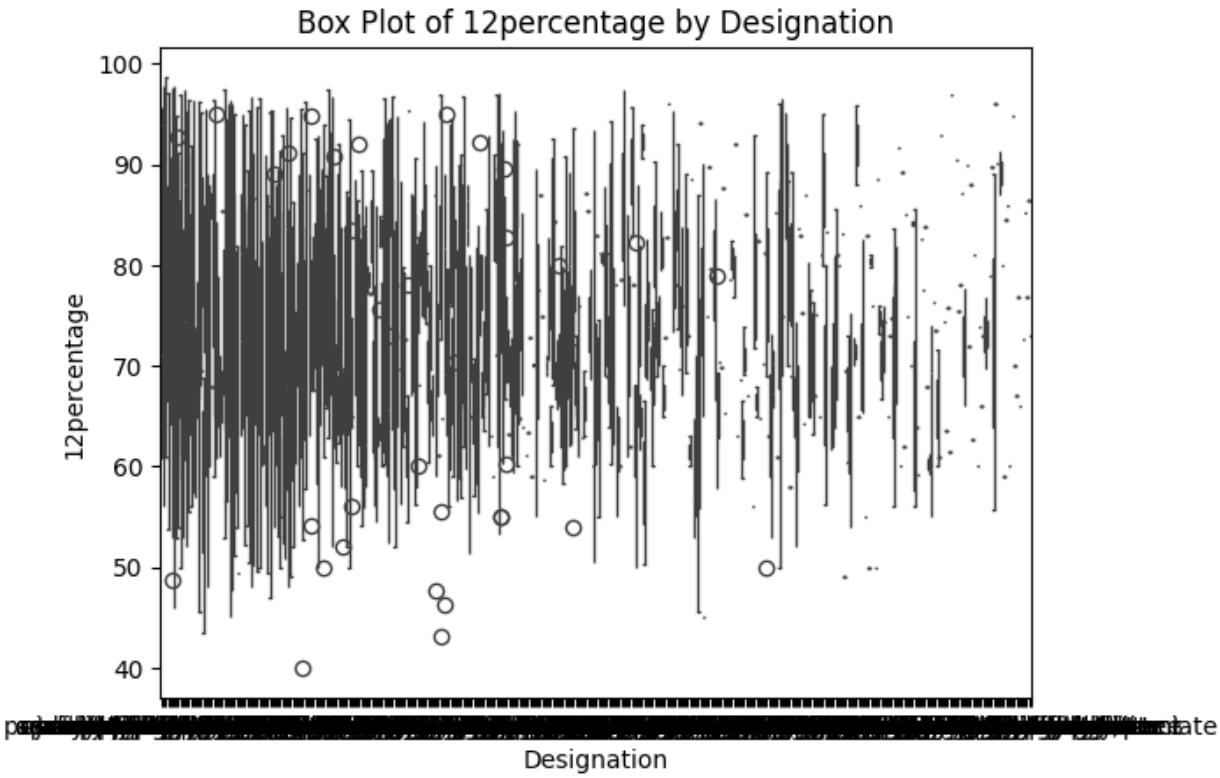

Box Plot of openess_to_experience by DOL
2
0
−2
−4
−6
openess_to_experience
DOL

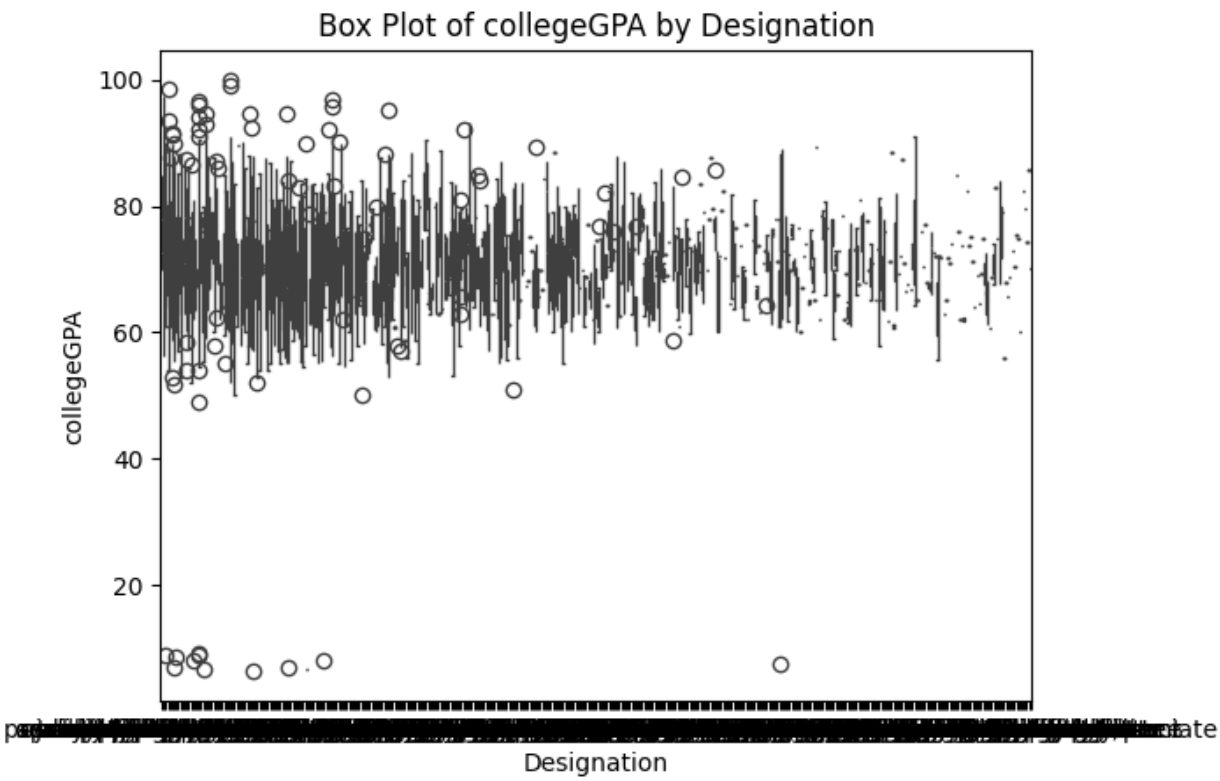

Box Plot of Age by DOL
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
DOL

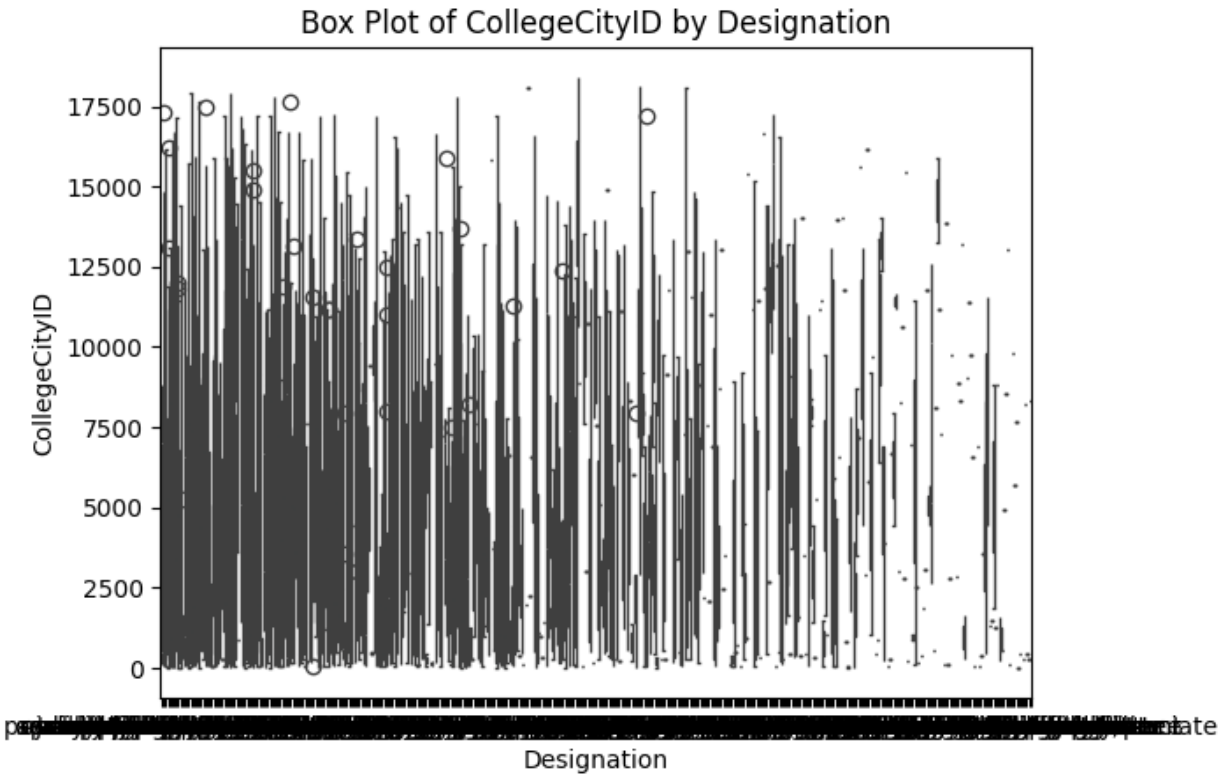

Box Plot of Salary by Designation
1e6
Salary
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Designation

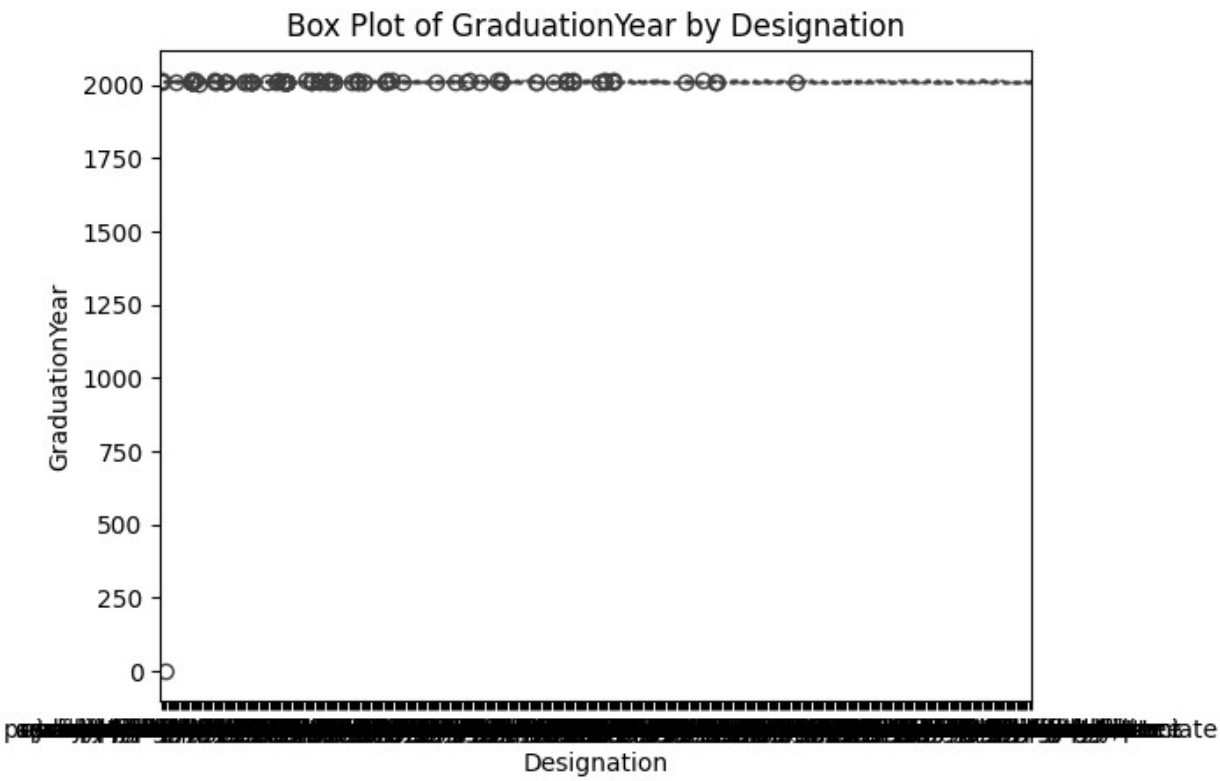

Box Plot of 10percentage by Designation
10percentage
100
90
80
70
60
50
Designation

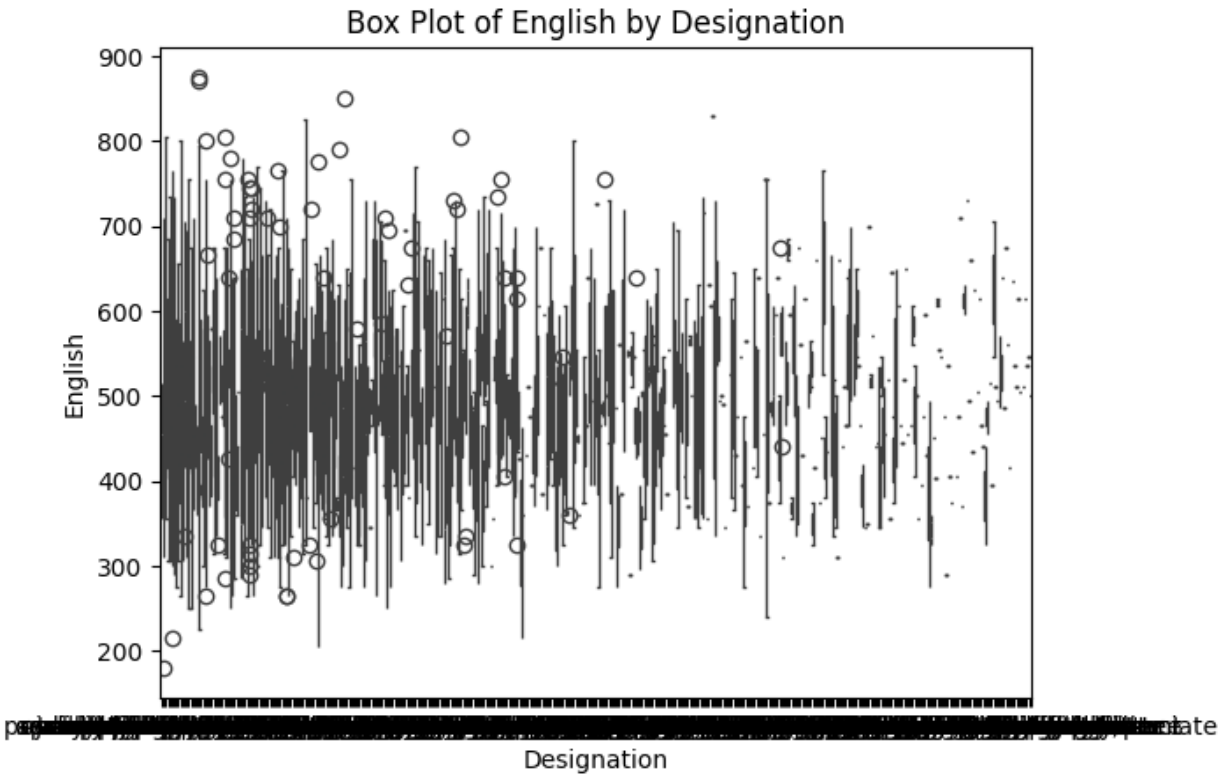

Box Plot of 12graduation by Designation
12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
Designation

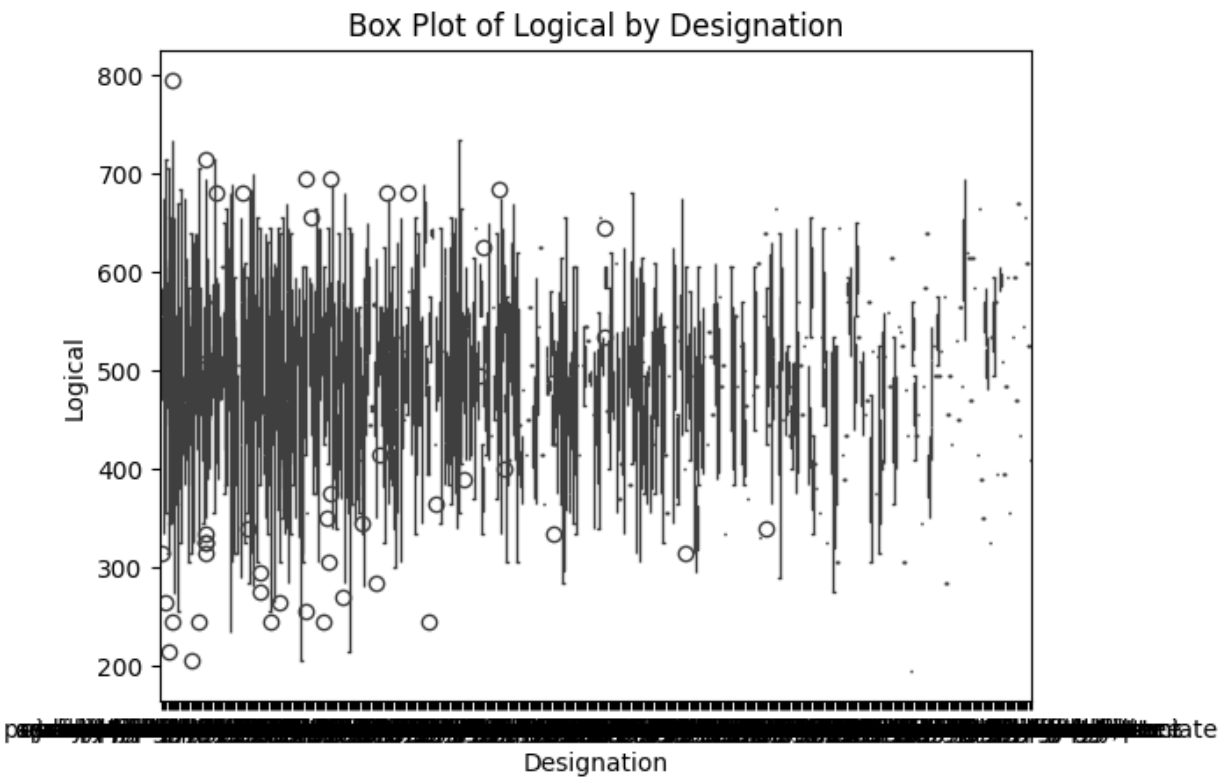

Box Plot of 12percentage by Designation
12percentage
100
90
80
70
60
50
40
Designation

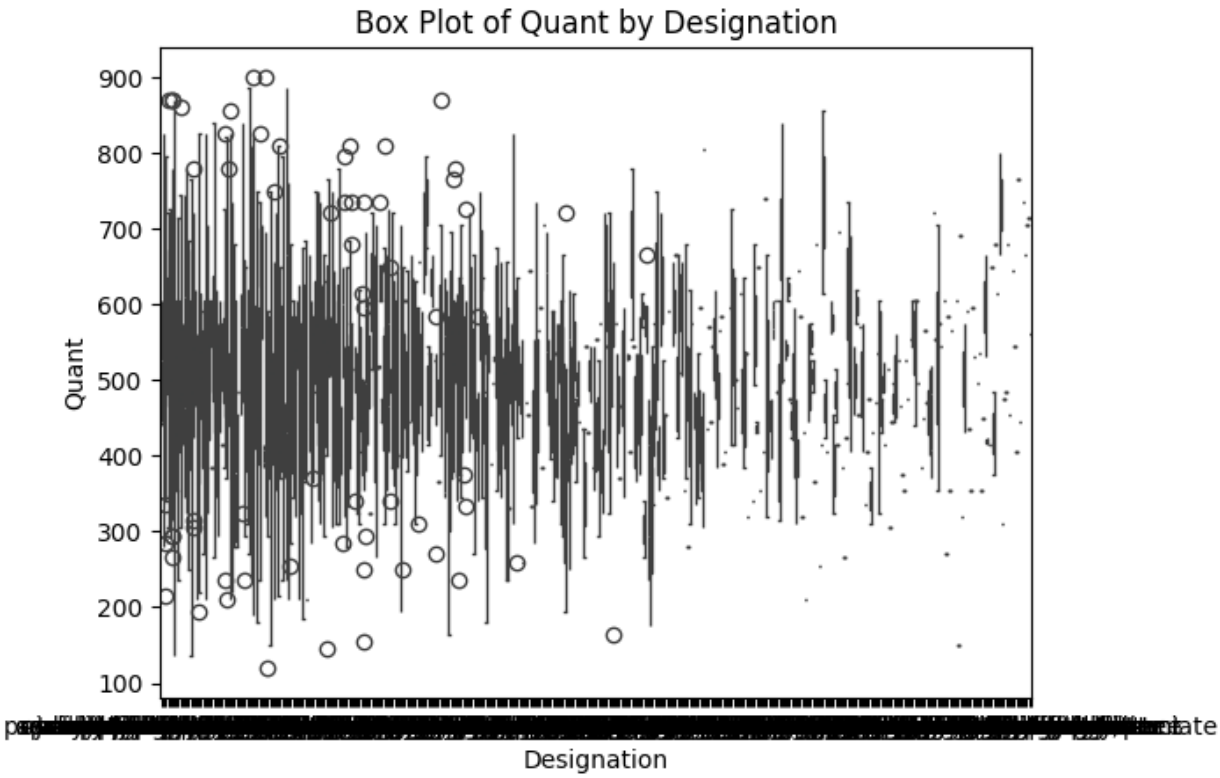

Box Plot of collegeGPA by Designation
collegeGPA
100
80
60
40
20
Designation

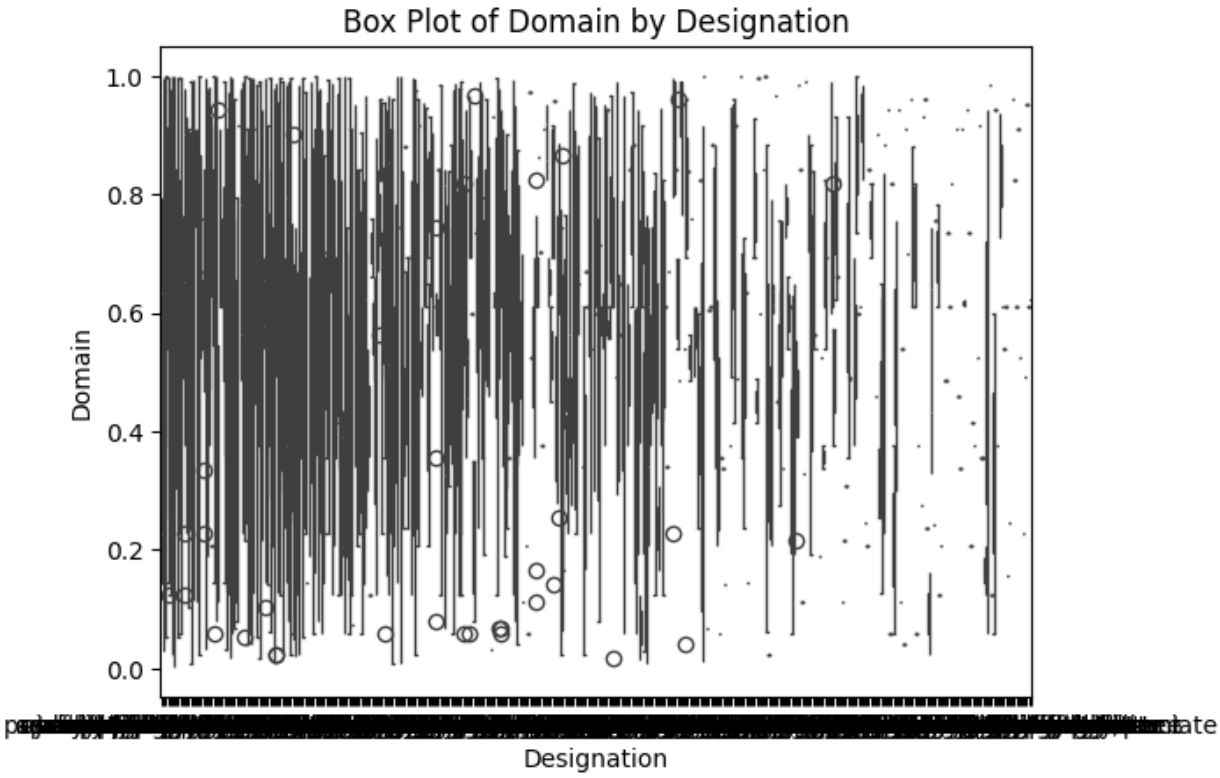

Box Plot of CollegeCityID by Designation
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
Designation

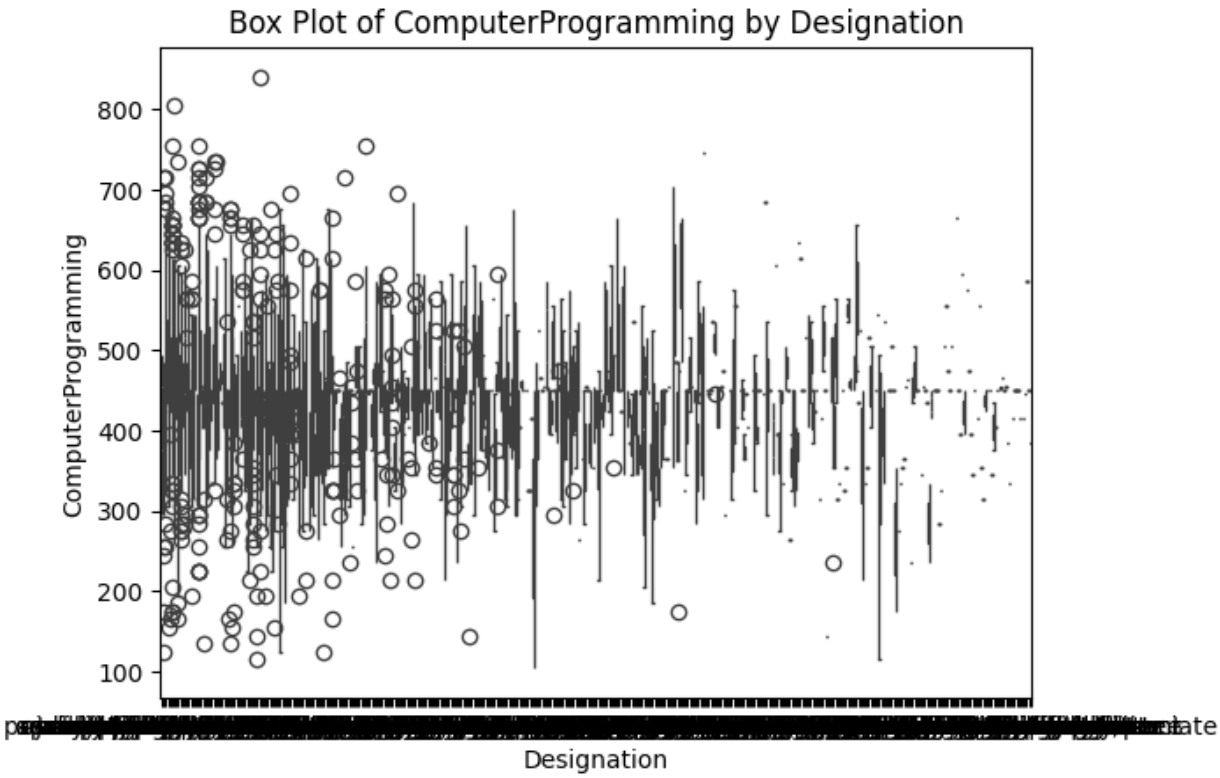

Box Plot of GraduationYear by Designation
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
Designation

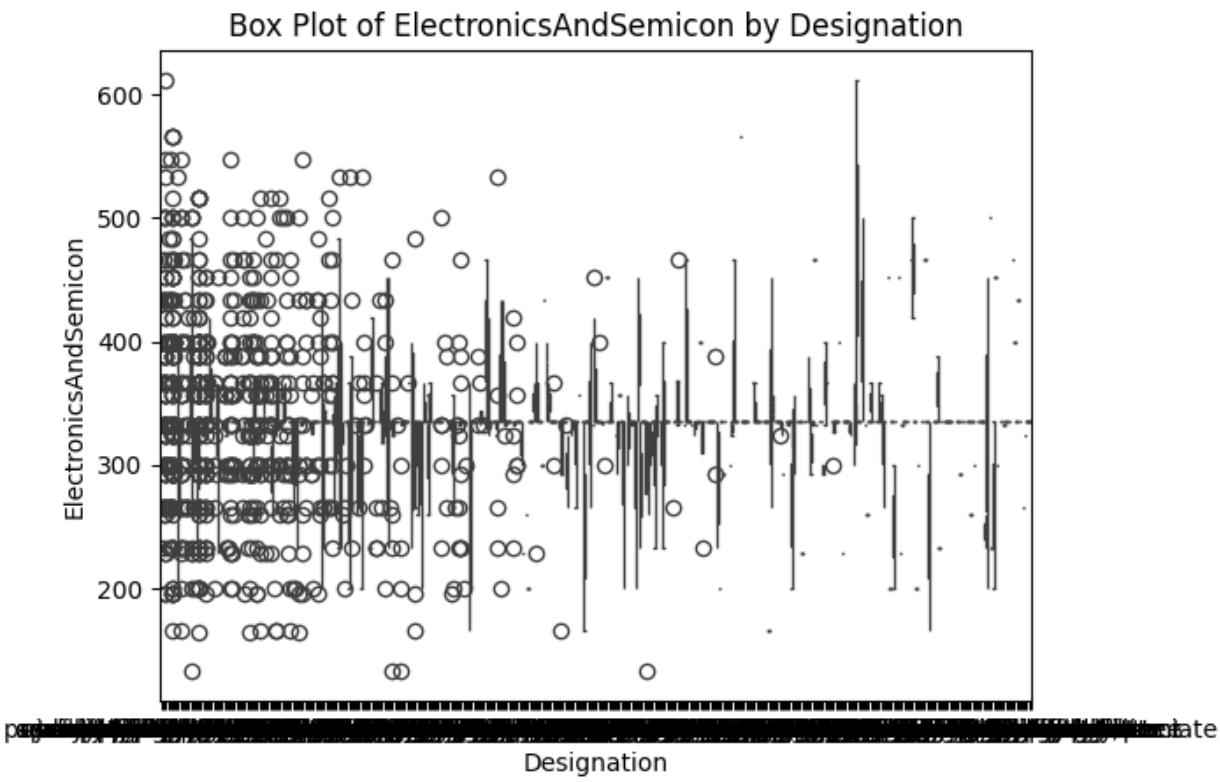

Box Plot of English by Designation
English
900
800
700
600
500
400
300
200
Designation

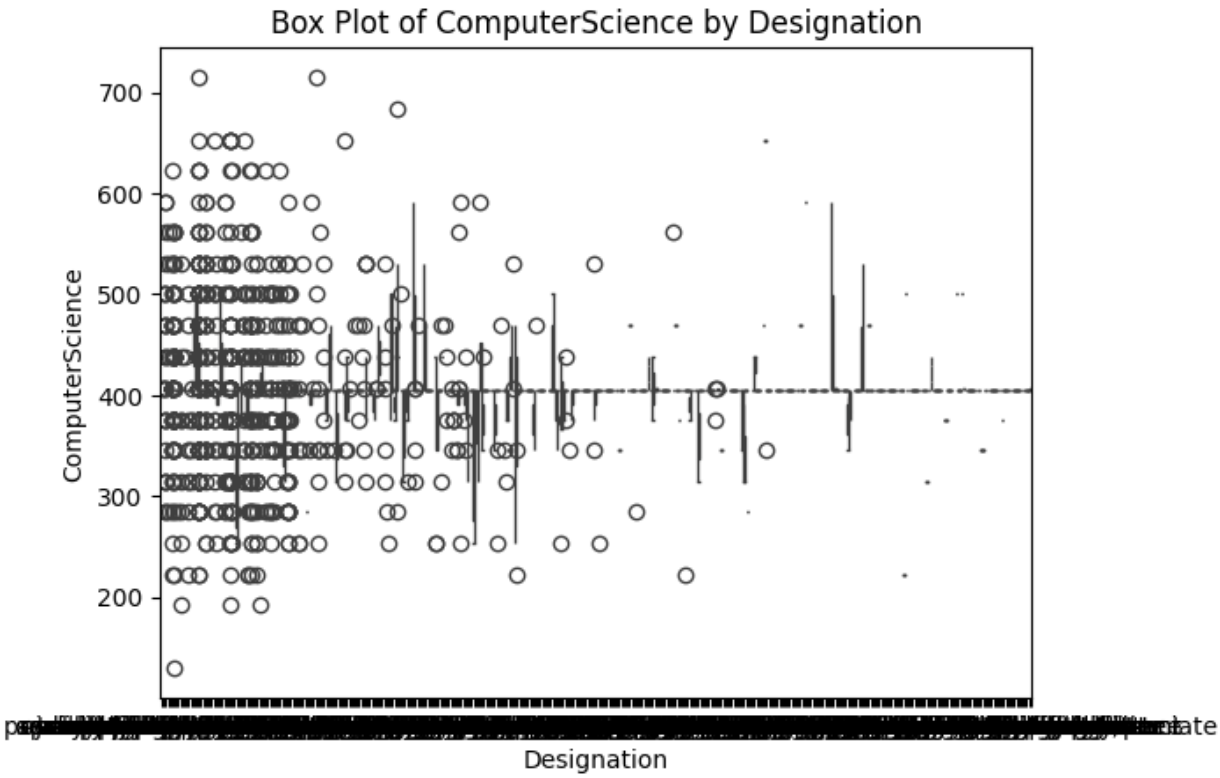

Box Plot of Logical by Designation
Logical
800
700
600
500
400
300
200
Designation

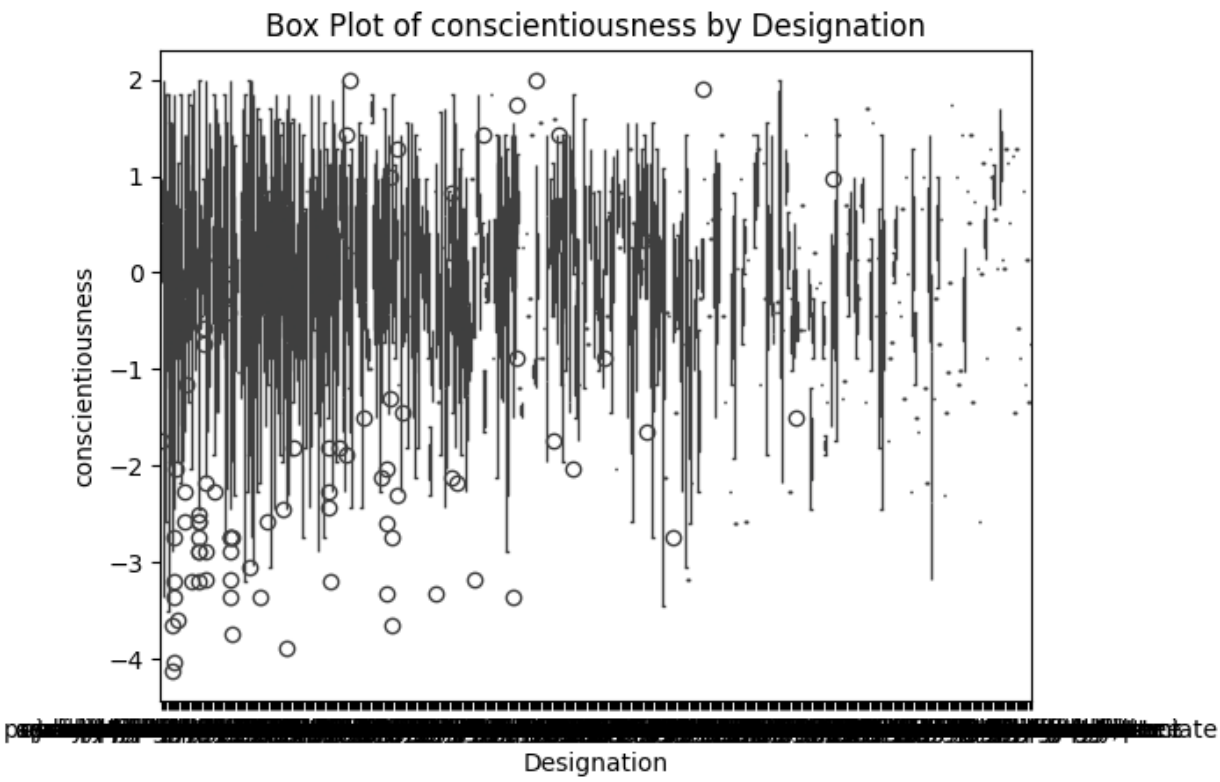

Box Plot of Quant by Designation
Quant
900
800
700
600
500
400
300
200
100
Designation

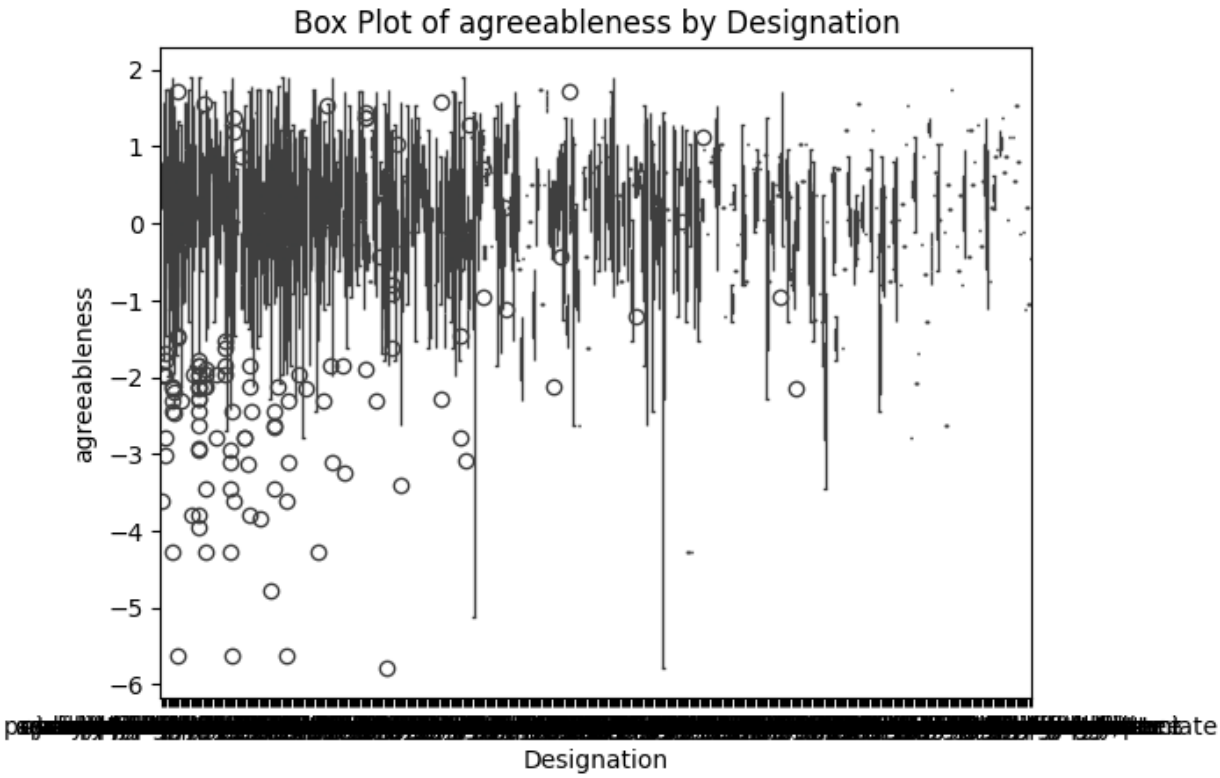

Box Plot of Domain by Designation
Domain
1.0
0.8
0.6
0.4
0.2
0.0
Designation

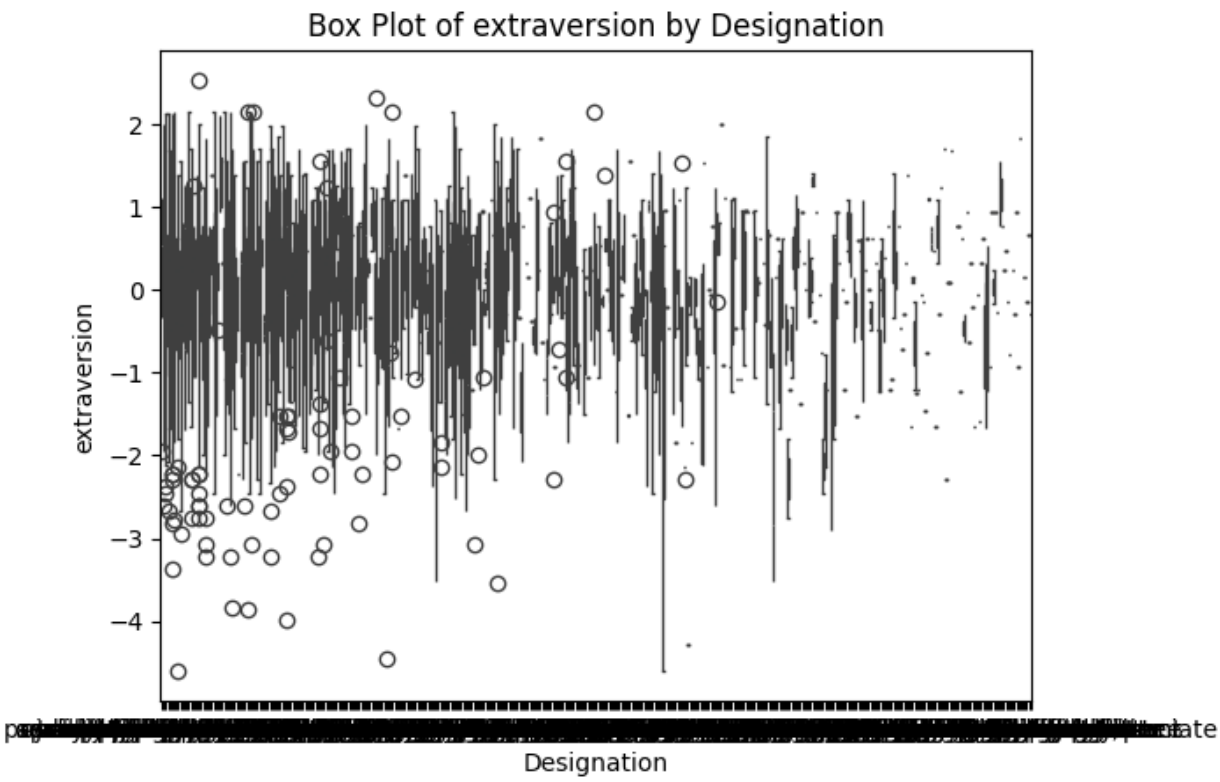

Box Plot of ComputerProgramming by Designation
ComputerProgramming
800
700
600
500
400
300
200
100
Designation

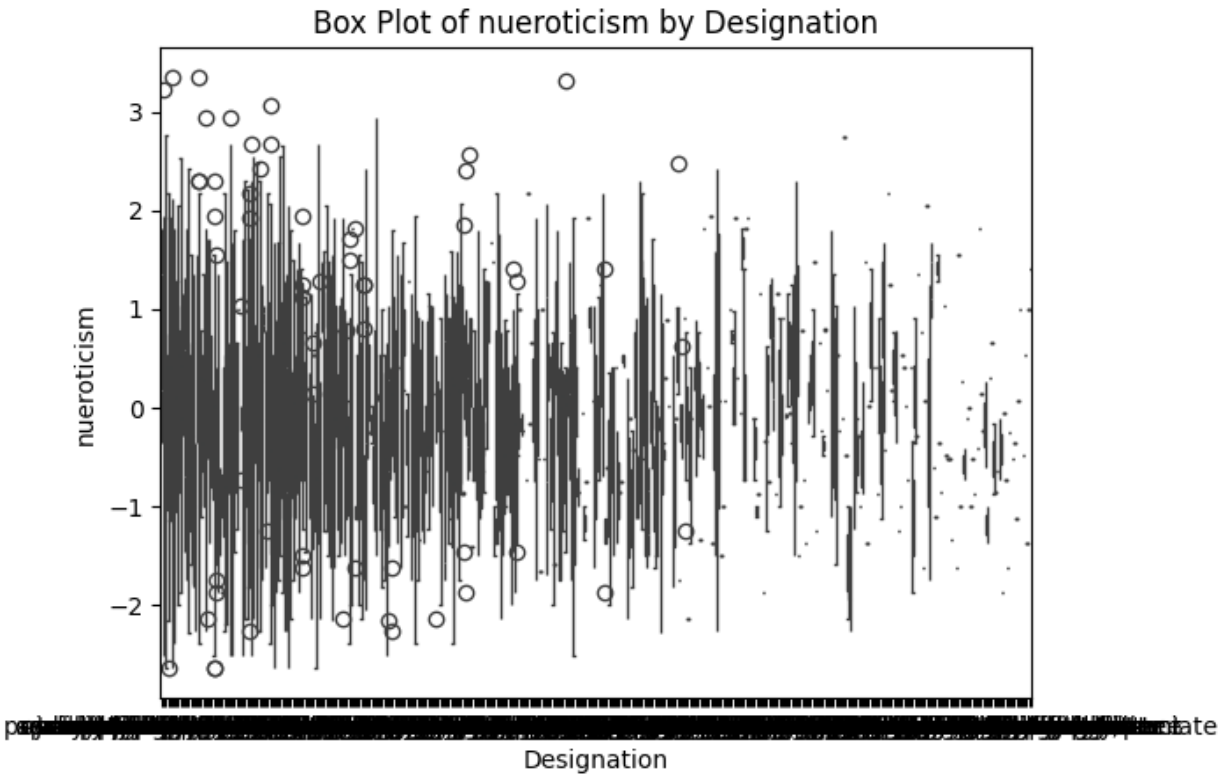

Box Plot of ElectronicsAndSemicon by Designation
ElectronicsAndSemicon
600
500
400
300
200
Designation

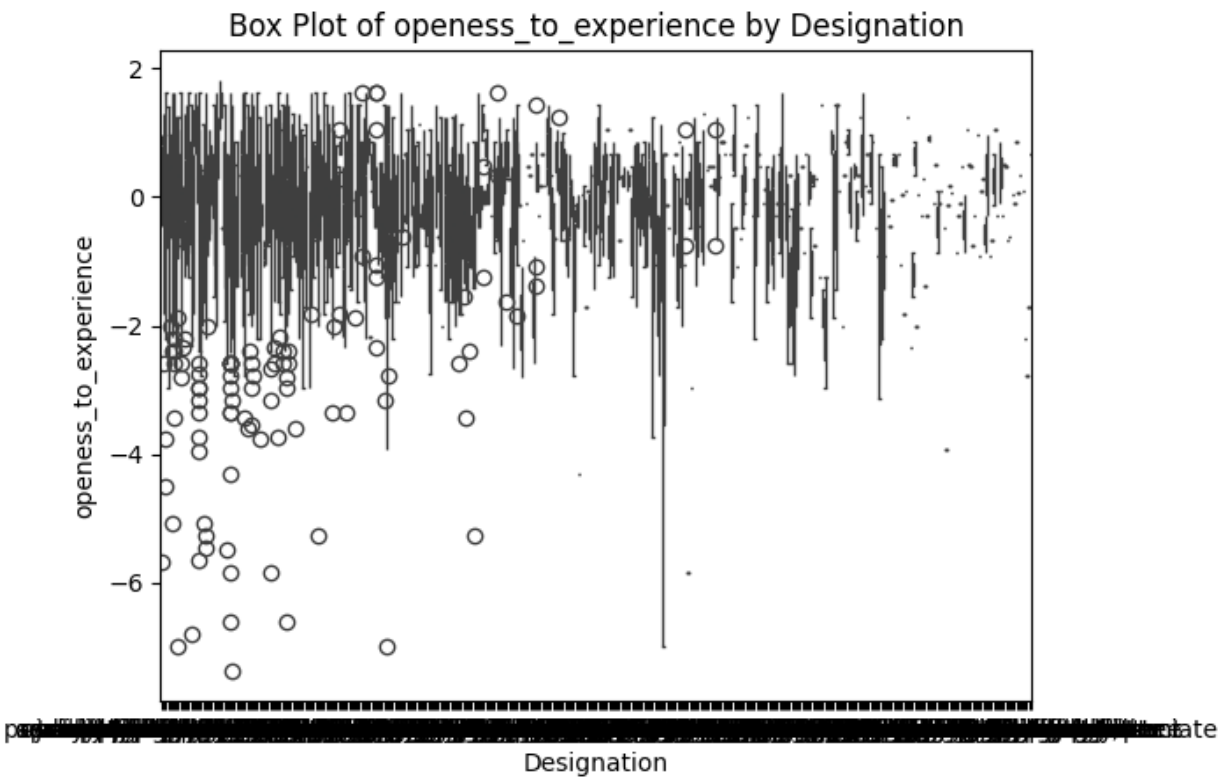

Box Plot of ComputerScience by Designation
ComputerScience
700
600
500
400
300
200
Designation

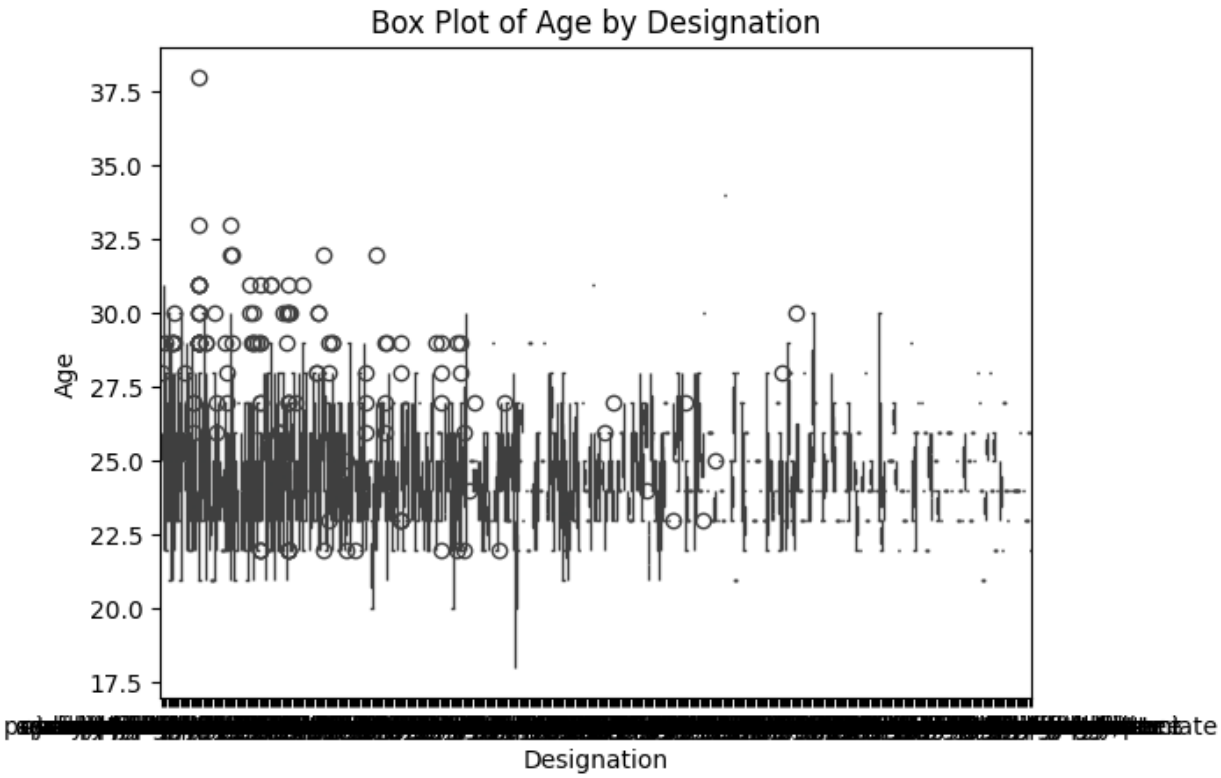

Box Plot of conscientiousness by Designation
conscientiousness
2
1
0
−1
−2
−3
−4
Designation

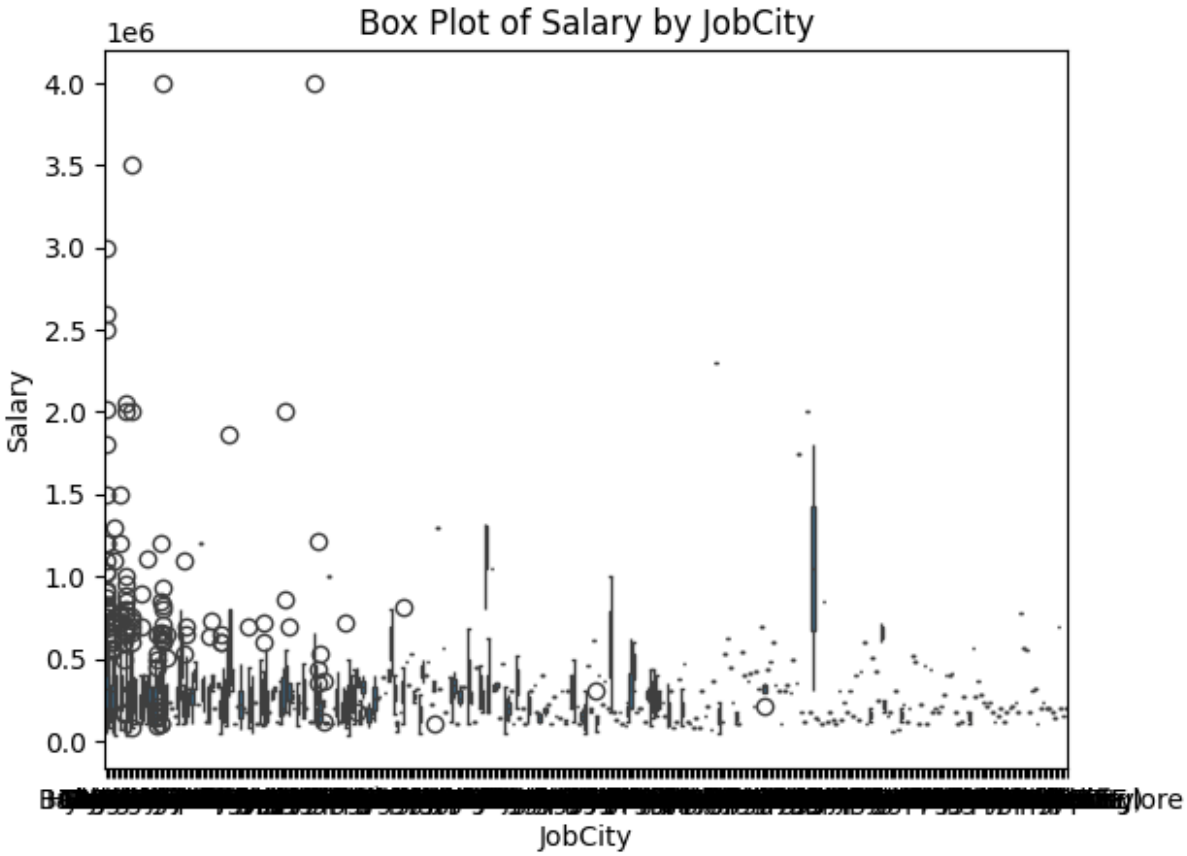

Box Plot of agreeableness by Designation
agreeableness
2
1
0
−1
−2
−3
−4
−5
−6
Designation

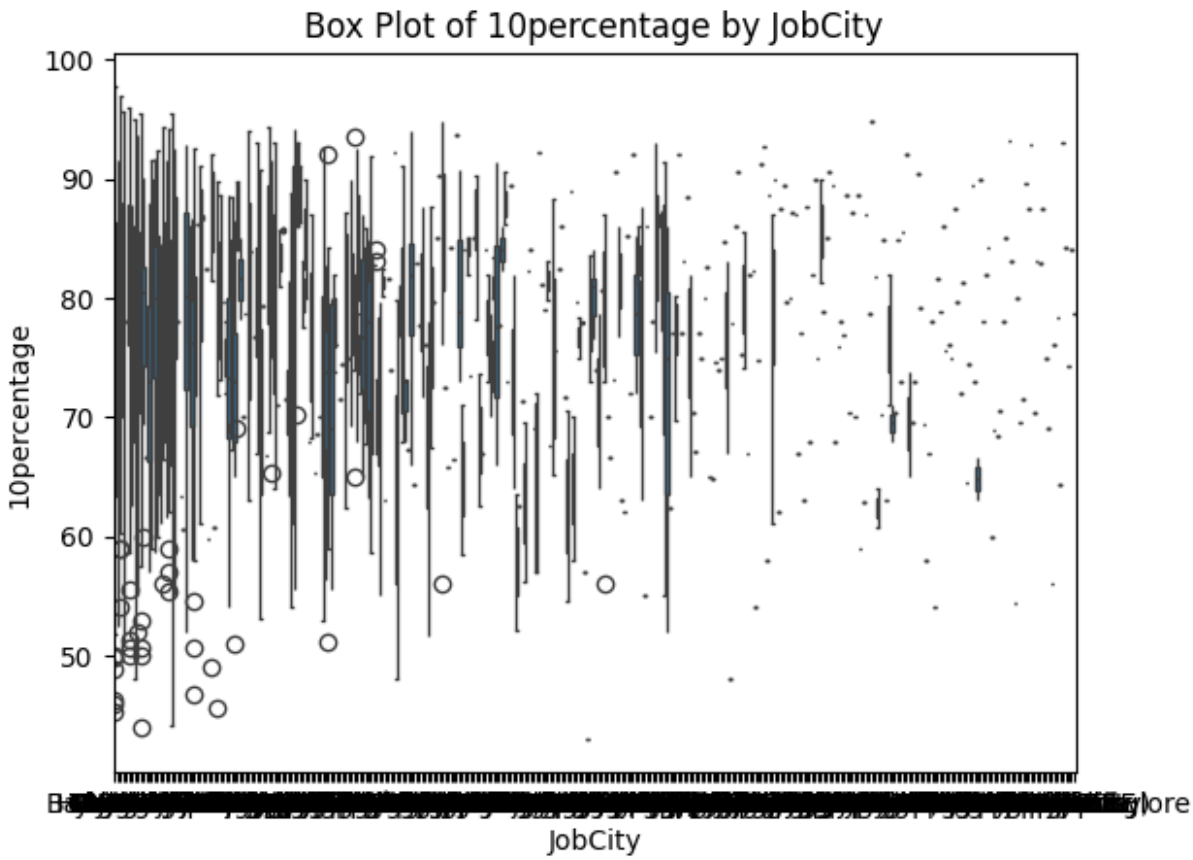

Box Plot of extraversion by Designation
extraversion
2
1
0
−1
−2
−3
−4
Designation

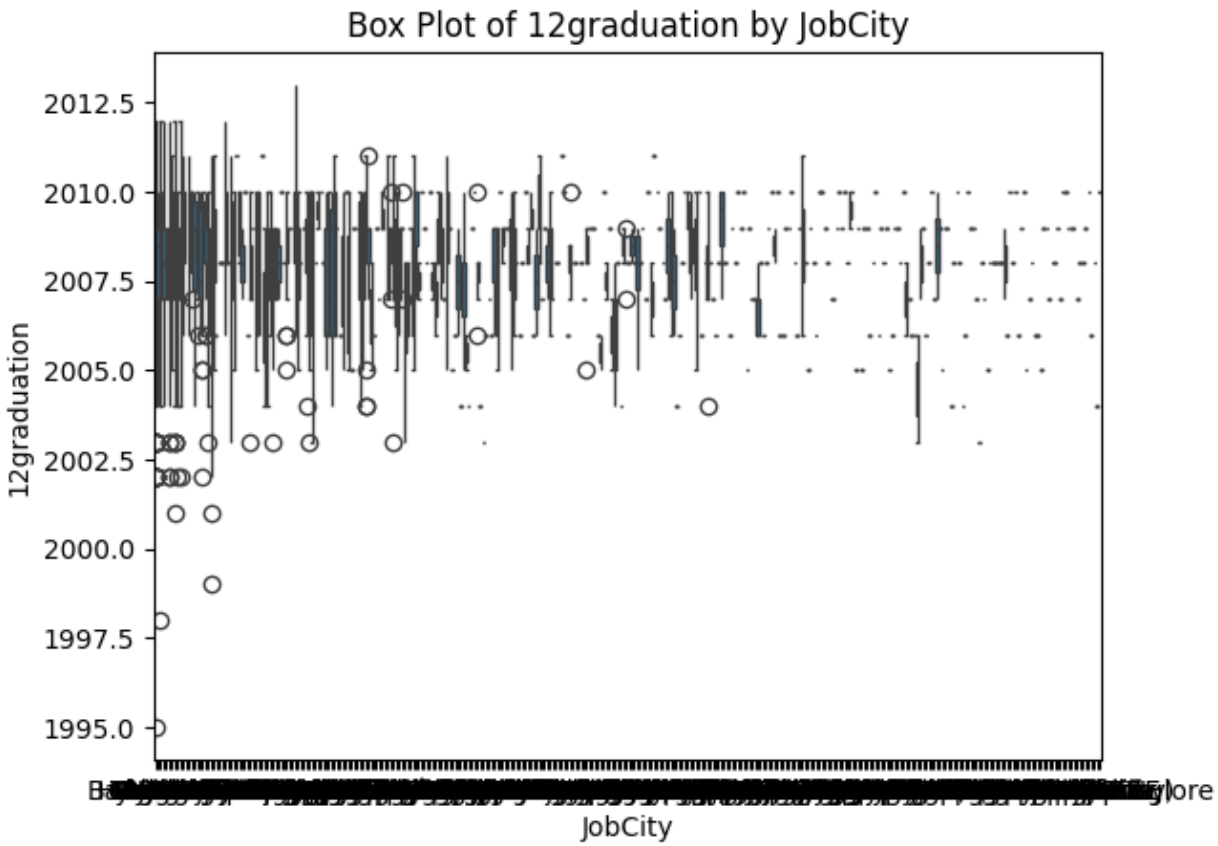

Box Plot of nueroticism by Designation
nueroticism
3
2
1
0
−1
−2
Designation

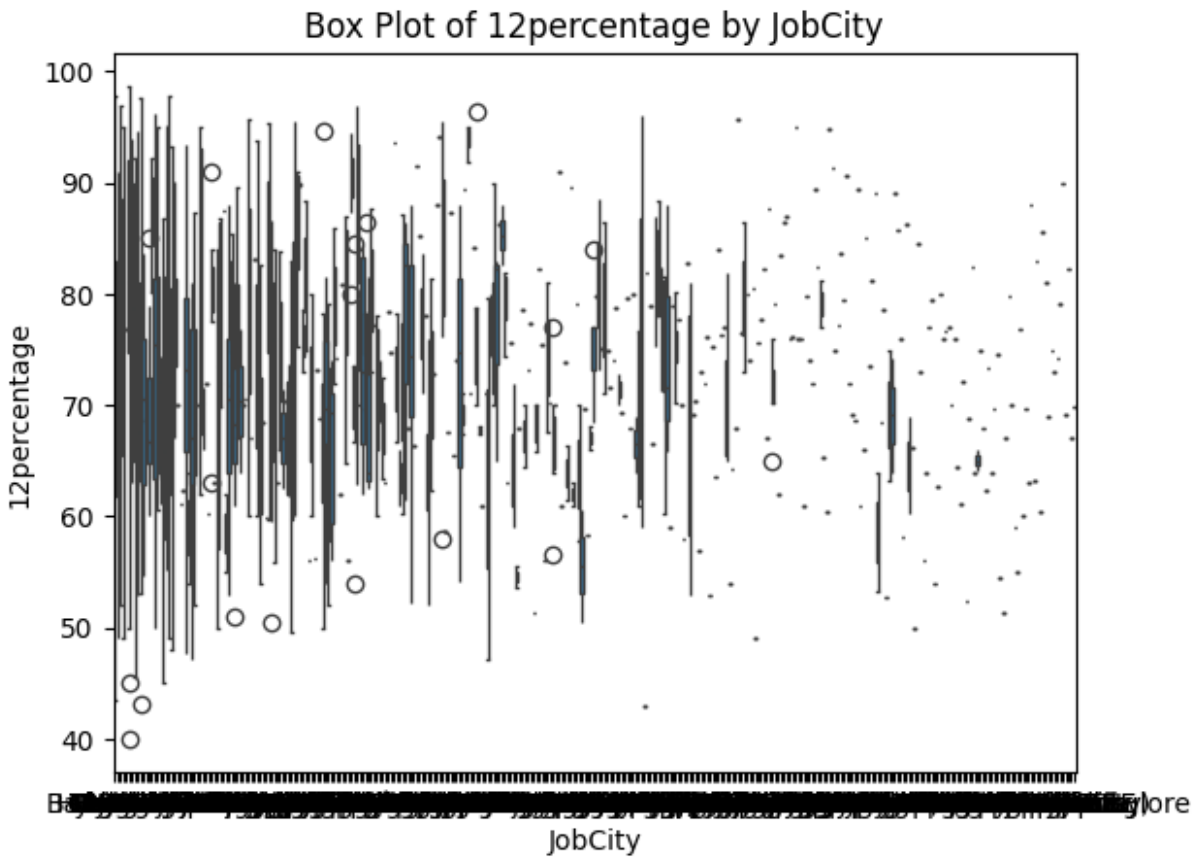

Box Plot of openess_to_experience by Designation
openess_to_experience
2
0
−2
−4
−6
Designation

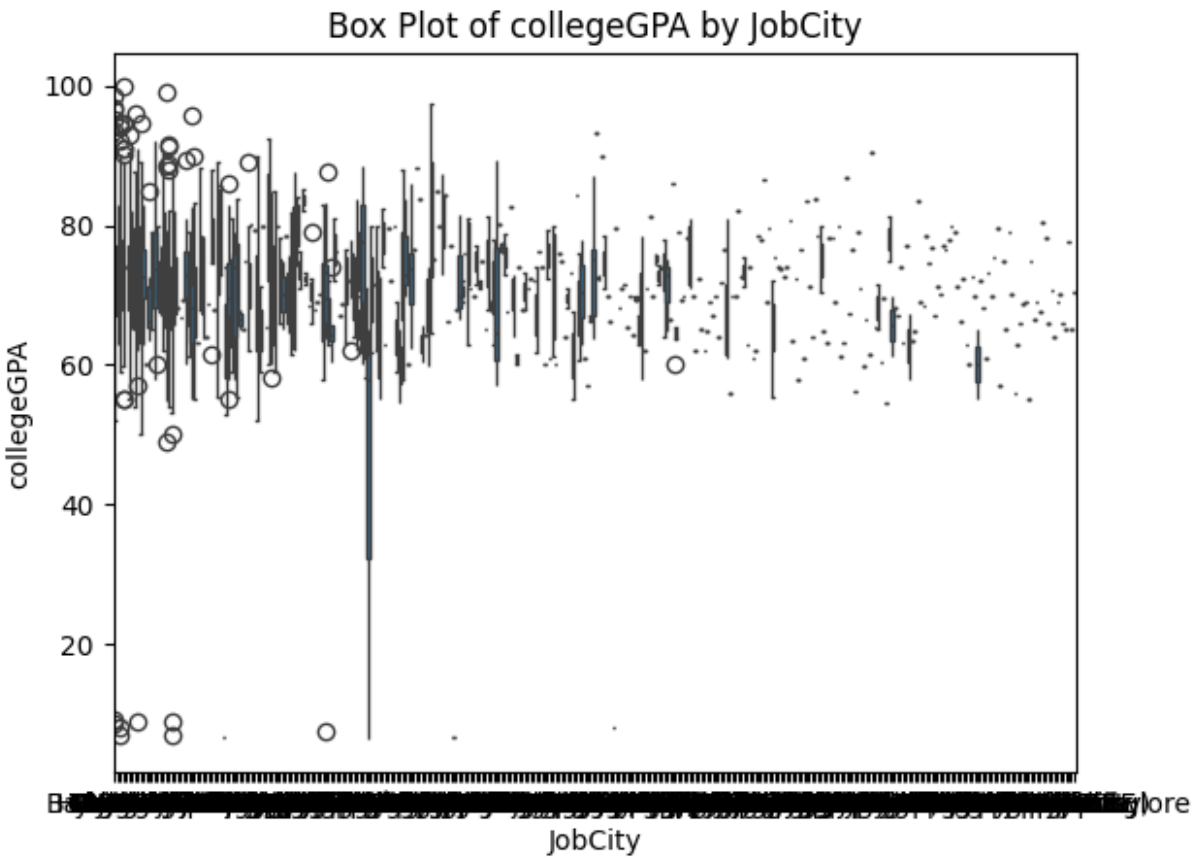

Box Plot of Age by Designation
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
Designation

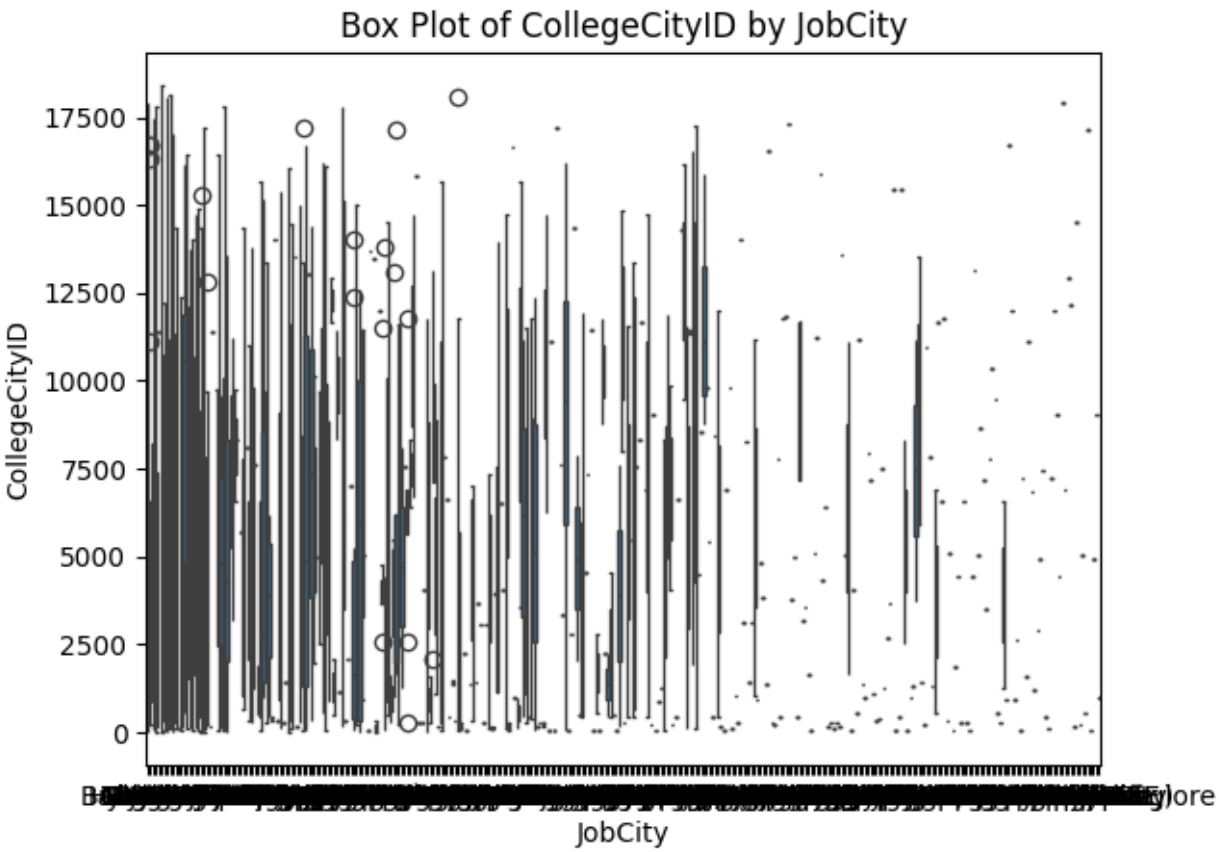

Box Plot of Salary by JobCity
1e6
Salary
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
JobCity

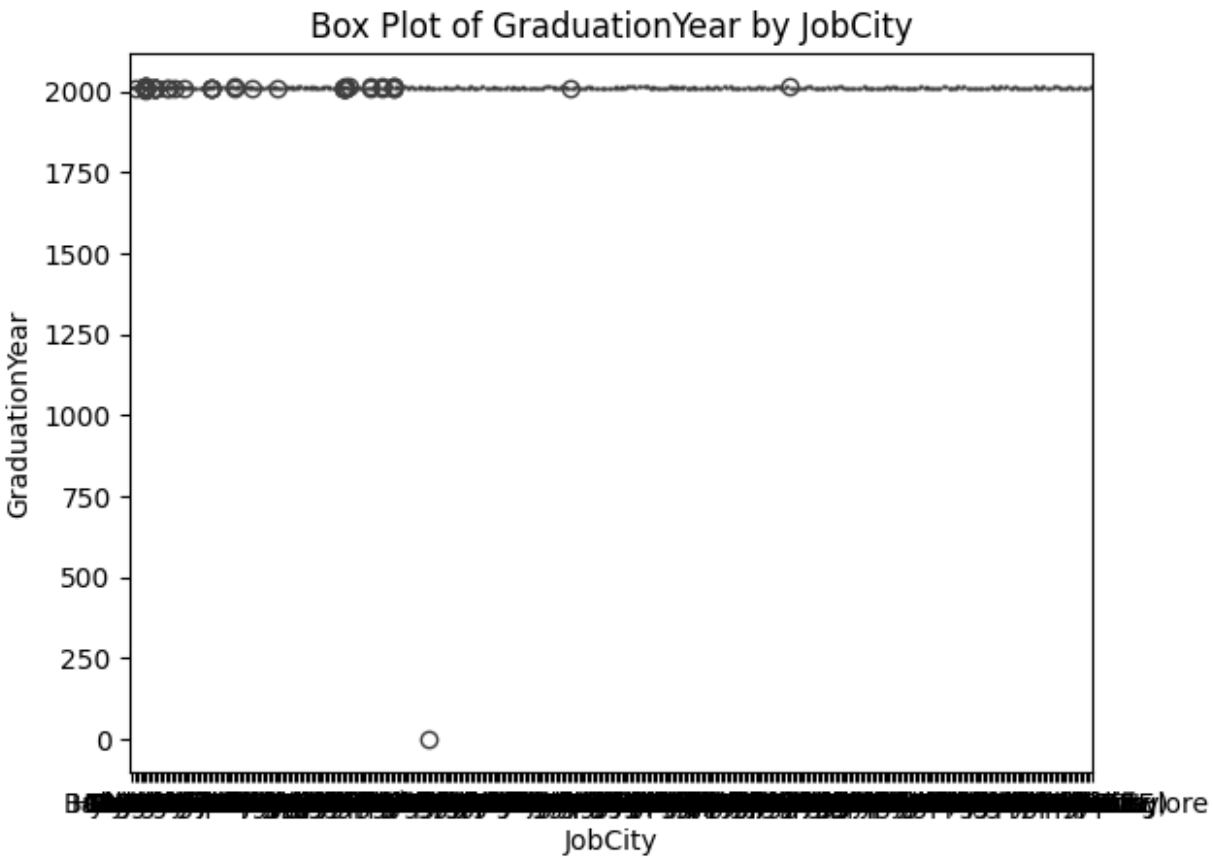

Box Plot of 10percentage by JobCity
100
90
80
70
60
50
10percentage
JobCity

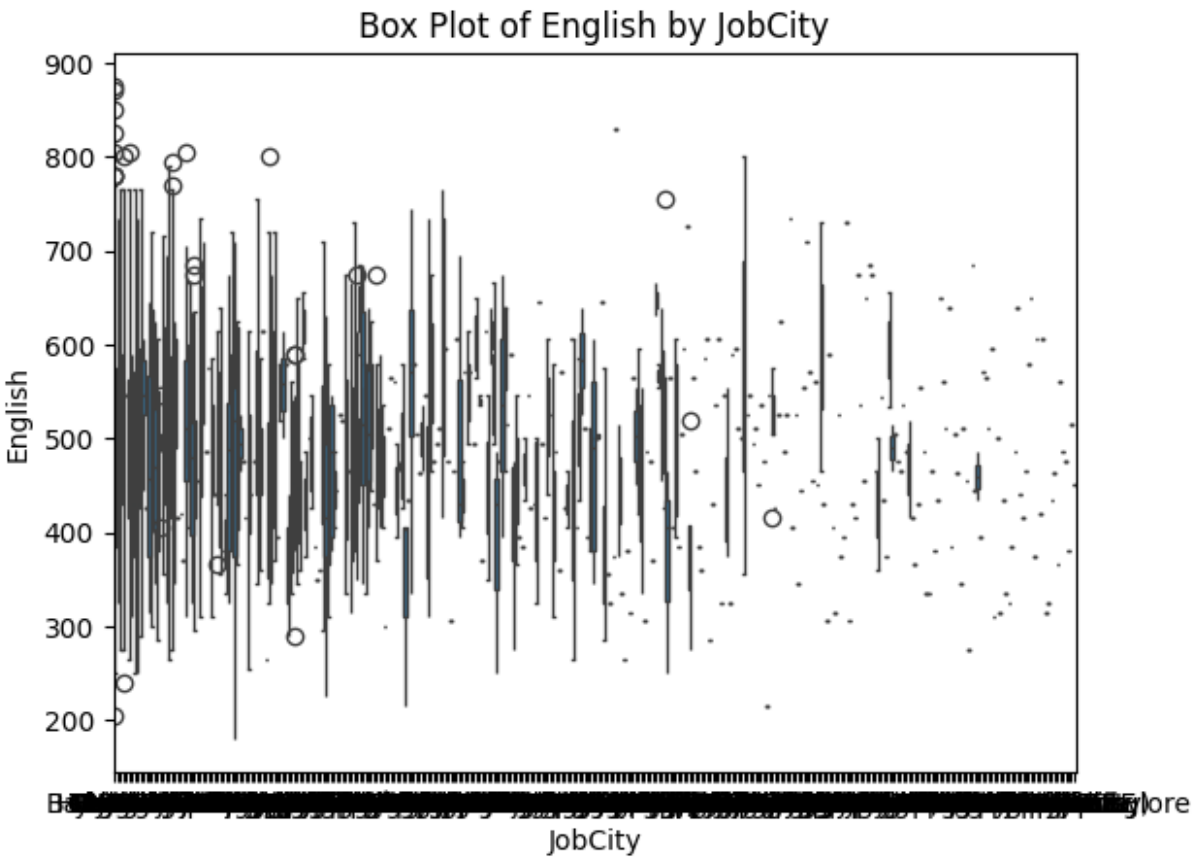

Box Plot of 12graduation by JobCity
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
12graduation
JobCity

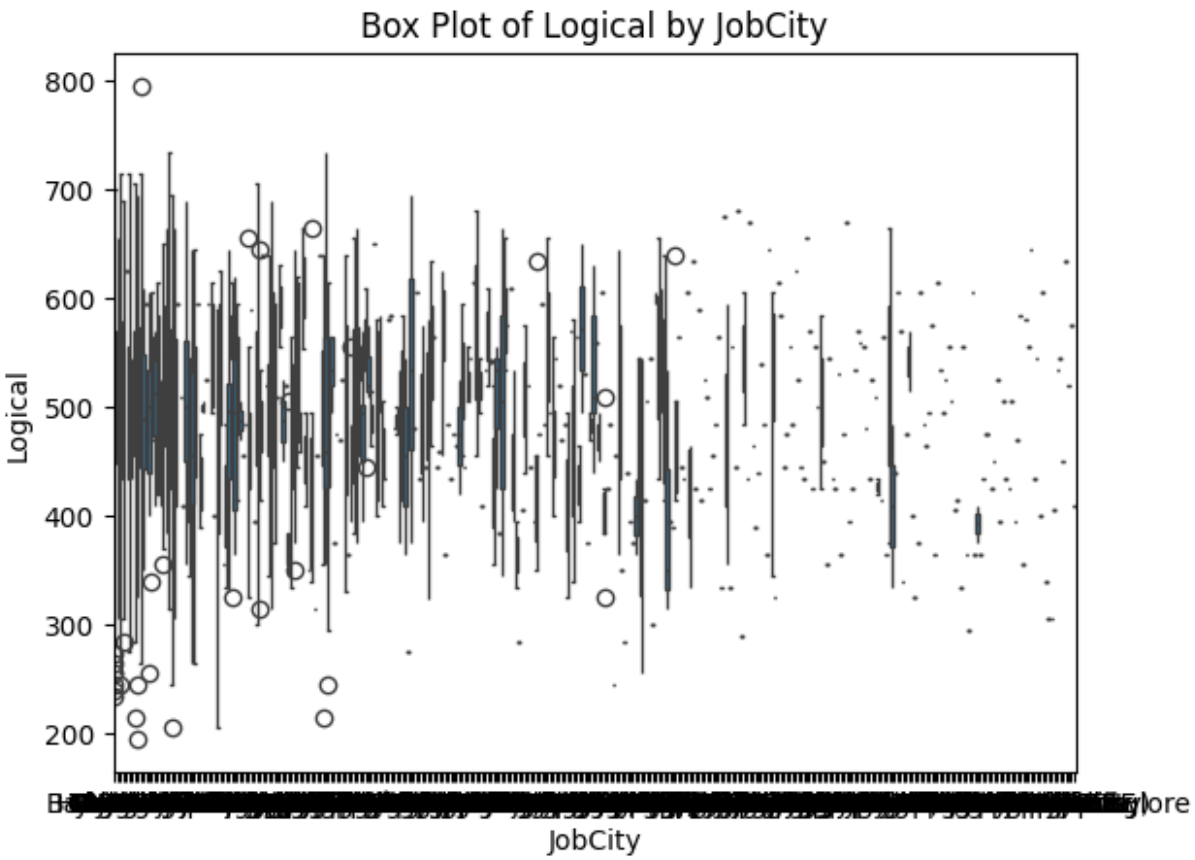

Box Plot of 12percentage by JobCity
12percentage
100
90
80
70
60
50
40
JobCity

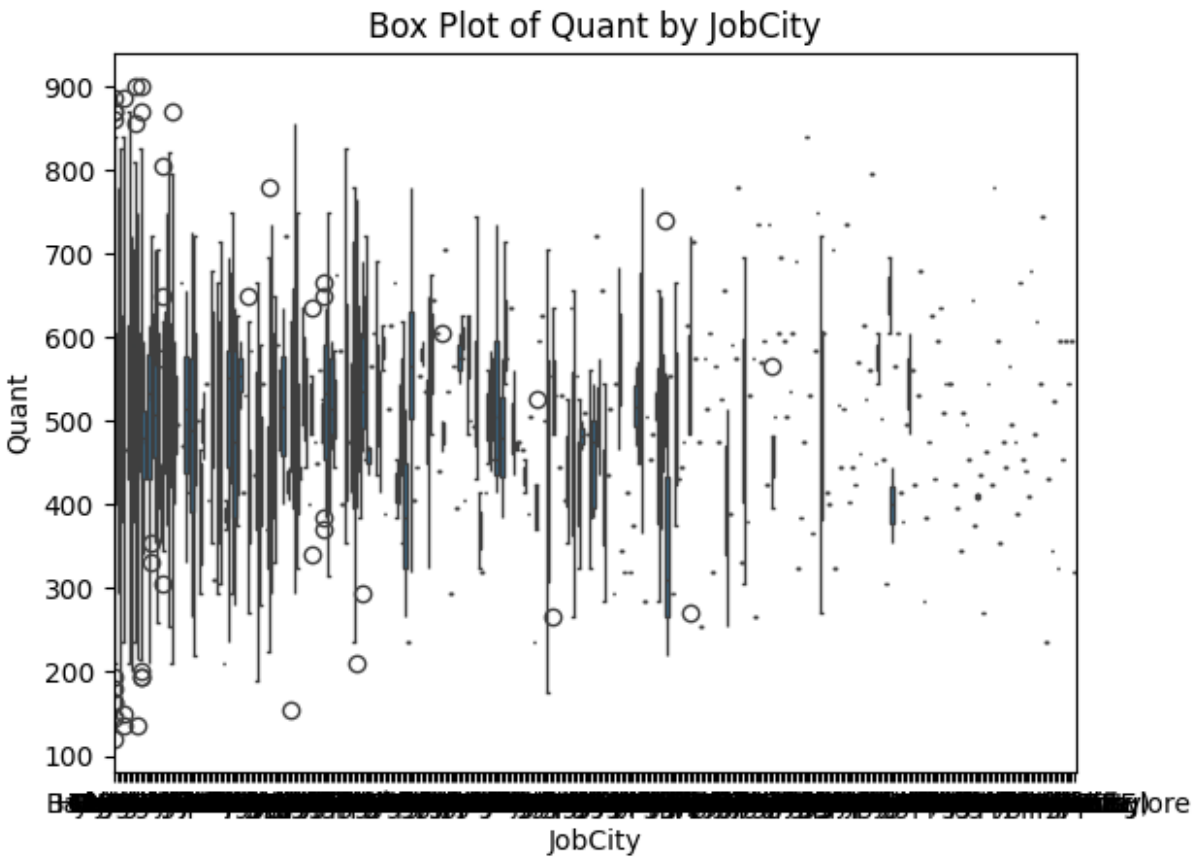

Box Plot of collegeGPA by JobCity
collegeGPA
100
80
60
40
20
JobCity

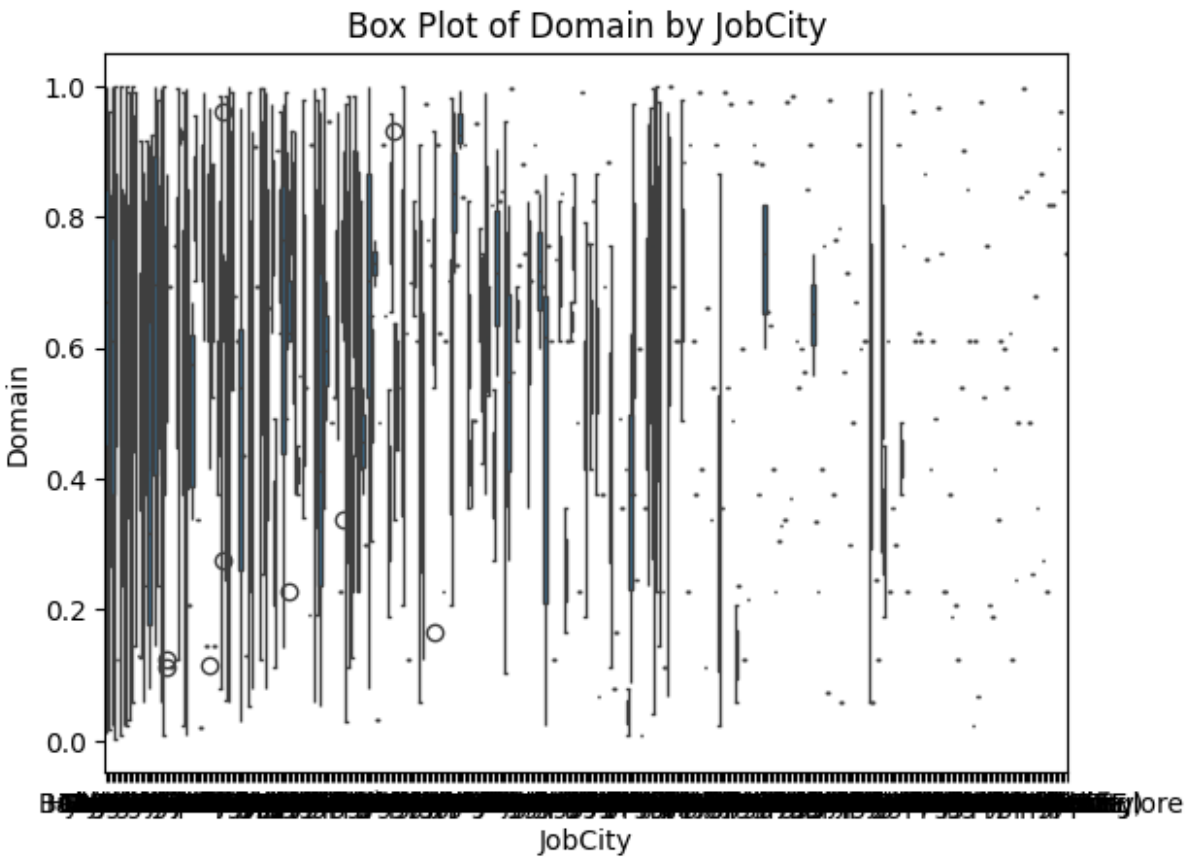

Box Plot of CollegeCityID by JobCity
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
JobCity

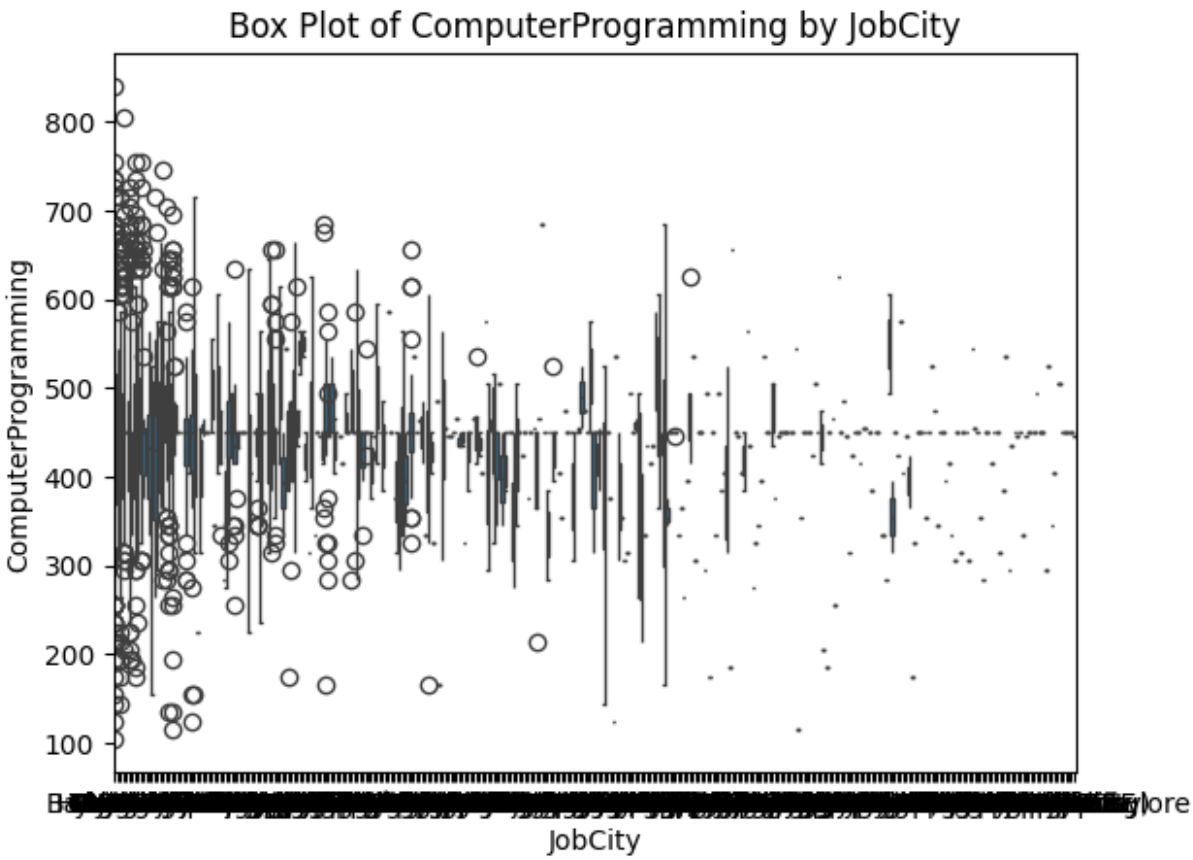

Box Plot of GraduationYear by JobCity
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
JobCity

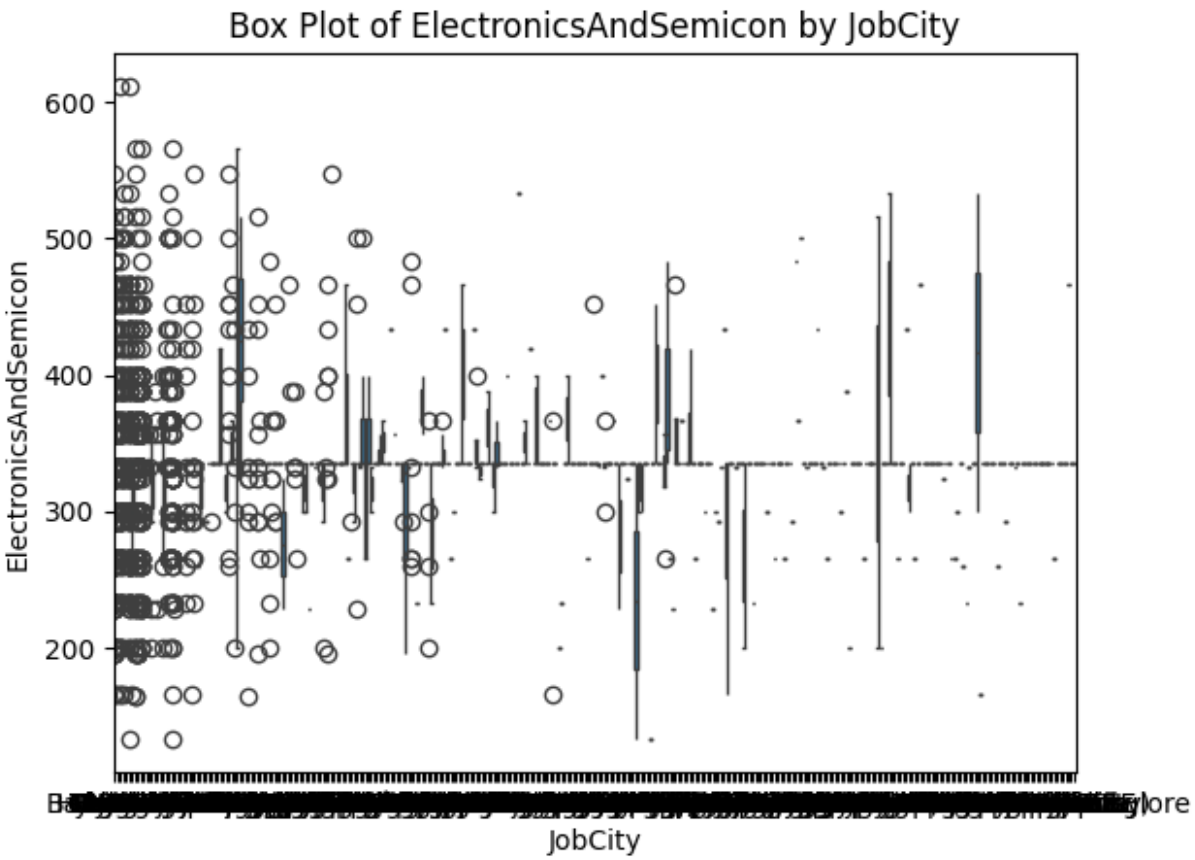

Box Plot of English by JobCity
English
900
800
700
600
500
400
300
200
JobCity

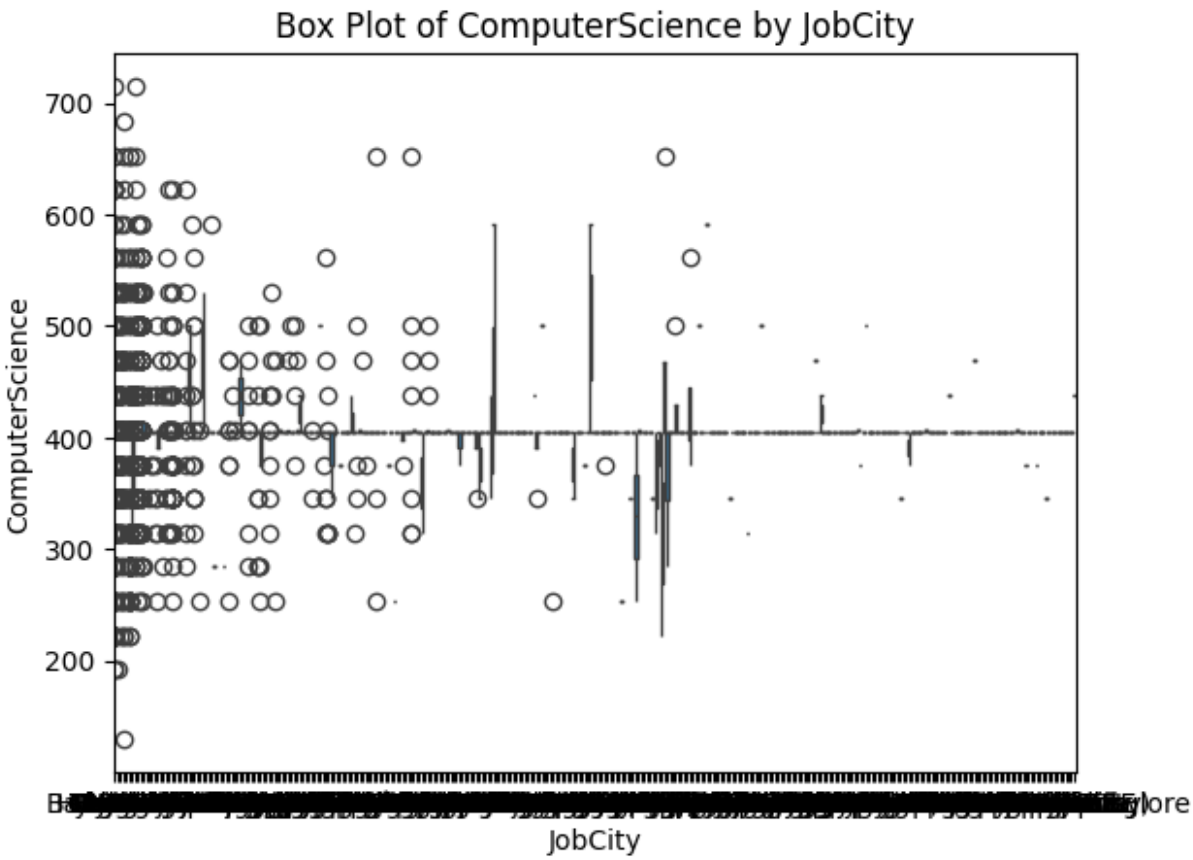

Box Plot of Logical by JobCity
Logical
800
700
600
500
400
300
200
JobCity

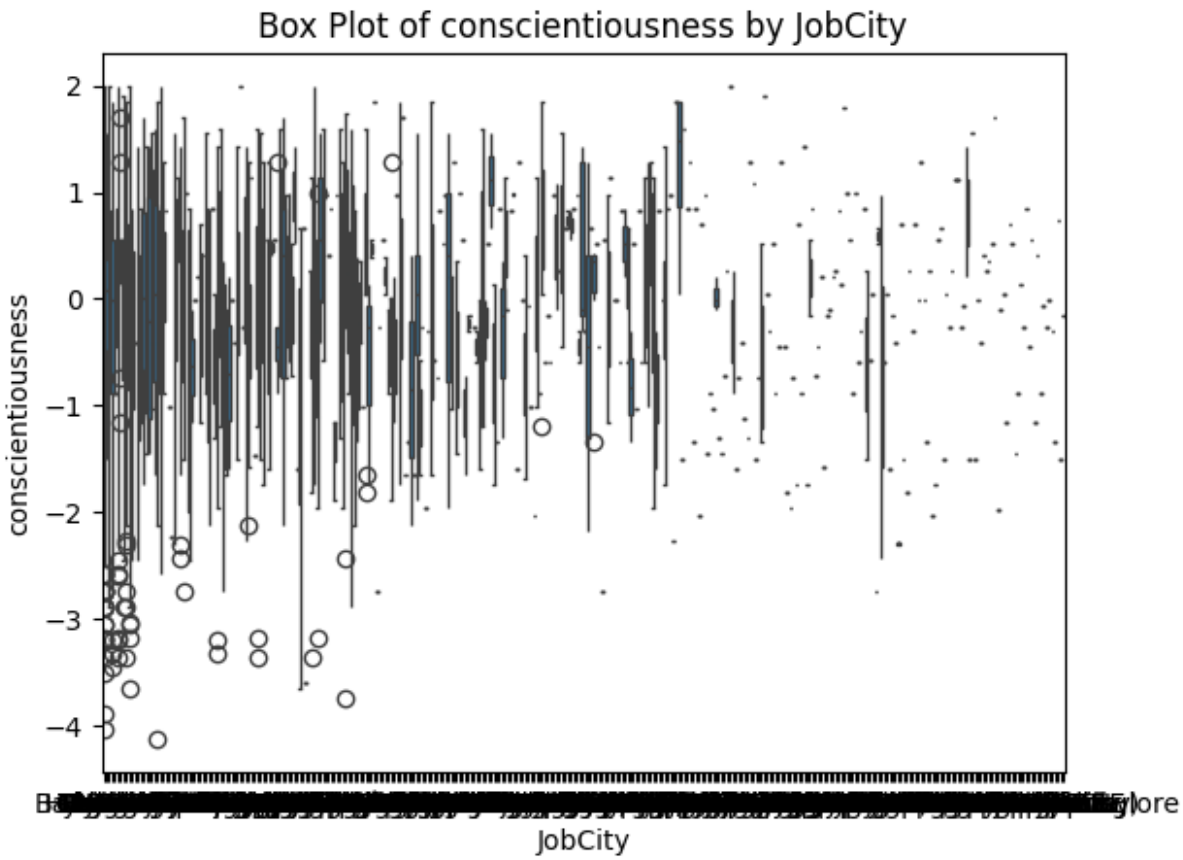

Box Plot of Quant by JobCity
Quant
900
800
700
600
500
400
300
200
100
JobCity

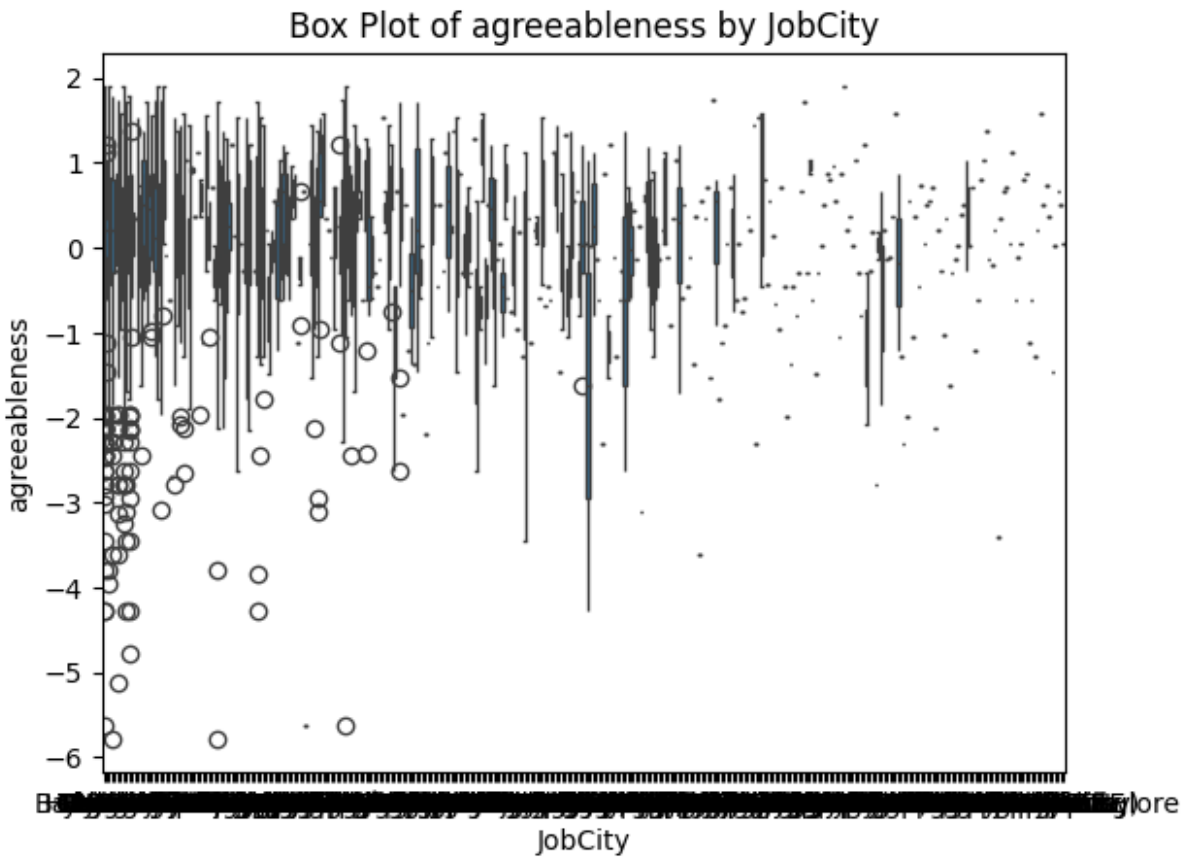

Box Plot of Domain by JobCity
Domain
1.0
0.8
0.6
0.4
0.2
0.0
JobCity

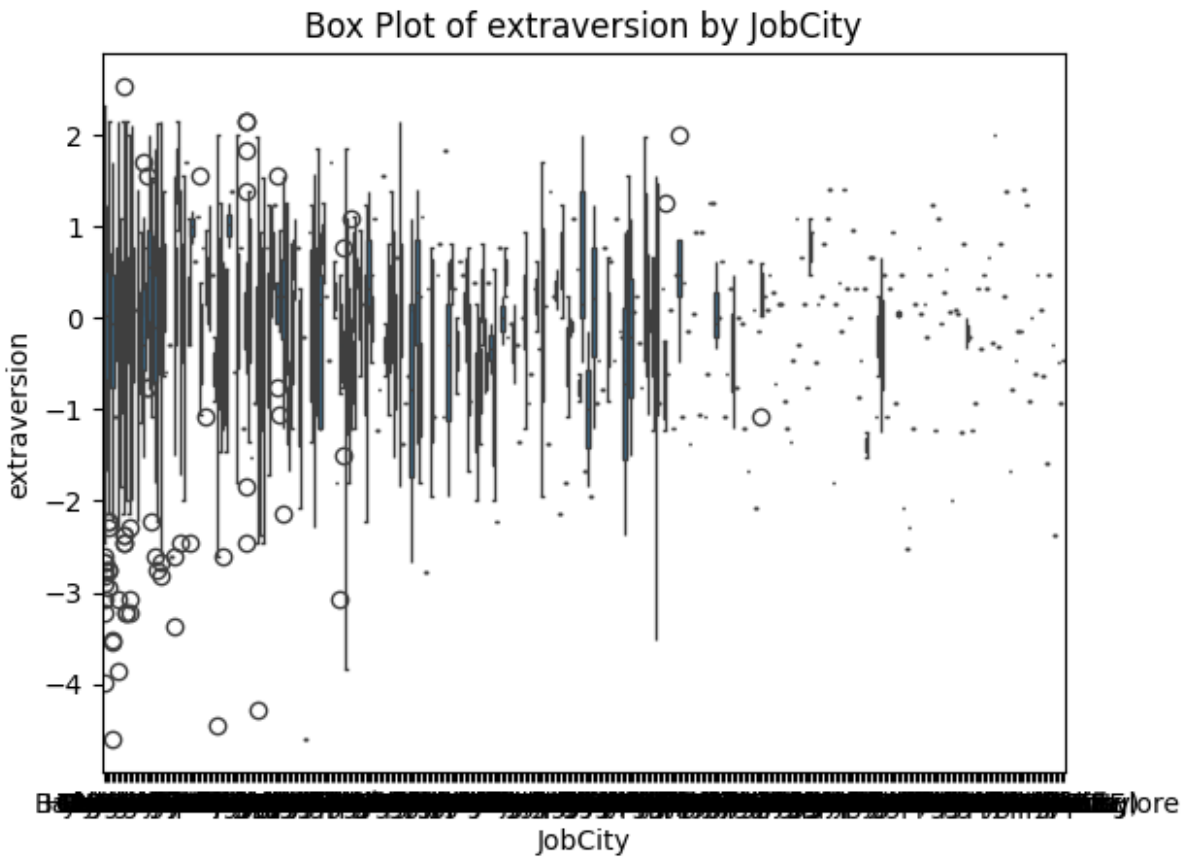

Box Plot of ComputerProgramming by JobCity
ComputerProgramming
800
700
600
500
400
300
200
100
JobCity

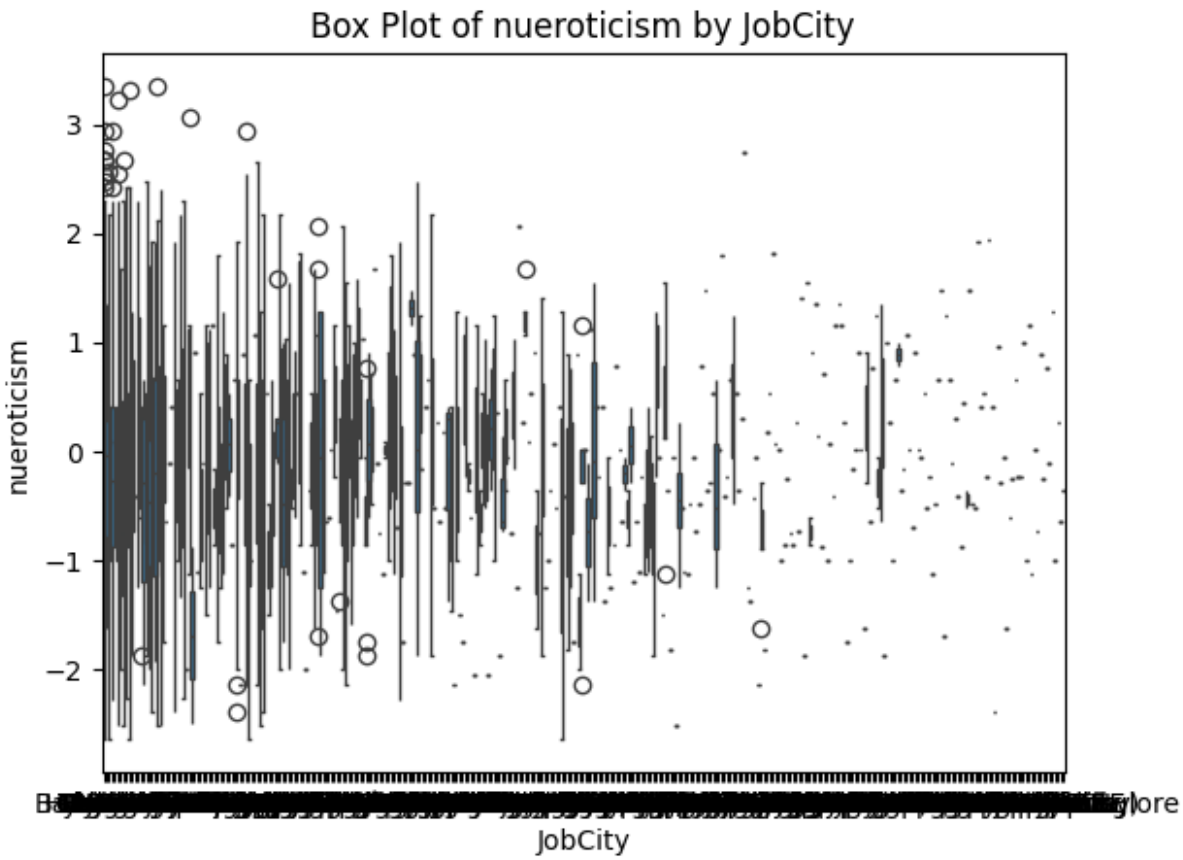

Box Plot of ElectronicsAndSemicon by JobCity
ElectronicsAndSemicon
600
500
400
300
200
JobCity

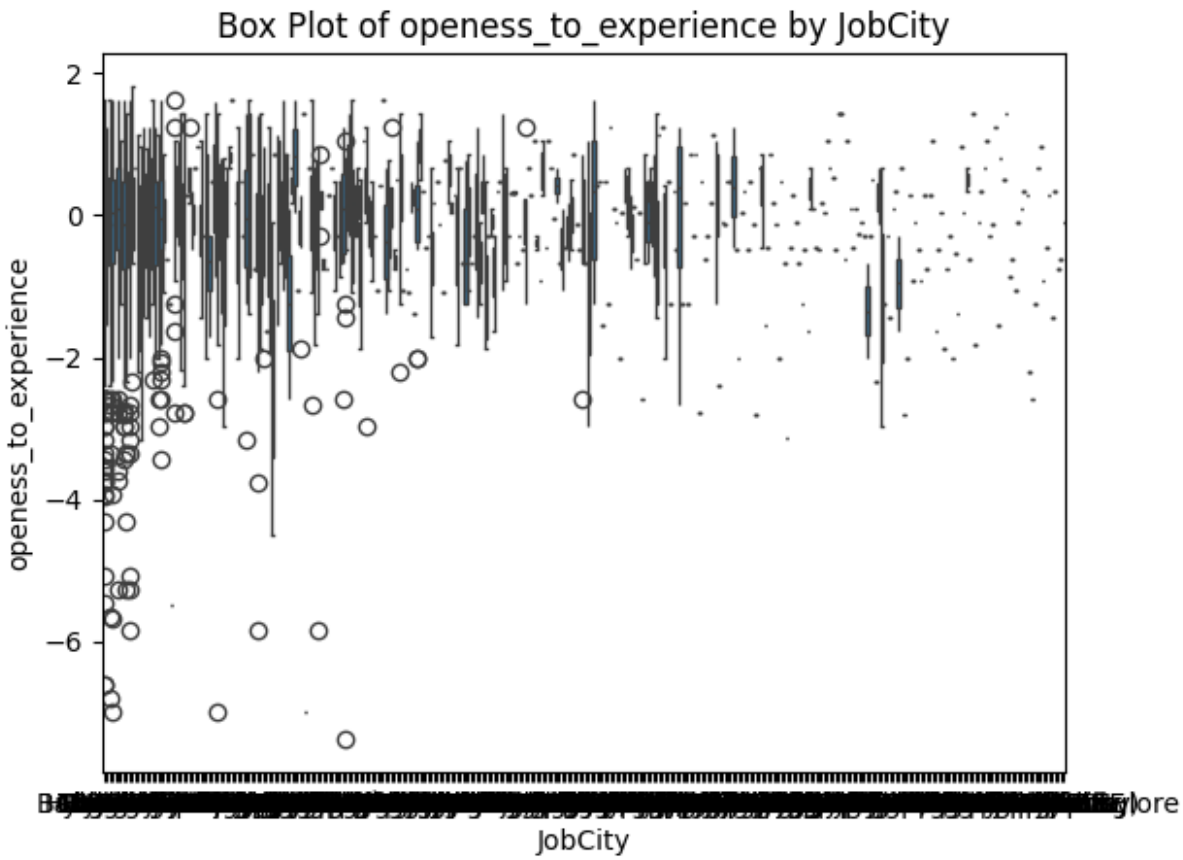

Box Plot of ComputerScience by JobCity
ComputerScience
700
600
500
400
300
200
JobCity

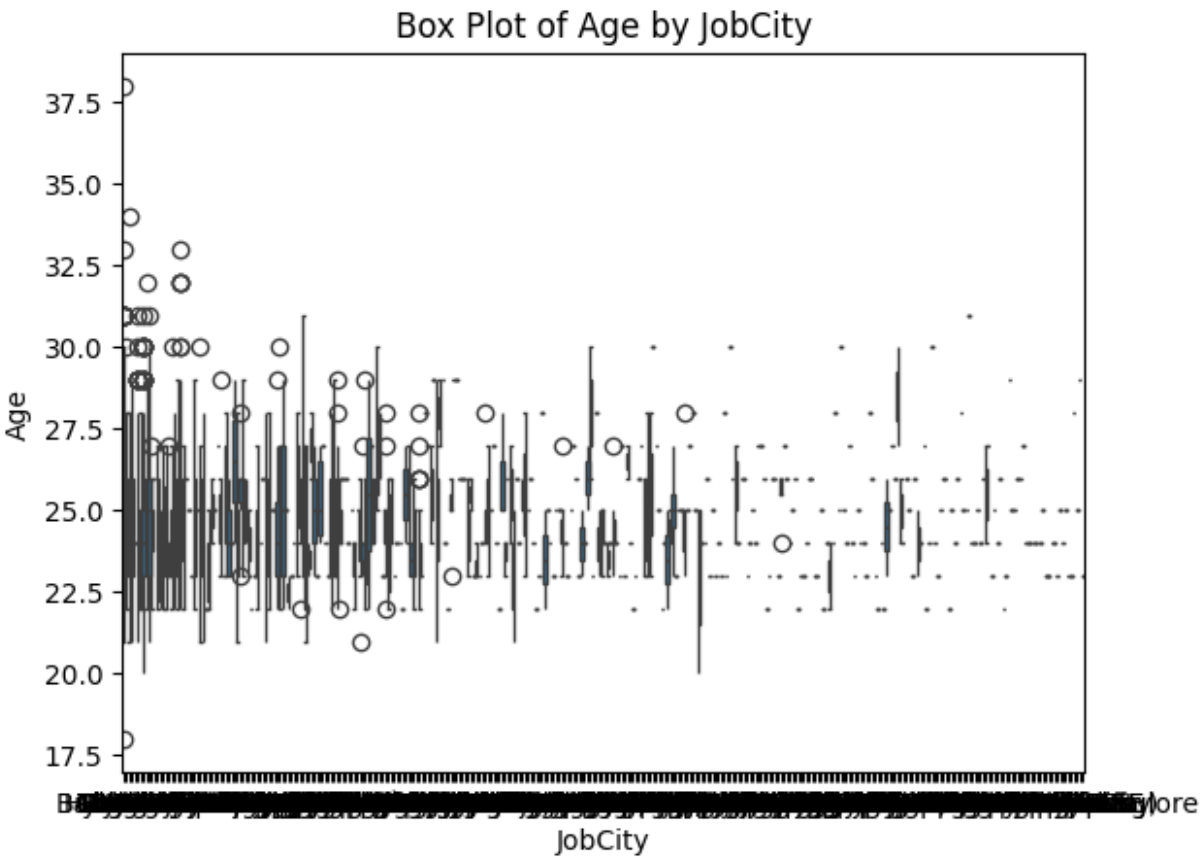

Box Plot of conscientiousness by JobCity
conscientiousness
2
1
0
−1
−2
−3
−4
JobCity

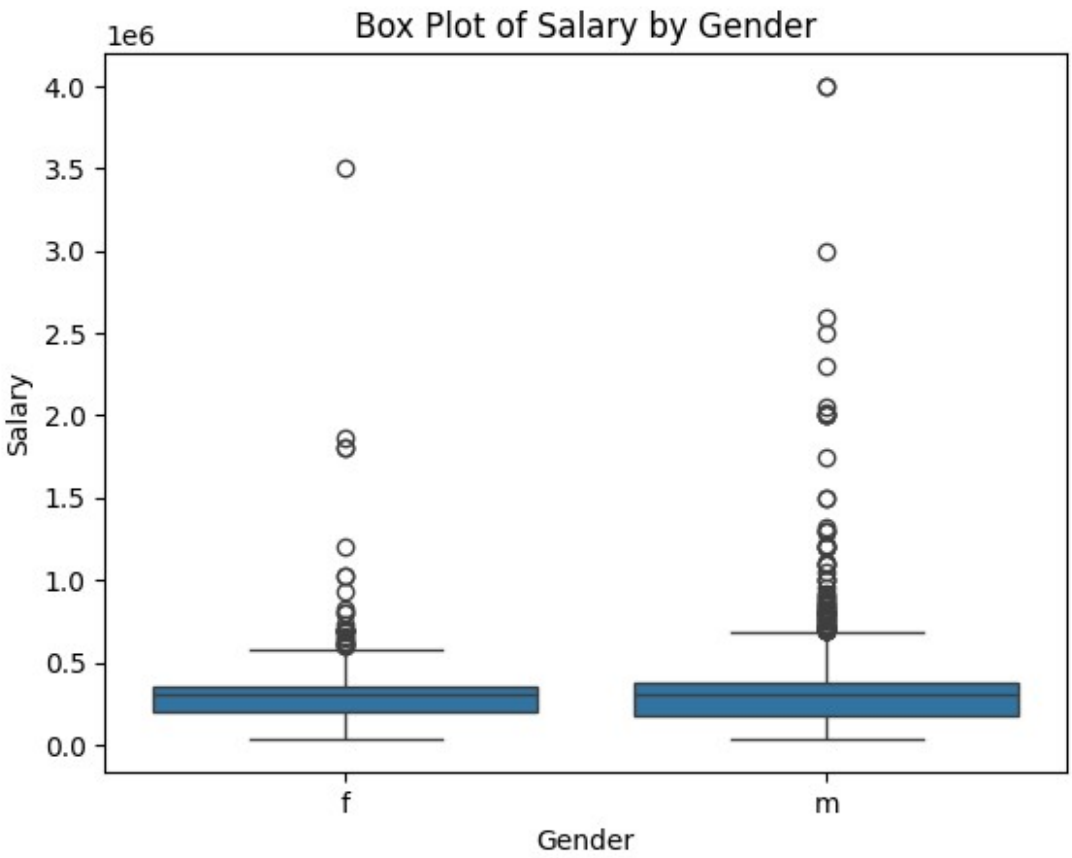

Box Plot of agreeableness by JobCity
agreeableness
JobCity
2
1
0
−1
−2
−3
−4
−5
−6

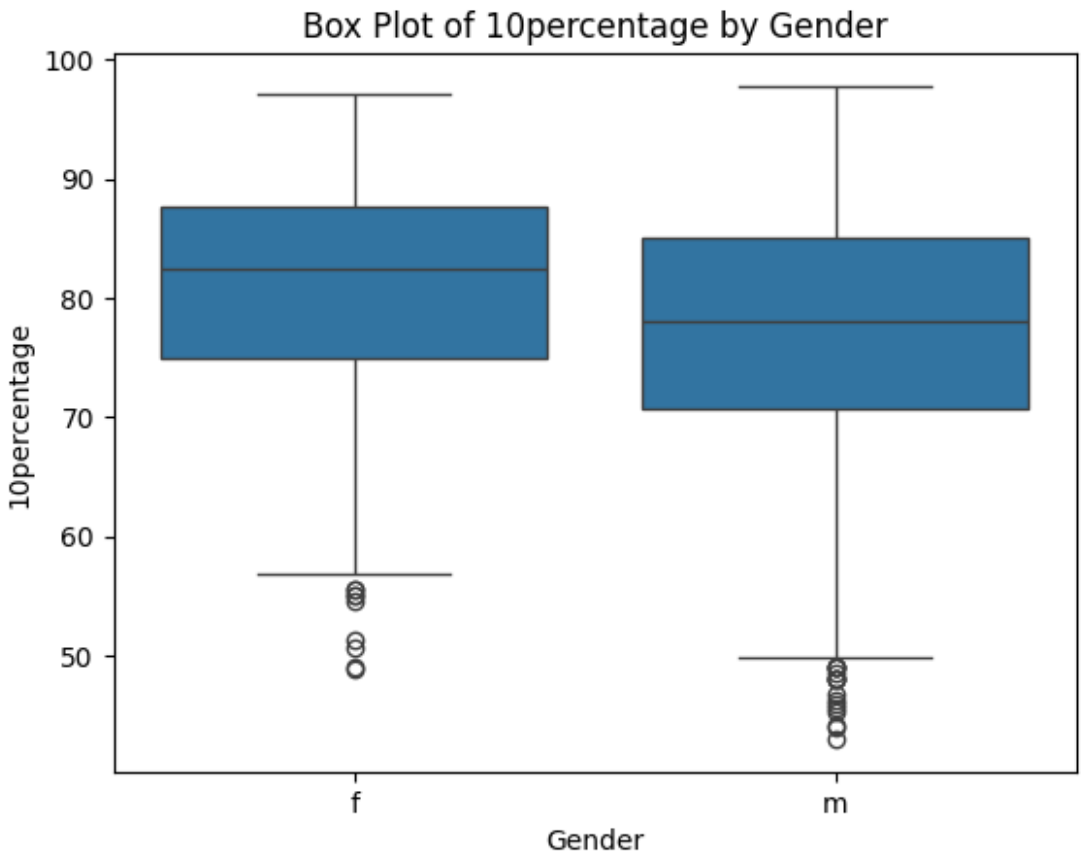

Box Plot of extraversion by JobCity
extraversion
2
1
0
−1
−2
−3
−4
JobCity

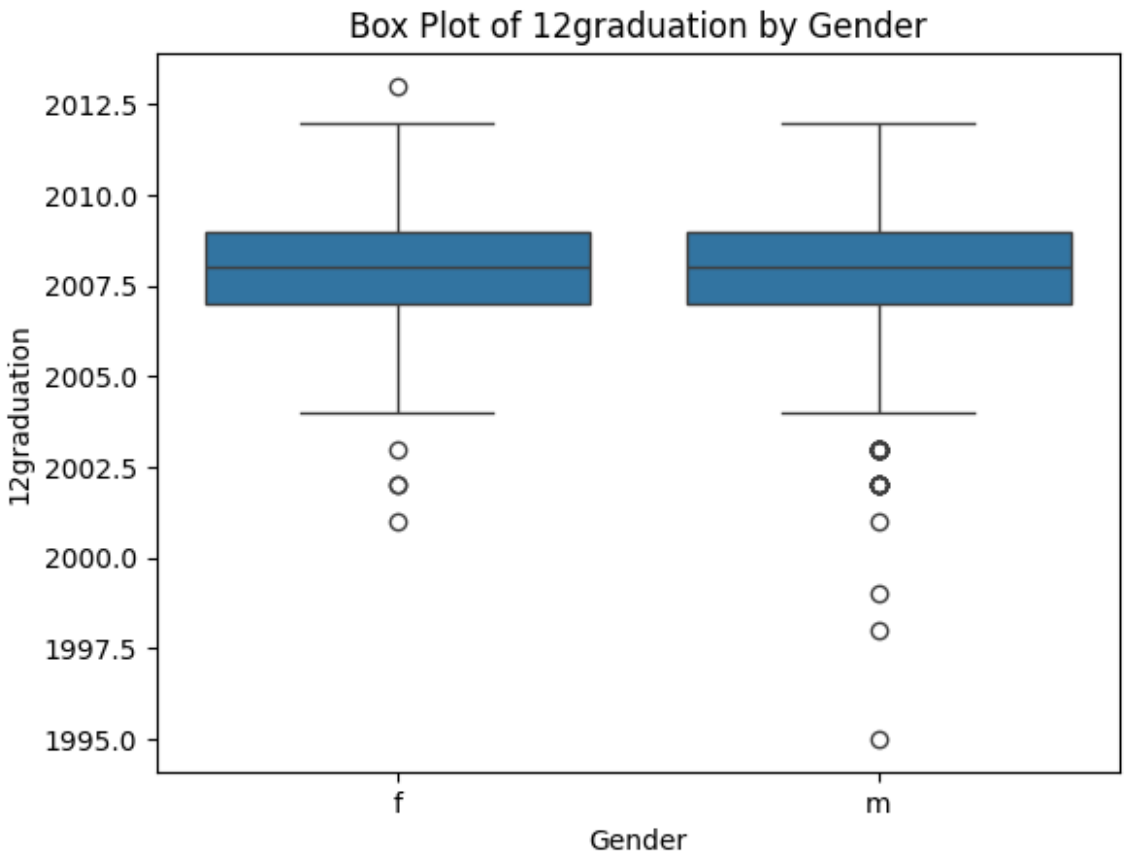

Box Plot of nueroticism by JobCity
nueroticism
3
2
1
0
−1
−2
JobCity

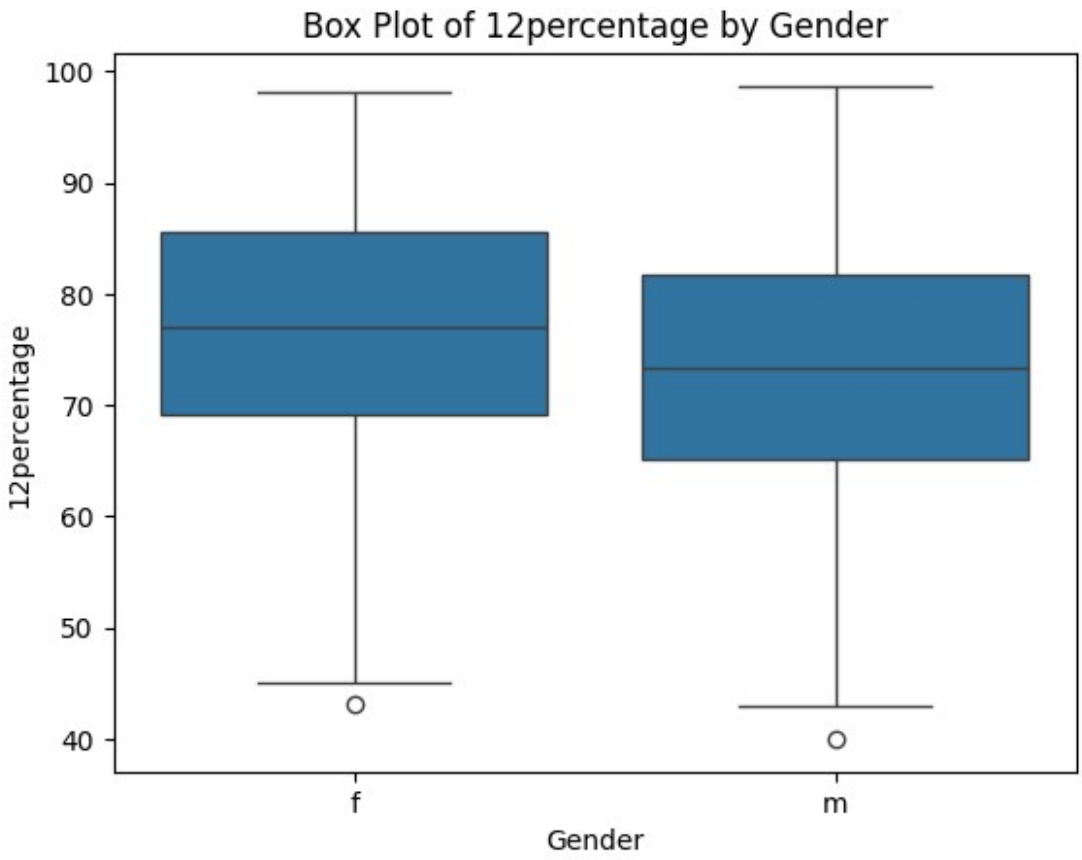

Box Plot of openess_to_experience by JobCity
openess_to_experience
2
0
−2
−4
−6
JobCity

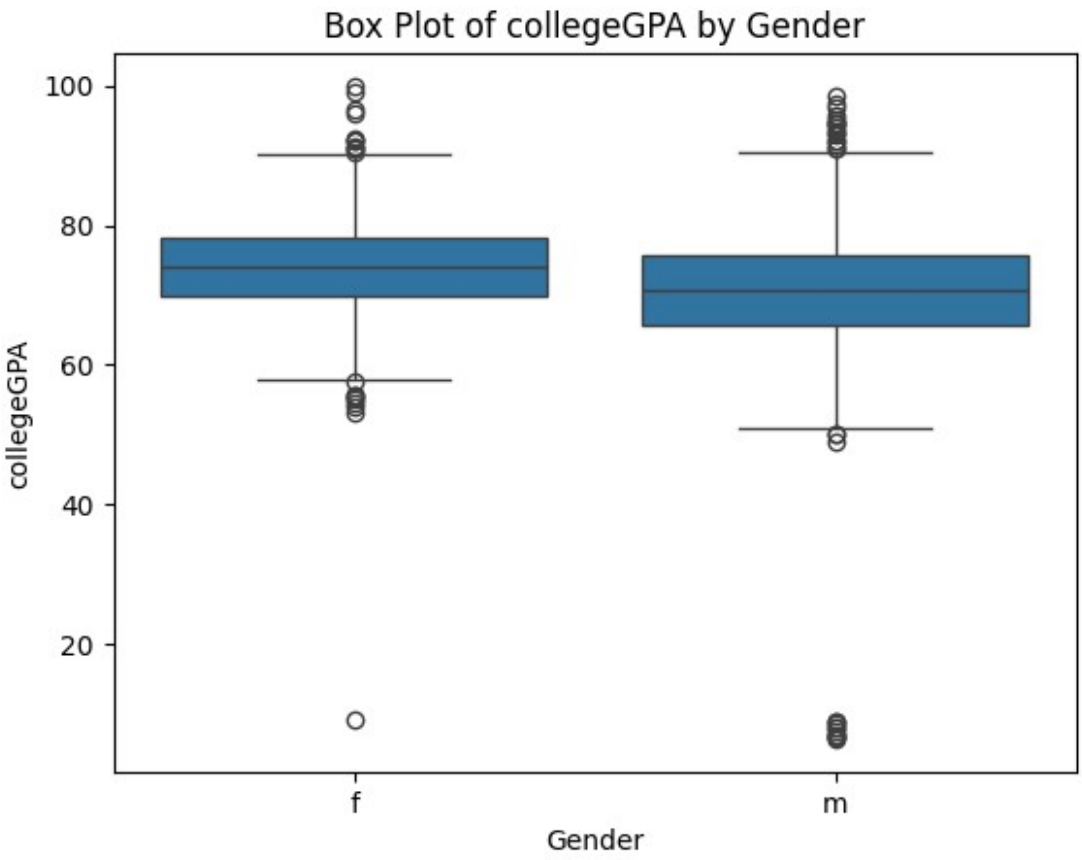

Box Plot of Age by JobCity
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
JobCity

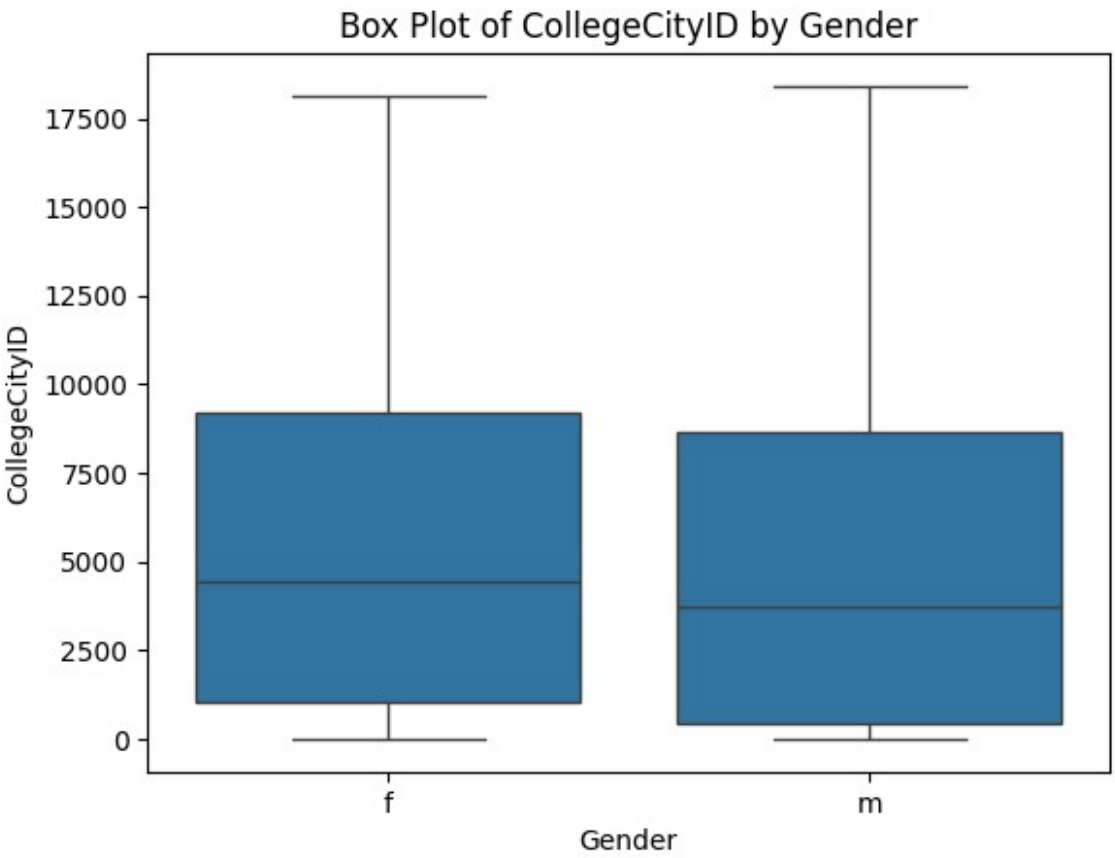

Box Plot of Salary by Gender
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Salary
f
m
Gender

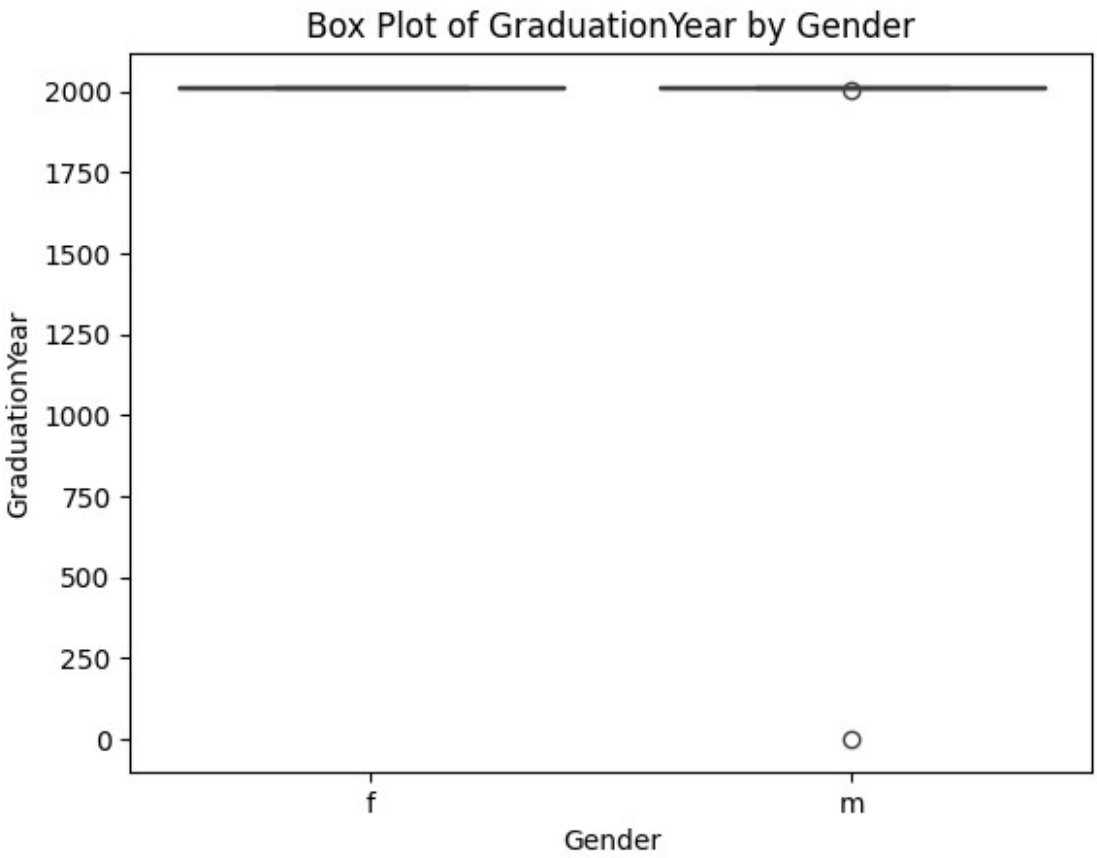

Box Plot of 10percentage by Gender
100
90
80
70
60
50
10percentage
f
m
Gender

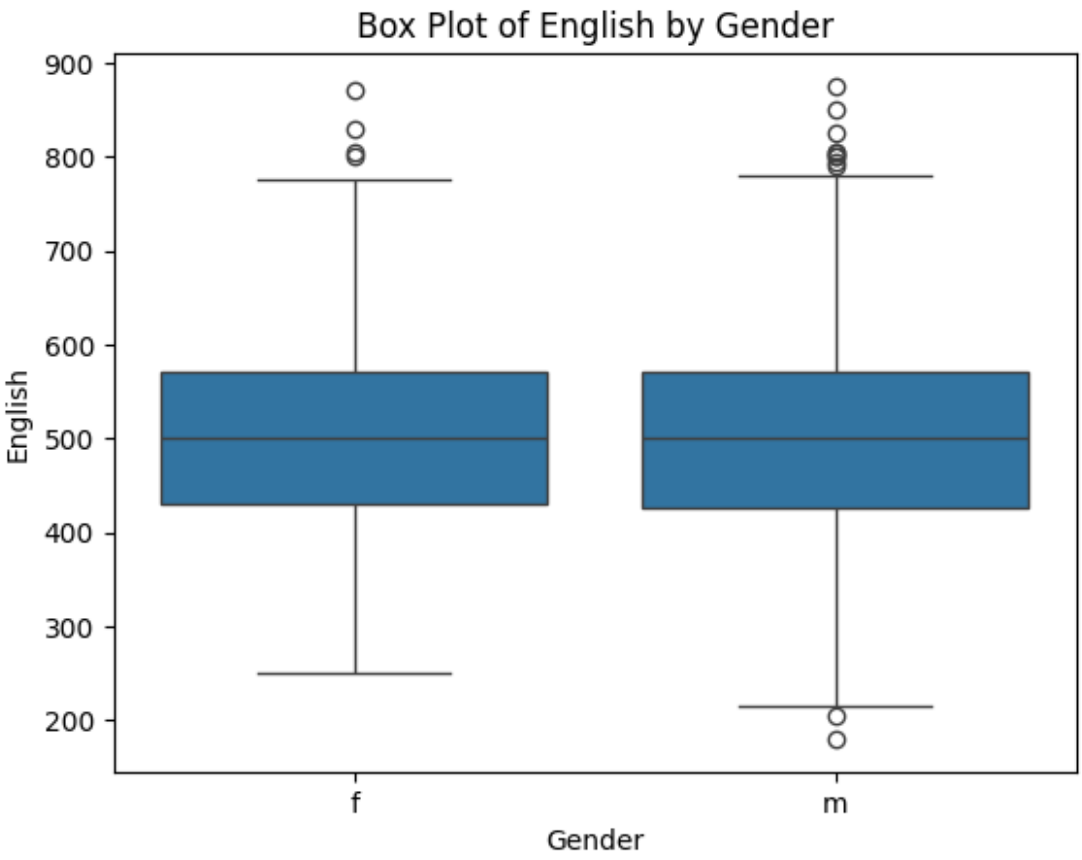

Box Plot of 12graduation by Gender
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
12graduation
f
m
Gender

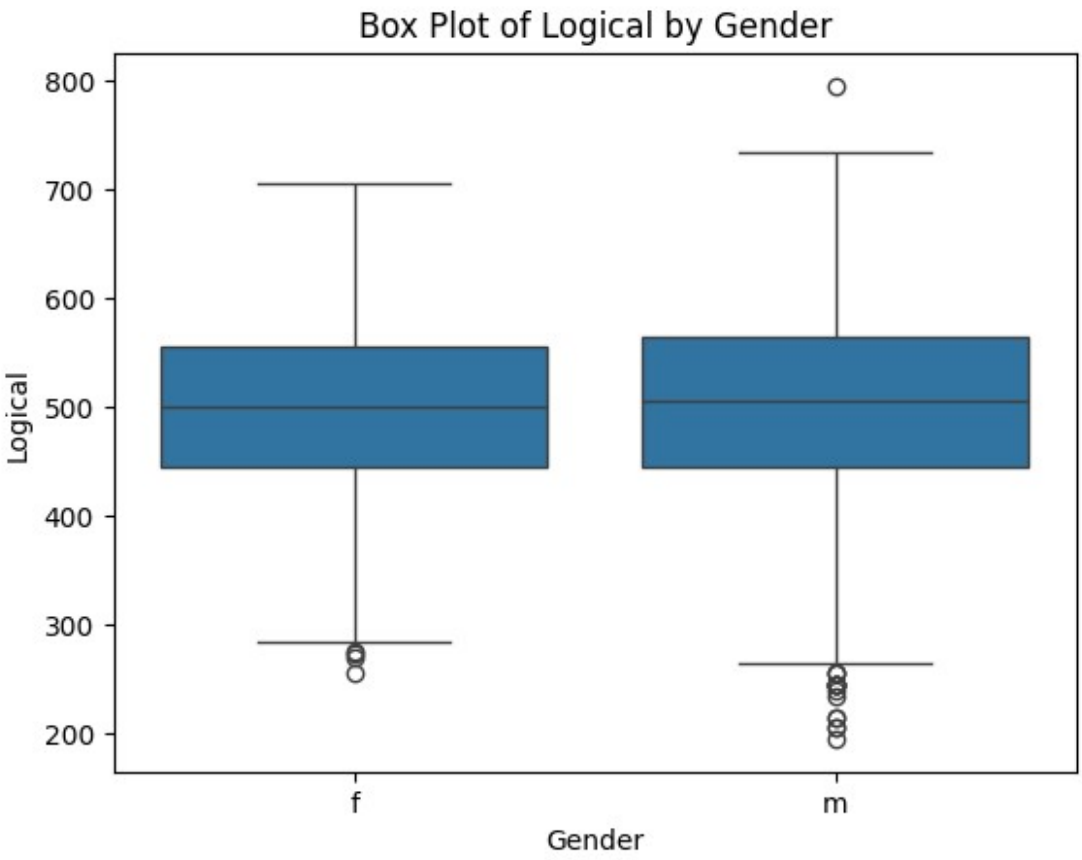

Box Plot of 12percentage by Gender
100
90
80
70
60
50
40
12percentage
f
m
Gender

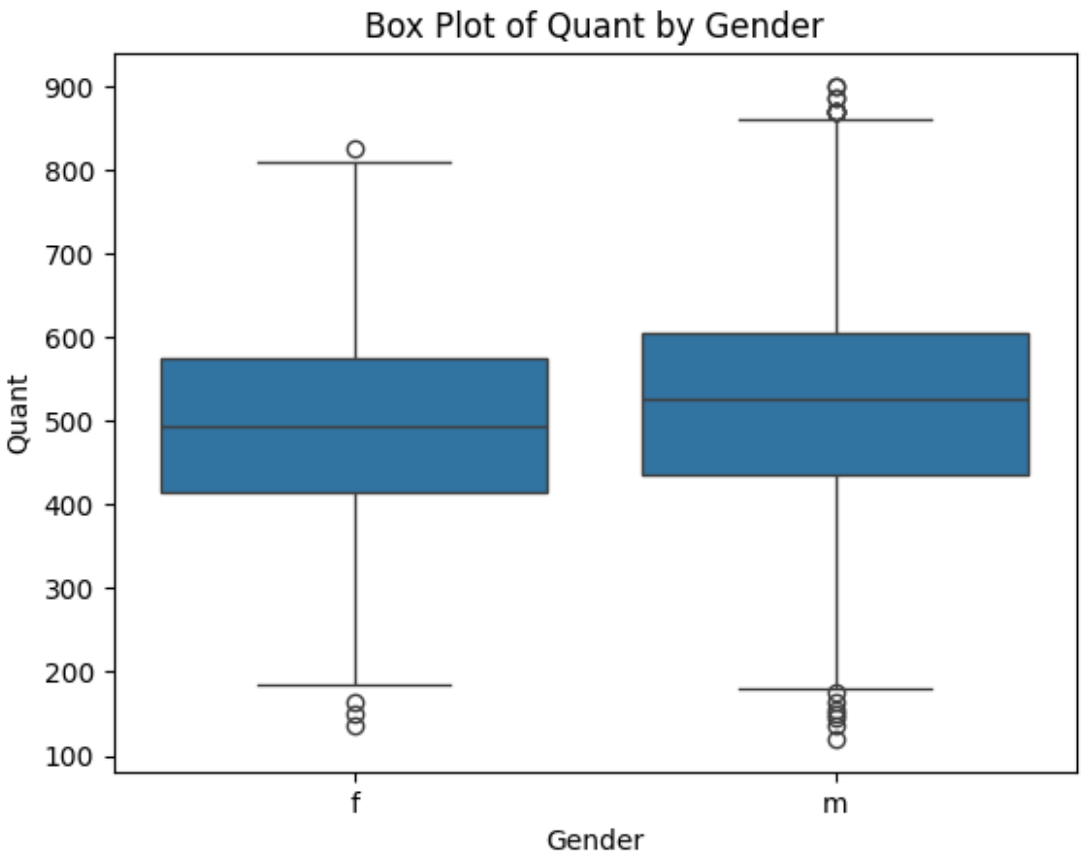

Box Plot of collegeGPA by Gender
100
80
60
40
20
collegeGPA
f
m
Gender

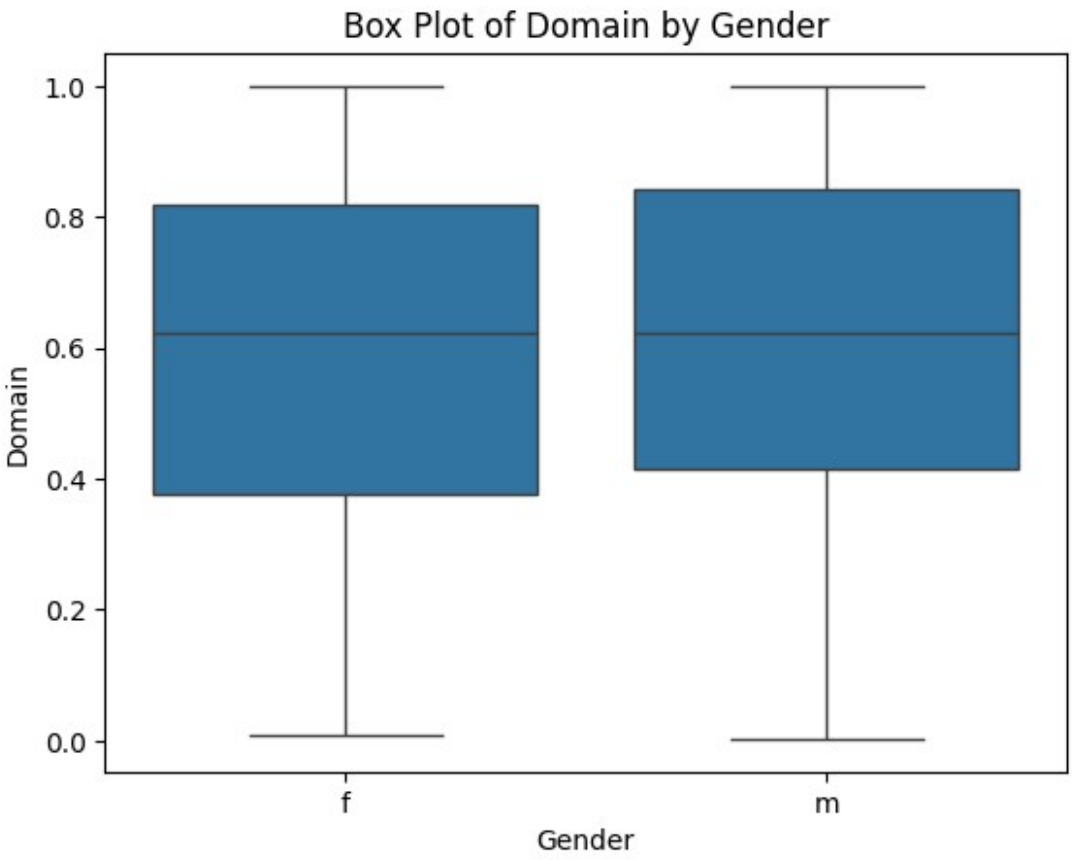

Box Plot of CollegeCityID by Gender
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
f
m
Gender

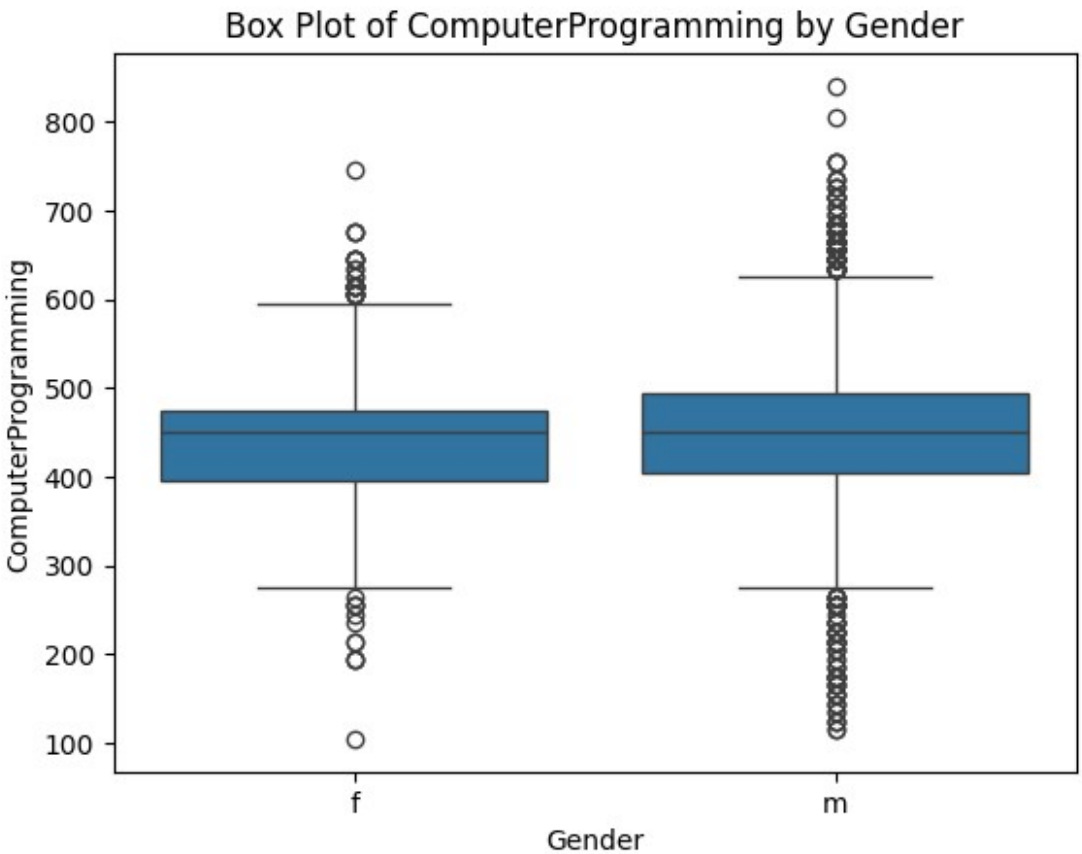

Box Plot of GraduationYear by Gender
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
f
m
Gender

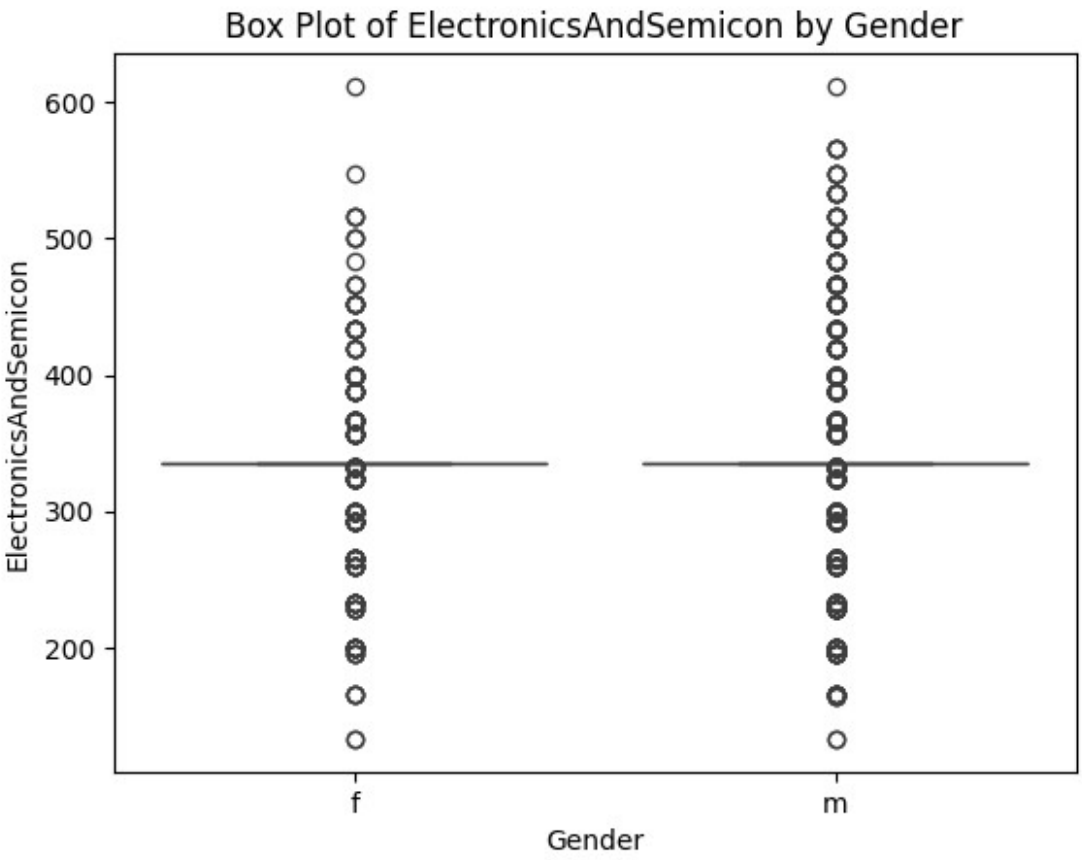

Box Plot of English by Gender
900
800
700
600
500
400
300
200
English
f
m
Gender


Box Plot of Logical by Gender
800
700
600
500
400
300
200
Logical
f
m
Gender

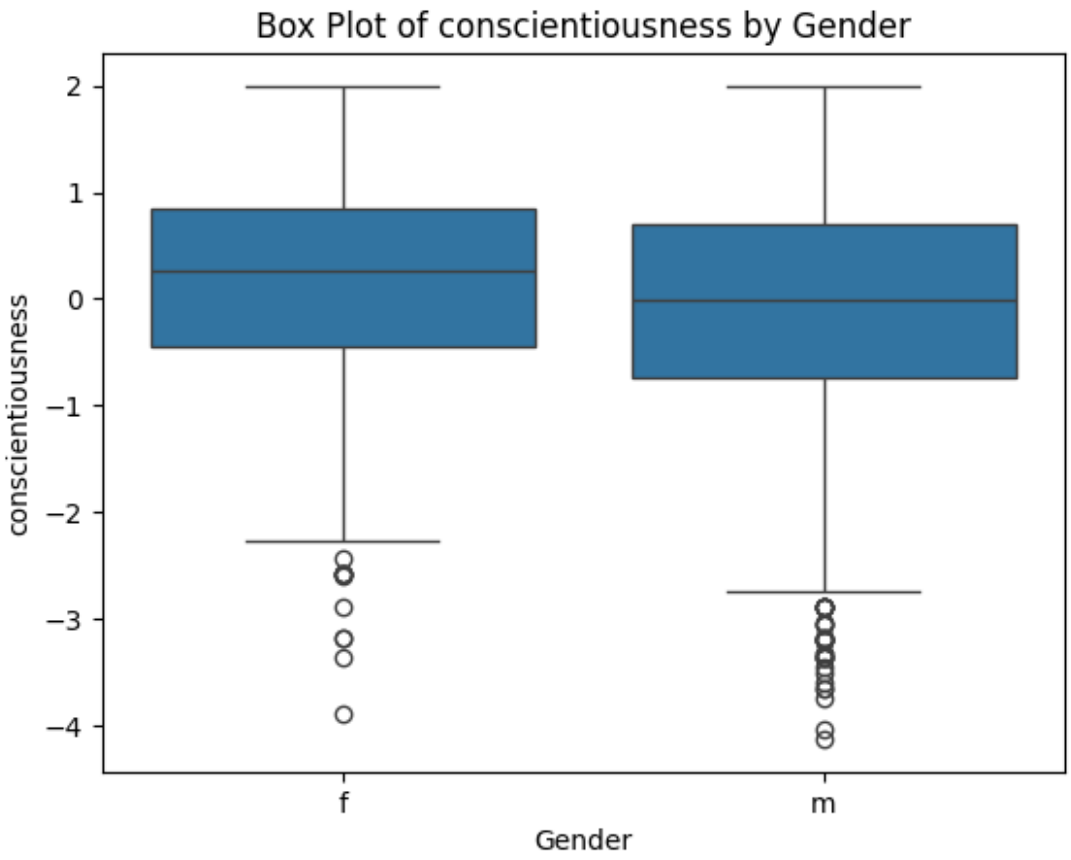

Box Plot of Quant by Gender
900
800
700
600
500
400
300
200
100
Quant
f
m
Gender

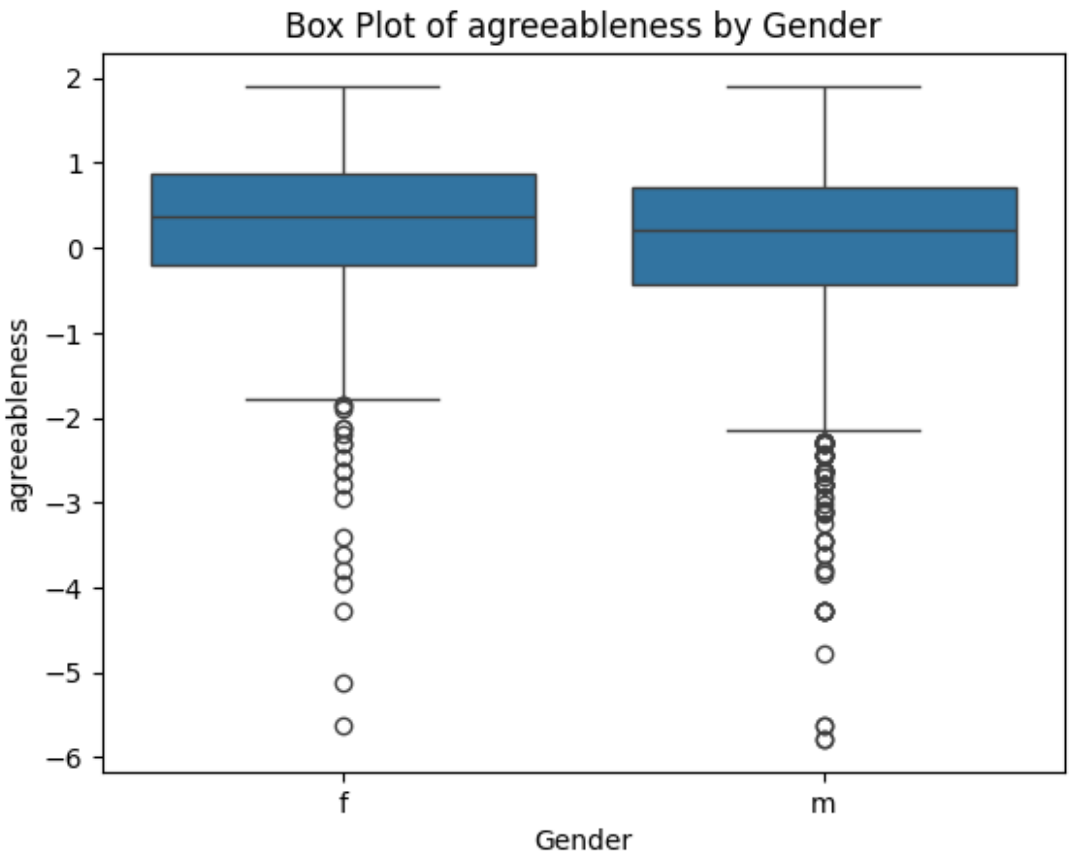

Box Plot of Domain by Gender
1.0
0.8
0.6
0.4
0.2
0.0
Domain
f
m
Gender

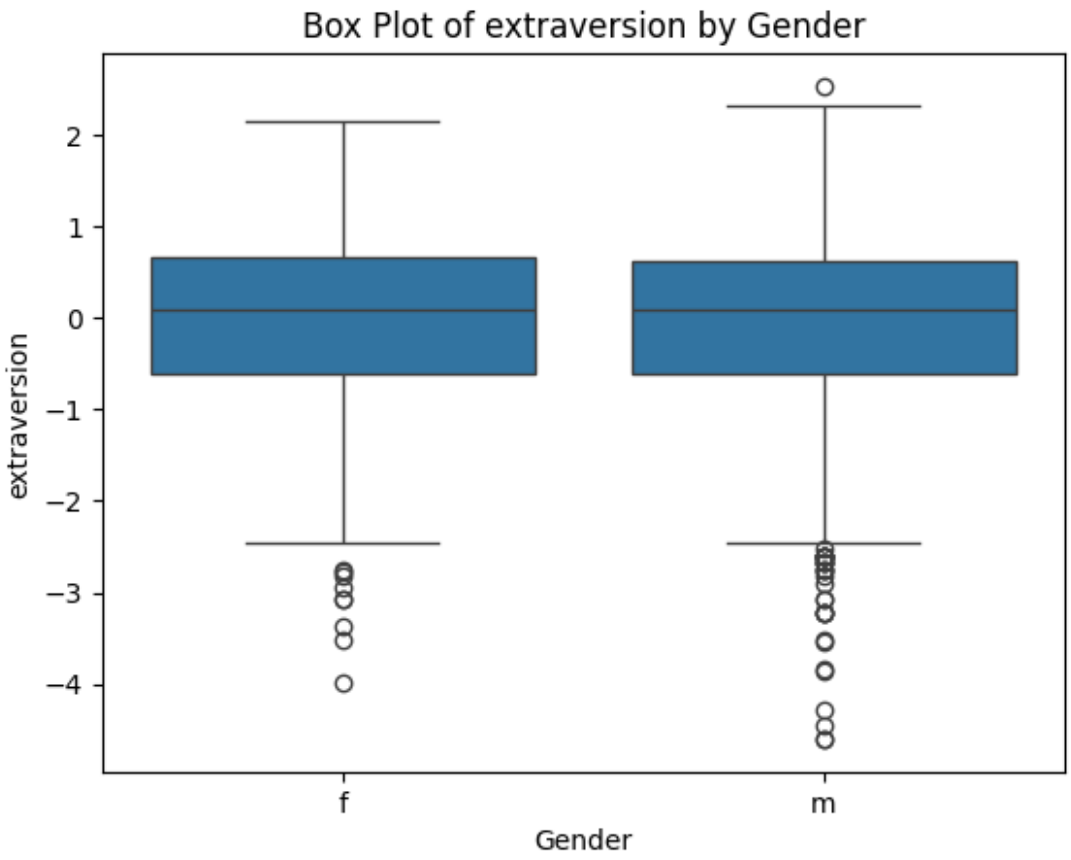

Box Plot of ComputerProgramming by Gender
ComputerProgramming
800
700
600
500
400
300
200
100
f
m
Gender

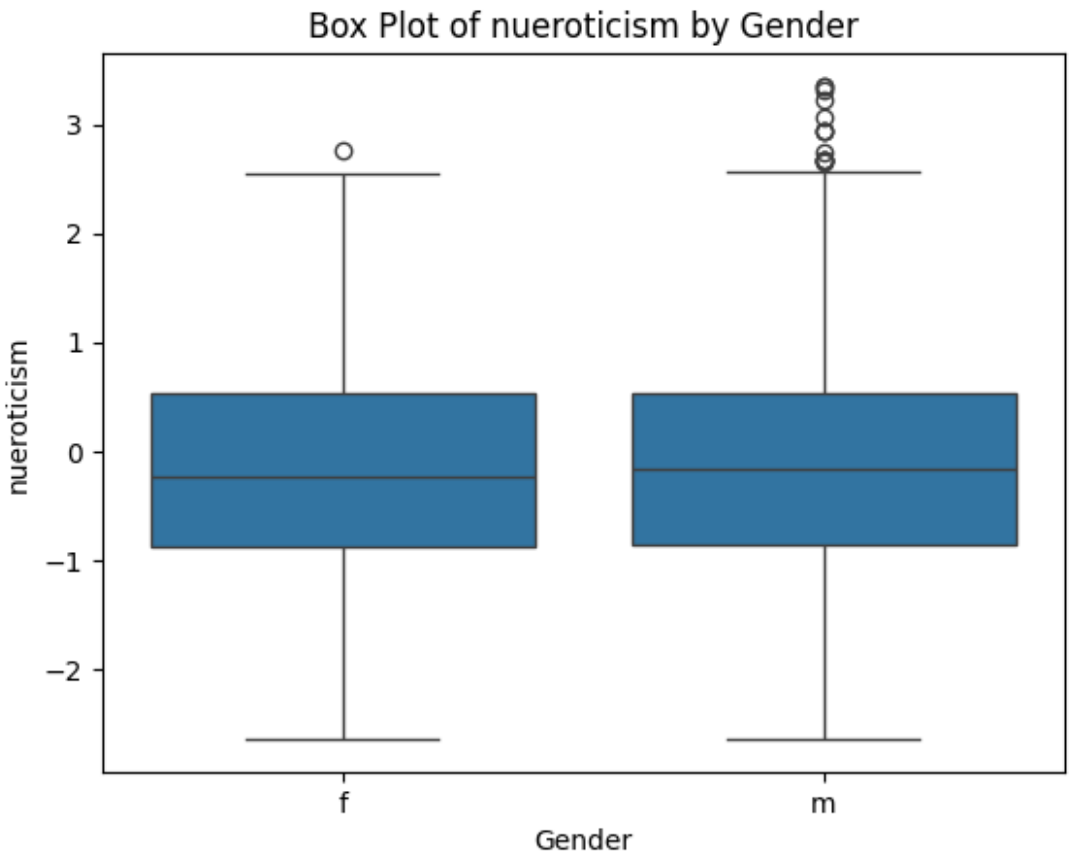

Box Plot of ElectronicsAndSemicon by Gender
ElectronicsAndSemicon
600
500
400
300
200
f
m
Gender

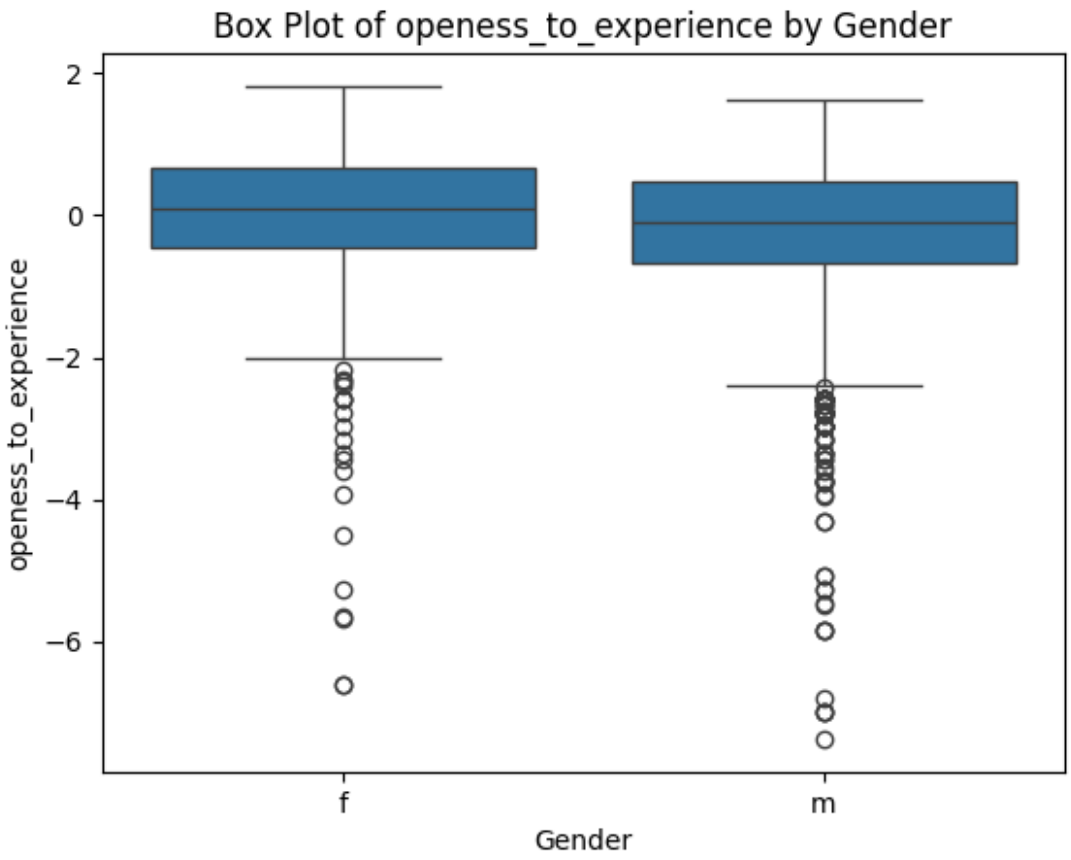
Box Plot of ComputerScience by Gender
ComputerScience
700
600
500
400
300
200
f
m
Gender

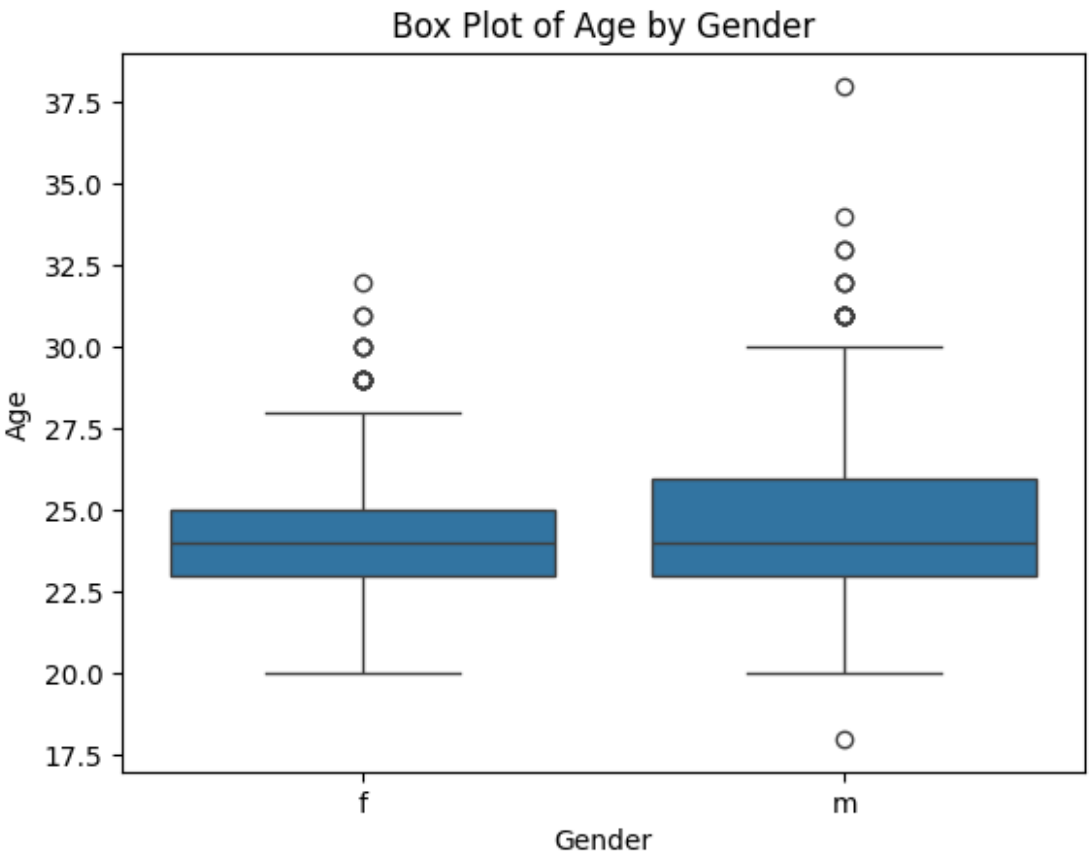

Box Plot of conscientiousness by Gender
conscientiousness
2
1
0
−1
−2
−3
−4
f
m
Gender

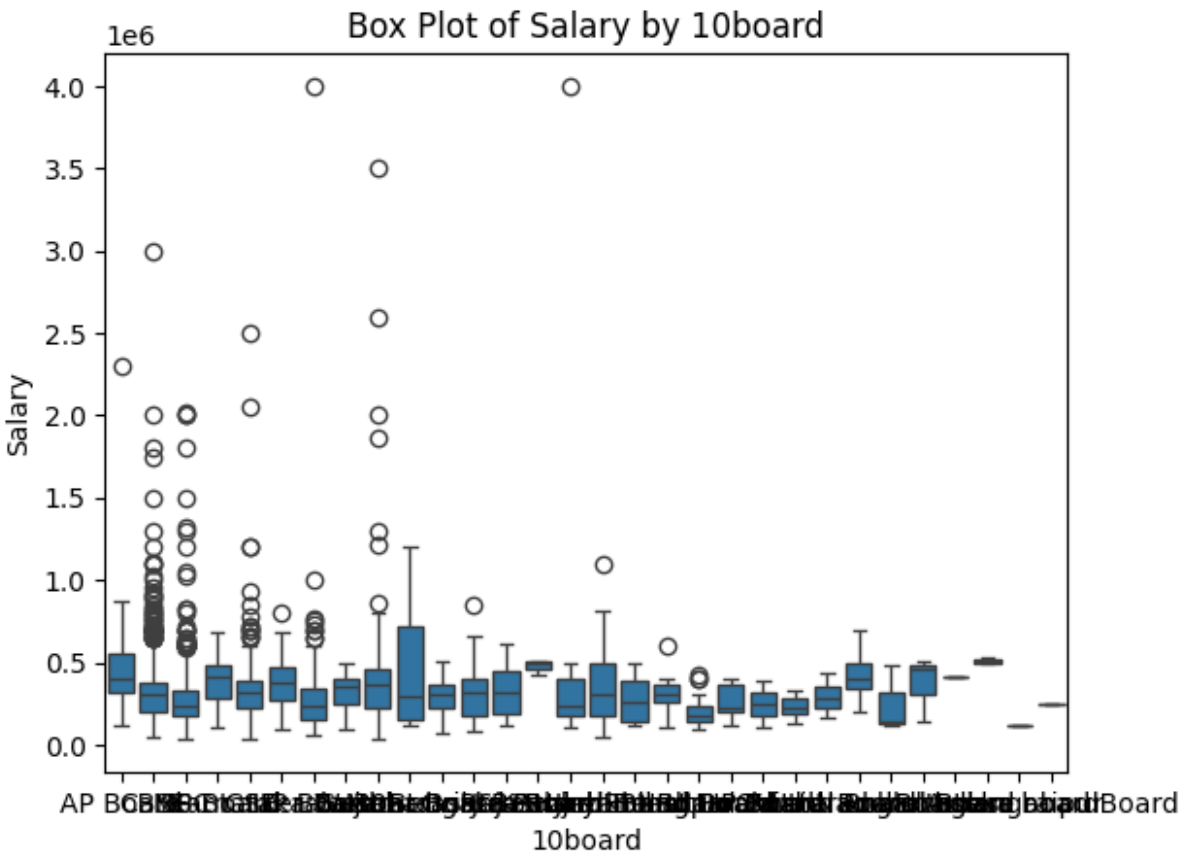

Box Plot of agreeableness by Gender
agreeableness
Gender
f
m
2
1
0
−1
−2
−3
−4
−5
−6

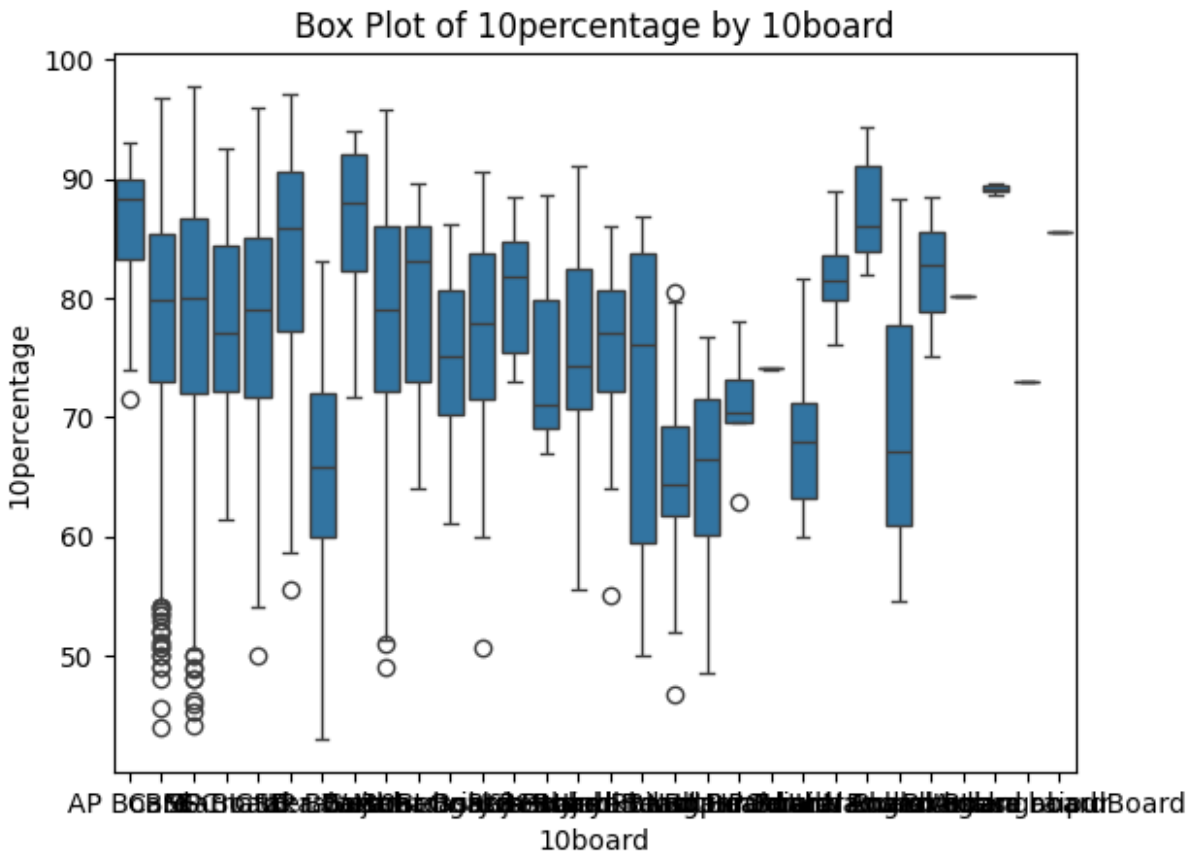

Box Plot of extraversion by Gender
extraversion
2
1
0
−1
−2
−3
−4
f
m
Gender

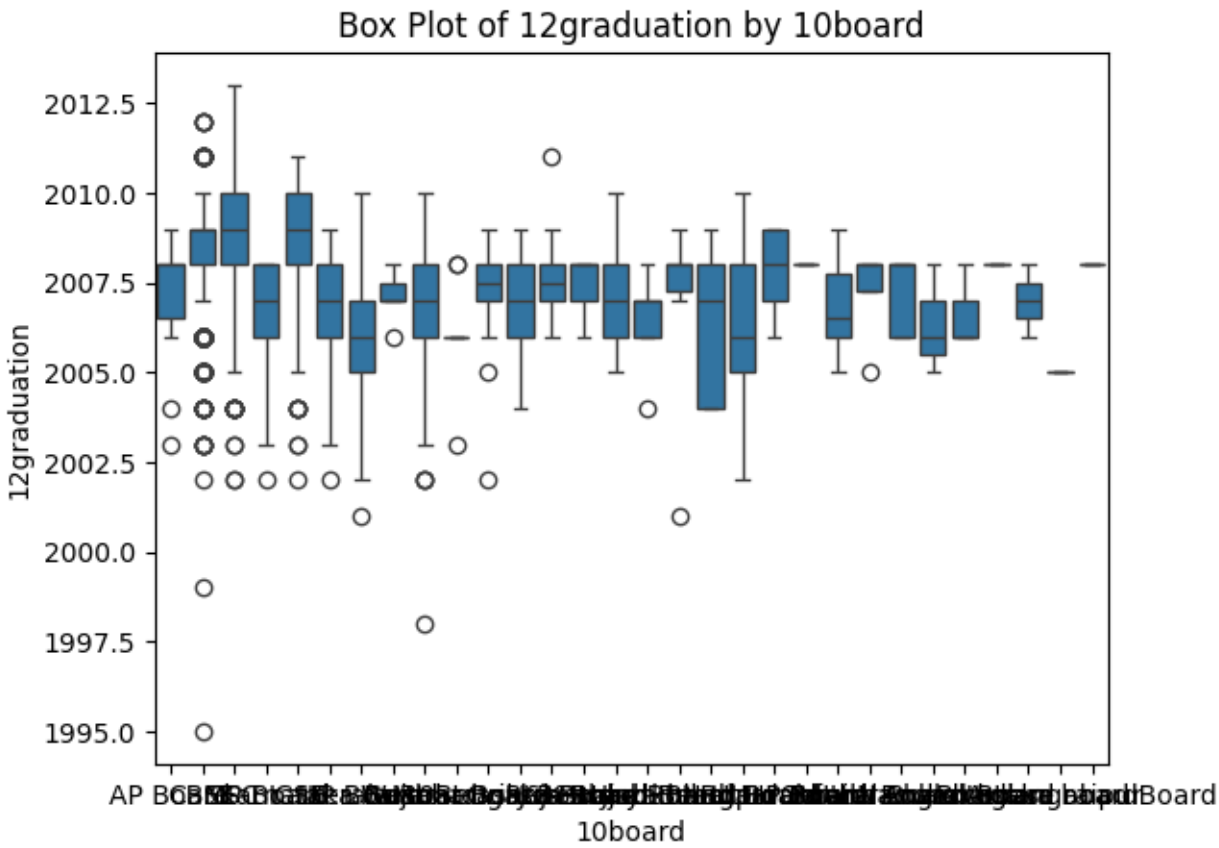

Box Plot of nueroticism by Gender
nueroticism
3
2
1
0
−1
−2
f
m
Gender

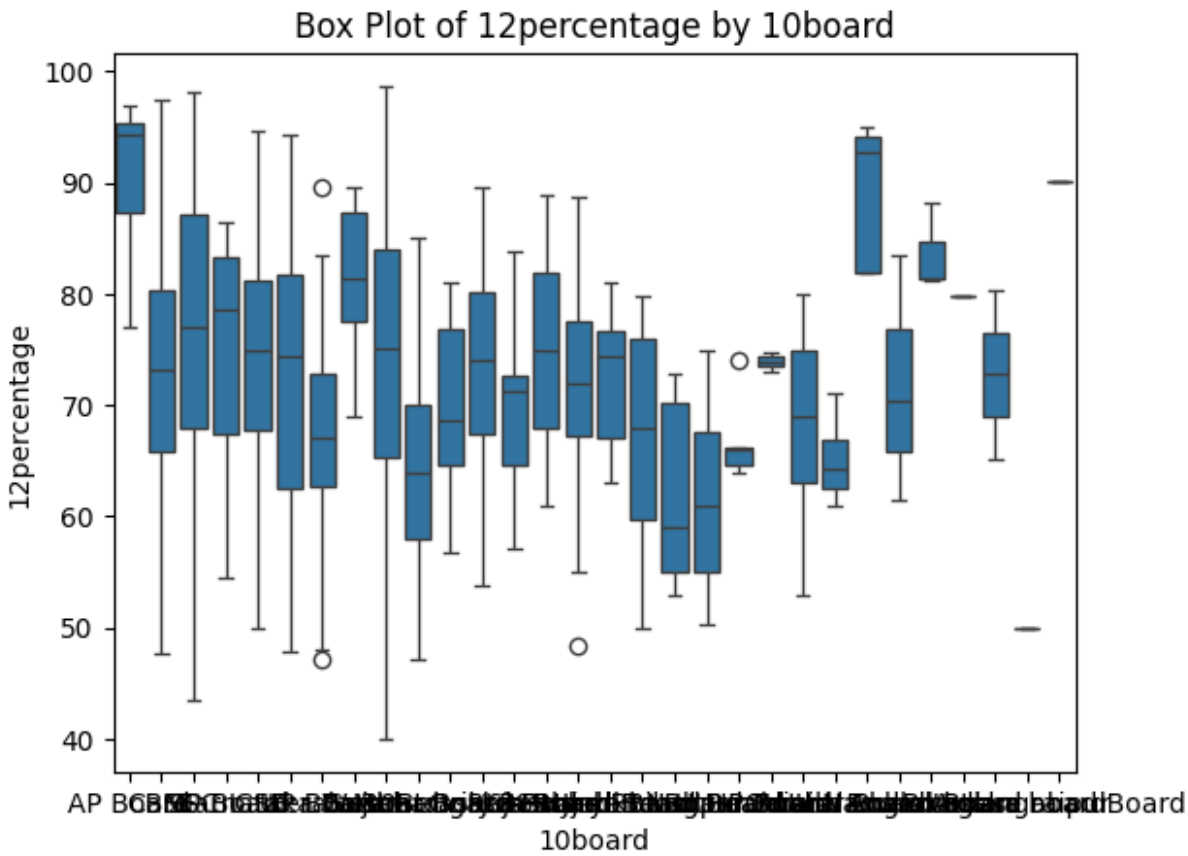

Box Plot of openess_to_experience by Gender
2
0
−2
−4
−6
openess_to_experience
f
m
Gender

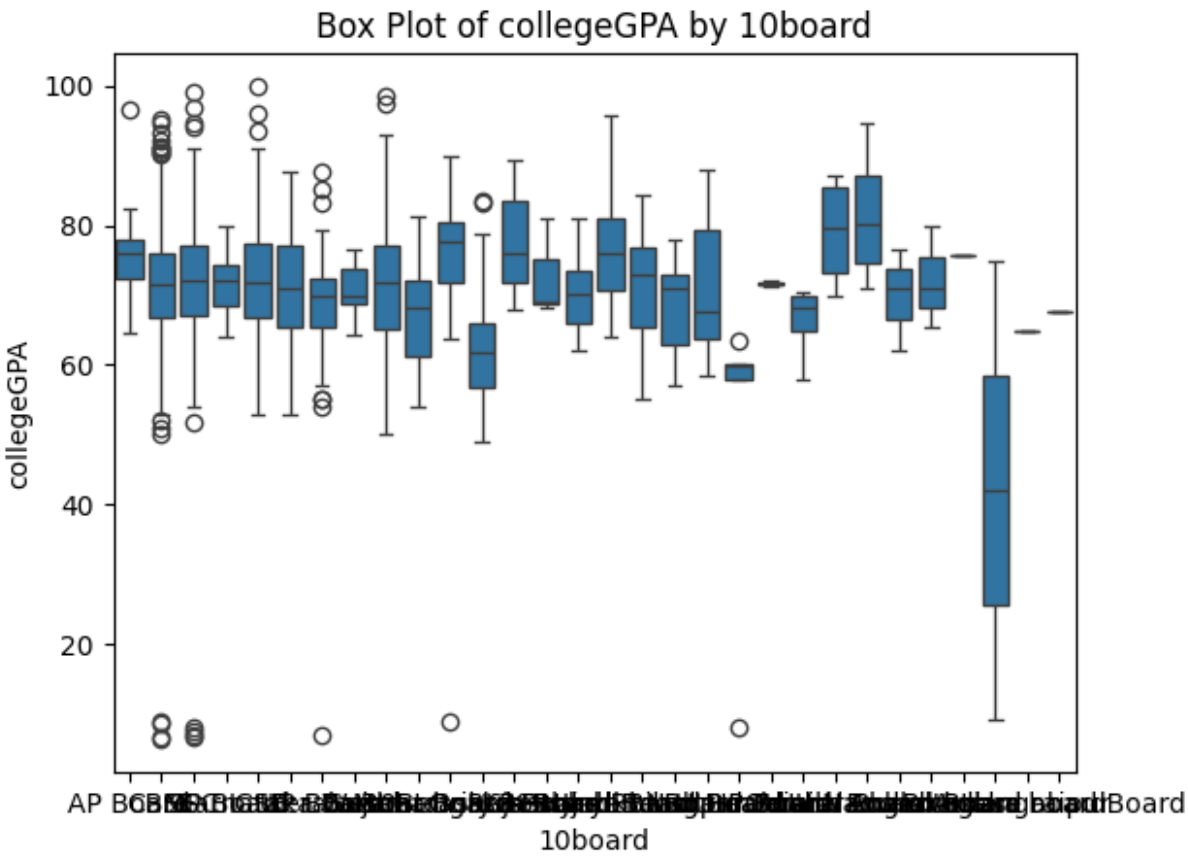

Box Plot of Age by Gender
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
Age
f
m
Gender

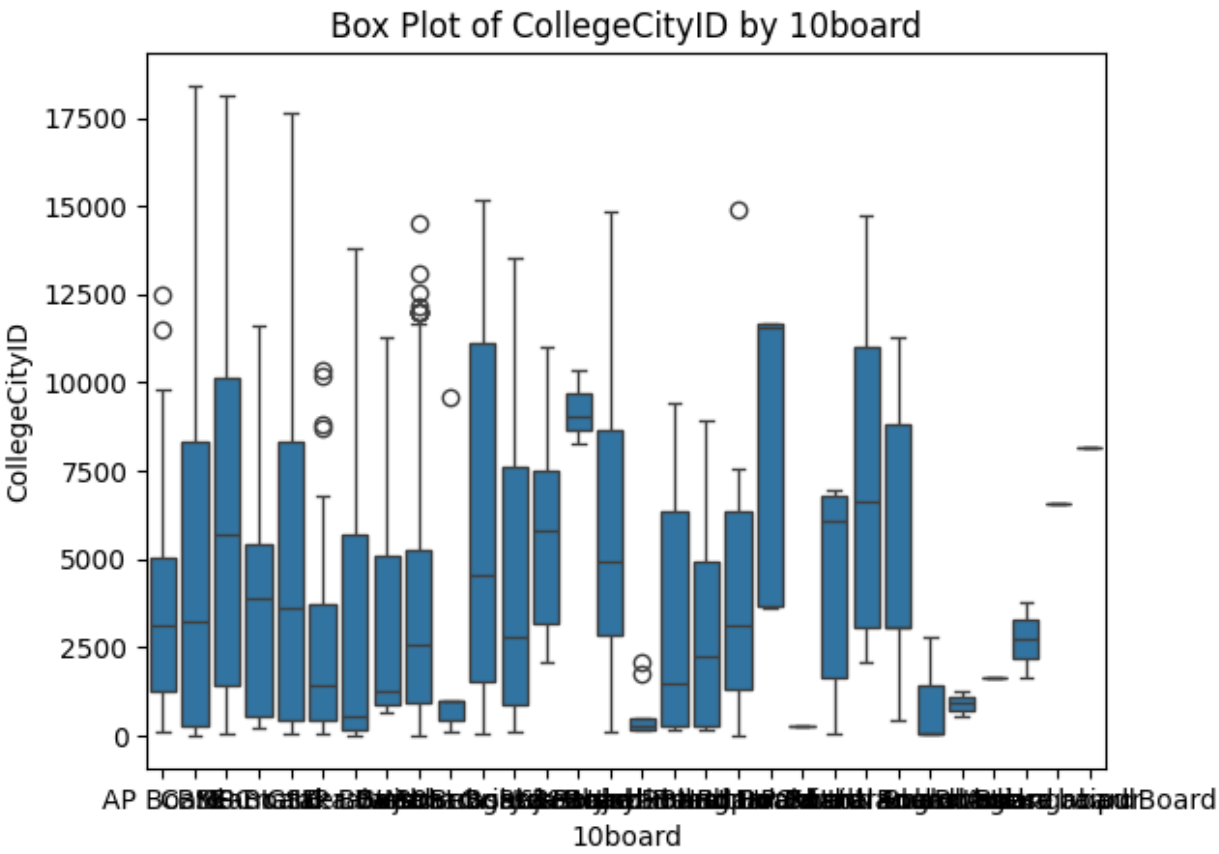

Box Plot of Salary by 10board
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Salary
10board

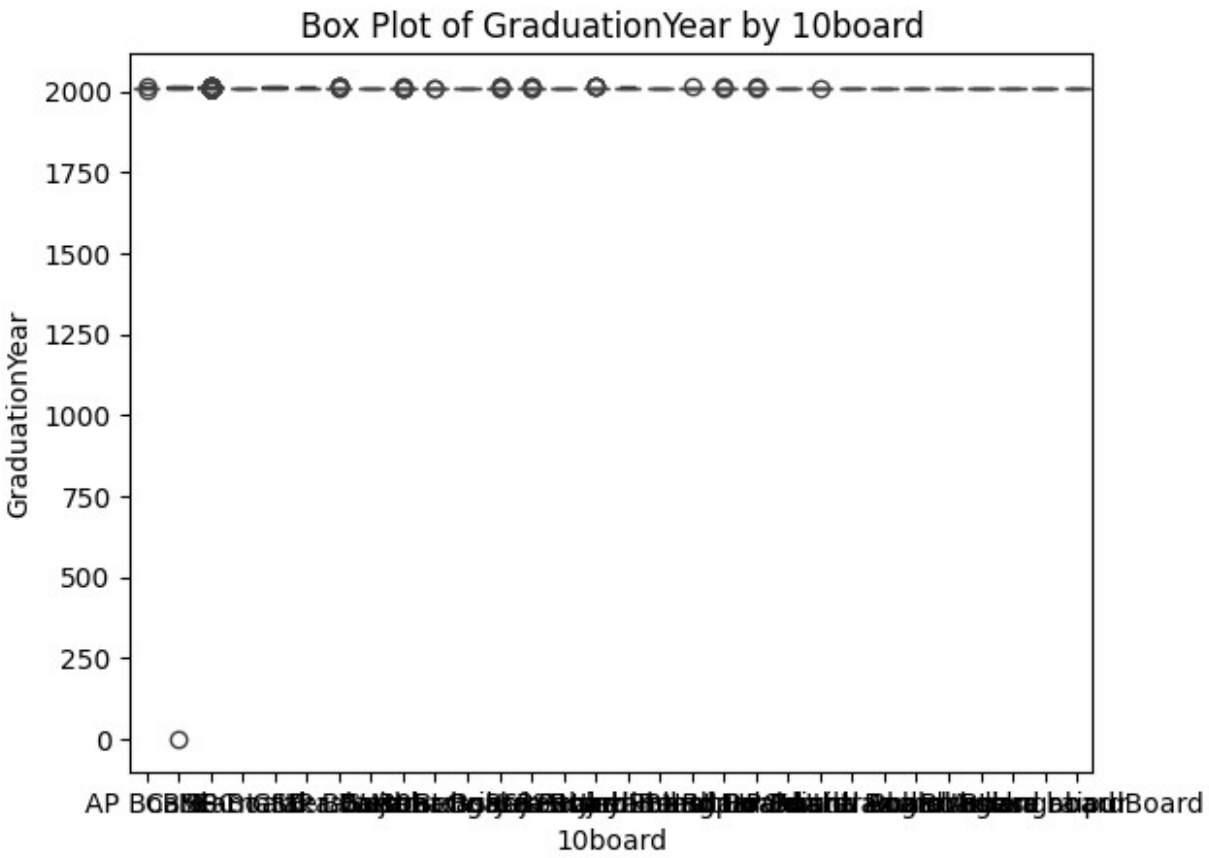

Box Plot of 10percentage by 10board
10percentage
100
90
80
70
60
50
10board

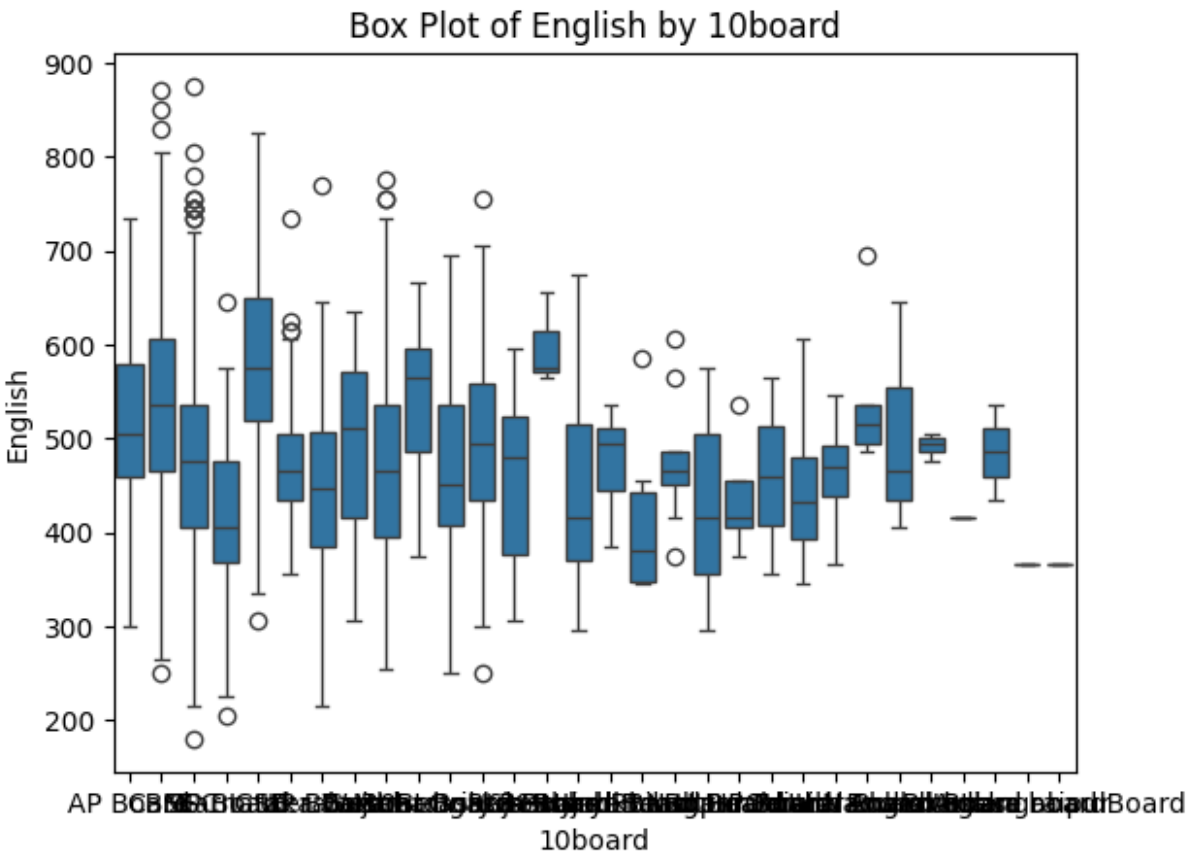

Box Plot of 12graduation by 10board
12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
10board

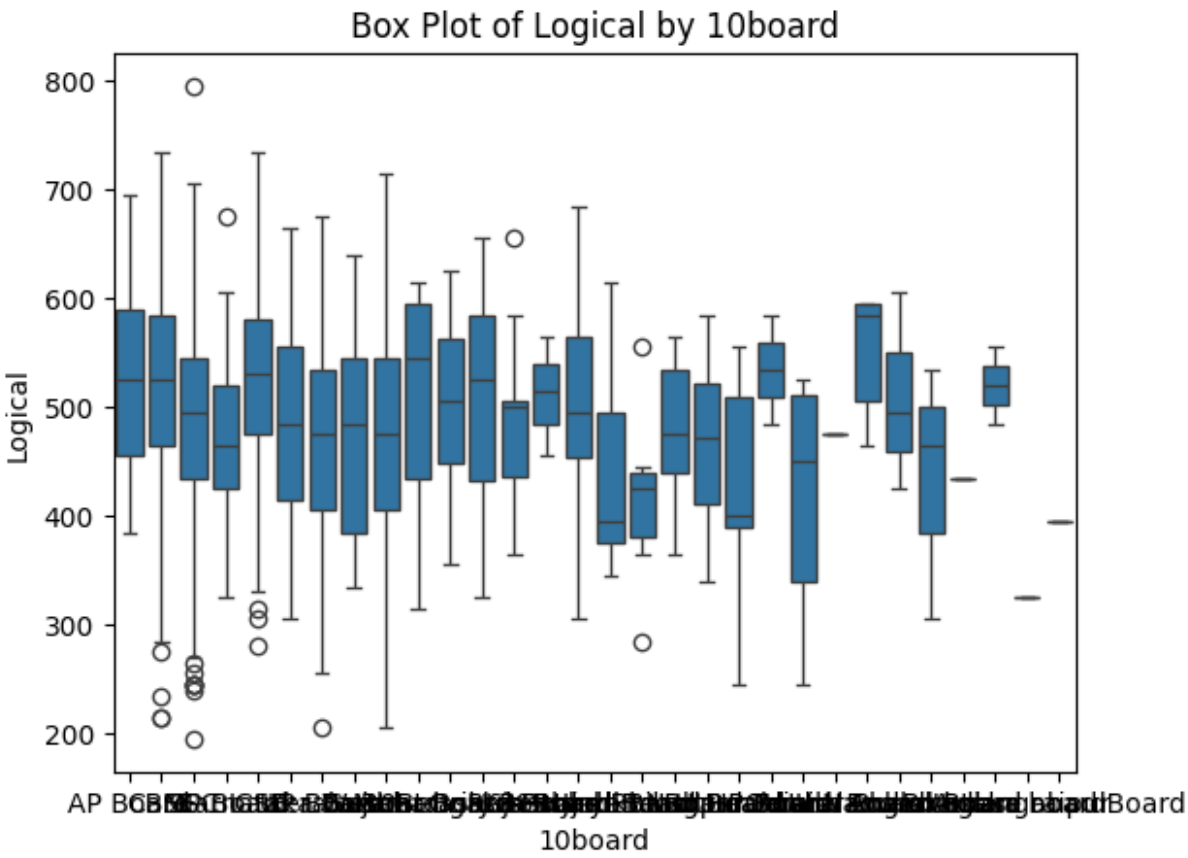

Box Plot of 12percentage by 10board
12percentage
100
90
80
70
60
50
40
10board

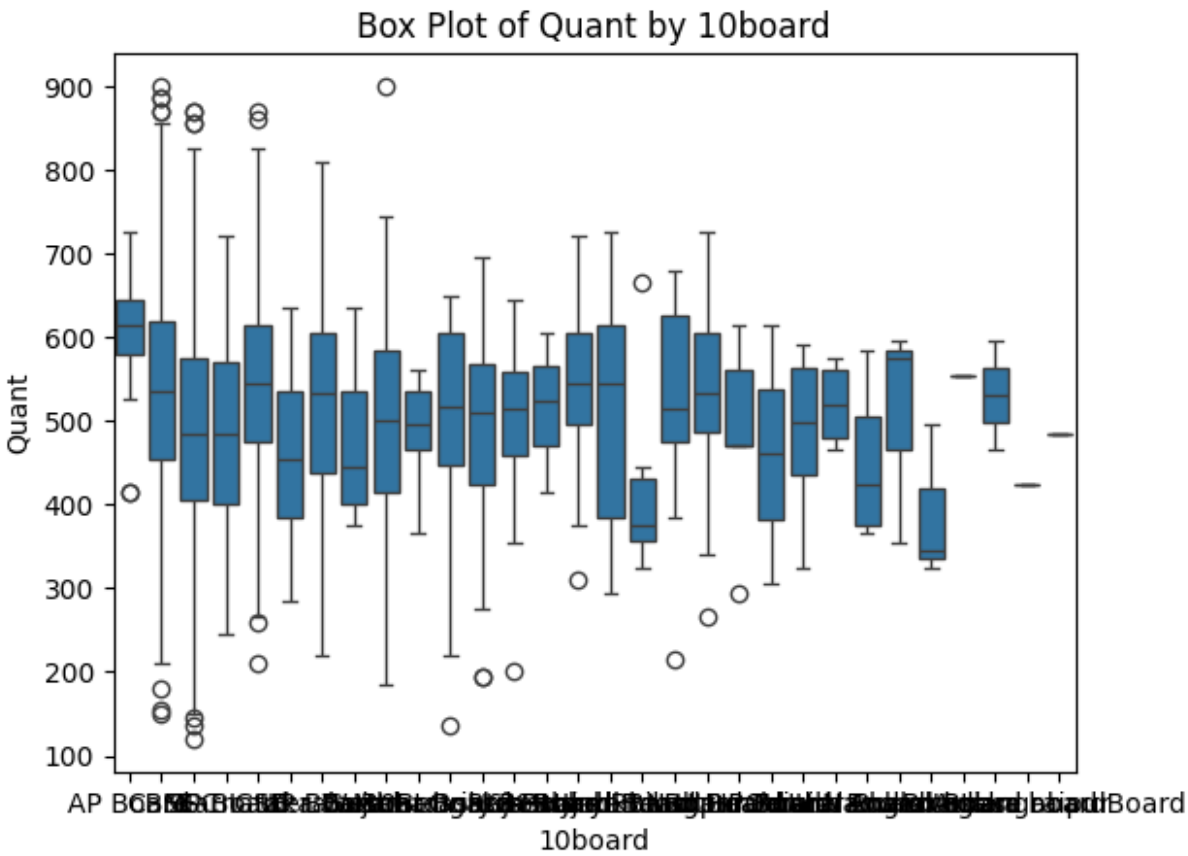

Box Plot of collegeGPA by 10board
collegeGPA
100
80
60
40
20
10board

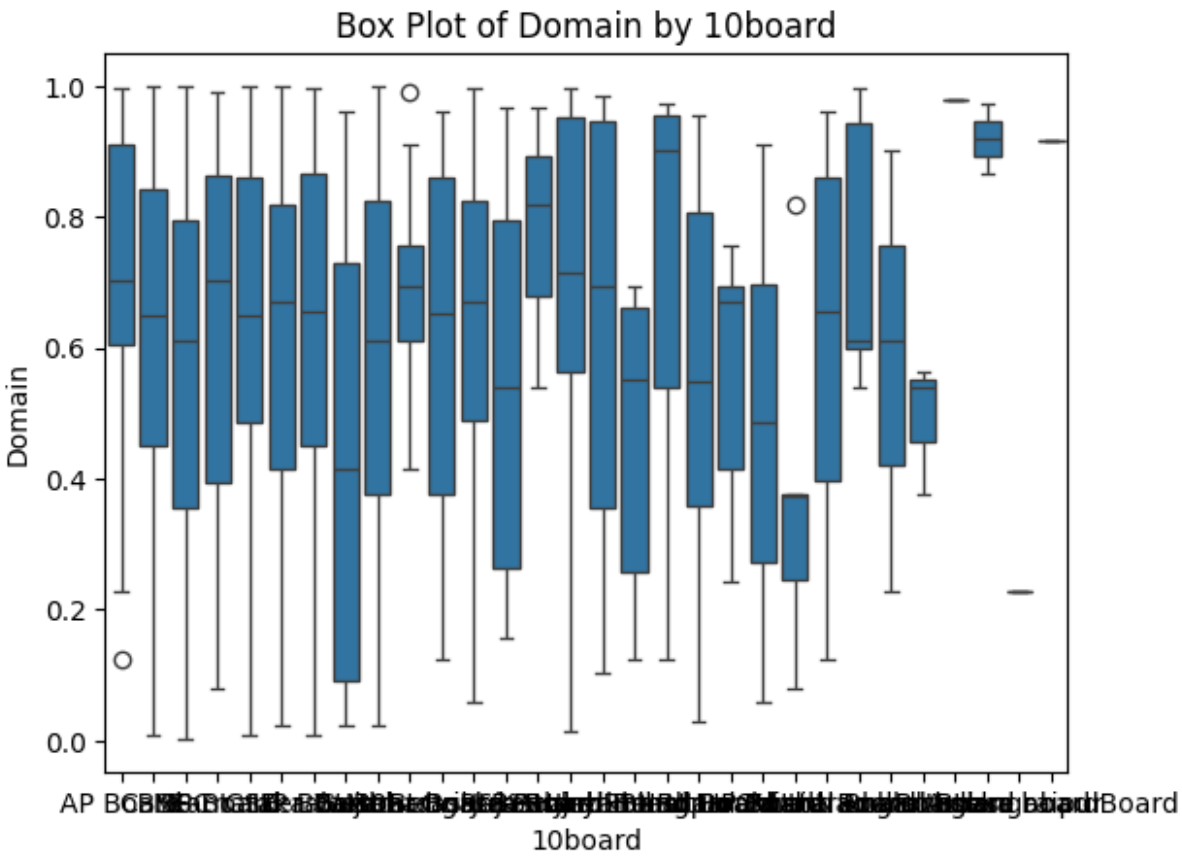

Box Plot of CollegeCityID by 10board
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
10board

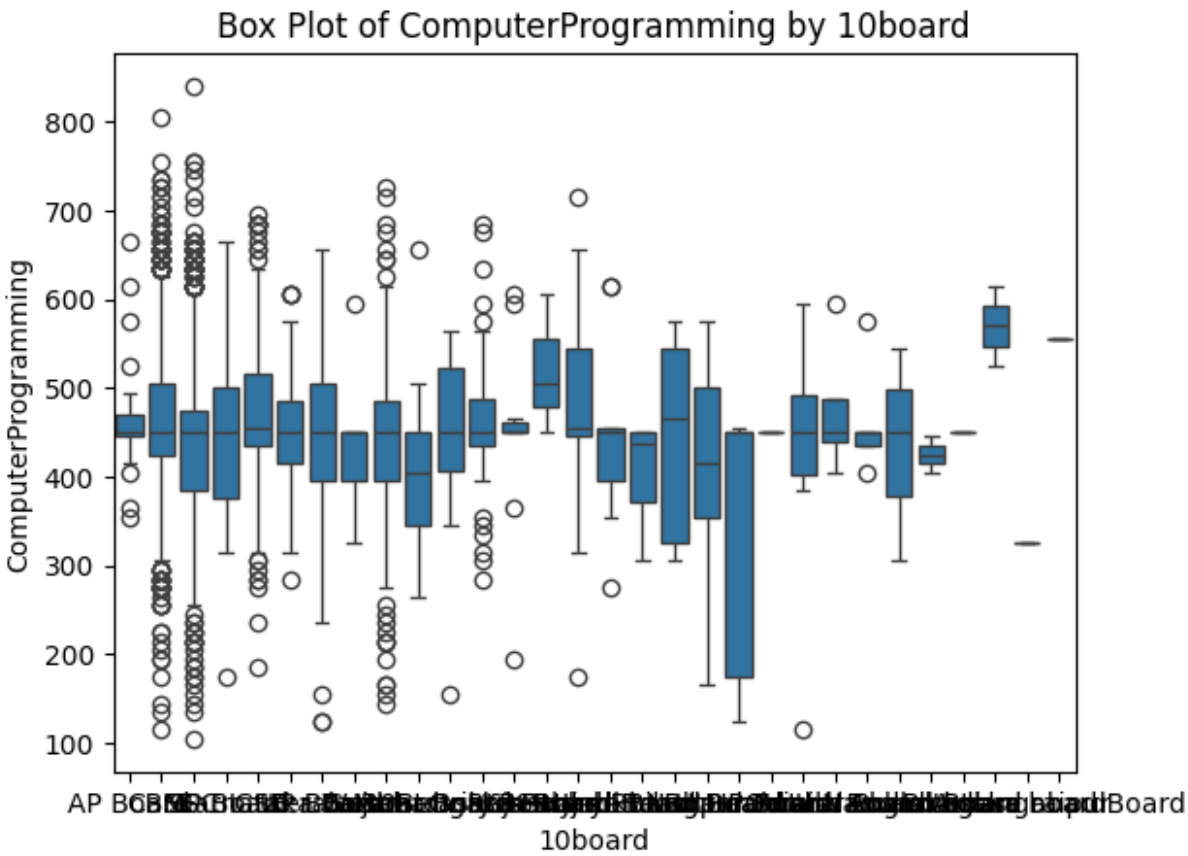

Box Plot of GraduationYear by 10board
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
10board

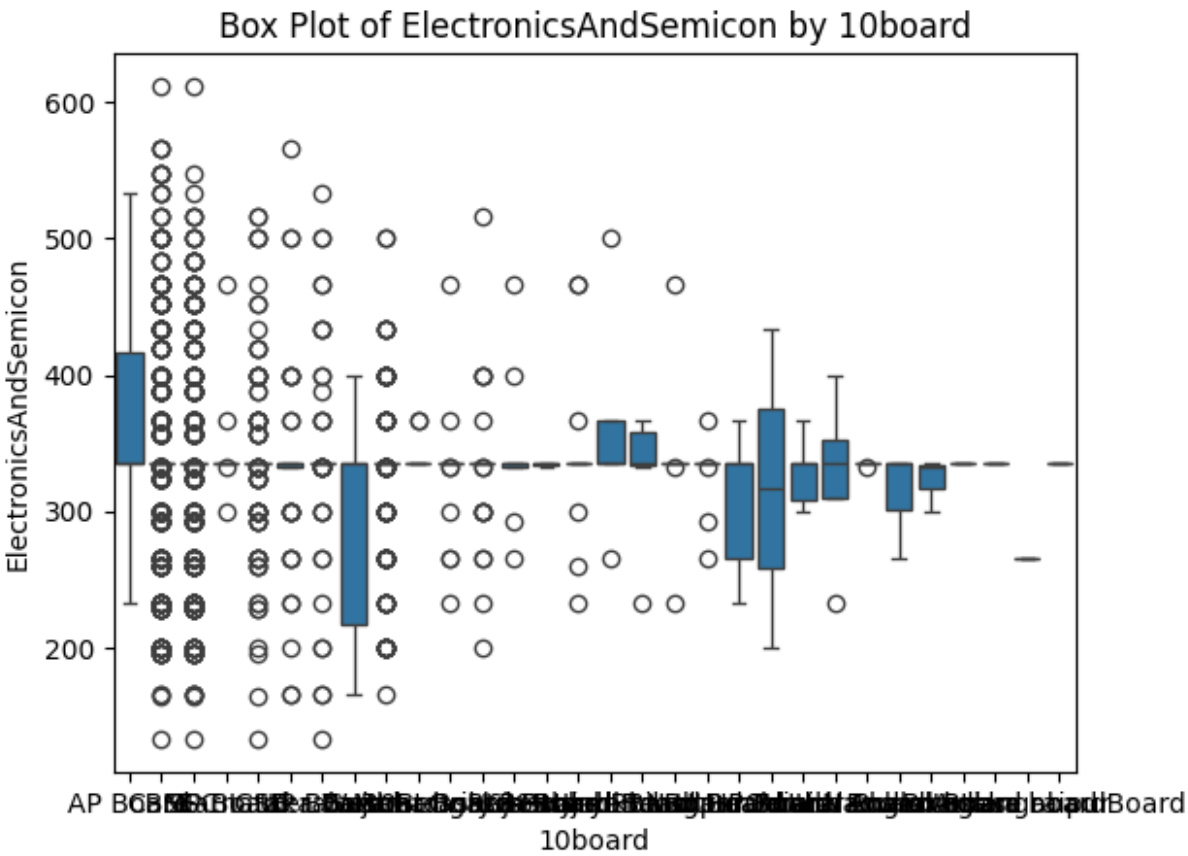

Box Plot of English by 10board
English
900
800
700
600
500
400
300
200
10board

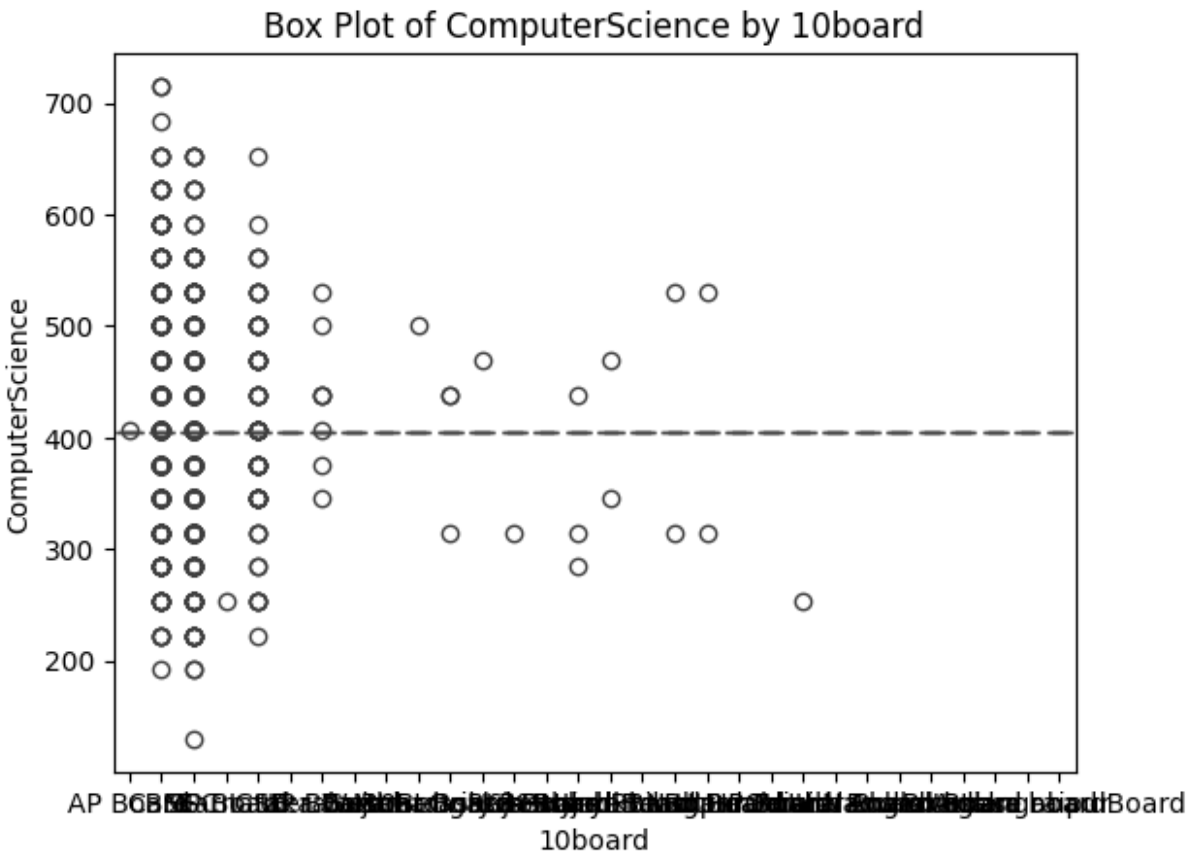

Box Plot of Logical by 10board
Logical
800
700
600
500
400
300
200
10board

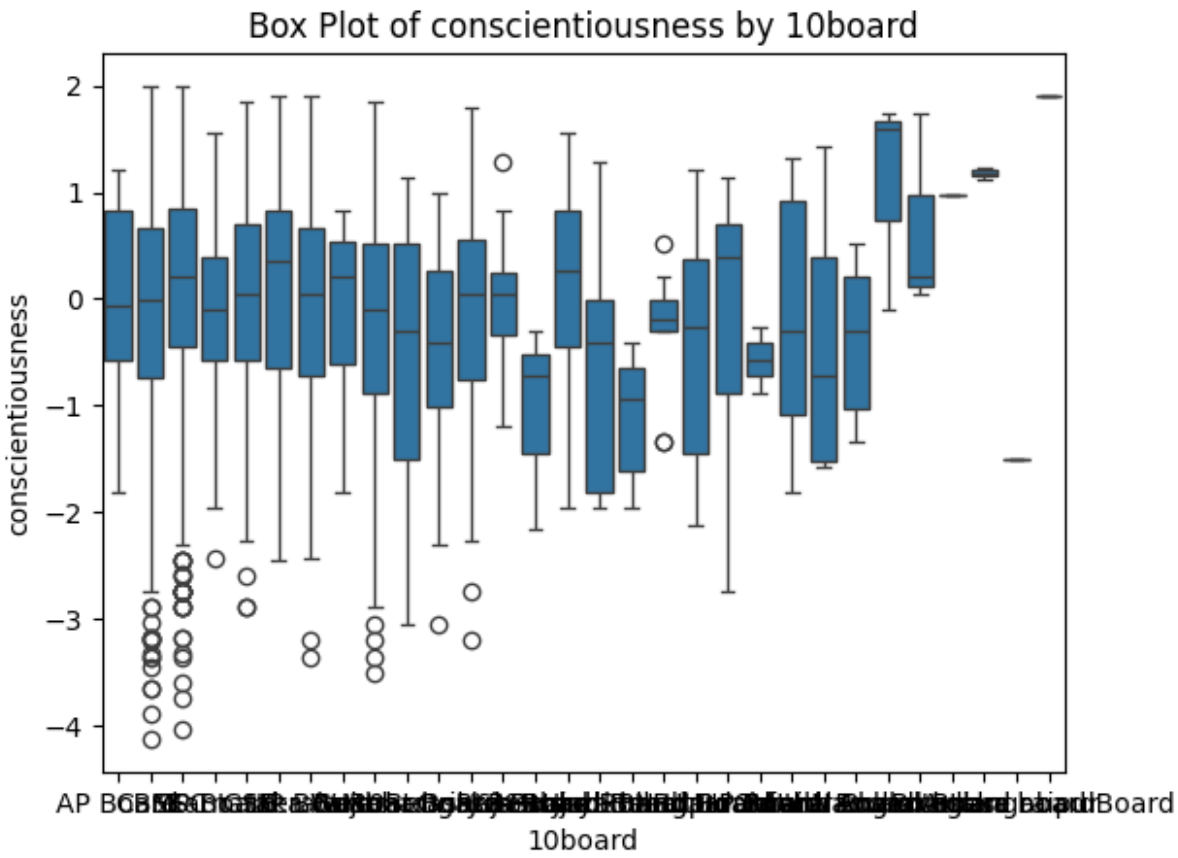

Box Plot of Quant by 10board
Quant
900
800
700
600
500
400
300
200
100
10board

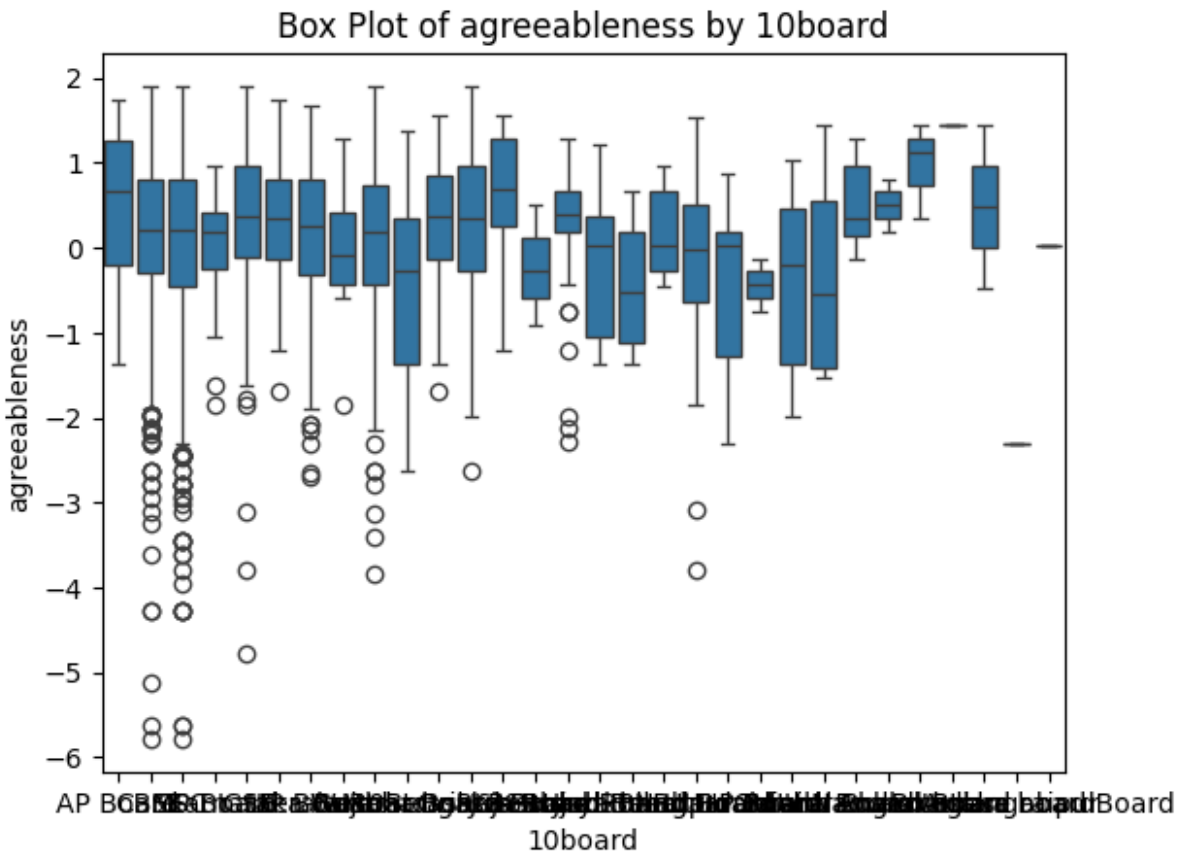

Box Plot of Domain by 10board
Domain
1.0
0.8
0.6
0.4
0.2
0.0
10board

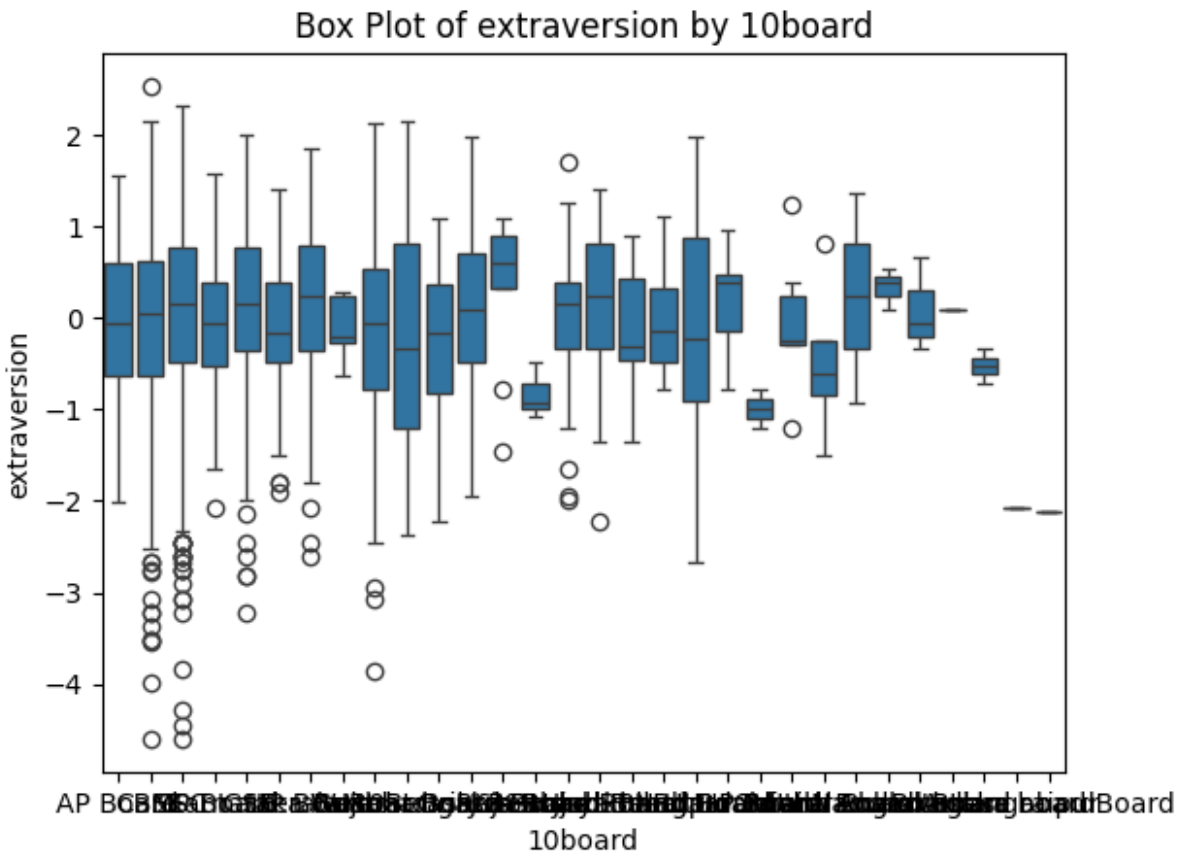

Box Plot of ComputerProgramming by 10board
ComputerProgramming
800
700
600
500
400
300
200
100
10board

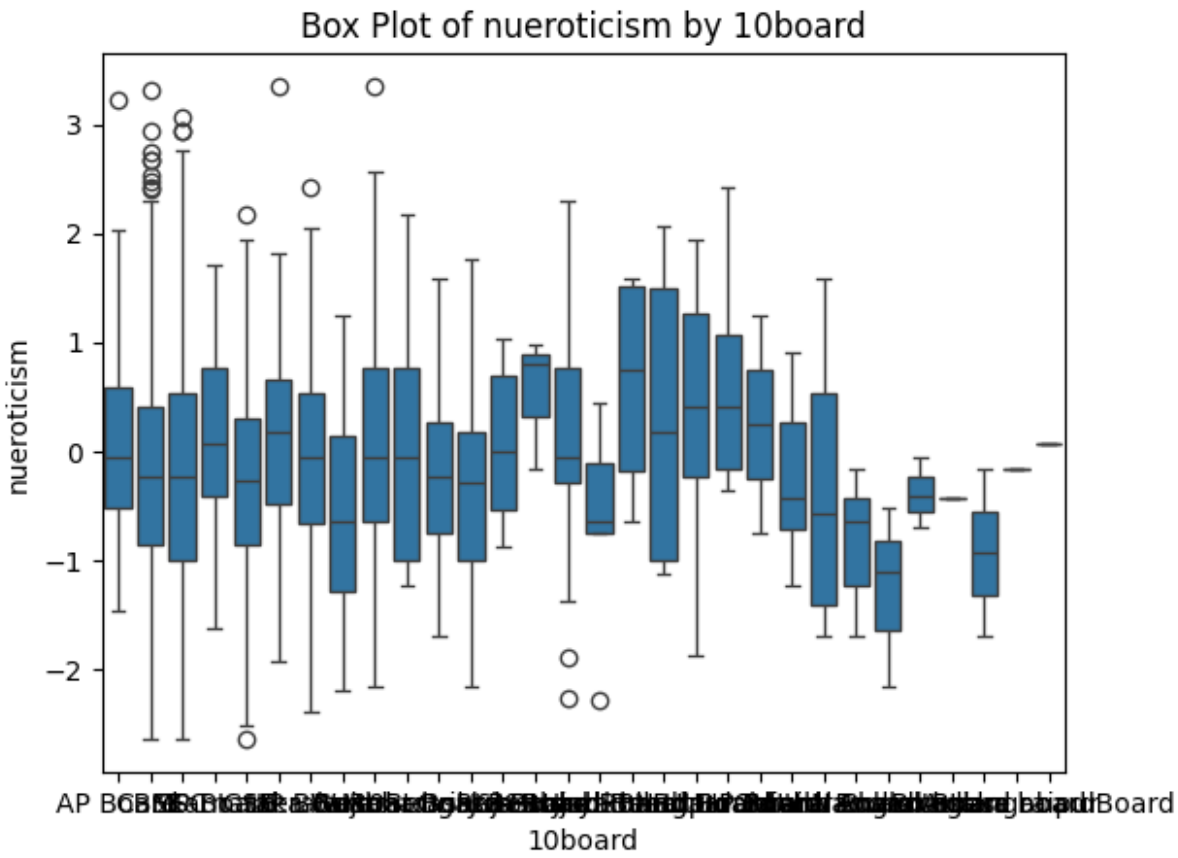

Box Plot of ElectronicsAndSemicon by 10board
ElectronicsAndSemicon
600
500
400
300
200
10board

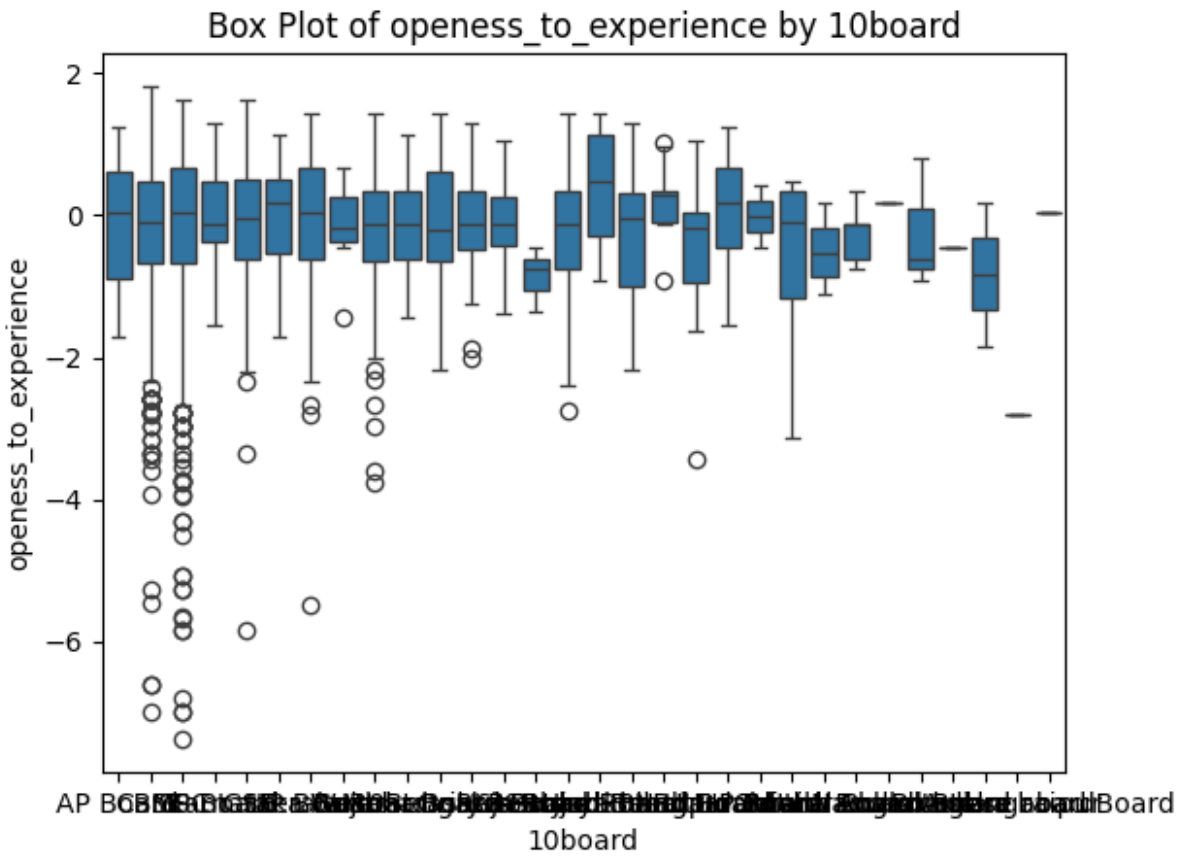

Box Plot of ComputerScience by 10board
ComputerScience
700
600
500
400
300
200
10board

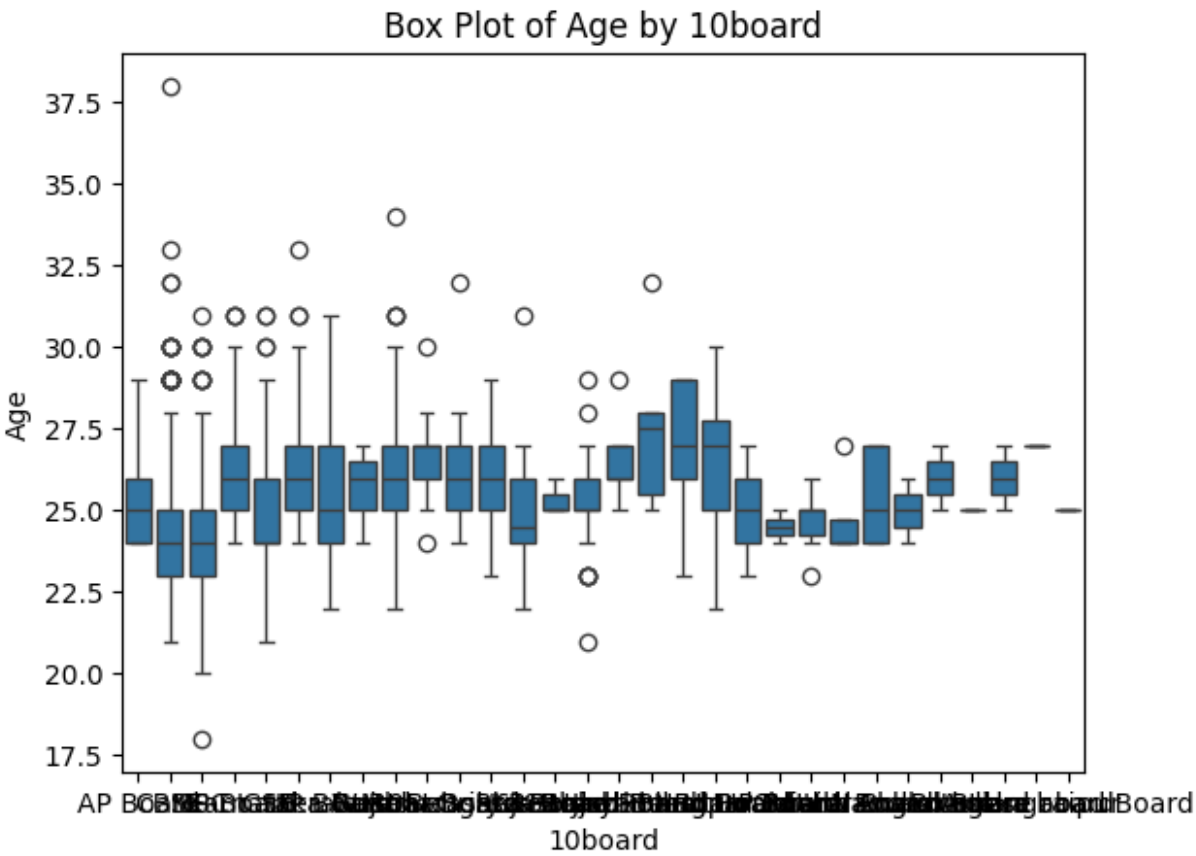

Box Plot of conscientiousness by 10board
conscientiousness
2
1
0
−1
−2
−3
−4
10board

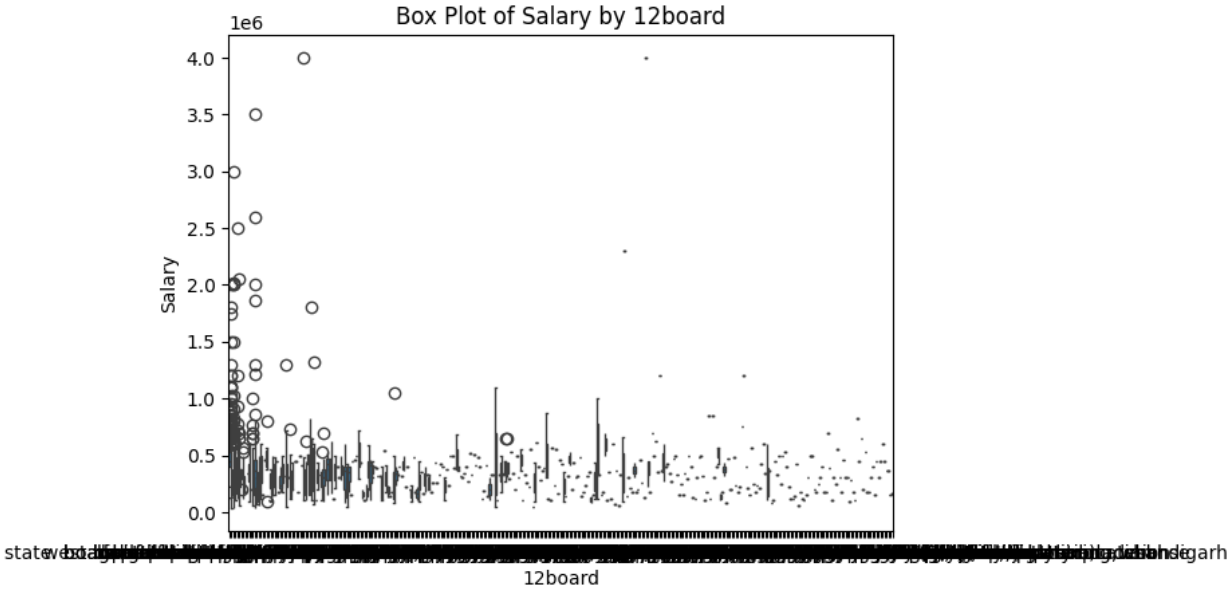

Box Plot of agreeableness by 10board
agreeableness
10board
2
1
0
−1
−2
−3
−4
−5
−6

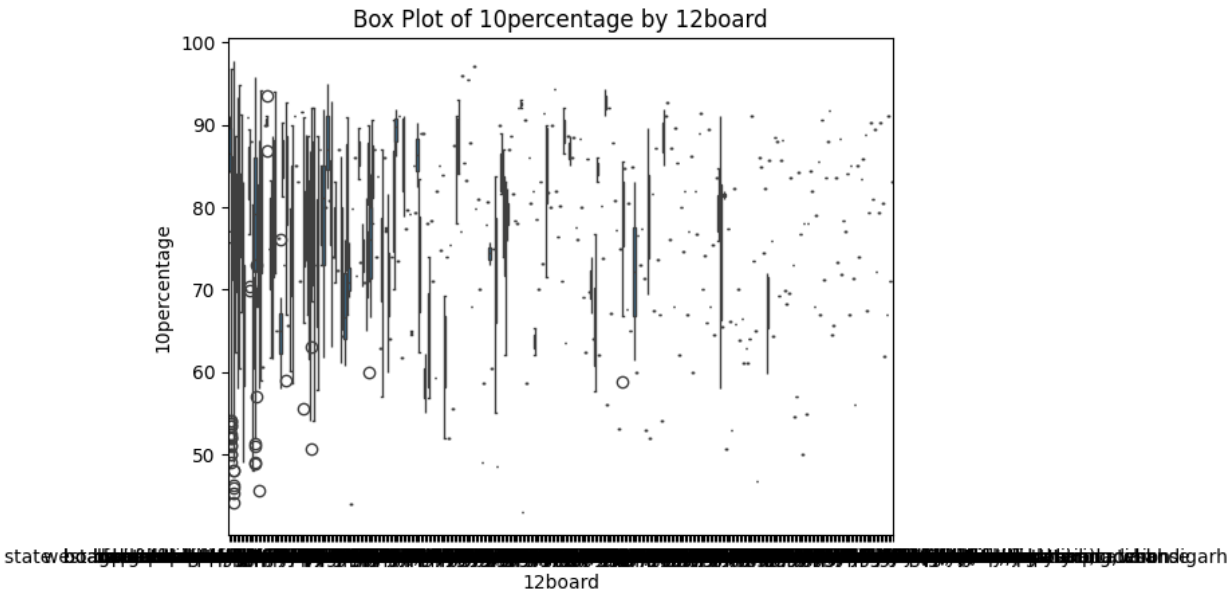

Box Plot of extraversion by 10board
extraversion
2
1
0
−1
−2
−3
−4
10board

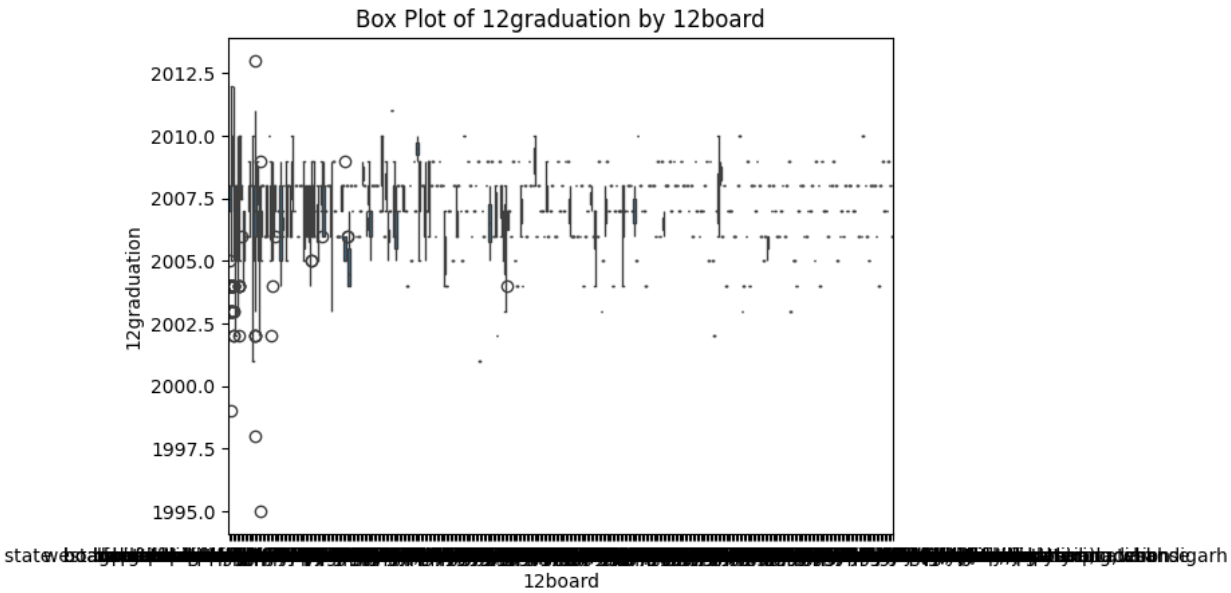

Box Plot of nueroticism by 10board
nueroticism
3
2
1
0
−1
−2
10board

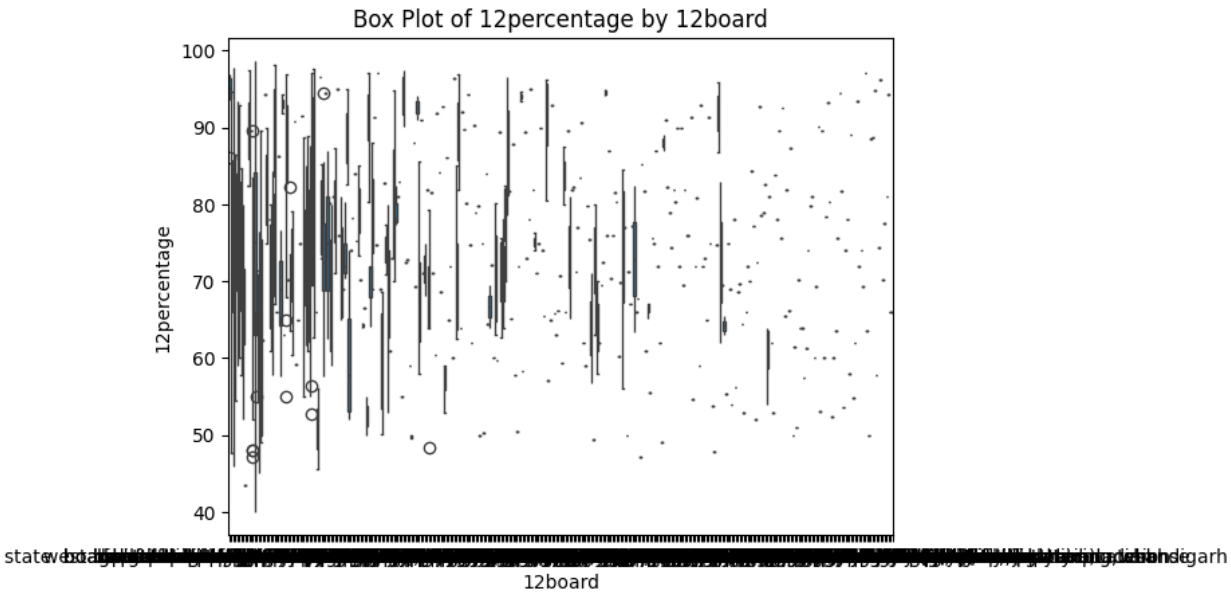

Box Plot of openess_to_experience by 10board
openess_to_experience
2
0
−2
−4
−6
10board

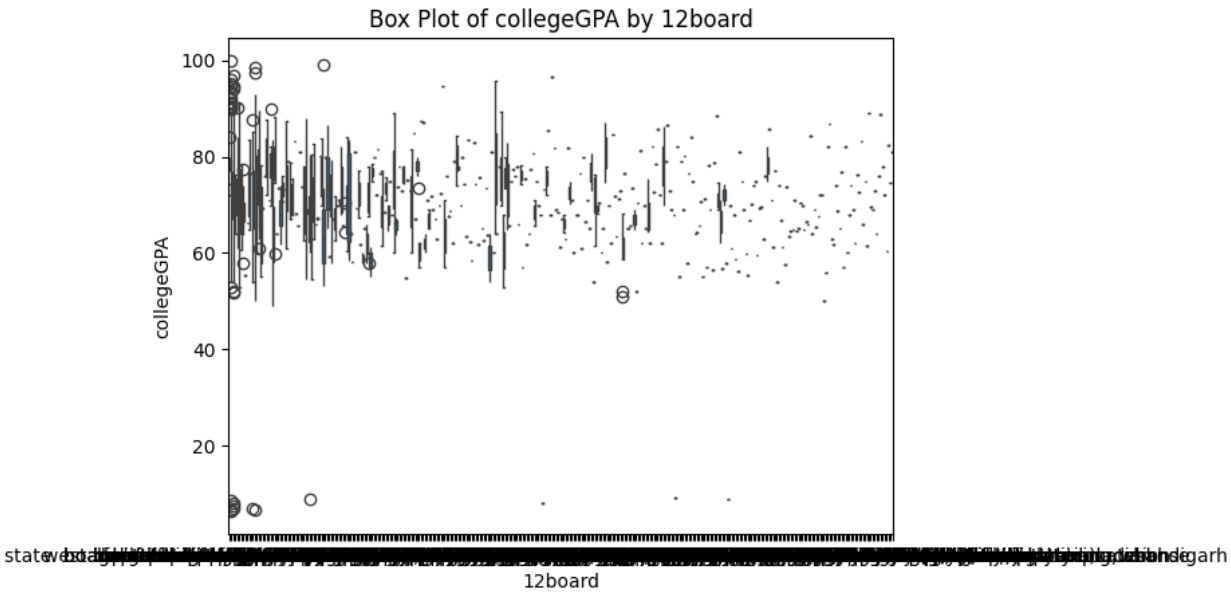

Box Plot of Age by 10board
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
10board

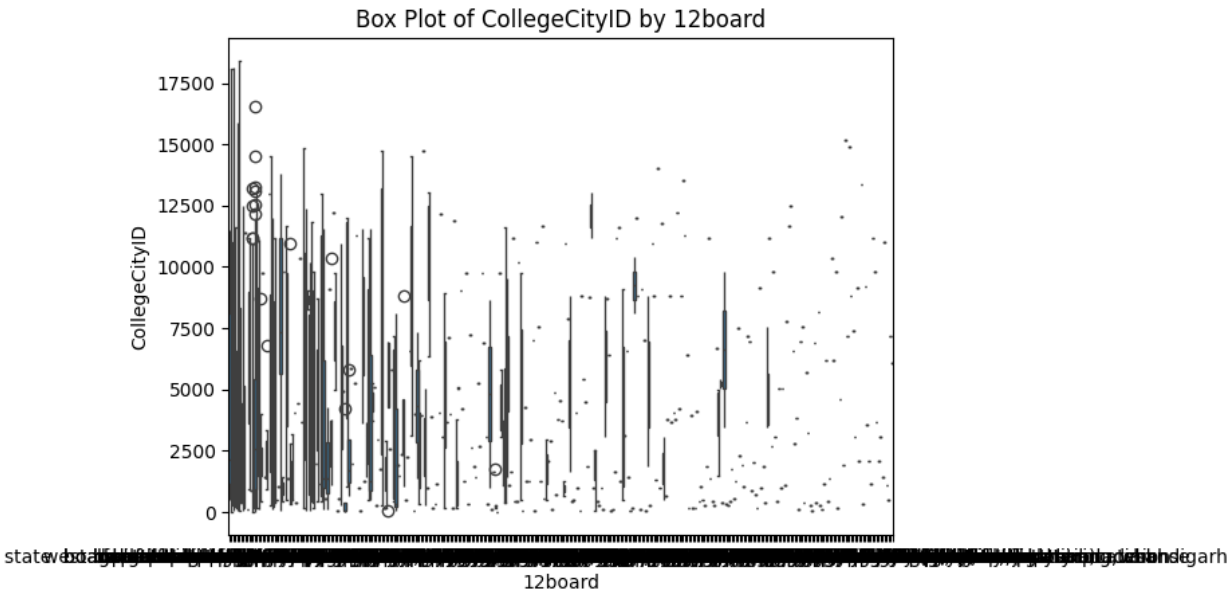

Box Plot of Salary by 12board
1e6
Salary
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
12board

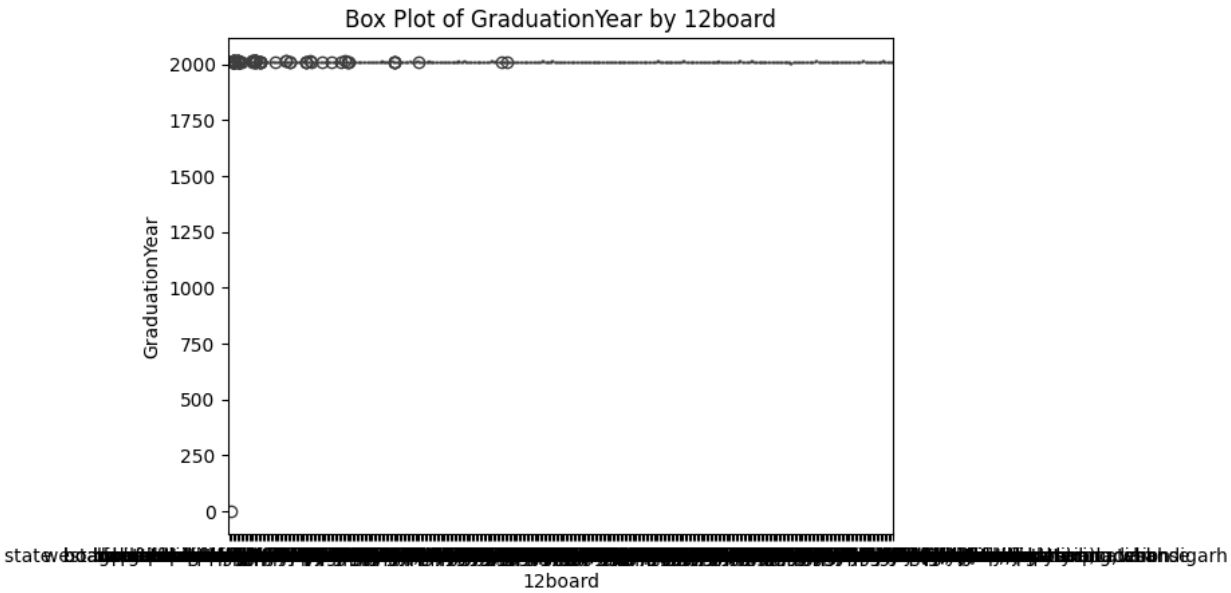

Box Plot of 10percentage by 12board
10percentage
100
90
80
70
60
50
12board

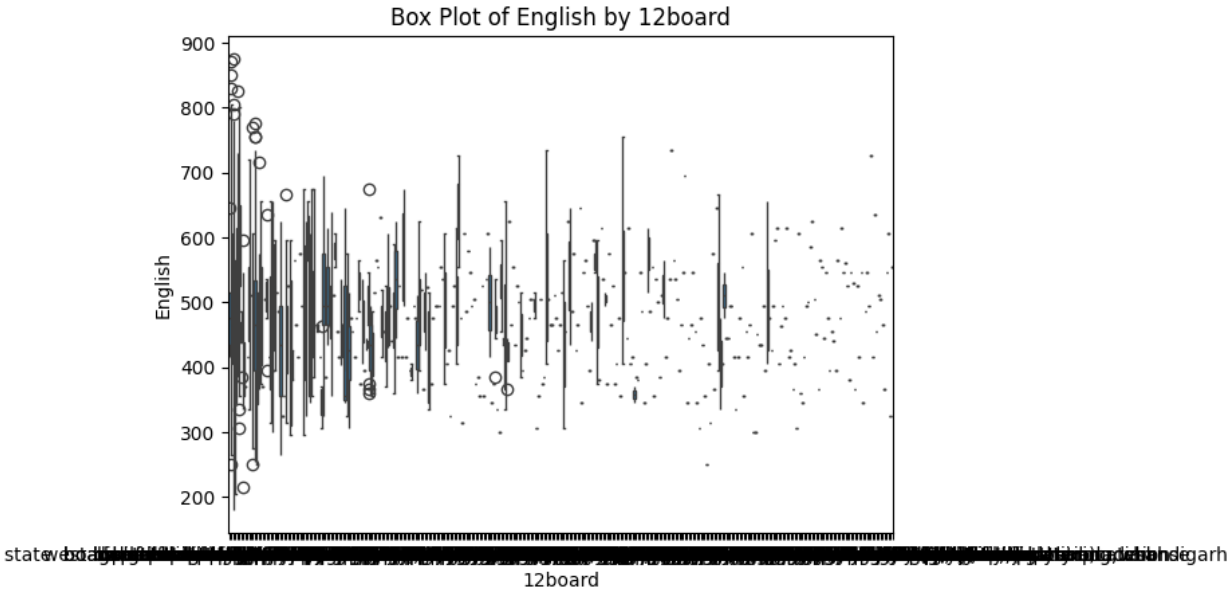

Box Plot of 12graduation by 12board
12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
12board

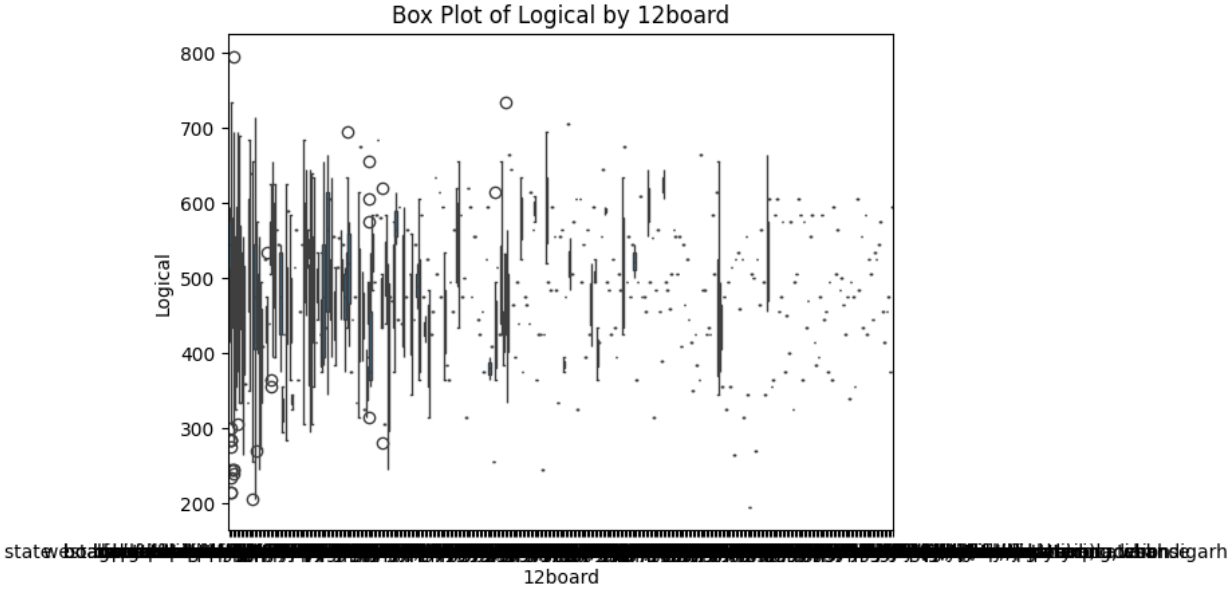

Box Plot of 12percentage by 12board
12percentage
100
90
80
70
60
50
40
12board

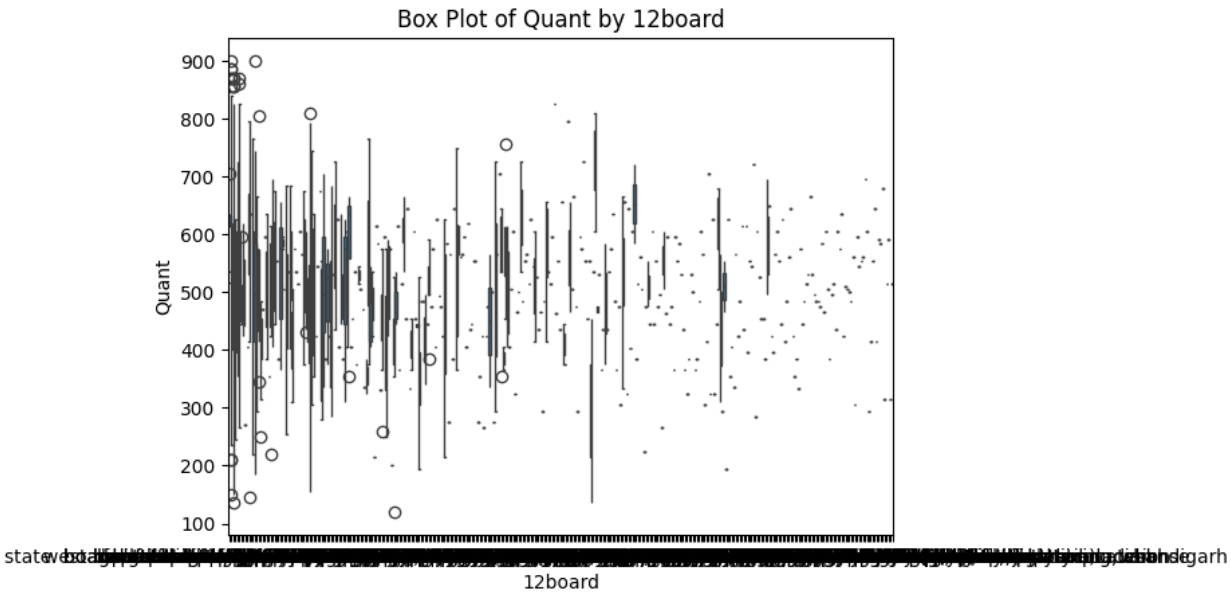

Box Plot of collegeGPA by 12board
collegeGPA
100
80
60
40
20
12board

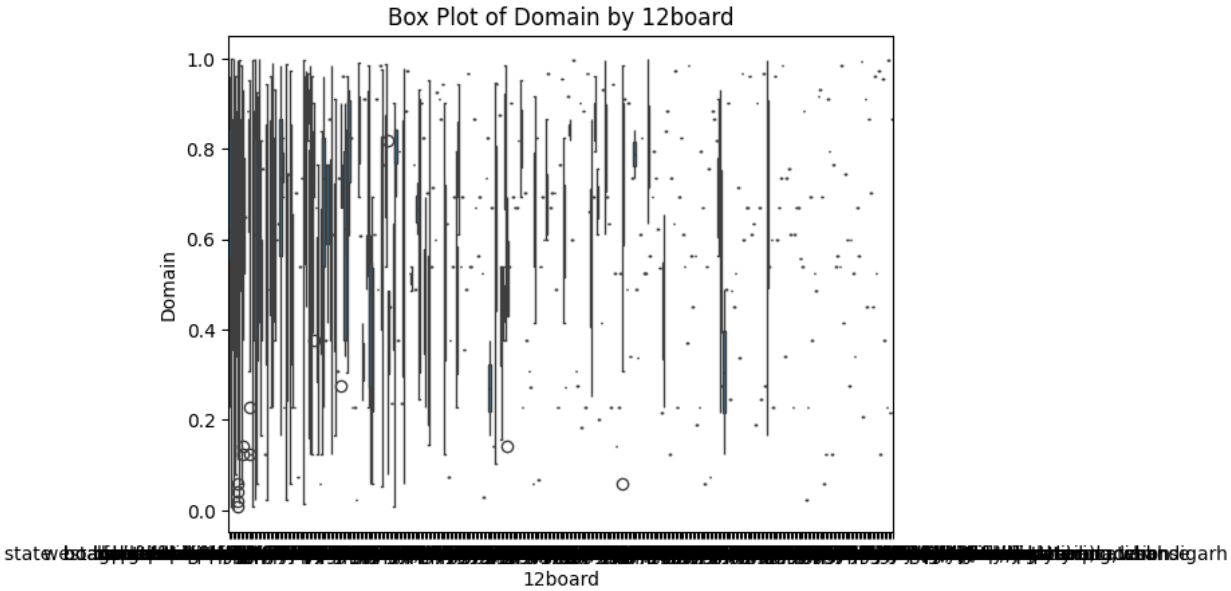

Box Plot of CollegeCityID by 12board
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
12board

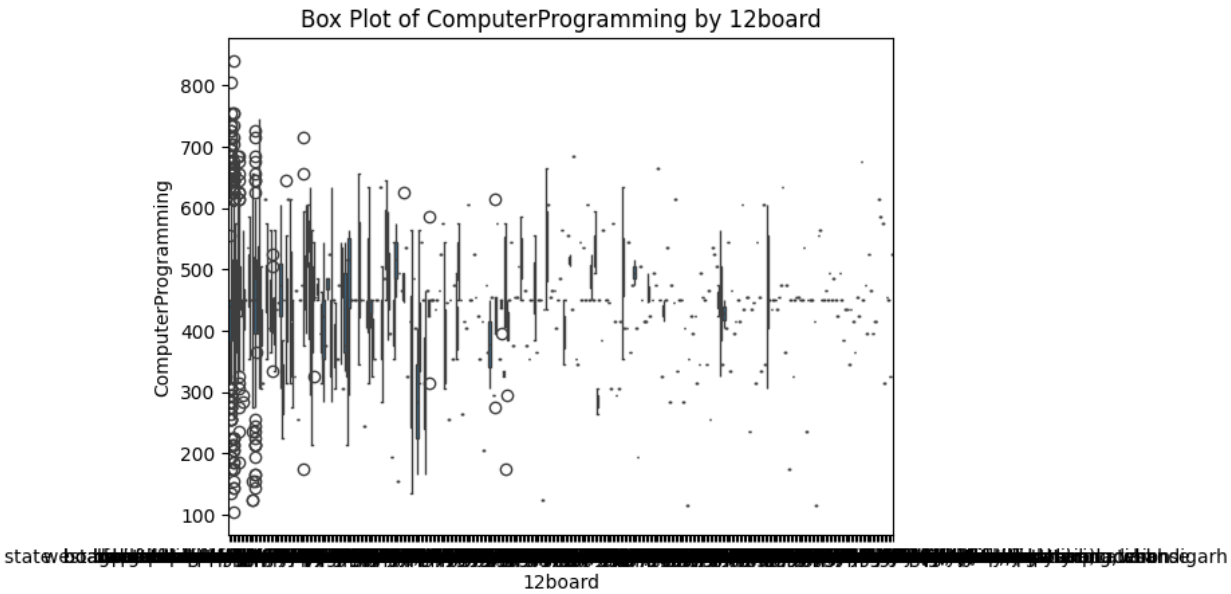

Box Plot of GraduationYear by 12board
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
12board

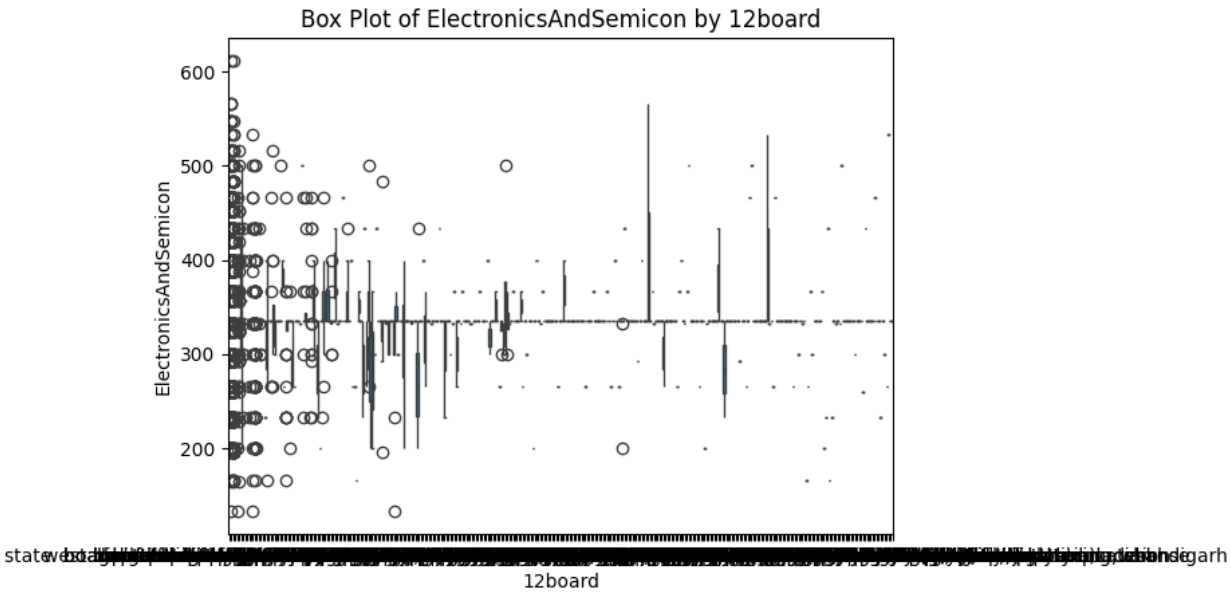

Box Plot of English by 12board
English
900
800
700
600
500
400
300
200
12board

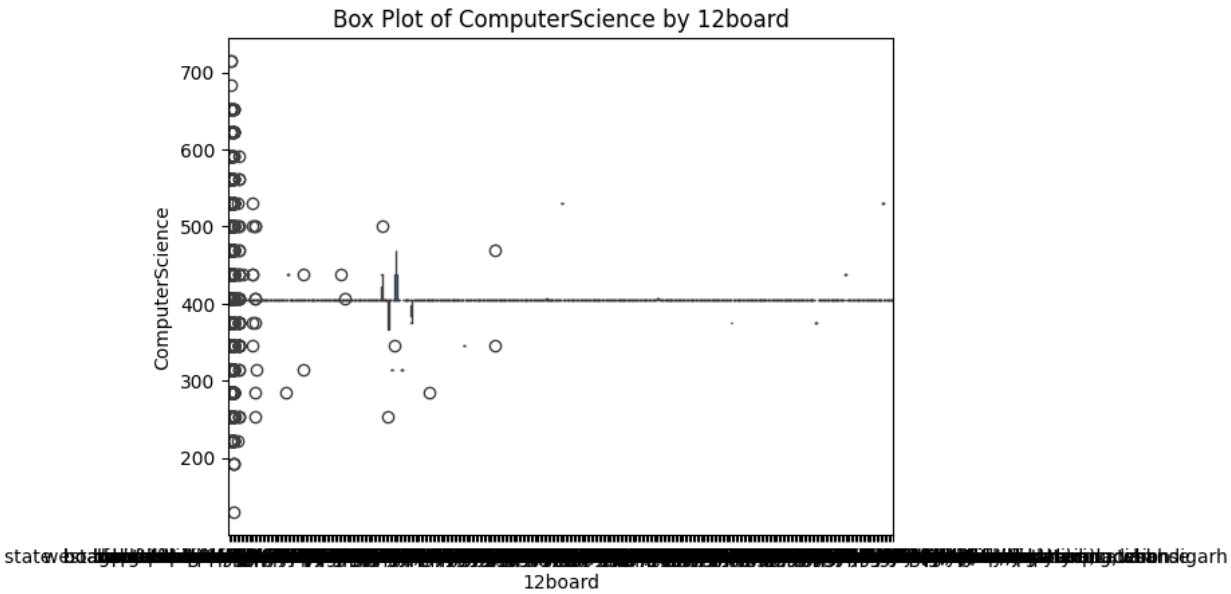

Box Plot of Logical by 12board
Logical
800
700
600
500
400
300
200
12board

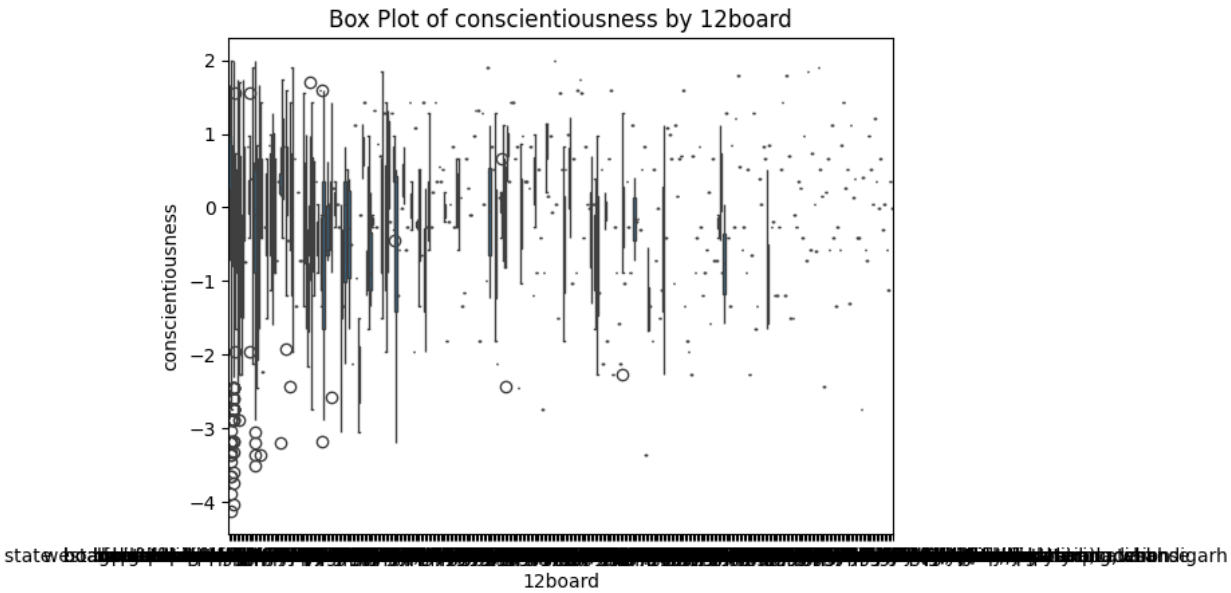

Box Plot of Quant by 12board
Quant
900
800
700
600
500
400
300
200
100
12board

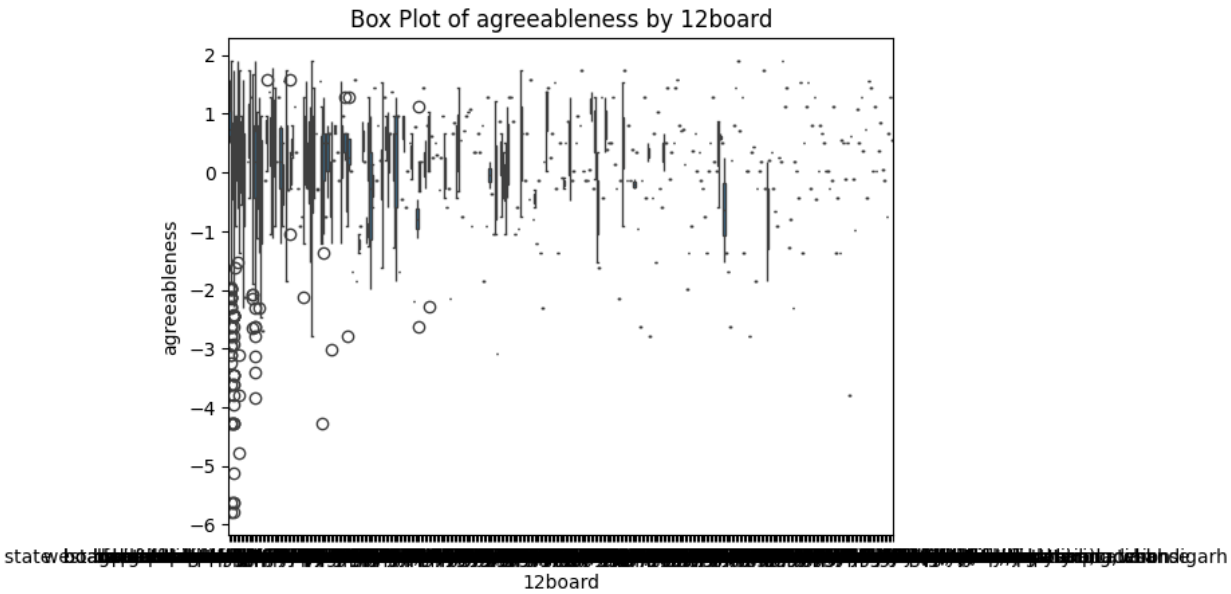

Box Plot of Domain by 12board
Domain
1.0
0.8
0.6
0.4
0.2
0.0
12board

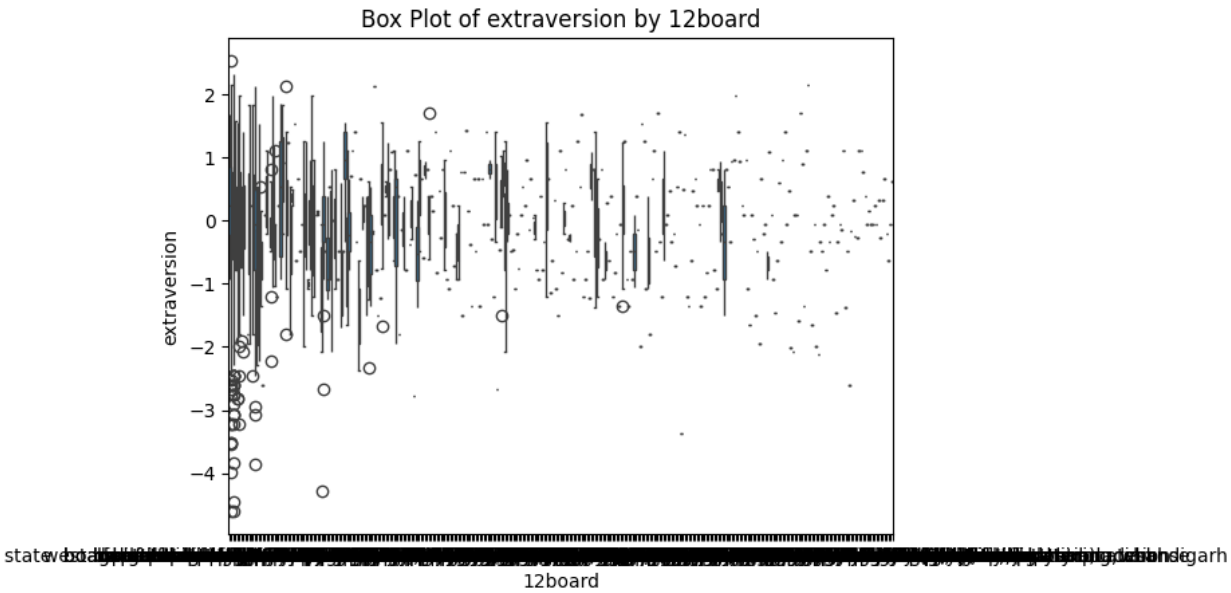

Box Plot of ComputerProgramming by 12board
ComputerProgramming
800
700
600
500
400
300
200
100
12board

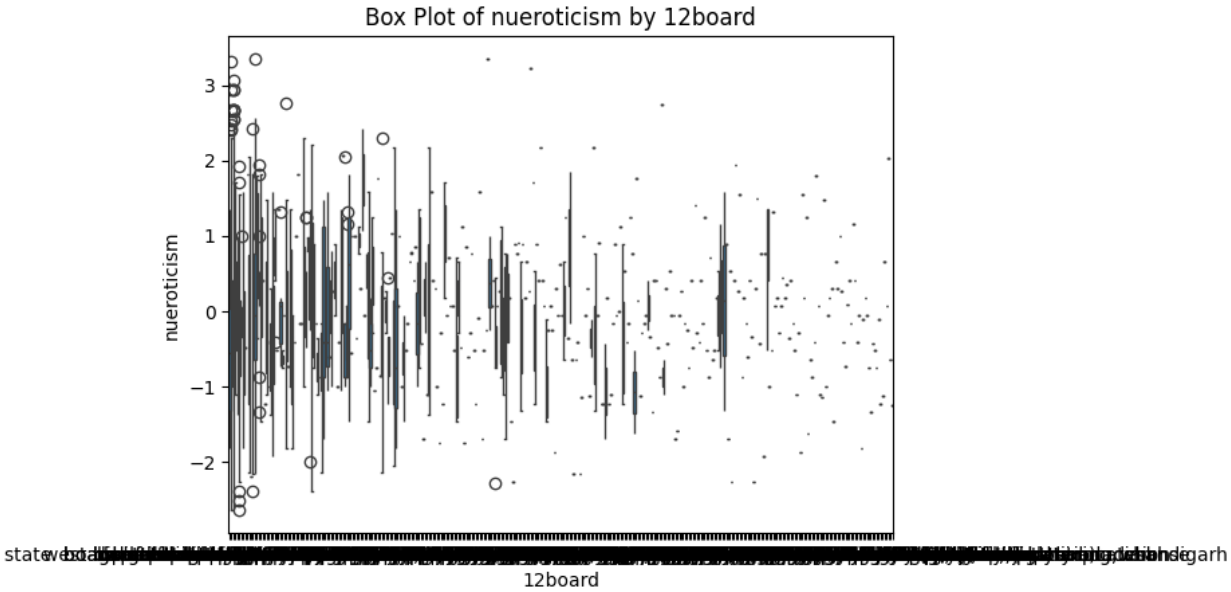

Box Plot of ElectronicsAndSemicon by 12board
ElectronicsAndSemicon
600
500
400
300
200
12board

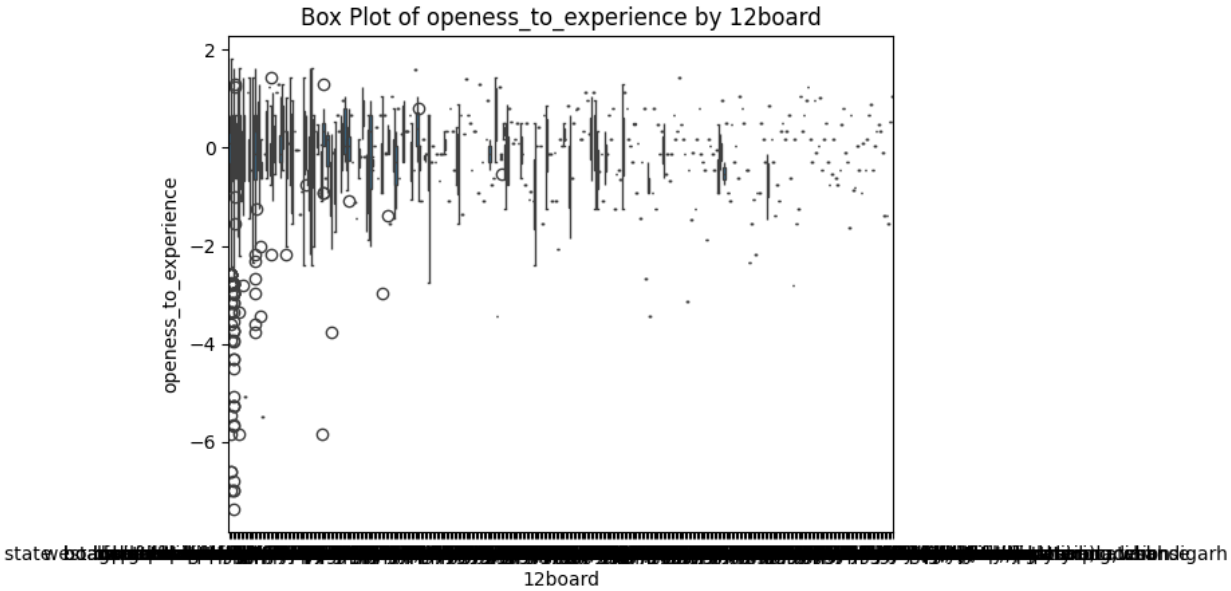

Box Plot of ComputerScience by 12board
ComputerScience
700
600
500
400
300
200
12board

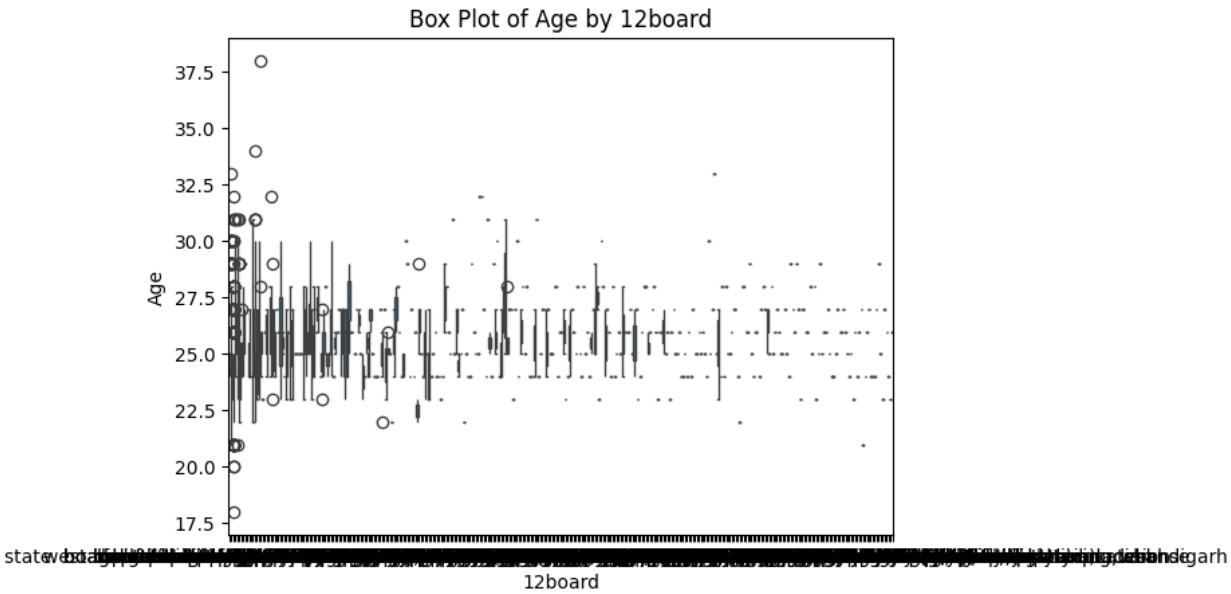

Box Plot of conscientiousness by 12board
conscientiousness
2
1
0
−1
−2
−3
−4
12board

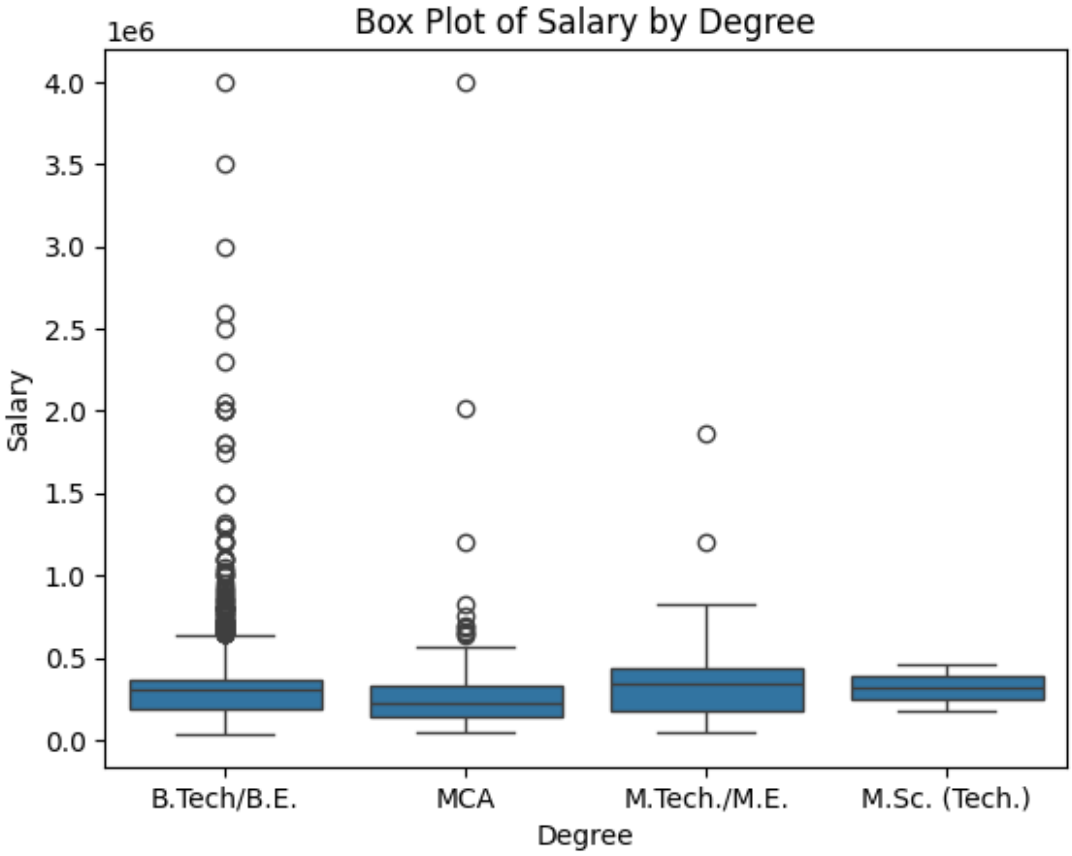

Box Plot of agreeableness by 12board
agreeableness
2
1
0
−1
−2
−3
−4
−5
−6
12board

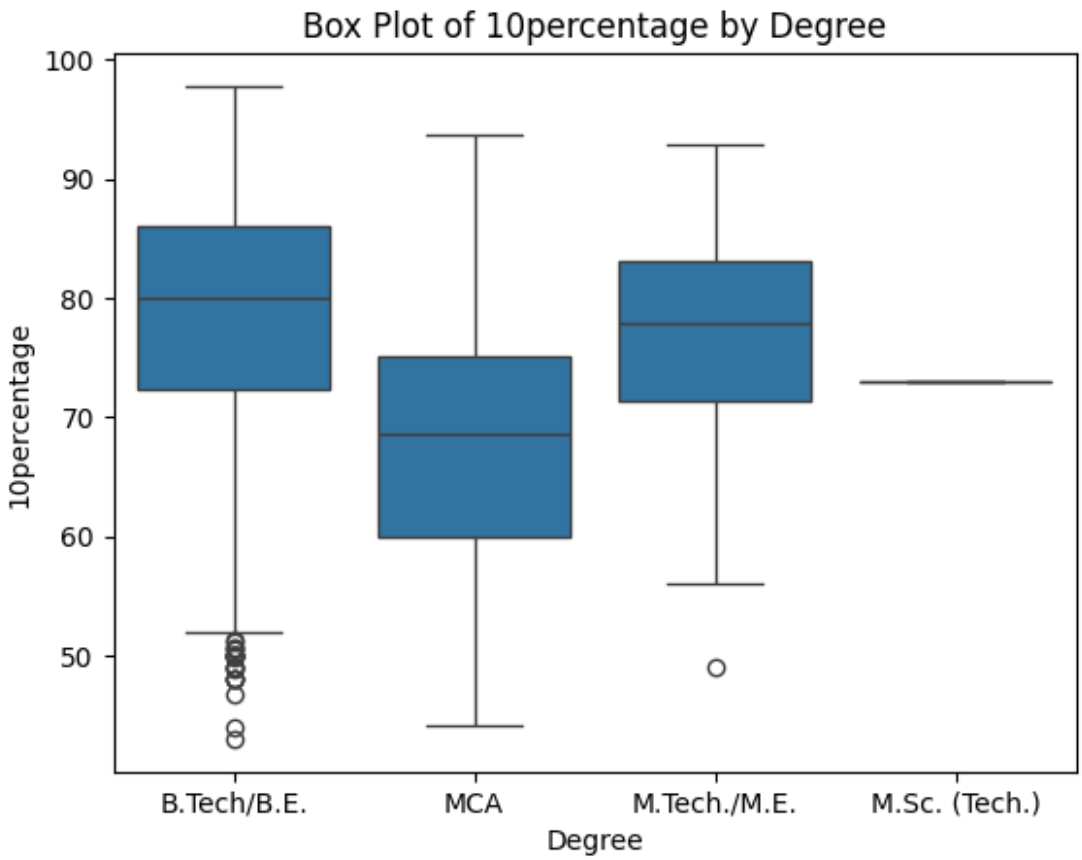

Box Plot of extraversion by 12board
extraversion
2
1
0
−1
−2
−3
−4
12board

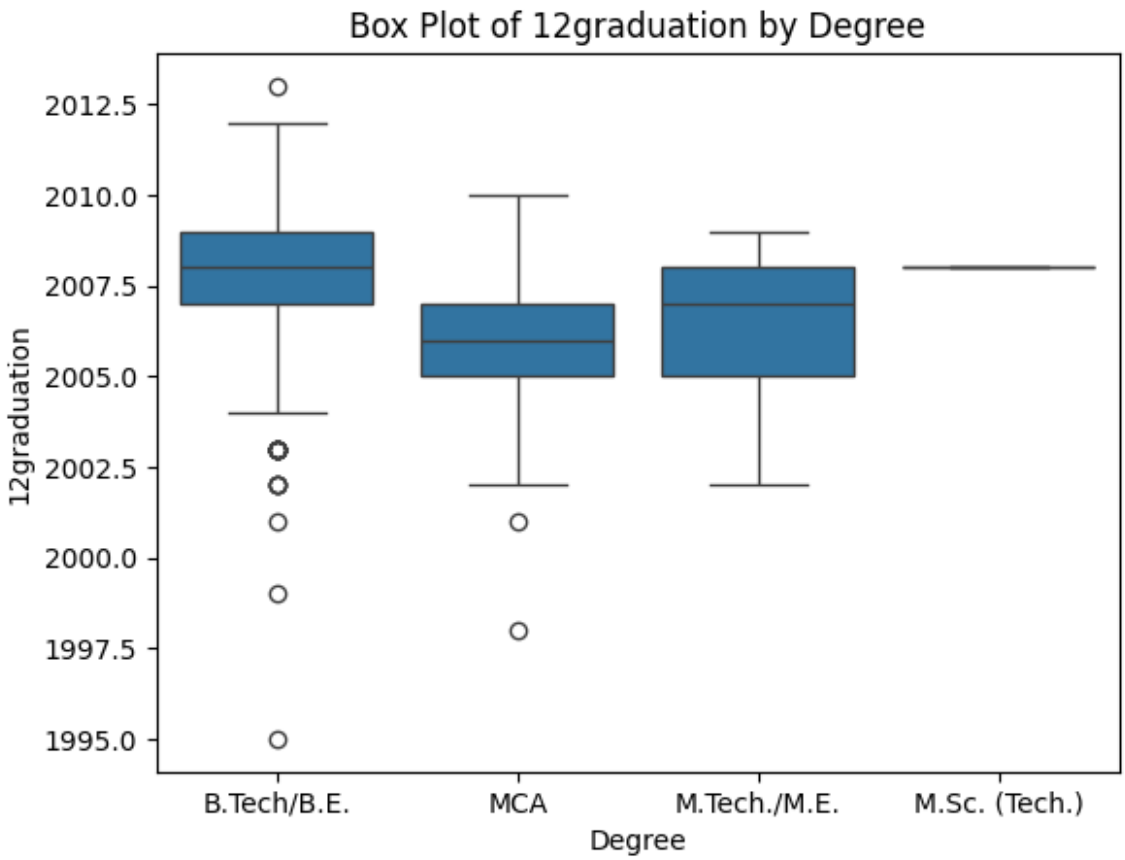

Box Plot of nueroticism by 12board
nueroticism
3
2
1
0
−1
−2
12board

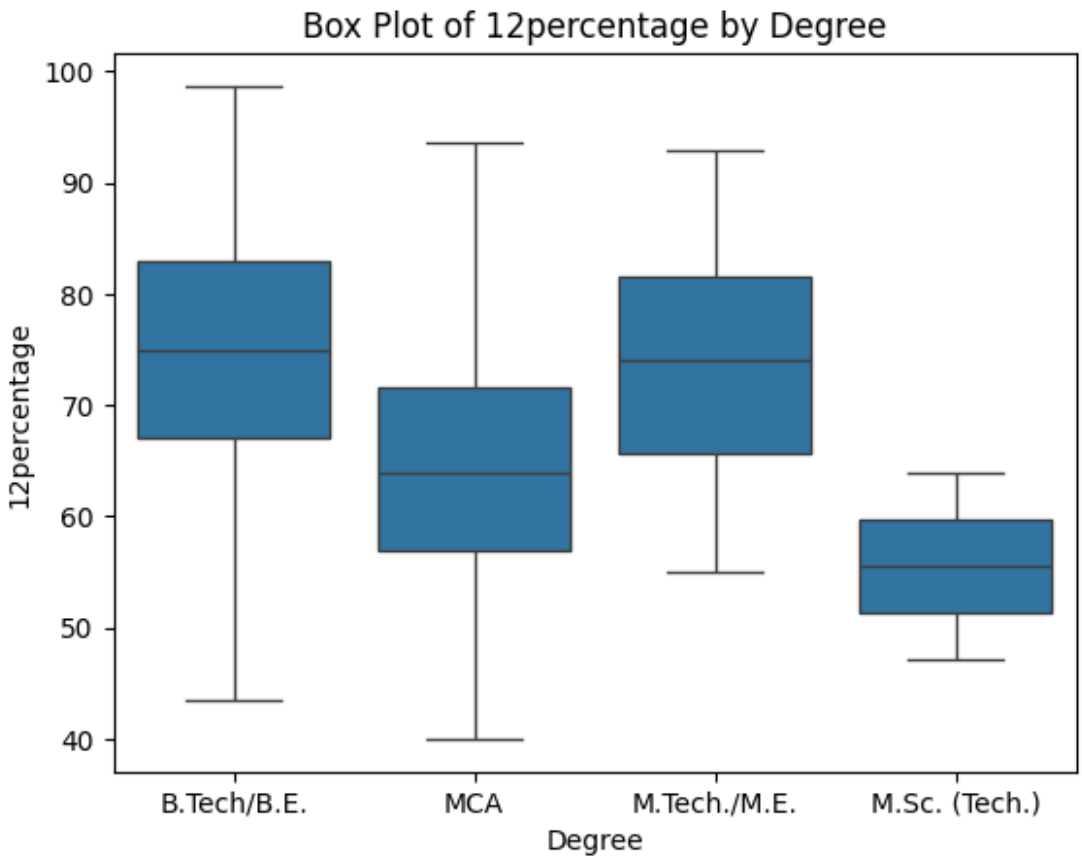

Box Plot of openess_to_experience by 12board
openess_to_experience
2
0
−2
−4
−6
12board

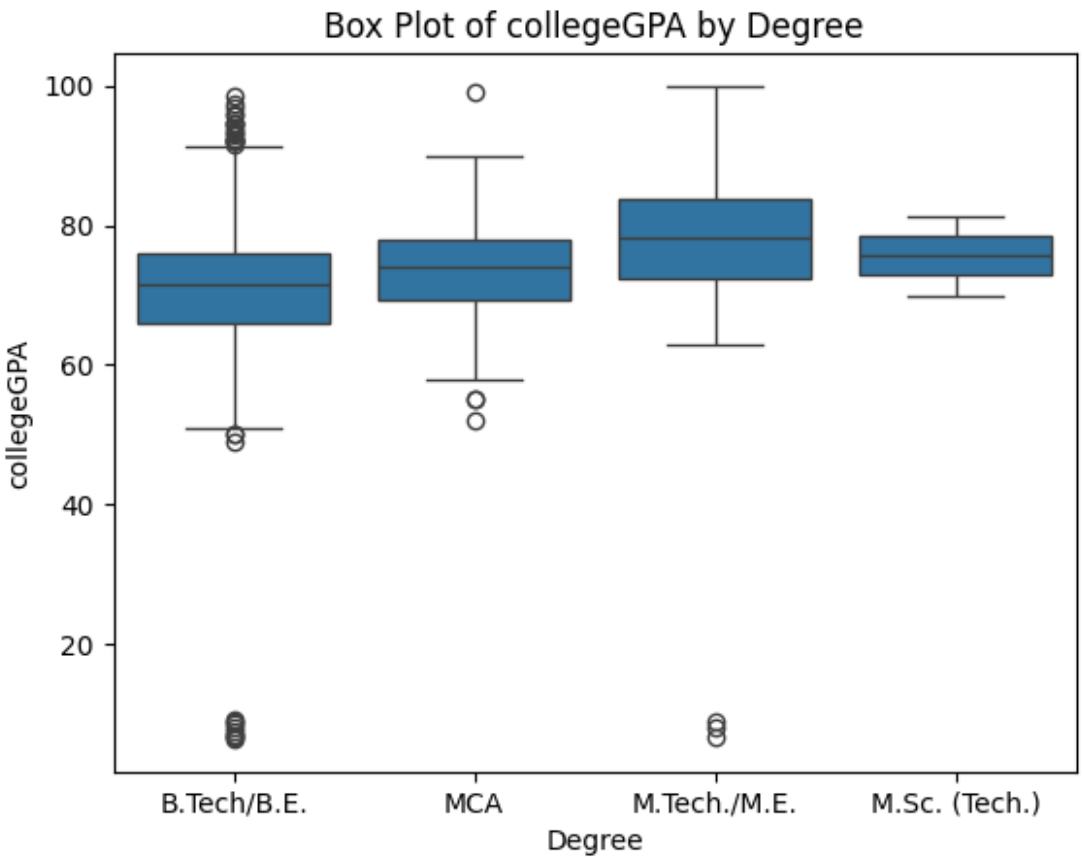

Box Plot of Age by 12board
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
12board

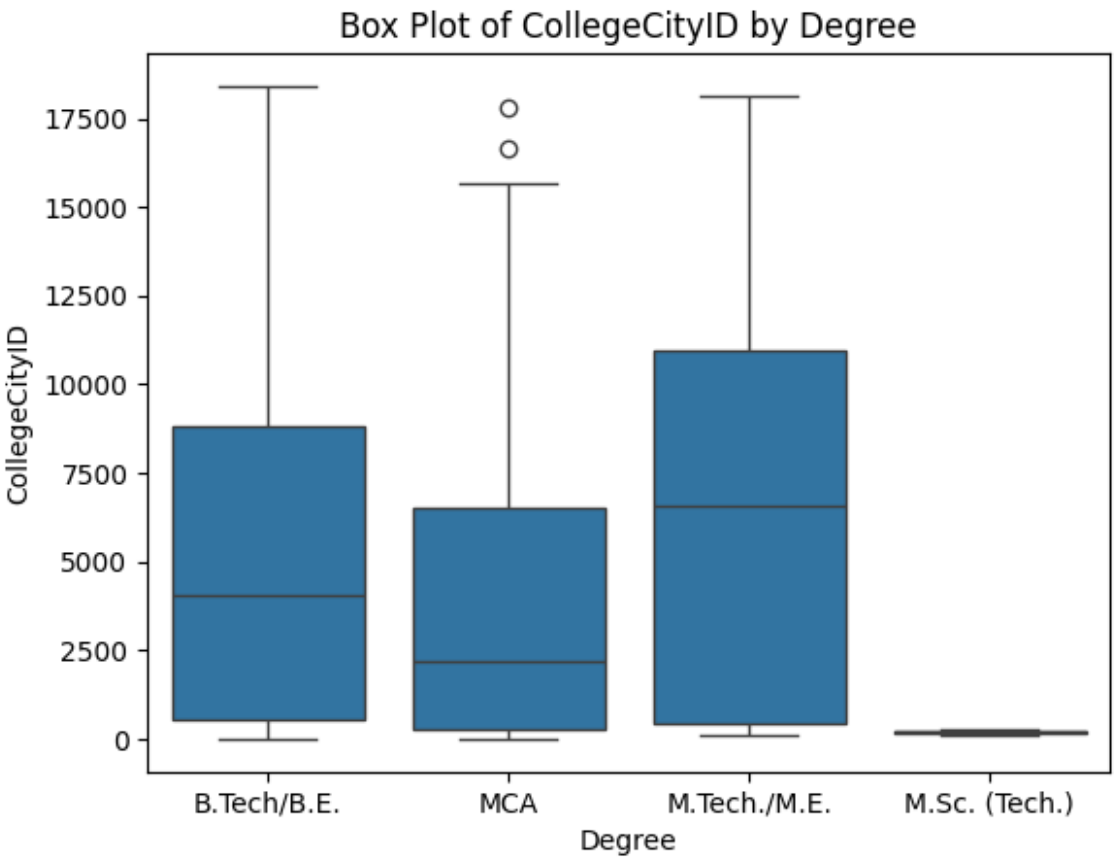

Box Plot of Salary by Degree
1e6
Salary
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

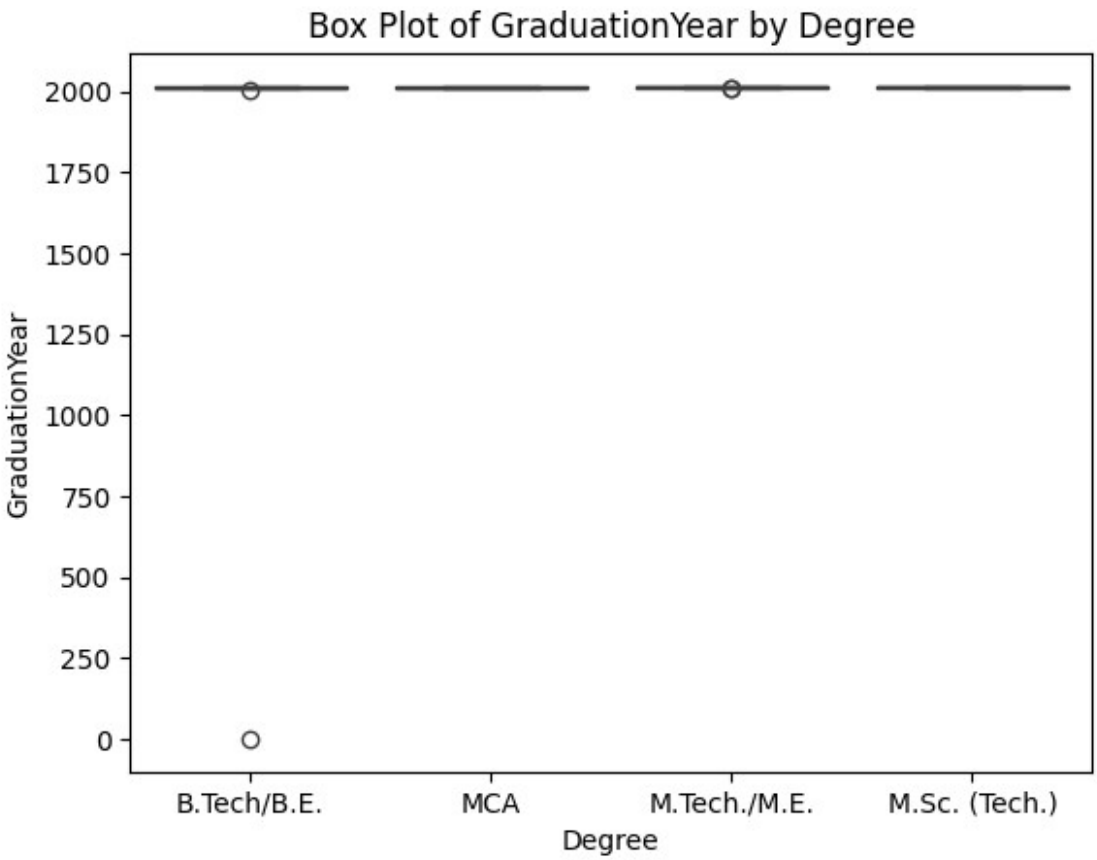

Box Plot of 10percentage by Degree
100
90
80
70
60
50
10percentage
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

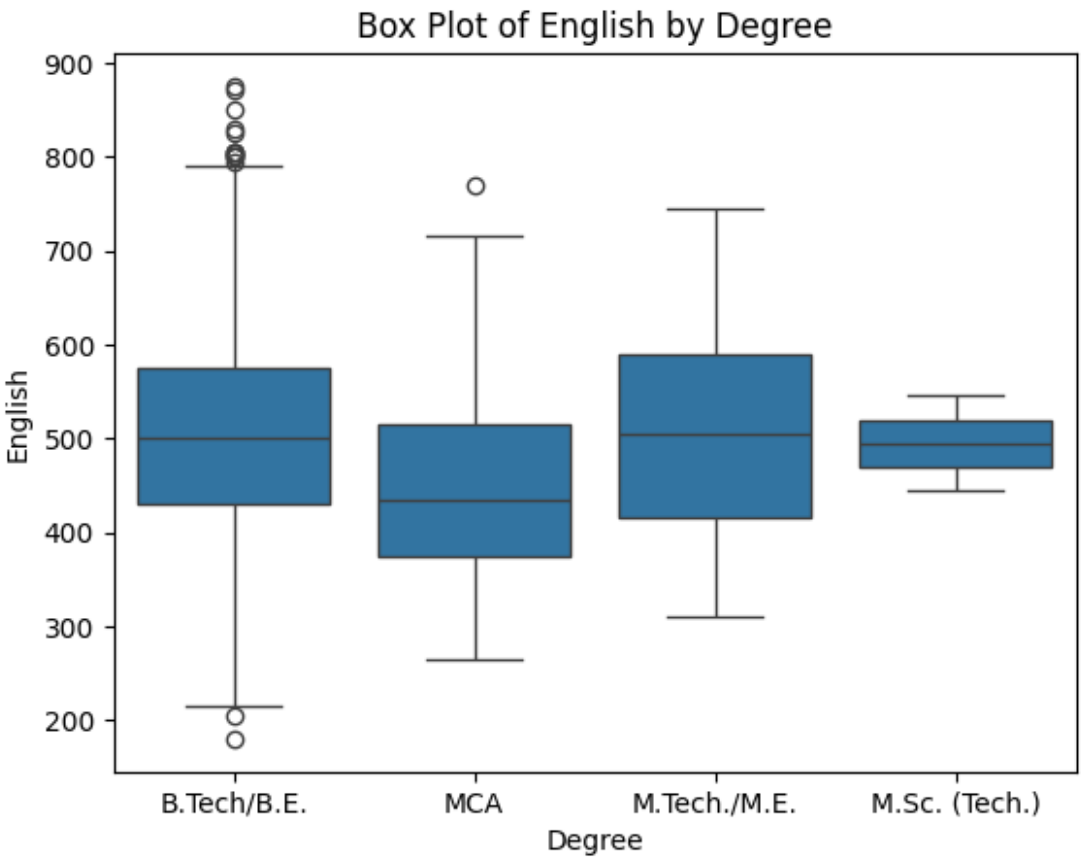

Box Plot of 12graduation by Degree
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
12graduation
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

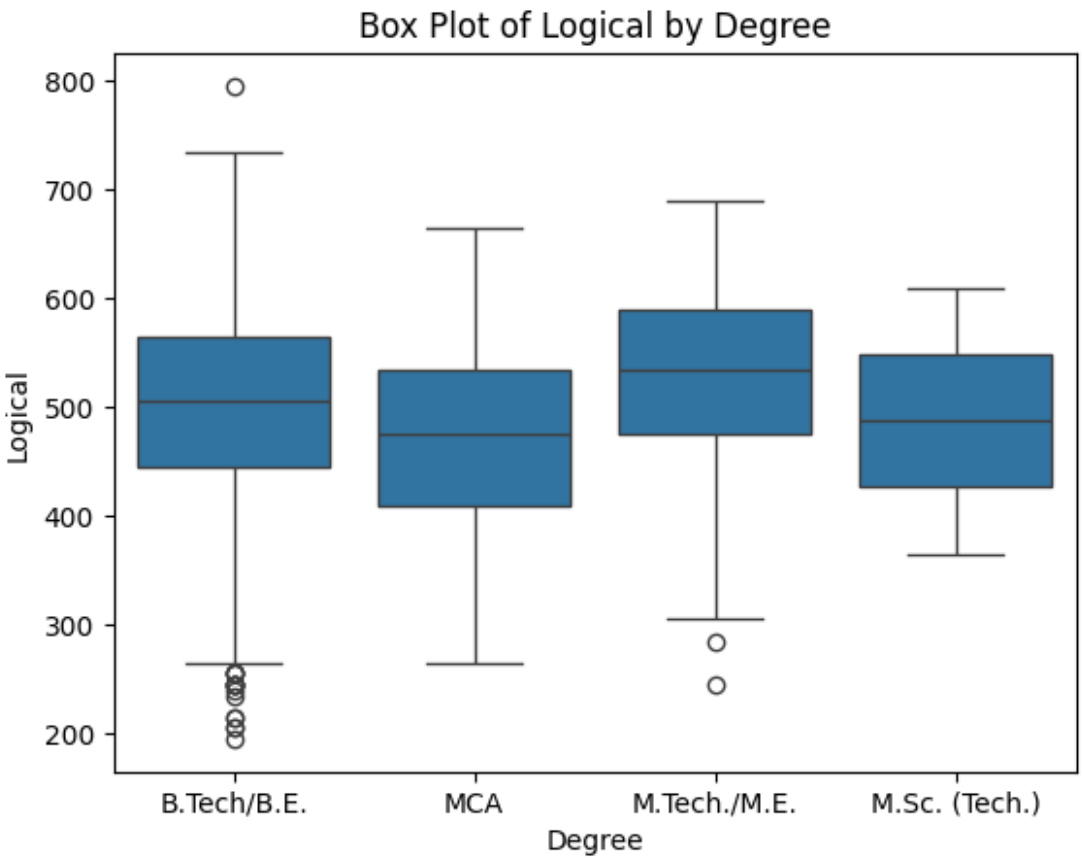

Box Plot of 12percentage by Degree
12percentage
100
90
80
70
60
50
40
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

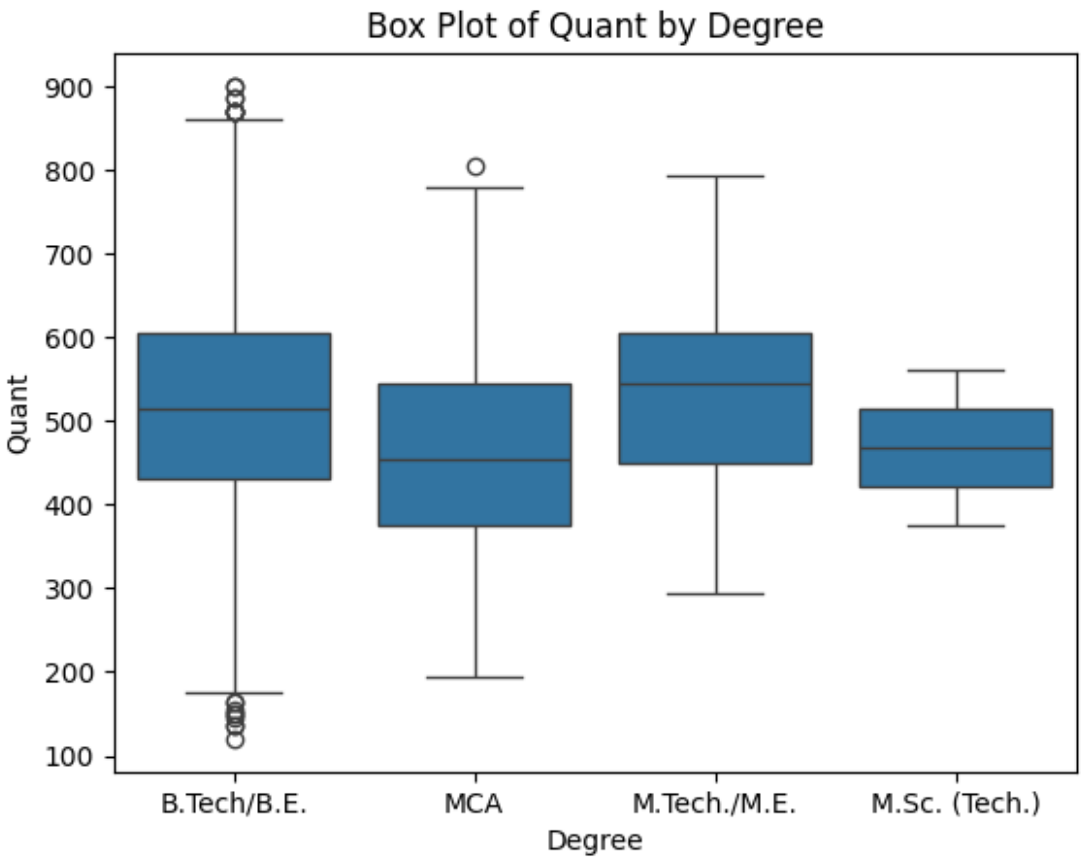

Box Plot of collegeGPA by Degree
100
80
60
40
20
collegeGPA
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

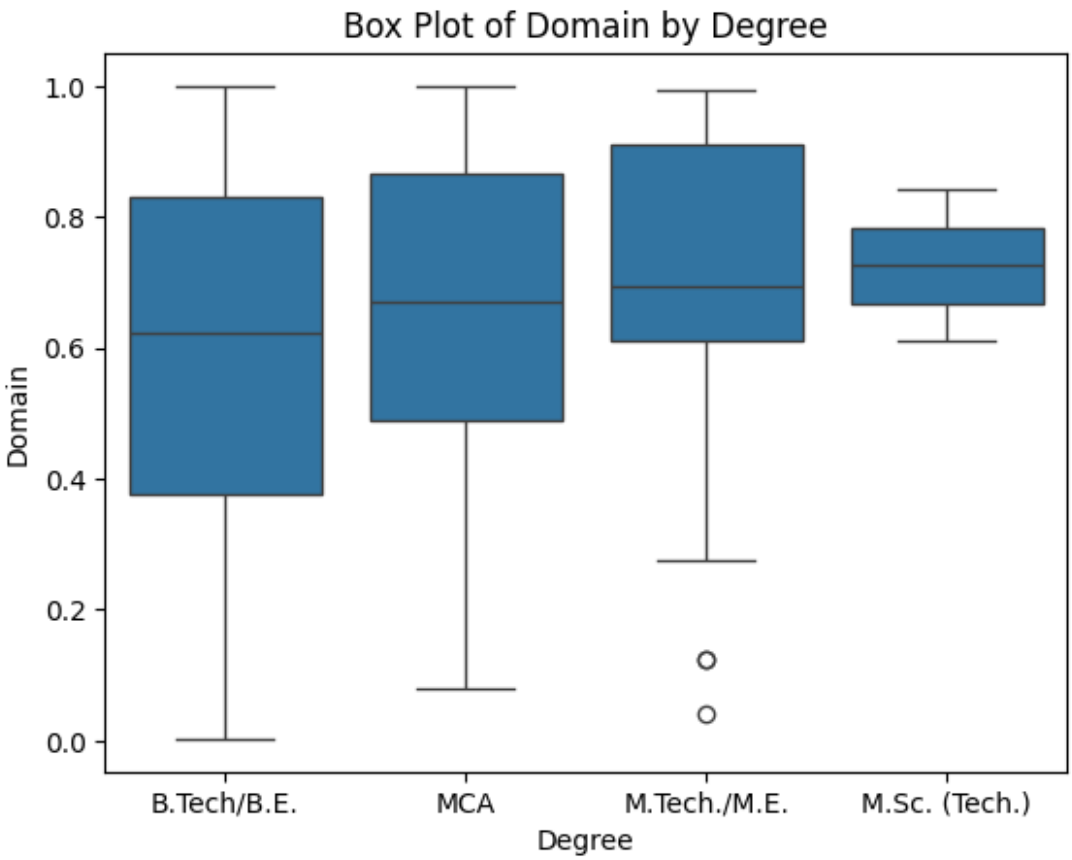

Box Plot of CollegeCityID by Degree
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

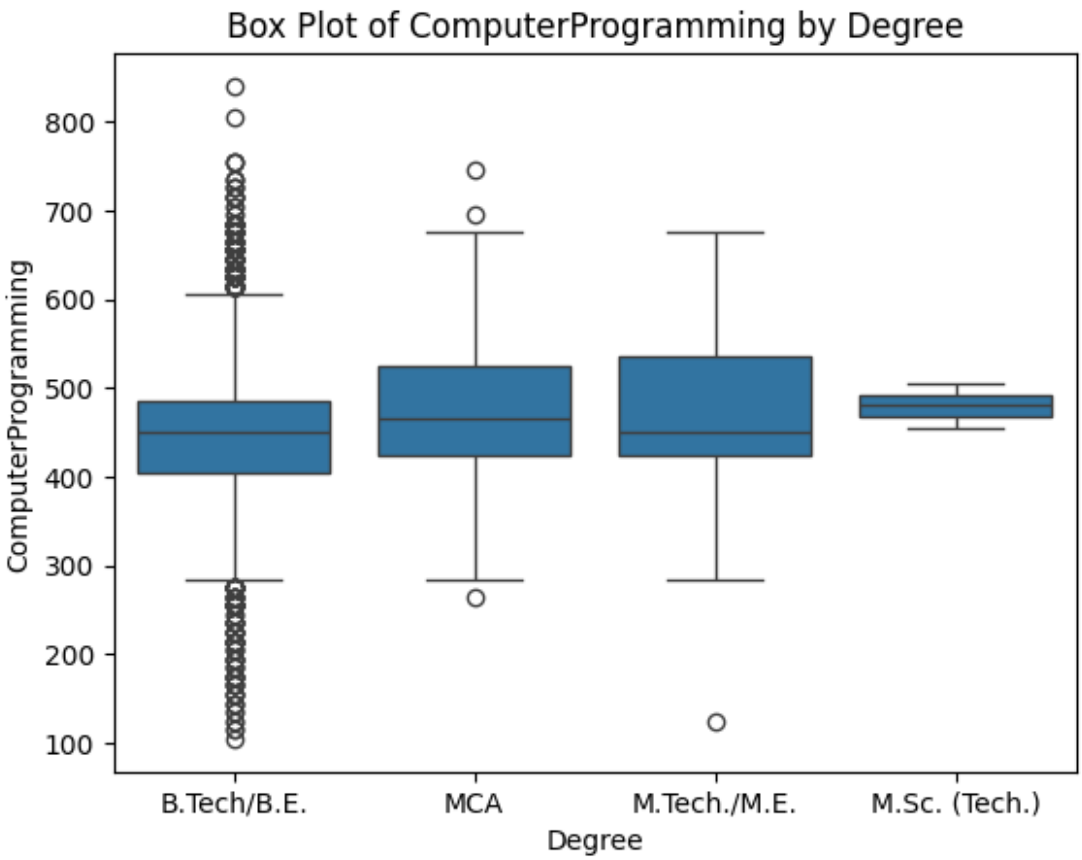

Box Plot of GraduationYear by Degree
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

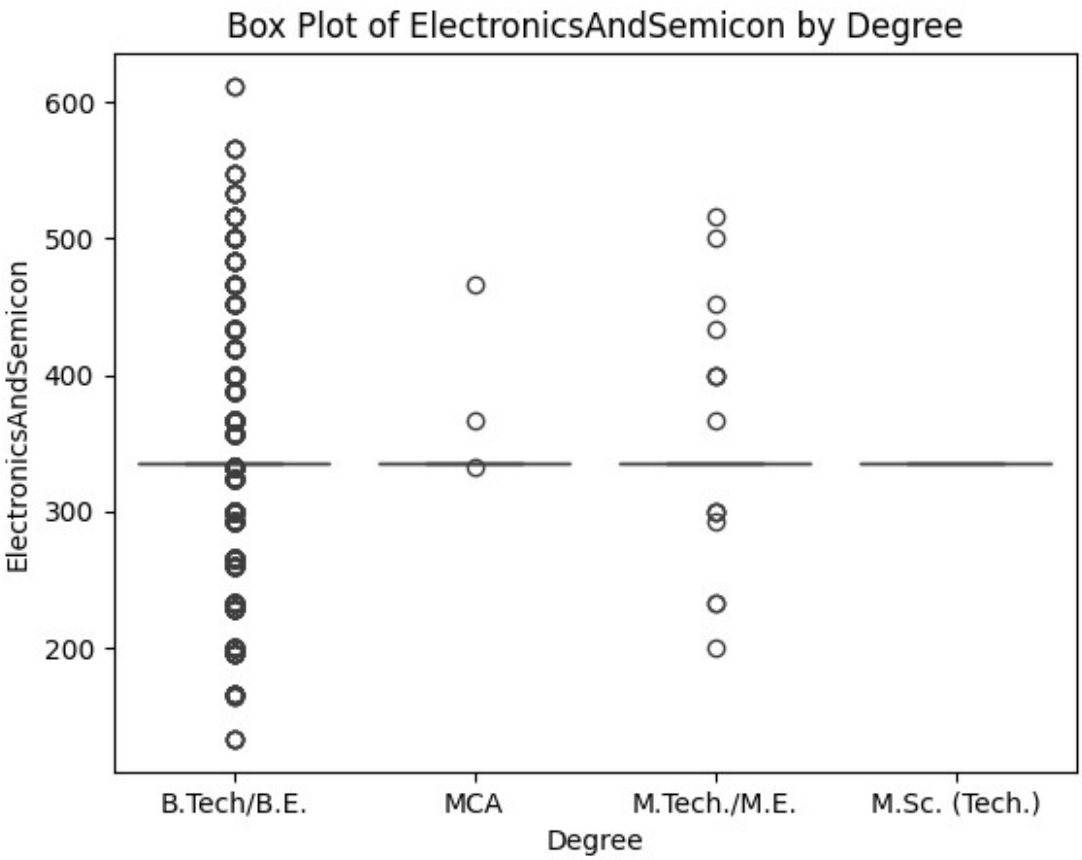

Box Plot of English by Degree
900
800
700
600
500
400
300
200
English
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

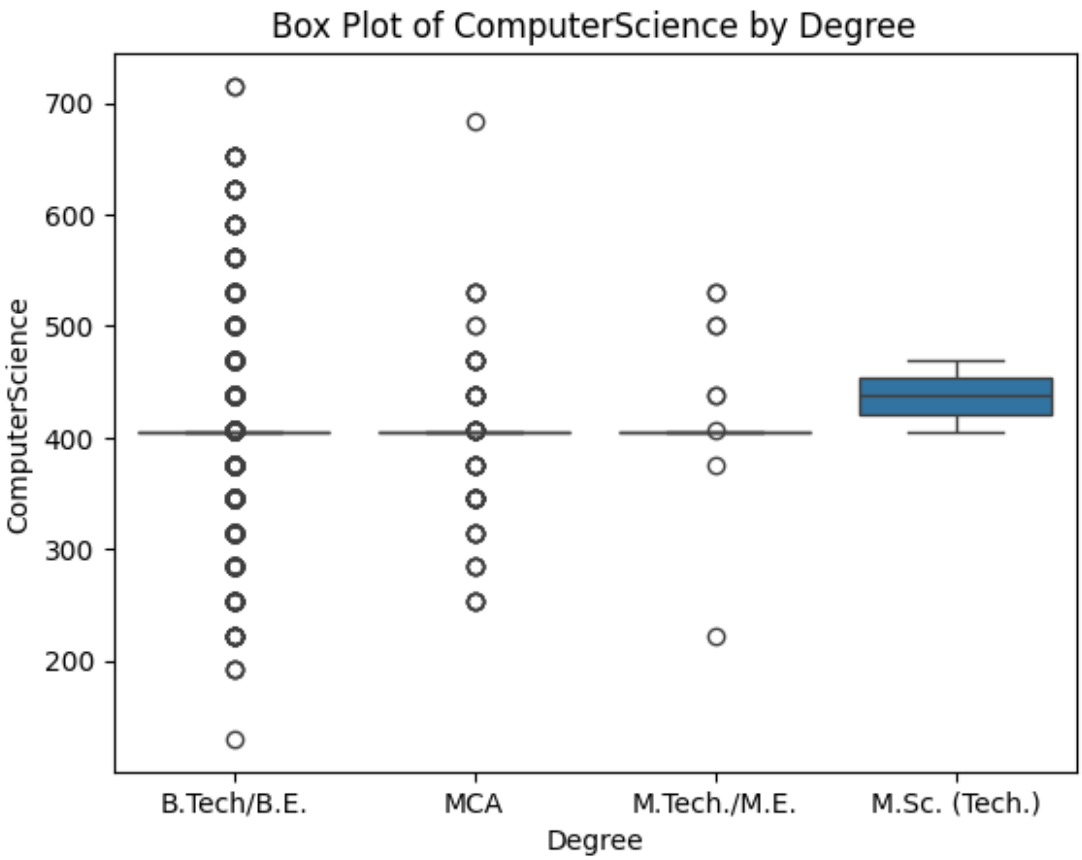

Box Plot of Logical by Degree
800
700
600
500
400
300
200
Logical
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

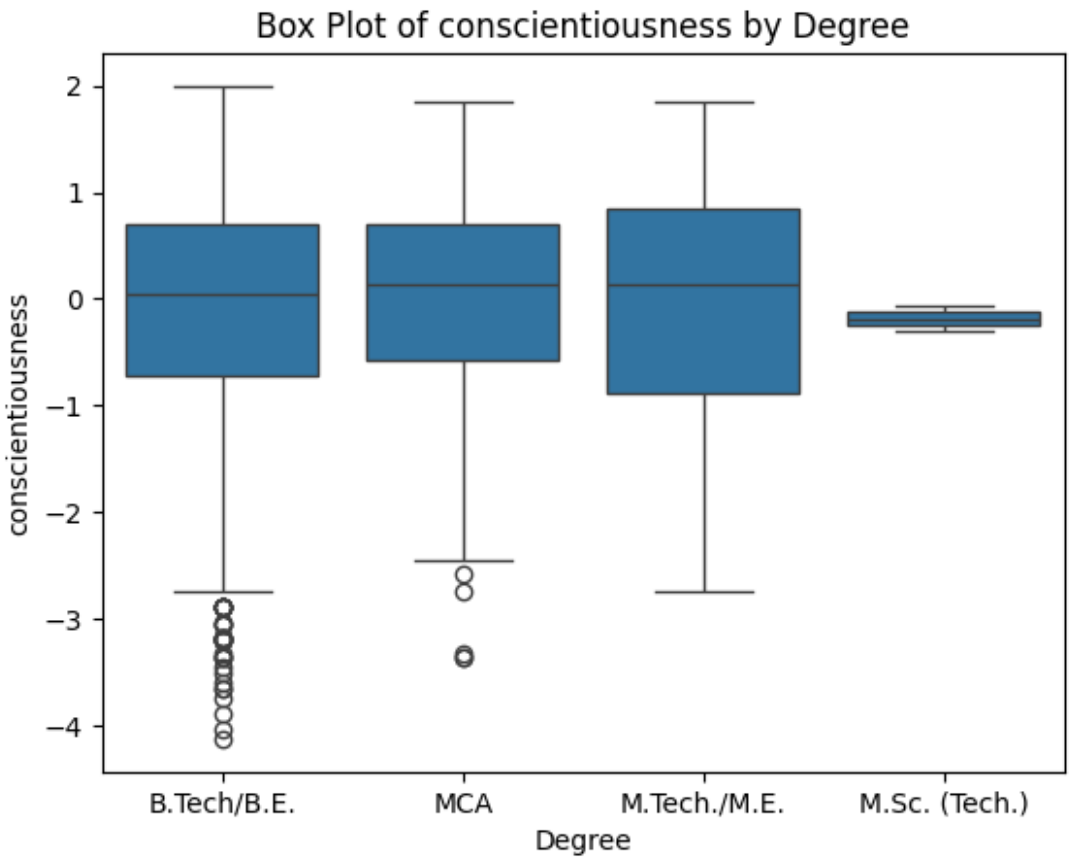

Box Plot of Quant by Degree
Quant
900
800
700
600
500
400
300
200
100
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

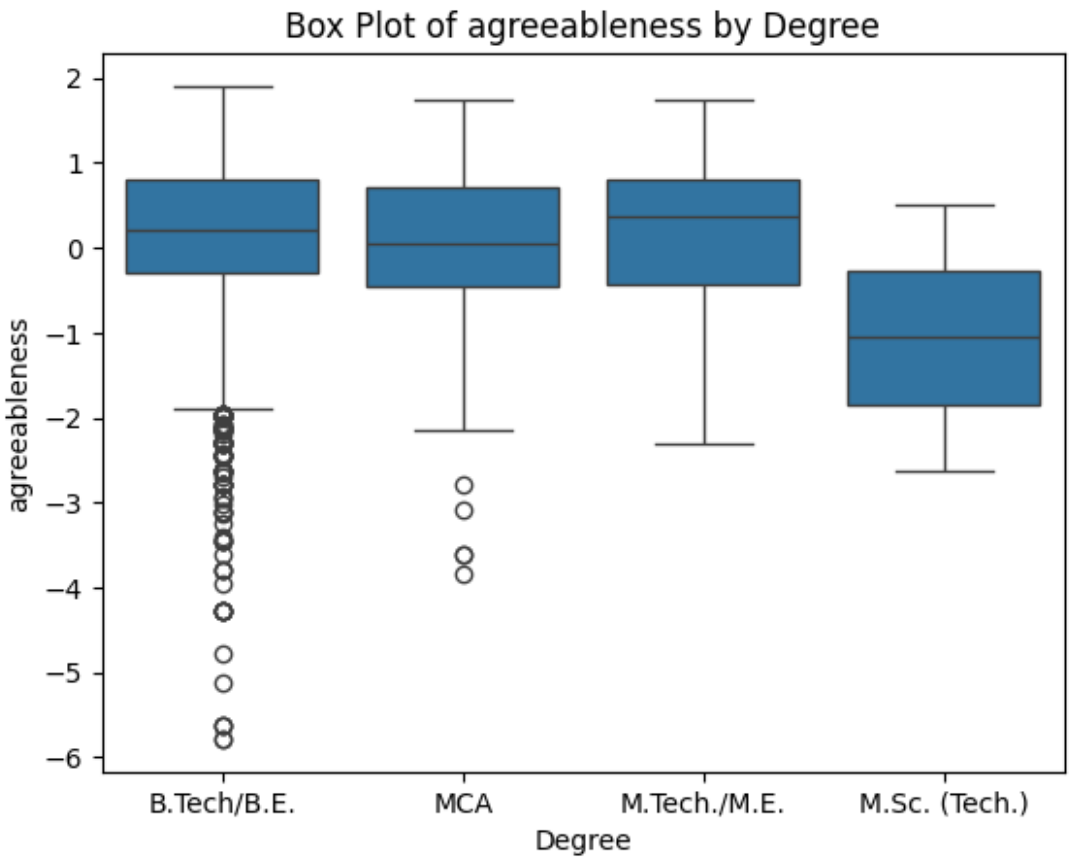

Box Plot of Domain by Degree
1.0
0.8
0.6
0.4
0.2
0.0
Domain
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

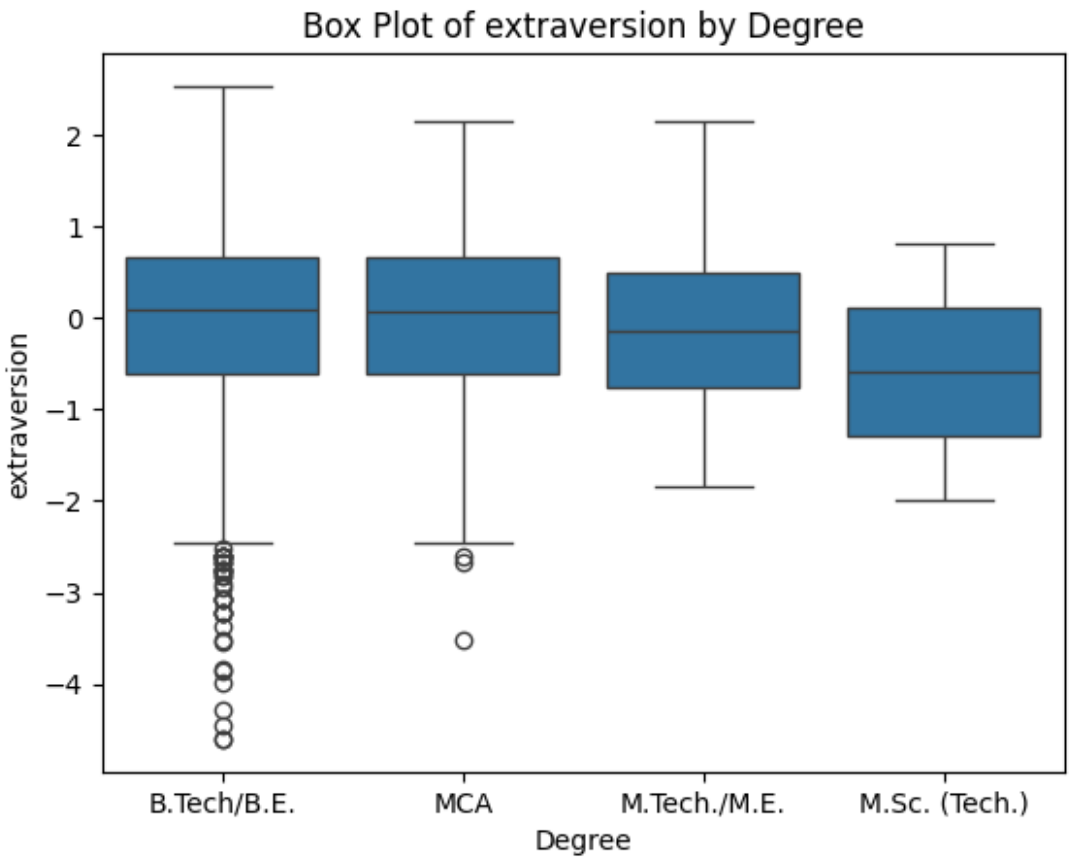

Box Plot of ComputerProgramming by Degree
ComputerProgramming
800
700
600
500
400
300
200
100
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

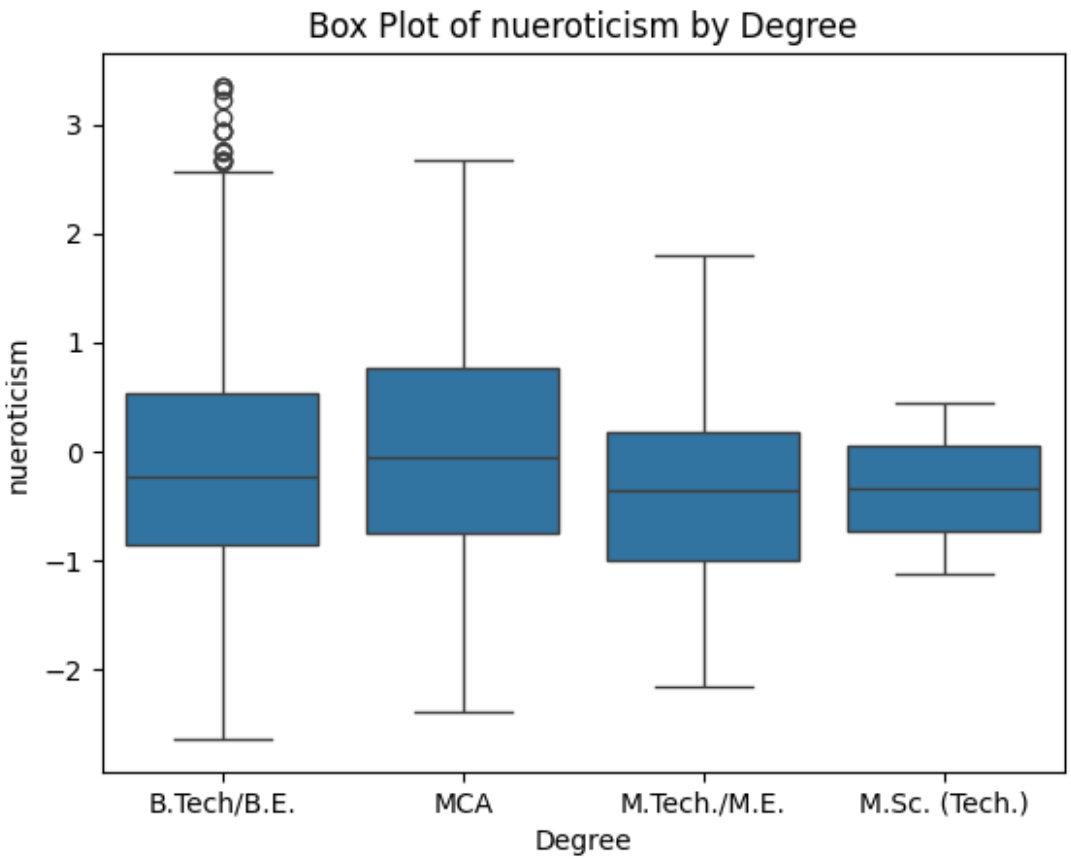

Box Plot of ElectronicsAndSemicon by Degree
ElectronicsAndSemicon
600
500
400
300
200
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

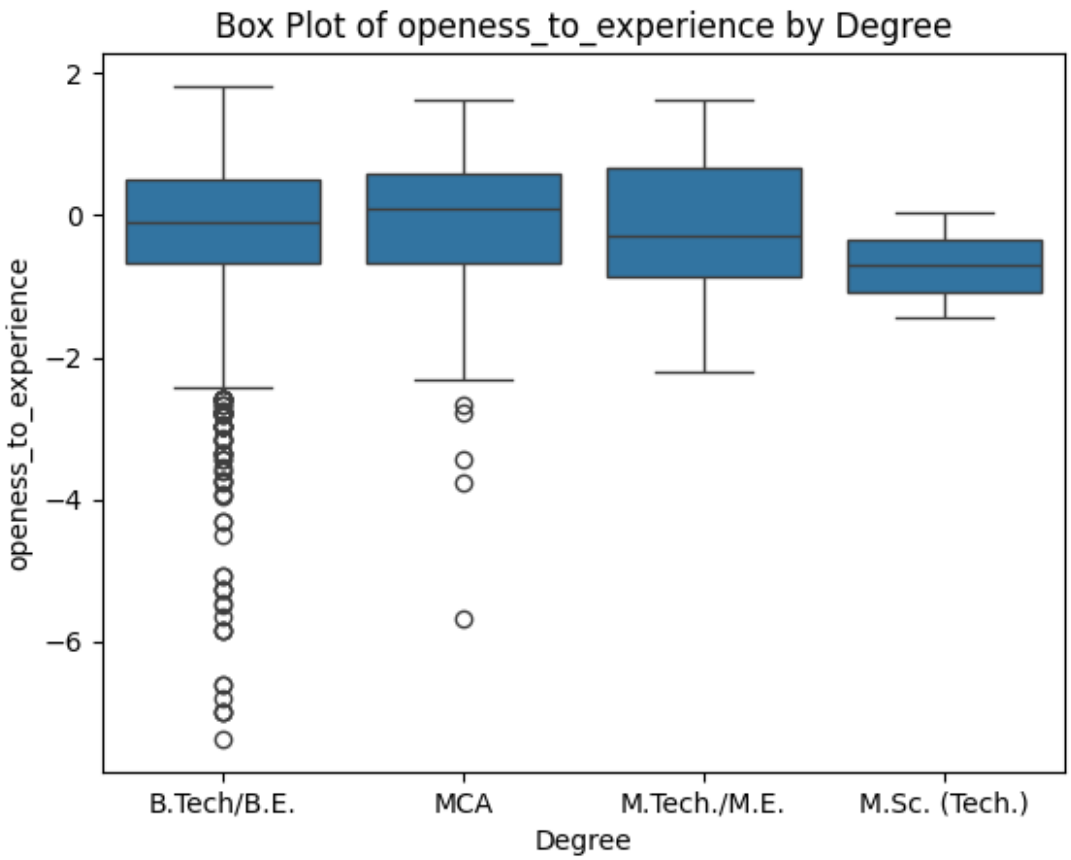

Box Plot of ComputerScience by Degree
ComputerScience
700
600
500
400
300
200
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

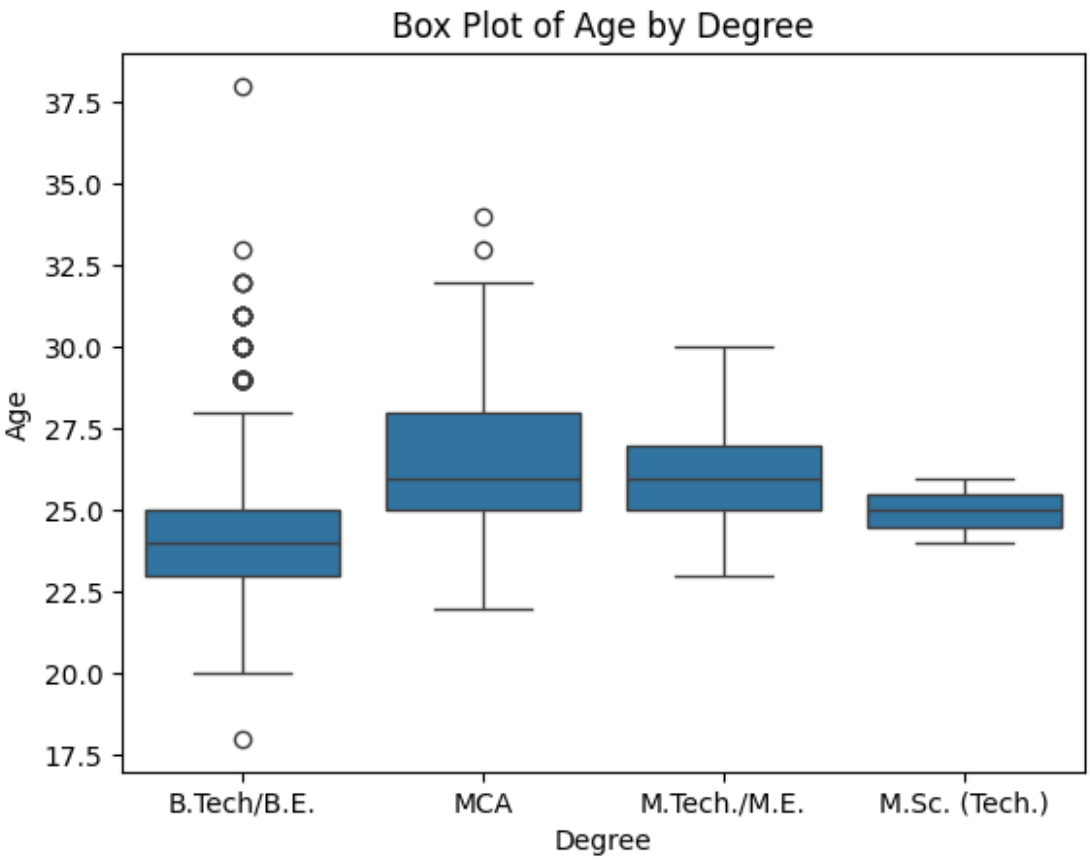

Box Plot of conscientiousness by Degree
conscientiousness
2
1
0
−1
−2
−3
−4
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

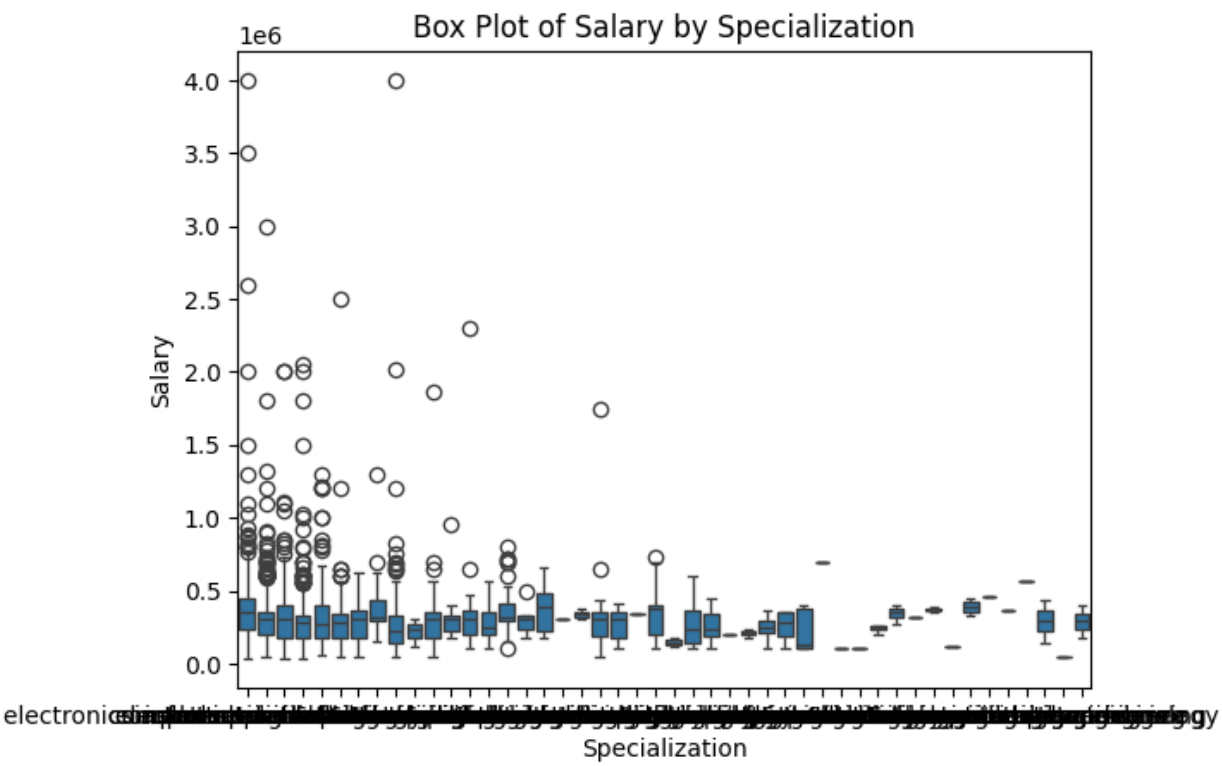

Box Plot of agreeableness by Degree
agreeableness
2
1
0
−1
−2
−3
−4
−5
−6
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

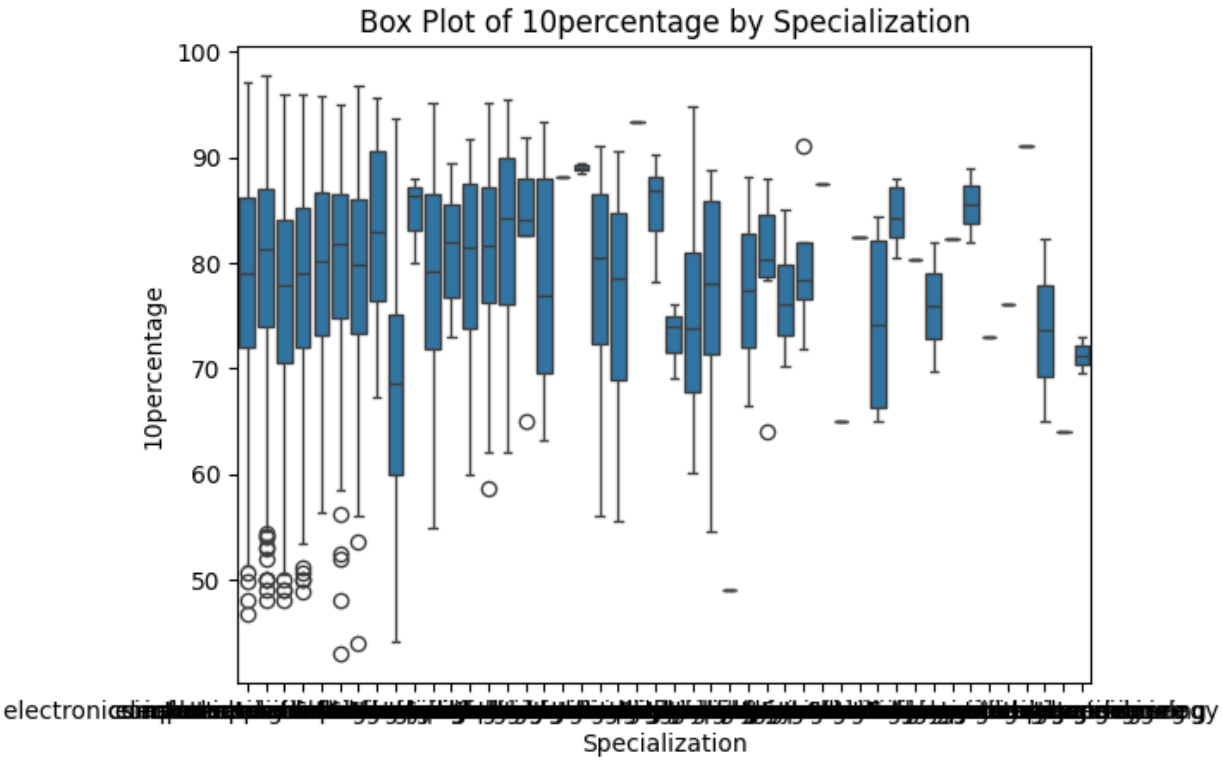

Box Plot of extraversion by Degree
extraversion
2
1
0
−1
−2
−3
−4
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

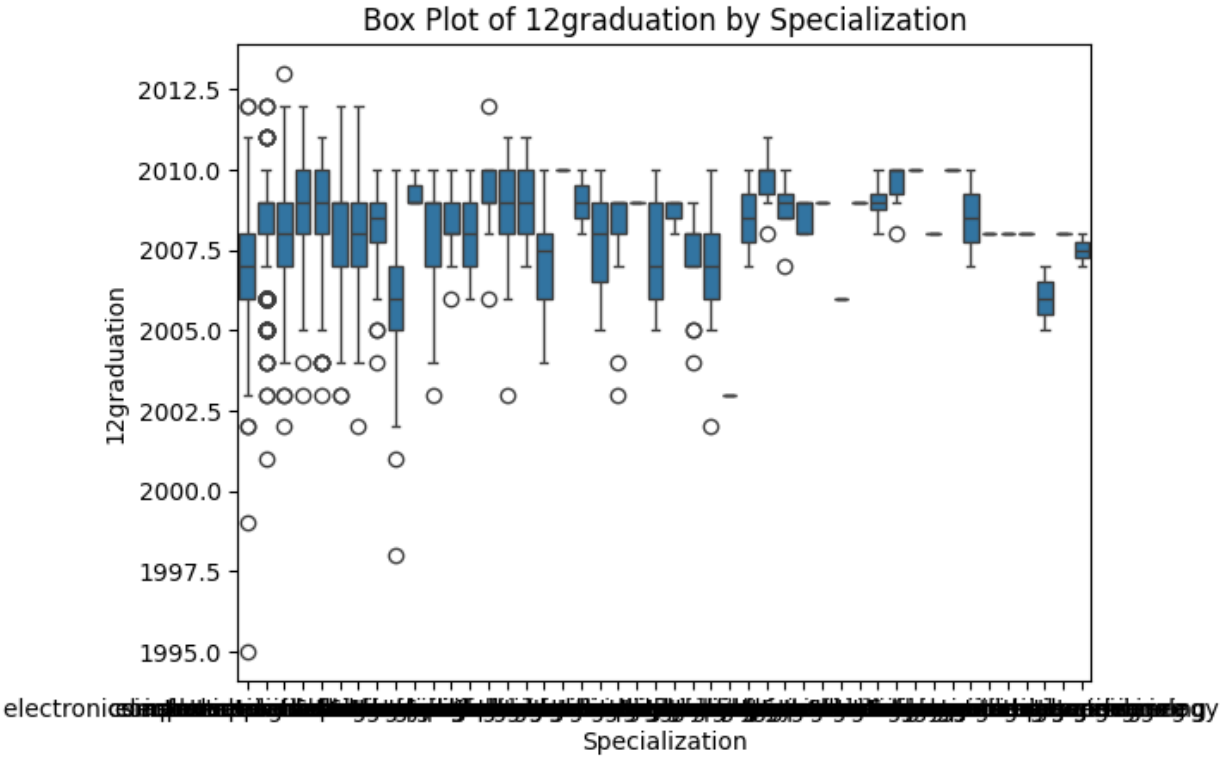

Box Plot of nueroticism by Degree
nueroticism
3
2
1
0
−1
−2
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

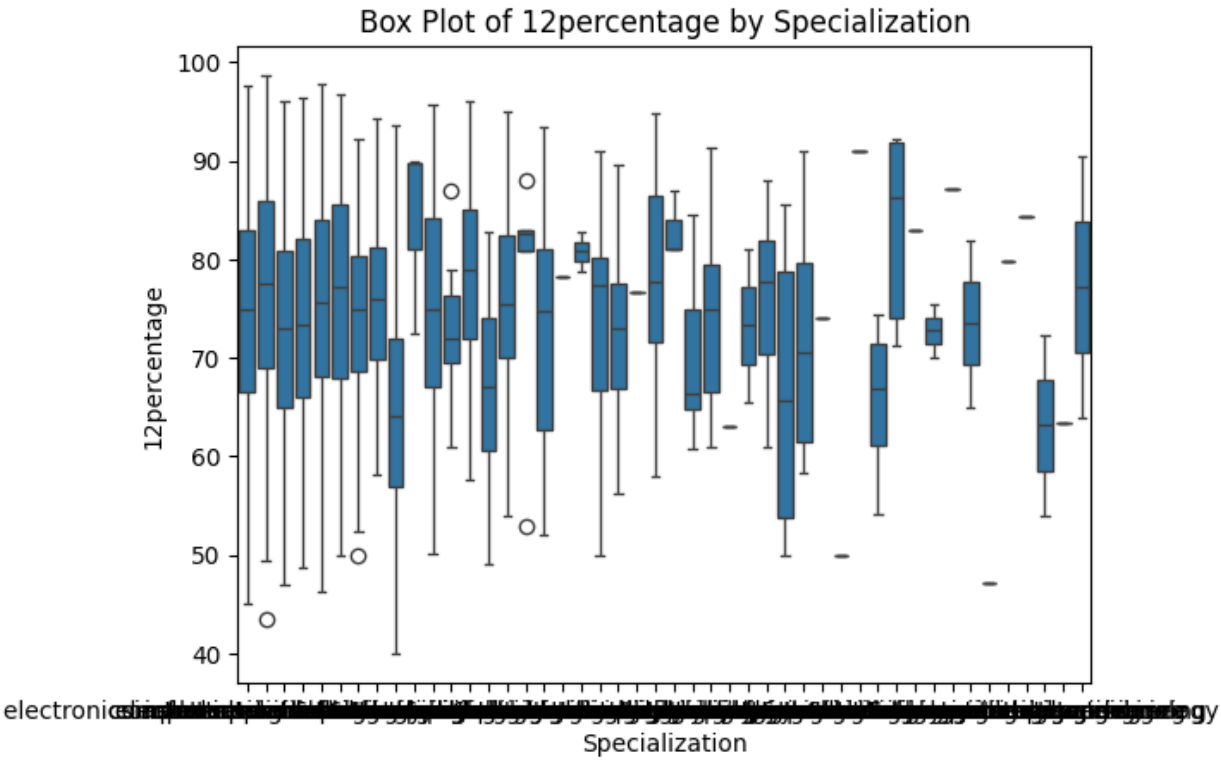

Box Plot of openess_to_experience by Degree
2
0
−2
−4
−6
openess_to_experience
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

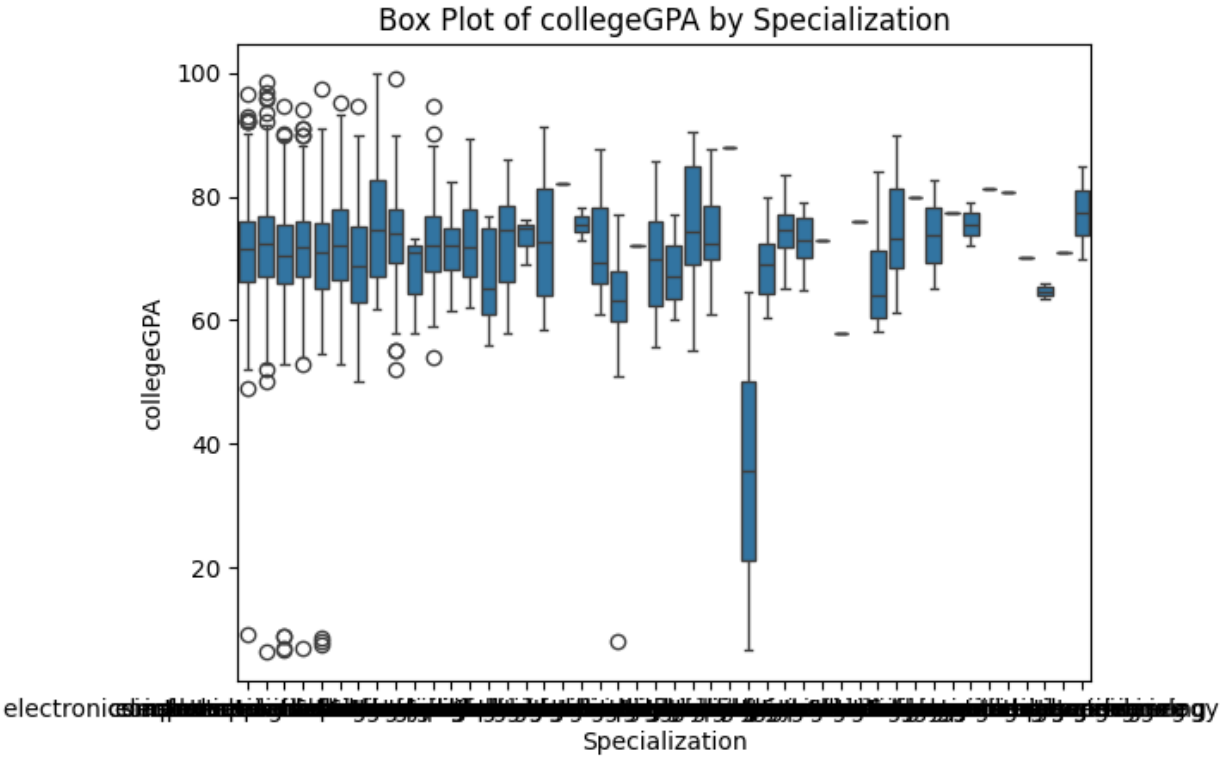

Box Plot of Age by Degree
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

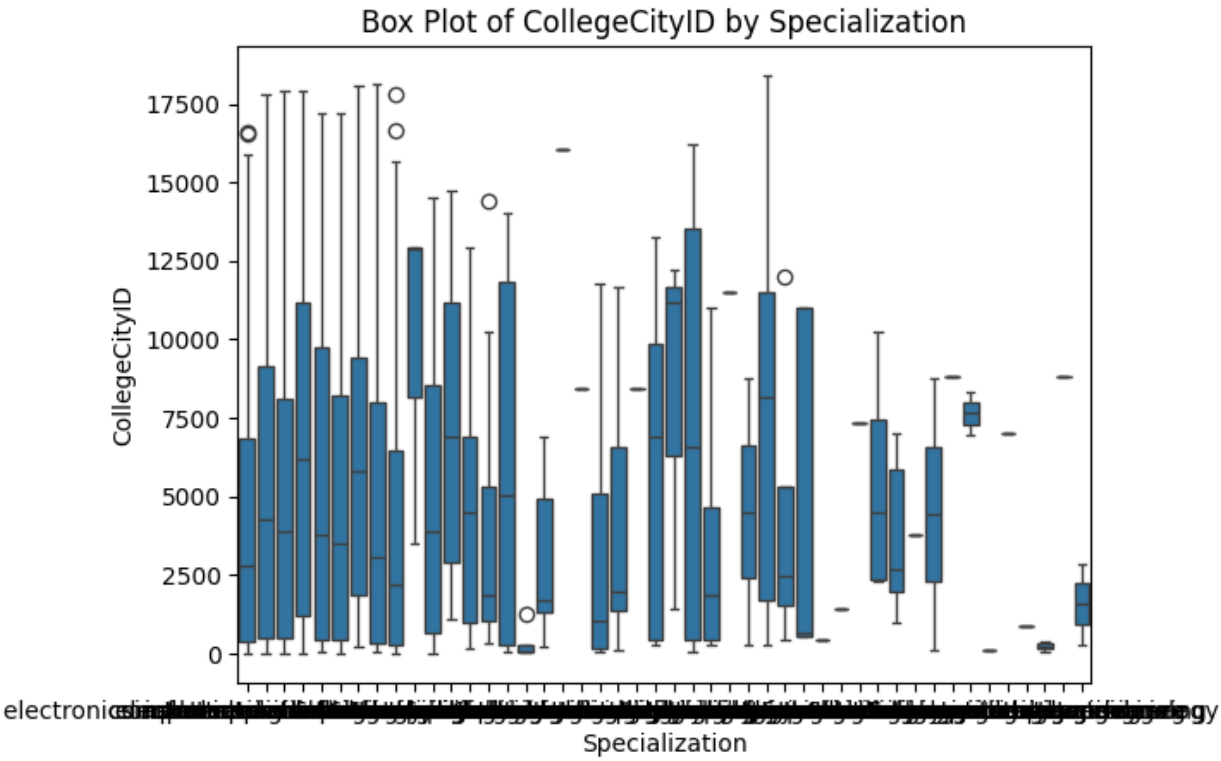

Box Plot of Salary by Specialization
1e6
Salary
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Specialization

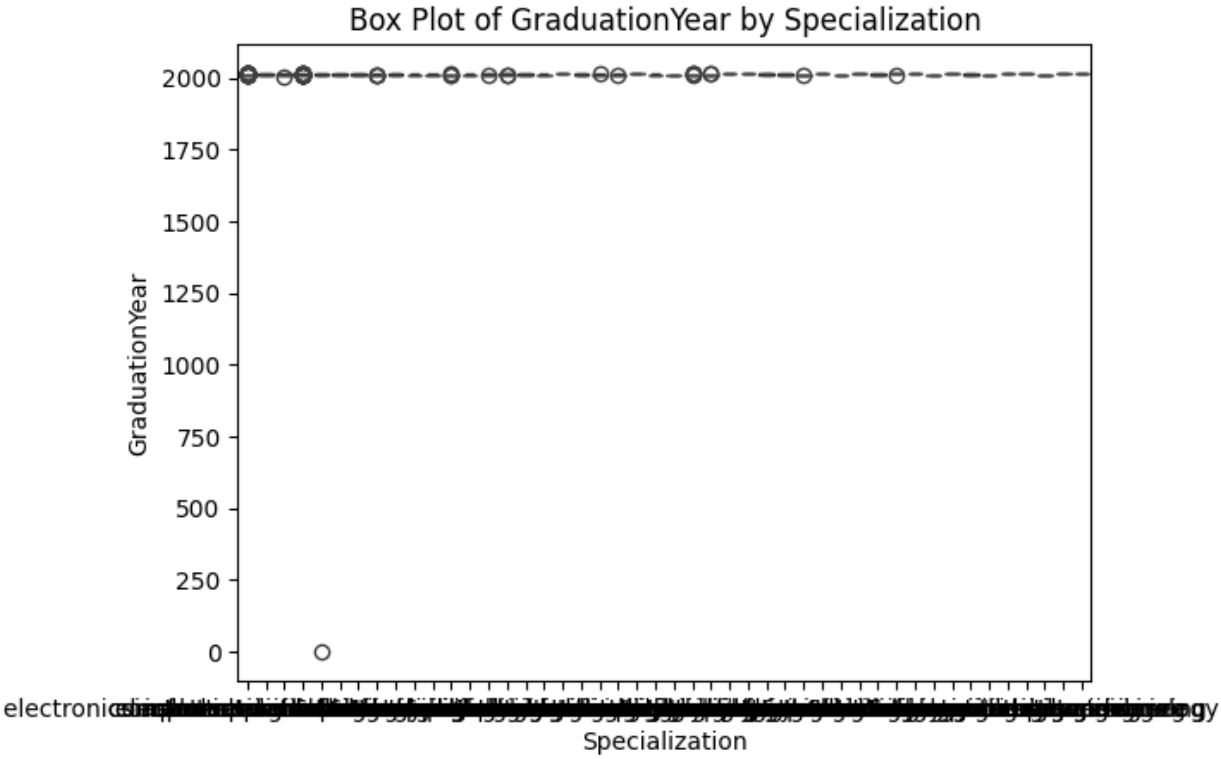

Box Plot of 10percentage by Specialization
10percentage
100
90
80
70
60
50
electronic
Specialization

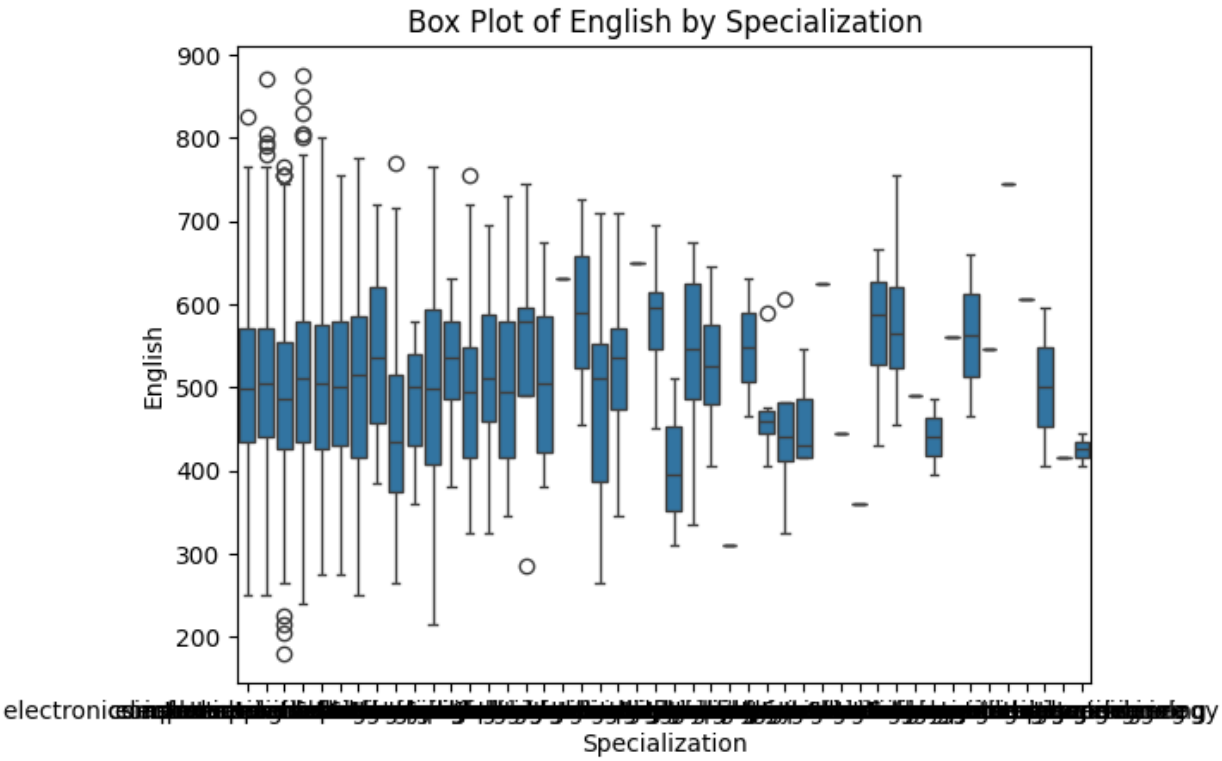

Box Plot of 12graduation by Specialization
12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
electronic
Specialization

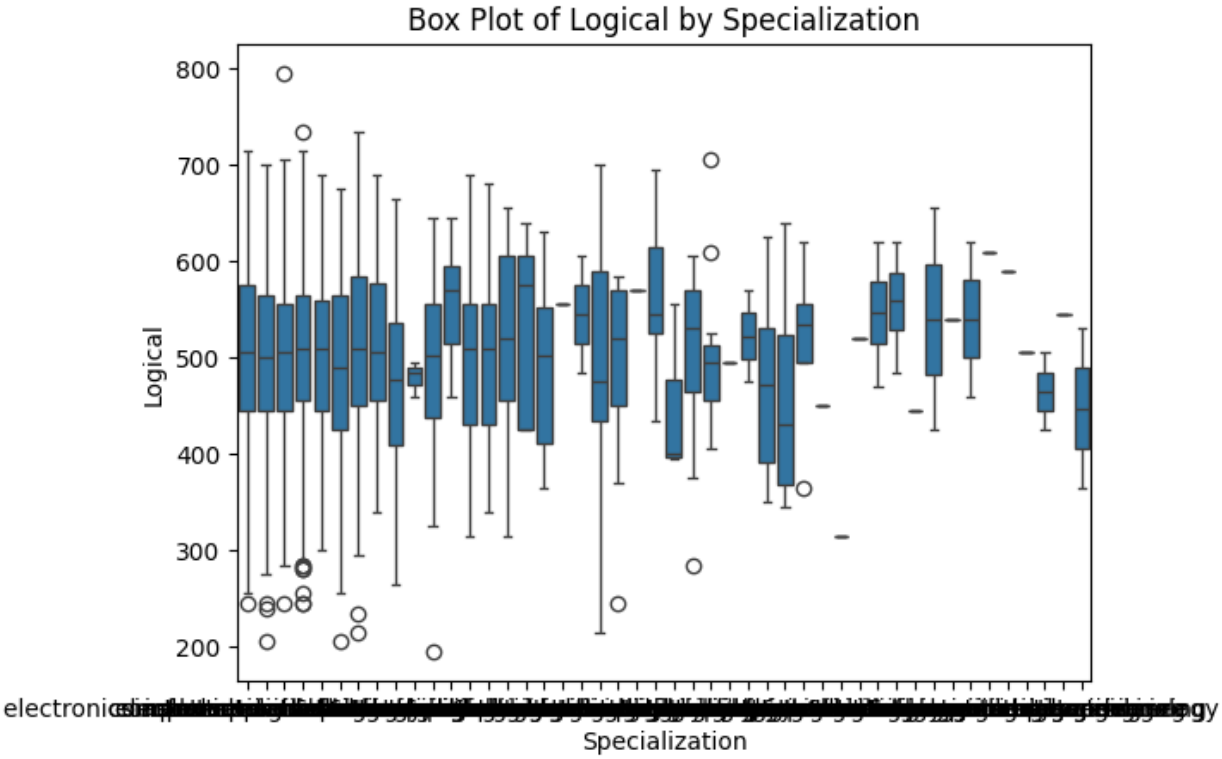

Box Plot of 12percentage by Specialization
12percentage
100
90
80
70
60
50
40
electronic
Specialization

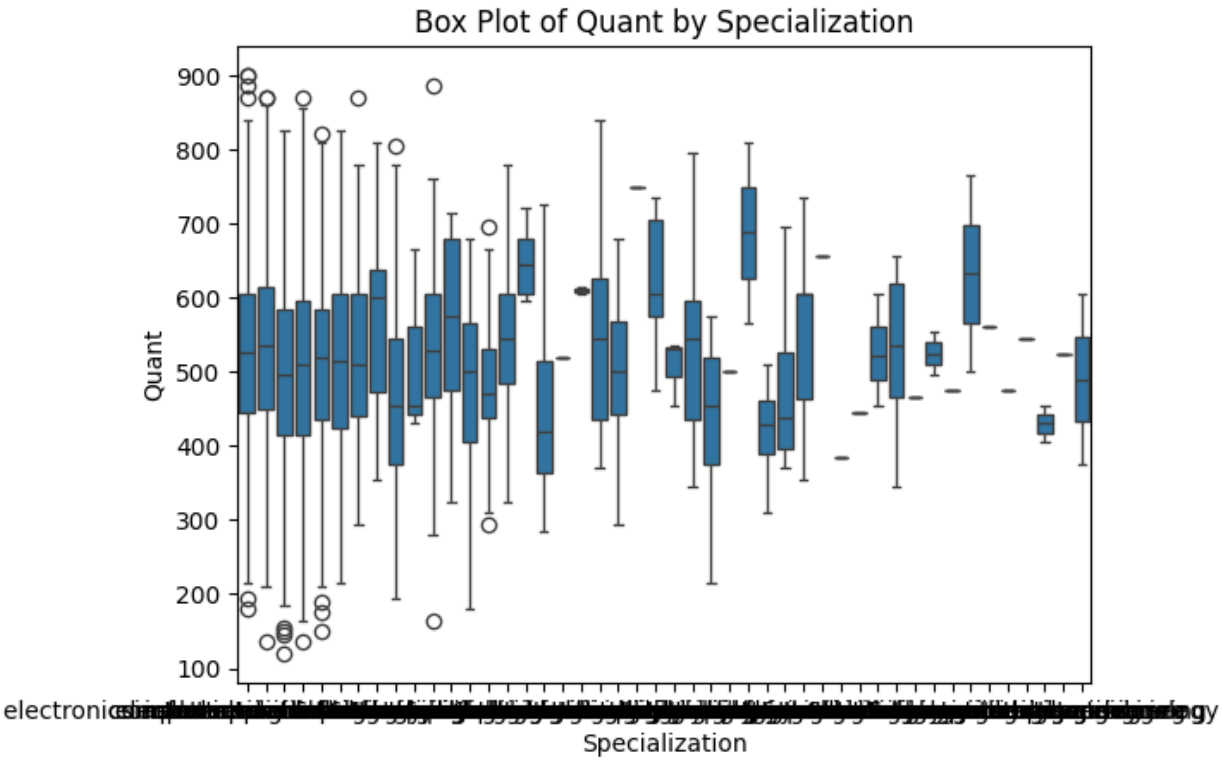

Box Plot of collegeGPA by Specialization
collegeGPA
100
80
60
40
20
electronic
Specialization

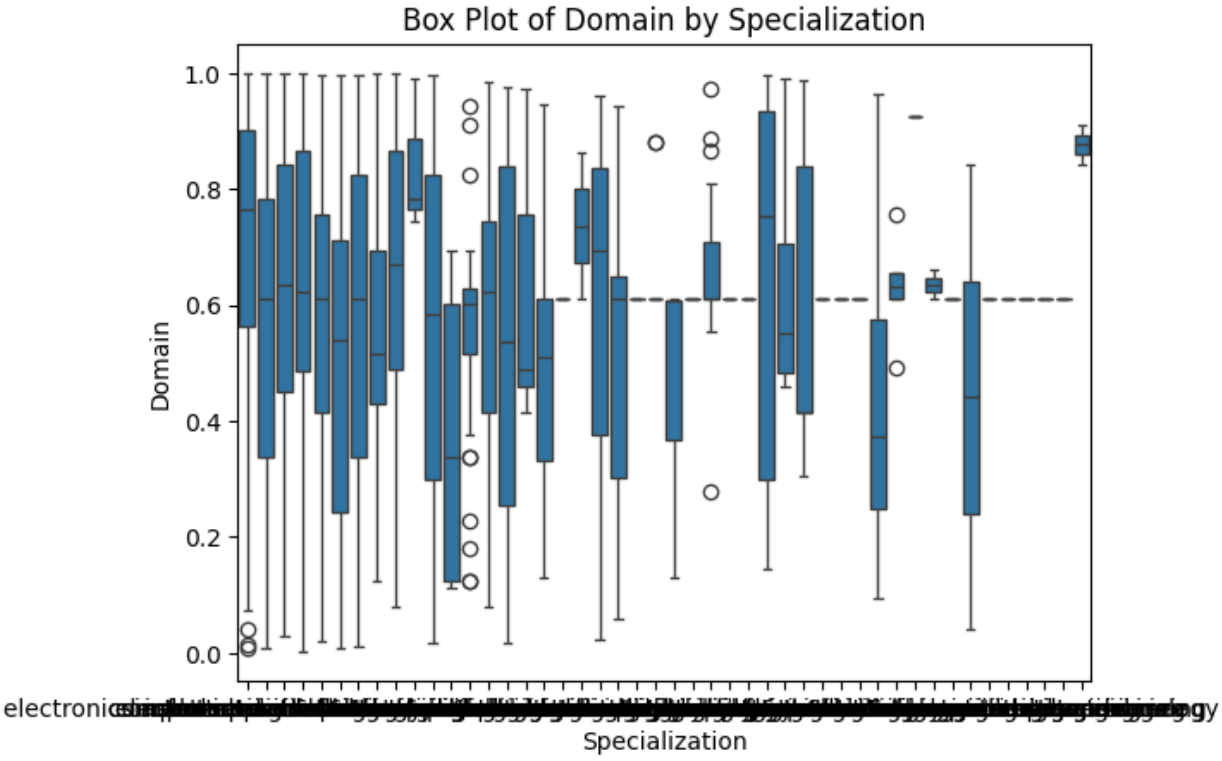

Box Plot of CollegeCityID by Specialization
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
electronic
Specialization

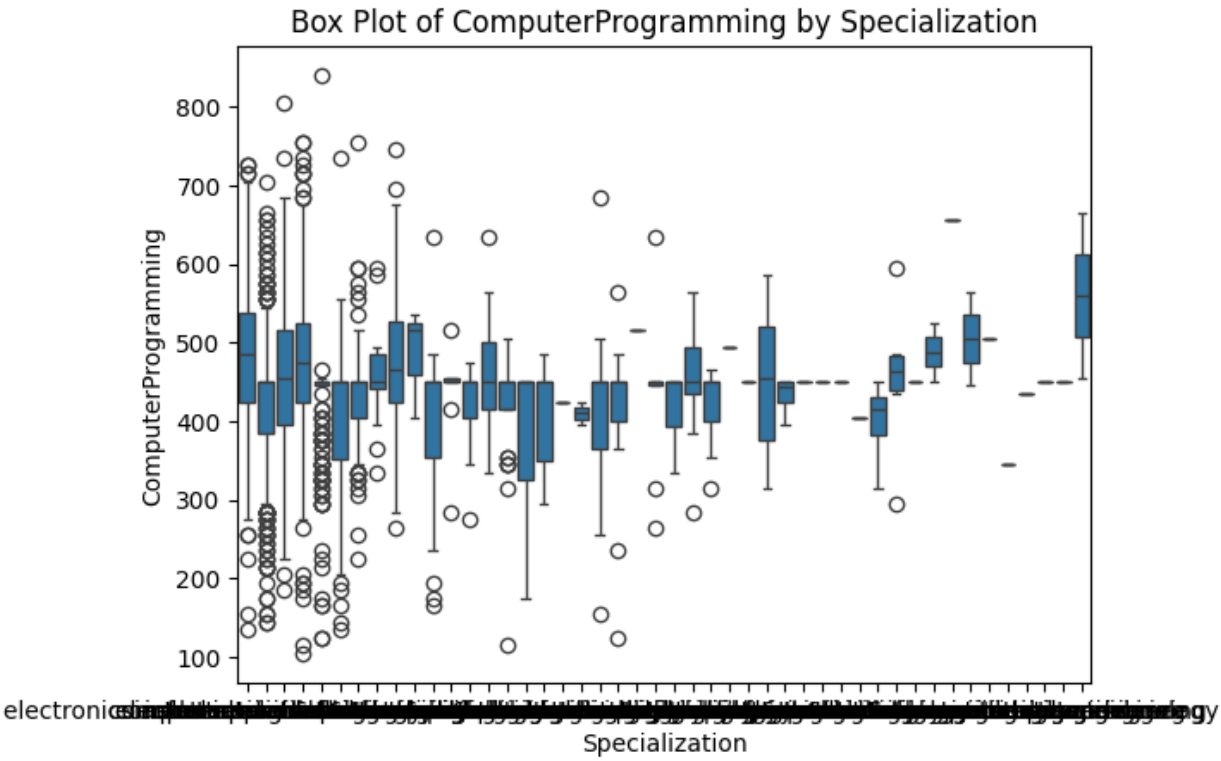

Box Plot of GraduationYear by Specialization
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
electronic
Specialization

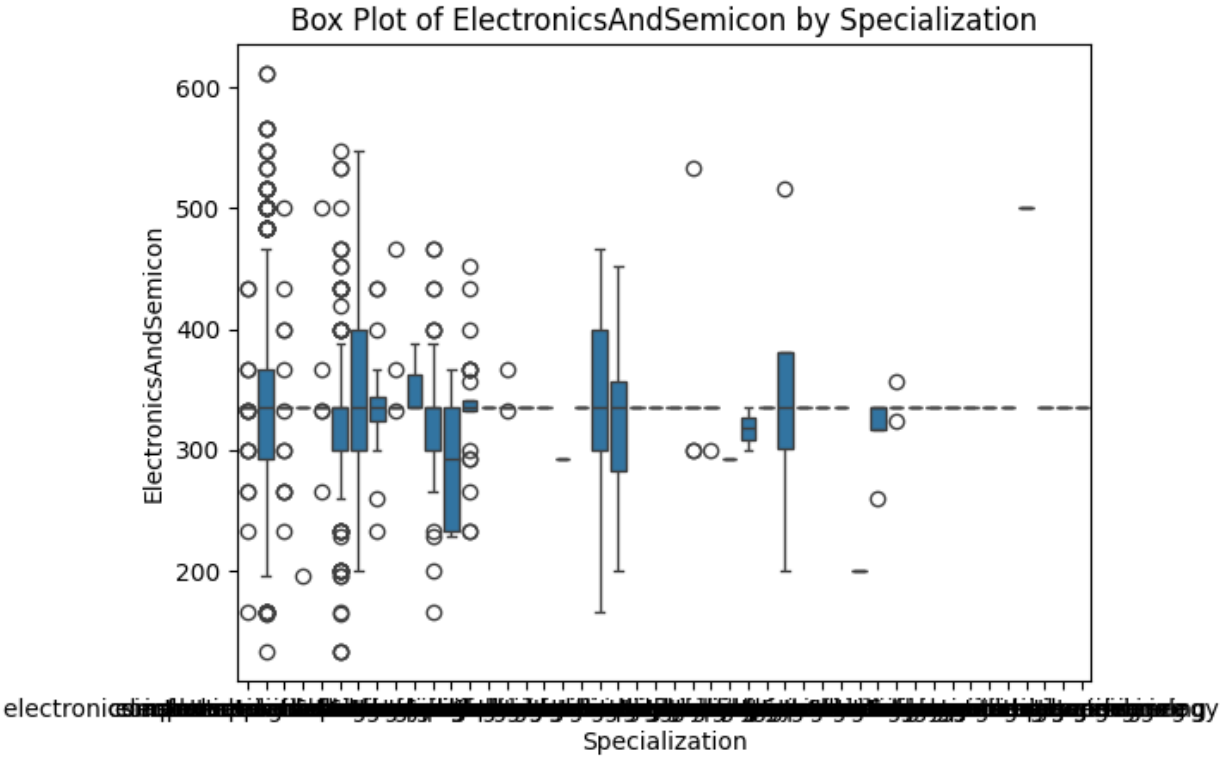

Box Plot of English by Specialization
English
900
800
700
600
500
400
300
200
electronic
Specialization

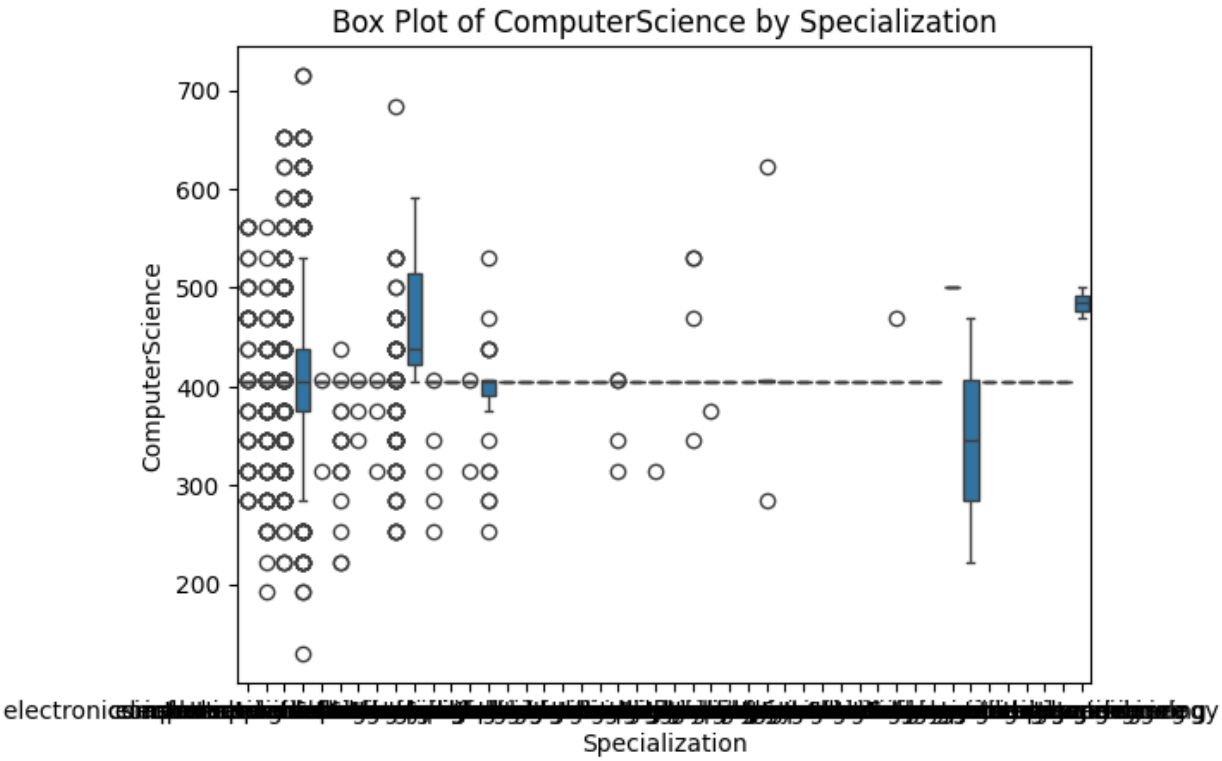

Box Plot of Logical by Specialization
Logical
800
700
600
500
400
300
200
electronic
Specialization

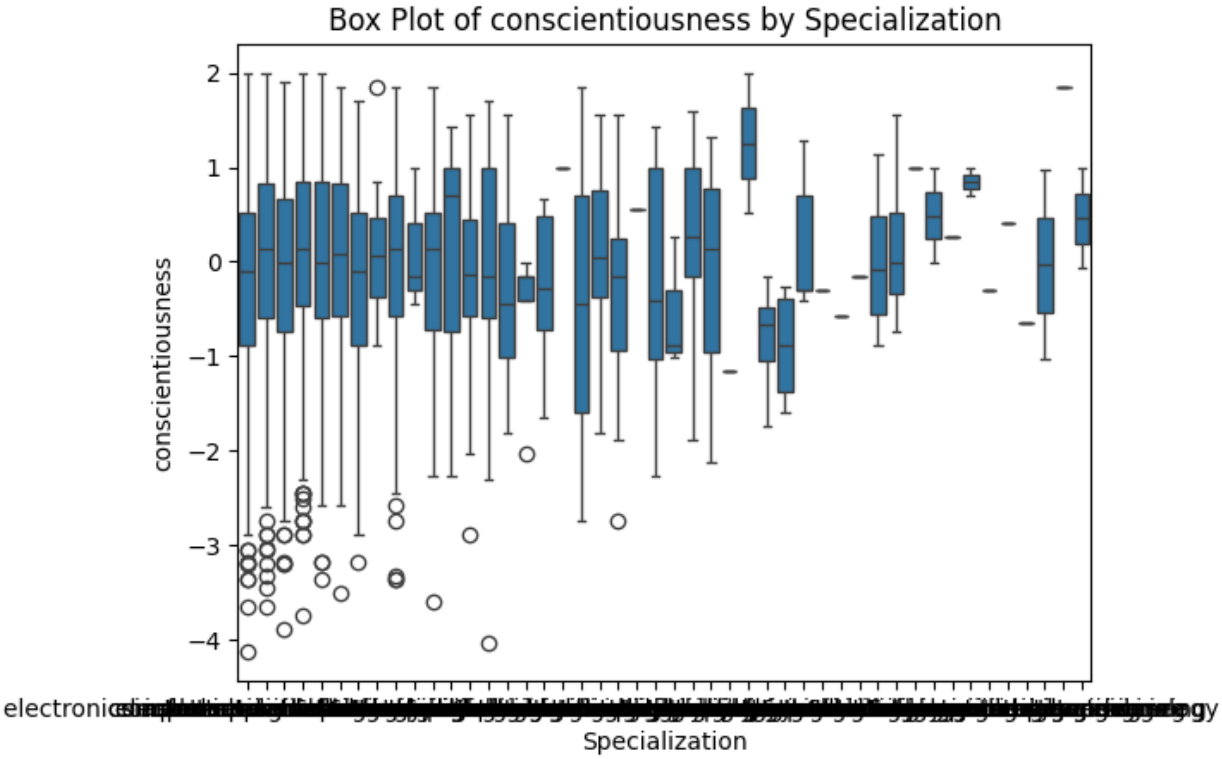

Box Plot of Quant by Specialization
Quant
900
800
700
600
500
400
300
200
100
Specialization

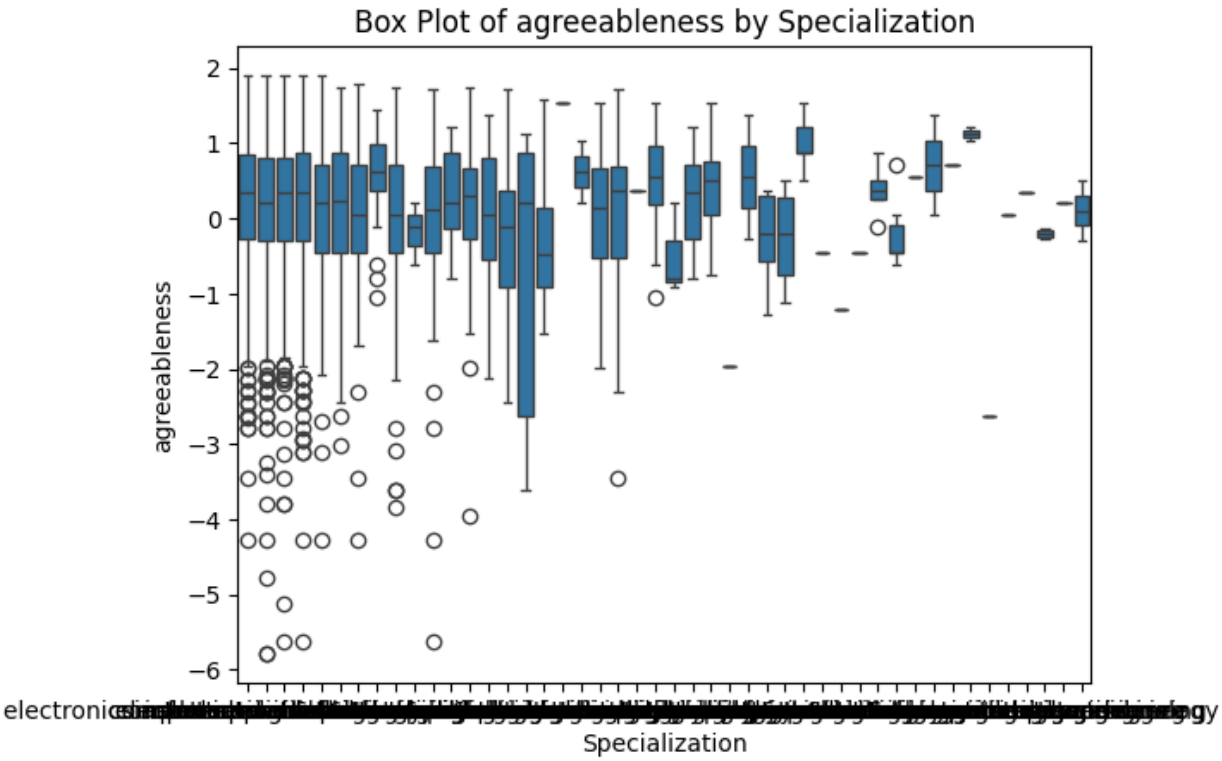

Box Plot of Domain by Specialization
Domain
1.0
0.8
0.6
0.4
0.2
0.0
Specialization

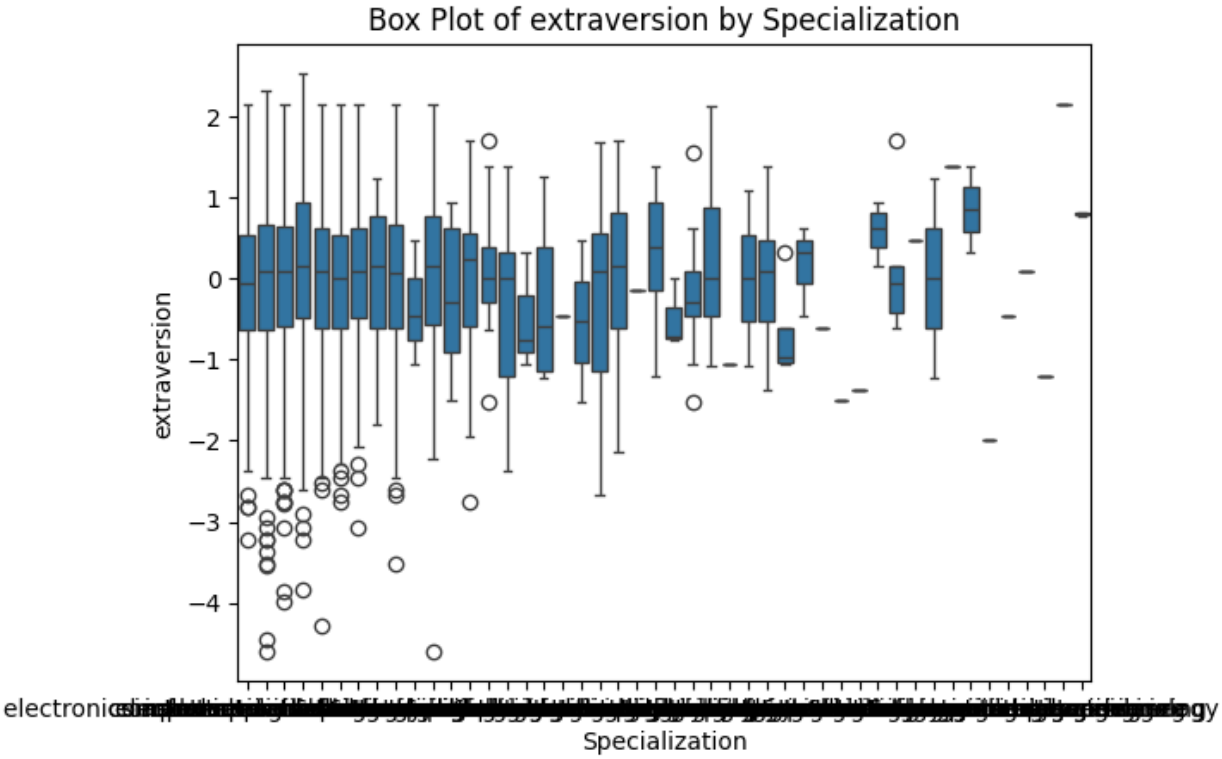

Box Plot of ComputerProgramming by Specialization
ComputerProgramming
800
700
600
500
400
300
200
100
electronic
Specialization

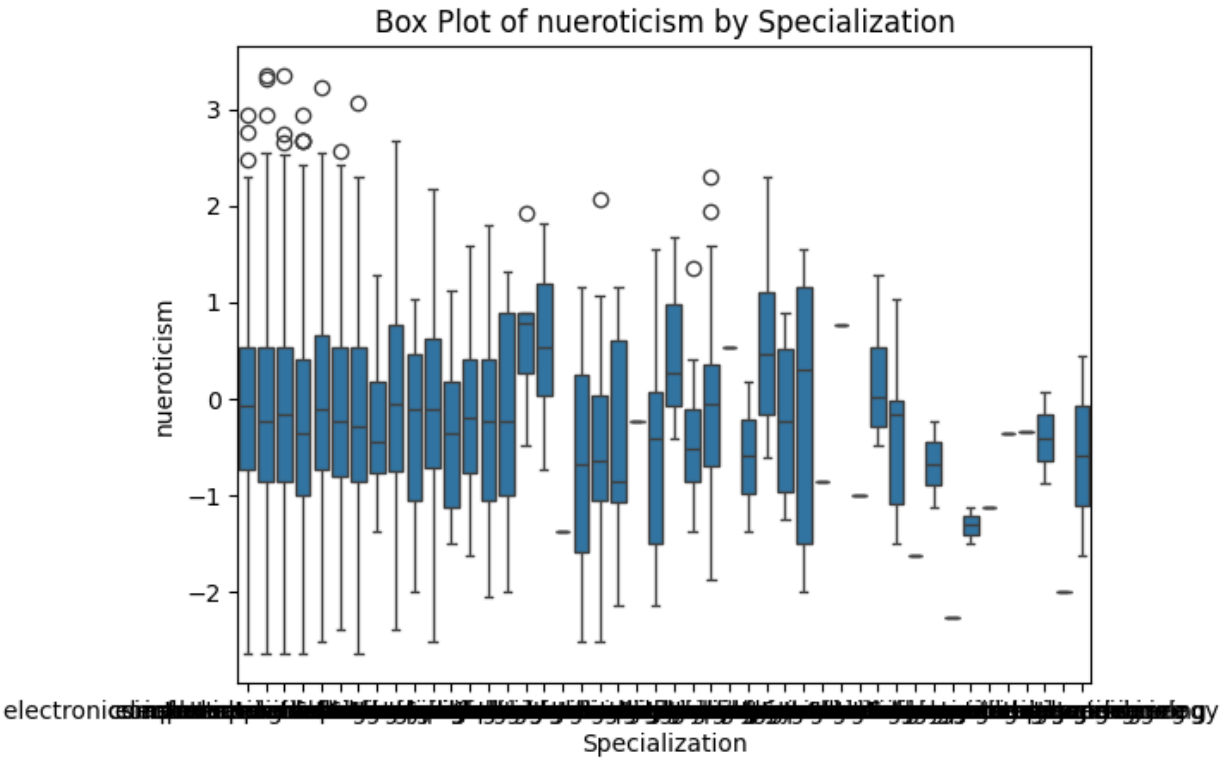

Box Plot of ElectronicsAndSemicon by Specialization
ElectronicsAndSemicon
600
500
400
300
200
electronic
Specialization

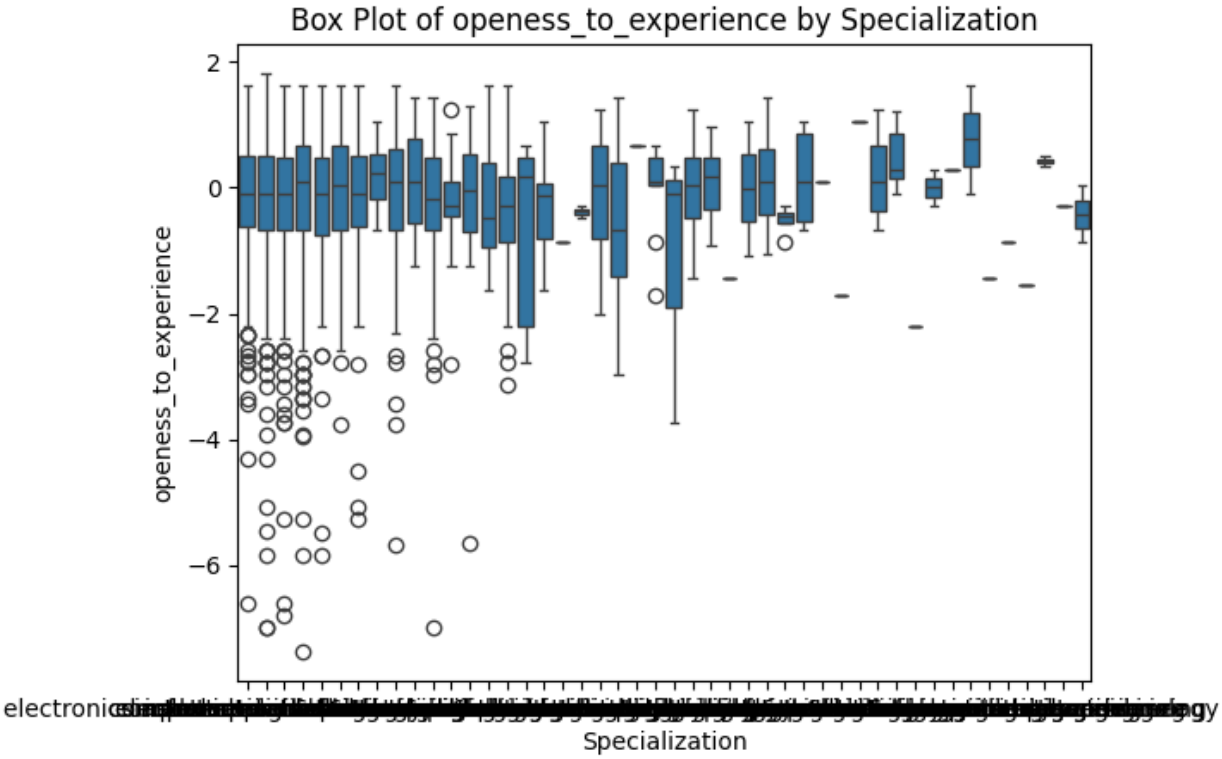

Box Plot of ComputerScience by Specialization
ComputerScience
700
600
500
400
300
200
Specialization

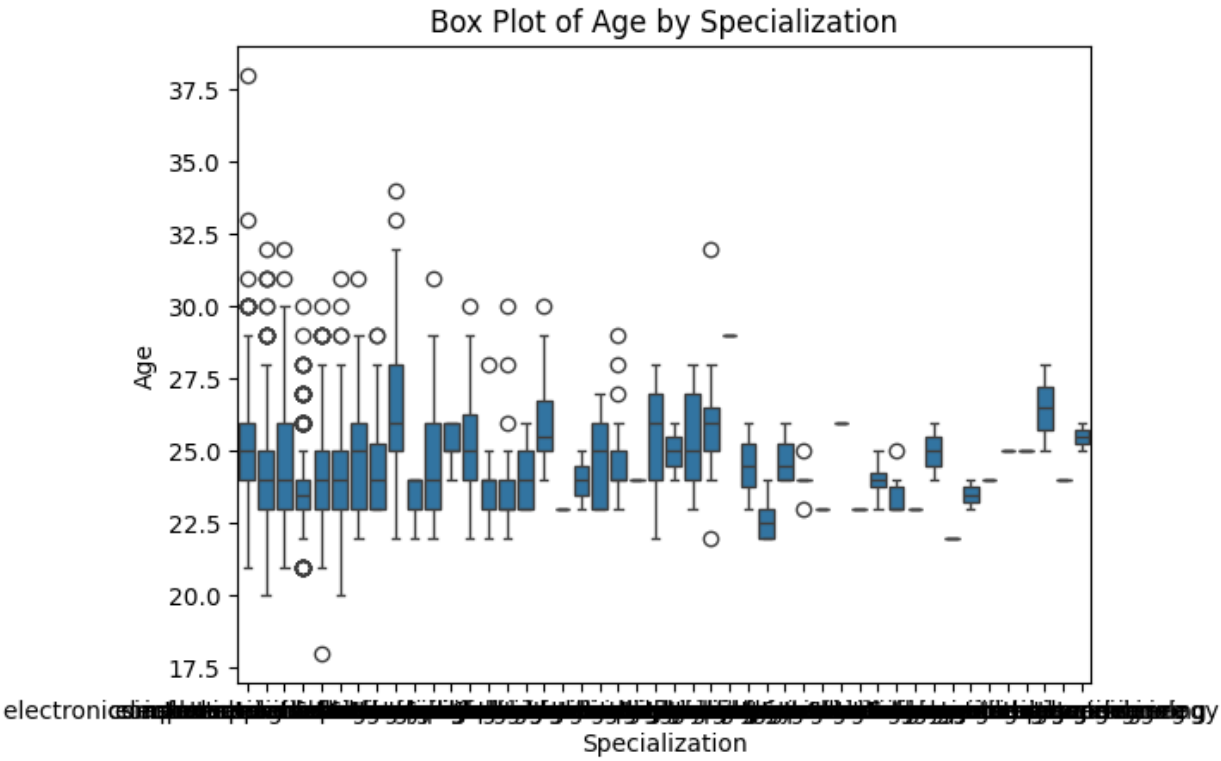

Box Plot of conscientiousness by Specialization
conscientiousness
2
1
0
−1
−2
−3
−4
Specialization

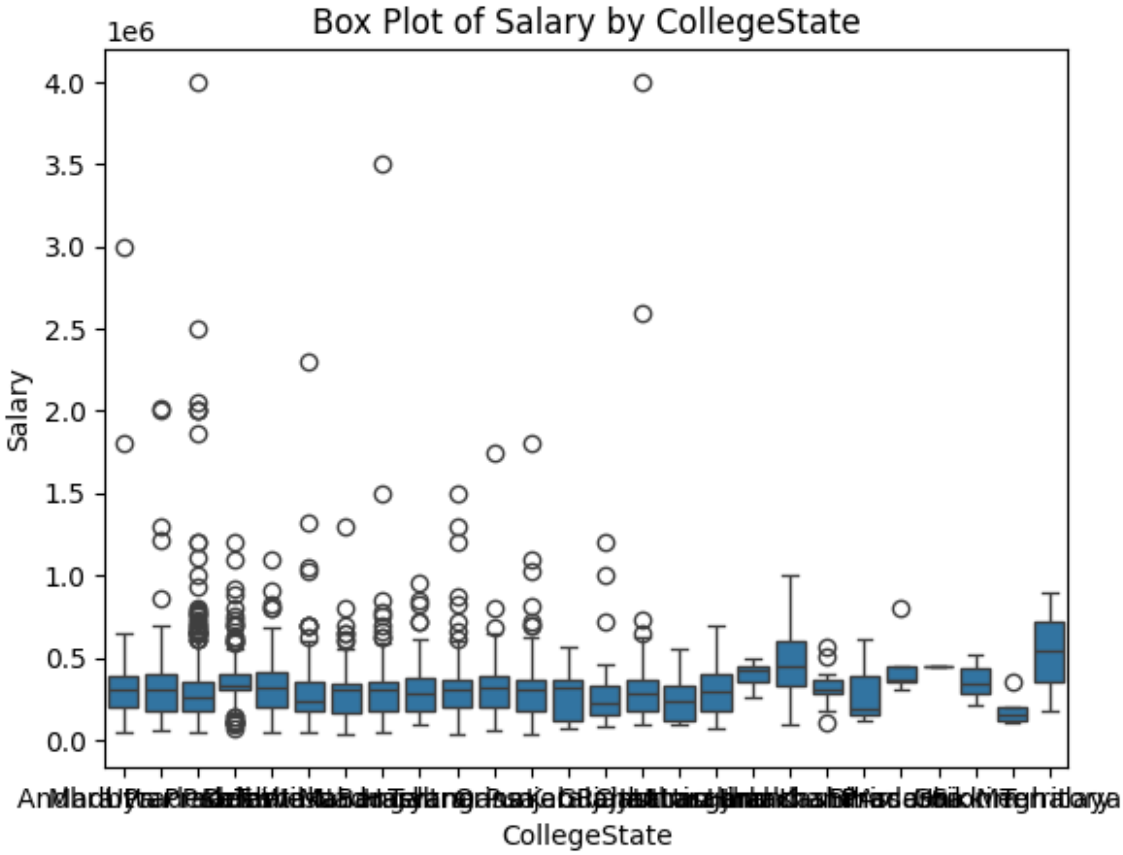

Box Plot of agreeableness by Specialization
agreeableness
2
1
0
−1
−2
−3
−4
−5
−6
electronic
Specialization

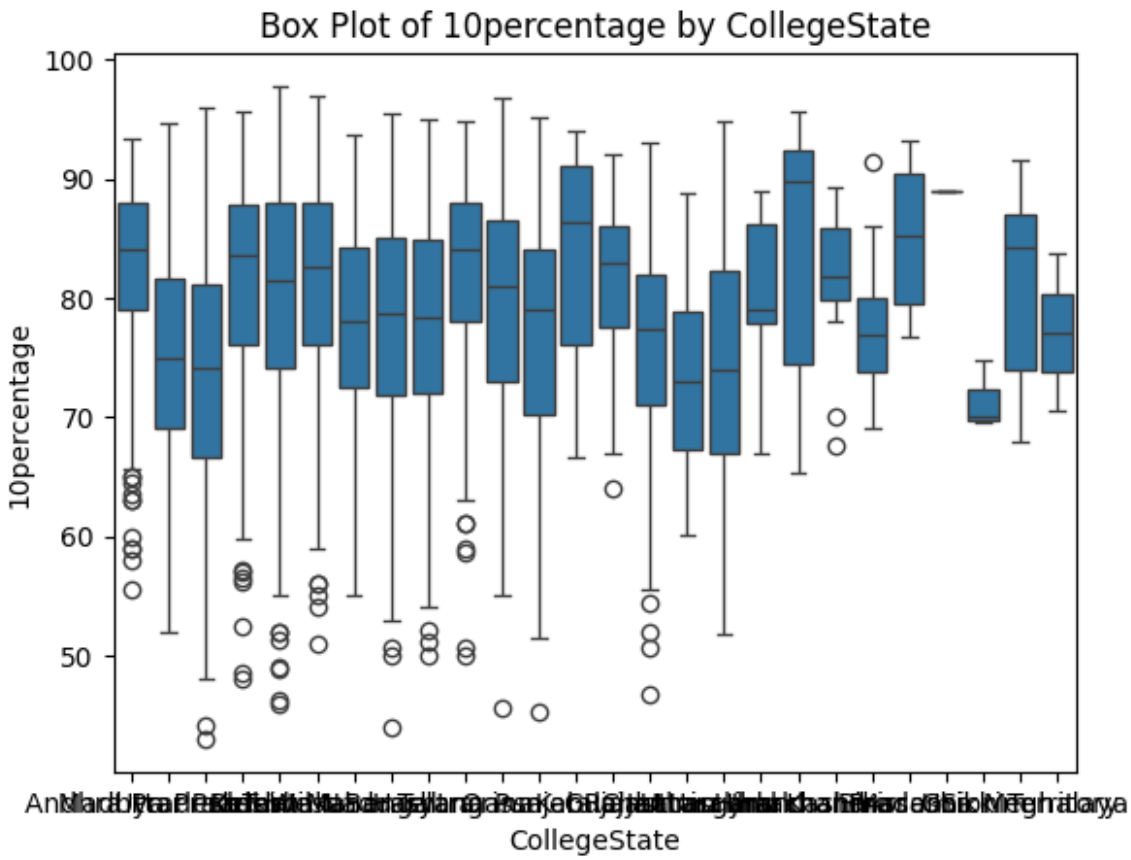

Box Plot of extraversion by Specialization
extraversion
2
1
0
−1
−2
−3
−4
electronic
Specialization

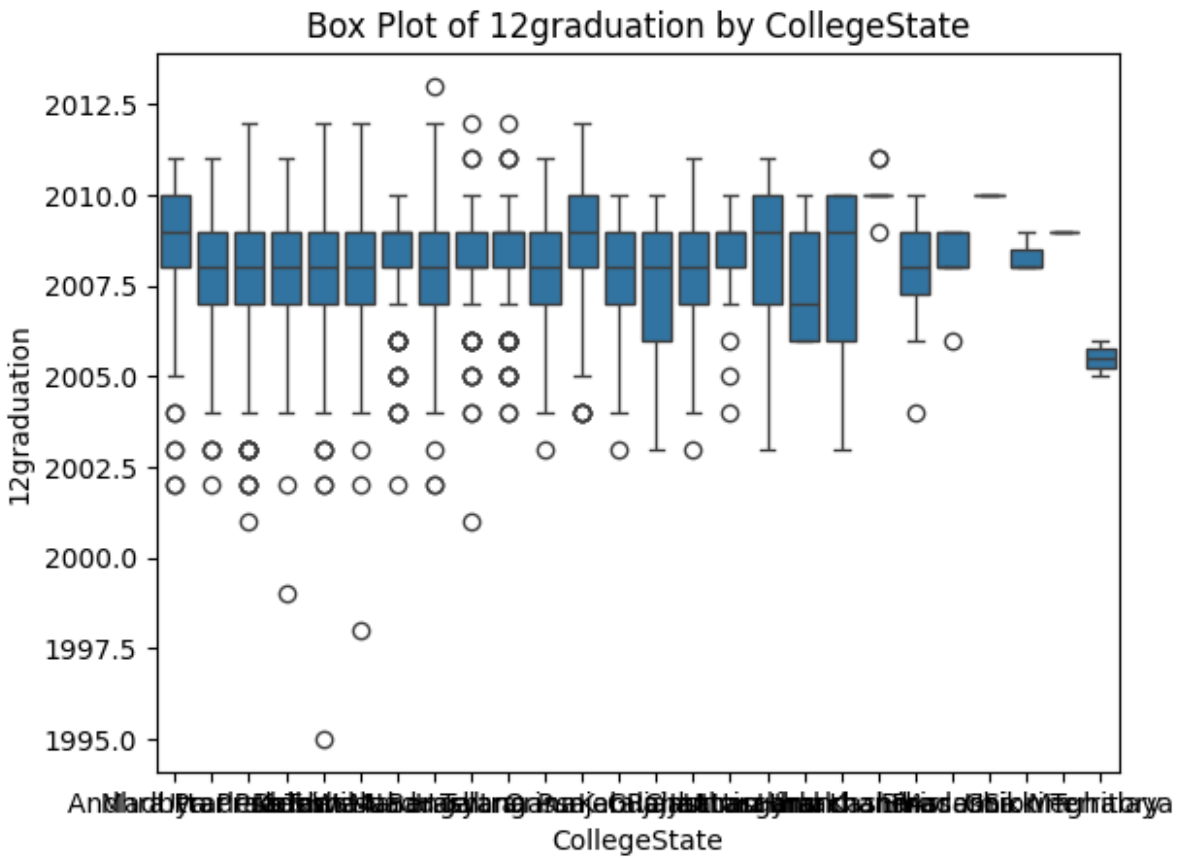

Box Plot of nueroticism by Specialization
nueroticism
3
2
1
0
−1
−2
electronic
Specialization

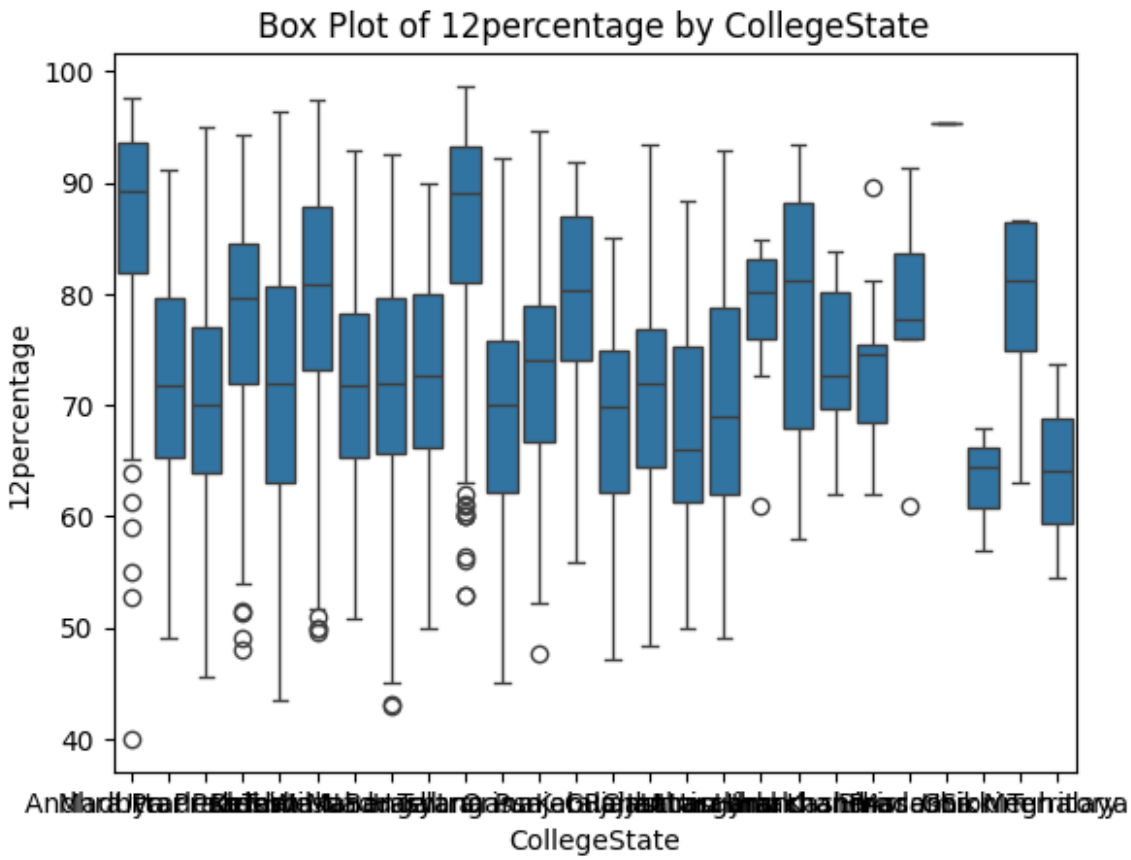

Box Plot of openess_to_experience by Specialization
openess_to_experience
2
0
−2
−4
−6
electronic
Specialization

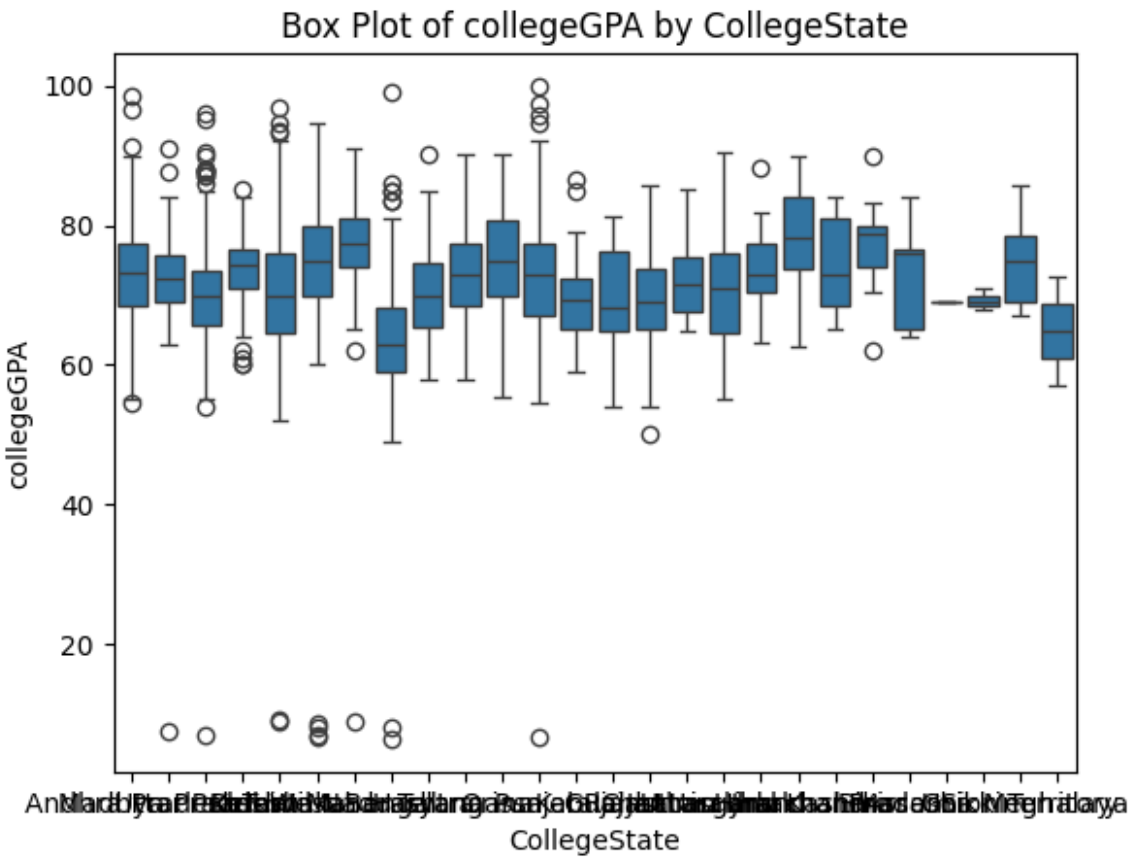

Box Plot of Age by Specialization
Age
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
electronics
Specialization

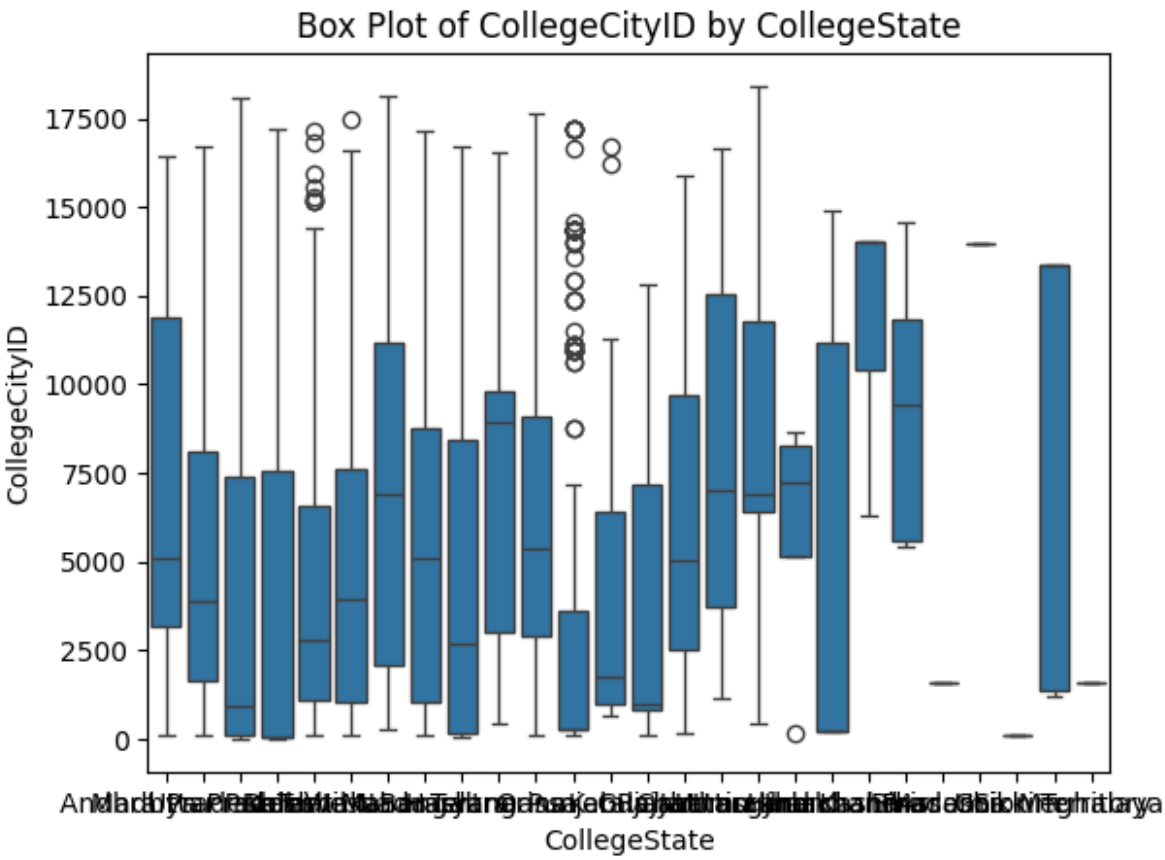

Box Plot of Salary by CollegeState
1e6
Salary
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
CollegeState

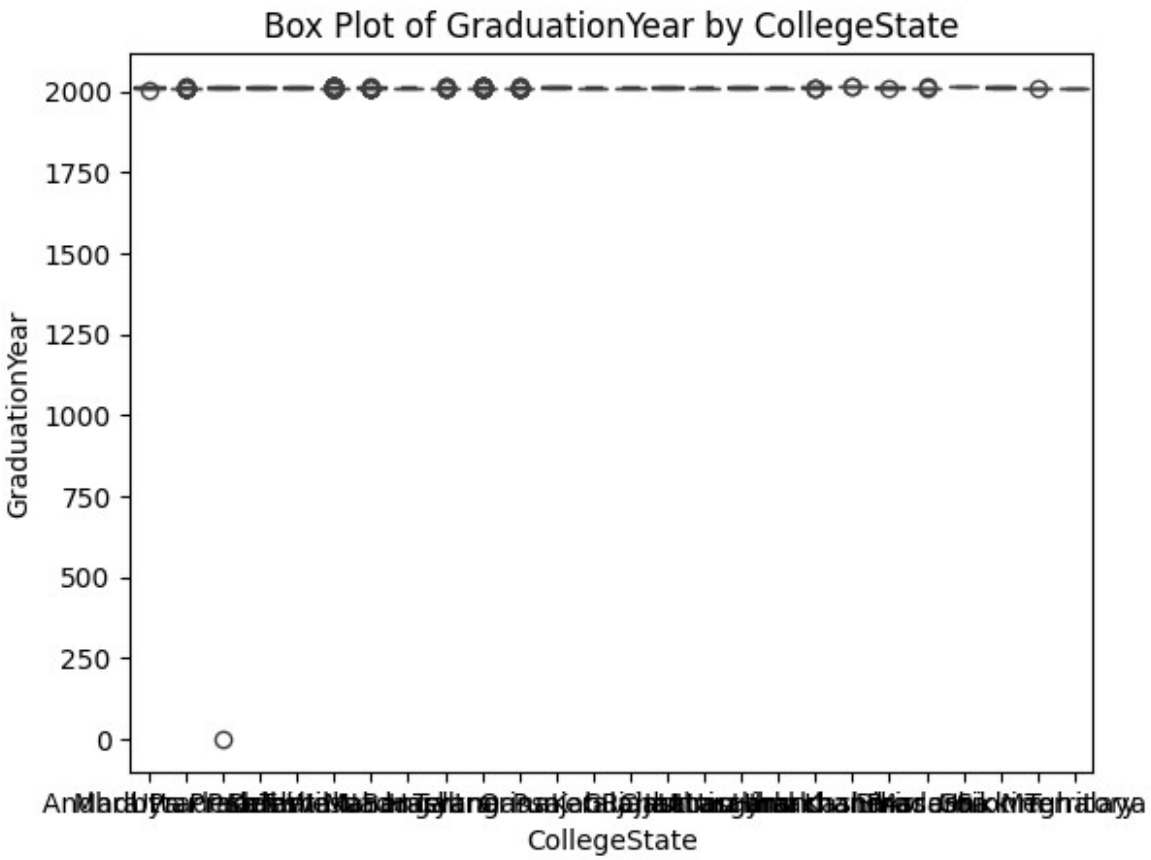

Box Plot of 10percentage by CollegeState
10percentage
100
90
80
70
60
50
CollegeState

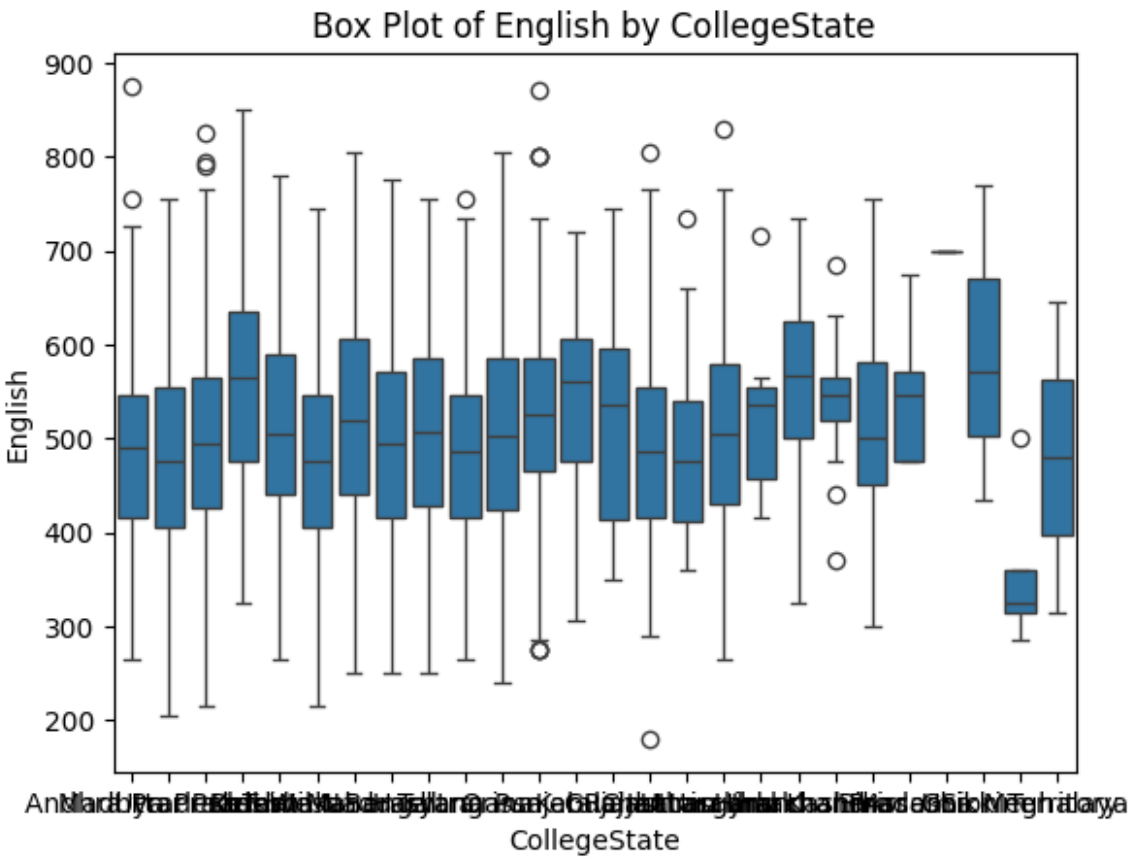

Box Plot of 12graduation by CollegeState
12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
CollegeState

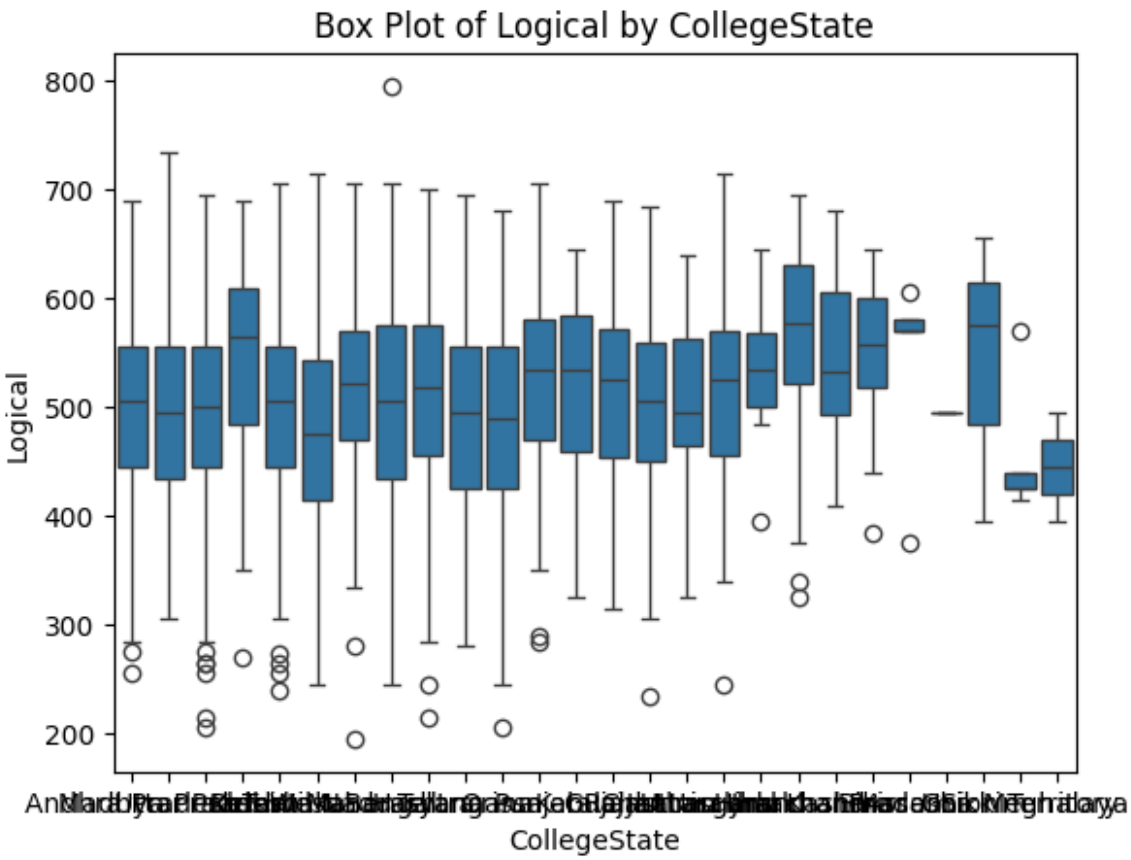

Box Plot of 12percentage by CollegeState
12percentage
100
90
80
70
60
50
40
CollegeState

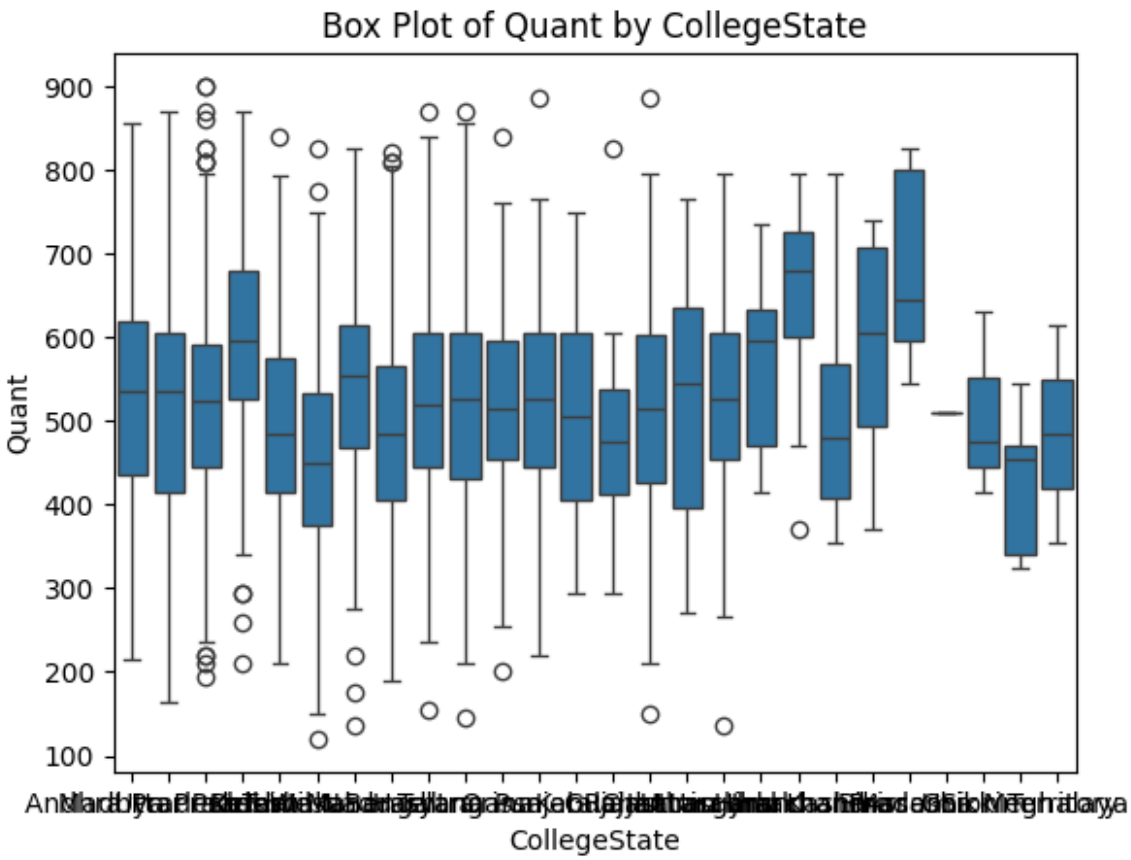

Box Plot of collegeGPA by CollegeState
collegeGPA
100
80
60
40
20
CollegeState

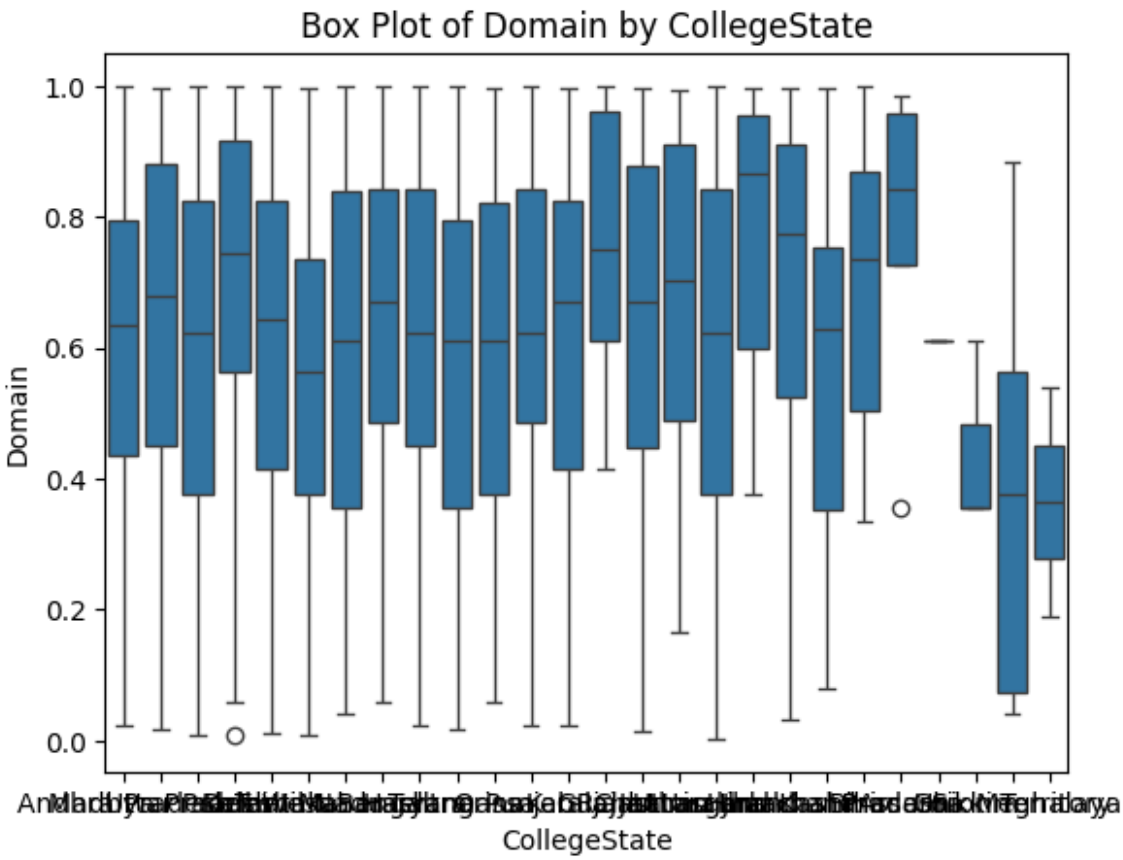

Box Plot of CollegeCityID by CollegeState
CollegeCityID
17500
15000
12500
10000
7500
5000
2500
0
CollegeState

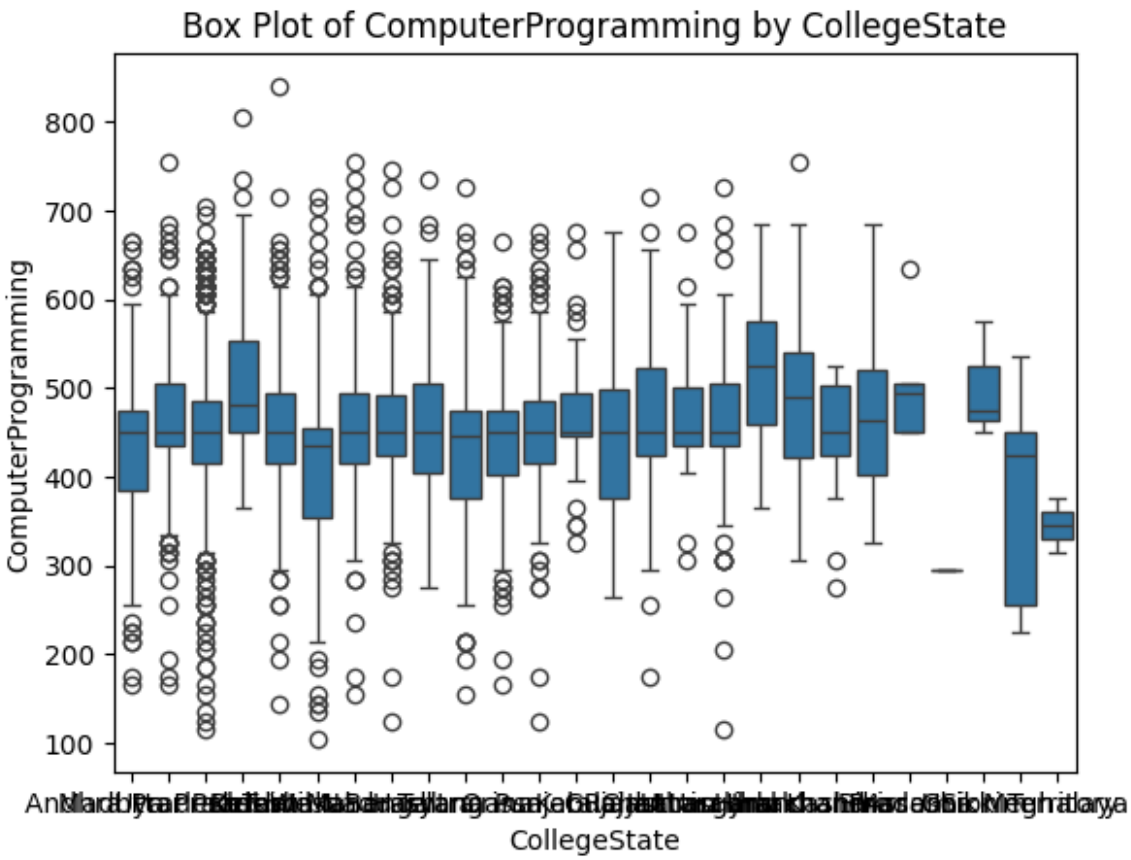

Box Plot of GraduationYear by CollegeState
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
CollegeState

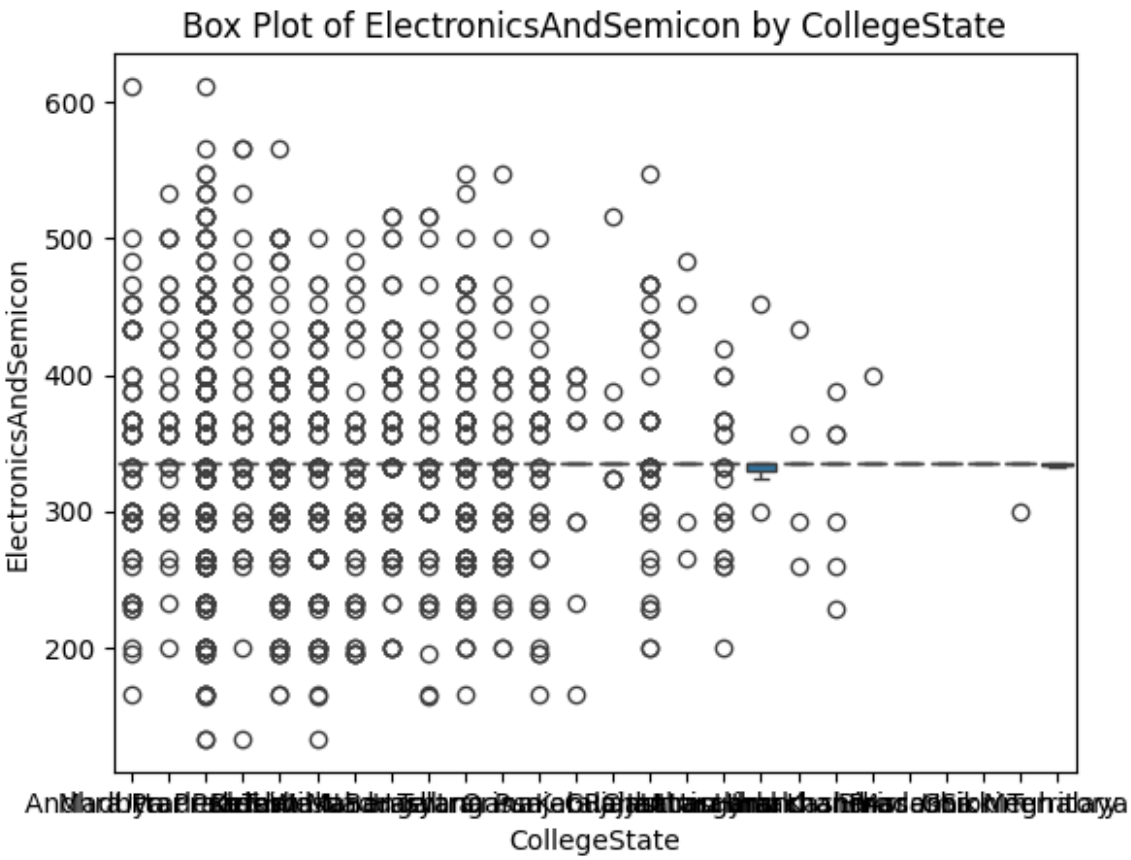

Box Plot of English by CollegeState
English
900
800
700
600
500
400
300
200
CollegeState

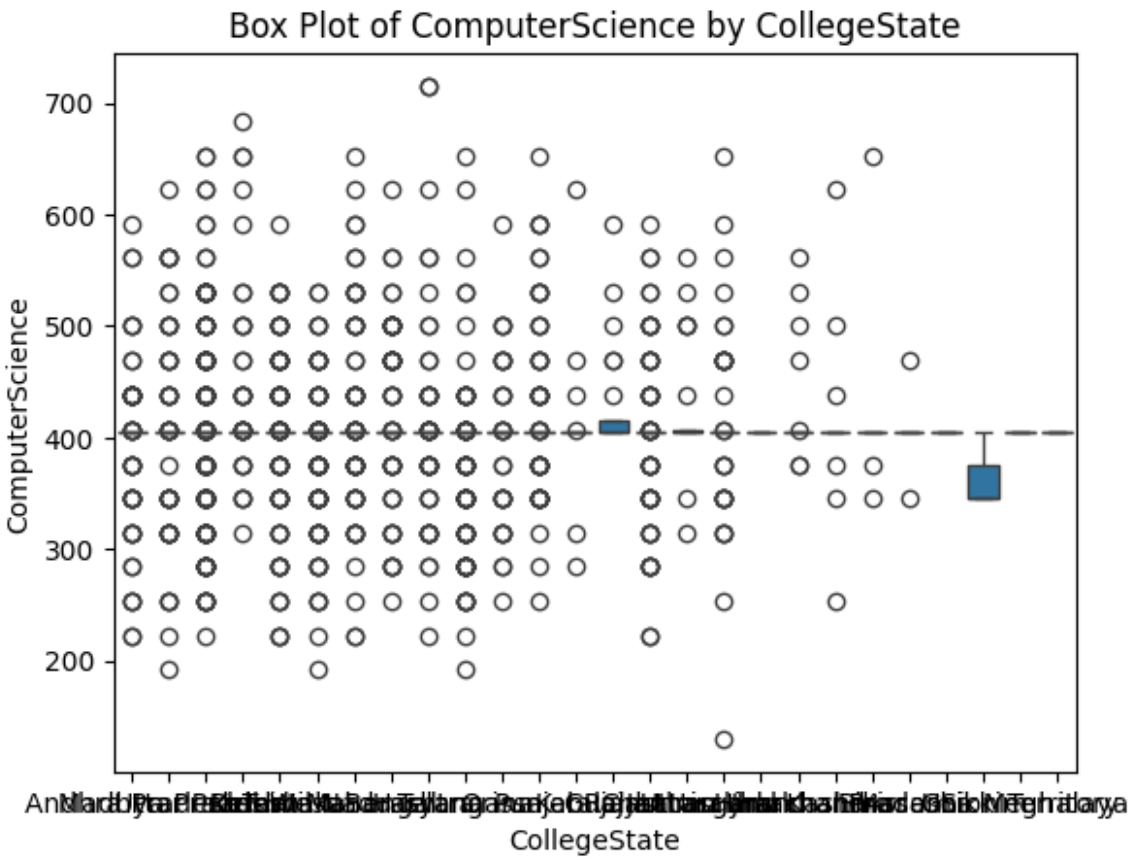

Box Plot of Logical by CollegeState
Logical
800
700
600
500
400
300
200
CollegeState

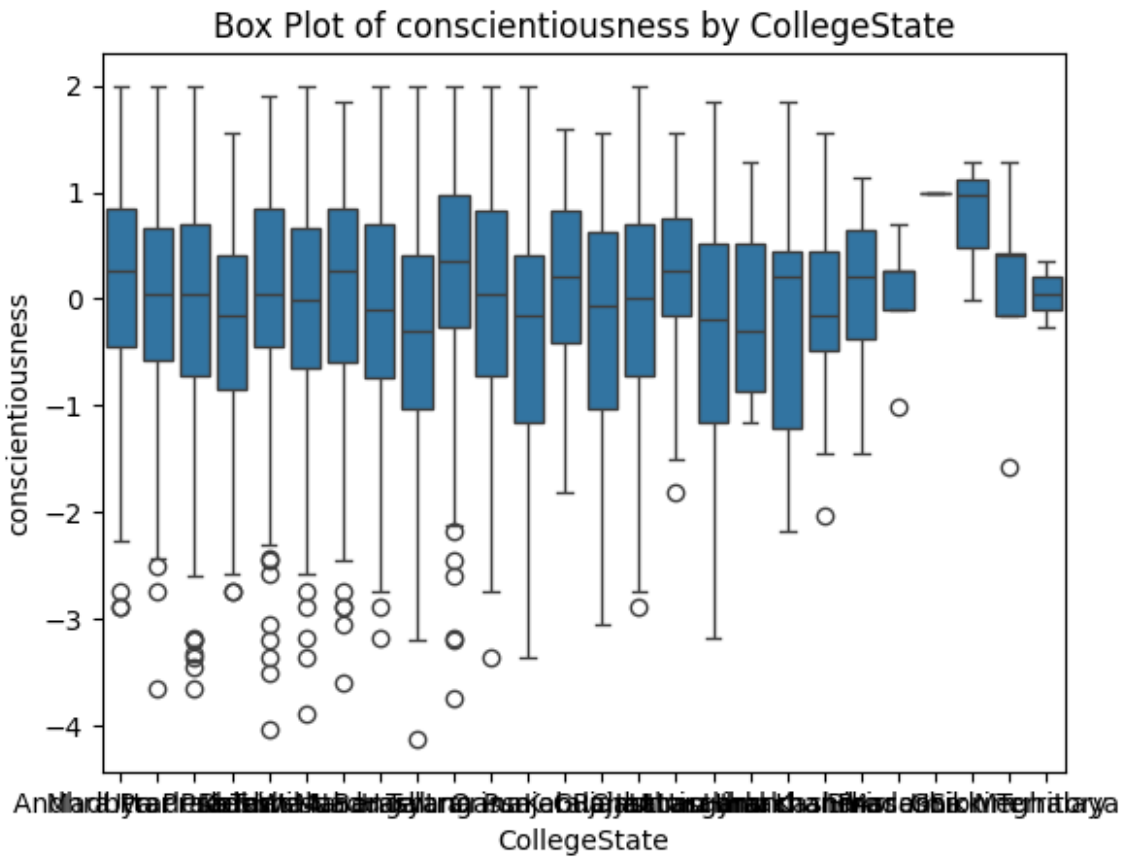

Box Plot of Quant by CollegeState
Quant
900
800
700
600
500
400
300
200
100
CollegeState

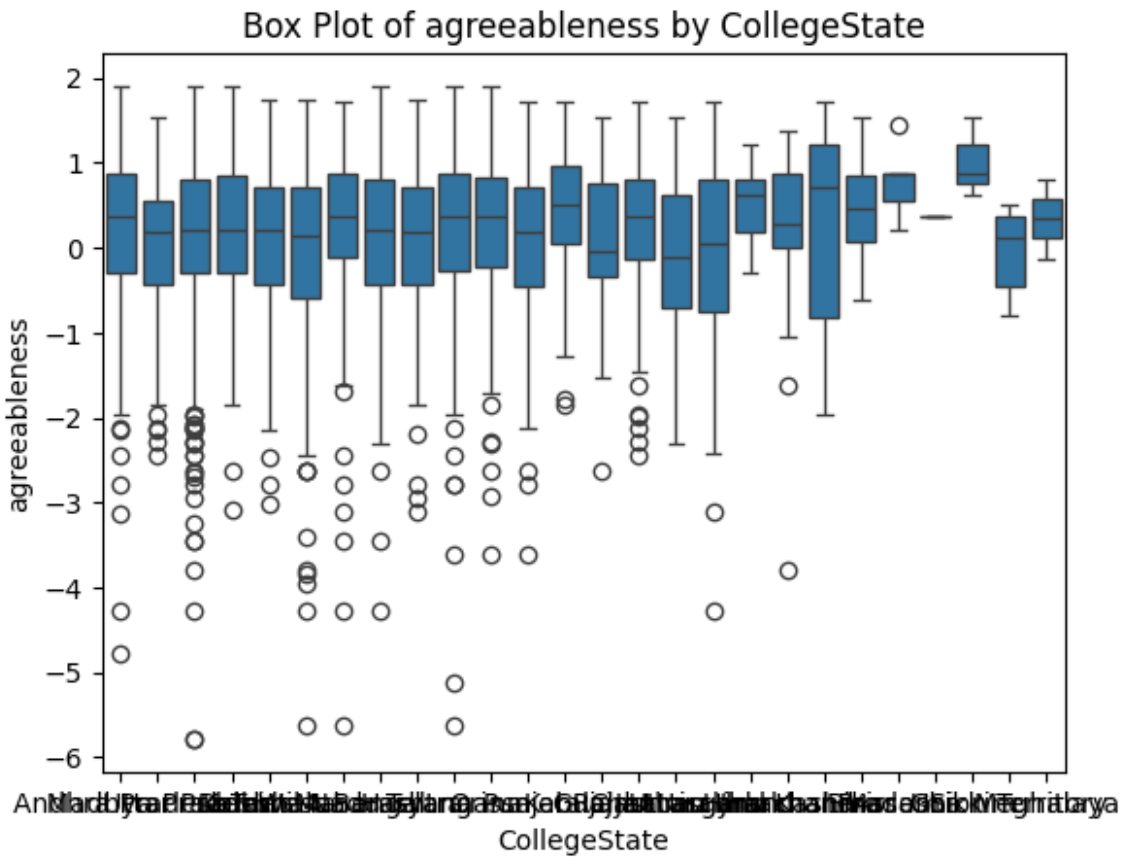

Box Plot of Domain by CollegeState
Domain
1.0
0.8
0.6
0.4
0.2
0.0
CollegeState

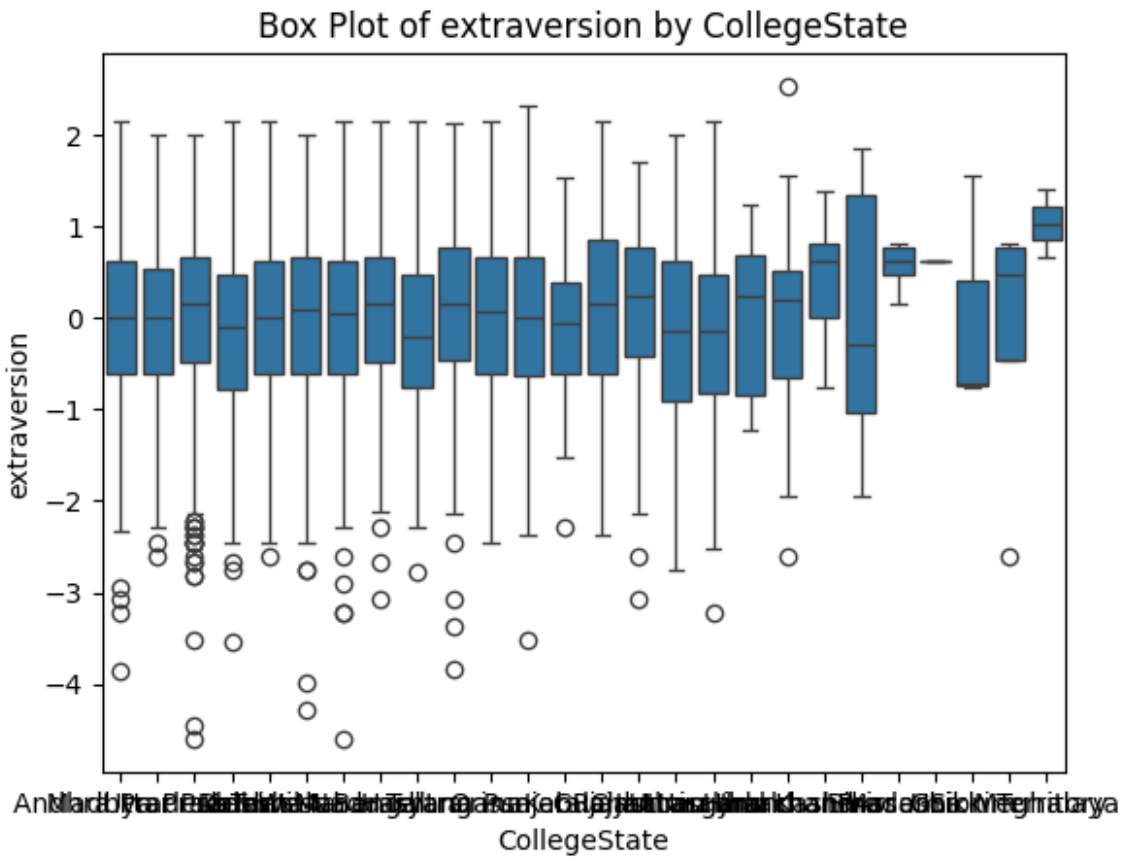

Box Plot of ComputerProgramming by CollegeState
ComputerProgramming
800
700
600
500
400
300
200
100
CollegeState

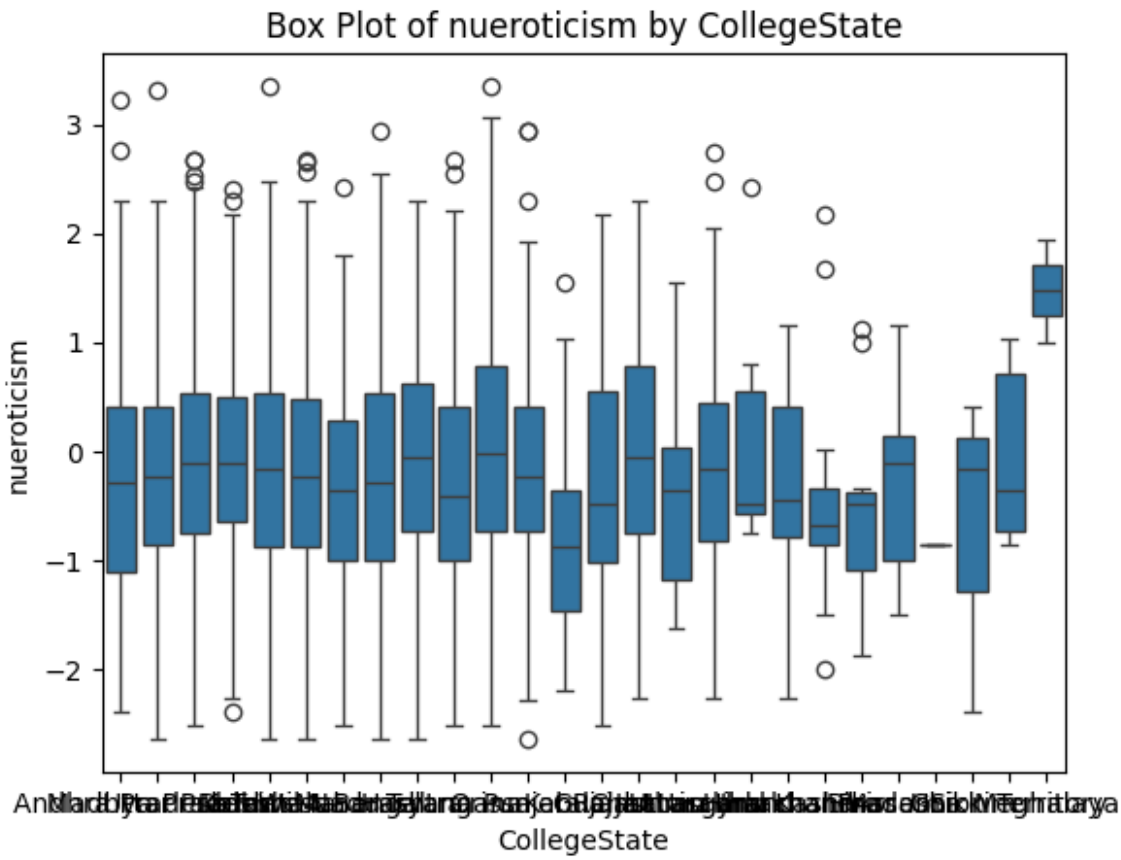

Box Plot of ElectronicsAndSemicon by CollegeState
ElectronicsAndSemicon
600
500
400
300
200
CollegeState

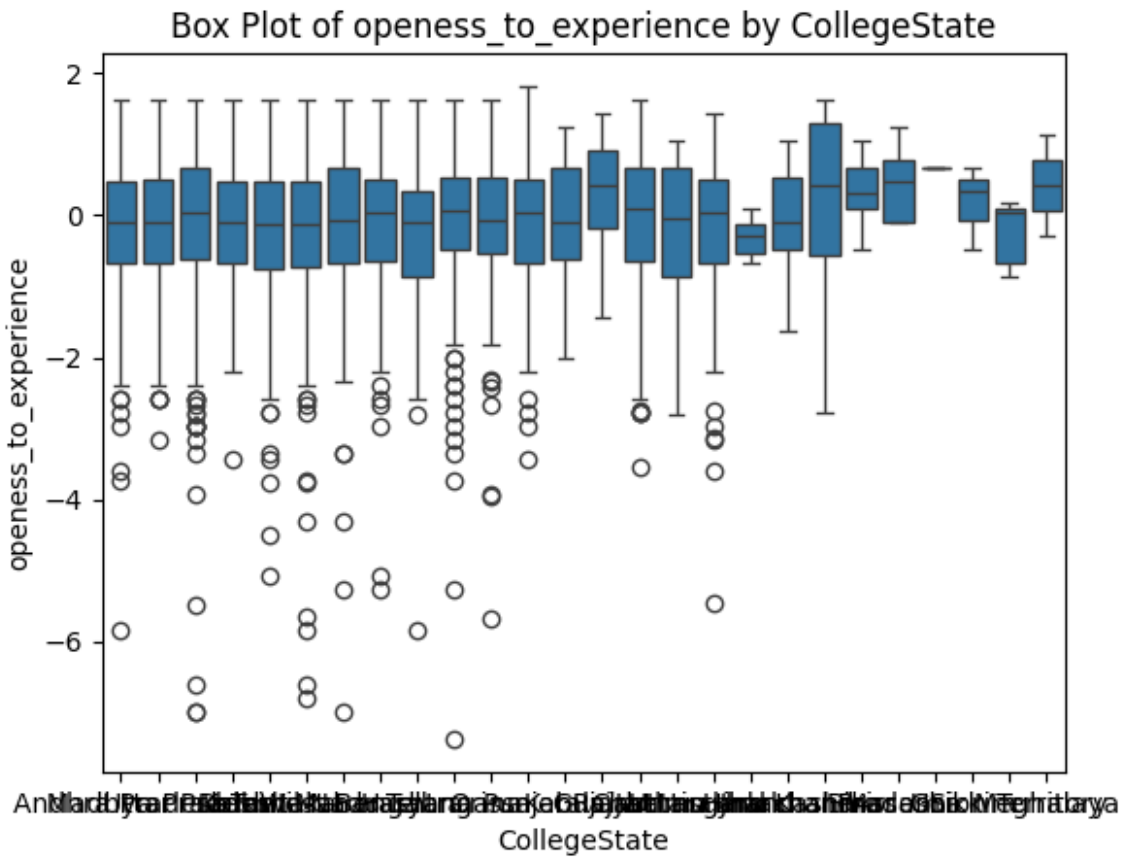

Box Plot of ComputerScience by CollegeState
ComputerScience
700
600
500
400
300
200
CollegeState

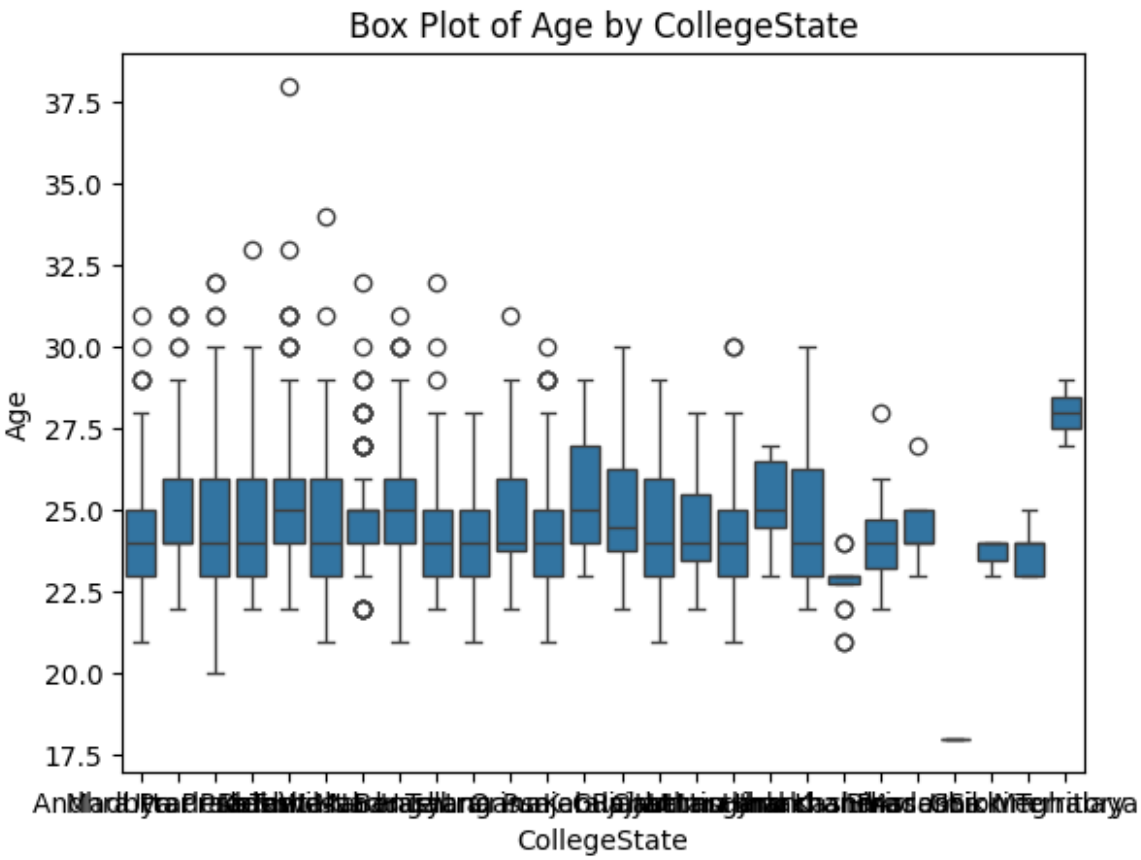

Box Plot of conscientiousness by CollegeState
conscientiousness
2
1
0
−1
−2
−3
−4
CollegeState


Box Plot of agreeableness by CollegeState
agreeableness
2
1
0
−1
−2
−3
−4
−5
−6
CollegeState

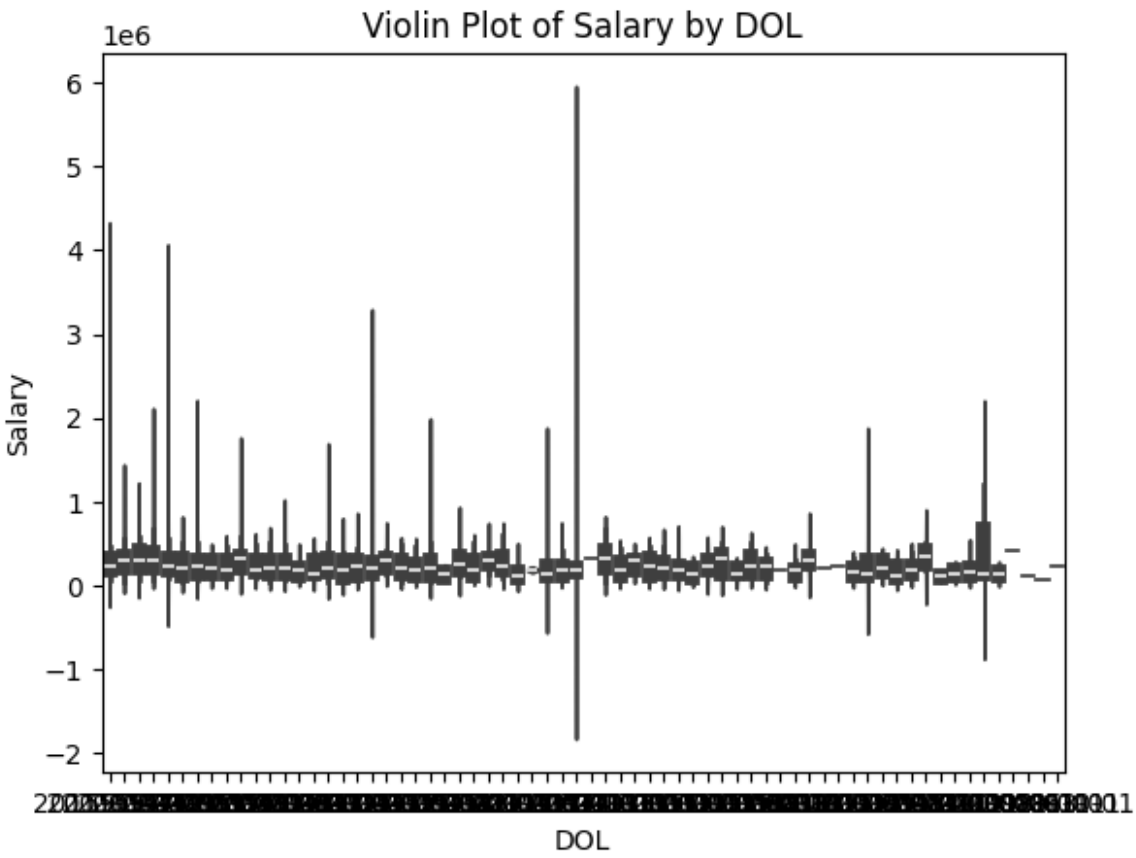

Box Plot of extraversion by CollegeState
extraversion
2
1
0
−1
−2
−3
−4
CollegeState

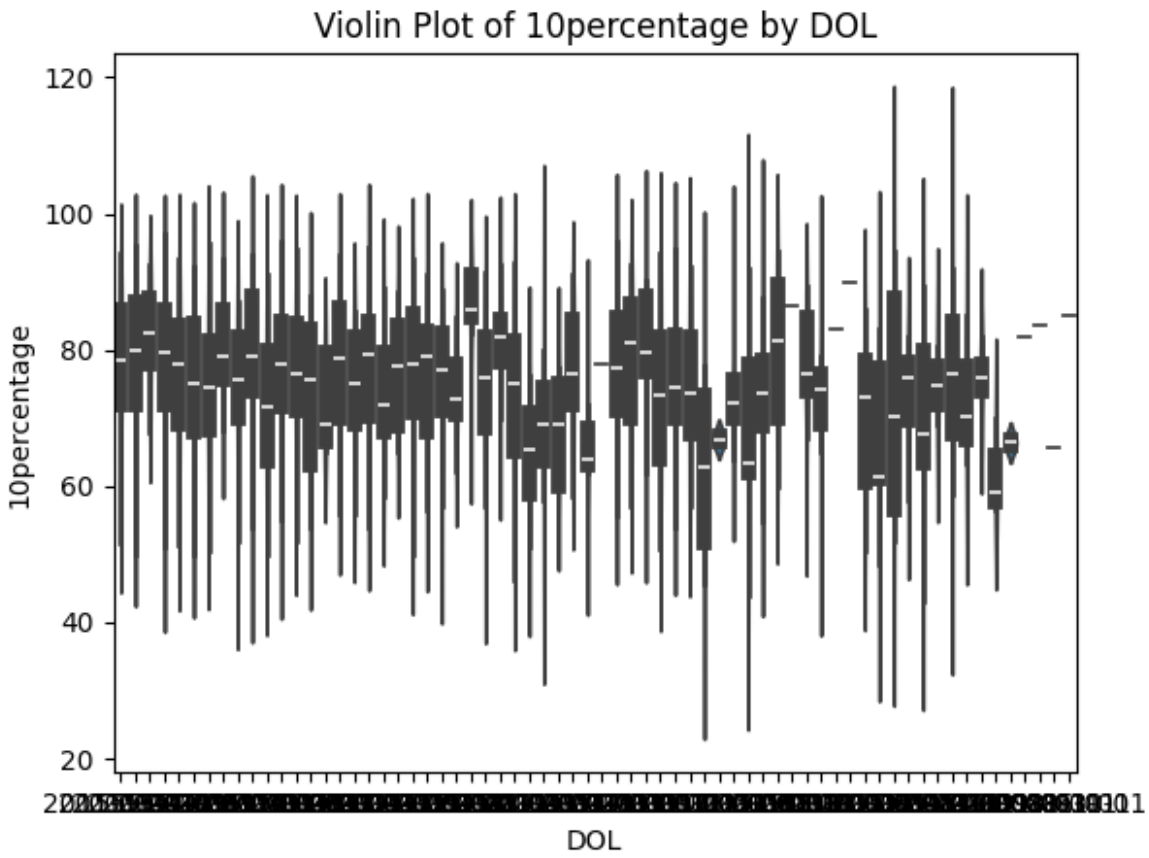

Box Plot of nueroticism by CollegeState
nueroticism
3
2
1
0
−1
−2
CollegeState

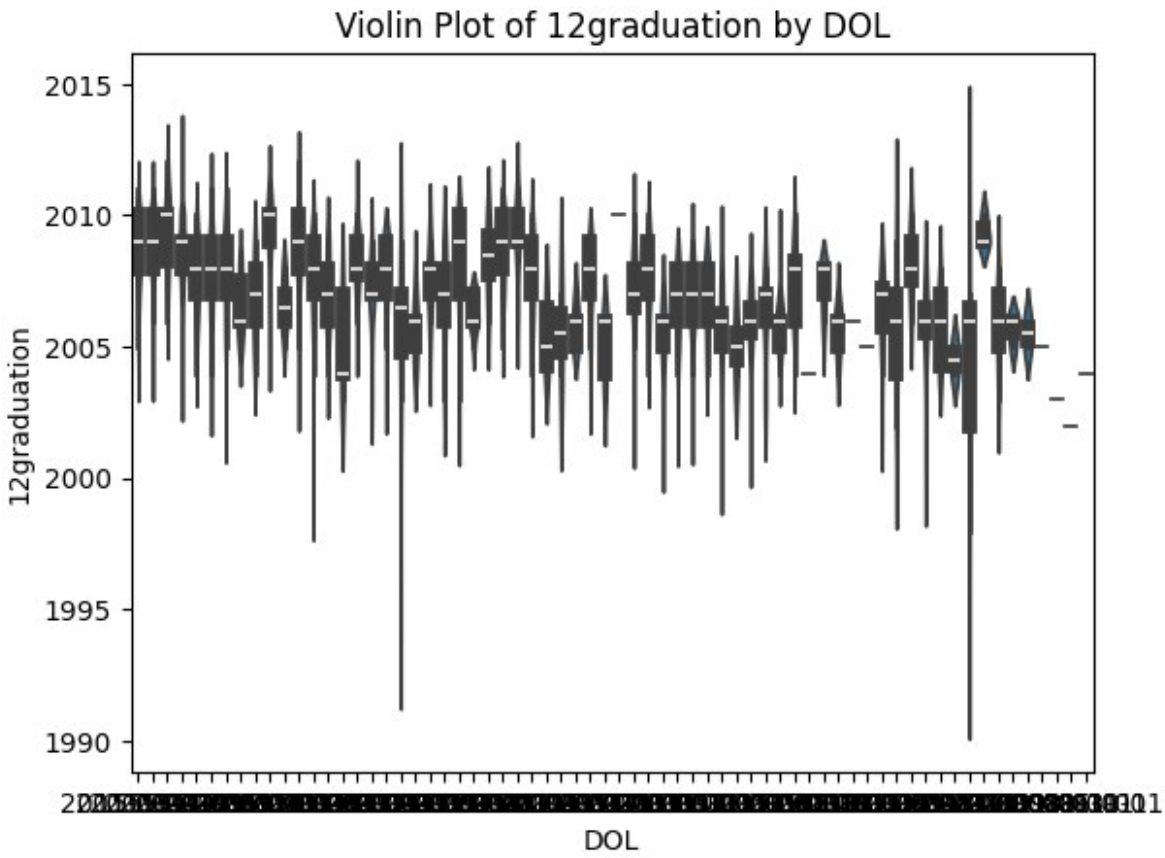

Box Plot of openness_to_experience by CollegeState
openness_to_experience
2
0
−2
−4
−6
CollegeState

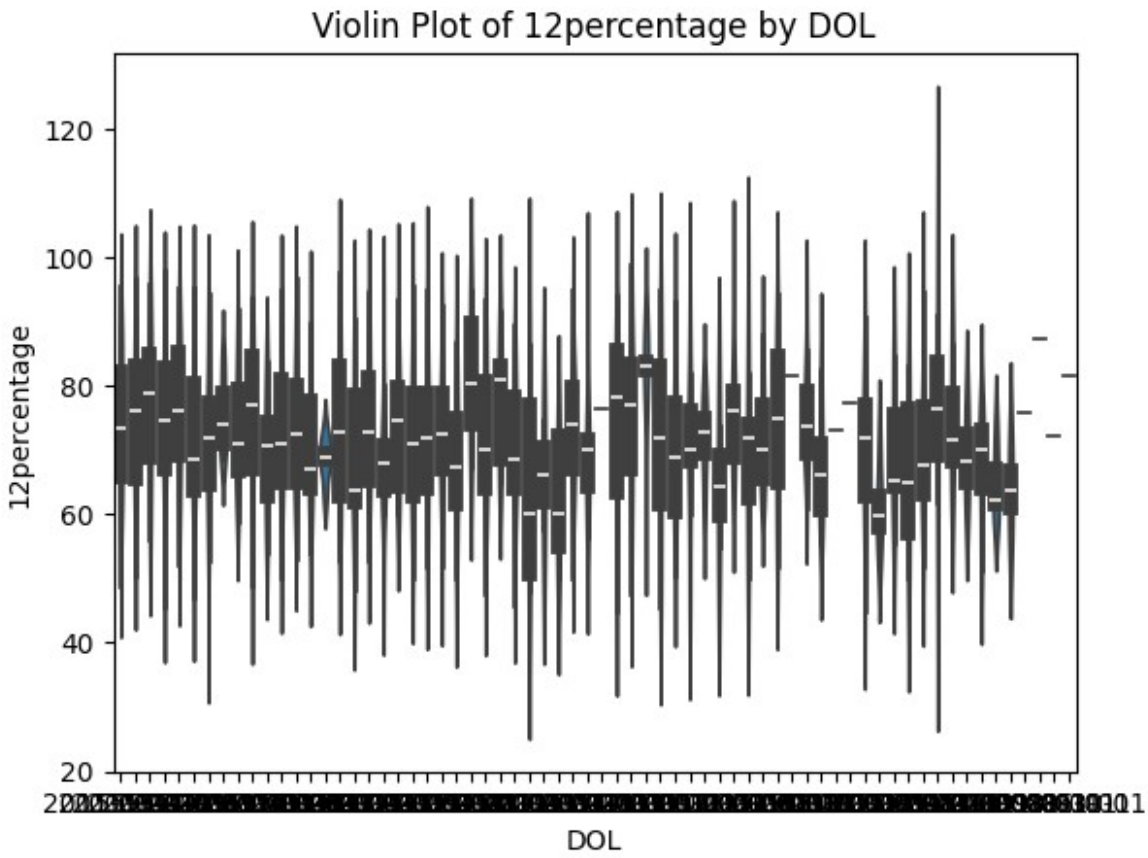


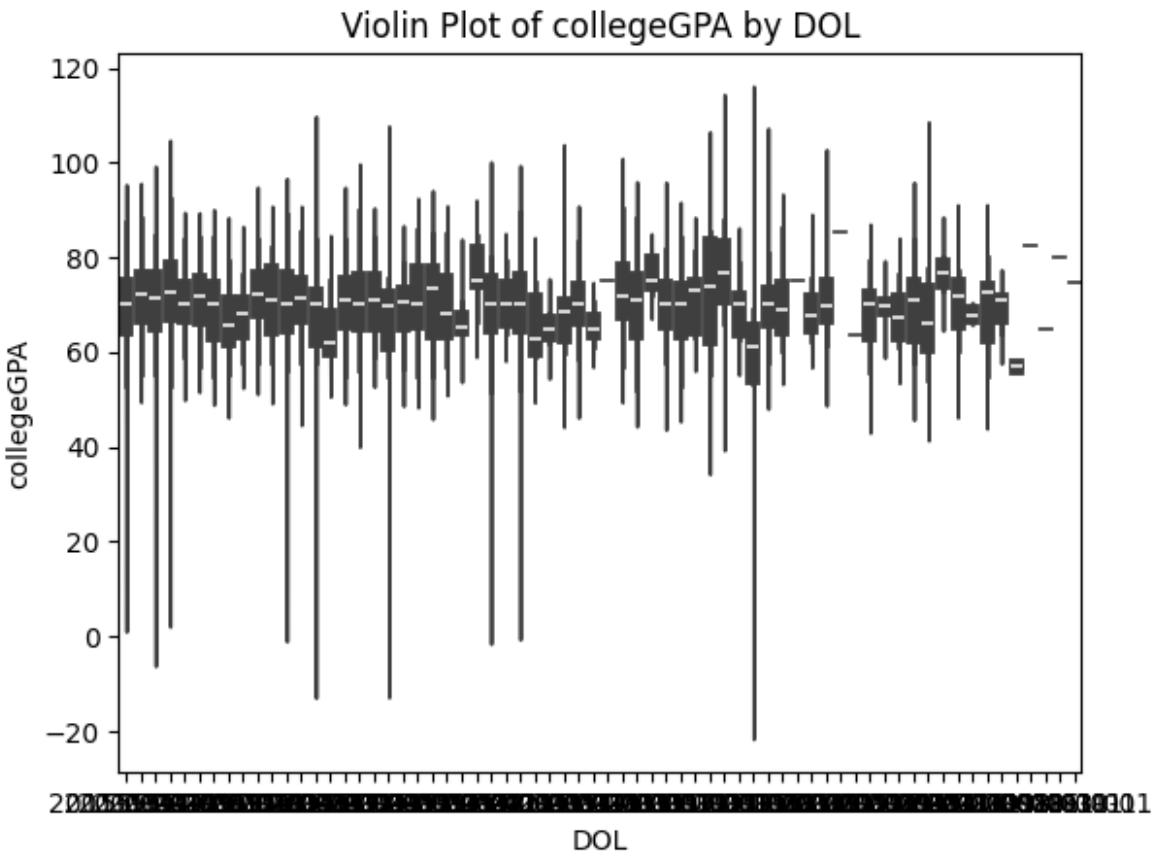

```
# Violin Plots for Categorical vs Numerical Columns
for col in categorical_columns:
    for num_col in numerical_columns:
        plt.figure()
        sns.violinplot(x=df[col], y=df[num_col])
        plt.title(f"Violin Plot of {num_col} by {col}")
        plt.show()
```

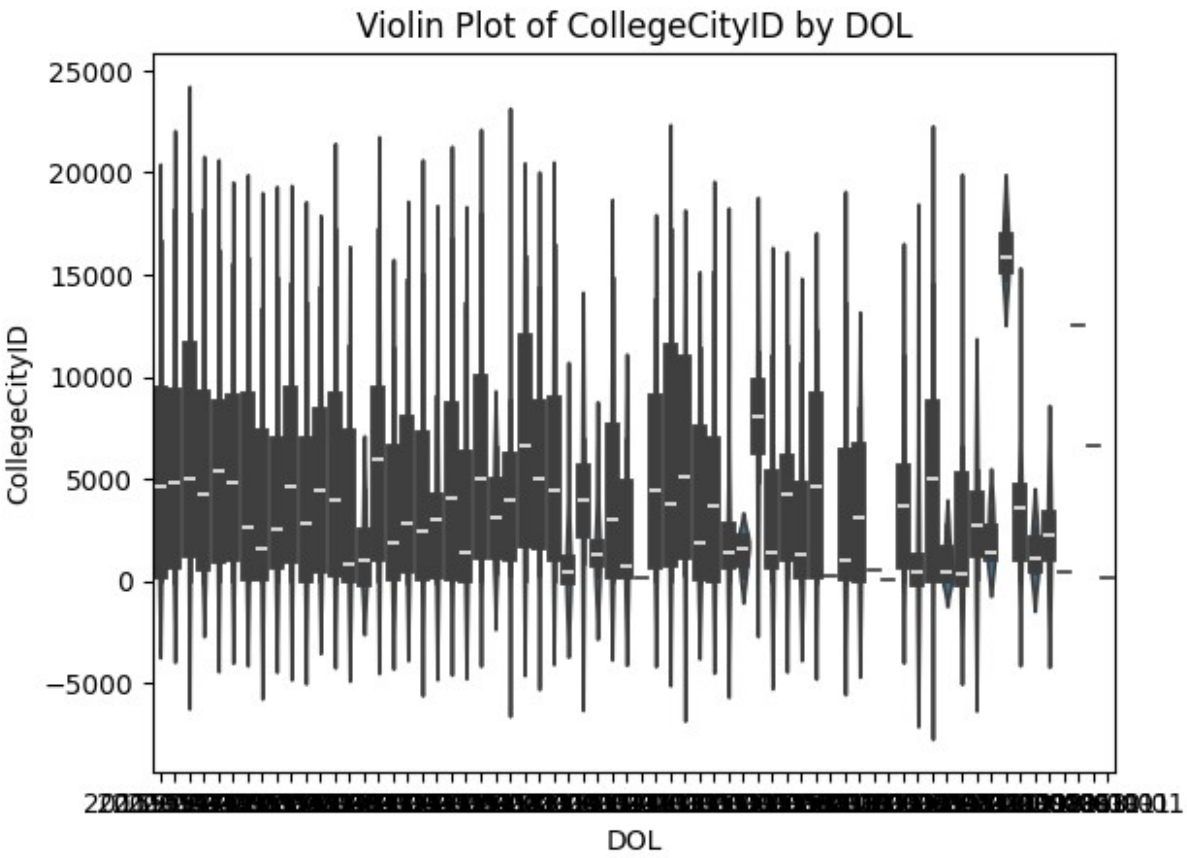

Violin Plot of Salary by DOL
1e6
6
5
4
3
2
1
0
−1
−2
Salary
DOL

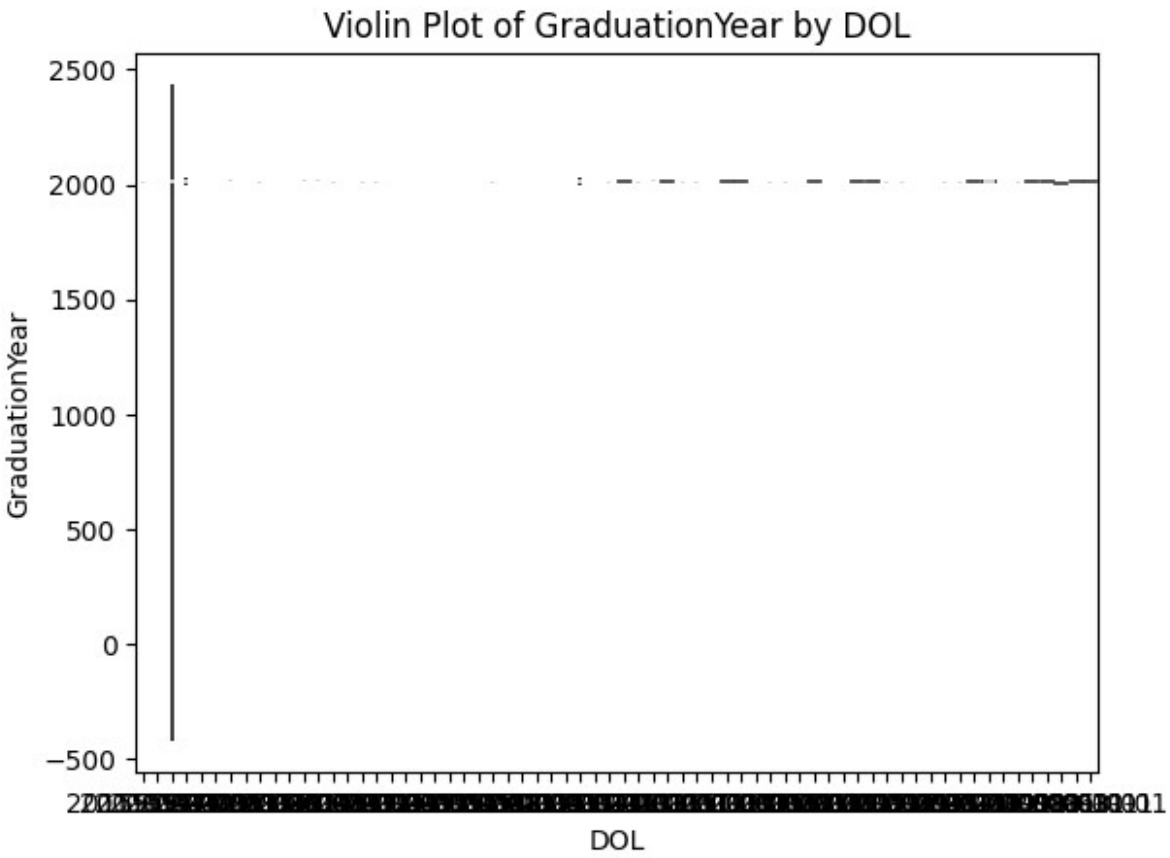

Violin Plot of 10percentage by DOL
120
100
80
60
40
20
10percentage
DOL

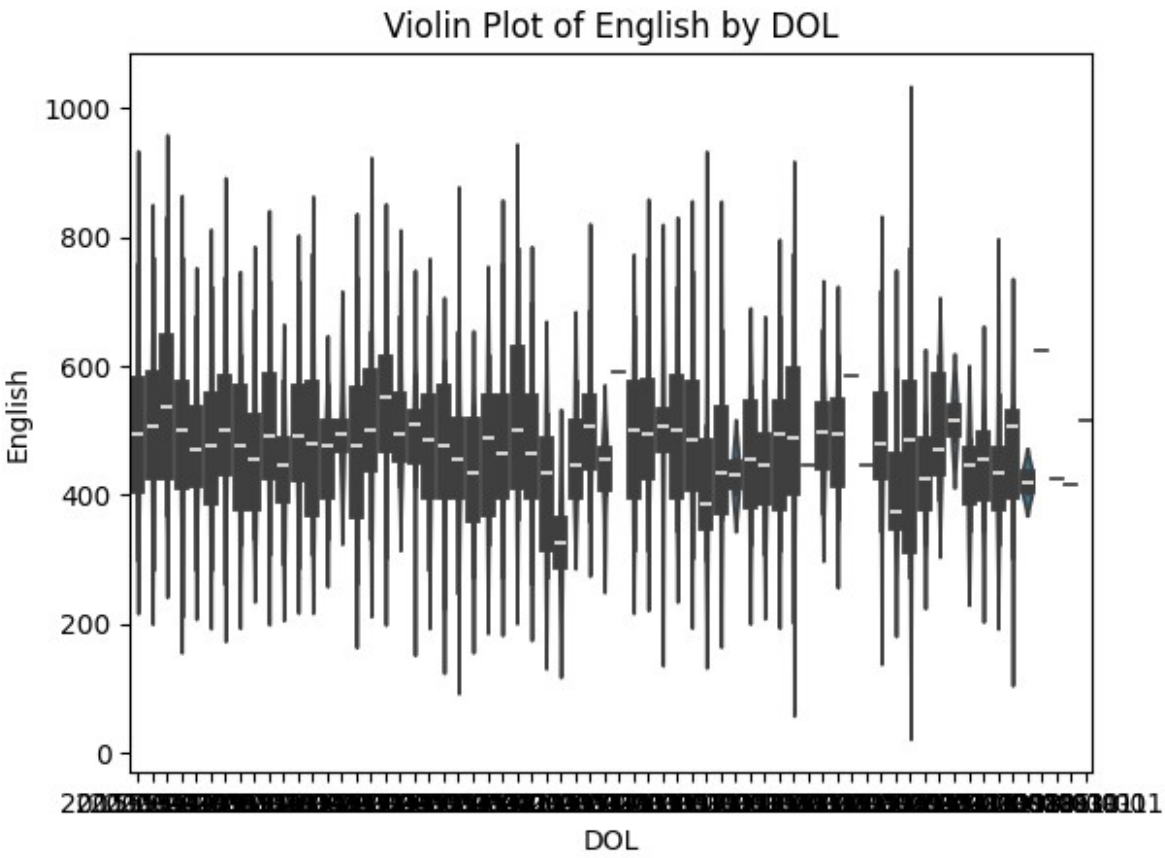

Violin Plot of 12graduation by DOL
2015
2010
2005
2000
1995
1990
12graduation
DOL

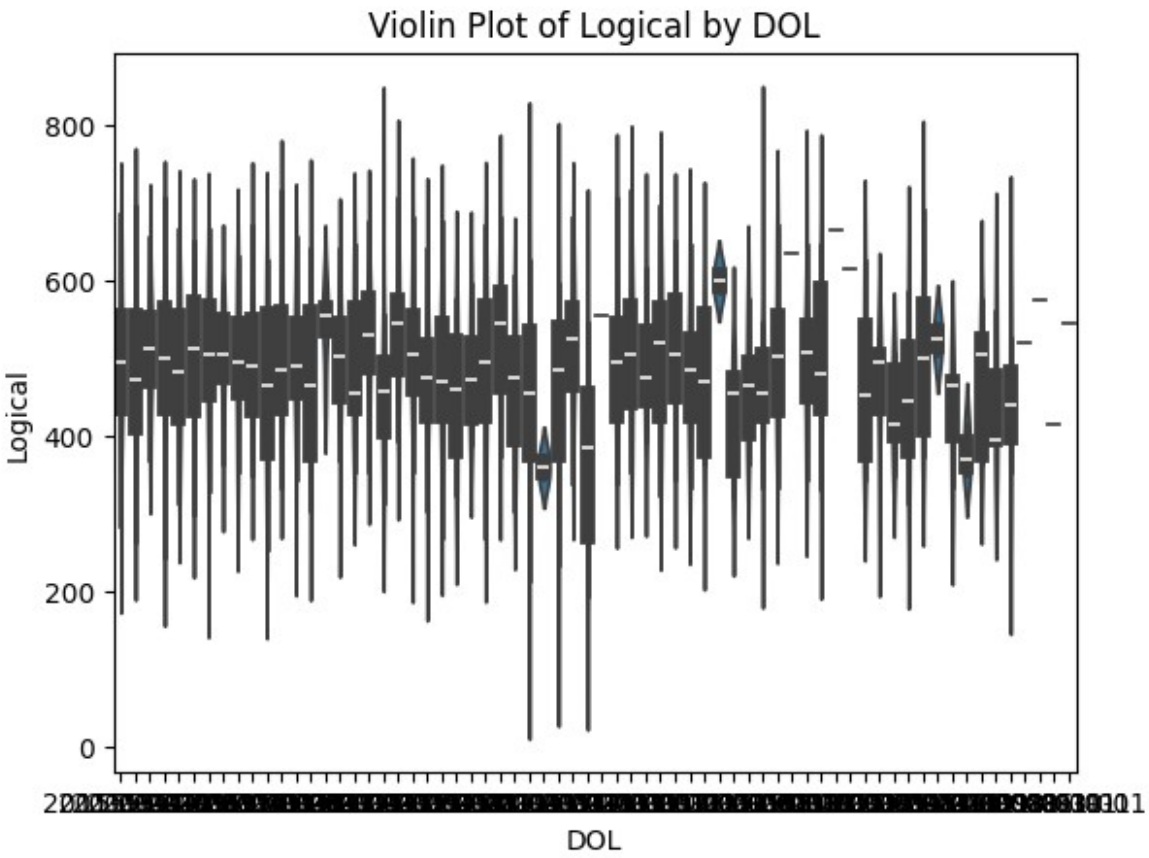

Violin Plot of 12percentage by DOL
12percentage
120
100
80
60
40
20
DOL

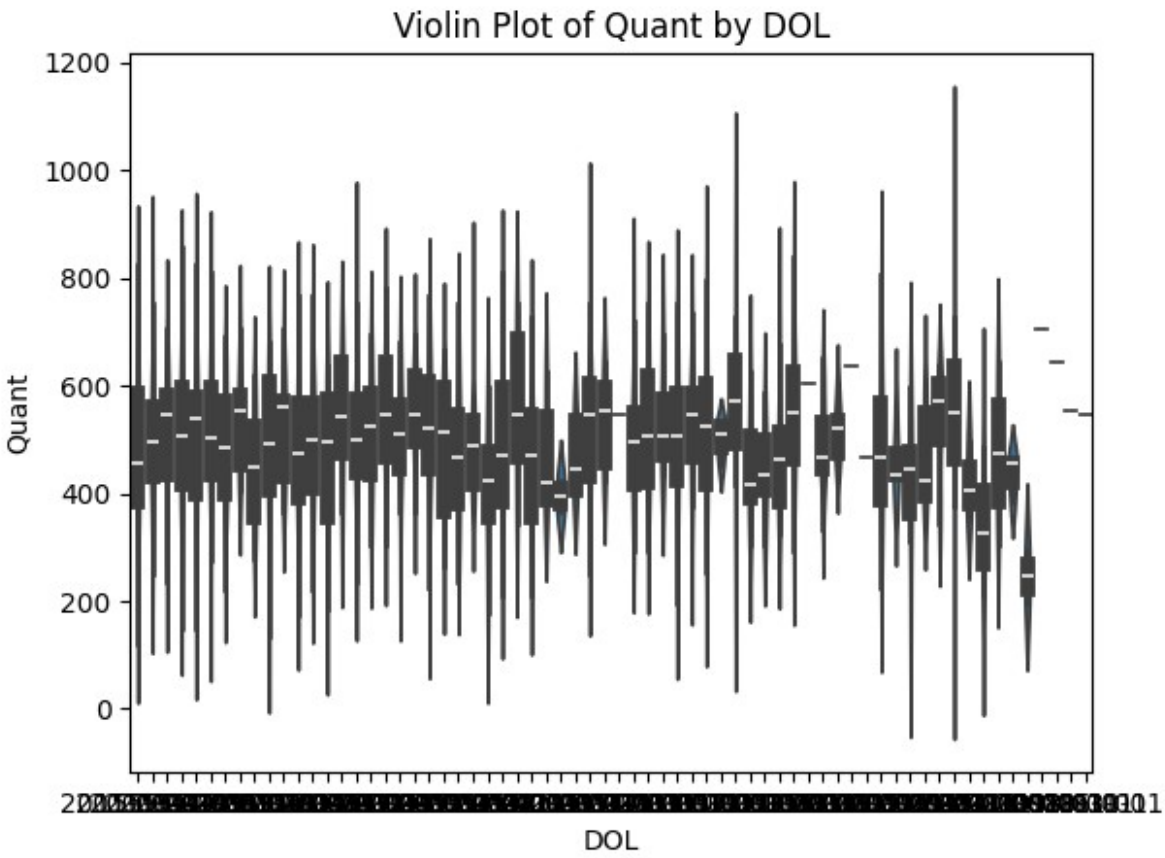

Violin Plot of collegeGPA by DOL
collegeGPA
120
100
80
60
40
20
0
−20
DOL

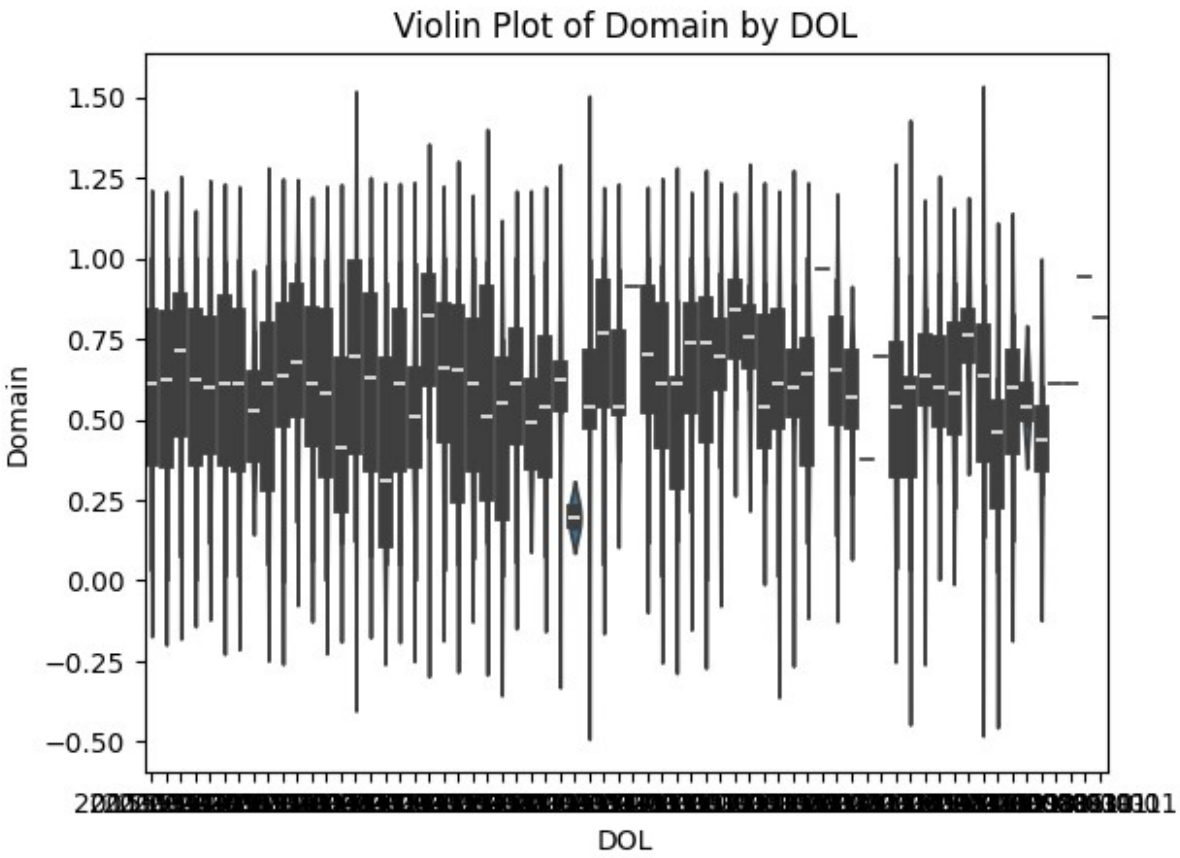

Violin Plot of CollegeCityID by DOL
25000
20000
15000
10000
5000
0
−5000
CollegeCityID
DOL

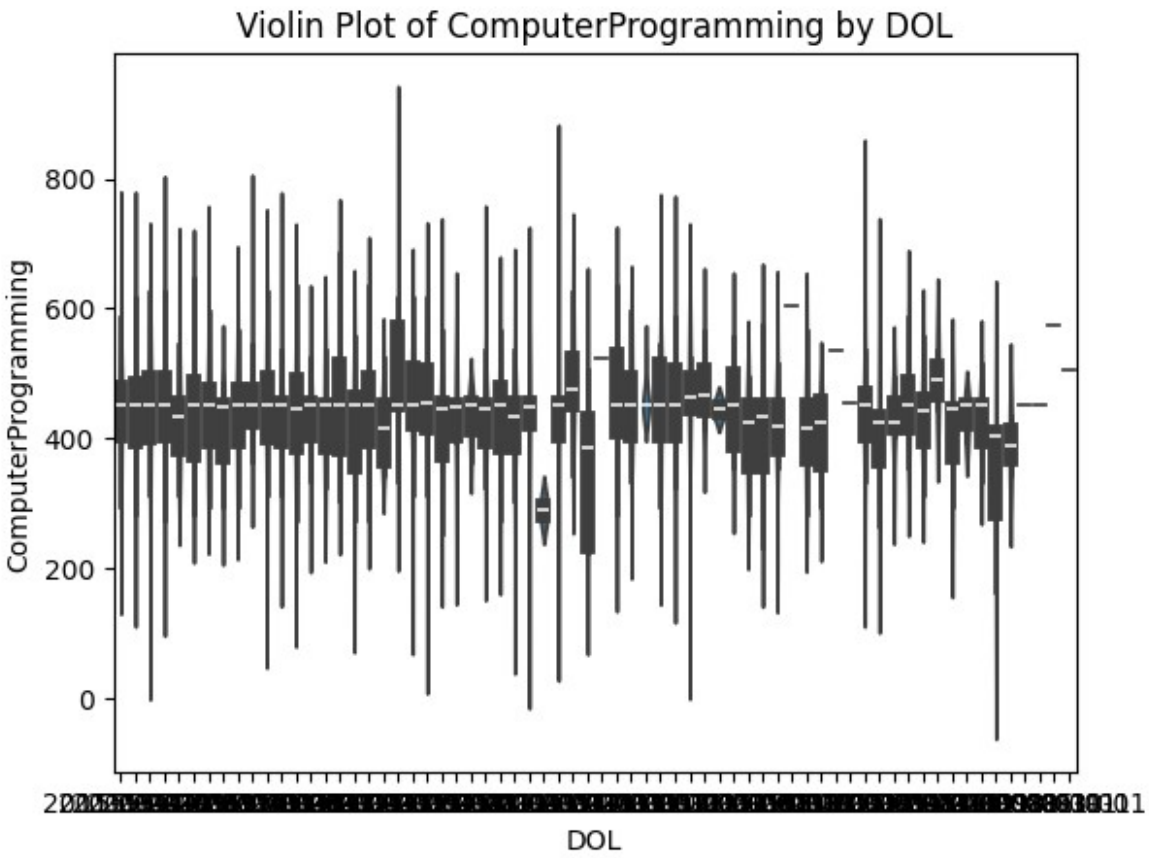

Violin Plot of GraduationYear by DOL
2500
2000
1500
1000
500
0
−500
GraduationYear
DOL

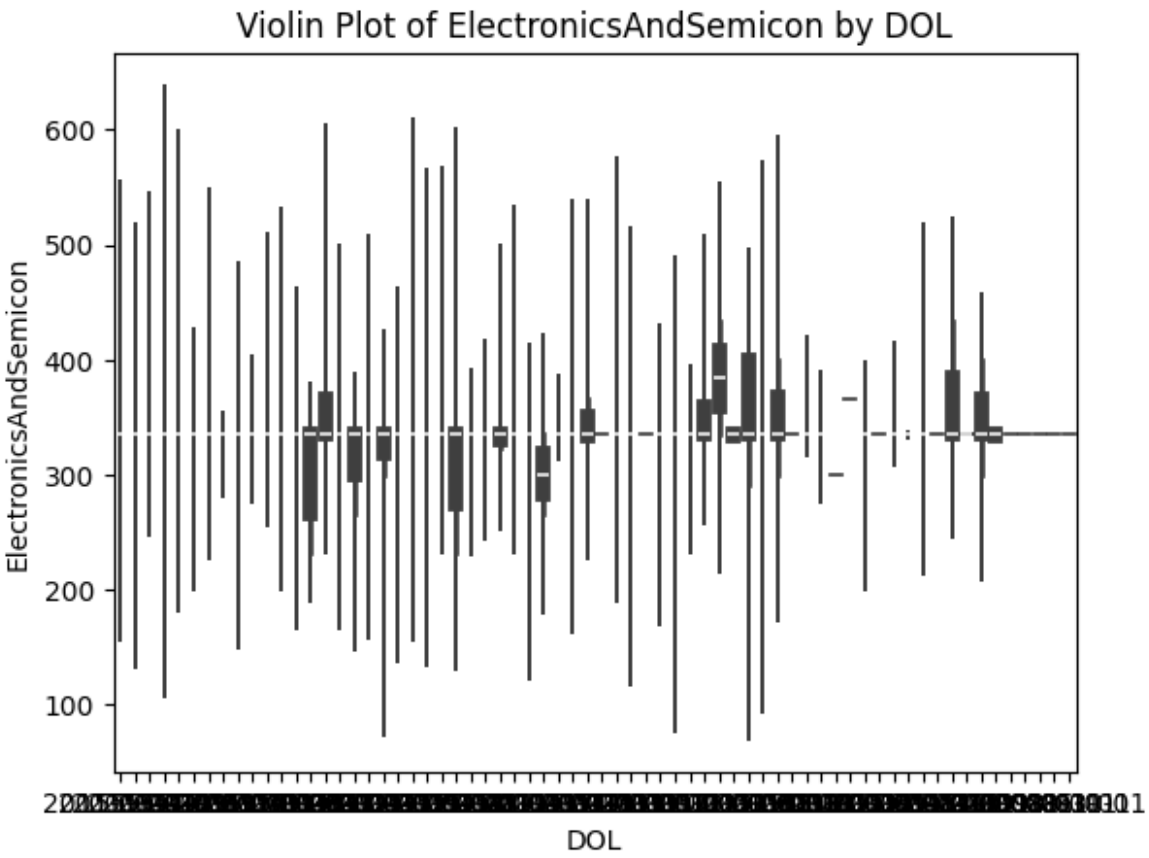

Violin Plot of English by DOL
English
1000
800
600
400
200
0
DOL

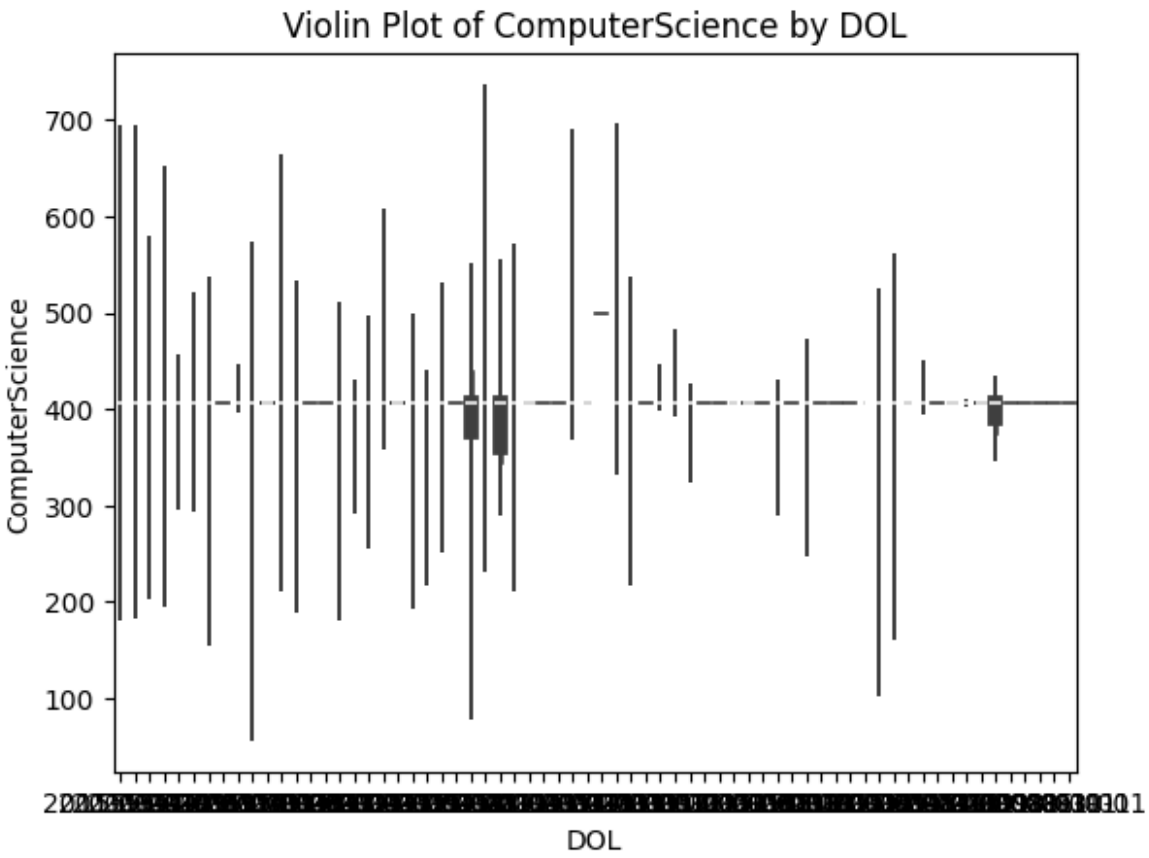

Violin Plot of Logical by DOL
Logical
800
600
400
200
0
DOL

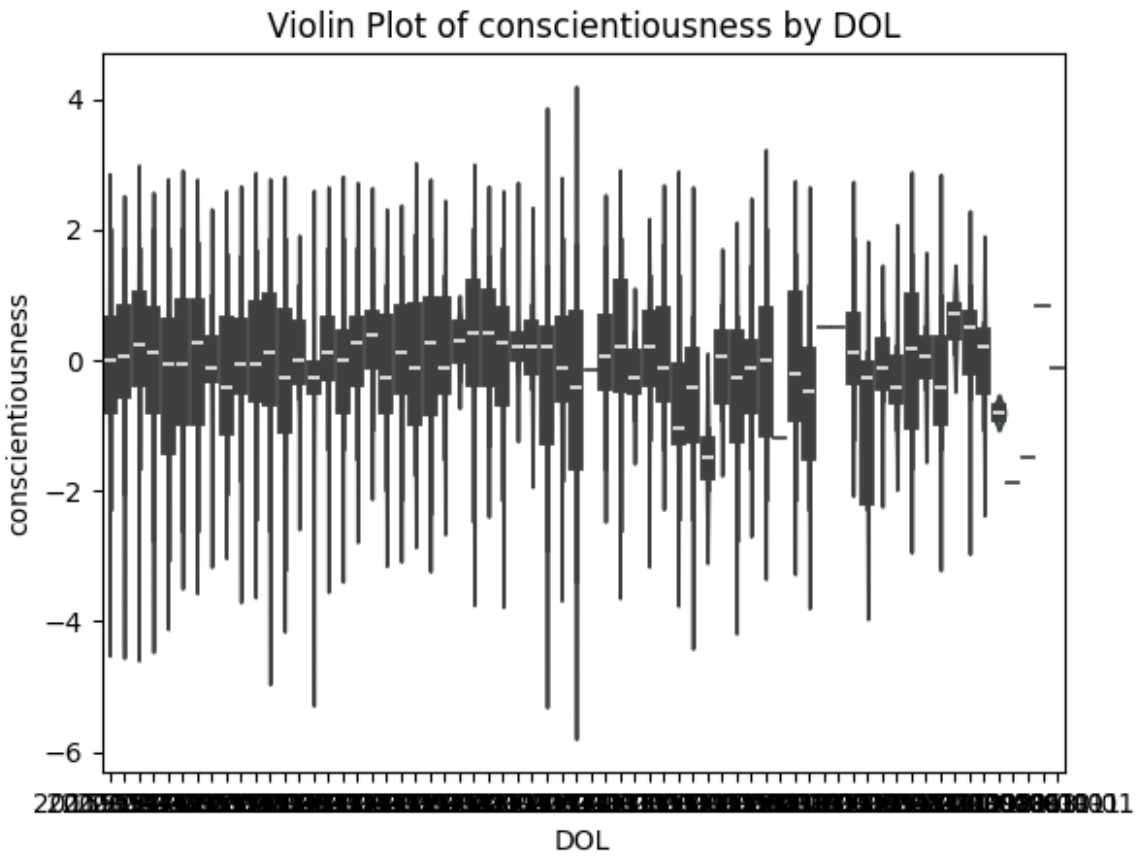

Violin Plot of Quant by DOL
1200
1000
800
600
400
200
0
Quant
DOL

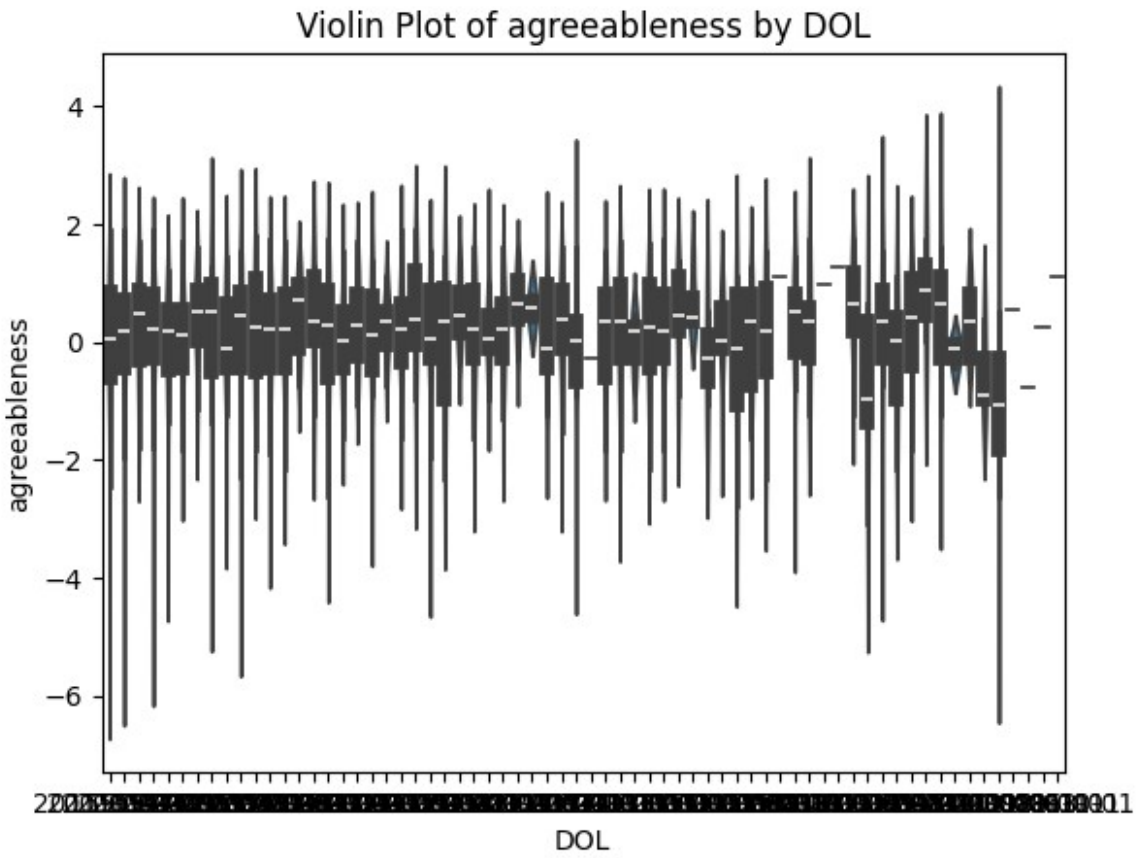

Violin Plot of Domain by DOL
Domain
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
DOL

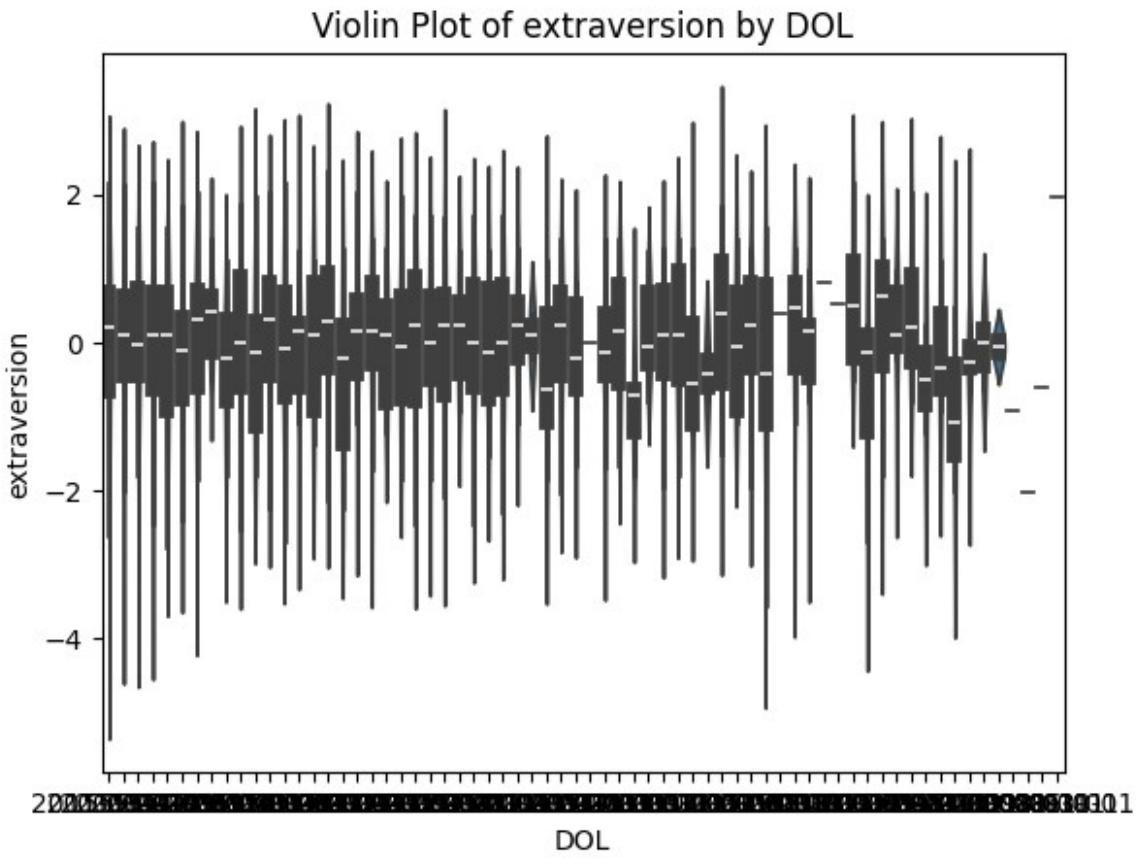

Violin Plot of ComputerProgramming by DOL
ComputerProgramming
800
600
400
200
0
DOL

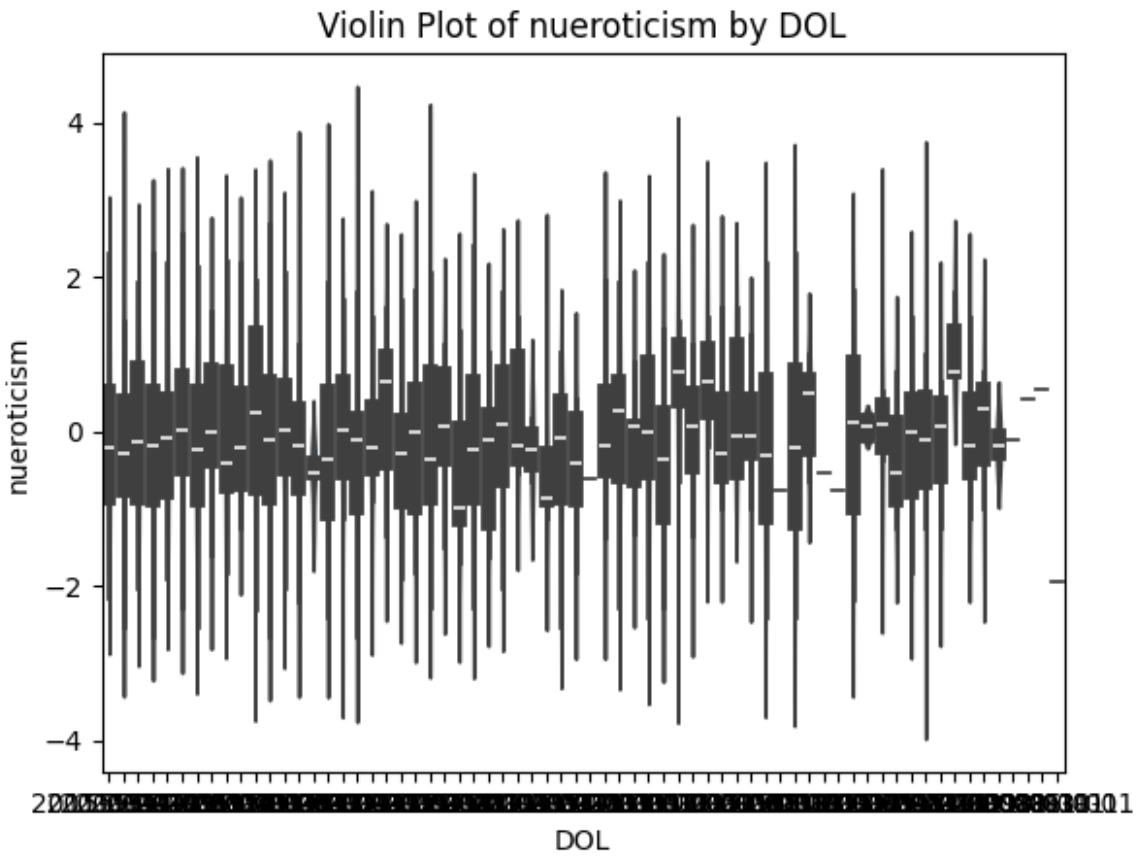

Violin Plot of ElectronicsAndSemicon by DOL
ElectronicsAndSemicon
600
500
400
300
200
100
DOL

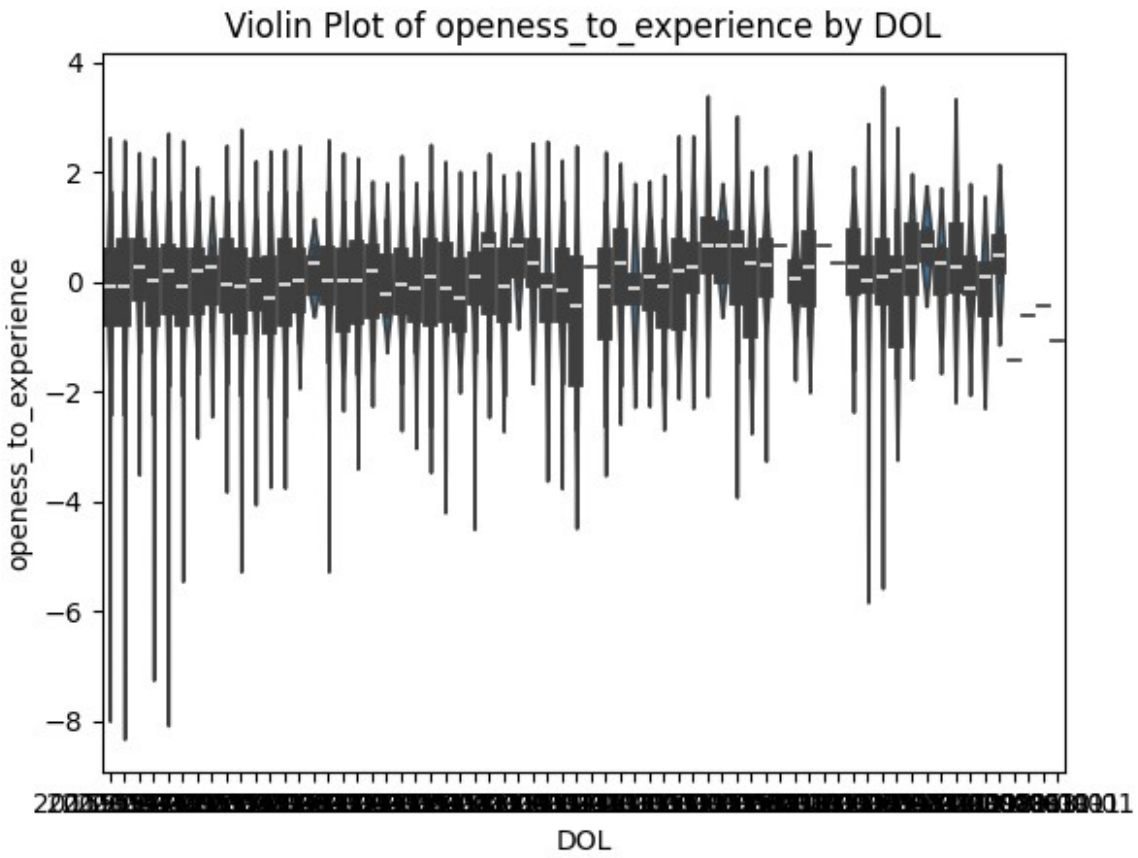

Violin Plot of ComputerScience by DOL
ComputerScience
700
600
500
400
300
200
100
DOL

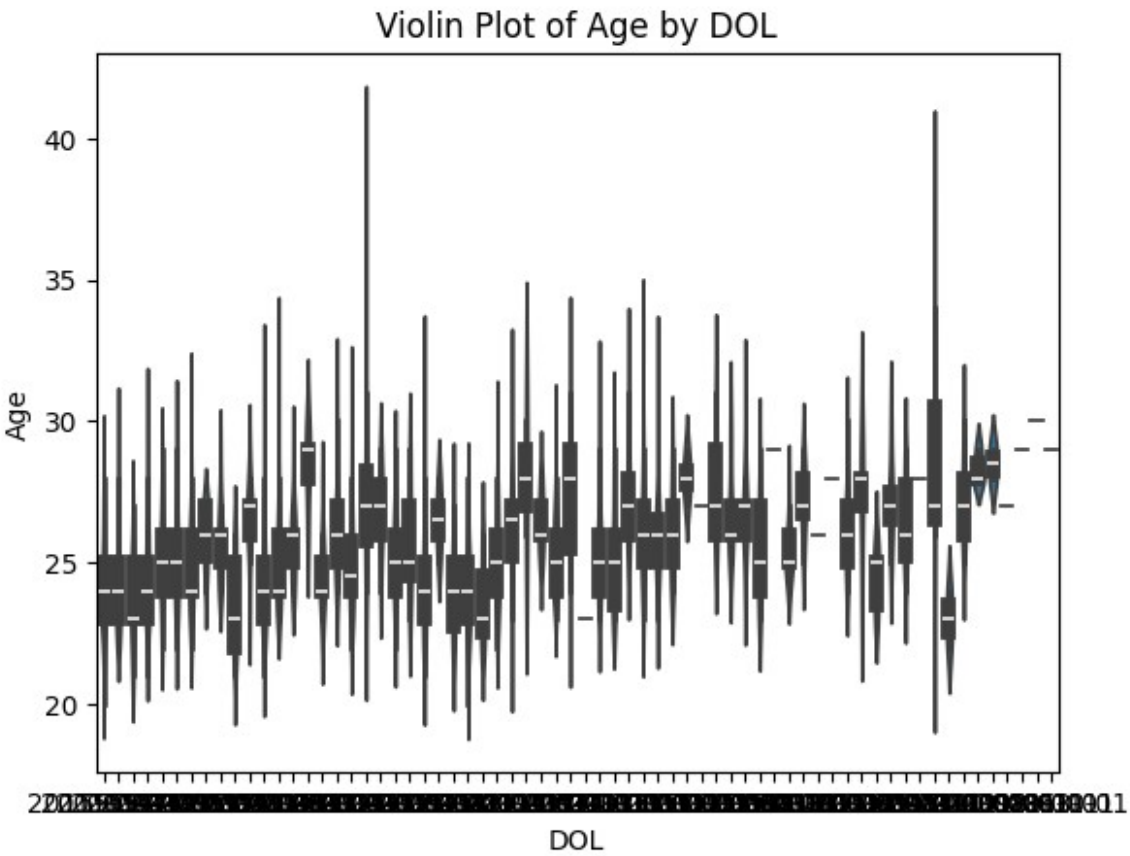

Violin Plot of conscientiousness by DOL
conscientiousness
4
2
0
−2
−4
−6
DOL

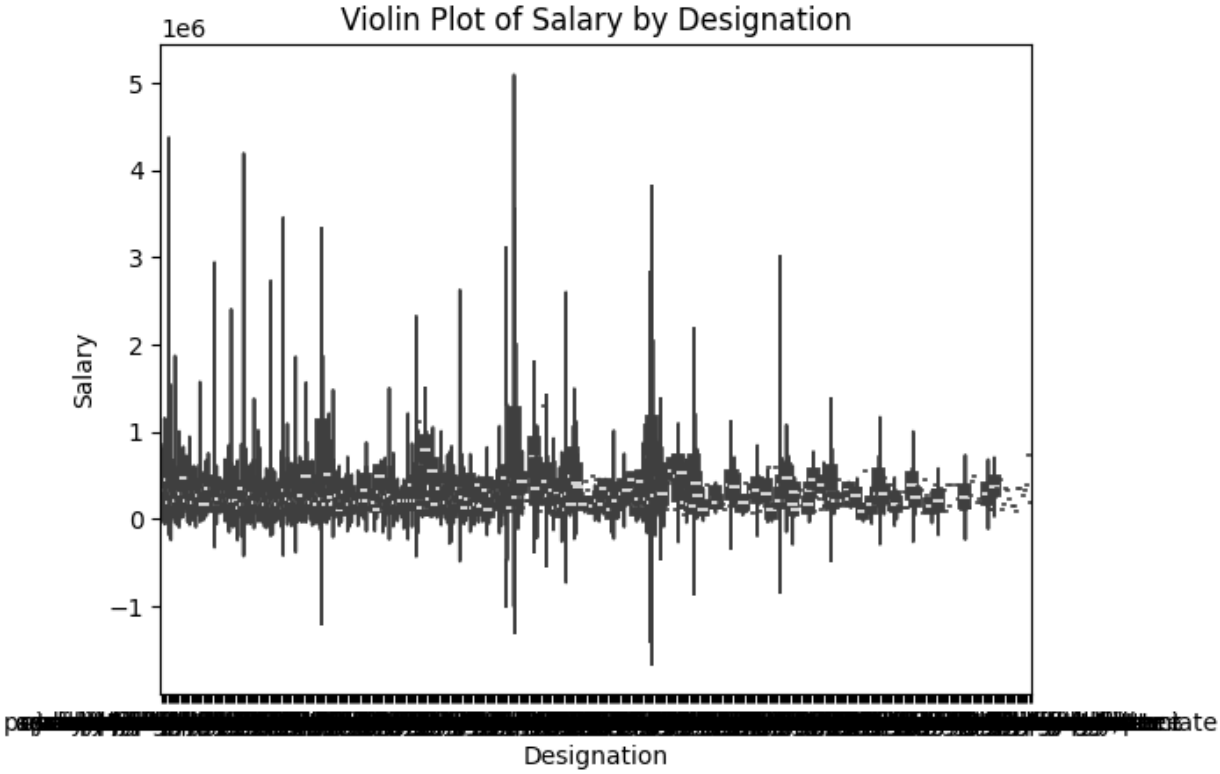

Violin Plot of agreeableness by DOL
agreeableness
4
2
0
−2
−4
−6
DOL

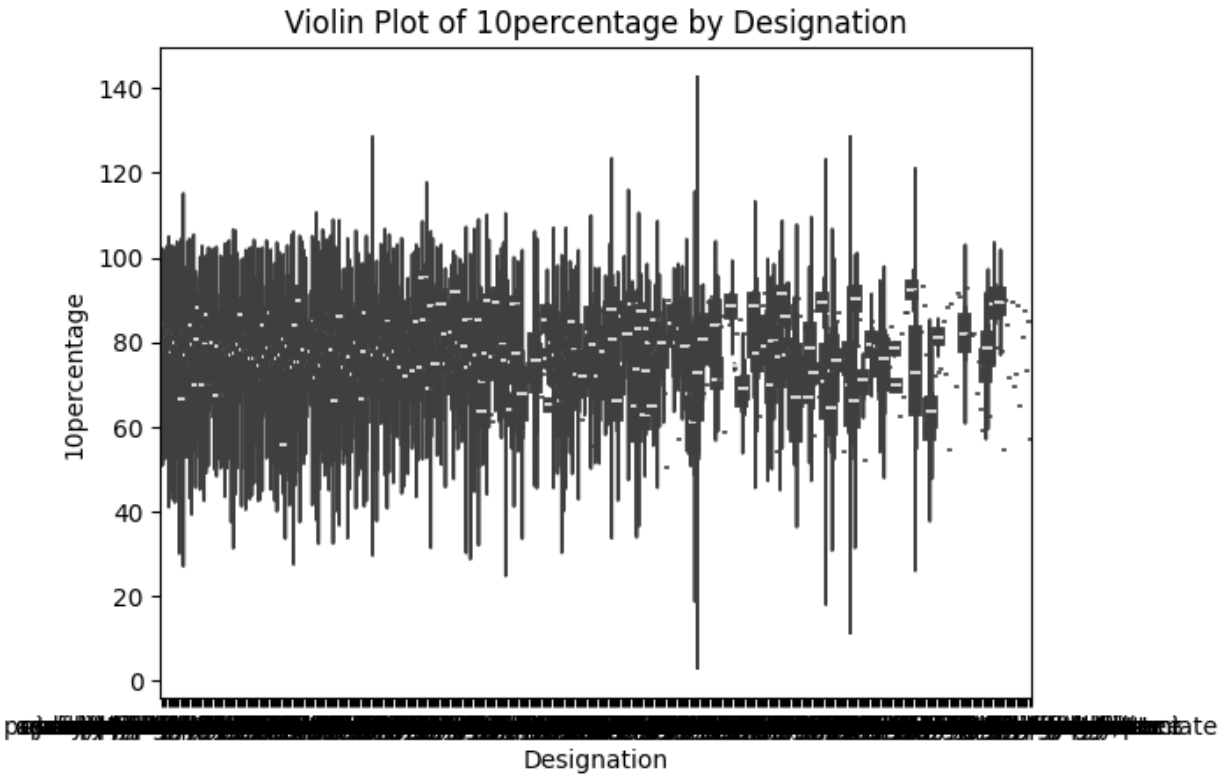

Violin Plot of extraversion by DOL
extraversion
2
0
−2
−4
DOL

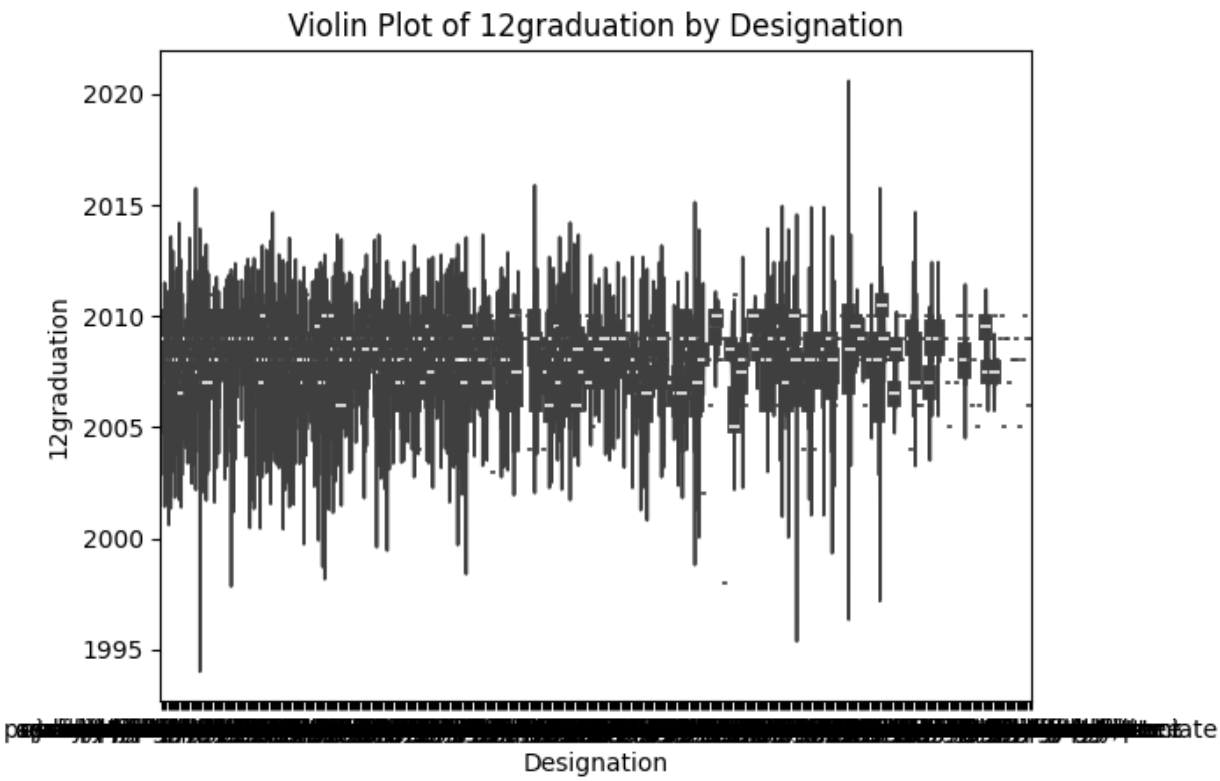

Violin Plot of nueroticism by DOL
nueroticism
4
2
0
−2
−4
DOL

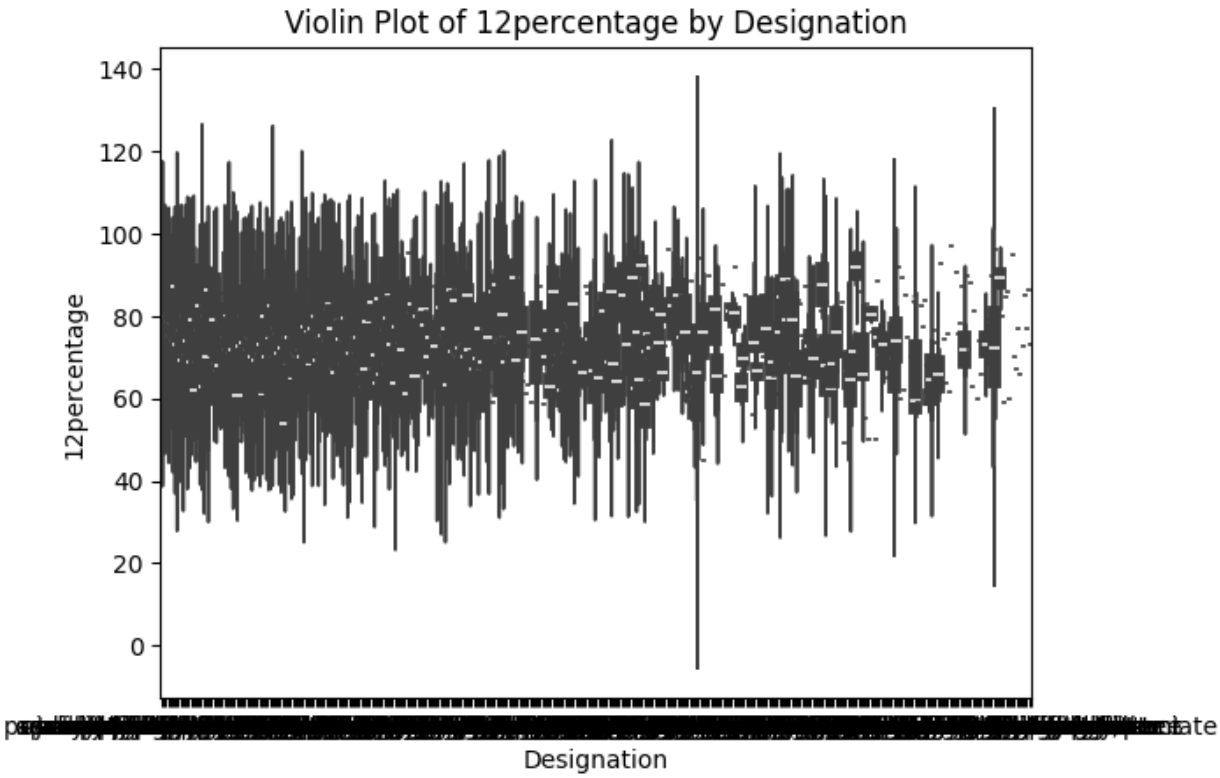

Violin Plot of openess_to_experience by DOL
openess_to_experience
4
2
0
−2
−4
−6
−8
DOL

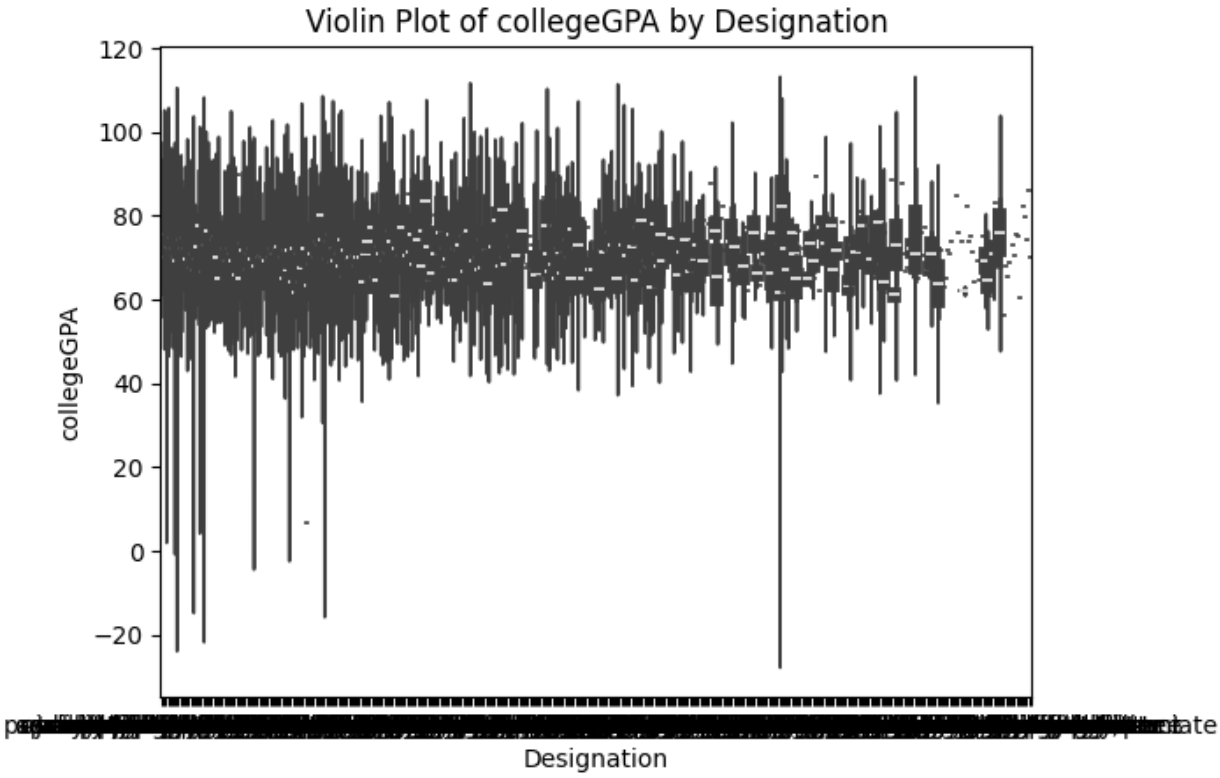

Violin Plot of Age by DOL
Age
40
35
30
25
20
DOL

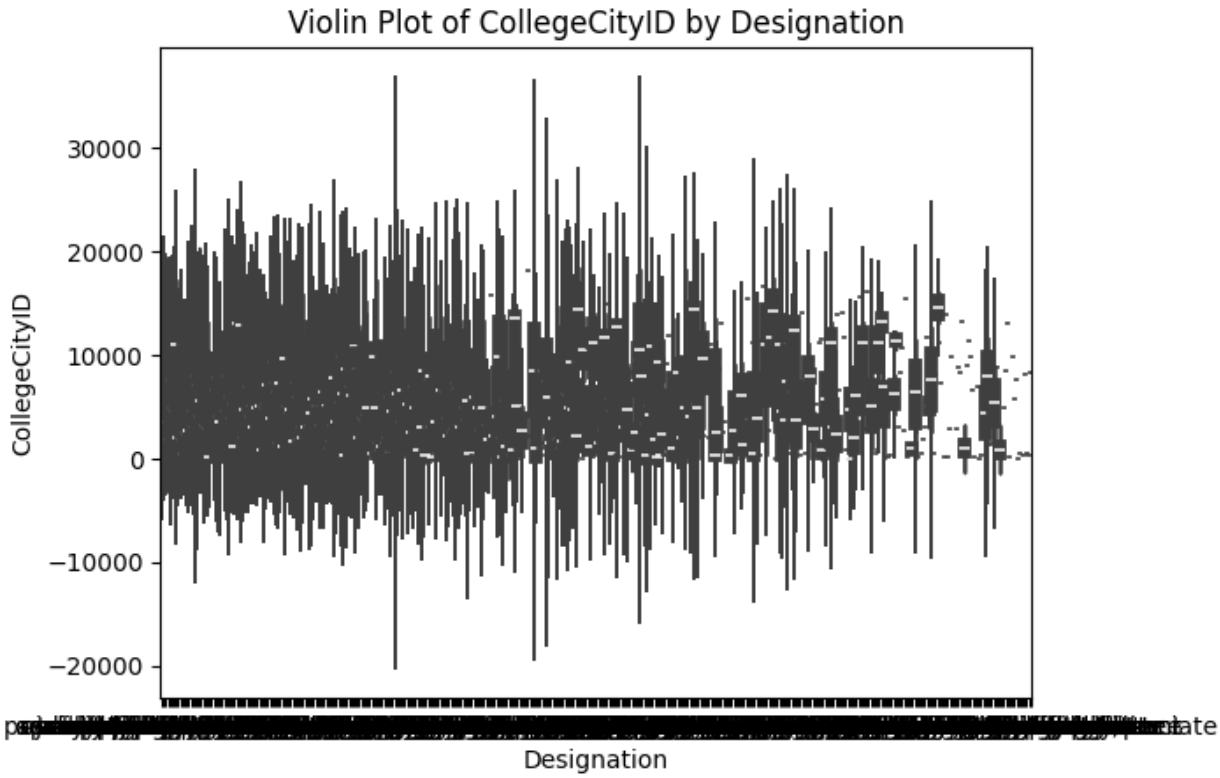

Violin Plot of Salary by Designation
1e6
Salary
5
4
3
2
1
0
−1
Designation

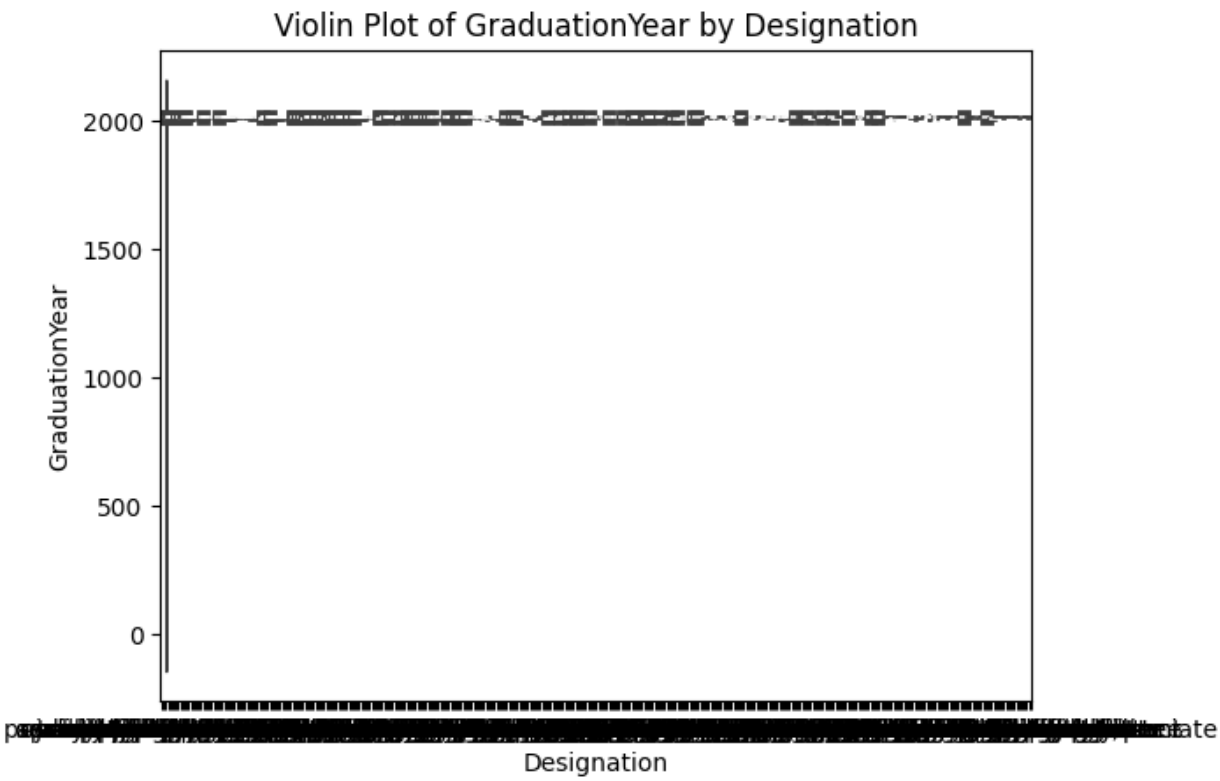

Violin Plot of 10percentage by Designation
10percentage
140
120
100
80
60
40
20
0
Designation

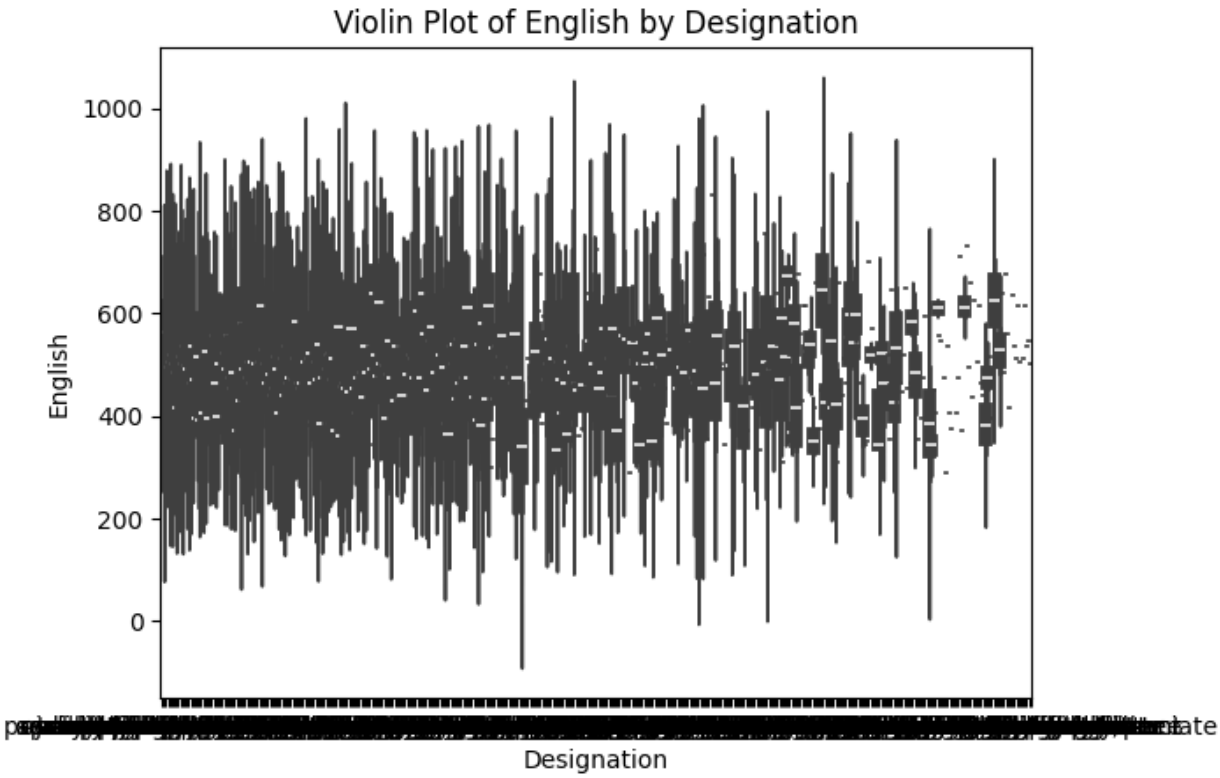

Violin Plot of 12graduation by Designation
12graduation
2020
2015
2010
2005
2000
1995
Designation

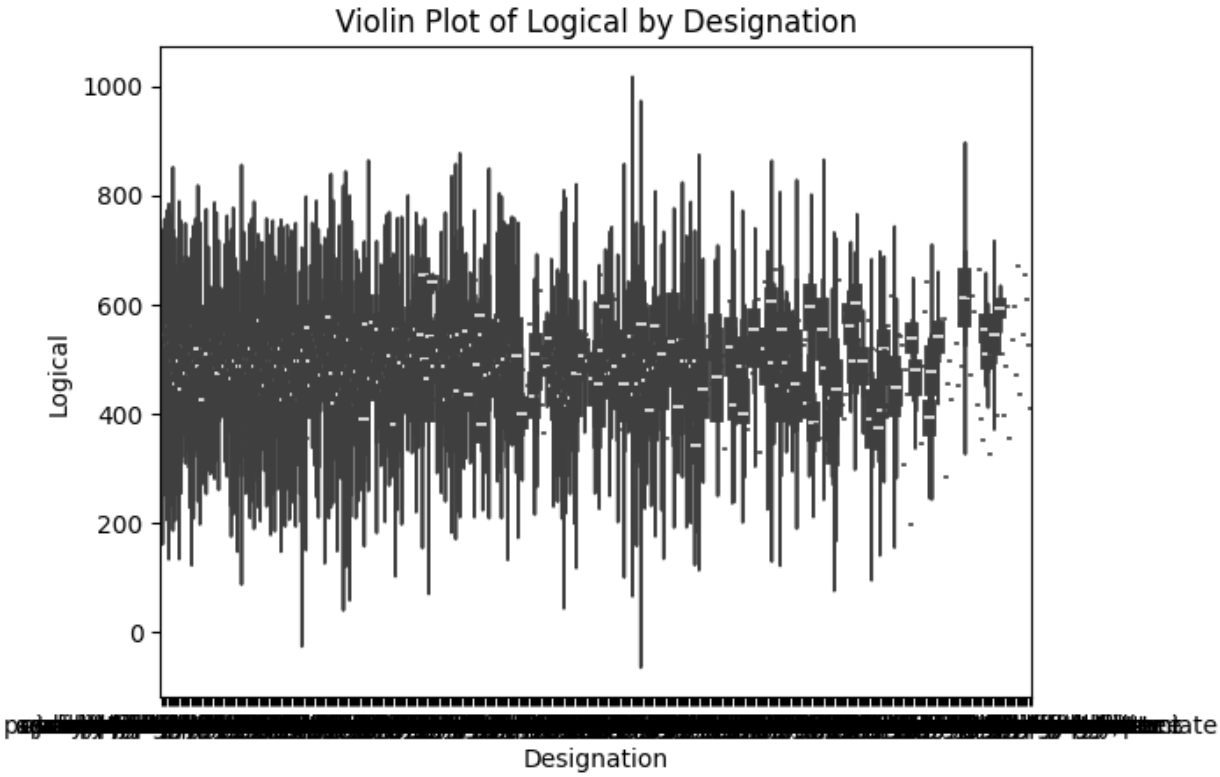

Violin Plot of 12percentage by Designation
12percentage
140
120
100
80
60
40
20
0
Designation

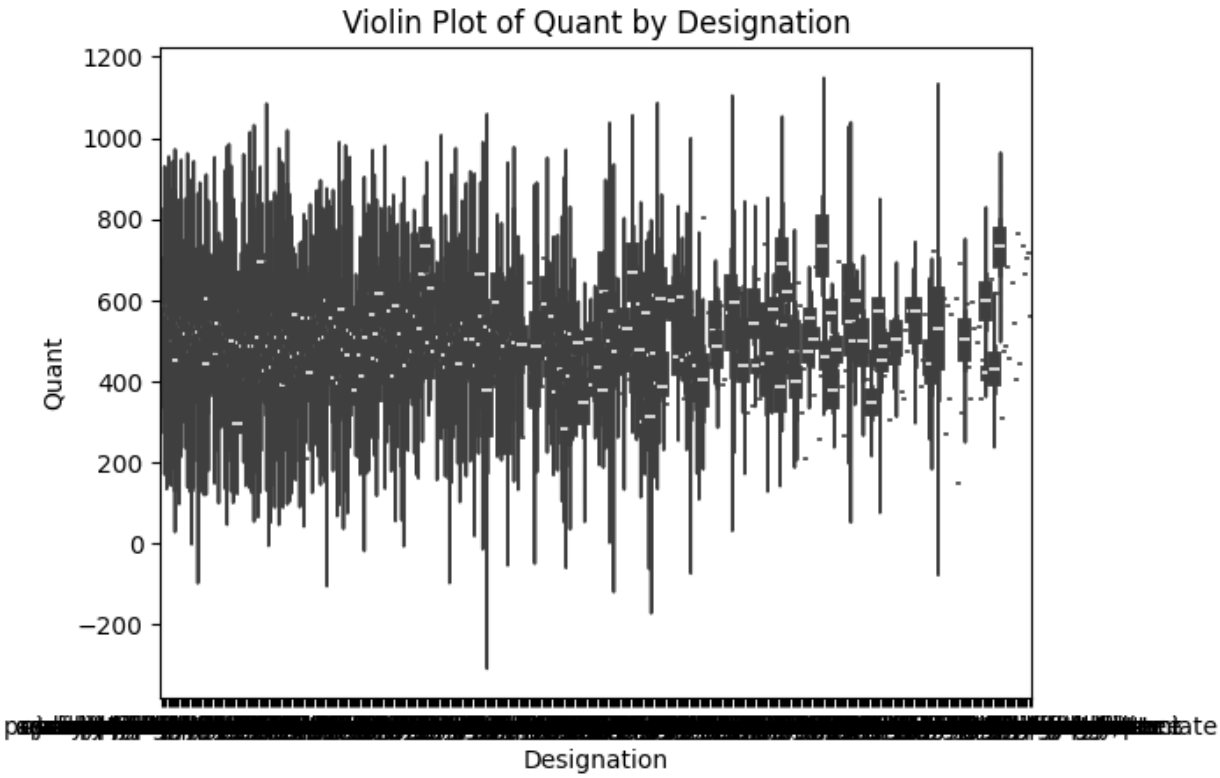

Violin Plot of collegeGPA by Designation
collegeGPA
120
100
80
60
40
20
0
−20
Designation

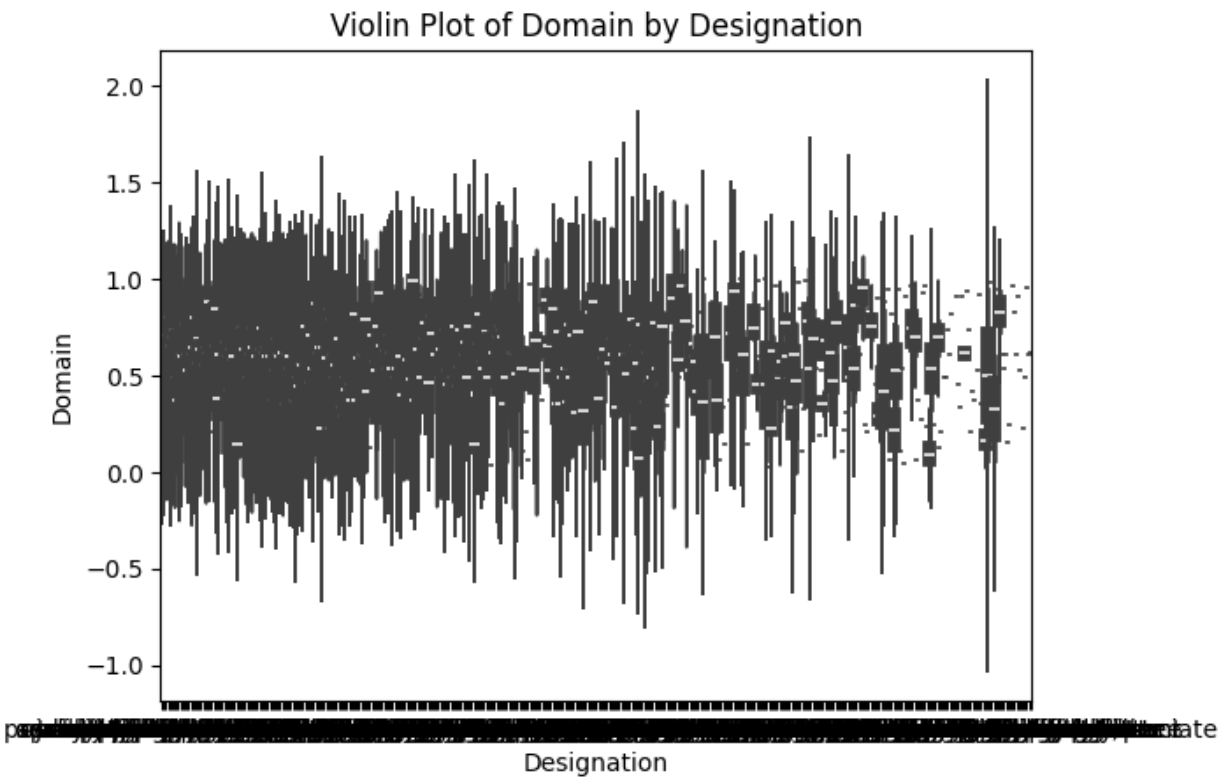

Violin Plot of CollegeCityID by Designation
CollegeCityID
30000
20000
10000
0
−10000
−20000
Designation

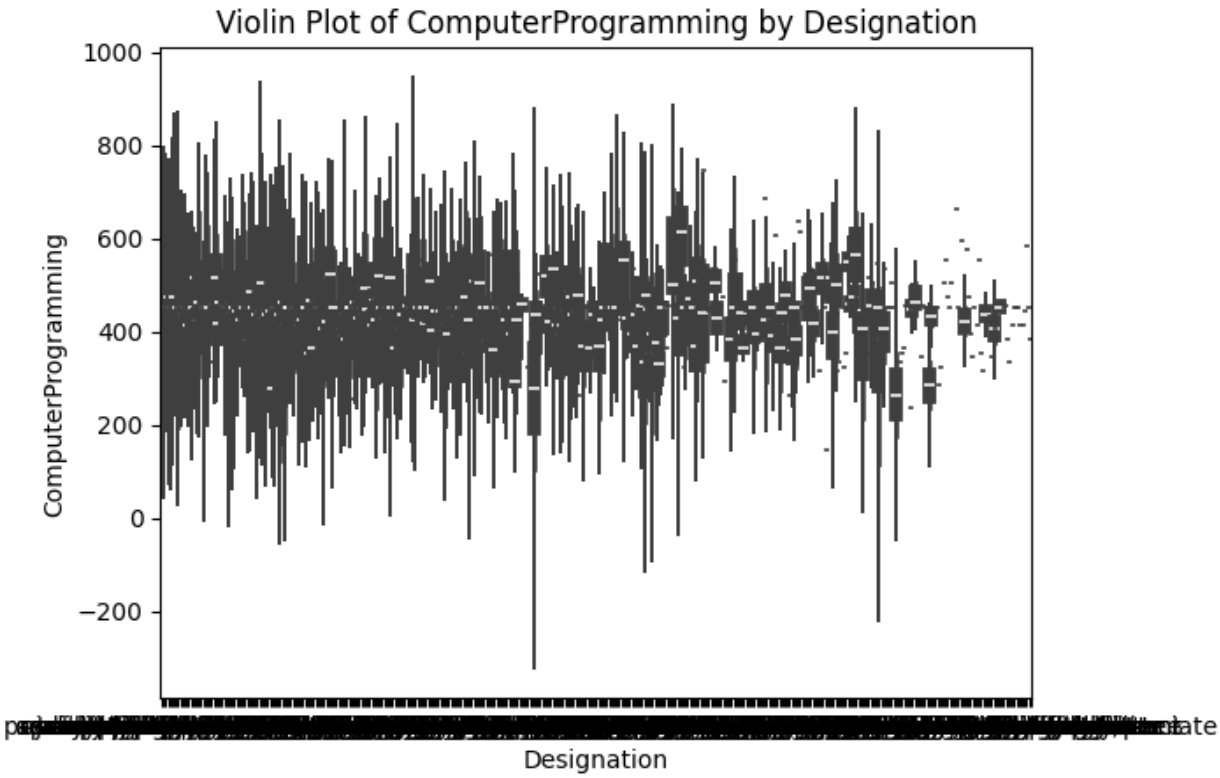

Violin Plot of GraduationYear by Designation
GraduationYear
2000
1500
1000
500
0
Designation

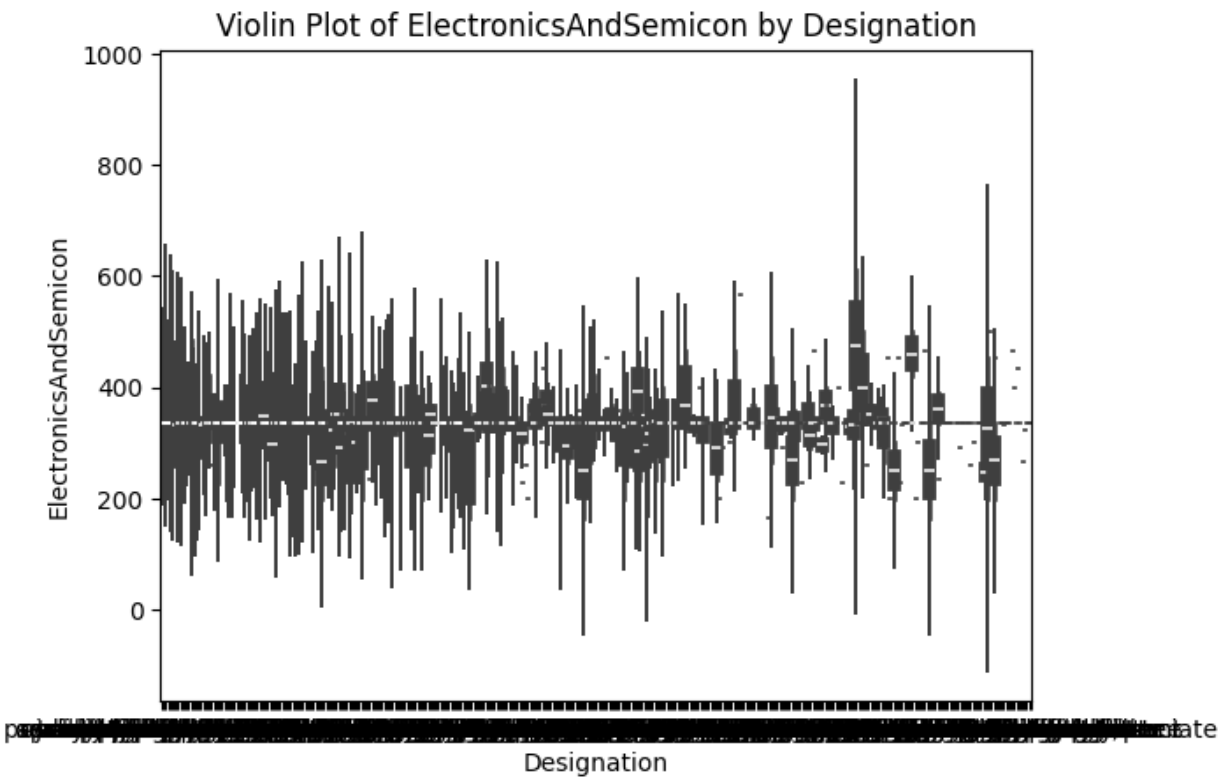

Violin Plot of English by Designation
English
1000
800
600
400
200
0
Designation

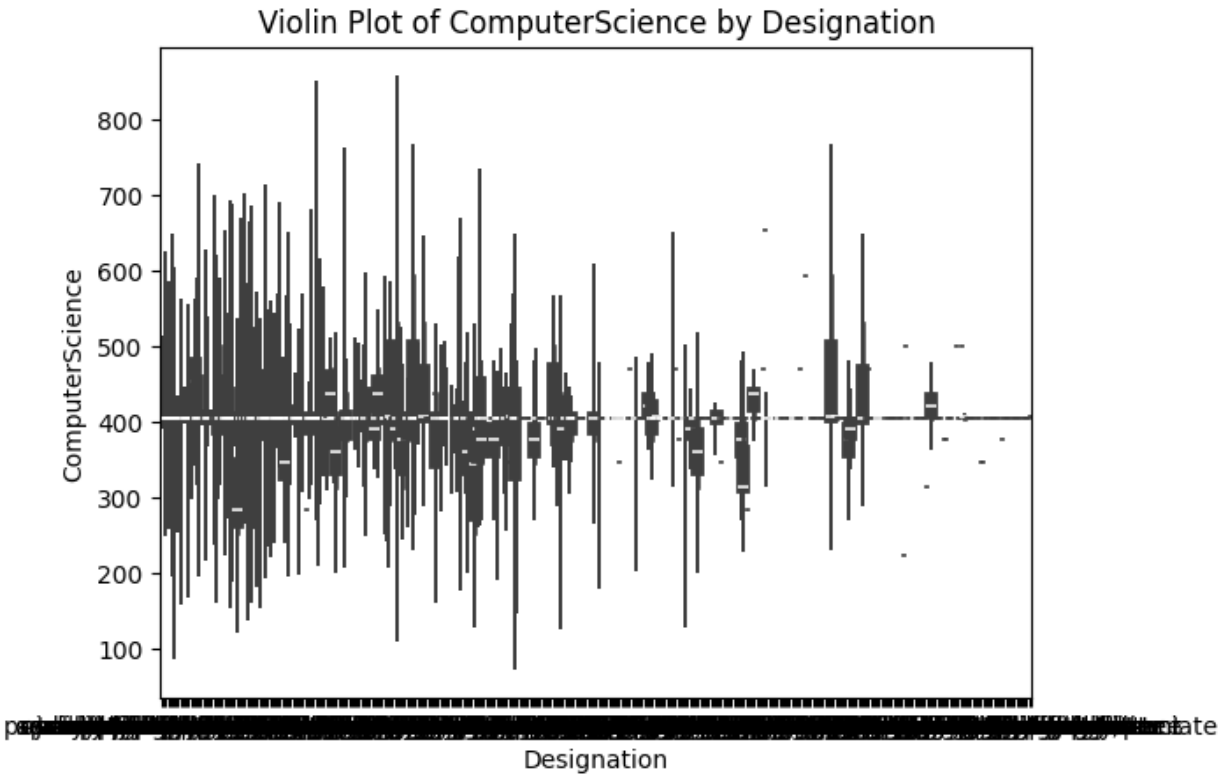

Violin Plot of Logical by Designation
Logical
1000
800
600
400
200
0
Designation

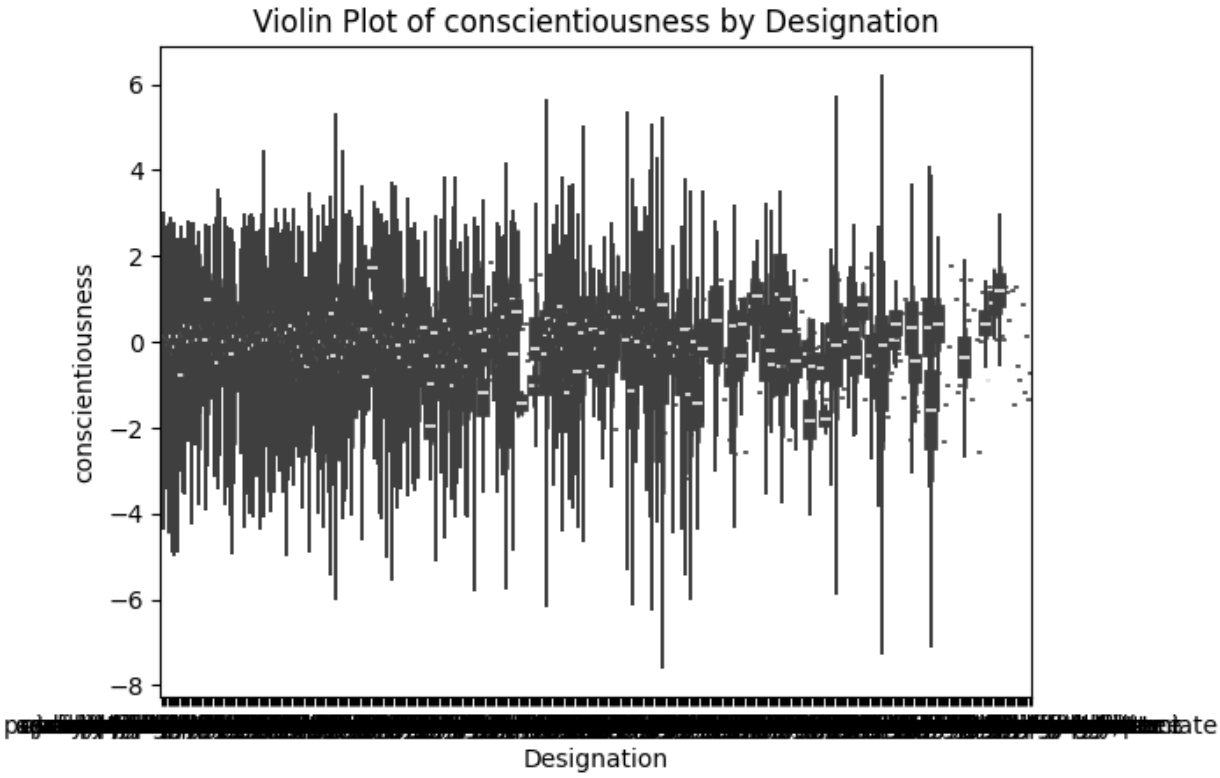

Violin Plot of Quant by Designation
Quant
1200
1000
800
600
400
200
0
−200
Designation

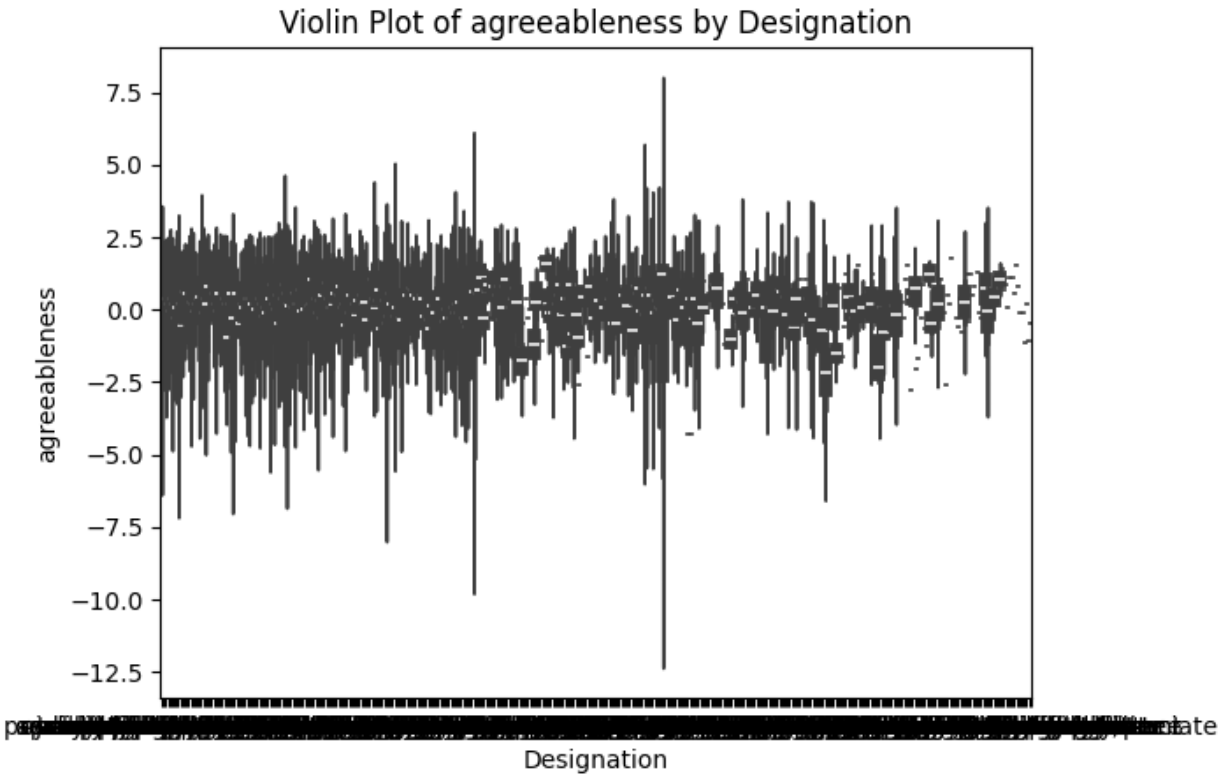

Violin Plot of Domain by Designation
Domain
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
Designation

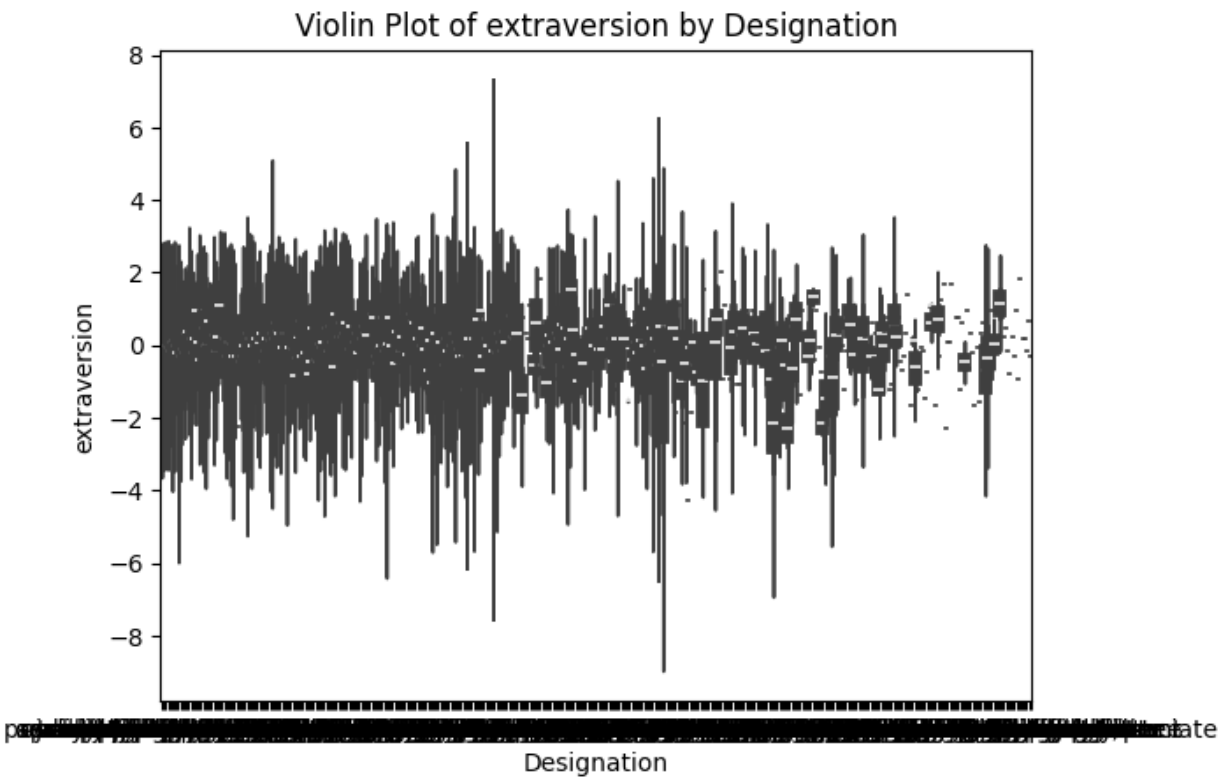

Violin Plot of ComputerProgramming by Designation
ComputerProgramming
1000
800
600
400
200
0
−200
Designation

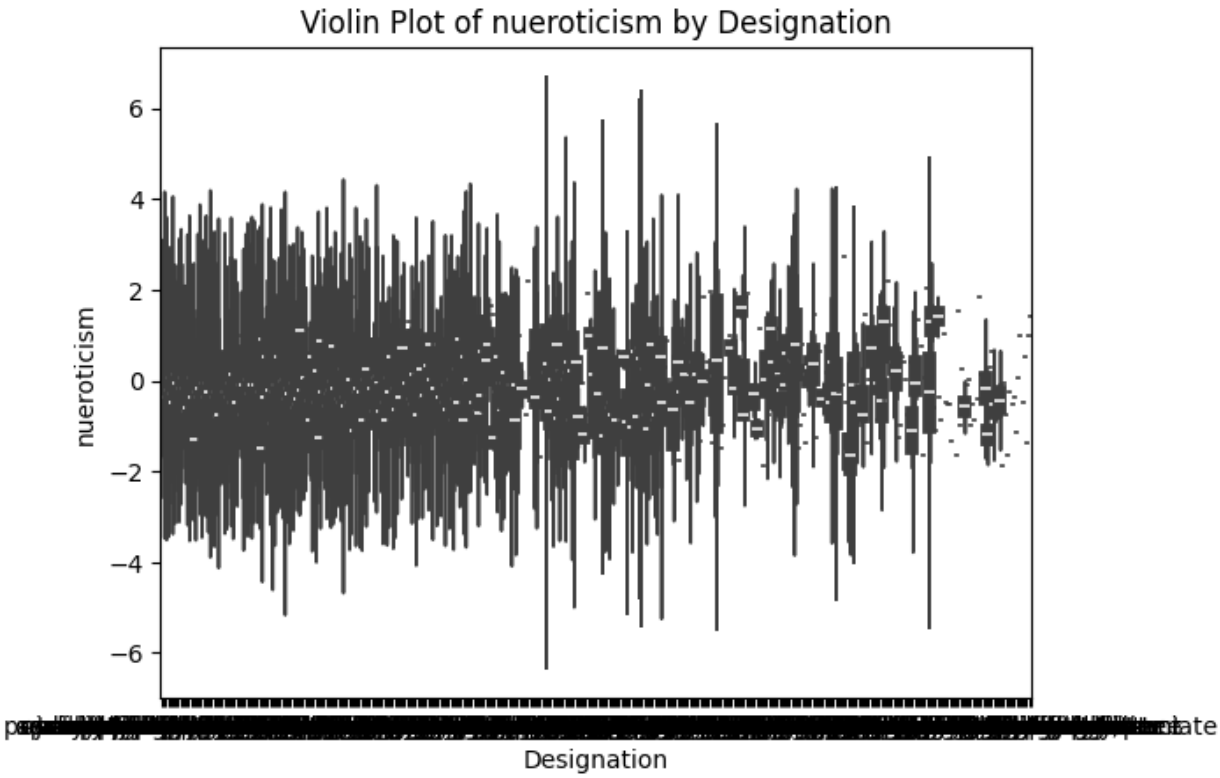

Violin Plot of ElectronicsAndSemicon by Designation
ElectronicsAndSemicon
1000
800
600
400
200
0
Designation

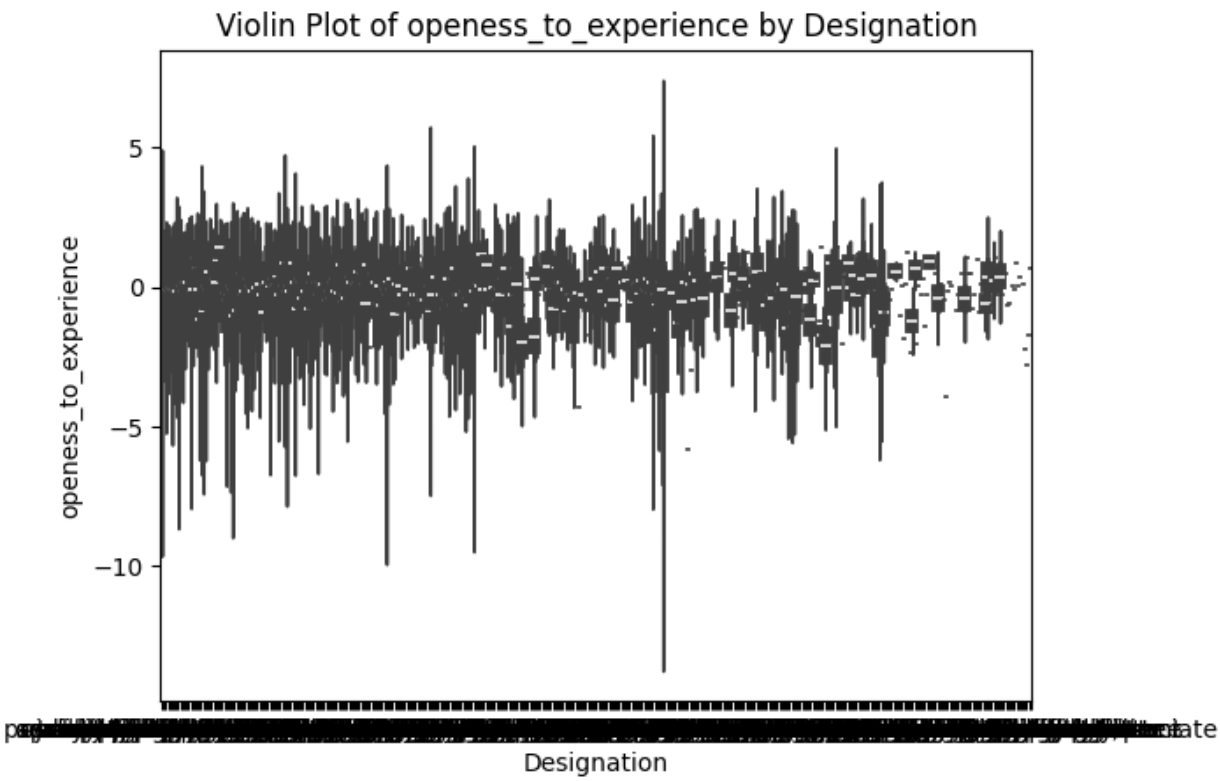

Violin Plot of ComputerScience by Designation
ComputerScience
800
700
600
500
400
300
200
100
Designation


Violin Plot of conscientiousness by Designation
conscientiousness
6
4
2
0
−2
−4
−6
−8
Designation

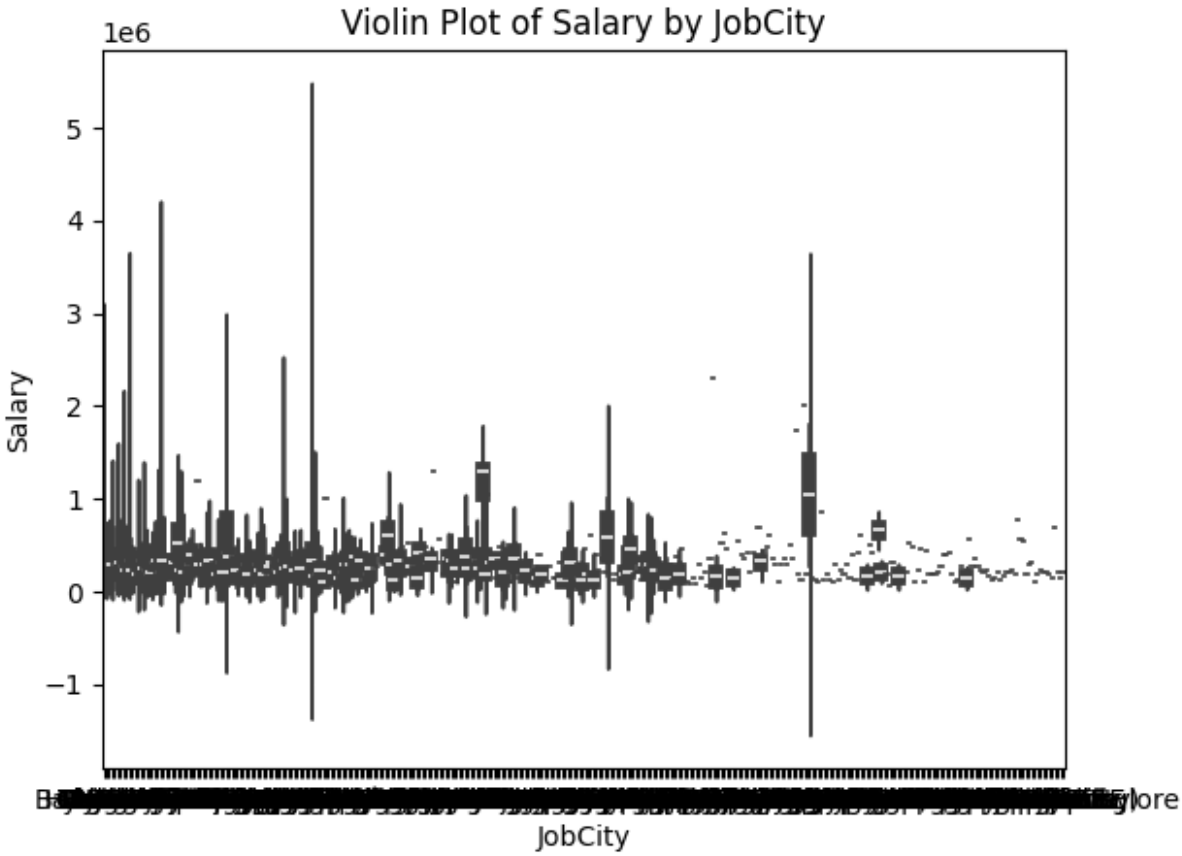

Violin Plot of agreeableness by Designation
agreeableness
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−12.5
Designation

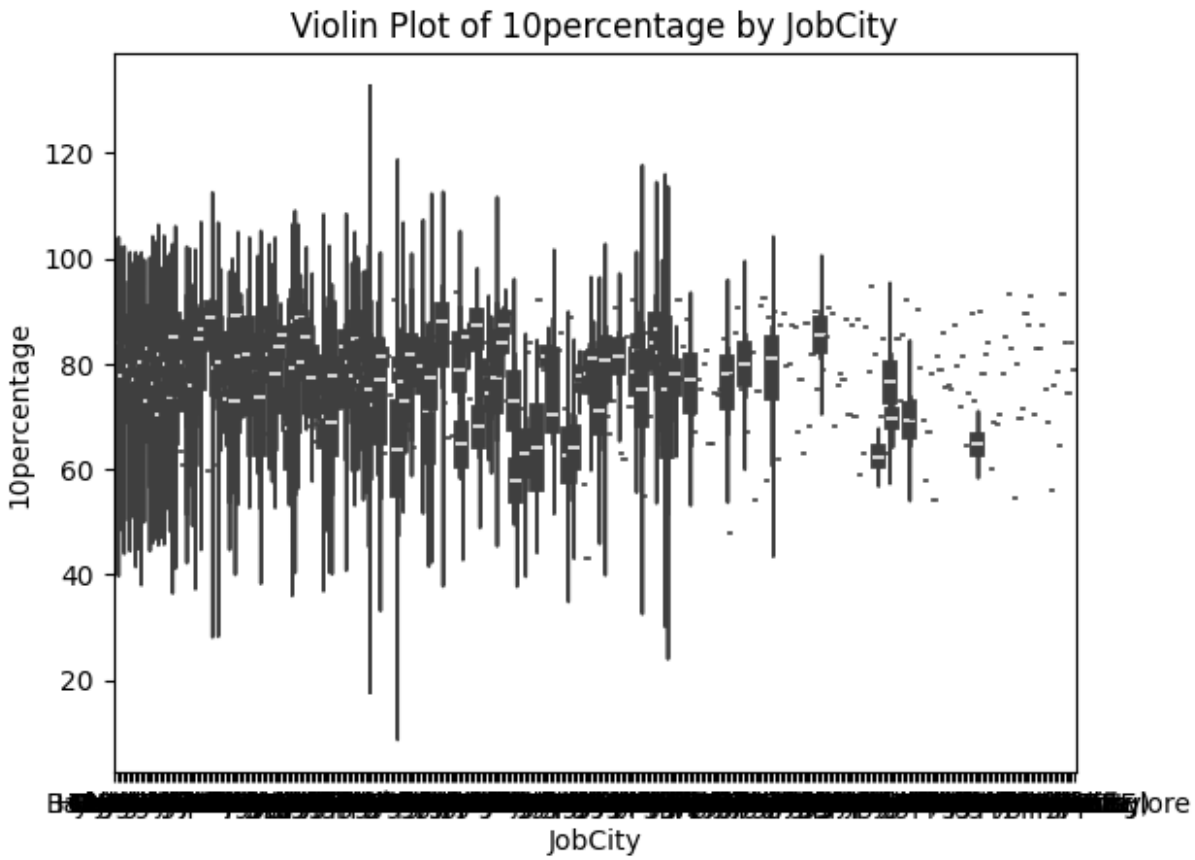

Violin Plot of extraversion by Designation
extraversion
8
6
4
2
0
−2
−4
−6
−8
Designation

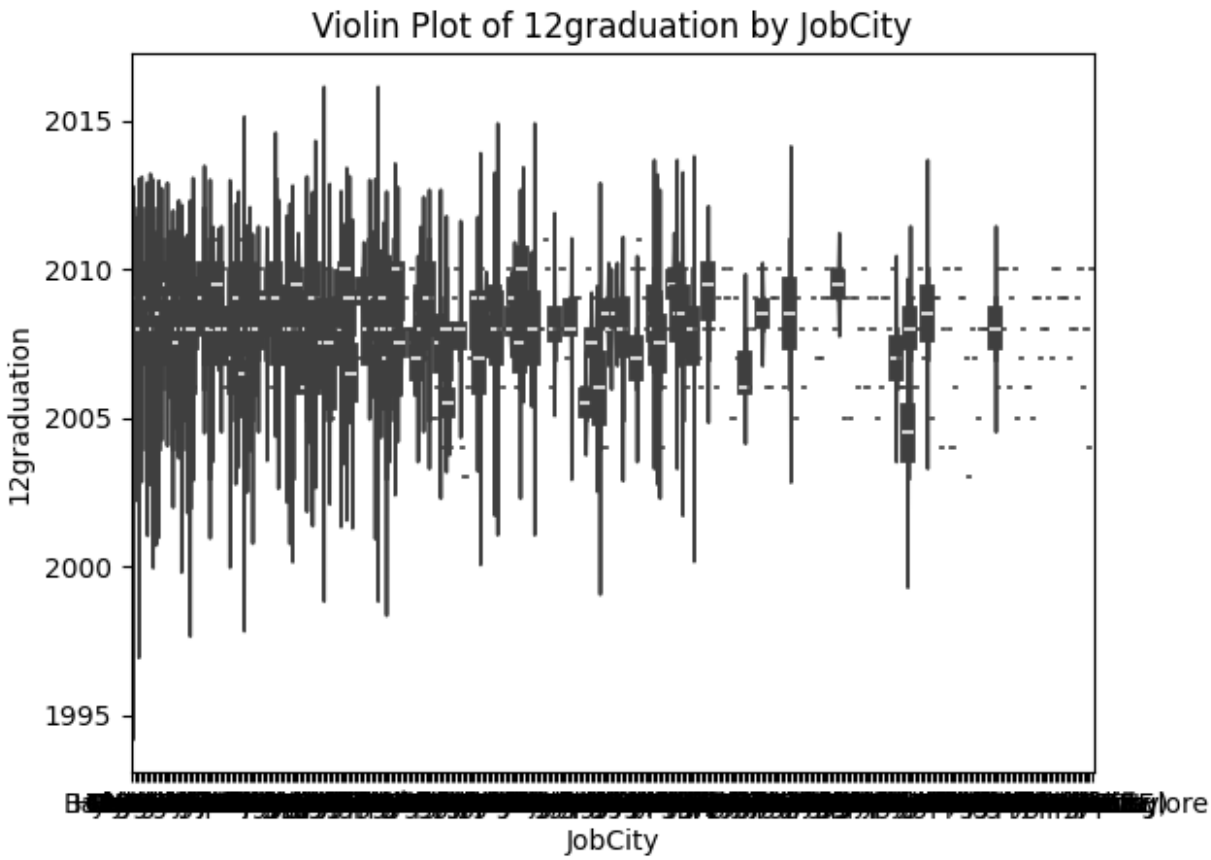

Violin Plot of nueroticism by Designation
nueroticism
6
4
2
0
−2
−4
−6
Designation

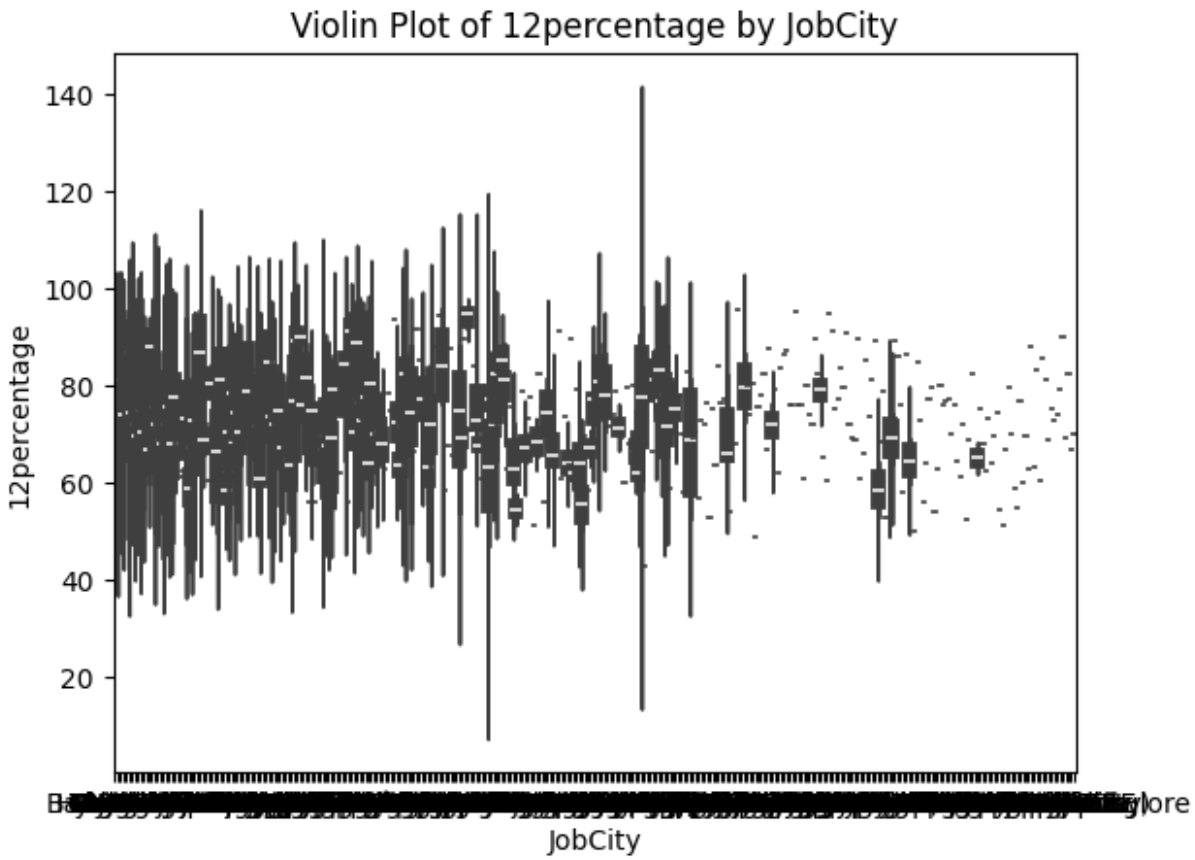

Violin Plot of openess_to_experience by Designation
openess_to_experience
5
0
−5
−10
Designation

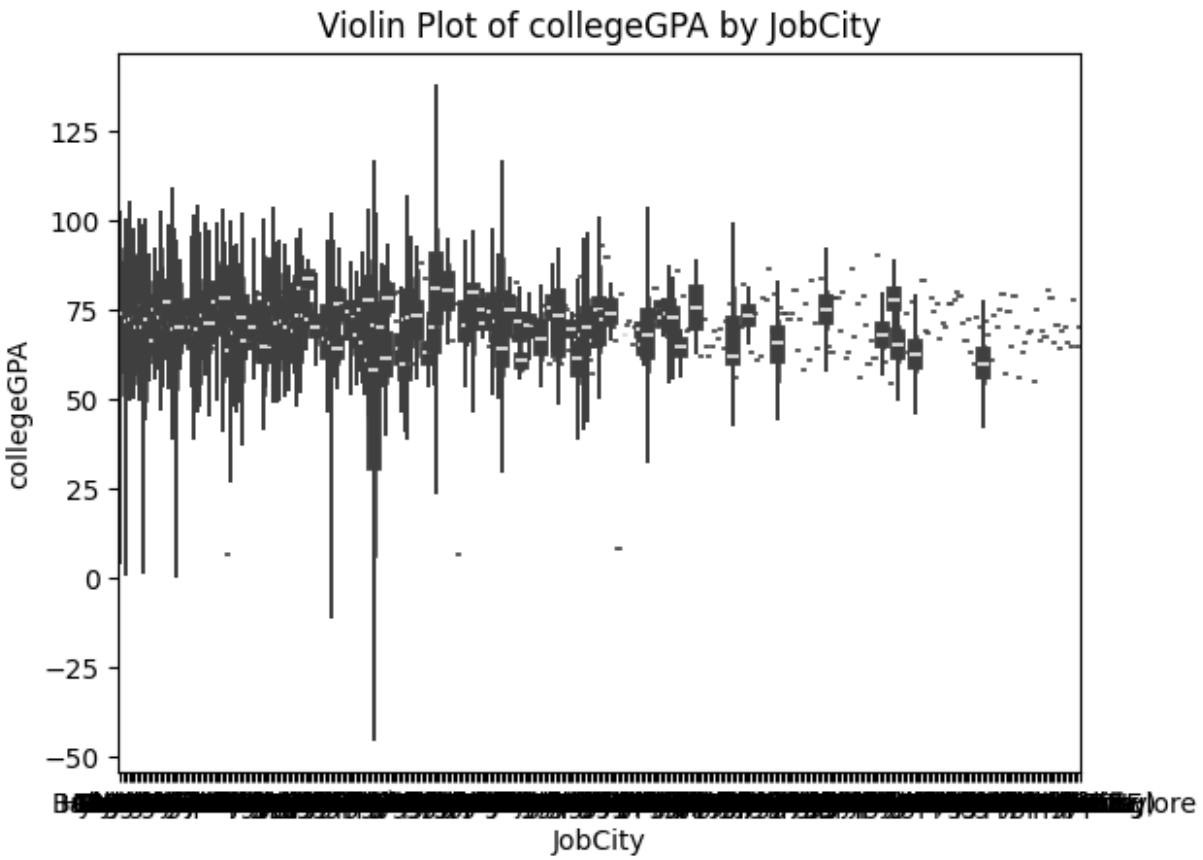
Violin Plot of Age by Designation
Age
40
35
30
25
20
15
Designation

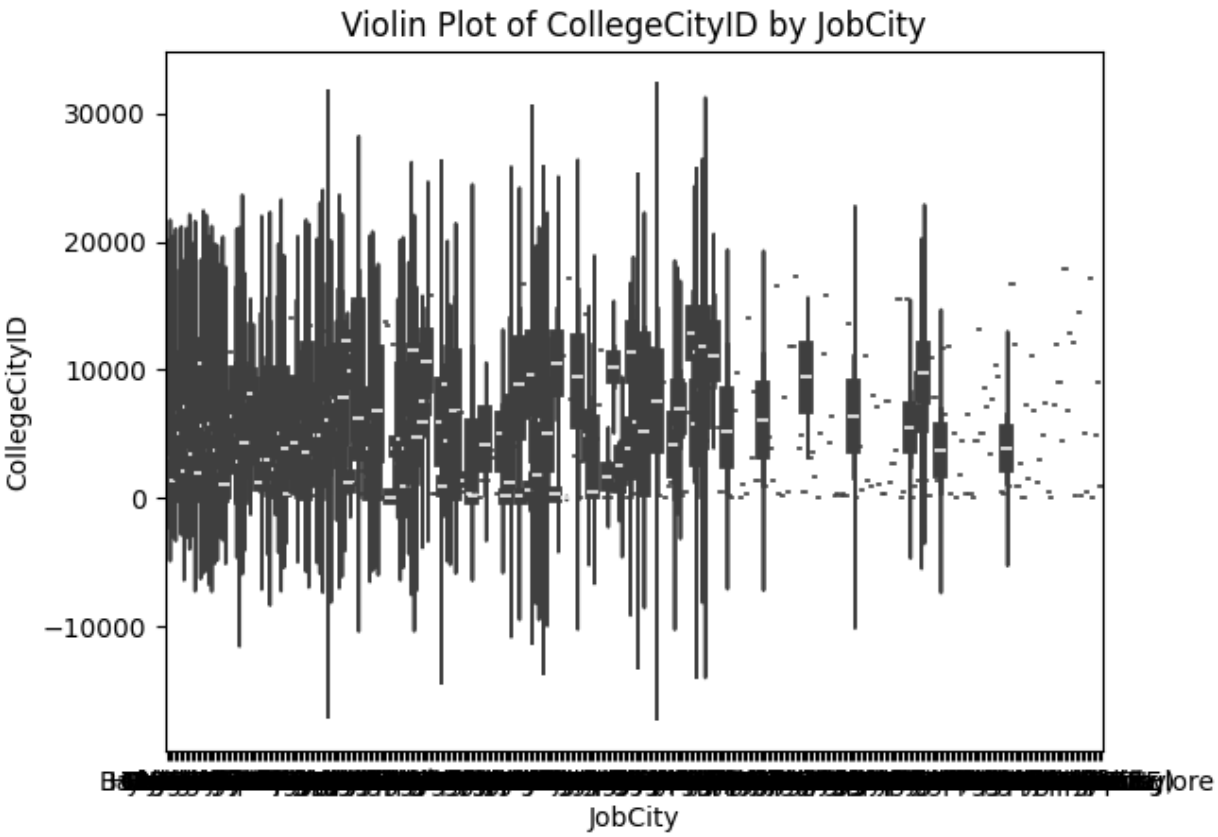

Violin Plot of Salary by JobCity
1e6
Salary
5
4
3
2
1
0
−1
JobCity

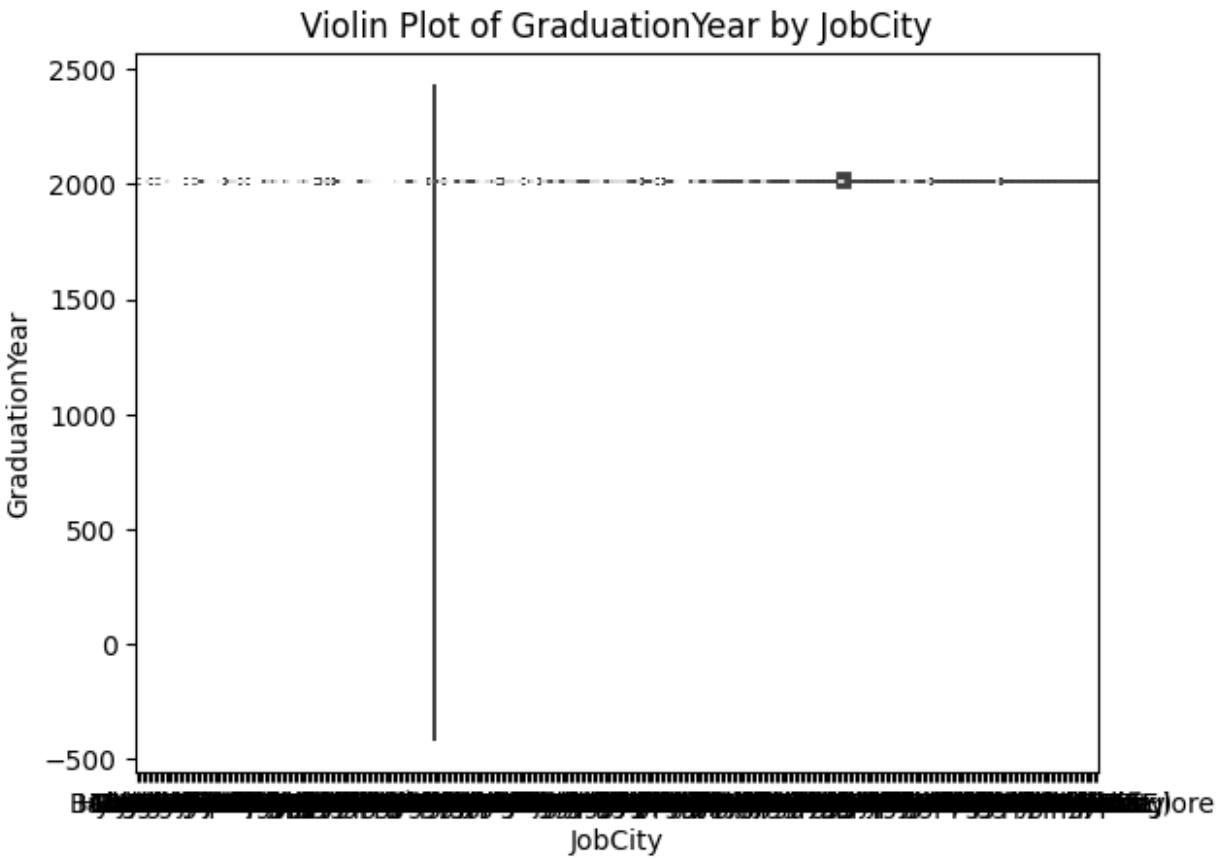

Violin Plot of 10percentage by JobCity
10percentage
120
100
80
60
40
20
JobCity

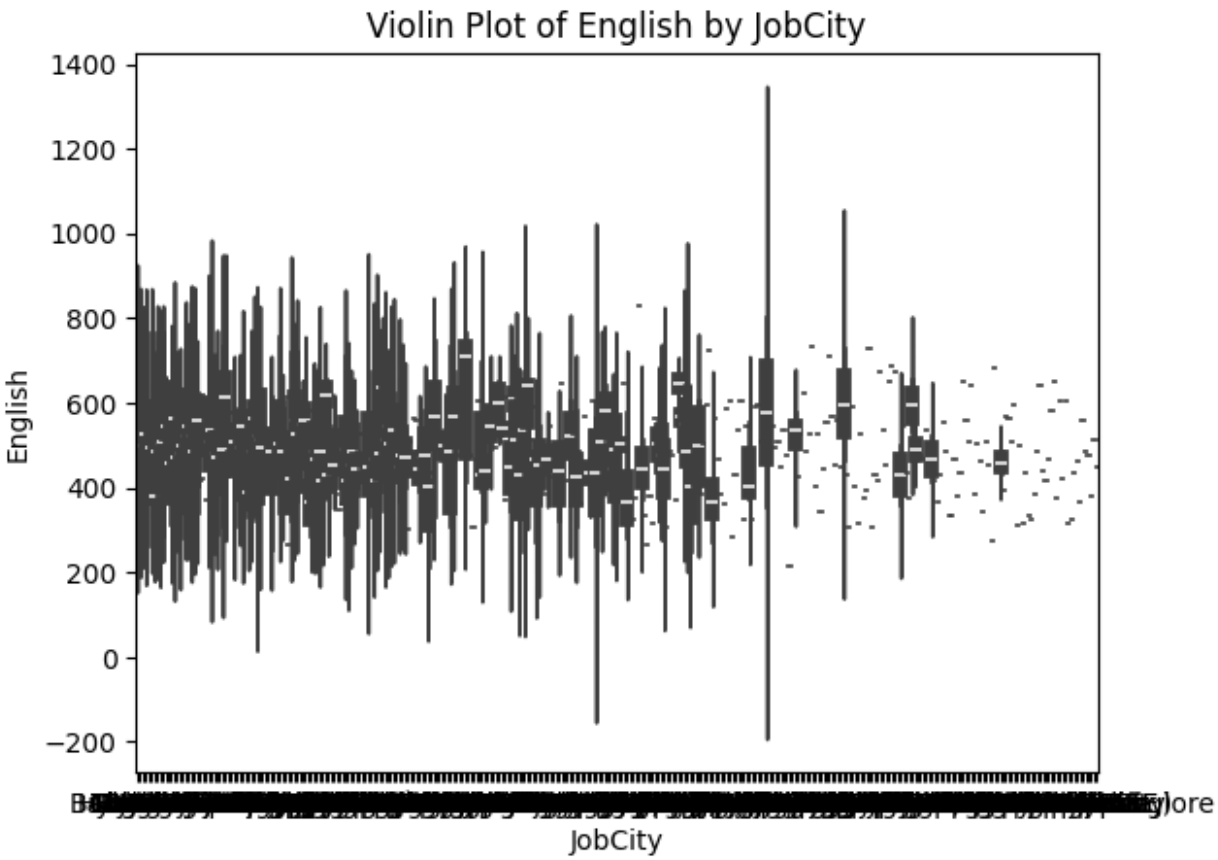

Violin Plot of 12graduation by JobCity
2015
2010
2005
2000
1995
12graduation
JobCity

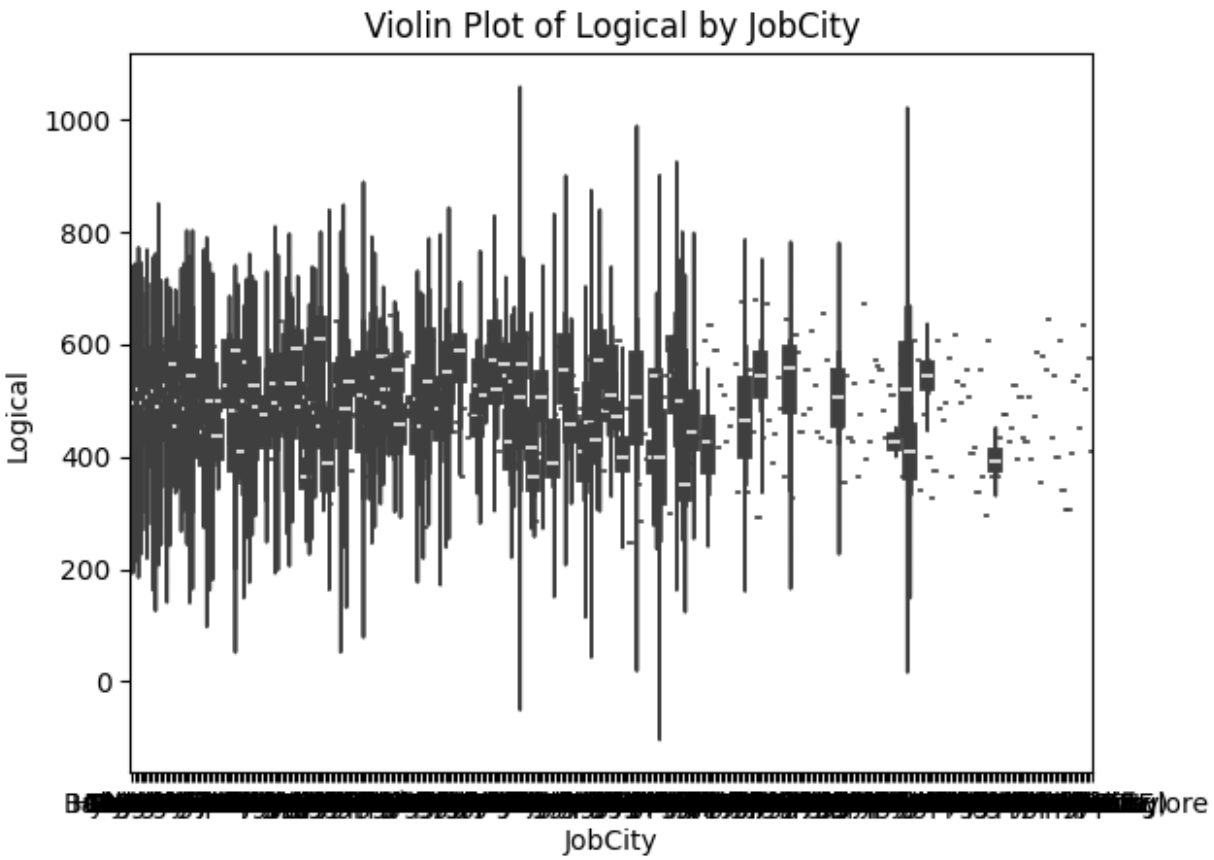

Violin Plot of 12percentage by JobCity
12percentage
140
120
100
80
60
40
20
JobCity

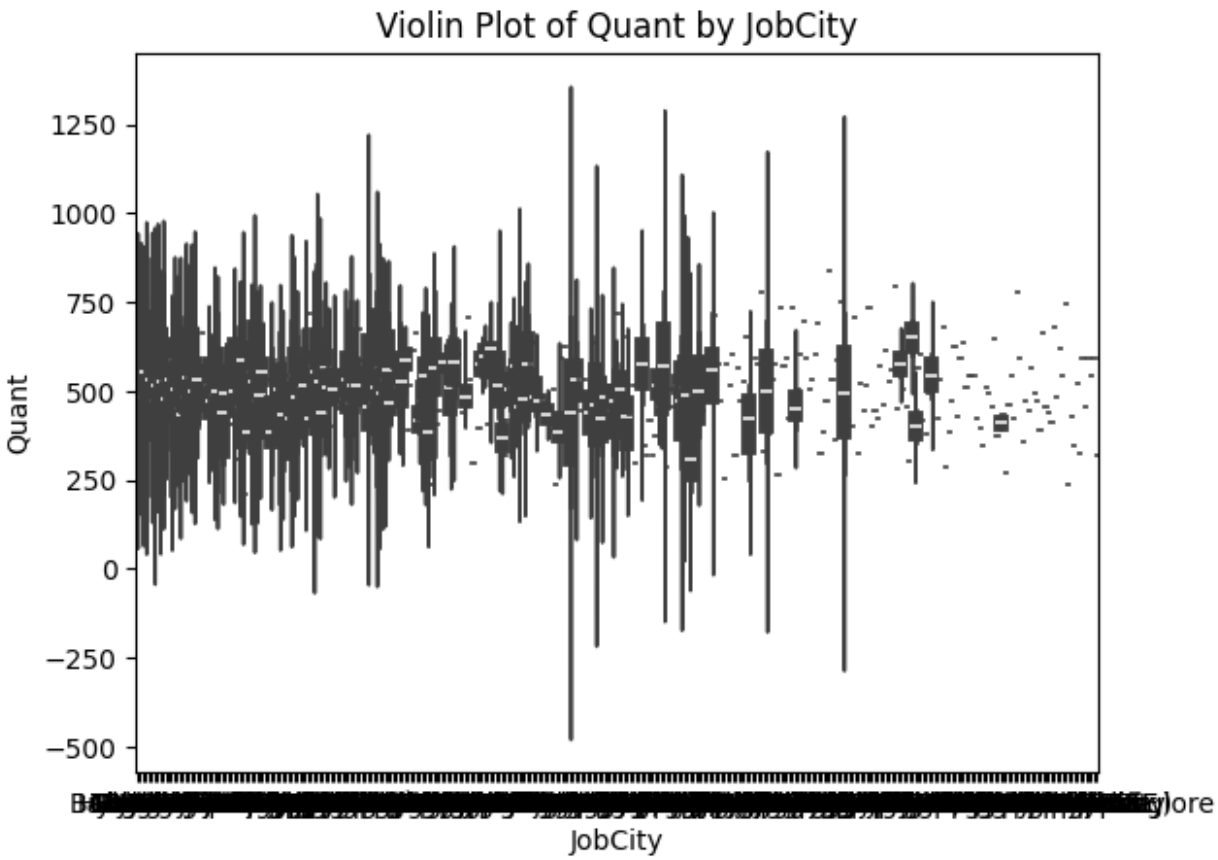

Violin Plot of collegeGPA by JobCity
collegeGPA
125
100
75
50
25
0
−25
−50
JobCity

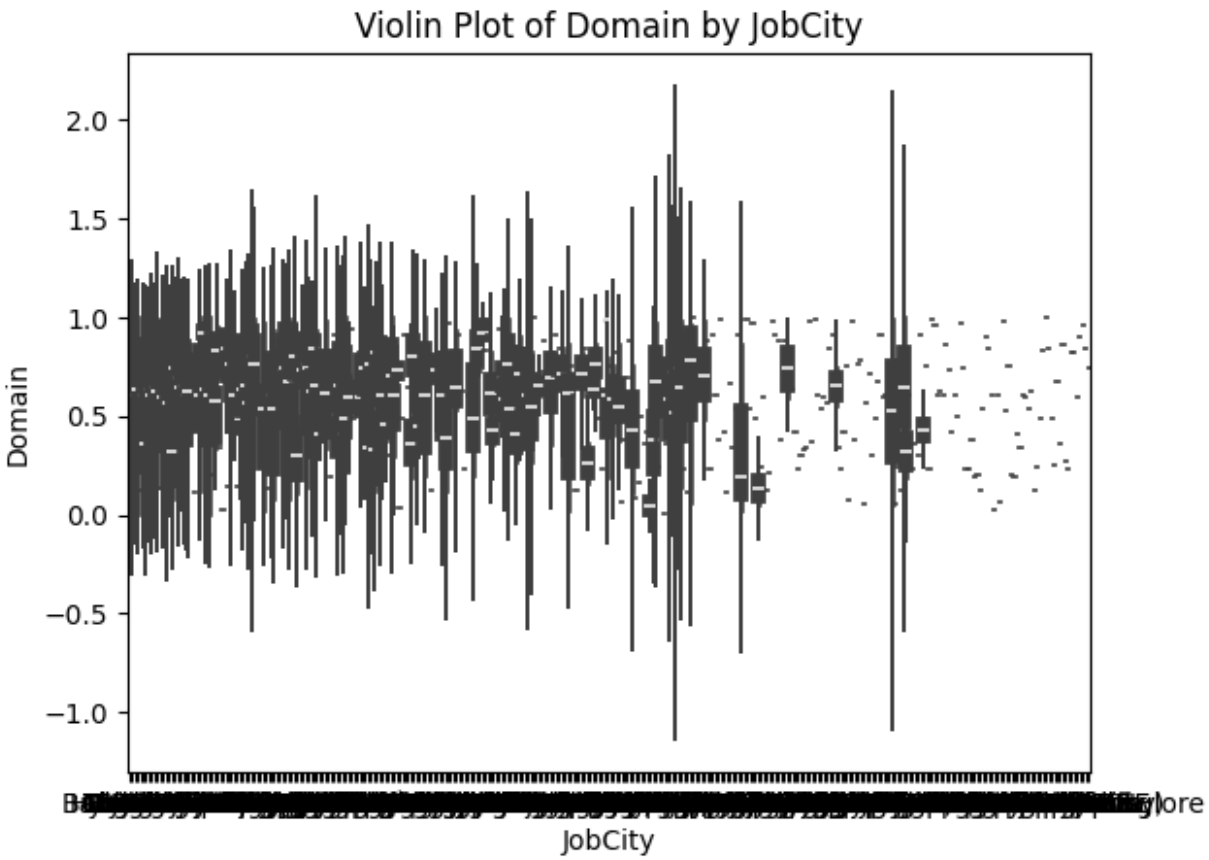

Violin Plot of CollegeCityID by JobCity
30000
20000
10000
0
−10000
CollegeCityID
JobCity

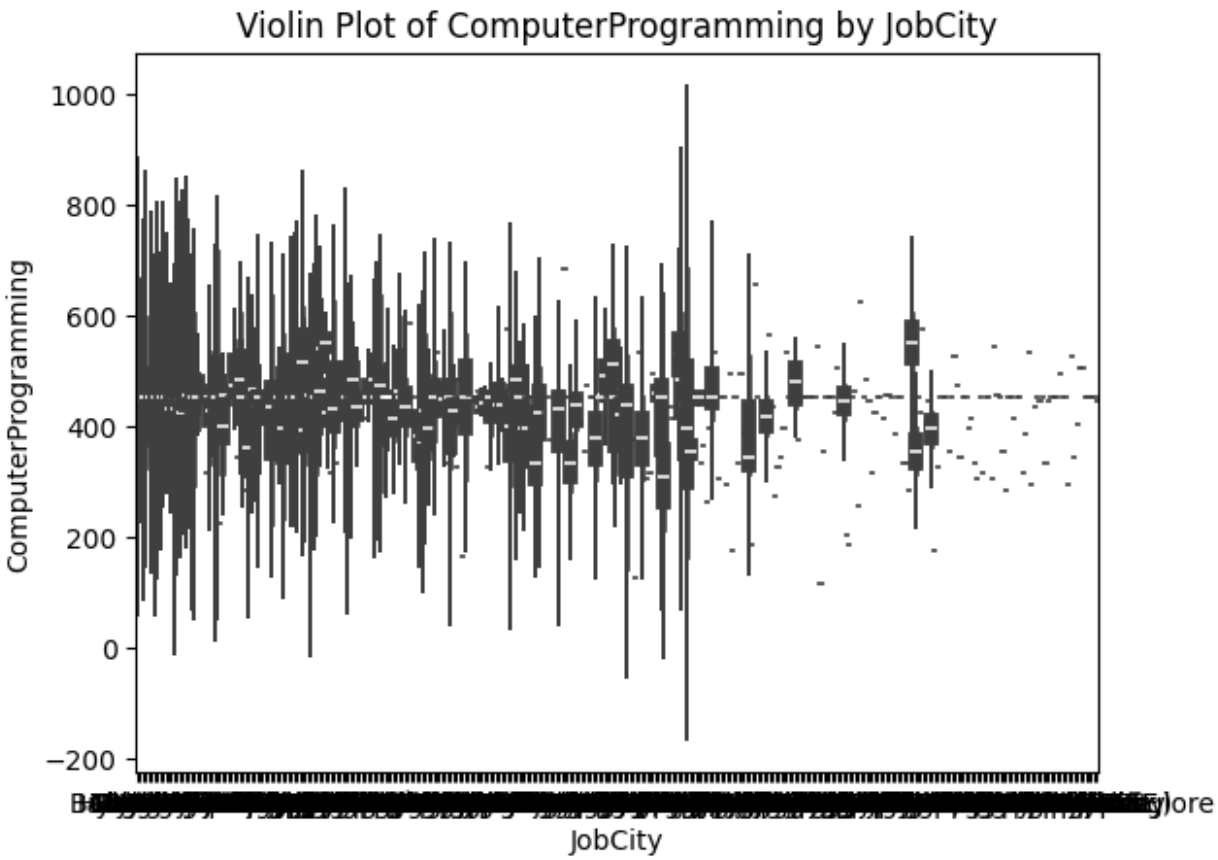

Violin Plot of GraduationYear by JobCity
2500
2000
1500
1000
500
0
−500
GraduationYear
JobCity

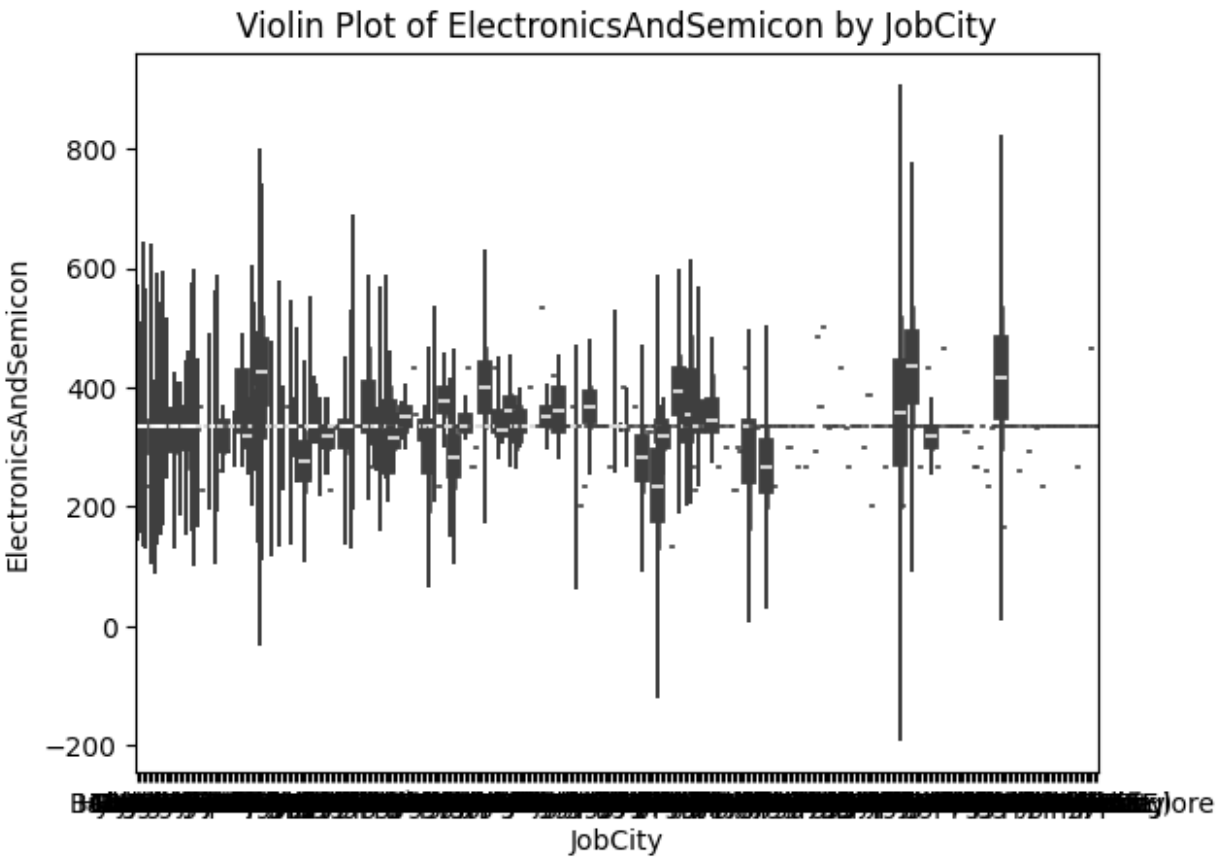

Violin Plot of English by JobCity
1400
1200
1000
800
600
400
200
0
−200
English
JobCity

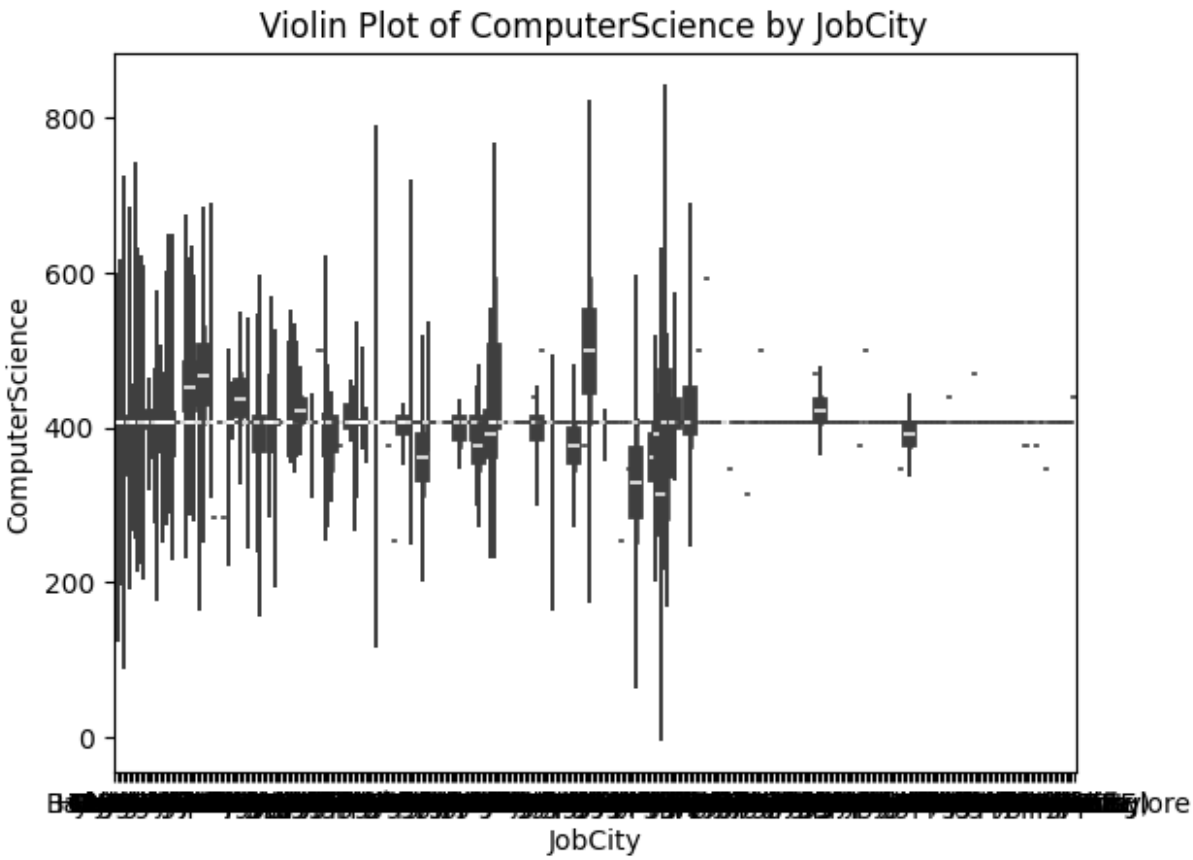

Violin Plot of Logical by JobCity
Logical
1000
800
600
400
200
0
JobCity

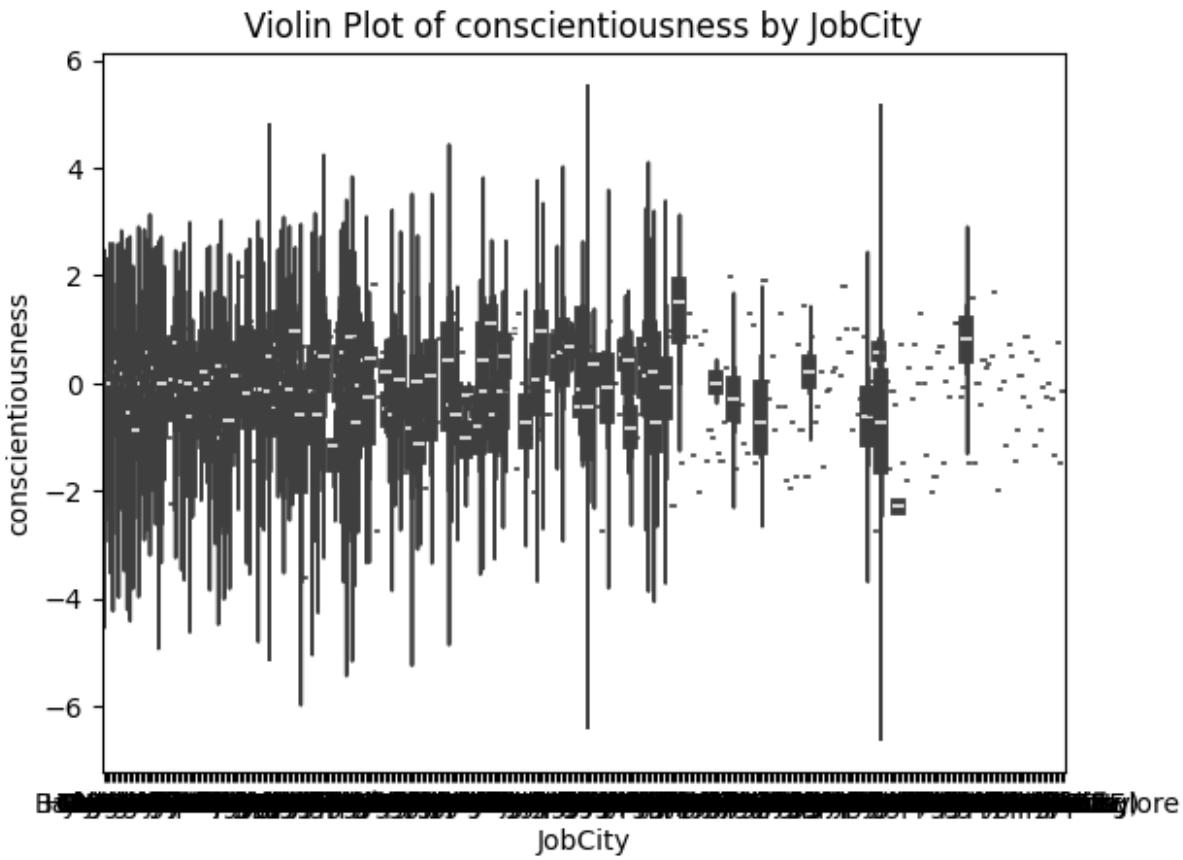

Violin Plot of Quant by JobCity
Quant
1250
1000
750
500
250
0
−250
−500
JobCity

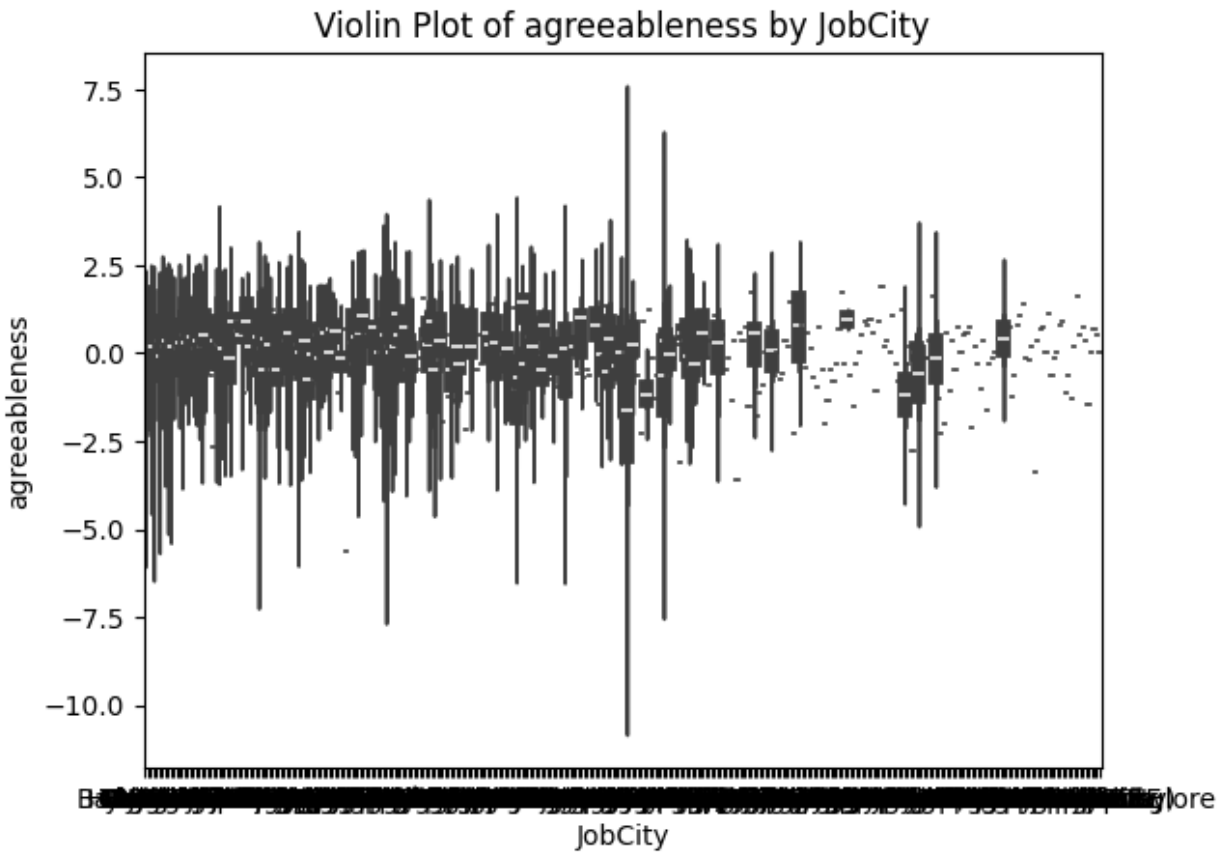

Violin Plot of Domain by JobCity
Domain
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
JobCity

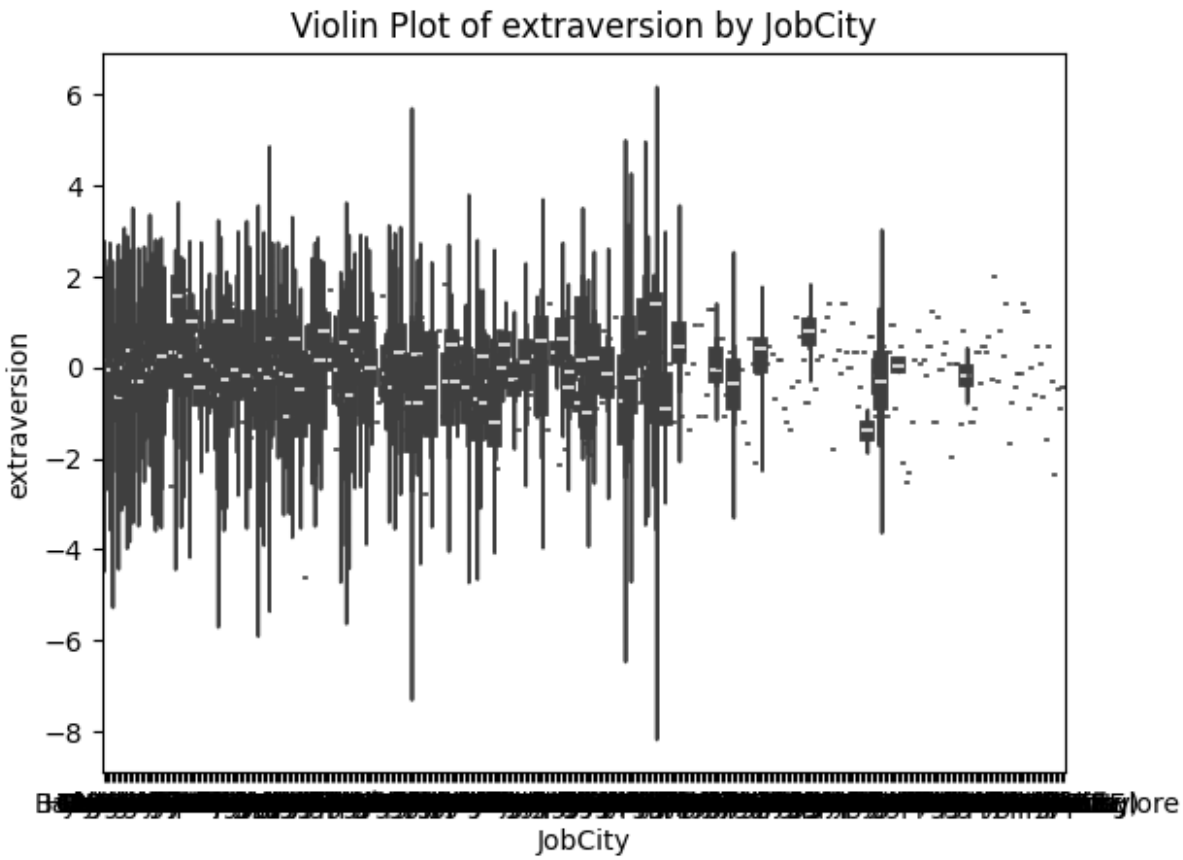

Violin Plot of ComputerProgramming by JobCity
ComputerProgramming
1000
800
600
400
200
0
−200
JobCity

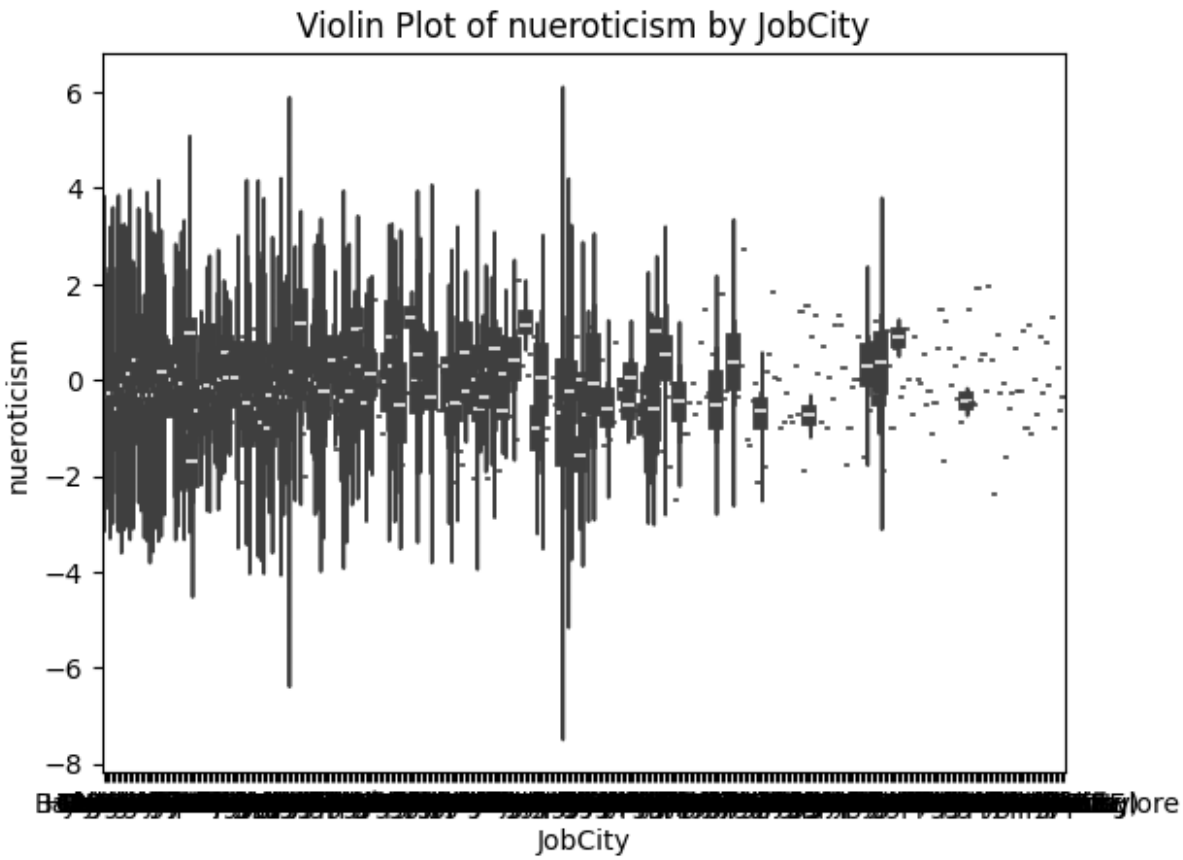

Violin Plot of ElectronicsAndSemicon by JobCity
ElectronicsAndSemicon
800
600
400
200
0
−200
JobCity


Violin Plot of ComputerScience by JobCity
ComputerScience
800
600
400
200
0
JobCity

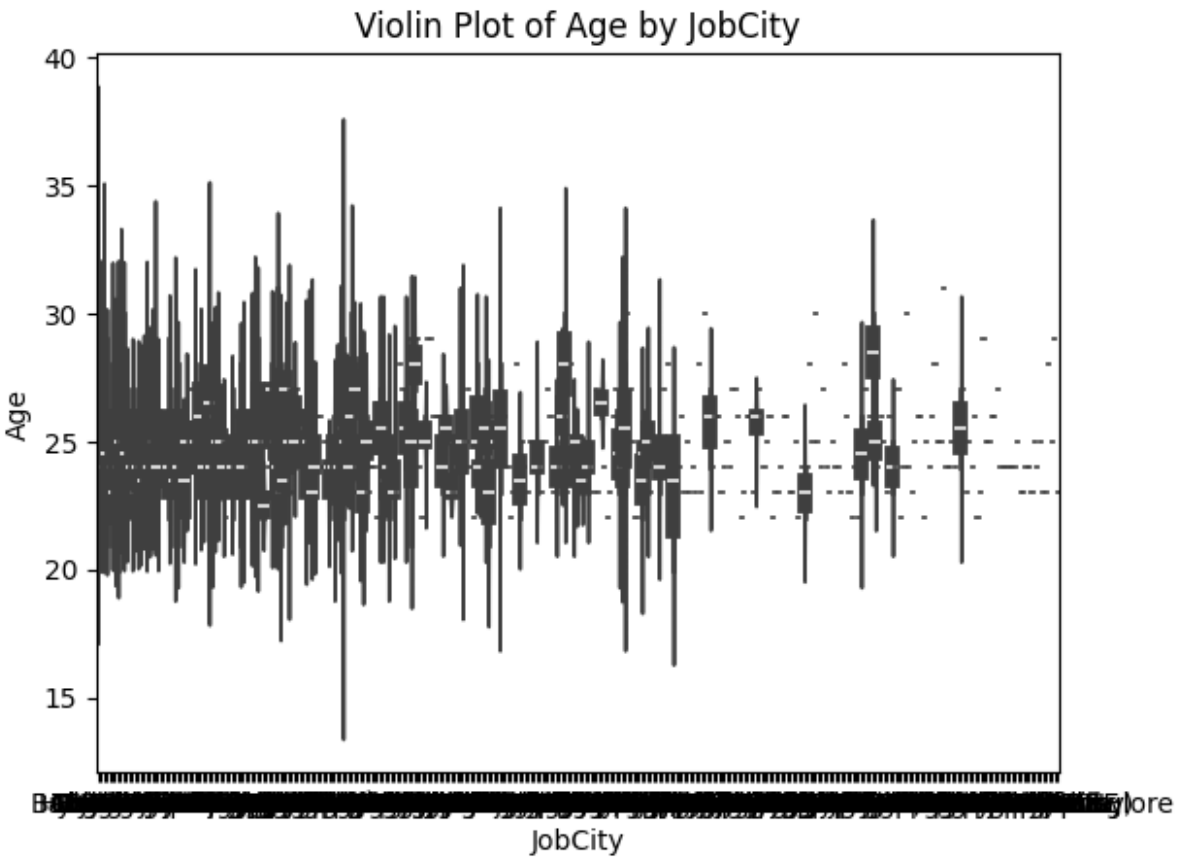

Violin Plot of conscientiousness by JobCity
conscientiousness
6
4
2
0
−2
−4
−6
JobCity


Violin Plot of agreeableness by JobCity
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
agreeableness
JobCity


Violin Plot of extraversion by JobCity
extraversion
6
4
2
0
−2
−4
−6
−8
JobCity


Violin Plot of nueroticism by JobCity
nueroticism
6
4
2
0
−2
−4
−6
−8
JobCity

Violin Plot of openness_to_experience by JobCity
openness_to_experience
6
4
2
0
−2
−4
−6
−8
−10
JobCity


Violin Plot of Age by JobCity
40
35
30
25
20
15
Age
JobCity

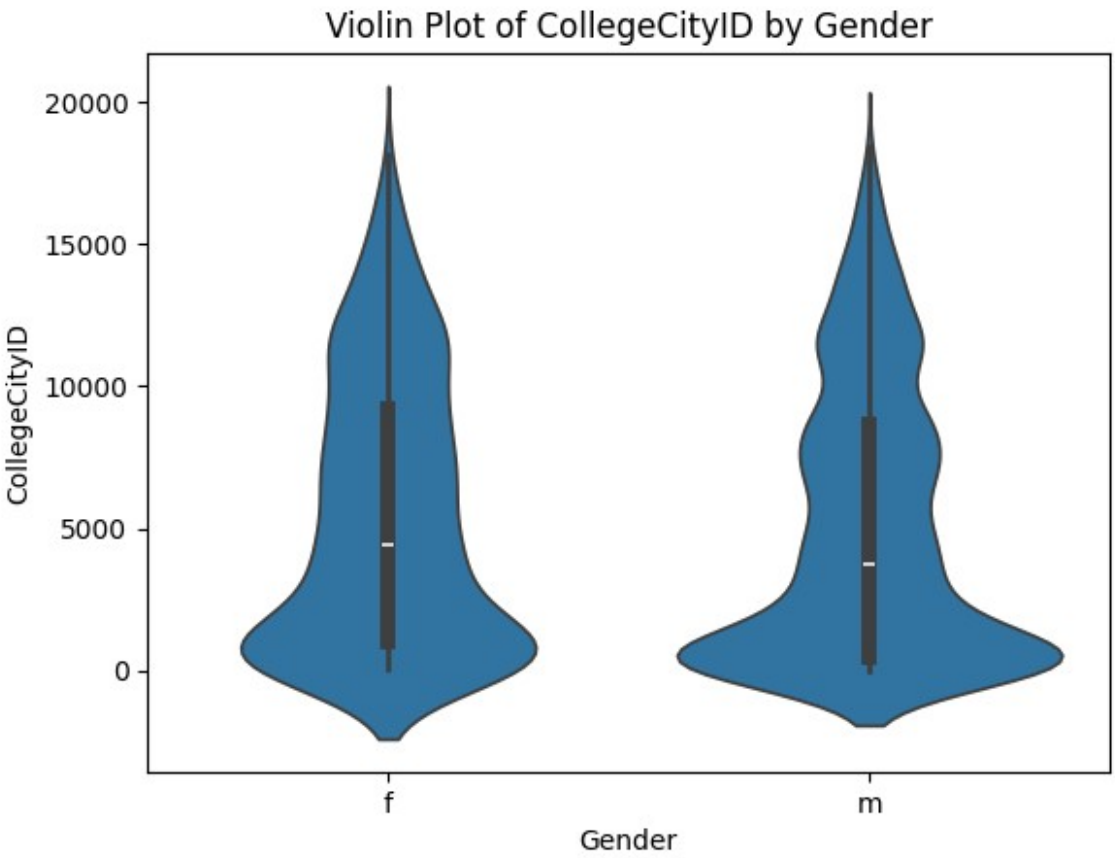
Violin Plot of Salary by Gender
1e6
4
3
2
1
0
Salary
f
m
Gender

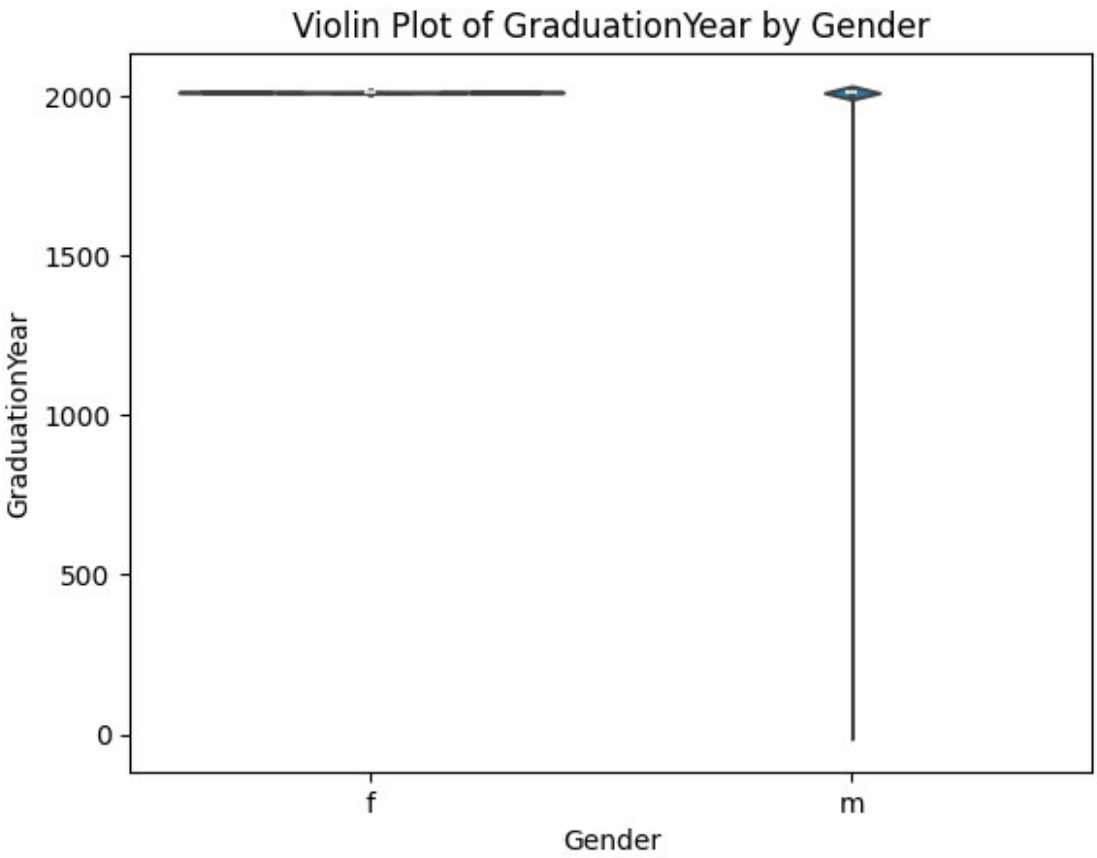
Violin Plot of 10percentage by Gender
100
90
80
70
60
50
40
10percentage
f
m
Gender

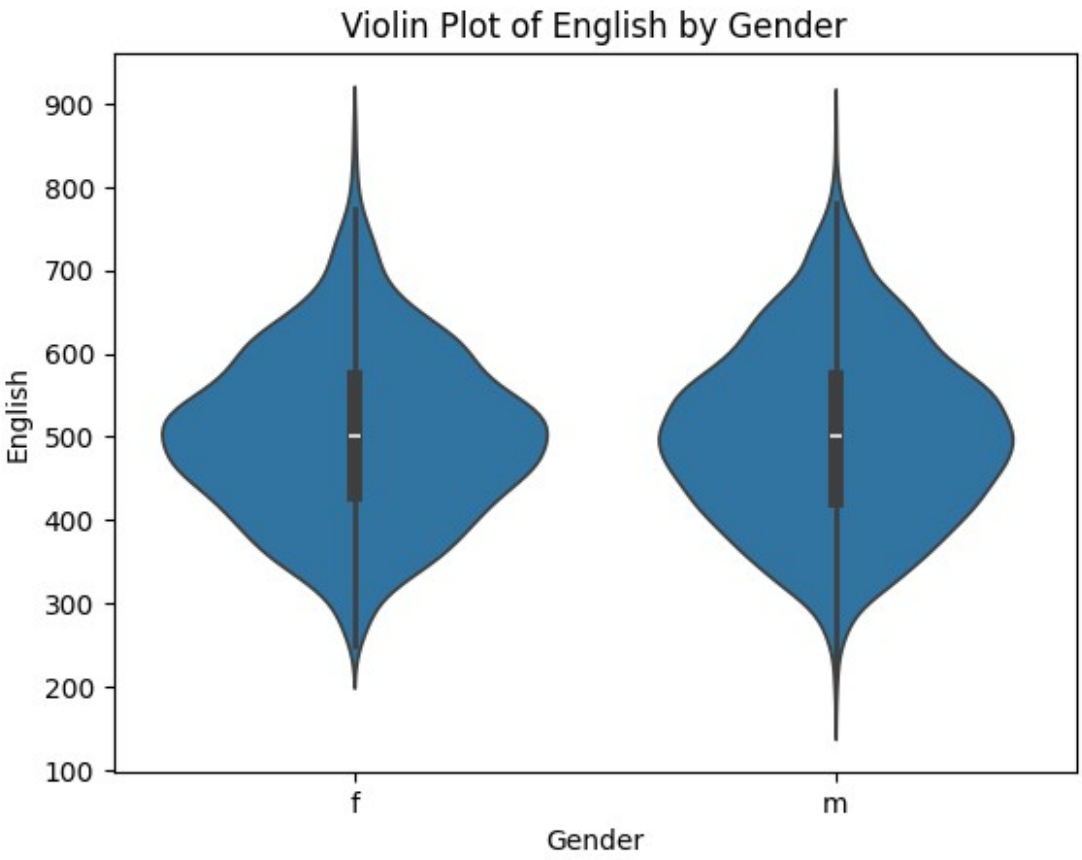
Violin Plot of 12graduation by Gender
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
12graduation
f
m
Gender

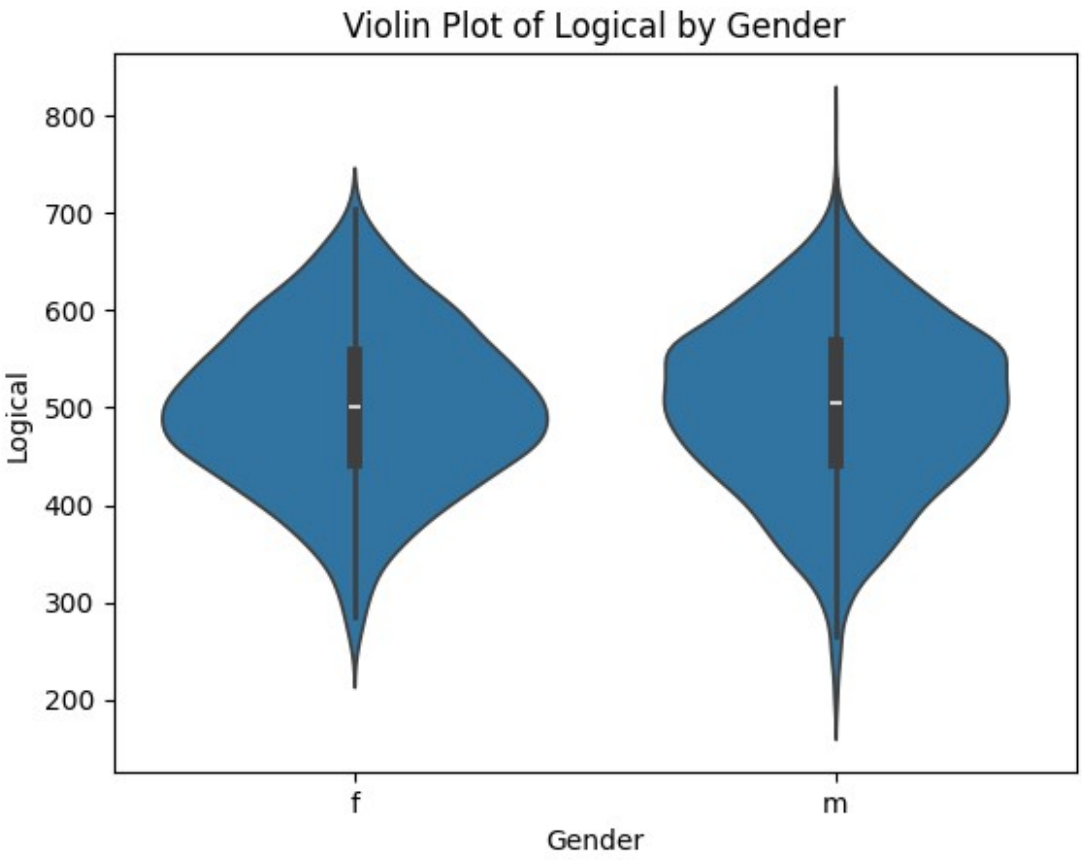
Violin Plot of 12percentage by Gender
12percentage
100
90
80
70
60
50
40
f
m
Gender

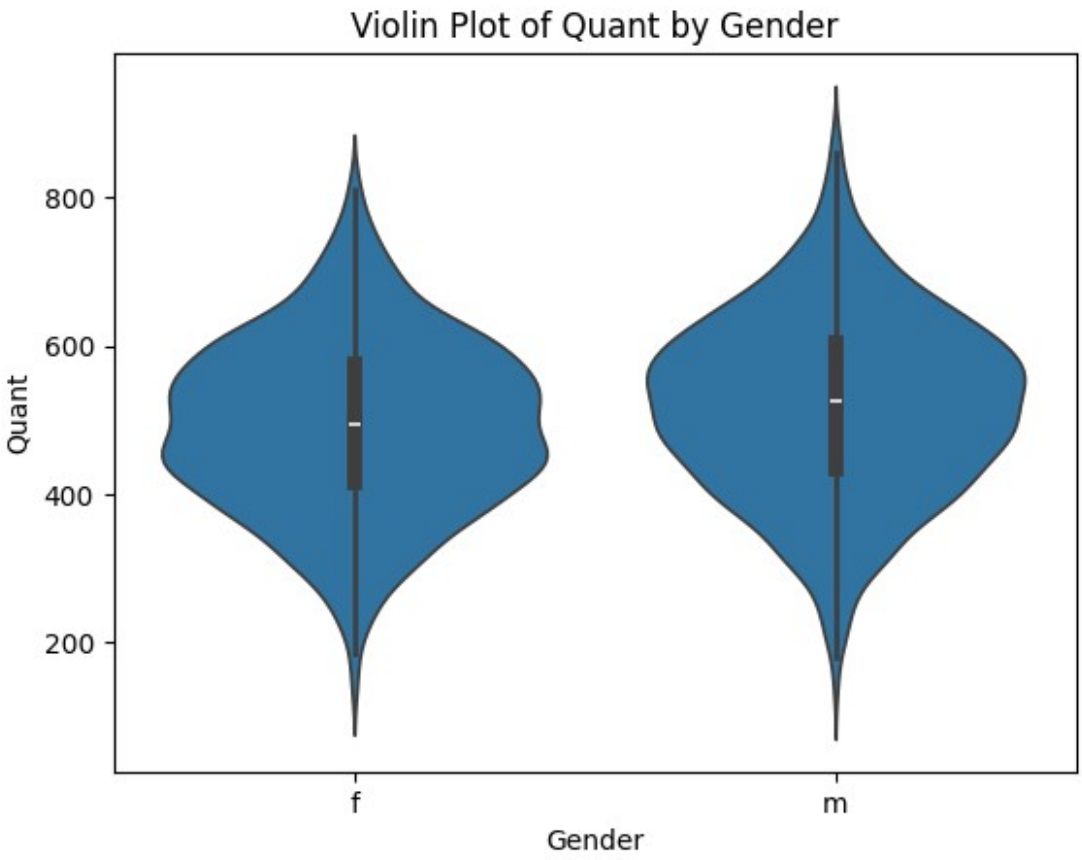
Violin Plot of collegeGPA by Gender
100
80
60
40
20
0
collegeGPA
f
m
Gender

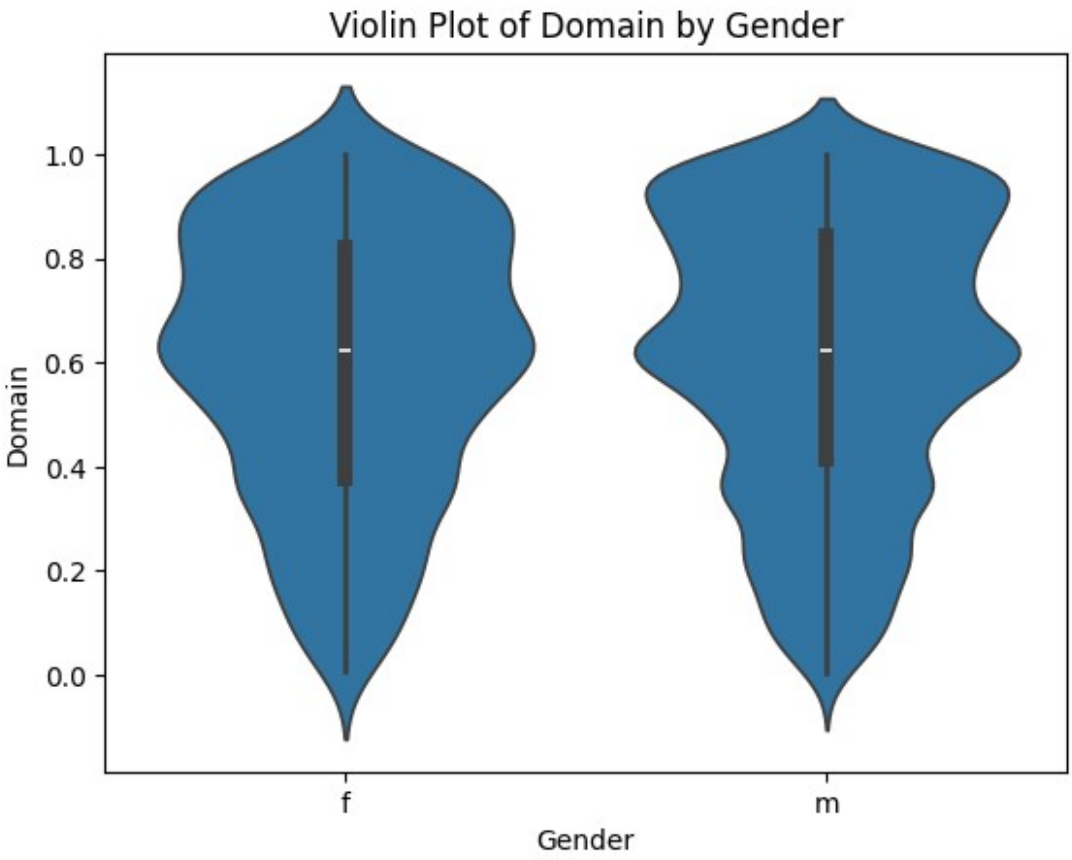

Violin Plot of CollegeCityID by Gender
20000
15000
10000
5000
0
CollegeCityID
f
m
Gender

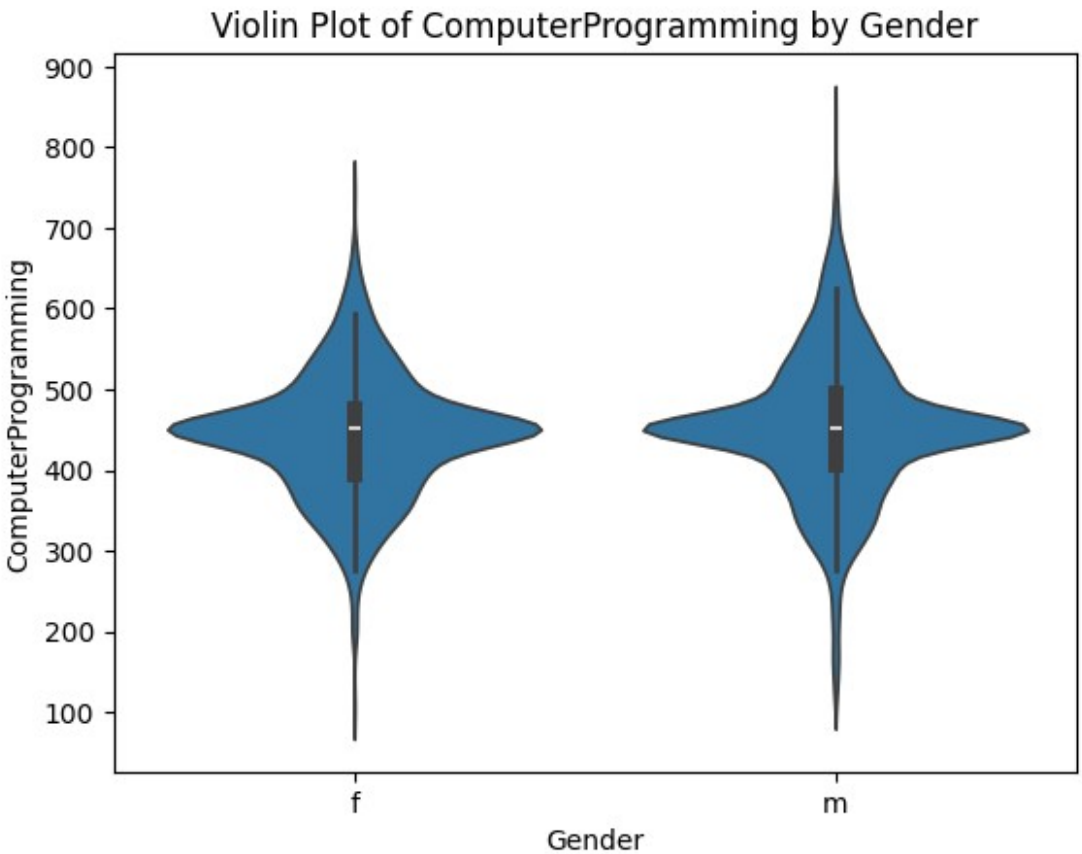

Violin Plot of GraduationYear by Gender
2000
1500
1000
500
0
GraduationYear
f
m
Gender

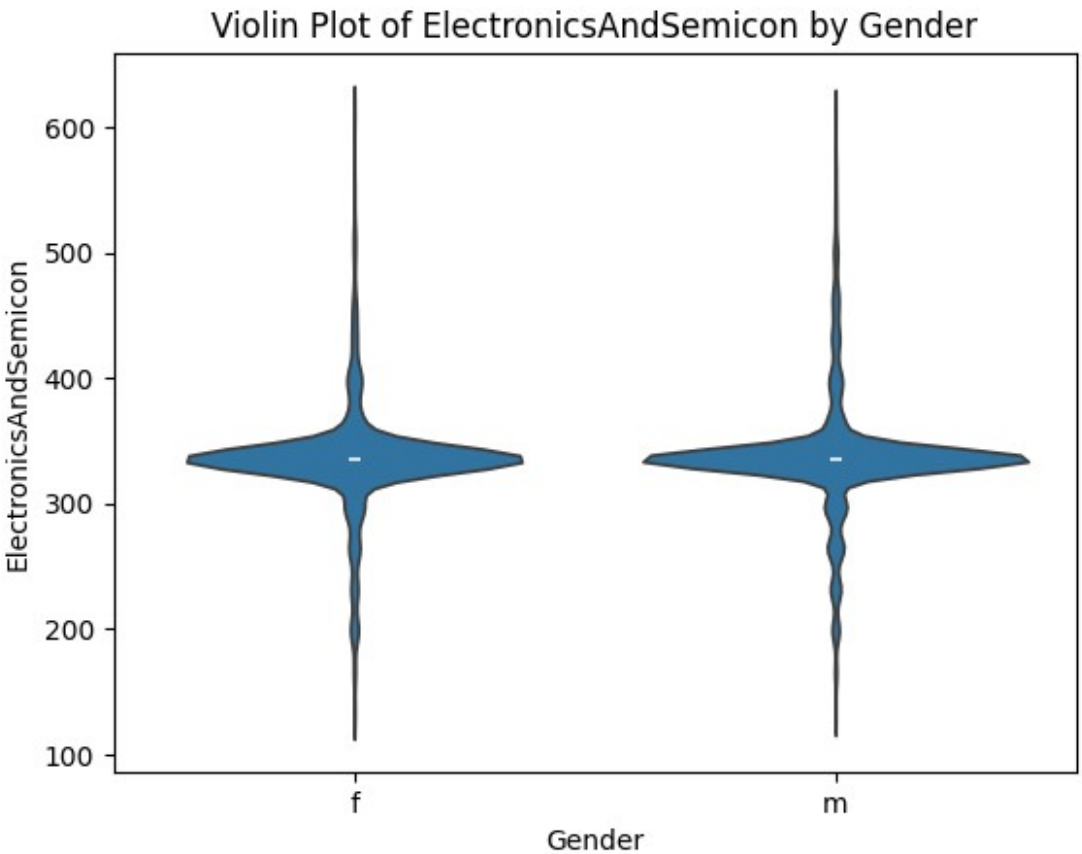

Violin Plot of English by Gender
English
900
800
700
600
500
400
300
200
100
f
m
Gender

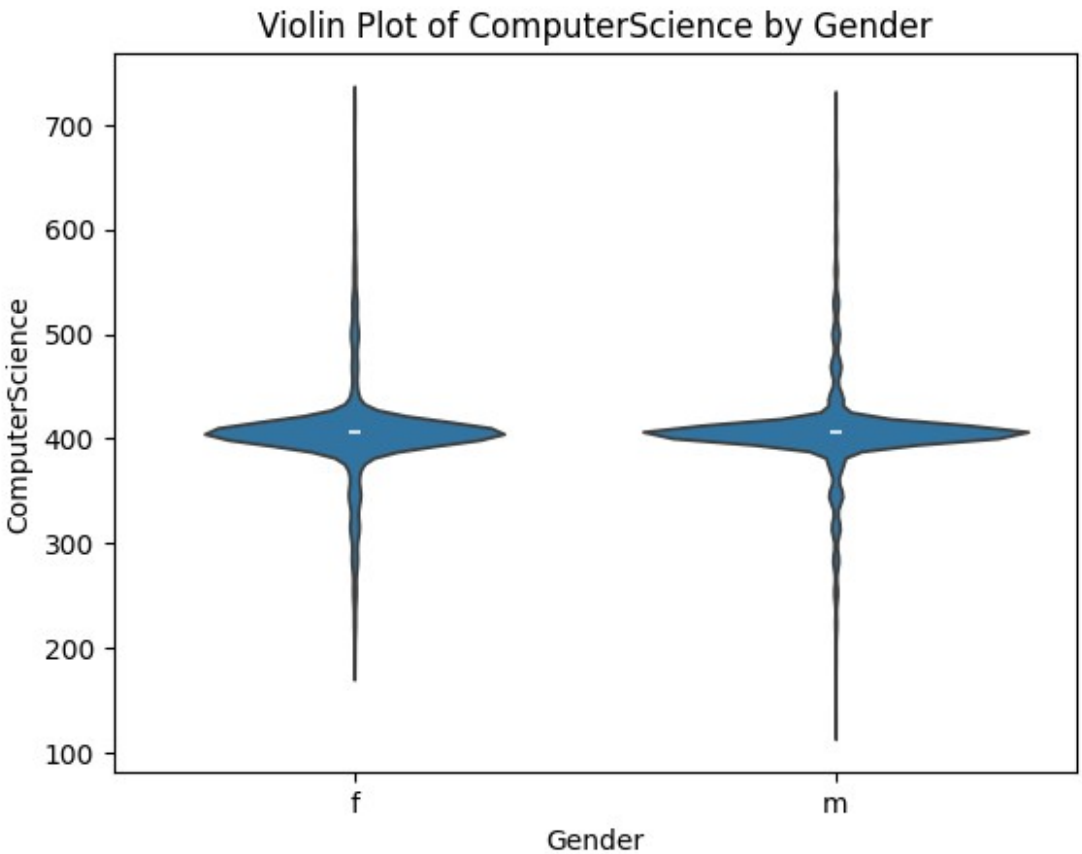

Violin Plot of Logical by Gender
800
700
600
500
400
300
200
Logical
f
m
Gender


Violin Plot of Quant by Gender
800
600
400
200
Quant
f
m
Gender

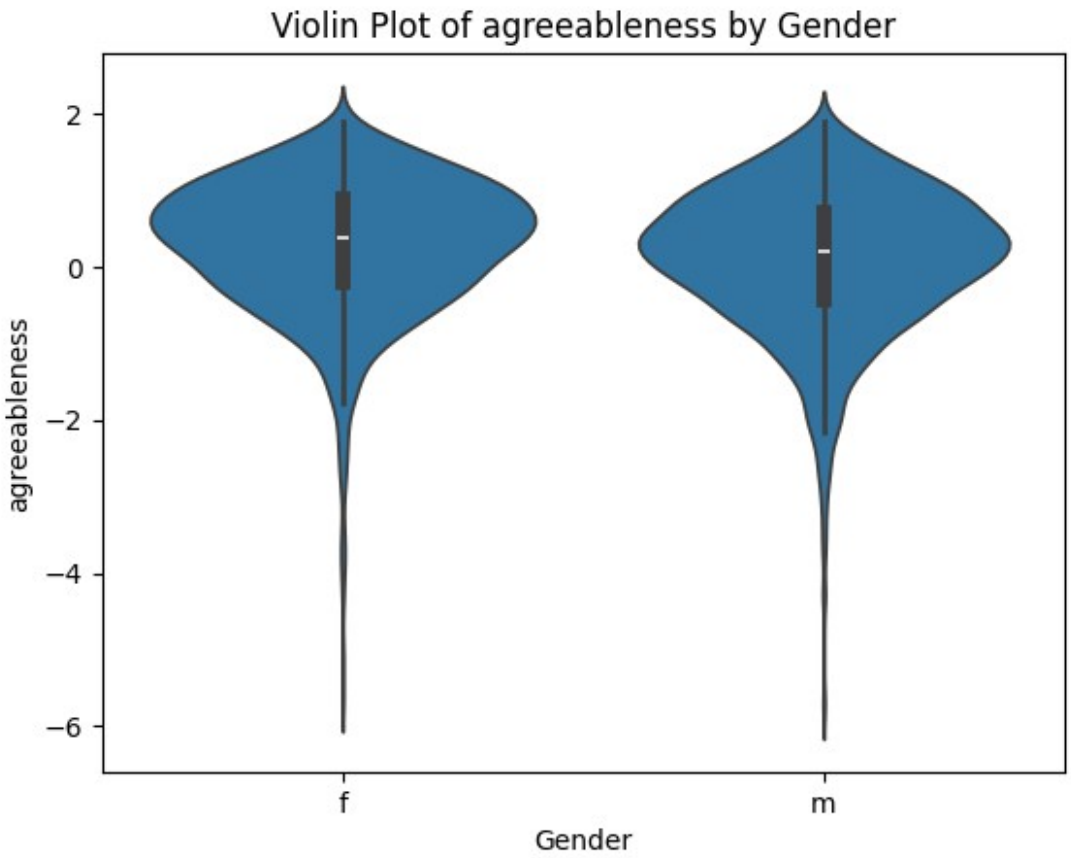

Violin Plot of Domain by Gender
Domain
1.0
0.8
0.6
0.4
0.2
0.0
f
m
Gender

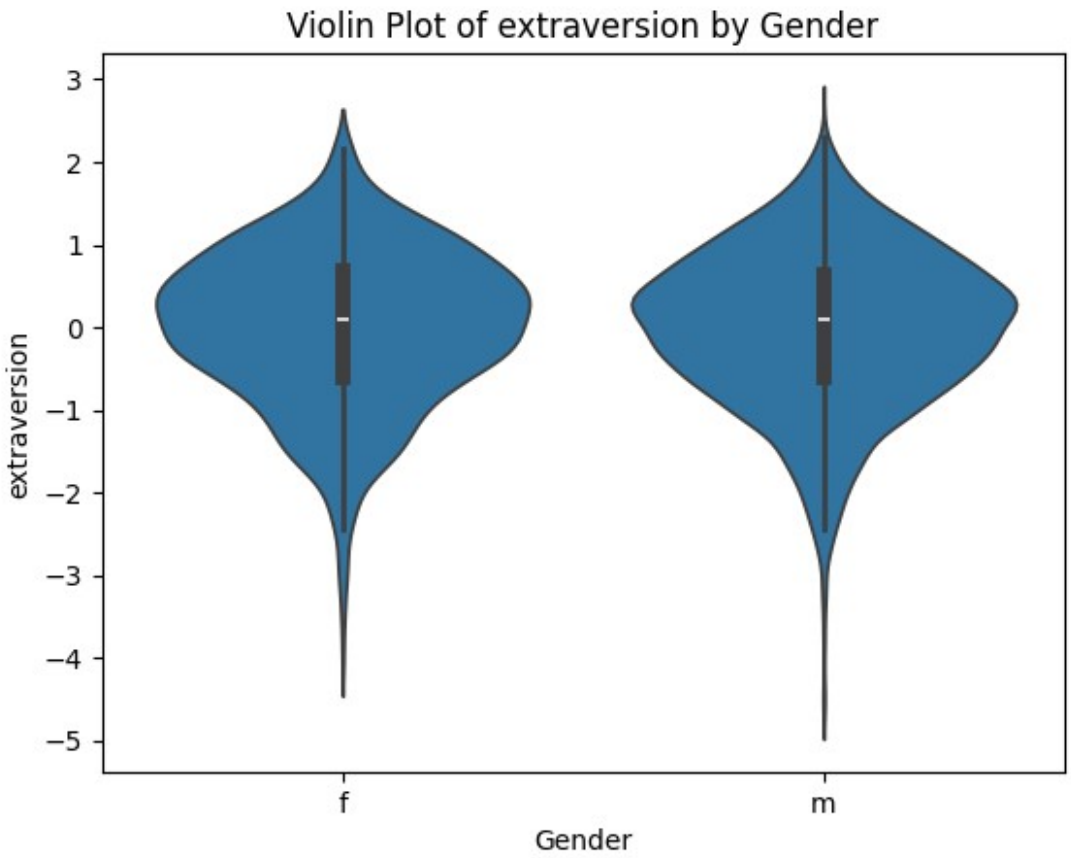

Violin Plot of ComputerProgramming by Gender
ComputerProgramming
900
800
700
600
500
400
300
200
100
f
m
Gender

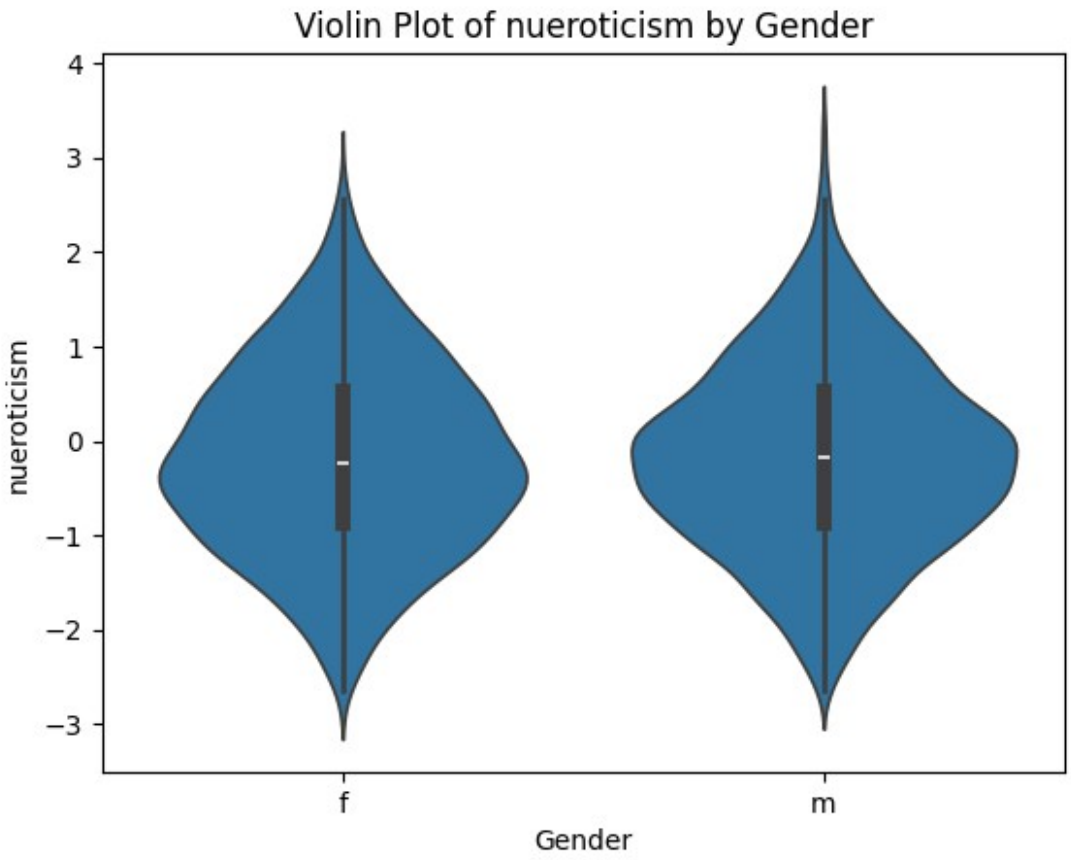

Violin Plot of ElectronicsAndSemicon by Gender
ElectronicsAndSemicon
600
500
400
300
200
100
f
m
Gender

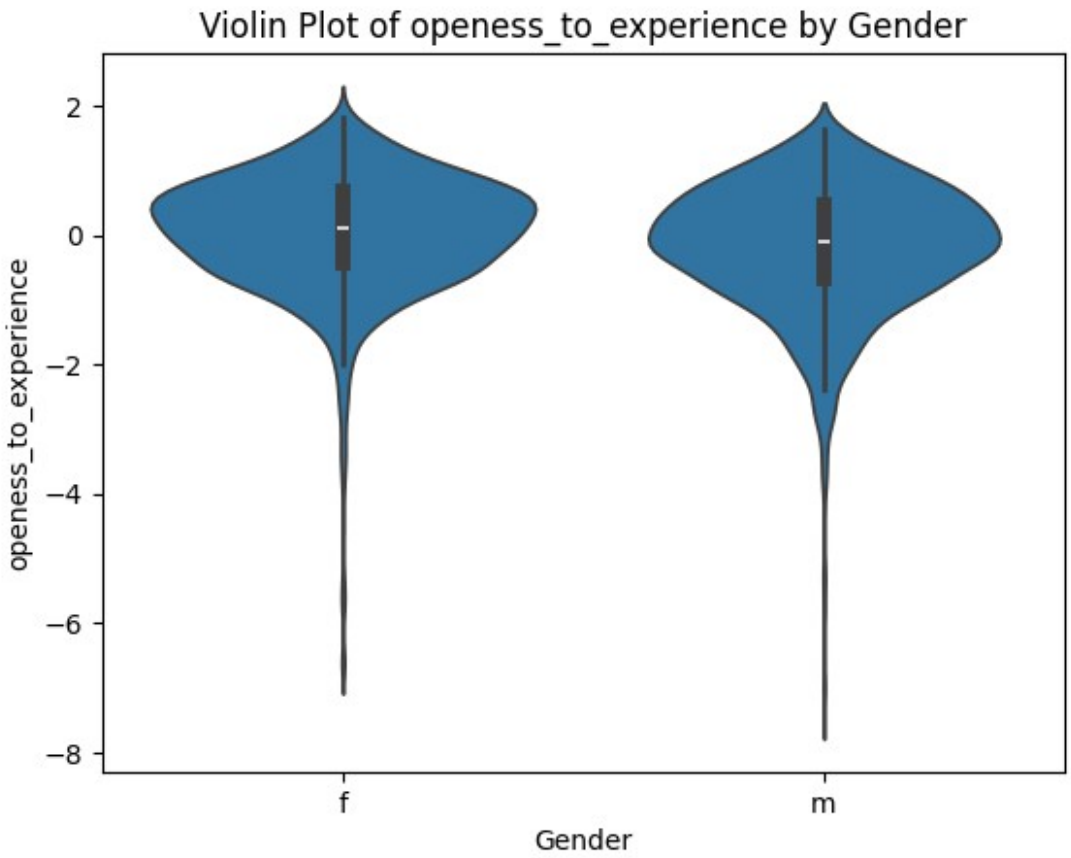

Violin Plot of ComputerScience by Gender
ComputerScience
700
600
500
400
300
200
100
f
m
Gender

Violin Plot of conscientiousness by Gender
conscientiousness
2
1
0
−1
−2
−3
−4
f
m
Gender

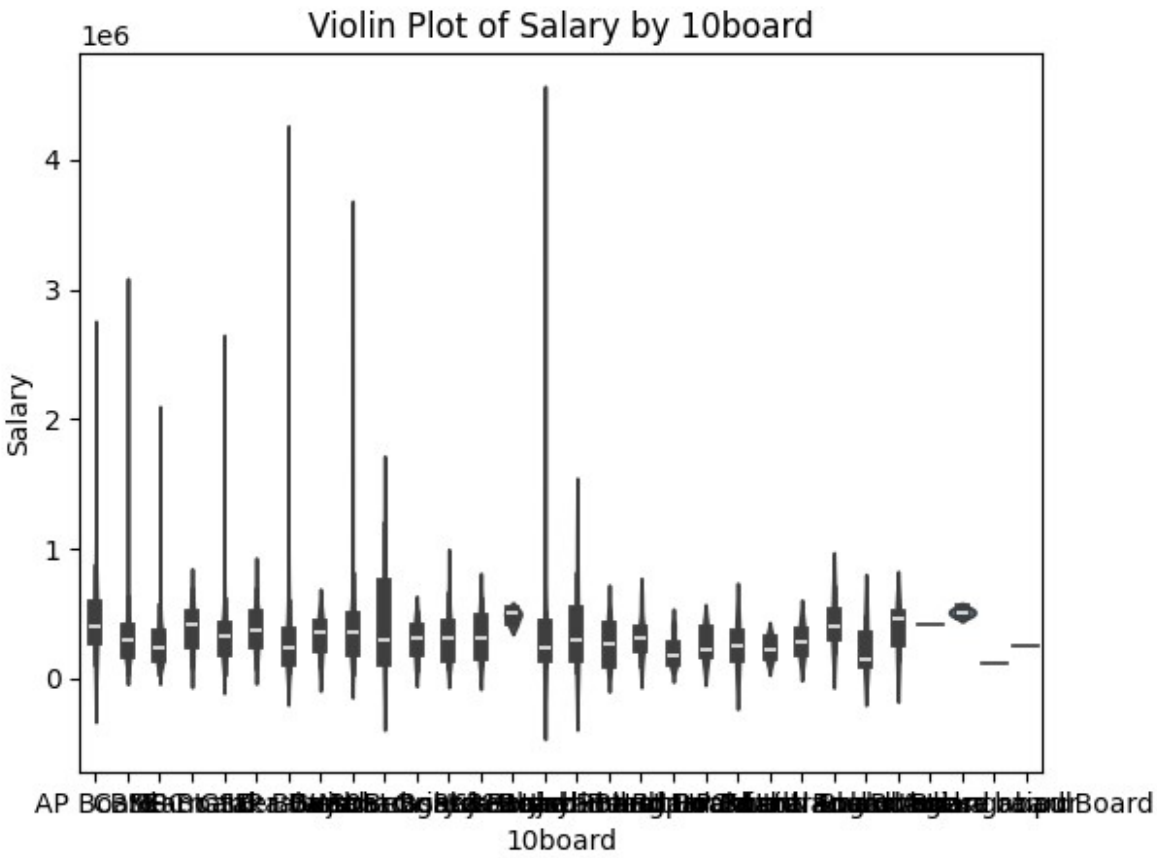

Violin Plot of agreeableness by Gender
2
0
−2
−4
−6
agreeableness
f
m
Gender

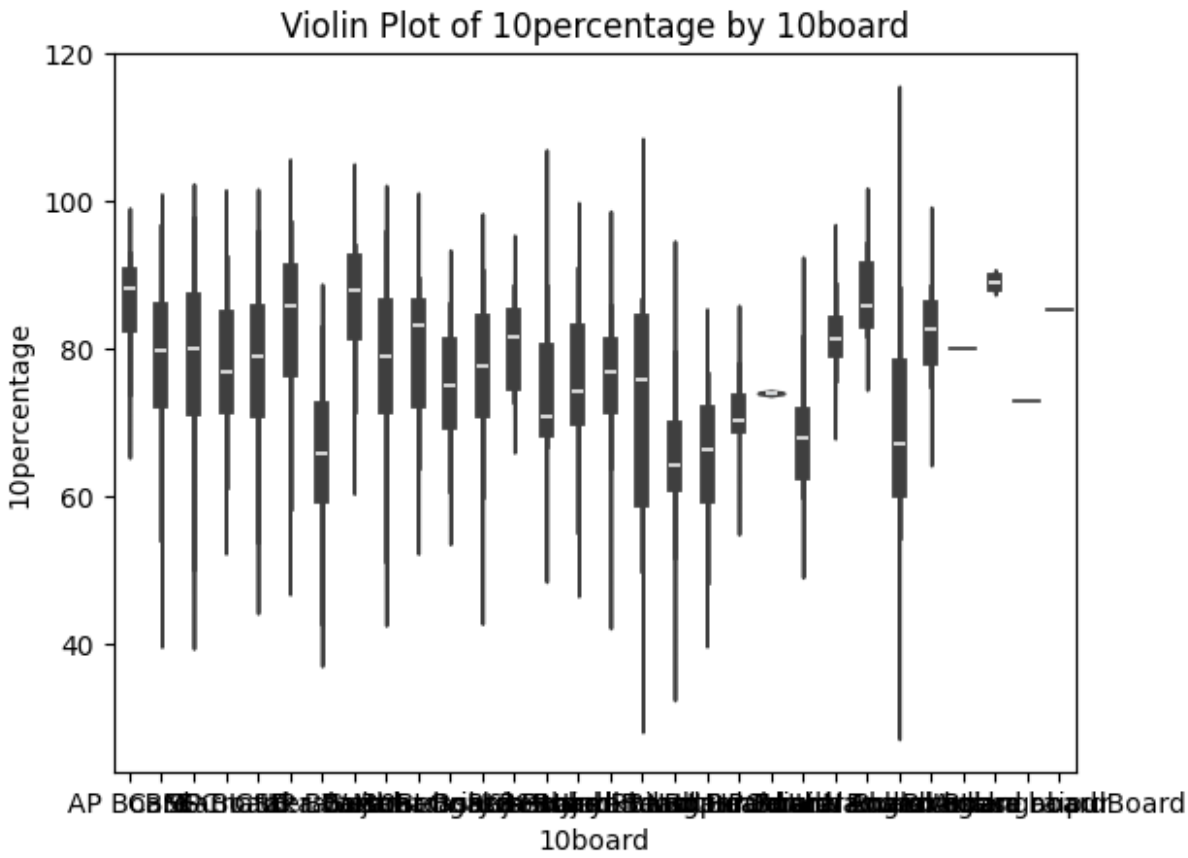

Violin Plot of extraversion by Gender
extraversion
3
2
1
0
−1
−2
−3
−4
−5
f
m
Gender

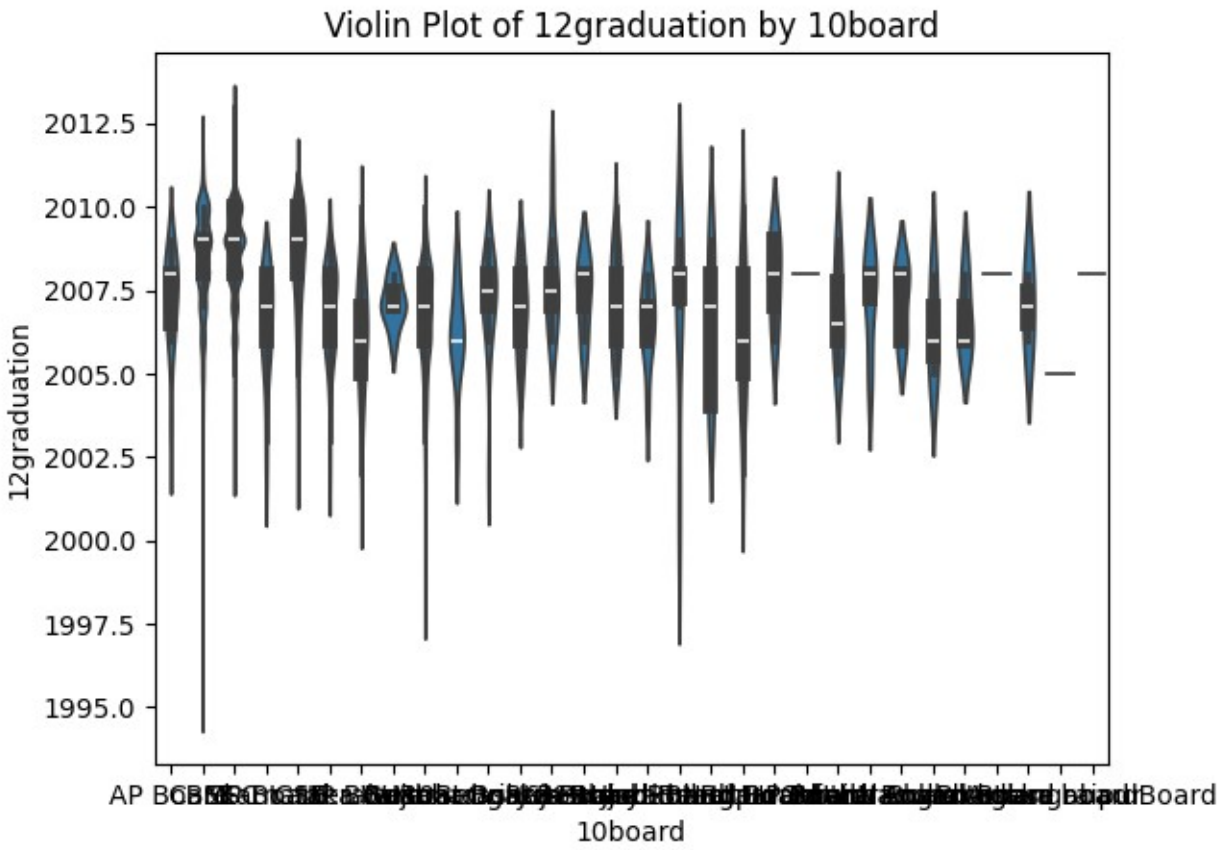

Violin Plot of nueroticism by Gender
nueroticism
4
3
2
1
0
−1
−2
−3
f
m
Gender

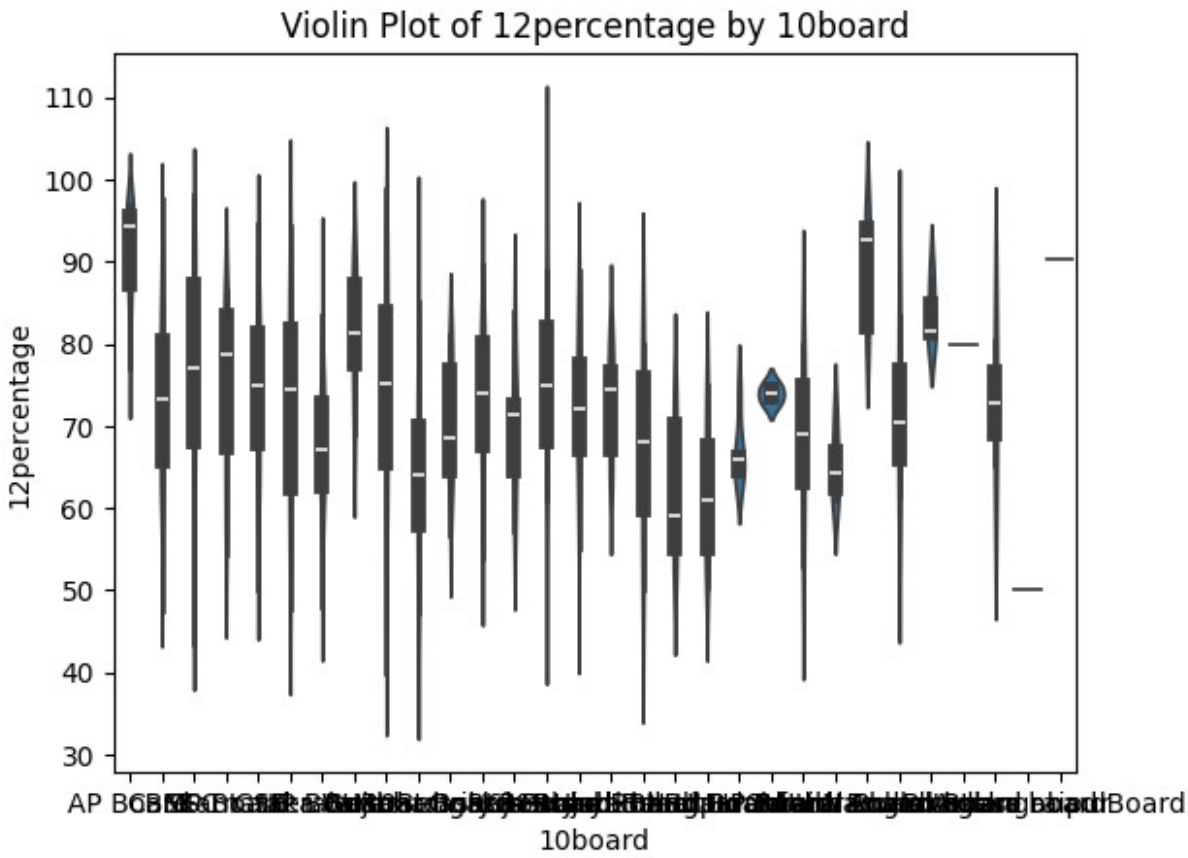

Violin Plot of openess_to_experience by Gender
2
0
−2
−4
−6
−8
openess_to_experience
f
m
Gender

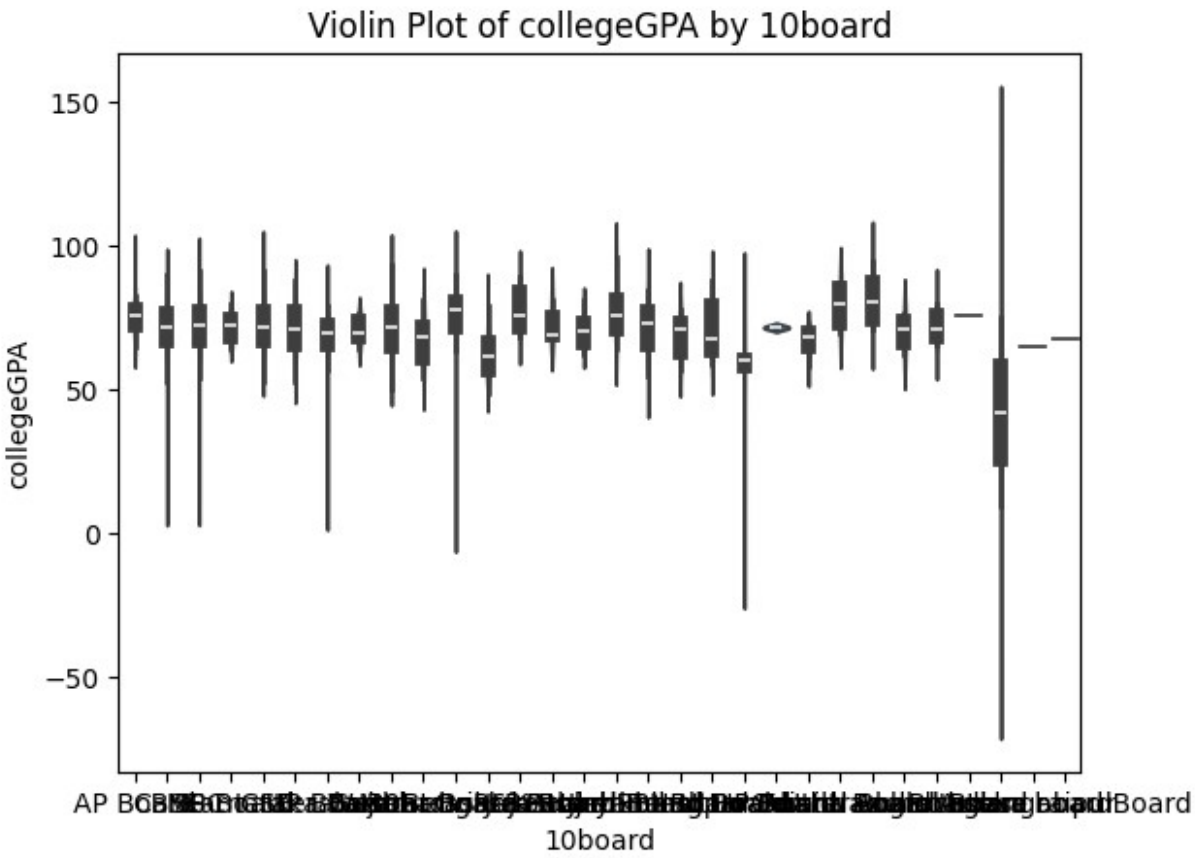
Violin Plot of Age by Gender
Age
35
30
25
20
f
m
Gender

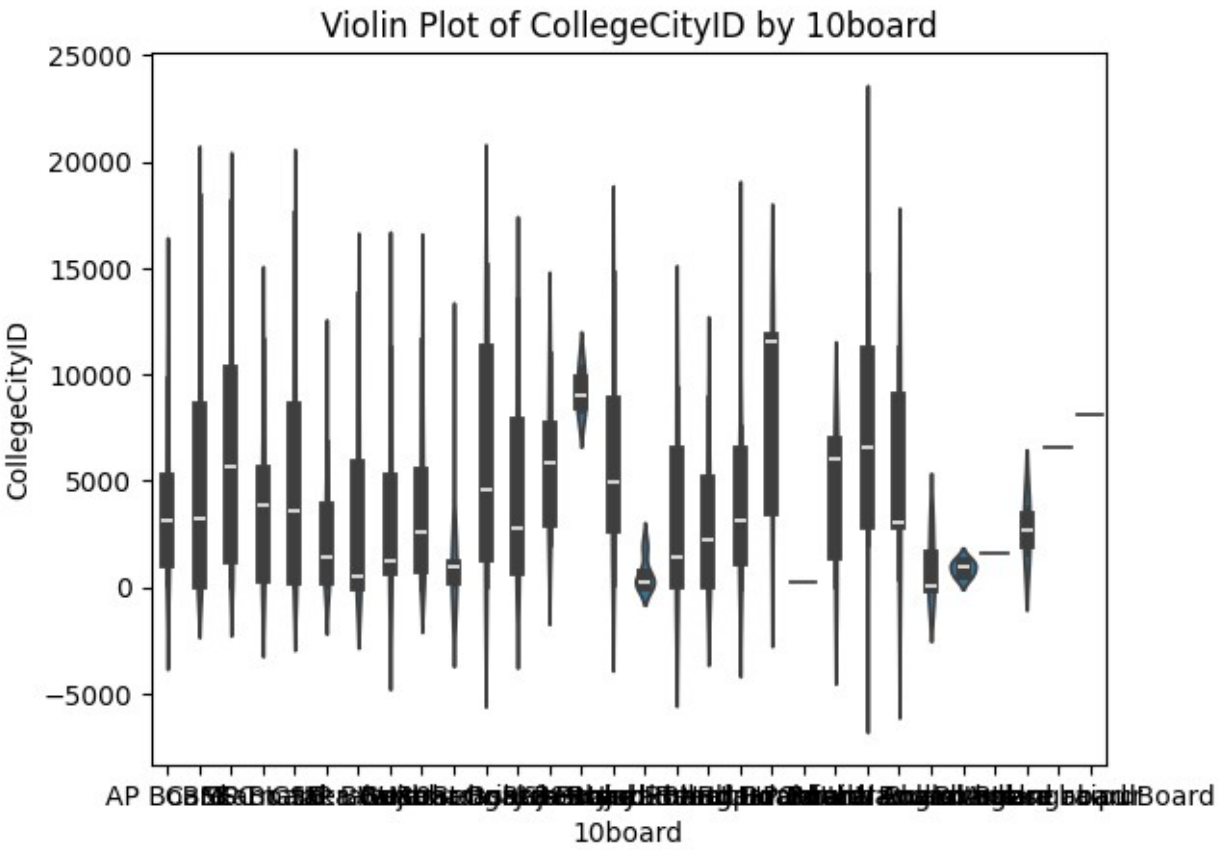

Violin Plot of Salary by 10board
1e6
4
3
2
1
0
Salary
10board

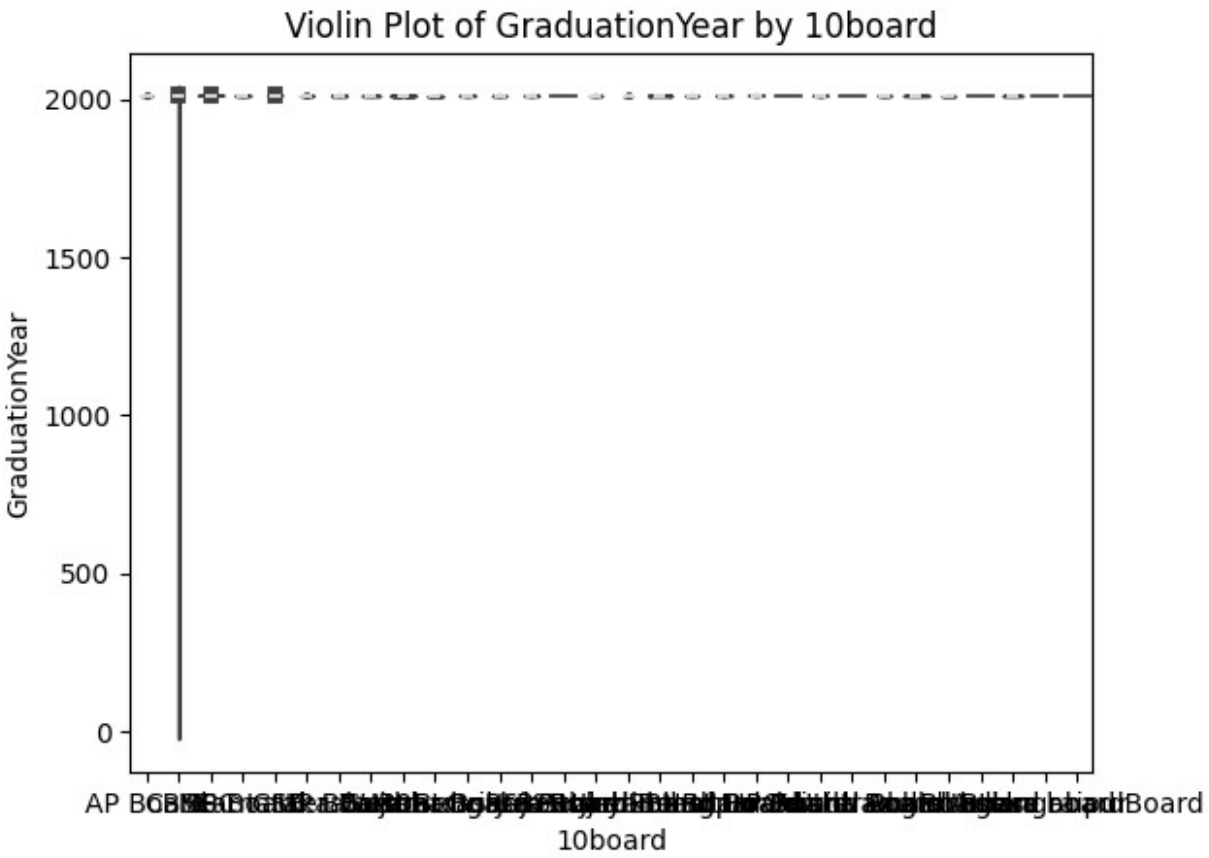

Violin Plot of 10percentage by 10board
120
100
80
60
40
10percentage
AP Board
10board

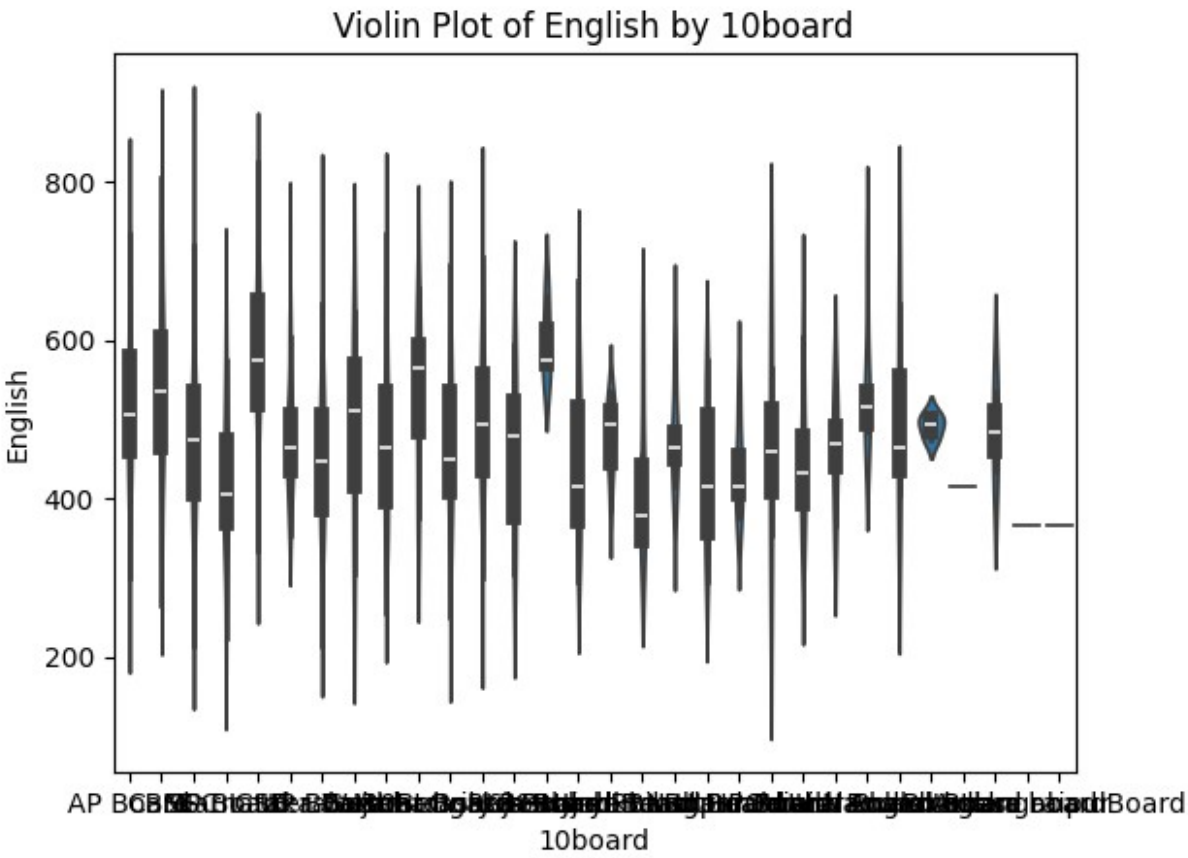

Violin Plot of 12graduation by 10board
12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
10board

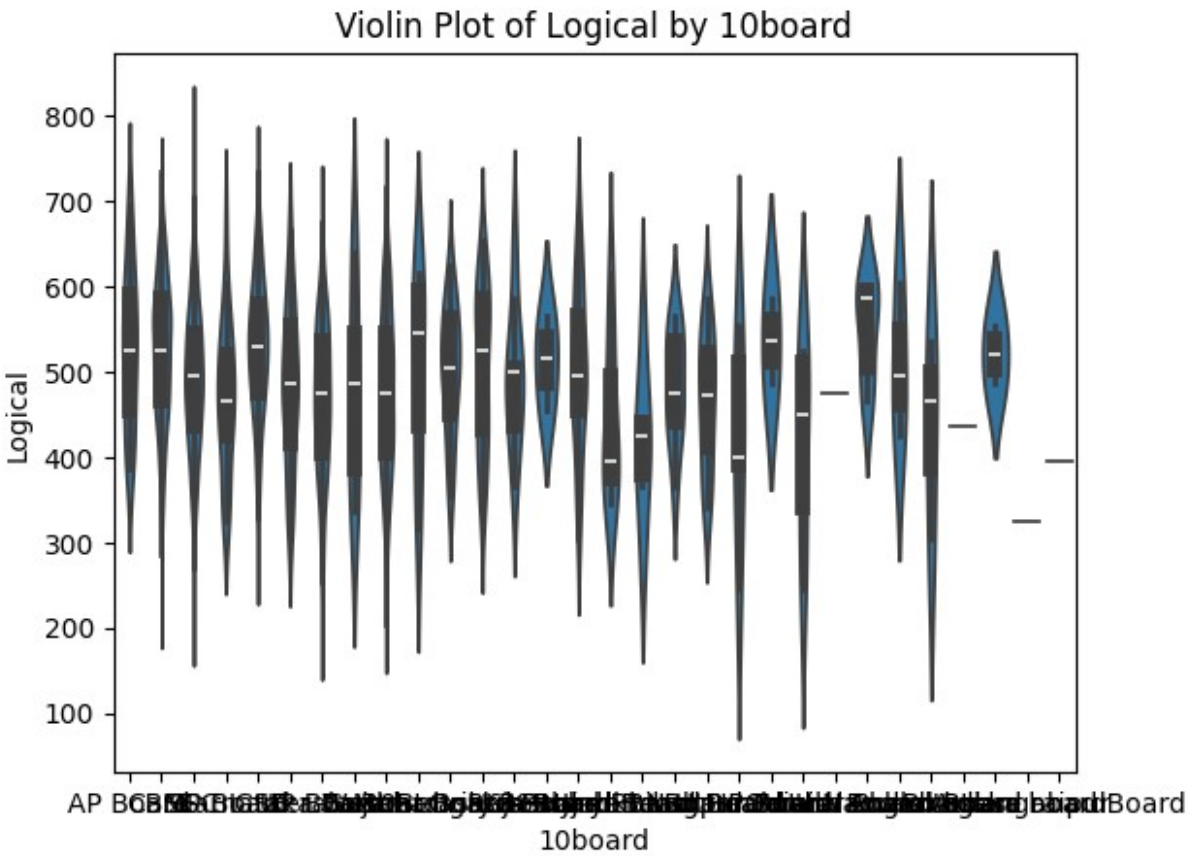

Violin Plot of 12percentage by 10board
12percentage
110
100
90
80
70
60
50
40
30
10board

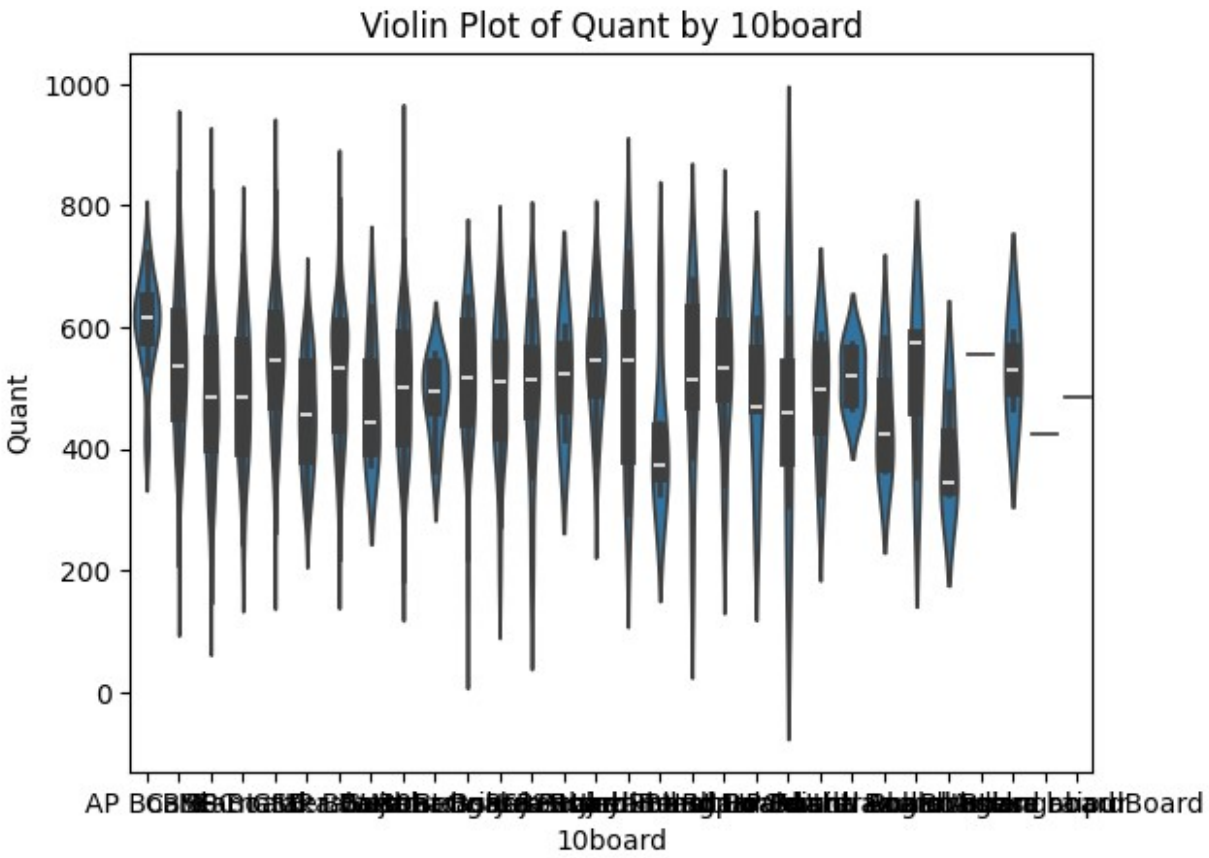

Violin Plot of collegeGPA by 10board
150
100
50
0
−50
collegeGPA
10board

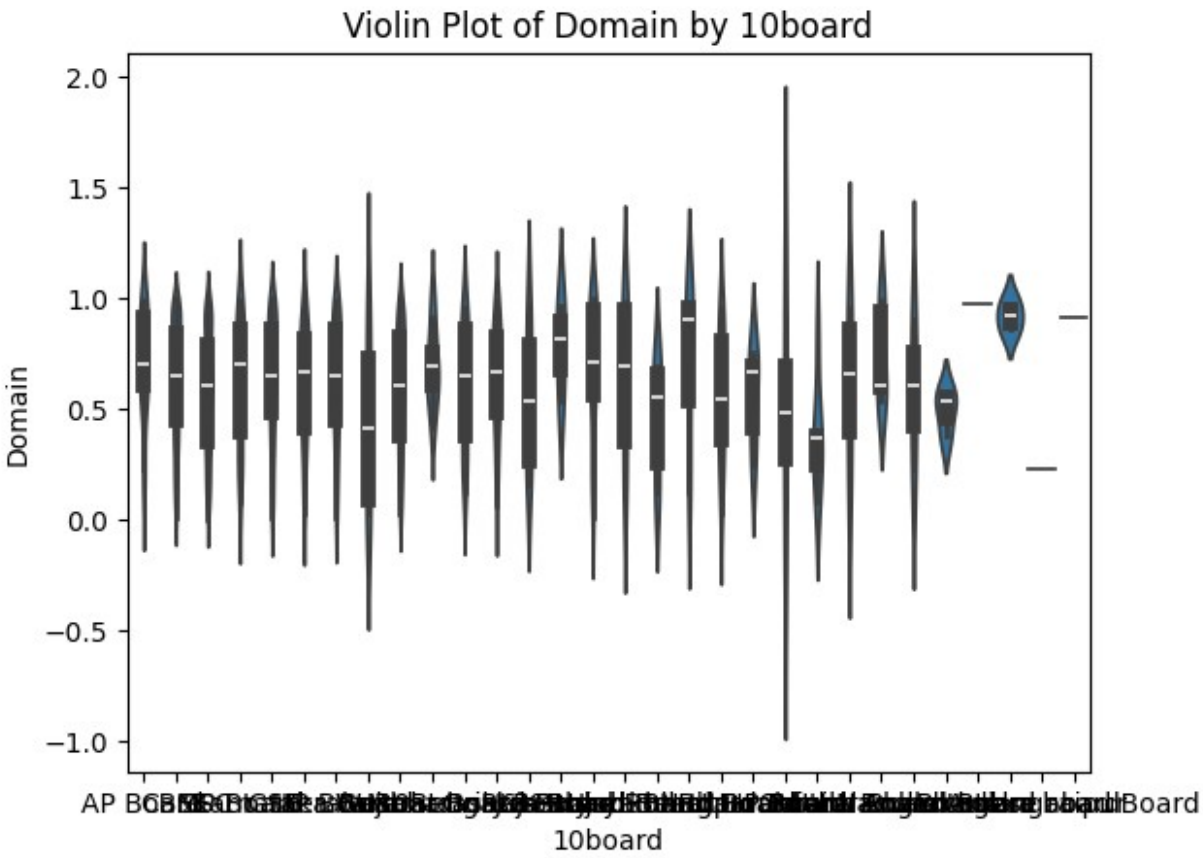

Violin Plot of CollegeCityID by 10board
25000
20000
15000
10000
5000
0
−5000
CollegeCityID
10board

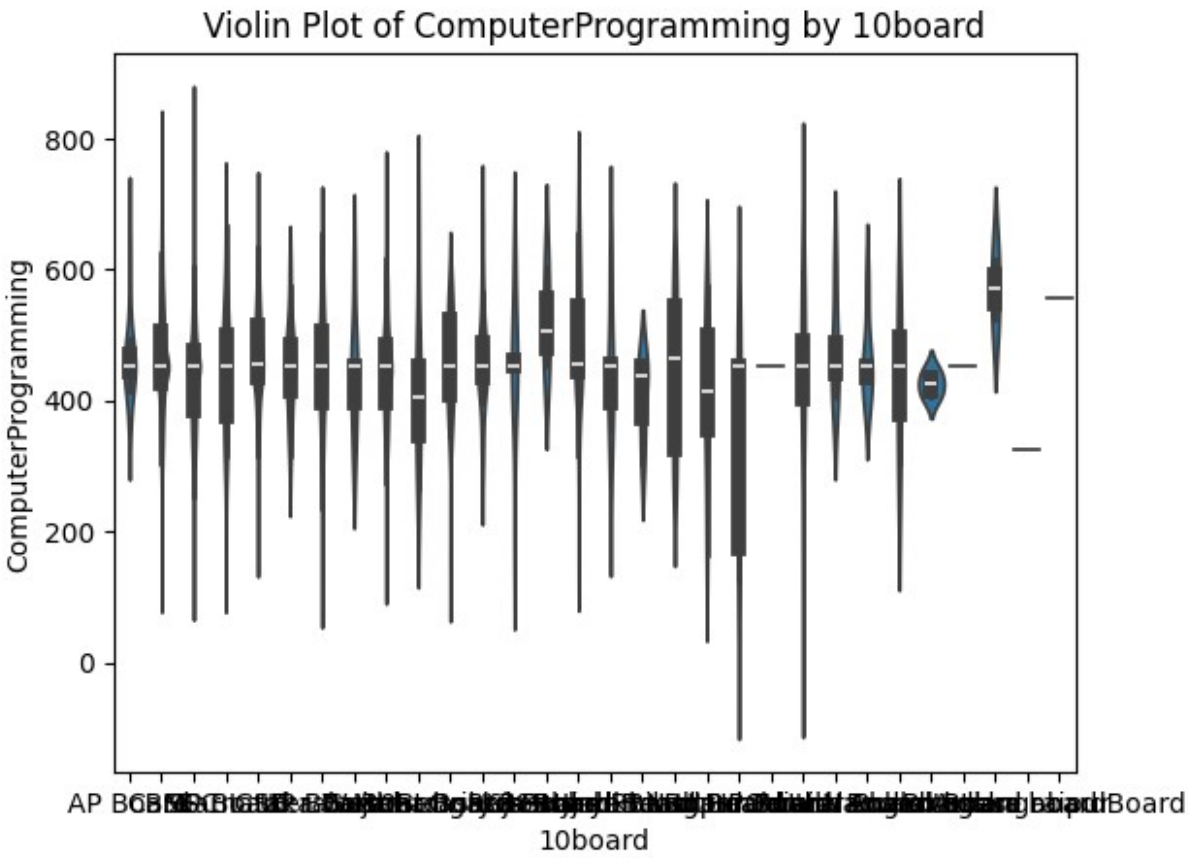

Violin Plot of GraduationYear by 10board
GraduationYear
2000
1500
1000
500
0
AP Board
10board

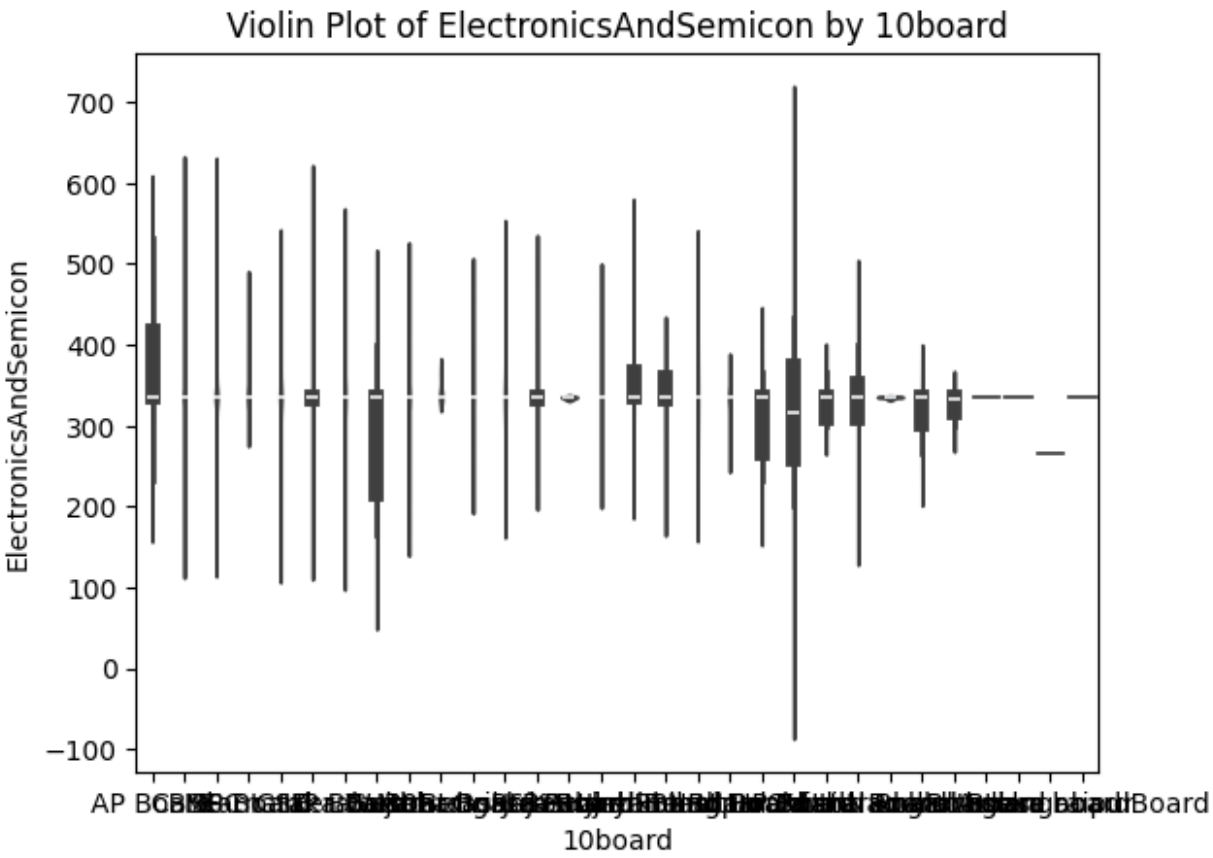

Violin Plot of English by 10board
English
800
600
400
200
10board

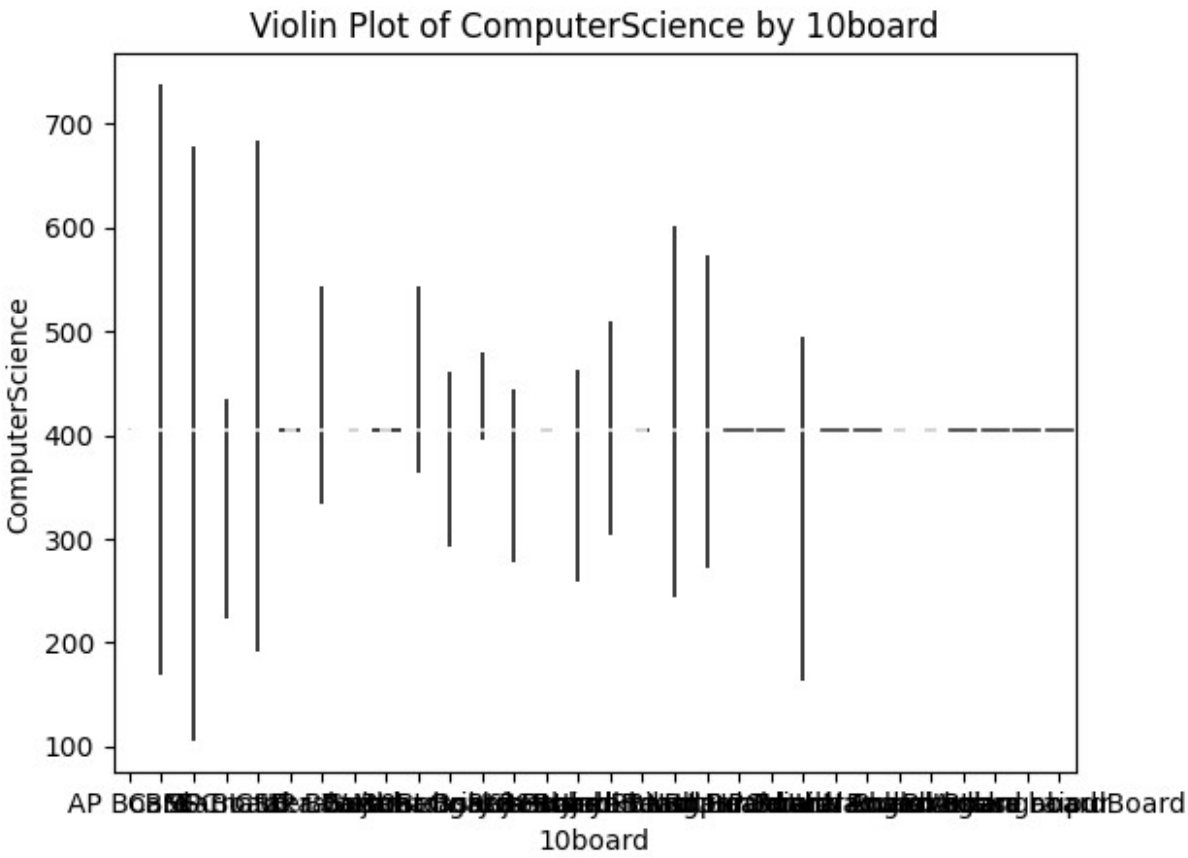

Violin Plot of Logical by 10board
Logical
800
700
600
500
400
300
200
100
10board

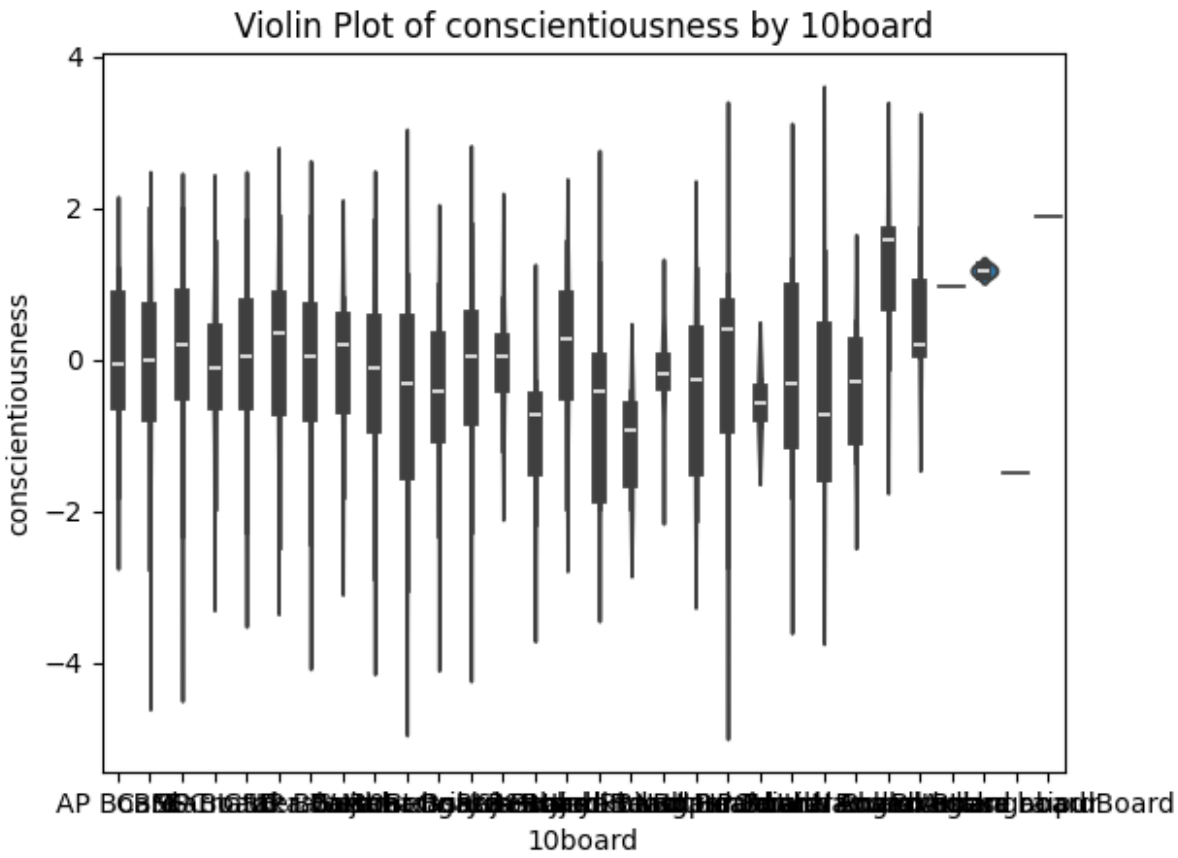

Violin Plot of Quant by 10board
Quant
1000
800
600
400
200
0
10board

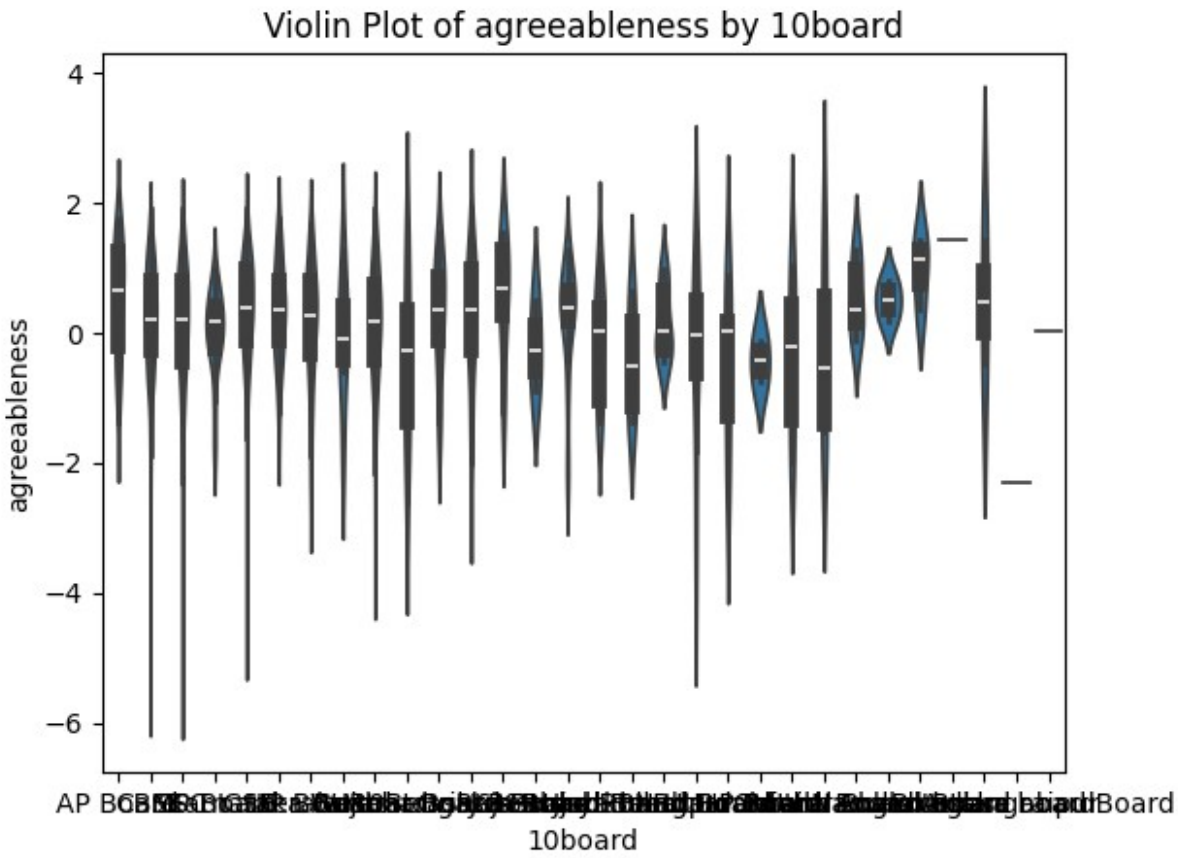

Violin Plot of Domain by 10board
Domain
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
AP Board
10board

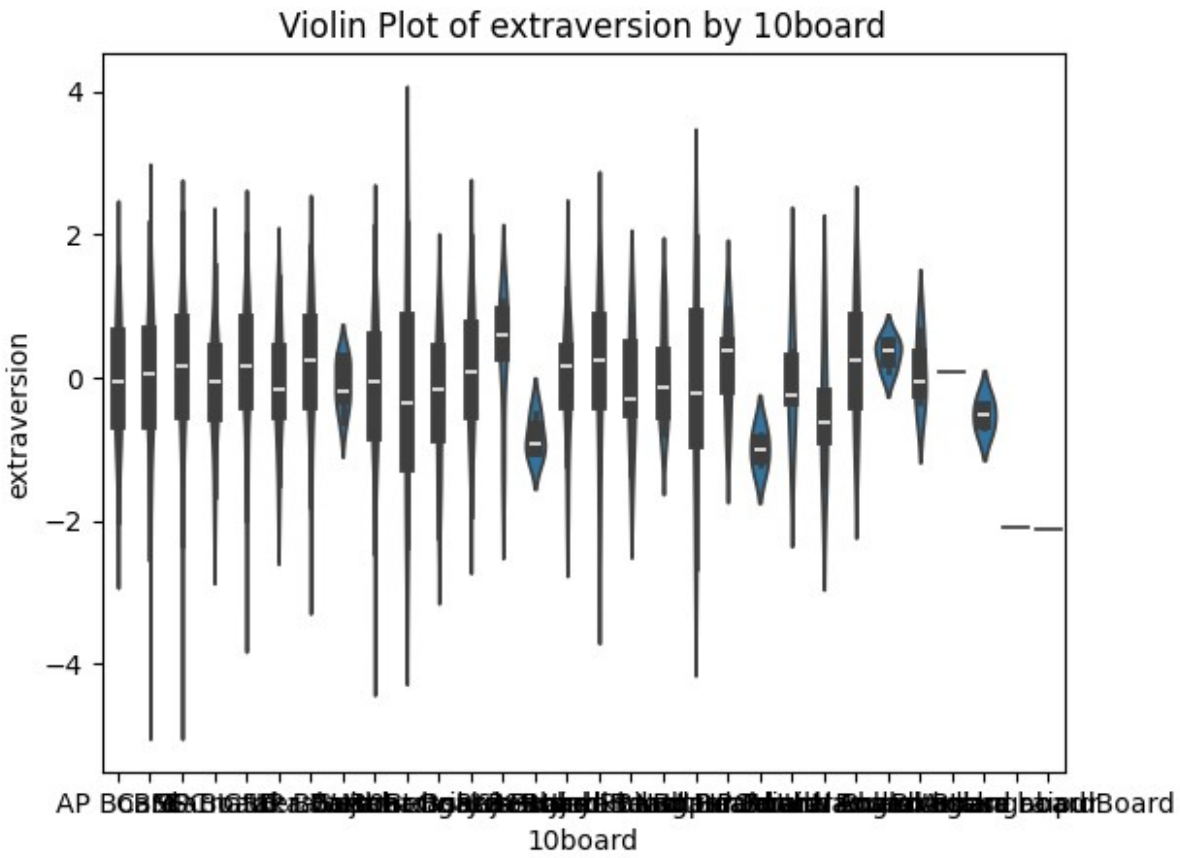

Violin Plot of ComputerProgramming by 10board
ComputerProgramming
800
600
400
200
0
AP Board
10board

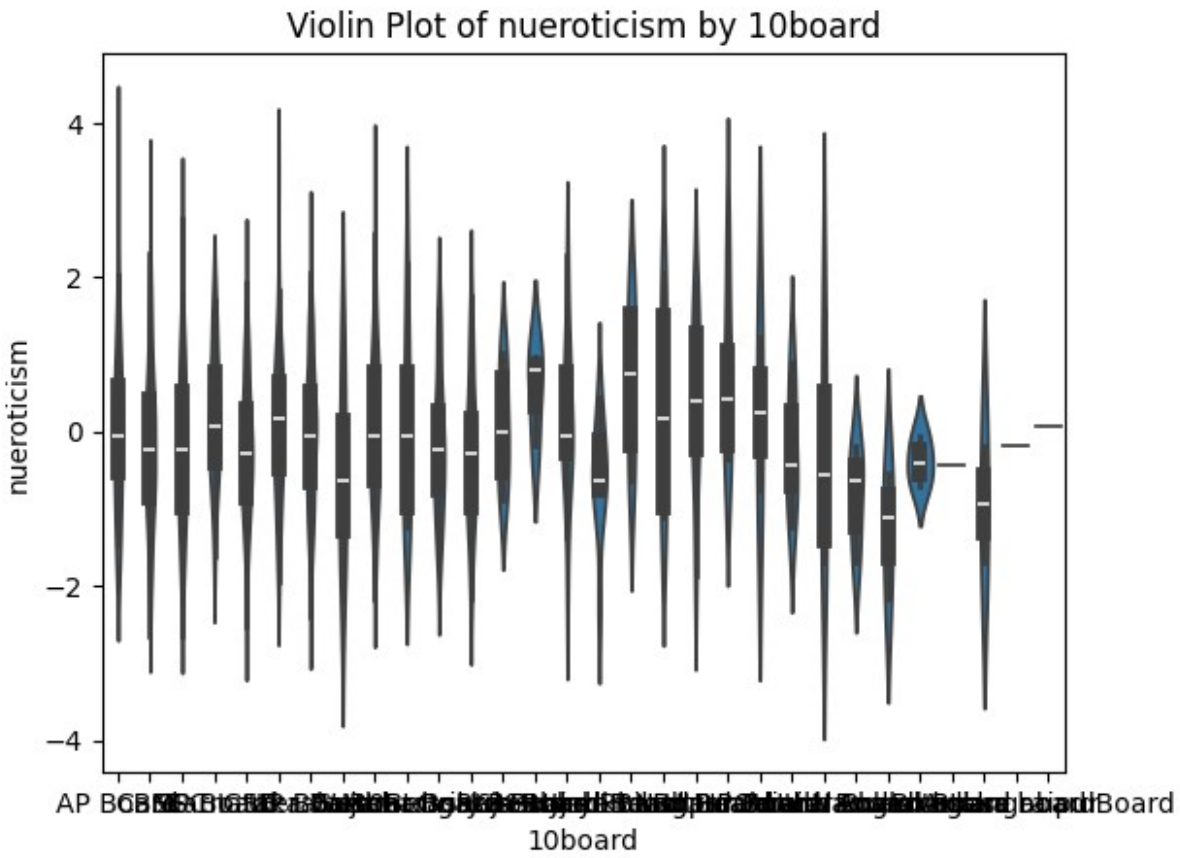

Violin Plot of ElectronicsAndSemicon by 10board
ElectronicsAndSemicon
700
600
500
400
300
200
100
0
−100
10board

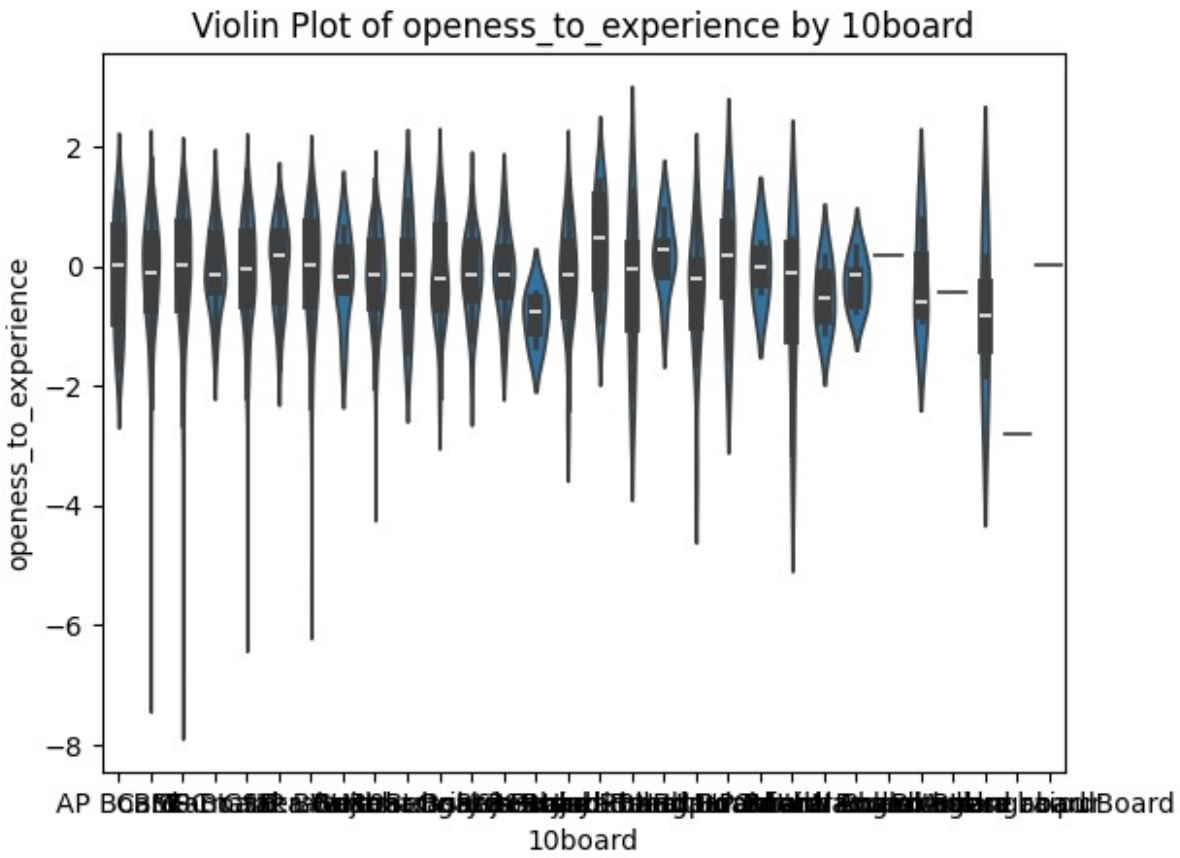

Violin Plot of ComputerScience by 10board
ComputerScience
700
600
500
400
300
200
100
10board

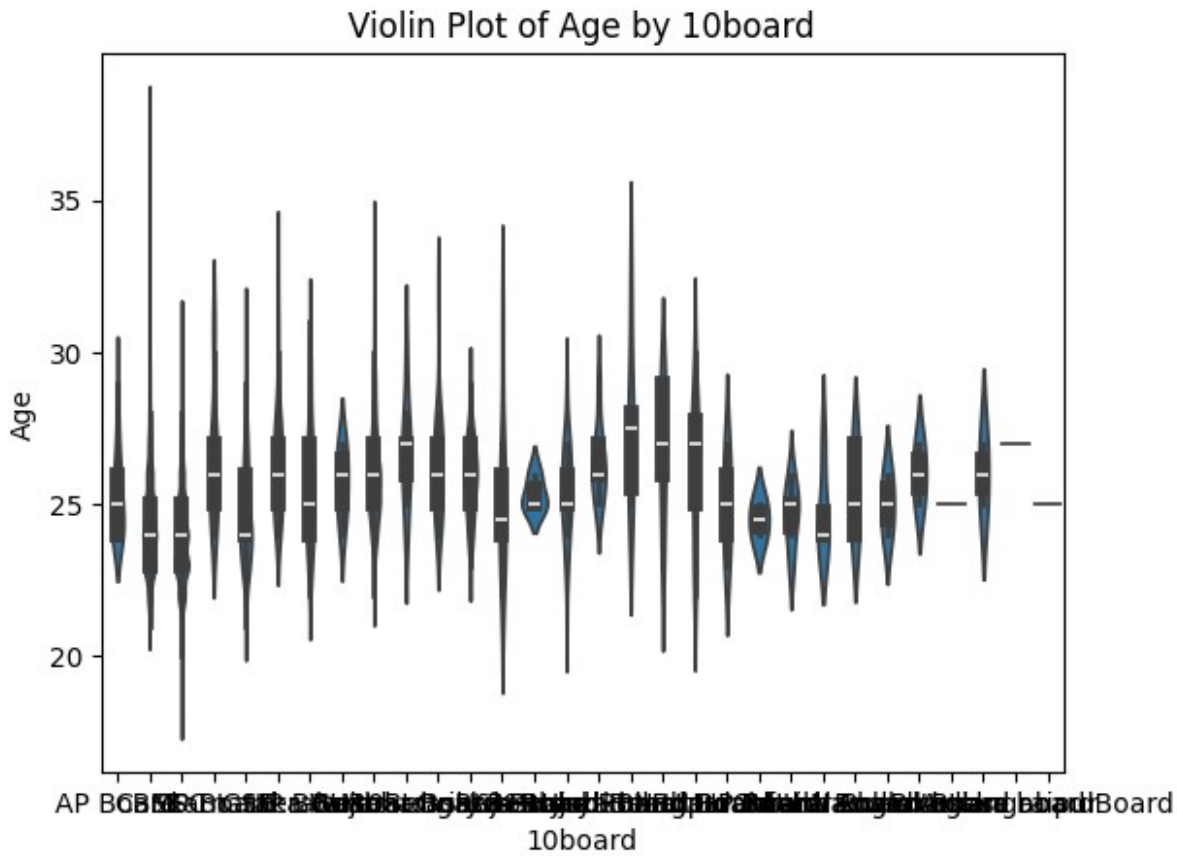

Violin Plot of conscientiousness by 10board
conscientiousness
4
2
0
−2
−4
10board

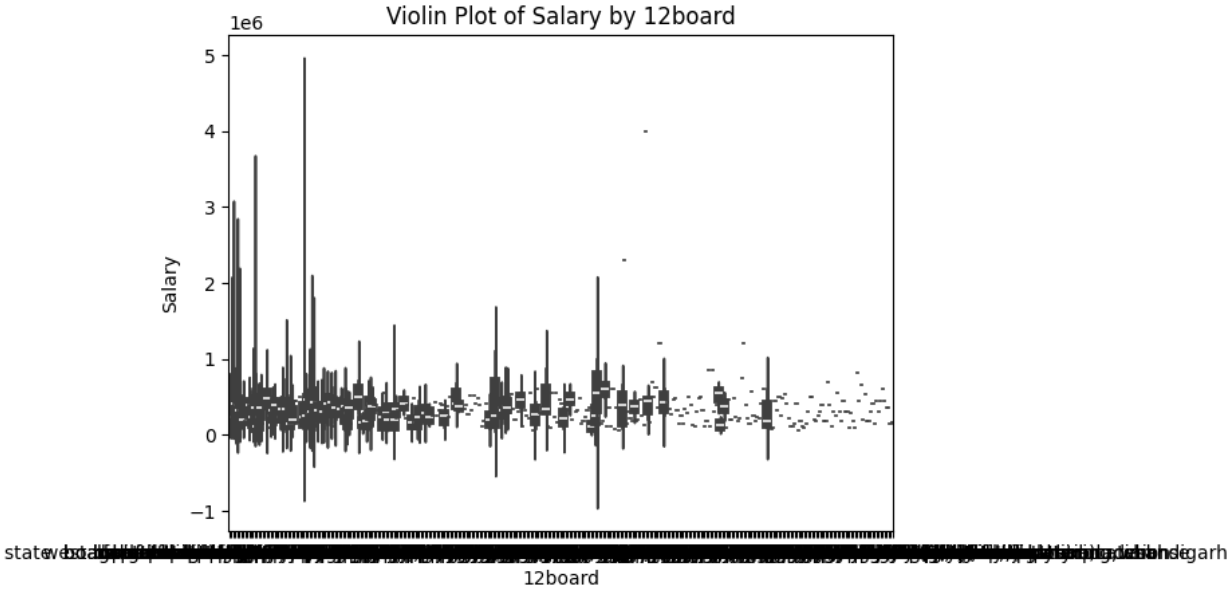

Violin Plot of agreeableness by 10board
agreeableness
4
2
0
−2
−4
−6
10board

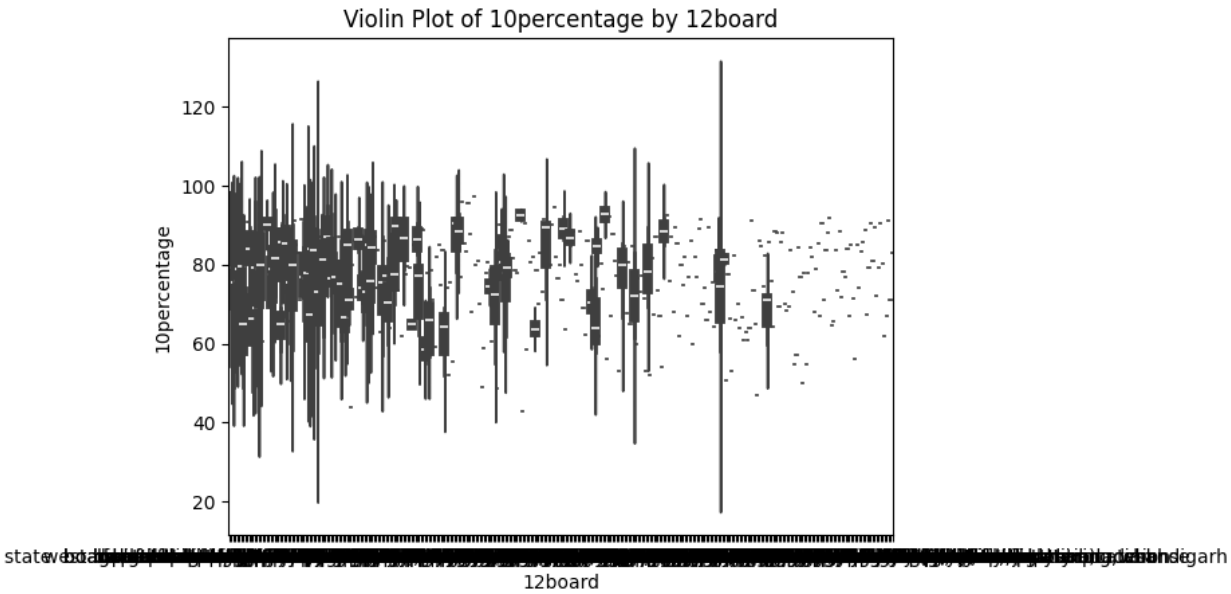

Violin Plot of extraversion by 10board
extraversion
4
2
0
−2
−4
10board

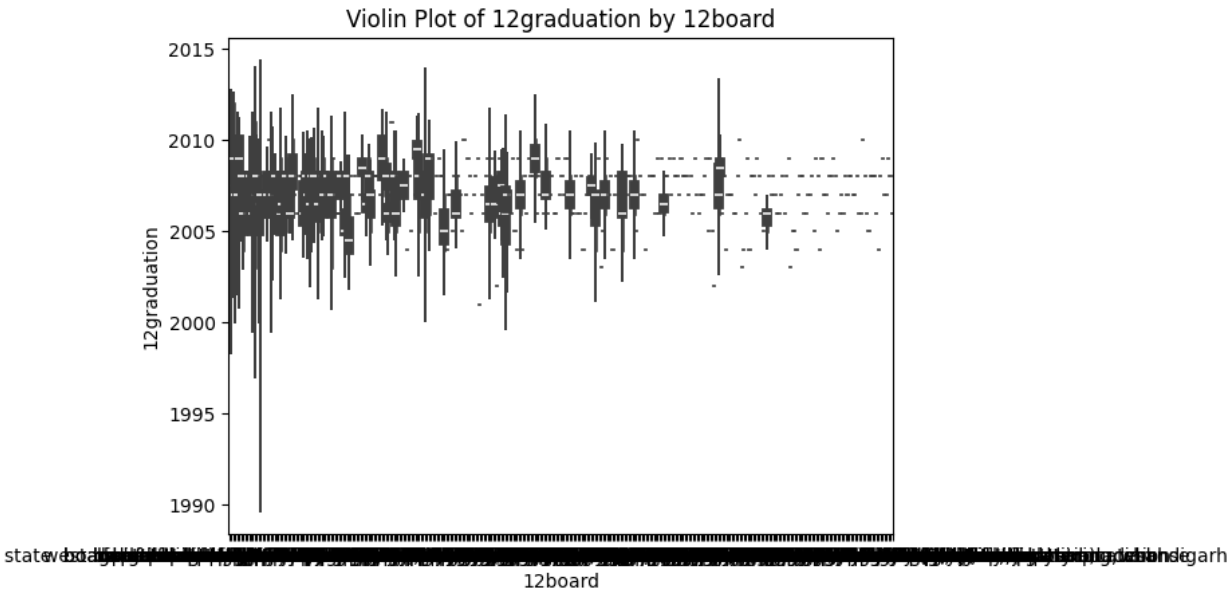

Violin Plot of nueroticism by 10board
nueroticism
4
2
0
−2
−4
10board

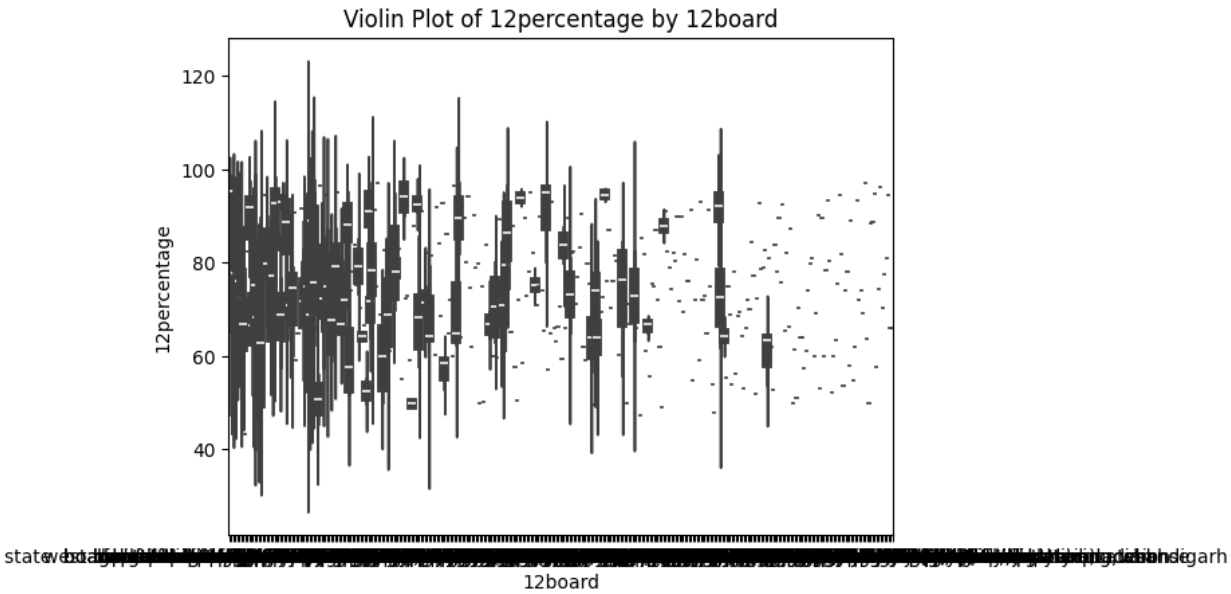

Violin Plot of openess_to_experience by 10board
openess_to_experience
2
0
−2
−4
−6
−8
AP Board
10board

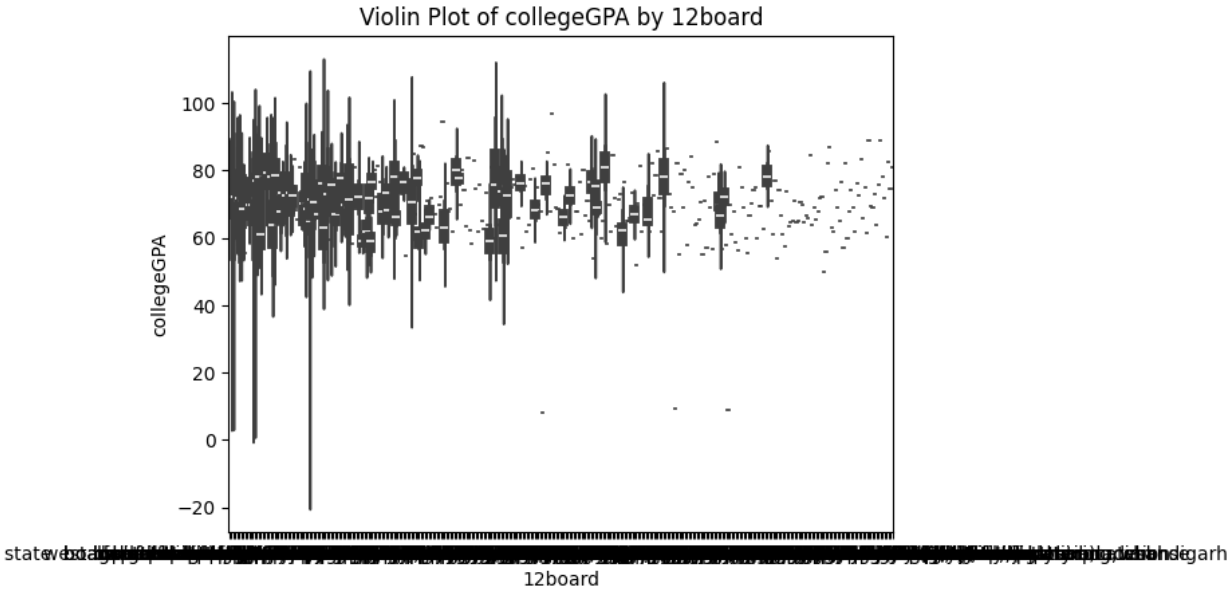

Violin Plot of Age by 10board
Age
35
30
25
20
10board

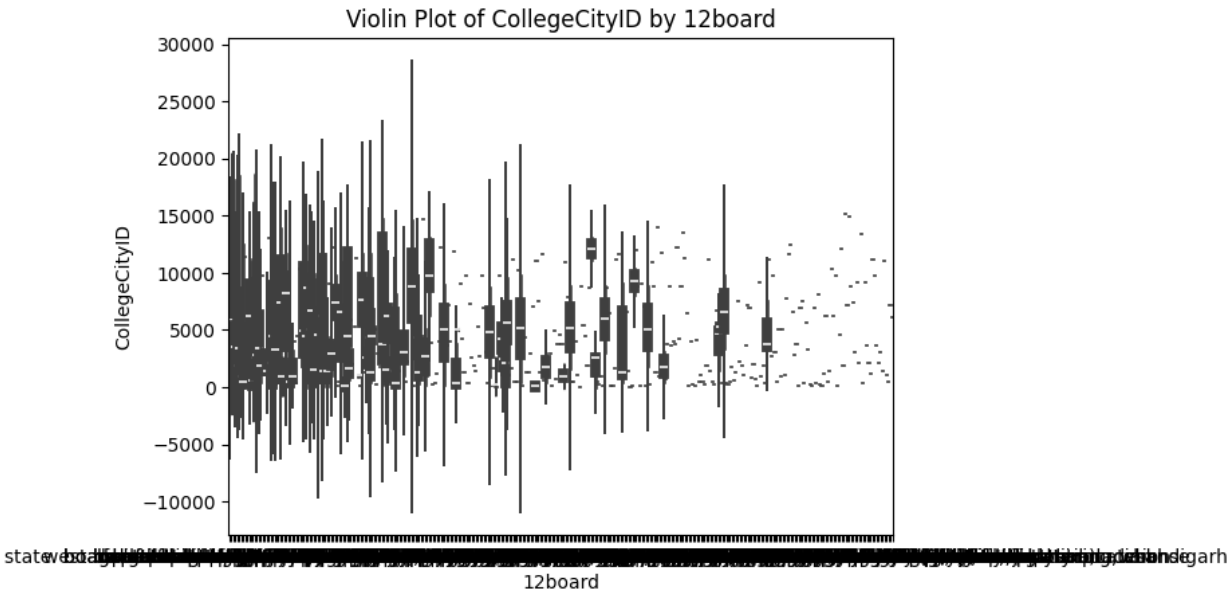

Violin Plot of Salary by 12board
1e6
Salary
5
4
3
2
1
0
−1
12board


Violin Plot of 10percentage by 12board
10percentage
120
100
80
60
40
20
12board

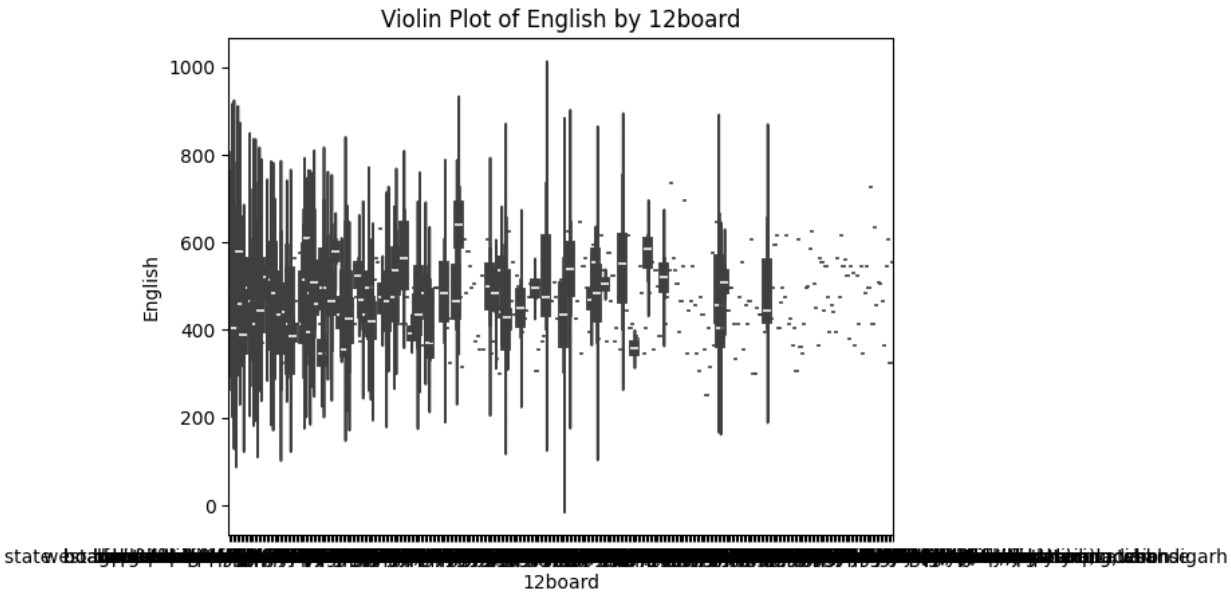

Violin Plot of 12graduation by 12board
12graduation
2015
2010
2005
2000
1995
1990
12board

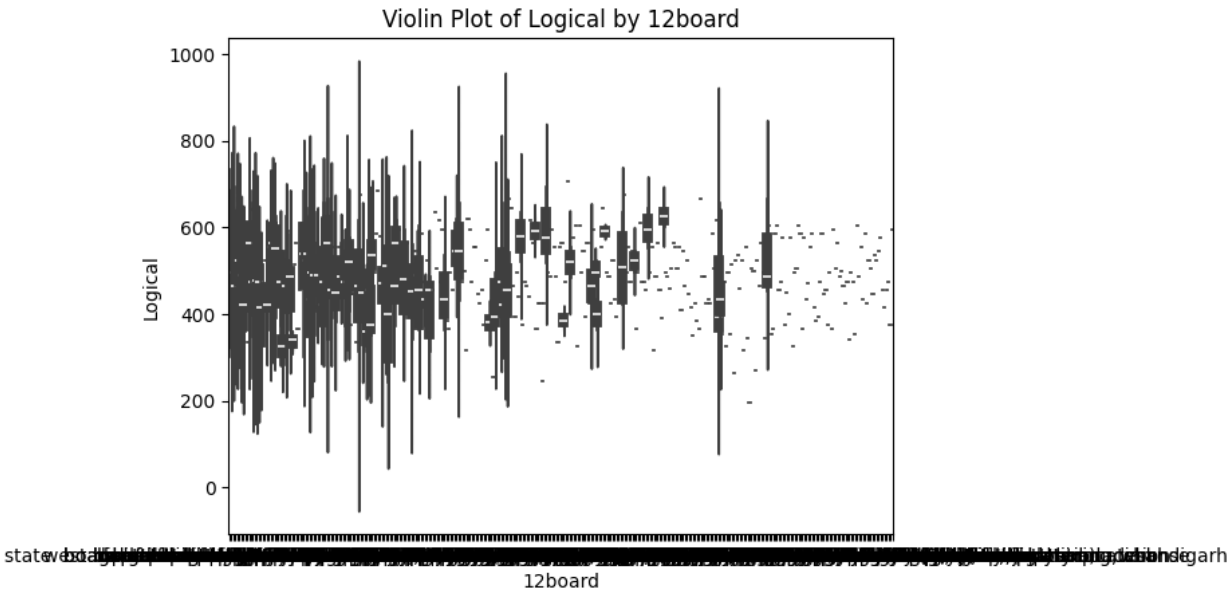

Violin Plot of 12percentage by 12board
12percentage
120
100
80
60
40
12board

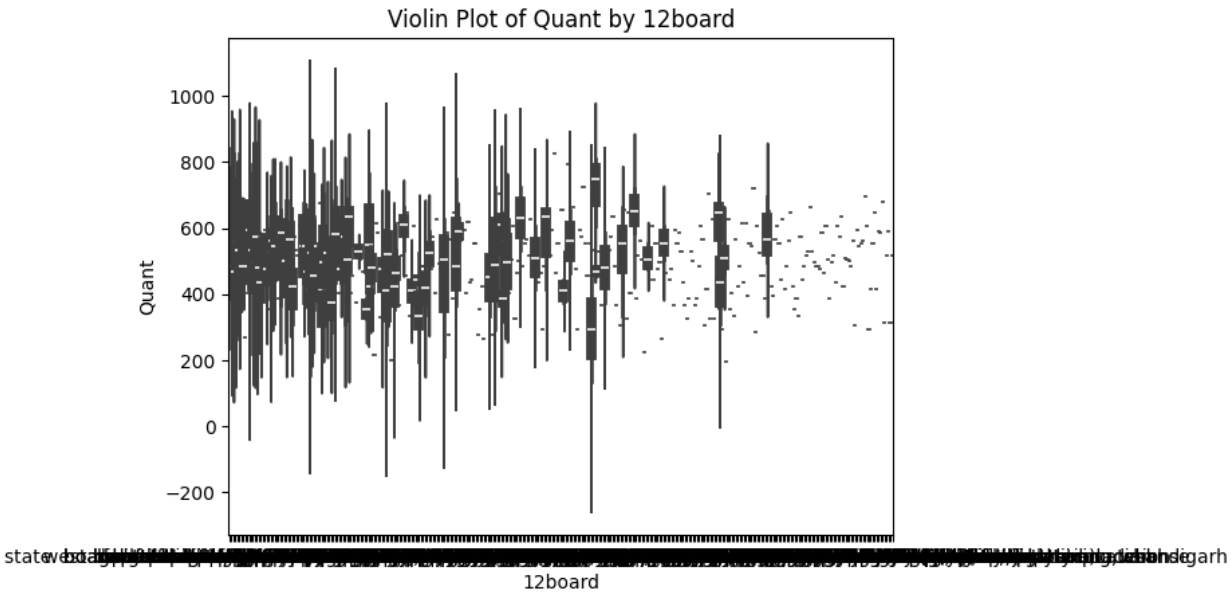

Violin Plot of collegeGPA by 12board
collegeGPA
100
80
60
40
20
0
−20
12board

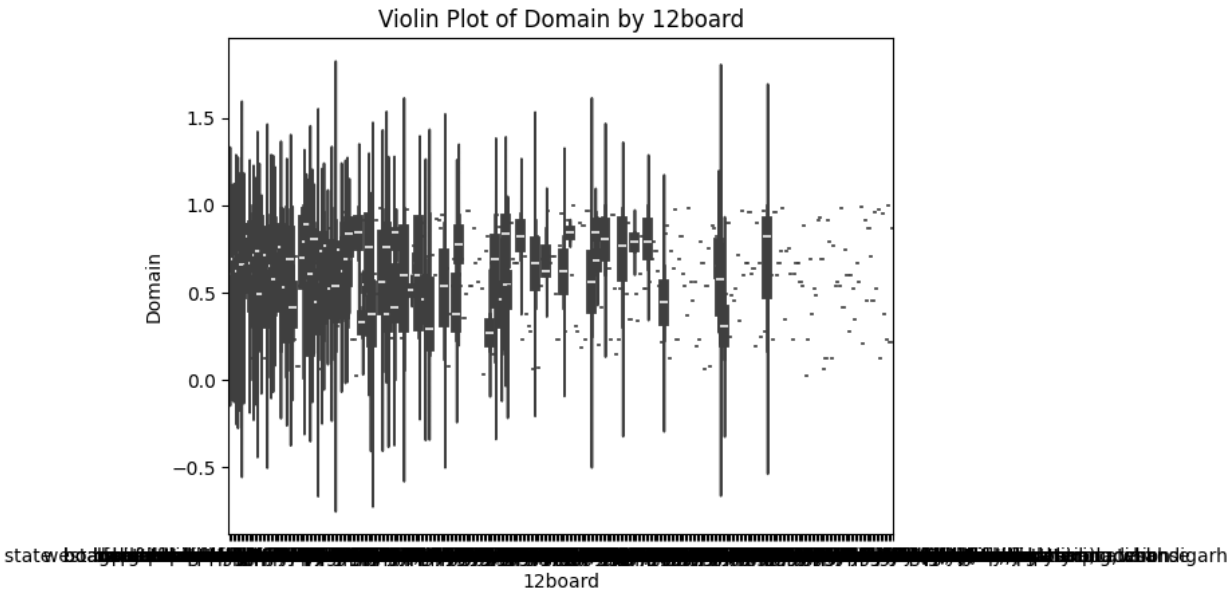

Violin Plot of CollegeCityID by 12board
CollegeCityID
30000
25000
20000
15000
10000
5000
0
−5000
−10000
12board

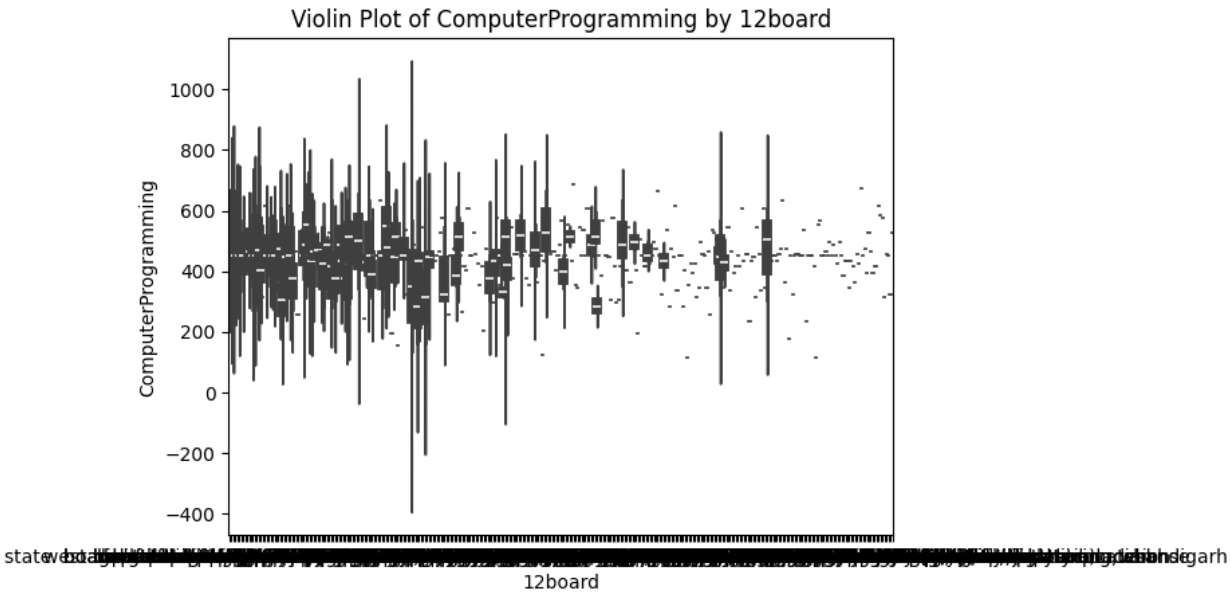
Violin Plot of GraduationYear by 12board
GraduationYear
2000
1500
1000
500
0
12board

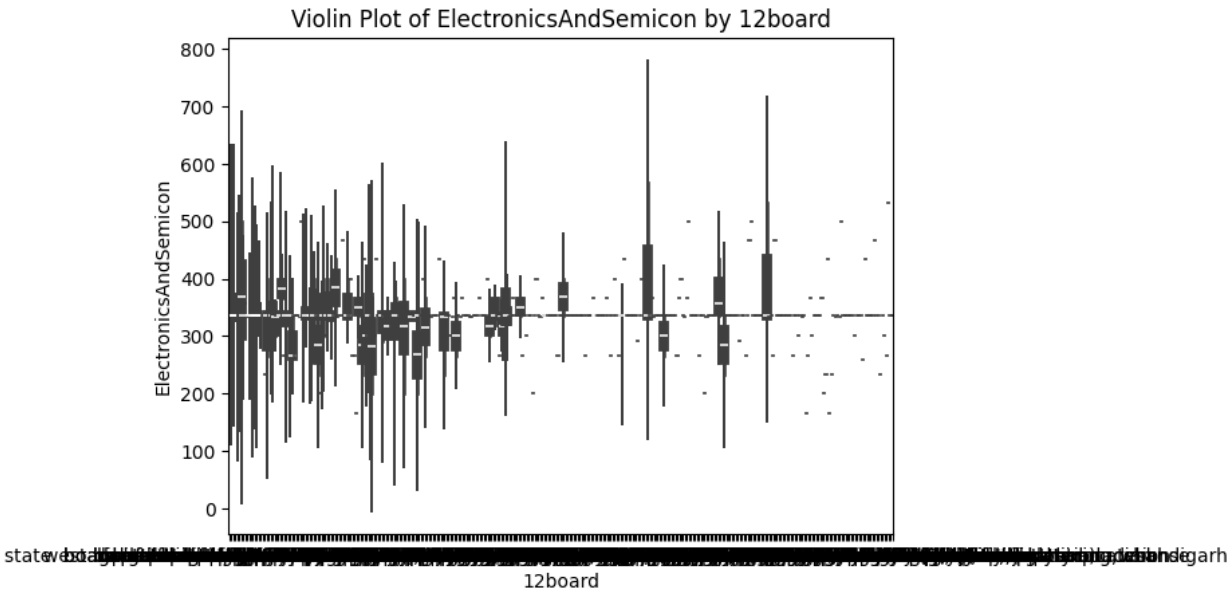

Violin Plot of English by 12board
English
1000
800
600
400
200
0
12board

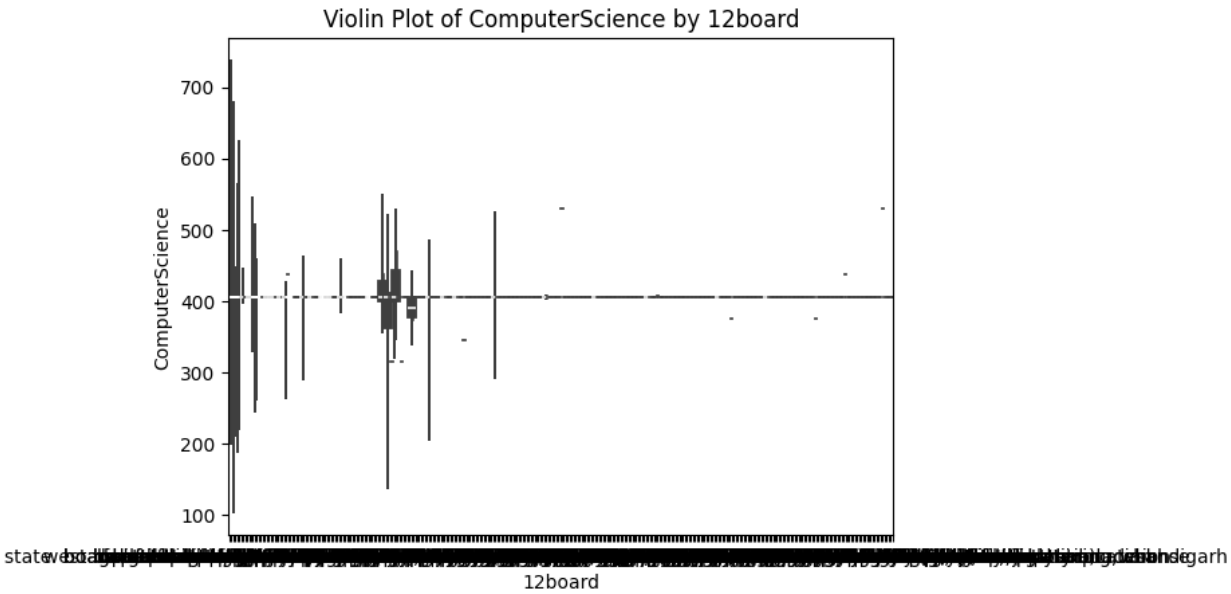

Violin Plot of Logical by 12board
Logical
1000
800
600
400
200
0
12board

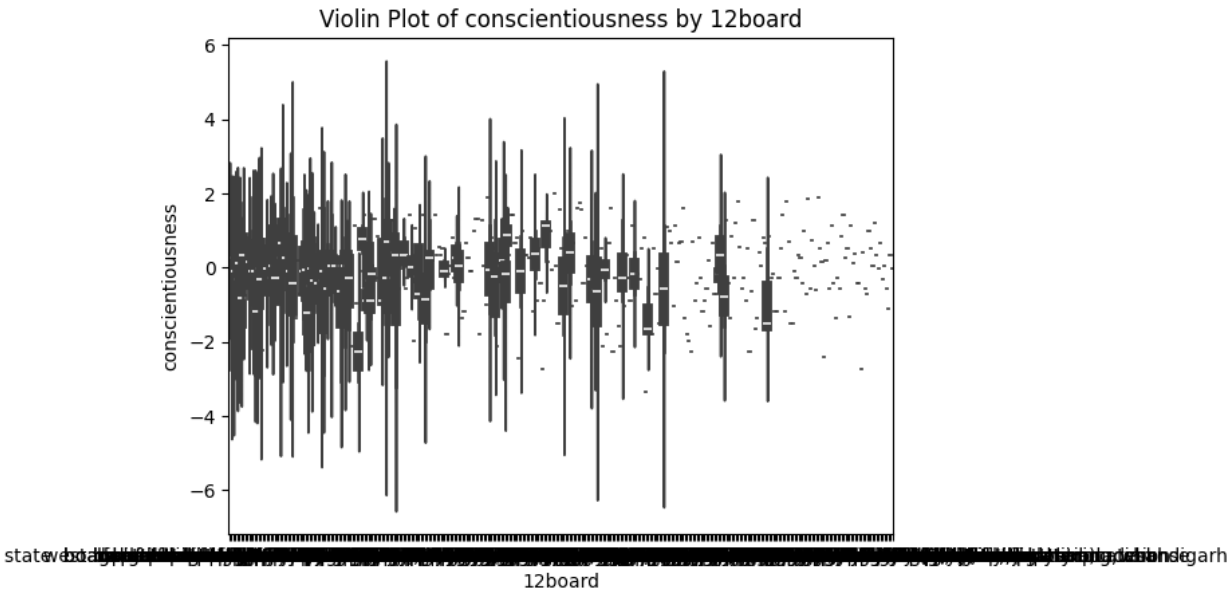

Violin Plot of Quant by 12board
Quant
1000
800
600
400
200
0
−200
12board

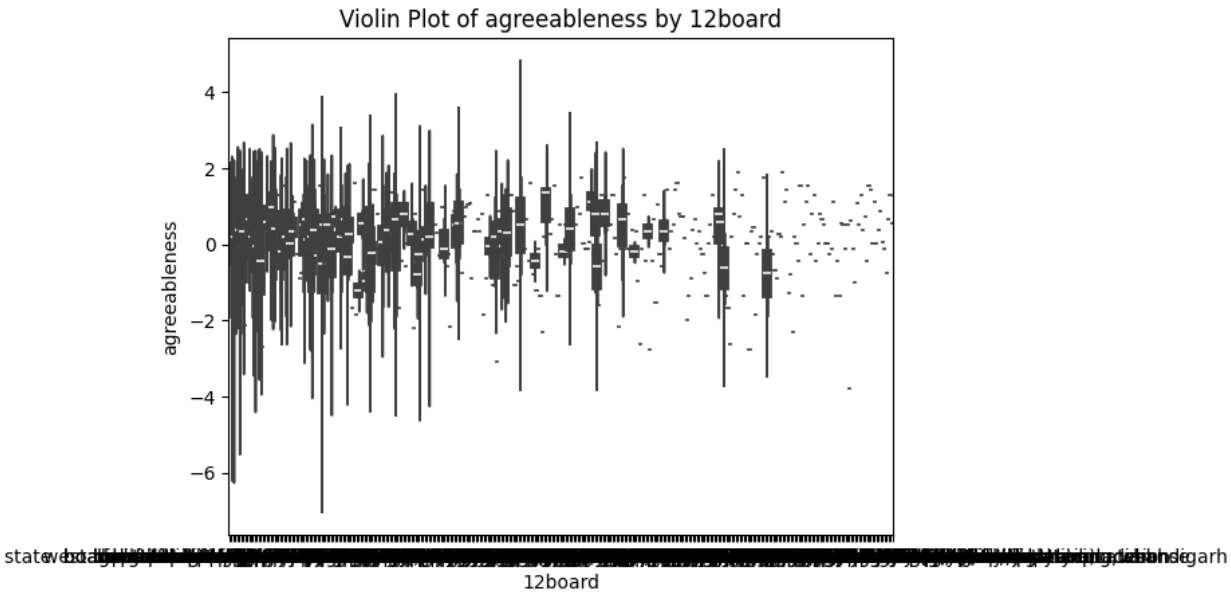

Violin Plot of Domain by 12board
Domain
1.5
1.0
0.5
0.0
−0.5
12board

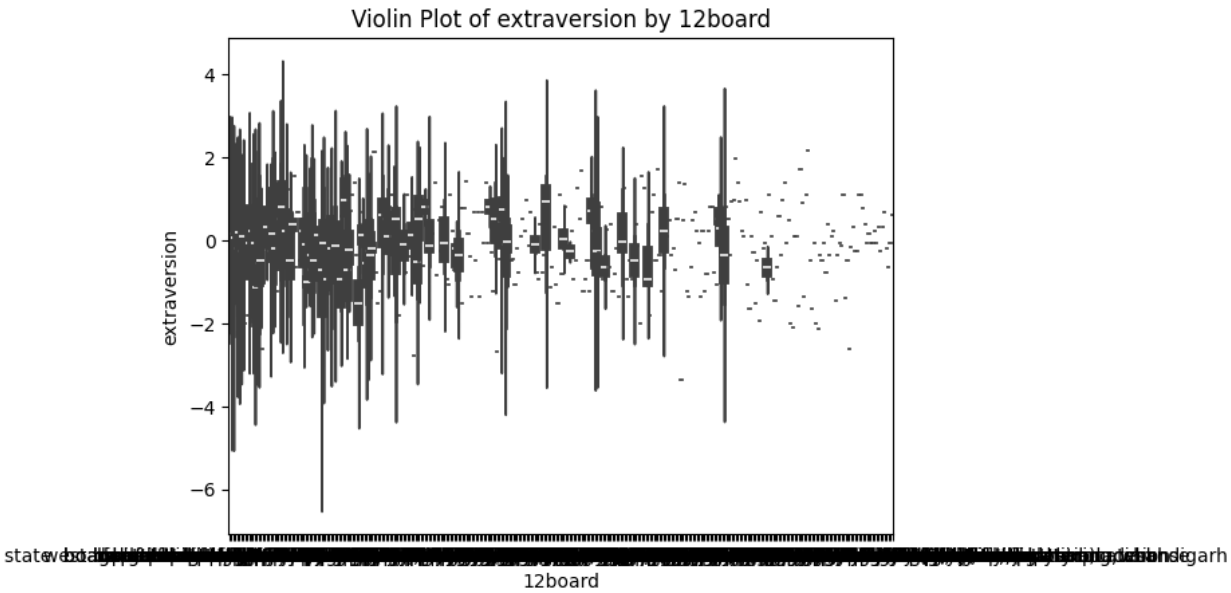

Violin Plot of ComputerProgramming by 12board
ComputerProgramming
1000
800
600
400
200
0
−200
−400
12board

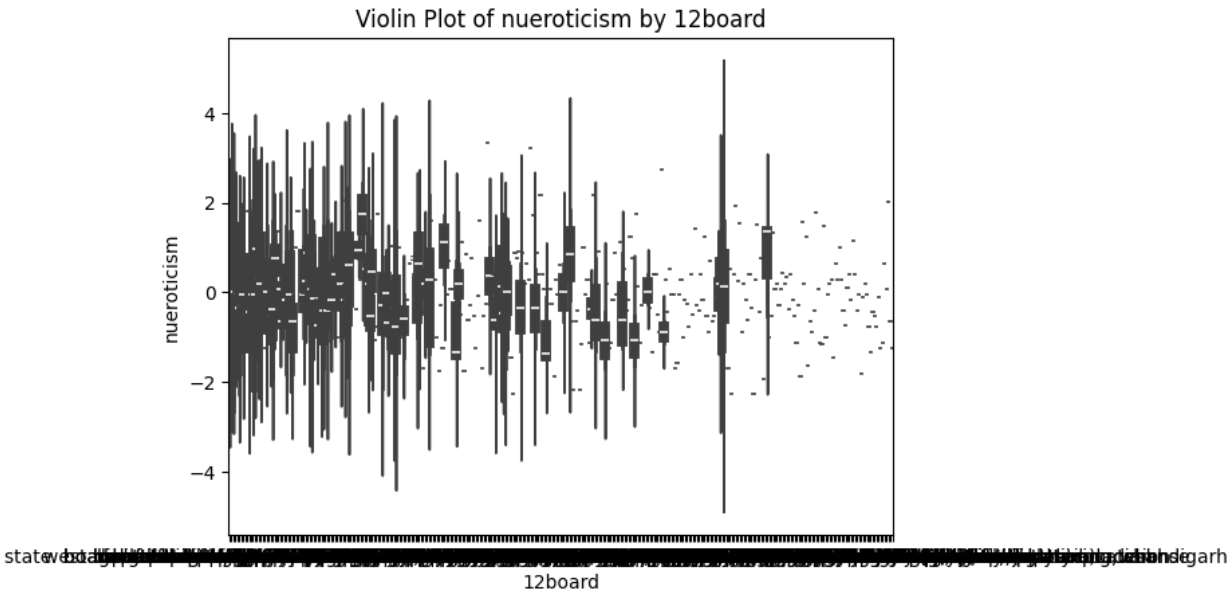

Violin Plot of ElectronicsAndSemicon by 12board
ElectronicsAndSemicon
800
700
600
500
400
300
200
100
0
12board

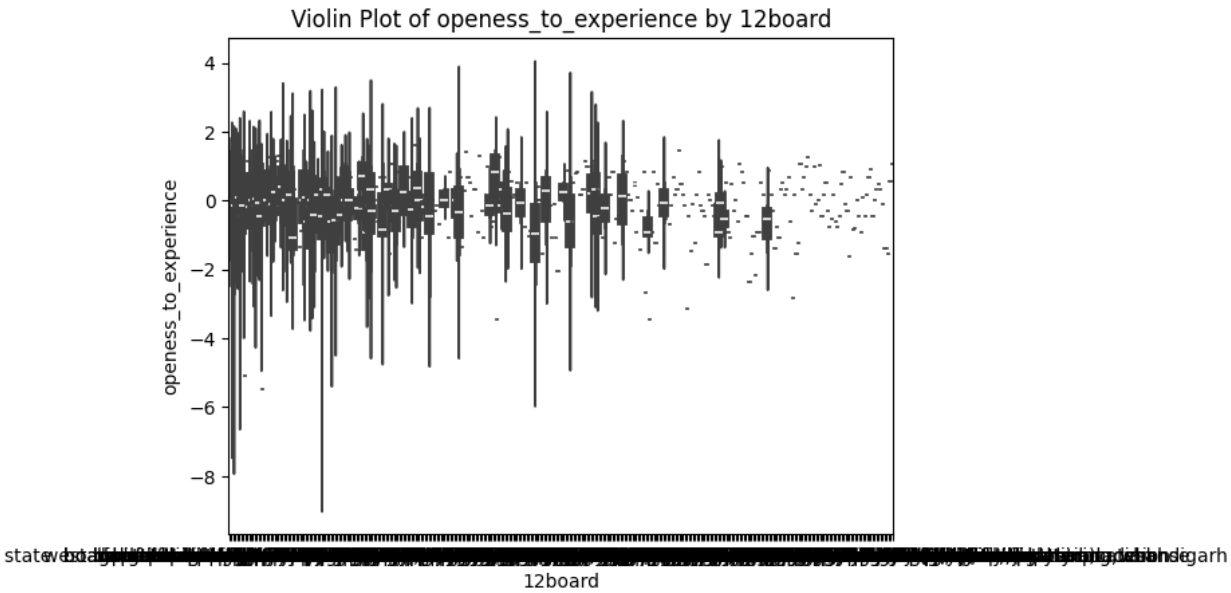

Violin Plot of ComputerScience by 12board
ComputerScience
700
600
500
400
300
200
100
12board

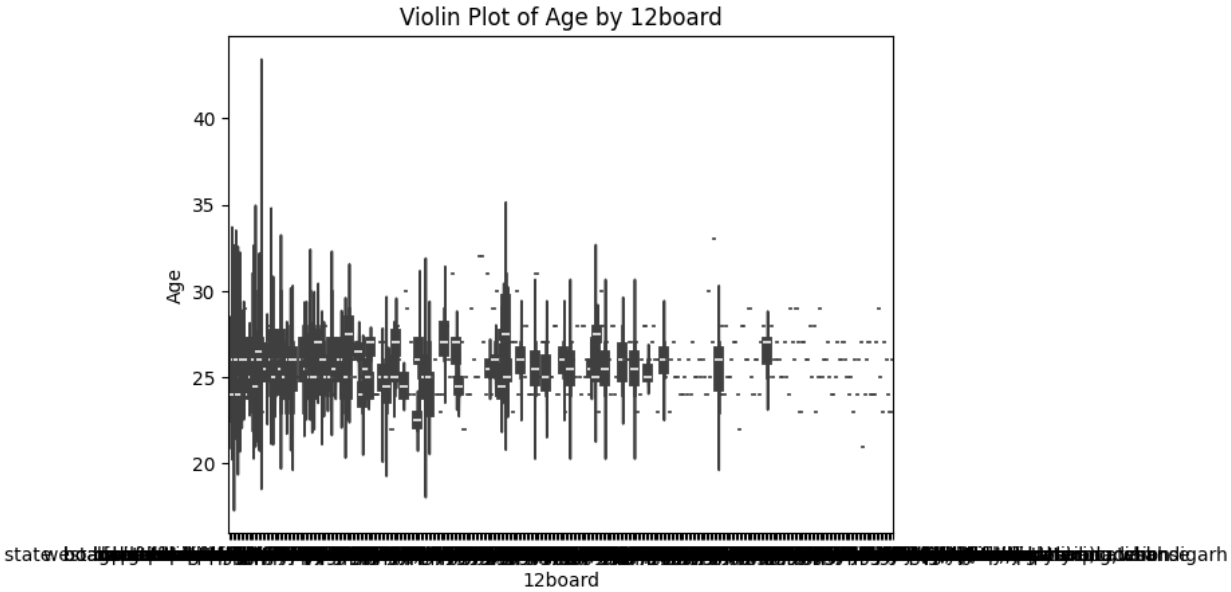

Violin Plot of conscientiousness by 12board
conscientiousness
6
4
2
0
−2
−4
−6
12board

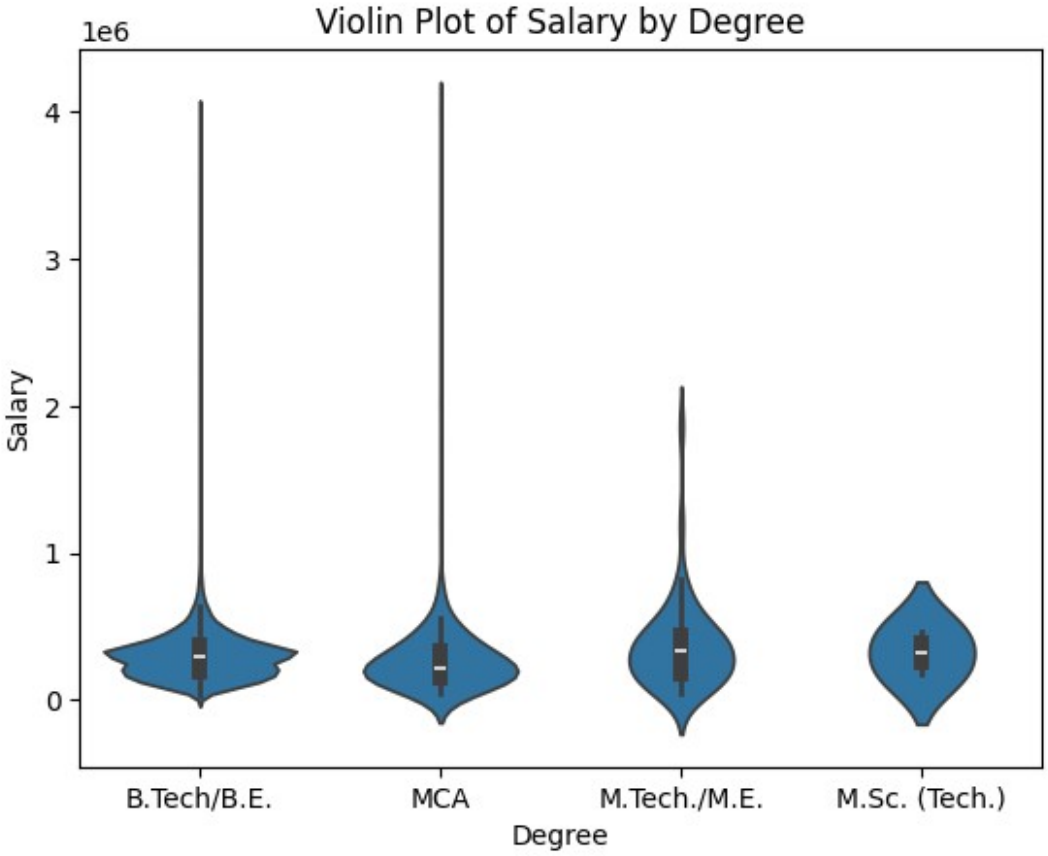

Violin Plot of agreeableness by 12board
agreeableness
4
2
0
−2
−4
−6
12board

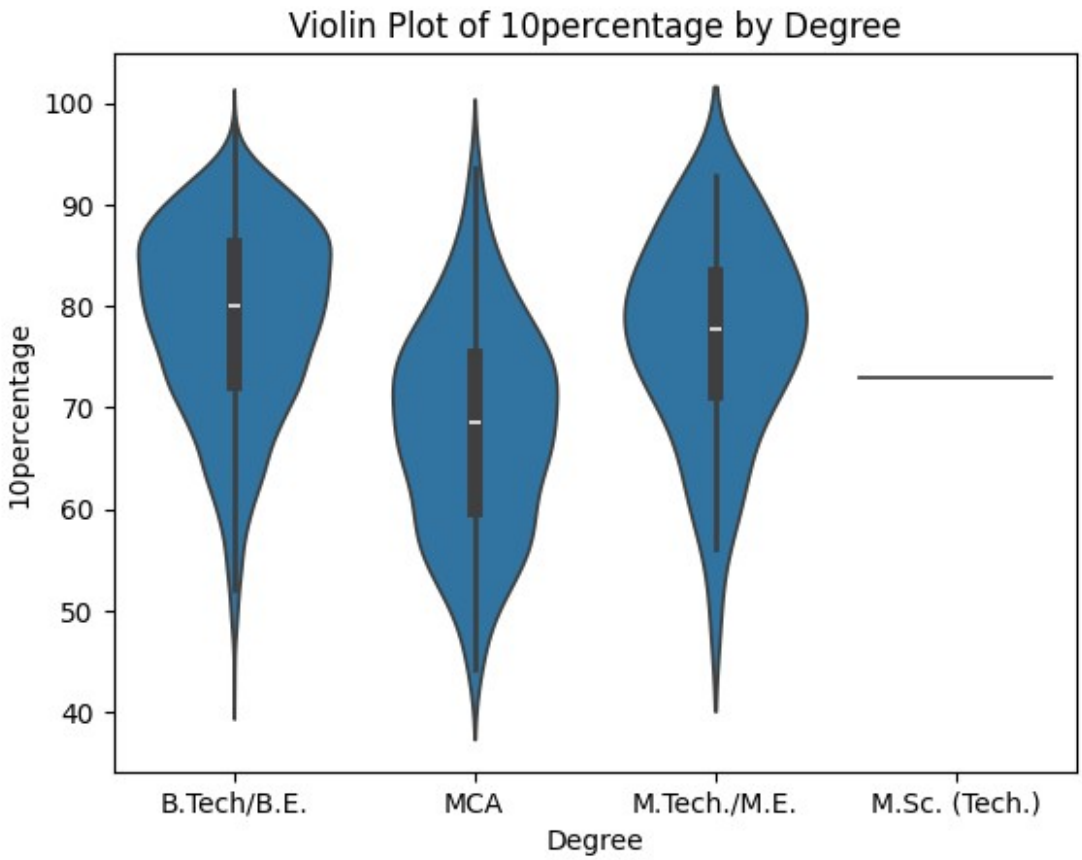

Violin Plot of extraversion by 12board
extraversion
4
2
0
−2
−4
−6
12board

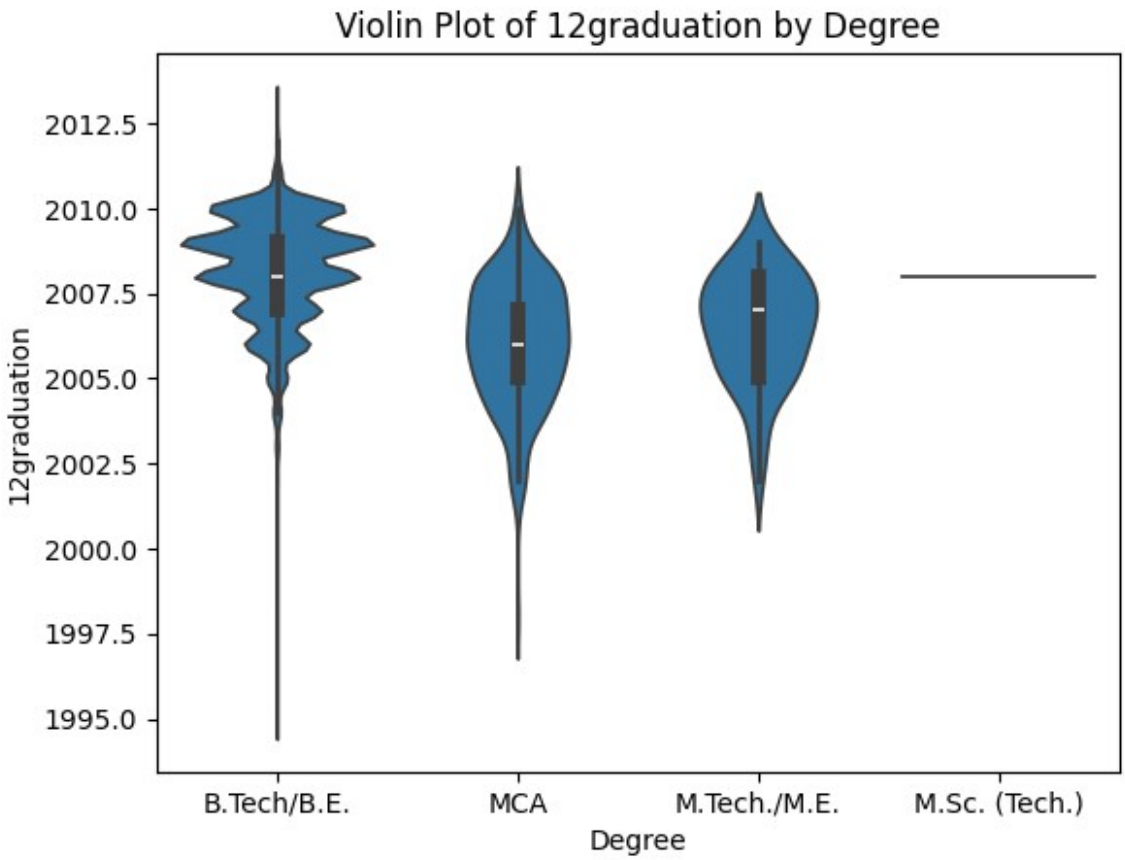

Violin Plot of nueroticism by 12board
nueroticism
4
2
0
−2
−4
12board

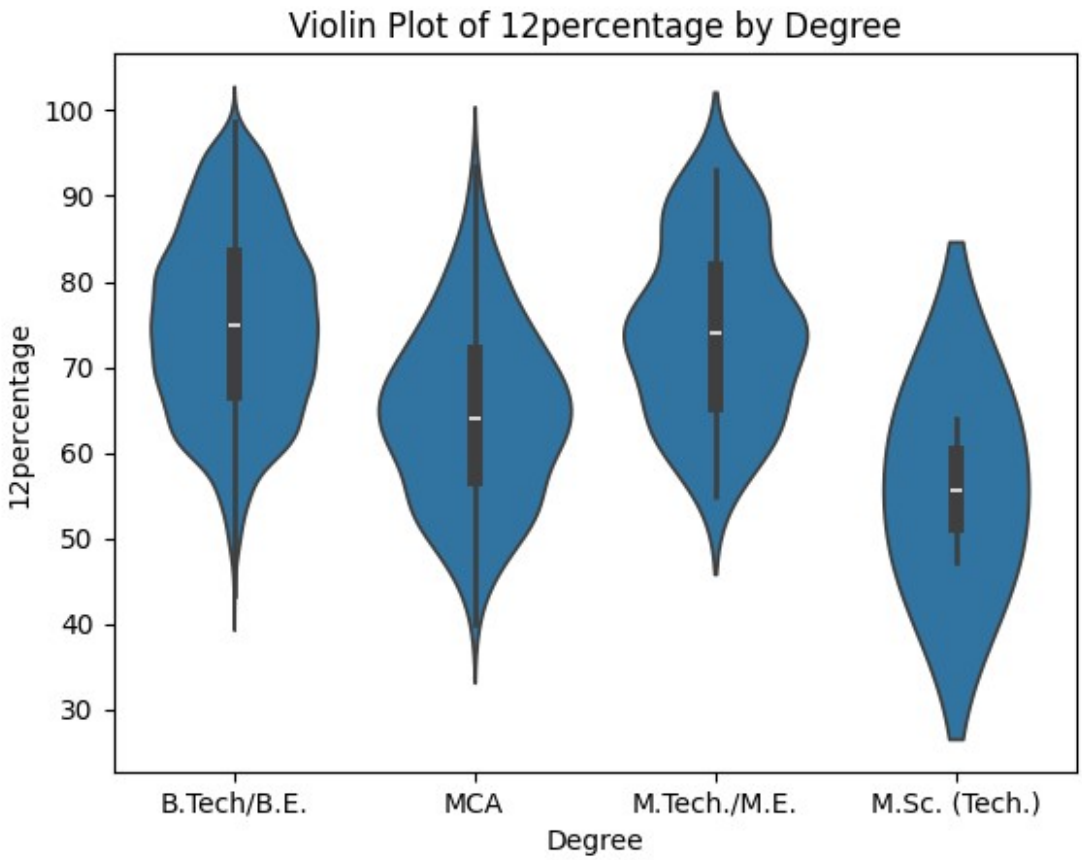

Violin Plot of openess_to_experience by 12board
openess_to_experience
4
2
0
−2
−4
−6
−8
12board

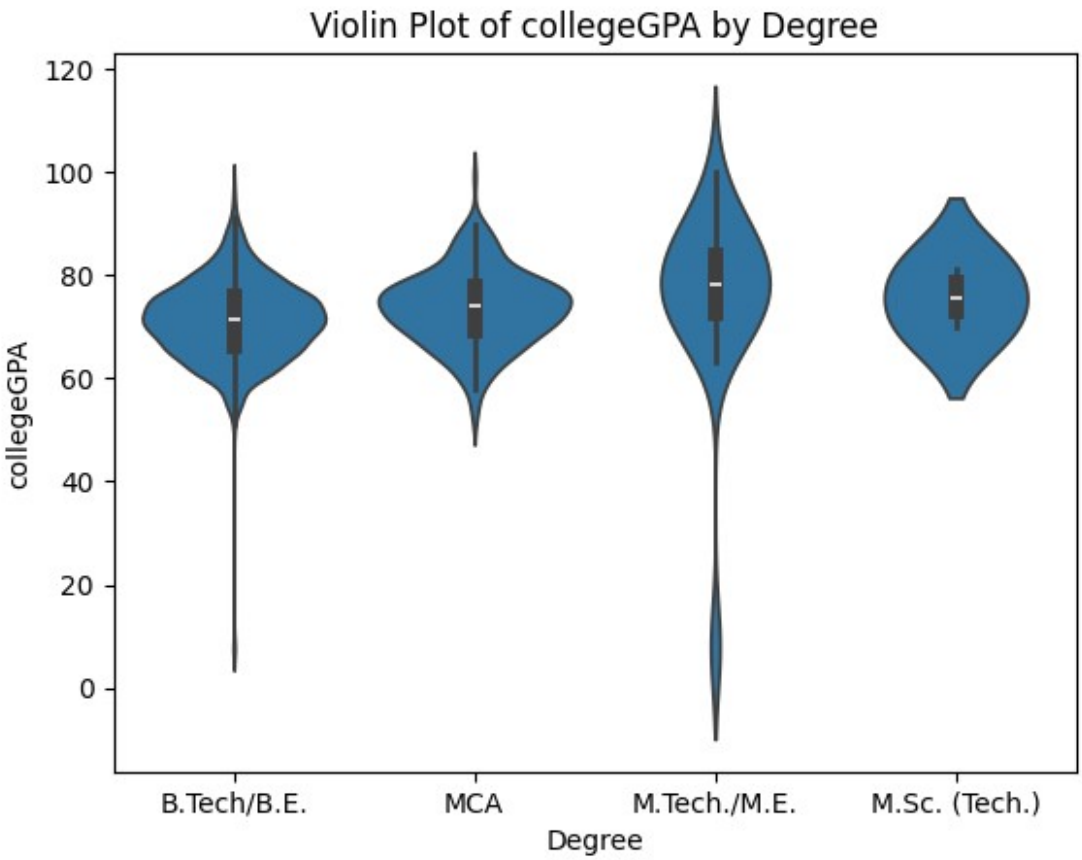

Violin Plot of Age by 12board
Age
40
35
30
25
20
12board

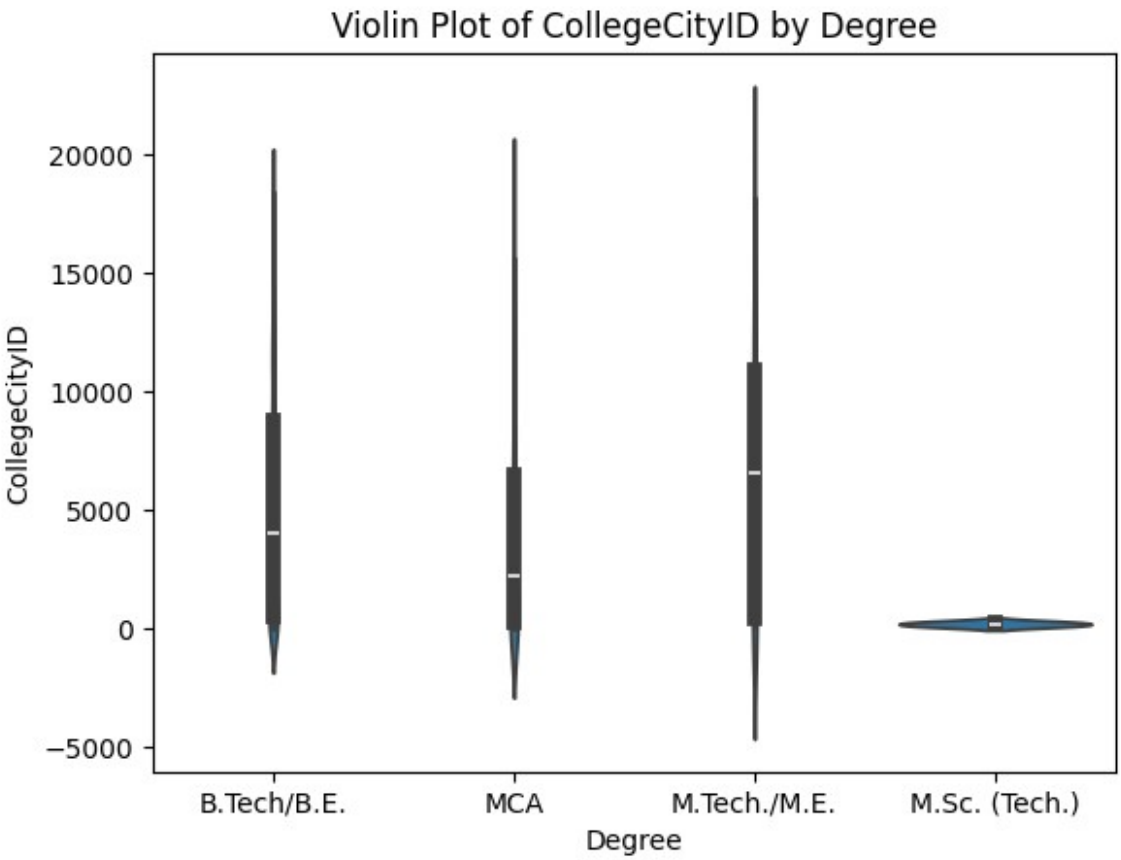

Violin Plot of Salary by Degree
1e6
Salary
4
3
2
1
0
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

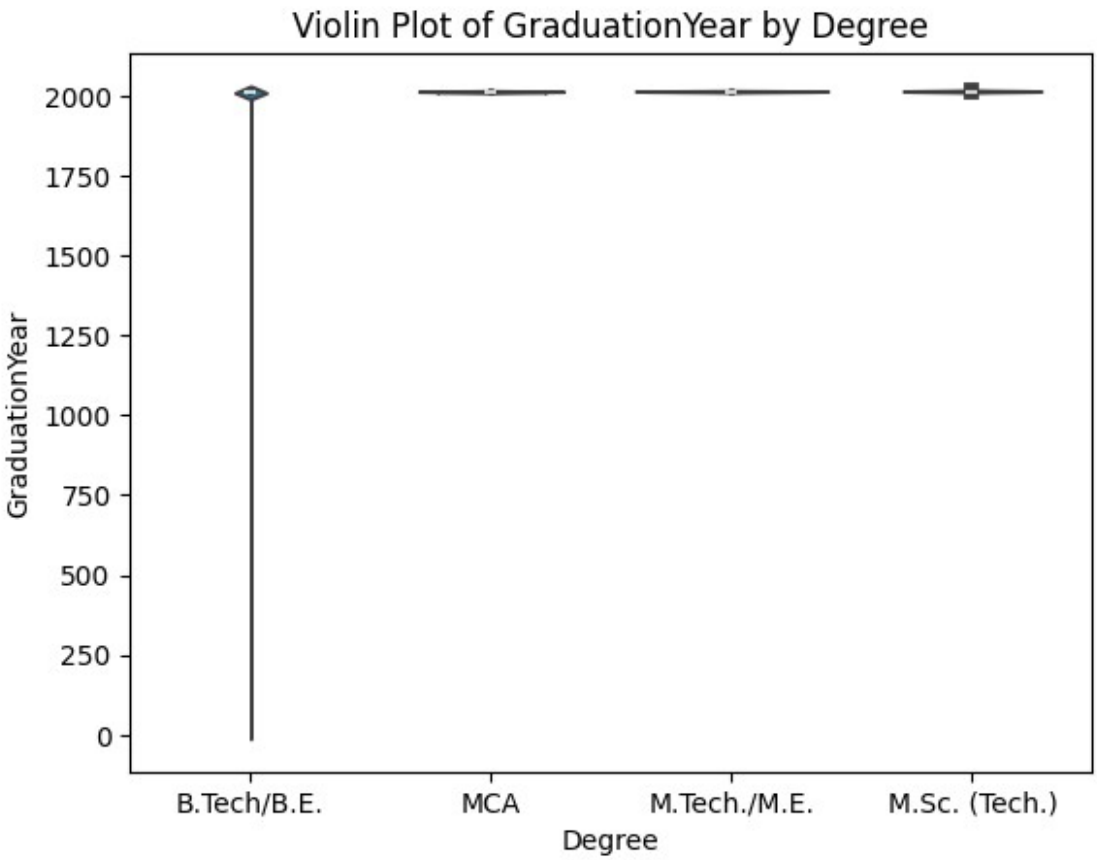

Violin Plot of 10percentage by Degree
10percentage
100
90
80
70
60
50
40
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

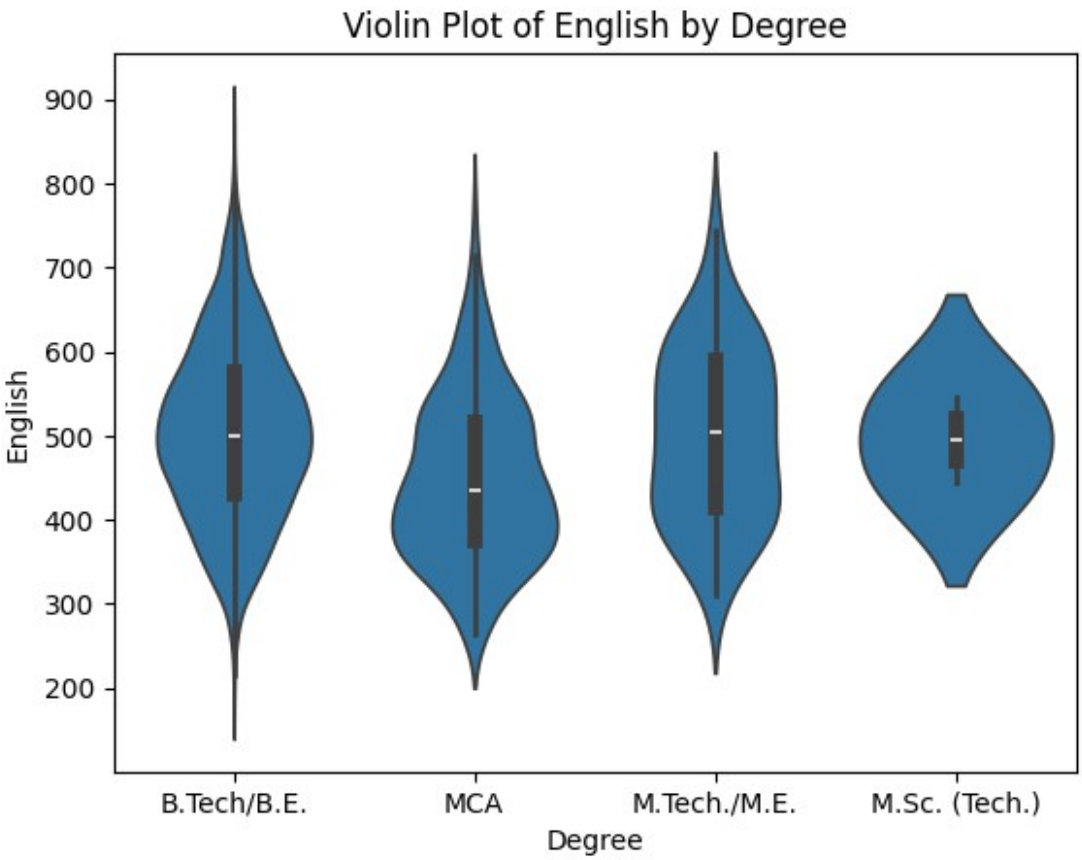

Violin Plot of 12graduation by Degree
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
12graduation
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

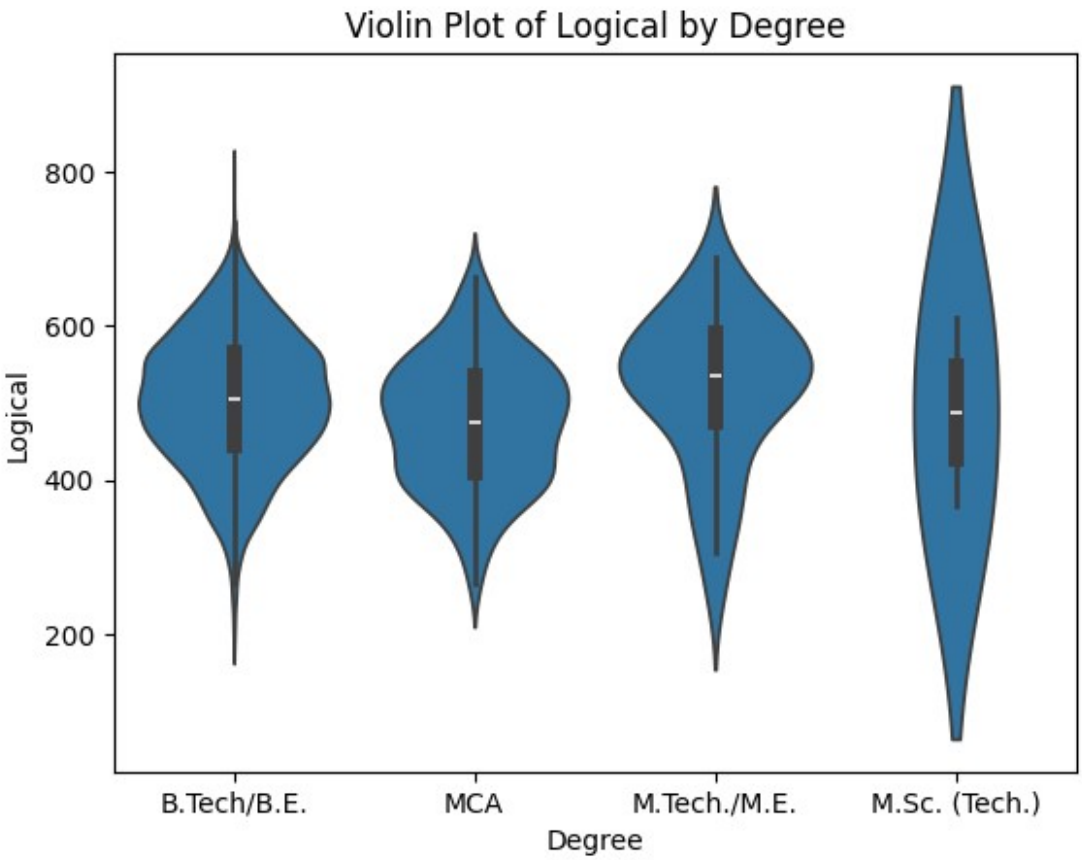

Violin Plot of 12percentage by Degree
12percentage
100
90
80
70
60
50
40
30
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

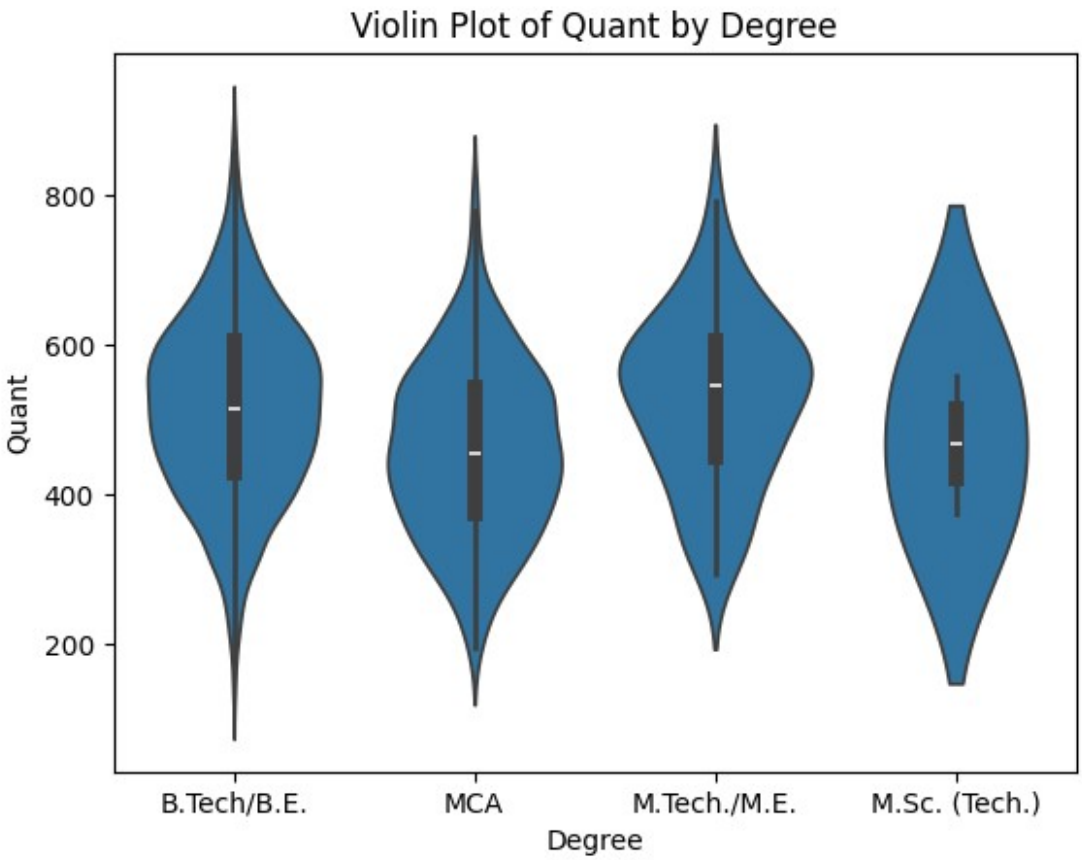

Violin Plot of collegeGPA by Degree
120
100
80
60
40
20
0
collegeGPA
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

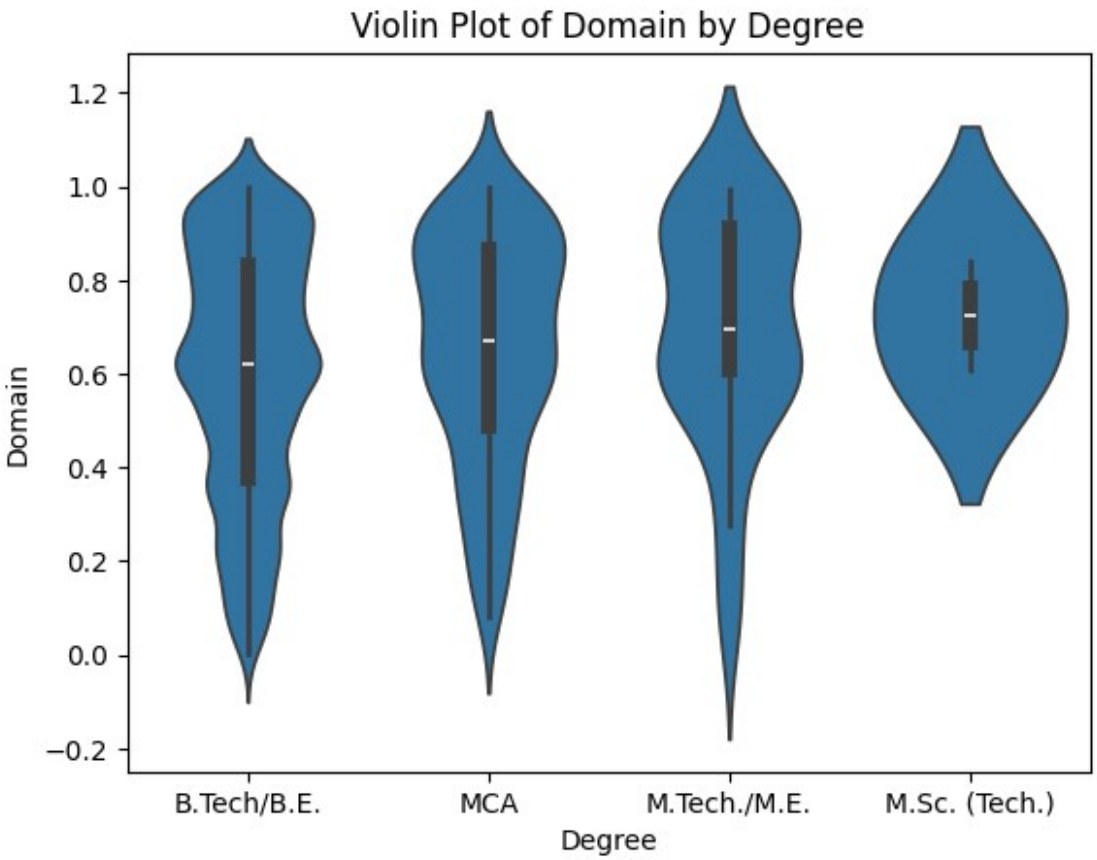

Violin Plot of CollegeCityID by Degree
CollegeCityID
20000
15000
10000
5000
0
−5000
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

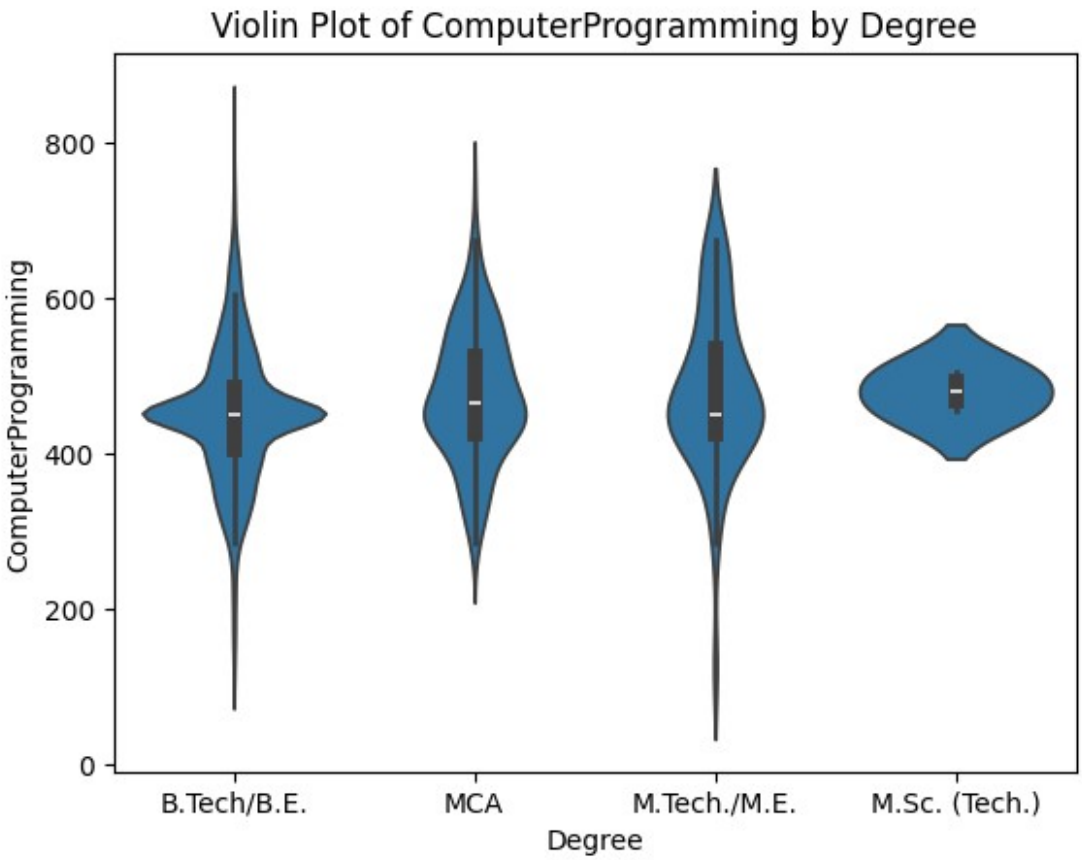

Violin Plot of GraduationYear by Degree
GraduationYear
2000
1750
1500
1250
1000
750
500
250
0
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

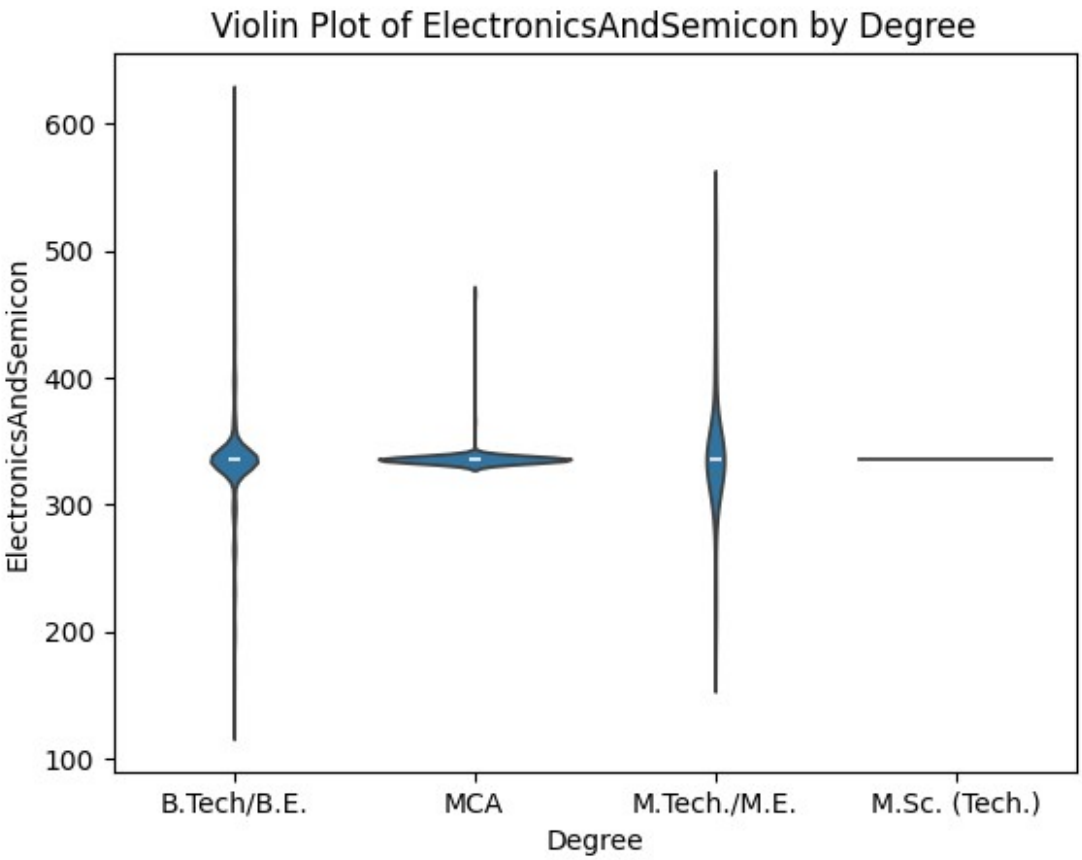

Violin Plot of English by Degree
English
900
800
700
600
500
400
300
200
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

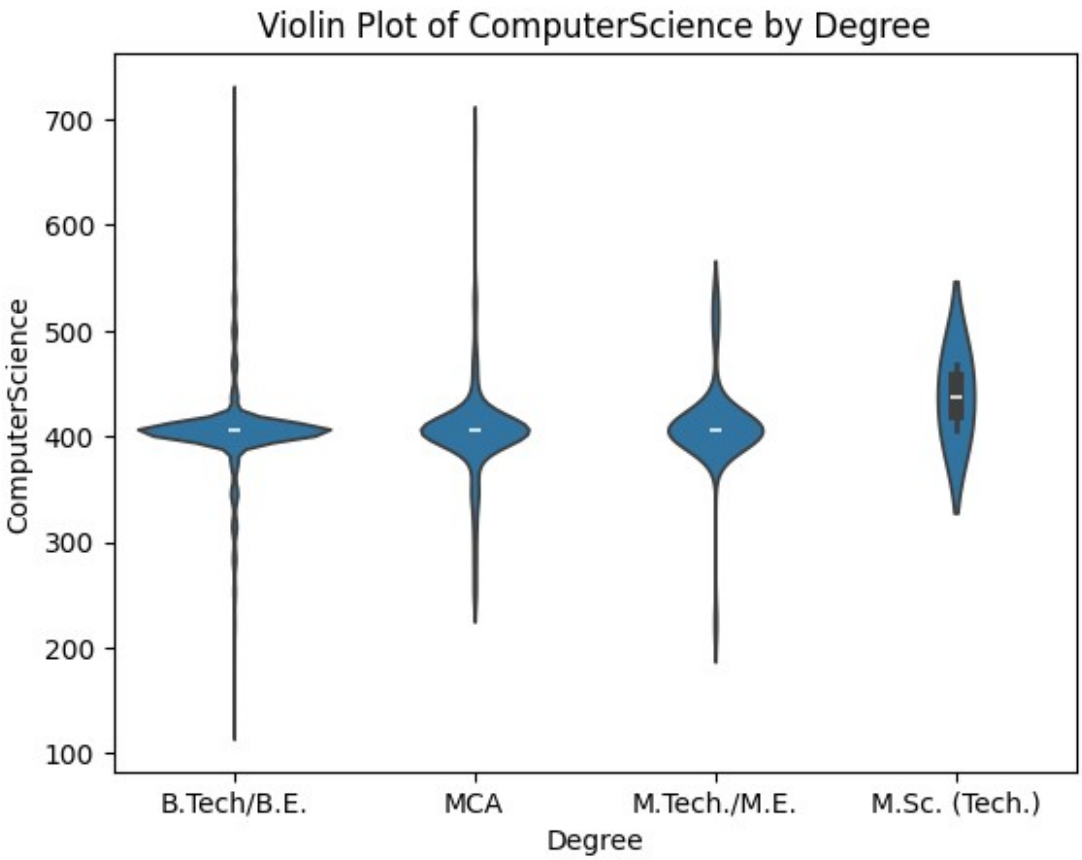

Violin Plot of Logical by Degree
800
600
400
200
Logical
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

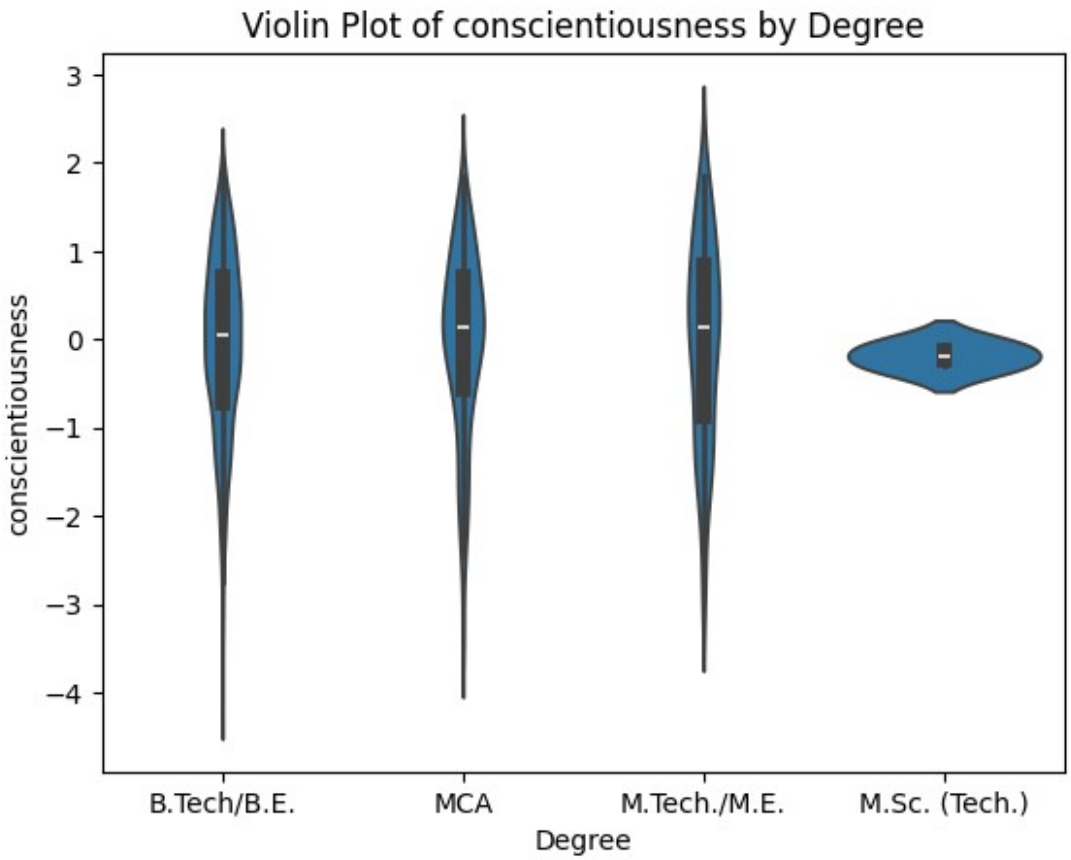

Violin Plot of Quant by Degree
Quant
800
600
400
200
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

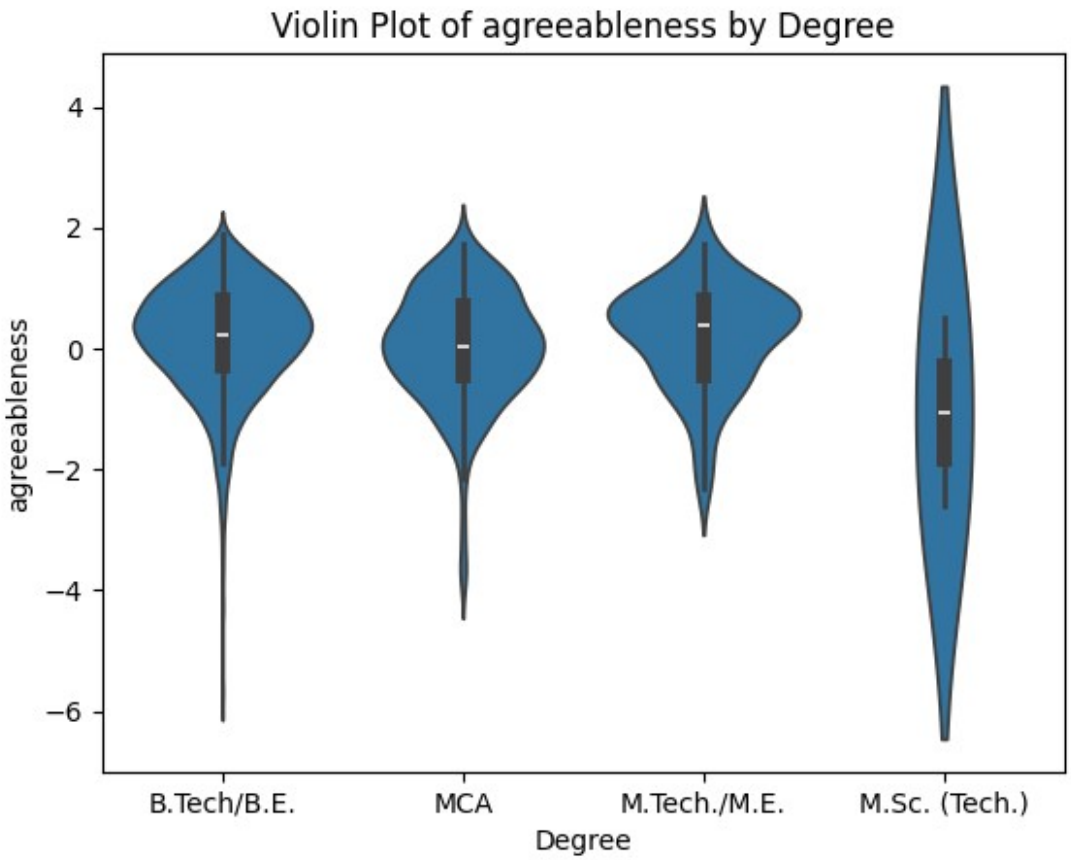

Violin Plot of Domain by Degree
Domain
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

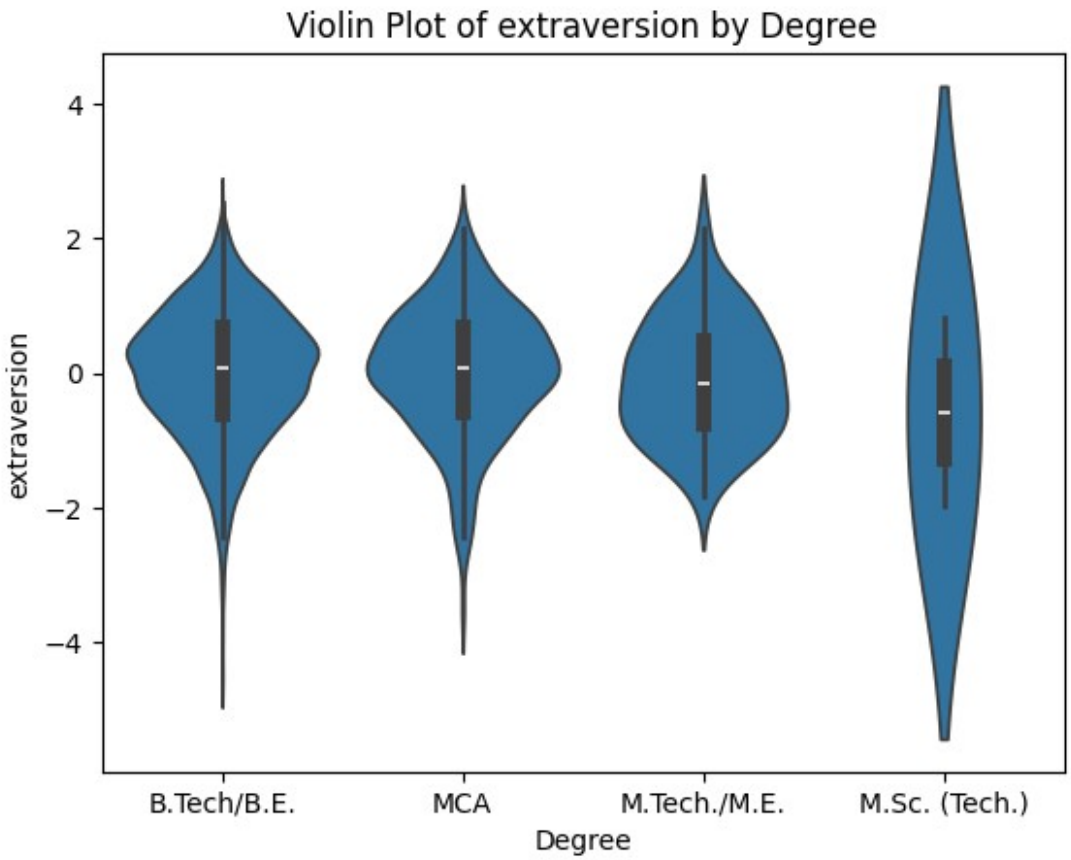

Violin Plot of ComputerProgramming by Degree
ComputerProgramming
0
200
400
600
800
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

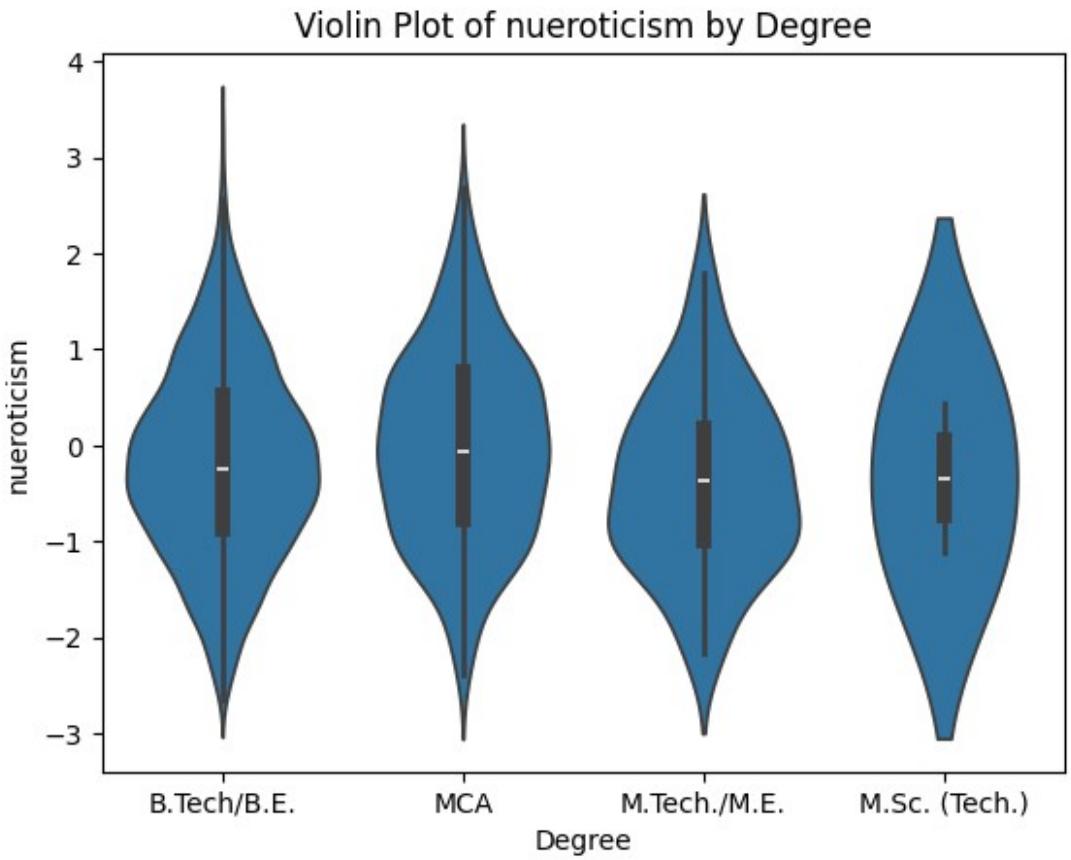

Violin Plot of ElectronicsAndSemicon by Degree
ElectronicsAndSemicon
600
500
400
300
200
100
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

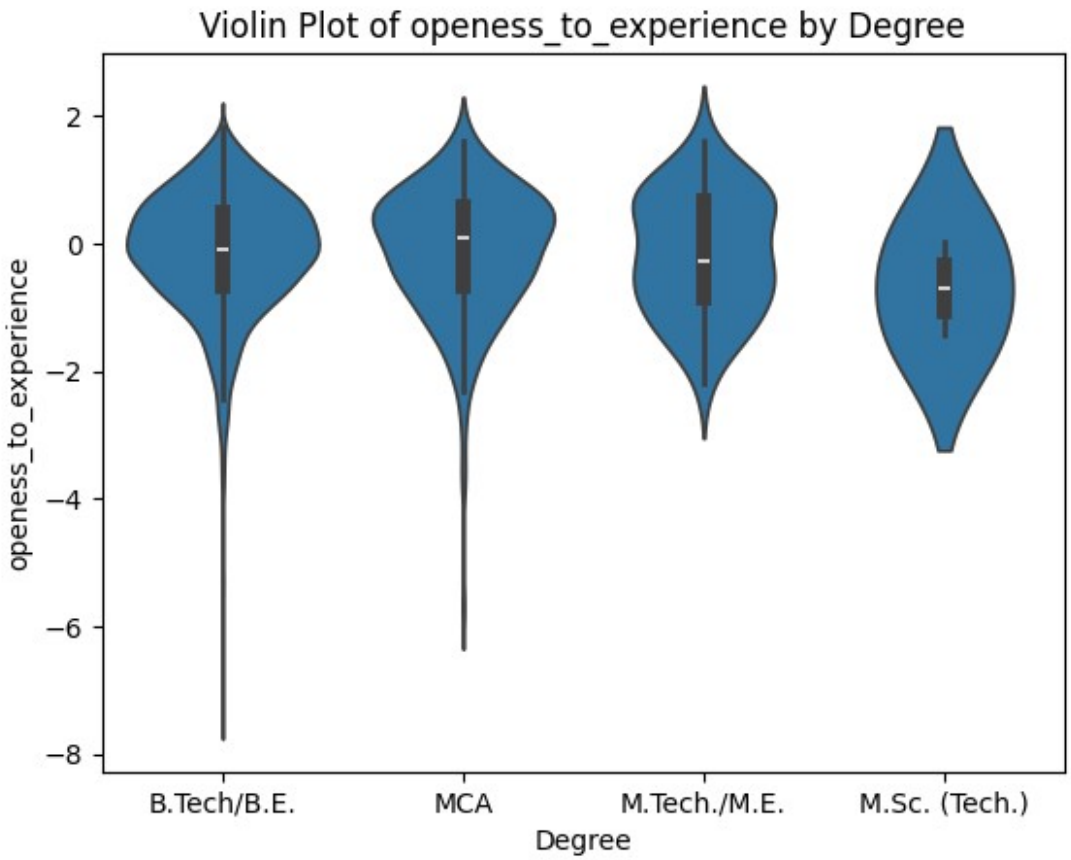

Violin Plot of ComputerScience by Degree
ComputerScience
700
600
500
400
300
200
100
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

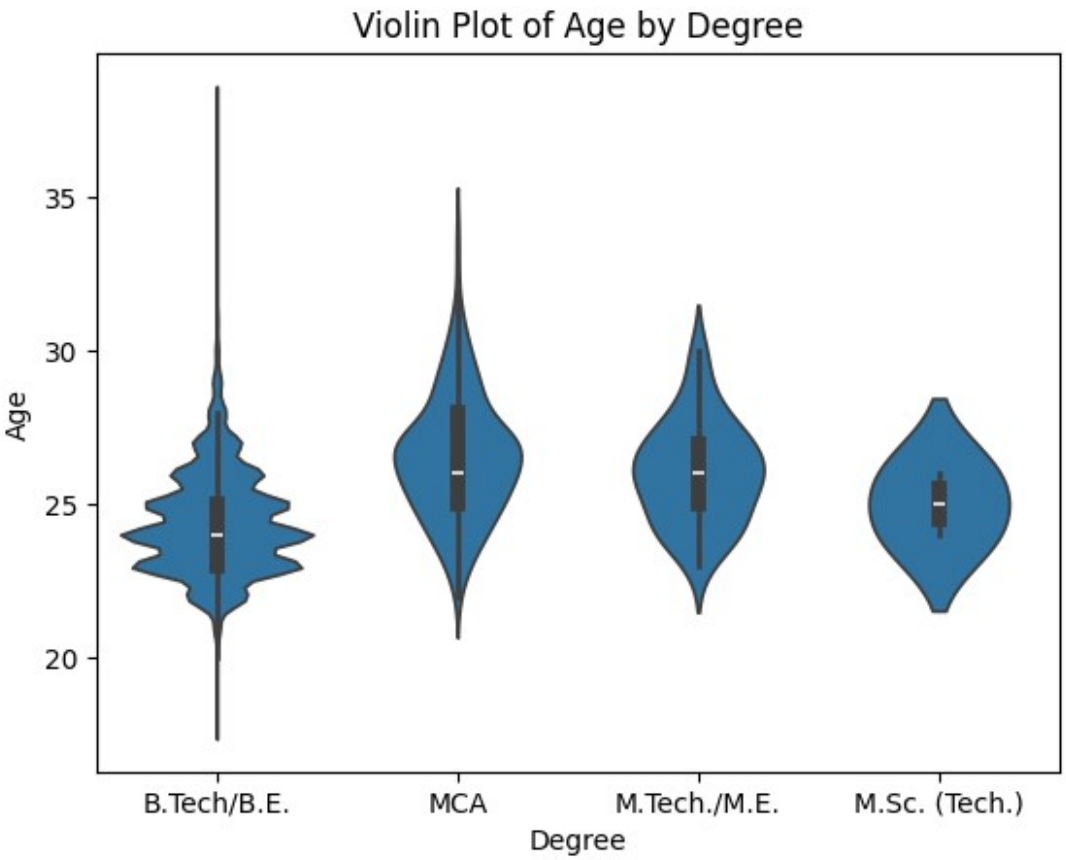

Violin Plot of conscientiousness by Degree
conscientiousness
3
2
1
0
−1
−2
−3
−4
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

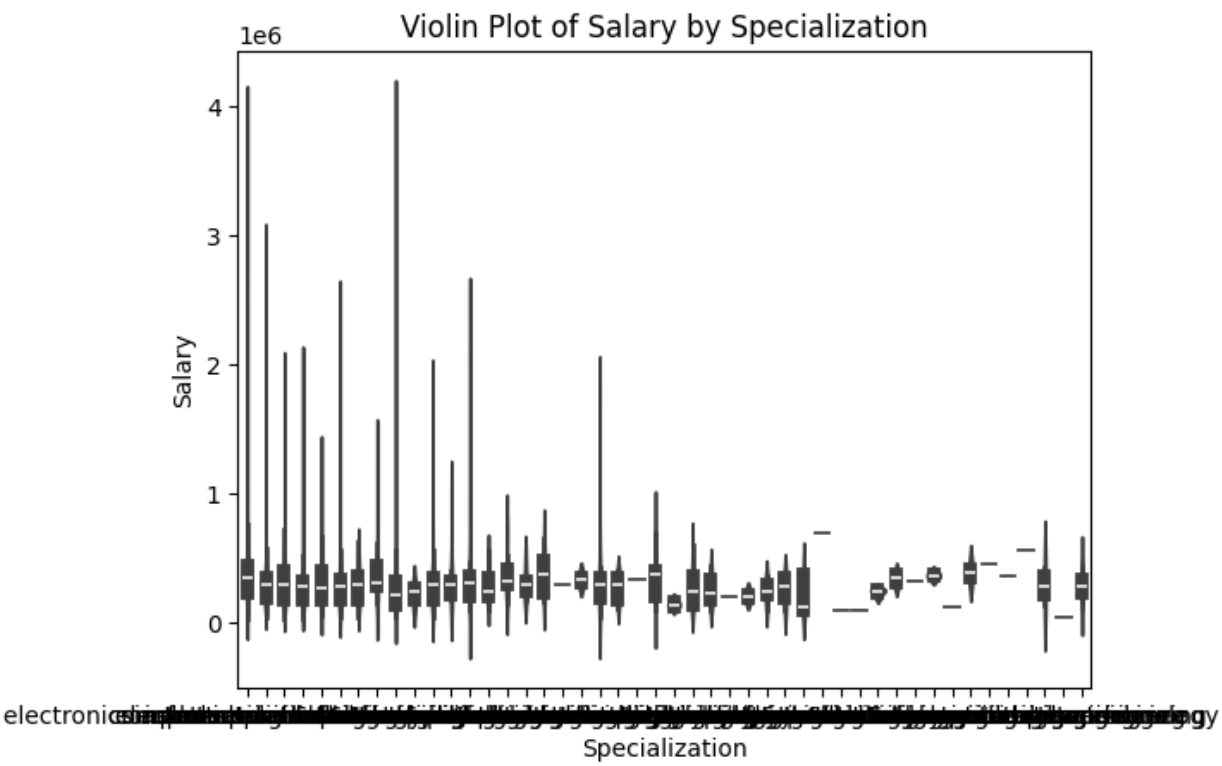

Violin Plot of agreeableness by Degree
agreeableness
4
2
0
−2
−4
−6
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

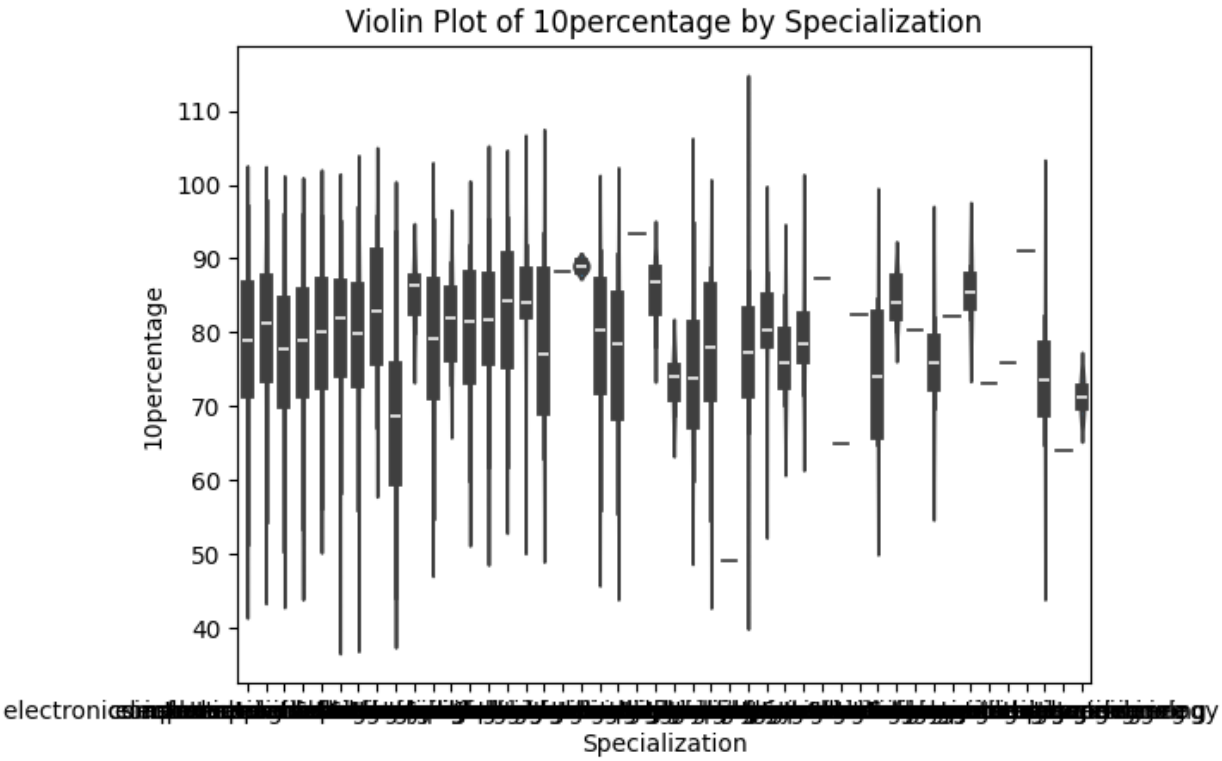

Violin Plot of extraversion by Degree
extraversion
4
2
0
−2
−4
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

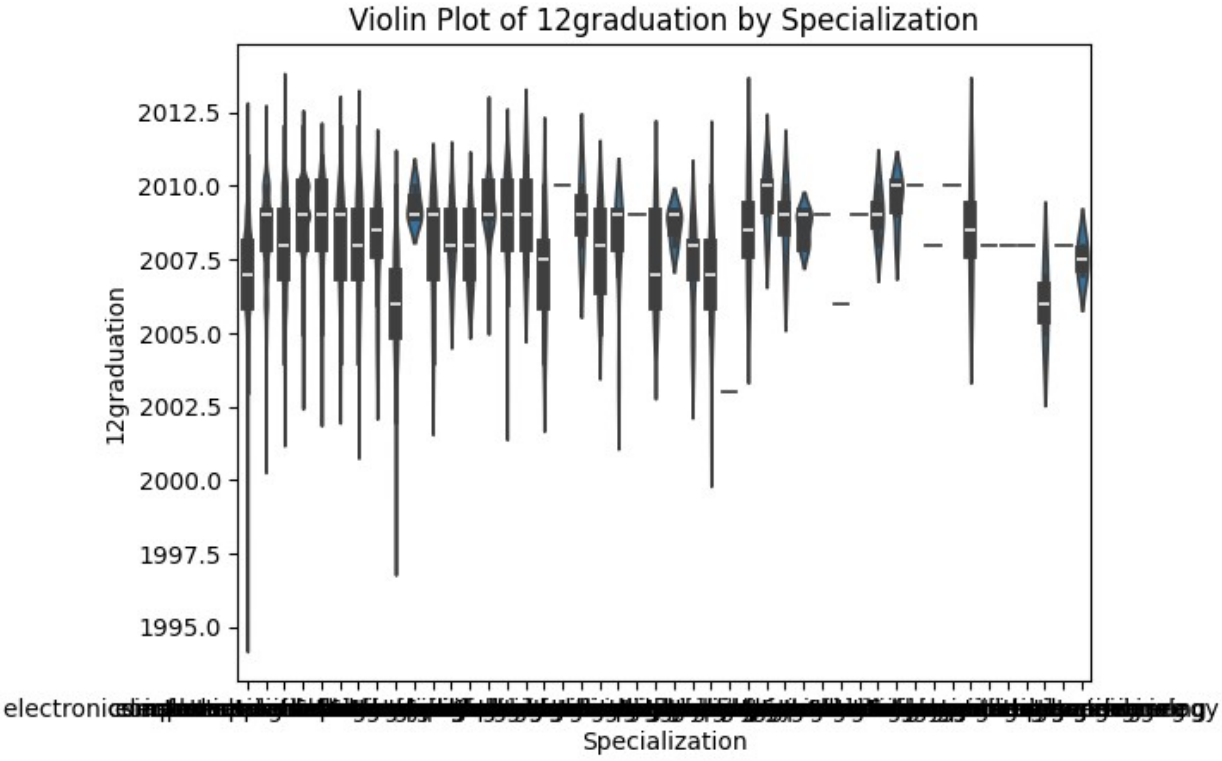

Violin Plot of nueroticism by Degree
nueroticism
4
3
2
1
0
−1
−2
−3
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

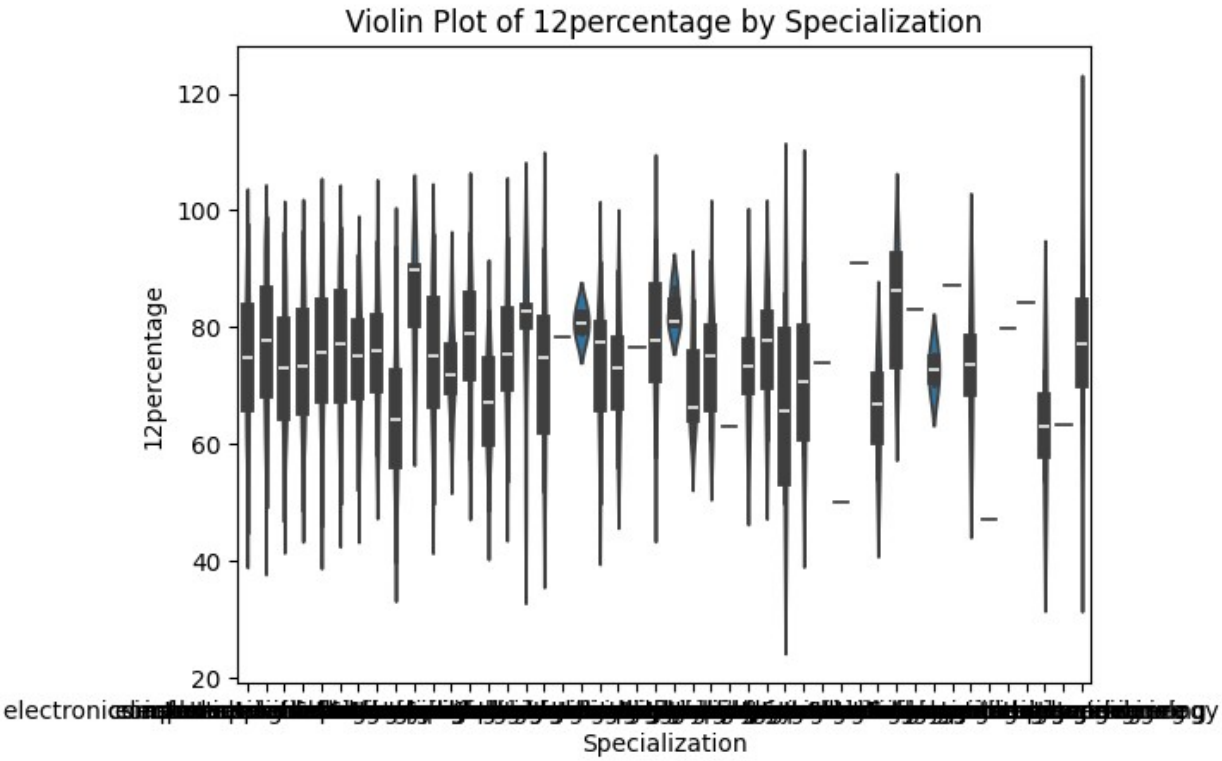

Violin Plot of openess_to_experience by Degree
openess_to_experience
2
0
−2
−4
−6
−8
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

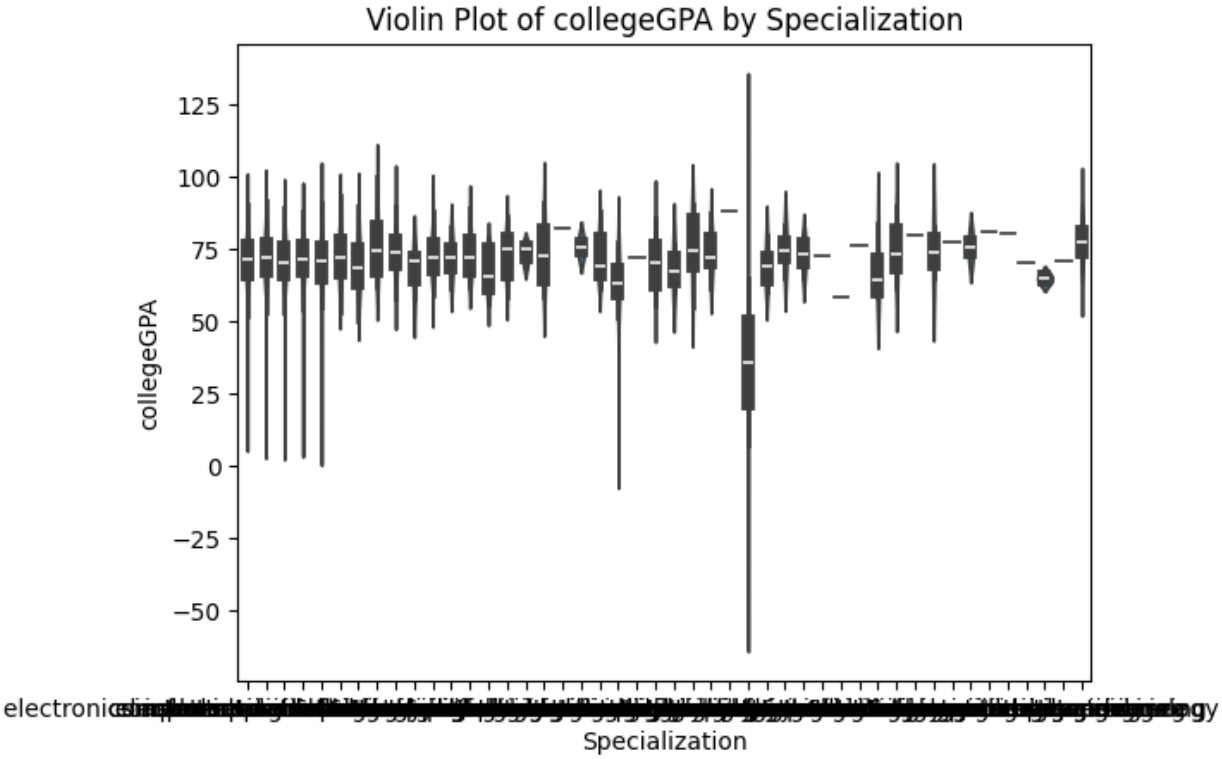

Violin Plot of Age by Degree
Age
35
30
25
20
B.Tech/B.E.
MCA
M.Tech./M.E.
M.Sc. (Tech.)
Degree

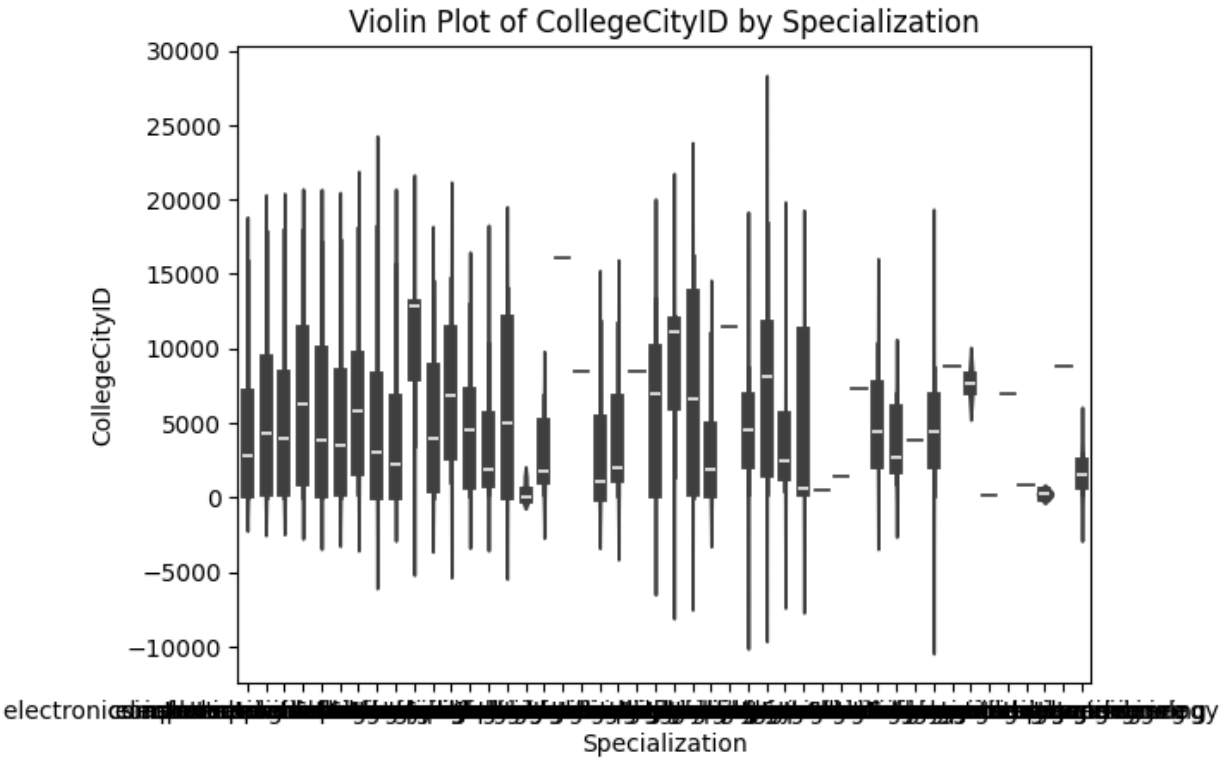

Violin Plot of Salary by Specialization
1e6
4
3
2
1
0
Salary
electronic
Specialization

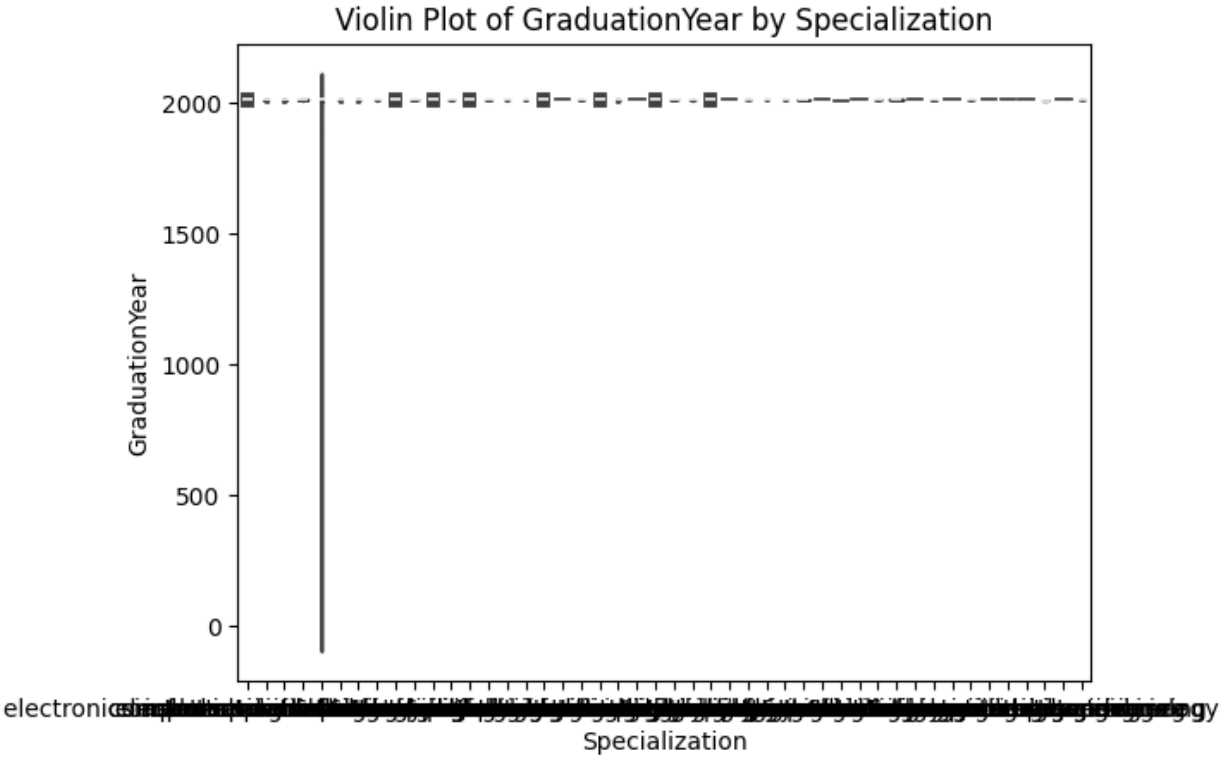

Violin Plot of 10percentage by Specialization
10percentage
110
100
90
80
70
60
50
40
electronic
Specialization

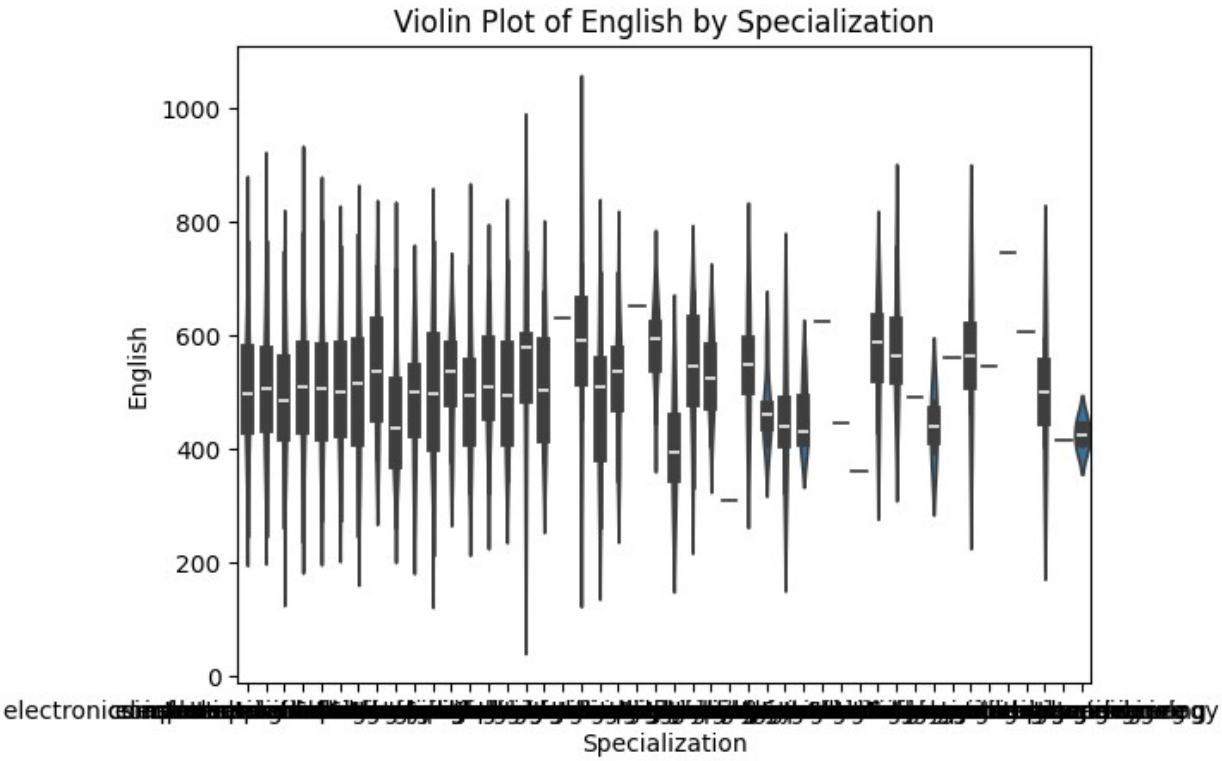

Violin Plot of 12graduation by Specialization
12graduation
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
electronic
Specialization

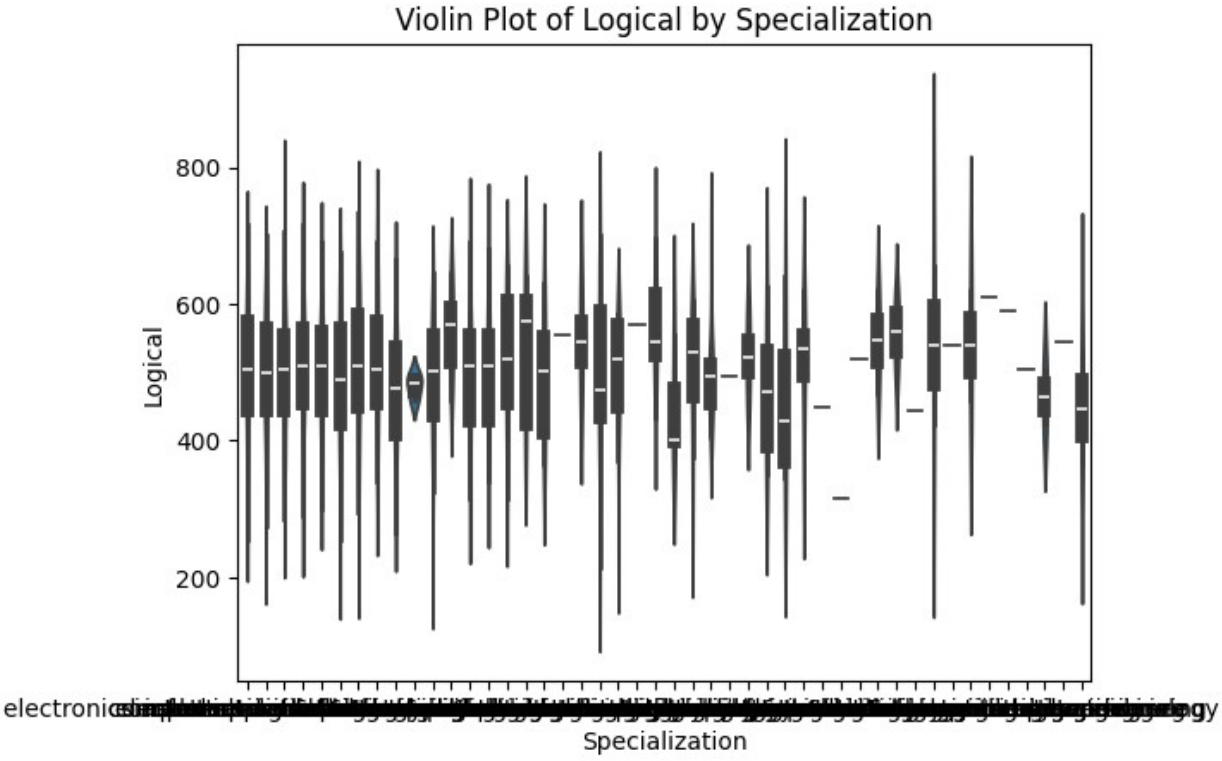

Violin Plot of 12percentage by Specialization
12percentage
120
100
80
60
40
20
electronic
Specialization

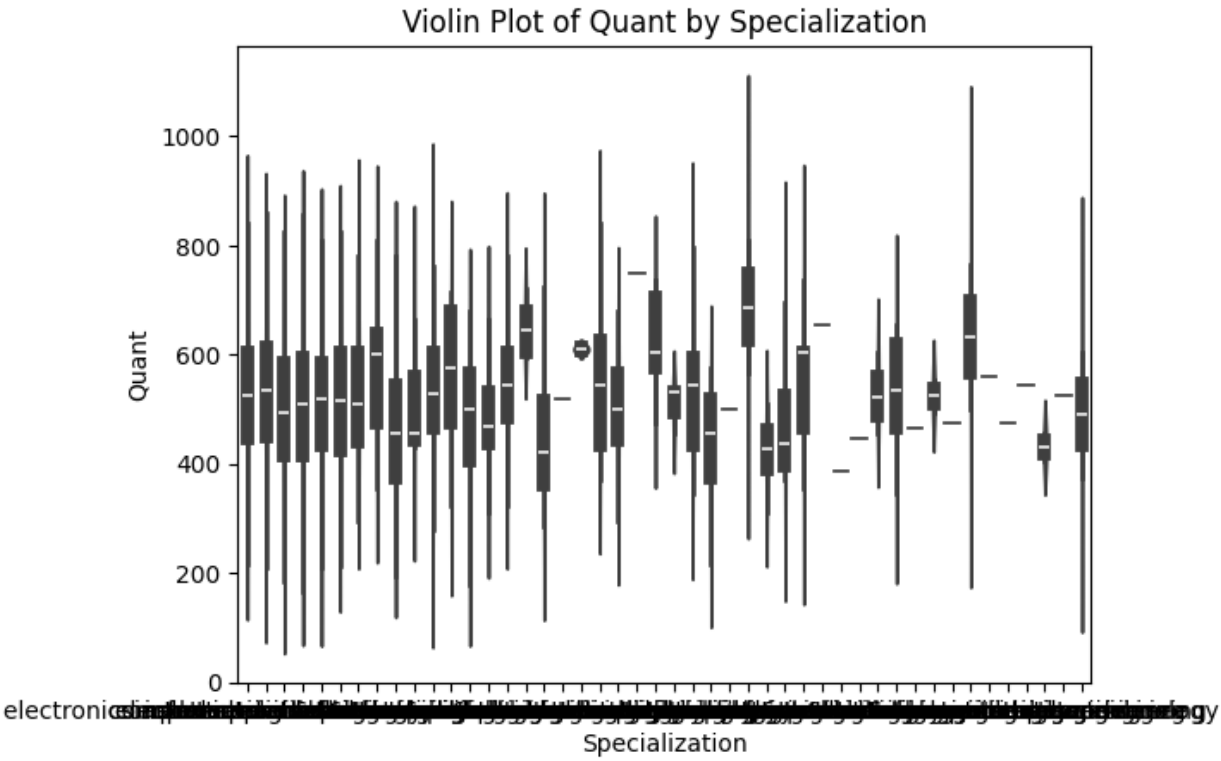

Violin Plot of collegeGPA by Specialization
collegeGPA
125
100
75
50
25
0
−25
−50
electronic
Specialization

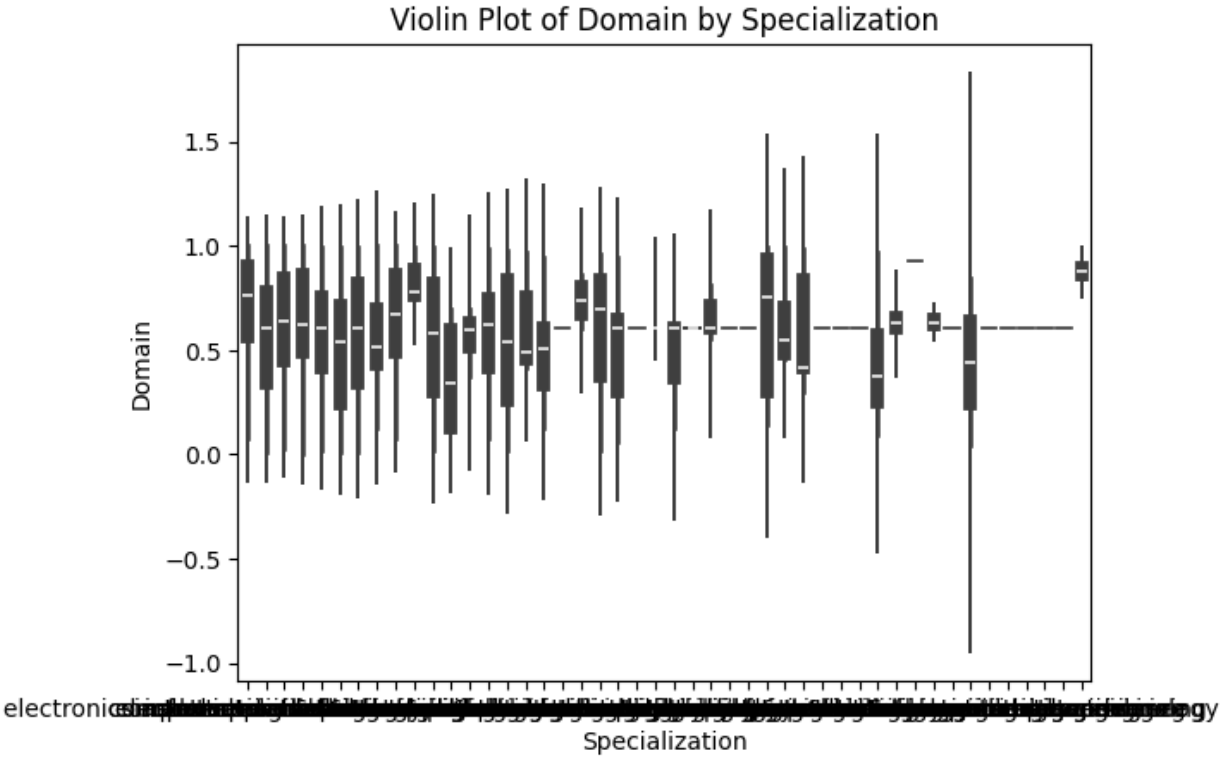

Violin Plot of CollegeCityID by Specialization
CollegeCityID
30000
25000
20000
15000
10000
5000
0
−5000
−10000
electronic
Specialization

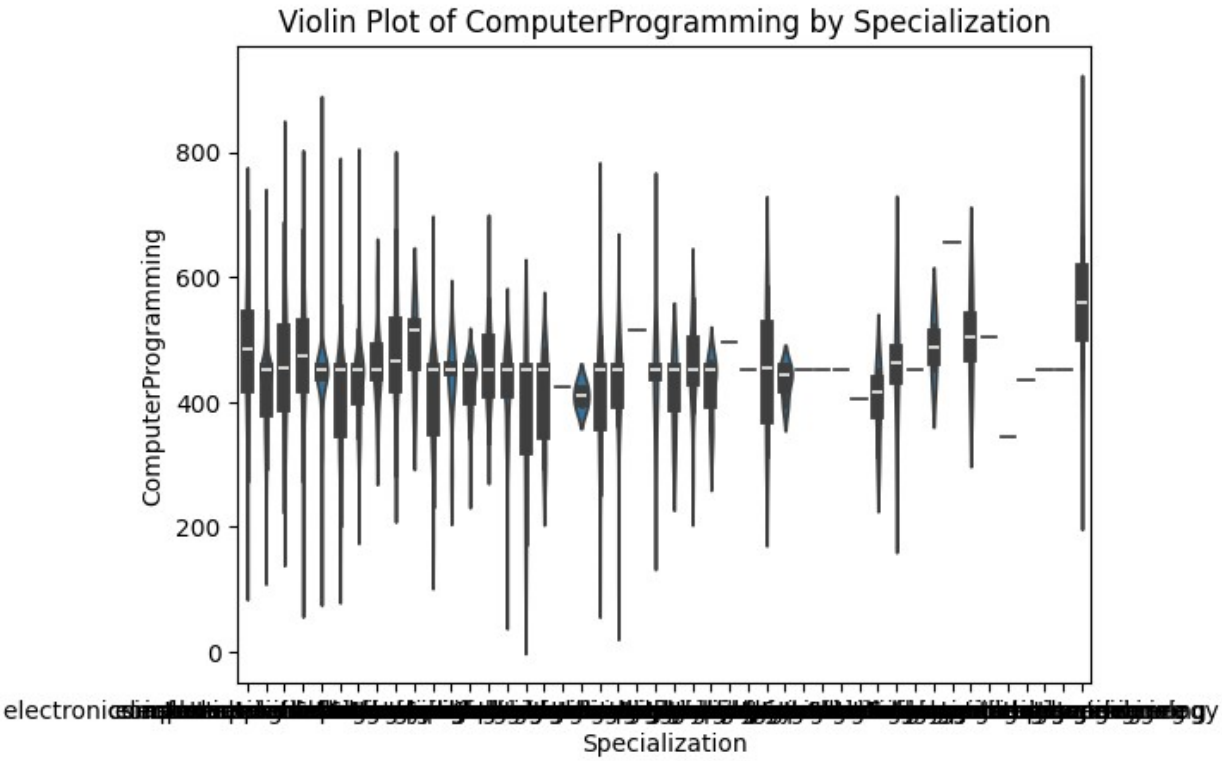

Violin Plot of GraduationYear by Specialization
GraduationYear
2000
1500
1000
500
0
electronic
Specialization

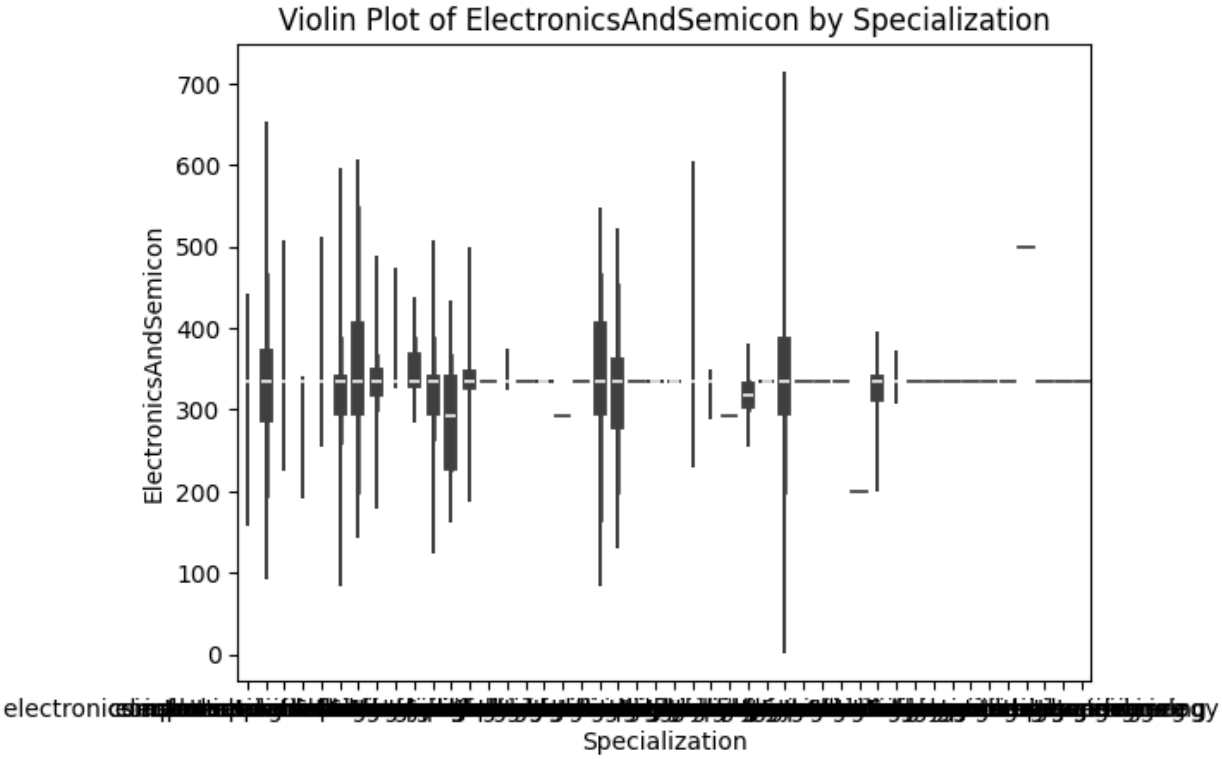

Violin Plot of English by Specialization
English
1000
800
600
400
200
0
electronic
Specialization

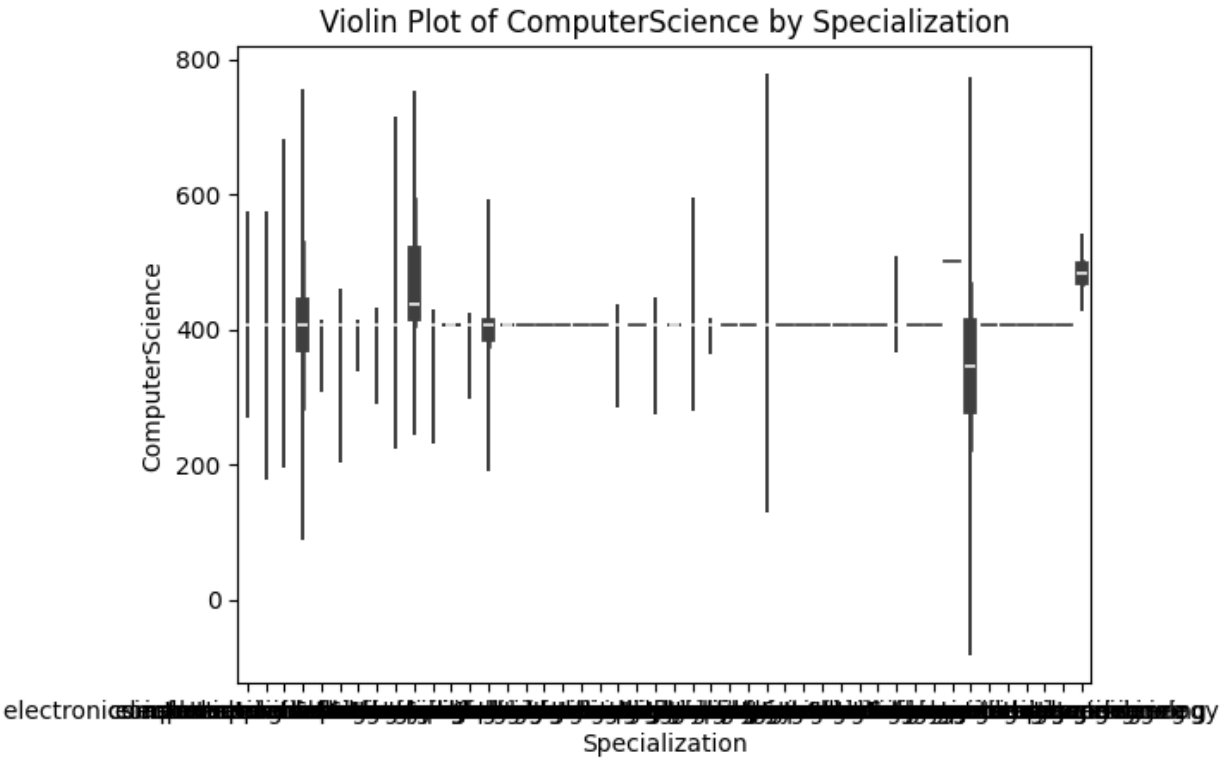

Violin Plot of Logical by Specialization
Logical
800
600
400
200
electronic
Specialization

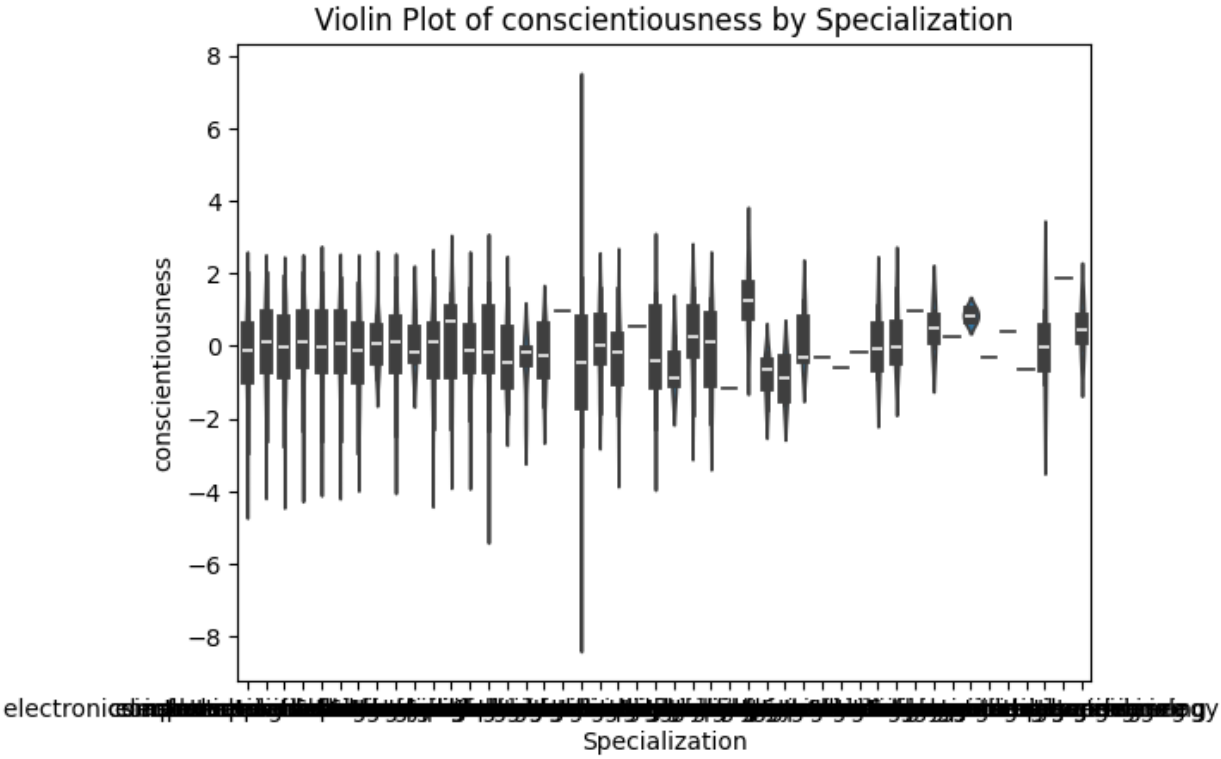

Violin Plot of Quant by Specialization
Quant
1000
800
600
400
200
0
Specialization

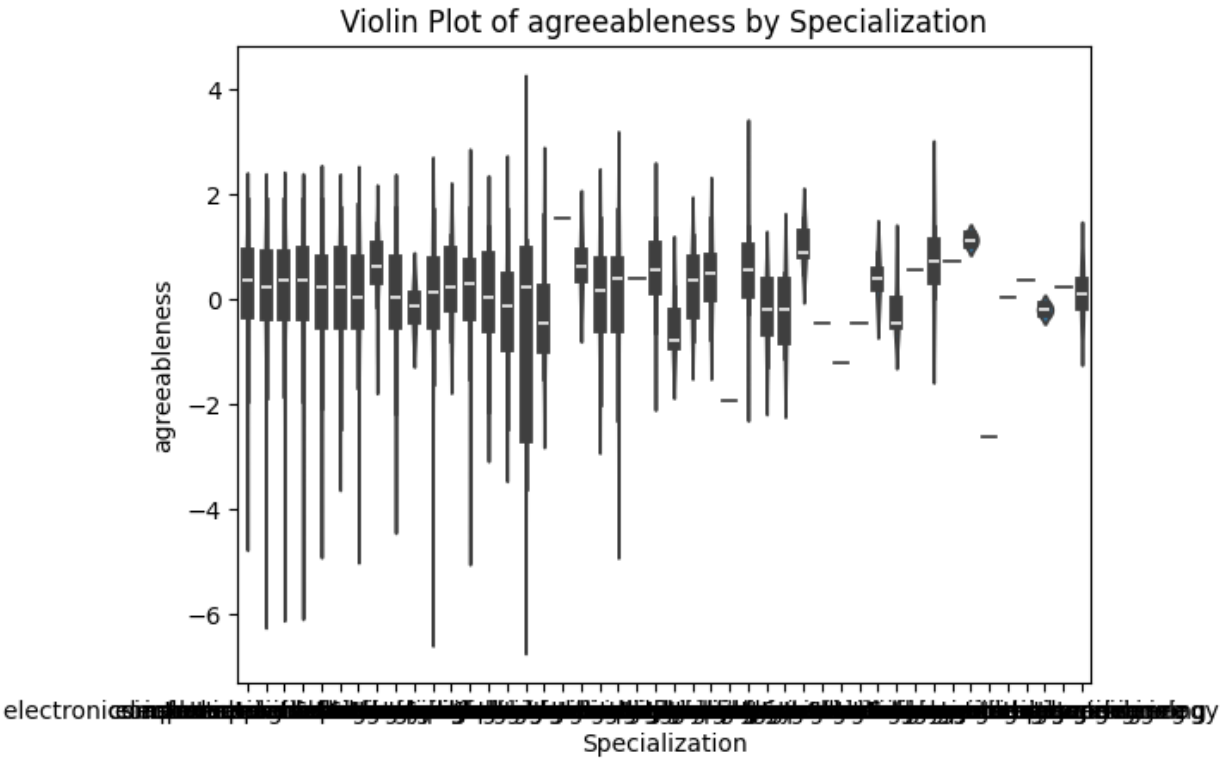

Violin Plot of Domain by Specialization
Domain
1.5
1.0
0.5
0.0
−0.5
−1.0
Specialization

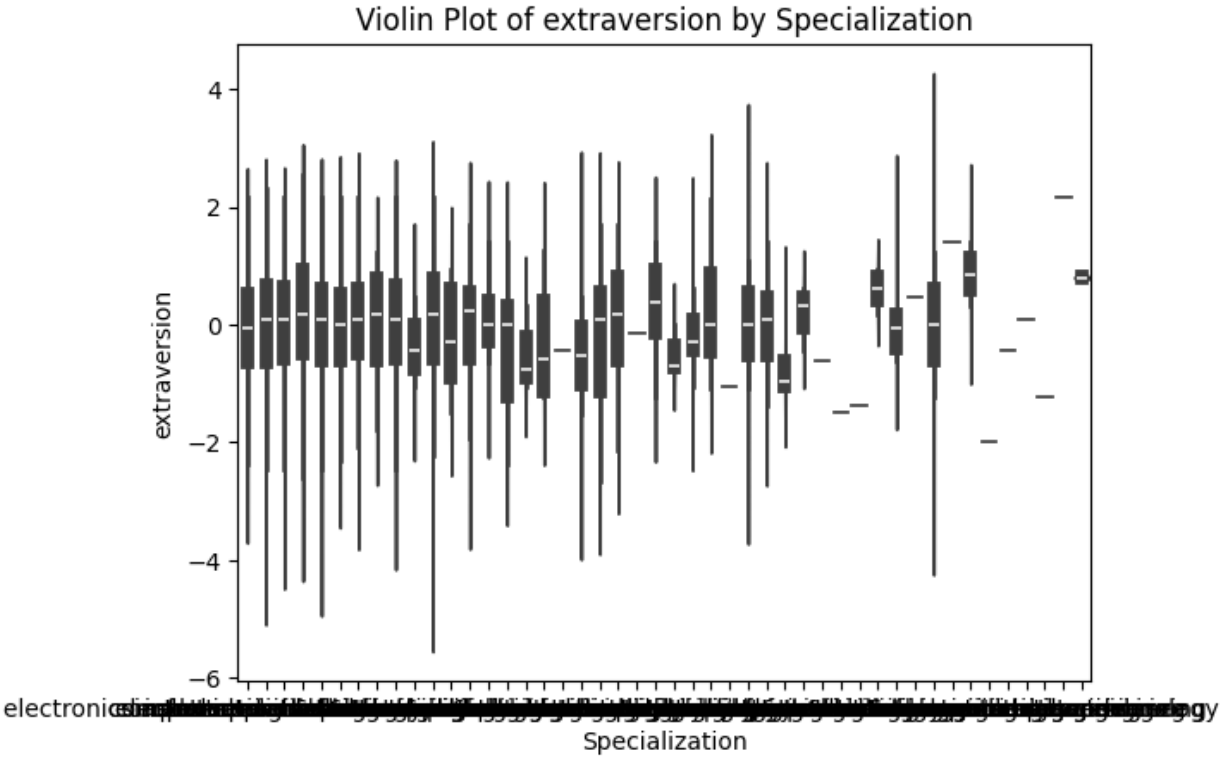

Violin Plot of ComputerProgramming by Specialization
ComputerProgramming
0
200
400
600
800
Specialization

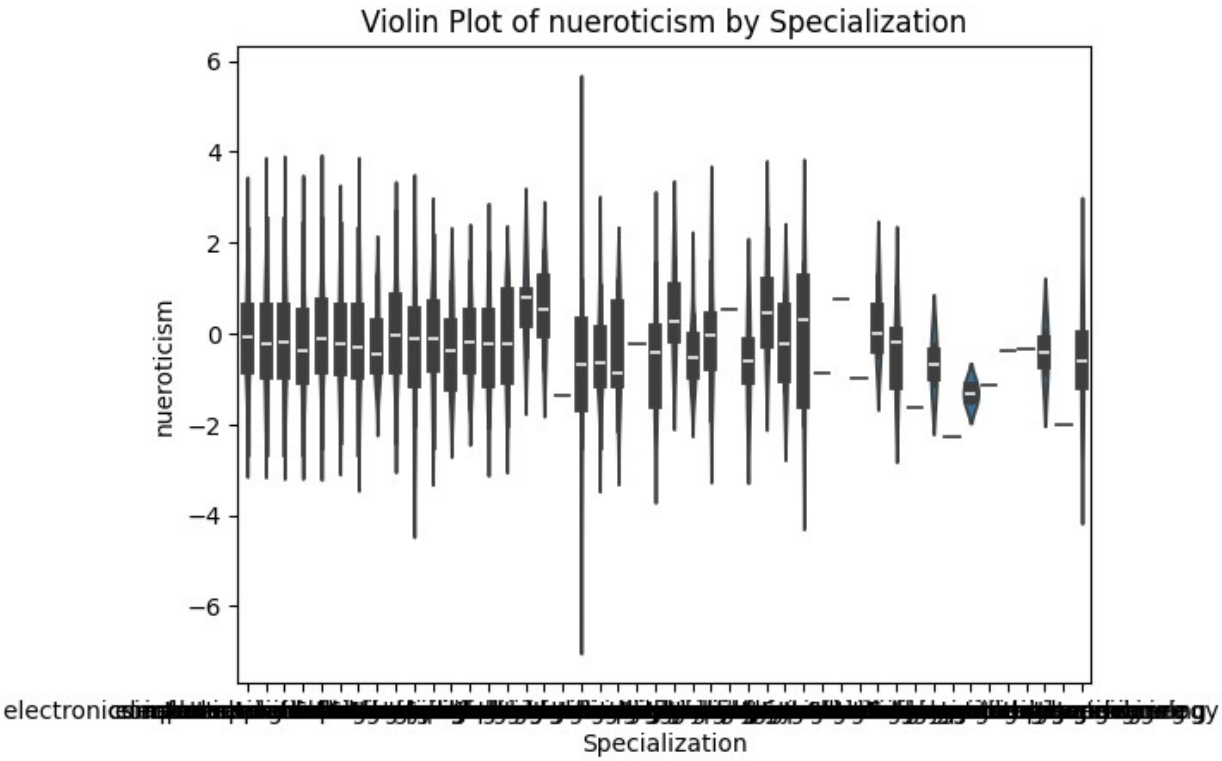

Violin Plot of ElectronicsAndSemicon by Specialization
ElectronicsAndSemicon
0
100
200
300
400
500
600
700
Specialization

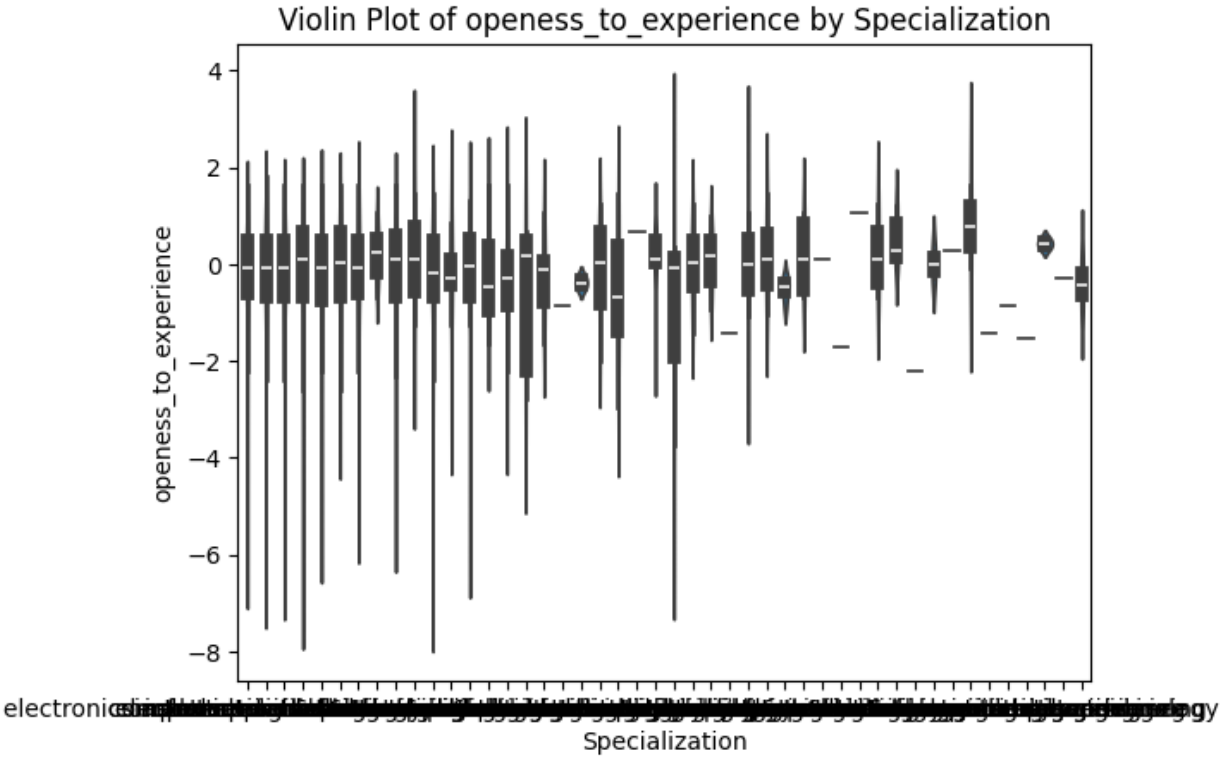

Violin Plot of ComputerScience by Specialization
ComputerScience
800
600
400
200
0
electronic
Specialization

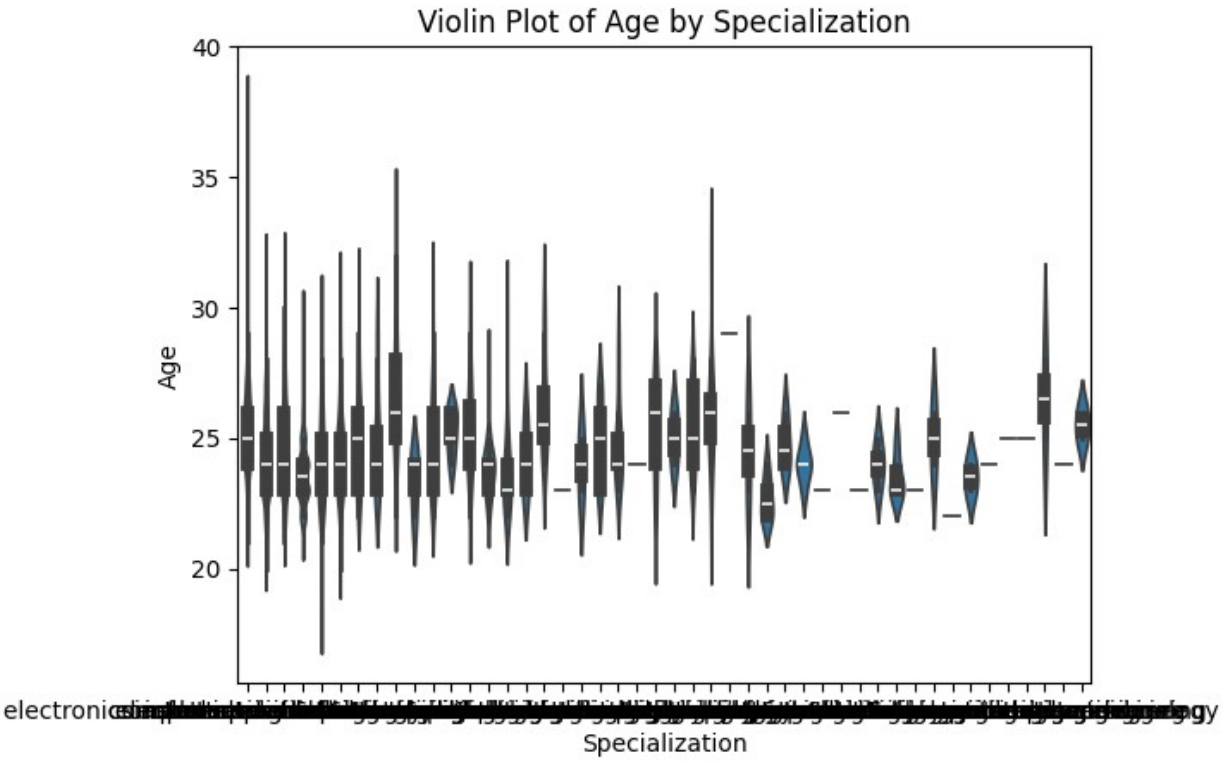

Violin Plot of conscientiousness by Specialization
conscientiousness
8
6
4
2
0
−2
−4
−6
−8
electronic
Specialization

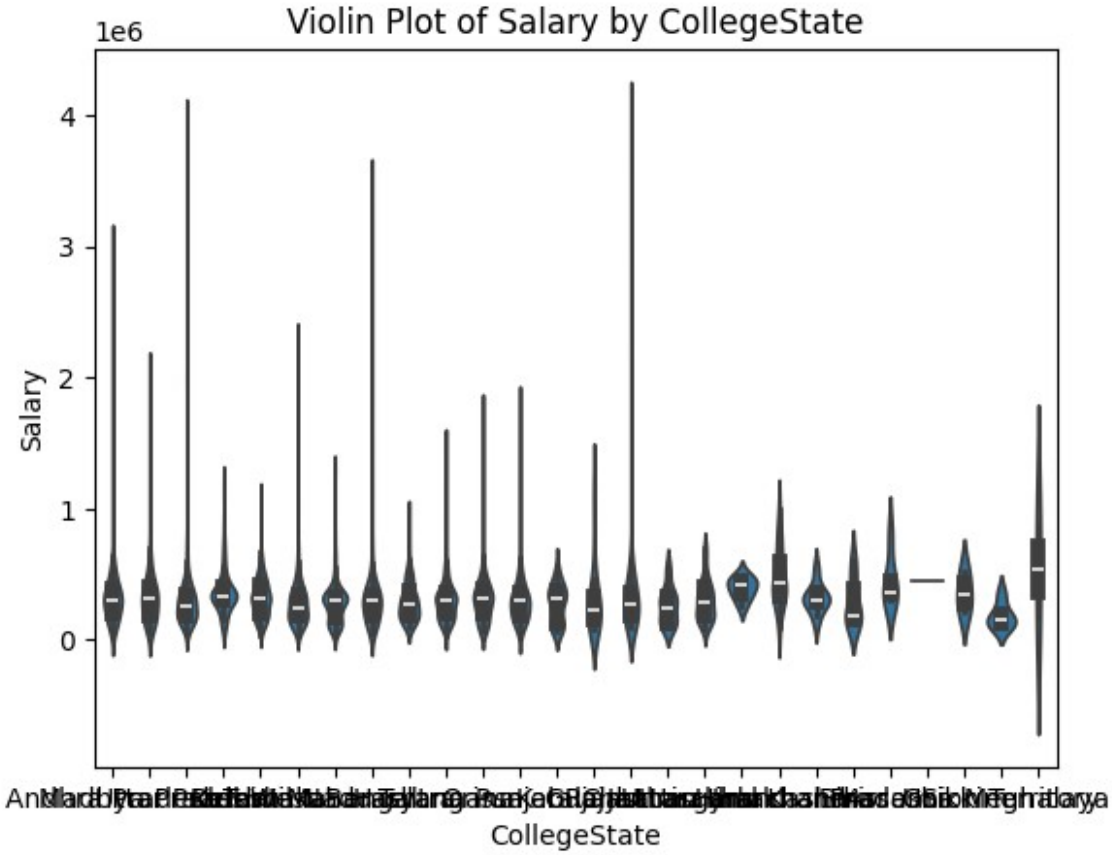

Violin Plot of agreeableness by Specialization
agreeableness
4
2
0
−2
−4
−6
electronic
Specialization

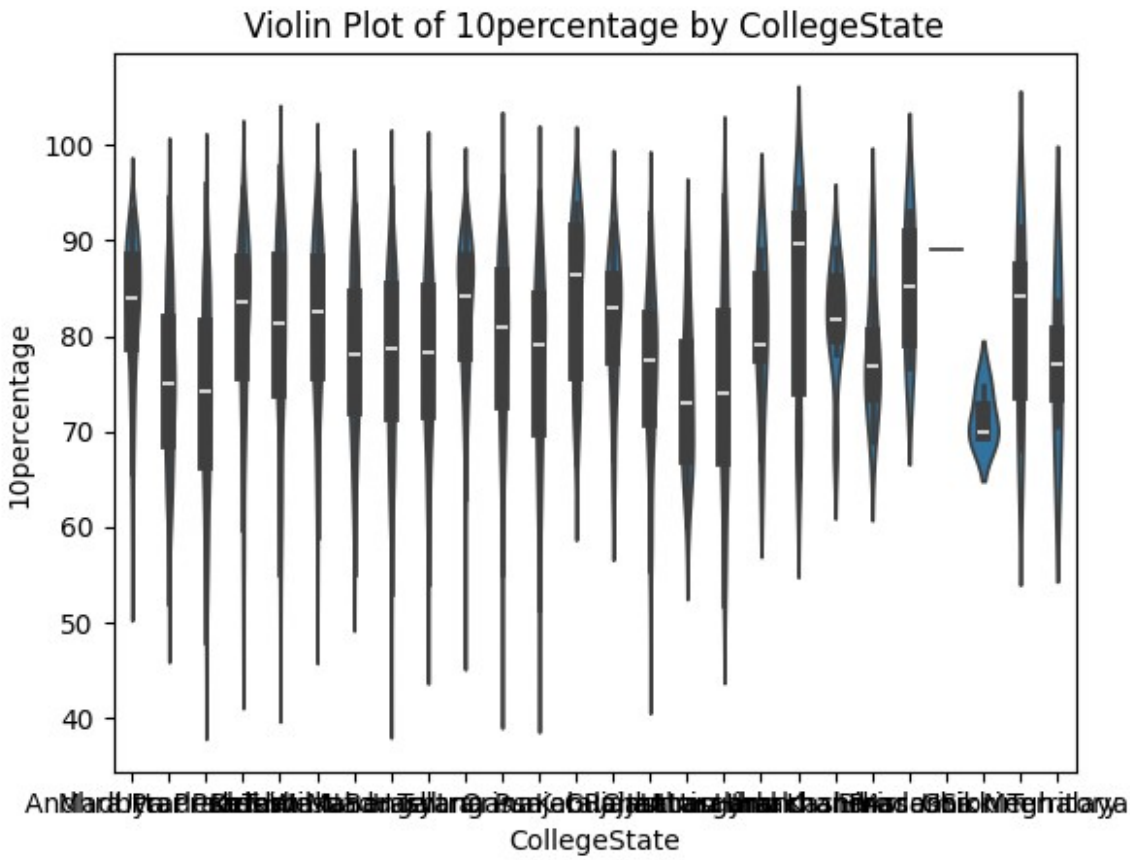

Violin Plot of extraversion by Specialization
extraversion
4
2
0
−2
−4
−6
electronic
Specialization

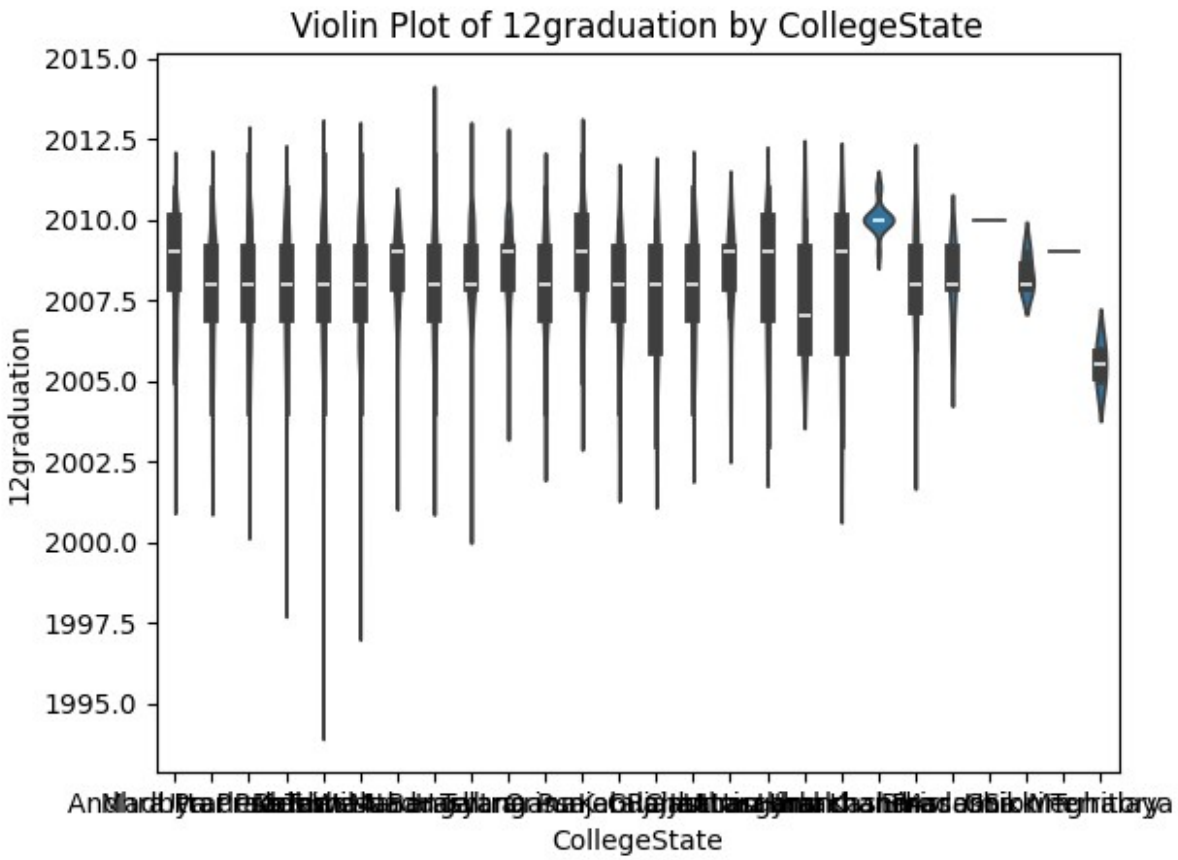

Violin Plot of nueroticism by Specialization
nueroticism
6
4
2
0
−2
−4
−6
electronic
Specialization

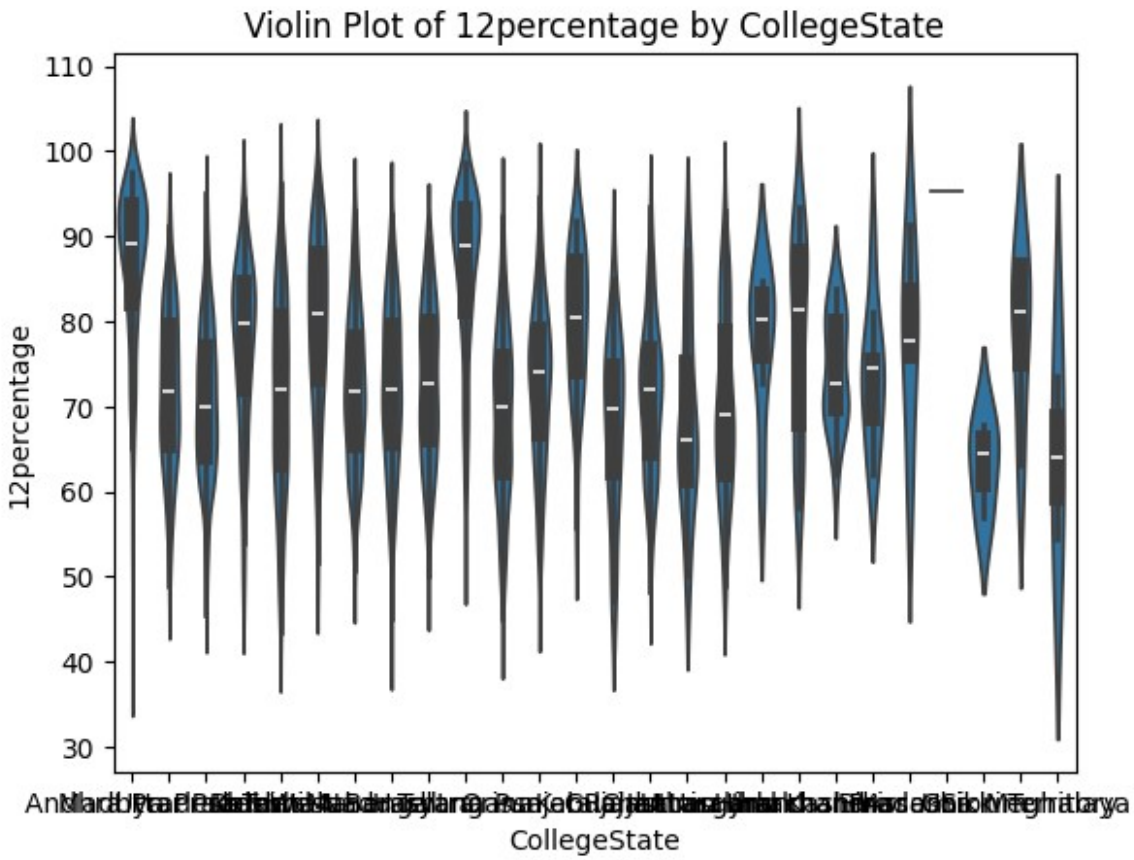

Violin Plot of openess_to_experience by Specialization
openess_to_experience
4
2
0
−2
−4
−6
−8
electronic
Specialization

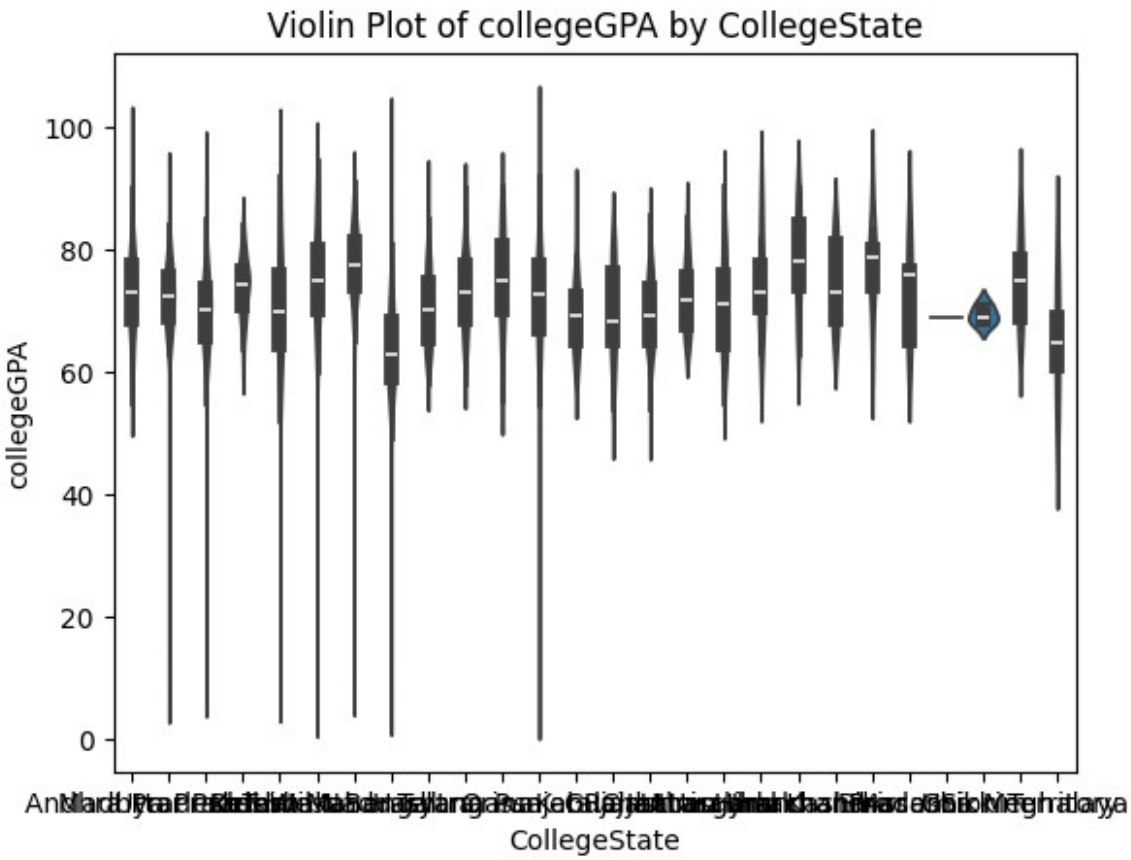

Violin Plot of Age by Specialization
40
35
30
25
20
Age
electronic
Specialization

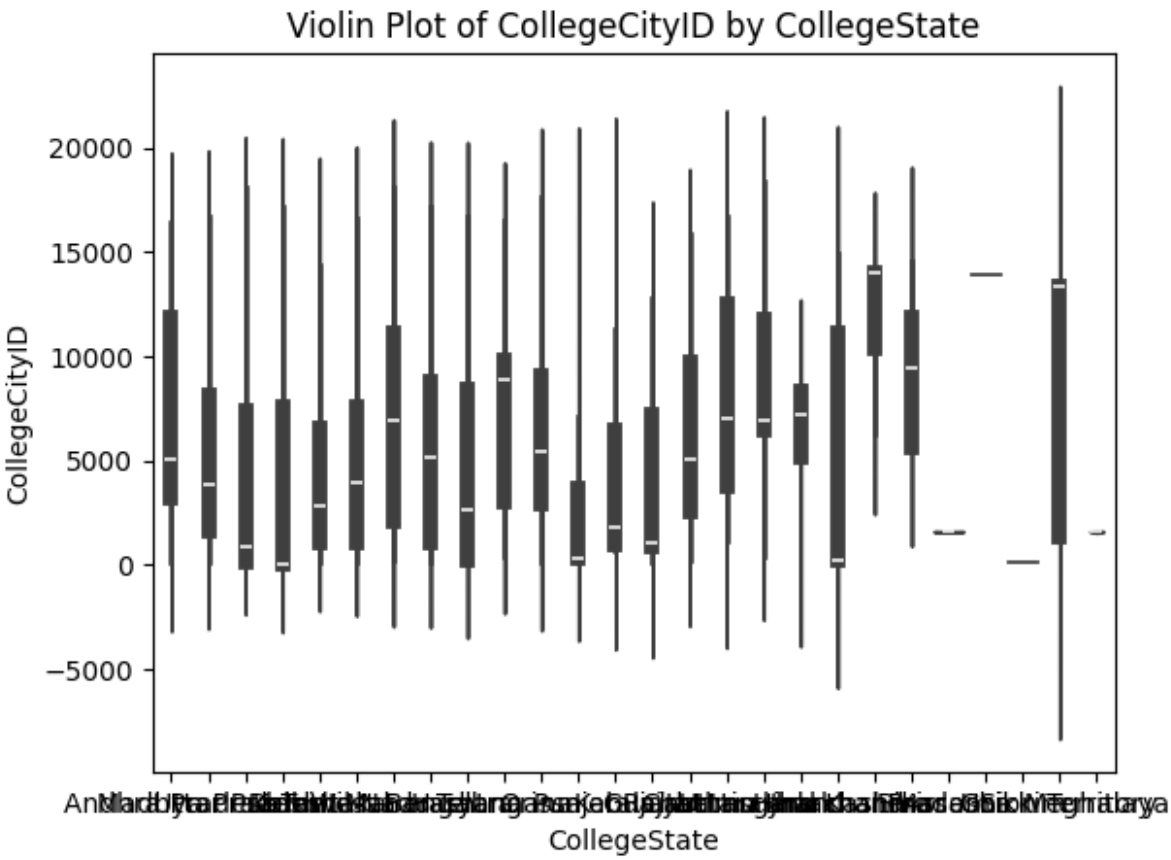

Violin Plot of Salary by CollegeState
1e6
4
3
2
1
0
Salary
CollegeState

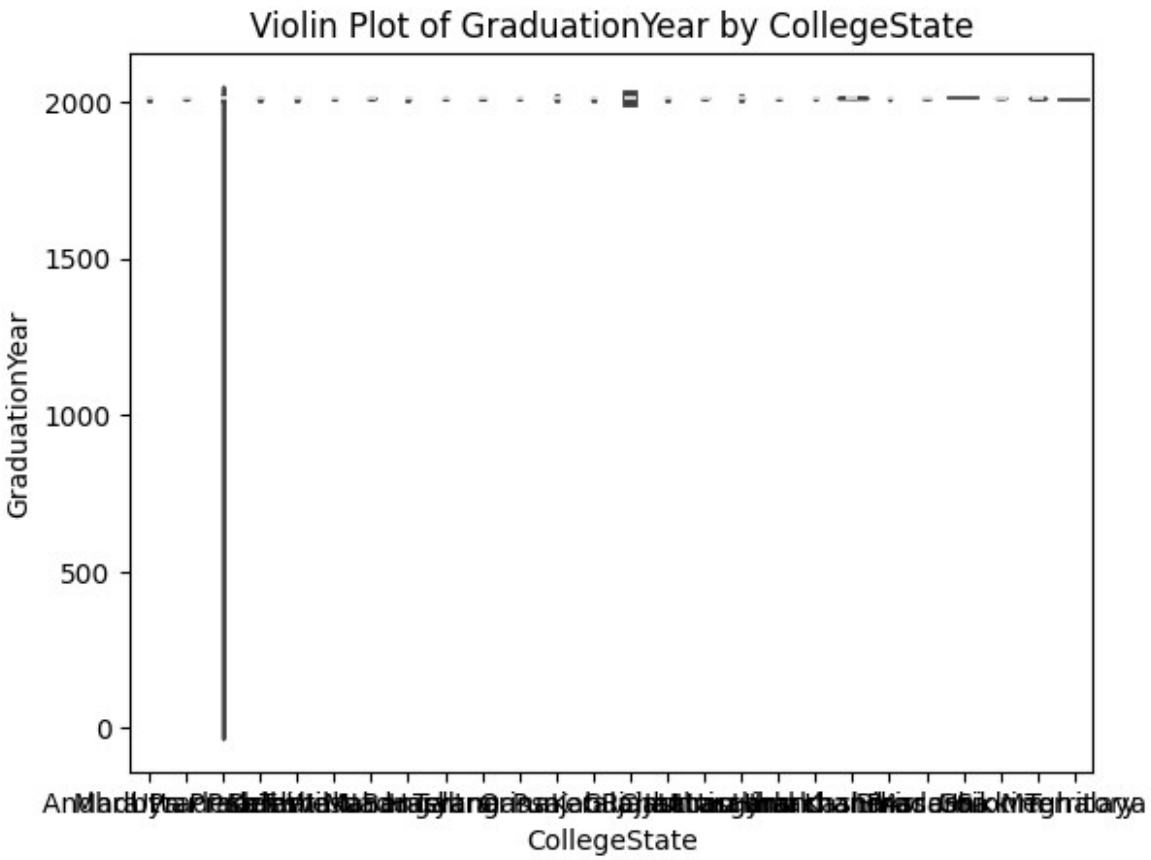

Violin Plot of 10percentage by CollegeState
10percentage
100
90
80
70
60
50
40
CollegeState

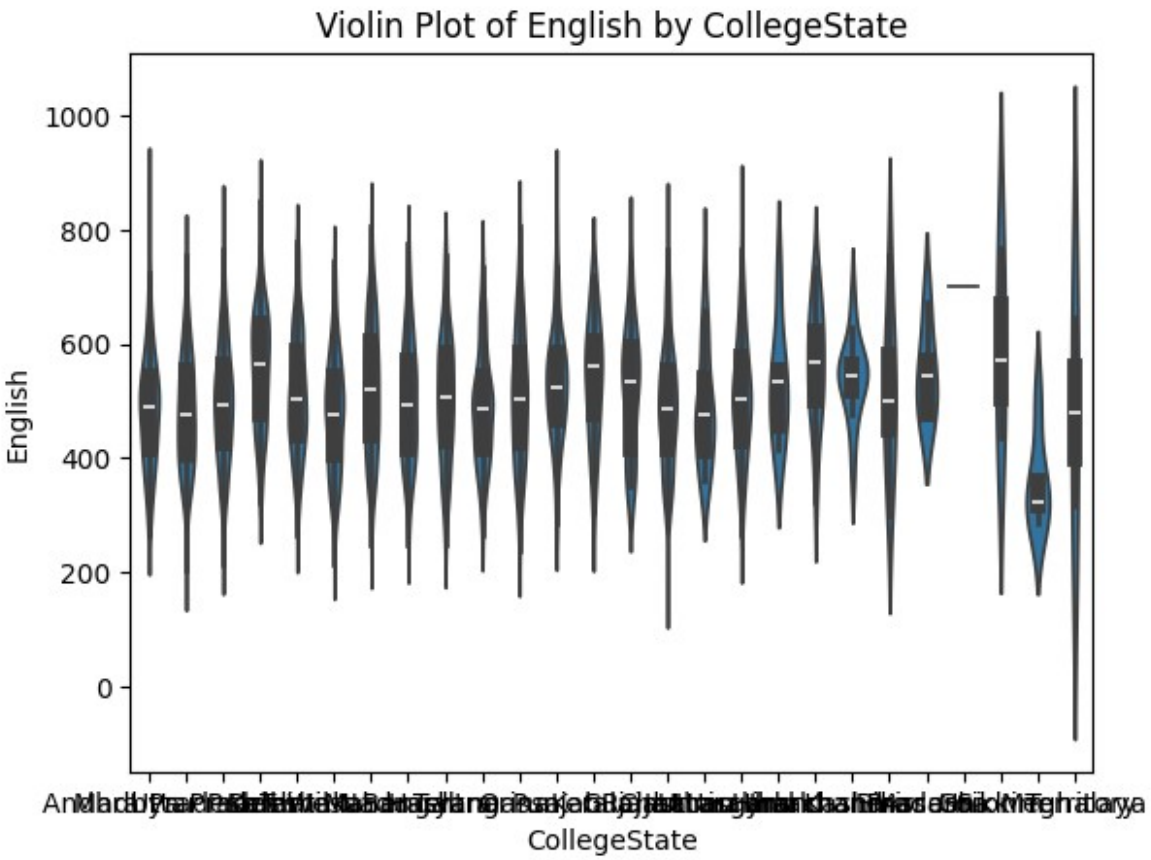

Violin Plot of 12graduation by CollegeState
2015.0
2012.5
2010.0
2007.5
2005.0
2002.5
2000.0
1997.5
1995.0
12graduation
CollegeState

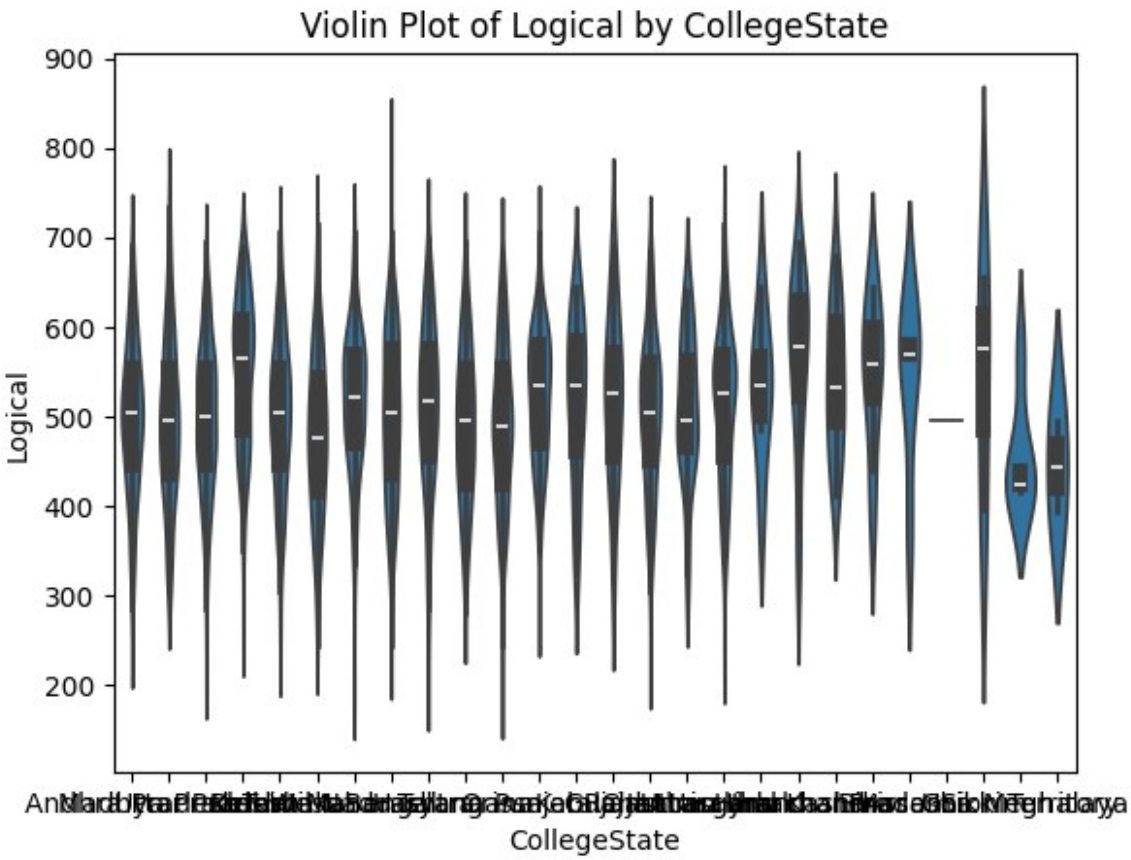

Violin Plot of 12percentage by CollegeState
12percentage
110
100
90
80
70
60
50
40
30
CollegeState

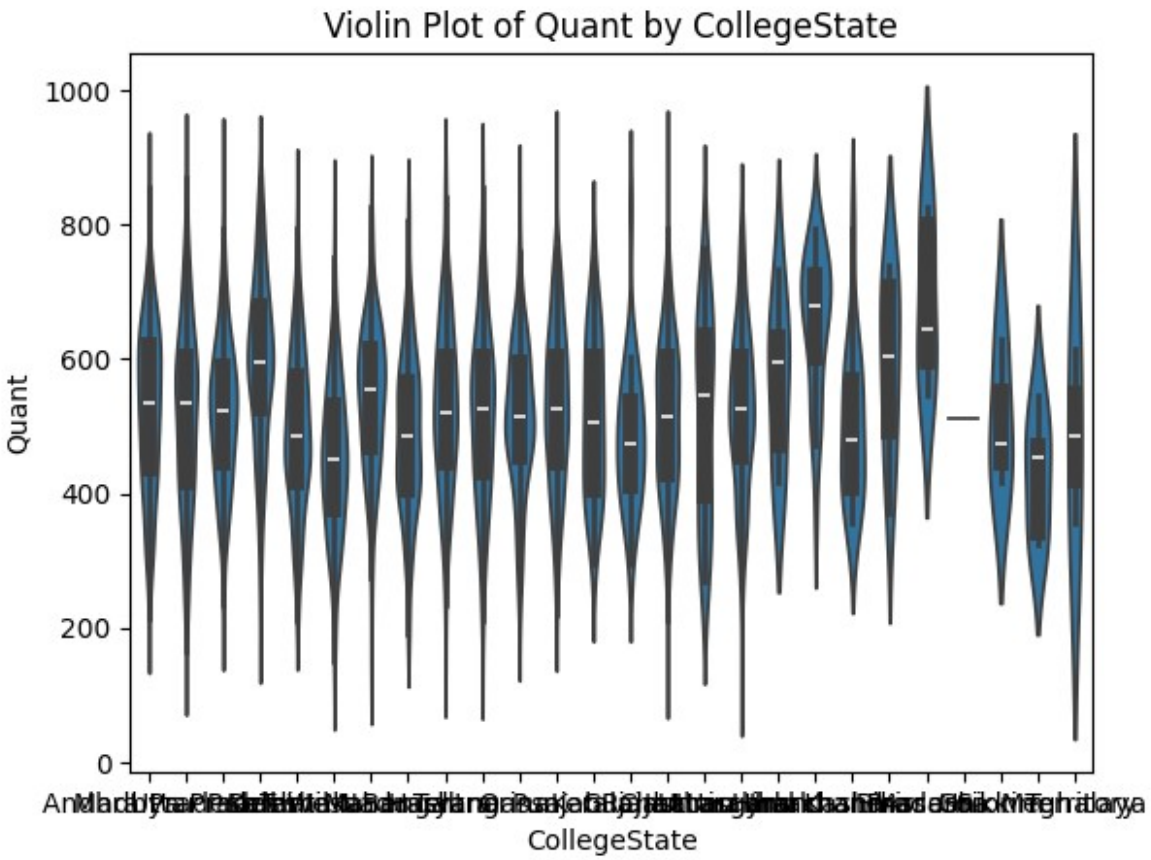

Violin Plot of collegeGPA by CollegeState
collegeGPA
0
20
40
60
80
100
CollegeState

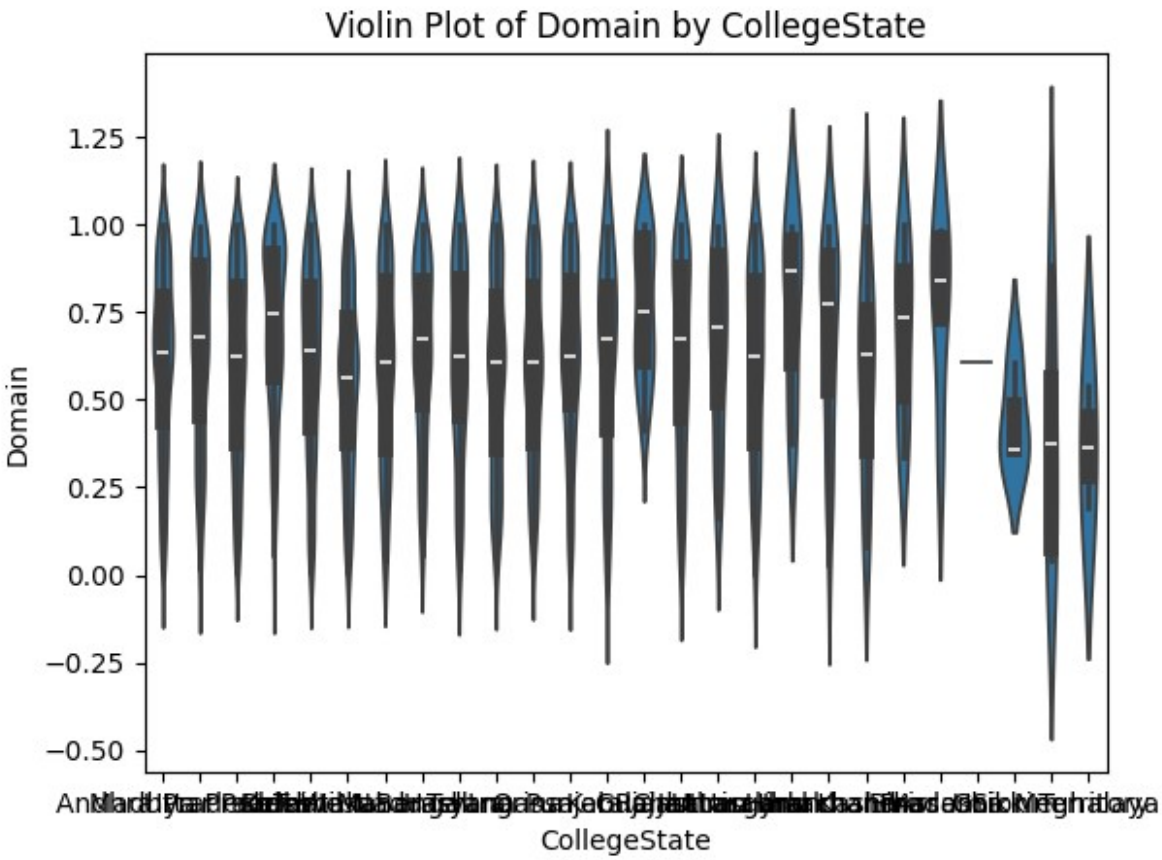

Violin Plot of CollegeCityID by CollegeState
CollegeCityID
20000
15000
10000
5000
0
−5000
CollegeState

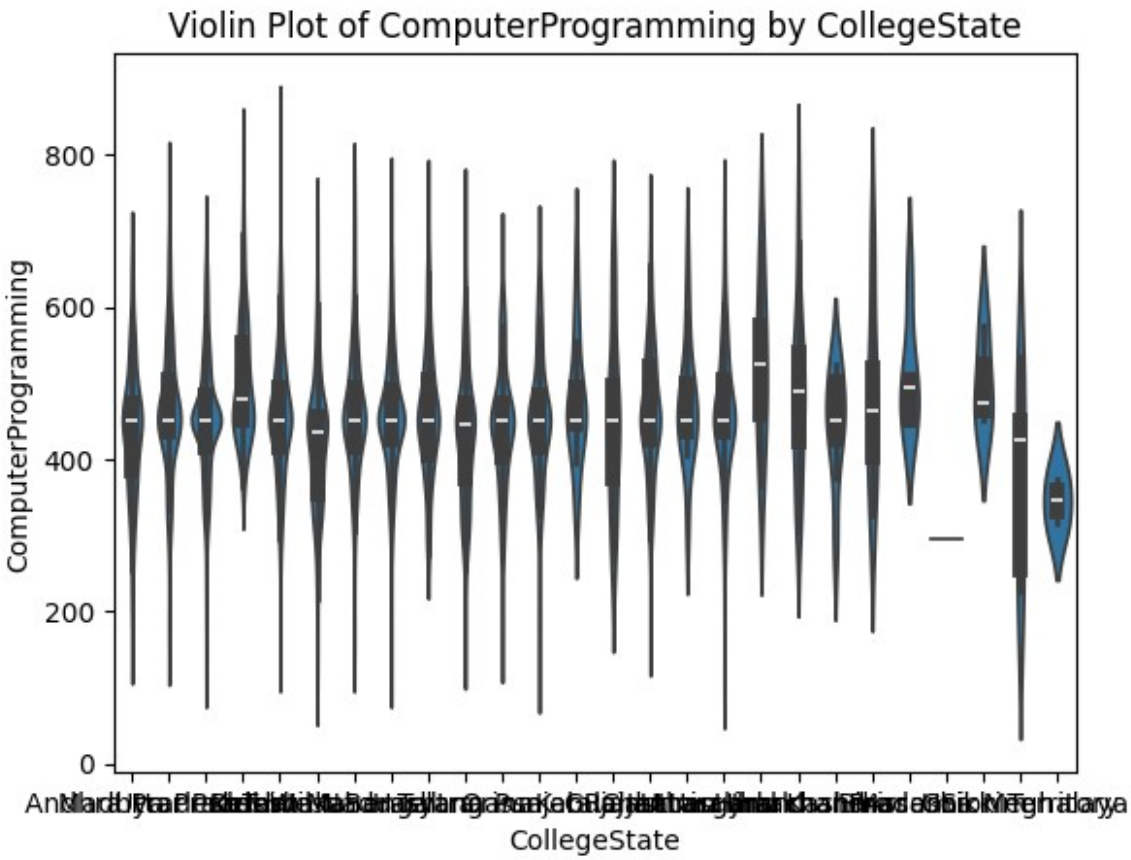

Violin Plot of GraduationYear by CollegeState
2000
1500
1000
500
0
GraduationYear
CollegeState

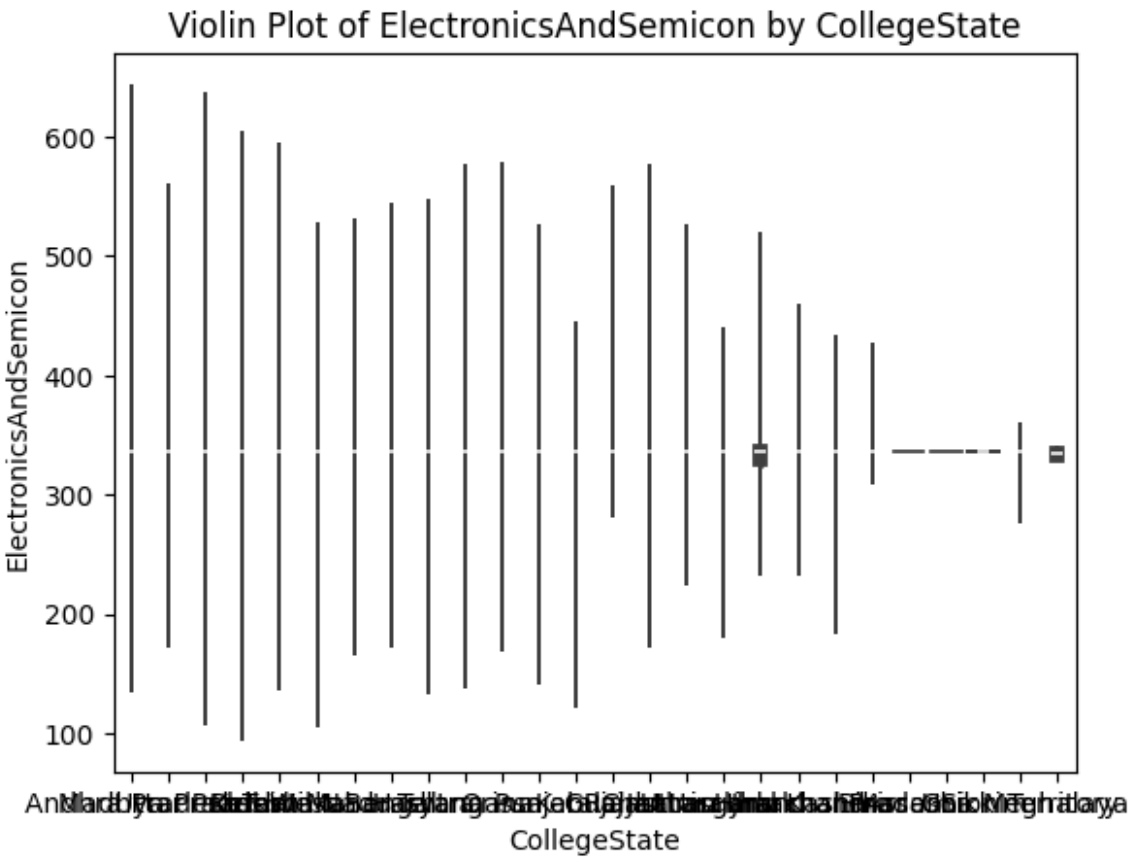

Violin Plot of English by CollegeState
English
1000
800
600
400
200
0
CollegeState

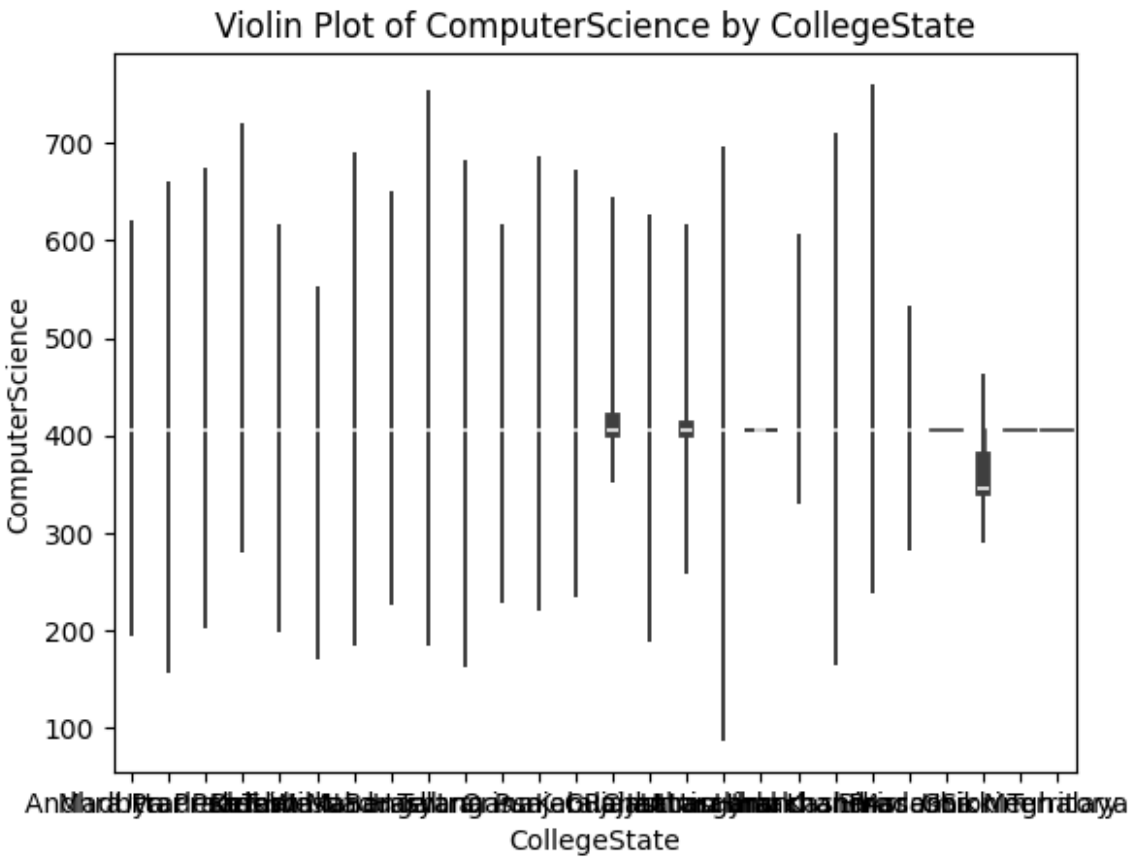

Violin Plot of Logical by CollegeState
Logical
900
800
700
600
500
400
300
200
CollegeState

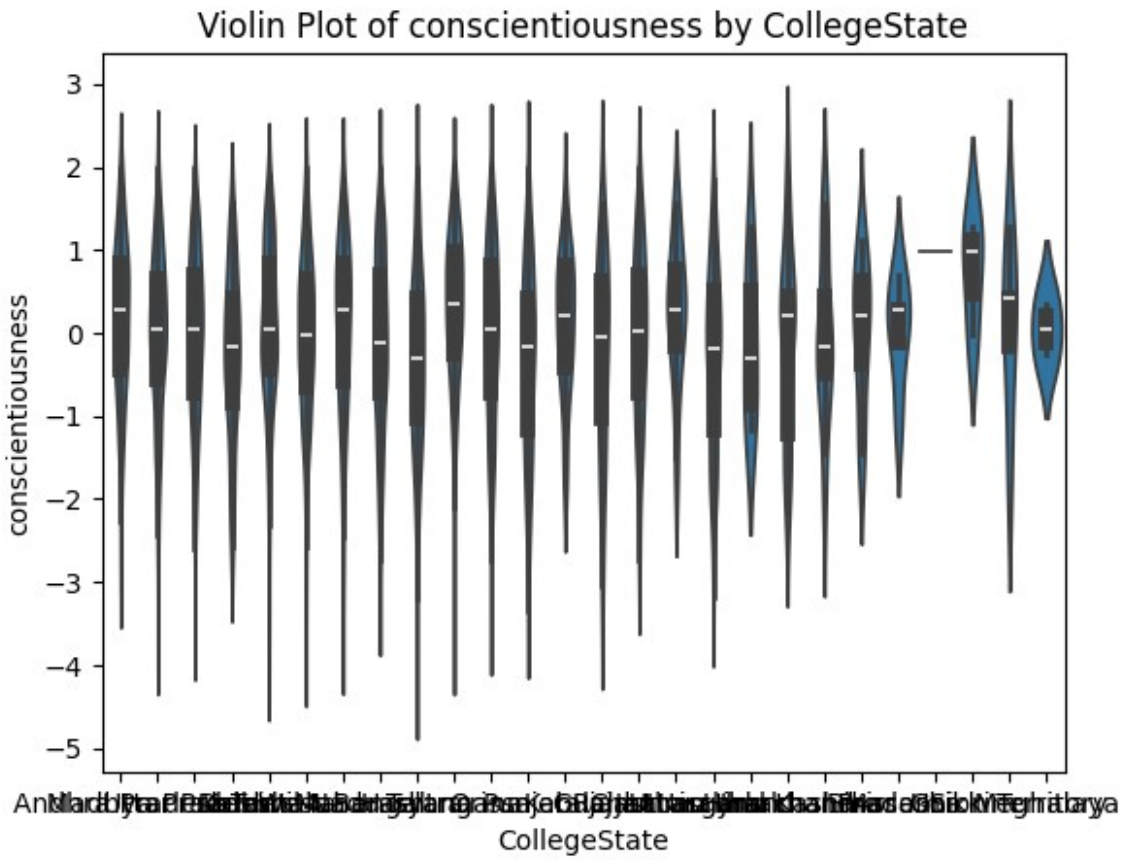

Violin Plot of Quant by CollegeState
Quant
1000
800
600
400
200
0
CollegeState

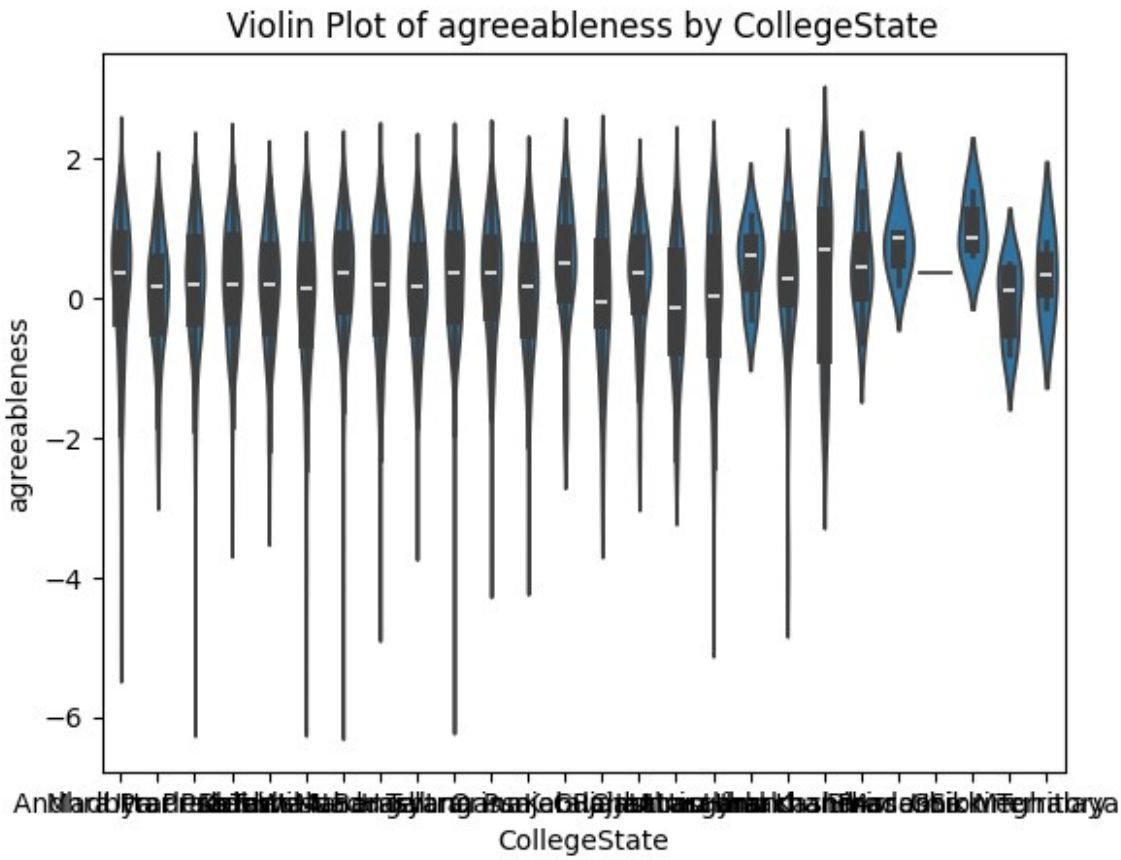

Violin Plot of Domain by CollegeState
Domain
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
CollegeState

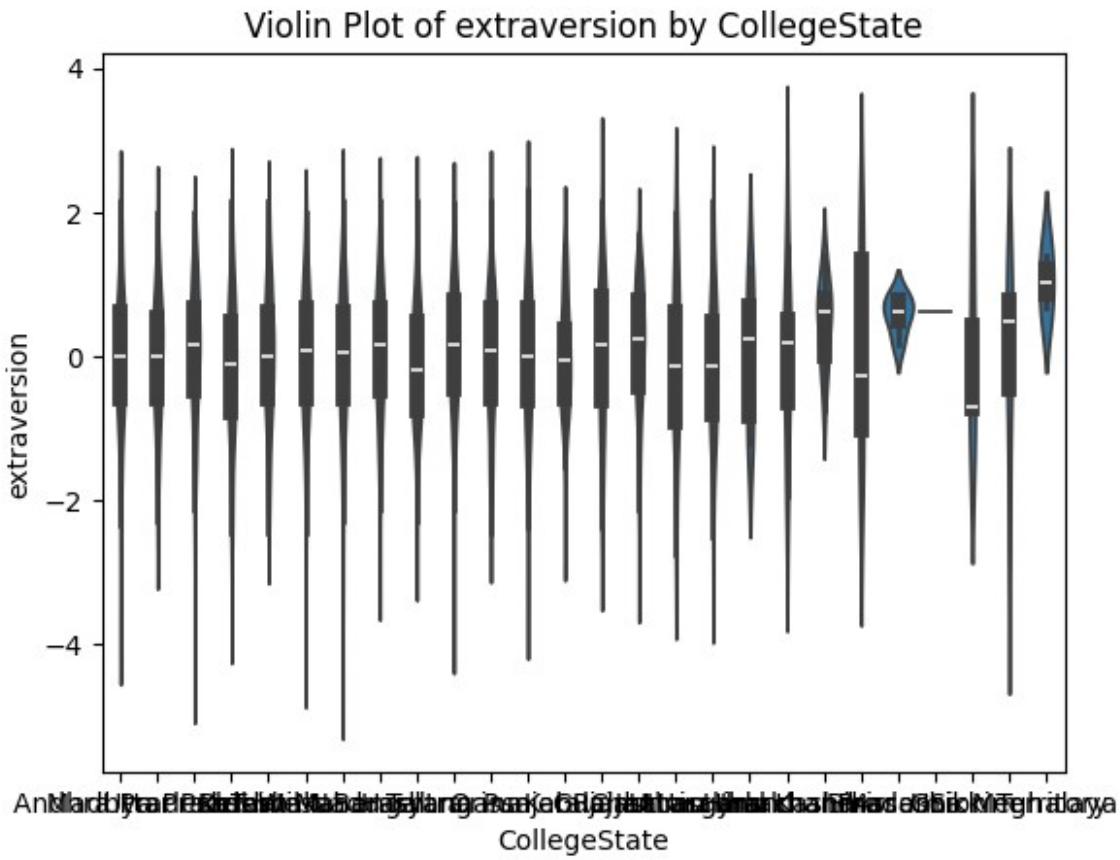

Violin Plot of ComputerProgramming by CollegeState
ComputerProgramming
0
200
400
600
800
CollegeState

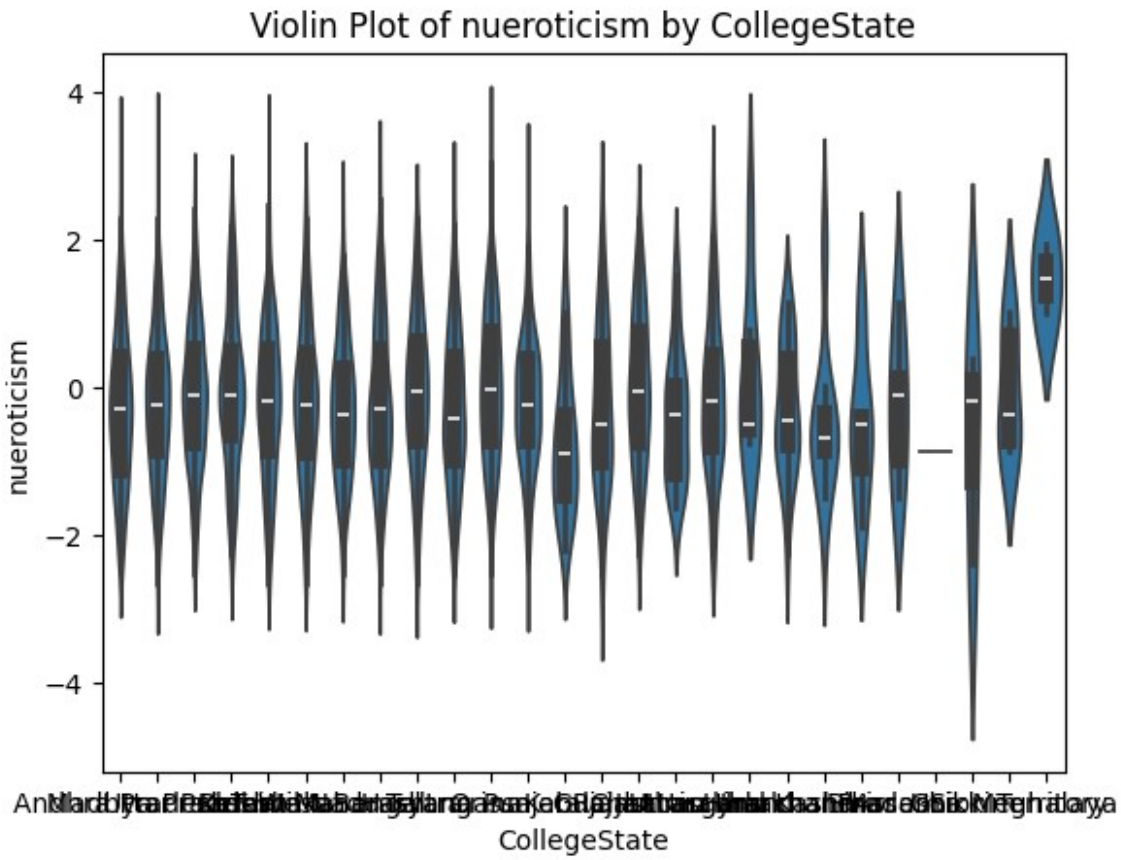

Violin Plot of ElectronicsAndSemicon by CollegeState
ElectronicsAndSemicon
600
500
400
300
200
100
CollegeState

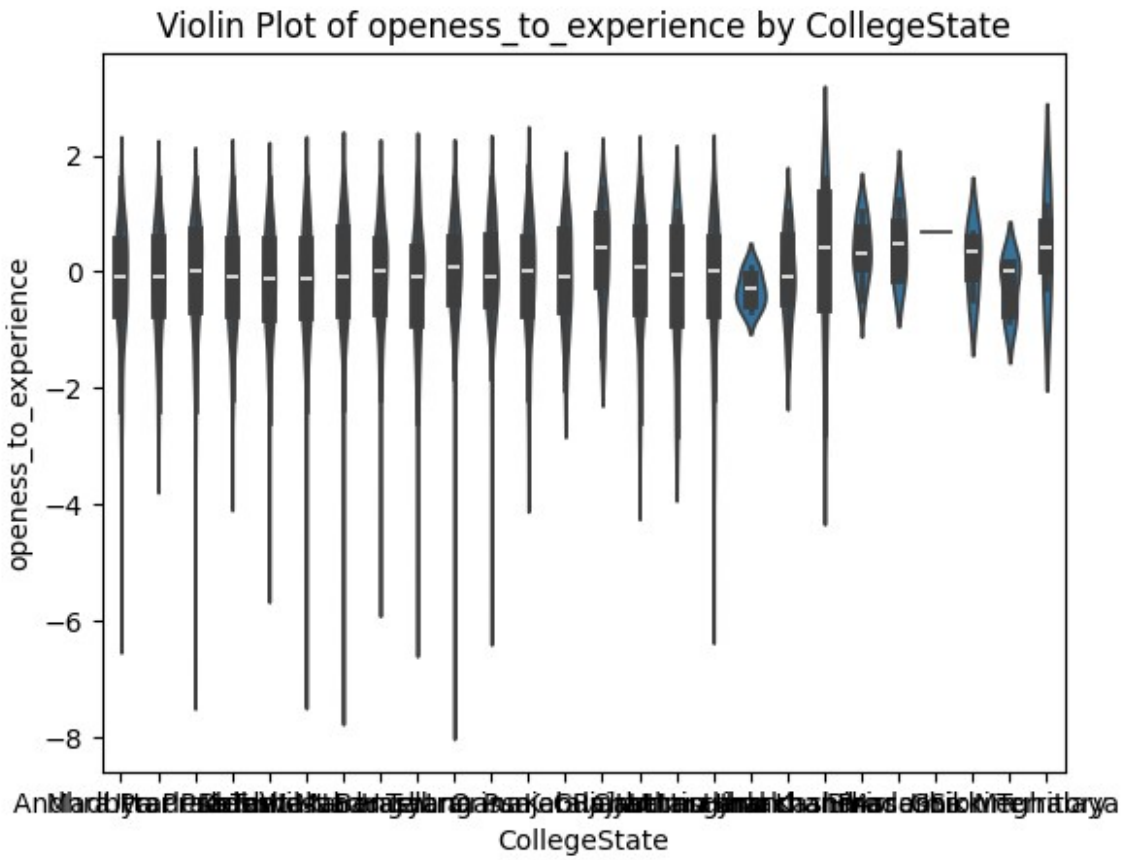

Violin Plot of ComputerScience by CollegeState
ComputerScience
700
600
500
400
300
200
100
CollegeState

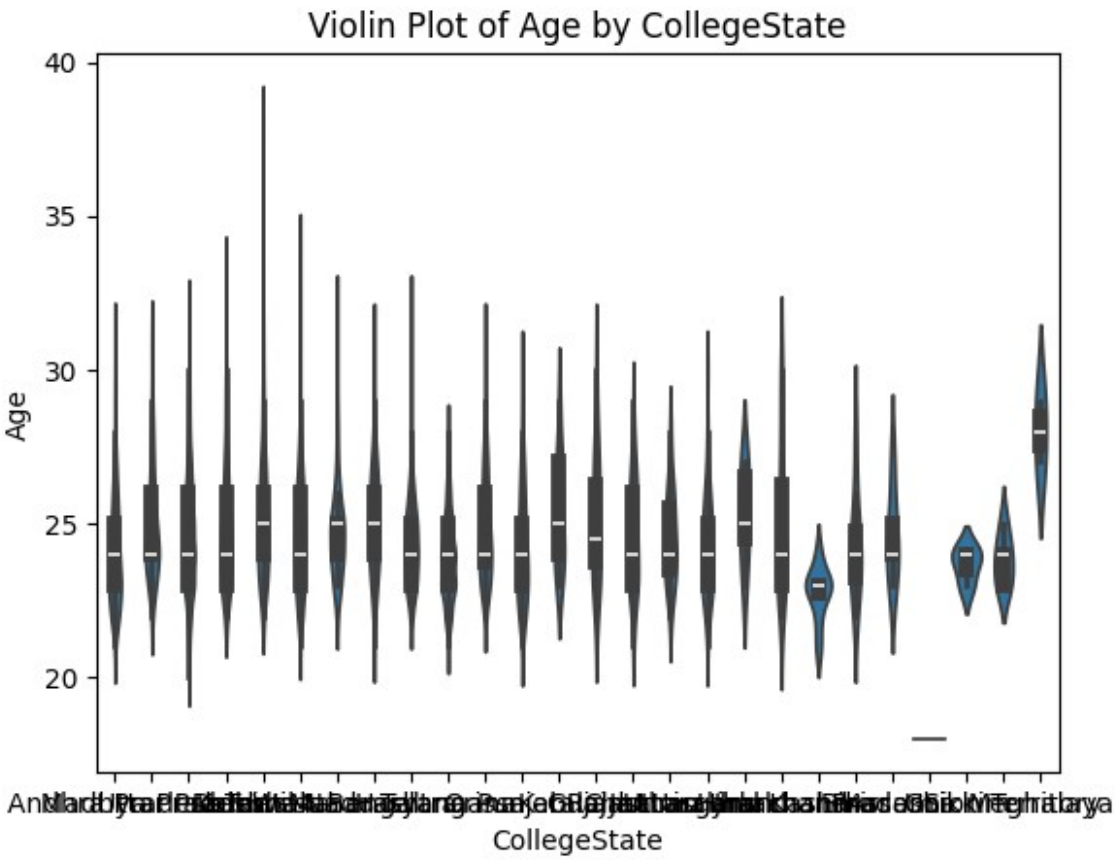

Violin Plot of conscientiousness by CollegeState
conscientiousness
3
2
1
0
−1
−2
−3
−4
−5
CollegeState


Violin Plot of agreeableness by CollegeState
agreeableness
2
0
−2
−4
−6
CollegeState


Violin Plot of extraversion by CollegeState
extraversion
4
2
0
−2
−4
CollegeState


Violin Plot of nueroticism by CollegeState
nueroticism
4
2
0
−2
−4
CollegeState


Violin Plot of openess_to_experience by CollegeState
openess_to_experience
2
0
−2
−4
−6
−8
CollegeState




```
# prompt: conclusion from above

**Conclusion:**

The data cleaning and univariate analysis revealed several insights
about the distribution of the numerical variables in the dataset.
There are some outliers present in the data, particularly for salary
and workex. The correlation matrix and heatmap showed that there are
several positive correlations between the numerical variables,
suggesting that these variables may be related to each other and could
potentially be used together in a predictive model. The scatter plots,
box plots, and violin plots provided further visual confirmation of
these relationships and also showed how the categorical variables
relate to the numerical variables.

Overall, the analysis suggests that there is valuable information in
the dataset that could be used to build a predictive model. However,
further data cleaning and feature engineering may be necessary before
building the model.
```